Methane Monitoring and Analysis Report

1. Introduction

This report presents an analysis of methane concentration levels measured by sensors at different locations, along with wind data to understand how methane spreads. The goal was to identify high emission areas, predict methane levels, and validate wind data accuracy. I have also suggested practical steps to improve methane management and data accuracy.

2. Objectives

- Understand methane concentration patterns across different locations.
- Identify potential emission hotspots using clustering techniques.
- Predict methane levels using machine learning models.
- Validate wind data accuracy by comparing it with external sources.
- Suggest ways to control methane emissions and improve data accuracy.

3. Data Description

3.1 Methane Data

The methane data was collected from multiple sensors placed at different locations. The main details included:

- Sensor ID: To identify each sensor.
- Timestamp: When the data was recorded.
- Methane Concentration (ppm): The amount of methane detected.

3.2 Wind Data

Wind data was collected from local anemometers and cross-verified with external sources like the OpenWeatherMap API. The key details were:

- Wind Speed (m/s): How fast the wind was blowing.
- Wind Direction (°): The direction in which the wind was blowing.

4. Methodology

4.1 Data Preprocessing

The data was cleaned to handle any missing or inconsistent information. Then, methane and wind datasets were merged using timestamps to align the data correctly.

4.2 Geospatial Analysis

Interactive maps and heatmaps were used to visualize the sensor locations and methane levels. Wind data was also displayed to see how it influenced methane dispersion.

4.3 Predictive Analytics

Different machine learning models were tested to predict methane levels, including:

- Random Forest Regressor

- Gradient Boosting Regressor

- XGBoost: This model performed the best.

- ARIMA: Used for time-series forecasting.

These models were evaluated based on their ability to predict methane levels accurately using metrics like Mean Squared Error (MSE) and R-Squared Score (R^2).

4.4 Clustering and Interpolation

To find areas with higher methane concentrations, the following techniques were used:

- DBSCAN: Effective for detecting outliers and hotspots.

- KMeans: Helped in segmenting different emission areas.

For interpolation, two methods were tested:

- IDW (Inverse Distance Weighting): Quick but not very smooth.

- Kriging: Gave smoother and more reliable results.

4.5 External Validation

To verify if the wind data was reliable, local wind sensor readings were compared with data from the OpenWeatherMap API. Some differences were found, suggesting that the sensors might need recalibration.

5. Key Findings

5.1 Methane Concentration Patterns

- Some areas consistently showed high methane levels, indicating potential emission sources.

- Wind direction played a significant role in spreading methane across different locations.

5.2 Model Performance

- XGBoost was the most accurate model for predicting methane levels.

- ARIMA effectively captured time-based trends in the data.

5.3 Clustering Insights

- DBSCAN was useful for spotting outlier regions with high methane concentrations, possibly pointing to leaks.

- Kriging provided the best spatial maps, showing smooth transitions in methane levels across different areas.

5.4 Wind Data Accuracy

- The comparison between local wind data and API data showed slight differences, suggesting that some sensors might need recalibration.

6. Challenges Faced

- Data Gaps: Some timestamps had missing data, affecting analysis accuracy.

- Sensor Calibration: Differences between local and external wind data indicated a need for regular calibration.

- Computational Demands: Techniques like Kriging and XGBoost required significant computational power.

7. Recommendations

1. Regular Sensor Calibration: Routine calibration against reliable external sources to ensure data accuracy.

2. Targeted Emission Control: Focus on areas identified as hotspots for detailed inspections and mitigation efforts.

3. Increase Sensor Density: Adding more sensors in high-risk areas could improve spatial coverage and analysis accuracy.

4. Optimize Computational Resources: Optimizing algorithms or using more efficient data handling techniques might help manage high computational demands.

5. Incorporate More Data Sources: Including additional external data sources like satellite data could improve validation accuracy.

8. Conclusion

This analysis shows that wind patterns significantly influence methane dispersion, and accurate wind data is crucial for reliable methane monitoring. Addressing the identified challenges such as sensor calibration and data gaps can significantly improve the accuracy of methane monitoring. Implementing the recommendations can help control methane emissions better and enhance the reliability of the data used for analysis.

Prepared by: Akanksha Kumari

Date: 02/03/2025