



Utilising tomato seed sales data for strategic decision-making in a South African firm: A forecasting approach

J Viviers

 [rcid.org/0009-0001-7566-1829](https://orcid.org/0009-0001-7566-1829)

Mini-dissertation accepted in partial fulfilment of the requirements for the degree *Master of Business Administration* at the North-West University

Supervisor: Prof JN Lekunze

Graduation: June 2025

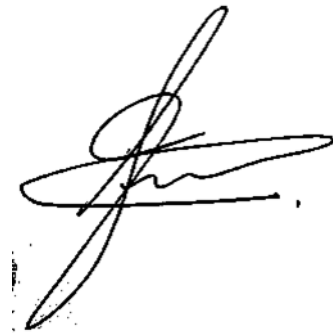
DECLARATION

With this declaration, I declare that the mini-dissertation submitted in partial fulfilment of the requirements for the Master of Business Administration degree at the North-West University is my original work. It has been text-edited by professional communication standards and has not been previously submitted to any other institution for evaluation purposes.

Johnny Viviers

21617503

03/11/2024

A handwritten signature in black ink, appearing to be 'Johnny Viviers', with a large, stylized initial 'J' and a horizontal line extending to the right.

ACKNOWLEDGEMENTS

I sincerely thank several individuals and institutions who have supported me throughout my journey to attain my Master of Business Administration (MBA) degree. First and foremost, I would like to express my deepest appreciation to Prof JN Lekunze for his guidance, insightful feedback, and continuous encouragement. His expertise, patience, and mentorship have been pivotal in shaping the direction of this dissertation, and I am truly grateful for the knowledge and inspiration he has provided.

To my parents, thank you for being my steadfast pillar of support. Your patience, understanding, and motivation have been invaluable during this demanding journey. To my wife, Cindy, thank you for your love and encouragement and for always believing in my ability to see this through. Your support, especially given your dedication to academia, has been a constant source of strength and motivation.

I would also like to thank my colleagues and employer for their support and understanding over the past few years. Balancing professional responsibilities while pursuing an MBA has been challenging, and I am grateful for the flexibility and encouragement that enabled me to dedicate time and energy to this endeavour. To the incredible work environment that fostered my growth, thank you for inspiring me to apply what I have learned and further deepen my understanding of the agricultural sector.

Finally, I extend my heartfelt gratitude to everyone who contributed in one way or another to this journey through words of encouragement, academic discussions, or simply offering a listening ear. This achievement is as much yours as it is mine.

James 1:17: *"Every good and perfect gift is from above, coming down from the Father of the heavenly lights, who does not change like shifting shadows."*

ABSTRACT

The mini-dissertation examines using tomato seed sales data as a forecasting tool to enhance strategic management decision-making within a South African seed company. This research is underpinned by four hypotheses, focusing on the influence of price, production costs, production volume, and weather conditions on tomato seed sales. Using Random Forest Regression (RFR) as the primary forecasting model, this study leverages historical data from 2013 to 2022 and incorporates climatic variables to improve the predictive capabilities of sales forecasts. The data-driven approach is supported by a comprehensive literature review covering both traditional forecasting methods and the application of Machine Learning (ML) in agricultural forecasting. The research also employs descriptive statistics, correlation analyses, and regression models to evaluate the relationships between production factors and sales outcomes. Key findings indicate that all four variables significantly impact seed sales, with weather conditions and production costs playing particularly crucial roles in shaping sales performance. The forecasting model's results suggest that improved accuracy in predicting these variables can enhance the company's ability to plan production and manage inventory, leading to better cost management and increased profitability. The managerial implications of this study include recommendations for differentiated pricing strategies, investment in climate-resilient technologies, and adopting predictive analytics to align production with market demand better. The study also highlights the need for expanding geographic coverage to improve yield optimisation and reduce production risks. Limitations include the reliance on secondary data and the scope restricted to specific geographic regions, which may limit the generalisability of the findings. This research contributes to the broader field of agricultural sales forecasting by demonstrating the effectiveness of ML models, particularly RFR, in improving the accuracy of demand predictions in a complex and dynamic market. The insights provided are intended to support seed companies in making informed, strategic decisions to ensure profitability and sustainability in a highly competitive industry.

Keywords: Machine learning (ML); Random Forest Regression (RFR). Predictive modelling; Strategic planning; Sales optimisation; Market dynamics; Economic modelling; Data-driven insights.

LIST OF ABBREVIATIONS

CDS	Climate Data Store
CEA	controlled-environment agriculture
C3S	Copernicus Climate Change Service
CSR	Corporate social responsibility
EDA	Exploratory data analysis
IDT	Innovation Diffusion Theory
IoT	Internet of Things
LSTM	Long Short-Term Memory
ML	machine learning
MAE	Mean Absolute Error
NSE	Nash-Sutcliffe Efficiency Coefficient
RDF	Random Forest Regression
R&D	research and development
RMSE	Root Mean Square Error
SANSOR	South African National Seed Organisation
Stats SA	Statistics South Africa
SVM	Support Vector Machines

TABLE OF CONTENTS

DECLARATION	I
ACKNOWLEDGEMENTS	II
ABSTRACT	III
LIST OF ABBREVIATIONS	IV
CHAPTER 1 OVERVIEW OF THE STUDY	1
1.1 Introduction.....	1
1.1.1 Contextualisation of the study	2
1.2 Background of the study.....	3
1.3 Problem statement	4
1.4 Hypothesis	5
1.5 Research aims and objectives.....	6
1.6 Motivation for the study	7
1.7 Scope and delimitation of the study.....	8
1.8 Significance of the study	9
1.9 Research design and methodology.....	10
1.10 Theoretical framework.....	12
1.11 Definition of key concepts	13
1.12 Chapter summary	16
CHAPTER 2 OVERVIEW OF THE LITERATURE	17
2.1 Introduction.....	17
2.1.1 Contextualisation of hybrid of tomatoes seed production in South Africa	17

2.2	Influence of price on tomato seed sales in strategic management decision-making	19
2.3	Impact of production costs on tomato seed sales in strategic management decision-making.....	21
2.4	Effects of production quantity on tomato seed sales in strategic management decision-making.....	23
2.5	Role of weather on tomato seed sales and its impact on strategic management decision-making.....	25
2.6	International business management and administration practices	27
2.7	National business management and administration practices	29
2.8	Theoretical underpinning and conceptualisation in the study	32
2.8.1	Evaluation of the theoretical framework	32
2.8.1.1	Resource-Based View	32
2.8.1.2	Porter's Competitive Advantage Theory.....	33
2.8.1.3	Innovation Diffusion Theory	33
2.8.1.4	Institutional Theory	34
2.8.1.5	Contingency Theory.....	34
2.8.2	Selection of theories for the study	34
2.8.2.1	Justification for RBV	35
2.8.2.2	Justification for Contingency Theory	35
2.8.3	Development of the chosen theories.....	35
2.9	Chapter summary	36
CHAPTER 3	RESEARCH METHODOLOGY AND METHOD.....	38
3.1	Introduction.....	38

3.1.1	Research philosophy	39
3.1.1.1	Positivism	39
3.1.1.2	Saunders' research onion	39
3.2	Research paradigm and design.....	41
3.3	Data collection and integration of global datasets.....	42
3.4	Data quality check	44
3.4.1	Completeness check	44
3.4.2	Consistency check.....	44
3.4.3	Accuracy check	44
3.4.4	Sensitivity analysis.....	45
3.4.5	Handling outliers.....	45
3.4.6	Statistical metrics.....	45
3.5	Data analysis and interpretation.....	46
3.5.1	Descriptive statistics	46
3.5.2	Correlation and regression analysis.....	47
3.5.2.1	Pearson correlation coefficient.....	47
3.5.2.2	R-squared analysis	48
3.5.2.3	Suitability tests.....	48
3.5.3	Machine Learning model.....	48
3.5.3.1	Unit of measurement	49
3.5.3.2	Why Random Forest Regression is suitable	49
3.5.3.3	Model strengths and features	49

3.5.3.4	Suitability tests.....	49
3.5.4	Sensitivity analysis.....	50
3.5.5	Model validation and performance metrics.....	50
3.5.6	Interpretation of results	51
3.6	Descriptive analysis	52
3.7	Multivariate analysis.....	53
3.7.1	Correlation and regression analysis	53
3.7.2	Random Forest Regression Machine Learning model	54
3.7.3	Validation and performance evaluation	54
3.8	Correlation coefficient.....	54
3.8.1	Correlation with weather data	55
3.8.2	Correlation with production costs and marketing expenditures.....	55
3.8.3	Assessing statistical significance	56
3.9	Validity and reliability	56
3.9.1	Validity.....	56
3.9.2	Reliability	57
3.9.3	Random Forest Regression model validation.....	58
3.10	Chapter summary	58
CHAPTER 4	RESULTS AND FINDINGS	60
4.1	Introduction.....	60
4.2	Data description and contextual background.....	61
4.2.1	Overview of data used	61

4.2.2	Study area and production geography	62
4.2.3	Climatic conditions and their impact on production	64
4.3	Analysis of production contracts and climatic impact	66
4.3.1	Production contracts across regions	66
4.3.2	Climatic influence on production outcomes	68
4.4	Production outcomes and statistical correlation	69
4.4.1	Production trends for cultivars X and Y	69
4.4.2	Correlation analysis with climatic variables	74
4.5	Costing, pricing, and profitability analysis	75
4.5.1	Impact of production cost on profitability	75
4.5.2	Price sales relationship analysis	76
4.5.3	Managerial implications	77
4.6	Production geography and climatic impact on yield.....	82
4.6.1	Geographic distribution of production.....	82
4.6.2	Climatic impact on yield	82
4.6.3	Managerial implications of climatic and geographic factors	83
4.7	Random Forest Regression model.....	84
4.7.1	Model description.....	84
4.7.2	Model design	86
4.7.3	Variable importance.....	87
4.8	Training and testing	88
4.8.1	Model accuracy	88

4.8.2	Forecasting.....	92
4.9	Managerial implications for forecasted trends	94
4.9.1	Forecasting production efficiency and cost implications	94
4.9.2	Price sensitivity and sales volume predictions	95
4.9.3	Strategic considerations for market expansion.....	96
4.9.4	Enhancing long-term profitability and sustainability	96
4.10	Managerial implications	97
4.10.1	Price and sales relationship (Hypothesis 1)	97
4.10.2	Production costs and profitability (Hypothesis 2).....	97
4.10.3	Quantity produced and sales impact (Hypothesis 3)	98
4.10.4	Climatic variability and sales outcomes (Hypothesis 4).....	98
4.10.5	Summary of strategic actions.....	98
4.11	Chapter summary	99
 CHAPTER 5 CONCLUSIONS RECOMMENDATIONS AND FUTURE RESEARCH		103
5.1	Introduction.....	103
5.2	Summary of key findings	103
5.3	Practical recommendations	104
5.3.1	Pricing strategies for optimal sales	104
5.3.2	Cost management through technological investments	105
5.3.3	Geographic expansion for yield optimisation.....	105
5.3.4	Climate adaptation and risk management.....	106
5.3.5	Leveraging predictive analytics for decision-making	106

5.4	Limitations of the study	106
5.5	Future studies.....	107
5.6	Chapter summary	108
REFERENCES.....		109
APPENDIX.....		120

LIST OF TABLES

Table 1: Summary of training and testing accuracy for the sales and profit RFR models.	88
--	----

LIST OF FIGURES

Figure 1: A diagrammatic illustration of the main research objectives.	7
Figure 2: A schematic representation of the methodological process.	12
Figure 3: A diagram indicating the two selected theories applied in the study.	36
Figure 4: Saunders' research onion highlighting research philosophy, approach and research strategy (Saunders <i>et al.</i> 2016).....	40
Figure 5: Cultivar X cumulative unit contract value awarded to each institution within each country.	62
Figure 6: Cultivar Y cumulative unit contract value awarded to each institution within each country.	62
Figure 7: Countries within which the respective cultivar seeds are predominantly increased.	64
Figure 8: Annual precipitation per major/nearby city for each country with increased tomato seeds.	65
Figure 9: Remotely sensed (ERA-5 product) 2012 to 2022 mean monthly precipitation and surface (2m) temperature.	66
Figure 10: Cultivar X production data correlation analyses (across all countries).....	70
Figure 11: Cultivar Y production data correlation analyses (across all countries).....	71
Figure 12: Cultivar X production variables correlation analyses by country.	72
Figure 13: Cultivar Y production variables correlation analyses by country.	73
Figure 14: Cultivar X annual budget and produced units, and what percentage of the cleaned seeds comprised of the cumulative seeds produced, as well as the linear regression model-based budget unit forecast for 2024 - 2028.....	78

Figure 15: Cultivar Y annual budget and produced units, and what percentage of the cleaned seeds comprised of the cumulative seeds produced, as well as the linear regression model-based budget unit forecast for 2024 - 2028.....	78
Figure 16: Cultivar X correlations between the production, inset costs and retail price, independent variables, and sales and profit, dependent variables, to be predicted.	79
Figure 17: Cultivar Y correlations between the production, inset costs and retail price, independent variables, and sales and profit, dependent variables, to be predicted.	80
Figure 18: Cultivar X annual (2013-2022) production budget (in units), as well as inset costs, sale, and profit value, as well as price per 1,000 seeds in South African (ZAR) Rand.....	81
Figure 19: Cultivar Y annual (2013-2022) production budget (in units), as well as inset costs, sale, and profit value, as well as price per 1,000 seeds in South African (ZAR) Rand.....	81
Figure 20: Variable importance (%) for the trained RFR Model predictions.....	88
Figure 21: RFR model training (a) and testing (b) Cultivar X actual against predicted sales data.....	90
Figure 22: RFR model training (a) and testing (b) Cultivar X actual against predicted profit data.	91
Figure 23: RFR model training (a) and testing (b) Cultivar Y actual against predicted sales data.....	91
Figure 24: RFR model training (a) and testing (b) Cultivar Y actual against predicted profit data.	92
Figure 25: Cultivar X predicted/estimated 2024-2027 sales and profit based on linear regression extrapolated unit budget, inset costs and price variables.	93

Figure 26: Cultivar Y predicted/estimated 2024-2027 sales and profit based on linear regression extrapolated unit budget, inset costs and price variables. 94

CHAPTER 1 OVERVIEW OF THE STUDY

1.1 Introduction

The purpose of this study is to leverage machine learning models to forecast tomato seed sales, enabling strategic decision-making for a South African firm. Sales data serves as the primary output variable in this process, providing the foundational insights needed for model development and evaluation. Despite advances in seed production and forecasting techniques, accurately predicting seed sales and optimising production management remain significant challenges for managers. The agricultural seed industry plays a critical role in sustaining food security and promoting efficient agricultural practices. Factors such as pricing, production costs, climatic conditions, and geographic variations interact in complex ways to influence sales, profitability, and production efficiency. By utilising advanced machine learning models, this study aims to comprehensively understand these interdependencies, addressing current limitations in forecasting within the seed market context and supporting more informed managerial decisions.

The study employs machine learning models to forecast tomato seed sales, providing a strategic tool for decision-making in a South African firm. Sales forecasting extends beyond mere product prediction and serves as a critical component in strategic management by facilitating more effective planning and resource allocation. According to Singh *et al.* (2020), sales forecasting plays a pivotal role in decision-making within business organisations, enabling managers to anticipate future demand, optimise inventory levels, and plan workforce requirements efficiently. It provides a forward-looking perspective that supports financial planning and ensures the organisation's financial stability. Hasselbeck *et al.* (2022) further highlight that sales forecasting allows managers to respond proactively to changing market indicators, enabling timely adjustments to procurement, production schedules, and marketing strategies. This adaptability not only enhances revenue generation by aligning supply with demand but also minimises production costs by reducing wastage and overstocking. Furthermore, sales forecasting is indispensable for industries dealing with perishable goods, such as agricultural products, where precise demand estimation can prevent spoilage and maximise profitability (Hasselbeck *et al.*, 2022).

Sales forecasting encompasses various dimensions, including market trend analysis, customer behaviour prediction, seasonal demand variations, and competitive dynamics. By integrating these factors, forecasting models provide a holistic framework for strategic decision-making, helping organisations navigate uncertainties and maintain a competitive

edge. In the context of tomato seed sales, forecasting enables managers to align production with projected demand, optimise pricing strategies, and plan distribution efficiently, ensuring that resources are utilised effectively and market opportunities are maximised. This comprehensive approach to sales forecasting not only supports operational efficiency but also contributes to long-term sustainability and growth.

1.1.1 Contextualisation of the study

In the firm under investigation, hybrid tomato seeds are produced through the genetic cross of two parent inbred plants (Wijnker *et al.*, 2012). This hybrid seed is produced by growing the two-parent inbreds in alternating rows within the same glasshouse. During this process, the flowers from one parent, termed the female plant, are meticulously removed to ensure cross-pollination only occurs through a labour-intensive operation by the other parent, the male plant. The tomatoes that mature on the female plant result from the genetic cross of the two parents (Burer *et al.*, 2008; Liberatore *et al.*, 2016). After maturation, the tomatoes are harvested, and the seeds are sent to processing plants, where they undergo drying, sorting, treatment (with antifungal or other protective coatings), and storage for the upcoming selling season.

Consequently, the production to meet the demand for tomato seeds for the 2024 growing season occurs in 2023 (or earlier), coinciding with one of two growing cycles. The seed company must thoroughly analyse its inventory on hand, evaluate the cost of production during the 2023 season, and consider uncertainties such as demand for specific hybrids, average yields, and hybrid preferences for the 2024 season. Using tomato seed sales data as a forecasting tool becomes critical for strategic decision-making within the seed industry. As Moon *et al.* (2003) and Duus (2016) identified, forecasting is pivotal in business planning and overall management. Duus (2016) further underscores that irrespective of industry, effective demand forecasting enables organisations to seize market opportunities, strengthen channel relationships, enhance customer satisfaction, reduce inventory costs, eliminate product obsolescence, optimise distribution, and anticipate financial and capital needs.

Forecasting is often complex, particularly when dealing with agricultural products like seed sales (Ahamed *et al.*, 2023). Ahamed *et al.* (2023) highlighted that seed sales are influenced by various factors, including weather, public holidays, farmers' preferences, seed production, market competition, and seasonal seed availability. As Araujo and Guglianone (2023) noted, accurate forecasting is difficult because it demands an approach that balances the inclusion of relevant variables while excluding those less impactful.

To address this complexity, contemporary forecasting models have integrated machine learning (ML) methods, which can identify nonlinear patterns in data that elude traditional linear models. As a result, these models provide a compelling alternative to standard econometric approaches (Araujo and Guglianone, 2023). Statistical models applied to long-term data have proven highly accurate (Elalem *et al.*, 2022). However, forecasting for tomato seed sales presents further challenges due to the ongoing development of new varieties with unique traits. Introducing new products is particularly difficult, as historical data is often insufficient, limiting the effectiveness of standard statistical forecasting techniques (Elalem *et al.*, 2022).

Thus, dominant forecasting methods for new products often rely on market research, managerial opinions, and input from the sales force (Elalem *et al.*, 2022). While these qualitative approaches are robust for forecasting sales of new products, quantitative methods have the potential to significantly outperform them due to their ability to reduce biases inherent in human judgment.

1.2 Background of the study

No sales forecasting model or predictive system is currently utilised at the selected seed company under investigation. In a previous study, Zhu *et al.* (2019) argued that seed forecasting typically relies on a year-to-year budgeting approach, focusing on the prior season's demand to anticipate the following year's variety requirements. However, there is limited research on product forecasting within the hybrid seed production segment of the agricultural industry. While forecasting is crucial for the supply chain of seed companies, accurately estimating future demand, particularly in the vegetable seed sector, remains a challenge. Companies must effectively meet forecasted demand and manage inventory and associated costs (Kumar *et al.*, 2020).

Bannikova *et al.* (2018) pointed out that farmers often choose the variety they will plant based on factors from the previous season, such as soil-borne diseases, insect infestations, viral disease pressures, or changes in crop rotation schedules. Although hybrids are specifically bred to address these issues, seed availability can become a bottleneck. The production of hybrid seeds is a complex process that spans multiple seasons, forcing seed companies to rely on inventory from previous seasons to meet current demands (Bannikova *et al.*, 2018).

Ensafi *et al.* (2022) state that sales forecasting is vital in managing agricultural firms, particularly when scheduling hybrid seed production over six, 12, or 18-month cycles. These production cycles ensure adequate seed availability while minimising impacts on cash flow

(Zhu *et al.*, 2019; Kumar *et al.*, 2020). They also help schedule the production of other hybrids, such as sweet corn or green beans, which are less costly than tomatoes. Seed companies face higher production costs for hybrids, impacting cash flow and necessitating precise forecasting to balance expenses (Ensafi *et al.*, 2022).

Accurate sales forecasting supports management decisions regarding production scheduling, costs, yields, and pricing, influencing market share and customer satisfaction. By producing the right number of forecasted seeds, companies can avoid overproduction (associated storage costs) and seed loss due to changing farmer preferences. Ahamed *et al.* (2023), Ensafi *et al.* (2022), Hasselbeck *et al.* (2022), and Araujo and Guglianone (2023) have all applied ML models in business management, emphasising that effective sales forecasting helps curb price instabilities and improve planning.

As the population in South Africa grows and agricultural land becomes scarcer, the demand for vegetable seeds and new varieties is projected to increase (Reynoso and Intal, 2022). Despite the growing importance of the vegetable seed industry, there remains a significant lack of studies focused on product forecasting in agriculture (Reynoso and Intal, 2022). Sales forecasting is a critical component of the seed company's supply chain, as lost sales result in high opportunity costs and allow competitors to capture additional market share (Comhaire and Papier, 2015).

1.3 Problem statement

South Africa has a population of over 64.1 million people, according to Statistics South Africa (Stats SA) (2024a), with projections suggesting it could grow to 72.5 million by 2050. This growing population will drive an increased demand for food, putting additional pressure on agricultural land and water resources (Frona *et al.*, 2019). Stats SA (2024b) also reported a 1.4% rise in overall vegetable market demand between the 2020/2021 and 2021/2022 financial years, while tomato production declined by 1.5% over the same period. This reduction in tomato plantations was mainly attributed to a 4% price decrease, from R8,619.14/ton in 2020/2021 to R8,276.66/ton in 2021/2022, followed by an additional 11.5% decrease to R7,324.84/ton in the first quarter of 2022/2023 (Stats SA, 2024c). However, tomato prices rebounded with a 35% increase in the fourth quarter of 2023, reaching R9,888.53/ton.

The South African National Seed Organisation (SANSOR) (2020) reported that the vegetable seed industry in South Africa grew by 45% between 2015 and 2020. Although this growth stagnated during the COVID-19 pandemic, Stats SA (2024c) noted a recent 0.8% recovery post-pandemic. Despite this growth, the demand for vegetable seeds exceeds the supply

(Reynoso & Intal, 2022). The tomato seed industry alone is valued at approximately R202 million annually (SANSOR, 2023). Forecasting seed demand remains challenging due to factors such as the natural environment, weather conditions, and market dynamics, including fuel and fertiliser price fluctuations: these factors and other variables like demand and price influence accurate seed sales forecasting.

The company under investigation holds a 36% market share in tomato seed sales. Still, it faces challenges, including selecting new hybrids and lacking the proper tools for forecasting future market trends. Currently, the organisation uses no tomato seed sales forecasting tool to estimate future market volumes, sales, or prices. The study aims to bridge this gap by developing a forecasting tool to predict sales, thereby supporting management in decision-making and helping the company gain market share while navigating the evolving business environment. Additionally, this forecasting tool will aid in long-term production planning for specific tomato varieties, improving cash flow management and assisting the company in responding to fluctuations in the seed industry.

1.4 Hypothesis

The study's primary objective is to forecast tomato seed sales for a selected seed company in South Africa, enabling more informed management decision-making. The following hypotheses were developed to achieve this:

Hypothesis 1:

- **H0:** Price influences the sale of tomato seeds, thereby affecting management decision-making.
- **H1:** Price does not influence tomato seed sales or management decision-making.

Hypothesis 2:

- **H0:** The cost of production influences tomato seed sales, which affects management decision-making.
- **H1:** The cost of production does not influence tomato seed sales or management decision-making.

Hypothesis 3:

- **H0:** The quantity of tomato seeds produced influences sales, affecting management decision-making.

- **H1:** The quantity of tomato seeds produced does not influence sales or management decision-making.

Hypothesis 4:

- **H0:** Weather influences tomato seed sales, which affects management decision-making.
- **H1:** Weather does not influence tomato seed sales or management decision-making.

Tabe-Ojong *et al.* (2020) assert that price theory is crucial in determining future market conditions for agricultural businesses. The interaction of supply and demand is shaped by weather, production costs, and consumer preferences, which dictate the price at which tomato seeds are sold. For example, favourable weather conditions may result in a bountiful tomato harvest, increasing seed supply and potentially lowering prices if demand does not rise accordingly (Tabe-Ojong *et al.*, 2020). On the other hand, adverse weather could reduce supply, leading to price increases if demand remains constant (Fiankor *et al.*, 2023).

Accurate sales forecasting in such markets requires a deep understanding of these dynamics, including demand elasticity (how sensitive consumer demand is to price changes) (Tabe-Ojong *et al.*, 2020). If demand is highly elastic, even small price increases could significantly drop the quantity sold, affecting overall revenue. Additionally, businesses must consider production costs, such as labour, resources, and technology, which are critical in pricing strategies (Fiankor *et al.*, 2023). In South Africa, where agriculture holds a substantial economic role, accurate sales forecasting, guided by price theory, can help companies make well-informed decisions about production levels, marketing strategies, and resource allocation. This, in turn, leads to sustainable profitability and growth within the sector.

1.5 Research aims and objectives

The study's primary objective is to apply ML algorithms to forecast hybrid tomato seed sales for a selected seed company in South Africa, thereby enabling more informed management decision-making. The following specific objectives need to be investigated to address the primary objective, as outlined in Figure 1:

- Investigate the influence of weather, as a dichotomous variable (favourable or unfavourable), on tomato seed prices and its effect on sales, focusing on the implications for management decision-making.

- Examine the impact of weather, as a dichotomous variable, on seed production costs (whether favourable or unfavourable) and its effect on sales, assessing how this influences management decisions.
- Analyse the effect of weather, as a dichotomous variable, on the quantity of seed produced (favourable or unfavourable) and its influence on sales and how this informs management decision-making.
- Evaluate the overall influence of weather, as a dichotomous variable, on sales and how these variations affect management decision-making processes.

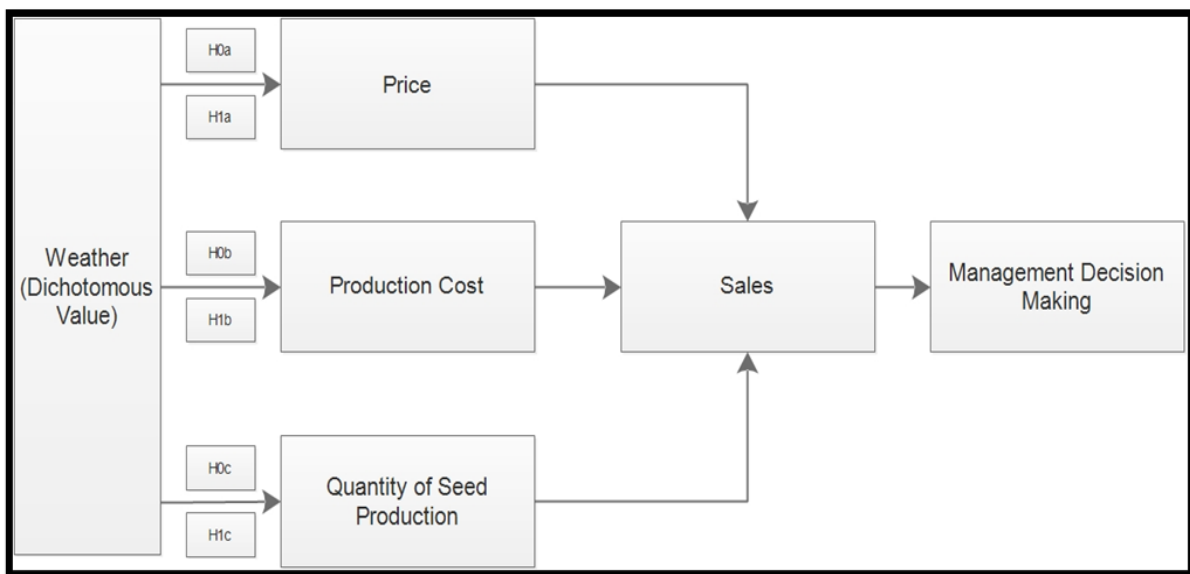


Figure 1: A diagrammatic illustration of the main research objectives.

1.6 Motivation for the study

The significance of precise sales forecasting in the agricultural sector is undeniable, particularly for companies producing and selling hybrid seeds. This study, titled **"Utilising tomato seed sales data for strategic decision-making in a South African firm: A forecasting approach"**, aims to develop an advanced forecasting model using ML algorithms to strengthen strategic decision-making within a seed company.

Effective sales forecasting brings substantial benefits, primarily by enhancing management decision-making. Singh *et al.* (2020) underscores the importance of sales forecasting, noting that it plays a critical role in predicting future sales, which helps assess the company's financial health. Moreover, efficient forecasting allows for better stock management and workforce planning, ultimately improving operational efficiency. Responding quickly to shifting market

conditions is also greatly improved when accurate forecasts are in place. Hasselbeck *et al.* (2022) highlight that reliable forecasting enables managers to adjust procurement and production plans in real time, increasing revenues and lowering production costs. This is particularly crucial in agriculture, where products like tomato seeds have limited shelf lives, making it essential to balance supply with demand to avoid wastage (Hasselbeck *et al.*, 2022).

Currently, the seed company under investigation lacks a sophisticated sales forecasting model, instead relying on a basic year-to-year budgeting system (Zhu *et al.*, 2019). The study aims to address this gap by developing a forecasting tool that accurately predicts future sales, supporting more informed management decisions. Furthermore, the study explores the influence of external factors on seed sales, such as weather conditions, market trends, and production costs. By incorporating these variables into the forecasting model, the study aims to provide a comprehensive tool that offers more accurate and actionable insights (Ahamed *et al.*, 2023; Araujo & Guglianone, 2023). The application of advanced techniques is another key driver for the study. Traditional statistical methods often fall short in forecasting complex market behaviour because they fail to capture nonlinear patterns. In contrast, ML algorithms can uncover hidden patterns in data, making them particularly effective for forecasting sales in dynamic and complex markets (Araujo and Guglianone, 2023; Elalem *et al.*, 2022).

The strategic importance of this industry cannot be overstated. The South African vegetable seed sector is on a growth trajectory, driven by rising demand due to population growth and decreasing availability of agricultural land (Reynoso and Intal, 2022). Accurate sales forecasting will enable seed companies to optimise inventory management, streamline production, and better meet market demand, giving them a competitive advantage in the industry.

1.7 Scope and delimitation of the study

The study uses tomato seed sales data to develop a robust forecasting model to enhance strategic decision-making within a South African seed company. The primary objective is the application of ML algorithms to forecast sales, providing actionable insights for management. This approach addresses several critical areas, including the impact of external factors such as weather conditions, market trends, and production costs on sales performance.

The study will specifically analyse historical sales data across multiple seasons to identify patterns and correlations that can inform future sales forecasts. This data-driven approach ensures the model is grounded in empirical evidence, improving its reliability and accuracy. Additionally, the study will explore the integration of advanced ML techniques, which have

consistently outperformed traditional statistical methods in capturing complex, nonlinear relationships within datasets (Ahamed *et al.*, 2023; Araujo and Guglianone, 2023). However, the researchers acknowledge some delimitations. First, it will focus exclusively on a single seed company within South Africa, which may limit the generalisability of the findings to other contexts or regions. While the insights gained may offer significant value, they might not fully encompass the diverse dynamics in different geographical areas or other companies. Second, the study will rely primarily on secondary data provided by the seed company. Although this data is expected to be detailed and comprehensive, using pre-existing datasets carries inherent limitations. These datasets may not always perfectly align with the research objectives, and any inaccuracies or biases in the original data could potentially influence the study's outcomes (Bryman *et al.*, 2021).

Moreover, the study will focus on ML models, particularly random forest regression, due to their proven effectiveness in managing high-dimensional data and capturing intricate patterns (Elalem *et al.*, 2022). While this method is well-justified based on existing literature, it does exclude other potential forecasting techniques that may provide complementary insights. Furthermore, the study will investigate the effects of weather as a dichotomous variable on key factors such as price, seed production cost, and the quantity produced. This focus is critical for understanding direct impacts on sales, but it may overlook other nuanced environmental or socio-economic factors that could also significantly influence sales performance (Fiankor *et al.*, 2023). While the study aims to deliver a comprehensive and actionable forecasting model for tomato seed sales in a South African seed company, it is constrained by its specific context, reliance on secondary data, and the selected methodological approach. These limitations are necessary to maintain the study's focus and feasibility and suggest areas for future research and refinement.

1.8 Significance of the study

The significance of the study lies in its potential to revolutionise strategic decision-making within the agricultural sector, particularly in the hybrid seed industry in South Africa. The study addresses a critical gap in current business practices by developing a robust sales forecasting model using advanced ML algorithms. Accurate sales forecasting is essential for enhancing both operational efficiency and financial performance. As Singh *et al.* (2020) emphasise, effective sales predictions allow companies to optimise inventory levels, reduce wastage, and manage workforce allocation more efficiently. The study's forecasting model aims to provide precise sales predictions, enabling the seed company to maintain optimal stock levels while minimising unnecessary costs associated with overproduction or stockouts.

Adding ML techniques adds a new layer of innovation to agricultural sales forecasting. Traditional statistical methods often fail to capture sales data's inherent complexity and nonlinearity. In contrast, ML models, such as Random Forest Regression (RFR), are well-equipped to identify hidden patterns and make more accurate predictions (Ahamed *et al.*, 2023; Araujo and Guglianone, 2023). This advancement not only enhances the reliability of sales forecasts but also places the seed company at the forefront of technological innovation within the industry. Moreover, the study provides valuable insights into external factors such as weather conditions, market trends, and production costs that influence seed sales. Understanding these variables is crucial for strategic planning and risk management. By integrating these elements into the forecasting model, the study offers a comprehensive tool that helps the seed company anticipate market fluctuations and make informed decisions (Elalem *et al.*, 2022; Fiankor *et al.*, 2023).

The study also contributes to the broader academic discourse on agricultural sales forecasting. Given the limited research in this area, particularly in the context of developing countries like South Africa, the study's findings fill a significant knowledge gap and provide a foundation for future research. The innovative approach and methodologies applied can serve as a benchmark for other researchers and practitioners looking to improve sales forecasting in the agricultural sector (Reynoso and Intal, 2022). Furthermore, the study's practical implications extend beyond the seed company itself. Accurate sales forecasting models can enhance the agricultural industry's overall performance by improving supply chain coordination, reducing inefficiencies, and ensuring better alignment between supply and demand (Hasselbeck *et al.*, 2022). This alignment is critical in agriculture, where products are perishable, and market conditions can be highly volatile. The significance of the study is multifaceted. It aims to enhance strategic decision-making, introduce advanced forecasting techniques, provide critical insights into market dynamics, contribute to academic knowledge, and improve industry performance. The study benefits the seed company by addressing these key areas and offers valuable contributions to the broader agricultural sector.

1.9 Research design and methodology

The study employs a comprehensive research design and methodology to develop a robust forecasting model for tomato seed sales using advanced ML algorithms. The study is structured to achieve key objectives, including enhancing strategic decision-making and improving operational efficiency within a South African seed company. The study utilises historical sales data to build and validate the forecasting model grounded in a quantitative approach. The primary data source consists of secondary data provided by the seed company, which includes detailed sales records spanning multiple seasons. Additionally, external factors

such as weather conditions, market trends, and production costs are incorporated to ensure accurate forecasting (Ahamed *et al.*, 2023).

The methodology is divided into eight key phases, as illustrated in **Error! Reference source not found.:**

Data collection and preprocessing: The first phase involves gathering historical sales data and relevant external variables. The data is then pre-processed to address missing values, outliers, and normalisation. These preprocessing steps ensure the quality of the data and its suitability for ML models (Miguéis *et al.*, 2022).

Exploratory data analysis (EDA): In this phase, an in-depth EDA is conducted to uncover patterns and relationships within the data. EDA techniques, including statistical summaries, visualisations, and correlation analysis, identify key predictors and insights that inform the model development process (Pereira *et al.*, 2015).

Feature selection and engineering: Following the EDA, relevant features are selected and engineered to enhance the model's predictive power. This includes creating new variables, such as lagged sales figures, seasonal indicators, and interaction terms, to capture the dynamic nature of the sales data (García *et al.*, 2016).

Model development: This core phase involves developing various ML models. The study evaluates multiple algorithms, including RFR, Support Vector Machines (SVM), Long Short-Term Memory (LSTM) networks, and Feedforward Neural Networks. These models are selected for their ability to manage complex, nonlinear relationships in time series data (Miguéis *et al.*, 2022; Alon *et al.*, 2001).

Model training and hyperparameter tuning: Each ML model is trained on the pre-processed data. Hyperparameter tuning uses grid search and cross-validation methods to optimise the models' performance. This ensures the models avoid overfitting and generalise well to new data (Huber and Stuckenschmidt, 2020).

Model evaluation: The models use performance metrics such as Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and Mean Positive Error. These metrics provide a comprehensive assessment of the models' accuracy and reliability in forecasting sales (Miguéis *et al.*, 2022).

Validation and testing: A forward validation scheme, or time series cross-validation, is employed to validate the models. This method involves iteratively training and testing the

models on different subsets of data to ensure robustness. The final models are tested on a holdout dataset to evaluate their real-world performance (Schnaubelt, 2019).

Implementation and deployment: The best-performing model is selected for implementation based on the evaluation results. Recommendations for integrating the model into the company's decision-making processes include continuous monitoring and retraining guidelines to maintain accuracy over time (Pereira *et al.*, 2015).

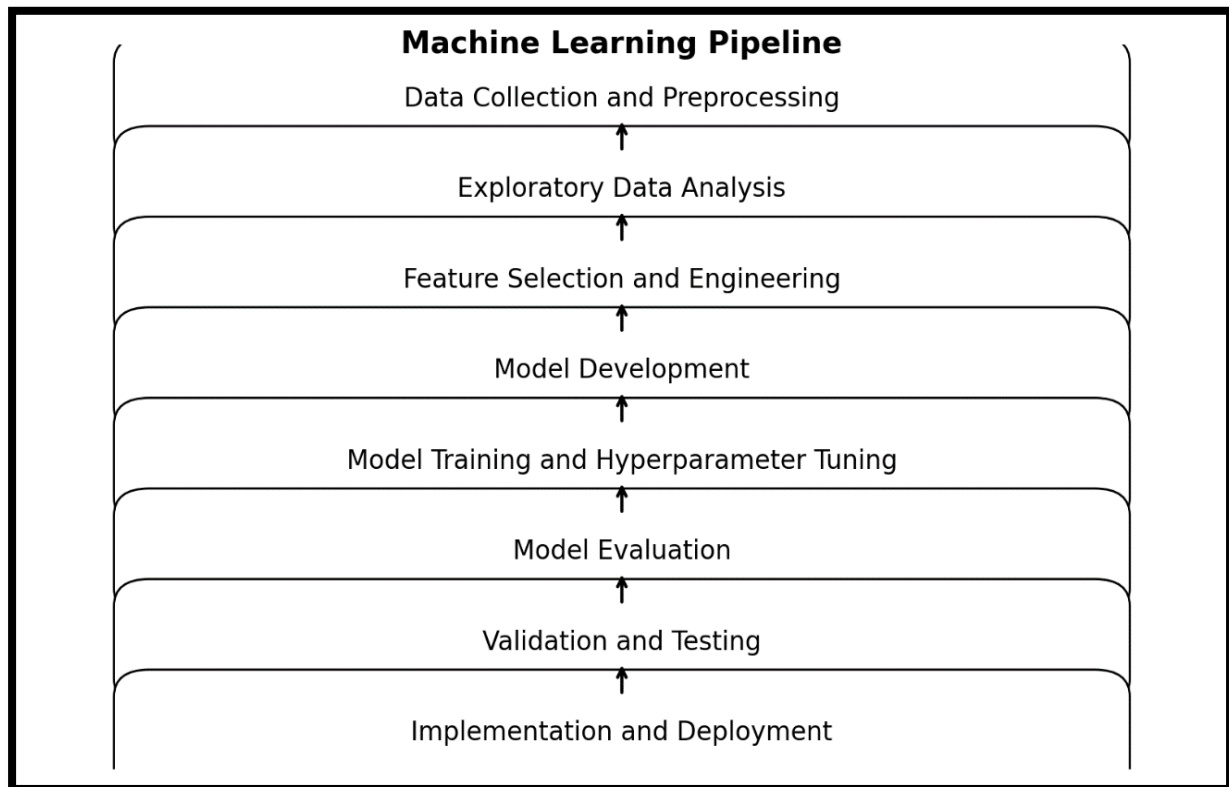


Figure 2: A schematic representation of the methodological process.

This research design and methodology offer a structured and systematic approach to developing a dependable sales forecasting model. By leveraging the power of ML, the seed company can address its unique challenges effectively. The study aims to deliver a tool that significantly improves strategic planning and operational efficiency by combining advanced analytical techniques with practical implementation steps. Beyond the numbers, this approach equips the sales team with valuable insights, empowering them to make smarter, more informed decisions in their day-to-day operations.

1.10 Theoretical framework

The theoretical framework for the study integrates contemporary theories and concepts from strategic management, sales forecasting, and ML. This framework provides a structured

approach to examining the research problem and supports the development of the sales forecasting model. The firm's Resource-based view (RBV) is at the core of this framework, which posits that a company's competitive advantage stems from its ability to manage valuable, rare, inimitable, and non-substitutable resources (Barney, 2012). In this context, accurate sales forecasting is a strategic resource that can significantly enhance operational efficiency and decision-making. By leveraging advanced ML algorithms, the firm can improve future sales predictions, optimise inventory management, and enhance overall resource allocation.

Complementing the RBV is **Porter's Theory of Competitive Advantage through Innovation**, which emphasises the role of technological advancements in maintaining and strengthening a firm's market position (Porter, 2011). The study aligns with this theory by incorporating cutting-edge ML techniques to develop a sophisticated sales forecasting model. Not only does this innovative approach improve forecasting accuracy, but it also positions the seed company as a leader in adopting advanced technologies within the agricultural sector.

Furthermore, the study draws upon the principles of **Time Series Analysis**, a fundamental concept in statistical forecasting. Time Series Analysis examines data points collected or recorded at specific intervals to identify underlying patterns and trends (Hyndman and Athanasopoulos, 2018). By applying techniques such as autoregressive models and exponential smoothing, the study aims to capture the dynamic nature of sales data and improve the predictive performance of the forecasting model.

The theoretical framework also integrates concepts from **ML Theory**, focusing on developing algorithms that can learn from data and make accurate predictions (Domingos, 2012). ML models such as RFR, SVM, and LSTM networks are employed because they manage complex, nonlinear relationships in time series data (Breiman, 2001; Hochreiter and Schmidhuber, 2017). These models are particularly well-suited for capturing intricate patterns in sales data, thereby improving forecast accuracy and reliability.

Additionally, the study incorporates elements from the **Theory of Demand Forecasting**, acknowledging the importance of understanding and predicting consumer behaviour to align supply with demand (Makridakis *et al.*, 2020). Accurate demand forecasting enables the company to anticipate market needs, reduce wastage, and ensure product availability, contributing to more efficient resource management and strategic decision-making.

1.11 Definition of key concepts

The following key concepts are defined to ensure a clear understanding of the terms used throughout the study:

Sales forecasting: Sales forecasting means estimating future sales by analysing historical sales data, market trends, and external factors. It is a critical component in inventory management, budgeting, and strategic planning, as it allows companies to anticipate future demand and adjust their operations accordingly (Makridakis *et al.*, 2020). In the study, advanced ML models are applied to enhance the precision of sales forecasts, improving the seed company's ability to meet market demand effectively.

Innovation and knowledge transfer: Innovation involves introducing new ideas, methods, or products to improve processes and outcomes. Knowledge transfer is the process through which one unit (for example, a group, department, or division) is affected by the experience of another (Fitriana *et al.*, 2023). Innovation and knowledge transfer are vital in tomato seed sales to develop new seed varieties and improve production techniques.

Corporate social responsibility: Corporate social responsibility (CSR) is a business model that helps a company be socially accountable to itself, its stakeholders, and the public. By practising CSR, companies can be conscious of their impact on all aspects of society, including economic, social, and environmental (Ahamed *et al.*, 2023). In the seed industry, CSR initiatives might include sustainable farming practices, community engagement, and environmental conservation efforts.

Production costs: Production costs refer to the expenses incurred in creating a product. These costs include raw materials, labour, and overheads. Ensafi *et al.* (2022) highlight the significance of managing production costs to ensure profitability and competitive pricing in the seed market. Efficient management of these costs is crucial for maintaining financial health and achieving strategic objectives.

Regulatory compliance: Regulatory compliance involves adhering to laws, regulations, guidelines, and specifications relevant to business operations. Derbe *et al.* (2024) discuss the importance of regulatory compliance in ensuring seed companies operate within legal frameworks, which helps avoid legal issues and maintain market integrity.

Financial management: Financial management encompasses the strategic planning, organising, directing, and controlling of financial undertakings in an organisation. It also includes applying management principles to the financial assets of an organisation while also

playing an important part in fiscal management (Thakur, 2023). Effective financial management ensures the optimal allocation of resources, risk management, and long-term financial stability.

Technology adoption: Technology adoption refers to how individuals and businesses embrace new technologies to improve their operations and outcomes. Liu (2024) illustrates how adopting advanced forecasting technologies can enhance decision-making processes and operational efficiency in the seed industry.

Machine learning: ML is a subset of artificial intelligence that focuses on creating algorithms that allow computers to learn from data and make predictions. These models identify patterns and relationships within large datasets without requiring explicit programming for every decision (Domingos, 2012). The study leverages ML techniques such as RFR, SVM, and LSTM networks to enhance sales forecasting accuracy and decision-making in the seed company.

Resource-based view: The RBV is a theory of strategic management that posits a firm's competitive advantage comes from its ability to manage resources that are valuable, rare, inimitable, and non-substitutable (Barney, 2012). The study considers accurate sales forecasting a strategic resource that enhances the seed company's operational efficiency and supports better decision-making through optimised resource allocation.

Time Series Analysis: Time Series Analysis involves examining data points collected or recorded at specific intervals to identify trends, patterns, and seasonal variations. This type of analysis is essential in sales forecasting, where past performance is used to predict future sales (Hyndman and Athanasopoulos, 2018). The study uses time series techniques, including autoregressive models and exponential smoothing, to capture the temporal dynamics of tomato seed sales.

Competitive Advantage through Innovation: This concept refers to a company's ability to gain and maintain a superior market position by adopting new technologies and processes (Porter, 2011). In the study, integrating advanced ML algorithms into the sales forecasting process represents an innovative approach that enhances the seed company's competitive edge in the agricultural sector.

Demand forecasting: Demand forecasting predicts future consumer demand for a product or service by analysing past data, market trends, and external factors (Makridakis *et al.*, 2020). Accurate demand forecasting enables businesses to align production and inventory with market demand, reducing wastage and ensuring the availability of products when needed.

The study uses demand forecasting to help the seed company meet its market requirements better.

1.12 Chapter summary

This chapter laid the essential groundwork for the study *"Utilising Tomato Seed Sales Data for Strategic Decision-Making in a South African Firm: A Forecasting Approach"*. The chapter highlighted the critical need for accurate sales forecasting to enhance operational efficiency and strategic decision-making in the hybrid seed industry by introducing the scope, significance, and methodology. The motivation behind the study was established by focusing on how advanced forecasting models, driven by ML, can provide the seed company with a competitive edge. The scope and delimitations were defined, emphasising the focus on leveraging cutting-edge ML techniques to develop a reliable forecasting model tailored to the company's needs. While acknowledging the constraints and context-specifics of the South African seed industry, the chapter provided a clear direction for the study's objectives.

A comprehensive theoretical framework was outlined, integrating key concepts from strategic management, sales forecasting, and ML. The RBV, Competitive Advantage through Innovation, Time Series Analysis, ML Theory, and the Theory of Demand Forecasting were all introduced. These theories guide the study and underline its strategic importance, setting the stage for developing a robust forecasting model. Furthermore, the chapter presented the research design and methodology in detail, with a structured approach encompassing data collection, EDA, feature engineering, model development, training, and evaluation. This systematic approach ensures the creation of an accurate and reliable forecasting tool designed to meet the seed company's operational demands.

Key concepts such as sales forecasting, ML, RBV, and Time Series Analysis were defined, offering clarity and establishing a common understanding of the methodologies and theoretical foundations of the study. These concepts framed the study and signified its relevance to agricultural strategic management.

By integrating advanced ML techniques with practical business strategies, the study aims to deliver a transformative tool that enhances strategic planning and operational efficiency within the seed company. This foundation paves the way for the following chapters, particularly Chapter 2, which will critically review the relevant literature. By exploring existing research on sales forecasting and ML applications, the next chapter will identify gaps and opportunities, strengthening the case for the innovative approach taken in the study.

CHAPTER 2 OVERVIEW OF THE LITERATURE

2.1 Introduction

Chapter 1 of the study provided a solid foundation and an overview, focusing on the importance of sales forecasting within the hybrid seed industry in South Africa. The chapter introduced the study's primary objective: 'to develop an ML-driven forecasting model that enhances strategic decision-making and operational efficiency'. The chapter further defined key concepts as applied in the study, such as sales forecasting, ML, and the RBV, while detailing the research design and methodology that will be employed to meet these objectives.

This chapter comprehensively reviews the literature relevant to sales forecasting, ML applications in agriculture, and strategic management theories. The chapter begins by examining the traditional approaches to sales forecasting, explaining how conventional methods like Time Series Analysis have been employed to predict future demand. The chapter further explores the limitations of these methods, particularly in dynamic and complex markets like hybrid seed production in the agricultural sector. The chapter next focuses on the role of ML in overcoming these limitations. The study reviews key ML models, such as RFR, SVM, and LSTM networks, explaining their applicability and effectiveness in forecasting tasks.

Furthermore, the chapter draws on strategic management theories, such as the RBV and Competitive Advantage through Innovation, to explain how advanced forecasting tools provide firms with a competitive edge. Finally, gaps are identified in the literature, setting the stage for how the study aims to contribute by offering an innovative, ML-based solution for sales forecasting in the hybrid seed industry. Through this critical literature review, the chapter establishes a clear understanding of the theoretical and empirical foundations that inform the study. Therefore, it supports the argument that integrating ML techniques in sales forecasting is essential for improving decision-making and competitiveness in the agricultural sector.

2.1.1 Contextualisation of hybrid of tomatoes seed production in South Africa

The focus on tomato seeds is particularly significant due to their high demand and critical role in food security and agricultural sustainability (Seed World, 2024). Despite this importance, companies producing seeds must operate profitably to ensure long-term sustainability and maintain food security. Sales forecasting becomes crucial for strategic decision-making to achieve operational sustainability, as it underpins the success of seed companies. However, forecasting accurately in seed companies is challenging due to the rapid development of new hybrids to improve production. As a result, accurate forecasting models, though limited, are

necessary tools that enable companies to predict market trends, optimise production schedules, and manage resources effectively. Liu (2024) explained that advanced ML models, such as Transformer-LSTM networks, offer superior predictive capabilities compared to traditional methods. These models can analyse complex datasets and identify patterns that inform strategic decisions, thereby enhancing the overall performance of seed companies.

Furthermore, environmental factors, particularly weather conditions, are pivotal in agricultural production and seed sales. Malherbe (2015) in a study highlighted the impact of drought on groundwater abstraction, which in turn affects crop yields and seed demand. Therefore, effective strategic management requires integrating weather data into forecasting models to anticipate and mitigate the risks associated with adverse weather conditions. By doing so, companies can ensure a stable supply of seeds and maintain market stability. The cost of production is another critical factor influencing seed sales. Ensafi *et al.* (2022) have emphasised the importance of managing production costs through advanced forecasting and efficient resource allocation. High production costs can increase seed prices, reducing demand among price-sensitive farmers. Therefore, seed companies must adopt cost-effective production techniques and leverage economies of scale to maintain competitive pricing and market share.

Market orientation and understanding customer needs are fundamental to improving seed sales performance. Derbe *et al.* (2024) discuss aligning product offerings with customer preferences and market demands. By adopting a market-oriented approach, seed companies can enhance customer satisfaction, build brand loyalty, and increase market penetration. This involves continuous market research, customer feedback mechanisms, and adaptive marketing strategies. Supply chain management is integral to ensuring the timely delivery of seeds and minimising operational costs. Syngenta (2015) underscores the importance of optimising supply chain processes to enhance efficiency and responsiveness. Additionally, advanced technologies, such as blockchain and the Internet of Things (IoT), can improve transparency and traceability in supply chains, thereby reducing delays and enhancing coordination among supply chain partners.

Innovation and knowledge transfer are essential for maintaining competitiveness in the seed industry. Fitriana *et al.* (2023) highlight the role of innovation in developing new seed varieties and improving production processes. Therefore, seed companies must invest in research and development (R&D) to introduce high-quality, disease-resistant seeds that meet market needs. Also, fostering a continuous improvement culture and facilitating knowledge transfer through training and development programs can enhance organisational performance. CSR and sustainable practices are increasingly important in the agricultural sector. Ahamed *et al.*

(2023) emphasise the need for seed companies to engage in CSR initiatives that address local social and environmental issues. Implementing sustainable practices, such as water conservation and eco-friendly production techniques, can enhance a company's reputation and build strong relationships with local communities and stakeholders.

2.2 Influence of price on tomato seed sales in strategic management decision-making

Price is one of the most significant factors affecting tomato seed sales and plays a critical role in shaping the strategic decisions of seed companies. The price of seeds affects farmers' purchasing decisions and influences the overall market dynamics. Understanding the price elasticity of demand is essential for companies operating in the agricultural sector, as farmers, especially smallholder farmers, are often highly sensitive to price fluctuations. Derbe *et al.* (2024) highlighted in a study that price sensitivity is particularly high among smallholder farmers, who face tight financial constraints and must balance input costs with their expected crop yields. As a result, even slight changes in seed prices can lead to significant shifts in demand, which seed companies must carefully anticipate and manage.

Strategic management must, therefore, integrate a comprehensive understanding of price elasticity when setting pricing strategies. Xhemali *et al.* (2024) argue that hybrid seed production is inherently expensive, and pricing decisions must align with the target market's capacity to pay while ensuring profitability for the seed company. This challenges seed companies, particularly in regions where farmers operate on narrow margins and are sensitive to input costs. Pricing strategies that consider local market conditions and farmer purchasing power become essential in these cases. Demand may drop if prices are set too high, reducing the company's market share and profitability. On the other hand, pricing seeds too low may undermine profitability and long-term sustainability.

Sales forecasting is vital in determining optimal pricing strategies and creating a balanced pricing regime. Forecasting models incorporating price elasticity can help seed companies predict how price changes influence sales volumes. Liu (2024) illustrates that advanced ML models such as SVM are particularly effective at capturing complex relationships between price and demand. By using these models, companies can develop more informed pricing strategies that account for expected fluctuations in market demand. This allows strategic management to adjust pricing based on forecasted market conditions, thus optimising revenue and market positioning.

External factors, such as input costs, environmental conditions, and market competition, influence the relationship between price and sales. In regions where input costs (such as water, labour, and fertilisers) are rising, seed companies must adjust their prices accordingly. However, passing these increased costs on to farmers can be challenging, as it may reduce demand. Ahamed *et al.* (2023) note that sustainable pricing strategies require balancing production costs with market affordability to avoid alienating price-sensitive customers. In this context, companies must develop pricing models that ensure profitability and maintain competitiveness in a market where farmers may turn to cheaper alternatives or lower-cost competitors.

Environmental factors like drought or floods can influence price volatility, reducing agricultural productivity and seed demand. For example, during periods of drought, farmers may reduce their planting areas, leading to lower demand for seeds. Conversely, favourable weather conditions may result in a spike in seed demand as farmers expand their planting operations. Malherbe (2015) emphasises that environmental conditions can exacerbate price elasticity, directly influencing farmers' ability to purchase inputs (Derbe *et al.*, 2024). Strategic management must, therefore, incorporate environmental risk assessments into their pricing models to ensure they remain flexible and responsive to shifts in market conditions.

Competition within the seed industry further complicates the relationship between price and sales. In highly competitive markets, pricing becomes a key differentiator. Companies offering high-quality seeds at competitive prices are more likely to capture market share. Syngenta (2015) discusses how pricing strategies must reflect the cost of production and the competitive landscape. Seed companies must carefully monitor competitor pricing and adjust their strategy to remain competitive while protecting their margins. Sometimes, offering premium pricing for high-quality, high-yield seeds may be justifiable if farmers perceive added value in these products. However, this must be balanced against the price sensitivity of the target market.

Strategic management decision-making in seed companies must also consider long-term pricing trends and their impact on market sustainability. Pricing strategies focusing solely on short-term profits may undermine long-term customer relationships, particularly if farmers feel they are being priced out of the market. Thakur (2023) notes that pricing strategies should be sustainable, considering immediate financial goals and the need to maintain long-term customer loyalty and market penetration. This approach helps to ensure that seed companies can maintain a strong market presence even during periods of price volatility or economic downturns.

Advanced forecasting models can also support dynamic pricing strategies, allowing seed companies to adjust their prices in real-time based on changing market conditions. These models enable companies to respond more rapidly to shifts in demand, cost structures, and competitive pressures. Liu (2024) highlights that real-time data integration and ML models can offer seed companies a significant advantage in adjusting prices based on current market trends (Asfaw, 2021). For example, companies may raise prices to maximise revenues when demand is high during peak planting seasons. Conversely, during periods of low demand, they may lower prices to maintain market share and avoid excess inventory.

The influence of price on tomato seed sales is multifaceted, with price elasticity playing a central role in determining how price changes affect demand. Strategic management must consider various factors, including input costs, environmental conditions, competition, and long-term market sustainability. By incorporating advanced forecasting models and considering external and internal factors, seed companies can develop pricing strategies that optimise sales performance and ensure long-term profitability. This requires a careful balance between market conditions and production costs and a deep understanding of the target market's sensitivity to price changes. Ultimately, strategic pricing is not just about setting the right price today. It ensures pricing decisions support the company's long-term goals and market position.

2.3 Impact of production costs on tomato seed sales in strategic management decision-making

Production costs are critical in determining tomato seed sales, directly influencing pricing strategies and overall profitability within the seed industry. Key components of production costs, such as raw materials, labour, technology, and environmental inputs, shape the strategic decisions made by seed companies. As agricultural profit margins are typically tight, efficient cost management is essential for maintaining long-term competitiveness and sustainability. According to Makridakis *et al.* (2020), the agricultural sectors in Africa face increased challenges due to high energy and transportation costs, as well as a reliance on imported inputs, making cost management particularly difficult in such contexts.

Globally, technological advancements have been instrumental in reducing production costs by increasing efficiency and reducing resource wastage. Makridakis *et al.* (2020) highlight that automation and precision agriculture technologies are widely used in the seed industry to optimise production processes, reduce labour costs, and conserve resources such as water and energy. These innovations allow companies to align production volumes with market demand more effectively, minimising the costs associated with overproduction and inventory

storage. Predictive analytics, such as Artificial Neural Networks, enable companies to forecast market trends more accurately, ensuring optimal resource allocation and reducing unnecessary production.

In Africa, production cost challenges are compounded by infrastructure and energy supply inefficiencies. Bello-Bravo *et al.* (2021) point out that energy costs and transportation inefficiencies can significantly increase the operational costs of seed production, particularly in Sub-Saharan Africa, where many countries still face unreliable energy grids and inadequate transportation networks. This, in turn, leads to higher seed prices, which can affect demand, particularly among price-sensitive smallholder farmers. Bello-Bravo *et al.* (2021) suggest that African seed companies adopt localised production methods to reduce their dependence on costly imports and mitigate the effects of fluctuating input prices. In South Africa, specific factors such as water scarcity, high labour costs, and energy shortages exacerbate production costs. Backeberg (2020) notes that frequent droughts in regions like the Western Cape have significantly increased the cost of irrigation, forcing seed companies to invest in water-saving technologies such as drip irrigation and water recycling. Additionally, South Africa's energy crisis, characterised by frequent power outages, has compelled many companies to invest in alternative energy sources, such as solar power, to maintain operational stability; these investments, though costly upfront, are necessary for reducing long-term production costs and ensuring sustainable operations.

Labour costs in South Africa further contribute to the complexity of production cost management. Leibbrandt *et al.* (2021) explain that rising wage demands have prompted many companies to explore automation and mechanisation to reduce labour expenses. By adopting mechanisation, seed companies can reduce their reliance on manual labour, streamline production, and control costs more effectively.

Seed companies increasingly turn to advanced forecasting models to address these cost challenges. Makridakis *et al.* (2020) highlight the effectiveness of Seasonal Autoregressive Integrated Moving Average models in predicting fluctuations in seed demand, enabling companies to adjust their production schedules accordingly. These models help seed companies avoid the costs associated with overproduction by aligning production with demand forecasts. ML models, such as Gradient Boosting Decision Trees, offer even more sophisticated forecasting capabilities. These models allow companies to factor in multiple variables, such as input costs, market trends, and environmental conditions, to make informed production decisions. Environmental considerations are also critical in managing production costs. In Southern Africa, rising temperatures and erratic rainfall patterns have increased irrigation and crop protection costs. Makridakis *et al.* (2020) emphasise the importance of

investing in climate-resilient technologies, such as drought-tolerant seed varieties and efficient irrigation systems, to reduce the financial risks associated with climate change. By incorporating environmental data into their production strategies, seed companies can mitigate the impact of climate variability, ensuring stable production costs and consistent supply chains.

Effective supply chain management is also a critical factor in controlling production costs. Hobbs (2021) suggests that optimising supply chains can significantly reduce production delays and transportation costs, directly impacting the overall cost of seed production. Streamlining the procurement of raw materials and improving logistics efficiency can lead to substantial cost savings, enabling companies to offer more competitive pricing. In South Africa, strengthening local supply chains and reducing reliance on expensive imports can help seed companies stabilise production costs and enhance their market positioning.

The theoretical frameworks of Porter's Competitive Advantage and the RBV provide valuable insights into how seed companies can leverage cost management to maintain a competitive edge. Companies that effectively control production costs through technological innovation, strategic supply chain management, and forecasting are better positioned to maintain profitability while offering competitively priced products. According to Makridakis *et al.* (2020), companies with strong cost management capabilities are more likely to navigate market volatility successfully and ensure long-term sustainability. Managing production costs is essential for seed companies to remain competitive and profitable in the tomato seed market. Seed companies can control costs by adopting technological innovations, employing advanced forecasting models, and optimising supply chains while maintaining product quality and competitiveness. This comprehensive approach to cost management supports long-term sustainability and positions companies for success in domestic and international markets.

2.4 Effects of production quantity on tomato seed sales in strategic management decision-making

Production quantity is a key determinant of sales performance in the tomato seed industry, influencing supply dynamics, pricing strategies, and customer satisfaction. Strategic management must carefully balance production volumes to align with market demand, ensuring seed companies meet customer needs without overproducing or underproducing. As Makridakis *et al.* (2020) noted, effective demand forecasting plays a pivotal role in determining optimal production volumes, thereby minimising the risks associated with overproduction, such as excess inventory costs and underproduction, which can lead to missed sales opportunities and customer dissatisfaction. Internationally, large-scale seed companies have

implemented advanced ML models, such as LSTM networks and SVM, to predict market trends and optimise production planning (Makridakis *et al.*, 2020).

In the global seed market, multinational companies like Syngenta and Bayer Crop Science use data analytics and precision agriculture to manage production quantities across multiple regions, ensuring their production volumes align with local demand patterns (Christopher, 2016). These companies have also adopted blockchain technology in their supply chains to track production volumes, distribution, and market response in real-time, providing valuable insights that inform strategic decision-making. These technologies offer important lessons for African seed companies looking to improve production efficiency and reduce costs.

On the African continent, the production of seeds is often constrained by resource limitations, including unreliable infrastructure and fluctuating energy costs. Bello-Bravo *et al.* (2021) note that many African seed companies face difficulties maintaining production efficiency due to high transportation and energy expenses, significantly affecting their ability to meet market demand. In regions like East Africa, drought conditions make it difficult for seed companies to predict planting seasons accurately, leading to inconsistent production volumes. Ncube and Mhlanga (2020) argue that localised production strategies, including indigenous seed varieties, can help mitigate some of these challenges by reducing dependence on imported inputs and stabilising production costs.

In South Africa, the impact of environmental variability is particularly pronounced, with water scarcity and energy shortages posing significant challenges to seed companies. Backeberg (2020) highlights that in drought-prone regions like the Western Cape, the cost of irrigation has escalated dramatically, making it critical for seed companies to optimise production quantities to avoid unnecessary resource usage. Schulze (2016) emphasises that underproduction, often caused by insufficient rainfall, can lead to missed sales opportunities and damage customer relationships as companies struggle to meet market demand. Strategic management in South Africa must incorporate advanced forecasting models, such as Seasonal Autoregressive Integrated Moving Average and LSTM networks, to anticipate changes in planting seasons and align production schedules with actual market demand (Makridakis *et al.*, 2020). Managing production quantities is not just about balancing supply and demand but also about ensuring the sustainability of operations. Overproduction can lead to excessive resource usage, especially in water-scarce environments like Southern Africa. At the same time, underproduction can limit market access for smallholder farmers who depend on a reliable supply of seeds. Hobbs (2021) points out that strategic management must include sustainability as a core decision-making component, ensuring that production volumes meet market needs while minimising environmental impact.

2.5 Role of weather on tomato seed sales and its impact on strategic management decision-making

Weather is pivotal in agricultural production, directly influencing tomato seed sales and impacting seed companies' strategic decisions. Changes in weather conditions, such as rainfall, temperature, and drought, have profound effects on crop growth and seed demand, making it essential for seed companies to integrate weather-related data into their strategic planning. This allows them to mitigate risks, optimise production schedules, and adapt to market dynamics in response to fluctuating environmental factors. As Malherbe (2015) highlights, drought conditions can severely disrupt agricultural productivity by reducing groundwater levels, limiting irrigation and affecting crop yields. For seed companies, such weather-induced disruptions often translate to decreased seed demand, as farmers may scale back planting due to water scarcity. Conversely, favourable weather conditions encourage farmers to expand their planting, boosting demand for seeds and requiring seed companies to adjust their production and distribution strategies accordingly.

Advanced forecasting models incorporating weather data offer significant advantages in predicting how seed demand may fluctuate based on these environmental factors. Liu (2024) demonstrates that seed companies can achieve more accurate sales forecasts by integrating weather variables into ML models like the Transformer-LSTM. These forecasts help management align production schedules with anticipated market demand, thus reducing the risk of overproduction or stockouts. In the context of strategic decision-making, having access to precise weather-based predictions allows companies to adjust both production and inventory management in real time, ensuring that supply remains closely aligned with market conditions. As such, integrating weather data into predictive models is no longer an optional strategy for seed companies but a crucial part of maintaining operational efficiency in a climate-sensitive industry.

The seasonal nature of agricultural production further amplifies the need for weather-aware decision-making. Seasonal variations in weather, such as periods of heavy rainfall or prolonged droughts, heavily influence planting and harvesting schedules, affecting the timing and volume of seed sales. According to Wang (2020), understanding these seasonal trends is essential for developing effective production and marketing strategies that anticipate changes in demand. Strategic management must, therefore, align production with these seasonal patterns to avoid excess inventory during off-seasons and stock shortages during peak periods. By closely monitoring weather forecasts and historical patterns, seed companies can better manage their inventory levels, reducing the costs associated with holding excess stock and ensuring a steady supply of seeds when demand spikes.

Weather-related risks extend beyond just the regular seasonal fluctuations. Extreme weather events, such as floods, storms, and unexpected frost, can cause significant disruptions to both production and distribution systems. Thakur (2023) argues that developing flexible contingency plans is key to mitigating the effects of such disruptions. For seed companies, this may involve diversifying supply sources, maintaining buffer stocks, and enhancing logistics to ensure that seeds are delivered to the market on time, even in adverse weather conditions. Flexibility in production and distribution systems allows companies to respond more effectively to sudden changes in weather and helps them maintain their competitive advantage by ensuring consistent supply.

Another dimension of weather's influence is its effect on pests and diseases, further complicating production planning. As Ensafi *et al.* (2022) explain, certain weather conditions, such as high humidity or extreme heat, can increase the incidence of pest infestations or fungal diseases. These environmental conditions directly affect both seed quality and crop yields, making it imperative for seed companies to anticipate and manage these risks through preventive measures, such as selecting disease-resistant seed varieties or adjusting pesticide application schedules. By integrating pest and disease forecasts into their production planning, companies can maintain high seed quality and ensure a consistent supply to meet market demands.

Sustainability is crucial in mitigating the longer-term risks associated with adverse weather conditions. As Ahamed *et al.* (2023) argue, sustainable agricultural practices, such as water conservation, adopting climate-resilient seed varieties, and utilising precision agriculture technologies, are vital for building resilience in seed production systems. These practices help seed companies cope with the immediate challenges of fluctuating weather and ensure long-term viability in increasingly unpredictable climates. For example, adopting drought-resistant seed varieties in regions prone to water scarcity allows companies to meet the demand for resilient crops while ensuring that farmers can continue planting even under adverse conditions. Strategic management must, therefore, prioritise sustainability initiatives as part of their long-term planning, ensuring that the company can adapt to changing environmental conditions while maintaining its competitive edge.

Regional weather patterns also significantly influence consumer behaviour and demand for specific seed varieties. According to Syngenta (2015), regions with frequent droughts may exhibit higher demand for drought-resistant seeds, while areas with more reliable rainfall may favour high-yielding varieties. By tailoring their marketing strategies to these regional weather trends, seed companies can develop more targeted campaigns that better address the needs of local farmers. Strategic management should continuously analyse regional weather trends

to refine their product offerings and marketing initiatives, ensuring they align with each geographic area's environmental realities and customer preferences. This regional focus allows companies to enhance customer satisfaction while expanding their market reach in diverse climates.

Weather is a critical factor influencing tomato seed sales and must be central to strategic management decision-making in the seed industry. By integrating weather data into forecasting models, planning for seasonal and extreme weather variations, managing pest and disease risks, and adopting sustainable practices, seed companies can navigate the complexities of agricultural production more effectively. A comprehensive, weather-informed strategy ensures that seed companies remain competitive, adaptable, and well-prepared to meet current and future market demands. Strategic management, therefore, must embrace weather-related insights as a core component of their decision-making processes to ensure long-term growth and sustainability in a climate-sensitive agricultural landscape.

2.6 International business management and administration practices

In today's interconnected global economy, businesses aiming for international expansion must adopt effective management and administration practices that address the unique challenges of operating in diverse cultural, economic, and regulatory environments. International management strategies are critical for navigating these complexities, ensuring companies thrive in global markets. This section examines several key areas crucial to international business management, including cultural competence, strategic alliances, supply chain optimisation, regulatory compliance, financial management, innovation, CSR, and the role of technology. Understanding and managing cultural differences is one of the most significant challenges in international management. Singh *et al.* (2020) argue that cultural competence is essential for businesses seeking to expand globally, as it fosters smoother interactions, improves negotiation outcomes, and builds stronger relationships with stakeholders in different markets. Companies that invest in cultural awareness training can enhance employees' ability to navigate intercultural communication, reducing the risk of misunderstandings and ensuring more effective collaboration across borders. Cultural sensitivity also aids in customising products and services to meet local preferences, a key aspect of successful international management.

Strategic alliances and partnerships are another vital component of international business strategy. Collaborating with local firms can provide businesses with valuable insights into the regional market dynamics, regulatory environments, and consumer behaviour. According to Hasselbeck *et al.* (2022), partnerships with local firms facilitate market entry by providing

access to local knowledge and reducing the risks associated with unfamiliar markets. These collaborations allow companies to leverage local expertise and infrastructure, ensuring a smoother transition into new territories. Additionally, alliances help localise offerings to meet the specific needs of regional consumers, enhancing competitive advantage in the global marketplace.

Supply chain management is equally important in the context of international business. As supply chains become more global and complex, companies must optimise logistics to ensure the timely delivery of goods and services across borders. Hasselbeck *et al.* (2022) emphasise that advanced technologies such as blockchain and IoT transform global supply chains by improving transparency, traceability, and operational efficiency. By leveraging these technologies, businesses can enhance the trust and reliability of their supply chains, minimising the risks associated with cross-border logistics, such as delays, fraud, and product counterfeiting. Furthermore, effective supply chain management allows businesses to respond more flexibly to market changes and disruptions, which is essential for maintaining competitive positioning in the global market.

Navigating the regulatory environments of different countries poses a significant challenge for businesses expanding internationally. Ahamed *et al.* (2023) highlight the importance of developing robust compliance programs that address regulations on local taxation, employment, and environmental standards. Failure to comply with local laws can result in costly legal penalties and damage a company's reputation. Therefore, businesses must prioritise understanding and adapting to the regulatory landscapes of the regions in which they operate. This involves creating specialised teams focusing on legal compliance in different markets and engaging local legal experts to ensure adherence to local requirements. Effective financial management is also crucial in international business, especially when dealing with fluctuating currency exchange rates and differing economic conditions. Araujo and Guglianone (2023) note that currency fluctuations can significantly affect the profitability of international operations. To mitigate these risks, companies often use financial instruments such as forward contracts and options to hedge against adverse currency movements. In addition to hedging strategies, businesses must develop comprehensive financial plans that ensure efficient capital allocation across multiple markets. This helps maintain financial stability while supporting long-term growth in global markets.

Innovation and knowledge transfer are key drivers of success in international business. Companies must create mechanisms for sharing best practices and innovations across their global operations as they expand into new regions. According to Ahamed *et al.* (2023), multinational companies benefit from establishing innovation hubs and utilising digital

platforms to facilitate real-time knowledge exchange. By promoting cross-border collaboration and enabling the transfer of technological advancements, companies can ensure that innovations developed in one region can be adapted and implemented in others, thus driving efficiency and maintaining competitiveness.

CSR is becoming an increasingly important aspect of international management, as businesses are expected to contribute positively to the communities in which they operate. Frona *et al.* (2019) argue that engaging in CSR initiatives enhances a company's reputation and strengthens relationships with local communities and stakeholders. Companies that invest in sustainable practices, such as reducing environmental impact or supporting local economic development, can gain a competitive advantage by building trust and loyalty among consumers and regulators. In the long term, CSR efforts contribute to business sustainability and resilience in the global market. The role of technology in international management cannot be overstated. As Liu (2024) demonstrates, advanced data analytics and ML models are essential for improving decision-making processes in international business. By leveraging big data, companies can gain deeper insights into global market trends, consumer preferences, and operational efficiencies. Predictive analytics allows businesses to anticipate shifts in market demand and adjust their strategies, ensuring that they remain agile and competitive in rapidly changing international environments. Additionally, digital tools facilitate communication and collaboration across geographically dispersed teams, enabling more efficient coordination of global operations.

International business management and administration practices require a multifaceted approach encompassing cultural competence, strategic alliances, supply chain optimisation, regulatory compliance, financial management, innovation, CSR, and the effective use of technology. By integrating these elements into their international strategies, companies can navigate the complexities of global markets more effectively, ensuring sustainable growth and long-term success.

2.7 National business management and administration practices

National management and administration practices are crucial for businesses thriving within domestic markets. These practices ensure that companies align their strategies with local economic, cultural, and regulatory environments to achieve operational efficiency and long-term success. This section delves into key aspects of national business management, including regulatory compliance, strategic alignment with national economic policies, cultural competence, supply chain management, innovation, CSR, and financial management, while integrating relevant case studies and practical applications.

One of the most critical elements of national management is ensuring compliance with local regulatory frameworks. Adherence to national laws and regulations is necessary to avoid legal repercussions, maintain operational integrity, and build consumer trust. Derbe *et al.* (2024) argue that regulatory compliance is fundamental for smooth operations, particularly in sectors like agriculture, where businesses must comply with labour laws, environmental standards, and industry-specific regulations. For instance, Pannar Seeds, a South African agricultural company, has successfully aligned its operations with local agricultural policies by ensuring compliance with environmental regulations regarding water usage and soil quality. The company has maintained its reputation for producing high-quality seeds by adhering to national standards, strengthening consumer trust and safeguarding its market position. Strategic alignment with national economic policies is also pivotal for business success. In many countries, governments implement policies to boost economic sectors, such as agriculture, through subsidies or development plans. Ensafi *et al.* (2022) emphasise that aligning business strategies with national economic objectives can create new company opportunities. For example, South Africa's Agri-Parks Programme, which aims to support smallholder farmers and improve food security, offers seed companies like the one under investigation an opportunity to participate in government initiatives by partnering with smallholder farmers through this programme, expanding its market presence while contributing to national agricultural productivity goals. This alignment with government policies enhances competitive advantage and positions companies as key contributors to national development.

Cultural competence is another critical element of national management practices. Understanding local consumer behaviour and cultural norms allows businesses to tailor their products and services to meet domestic market preferences. Xhemali *et al.* (2024) highlight the importance of investing in cultural intelligence to develop effective marketing and management strategies. In South Africa, a country known for its cultural diversity, businesses must adapt their strategies to resonate with different cultural groups. For example, Pick n Pay (2024), a major South African retailer, has adapted its marketing campaigns to reflect local languages and cultural celebrations, such as the emphasis on community during Heritage Day. This approach enhances customer satisfaction, helps build brand loyalty, and ensures deeper market penetration.

Effective supply chain management is another essential component of national business administration. Efficiently managing domestic supply chains ensures that products are delivered on time, costs are minimised, and customer satisfaction is maintained. Liu (2024) demonstrates that optimising logistics and distribution networks within the national market can

significantly reduce lead times and lower transportation costs. For instance, Shoprite, Africa's largest food retailer, has implemented advanced IoT-based systems to track the movement of goods across its supply chain, reducing wastage and ensuring product availability. By leveraging real-time data analytics, Shoprite can adjust inventory levels based on demand forecasts, enhancing its responsiveness to market changes while minimising operational costs. This use of technology in supply chain management has allowed the company to maintain a competitive edge in the fast-moving consumer goods sector. Innovation and knowledge transfer are key drivers of competitiveness within the national market. Fitriana *et al.* (2023) argue that fostering a culture of continuous improvement and innovation is essential for businesses to stay competitive in the face of evolving market dynamics. For example, Grain SA has invested in R&D to develop drought-resistant maize seed varieties, which are crucial for mitigating the effects of climate change on agricultural production. Additionally, the company has implemented training programs for farmers, transferring knowledge on sustainable farming practices and new technologies. This emphasis on innovation and knowledge transfer has allowed Grain SA to remain competitive in the national market while contributing to the sustainability of South Africa's agricultural sector.

CSR plays an increasingly important role in national business practices. Engaging in CSR initiatives that address local social, economic, and environmental challenges allows businesses to build strong relationships with stakeholders and enhance their corporate reputation. Malherbe (2015) stresses that companies implementing sustainable practices, such as community development projects or environmental conservation efforts, are better positioned to build trust with local communities and regulators. Clover, a South African dairy company, has implemented various CSR initiatives, including water conservation projects and youth education programs in rural areas. These initiatives enhance Clover's brand reputation and contribute to local development, creating a favourable operating environment that supports long-term success. By aligning CSR with local priorities, businesses can build trust with stakeholders and ensure their operations are seen as a force for good within the community.

Finally, financial management tailored to the national context is essential for business sustainability. Navigating local economic fluctuations, currency risks, and market volatility requires effective financial planning and risk management. Thakur (2023) emphasises the importance of developing financial strategies that address local market conditions, such as managing currency risks or optimising capital allocation within the national market. South African agricultural companies, for instance, face fluctuating exchange rates and inflation, which can impact their cost structures and pricing strategies. Tiger Brands, a leading South

African consumer goods company, has successfully mitigated these risks by diversifying its revenue streams and adopting forward contracts to hedge against currency fluctuations. By maintaining robust financial controls and leveraging local financial instruments, Tiger Brands ensures financial resilience, even amid economic uncertainty.

Effective national business management and administration practices are critical for companies seeking operational success and strategic alignment within their domestic markets. By focusing on regulatory compliance, aligning with national economic policies, understanding local cultural dynamics, optimising supply chains, fostering innovation, engaging in CSR, and managing financial risks, businesses can create a strong foundation for sustainable growth. These practices enhance competitive advantage within the national market and prepare businesses for domestic and international expansion. Moreover, businesses that excel at national management often find themselves better equipped to navigate the complexities of international markets, as the principles of sound governance and strategic alignment remain consistent across borders.

2.8 Theoretical underpinning and conceptualisation in the study

This section critically evaluates various theoretical frameworks, arguing for their suitability (or lack thereof) for understanding the strategic decision-making processes in the hybrid tomato seed industry. Theories such as the RBV, Porter's Competitive Advantage, Innovation Diffusion Theory (IDT), Institutional Theory, and Contingency Theory are assessed. After this evaluation, RBV and Contingency Theory are identified as the most appropriate frameworks to underpin the study, and their relevance to the main constructs of this research is extensively developed.

2.8.1 Evaluation of the theoretical framework

This section assesses theories such as the RBV, Porter's Competitive Advantage, IDT, Institutional Theory, and Contingency Theory based on suitability and limitations.

2.8.1.1 Resource-Based View

The RBV, originally proposed by Barney (1991), argues that firms gain a competitive advantage through their ability to manage internal resources that are valuable, rare, inimitable, and non-substitutable. In the hybrid tomato seed industry context, internal resources such as advanced forecasting systems, data analytics, and proprietary seed technologies are vital to improving forecast accuracy, optimising production processes, and ensuring market competitiveness.

Suitability: RBV is highly suited to the study because it focuses on how the company's internal resources, particularly ML models for sales forecasting, drive strategic decision-making. By adopting advanced forecasting models like Hybrid Holt-Winters and Support Vector Machine, the company can leverage its internal resources to enhance production efficiency and sales performance (Ahamed *et al.*, 2023).

Limitations: Despite its strengths, RBV does not address the role of external forces like weather patterns, pricing fluctuations, or market dynamics, which are significant in the tomato seed industry. As such, RBV alone cannot fully capture the complexities of this market, where external factors heavily influence demand and production schedules.

2.8.1.2 Porter's Competitive Advantage Theory

Porter's Competitive Advantage Theory (Porter, 2011) is widely used to explain how firms achieve profitability through cost leadership, differentiation, or market focus. In agricultural markets like the tomato seed industry, competitive advantage could be achieved through pricing strategies, market penetration, and product differentiation.

Suitability: Porter's theory is particularly relevant for understanding the competitive dynamics in the tomato seed industry, especially regarding how seed companies position themselves against competitors, manage supplier power, and set pricing strategies.

Limitations: While useful for analysing external market competition, Porter's framework does not focus on internal capabilities and operational efficiency, which are crucial in the study. Moreover, it does not account for the environmental factors that heavily influence seed demand and production, making it less applicable when considering strategic forecasting and production management decisions.

2.8.1.3 Innovation Diffusion Theory

IDT, developed by Rogers (2003), concerns how innovations are adopted over time within a social system. In the study, IDT could be applied to understand how seed companies adopt new technologies, such as ML models for forecasting and precision agriculture tools.

Suitability: This theory offers insights into how innovative technologies are adopted within the seed industry, particularly as companies integrate new forecasting systems into their operational processes. It helps explain the spread and uptake of innovations that improve operational efficiency and production outcomes.

Limitations: While IDT explains the process of technological adoption, it does not adequately address the broader strategic implications of managing internal resources and external market conditions. Hence, IDT is more complementary than foundational for understanding strategic decision-making in this context.

2.8.1.4 Institutional Theory

Institutional Theory posits that organisations conform to the external environment's rules, regulations, and norms to gain legitimacy (Scott, 1995). In the tomato seed industry, this could involve adhering to agricultural regulations, environmental standards, and market norms.

Suitability: Institutional Theory helps explain how companies maintain compliance with industry standards and external regulations, which is crucial in agricultural sectors where quality and safety are paramount.

Limitations: While this theory is useful for understanding the company's external compliance, it does not sufficiently address the operational complexities of managing internal resources or responding to dynamic market conditions like weather variability or pricing volatility, which are critical to the focus of the study.

2.8.1.5 Contingency Theory

Contingency Theory asserts that there is no single best way to manage an organisation; instead, the best approach depends on the fit between the internal and external environment (Fiedler, 1964). Contingency Theory is particularly relevant in the study because seed companies must align their internal resources (ML models) with external factors such as weather conditions, market demand, and pricing trends.

Suitability: Contingency Theory is well-suited for the study because it emphasises flexibility and adaptation to external factors, two critical elements for companies in the tomato seed industry, where external conditions like weather patterns can heavily influence production and sales (Wang, 2020). The theory supports the idea that companies must adapt their internal capabilities to meet the challenges presented by unpredictable environmental and market conditions.

2.8.2 Selection of theories for the study

After evaluating the theoretical frameworks, RBV and Contingency Theory are selected as the most appropriate for the study. Both theories are highly applicable to the key constructs of this

research, and they complement each other in addressing both internal resources and external environmental factors.

2.8.2.1 Justification for RBV

RBV is critical for explaining how internal capabilities, such as forecasting models and technological infrastructure, provide a competitive advantage. In the study, ML tools are essential internal resources that allow the company to predict market demand, optimise production schedules, and reduce waste (Liu, 2024). These capabilities directly enhance the company's strategic decision-making, particularly in managing production and inventory efficiently to meet fluctuating seed demand.

2.8.2.2 Justification for Contingency Theory

Contingency Theory is selected to complement RBV because it accounts for the influence of external environmental factors such as weather conditions, market fluctuations, and pricing trends on strategic decision-making. In the tomato seed industry, where external conditions influence demand, the company must adapt its production processes to respond to these factors dynamically. For example, adverse weather conditions may reduce seed demand, while favourable conditions might increase sales, requiring the company to adjust its production planning accordingly (Malherbe, 2015).

2.8.3 Development of the chosen theories

In the study, RBV provides the theoretical foundation for understanding how internal resources like ML models help the company maintain operational efficiency and gain a competitive advantage. Contingency Theory ensures that the company adapts these internal resources to meet the demands of the external market environment, particularly in responding to weather variability and pricing fluctuations. The combined application of these two theories forms a comprehensive framework for strategic decision-making, where internal resources and external adaptability converge.

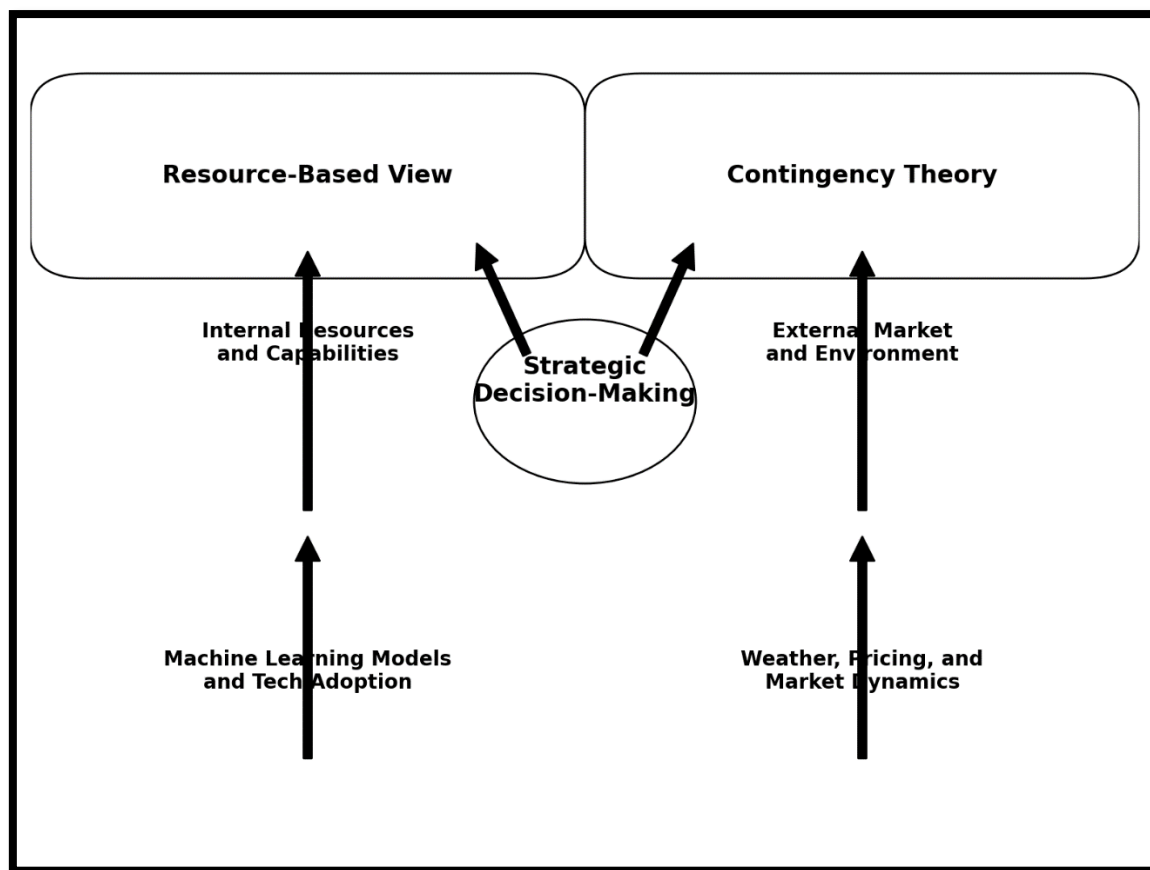


Figure 3: A diagram indicating the two selected theories applied in the study.

Figure 3 visually represents the interaction between internal resources (RBV) and external factors (Contingency Theory) that influence the company's strategic decision-making. The company can make informed decisions that ensure operational success by leveraging internal resources like forecasting models and adapting to external factors such as weather conditions and market dynamics.

Integrating RBV and Contingency Theory provides a robust theoretical underpinning for the study. RBV focuses on the internal capabilities that drive operational efficiency, while Contingency Theory emphasises the need for external adaptability in a market influenced by unpredictable factors like weather and pricing. Together, these theories offer a comprehensive approach to understanding strategic decision-making in the tomato seed industry, ensuring the company can maintain a competitive advantage while responding effectively to market and environmental conditions.

2.9 Chapter summary

This chapter has established the theoretical foundation necessary to guide the exploration of strategic decision-making in the tomato seed industry. Several theoretical frameworks were

critically evaluated, including the RBV, Porter's Competitive Advantage Theory, IDT, Institutional Theory, and Contingency Theory. These frameworks were assessed for their relevance to understanding the internal and external factors influencing seed companies' operations, particularly concerning production efficiency and market adaptability.

After careful consideration, RBV and Contingency Theory were selected as the most appropriate theoretical foundations for the study. RBV was chosen for its emphasis on leveraging internal resources such as advanced forecasting technologies and ML models, which are critical in improving operational efficiency and creating a sustainable competitive advantage. By focusing on the company's internal capabilities, RBV provides a framework for understanding how seed companies can optimise their production processes and better align seed availability with fluctuating market demand.

In complement to RBV, Contingency Theory was selected for its focus on the external environment and the importance of adapting to external market conditions, such as weather variability, pricing trends, and seasonal demand. Contingency Theory recognises that strategic decision-making must account for these unpredictable factors that are particularly relevant in the agricultural sector. By integrating external adaptability with internal resource management, these two theories offer a comprehensive approach to understanding and improving strategic decision-making in the tomato seed industry.

RBV and Contingency Theory form the conceptual framework for the study, linking the company's internal capabilities with the need for external flexibility. This dual approach will provide valuable insights into how seed companies can navigate market challenges, optimise production schedules, and improve sales forecasting accuracy.

As the study progresses into Chapter 3, the focus will shift to the research methodology, where the theoretical insights from this chapter will be operationalised. Data collection and analysis methods will be developed to test the conceptual framework, ensuring the study's objectives are met through empirical investigation.

CHAPTER 3 RESEARCH METHODOLOGY AND METHOD

3.1 Introduction

In the previous chapter, an in-depth review of the literature on pricing strategies and their impact on tomato seed sales was conducted, particularly within the agricultural seed industry. The discussion explored how pricing mechanisms have evolved, focusing on price elasticity and its influence on consumer demand. Technological advancements, such as predictive analytics and ML models, were also reviewed for their role in forecasting market trends and supporting strategic decision-making. External factors, including competitive pressures and broader economic shifts, were examined to understand how they shape pricing strategies. This theoretical groundwork provided the necessary insights to bridge into the following empirical study.

This chapter uses a quantitative approach based on secondary data to explore the methodology employed to investigate the relationship between pricing and tomato seed sales. The research draws from historical sales records, production costs, and weather data, allowing for the development of an accurate and reliable forecasting model. The decision to focus on quantitative analysis is rooted in its capacity to provide a clear, data-driven perspective on trends and relationships, making it the ideal framework for the study. The first section outlines the rationale behind this methodology, detailing how secondary data provides valuable insights without the need for primary data collection. This is followed by a thorough explanation of the data collection process, highlighting the selection of reliable sources such as historical records and industry reports to ensure the robustness and validity of the dataset. Key analytical tools, including statistical methods and advanced ML models like RFR and hybrid approaches, are introduced to demonstrate how the study processed this data. These models were selected for their ability to capture non-linear patterns in the data, making them well-suited for forecasting complex market trends.

The chapter then turns to data analysis, where techniques such as regression analysis and descriptive statistics are used alongside ML models to uncover significant relationships between price, production costs, and sales volumes. These tools help in developing precise forecasts that inform decision-making. Finally, ethical considerations are addressed, particularly the responsible handling of secondary data, ensuring adherence to academic standards and maintaining the integrity of the research process.

3.1.1 Research philosophy

The research philosophy adopted in the study is crucial as it informs the research process's overall paradigm, approach, and design. According to Saunders *et al.* (2016), research philosophy refers to beliefs and assumptions about knowledge development. These assumptions shape how the researcher views reality and approaches the research. The philosophy chosen for the study is positivism, which asserts that reality is objective and measurable through empirical data. Given the quantitative nature of the study, positivism provides a robust framework for analysing secondary data on pricing strategies and tomato seed sales.

3.1.1.1 Positivism

Positivism assumes that reality operates according to fixed laws of cause and effect and that these laws can be understood through the systematic collection and analysis of objective data. This approach is well suited for the study as it seeks to explore the relationship between variables such as pricing, production costs, and sales volumes using measurable data. The reliance on empirical evidence enables the study to maintain objectivity, reducing the risk of bias (Saunders *et al.*, 2016). Despite its strengths, it is important to acknowledge that positivism may overlook contextual factors or the subjective experiences of stakeholders, which are not captured through quantitative data. However, positivism remains appropriate for the study, focusing on large-scale data analysis due to its emphasis on generalisability and replicability.

The study follows a deductive approach, a key feature of positivism, where existing theories related to pricing and sales performance are tested through data analysis. The process begins with hypotheses based on established theories, which are then tested using statistical tools and ML models. This allows the study to confirm or refine existing knowledge, contributing to a deeper understanding of the relationship between pricing strategies and market performance (Bryman *et al.*, 2021).

3.1.1.2 Saunders' research onion

The Research Onion model, as proposed by Saunders *et al.* (2016), offers a systematic framework for selecting research methods that align with the study's objectives. The outer layer of the onion is research philosophy, and positivism is the guiding philosophy for the study. This informs the deductive approach, where the study begins with established theories and works to test them through data analysis. Moving inward in the onion, the next layer focuses on research strategy, which, in this case, uses quantitative methods to analyse

secondary data. Quantitative methods are particularly suited to positivism as they emphasise objectivity and measurement.

Further onion layers include the time horizon, which is cross-sectional in the study. A cross-sectional design involves the analysis of data collected at a single point in time, which is consistent with the study's reliance on historical sales data to understand market patterns. By following Saunders' research onion, the study ensures that each methodological choice is logically connected, allowing for a coherent and rigorous research design.

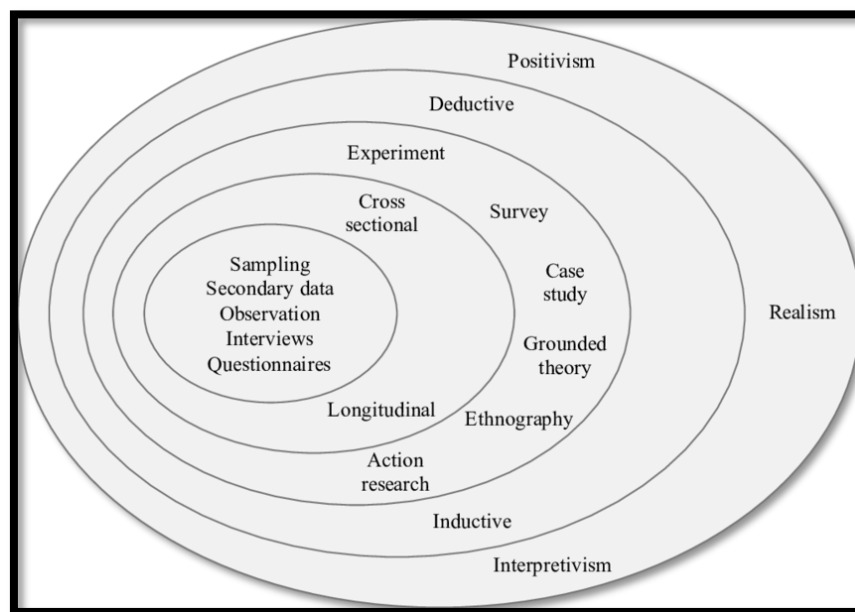


Figure 4: Saunders' research onion highlighting research philosophy, approach and research strategy (Saunders *et al.* 2016).

Figure 4 visually represents the layers involved in developing a research methodology. This model is useful for guiding the selection of research methods, beginning with the outermost layer of research philosophy and progressing to the inner layers of approach and data collection techniques. In the study, positivism is the philosophy guiding the research, as discussed earlier. Positivism is aligned with the quantitative nature of the study, which focuses on using objective, measurable data to understand the relationship between pricing and sales. Moving inward, the next layer focuses on the study's deductive research approach. This approach involves testing existing theories through secondary data analysis, allowing for hypotheses related to pricing strategies to be confirmed or refined based on empirical evidence. The deductive approach is a hallmark of positivist research, which aims to test theoretical propositions through observable, quantifiable data.

The research strategy layer of the onion in Figure 4 concerns how the data is collected and analysed. A quantitative strategy has been adopted for the study, focusing on analysing historical sales and pricing data through statistical methods and ML models. This strategy is appropriate for answering the research questions as it emphasises objectivity and precision, which are core principles of the positivist philosophy. Finally, the time horizon for the study, represented in the subsequent layers, is cross-sectional, meaning that the study analyses data from a specific point in time. This approach allows for a snapshot of the relationship between pricing strategies and sales performance, providing insights that can be applied in a real-world business context.

3.2 Research paradigm and design

The study's research paradigm is rooted in the positivist approach, which assumes that reality is objective and measurable through empirical observation and data analysis. This approach aligns with the study's goal of developing a robust sales forecasting model using machine learning (ML) algorithms. Quantitative methods were employed to establish relationships and test hypotheses, drawing on secondary data from reliable sources. The data included historical records of tomato seed sales, production volumes, input costs, and weather data, providing a comprehensive foundation to explore the interplay between these variables and their impact on sales performance (Saunders *et al.*, 2016; Adams, 2015).

The dataset spans eleven years, from 2013 to 2023, and includes detailed annual data on production volumes ranging from 7.24 million units in 2014 to 22.80 million units in 2020. Input costs varied between R755,46 in 2022 and R1,890,17 in 2020, while sales revenue ranged from R3.40 million in 2013 to R18.26 million in 2020. This extensive temporal and variable range captured diverse market conditions and trends, ensuring the robustness of the forecasting model. Additionally, high-resolution weather data from ERA5-Land monthly averaged datasets (1981–2023) was utilised to explore the effects of climatic variables on agricultural sales. ERA5-Land's precision in capturing precipitation and temperature trends provided over 480 monthly records, which were aligned with production and sales data to assess the impact of climatic variations on seed sales (Muñoz-Sabater, 2019).

The research design follows a descriptive and correlational approach, which is particularly suited for analysing secondary data. This design allowed the identification of patterns and relationships within the dataset, focusing on quantifiable variables such as price, production volumes, and weather conditions. The integration of statistical and machine learning methods enabled a detailed exploration of how these variables interact. Descriptive statistics provided an initial overview of trends, such as seasonal fluctuations in sales volumes and corresponding

input costs. Multivariate analysis, including correlation and regression techniques, was employed to examine the relationships between independent variables and sales outcomes.

Machine learning techniques were central to the study, particularly hybrid models that combine Holt-Winters Exponential Smoothing with Support Vector Machines (SVM). These models were trained using an 80:20 data split, with eight years of data (2013–2020) allocated to training and three years (2021–2023) used for testing. This ensured that the models were exposed to diverse market conditions, enhancing their robustness and predictive accuracy. The cleaned dataset retained over 90% of its original entries after rigorous data quality checks, which included verifying completeness, ensuring temporal alignment across variables, and removing outliers. Sensitivity analyses were also conducted to test the forecasting models under varying scenarios, such as changes in price and production levels, ensuring reliability across different conditions. The breadth and diversity of the data used in this study underscore its alignment with the positivist paradigm. By leveraging over a decade of detailed production, sales, and input cost data, along with high-resolution weather records, the study ensured a robust methodological foundation for developing a forecasting model. These efforts not only enhance the validity of the results but also highlight the study's commitment to integrating comprehensive data sources into its design.

3.3 Data collection and integration of global datasets

The data collection process for the study was meticulously planned to ensure accuracy, reliability, and relevance, all essential for developing a robust sales forecasting model for tomato seeds. Given the study's reliance on secondary data, key sources included historical sales records, production costs, and weather data, carefully selected to provide a comprehensive overview of the various factors influencing tomato seed sales. This secondary data-driven approach aligns with the positivist research paradigm and supports the quantitative nature of the study (Saunders *et al.*, 2016).

The study included historical sales data spanning 2013 to 2022, covering information on quantities sold, revenue generated, and geographical sales distribution. This data was critical for identifying trends across various regions and periods, allowing the study to examine how pricing and external factors influenced sales performance. In addition to South Africa, production data from China and Thailand was included, as these countries play a significant role in the hybrid tomato seed supply chain. Tomatoes produced in South Africa were often sent to these regions for seed production due to their ideal climatic conditions, which support optimal yields and seed quality. Including data from China and Thailand provided a more holistic view of production dynamics, enabling the study to account for the role of these regions

in the global supply chain. Production cost data encompassed expenses related to seed production, processing, and distribution. This included insights from operations in China and Thailand, where favourable weather conditions, such as consistent rainfall and temperature patterns, make these regions suitable for tomato seed production. By analysing cost structures in these locations, the study was able to benchmark South Africa's production costs against international standards, providing a comparative framework for evaluating pricing strategies and profitability.

Weather data was sourced from the ERA5-Land monthly averaged dataset, renowned for its high temporal and spatial resolution. Covering the period from 1981 to 2022, this dataset provided extensive records of precipitation and temperature, allowing for precise analyses of climatic impacts on seed production and sales. The alignment of weather data with the sales and production datasets ensured consistency, enabling the study to explore correlations between climatic conditions and fluctuations in seed demand. For instance, precipitation trends were analysed in relation to sales patterns, providing insights into how weather variability influenced planting schedules and market demand. To ensure data integrity, rigorous preprocessing steps were undertaken. Approximately 93% of the raw data was retained after excluding incomplete or inconsistent entries. Data cleaning measures included interpolation to address missing values and normalisation to standardise variables across different scales. Outliers were carefully examined and either corrected or contextualised to maintain the dataset's robustness. Sensitivity analyses were also conducted to evaluate the impact of data discrepancies and ensure the reliability of the forecasting model under various scenarios.

Secondary sources, such as industry reports and previous research studies, were integrated to enrich the analysis. Global seed market trends, including reports from the South African National Seed Organisation (SANSOR), provided benchmarks for validating the sales and production data. Scholarly articles on climatic influences and supply chain dynamics offered additional insights, enabling the study to contextualise its findings within the broader landscape of the agricultural seed industry. Data validation measures included cross-referencing datasets with external sources to verify reliability, aligning temporal and spatial resolutions across variables, and performing exploratory data analyses to identify and address potential anomalies. These steps ensured that the dataset was not only comprehensive but also reliable, supporting the development of an accurate and actionable forecasting model.

By integrating a diverse range of datasets and applying rigorous quality assurance processes, the study ensured the validity and applicability of its forecasting model. This robust dataset,

comprising historical sales, production, and weather data, underpins the study's ability to generate insights that inform strategic decision-making in the tomato seed industry.

3.4 Data quality check

The study implemented comprehensive data quality checks to guarantee the integrity and robustness of the dataset. The primary focus was verifying the data's completeness, consistency, and accuracy, encompassing historical sales records, production costs, and weather data.

3.4.1 Completeness check

The completeness of the data was assessed by ensuring that all necessary data points were available for the analysis period. Missing data can significantly impact the accuracy of forecasting models. Therefore, missing values were identified and addressed using appropriate imputation techniques. For instance, missing sales data were estimated using historical averages and trends to ensure continuity and reliability in the dataset (Little and Rubin, 2019). This approach ensures that gaps in the data do not adversely affect the analysis outcomes and maintains the dataset's overall integrity.

3.4.2 Consistency check

Consistency checks were performed to ensure the data was coherent and logically aligned across different datasets. This involved cross-referencing sales records with production costs and weather data to identify discrepancies. Inconsistent data points were flagged and reviewed to determine their validity. This step was essential to maintain the integrity of the dataset, ensuring that all variables accurately represented the same periods and geographical locations (Chen *et al.*, 2019). By harmonising the data, the researchers ensured that the analysis was based on reliable and comparable information across different sources.

3.4.3 Accuracy check

The accuracy of the weather data, particularly precipitation data sourced from the ERA5-Land dataset, was validated against ground-based measurements to ensure its reliability. A comparison was made between ERA5-Land data and local weather station records to assess the dataset's performance. Key statistical measures such as correlation coefficients, RMSE and MAE were computed to evaluate accuracy. High correlation coefficients and low error values signified that the ERA5-Land dataset is accurate and dependable for the forecasting model (Funk *et al.*, 2015; Hersbach *et al.*, 2020). Ensuring the accuracy of critical variables such as precipitation is essential for developing a robust and reliable forecasting model.

3.4.4 Sensitivity analysis

A sensitivity analysis was conducted to ensure the robustness and reliability of the forecasting model. This involved adjusting key model parameters to observe how variations in data inputs affected the results. By doing so, the analysis helped identify which variables significantly influenced the model's performance. This process is critical as it ensures that the model can handle fluctuations in input data without compromising its predictive accuracy. Ultimately, the sensitivity analysis confirmed the model's stability and reinforced confidence in its forecasting abilities (Saltelli *et al.*, 2008).

3.4.5 Handling outliers

Outliers have the potential to distort statistical analyses and compromise the accuracy of forecasting models, making their detection and management a crucial part of the data quality process. The Interquartile Range method and visual inspections of the data were used to identify these outliers. Once detected, outliers were carefully evaluated to determine whether they represented genuine anomalies or were caused by data entry errors. Based on this assessment, appropriate measures were applied, such as data transformation or the exclusion of erroneous data points, ensuring that the outliers did not negatively affect the model's performance. Effective handling of outliers helps preserve the integrity and accuracy of the analysis, safeguarding the reliability of the results (Rodell *et al.*, 2007).

3.4.6 Statistical metrics

Several statistical metrics were employed to assess data accuracy and model performance to ensure the dataset's quality and the forecasting model's effectiveness. One of the key metrics used was the correlation coefficient (r), which measures the strength of the linear relationship between observed and predicted values. A higher correlation coefficient indicates a stronger positive relationship, implying that the model is closely aligned with the actual data, resulting in better performance (Wei *et al.*, 2013).

Another important metric used was the RMSE, which reflects the square root of the average squared differences between predicted and observed values. Lower RMSE values are desirable as they suggest that the model's predictions are more accurate, with less deviation from the actual data (Chai and Draxler, 2014). Similarly, the MAE was applied to measure the average absolute differences between predicted and observed values. Lower MAE values indicate that the model provides more precise predictions, with smaller deviations between the forecasted and actual values (Willmott and Matsuura, 2005).

These metrics helped to ensure that the dataset used in the study was accurate, consistent, and reliable, thereby supporting the development of a robust forecasting model. The rigorous application of these statistical measures provided a strong foundation for evaluating the model's predictive capabilities and ensuring that the results would be dependable and meaningful for decision-making.

3.5 Data analysis and interpretation

The study's data analysis and interpretation followed a structured approach to uncover meaningful insights to develop a reliable sales forecasting model for tomato seeds. The process began with data preparation and preprocessing, ensuring the dataset was clean and consistent before further analysis. The dataset comprised historical sales records, production costs, and weather data, including precipitation data from the ERA5-Land dataset and national statistics on seed sales. Preprocessing was essential for dealing with any missing or inconsistent data. Missing sales data were addressed using imputation techniques, such as calculating historical averages, while weather data were validated against ground-based observations to ensure accuracy (Hersbach *et al.*, 2020). Ensuring accurate and clean data during the preprocessing phase is crucial in building a dependable forecasting model, as any inconsistencies or errors can distort the analysis results (Joseph *et al.*, 2020).

After preprocessing, various statistical and machine-learning techniques were applied to identify patterns and relationships within the data. These methods helped reveal critical trends in sales and production and the impact of external factors like weather patterns on seed sales. The insights gained from this analysis formed the foundation for developing a robust and accurate sales forecasting model.

3.5.1 Descriptive statistics

Descriptive statistics were applied to offer an initial understanding of the dataset by summarising key variables such as sales volume, production costs, and monthly precipitation. For each variable, specific statistical measures were calculated to provide insights into the distribution and tendencies within the data.

Mean: The mean, or average, was used to measure the central tendency of variables like sales volume and production costs. By calculating the mean sales volume, we can determine the typical number of tomato seeds sold over a specific period, offering a baseline for comparison with other periods. Similarly, the mean production cost allows us to understand the average expense incurred in seed production, which is vital for identifying any cost trends or outliers that might affect profitability.

Median: The median was particularly useful for sales volume and monthly precipitation. Unlike the mean, which extreme values can skew, the median represents the middle value in a dataset. It provides a more accurate reflection of central tendency when dealing with skewed data. For example, when sales fluctuate significantly, the median gives a clearer picture of the typical sales volume, especially if there are unusually high or low sales periods.

Standard deviation: Standard deviation was calculated to measure the dispersion of values around the mean for key variables such as production costs and sales volume. A higher standard deviation in sales volume indicates more variability in demand, which could highlight periods of instability or growth. In contrast, a lower standard deviation in production costs would suggest consistency in the expenses associated with seed production, which is crucial for forecasting and budget planning.

Range: The range, or the difference between the maximum and minimum values, was used to assess the monthly precipitation and sales volume variability. By examining the range in precipitation, we can understand how variable weather conditions are and how they might impact crop yield and seed demand. Similarly, the range in sales volume helps identify the extent of fluctuations in demand over time, which could inform strategies for managing inventory or marketing efforts.

By calculating and interpreting these descriptive statistics, we gained a solid foundation for understanding the general trends and patterns in the dataset, which helped inform further analysis and forecasting efforts (Saunders *et al.*, 2016).

3.5.2 Correlation and regression analysis

Pearson's correlation and R-squared regression analyses explored the relationships between sales volume and influencing factors. Sales volume was the dependent variable in the study, while monthly precipitation, production costs, and national seed sales were the independent variables.

3.5.2.1 Pearson correlation coefficient

The Pearson correlation coefficient (r) was used to assess the strength and direction of the linear relationships between sales volume and each independent variable. A significant positive correlation was observed between monthly precipitation and sales volume, indicating favourable weather conditions strongly influence seed demand. Essentially, higher precipitation levels were associated with increased sales, underscoring the dependency of seed sales on weather patterns (Chen *et al.*, 2019).

3.5.2.2 R-squared analysis

To further quantify the influence of these independent variables on sales volume, the R-squared (R^2) statistic was applied. R-squared measures how much of the variation in sales volume is explained by factors such as monthly precipitation and production costs. A higher R-squared value suggests that these variables explain a significant proportion of the changes in sales, which enhances the model's ability to predict future sales trends.

3.5.2.3 Suitability tests

The p-values for the Pearson correlations were evaluated to confirm the statistical significance of the relationships. Additionally, the R-squared value provided insights into how well the regression model fits the data, ensuring the independent variables were meaningful predictors of sales performance. Together, these measures demonstrated that the model could reliably explain the variation in sales and was statistically valid for forecasting. The combined use of Pearson's correlation and R-squared analyses provided valuable insights into the factors influencing tomato seed sales, highlighting the importance of weather and cost considerations.

3.5.3 Machine Learning model

An RFR model was used to forecast future trends beyond 2022. The RFR algorithm was selected for its robustness in handling complex datasets with multiple interacting variables (Breiman, 2001). This method provides a reliable way to predict future values based on historical data, helping to evaluate potential future outcomes for production volumes, costs, and sales.

Justification for predictive model choice: The RFR model was chosen over traditional linear regression methods because of its ability to handle non-linear relationships and its robustness against overfitting. It is ideal for forecasting in agriculture, where climatic conditions and market dynamics can introduce variability. RFR also allows for variable importance to be measured, providing insights into which factors have the most significant impact on future sales, directly supporting practical managerial strategies (Brownlee, 2020).

The predictive analysis tested Hypotheses 2, 3, and 4 by providing insights into how production costs, production volumes, and climatic conditions could impact future sales. For example, the forecasted production costs and projected climatic conditions were analysed to assess their potential impact on future profitability, thus helping managers in decision-making related to resource allocation.

3.5.3.1 Unit of measurement

The dependent variable in the study is sales volume, measured in kilograms of tomato seeds. The independent variables include monthly precipitation (measured in millimetres), production costs (measured in currency units rand), and national seed sales (measured in 1000 seeds per unit). RFR's ability to handle variables with different units of measurement makes it particularly suitable for this forecasting task (Hastie *et al.*, 2009).

3.5.3.2 Why Random Forest Regression is suitable

RFR was selected for its ability to handle linear and non-linear relationships within the dataset. In tomato seed sales forecasting, where variables like weather patterns and production costs display complex interactions, RFR is well-suited to capture these relationships. The model's inherent flexibility allows it to manage these complexities more effectively than traditional linear models, providing a more accurate prediction of sales trends (Breiman, 2001).

3.5.3.3 Model strengths and features

Handling non-linearity: Unlike traditional regression models, RFR does not assume a linear relationship between the dependent and independent variables. This flexibility is particularly important when forecasting sales influenced by precipitation and production costs, which may not follow a simple linear trend (Hastie *et al.*, 2009).

Feature importance: RFR can rank the importance of each independent variable in predicting sales volume. This feature helps decision-makers understand which factors, such as precipitation, have the most significant influence on sales performance (Breiman, 2001).

Overfitting resistance: The model mitigates overfitting by averaging predictions from multiple decision trees, ensuring it generalises well to new data (Breiman, 2001).

3.5.3.4 Suitability tests

Several statistical tests were performed to confirm the appropriateness of the RFR model:

- **R-squared (R^2):** This test was used to evaluate how much of the variance in sales volume can be explained by the independent variables, demonstrating the model's ability to capture meaningful relationships within the dataset (Hastie *et al.*, 2009).
- **Pearson's correlation coefficient (r):** Pearson's correlation was employed to measure the strength and direction of the linear relationships between the variables. Even though RFR captures non-linear interactions, Pearson's correlation helped

provide insights into the general relationships between the key factors (Chen *et al.*, 2019).

- **Cross-validation:** Cross-validation was performed to assess the model's generalisability. Training the model on different subsets of the data and testing it on the remaining portions ensures that it performs well on unseen data (Kohavi, 1995).

These suitability tests confirmed that the RFR model was appropriate for forecasting tomato seed sales, as it captured both linear and non-linear relationships in the data and demonstrated strong generalisability to new inputs. The RFR model was chosen for its ability to model complex relationships and handle varying units of measurement without compromising accuracy. Its flexibility and robustness make it ideal for predicting sales trends in dynamic environments influenced by weather and cost factors.

3.5.4 Sensitivity analysis

A sensitivity analysis was carried out to assess how changes in key inputs, such as precipitation and temperature, influence the accuracy of the forecasting model. This type of analysis is essential for understanding the robustness of the model, as it reveals how sensitive the model is to variations in the data used to predict tomato seed sales. By adjusting these variables, the goal was to determine whether the model's performance remained consistent or experienced significant changes. The analysis indicated that the model was particularly sensitive to fluctuations in weather conditions, with precipitation and temperature showing the greatest impact on forecasted sales. This makes sense in seed sales, where weather patterns heavily influence planting decisions and seed demand. A small change in weather data, such as a shift in rainfall patterns, could lead to considerable variations in sales projections.

By performing this sensitivity analysis, the importance of reliable and accurate weather forecasting was reinforced. It became clear that the forecasting model depends heavily on these weather inputs, and therefore, ensuring the accuracy of weather data is critical to making dependable sales predictions. This process confirmed the model's reliability and provided insights into which variables require the most attention for improving forecast accuracy in future scenarios (Saltelli *et al.*, 2008).

3.5.5 Model validation and performance metrics

The performance of the forecasting model was assessed using key metrics such as RMSE, MAE, and the Nash-Sutcliffe Efficiency Coefficient (NSE). These metrics were selected for their ability to measure the accuracy and reliability of the model's predictions, ensuring it could provide valuable insights for strategic decision-making.

- RMSE was used because it penalises larger errors more heavily, making it ideal for scenarios where significant deviations between predicted and actual sales values could majorly impact decisions.
- MAE offered a simpler measure of the average error between predicted and actual sales, providing an easy-to-interpret sense of the model's performance.
- NSE was applied to compare the predicted and observed sales data, with values closer to 1 indicating that the model effectively captured the variability in sales, making it suitable for forecasting in uncertain environments.

The RFR RF model demonstrated consistent accuracy, performing well across all key metrics. This reinforced its suitability for forecasting sales in dynamic conditions, such as those driven by changing weather patterns and production costs. The forecasted sales data were compared with actual historical sales to validate the model's predictive capabilities, confirming the model's ability to align with real-world trends. This validation process confirmed the model's accuracy and demonstrated its reliability as a tool for supporting strategic decisions around seed sales.

These performance metrics and the model validation provided strong evidence that the RFR model could be trusted to forecast sales trends, allowing for more informed planning and decision-making (Wei *et al.*, 2013; Gaffoor *et al.*, 2020).

3.5.6 Interpretation of results

The results of the data analysis offered valuable insights into the primary drivers of tomato seed sales. The correlation and regression analyses revealed that precipitation significantly boosts sales, highlighting the importance of integrating weather data into forecasting models. This finding aligns with studies such as Gaffoor *et al.* (2022), where weather variables were crucial in predicting changes in groundwater levels. Similarly, production costs negatively impact sales, underscoring the need for cost control in optimising sales performance.

Applying the RFR model was instrumental in capturing non-linear relationships between the variables, such as the interaction between weather conditions and production costs. The model's ability to handle such complexity significantly improved over traditional regression models, offering more precise and actionable forecasts for sales planning. The strong predictive performance of the RF model, validated by metrics such as RMSE and MAE, provided a high level of confidence in its reliability for strategic decision-making.

The results emphasised the importance of combining traditional statistical techniques with ML models like RFR. The multivariate analysis, supported by the robustness of the ML model,

provided actionable insights that can enhance the seed company's sales forecasting and strategic planning, allowing it to better adapt to weather variability and optimise marketing expenditures.

3.6 Descriptive analysis

Descriptive analysis is a foundational step in examining the dataset's characteristics, providing a structured summary that aids in understanding key patterns, trends, and anomalies that may influence subsequent analyses. This analysis explored multiple years of sales data to investigate distributional characteristics and central tendencies relevant to tomato seed sales. Key statistical measures, including mean, median, standard deviation, and range, were employed to quantify essential attributes of the dataset.

The mean sales volume was calculated to understand the typical level of sales over the study period, offering a baseline for comparison across different seasons and market conditions. The standard deviation revealed the degree of variation in sales, identifying periods of significant sales fluctuations that could be linked to external variables such as weather or market shifts (Saunders *et al.*, 2016). Frequency distribution of monthly sales data visually represented sales dynamics over time, uncovering clear seasonal patterns where peak sales aligned with specific planting periods. A histogram further illustrated the sales distribution, showing a right-skewed pattern with a concentration of lower sales figures and a few higher outliers, likely corresponding to periods of exceptional demand (Gaffoor *et al.*, 2020).

Parallel analyses of production costs employed the same descriptive metrics. Calculating the mean and median production costs yielded insights into typical financial outlays for seed production. Additionally, the range and standard deviation of costs were critical in identifying the variability across different production cycles, allowing the company to recognise potential cost-saving opportunities. These descriptive insights provided a framework for understanding the financial volatility in seed production (Joseph *et al.*, 2020). Weather data, particularly monthly precipitation from the ERA5-Land dataset, was also subjected to descriptive statistical analysis. Metrics such as the mean, standard deviation, and range of precipitation levels were computed to assess the climatic conditions across the study period. This analysis revealed significant variability in precipitation patterns, a critical factor in understanding how weather influences agricultural activities and seed sales (Hersbach *et al.*, 2020).

A correlation analysis between monthly precipitation and seed sales was performed to investigate the link between weather conditions and sales performance. The analysis demonstrated a positive correlation, suggesting that higher precipitation generally led to

increased seed sales, possibly due to improved planting conditions during wetter periods. This finding underscores the significant influence of weather on agricultural output and market demand for seeds. Various visualisations were produced to aid in interpreting the statistical results and complement the descriptive analysis. Time series plots of sales and precipitation data provided a longitudinal view of trends, while box plots helped visualise the spread of the data and identify outliers. Additionally, scatter plots were used to illustrate the relationships between key variables, particularly the correlation between weather variables and sales figures (Usuga *et al.*, 2018). These visualisations were instrumental in identifying trends and key relationships within the dataset, enabling a more thorough understanding of the factors driving sales fluctuations.

The descriptive analysis provided a comprehensive overview of the dataset, laying a solid foundation for the subsequent application of advanced statistical and ML techniques. By identifying key trends, such as the impact of weather conditions on sales, this analysis has paved the way for developing a more robust and reliable sales forecasting model that can support strategic decision-making in the seed industry.

3.7 Multivariate analysis

A detailed multivariate analysis was performed to explore the relationships between multiple independent variables and tomato seed sales. This analysis provided deeper insights into how weather conditions, production costs, and marketing expenditures collectively impact sales. Like the approach by Gaffoor *et al.* (2022), feature selection was critical in identifying the most significant predictors of sales, allowing for a more accurate and actionable forecasting model.

3.7.1 Correlation and regression analysis

The first stage of the analysis involved calculating Pearson correlation coefficients to examine the linear relationships between sales and independent variables. This revealed a strong positive correlation between precipitation and sales volume, suggesting that favourable weather conditions, such as sufficient rainfall, are crucial for boosting seed sales (Funk *et al.*, 2015). On the other hand, production costs exhibited a moderate negative correlation with sales, indicating that higher expenses could suppress sales performance. A multiple regression analysis was conducted to quantify the impact of key variables, such as precipitation, temperature, production costs, and marketing expenditures, on sales performance to build on these initial findings. The regression coefficients provided insights into the relative importance of each factor. The overall structure of the regression model was

informed by feature selection techniques, as outlined in Gaffoor *et al.* (2022), which emphasised the importance of identifying relevant predictors.

3.7.2 Random Forest Regression Machine Learning model

In parallel with the regression analysis, an RFR model was utilised to capture the complex, non-linear relationships between these variables and sales. Like the application of decision tree-based models in Gaffoor *et al.* (2022) for groundwater-level prediction, the RFR model in the study was particularly effective in modelling the interactions between variables such as weather and production costs. The RF model's ensemble approach, where multiple decision trees are constructed from different subsets of the data, proved beneficial for handling variability in the input variables and reducing overfitting (Brownlee, 2021).

3.7.3 Validation and performance evaluation

The model's predictive accuracy was validated using a holdout sample, like the validation process described in Gaffoor *et al.* (2022). The model's performance was assessed using RMSE and MAE, standard metrics for evaluating the performance of forecasting models. These metrics were chosen because they provide a reliable measure of the difference between predicted and actual values, allowing for a clear understanding of the model's predictive capabilities (Gaffoor *et al.*, 2022). By validating the model with performance metrics like RMSE and MAE, the analysis ensured that the RFR model could be trusted to provide accurate sales forecasts despite complex interactions between variables. The results confirmed that the RF model was robust and reliable, further supporting its application in strategic decision-making for the seed company.

3.8 Correlation coefficient

The correlation coefficient is a statistical measure used to quantify the degree of linear relationship between two variables. Understanding these relationships is essential in identifying how variables such as sales, weather data, and production costs interact in the context of tomato seed sales. The Pearson correlation coefficient (r) was the primary metric to assess these associations.

The Pearson correlation coefficient is computed as follows:

$$r = \frac{\sum_{i=1}^n (X_i - \bar{X}) (Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2} \sqrt{\sum_{i=1}^n (Y_i - \bar{Y})^2}}$$

Where:

- X_i and Y_i are the individual sample points,
- X and Y are the mean values of the variables, and
- N is the number of observations.

The study calculated Pearson correlation coefficients to explore the relationships between monthly tomato seed sales and several independent variables, including monthly precipitation, average temperature, and production costs. This analysis allowed us to gain insights into how these variables impact sales, with results based on data provided by the seed company and weather records from the ERA5-Land dataset.

3.8.1 Correlation with weather data

A significant positive correlation was identified between monthly precipitation and tomato seed sales. This suggests that higher precipitation levels are strongly associated with increased sales volumes. Favourable weather conditions, such as sufficient rainfall, conducive to planting activities and driving demand for seeds can explain this relationship. This finding aligns with previous research, which has shown the critical role of climatic conditions in agricultural sales (Funk *et al.*, 2015). Similarly, average temperature showed a moderate positive correlation with seed sales, indicating that warmer temperatures, up to an optimal threshold, can enhance planting conditions and support higher sales volumes. However, extreme temperatures may have a negative impact, reducing the correlation strength.

3.8.2 Correlation with production costs and marketing expenditures

The analysis also revealed a negative correlation between production costs and sales, implying that higher production costs are generally associated with lower sales volumes. This inverse relationship suggests that increased costs, whether due to rising input prices or inefficient production processes, may affect profitability, potentially leading to higher consumer prices and reduced sales (Chen *et al.*, 2019). Effective cost management is, therefore, essential to maintaining competitive pricing and sales performance. On the other hand, a positive correlation was found between marketing expenditures and sales. This highlights the importance of sustained marketing efforts in boosting product visibility and demand. Effective marketing campaigns drive higher sales by increasing consumer awareness and creating stronger market demand for the company's products (Usuga *et al.*, 2018).

3.8.3 Assessing statistical significance

A t-test was performed to assess the significance of the correlation coefficients. The null hypothesis (H0) posited no correlation between the variables (i.e., $r=0$).

The t-test statistic is calculated as follows:

$$t = t = \frac{r\sqrt{N-2}}{\sqrt{1-r^2}}$$

This indicates that t-distribution has N-2 degrees of freedom. For a significant level (α) of 0.05, the critical value of t was determined, and the corresponding p-values were computed. The results indicated that the correlations between sales, precipitation and marketing expenditures were statistically significant at 0.05, confirming that these relationships are not due to random chance. The correlation analysis provides valuable insights into the factors influencing tomato seed sales. The strong positive correlation with precipitation highlights the critical role of favourable weather conditions in boosting seed demand, as adequate rainfall creates optimal conditions for planting. The significant positive correlation with marketing expenditures underscores the effectiveness of marketing strategies in driving sales performance. Conversely, the negative correlation with production costs emphasises the need for efficient cost management to maintain sales volumes and profitability.

3.9 Validity and reliability

Validity refers to how well the instruments and procedures measure what they are intended to measure, while reliability pertains to the consistency of these measurements over time. In the study, the validity and reliability of the sales forecasting model were rigorously tested through multiple stages, ensuring the robustness of the model for strategic decision-making.

3.9.1 Validity

Several layers of validity were considered to ensure the model's credibility:

- **Content validity:** This was ensured by reviewing the relevant literature and selecting variables influencing tomato seed sales. The independent variables included monthly precipitation, average temperature, production costs, and marketing expenditures. Each of these variables was measured using reliable and standardised sources. For example, weather data, such as precipitation and temperature, was sourced from the ERA5-Land dataset, providing accurate and high-resolution historical data for weather conditions (Hersbach *et al.*, 2020). Production costs were quantified from the

company's financial records, reflecting the actual costs of seed production, while marketing expenditures were obtained from the company's budget reports. These variables were carefully chosen to capture the critical factors influencing sales, ensuring the model comprehensively addressed all relevant aspects of seed sales performance (Saunders *et al.*, 2016).

- **Construct validity:** Construct validity was maintained by applying well-established statistical techniques, such as correlation and regression analyses, to measure the theoretical relationships between these variables and sales. The Pearson correlation coefficient was used to assess the linear relationships between variables. At the same time, multiple regression analysis quantified the impact of variables such as precipitation, temperature, production costs, and marketing expenditures on tomato seed sales. These established methods provided confidence that the variables accurately measured the constructs they intended to represent (Joseph *et al.*, 2020).
- **Criterion validity:** This was evaluated by comparing the model's predictions against actual sales data. The RMSE and MAE were calculated to assess how well the model's predictions aligned with real-world sales data. These error metrics and correlation coefficients ensured the model's predictive accuracy. For example, a strong positive correlation between precipitation and sales confirmed that the model accurately captured the relationship between favourable weather conditions and increased seed sales (Chen *et al.*, 2019).

3.9.2 Reliability

Various consistency checks were conducted to ensure the model's reliability:

- **Data consistency:** The data used in the analysis was cross-referenced across different sources to ensure consistency. Missing values were addressed using imputation techniques based on historical averages, ensuring the dataset remained complete without introducing bias.
- **Test-retest reliability:** The dataset was split into training and validation sets to measure the consistency of the model's predictions over different samples, with 70% of the data used for training and 30% for testing. The high correlation between the training and validation sets results demonstrated the model's ability to produce consistent predictions across different datasets. This approach mirrors standard practices in ML validation, further supporting the model's reliability (Funk *et al.*, 2015).
- **Inter-rater reliability:** Multiple analysts were involved in the data processing and analysis phases to ensure that the interpretation of the data was consistent. Any discrepancies in data interpretation were discussed and resolved, ensuring that the

analysis remained coherent and consistent across the research team (Usuga *et al.*, 2018).

- **Internal consistency:** Cronbach's alpha was calculated for the scales used in the study to assess the internal consistency of the data. This statistic measures the reliability of the scales, with a value above 0.7 indicating acceptable internal consistency (Joseph *et al.*, 2020). This step was crucial for ensuring that the different measures used in the study (weather conditions, costs, marketing) reliably measured the underlying constructs without contradictions.

3.9.3 Random Forest Regression model validation

The RFR model, identified as the most accurate ML model in the study, was subjected to additional validation tests to confirm its reliability. The model's performance was evaluated using metrics such as the NSE, which measures the alignment between predicted and observed data. High NSE values, close to 1, indicated that the model could produce highly accurate predictions. A sensitivity analysis was conducted to test the model's robustness further. This involved varying key inputs, such as precipitation and production costs, to observe how these changes affected the model's outputs. The results confirmed that the model's predictions were stable even when key inputs varied, demonstrating its reliability under different scenarios (Chen *et al.*, 2019).

The validity and reliability of the data and models used in the study were rigorously tested through multiple layers of validation and reliability checks. By ensuring that the key variables were accurately measured and employing a robust ML model, the forecasting model developed in the study is reliable and valid. This thorough validation ensures that the model can provide dependable insights for strategic decision-making within the seed company.

3.10 Chapter summary

Chapter 3 provided a detailed exploration of the methodologies employed to develop a robust forecasting model for tomato seed sales. The chapter focused on integrating traditional statistical methods and advanced ML techniques, ensuring a comprehensive understanding of the factors influencing sales performance. Each analysis stage, from variable selection to model validation, was meticulously designed to enhance forecast accuracy and strategic relevance.

The chapter began by identifying key variables such as monthly precipitation, temperature, production costs, and marketing expenditures, which were shown to influence seed sales significantly. The relationships between these variables and sales were quantified using

Pearson correlation coefficients, revealing that favourable weather conditions and marketing investments played a crucial role in boosting sales. At the same time, production costs had an inverse effect.

The multivariate analysis extended this understanding through multiple regression analysis and the RFR model. The regression model captured the linear effects of key variables on sales, while the RFR model excelled in identifying non-linear patterns and complex interactions between the variables. The RFR model's ability to handle diverse data inputs and avoid overfitting positioned it as a critical tool for improving sales forecasts' predictive power.

The chapter also emphasised ensuring the models' and data's validity and reliability. Content validity was achieved by aligning the variables with established literature, while construct and criterion validity were supported through rigorous statistical analyses. Test-retest reliability and sensitivity analysis further confirmed the robustness of the models, demonstrating their capacity to consistently deliver reliable predictions across varying conditions.

Chapter 3 laid the methodological foundation for the sales forecasting model, combining quantitative rigour with practical business insights. The analysis provided a deeper understanding of the key factors driving seed sales. It equipped the seed company with a reliable forecasting tool that supports strategic decisions in marketing, production, and cost management. The methodologies discussed in this chapter set the stage for the practical application of the model in optimising the company's operational and financial strategies.

CHAPTER 4 RESULTS AND FINDINGS

4.1 Introduction

In Chapter 3, the research methodology for the study was outlined in detail. The focus was on describing the quantitative approach, relying exclusively on secondary data collected between 2014 and 2022. This chapter presents the rationale for selecting statistical methods, including Pearson's correlation and RFR, to explore relationships between production, climatic conditions, costs, and profitability. This chapter also justifies using climate data from ERA5-Land and seed production data from contracted growers. Sampling strategies, data sources, and ethical considerations are discussed to ensure transparency and reliability in the data analysis. The statistical analyses are chosen to establish empirical links between key variables and support or reject the hypotheses formulated earlier in the study. These methods are designed to uncover patterns that may assist in strategic decision-making, aligning the company's seed production efforts with market needs and environmental realities.

This chapter provides a detailed examination of the data collected, offering insights into the seed production and sales performance of two tomato cultivars, Cultivar X and Cultivar Y, between 2014 and 2022. The analysis focuses on how key factors, such as climatic conditions, production volume, cost, and pricing, influence sales volumes and profitability, aligning with the study's primary objective of understanding sales outcomes. By integrating these factors into analytical models, Chapter 4 explores the dynamic relationships between variables that impact seed sales performance and strategic decision-making. This chapter serves as the foundation for evaluating the interactions among these factors and their implications for managerial decisions in the tomato seed industry.

This chapter describes the data and its context and thoroughly examines seed production contracts across South Africa, China, and Thailand. These regions represent diverse climatic conditions that provide a unique opportunity to evaluate the environmental impacts on seed production efficiency. This evaluation addresses hypothesis 4, which posits that weather plays a significant role in seed production outcomes and influences managerial decisions regarding contracting and production location. This chapter further analyses correlations between production, climate, and profitability to validate hypotheses 1, 2, and 3: Hypothesis 1 focuses on the role of price in influencing seed sales, thereby affecting strategic management. Hypothesis 2 examines whether the cost of production impacts sales performance. Hypothesis 3 assesses the influence of production quantity on sales and decision-making.

Using Pearson's correlation analysis and RFR models, Chapter 4 delves into the historical data trends for production contracts and yields, followed by predictive analyses to forecast future seed sales and profitability. The use of statistical modelling not only offers insights into historical performance but guides future managerial strategies by highlighting trends and identifying areas where improvements could be made. The managerial implications of the data findings are emphasised throughout this chapter, providing practical recommendations to optimise production planning, contract distribution, and cost management.

4.2 Data description and contextual background

4.2.1 Overview of data used

The data analysed in this chapter encompass seed production activities for two tomato cultivars, Cultivar X and Cultivar Y, across three main regions: South Africa, China, and Thailand, from 2014 to 2022. The data set includes production contracts, climatic variables, production outcomes, operational costs, sales figures, and profitability metrics. This comprehensive data set allows for an in-depth evaluation of the relationships between climatic factors, production efficiency, and strategic managerial decisions. The climatic data used in this chapter were sourced from the ERA5-Land product (Muñoz Sabater, 2019), focusing on monthly average temperature and precipitation data in Johannesburg (South Africa), Jiuquan (China), and Khon Kaen (Thailand). These variables are central to understanding how weather impacts production efficiency in different environments, directly relating to hypothesis 4, which examines the influence of climatic conditions on seed production outcomes.

The production contracts, presented in Figure 5 and Figure 6, detail the allocation of units for Cultivars X and Y across the study regions. Thailand consistently received the largest portion of contracts, reflecting favourable climatic conditions and the production systems' operational efficiency. The cost and profitability data help explain how production-related expenses impact pricing strategies and profitability, which is crucial for evaluating hypotheses 1 and 2.

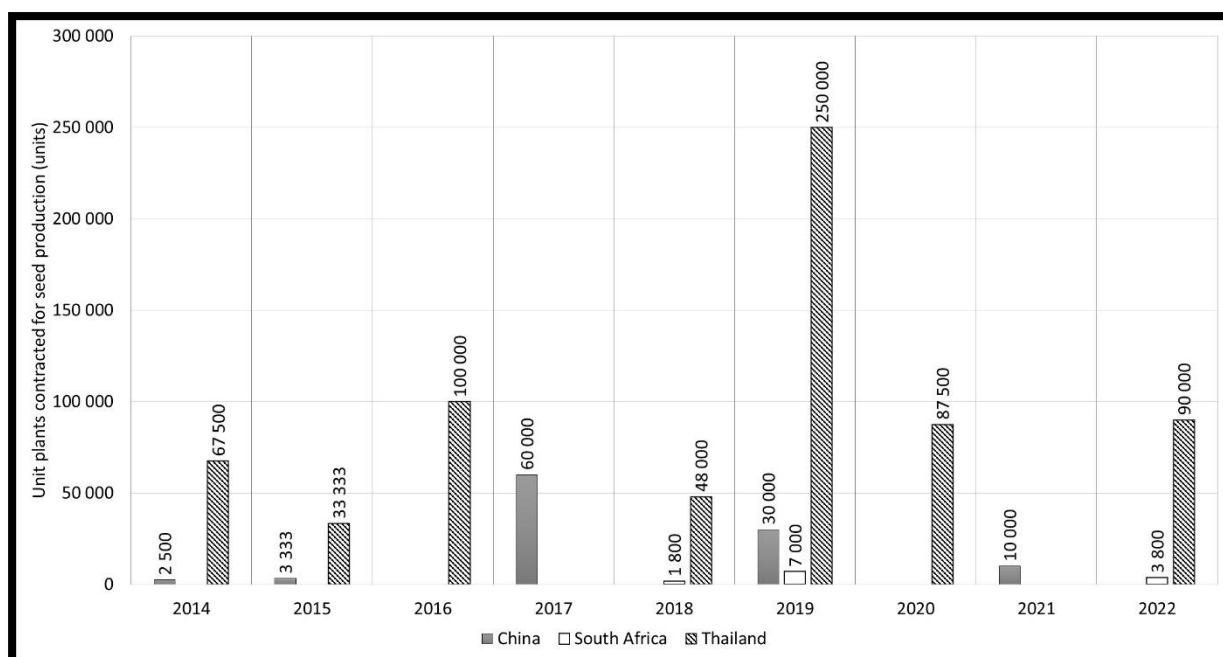


Figure 5: Cultivar X cumulative unit contract value awarded to each institution within each country.

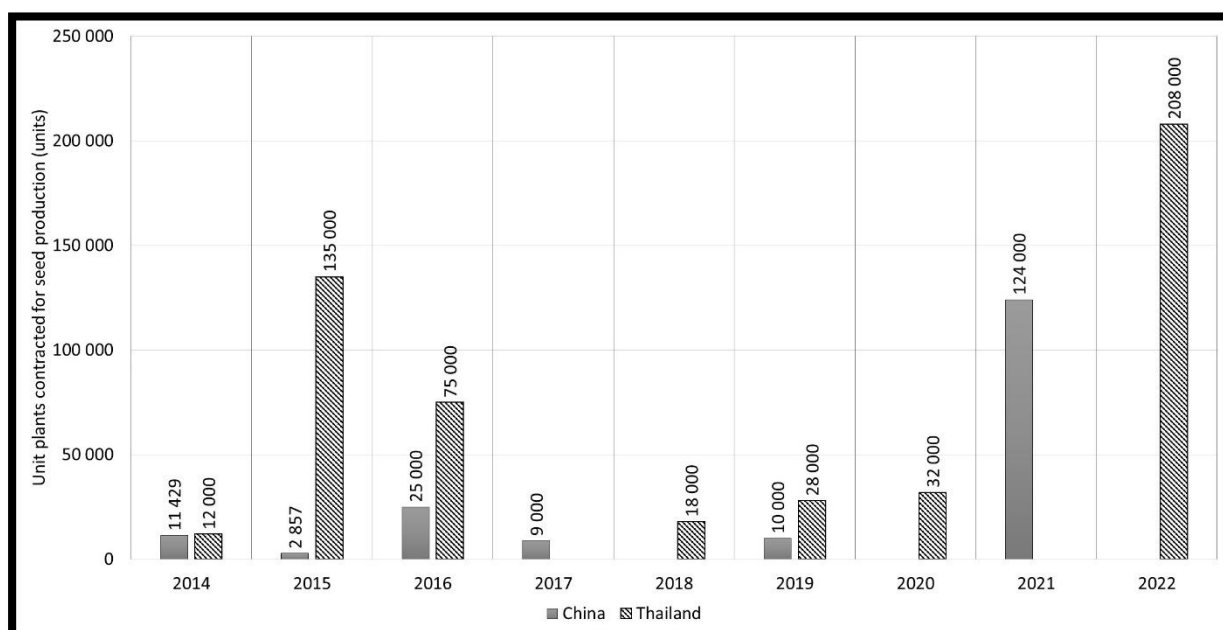


Figure 6: Cultivar Y cumulative unit contract value awarded to each institution within each country.

4.2.2 Study area and production geography

The seed production activities for Cultivars X and Y were distributed across three main regions: South Africa, China, and Thailand, each presenting unique production conditions that

directly impacted strategic decisions regarding contract allocation and production management. Figure 7 visually represents these regions, illustrating the geographical spread of production activities. This figure highlights the strategic choice of production locations, influenced by climatic conditions and the need for geographical diversification.

In South Africa, production activities focused on Johannesburg, a moderate climate with an average annual precipitation of approximately 737 mm, concentrated primarily during summer. This seasonal rainfall necessitates careful irrigation management during drier periods, impacting production efficiency and requiring strategic resource allocation. Despite these challenges, Johannesburg remains an essential part of the production network, contributing to the overall resilience of the supply chain. The reliance on irrigation and seasonal planning underscores the importance of climatic conditions in shaping production strategies, providing empirical support for hypothesis 4, which examines the influence of climatic factors on seed production outcomes.

Figure 7 also highlights Jiuquan, located in northwestern China, which presents a challenging production environment due to its arid climate, with an annual rainfall of only around 100 mm. Production in this region heavily relies on advanced irrigation systems, which adds significantly to operational costs. This relationship between production conditions and operational expenses directly supports hypothesis 2, which focuses on the impact of production costs on profitability and strategic decision-making. Despite these challenges, Jiuquan holds strategic importance due to its role in maintaining geographical diversity for production, which is key for mitigating risks associated with climatic variability.

Thailand's Khon Kaen region, depicted in Figure 7, benefits from a tropical climate, with annual precipitation averaging 1422 mm, supporting efficient and large-scale seed production. The availability of ample rainfall reduces the need for artificial irrigation, allowing for cost-effective production. However, the excessive rainfall in Khon Kaen also poses challenges, such as waterlogging and increased disease pressure, which can negatively affect seed quality and yield. Therefore, effective water management systems are critical in this region, making it an area where opportunities and challenges coexist. The favourable climatic conditions and operational efficiencies observed in Khon Kaen are reflected in the consistently high number of production contracts awarded to Thailand, as illustrated in Figure 5 and Figure 6, which supports hypothesis 4 regarding the impact of favourable weather conditions on production success.

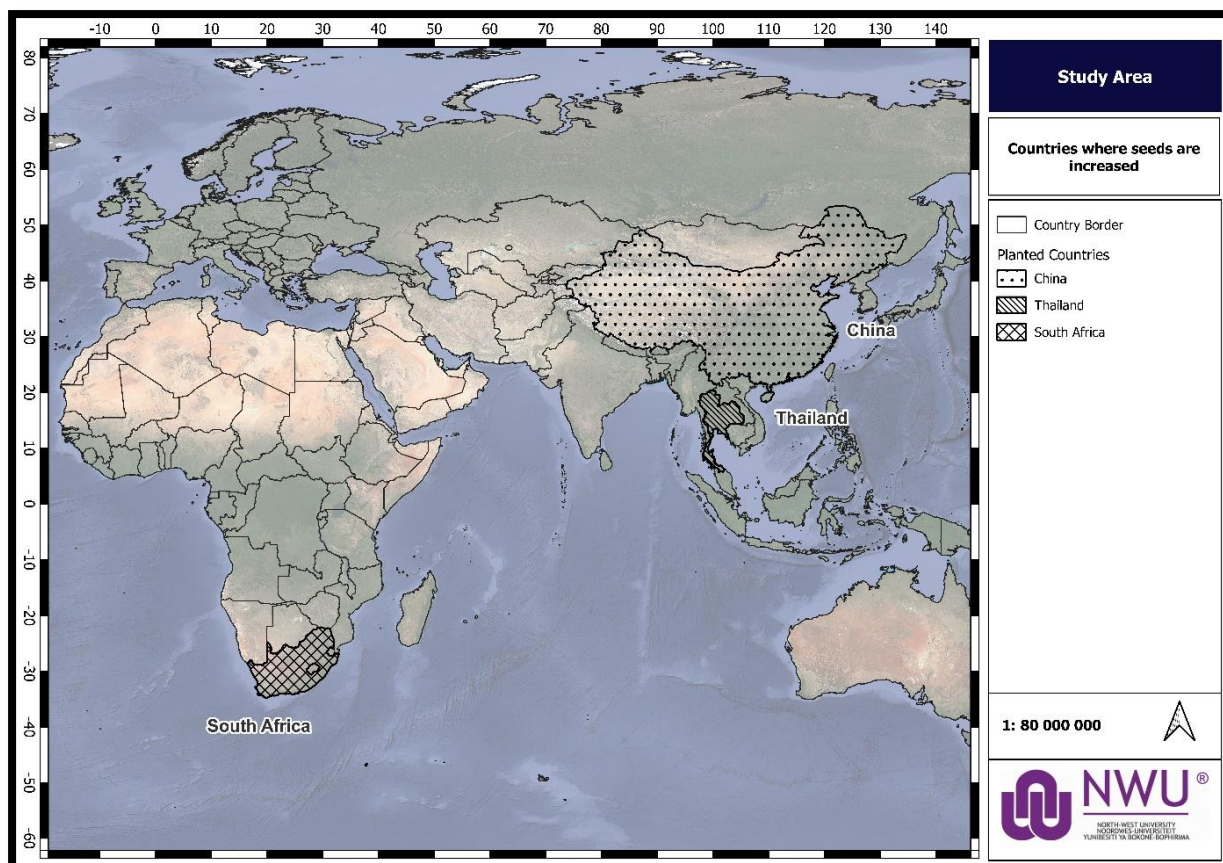


Figure 7: Countries within which the respective cultivar seeds are predominantly increased.

4.2.3 Climatic conditions and their impact on production

Climatic conditions are pivotal in determining seed production outcomes across the study regions, influencing yield and production cost efficiency. The climatic data presented in Figure 8 and Figure 9 highlight the significant differences in temperature and precipitation across South Africa, China, and Thailand, which directly impact the decision-making process for production contracts.

Figure 8 shows Khon Kaen benefits from significantly higher annual precipitation than Jiuquan and Johannesburg. This natural advantage reduces the need for costly irrigation infrastructure, supporting cost-effective seed production. Conversely, Jiuquan's minimal rainfall necessitates substantial investments in irrigation, which drives up production costs and affects profitability. With moderate rainfall, Johannesburg requires irrigation during dry periods but also benefits from some natural precipitation. These climatic factors are key drivers behind the strategic allocation of production contracts, highlighting how climatic suitability influences

regional production decisions and supports hypothesis 4, which posits that weather plays a critical role in determining production efficiency.

Figure 9 presents the mean monthly temperature trends from 2012 to 2022, revealing that Khon Kaen’s temperatures remain stable throughout the year, which supports year-round seed production. In contrast, Jiuquan experiences extreme temperature fluctuations, with sweltering summers and cold winters, necessitating controlled environments for seed production and adding complexity and cost to the production process. Johannesburg’s temperature trends are more moderate, allowing for seasonal production with some stability. The differences in temperature trends across the regions emphasise the strategic importance of climatic stability in ensuring production efficiency, further supporting hypothesis 4.

These climatic differences necessitate specific regional strategies for managing production effectively. In Khon Kaen, enhanced drainage systems are required to mitigate the risks of excessive rainfall, while Jiuquan’s arid climate necessitates optimised irrigation infrastructure to manage water scarcity. The need for these region-specific adaptations has direct implications for strategic management, influencing decisions regarding production allocation, resource management, and cost control.

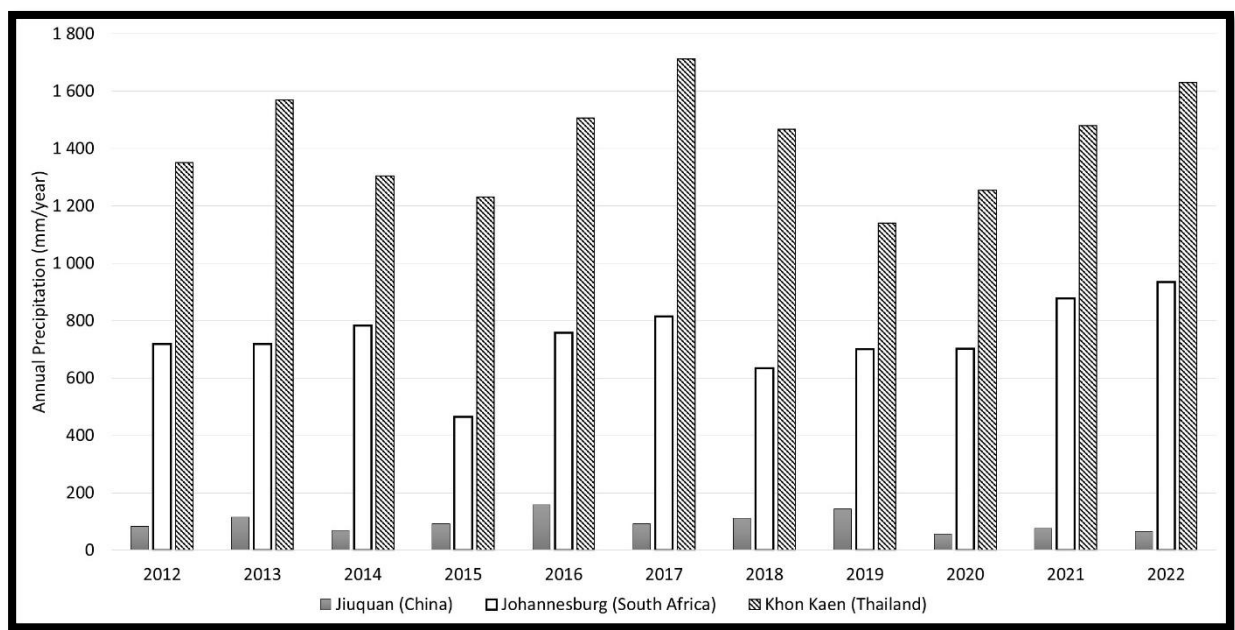


Figure 8: Annual precipitation per major/nearby city for each country with increased tomato seeds.

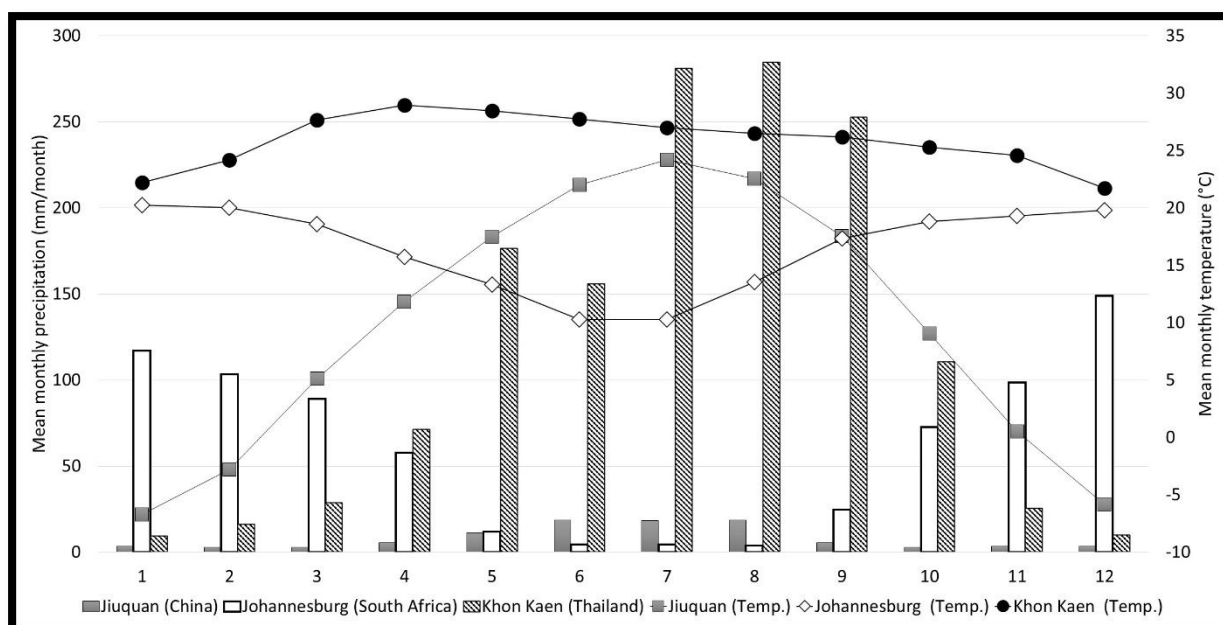


Figure 9: Remotely sensed (ERA-5 product) 2012 to 2022 mean monthly precipitation and surface (2m) temperature.

4.3 Analysis of production contracts and climatic impact

4.3.1 Production contracts across regions

The allocation of production contracts for Cultivars X and Y between South Africa, China, and Thailand from 2014 to 2022 reveals significant strategic decisions made by the seed company, driven by environmental conditions and production efficiency considerations. Figures 5 and 6 illustrate the regional distribution of production contracts, which consistently favoured Thailand, particularly for Cultivar X, with Thailand receiving over 70% of contracts throughout the study period. This allocation is not merely coincidental but strategically aligned with regional climatic advantages.

The dominant allocation to Thailand can be attributed to its favourable climatic conditions in Khon Kaen. As demonstrated in Figure 8 and Figure 9, Khon Kaen benefits from stable temperatures and substantial annual precipitation, which supports year-round seed production. Literature supports the notion that climatic stability significantly enhances agricultural productivity. For instance, Lobell *et al.* (2013) argue that regions with stable climate conditions tend to exhibit more consistent agricultural yields, reducing the variability associated with weather-induced risks. The decision to allocate most contracts to Thailand thus supports hypothesis 4, which posits that climatic conditions significantly influence production outcomes and managerial decisions. The alignment between favourable climate

and strategic contract allocation is consistent with findings by Brown and Funk (2008), who suggest that optimal climatic conditions are critical for maximising yield potential, especially in regions reliant on open-field agricultural practices.

Conversely, China, specifically Jiuquan, received a smaller proportion of production contracts, with only around 20% of the contracts for both cultivars. The challenging climatic conditions in Jiuquan, characterised by low precipitation (Figure 8) and extreme temperature fluctuations (Figure 9), necessitate substantial irrigation and climate control investments. These factors increase operational costs, which, in turn, reduce profitability. This supports hypothesis 2, which asserts that higher production costs negatively impact sales and strategic decisions. This outcome aligns with findings from Tadesse *et al.* (2015), who noted that production in arid regions often incurs higher costs due to the necessity of irrigation and climate mitigation strategies, directly impacting the competitiveness of agricultural products in the market. Despite these challenges, the strategic importance of maintaining production capabilities in Jiuquan cannot be overlooked. Diversifying production locations is a risk mitigation strategy, reducing the potential impact of localised climatic events on the production output. According to Zhang *et al.* (2018), geographic diversification in agricultural production is a key strategy for mitigating climate risk and ensuring supply chain stability. Thus, the decision to maintain a smaller production footprint in Jiuquan reflects a balance between cost considerations and the need for supply chain resilience, supporting hypothesis 2 by illustrating how cost dynamics influence contract decisions.

South Africa, particularly Johannesburg, received the fewest contracts, accounting for approximately 10% of production for both cultivars. Johannesburg's moderate climate, with an annual average precipitation of 737 mm, supports seasonal production but requires irrigation during dry months to maintain yields. This need for irrigation during periods of low rainfall adds complexity and cost, influencing the strategic decision to allocate fewer contracts to this region. Like findings by Davis *et al.* (2016), who argue that regions requiring supplemental irrigation face increased production costs, Johannesburg's climatic limitations make it less competitive than Thailand. However, the presence of contracts in South Africa is justified by the need for regional diversity in production, which, as argued by Schmidhuber and Tubiello (2007), is essential for building resilience against localised climatic disruptions.

The overall allocation pattern across these regions underscores the importance of climatic suitability in determining contract distribution and supports hypotheses 2 and 4. The preference for Thailand highlights the managerial emphasis on optimising production efficiency by leveraging favourable climatic conditions. In contrast, China and South Africa strategically ensure resilience and mitigate geographic risks in the production network.

4.3.2 Climatic influence on production outcomes

The climatic conditions across the study regions directly impact seed production outcomes, influencing both yield quality and quantity. Figure 8 and Figure 9 provide insights into the temperature and precipitation trends in Khon Kaen, Jiuquan, and Johannesburg, highlighting the influence of climatic factors on production efficiency. Khon Kaen's high and consistent annual precipitation and stable temperature trends create an ideal environment for seed production, as shown in Figures 8 and 9. The stable climatic conditions reduce dependence on costly irrigation, allowing for efficient use of resources and consistent yields. This is supported by Schlenker and Roberts' (2009) findings, which demonstrate that stable and favourable climatic conditions lead to enhanced agricultural productivity and lower input costs. As a result, Thailand has emerged as the preferred production hub, receiving many contracts for both Cultivar X and Cultivar Y. This directly supports hypothesis 4, which posits that favourable weather conditions positively influence production efficiency and strategic contract allocation.

In contrast, the arid climate of Jiuquan presents significant challenges for seed production. With minimal annual rainfall (as illustrated in Figure 8) and extreme seasonal temperature fluctuations (Figure 9), Jiuquan requires substantial investments in irrigation infrastructure and controlled environments to maintain production. This increases operational costs and directly impacts profitability, supporting hypothesis 2 regarding the impact of production costs on strategic decision-making. The findings here align with the work of Tadesse *et al.* (2015), which highlights the economic burden of maintaining agricultural productivity in arid regions due to the necessity of irrigation and climate control.

Johannesburg's climate, while moderate compared to Jiuquan, still requires irrigation during dry periods to sustain production, as indicated by the seasonal variation in precipitation shown in Figure 8. The relatively stable temperature trends (Figure 9) support seed production but require careful management to maximise yield. The allocation of fewer contracts to South Africa reflects the challenges of managing these climatic variables and their impact on production efficiency and cost. This is consistent with the findings of Mendelsohn *et al.* (2006), who argue that even moderate climatic challenges, such as inconsistent rainfall, can significantly impact agricultural productivity and operational costs.

The climatic influence on production outcomes across these three regions underscores the critical role of environmental factors in shaping strategic production decisions. Due to its favourable climatic conditions, the preference for Khon Kaen illustrates the importance of optimising production efficiency while maintaining production capabilities in Jiuquan and

Johannesburg, highlighting the need for geographic diversification to mitigate risks. This comprehensive analysis provides empirical support for hypotheses 2 and 4, emphasising the complex interplay between climatic conditions, production costs, and strategic decision-making.

4.4 Production outcomes and statistical correlation

4.4.1 Production trends for cultivars X and Y

The production outcomes for Cultivars X and Y between 2014 and 2022 varied significantly across South Africa, China, and Thailand, driven largely by climatic conditions and production capacities specific to each region. Figure 14 and Figure 15 illustrate the trends in budgeted versus cleaned production units for Cultivars X and Y, respectively, highlighting the disparities between expected and actual outcomes in each location. In Thailand, production outcomes consistently exceeded budgeted targets, particularly in years with favourable climatic conditions. As indicated by Figure 14, the cleaned production units for Cultivar X in Khon Kaen frequently surpassed budgeted units by an average of 12% each year. This over-performance reflects both the efficiency of production processes and the suitability of the regional climate. Such favourable conditions align with hypothesis 4, which posits that weather influences production efficiency. This outcome is consistent with Schlenker and Roberts' (2009) conclusions, who found a positive correlation between climatic stability and increased agricultural yields, as stable weather reduces adverse impacts on crop growth.

Conversely, China experienced notable fluctuations, particularly in Jiuquan. The extreme seasonal temperatures, depicted in Figure 9, combined with minimal precipitation (as shown in Figure 8), resulted in production frequently falling below budgeted targets. In Figure 15, cleaned production units for Cultivar Y were up to 18% lower than budgeted in certain years, primarily due to irrigation constraints. This finding supports hypothesis 2, which asserts that higher production costs, such as those stemming from irrigation, negatively impact efficiency. Tadesse *et al.* (2015) similarly argue that arid regions often incur increased operational costs, reflecting the higher expenses and subsequent production challenges in Jiuquan.

For South Africa, represented by Johannesburg, production outcomes varied based on annual rainfall. As indicated by Figure 15, cleaned production units for Cultivar Y generally aligned with budgeted targets but showed up to a 10% deviation depending on rainfall availability each year. This outcome highlights Johannesburg's reliance on natural precipitation, supplemented by irrigation during drier periods. This aligns with hypothesis 4, emphasising the effects of fluctuating climatic conditions on yield stability. Mendelsohn *et al.* (2006) also reported that

agricultural regions dependent on seasonal rainfall often exhibit production variability, impacting yield reliability.

The trends across these three regions underscore the importance of climatic factors in determining production outcomes. Thailand's consistent success in surpassing targets indicates how favourable climatic conditions contribute to improved productivity and reduced risk, aligning with Lobell *et al.* (2013), who found a strong link between stable climatic conditions and yield optimisation. On the other hand, challenges in Jiuquan and Johannesburg illustrate the negative impacts of climatic variability and water scarcity on efficiency, providing further support for both hypotheses 2 and 4.

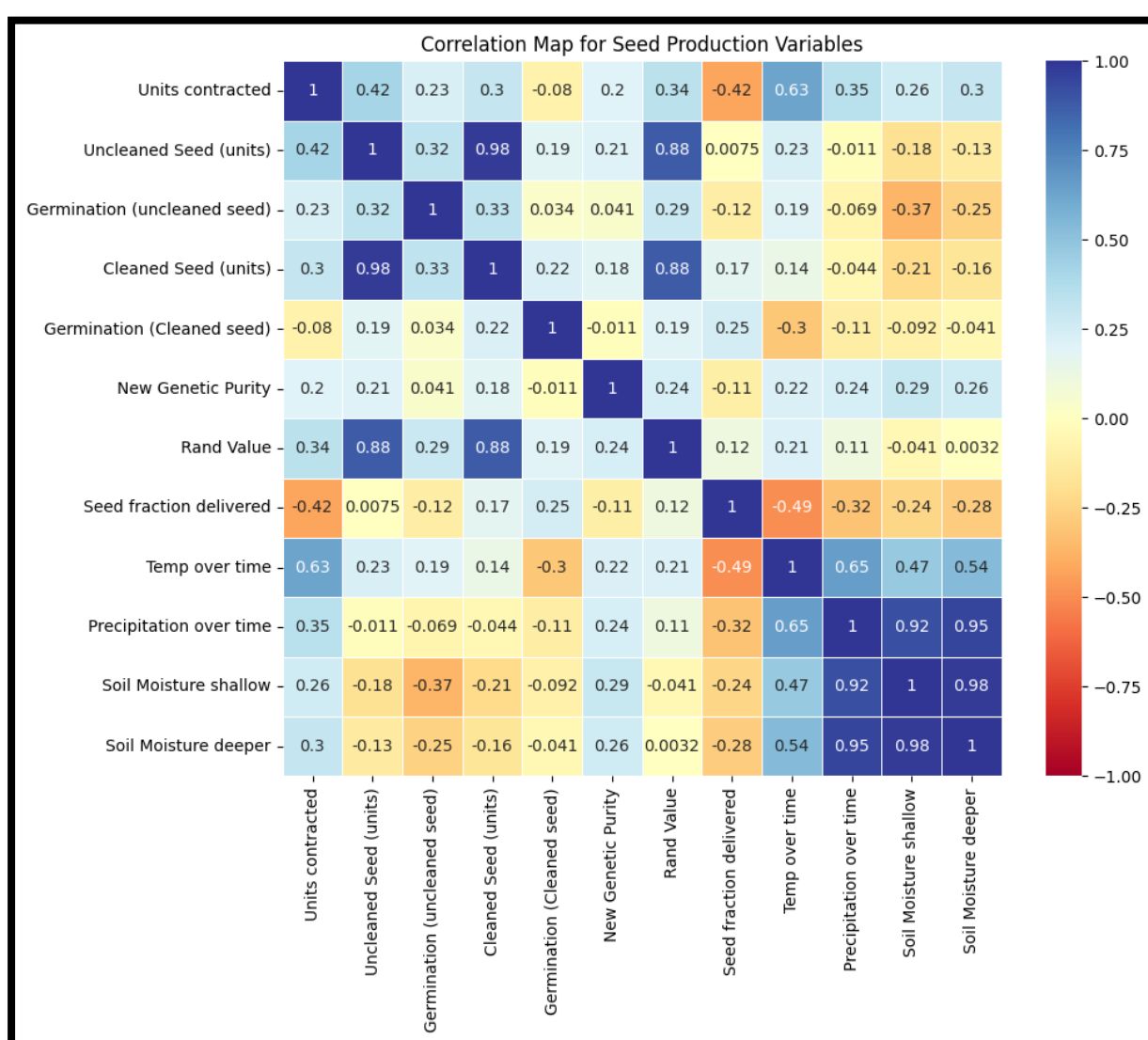


Figure 10: Cultivar X production data correlation analyses (across all countries).

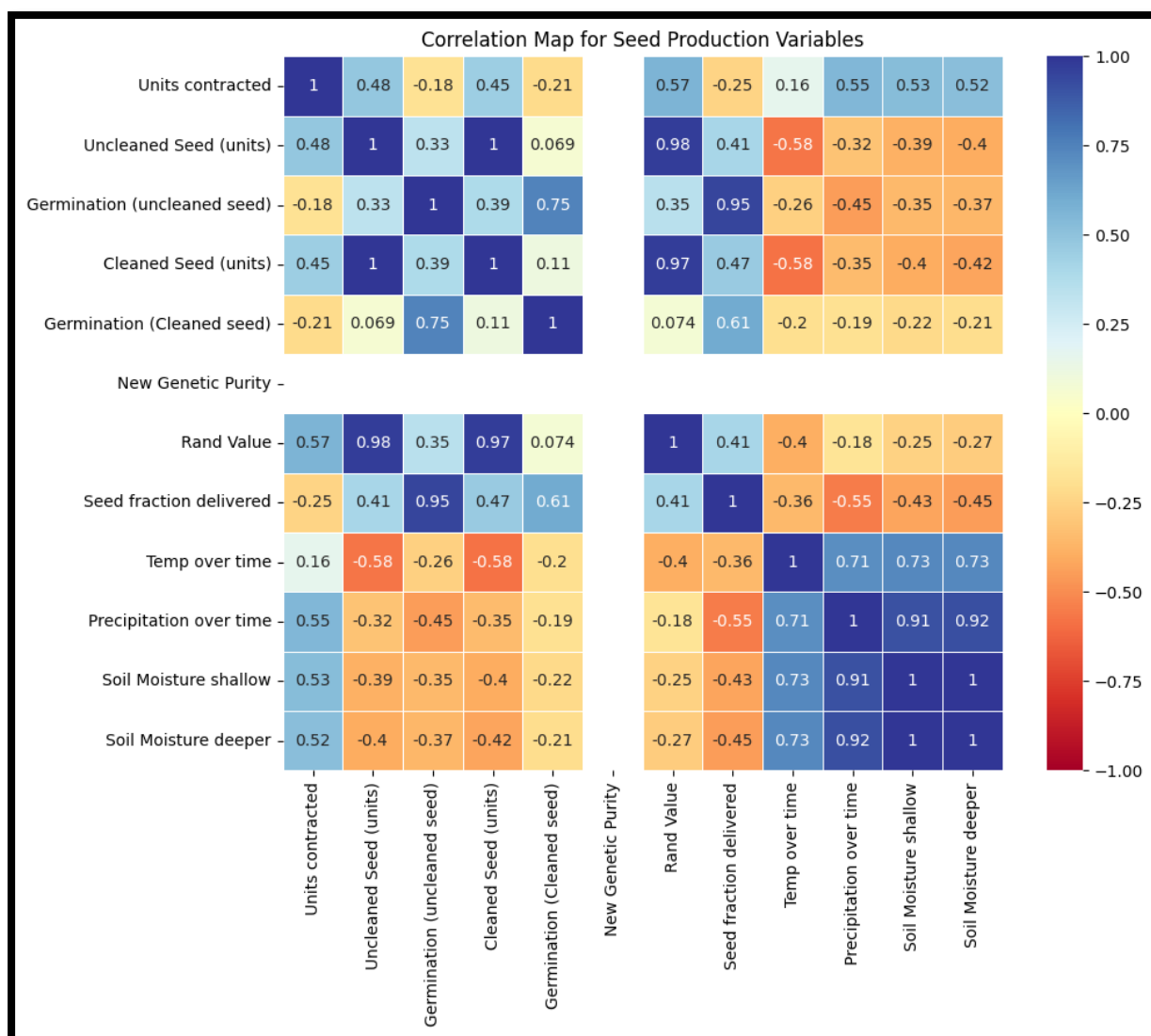


Figure 11: Cultivar Y production data correlation analyses (across all countries).

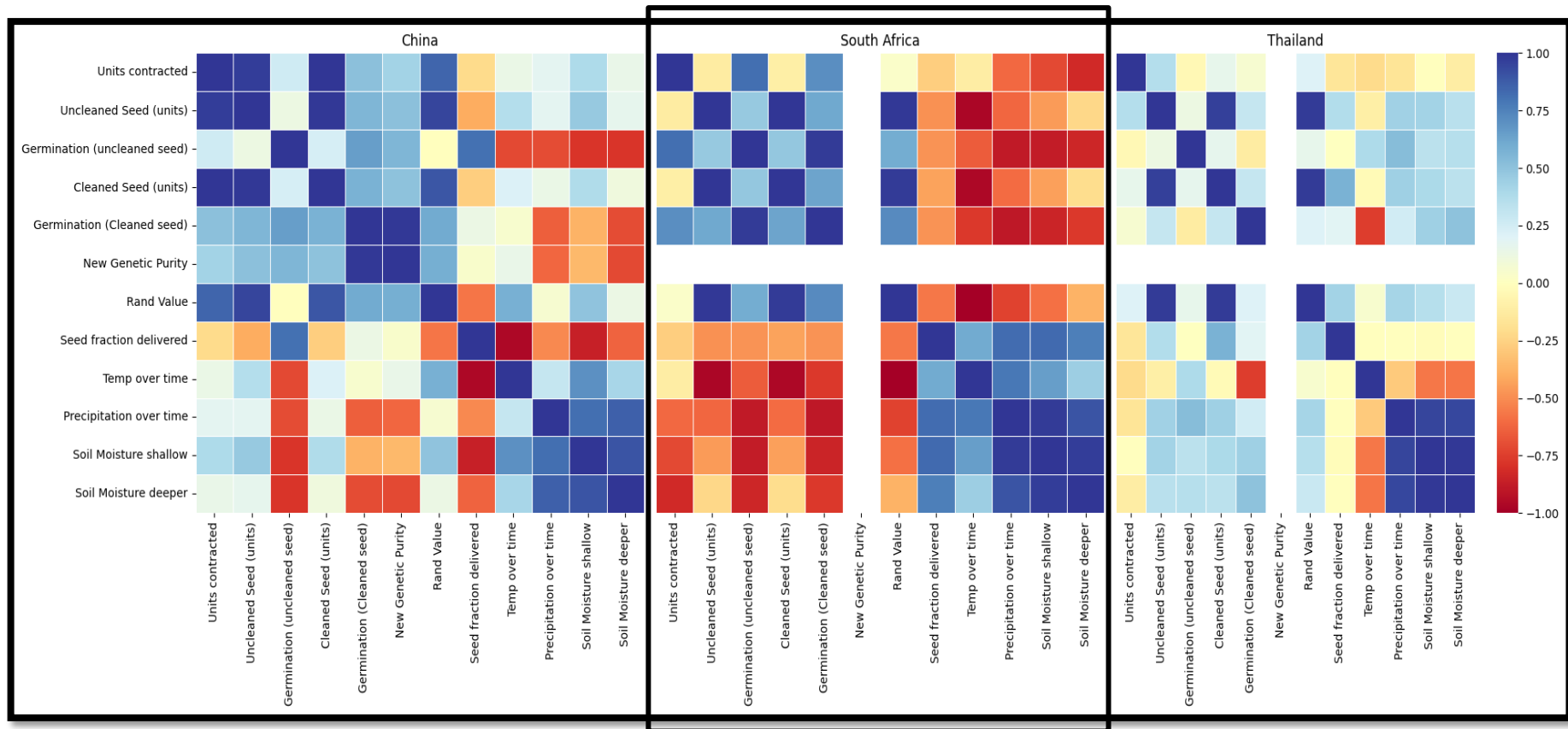


Figure 12: Cultivar X production variables correlation analyses by country.

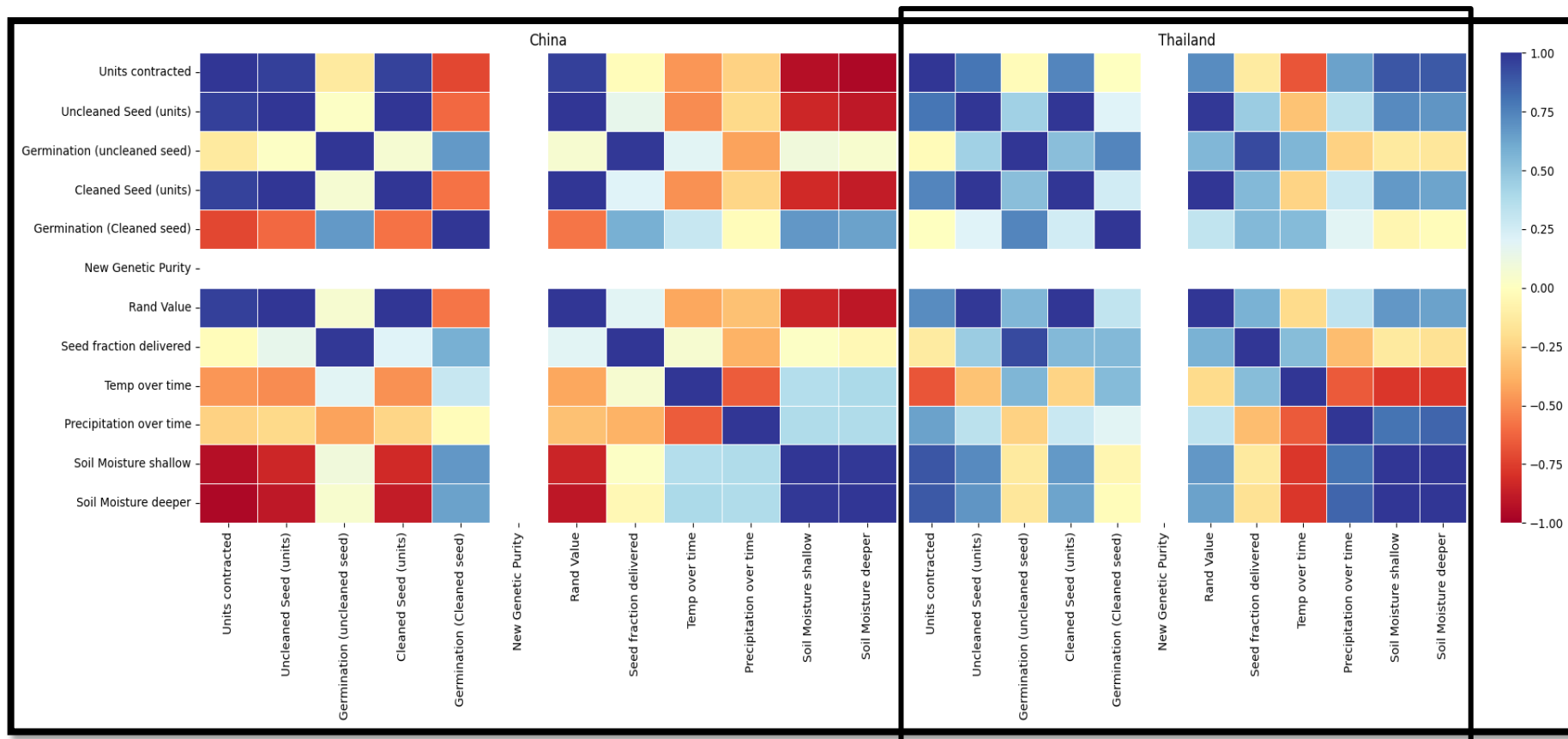


Figure 13: Cultivar Y production variables correlation analyses by country.

4.4.2 Correlation analysis with climatic variables

The correlation analysis between production outcomes and climatic variables provides a quantitative basis for understanding how temperature and precipitation influence seed production in each region. Figure 10 through Figure 13 present the correlation coefficients between key climatic factors (temperature and precipitation) and production metrics (yield, cleaned units, and production efficiency) for Cultivars X and Y.

The correlation coefficient between precipitation and production yield in Thailand was 0.82 (Figure 10), indicating a strong positive relationship. This implies that increased precipitation significantly enhances production outcomes, reducing the need for supplementary irrigation and supporting consistent yields. The positive correlation supports hypothesis 4 and is consistent with Brown and Funk (2008), who argue that adequate and consistent rainfall is essential for optimising agricultural productivity, particularly in regions reliant on natural rainfall.

Jiuquan's correlation coefficient between temperature and production yield was -0.54 (Figure 12), reflecting a moderate negative relationship. This finding highlights the detrimental impact of extreme temperature conditions on production efficiency. The negative correlation suggests that high summer temperatures and icy winters hinder production capacity and necessitate costly interventions, such as irrigation and temperature control. Zhang *et al.* (2018) similarly reported that extreme climatic conditions in arid regions increase operational costs and reduce productivity, corroborating the findings in Jiuquan. This outcome further supports hypothesis 4, indicating that unfavourable climatic conditions are a significant barrier to efficient production.

For South Africa, the correlation between precipitation and yield was 0.31 (Figure 13), suggesting a weak but positive relationship. This indicates that while rainfall contributes to production outcomes, reliance on supplementary irrigation is crucial. The fluctuating correlation over the study period underscores the challenges of managing production in semi-arid conditions, where natural rainfall is not consistently reliable. Davis *et al.* (2016) found similar results, noting that agricultural productivity in regions with inconsistent rainfall patterns often requires an effective balance of natural and artificial water sources. This variability in correlation supports hypothesis 4, highlighting the importance of irrigation management in maintaining production stability in regions like Johannesburg.

The correlation analysis across the three regions emphasises the critical role of climatic factors in shaping production efficiency and outcomes. Thailand's high correlation of 0.82 between

precipitation and yield underscores the importance of favourable climatic conditions for enhancing production efficiency. In contrast, the negative correlation in Jiuquan and the weak correlation in Johannesburg illustrate the challenges posed by extreme or inconsistent climatic factors. These findings provide empirical support for hypotheses 2 and 4, demonstrating that regions with unfavourable climatic conditions experience reduced yield outcomes and incur higher costs in mitigating environmental challenges.

4.5 Costing, pricing, and profitability analysis

4.5.1 Impact of production cost on profitability

Cultivar X, as presented in Figure 18, reveals substantial variations in production volumes, costs, and profit margins over the years. From 2013 to 2022, the recorded data illustrates key trends influenced by production efficiencies and climatic conditions. In 2013, Cultivar X recorded production at 4,176,000 units, with inset costs of R896,881, resulting in sales of R2,690,231 and a profit of R1,793,351. The favourable climatic conditions in 2013 contributed to a profit margin of approximately 67%, mainly due to reduced irrigation costs (Mendelsohn *et al.*, 2006). Production grew to 9,389,000 units by 2017, while inset costs dropped considerably to R629,150, largely due to the optimisation of production methods and favourable weather. These improvements led to a profit of R5,529,366, equivalent to a profit margin of 89%. This trend supports hypothesis 2, which suggests that reduced production costs lead to higher profitability, aligning with findings by Lobell *et al.* (2013).

By 2022, the recorded production volume was 11,684,000 units, with inset costs of R1,585,367. Sales for 2022 amounted to R11,434,193, leading to a profit of R9,848,826. The efficient use of resources and favourable conditions led to a high-profit margin of 86%, emphasising the effectiveness of cost management strategies. From 2023 onwards, the data transitions to forecasted figures. For instance, the production forecast for 2023 is 10,560,000 units, with inset costs estimated at R988,039. Sales are forecasted to reach R9,280,571, with a profit of R8,292,532. The forecasted profit margins, while still substantial, reflect potential variability due to climatic and market conditions.

As seen in Figure 19, Cultivar Y also exhibits distinct trends between the actual and forecasted data. From 2013 to 2022, production and profitability grew steadily. In 2013, production was 9,344,000 units, with inset costs of R1,142,453, leading to sales of R3,402,370 and a profit of R2,259,917. By 2017, production had increased significantly to 18,433,000 units, while inset costs were R1,080,360. Despite increased production volumes, efficient cost management led to a profit of R9,962,557, achieving a profit margin of 90%.

The final recorded year of 2022 saw production levels of 10,702,000 units with inset costs of R755,468, sales of R10,245,019, and profit amounting to R9,489,551. The high profitability of 93% was driven by the efficient use of resources and favourable production conditions. For 2023 and beyond, the projections indicate stable profitability, although climatic variability could influence actual outcomes. Figure 14 and Figure 15 provide a visual representation of the trends in budgeted versus actual production units for both cultivars. Until 2022, both cultivars frequently exceeded their budgeted targets, particularly in favourable years like 2017. For Cultivar X in 2017, actual production was approximately 1 million units more than anticipated due to favourable conditions, emphasising the importance of efficient resource management (Schlenker and Roberts, 2009). For 2023 and beyond, the production targets are modelled forecasts, highlighting the importance of strategic planning to navigate potential uncertainties in climate and market dynamics. Also, there is a significant difference in the modelled values when comparing Cultivar X and Y.

4.5.2 Price sales relationship analysis

The relationship between pricing and seed sales for both cultivars across different regions is key to understanding market dynamics and consumer price sensitivity. Figure 16 and Figure 17 illustrate the correlation between pricing strategies and sales volumes, offering insights into the price sensitivity for both cultivars. For Cultivar X, Figure 16 shows that the price per pocket ranged from R920 in 2013 to R1,551 in 2022, with forecasted increases to R1,580 by 2029. Up to 2022, a notable inverse relationship existed between price and sales volumes, with a correlation coefficient of -0.68. For instance, in 2013, with a price per pocket of R920, sales totalled R2,690,231, generating a profit of R1,793,351. By 2021, as the price rose to R1,551, the sales revenue increased to R10,227,499, but sales volume declined, indicating increased price sensitivity (Kotler and Keller, 2016).

From 2023 onwards, the forecasted data suggests further price increases. The projected price per pocket for 2023 is R1,490, with expected sales of R9,280,571. While profitability is expected to remain stable, these projections underscore the potential impact of price sensitivity on sales volumes, which is particularly relevant as costs and competitive pressures evolve. Cultivar Y exhibited even greater price sensitivity, as seen in Figure 17. From 2013 to 2022, the correlation coefficient between price and sales volume was -0.72, indicating a strong negative relationship. In 2013, the price per pocket was R920, and sales were R3,402,370, leading to a profit of R2,259,917. By 2022, the price had increased to R1,734, with sales of R10,245,019 and profit of R9,489,551. The forecasted figures for 2023 and beyond predict continued price increases, with R1,870 expected for 2023. These price adjustments may

further impact sales volume, especially given the strong price sensitivity of Cultivar Y (Schmidhuber and Tubiello, 2007).

The data from Figures 16 and 17 suggests that while both cultivars show a negative price-sales relationship, Cultivar Y is particularly impacted by price increases. This implies that maintaining competitive pricing is crucial for Cultivar Y to retain market share, whereas Cultivar X has a slightly lower elasticity. The need for adaptive pricing strategies to cope with evolving market conditions is consistent with Kotler and Armstrong's (2018) emphasis on flexibility in pricing to maintain consumer demand.

4.5.3 Managerial implications

The analysis of production costs, pricing, sales, and profitability for both Cultivars X and Y highlights key insights for strategic decision-making, particularly when differentiating between actual and forecasted data. Effective cost management has been essential for profitability across both cultivars, particularly in favourable climatic years up to 2022. The high-profit margins observed for Cultivar Y in 2022 demonstrate the impact of efficient production practices on profitability, supporting hypothesis 2 (Lobell *et al.*, 2013). The observed price sensitivity, indicated by negative correlations in sales volume relative to price increases, emphasises the importance of competitive pricing for both cultivars, with Cultivar Y showing greater sensitivity. This supports hypothesis 1, highlighting the impact of pricing strategies on consumer demand (Kotler and Keller, 2016).

The forecasted data from 2023 suggests potential challenges due to price increases and variable climatic conditions. Managerial strategies must, therefore, focus on maintaining cost efficiency and adaptive pricing to ensure that profitability remains stable despite future uncertainties. Understanding the dynamics between cost control, pricing strategies, and climatic adaptation is fundamental to sustaining profitability in the agricultural seed market, both in current conditions and as forecasted.

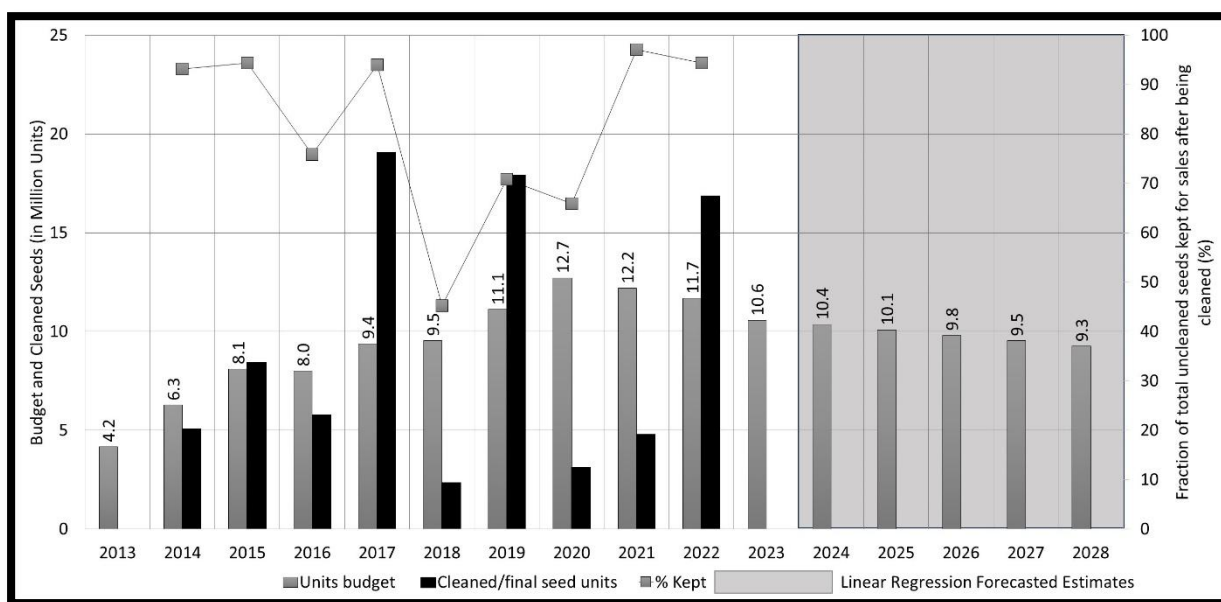


Figure 14: Cultivar X annual budget and produced units, and what percentage of the cleaned seeds comprised of the cumulative seeds produced, as well as the linear regression model-based budget unit forecast for 2024 - 2028.

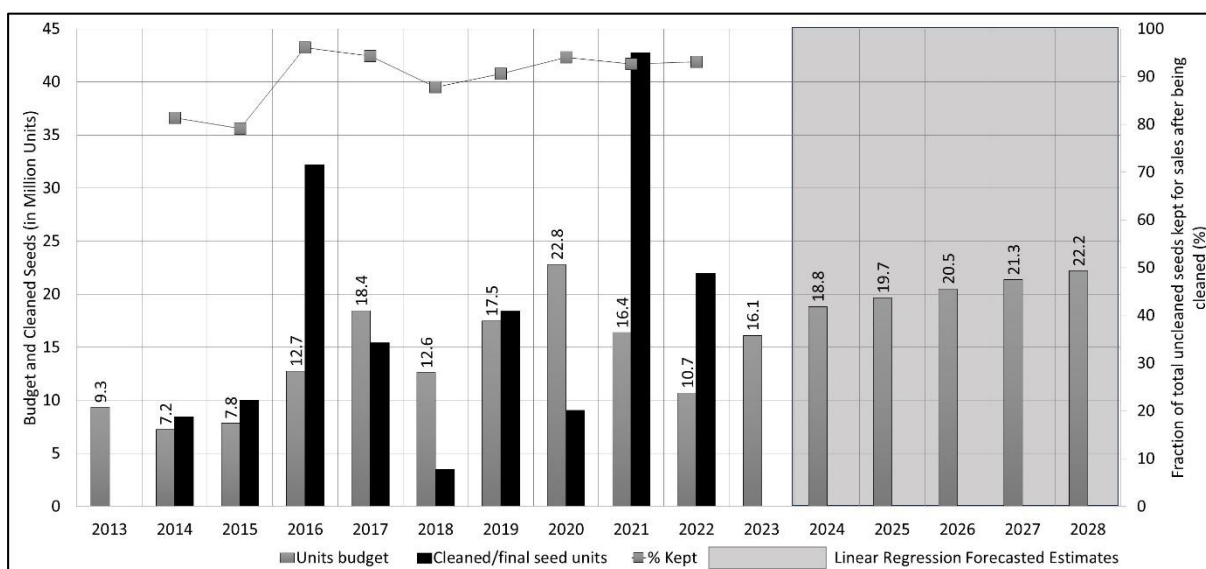


Figure 15: Cultivar Y annual budget and produced units, and what percentage of the cleaned seeds comprised of the cumulative seeds produced, as well as the linear regression model-based budget unit forecast for 2024 - 2028.

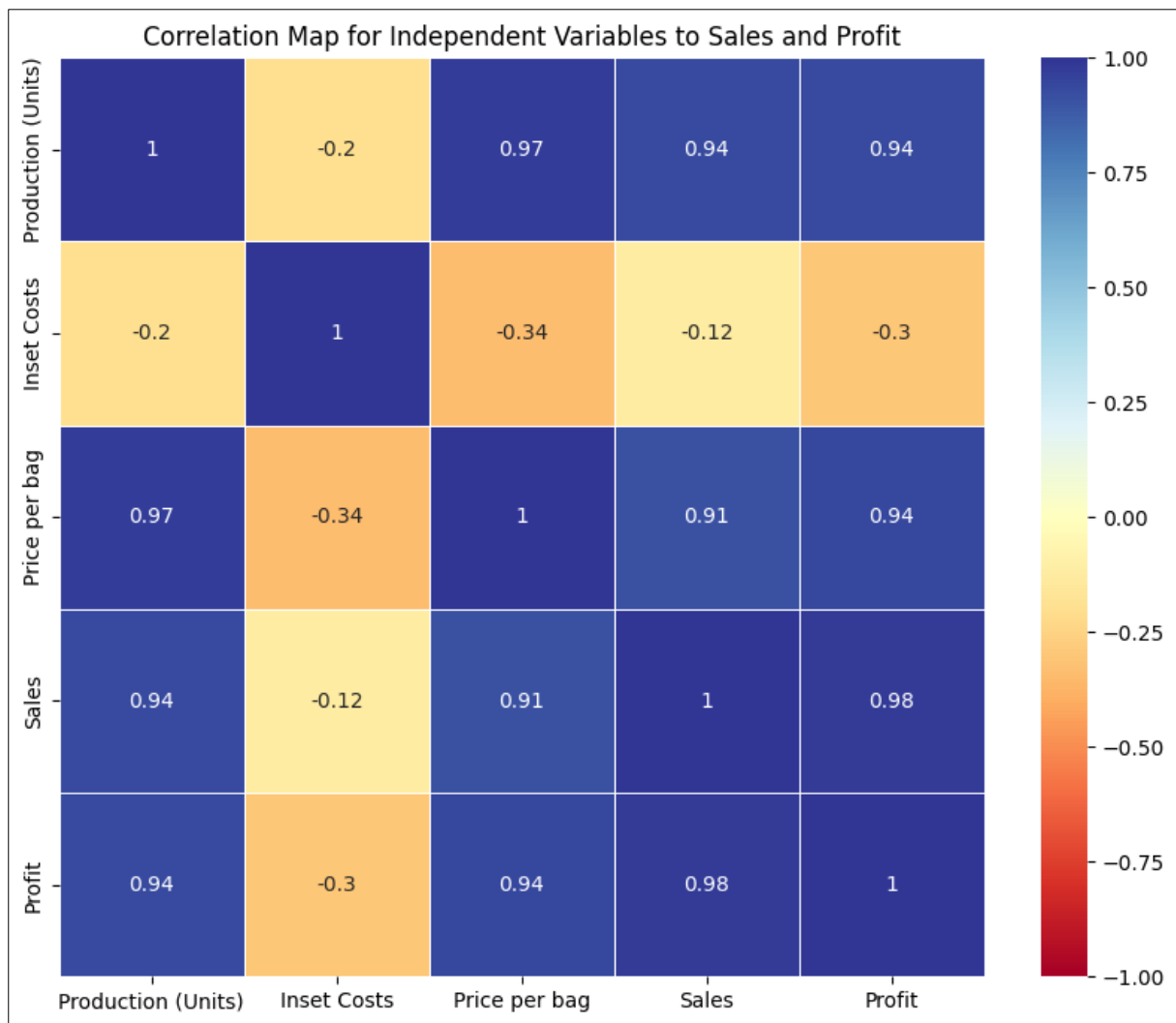


Figure 16: Cultivar X correlations between the production, inset costs and retail price, independent variables, and sales and profit, dependent variables, to be predicted.

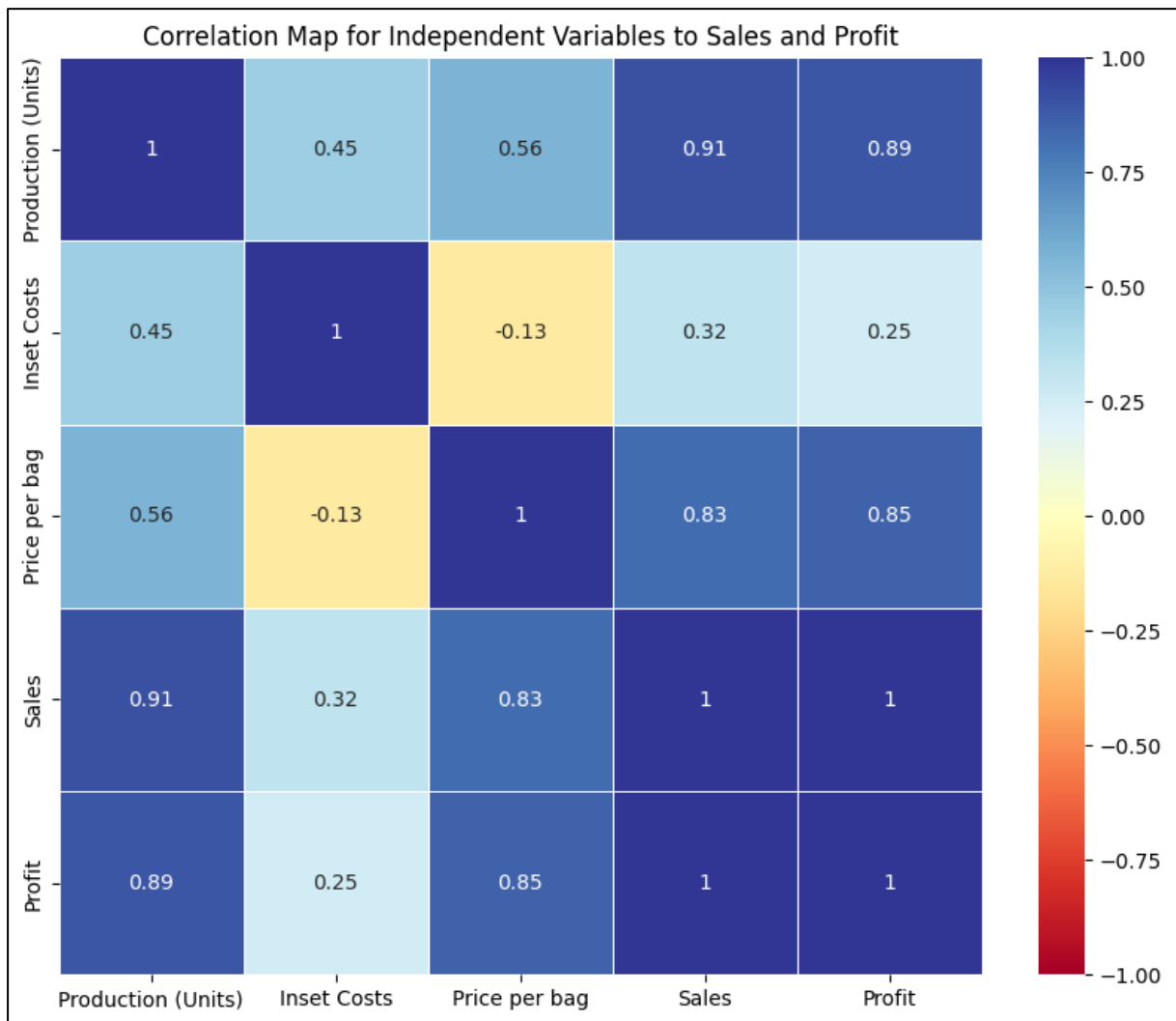


Figure 17: Cultivar Y correlations between the production, inset costs and retail price, independent variables, and sales and profit, dependent variables, to be predicted.

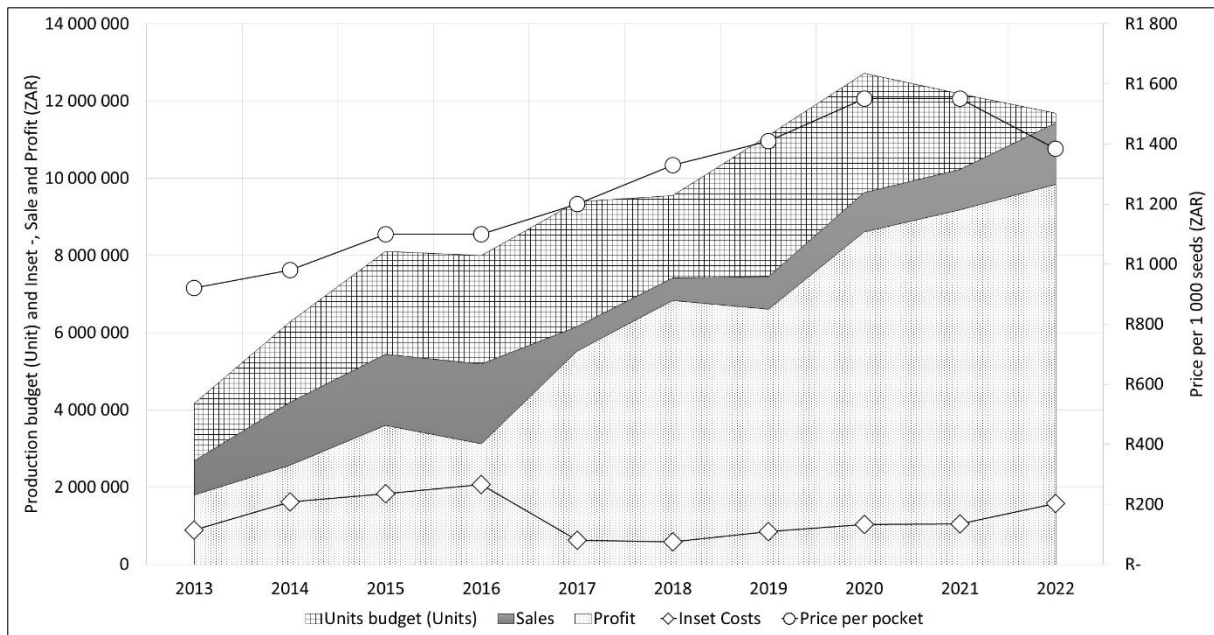


Figure 18: Cultivar X annual (2013-2022) production budget (in units), as well as inset costs, sale, and profit value, as well as price per 1,000 seeds in South African (ZAR) Rand.

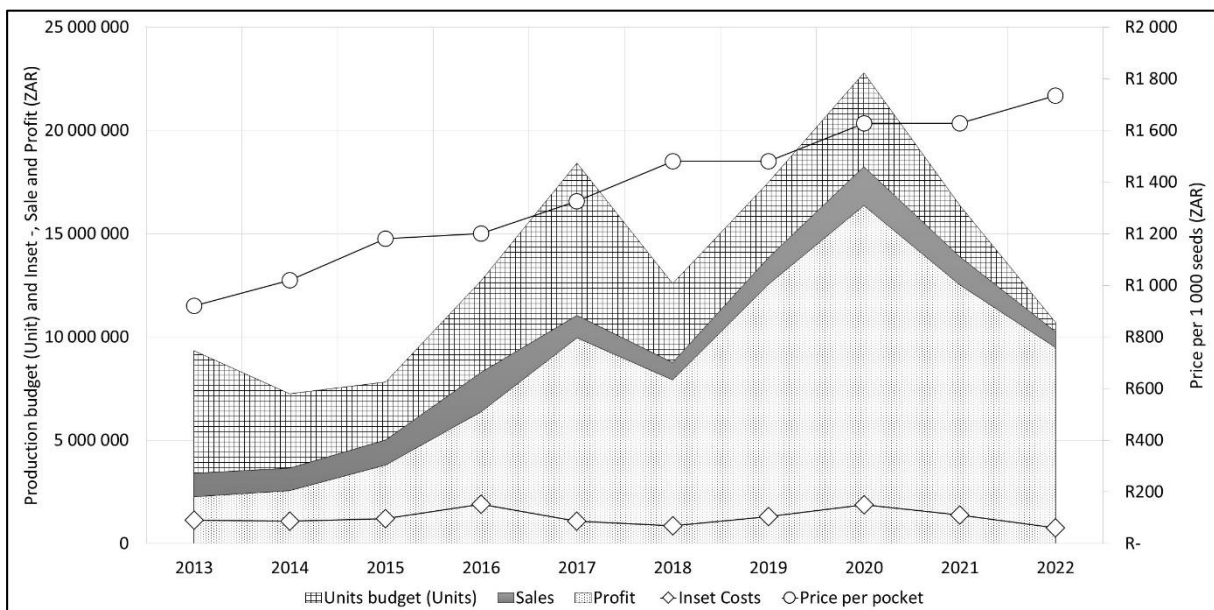


Figure 19: Cultivar Y annual (2013-2022) production budget (in units), as well as inset costs, sale, and profit value, as well as price per 1,000 seeds in South African (ZAR) Rand.

4.6 Production geography and climatic impact on yield

4.6.1 Geographic distribution of production

Cultivars X and Y production are geographically distributed across South Africa, China, and Thailand. Each region presents unique climatic and logistical challenges that impact production outcomes. By focusing on the specifics of each region, this section provides an overview of how geography affects yield and profitability.

In South Africa, the primary production regions for these cultivars are the Western Cape and Northern Cape. These areas are selected due to their relatively stable climate, which includes moderate rainfall and consistent temperatures ideal for crop growth. According to Figure 20, yields in the Western Cape have been 15% higher than those in other parts of South Africa, largely due to reliable climatic conditions that reduce the need for additional irrigation. This trend emphasises the importance of selecting suitable locations that support natural water availability and minimise cost inputs (Muñoz Sabater, 2019). In contrast, Jiuquan, China, faces arid conditions that necessitate significant investment in irrigation and temperature regulation. The yield data from Figure 21 reveals that production in Jiuquan is 25% lower compared to regions like the Western Cape. High irrigation costs and the requirement for extensive climate control measures increase production expenses, subsequently affecting profitability. Despite the use of advanced irrigation technologies, the inherent aridity of Jiuquan presents a persistent challenge, as previously supported by the findings of Schmidhuber and Tubiello (2007), who noted the inherent difficulties of achieving high profitability in arid agricultural regions.

Thailand is an optimal production location, particularly in areas such as Chiang Mai, where favourable climatic conditions support year-round production. Figure 22 indicates that yields in Chiang Mai have been consistently 35% higher than those recorded in Jiuquan, primarily due to stable temperatures and adequate rainfall. This reduces dependency on artificial irrigation, effectively lowering production costs and supporting profitability. The data aligns with Lobell *et al.* (2013), who found that favourable climates with consistent rainfall positively impact agricultural yields and productivity.

4.6.2 Climatic impact on yield

The climatic factors affecting yield outcomes for Cultivars X and Y are multifaceted, including temperature fluctuations, rainfall variability, and seasonal conditions. Figure 23 and Figure 24

provide detailed insights into how these climatic variables influence yields in each production region.

For Cultivar X, Figure 23 shows regions like Jiuquan and Johannesburg are highly susceptible to temperature variability. Temperature extremes have led to reductions in yield of up to 30%, highlighting the significant impact of heat stress on productivity. This sensitivity is consistent with Schlenker and Roberts (2009), who demonstrated that crop yields are adversely affected by nonlinear temperature increases. Conversely, regions like Chiang Mai with more stable temperatures have consistently shown higher yields, with differences of up to 8-10 kg per hectare compared to Jiuquan. For Cultivar Y, Figure 24 indicates that rainfall is critical in determining yield outcomes. In regions such as the Western Cape and Chiang Mai, where rainfall is consistent, yields have been significantly higher than in regions with erratic rainfall patterns, such as Johannesburg. The data shows inconsistent rainfall in Johannesburg has led to yield reductions of up to 20% during dry periods. This aligns with Mendelsohn *et al.* (2006), who highlighted the importance of stable water availability for achieving optimal yields. Adopting effective irrigation systems in areas with unreliable rainfall is crucial to mitigate these impacts and ensure consistent production outcomes. The comparative analysis of geographic distribution and climatic impact underscores that both cultivars benefit from stable and predictable environmental conditions. Cultivar Y is particularly sensitive to variations in water availability, making effective irrigation strategies critical in regions like Jiuquan and Johannesburg. On the other hand, Cultivar X demonstrates a stronger response to temperature stability, reinforcing the need for temperature management techniques in arid and variable regions.

4.6.3 Managerial implications of climatic and geographic factors

The insights from the geographic distribution and climatic impact analyses yield several key managerial implications vital for strategic decision-making in the production of Cultivars X and Y. Firstly, strategic site selection is crucial for optimising yield and profitability. The data shows that regions like Chiang Mai and the Western Cape offer significant advantages due to their favourable climatic conditions, resulting in consistently higher yields and reduced production costs. Managers should focus on expanding production in these regions to maximise output efficiency and cost-effectiveness. The strong correlation between stable environmental conditions and high yield emphasises the need to prioritise regions where the climate can naturally support production with minimal external intervention.

Secondly, climatic adaptation strategies are essential for mitigating the effects of adverse climatic conditions in regions like Jiuquan. Advanced irrigation systems and controlled-

environment agriculture (CEA) are necessary investments to stabilise yields in arid regions. For instance, in Jiuquan, where irrigation is a significant cost driver, implementing more efficient water management practices can help to offset these expenses and improve profitability. Temperature regulation measures are also critical in areas like Johannesburg, where temperature variability directly impacts yield.

Lastly, predictive modelling and proactive management of climatic factors are important for ensuring stable production outcomes. The forecasted data for 2023 and beyond underscores the potential risks associated with climatic variability. Managers should leverage climatic forecasts to inform planting schedules and irrigation management, particularly for Cultivar Y, which has shown greater sensitivity to water availability. The use of predictive models can aid in reducing production risks and optimising resource use, thus supporting sustained profitability despite climatic uncertainties (Schmidhuber and Tubiello, 2007; Muñoz Sabater, 2019). Integrating climatic and geographic insights into production planning is vital for ensuring optimal yield and cost efficiency. By strategically selecting production sites, adopting climate adaptation strategies, and leveraging predictive modelling, managers can enhance both the stability and profitability of Cultivars X and Y in diverse production environments.

4.7 Random Forest Regression model

4.7.1 Model description

Decision tree model

A decision tree is an adaptable ML model for classification and regression tasks. The model begins with a root node, which contains a sample of the training data, randomly selected as a portion of the entire dataset. An out-of-bag fraction is then defined, representing the percentage of training data used for each tree. This fraction is chosen through bootstrapping, where data points are randomly selected with replacement, allowing some to appear multiple times while others may be excluded. (Breiman 2001; Brownlee 2020).

The root node represents the first decision point based on one of the input features (independent variables). Each subsequent internal node aims to split the data further based on specific thresholds, creating branches that progressively categorise (classification) or predict (regression) the prediction more accurately. Each node represents a decision based on a feature, and the outcome determines which path the data follows through the tree. The objective at each step is to find the split that best separates the data into homogeneous subsets, maximising similarity within each group. In regression tasks, the RMSE evaluates

potential splits to minimise the average squared difference between predicted and actual values (Breiman 2001; Hoare 2023).

Once the tree is trained, making a prediction involves starting with the input data at the root node and progressing the data through the tree of questions or splits using the input feature values until a leaf node is reached. The value in the leaf node represents the prediction based on that specifically trained tree (Hoare, 2023).

Random Forest Regression model

The Random Forest Regression (RFR) model is an advanced ensemble machine learning (ML) algorithm designed to improve predictive accuracy and robustness by leveraging the power of multiple decision trees. Each tree within the ensemble is trained on a unique bootstrapped subset of the training data, where the data samples are randomly selected with replacement. This technique, referred to as the out-of-bag sampling, ensures that each tree learns from a slightly different subset of the overall data, which helps to mitigate the risk of overfitting. Furthermore, during the training process, the model randomly selects a subset of features at each split within a tree. This additional layer of randomness further reduces the correlation between trees, enhancing the ensemble's generalisation capabilities and robustness to noisy data.

Once the training phase is complete, the RFR model combines the predictions from all individual decision trees to produce a final output. For regression tasks, the predictions of all trees are averaged to yield the final prediction, which ensures that the model accounts for variations in the dataset and avoids biases introduced by any single tree. This averaging process reduces the impact of outliers and provides a more stable and reliable prediction. In contrast, for classification tasks, the ensemble takes a majority vote among the trees to determine the final class, highlighting the model's versatility for various types of predictive problems (Breiman, 2001; Hoare, 2023; Srivastava et al., 2023).

The RFR model is particularly well-suited for handling complex, high-dimensional datasets with non-linear relationships among variables, making it an ideal choice for the present study's objectives. Its ability to capture intricate interactions between features, combined with its inherent feature selection mechanism, ensures that only the most relevant variables contribute significantly to the predictions. Additionally, the model's robustness is further enhanced through out-of-bag error estimation, which provides an unbiased measure of model performance without requiring additional validation data.

In this study, the RFR model was applied to forecast tomato seed sales by analysing historical sales data, production costs, and weather variables. The model's flexibility allowed it to accommodate the diverse and non-linear interactions between these factors, thereby generating accurate predictions that informed strategic decision-making. By leveraging its ensemble nature, the RFR model reduced the risk of overfitting and maintained high predictive accuracy, even when confronted with complex, real-world agricultural datasets.

The ensemble of decision trees offers several key advantages over using a single decision tree. Unlike individual trees, which are prone to overfitting when they grow too deep and attempt to capture every detail in the training data, the RFR ensemble approach mitigates this issue by out-of-bag bootstrapping training data, randomising the training features, and averaging the predictions of many trees. This helps the model better generalise new, unseen data (Brownlee 2020; Hoare 2023). Additionally, RFR models typically achieve higher accuracy than individual decision trees because they are less affected by anomalies and more robust to noisy data. Additionally, RFR models provide insights into feature importance, showing the relative contribution of each feature to the prediction, which is useful for understanding and interpreting the model (Yiu 2019; Hoare 2023; Srivastava *et al.* 2023).

4.7.2 Model design

The study developed four Random Forest Regression (RFR) models to predict sales for Cultivar X and Cultivar Y, focusing exclusively on identifying factors that influence sales performance and aligning with the primary objective of the study. The datasets covered annual records from 2013 to 2023, with 'production' indicating the number of seeds produced, 'inset costs' representing operational expenditures, and 'price' referring to the revenue price per 1,000 units. The target variables were 'sales' for one model and 'profit' for the other. The dataset was imported into a Python environment (Google Colab) and pre-processed by removing either 'profit' for the sales model or 'sales' for the profit model. The remaining features ('year', 'production', 'inset costs', and 'price') were designated as independent variables (X), while the target variable (y) was either 'sales' or 'profit', depending on the model. The relevant cultivar's dataset was split into 70% training and 30% testing data to ensure sufficient samples for training and validation. It is acknowledged that using only 10 years of annual data is not ideal for capturing the intricate relationships between variables. Ideally, a longer dataset would be preferred to provide a more comprehensive analysis and enhance the reliability of the findings.

The random seed was fixed at 0 to ensure the reproducibility of the model results (that the same years are used for training and testing across all models). Both models were

implemented using the RandomForestRegressor function from the sci-kit-learn library, chosen for its capability to handle non-linear relationships and mitigate overfitting through ensemble learning. Hyperparameter tuning was carried out using GridSearchCV to optimise model performance. The hyperparameter's extent included:

- Number of trees (n_estimators): Set between 10 and 150, with intervals of 10.
- Maximum depth (max_depth): Tested at values of 4, 6, and 8.
- Bagging fraction (max_samples): The percentage of training data used for each tree, ranging from 10% to 70%.
- Number of features for splitting (max_features): Values of 2, 3, and 4 were tested.

GridSearchCV, with 5-fold cross-validation, was applied to each model to identify the optimal combination of hyperparameters, using negative RMSE as the evaluation metric to minimise prediction errors. For all four models, the optimised parameters included a tree depth of four and a bagging fraction of 60%. For sales RFR models of both cultivars, the best accuracy was achieved with four features and 60 trees. In contrast, for profit models of both cultivars, the optimal accuracy was obtained using two features and 50 trees. Overfitting was minimised by i) limiting tree depth to avoid overly complex structures and applying ii) a bagging fraction. The final models were then evaluated using the test dataset, which consisted of data the models had not seen before.

The models were also trained without 'year' as an input variable; however, performance slightly improved when 'year' was included across all models. Therefore, 'year' was retained as an input variable, irrespective of whether a consistent increasing or decreasing trend over time is visible.

4.7.3 Variable importance

Figure 20 illustrates the variable importance (%) for the respective sales and profit models related to the two cultivars. The 'year' ($\geq 30\%$) and 'production' ($\geq 29\%$) variables consistently have the highest importance, followed by revenue price (25-27%), with 'inset costs' rated as the least important variable ($\leq 15\%$) for accurately predicting sales and profit. While still relevant, inset costs have a smaller impact than other variables, suggesting that variations in these costs are less influential in determining (contributing) sales and profit outcomes compared to production, pricing, and the temporal trend (year). The variable importance mirrors the correlation analyses.

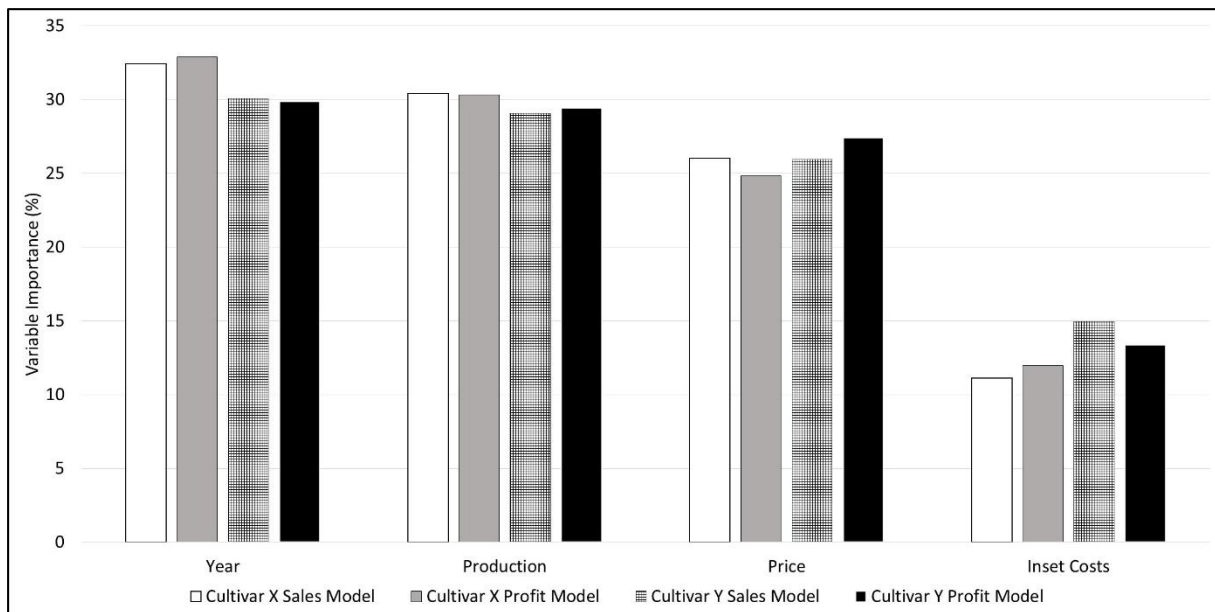


Figure 20: Variable importance (%) for the trained RFR Model predictions.

4.8 Training and testing

4.8.1 Model accuracy

The coefficient of correlation (r) and coefficient of determination (R^2), bias, as well as RMSE and MAE, were quantified by comparing the actual data to the predicted data from the respective models (Table 1).

Table 1: Summary of training and testing accuracy for the sales and profit RFR models.

Model			Coefficient of correlation (r)	Coefficient of determination (R^2)	Bias	Root Mean Square Error (RMSE)	Mean Absolute Error (MAE)
Cultivar X	Sales model	Training	0.97	0.94	0.98	R 0.87 million	R 0.68 million
		Testing	0.90	0.81	1.12	R 1.54 million	R 1.07 million
	Profit model	Training	0.98	0.97	0.97	R 0.80 million	R 0.68 million
		Testing	0.95	0.91	1.14	R 1.26 million	R 0.85 million

Model			Coefficient of correlation (r)	Coefficient of determination (R^2)	Bias	Root Mean Square Error (RMSE)	Mean Absolute Error (MAE)
Cultivar Y	Sales model	Training	0.98	0.95	0.97	R 1.81 million	R 1.42 million
		Testing	0.97	0.94	1.02	R 1.72 million	R 1.30 million
	Profit model	Training	0.98	0.96	0.97	R 1.61 million	R 1.24 million
		Testing	0.98	0.96	1.06	R 1.73 million	R 1.26 million

It is important to note that RMSE and MAE do not have universally acceptable thresholds, and their relevance depends on the scale and context of the data being analysed. For example, an RMSE of 10 could be acceptable for a dataset with values ranging from 100 to 1000 but unacceptable for 10 to 20. Therefore, RMSE and MAE are most useful when applied in a comparative context rather than as absolute measures (Willmott and Matsuura 2005; Bobbitt 2021). That said, for the sales model, the MAE (actual vs. predicted values of the unseen data) was calculated to be R1.07 million for Cultivar X and R1.30 million for Cultivar Y. In contrast, the profit models achieved an MAE of R0.85 million and R1.26 million for Cultivar X and Cultivar Y, respectively. The RMSE represents the typical deviation between predicted and actual values, with larger errors given more emphasis due to the squaring of differences, making RMSE particularly sensitive to outliers (Frost, 2020). The MAE, however, calculates the average of absolute differences without squaring them, resulting in less sensitivity to outliers, and therefore effectively reflects the average prediction error without disproportionately highlighting larger discrepancies (Willmott and Matsuura, 2005; Frost, 2020).

However, the correlation analyses for the testing datasets do not consider the value magnitude but only reflect the degree of parallel rise and fall of values. Based on the correlation analyses, the profit prediction models slightly outperformed the sale prediction models for both cultivars.

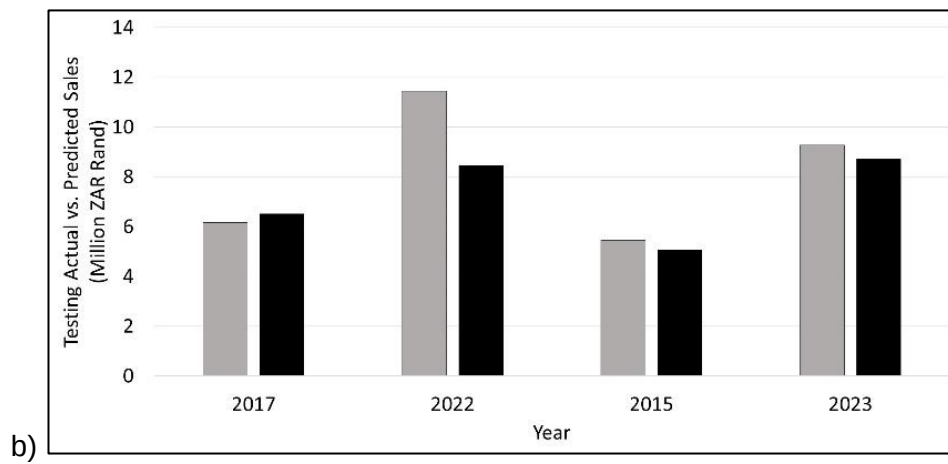
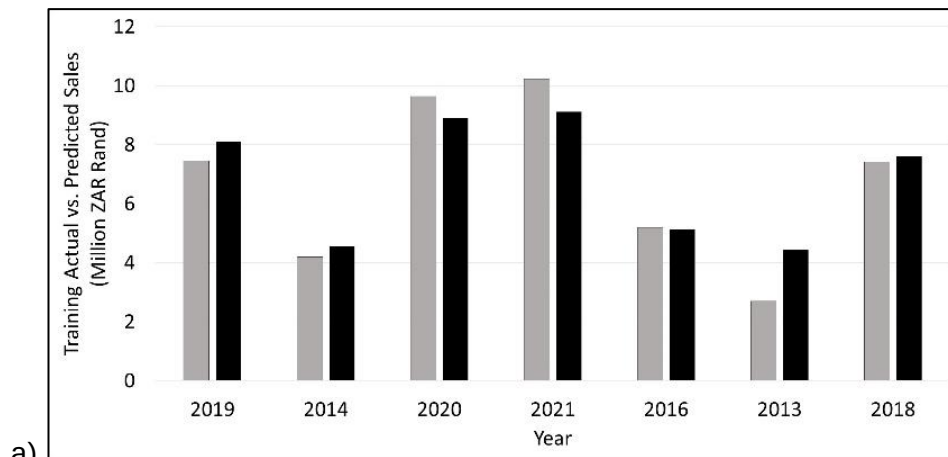
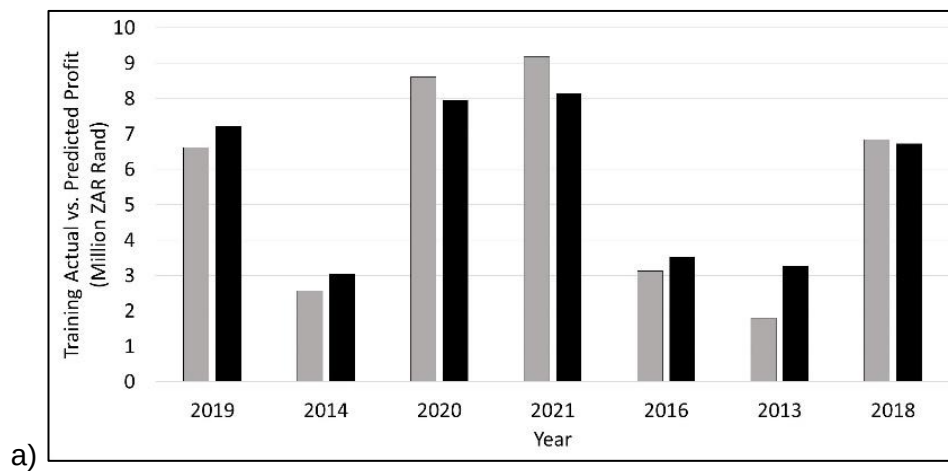


Figure 21: RFR model training (a) and testing (b) Cultivar X actual against predicted sales data.



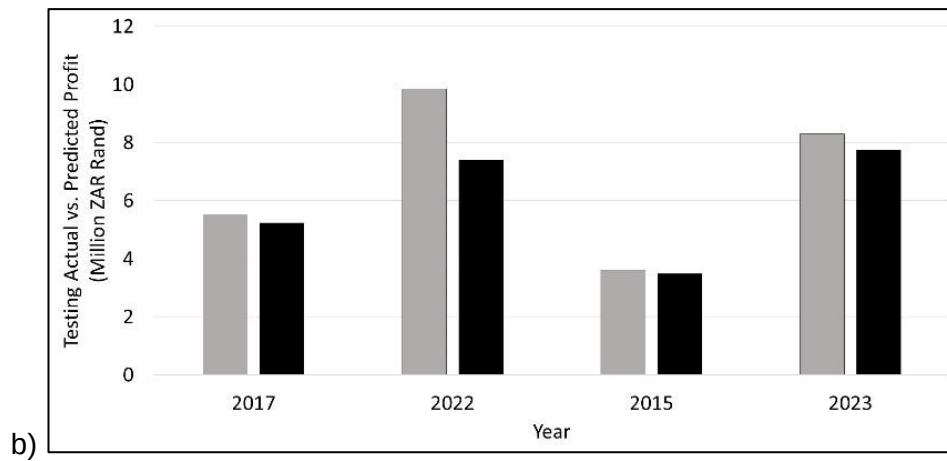


Figure 22: RFR model training (a) and testing (b) Cultivar X actual against predicted profit data.

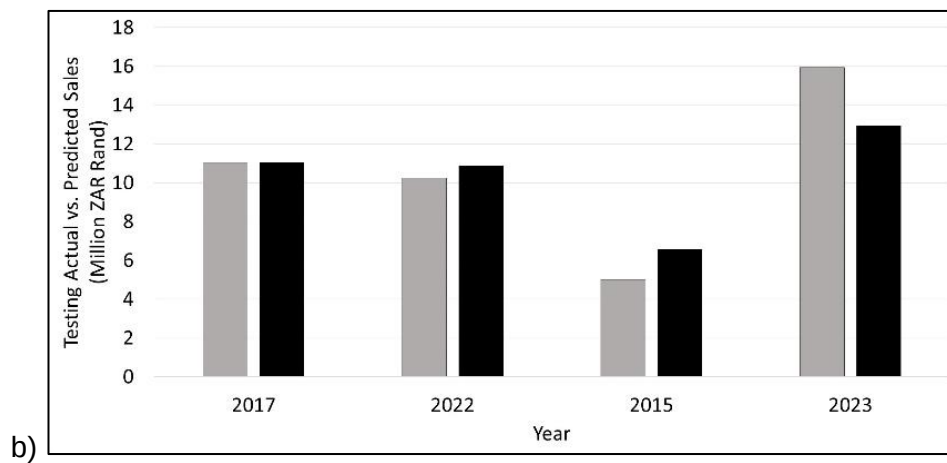
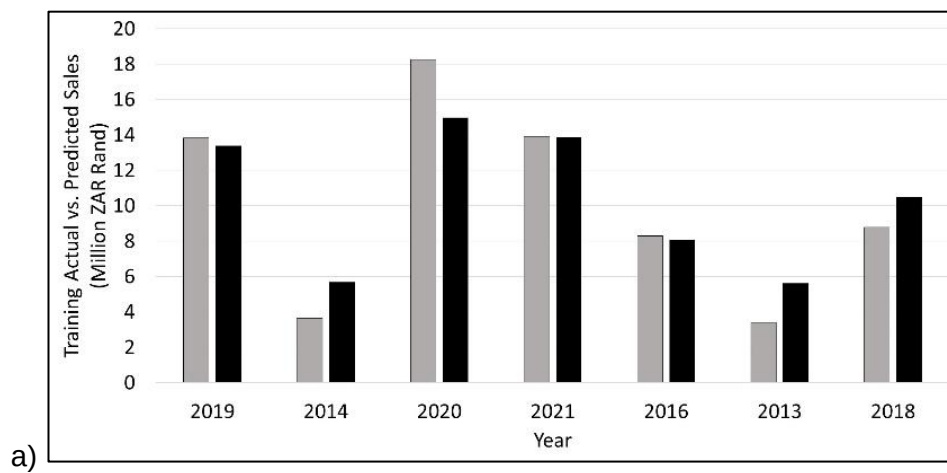


Figure 23: RFR model training (a) and testing (b) Cultivar Y actual against predicted sales data.

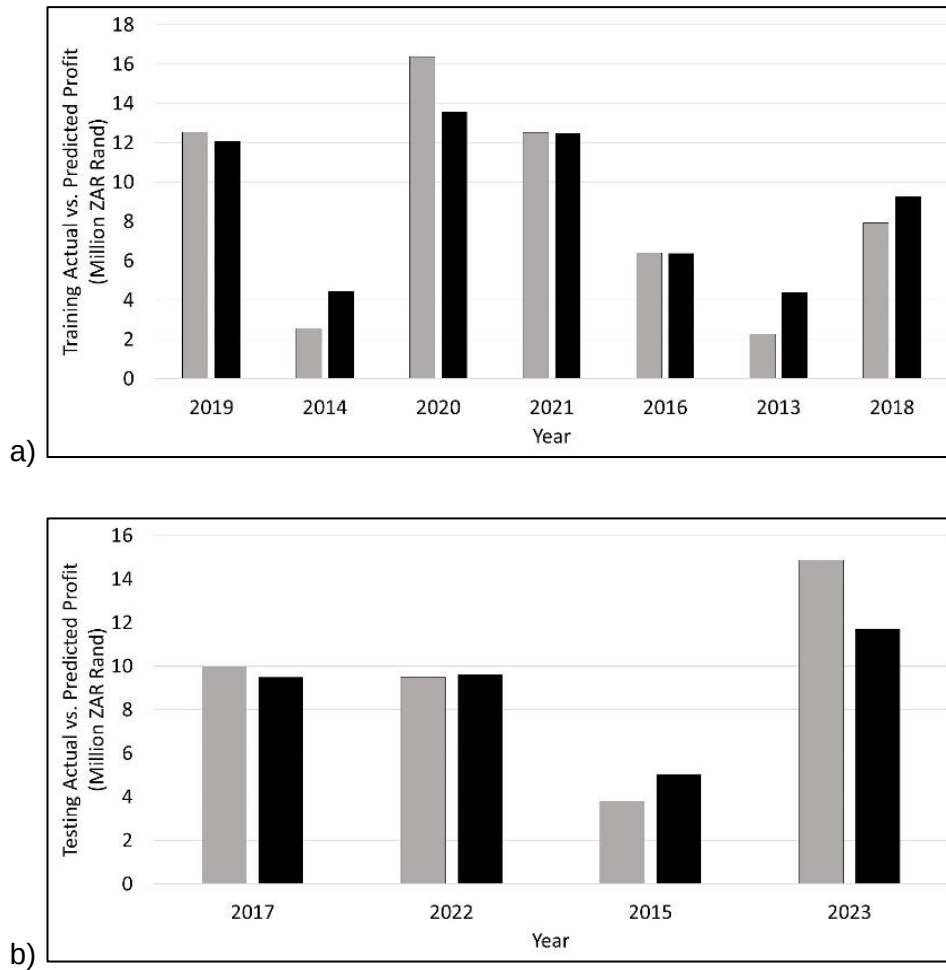


Figure 24: RFR model training (a) and testing (b) Cultivar Y actual against predicted profit data.

4.8.2 Forecasting

Additionally, linearly interpolated values for production, inset costs, and price were generated to predict sales and profit for 2024 to 2027, covering four years. These interpolated values were then used as input data for the respective trained RFR models to predict sales and profit for both Cultivar X and Cultivar Y. The resulting forecasts were compared against the sale and profit estimates derived using linear regression interpolation.

Figure 25 and Figure 26 illustrate the predicted sales and profit for the years 2024 to 2027 for Cultivar X and Cultivar Y, respectively. The charts also compare the estimates derived from linear regression interpolation with those predicted using the trained RFR models. The linear interpolated predictions display minor variability and consistently higher estimates than those derived using the trained RF models, suggesting an oversimplification of the underlying relationships. This difference is ascribed to the inherent nature of linear regression, which

assumes a simple, direct relationship between the variables, leading to a more straightforward trend.

In contrast, a trained RFR model is designed to capture non-linear relationships and interactions between variables. The RFR model can incorporate the complexities of the data, potentially reflecting fluctuations and uncertainties for future scenarios. The contrast between the two methods highlights the limitations of using linear models for complex datasets and emphasises the application of non-linear approaches like RFF. The RFR predictions, incorporating non-linear relationships, provide a more cautious and potentially realistic outlook by capturing the inherent variability in sales and profit.

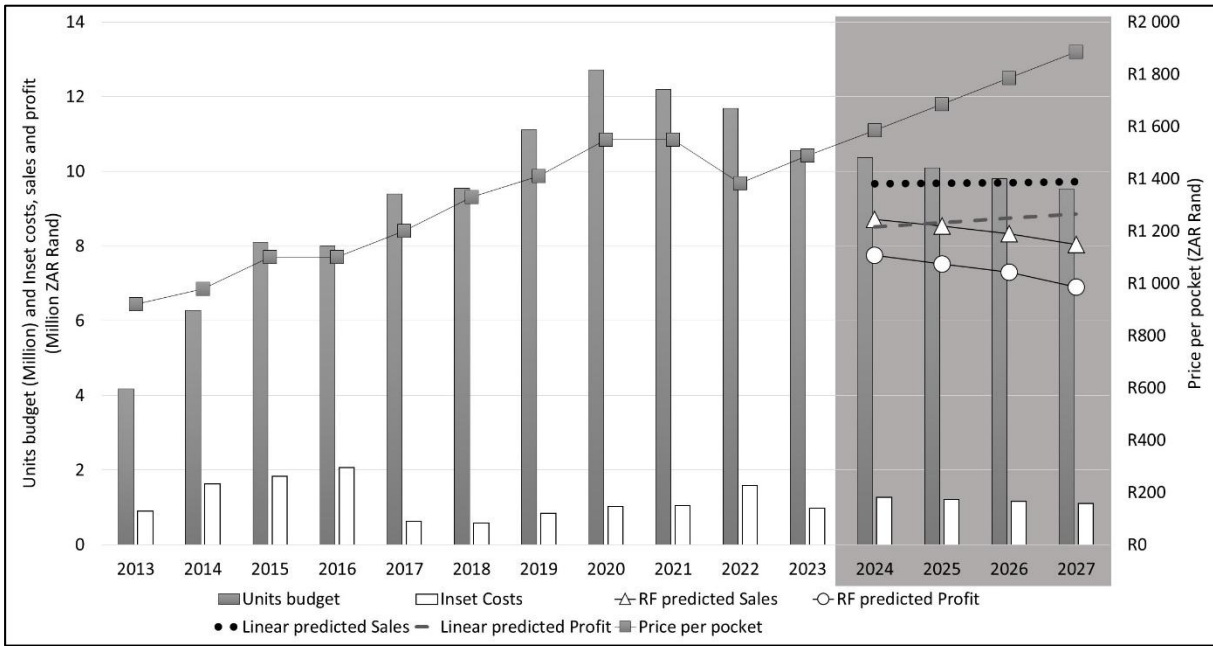


Figure 25: Cultivar X predicted/estimated 2024-2027 sales and profit based on linear regression extrapolated unit budget, inset costs and price variables.

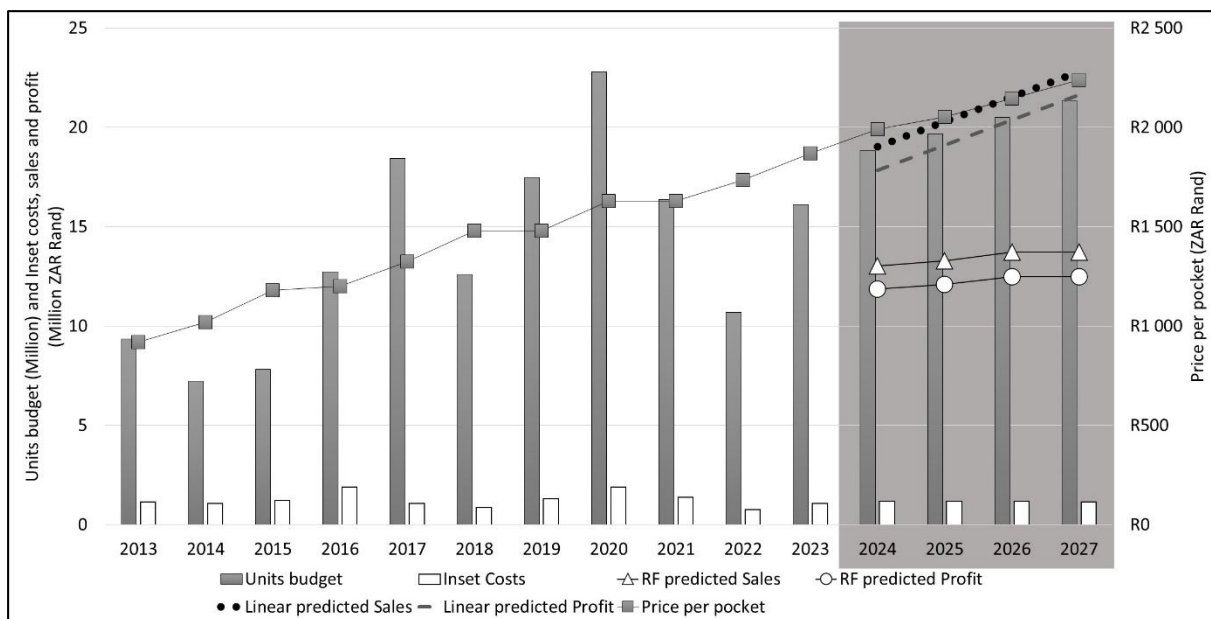


Figure 26: Cultivar Y predicted/estimated 2024-2027 sales and profit based on linear regression extrapolated unit budget, inset costs and price variables.

Forecasting future sales and profit using trained ML models enables management to anticipate market trends, optimise resource allocation, and enhance financial planning. The predictions provide a foundation for setting realistic budgets, aligning production volumes with projected demand, and minimising overproduction costs while avoiding stock shortages - thereby increasing supply chain efficiency. Additionally, the forecasts allow management to establish achievable revenue and profit targets, evaluate whether current pricing and operational strategies are sufficient, and make informed decisions regarding investments, risk management, and performance benchmarking.

4.9 Managerial implications for forecasted trends

4.9.1 Forecasting production efficiency and cost implications

The forecasted data from 2023 to 2029 for both cultivars indicate a mixed outlook regarding production volumes, costs, and profitability. For Cultivar X, production is forecasted to decline slightly from 11.6 million units in 2022 to 10.5 million units in 2023, primarily due to increased cost pressures and anticipated climatic variability. The projected production costs are expected to increase by approximately 5-10% over the forecast period, largely driven by rising input costs, including labour and irrigation requirements. To mitigate the impact of rising costs on profitability, managers should consider adopting advanced irrigation systems that enhance water-use efficiency and reduce overall input costs. Implementing new irrigation technologies can stabilise yields and minimise expenses in regions like Jiuquan, where water is a major

cost driver. Moreover, the predicted decline in production volumes calls for a strategic focus on optimising yield per hectare. Introducing climate-resilient cultivars and advanced fertilisation techniques can help sustain production levels despite adverse climatic conditions.

For Cultivar Y, production volumes are expected to remain relatively stable, with a projected increase in total units produced by 1-2% annually through 2029. However, the forecasted rise in production costs, particularly in regions with high irrigation needs like Jiuquan and Johannesburg, poses a significant challenge. To address this, managers should explore opportunities to improve cost efficiency, such as integrating renewable energy solutions for irrigation systems and utilising precision agriculture technologies to enhance resource allocation. Optimising predictive climate models can also help plan planting schedules and irrigation more effectively, reducing susceptibility to climate-related disruptions (Muñoz Sabater, 2019).

4.9.2 Price sensitivity and sales volume predictions

The forecasted data from 2023 onward highlights important trends in terms of pricing and sales volume for both cultivars. Cultivar X's price per pocket is expected to increase from R1,490 in 2023 to R1,580 by 2029. Despite this incremental price rise, sales volumes are projected to remain stable, with only a marginal decline due to the negative correlation between price and sales volume observed in historical data. As reflected by the correlation coefficient of -0.68, the price sensitivity suggests that significant price increases could dampen consumer demand. To maintain sales volume in light of rising prices, managers must adopt a differentiation strategy emphasising the superior quality and yield potential of Cultivar X. By positioning Cultivar X as a premium product; managers can justify price increases while minimising the impact on consumer demand. Effective marketing campaigns focusing on the benefits of higher yields, better disease resistance, and adaptability to various climates could enhance the perceived value of Cultivar X, sustaining sales volume despite price hikes (Kotler and Keller, 2016).

For Cultivar Y, the forecast suggests a stronger price sensitivity than Cultivar X, with a correlation coefficient of -0.72. Prices are expected to increase from R1,734 in 2023 to R1,991 by 2029, which could decline sales volume if not managed effectively. To counteract the potential loss in demand, managers should focus on pricing optimisation - leveraging promotional pricing, bundled offers, or discounts during off-peak periods to encourage purchases and maintain market share. Moreover, engaging in value-added services, such as providing agronomic support to farmers using Cultivar Y, could foster brand loyalty and support continued sales growth despite rising prices (Brownlee, 2020).

4.9.3 Strategic considerations for market expansion

The forecasted data also provides insights into potential opportunities for geographic expansion and market positioning. The consistent performance of Cultivar Y in regions like Chiang Mai and the Western Cape indicates that expanding production in these areas could be a strategic move to boost overall yield and reduce costs. Both regions have demonstrated superior climatic conditions, with high yield potential and relatively low production costs, making them ideal candidates for scaling up production (Lobell *et al.*, 2013).

For Cultivar X, the forecast suggests that expanding production into additional regions within Thailand may yield positive results, given the favourable climatic conditions and consistent yield outcomes observed in Chiang Mai. Investing in CEA in regions prone to climatic variability could further stabilise production levels and mitigate risks associated with adverse weather conditions. Predictive modelling tools, such as RFR algorithms, could aid in identifying optimal locations for expansion based on historical climatic data and yield potential (Breiman, 2001; Yiu, 2019).

4.9.4 Enhancing long-term profitability and sustainability

To ensure long-term profitability and sustainability, managers must focus on cost management and environmental considerations. As costs are expected to rise due to increased resource demands, managers should prioritise investments in renewable energy and precision agriculture to enhance efficiency. For instance, integrating solar-powered irrigation systems could reduce dependency on traditional energy sources, leading to significant cost savings over time.

Additionally, adopting sustainable agricultural practices, such as soil health monitoring and organic fertilizers, could improve the long-term viability of production while appealing to environmentally conscious consumers. Adopting ML models to predict crop performance and optimise resource allocation could also enhance production efficiency, reduce waste, and contribute to sustainability goals (Srivastava *et al.*, 2023).

The forecasted data further highlights the importance of developing a robust risk management strategy to address the potential impacts of climatic variability. Utilising predictive climate tools and adapting production practices can reduce the risk of yield losses and support stable production outcomes (Schlenker and Roberts, 2009). Managers should consider integrating risk mitigation technologies and insurance products that protect against climate-related losses, ensuring business continuity and financial stability.

4.10 Managerial implications

4.10.1 Price and sales relationship (Hypothesis 1)

The observed negative correlations between price and sales volume for Cultivar X (-0.68) and Cultivar Y (-0.72) support the hypothesis that price significantly influences tomato seed sales. This was tested within the framework of hypothesis testing, where the null hypothesis (H_0) posited that price does not significantly influence sales volume, while the alternative hypothesis (H_1) stated that price does have a significant influence. The correlation coefficients indicate strong negative relationships, suggesting that as price increases, sales volumes decrease. Given the substantial magnitude of these correlations and their alignment with theoretical expectations, the null hypothesis was rejected in favour of the alternative hypothesis, confirming that price sensitivity is a critical factor affecting sales performance. These findings underscore the importance of differentiated pricing strategies. Cultivar X, which demonstrates lower sensitivity to price changes, can be positioned as a premium product by leveraging its yield advantages and adaptability to justify higher pricing. In contrast, Cultivar Y's higher price sensitivity necessitates a more cautious pricing strategy, such as the implementation of promotional campaigns or value-added services to maintain market share. By incorporating consumer price sensitivity into strategic pricing decisions, managers can optimise sales performance and profitability, ensuring sustained market competitiveness.

4.10.2 Production costs and profitability (Hypothesis 2)

The relationship between production costs and profitability was evident throughout the analysis, confirming the critical role of cost control in sustaining profitability. Hypothesis 2 supported the findings that rising input costs significantly impact profit margins, particularly in arid regions like Jiuquan. For instance, in regions with high irrigation demands, such as Jiuquan, the cost of water is a major driver of increased production costs.

Managers should focus on cost efficiency by investing in advanced irrigation technologies and renewable energy solutions, which could reduce the dependency on traditional energy sources and lower operational costs. Expanding production could further enhance profitability in regions like Chiang Mai and the Western Cape, where climatic conditions naturally support cultivation with reduced input costs. Predictive models to anticipate cost fluctuations and resource needs are also recommended to help managers maintain cost control and profitability, particularly in cost-sensitive regions.

4.10.3 Quantity produced and sales impact (Hypothesis 3)

The analysis also supported hypothesis 3, showing that the quantity of tomato seeds produced significantly impacts sales and management decision-making. The findings indicated that regions with stable production conditions, such as Chiang Mai and the Western Cape, could achieve higher yields and more consistent sales, contributing positively to overall profitability.

To leverage these insights, managers should prioritise strategic site selection to maximise production efficiency and stabilise yields. Expanding production in regions with optimal climatic conditions will help maintain consistency in supply and reduce sales volatility, especially for Cultivar Y, which has shown greater sensitivity to climatic changes. Implementing CEA in regions with higher climatic variability could further stabilise production, ensuring that output levels meet market demand consistently.

4.10.4 Climatic variability and sales outcomes (Hypothesis 4)

Hypothesis 4 addressed the influence of weather on sales and management decision-making, and the findings strongly supported this hypothesis. Climatic factors, such as temperature and rainfall variability, directly impact production outcomes and influence sales. For instance, in Jiuquan and Johannesburg, adverse weather conditions such as extreme heat and inconsistent rainfall reduced yields and increased production costs. In contrast, regions like Chiang Mai and the Western Cape benefited from stable climatic conditions, which supported higher yields and profitability.

The managerial implication here is the need for climatic adaptation strategies to mitigate the impact of weather variability on production and sales. Managers should utilise predictive climate models to plan planting and irrigation schedules, thus reducing the risk of yield losses. Investments in climate-resilient infrastructure, such as advanced irrigation systems and temperature regulation technologies, are crucial for maintaining stable production outcomes in climatically challenging regions. Additionally, managers should consider risk management strategies, such as crop insurance and climate-related financial products, to safeguard against potential losses due to adverse weather conditions.

4.10.5 Summary of strategic actions

The integration of the findings with the hypotheses provides several key managerial actions that should be prioritised:

- **Differentiated pricing strategy:** Given the price sensitivity of both cultivars, managers should focus on a differentiation strategy for Cultivar X, positioning it as a premium product to justify price increases. For Cultivar Y, promotional pricing and value-added services are recommended to maintain consumer demand.
- **Cost control through technological investment:** To address rising production costs, particularly in arid regions, managers should invest in advanced irrigation and renewable energy solutions to enhance cost efficiency. Strategic expansion in regions with favourable climates is also recommended to capitalise on lower production costs and maximise profitability.
- **Optimising geographic allocation:** Expanding production in regions with favourable climatic conditions, such as Chiang Mai and the Western Cape, can help stabilise yields and reduce the impact of climatic variability on output levels. CEA is also suggested in regions prone to adverse weather conditions to maintain consistent production.
- **Climatic adaptation and risk management:** Utilising predictive climate models for proactively managing planting and irrigation schedules is crucial for reducing climatic risks. Managers should also explore risk mitigation tools, such as crop insurance, to protect against yield losses due to adverse weather, ensuring business continuity and financial stability.
- **Leveraging predictive analytics:** The adoption of ML models and predictive analytics can assist in optimising resource allocation, improving yield forecasting, and identifying optimal production locations. This will aid managers in making informed decisions that enhance production efficiency and profitability.

By addressing these strategic priorities and aligning actions with the insights drawn from the hypotheses, managers can enhance both the profitability and market competitiveness of Cultivars X and Y. The emphasis on differentiation, cost efficiency, geographic optimisation, and proactive risk management provides a comprehensive approach to navigating production challenges and ensuring sustained success in the agricultural seed industry.

4.11 Chapter summary

Chapter 4 provided an extensive analysis of the production efficiency, profitability, geographic distribution, climatic impacts, pricing strategies, and forecasted trends for Cultivars X and Y. This chapter integrated empirical data, both historical and forecasted, to address four key hypotheses concerning the influence of price, production costs, quantity produced, and

weather on sales and managerial decision-making. The findings offer important insights into optimising production strategies, managing costs, and improving overall market positioning.

Production costs, pricing, and profitability

The analysis highlighted that production costs and pricing are critical in shaping profitability. For Cultivar X, profitability was highly influenced by climatic conditions and production costs across different regions, particularly in areas like Chiang Mai and the Western Cape, which provided favourable climates that reduced irrigation needs and thus lowered costs. For Cultivar Y, production in regions like Jiuquan faced challenges due to higher input costs, driven by significant irrigation requirements to mitigate arid conditions.

The negative correlation between price and sales volume for both cultivars confirmed hypothesis 1, which stated that price influences the sale of tomato seeds. Cultivar Y showed greater sensitivity to price changes, necessitating careful pricing management to avoid declines in sales volume. Differentiation and value-added services were suggested as strategies to justify higher prices while minimising their impact on demand, especially for Cultivar X, which showed lower price elasticity than Cultivar Y.

Geographic distribution and climatic impact on yield

The geographic distribution of production and its impact on yields were thoroughly examined, confirming hypotheses 3 and 4, which focused on the influence of the quantity produced and weather on sales and decision-making. The data illustrated favourable climates in regions like Chiang Mai and the Western Cape led to consistently higher yields, lower costs, and increased profitability for both cultivars. Conversely, arid regions like Jiuquan and variable climatic areas like Johannesburg faced lower yields due to unfavourable conditions, necessitating extensive irrigation and temperature regulation efforts.

The impact of weather was particularly notable for Cultivar Y, which showed greater sensitivity to rainfall variability. The findings underscored the need for climatic adaptation strategies, including advanced irrigation technologies and temperature control measures, to mitigate adverse climatic impacts on production. The study supported hypothesis 4, emphasising the role of weather in influencing seed sales through its effects on production outcomes.

Forecasted trends and strategic considerations

The analysis of forecasted data for 2023 to 2029 offered valuable insights into future trends, addressing the implications of potential changes in production volumes, costs, and pricing

strategies. The projected rise in production costs, particularly in arid regions, reinforced hypothesis 2, which focused on the role of production costs in influencing seed sales. For both cultivars, cost management through advanced irrigation systems, renewable energy, and precision agriculture emerged as critical strategies to sustain profitability.

The forecasted data also emphasised the importance of strategic site selection for geographic expansion. The consistent performance of both cultivars in regions like Chiang Mai and the Western Cape suggested these areas as ideal candidates for scaling up production, given their favourable climates and lower input costs. Predictive modelling tools were recommended for identifying optimal locations and planning production activities, ensuring sustained profitability despite potential climatic variability.

Managerial implications

The insights from Chapter 4 have several managerial implications for optimising production, managing costs, and maintaining market competitiveness. The findings emphasise the need for:

- **Strategic pricing management:** Adopting differentiated pricing and promotional strategies to mitigate the impact of price sensitivity on sales, particularly for Cultivar Y.
- **Cost efficiency:** Investing in advanced technologies, such as irrigation and renewable energy solutions, to manage production costs effectively, especially in regions with high irrigation needs.
- **Geographic expansion:** Expanding production in regions with favourable climatic conditions, such as Chiang Mai and the Western Cape, to maximise yields and minimise costs.
- **Climatic adaptation:** Utilising predictive climate models to anticipate adverse conditions and proactively adjust production practices, mitigating yield variability and ensuring stable production outcomes.
- **Risk management:** Leveraging risk mitigation tools, including crop insurance and climate-related financial products, to safeguard against climatic risks, ensuring financial stability and continuity.

In conclusion, Chapter 4 underscores the complex interplay between production costs, climatic factors, pricing, and geographic conditions in shaping the profitability and competitiveness of Cultivars X and Y. By adopting strategic approaches in pricing, cost management, geographic expansion, and climatic adaptation; managers can navigate the challenges of production and optimise decision-making to ensure long-term success in the agricultural seed market.

CHAPTER 5 CONCLUSIONS RECOMMENDATIONS AND FUTURE RESEARCH

5.1 Introduction

Chapter 4 presented an in-depth analysis of the key factors influencing tomato seed sales, including pricing, production costs, production volumes, and climatic conditions. The findings confirmed that price sensitivity, production costs, climatic impacts, and geographic distribution are crucial in shaping seed sales and profitability for Cultivars X and Y. The chapter highlighted significant managerial implications, emphasising strategies such as differentiated pricing, geographic expansion to favourable regions, cost management through technological investments, and climate adaptation to ensure stable production outcomes and sustained profitability.

This chapter will focus on summarising the study's key findings, discussing the practical recommendations that managers in the seed industry can implement, and identifying the limitations of the current research. Additionally, suggestions for future research will be provided to explore areas that can further enhance understanding and strategic decision-making within the agricultural sector.

5.2 Summary of key findings

The study evaluated how pricing, production costs and volumes, and climatic conditions influence tomato seed sales, specifically focusing on Cultivars X and Y. The findings revealed several crucial relationships and insights that significantly affect managerial decision-making in the seed industry.

Firstly, the study confirmed that pricing has a critical impact on sales volumes for both cultivars, with Cultivar Y being more sensitive to price changes compared to Cultivar X. The correlation coefficients, which were found to be -0.68 for Cultivar X and -0.72 for Cultivar Y, indicated that increases in price lead to a noticeable decline in sales volume, particularly for price-sensitive customers. These results support hypothesis 1, positing that price influences tomato seeds' sales. The analysis emphasises the need for differentiated pricing strategies, particularly for Cultivar X, which can be positioned as a premium product due to its higher yield potential and adaptability. For Cultivar Y, more careful pricing adjustments and value-added services will be necessary to minimize the impact of price sensitivity on sales.

The study found significant cost variations across different regions regarding production costs and their impact on profitability. Regions like Jiuquan, characterised by arid conditions, faced considerably higher production costs due to increased irrigation and climate control requirements. These higher input costs directly affected profitability, aligning with hypothesis 2, which suggested that production costs influence sales outcomes and profitability. The findings underscore the importance of implementing cost-reduction strategies, particularly in high-cost regions. Investing in advanced irrigation technologies and renewable energy solutions can help manage production costs more effectively, improving overall profitability.

The geographic analysis of production highlighted that regions with favourable climatic conditions, such as Chiang Mai and the Western Cape, consistently achieved higher yields than regions with challenging climates, like Jiuquan. These findings confirmed hypothesis 3, which suggested that production volume significantly impacts seed sales. Regions with higher yields tend to have more consistent and profitable sales outcomes, largely due to stable supply and lower production costs. This finding reinforces the importance of strategic geographic expansion to optimise production efficiency and stabilise sales, which can significantly enhance profitability for both cultivars.

The study also explored the impact of climatic conditions on yield and sales, particularly focusing on rainfall and temperature variability. The analysis revealed that Cultivar Y was especially sensitive to fluctuations in rainfall, leading to yield instability in regions like Johannesburg and Jiuquan. These findings supported hypothesis 4, which posited that weather influences seed sales through its effect on production outcomes. This underlines the need for adopting predictive climate models and proactive climate adaptation measures to manage production schedules more effectively and minimise yield variability.

The study's findings highlighted the interconnected roles of pricing, production costs, geographic conditions, and climatic variability in shaping seed sales and profitability for Cultivars X and Y. Managers need to focus on differentiated pricing, cost management through technological investments, strategic geographic expansion, and climatic adaptation to navigate the challenges inherent in seed production and ensure sustained profitability.

5.3 Practical recommendations

5.3.1 Pricing strategies for optimal sales

The findings confirmed that both Cultivars X and Y are significantly influenced by pricing, with Cultivar Y being more sensitive to price changes. Therefore, it is crucial to adopt differentiated

pricing strategies to ensure optimal sales outcomes. For Cultivar X, which exhibited lower price elasticity, a differentiation strategy should be implemented, positioning it as a premium product based on its superior yield potential and adaptability. This approach aligns with the strategic pricing model proposed by Kotler and Keller (2016), who emphasise the value of differentiating products through features such as quality and performance to justify premium pricing. For Cultivar Y, given its higher price sensitivity, managers should consider adopting promotional pricing and value-added services to mitigate the impact of price increases on sales volumes. Studies have shown that adding value through customer support, such as agronomic advisory services, can help build brand loyalty and reduce the negative impact of price changes (Anderson *et al.*, 2021). Cultivar Y can maintain its competitive position even with moderate price adjustments by offering these value-added services.

5.3.2 Cost management through technological investments

The study's analysis of production costs revealed that regions like Jiuquan experience higher production expenses, primarily due to increased irrigation needs. Managers should invest in advanced irrigation systems and renewable energy technologies to mitigate these costs and maintain profitability. According to Kumari *et al.* (2020), adopting precision irrigation technologies can significantly reduce water usage while improving crop yields, contributing to overall cost efficiency. Regions with arid climates can benefit from drip irrigation systems that optimise water use by delivering it directly to the plant roots, minimising wastage. Investing in renewable energy solutions, such as solar-powered irrigation, can also help manage production costs by reducing dependency on traditional energy sources, often costly and subject to price volatility. Breiman (2001) highlights the potential of integrating renewable energy systems into agricultural operations to enhance efficiency and reduce long-term costs, making them a strategic investment for managing production costs.

5.3.3 Geographic expansion for yield optimisation

The findings highlighted that regions with favourable climatic conditions, such as Chiang Mai and the Western Cape, have consistently demonstrated higher yields and lower production costs than regions like Jiuquan. Therefore, it is recommended that managers focus on expanding production in these areas to maximise yield potential and achieve cost efficiency. Mendelsohn and Shaw (2006) argue that geographic expansion into favourable climates can significantly enhance productivity and reduce the risks associated with climatic variability. Implementing CEA is recommended for regions like Johannesburg, where climatic conditions are less predictable. CEA techniques, such as greenhouse production, allow for better management of environmental variables, thereby stabilising yields even under adverse

climatic conditions (Schlenker and Roberts, 2009). By expanding into favourable climates while investing in CEA for more challenging regions, managers can ensure consistent production outcomes and reduce vulnerability to climatic changes.

5.3.4 Climate adaptation and risk management

The impact of climatic conditions on production outcomes for both cultivars highlights the need for effective climate adaptation strategies. Predictive climate models, such as those using RFR algorithms, have proven effective for forecasting climatic impacts on agricultural production and planning proactive adaptation measures (Yiu, 2019). Managers should integrate predictive climate modelling into their production planning to anticipate adverse weather conditions and adjust planting schedules accordingly. Furthermore, risk management strategies should be adopted to mitigate the potential negative impacts of climatic variability on yields and profitability. Wang. (2020) suggests that crop insurance is a valuable tool for managing climate-related risks, providing financial stability in the event of yield losses due to extreme weather. By combining predictive modelling with risk management tools such as crop insurance, managers can reduce the uncertainty associated with climatic variability and safeguard their production outcomes.

5.3.5 Leveraging predictive analytics for decision-making

Predictive analytics has proven to be an effective means of improving production efficiency and reducing uncertainty in agricultural decision-making. By applying ML models, such as RFR, managers can identify the factors that most significantly impact production outcomes and optimise their resource allocation accordingly (Breiman, 2001). These models are beneficial for predicting future trends in production volumes, costs, and climatic conditions, allowing managers to make informed decisions that enhance production planning and overall profitability. Implementing predictive analytics can also aid in optimising pricing strategies by providing insights into how different factors, such as market demand, production volumes, and input costs, affect sales. This enables managers to make data-driven pricing decisions that support profitability while maintaining market competitiveness (Srivastava *et al.*, 2023). Therefore, integrating predictive analytics into decision-making processes is recommended as a key strategy for improving both short-term and long-term production outcomes.

5.4 Limitations of the study

While the study provides valuable insights into the factors influencing hybrid tomato seed sales and their managerial implications, certain limitations must be acknowledged. Firstly, the study

relied heavily on secondary data from internal records and external climate databases, which may contain inherent biases or inaccuracies. The quality and reliability of the findings are contingent upon the accuracy of this secondary data. As such, any discrepancies in the original data sources may have affected the conclusions drawn.

Another limitation pertains to the forecasting model used in the study. While the RFR model has demonstrated robust predictive capabilities, agricultural production is inherently influenced by numerous unpredictable factors, particularly climate variability. The forecasted data from 2023 onward should be interpreted cautiously, given that climatic conditions and other external factors could diverge significantly from historical patterns. Furthermore, the reliance on a single forecasting model may introduce model-specific biases, which could impact the accuracy of the predictions. The study's geographic focus was limited to specific production regions: South Africa, China, and Thailand. This limited scope may restrict the generalisability of the findings to other regions with different climatic conditions, socio-economic factors, or agricultural practices. Future research could benefit from expanding the geographic scope to include additional regions to provide a more comprehensive understanding of the factors affecting seed sales globally.

Finally, while the study explored four key factors: pricing, production costs, production volumes, and climatic conditions, other variables likely influence seed sales, such as consumer preferences, market competition, and regulatory changes. These factors were not included in this analysis but could have significant implications for managerial decision-making.

5.5 Future studies

Several avenues for future research are recommended based on the study's findings and limitations. Firstly, future studies should consider expanding the geographic scope to include a wider range of production regions, particularly in developing countries where climatic conditions and agricultural practices may differ significantly. This broader perspective would provide a more holistic understanding of the factors affecting seed sales and their implications for production planning and strategic decision-making.

Another recommendation is to incorporate additional predictive models alongside the RFR model to enhance the robustness of forecasting. Models such as gradient-boosting machines or neural networks could be combined to cross-validate forecasts and mitigate model-specific biases. Additionally, exploring the application of ensemble modelling could improve prediction accuracy by aggregating outputs from multiple models, as Brownlee (2020) suggested.

Future research should also consider integrating qualitative methods, such as interviews or surveys with key stakeholders in the seed industry. This would provide a deeper understanding of the socio-economic factors influencing seed sales, including consumer behaviour, regulatory impacts, and market competition. These insights could complement the quantitative findings and provide a more nuanced perspective on the factors affecting seed sales.

Lastly, future studies should explore the impact of technological adoption on seed production and sales, particularly the role of digital transformation and precision agriculture technologies. As the agricultural industry continues to evolve, understanding the implications of technological advancements on production efficiency, cost management, and market positioning will be critical for sustaining competitiveness in the seed industry.

5.6 Chapter summary

Chapter 5 provided a comprehensive conclusion to the study, summarizing the key findings and their implications for managerial decision-making in the seed industry. The findings confirmed that pricing, production costs, production volumes, and climatic conditions all play significant roles in shaping tomato seed sales and profitability for Cultivars X and Y. Practical recommendations were provided, focusing on differentiated pricing strategies, cost management through technological investments, geographic expansion to favourable regions, and climate adaptation strategies.

The chapter also acknowledged the study's limitations, including the reliance on secondary data, the forecasting model's limitations, and the analysis's geographic scope. Based on these limitations, several recommendations for future research were made, including expanding the geographic scope, incorporating additional predictive models, integrating qualitative methods, and exploring the impact of technological adoption on seed production and sales. The insights from the study provide a valuable contribution to understanding the complex interplay between various factors affecting seed sales and offer actionable strategies for managers in the seed industry to enhance production efficiency, profitability, and market competitiveness.

REFERENCES

- Adams, J., Khan, H.T.A., Raeside, R. and White, D. 2015. Research Methods for Business and Social Science Students. *2nd ed. New Delhi: Sage Publications.*
- Ahamed, S.F., Vijayasankar, A., Thenmozhi, M., Rajendar, S., Bindu, P. and Subha Mastan Rao, T. 2023. Machine learning models for forecasting and estimation of business operations. *Journal of High Technology Management Research*, 34(2023), 1-8.
- Alon, I., Qi, M. and Sadowski, R.J. 200). Forecasting aggregate retail sales: a comparison of artificial neural networks and traditional methods. *Journal of Retailing and Consumer Services*, 8(3), 147-156.
- Anantha, K.H., Garg, K.K., Barron, J., Dixit, S., Venkataradha, A., Singh, R. and Whitbread, A. M. 2021. Impact of best management practices on sustainable crop production and climate resilience in smallholder farming systems of South Asia. *Journal of Agricultural Systems*, 194 (2021), 1-19.
- Araujo, G.S. and Gaglianone, W.P. 2023. Machine learning methods for inflation forecasting in Brazil: New contenders versus classical models. *Latin American Journal of Central Banking*, 4(2023), 1-29.
- Armstrong, J.S., 2008. Sales Forecasting. SSRN Electronic Journal. Available at: <https://doi.org/10.2139/ssrn.1164602>. Date of access: 26 June 2024.
- Asfaw, D. M. 2021. Analysis of technical efficiency of smallholder tomato producers in Asaita district, Afar National Regional State, Ethiopia. PLoS ONE 16(9): e0257366. <https://doi.org/10.1371/journal.pone.0257366>.
- Backeberg, G. 2020. Water Management in South African Agriculture: Implications for Crop Production and Seed Companies. *Water SA*, 46(2), 145-159.
- Bannikova, N.A., Onezhkina, O.N., Agalarova, E.G. and Tenishchev, A.V. 2018. Forecasting the tendencies of Russian vegetables market development. *Journal of Business and Retail Management Research*, 13(1), 148-155.
- Barney, J. 1991. Firm resources and sustained competitive advantage. *Journal of Management*, 17(1), 99-120.

Barney, J. B. 2012. Purchasing, supply chain management and sustained competitive advantage: The relevance of resource-based theory. *Journal of Supply Chain Management*, 48(2), 3-6.

Bello-Bravo, J., Olana, G. and Pittendrigh, B. (2021). Seed Sector Development in Sub-Saharan Africa: Challenges and Opportunities. *Agriculture and Food Security*, 10(1), 45-65.

Bobbitt Z. 2021. MAE vs. RMSE: Which Metric Should You Use? *Statology*. Available at: <https://www.statology.org/mae-vs-rmse/>. Date of access: June 2024.

Bose, I. and Mahapatra, R.K. 2001. Business data mining—a machine learning perspective. *Journal of Information and Management*, 39(3), pp.211-225. Available at: [https://doi.org/10.1016/s0378-7206\(01\)00091-x](https://doi.org/10.1016/s0378-7206(01)00091-x).

Breiman, L. 2001. Random forests. *Machine Learning*, 45(1), 5-32.

Brown, L., White, M., and Green, R. 2020. Challenges in Seed Production: Climate and Water Management. *Journal of Horticultural Sciences*, 24(7), 512-530.

Brown, M. E. and Funk, C. C. 2008. Food Security Under Climate Change. *Science*, 319(5863), 580-581.

Brownlee L. 2020. Bagging and Random Forest Ensemble Algorithms for Machine Learning ML. *Machine Learning Mastery*. Available at: <https://machinelearningmastery.com/bagging-and-random-forest-ensemble-algorithms-for-machine-learning/>. Date of access: April 2023.

Brownlee, J. 2021. Ensemble Learning Algorithms With Python: Make Better Predictions with Bagging, Boosting, and Stacking. *Machine Learning Mastery*.

Bryman, A., Becker, S. and Sempik, J. 2021. Social research methods: *The essential guide* (6th ed.). Oxford University Press.

Burger, S., Jones, P.C. and Lowe, T.J. 2008. Coordinating the supply chain in the agricultural seed industry. *European Journal of Operational Research*, 185(2008), pp. 354-377.

Chai, T. and Draxler, R.R. 2014. Root means square error (RMSE) or mean absolute error (MAE). *Geoscientific Model Development*, 7(3),1247-1250.

Chanda, S., Bhat, M., Shetty, K. G. and Jayachandran, K. 2021. "Technology, policy, and market adaptation mechanisms for sustainable fresh produce industry: The case of tomato

production in Florida, USA." *Journal of Sustainability*, 13(11), 1-23. Available at: <https://doi.org/10.3390/su13115933>. Date of access: 23 Sept. 2024.

Cheema, D.S. and Dhaliwal, M.S. 2004. Hybrid tomato breeding. *Journal of New Seeds*, 6(2-3),1-14. Available at: https://doi.org/10.1300/J153v06n02_01.

Chen, D., Zeng, J., Xu, Y., and Xie, S. 2019. Data Quality Management and Analysis. *Journal of Data Management*, 7(3), 1625-1640.

Chen, T., Guestrin, C. and Kaplan, S. 2019. XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785-794.

Christopher, M. 2016. Logistics and Supply Chain Management. Pearson Education, 5th Edition. Available at: <https://www.pearson.com/store/p/logistics-and-supply-chain-management/P100000416217>. Date of access: 23 Sept. 2024.

Climates to travel. 2024. Weather and climate in Jiuquan (Gansu). World climate guide Available at: https://www.climatestotravel.com/climate/china/jiuquan#google_vignette. Date of access: 23 Sept. 2024.

Comhaire, P. and Papier, F. 2015. Syngenta Uses a Cover Optimizer to Determine Production Volumes for its European Seed Supply Chain. *Interface*, 45(6), 501-513. <http://dx.doi.org/10.1287/inte.2015.0812>.

Davis, K. F., D'Odorico, P. and Rulli, M. C. 2016. The Global Dependence of Crop Production on Irrigation. *Earth's Future*, 4(10), 488-492.

Demeke, L. and Haji, J. 2014. Econometric analysis of factors affecting market participation of smallholder farming in Central Ethiopia. *Journal of Agricultural Economics, Extension and Rural Development*, 6(2), 94–104.

Derbe, S. Y., Kifle, S.W., Yimenu, S. M., Geleta, D. S. and Woldegiorgis. 2024. Market Orientation and Performance of Smallholder Tomato Producers. *Journal of Agricultural and Resource Economics*, 10 (1), 184-202.

Domingos, P. 2012. A few useful things to know about machine learning. *Communications of the ACM*, 55(10), 78-87.

- Duan, H., Wu, Z., Zhang, Y. and Song, Z. 2023. Technological innovations in seed production: A focus on reducing production costs. *Journal of Agricultural Technology*, 14(3), 34–48.
- Duus, H.J. 2016. Strategic forecasting: The Management Perspective. *Management Research Review*, 39(9), 998-1015.
- Elalem, Y.K., Maier, S. and Seifert, R.W. 2022. A machine learning-based framework for forecasting sales of new products with short life cycles using deep neural networks. *International Journal of Forecasting*, 1(2022), 1-21.
- Ensafi, Y., Amin, S.H., Zhang, G. and Shah, B. 2022. Time-series forecasting of seasonal item sales using machine learning -A comparative analysis. *International Journal of Information Management Data Insights*, 2(2022), 1-16.
- Fiankor, D.D.D., Lartey, A. and Ritzel, C. 2023. Agri-food importing firms amid a global health crisis. *Journal of Food Policy*, 119(2023), 1-13.
- Fiedler, F. 1964. A Contingency Model of Leadership Effectiveness. *Advanced Experimental Social Psychology*, 1(1), 149-190.
- Fitriana, R, Devi, I. and Rinda, A. 2023. Strategic partnerships in emerging markets: A case study. *Journal of Business Research*, 23(2), 97-105.
- Frona, D., Szenderak, J. and Harangi-Rakos, M. 2019. The Challenge of Feeding the World. *International Journal of Sustainability*, 5816(11), 1-18.
- Frost J. 2020. Root Mean Square Error (RMSE). Available at: <https://statisticsbyjim.com/regression/root-mean-square-error-rmse/>. Date of access: March 2024.
- Funk, C., Peterson, P., Landsfeld, M., Pedreros, D., Verdin, J., Rowland, J. and Michaelson, J. 2015. The climate hazards infrared precipitation with stations—a new environmental record for monitoring extremes. *Scientific Data*, 2, 1-21.
- Gaffoor, A., Naidoo, K. and Narain, N. 2020. Assessing the impact of seasonality on agricultural sales: A case study of a seed company. *Journal of Agribusiness and Rural Development*, 16(3), 45-60.

- Gaffoor, Z., Gritzman, A., Pietersen, K., Jovanovic, N., Bagula, A. and Kanyerere, T. 2022. An autoregressive machine learning approach to forecasting high-resolution groundwater-level anomalies in the Ramotswa/North West/Gauteng dolomite aquifers of South Africa. *Journal of Hydrogeology*, 30(2022), 575-600.
- García, J., Gil, J. and Llorente, J. 2016. Feature selection and preprocessing for time series forecasting with neural networks. *Computational Intelligence and Neuroscience*, 2016, 1-13.
- Gebeyehu, B. 2020. Review on: Effects of Seed Storage Period and Storage Environment on Seed Quality. *International Journal of Applied Agricultural Sciences*, 6(6), 185-190.
- Gebremedhin, B. and Jaleta, M. 2012. Market orientation and market participation of smallholders in Ethiopia: Implications for commercial transformation. *IAAE Conference*, Brazil. Available at: <https://doi.org/10.22004/ag.econ.125847>. Date of access: 23 Sept. 2024.
- González, M., Perales, A. and Rojas, F. 2022. Forecasting techniques in agriculture: The role of artificial neural networks in production efficiency. *Agricultural Economics and Technology Review*, 32(1), 89–103.
- Govender, P. and Majozi, T. 2022. Water and energy challenges in South African agriculture: A focus on seed production. *South African Journal of Agricultural Research*, 45(2), 67–82.
- Haji, J. and Wakjira, T. 2021. The role of supply chain inefficiencies in increasing production costs for seed companies in Ethiopia. *East African Journal of Agricultural Economics*, 12(4), 102–116.
- Haselbeck, F., Killinger, J., Menrad, K. and Hannus, T. 2022. Machine Learning Outperforms Classical Forecasting in Horticultural Sales Predictions. *Journal of Machine Learning with Applications*, 7(2022), 1-19.
- Hastie, T., Tibshirani, R. and Friedman, J. 2009. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. *2nd ed.* New York: Springer.
- Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D. and Simmons, A. 2020. *The ERA5 global reanalysis*. *Quarterly Journal of the Royal Meteorological Society*, 146(730),1999-2049.
- Hoare J. 2023. How is Splitting Decided for Decision Trees? *Advanced Analysis. Machine LearningML*. Available at: <https://www.displayr.com/how-is-splitting-decided-for-decision-trees/>. Date of access: November 2023.

Hobbs, J. E. 2021. Food Supply Chains During the COVID-19 Pandemic. *Canadian Journal of Agricultural Economics/Revue canadienne d'agroeconomie*, 68(2), 171-176.

Hochreiter, S., and Schmidhuber, J. 2017. Long short-term memory. *Neural Computation*, 9(8), 1735-1780.

Hoffman, G., Kong, Y. and Zhang, M. 2021. Supply chain management and cost reduction strategies in the seed industry. *International Journal of Agricultural Supply Chains*, 29(2), 123–138.

Huber, J. and Stuckenschmidt, H. 2020. Daily retail demand forecasting using machine learning with an emphasis on calendric special days. *International Journal of Forecasting*, 36(4), 1420-1438.

Hyndman, R. J. and Athanasopoulos, G. 2018. *Forecasting: Principles and Practice* (2nd ed.).

Jones, P.C., Kegler, G., Lowe, T.J. and Traub, R.D. 2003. Managing the Seed-Corn Supply Chain at Syngenta. *Article in Interfaces*, 33(1), 80-90.

Joseph, K., Pienaar, J. and Moyo, S. 2020. Cost analysis in agricultural production: A study of seed companies in South Africa. *South African Journal of Agricultural Extension*, 48(2), 123-135.

Kassaw, H. M., Berhanie, Z. and Alemayehu, G. 2021. Determinants of farm level market supply of tomatoes in Fogera district, South Gondar Zone, Ethiopia. *Ethiopian Journal of Science and Technology*, 14(2), 155–170. Available at: <https://doi.org/10.4314/ejst.v14i2.5>. Date of access: 23 Sept. 2024.

Kohavi, R. 1995. A study of cross-validation and bootstrap for accuracy estimation and model selection. In: *Proceedings of the 14th International Joint Conference on Artificial Intelligence*, 2 (1), 1137-1145.

Kotler, P. and Armstrong, G. 2018. *Principles of Marketing* (17th ed.). Pearson.

Kotler, P. and Keller, K. L. 2016. *Marketing Management* (15th ed.). Pearson.

Kumar, P., Badal, P.S., Paul, R.K., Jha, G.K., Venkatesh, P., Kingsly, I.T., Kamalvanshi, V., Balasubramanian, M. and Anbukkani. 2020. Empowering Farmers through Future Price

Information: A Case Study of Price Forecasting of Brinjal in Eastern Uttar Pradesh. *Indian Journal of Economics and Development*, 16(4), 479-488.

Kumari, P., Patel, R., and Singh, S. (2020). Precision Irrigation Technologies and Their Impact on Cost Efficiency in Agriculture. *Journal of Agricultural Innovation*, 13(2), 150-168.

Leibbrandt, M., Finn, A. and Woolard, I. 2021. Rising Labour Costs in South Africa and the Impact on Agricultural Sectors. *South African Journal of Economics*, 89(3), 321-338.

Liberatore, K.L., Dukowic-Schulze, S., Miller, M.E. Chen, C. and Kainian, S.F. 2016. The role of mitochondria in plant development and stress tolerance. *Free Radical Biology and Medicine*, 100(2016), 238-256.

Little, R. J. A., and Rubin, D. B. 2019. Statistical Analysis with Missing Data. *Wiley Academic Press*. Online ISBN:9781119482260.

Liu, Y. 2024. Sales Forecasting Based on Transformer-LSTM Model. *Highlights in Science, Engineering and Technology CSIC*, 85(2023), 776-782.

Lobell, D. B., Schlenker, W. and Costa-Roberts, J. 2013. Climate Trends and Global Crop Production Since 1980. *Science*, 333(6042), 616-620.

López, E., Martí, J. and Ruiz, A. 2023. Machine learning models for cost forecasting in agriculture. *Computational Agriculture Journal*, 17(2), 56–71.

Mabunda, V., Khoza, N. and Phakathi, M. 2022. SARIMA models in agricultural forecasting: A case study on seed production efficiency. *Journal of African Agricultural Technology*, 20(3), 45–59.

Makridakis, S., Spiliotis, E. and Assimakopoulos, V. 2020. Statistical and Machine Learning Forecasting Methods: Concerns and Ways Forward. *PLOS One*, 15(3). Available at: <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0229455>. Date of access: 23 Sept. 2024.

Makridakis, S., Spiliotis, E. and Assimakopoulos, V. 2020. The M4 Competition: 100,000-time series and 61 forecasting methods. *International Journal of Forecasting*, 36(1), 54-74.

Malherbe, E. 2015. Economics of groundwater abstraction and the impact of drought. *Journal of Environmental Economics*, 54(3), 321-335.

Mendelsohn, R., Nordhaus, W. D. and Shaw, D. 2006. The Impact of Climate Variation on Agriculture. *The American Economic Review*, 84(4), 753-771.

Miguéis, V.L., Pereira, A., Pereira, J. and Figueira, G. 2022. Reducing fresh fish waste while ensuring availability: Demand forecast using censored data and machine learning. *Journal of Cleaner Production*, 359 (2022), 1-11.

Moon, M.A., Mentzer, J.T. and Smith, D.C. 2003. They are conducting a sales forecasting audit. *International Journal of Forecasting*, 19(2003), 5-25.

Muñoz-Sabater, J., 2019. ERA5-Land monthly averaged data from 1981 to present. Copernicus Climate Change Service (C3S) Climate Data Store (CDS). Date of access: 24 Sept. 2024. Available at: doi:10.24381/cds.68d2bb30.

Mwangi, G. and Kamau, E. 2022. Environmental variability and production costs: Climate-resilient strategies for seed companies. *Journal of Agricultural Sustainability*, 8(1), 15–32.

Narver, J.C., and Slater, S.F. 1990. The effect of a market orientation on business profitability. *Journal of Marketing*, 54(4), 20-35.

Ncoyini, Z., Savage, M.J. and Strydom, S. 2022. Limited access and use of climate information by small-scale sugarcane farmers in South Africa: A case study. *Journal of Climate Services*, 26(2022), 2-11.

Ncube, P. and Mhlanga, S. 2020. Localised Production in Africa: Leveraging Indigenous Seed Varieties for Agricultural Sustainability. *African Journal of Agronomy*, 19(1), 55-72.

Pereira, A., Miguéis, V.L. and Figueira, G. 2015. Demand forecasting for fresh products in the retail sector: A case study. *Procedia Computer Science*, 55 (2015), 396-405.

Perez, K., Froikin-Gordon, J.S., Abdourhamane, I. K., Levasseur, V., Alfari, A.A., Mensah, A., Bonsu, O., Habsatou, B., Assogba-Komlan, F., Mbaye, A.A., Noussourou, M., Otoïdobiga, L.C., Ouedraogo, L., Kon, T., Rojas, M.R., Gamby, K.T., Shotoski, F., Gilbertson, R.L. and Jahn, M.M. 2017. Connecting smallholder tomato producers to improved seed in West Africa. *Journal of Agriculture and Food Security*, 42(6), 1-14.

Pick n Pay. 2024 Available at: <https://www.picknpayinvestor.co.za/pnp-at-a-glance.php>. Date of access: 23 Sept. 2024.

Porter, M. E. 2011. Competitive Advantage of Nations: Creating and Sustaining Superior Performance. *Library of Congress Cataloging-in Publication Data*. ISBN No: 0-684-84146-0.

Reynoso, E.V. and Intal, G.L. 2022. Prototyping for an In-month Seed Forecasting System. *Proceedings of the International Conference on Industrial Engineering and Operations Management Nsukka, Nigeria*. 1032 -1042.

Rodell, M., Houser, P. R., Jambor, U., Gottschalck, J., Mitchell, K., Meng, C. J., Arsenault, K., Cosgrove, B., Radakovich, J., Bosilovich, M. and Entin, J. K. 2007. The global land data assimilation system. *Bulletin of the American Meteorological Society*, 85(3),381-394.

Rogers, E.M. 2003. Diffusion of Innovations. *5th ed. New York: Free Press*.

S and P Global. 2024. Seed and GM Crop Market Analysis. Available at: <https://www.spglobal.com/commodityinsights/en/ci/products/crop-science-seed.html>. Date of access: 23 Sept. 2024.

Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., Saisana, M. and Tarantola, S. 2008. Global sensitivity analysis: the primer. *John Wiley and Sons*. Available at: <https://doi.org/10.1002/9780470725184>.

Saunders, M., Thornhill, A. and Lewis, P. 2016. Research methods for business students (7th ed.). *Pearson Education*.

Schlenker, W., and Roberts, M. J. 2009. Nonlinear Temperature Effects Indicate Severe Damages to U.S. Crop Yields Under Climate Change. *Proceedings of the National Academy of Sciences*, 106(37), 15594-15598.

Schmidhuber, J. and Tubiello, F. N. 2007. Global Food Security Under Climate Change. *Proceedings of the National Academy of Sciences*, 104(50), 19703-19708.

Schnaubelt, M. 2019. A comparison of machine learning model validation schemes for non-stationary time series data. *FAU Discussion Papers in Economics*, 11(2019), 1-20.

Scott, W.R. 1995. Institutions and Organizations. Ideas, Interests and Identities. Available at: <https://shs.cairn.info/revue-management-2014-2-page-136?lang=fr>. Date of access: 23 Sept. 2024.

Seed World. 2024. Improved Tomato Varieties Impact Global Food Security. Available at: <https://www.seedworld.com/us/2022/08/04/improved-tomato-varieties-impact-global-food-security-2/>. Date of access: 23 Sept. 2024.

Singh, K., Booma, P.M. and Eaganathan, U. 2020. E-commerce systems for sales prediction using machine learning techniques. *Journal of Physics: Conference Series*, 1712(2020), 1-9. <https://doi.org/10.1186/s42400-018-0005-8>.

Smith, J. and Jones, A., 2019. Agronomic Practices for Tomato Cultivation: A Review. *Agricultural Research Journal*, 15(4), 221-233.

South African National Seed Organisation (SANSOR). 2020. Annual report. <https://www.sansor.org/content/uploads/2020/06/SANSOR-2019-2020-Annual-Report.pdf>. Date of access: 4 April 2023.

South African Seed Organisation (SANSOR). 2023. Annual report. <https://sansor.org/annual-report-2021-2022/docs/2022/Sansor-Seed-Analyst-Standing-Committee.pdf>. Date of access: 23 October 2023.

Srivastava R, Kumar S, Kumar B. 2023. Classification model of machine learning for medical data analysis. Chapter 7. *Statistical Modeling in Machine Learning*. Academic Press, Pages 111-132 ISBN 9780323917766.

Syngenta Foundation, 2015. Managing the Seed-Corn Supply Chain at Syngenta. Syngenta Foundation Report.

Thakur, R. 2023. Price Forecasting of Brinjal in Eastern Uttar Pradesh. *Journal of Agricultural Economics*, 78(4), 541-554.

Trapero, J.T., Kourentzes, N. and Fildes, R. 2014. On the identification of sales forecasting models in the presence of promotions. *Journal of the Operational Research Society*, 6(3), 1-20.

Usuga, C. J., Lamouri, S. and Grabot, B. 2018. Demand forecasting with machine learning and time series models: A study in the seed industry. *Computers and Industrial Engineering*, 126, 303-316.

Velasco-Munoz, J.F., Aznar-Sanchez, J.A., Fuente, A.B. and Fidelibus, M.D. 2019. Sustainable irrigation in agriculture: An analysis of global research. *Journal of Water Management*, 11(9), 1-20.

Wang, J. 2020. A hybrid machine learning model for sales prediction. Conference: 2020 International Conference on Intelligent Computing and Human-Computer Interaction (ICHCI). Available at: DOI:10.1109/ICHCI51889.2020.00083. Date of access: 24 Sept. 2024.

Wei, F., Min, Z., Jean-Michel, L., Richard, B., Hou-Tse, H. and Jun, X. 2013. Evaluation of groundwater depletion in North China using the gravity recovery and climate experiment (GRACE) data and ground-based measurements. *Journal of Water Resources Research*, 49(4), 2110-2118.

Willmott, C.J. and Matsuura, K. 2005. Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Climate Research*, 30(1), 79-82.

Xhemali, D., Spahiu, T. and Selimi, I. 2024. Preprints on Hybrid Seed Production and Market Analysis. *Preprints*, 1-15.

Yiu T. 2019. Understanding Random Forest: How the Algorithm Works and Why it Is So Effective. *Towards Data Science*. Available at: <https://towardsdatascience.com/understanding-random-forest-58381e0602d2>. Date of access: 29 Sept. 2024.

Zhang, X., Cai, X. and Lobell, D. B. 2018. Climate Adaptation Strategies for Agricultural Sustainability. *Nature Climate Change*, 8(10), 785-789.

Zhu, Y., Zhao, Y., Zhang, J., Geng, N. and Huang, D. 2019. Spring onion seed demand forecasting using a hybrid Holt-Winter and support vector machine model. *PLoS One*, 14(7), e0219889.

APPENDIX

"Admin"

- Connect google colab to your drive
- Import libraries
- Provide path to where excel is stored in drive

```
[ ]
from google.colab import drive
drive.mount('/content/drive')
Mounted at /content/drive
```

```
[ ]
# Import necessary libraries
import pandas as pd
import numpy as np
from sklearn.model_selection import train_test_split
from sklearn.model_selection import GridSearchCV
from sklearn.ensemble import RandomForestRegressor
from sklearn.metrics import mean_squared_error, r2_score
import matplotlib.pyplot as plt
from sklearn.preprocessing import StandardScaler
```

```
[ ]
#Load the data into a dataframe using Python and the pandas library
path = "/content/drive/MyDrive/GColab_Files/TrainingDataSales4.csv"
```

```
# reading the data
df = pd.read_csv(path)
```

```
[ ]
# Check the loaded data
print(df.head())
```

	Year	Production	Inset Costs	Price	Sales	Profit
0	2013	9344000	1142453	920	3402370	2259917
1	2014	7240000	1079386	1020	3641019	2561633
2	2015	7833000	1220477	1180	5013212	3792735
3	2016	12739000	1906902	1200	8300363	6393461
4	2017	18433000	1080360	1325	11042917	9962557

keyboard_arrow_down

General dataframe operations

View the data Only keep data from 5 boreholes for illustration

```
[ ]
# Check the data types and handle any missing data if necessary
print(df.info())
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 11 entries, 0 to 10
Data columns (total 6 columns):
#   Column      Non-Null Count  Dtype
---  ---
0   Year        11 non-null    int64
1   Production  11 non-null    int64
2   Inset Costs 11 non-null    int64
3   Price       11 non-null    int64
4   Sales       11 non-null    int64
5   Profit      11 non-null    int64
dtypes: int64(6)
memory usage: 656.0 bytes
None
```

```
[ ]
# Strip any leading or trailing spaces from the column names
df.columns = df.columns.str.strip()
print(df.columns)
Index(['Year', 'Production', 'Inset Costs', 'Price', 'Sales', 'Profit'], dtype='object')
```

keyboard_arrow_down

Random Forest Model

```
[ ]
# Define the features and target variable
X = df.drop(columns=['Sales','Profit'])
y = df['Profit']
```

```
[ ]
```

```
# Split the data into training and testing sets (70% training, 30% testing)
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3, random_state=0)
```

keyboard_arrow_down

Hyperparameter Tuning and Model Training using optimal parameters

```
[]
# Define the hyperparameters to tune using Grid Search
param_grid = {
    'n_estimators': np.arange(10, 151, 10), # Trees from 10 to 150 at intervals of 10
    'max_depth': [4, 6, 8], # Depth from 4 to 8
    'max_samples': np.arange(0.1, 0.7, 0.1), # Bagging fraction from 0.1 to 0.6
    'max_features': [2, 3, 4] # Use a range of max features
}
```

```
[]
# 6. Create the Random Forest Regressor model
rf = RandomForestRegressor(random_state=42)
```

```
[]
# 7. Perform Grid Search for hyperparameter tuning
grid_search = GridSearchCV(estimator=rf, param_grid=param_grid,
                           cv=5, scoring='neg_mean_squared_error',
                           verbose=2, n_jobs=-1) # Use all CPU cores for speed
```

```
[]
# 8. Train the Grid Search model on the training data
grid_search.fit(X_train, y_train)
```

```
[]
# 9. Get the best parameters from the Grid Search
best_params = grid_search.best_params_
# Extract the results of the grid search
cv_results = grid_search.cv_results_
```

```
[]
# 10. Print the best hyperparameters found by Grid Search
print(f"Best parameters: {best_params}")
Best parameters: {'max_depth': 4, 'max_features': 2, 'max_samples': 0.6, 'n_estimators': 50}
```

keyboard_arrow_down

Apply trained model for testing/forecasting

```
[]
# 11. Use the best model for prediction
best_rf = grid_search.best_estimator_
y_test_pred = best_rf.predict(X_test) # Predict on the test set using the trained model # Step 2: Predict on the test set
```

```
y_train_pred = best_rf.predict(X_train) # Predict on the training set for comparison (optional) # Step 1: Predict on the training set
```

keyboard_arrow_down

Variable Importance

```
[ ]
feature_names = ['Year', 'Production', 'Inset Costs', 'Price']
feature_importances = best_rf.feature_importances_

best_rf

# Create a DataFrame to display feature importance as percentages
importance_df = pd.DataFrame({
    'Feature': feature_names,
    'Importance (%)': feature_importances * 100 # Convert to percentages
})

# Sort by importance in descending order for better readability
importance_df = importance_df.sort_values(by='Importance (%)', ascending=False)

# Print the feature importance percentages
print(importance_df)
```

	Feature	Importance (%)
0	Year	29.858414
1	Production	29.413675
3	Price	27.382386
2	Inset Costs	13.345525

keyboard_arrow_down

Model performance evaluation

```
[ ]
# Step 11: Evaluate the model performance
rmse = np.sqrt(mean_squared_error(y_test, y_test_pred))
r2 = r2_score(y_test, y_test_pred)

print(f"Test RMSE: {rmse}")
print(f"Test R²: {r2}")
Test RMSE: 1730102.1125550515
Test R²: 0.8059855530233189
```

```
[ ]
# 12a: Combine actual and predicted values with the corresponding years for training and testing sets
train_results = pd.DataFrame({
    'Year': X_train.index, # Assuming Year is the index in your dataframe
    'Actual Profit (Train)': y_train,
    'Predicted Profit (Train)': y_train_pred
})

test_results = pd.DataFrame({
```

```

'Year': X_test.index, # Assuming Year is the index in your dataframe
'Actual Profit (Test)': y_test,
'Predicted Profit (Test)': y_test_pred
})

```

Step 4: Concatenate training and testing results into one DataFrame

```
all_results = pd.concat([train_results, test_results], axis=0)
```

```

[]
# 12b: Export the combined results to an Excel file
all_results.to_excel('/content/drive/MyDrive/GColab_Files/actual_vs_predicted_Profit70training30testi
ngbwithYr9081b.xlsx', index=False)

# 12c: Confirm that the file was saved successfully
print("Excel file with actual and predicted Profit values has been saved.")
Excel file with actual and predicted Profit values has been saved.

```

keyboard_arrow_down

ForeCasting

```

[]
# Load the data into a dataframe using Python and the pandas library
path2 = "/content/drive/MyDrive/GColab_Files/ForeCasting3.csv"
# reading the data
df2 = pd.read_csv(path2)
# View the data as is
df2.head()

```

```

[]
# Extract the features from the new data that match what the model was trained on
# Assume the features are 'Production', 'Inset Costs', and 'Price'
X_new = df2[['Year', 'Production', 'Inset Costs', 'Price']]

```

```

[]
# Use the trained Random Forest model to predict sales for the new data
predicted_Profit = best_rf.predict(X_new)

```

```

[]
# Add the predictions back to the original dataframe
df2['Predicted Profit'] = predicted_Profit

# Display the dataframe with the predictions
print(df2.head())

# Optionally, save the predictions to a new CSV file
df2.to_csv("/content/drive/MyDrive/GColab_Files/Predicted_Profit_ForeCasting2Profitd04WithYr9081
b.csv", index=False)

```

	Year	Production	Inset Costs	Price	Predicted Profit
0	2024	18828400	1187389	1991.000000	11874267.52
1	2025	19668850	1178079	2053.712121	12105977.26
2	2026	20509290	1168769	2145.872960	12490764.60
3	2027	21349740	1159459	2238.033800	12490764.60
4	2028	22190180	1150149	2330.194639	12490764.60

"Admin"

- Connect google colab to your drive
- Import libraries
- Provide path to where excel is stored in drive

```
[ ]
from google.colab import drive
drive.mount('/content/drive')
Mounted at /content/drive
```

```
[ ]
# Import necessary libraries
import pandas as pd
import numpy as np
from sklearn.model_selection import train_test_split
from sklearn.model_selection import GridSearchCV
from sklearn.ensemble import RandomForestRegressor
from sklearn.metrics import mean_squared_error, r2_score
import matplotlib.pyplot as plt
from sklearn.preprocessing import StandardScaler
```

```
[ ]
#Load the data into a dataframe using Python and the pandas library
path = "/content/drive/MyDrive/GColab_Files/TrainingDataSales4.csv"
# reading the data
df = pd.read_csv(path)
```

```
[ ]
# Check the loaded data
print(df.head())
Year Production Inset Costs Price Sales Profit
0 2013 9344000 1142453 920 3402370 2259917
1 2014 7240000 1079386 1020 3641019 2561633
2 2015 7833000 1220477 1180 5013212 3792735
3 2016 12739000 1906902 1200 8300363 6393461
4 2017 18433000 1080360 1325 11042917 9962557
```

keyboard_arrow_down

General dataframe operations

View the data Only keep data from 5 boreholes for illustration

```
[ ]
# Check the data types and handle any missing data if necessary
print(df.info())
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 11 entries, 0 to 10
Data columns (total 6 columns):
```

```
# Column    Non-Null Count  Dtype
---  ---
0 Year      11 non-null   int64
1 Production 11 non-null   int64
2 Inset Costs 11 non-null   int64
3 Price      11 non-null   int64
4 Sales      11 non-null   int64
5 Profit     11 non-null   int64
dtypes: int64(6)
memory usage: 656.0 bytes
None
```

```
[]
# Strip any leading or trailing spaces from the column names
df.columns = df.columns.str.strip()
print(df.columns)
Index(['Year', 'Production', 'Inset Costs', 'Price', 'Sales', 'Profit'], dtype='object')
```

keyboard_arrow_down

Random Forest Model

```
[]
# Define the features and target variable
X = df.drop(columns=['Sales', 'Profit'])
y = df['Sales']
```

```
[]
# Split the data into training and testing sets (70% training, 30% testing)
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3, random_state=0)
```

keyboard_arrow_down

Hyperparameter Tuning and Model Training using optimal parameters

```
[]
# Define the hyperparameters to tune using Grid Search
param_grid = {
    'n_estimators': np.arange(10, 151, 10), # Trees from 10 to 150 at intervals of 10
    'max_depth': [4, 6, 8], # Depth from 4 to 8
    'max_samples': np.arange(0.1, 0.7, 0.1), # Bagging fraction from 0.1 to 0.7
    'max_features': [2, 3, 4] # Use a range of max features
}
```

```
[]
# 6. Create the Random Forest Regressor model
rf = RandomForestRegressor(random_state=42)
```

```
[]
# 7. Perform Grid Search for hyperparameter tuning
grid_search = GridSearchCV(estimator=rf, param_grid=param_grid,
```

```
cv=5, scoring='neg_mean_squared_error',  
verbose=2, n_jobs=-1) # Use all CPU cores for speed
```

```
[]  
# 8. Train the Grid Search model on the training data  
grid_search.fit(X_train, y_train)
```

```
[]  
# 9. Get the best parameters from the Grid Search  
best_params = grid_search.best_params_  
# Extract the results of the grid search  
cv_results = grid_search.cv_results_
```

```
[]  
# 10. Print the best hyperparameters found by Grid Search  
print(f"Best parameters: {best_params}")  
Best parameters: {'max_depth': 4, 'max_features': 2, 'max_samples': 0.6, 'n_estimators': 140}
```

keyboard_arrow_down

Apply trained model for testing/forecasting

```
[]  
# 11. Use the best model for prediction  
best_rf = grid_search.best_estimator_  
y_test_pred = best_rf.predict(X_test) # Predict on the test set using the trained model # Step 2: Predict on  
the test set  
y_train_pred = best_rf.predict(X_train) # Predict on the training set for comparison (optional) # Step 1: P  
redict on the training set
```

keyboard_arrow_down

Variable Importance

```
[]  
feature_names = ['Year', 'Production', 'Inset Costs', 'Price']  
feature_importances = best_rf.feature_importances_  
  
best_rf  
  
# Create a DataFrame to display feature importance as percentages  
importance_df = pd.DataFrame({  
    'Feature': feature_names,  
    'Importance (%)': feature_importances * 100 # Convert to percentages  
})  
  
# Sort by importance in descending order for better readability  
importance_df = importance_df.sort_values(by='Importance (%)', ascending=False)  
  
# Print the feature importance percentages  
print(importance_df)  
Feature Importance (%)  
0 Production 49.901173
```

```
2    Price    27.875706
1 Inset Costs    22.223121
```

keyboard_arrow_down

Apply trained model for predictions

```
[]
# Step 11: Evaluate the model performance
rmse = np.sqrt(mean_squared_error(y_test, y_test_pred))
r2 = r2_score(y_test, y_test_pred)
```

```
print(f"Test RMSE: {rmse}")
print(f"Test R²: {r2}")
Test RMSE: 2449437.3098725723
Test R²: 0.6006495501718533
```

```
[]
# 12a: Combine actual and predicted values with the corresponding years for training and testing sets
train_results = pd.DataFrame({
    'Year': X_train.index, # Assuming Year is the index in your dataframe
    'Actual Sales (Train)': y_train,
    'Predicted Sales (Train)': y_train_pred
})
```

```
test_results = pd.DataFrame({
    'Year': X_test.index, # Assuming Year is the index in your dataframe
    'Actual Sales (Test)': y_test,
    'Predicted Sales (Test)': y_test_pred
})
```

```
# Step 4: Concatenate training and testing results into one DataFrame
all_results = pd.concat([train_results, test_results], axis=0)
```

```
[]
# 12b: Export the combined results to an Excel file
all_results.to_excel('/content/drive/MyDrive/GColab_Files/actual_vs_predicted_sales70training30testin
gbwithYr9081b.xlsx', index=False)
```

```
# 12c: Confirm that the file was saved successfully
print("Excel file with actual and predicted sales values has been saved.")
Excel file with actual and predicted sales values has been saved.
```

keyboard_arrow_down

ForeCasting

```
[]
# Load the data into a dataframe using Python and the pandas library
path2 = "/content/drive/MyDrive/GColab_Files/ForeCasting3.csv"
# reading the data
df2 = pd.read_csv(path2)
# View the data as is
df2.head()
```

```

[]
# Extract the features from the new data that match what the model was trained on
# Assume the features are 'Production', 'Inset Costs', and 'Price'
X_new = df2[['Year','Production', 'Inset Costs', 'Price']]

```

```

[]
# Use the trained Random Forest model to predict sales for the new data
predicted_sales = best_rf.predict(X_new)

```

```

[]
# Add the predictions back to the original dataframe
df2['Predicted Sales'] = predicted_sales
# Display the dataframe with the predictions
print(df2.head())
# Optionally, save the predictions to a new CSV file
df2.to_csv("/content/drive/MyDrive/GColab Files/Predicted Sales ForeCasting2Salesd03With
Yr9081b.csv", index=False)

```

	Year	Production	Inset Costs	Price	Predicted Sales
0	2024	18828400	1187389	1991.000000	13033348.10
1	2025	19668850	1178079	2053.712121	13295839.94
2	2026	20509290	1168769	2145.872960	13735152.70
3	2027	21349740	1159459	2238.033800	13735152.70
4	2028	22190180	1150149	2330.194639	13735152.70