

Efficient Predictive Model for Determining Critical Factors Affecting Commodity Price: The Case of Coffee in Ethiopian Commodity Exchange (ECX)

Worku Abebe Degife¹

¹ Faculty of Informatics, University of Gondar, Gondar, Ethiopia
Email: workuabebeis@gmail.com, worku.com@uog.edu.et

Dr.ing. Abiot Sinamo (PhD)²

² Ethiopian Institute of Technology-Mekelle (EiT-M), Mekelle University, Ethiopia
Email: Abiotsinamo35@gmail.com

Received: 24 September 2019; Accepted: 25 October 2019; Published: 08 November 2019

Abstract—In this paper, we have focused on the data mining technique on market data to establish meaningful relationships or patterns to determine the determinate critical factors of commodity price. The data is taken from Ethiopia commodity exchange and 18141 data sets were used. The dataset contains all main information. The hybrid methodology is followed to explore the application of data mining on the market dataset. Data cleaning and data transformation were used for preprocessing the data. WEKA 3.8.1 data mining tool, classification algorithms are applied as a means to address the research problem. The classification task was made using J48 decision tree classification algorithms, and different experimentations were conducted. The experiments have been done using pruning and unpruning for all attributes. The developed models were evaluated using the standard metrics of accuracy, ROC area. The most effective model to determine the determinate critical factors for the commodity has an accuracy of 88.35% and this result is a good experiment result.

The output of this study is helpful to support decision-making activities in the area of the Ethiopia Commodity Exchange. The study support commodity suppliers to take care of the determinant factors work towards maintaining quality. Ethiopia Commodity Exchange (ECX), as the main facilitator of commodity exchanges, can also use the model for setting price ranges and regulations.

Index Terms—Commodity Price, Predictive Model, J48 decision tree

I. INTRODUCTION

Modern commodity exchanges dates back to the trading of Rice futures in the 17th century in Osaka, Japan[1]. With the liberalization of agricultural trade in

many countries, and the withdrawal of government support to agricultural producers a new need arises for price discovery and even physical trading mechanisms, a need that can often be met by commodity exchanges. Hence, the rapid creation of new commodity exchanges, and the expansion of existing ones have increased over the past decade.

At present, there are major commodity exchanges globally and a large number of brand new exchanges have been created during the past decade in developing countries[1].

Although the major task of ECX is protecting its customers through the modern trading system, controlling the factors affecting commodity prices should also be taken into consideration for smooth and healthy exchanges between customers. In this regard number of factors can be considered as determinants but there is no a clear indicator that can directly be inferred out of the available commodity exchange data. Therefore, as one means of potential intervention, data mining research can be conducted to unveil the unseen pattern that can potentially be useful for regulating commodity exchange trends.

II. LITERATURE REVIEW

2.1. Data Mining

Data mining is the process to discover interesting knowledge from large amounts of data[3]. Nowadays, data stored in market databases are growing in gradually rapid way. Therefore, it is necessary to analyse this huge amount of data and extract useful information from it. Data mining is the process that results in the discovery of new patterns in large data sets. The goal of the data mining process is to extract knowledge from an existing data set and transform it into a human understandable

formation for advance use. It is the process of analyzing data from different perspectives and summarizing it into useful information. There is no restriction to the type of data that can be analyzed by data mining. We can analyze data contained in a relational database, a data warehouse, a web server log or a simple text file. Analysis of data in effective way requires understanding of appropriate techniques of data mining. [3,4,5,6] In addition to this, data mining technology can generate new business opportunities by providing automated process of finding predictive information in large databases and discovery of previously unknown patterns.

A large data collection is required for producing information. Only data retrieving is not enough, rather we need a means to automate the aggregation of data, information extraction, and recognize discovery patterns in The source data. Files, databases and other repositories consists of huge amount of data, hence it is necessary to develop a prevailing tool for analysis and explanation of data and extracting interesting knowledge to facilitate in decision making. Data mining can solve all of the tasks[7].

Data mining is a method of extracting unknown projecting information from large databases which is a widespread technology that helps organizations to focus on the most important information in data repositories with great potential[7,8,9,10]. Data analysis tools predict future trends and behavior, helping organizations in active business solutions to knowledge driven decisions [6,11,12]. Intelligent data analysis tools produce a database to search for hidden patterns, finding projecting information that may be missed due to beyond experts' prediction. The task of data mining are varied and distinct because there are many patterns in a large database. Deferent kinds of methods and techniques are needed to find deferent kinds of patterns[10,12,13].

2.2. Related Works

The first work in this regard is the work done by Ticlavilca et al [14]. This work applied a MVRVM model to develop multiple-time-ahead predictions with confidence intervals of monthly agricultural commodity prices. The predictions are one, two and three months ahead of prices of cattle, hogs and corn. The MVRVM is a regression tool extension of the RVM model to produce multivariate outputs. The statistical test results indicate an overall good performance of the model for one and two month's prediction for all the commodity prices. The performance decreased for the three-month prediction of the three commodity prices. The MVRVM model outperforms the ANN most of the time with the exception of corn price prediction two and three months ahead. However, the bootstrap histograms of the MVRVM model show narrow confidence bounds in comparison to the histograms of the ANN model for the three commodity price forecasts. Based on this study, the MVRVM is more robust.

From the study in [15], the author discussed, determining the Stock market forecasts has always been challenging work for business analysts. In the paper, the author attempted to make use of these huge chaotic in nature data to predict the stock market indices. Moreover, if we combine both these chaotic data and numeric time series analysis, the accuracy in predictions can be achieved. Investors can use this prediction model to take trading decision by observing market behavior. Enhancements of this system are focused to help in improving more accurate predictability in stock market regardless how chaotic the stock market data can be.

The study of Santoso and Rusdianto [16] resulted with a "Hybrid Clustering Method for Stock Price and Commodity Price". They find out that, combination between K-Means Clustering and Principal Component Analysis can give better analysis and classification. The results of using those two methods are a dimension reduced cluster (compact cluster). The result is supported by some findings in the reality through having some observations in the news and daily reports of the company conditions. In their summary, the study has shown that it is an effective method to use a hybrid method to cluster the stock price and commodity price using K-Means Clustering and Principal Component Analysis. It is supported that the result of the Principal Component Analysis can have a direct impact on the number and type of dietary patterns revealed in the data. They also mentioned that, reducing the dimension of the cluster is an important task and it can be implemented by using Principal Component Analysis. Moreover, for further research works, they suggested to combine the K-Means algorithm, Principal Component Analysis, and Neural Network to have better solutions. By conducting the Neural Network to the established clusters, it will give the exact information on how to identify the information in every cluster.

III. PROBLEM OF THE STATEMENT

Knowing how an industry is influenced by market trends is essential to stay competitive and meeting consumers' needs. In order to keep a company ahead of the competition, it is also important to utilize market trend analysis that is the process of evaluating changes to a given market. As a pioneer and modern marketing platform in Ethiopia [2]. ECX should also determine factors affecting commodities prices. Although the global market plays a significant role in this aspect, there is no a clear cutting rule explicitly list the deterrent factors and indicate the relationships within the factors.

As a researcher knowledge there was no any work that studded determinate factors of the commodities price which one considered in ECX. The study will help the organization (ECX) to take action according to the discovered knowledge and experts and suppliers also use for decision making.

IV. EXPERIMENTATION DESIGN

In this research number of experiments for J48 were conducted and high accuracy values are recorded. A total of 18141 datasets with 7 independent attributes and one dependent variable are used throughout all the experiments. Once the experimental setups are established, building model with a number of parameters that govern the model generation process would be the next task.

Table 1 Sample values of final attributes after pre-processed

Reference	Season	Commodity_Name	Description	Origin	Volume	Warehouse	Closing_Price	
1	One	Winter	Export_Coffee	Rank5	Kaffa	X	Bedele	Low
2	One	Spring	Export_Coffee	Rank5	Wollega	Y	Gimbi	Low
3	One	Spring	Specialty_Coffee	Rank1	Yirgacheffe	Y	Dilla_Havassa	High
4	One	Spring	Specialty_Coffee	Rank2	Gelana_Abaya	X	Dilla_Havassa	High
5	One	Spring	Specialty_Coffee	Rank1	Yirgacheffe	X	Dilla_Havassa	Medium
6	One	Spring	Specialty_Coffee	Rank1	Sidama	X	Dilla_Havassa	High
7	One	Summer	Export_Coffee	Rank5	Harar	W	DireDawa	High
8	One	Summer	Export_Coffee	Rank4	Harar	W	DireDawa	High
9	One	Spring	Export_Coffee	Rank4	Harar	X	DireDawa	High
10	One	Spring	Export_Coffee	Rank5	Sidama	X	Dilla_Havassa	Medium
11	One	Spring	Specialty_Coffee	Rank2	Sidama	W	Dilla_Havassa	High
12	One	Spring	Export_Coffee	Rank5	Kaffa	Y	Bonga	Low
13	One	Autumn	Export_Coffee	Rank3	Sidama	Z	Dilla_Havassa	High
14	Same_year	Autumn	Export_Coffee	Rank4	Harar	X	DireDawa	High
15	One	Spring	Export_Coffee	Rank5	Kaffa	W	Dilla_Havassa	Low
16	One	Winter	Export_Coffee	Rank5	Kaffa	X	Bedele	Medium
17	One	Spring	Export_Coffee	Rank5	Sidama	W	Dilla_Havassa	Low
18	One	Winter	Specialty_Coffee	Rank2	Sidama	X	Dilla_Havassa	Medium
19	Same_year	Winter	Specialty_Coffee	Rank1	Sidama	X	Dilla_Havassa	High
20	One	Summer	Export_Coffee	Rank5	Sidama	W	Dilla_Havassa	High
21	One	Autumn	Export_Coffee	Rank5	Wollega	Y	Gimbi	Low
22	One	Spring	Export_Coffee	Rank3	Harar	W	DireDawa	High
23	One	Winter	Export_Coffee	Rank5	Harar	W	DireDawa	High

4.1. Model Building Using J48 Decision Tree

Model building is an iterative process. Therefore, it is important to conduct different experiments to find the optimal model to address the problem. In this study, different experiments are conducted by altering parameters of the J48 decision tree but only some of the experiments are presented here which score high accuracy [17,118]. (Compare to each other, this scenario is also used for other selected algorithm). J48 algorithm contains some parameters that can be changed to further improve classification accuracy. Initially the classification model is built with the default parameter values of the J48 algorithm. Table 2 summarizes the default parameters with their values for the J48 decision tree algorithm.

Table 2 Some of the J48 algorithm parameters and their default values

Parameter	Description	Default Value
ConfidenceFactor	The confidence factor used to for pruning (smaller values incur more pruning)	0.25
MinNumObj	The minimum number of instances per leaf	2
Unpruned	Whether pruning	False

By changing the different default parameter values of the J48 algorithm, the experimentations of the decision tree model-building phase are approved.

Table 3 Values of parameters used for J48 algorithm

Experiments	Parameters			
	Pruned	Confidence factor	(min Numobj)	Test option
Experiment #1	True	0.25	2	10 fold cross validation
Experiment #2	True	0.5	5	10 fold cross validation
Experiment #3	True	0.5	2	80% percentage split
Experiment #4	False	0.25	2	10 fold cross validation
Experiment #5	False	0.25	5	10 fold cross validation
Experiment #6	False	0.5	5	80% percentage split

The performance measures (in terms of accuracy, ROC, and other effectiveness measures) of the above six experiments of J48 decision tree algorithms are organized in Table 4 below.

Table 4 Experimentation result of J48 Algorithms

Performance measurements	Experiments					
	#1	#2	#3	#4	#5	#6
Accuracy (%)	88.31	88.35	88.00	88.34	88.32	87.95
Numbers of leaves	77	142	197	197	134	134
Size of tree	98	182	253	253	172	172
Time taken	0.08	0.06	0.08	0.08	0.06	0.05
ROC area	0.919	0.969	0.972	0.972	0.970	0.972
CCI	16022	16051	3193	1629	16023	3191
ICCI	2119	2110	435	2112	2118	437

Key: CCI: Correctly classified Instance, ICCI (Incorrectly classified Instance), ROC: Relative Optical character curve.

On presented Table 4 the result of each experiment developed. The experiment was designed to evaluate the performance of a J48 classifier Unpruned and pruned tree. The J48 pruned achieved more accuracy out of the 6 experiments. As the result Experiment #2 (Building decision tree pruned and testing option 10 cross validation) is best based on accuracy which registered 88.35% and correctly classified instance which is accounts 16031 out of 18141. And The Number of leaves and the size of the tree have a value of 142 and 182 respectively.

Therefore, the researcher selected Experiment #2 (pruned J48 decision tree with 10 fold cross validation) for comparison with other classification algorithms out of the 8 experiments.

In addition to this, the researcher attempted to use different parameters out of that listed in the table 2 to increase the accuracy of the model and to minimize the number of leaves and tree for all experiments. With this objective in mind, the MinNumObj (minimum number of objects in a leaf) parameter was tried with a value 10, 15 and 20. The confidence factor also tried with a value 0.6, 0.7... But the result is not much improved when compare it with Experiment#2. This is because if the value of MinNumObj increases, the number of the leaves and size of the tree also decreases but the accuracy and the performance of the model decrease. Consequently, if the value of confidence factor increases, the number of the leaves and size of the tree also increases as well as accuracy and the performance of the model decrease. This is because smaller values incur more pruning.

V. CONCLUSIONS

In this study, DM techniques have been used with the aim of identifying and critical determinate factors for commodities price. The hybrid process model was followed during undertaking experimentation and discussion. The data set used in this study has been taken from ECX market data. After taking the data, it has been preprocessed and prepared in a format suitable for the DM tasks.

The study was conducted using classification techniques namely decision tree, for model building and experimentation J48 Experimentation was conducted using two test options (10-fold cross validation and 80 % percentage split) for each experiment. Various experiments were made by making adjustments on the modeling parameters in order to come up with meaningful results. J48 algorithms performed with accuracy of 88.35%. The results from this study can contribute towards encouraging and supporting the decision making process for marketing organizations and marketers. To conclude, the study showed that the problem of fluctuating the commodities price can be solve using data mining techniques.

VI. RECOMMENDATIONS

This study has provided a predictive model for determining critical factors of commodity price in ECX. Based on the results and findings obtained from the study, the researcher forwards the following recommendations and potential future works.

- The output of this research is helpful to the domain experts of ECX and suppliers for identifying the determinant factors for commodities price. This can help for solving problems rising during trading activity with regard to price fluctuations.

The model which is developed in this research generates various patterns and rules that can be used by ECX branches in deferent location of the country.

Market policy makers, suppliers and planners can have useful insight for future planning and special intervention program based on the findings of the study.

In this study the scope was limited to the commodity namely coffee, further study can be conducted by considering different commodities including local coffee.

This study can be used as an input for the development of a full-fledged model that can be used for price prediction by developing a full-fledged decision support system. Graphical User Interface as a prototype of the proposed model

Other methodology or classification algorithms (neural network and Support Vector Machine) can also be used for further research work.

ACKNOWLEDGEMENT

We would like thank to Ethiopian Commodity Exchange (ECX) higher management for their willingness to access the data and assigned the experts. We thank to Ethiopian Commodity Exchange (ECX) experts for their support and valuable comments.

REFERENCE

- [1] C. J. Santana-Boado, Leonela; Brading, 'Commodity exchanges in a globalized economy', no. September, pp. 1–8, 2000.
- [2] M. A. Hernandez, S. Lemma, and S. Rashid, 'The Ethiopian commodity exchange and the coffee market: Are local prices more integrated to global markets?', no. April 2008, 2015.
- [3] H. Patel and D. Patel, 'A Brief survey of Data Mining Techniques Applied to Agricultural Data', vol. 95, no. 9, pp. 6–8, 2014.
- [4] T. Edition and T. C. Corporation, *Introduction to Data Mining and .*
- [5] M. Gandhi and G. Vishwavidyalaya, 'Data mining Techniques for Predicting Crop Productivity – A review article', vol. 4333, pp. 98–100, 2011.
- [6] G. Marketos, K. Padiaditakis, and Y. Theodoridis, 'Intelligent Stock Market Assistant using Temporal Data Mining', pp. 1–11.
- [7] D. Das and M. S. Uddin, 'T he S Tock M arkeT', vol. 4, no. 1, pp. 117–127, 2013.
- [8] D. V. Setty, 'A Review on Data Mining Applications to the Performance of Stock Marketing', vol. 1, no. 3, pp. 33–43, 2010.
- [9] S. Tiwari and A. Gulati, 'Prediction of Stock Market from Stream Data Time Series Pattern using Neural Network and Decision Tree', vol. 7109, pp. 99–102, 2011.
- [10] S. Dzeroski, 'Relational data mining', in *Data Mining and Knowledge Discovery Handbook*, Springer, 2009, pp. 887–911.
- [11] K. V Nesbitt and S. Barrass, 'Patterns in Stock', pp. 45–55, 2004.
- [12] Niketa Gandhi, Leisa Armstrong, "Applying Data Mining Techniques to predict yield of Rice in Humid Subtropical Climatic Zone of India". 978-9-3805-44212/16/\$31.00_c 2016

- [13] A.T.M Shakil Ahamed, Navid Tanzeem Mahmood, Nazmul Hossain, Mohammad Tanzir Kabir, Kallal Das, Faridur Rahman, Rashedur M Rahman," *Applying Data Mining Techniques to Predict Annual Yield of Major Crops and Recommend Planting Different Crops in Different Districts in Bangladesh*", 15978-1-4799-8676-7/15/\$31.00 copyright 2015 IEEE SNPD 2015, June 1-3 2015,
- [14] A. M. Ticlavilca, D. M. Feuz, and M. Mckee, 'Forecasting Agricultural Commodity Prices Using Multivariate Bayesian Machine Learning Regression by Andres M . Ticlavilca , Dillon M . Feuz , and Mac McKee', 2010.
- [15] M. P. Naeini, H. Taremian, and H. B. Hashemi, 'Stock market value prediction using neural networks', in *Computer Information Systems and Industrial Management Applications (CISIM), 2010 International Conference on*, 2010, pp. 132–136.
- [16] H. Pan, C. Tilakaratne, and J. Yearwood, 'Predicting Australian stock market index using neural networks exploiting dynamical swings and intermarket influences', *J. Res. Pract. Inf. Technol.*, vol. 37, no. 1, pp. 43–56, 2005.
- [17] Monali Paul, Santosh K. Vishwakarma, Ashok Verma," Analysis of Soil Behaviour and Prediction of Crop Yield using Data Mining Approach" ,2015 *International Conference on Computational Intelligence and Communication Networks*, 978-1-5090-0076-0/15 \$31.00 © 2015
- [18] Ramesh A. Medar,Vijay S. Rajpurohit, "A survey on Data Mining Techniques for crop yield prediction", *International Journal of Advance Research in Computer Science and Management Studies*. Volume 2 , Issue 9, Sept 2014.

Authors' Profiles



Worku A. Degife (MSc) lecturer at University of Gondar and also currently he studying his PhD in Taiwan at National Taiwan University of science and Technology, in major of Information Management. Mr. Worku A. received his MSc. in Information Technology from University of Gondar, Gondar, Ethiopia in 2017. His main research fields are data mining, Neural network, machine learning, data mining and knowledge based system, big data management and security, reinforcement learning and AI.



Dr.ing. Abiot Sinamo (PhD), is an Asst. Professor in Assistant Professor of Intelligent Systems and ERP Systems School of Computing Ethiopian Institute of Technology-Mekelle (EiT-M), Mekelle University. Currently, he is Director General, ICT Sector FDRE Ministry of Innovation and Technology His main research fields are, Artificial Intelligence, Business Analytics, Big data Analytics, Green Computing, Engineering, Enterprise Systems, ERP Systems, Very Large Business Applications(VLBAs), Knowledge based Systems, Expert Systems, Intelligent Systems, Robotics, Natural Language Processing, Machine Learning, Image Processing, Computer Vision

How to cite this paper: Worku Abebe Degife, Dr.ing. Abiot Sinamo, " Efficient Predictive Model for Determining Critical Factors Affecting Commodity Price: The Case of Coffee in Ethiopian Commodity Exchange (ECX)", *International Journal of Information Engineering and Electronic Business(IJIEEB)*, Vol.11, No.6, pp. 32-36, 2019. DOI: 10.5815/ijieeb.2019.06.05