

PREDICTING THE PRICE OF NATURAL GAS USING MACHINE LEARNING APPROACH

Milestone 1: Project Initialization and Planning Phase

The "Project Initialization and Planning Phase" marks the project's outset, defining goals, scope, and stakeholders. This crucial phase establishes project parameters, identifies key team members, allocates resources, and outlines a realistic timeline. It also involves risk assessment and mitigation planning. Successful initiation sets the foundation for a well-organized and efficiently executed machine learning project, ensuring clarity, alignment, and proactive measures for potential challenges.

Activity 1: Define Problem Statement

Problem Statement: The natural gas market is characterized by significant price volatility due to a variety of complex, interrelated factors such as weather patterns, geopolitical events, supply and demand dynamics, and market speculation. Traditional statistical and econometric methods often fall short in accurately predicting these price movements because they are not well-equipped to handle the non-linear and high-dimensional nature of the influencing variables. Therefore, the specific problem we aim to address is:

How can machine learning techniques be applied to predict natural gas prices more accurately than traditional methods by effectively capturing and modeling the complex relationships among the various influencing factors?

Ref. template: [Click Here](#)

Predicting the price of natural gas Problem Statement Report:

Activity 2: Project Proposal (Proposed Solution)

To address the problem of accurately predicting natural gas prices, we propose a comprehensive solution leveraging advanced machine learning techniques. We use few key features like Data Collection and Preprocessing, Feature Engineering and Selection. Model Development, Model Training and Validation, Model Evaluation and Comparison. Deployment and Visualization, Continuous Improvement and Adaptation, Implementation Plan, Expected Outcomes.

The solution will be generalised to provide real-time prediction estimates. Continuous model training and validation will ensure the system adapts to changing conditions and maintains high accuracy.

Ref. template: [Click Here](#)

Predicting the price of natural gas Project Proposal Report:

Activity 3: Initial Project Planning

Initial Project Planning involves outlining key objectives, defining scope, and identifying stakeholders for a price prediction system. It encompasses setting timelines, allocating resources, and determining the overall project strategy. During this phase, the team establishes a clear understanding of the dataset, formulates goals for prediction accuracy, and plans the workflow for data collection and preprocessing. Effective initial planning lays the foundation for a systematic and well-executed project, ensuring successful outcomes.

Ref. template: [Click Here](#)

Predicting the price of natural gas Initial Project

Planning Report:

Milestone 2: Data Collection and Preprocessing Phase

The Data Collection and Preprocessing Phase involves executing a plan to gather relevant loan application data from Kaggle, ensuring data quality through verification and addressing missing values. Preprocessing tasks include cleaning, encoding, and organizing the dataset for subsequent exploratory analysis and machine learning model development.

Activity 1: Data Collection Plan, Raw Data Sources Identified, Data Quality Report

The dataset for “Predicting the price of natural gas” is sourced from Kaggle. The downloaded data set is not suitable for training the machine learning model as it might have so much randomness so we need to clean the dataset properly in order to fetch good results. This activity includes the following steps.

- Handling missing values
- Handling categorical data
- Handling Outliers

Ref. template: [Click Here](#)

Predicting the price of natural gas Data Collection Report:

Activity 2: Data Quality Report

The dataset for " Predicting the price of natural gas using machine learning approach " is sourced from Kaggle. It includes customer, stock details. Data quality is ensured through verification, addressing missing values, and maintaining adherence to ethical guidelines, establishing a reliable foundation for predictive modeling.

Ref. template: [Click Here](#)

Predicting the price of natural gas Data Quality Report:

Activity 3: Data Exploration and Preprocessing

Data Exploration and Preprocessing for predicting the price of natural gas involves thoroughly examining the collected historical predicting price data to understand its structure, quality, and key characteristics. This phase includes identifying and handling missing values, outliers, and inconsistencies to ensure data integrity. Key features such as price details and external factors like price, performances are extracted and transformed as needed.

Ref. template: [Click Here](#)

Predicting the price of natural gas Data Exploration and Preprocessing
Report:

Milestone 3: Model Development Phase

The Model Development Phase involves selecting appropriate machine learning algorithms training them on the pre processed historical price data. This phase includes evaluating and selecting models(Demand Forecasting, Linear Regression ,KNN ,Decision Tree, XGBoost, Ridge, SVM), hyperparameter tuning, cross-validation, and model evaluation to ensure high prediction accuracy. The development process also involves feature selection and engineering to enhance model performance. Once the model is trained, it is validated using a separate test dataset to confirm its predictive capabilities.

Activity 1: Feature Selection Report

The feature selection process in their model development phase involves evaluating various features such as price,date,year,day.Each feature is assessed for its relevance to the price prediction model,aiming to streamline decision-making and enhance transparency

Ref.Template: [Click Here](#)

Predicting the price of natural gas Feature Selection Report:

Activity 2:Model Selection Report

The Model Selection Report details the rationale behind choosing Random Forest Regressor, Decision Tree, and Linear Regression models for prediction. It considers each model's strengths in handling complex relationships, interpretability, accuracy, adaptability, and overall predictive performance, ensuring an informed choice aligned with project objectives.

Ref. Template: [Click Here](#)

Predicting the price of natural gas Model Selection Report:

Activity 3: Initial Model Training Code, Model Validation and Evaluation Report

The initial model training in the predicting the price of natural gas project, we utilized historical price data to train a gradient boosting regressor. The code included data preprocessing steps such as handling missing values and encoding categorical variables. We split the dataset into training and validation sets, using cross-validation to optimize hyperparameters and ensure robust performance.

Ref. Template: [Click Here](#)

Predicting the price of natural gas Model Development Phase Template:

Milestone 4: Model Development and Tuning Phase

During the model optimization and tuning phase for predicting the price of natural gas, the focus is on refining and enhancing the predictive performance of the models developed in earlier stages. This involves fine-tuning hyperparameters, such as learning rates and tree depths for gradient boosting models, and adjusting architecture and layers for neural networks. Techniques like grid search and randomized search are employed to explore various combinations of hyperparameters efficiently.

Activity 1: Model Optimization Tuning Documentation

The Random Forest Regressor model was selected for its superior performance, exhibiting high accuracy and r^2 score during hyperparameter tuning. Its ability to handle complex relationships, minimize overfitting, and optimize predictive accuracy aligns with project objectives, justifying its selection as the final model.

Ref. Template: [Click Here](#)

Predicting the price of natural gas Model Optimization and Tuning Phase Template:

Activity 2: Performance Metrics Comparison Report

The Performance Metrics Comparison Report contrasts the baseline and optimized metrics for various models, specifically highlighting the enhanced performance of the Random Forest Regressor Model. This assessment provides a clear understanding of the refined predictive capabilities achieved through hyperparameter tuning.

Activity 3: Final Model Selection Justification

The Final Model Selection Justification articulates the rationale for choosing Random Forest Regressor Model as the ultimate model. Its exceptional accuracy, ability to handle complexity, and successful hyperparameter tuning align with project objectives, ensuring optimal loan approval predictions.

Milestone 5: Project Files Submission and Documentation

For project file submission in Github, Kindly click the link and refer to the flow.

[Click Here](#)

For the documentation, Kindly refer to the link. [Click Here](#)

Milestone 6: Project Demonstration

In the upcoming module called Project Demonstration, individuals will be required to record a video by sharing their screens. They will need to explain their project and demonstrate its execution during the presentation.