Fraud Prevention on Vending Machines

By Yonatan Rojas
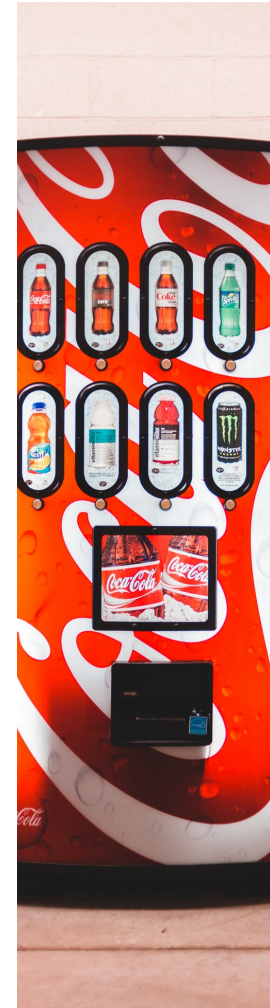
The Problem

- ✔ **Fraud within vending machine operations and the associated delivery routes poses a significant challenge for companies.**

- ✔ **Unauthorized access, tampering, and theft compromise the integrity of transactions and impact the overall efficiency and profitability of the vending business.**

Possible Solution

- Machine learning can provide solutions by analyzing patterns and anomalies in data to predict potential vulnerabilities.

- By leveraging historical data on transactions, machine statuses, and driver routes, the system can identify unusual activities and trigger alerts or interventions to prevent fraudulent behavior.

- Research on similar applications of machine learning in fraud detection, such as anomaly detection algorithms and predictive modeling, will inform the project approach.

Impact

The anticipated impact of the project includes a reduction in financial losses due to fraud, increased operational efficiency, and enhanced customer trust.

Quantifying the scale of the problem will involve assessing the historical losses attributed to fraud and estimating potential savings through the implementation of a successful fraud detection system.

EDA

*Data to be provided

Data Processing

Data Cleaning

Remove any duplicate entries, missing values, or outliers from the transaction, machine status, and route data to ensure data quality. Normalization/Standardization: Scale numerical features to a standard range to improve the performance of machine learning algorithms.

Feature Encoding

Convert categorical variables, such as product types or machine IDs, into numerical representations.

Temporal Aggregation

Aggregate transaction and machine status data over specific time intervals (e.g., hourly or daily) to identify patterns and trends.

Feature Engineering

Transaction Features

Create features such as transaction frequency, total sales per machine, and average transaction value to capture the vending machine's usage patterns.

Route Features

Derive features like total distance traveled, number of stops, and deviations from planned routes to analyze driver behavior and route efficiency.

Time-Based Features

Extract temporal features such as day of the week, time of day, and seasonal trends to account for temporal variations in vending activity.

Anomaly Indicators

Generate features to flag anomalies, such as sudden drops in transaction volume or irregular machine status updates, which could indicate potential fraudulent activities.

Baseline Modeling

Algorithm Selection

Choose baseline machine learning algorithms suitable for the problem, such as logistic regression, decision trees, or random forests, based on the nature of the data and the desired interpretability of the model.

Model Training

Split the preprocessed data into training and testing sets to train the baseline models on a subset of the data and evaluate their performance on unseen data.

Model Comparison

Compare the performance of different baseline models to establish a benchmark for further model improvement and optimization

That's it!
Thank you.