International Journal of Innovations & Advancement in Computer Science
IJIACS
ISSN 2347 – 8616
Volume 4, Issue 9
September 2015

Forecasting of Aluminium Prices using Data Mining Techniques

Viraj Gada
D.J Sanghvi COE
Vile-Parle(West)
Mumbai,India

Apoorva Dhakras
D.J Sanghvi COE
Vile-Parle(West)
Mumbai,India

Khushali Deulkar
D.J Sanghvi COE
Vile-Parle(West)
Mumbai,India

ABSTRACT

Data mining is a broad topic involving concepts of artificial intelligence, machine learning, statistics and database systems. Currently, there has been an extensive research into the prospects of data prediction for future using data mining algorithms and tools. Much of the existing work has been used to predict the prices of essential commodities and their future sales. However, a prediction into the key factors determining the future sales or prices has been limited. Aluminium is an essential commodity which is used as raw material for a wide range of products and heavy machineries. However, there has been a limited research into the factors which determine the trends and reversal of aluminium prices. As a result, it has become very difficult to estimate the prices of aluminium in the face of limited research on the factors driving its prices. Open Interest of any commodity is one such factor which plays a major role in determining the bearish/bull run of a commodity in the market. The focus of this paper is to conduct experiments on prediction of open interest values using three best algorithms which are best suitable to the data set available. Using the experimental results obtained by training data on weka software, we have attempted to suggest the best algorithm for open interest value prediction for aluminium. The proposed algorithm can potentially be used to provide an estimate for the future prices of aluminium and can revolutionize the industry.

Keywords

Aluminium, Open Interest, Multilayer Perceptron, Gaussian Process, Decision Stump,Volume, Data Mining.

INTRODUCTION

Aluminum is the chief raw material that amounts to value greater than $148 billion to the U.S. economy in the form of aluminium industry, nearly 1 percent of GDP. It is a rare materials ,products of which are used in almost every household in the country, and more than 670 thousand American jobs are provided by this industry. The aluminum industry's technology and scientific advances have helped spur the growth of U.S businesses whilst also helping them compete at par with other leading producers . From construction stuff to packaging of food, new mixtures of elements to high-end appliances, the aluminum industry provides its users with the technology necessary to build and develop hundreds of industries. However, in recent times, there has been an increasing uncertainty in the prices of aluminium. This is a worrying prospect as the fluctuating prices of Aluminium can lead to huge losses to industry in terms of production. It is with this observation that we need an algorithm that can predict the prices of aluminium in market with increased accuracy. In this paper, we have successfully implemented and predicted the prices of aluminium by using three best algorithms to train data and successfully predict the prices of Aluminium. They are decision stump, multilayer perceptron and gaussian process algorithms. These algorithms predict the future prices of aluminium with varying accuracy. Taking the root relative squared error into accont, we have also tried to find out the most accurate of the three algorithms. The implementation of this algorithm will help in predicting the future prices of aluminium and thus help the industry to estimate its manufacturing potential.

DATA MINING TECHNIQUES

In this paper, we have implemented three data mining techniques with an aim to obtain the best possible results with minimum errors. The techniques used are decision stump, multilayer perceptron and gaussian process algorithms.The explanation of these algorithms along with their results has been done. The reason these three algorithms have been selected is that the results obtained via these algorithms were most pertinent

International Journal of Innovations & Advancement in Computer Science
IJIACS
ISSN 2347 – 8616
Volume 4, Issue 9
September 2015

and accurate for the data presented. This algorithms have also been implemented on a trial basis in the market but have yet lacked the results to back their efficiency. Here, however, we have presented this data with the results obtained from experimentation thus providing proof of their accuracy and efficacy.

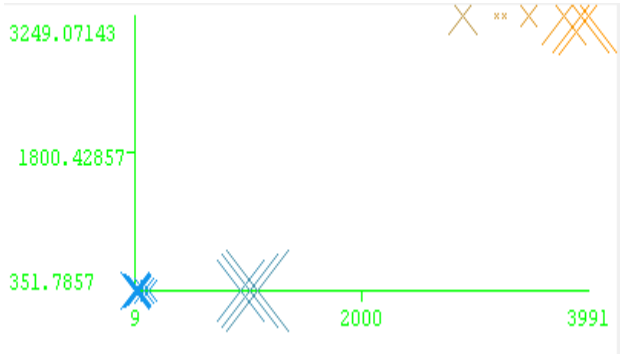
DECISION STUMP ALGORITHM

The term "decision stump" was coined in a 1992 ICML paper by Wayne Iba and Pat Langley.[1] A decision stump is also known as a weak learner ensemble i.e many weak learners combine to form a better learner. Decision stump is also known as 1-R algorithm. The decision stump is about creating a decsion tree with one internal node (the root) which is immediately connected to the terminal nodes (its leaves). It's a very simple algorithm and is eventually used with bagging and boosting algorithms. This machine learning model involves taking data points and then classifying them In the current implementation, we have considered the four important factors which play a key role in determining the prices of aluminium:[2]

- Open - Orders that are entered by investors but not transacted are deemed to be open until they expire or are filled.
- High - Upper range of prices at which aluminium is available.
- Low - lower range of prices at which aluminium is available.
- Volume -In an entire market there are a lot of shares and contracts which are traded in a particular duration as securities.These shares are the amount that is transferred from sellers to buyers as monitoring of activity. Hence if a buyer buys 2000 units from seller, then for that duration based on the transaction the volume will increase by 2000 units.
- Open Interest- On a particular day, there will always be some amount of options and/or future contracts that will not be traded. These are called Open Interest

To obtain results out of the given dataset, around 66% was used as a training set and the remaining 33% as a test dataset.

Fig 1: Graph Forecasting Open Interest Values using Decision Stump Algorithm



X=Open Interest

Y=Predicted Open Interest

Colour variations from blue to orange varies as the value of open interest increases

From these results obtained from Figure 1 and table 1, it is clear that the algorithm was able to predict data with a root relative squared error of 27.1752 %. Also, the graph estimates the future values of aluminium based on these factors.

Multilayer Perceptron Algorithm

When data flows in one direction from input to output layer, then that neural network is called feedforward neural network. Multi Layer perceptron (MLP) is a type of feedforward neural network which has one or greater layers that can exist between input and output layer. Backpropagation learning algorithm is used to train this type of algorithm. The uses of MLP are for classification of patterns, recognition, prediction approximation,etc.[3]

Algorithm

1.Initialize weights and choose a learning rate η

2. Until network is trained, do the following:

For each training example(input pattern and target outputs):

Do forward pass through net(with fixed weights) to produce outputs assuming J hidden layer nodes and N inputs for a 2-layer MLP:

$$y = f\left(\sum_{j=0}^{J} w_{jk}\, O_j\right)$$

International Journal of Innovations & Advancement in Computer Science
IJIACS
ISSN 2347 – 8616
Volume 4, Issue 9
September 2015

where Oj is output from each hidden node j:

$$O_j = f\left(\sum_{i=0}^{N} w_{ij}\, x_i\right)$$

For each output unit k, compute deltas:

$$\delta_k = (y_{target} - y_k) y_k (1 - y_k)$$

For hidden units j(from last to first hidden layer , for the case of more than 1 hidden layer) compute deltas:

$$\delta_j = Oj(1 - Oj)\left(\sum_k w_{jk}\ \delta_k\right)$$

For all weights, change weight by gradient descent,

$$\Delta w_{ij} = \eta \delta_j y_i$$

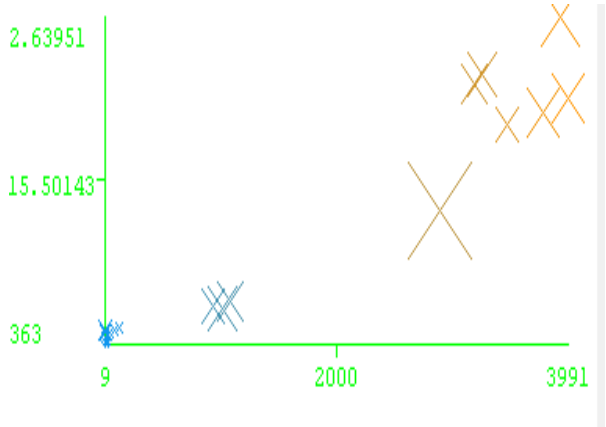
Specifically, for 2- layer MLP, for weight from input layer unit i to hidden layer unit j, the weight changes by

$$\Delta w_{ij} = \eta \delta_j x_i$$

And for weight from hidden layer unit j to output layer unit k, weight changes by

$$\Delta w_{jk} = \eta \delta_j o_j$$

Fig 2: Graph Forecasting Open Interest Values using Multilayer Perceptron Algorithm



X=Open Interest

Y=Predicted Open Interest

Colour variations from blue to orange varies as the value of open interest increases

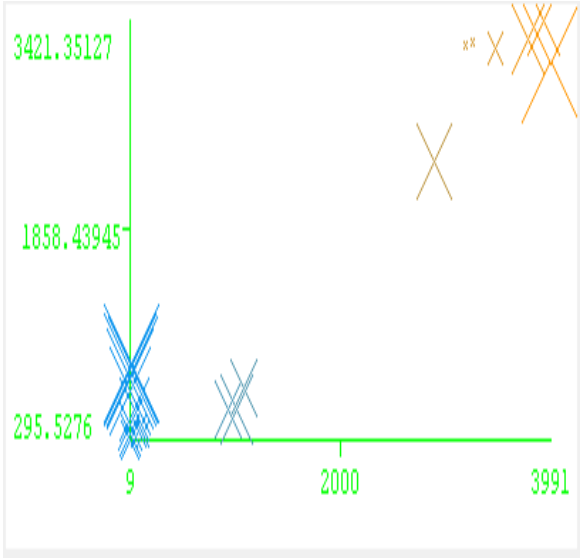
This algorithm provides a root relative squared error of 25.4181% as can be obtained from Table 1

Gaussian Process*:*

Multivariate Gaussian distribution is extended to infinite dimensioanility by Gaussian Process. Formally, any finite subset of the range follows a multivariate Gaussian distribution that is generated by multivariate Gaussian located throughout some domain. Now, assume a single point sampled from multi variate (n-variate) Gaussian distribution after enough thought. This can always be imagined as n observations in an arbitrary data set y={y1,y2,y3…yn}. Hence, working backwards,this data set can be partnered with a GP. Thus GPs are as universal as they are simple.[5]

In probability so far and theory and statistics, Gaussian processes are a family of statistical distributions (not necessarily stochastic processes in which a role is played by time). With a normally distributed random variable is associate every point in an input variable in a gaussian space. Also, multivariate normal distribution is acquired by every finite collection of these random variables. Thus a distribution over functions is the distribution of a Gaussian Process. [6]

Fig 3: Graph Forecasting Open Interest Values using Gaussian Process Algorithm



The algorithm makes a root relative squared error of 34.21% as seen from values in Table 1.

X=Open Interest ;

Y=Predicted Open Interest

Colour variations from blue to orange varies as the value of open interest increases

TABLE 1:

Comparison Chart Of Values Of Algorithms Used:

Factors	Decision Stump	Multilayer Perceptron	Gaussian Process
Corelation-co-efficient	0.9686	0.9675	0.9607
Mean Absolute error	378.9935	311.6241	477.4797
Root mean squared error	427.1242	399.5074	537.7285
Relative absolute error	26.4531%	21.7508%	33.3273%
Root relative squared error	27.1752%	25.4181%	34.2122%
Total number of Instances	22	22	22

Co-relation Co-efficient:

A number between +1 and −1 calculated so as to represent the linear interdependence of two variables or sets of data.

Mean absolute error:

In statistics, the mean absolute error (MAE) is a quantity used to measure how close forecasts or predictions are to the eventual outcomes

Relative absolute error:

The relative absolute error is very similar to the relative squared error in the sense that it is also relative to a simple predictor, which is just the average of the actual values.

Root relative squared error:

The root-mean-square error(RMSE) is a commonly used measure of the differences between values (sample and population values) forecasted by an estimator and the values actually observed.

CONCLUSION:

This paper presents a new approach which is the use of multi-layer perceptron algorithm, decision stump and gaussian process to forecast the prices of aluminium. We applied these algorithms in weka software and obtained the results with precision through training and testing the classification model with real data set. The predictions of prices by using sample data was also obtained. With the results obtained, we can predict that multilayer perceptron algorithm gave us the best results with minimum error and maximum efficiency.This implementation of the most efficient neural network algorithm on a commodity like aluminium opens up new vistas for predicting the prices and could thus help in the greater development of industry. This system if implemented is likely to have a greater impact in the aluminium market.

REFERENCES

[1] Carlos Guestrin. 2007. Decision Trees

[2]www.investopedia.com

[3] https://en.wikipedia.org/wiki/Multilayer_perceptron

[4] http://www.cs.bham.ac.uk/~jxb/INC/l7.pdf

[5] http://www.robots.ox.ac.uk /~mebden/reports/GPtutorial.pdf

[6] https://en.wikipedia.org/wiki/Gaussian_process

[7] http://www.gepsoft.com /gxpt4kb/Chapter10/Section1/SS07.htm

.