E-COMMERCE PLATFORM FOR FARMERS

Final Report 2021_135

Jayaraman Rajagopalasarma IT16037984

Keeyani Sivagnanam IT18179118

Lathusha Sritharan IT18181784

Manikavasakar Anjanan IT18186048

B.Sc. (Hons) Degree in Information Technology
Faculty of computing

Sri Lanka Institute of Information Technology
Sri Lanka

October 2021

E-COMMERCE PLATFORM FOR FARMERS

Jayaraman Rajagopalasarma IT16037984

Keeyani Sivagnanam IT18179118

Lathusha Sritharan IT18181784

Manikavasakar Anjanan IT18186048

Faculty of computing

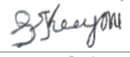
Sri Lanka Institute of Information Technology
Sri Lanka

October 2021

DECLARATION

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

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

Student Name	Student ID	Signature
Jayaraman Rajagopalasarma	IT16037984	
Keeyani Sivagnanam	IT18179118	
Lathusha Sritharan	IT18181784	
Manikavasakar Anjanan	IT18186048	

The above candidate is carrying out research for the undergraduate Dissertation under my supervision.

…….. ……..

Signature of the Supervisor: Date:

ABSTRACT

Human body uses water in every cells, organs and tissues to help regulate body functions and temperature. It is vital that the water used by the body organs are of good quality. In Sri Lanka, reportedly 59.4% of population depends on water from natural sources which grabs the attention to make sure that these people are receiving and dealing with a safe water with a good quality for the usage. Although government is taking necessary action to provide a better quality of water, there has been always a need for a better educational session to educate people about the

importance of maintaining water quality, importance of using a better-quality water and necessary precautions to be taken to avoid any health hazards. Taking this issue into consideration, it is mainly recognized that a smart solution must be implemented in order to solve the identification of water quality problem. E-Tongue: a smart device to predict safe consumption of ground water is an attempt to assist any kind of users to identify the water quality of a groundwater sample in real time by analyzing the water quality parameters to predict the Water Quality Index. This task is achieved by designing a hardware device that embeds a set of sensors to read the value of water quality parameters which will be then transferred to cloud environment for an easy access by the mathematical model to process and identify the WQI value. It will then predict the water parameter readings that could be changed in future along with any disease outbreak possibilities. All the outputs will be finally displayed through mobile application.

Keywords: *water quality, Water Quality Index (WQI), hardware device, mobile application.*

ACKNOWLEDGEMENT

First and foremost, we would like to express our sincere gratitude to our supervisor Ms. Anjali Gamage for the constant guidance and support which helped us at all times for the successful completion of us undergraduate research. Besides our supervisor, our sincere thanks also go to Mr Dammika De Silva, the co-supervisor of this research project for being willing to help whenever it was needed.

Table of Contents

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

Abbreviations	Description
ML	Machine Learning
AI	Artificial Intelligence
IoT	Internet of Things
NLP	Natural Language Processing
LSTM	Long-Short-Term-Memory
COD	Chemical Oxygen Demand
SDLC	Software Development Life Cycle

1. INTRODUCTION

1.1 Introduction

Figure 1: 1 E-commerce Platform

1.2 Background Literature

In Wu, W., L, Zhou & Wu, Y., Z paper ‘’A New Architecture in Retrieval-based Chatbots for Multi-turn Response Selection” in 2016. In here from retrieval chatbots they are looking at multiturn speech response options. Existing work combines presentations in the set or pairs the discourses with a unique setup vector at the end, which can be lost by organizations between important key details or lectures. They suggest successive planning to address both of these issues.

SMN Distributes critical planning data from each link such as a convolution vector and integration functions and which is begins by illustrating the response to each statement within the set at different levels of granularity. In the orchestrated in the order of dreary neural arrangements show the correlation between pronunciation collected by the victors. Experimental assumptions with two open data sets suggest that SMN can bypass the preferred technology to respond to multiple-time chats. The final linking school is included in the RNN protected areas.

Research contains ***“Design, implementation of a chatbot in the context of customer support”*** in 2018. Which is starting from a subjective definition of the chatbot’s targets, a set of client issues, goals were laid down on research paper in chapters a versatile computer program arrangement was planned and created for preparing and conveying chatbots. This arrangement makes it conceivable to extricate ticket information powerfully from an existing Zendesk environment and name them utilizing labels connected to the tickets as showed up in a chapter. In expansion, the chatbot concept was rotted into a few modules which light up a particular plan issue.

Of those modules, a few of them make utilize neural organize models. These models' structure was legitimately analyzed in a chapter.

Additionally, it clears out the entryway open for extra languages' back and client issues. It isn't a come full circle arrangement but considering the necessities, it works well in the hone. Those are computer programs that reenact a genuine discussion between people to answer questions or do errands, giving the impression that the individual is communicating to somebody else and not with a computer program. For rural purposes, it is vital that the data around field conditions, such as discuss and soil temperature, discuss relative mugginess, soil dampness, precipitation, wind speed, and other related factors, be quick and effortlessly accessible for utilize by developing organization frameworks, by masters, or the agriculturist itself in decision-making forms.

The chatbot was made centered on the look and show of data gotten from a Remote Sensor Organize conveyed on a vineyard. It can urge to data collected by EKO field sensors, based on Wire Bot API, getting user experience after an interaction over the Wire application. The IBM Watson cognition organization organize was additionally utilized for moving forward the client involvement by enabling the utilize of characteristic dialect amid the discussion encounter, giving deliberate location.

Encourage improvements are organized for Chatbot, such as the expansion to other informing stages, the execution of discourse communication capacity, picture classification, and nonstop information examination. It gets to be conceivable to work towards the evasion of damaging circumstances to rural preparations. and trusted that with interpretive capacity over the mass of available information, an early revelation of maladies in crops, vitality, and water squander lessening, progressed organization capabilities for the agriculturist.

1.2.1 background about device

This is an enormous water amount and if 1% of water misfortune is spared it will be adequate to give water to 20,000 families [p1]. The related cost that the nation could be spared will be roughly Rs. 3,000 million every year. This sparing is adequate to build 1

km of interstates for every year. Along these lines, if more levels of water misfortunes can be wiped out the advantages are so high. Supply of consumable water is a costly issue and accordingly, the capital speculation required for the arrangement of consumable water surpasses Rs. 175,000 to 300,000 for each family. Consequently, the measure of cash required for water supply is enormous and this has gained the ground of providing water slower than usual. Simultaneously if the water created is utilized for expected purposes, without wastage and ill-advised utilization, the inclusion of the population can be expanded without going for additional speculations [p1].

Sri Lanka is an agriculture country. A relatively huge amount comparing to the total population in Sri Lanka are farmers. 27.1 percent of Sri Lanka's population is involved in agricultural pursuits. In 2020, agriculture contributed for 7.4% of GDP (gross domestic product). Sri Lanka is mostly an agricultural country, with a concentration on vegetable farming. Each leaf should be free of pests and illnesses in order to produce healthy plants. Farmers may have trouble cultivating plants if they do not have adequate understanding of the corresponding pesticides. The majority of Sri Lankan farmers have little knowledge about plant management. Plant diseases are one of the major problems which farmers are facing today. Many experience farmers know about diseases and how to manage them. But today many new Youngers are involved in farming. Also well educated people are interested in farming and home gardening. In this situation new comers to farming don't know about plant diseases and how to use pesticide management to particular diseases. So they need someone who has knowledge about plant diseases and pesticides management to treat the affected plants. To fix the situation, they must meet and seek guidance from an agriculture department officer or an agricultural specialist. Farmers must demonstrate sample-affected leaf to the experts, which takes time. Farmers must thus apply proper pesticides; otherwise, they may encounter challenges such as decreased agricultural production quality and quantity. it is time consuming. To control above problems, pesticides functions will be implement. Recognize the indications of insect attack and nutritional shortages in the leaves and find a remedy early on will assist to reduce the impact of faults in the plant. When symptoms develop, samples of the afflicted plant parts are gathered and analyzed using the conventional method, which is time consuming, labor

involved, and costly. Therefore, a system has been implemented to detecting plant diseases and suggests pesticides in this research project.

The system's goal is to provide precise and scalable visual cues to identify Vegetables plant diseases, particularly pumpkin, tomato, chili plants diseases, and to provide a solution for farmers by providing exact information about diseases, pests, nutrient imbalance, and how to apply fertilizer and pesticides effectively. Detection of disease is achieved by image processing. Here, some technologies are used to data training. Machine Learning and Deep Learning which are the branches of Artificial Intelligence. Deep Learning is also type of Machine Learning. Deep learning is modeled using architecture modeled after the human brain. Where as in machine learning the human involvement is minimized most of the time. Deep Learning is used for data training and testing. And, in this research TensorFlow library, Keras library and NumPy library are used in python language. TensorFlow combines a variety of techniques and models to allow users to build deep neural networks for applications such as classifying as well as identification of image. The models generated by the Image Data Genetor in Keras Framework may be augmented using image data.. Image data augmentation is used to increase the size of the training dataset in order to improve the model's performance and generalization capabilities. In order to do numeral computation and store structure such as matrix and multi-dimensional array a python library called NumPy is used. This can perform various forms of statistical and algebraic computations.

Convolutional neural network (CNN) algorithm is used in this research. The overall theory of the functions of CNN is that it is able to take a picture as an input, then assign weights and biases to depending on the various aspects of images and identify them. Only a subarea of the preceding layer's input is received by neurons in a convolutional layer. Each neuron in a completely linked layer receives information from all elements of the preceding layer. A CNN extracts characteristics from photos and uses them to make decisions. Feature detection is taught to CNNs over tens or hundreds of hidden layers.

1.2.2. Background literature of vegetable price prediction.

Not only the manufactures of agricultural crops but also the other parties involved in a market chain can benefit from the knowledge which can be achieved from the

analysis and forecast which can done using the historical data of variety of vegetables. But there is only a handful number of researches done in analyzing and forecasting. And analyzing of the markets and considering multivariate factors are done in a minimum number, even from these researches which are considered. As timeseries forecasting deeply depends on type of problem than a algorithm which is used.

Kasturi Kanchymalay et el aimed a study in order to find the relationships between the price of crude oil , other vegetable oil selected for this study such as soybean coconut oil, and olive oil, rapseed oil and crude oil price (CPO) monthly exchange rate. After the analysis forecasting was done. The analysis showed a positive relationship between Soya bean oil. Three following three machine learning model were used: support vector regression (SVR), multi-layer perception, and Holt Winter exponential smoothing and SVR showed better results comparing to the others.

Xiong et el [42] forecasted vegetables by considering the seasonal factors of time series data. An Extreme Learning Machines (ELM) was used for this purpose. And they employed a decomposition procedure based on Loess in order to decompose the data. After decomposition they do a univariate forecast of each component such as trend and remainder separated and at last they integrate them in order to get the results. And they have used Diebold Mariano test and to accuracy measures as evaluation metrics. And concluded that STL-ELM method as promising model to forecast vegetables with high seasonality prices range.

As per some studies, it states that univariate models performs better than multivariate forecast, and in some studies multivariate shows better results. Jasleen et el[43] has done a univariate and multivariate forecast on timeseries model in order to analyze air quality. In this study they have used ARIMA for univariate and Vector Autoregression (VAR) for multivariate prediction. And after conducting the experiments they analyzed the model and evaluated them using a number of metrics and at last they have observed that ARIMA model produced better results. But a study done by Tinker et el[44] on forecasting soybean prices , the results showed different result than the latter. In this study.

1.2 Literature Survey

A number of researchers give a brief review of ways to detecting plant illness, nutrient insufficiency, and insect assault using image processing with machine learning in this paper. This section discusses the conclusions made from the earlier researcher's study publications on the subject.

In 1998[1], this study examines and analyzes several techniques for handwritten character recognition using a conventional handwritten digit recognition problem. Convolutional neural networks have been proven to surpass all other approaches when it comes to dealing with the variety of 2D forms. Modeling of language, segmentation of recognition, field extraction,are all components of real-world document recognition systems. Graph transformer networks (GTN), a novel learning paradigm, in order to reduce the overall metric of performance the gradient based techniques which enables multi component modules to be taught all over the world.. Descriptions of 2 online handwriting recognition systems are provided. Global training and graph transformers present benefits according to the experiments done. Also discussion on reading a bank check using graph transformer also has been done. With the combination global training methods and CNN character recognizer the checks which are personal as well as business achieves accuracy on records. It is in commercial use and reads millions of checks every day.

In 2009 [2], the focus of this study is on obtaining paddy characteristics from off-line images. Image capture is followed by automated using a local entropy threshold as a threshold and the Otsu technique to transform RGB pictures to binary images. The noise is removed using a morphological algorithm and the region filling approach. Then, from paddy leaf pictures, image features such as lesion kind, the colors of border, spot and diseased paddy leaf are retrieved. As a result of using the production rule approach, paddy illnesses may be identified with an accuracy rate of 94.7 percent.

In 2020 [3], the study focuses on identifying the patterns in plants which can be useful in finding the diseases using visual elements that are common in the diseased plants.. To have a prolonged healthy agricultural activity, constant care of health and detecting diseases in plants is crucial. But constantly watching over the plant in real life could be tedious task in real life, especially if there is a huge number of plants to be scanned with

our bare eyes and if it happens in action it might take a lot of the valuable time. Also lack of knowledge about plant diseases and how they visually affect a plant make the task of detecting a diseases almost an impossible chore So as a solution to this complication, image processing used to identify the patterns which are similar to the dataset which has pictures of the diseased leaves and the ones which are fed for identifying the disease. The data set contains a varied range of plants in form of a picture. And further the users in this study are led into commercial websites where pesticides are up for sale with their user manual is present. This activity is considered to be beneficial when considering the evaluation of MRPs of pesticides and to cure the disease using it. Also this study gives a hands to those who grow their plants in an green house environment.

In 2016 [4], the objectives of this article is to provide an insight of how image processing technology can be applied to represent various plant diseases. With the use of image processing, the diseases cause by microorganisms such as virus, bacteria and fungus are easily identified. The use of only one's eyes to identify illnesses is ineffective. As pesticides are not cleaned correctly, they create dangerous chronic illnesses in humans. Also when there is an over the limit usage, it can lead to more harms such as loss of quality and also loss in the amount of production. So, image processing methodologies might be useful in agricultural applications for detecting and classifying illnesses.

In 2020 [5], Yallappa D et al presented a basic drone vehicle and improved its performance for spraying pesticides on crop plants. The major experimental investigation was conducted on crops such as groundnut and paddy. BLDC motors, six cells in two lithium polymer batteries with 8000mAh capacity, 5kg payload capacity, a 12v dc motor, a camera for front view, a fluid tank to spray pesticide with a 5 liter capacity, GPS, and other components make up this drone system. The major focus will be on evaluating the performance characteristics of a drone-mounted sprayer as well as the economics of its operation. This technique is utilized in areas where humans cannot readily travel, such as orchard crops. The field efficiency is between 62.84 and 60.00, and the forward speed is at 3.6 km/h1.

In 2019 [6], in his study effort, Santhosh Kumar S et al highlighted the major important concerns and different obstacles that arise when analyzing plant diseases. Various plant diseases, such as rust, yellow leaf, rotting of leaf, and leaf curling, are described in depth,

along with their symptoms and appearance. Using the Gabor wavelet transform and the hybrid clustering approach, mobile-based client-server architecture is utilized to identify leaf disease. The ANN model, a deep learning approach, was utilized to find fungus of different kinds in cucumber plant leaves. It aids in illness categorization and detection. Automation, multi-feature, and genetic algorithms are used to identify plant diseases. The literature review detailed the BP neural network, the method for segmenting plant leaf disease, and the Gabor wavelet transform, concluding that image processing should be the major perspective for disease detection in plants.

In 2017 [7], This study paper provides a brief overview of the most deadly cucumber mildew, Downey mildew, which spreads rapidly in cucumber plants. There is a quick comparison between the human eye assessment of illness and the image analysis approach. A developed downy mildew spot extraction algorithm has been suggested, which employs the leaf scanning approach to determine the illness index. The mildew picture has an accuracy of 98.3 percent.

In 2015 [8], this study offers a computationally efficient approach for detecting paddy leaf disease. Segmenting of images, feature extraction, and classification are the three steps of the suggested methods. The image segmentation technique K-means cluster algorithm is utilized to find diseased leaf sections. The paddy leaf picture is used to extract features in thefeature extraction step. These characteristics are utilized as input to the classifier in order to classify the data. The classifier is utilized as an artificial neural network in this experiment. For many years, numerous researchers have been researching on real-time plant leaf diseases. This project will be used to identify leaf disease in real time in the future. Farmers would benefit greatly from this initiative since it will allow them to detect paddy illnesses at an early stage.

In 2010 [9], the study focuses on a framework which enables the detection of plant leaf diseases. According to studies, depending solely on professional observation with the naked eye to diagnose such illnesses might be too expensive, particularly in underdeveloped nations. Providing quick, automated, low-cost, and accurate image-processing-based solutions for that task can be quite feasible. The suggested framework is image-processing-based and consists of the key phases listed below. The pictures at

hand are segmented using the K-Means approach in the first phase, and then fed through a pre-trained neural network in the second. They use a collection of leaf pictures from Jordan's Al-Ghor region as a testbed. The results of our experiments show that the suggested method can greatly aid in the accurate and automated detection of diseased leaves. The statistical based Neural network classifier was able to detect and also categorize the disease with the accuracy of 93%.

In 2013 [10], this study is structured in away to analyze and categorize, brown pot , leaf blast, bacterial blight and tungro which are the most common Indonesian paddy diseases with the help of fractal descriptors. The pictures of the lesions were manually extracted. Afterwards, the S component descriptors out of each lesion image were used in a stochastic neural network classification technique. When it came to detecting illnesses, this approach had an accuracy of at least 83.00 percent. If coupled with additional characteristics, this approach has the potential to be utilized as one of the features, particularly when two illnesses of similar hue are involved.

In 2018 [11], the main aim of their study is to classify and detect diseases of plants using image of the leaves with the help of software system which consist image processing, by testing and improving the image processing algorithm. This study also suggest that identifying the illness can be consumes a lot of time and expensive as it might require a lot of labor specifically in the rural areas of the countries as it should be done with just the observation of the normal eyes scanning carefully in order to detect a disease. As a result, we offer a method based on image processing that is quick, automated, inexpensive, and accurate. The first process consists of creation of color transformation structure for the data which is the image of the leaf. Second, this structure is changed by the application of color space transformation.

Then the pictures are segmented using the K-means clustering. On the third step, texture characteristics of segmentation of the diseased object is estimated.. On the final stage the received feature a fed into the already trained neural network model.

In this paper [12], Nikita Goel et al. presented a model that uses a trained collection of pomegranate leaf pictures to inform about the illness of the leaf. This study introduces a new computer-assisted segmentation and classification approach. Obtaining image,

picture enhancement, segmentation of image, extracting features, classifying image, and image accuracy are included in the first stage of this article, which aids in pre-processing. The k-means method is employed on all pictures in dataset in the second stage. The feature extraction step, which includes color and form characteristics, is completed in the third stage.

Gina S. Tumang spoke on the most prevalent illnesses that affect mango trees. Anthracnose, fruit borer, and sooty mold are the culprits. The illnesses are detected utilizing image processing techniques such as multi-SVM and GLCM. For image classification, a support vector machine classifier is employed. Image capture, image pre-processing, segmenting using k-means algorithm, feature extraction using gray level co-occurrence matrix, are all utilized. This method of detecting illness in plants was shown to be 85 percent accurate.

In this paper [14], an investigation was conducted by author ananthi, s.vishnovarthini into the categorization of plant leaves and the identification of illnesses that damage them. They focused on controlling pests that cause illnesses that reduce crop output, and they mostly summarized machine learning and image processing approaches such as BPNN, SVM, and KNN.

In 2018 [15], Tomatoes are one of India's most important agricultural crops. In that situation, early illness identification in the plant is critical. The tomato plant diagnosis would be aided by image processing and a prediction model based on IoT sensors., Extreme Learning Machine ,Support Vector Machines, ANN classifier, Deep learning, and other modeling approaches have been created in response to a variety of symptoms and illnesses identified in tomato plants. The majority of the times, pictures are utilized as system input.

1.3 Research Gap

After review of previous research and literature, Below, mentioned research gaps are identified:

- Srilankan farmers facing communication problems while selling and buying farming products.

- Because they don’t have enough platform to easily getting different kinds of Srilankan languages vegetables.
- Sometimes they have an idea about a product and do not know the particular area people using names of that vegetable.
- Investigators also compared specific methods of a chatbot connector. But most farmers do not have enough knowledge to understand those effects.
- There is no current investigation on the comparison of statistical and deep learning model for analyzing and forecasting data for vegetables of Sri Lanka

And agricultural people are not familiar with English typing without spelling mistakes. So, if they use some main keywords of the vegetable features in spelling mistakes, our system will correct those words and suggest the correct output. For this purpose, the NLP model is using. But these features containing platforms are not satisfied to our Srilankan farmers.

Concurring to the writing study over conducted it is obvious that space particular image processing part in e-commerce stage may be a much required implementation since no investigates have been done focusing on that. Up to now within the investigates that was conducted beneath space particular discourse acknowledgment. There have been several studies performed to detect the illness, but the most of them have been connected to paddy fields. In addition, various studies have been conducted on vegetable plants such as tomato and bitter grout. To accomplish the desired outcome, the leaf must be plucked off the plant. Those are also concentrating solely on plant disease. Those who are not offering any remedy for crops to heal illnesses, and some of those who are advocating pesticides. Looking above research papers there are many applications using image processing. But there is no online platform for farmers to identify diseases using image processing. In our research we will build a platform for farmers which help them search pesticide by images. There are numerous types of studies that use image processing to identify crop illnesses and leaves, but this method uses both image processing and deep learning. Disease categorization and image identification are critical technological and economic issues in agriculture. The method primarily focuses on vegetable-type plants. The system will detect

crop leaves, and then it will use deep learning to identify the illness. The system suggests appropriate pesticides based on the illness.

The majority of study is discovering solutions by plucking the plant leaf such that the plant's growth rate and product rate are both reduced. They are not attempting to discover a way to obtain the desired outcome without plucking a leaf off the plant; instead, they should enhance the plant-growing rate and product rate. In this scenario, take a photo using your Android phone and get the results right away.

Price prediction

By using multivariate forecast the problem can be solved. But in order to do so, an appropriate method should be verified. In order to do that, this research studies two methods VAR which is statistical method and LSTM which is a deep learning model.
There is a lack of researches on multivariate forecasting which compares these two models in terms of the Sri Lanka as well as time series forecasting using retail and whole sale prices.

1.4 Research problem

There is no internet platform that helps farmers to communicate. And further they may face difficulties when deciding on price on an online platform. Some farmers do not know which product to grow and which product gives them the most profit. Sometimes what happens is that many farmers grow the same product in one place so the price will go down. Therefore, our research focuses on these problems that a farmer may face.
A large percentage of the country depends on agriculture. Therefore, modern agriculture is very important and thus will lead the farmers of our country to profit. But these days the percentage of total GDP is declining. Below is a specific problem we have seen with this change.

- Inadequate funding of farmers through their pre-determined crop planting program.
- The competition is against smallholder farmers in marketing the same crop at a fair price.

- Lack of a single smart platform to guide farmers on the proper harvesting of crops at a number of concerns.

To overcome the above problem, we plan to launch the app. We hope that this will reduce the problem of rural farmers and promote more profitable crops. It helps farmers to make decision.

The research gap described above gives an idea of the difference between the present features provided by pest and disease control that we are discussing and the benefits gained when the system is developed for testing purposes. As a result, the flowing is a quick summary of the research topic that we expect to answer once this technique is completed.

- This study is being presented in order to create a system that gives solutions for farmers. After identifying the leaf, the system examines it to determine whether or not it has been impacted by pests. The system recognizes the leaf based on the shape of the bitten type as well as the symptoms of the leaf. The system then offers certain insecticides to treat the illness or eliminate the pests based on the results. Pesticides are often used to decrease infection from the damaged leaf. However, certain insecticides are unable to heal the damaged leaf or are ineffective in their function. The system determines whether or not the insecticides used are effective. If it fails to operate properly, the system recommends other insecticides to cure the illness.

- Some farmers spend effort determining the related pest or illness because they must apply pesticides or fertilizer to treat the disease/control insect assault. Farmers in this scenario must seek guidance from agricultural experts. As a result, those farmers are undoubtedly assisting our research and testing efforts.

Vegetable multivariate forecasting.

There is unavailability of forecasted prices available on main market of Sri Lanka.

This research focuses on price analysis and forecast of Sri Lankan market in order to help agricultural participants of Sri Lanka to get an insight. Deep learning and statistical not been used in order to analyze and explore the prices markets of Sri Lanka. Previous literature pays less attention to various markets. So, this research explores this part of vegetable price multivariate timeseries analysis and forecast.

2 Research Objectives

2.1 Main Objectives

Developing an eCommerce platform for farmers is the main objective of this component. This will provide good guidance for farmers and they can get the accurate vegetable name in Tamil, Sinhaha, and English languages. They no need to worry about their English knowledge. Because when they type description spelling mistakenly the system will correct that words and provide correct output.

The primary goal of this project is to improve the livelihood of vegetable plants by identifying specific disease and pest attacks in each phase of the plant's life cycle, recommending organic and inorganic pesticides, fertilizer for the Cucurbita plantation, and acquiring efficiency in the use of fertilizers and pesticides.

- Produce a multivariate forecasting model which performs better analysis and forecast for Sri Lankan market.

2.2 Sub Objectives

- Perform a thorough research on marketing availability which is needed for farmers.

- To answer questions accordingly configure the appropriate classification algorithm.
- Cleaning the dataset and creating a frequent data dictionary.
- Providing output in three languages in English characters.

- Identify plant affected by pest diseases or not.
- Identify a method to increase the efficiency of using pesticide or insecticide
- Taking photo of that affected area.
- Finding disease using image processing
- Suggest the pesticide
- Plot the price series and obtain descriptive measures. Graph enables to make hypothesis which and assumption which can help in our analysis and prediction.
- Observation is done one this objective and variation in one series leads can lead to variation in another series which leads to deeper understanding of the problem.
- Using evaluation metrics select the most suitable model for multivariate forecasting.
- Prediction the future values
- The results are presented through and android app which could be useful to the users of the system.
-

3. METHODOLOGY

Artificial Intelligence is responsible for the development of new products such as private cars and chatbots. The latest developments in Natural Language Processing (NLP) have made conversational discussions, also known as real helpers, a great option to improve the farmer's knowledge. Answering frequently asked questions, filing claims, checking the order status, and getting feedback from farmers are among the most commonly used negotiation cases. Creating a chatbot that provides a good experience for farmers requires collaboration from a team that includes business analysts, service builders, data scientists,

machine learning engineers, and software developers. The chatbot development approach incorporates a wide range of frameworks and approaches including design thinking, new AI sprints, and software development.

A. Data processing before processing data is a method used to convert raw data into a pure data set. Data is collected from a variety of sources, collected in raw format that cannot be analyzed. By using different techniques such as converting lost values and null values, we can convert data into an understandable format. The final step in data expansion is the separation of training and testing data. Details are often unequally classified because model training often requires as much detail as possible. The training database is the first database used to train ML algorithms to read and produce accurate predictions. Figure 1. shows a few lines of data used.

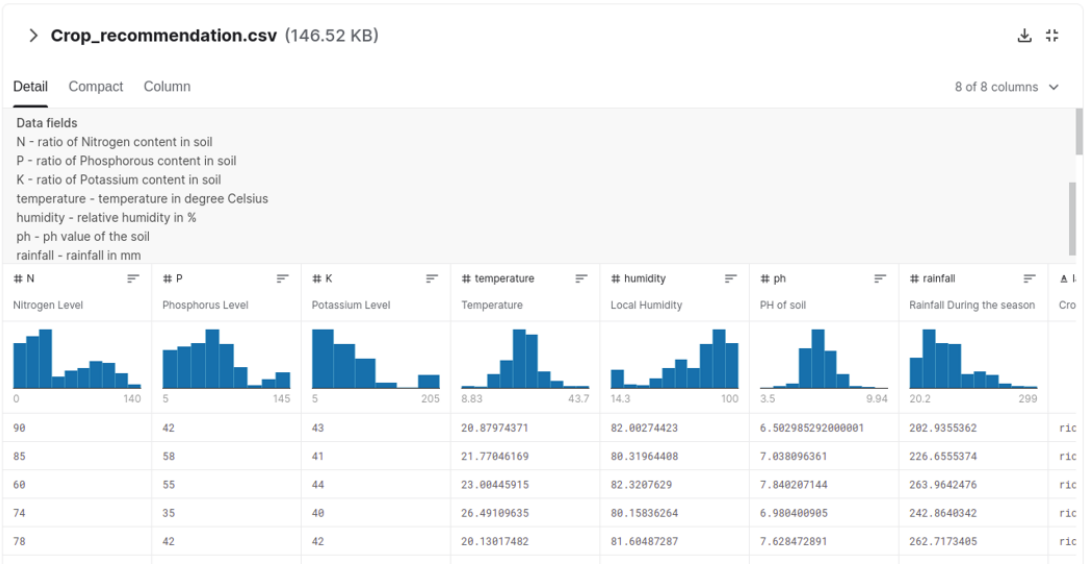


Fig. 3.1. Kaggle dataset

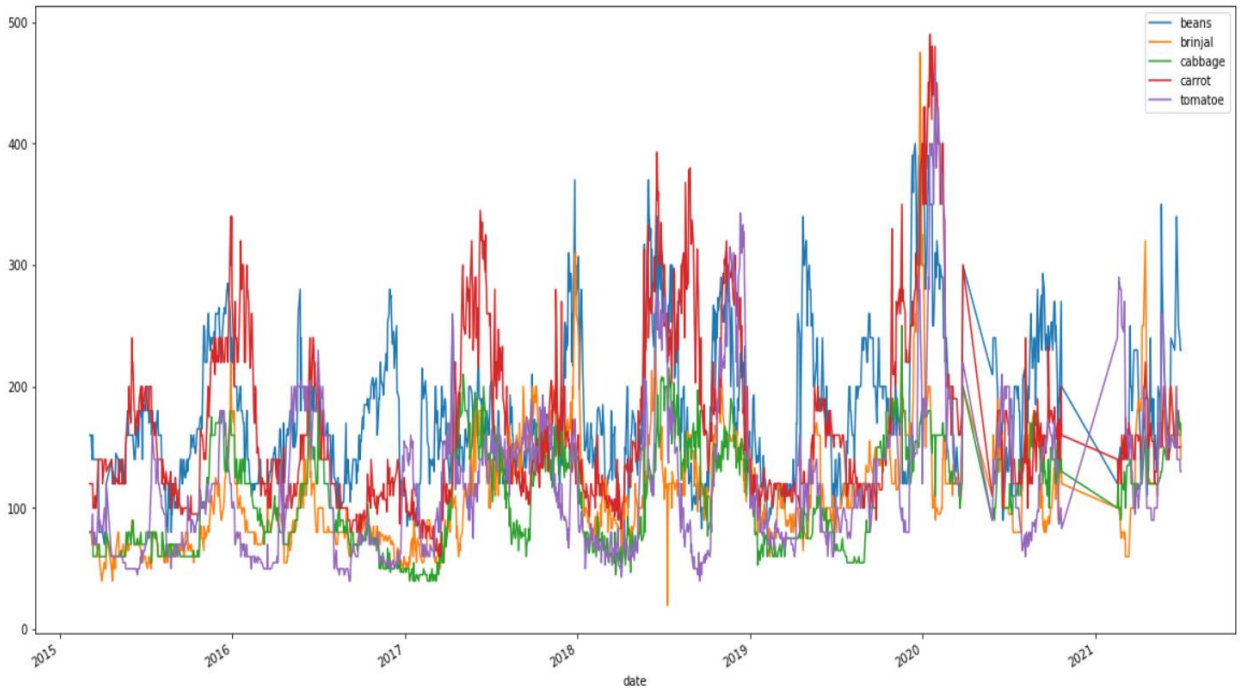


Fig. 3.2. data library of Central bank.

B. Factors affecting crop production and production There are many factors that affect the yield of any crop and its production. These are basically factors that help predict the production of any annual crop. In this paper we include such things as Temperature, Rain, Location, Humidity and PH (Figure 1 shows the predictive properties of a plant name and its calculation)

```
dataset.describe()
```

	N	P	K	temperature	humidity	ph	rainfall
count	2200.000000	2200.000000	2200.000000	2200.000000	2200.000000	2200.000000	2200.000000
mean	50.551818	53.362727	48.149091	25.616244	71.481779	6.469480	103.463655
std	36.917334	32.985883	50.647931	5.063749	22.263812	0.773938	54.958389
min	0.000000	5.000000	5.000000	8.825675	14.258040	3.504752	20.211267
25%	21.000000	28.000000	20.000000	22.769375	60.261953	5.971693	64.551686
50%	37.000000	51.000000	32.000000	25.598693	80.473146	6.425045	94.867624
75%	84.250000	68.000000	49.000000	28.561654	89.948771	6.923643	124.267508
max	140.000000	145.000000	205.000000	43.675493	99.981876	9.935091	298.560117

Fig. 3.3. data set.

Data description for the analysis is given below.

```
df.describe()
```

	pettah_wholesale	pettah_retail	dambulla_wholesale	dambulla_retail
count	1451.000000	1386.000000	1471.000000	1457.000000
mean	81.956582	112.790043	73.330387	93.238847
std	41.594116	45.985117	40.604375	41.895826
min	20.000000	20.000000	9.000000	24.000000
25%	50.000000	80.000000	45.000000	65.000000
50%	73.000000	100.000000	65.000000	85.000000
75%	100.000000	140.000000	95.000000	115.000000
max	400.000000	475.000000	355.000000	375.000000

Fig. 3.4. data description.

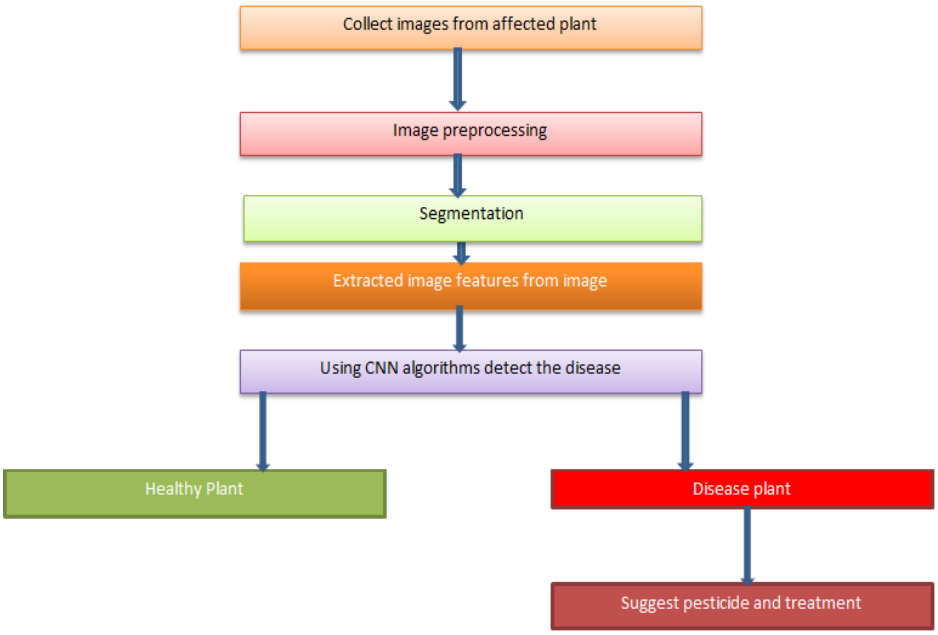


Fig. 3.5. Image Pre-Processing

Image Pre-Processing

Image pre-processing is an important stage in image processing. This technique will remove sounds in the photos, such as hair, clothes, and other artifacts. The major objective of image processing is to remove distracting aspects from damaged plants' images in order to improve image quality**.**

Image Processing

Image processing to identify plant diseases is the most often addressed topic at various stages. To train and test the data set, a combination of DL and ML approaches are utilized. CNN algorithms are used to extract and categorize information from pictures. Mainly, five steps are used to get accurate output.

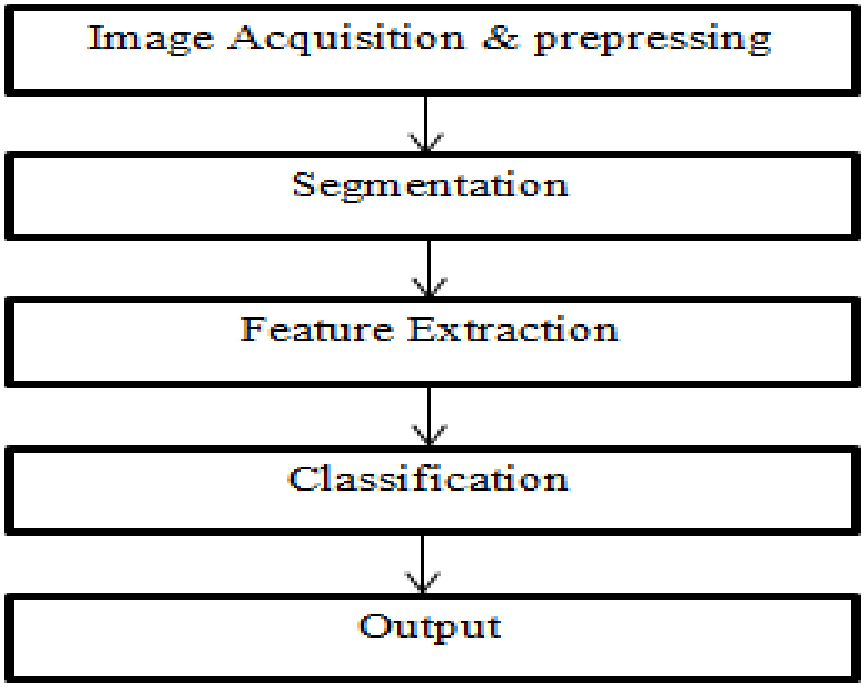


Fig. 3.6. Image processing.

Obtaining images

Images of plant leaves are photographed with professional camera or mobile smart phones in order to achieve the optimal resolution and size of image required for the study. Also the images could acquire from the internet if they satisfy the requirements. The recorded image was preprocessed to improve image quality, such as removing the non-uniform background and resizing the image with the dimensions as RGB training data. To eliminate the noise from the picture, nonlinear spatial filtering is used.

Image Segmentation

This step requires optimizing image in a way to make them viable for analyzation. So, accordingly the representation of the image is reduced. This phase is also the essential method to image processing because it is the foundation of feature extraction. Otsu' algorithm and k-means clustering are among some of the methods in which image may get segmented. Based on a collection of characteristics, the k-means clustering algorithm

divides objects or pixels into K groups. The reduction of total of the squares of distance between items and their groups help in categorizing.

First, impacted leaves will be detected, followed by affected parts and green color pixels. The threshold equation will be used to demonstrate the threshold value of the green color pixels.

$$g(x,y) = \begin{cases} 1, & if\ f(x,y) \geq T \\ 0, & if\ f(x,y) < T \end{cases}$$

Fig. 2.7. K-Means methods.

Enhance the image utilizing K means methods, which result in a high-quality image for detecting nutritional insufficiency. Edge detection technology will be utilized to determine the insect's carrying pattern, which will be used to detect the pest.

Image Feature extraction and Classification

Following segmentation, features from the targeted areas are retrieved using feature extraction techniques. Skewness Asymmetry in the distribution of pixels in the given window around its mean

$$m = \frac{1}{mn} \sum_{i=1}^{m} \sum_{j=1}^{n} p(i,j)$$

$$m = \frac{1}{mn} \sum_{i=1}^{m} \sum_{j=1}^{n} \left(\frac{p(i,j) - m}{\sigma}\right)^3$$

Fig. 3.7. K-Means Algorithm

Pixels in the same neighborhood have different intensities.

$$C = \sum_{i,j} (i-j)^2 P(i,j)$$

Fig. 3.8. K-Means Algorithm formula

The component extracts the image model using a deep learning method and approach after extracting the features from the damaged leaf picture. As a result, the picture model is reliant on an external library. Tensor flow is a toolkit for dealing with machine learning techniques like the convolutional neural network (CNN).

It's used to classify images, such as determining whether or not a leaf is healthy. The neural networks (CNN) method learns a complicated image in small layers, and each image will have many layers. In addition, keras is most likely tensorflow when it comes to deep neural networks. For this study, we collected more than 4000 plant diseases images and healthy images from field and internet for data training and testing.

Models Used for price prediction.

Models Used in the research

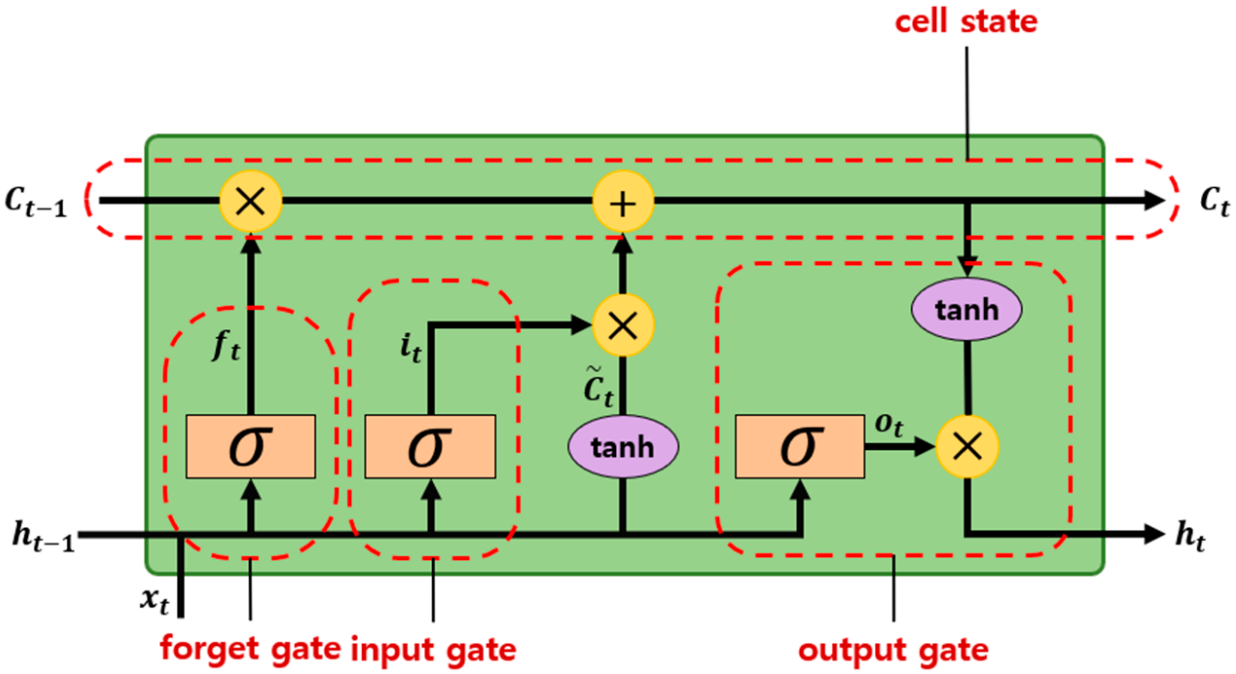


Fig. 3.9. LSTM process

Recurrent neural network(RNN) consist of special type of network called Long short term memory (LSTM). Modeling of language, text recognition, machine translation, and speech recognition are the few fields where RNN have been successfully employed. Solving current stop problems using previous stop information is one of the key features of RNN [11]. But a long tern dependency problem arises when there is a context where there is an increment in the gap of the two different type of data causing the prediction data to become longer. In order to solve this long term dependent and vanishing gradient problem. Hochreiter et el [12], proposed Long Short Term Memory (LSTM). It consists of component named as gates which are namely input gate, forget gate, output gate, and cell gate, all which contained within in a single layer of neural network. The figure above presents the structure as mentioned. The operation behind the forget gate is given in equation 1. The process starts in forget gate as it gets the hidden state of the previous step which is denoted as ht-1 and the input of the current step which is denoted as xt , and with that it execute matrix multiplication with the learnable gate which is denoted bt Wf. After this a learnable forget gate bias value noted as bf is added and the output is received

through the sigmoid function as it produces a value in between zero and one. The main reason for this is that when the ft which is the output value is close to one, more information is saved in the cell state of the previous time step and as it nears zero , information is avoided.

Equation 1

$Ft = \sum(Wf(ht-1,Xt) + bf)$

The Equation 2 and 3 shows how the latest information will be stored in the cell state by the input gate. This can be divided in to two sub parts. The first part is the input gate's weight(Wi) and the input gate's bias B_i will decide which value is going to be updated as presented by the equation one. In the second equation in order to obtain the output first the matrix multiplication operation is executed by the multiplication of the x_t and h_{t-1} by the weight W_i, then bias is added. Same as before at last Sigmoid function is employed to get the results. The latter part which is presented by equation three forms a Candidate vector denoted by ^Ct which is integrated to the cell state. Same as before but using tanh function ^Ct is obtained by the multiplying of the x_t and h_{t-1} by the weight Wi, and then adding bias $b_{c..}$ The estimated i_t and C_t help the cell state of the precious step to be get updated. As given by equation 4 the update process of a cell state is done by forgetting. Forgetting done by the multiplication of f_t which is the result of the forget gate which was calculated by the by the cell state of the previous step (C_{t-1}) and then adding the new candidate which is i_t multiplied by $\hat{C}_{t.}$

Equation 2

$I_t = \sum(W_i * [h_{t-1},x_t] + b_i)$

Equation 3

$\hat{C}_t = \tanh(W_c * h_{t-1}| x_t] + b_{c)}$

Equation 4

$C_t = f_t * C_{t-1} + i_t * \hat{}C_t$

The result is decided as the output by the output gate as the final stage. Equation 5 and Equation 6 presents the operation done with in the output gate. With the multiplication shown by the equation 5, a matrix multiplication is performed by multiplying $[h_{t-1}, x_t]$ by the learnable output gate weight (W0) and the learnable output gate bias (b_0) being added at next. Then, sigmoid function is used to get the output for the value which was added. A value between -1 to 1 is assigned to the cell state by the tanh function. Equation 6 shows that the hidden state is derived from the value obtained from the Equation 5 which is the 0_t and the cell state value.

Equation 5

$0_t = \sum(W_0[h_{t-1}, x_t] + b_0)$

Equation 6

$H_t = 0_t * \tanh(C_t)$

Vector Auto Regression VAR

The usage of this statistical model is to forecast timeseries which two or more in number where they have relationship in between them. The term 'autoregressive' refers to how each time-series variable is modelled as a function of its previous values, with lags acting as predictors. The VAR model is bidirectional, which distinguishes it from ARMA, ARIMA, and other models. That is, predictors influence Y value, and Y value influences predictors.

The VAR model can be represented as a system of equations, with one equation for each variable that can be represented as vectors. If we have a vector of time series data Yt, we can express a VAR model with k variables and p lags mathematically.

Equation 1

$Yt = \alpha + \beta 1Yt\text{-}1 + \beta 2Yt\text{-}2 + \ldots\ldots + \beta pYt\text{-}p + Et$

Where Yt, B0, and are k 1column vectors and B0, B1, B2, Bp are k coefficient matrices. The simplest VAR model for three variables with lag is as follows,

$$\begin{bmatrix} y_{1(t)} \\ y_{2(t)} \\ y_{3(t)} \end{bmatrix} = \begin{bmatrix} b_{10} \\ b_{20} \\ b_{30} \end{bmatrix} + \begin{bmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{bmatrix} \begin{bmatrix} y_{1(t-1)} \\ y_{2(t-1)} \\ y_{3(t-1)} \end{bmatrix} + \begin{bmatrix} \epsilon_{1(t)} \\ \epsilon_{2(t)} \\ \epsilon_{3(t)} \end{bmatrix}$$

Testing images using classification techniques

Determining whether a leave is healthy or not using the image provided is done in this classifying phase. Based upon the categorization arrived using the prior training the unhealthy ones are identified. A classifier is created in MATLAB environment in this study.

Among the huge varieties of classifiers such as Back propagation network, Naive Bayes, Decision Trees, Support Vector Machine (SVM), and K-nearest neighbor, SVM is chosen even though each one of them has its pros and cons, this is because of the reliable approach that is found in the SVM.

Convolutional Neural Network

A Deep Learning algorithm is a Convolutional Neural Network (CNN).

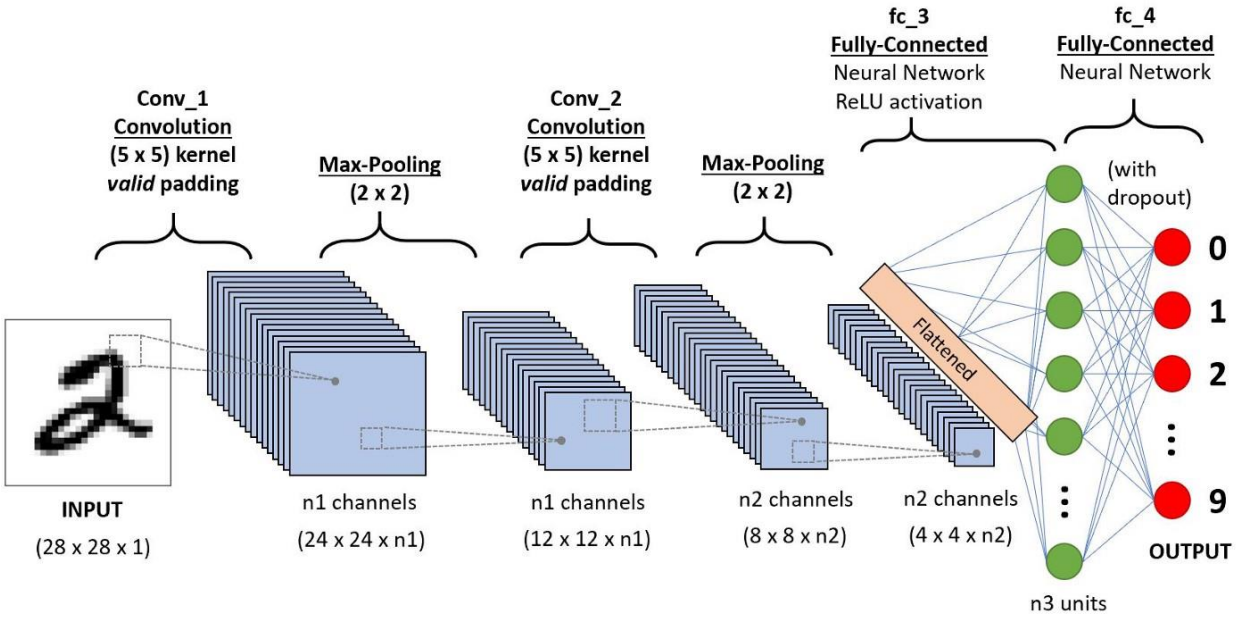


Figure 3.10 CNN Algorithm

The convolutional neural network (CNN), often known as CNN, It is a type of deep learning neural network. In a nutshell, consider CNN to be a machine learning system that can take an input picture, assign significance (learnable weights and biases) to various aspects/objects in the image, and distinguish between them. Because of its great accuracy, CNNs are utilized for picture categorization and identification. The CNN uses a hierarchical model that builds a network, similar to a funnel, and then outputs a fully-connected layer in which all neurons are linked to each other and the output is processed.

Image classification process

The test's input picture stream is initially gathered into a target directory for the study aim. The CNN model is then used to binary classification, with the classification image directory being divided into our "Positive" and "Negative" outcome biases. The positive bias directory contains photos with strong features that may be used to demonstrate the influence of goal research aims, whereas the negative bias directory contains images that are comparable but do not correspond to the target research bias.

After configuring the image directory containing our pre-training pictures, we use the Keras API's dataset generator method to establish an image feed with the following target parameters:

The pictures are then processed using CNN to build our machine learning model, which can then categorize the raw image data from the source capture device. There are layers in the CNN model that need to be parameterized. Following are the layers: Layers: Convolutional, Pooling, and Fully Connected

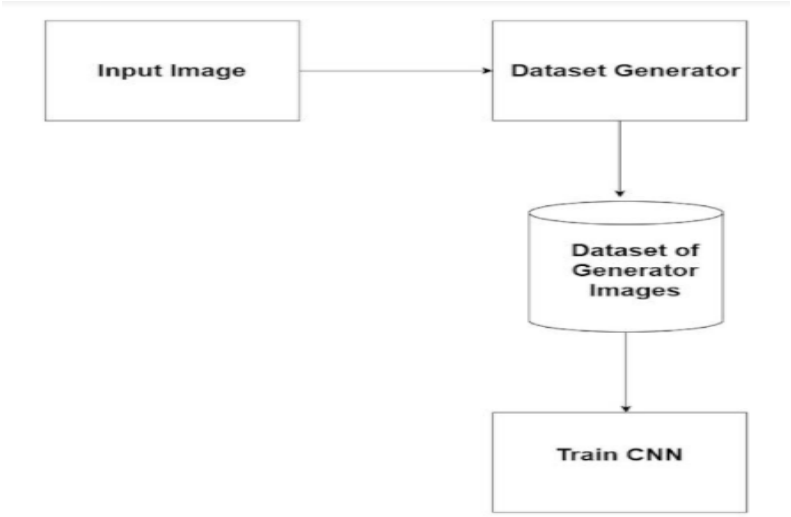


Figure 3.11 Data Training

I used Google colab for data training and testing. For deep learning tasks, Google Colab is a great tool. It's a hosted Jupyter notebook that doesn't require installation and includes a great free version that provides you free access to Google processing resources like GPUs and TPUs. In only a few lines of code, you can import an image dataset, train an image classifier on it, and assess the model with Colab. Colab notebooks run code on Google's cloud servers, allowing you to take advantage of Google hardware, such as GPUs and TPUs, independent of your machine's capabilities.

TensorFlow is a Python library for fast numerical computing created and released by Google. It is a foundation library that can be used to create Deep Learning models directly or by using wrapper libraries that simplify the process built on top of TensorFlow. And an open source platform for decentralized data machine learning and other computations. TensorFlow Probability is a library for statistical analysis and probabilistic reasoning. Tensor2Tensor is a deep learning model and dataset library that aims to make deep learning more accessible and ML research faster.

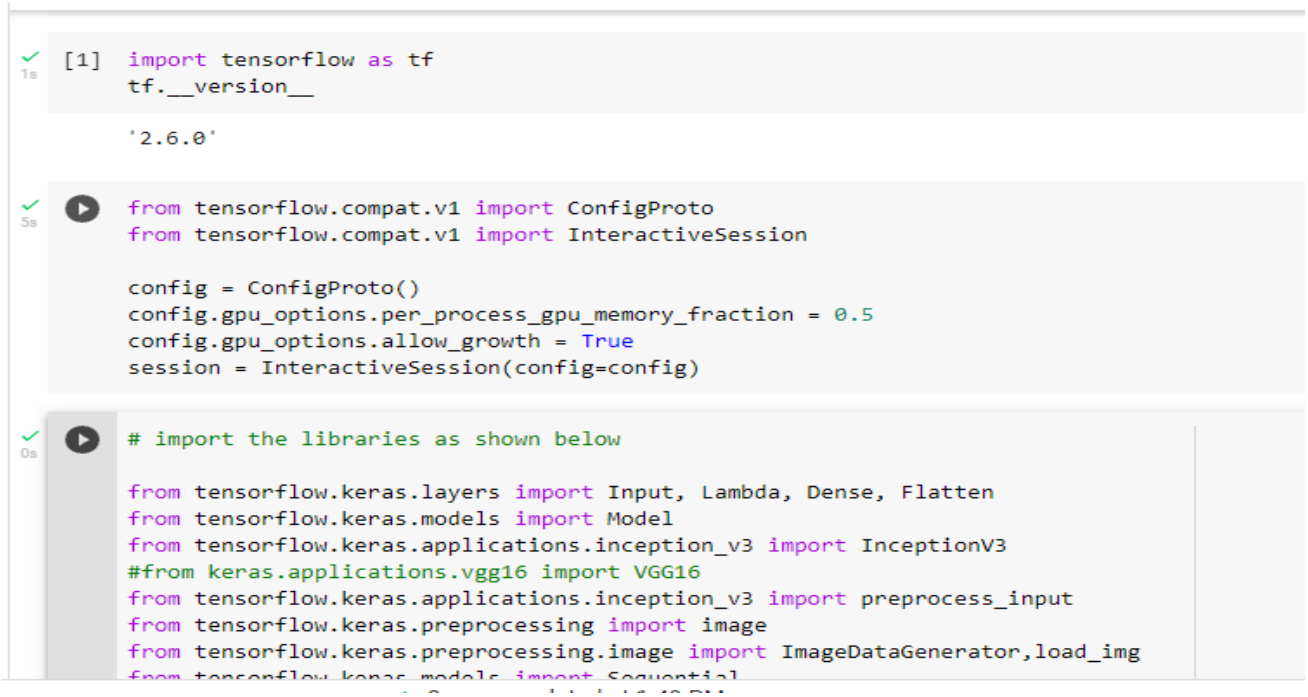


Figure 3.12 library version 2.6.0

I imported tensorflow library version 2.6.0 for model training. And also imported

```
[4]  # re-size all the images to this
     IMAGE_SIZE = [224, 224]

     train_path = '/content/drive/MyDrive/plant diseases/Dataset/train'
     valid_path = '/content/drive/MyDrive/plant diseases/Dataset/test'
```

This is used for resize all images.

4

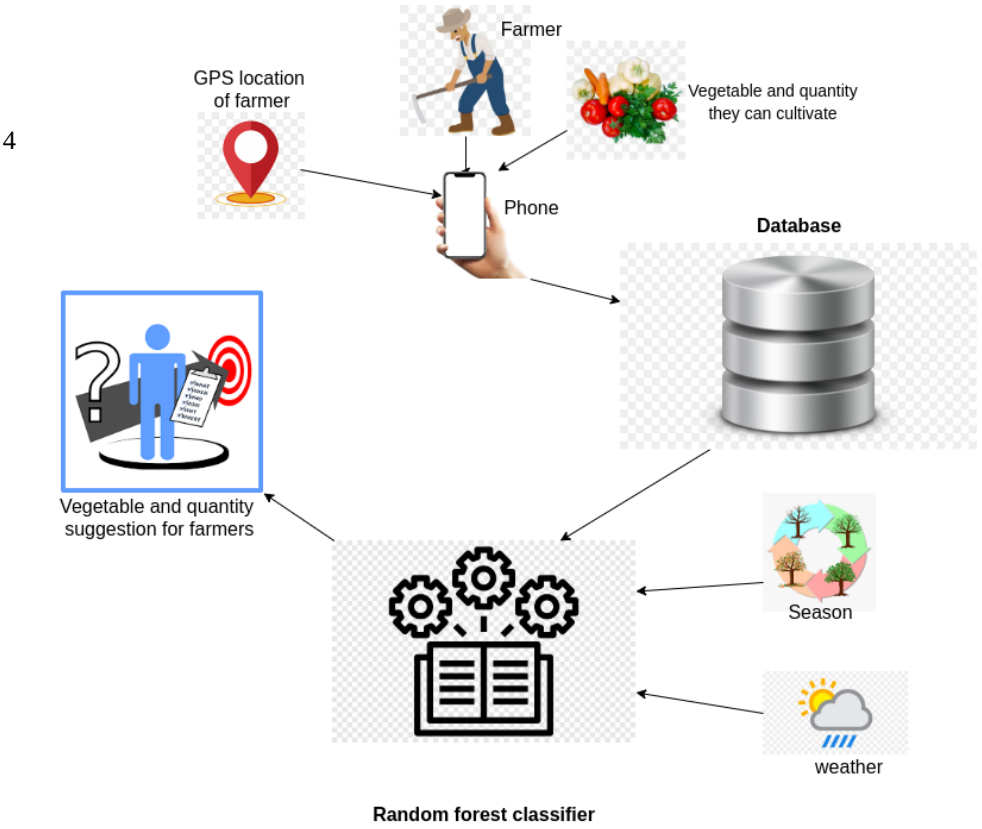


Figure 3.13 System architecture

4.1 Development Process

Software Development Life Cycle (SDLC) is a process used by the software industry to use any type of software largely. Therefore, we must follow certain procedures to achieve the ultimate goal of our research object.

This part is made under this life cycle in a very special way. Consistent with the adoption after the number 5 that appears below, the required requirements were divided into a useful presentation in a very large way. Within the emphasis handle, each switch module needs to perform the required requirements, configuration, and use, and in a visual way.

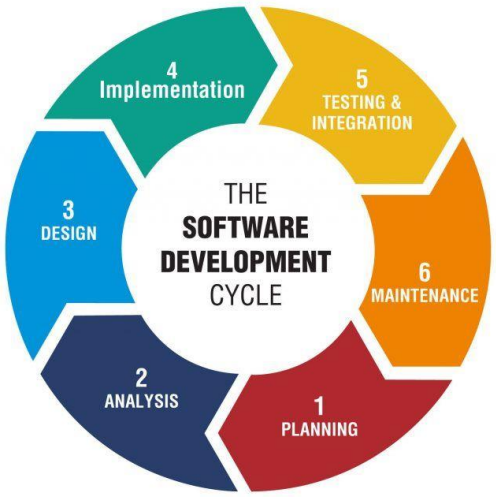
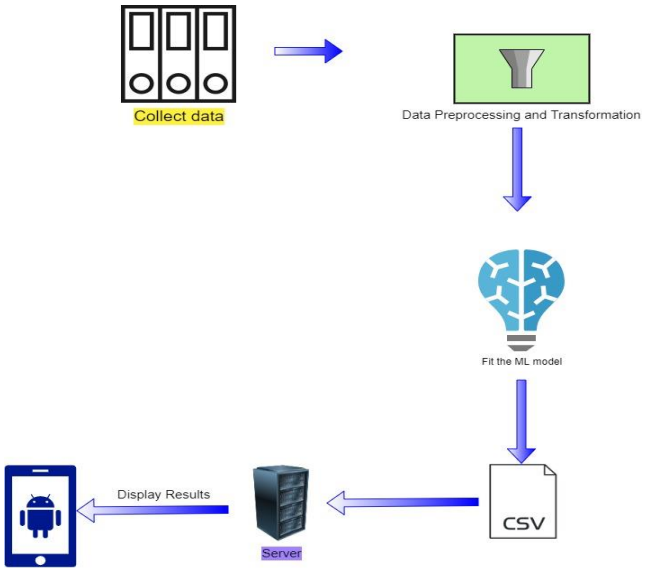


Figure 3.14: Development process of the system.

System overview diagram for vegetable price prediction



4.2 Requirement Gathering

Many research papers are focusing on farmers' problems to find a solution and what kind of system we will use that gives the best results with the highest accuracy. By analyzing and gaining a better understanding of the best solutions for farmers. The following requirements are considered when using classification models.

4.2.1 Functional and non-functional requirements

- **Functional requirements**
    - Getting the simple English description from the farmer and correct the spelling mistake.
    - Suggest the correct vegetable name in Tamil, Sinhaha, English languages.
    - Entered dataset wants to be clean by using special features of NLP. Such as lemmatization & stemming, stop word removal, named entity recognition algorithm, and tokenizing.
    - Used to find plant diseases and diseases
    - Used to find pesticides
    - Used to find pesticide using method
    - Used to find management of diseases
    - Used to analyzed the market prices of vegetables and produce prediction.
    - Used to display the predictions in way useful to Sri Lankan Farmers.
    -
    -

- **Non-functional requirements**
    - It should be filled the gap between human communication and machine understanding.
    - It wants to be user-friendly and a good assistion for farmers.

- Come up with the appropriate output.
- Should be made farmers feel like good satisfaction.
- Less manual work
- Take less time to display a forecast through the android system.
- Accuracy of prediction**.**
-

4.2.2 User Requirements

- By providing a simple description getting the actual name of the products and diseases.
- Getting suggestions for disease.
- Providing marketing suggestions.
- Less manual work
- Take less time to detect diseases
- Take less time to get accurate result
- Accuracy of prediction.

4.3 Data collection

Collecting the dataset is one of the noticeable works that has got to be wiped out of this component. In this pandemic situation, I faced difficulties to meet farmers and their society directly. I choose some farming societies from different areas in Sri Lanka and contact them using social media such as Whatsapp, Viber, Facebook. And joined some Srilankan farmers Facebook groups and shared my google form and asked them to fill that. After that, I collected some vegetable

details and get some more details about the main problems they are facing, what they are expecting to solve those problems like information collected from them. The vegetable portrayal dataset which we collected from a distinctive range set of agriculturists and their relations. The dataset contains about 50 different sorts of vegetables. The taking after figure appears as the outline of the dataset.

Price prediction

Data is collected for 5 vegetables of Dambula and Petttah market. The data is daily prices from March 2015 to June of 2021.

	date	pettah_wholesale	pettah_retail	dambulla_wholesale	dambulla_retail
0	3/6/2015	60.0	80.0	60.0	85.0
1	3/9/2015	55.0	80.0	45.0	70.0
2	3/11/2015	50.0	70.0	40.0	65.0
3	3/12/2015	55.0	75.0	40.0	65.0
4	3/13/2015	50.0	70.0	48.0	65.0
...	...	...	...	...	...
1504	6/23/2021	150.0	180.0	155.0	175.0
1505	6/25/2021	120.0	160.0	145.0	165.0
1506	6/28/2021	120.0	160.0	NaN	NaN
1507	6/29/2021	120.0	160.0	83.0	103.0
1508	6/30/2021	110.0	140.0	68.0	88.0

Figure 4.3.1: dataset with Descriptions and vegetable names.

The data set has 10 tomato leaf diseases and 4 paddy leaf diseases.

Tomato Diseases

1. Bacterial Spot

2. Early Blight
3. Late Blight
4. Leaf Mold
5. Septoria Leaf Spot
6. Spider Mites
7. Target Spot
8. Tomato Mosaic Virus
9. Tomato Yellow Leaf Curl Virus
10. Healthy Leaf

Paddy Leaf Diseases

1. Brown spot
2. Hispa
3. Leaf Blast
4. Healthy

The raw data gathering stage is when the process of analyzing image processing techniques begins. Image data is the sort of raw data that is fed into the system. In this phase, the injured plants are photographed using a smartphone camera.

4 TESTING & IMPLEMENTATION

4.1 Testing

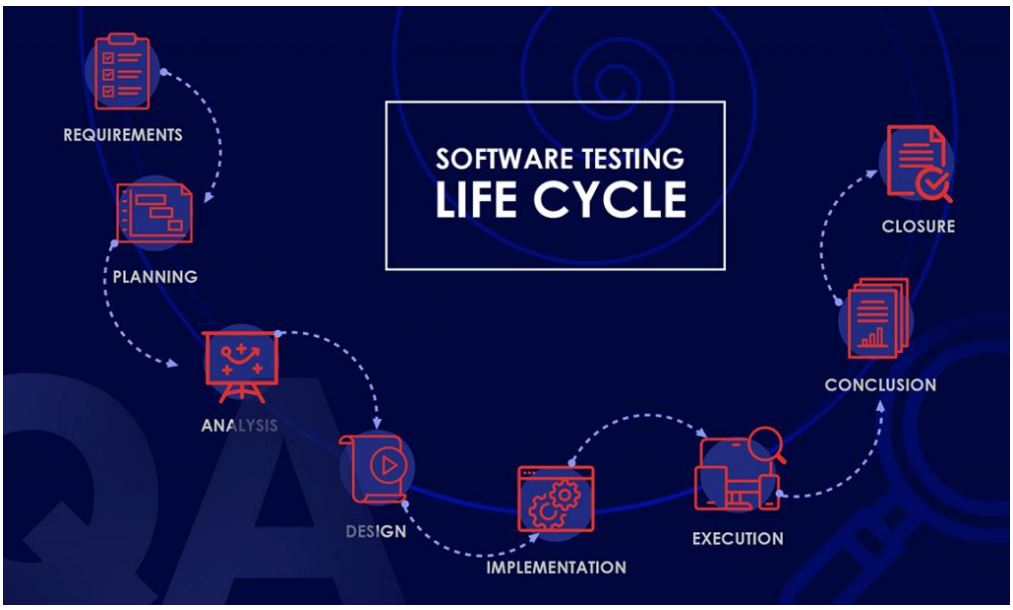


Figure 4.1.1: Software testing phases.

In addition to Figure 5.1.1, there is an undoubted approach to system testing at a given time in each phase of improvement, system testing, and bug fixing. By testing, large-scale applications complete the type of additional quality and the best feature. Computer system testing can be particularly important within the developmental life cycle to detect abnormal absconds and abnormalities between stages of development. And it is very important to make it without any doubt about the quality of the item. is required for the compelling operation of an object or a computer program, contrary to popular belief. it is required that every arrangement we take after the SDLC be subjected to non-compliant testing categories.

1. Unit Testing
   In this arrangement, each unit of the person is tested and performed without any doubt that each function works correctly without interruptions. As a result of this section, it passes through the famous way as a flawless part and will be prepared for the part to be directed and the main part. This part of the test for

the most part comes under the examination of the white boxes, which is contrary to popular belief.

2. Module Testing

   It is defined as another type of system test, which is performed by looking at the subroutine, subprogram, and class.

3. Integration testing

   In this experiment, all individual components or units were assembled and tested as a whole. These test sites test the compatibility of each component or system that contains usability requirements.

4. System Testing

   When the complete outline was the links, it was made. All the required requirements of the framework are assessed in this assessment. End-to-end data testing is the reason for this type of test. Once the bug has been detected from this test program, then the bug will be fixed by the people in the group

5. User Acceptance Testing

   The client or end-user performed this test. It will be done at the end of the test phase. Customer input and input are provided by targeted customers. The main purpose of this test is to determine whether the intended customer needs are being met or not

6. Maintenance

   The final stage is the final stage of the SDLC exhibition. Here, maintenance of a system restructuring, repair, and repair system is considered one of the operations performed in this section.

The generated part was passed on to all test stages without error. The complete framework should be separated when conducting an assessment phase that can be a visual approach to assessment.

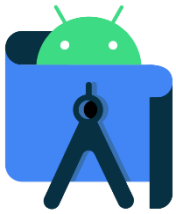
4.2 Component Overview

4.2.1 Developing a smart gadget to get water samples for predictions purposes.

Software Boundaries

- AndroidStudio

  One of the popular official IDE for Android Operating System (OS) of Google' and that is built by JetBrain IntelliJ IDEA. It provides enhanced features of the software that is designed especially for android application to develop development.



- Pycharm

  It is a popular IDE for computer programming, which is specially used for Python programming language developing environment. It is developed and delivered by the Czech company Jetbrains. It has enhanced features for graphical debuggers, code analysis, testing, and version controller.

- GitLab

  It is a DevOps platform for open source and end-to-end software development that builds in version control, Continuous Integration and Deployment, code review, and issue tracking. It provides self-host on our own servers or on cloud providers.

- 

Django that promotes rapid development and clean, and proper design. It was formed to take care of much of the inconvenience of software web development, allowing you to focus on writing application without having to start from scratch. It is both free and open.

Hardware Boundaries

For backend high-performance server machine needed. To run this application an android mobile is needed with the below requirements.

- ✦ The processor speed should be 1.2 GHz or later.
- ✦ Ram should be a minimum of 2GB.
- ✦ Internal storage 1GB or more than that should be available.
- ✦ Screen resolution 1280 x 720 or higher.

Front End

Technologies used: Android and Java

Platform used: Android Studio

Android Studio is the official integrated development environment for Google's Android operating system, built on JetBrains' IntelliJ IDEA software and designed specifically for Android development [21]. When using this platform for front end development, we need to create separate activities to define specific functionality.

As an initial step, an activity was created to display the readings obtained from the device. Once an activity has been created, the platform will generate a layout file which is linked to the created activity. The layout is defined using an Extensible Markup Language (XML) file where we can work on the design of the user interface.

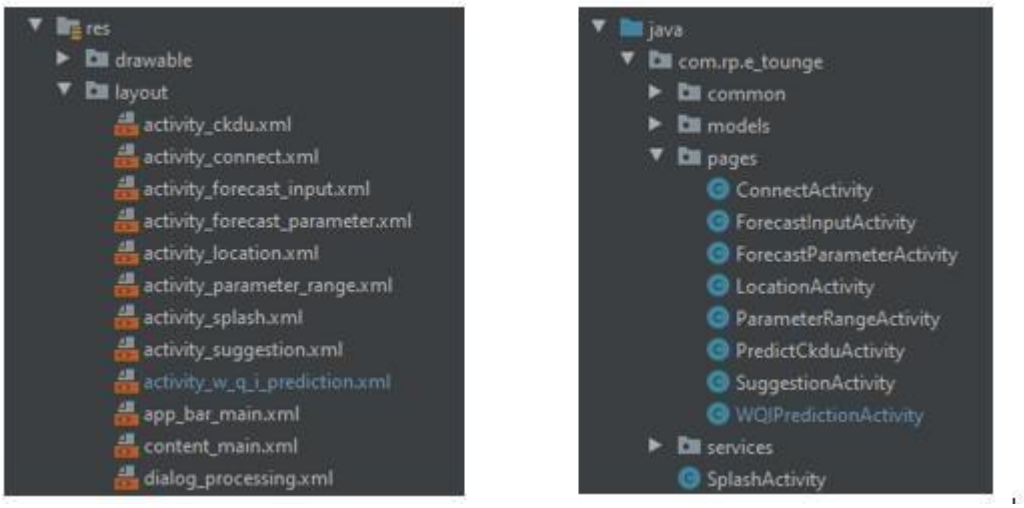


Figure 4.2.1: *Android studio classes and activities*

When implementing the backend server to connect with the front end, a method is called where the server API is called to obtain data and predicted result to be displayed in the user interface.

```
public interface ApiServices {

    @GET("/resultParameterForMonth")
    Call<List<PredictWaterParameters>> getJson(@Query("Year") String year, @Query("Site") String site,
                                               @Query("month") String month);

    @GET("/resultWQI")
    Call<List<WQI>> getJsonWQI(@Query("Temperature") float Temperature, @Query("pH") float pH,
                               @Query("Turbidity") float Turbidity, @Query("Conductivity") float Conductivity,
                               @Query("TDS") float TDS);


    @GET("/resultCKD")
    Call<CkdResult> getCkduResult(@Query("site") String site,
                                  @Query("temperature") float temperature,
                                  @Query("ph") float ph,
                                  @Query("turbidity") float turbidity,
                                  @Query("tds") float tds,
                                  @Query("ec") float conductivity);


    @GET("/parameters/getParameters")
    Call<SensorDataResponse> getSensorData();
}
```

Figure 4.2.2: Api services

Back End

Technologies used: Python 3.8.5, NodeJS

Platform used: PyCharm, Visual Studio Code

The backend of the system comprises of 2 different components. The results from these 2 backend servers must be directed to the front end for the users to view the data.
As the device reads parameter readings from the sample, it is designed to store the reading values directly to the cloud where the AWS DynamoDB is hosted. The device is designed in a way where it sends 5 rows of data from five timestamps with equal intervals. As this particular technique ensures the quality of the data that we are obtaining, we have created a backend server using NodeJS to obtain the average of the stored 5 rows and to direct them to the python model which was hosted in another server.

```
AWS.config.update(awsConfig);
let docClient = new AWS.DynamoDB.DocumentClient();

var ParameterController = function(){
    this.getData =()=>{
        var params = {
            TableName : 'ESP8266TEST',
            IndexName : 'device-index',
            KeyConditionExpression : 'device = :deviceVal',
            ExpressionAttributeValues : {
            ':deviceVal' : 4468785
            }
        };
        docClient.query(params, function(err, data) {
            if (err) {
                console.error("Unable to read item. Error JSON:", JSON.stringify(err,
                        null, 2));
            } else {
                console.log("GetItem succeeded:", JSON.stringify(data, null, 2));
            }
        });
    }
}
```

Figure 4.2.3: Database data fetching api call

4.3 Testing

A. WQI prediction - Functionality Test

The testing procedure of the above elaborated implementation is explained below.

Front-end testing

Test case ID	001
Test case scenario	Check to fetch user current location
Test steps	a. User navigate to the home page b. Tab locate me button c. Display the current user location
Test data	Location
Expected result	User location should be shown in the home screen
Actual result	As expected
Pass/Fail	Pass

Test case ID	002
Test case scenario	Search user entered location

Test steps	a. The user navigates to the forecast page b. Enter the user's desired location c. Select the water resource area d. Submit
Test data	Location = Galle, Sri Lanka
Expected result	The location should be shown on the home screen
Actual result	As expected
Pass/Fail	Pass

Test case ID	003
Test case scenario	Show warning alert on water quality parameter values based on range
Test steps	a. The user navigates to the home page b. User select the user location c. The user navigates to the predict CKDu page
Test data	Temperature =27.36 Ph =6.94 Turbidity =14.15 TDS = 162.80 EC =0.2
Expected result	The color of the card should be change accroding to the range of the water quality parameter
Actual result	As expected
Pass/Fail	Pass

C. Forecasting parameters – front end testing

Test case ID	001
Test case scenario	Check to fetch user current location
Test steps	d. User navigate to the home page e. Tab locate me button f. Display the current user location
Test data	Location
Expected result	User location should be shown in the home screen
Actual result	As expected
Pass/Fail	Pass

Test case ID	002
Test case scenario	Search user desired location
Test steps	e. The user navigates to the forecast page f. Enter the user's desired location g. Select the water resource area h. Submit
Test data	Location = Nochchiyagama, Sri Lanka
Expected result	The location should be shown on the home screen
Actual result	As expected
Pass/Fail	Pass

Test case ID	003
Test case scenario	Check calendar

Test steps	a. The user navigates to the forecast page b. Hit the calendar button c. Select month and year d. Submit
Test data	Month = December, Year = 2022
Expected result	Display year and month
Actual result	As expected
Pass/Fail	Pass

Test case ID	004
Test case scenario	Check the submit button disabled when the calendar field empty.
Test steps	a. The user navigates to the forecast screen b. User enter location or site c. Keep the calendar field empty
Test data	Location = Nochchiyagama, Sri Lanka
Expected result	The alert dialog should be shown as “Input expected year and month”
Actual result	As expected
Pass/Fail	Pass

Teat cases for image processing

Test case ID	01
Test case Description	Verify tomato healthy leaf will be identify by the system
Pre-Condition	Install the app in mobile
Test Steps	1. Open the App 2. Click “detect diseases “ button 3. Browse the image 4. Select the image 5. Click “confirm” button
Test Input	Select tomato leaf mold diseases 
Expected Result	The system should give the output as “ tomato leaf Mold” and suggest pesticides and management to cure the disease
Actual Result	tomato leaf Mold
Result	Pass

Test case ID	02

Test case Description	Verify tomato diseases will be identify by the system
Pre-condition	Install the app in mobile
Test Steps	1. Open the App 2. Click “detect diseases “ button 3. Browse the image 4. Select the image 5. Click “confirm” button
Test Input	Select healthy leaf image 
Expected Result	The system should give the output as “ tomato healthy leaf ”
Actual Result	Tomato Healthy Leaf
Result	pass

Test case ID	03
Test case Description	Verify the app shows the 90 day prediction of vegetables
Pre-condition	Install the app in mobile
Test Steps	1. Open the App 2. Click “price prediction “ button 3. Browse the vegetables 4. Select the vegetab;e 5.
Test Input	capture an object
Expected Result	System should give an error message
Actual Result	Display error message
Result	pass

Test case ID	04

Test case Description	Verify unknown images identified by the system
Pre-condition	Install the app in mobile
Test Steps	1. Open the App 2. Click “detect diseases “ button 3. Browse the image 4. Select the “take photo” button 5. Capture a different object 6. Click “confirm” button
Test Input	Selecting the vegetable
Expected Result	System should show the dates and prices of the vegetable
Actual Result	Display prices
Result	pass

5. RESULT & DISCUSSIONS

5.1 Results

The implementation of research work begins with uploading a database of plants already collected. It continues to import the required libraries and packages and continues to process and prioritize data. Data is spit out for test data and professional data. Finally, a model is developed using the required AI techniques that will provide the best possible

crop to be grown in a particular country. Results Analysis the research work is being implemented using the plant database collected on the kaggle.com, website. It contains a variety of crops, such as, wheat, corn, rice, pea, sorghum, green gram, dove pea, sugarcane. Installed with a few predictive parameters such as, temperature, pH value, rainfall, humidity relative to the location. With a predictive model, in-depth learning algorithms and machine learning require two types of data, namely, a test set and a trained set. Trained data are research data collected that have been collected from past events. While current research data is experimental data. After applying the code, the answer is clearly represented by each of the parameters taken at the beginning. Which are include the use of fertilizers, environment, pesticides and water. And yield is predicted on the basis of the following data:

Figure 7.

```
dataset.head()
```

	N	P	K	temperature	humidity	ph	rainfall	label
0	90	42	43	20.879744	82.002744	6.502985	202.935536	rice
1	85	58	41	21.770462	80.319644	7.038096	226.655537	rice
2	60	55	44	23.004459	82.320763	7.840207	263.964248	rice
3	74	35	40	26.491096	80.158363	6.980401	242.864034	rice
4	78	42	42	20.130175	81.604873	7.628473	262.717340	rice

Figure 8 Data analysis

```
plt.show()
```

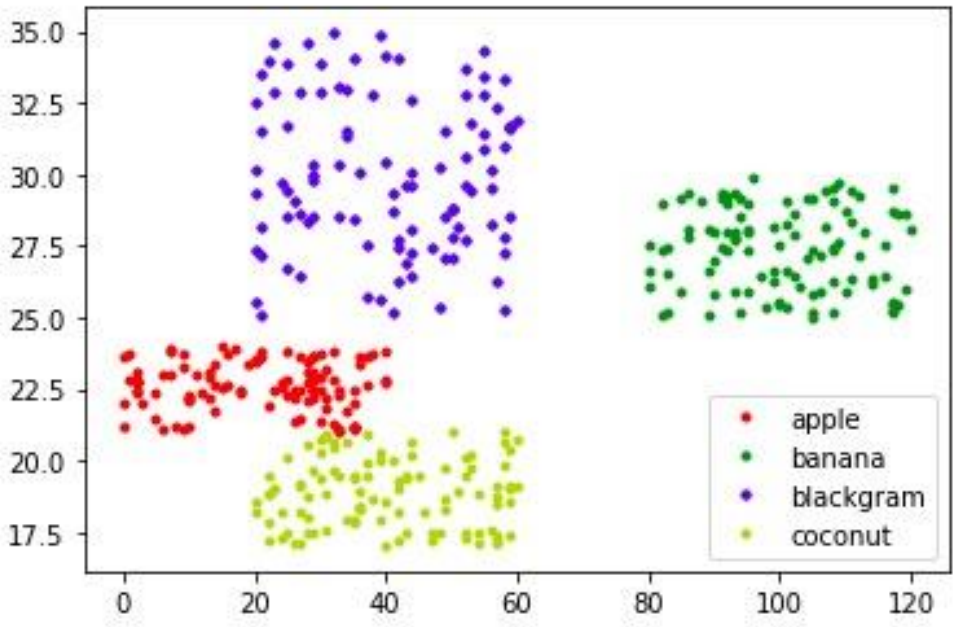


Figure 5.1: Data Analysis

The research project's final product was a mobility and Android platform-based application. It is an APK file that can be installed on Android devices. A single app with several outcomes is included.

• Identify insect attacks
• Recognize illness
• Suggest pesticides and fertilizer
• Suggest diseases management cultural method

Keras and Tensorflow libraries are used to train image models. The dataset training procedure is depicted in the image below.

```
  validation_steps=len(test_set)
)
```

```
/usr/local/lib/python3.7/dist-packages/keras/engine/training.py:1972: UserWarning: `Model.fit_generator` is deprecated and will be removed in a future version.
  warnings.warn('`Model.fit_generator` is deprecated and '
Epoch 1/7
1147/1147 [==============================] - 5957s 5s/step - loss: 4.1895 - accuracy: 0.6841 - val_loss: 3.3645 - val_accuracy: 0.7708
Epoch 2/7
1147/1147 [==============================] - 2241s 2s/step - loss: 3.4601 - accuracy: 0.7890 - val_loss: 3.8129 - val_accuracy: 0.7972
Epoch 3/7
1147/1147 [==============================] - 2243s 2s/step - loss: 3.3702 - accuracy: 0.8192 - val_loss: 5.3741 - val_accuracy: 0.7692
Epoch 4/7
1147/1147 [==============================] - 2249s 2s/step - loss: 3.1343 - accuracy: 0.8388 - val_loss: 3.3554 - val_accuracy: 0.8336
Epoch 5/7
1147/1147 [==============================] - 2253s 2s/step - loss: 3.0341 - accuracy: 0.8527 - val_loss: 3.2981 - val_accuracy: 0.8493
Epoch 6/7
1147/1147 [==============================] - 2253s 2s/step - loss: 2.8682 - accuracy: 0.8663 - val_loss: 5.1393 - val_accuracy: 0.8028
Epoch 7/7
1147/1147 [==============================] - 2256s 2s/step - loss: 2.8938 - accuracy: 0.8707 - val_loss: 3.6258 - val_accuracy: 0.8537
```

Figure 5.2: Dataset training

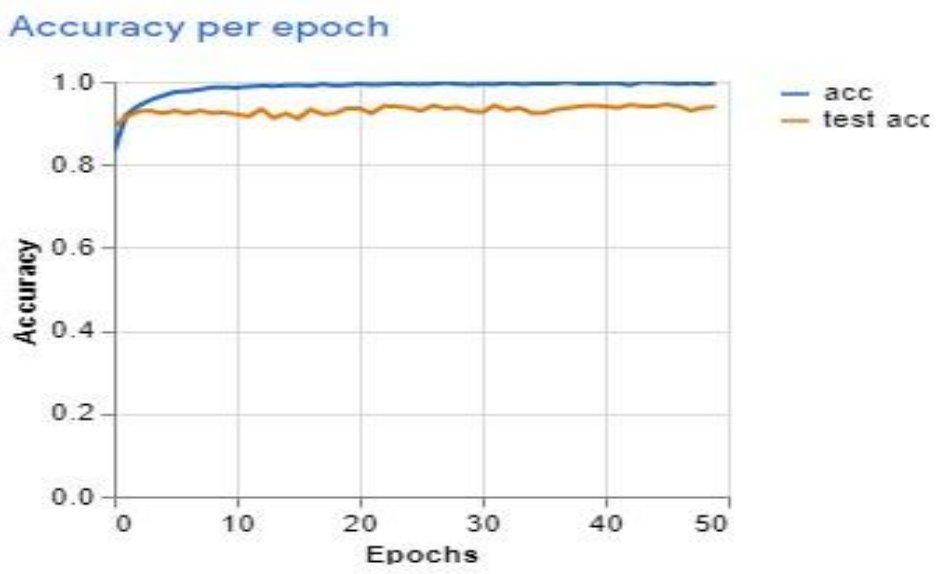


Figure 5.3: Accuracy

The figure shows accuracy of tomato leaf diseases detection in both of training and testing data.

Figure depicts the tomato disease samples and accuracy. When utilizing 50 epochs in data training, the most accurate result is close to 98 percent.

Output

Using the training data file, created python file to get accurate result

Discussion

Such research is susceptible to potential threats to validity can be external, construct reliability and validity. The construct validity and external validity are addressed for this research since the initial search string was broad, and nearly 600 publications in total the query returned a substantial number of studies. whole scope of the SLR is covered by the search string. For the SLR reliability, the validity can be considered well-addressed since the process of the SLR has been described replicable and clearly. If this SLR is replicated, , but the differences would be a result of different personal judgments, and it could return slightly different selected publications. Whatever it is not like that the overall findings would change.

Important publications may be published. Many comparisons extensive and could be used, and researchers may return new lessons. However, the search string has indicated a broad enough search and led to a high number of indexes.

Another issue that could be detrimental to performance in the way analysis is done. For example, sometimes a few examples of features are described and not all publications report on the type of test parameters used. So, the information needed to deal with not

available on research paper questions. In this way, the data not necessary to answer the total number of selected books research questions were taken from a few published numbers rather than that. The authors may have been contacted to find out more about the publication, and that too may not resolve all problems, but this course of action did not take place within the context of this research.

The second most used algorithm is Linear Regression. Which is used in most cases to determine whether the proposed algorithm is better than Linear Regression as a benchmarking algorithm. So, it does not mean that it is the most efficient algorithm, even if it is shown in many articles. In fact, DL is a Mechanical Learning type, has been believed to be very promising and used to predict the recent harvest. In this research, we also identified a number of in-depth study subjects. DL methods contains several promising features, which are high performance and automatic feature extraction. Further research into the use of, DL methods expected by us in some problematic domains in predicting yields soon due to the high performance of DL algorithms.

Both integration and editing techniques are used among the selected publications. Since images are used to assemble in those books, the publication is related to the mechanical perspective instead of ML using numerical databases. The use of integration algorithms for this problem can be investigated in detail to find different research perspectives on the problem.

Algorithms and groups designed for features to visualize algorithms and key features. As a result of this decision, clarity is maintained, but detailed data is missing. PH, rainfall, temperature and soil type factors are the most commonly used. In addition, there are also features that are used in certain subjects to those features that are used in a number of subjects. Gamma rays, MODIS-EVI, precipitation, humidity, image time, pH value, irrigation, leaf area, NDVI, EVI, and plant details are those features. There are also studies that use boron, magnesium, potassium, nitrogen, sulfur, zinc, and calcium kind of nutrients. The most widely used features are not always the same types of data. For

example in temperature, many factors such as high temperatures and low temperatures are also used but which measured as moderate temperature.

In selected papers no multiple test methods were reported. As a model quality measure almost, all studies used RMSE. Other test parameters are MAE, R2, and MSE. Most of these frameworks look like some of the parameters mentioned earlier, some of the parameters used in certain subjects, with little difference. Most models for their test parameters have results with high accuracy values, which means that the model makes accurate predictions. The method of verifying the tenfold crossing was preferred by investigators as a test method.

In the article challenges were reported based on explicit statements. But may be other challenges not mentioned in the papers shown. Challenges are especially in the improving the performance model of the area. When more information is collected for training and testing, more can be said about model accuracy. Implementation of models in farm management systems is another challenge. When making applications that the farmer can use, models can be useful not only in making decisions, but also in the growing season. When certain limits of a area are measured and felt, the prediction will be more accurate.

Android Application: The Android application is designed to query machine learning results. The app is compatible with Android OS version. Pages written in Java language. The app has a simple, easy-to-use application that requires only a few taps to get the results you want. Providing only the location and location of the field the Android app provides the name of the plant to be grown there. By accessing user-entered information, the app will query machine learning analysis. Using location, API will extract weather data. Returned weather data is obtained by machine learning separator to yield calculation and predict yield. The output is downloaded by the server to display the result in the application. The main tasks in the application process were particular

account creation, data entry and download results. Account creation helps the user to actively engage with the visual interface of the app. The user fills a field on the home

page to proceed with the results function. The retrieved data transmitted to the machine learning model and plant name are predicted by the calculated yield value.

Figure 6

```
log_reg.score(X_test, y_test)
```

```
0.9727272727272728
```

```
log_reg.score(X,y)
```

```
0.9745454545454545
```

The system's main output is to control pests and diseases while also recommending insecticides and fertilizers. To address this issue, we developed a system that would monitor pest and disease activity and then intelligently recommend pesticides and fertilizers. Using image classification, the algorithm primarily recognizes the affected leaf and identifies illness on the leaf. When examining the photograph, the attention was mostly on the leaf's form and color. Deep learning CNN algorithms are used to do this. When recognizing an image, a dataset model must be compared, therefore construct an image model with the tensorflow and keras libraries.

Here we use all the data from books which is released by Sri Lanka agriculture Department and ministry of Agriculture. So It will help farmers in Sri Lanka to diseases management.

3.2 Research Findings

When revisiting the literature survey at the beginning of the document, it was stated obvious that machine learning approaches are way better and efficient when applying to the solutions that most of the current problems leads to, and solving issues regarding water quality management is no exception.

3.2.1 Device Implementation

Selecting the best combinations of sensors.

The key component of the device was the sensors. Analyzing literature surveys and earlier findings we were able to out the most suitable sensors which is need for predictions. They were ph probe, turbidity, temperature, and electric conductivity and by using conductivity

there was a possibility of calculating total dissolved solids. Since conductivity sensors are much expensive. We had to build the conductivity sensor by utilizing basic concepts of science.

Among these parameters conductivity and turbidity depend on environment temperature. We took this also into consideration to give an accurate result by calibrating the sensors accordingly.

3.2.2 WQI Prediction

By considering the results from the conducted literature survey, the focus of this research study was to propose, design and implement a solution that is faced by the currently available systems. By looking at all the involving factors and limitations, the solution proposed should be an asset to the public where the ultimate focus is flown sent towards them.

As the traditional models which are being practiced specially in our country is more manual and time consuming, the aim is to fashion a device with the inclusion of mathematical models so that the precise results could be driven directly on time. That being said, the scope of this study to introduce the concept of WQI to the currently working bater bodies where it could be used to calculate the quality of the water samples more accurately without any contradiction in the results. The most challenging concept starts at the very beginning when selecting some water quality parameters to be used in the machine learning model. With the presence of 30 different water quality parameters and 15 different parameters among them which are being used by the industrial professionals, 5 main water quality parameters such as pH, turbidity, temperature, electrical conductivity and total dissolved solids are selected for our study based on the correlation of the parameters and the importance they carry.

Preparing a suitable dataset for the machine learning approach was something which needed extra care. As discussed earlier, the concept of this single numerical unit, WQI was a new branding and it needed a lot of background work and understanding to pick a suitable one among the available ones which reflects the nature of water that are present in the country and for the purpose of introducing the Water Quality Index concept to the industry, we had to manually calculate WQI of each parameter set to deliberately show

how difficult it is as it involves a lot of complex mathematical procedures and the usage of graphs. These manually calculated WQIs were added correspondingly to their respective parameter set to prepare the final data.

In the end, after a thorough study and much deliberation, seven regression machine learning models such as linear, ridge, lesso, elastic net, random forest, kNearest neighbors and artificial neural network regression models were handpicked for completing the initial objective. All of the above-mentioned models were trained individually using the same dataset which was obtained after a hectic search in the industry, to select the best model out of it based on their results.

Individual model's test data splits were evaluated and compared, and Random Forest Regression model was selected to carry on this study because of its Highest R-Squared value and its lowest root mean square error value. As this process was a success, the selected RFR model was put into test and the number of trees were adjusted till an acceptable accuracy percentage was obtained.

3.2.3 Forecasting Parameters

Analyzing the technology area

According to the literature survey, our final system should be the mobile application and figure out the specific features, In order to develop the prediction model, need to use the real dataset that consists water quality parameter for a certain location. Therefore, historical data of the water quality parameter level for each site has to be combined to develop the model that is going to predict the parameter values for the future.

Model training is a prominent basic procedure to develop an ML model in which performance depends on the learning outcome of the dataset. When trained the model, have to complete each phase of the ML model development cycle. The incomplete phrase is not going to give the best model to the completed phase model. Data need to be preprocessed before starting the model training stage. Mean normalization and feature scaling are done at the preprocessing phase.

Choosing the best algorithm

The developed system able to forecast the water quality parameters for a given location. It is fully managed by the ML techniques. Before achieving the main goal of the system,

need to satisfy other mandatory requirements, In order to find the best approach, training the ML model with most appropriate algorithm and optimize the accuracy level of it, all the pre-trained model should be examined with test data set then obtaining the output.

CKDu Outbreak Prediction

Main objective of this research study is to create smart tool to predict safe consumption of ground water. Along with the research project objective, main objective of this individual component is to alert user if there is a possible risk of exposing to CKDu.

Main approach for the component's solution is supervised learning techniques. Training model is the major functionality of it and the whole result of the prediction is depends on the algorithm that is used for the model training. Therefore, selecting best suited model plays the significant importance. To achieve that dataset needs to be trained with multiple algorithms that are suitable for the given scenario and select the best algorithm based on the accuracy of each model. The likelihood of increasing the accuracy of the prediction can be affected by the size of the dataset.

It is plausible that CKDu is multifactorial. Agricultural practices, geographical area distribution and number of other factors are suspected to be the cause of it. Based on literature survey and finding the people who uses shallow wells in close proximity to irrigation systems for agriculture are more affected by the CKDu comparing to the other areas. Hence, we can assume that values of water quality parameters or quality of the water can be one of the causes for the CKDu. Therefore, this research study gives the prediction of possible risk of exposing to the CKDu by analyzing the water quality parameters, location and the past data of the CKDu patients.

With the usage of aforementioned methodologies, this study has compared two widely used models in order to perform multivariate analysis and forecasting consisting of variants such as whole sale and retail prices of vegetables from two Sri Lankan market namely Pettah and Dambulla. Table 1 shows the performance metric used for each variant. In this study, we used daily data of five main vegetables grown in Sri Lanka, and calculated their Mean Absolute Error MAE. The experiment result is shown below.

In here, the same variable but for LSTM shows a different scale among the other metric value. Which makes LSTM error metric a little higher than the VAR evaluation metric.

Vegetable Type	Metric	LSTM	VAR
Beans	MAE	30	43
	MSE	1867	3059
Carrot	MAE	36	32
	MSE	2509	1631
Cabbage	MAE	18	25
	MSE	718	986
Tomato	MAE	24	40
	MSE	1608	2642
Brinjal	MAE	37	52
	MSE	2817	2428
Average	MAE	29	40
	MSE	1903	2149

The above tables shows the performance metric in order to identify the best suited model in order to forecast vegetable price prediction. The results shows that LSTM has the lowest MAE and MSE except for the Carrot vegetable. But as we calculate the total average we can conclude that LSTM would perform better for the price prediction.

The research project's final product was a mobility and Android platform-based application. It is an APK file that can be installed on Android devices. A single app with several outcomes is included.

- Identify insect attacks
- Recognize illness
- Suggest pesticides and fertilizer
- Obtain assistance from agro experts

Keras and Tensorflow libraries are used to train image models. The dataset training procedure is depicted in the image below.

Such research is susceptible to potential threats to validity can be external, construct reliability and validity. The construct validity and external validity are addressed for this research since the initial search string was broad, and nearly 600 publications in total the query returned a substantial number of studies. whole scope of the SLR is covered by the search string. For the SLR reliability, the validity can be considered well-addressed since the process of the SLR has been described replicable and clearly. If this SLR is replicated, , but the differences would be a result of different personal judgments, and it could return slightly different selected publications. Whatever it is not like that the overall findings would change.

Important publications may be published. Many comparisons extensive and could be used, and researchers may return new lessons. However, the search string has indicated a broad enough search and led to a high number of indexes. Another issue that could be detrimental to performance in the way analysis is done. For example, sometimes a few examples of features are described and not all publications report on the type of test parameters used. So, the information needed to deal with not available on research paper questions. In this way, the data not necessary to answer the total number of selected books research questions were taken from a few published numbers rather than that. The authors

may have been contacted to find out more about the publication, and that too may not resolve all problems, but this course of action did not take place within the context of this research.

6.CONCLUSION

We used crop prediction methodology, which showed high efficiency in using large data sets of corn kernels. This method has been used to generate yield predictions based on genotype and genetic data on Artificial neural networks. Deep well-designed networks are able to study the complex relationships between genes, their interactions, and environmental conditions from historical data and to make accurate predictions of new plants planted in new climatic zones. The performance of the model was found to be more sensitive compared to the weather, the importance of the weather forecast strategies suggested by this.

Its black box material is the main limit of the proposed model, many mechanical learning methods are distributed for this. Which model is a complex structure that makes it difficult to produce experimental concepts so that natural understanding can provide for this. We made a feature selection based on the ANN-trained model using the back-to-

back model to make the model under the black box. Genetic selection methods have shown that genetic factors have had a greater impact on plant fruits than genotype effectively and identified important factors. Our future research to overcome this limitation by looking at the most advanced models is more descriptive and accurate.

The system's main function is to detect insect assaults and nutritional deficiencies and recommend pesticides for farmers to purchase through our "E-commerce for Farmers system." Methods concentrate on image processing techniques used to damaged or diseased leaf images, as well as the application of machine learning. CNN and K means clustering approach detection methods may be efficient and accurate in classifying the ailment and suggesting solutions. For automated detection, this approach is efficient and accurate.

7. REFERENCES

[1] Brownlee, J. (2019). "How to Create an ARIMA Model for Time Series Forecasting in Python," *Machine Learning Mastery*. Available at: https://machinelearningmastery.com/arimafor-time-series-forecasting-with-python [Accessed 16 Sep. 2019].

[2] Jia, J. (2003). The study of the application of a keywords-based chatbot system on the teaching of foreign languages. In *arXiv [cs.CY]*. http://arxiv.org/abs/cs/0310018

[3] Matney, L. (2016, August 3). Kik users have exchanged over 1.8 billion messages with the platform's 20,000 chatbots. *TechCrunch*. http://techcrunch.com/2016/08/03/kik-users-have-exchanged-over-1-8-billion-messages-with-the-platforms-20000-chatbots/

[4] Huang, J., Zhou, M., & Yang, D. (n.d.). *Extracting chatbot knowledge from online discussion forums*. Ijcai.Org. Retrieved October 3, 2021, from https://www.ijcai.org/Proceedings/07/Papers/066.pdf

[5] Joachims, T. (2002). Optimizing search engines using clickthrough data. *Proceedings of the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD '02*

[6] Hutapea, A. (n.d.). *CHATBOT: Architecture, Design, & Development*. Retrieved October 3, 2021, from https://www.academia.edu/37082899/CHATBOT_Architecture_Design_and_Development

[7] Abdul-Kader, S. A., & Woods, J. (n.d.). *Survey on Chatbot Design Techniques in Speech Conversation Systems*. Thesai.Org. Retrieved October 3, 2021, from https://thesai.org/Downloads/Volume6No7/Paper_12-Survey_on_Chatbot_Design_Techniques_in_Speech_Conversation_Systems.pdf

[8] Abdul-Kader, S. A., & Woods, J. (n.d.). *Survey on Chatbot Design Techniques in Speech Conversation Systems*. Thesai.Org. Retrieved October 3, 2021, from https://thesai.org/Downloads/Volume6No7/Paper_12-Survey_on_Chatbot_Design_Techniques_in_Speech_Conversation_Systems.pdf

[9] Sharma, R., & Patel, M. (2018). Review on Chatbot Design Techniques in Speech Conversation Systems. *International Advanced Research Journal in Science, Engineering and Technology*, *5*(9), 37–46.

[10] Mikic, F. A., Burguillo, J. C., Llamas, M., Rodriguez, D. A., & Rodriguez, E. (2009). CHARLIE: An AIML-based chatterbot that works as an interface among INES and humans. *2009 EAEEIE Annual Conference*.

[11] (N.d.). Researchgate.Net. Retrieved October 3, 2021, from https://www.researchgate.net/publication/304563737_NLAST_A_natural_language_assistant_for_students

[12] Tran, T. T. K., Bateni, S. M., Ki, S. J., & Vosoughifar, H. (2021). A review of neural networks for air temperature forecasting. *Water*, *13*(9), 1294.

[13] *Allison, P. d. (2002). Missing data. Thousand Oaks, CA Sage. - references - scientific research publishing*. (n.d.). Scirp.Org. Retrieved October 4, 2021, https://www.scirp.org/(S(351jmbntvnsjt1aadkposzje))/reference/ReferencesPapers.aspx?ReferenceID=2335964

[14] Bączkiewicz, A., Wątróbski, J., Sałabun, W., & Kołodziejczyk, J. (2021). An ANN model trained on regional data in the prediction of particular weather conditions. *Applied Sciences (Basel, Switzerland)*, *11*(11), 4757.

[15] Bengio, Y., Simard, P., & Frasconi, P. (1994). Learning long-term dependencies with gradient descent is difficult. *IEEE Transactions on Neural Networks*, *5*(2), 157–166.

[16] Bou-Rabee, M., Sulaiman, S. A., Saleh, M. S., & Marafi, S. (2017). Using artificial neural networks to estimate solar radiation in Kuwait. *Renewable and Sustainable Energy Reviews*, *72*(C), 434–438.

[17] Lee, I., Choi, H., Bang, K.-S., Kim, S., Song, M., & Lee, B. (2017). Effects of forest therapy on depressive symptoms among adults: A systematic review. *International Journal of Environmental Research and Public Health*, *14*(3). https://doi.org/10.3390/ijerph14030321

[18] Burgueño, J., Crossa, J., Cornelius, P. L., & Yang, R.-C. (2008). Using factor analytic models for joining environments and genotypes without crossover genotype × environment interaction. *Crop Science*, *48*(4), 1291–1305.

[19] Atanda, S. A., Olsen, M., Crossa, J., Burgueño, J., Rincent, R., Dzidzienyo, D., Beyene, Y., Gowda, M., Dreher, K., Boddupalli, P. M., Tongoona, P., Danquah, E. Y., Olaoye, G., & Robbins, K. R. (2021). Scalable sparse testing genomic selection strategy for early yield testing stage. *Frontiers in Plant Science*, *12*, 658978.

[20] Bustami, R., Bessaih, N., Bong, C., & Suhaili, S. (n.d.). *Artificial neural network for precipitation and water level predictions of bedup river*. Iaeng.Org. Retrieved October 4, 2021, from http://www.iaeng.org/IJCS/issues_v34/issue_2/IJCS_34_2_10.pdf

[21] Bustos-Korts, D., Romagosa, I., Borràs-Gelonch, G., Casas, A. M., Slafer, G. A., & van Eeuwijk, F. (2018). Genotype by Environment Interaction and Adaptation.

In *Encyclopedia of Sustainability Science and Technology* (pp. 1–44). Springer New York.

[22] (N.d.). Researchgate.Net. Retrieved October 4, 2021, from https://www.researchgate.net/publication/43489119_Genotype_by_environment_interactions_affecting_grain_sorghum_I_Characteristics_that_confound_interpretation_of_hybrid_yield

[23] Casadebaig, P., Trépos, R., Picheny, V., Langlade, N. B., Vincourt, P., & Debaeke, P. (n.d.). *Increased genetic diversity improves crop yield stability under cli- mate variability: a computational study on sunflower*. Arxiv.Org. Retrieved October 4, 2021, from https://arxiv.org/ftp/arxiv/papers/1403/1403.2825.pdf

[24] Peixoto, M. A., Malikouski, R. G., Nascimento, E. F. do, Schuster, A., Farias, F. J. C., Carvalho, L. P., Teodoro, P. E., & Bhering, L. L. (2022). Genotype plus genotype by-environment interaction biplot and genetic diversity analyses on multi-environment trials data of yield and technological traits of cotton cultivars. *Ciencia Rural*, *52*(2). https://doi.org/10.1590/0103-8478cr20201054

[25] Crossa, J., Yang, R.-C., & Cornelius, P. L. (2004). Studying crossover genotype × environment interaction using linear-bilinear models and mixed models. *Journal of Agricultural, Biological, and Environmental Statistics*, *9*(3), 362–380.

[26] Y. Lecun, L. Bottou, Y. Bengio, P. Haffner, "Gradient-based learning applied to document recognition," Proceedings of the IEEE, 1998.

[27] Kurniawati, N. N. et al, "Investigation on image processing techniques for diagnosing paddy diseases" (2009).

[28] Mr.V Suresh, D Gopinath, M Hemavarthini., K Jayanthan, Mohana Krishnan, "Plant Disease Detection using Image Processing., " International Journal of Engineering Research & Technology (IJERT), 2020.

[29] Shivani K. Tichkule, Dhanashri. H. Gawali, "Plant diseases detection using image processing techniques" Online International Conference on Green Engineering and Technologies (IC-GET), 2016.

[30] Yallappa D, "development and evaluation of drone mounted sprayer for pesticide applications to crops" IEEE, 2017.

[31] Santhosh Kumar.S, B.K.Raghavendra, "Diseases Detection of Various Plant Leaf Using Image Processing Techniques" A Review, IEEE, 2019.

[32] Jingzhu Li, Peng Wang, changing geng, "The Disease Assessment of Cucumber Downy Mildew Based on Image Processing" IEEE, 2017.

[33] RadhikaDeshmukh, ManjushaDeshmukh, "Detection of Paddy Leaf Diseases" ICAST, 2015.

[34] Dheeb Al Bashish, Malik Braik, Sulieman Bani-Ahmad, “A framework for detection and classification of plant leaf and stem diseases” IEEE, 2010.

[35] Auzi Asfarian, Yeni Herdiyeni, Aunu Rauf, Kikin Hamzah Mutaqin, “Paddy diseases identification with texture analysis using fractal descriptors based on Fourier spectrum,” The International Conference on Computer, Control, Informatics and its Applications (IC3INA), IEEE, 2013.

[36] Niket Amoda, Bharat Jadhav, Pradnya Kurle, Sharon Kunder, Smeeta Naikwadi, “DETECTION AND CLASSIFICATION OF PLANT DISEASES BY IMAGE PROCESSING” IJERT, 2018

[37] Nikita Goel, Dhruv Jain, Adwitiya Sinha, “Prediction Model for Automated Leaf Disease Detection & Analysis” IEEE, 2019.

[38] Gina S. Tumang, “Pests and Diseases Identification in Mango using MATLAB”, IEEE, 2019.

[39] Ananthi, S.Vishnu Varthini, “Detection & Classification of Plant Leaf Diseases”, International Journal of Research in Engineering & Applied Sciences, Volume 2, Issue 2, February 2012.

[40] Shradha Verma, Anuradha Chug, Amit Prakash Singh, “Prediction Models for Identification and Diagnosis of Tomato Plant Diseases”, IEEE, 2018.

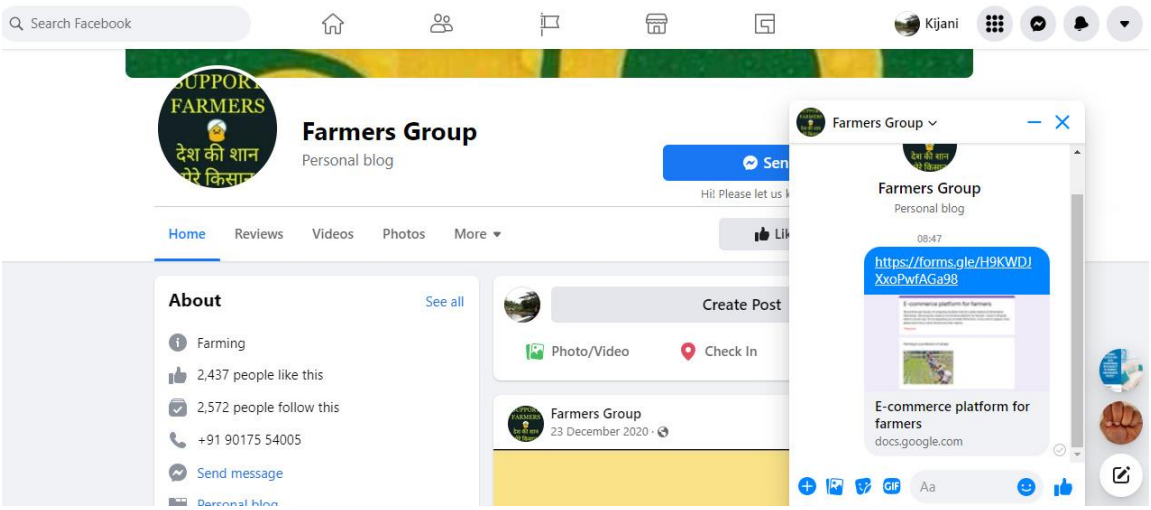
[41] Kanchymalay, K., Salim, N., Sukprasert, A., Krishnan, R. and Hashim, U., 2017. Multivariate Time Series Forecasting of Crude Palm Oil Price Using Machine Learning Techniques. *IOP Conference Series: Materials Science and Engineering*, 226, p.012117.

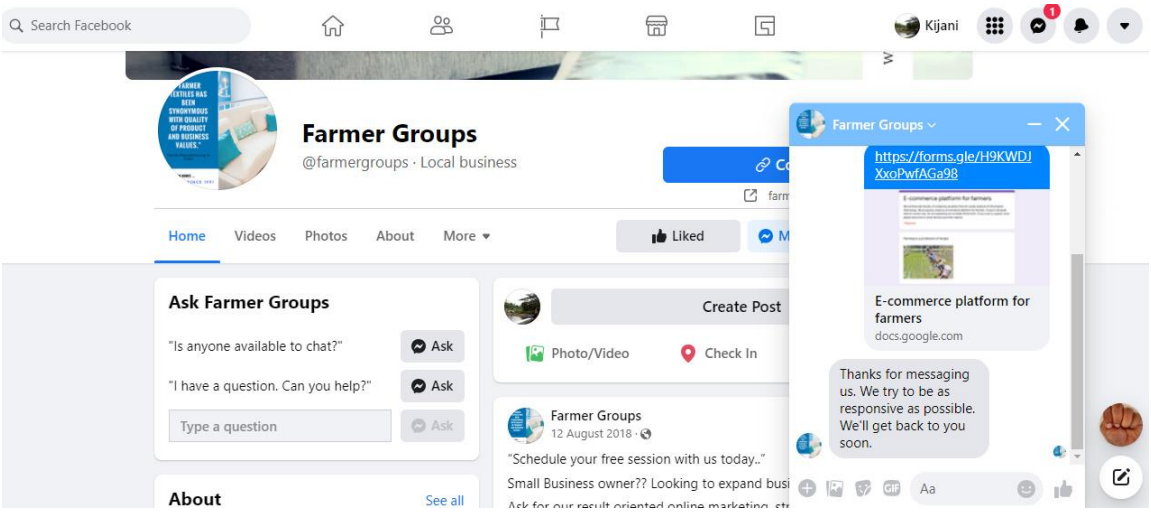
[42] Xiong, T., Li, C. and Bao, Y., 2018. Seasonal forecasting of agricultural commodity price using a hybrid STL and ELM method: Evidence from the vegetable market in China. *Neurocomputing*, 275, pp.2831-2844.

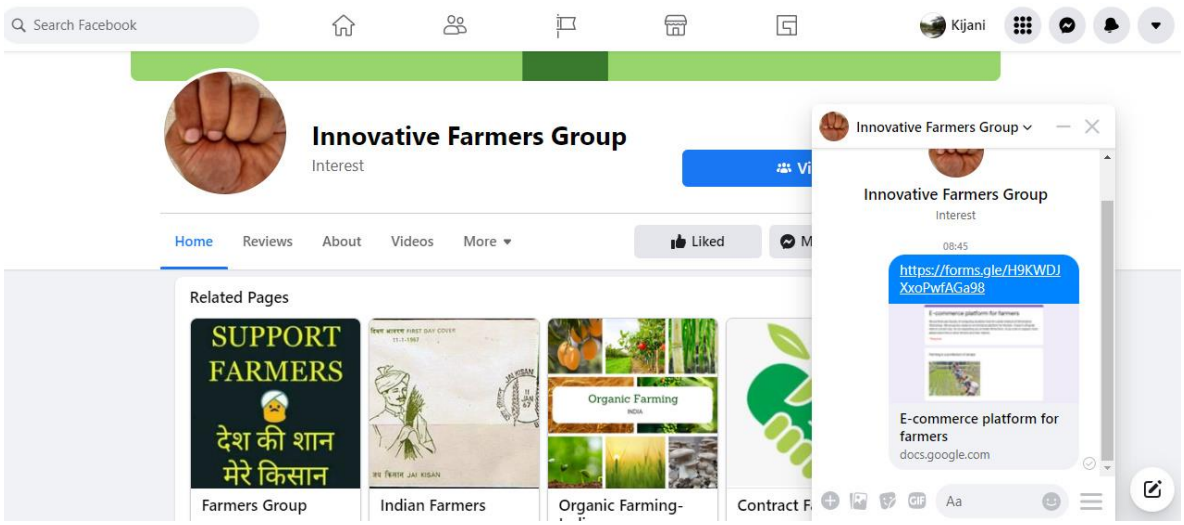
[43] J. K. Sethi and M. Mittal, "Analysis of Air Quality using Univariate and Multivariate Time Series Models," 2020 10th International Conference on Cloud Computing, Data Science & Engineering (Confluence), 2020, pp. 823-827, doi: 10.1109/Confluence47617.2020.9058303.

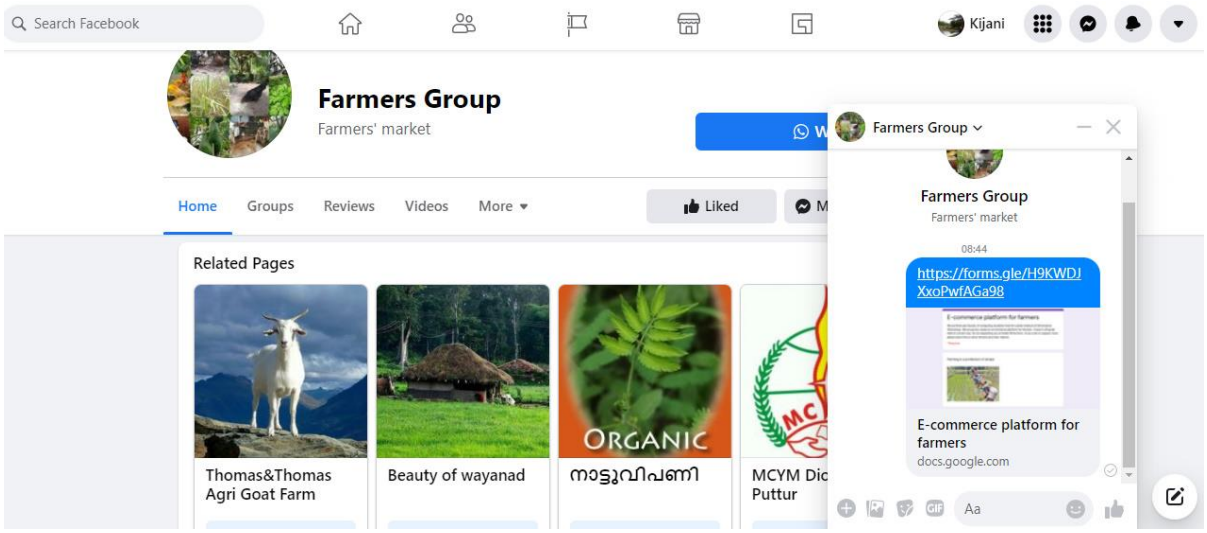
APPENDICES

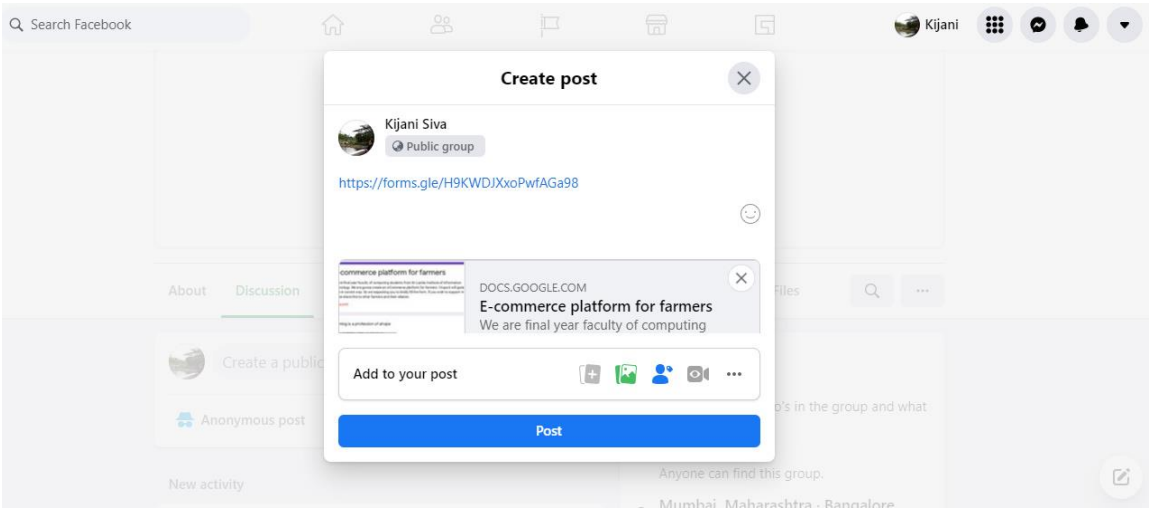
Gathering some information from social media farming groups

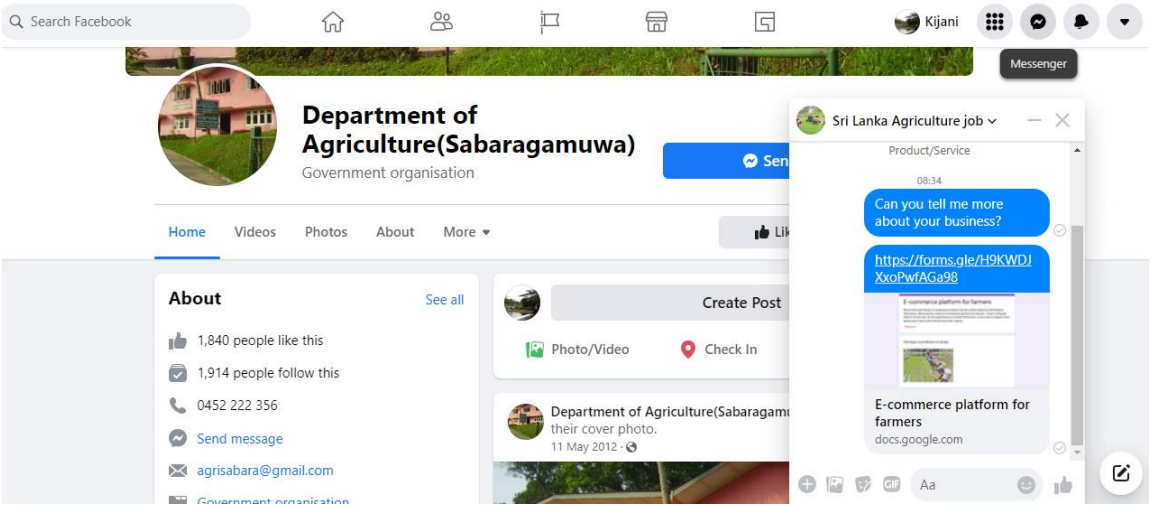





Search Facebook
Kijani
Innovative Farmers Group
Interest
Home
Reviews
About
Videos
More
Liked
Related Pages
SUPPORT FARMERS
देश की शान
मेरे किसान
Farmers Group
Indian Farmers
Organic Farming
Organic Farming-
Contract F
Innovative Farmers Group
Innovative Farmers Group
Interest
08:45
https://forms.gle/H9KWDJXxoPwfAGa98
E-commerce platform for farmers
docs.google.com
Aa


Search Facebook
Kijani
Farmers Group
Farmers' market
Home
Groups
Reviews
Videos
More
Liked
Related Pages
Thomas&Thomas Agri Goat Farm
Beauty of wayanad
ORGANIC
നാട്ടുവിപണി
MCYM Dic
Puttur
Farmers Group
Farmers Group
Farmers' market
08:44
https://forms.gle/H9KWDJXxoPwfAGa98
E-commerce platform for farmers
docs.google.com
Aa


Search Facebook
Kijani
Create post
Kijani Siva
Public group
https://forms.gle/H9KWDJXxoPwfAGa98
DOCS.GOOGLE.COM
E-commerce platform for farmers
We are final year faculty of computing
Add to your post
Post
About
Discussion
Anonymous post
New activity


Search Facebook
Kijani
Messenger
Department of Agriculture(Sabaragamuwa)
Government organisation
Home
Videos
Photos
About
More
About
See all
1,840 people like this
1,914 people follow this
0452 222 356
Send message
agrisabara@gmail.com
Create Post
Photo/Video
Check In
their cover photo.
11 May 2012
Sri Lanka Agriculture job
Product/Service
08:34
Can you tell me more about your business?
https://forms.gle/H9KWDJXxoPwfAGa98
E-commerce platform for farmers
docs.google.com
Aa