HeatEx FanGuard: Predictive Maintenance System

1. Introduction

HeatEx FanGuard is a predictive maintenance solution designed to optimize the operation of heat exchange systems by monitoring IoT sensor data in real-time and predicting potential failures. This system helps in reducing maintenance costs, preventing unexpected downtime, and ensuring the smooth operation of critical industrial equipment.

1.1 Business Problem

Heat exchange systems are essential in various industries, such as manufacturing, chemical processing, and HVAC systems. Unplanned downtime due to system failures can lead to significant financial losses, operational disruptions, and safety hazards. Traditional maintenance approaches are either reactive (fix after failure) or preventive (based on time or usage), both of which can be costly and inefficient.

Objective: HeatEx FanGuard aims to develop a predictive maintenance system that forecasts potential failures using real-time sensor data, enabling proactive maintenance and reducing operational costs.

2. Solution Overview

HeatEx FanGuard utilizes the Google Cloud Platform (GCP) for its end-to-end pipeline, from data ingestion to model deployment and monitoring.

2.1 Architecture Diagram

The architecture involves:

- Google Cloud Pub/Sub: Real-time ingestion of IoT sensor data.

- Google Cloud Dataflow: Real-time processing and feature engineering.
- Google BigQuery: Storing processed data for analysis and training.
- Vertex AI: Training and deploying machine learning models.
- Google Cloud Storage (GCS): Storing model artifacts and temporary data.
- Google Cloud Monitoring: Monitoring model performance and setting up alerts.

3. Data Ingestion and Processing

3.1 Data Source

The data is generated from IoT sensors installed on heat exchange systems. Each sensor captures various parameters, including:

- Timestamp: The time of data recording.
- Sensor ID: Unique identifier for each sensor.
- Temperature: Temperature reading in Celsius.
- Pressure: Pressure reading in Pascals.
- Humidity: Humidity percentage.
- Velocity: Airflow velocity in meters per second.
- Status: Indicator of sensor health (e.g., normal, warning, critical).

3.2 Data Ingestion

Data is ingested in real-time using Google Cloud Pub/Sub. The script 'publish_sensor_data.py' publishes synthetic sensor data to a Pub/Sub topic ('sensor-data-topic').

3.3 Data Processing

Data processing is handled by a Google Cloud Dataflow job that reads data from Pub/Sub, transforms it, and writes it to BigQuery. Key transformations include:

- Converting temperature from Celsius to Fahrenheit.
- Calculating rolling averages for temperature.
- Detecting anomalies based on threshold values.
- Creating new features like temperature_delta.

4. Exploratory Data Analysis (EDA)

The processed data stored in BigQuery is analyzed to understand patterns and generate insights that inform feature engineering and model development.

4.1 Data Distribution Graphs

- Temperature Distribution Before Preprocessing (Graph Included)
- Temperature Distribution After Preprocessing (Graph Included)
- Pressure Distribution Before Preprocessing (Graph Included)
- Pressure Distribution After Preprocessing (Graph Included)

4.2 Correlation Heatmap

- Correlation Heatmap After Preprocessing (Graph Included)

The correlation heatmap shows the relationships between different sensor readings, helping identify features that may influence the predictive model.

5. Feature Engineering and Model Development

5.1 Feature Engineering

Feature engineering includes creating new features that capture trends and patterns in the data:

- Rolling Averages: Calculating rolling averages for temperature to identify gradual changes.
- Temperature Delta: Difference between current temperature and the rolling average.
- Anomaly Flags: Binary flags indicating potential anomalies based on threshold values.

5.2 Model Selection

Multiple machine learning models are trained and evaluated for performance:

- Random Forest Classifier: A robust model for tabular data, optimized using GridSearchCV.
- Gradient Boosting Classifier: A powerful model for handling imbalanced datasets, also tuned using GridSearchCV.

10. Conclusion and Future Work

HeatEx FanGuard provides a scalable and effective solution for predictive maintenance using state-of-the-art cloud infrastructure and machine learning models. The system can be further enhanced by integrating more data sources, such as vibration sensors or external environmental data.

Future Enhancements:

- Integrate Additional Sensors: Incorporate more data points for enhanced predictive accuracy.
- Implement Deep Learning Models: Experiment with deep learning architectures for complex feature extraction.
- Automate Model Retraining: Use Cloud Scheduler to periodically retrain models with new data.