Final Project Report: Machine Learning Approach for Predicting the Price of Natural Gas

1. Introduction

1.1 Project Overview

The "Machine Learning Approach for Predicting the Price of Natural Gas" project aims to develop a predictive model that accurately forecasts natural gas prices. The project leverages machine learning techniques to analyze historical data and various contextual factors that influence gas prices.

1.2 Objectives

- Develop an accurate machine learning model to predict natural gas prices.
- Enhance forecasting efficiency and precision by incorporating historical and contextual data.
- Provide reliable insights for stakeholders, enabling data-driven decision-making.

2. Project Initialization and Planning Phase

2.1 Define Problem Statement

The current natural gas price prediction models are unreliable, leading to unexpected price fluctuations and customer distrust. Our project aims to develop a robust machine learning model that can predict natural gas prices by analyzing historical data and considering factors like supply, demand, geopolitical events, and economic indicators.

2.2 Project Proposal

This project proposes using advanced machine learning algorithms, specifically Decision Tree models, to improve the accuracy and reliability of natural gas price predictions. By incorporating historical data and external factors, the model will provide more precise forecasts, helping stakeholders make informed decisions.

2.3 Initial Project Planning

Team Members and Roles

- **Neha Sharma:** Problem statement, data collection, data loading, file merging.
- **Vikash Kumar:** Project proposal, handling data issues, managing project records.
- **Yash Bajpai:** Project planning, model training, tuning, comparison, selection, documentation.
- **Pranav Agarwal:** Problem solution report, data splitting, feature scaling, video documentation.

Timeline and Sprint Schedule

- **Sprint-0:** Project Initialization and Planning (2024/07/01 - 2024/07/04)
- **Sprint-1:** Data Collection (2024/07/04 - 2024/07/05)
- **Sprint-2:** Data Preprocessing (2024/07/05 - 2024/07/06)
- **Sprint-3:** Model Development (2024/07/06 - 2024/07/07)
- **Sprint-4:** Model Optimization and Tuning (2024/07/07 - 2024/07/08)
- **Sprint-5:** Project Executable Files (2024/07/08 - 2024/07/10)
- **Sprint-6:** Documentation and Demonstration (2024/07/10 - 2024/07/11)
- **Sprint-7:** Managing Project Record (2024/07/01 - 2024/07/11)

3. Data Collection and Preprocessing Phase

3.1 Data Collection Plan and Raw Data Sources Identified

Data was obtained from sources such as SmartInternz and Kaggle, covering the years 1997 to 2020. The data includes daily natural gas prices and relevant contextual information.

3.2 Data Quality Report

Data quality issues such as missing values and date formatting were identified and addressed. Missing data was handled by dropping rows with null values, and the date column was split into day, month, and year.

3.3 Data Preprocessing

- **Handling Missing Data:** Rows with missing values were dropped.
- **Handling Duplicate Data:** Duplicate data entries were identified and removed.
- **Feature Engineering:** The date column was split into day, month, and year for better analysis.
- **Data Scaling and Encoding:** Data was normalized and encoded for model training.

4. Model Development Phase

4.1 Model Selection Report

The Decision Tree model was selected due to its ability to capture non-linear relationships and its interpretability.

4.2 Initial Model Training Code, Model Validation, and Evaluation Report

Initial model training involved splitting the data into training and test sets, followed by training the Decision Tree model. Model performance was evaluated using accuracy metrics.

5. Model Optimization and Tuning Phase

5.1 Tuning Documentation

Hyperparameters of the Decision Tree model were fine-tuned to optimize performance.

5.2 Final Model Selection Justification

The Decision Tree model was chosen as the final model due to its higher accuracy and better performance compared to other models such as Linear Regression.

6. Results

6.1 Output Screenshots

Screenshots of model predictions, accuracy metrics, and evaluation reports were generated and included in the project documentation.

7. Advantages & Disadvantages

Advantages

- Improved accuracy in natural gas price predictions.
- Ability to incorporate various contextual factors.
- Enhanced decision-making for stakeholders.

Disadvantages

- Model complexity may increase with additional features.
- Requires continuous updating with new data.

8. Conclusion

The project successfully developed a machine learning model for predicting natural gas prices. The Decision Tree model provided accurate and reliable predictions, addressing the shortcomings of existing models. The project demonstrates the potential of machine learning in improving forecasting accuracy and stakeholder decision-making.

9. Future Scope

Future enhancements could include:

- Incorporating more diverse data sources.
- Using advanced algorithms like ensemble methods.
- Real-time data integration for more dynamic predictions.

10. Appendix

10.1 Source Code

The complete source code is available in the project repository.

10.2 GitHub & Project Demo Link

- GitHub Repository
- Project Demo

Date: 11 July 2024
Team ID: SWTID1720359900
Project Title: Machine Learning Approach for Predicting the Price of Natural Gas
Maximum Marks: 4 Marks