



IBM-Project-47385-1660798812

PROJECT DOCUMENTATION REPORT

CRUDE-OIL PRICE PREDICTION

TECHNOLOGY: ARTIFICIAL INTELLIGENCE

Team ID: PNT2022TMID46457

TEAM MEMBERS

Team Size: 4

Team Leader: SIVASAKTHI A

Team member: PRIYADHARSHINI B

Team member: SWETHA S

Team member: AFREEN YUSUFA A

1 INTRODUCTION:

1.1 PROJECT OVERVIEW:

This document is provided as a report for the project **Crude Oil Price Prediction**.

Crude oil is amongst the most important resources in today's world, it is the chief fuel and its cost has a direct effect on the global habitat, our economy and oil exploration, exploitation and other activities. Prediction of oil prices has become the need of the hour, it is a boon to many large and small industries, individuals, the government. The evaporative nature of crude oil, its price prediction becomes extremely difficult and it is hard to be precise with the same. Several different factors that affect crude oil prices.

1.2 PURPOSE:

The purpose of this document is to provide a clear-cut view of the project undertaken and produce a neat and greater understanding of the project.

2 LITERATURE SURVEY:

2.1 EXISTING PROBLEM:

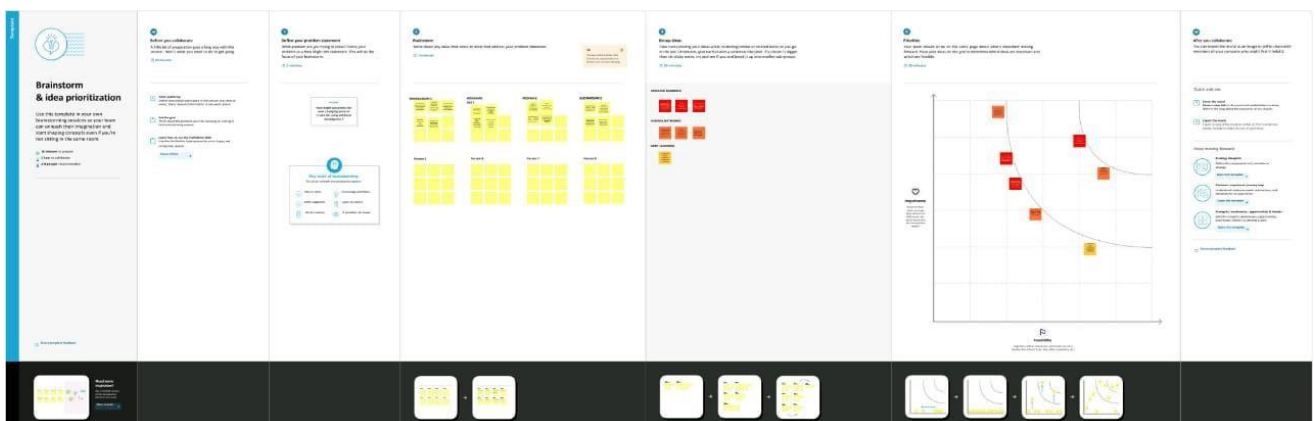
One of the most significant commodities in the world, crude oil is responsible for one-third of the world's energy use. It serves as the foundation for the majority of the items we use on a daily basis, ranging from plastics to transportation fuels. Since changes in the price of crude oil have a significant impact on national economies around the world, price forecasting can help reduce the risks brought on by oil price volatility. For a variety of stakeholders, including governments, public and private organisations, policymakers, and investors, price projections are crucial.

2.2 REFERENCES:

<https://drive.google.com/drive/folders/1yq9UqoGpyAQFKR6ARNFwpVMofYtOHdCm?usp=sharing>

It is required to forecast CRUDE OIL PRICE in international market. The input and output should also be shown as charts and/or dashboards in various formats (like day, week, work-week, month, quarter, year, etc.). The models should be built with comprehensive explanation of data (using EDA), trend analysis, assumptions, data cleaning and validation, data augmentation (if required). Performance of various models need to be clearly evaluated and best model needs to be recommended based on some robust evaluation criteria e.g., AIC (Akaike information criterion), Accuracy, RMSE, MSE etc.

3.1 EMPATHY MAPS:



3.3 PROPOSED SOLUTION:

S. No.	Parameter	Description
1	Problem Statement (Problem to be Solved)	The existing patrimony model of oil price prediction is not capable enough to deliver the accurate predicted prices as expected. Few factors can be described as the conjectural buying and selling , geopolitical, OPEC output, increased demand from important role in the prediction of the oil prices. Now problem arising with the current ANN and CNN models that are used as prediction model's are that they can't provide accurate results when the data is too big.
2	Idea / Solution description	1) LSTM clears about keeping the previous data and prediction which might be encouraging and more accurate. The possible results are comparatively inspiring. 2) The LSTM model will be updated whenever new oil price data are available, and provided to model, so the model continuously evolves over time, and can capture the changing pattern of oil prices.
3	Novelty / Uniqueness	1) Price forecasting can assist in minimising the risks associated with volatility in oil prices. 2) Price forecasts are very important to various stakeholders: governments, public and private enterprises, policymakers, and investors.
4	Social Impact / Customer Satisfaction	1) Brand activation 2) Innovative and schemes 3) Instant reward schemes 4) Personalized consumer purchase exchanges 5) Capability building of sales personnel.

5	Business Model (Revenue Model)	<p>1) The price of crude oil should be easily predictable from the equilibrium between demand and supply.</p> <p>2) Traders analyze demand and supply factors and take calculated positions. If their prediction comes true, traders close their position to book profits way before expiry.</p> <p>3) price of crude oil are changeable based from time to time.</p>
6	Scalability of the Solution	<p>1) hydrodynamic conditions in oilfield operations is suggested.</p> <p>2) Modern refineries typically use a high number of sensors that generate an enormous amount of data.</p> <p>3) Sustainable Solution for Crude Oil using Concentrated Solar Power Technology.</p>

3.4 PROBLEM SOLUTION FIT:

Problem-Solution fit canvas 2.0 Purpose / Vision: Crude Oil Price Prediction

Define CS, fit into CC	1. CUSTOMER SEGMENT(S) CS Who is your customer? I.e. working parents of 0-5 y.o. kids Our Target Customers are: → Private Organizations → Public Organizations → Government Enterprises Who, wants to understand the oscillations in Crude Oil Prices and gain both tangible and intangible benefits out of it.	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. There are several constraints that are encountered. Some of them are: → Extreme Disparity in prediction → Fluctuating Prices → Shortage of Availability	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 retooling There are several techniques that have been tried are: → GRNN → LSTM → Elman Neural Networks Pros: High Accuracy for Short-Term Predictions Cons: Long-Term Predictions cannot be made.	Explore AS, differentiate
	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. As far as the existing solutions are concerned, it is very much evident that they are not accurate enough for long-term prediction. Our solution does not concentrate only on short-term predictions but also strives to provide a realistic view and prediction of the Crude Oil prices in the distant future.	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. → Crude oil is one of the most important commodities in the world, accounting for one-third of global energy consumption. → It is a starting material for most of the products that we use in everyday life. → Crude oil price fluctuations have a great impact on global economies and thus forecasting can assist in minimising the risks associated with volatility in oil prices.	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, calculator usage and benefits; indirectly associated: customers spend free time on volunteering work (i.e. Greenpeace) The behaviour of the customer is assumed as follows: → The provision to select a date is given and if a date is selected, the price can be viewed on that selected date. → There is also a provision in which the price fluctuations can be observed for a range of dates, with the percentage of increase and decrease.	Focus on J&P, fit into BE, understand RC
Identify strong TR & EM	3. TRIGGERS TR What triggers customers to act? I.e. seeing their neighbor installing solar panels, reading about a more efficient solution in the news. Since the customers are spread out across, and are interlinked, seeing the benefits that one yields using our solutions that others are automatically attracted to use our solutions.	10. YOUR SOLUTION SL If you are working on an existing business, write down your current solution first. Fit it 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 behaviors. We are working on building Machine Learning Models based on ARIMA, LSTM and Exponential Smoothing, and simultaneously working on the pre-processing techniques for achieving better accuracy. The best out of these models will be picked and would be proposed as the final solution.	8. CHANNELS of BEHAVIOUR CH 8.1 ONLINE What kind of actions do customers take online? Extract online channels from #7 There is a separate forums for handling queries and addressing the grievances of the customers, which is available 24 x 7. 8.2 OFFLINE What kind of actions do customers take offline? Extract offline channels from #7 and use them for customer development. There is a little to no, offline interaction and support provided to the customers.	Extract online & offline CH of BE
	4. EMOTIONS: BEFORE / AFTER EM How do customers feel when they face a problem or a job and afterwards? I.e. loss, insecure, confident, in control, use it in your communication strategy & design. BEFORE: Lack of knowledge AFTER: Sufficient knowledge to exploit the price fluctuations.			

Problem-Solution fit canvas is licensed under a Creative Commons Attribution-NonCommercial-No Derivatives 4.0 license
 Created by Daria Heprakhina / Amaltama.com

AMALTAMA

4 REQUIREMENT ANALYSIS:

4.1 FUNCTIONAL REQUIREMENT:

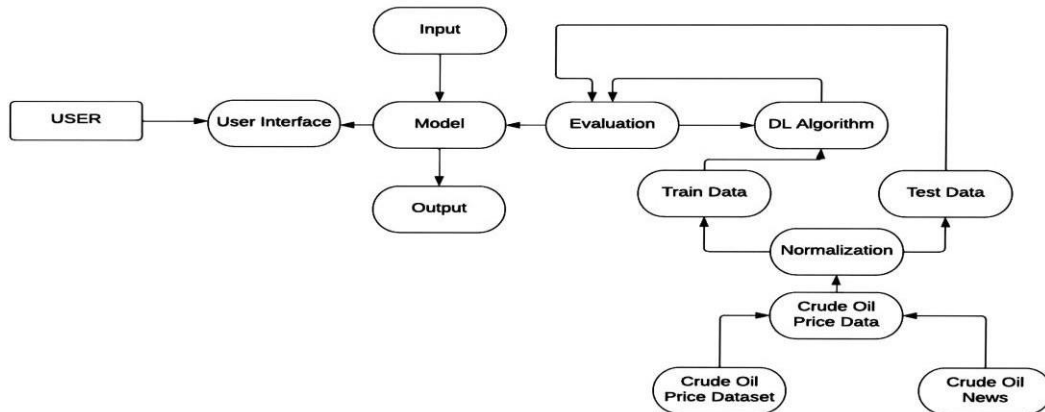
FRNO.	FUNCTIONAL REQUIREMENT	SUB REQUIREMENT
FR-1	User Registration	Registration through Form Registration through Gmail
FR-2	User Confirmation	Confirmation via Email Confirmation via OTP
FR-3	User Enquiry	Enter the date or range of dates
FR-4	User Visualization	Visualize the trend Enquire the prices Analyse the results
FR-5	User Endowment	See the results Gain Knowledge
FR-6	User Utilization	Use it in your idea Close the portal

4.2 NON-FUNCTIONAL REQUIREMENT:

FRNO.	NON-FUNCTIONAL REQUIREMENT	DESCRIPTION
FR-1	Usability	The application interface is easy to use and implement.
FR-2	Security	The credentials are secured and the result is encrypted.
FR-3	Reliability	The accuracy and reliability quotient is quoted to be high.
FR-4	Performance	The performance is uninterrupted and undeterred
FR-5	Availability	The data is freely available and the trend can be manually analysed
FR-6	Scalability	The predictions are scalable and reliable.

5 PROJECT DESIGN

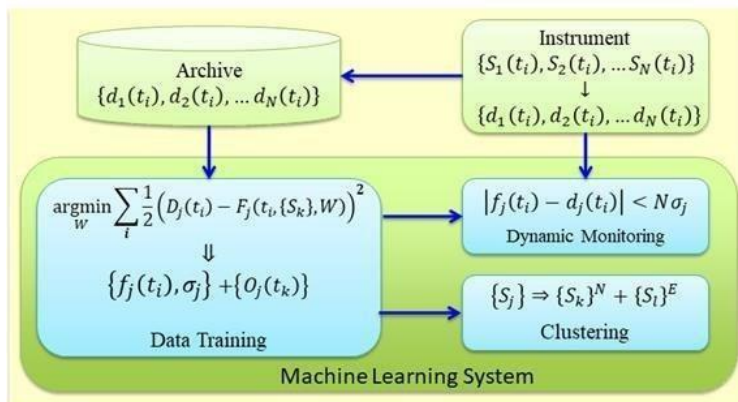
5.1 DATA FLOW DIAGRAM:



5.2 SOLUTION AND TECHNICAL ARCHITECTURE:

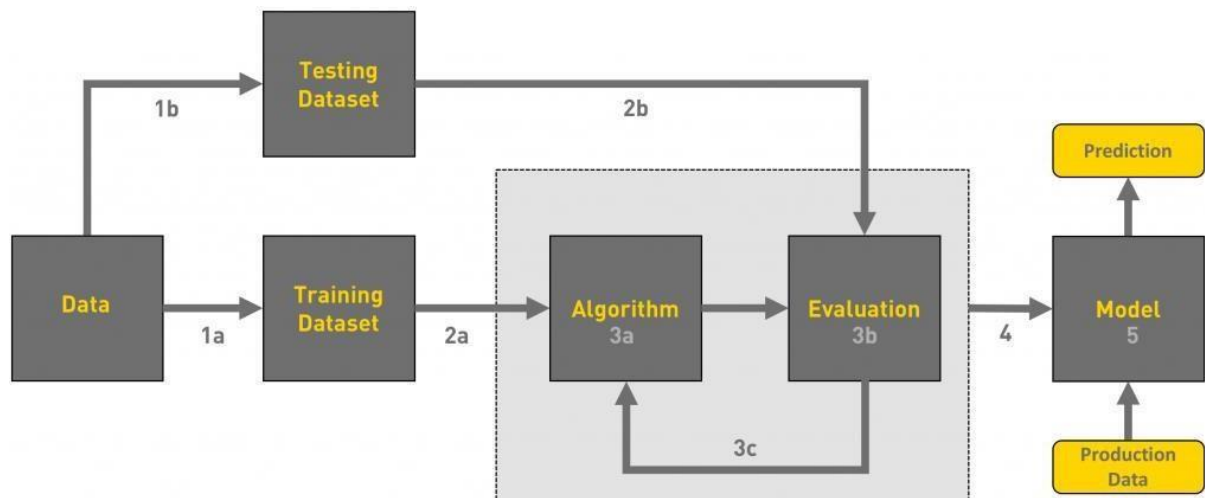
Context View

This view gives a high level representation of the system, the different user types and interactions with external entities. It describes the boundaries of the solution.



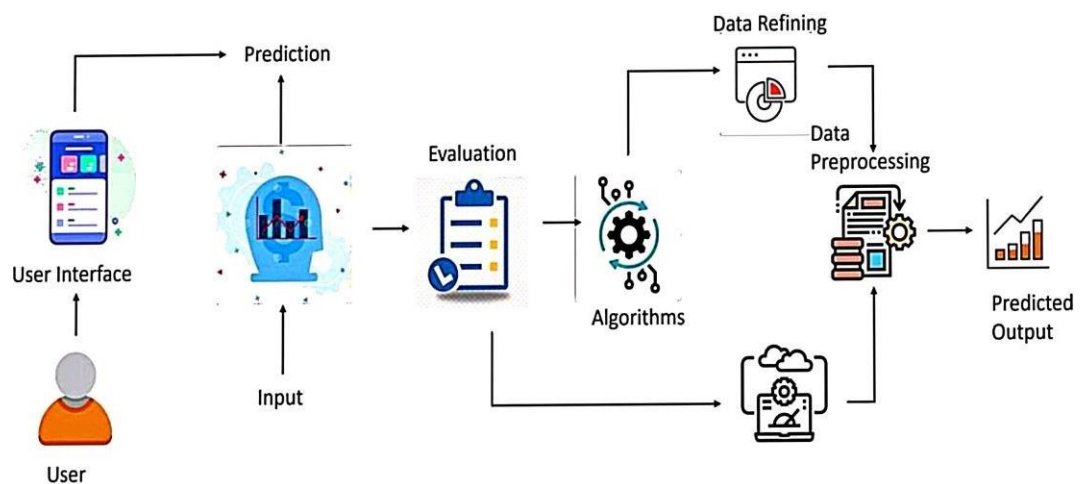
Project View

This section shows how key functionality relevant to the solution architecture maps to releases and milestones.



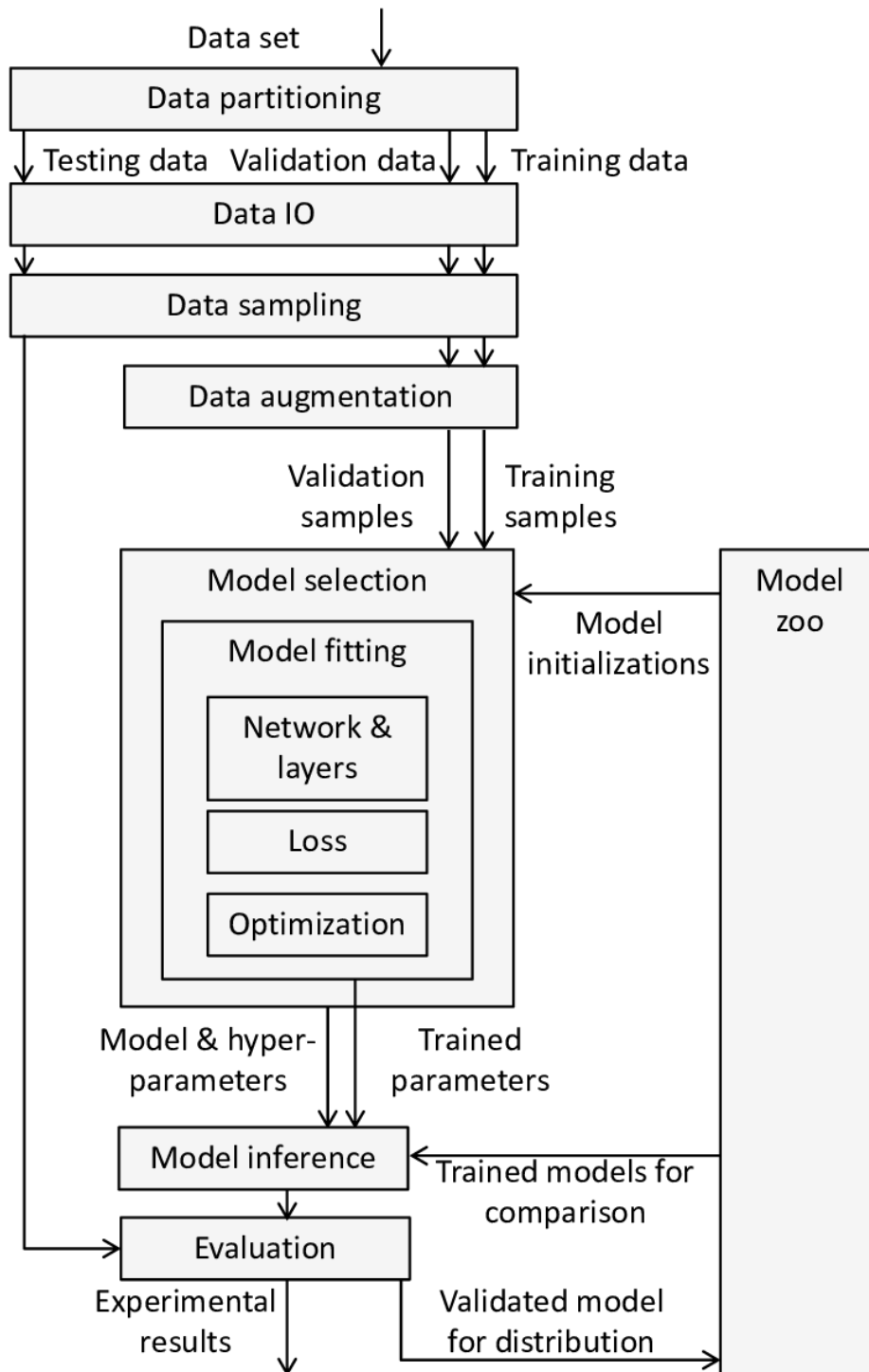
Functional View

This section describes the key functional areas of the project. The goal is to provide context around the architecture – all software performs some functionality and the definition of this functional scope is a very important factor to define the architecture.



Process View

The intent of the process view is to show how the various processing steps within the system fit together to implement the overall functional requirements. This is necessary if the system relies on workflow processes, forked or parallel processing mechanisms. The following processes are significant:



Non-Functional View

This section describes architecturally significant changes that enable the solution to achieve the agreed non-functional requirements (NFRs). Each change is mapped to the corresponding NFR category, which is based on the ISO/IEC 25010-2011 product quality model.

NFRs are documented and maintained in the Non-Functional Requirements Definition and will not be repeated here. In case of duplication, the Non-Functional Requirements Definition takes precedence.

Performance Easy tracking of records and updating can be done. All the requirements relating to performance characteristics of the system are specified in the section below. There are two types of requirements.

1. Static Requirements:

These requirements do not impose any constraints on the execution characteristics of the system. They are:

A) Number of Terminals: The software makes use of an underlying database that will reside at the same system, while the front end will be available to the administrative computer.

B) Number of Users: The number of users can be administrator only, but this software can be extended to applications for almost all staff members of the organization.

2. Dynamic Requirements:

These specify constraints on the execution characteristics of the system. They typically include response time and throughput of the system. Since these factors are not applicable to the proposed software, it will suffice if the response time is high and the transactions are carried out precisely and quickly. Reliability: The software will not be able to connect to the database in the event of the server being down due to a hardware or software failure.

3. Availability:

The software will be available only to administrator of the organization and the product as well as customer details will be recorded by him. He can add customers, update and delete them as well as add new products and manage them.

4. Security:

The security requirements deal with the primary security. The software should be handled only by the administrator and authorized users. Only the administrator has right to create new accounts and generating inventory. Only authorized users can access the system with username and password of administrator

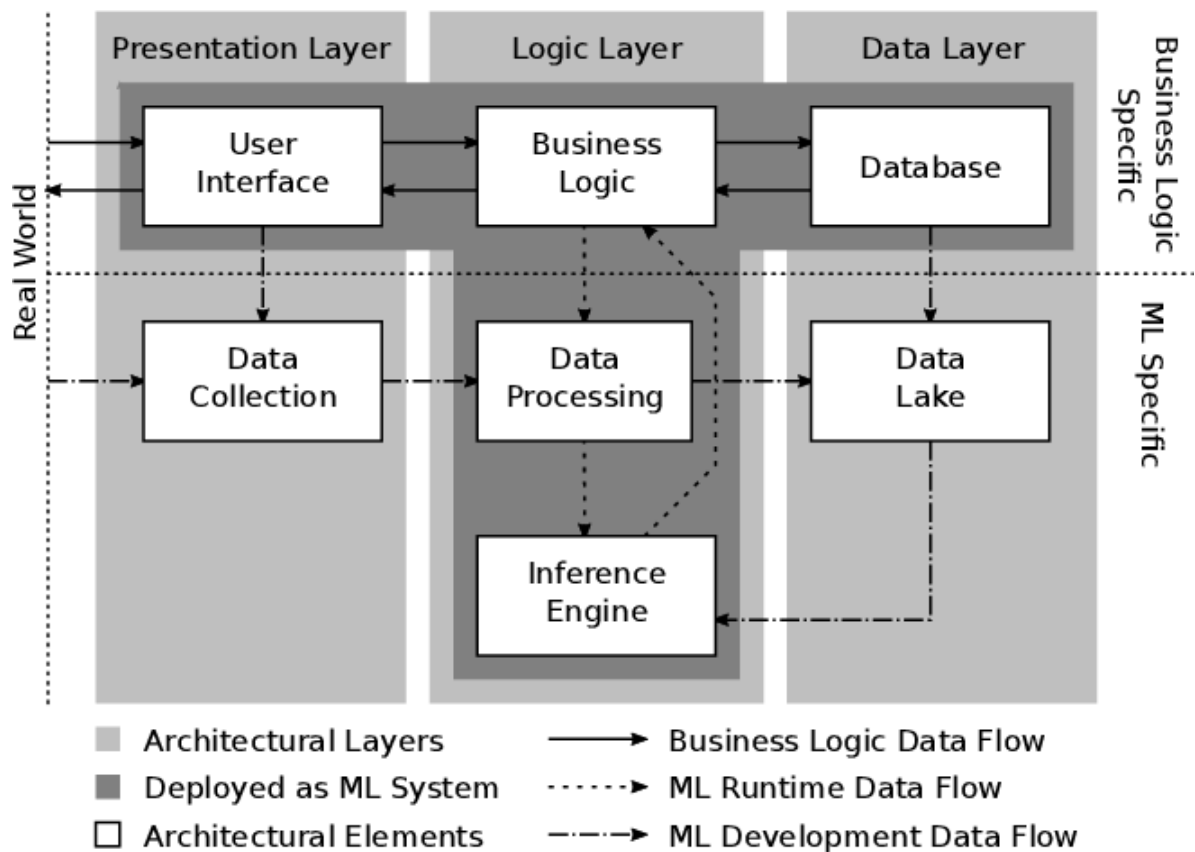
5. Maintainability:

Backups for database are available.

6. Portability:

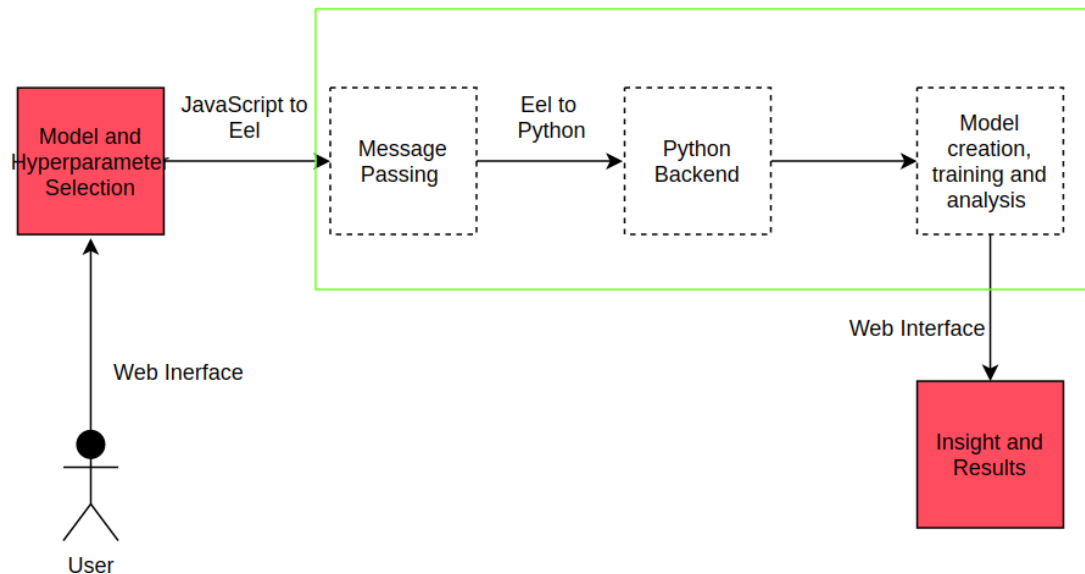
The Software is a web-based application and is built in Python and Nosql so it is platform independent and is independent of operating system.

Logical View



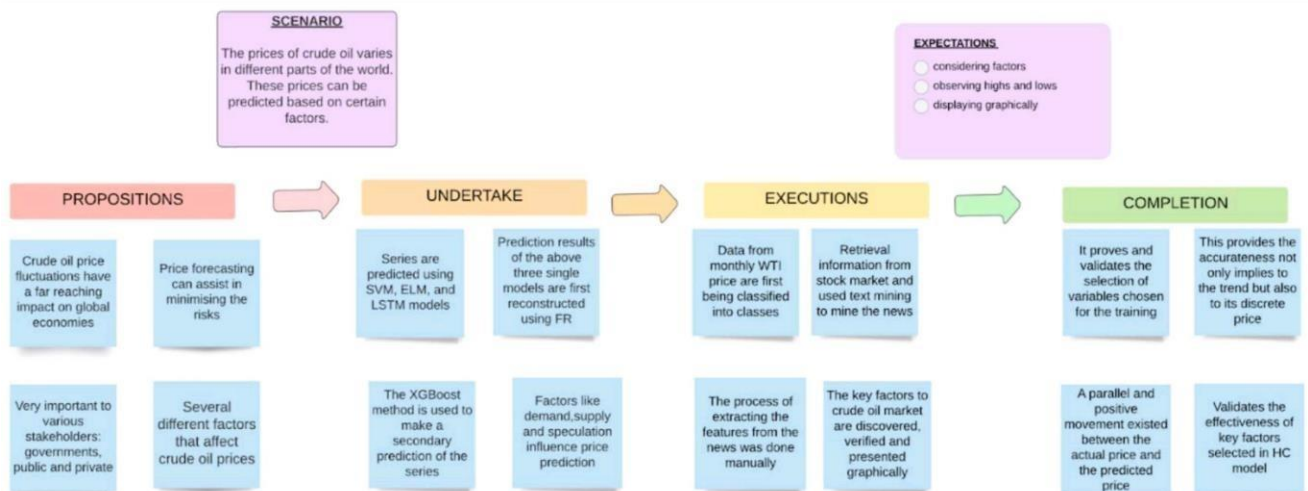
1 Interface View

This section describes the interfaces that will be required to the external system integration touch points

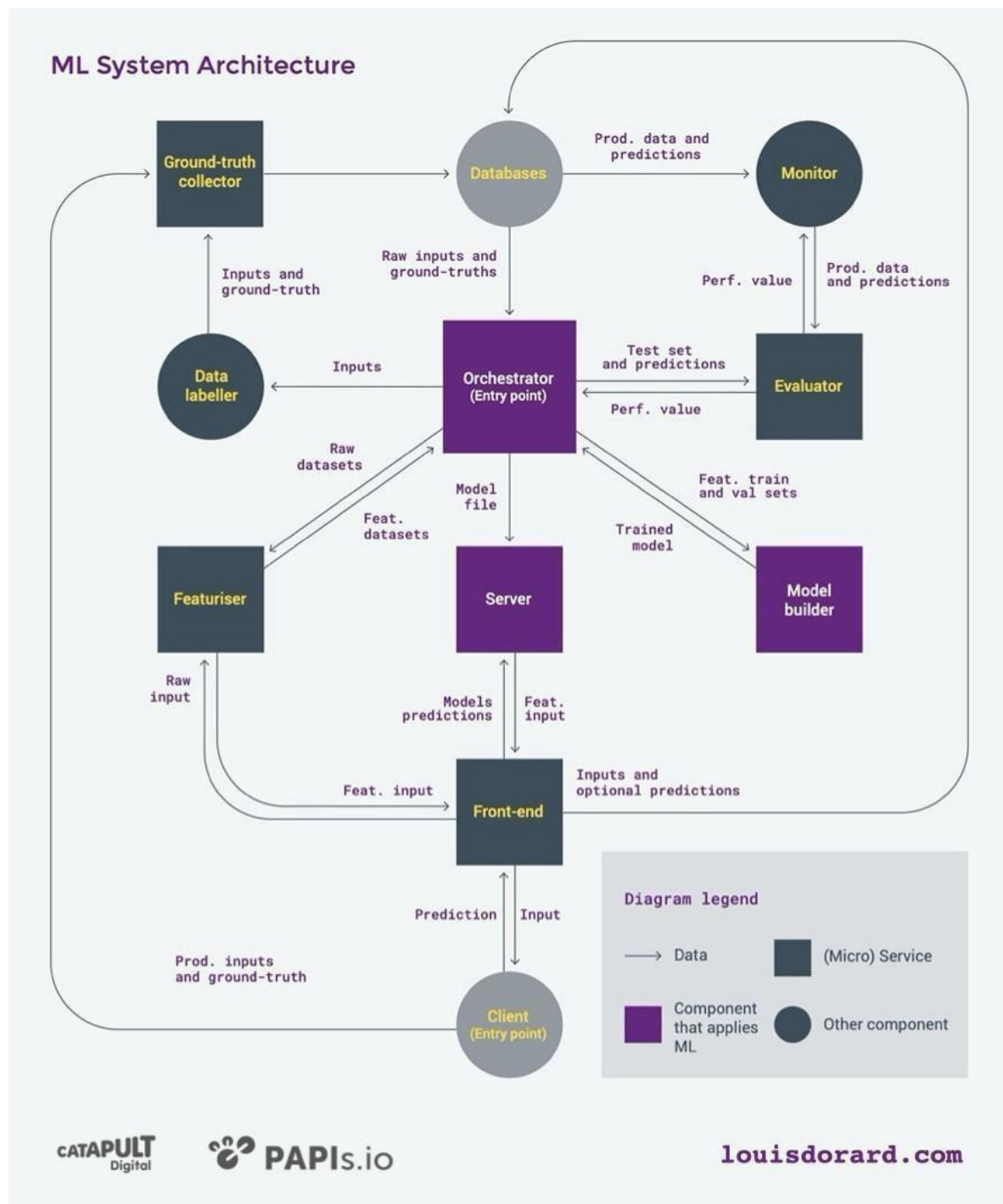


Design View

This section describes and explains any lower-level design concepts arising from the solution if required.

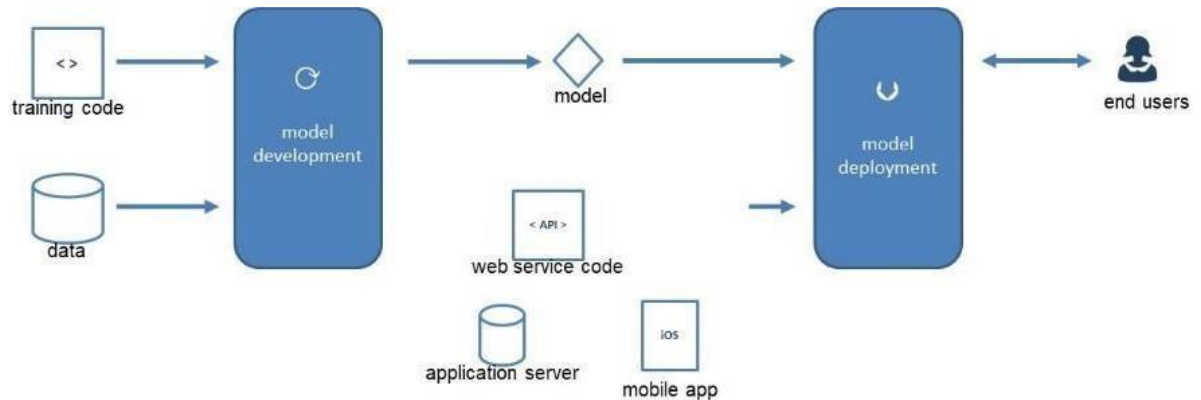


Physical View



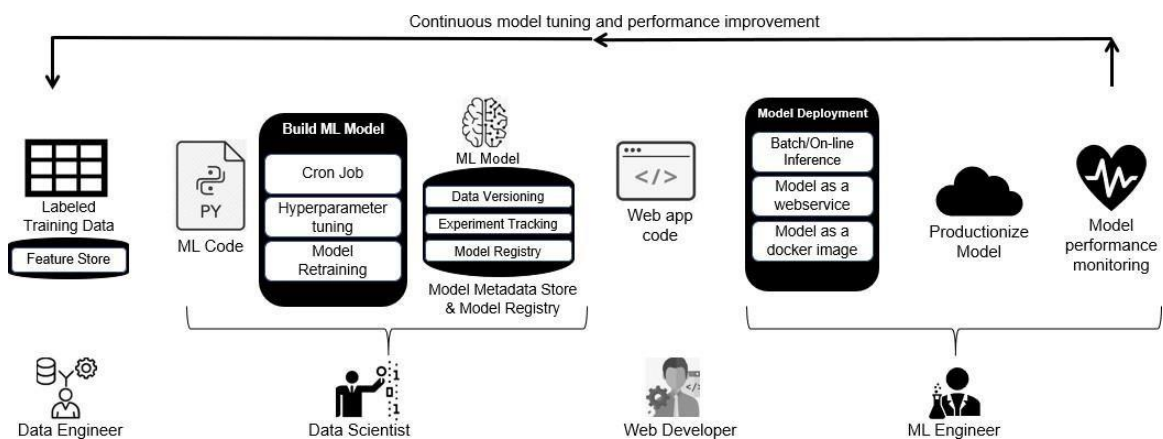
Deployment View

This section describes how code will be deployed in test environments and key considerations for the more complex Production go-live deployment.



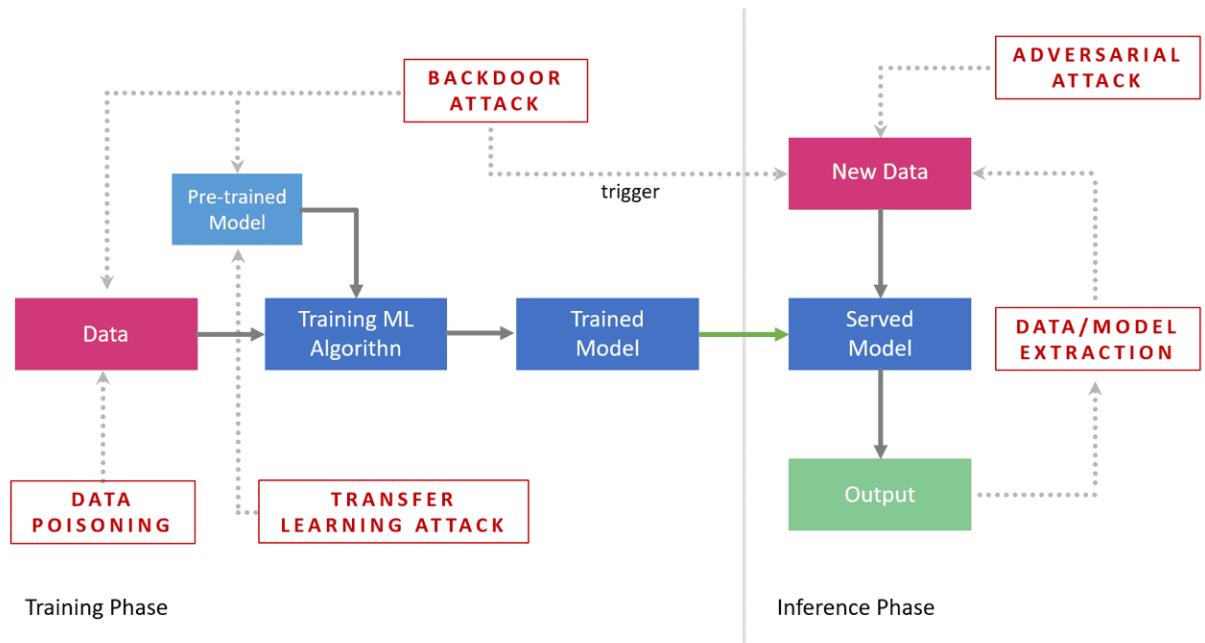
Operational View

This section describes how the architecture will support operational processes and activities.



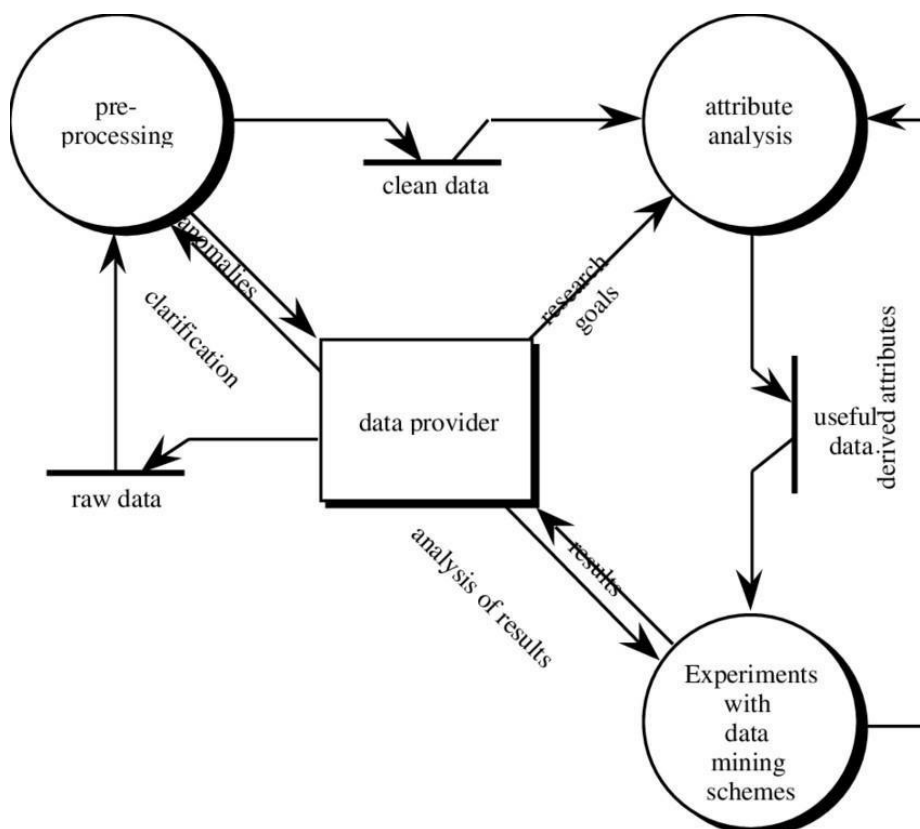
Security View

This section describes how the architecture addresses the different security aspects.



Data View

This section describes the important data model changes required to fulfil the requirements and the associated data flows.



5.3 USER STORIES:

USER TYPE	FUNCTIONAL REQUIREMENT	USER STORY	USER STORY / TASK	ACCEPTANCE CRITERIA	PRIORITY	RELEASE
Customer	Registration	1	Register for the Application through different vendors.	I can access my account / dashboard	High	Sprint-3
	Confirmation	2	Receiving Confirmation Mail	I can receive confirmation email & click confirm	High	Sprint-4
	Login	3	Log in into the application	Access to the account	High	Sprint-2
	Enquiry	4	Enter the range of dates	Plausible Range	High	Sprint-1
	Visualize	5	Visualize the Trend	Accuracy Check	High	Sprint-3
	Endowment	6	See the result	Prediction Check	High	Sprint-1
	Utilization	7	Log Out	Confirmation and Session Closure	High	Sprint-2
Administrator	Authority	1	Verify the imbalances	Session Dryness	High	Sprint-3

6 PROJECT PLANNING AND SCHEDULING:

6.1 SPRINT PLANNING AND ESTIMATION:

SPRINT	FUNCTIONAL REQUIREMENT	USER STORY	USER STORY / TASK	STORY POINTS	PRIORITY	MEMBERS
1	Registration	1	Register for the Application	2	High	2
2	Confirmation	2	Receiving Confirmation Mail	1	Medium	2
2	Login	3	Log in into the application	2	High	2
3	Enquiry	4	Enter the range of dates	2	Medium	2
4	Visualize	5	Visualize the Trend	2	High	2
3	Endowment	6	See the result	2	High	2
4	Utilization	7	Log Out	1	Medium	2

6.2 SPRINT DELIVERY SCHEDULE:

Sprint	Total Story Points	Duration	Sprint Start Date	Sprint End Date (Planned)	Story Points Completed	Sprint Release Date (Actual)
Sprint-1	20	6 Days	24 Oct 2022	29 Oct 2022	20	29 Oct 2022
Sprint-2	20	6 Days	31 Oct 2022	05 Nov 2022	20	05 Nov 2022
Sprint-3	20	6 Days	07 Nov 2022	12 Nov 2022	20	12 Nov 2022
Sprint-4	20	6 Days	14 Nov 2022	19 Nov 2022	20	19 Nov 2022

1. CODING & SOLUTIONING (Explain the features added in the project along with code)

1. A LSTM Price forecasting machine learning model
2. A User Interface for forecasting based on the past 10 days price

2. TESTING

1. Test Cases
2. User Acceptance Testing

3. RESULTS

1. Performance Metrics – RMSE – 2.78

4. ADVANTAGES & DISADVANTAGES

The model is able to forecasting the price accurately based on the past 10 days data.

But whenever the commodity is affected by external factors which are caused naturally then the predictions are bad. This cannot be predicted by machine learning model

5. CONCLUSION

Therefore the ml was deployed as a web app and the user interface is handy for stakeholders who do not have much knowledge in programming. The predicted value is displayed in the user interface.

6. FUTURE SCOPE

The machine learning model can be improved by converting it from univariate to multi variate model for better understanding the price value. This can result in a great boost in the prediction accuracy.