Crude oil is one of the most important commodities for almost every part of the world. The changes in its price have great impact on economies around the world.

Many forecasting methods have been developed to predict the price of crude oil, including conventional econometrics models and machine learning approaches, which can provide more accurate prediction than that of econometrics models but are often difficult to interpret economically. As a popular machine learning approach, decision tree models have great predictive power in some studies. Moreover, unlike other machine learning models that are considered as ‘black boxes’ due to difficulties to be interpreted economically, decision tree models are interpretable in theory (Loh, 2014). In this research, we develop various decision tree models to compare with such benchmark models as multiple linear regression (MLR) and ARIMA models for forecasting accuracy. A decision tree is an attractive machine learning method due to its efficiency, robustness, and relatively simple structure (Quinlan, 1986). According to J.R. Quinlan (1992), the prominent advantage of decision tree is that it allows for interpretation easily after making the prediction. The basic decision tree structure

The node on the very top of the tree in Figure 1 is called the root node, which contains the full training dataset where the first split occurs. The nodes at the end of the tree are called leaf nodes, whereas the nodes in between are called intermediate nodes. The root node and the intermediate nodes will split into two subsets based on certain attributes. A decision has to be made whether to split the node into two different nodes or to leave it as a leaf node. This process continues until the tree is fully grown



The node on the very top of the tree in Figure 1 is called the root node, which contains the full training dataset where the first split occurs. The nodes at the end of the tree are called leaf nodes, whereas the nodes in between are called intermediate nodes. The root node and the intermediate nodes will split into two subsets based on certain attributes. A decision has to be made whether to split the node into two different nodes or to leave it as a leaf node. This process continues until the tree is fully grown

***RESEARCH BACKGROUND*** :Crude oil price market prediction is known for its obscurity and complexity. Due to its high vacillation degree, unpredictable irregularity events, and the complex correlations involved between the factors in the market, it is indeed difficult to predict the movements of the crude oil price. E. Panas et. al [2] mentioned that crude oil market has strong evidence of chaos and develops as one of the most volatile market in the world. Corresponding to that, there are few numbers of research conducted for crude oil price prediction. Among the research model used are single statistical and econometric model, single Artificial Intelligence (AI) model and the hybrid. Formerly, Generalized Autoregressive Conditional Heteroskedasticity (GARCH) model and Naive Random Walk were among the statistical and econometric model used to predict the crude oil price. Research [3] successfully utilised a probabilistic model to predict the oil price. The research was conducted based on a case study about the probabilistic inheritance of Belief Network (BN) models.

The models are used to forecast crude oil price and then produce a probabilistic prediction for it [4]. The probabilistic prediction is actually generated by running Monte Carlo analyses on annual WTI average prices. For the purpose of simulation experiment in [4], the analysis done in this study is based on two assumptions of the timing when Iraq’s return to the market and the impact of oil exports from the Former Soviet Union. Three variables input are then used to define the scenarios; the probabilities of embargo ends, total demand and other world productions. The result from the simulation were robust and consistent with the annual average prices are almost certain in between USD$15.00 to USD$25.00 per barrel. There was only 0.75% out of the total scenarios, predicted price over the range. Other statistical model predictions made for crude oil price is by C. Morana [5]. This research used semi parametric approach suggested in [6] for short-term oil price prediction. It also used GARCH to employ oil price changes to predict the oil price distribution over short-term horizon. The approach used one-month-ahead daily Brent oil price which emphasised on periods with the high uncertainty (November 21, 1998 to January 21, 1999). Furthermore, the analysis of the forecasting is based on the last two months of the available data and according to the analysis the result was strayed from the actual. This is most likely linked to the widening of the forecast confidence interval. Nevertheless, the study offers improvement from [6]. Next statistical model used for predicting the crude oil market is by [7] where they predict monthly WTI spot price using relative inventories. This study used Relative Stock Model (RSTK) as the basis to predict the price by comparing two other alternative models; Naïve Autoregressive

(NAIV) forecast model and Modified Alternative (MALT) model. The only variable they used in this research is the petroleum inventories because of its independent practicality and it is readily available every month. RSTK model shows the best performance for both in and out of sample forecast compared to the other two models. It is also being used by the Energy Information Administration (EIA) with among other models to investigate the future market disruptions that derived from changes in demand and production. Nowadays, AI models are among the popular tools to be used for prediction. As an alternative tool to statistical and econometric models, AI offers recognition ability on complex patterns and also on providing intelligent reasoning and intelligent decision-making based on data. Among the single AI models used for predicting the crude oil price is Support Vector Machine (SVM) in [8] where for the task of time-series prediction, this research focused only on Support Vector Regression (SVR) model. In this study, they used monthly WTI price ranging from January, 1970 to December, 2003 as the only independent variable. There was no normalisation process involved in the investigation as to make it simple and moreover, SVM is resistant to noise. This model is evaluated with other two models; AutoRegressive Integrated Moving Average (ARIMA) and Backpropogation Neural Networks (BPNN).

