Predicting Equipment Failure and Scheduling Maintenance Proactively

Project Deliverable

- A GitHub repository with a python file (.py) with your solution.

Problem Statement

Equipment failure is a major cause of downtime in the telecommunications industry, which can result in significant financial losses and customer dissatisfaction. To minimize downtime and ensure optimal performance, it is crucial to identify potential equipment failures and schedule maintenance accordingly proactively. This requires the collection and analysis of large amounts of data generated by various equipment and network sensors.

The deliverable for this project is a data pipeline that can efficiently collect, clean, and analyze equipment and network sensor data. The pipeline should be designed to identify potential equipment failures and schedule maintenance proactively, minimizing downtime and improving overall equipment performance. The data pipeline will be built using Python and PostgreSQL and with the Postgres database hosted on Google Cloud.

Guidelines

Here are some guidelines and hints to help you create the data pipeline:

- **Data Extraction**: The data pipeline should be designed to collect data from various sources, including network sensors, equipment sensors, and maintenance records. Sample datasets for data extraction will be provided by the client and should be used for building the pipeline.
- **Data Transformation:** The collected data must be cleaned and transformed to ensure consistency and quality. This will involve removing duplicates, fixing missing data, and normalizing the data for consistency. You can also explore the following techniques:
    - Aggregation: Summarizing data into useful metrics such as the total number of equipment failures, average time between failures, etc.
    - Joining: Combining multiple datasets based on common fields or keys to create a unified view of the data.
    - Data enrichment: Combining internal data with external data sources such as weather data or other publicly available datasets to gain additional insights.
- **Data Analysis:** The cleaned data will be used to build machine learning models that can predict potential equipment failures and schedule maintenance proactively. The models will be designed to analyze equipment and network sensor data in real time to identify

anomalies and predict potential failures. You don't need to implement this step in the data pipeline.

- **Data Loading:** The resulting data will be stored in a PostgreSQL database.

Sample Datasets for Data Extraction

Sample datasets (https://bit.ly/3YNdO2Y) will be provided by the client for data extraction. The datasets will include equipment sensor data, network sensor data, and maintenance records. The datasets will be in CSV format and will include the following fields:

- Equipment sensor data: ID, date, time, sensor reading
- Network sensor data: ID, date, time, sensor reading
- Maintenance records: ID, date, time, equipment ID, maintenance type