American Scientific Research Journal for Engineering, Technology, and Sciences (ASRJETS)

ISSN (Print) 2313-4410, ISSN (Online) 2313-4402

http://asrjetsjournal.org/

Association Pattern Discovery of Import Export Items in Ethiopia

Mezgeb Manaye Gebru*

Unity University, Addis Ababa, Postcode 170253, Ethiopia

Email: mezgebmanaye@gmail.com

Abstract

This paper examines the application of data mining to detect association pattern of customs administration data with market price and currency rate exchange in Ethiopia. The association rule method of data mining is used in this paper to generate the interesting pattern from the data. This study was done to identify the relationships between attributes of custom data and market price to clearly understand the nature of import-export items in Ethiopia. The results of the experiments carried out using association rules revealed that the technique of data mining is applicable to generate knowledge from import and export items in custom administration. Algorithms such as Apriori, Tertius, PredictiveApriori and FliteredApriori were used to generate the associations. One of the resulting associations indicates that there is a strong link between market price and textiles imported. The implication of this research finding is that it clearly identified the association of import-export items with the market price and the effects of those items on the market price and currency rate in Ethiopia.

Keywords: Data mining; Association rule; Weka; Association pattern; Apriori; Tertius; PredictiveApriori; FliteredApriori; Algorithm.

I. Introduction and Background

The incapability of human beings to interpret and digest accumulated huge data and make use of them for decision-making has created the need for the development of new tools and techniques for automated and intelligent analysis. As a result, the discipline of knowledge discovery or data mining, which allows users to analyze data from many different dimensions or angles, categorize it, and summarize the relationships identified, has evolved into an important and active area of research [20].

----

* Corresponding author.

Data mining is an automated process employed to analyze patterns in data and extract information [29]. Central to data mining is the process of modeling data set. There are widely used generic modeling techniques: some of these are neural networks, Agent Networks, Genetic Algorithm, Decision trees, and hybrid models [3].

Data mining's extraction of meaning from huge databases is exactly what companies are looking for to increase profits through describing past trends and predicting the future ones [7]. It is about extracting interesting patterns from raw data. There is some agreement in the literature on what qualifies as a “pattern” (association rules and correlations [22,23,24,25] as well as clustering of the points [14], are some common classes of patterns sought), but only disjointed discussion of what “interesting” means. Most works on data mining study patterns to be extracted automatically, presumably for subsequent human evaluation to the extent in which they are interesting. Patterns are often deemed “interesting” on the basis of their confidence and support [22,23,24,25], information content [15,19], and unexpectedness [1].

Modern data mining is used intensively and extensively by financial institutions (for credit scoring and fraud detection) [18], retailers (for market segmentation and store layout) and manufacturers (for quality control and maintenance scheduling) [13].

Similarly, various studies have addressed the different aspects of customs administration and market price by using data mining techniques and other statistical approaches. Numerous data mining-related studies have been undertaken to analyze customs data, with results frequently varying depending on the socio-economic conditions and infrastructure of a given location.

II. Scope and limitation of the study

The scope of this research, in terms of purpose, is to examine the potentials of data mining techniques in investigating the effects of export import on the economy and import export items relationship using association rule data mining. Particularly, the study is limited to finding associations between export items such as, coffee, livestock, oil seeds and major import items such as, food, fuel, textiles that are stored from 2002 to 2017. Because these import and export items are dominant and major items in Ethiopia than other items. The research is limited to currencies only: Birr and USD. Moreover, the scope of data collection is limited to 15 years.

III. Related Works

Import-export activity of Ethiopia is strongly related to the country’s economy and policy. As this research focuses more on aspects that influence price from custom perspective, a review was conducted of prior works that dwelled on factors affecting the price of commodities.

A study by Elias entitled, “The Effect of Depreciation of Birr on Major Export Products of Ethiopia: The Case of Hides and Skins,” analyzed the effects of trade and exchange rate policies on one of Ethiopia’s agricultural export items, hides and skins. This study found out what happened in the export of hides and skins when the exchange rate depreciated continuously for the last 17 years. The empirical findings of the study revealed that real exchange rate was one factor, among many others, that affected the volume of export of hides and skins.

Hence, it was recommended that "policy towards liberalized exchange rate determination should be complemented by other policy measures which are in harmony with the economic agenda of export enhancement" [5]. This paper shows the impact of Birr exchange fluctuation on export, but not vice versa and the association between Birr and import product.

The Impact of Trade Liberalization on the Ethiopia's Trade Balance was presented by [8].The model used for the analysis of the impact of trade liberalization on trade balance was based on export equation of Santos-Reference [26]. This study examines the impact of trade liberalization on the Ethiopia's trade balance using the data over the period 1974 to 2009 from NBE (National Bank of Ethiopia). The country has undertaken serious trade reforms, either as a part of major macroeconomic reforms and commitments with international regulations, or by decisions driven by a process of internal adjustment for the last two decades. One of the anticipated gains from the trade liberalization policies adopted by Ethiopia is improved export performance. The author of this paper analyzed the impact of trade liberalization on Ethiopia's trade balance. However, when it was examined with the application of export equation, it showed that trade liberalization led to the deterioration of the trade balance or too fast of an increase in imports. Thus, "It was concluded that trade liberalization worsens trade balance as a function of more imports than exports" [26]. Even though the paper revealed that the trade liberalization worsens trade balance due to more imports than exports, it did not demonstrate the association between import/export items with the market price and currency exchange, and which item could highly affect the market price and currency rate exchange.

Another study conducted by International Food Policy Research Institute examined Foreign Exchange Rationing, Wheat Markets and Food Security in Ethiopia [16]. The study tried to depict the relation between Ethiopia's domestic cereal markets and the international market. According to this paper, there was a remarkable growth in Ethiopia's agricultural production and overall real incomes (GDP/capita) from 2004/05 to 2008/09. Due to this growth, the prices of major cereals (Teff, Maize, Wheat and Sorghum) fluctuated sharply in both nominal and real terms [16]. International prices of cereals also fluctuated widely, particularly between 2006 and 2008[16]. However, the links between Ethiopia's domestic cereal markets and the international market are by no means straightforward. Among the major staples, only wheat is imported or exported on a significant scale, and frequent changes in trade and macro-economic policies, movements in international prices and fluctuations in domestic production have at times eliminated incentives for private sector imports of wheat. "Domestic wheat prices have been above wheat import parity prices since May 2008, indicating that it would be profitable for private traders to import wheat if they had access to foreign exchange at the official exchange rate"[16].

A recent paper written by World Bank and World Customs Organization (WCO), which examined container clearance/release times in gateway of African Ports, is an example of the analysis of data extracted from Customs IT systems [21]. This study provided information on why clearance/release times "are widely recognized as a critical hindrance to economic development". It also puts in context the inter-relationships among the logistics performance of consignees, operational performance of port operators, and efficiency of customs clearance operations" [21]. Moreover, the paper depicts that customs data mining helped the authors to identify and describe clearance and forwarding agents, shippers and shipping line strategies and their impact on time release.

Reference [4] used data mining to develop custom intrusion detection filters. The approach was to develop custom filters that reduce the false alarm stream based on known “normal behavior” in a particular environment. They used commercial intrusion detection systems, but filtered out produced alarms that fit a pattern caused by normal operation at that site. The difficulty with this approach is building these filters, and determining what is a normal operation at a site. While much less costly than building a complete intrusion detection system, it still requires considerable human effort. To reduce this effort, they used data mining technology to discover alarms caused by normal operation. They developed custom filters data mining model based on sequences of alarms using sequential association mining. The idea is that a sequence of operations that are normal in a particular environment may contain operations that look like a potential intrusion. However, the complete sequence is unlikely to be duplicated in an intrusion, so alarms that are part of the complete sequence can be ignored. The problem is being able to identify such normal sequences. They used frequent episodes [4] to identify frequently occurring sequences of alarms. An episode is a sequence of alarms that occurs within a specified time window [4].

Reference [11] used cluster data mining rule to Study and Apply Data Mining to Structure Risk Analysis of Customs Declaration. The cluster method of data mining is used in this paper to divide the cargo into seven types; Thus, customs can put the mainly inspection force to the high-risk level cargo. The results show that this kind of method can be used to reform the operation mode of customs inspection. Through the basic cluster analysis of cargo, they found that multi-variance statistical analysis is a very important method, and more effective than the other tools, which can offer the basement of quantity analysis, and support the decision-making [11].

Another study by World Bank and World Customs Organization (WCO) in 2011 examined risk management systems using data mining in developing countries’ customs administrations[10].The researcher involved in this study used data mining (descriptive statistics) to be successful in accurately targeting declarations that present a risk of infraction, to carry out prior work on data analysis, on descriptive statistics. This work requires customs to identify the characteristics of declarations that, in a preceding period has resulted in an infraction, and then deducing the ‘statistical regularities’ in those infractions. The researcher used database covers twelve months. Data came from detailed declarations and monthly statements of customs infractions for the two main offices in Dakar. Finally, they combined these risk profiles to facilitate the right decision with regard to referring the declaration to a particular customs clearance channel. Generally, this paper concluded that risk management on the custom administration improved or facilitated the decision, but it did not show the association of import/export items with the market price and currency exchange.

Scheffer carried out a more specific study on ‘Finding Association Rules that Trade Support Optimally Against Confidence’ [27]. The methodology used was knowledge Discovery in Databases (KDD). KDD is the process of discovering useful knowledge from a collection of data. When evaluating association rules, rules that differ in both support and confidence have to be compared; a larger support has to be traded against a higher confidence. The solution, which they propose for this problem, is to maximize the expected accuracy that the association rule will have for future data. They determine the contributions of confidence and support to the expected accuracy on future data. Moreover, they present a fast algorithm that finds the n best rules that maximize the resulting criterion. "The PredictiveApriori algorithm returns the n rules which maximize the expected predictive accuracy; the user only has to specify how many rules he or she wants to be presented" [27].

Han, Karypis and Kumar presented scalable parallel data mining for association rule [6]. The algorithms of the Apriori class are based on the simple observation: if a given item set is not frequent, then none of its supersets can be frequent. They have a level-wise behavior: they start with k=1 by evaluating singleton item sets, and base the computations performed at step k on the results of the previous iteration k-1. This level-wise behavior has been often criticized because of the consequent multiple scans of the dataset, one for each level. A lot of research has been thus devoted to minimizing the number of the dataset scans.

Flach and Lachiche presented Confirmation-Guided Discovery of First-Order Rules with Tertius [17]. This paper deals with unsupervised discovery of rules in first-order logic. The researcher defined statistically well-founded confirmation measure to induce rules that are in some sense unusual or interesting. Then, they described a complete best-first search algorithm that uses an optimistic estimate of the best confirmation of possible refinements of a rule to prune the search. The algorithm works on a function-free first order Prolog representation, which enables application to a wide range of structured domains, and to include background knowledge as part of the heuristic evaluation [17]."Tertius is a full-fledged and powerful rule discovery system" [17].

From the above reviewed papers, we can understand that there is research gap on the area of association mining of import items and export items, and the relationship between those items and economy regarding market price and currency rate of different countries, Thus, in the current study, the researcher used association rule data mining technique to examine the relationship between major export items: coffee, livestock and oil seeds import items: food, fuel and textiles, and market price and currency rate.

Consequently, they attempted to address the following research question in this paper:

- What sort of association exists between major import-export items, and market price or currency exchange rate in Ethiopia?

IV. Method

This part of the paper discusses the data set, tools and approaches used for attribute selection, dimensionality reduction, and model building.

A. General Approach

The writers of this paper made use of Cross-Industry Standard Process for Data Mining (CRISP-DM) data mining methodology with appropriate modification to fit into the problem domain at hand. CRISP data mining process model describes commonly used approaches that expert data miners use to tackle problems. Polls [9] showed that it is the leading methodology used by data miners. Accordingly, based on situational analysis on this study, business and data understanding were the first tasks, followed by the identification of data pre-processing tasks relevant to the data-mining goal. The last stage was model building and evaluation along with a possible recommendation to integrate the resulting pattern or knowledge with the existing one.

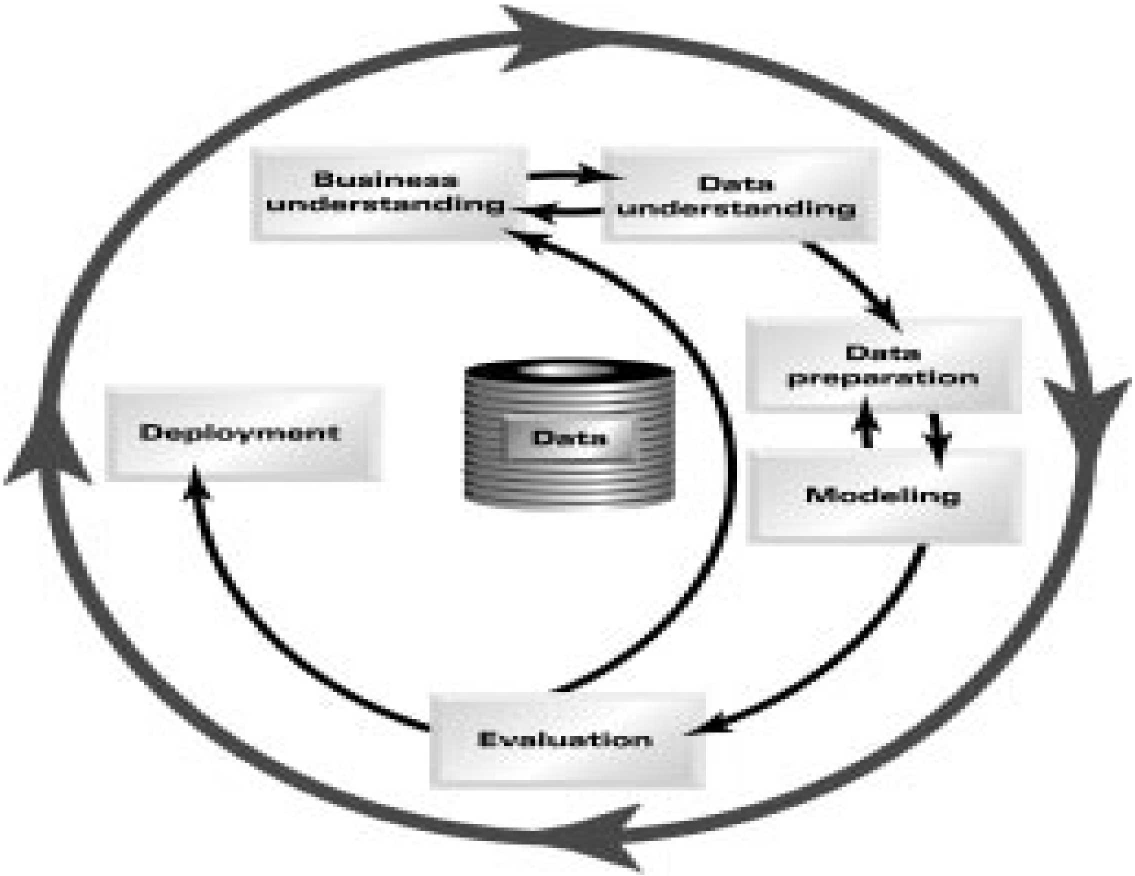

Figure 1: CRISP-DM Modeling Life cycle (Adopted from Wirth and Hipp, 2000)

B. Data Collection

This study used data obtained from Ethiopian Revenue and Customs Authority (ERCA), Central Statics Agency (CSA) and National Bank of Ethiopia based in Addis Ababa. The total dataset for the study contains import export records from 1997-2012, market price from 2000-2012 and currency rate from 1997-2012 respectively. Based on the availability of data, a total of more than 750,000 import export information was described with 10 attributes. The dataset has information related to import and export items, market price and currency exchange rate.

C. Tools

This study employed Weka data mining as a data analysis tool. Weka is a machine-learning system written using Java language. It was adopted for undertaking the experiment in data mining, which includes several algorithms that can be used for feature/attribute selection and model building.

There are also other data mining tools, one of which is Knowledge STUDIO. Knowledge STUDIO is an advanced data mining and predictive analytics suite for all phases of the model development and deployment cycle profiling, exploration, modeling, implementation, scoring, validation, monitoring and building scorecards all in a high performance visual environment. The models created can be exported in SQL, XML and SAS formats to allow them to be integrated into applications or systems.

In this study, Weka was selected as a tool because it is well suited for developing new machine learning schemes. It is open source software issued under the General Public License (GNU). It incorporates an association rule learner. It provides a number of data mining functionalities such as classification, clustering, association, attribute selection and visualization. It also includes about ten different methods for classification, three for clustering, and six for numeric prediction and several so called "metaschemes" (bagging, stacking, boosting . . .). In addition to the learning schemes, Weka also comprises several tools that can be used for datasets pre-processing [43]. On the other hand, Knowledge STUDIO is applicable only to Classification Discovery, Cluster Discovery, Data Visualization, and Discovery Visualization, but not to association rule discovery [21]. In addition to Weka, the researcher used MS-Excel for manual data processing.

D. Data mining techniques

According to Han and Kamber [18], data mining methods can be classified by the function they perform, (i.e. the kind of patterns to be found in data mining tasks) or according to the class of application they can be used in. Data mining technique can be classified into two categories: predictive and descriptive [18].

In predictive modeling tasks, one identifies patterns found in the data to predict future values [29]. Predictive modeling consists of several types of models such as classification, regression and AI-based models. Predictive models are built, or trained, using data for which the value of the response variable is already known [63].

Descriptive models belong to the realm of unsupervised learning [18,29]. Such models interrogate the database to identify patterns and relationships in the data. Clustering (segmentation) algorithms, pattern recognition models, visualization methods, among others, belong to this family of descriptive models [18,28].

In this study, the researcher used descriptive data mining technique because the objective of this study is to find out the association (pattern recognition models) of major import items and the effects of those items on market price.

E. Evaluation and Interpretation

As it is often the case, with the application of the learning algorithm, several association rules were discovered from the dataset. Considering the large number of discovered rules, it is imperative to select only those rules that are interesting in relation to the purpose of the research.

To accomplish this task, two different types of measure of interestingness have been adopted: objective and subjective measures of interestingness.

The objective measures of interestingness of a rule used in the research include confidence and support measures. While support of an association rule represents the percentage of the records satisfying the body and/or the consequent, Confidence of an association rule represents the percentage of the records satisfying both the body and the consequent to those satisfying only the body [65].

Although the objective measures of interestingness are useful in many respects, they often fail to capture all the complexities of the pattern discovery process [29]. Accordingly, subjective measures of interestingness were also employed. These measures depend mainly on the user who examines the pattern. The use of subjective measures is considered even more important in many data mining application due to the fact that one can discover a large number of rules that are interesting “objectively” but of little interest to the user [59].

Subjective measures of interestingness are classified into actionable and unexpected. According to the unexpectedness measure, a pattern is interesting if it is “surprising” to the user. On the other hand, the action ability measure, which is deemed as the essential subjective measure of interestingness, considers a pattern as interesting if the users can act on it to their advantage. Actionability is an important subjective measure of interestingness because users are mostly interested in the knowledge that permits them to do their jobs better by taking some actions in response to the newly discovered knowledge [59]. Hence, the researcher, together with domain experts, attempted to determine the subjective interestingness of the discovered regularities based on knowledge about the problem domain.

In this research, the researcher used the objective measures of interestingness, because this research studies about the association between major import and export items with the market price and currency rate using association data mining rule.

V. Experiment and Result

The first stage of the data mining process is to select the related data from many available databases to correctly describe a given business task [10]. There are at least three issues to be considered in the data selection. The first issue is to set up a concise and clear description of the problem. The second issue would be to identify the relevant data for the problem description. The third issue is that selected variables for the relevant data should be independent of each other. Variable independence means that the variables do not contain overlapping information.

Using the right data for data mining task is one of the primary keys for successful data mining [28]. For this study, the 12-year initial data were collected from three different government organizations (i.e. ERCA, CSA and NBE). The initial data collected from ERCA contain 704,573 raw of data with a total of 52.58MB of disk space.

Data preparation or preprocessing is always important in a machine learning and pattern recognition process. The purpose of data preparation is to clean the data as much as possible and to put it into a form that is suitable for selected data mining software. Starting from the data extracted from the source database maintained by ERCA, CSA and NBE, a number of transformations were performed before a working dataset was built. The activities during this phase included data cleaning, data selection, attribute or feature selection, transformation and aggregation, integration and formatting, discretization and binarization, dimensionality reductions, minimizing noises, handling missing values.

Accordingly, the data which was in a relational database format stored using different application program was first exported into a single table format of excel sheet. This is mainly because the Weka tool supports a single table data format for processing. Moreover, some attributes were removed since they were irrelevant to the objective of this research, redundant and had no values at all, and 10 features were used for many-sided analysis. Attributes creation through aggregation of attribute values of import export data set.

A. Data Mining Tool Selection

According to Paola Britos and his colleagues [48], a very important problem in the data mining process is detecting too late that the tool selected is inappropriate to do the objective of business. Data mining tool selection is normally initiated after the definition of problem to be solved and the related data mining goals. However, more appropriate tools and techniques can also be selected at the model selection and building phase. Selection of appropriate data mining tools and techniques depends on the main task of the data mining process.

The data mining software selected for this research is Weka (to find interesting patterns in the selected dataset). In addition, the methodologies used by data mining software to perform each of the data mining functions are also an important factor to consider. This means that the algorithms supported by the software should be known. The researcher selected four algorithms: apriori, tertius, PredictiveApriori and FilteredAssociator. These algorithms are used to identify interesting patterns and relationships out of the selected dataset. The researcher therefore compared the performance of those algorithms.

The reason for the selection of Weka data mining software is that it provides a number of data mining functionalities such as classification, clustering, association, attribute selection and visualization. Familiarity, easy and quick access were also the other reasons why this software was selected.

Weka is developed at the University of Waikato in New Zealand. “Weka” stands for the Waikato Environment of Knowledge Analysis. The system is written in Java, an object-oriented programming language that is widely available for all major computer platforms, and it has been tested under Linux, Windows, and Macintosh operating systems. Java allows us to provide a uniform interface to many different learning algorithms, along with methods for pre and post processing and for evaluating the result of learning schemes on any given dataset [65].

Weka includes a variety of tools for pre-processing a dataset, such as attribute selection, attribute filtering and attribute transformation, feeding into a learning scheme, and analyzing the resulting classifier and its performance. Weka is organized in packages that correspond to a directory hierarchy. The important packages of Weka are association, attribute selection, classifiers, clusters, estimators, and filters packages [67]. Moreover, beyond its popularity and frequent update it gets by volunteers, the negligence of cost also skewed our choice to this tool.

B. Loading the Data

In addition to the native ARFF data file format, WEKA has the capability to read in comma separated values ".csv" format files. This is fortunate since many databases or spreadsheet applications can save or export data into flat files in this format. In fact, once loaded into WEKA, the data set can be saved into ARFF format. The researcher loaded the data set into WEKA, performed a series of operations using WEKA's attribute and discretization filters, and then performed association rule mining on the resulting data set. While all of these operations can be performed from the command line, the researcher used the GUI interface for WEKA Explorer.

Weka is a collection of machine learning algorithms for solving real-world data mining problems. It is written in Java and runs on almost any platform. The algorithms can either be applied directly to a dataset or called from your own Java code. Weka computes basic descriptive, statistics and also other data types. It has good graphic data property visualizing mechanism as we can see below.

The following figures show visualization of dataset for all attributes.

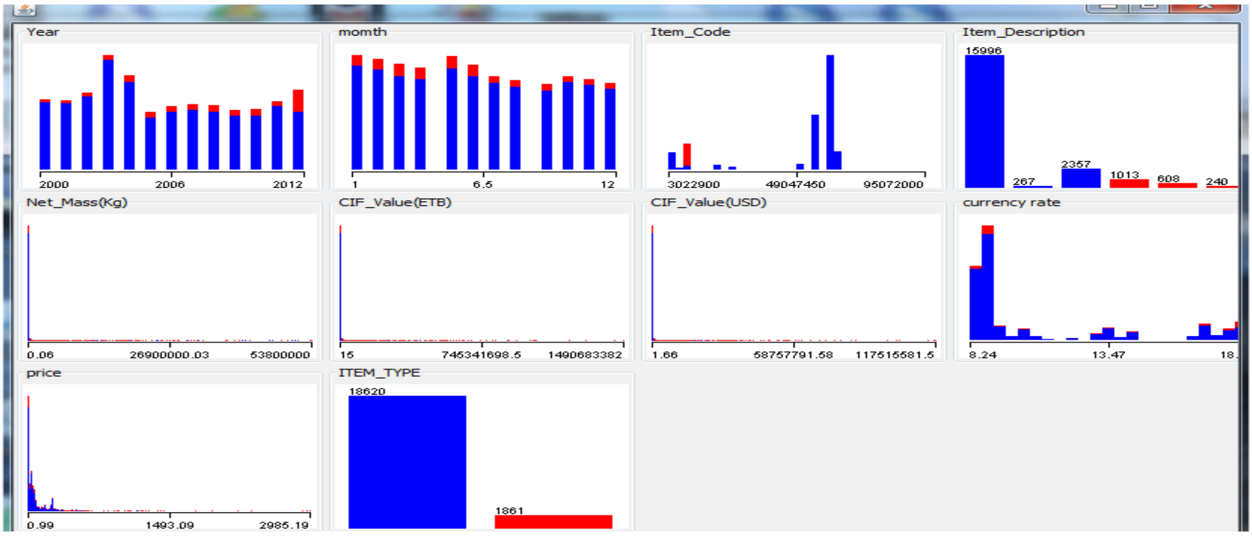

Figure 1: Visualized data for all attributes.

C. Model Building

The first task of the experiment was to understand and heuristically identify attributes or features related to the goal of the machine-learning task, which would be evaluated obviously by the machine learning process through attribute selection. Feature or attribute selection is deciding on the data to be used for analysis. The criteria for this include relevance to the data mining goals, quality and technical constraints such as limits on data volume or data types. It particularly covers selection of attributes (columns) in a table. The reason for selecting features pertain to the time it takes to build a model increases with the number of variables, and blindly including extraneous columns can lead to incorrect models [28]. Thus, given the data-mining task mentioned above, 10 attributes, which were identified as being relevant, were selected. Descriptions of the attributes are presented in Table 1.

D. Result

The results of the experiments confirmed that the techniques of data mining are applicable to generate knowledge from import and export items data in custom administration. The result indicates that more than 78% is imported textile. In addition, this imported textile is highly associated with price and currency rate. This indicates that imported textile is an item that is highly associated with the market price and by extension currency exchange rate in Ethiopia. Next to textile, food is an item that is associated with market price and currency rate. Besides, from the four association rule algorithms, Apriori is the fastest algorithm and Tertius is the slowest one, when compared with all other algorithms used in this study.

Table 1: List and description of selected attributes

Attribute Name	Data Type	Description	Remark
Month	Number	Export or Import month for each item.	
Year	Number	Export or Import Year for each item.	
Price	Currency	The price of each item in the market	Initially, it was named UNIT PRICE
Item_Code	Number	Code of the import or export items	Initially it was named HS_CODE
Item_Description	Categorical or nominal	Description of the import or export items	Initially, it was named HS_DESCRIPTION
Currency rate	Number	Monthly average currency exchange rate	Initially, it was named AVERAGE WEIGHTED RATE
CIF_Value (ETB)	Number	The total cost of imported or exported items in Ethiopian Birr	
CIF_Value (USD)	Number	The total cost of imported or exported items in US dollar	
ITEM_TYPE	Categorical or nominal	Show either item is imported or exported item.	Derived
Net_Mass (Kg)	Number	The net weight of the items	

VI. Conclusions and Recommendations

This study attempted to examine the possible application of data mining techniques, and especially association rule, to identify the association between import and export items and market price in Ethiopia. The study was conducted in five major phases; namely, problem domain understanding, data understanding, data preparation, model building, and evaluation. However, since data mining task is an iterative process, these steps were not followed strictly in linear order, but through spiral approach where significant improvement was achieved.

The result of the experiment shows that imported textile is an item highly associated with market price and currency exchange rate. The domain experts in custom administration should properly handle this item. This means that they should properly follow up this item in order to minimize its immediate implication on market price. The policy makers should also find solutions such as manufacturing those textiles in Ethiopia to replace the imported ones or address the issue of tax avoidance to control the market price and currency exchange rate inflation.

The writers of this paper used different association rule algorithms: Apriori, Tertius, PredictiveApriori and FliteredApriori to construct model. From these algorithms, Apriori and FliteredApriori generated similar model with almost equal time. The time it took to generate 100 rules by Apriori was 2.075 seconds, and that of FliteredApriori was 2.90 seconds. However, both Tertius and PredictiveApriori generated different model than Apriori and FliteredApriori. Tertius algorithm was the slowest algorithm when compared with all other algorithms that the researcher used in this research. It took more than an hour to generate 10 rules, and most of the generated rules were two itemset rules. PredictiveApriori generated the rule that only associated numerical data; it was slower than both Apriori and FliteredApriori algorithms. PredictiveApriori took 5 minutes to generate 100 rules. All the above time was taken by the system with core(TM) 2 Duo @ 2.00GHZ CPU speed, 2GB RAM, running on 32 bit windows7 operating system. However, given the fact that there was treatment in the data in the process of normalization (such as converting random values of a year into monthly based), the result might have been affected by the changes.

In general, the results from this study were encouraging. It was possible to implement data mining techniques on the import and export items of custom administration data and market price. It is the authors' belief that a more thorough study using data mining techniques can help to understand the association between import and export items and market price in Ethiopia. The results of this experiment could be employed as an input for the decision making process for decision makers in custom administration and consumer market associations.

References

[1] Alaaeldin M. Hafez, "Knowledge Discovery in Databases Alexandria University." Department of Computer Science and Automatic Control Faculty of Engineering Alexandria University, 2008.

[2] A. Larson and P. Sheth, "Federated Database Systems for Managing Distributed, Heterogeneous, and Autonomous Databases." Published by ACM Computing Surveys (CSUR), 1990.

[3] A. Savasere, E. Omiecinski and S. Navathe. "An Efficient Algorithm for Mining Association Rules in large Databases," Proc. 21st Int'l, 1995, Conf. of Very Large Data Bases.

[4] A. Silberschatz and A. Tuzhilin, "What makes patterns interesting in knowledge discovery systems." IEEE Trans. on Knowledge and Data Eng., 8, 6, 1996.

[5] Baeza-Yates, R. and Ribeiro-Neto, B." Modern Information Retrieval." Harlow, England: Addison Wesley, 1999.

[6] Berry, M and Linoff, G. "Mastering Data Mining: The Art of Science of Customer Relationship

Management.” John Willy and Sons Inc, New York, 2000.

[7] Bigus, J.P. “Data Mining with Neural networks in Solving Business Problems-From Application to Development to Decision Support.” New York: McGraw-Hill, 1996.

[8] Chris Clifton,GaryGengo, “Developing custom intrusion detection filters using data mining.” MILCOM. 21st Century Military Communications Conference Proceedings, 2000.

[9] Chris Rygielski, Jyun-Cheng Wang and David C. Yen.”Data mining techniques for customer relationship management.”Technology in Society 24, Elsevier Science Ltd, 2000.

[10] David L. Olson and Dursun Delen. “Advanced Data Mining Techniques.” Verlag Berlin Heidelberg: Springer, 2008.

[11] Deogun, Jitender S. (2001) “Data Mining: research Trends, Challenges, and Applications.” [On-line] Available: http://citeseer.nj.nec.com/deogun97data.html,

[12] D.Hawkins, “Identification of outliers”. London, Chapman and Hall, 1980.

[13] D Pyle, “Data Preparation for Data Mining” Morgan Kaufmann Inc. ISBN 1-55860-529-0, 1999.

[14] D Romano, “Data Mining Leading Edge: Insurance & Banking.” in Proceedings of Knowledge Discovery and Data Mining, Unicom, Brunel University, 1997

[15] ER Grilli, “Primary commodity prices, manufactured goods prices, and the terms of trade of developing countries”. World Bank Economic Review, 1988, vol. 2, issue 1, 1-47. Date: 1988.

[16] Fayyad, U. M., G. P. Shapiro, P. Smyth, “From Data Mining to Knowledge Discovery in Databases”, 0738-4602-1996, AI Magazine (Fall 1996): 37–53, 1996.

[17] Fernando Crespoa, Richard Weberb , “A methodology for dynamic data mining based on fuzzy clustering”, Fuzzy Sets and Systems Volume 150, Issue 2, Pages 267-284 ,1 March 2005.

[18] Han, Jiawei and Kamber, Micheline (2001). "Data Mining: Concepts and Techniques and Applications“. San Fransisico; Morgan Kufman Publishers.

[19] Hailegiorgis Biramo Allaro, “The Impact of Trade Liberalization on the Ethiopia's Trade Balance American”. Journal of Economics 2(5): 75-81 DOI: 10.5923/j.economics.20120205.02, 2012.

[20] Han,G.Karypis, and Kumar “Scalable parallel data mining for association rule”. IEEE transactions on knowledge and data engineering, 2000.

[21] Forrester Wave “Predictive Analytics and Data Mining Solutions”, Whitepaper, [on line]

http://www.rosebt.com/1/post/2012/06/predictive-analytics-and-data-mining-wave-2010.html,2010.

[22] H. Liu, F. Hussain, C. Lim, M. Dash." Discretization: An Enabling Technique". Data Mining and Knowledge Discovery 6:4 393-423, 2002.

[23] James D. Fearon, "Primary Commodity Exports and Civil War", Journal of Conflict Resolution (J CONFLICT RESOLUT) Department of Political Science. Stanford University.2013.

[24] Jerome H. Friedman, "Data Mining and Statistics: What's the Connection?" Int'l Conf. Scientific and Statistical Database Management. IEEE Computer Society. Friedman, Jerome H. 1998.

[25] Jiawei Han, Micheline Kamber, "Data Mining: Concepts and Techniques", Morgan Kaufmann Publishers, Champaign: CS497JH, fall 2001.

[26] Justin Zake, "Customs Administration Reform and Modernization in Anglophone Africa—Early 1990s to Mid-2010", [on line] https://www.imf.org/.../Customs-Administration-Reform-and-Modernization-in-anglo, August 1, 2011.

[27] Jordi Galí and Tommaso Monacelli, "Monetary Policy and Exchange Rate Volatility in a Small Open Economy", IGIER, Universit`a Bocconi and CEPR. First version received November 2002; final version accepted October 2004 (Eds.).

[28] Laporte, "Risk management systems: using data mining in developing countries customs administrations", World Customs Journal, 2011.

[29] Levin, Nissan and Zahavi, Jacob, "Data Mining", Tel Aviv University. December 1999.

[30] Li Yan-hai, Sun lin-yan, "Study and Applications of Data Mining to the Structure Risk Analysis of Customs Declaration cargo", e-Business Engineering, ICEBE IEEE International Conference on 2005.

[31] L. Breiman, J.H. Friedman, R.A.Olshen, and C.J. Stone, "Classification and Regression Trees". Wadsworth, Belmont, 1997.

[32] Maimon, Oded, kandel, Abe and Last, Mark. "Information Theoretic Fuzzy Approach to Knowledge Discovery in Databases". Available: http://www.csee.usf.edu/~mlast/papers/wsc_f2.pdf, 2002.

[33] Maria Halkidi, "Quality assessment and Uncertainty Handling in Data Mining Process" Available: http://www.edbt2000.unikonstanz. de/phd-workshop/papers/Halkidi.pdf, 2000.

[34] Mark Brown, John Brocklebank, "data mining", SAS Institute Inc., 1998.

[35] Maurizio Lenzerini, "Data Integration: A Theoretical Perspective" Dipartimento di Informatica e Sistemistica Universit `a di Roma "La Sapienza" Via Salaria 113, I00198 Roma, Italy, 2002.

[36] MJA Berry, John Willey & Sons," Data mining techniques: for marketing, sales, and customer support" ISBN: 978-0-471-17980-1. Jun 1997.

[37] M. O. Mansur, Mohd. Noor Md. Sap,"Outlier Detection Technique in Data Mining:" A Research Perspective, 2005.

[38] M. S. Chen, J. Han, P. S. Yu. , "Data mining: An overview from a database perspective." IEEE Trans. on Knowledge and Data Eng., 8, 6, pp. 866–884, 1996.

[39] Moghadam, M. B. and Rohanizadeh, S. S., "A Proposed Data Mining Methodology and its Application to Industrial Procedures". Journal of Industrial Engineering 4 pp 37-50., 2009.

[40] Narasimha Raju Alluri, "Evaluation of data mining methods to support data warehouse administration and montoring in SAP business warehouse", Faculty of Business Applications of Computer Science. University of applied sciences - furtwangen. 2005.

[41] Óscar Marbán, Gonzalo Mariscal and Javier Segovia "A Data Mining & Knowledge Discovery Process Model. In Data Mining and Knowledge Discovery in Real Life Applications" Book edited by: Julio Ponce and Adem Karahoca, ISBN 978-3-902613-53-0, pp. 438-453, I-Tech, Vienna, Austria, February 2009.

[42] Padhraic Smyth, "Breaking Out of the Black-Box: Research Challenges in Data Mining". June 2001.

[43] Palous, Jiri (Nd), "Machine Learning and Data Mining. Prague: Gerstner Laboratory for Intelligent Decision Making and Control" .Czech Technical University. Available: http://citeseer.nj.nec.com/506615.html, 2001.

[44] Paul R. Krugman, "Strategic trade policy and the new International economics", Available:https://econpapers.repec.org/bookchap/mtptitles/0262610450.htm 2008.

[45] Paul Dorosh and Hashim Ahmed, "Foreign Exchange Rationing, Wheat Markets and Food Security in Ethiopia, Development Strategy and Governance Division," International Food Policy Research Institute – Ethiopia Strategy Support Program 2, Ethiopia, October 2009.

[46] Philip K. chan,"Distributed data mining in credit card fraud detection", Submitted to IEEE Intelligent Systems' Special Issue on Data Mining, 1999.

[47] P. Smyth, R. M. Goodman, "Rule induction using information theory." Proc. Intl. Conference on Knowledge Discovery and Data Mining, 159–176, 1991.

[48] Paola Britos et al, "Tool Selection Methodology in Data Mining", Intelligent Systems Laboratory. School of Engineering. University of Buenos Aires.2006.

[49] Raghavan, V.V; Deogun, J.S. and Sever, H. "Knowledge Discovery and Data Mining: Introduction". Journal of American Society for Information Science, Vol. 49, 1998.

[49] Rea, Allan. "Data Mining: an introduction Student Notes". Available: http://www.pcc.qub.ac.uk/tec/courses/datamining/stu_notes/dm_book_1.html 2001.

[50] R. Agrawal, R. Srikant , "Fast algorithms for mining association rules". Proc. 20th Intl. Conference on Very Large Databases, 487–499, 1994.

[51] R. Agrawal, T. Imielinski, A. Swami., "Mining association rules between sets of items in a large database." Proc. ACM SIGMOD Intl. Conference on Management of Data,pp. 207–216, 1993.

[52] R. Agrawal and J.C. Shafer, "Parallel Mining of Association Rules," IEEE Trans. On Knowledge and Data Eng., 8(6):962-969, December 1996.

[53] R. Agrawal and R. Srikant, "Mining Sequential Patterns", In Proc. 11th Intl. Conf. On Data Engineering, Taipi, Taiwan, March 1995.

[54] R. Garner WEKA: "The Waikato Environment for Knowledge Analysis."

Department of Computer Science, University of Waikato, Hamilton. 2005

[55] Refas, Salim, and Thomas Cantens, "Why does cargo spend weeks in African Ports? The case of Doula, Cameroon", Policy Research Working Paper Series 5565, The World Bank, 2011.

[56] Rüdiger Wirth, Jochen Hipp "CRISP-DM: Towards a Standard Process Model for Data Mining", DaimlerChrysler Research & Technology FT3/KL. PO BOX 2360 89013 Ulm, Germany ruediger.wirth@daimlerchrysler.com. Jochen Hipp. Wilhelm-Schickard-Institute, University of Tübingen. Sand 13, 72076 Tübingen, Germany.2000.

[57] Santos-Paulino, A and A. P. Thirlwall., "The Impact of Trade Liberalization on Exports, Imports and the balance of Payments of Developing Countries," Economic Journal, Royal Economic Society, 114(493), F50-F72, 02. 2004.

[58] Sandeep Slave and Mathew Ninan, "Managing Data Quality", Amrita University Publication, 2005.

[59] Silberschhatz, Avi&Tuzhilns, Alexander. "What Makes Patterns Interesting In Knowledge Discovery System". New York University. Available: http://citeseer.nj.nec.com/silberschatz96what.html, 1996.

[60] S. Brin et al., "Dynamic Itemset Counting and Implication Rules for Market Basket Data," Proc. ACM SIGMOD Conf. Management of Data, ACM Press, New York, 1997, pp. 255–264, 1997.

[61] S. B. Kotsiantis, D. Kanellopoulos and P. E. Pintelas, "Data Preprocessing for Supervised Leaning",

International Journal Of Computer Science Volume 1 Number 2 2006 Issn 1306-4428, 2006.

[62]Tobias Scheffer, “Finding Association Rules that Trade Support Optimally Against Confidence”. Humboldt-Universität zu Berlin, Department of Computer Science. Unter den Linden 6, 10099 Berlin, Germany scheffer@informatik.hu-berlin.de. August 9, 2004.

[63] Herbert A. Edelstein. “Introduction to Data Mining and Knowledge Discovery”. Two Crows Corporation, Third Edition, 1999.

[64] Weiyang Lin et al, “Efficient Adaptive support Association rule Mining for Recommender System”, Kluwer Academic Publisher, 2001.

[64] Ian H. Witten and Frank, Eibe, “Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations”, 3rd edition, The Morgan Kaufmann Series in Data Management Systems, 2000.

[65] Wrobel, S. “Inductive logic programming for knowledge discovery in databases. Relational Data Mining”, European Workshop on Logics in Artificial Intelligence JELIA 2006.

[67] X. Wu et al, “Top 10 algorithms in data mining”, IEEE. International, 2008.

[68]Zaki, M., “Generating non-redundant association rules”. Proceedings of the International Conference on Knowledge Discovery and Data Mining, 2005.

[69]ELIAS A. ALI, “The Effect of Depreciation of Birr on Major Export Products of Ethiopia: The Case of Hides and Skins”, economics. Issue Date: 26-Jun-2012. Publisher: aau.2011.

[70]Anonymous, “Uncover gems of information”. SAS Institute. Retrieved from the World Wide Web on January 28, 2002. http://www.sas.com/products/miner/index.html

[71] KDnuggets . Polls,” Data Mining Methodology “, analytics Summit, Aug 2007.

[72] Peter A. Flach Nicolas Lachiche, “Confirmation-Guided Discovery of First-Order Rules with Tertius”, Machine Learning, 2001.

[73] E.M. Knorr and R.T. Ng. “A unified notion of outliers: Properties and computation”. In Proc. KDD 1997, pages 219–222, 1997.

[74] Trybula, Walter J. “Data Mining and Knowledge Discovery” Annual review of Information Science and Technology (ARIST); (32): 197 – 229, 1997.