Sri Lanka Tea Auction Price Forecast By Using Data Mining Techniques

B.P Wickramasooriya

169342C

Dissertation submitted to the Faculty of Information Technology, University of Moratuwa, Sri Lanka for the partial fulfilment of the Requirements of the Degree of Master of Science in Information Technology.

2019

Declaration

I declare that this is my own work and has not been submitted in any form for another degree or diploma at any university or other institution of tertiary education. Information derived from the published or unpublished work of others has been acknowledged in the text and a list of reference is given.

Name of the Student	Signature of the student
B.P. Wickramasooriya	……
	Date:

Supervised by

Name of the Supervisor	Signature of the Supervisor
Mr. S.C. Premaratne	……
	Date:

Acknowledgement

I would take this as an opportunity to express my gratitude for my supervisor Mr S.C. Premaratne, Senior Lecture, Faculty of Information Technology, University of Moratuwa, who spent his valuable time throughout the project for giving me the proper guidance and maximum supervision for make this research success. Also would like to thanks for the all lectures taught us in the Masters programme who gave their best to encourage us sharpen our knowledge throughout these two years as they were the illumination which lit up our path ways to the success.

Moreover, I wish to grace to the Director General, Top Management and others of the NERD Centre who gave me valuable concession to the Master's Program. Also praise for the class mates of Batch 10, who gave me wonderful hands every time whenever necessary.

Finally I would like to thank for my wife and parents who gave me enamours support during the period by taking all the duties and responsibilities from me and make the freedom to carry out the studies.

Abstract

This report presents the results and analysis of the research carried out to predict the auction price of tea in Sri Lanka. Sri Lanka Tea Board statistics say the price of tea often vary significantly with time of the year. The key factors that influence the price of tea are weather, climate change, plucking season, production capacity, export tonnage, US dollar exchange rate and crude oil price, as a result of that stakeholders in this industry are affected seriously and it also badly affects to the GDP of the country. Further, Sri Lanka Tea Board has no proper mechanism to evaluate these changes and predict tea auction price. There were a few research carried out to analyse and predict the price of tea based on the environmental and economic factors, hence this study may be beneficial for those who are involved in this industry. The factors related to tea price variations in the auction are overcome from investigating the correlation between the price of tea and the above key factors. This study used weekly tea auction price in the past seven years and prices & quantities of the above factors during that period. These past data were used to evaluate and identify the strength of correlation between the key factors and its variation pattern to predict the price of tea. The classification methods in data mining process were used for the predictions and was based on the analysis of correlation of key factors over the time period. Then regression models were used to forecast auction price. This analysis used several regression algorithm methods, but Regression by Discretization algorithm was identified as the best method among them to build an accurate prediction model. Mean Squared Error (MSE), Mean Absolute Percentage Error (MAPE), Relative, Absolute Error (RAE) and Root Relative Squared Error (RRSE) were used to test and evaluate the accuracy of the results. These methods are not hundred percent correct to check the price variations due to fundamental drives such as human errors, natural disasters etc. However, this prediction model offer eighty five percent accurate, which is reasonable to forecast tea auction price.

Table of Contents

List of Figures

Page

Chapter 1

Introduction

1.1 Prolegomena

The Ceylon tea is a famous brand in the world for the past decades, due to its premium quality, colour and taste. Tea production is one of the major agricultural exports and key source of foreign exchange for Sri Lanka, and accounts for 3% of the country GDP. Sri Lanka in 4th position in list of the largest tea producers and the 3rd largest tea exporter due to the size of the geographical area of harvesting. Sri Lanka has a larger tea customer base around the world. Russia, most of the Middle East countries like Saudi Arabia, Qatar, Jordan, Dubai and some of the European Union countries who are the biggest customers of Sri Lanka and exporting large amounts on a daily basis to fulfil their needs. Most of the above countries are developed and richest countries in the world and top of the table of crude oil producers and exporting.

1.2 Background and Motivation

Colombo tea auction carried out once in every week to auction most of the Sri Lankan tea production in every part of the county. There are various kinds of tea categories available based on its taste, type and its harvesting locations. The harvested area, mostly categorizes into three parts based on the elevations. High grown, mid grown and low grown are the three main elevations and its production are variable due to the various kinds of weather conditions, plucking seasons and labour force involvements. Therefore the auction quantity may also vary in every week in the auction and the exporting quantity will be direct proportion to the auction quantity.

All the international finance transactions in the world, dealing with the most recognizes monetary units like US dollars and the foreign revenue of the country will come as also in US dollars. So the Sri Lankan rupee exchange rate relative to the US dollars is a huge factor

when considering foreign transactions and the rate is always varied as the Sri Lanka is still a dependent and a developing country.

Crude oil prices are also playing a huge part of the country's economy as well as the agricultural field. Transportation of crops, fertilizers, labours and etc. has affected based on the fuel prices.

Tea auction price changing time to time throughout the year and no one can give the guaranteed price for the certain tea category due to the high demand from the buyers. There are many parties who are interested in the auction price because of the tea is one of the highly moving agricultural products in Sri Lanka as well as the worldwide. Sri Lanka Tea Board (SLTB) is the governing body for tea in Sri Lanka and the tea auction is exerted and fully monitoring according to the guideline prepared by them. They also does not have the mechanism or any tool to give a prediction or forecast to the auction price.

1.3 Problem Statement

The problem related to the tea auction price is, it has no method to know and forecast the auction price or mechanism to guess the price before the auction is happening. If it does, it will be beneficial for many parties who are involved in the industry. This research is about to build some kind of a model which is to give the predictions about tea auction price forecast for few weeks ahead. The other factors mentioned earlier, which are normally not considered with the price of tea auction and could be affected to the tea and may give a direct or indirect impact to the tea price.

Therefore, in this research consider some of the supporting factors which are directly influenced by the tea auction price. Those are US Dollar exchange rate, crude oil price in the world market, weekly tea production, quantity of tea exports weekly, weekly auction quantity and weather seasons. SLTB and the tea experts was mentioned the above factors are given the most influenced to the auction price fluctuates from time to time. Therefore, choosing the above factors to build some kind of a model which can give the predictions about tea auction price forecast.

There are only few research carried out to predict the price of tea auction for a few months ahead. Also, as same as for the agricultural products. But most of the researchers used seasonal time series methodologies to evaluate and identify the relationships to predict the prices. Seven years of historical data of the above factors and the auction prices were used as the dataset for analysis and make the predictions. Classification of the data mining technique and regression algorithms were used to build a suitable model. The model identifies the strength of correlation between the auction price and factors and best data patterns to give the predictions. Regression and Time Series models are preferred when the research dataset consisting of historical data which are used to analysed and give the predictions. These methods are certainly not having the ability to check all the price variations due to the fundamental drivers like human errors, natural disasters, etc.…, but offer a way to have a stable price prediction for a few weeks ahead.

1.4 Aim

The aim of the project is, use data mining techniques to give a statistically proven prediction for the Sri Lanka tea auction price for a few weeks ahead by analysing affective factors.

1.5 Objectives

- Finding the relevant historical data for the price of tea auction and all the other factors that use in this project
- Data pre-processing
- Analyse the dataset to identify the correlation between factors and the tea price
- Selecting the suitable algorithm
- Building a suitable model using variables
- Applying the model for training the dataset
- Test the dataset to find the accuracy
- Use time series technics by applying model to forecast the price

1.6 Structure of the Thesis

This chapter has introduced about what is the problem statement having in the tea auction price and the solution of tea auction price forecast for few weeks ahead by building a predictive model create using data mining techniques. To analyse all the affected variables with the tea auction price and identify the correlation between them. Then build a predictive model using the classification algorithms which forecast the price of tea auction. The result will be tested using various test cases to check the accuracy.

Chapter 2

Literature Review

2.1 Introduction

This chapter discusses about the previous findings and works has been done so far related to tea and auction price. As mentioned in the earlier chapter there were few researches has been carried out on the tea auction price. Therefore, we will analyse other studies such as stock market analyse and forecast, other agricultural product based decisions which are related to the prediction based results together with tea price analyses to select the best techniques and algorithms to carry out this research.

2.2 The Brief

The Sri Lankan Tea industry maintains the highest quality in the world tea market and has the capability to produce the cleanest tea in the world in terms of minimum pesticide residues. The total tea cultivation area is about 200 hectares in the country. The major tea growing areas are Kandy and Nuwara Eliya in Central Province, Badulla, Bandarawela and Haputale in Uva Province, Galle, Matara and Mulkirigala in Southern Province, and Ratnapura and Kegalle in Sabaragamuwa Province. There are main six principal regions, planting tea -Nuwara Eliya, Dimbula, Kandy Uda Pussellawa, Uva Province and Southern Province.

The export revenue has been increased over the last two years and slightly dropped in the year 2015 due to the economic crisis in few top importing countries. However the value of exports has been increased by 5.43% in the year 2014. The tea exports account for about 15% percent of the total exports and about 65% contributes for the total agricultural exports in the country. The tea sector is expected to achieve the export target of US $ 3,000 Million in the year 2020.

According to the report [1] Ministry of Plantation and Sri Lanka Tea Board have introduced a wide range of assistance and development programs for this sector since many years. The EDB also has initiated a number of assistance programmes to assist tea exporters in the country. Also the report identified some of the weakness in the tea industry such as, limited cultivated land area, weather patterns, impurities of bulk tea when processing, lack of infrastructure, high cost of production and high labour, packaging transport and high cost of investment for new technology are some of them.

Limited land area available for cultivation and production expansion, increasing competition from other beverages, scarcity of labour are some of the challenges identify during the study of [2]. Also climate change plays a big role when it comes to production and need to identify appropriate and effective climate adaptation measures through modelling and impact assessment. The VAR model appears to be the most appropriate method for modelling tea prices by incorporating a group of interacting time series variables in order to explain the dynamic relationships among these time series in the system.

2.3 Tea Auction

The summary report [3] from the Chamber of Commerce of Sri Lanka mentioned, where the Colombo tea auction placed in every week, Sri Lanka Tea Exports for the Month of August 2017 was recorded at 24.8MTs which is a decrease of 1.34% when compared to the same period of the previous year. The FOB value at Rs.810.93 was Rs.180.56 above the FOB value of 2016. As a result the value of the Tea Exports for the month of August has exceeded by Rs.4.2 billion over the previous year. The tea export quantity for the first seven months of the year has recorded a 3.75% decrease. However recording a 25% increase in export earnings due to the higher FOB value of Rs. 189.60 when compared to the same period last year

Xia, Liu and Zhou [4] Found that the price of the tea may varying due to quality of a tea. According to this study of Data Mining framework for tea price evaluation, they try to develop an automatic system to evaluate the tea price rather than using manual systems.

During the study, they have to use data mining techniques to detect data anomalies, feature selection and for the classification of the dataset.

The research [5] of Analysing Tea Auction Trends for Beneficial Seasonal Tea Production in Sri Lanka, Rajapakse focused on extracting seasonal demands from tea auction history records to better facilitate the brokers predict demand of upcoming sales. Association Rule Mining is the technique which carries out limited passes in a dataset to identify rules among discrete attributes and exposing relationships among them. They applied the Association Rule Mining technique on the auction data to see the impact of Grade, Gross Weight, Sale Week, Month, Factory Code and Selling Mark on the price of tea lots. The result of the study identified that some of the months have low prices on the tea because of all the tea factories release their stocks to the auctions and some months have produced low quality teas and the remaining tea gets a high price. According to this study the tea auction prices may vary throughout the year.

2.4 Tea Industry

The gaining insight of tea industry of Sri Lanka using data mining, the research conducted by [6], to investigate and analyse the relationships, patterns and trends between time, tea category, tea production, exports, and auction price. Various kinds of multiple data mining techniques (k means clustering, regression, time series analysis) have been used to analyse the obtained data set by Sri Lanka Tea Board. All the factors are taken individually and analysed the correlation exist between each and every factors to find the pattern. But for more accurate forecast it must be analysed the combination of all the related factors and identify the correlation between them.

In this study clustering has been used at the time, price and the three kinds of tea productions. The result of the cluster analysis showed that the low grown tea category has contributed a lot more than other high and medium grown category to price and production. The time series analysis based on Autoregressive Tree Model (ART) has been used in production, exports and price with time to explore the trends of the variables with the time. The result of the regression analysis shown that the production and exports are correlated and the price in independent of the other two variables.

Limitation the sizes of the data set have caused to limitation to perform the data mining tasks on different factors. To identify the dependencies in clear form of the factors, it should be tested as a combination. The result has produced a complex idea due to the various kinds of data mining techniques being used and each of the technique has produced the different kind of result.

2.5 Tea Price Predictions

Hettiarachchi and Banneheka [7] carried out a research about forecasting the unit price of tea at Colombo tea auction for the one month ahead by using time series regression and artificial neural network approaches. It was used the previous 160 months of tea prices in seven auction centres from worldwide and analysis with the Colombo tea auction price and build two models to get the most accurate output results. The results are assessed by using the Mean Squared Error (MSE), Mean Absolute Percentage Error (MAPE), coefficient of determination and correlation coefficient between observed and fitted values to get the forecasted price. The data set was divided into subsets and one used to train and other used for testing the results.

The time series regression model is one of the successful methods can use to give the predictions for longer periods if the resulting data may not affect by other conditions. But for the inherently noisy, unstable, chaotic tea auction price forecast for the short period, the using of only the time series regression model is not adequate. According to this research, they only used previous month's price list for predicting the future prices. But the results may be not accurate because there are so many other factors may affected by the tea price. Not analysing the correlation of other related factors, the prediction results may not come as accurate as the real price. Also the data set may be not big enough for the training and the above approaches will give faulty results at the end.

Mathematically combined previous prices and exogenous variables used by modern statistical models to predict the future prices. According to [8], time series forecast method SARIMA is examined under the statistical models and those are differ from fundamental

models not only by explaining part of the data, but also predicting the unexplainable part as a sequence of past observations without a comprehensive approach.

Senaviratne [9] has done a research to forecast the tea auction price using Box-Jenkins modelling approach. The data set consist between 1996 to 2015 monthly auctioned tea prices and the SARIMA model is used to analyse the dataset. As per the research, there are a high percentage of price variation between months due to the various reasons like seasonal weather conditions and other factors and SARIMA model is well suited to analyse the data variations when the time series data exhibit seasonality periodic fluctuations that recur with about the same intensity each year.

$$\varphi_P(B^s) = 1 - \varphi_1 B^s - \varphi_2 B^{2s} - \ldots - \varphi_P B^{PS},$$
$$\phi_p(B) = 1 - \phi_1 B - \phi_2 B^2 - \ldots - \phi_p B^p,$$
$$\delta_Q(B^s) = 1 - \delta_1 B^s - \delta_2 B^{2s} - \ldots - \delta_Q B^{Qs},$$
$$\theta_q(B) = 1 - \theta_1 B - \theta_2 B^2 - \ldots - \theta_q B^q,$$

2.1 SARIMA Model

Augmented Dickey- Fuller (ADF) test, Lagrange's Multiplier (LM) test, White's General test, Jarque-Bera (J-B) test were used to verify adequacy of the model. According to the observations made by analysing the result, the forecasted prices have more than 15% of the error rate with compared to the real price of the test set and was tested by using Mean Absolute Percentage Error.

Senaviratne NAMR was doing the research by only analysing the tea auction price historical data for predicting the price forecast for monthly basis. But the tea auction held once in a week and the tea auction price needs to predict on a weekly basis. Only time seasons are considered, but there is no indication of weather conditions are being to consideration and same as no other related factors.

ARMA model used to analyse the weekly tea auction price, the study carried by [10]. The data were tested for stability, autocorrelation and partial correlation teat and the ARMA model established. Due to the longer production period, the fluctuation of the price of

agricultural products has a great influence on the agricultural producers. As an agricultural product, tea's price is a cyclical fluctuation, which is closely related to the cyclical fluctuation of agricultural production.

During this study, used 2 years of previous auction prices for analysed and the last few weeks to test the results. The results shown some variation about the forecasted price and the real price. There is a considerable error rate. This is due to the small data set that used for the study. The data set was not enough to give the prediction or forecast the prices. At least it should have minimum 5 years back data set and then the result could be significantly different than these forecasted prices.

The study of forecasting, black tea auction prices of capturing common seasonal patterns by [11] used previous months tea auctions prices from 8 tea auction centres around the world as the data set. SECM and Vector Error Correction (VEC) models have been used to test and forecast the results. The accuracy was tested by using MSE and MAPE. Also, this study identified some seasonal co-integration relationship between some auction centres. The unit root test was confirmed about the zero frequency for data and the result suggests that tea auction prices are not stationary and include a common trend and auction prices are varying according to the seasonal patterns. But for this research, seasonally unadjusted data were used as only the single set of series. By combining the data set with the production data and other related factors, this will lead to predict the most accurate forecasting auction prices.

2.6 Why these factors

2.6.1 Crude oil

Sri Lanka's main tea export destinations are Russia, Middle East countries like the United Arab Emirates, Qatar, Jordan, and some of the European Union countries and most of them which are the top in the list of world crude oil producers. They are extracting large amount of oil in daily basis and their main revenue generate by selling those oils in the international oil markets in Singapore and Europe. Russia Exports US $400 billion of oil annually, which

is 2/3 of the total exports of the country and the shocks in the oil price will definitely make an impact to their economies and [12] Liudmila Popova explains that how the oil price shock effect on the economy of Russia in 2003 and what are the actions was get after to neutralize the situation and what are the policies was introduce to face the certain situations. Also Saudi Arabia, which totally depends on their economy by exporting crude oils. Whenever the shock comes in they handle the economy by controlling the expenditures as well as imports. In certain situations, the interest to import products getting reduced and that will cause to reduce the amount of imported products and the commodity price of the certain products.

The Middle East and Russian markets account for 62% of Sri Lanka's tea exports by volume and 59% by value. Fluctuation of oil price will give an impact on the amount of tea exports and the commodity price in the auction in Ceylon tea. Therefore, due to the above reasons choosing crude oil price as one of the affecting factors, which will give an impact to the auction price.

Tea review by F & W [13] discuss on the observation results on tea price trends. It is said that the increases of the consumption around the world is one of the main contributory factors to price change. According to the study, another factor that needs to focus is the steady growth of oil prices, which would have a positive impact on Sri Lankan tea price. It is clearly stated that oil price considered as a major factor affecting the price of tea in Sri Lanka when compared to other nations.

2.6.2 US Dollar Exchange Rate

The researchers have done many of the researches about how the dollar exchange rate volatility will effect on exports of the country and the outcome of the result gave a mix ideas about how was the impact when fluctuation the rate. The empirical literature reveals that the effects of exchange rate volatility on exports are ambiguous. While a large number of studies find that exchange rate volatility tends to reduce the level of trade, others find either weak or insignificant or positive relationships. Many of the above researches have done the research about the developed countries which has a highest GDP (Gross Domestic Price) and did the least about the developing countries by little attention. [14] E.M

Ekanayake investigate about the effect on exchange rate volatility on aggregate export volume of Sri Lanka, and found that the uncertainty of the exchange rate will decrease the amount of the export in Sri Lanka. Also express that Sri Lanka should follow some monetary policies to stabilize the Sri Lankan Rupee against the foreign exchange rates to keep the economy stable, which give huge influence to do exports and imports neutral way.

As mentioned the above gives a clearer understanding about how the US Dollar exchange will give an impact to the economy of the country like Sri Lanka which is still in development phase. Decrease the value of the SL rupee compared to other foreign finance will inflation of the country also influence badly in the economy. Therefore, due to the above reasons US dollar exchange rate can get as one of the key factors to give an influence to the tea auction price as tea is one of the major agricultural exports in Sri Lanka.

2.6.3 Tea Production

World tea production has increased significantly in the past decades, same as in Sri Lanka also increase its tea productions due to the various reasons like to supply to the demand of the population sizes, the increasing social acceptance of tea as the drink of choice, an increase in the area of tea cultivation, improved varieties of tea cultivation by selective breeding, advanced technology and improved cultivation practices. New parties who interested in tea industry also joined and make and increase the competitiveness of the industry by researching and identifying various kinds of new flavours to the tea also being some reasons to increase the tea production.

Mutegi, [15] has done the research to investigate the relationship between the monthly tea productions vs. tea price. The study used a distributed lag model to the dataset of the past 20 years of data obtained by the Central bank of Sri Lanka and Sri Lanka Tea Board. Researchers found that, the black tea has a significant and a detectable effect on black tea production which is two and a half months on an average.

There are 34 countries around the world who has produced tea and the biggest producing countries like China, India, and Sri Lanka are located in the South East Asia region. Most of the above countries as well as Sri Lanka increased the harvested area to increase the

production. People move to the upper area by cutting Hill Mountains and areas to spread harvest. Always do the researchers to investigate and identify new technologies to adapt and grow the quality tea trees as well as economical ways to harvest area to get this without losing its quality as a production. Due to the above reasons cause to increase the production in the last decades and the tea factories in Sri Lanka will bring the largest amount of its production to the auction. That will give some impact to the price at the auction. There is some small amount of direct sales to the local customers, but most of the tea production will present to the auction. Therefore, tea production chooses as one of the factors to consider when forecasting the tea price at the auction.

2.6.4 Tea Exports

Most of the auctioned tea quantities will go to process and send to the exports. Exporters purchasing their maximum to fulfil the needs of the people as well as the foreign requests. The exporters will purchase the presented tea stocks at the tea auction and tea auction price will make them to decide the amount of the quantity which they are going to purchase. Because the minimum quantity is 1000Kgs.

During the recent past years, some of the countries tea export quantity had some decreases and some may the increases its export quantity rapidly and Sri Lanka is among in the second group which increases their export quantities. Because of Ceylon tea never lose their premium quality and producers and exporters make really hard works to maintain the pure Ceylon tea quality. The Sri Lanka Tea Board always conducts various kinds of tests to make sure that the tea is up to the quality standards. Therefore the Sri Lanka does not meet to the large fluctuation when exporting.

According to the past data there are some variations in tea price compared to exports quantity. When the amount or quantity of exports are going high, the price has shown some increase and same happen when the vice versa. Therefore the amount of tea exports will give impact to the tea price at the auction and choose as a factor to forecasting the tea auction price.

2.6.5 Climate Seasons

Sri Lanka is situated near the equator and lucky to experience of some of the changing weather conditions throughout the year. Countries near the equator are rich in agricultural products mainly due to the above reason and the same applied to the Sri Lankan agricultural products and Ceylon tea. By looking to the weather conditions can identify 4 main season patterns that used as tea plucking seasons. Some may have a heavy rainy condition which has above 100mm daily rain percentage and some very sunny condition which is around 30 degrees of Celsius. Plucking tea leaves percentage are varied with the seasons, therefore weather seasons takes as the last factor to evaluate and predict tea auction price forecast.

According to the study [16], Challenges related to weather and climatic changes have greatly led to fluctuations in the tea industry earnings, in Kenya. Forest attack on the tea growing areas and dry and hot weather conditions which was not favourable for tea production. Further, Tea industries in Kenya rely on hand labour which is becoming expensive. Apparently, tea harvesting is very expensive

2.7 Regression

Han Siew and Jan Nording [17] has done the study for predict the stock price trend by using regression techniques. As per them the regression techniques are useful when predicting the dependent variable value based on the other independent variable values. They study various kinds of regression algorithms such as linear regression, additive regression, regression by discretization and simple linear regression and the results were obtained are much similar to the real values that they were used.

Also, they found that the use of different data types by transforming real numbers into categorical ordinal data can improve the outcome of the regression techniques and outcome are favourable when less structured data are transformed into more structured data in ordinal form.

2.8 Summary

This chapter describes the previous studies relate to the tea industry, tea auction, tea price and data mining techniques which are used to assess the results. Also mentioned the factors which are directly related to the tea auction price variation. The next chapter will be about data and technology related to this research.

Chapter 3

Data and Technologies used

3.1 Introduction

In here we discuss about the dataset used for research and what kind of technologies we used to mine the dataset and obtained the results.

3.2 Data

This research used previous historical auction price data as the dataset and used those to predict and forecast the auction price for a few weeks ahead. Also the historical data of other supporting affecting factors data also used to forecast the auction price accurately. The other supporting factors are the crude oil price, US dollar exchange rate, tea production, tea exports and weather seasons used to plucking tea.

Type of data found in the datasets used for the tea price analysis mainly consisted of continuous data. The climate, season data are the only categorical values found in the data set. The weather, climate during the year are categories into 4 seasons as mentioned by the meteorology department of Sri Lanka.

There are different kind of data sources are used for collecting the historical data sets. Tea auction price, tea production, tea export quantity and tea auctioned quantitative data obtained from the Sri Lanka Tea Board. Crude oil prices were collected from the Federal Reserve Bank from St. Louis and the US Dollar Exchange rates collected from Central Bank of Sri Lanka.

3.2.1 Tea Category

Mainly tea divide into three categories based on the elevations. Those are high grown teas, mid grown teas and low grown teas. Above 4000 feet elevation tea are categories into the high grown tea and normally whole area of central province and some areas in Uva

province use to harvest the high grown tea. Between 2000 feet to 4000 foot area used to harvest the mid grown tea and beyond the 2000 foot area used to harvest low grown teas. Tea also categorizes according to the tea types like green, CTC, but for this research used elevation categorization only.

3.2.2 Tea Auction price

The Colombo tea auction holds at once a week and doing tea auctions for tea categories separately. Therefore, prices are available in separately according to the elevations along with the auctioned quantities. The source of the data is the web site of Sri Lanka Tea Board's official website and it contain the auction prices along with the quantities on a weekly basis from 2011 September onwards.

3.2.3 Tea Production

Tea production data also mentioned in the SLTB web site and production also categorize in to 3 elevations. Production also varies from time to time and taken to study to find some useful patterns according to the auction price.

3.2.4 Tea Exports

Total tea export quantity from weekly basis according to the elevations are taken into study to analyse the correlation between the auction price and exports to identify the pattern. The source also a SLTB statistical department, which is the official data holder for Sri Lanka Tea.

3.2.5 Climate Seasons

According to the data provided by the Meteorology, Department of Sri Lanka, the 12 months of the year can be categorized into four climatic seasons. The temperature, level of humidity and natural rainfall differ from season to season and this makes some effects on

the productions, because there is elevation base tea plucking seasons and natural crop can be different from climate season to season. The climate, seasons as follows.

1. First inter monsoon season – March –April
2. Southwest monsoon season – May – September
3. Second inter monsoon season – October –November
4. Northeast monsoon season – December- February

3.2.6 US Dollar exchange rate

US Dollar exchange rate compared to Sri Lankan rupee is changing rapidly from past year due to the unstable income of foreign revenue, various kinds of environmental reasons and country's political decisions. The data set was obtained from the Central Bank of Sri Lanka and the rate is on per day basis. For the mining purpose the data were transformed to weekly basis price rates.

3.2.7 Crude oil price

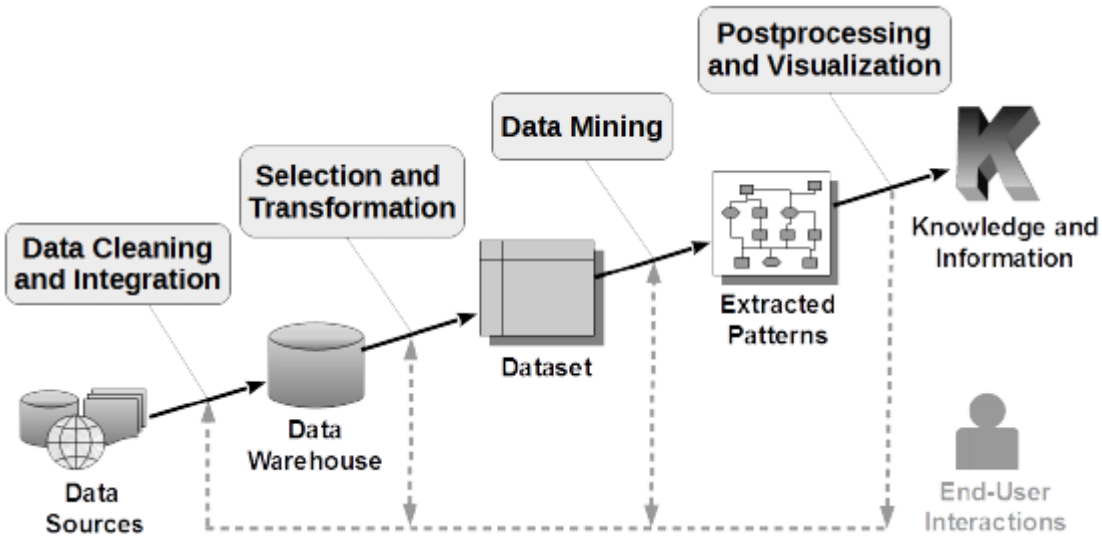
Price of the crude oil also changing rapidly due to the various reasons like political reasons of exporting countries, geographical and environmental reason. Therefore Europe oil market will always update their oil prices daily basis. Obtained data set from the Economic Research in Federal Reserve Bank of St. Louis which contained daily oil price from 2010 onwards and transform into weekly prices.

3.3 Technologies Used

There are various kinds of limited technologies were used for the research to analyse and build the model.

3.3.1 Data Mining

Data mining is used to extract the knowledge of, interest data from the large databases which is implicit, non-trivial, previously unknown and potentially useful data. Mainly data are taken from several databases and do pre-processing to remove dirt, noisy data and do integrations and store in one database call data warehouse which contain all the required data to be mined.



3.1 Data mining phases

Then select the data which need to mine and transform using normalization and discretization according to the requirement and take all in one stage. After that, can apply any suitable mining technique and searching for correlation with the factors to obtain a pattern and develop a model from the database and used it for knowledge extraction.

3.3.2 Weka

Weka can be used for the process of modelling data and it contains tools for data pre-processing, classification, regression, clustering, association rules and visualization. It is also well suited for developing new machine learning schemes. Weka (Waikato Environment) is an open source software written in Java language, developed at the

University of Waikato, New Zealand and it is free software licensed under the GNU General Public License.

The whole study was mainly done using WEKA software. From the data entering stage, all the other stages of data pre-processing, data normalization and transformation, classifier selection, training the data set and to testing the result, WEKA were used. The model was built also using this application and gave the clear output result for the predictions which can forecast the tea auction price.

3.3.3 Microsoft Excel

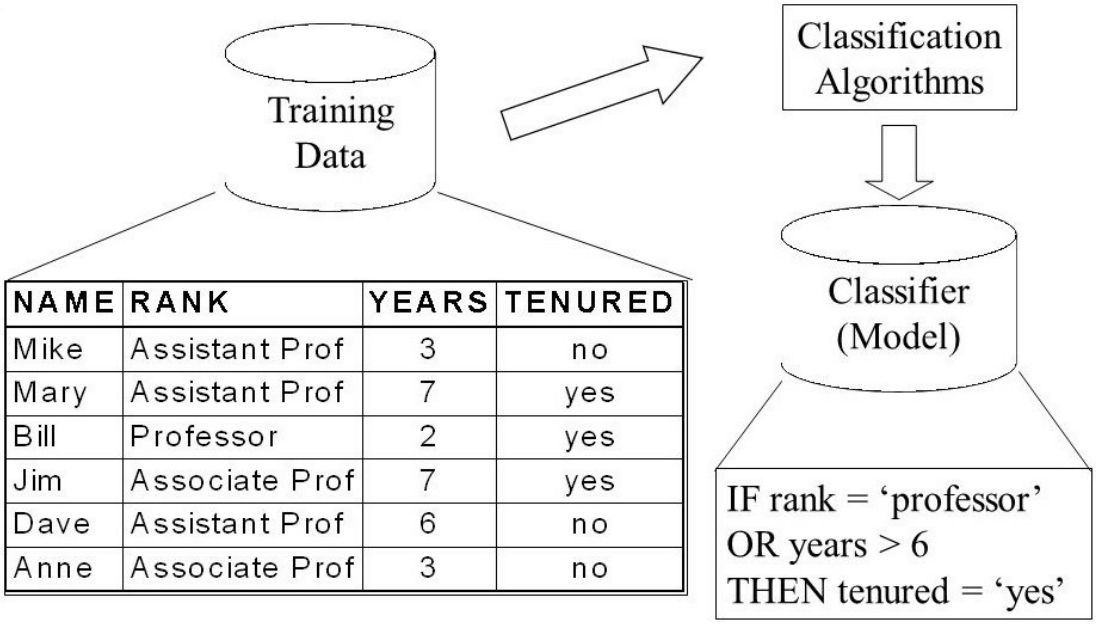
Microsoft Excel is a spreadsheet program that uses for doing the calculations and analysing the numerical data. Can use as the simple database to record the data. Features of pivot tables, graphing tools, macro programming are included in excel to do works more comfortably.

For this research Microsoft Excel has been used to record the data from various sources. After that processed the data by adding those to the table and merge the variable datasets into one common dataset based on the tea grown level. WEKA only read the .arff files which are the own outcome of WEKA and comma separated values (CSV) files. So the Excel help to create those CSV files which can read through the WEKA. Also, this helps to analyse the correlations of the variable with the class data by drawing the graphs. The various kinds of excel graphs are given the opportunity to select best graphs to show the result that make instant and clear idea for the readers.

3.3.4 Classification

In data mining, classification is a variety of data analysis used to extort models to describe the essential data module or to expect future data trends. The classification method is a two-step process. The first part is learning practice and model construction, in which training data will be analysed. The learned type or classifier shall be characterized in the shape of classification regulations. This describes a set of predetermined classes. Each

tuple/sample is assumed to belong to a predefined class, as determined by the class label attribute. The set of tuples used for model construction is called training set. The model is represented as classification rules, decision trees, or mathematical formulae. The other level of classification practice, in which test information to calculate approximately the exactness of classification style or classifier. If the exactness is acceptable, the regulations can be useful for classification of new data.



NAME	RANK	YEARS	TENURED
Mike	Assistant Prof	3	no
Mary	Assistant Prof	7	yes
Bill	Professor	2	yes
Jim	Associate Prof	7	yes
Dave	Assistant Prof	6	no
Anne	Associate Prof	3	no

3.2 Classification technique

Model usage is the 2nd step in classification. For classifying future or unknown objects, this is used. This model estimates the accuracy of the model. The known label of the test sample is compared with the classified result from the model. The test set is independent of training set [18]. For e.g.: If some tuples with certain data is given in the training dataset, in which these tables are distributed among different classes, then this dataset is used to further determine the class of the new tuple arrived for classification.

3.3.5 Regression

Regression analysis is a technique comes under the classification techniques which used to investigate the relationships between variables. The statistical significance or the degree of confidence of the relationship is also assessed in the process of regression analysis. The complex mathematical parts are avoided and only useful concepts regarding data analysis are included in the field of data mining. Data mining tools and packages include regression as a predictive data mining algorithm and have categorized it under Classification. WEKA includes regression as a method for main data mining problems and a standard data mining method with classification, clustering, association and attribute selection. Implementation of many regression schemes utilized in data mining exist nowadays and it includes multiple and simple linear regression, pace regression, a multi-layer perceptron, regression by discretization, support vector regression, locally weighted learning, decision stumps, regression and model trees and rules. Also, regression is utilized in pricing decisions where the factors affecting the prices can be figured out. The tea auction price analysis is also a price prediction where the regression can be used.

Regression is often used to determine how many specific factors such as the price of a commodity, interest rates, particular industries or sectors influence the price movement of an asset. The aforementioned CAPM is based on regression, and it is utilized to project the expected returns for stocks and to generate costs of capital. A stock's returns are regressed against the returns of a broader index, such as the S&P 500, to generate a beta for the particular stock. Beta is the stock's risk in relation to the market or index and is reflected as the slope in the CAPM model. The expected return for the stock in question would be the dependent variable Y, while the independent variable X would be the market risk premium.

Additional variables such as the market capitalization of a stock, valuation ratios and recent returns can be added to the CAPM model to get better estimates for returns. These additional factors are known as the Fama-French factors, named after the professors who developed the multiple linear regression model to better explain asset returns.

3.4 Summary

A different kind of historical datasets and technologies about to use in this research were discussed in this chapter. The next chapter will be how to use data mining methodology for the research.

Chapter 4

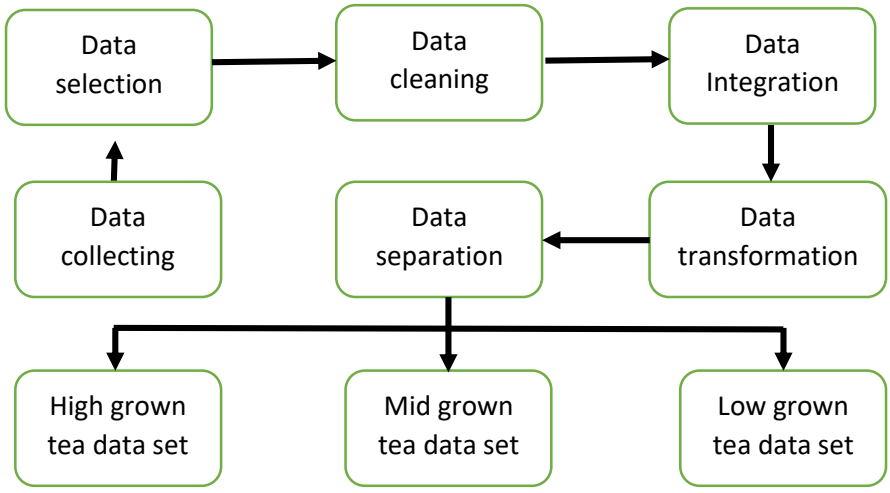
Data Mining

4.1 Introduction

In this chapter discuss about how the steps of KDD process are carried out to obtain results. The activities or the actions performed on the data set to produce the expected output during the phase of the KDD process.

4.2 Pre Data Mining Phases

According to the KDD process, the pre data mining phases are consisting with identifying the problem, identifying data sources, identifying data types, selecting data from databases, data cleaning and pre-processing and data integration. Following are the steps according to this research and describe as follows.



4.1 Pre data mining process

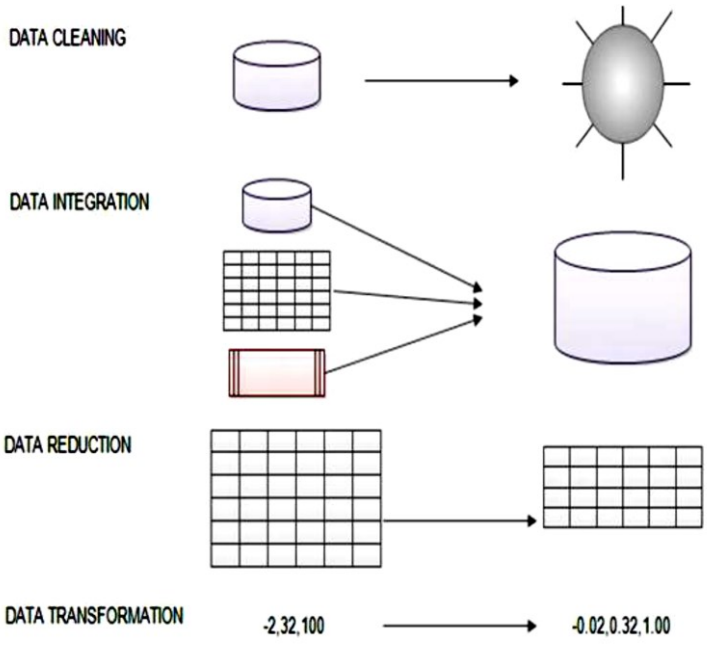
4.3 Select the Data

Selecting data is to decide what type of data is to be used for analysis. Criteria for selecting data from the available data source include relevant to the data mining goals, quality and technical constraints. Data selection deals with both attribute selection and the selection of records in a table from the identified data set.

In the tea auction price dataset, the time period for which data values were available for various factors differed from each other. One example is the auction price is available from the 2011 September onwards weekly basis and the crude oil price available from 2000 onwards daily basis. Therefore, in order to build the consistency between the data sets and we are dealing with the time series data for all factors where time is also playing a vital role. So the unnecessary rows with data are removed and the time period of 2011 September to 2018 April data were selected on a weekly basis, because the tea auction is held once in an every week. Crude oil price and the US dollar exchange rate data consist of daily basis and take the average of the weekly price of that and prepare the rows according to the auction price data to get the consistency of the dataset.

4.4 Data Pre-processing

Data in the real world is dirty, incomplete and noisy. Incomplete in lacking attribute values and lacking attributes of interest or containing only aggregate value noisy in terms of containing errors or outliers and inconsistent containing discrepancies in names or codes. Now the question arises why is the data dirty? Because incomplete data may come from ―not applicable‖ data value when data have to be collected and the major issue is a different consideration between the times when the data was analysed and human hardware and software issues are common. Noisy data may come from the when a human entered the wrong value at the time of data entry as Nobody is perfect. Errors in transmission of data and instruments that collect the faulty data. Inconsistent data may come from the different data sources. Duplicate records also need data cleaning.



4.2 Pre-Processing

Raw data are highly susceptible to noise, missing values, and inconsistency. The quality of data affects the data mining results. In order to help improve the quality of the data and, consequently, of the mining results raw data is pre-processed so as to improve the efficiency and ease of the mining process. Data pre-processing is one of the most critical steps in a data mining process which deals with the preparation and transformation of the initial dataset. Data pre-processing methods are divided into following categories.

- Data Cleaning
- Data Integration
- Data Transformation
- Data Reduction

4.4.1 Data Cleaning

Data of the real world consisting with the dirty data and containing errors and outliers. It can also be inconsistent and containing discrepancies. Quality data are needed in order to proceed through an accurate and smooth data mining process and that will be produce some quality results. Cleaning data phase consists of handling missing values. Identifying outliers, smoothing out noisy data and correcting of inconsistent data. Some of the production quantities and export quantities had missing values and figured out by handling missing values, methods according to the data mining process.

4.4.2 Data Integration

In this phase, data from multiple sources are combined into coherent storage. Identifying similar entities with different names while combining and detecting and resolving data value conflicts such as similar entities with different measurement in different units are some of the problem that needs to be solved during the integration process. Two methods used at two different times before data transformation and after data transformation, for integrating data are merging data and generating aggregate values. Both these methods were used with the tea auction price dataset. In order to construct the attribute climate, seasons during the data transformation stage, the weekly values needed to be generated. Due to this reason a data integration phase of generating aggregate values was conducted prior the data transformation phase. The data integration phase of merging data was conducted after the data transformation phase because all fields, including the weather, climate seasons should be present in a dataset or data table for factors to be merged.

4.4.2.1 Generate aggregate values

The crude oil price and the US dollar exchange rate data consisted of more than one value for a week due to the changes taking place in the price of crude oil and exchange rate several times per week. All the other factors, including price for the tea auction consisted of a single value per week. Since the major purpose was to compare all the factors with tea

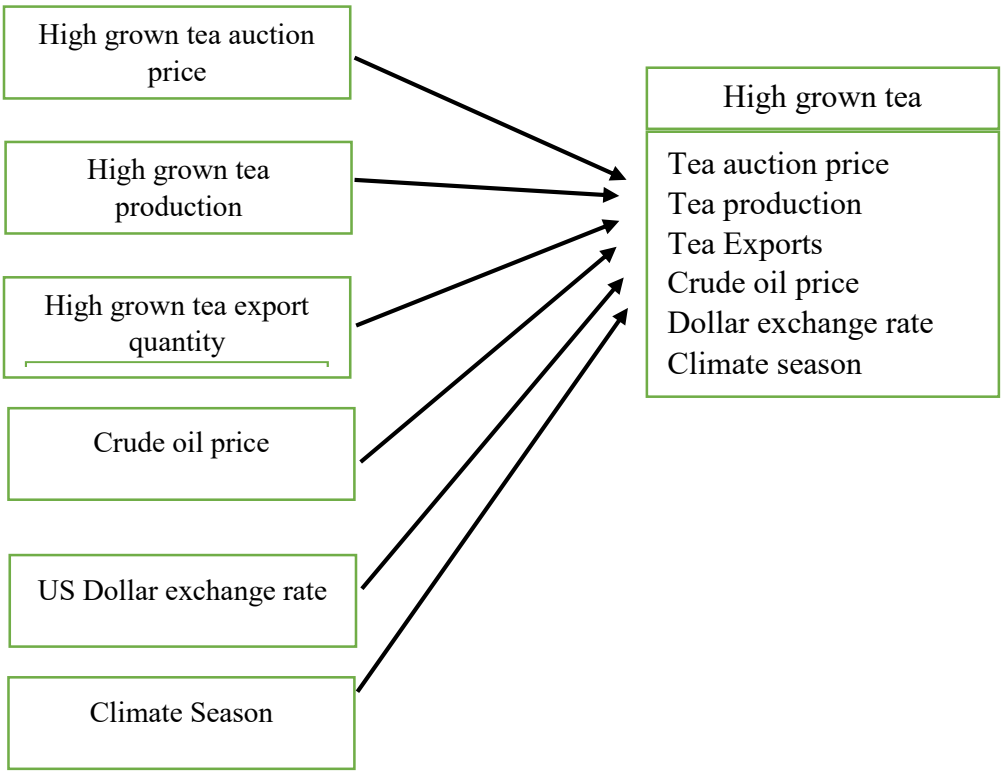
auction price, it is required to bring all the factor's data to be consistent with the tea auction price dataset, because the tea auction price acting as the class dataset in the data mining process. Due to that reason the average price of crude oil and exchange rate, which is an aggregate value was calculated with the use of excel and SQL queries for the each week. The result in a dataset after aggregating, with a single value for each week for the price of crude oil and the US dollar exchange rate which can be easily manipulated with all the other factors.

4.4.2.2 Merging data

When considering the datasets used in the tea auction price analysis, the data for different factors were in different data files in different data sources and organizations like crude oil price in Federal Reserve Bank St. Louis web site and the US dollar exchange rate from Central Bank of Sri Lanka. To perform the data mining process, there was a need to create a CSV format file that contains all the factors of data to include in the Weka and Rapid miner software which is used for mining purposes. Since all the factors had values against the time in their files, all factors brought into one single file with values listed under an attribute with the factor name against a single time column. The selected and cleaned data were integrated with the use of queries. Three CSV files were created for the different categories of tea. Those are

- High grown tea data
- Mid Grown tea data
- Low grown tea data

All the above tea category datasets consist of tea auction price for one category and the related tea production amount, tea export quantity and tea auction quantity for that category. Crude oil price, US dollar exchange rate and climate, seasons are the common values and those data are in all the datasets.



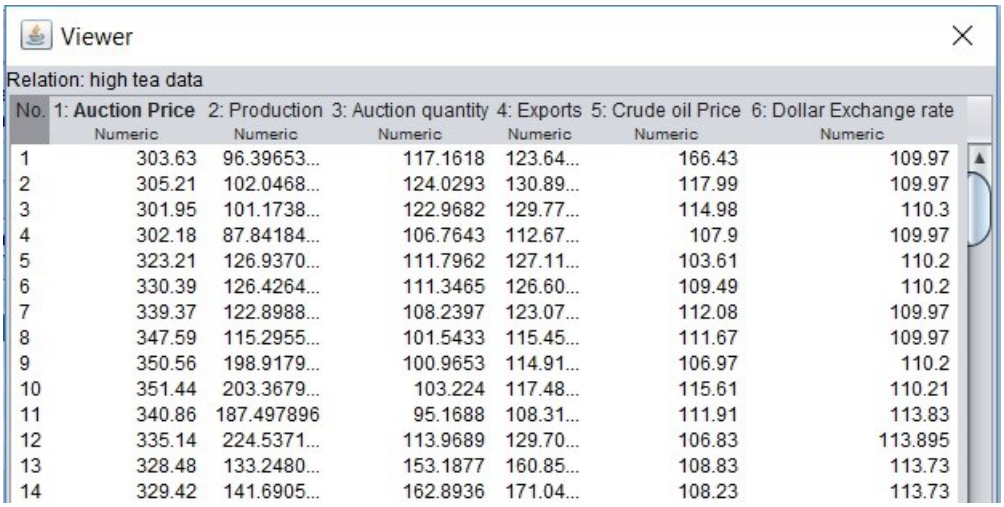
4.3 Merging data

4.4.3 Data Transformation

The phase of data transformation is used to consolidate data into forms suitable for data mining. This consists of syntactic modifications made to the data without changing its meaning. This is done mainly as it is required by most data mining tools and the data mining goal. The date field of the US dollar exchange rate and the crude oil price were modified and decomposed into separate fields. This was performed in order to apply consistency between all the factors of the dataset to incorporate merging of datasets based on a weekly basis and construct the new attribute climate season based on the month.

4.4.3.1 Data Normalization

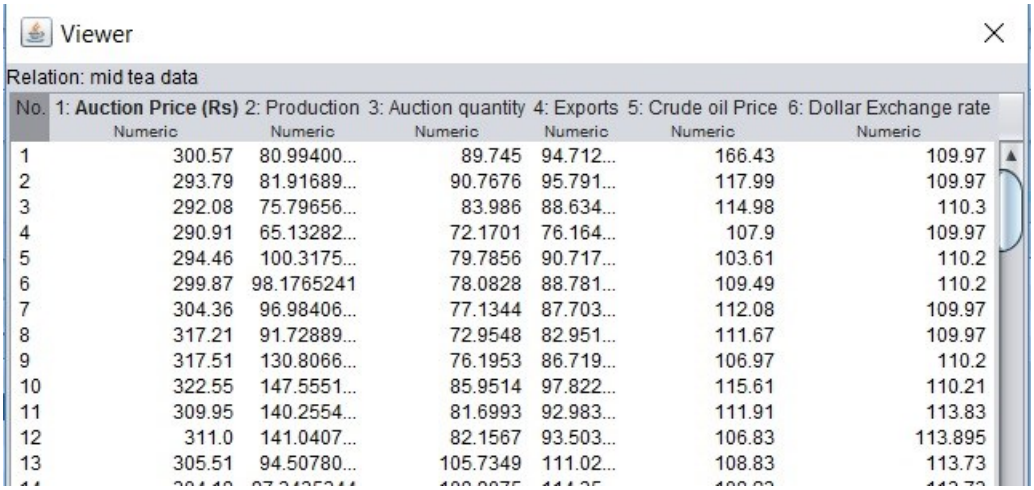
To acquire the accurate prediction model using data mining algorithms, all the data values of the dataset need to normalize into a common data range. So the tea auction price, US Dollar exchange rate and crude oil price keep as it is, because the tea auction price data range is in between 100 to 1000 range and also tea price prediction result need to get as the real value. Therefore quantity data values of production, auction and export data are normalized to 1000 range. After converting those values in three data sets, added to the WEKA explorer to build a model. Following listed how the WEKA viewer shows datasets.


Viewer

Relation: high tea data

No.	1: Auction Price	2: Production	3: Auction quantity	4: Exports	5: Crude oil Price	6: Dollar Exchange rate
	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric
1	303.63	96.39653...	117.1618	123.64...	166.43	109.97
2	305.21	102.0468...	124.0293	130.89...	117.99	109.97
3	301.95	101.1738...	122.9682	129.77...	114.98	110.3
4	302.18	87.84184...	106.7643	112.67...	107.9	109.97
5	323.21	126.9370...	111.7962	127.11...	103.61	110.2
6	330.39	126.4264...	111.3465	126.60...	109.49	110.2
7	339.37	122.8988...	108.2397	123.07...	112.08	109.97
8	347.59	115.2955...	101.5433	115.45...	111.67	109.97
9	350.56	198.9179...	100.9653	114.91...	106.97	110.2
10	351.44	203.3679...	103.224	117.48...	115.61	110.21
11	340.86	187.497896	95.1688	108.31...	111.91	113.83
12	335.14	224.5371...	113.9689	129.70...	106.83	113.895
13	328.48	133.2480...	153.1877	160.85...	108.83	113.73
14	329.42	141.6905...	162.8936	171.04...	108.23	113.73

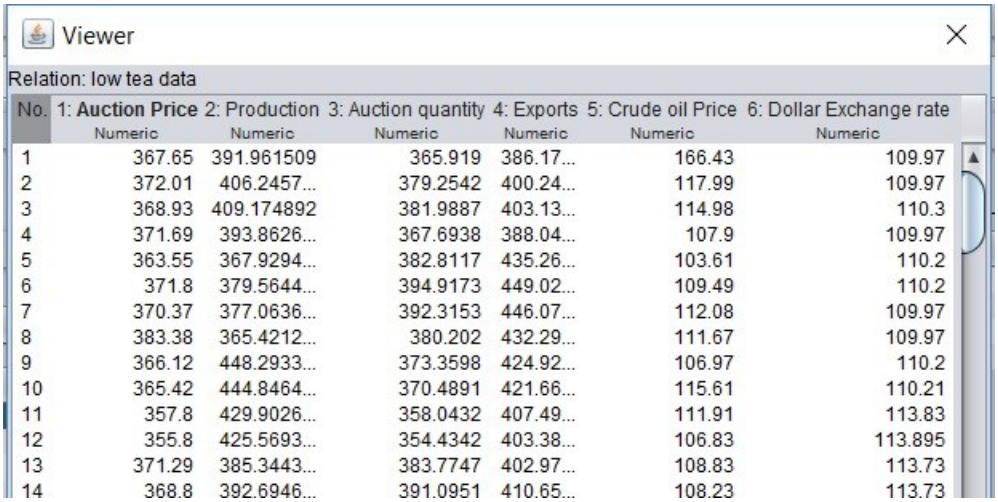
4.4 High grown tea data


Viewer

Relation: mid tea data

No.	1: Auction Price (Rs)	2: Production	3: Auction quantity	4: Exports	5: Crude oil Price	6: Dollar Exchange rate
	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric
1	300.57	80.99400...	89.745	94.712...	166.43	109.97
2	293.79	81.91689...	90.7676	95.791...	117.99	109.97
3	292.08	75.79656...	83.986	88.634...	114.98	110.3
4	290.91	65.13282...	72.1701	76.164...	107.9	109.97
5	294.46	100.3175...	79.7856	90.717...	103.61	110.2
6	299.87	98.1765241	78.0828	88.781...	109.49	110.2
7	304.36	96.98406...	77.1344	87.703...	112.08	109.97
8	317.21	91.72889...	72.9548	82.951...	111.67	109.97
9	317.51	130.8066...	76.1953	86.719...	106.97	110.2
10	322.55	147.5551...	85.9514	97.822...	115.61	110.21
11	309.95	140.2554...	81.6993	92.983...	111.91	113.83
12	311.0	141.0407...	82.1567	93.503...	106.83	113.895
13	305.51	94.50780...	105.7349	111.02...	108.83	113.73

4.5 Mid grown tea data

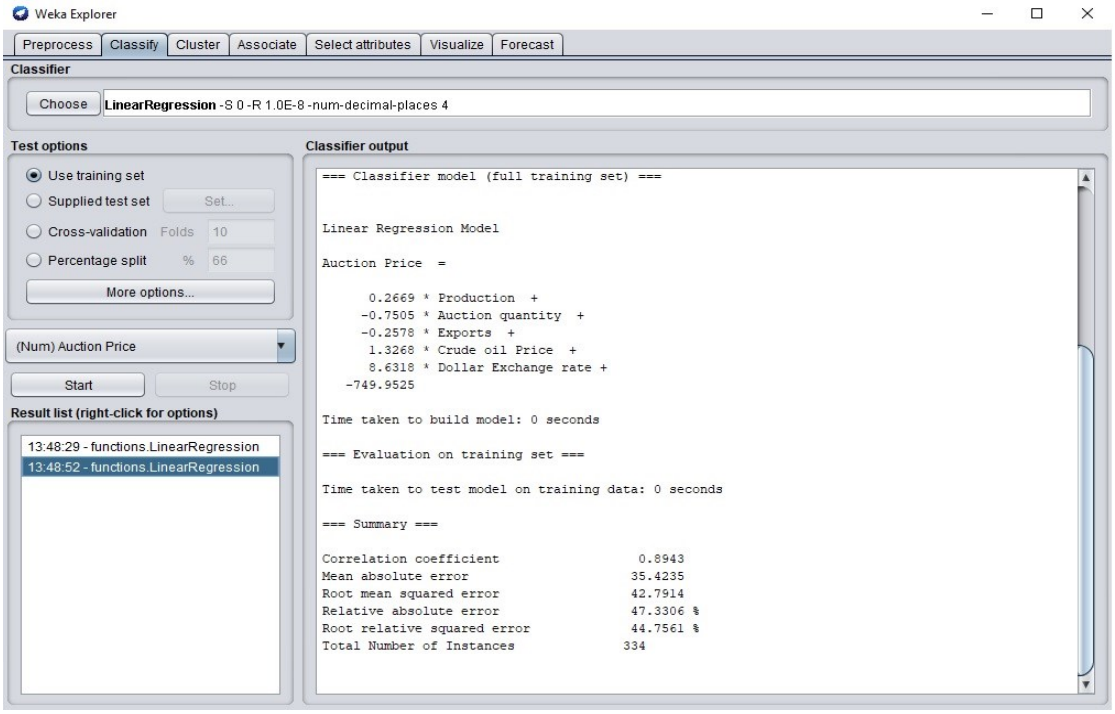


No.	1: Auction Price	2: Production	3: Auction quantity	4: Exports	5: Crude oil Price	6: Dollar Exchange rate
	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric
1	367.65	391.961509	365.919	386.17...	166.43	109.97
2	372.01	406.2457...	379.2542	400.24...	117.99	109.97
3	368.93	409.174892	381.9887	403.13...	114.98	110.3
4	371.69	393.8626...	367.6938	388.04...	107.9	109.97
5	363.55	367.9294...	382.8117	435.26...	103.61	110.2
6	371.8	379.5644...	394.9173	449.02...	109.49	110.2
7	370.37	377.0636...	392.3153	446.07...	112.08	109.97
8	383.38	365.4212...	380.202	432.29...	111.67	109.97
9	366.12	448.2933...	373.3598	424.92...	106.97	110.2
10	365.42	444.8464...	370.4891	421.66...	115.61	110.21
11	357.8	429.9026...	358.0432	407.49...	111.91	113.83
12	355.8	425.5693...	354.4342	403.38...	106.83	113.895
13	371.29	385.3443...	383.7747	402.97...	108.83	113.73
14	368.8	392.6946...	391.0951	410.65...	108.23	113.73

4.6 Low grown tea data

The three data sets were prepared in excel worksheet and entered to the WEKA explorer as a CSV (Comma Separate Value) format. Tea auctioned price is used as the class variable because the research is carried for predicting the tea price. High grown tea data set used to check the best algorithm and enter those data to build the suitable prediction model.

After training the entered high grown tea data set, we observed, there are high error rates in MAE and RMSE. Because still the values of the variables are in different stages and not in the same level.



4.7 Linear Regression algorithm

Therefore then we decide to fully normalize the dataset. Transforming the dataset by applying for all the variable and get all the values in between 0 and 1.

4.4.3.2 Min – Max Normalization

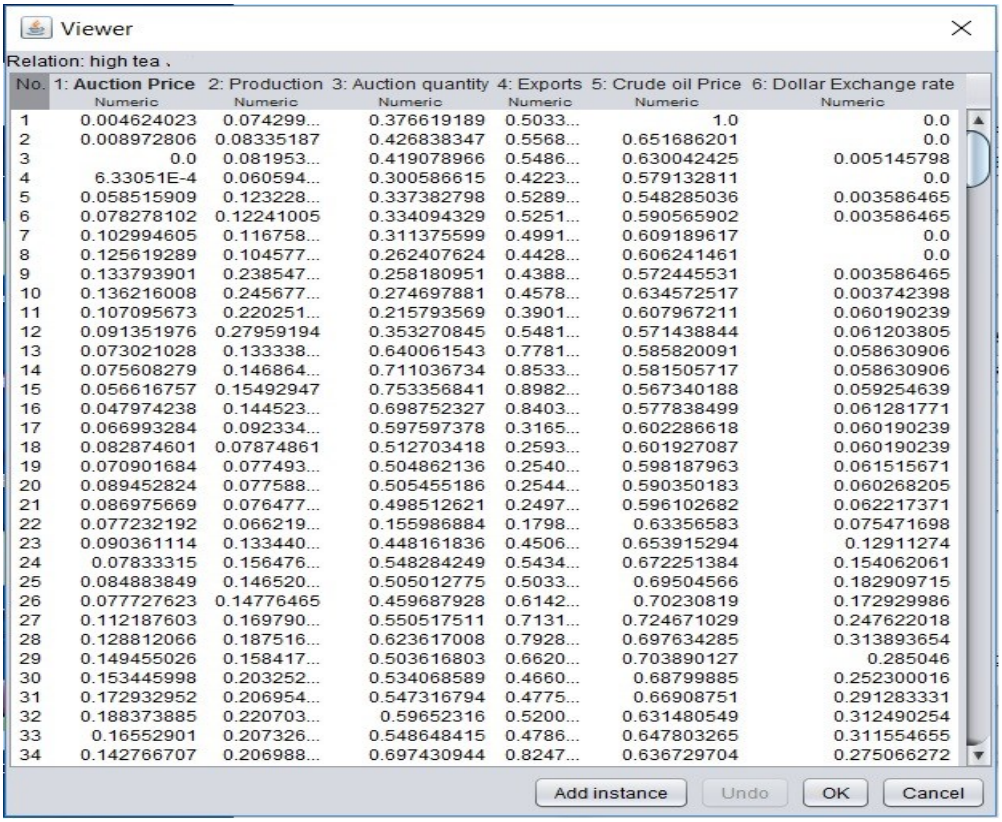
Min-Max normalization is a normalization strategy which linearly transforms x to y= (x-min)/(max-min), where min and max are the minimum and maximum values in X, where X is the set of observed values of x.

$$z = \frac{x - \min(x)}{\max(x) - \min(x)}$$

4.8 Min-Max Normalization Equation

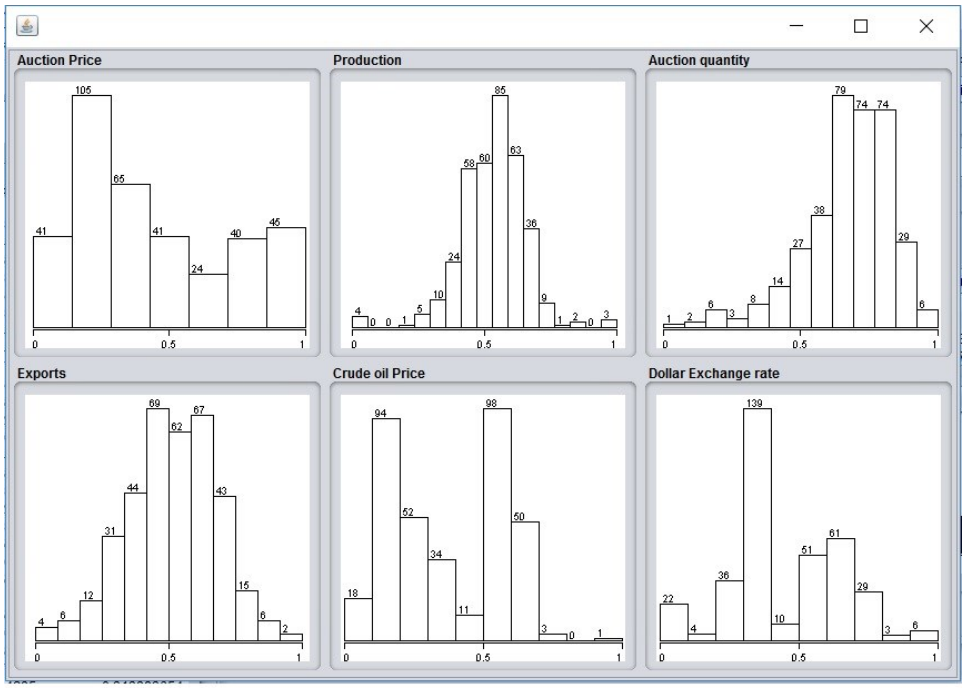
It can be easily seen that when x=min, then y=0, and When x=max, then y=1. This means, the minimum value of X is mapped to 0 and the maximum value in X is mapped to 1. So, the entire range of values of X from min to max is mapped to the range 0 to 1.

After transforming the dataset to min-max normalization, all the variable values comes to the range of 0 and 1.



4.9 High grown dataset after min max normalization

As seen in the above picture the maximum value of the dataset is 1 and the minimum value would be 0 and other values are varying between the 0 and 1. This will bring all the values into the common stage where can apply regression techniques to get the accurate result.

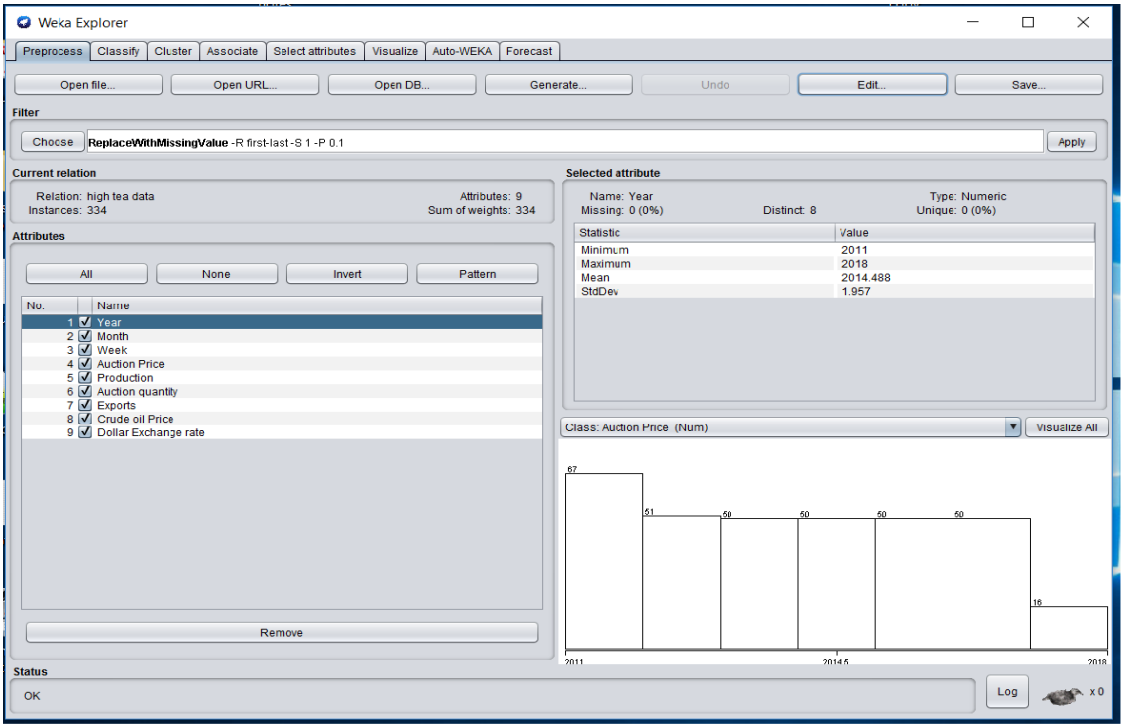


4.10 Data visualization after normalization

The above graph shows how the data is visualized after the min-max normalization are applied.

4.5 Handling missing values

There are various kinds of filters available in pre-process stage in Weka. In this dataset applied Replace Missing Values filter to fill two or three missing values which was in production quantity.

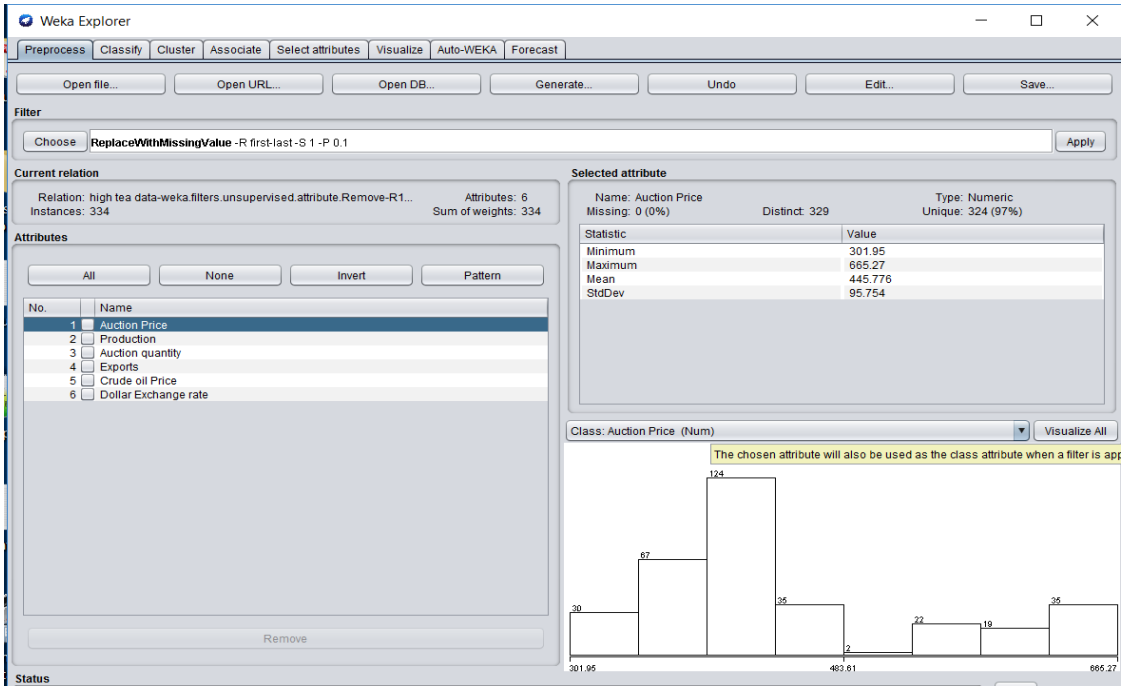


4.11 Applying missing values filter

4.6 Data Reduction

Complex data analysis and mining on huge amounts of data may take a very long time, making such analysis impractical or infeasible. Data reduction techniques have been helpful in analysing the reduced representation of the dataset without compromising the integrity of the original data and yet producing the quality knowledge. The concept of data reduction is commonly understood as either reducing the volume or reducing the dimensions (number of attributes). There are a number of methods that have facilitated in analysing a reduced volume or dimension of data and yet yield useful knowledge. Certain partition based methods work on partition of data tuples.

That is, mining of the reduced data set should be more efficient yet produce the same (or almost the same) analytical results.



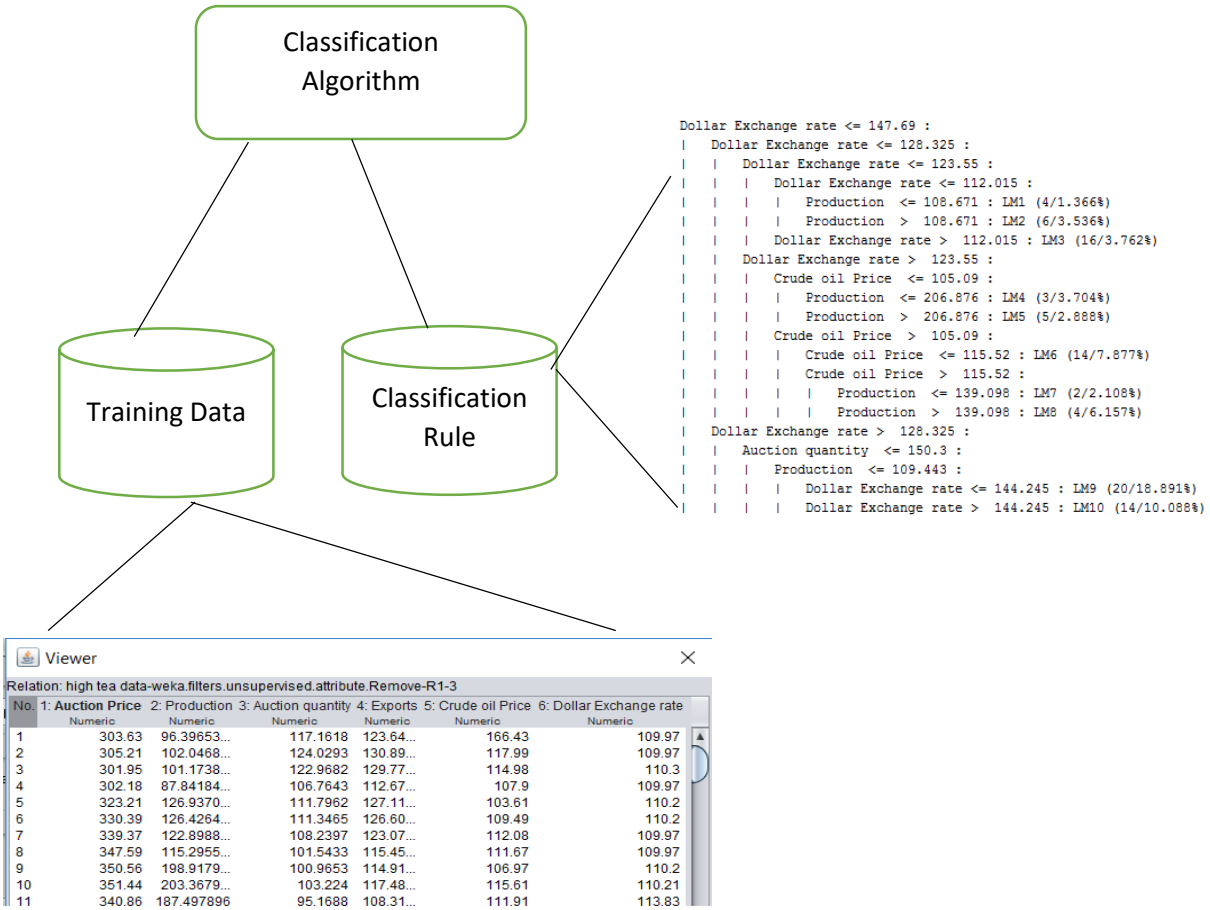
4.12 Removing date attributes

Before sending to the classification date attributes in every dataset are removed because the date fields do not need to classify and build the model.

4.7 Classification

Classification is the process of finding a set of models that describe and distinguish data classes and concepts, for the purpose of being able to use the model to predict the class whose label is unknown. Classification is a two-step process, first, it build classification models using training data. Every object of the dataset must be pre classified i.e. Its class label must be known; second the model generated in the preceding step is tested by assigning class labels to data objects in a test data set. Each tuple/sample is assumed to belong to a predefined class, as determined by the class label attribute. The model is represented as classification rules, decision trees, or mathematical formulae. The second step is model usage. It is for classifying future or unknown objects. It estimates accuracy of the model. The known label of the test sample is compared with the classified result from the model. Model construction describes a set of predetermines classes. The accuracy rate is the percentage of test set samples that are correctly classified by the model. The test

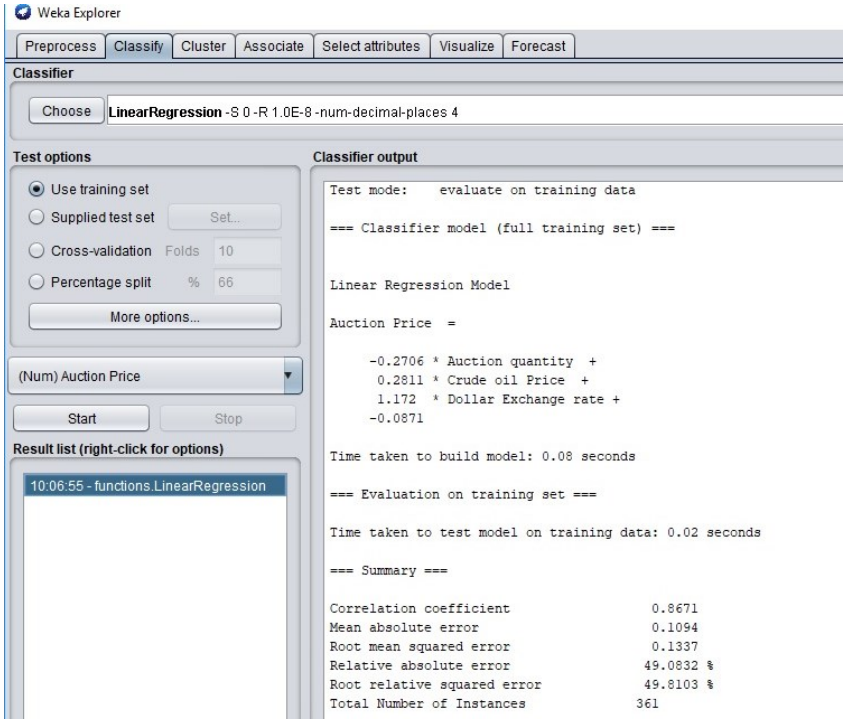
set is independent of training set, otherwise over fitting will occur. If the accuracy is acceptable, use the model to classify data tuples whose class labels are not known.



4.13 Classification Model

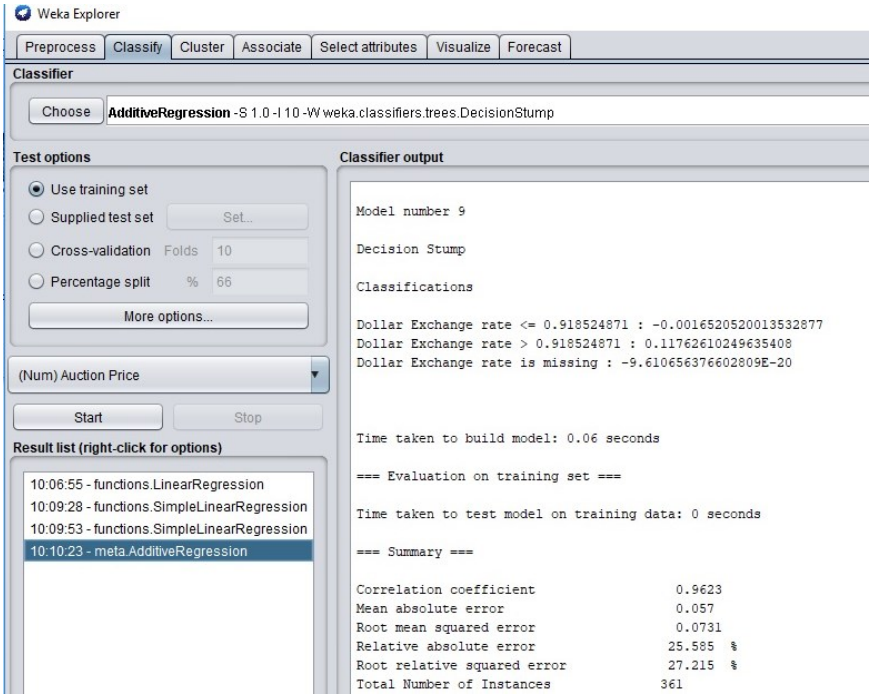
4.8 Selecting an algorithm

Classification algorithms are normally used to get the predictions from the dataset. Selecting a suitable classification algorithm for building a model that can give the predictions for the tea price forecast is one of an objective of this research. Regression algorithms analyse the class data with the variables and build a model of which variables are highly correlated with the class data to give the accurate results. Therefore, various kinds of regression related algorithms are applied for the dataset to select the best algorithm which gives the highest percentage of accurate results.

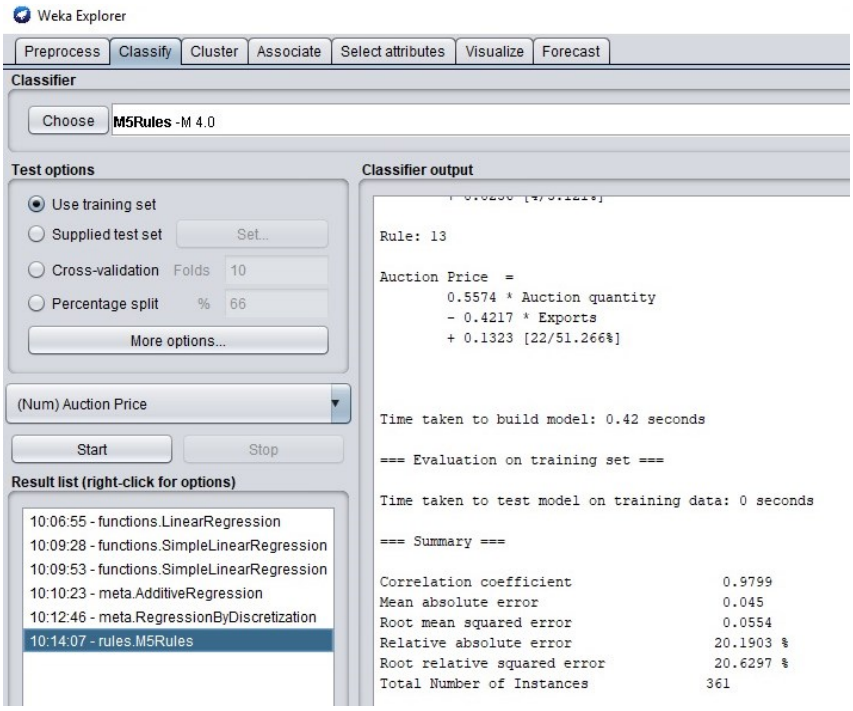


4.14 Linear regression applied for normalize data set

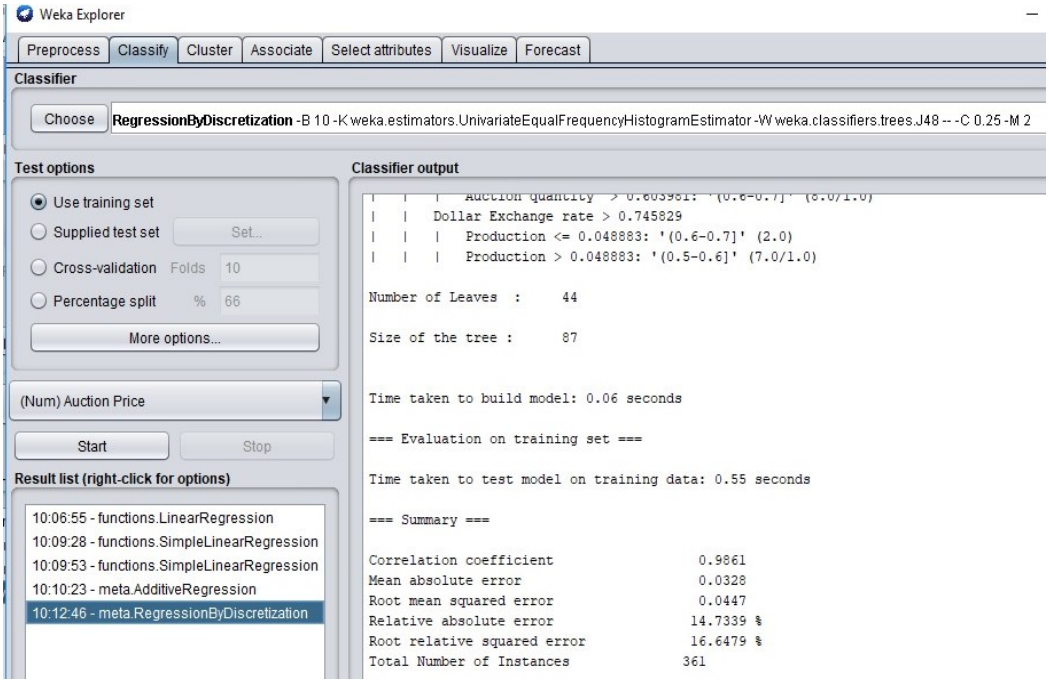
First applied the Linear Regression algorithm for training the high grown tea dataset.



4.15 Additive Regression error rate



4.16 M5 Rules error rate



4.17 Regression by Discretization error rate

Following are the results obtained by applying the preprocessed high grown data set as a training data to the various kinds of classification algorithm which related to the predictions.

Algorithm	**Mean Absolute Error**	**Root Mean Squared Error**	**Relative Absolute Error**	**Root Relative Squared Error**
Linear Regression	0.0859	0.1104	73.1802%	63.6309%
Simple Linear Regression	0.1034	0.129	88.0574%	74.4%
D14j Mlp Classifier	0.2178	0.2455	185.521%	141.5503%
SMOreg	0.0793	0.1175	67.5522%	67.7269%
LWL	0.0732	0.0921	62.3851%	53.1164%
Additive Regression	0.0512	0.0661	43.5893%	38.0896%
Random Sub Space	0.042	0.0543	36.1533%	31.2974%
Regression by Discretization	0.0296	0.0392	25.2292%	22.6001%
Stacking	0.1174	0.1734	100%	100%
Decision Table	0.0618	0.0874	52.6221%	50.4116%
M5Rules	0.0441	0.0596	37.5256%	34.3618%
Hot Winters	0.5203	0.5831	233.5032%	217.1671%
M5P	0.0463	0.0603	39.475%	34.4639%
REP Tree	0.0353	0.0514	30.8591%	29.154%

4.18 Algorithm error rate table

As per the above graph shown, from the various classification algorithms, the Regression by Discretization algorithm shows the lowest percentage of Relative Absolute Error and the Root Relative Squared Error. Both errors are distinguishing the error rate of correlation between the class data and the affected variable data.

Therefore, select the Regression by Discretization algorithm as the best algorithm with lower error rate percentages among the above classification algorithms, and used for building the prediction model for three separate data sets based on the elevation.

4.9 Summary

In this chapter briefly discuss about what kind of data mining techniques are applied to the data set, what are the methodologies were used to analyse the dataset and how to select the best algorithm by experimenting the normalize dataset. Finally Regression by discretization algorithm select as the best suitable algorithm with less error to apply and build the prediction model to forecast the tea auction price. Next chapter will describe how to build the model using WEKA.

Chapter 5

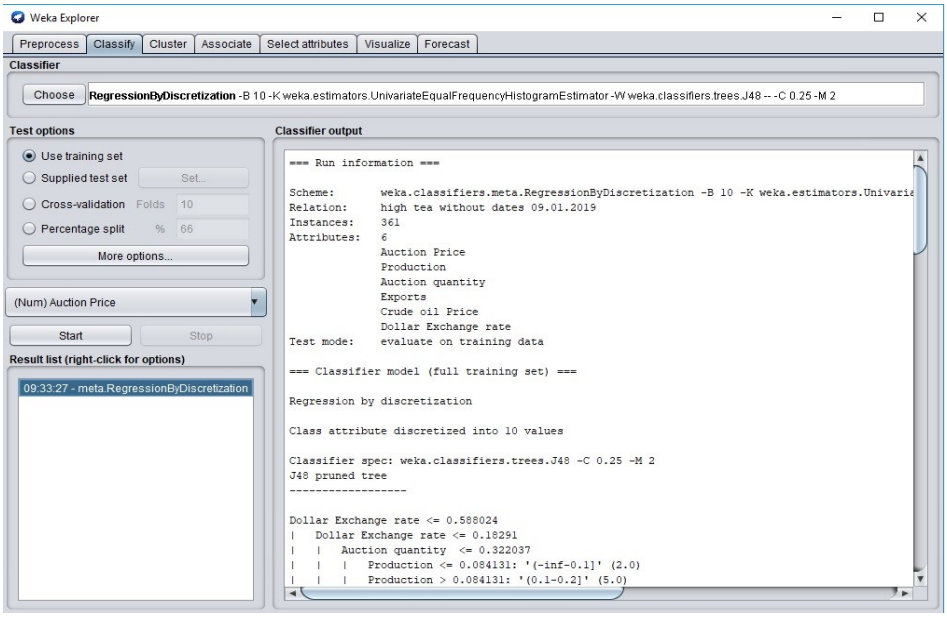
Building a Model for Prediction

5.1 Introduction

This chapter describes how the predictive model was built by training the dataset using WEKA analyser.

5.2 Model Build

After selecting the Regression by Discretization as the classification algorithm, then build the model using that. Preprocessed high grown tea data set was entered to the WEKA explorer, then selecting the tea auction price as the class variable and select the above prediction algorithm and start to build the model. 2/3 of the dataset were used for training the data to build models. The WEKA, build the model according to the data set by analysing the correlation of the variables with the class data and build a prediction model which can forecast the tea auction price.



5.1 Regression by Discretization model

The following display the model information after builds by WEKA.

=== Run information ===

Scheme: weka.classifiers.meta.RegressionByDiscretization –B 10 –K weka.estimators.UnivariateEqualFrequencyHistogramEstimator –W weka.classifiers.trees.J48 -- -C 0.25 –M 2

Relation: high tea data training set
Instances: 275
Attributes: 6
Auction Price
Production
Auction quantity
Exports
Crude oil Price
Dollar Exchange rate
Test mode: evaluate on training data
=== Classifier model (full training set) ===

Regression by discretization
Class attribute discretized into 10 values
Classifier spec: weka.classifiers.trees.J48 –C 0.25 –M 2
J48 pruned tree

Dollar Exchange rate <= 0.18291
| Auction quantity <= 0.322037
| | Production <= 0.084131: '(-inf-0.093658]' (2.0)
| | Production > 0.084131: '(0.093658-0.187317]' (5.0)
| Auction quantity > 0.322037: '(-inf-0.093658]' (19.0)
Dollar Exchange rate > 0.18291
| Dollar Exchange rate <= 0.568065
| | Auction quantity <= 0.658883
| | | Crude oil Price <= 0.65521
| | | | Production <= 0.080729
| | | | | Dollar Exchange rate <= 0.487759
| | | | | | Production <= 0.028172: '(0.187317-0.280975]' (6.0)
| | | | | | Production > 0.028172: '(0.093658-0.187317]' (6.0/1.0)
| | | | | Dollar Exchange rate > 0.487759: '(0.374634-0.468292]' (8.0)
| | | | Production > 0.080729
| | | | | Crude oil Price <= 0.551233
| | | | | | Auction quantity <= 0.614268
| | | | | | | Dollar Exchange rate <= 0.514736
| | | | | | | | Exports <= 0.573701

| | | | | | | | | Crude oil Price <= 0.146761: '(0.187317-0.280975]' (11.0/1.0)
| | | | | | | | | Crude oil Price > 0.146761
| | | | | | | | | | Dollar Exchange rate <= 0.317324: '(0.187317-0.280975]' (8.0/1.0)
| | | | | | | | | | Dollar Exchange rate > 0.317324
| | | | | | | | | | | Crude oil Price <= 0.377795
| | | | | | | | | | | | Dollar Exchange rate <= 0.35927: '(0.280975-0.374634]' (5.0)
| | | | | | | | | | | | Dollar Exchange rate > 0.35927: '(0.187317-0.280975]' (9.0/3.0)
| | | | | | | | | | | Crude oil Price > 0.377795: '(0.374634-0.468292]' (4.0/1.0)
| | | | | | | | | Exports > 0.573701: '(0.280975-0.374634]' (15.0/2.0)
| | | | | | | | Dollar Exchange rate > 0.514736: '(0.280975-0.374634]' (27.0/4.0)
| | | | | | Auction quantity > 0.614268
| | | | | | | Production <= 0.173122: '(0.187317-0.280975]' (2.0)
| | | | | | | Production > 0.173122: '(0.093658-0.187317]' (3.0/1.0)
| | | | | Crude oil Price > 0.551233
| | | | | | Auction quantity <= 0.142852: '(0.187317-0.280975]' (2.0)
| | | | | | Auction quantity > 0.142852
| | | | | | | Production <= 0.217593
| | | | | | | | Exports <= 0.785721: '(0.280975-0.374634]' (59.0/17.0)
| | | | | | | | Exports > 0.785721: '(0.374634-0.468292]' (4.0)
| | | | | | | Production > 0.217593
| | | | | | | | Auction quantity <= 0.509265: '(0.374634-0.468292]' (2.0)
| | | | | | | | Auction quantity > 0.509265
| | | | | | | | | Exports <= 0.582455: '(0.187317-0.280975]' (2.0)
| | | | | | | | | Exports > 0.582455: '(0.280975-0.374634]' (4.0/1.0)
| | | Crude oil Price > 0.65521: '(0.093658-0.187317]' (5.0)
| | Auction quantity > 0.658883
| | | Crude oil Price <= 0.580571
| | | | Crude oil Price <= 0.236931: '(0.187317-0.280975]' (4.0)
| | | | Crude oil Price > 0.236931
| | | | | Crude oil Price <= 0.55979: '(0.093658-0.187317]' (23.0/1.0)
| | | | | Crude oil Price > 0.55979: '(0.187317-0.280975]' (3.0/1.0)
| | | Crude oil Price > 0.580571
| | | | Auction quantity <= 0.836503
| | | | | Production <= 0.225278
| | | | | | Production <= 0.199451: '(0.280975-0.374634]' (2.0)
| | | | | | Production > 0.199451: '(0.093658-0.187317]' (4.0/1.0)
| | | | | Production > 0.225278: '(0.280975-0.374634]' (4.0)
| | | | Auction quantity > 0.836503: '(0.187317-0.280975]' (3.0/1.0)
| Dollar Exchange rate > 0.568065
| | Dollar Exchange rate <= 0.588024
| | | Production <= 0.094097
| | | | Crude oil Price <= 0.158841: '(0.468292-0.561951]' (2.0)

| | | | Crude oil Price > 0.158841: '(0.561951-0.655609]' (2.0)
| | | Production > 0.094097: '(0.280975-0.374634]' (4.0/1.0)
| | Dollar Exchange rate > 0.588024
| | | Dollar Exchange rate <= 0.620303: '(0.655609-0.749268]' (8.0/1.0)
| | | Dollar Exchange rate > 0.620303
| | | | Production <= 0.078749: '(0.749268-0.842926]' (2.0)
| | | | Production > 0.078749: '(0.842926-inf)' (6.0/2.0)

Number of Leaves : 35
Size of the tree : 69
Time taken to build model: 0.08 seconds

=== Evaluation on training set ===

Time taken to test model on training data: 0.39 seconds

=== Summary ===

Correlation coefficient	0.9742
Mean absolute error	0.0277
Root mean squared error	0.0367
Relative absolute error	25.2292 %
Root relative squared error	22.6001 %
Total Number of Instances	275

For the 249 number of data instances, WEKA generated the above prediction model which has 25.2% of relative absolute error and 22.6% of root relative squared error rate. Also the correlation coefficient 0.9742 which means the highest percentage of positive relationship having between the class and the variable.

The correlation coefficient measures the degree of relationship or correlation between the variables. This is normally used in regression analysis and calculate the linear relationship between the variables. This cannot calculate non-linear relationships.

$$r = \frac{n(\sum xy) - (\sum x)(\sum y)}{\sqrt{[\, n\sum x^2 - (\sum x)^2 \,][\, n\sum y^2 - (\sum y)^2 \,]}}$$

5.2 Correlation Coefficient Formula

The result of correlation coefficient comes in the range of +1 to -1. The result of positive one means the class variable has an extremely good relationship with the variables. If the variables are increasing then the class variable also change positively according to that. If the result is negative one, the class data have the opposite reaction to the variables. And the zero value means there are no relationship between the variables.

In here the build model displays 0.9742 correlation coefficient value, which has the highest positive relationship with the variables. If the variables of production, export quantity, auction quantity, crude oil price and the US dollar exchange rate values are changing positively or negatively, the tea auction price also increase or decrease according to the change.

5.3 Summary

This chapter annotate about how the model built using a classification algorithm and the next step is to test and evaluate the accuracy of the model using test data set.

Chapter 6

Test and Evaluation

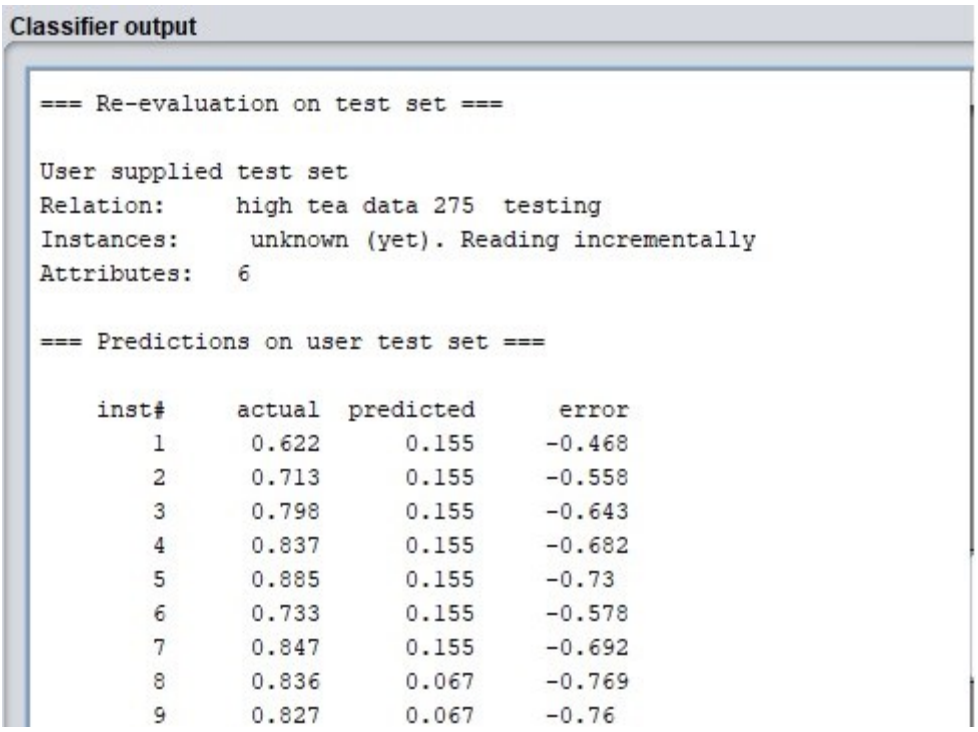
6.1 Introduction

In this chapter describe how the built forecasting model was tested and evaluated using remaining dataset. As mentioned in the earlier chapters, the data set was divided into 2 parts and one was used as a training data for building the forecasting model and the remaining data set is used as test data to check the accuracy of the model predictions.

6.2 Test the Model

2/3 quota of the full data set has been used for the training purpose to build a model using a suitable classification algorithm and the remaining 1/3 data set was kept for test the output. To test the model, prepared preprocessed remaining dataset was inserted to WEKA as test data set and it gave the tea price predictions for the steps ahead. The three built models were tested separately by entering three datasets according to the elevations.

The high grown tea test data set was inserted as the test data to the WEKA after load the prepared model to get the predictions. Output predictions set as the text type. After settings has been made, reevaluate the data set using the model.



6.1 Classifier output

==========

=== Model information ===

Filename: 275 high 29.01.2019.model
Scheme: weka.classifiers.meta.RegressionByDiscretization -B 10 -K weka.estimators.UnivariateEqualFrequencyHistogramEstimator -W weka.classifiers.trees.J48 -- -C 0.25 -M 2
Relation: high tea data 275 training
Attributes: 6
Auction price
Production
Auction Quantity
Export Quantity
Crude oil Price
Dollar Exchange Rate

=== Classifier model ===

Regression by discretization
Class attribute discretized into 10 values
Classifier spec: weka.classifiers.trees.J48 -C 0.25 -M 2
J48 pruned tree

```
------------------

Dollar Exchange Rate <= 0.28561
|  Auction Quantity <= 0.32932
|  |  Production <= 0.084131: '(-inf-0.1]' (2.0)
|  |  Production > 0.084131: '(0.1-0.2]' (5.0)
|  Auction Quantity > 0.32932: '(-inf-0.1]' (19.0)
Dollar Exchange Rate > 0.28561
|  Dollar Exchange Rate <= 0.887022
|  |  Auction Quantity <= 0.673784
|  |  |  Crude oil Price <= 0.65521
|  |  |  |  Production <= 0.080729
|  |  |  |  |  Dollar Exchange Rate <= 0.761626
|  |  |  |  |  |  Production <= 0.028172: '(0.2-0.3]' (6.0)
|  |  |  |  |  |  Production > 0.028172: '(0.1-0.2]' (6.0/1.0)
|  |  |  |  |  Dollar Exchange Rate > 0.761626: '(0.4-0.5]' (8.0)
|  |  |  |  Production > 0.080729
|  |  |  |  |  Crude oil Price <= 0.551233
|  |  |  |  |  |  Auction Quantity <= 0.628159
|  |  |  |  |  |  |  Dollar Exchange Rate <= 0.80375
|  |  |  |  |  |  |  |  Export Quantity <= 0.573701
|  |  |  |  |  |  |  |  |  Crude oil Price <= 0.146761: '(0.2-0.3]' (11.0/1.0)
|  |  |  |  |  |  |  |  |  Crude oil Price > 0.146761
|  |  |  |  |  |  |  |  |  |  Dollar Exchange Rate <= 0.495495: '(0.2-0.3]' (8.0/1.0)
|  |  |  |  |  |  |  |  |  |  Dollar Exchange Rate > 0.495495
|  |  |  |  |  |  |  |  |  |  |  Crude oil Price <= 0.377795
|  |  |  |  |  |  |  |  |  |  |  |  Dollar Exchange Rate <= 0.560993: '(0.3-0.4]' (5.0)
|  |  |  |  |  |  |  |  |  |  |  |  Dollar Exchange Rate > 0.560993: '(0.2-0.3]' (9.0/3.0)
|  |  |  |  |  |  |  |  |  |  |  Crude oil Price > 0.377795: '(0.4-0.5]' (4.0/1.0)
|  |  |  |  |  |  |  |  Export Quantity > 0.573701: '(0.3-0.4]' (15.0/2.0)
|  |  |  |  |  |  |  Dollar Exchange Rate > 0.80375: '(0.3-0.4]' (27.0/4.0)
|  |  |  |  |  |  Auction Quantity > 0.628159
|  |  |  |  |  |  |  Production <= 0.173122: '(0.2-0.3]' (2.0)
|  |  |  |  |  |  |  Production > 0.173122: '(0.1-0.2]' (3.0/1.0)
|  |  |  |  |  Crude oil Price > 0.551233
|  |  |  |  |  |  Auction Quantity <= 0.146083: '(0.2-0.3]' (2.0)
|  |  |  |  |  |  Auction Quantity > 0.146083
|  |  |  |  |  |  |  Production <= 0.217593
|  |  |  |  |  |  |  |  Export Quantity <= 0.785721: '(0.3-0.4]' (59.0/17.0)
|  |  |  |  |  |  |  |  Export Quantity > 0.785721: '(0.4-0.5]' (4.0)
|  |  |  |  |  |  |  Production > 0.217593
|  |  |  |  |  |  |  |  Auction Quantity <= 0.520782: '(0.4-0.5]' (2.0)
|  |  |  |  |  |  |  |  Auction Quantity > 0.520782
|  |  |  |  |  |  |  |  |  Export Quantity <= 0.582455: '(0.2-0.3]' (2.0)
|  |  |  |  |  |  |  |  |  Export Quantity > 0.582455: '(0.3-0.4]' (4.0/1.0)
|  |  |  Crude oil Price > 0.65521: '(0.1-0.2]' (5.0)
|  |  Auction Quantity > 0.673784
```

```
|   |   |   Crude oil Price <= 0.580571
|   |   |   |   Crude oil Price <= 0.236931: '(0.2-0.3]' (4.0)
|   |   |   |   Crude oil Price > 0.236931
|   |   |   |   |   Crude oil Price <= 0.55979: '(0.1-0.2]' (23.0/1.0)
|   |   |   |   |   Crude oil Price > 0.55979: '(0.2-0.3]' (3.0/1.0)
|   |   |   Crude oil Price > 0.580571
|   |   |   |   Auction Quantity <= 0.85542
|   |   |   |   |   Production <= 0.225278
|   |   |   |   |   |   Production <= 0.199451: '(0.3-0.4]' (2.0)
|   |   |   |   |   |   Production > 0.199451: '(0.1-0.2]' (4.0/1.0)
|   |   |   |   |   Production > 0.225278: '(0.3-0.4]' (4.0)
|   |   |   |   Auction Quantity > 0.85542: '(0.2-0.3]' (3.0/1.0)
|   Dollar Exchange Rate > 0.887022
|   |   Dollar Exchange Rate <= 0.918188
|   |   |   Production <= 0.094097
|   |   |   |   Crude oil Price <= 0.158841: '(0.5-0.6]' (2.0)
|   |   |   |   Crude oil Price > 0.158841: '(0.6-0.7]' (2.0)
|   |   |   Production > 0.094097: '(0.3-0.4]' (4.0/1.0)
|   |   Dollar Exchange Rate > 0.918188
|   |   |   Dollar Exchange Rate <= 0.96859: '(0.7-0.8]' (8.0/1.0)
|   |   |   Dollar Exchange Rate > 0.96859
|   |   |   |   Production <= 0.078749: '(0.8-0.9]' (2.0)
|   |   |   |   Production > 0.078749: '(0.9-inf)' (6.0/2.0)


Number of Leaves  :     35
Size of the tree :      69


=== Re-evaluation on test set ===

User supplied test set
Relation:     high tea data 275  testing
Instances:     unknown (yet). Reading incrementally
Attributes:   6

=== Predictions on user test set ===

   inst#     actual  predicted      error
      1      0.622      0.155     -0.468
      2      0.713      0.155     -0.558
      3      0.798      0.155     -0.643
      4      0.837      0.155     -0.682
      5      0.885      0.155     -0.73
      6      0.733      0.155     -0.578
      7      0.847      0.155     -0.692
      8      0.836      0.067     -0.769
      9      0.827      0.067     -0.76
     10      0.721      0.067     -0.654
```

11	0.46	0.067	-0.393
12	0.387	0.067	-0.32
13	0.292	0.067	-0.225
14	0.249	0.067	-0.182
15	0.225	0.067	-0.158
16	0.269	0.067	-0.202
17	0.286	0.067	-0.219
18	0.254	0.067	-0.187
19	0.318	0.155	-0.163
20	0.381	0.067	-0.314
21	0.455	0.067	-0.388
22	0.504	0.155	-0.349
23	0.514	0.155	-0.359
24	0.703	0.067	-0.635
25	0.587	0.067	-0.52
26	0.624	0.155	-0.469
27	0.773	0.155	-0.618
28	0.742	0.155	-0.587
29	0.787	0.155	-0.632
30	0.891	0.155	-0.736
31	0.948	0.155	-0.793
32	0.999	0.155	-0.845
33	1.00	0.155	-0.845
34	0.997	0.155	-0.842
35	0.892	0.155	-0.737
36	0.894	0.155	-0.739
37	0.863	0.155	-0.708
38	0.896	0.155	-0.741
39	0.921	0.155	-0.766
40	0.95	0.155	-0.795
41	0.919	0.155	-0.764
42	0.885	0.155	-0.73
43	0.879	0.155	-0.724
44	0.799	0.155	-0.644
45	0.851	0.155	-0.696
46	0.893	0.155	-0.739
47	0.819	0.155	-0.664
48	0.945	0.067	-0.878
49	0.96	0.155	-0.805
50	0.914	0.155	-0.759
51	0.905	0.155	-0.75
52	0.871	0.155	-0.716
53	0.804	0.155	-0.649
54	0.728	0.155	-0.573
55	0.655	0.067	-0.588
56	0.552	0.155	-0.397
57	0.471	0.067	-0.404

58	0.401	0.067	-0.334
59	0.471	0.067	-0.404
60	0.514	0.155	-0.359
61	0.467	0.067	-0.4
62	0.413	0.067	-0.346
63	0.272	0.067	-0.205
64	0.146	0.067	-0.079
65	0.155	0.155	-0
66	0.049	0.155	0.106
67	0.054	0.155	0.101
68	0.000	0.155	0.155
69	0.116	0.155	0.039
70	0.218	0.363	0.145
71	0.313	0.155	-0.158
72	0.313	0.155	-0.158
73	0.422	0.155	-0.267
74	0.497	0.259	-0.238
75	0.653	0.155	-0.498
76	0.592	0.155	-0.437
77	0.264	0.155	-0.109
78	0.585	0.155	-0.43
79	0.494	0.155	-0.339
80	0.523	0.155	-0.368
81	0.513	0.155	-0.358
82	0.689	0.155	-0.534
83	0.827	0.155	-0.672
84	0.819	0.155	-0.664
85	0.748	0.769	0.021
86	0.643	0.924	0.281

=== Summary ===

Correlation coefficient	0.1185
Mean absolute error	0.4847
Root mean squared error	0.5425
Total Number of Instances	86

==========

Above display the results obtained from the model by testing the remaining data set. The total number of 86 instances were used for the whole dataset to test the accuracy of the model. According to the model test, that shows less error rates of Mean Absolute Error and the Root mean squared error which is 0.4847 and 0.5425 respectively.

6.3 Evaluate the Results

By converting the above result of the real values, can get a clearer idea of how much the values of the data are deviating from the real values. To do that output of the prediction model export to the excel workbook and converted values to real values by applying the standard deviation formula.

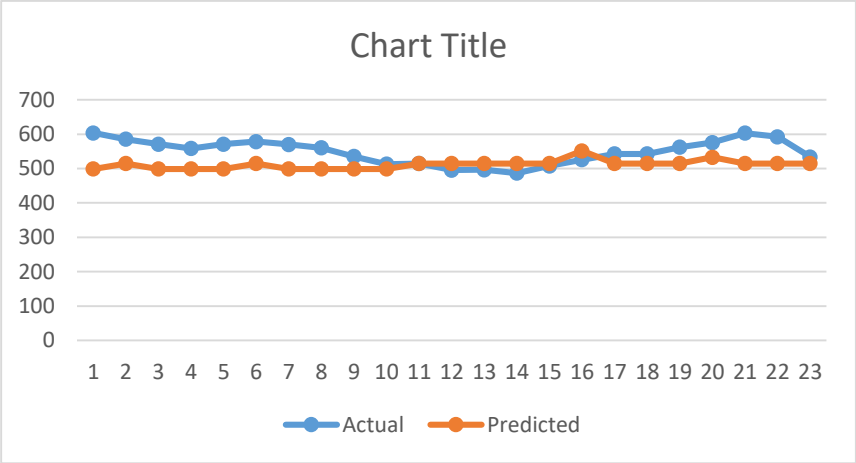
$$SD = \frac{X - X\,min}{X\,max - X\,min}$$

6.2 Data conversion formula

	A	B	C	D	E	F	G
1	Actual		Predicted		Difference		percentage
56	603.7255		498.8321		104.8933		17.37434
57	585.3513		514.5305		70.82083		12.09886
58	570.9017		498.8321		72.06956		12.62381
59	558.4144		498.8321		59.58226		10.6699
60	570.9017		498.8321		72.06956		12.62381
61	578.5725		514.5305		64.04201		11.06897
62	570.1881		498.8321		71.356		12.51447
63	560.5551		498.8321		61.72294		11.01104
64	535.4021		498.8321		36.56995		6.830371
65	512.9249		498.8321		14.09281		2.747538
66	514.5305		514.5305		0		0
67	495.6211		514.5305		18.9093		3.815273
68	496.5131		514.5305		18.0174		3.628787
69	486.88		514.5305		27.6505		5.67912
70	507.5732		514.5305		6.9572		1.370679
71	525.769		551.6356		25.8666		4.919765
72	542.7161		514.5305		28.18562		5.193438
73	542.7161		514.5305		28.18562		5.193438
74	562.1606		514.5305		47.63013		8.472691
75	575.5398		533.083		42.45682		7.376869
76	603.3687		514.5305		88.83822		14.7237
77	592.4869		514.5305		77.95643		13.15749

6.3 output data error rate

The above excel sheet shows how much the predicted auction prices are deviating from the real auction prices. The overall error percentage for 86 instances of the test dataset is 14.02%. That means the model has a better accuracy rate for forecast the tea auction price.



6.4 Deviation of error rate

The degree of correlation of the variables of tea production, tea auction quantity, tea export quantity, crude oil price and the US dollar exchange rate with the class variable of tea auction price are having the best relationship. The above error percentage rate shows the accuracy of the prediction model and the accuracy of the variable data.

But the downgrade of this test is, when take the result as an instance by instance the error percentages are not in the same linear line. Those are varied from 0 to 20% rate, according to the variation of the variable data. If the variations of the instance are having high error percentage, the deviate value is also high from the real price. In those situations the price difference may reach the Rs.100/- mark and that will be a problem for people who are willing to get predictions from the model.

6.4 Summary

This chapter discussed how to test and evaluated the data obtained from the classification model as a result. This shows how much of accuracy is having the model and what are the error rate of single instances as well. Next chapter is a discussion and conclusion and descriptive summary of the whole project, what are the limitations and future works based on this study.

Chapter 7

Discussion and Conclusion

7.1 Introduction

This chapter will discuss about the summary of the research has been done to yet, what are the limitations having, what kind of systems that can build and what are the future works can do based on this study.

7.2 Discussion

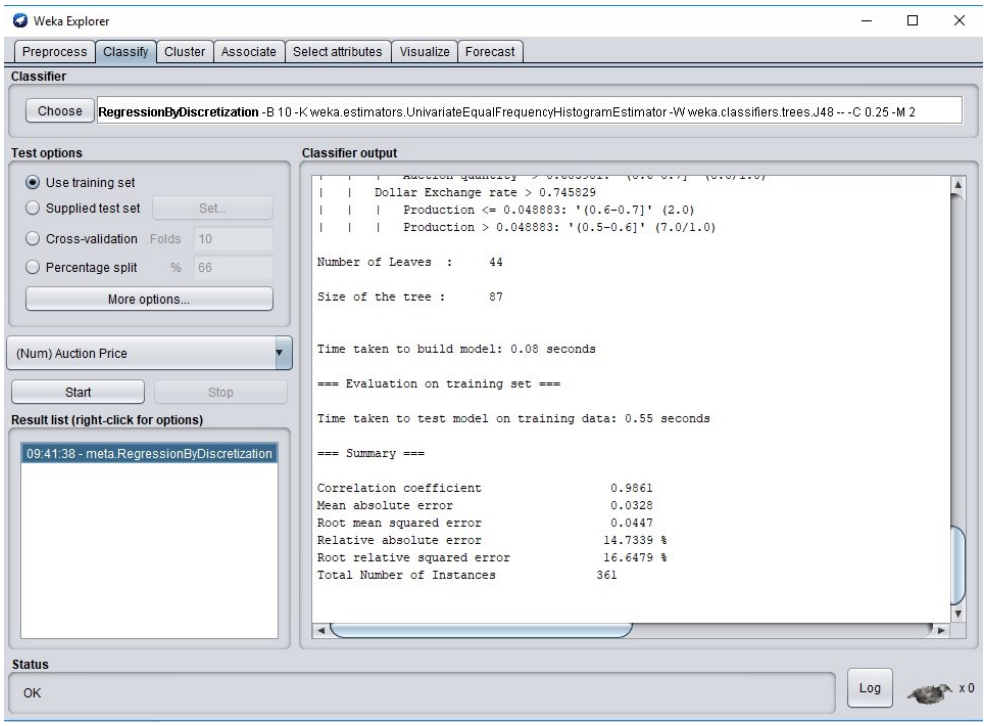
Tea is one of the major agricultural exports in Sri Lanka and earn considerable foreign remittance to the country. Tea auction is happening on every week and price will be fluctuated in every time. Lots of parties, including tea estate owners, auctioneers, exporters and customers are extremely causes of the tea auction prices and very much consent if there is any price forecast available. The Sri Lanka tea board is the official governing body of tea of Sri Lanka and has no mechanism yet to predict the price forecast for future purposes. Therefore, this research is to find the mechanism for predicting the tea auction price forecast for future purposes. In here analyses the factors that are most related to the tea price variance to build the suitable model for forecasting.

7.3 Conclusion

According to the results obtained from the model, it gave only the 14.01% error rate. That means the predicted auction prices are having over 85% accuracy rate compared to the real price. Also, this will prove the selected variables are having a high degree of relationship to decide the tea auction price in Sri Lanka. The Regression by Discretization is much more suitable than any other classification algorithm to predict the tea auction price with the correlation of the affecting variables.

7.4 Limitations

There were a few limitations occurred while carrying out this research. The historical data of tea price in the source are only available for 360+ weeks. So the dataset has to divide for 2 sections for training and testing purposes. Only 275 instances from the data set were used to the training and build the model. It gave the 25% of Relative absolute error and 22% of root relative squared error rate. But if used all the dataset for training, the error rates are 14% of Relative absolute error and 16% of root relative squared error. This will show if the dataset is larger than this, the result may be more accurate than this. But still this research gives a higher percentage of prediction accuracy.



7.1 Regression by discretization applying full data set

Another limitation faced during the research was, the accuracy of the predictions may gradually go down if it used to forecast prices for more steps to a head. Because the model was built using correlations of affected factors data and the previous auction price. Therefore the most accurate result would be 1 step ahead of the current data.

7.5 Future works

This thesis only builds a predictive model up to this stage, which can forecast the tea auction price. The next level of the research is to develop a software based on this study. The model pattern and result identified during the research has to be used as the base of the new software. Simple web based system is preferred when develop the software, because of its smoothness and lightness and easy access. The system should include the facilities such as previous data entering, number of forecasting steps required, graphs which display the data pattern of variables, etc.… which help to decision making for the user. By introducing such system will help to the tea industry and people who involved in Sri Lanka. And for the people who are interested in joining to the tea industry by making business opportunities also may get the help of this software.

This research covers only the elevation average tea auction prices for build a predictive model. One elevation has various kinds of tea categories and three elevations had different tea categories which are not similar to each elevation. For further studies to extend this research, analyze to the depth of those tea category, price and build a predictive model which can forecast the tea auction price.

7.6 Summary

This chapter discusses about why the research has been done, what the limitations are observed during the study and what are the future works that are available based on the study.

References

[1] S. L. E. D. Board, "Industry capability report - Tea," 2016.

[2] R. D. Gunathilaka and G. A. Tularam, "The Tea Industry and a Review of Its Price Modelling in Major Tea Producing Countries," *Journal of Management and Strategy,* vol. 7, p. 21, 2016.

[3] T. C. C. o. Commerce, "The Colombo Tea Auction," 2017.

[4] F. Xia, Z. Liu, and Q. Zhou, "A Data Mining Framework for Tea Price Evaluation," in *Semantic Computing (ICSC), 2016 IEEE Tenth International Conference on*, 2016, pp. 366-369.

[5] D. Rajapakse, "Analysing Tea Auction Trends for Beneficial Seasonal Tea Production in Sri Lanka," 2015.

[6] H. Fernando, W. Tissera, and R. Athauda, "Gaining insights to the tea industry of Sri Lanka using data mining," in *Proceedings of the International MultiConference of Engineers and Computer Scientists*, 2008.

[7] H. Hettiarachchi and B. Banneheka, "Time series regression and artificial neural network approaches for forecasting unit price of tea at Colombo auction," *Journal of the National Science Foundation of Sri Lanka,* vol. 41, 2013.

[8] R. Weron, "Electricity price forecasting: A review of the state-of-the-art with a look into the future," *International journal of forecasting,* vol. 30, pp. 1030-1081, 2014.

[9] N. Senaviratna, *Forecasting Tea Auction prices in Sri Lanka: Box-Jenkins modeling approach* vol. 6, 2016.

[10] H. Liu and S. Shao, "India's Tea Price Analysis Based on ARMA Model," *Modern Economy,* vol. 7, p. 118, 2016.

[11] D. Induruwage, C. Tilakaratne, and S. Rajapaksha, "Forecasting Black Tea Auction Prices by Capturing Common Seasonal Patterns," *Sri Lankan Journal of Applied Statistics,* vol. 16, 2016.

[12] L. Popova, F. Jabalameli, and E. Rasoulinezhad, "Oil Price Shocks and Russia's Economic Growth: The Impacts and Policies for Overcoming Them," *World Sociopolitical Studies,* vol. 1, pp. 1-31, 2017.

[13] F. Walkers, "Tea Review," 2014.

[14] M. De Silva, U. Jayasinghe-Mudalige, J. Edirisinghe, H. Herath, and J. Udugama, "Assesing the production Vs. Price relationship of black tea in Sri Lanka: An applicaton of Koyck's Geometric-Lag model," *Economic Research,* vol. 2, pp. 43-53, 2014.

[15] M. N. Mutegi, "An Empirical investigation of factors influencing tea export earnings in Kenya," 2015.

[16] E. Nyaga and W. Doppler, "Use of Dynamic Models to Assess Impact of Changing Tea Prices on Family Income of Smallholders in Kenya," *Journal of applied sciences,* vol. 9, pp. 1647-1657, 2009.

[17] E. Ekanayake and D. Chatrna, *The Effects of Exchange Rate Volatility on Sri Lankan Exports: An Empirical Investigation* vol. 11, 2010.

[18] J. Han, J. Pei, and M. Kamber, *Data mining: concepts and techniques*: Elsevier, 2011.