International Journal of Computer Information Systems and Industrial Management Applications.
ISSN 2150-7988 Volume 15 (2023) pp. 550-565

Best Profitable Crops Prediction with Profit, Cost, and Farmland Optimization using Machine Learning

A.W.H.P. Alagalla[1] and Lokesha Weerasinghe[2]

[1] Faculty of Graduate Studies and Research, Sri Lanka Institute of Information Technology,
Colombo 00300, Sri Lanka
heshanalagalla@gmail.com

[2] Faculty of Computing, Sri Lanka Institute of Information Technology,
Malabe 10115, Sri Lanka
lokesha.w@sliit.lk

Abstract*: Agriculture is the fundamental practice that sustains human life by cultivating plants for human consumption. That is why since ancient history every government has made its maximum effort to strengthen the Agriculture. But now Agriculture is facing several new challenges due to many reasons. Especially Production shortages, Climate change, Cost of production, Fluctuating Market prices, and Provincial/Regional conditions have influenced to the production of the farmers. In Sri Lanka when any crop gains profit in previous season, majority of farmers tent on to cultivate that crop in the next season. But that could result in an increase in the market supply causing a depletion the market price. But there should be a variety in selecting suitable crops in order to optimize profitability and crop rotation could be practiced optimizing yield potential. This research tries to predict the best crop for the Local and Export Market in Sri Lanka by Provincial/District/Regional wise to gain more profits using Machine Learning. This solution is designed to analyze the total profit by considering cost factors for production materials and employee wages. This facilitates farmers to identify suitable crops for local and export markets easily. Research elaborated on how the Machine Learning model should develop using special Algorithms for the prediction and classification of Data using many Statistical Concepts and Computational Algorithms. One of the most important and novel factors of this research is that it introduces a technique to optimize the land area of the farmer using predicted two crops. Because this research tries to encourage the farmers to gain a suitable and satisfactory income without affecting the market prices. Apart from the acquired profits, it could also be identified as a suitable solution for the supply chain issues in the Sri Lankan market. Furthermore, this research produces a profit maximization technique to enhance the Sri Lankan farmer's livelihood and economy.

***Keywords*:** Machine Learning, Best Crop Selection, LSTM, Time Series, Production Forecasting

I. Introduction

Agriculture has been one of the main occupations in the world since ancient times. There isn't anything wrong to recognize Agriculture as the best occupation in the world that gives life to every other person who are doing other occupations. If there is a world without Agriculture no one will survive on the planet earth. When it comes to the countries of the world point of view from ancient history, Agriculture played the most important role in the day-to-day life of humans. Even today it hasn't changed. Even though humans manufactured a huge number of high technological devices, people can't eat those things except the Food cultivated by agricultural community. As a result, most of the nations in the world work on development of agriculture related technologies and devices. Apart from that, the United Nations (UN) has initiated worldwide novel programs to strengthen the quality of the life of farmers and to improve agricultural production by introducing many concepts through the World Food Program. The World Food Program has introduced advanced technologies to enhance the agricultural products of a country by combining technologies in their experiments. Developed countries like United States, United Kingdom, Germany, China, and Israel have shown magnificent progress through agriculture technology. These countries continue research and development by combining agriculture and smart technologies and develop massive analytical platforms to get the maximum benefit from cultivation.

This research was planned to focus on multiple issues raised by the farming community in Sri Lanka. Those problems were answered by Information Technology and Computer Science-based applications. As a technological solution, Machine Learning is used to predict the best crop for the desired districts by analyzing various datasets that have been gathered by the Department of Statistics in previous years. This research tries to give a guided plan to farmers to get an idea of which crop should be cultivated in their area to get maximum profit in the ongoing season. For that purpose, research was planned to use Machine Learning and related packages to build analytical models. The Research planning to build an analytical model using machine learning and related statistical packages to analyze the major decision criteria of Sri Lankan agricultural product.

In present, climate change has become the major environmental issue for every business sector. As a result of global warming, a more heat waves are sweeping the US, Canada, and Europe in recent years and there are considerable negative impacts for other countries as well. Favorable

climate conditions are an important factor for cultivation. Fluctuation in expected yield has become a key criterion which creates supply uncertainties. Globalization, which is fueled by growth of information technology has led to interdependency of nations. Most of the countries in the world heavily depend on imports, as globalization diminished self-sufficient nations in the world. Sri Lanka is developing country, which struggling lot with inflation and economic recession in the country since from recent past. There was trade imbalance in the country from decade. According to the statistics from last two decades [23], there is fluctuation in annual change of imports and exports in Sri Lankan agricultural sector. However, Agricultural import share in total import in Sri Lankan shown a steady flow during past two decade with average of 8.35 percentage [24]. Due to this imbalance price drops can be observed in most food items comparative to production cost. Farmers spend more on fertilizers, equipment, seeds, and labor force. When there is price drops, demand uncertainties and other such issues, farmers try to move with more demanding crops which has attractive markets. But the framers community in Sri Lanka, has no proper guidance, any data driven approach or any reliable technical solution to select bets crop for their lands.

An analytical model was developed to predict the best crop. For this purpose, Python language and its statistical libraries were used by the authors. As a supportive platform for python, Anaconda was utilized in this model development. By using the above technologies, the following goals and objectives were focused by this study. The following objectives were defined after analyzing previous similar research works to build the research gap.

- Recognize the climate patterns in the Sri Lankan region that affect agriculture in Yala and Maha seasons.
- Identify the Market price trending pattern in various districts.
- Analyze the production pattern from various districts and regions.
- Predict the best profitable crops to be planted in the desired area/district or region.
- Best Crops should be produced for local and export markets.
- The cost factor for production will be very crucial for the solution to maximize profits. Because that cost factor should be analyzed.
- Importations to the local markets and their effect on the local market should be analyzed.
- The solution should be a cost-effective and more profitable technique.
- Farmers should have an option to analyze the data through the Web Dashboard/Interface.
- Farmers should have the option to get a list of crops that can be cultivated on their owned Land by considering Profit maximization and Land Optimization.
- A complete solution should be specific to Sri Lanka

The possibility should be there to expand the model in a worldwide manner.

II. Literature Review

A. Agriculture Production and World Economy

Agriculture is an important industrial sector for all nations. Most of the developed economies have a strong contribution from their agriculture sector toward the national economy. Several studies around the world have shown the essence of the development of the agriculture sector as it heavily contributes to poverty and reduces hunger of people. Rather than in the early days, there is sudden fluctuation in the environment. These environmental changes like natural disasters and climate changes have drastically influenced the agriculture sector. And in present farmers and agriculture-related business companies struggle with many issues like the choice of the correct crop, appropriate land, the anticipated yield from their cultivation, etc. Manual prediction has less accuracy, and frequent and unexpected changes in the climate make it difficult to move with manual prediction. Hence research community showing increasing interest in accurate predictions in the field of agriculture [6].

Though it is essential to have accurate predictions, agriculture-related Prediction is not an easy thing to do for any researcher due to its complexity. But some researchers have taken the challenge to find this tremendous task and make it a reality. When analyzing world research, some scientists have tried to do predictions using supervised machine-learning techniques for various datasets that have been obtained from different sources. Machine-learning algorithms detect patterns and learn how to make predictions and recommendations by processing data and experiences, rather than by receiving explicit programming instruction. The algorithms also adapt in response to new data and experiences to improve efficacy over time. There are different types of machine learning like supervised learning, unsupervised learning, and reinforcement learning.

In supervised learning, algorithms work together with the human touch. Here human feedback is always linked to improving the relationship between inputs and output obtained through analysis. Usually, this link relates to several statistical methods like linear regression, naïve base, logistic regression, and decision three.

Unsupervised learning is another algorithm, it focuses more on input data without being given a clear output variable. It helps to identify a group of data that behave with a similar pattern. Usually, this utilizes algorithms like Gaussian mix clustering and K- mean clustering.

In reinforcement learning, these algorithms try to minimize the tasks that need to perform, the recurrent auto-correction method is assigned to this algorithm. This is heavily used in optimization strategies and the field of robotics.

Machining learning is an evolving area, as a result, today, it has several variants. Deep learning is a hot topic that links with both machine learning and artificial intelligence. Deep learning is an advanced technique, which has the capability of learning from a given data set.

Deep learning is a part of machine learning and provides

appropriate modeling with several processing layers. And the link of human behavior through artificial intelligence has added more value to deep learning. Ultimately deep learning can generate intelligent decisions through advanced analytics base predictions.

B. Applications of Prediction Techniques in Agriculture

Agriculture-related Prediction isn't an easy thing to do for any researcher due to its complexity. But some researchers have taken the challenge to find this tremendous task and make it a reality. When analyzing world research, some scientists have tried to do predictions using supervised machine-learning techniques for various datasets that have been obtained from different sources. According to [4], they have elaborated on the more scientific way to analyze the Previous Data in Bangladesh related to Agriculture. From that, they emphasized the Supervised Machine Learning Technique as the best way to predict the future harvest accurately. They took six crops for the analysis that is comprising, potatoes, Jute, wheat, and three rice types. Researchers have taken data regional-wise across Bangladesh and it comprised a data of Production-based. To analyze the data they used the Decision Tree Algorithm with the KNN Regression. They have used Temperature values and Rainfall values in Bangladesh in the supervised learning algorithms and analyzed how the same regions perform by production. Apart from above mentioned basic Decision Trees [4] built a specific machine learning model for the solution by Iterative Dichotomiser 3. That algorithm absorbs the attributes and relevant data from the dataset and classifies those data according to the Proposed Models. KNN regression was used to measure the Euclidean Distances from adjacent values of the dataset based on the number K value's optimal criteria. When it comes to the result and outcomes their research mainly contributes to the prediction by multiple classifies and they have shown the comparison of those classifier errors to get an idea about the best classifier to select for the prediction purpose. As final point authors said some of the crops clearly show ID3 as better accuracy and some of them show KNN as better accuracy for the prediction. Because they combine both classifiers into a hybrid model and perform the calculations.

The research examined by [5] has introduced another important technique based on Agriculture Sector in Iran. There is a very special climate condition mainly dominated by the draught. The motivation for the research was to increase the production of Crops to face Food safety problems for a rising population. However, they have proposed Machine Learning and Artificial Neural Networks to find the solution to predict the Harvest in the Future. They used SVM Classifier Model to detect the production of the rice. Apart from that, they tried to analyze the diseases that infected rice and strawberry crops using ANN Models. Mainly researchers put two variables for yields and losses for Agricultural Products and Live Animals and Livestock yield for Livestock Production in Iran. Although Multi-Layer Perception is proposed to use for the back-Propagation technique. It is a type of Supervised Learning algorithm that is easily deployed to problems. it is the three-layer architecture that is connected from the existing layer to the next layer. Another model that is used by [5] is ANFIS. That means an adaptive network-based fuzzy inference system that is one type of Hybrid concept that uses fuzzy logic to develop the algorithms. According to [6] developing a prediction platform is more accurate by using review protocols. It has collected data about soil moisture, type of soil, variety of plants, plant progress, production of the harvest, and many others. From that researchers tried to decide how the soil types of different regions affecting to the final production of the harvest. Apart from that, they have planned to build a technique to analyze how fertilizers affect the production of the harvest of that respective season. The authors tested many algorithms to find the best technique to predict the desired result. Neural Networks, Linear Regression, Random Forest, Support Vector Mechanisms, and Gradient Boosting Trees are a few of the used algorithms and concepts in this investigation. To check the accuracy of the prediction model it has elaborated on the techniques used. Root Mean Square Error, R-Squared, Mean Absolute Error and Mean Square Error have been experimented with within the project scope. And research conducted by [7] has elaborated on the feasibility of Deep Learning with Deep Learning. They suggested machine learning techniques with Multivariate Regression, Decision Trees and Association rule mining are suitable concepts to achieve the target of Crop production prediction.

When comes to India, is a highly agricultural country compared to other countries. Most of the people in rural areas are depending on the paddy field to survive. But Indian Climate is not the same as around the whole sub-continent. Some seasons in north India have very dry situations and sometimes have flooding and rainy situation. Because climate plays a huge role in India. Because of that [8] have done a deep analysis of the effectiveness of the Crop production prediction according to the Climate Situation. They built a software application called Crop Adviser to predict the best product for the best climate situation. That is very important for the farmers to get an idea about the farming techniques according to the weather. Apart from that, they are developing this application to find out what is the best influencing weather factor for crop production in relevant areas. They have found there is a big pesticide usage in the paddy and it is negatively interrelated with the harvest. Pest usage and meteorological data provided by the Indian meteorological department have been used to optimize the result using SVM. And K – the nearest neighbor of the dataset calculated by applying the weather condition variable to the SVM with the combination of the GPS Technologies. Here rain falling incidents in the future are forecasted by using many Data mining techniques and after applying the algorithm it says heavy rainfall has significantly affected Crop production in the season. The drought period and high rain situations at the peak of the crop production phases were reasoned to the reduction of the significant crop production. As a variable for the desired research, Maximum & Minimum temperature, highest and lowest rainfalls, Evapotranspiration condition, percentage of Cloud cover, and frequency of Wetness have been selected by the researchers. The decision Tree Algorithm was used to calculate the prediction process [8]. In the presentation mode, they elaborated the results by web

interface for various categories. For example how these harvests were received from different cloud cover conditions, temperature conditions, and rain conditions. Statistical analyses were conducted to check the accuracy of the received analytical output from the algorithms.

The review done by [9] emphasizes the importance of Deep Learning for Crop Management. To prove their idea they focused to review the multiple research papers that have been written regarding crop management by machine learning deep learning. The papers reviewed in [9] elaborated on techniques used to find the maximum temperature timeframes by using an SVM classifier and other ML algorithms. Predicting the temperature throughout the year was suggested conducted by Artificial Neural Networks. But there was some issue to find the best-executed concept and algorithm for performance indexes of crop management conducted by Random Forest forecasting. It was performed with 87.5% accuracy within the experiment boundaries. Indian Crop management and compatibility are was checked with the ANN Back-propagation techniques. But researchers emphasized this technique is difficult to implement because of its high complexity. Although some of the researchers suggested time series should be integrated into the functions that need to capture more long-term dependencies. Apart from that Indian Authorities have done a comprehensive analysis and built a decision-making system to predict the climate and crop inter-relationship using Association rule mining. By using the above algorithms China predicted they were facing a reduction of the Maize Harvest by 13% in 2050. Africa got the result from using above mentioned concepts and algorithms to select more appropriate land to get better production in the future. Finland, Germany, and many other European countries gained the conclusion that they reduced their crops by 5-6% if the temperature increased by 1 degree of Celsius. Because of that [9] has clearly shown evidence and importance of Machine learning or deep learning for the prediction in the agriculture field. Several countries everywhere in the world now slowly moving to the usage of Machine Learning to increase their production and profits [9].

Similar to the above research [10] introduces another new innovative technique to use Machine learning in the field of agriculture. They proposed the data gathered from the respective data sources should be divided into two parts. They emphasize the importance of allocating 75% for training purposes. The other 25% should be allocated to testing purposes. They encourage that technique due to the performance and results gained from the previous research. And data they have gathered from various sorts of sources e easy to recognize as descriptive and diagnostic data. Supervised Learning Algorithms are used in projects when there is a technique to rain previous data and predict the result using new data. Those data are mainly recognized as classified data. But Unsupervised Data isn't classified data. Because [10] suggested conducting deep analysis before applying any algorithm to the research, researchers should get a clear idea about what sort of research they are planning to do. The authors of this research used Random Forest Classifier for the Prediction purpose because it is capable of doing classification and regression with the same technique. Then summarizes the data and integrates it with the constraints used for prediction. After that, they created a web interface to present the calculated results in a user-friendly manner. In a conclusion, they emphasized their accuracy of reduction that is reaching to above 75% percentage. It is a better percentage when it reached the agriculture prediction. Another report elaborated by [11] has deeply elaborated the technique that is used to find the missing values of the data. They used feature scaling techniques to find the values and prepare the full prediction of production.

When it comes to the economy of the farmers' Agriculture production is very important to the Farming society. Not only for the farmers, it was a highly necessary thing for the Economy of the country. [12] Mentioned they were done informative research to analyze prices of specific products in agriculture according to the market prices of previous years. They have divided the whole research into two parts. The first part was a development of a machine learning technique to predict the price of the crop in the market. Another part is developed to predict the Harvest production in the relevant area. They have used several variables in the algorithms comprising weather conditions, history of the production, and soil information. They built a specific model to calculate the statistical values of the algorithm by using the Random Forest technique. Apart from that Crop Price Prediction part was done by using unsupervised learning algorithms. Maximum Trade Value, Lowest Trade Value, Trade compared to production, and many other variables have been used in the proposed system. Other than the above Cultivation cost and seed cost was also gathered from the Farmers to design this. For the prediction purpose, researchers elaborated and encourage to use Naïve Bayes Algorithm as a Classification technique. Apart from the Naïve Byes K Nearest Neighbors (KNN) are also used to predict crop prices. The authors of [11] introduced a new model for this research and is comprised of nonlinear functions that have very useful characteristics regarding machine learning/Deep Learning. Other than that, it has been agreed that the proposed methodology should use the concepts and algorithms of logistic regression, neural networks, and decision trees. Additionally, researchers attempted to analyze regional crop prices in China and the US. It examines the effects of those nations' imports and exports on the harvest's market prices. Many graphs have been used to analyze and compare prices, especially in the soybean industry. Although Researchers developed a model to compare imports and exports from China and the United States with regional pricing structures. The algorithms for Linear Regression, Decision Trees, XGboost, and Neural Nets compare actual values and estimated values since getting a sense of the precision and error rates of the desired machine learning algorithms is very helpful.

In machine learning techniques, forecasting and prediction are extremely valuable processes. The Partial Least Square (PLS), ANN, and Autoregressive Integrated Moving Average (ARIMA) are recommended for the proposed system [13]. By comparing the algorithms' efficacy using the same market data set, researchers have done an outstanding job. As a result, it can evaluate the chosen algorithm's precision, effectiveness, and productivity and choose the best algorithm for the project under consideration. Additionally, some researchers emphasize how well convolutional neural networks and deep

learning work together. A multi-processor cloud-based script has been used to collect agriculture data from both public and private sector organizations. Three crops are the main focus of the research. They are tomatoes, Chile, and brinjal. The study suggests a brand-new framework with precise predictive performance. The deep learning idea aims for each ML algorithm's phase with the lowest coefficient of variance.

In recent years, the majority of nations, in particular government-backed organizations, have also funded this kind of crop selection research globally. Research by Thomas (2020) demonstrates the distinctions between machine learning-based prediction and manual reduction, as well as the benefits of each approach. The initial researchers described in detail how they intended to carry out a feature selection process using the Sequential Forward feature Selection (SFFS) and Recursive Feature Elimination (RFE) techniques. And after narrowing down the dataset into the chosen set, they discussed the potential use of KNN, Naive Bayes, Support Vector Machine, Decision Tree, and Random Forest for the prediction process. The functions in the model that were used in the ML process to gauge the precision and accuracy of the results have been made very clear. The best classifier was selected for the analysis of the results after they compared the classifiers used in the process. Because those who use this model to understand crops and farming have no trouble making their own decisions on their own. Then it made sense for them to increase crop production and make a lot of money from farming. That will greatly benefit Sri Lanka's economy and its capacity to grow economically [14].

The methods for creating the best crop predictions using the Sri Lankan context were discussed in detail [15]. They have made use of data from the departments of agriculture and statistics. In order to forecast the optimal crop for each region, they have used machine learning algorithms. On the other hand, they examined the Auto Regressive Integrated Moving Average (ARIMA) time series technique. In addition, they employed a method called linear programming to optimize the prices of the crops that fell under their purview. Additionally, they have attempted to analyze the best crop regionally by combining an ML and statistical model. However, they didn't use cost factors as inputs in their models. Only Market Price was used for their optimization.

For quick decision-making, leaders at the public and provincial levels (such as the EU level) must take on a fundamental task: estimating crop yield. Ranchers can decide what to develop and when to develop it with the aid of an accurate harvest yield forecast model. There are various methods for determining editing yield expectations. This audit article has investigated what has been done regarding the use of AI in crop yield writing. One of the criteria we used to reject a distribution during our analysis of the recovered distributions was that it was an overview or traditional audit paper. In fact, the segment discussed in this sentence refers to those barred distributions as related work. The Papers conducted an audit that focused on assessing nitrogen status using AI. The paper makes the case that quick developments in detecting technologies and ML techniques led to financially astute farming arrangements. The article promotes taking a broad view to see new boundaries that can produce record crop yields. A survey paper on the application of machine learning in the horticultural field was distributed. Distributions that focused on cropping the board, housing the executives, irrigating the executives, and sowing the board were used in the investigation. The readiness of organic products to select the ideal gather time and yield expectation was the focus of a survey conducted by researchers (Li et al., 2018). They focused on the challenges and philosophies that are encountered in the agrarian area in the field of picture handling and AI, particularly in the identification of illnesses. The authors discussed a few AI techniques and how they apply to plant science. A survey paper on information mining's general application in the farming sector was distributed by researchers. They assumed that additional research should be conducted in order to understand how information mining into complicated horticultural datasets could be carried out. They examined the various information mining techniques used for crop yield prediction and assumed that information mining techniques could be used to address the harvest yield forecast. This paper is the primary SLR that focuses on the use of AI in the harvest yield prediction issue, according to our analysis of survey articles, particularly the large ones that are introduced in this segment.

Output	Input	Algorithm	Resource
Productior	Temperature values and Rainfall values	Decision Tree Algorithm with the KNN Regression	[4]
	Soil type, variety of plants, plant progress, soil moisture	SVM Classifier, Neural Networks, Linear Regression, Random Forest, Support Vector Mechanisms, and Gradient Boosting Tree, Multivariate Regression, Decision Trees	[6]
	temperature, highest and lowest rainfalls, Evapotranspiration condition, percentage of Cloud cover, frequency of Wetness	Decision Tree Algorithm	[8]

Table 1: Applications of ML Algorithm

III. Methodology

This section tries to explain the various techniques and the organized process that the researcher used. There are several statistical techniques and some important technical aspects because the primary research area is relevant to machine learning and a crucial area of mathematics known as the "science of data" is statistics. Data are therefore essential to this study. This chapter covers all statistical techniques as well as the various types of data, data sources, and data collection methods. Figure 1 illustrates the key phases of this research effort, and the section that follows gives a brief explanation of its key contents.

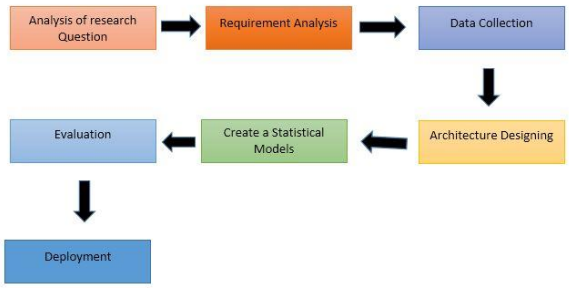


Figure 1: Research Methodology

A. Applications of Prediction Techniques in Agriculture

Analyzing the research question can give a thorough understanding of the research approach. The researcher has chosen a deductive research approach for the study in light of the problem's context. Because this study uses a top-to-bottom methodology, it was pertinent to test hypotheses before making decisions. The frameworks created by the variables identified from the knowledge gained through previous research work allow the deductive method to also reach a research conclusion. In order to move from existing theories to a research framework and hypothesis testing, researchers employ the deductive research approach. The inductive approach is a different research strategy. A bottom-to-top approach is an inductive approach. As a result, the researcher must make observations, carefully examine the data set, and then decide on a hypothesis based on the patterns in the data. Finally, the researcher could put the theory to the test and develop a theory based on the data. This strategy demanded more time and effort because it is complex. Given the time constraints and the abundance of credible literature sources, this study employs a deductive research methodology.

Another stage of the research methodology is the research strategy. The method for managing the research study is contained in this phase. There are a few methods for conducting investigations. Survey methodologies, Case Studies, Experimental methodologies, Action research, and Grounded hypotheses can be used as per their applicability. These strategies are chosen to determine how to lead the exploration. The survey method of information gathering has been attempted in this investigation. The survey was carried out to compile pertinent data from the data sources available. Particularly from the statistical reports released by government research organizations and institutes. As a result, the study relied on secondary data rather than primary data, which would require the researcher to gather new information for his study. This data from the study's secondary data set allows the final product to satisfy the study's predetermined general and specific objectives.

B. Research Choice

Three basic types of research choices are mono strategy, multi-strategy, and blended strategy. Utilizing only subjective, qualitative, or quantitative data and data analysis techniques is known as a mono strategy for the examination. The term "blended strategy" refers to a method that uses both techniques for the investigation and for translating the results. The use of both subjective and quantitative strategies should be evident in the multi-technique. Since the investigation falls under the quantitative strategy, the analyst in this study used the mono technique. However, the study uses a variety of statistical methods, which are classified as quantitative research.

C. Data Gathering and Analysis

A suitable data source or data set have been identified within the project's parameters after the research question has been examined. There should be a precise dataset featuring Sri Lankan agriculture available, and this research should be conducted specifically within the Sri Lankan context. Due to the majority of Sri Lankan departments using manual data reporting, it can be challenging at times. Data in soft copy is extremely uncommon in the Agriculture Department. For this reason, more effort should be put into locating a reliable dataset for the project.

By examining the production supply, climate, and market prices in Sri Lanka, this research is primarily focused on creating a machine-learning model to forecast the best crop for the desired regions/districts. Finding the dataset or data source for the project that contains Sri Lankan Data is the first step in conducting the research. The Department of Census and Statistics of the Sri Lankan Government and the Socio-Economic Planning Centre, Department of Agriculture, Peradeniya, Kandy, will both produce secondary data sets for this research. It contains a variety of information about Sri Lanka's agricultural sector. With 25 districts, this data covered the entire country of Sri Lanka. These are the Dataset's primary focal points, which are crucial to this study.

Agricultural data this dataset comprised with

- Size of the land used for rice production in recent years
- Size of the Land used for Other Crops in recent years
- Production of each crop type in tons
- The yield of each crop (district-wise)
- The cost of cultivation derives from Trends of changes in Average annual wages and in recent years cultivation cost
- Whether data
- Market Price data for each crop – Farm gate Price and Retail Price
- Import Quantities and Market Prices of Rice

D. Design Architecture

This research tries to propose a new architecture to predict the desired results using data gathered in the above section. The whole architecture was comprised of the main components and it is displayed in the Figure 2. Those are

1. Extracting Data from the dataset
2. Feature selection for the desired scope
3. Design a Statistical Model for every factor needed to absorb
4. Execute the Prediction process
5. Evaluation of Manual and predicted Analysis.

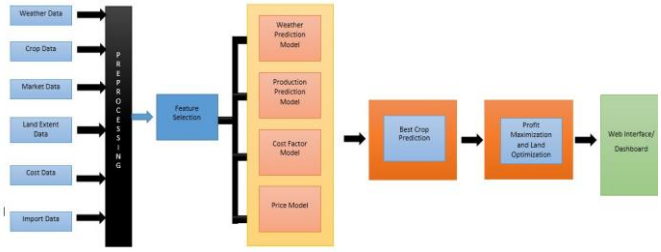


Figure 2: Design Architecture

Recurrent Neural Networks (RNN) are anticipated to be used for the project's Machine Learning Parts. Additionally, to specialize in the Neural Network Elements of the ML Procedures, Gate Recurring Unit Networks (GRU) were used. Decision Trees have been used in place of Naive Bayes (NB) and Support Vector Machines (SVM) for statistical calculations.

The Partial Least Square Regression (PLS) Technique will likely be used for Time Series Analysis Parts, among other techniques. In addition, numerous other Techniques will be tested to determine the most effective strategy.

E. Data Extraction

The extraction process comes first after architecture design. Recognizing the most crucial data from the original dataset is crucial for the project. This procedure included data cleaning, data type conversion, finding missing values, and correction. It also involved removing redundant values. Additionally, a thorough analysis of the dataset's attributes is required. A huge number of CSV files make up this dataset. Because several CSV files had to be run concurrently while the model was running.

F. Developing Statistical Models

The main goal of this analysis is to cover several topics. That range of data is not contained in a single CSV file. Since creating a sub-model for each of the different targets is the first step, and incorporating those sub-models into the final statistical model is the second. Crop data needs to be extracted and trained based on production, market price, and climate. Analysis should also be done of the Farmers' earnings from selling those crops due to the fact that this mode's appearance was extremely complex. The first step in doing that is to identify the project's key characteristics.

This article makes several new statistical models to predict various types of data in order to achieve that. This indicates that some of the data cannot be directly accessed from the dataset since different statistical models should be created to forecast desired results. Additionally, it is planned to incorporate those models into the final model after taking results from the first phase of statistical models. In this study, some models will be built using machine learning, while others will be created using time series models.

G. Weather Prediction Model

Used Technique:

- ❖ For Temperature and Humidity Prediction – ARIMA Time Series Technique
- ❖ For Rainfall prediction – LSTM

The weather is important for this research because climate change has a significant impact on plant growth. There are many differences, particularly when examining the Sri Lankan context of plantations and other types of agriculture that include rice, vegetables, and fruits. Climate plays a significant role in rice cultivation in Sri Lanka. The cause is that Sri Lanka grows rice primarily in two different zones. The first is a Dry Zone, while the second is a Wet Zone. The dry Zone includes large-scale rice-producing areas like Ampara, Anuradhapura, Jaffna, Batticaloa, and Polonnaruwa. The Wet Zone included the districts of Colombo, Gampaha, Kalutara, Kegalle, Galle, etc. The climates of these two areas are completely dissimilar. Farmers in the Dry Zone typically use lake water to cultivate paddy. During the rainy season, lake water is not necessary. However, during the dry season, farmers used lake water for farming. However, in the wet zone, farmers typically use rainwater for their farming operations. As a result of these, the climate is a crucial element in agriculture.

Research focuses on what is the best method to predict future behavior before choosing the ARIMA model. There are primary Time Series-based techniques to obtain the prediction, according to the analysis. Auto-Regressive Integrated Moving Average (ARIMA), Seasonal Auto-Regressive Integrated Moving Average (SARIMA), K-Neighbor Neural Networks, Bayesian Networks, and Recurrent Neural Networks. ARIMA or SARIMA have been chosen for development due to the Complexity and Usability of the Current Environment. Given that SARIMA is primarily used during one season, ARIMA was chosen over SARIMA. However, research is required to create a model for the various seasons for various districts.

An ARIMA model is characterized by 3 terms: p, d, q
Where,

- p is the order of the AR term
- q is the order of the MA term
- d is the number of differences required to make the time series stationary

A pure Auto-Regressive (AR only) model is one where Y (t) depends only on its lags. That is, Y (t) is a function of the 'lags of Y (t)

$$Y_t = \alpha + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + .. + \beta_p Y_{t-p} + \epsilon_1 \quad (1)$$

y_t - Temperature of period t
$y_{(t-1)}$ - Lag1 of the temperature series
β_1 – Coefficient of the lag1 (Have to identify through analysis output)
α - intercept (have to estimate by the model)
ϵ_1 - lagged forecasted error term some time (noise) (have to estimate by the model)

The moving Average (MA only) model is one where Y (t) depends only on the lagged forecast errors/ noise. Thus y (t) in down just consider the series of error terms in forecasted temperature values.

$$Y_t = \alpha + \epsilon_t + \phi_1 \epsilon_{t-1} + \phi_2 \epsilon_{t-2} + .. + \phi_q \epsilon_{t-q} \quad (2)$$

y_t - Forecasted error term for t
$\epsilon_1 \ldots \epsilon_{(1-q)}$ - errors of the autoregressive models of the respective lags
α – mean of the error terms
$\emptyset_t$ - Coefficient of the t^{th} error term

ARIMA Model

Predicted Y_t = Constant + Linear combination Lags of Y (up to p lags) + Linear Combination of Lagged forecast errors (up to q lags)

$$Y_t = \alpha + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + .. + \beta_p Y_{t-p} \epsilon_t + \phi_1 \epsilon_{t-1} + \phi_2 \epsilon_{t-2} + .. + \phi_q \epsilon_{t-q} \quad (3)$$

P – is corresponding to the lag series or series of temperature
Q - is the corresponding series of the error term

ARIMA (P, q, d) therefore have to find values for p, q, d through the analysis

To predict the best crop, it was necessary to gather information about rainfall, temperature, and humidity for the research project. The datasets from the Meteorological Department were used to collect weather data. Additionally, a Time Series-based Model will be used to analyze those data in order to forecast future weather and climate changes across various regions. The majority of this statistical model is made up of the Auto Regressive Integrated Moving Average. To create the mathematical and statistical model for the ARIMA-based temperature model, this research used two variables. The two variables in question are the date and the temperature. This model is used to forecast the coming temperatures in various regions of the nation. In addition, humidity can be used to forecast future humidity. It also employs the ARIMA model. In order to forecast future rainfall for various districts and time periods, an ARIMA-based model was also developed for rainfall. A weather model provides specific information regarding the various weather patterns in various districts for various reasons. The farmers must choose the best month to begin the Yala or Maha Season. The traditional methods to begin the Yala and Maha Season have existed for a long time and are the reason for that. However, the current rainy season has changed as a result of climate change, the modern farmers cannot cultivate using outdated data and patterns.

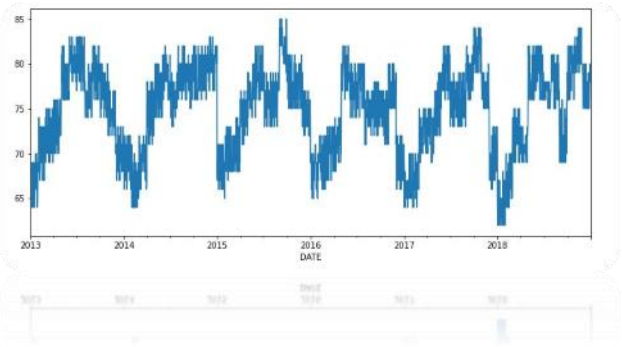


Figure 3: Temperature Variation

H. Cost Prediction Model

Used Technique: ARIMA Time Series Technique

The proposed techniques' production costs are yet another crucial consideration. The reason is that if farming costs are high, farmers will not make a profit from the plantation that is satisfactory. There were several components to this cost. The costs of fertilizer, chemicals, and laborers' wages are considered to be the cost factors. The price can change depending on the time. The cost of living is rising daily, which is the cause. Due to the fact that laborers in paddy fields and other plantations were compelled to demand higher wages in accordance with market prices and demand, material costs and fertilizer costs were also rising daily. To predict the future cost pattern, this model is employed. Another time-series-based model was used to make the prediction. Then, when any farmer chooses a Season or Month that is anticipated to be a growing season and where they have land to grow crops. The model processed relevant data and after that Cost of the Production was predicted by this Model.

I. Market Price and Profit Prediction Model

Used Technique: ARIMA Time Series Technique

For the research to be analyzed, the crop market price is crucial. The market prices for various crops during the previous growing season had a significant impact on farmers' decisions regarding the crops they would plant the following year. One of the main challenges in conducting this research is also this. The reason is that these farming-related decisions had an impact on crop market prices the following year. In other words, if potato farmers made a lot of money the

previous season, other farmers may try to grow the same potatoes since the harvest of potatoes increased. But the demand dropped. Additionally, the market price dropped. It causes the farmers a serious problem. Because of this, a system should exist to predict future changes in market prices for the various types of crops.

This model is used in particular to close that gap. Machine learning algorithms will be used to complete this task along with a combination of statistical libraries. The cost of production was considered when calculating profit as pure profit. Retail prices and import factors (rice) were used as inputs for this farm gate price. Different types of machine learning and statistical analysis-based algorithms were examined before choosing a method to predict future market price fluctuations. An ARIMA-based algorithm has been chosen in this case to forecast the upcoming market pattern. This won't show up on the dashboard. After determining the total cost in the Production Prediction and Auto Load Cost Option, it will send to the main dashboard.

J. Best Crop Prediction Model

Used Technique:

As an Overall Model, this is a combination of 3 models. Those are ARIMA, LSTM, and Statistical Classifiers (Random Forest, Naïve Bayes, Logistic Regression, SVM, Decision Trees, XGBoost)

The majority of farmers find it to be a very difficult task to choose the best crop. They still use conventional methods to choose the crop they want to cultivate today. They typically use two methods when cultivating crops. The first is that if their ancestors are growing a certain crop, the new generation will attempt to grow that same crop repeatedly. Another type is when they examine and watch how other farmers cultivate their land and make money the previous season. then choose the best crop to grow the following season. But when it comes to the profit-based mechanism, this is a very ineffective strategy. The goal of this model was to recommend the top two crops for the targeted districts. This model was created through the analysis of various factors. The primary variables in this model are temperature, humidity, rainfall, land area, and expected production. Here, the research's independent variables are temperature, humidity, rainfall, and extent. The other desired factors have been dependent on production, which is a dependent variable. According to the two seasons, the model is analyzing and processing data from 25 districts in Sri Lanka.

Long-short-term memory (LSTM), the ARIMA technique, and statistical classifier techniques were primarily combined to develop the machine learning process. This is why temperature and humidity predictions were processed using the ARIMA time series method. The LSTM technique was used to process the rainfall prediction. Because it is obvious that it is more effective to integrate with the ML algorithms, it has been used by more researchers in earlier research papers. It is a highly efficient and more reliable technique to get and temporarily store the values during the execution of multiple datasets. A type of artificial neural network called LSTM is popular in deep learning and machine learning. It includes classification, clustering, processing, and time series-related prediction algorithms. The ability to regularly track historical data is a strong suit of LSTM. The model will then be updated using long-term dependencies. It will be referred to as an RNN (Recurrent Neural Network).

Processed Data will automatically send to the Classifier Techniques by the Model. It was comprised of the 6 main classifier Techniques including Random Forest, SVM, Logistic regression, Naïve Bayes, and XGBoost.

An example of the LSTM Model is displayed following. Some of the Variables that are used to develop the model are mentioned here.

- $x_t \in R_d$ - input variable vector to proposed LSTM unit
- The input vector is a combination of the following data
  - X_1 - Temperature
  - X_2 - Humidity
  - X_3 - Rainfalls
  - X_4 - Production
  - X_5 – Land
- $f_t \in (0, 1)_h$ - This variable is known as the forget gate's activation vector
  - $i_t \in (0, 1)_h$ - it is widely known as the input or update gate's activation vector
  - $O_t \in (0, 1)_h$ - activation vector of the output gate
  - $h_t \in (-1, 1)_h$ This function or set is recognized as the hidden state vector. Apart from that, it is known as the LSTM Unit of the output vector.

In addition, statistical software has been used to evaluate the predictive model's precision. The reason is that sometimes there may be many differences shown in the results. These are a few of the statistical packages used for this model because research must determine the best accuracy-based statistical technique for the prediction.

K. Profit Maximization Model using Land Optimization with Crops

Artificial Neural Networks Based Optimization Technique is expected to build to predict the best way to maximize the profit. For example, if a farmer logs into the dashboard and selects his Farmer Type, Region, and Size of the Land he owned to cultivate Proposed System Executing all models and predicts the best combination of crops and how many lands should be allocated for the desired crops to maximize the profits. This option is highly useful to the research because farmers can decide what sort of crops can be easily cultivated together and what sort of percentage should use for respective crops. Then it will be reasoned to make variety between the different parts of the country and cultivate other crops rather than cultivating the same crop. Then the market price will not be decreased and it will stabilize in the satisfactory price range due to the optimized plantation. This is one of the main models that should be developed very carefully as it is a highly sensitive thing for the farmers. If there is a failure whole season will be destroyed. Because the accuracy of the model will be important to this model also.

L. Hardware & Software Requirement

Hardware	Software
PC/Laptop Windows 7 or Higher Operating System Neural Networks	Python 3 PyQT Anaconda ScikitLearn, Numpy, SciPy, and other ML Libraries Tensorflow and Keras

Table 2: Hardware and Software Requirements

IV. Results and Discussion

This section goes into more detail about the findings of the study and the outputs of the programs. There are numerous ways to divide up the results. These include forecasting for temperature, humidity, precipitation, the best crops, and land use. Results are shown in two additional sections after that. These screenshots of a functioning system display and statistical accuracy output. In this section, the acf and pacf graphs of the pertinent dataset are first described. These graphs are used to determine whether the dataset is appropriate for the ML process.

A. Statistical Results

1) Temperature Dataset – Autocorrelation Graph and Partial Autocorrelation Graph

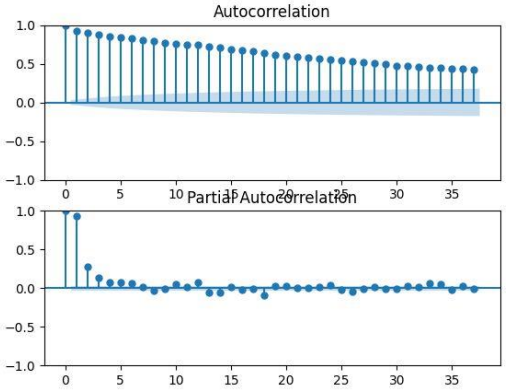


Figure 4: Acf and Pacf Graphs for Temperature

These autocorrelation graphs are used to test the suitability of the dataset for implementation. Also, the Partial Autocorrelation graph shows the behavior and distribution of the data.

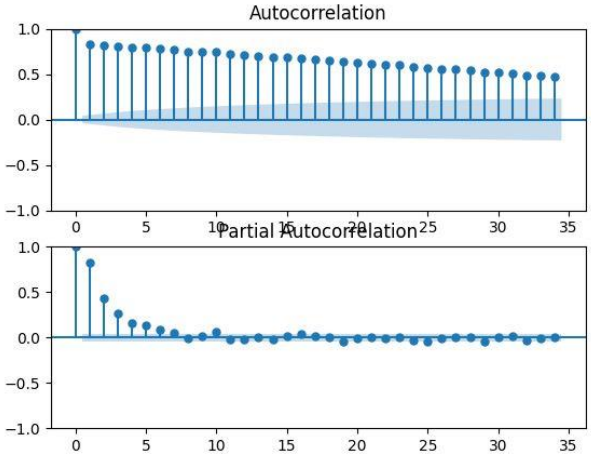


Figure 5: Acf and Pacf Graphs for Humidity

The above two graphs are belonging to the Humidity Dataset and its suitability for development.

2) Predicted Graph for Temperature

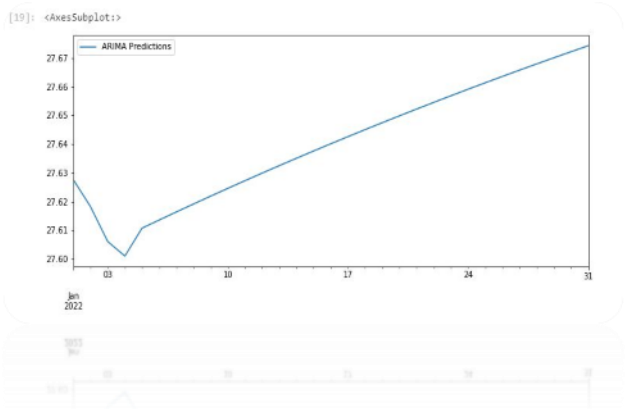


Figure 6: Temperature Prediction Values Graph

These are the Temperature variation graph for the Colombo District – January 2022 month.

3) Rainfall value Variation

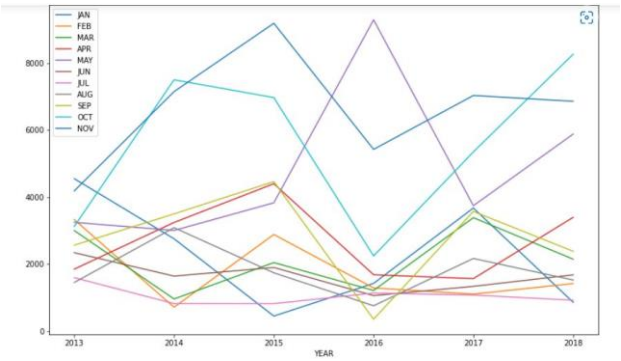


Figure 7: Rainfall Variation Graph

This graph showing how the difference between actual and predicted fluctuations for the Rainfall dataset.

4) Rainfall Prediction (Actual Values and Predicted Values)

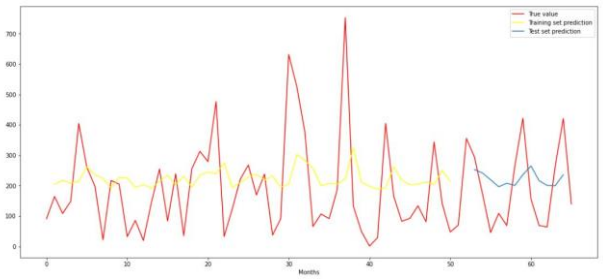


Figure 8: Rainfall prediction Values Graph

5) Rainfall Histogram

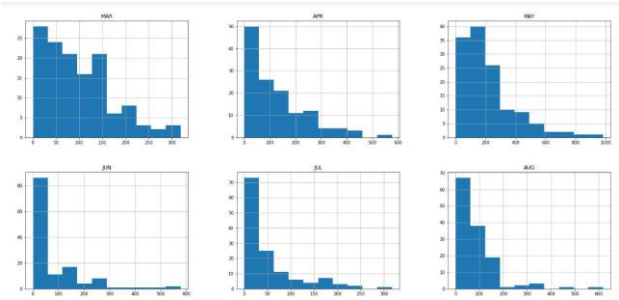


Figure 9: Rainfall Histogram

6) Error/Loss Graph for LSTM Model

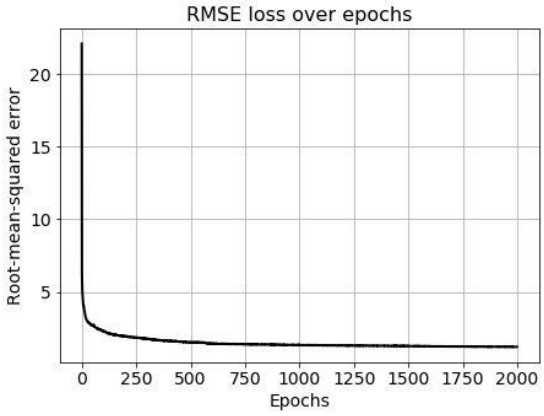


Figure 10: Error and Loss Graph for LSTM Model

This graph is incredibly helpful for determining the function's error rate. To decrease error and boost accuracy, this graph could be used to compare the training and testing loss graphs.
Several methods can be used to gauge model accuracy. Using the Symmetric Mean Absolute Percentage Error (SMAP) and the Root Mean Square Error, the author intends to assess the model's accuracy. SMAP has a greater sensitivity to smaller observations.

$$RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(y_{t-}\hat{y}_t)^2} \quad (4)$$

n = number of periods
t = time period
y_t = Prediction value of (temperature, rainfall, crop)
$y\hat{}_t$ = Actual value of (temperature, rainfall, crop)

$$SMAPE = \frac{100\%}{n}\sum_{t=1}^{n}\frac{|y_{t-}\hat{y}_t|}{(|y_t| + |\hat{Y}_t|)/2} \quad (5)$$

B. Statistical Results of the Model

1) LSTM Processing Error Results

Feature	Criteria	Error Rate
Rainfall Prediction	RMSE	29.4
	SMAPE	26.6%

Table 3: LSTM Error Rate

These are the results of the Error checking using RMSE and SMAPE Techniques. It will clearly show that the error rate is in between the satisfactory range for the accuracy of prediction.

2) Statistical Algorithms Error Rates

Feature	Technique	DT	NB	SVM	LR	RF	XG
Best Crop Selection	RMSE	29.5	27.2	28.1	30.5	18.1	24.2
	SMAPE	26.7%	25.8%	26.6%	27.6%	15.4%	17.9%

Figure 11: Statistical Algorithm Error Rates

DT – Decision Trees
NB – Naïve Bayes
SVM – Support Vector Machine
LR – Logistic Regression
RF – Random Forest
XG – XGBoost

The goal of this research was to determine the most effective algorithm and method for predicting the best crop and land optimization. The research used the statistical methods mentioned above to accomplish that. The above table lists every result that the error-checking analysis produced. The results above show that Random Forest has the lowest error

rate of all the classifiers. Additionally, the majority of classifiers produce the same outcome for the best crop prediction.

The goal of this research was to determine the most effective algorithm and method for predicting the best crop and land optimization. The research used the statistical methods mentioned above to accomplish that. The above table lists every result that the error-checking analysis produced. The results above show that Random Forest has the lowest error rate of all the classifiers. Additionally, the majority of classifiers produce the same outcome for the best crop prediction.

C. System Screenshots of Results

1) Production Prediction

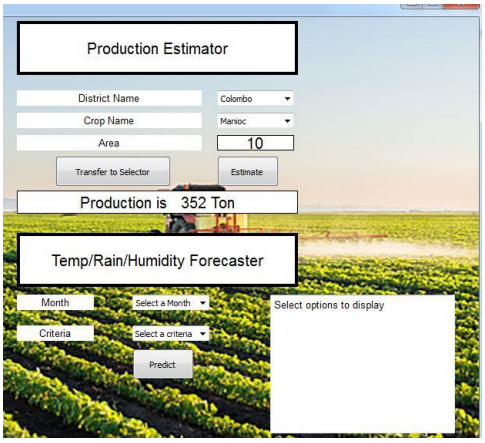


Figure 12: Production Prediction Dashboard

2) Best Crop Prediction and Land Optimization

This is the main aim of the study, and to achieve it, a variety of important issues and traits must be addressed. This is the Best Crop Selector Dashboard's main user interface. It's made with the PyQT package. Text boxes for temperature, humidity, rainfall, extent, expected output, and expected cost are included.
The best crop forecast for each district is just one aspect of this research. Research is concentrating on improving its research capability for novelty purposes. That implies that it is possible to cultivate some crops alongside other crops. Because farmers have the choice to grow the two best crops in order to maximize profit from the plantation if that is a possibility. These are some images from the Land Optimization category.

Figure 13: Best Crop Selection and Land Optimization for Colombo District

This image shows the dashboard. Additionally, it expresses in plain terms how the best two crops are anticipated using the suggested model. The user had Colombo as their district and Maha Season as their preferred season when the aforementioned screenshot was taken. Mangoes and green chilies are the ideal crops for Colombo District during the Maha Season, according to the algorithm. These values are selected after analyzing all the factors of temperature, humidity, rainfall, extent, expected cost, and expected production. (This screenshot shows the best crop selection along with the part about optimizing the use of the land.) Farmers can choose the region and the season that best suits their needs. Farmers only need to enter the season and district here. The predicted temperature, rainfall, and humidity were then displayed after pressing the Load the Data button. Production can then be input manually or based on an estimator's predicted guess. Cost is the price of producing 1 kilogram of that particular crop. The Colombo District data is primarily displayed in the screenshot above. Linear programming was used to develop the optimizer component. And it demonstrates that green chilies should be grown on 90% of the land, while manioc should be grown on 10% of the land

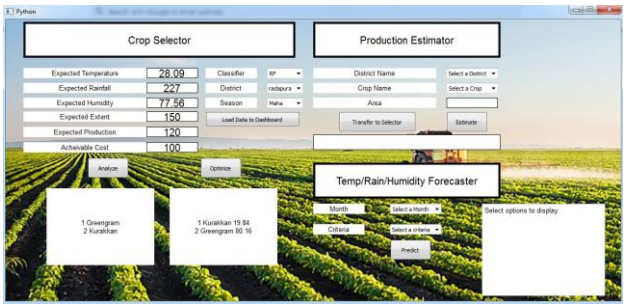


Figure 14: Best Crop Selection and Land Optimization for Anuradhapura District

This graph unequivocally demonstrates how the district change has altered the best crop results. Anuradhapura was chosen as the new district, and the Maha Season is still in effect. The dashboard is showing the top two crops, Greengram and Kurakkan, in accordance with the algorithm. Kurakkan should cultivate 19.84% of the land, while Greengram should cultivate 80.16 percent of the land, according to the optimizer.

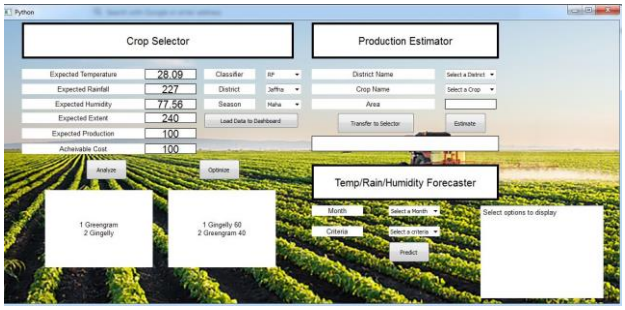


Figure 15: Best Crop Selection and Land optimization for Jaffna District

Another significant and well-known district in Sri Lanka is Jaffna. Farmers in Jaffna always carry on their farming without any interruptions, even during the country's heavy rain seasons in other regions. The fact that many vegetables and other crops are native to Jaffna is another very significant fact. They always deliver their harvest to markets in Dambulla, Colombo, and other places. The graph shows that the best products or crops for Jaffna farmers to grow and earn the most money are ginseng and green gram. This prediction comes true as expected when looking at the Northern Provinces' oil usage and production. The reason is that people in the North frequently consume food and products that are related to oil due to the ease with which farmers can sell their manufactured goods and crops to consumers. Also, this is due to the accuracy of predictions when compared to actual circumstances. According to the optimized result, the farmer should cultivate Greengram on 40% of the provided land and Gingelly on 60% of it.

Information about the Nuwaraeliya District is shown in the following figure. Kurakkan and sweet potatoes are the two best crops to grow in the area, evidence suggests. Most of it is accurate. The explanation is that sweet potatoes are very effective in cold climates. The Optimizer Program suggests that Kurakkan should be grown in 70.76% of the area and sweet potatoes in 29.24% of the area.

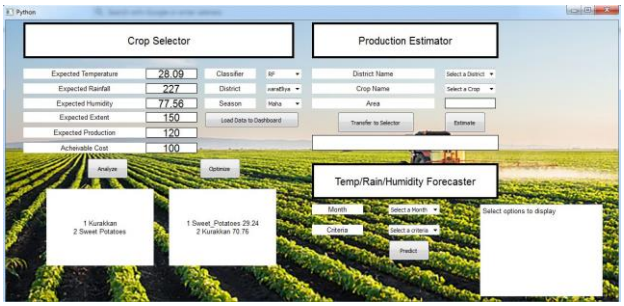
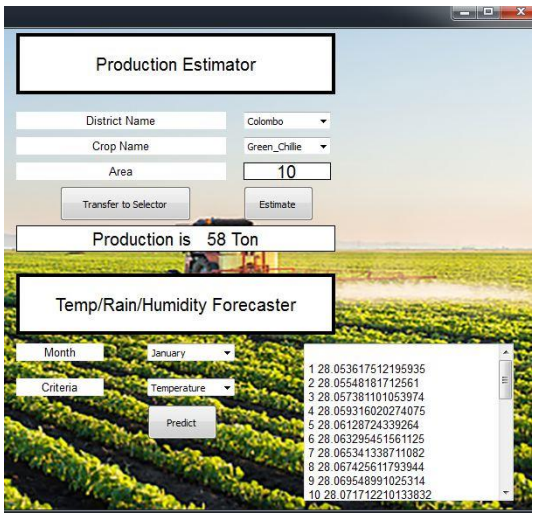


Figure 16: Best Crop Selection and Land Optimization for the Nuwara Eliya District



3) Temperature/ Prediction

Figure 17: Temperature Prediction Dashboard

4) Humidity Prediction

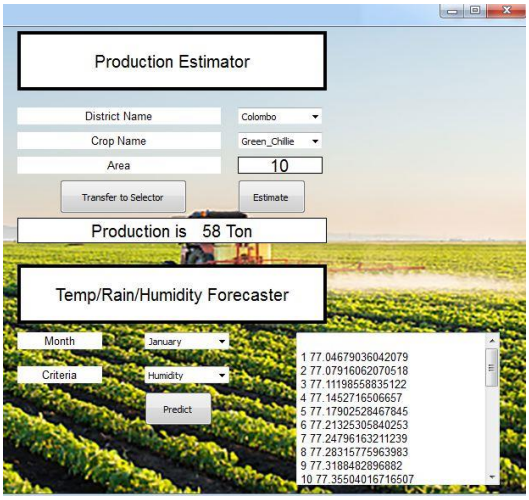


Figure 18: Humidity Prediction Dashboard

D. Testing & Validation

The testing phase cannot be avoided when it comes to software engineering and computer science. The reason is that software development never assures that the final product will be free of errors and bugs. Carrying out a software testing cycle and validation phase is absolutely necessary. Similarly, projects involving computer science or information technology require a testing and validation phase to determine whether the final output is accurate or inaccurate. The methods used to verify the accuracy of the final output or result are then the main focus of this section. This means that the best two crop names for the different districts in Sri Lanka were produced as the research's final product. Users or developers are then unable to determine whether this response is correct or incorrect. The agricultural community should confirm the reason. They are the ones who interact with farms, plantations, and paddy fields the most.

Technique 1

It is essential to test the results using the accepted procedure for this study. The testing phase will be conducted using one of two techniques. Testing on the second dataset is the first technique. This indicates that the data in question is old. The outcomes for the following year can then be predicted using a specific set of years. The following year's data is already available. That implies that it will be necessary to forecast the best crop for 2019. The dataset already contains information on crop production for 2019. The Results can then be easily compared to determine whether they are correct or incorrect.

Technique 2

Getting input from the farmer community is the next technique. The differences between the Real Farmer's Ideas and the predicted ideas can then be analyzed. To achieve this, the research employs telephone interviews to gather the Best Two Crops for each District from the perspective and experience of Farmers. These are a few of the results of the feedback in chart form. This graph was created to show how closely farmers agree with the Proposed Model's predictions.

1) Colombo District

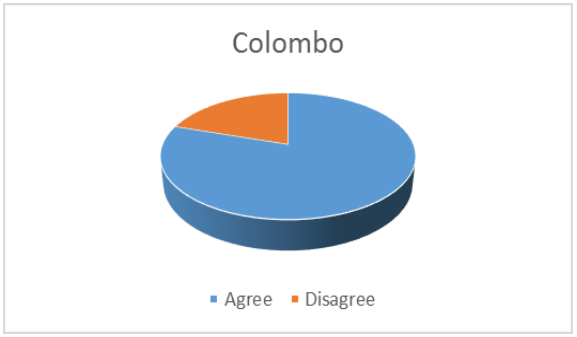


Figure 19: Colombo District Feedback

2) Nuwara Eliya District

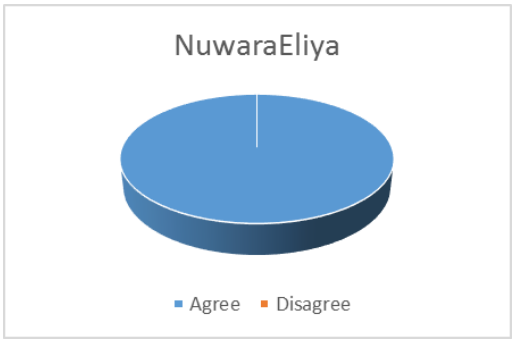


Figure 20: NuwaraEliya District Feedback

3) Anuradhapura District

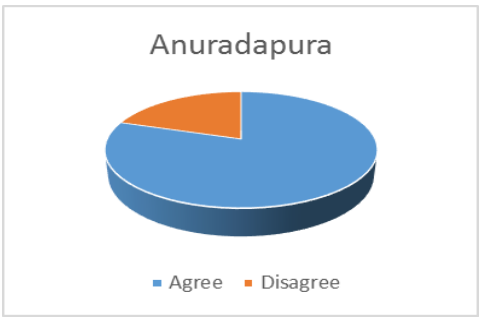


Figure 21: Anuradhapura District Feedback

4) Polonnaruwa District

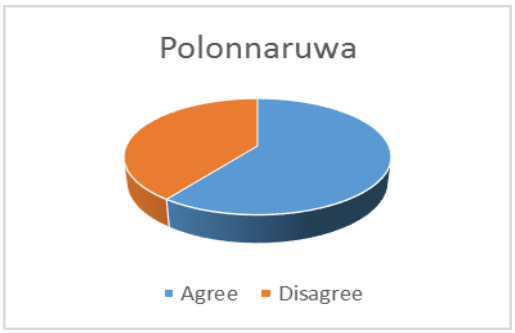


Figure 22: Polonnaruwa District Feedback

5) Hambantota District

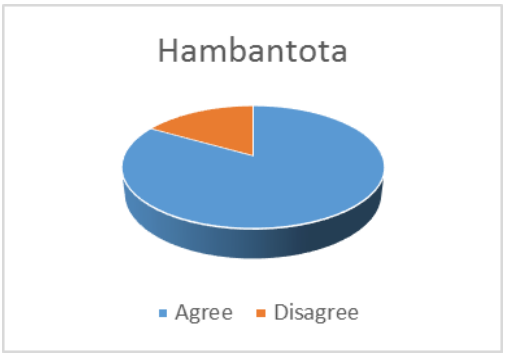


Figure 23: Hambantota District Feedback

6) Jaffna District

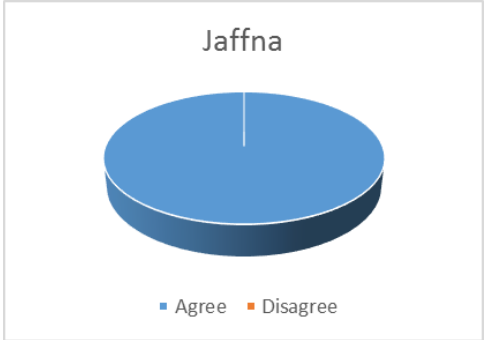


Figure 24: Jaffna District Feedback

This technique is comprised of taking feedback from the Agricultural Experts in the field. I have taken advice from the 2 Agricultural Experts to enhance the accuracy of the Research. Then First get Information from the Agricultural Expert during the Development of the Project. Resources Persons are

- ❖ Agriculture Research Officer. Department of Agriculture, Sri Lanka
- ❖ Subject Matter Officer (Paddy & Home Garden), Department Of Agriculture, Ratnapura

The Feedback taken from the Experts in the Industry also showing the predicted crops and optimization criteria is suitable to deploy in the real world. And they requested to expand this research to more crop types in the future.

V. Conclusion

The primary goal of this research is to analyze the current issues and difficulties facing Sri Lankan agriculture and plantations. The difficulty in obtaining the greatest profit as a result of improper cultivation organization is, in accordance with the analysis, the most significant issue. That indicates that all farmers have chosen to grow the crop that has yielded the highest profit. However, that was a justification for lowering the profit and market price. The best answer to forecasting the best crop will be provided by this proposed machine learning model. A method to effectively optimize farmers' land area is also presented by research. And numerous statistical models have been used to integrate and validate the models. The secondary data testing methodology was used to validate the entire model. Although the farmers' opinions were solicited to confirm the findings. Due to this, it was very advantageous for Farmers to use this model to foretell the best two crops to cultivate and the percentages they should use for each crop with land optimization in order to obtain the highest possible profit from the market.

VI. References

[1] F. Ceballos, S. Kannan and B. Kramer, Crop prices, farm incomes, and food security during the COVID-19 pandemic in India: Phone-based producer survey evidence from Haryana State, Wiley, 2021.

[2] T. Pettinger, "What causes price fluctuations in agricultural markets?," 2017. [Online].

[3] M. B. Ranathilaka and B. Andri, "Market Efficiency on Vegetable Commodities in Developing Country: Case Study from Dambulla Wholesale Market in Sri Lanka," *Science Alert*, 2017.

[4] S. Khaki, "Crop Yield Prediction Using Deep Neural Networks," *Frontier*, 2019.

[5] T. Selvanayagam, S. Suganya, P. Palendrarajah, M. P. Manogarathash, A. Gamage and D. Kasthurirathna, "Agro-Genius: Crop Prediction using Machine Learning," *International Journal of Innovative Science and Research Technology*, 2019.

[6] U. Moiz and I. Hussaina, "Prediction of Wheat Production Using Machine Learning Algorithms in northern areas of Pakistan," *ELSIVIER*, pp. 120-270, 2022.

[7] W. M. Brown, A. Semin, M. Hebenstreit, S. Khvostov, K. Raman and S. J. Plimpton, "Increasing molecular dynamics simulation rates with an 8-fold increase in electrical power efficiency," *SC '16 Proc. of the Int. Conf. for High Perf. Computing, Networking, Storage and Analysis*, 2016.

[8] A. Vaskuri, H. Baumgartner, P. Karha, G. Andor and E. Ikonen, "Modeling the spectral shape of InGaAlP-based red light-emitting diodes," *Journal of Applied Physics*, 2015.

[9] A. K. Tripathy, J. Adinarayana, D. Sudharshan, K. Vijayalakshmi, S. N. Merchant and U. B. Desai, "Data Mining and Wireless Sensor Network for Groundnut Pest/Disease Interaction and Predictions - A Preliminary Study," *International Journal of Computer Information Systems and Industrial Management Applications*, vol. 5, pp. 427-437, 2013.

[10] D. James and R. Hans, "Agro-Selection," 05 05 2001. [Online]. Available: https://en.wikipedia.org/wiki/Portal:Science. [Accessed 19 04 2022].

[11] T. Shakoor, K. Rahman, S. N. Rayta and A. Chakraborty, "Agricultural Production Output Prediction Using Supervised Machine Learning Techniques," *IEEEXPlore*, 2017.

[12] S. Nosratabadi, S. Ardabili and C. Mako, "Prediction of Food Production Using Machine Learning Algorithms of Multilayer Perceptron and ANFIS," *MDPI*, 2021.

[13] T. Klompenburg, A. Kassahun and C. Catal, "Crop yield prediction using machine learning: A systematic literature review," *Computers and Electronics in Agriculture*, vol. 177, 2020.

[14] K. T. Thomas, S. Varsha and J. Thomas, "Crop Prediction Using Machine Learning," *International Journal of Future Generation Communication and Networking*, vol. 13, 2020.

[15] S. Veenadhari, B. Misra and C. D. Singh, "Machine learning approach for forecasting crop yield based on climatic parameters," *IEEEXPlore*, 2014.

[16] S. B. Kamatchi and R. Parvathi, "CROP-CLIMATE MANAGEMENT: A REVIEW," *Journal of Critical Review*, vol. 07, no. 06, 2020.

[17] M. Champaneri, C. Chandvidkar, D. Chachpara and M. Rathod, "CROP YIELD PREDICTION USING MACHINE LEARNING," *IEEEXPlore*, 2020.

[18] A. Pande, V. Pagare, P. Palle and D. Dalgade, "Crop Yield Prediction based on climatic Parameters," *International Research Journal of Engineering and Technology*, vol. 06, 2019.

[19] P. Samuel, B. Sahithi, T. Saheli and D. Ramanika, "Crop Price PredictionSystem using Machine learning Algorithms," *Journal of Software Engineering and Simulation*, 2020.

[20] R. Vishnu and A. Kishor, "Crop Price Prediction and Forecasting System using Supervised Machine Learning Algorithm," *International Journal of Advanced Research in Computer and Communication Engineering,* vol. 09, 2020.

[21] H. Guo, A. Woodruff and A. Yadav, "Improving Lives of Indebted Farmers Using Deep Learning: Predicting Agricultural Produce Prices Using Convolutional Neural Networks," *ResearchGate,* 2020.

[22] A. Suruliandi and G. Mariammal, "Crop prediction based on soil and environmental characteristics using feature selection techniques," *Mathematical and Computer Modelling of Dynamical Systems,* 2021.

[23] I. Jariya and M. Hasan, "ANALYSIS OF BALANCE OF TRADE: SRI LANKAN CONTEXT," *Journal of Management, SEU,* 2018.

[24] N. Dissanayaka and M. Thibbotuwana, "SRI LANKA'S AGRI- FOOD TRADE:STRUCTURE, OPPORTUNITIES,CHALLENGES & IMPACTS OF COVID-19," *Feed the Future,* 2021.

Author Biographies

Heshan Alagalla is a Lecturer who is working at the National Apprenticeship and Industrial Training Authority in Sri Lanka. He is qualified with bachelor's degree in Computer Science and Electronics. He has also obtained Master of Science degree in Information Technology. His research involves machine learning, big data analytics, blockchain technology and robotics.

Lokesha Weerasinghe has B.Sc degree and MSc degree in the field of Information Technology and Computer Science. Her major expertize areas are artificial intelligence, machine learning, image processing and computer vision. She is a Lecturer who is currently working in the Faculty of Computing at Sri Lanka Institute of Information Technology