2020 International Conference on Computational Performance Evaluation (ComPE)
North-Eastern Hill University, Shillong, Meghalaya, India. Jul 2-4, 2020

Solving Onion Market Instability by Forecasting Onion Price Using Machine Learning Approach

Md. Mehedi Hasan
Computer science and engineering
Daffodil International University
Dhaka, Bangladesh
mehedi15-8421@diu.edu.bd

Muslima Tuz Zahara
Computer science and engineering
Daffodil International University
Dhaka, Bangladesh
muslima15-8329@diu.edu.bd

Md. Mahamudunnobi Sykot
Computer science and engineering
Daffodil International University
Dhaka, Bangladesh
mahamudunnobi15-8266@diu.edu.bd

Rubaiya Hafiz
Computer science and engineering
Daffodil International University
Dhaka, Bangladesh
rubaiya.cse@diu.edu.bd

Mohd. Saifuzzaman
Computer science and engineering
Daffodil International University
Dhaka, Bangladesh
saifuzzaman.cse@diu.edu.bd

***Abstract*— Price is the key factor in financial activities. Unexpected fluctuation in price is the sign of market instability. Nowadays Machine learning provides enormous techniques to forecast price of products to cope up with market instability. In this paper, we look into the application of machine learning approach to forecast the price of onion. The forecast is based on the data collected from Ministry of Agriculture, Bangladesh. For making prediction we used machine learning algorithms e.g. K- Nearest Neighbor (KNN), Naïve Bayes, Decision Tree, Neural Network (NN), Support Vector Machine (SVM). Then we assessed and compared our techniques to find which technique provides the best performance in term of accuracy. We find all of our techniques provide analogous performance. By above mentioned techniques we seek to classify whether the price of onion would be preferable (low), economical (mid), expensive (high).**

***Keywords*— Onion price, Data Analysis, Machine learning, Forecasting, Classification.**

I. INTRODUCTION

Onion is a very essential food in our daily diet. In Bangladesh about 2.5 billion-ton onions are needed for a year. Unexpected changes in onion price has become a crucial concern in the economic environment of Bangladesh. At the end of 2019 Bangladesh experienced a dramatic change in onion price. In the year 2019 onion price was 28 Taka per kg on 1st January but at 17th November it was 228 Taka per kg. This change received a big attention for its unusual behavior. From the price we found that fluctuation range was very high. This price is not bearable to the unprivileged people of Bangladesh. Financial forecasting is intricate due to unstructured nature and uncertainty of data. Besides this the forecasting result is influenced by factors like weather conditions, productivity, storage limitation, transportation, supply-demand ratio making forecasting more intricate.

On this era of artificial intelligence, machines are behaving intelligently. Einav et al. [1] describe machines are used vastly for prediction purpose. Financial area is the most promising sector for applying machine learning. For this reason, we pursue to forecast onion price by applying Machine Learning (ML) approach on our collected data of onion price. Geron [2] showed that for machine learning applications variety of tools are available like Scikit-Learn, Tensorflow, Matplotlib, Pandas, Numpy. For making dataset convenient to use there are different feature selection algorithms which can be used. Making prediction of onion price is not a simple task. We discussed previously the number of factors that influenced the prediction. Machine learning enriches us by giving the best prediction techniques like supervised learning algorithms, unsupervised learning algorithms, reinforcement learning algorithms. As our work is prediction type we picked supervised learning algorithm to make forecasting. The above-mentioned ML techniques can be applied for forecasting onion price. This is the main goal of our work.

This paper is organized as follows. In Section II review of related work is given. In section III proposed methodology is provided. In section IV we find expected result and compare different result that we achieved through different ML techniques to predict onion price. Finally, we conclude our paper with a conclusion containing future scopes.

II. LITERATURE REVIEW

Several studies have conducted in the field of machine learning for forecasting purpose. Enormous number of studies have been done to predict future price of merchandise goods to solve the related instability. These studies have focused on problem and for solving the problem, they applied machine learning techniques. Gabralla et al.[3] worked on forecasting crude oil price by using machine learning based model. The dependency of desired output on information contents of input has been showed in this paper. For features selections three algorithms have been used. Before forecasting dataset is split into training and testing data randomly then made four groups. For the purpose of forecasting IBL , K star , and SMO regression model have been used. Finally, the authors constructed an ensemble model for better prediction. On numerical time series data machine learning approach works fruitfully. So it is a good stand for us to predict onion price by machine learning approach.

To find the demand and supply of the raw material of paper pulpwood in Tamil Nadu, India SVM algorithm have been applied by Anandhi et al.[4] Demand and supply of pulpwood have been calculated by collecting and analyzing fifteen years dataset of pulpwood. After analyzing the dataset the authors found demand-supply of forthcoming year. Supply

978-1-7281-6644-5/20/$31.00 ©2020 IEEE

represented the shortage of pulpwood. Our work will include more algorithm besides SVM and compare the accuracy rate. Tanizaki et al.[5] have done a quite similar work in this aspect using POS data for forecasting customers in a restaurant in Japan. Different Machine learning and Statistical analysis such as Bayesian, Boosted, Decision and Stepwise methods have been used on this work. Data usage rate was 40% to 100%. Getting the highest accuracy rate 91.7% after Bayesian algorithm was implemented in dataset where data using rate was 100%. In contrast we got 98.17% accuracy by Neural Network algorithm where data using rate was 30%.

Sinta et al.[6] have done another equivalent work in the field of prediction to predict price of rice using ensemble K-Nearest Neighbors Method in Indonesia using monthly rice price data of 14 years. The authors of this paper discussed about comparison between K-Nearest Neighbors and ensemble K-Nearest Neighbors and showed KNN is less efficient than ensemble KNN because of time series data. The authors tested the accuracy using variation of training data. After prediction, we compared prediction result with actual result and found it very close to the actual price. In contrast to this work we used multiple algorithms on the dataset comprised by daily onion price data around 730 observations.

For decision function, SVM is very specific type of learning algorithm. This work has done financial market prediction by Huang et al.[7] using SVM. The authors observed 676 pair of data for forecasting stock market movement. Applying different classification method in which combining model performed the better result 75%. Finally, a proposed model come which integration of SVM with other classification method has done. In contrast to this work we have got better accuracy and we verified our accuracy with another machine learning algorithm.

Rafieisakhaei et al. [8] made relationship between oil price and demand-supply changes. The authors sought a numerical solution by building a model including the casual loop and mathematical formula where oil price, oil demand and oil supply was the main variable. They proposed different types of loops where global oil demand, oil price, oil supply were the key factors. Finding oil price demand-supply ratio by regression analysis on the price of historic data. In contrary to this work, we have been drawn our solution by machine learning approach.

Similar work has done on the field of agriculture to predict rice production by Gandhi et al. [9] using SVM approach. Main purpose of this paper was to cope up rice production with increasing population in India. On the methodology data mining, SVM approach, SMO algorithm and confusion matrix were used. Prior to applying algorithms, the author used data mining techniques to assemble data by omitting unwanted data. After that SVM approach was used for making needed function which was used to making labeled training data. Then the result was measured SMO algorithm. Finally, using confusion matrix accuracy of proposed model was calculated. On the contrary to this work, we used machine learning approach to assemble data for finding accuracy.

III. Methodology

The methodology of our work includes data collection, data analysis, algorithm implementation and evaluation.

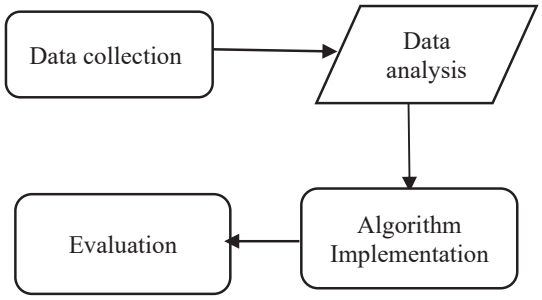

Fig. 1. Research Methodology.

Fig. 1 illustrates the methodology. We achieved our desired result by following stpes of the methodology.

A. Data collection:

We collected our needed data from Ministry of Agriculture website. We collected 730 samples day to day of onion price. These data is numerical, unstructured, and time series data. For working on this data we followed the next steps.

B. Data Analysis:

Before applying algorithms we assembled our data into structured format. We considered year, month, date, season, location, price, and category which is a derived attribute extracted from price as parameters. In our collected dataset data was not in a clean condition as dataset was mixed with unwanted data. We had to remove unuanted data from our dataset before algorithm implementation. For this purpose we used the popular ML tool panda to omit unwanted data for making desired data frame.

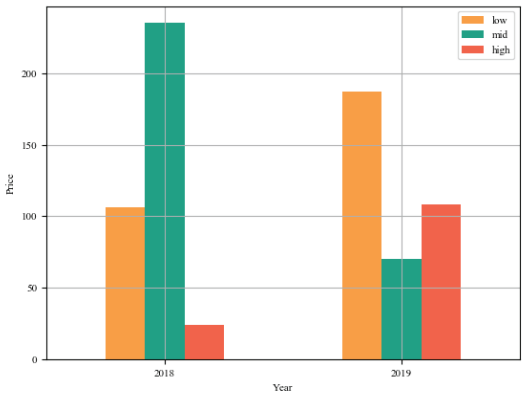

Fig. 2. Yearly price range.

Fig 2: shows price of onion of two years 2018 and 2019 where we plotted year on the x-axis and price on the y-axis. We plotted into the graph as low, mid , high after classifying the price into these three classes. The orange bar represents low price, green represents mid price and red bar represents high price. From the Fig. 2 we can easily determine how many days of a year the price was low, mid and high.

Fig.3. Monthly onion price.

Fig. 3 illustrates line graph representing the month on the x-axis and corresponding price on the y-axis. Fig. 3 shows price fluctuation visually. From this graph we made an insight from the data by data analysis. We noticed when price of onion was high at the beginning of the year, price would be low at the end of the year. Following the beginning of next year price of onion was low and we also noticed at the end of this year the price would be high. We noticed that the previous years also maintained this structure. Considering this structure we took year code as a new parameter that helped us to predict onion price more accurately.

C. *Algorithm implementation:*

At first we classified the onion price into 3 classes and then applied supervised learning algorithm techniques more specificly classification algorithms. Completing the data analysis part now we have the prepared dataset for algorithm implementation. In this portion we applied every algorithm step by step and observed the result of these algorithms. Then we calculated the accuracy rate of these algorithms.

We used different machine learning algorithms for prediciton of onion price. These algorithms are KNN, Naïve Bayes, Decision tree, SVM, and Neural Network.

D. *Evaluation:*

At the beginning of our work we prepared a dataset then we implemented five different ML algorithms. Every algorithms genereated forcasting results which were nearly expected. By surpassing this process on the experimental result section we will work on the result for further observation.

IV. Experimental Result

On the assembled dataset by applying the selected algorithm we calculated the performance of the algorithms and inserted them into the accuracy table for further comparison.

TABLE I. Accuracy Table

Data usage rate	Algorithms				
	KNN	*Naïve Bayes*	*Decision tree*	*SVM*	*NN*
30%	88.58%	68.49%	97.72%	77.17%	98.17%
40%	87.67%	68.84%	95.89%	80.14%	92.47%
50%	83.84%	67.67%	96.16%	74.25%	97.53%
60%	84.47%	67.81%	97.03%	76.94%	97.95%
70%	80.04%	67.71%	95.30%	70.65%	87.28%

Table I shows prediction results containing columns data usage rate which is 30% to 70%, The following columns show the five models with corresponding results. Yellow box hits the highest accuracy of corresponding model. Keller et al.[10] explained advantages of KNN. This algorithm is most widely used machine learning algorithm as it is easy to implement and a suboptimal algorithm. For KNN model we got prediction rate 88.58% when the data usage rate was 30%. Patil et al[11] described how Naïve Bayes makes prediction by using a data compared with each classes of given dataset. This algorithm uses Bayes theory to make classification. By applying Naïve Bayes on our dataset we achieved 68.84% accuracy at 40% data usage rate. Safavian et al.[12] evaluated that Decision tree shows splendid performance when the dataset is complex. We got 97.72% accuracy by using Decision tree algorithm when data usage rate 30%. SVM is the inclinable tool for financial forecasting. Pradhan.[13] explained basic about SVM. SVM gives 80.14% accuracy at 40% data usage rate. For solving classification problem NN is a good choice. KangaraniFarahani et al.[14] explained that NN is mainly used for creating complex prediction function dynamically and NN is more compatible with the denoising data. In our work Neural Network model achieved the highest forecasting rate 98.17% using 30% data that indicated by purple color border box.

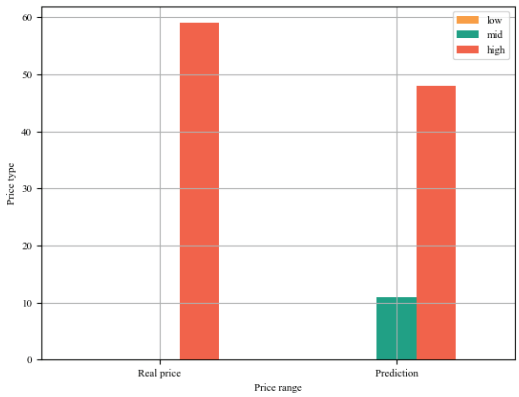

Fig. 4. Comparison between real price and forecasting price.

Fig 4 shows visualization of actual vs predicted price. For comparison we collected around 60 pair of recent (January, February-2020) dataset and plotted the price range in the bar graph. The graph produces high price in its real price part represented by red bar as we have high price rate in every day of January, February 2020. So we expect high price also in our prediction. The prediction part produced almost all high price represented by red bar with a very few medium price

represented by green bar. On the prediction part the high value price which is represented by red bar contains 81.4% where the mid value price which is represented by green contains 18.6%.

V. Conclusion:

In our work each algorithm performs closely. We investigated five algorithms performance and found the best algorithm for a better onion price prediction.

Finally we have a model by which we can expectedly forecast the future price of onion. Depending on this forecasting price we can calculate the demand and supply of onion, as we know demand-supply plays the main role in market equilibrium state.

Now if we can calculate the future demand-supply of onion based on this prediction we can maintain equilibrium state in onion market which will help us to remove onion market instability. The main limitation of our work is unusual behavior of data and low number of records.

VI. Future work:

In future work we will apply more sophisticated algorithm on our dataset. We will increase the number of our data by adding other cities of Bangladesh and make forecast of onion price for all over Bangladesh. Our work can be used for forecasting the optimal price range of any financial product based on related dataset. We will also work on the term linear programming to balance production-demand-supply chain of onion with ML approach to reduce the instability in onion market.

References

[1] L. Einav and J. J. S. Levin, "Economics in the age of big data," vol. 346, no. 6210, p. 1243089, 2014.

[2] A. Géron, *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems*. O'Reilly Media, 2019.

[3] L. A. Gabralla, R. Jammazi, and A. Abraham, "Oil price prediction using ensemble machine learning," in *2013 International Conference on Computing, Electrical and Electronic Engineering (ICCEEE)*, 2013, pp. 674-679: IEEE.

[4] V. Anandhi, R. M. J. I. j. o. f. c. Chezian, and communication, "Support vector regression to forecast the demand and supply of pulpwood," vol. 2, no. 3, p. 266, 2013.

[5] T. Tanizaki, T. Hoshino, T. Shimmura, and T. J. P. C. Takenaka, "Demand forecasting in restaurants using machine learning and statistical analysis," vol. 79, pp. 679-683, 2019.

[6] D. Sinta, H. Wijayanto, and B. J. A. M. S. Sartono, "Ensemble k-nearest neighbors method to predict rice price in Indonesia," vol. 8, no. 160, pp. 7993-8005, 2014.

[7] W. Huang, Y. Nakamori, S.-Y. J. C. Wang, and o. research, "Forecasting stock market movement direction with support vector machine," vol. 32, no. 10, pp. 2513-2522, 2005.

[8] M. Rafieisakhaei, B. Barazandeh, and M. Tarrahi, "Analysis of supply and demand dynamics to predict oil market trends: A case study of 2015 price data," in *SPE/IAEE Hydrocarbon Economics and Evaluation Symposium*, 2016: Society of Petroleum Engineers.

[9] N. Gandhi, L. J. Armstrong, O. Petkar, and A. K. Tripathy, "Rice crop yield prediction in India using support vector machines," in *2016 13th International Joint Conference on Computer Science and Software Engineering (JCSSE)*, 2016, pp. 1-5: IEEE

[10] J. M. Keller, M. R. Gray, J. A. J. I. t. o. s. Givens, man,, and cybernetics, "A fuzzy k-nearest neighbor algorithm," no. 4, pp. 580-585, 1985.

[11] J. J. I. J. o. A. E. R. COE, "Performance comparison of Naïve Bayes and J48 classification algorithms," vol. 7, no. 11, p. 2012, 2012.

[12] S. R. Safavian, D. J. I. t. o. s. Landgrebe, man,, and cybernetics, "A survey of decision tree classifier methodology," vol. 21, no. 3, pp. 660-674, 1991.

[13] A. J. I. J. o. E. T. Pradhan and A. Engineering, "Support vector machine-A survey," vol. 2, no. 8, pp. 82-85, 2012

[14] M. KangaraniFarahani and S. Mehralian, "Comparison between artificial neural network and neuro-fuzzy for gold price prediction," in 2013 13th Iranian Conference on Fuzzy Systems (IFSC), 2013, pp. 1-5: IEEE.