CRUDEOIL PRICE PREDICTION USING ARTIFICIAL INTELLIGENCE

Abstract:Crudeoilistheworld'smostleadingfuel. The main advantages of crude oil are ithas high density, it is easily available. Oil isusedinalmostalltheindustries.OilisaConstant Power Source. The main aim of thisprojectistofindthedifferentmodelsthateffici ently fit the data points and predict theprice of fuel with the help of machine learningmodels. This project works on comparing thedifferentsupervisedlearningmodelsandbrin gsaconclusionbasedontheefficiency.Wehaveus ed3supervisedlearningmodelsnamelyRandom ForestRegression,LinearRegression and Decision Tree Regression toknowwhichgivesbestintermsofaccuracyandp erformance.Thesealgorithmsgiveanumericval ueasoutput.Sowecancomparetheoutputofthese modelswiththeactualmodels.Now-a-days the oil price has been increasing in leapsandboundsduetocertainreasonslikeinflati onthroughout the world. Hence theseare derived or extracted from petroleum. Topredict the values of petroleum like petroleumand Diesel within the future, we've decided touse the Machine Learning algorithms. We useperformance metrics to find the performanceofthesupervisedlearningmodelsb asedontheirerroredvalue.Inthiswaywecancom parediferentalgorithmsandfindthebestonefor ourproblem statement.

Keywords:Python,Algorithm,Price,Deep learning, Data, Prediction, Crudeoil,Testand Train,Models.

1. Introduction

Oilenergyisveryreliablewhencomparedtoo thersourcessuchassolarandwindenergy. Some machine learning models fitthedatasetefficientlydependinguponthet ypeofdatapointsprovided.Themainaimofth isprojectistofindthedifferentmodelsthateffi cientlyfitthedatapointsandpredicttheprice offuelwiththehelpofmachinelearningmode ls.Thisprojectworks on comparing the differentsupervisedlearningmodelsa ndbringsaconclusionbasedontheefficiency .Wehaveused3supervisedlearningmodelsn amely, Random Forest Regression,Linear Regression and Decision TreeRegressiontoknowwhichgivest hebestintermsofaccuracyandperformance. Thesealgorithmsgiveanumericvalueasoutp ut.Sowecancomparetheoutputofthesemode lswiththeactualmodels.Now-a-days the oil price has been increasing inleapsandboundsduetocertainreasonslike inflation throughout the world. Hencethese are derived or extracted frompetroleum . The sources of crude oil forIndiacomefromneighbouringcountriess uch as Dubai and Saudi-Arabia. Topredict the values of petroleum likepetroleumandDieselwithinthefu ture,we've decided to use the MachineLearningalgorithmsandapp lyensemble

learning.Ensemblelearningisatechniquewhereweusedifferentalgorithmsorsingle algorithms many times. In this waywe can compare different algorithms andfindthebestoneforourproblemstatement.

2.LiteratureReview

S.N.Abdullah,X.Zeng[1]proposed that among the main factors thataffect the volatility of crude oil are thedemand and supply of the oil, populationand economicalaspects.

Generalized Autoregressive ConditionalHetroskedasticity(GARCH)modelandNaïveRandomwalkwereamongthestatisticalandeconometricmodel used to predict the crude oil price.Themodelsareusedtoforecastethecrudeoil price and then produce a probabilisticpredictionforit.Theprobabilisticpredictionisactuallygeneratedbyrunning MonteCarloanalysisonannualWTIaverage prices.Otherstatisticalmodelpredictionsmadeforcrudeoilpriceisby C.Morana.

This research used semi parametricapproach suggested in for short term oilpriceprediction.

Wei-YinLoh,UniversityofWisconsin,Madison, USA[2]proposedthatregressionlearningisa machinelearning approach that aims to accuratelypredictthevalueofcontinuousoutputvariables from certain independent inputvariables,viaautomaticestimationoftheirlatent relationship from data. Tree basedregression models are popular in literatureduetotheirflexibilitytomodelhigherordernon-linearityandgreatinterpretability.

Conventionally, regression treemodels are trained in a two stageprocedure, i.e, recursive binarypartitioning and is employed to produce atreestructure,followed by apruning process of removing insignificant leaves,withthepossibilityofassigning multivariant functions to terminal leavesto improve generalization. The primarygoal of applying a regressive analysis isusuallytoobtain priceprediction.

Mr.BrijainRPatel,erMr.KushikK Rana[4] proposed that researchers havedevelopedvariousdecisiontreealgorithms over a period of time withenhancementinperformanceandability to handle various types of data . someimportant algorithms are discussed below.CHID:CHAID(Chi-squared automaticinteraction detector) is a fundamentaldecision tree learning algorithm. It wasdevelopedby GordonVKassin 1980.
CHAID is easy to interpret, easy tohandle and can be used for classificationand detection of interaction betweenvariables. CHID is an extension ofAID(Automatic Interaction Detector) andTIDE(Theta Automatic InteractionDetector) procedures. Shen Rong, ZhangBao-wen[7] proposed that linearregression analysis can be divided intosimple linear regression and multiplelinear regression . It mainly analysessimple linear regression model that is theanalysis method of studying the relationsbetween independent variable anddependentvariable.To set uplinearregression analysis model Python3.6 isused and introduced pandas analysispackage and established more advanceddata structure and data analysis packageoftool.