SMART ASSISTANCE FOR THE FLORICULTURE INDUSTRY

2023-133

Project Proposal Report

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

February 2023

SMART ASSISTANCE FOR THE FLORICULTURE INDUSTRY

2023-133

Project Proposal Report

Prabhashi P.A.N.
IT20017088

Dr. Sanvitha Kasthuriarachchi
Dr. Lakmini Abeywardhana

B.Sc. (Hons) 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 proposal 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.

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)   2023.05.07

…………………………….   ..……………………….

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 supply, and Geospatial technology for mapping and visualizing production and distribution data is required to overcome these obstacles. The proposed system aims to develop a supply and demand prediction system for the floriculture industry, utilizing GPS technologies to identify the nearest and most suitable suppliers. Additionally, a recommendation system for packaging needs throughout delivery will be implemented. 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 and supply, Geospatial technology, GPS technologies, recommendation system

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

ANN	Artificial Neural Network
GPS	Global Positioning System
LightGBM	Light Gradient Boosting Machine
ARIMA	Auto Regressive Integrated Moving Average
SVM	Support Vector Machine
LSTM	Long Short-Term Memory
KNN	K-Nearest Neighbors
IoT	Internet of Things
GRU	Gated Recurrent Unit

GIS	Geographic Information System
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

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 supply 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 and supply prediction are crucial for the success of the floriculture industry in Sri Lanka.

Forecasting accurate demand helps the industry to identify the types of flowers, foliage, and ornamental plants that are in demand in different regions of the world. This enables the industry to focus on producing the right products in right quantities, which in turn helps to reduce wastage and increase profitability. Accurate demand forecasting also allows the industry to plan production cycles, manage inventory levels, and optimize logistics operations. On the other hand, predicting supply is also important for the floriculture industry because it helps to ensure a consistent supply of products throughout the year. Supply prediction can be used to identify potential shortages or surpluses of certain products, allowing the industry to adjust production schedules and manage inventory levels accordingly. Predicting the supply can also help the industry to optimize logistics and transportation operations, reducing costs and improving efficiency.

By utilizing GPS technologies, growers and exporters can identify the nearest and most suitable suppliers, which can help streamline the supply chain, reduce transportation costs, and time, and improve the overall efficiency of production and distribution processes. Moreover, GPS technologies can assist in identifying the most suitable suppliers based on factors such as price, quality, and availability, enabling growers and exporters to make informed decisions about their suppliers. This can lead to more stable and reliable supply chains, ultimately improving customer satisfaction and increasing competitiveness in the global market.

By using advanced technology such as predictive analytics software, the recommendation system can analyze various factors such as the type of product, transportation method, and duration of transportation to determine the most suitable packaging material for each product. This can help minimize product damage and improve the overall quality of the products during transportation.

There have been several types of research in these components for forecasting market demand, supply prediction, demand forecasting, and more. However, the majority of them have either not done so in the floriculture industry or have only done so partially.

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 other 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 the 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. In order 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 kind of machine learning techniques and other advanced techniques used various predictions like demand forecasting, and supply predictions. 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 rapid growth of the floriculture industry in Sri Lanka, inaccurate demand forecasting and supply prediction can result in lost sales opportunities, excess inventory, and wastage of resources such as water, fertilizer, and labor. Furthermore, the lack of utilization of GPS technologies in the industry can lead to inefficient transportation and logistics operations. As a result, there is a need for an advanced technology-based system that can accurately predict demand and supply, recommend suitable packaging, and utilize GPS technologies to improve transportation logistics in the floriculture industry of Sri Lanka.

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, the majority of 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.

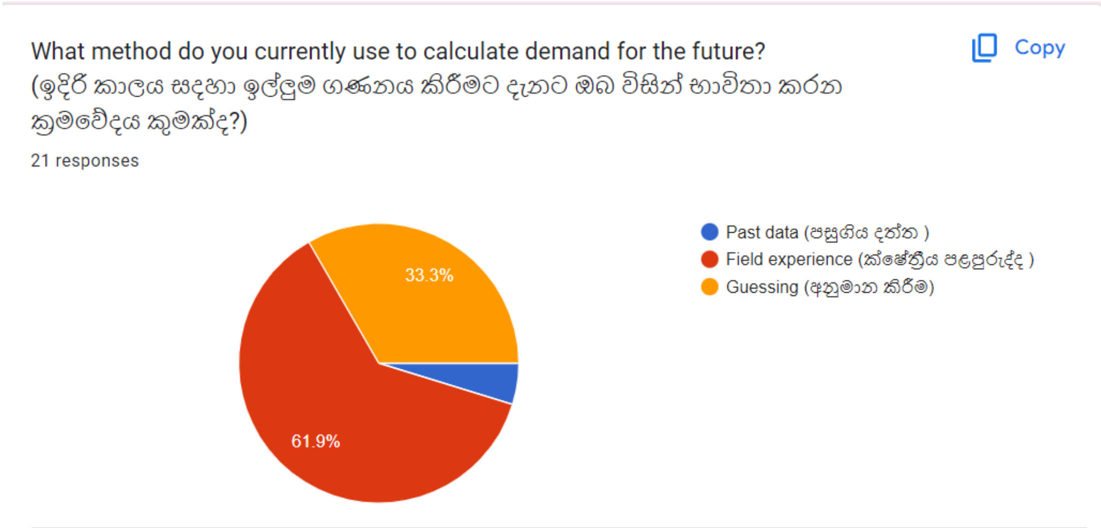

Figure 1-1 Survey to find the method that currently used to calculate the demand

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)

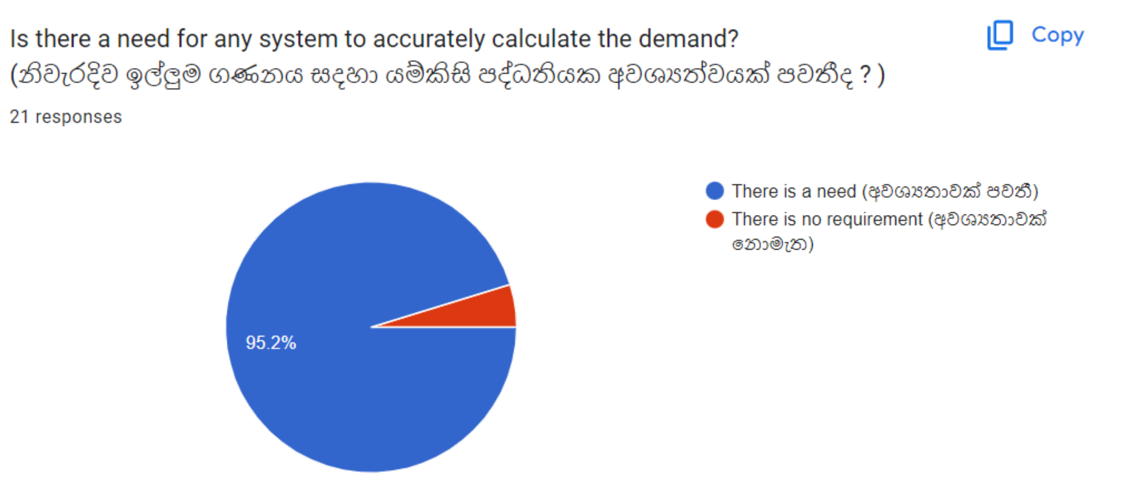

Figure 1-2 Survey to identify whether the system to calculate the demand accurately is needed or not

As stated by respondents in Figure 1-3, the methods they use to identify relevant external suppliers and determine whether they can fulfil an order in the floriculture industry in Sri Lanka. The responses indicate that the majority of respondents (61.9%) rely on their field experience to identify relevant suppliers and assess their ability to fulfil orders. On the other hand, 38.1% of respondents use guessing to identify suppliers and assess their capacity to fulfil orders. This suggests that there may be room for improvement in the methods used to evaluate suppliers and ensure that they can meet the demands of the business.

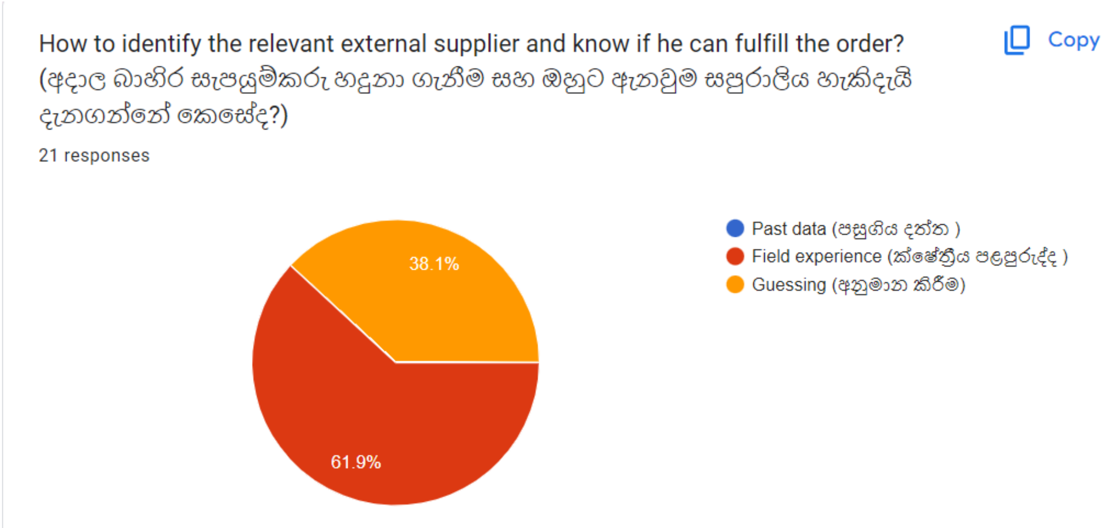

Figure 1-3 Survey to find how they find the most suitable external supplier

In accordance with the respondents about their need for software to calculate the amount of storage material required for plant transportation in Figure 1-4, the responses indicate that the vast majority of respondents (90.5%) believe that such software is needed. This suggests that there may be challenges or limitations to the current methods used to calculate storage requirements for plant transportation and that a more systematic and automated approach could be beneficial for businesses in various ways, like improving efficiency and more.

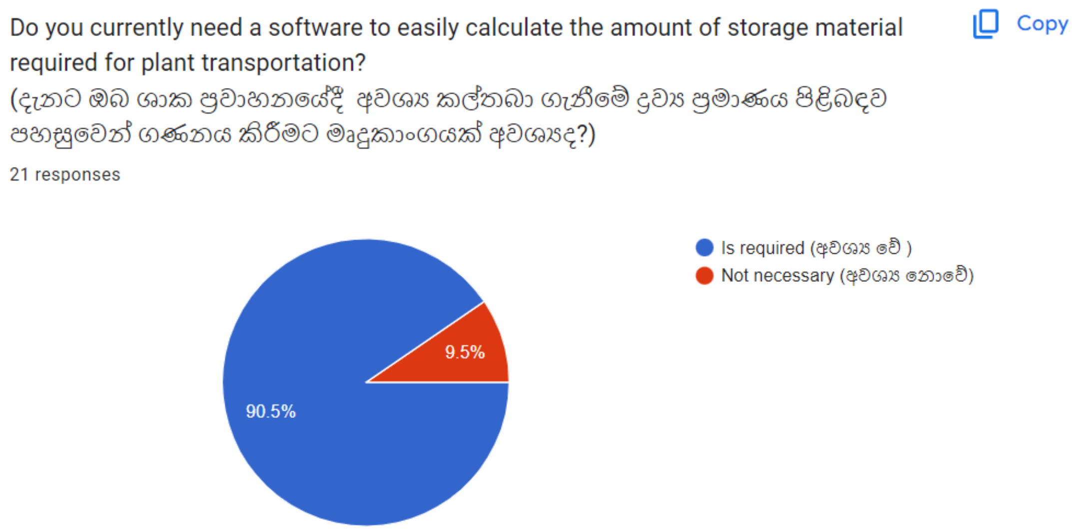

Figure 1-4 Survey to find that need of a software to calculate the storage material for transportation

The results show that the floriculture industry in Sri Lanka faces challenges in demand forecasting, supply prediction, finding the nearest vendor and packaging systems. The majority of respondents rely on personal experience to forecast demand and supply and then identify relevant suppliers, indicating a need for more systematic and data-driven approaches. Developing automated tools to forecast demand and supply and calculate storage requirements or packaging requirements could improve decision-making and reduce waste and costs. The industry could also benefit from identifying reliable suppliers and using data-driven transportation logistics to improve sustainability and competitiveness. By addressing these challenges, the industry could improve its sustainability and competitiveness and contribute to business growth and development.

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 these researches have been done in agriculture, not floriculture. Additionally, none of these will be discovered in Sri Lanka.

The substantial majority of these technological advancements are novel to our country, as the majority 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 forecast 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 of the 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 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]

Overall most of this research is done in the agricultural sector and also done in India, which means most of these technologies and ideas will be novel to Sri Lanka. Most importantly, it will be a whole new idea when it comes to the floriculture industry.

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	Supply Prediction using historical data	Using GPS technologies to point out the nearest and suitable supplier	Recommendation system for the packaging needs throughout the delivery	Mobile Application
[2]	✓	-	-	-	✓
[3]	✓	-	-	-	-
[4]	✓	✓	-	-	✓
[5]	-	✓	-	-	-
[6]	-	-	✓	-	✓
[7]	✓	-	-	-	-
[8]	✓	-	-	-	✓
[9]	✓	✓	-	-	-
[10]	-	-	-	-	-
[11]	-	-	-	-	✓

[12]	-	-	✓	-	-
Proposed System	✓	✓	✓	✓	✓

2 OBJECTIVES

2.1 Main Objective

Predicting the supply and demand, Finding the nearest vendor, and creating a Recommendation system for packaging and transportation

2.2 Specific Objectives

- Forecasting the demand for local and international markets considering multiple factors
- Revolutionizing the Floriculture Industry: Streamlining Order Fulfilment through Predictive
- Analytics and Efficient Supplier Management
- Finding the nearest vendor
- Predict seasonal and species-specific demand, estimate supply, and locate the closest vendor.
- Mobile application development

3 METHODOLOGY

3.1 Overall System Architecture

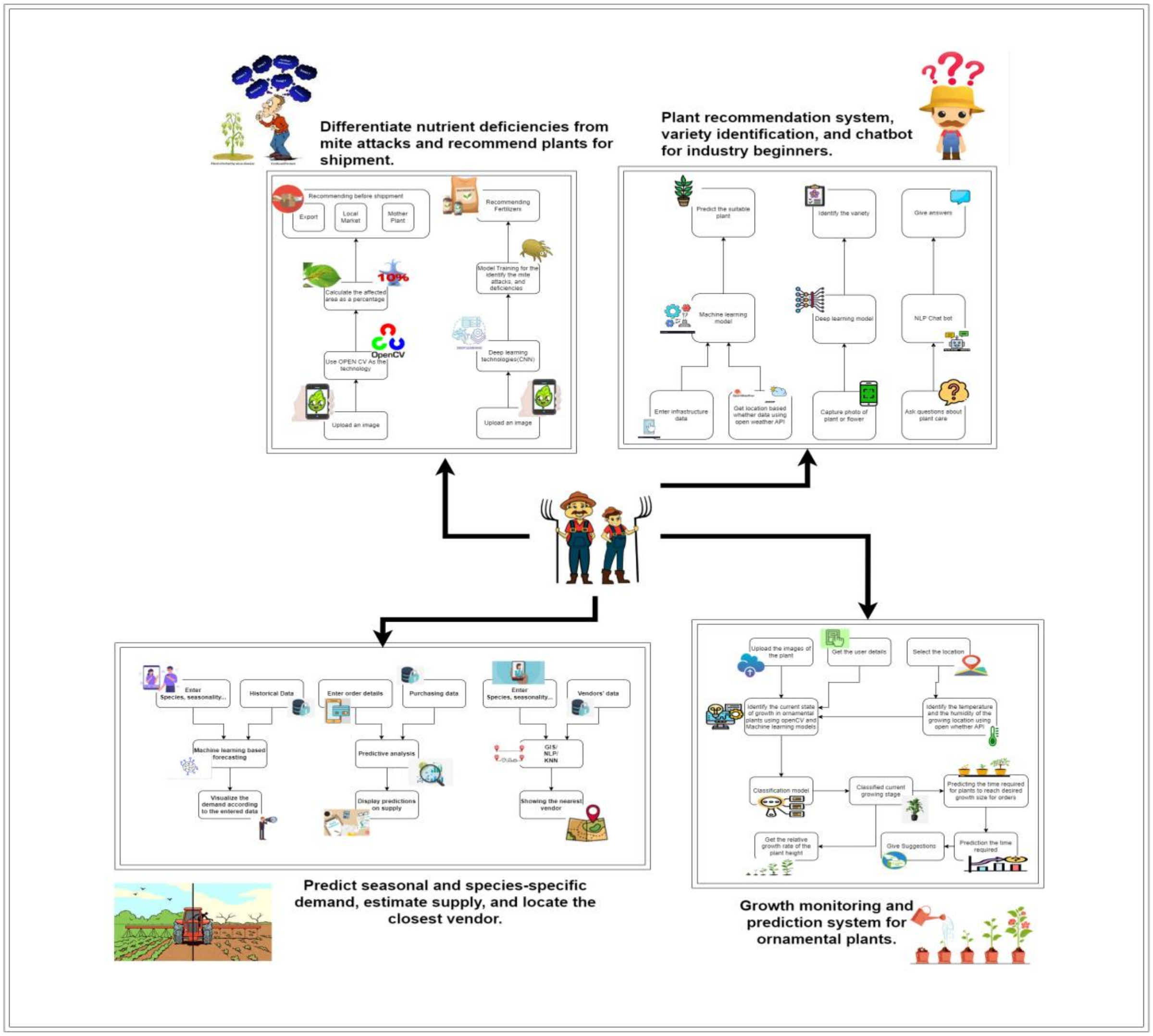

Figure 3-1 Overall system diagram

3.2 Individual System Architecture

In the Figure 3-2 it explains the architecture of the individual component which contain demand prediction, supply prediction and the finding the nearest vendor.

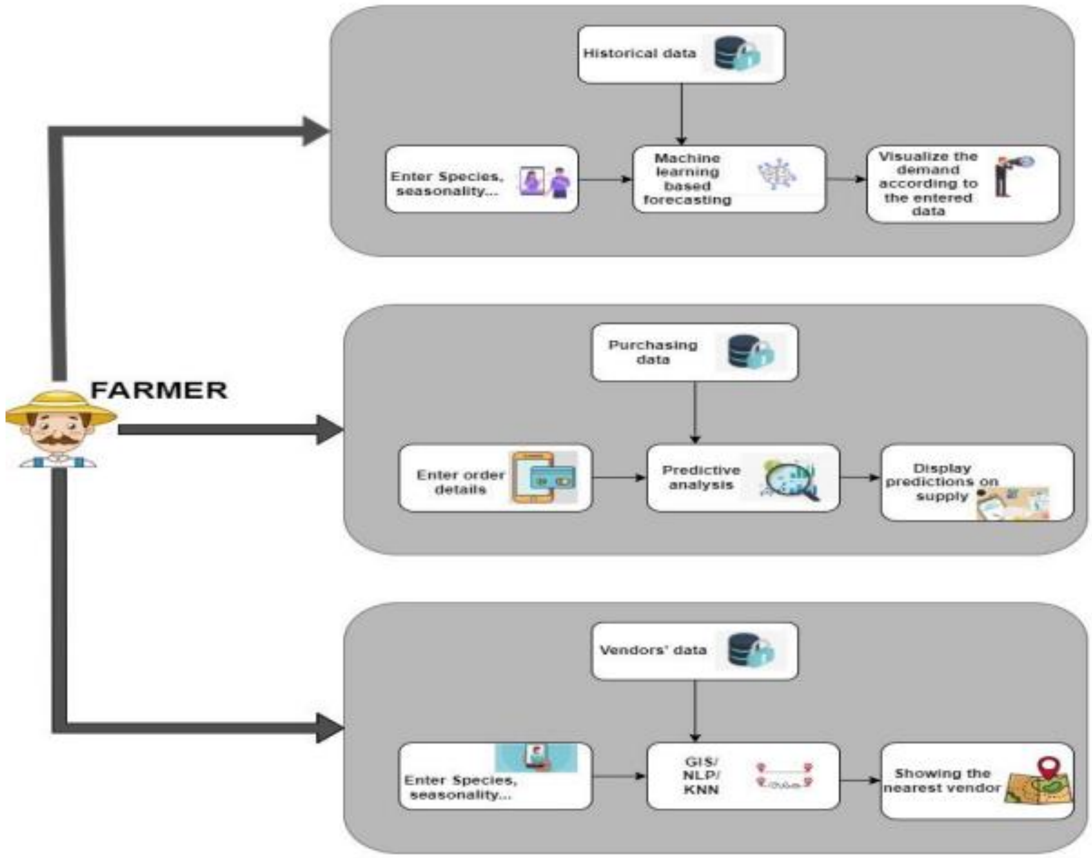

Figure 3-2 Individual system diagram

And this suggested method will be doing in the phrase of the agile methodology.

Table 3-1 This will be the summarize of the technologies that going to use on the proposed system.

Table 3-1 Summary of tools and technologies used

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

3.2.1 Demand prediction

- Data Collection: First, identify the relevant data that is needed, such as historical sales data, order details, countries, seasonality according to the order, market trends, customer behavior and more details. Then need to determine the time frame and the length of the sales cycle, and the seasonality of the floriculture industry. Then collect all the data according to the plan and organize a single data set
- Data Preprocessing and cleaning: Check for missing data and outliers and decide on an appropriate strategy for handling them, such as imputation or removal. The strategy will be changed according to the data set. Then identify any data quality issues, such as duplicate entries or inconsistent formatting, and resolve them. Ultimately, transform the data into an acceptable format for analysis by, for example, converting category variables to numerical variables based on the model that will be applied.
- Feature selection and feature engineering: Identify the relevant features that may impact demand, such as weather data, holidays, events, and more. Use

exploratory data analysis and correlation analysis to identify the most important features. At last, create new features, such as lagged variables, rolling averages, or seasonality indicators, to capture any patterns or trends in the data.

- 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: Within the dataset, separating the training set from the testing set, ensuring that the data is randomly sampled and stratified, then reserve a portion of the data as a validation set to use for model selection and hyper parameter tuning.
- Applying Feature Scaling: Normalizing the data using techniques such as MinMaxScaler, StandardScaler, or RobustScaler to ensure that the data is on a similar scale and avoid any feature dominating the model.
- Model Building: Selecting the range of models suitable for the task of demand prediction, such as Linear Regression, Random Forest, or Gradient Boosting. Use the training dataset to fit each model and assess their performance on the validation dataset. Then select the best-performing model based on the evaluation metric(s) of interest, such as R2, MSE, or MAE.
- Model Accuracy and Selecting the Best Model: Use the testing dataset to evaluate the final model's performance, which will indicate the model's effectiveness when applied to new data. Use the appropriate metrics to evaluate the model's performance and compare it to a baseline model. Included among relevant metrics are R2, MSE, and MAE. In the event that the model's performance is subpar, we will replace it by go back to the feature engineering or model selection stage and iterate until a satisfactory result is achieved. According to this plan, the model will be much more accurately done.

3.2.2 Supply prediction

- Data Collection: For the supply prediction, the necessary data such as inventory levels, production schedules, weather forecasts and, most importantly, supplier information will be needed. Then, after determining the timeframe as previously, collect data from Omega Green Pvt Limited with their permission.
- Data Preprocessing and cleaning: Like previously, this is also going to check for missing data and outliers and decide on an appropriate strategy for handling them, such as imputation or removal. Then look for the duplicate inconsistencies and resolve them.
- Feature selection and feature engineering: As previously that, the features that are going to impact the supply, like production capacity, inventory levels, and weather conditions, need to identify them and find the pattern in them. Then using that knowledge in the model when developing.
- Handling Imbalance with Stratified Sampling: The class imbalance stratified sampling can still be used to ensure that each class is represented in training and testing datasets in proportion to their overall frequency. This can help to prevent bias towards the more common packaging or transportation options and improve the overall accuracy of the model.
- Splitting the Dataset: The dataset should be divided into three groups: training, validation, and testing. Randomly sample the data for each set, ensuring that each set is representative of the overall dataset. Use the training set to train the model, the validation set to fine-tune the model and choose the appropriate parameters, and the testing set to evaluate the trained model's overall performance.
- Applying Feature Scaling: The same feature scaling will add accordingly
- Model Building: There are multiple ways to predict the supply, but the most suitable will be chosen according to the data set that is going to be prepared.

- Linear Regression Model: Linear regression is a simple yet effective model that can be used to predict the supply of flowers based on various input features such as weather conditions, market trends, and past sales data. This model assumes that there is a linear relationship between the input characteristics and the target variable.

- Random Forest Model: A random forest is a method that can be used for supply forecasting within the context of ensemble learning. It requires the construction of multiple decision trees during the training phase, followed by the output of the modes of the classes (during classification) or prediction means (during regression) for each individual tree. This model can handle complex relationships between input features and target variables and can also handle outliers and missing data.

- Gradient Boosting Model: Gradient boosting is another ensemble learning method that can be used for supply prediction. It builds an ensemble of weak models in a stage-wise manner and optimizes a loss function by adding models sequentially. This model was developed with the ability to handle non-linear relationships between input features and target variables, as well as absent data.

- Long Short-Term Memory (LSTM) Model: This model is a type of RNN model that can be used for predict the supply. This model can learn from historical data to predict future supply levels based on past supply trends, as well as other relevant input features such as weather and market conditions.

- Model Accuracy and Selecting the Best Model: Use appropriate evaluation by utilizing metrics such as mean squared error (MSE), mean absolute error (MAE), or R-squared to determine how well a model performs on a testing set. Either the error's magnitude or the R-squared value enhances the quality of the model in direct proportion.

3.2.3 Finding the nearest vendor

- Data Collection: Identify relevant data sources, such as Geolocation data, vendor information, user preferences, behavior and more. Then determine the appropriate format for the data collection, such as real-time streaming or batch processing, based on the requirements of the mobile application.
- Data Preprocessing and cleaning: Removing any irrelevant data and handle missing data. This step ensures that only relevant data is included in the analysis.
- Feature selection and feature engineering: Remove any irrelevant data and handle missing data. This step ensures that only relevant data is included in the analysis. Create new features from the existing ones, such as calculating the total weight or volume of the shipment, to improve the accuracy of the system.
- Handling Imbalance with Stratified Sampling: Encode categorical variables, such as product type and delivery options, into numerical values that can be used by the model. The numeric variables must be normalized or scaled to ensure they are all on the same scale and to prevent any one characteristic from dominating the others. Identify and handle any outliers or extreme values in the data that may affect the accuracy of the model. If there is an unequal distribution of data across different products or regions, use techniques such as stratified sampling to ensure that each class is represented in training and testing datasets in proportion to their overall frequency.
- Splitting the Dataset: Dividing the preprocessed data into training, testing, and validation sets.

- Applying Feature Scaling: Normalize data using techniques such as MinMaxScaler, StandardScaler, or RobustScaler to ensure that data is on a similar scale and avoid any feature dominating the model.
- Model Building: Depending on the task's objectives and the available data, a variety of models should be considered. k-nearest neighbors (KNN), support vector machines (SVM), decision trees, and random forests are examples of such models. The subsequent phase is to evaluate each model's performance using the appropriate performance metrics, such as accuracy, precision, recall, F1-score, and ROC-AUC score, respectively.
- Model Accuracy and Selecting the Best Model: When determining the efficacy of a model, it is essential to employ pertinent performance indicators such as mean squared error (MSE), mean absolute error (MAE), accuracy, precision, recall, F1-score, or ROC-AUC score. Then, compare model performance by evaluating each model's performance based on the selected metrics, and select the model that produces the best results when applied to the validation set.

3.2.4 Recommendation system for packaging

- Data Collection: Product information, transportation data, packaging data, and environmental data. As data, they can serve as the foundation for the construction of this model.
- Data Preprocessing and cleaning: Need to do the data cleaning as previously by checking for missing values, duplicate entries, and outliers. If you find any, you can either remove them or impute them based on appropriate methods.
- Feature selection and feature engineering: Identify the key features that could potentially impact packaging and transportation recommendations, such as the product's weight, dimensions, delivery location, and type of packaging material. Then create new features, such as the distance between the customer's location and the delivery location, time of day, season, and weather conditions, to enhance the accuracy of the model.

- Handling Imbalance with Stratified Sampling: Using these methods, the imbalance will be handled in the model. One common approach is to oversample the minority class to create a balanced dataset. This can be achieved by replicating the minority class samples or by synthesizing new samples using techniques such as Synthetic Minority Oversampling Technique (SMOTE). It is also possible to create a balanced dataset by arbitrarily under sampling the class that constitutes the majority of the population. To achieve this, the number of samples collected from the majority class must be decreased until it is equal to the total number of samples collected from the minority class. Stratified Sampling: Using stratified sampling during the dataset split can ensure that each class is represented in training and testing datasets in proportion to their overall frequency.
- Splitting the Dataset: Splitting the data set relevantly according to the model and the data set.
- Applying Feature Scaling: Normalizing the data set using the most appropriate method, like previous models.
- Model Building: Select a range of models suitable for the task of recommending packaging and transportation options, such as collaborative filtering, content-based filtering, or a hybrid model. Use the training dataset to fit each model and assess their performance on the validation dataset. Select the model with the highest performance based on the metric(s) that are most important to you, such as accuracy, precision, or recall.
- Model Accuracy and Selecting the Best Model: Using the most appropriate one according to the data set and considering the need
    - Collaborative filtering: This model is based on the assumption that individuals with similar preferences in the past will continue to hold those preferences in the foreseeable future. Through the use of user-item interactions, collaborative filtering is able to recommend products that

other users who are similar to the current user have previously purchased or relished.

- Content-based filtering: This model recommends items based on the similarity of their features or characteristics to items that a user has previously liked or interacted with. This approach relies on item features such as product descriptions, categories, attributes.
- Hybrid filtering: This model is based on the assumption that individuals with similar preferences in the past will maintain those preferences in the foreseeable future. Collaborative filtering is capable of recommending to the current user items that other analogous users have previously purchased or appreciated. This is achieved through the use of user-object interactions.
- Matrix Factorization: This model factorizes the user-item interaction matrix into lower-dimension matrices, which allows it to find latent features that can be used to make recommendations.
- Deep Learning Models: These models can be used to learn complex patterns in user-item interactions by using deep neural networks, which can improve the accuracy of recommendations.

3.3 Commercialization & Business Plan

The commodity version will be as follows:

- Prediction of flowering period for ornamental plants
- Estimation of optimal growth conditions (e.g., temperature, humidity, light intensity) for different plant species
- Identification of plant diseases and pests and recommendation of appropriate treatments
- Recommendation of the most efficient irrigation and fertilization practices based on plant type, location, and weather conditions
- Forecasting the supply and demand for specific flower varieties in different regions and markets

The premium version will be developed as follows:

- Prediction of the optimal harvest time for different flower species
- Providing suggestions for improving plant health and reducing the risk of disease or pest infestation
- Identification of the best-suited soil type and nutrient requirements for different plant varieties
- Integration of a virtual garden planner to help users plan and design their flower gardens
- Integration of a vendor management system to help users find the nearest and most reliable vendors for their flower-sourcing needs
- Automated tracking and monitoring of inventory levels to ensure timely replenishment of flower stocks and avoid stock outs

4 PROJECT REQUIREMENTS

4.1 Functional Requirements

- Implement data preprocessing techniques to clean and prepare the datasets for analysis.
- Develop machine learning models for demand and supply forecasting using methods such as regression and time series analysis. Visualize the results to gain insights into trends and patterns.
- Integrate geolocation data to enable the system to identify the nearest vendors and provide location-based recommendations for packaging and transportation.
- Use feature selection and engineering techniques to identify the most relevant variables that impact packaging and transportation and develop a recommendation engine to suggest optimal packaging strategies based on the plant's lifespan and other factors

4.2 Non-Functional Requirements

- Security: To protect the privacy of its users and prevent unauthorized access, the mobile application should support encrypted communication and data storage.
- Reliability: The application should be reliable and available at all times to ensure that users can access the features and functionalities whenever they need them.
- Scalability: As the application acquires additional users and data sets, the system must be designed to accommodate the growing number of users and data sets.
- Usability: The application should be user-friendly and easy to navigate, with clear instructions and guidelines for users who may not be familiar with the technology.
- Compatibility: The application should be compatible with a wide variety of mobile devices and operating systems in order to attain the highest level of accessibility possible.

4.3 System Requirements

- Python will be utilized to implement the backend, which will be used for model construction and training, model deployment, and data preparation.

- Machine learning models for supply and demand prediction and locating the closest vendor will be developed and trained using TensorFlow, an open-source machine learning framework.

- Spacy, an open-source package for sophisticated natural language processing, will be used to build and deploy the NLP pipeline for the recommendation system.

- VS Code will be used as the code editor for this implementation, and it will be utilized to write and debug the code.

- Data analysis, data visualization, and machine learning experimentation will be conducted using Jupiter Notebook.

- React Native, a cross-platform development tool, will be used to create the mobile application for iOS and Android.

4.4 User Requirements

- The user needs to be able to check the supply and demand according to a given time frame and make wise decisions using it.
- Users need to be able to get the chance to find the nearest vendor according to the user-given data.
- Users need to be able to get ideas by recommendations for the packaging accordingly.

5 GANTT CHART

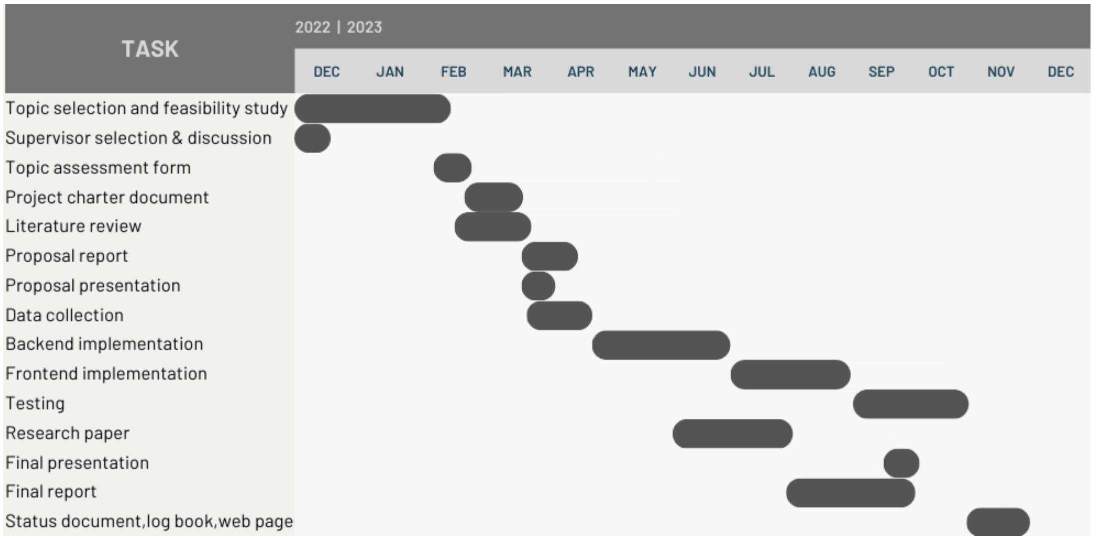

Figure 5-1 Gantt chart

6 WORK BREAKDOWN STRUCTURE

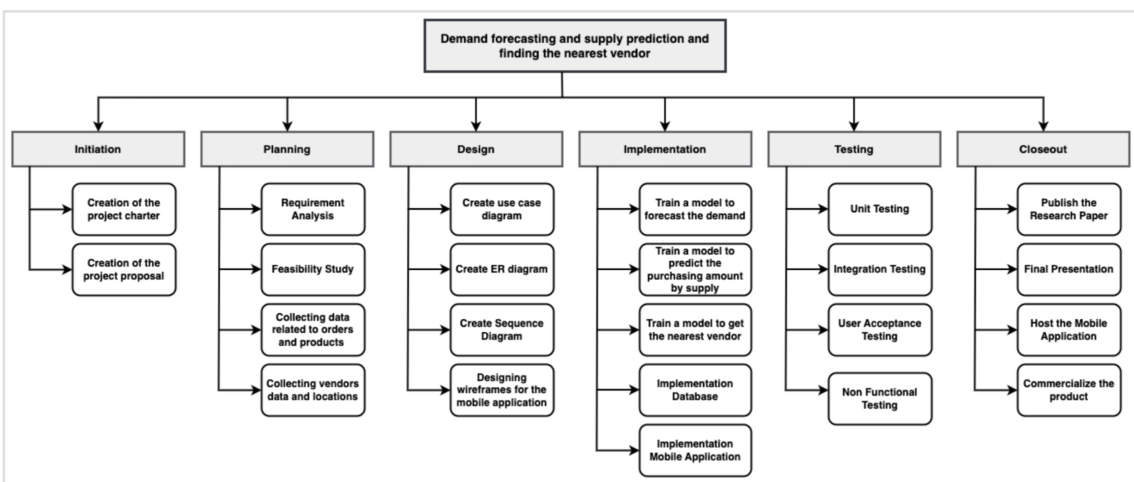

Figure 6-1 Individual work breakdown structure

7 BUDGET AND BUGET JUSTIFICATION

The following table (Table 7-1) will show the overall budget for the project approximately

Table 7-1 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

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 A: Survey on Floriculture Industry

Appendix 1-Survey for floriculture

https://forms.gle/mknPpztYp63e2wJE8