MSC IN IT

Machine Learning and Data Science for Decision Making in Crop Cultivation

S.P.T.P. Senadheera

189472R

Faculty of Information Technology

University of Moratuwa

Sri Lanka

August 2021

Declaration

I the undersigned solemnly declare that the project report "**Machine Learning and Data Science for Decision Making in Crop Cultivation**" is based on my own work carried out during the course of my study under the supervision of **Mrs. Indika Karunarathne**

I. I assert the statements made and conclusions drawn are an outcome of my research work. I further certify that
II. The work contained in the report is original and has been done by me under the general supervision of my supervisor.
III. The work has not been submitted to any other Institution for any other degree/diploma/certificate in this university or any other University of India or abroad.
IV. We have followed the guidelines provided by the university in writing the report.
V. Whenever we have used materials (data, theoretical analysis, and text) from other sources, we have given due credit to them in the text of the report and giving their details in the references.

S.P.T.P. Senadheera ....
Student Name Date

....
Student Signature

Supervised by

Mrs. Indika Karunarathne
Supervisor Name

....
Supervisor Signature

....
Date

Acknowledgement

This project would not have been possible without the support of many people. Many thanks to my adviser, Mrs. Indika Karunarathne, who read my numerous revisions and helped to make some sense of the confusion. Also thanks to the authorities of Department of Agriculture, Sabaragamuwa Province, who offered relevant details and support. Thanks to the University of Moratuwa, Faculty on Information Technology for awarding me a Dissertation Completion Fellowship to complete this project. And finally, thanks to parents, and numerous friends who endured this long process with me, always offering support and love.

Table of Contents

Abstract

In many situations, farmers face lots of difficulties due to not gaining enough income for their harvest as a result of it, considerable amount of harvest is wasted because of not having enough return on investment. Although government introduced many approaches, they seem to have been unable to address the issues of crop prediction and sustainable agriculture as farmers complain that the policies created by the authorities never match with the expectations of farmers' community. Considerable amount of population still engage in agricultural sector. Therefore, their contribution to the GDP has been representing an important portion through last few decades. Therefore this research has been conducted to support farmers to make optimum decisions based on various factors such as weather, price, and potential production to make cultivations. The research is based on two basic data mining models (Price and Expected Production) to predict the price and Production for a selected crop based on a particular district. The prediction outcome is presented using a mobile application with user friendly interfaces. Price model is based on the production (MT), per capita consumption, imports and inflation with the accuracy of 77.93% and the Production model is based on the climatic conditions such as humidity, wind speed, sun hours, rainfalls and temperature with the accuracy of 65.08%. The production model is to be further developed by adding the feature of soil moisture as it is an important factor to decide which crop is to be cultivated. Furthermore, an additional model is to be added to predict the amount of fertilizer to be added based to get the optimum yield.

Keywords:-crop, price, production, model, prediction, yield, fertilizer, moisture, accuracy

List of Figures

List of Tables

Chapter 1

Introduction

In many situations, farmers face lots of difficulties due to not gaining enough income for their harvest even to cover their cost. Even though past few governments have introduced some solutions, rather than the farmers, few intermediates and companies took undue advantage and even farmers didn’t have enough knowledge or information and a specific plan on crop cultivation based on the market price and the other natural conditions. In addition, farmers have had to deal with severe agricultural damage and destruction of food crops caused by natural disasters over the past few years.

Although government introduced many approaches, they seem to have been unable to address the issues of crop prediction and sustainable agriculture. Most of the occasions, farmers complain that the government policies do not prioritize nor have a dedicated strategy to support them. And also there seems to be lack of sources for farmers to make a decision on crop cultivation since they don’t have enough reliable resources to get the information for crop prediction.

Climatic changes and price fluctuation are among major challenge faced by the Sri Lankan Farmers. According to the 2019 Central Bank’s Annual Report, Agriculture contributes 7% for of Gross Domestic Product (GDP) and 31.2% of population still engage in Agricultural Sector [1, pp. 9,27]. Agriculture’s contribution is also very significant in Sri Lanka’s Exports sector. According to Central Bank’s Annual Report in 2019, Rs. Bn. 418.9 of Exports represents Agriculture [1, p. 10]. Even though Agriculture holds such important portion of the economy, the main link of agriculture, the farmers face difficulties in crop production. As mentioned above the price fluctuation and lack of proper ways to get information of price fluctuation cause heavily on crop prediction and crop planning. Below Table describes the daily wholesale Price levels of few selected vegetables for selected dates throughout first 4 months of 2020 [2].

Table 1 Daily Price Report-CBSL-2019

Vegetable	Prices for Selected Dates							
	Jan 31		Feb 28		March 31		April 30	
	Pettah	Dambulla	Pettah	Dambulla	Pettah	Dambulla	Pettah	Dambulla
Beans	103.00	95.00	90.00	95.0	103.00	75.00	240.00	243.00
Carrot	90.00	125.00	85.00	83.00	90.00	105.00	80.00	78.00
Cabbage	47.00	43.00	35.00	15.00	47.00	23.00	45.00	33.00

Therefore, after considering the above facts this research has been aimed towards a decision support system which helps farmers to plan their cultivations based on the market price and the potential production based on environment factors.

Chapter 2

Background and Motivation

By 2018, the Department of Agriculture has embarked on a vision to establish a vibrant and vibrant agricultural sector for food security and national prosperity to boost national food production in line with the 2030 Agenda for Sustainable Development. [3, p. 7]. Smooth food supply for a reasonable market price plays a major role to achieve that mission. Increasing the productivity of crop production, cost minimization and crop diversification using modern technology are very important to achieve above milestone.

In order to achieve the above goal, ministry has implemented various Food Production and Drive Programs from 2018-2020. These programs include

- Off Season Product Cultivation Promotion Programs.
- Productivity Improvement Programs
- Supply Chain Management Programs for Agricultural Products.

Etc [3, p. 32].

In addition, the department has implemented research and development programs to introduce new crop varieties such as Cluster onions, one Mung bean, two varieties of Woodapple and two varieties of Pummelo etc. [3, p. 68]. Apart from that the department seems to have taken some steps to encourage farmers to establish 15 fruit orchards and to distribute and distribute agricultural equipment to maintain existing fruit villages. [3, p. 68].

But the agricultural sector, including thousands of farmers, faces many challenges. Although many applications and research programs have been taken up by the relevant authorities, it seems that most research programs are kept in laboratories and do not have sufficient access to farmland. [4]. In addition, some projects or programs do not have the resources and resources to change the needs of the farming community in order to find answers to important questions such as how much will be produced and how much they will cost. Price instability is one of major issues faced by the farming community. Continuously changing climatic conditions increases this issue. Nearly 72% of Sri Lankan population lives in rural areas and 30% of labor force lie in paddy and the other agricultural cultivations, agricultural production including paddy and the other vegetables is the main occupation. Recent researches suggest that the price instability not only affects famers' income, but also makes an impact on country's exports [5].

Below table indicates the crop production of Sri Lanka from 2008 to 2017 [6]

Crop	Production, '000 metric tons (MT)									
	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017
Paddy	3,875	3,652	4,301	3,894	3,846	4,621	3,381	4,819	4,420	2,383
Other Field Crops										
Manioc	240	278	283	293	291	303	310	324	324	306
Maize	112	130	167	138	202	209	241	261	244	196
Potatoes	75	62	52	59	72	79	83	97	96	73
Big Onions	57	82	59	46	84	70	111	89	65	53
Red Onions	50	46	62	72	74	56	63	61	64	58
Groundnuts	10	13	14	17	22	27	25	29	24	22
Minor Export Crops										
Cinnamon	15	16	16	18	17	18	18	18	20	22
Pepper	13	16	17	11	19	28	19	28	18	30
Cloves	8	3	10	6	4	6	3	5	2	6
Plantation Crops										
Tea	319	291	331	328	328	340	338	329	293	308
Rubber	129	137	153	158	152	130	99	87	79	83
Coconut	380	367	251	386	351	379	716	552	765	466
Source: Central Bank of Sri Lanka. Economic and Social Statistics 2018.										

Figure 1 Crop Production in Sri Lanka

Since this research is based on crop production planning based on the market price, the producers (Farmers) need to understand the upcoming weather conditions, price fluctuations. But it reported that Yield stagnation in almost all other food crops productivity limit food supplies .The Below Chart indicates the price fluctuations of the selected crops taken from the Central Bank's Daily Price Report.

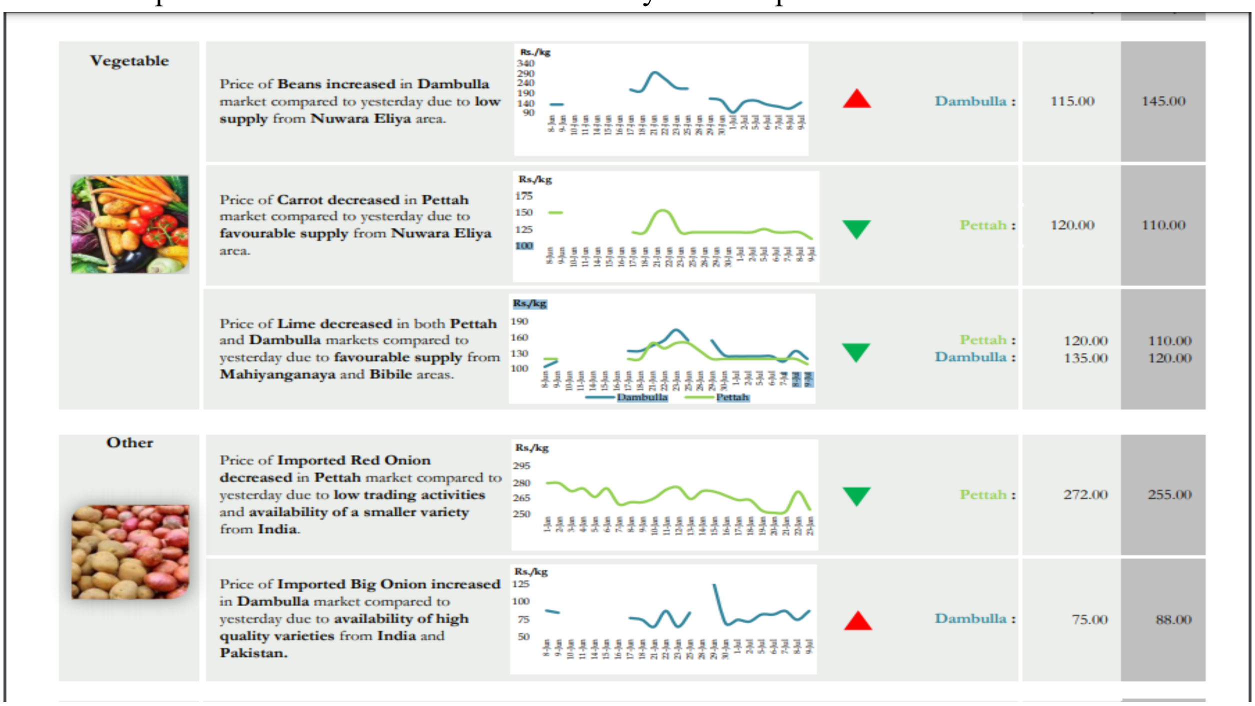


Figure 2 Daily Price Fluctuation of Selected Crops (CBSL-Daily Price Report-2021)

Above chart suggests that the price is highly unstable and which is fluctuated daily. But this kind of information is barely accessed by the farmers therefore they just plan their crops based on their budget, previous experiences and with limited experiences. But the key factors like market price and the environmental conditions are changed daily for which the farmers don’t have enough access. Researchers suggest that poor crop prediction and harvesting practices have resulted in 20-40% loss of perishable foods such as fruits, vegetables, milk etc. [6]. A recent report in the local newspaper “Aruna” on July 9, 2021 indicates that the loss of vegetables and fruits after harvest is about 40%.

Figure 3 Crop Wastage Amount on Media

Agricultural officers point out that there is no proper mechanism to convey message of market demand, crop price information to the farmers, crop suggestions based on climatic conditions. Department of agriculture website points out that there are lots of options to go for farmers as cultivations as they have decent market prices but most of the small scale farmers don’t have proper access to this information [7].

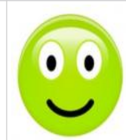
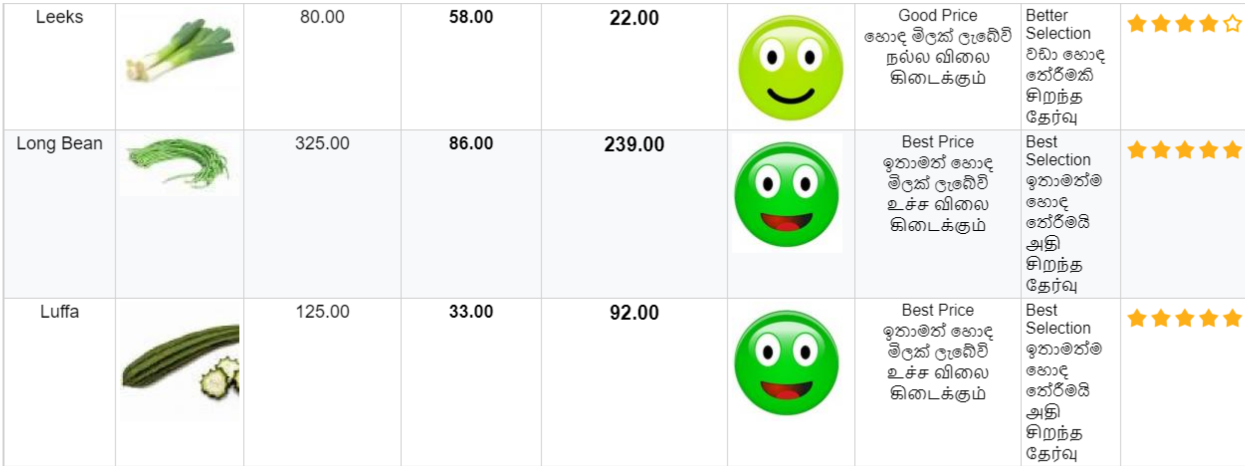
		/ இரண்டு வாரங்களுக்கான Recomended biweekly Ext.(ha.)	அளவு (ஹெக்.)/ These 2 week Cul.Ext. (ha.)	extent that can be cultivated within these two weeks		விலை எதிர்வுகூறல் Price Prediction for farmers who cultivate within these 2 weeks	Recomendation	
Beans		350.00	**257.00**	**93.00**		Good Price හොඳ මිලක් ලැබේවි நல்ல விலை கிடைக்கும்	Better Selection වඩා හොඳ තේරීමකි சிறந்த தேர்வு	★★★★☆
Beet Roots		100.00	**41.00**	**59.00**		Good Price හොඳ මිලක් ලැබේවි நல்ல விலை கிடைக்கும்	Better Selection වඩා හොඳ තේරීමකි சிறந்த தேர்வு	★★★★☆
Bitter gourd		135.00	**56.00**	**79.00**		Good Price හොඳ මිලක් ලැබේවි நல்ல விலை கிடைக்கும்	Better Selection වඩා හොඳ තේරීමකි சிறந்த தேர்வு	★★★★☆
Brinjal		450.00	**126.00**	**324.00**		Best Price ඉතාමත් හොඳ මිලක් ලැබේවි உச்ச விலை கிடைக்கும்	Best Selection ඉතාමත්ම හොඳ තේරීමයි அதி சிறந்த தேர்வு	★★★★★

Figure 4 Price and Suggested Amount of Hectare 1

Cabbage		175.00	**98.00**	**77.00**		Good Price හොඳ මිලක් ලැබේවි நல்ல விலை கிடைக்கும்	Better Selection වඩා හොඳ තේරීමකි சிறந்த தேர்வு	★★★★☆
Capsicum		140.00	**64.00**	**76.00**		Good Price හොඳ මිලක් ලැබේවි நல்ல விலை கிடைக்கும்	Better Selection වඩා හොඳ තේරීමකි சிறந்த தேர்வு	★★★★☆
Carrot		125.00	**108.00**	**17.00**		General Price සාමාන්‍ය මිලක් ලැබේවි சாதாரண விலை கிடைக்கும்	Good Selection හොඳ තේරීමකි நல்ல தேர்வு	★★★☆☆
Cucumber		125.00	**43.00**	**82.00**		Best Price ඉතාමත් හොඳ මිලක් ලැබේවි உச்ச விலை கிடைக்கும்	Best Selection ඉතාමත්ම හොඳ තේරීමයි அதி சிறந்த தேர்வு	★★★★★

Figure 5 Price and Suggested Amount of Hectare 2



Leeks		80.00	**58.00**	**22.00**		Good Price හොඳ මිලක් ලැබේවි நல்ல விலை கிடைக்கும்	Better Selection වඩා හොඳ තේරීමකි சிறந்த தேர்வு	★★★★☆
Long Bean		325.00	**86.00**	**239.00**		Best Price ඉතාමත් හොඳ මිලක් ලැබේවි உச்ச விலை கிடைக்கும்	Best Selection ඉතාමත්ම හොඳ තේරීමයි அதி சிறந்த தேர்வு	★★★★★
Luffa		125.00	**33.00**	**92.00**		Best Price ඉතාමත් හොඳ මිලක් ලැබේවි உச்ச விலை கிடைக்கும்	Best Selection ඉතාමත්ම හොඳ තේරීමයි அதி சிறந்த தேர்வு	★★★★★

Figure 6 Price and Suggested Amount of Hectare 2

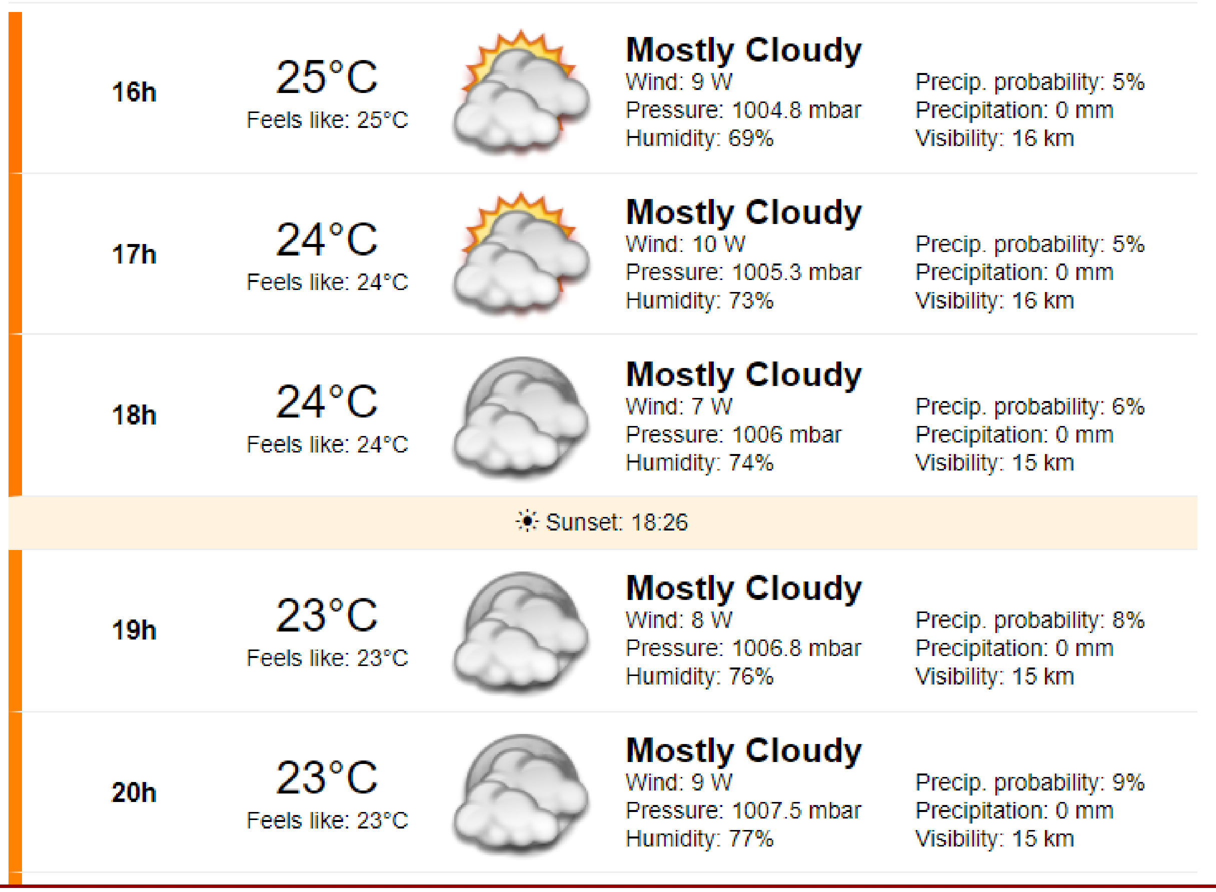
According to the above images, few vegetable items can be identified as they have a good demand and higher price levels but farmers are not aware of them [7]

Table 2 Expected Price Levels and % Hectares

Crop	% of Hectares cultivated	Price Levels
Beet Root	41	Good Price
Bitter Guard	41	Good Price
Brinjol	28	Very Good Price
Cucumber	34	Very Good Price
Long Bean	21	Very Good Price
Luffa	26	Very Good Price

Above images also indicate that the most of upcountry farmers tend to grow carrot as around 86% of suggested hectares have been cultivated even though the price level and the demand is very low [7].

Weather is also another key factor to decide crop. From the past, farmers have been using their previous experiences on weather conditions to plan crops in future. But the climatic conditions are changed frequently. These climatic changes make a huge impact on the environmental factors to grow crops. In such farmers' previous experiences on weather conditions might not be enough. As a result of it, they need a proper mechanism to predict future weather conditions to decide whether a particular crop is ideal to be cultivated. Below are pictures of the daily and hourly weather of Nuwara Eliya with Badulla taken from the Official Website of the Department of Meteorology in Sri Lanka. Such weather information is very important for farmers because those climates play an important role in making decisions for planting crops.



16h	25°C Feels like: 25°C	Mostly Cloudy	Wind: 9 W Pressure: 1004.8 mbar Humidity: 69%	Precip. probability: 5% Precipitation: 0 mm Visibility: 16 km
17h	24°C Feels like: 24°C	Mostly Cloudy	Wind: 10 W Pressure: 1005.3 mbar Humidity: 73%	Precip. probability: 5% Precipitation: 0 mm Visibility: 16 km
18h	24°C Feels like: 24°C	Mostly Cloudy	Wind: 7 W Pressure: 1006 mbar Humidity: 74%	Precip. probability: 6% Precipitation: 0 mm Visibility: 15 km
Sunset: 18:26				
19h	23°C Feels like: 23°C	Mostly Cloudy	Wind: 8 W Pressure: 1006.8 mbar Humidity: 76%	Precip. probability: 8% Precipitation: 0 mm Visibility: 15 km
20h	23°C Feels like: 23°C	Mostly Cloudy	Wind: 9 W Pressure: 1007.5 mbar Humidity: 77%	Precip. probability: 9% Precipitation: 0 mm Visibility: 15 km

Figure 7 Hourly Weather Forecast

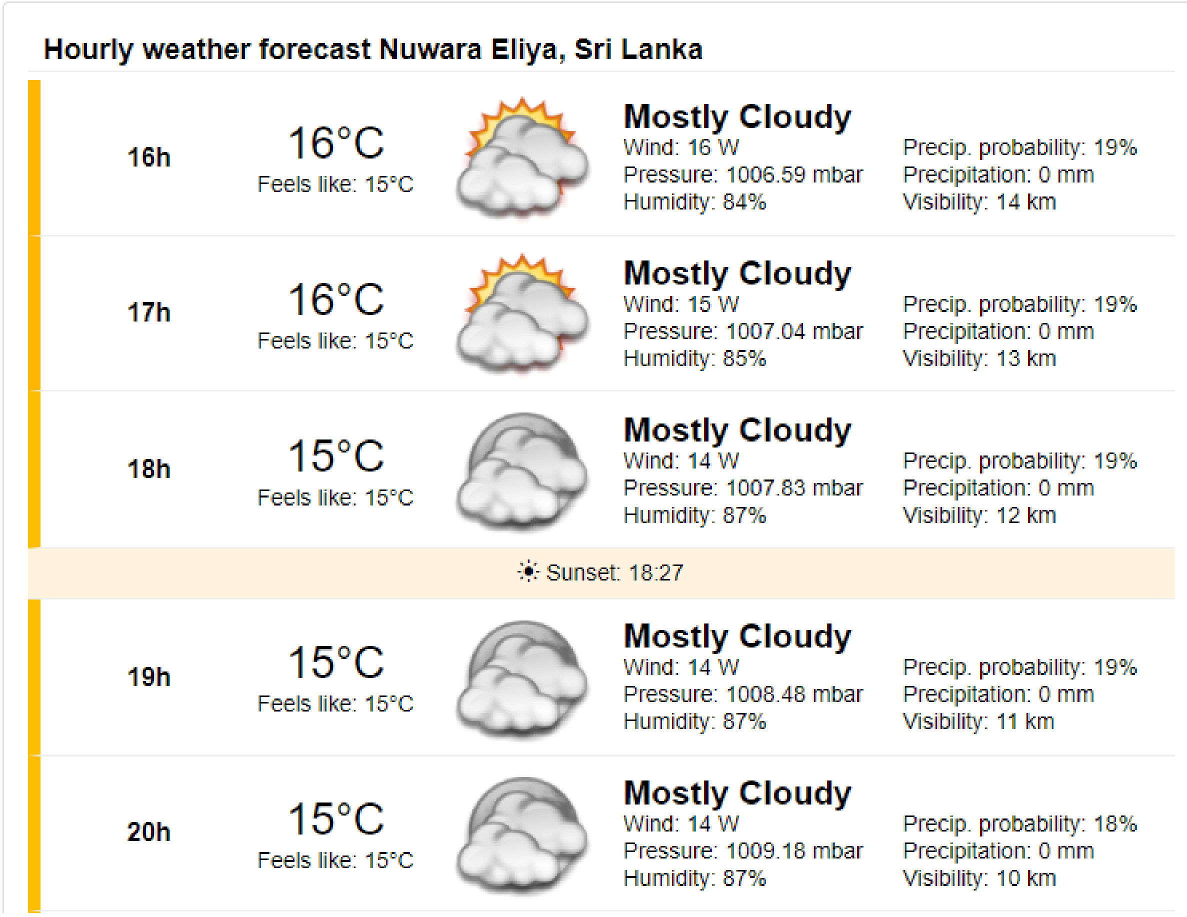


Figure 8 Hourly Weather Forecast- Nuwara Eliya District

This study therefore aims to bring the price of sensitive information and future weather conditions to the farmer so that he can make good crop decisions by taking into account the insufficient attention in agricultural divisions to harvest crops that have a better chance of earning money.

Chapter 3

Aim and Objectives

Aim

To present successful crop prediction model in user friendly manner based on the market price and weather conditions

Objectives

1. Data collection of crop prices from reliable data sources.
2. Analysis of crop prices to suggest crop plan based on the current inflation rate
3. Conveying crop plans via a mobile application for farmers for crop prediction with upcoming weather forecast.
4. Successful prediction of market price movement for a particular crop with potential production amount and the potential income

Chapter 4

Methodology

The application provides some useful insights on a particular crop based on two data model

- Price Prediction Model
- Potential Production prediction model

Building Price Model

Data Integration

The price forecast model has been developed and tested based on factors such as Average Production, Average Imports, inflation and consumption. Data collection has been a challenging part of modeling but with the resources of the Department of Statistics and Statistics, the World Bank [8], Helgi Library [9], Annual Statistics of Department of Agriculture [10, p. 38], the model has been built successfully to predict the price.

Table 03: Movements of the NCPI (Base: 2013=100)

		All Item			
				Inflation %	
Year	Month	Index Number	% Change Month on Month	Y on Y	12 Month Moving Avg.
2020	January	137.0	1.5	7.6	4.1
	February	137.0	0.0	8.1	4.5
	March	135.2	-1.3	7.0	4.9
	April	134.8	-0.3	5.9	5.1
	May	135.4	0.4	5.2	5.2
	June	137.3	1.4	6.3	5.6
	July	137.3	0.0	6.1	5.9
	August	137.8	0.4	6.2	6.1
	September	138.9	0.8	6.4	6.2
	October	139.1	0.1	5.5	6.2
	November	139.8	0.5	5.2	6.3
	December	141.2	1.0	4.6	6.2
2021	January	142.1	0.6	3.7	5.8

Figure 9 Monthly Inflation

Above table indicates the monthly calculated inflation based on NCPI (National Consumer Price Index) [11, p. 4]. But since the inflation had be calculated daily to match the daily price levels, average inflation has been calculated from the inflations which have been recorded from the beginning and the end of a particular month and using the standard deviation which has been calculated using the average inflation random realistic values have been generated to as daily inflation for each month using a web application [12]

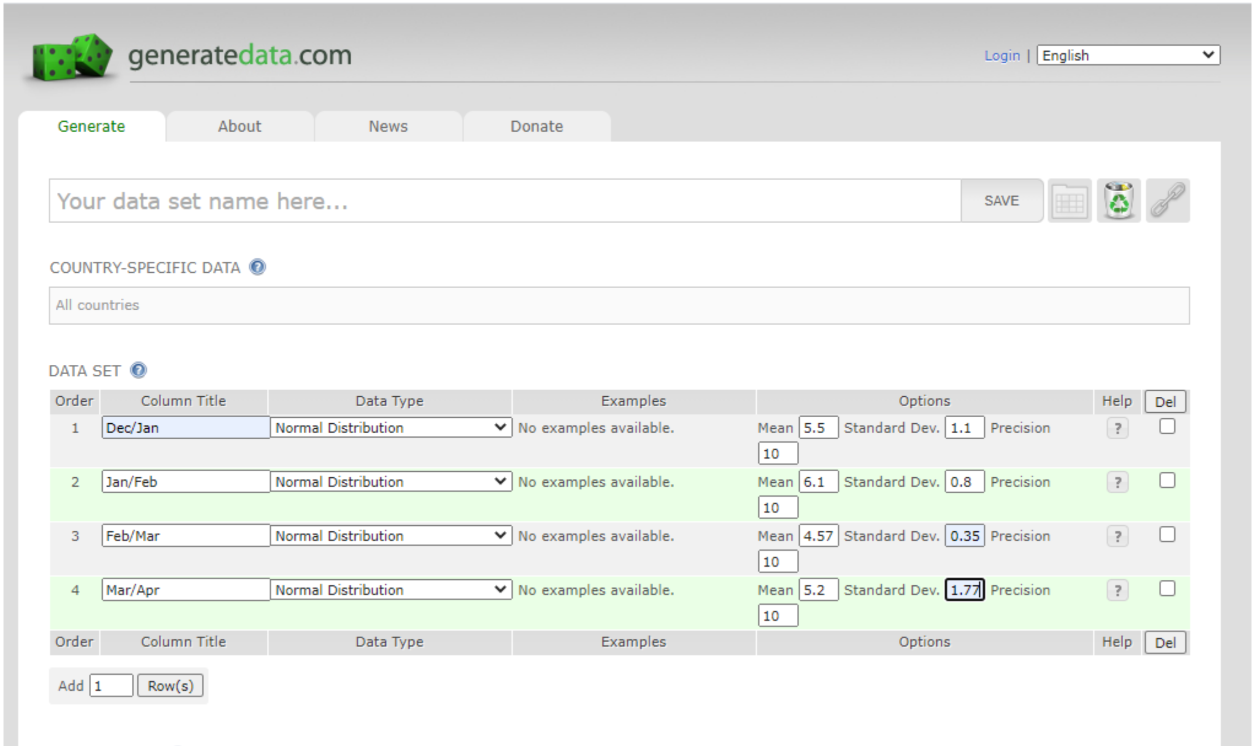


Figure 10 Random Realistic Number Generation

After gathering the information, the dataset has been created based on which the model was created to predict the price.

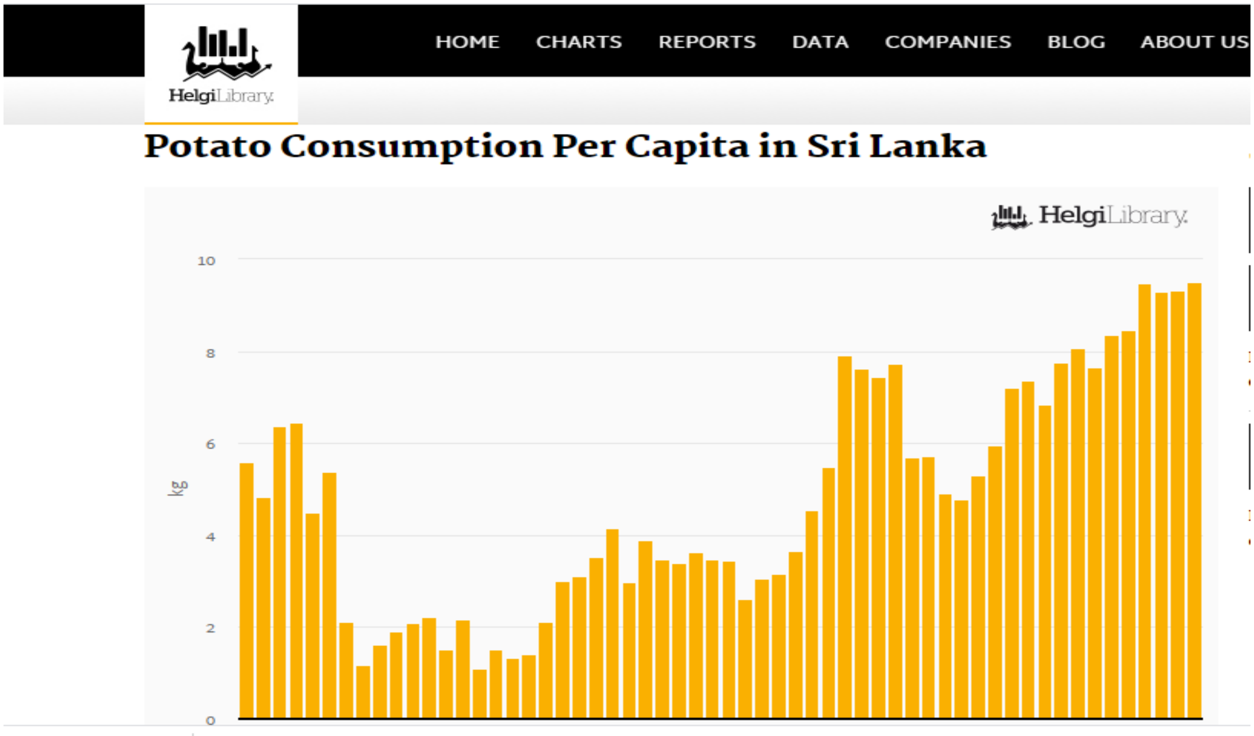


Figure 11 Helgi Library Website is a reliable source to get consumption details

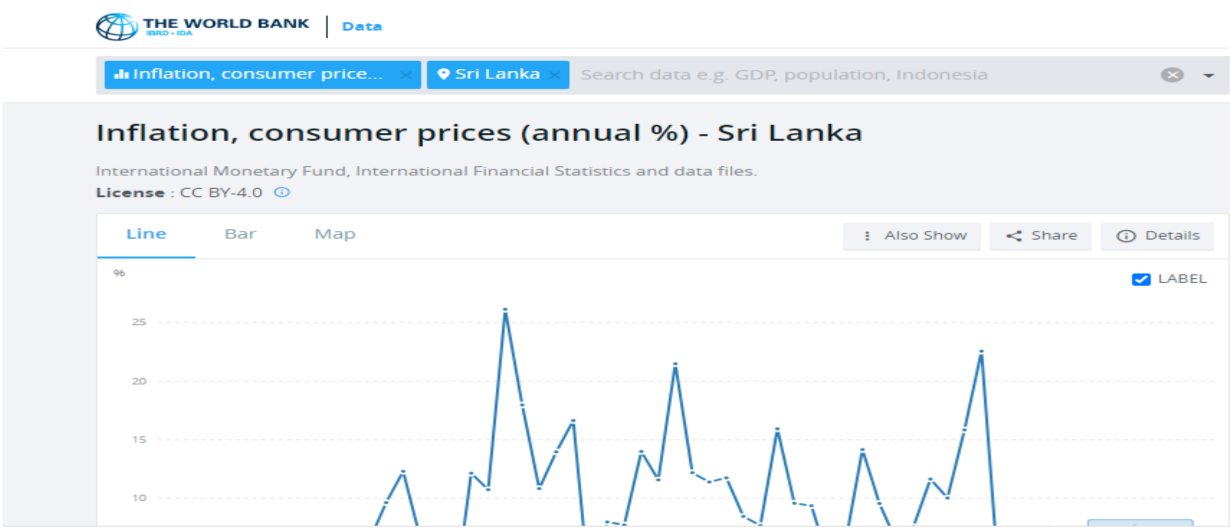


Figure 12 Inflation Statistics from World Bank Data

4.2 IMPORTS OF OTHER FIELD CROPS

Crop	2016			2017*		
	Quantity (t)	Value (Rs.'000)	CIF Value (Rs/kg)	Quantity (t)	Value (Rs.'000)	CIF Value (Rs/kg)
Dry chiili	51,018	13,477,137	264.16	51,692	7,881,723	152.48
Big onion	215,593	6,796,127	31.52	232,318	12,117,950	52.16
Red onion	19,842	1,306,678	65.85	10,911	1,284,965	117.76
Potato (seed)	1,841	241,043	130.96	1,611	236,074	146.50
Potato (other)	148,081	4,596,036	31.04	151,438	5,440,129	35.92
Maize (seed)	1,112	603,130	542.48	1,432	918,204	641.00
Maize (other)	41,946	1,571,524	37.47	179,589	6,627,909	36.31
Kurakkan	2,178	123,328	56.61	2,918	242,782	83.20
Gingelly	129	18,930	146.69	223	27,221	122.32
Soya bean	7,126	433,880	60.89	3,176	214,014	67.38
Green gram	13,862	2,265,997	163.47	15,541	2,269,422	146.03
Cowpea	6,055	649,695	107.29	7,345	993,971	135.53
Ground nut	3,095	552,955	178.67	3,876	689,463	17789
Red lentil (whole)	130,555	16,672,774	127.71	136,432	13,627,742	99.89
Red lentil (split)	21,537	3,311,814	153.78	26,744	3,342,699	124.99
Yellow lentil (whole)	1,738	237,494	136.65	417	48,916	117.36
Yellow lentil (split)	511	75,048	146.90	2,286	385,125	168.47
Black gram	11,991	3,265,393	272.32	12,767	1,993,448	156.14
Chickpea (whole)	20,879	3,687,625	176.62	19,985	4,222,886	211.30
Chickpea (split)	6,078	696,809	114.64	6,552	755,806	115.36

Source: Sri Lanka Customs ; (Statistics Division) * Provisional

Figure 13 Imports Statistics

With above mentioned data resources, the model has been successfully created to predict the price as follows.

Average Local Production (MT)	Imports	Inflation	Consumption	Price Fluctuation
1709.26	146.7	5.2	8.2	150
1882.58	146.6	5.4	8.6	150
2030.51	146.8	5.1	8.2	150
1643.94	146.6	5.5	8.7	150
2115.97	146.3	5.5	8.4	150
791.87	146.5	5.5	8.6	150
2064.21	147.2	4.9	8.4	150
1978.6	146.2	5.3	8.6	157.82
2462.62	147.1	5	8.4	158.27
2036.84	146.9	5.1	8.5	155.86
1964.2	146.2	5.4	8.4	157.09
1749.66	146.6	4.8	8.8	157.85
1321.79	146.3	5.2	8.5	156.65
3002.44	146.5	5.1	8.6	150.95
2056.38	146.8	5.2	8.5	149.92
17.44	149.6	5.7	8.4	150.96
32.77	150.2	5.2	8.3	150.41
32.84	149.8	5.2	8.6	150.76
13.49	150.1	5.3	8.5	151.26
22.92	149.9	5.7	8.6	150.82
37.77	149.8	4.8	8.9	140.45
25.38	148.9	5.3	8.4	140.07
31.95	149.8	5.6	8.3	145.44
29.21	149.6	5.4	8.4	135.61

Figure 14 Created Data Model

Prediction Based on the Model

According the price model price is the dependent variable is predicted based on Annual Production, Imports, Consumption and the inflation. With the nature of the prediction, it is obvious that the multiple linear regression model is the ideal platform predict the price [13]. Since python is very popular for data analysis with Pandas [14]. The price prediction has been conducted using Python Pandas.

```python
#Multiple Linear Regression Model to predict Price
DataSetPrice = pd.read_csv('E:/MSc IT/2Nd Part/FinalProject/NewDataSets/ExcelFilesSet2/CSVFiles/PriceDataSet.csv')
print(DataSetPrice)

#Fetching data from worldbank data using pandas datareader package
data = wb.download(indicator='FP.CPI.TOTL.ZG',country='LK', start=2020, end=2020)
data = data.reset_index(1)
data.columns = ['Year', 'Inflation']
# # #print(type(data))
inflation = data['Inflation']
DataSetPrice.isnull().values.any()
# # # # #
DataSetPrice = DataSetPrice.dropna()#Dropping null values of the dataset
# # # # #
print(DataSetPrice.isnull().values.any())
# # # # #
inputInflation = DataSetPrice.drop(['Price Fluctuation'],axis = 'columns')
# #inputInflation = DataSetPrice['Inflation']
targetPrice = DataSetPrice['Price Fluctuation']
# # # # # # #
print("Input Information")
print(inputInflation)
print(" ")
print("Price Information")
print(targetPrice)
```

Figure 16 Python Code to build and predict the model

```
     Average Local Production (MT)  Imports  ...  Consumption  Price Fluctuation
0                          1709.26    146.7  ...          8.2             150.00
1                          1882.58    146.6  ...          8.6             150.00
2                          2030.51    146.8  ...          8.2             150.00
3                          1643.94    146.6  ...          8.7             150.00
4                          2115.97    146.3  ...          8.4             150.00
..                             ...      ...  ...          ...                ...
679                           0.00    153.4  ...          9.4              89.71
680                           0.00    155.7  ...          9.4              92.66
681                           0.00    153.9  ...          9.4              90.84
682                           0.00    152.8  ...          9.4              92.38
683                           0.00    155.0  ...          9.5              90.81

[684 rows x 5 columns]
False
```

Figure 15 Independent Variables of the model

Price Model Prediction Results

```
     Average Local Production (MT)  Imports  ...  Consumption  Price Fluctuation
0                          1709.26    146.7  ...          8.2             150.00
1                          1882.58    146.6  ...          8.6             150.00
2                          2030.51    146.8  ...          8.2             150.00
3                          1643.94    146.6  ...          8.7             150.00
4                          2115.97    146.3  ...          8.4             150.00
..                             ...      ...  ...          ...                ...
679                           0.00    153.4  ...          9.4              89.71
680                           0.00    155.7  ...          9.4              92.66
681                           0.00    153.9  ...          9.4              90.84
682                           0.00    152.8  ...          9.4              92.38
683                           0.00    155.0  ...          9.5              90.81

[684 rows x 5 columns]
False
```

Figure 17 Displaying the contents of the CSV file

```
Input Information
     Average Local Production (MT)  Imports  Inflation  Consumption
0                          1709.26    146.7        5.2          8.2
1                          1882.58    146.6        5.4          8.6
2                          2030.51    146.8        5.1          8.2
3                          1643.94    146.6        5.5          8.7
4                          2115.97    146.3        5.5          8.4
..                             ...      ...        ...          ...
679                           0.00    153.4        5.1          9.4
680                           0.00    155.7        5.4          9.4
681                           0.00    153.9        4.9          9.4
682                           0.00    152.8        4.7          9.4
683                           0.00    155.0        5.7          9.5

[684 rows x 4 columns]

Price Information
0      150.00
1      150.00
2      150.00
3      150.00
4      150.00
        ...
679     89.71
680     92.66
681     90.84
```

Figure 18 Price Information is the dependent variable which predicted based on the inputs shown in above figure

```
Name: Price Fluctuation, Length: 684, dtype: float64
[90.56796237]
0.7793312789396525
```

Figure 19 Accuracy of the model

Data supplied to the model to predict the value and accuracy of the model has been tested. The accuracy of the model has been approximately 78% reasonable as the result of the model is almost 78% accurate.

Production Prediction Model

Data Integration

Predicting the potential production is the other key component of the application. Production model is based on weather factors such as Temperature, Windspeed, Rainfalls, Sunhours, Humidity, Sunhours, Cloud. The weather information is available on www.worldweatheronline.com webite [15]. Country based customized weather information can be extracted for each month under district or regional base from the website [15].Below images indicate the information taken district wise to build the production model.

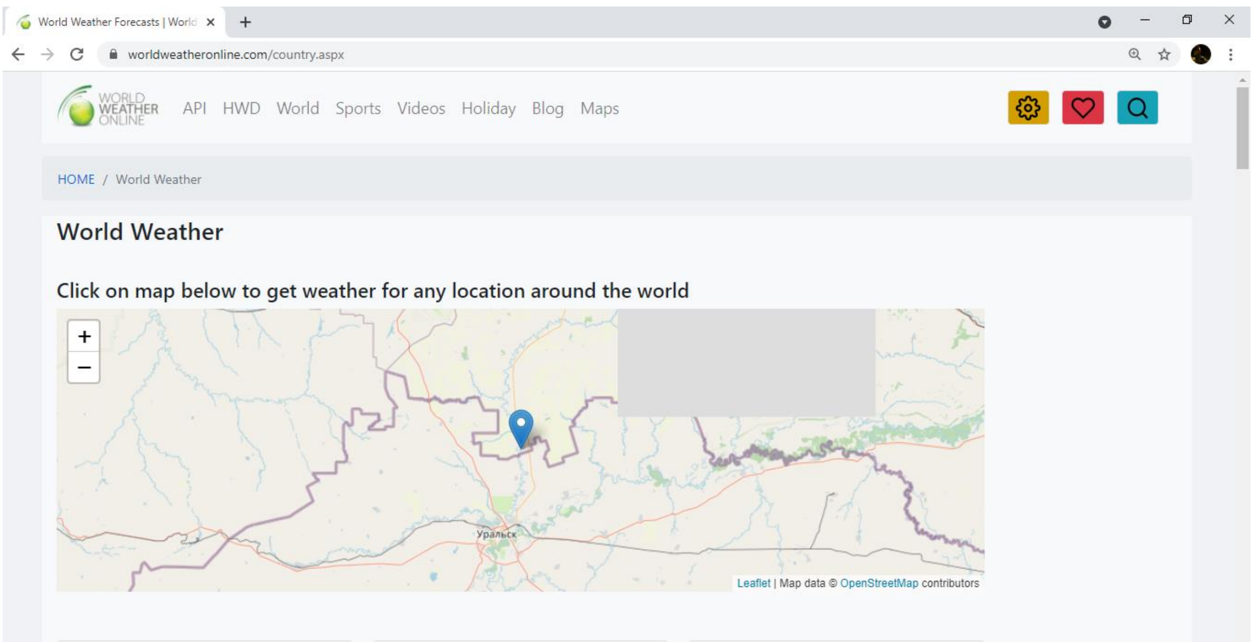


Figure 20 Home Page of the website from which the weather information has been extracted.

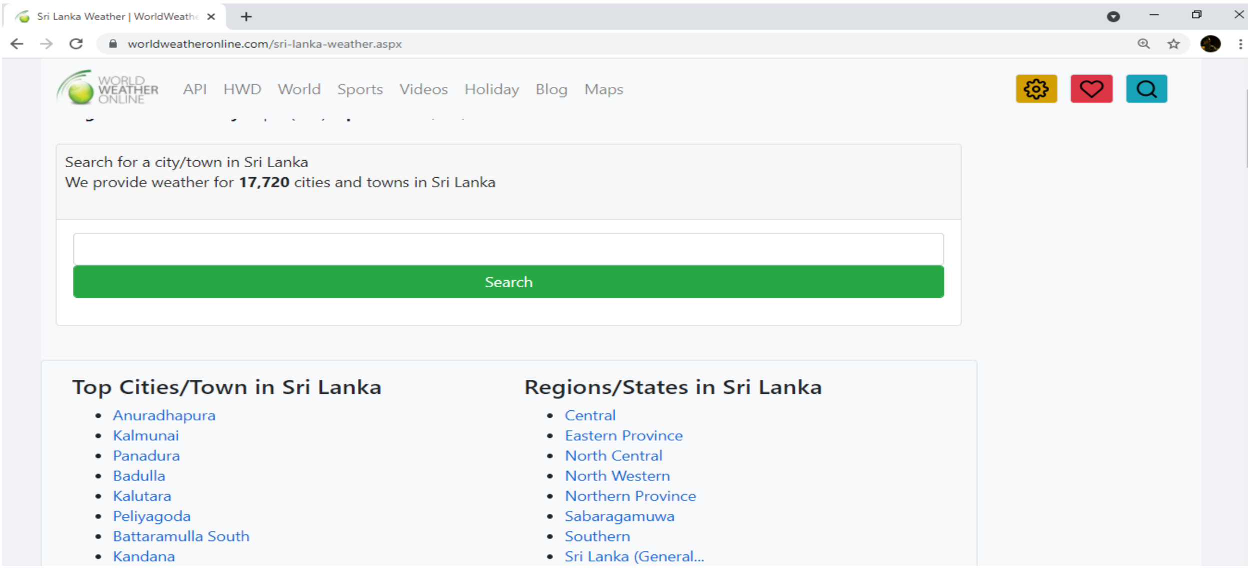


Figure 21 District, City and Regional wise weather information has been extracted from this Interface.

Below Images indicate the weather information which has been extracted from Nuwara Eliya District

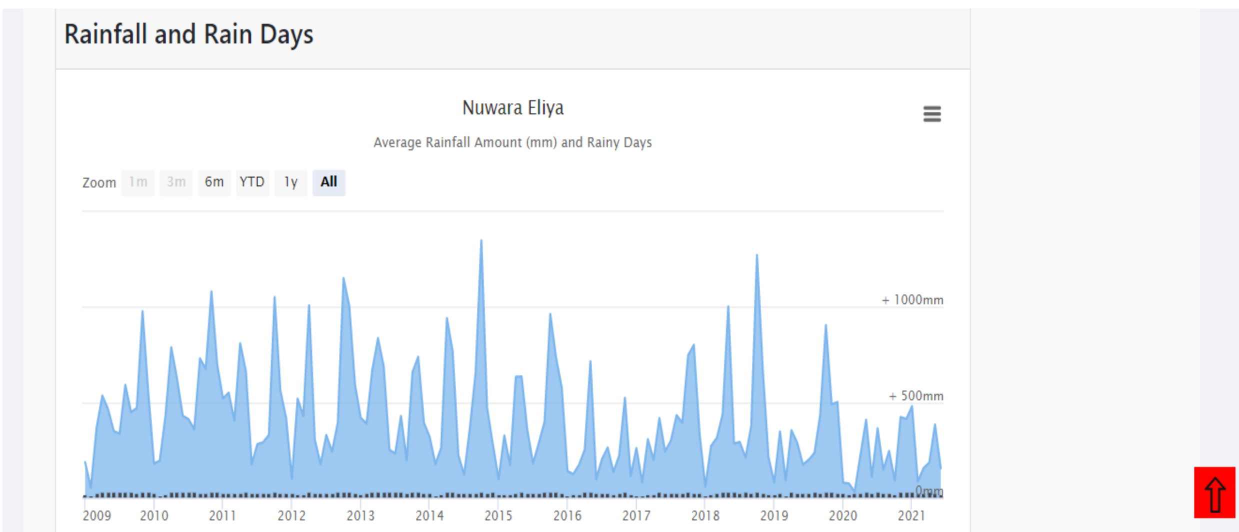


Figure 22 Average Rain falls

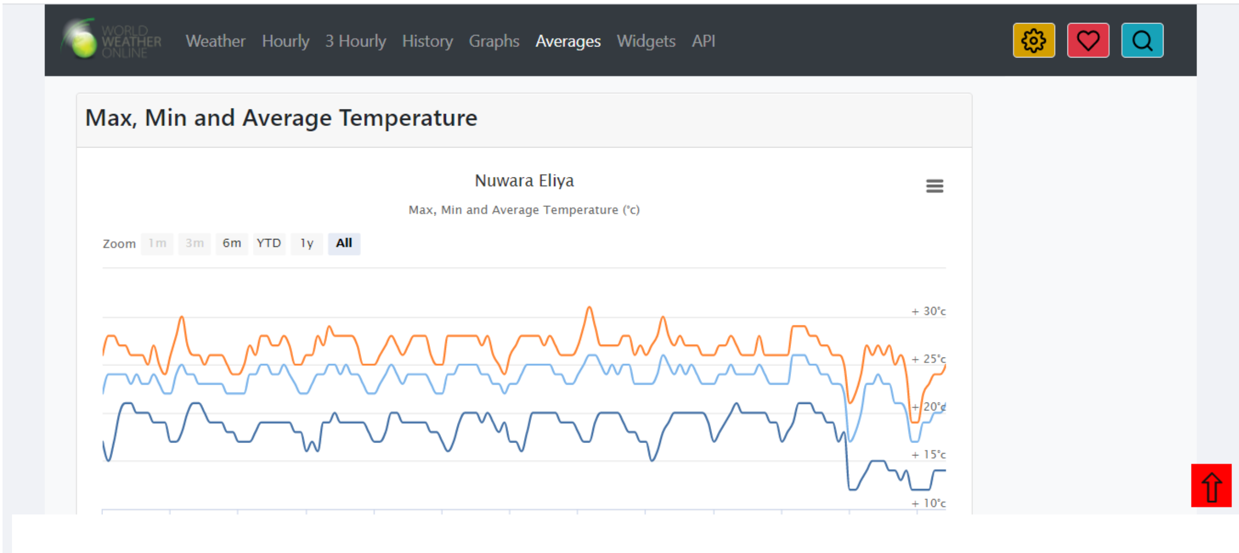


Figure 23 Average Rail Falls

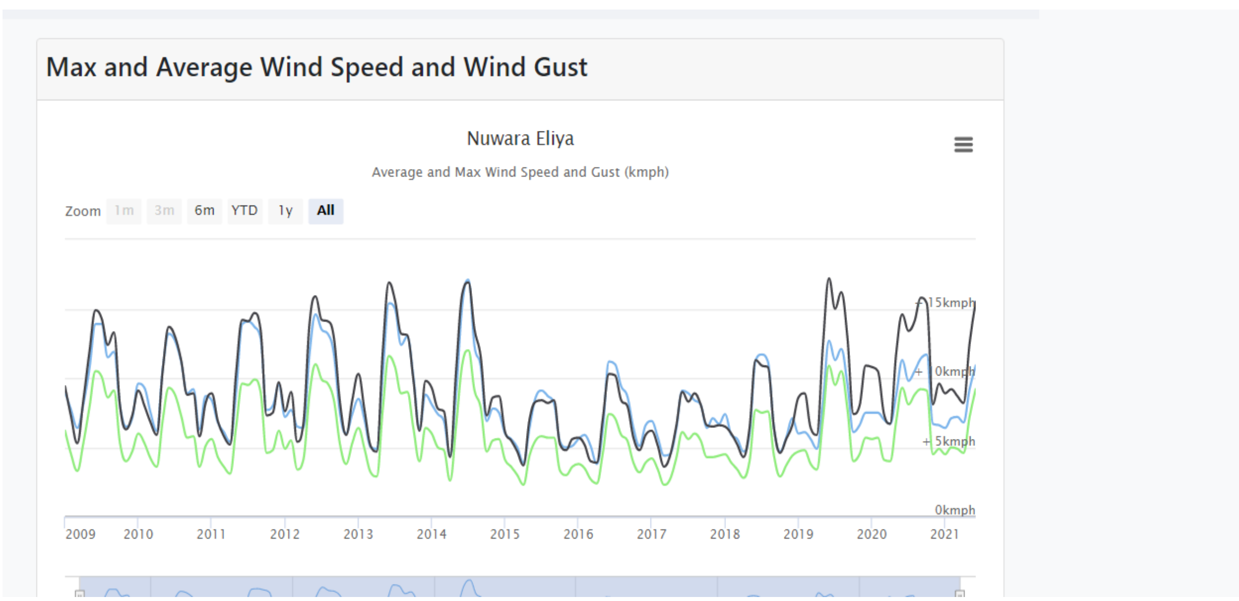


Figure 24 Average Wind Speed

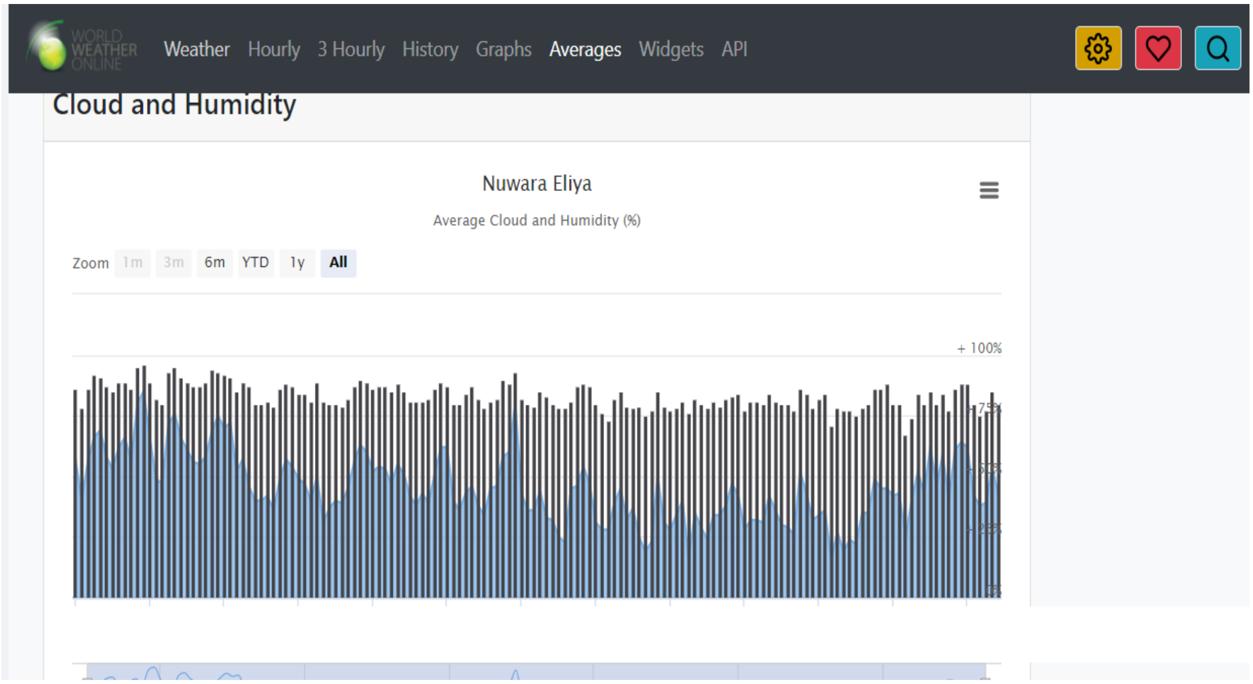


Figure 26 Average Humidity

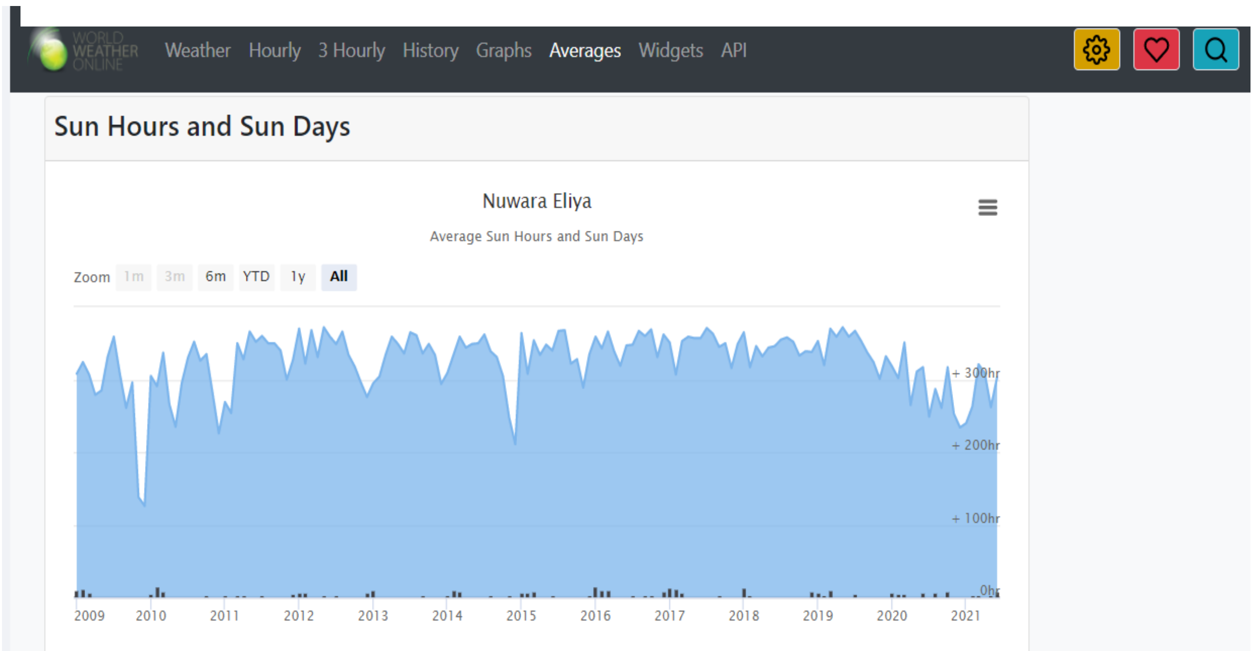


Figure 25 Average Sun Hours

Monthly data for a particular year has been extracted and using the mean value and standard deviation, random daily values have been generated to get daily weather conditions to build the model [12].

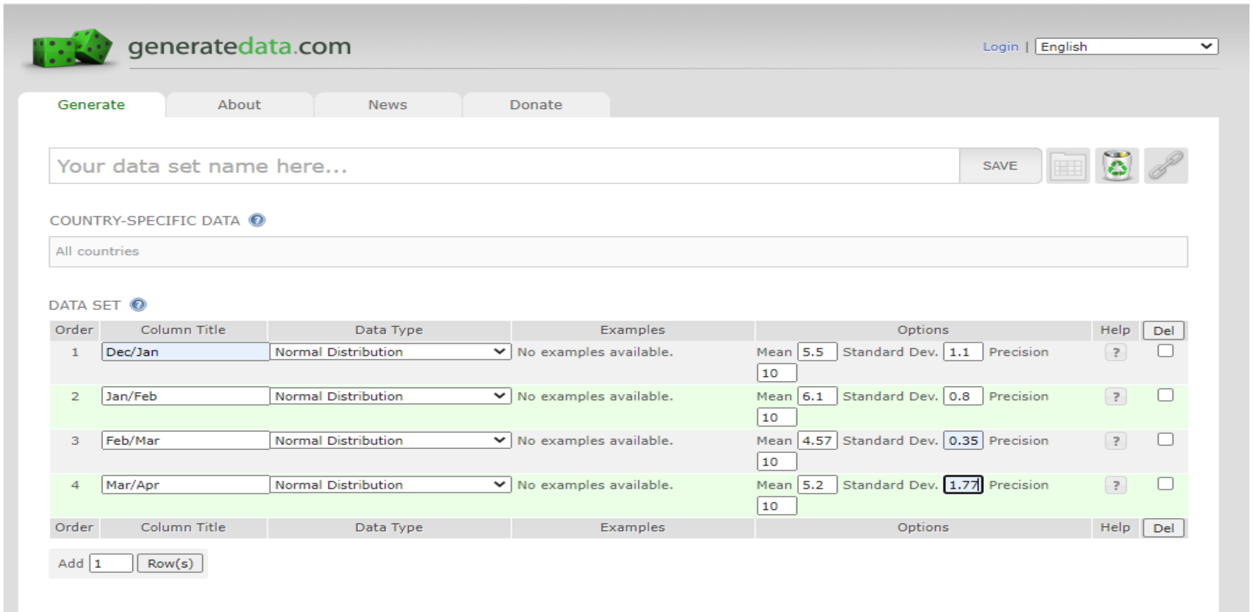
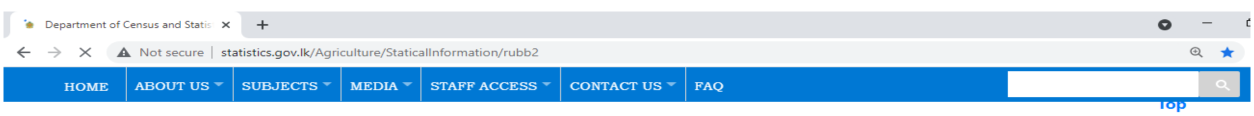


Figure 27 Random Realisric Number Generation

Weather data compiled from www.worldweatheronline.com as input. Historical details are taken from the official website of the Department of Statistics and Statistics [16].

District wise monthly Production information for each year for a particular crop is available in that website.



CO3 – 3: Monthly Production of Potato

Unit: MT .

Year	District	Nov/Dec	Dec/Jan	Jan/Feb	Feb/Mar	Mar/Apr	Apr/May	May/Jun	Jun/Jul	Jul/Aug	Aug/Sep	Sep/Oct	Oct/Nov	Total
2006/2007	National Total	3,126	922	848	10,001	15,845	8,886	844	816	2,348	19,207	9,342	2,910	75,263
	Nuwara Eliya	1,738	757	185	-	95	7,504	844	816	453	346	425	2,283	15,446
	Badulla	1,388	165	663	10,001	15,750	1,382	-	-	1,895	18,861	8,917	627	59,649
	Other District	-	-	-	-	-	-	-	-	-	-	-	-	168
2007/2008	National Total	5,003	2,134	426	7,216	8,314	5,268	2,487	595	6,819	19,781	11,256	4,862	74,398
	Nuwara Eliya	2,812	1,917	426	887	3,040	2,321	2,487	595	630	2,382	1,678	2,454	21,629
	Badulla	2,191	217	-	6,329	5,274	2,947	-	-	6,189	17,399	9,578	2,408	52,532
	Other Districts	-	-	-	-	-	-	-	-	-	-	-	-	237
2008/2009	National Total	1,506	581	-	14,036	10,537	2,462	1,387	-	-	1,348	15,479	13,001	60,848

Figure 28 District Wise Annual Crop Production Information

Based on such information, the data model to predict the production has been built as follows.

	A	B	C	D	E	F	G
1	Average T	Average R	Average V	Average C	Average H	Average S	Average Production (MT)
2	21	75.75	3.65	41.8	82.92	202.55	680.31
3	21	80.28	3.66	41.75	82.02	203.18	772.1
4	21	72.17	3.65	41.8	82.39	200.05	1037.58
5	21	85.81	3.61	41.73	82.81	202.59	1055.7
6	21	73.51	3.67	41.67	83	199.2	1030.55
7	21	73.1	3.63	41.62	81.93	200.2	838.42
8	21	88.93	3.62	41.51	82.03	204.4	821.13
9	21	88.37	3.65	41.66	82.88	205.45	754.47
10	21	88.28	3.66	41.88	82.24	204.54	872.02
11	21	91.65	3.64	41.73	82.14	202.78	710.94
12	21	75.32	3.66	41.69	81.95	203.17	730.65
13	21	67.34	3.66	41.65	82.46	201.24	964.4
14	21	71.15	3.66	41.79	82.06	198.91	780.52
15	21	70.22	3.67	41.78	83.21	197.11	867.04
16	21	69.72	3.67	41.75	83.05	202.69	610.35
17	20.83	54.45	3.75	37.55	81.47	229.57	275.38
18	20.77	54.14	3.75	35.89	79.56	223.37	295.93
19	20.86	55.34	3.73	36.53	78.87	226.87	307.84
20	20.71	56.13	3.76	39.39	82.03	232.4	292.77
21	20.74	55.27	3.74	36.46	79.68	225.77	249.37
22	20.72	55.39	3.78	36.38	82.05	229.68	277.3
23	20.75	55.48	3.76	36.49	81.15	225.49	286.45
24	20.77	54.73	3.74	35.87	79.29	231	302.8
25	20.79	55.98	3.74	37.93	81.27	229.6	351.17

Figure 29 The Production Model

Prediction Based on Model

Depending on the production model the product variability is predicted based on weather conditions such as rainfall, temperature, sun hours, humidity, wind speed and normal cloud [13]. s python is very popular in data analysis with Pandas [14], a production prediction has been made using Python Pandas.

```python
import pandas as pd#Importiing pandas data package
import numpy as np#importing numpy package
from sklearn import linear_model#Importing leniar regression model
from sklearn.linear_model import LinearRegression
import math as mts#importing math module
import scipy
from scipy.stats.stats import pearsonr
import matplotlib.pyplot as plt#Importing the matplotlib package for data visualization
from sklearn.model_selection import train_test_split #To train and test the model
import seaborn as sns
from sklearn.metrics import r2_score
from sklearn import tree#Imorting decision tree algorithm
import pandas_datareader as pdr
from pandas_datareader import wb

#Multiple Linear Regression Model predict the production

DataSetProduction = pd.read_csv('E:/MSc IT/2Nd Part/FinalProject/NewDataSets/CSV Files/Updated Excel Data Sets/UpdatedCSVFiles/NewUpdatedDataCopy.csv')
#print(DataSetProduction)

#print(DataSetProduction.columns)
DataSetProduction.isnull().values.any()
# #
DataSetProduction = DataSetProduction.dropna()#Dropping null values of the dataset
# #
print(DataSetProduction.isnull().values.any())
```

Figure 31 Importing Pandas Library and other Packages for Data Mining

```python
DataSetProduction = pd.read_csv('E:/MSc IT/2Nd Part/FinalProject/NewDataSets/CSV Files/Updated Excel Data Sets/UpdatedCSVFiles/NewUpdatedDataC
#print(DataSetProduction)

#print(DataSetProduction.columns)
DataSetProduction.isnull().values.any()
# #
DataSetProduction = DataSetProduction.dropna()#Dropping null values of the dataset
# #
print(DataSetProduction.isnull().values.any())
print(DataSetProduction.columns)

inputProductionData = DataSetProduction.drop(['Average Production (MT)'],axis = 'columns')
targetProductiontData = DataSetProduction['Average Production (MT)']

regLinearProduction = linear_model.LinearRegression()#Object of linear regression model
input_weather_train_data,input_test_weather_data,target_train_production_data,target_test_production_data = train_test_split(inputProductionData,targetProd
# # # #training the data set with the linear model by fitting
#regLinearProduction.fit(input_test,target_test)

regLinearProduction.fit(input_weather_train_data,target_train_production_data)

output = regLinearProduction.predict([[20.9,55.0,3.5,44.2,83.9,205.0]])
#print(output)
print(regLinearProduction.score(input_weather_train_data,target_train_production_data))#Predicting the accuracy of model
#print(model.score(input_test,target_test))#Predicting the accuracy of model
```

Figure 30 Training, Predicting and evaluating the model

Production Model Prediction Results

```
C:\Users\Thilly\anaconda3\python.exe C:/Users/Thilly/PycharmProjects/UniFinalProject/FarmerDataModfel.py
     Average Temperature (Celcius)  ...  Average Production (MT)
0                             21.0  ...                   680.31
1                             21.0  ...                   772.10
2                             21.0  ...                  1037.58
3                             21.0  ...                  1055.70
4                             21.0  ...                  1030.55
..                             ...  ...                      ...
679                           21.5  ...                     0.00
680                           21.5  ...                     0.00
681                           21.5  ...                     0.00
682                           21.5  ...                     0.00
683                           21.5  ...                     0.00
```

Figure 33 Importing Production Model Dataset to Python

```
Index(['Average Temperature (Celcius)', 'Average Rainfalls (mm)',
       'Average Wind Speed(kmph)', 'Average Could (%)', 'Average Humidity(%)',
       'Average Sun Hours (hr)', 'Average Production (MT)'],
      dtype='object')
     Average Temperature (Celcius)  ...  Average Sun Hours (hr)
0                             21.0  ...                  202.55
1                             21.0  ...                  203.18
2                             21.0  ...                  200.05
3                             21.0  ...                  202.59
4                             21.0  ...                  199.20
..                             ...  ...                     ...
679                           21.5  ...                  232.10
680                           21.5  ...                  230.79
681                           21.5  ...                  233.65
682                           21.5  ...                  234.82
683                           21.5  ...                  235.68

[684 rows x 6 columns]
0       680.31
1       772.10
2      1037.58
3      1055.70
4      1030.55
        ...
679       0.00
680       0.00
```

Figure 32 Splitting the Model into Independent and Dependent variables

```
Name: Average Production (MT), Length: 684, dtype: float64
6378.8288013062065
[160717.0941425]
167095.9229438067
0.6508591659852526
```

Figure 34 Evaluation results of the production model

Data supplied to the model to predict the value and accuracy of the model has been tested. The accuracy of the model has been approximately 63% reasonable as the result of the model is almost 63% accurate.

Evaluation of Simple and Multiple Linear Regression Models to do the predictions

Regression analysis is a mathematical method that allows us to determine the strength and relationship of two dynamic objects. The reversal is not limited to two variables; we can have 2 or more variants showing relationships. Results from regression assist in predicting unknown value depending on the relationship and predictive variability [17]. As an example the height and weight of person have a higher relationship. Taller person is heavier. This relationship helps to predict the unknown weight or height based on the height or weight given as independent variables. Regression can be formed in two ways.

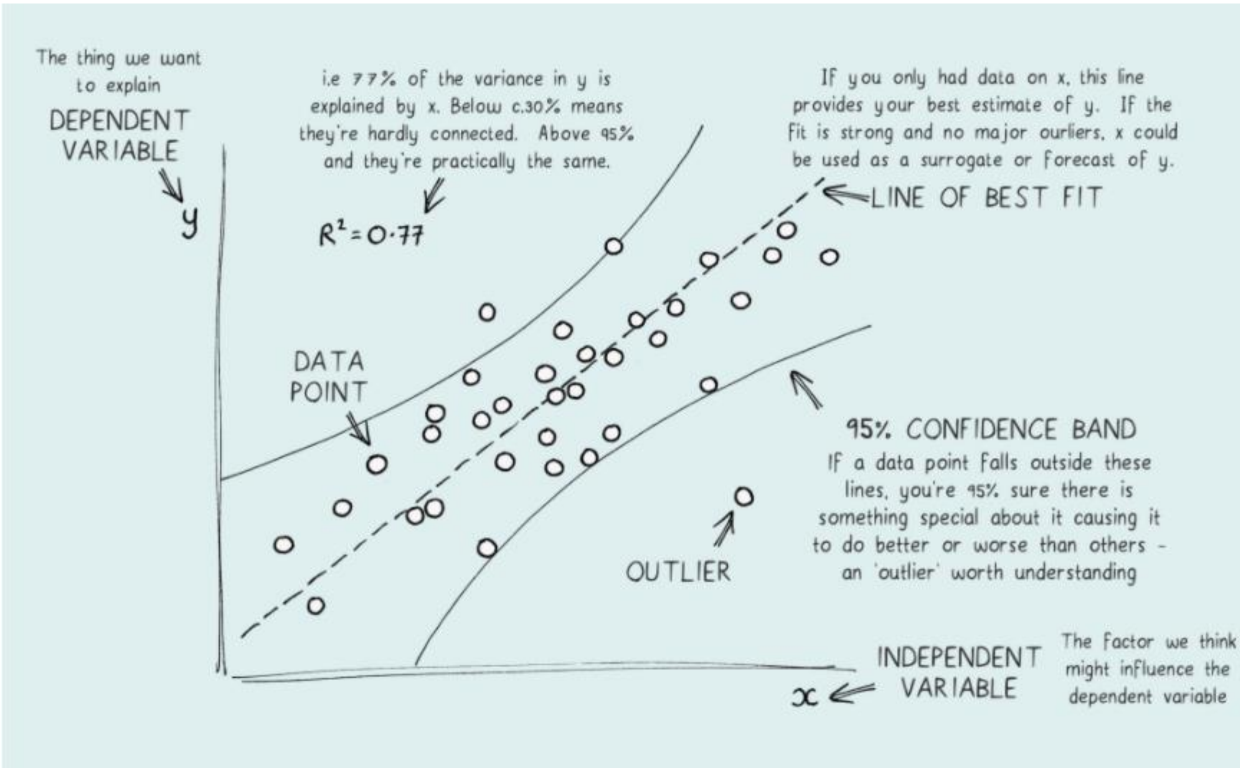


Figure 35 Sample Image of a Linear Regression Model [17]

- Linear Regression Model-Determining the relationship between single independent variations and single variables
- Multiple Linear Regression Model.-Finding relationships between multiple independent variables and single variables.

Outcomes have been predicted by incorporating these variables to describe relationship-dependent variability and independent variability.

More on Simple Linear Regression Model

Simple linear regression is used to measure the relationship between two variables [18].Simple linear regression model is useful when s particular researcher wants to;

- Predict the strength of the relationship between dependent variables and independent variables.
  - Eg: -The relationship between rain and humidity
- The amount of variation depends on the specific value of the individual variant (eg the amount of moisture at a certain rainfall level).

Simple regression model can be performed on qualitative variables by assigning on the qualitative measures as follows.

Imagine that one researcher wants to measure the happiness in terms of income. Here the income level can range between two acceptable prices such as $ 15k to $ 75k. Happiness is a quality variation but can be measured by providing ratings ranging from 1-10.

- Below assumptions can be made on simple linear regression [18].
- The size of the error in our forecast does not change significantly in the independent variance values.
- Database observations were collected using statistical measurement methods, and there is no hidden relationship between the observations.
- The data follows the general distribution.
- The relationship between independent and dependent variables is straightforward: a line of direct equation with data points is a straight line (rather than a curve or some kind of collecting element).

Above all assumptions can be applied in parametric linear regression model and if these assumptions are not met with the dataset, non-parametric linear regression model can be applied.

Performing the Linear Regression

$$y = \beta_0 + \beta_1 X + \varepsilon$$

Figure 36 Simple Linear Equation

Y = Predicted variable

β_0=Intercept, the y value when x is 0

X=Independent variable based on which the y value is changed

Ɛ=Estimated error or the estimated variation of the regression coefficient

Line reversal finds the correct line of data in your data by searching the return coefficient (B1) which reduces the total error (e) of the model [13].

More on Multiple Linear Regression Model

Multiple line regression is used to measure the relationship between two or more independent variables and one dependent variable. The Multiple Linear Regression Model can be used there;

- How strong is the relationship between two or more independent varieties and one variety (eg how rainfall, temperature, and additional amount of fertilizer affect plant growth).
- The amount of variation depends on a certain number of independent variants (eg expected crop yields at specific rainfall, temperature, and fertilizer addition) [13].

My main research project is exactly suited to be implemented using multiple regression model since it has been developed based on two models (Price and Production) with multiple dependent variables.

Below assumptions can be made on simple linear regression [13].

- The size of the error in our forecast does not change significantly in the independent variance values.
- Database observations were collected using valid mathematical methods, and there are no hidden relationships between variables.
- The data follows the general distribution.
- A line of direct equation with data points is a straight line, rather than a curve or some kind of collecting element.

In multiple linear regression model, it is possible that some of the variables that are represented are actually related to each other, so it is important to evaluate them before developing a regression model. If two independent variables are closely related (r2> ~ 0.6), only one of the regression models will be used [13].

Performing Multiple Linear Regression

$$y = \beta_0 + \beta_1 X_1 + ... + \beta_n X_n + \varepsilon$$

Figure 37 Multiple Linear Equation

Y = Predicted variable

β_0=Intercept, the y value when x is 0

$\beta_1 X_1$= The regression coefficient (B_1) of the first independent variable (**X_1**) (a.k.a. the effect that increasing the value of the independent variable has on the predicted **y** value)

$\beta_n X_n$=The regression coefficient (B_n) of the nth independent variable (**X_n**)

Ɛ=Estimated error or the estimated variation of the regression coefficient

To find the most accurate line for each individual variation, several duplicate lines include three items:

- Regression coefficients lead to very small model errors.
- The total t number of the model.
- Corresponding p-value (how is it possible that the t-number occurred by accident if the hypothesis was not at all of the non-relationship between independent and reliable variables was true).

By researching the features of the simplest and most flexible retreat models, the Multiple Regression Model has been used to predict the price and potential price of a given crop.

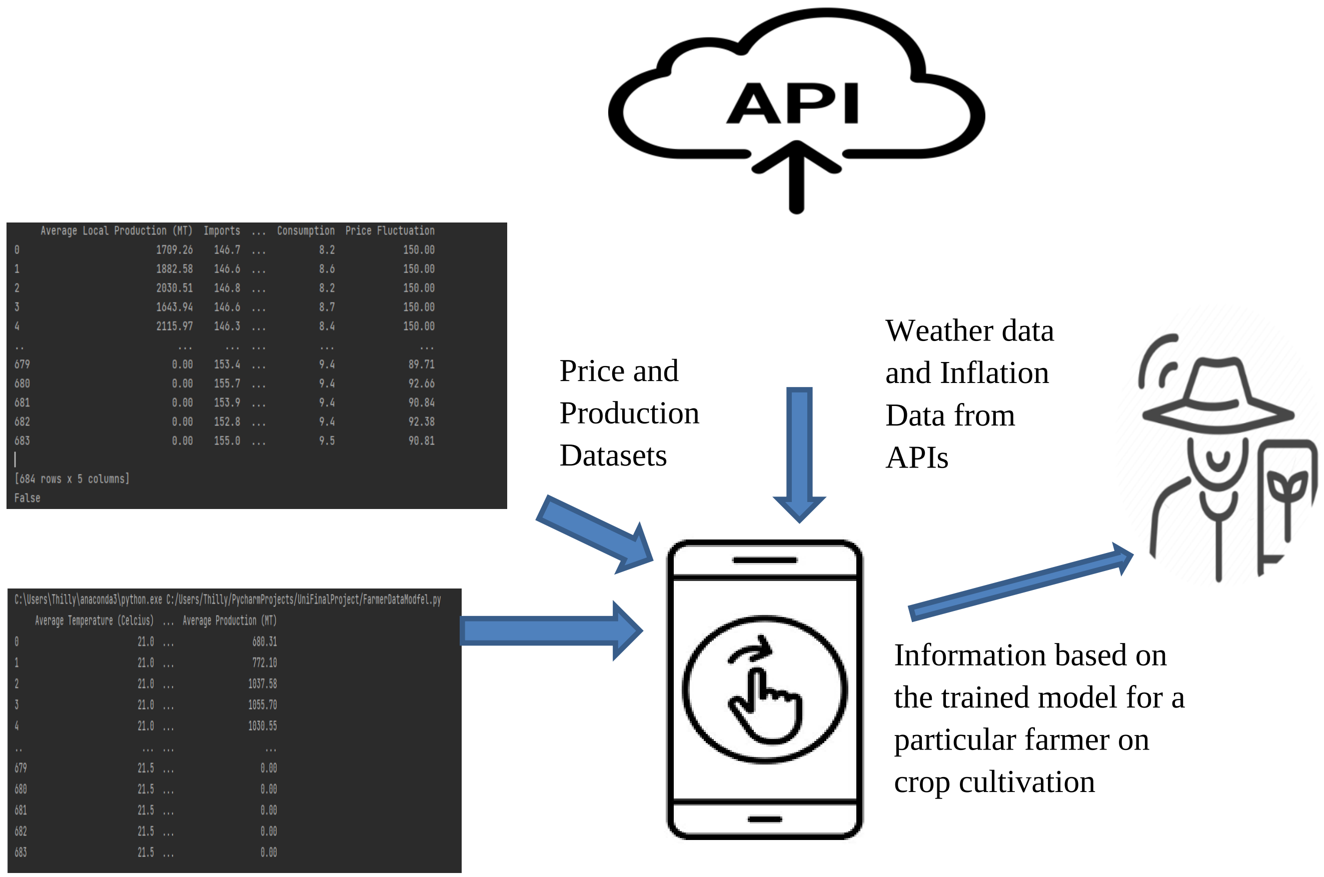
Chapter 5

Technical Implementation

This chapter discusses about the representation of the results of Price Prediction model and Production Prediction via a user interface. As discussed in earlier chapter, the data analysis has been conducted using Pandas data analyzing flatform in Python. The ultimate result is a mobile application which presents the results in user friendly manner for farmers to decide a crop. Below technologies have been used to implement the mobile application

- Kivy-
    - Python open source library for rapid application development that use the new user interface, like multi-touch apps [19].
- Pandas
    - A fast, powerful, flexible and easy-to-use tool for open source data analysis and deceptive tool, built on top of Python [20]
- Pandas Datareader
    - Pandas datareader is a sub-package that allows one to create dataframe from various online data sources, currently including: Yahoo! Finance.
    - In this project, Pandas Datareader has been used to fetch the real-time inflation of Sri Lanka from Worldbank Data official site.
- XAMPP local server with mysql database
    - XAMPP is a software distribution that offers Apache web server, MySQL database (actually MariaDB), Php and Perl (as a result of command line and Apache modules) all in one package. Available for Windows, MAC and Linux programs. No configuration required to integrate Php and MySQL [21].

With the above mentioned technologies and flat forms, a user friendly presentation of the data models have been presented for the farmers for decision making. Below figure suggests the structure of the mobile application


API
Average Local Production (MT) Imports ... Consumption Price Fluctuation
0 1709.26 146.7 ... 8.2 150.00
1 1882.58 146.6 ... 8.6 150.00
2 2030.51 146.8 ... 8.2 150.00
3 1643.94 146.6 ... 8.7 150.00
4 2115.97 146.3 ... 8.4 150.00
.. ..
679 0.00 153.4 ... 9.4 89.71
680 0.00 155.7 ... 9.4 92.66
681 0.00 153.9 ... 9.4 90.84
682 0.00 152.8 ... 9.4 92.38
683 0.00 155.0 ... 9.5 90.81
[684 rows x 5 columns]
False
C:\Users\Thilly\anaconda3\python.exe C:/Users/Thilly/PycharmProjects/UniFinalProject/FarmerDataModfel.py
Average Temperature (Celcius) ... Average Production (MT)
0 21.0 ... 680.31
1 21.0 ... 772.10
2 21.0 ... 1037.58
3 21.0 ... 1055.70
4 21.0 ... 1030.55
.. ..
679 21.5 ... 0.00
680 21.5 ... 0.00
681 21.5 ... 0.00
682 21.5 ... 0.00
683 21.5 ... 0.00
Price and Production Datasets
Weather data and Inflation Data from APIs
Information based on the trained model for a particular farmer on crop cultivation

More on Technical Implications

Kivy

Figure 38 Kivy Logo

Kivy is an open python library for fast-paced development apps for developing new specialized user connectors for mobile devices. Kivy Python Library has the features below

- Cross Flatform
  Kivy works on all flat formats like Windows, Linux, iOS, Rasberry Pi and even Android. Kivy codes can be generated in all of the aforementioned platforms. Kivy also receives input from WM_Pen, Mac OS X Trackpad and Magic Mouse, Mtdev, Linux Kernel HID, TUIO. Multi-touch mouse simulator etc.

- Business Friendly
  Kivy is a 100% free open source base. The tool kit is technically advanced, supported and operated. You can use it in a commercial product. The framework is stable and has a well-written API, as well as a planning guide to help you get started [19].

- GPU Accelerated
  In addition to OpenGL ES 2, the graphics engine is built using a modern and fast pipeline. The tool kit contains more than 20 highly free widgets. Most parts are written in C using python, and were tested with a retrospective test.

More on Pandas

Figure 39 Pandas Logo

Pandas is a software library used for python for data fraud purposes. Provides a variety of data structure functions and management tools, data representation using mathematical and graphical methods. It is a free package with software released under the BSD three-clause license. The term ‘Pandas’ is derived from the term “panel data”, which is the economic name of data sets that include the occasional recognition of the same people.

It is also called "Python data analysis" as it works on the hidden python form of data mining.

Pandas Library Features

- Dataframe, a device used to manipulate data with integrated identification.
- Tools are available to read and record data between memory data structures and different file types.
- Ability to rotate and rebuild data sets.
- A hierarchical axis index to work with high-resolution data in a low-level data structure.
- Syncing data and handling missing data
- Integrating and joining data sets.
- Label-based cutting, advanced reference and reset of big data sets.
- Management and formatting of data structures such as column insertion and deletion.

More on Panadas Datareader

Figure 40 Pandas Datareader Logo

Pandas datareader is a sub-package that makes it easy to create a data name from various online sources. Such online sources include:

- Yahoo! Finance
- Google Finance
- St. Louis FED (FRED)
- Kenneth French Library
- World Bank
- Google Analytics
- Tiingo
- IEX
- Alpha Vantage
- Econdb

Pandas Datareader provides remote data access to date pandas and works with many types of panda releasess for the purpose of data fraud.

```
pip install pandas-datareader
```

Figure 41 Installing Panadas Datareader in system environment

```
import pandas_datareader as pdr
pdr.get_data_fred('GS10')
```

Figure 42 Importing Pandas Datareader to the python environment

In this particular research project, the pandas datareader has been used to fetch the inflation information from the worldbank data sources.

```
#Fetching data from worldbank data using pandas datareader package
data = wb.download(indicator='FP.CPI.TOTL.ZG',country='LK', start=2020, end=2020)
data = data.reset_index(1)
data.columns = ['Year', 'Inflation']
# # #print(type(data))
inflation = data['Inflation']
DataSetPrice.isnull().values.any()
# # # # #
```

Figure 43 Fetching inflation rate from worldbank datasource

Indicators are available to access to explore data from different internet resources. Indicator is mostly available or embedded in the url of a particular data resources. Below code segment of indicates how the inflation details of Sri Lanka for a particular fiscal year has been fetched using pandas datareader.

More on XAMPP Local Server

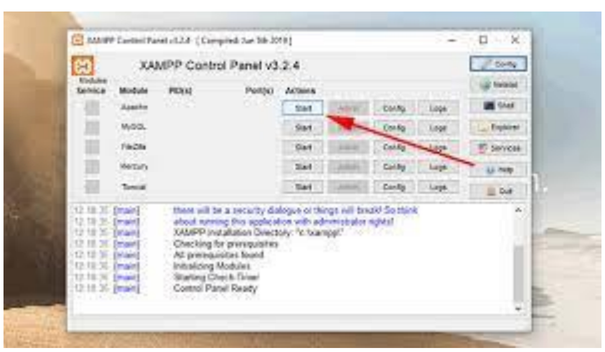


Figure 44 Xampp Environment

XAMPP stands for cross-platform, Apache, MySQL, PHP and Perl, and allows a specific developer to build and test a WordPress site or other php or mysql applications that are not connected to online, on a web server on your computer. This simple and easy solution works on Windows, Linux, and Mac - which is why this has a feature called "cross-platform". For this specific research project xampp helps to keep the details of farmers in their area for testing purposes. But the database will be downloaded to the remote server as the system is installed in the real world.

←T→	f_name	l_name	nic	district	password	contact_no
Edit Copy Delete	Bimal	Chandana	661171030v	Nuwara Eliya	Bimalc123	0717-95123
Edit Copy Delete	Bimal	Chandana	721171030v	Nuwara Eliya	cxbsfndfg	0717951239
Edit Copy Delete	Amal	Kumara	771171030v	Badulla	Amalkumara	0711-25263
Edit Copy Delete	dsgsdg	sdfsdf	881171030v	Badulla	qgq	0711-25263
Edit Copy Delete	Amal	Bimsara	891171030v	Badulla	amal@123	0711252632

Figure 45 Information of the registered farmers have been stored in local server

Weather API

Weather conditions are important to predict the product. The climate varies from region to region. As a result, the need for reliability of the Program Installation Program was always needed. OpenWeatherMap is a type of online service, run by OpenWeather Ltd. This Weather API provides global weather data, including current weather data, forecasts, nowcast and weather data for any location. The company provides timely updated forecasts for any location. The convolutional learning model is a model used to use weather broadcasting services and data from aircraft weather stations, ground radar stations, weather satellites, remote satellites, METAR and automated weather stations. Below statistics show some of the products provided by the API service

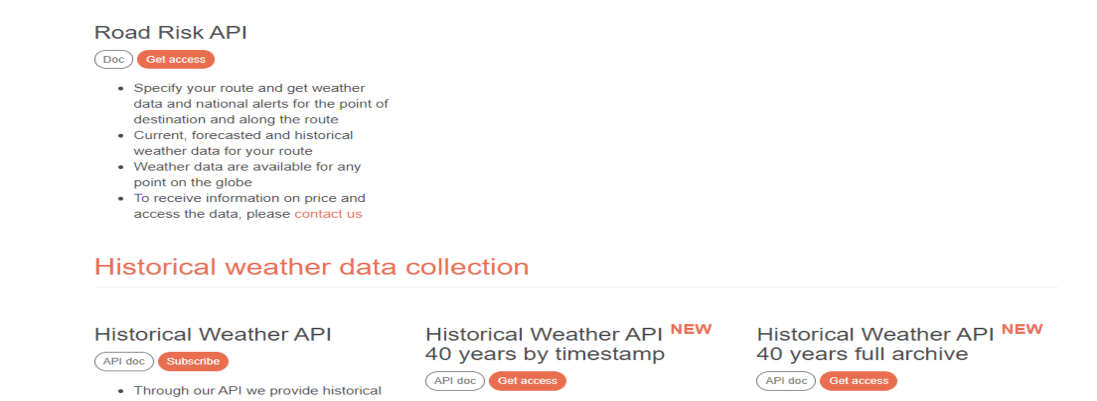
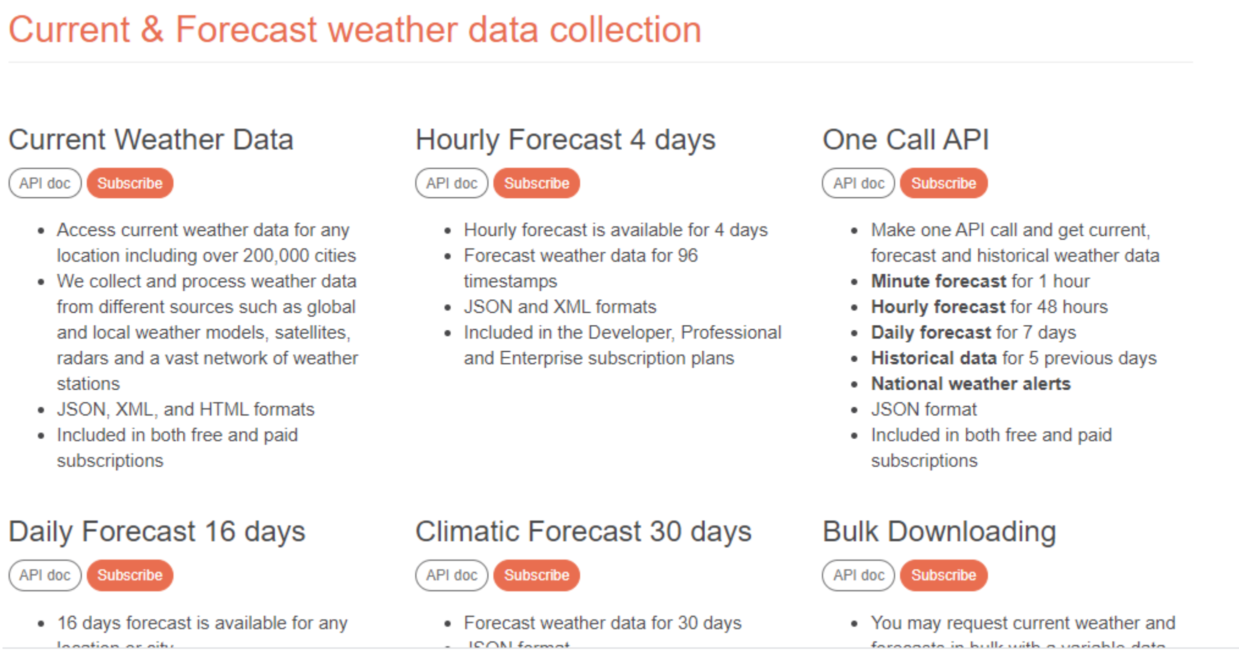


Figure 46 Services of openweathermap API



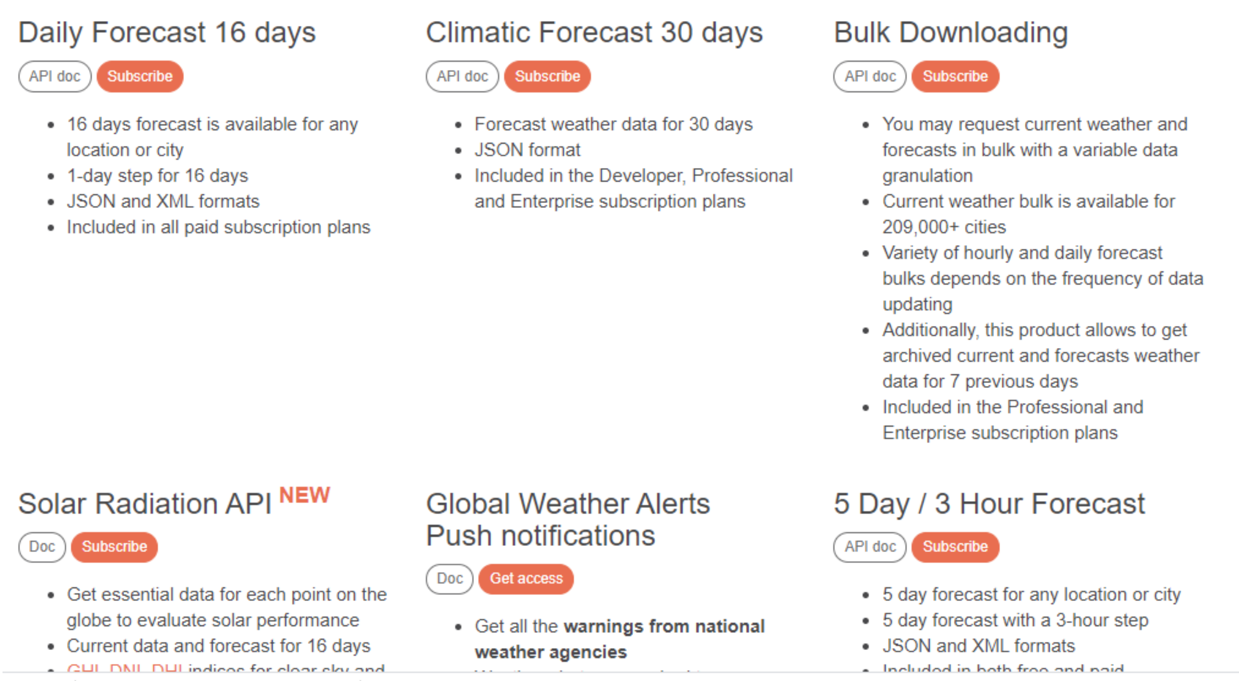


Figure 47 Services of openweathermap API

Source Codes to fetch weather data from OPenweathermap API

```python
if self.crop == "Potato":
    if self.Dstrict == "Badulla":
        api_link = "http://api.openweathermap.org/data/2.5/forecast?q=" + self.Dstrict + "&appid=" + self.APIBD
        api_link = requests.get(api_link)
        api_data = api_link.json()
```

Figure 48 Accessing Weather Information Based on the API key Passed for the url

```python
listTemp = []
listRainfalls = []
listWindSpeed = []
listCloud = []
listHumidity = []
#listPriseData = []

# Variables to handle weather forcasr
totTemp = 0
totHumidity = 0
totCld = 0
totWind = 0
sunRise = api_data['city']['sunrise']
sunSet = api_data['city']['sunset']
totRainFalls = 0
#totPrice = 0.0
```

Figure 50 Source Code to Store Weather data fetched from API

```python
# For loop to load temperature from api
for temp in range(30):
    listTemp.append((api_data['list'][temp]['main']['temp']))

# For loop to load humidity from api
for hum in range(30):
    listHumidity.append(api_data['list'][hum]['main']['humidity'])

# For loop to load cloud from api
for cld in range(30):
    listCloud.append(api_data['list'][cld]['clouds']['all'])
    # print(listCloud[0])

# For loop to load WindSpeed from api
for wnd in range(30):
    listWindSpeed.append(api_data['list'][wnd]['wind']['speed'])

# For loop to load Rainfalls from api
for x in range(0, 30):
    if "rain" in api_data['list'][x]:  # Checking whether the key is available in the list
        listRainfalls.append(api_data['list'][x]['rain']['3h'])
    else:
        totRainFalls = 0  # If that particular item is not available the default value is assigned as 0
    # print(totRainFalls)
```

Figure 49 Source Code to Manipulated Weather data fetched from the API

```python
# Formatting weather data with 2 decimal points
print("Average Temperature is ", "{:.2f}".format(self.avgTemp))
print("Average Humidity is ", "{:.2f}".format(self.avgHumidity))
print("Average cloud as % is ", "{:.2f}".format(self.avgCld))
print("Average Windspeed is ", "{:.2f}".format(self.avgWindSpeed))
print("Average sun hours ", "{:.2f}".format(self.avgSunHrs))
print("Average Rainfalls is in mm", "{:.2f}".format(self.avgRf))
#print("{:.2f}".format(self.avgPrice) , "/=")
self.ultimateResult = FarmerDataModfel.regLinearProduction.predict([[self.avgTemp , self.avgRf ,
                                                    self.avgWindSpeed , self.avgCld ,
                                                    self.avgHumidity , self.avgSunHrs]])

if self.ultimateResult[0] < 0:
    self.ultimateResult[0] = self.ultimateResult[0]*-1
else:
    self.ultimateResult[0] = self.ultimateResult[0]
    # print(type(APIBD))
    # print(Dstrict)
```

Figure 51 Predicting the Production based on Weather data fetched from the API

Based on above approaches, the information is presented via a mobile application interface for farmers to make decision on cultivating a selected crop based on the district where he/she lives

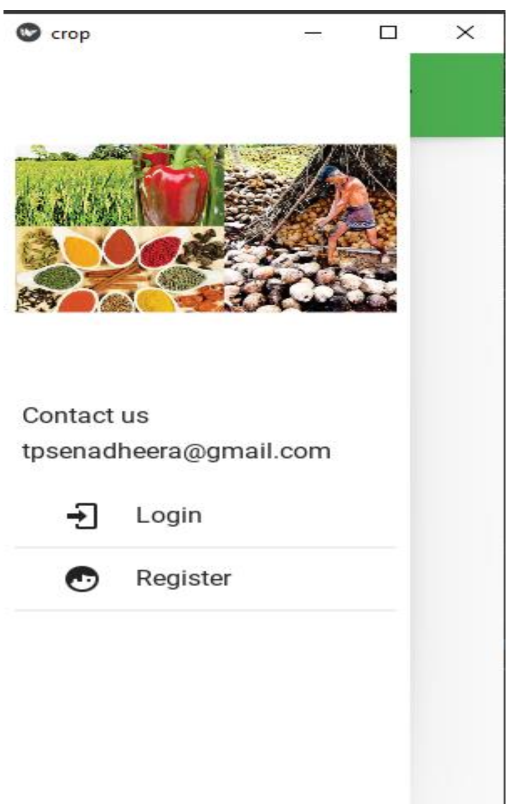


Figure 53 Home Interface

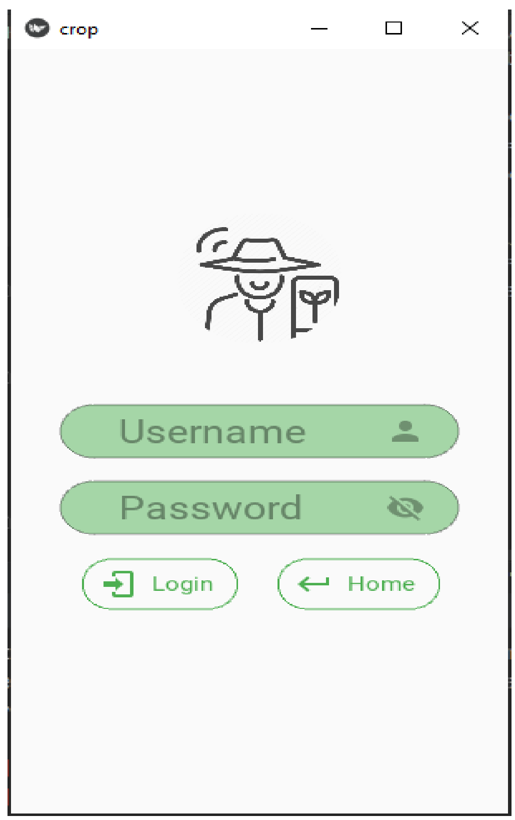


Figure 52 Login Window

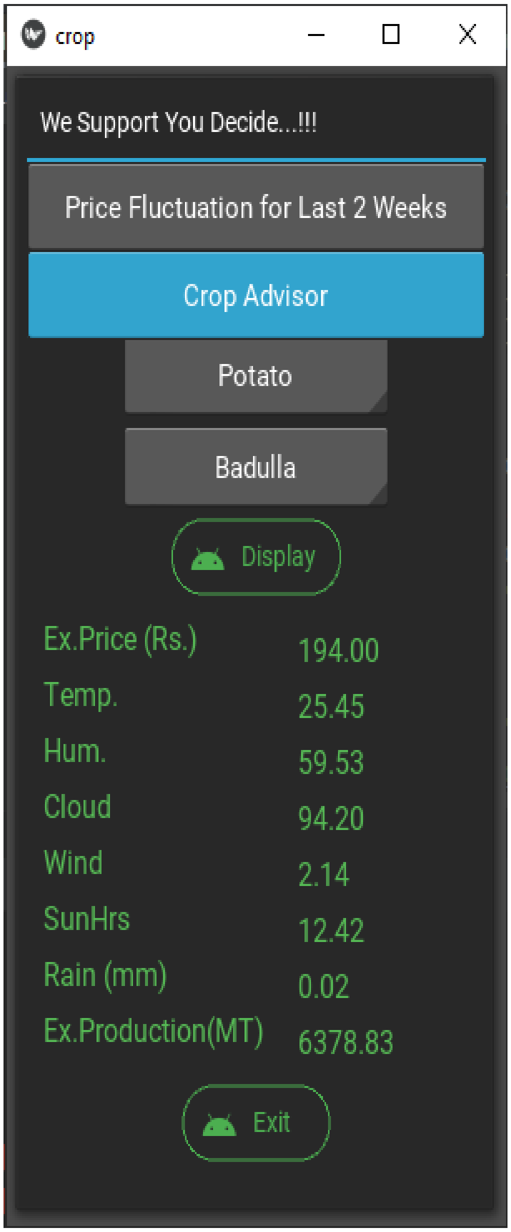


Figure 54 Window for Price and Production Prediction

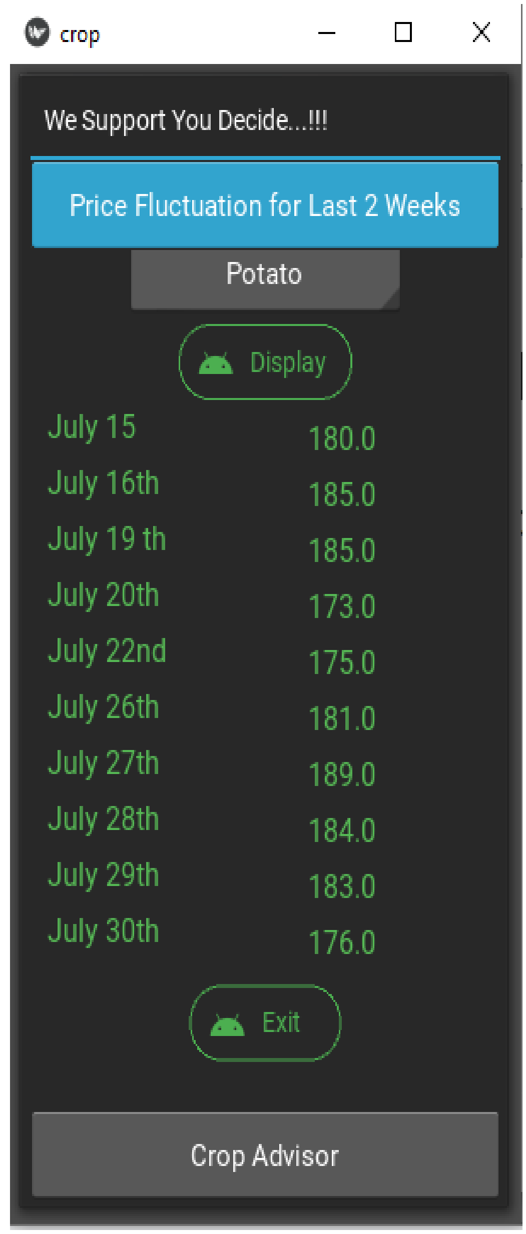


Figure 55 Window to Show Price fluctuation for last 2 weeks

Chapter 6

Future Developments and Benefits of the Application

Introduction

Successful implementation of this project will be an effective communication medium between for farmers on crop prediction and planning since the application provides decision making support facility based on the upcoming weather conditions and price prediction. Agricultural sector in Sri Lanka desperately needs a proper decision support system as the factors which affect the agriculture changes dynamically.
As the name suggests, this is a decision-making system (DSS). Such systems support the managers/ users of the system to make decisions by assessing large amount of information from various sources. As the name suggests, this crop prediction decision support system provides information on upcoming climatic condition, expected production by district, and price for a particular crop. Based on that information a farmer can decide which crop is ideal to cultivate in future.

Characteristics of a Crop Prediction Decision Support System

- Supporting for farmers by providing price, production and climatic conditions structured way.
- Intelligent and accurate prediction on price, production for selected crop, district wise climatic conditions.
- Integrating different levels and types of critical information from various sources to a common flat form such as weather, price and production..
- Support for interdependent or sequential decisions on crop prediction based on the available information.
- Support for intelligence, design, choice, and implementation.
- Support for variety of crop prediction approaches with the availability of information.
- Decision Support System is adaptive over time as it fetches data from reliable APIs around the world to do the predictions.

Benefits of DSS in Crop Prediction

On the whole, the ultimate aim is to build useful models for crop cultivation. Especially the upcountry farmers suffer a lot because of lack of return on investment, crop wastage since lack of information and some of data which cause production change dynamically. So this decision support mobile application presents enough data for decision making using two specific models (Price and Production)

- Improves efficiency and speed of decision-making activities to select a crop based on accurate models for price and production.
- Increases the control, capability of futuristic decision-making of the crop cultivation to gain optimum return on investment.
- Facilitates communication between farmer and major data sources like weather and price for smooth decision making.
- Encourages learning or training on how does the price behaves and how do climatic changes make an effect on crop cultivation.
- It reveals new cultivation approaches which can be unusual but can earn more profits after making decision.
- Helps automate decision making processes system provides enough information district wise on crops based on price climate and production.

Chapter 7

Results and Discussion

At the moment, the application provides information of Expected Price, Expected Production (MT) and the Price fluctuation of a particular selected crop for last two weeks. The accuracies of the models of Price and Production are recorded as **77.93% and 65.08%** respectively. Below chart indicates the behavior of the price model with randomly selected dates and a clear comparison can be done between the predicted price and the actual price.

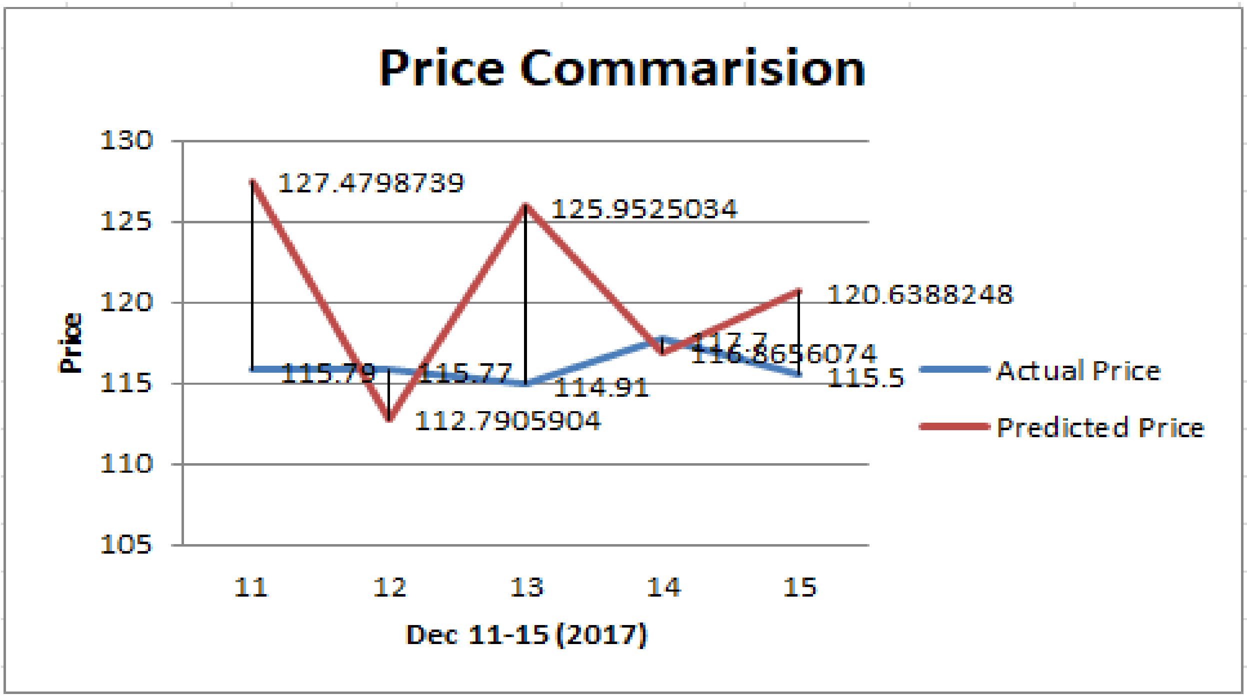


Figure 56 Actual Price Vs the Predicted Price

The accuracies of the Price and the Production models are in the modest levels as they have to be improved furthermore to provide precise information for farmers to do decision making in crop cultivation. Below mentioned additional improvements are to be suggested to improve the overall performance of this decision support mobile application

- Improving the accuracy and Performance of the Production model by adding soil moisture feature which provides an additional dimension for farmers in deciding a crop to cultivate

- Adding and Additional Model to predict amount of fertilizer to be fed for a particular crop based on the yield so farmers are given additional information on amount of fertilizer required for a particular crop.
- Improving the accuracy of the Price Model can be further done by adding the feature of the prices of imported crops which will be added in future versions of the mobile application
- Facilitating farmers by suggesting potential crops based on the climatic condition of a particular district so the farmers can choose a particular crop with the available resources, potential price and by considering the expected climatic condition in future.

Chapter 8

Future Development

This Mobile Support Decision application is developed under the Phased Software Development Approach section because each category ends with a defined release. Categories are executed in accordance with the order specified by the software development process model. The main reason for using a phased approach to software development process breaks down the problem of building software into successful phases, each with different concerns about software development. [22]. Earlier stage of this mobile application facilitates farmers to make decisions based on the information given under selected crop for a selected district based on the price and expected production. Below list indicates the expected future developments for this application.

- Increasing more components to the production model by including more independent variables making this production model more accurate by predicting the potential amount of potato can be grown by each farmer. In order to improve the price prediction model in such a way below components are expected to be added
    - Soil Moisture

Soil Moisture is very important to predict the conditions for crop cultivation. All plants vary in their water requirements depending on their size and growth and the length of their ripeness and the season of high growth. Perhaps no other major crop differs in its sensitivity to water pressure depending on the stage of growth than potatoes. Potato requires cool temperature and high moisture. Therefore it needs a proper irrigation and proper way to track soil moisture. According to the Food Research project under the leadership of Dr. Steve Thomas, of Lincoln, that water is a factor in closing the potato harvesting gap, that management practices change the capture and retention of groundwater, and that soil parameters affect water use over time [23]. His research targets on finding more ways on storing more in soil. The research team found that the potato is high water stressed crop as its roots are relatively small. As a result, potatoes need a high water content in the soil because potato strength is limited to extracting water from deep layers of soil. The research has been conducted to find the optimum way to extract and maintain the soil moisture for potato cultivation. Research looked at flatbed or the traditional ridge and furrow to capture which method reduces water losses through evaporation. The experiment was performed on a PFR test site using two types of beds (ridge / furrow and flatbed) and irrigation regimens such as full irrigation, double irrigation weekly to replace ground water shortages and low irrigation in the top 400mm. wilting. This test was performed at the PFR test site using a variety of bed methods such as ridge / furrow and flatbed and two irrigation regimens such as full irrigation, double irrigation to replace ground water shortages and low irrigation irrigation when the ground water 400mm was near the area. wilting. This showed an increase in yield (seven tons / ha) on flatbed plots compared to reed / ditch sites under full irrigation. Likewise there are enough researches have been conducted to examine the importance of soil moisture to grow a particular crop but in this particular sector the points are presented on the importance of soil moisture to potato. So in future releases the production model will be more improved by predicting the soil moisture to grow a selected crop.

- Improving the System by Introducing a new Model to decide amount fertilizer to be added for a particular product. This will help farmers to decide the amount of fertilizer required for a particular crop. This model will be built with the factors such as yield, amount of materials in the soil and the amount of fertilizer fed in to a particular crop. The dependent variable will be the amount of fertilizer to be fen and the independent variables are yield, organic materials of the soils. The model will suggest that which amount of fertilizer will be required for a particular crop for an optimum yield.
- Suggesting more optional crops based on the environmental and other conditions to grow.
    - At the moment farmer has to choose the crop and based on the upcoming climatic conditions and other factors, expected production and price will be given so the farmers can decide whether he / she can go for that particular crop. But the upcoming release will suggest the crop, amount of fertilizer required based on the region/ district in Sri Lanka.

Chapter 9

References

Bibliography

[1] Central Bank Of Sri Lanka, "Economics and Social Statistics of SriLanka," Statistics Department,Central Bank of Sri Lanka, Colombo 01, 2019.

[2] Central Bank of Sri Lanka, "Daily Price Report," [Online]. Available: https://www.cbsl.gov.lk/en/statistics/economic-indicators/price-report.

[3] Ministry of Agriculture, Rural Economic Affairs,Livestock, Irrigation, Fisheries and Aquatic,Resources Development, Performance Report 2018 - Agriculture Sector, 2018.

[4] Subday Observer, "Agriculture sector challenges and reforms," 16 December 2018. [Online]. Available: http://www.sundayobserver.lk/2018/12/16/thoughts/agriculture-sector-challenges-and-reforms.

[5] M. Henegedara, Price Instability and Change of Terms of Trade in Small Farming Sector in SriLanka(With Special Reference to Cultivation of Paddy and Vegetables), IJDR, 2016.

[6] Ministry of Agriculture, Rural Affairs, Irrigation, and Fisheries, and Aquatic Resources Development, SRILANKA Overchaining Agricultural Policy, 2018.

[7] "Early Warning Messages for Vegetable Farmers (Price Prediction and Crop Selection)," [Online]. Available: http://yala2021.croplook.net/tabs_country_original_4public/.

[8] WORLD BANK DATA GROUP, "Inflation, consumer prices (annual %) - Sri Lanka," [Online]. Available: https://data.worldbank.org/indicator/FP.CPI.TOTL.ZG?locations=LK.

[9] Helgi Analytics, "Potato Consumption Per Capita in Sri Lanka," [Online]. Available: https://www.helgilibrary.com/indicators/potato-consumption-per-capita/sri-

lanka/#:~:text=Potato%20consumption%20per%20capita%20reached,time%20low%20of%20kg%20in.

[10] Socio Economics and Planning Centre,Department of Agriculture,Peradeniya, AgStat, Peradeniya: Socio Economics and Planning Centre,Department of Agriculture,Peradeniya, 2019.

[11] Department of Census and Statistics,Ministry of Finance, NATIONAL CONSUMER PRICE INDEX (NCPI), Colombo, 2021.

[12] "generatedata.com," [Online]. Available: http://generatedata.com/.

[13] Scribber, "An introduction to multiple linear regression," 2021. [Online]. Available: https://www.scribbr.com/statistics/multiple-linear-regression/.

[14] Pandas Development Team, "What's new in 1.3.0 (July 2, 2021)," 2021. [Online]. Available: https://pandas.pydata.org/docs/whatsnew/v1.3.0.html.

[15] "https://www.worldweatheronline.com/," [Online]. Available: https://www.worldweatheronline.com/sri-lanka-weather.aspx.

[16] Agriculture and Environment Statistics Division,Department of Census and Statistics, Sri Lanka., "Estimated Extent and Production of Potato (solanum tuberosum)," [Online]. Available: http://www.statistics.gov.lk/Agriculture/StaticalInformation/rubb2.

[17] Towards Data Science Inc, "Linear Regression Explained," [Online]. Available: https://towardsdatascience.com/linear-regression-explained-1b36f97b7572.

[18] Scribber, "An introduction to simple linear regression," [Online]. Available: https://www.scribbr.com/statistics/simple-linear-regression/.

[19] Kivy, "Kivy:Cross-flatform PythonFramework for NUI," [Online]. Available: https://kivy.org/#home.

[20] Pandas Development Team, "pandas," [Online]. Available: https://pandas.pydata.org/.

[21 West Chester University, "XAMPP," [Online]. Available:

] https://www.cs.wcupa.edu/rkline/index/xampp.html.

[22] MBA Knowledge Base, "A Phased Approach to Software Development," [Online]. Available: https://www.mbaknol.com/information-systems-management/a-phased-approach-to-software-development/.

[23] University of Nebraska,Lincoln, "Potato Growth and Irrigation Scheduling," [Online]. Available: https://cropwatch.unl.edu/potato/plant_growth.

[24] Scribber, "An introduction to multiple linear regression," [Online]. Available: https://www.scribbr.com/statistics/multiple-linear-regression/.

[25] University Of Nebraska–LINCOLN, "Potato Growth and Irrigation Scheduling," [Online]. Available: https://cropwatch.unl.edu/potato/plant_growth.