

Literature Survey

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 Heteroskedasticity(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 forecast 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. Vol 12, Issue 7, July/2021 ISSN NO:0377-9254 PageNo:251 www.jespublication.com 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 multi variant functions to terminal leaves to improve generalization. The primary goal of applying a regressive analysis is usually to obtain precise prediction.

Mr. Brijain R Patel, erMr. Kushik K Rana[4] proposed that reserchers 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.