Raul G. Garcia III

Dr. Jeffry Babb, Ph.D.

Business Intelligence & Decision Support Systems – 4310

28 July 2024

Developing a Predictive Flight Delay Dashboard

In today's fast-paced aviation industry, maintaining punctuality is paramount for customer satisfaction and operational efficiency. As a member of the Business Intelligence (BI) team, I have been assigned the crucial task of developing a reporting dashboard to monitor and predict flight delays. This paper outlines the steps to collect data, train an Azure Machine Learning (ML) model, export endpoints to load trained CSV data into PowerBI, and create visualizations to identify root causes of flight delays, ultimately aiming to implement operational improvements for carriers.

Capturing historical flight data on flights that departed within the last 60 minutes, will allow us to train the ML model to generate predictive information. This data, received hourly, contains flight details such as origin, destination, flight number, and departure delay. The dataset is timestamped in Coordinated Universal Time (UTC) for consistency. To ensure the accuracy and completeness of our data, we employ a robust data pipeline that automates the extraction, transformation, and loading (ETL) process. The pipeline extracts data from various sources, including airline databases and air traffic control systems, transforms it into a standardized format, and loads it into our centralized data warehouse.

Once we have amassed a significant amount of historical flight data, the next step is to train an Azure ML model to predict flight delays. The model will utilize features such as departure time, carrier, origin, destination, and historical delay patterns. Here’s a detailed outline of the process:

1. **Data Preprocessing**: Cleanse the data to handle missing values, remove outliers, and encode categorical variables. Split the data into training and testing sets to evaluate the model’s performance.
2. **Feature Engineering**: Create new features that could enhance the model’s predictive power. For instance, weather conditions, day of the week, and holiday schedules might impact flight delays.

3. **Model Selection**: Azure ML’s automated machine learning capabilities can assist in identifying the most suitable model.
4. **Training and Validation**: Train the model on the training dataset and validate its performance on the testing dataset. Use metrics such as accuracy, precision, recall, and F1-score to evaluate the model.
5. **Deployment**: Once the model achieves satisfactory performance, deploy it as a web service in Azure. This deployment will generate endpoints that can be consumed by other applications.

With the model deployed, we need to integrate its predictions into PowerBI for visualization. This involves the following steps:

1. **Data Export**: Use the generated endpoints to fetch predictions from the Azure ML model. The predictions, along with the original flight data, are exported as CSV files.
2. **PowerBI Integration**: Import the CSV files into PowerBI. PowerBI’s data connectivity features facilitate seamless integration with Azure ML endpoints, allowing real-time data updates, or you can connect to the CSV file from your machine directory.

Creating insightful visualizations is critical for identifying root causes of flight delays and implementing improvements. Here’s how we approach this in PowerBI:

1. **Dashboard Design**: Design an intuitive dashboard that displays key metrics such as the percentage of delayed flights, average delay time, and delay distribution by carrier, origin, and destination.
2. **Interactive Visualizations**: Implement interactive charts and graphs that allow users to drill down into specific time periods, carriers, and routes. For example, a heatmap can highlight routes with the highest delay frequencies.
3. **Root Cause Analysis**: Use visual analytics to identify patterns and correlations. For instance, a line graph showing delay trends over time can reveal peak delay periods. A bar chart comparing delays across carriers can pinpoint underperforming airlines.

4. **Operational Improvements**: Based on the insights gained, collaborate with carriers to develop and implement strategies to reduce delays. This might involve optimizing flight schedules, improving ground operations, or enhancing communication with air traffic control.

Building a predictive flight delay dashboard involves a systematic approach to data collection, model training, and visualization. By leveraging Azure ML and PowerBI, we can gain actionable insights into the root causes of flight delays and work towards improving operational efficiency. This initiative not only enhances customer satisfaction but also strengthens the overall reliability of our flight schedules, ensuring a smoother and more punctual travel experience for passengers.