Turkish Online Journal of Qualitative Inquiry (TOJQI)

Volume 12, Issue 3, June 2021: 2468-2474

Research Article

Accuracy Acquirement for Petrol Price Prediction Using Machine Learning Enhanced Random Forest Algorithm

Vijay Anand M[1] , Devi.T[2] , N. Poornima[3], Gnanavel R[4]

Abstract

In day today life Worldwide the most trending analysis is on Petrol price deviations along forecasting analysis required for everyday activities. Production varies in each states and countries which needs analysis that helps to reduce the demand in fuels and gases. The vital and global need for transportation is fuel needs and on time availability in each and every place. Global positioning system (GPS) updated satellite locations for vehicle navigation to increase the required availability of fuels. Machine Learning (ML) an accurate technique to consume the fuel usage based on the travel distance and prediction using consumption of petrol. We propose a prediction model using random forest algorithm for statistical analysis and attribute based on the dataset collected. We have collected a petrol price variation data from bank bazar and gathered the price variation based on highest and lowest changes. From 2019 to 2021 march petrol price as highest and lowest at the end of year and its performance mentioned as Rise, Decrease, Stable, Unstable as an attributes for consumption range. Based on the vehicle's size such as light weight for travel like cars, jeep and heavy weight vehicles such as container and Lorries the fluently changes takes place. According to the world availabilities of vehicles the traffic increased day by day, also decision making system also fluctuate based on the traffic. Using random forest algorithm 87% of accuracy achieved and data attributes splits the data according to the decision making points and avoid overfitting to the values which identifies the Euclidean distance according to the independent parameters which produces accuracy in proposed model.

Keywords: GPS, Machine learning (ML), Random Forest algorithm, statistics analysis, Euclidean distance, prediction model.

[1]Professor, Department of Computer Science and Engineering, Saveetha Engineering College, Chennai, India.
[2]Assistant Professor, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Chennai, India
[3]Assistant Professor, School of Computing, SRM Institute of Science and Technology, Kattankulathur,Chennai.
[4]Assistant Professor , Department of Computer Science and Engineering, Rajalakshmi Institute of Technology, Chennai, India.
vijaydeepi1212@gmail.com, devit.sse@saveetha.com,rgvelu22@gmail.com, poornimn1@srmist.edu.in

1. Introduction

Randomness characterizes the price of oil which is being influenced by several factors. Time series analysis is one among the traditional model used for forecasting works based on the price obtained from the historical series. These are later decomposed as trends and other components including seasonal as well as random elements and the decisions are being made depending on the historical as well as future trends [1]. Model makes decisions based on historical trends as well as prediction of future trends for investing purpose also.

Sequence of price in trends for petrol is based on the frame of time that helps in understanding the data relationships working on the scale of time [2]. Trend prediction occurs is particular interval of time. Characterization of trend is considered and from the changes in price of oil, price of oil and its randomness is being verified [3]. Another transformation model termed as Markov system including the autoregressive coefficients that are dynamic in nature analyzed the oil price benchmarks. Information particles that are fuzzy in nature can also be employed in predicting the time series trends including temperature as well as financial data [4].

Petrol plays an important role in day to day life as it helps in modes of transport of people from one place to another [5]. Predicting the price of petrol is computationally complex, as the factors concerned with them are more in number. Depending on several factors such as requirements of the dealers and the dataset chosen and classification analysis is performed for development of predictive model [6]. The model makes use of the Random Forest algorithm for classification purpose in order to predict the increase in price range in a given specific period of time. One of the significant supervised learning algorithm is Random Forest algorithm which can be utilized both for problems such as classification as well as regression [7]. The hyperparameters of the decision tree is similar to that of a random forest [8]. Combination of decision tree and bagging classifier is not necessary as the classifier class can be made utilized for the random forest algorithm [9]. Regressor can be used for tasks such as regression.

A bag with several decision tree of different hyperparameters set and training can be performed on several data subsets. Decision finally can be made based on the trees that are available majority and finally being chose by random forest. The main reason for usage of random forest algorithm in predicting petrol price is that overfitting risk is being reduced and the time taken for training is also less. Also, the accuracy of the classifier is high as its works effectively when the size of database is huge. Additionally, missing data can also be predicted accurately.

For development of the predictive model, the petrol dataset used is first preprocessed that can be utilized for model training. Both the training dataset and the testing dataset is being used for model and the training process begins for the model. This is done by random forest method and the classifier accuracy improvement is noted for forecasting the petrol price.

The paper is organized as follows, Section 2 deals with the materials and methods, Section 3 presents the enhanced random forest algorithm, Section 4 explains the proposed system, Section 5 discusses the results and Section 6 concludes the paper.

2. Materials and methods

For our research, the algorithm used is random forest for the purpose of predicting the price of petrol. The petrol dataset is subjected to pre-processing first followed by model training. Several

processes are performed further to increase the accuracy of the predictive model for petrol pricing.

2.2 Dataset

Our dataset is created from bankbazaar website where the price of fuel has been updated on daily basis and consolidated its modified value for each month. Also the data have the parameters such as highest price, lowest prices, performance level and modified rate for each month, based on the performance it is mentioned fluctuated or not. Also on high demand the rate is also defined as stable, unstable, fluctuate, rise or decrease. Random forest algorithm formed a group and decision trees has been initiated and training dataset grouped based on forest and accurate classification done for predicting the size for multiple classes and labels.

3. Enhance Random Forest Algorithm (ERFA)

A supervised algorithm which suits for decision trees and time consumption based on classifying the hyper parameters. It builds a forest where the attributes based on groups and clusters can be classified and applied with regression analysis. Enhance random forest algorithm (ERFA) which builds a building block for all tress that mapped as features to summarize the attributes according to its group. Classifier used to cluster the parameters to form the decision trees and reduce the time for searching the features repeatedly. We design a model based on ERFA that build a prediction model and create a subset according to the cluster splits from each node. Based on the searching node the threshold value suggests the node to visit its next trees and recommend the price analysis. When data is being preprocessed it needs to remove the impurity data such as noise and unnecessary NA values, separated commas etc. In that way Enhanced random forest algorithm shortly remove those features from the petrol price dataset. The outcome which relies on its branches from the trees is based on the class labels. Features that doesn't have next children to identify the label they can be categorized as leaf for deciding the next process. Machine learning model mostly suits in classification and feature mapping for providing best accuracy.

4. Proposed system

Petrol production always deviates in Chennai makes a greater demand for fuel. As travelling modes and transportation vehicles are growing faster the need of fuels also grows higher. More profit for consumer might be better price and stability in cost of fuels. Not only district decides the price but also international market makes the decision for price fixing.

The main objective of our research is

i) To obtain a prediction model for prediction of petrol price that varies in day to day life

ii) To provide an overview of price difference and propose an enhanced random forest algorithms for accurate price prediction

iii) To involve the proposed model using supervised deep learning techniques by testing and validating the classified attributes from dataset, provided an excellent accuracy for predicting the petrol price.

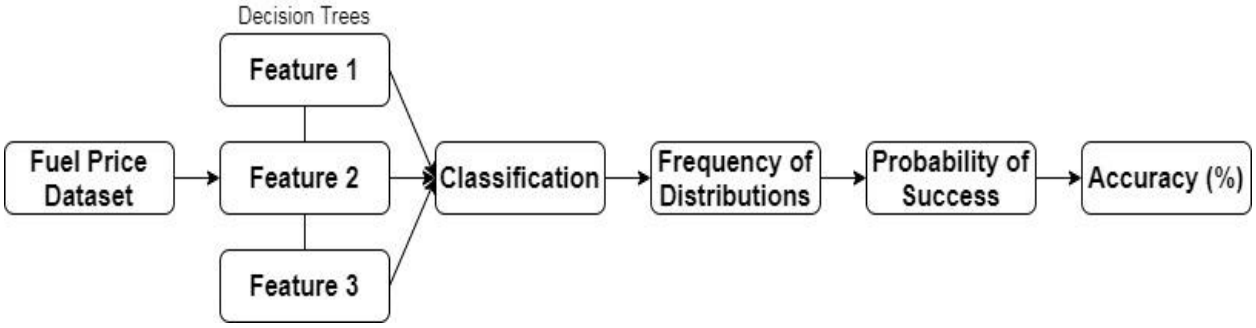

Figure 1. Proposed System

When there is major difference on basis of every months price modification there will be multiple trees formed to make decision. Based on each time output that have been generated from different decision trees the results also varies based on the highest prioritized optimum value as best value. Binary representation as 0 and 1 for all optimum results will make the operation to decide the best solution. The classification based on attribute modified in percentage predicts the accuracy by applying on each multiple decision trees. According to the trees matched there will be low unstable whereas the under fitting or over fitting will be reduced, such that accuracy will be comparatively higher. Missing data from the data set also estimated based on the color and shape of the trees. Error rate and its sequence changed according to the lifespan of the trees usage. As training dataset which classifies as column in X variable which accepted for prediction of price and Y variable as rows corresponding labels that matches the price attributes are allowed to fit model, normalized according to the batch size to get the accurate price for fuel.

4.1 Algorithm

Time-based consumption tree for decision making

a) Input: based on the input create an instance and classify the trees such as T1,T2,...Tn

b) Output: According to the trained attributes the model from trained set values apply the regression and identify the test set (T_s)

Step 1: Use various degrees according to the regression in the estimator and predict the best suit parameter according to the highest depth

Step 2: random regression as enhanced model predict the accuracy and tune the hyper_parameters in terms of attributes used in prediction for performance.

Step 3: Revisit all windows in the required space for identifying and restore the last visited fitting parameters.

Step 4: the fitted coefficient with the least errors in correspondence to the price variation window point will be analyzed

Step 5: Obtained value predicts the exact price on the basis of performance level.

4.2 Prediction model for petrol Dataset

Table 1: Dataset based on various period with maximum and minimum variation

S.NO	YEAR	MONTH	PETROL MAX.PRICE(IN.RS)	PETROL MIN PRICE(IN.RS)	LEVEL OF PERFORMANCE	% MODIFIED
1	2019	DECEMBER	73.18	76.01	RISE	3.86
2	2020	JANUARY	79.04	76.15	DECREASE	-2.49
3	2020	FEBRUARY	76.01	74.64	DECREASE	-1.8
4	2020	MARCH	74.79	72.26	DECREASE	-2.99
5	2020	APRIL	72.37	72.26	RISE	0.02
6	2020	MAY	75.78	72.28	RISE	4.48
7	2020	JUNE	83.63	75.52	RISE	10.7
8	2020	JULY	83.63	83.63	STABLE	0
9	2020	AUGUST	85.62	85.05	RISE	1.71
10	2020	SEPTEMBER	85.09	84.14	DECREASE	-1.11
11	2020	OCTOBER	84.14	84.14	STABLE	0
12	2020	NOVEMBER	85.31	84.14	UNSTABLE	0.57
13	2020	DECEMBER	86.51	85.31	RISE	1.37
14	2021	JANUARY	88.8	86.48	RISE	2.68
15	2021	FEBRUARY	92.58	88.8	RISE	4.25
16	2021	MARCH	93.1	90.8	DECREASE	-2.3

5. Results and Discussion

Using our dataset for petrol price prediction model using enhanced random forest algorithm we have used the modified value for feature analysis. Based on the performance the pandas package has been used as library files and CSV file with all attributes such as highest price, lowest price, performance, month and year and how price have been modified is classified to consume data and visualized. Evaluating the data as the performance classification all 4 data as features are grouped and labeled for training the parameters as well testing. When all the data frame are removed with missing data and not applicable form the price column alone focused according to its labels axis. To train each built tree the groups have been split and decision trees are called to fit the proposed model.by ensemble the generated class the regression value are estimated and predicted.

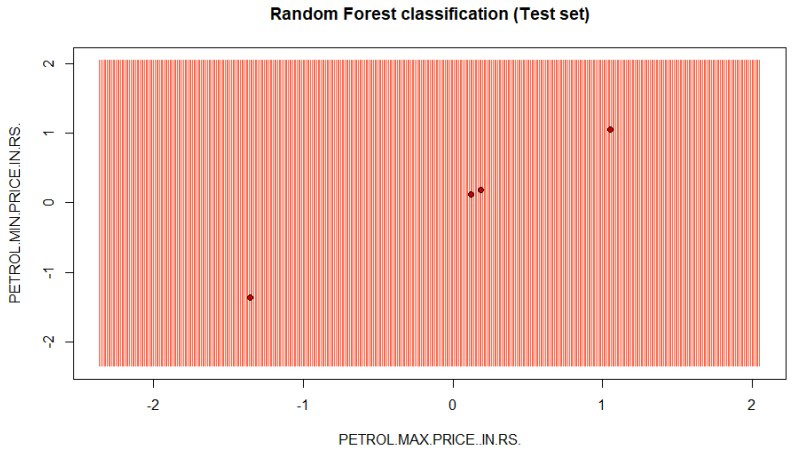

Figure 2. Comparison of attributes based on test set using Enhanced Random forest classification

Figure 2 shows the results of fuel on its independent variables as y-axis defined as minimum price of petrol on month basis where as another feature on x-axis are showing the highest price for predicting the accuracy of price variation on the current year. The performance decides the trees group and fit according to the modification of model.

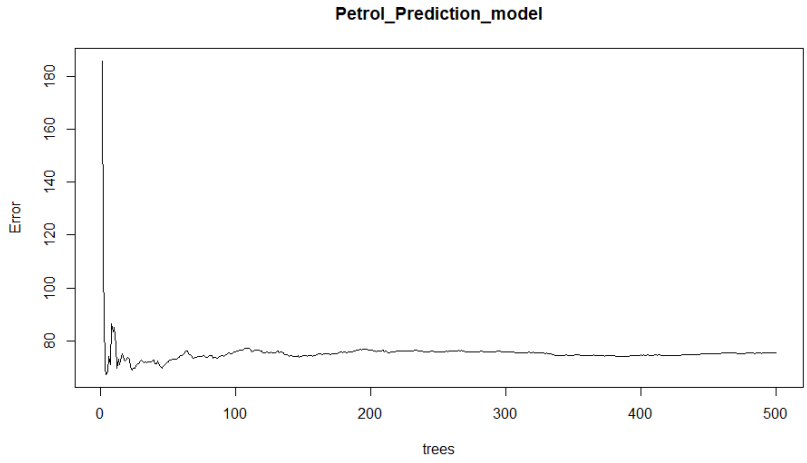

Figure 3: Petrol prediction based on error rates

Figure 3. Shows the petrol prediction error prediction and trees formed according to the fit in model proposed for fuel attributes that have been used for analysis based on performance.

6. Conclusion

Price of petrol is unstable as it keeps on increasing, price prediction is much important. Price forecasting is an Random forest predictive algorithm is used for predicting petrol price. The accuracy of the predictive model is high compared to the existing systems and helps in predicting the rise of petrol price for a specific period of time. The model is found to be robust when it comes to prediction based on the experimental results. It also helps the importers mainly to

reduce the financial loss and also helps in organizing the process of import by usage of the model.

Reference

1. Noguera, J. Oil prices: Breaks and trends. Energy Econ. 2013, 37, 60–67.

2. Zhang, Y.; Zhang, L. Interpreting the crude oil price movements: Evidence from the Markov regime switching model. Appl. Energy 2015, 143, 96–109.

3. Luo, C.; Tan, C.H.; Zheng, Y.J. Long-term prediction of time series based on stepwise linear division algorithm and time-variant zonary fuzzy information granules. Int. J. Approx. Reason. 2019, 108, 38–61.

4. Hamdi, Y.; Reem, A. A novel trend based SAX reduction technique for time series. Expert Syst. Appl. 2019, 130, 113–123.

5. Mello, C.E.; Carvalho, A.S.T.; Lyra, A.; Pedreira, C.E. Time series classification via divergence measures between probability density functions. Pattern Recogn. Lett. 2019, 125, 42–48.

6. Gao, X.; Fang, W.; An, F.; Wang, Y. Detecting method for crude oil price fluctuation mechanism under different periodic time series. Appl. Energy 2017, 192, 201–212.

7. Zhang, M.; Jiang, X.; Fang, Z.; Zeng, Y.; Xu, K. High-order Hidden Markov Model for trend prediction in financial time series. Physica A 2019, 517, 1–12.

8. Huang, P.; Ni, Y. Board structure and stock price informativeness in terms of moving average rules. Q. Rev. Econ. Financ. 2017, 63, 161–169.

9. Brock,W.; Lakonishok, J.; Lebaron, B. Simple technical trading rules and the stochastic properties of stock returns. J. Financ. 1992, 47, 1731–1764.