CRUDE OIL PRICE PREDICTION USING ARTIFICIAL INTELLIGENCE

K.P.Velmurugan [1], K.Nithish [2], S.Naveen Kumar [3] M.Yogesh Kumar [4]

[1 2 3 4] Department of Computer science and Engineering,

[1 2 3 4] Dhanalakshmi Srinivasan Engineering,Perambalur, Tamilnadu, India,

[1] kp.velmurugan2002@gamil.com, [2] nithinithish790@gmail.com, [3] neevan0401@gmail.com,

[4] yokeshiva@gmail.com.

Abstract: Crude oil is the world's most leading fuel. The main advantages of crude oil are it has high density, it is easily available. Oil is used in almost all the industries. Oil is a Constant Power Source. The main aim of this project is to find the different models that efficiently fit the data points and predict the price of fuel with the help of machine learning models. This project works on comparing the different supervised learning models and brings a conclusion based on the efficiency. We have used 3 supervised learning models namely Random Forest Regression, Linear Regression and Decision Tree Regression to know which gives best in terms of accuracy and performance. These algorithms give a numeric value as output. So we can compare the output of these models with the actual models. Now-a-days the oil price has been increasing in leaps and bounds due to certain reasons like inflation throughout the world. Hence these are derived or extracted from petroleum. To predict the values of petroleum like petroleum and Diesel within the future, we've decided to use the Machine Learning algorithms. We use performance metrics to find the performance of the supervised learning models based on their errored value. In this way we can compare different algorithms and find the best one for our problem statement.

Keywords: Python, Algorithm, Price, Deep learning, Data, Prediction, Crude oil, Test and Train, Models.

1. Introduction

Oil energy is very reliable when compared to other sources such as solar and wind energy. Some machine learning models fit the dataset efficiently depending upon the type of data points provided. The main aim of this project is to find the different models that efficiently fit the data points and predict the price of fuel with the help of machine learning models. This project works on comparing the different supervised learning models and brings a conclusion based on the efficiency. We have used 3 supervised learning models namely, Random Forest Regression, Linear Regression and Decision Tree Regression to know which gives the best in terms of accuracy and performance. These algorithms give a numeric value as output. So we can compare the output of these models with the actual models. Now-a-days the oil price has been increasing in leaps and bounds due to certain reasons like inflation throughout the world. Hence these are derived or extracted from petroleum . The sources of crude oil for India come from neighbouring countries such as Dubai and Saudi-Arabia. To predict the values of petroleum like petroleum and Diesel within the future, we've decided to use the Machine Learning algorithms and apply ensemble

learning. Ensemble learning is a technique where we use different algorithms or single algorithms many times. In this way we can compare different algorithms and find the best one for our problem statement.

2.Literature Review

S. N. Abdullah, X. Zeng[1] proposed that among the main factors that affect the volatility of crude oil are the demand and supply of the oil, population and economical aspects.

Generalized Autoregressive Conditional Hetroskedasticity(GARCH) model and Naïve Random walk were among the statistical and econometric model used to predict the crude oil price. The models are used to forecaste the crude oil price and then produce a probabilistic prediction for it. The probabilistic prediction is actually generated by running Monte Carlo analysis on annual WTI average prices. Other statistical model predictions made for crude oil price is by C. Morana.

This research used semi parametric approach suggested in for short term oil price prediction.

Wei-Yin Loh, University of Wisconsin, Madison, USA[2] proposed that regression learning is a machine learning approach that aims to accurately predict the value of continuous output variables from certain independent input variables, via automatic estimation of their latent relationship from data. Tree based regression models are popular in literature due to their flexibility to model higher order non-linearity and great interpretability.

Conventionally, regression tree models are trained in a two stage procedure, i.e, recursive binary partitioning and is employed to produce a tree structure, followed by a pruning process of removing insignificant leaves, with the possibility of assigning

multivariant functions to terminal leaves to improve generalization. The primary goal of applying a regressive analysis is usually to obtain price prediction.

Mr. Brijain R Patel, erMr. Kushik K Rana[4] proposed that researchers have developed various decision tree algorithms over a period of time with enhancement in performance and ability to handle various types of data . some important algorithms are discussed below. CHID:CHAID(Chi-squared automatic interaction detector) is a fundamental decision tree learning algorithm. It was developed by Gordon V Kass in 1980. CHAID is easy to interpret, easy to handle and can be used for classification and detection of interaction between variables. CHID is an extension of AID(Automatic Interaction Detector) and TIDE(Theta Automatic Interaction Detector) procedures. Shen Rong, Zhang Bao-wen[7] proposed that linear regression analysis can be divided into simple linear regression and multiple linear regression . It mainly analyses simple linear regression model that is the analysis method of studying the relations between independent variable and dependent variable. To set up linear regression analysis model Python3.6 is used and introduced pandas analysis package and established more advanced data structure and data analysis package of tool.