CRUDE OIL PRICE PREDICTION

SUBMITTED BY

SARAVANAN.K

AJIN.V

PRAVEEN KUMAR.C

ARUNSAMUEL

LOGESHWARAN.G

DEPARTMENT OF COMPUTER SCIENCE

& ENGINEERING

UDAYA SCHOOL OF ENGIMEERING

VELLAMODI

TEAM ID:PNT2022TMID52057

CONTENTS

INTRODUCTION

1.1 PROJECT OVERVIEW

Crude oil is the world's leading fuel, and its prices have a big impact on the global environment, economy as well as oil exploration and exploitation activities. Oil price forecasts are very useful to industries, governments and individuals. Although many methods have been developed for predicting oil prices, it remains one of the most challenging forecasting problems due to the high volatility of oil prices. In this paper, we propose a novel approach for crude oil price prediction based on a new machine learning paradigm called stream learning. The main advantage of our stream learning approach is that the prediction model can capture the changing pattern of oil prices since the model is continuously updated whenever new oil price data are available, with very small constant overhead. To evaluate the forecasting ability of our stream learning model, we compare it with three other popular oil price prediction models. The experiment results show that our stream learning model achieves the highest accuracy in terms of both mean squared prediction error and directional accuracy ratio over a variety of forecast time horizons.

a. PURPOSE

The world's environment is affected by the oil price falling. With the drop of oil prices, the fuel bills are lowered. As a result, consumers are very likely to use more oil and thus increase the carbon emission. In addition, there is less incentive to develop renewable and clean energy resources. On the other hand, sustained low oil prices could lead to a drop in global oil and gas exploration and exploitation activities. Fluctuating oil prices also play an important role in the global economy. The fall in oil prices would result in a modest boost to global economic activity, although the owners of oil sectors suffer income losses. Recent research from the World Bank shows that for every 30% decline of oil prices, the global GDP (Gross Domestic Product) would be increased by 0.5%. At the same time, the drop of oil prices would reduce the cost of living, and hence the inflation rate would fall.

2.literature survey

2.1 EXISTING PROBLEM

We divide crude oil price forecasting approaches into three categories: (1) heuristic approaches; (2)LSTM; and (3) machine learning techniques. Heuristic approaches for oil price prediction include professional and survey forecasts, which are mainly based on professional knowledge, judgments, opinion and intuition. Another heuristic approach, the so-called no-change forecast, uses the current price of oil as the best prediction of future oil prices. Despite its simplicity, the no-change forecast appeared to be a good baseline approach for oil price prediction and was better than other heuristic judgmental approaches (Alquist et al., 2013).Long short-term memory (LSTM) is an artificial neural network used in the fields of artificial intelligence and deep learning. It has feedback connections and is also applicable to tasks such as unsegmented, machine translation, robot control.Recurrent neural networks (RNN) identifies to be the most powerful and impactful models for processing time-series based sequential data. LSTM variants can be used for other task as well other than prediction such as speech, handwriting and polyphonic modelling. LSTM networks are well-suited to classifying, processing and making predictions based on time series data, since there can be lags of unknown duration between important events in a time series.

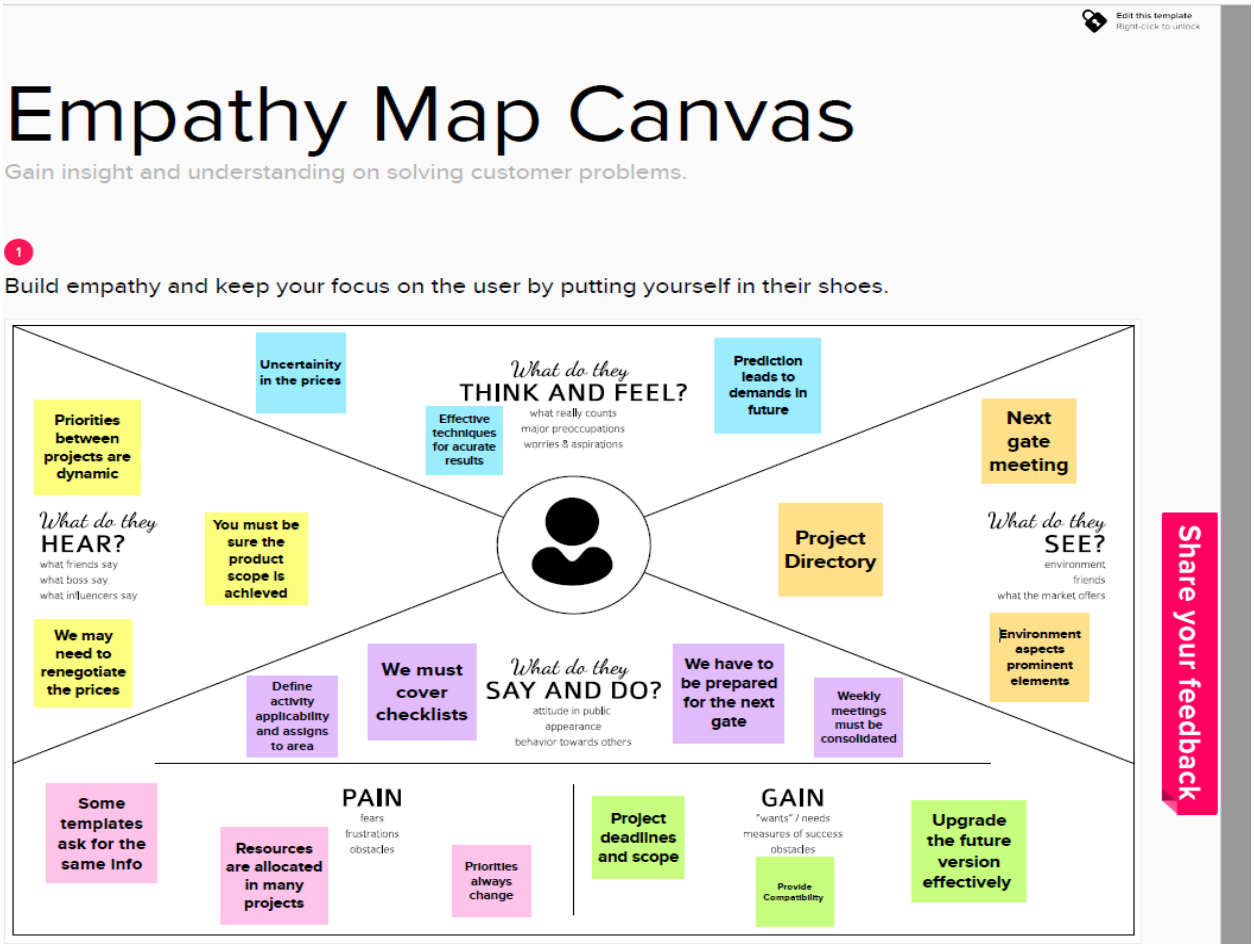
Several machine learning techniques were proposed for oil price prediction, such as artificial neural networks (ANN) (Yu et al., 2008, Kulkarni and Haidar, 2009), and support vector machine (SVM) (Xie et al., 2006). These are nonlinear models which may produce more accurate predictions if the oil price data are strongly nonlinear (Behmiri and Pires Manso, 2013). However, these machine learning techniques, like other traditional machine learning techniques, rely on a fixed set of training data to train a machine learning model and then apply the model to a test set. Such an approach works well if the training data and the test data are generated from a stationary process, but may not be effective for non-stationary time series data such as oil price data.

2.2 problem statement definition

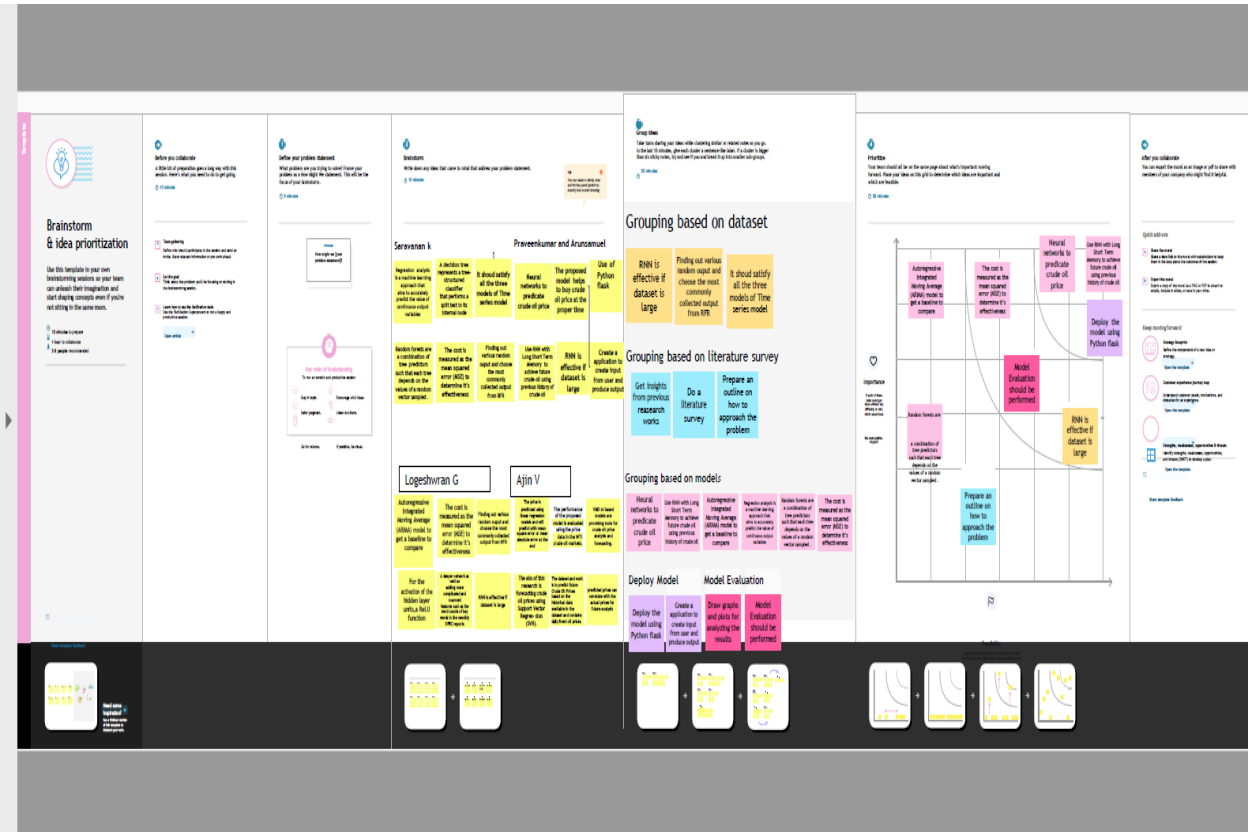
Crude oil is the world's leading fuel, and its prices have a big impact on the global environment,economy as well as oil exploration and exploitation activities. Oil price forecasts are very useful to governments,industry individuals. Although many methods have been developed for predicting oil prices, it remains one of the most challenging forecasting problems due to the high volatility of oil prices. The volatility of crude oil market and its chain effects to the world economy augmented the interest and fear of individuals, public and private sectors. Previous statistical and econometric techniques used for prediction, offer good results when dealing with linear data. Nevertheless, crude oil price series deal with high nonlinearity and irregular events. The continuous usage of statistical and econometric techniques including AI for crude oil price prediction might demonstrate demotions to the prediction performance.

3.IDEATION & PROPOSED SOLUTION

3.1 EMPATHY MAP CANVAS



3.2 IDEATION AND BRAINSTORMING

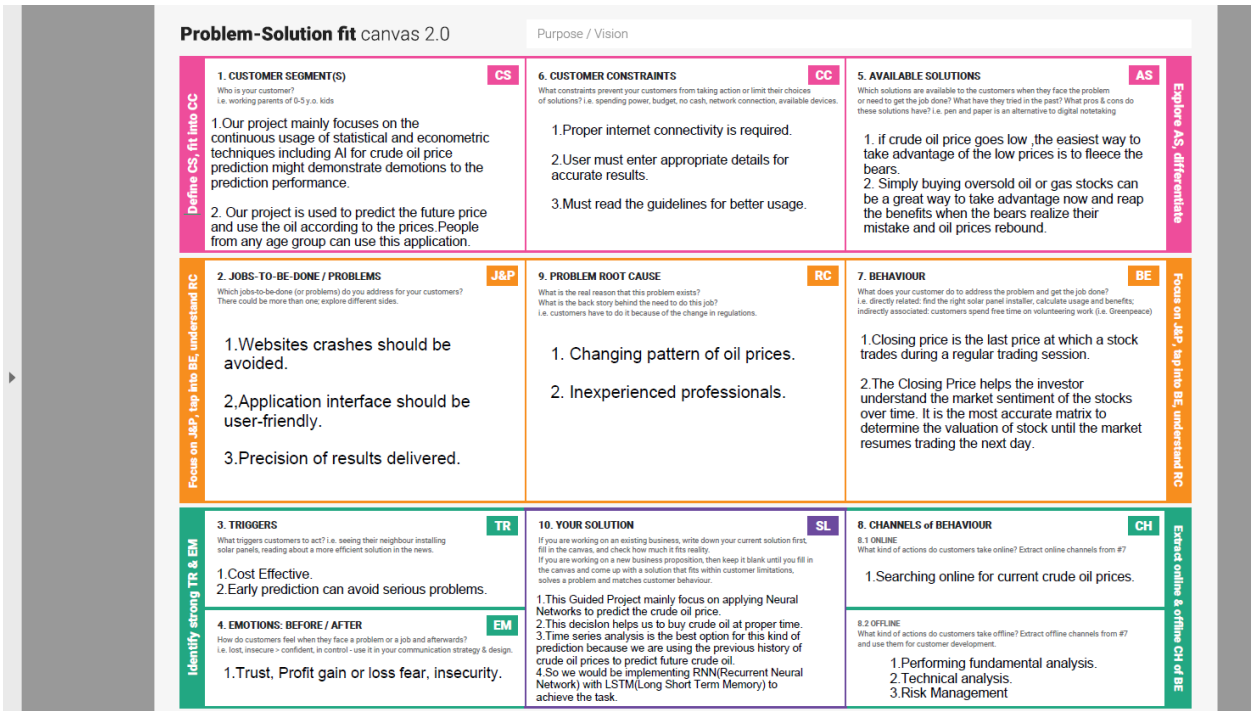


3.3 PROPOSED SOLUTION

S.No.	Parameter	Description
1.	Problem Statement (Problem to be solved)	Crude oil is the world's leading fuel, and its prices have a big impact on the global environment and its forecasts are very useful to governments,industry is individuals.The continuous usage of statistical and econometric techniques including AI for crude oil price prediction might demonstrate demotions to the predictionperformance.
2.	Idea / Solution description	RNN is used with long short term memory to achieve future crude oil using previous history of crude oil.The cost is measured as the mean squared error to determine it's effectiveness.The performance of the proposed model is evaluated using the price data in the WTO crude oil materials
3.	Novelty / Uniqueness	● Crude oil price fluctuations have a far reaching impact on global economies and thus price forecasting can assist in minimising the risks associated with volatility in oil prices. ● Price forecasts are very important to various stakeholders: governments, public and private enterprises, policymakers, and investors.

4.	Social Impact / Customer Satisfaction	● It is used to predict the future price and use the oil according to the prices. ● this price has direct effects on several goods and products and its fluctuations affect the stock markets. ● Oil prices are not only driven by economic variables, but they are also affected by key events ..
5.	Business Model (Revenue Model)	● It can help decision makers – either firms, private investors, or individuals – when choosing to buy or sell the crude oil ● crude oil is one of the most profitable trading commodities for traders. ● **RNN and LSTM models are used as the benchmark model to predict the crude oil** **prices.**
6.	Scalability of the Solution	● PCA, MDS and LLE methods are used to reduce the dimensions of the data ● Improve the accuracy of the RNN and LSTM models.

3.4 PROBLEM SOLUTION FIT


Problem-Solution fit canvas 2.0
Purpose / Vision
Define CS, fit into CC
1. CUSTOMER SEGMENT(S)
CS
Who is your customer?
i.e. working parents of 0-5 y.o. kids
1.Our project mainly focuses on the continuous usage of statistical and econometric techniques including AI for crude oil price prediction might demonstrate demotions to the prediction performance.
2. Our project is used to predict the future price and use the oil according to the prices.People from any age group can use this application.
6. CUSTOMER CONSTRAINTS
CC
What constraints prevent your customers from taking action or limit their choices of solutions? i.e. spending power, budget, no cash, network connection, available devices.
1.Proper internet connectivity is required.
2.User must enter appropriate details for accurate results.
3.Must read the guidelines for better usage.
5. AVAILABLE SOLUTIONS
AS
Which solutions are available to the customers when they face the problem or need to get the job done? What have they tried in the past? What pros & cons do these solutions have? i.e. pen and paper is an alternative to digital notetaking
1. if crude oil price goes low ,the easiest way to take advantage of the low prices is to fleece the bears.
2. Simply buying oversold oil or gas stocks can be a great way to take advantage now and reap the benefits when the bears realize their mistake and oil prices rebound.
Explore AS, differentiate
Focus on J&P, tap into BE, understand RC
2. JOBS-TO-BE-DONE / PROBLEMS
J&P
Which jobs-to-be-done (or problems) do you address for your customers? There could be more than one; explore different sides.
1.Websites crashes should be avoided.
2,Application interface should be user-friendly.
3.Precision of results delivered.
9. PROBLEM ROOT CAUSE
RC
What is the real reason that this problem exists? What is the back story behind the need to do this job? i.e. customers have to do it because of the change in regulations.
1. Changing pattern of oil prices.
2. Inexperienced professionals.
7. BEHAVIOUR
BE
What does your customer do to address the problem and get the job done? i.e. directly related: find the right solar panel installer, calculate usage and benefits; indirectly associated: customers spend free time on volunteering work (i.e. Greenpeace)
1.Closing price is the last price at which a stock trades during a regular trading session.
2.The Closing Price helps the investor understand the market sentiment of the stocks over time. It is the most accurate matrix to determine the valuation of stock until the market resumes trading the next day.
Focus on J&P, tap into BE, understand RC
Identify strong TR & EM
3. TRIGGERS
TR
What triggers customers to act? i.e. seeing their neighbour installing solar panels, reading about a more efficient solution in the news.
1.Cost Effective.
2.Early prediction can avoid serious problems.
4. EMOTIONS: BEFORE / AFTER
EM
How do customers feel when they face a problem or a job and afterwards? i.e. lost, insecure > confident, in control - use it in your communication strategy & design.
1.Trust, Profit gain or loss fear, insecurity.
10. YOUR SOLUTION
SL
If you are working on an existing business, write down your current solution first, fill in the canvas, and check how much it fits reality.
If you are working on a new business proposition, then keep it blank until you fill in the canvas and come up with a solution that fits within customer limitations, solves a problem and matches customer behaviour.
1.This Guided Project mainly focus on applying Neural Networks to predict the crude oil price.
2.This decision helps us to buy crude oil at proper time.
3.Time series analysis is the best option for this kind of prediction because we are using the previous history of crude oil prices to predict future crude oil.
4.So we would be implementing RNN(Recurrent Neural Network) with LSTM(Long Short Term Memory) to achieve the task.
8. CHANNELS of BEHAVIOUR
CH
8.1 ONLINE
What kind of actions do customers take online? Extract online channels from #7
1.Searching online for current crude oil prices.
8.2 OFFLINE
What kind of actions do customers take offline? Extract offline channels from #7 and use them for customer development.
1.Performing fundamental analysis.
2.Technical analysis.
3.Risk Management
Extract online & offline CH of BE

4. REQUIREMENT ANALYSIS

4.1 FUNCTIONAL REQUIREMENTS

FR No.	Functional Requirement (Epic)	Sub Requirement (Story / Sub-Task)
FR-1	User Application	User Direct Open With Google Play Store App User Can Download The Crude Oil Price
FR-2	User Products Available	User Using The Application There Are So Many Products In Crude Oil Price App User Update The Energy And Oil Price Instant The Application
FR-3	User Additional Features	User Can Read Latest News And View Oil Price Charts User View Major Energy Quotes User Can Using A Multiple Color Themes
FR-4	User Exceptions	User Can Exchange Rates And Currancy Converter

4.2 NON-FUNCTIONAL REQIREMENTS

FR No.	Non-Functional Requirement	Description
NFR-1	**Usability**	Used to improve to the Accuracy of crude oil price prediction
NFR-2	**Security**	In the rising oil price can even shift economical/political power from oil importers to oil exporters communications will be secured
NFR-3	**Reliability**	Reliability of the pointing towards high –risk components
NFR-4	**Performance**	Performance of the this project is to improve to the accuracy of crude oil price prediction

NFR-5	**Availability**	The Availability Solution is More Benefit for and the Importers and exporters in the crude oil price prediction.
NFR-6	**Scalability**	The scalability are 90%-95%

5. PROJECT DESIGN

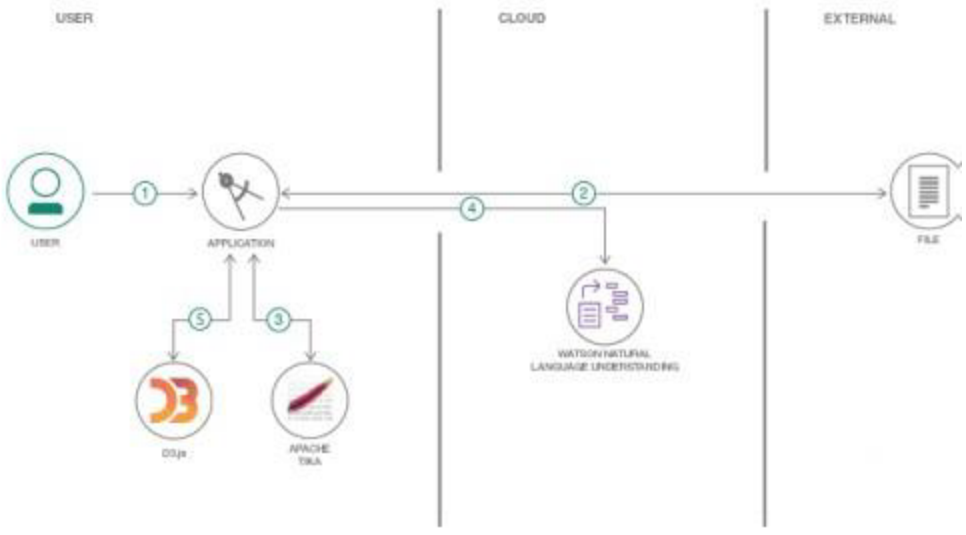
5.1 DATA FLOW DIAGRAM

Data Flow Diagrams:

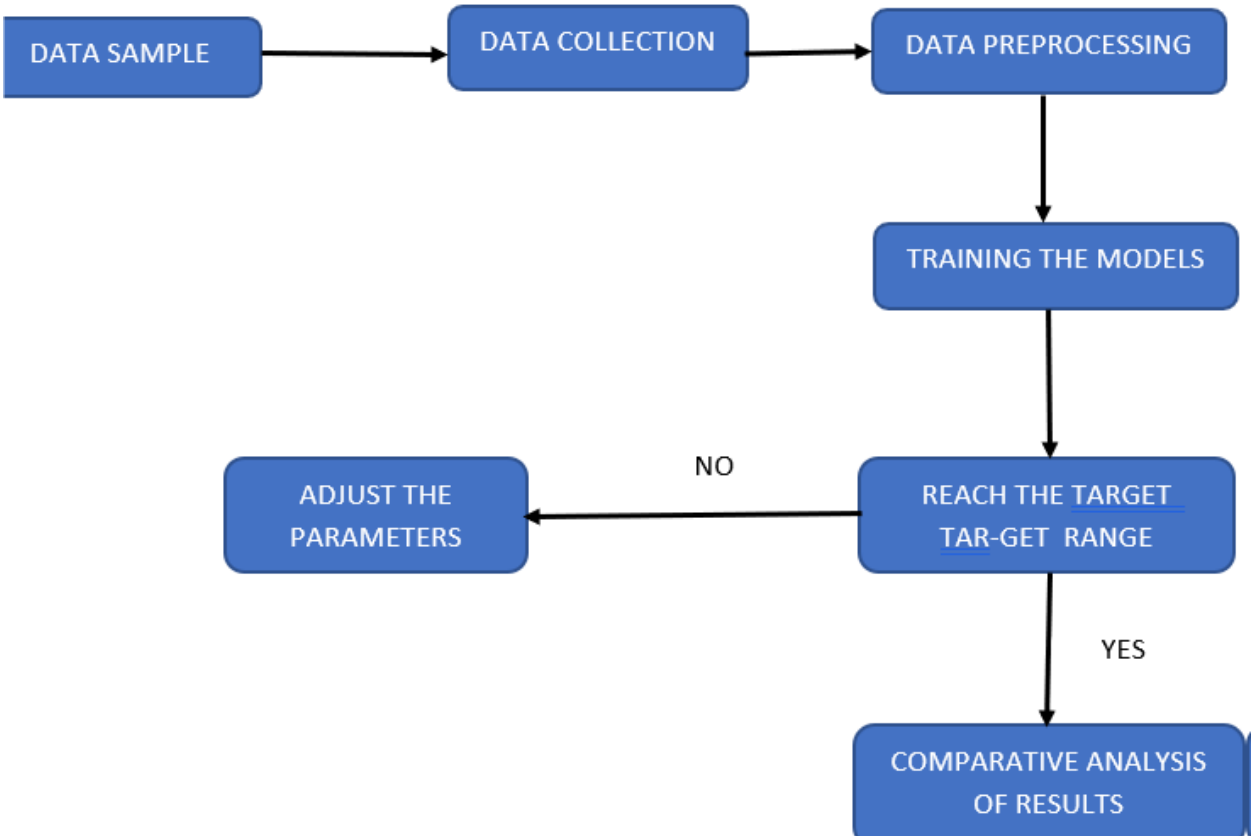
A Data Flow Diagram (DFD) is a traditional visual representation of the information flows within a system. A neat and clear DFD can depict the right amount of the system requirement graphically. It shows how data enters and leaves the system, what changes the information, and where data is stored.

Example: DFD Level 0 (Industry Standard

Example: (Simplified)

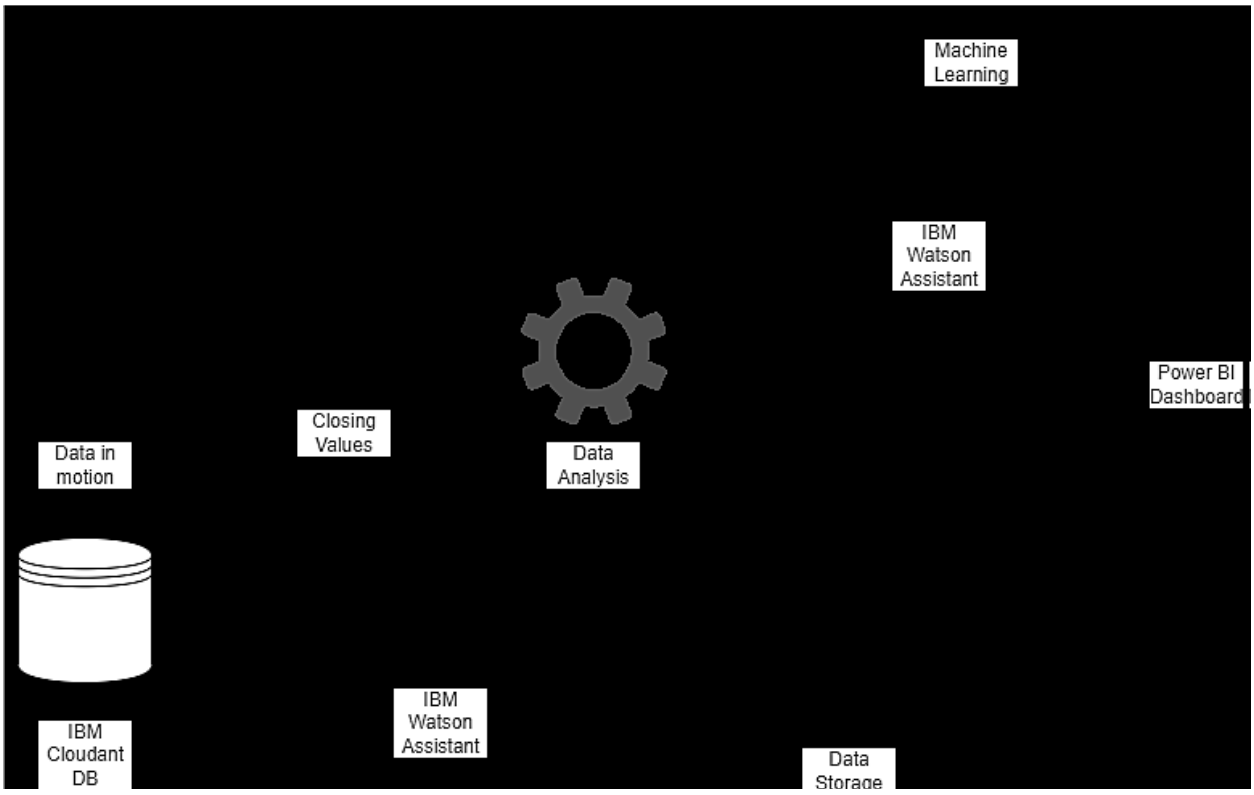


CRUDE OIL PRICE PREDICTION

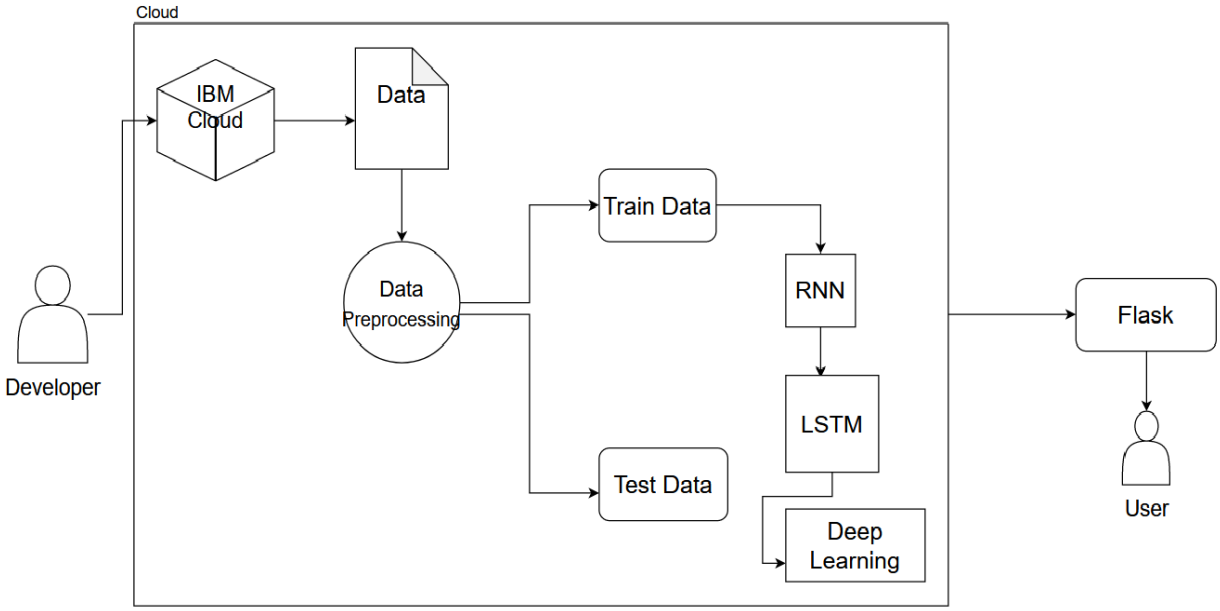


5.2 SOLUTION AND TECHNICAL ARCHITECTURE

SOLUTION ARCHITECTURE



TECHNICAL ARCHITECTURE



6.PROJECT PLANNING AND SCHEDULING

6.1 SPRINT PLANNING AND ESTIMATION

Sprint	Functional Requirement (Epic)	User Story Number	User Story / Task	Story Points	Priority
Sprint-1	Registration	USN-1	As a user, I can register for the application by entering my email, password, and confirming my password.	10	High
Sprint-1		USN-2	As a user, I will receive confirmation email once I have registered for the application	10	High
Sprint-1	Login	USN-3	As a user, I can log into the application by entering email & password.	15	High
Sprint-2	Input Necessary Details	USN-4	As a user, I can give Input Details to Predict Likeliness of crude oil	15	High
Sprint-2	Data Pre-processing	USN-5	Transform raw data into suitable format for prediction.	15	High
Sprint-3	Prediction of Crude Oil Price	USN-6	As a user, I can predict Crude oil using machine learning model.	20	High
Sprint-3		USN-7	As a user, I can get accurate prediction of crude oil	5	Medium

Sprint-4	Review	USN-8	As a user, I can give feedback of the application.	20	High

6.2 SPRINT DELIVERY SCHEDULE

TITLE	DESCRIPTION	DATE
Literature Survey & Information Gathering	Literature survey on the selected project & gathering information by referring the, technical papers, research publications etc.	28 SEPTEMBER 2022
Prepare Empathy Map	Prepare Empathy Map Canvas to capture the user Pains & Gains, Prepare list of problem statements	24 SEPTEMBER 2022 24 OCTOBER 2022
Ideation	List the by organizing the brainstorming session and prioritize the top 3 ideas based on the feasibility & importance.	25 SEPTEMBER 2022
Proposed Solution	Prepare the proposed solution document, which includes the novelty, feasibility of idea, business model, social impact, scalability of solution, etc.	23 SEPTEMBER 2022
Problem Solution Fit	Prepare problem - solution fit document.	30 SEPTEMBER 2022
Solution Architecture	Prepare solution architecture document.	28 SEPTEMBER 2022

7.RESULTS

7.1 PERFORMANCE METRICS

We use two standard performance metrics in the oil price prediction literature for comparing different oil price prediction models. The first metric is Mean Squared Prediction Error (MSPE). MSPE of a prediction model measures the average of the squares of the prediction errors. The prediction error is the difference between the true value and the predicted value. Let y_1, y_2, …, y_n be the true oil prices and yˆ1, yˆ2, …, yˆn be the predicted oil prices under an oil price prediction model, then the MSPE of that model is:MSPE=1n∑i=1n(yˆi−yi)2
For comparison purposes, we use the no-change model as the baseline model and express the MSPE of another model as a ratio relative to the MSPE of the no-change model. If the MSPE ratio of a model is less than 1, then the model is more accurate than the no-change model in terms of MSPE.

The second metric is Directional Accuracy Ratio (DAR), which measures the accuracy of predicting the direction of oil price change (i.e., whether oil price increases or decreases in the next time slot). It can be computed as follows:DAR=1n∑i=1ndtwhere dt=1 if (yˆt−yt−1)(yt−yt−1)>0 and dt=0 otherwise. Note that if we do a random guess of the oil price direction by tossing a fair coin, the DAR would be 0.5. Thus, if the DAR of a model is greater than 0.5, then the model is better than a random guess.

8. CONCLUTION

Forecasting crude oil prices is a very challenging problem due to the high volatility of oil prices. In this paper, we developed a new oil price prediction approach using ideas and tools from stream learning, a machine learning paradigm for analysis and inference of continuous flow of non-stationary data. Our stream learning model will be updated whenever new oil price data are available, so the model continuously evolves over time, and can capture the changing pattern of oil prices. In addition, updating the model requires only a small constant time per new data example, as opposed to re-training the model using the entire training data set. The experiment results show that our stream learning model outperformed three other popular oil price prediction models over a variety of forecast time horizons.

9. FUTURE SCOPE

Crude oil is the world's leading fuel, and its prices have a big impact on the global environment, economy as well as oil exploration and exploitation activities. Oil price forecasts are very useful to industries, governments and individuals. Although many methods have been developed for predicting oil prices, it remains one of the most challenging forecasting problems due to the high volatility of oil prices. In this paper, we propose a novel approach for crude oil price prediction based on a new machine learning paradigm called stream learning. The main advantage of our stream learning approach is that the prediction model can capture the changing pattern of oil prices since the model is continuously updated whenever new oil price data are available, with very small constant overhead. To evaluate the forecasting ability of our stream learning model, we compare it with three other popular oil price prediction models. The experiment results show that our stream learning model achieves the highest accuracy in terms of both mean squared prediction error and directional accuracy ratio over a variety of forecast time horizons.

10. APPENDIX

Python:

Python is an interpreted, high-level, general purpose programming language created by Guido Van Rossum and first released in 1991, Python's design philosophy emphasizes code Readability with its notable use of significant White space. Its language constructs and object oriented approach aim to help programmers write clear, logical code for small and large-scale projects. Python is dynamically type and garbage collected. It supports multiple programming paradigms, including procedural, object oriented ,and functional programming.

Keras :

Keras is a powerful and easy-to-use free open source Python library for developing and evaluating **deep learning** model **.**It wraps the efficient numerical computation libraries **Theano** and **TensorFlow** and allows you to define and train neural network models in just a few lines of code. It uses libraries such as Python, C#,C++ or standalone machine learning toolkits. Theano and TensorFlow are very powerful libraries but difficult to understand neural network.Keras is based on minimal structure that provides a clean and easy way to create deep learning models based on TensorFlow or Theano. Keras is designed to quickly define deep learning models. Well, Keras
is an optimal choice for deep learning applications.

Steps for creating a keras model:

1)First we must define a network model.
2)Compile it, which transforms the simple sequence of layers into a complex group of matrix operations.
3)Train or fit the network.
To import: from keras.models import Sequential
From keras.layers import Dense, Activation, Dropout

TensorFlow:

TensorFlow is a Python library for fast numerical computing created and released by Google. It is a foundation library that can be used to create Deep Learning models directly or by using wrapper librarie sthat simplify the process built on top of **TensorFlow.** TensorFlow tutorial is designed for both beginner and professionals. Our tutorial provides all the basic and advanced concept of machine learning and deep learning concept such as deep neural network, image processing and sentiment analysis. TensorFlow is one of the famous deep learning frameworks, developed by **Google** Team. It is a free and
open source software library and designed in **Python** programming language, this tutorial is designedin such a way that we can easily implements deep learning project on TensorFlow in an easy andefficient way. Unlike other numerical libraries intended for use in Deep Learning like **Theano**,**TensorFlow** was designed for use both in research and development and in production systems. It canrun on single CPU systems, GPUs as well as mobile devices and largescale distributed systems ofhundreds of machines.

Numpy:

NumPy is a Python library used for working with arrays. It also has functions for working in domain oflinear algebra, Fourier transform, and matrices. Numpy which stands for Numerical Python, is a libraryconsisting of multidimensional array objects and a collection of routines for

processing those arrays.Using NumPy, mathematical and logical operations on arrays can be performed. This tutorial explains
the basics of NumPy such as its architecture and environment. It also discusses the various array functions, types of indexing, etc. It is an opensource project and you can use it freely.
NumPy stands for Numerical Python. NumPyaims to provide an array object that is up to 50x faster than traditional Python lists. The array object inNumPy is called **ndarray**, it provides a lot of supporting functions that make working with **ndarray** very easy. Arrays are very frequently used in data science, where speed and resources arevery important.

Pillow:

Pillow is a free and ope nsource library for the Python programming language that allows you to easily create &s manipulate digital images. Pillow is built on top of PIL (Python Image Library). PIL is one of the important modules for image processing in Python. However, the PIL module is not supported since 2011 and does n't support python 3.
Pillow module gives more functionalities, runs on all major operating system and support for python
3. It supports wide variety of images such as "jpeg", "png", "bmp", "gif", "ppm", "tiff". You can do almost anything on digital images using pillow module. Apart from basic image processing functionality, including point operations, filtering images using built-in convolution kernels, and color space conversions.

Tkinkter:

Tkinter is the standard **GUI library** for Python. Python when combined with Tkinter provides a fast and easy way to create **GUI applications**. Tkinter provides a powerful object-oriented interface to the Tk GUI toolkit.We need to import all the modules that we are going to need for training our model. The Keras library already contains some datasets and MNIST is one of them. So we can easily import the dataset through Keras. The mnist.load_data() method returns the training data, its labels along with the testing data and its labels.

Jupyter Notebook:

Jupyter Lab is a web-based interactive development environment for Jupyter notebooks, code, and data. JupyterLab is flexible: configure and arrange the user interface to support a wide range of workflows in data science, scientific computing, and machine learning.JupyterLab is extensible and modular: writeplugins that add new components and integrate with existing ones.

Machine Learning:

Machine learning is a method of data analysis that automates analytical model building. It is a branch of artificial intelligence based on the idea that systems can learn from data, identify patterns and make decisions with minimal human intervention.

Deep Learning:

Deep learning is an artificial intelligence (AI) function that imitates the workings of the human brain in processing data and creating patterns for use in decision making. Deep learning is a subset of machine learning in artificial intelligence that has networks capable of learning unsupervised from data that is unstructured or unlabeled. Also known as deep neural learning or deep neural network.

Neural Networks:

A neural network is a series of algorithms that endeavors to recognize underlying relationships in a set of data through a process that mimics the way the human brain operates. In this sense, neural networks refer to systems of neurons, either organic or artificial in nature.

GitHub link
https://github.com/IBM-EPBL/IBM-Project-51115-1660972139