

# Future forecasting and analysis of Sri Lankan tea exports in terms of driving forces using data mining concept

R Laveena<sup>1#</sup> and UP Kudagamage<sup>1</sup>

<sup>1</sup>Department of Computing and Information Systems Faculty of Applied Sciences,  
Sabaragamuwa University of Sri Lanka Belihuloya, Sri Lanka

#rlaveena@std.appsc.sab.ac.lk

**Abstract:** In order to identify new research directions and gaps in the body of knowledge for the time period under consideration, this study concentrated on collecting data specifically related to the Future Forecasting and Analysis of Sri Lankan Tea Exports in Terms of Driving Forces Using Data Mining Concepts. For individuals working in the sector, this research analyzes and projects tea exports depending on the types of tea exported. Finding the elements that lead to fluctuations in tea export volume is made easier by examining the link between tea export and important variables. The prices and volumes of various tea types over that time period, as well as monthly data on tea exports from the previous ten years, were all used in this study. These historical data were utilized to assess and establish the strength of the correlation between the important variables and their patterns of variation in order to forecast tea export volume using WEKA software. Out of a variety of prediction and forecasting techniques, the Multilayer Perceptron, a form of feed forward Artificial Neural Network, was determined to be the most efficient method for creating an accurate prediction model. A confusion matrix was used to gauge the accuracy of the results. With a 98 percent accuracy rate, this forecasting model is suitable for predicting the volume of tea exports. It is also discovered that year, month, and tea types have the highest level of connection among the components in determining Sri Lankan tea export.

**Keywords:** Economic conditions in Sri Lanka, machine learning, and varieties of tea

## 1. Introduction

With the planting of a batch of tea seeds at the Royal Botanic Gardens in Peradeniya in 1839, tea was first brought to Sri Lanka. On 19 acres of land on the Loolkandura estate in Hewaheta, James Taylor began the first commercial tea planting in 1867. A complicated input of many cultural practices was needed for the tea crop, which required substantially more effort. Following water in terms of global consumption, tea is the most

manufactured beverage. One of the key industries in Sri Lanka's economy that generates significant export revenue is tea. Since it generates the most net foreign exchange on the island today, tea is the most significant agricultural industry. In comparison to Sri Lanka's garment and textile industries, tea exports generate twice as much net foreign cash. As a result, it is believed that the tea industry will continue to be important to Sri Lanka's economy going forward. Employment is one of the ways the tea industry specifically benefits the national economy. Tea's laborintensive production structure results in a high degree of employment. The tea business has supported Sri Lanka's economy for a number of years. The island is definitely veering away from other criteria, like comparative advantages, as evidenced by the present tea industry view. For more than millions of islanders, the tea industry provides their primary source of income. Tea in the forms of bulk, packaged, tea bags, black, green, and white tea are among the varieties exported by the country. For Sri Lanka to maintain a competitive edge in the market, it is necessary to adhere to a number of policies and initiatives. As a result, it is necessary to address the problems in the tea sector and grow the Sri Lankan tea business by utilizing cutting-edge technologies. To increase the competitiveness of Ceylon tea exports globally, it is necessary to further investigate how all the independent features affect tea output and how their positive and negative association through time. By examining the elements that affect tea export, this study's main goal is to increase tea export in Sri Lanka. It gives a brief overview of the important variables that affect Sri Lanka's export of tea as well as their ramifications. This study identifies and analyzes the role that tea export plays in Sri Lanka's economy. On the international market, the major rivals are also identified and their contributions to the tea market are discussed. The study identified the major tea types that are shipped from Sri Lanka to the international market and determined the kinds with the highest and lowest export demand, as well as the proper explanations for each. The goal of this study is to make the research's findings highly accurate. Additionally, this will support Sri

Lanka's Department of Export's efforts to remain competitive on the world stage. The goal of this research is to build a platform that can forecast Sri Lankan tea exports when a specific set of crucial conditions are met.

## 2. Methodology and Experimental Design

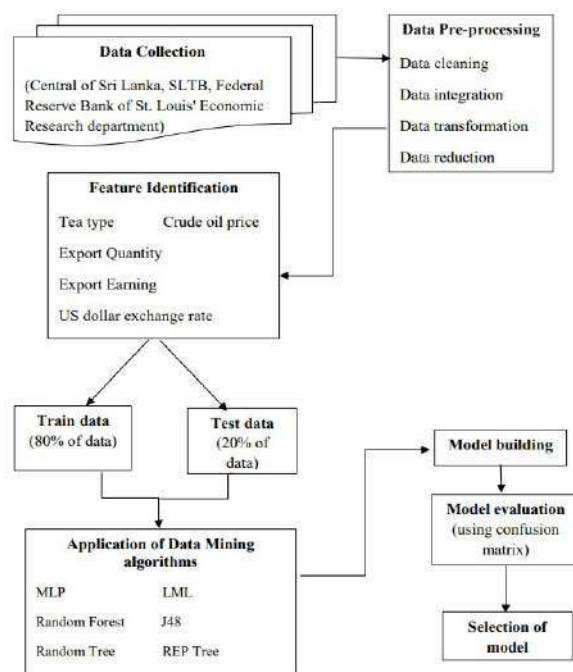


Figure 1: Overall Methodology

### A. Data collection

Multiple types of data inputs are employed for the compilation of data sets for the previous 10 years, which span from 2011 to 2021. The central bank of Sri Lanka supplied information on several tea sorts including Green tea, Tea in Bulk, Tea in packets, Tea in bags, and Instant tea, while the official website of the Sri Lanka Tea Board statistical section gave data on tea export prices and quantities. The Economic Research division of the Federal Reserve Bank of St. Louis provided the data on crude oil prices, which included daily oil prices from 2011 onward and were converted into monthly prices. Additionally, information about the US dollar exchange rate was gathered from Sri Lanka's central bank. 630 different occurrences were taken into account for the study out of the data collection.

### B. Data Pre-processing

Data is dated, imperfect, and noisy in the actual world. Unsatisfactory in that it lacks important feature values and attribute values, or just has aggregate values, noisy in that it contains errors or outliers, and contradictory in that it includes name or code inconsistencies. Why is the data outdated is the next question. Since non-applicable data values might lead to incomplete data when data must be acquired, and the main problem is a variation in thinking between the times when the data was evaluated, together

with human hardware and software concerns. Everyone makes mistakes while entering data, therefore noisy data can happen when a human enters the incorrect value. faulty data collection tools and data transmission issues. There are many different sources of inconsistent data. Duplicate data calls for more data cleansing. With raw data, common issues include noise, missing numbers, and consistency. The caliber of the initial data has an impact on data mining outcomes. Preparing and modifying the initial dataset is one of the most crucial processes in the data mining process since pre-processing data helps to improve data quality, which in turn improves the mining results. The data pre-processing techniques are illustrated in the list below.

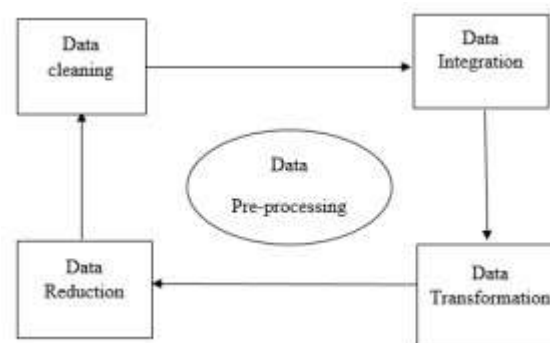


Figure 2: Data Pre-Processing Techniques

During the data cleaning stage, missing values are handled with. It is necessary to perform the duties of identifying outliers, smoothing noisy data, and attempting to repair erroneous data. The determination of some production and export volumes was made utilizing data mining techniques for addressing missing information.

#### 1) Data Integration:

This step involves combining and logically storing data that has been collected from several sources. During the integration process, challenges like identifying similar entities with multiple labels while merging and resolving data value conflicts involving similar entities with various metrics in multiple units are some of the issues that must be resolved. Both during data transformation, there are several periods when merging data and producing aggregate values are employed as ways of data integration. Both of these approaches were used using the dataset on tea exports.

#### 2) Data Transformation:

This step involves combining and logically storing data that has been collected from several sources. During the integration process, challenges like identifying similar entities with multiple labels while merging and resolving data value conflicts involving similar entities with various metrics in multiple units are some of the issues that must be resolved. Before and after data transformation, there are several periods when merging data and producing aggregate values are employed as ways of data integration. Both approaches were used using the dataset on tea exports. The purpose of data mining and various data mining

technologies both call for this, thus it is done. This was done in order to enable for monthly dataset unification and to make sure that the dataset's determining criteria were constant.

### 3) *Data Reduction:*

It is used to describe either a reduction in data volume or quality (number of attributes). Due to the fact that the data was manually entered and had few features, there were zero values since further processing might have been done without reducing the data.

### C. *Feature Identification*

The characteristics that have an impact on the research study from the data collected have been determined, including export quantity, export earnings, US dollar exchange rate, and crude oil price.

#### 1) *Tea Export:*

To establish the pattern of correlation between export volume and sales, which has a significant impact on forecasting the tea export in Sri Lanka in terms of influencing factors, the total amount of exported tea, expressed in kilograms, is studied monthly.

#### 2) *Price:*

The information provided here pertains to the profits made by Sri Lankan tea exports to foreign nations in LKR. When predicting tea export, this attribute is crucial.

#### 3) *Tea Type:*

Based on taste, elevation, and structure, tea can be categorized into a number of groups. Some of them include Green, Black, White, Instant, RTD and CTC, High grown tea, Mid grown tea, etc., but only five types—Tea in bags, Tea in packets, Tea in bulk, Instant, and Green tea—were used in this analysis for forecasting that is used to analyze the export in accordance with the types of the tea that are exported from Sri Lanka.

#### 4) *USD:*

According to this information, the US dollar's value relative to the Sri Lankan rupee has already been shifting significantly over the past few years because of both the country's political policies and its erratic foreign revenue. This had a respectable impact on Sri Lanka's export of tea.

#### 5) *Crude Oil Price:*

Numerous elements, such as regional shifts in exporting nations, topographical characteristics, and environmental variables, contribute to the swift price fluctuations of crude oil. The European oil market will therefore daily adjust its oil prices as a result. One of the key elements with a significant impact on the nation's tea exports is this characteristic.

### D. *Model Building for Tea Export Prediction*

Typically, classification algorithms are used to extract forecasts from a dataset. Finding a suitable classification method is one of the objectives of this project in order to create a model for predicting tea exports. The class data is analyzed by algorithms along with the different influential elements to produce accurate results. The algorithms then construct a model wherein the parameters are closely related to the class data. The model for predicting tea exports is therefore built and analysed using a variety of classification methods. Classification techniques were trained using the pre-processed tea export data set, which was entered. In order to train the models, information from 80% of the dataset were utilised. A proposed methodology that can estimate the volume of tea exports based on the influencing factors was developed in this study after analyzing how closely the factors matched the class data.

### E. *Model Evaluation*

The leftover 20% of the data set was maintained for testing, and the remaining 80% was utilized for building a model using an appropriate classification technique. The remaining pre-processed dataset was added as a test data set to the model to test it, and the tea export was created for the subsequent steps. The created models were each tested independently by inserting datasets. In order to acquire predictions, the constructed model was loaded after the test data set for the tea export industry was loaded. Utilizing a variety of criteria, including the confusion matrix and its derivations, the performance and efficiency of the algorithms were assessed.

## 3. **Results**

A total of 126 occurrences from the full dataset were used to assess the model's precision. The model that provides the highest accuracy was selected for the study after numerous classification algorithms were examined for their Accuracy, Precision, Recall, F-Measure, MCC, ROC Area, and PRC Area. The datasets collected were used to train nine distinct classifiers. The summary findings from using the preprocessed data set as training data for a variety of different classifiers linked to the forecasts are provided in Table 1. The results show that, in terms of affecting factors, the nine distinct classifiers have performed comparably better in predicting Sri Lanka's tea export, with all developed models achieving more than 90% accuracy. Among the numerous classification methods, the Multilayer Perceptron algorithm and Random Forest had the highest accuracy (98%) and the lowest percentages of Mean Absolute Error and Root Relative Mean Squared Error. The proportion of Mean Absolute Error and Root Relative Mean Squared Error of the built model was taken into consideration to determine the best design methodology for tea export prediction of Sri Lanka in terms of influencing factors because the Accuracy, Recall, and F-Measure of the two classifiers are identical.

The study employed an ensemble approach to improve the outcome and developed prediction models using the first five highest accuracy forecasting techniques. The newly built model provided an accuracy of 97.6%, which is less accurate than Multilayer Perceptron and Random Forest classifiers. In order to create the forecasting models for tea export, the study used the Multilayer Perceptron technique.

#### A. Attribute Ranking

The research hypotheses used for the investigation are ordered according to their impact on the prediction of tea export as shown in the following table, which is based on the analysis of data of the finalized dataset used for the prediction of Tea export in ways that affect factors. Within research study, it was discovered that the elements year, month, type of tea, and export earnings had the most influence on dependent variable export volume, while the factors US dollar exchange rate and crude oil price had the least.

Table 1: Attribute Ranking

Attribute	Rank
Year	1
Month	2
Type	3
Export earning	4
US Dollar exchange rate	5
Crude oil	6

#### B. Results according to tea types

The study's primary focus is the forecasting of tea export volumes for several tea varieties, including green tea, instant tea, tea in bags, tea in packages, and tea in bulk. According to the data analysis, tea in packages was sent more frequently per month than instant tea, but both were exported on a monthly basis. The second and third largest quantities of tea are exported in bulk and bags, respectively. These results demonstrate that tea in packets plays a significant part in Sri Lanka's export of tea in terms of influencing factors and offers a comprehensive economic rate to the nation.

#### 4. Discussion

The main objective of this research was to examine the elements that affect tea exports in order to increase Sri Lankan tea exports. The study identified the major factors that affect Sri Lanka's tea export and how they affect it. Additionally, the major rivals in the world market are named and their commitments to the tea industry are assessed in terms of export volumes and export prices. The study identified the main tea kinds exported by Sri Lanka to the international market, along with the varieties with the maximum and minimum export demand. While instant tea is at least shipped from Sri Lanka, tea in packets is known to be exported in massive quantities. Nine alternative

classifiers are employed in the search for the optimum performance model for tea export forecasting. The Multilayer Perceptron classifier was identified as the high accuracy model to carry out the investigation by analyzing the accuracy, recall, F-Measure, mean absolute error, and Root mean squared error with the aid of the confusion matrix. The year, month, tea kind, and export revenues all play a vital role in the forecasting of tea exports, according to the final dataset that was utilized for the study.

#### 5. Conclusion

With large foreign exchange revenues, tea has emerged as one of Sri Lanka's most important export crops. The amount and price of tea exported varies from month to month, and it only happens once per month. An extensive range of stakeholders, including 40 tea landowners, brokers, distributors, and consumers, are fervent proponents of tea export. The Sri Lanka Tea Board is the government's national regulatory organization for tea, however at the moment there isn't a mechanism in place to aid in forecasting exports based on different types of tea. Because of this, the aim of this research is to identify a methodology for predicting tea export estimates in the next years. In order to create a suitable prediction model, the features that affect tea export variation were examined in this paper. The Multilayer Perceptron is far more suitable than any classification system for predicting tea export well with the correlation of the affecting parameters. According to the results, the model only produced an error rate of 1.5873 percent. As a result, when compared to the original number, the forecasted export volume is more than 98 percent accurate. Additionally, it would show that the factors have a higher degree of association in affecting Sri Lanka's export of tea. It has been determined that the features Year, Month, and Tea Type have the most impact on Sri Lanka's export of tea, correspondingly.

#### 6. Recommendations for Further Research

This study only extends as far as to create a forecasting model for tea exports. The more the study can produce software based on its conclusions. The model pattern and findings from the research should be the foundation around which the software application is developed. To help the user come to conclusions, the platform must provide elements like previous data entry, the number of forecasting steps required, and graphs that show the data pattern of various components. The establishment of such a structure will be advantageous to the industry as well as those engaged in Sri Lanka. This program can also be useful to people who want to enter the tea industry by developing business opportunities. To create a prediction model, this study primarily determines the types of tea that are exported as well as historical data on tea export. Beyond the scope of this analysis, analyze the specifics of those, auction prices, and tea elevation, and create a predicting model that can predict the volume of tea exported together with auction price in subsequent studies.

Table 2: Comparison of Classifiers - Accuracy And Error Rate

Classifier	Accuracy	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Mean absolute error%	Root mean squared error%
Naive Bayes	92.06	0.932	0.92	0.921	0.853	0.991	0.99	0.0913	0.235
Multilayer Perceptron	98.41	0.985	0.98	0.984	0.969	1.0	1.0	0.023	0.124
Simple Logistic	96.825	0.968	0.96	0.968	0.936	0.999	0.99	0.038	0.123
Decision table	93.65	0.944	0.93	0.937	0.880	0.990	0.99	0.08	0.203
REP Tree	96.03	0.964	0.96	0.960	0.923	0.996	0.99	0.047	0.150
Random Tree	97.61	0.977	0.97	0.976	0.953	0.978	0.96	0.023	0.154
Random Forest	98.41	0.985	0.98	0.984	0.969	0.998	0.99	0.037	0.126
LMT	96.85	0.968	0.96	0.968	0.936	0.999	0.99	0.038	0.123
J48	97.61	0.977	0.97	0.976	0.953	0.997	0.99	0.03	0.143

## References

A. Disdier, L Fontagne, M Mimouni, 2008. The impact of regulations on agricultural trade: evidence from the SPS and TBT agreements. *American Journal of Agricultural Economics*, Volume 90, pp. 336-350.

Anon., 2016. *Industry Capability Report- tea*, Sri Lanka: Sri Lanka Export Development Board (EDB).

Anon., 6 August 2020. *Economic and social statistics of Sri Lanka 2020*, s.l.: Statistics Department, Central Bank of Sri Lanka.

Ariyawardana, A., 2001. Performance of the Sri Lankan value added tea producers. In: *An integration of resource and strategy perspectives*. s.l.:s.n., p. 314.

D. Kasthuriratne, N. Poole, 2006. *Creating Value for Competitive Advantage in supply Relationship: The case of the Sri Lankan Tea Industry*. Germany, 99th EAAE Seminar.

D.Kasthurirathne, 2008. An Overview of Sri Lankan Tea Industry: An explanatory case study. *The Marketing Review (Winter)*, pp. 367-381.

H C Fernando, W M R Tissera, R I Athauda, 19-21 March 2008. *Gaining insights to the Tea Industry of Sri Lanka using Data Mining*. Hong Kong, Proceedings of the International Multi Conference of Engineers and Computer Scientists 2008.

Jayarathne, P., 2018. *Effect of Governance and policy on the management of a sustainable agri-supply supply network: An example from the sri lankan tea industry*. s.l., s.n., pp. 196-201.

Lei, L., 2017. *Effects of standards on tea exports from developing countries : Comparison of China and Sri Lanka*. s.l.:IDE Discussion Papers 642, Institute of developing economies, Japan external trade organization (JETRO).

Thusara, C., 22 - 24 May 2015. *Sri Lankan Tea Industry: Prospects and Challenges*. Dubai, s.n.

X. Martin, Nyaga Mutegi, November 2015. *An empirical investigation of factors influencing tea export earnings in Kenya (1973 - 2013)*. Kenya: University of Nairobi.

## Acknowledgment

I would like to express my sincere thanks also go to all the academic and non-academic staff and my friends of the Department of Computing and Information Systems, Faculty of Applied Sciences, Sabaragamuwa University of Sri Lanka for their support and wishes for the successful completion of this research study.

## Author Biography



Laveena Rajendran is a graduate with BSc Special in Computing and Information systems, from Sabaragamuwa University of Sri Lanka. Her current field placement is demonstrator at the department of Computing and information systems, faculty of Applied sciences, Sabaragamuwa University of Sri Lanka.



Miss Upeksha Kudagamage graduated from the Department of Computing and information systems, faculty of Applied sciences, Sabaragamuwa University of Sri Lanka. She is currently working as a Lecturer (Probationary) in Information Systems attached to the Department of Computing and Information Systems, Faculty of Applied Sciences and she is pursuing her Master in Computer Science at University of Colombo School of Computing.