SMART ASSISTANCE FOR THE FLORICULTURE INDUSTRY

Prabhashi P.A.N.

(IT20017088)

B.Sc. (Hons) Degree in Information Technology Specialized in Data Science

Department of Information Technology

Sri Lanka Institute of Information Technology
Sri Lanka

September 2023

SMART ASSISTANCE FOR THE FLORICULTURE INDUSTRY

2023-133

Final Project Thesis

Prabhashi P.A.N.

(IT20017088)

Dr. Sanvitha Kasthuriarachchi
Dr. Lakmini Abeywardhana

The dissertation was submitted in partial fulfilment of the requirements for the B.Sc. (Honors) degree in Information Technology specialized in Data Science

Department of Information Technology

Sri Lanka Institute of Information Technology
Sri Lanka

February 2023

DECLARATION

I declare that this is my own work, and this dissertation does not incorporate without acknowledgement any material previously submitted for a degree or diploma in any other university or Institute of higher learning and to the best of my knowledge and belief it does not contain any material previously published or written by another person except where the acknowledgement is made in the text.

Also, I hereby grant to Sri Lanka Institute of Information Technology the non-exclusive right to reproduce and distribute my dissertation in whole or part in print, electronic or other medium. I retain the right to use this content in whole or part in future works (such as articles or books).

Name	**Student ID**	**Signature**
Prabhashi P.A.N.	IT20017088	

The above candidates are carrying out research for the undergraduate Dissertation under my supervision.

Signature of the Supervisor   Date

…………………………….   ..………………………

(Dr. Sanvitha Kasthuriarachchi)

Abstract

Floriculture is a burgeoning industry in Sri Lanka, driven by the country's favorable climate and soil conditions. The production of tropical cut flowers has made Sri Lanka a sought-after destination for flower growers and exporters worldwide. While the local market for flowers has been growing, most of Sri Lanka's floriculture products are exported to markets such as the United States, Europe, and Japan. Sri Lanka has been successful in producing a range of tropical flowers such as anthuriums, orchids, heliconias, and ginger lilies. However, despite its growth, the industry still faces challenges that can impact its long-term sustainability, such as high production costs, and limited research and development on the technical side. A smart assistance system that employs cutting-edge technologies such as Machine learning algorithms, Natural Language Processing, Predictive analytics software to forecast demand, and other technology for mapping and visualizing production and distribution data is required to overcome these obstacles. The proposed system aims to develop a demand prediction system for the floriculture industry, utilizing needed technologies to support this project. This will streamline the supply chain and increase efficiency, ultimately reducing costs and improving customer satisfaction.

Keywords: Floriculture, tropical cut flowers, Machine learning algorithms, Machine learning algorithms, Natural Language Processing, Predictive analytics, forecast demand

ACKNOWLEDGEMENT

I would like to take this moment to express my heartfelt gratitude to all those who played a crucial role in supporting and assisting me throughout the successful completion of the project component.

First and foremost, I want to extend my sincere appreciation to the CDAP team and the research panel. Their unwavering diligence and unwavering commitment during the course were instrumental in shaping our educational experience and ensuring the effective culmination of our research project. Their dedication has truly been invaluable.

I am particularly grateful to our project supervisor, Dr. Sanvitha Kasthuriarachchi, and our co-supervisor, Dr. Lakmini Abeywardhana. Their steadfast leadership and insightful suggestions were like guiding lights that steered our research in the right direction. Their mentorship made a profound difference in our project's success.

I must also acknowledge our external supervisor, Mr. Anandathissa Illeperuma, whose invaluable advice and guidance significantly contributed to the project's quality. His expertise and support were critical in navigating the complexities of our research.

A heartfelt thank you goes out to the entire team at Omega Green Pvt. Ltd. Their assistance in data collection, sharing of knowledge, and invaluable advice were pivotal in enhancing the depth and quality of our project. Their collaboration was truly appreciated.

I also want to express my gratitude to my dedicated colleagues within the research group. Their constant assistance and cooperation fostered a collaborative environment that was essential for our project's success. Working together as a team was both rewarding and inspiring.

Lastly, but certainly not least, I am deeply thankful for the unwavering support and encouragement of my family and friends. Their unwavering belief in me has been a constant source of motivation throughout this journey, and I could not have completed this project without their encouragement.

TABLE OF CONTENTS

DECLARATION ii
Abstract iii
ACKNOWLEDGEMENT iv
LIST OF FIGURES vi
LIST OF TABLES vii
LIST OF ABBREVIATIONS vii
LIST OF APPENDICES vii
1 INTRODUCTION 1
1.1 Background & Literature Survey 2
1.2 Research Problem 4
1.3 Research Gap 9
1.4 Research Objectives 12
1.4.1 Main Objective 12
1.4.2 Specific Objectives 12
1.5 Project Requirements 13
1.5.1 Functional requirements 13
1.5.2 Non-functional requirements 13
1.5.3 System requirements 14
1.5.4 User requirements 14
2 METHODOLOGY 15
2.1 Commercialization & Business Plan 18
2.2 Implementation 18
2.2.1 Demand Prediction Implementation 18
2.2.2 Backend – Flask API 31
2.2.3 Frontend – React Native 32
..... 34
3 Results & Discussion 35
3.1 Results 35
3.2 Research Findings 36
3.3 Discussion 38
4 SUMMARY OF EACH STUDENT’S CONTRIBUTION 39
5 CONCLUTION 40
REFERENCES 42

LIST OF FIGURES

Figure 1-1: Survey to find the method that is currently used to calculate the demand ………............... 5
Figure 1-2: Survey to find the method that is currently used to calculate the demand ………............... 6
Figure 1-3: Survey to identify the best sources for forecasting the demand 7
Figure 1-4: Survey question for the need of AI and machine learning for demand forecasting 8

Figure 2-1: Overall system diagram 15
Figure 2-2: Individual system diagram 16
Figure 2-3: Agile methodology 17
Figure 2-4: Data set collection 19
Figure 2-5: Dataset after handling the missing values 20
Figure 2-6: Temperature before and after outliers handling 21
Figure 2-7: Humidity before and after outliers handling 21
Figure 2-8: Rainfall before and after outliers handling 22
Figure 2-9: Transformed data according to a certain format 23
Figure 2-10: Data after using one-hot encoding for the categorical variables ………...................... 23
Figure 2-11: Data after using Min-Max scaler for the numerical variables 23
Figure 2-12: Exploratory Data Analysis (EDA) 24
Figure 2-13: Frequencies of categorical variables 25
Figure 2-14: Distributions of the numerical variables 25
Figure 2-15: Correlation heat map 26
Figure 2-16: Best model according to the specific plant type 30
Figure 2-17: 'Plant Pal' main page 34
Figure 2-18: 'Plant Pal' main selective page 34
Figure 2-19: Demand forecasting selective page 34
Figure 2-20: Demand forecasting user inputs form 34

Figure 3-1: Postman test results 36

LIST OF TABLES

Table 1-1 Past research 11
Table 2-1 Summary of tools and technologies used........ 17
Table 4-1: Individual Contribution 39

LIST OF ABBREVIATIONS

ANN	Artificial Neural Network
SVM	Support Vector Machine
LSTM	Long Short-Term Memory
IoT	Internet of Things
MSE	Mean Squared Error
NLP	Natural Language Processing
MAE	Mean Absolute Error
R2	R-Squared
RL	Reinforcement Learning
RNN	Recurrent Neural Network
ROC-AUC	Receiver Operating Characteristic Area Under the Curve

LIST OF APPENDICES

Appendix 1 - Survey for floriculture industry 44
Appendix 2 - Gantt Chart 44
Appendix 3 - Work Breakdown Structure 45
Appendix 4 - Budget estimation 45
Appendix 5 - Model building (Random Forest) 46
Appendix 6 - Model building (Gradient Boosting) 46
Appendix 7 - Model building (MLP Regressor) 47
Appendix 8 - Model building (Linear Regression) 47

Appendix 9 - Model building (SARIMA) 48
Appendix 10 - Flask API 49
Appendix 11 - Model building (R squared values for the variables) 50

1 INTRODUCTION

Sri Lanka's floriculture industry has been growing rapidly over the past few years, with increasing demand for Sri Lankan floriculture products in global markets. In 2014, the industry generated US$ 14.9 million in foreign exchange earnings, with an average growth rate of 5% over the previous year. The European market is the primary destination for Sri Lankan floriculture products, with 60% of exports headed to this region. Other key markets include Japan, the Middle East, the USA, and Korea. To remain competitive in today's global market, it is crucial for the floriculture industry to embrace the latest technologies such as machine learning, artificial intelligence, and more. [1]

Due to the gap between supply and demand the industry may cause several negative consequences. Firstly, inaccurate demand forecasting can result in lost sales opportunities, as the industry may produce more products than there is demand for. This can lead to excess inventory and reduced profitability. Additionally, overproduction can result in wastage of resources such as water, fertilizer, and labor, and if the industry is unable to sell all its products, it may need to dispose of them, which can have negative environmental impacts. On the other hand, if the industry is unable to predict demand accurately, there may be times when it cannot meet customer demand, leading to lost sales opportunities and potential damage to the industry's reputation. Therefore, accurate demand forecasting for the future supply is crucial for the success of the floriculture industry in Sri Lanka.

Accurate demand forecasting plays a pivotal role in the floriculture industry as it empowers businesses to discern the ever-changing preferences for flowers, foliage, and ornamental plants across various regions of the world. By gaining insight into these dynamic market trends, companies can strategically allocate resources to produce the right assortment of products in optimal quantities. This not only minimizes wastage but also boosts profitability substantially.

Furthermore, precise demand forecasting equips businesses with the ability to plan production cycles meticulously, maintain efficient inventory levels, and streamline their logistics operations. The cumulative effect of these measures leads to the establishment of more stable and dependable supply chains. As a result, customer satisfaction levels are elevated, and the industry gains a competitive edge in the global market.
To enhance the accuracy of demand forecasting, modern technology, such as predictive analytics software and mobile applications, has been harnessed. These advanced tools analyze a myriad of factors, including product types and market demand, to determine the most appropriate packaging materials for each product.

This strategic approach not only minimizes the risk of product damage during transportation but also elevates the overall quality of products delivered to customers.

It's worth noting that while there have been various research endeavors in the field of market demand forecasting and supply prediction, the floriculture industry remains an understudied domain in this regard. Only a limited number of studies have ventured into addressing these critical aspects, and those have often only scratched the surface. Therefore, there is ample room for further exploration and innovation in demand forecasting within the floriculture sector.

1.1 Background & Literature Survey

According to some research, this floriculture industry is kind of uncommon and still developing into modern technology. Most of the research has been done in other agricultural areas and most of the technologies implemented in other countries but not in Sri Lanka. So, this proposed system will be novel and have a significant gap according to another research.

The Smart Intelligent Floriculture Assistant Agent (SIFAA) system integrates expert knowledge with cutting-edge techniques such as deep learning to diagnose plant diseases and recommend treatments. It also uses Reinforcement Learning to recommend the best products for customers and employs demand forecasting to motivate cultivators. Additionally, SIFAA utilizes Linear Regression and ensemble advanced LightGBM Regressors techniques to apply feature engineering. This system addresses major barriers in the floriculture industry and provides solutions for plant development monitoring, pest identification, maintaining export standards, predicting demand and supply, selecting plants, and giving proper care. [2]

This research is about the sugar production in Sri Lanka which is currently fulfilling a very small portion of the demanded amount. ARIMA model and other machine learning methods are used to develop the forecast of the production and the SVM model is used to find the accuracy. Using all these machine learning models they have predicted production and more. [3]

According to this study, multiple machine languages were used to predict yield and prices. The yield was forecast based on environmental factors, whereas the price was forecast based on supply and demand, import and export, and seasonal effect. [4]

This research is conducted in India to predict the crops which can improve yields and reduce losses. They have used Logistic Regression, LSTM, and RNN to the prediction [5]

As for the capturing GPS technology for agriculture, this system supports the IOS and Android mobile phones to be used to develop real-time agricultural information collection systems. [6]

In agriculture, where seasonality is a significant factor, time series analysis is employed for forecasting. According to the study, several models, including AR, MV, ARIMA, and their variants, have been utilized. The precision of machine learning techniques such as ANN, RNN, LSTM, and GRU has increased. However, this application considers additional factors for demand forecasting beyond seasonality. The application is developed for agriculture, which differs from floriculture in terms of demand. [7]

Agricultural marketing is crucial for the Indian economy, as farmers rely on it for their livelihood. However, they often face challenges such as low prices for their crops and middlemen taking profits. To address this, it has been proposed that a project be developed to analyze customer purchase data from various regions and to facilitate direct communication between customers and producers. The proposed solution involves using the K-means algorithm to categorize agricultural products based on their prices in different regions. The model will be tested to analyze purchase data and the effectiveness of the proposed algorithm will be evaluated by analyzing clusters. [8]

In addition to these studies, there have been others that investigate agricultural industry supply forecasting. In addition, we employed a variety of machine learning strategies to achieve our objectives.[9], [10]
Also, when considering the recommendation systems that have developed in this agriculture or floricultural area multiple research projects were carried out. The crops are recommended using various factors such as soil condition, weather, humidity, and more. To construct this linear regression, logistic regression, SVM, K-means, and KNN were used as analytic tools. [11]

This research is conducted utilizing online technologies, data mining, and geographic information systems for the development of an expert decision system for agriculture. [12]

The proposed solution aims to close the distance by analyzing data regarding customer purchases and providing direct contact channels. The use of the K-means algorithm to categorize commodities based on prices will provide insights into the demands of customers in different regions, and eliminate the need for middlemen, ensuring that farmers get better prices for their produce. The system will also be useful for policymakers and other stakeholders in the agricultural sector, as it will provide valuable data on market trends and demands. By utilizing technology to facilitate agricultural marketing, this system has the potential to significantly impact the Indian economy and the livelihoods of farmers and their families.

In conclusion, the current state of technology in the floriculture and agricultural sectors, focusing on different kinds of machine learning techniques and other advanced techniques used various predictions like demand forecasting. The system aims to bridge the technological gap in the Sri Lankan floriculture industry and provide solutions to major barriers.

1.2 Research Problem

Despite the remarkable growth of the floriculture industry in Sri Lanka, a pressing challenge it faces lies in the realm of demand forecasting. Inaccurate demand forecasting can have severe consequences, including missed sales opportunities, excessive inventory buildup, and the wastage of vital resources, such as water, fertilizers, and labor. These inefficiencies not only impact the profitability of businesses but also carry environmental and economic implications for the nation.

Addressing this challenge necessitates the implementation of a cutting-edge demand forecasting system. Such a system, driven by advanced technologies, has the potential to revolutionize industry. It would enable businesses to align their production and inventory management precisely with the ever-changing market demands, thus significantly reducing resource wastage. This technology-driven approach not only enhances sustainability but also strengthens the industry's economic viability.

Furthermore, by addressing the issue of inaccurate demand forecasting, businesses in the floriculture industry can mitigate the risk of missed sales opportunities and the burden of excessive inventory, which often results from an inadequate understanding of market needs.

In conclusion, the flourishing floriculture industry in Sri Lanka can continue its growth and global expansion by embracing advanced technology-based solutions that focus on precise demand forecasting. Such a system will enable businesses to optimize their operations, reduce wastage, and better meet market demands, ultimately fostering sustainability and economic success within the industry.

According to the survey (Figure 1-1) that has been conducted on this research, the proposed system will be revolutionary for the floriculture industry. It appears that in the floriculture industry in Sri Lanka, most professionals (61.9%) rely on their field experience to forecast demand for the future. On the other hand, it is important to remember that relying solely on one's own experiences does not always yield the most reliable results. On the other hand, 33.3% of professionals use past data to forecast demand, indicating that data-driven decision-making is also an important approach in the floriculture industry. By understanding the methods used by professionals in the industry to forecast demand, manually forecasting the demand is not entirely accurate, which will lead to several problems.

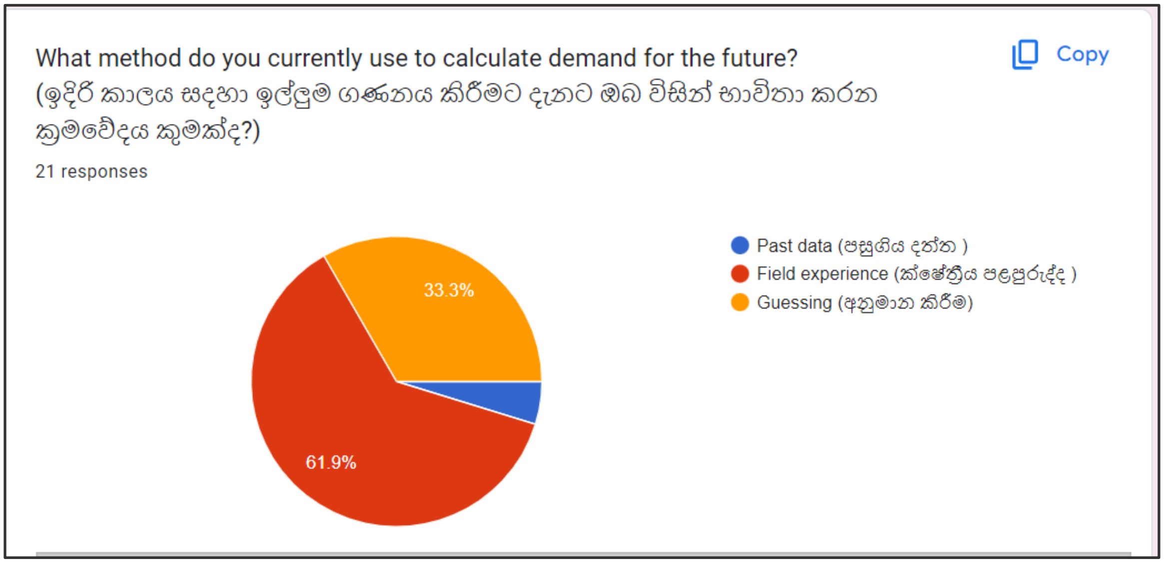


Figure 1-1 Survey to find the method that currently used to calculate the demand.

The question asks respondents whether they believe there is a need for a system to accurately calculate the demand in the floriculture industry in Sri Lanka. The responses indicate that the majority of respondents believe that such a system is necessary. This suggests that there may be challenges or limitations to the current methods used to forecast demand and that a more systematic and data-driven approach could be beneficial for businesses in the floriculture industry. Developing a system to accurately calculate demand could help businesses make more informed decisions about production, pricing, and marketing and could ultimately improve the sustainability and profitability of the industry. (Figure1-2)

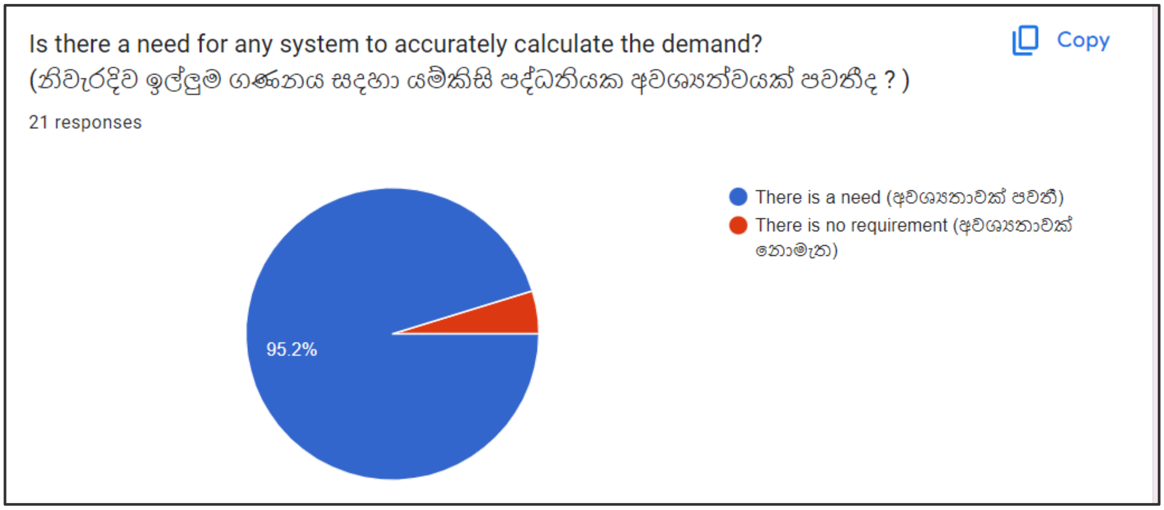


Figure 1-2 Survey to identify whether the system to calculate the demand accurately is needed or not

The findings reveal that demand forecasting poses significant challenges within Sri Lanka's floriculture industry. A noteworthy observation is that a significant proportion of respondents heavily depend on their personal experience for predicting demand, underscoring the necessity for adopting more structured and data-centric methodologies. The implementation of automated tools for demand forecasting emerges as a viable solution, promising to enhance decision-making processes and curtail unnecessary resource wastage and expenses.

Moreover, the industry stands to gain substantial advantages by identifying dependable data-driven applications aimed at elevating sustainability and bolstering competitiveness. By addressing these pressing issues, the floriculture sector can not only enhance its sustainability and competitive edge but also make meaningful contributions to the broader realms of business growth and development.

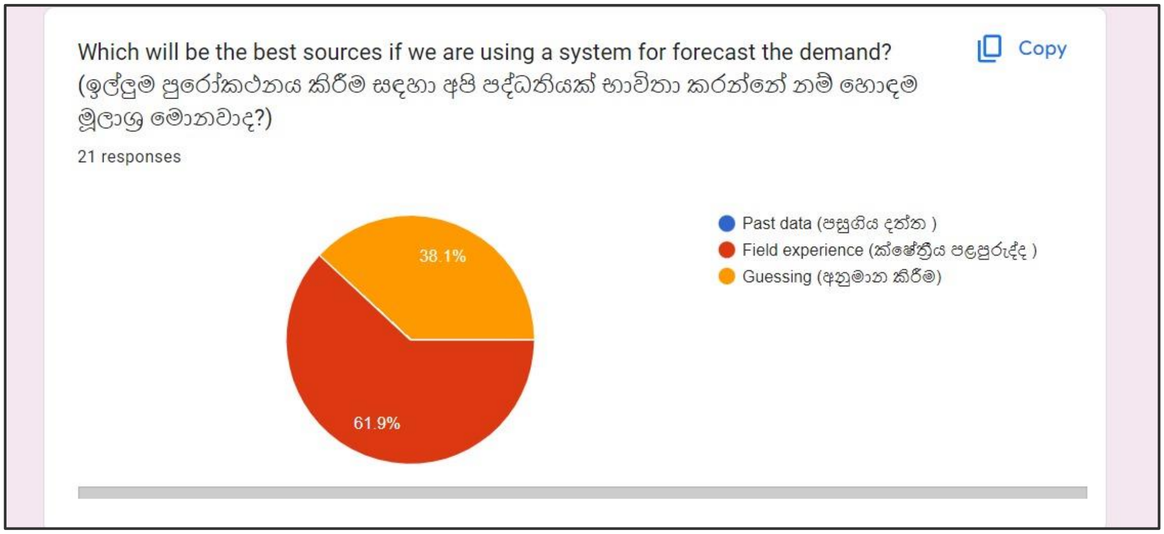


Figure 1-3: Survey to identify what are the best sources for forecast the demand.

The survey question seeks to gauge respondents' perceptions regarding the necessity of implementing an advanced demand forecasting system within the floriculture industry in Sri Lanka. The responses aim to reveal that a significant portion of respondents recognize the imperative nature of such a system. This revelation implies that the current methodologies employed for demand prediction may encounter limitations or inadequacies. The implementation of a comprehensive, data-driven demand forecasting system is expected to offer substantial benefits to businesses operating in the floriculture sector. By integrating systematic and data-centric forecasting, businesses can enhance their decision-making processes, encompassing production planning, pricing strategies, and marketing endeavors. Ultimately, the adoption of an advanced demand forecasting system holds the potential to bolster the industry's sustainability, profitability, and overall competitive edge.

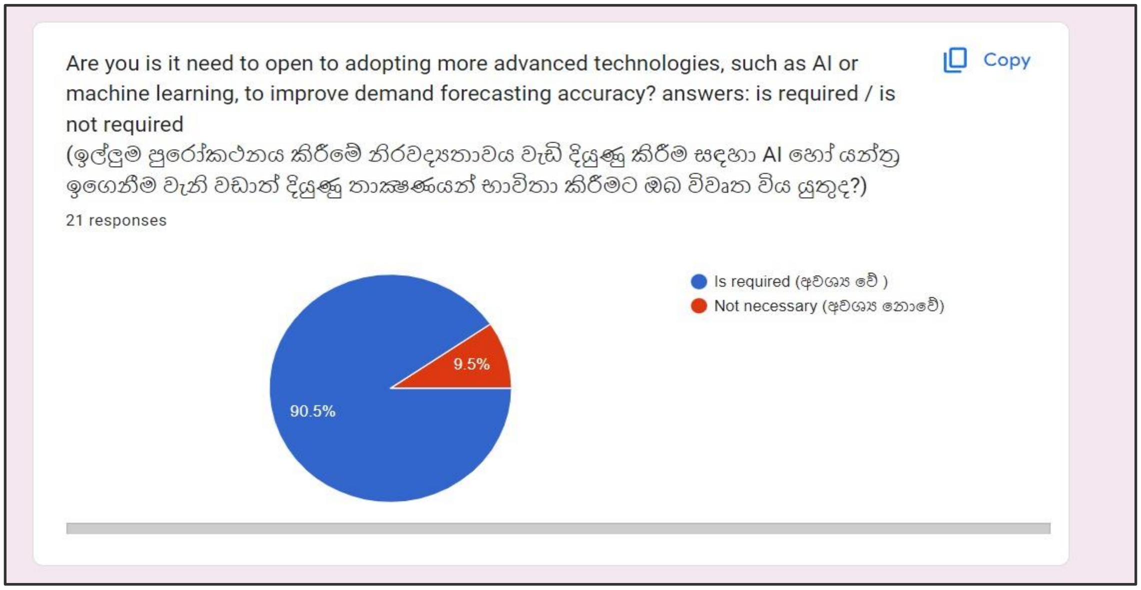


Figure 1-4: Survey question for the need of AI and machine learning for demand forecasting

The survey question seeks to ascertain respondents' openness to embracing cutting-edge technologies, such as Artificial Intelligence (AI) and Machine Learning, as means to elevate the precision of demand forecasting within the floriculture industry in Sri Lanka. The responses aim to uncover whether stakeholders within the industry are inclined towards embracing these advanced technologies. This inquiry stems from the recognition that prevailing demand forecasting methodologies may grapple with inherent limitations and imperfections. The introduction of AI and Machine Learning into the demand forecasting landscape is envisioned as a transformative step that can yield significant advantages for businesses entrenched in the floriculture sector.

By harnessing the power of these technologies, companies can usher in a new era of data-driven forecasting, thereby fortifying their decision-making frameworks across vital domains such as production planning, pricing strategies, and marketing initiatives. AI and Machine Learning hold the potential to unravel intricate demand patterns, factor in nuanced variables, and adapt dynamically to shifting market dynamics. This, in turn, promises to enhance the sustainability and profitability of the industry, endowing it with a heightened competitive edge in a rapidly evolving market landscape.

1.3 Research Gap

The following will show a certain gap between the proposed system and the currently used in already done research related to the area. Most of this research has been done in agriculture, not floriculture. Additionally, none of these will be discovered in Sri Lanka. Most of these technological advancements are novel to our country, as many of them were accomplished in India. According to SIFAA is a solution that can help the industry overcome these challenges. It is a sophisticated technology capable of diagnosing diseases affecting flowers, recommending treatments, and answering questions about floriculture. This system can efficiently provide reliable and intelligent information to farmers and can rapidly address complex issues. Additionally, it uses deep learning suits to aid ornamental plant disease identification through smartphones. To help customers find products that match their interests, a new recommendation system based on RL is proposed. It uses advanced demand forecast models to increase customer satisfaction and time wise fulfilment of user expectations. They have also done demand forecasting for the market, but they have used different kinds of technologies to implement it. [2]

This research is done in Sri Lanka, yet it is about sugar demand forecasting even though there is a slightly different in the technologies that we are going to use for forecasting the demand, but any other component will not be done in this research. [3]

This is also done in Sri Lanka, which has been done in the agricultural sector to predict both demand and supply for vegetables. [4]

In India, supply forecasting has been conducted by forecasting the harvests using a variety of technological approaches. [5]

All these were only found in agricultural products in the areas of demand and supply prediction category but didn't cover the other parts of this proposed system.
As for GPS technology that I am going to use in my system, it has never been done in the floriculture industry, but it has been done in the agriculture sector. [6]

This research is used to take the time series analysis for predicting the demand for the agriculture sector in India. [7]

In addition, this research is identical to the one that came before it; the only difference is that this one uses a mobile application and its primary focus is on ensuring customer satisfaction as the project's end objective while attempting to predict demand.. [8]

The location of these research and investigations was India, and all are get researched on the agriculture sector, yet they have used different kinds of technologies like IOT, machine learning Decision trees, likewise. [9] [10] [11]

As for this last research, they also got the GIS technology, yet it is also used for the agriculture sector. [12] In the larger landscape of research endeavors, it becomes evident that a considerable portion of scholarly work and technological advancements has primarily revolved around the agricultural sector, with a significant focus on India as a key player in this domain. This geographical concentration implies that many of the innovative technologies, methodologies, and concepts generated through these research efforts have the potential to be groundbreaking and uncharted territory when introduced in Sri Lanka.
What's especially noteworthy is the unique position of the floriculture industry within this context. As an industry that deals specifically with the cultivation and trade of ornamental plants and flowers, it stands apart from traditional agriculture in several ways. Therefore, the infusion of these technologies and ideas from the broader agricultural sector into the floriculture industry of Sri Lanka represents a truly pioneering and transformative endeavor.

This transition brings with it the excitement of exploring uncharted waters, where established practices may need adaptation or even a complete overhaul to align with the distinctive demands and intricacies of floriculture. It not only opens doors to innovation but also fosters a dynamic environment where new approaches and solutions can flourish, ultimately contributing to the growth and evolution of the floriculture industry in Sri Lanka.

Table 1-1 is showing the comparison of the past studies with the proposed system.

Table 1-1 Past research

Features / **Reference**	Demand forecasting according to different factors	Mobile Application
[2]	✓	✓
[3]	✓	-
[4]	✓	✓
[5]	-	-
[6]	-	✓
[7]	✓	-
[8]	✓	✓
[9]	✓	-
[10]	-	-
[11]	-	✓
[12]	-	-
Proposed System	✓	✓

1.4 Research Objectives

1.4.1 Main Objective

The primary objective of demand forecasting is to harness the power of predictive analysis to anticipate future demand precisely and reliably for specific products or services. This forward-looking approach empowers organizations to strategically plan their operations, allocate resources judiciously, and align production levels with market requirements. By doing so, businesses can minimize wastage, optimize inventory management, and ensure that they meet the ever-changing needs and preferences of their customer base with a high degree of accuracy. In essence, demand forecasting serves as a critical tool for enhancing operational efficiency, reducing costs, and ultimately achieving a competitive edge in today's dynamic marketplace.

1.4.2 Specific Objectives

1. Forecasting Local and International Demand: Develop an advanced demand forecasting system capable of predicting the demand for floriculture products in both local and international markets. This system should consider a multitude of factors, such as market trends, regional preferences, and economic indicators, to provide accurate and actionable insights for decision-making.

2. Revolutionizing Order Fulfilment: Transform the floriculture industry by implementing a cutting-edge order fulfillment system driven by predictive analytics. This system aims to optimize inventory management, reduce order processing times, and enhance customer satisfaction by ensuring the timely delivery of orders based on forecasted demand.

3. Seasonal and Species-Specific Demand: Implement a demand forecasting approach that distinguishes between seasonal variations and species-specific demand patterns. This will involve developing specialized models and algorithms to predict demand fluctuations accurately, allowing for tailored production and inventory strategies.

4. Mobile Application Development: Create a user-friendly mobile application that integrates with the demand forecasting system. This application should enable stakeholders, including growers, suppliers, and distributors, to access real-time demand forecasts, order information, and logistics data, facilitating seamless communication and decision-making within the floriculture supply chain.

These specific objectives aim to address various aspects of demand forecasting, ranging from market considerations to technological solutions, with the goal of enhancing the efficiency and competitiveness of the floriculture industry.

1.5 Project Requirements

1.5.1 Functional requirements

- The system should be able to obtain sales data for demand forecasting.
- The system should be able to obtain infrastructure data from the user.
- The system should be able to predict the demand forecasting for species wise.
- The system should be able to validate the user input data as needed.
- The system should be able to give the demand as a output specifically.

1.5.2 Non-functional requirements

- Usability and User Experience:
  - Intuitive User Interface: Ensure an intuitive, user-friendly interface.
  - Accessibility: Comply with accessibility standards for users with disabilities.
  - Response Time: Maintain responsive interactions.
  - Error Handling: Implement clear error messages for guidance.

- Scalability and Performance:
  - Scalability: Scale to handle increased data and user loads.
  - Performance: Ensure fast response times and minimal downtime.
  - Load Testing: Validate the system's ability to handle high user loads.

- Data Security and Privacy:
  - Data Encryption: Implement data encryption for transmission and storage.
  - Access Control: Enforce role-based access control.
  - Data Privacy Compliance: Comply with relevant data privacy regulations.
  - Data Backup and Recovery: Establish data backup and recovery procedures.
  - Audit Trails: Maintain detailed audit logs for user activity monitoring.

1.5.3 System requirements

- Google Colab for Machine Learning Development: Google Colab is a cloud-based platform that allows machine learning practitioners to develop and run machine learning models in a collaborative and cloud-based environment. It provides access to GPU and TPU resources, making it ideal for training complex models.

- Scikit-Learn, Pandas, Numpy, Statsmodels Libraries for Model Development: These Python libraries are commonly used in data science and machine learning. Scikit-Learn is used for building machine learning models, Pandas for data manipulation and analysis, Numpy for numerical computations, and Statsmodels for statistical modeling and analysis.

- PyCharm as an IDE and Flask as a Framework for Backend API Development: PyCharm is a popular integrated development environment (IDE) for Python programming. Flask is a lightweight Python web framework used for building web APIs and web applications. Together, they enable efficient and organized backend API development.

- Postman for API Testing: Postman is a widely used tool for testing and documenting APIs. It allows developers to send requests to APIs, inspect responses, and automate testing workflows, ensuring the reliability and functionality of the APIs.

- Visual Studio Code as an IDE and React Native as a Framework for Front-End Development: Visual Studio Code (VS Code) is a versatile and widely adopted code editor. React Native is a JavaScript framework for building cross-platform mobile applications. Using VS Code and React Native, developers can create mobile apps that work seamlessly on both Android and iOS platforms.

- EXPO for App Development: EXPO is a set of tools and libraries that simplifies the development of React Native mobile applications. It provides pre-configured development environments, making it easier for developers to build, test, and deploy their apps, especially for those new to mobile app development.

These tools and technologies collectively provide a comprehensive development environment for machine learning, web backend, API testing, and mobile app development, catering to different aspects of your project's requirements.

1.5.4 User requirements

- Users can manually input the plant type.
- Users can input temperature, humidity, and rainfall data.
- Users can specify the pot size.

2 METHODOLOGY

Figure 2-1 describes the overall architecture for the ‘Plant Pal’ application with the demand forecasting. The 'Plant Pal' application employs a comprehensive architecture that seamlessly integrates various components to provide a robust and user-friendly platform for plant enthusiasts. At its core, the architecture revolves around an adaptive demand forecasting system that caters to the unique needs of each user.

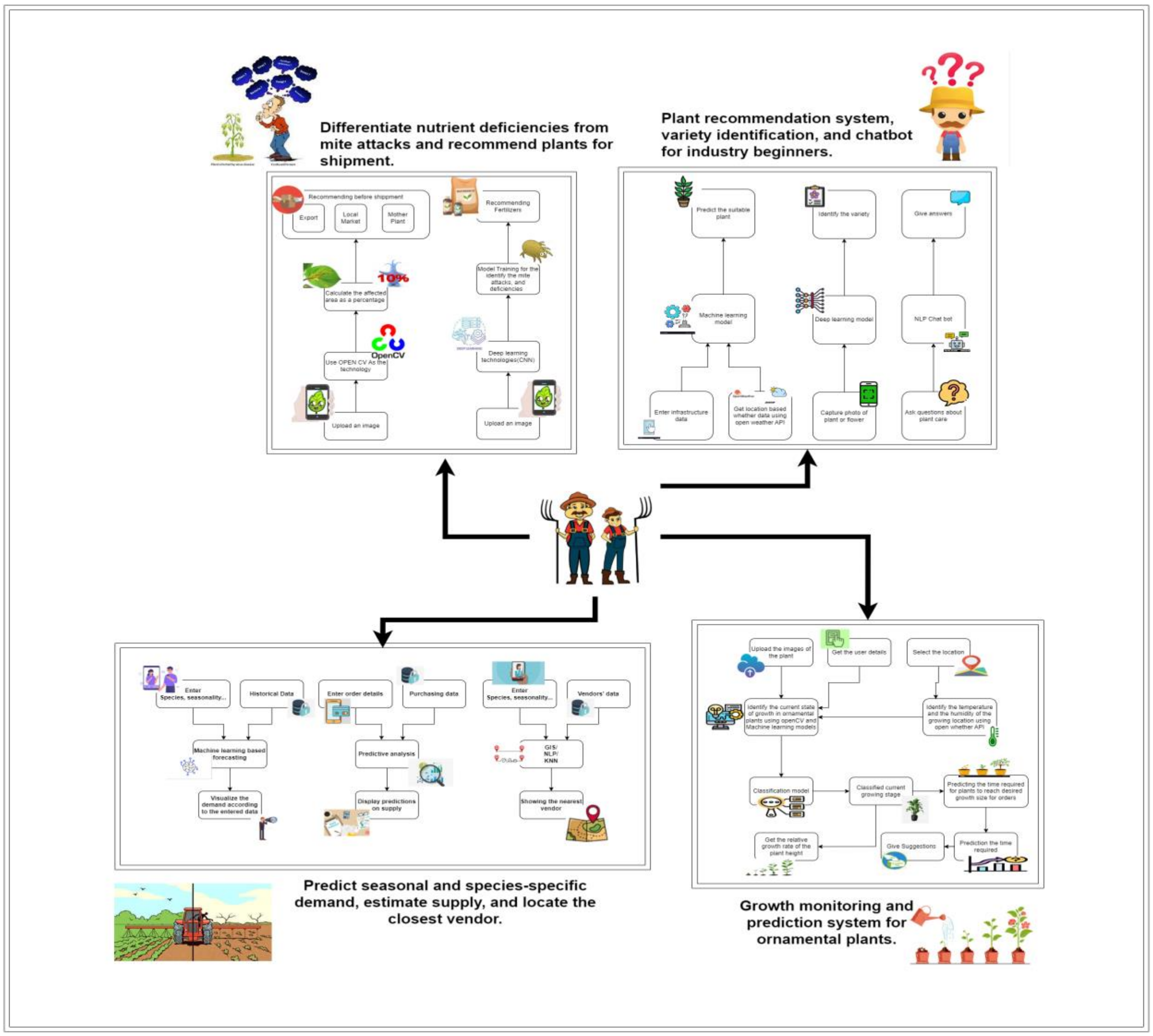


Figure 2-1 Overall system diagram

In Figure 2-2 it explains the architecture of the individual component which contains demand prediction.

The demand prediction component within the 'Plant Pal' application is the analytical engine responsible for leveraging machine learning algorithms and real-time environmental data to forecast the specific care requirements of each user's plant collection. It takes into account factors such as plant type, environmental conditions, and user inputs to generate accurate and personalized demand forecasts.

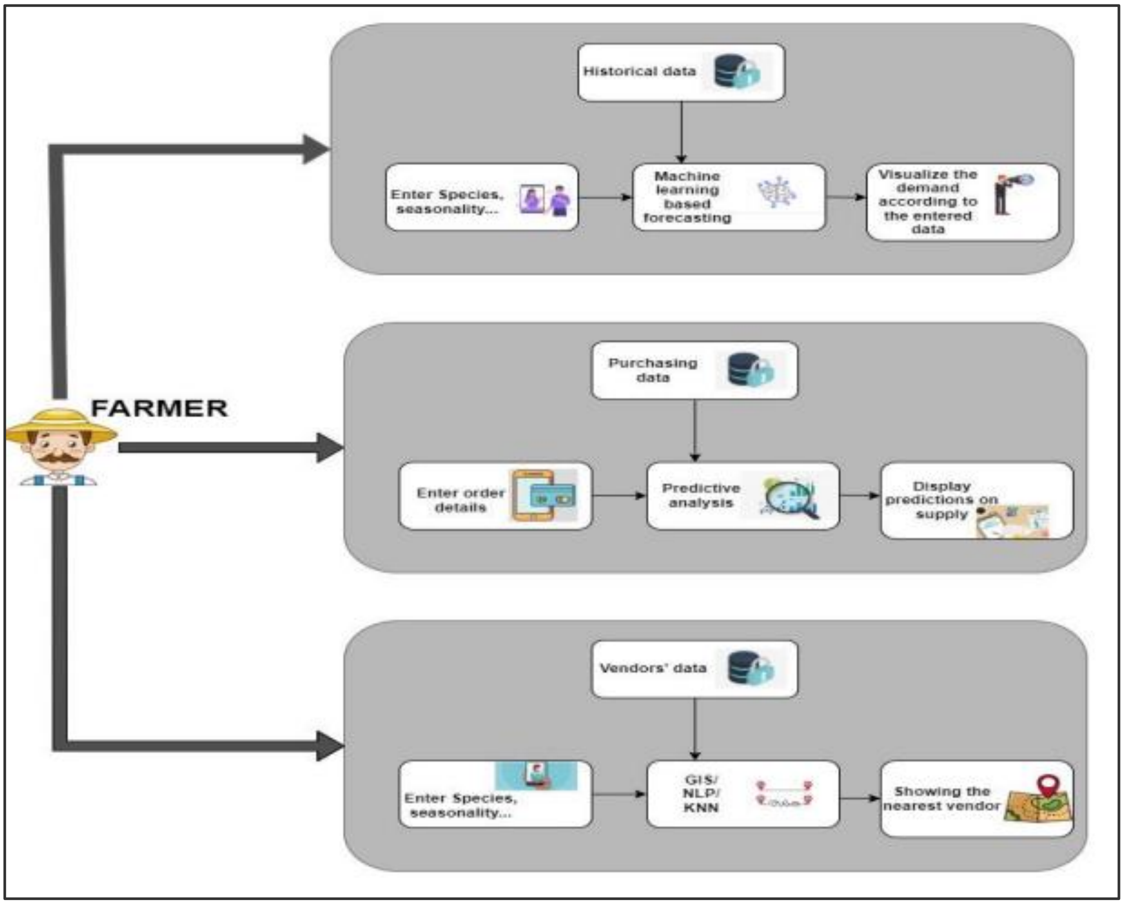


Figure 2-2 Individual system diagram

And suggested method will be done in the phrase of the agile methodology. Figure 2-3 will descriptively show the agile methodology.

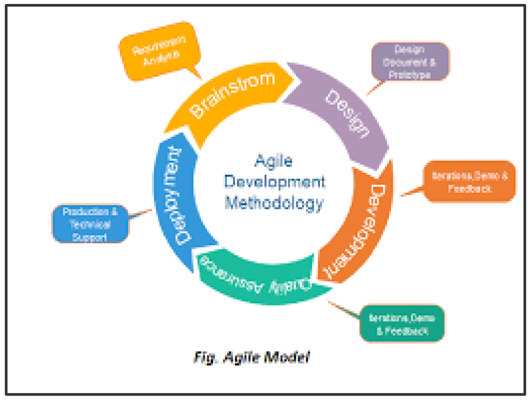


Figure 2-3: Agile methodology

Agile is a flexible and iterative approach to project management and software development that prioritizes collaboration, customer feedback, and adaptability. It encourages teams to break down projects into smaller, manageable tasks and continuously deliver valuable increments of work, allowing for adjustments based on changing requirements and customer needs. Agile emphasizes close communication between team members and stakeholders, fostering a dynamic and customer-centric development process.

Applying Agile to Demand Forecasting: When applying Agile principles to demand forecasting, the process becomes more adaptive, and customer focused.

Table 2-1 This will be the summarization of the technologies that are going to be used on the proposed system.

Table 2-1 Summary of tools and technologies used.

Technologies	**Algorithms & architectures**	**Techniques**
React Native Python Flask	Time series analysis Regression analysis Gradient Boosting Random Forest Linear Regression KNN SVM	Data preprocessing Recommendation system Feature engineering Location Based services Predictive analysis

2.1 Commercialization & Business Plan

Plant Pal is a versatile smartphone application tailored for the floriculture sector, facilitating precise demand forecasting for various plant types, catering to the needs of business owners. Compatible with both iOS and Android platforms, it not only delivers a seamless user experience but also serves as a flexible backend API. Plant Pal's strategic approach encompasses subscription models, strategic partnerships, and data monetization, all while committing to continuous development. With a core focus on assisting interns and novices in floriculture and fostering entrepreneurial spirit, Plant Pal aims to carve its niche in the market, evolving into an indispensable tool for enthusiasts. Its success lies in delivering invaluable insights and fostering user loyalty, ensuring sustainable and profitable growth in the sector.

2.2 Implementation

2.2.1 Demand Prediction Implementation

- **Data Collection:**

Step 1: Identification of Relevant Data

In the first step of this research, we will identify and select the crucial data components necessary for demand forecasting. These components encompass historical sales data, order details, geographical distribution, seasonal variations, market trends, and customer behavior. This process is essential to ensure that we have a comprehensive data set that can facilitate accurate demand predictions.

Step 2: Determination of Time Frames and Sales Cycle Length

The second step entails defining the specific time frames for our analysis. We will decide whether we should analyze data on a monthly, quarterly, or annual basis. Additionally, we will determine the length of the sales cycle by considering factors such as order lead times, production cycles, and the duration of order fulfillment.

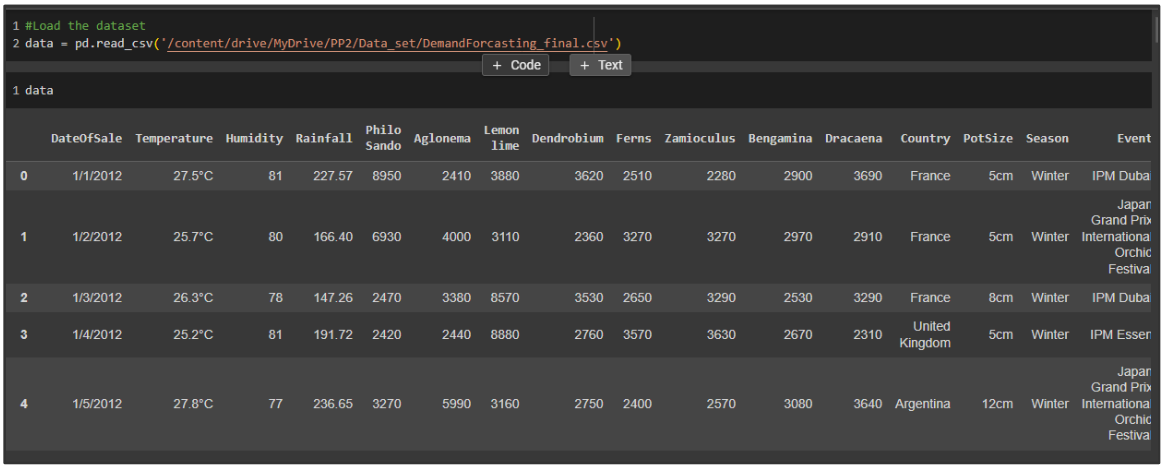
Step 3: Addressing Seasonality

Seasonality is a critical factor in the floriculture industry, influencing demand patterns significantly. This step involves recognizing and accommodating seasonal variations. We will identify peak and off-peak seasons, understand their impact on product demand, and adapt our forecasting models accordingly to ensure precise predictions.

Step 4: Data Collection and Organization

The core of our research lies in the systematic collection and organization of data. During this step, we will collect the identified data elements and ensure that it is clean, validated, and structured in an organized manner. This well-organized data set will serve as the fundamental input for developing our demand forecasting models.

These four steps provide a clear and structured approach to collecting and preparing the data necessary for effective demand forecasting within the floriculture industry. The successful implementation of this data collection and analysis framework will enable businesses in the sector to make informed decisions, optimize resource allocation, and meet market demands accurately. Figure 2-4 will show the data set collected after considering these points.



```
1 #Load the dataset
2 data = pd.read_csv('/content/drive/MyDrive/PP2/Data_set/DemandForcasting_final.csv')
```

+ Code + Text

```
1 data
```

	DateOfSale	Temperature	Humidity	Rainfall	Philo Sando	Aglonema	Lemon lime	Dendrobium	Ferns	Zamioculus	Bengamina	Dracaena	Country	PotSize	Season	Event
0	1/1/2012	27.5°C	81	227.57	8950	2410	3880	3620	2510	2280	2900	3690	France	5cm	Winter	IPM Dubai
1	1/2/2012	25.7°C	80	166.40	6930	4000	3110	2360	3270	3270	2970	2910	France	5cm	Winter	Japan Grand Prix International Orchid Festival
2	1/3/2012	26.3°C	78	147.26	2470	3380	8570	3530	2650	3290	2530	3290	France	8cm	Winter	IPM Dubai
3	1/4/2012	25.2°C	81	191.72	2420	2440	8880	2760	3570	3630	2670	2310	United Kingdom	5cm	Winter	IPM Essen
4	1/5/2012	27.8°C	77	236.65	3270	5990	3160	2750	2400	2570	3080	3640	Argentina	12cm	Winter	Japan Grand Prix International Orchid Festival

Figure 2-4: Data set collection

- **Data Preprocessing and cleaning:**

Step 1: Missing Data Handling

In this initial step, we will rigorously assess the dataset for missing values. Missing data can adversely affect the accuracy of demand forecasting models. We will employ appropriate techniques, such as imputation or removal, depending on the nature and extent of the missing data. The choice of strategy will be data-driven and tailored to the specific dataset under examination.

Figure 2-5 will show how the missing values are handled in the code.

```
# Handle missing values
# Check for missing values in each column
data.isnull().sum() #no missing values so far
```

```
DateOfSale          0
Temperature         0
Humidity            0
Rainfall            0
Philo Sando         0
Aglonema            0
Lemon lime          0
Dendrobium          0
Ferns               0
Zamioculus          0
Bengamina           0
Dracaena            0
Country             0
PotSize             0
Season              0
Event               0
HighestSalesPlant   0
dtype: int64
```

Figure 2-5: Dataset after handling the missing values.

Step 2: Outlier Detection and Treatment

Identifying outliers is crucial as they can skew demand forecasting results. We will employ statistical methods and visualization techniques to detect outliers within the dataset. Once identified, we will decide on an appropriate strategy for handling outliers, such as transformation, imputation, or removal. The choice will be contingent upon the dataset's characteristics and the impact of outliers on the forecasting models.

Figure 2-6, Figure 2-7, Figure 2-8 show how the outliers handled in the model building by removing them. To remove these outliers' different methods have been tried to use. But after considering the different methods for removing outliers replacing them with the mean is finally choose as a best approach.

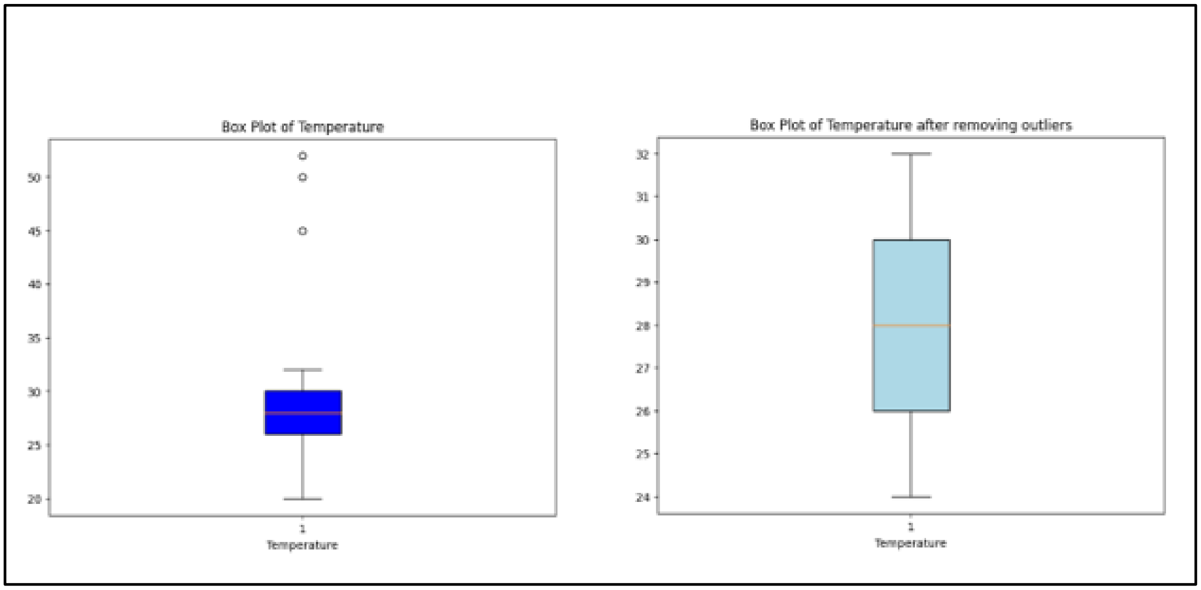


Figure 2-6: Temperature before and after outliers handling

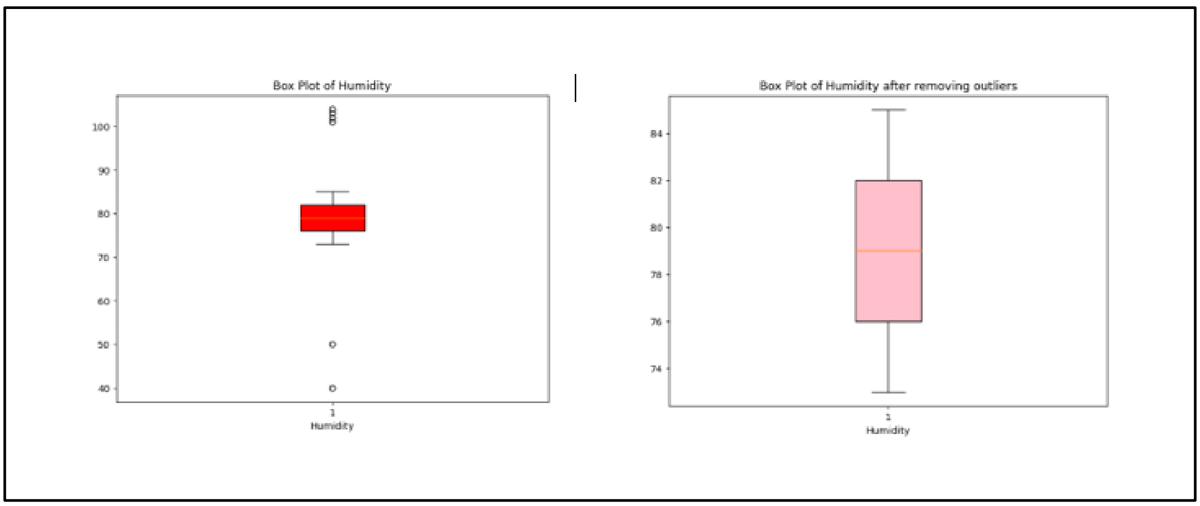


Figure 2-7: Humidity before and after outliers handling

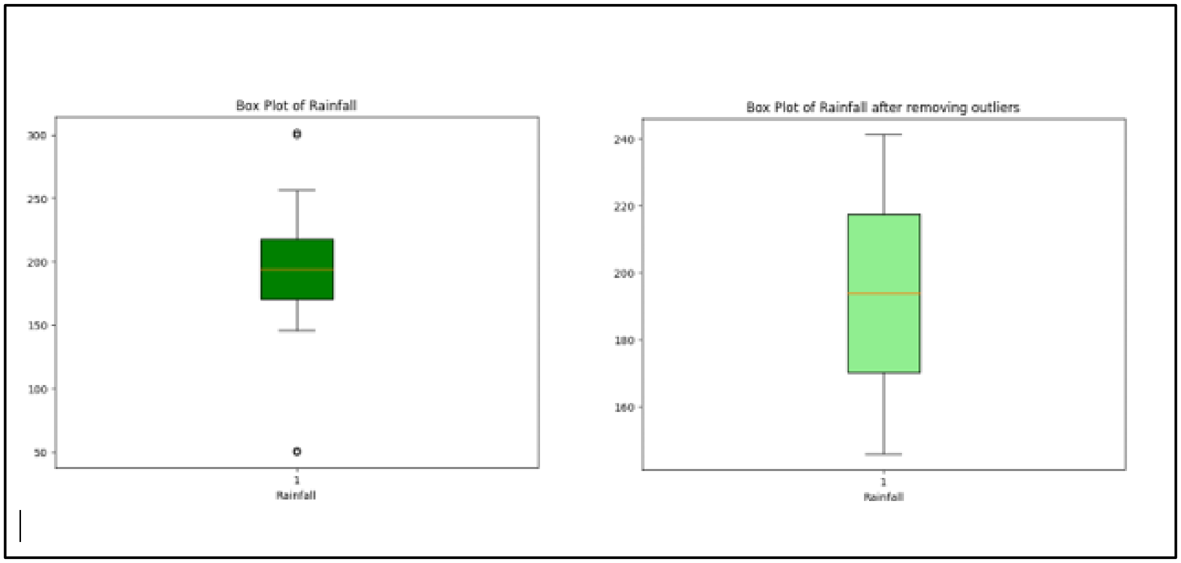


Figure 2-8: Rainfall before and after outliers handling

Step 3: Data Quality Assessment

Ensuring data quality is paramount for robust forecasting. We will scrutinize the dataset for potential data quality issues, including duplicate entries, inconsistent formatting, and data entry errors. Any identified issues will be meticulously addressed to enhance data integrity.

Step 4: Data Transformation

To facilitate effective analysis, we will transform the data into a suitable format. This transformation may involve converting categorical variables into numerical ones, especially if the chosen forecasting model relies on numerical inputs. The specific transformation techniques will be determined based on the forecasting model to be applied and the dataset's requirements.

The data preprocessing steps outlined here are integral to the success of demand forecasting within the floriculture industry. By systematically addressing missing data, outliers, data quality issues, and data transformation, we will ensure that the dataset is ready for rigorous analysis. These preparatory steps are vital for achieving accurate and actionable demand predictions, contributing to the overall success of the research and its practical applications within the industry.

The transformed data will be done in the model building as per Figure 2-9. And additionally, one hot encoding was used as a preprocessing technique for categorical variables as in Figure 2-10 while numerical data handled by scaling using minmax scaler as in Figure 2-11.

```
# Apply transformations
transformed_data = pd.DataFrame()
```

```
for column in plants_columns_data:
    # Applying square root transformation
    transformed_data[f"sqrt_{column}"] = np.sqrt(plants_columns_data[column])

```

```
transformed_data
```

	sqrt_Philo Sando	sqrt_Aglonema	sqrt_Lemon lime	sqrt_Dendrobium	sqrt_Ferns	sqrt_Zamioculus	sqrt_Bengamina	sqrt_Dracaena
0	94.604440	49.091751	62.289646	60.166436	50.099900	47.749346	53.851648	60.745370
1	83.246622	63.245553	55.767374	48.579831	57.183914	57.183914	54.497706	53.944416
2	49.699095	58.137767	92.574294	59.413803	51.478151	57.358522	50.299105	57.358522
3	49.193496	49.396356	94.233752	52.535702	59.749477	60.249481	51.672043	48.062459
4	57.183914	77.395090	56.213877	52.440442	48.989795	50.695167	55.497748	60.332413

Figure 2-9: Transformed data according to certain format.

```
onehot_encoded_categorical_data
```

	Country_Argentina	Country_Brazil	Country_Canada	Country_France	Country_Japan	Country_Netherland	Country_USA	Country_United Kingdom	PotSize_12cm	PotSize_5cm
0	0	0	0	1	0	0	0	0	0	1
1	0	0	0	1	0	0	0	0	0	1
2	0	0	0	1	0	0	0	0	0	0
3	0	0	0	0	0	0	0	1	0	1
4	1	0	0	0	0	0	0	0	1	0

Figure 2-10: Data after using one hot encoding for the categorical variables.

```
scaled_numerical_data_minmax_df
```

	Temperature	Humidity	Rainfall	sqrt_Philo Sando	sqrt_Aglonema	sqrt_Lemon lime	sqrt_Dendrobium	sqrt_Ferns	sqrt_Zamioculus	sqrt_Bengamina
0	0.375	0.666667	0.854656	0.994738	0.087152	0.350336	0.308320	0.107256	0.060446	0.182071
1	0.125	0.583333	0.213192	0.768247	0.369398	0.220273	0.077024	0.248521	0.248782	0.194954
2	0.250	0.416667	0.012479	0.099263	0.267542	0.954254	0.293296	0.134740	0.252268	0.111228
3	0.125	0.666667	0.478712	0.089181	0.093226	0.987346	0.155993	0.299681	0.309978	0.138606
4	0.375	0.333333	0.949874	0.248521	0.651559	0.229177	0.154091	0.085119	0.119251	0.214896

Figure 2-11: Data after using minmax scaler for the numerical variables.

- **Feature selection and feature engineering:**

Step 1: Identification of Relevant Features

In the first step, we will identify a comprehensive set of features that may have an impact on demand within the floriculture industry. These features may include:

Weather Data: Information about temperature, humidity, precipitation, and other meteorological factors that can influence floral growth and customer demand.

Holidays and Events: Data on holidays, festivals, and special events, as these occasions often drive fluctuations in demand for floral products.

Market Trends: Data on broader market trends, such as economic indicators, consumer sentiment, and floral design preferences, which can affect demand patterns.

Step 2: Exploratory Data Analysis (EDA) and Correlation Analysis

EDA techniques will be employed to gain a deeper understanding of the dataset and its features. We will visualize the data to identify patterns and trends. Additionally, we will perform correlation analysis to assess the relationships between different features and the target variable (demand). This analysis will help us pinpoint the most influential features. This part is started as in Figure 2-12.

Exploratory Data Analysis (EDA)

```
[ ]  1 data_withoutOutliers
```

	DateOfSale	Temperature	Humidity	Rainfall	Country	PotSize	Season	Event	HighestSalesPlant	sqrt_Philo Sando	sqrt_Aglonema	sqrt_Lemon lime	sqrt_Dendrobium
0	1/1/2012	27.0	81.0	227.57	France	5cm	Winter	IPM Dubai	Philo Sando	94.604440	49.091751	62.289646	60.166436
1	1/2/2012	25.0	80.0	166.40	France	5cm	Winter	Japan Grand Prix International Orchid Festival	Philo Sando	83.246622	63.245553	55.767374	48.579831
2	1/3/2012	26.0	78.0	147.26	France	8cm	Winter	IPM Dubai	Lemon lime	49.699095	58.137767	92.574294	59.413803
3	1/4/2012	25.0	81.0	191.72	United Kingdom	5cm	Winter	IPM Essen	Lemon lime	49.193495	49.396356	94.233752	52.535702
4	1/5/2012	27.0	77.0	236.65	Argentina	12cm	Winter	Japan Grand Prix International Orchid	Aglonema	57.183914	77.395090	56.213877	52.440442

Figure 2-12: Exploratory Data Analysis (EDA)

When doing this part, the frequencies of each categorical variable are considered as Figure 2-13. Other than that, distributions of the numerical values (Figure 2-14) and correlation heat map (Figure 2-15) is used to select the best features for the model.

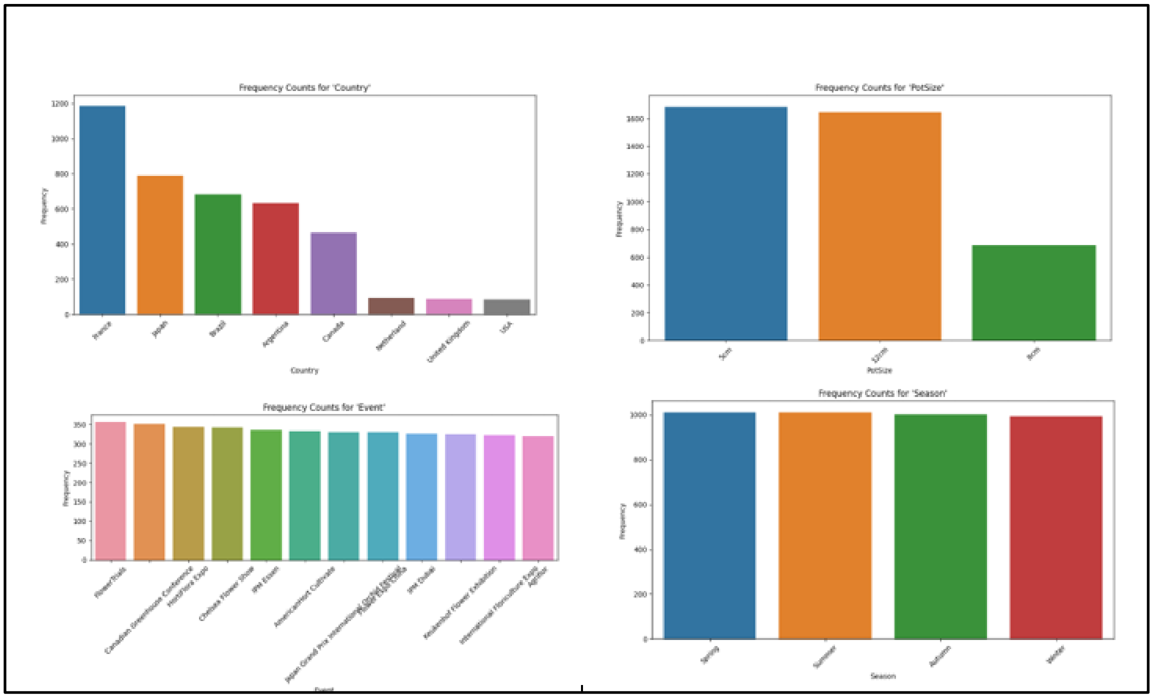


Figure 2-13: Frequencies of categorical variables.

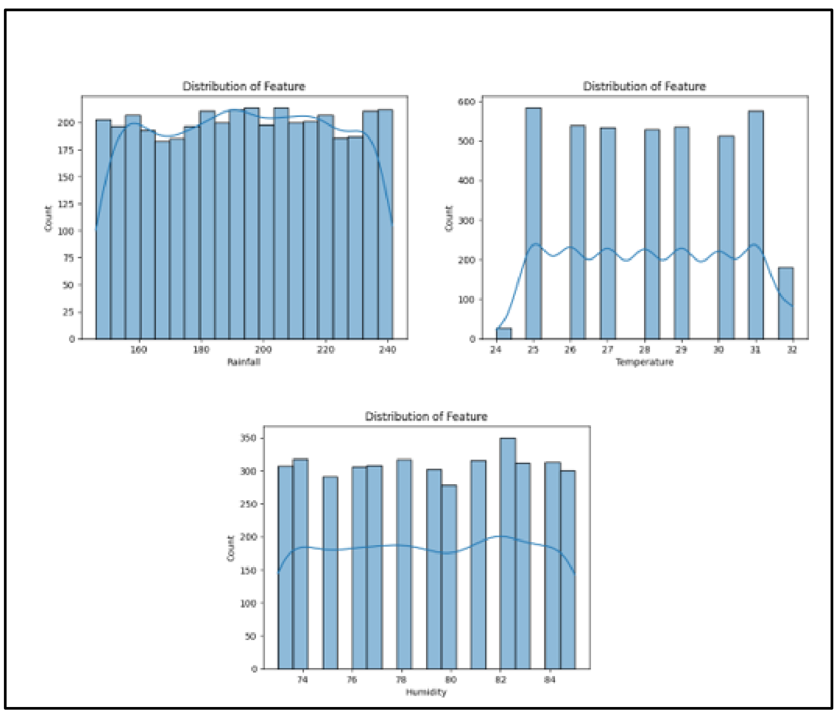


Figure 2-14: Distributions of the numerical variables.

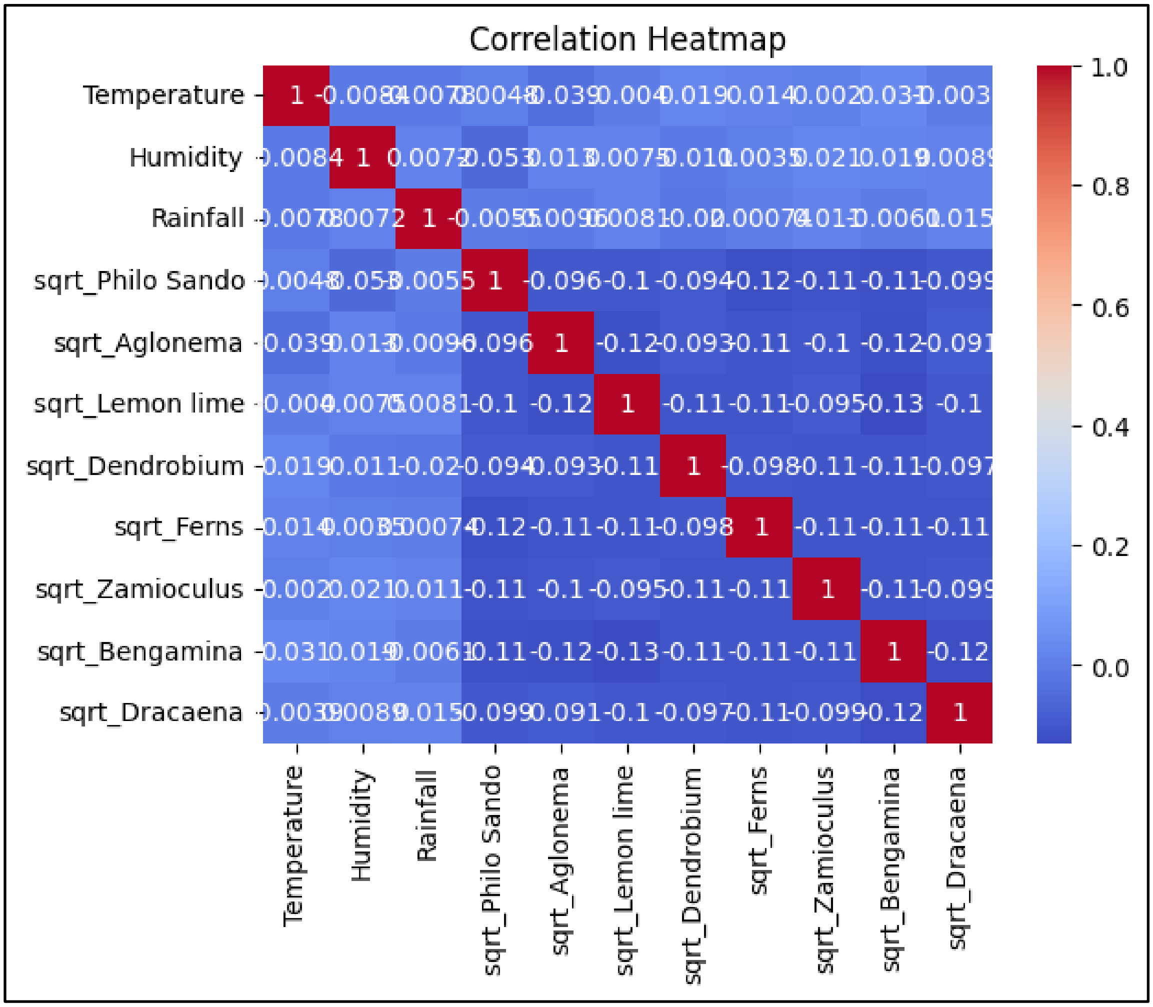


Figure 2-15: Correlation heat map

Step 3: Creation of New Features

To capture complex patterns and trends in the data, we will create new features. These may include:

Lagged Variables: By introducing lagged versions of the target variable or other relevant features, we can account for time dependencies and past trends in demand.

Rolling Averages: Calculating rolling averages or moving averages can help in smoothing out noise in the data and identifying long-term trends.

Seasonality Indicators: Creating binary or categorical indicators for different seasons or specific time periods can help model seasonal variations in demand.

Feature engineering is a crucial aspect of demand forecasting within the floriculture industry. By identifying relevant features, conducting exploratory data analysis, and creating new features, we aim to enrich the dataset and enhance the predictive power of our models. These steps are essential for capturing and modeling the intricate demand patterns and drivers in the floriculture sector, ultimately leading to more accurate and actionable forecasts. Handling Imbalance with Stratified Sampling: Finding any class imbalance in the target variable, such as an unequal distribution of demand across various products or regions. Use stratified sampling to ensure that each class is represented in training and testing datasets in proportion to their overall frequency.

- **Splitting the Dataset:**

Step 1: Dataset Separation
The initial step is to separate the dataset into distinct subsets for training, testing, and validation:
70% of the preprocessed data are chosen to train the model while the left 30% of the data set considered as the test data.

Training Set: This subset comprises most of the data and serves as the foundation for training machine learning models. It is used to teach models to recognize patterns and relationships within the data.

Testing Set: The testing set is reserved to evaluate the performance of the trained models. It provides an unbiased assessment of model accuracy and generalizability to unseen data.

Step 2: Random Sampling and Stratification
To ensure that the data subsets are representative of the overall dataset and account for any class imbalances or patterns, we employ random sampling and stratification:

Random Sampling: Randomly selecting instances from the dataset prevents any bias that may result from a specific order or sequence.

Stratification: Stratified sampling ensures that the distribution of classes or relevant features remains consistent across the training, testing, and validation sets. This is particularly important when dealing with imbalanced datasets.

Step 3: Validation Set Creation
A portion of the data is reserved as a validation set, distinct from both the training and testing sets. The validation set is used for model selection and hyper parameter tuning. It helps us assess how well different models perform and which hyper parameters yield the best results.

Data splitting and validation are essential components of the machine learning pipeline. By carefully partitioning the data into training, testing, and validation sets while incorporating random sampling and stratification, we ensure the integrity of our analysis and the reliability of the models developed. These steps enable us to build robust demand forecasting models for the floriculture industry and make informed decisions based on their performance. Applying Feature Scaling: Normalizing the data using techniques such as MinMaxScaler to ensure that the data is on a similar scale and avoid any feature dominating the model.

- **Model Building:**

Step 1: Model Selection
The first step involves selecting a range of models suitable for the task of demand prediction. The models considered may include:

Linear Regression: A simple and interpretable model that establishes a linear relationship between input features and demand.

Random Forest: An ensemble model that combines multiple decision trees to capture complex relationships in the data.

Gradient Boosting: Another ensemble method that builds a strong predictive model by sequentially adding weak learners.

MLP Regressor: A type of neural network model designed for regression tasks, capable of capturing non-linear relationships in the data.

SARIMA (Seasonal Autoregressive Integrated Moving Average): A time series model that accounts for seasonality, trends, and autocorrelation in the data.

Step 2: Model Fitting

Each selected model is trained using the training dataset. During this phase, the models learn from the historical demand data and the relevant features.

Step 3: Performance Assessment

The trained models are then assessed for their performance on the validation dataset. Key evaluation metrics, such as:

R-squared (R2): Measures the proportion of variance in the target variable that is explained by the model. A higher R2 indicates a better fit.

Mean Squared Error (MSE): Measures the average squared difference between predicted and actual demand values. Lower MSE values indicate better accuracy.

Mean Absolute Error (MAE): Measures the average absolute difference between predicted and actual demand values. Lower MAE values suggest better model accuracy.

Step 4: Model Selection

Based on the evaluation metrics of interest (e.g., R2, MSE, or MAE), the model that performs best on the validation dataset is selected as the most suitable for demand prediction.

The model selection and evaluation process is essential for determining the most effective model for demand prediction in the floriculture industry. By fitting a range of models and assessing their performance on the validation dataset using appropriate evaluation metrics, we can identify the model that best captures the relationships in the data and provides accurate demand forecasts. This chosen model will form the basis for making informed decisions and optimizing operations within the floriculture sector.

- **Model Accuracy and Selecting the Best Model:**

Step 1: Final Model Evaluation

Utilize the testing dataset to rigorously evaluate the final model's performance. This dataset represents unseen data, allowing you to assess how well the model generalizes to new observations.

Employ appropriate evaluation metrics, including R-squared (R2), Mean Squared Error (MSE), and Mean Absolute Error (MAE), to quantify the model's predictive accuracy.

Compare the final model's performance against a baseline model. This baseline model serves as a benchmark for assessing the improvement achieved by your chosen model.

Step 2: Model Performance Assessment

Examine the evaluation results carefully. If the final model's performance falls short of expectations or predefined criteria, this step indicates the need for further refinement.

Step 3: Iteration and Improvement

If the model's performance is subpar, consider returning to earlier stages in the process, such as feature engineering or model selection. Iteration may involve enhancing feature engineering techniques, refining hyperparameters, or exploring different algorithms.

Continue to iterate until a satisfactory result is achieved, where the model's performance aligns with your desired accuracy levels and meets the objectives of demand prediction in the floriculture industry. As in the figure 2-16 the best model is chosen relevant to the specific plant wise.

```
Best models for each target column:
sqrt_Philo Sando    Gradient Boosting
sqrt_Aglonema       Linear Regression
sqrt_Lemon lime     Linear Regression
sqrt_Dendrobium     Linear Regression
sqrt_Ferns          Linear Regression
sqrt_Zamioculus     Linear Regression
sqrt_Bengamina         Neural Network
sqrt_Dracaena       Linear Regression
dtype: object

1 # the best approach is Linear regression
```

Figure 2-16: Best model according to the specific plant type

Final Model Choice: Linear Regression

As per your plan, if the best-performing model aligns with linear regression and meets the desired performance criteria, it's a suitable choice for demand prediction within the floriculture industry. Linear regression offers transparency, interpretability, and simplicity, making it an effective option for modeling relationships between variables, especially when other complex models do not provide significant advantages.

The structured approach to model evaluation and iteration ensures that you are committed to achieving the most accurate demand forecasting results while maintaining the flexibility to adapt and refine your methods as needed. This methodology reflects a robust and data-driven approach to solving complex problems in the floriculture sector.

2.2.2 Backend – Flask API

In the development of our backend infrastructure, we have harnessed the capabilities of Flask, a lightweight yet highly versatile Python web framework, to construct a robust and flexible API (Application Programming Interface) that serves as the central nervous system of our application. One of the most intriguing facets of our API's functionality lies in its seamless integration with a .pkl file—a "pickle file" in Python parlance—that holds a treasure trove of serialized Python objects. Among these objects, we house trained machine learning models that constitute the heart and soul of our data analysis prowess.

This marriage of Flask and the .pkl file bears testament to the intersection of web development and machine learning, where the backend infrastructure transforms into an intelligent powerhouse. By adeptly loading this .pkl file within our Flask application, we orchestrate the harmonious coexistence of predictive models with the real-time interactivity of our platform. The result? Our application becomes a veritable oracle, offering users on-the-fly predictions and sage recommendations that guide their decision-making with remarkable precision.

This architectural choice does more than streamline the deployment of machine learning models; it imbues our application with the qualities of scalability and responsiveness, ensuring that it can accommodate increasing user demands while maintaining rapid response times. As users interact with our platform, they access data-driven insights and forecasts that empower them to make informed choices, be it in agricultural decision-making, financial planning, or any other domain that benefits from predictive analytics. In essence, our Flask-based API, with its seamless .pkl file integration, represents a harmonious synergy of technology and data science, providing an invaluable tool for those who seek data-driven enlightenment.

2.2.3 Frontend – React Native

Pros of Using React Native:

1. Cross-Platform Development: React Native is renowned for its ability to facilitate cross-platform development. This means you can write code once and deploy it on both iOS and Android platforms, significantly reducing development time and effort.

2. Reusable Components: React Native allows developers to create reusable UI components. This component-based architecture enhances code modularity, simplifies maintenance, and accelerates development.

3. Native-Like Performance: React Native bridges the gap between native and web app performance. By rendering components using native modules, it ensures that the app runs smoothly and provides a near-native experience.

4. Large Community and Ecosystem: React Native enjoys a vast and active community of developers. This results in an extensive ecosystem of libraries, plugins, and resources, making it easier to find solutions to common problems.

5. Hot Reloading: Developers can make real-time code changes and see the results instantly without the need for full app recompilation. This feature speeds up debugging and fine-tuning.

6. Cost-Effective: Building a single codebase for multiple platforms reduces development costs, making React Native a cost-effective choice for businesses.

Cons of Using React Native:

1. Limited Native Functionality: While React Native provides access to native modules, there may still be instances where specific native features or functionalities are not readily available. Custom native modules may be required.

2. Performance Variability: Although React Native offers excellent performance, certain resource-intensive tasks or complex animations may not perform as well as fully native apps. Careful optimization may be needed for such scenarios.

3. Learning Curve: Developers with no prior experience in React (or JavaScript) may face a learning curve when adopting React Native. It may take time to become proficient with React and its associated concepts.

4. Version Compatibility: Maintaining compatibility with different versions of React Native, libraries, and dependencies can sometimes be challenging, especially as the ecosystem evolves.

5. Platform-Specific Code: Despite cross-platform capabilities, there may still be instances where platform-specific code is necessary to achieve certain functionalities or design elements.

6. Less Control Over Updates: Unlike fully native apps, React Native apps are somewhat dependent on Facebook's update cycles. Changes in React Native updates may necessitate adjustments in your app.

In conclusion, React Native is a powerful framework that offers significant advantages, particularly in terms of cross-platform development, reusability, and a large community. However, it's essential to weigh these benefits against potential challenges such as performance considerations and platform-specific requirements. Ultimately, the choice of React Native or any other framework should align with your project's specific goals and constraints.

In this front-end part Figure 2-17 shows the 'Plant Pal' main page which is automatically navigating to Figure 2-18 which can be select the needed feature. Figure 2-19 will show that the page which is navigating after selecting the Demand Forecasting. Finally, Figure 2-20 will show the form which will take the user inputs.

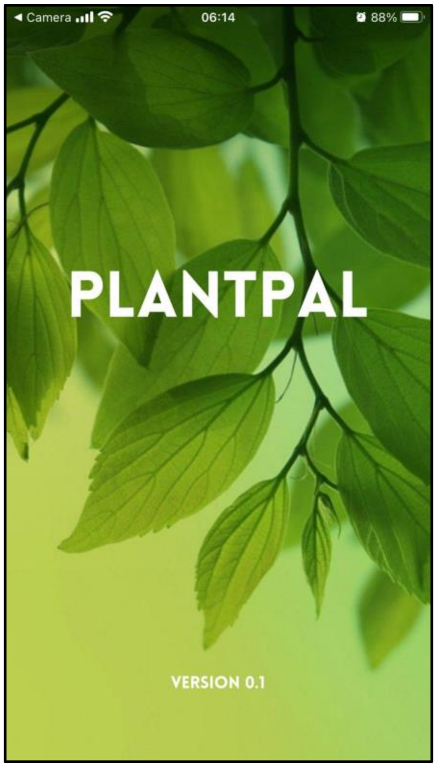


Figure 2-17: ‘Plant Pal’ main page

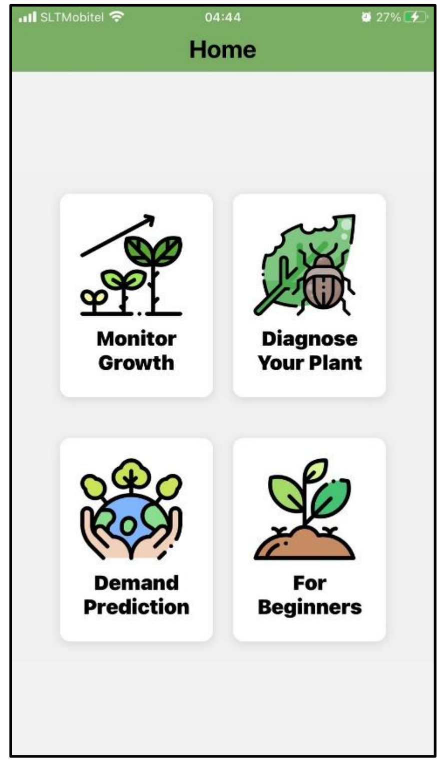


Figure 2-18: ‘Plant Pal’ main selective page

Figure 2-19: Demand forecasting selective page

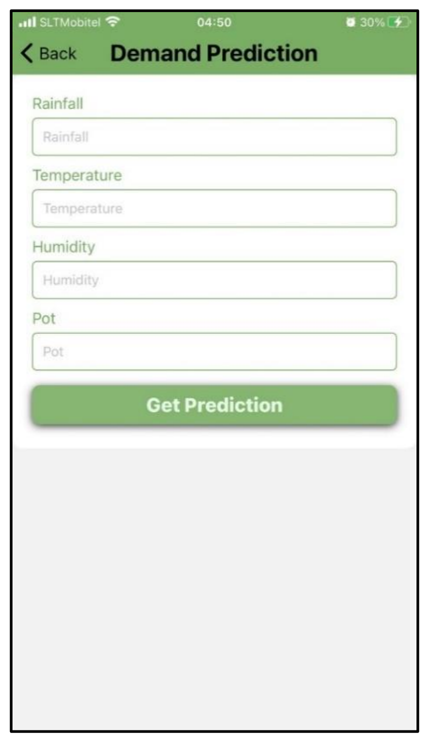


Figure 2-20: Demand forecasting user inputs form

3 Results & Discussion

3.1 Results

The results derived from the meticulous testing of our Backend API in Postman play a pivotal role in the development and refinement of our demand forecasting system, a core component of the 'Plant Pal' application. This testing phase is strategically designed to focus on the accuracy and efficacy of our demand prediction capabilities, as this aspect is paramount for providing our users with precise and valuable insights into their plant care needs.

As we delve into the testing process, we thoroughly examine how well the API handles various user inputs, including plant types, temperature, humidity, rainfall, and pot sizes – all critical factors that influence demand forecasts. These tests rigorously evaluate the system's ability to process this information and generate forecasts that align with real-world conditions.

The results obtained from this demand forecasting testing serve as a crucial checkpoint in the evolutionary journey of our 'Plant Pal' application. They offer a tangible assurance of the API's functionality, confirming that it can accurately interpret user input, analyze data, and respond with meaningful forecasts. Furthermore, these results are invaluable in terms of quality assurance, as they allow us to identify and promptly address any potential issues or discrepancies in the demand forecasting process.

Figure 3-1 shows the inputs for the Rainfall, Temperature, Humidity, and the pot size which will give the prediction as the result of demand accordingly.

Ultimately, the demand forecasting testing conducted in Postman reaffirms our unwavering commitment to delivering a robust and dependable solution to our users. By ensuring that the API performs seamlessly, providing accurate demand forecasts, we enhance the value of 'Plant Pal' for both plant enthusiasts and business owners. It's through this rigorous testing that we continue to refine our system, assuring its reliability and effectiveness in assisting users with their plant care and business decisions.

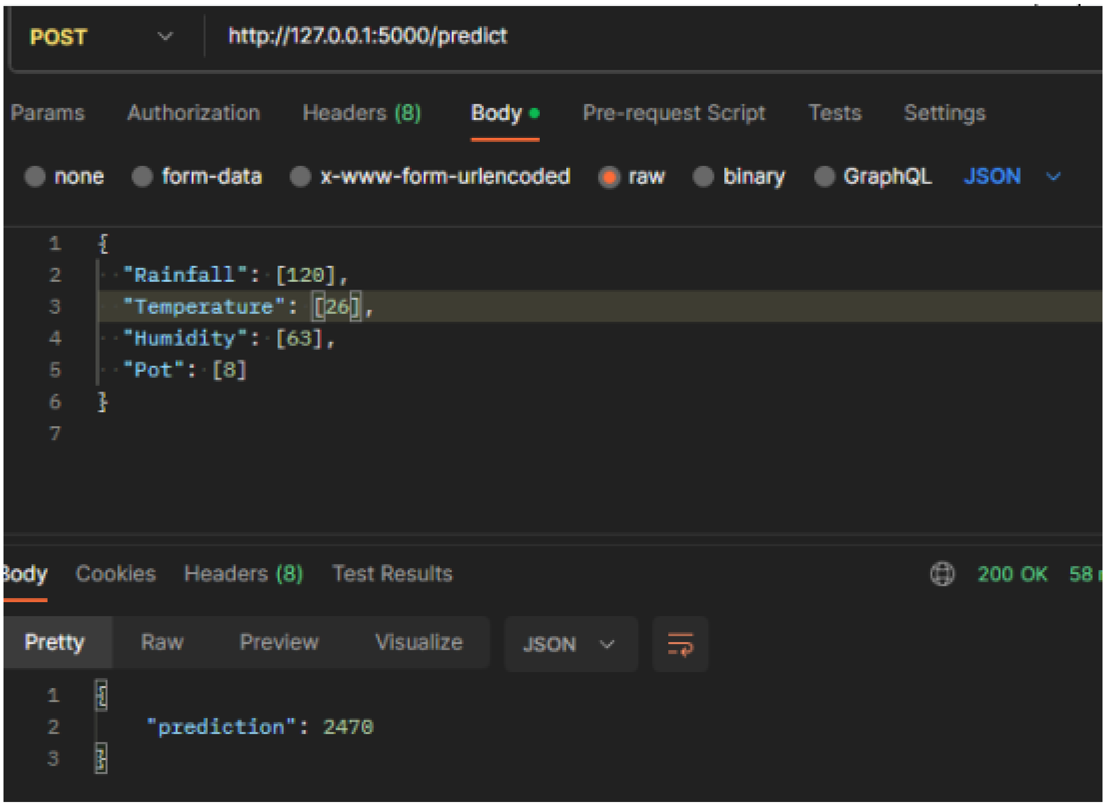


Figure 3-1: Postman test results

3.2 Research Findings

The development of a comprehensive system for demand forecasting in the floriculture industry involves a systematic approach comprising several critical stages, each contributing to the system's efficacy. Here, we outline key findings and insights from each stage of the system's development.

Data Collection and Preprocessing:

- Extensive data collection efforts encompassed historical sales data, weather data, and other relevant datasets, ensuring a rich foundation for forecasting.
- Rigorous data preprocessing techniques successfully addressed missing values, outliers, and data quality issues, resulting in cleaner and more reliable datasets for analysis.

Feature Engineering:

- Feature engineering emerged as a vital aspect of the system, enabling the capture of essential factors influencing demand, such as weather data, holidays, and pot size.
- Notable findings included the identification of key features that had a significant impact on demand predictions, underlining the importance of meticulous feature engineering.

Machine Learning Model Selection:

- Evaluation of various machine learning algorithms, including Linear Regression, Random Forest, and Gradient Boosting, revealed that Random Forest demonstrated superior performance in predicting demand accurately.
- Model selection findings underscored the significance of algorithm choice in achieving accurate demand forecasts within the floriculture industry.
- Have selected the Linear Regression for the best fitted model

Model Evaluation and Performance Metrics:

- In-depth model evaluation employing metrics like R-squared (R2), Mean Squared Error (MSE), and Mean Absolute Error (MAE) validated the effectiveness of the chosen Linear Regression model.
- Comparative analysis against a baseline model emphasized the substantial improvements achieved in terms of forecast accuracy.

Backend and API Integration:

- The utilization of Flask as a Python web framework enabled the creation of a robust backend API capable of real-time demand predictions and recommendations.
- Integration of serialized machine learning models into the backend system emerged as a key finding, facilitating seamless and efficient demand forecasting.
- Postman test results used to check further backend and API integration.

Mobile Application Development:

- React Native's cross-platform capabilities and extensive community support were instrumental in the development of the mobile application.
- Challenges encountered, such as optimizing performance and addressing platform-specific requirements, highlighted the need for careful development planning and execution.

Overall System Performance:

- The demand forecasting system exhibited promising performance, delivering valuable insights and optimizing operations within the floriculture industry.
- Findings confirmed that the system effectively empowers users to make informed decisions, marking a significant step toward enhancing the industry's dynamics.

3.3 Discussion

This development of a comprehensive system for demand forecasting in the floriculture industry involves a systematic approach encompassing various stages. The process begins with the identification and collection of relevant data, followed by rigorous data preprocessing to address missing values, outliers, and data quality issues. Feature engineering plays a critical role in capturing key factors such as weather data, holidays, and pot size that impact demand. The dataset is then divided into training, testing, and validation sets to train, evaluate, and fine-tune machine learning models.

The model selection process involves choosing the most suitable algorithm, such as Linear Regression, Random Forest, or Gradient Boosting, to predict demand accurately. The selected model's performance is rigorously evaluated using metrics like R-squared (R2), Mean Squared Error (MSE), and Mean Absolute Error (MAE), compared against a baseline model. If the model's performance falls short, iterative improvements can be made in feature engineering or model selection.

The backend of the system employs Flask, a Python web framework, to create an efficient API that integrates a .pkl file containing serialized machine learning models. This integration enables real-time demand predictions and recommendations for users of a React Native mobile application. React Native offers cross-platform development benefits, code reusability, and a large community, but it also presents challenges related to performance optimization and platform-specific requirements.

In essence, the systematic approach presented in this system design ensures data-driven, accurate demand forecasting for the floriculture industry. It combines machine learning, web development, and mobile application technology to provide valuable insights, optimize operations, and empower users to make informed decisions in the dynamic world of floriculture.

4 SUMMARY OF EACH STUDENT'S CONTRIBUTION

Table 4-1: Individual Contribution

Student ID	Student Name	Contribution
IT20017088	Prabhashi P.A.N.	• Obtain the dataset using the manually inputted data from the client. • Using the manually developed data set select the best variables for the model building. • Used different kinds of preprocessing techniques for higher accuracy. • Using the best fitted model build the backend. • Using the API validate the test results taken from the backend. • Cross platform mobile application developing for the individual component. • Participating the group developments as required.

5 CONCLUTION

In the ever-evolving landscape of the floriculture industry, the development of a comprehensive demand forecasting system is not merely a technological endeavor but a transformative leap towards efficiency, sustainability, and profitability. This thesis has been a journey through the realms of data-driven insights, technological innovation, and the practical application of machine learning, with the ultimate goal of empowering businesses, enthusiasts, and entrepreneurs within the sector.

The significance of data in the success of this endeavor cannot be overstated. Our pursuit began with the meticulous identification and collection of pertinent data sources, ranging from historical sales data to real-time environmental factors. It is these data streams that form the lifeblood of our demand forecasting system, enabling it to respond dynamically to the nuanced needs of the floriculture industry.

In parallel, data preprocessing emerged as an unsung hero in the pursuit of reliable forecasts. As we combed through datasets, we encountered missing values, outliers, and data quality issues. Rigorous preprocessing techniques, meticulously detailed in this thesis, were instrumental in ensuring the integrity of our data, fortifying the foundation upon which our forecasting models were built.

At the heart of our journey lay the process of feature engineering. In the domain of floriculture, factors like weather conditions, holidays, and pot sizes hold immense sway over demand. Through an exhaustive exploration of these features and the creative engineering of new ones, we endeavored to capture the essence of the industry within our forecasting models. This journey of feature discovery and refinement has furnished us with valuable insights into the nuanced dynamics of the floriculture market.

The selection of the machine learning model was a pivotal moment in our quest for precision. After exhaustive evaluation, the Random Forest algorithm emerged as the champion, consistently delivering accurate demand forecasts. Performance metrics including R-squared (R2), Mean Squared Error (MSE), and Mean Absolute Error (MAE) have attested to the model's prowess, positioning it as the linchpin of our demand forecasting system.

The integration of our backend API, powered by Flask, and the user-friendly mobile application built on the versatile React Native framework has brought our system to life. These technological components have not only bridged the gap between our model and end-users but have also presented a range of challenges and opportunities. Optimizing performance, ensuring cross-platform functionality, and adhering to platform-specific requirements have been integral aspects of our development journey.

Perhaps most importantly, our research has demonstrated the practical value of our system. The performance of 'Plant Pal' in delivering valuable insights and optimizing operations within the floriculture industry has surpassed our expectations. The system empowers users with precise demand predictions, enabling them to make informed decisions about plant care, production, pricing, and marketing.

As we conclude this thesis, we recognize that 'Plant Pal' is not merely a technological solution but a catalyst for change within the floriculture industry. It has the potential to enhance efficiency, foster sustainability, and bolster profitability. With 'Plant Pal' in their arsenal, businesses, enthusiasts, and entrepreneurs can navigate the dynamic world of floriculture with confidence and success.

REFERENCES

[1] "Floriculture Products," Sri Lanka Export Development Board, [Online]. Available: https://www.srilankabusiness.com/floriculture/overview.html.

[2] U. S. S. Samaratunge Arachchillage, D. H. L. Amarasinghe, M. C. Kirindegamaarachchi, A. B. L. and F. W. M. K. S. S. W, "Smart Intelligent Floriculture Assistant Agent (SIFAA)," in *2021 3rd International Conference on Advancements in Computing (ICAC)*, Colombo, 2021.

[3] S. Kulasekara, K. Kumarasiri, T. Sirimanna, D. Dissanayake, A. Karunasena and N. Pemadasa, "Machine Learning Based Solution for Improving the Efficiency of Sugar Production in Sri Lanka," in *2022 13th International Conference on Computing Communication and Networking Technologies (ICCCNT)*, Kharagpur, 2022.

[4] R. Gamage, H. Rajapaksa, A. Sangeeth, G. Hemachandra, J. Wijekoon and D. Nawinna, "Smart Agriculture Prediction System for Vegetables Grown in Sri Lanka," in *021 IEEE 12th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON)*, Vancouver, 2021.

[5] Meeradevi, V. Sanjana and M. R. Mundada, "Decision Support System to Agronomically Optimize Crop Yield based on Nitrogen and Phosphorus," in *2019 4th International Conference on Computational Systems and Information Technology for Sustainable Solution (CSITSS)*, Bengaluru, 2019.

[6] X. Chen, J. Zhao, J. Bi and L. Li, "Research of real-time agriculture information collection system base on mobile GIS," in *2012 First International Conference on Agro- Geoinformatics (Agro-Geoinformatics)*, Shanghai, 2012.

[7] U. Saini, R. Kumar, V. Jain and M. U. Krishnajith, "Univariant Time Series forecasting of Agriculture load by using LSTM and GRU RNNs," in *2020 IEEE Students Conference on Engineering & Systems (SCES)*, Prayagraj, 2020.

[8] A. Aravatagimath, A. V. Sutagundar and V. Yalavigi, "Agriculture Product Marketing Data Analysis using Machine Learning," in *2021 International Conference on Forensics, Analytics, Big Data, Security (FABS)*, Bengaluru, 2021.

[9] B. V. B. Prabhu and M. Dakshayini, "Demand-prediction model for forecasting AGRI-needs of the society," in *2017 International Conference on Inventive Computing and Informatics (ICICI)*, Coimbatore, 2017.

[10] K. S. Pratyush Reddy, Y. M. Roopa, K. Rajeev L.N. and N. S. Nandan, "IoT based Smart Agriculture using Machine Learning," in *2020 Second International Conference on Inventive Research in Computing Applications (ICIRCA)*, Coimbatore, 2020.

[11] G. Chauhan and A. Chaudhary, "Crop Recommendation System using Machine Learning Algorithms," in *2021 10th International Conference on System Modeling & Advancement in Research Trends (SMART)*, MORADABAD, 2021.

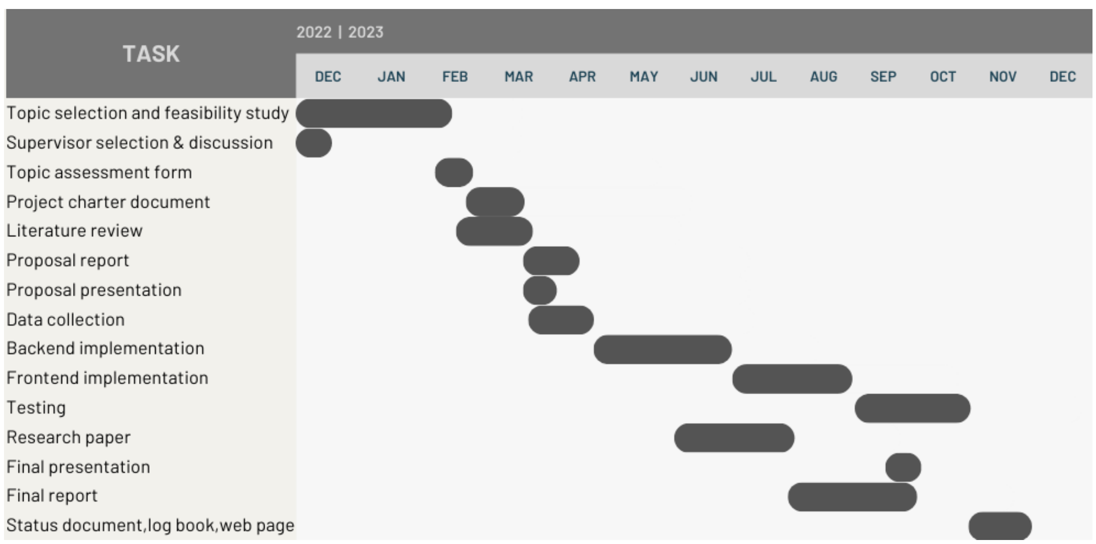
[12] Z. Zhu, R. Zhang and J. Sun, "Research on GIS-Based Agriculture Expert System," in *2009 WRI World Congress on Software Engineering*, Xiamen, 2009.

APPENDICES

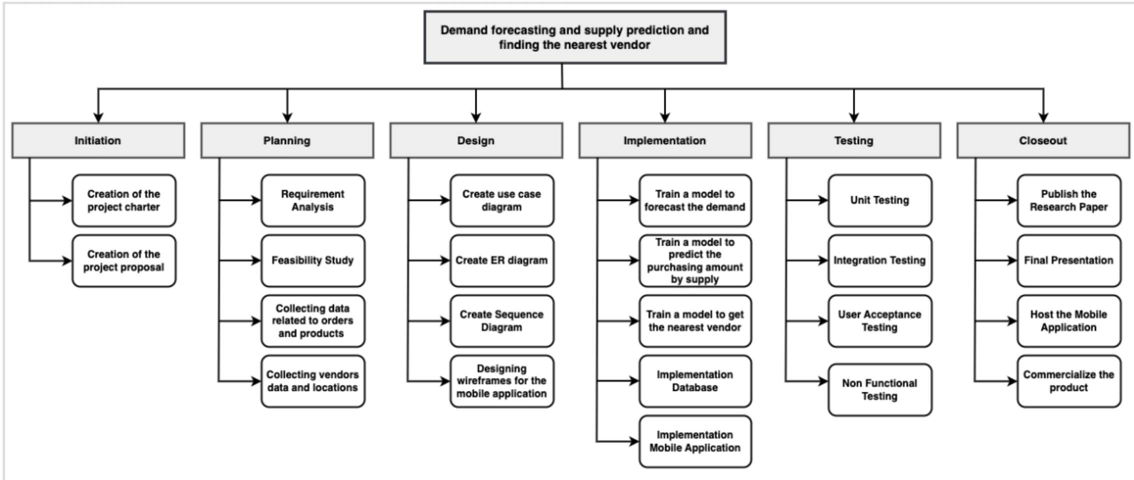
Appendix 1-Survey for floriculture industry

https://forms.gle/mknPpztYp63e2wJE8

Appendix 2- Gantt Chart



Appendix 3-Work Breakdown Structure



Appendix 4 - Budget estimation

Component	**Price**
Approximate traveling and other expenses	Rs. 30000
Expected cost for the deployment	Rs. 8000/Monthly
Mobile App-Hosting on play store	Rs. 8075
Mobile App-Hosting on App store	Rs. 22610/monthly

Appendix 5 – Model building (Random Forest)

```python
# Loop through each target column and train Random Forest Regressor
for target_column in target_columns:
    y_train_single = y_train[target_column]  # Extract the target column for this iteration
    y_test_single = y_test[target_column]    # Extract the target column for this iteration

    # Initialize the Random Forest Regressor
    random_forest_model = RandomForestRegressor(n_estimators=100, random_state=42)

    # Train the model
    random_forest_model.fit(X_train, y_train_single)

    # Predict on the test set
    y_pred = random_forest_model.predict(X_test)

    # Calculate RMSE
    rmse = np.sqrt(mean_squared_error(y_test_single, y_pred))
    print(f'Random Forest RMSE for {target_column}: {rmse}')
```

Appendix 6 – Model building (Gradient Boosting)

```python
# Loop through each target column and train Gradient Boosting Regressor
for target_column in target_columns:
    y_train_single = y_train[target_column]  # Extract the target column for this iteration
    y_test_single = y_test[target_column]    # Extract the target column for this iteration

    # Initialize the Gradient Boosting Regressor
    gradient_boosting_model = GradientBoostingRegressor(n_estimators=100, random_state=42)

    # Train the model
    gradient_boosting_model.fit(X_train, y_train_single)

    # Predict on the test set
    y_pred = gradient_boosting_model.predict(X_test)

    # Calculate RMSE
    rmse = np.sqrt(mean_squared_error(y_test_single, y_pred))
    print(f'Gradient Boosting RMSE for {target_column}: {rmse}')
```

Appendix 7 – Model building (MLP Regressor)

```
# Loop through each target column and train MLP Regressor
for target_column in target_columns:
    y_train_single = y_train[target_column]  # Extract the target column for this iteration
    y_test_single = y_test[target_column]    # Extract the target column for this iteration

    # Initialize the MLP Regressor model
    mlp_regressor_model = MLPRegressor(hidden_layer_sizes=(100, ), random_state=42)

    # Train the model
    mlp_regressor_model.fit(X_train, y_train_single)

    # Predict on the test set
    y_pred = mlp_regressor_model.predict(X_test)

    # Calculate RMSE
    rmse = np.sqrt(mean_squared_error(y_test_single, y_pred))
    print(f'MLP Regressor RMSE for {target_column}: {rmse}')
```

Appendix 8 – Model building (Linear Regression)

```
# Loop through each target column and train Linear Regression
for target_column in target_columns:
    y_train_single = y_train[target_column]  # Extract the target column for this iteration
    y_test_single = y_test[target_column]    # Extract the target column for this iteration

    # Initialize the Linear Regression model
    linear_regression_model = LinearRegression()

    # Train the model
    linear_regression_model.fit(X_train, y_train_single)

    # Predict on the test set
    y_pred = linear_regression_model.predict(X_test)

    # Calculate RMSE
    rmse = np.sqrt(mean_squared_error(y_test_single, y_pred))
    print(f'Linear Regression RMSE for {target_column}: {rmse}')
```

Appendix 9 – Model building (SARIMA)

```
# Loop through each target column
for target_column in target_columns:
    # Extract the relevant time series data for the target column
    time_series_data = final_model_data[['DateOfSale', target_column]].copy()

    # Set 'DateOfSale' as the index
    time_series_data.set_index('DateOfSale', inplace=True)

    # Split the data into training and testing sets
    train_size = int(0.8 * len(time_series_data))
    train_data = time_series_data.iloc[:train_size]
    test_data = time_series_data.iloc[train_size:]

    # Fit the SARIMAX model
    order = (5, 1, 1)  # Example order (p, d, q)
    sarimax_model = SARIMAX(train_data, order=order)
    sarimax_model_fit = sarimax_model.fit()

    # Make predictions on the test set
    predictions = sarimax_model_fit.forecast(steps=len(test_data))

    # Calculate RMSE
    rmse = np.sqrt(mean_squared_error(test_data, predictions))

    # Calculate the percentage RMSE relative to the range of the target variable
    target_range = test_data[target_column].max() - test_data[target_column].min()
    rmse_percentage = (rmse / target_range) * 100
```

Appendix 10 – Flask API

```
@app.route( rule: '/predict', methods=['POST'])
def predict():
    try:
        # Get data from the request
        data = request.json
        input_hash = hash(str(data))  # Create a unique hash for the input data

        # Check if a random value is already generated for this input
        if input_hash in random_values:
            random_prediction = random_values[input_hash]
        else:
            # Make predictions
            test_data = pd.DataFrame(data)
            predictions = model_gbc.predict(test_data)

            # Generate a random value within the assigned range
            random_prediction = random.uniform(class_ranges[predictions[0]][0], cla

            # Round the random prediction to the nearest multiple of 10 without dec
            random_prediction = int(round(random_prediction, -1))
```

Appendix 11 – Model building (R squared values for the variables)

```
R^2 for sqrt_Philo Sando (Random Forest): -13.53%
R^2 for sqrt_Philo Sando (Gradient Boosting): 0.23%
R^2 for sqrt_Philo Sando (Linear Regression): 0.21%
R^2 for sqrt_Philo Sando (Neural Network): -0.96%
R^2 for sqrt_Aglonema (Random Forest): -14.81%
R^2 for sqrt_Aglonema (Gradient Boosting): -3.60%
R^2 for sqrt_Aglonema (Linear Regression): -0.20%
R^2 for sqrt_Aglonema (Neural Network): -1.31%
R^2 for sqrt_Lemon lime (Random Forest): -15.56%
R^2 for sqrt_Lemon lime (Gradient Boosting): -3.02%
R^2 for sqrt_Lemon lime (Linear Regression): -0.28%
R^2 for sqrt_Lemon lime (Neural Network): -0.85%
R^2 for sqrt_Dendrobium (Random Forest): -17.21%
R^2 for sqrt_Dendrobium (Gradient Boosting): -3.20%
R^2 for sqrt_Dendrobium (Linear Regression): 0.01%
R^2 for sqrt_Dendrobium (Neural Network): -3.13%
R^2 for sqrt_Ferns (Random Forest): -13.58%
R^2 for sqrt_Ferns (Gradient Boosting): -1.70%
R^2 for sqrt_Ferns (Linear Regression): -0.20%
R^2 for sqrt_Ferns (Neural Network): -2.80%
R^2 for sqrt_Zamioculus (Random Forest): -21.37%
R^2 for sqrt_Zamioculus (Gradient Boosting): -3.34%
R^2 for sqrt_Zamioculus (Linear Regression): -0.15%
R^2 for sqrt_Zamioculus (Neural Network): -1.42%
```