Project Initialization and Planning Phase

Date	8 July 2024
Team ID	SWTID1720096271
Project Title	Machine learning approach for Predicting the price of natural gas
Maximum Marks	3 Marks

Project Proposal (Proposed Solution) template

This project proposal outlines a solution to address a specific problem. With a clear objective, defined scope, and a concise problem statement, the proposed solution details the approach, key features, and resource requirement

s, including hardware, software, and personnel.

Project Overview	
Objective	Develop a Machine learning model that accurately predict future prices of Natural Gas based on time-based features and past prices.
Scope	Data Collection and Preprocessing: • Gathering historical natural gas price data. • Clean and preprocess the data • Generating time-based features (day, month, year, day of the week). Model Development: • Evaluating machine learning algorithms like Random Forest, Gradient Boosting Machines, and Neural Networks. • Training and tuning models using historical price data. Model Evaluation: • Validating model performance using metrics like Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE). Model Deployment: • Deploy the best-performing model for real-time price prediction. • Develop an interface for stake holders to access prediction.
Problem Statement	

<table>
<tr><td>Description</td><td>The natural gas market is characterized by significant price volatility influenced by factors such as weather conditions, economic indicators, and geopolitical events. Accurate price predictions are essential for energy company planners and financial advisors to optimize procurement strategies, manage risks, and provide reliable investment advice.

This project aims to utilize historical natural gas price data to build a machine learning model capable of making accurate price forecasts. The model will incorporate advanced techniques to handle time-series data, including the creation of lagged features, rolling statistics, and relevant time-based attributes.</td></tr>
<tr><td>Impact</td><td>By providing a reliable and accurate forecasting tool, this project will help stakeholders navigate the complexities of the natural gas market, mitigate risks, and capitalize on market opportunities, ultimately leading to better financial outcomes and operational efficiency.</td></tr>
<tr><td colspan="2">Proposed Solution</td></tr>
<tr><td>Approach</td><td>1.Data Collection
2.Data Preprocessing
Reading the Dataset
Handling Missing Values
Label Encoding and One Hot Encoding
Data Visualization
Splitting Dataset into Dependent and Independent Variable.
Splitting Dataset into Train Set and Test Set
3.Model Building
Train the Model with Descision Tree Algorithm
Test the Model
4.Application Building
Build HTML Page
WEB Page
Build Python Code
Run the App
Output</td></tr>
<tr><td>Key Features</td><td>Advanced Feature Engineering:</td></tr>
</table>

	• **Time-Based Features**: Utilize detailed time-based features (day, month, year, day of the week) to capture seasonal and temporal patterns in natural gas prices. • **Lagged Features and Rolling Statistics**: Incorporate past price data through lagged features and rolling statistics (e.g., moving averages, standard deviations) to detect trends and volatility, enhancing predictive accuracy. **User-Friendly Deployment**: • **API and Interface Development**: Develop an intuitive API and user interface using frameworks like Flask allowing users to easily access and interpret predictions in real-time. • **Interactive Visualization**: Provide interactive visualizations for users to explore historical data and forecast results, aiding in better decision-making. **Practical Utility for Stakeholders**: • **Customized Insights for Different Users**: Tailor the solution to provide specific insights for different stakeholders, such as energy company planners and financial advisors, enhancing its practical utility and relevance. • **Risk Mitigation and Opportunity Identification**: Enable users to better navigate market volatility by providing predictive insights that help in mitigating risks and identifying profitable opportunities.

Resource Requirements

Resource Type	Description	Specification/Allocation
Hardware		
Computing Resources	CPU/GPU specifications, number of cores	4 x Intel Xeon cores, 1 x NVIDIA V100 GPU
Memory	RAM specifications	16 GB
Storage	Disk space for data, models, and logs	1 TB SSD
Software		

Frameworks	Python frameworks	Flask
Libraries	Additional libraries	scikit-learn, pandas, numpy, matplotlib, TensorFlow/Keras
Development Environment	IDE, version control	Jupyter Notebook, Git
Data		
Data	Source, size, format	Kaggle dataset, Github Repositories