



Data Mining Forecasting Oil and Gas  
Development Company Ltd. Share Prices Using  
Orange

---

Muhammad Farooq Ishaq

EasyChair preprints are intended for rapid dissemination of research results and are integrated with the rest of EasyChair.

June 10, 2020

# DATA MINING FORECASTING Oil and Gas Development Company Ltd. Share Prices Using Orange.

MUHAMMAD FAROOQ ISHAQ

RIPHAHA INTERNATIONAL UNIVERSITY, LAHORE, PAKISTAN

Muhammad.farooq613@gmail.com

## ABSTRACT

Data Mining is one of the emerging technology that is being use in the field of Data Science. In data mining, we can be used multiple algorithms of machine learning. There are many tools available for the purpose of data mining which one of the Orange3. The purpose of this paper is to use Orange3 for the Analysis of Oil and Gas share prices of the stock exchange. In this paper, four algorithms are being used for the purpose of predicting the share prices of OGDCL in the exchange market. The result of both algorithms are compared by each other and then it was found the best result of the algorithm is Naïve Bayes and Neural Network. It was found the most accurate and perfect result gives the Naïve Bayes and Neural Network. For the given Dataset. In the future, I will do different analyses for the purpose of predict accurate share price results.

**Keyword** → Data Mining, Orange3, Forecasting, OGDCL, Share Prices, Algorithms.

## 1. INTRODUCTION

With the volatile growth of data such as, from from terabytes to petabytes, the importance of data mining is increasingly day by day. In this way data mining is the

extraction of data from large dataset, and changing it into more meaningful from with the help of different tools and methods. It can also be distinct as the grouping of data base management system. And computer science.[1] Which is include the data science, artificial intelligence, machine learning etc.

Data mining is also called knowledge detection in database as it is the process of removing important knowledge from the huge amount of data by using the different analysis or used pattern techniques.

The data mining process is multiple steps such as selecting the target data, preprocessing the data, removing extracting the data and applying the different machine learning algorithms and interrupting the accurate and perfect results. [2]

In this way one of the most wide-ranging tool is Orange3, and orange is a collection of python-based module which is written in C++. Which is open source and free of cost for using the machine learning, data mining and artificial intelligence. Orange provides the different algorithms and techniques for the purpose of finding the different results in different fields. In orange different algorithms categories in different model. It required a specific extension dataset to find the specific and appropriate targets.[3]

## 2. LITERATURE REVIEW

A lot of research and practical work had been done on the data mining. Dataset obtain by stock exchange market websites. Different algorithms provide the ORANGE tool. In this section summarized the previous work that is done in the field of data mining. Four different tools are being used in this research paper for the purpose of predict or forecasting the price of stock exchange in Pakistan.[4]

One open source data mining tool being used to apply the multiple machine learning algorithms. In orange different algorithms available like KNN, SVM, NAÏVE BAYES, NEURAL NETWORK, LOGISTIC REGRESSION, LINEAR REGRESSION, and TREE etc. These models being used for the different purpose of finding the results.[5]

Data mining techniques can also be applied in education system in order to prediction the results of students and predict the drop out rat of the students, predict the students' performance. We can analyses the students data for the purpose research. Data mining is a wide range field because in future data mining is necessarily for everyone. Data mining is need of business. In data mining being used the orange were evaluate the prediction, test and score, confusion matrix, calibration curve.

The survey of literature and classification of data mining articles from 2015 to 2018, done in this paper, which is allows us to regulate the number of data mining techniques that is established with the passage of time. [2] However, data mining is applied in different fields like Bio, Bioscience, Biochemistry, Earth, Math, Physics, Chemistry, Medical, Social, Planetary, Accounting, Decision science, Health, Finance, Chemical science, Economics, astronomy, nursing, energy, agriculture, Biological, Engineering, veterinary, pharmacy.[6]

Data mining is used in classification like gender, birds, flowers, and datamining is being used for the prediction of diseased liked breast cancer, etc. However, we can predict the share prices of stock exchange of oil and gas company limited Pakistan.

### **3. RESEARCH QUESTION**

In this paper were faced as well as many problems .Firs of all how to collect data for dataset. I researched the dataset for research of this paper from many websites but no found at last I were found the data in the stock exchange market websites. Which were the algorithms use for prediction of OGDCL. How to predict and which factors are affected in Prediction. How to study for research write a paper.

### **4. METHODOLOGY**

#### **A. Orange3**

Orange is collection of OOP-Based module that sit over the core library C++ Objects. It is an open source and free of cost for each user. Orange widget is a GUI (Graphical User Interface) for data mining and machine learning. By using orange no need for any kind of programming language. That is very easy to apply in machine learning algorithms. In

orange provides add-ons for neural network, time series, bioinformatic, text mining, educational, Image Analytics, prototypes, text able, etc.[7]

Orange provides different algorithms for data mining , classification, time series, forecasting, predicting, etc. However, this paper deals the four algorithms and in this purpose Orange 3.24.0

## **B. ORANGE**

### **Forecasting Algorithms**

This Paper is providing the Relative analysis of four algorithms. The forecast environment of orange allows the user to make the predicting the

share price of stock exchange market of OGDCL.

- 1) Prediction
- 2) Test and Score
- 3) Confusion Matrix
- 4) Calibration Plot

## **C. OGDCL Stock Exchange Dataset**

OGDCL (Oil and Gas Development Company Limited) is one of the top company which provide the Oil and Gas in Pakistan (PSX) as well as the London exchange Stock Exchange market. Data were taken by (PSX) website that is 1<sup>st</sup> February 2020 to 30<sup>th</sup> April . This data has been a sample data and were prepared for predict the price in orange .

## **5. Results and Discussion**

The data is loaded in orange3 by File in .csv format. Following are the attributes being used in the datasets.

- i. OGDCL
- ii. Date
- iii. Open
- iv. High
- v. Low
- vi. Close

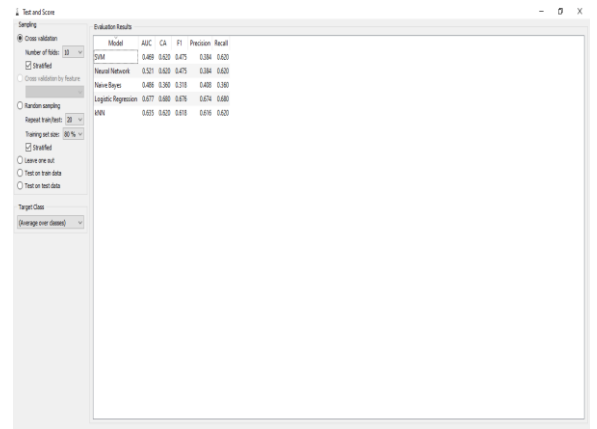
These attributes open, high, low, close, prices of OGDCL has been taken in future prediction

without holidays. In this way the Minimum Value, Maximum Value, Shows in the table. Four models were compared it was found The best algorithm is Neural network which gives the best classification accuracy is 0.878 and and NAÏVE BAYES accuracy is 0.796 other models are less results.

Table 1: Minimum, Maximum, Mean, StdDev Values of attributes.

	Minimum	Maximum	Mean	StdDev
Open	76	137.25	106.132	19.295
High	76	137.5	108.143	18.935
Low	73.25	133.2	103.815	19.171
Close	75.01	134.83	105.648	18.844

Figure 2: Shows the Test Datasets Analyses the Attributes.



Figures 3 Shows The Test Results

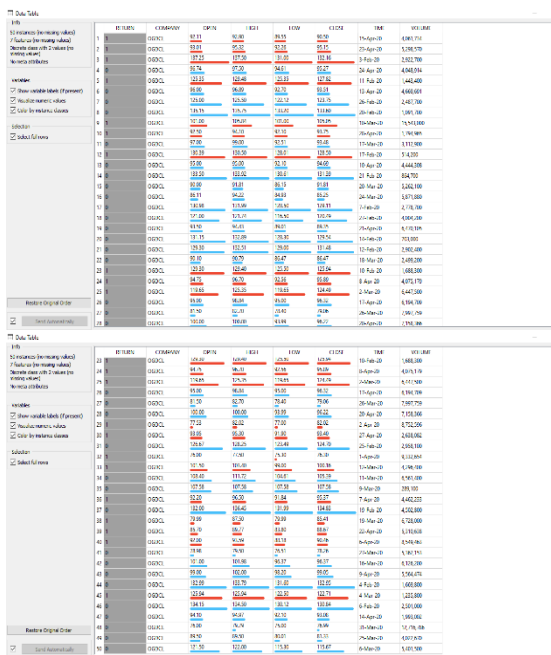


Figure 1: Shows the Train Datasets Analyses the Attributes.

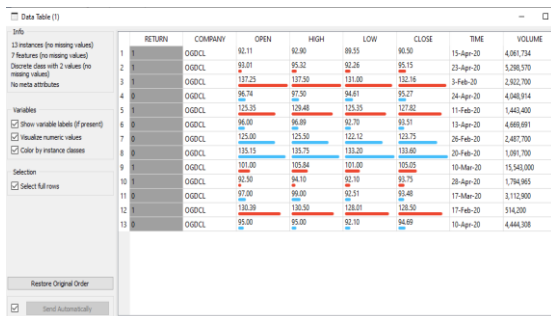
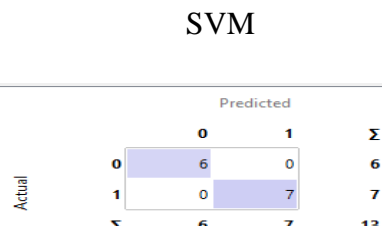
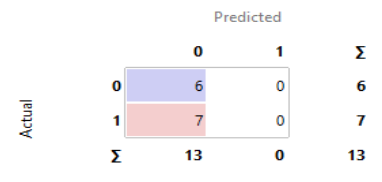
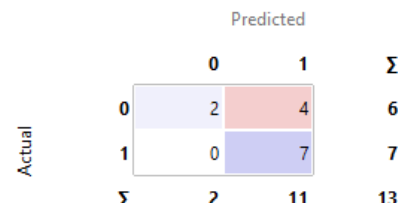


Figure 4: Shows The Prediction Results of OGDCL



### Neural Networks



### Naïve Bayes

		Predicted		
		0	1	$\Sigma$
Actual	0	6	0	6
	1	2	5	7
$\Sigma$		8	5	13

### Logistic Regression

		Predicted		
		0	1	$\Sigma$
Actual	0	4	2	6
	1	3	4	7
$\Sigma$		7	6	13

### KNN

Model	AUC	CA	F1	Precision	Recall
SVM	0.80	0.82	0.79	0.73	0.82
NeuralNetwork	1.00	1.00	1.00	1.00	1.00
NaiveBayes	1.00	0.82	0.81	0.84	0.82
KNN	0.78	0.81	0.81	0.83	0.81
Logistic Regression	1.00	0.88	0.84	0.83	0.88

Figure 5: Shows the Forecasting OGDCL Share price

Thus, after the analysis and comparison of Naïve Bayes and Neural Network models, it was found that Neural Network is a better model for forecasting the values of open, high, low and close prices of OGDCL stock data. The values predicted by Neural Network were found to be more accurate, when compared with actual prices of next ten days. Also, Neural Network had the lower error rate than the Remaining algorithm, as shown in

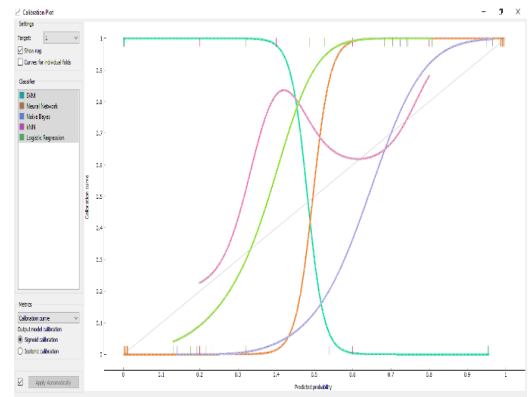


Figure 6: Shows the Future Predicted Probability.

## 6. CONCLUSION AND FUTURE WORK

Although there are many financial, economic and fundamental factors involved in the prediction of stock exchange share prices analysis, but technical analysis and prediction can be done by applying the machine learning algorithms by using orange3.

According to the analysis done in this paper. Neural Network Can be used for the forecasting the future prediction of stock exchange datasets and the that factors which is affected on that results.[8]

However, In future other predictions may be used on the this example of datasets like daily, weekly, monthly, to find it there is any other better option available. And in future others product predictions may be use on this sample of datasets. So, that's simple and unique work is possible for future.

### References

1. Covell, D.G., *Data Mining Approaches for Genomic Biomarker Development: Applications Using Drug Screening Data from the Cancer Genome Project and the Cancer Cell Line Encyclopedia*. PLoS One, 2015. **10**(7): p. e0127433.
2. Shen, D., et al., *Machine Learning and Data Mining in Medical Imaging*. IEEE J Biomed Health Inform, 2015. **19**(5): p. 1587-8.

3. Wan, J., et al., *Data Mining and Expression Analysis of Differential lncRNA ADAMTS9-AS1 in Prostate Cancer*. *Front Genet*, 2019. **10**: p. 1377.
4. Bae, S.M., S.A. Lee, and S.H. Lee, *Prediction by data mining, of suicide attempts in Korean adolescents: a national study*. *Neuropsychiatr Dis Treat*, 2015. **11**: p. 2367-75.
5. Mani, A., et al., *Data mining strategies to improve multiplex microbead immunoassay tolerance in a mouse model of infectious diseases*. *PLoS One*, 2015. **10**(1): p. e0116262.
6. Kim, Y.M., *Comparing Medical Comorbidities Between Opioid and Cocaine Users: A Data Mining Approach*. *Addict Health*, 2019. **11**(4): p. 223-233.
7. Wu, C.T., et al., *Applying Data Mining Techniques for Predicting Prognosis in Patients with Rheumatoid Arthritis*. *Healthcare (Basel)*, 2020. **8**(2).
8. Shouval, R., et al., *Prediction of Allogeneic Hematopoietic Stem-Cell Transplantation Mortality 100 Days After Transplantation Using a Machine Learning Algorithm: A European Group for Blood and Marrow Transplantation Acute Leukemia Working Party Retrospective Data Mining Study*. *J Clin Oncol*, 2015. **33**(28): p. 3144-51.