

Use of Data Mining Techniques to Forecast Fuel Price

Arundhati M. Gaikwad¹ Aarti Bhosale² Mrunmaye Adsul³ Pooja Shinde⁴

^{1,2,3,4}MKSSS'S Cummins College of Engineering for Women, Pune, India

Abstract— One of the major sectors that is driving the economy of the nation is the fuel industry. Recently, the changes in the pricing of fuels have been evolving and there are no definite indicators or any other entities that can help in predicting how the fuel prices can be determined. This where the data mining concepts and techniques help in doing the same. Data Mining is an analytical process designed to explore data in search of hidden relationships between variables, and then to validate the results by applying the patterns found, to a new subset of data. This paper, gives the use of data mining techniques and use them into a processed form to predict the future fuel prices. We propose this paper which introduces the use of Bayesian classification, for the purpose of prediction of future fuel prices. This method stands out to be more relevant and accurate than other methods which can be considered for the same application. This proposed method is attractive as the results can be obtained using a mobile phone.

Key words: Android Application, Fuel Prices, Data Mining, Classification, Prediction, Stock

I. INTRODUCTION

The fuel industry affects the economic sector majorly. The adverse effects of the fluctuation in fuel prices was noticed after the disintegration of the USSR (Soviet Union) due the decrease in the oil prices suddenly in early 80's. There is need for a statistical analysis of trends and pattern of fluctuation of fuel prices. This will allow the business to manipulate their finance accordingly. Data mining technique is used to identify patterns and establish relationships to solve problems through data analysis. Data mining tools are used to predict future trends. We propose to data mining classification technique for prediction of future fuel prices. This paper proposes use of Naive Bayes classification algorithm for prediction. There are techniques suggested before for oil price prediction, but there is no user friendly application available. This paper proposes to utilize the structured data of historic fuel prices as training dataset for Bayesian classification. The structured data utilization allows more accuracy in the proposed system compared to other classification techniques available. This can be made more attractive and user friendly by using mobile phone application to view the end result.

II. CLASSIFICATION ALGORITHM

Classification consists of predicting a certain outcome based on a given input. In order to predict the outcome, the algorithm processes a training set containing a set of attributes and the respective outcome, usually called goal or prediction attribute. The algorithm tries to discover relationships between the attributes that would make it possible to predict the outcome. Naive Bayes builds classification models in the form of a training dataset. It uses the training dataset for reference and calculates the prediction based on the trends available. The final result is

more accurate compared to other predictive analytical techniques.

III. NAÏVE BAYES ALGORITHM

The Bayesian Classification represents a supervised learning method as well as a statistical method for classification. Assumes an underlying probabilistic model and it allows us to capture uncertainty about the model in a principled way by determining probabilities of the outcomes. It can solve diagnostic and predictive problems. This algorithm uses a training dataset for reference and all the attributes in the dataset are utilized for the prediction. This confirms the accuracy of the solution.

IV. MATHEMATICAL MODEL

Let the system be S ,
then $S = \{ s, e, X, Y, F, DD, NDD, \text{Success}, \text{Failure}, \text{MEMshared}, \text{CPUcorecount} \}$

Here, $S' = \{ s, e, X, Y, F, DD, NDD, \text{Success}, \text{Failure} \}$

Also, $S' C S$

Where,

- s = start condition
- e = end condition
- X = set of inputs
- Y = set of outputs
- $F = \{ F_{me}, F_{friend} \}$
where, F_{me} = principle functions
 F_{friend} = associate functions
- DD = deterministic data
- NDD = non-deterministic data
- Success = desired output is generated
- Failure = desired output is not generated

A. Initialization

Username= {List of all usernames from db}
Password= {List of all passwords from db}
Manager_Info={id,name,username,password}
Admin_Info={unique_id,name,substation_name}
Fuel_info= {price, date}
Database= {Manager_Info, Admin_Info, Fuel_Info, Predicted_Dataset}
CPU_corecent = {1,2,4,6,8}

B. Start State

$s = \{ \text{initial_val_regs} | \text{create data of previous fuel prices and creation of petrol substations} \}$

C. End state

$e = \{ \text{Display predicted dataset} \}$

D. Set of Input

$X = \{ X_1, X_2, X_3 \}$
 $X_1 = \{ \text{username, password} \}$
 $X_2 = \{ \text{unique_id, substation_name} \}$
 $X_3 = \{ \text{previous fuel_prices} \}$

E. Set of output

$$Y=\{Y1,Y2,Y3,Y4\}$$

F. Set of functions

$$F=\{F_me,F_friend\}$$

1) $F_me=\{F_K1,F_K2,F_K3,F_K4\}$

- a) $F_K1=Register_manager\{Username,Password\}$
= Login where Username belongs to Password
- b) $F_K2 = Admin_Info\{unique_id, substation_name\}$
= Authenticate_Admin where Admin_Info belongs to Admin_info
- c) $F_K3 = Fuel_info$ (price,date)
= save where Fuel_info belongs to Fuel_info
- d) $F_K4 = Predict_data(prevData)$
=find_Probability

2) $F_friend = \{F_AS1,F_AS2\}$

- a) $F_AS1= valid_user$ (username)
where username belong to Username
=login if_valid
=Error if_invalid

G. $DD = \{fuel_info,Admin_info\}$

H. $Success = \{Y1,Y2\}$

- $Y1 =$ User login successful
- $Y2 =$ Prediction given successful

I. $Failure = \{Y3,Y4\}$

- $Y3 =$ User Login issue unsuccessful
- $Y4 =$ Prediction not given

V. SYSTEM ARCHITECTURE

The system architecture (see Figure 1) describes the flow of the project work. The first step in the process is the collection of data needed for the work. Here the dataset used is historic fuel prices of different companies, which is collected in the first step. The next step in the process is preprocessing of the data. Here we will convert the raw data into understandable format. Now the preprocessed data is classified into training dataset to predict the future fuel prices of different companies.

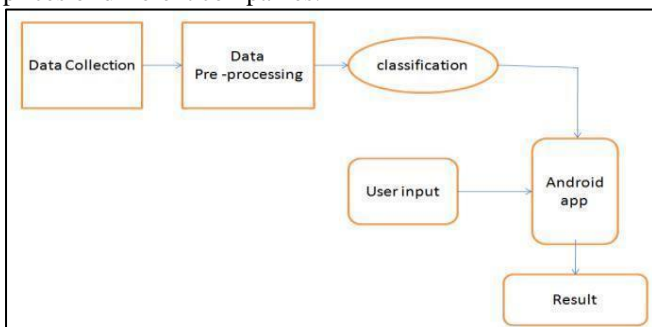


Fig. 1: System Architecture

The user can then enter the details to know his results for the prediction into an android app installed in his mobile device. The attributes entered by the user is compared with the training dataset and the results will be generated.

VI. TECHNOLOGIES

A. Android Studio

Android studio is the IDE used for android application development. It supports java language for development of the application. It comes with built-in libraries and gradle for designing the application.

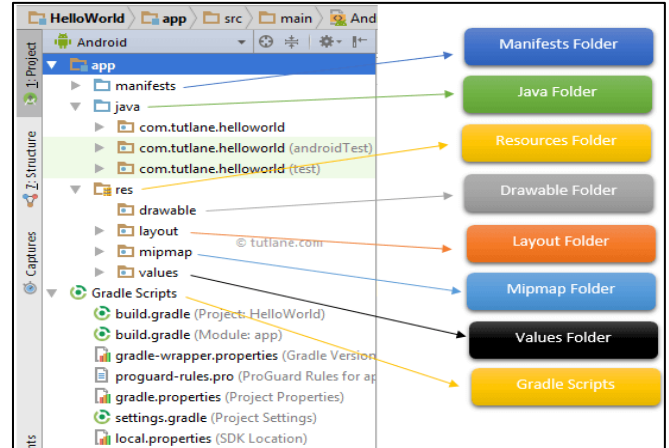


Fig. 2: Android Studio

B. XAMPP Server

Android studio gives view and working of the application on the computer. For the application to work on the mobile phone it needs XAMPP server for connectivity. The server comes with some built-in facilities for database creation and database maintenance. XAMPP server also supports SQL Querying for databases.

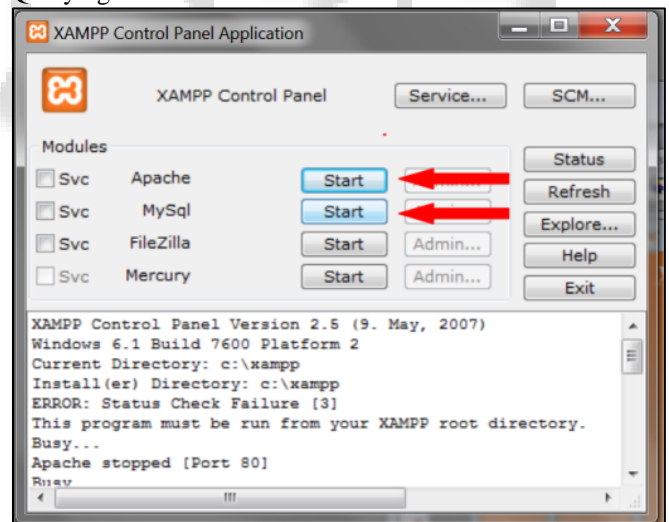


Fig. 3: XAMPP Server Modules

C. JAVA

Java language is supported by Android studio and we are using it for our mobile application development.

VII. CONCLUSION

In this paper we presented the application of data mining concepts and the use of Bayesian Classification for future fuel price prediction. It aims at the study of Bayesian Classification algorithm which undergoes utilization of training dataset for the prediction of future fuel prices. This algorithm used for classification also produces maximum

accuracy when compared to other algorithms of classification. The proposed system of prediction of future fuel prices can be made available to the users with the help of an android phone application.

REFERENCES

- [1] P. Yasodha, M. Kannan, "Analysis of a Population of Diabetic Patient Databases in Weka Tool", International Journal of Scientific & Engineering Research Volume 2, Issue 5, May-2011
- [2] Han, J., Kamber, M.: Data Mining; Concepts and Techniques, Morgan Kaufmann Publishers (2000).
- [3] Folorunso O and Ogunde A. O, "Data Mining as Technique for Knowledge Management in Business Process Redesign" The Electronic Journal of Knowledge Management Volume 2 Issue 1, pp 33-44 (2004)
- [4] <http://software.ucv.ro/~cmihaescu/ro/teaching/AIR/docs/Lab4-NaiveBayes.pdf>
- [5] ibm.im.overview.doc/c_naive_bayes_classification.html
- [6] Kaur H, Wasan SK, "Empirical Study on Applications of Data Mining Techniques in Healthcare", Journal of Computer Science, 2(2):194-200, 2006
- [7] Margaret H. Dunham, "Data Mining Techniques and Applications", McGraw-Hill Publishers
- [8] U.M. Piatetsky-Shapiro and G. Smyth "From Data Mining to Knowledge Discovery : An Overview", 1996, pp.1 -36
- [9] Szakacs-Simon, P. Dept of Autom., "Transilvania" Univ., Brasov, Romania Moraru, S.A.; Perniu, L. "Android application developed to extend health monitoring device range and real-time patient tracking" International Journal of Advanced Research in Computer Science and Software Engineering
- [10] Rohanizadeh S "A proposed data mining methodology application to industrial procedures"
- [11] Anita Thakur, Aishwarya Tiwari, Saswat Kumar, Aditya Jain, Jagjot Singh "NARX Based Forecasting of Petrol Price"
- [12] M.S. Khan, A. Saikia, Dr. Sushanta Tripathy "Prediction of Petroleum Price in India"
- [13] Neha Sehgal, Krishan K Pandey "The Drivers of Oil Prices – A MI3 Algorithm approach"
- [14] Professional Android Application Development 4 by pp-1-137,209-284,389.
- [15] Developing Android application with eclipse(pdf) by Frank Ableson
- [16] Introducing Mobile Application development(pdf) by Ahmed Misbah.
- [17] www.developer.android.com