Title

MSc Research Project
Programme Name

Forename Surname
Student ID: XXX

School of Computing
National College of Ireland

Supervisor: XXX

National College of Ireland

MSc Project Submission Sheet

School of Computing

Student Name:	…Tanay Kasar…. ………………………………………………………………………………………………………
Student ID:	…x17156700…………………………………………………………………………………………………… …….……
Programme:	Msc in Data Analytics……………………………………………………………… **Year:** …………………………..
Module:	……………………………………………………………………………………………………….………
Supervisor:	……………………………………………………………………………………………………….………
Submission Due Date:	……………………………………………………………………………………………………….………
Project Title:	……………………………………………………………………………………………………….………
Word Count:	……………………………………… **Page Count**…………………………………………….……..

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.
<u>ALL</u> internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature: ……

Date: ……

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST

Attach a completed copy of this sheet to each project (including multiple copies)	□
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	□
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	□

Assignments that are submitted to the Programme Coordinator Office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	

Date:	
Penalty Applied (if applicable):	

Time-Series Forecasting on Food and Agricultural Production of World

Tanay Kasar
x17156700

Abstract

Agriculture domain is one of the most important domains. Industries rely on agricultural production for raw materials. This research serves as cross-section between Agriculture domain with Data Analytics and Machine learning models. The data was obtained from 'Food and Agriculture Organisation (FAO)'. It contains the data on 200 countries from year 1961-2013. The food items mentioned in the data are classified as the food produced for human population and the food available for other living beings other than humans. The objective is to obtain Insights on Food and Agriculture production of world. Forecasting the production of coming 30 years. Time-Series forecasting is carried out by using ARIMA models, Holt-Winter's model, Exponential model. The results of these model were compared to select the most reliable model. The insights and forecasting obtained can be useful in decision making and policy implementation in the Agriculture domain. The forecasted values will be useful for understanding the potentials and how much a country must produce to satisfy the needs of their population.

1 Introduction

Agriculture has always been a primary occupation of the human civilization. When humans learned the process of agriculture, their lives became stable. All over the earth the tropical covering of the vegetation changes from place to place. This diversity in vegetation is due to geographical factors, region in which the land lies, the level of fertility. The amount of crops that a country produces depends upon the extent of aid provided by the government. There are some regions where people strive to get one meal for the day. Our aim in this research is to understand and analyze the trend followed by the agricultural data. This research serves as a cross-section between Agricultural domain and Data Analytics. The concepts and principles on Data Analytics can be applied on the historical agriculture data. It can be used to understand the nature of the data and most importantly to forecast the outcome with respect to time.

Agricultural domain is one of the most important domain, which is linked to a number of different domains. Thus the changes in agricultural productivity will also reflect on other productions. This co-relation exists because agricultural provides raw material to numerous industries. This makes agricultural domain very crucial. The data acquired can be used for making predictions and getting insights on the agriculture domain. Thus the results can be useful while implementing agricultural policies for the countries according to their production. Overall this research is useful for the purpose of understanding the scenario of agricultural economy and also to implement policies which can improve the productivity of a nation. In the later sections of this report the related works , implementation and evaluation have been discussed.

2 Related Work

In this section all the papers that were referred before implementation have been discussed. The referred research papers were selected with respect to agricultural domain. The research carried on forecasting models like ARIMA and other exponential models were studied. In this section all the advantages and disadvantages of the particular models are discussed. Each reference is critically reviewed to get a better understanding of the work done in the past.

2.1 Work done in Agricultural Domain

In this sub-section all the related work in agricultural domain has been discussed. In the following research paper the author has specified the implementation of machine learning models. The model was implemented and trained on the agriculture data. The model's accuracy was good, but the model faced problems in accepting he large data (McQueen et al., 1995). In this series the data will be pre-processed to eliminate all the unwanted data, such that the model accepts the complete data. For this purpose R programming was used for data processing.

In the next referred research the author has implemented time series to study the agriculture data. The data used in this data was of just 14 countries. An attempt was made to find cross-section between agricultural field with industry domain (Diop and Kengne, 2017). This concept of the cross-section can also be implemented on this concerned project, but the limitation of data hinders the execution. Our data contains the data of 200 countries, if the data of gdp of all countries is available then the cross-section between agriculture and gdp of the country can be known (Bernard and Jones, 1996).

In the following paper the use of satellite images to get intel on the land properties was done. The information coming from the images was collected and a dataset was formed. Clustering models were used to classify images depending on their properties. The images obtained from the space had very low resolution, therefore the information conveyed from each image was less.(Goncalves et al., 2014). There is a scope of applying clustering models on the data. But the clustering models failed to classify the agricultural products produced by the countries.

The following research paper focuses on the production of soya bean over past years. The author has used forecasting models to predict the future. The accuracy of these model was good. The author compared a number of model on their performance (Savla et al., 2015). In the proposed research the forecasting models like ARIMA and exponential were used to forecast the production of the countries . In the next paper the author has applied data mining techniques on soil data. The model was trained on the soil data. The data fed to the model was not sufficient, thus the model was not trained properly. This was the reason the accuracy of this model was low (Gandge and Sandhya, 2017). The data used in this research is vast, thus the models gave better results.

Land area is a very important factor when it comes to agriculture. The next paper focuses on the prediction of Sugarcane crop grown in India and also to find a co-relation between production and land area. The author had combined the data on production and land area. The concepts of regression analysis were used. In conclusion this paper states that the production depends on the area of cultivable land (Grover and Johari, 2016). The data on the land area on a global level isn't available, therefore the dependency of land area cannot be found.

2.2 Applications of ARIMA and other models.

For implementing time series, a number of machine learning models are used. The research papers which had the implementation of these models are discussed in this section. Time series analysis can be classified as time domain and frequency domain analysis. For time series analysis linear models are used. ARIMA model is preferred due to its capabilities to solve the problem of auto-correlation (Ferenti, 2017). ARIMA can also be combined with other models. In the following paper Arima model was combined with "Multilayer Perceptron". This was dome to boost the accuracy of the result (Yujun, Yimei, & Jianping, 2016). This combination proved to be successful, this proves that ARIMA model is flexible. This makes it easy to interlink it with other models (Hirata et al, 2015). Thus ARIMA model was selected, because the concerned research makes the use of multiple models.

The next paper compares ARIMA model with other auto-regressive models. A Solar-grid data was used to train the models. The task was to predict the damage caused to instruments with time. ARMA model was compared with ARIMA model, it was found that ARIMA model performed better than ARMA model. The life of the instruments was predicted using ARIMA model (Colak et al., 2015). As in the next paper the author has proposed the use of forecasting model to predict the time when the components in air-craft must be replaced. The functioning of instruments used in Aircrafts must be closely monitored. If these instruments fail, the loss of life will be unimaginable. Thus a model was to chosen which has good forecast with good accuracy. Therefore ARIMA model was used, and it gave good results with prediction (Yang et al., 2017). By the work done in the discussed research paper it can be concluded that ARIMA model is well suited to perform the task. This is because of its moving average and auto-regressive techniques. Holt-winter's model is used for forecasting tasks (Foutz and Lee, 2000). It is suggested that holt-winters model should be used with exponential smoothing. Holt-Winter's model uses the initial values to fit the next values. It plots the time-series of observed values with the fitted values (Kazempour, 2007).

3. Research Methodology

The selection of the methodology depends on the approach and domain you want to work with. Since the concerned research is based on the commercial impact of agriculture, CRISP-DM methodology was selected. CRISP-DM is the full-form of Cross Industry process for 'Data Mining'. An approach which is very structural, useful for implementing data mining models and projects. The process flow diagram of CRISP-DM is shown below. This methodology was adopted due to the nature of the data. The understanding of the objectives and then transform the knowledge obtained into a data mining solution which solves the concerned problem. After the implementation of the models, the models should be tested on the data. To evaluate whether the results solves the problem[1].

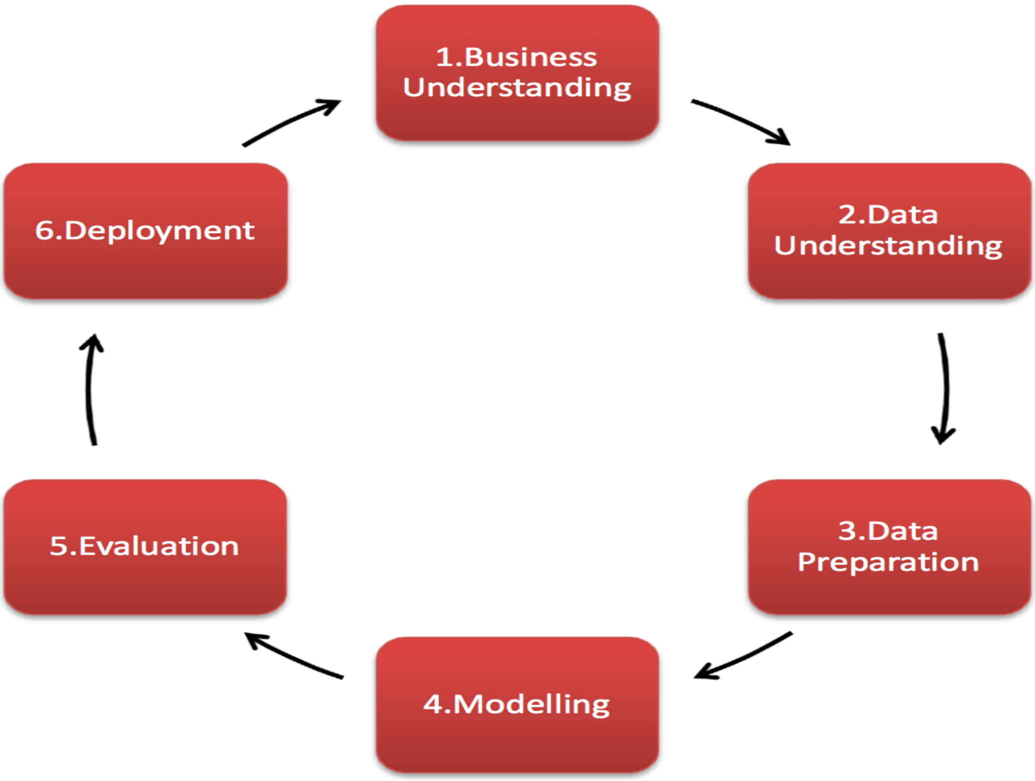

Figure 1: CRISP-DM methodology

3.1 Application of CRISP-DM methodology on the Research

From the process flow we acknowledge that there are these 6 events which follow each other. While implementing any data mining project, the data must undergo these processes. The order of these steps does not matter, but the execution of these steps accurately is vital. In this section each of the steps derived from CRISP-DM methodology is discussed.

1. <u>Business Understanding:</u> This first step is to understand and visualise what is to be achieved from a commercial perspective. The objectives must be set and the required components should be balanced accordingly. The objective is to target the factors which are of significance. In this proposed the aim is to give insights on the agricultural economy, and also to show its commercial value. From the agricultural data all of the important factors were identified.

[1] https://www.sv-europe.com/crisp-dm-methodology/

2. Data Understanding: This step concerns with exploration of data. Exploration of data means targeting the key attributes in the data and their significance with one another and the information it conveys. Our data consists the information of a country's production. The data of 200 countries is available from the year 1961-2011. It also shows what types of crops a country grows.

3. Data Preparation: In this the data needs to be transformed in such a way that it serves the purpose. It may include a number steps depending on the nature of the data and how the data needs to altered. For the agricultural data, a number of columns were renamed according to the convenience. There were a number of missing values in the data, the columns which had few missing values were interpolated. While the columns with many missing values were deleted. For time-series analysis the transpose of data was required therefore a transpose of the data was obtained.

4. Modelling: For fitting the model, the main data needs to be partitioned into test data and training data. After the model is fitted using training and test data, its accuracy should be deduced. If the accuracy is fine then the model can be used in implementation. If the model performance is not satisfactory, different alternatives should be used to improve the performance. In this research the use of ARIMA, exponential models are used.

5. Evaluation: This step deals with evaluation of the model. Evaluation is done to find out whether the model satisfies the conditions specified in the problem statement. The model should be implemented with the actual data. When all the evaluations are convincing, the model is finalised.

6. Deployment: This is the final step of the project. This step focuses on summarising all the results that were obtained after the execution of the above steps. A better organising of results is very vital, such that the interpretations make sense to the viewers and the organisation.

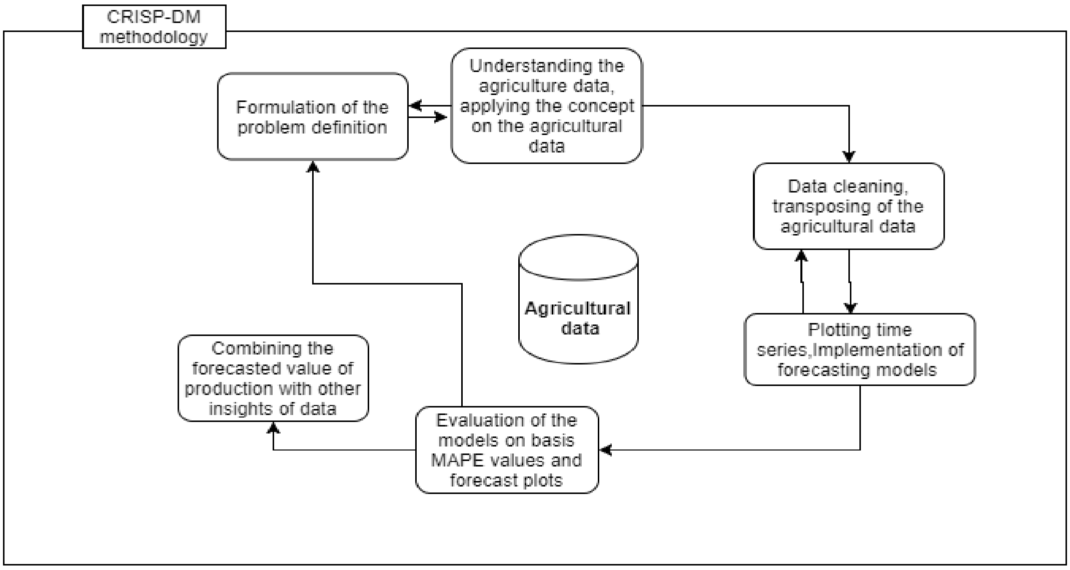

Figure 2: Application of CRISP-DM methodology

4.Design Specification

In this section all the techniques, frameworks used to implement the project are described.
All the processing and implementation of the models was executed using R-programming language. The execution using R was implemented in R-studio. R-studio is an 'Integrated Development Environment(IDE)' which supports programming and statistical computing. The version of the R-studio used is Version 1.1.419. R-studio supports installation all the packages, which carry the required functions. Library packages like ggplot2, Forecast, Graphics, cowplot and many more.

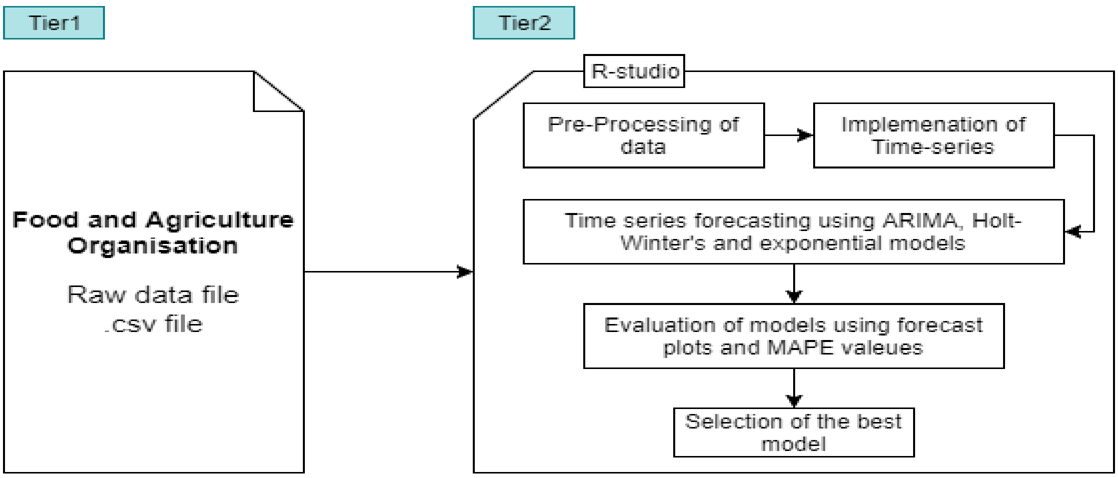

Figure 3: 2-Tier Design layout

The design is consists of 2 tiers. The first tier includes acquiring the data. In this case the data was obtained from the 'Food and Agricultural Organisation'. The Tier 1 also encompasses the time spent on business understanding. Tier 2 includes all the execution done on the data, from pre-processing to the implementation of data.

5. Implementation

In this section all the steps that were executed to achieve the goal have been explained. The process flow of the project has been displayed below. The process begins with getting the raw data files, followed by pre-processing steps. The process follows the steps mentioned in CRISP-DM methodology. In this section the nature of data is explained along with all the insights obtained on the agricultural economy. The process of implementing Time series forecasting is described in the later part of this section.

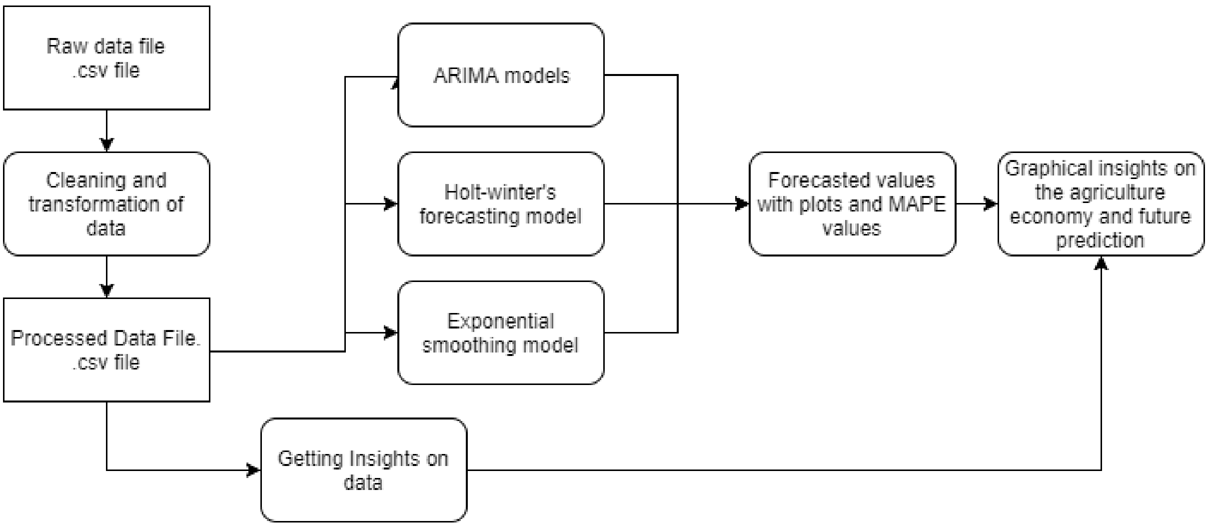

Figure 4: Process flow of Implementation

5.1 The Data

The process begins with acquiring the data. The data used for this research was obtained from 'Food and Agriculture Organization(FAO)'. The dataset contains the agricultural data from 245 countries, it contains the information of a country's production. The information from the year 1963-2014 is available, it contains the types of crops produced by each country along with is quantity. From the data it can be concluded that there are two major classifications, Food and Feed. Food represents all the stuff that is available for feeding the human population. While Feed represents the stuff available for the living stalk of the country, that is the food available as fodder for animals. The data acquired needs to cleaned and processed in such away that the data becomes fit for further deploying.

5.2 Data Cleaning and Processing

Before we use the data for our main purpose, pre-processing of it is necessary. This is done to get rid of missing values and the unwanted components of our data. Our data's timeline extends from the year 1963-2014.Between these years a number of new nations were formed especially in the year 1992 when U.S.S.R was dissolved into Russia and a number of small countries. Therefore the data of such countries is available from the year-1992. In the later sections we will inspect about the changes in production due to the dissolving of U.S.S.R. Data cleaning involves renaming some components and taking care of the missing values. This ensures smooth and efficient fitting of models on our data. The next important step is to rename some of the column names, the columns which are significant. This renaming of columns provides better readability and understanding. The columns of area code, item code, elements were renamed. The graph displayed below shows the percentage of missing values in each year. In the beginning the number of missing values were more, but after the emergence of newer countries the number of missing values decreased. In most of the execution steps missing values were replaced by 0.

5.3 Transforming the data for time series.

This step is necessary to transform our dataset, to make it suitable for plotting the time series. The `component of time should be as row heads but in our data set the years were represented as columns. Thus the transpose of the dataset was obtained. Along with the transpose all of the production of a country was added for each year. After executing the below displayed R-code we get our newly transformed dataset which is suitable for plotting time series and fitting forecasting models. The models will be implemented using this transformed data. The data is without any missing values.

5.4 Analysis on Agricultural Data

In this section all the analysis done on the data to achieve the desired output is explained. After the data is clean pre-processed it is fit for further analysis. The research aims at getting maximum information from the data, which can be useful for decision makers. To draft policies for the welfare of the agricultural economy. This research also focuses on implementing time series for forecasting the future production of country. In addition, the reason behind the rise in global agricultural production will be discussed.

5.4.1 The trend from 1961-2013

Before implementing the forecasting models, it is necessary to go through the statistical values of past years. There are more than 200 countries, thus we need to target the countries which produce more. The types of crops and edibles that dominate the world production will be found. The production by the country is classified into food and feed. Food is the edible stuff for human population and Feed represents the stuff for animals and other living animals.

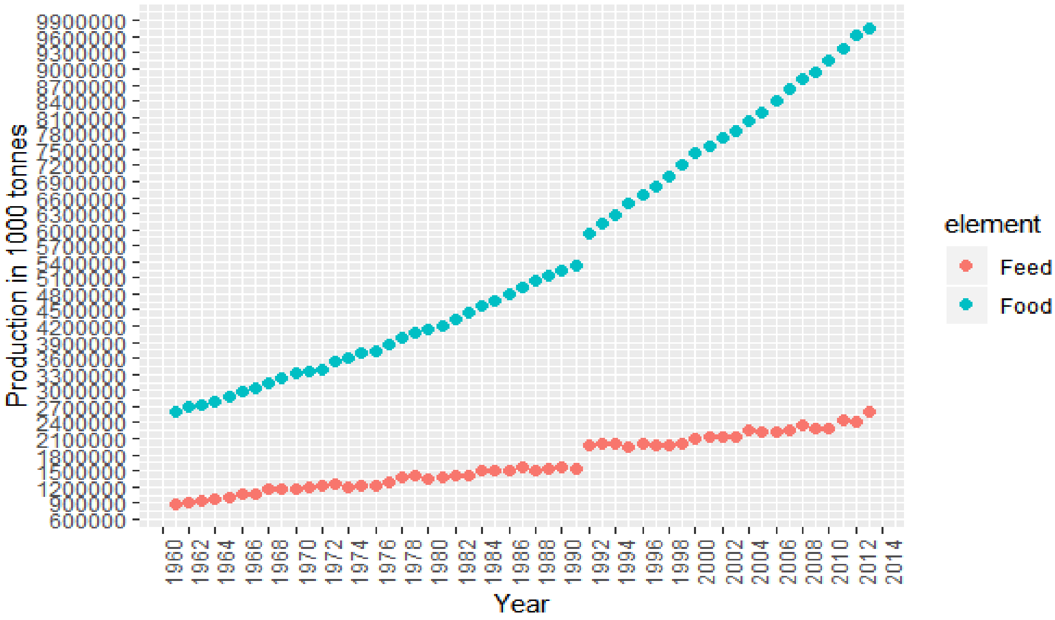

Figure 5: The trend of data

The above displayed plot shows the trend of data followed with time. It is observed that the production of food items has grown at a higher rate as compared to feed items. This occurred due to rise in population of the world, thus the production of the food items increased so dramatically, It is also observed that there is sudden increase in the production of both food and feed items during the year 1992. The reason behind this anomaly will be discussed later in the coming sections

A. Top producers of food:

For finding the top producers of food the code was written in R which was executed in R=studio. Using R the data was explored to a greater extent. For attaining the required output the use of R-packages were installed. The starting year is 1961, the aim is to observe the trend of the data till the year 2013.To find out the top producers the total production of each country throughout the years was added. This was followed by plotting of the output. The plot of last 5 years is plotted to show any change in the trend in recent times. From the plot obtained it can be concluded that China, India, The United States of America. Brazil are among the top producers of food since the year 1961. In the other graph was to show the data of past 5 years (2008-2013). By obtaining this result, the major producing countries are known. And thus for further analysis these countries will be focused.

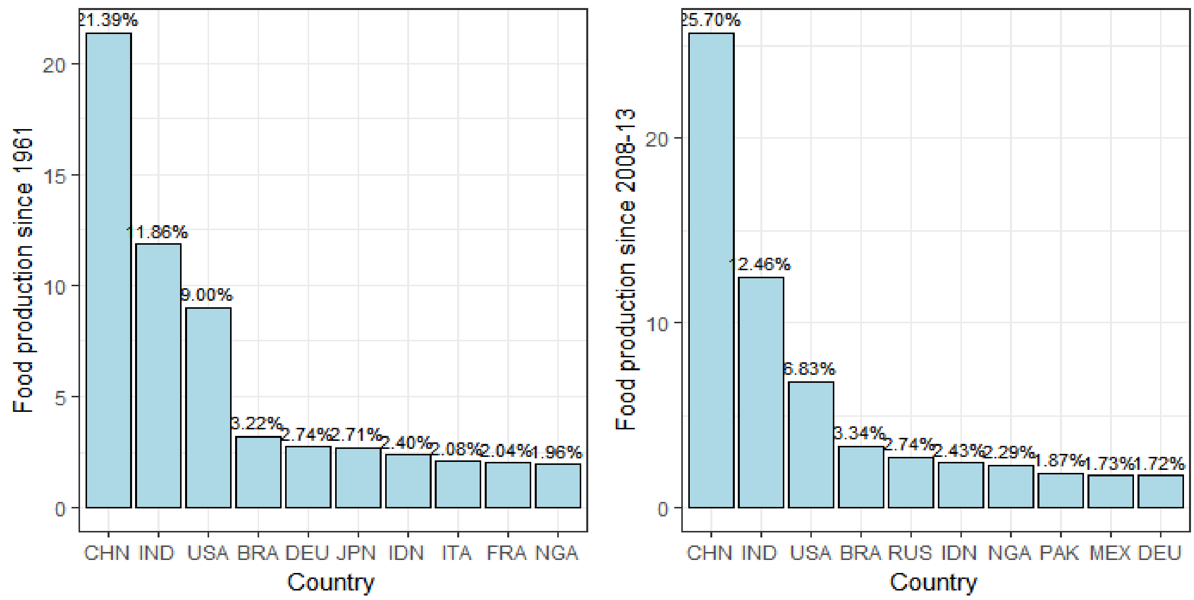

Figure 6: Top Producers of food

B. Top items produced

To get a better understanding of the data, the production of crops and food stuff must be calculated. To achieve this all of the production of each type of food stuff was added and the result was plotted.

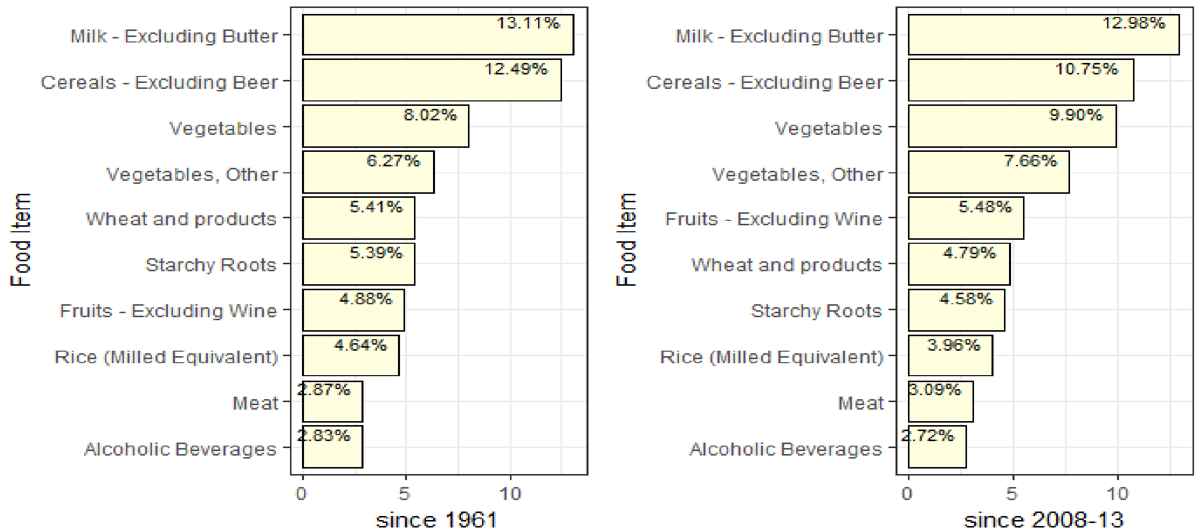

Figure 7: Top Food items produced

From the obtained visualization we can conclude that Milk, Cereals, Vegetables are among the food stuffs which are produced on a large scale since 1961, There hasn't been any significant change in the production in recent years. It is also observed the production of Fruits and Wheat has risen in the recent times.

C. The sudden rise in production in year 1992

Before the year 1992 the production of the countries was growing exponentially at certain rate. But during the year 1992 the production went up. It was difficult to find the reason behind this rise. While cleaning the data it was observed that there are some countries, which had their data from the year 1992. After studying about these countries it was found that these were the countries which were formed after the dissolving of U.S.S.R. Amon these countries Russia is the only large country.

	area	`1991`	`1992`
	<fct>	<dbl>	<dbl>
1	Armenia	0	4564
2	Azerbaijan	0	8624
3	Belarus	0	44161
4	Bosnia and Herzegovina	0	5319
5	Croatia	0	9525
6	Estonia	0	4771
7	Georgia	0	6315
8	Kazakhstan	0	43816
9	Kyrgyzstan	0	9041
10	Latvia	0	8167
11	Lithuania	0	9667
12	Republic of Moldova	0	11521
13	Russian Federation	0	455486
14	Slovenia	0	4087
15	Tajikistan	0	5550
16	The former Yugoslav Republic of Macedonia	0	3427
17	Turkmenistan	0	7096
18	Ukraine	0	157361
19	Uzbekistan	0	32167

Figure 8: Emergence of New Nations

3.5. Time Series Analysis

Time series analysis is a process of harnessing of meaningful information from the data. In time series forecasting models are used to forecast the future occurring values on the basis of data. Regression analysis is often misunderstood as Time series analysis, but regression analysis is time independent. As mentioned earlier time series divided into time domain and frequency domain. Time domain analysis comprises of auto-regressive techniques, along with cross-section analysis.

3.5.1 Forecasting using ARIMA and other models

It is a model based on auto-regression concepts. It is mostly used for predicting future values. Depending on the composition of the data it is classified as seasonal or non-seasonal data. The data used in our data doesn't show any seasonality. ARIMA models are classified on the basis of their order (p,q,k). The order of ARIMA model for non-seasonal data is given by the equation shown below.

AIC = -2log(L) + 2(p + q + k)

Where AIC = 'Akaike Information Criterion'

L=Likelihood, **p**= auto-regressive part

q=moving average, **k**=intercept of ARIMA model. If k=1 then c is not equal to zero. If k=0 then c = 0, thus ARIMA model doesn't intercept. Thus to avoid this issue the equation of AIC is written as follows.

$$AICc = AIC + (2(p + q + k)(p + q + k + 1)) \;/\; (T - p - q - k - 1).$$

$$BIC = AIC + (\log(T) - 2)(p + q + k)$$

BIC stands for 'Bayesian Information Criterion'

After the execution of the pre-processing steps, the data is ready for further model implementation of model. The transformed dataset contained the sum total of a country's production from the year 1961 to 2013. For testing the model's performance one country's production was taken into consideration. ARIMA models of different orders were implemented to get the clarity in the results. Exponential models along with Holt-winter's models were also used to test the forecasting value. After the implementation, the results of the models were compared. In the earlier sections it was acknowledged that China, India, Brazil and the USA are among the top producers of the world. Therefore the data from these countries was used to train the data. The results obtained from the implementations are explained in the results section.

3 Evaluation

In this section all the implementation carried so far will be evaluated, in order to find the precision of the result. Models like ARIMA, Holt-winters and other exponential models were implemented using the agricultural data. All of the models were implemented with the purpose of predicting the future values of the production in the coming 30 years. In this section the results obtained from models will be compared to find out the best suitable model. Data of a single country was fed to the data. It can be better to predict the future values of a top producing country.

3.1 Case Study 1: ARIMA model

The data used for training the model was without any missing values. In order to evaluate the performance, the data of one country was selected. Since China has been the leading producer of the world. The data of China was taken into consideration. In this case study ARIMA model of order (0, 2, 1). After the execution plots of the models results were obtained. Accuracy of the model was obtained by using accuracy() function, which is included in the Forecast library package. The MAPE ('Mean Absolute Percentage Errors') value obtained for this particular model was **2.405**. Along with this 30 future values of China's production.

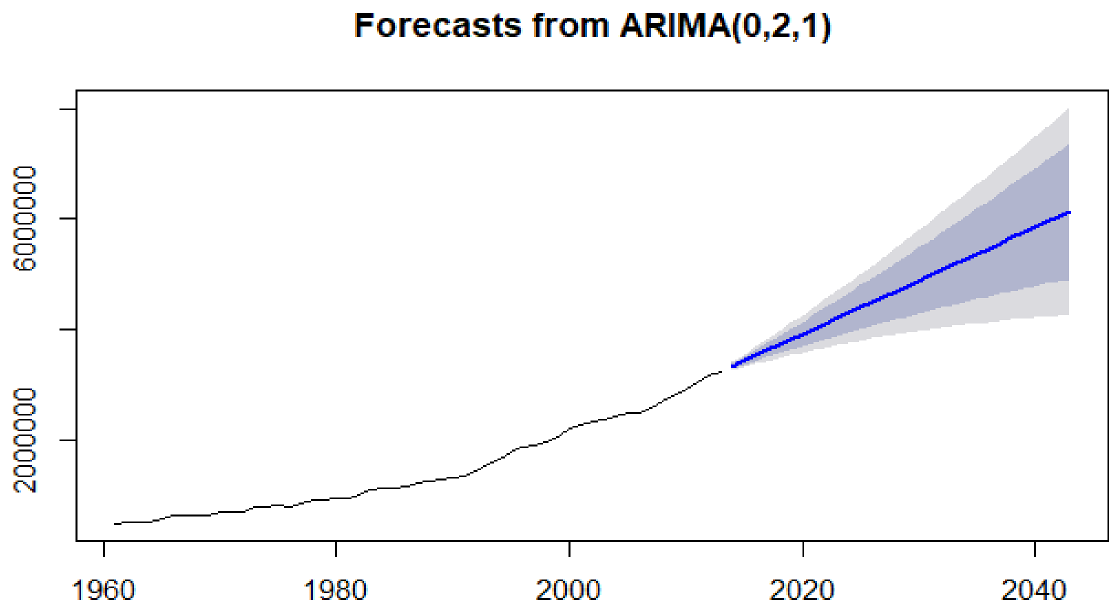

Figure 9: Forecast plot of ARIMA (0,2,1) model

3.2 Case Study 2 : ARIMA model combined with Simple Linear Regression

An attempt was made to combine the aspects of simple linear regression model with the moving average capabilities of ARIMA. This was done to find out, whether it performs better than ARIMA model. The regression model was implemented first using China's production data. The output of regression was named global prediction. This global component was subtracted from the actual time series to get local prediction value. This local prediction component was use to train the ARIMA model. The model was successfully executed. Future value and plots were obtained. The MAPE values of this model was **1.908**

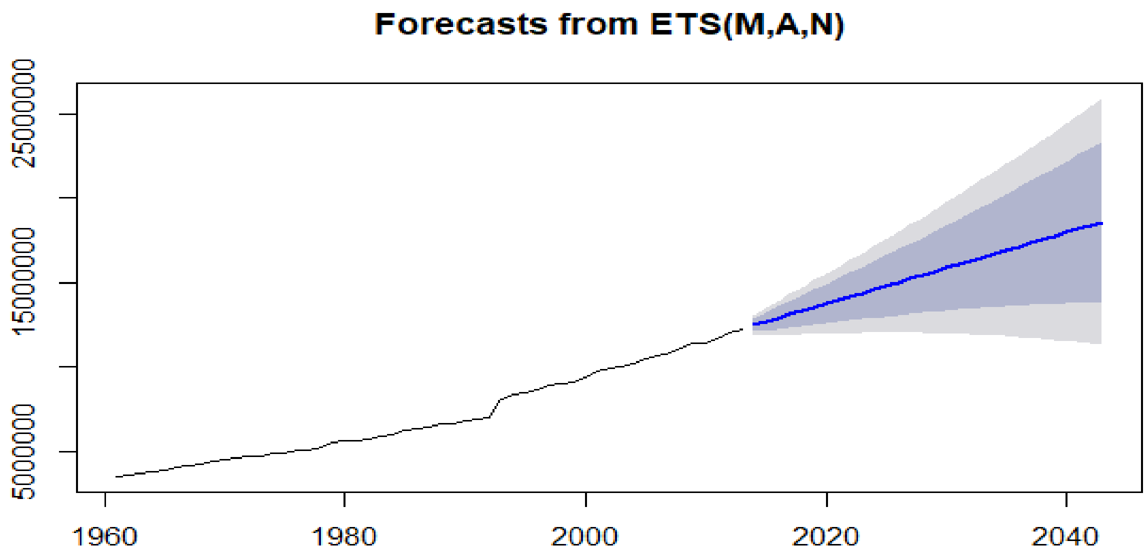

Figure 10: Forecast of Regression and ARIMA

3.3 Case Study 3: Holt-Winter's forecasting with Exponential smoothing model

Holt-Winter's is yet another popular model used for forecasting. The exponential smoothing function ets() does the task of estimation of initial states with smoothing parameters with optimization of likelihood function. Holt-winter's relies on heuristic values. Both the models were trained using China's production data. Holt-Winter's investigates the initial values and plots the fitted values with the observed values. The plot obtained from the Holt-Winters is displayed below.

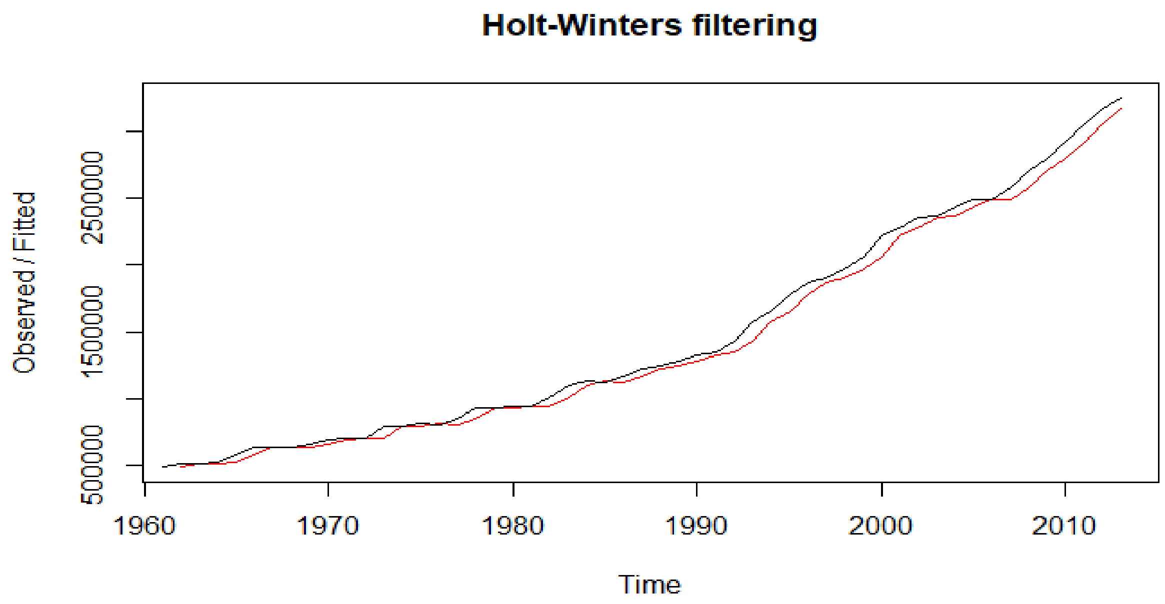

Figure 11: Observed/Fitted values

In the plot the black graph is for the observed values, while the red is for the fitted values. It can be seen that the observed values and fitted values are close. The forecasted values obtained were plotted. The MAPE value for Holt-Winters model is **3.658.** The forecasting of the exponential model was also obtained. The MAPE value observed was **2.482**. The forecast plots of Holt-Winters is displayed below.

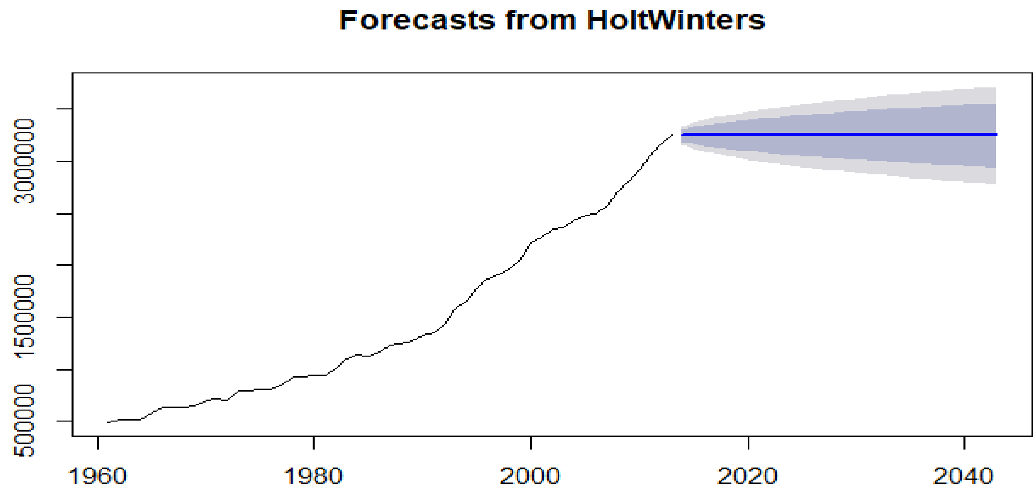

Figure 12: Forecast plot of Holt-Winters

3.4 Discussion

In this section all the case studies discussed above are compared to select the best model. After the implementation of the forecasting models, it was found that ARIMA model combined with regression achieved the least MAPE value among all models. Therefore based on the MAPE values ARIMA with regression will give the most accurate forecast. The results of all the models are included in the table displayed below.

Model name.	MAPE value	MASE value
ARIMA	2.405	0.54
ARIMA with regression	1.99	0.42
Holt-Winter's	3.658	1.00
Exponential forecasting	2.482	0.553

Table 1: Comparison of the forecasting models

Where,

MAPE: Mean Absolute Percentage Error
MASE: Mean Absolute Scaled Error

From the table it can be concluded that ARIMA with regression obtained the best results. Thus the forecasted values obtained from this model will be more precise, compared to other models.

4 Conclusion and Future Work

With the execution of all steps described in the above sections, the insights on the agriculture along with forecasted values. By gaining the future prediction of the countries we get to know how much a country will produce in the next 30 years. In the related work section the work done in agriculture domain was discussed. From the implementation section we get the insights that, China, India, Brazil are the leading producing countries of the world. Milk, Cereals and Vegetables are among the top food items to be produced in the world. As for the feed items, China is the leading producer followed by USA, Brazil and Russia. Cereals, Maize products, Starchy roots are among the most produced feed items. It was observed that India was among the top producer of food items. But when it comes to production of feed items, it doesn't top the list. This shows that India focuses much on the production of food for human population. In the year 1992 the global production went up suddenly. This was due to emergence of newer nations because of dissolving of U.S.S.R. It was acknowledged that for the task of forecasting, machine learning models were preferred. It was observed that ARIMA model was preferred number of times. Models like Holt-Winters and exponential models were also implemented for forecasting. After comparing the results ARIMA model with regression, it was acknowledged that ARIMA with regression performed well than other models. By giving the least MAPE value. Thus if the decision makers want to implement policies relating to agricultural production this research can serve as a base.

An attempt was made to implement clustering model to target the countries which follow the same graph of data over the time. If this is achieved in future, the country's production can be related to other country. The solutions implemented on a particular country can also be applicable on a country which belongs to the same cluster. Multi-variate time series can be implemented, targeting multiple countries at a time. The implementation of multi-variate time-series will give fast results and more insights on agricultural economy.

References

Bernard, A. and Jones, C. (1996). Productivity Across Industries and Countries: Time Series Theory and Evidence. The Review of Economics and Statistics, 78(1), p.135.

Yujun, Y., Yimei, Y., & Jianping, L. I. (2016). RESEARCH ON FINANCIAL TIME SERIES FORECASTING BASED ON SVM. 2016 13th International Computer Conference on Wavelet Active Media Technology and Information Processing (ICCWAMTIP), 346–349. https://doi.org/10.1109/ICCWAMTIP.2016.8079870

McQueen, R., Garner, S., Nevill-Manning, C. and Witten, I. (1995). Applying machine learning to agricultural data. Computers and Electronics in Agriculture, 12(4), pp.275293.

Savla, A., Israni, N., Dhawan, P., Mandholia, A., Bhadada, H. and Bhardwaj, S. (2015). Survey of classification algorithms for formulating yield prediction accuracy in precision agriculture. 2015 International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS).

Gandge, Y. and Sandhya (2017). A study on various data mining techniques for crop yield prediction. 2017 International Conference on Electrical, Electronics, Communication, Computer, and Optimization Techniques (ICEEECCOT).

Lourenco, L., de Camargo Salles, M., Gemignani, M., Gouvea, M. and Kagan, N. (2017). Time Series modelling for solar irradiance estimation in northeast Brazil. 2017 IEEE 6th International Conference on Renewable Energy Research and Applications (ICRERA).

Grover, P. and Johari, R. (2018). PAID: Predictive agriculture analysis of data integration in India. 2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom).

Hirata, T., Kuremoto, T., Obayashi, M., Mabu, S. and Kobayashi, K. (2015). Time Series Prediction Using DBN and ARIMA. 2015 International Conference on Computer Application Technologies.

Huang, Y. and Nashrullah, M. (2016). SVM-based Decision Tree for medical knowledge representation. 2016 International Conference on Fuzzy Theory and Its Applications (iFuzzy).

Ferenti, T. (2017). Biomedical applications of time series analysis. 2017 IEEE 30th Neumann Colloquium (NC)

Xie, J. (2009). Time Series Prediction Based on Recurrent LS-SVM with Mixed Kernel. 2009 Asia-Pacific Conference on Information Processing, 1(2), 113–116. https://doi.org/10.1109/APCIP.2009.37

Kazempour, M. (2007). Stationary Forecasting; Using Holt-Winter and a Modification of Holt-Winter. *Biometrical Journal*, 32(3), pp.347-356.

Grillenzoni, C. (1993). ARIMA Processes with ARIMA Parameters. *Journal of Business & Economic Statistics*, 11(2), p.235.

Comparison of Forecasting Energy Consumption in Shandong, China Using the ARIMA Model, GM Model, and ARIMA-GM Model. (2017). *Sustainability*, 9(7), p.1181.

Luo, H. and Wang, S. (2017). Based on the PCA-ARIMA-BP hybrid model of stock price prediction research. *ANZIAM Journal*, 58, p.162.

Elo, L. and Schwikowski, B. (2011). Mining proteomic data for biomedical research. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 2(1), pp.1-13.

Kumar, M. and Thenmozhi, M. (2014). Forecasting stock index returns using ARIMA-SVM, ARIMA-ANN, and ARIMA-random forest hybrid models. *International Journal of Banking, Accounting and Finance*, 5(3), p.284.

Foutz, R. and Lee, H. (2000). Adaptive Fourier Series and the Analysis of Periodicities in Time Series Data. *Journal of Time Series Analysis*, 21(6), pp.649-662.

Diop, M. and Kengne, W. (2017). Testing Parameter Change in General Integer-Valued Time Series. *Journal of Time Series Analysis*, 38(6), pp.880-894.

Marček, D. (2015). Hybrid ARIMA/RBF Framework for Prediction BUX Index. *Journal of Computer and Communications*, 03(05), pp.63-71.

Kriegel, H., Kröger, P. and Zimek, A. (2008). Detecting clusters in moderate-to-high dimensional data. *Proceedings of the VLDB Endowment*, 1(2), pp.1528-1529.

Kolassa, S. (2011). Combining exponential smoothing forecasts using Akaike weights. *International Journal of Forecasting*, 27(2), pp.238-251.

McQueen, R., Garner, S., Nevill-Manning, C. and Witten, I. (1995). Applying machine learning to agricultural data. Computers and Electronics in Agriculture, 12(4), pp.275293.