Default prediction model

Objective:

Our objective was to create a model that enables the company to predict client defaults, by leveraging the pump's IoT data and combining it with the client's payment data.

The objective of this project was to develop a predictive model to help SunCulture identify potential client defaults. By combining IoT data from pumps with client payment data, the model provides actionable insights to enable proactive measures, such as customer outreach or payment plan adjustments.

Processing Financial Information

We were given a database with the following information:

- Device ID: unique identifier of the pump
- Expected amount: expected pay
- Expected date: expected date of pay
- Amount paid
- Timestamp: date of payment

To use this data, we aggregated by grouping according to:

- Device ID
- Year
- Month

This way we can get an understanding of the financial performance of each pump over time.

Understanding customer defaults:

- Defaults were defined based on the relationship between expected payments and actual payments made by the pumps**.**
- **Cumulative Default:** This represents whether the total (cumulative) payments made up to the current month are less than the cumulative expected payments up to that month. For this, we calculated cumulative sums for both the expected amounts and the actual payments made. This involved aggregating payments month by month for each pump, allowing us to see not just individual monthly performance but also the broader financial picture over time.
- **Month Default:** evaluates whether the payments made in the current month are less than the expected payments for that same month.

- **New Default Status**: Highlights devices that have transitioned into cumulative default status in the current month compared to the previous month.

Processing IoT Data

The IoT dataset provided telemetry data from each pump, reflecting usage patterns and operational behavior. We conducted to main tasks:

- **Filtering**: Relevant data was selected to align with the financial dataset's timeline and scope.
- **Aggregation**: Device usage metrics were summarized by month to capture trends and patterns for each pump.

Combining IoT and payment data

The predictive model integrates aggregated IoT metrics with payment history. By analyzing the relationships between device performance and payment behavior, the model identifies patterns that are predictive of default.

Model Overview

The model uses a Random Forest Classifier predictor. The features incorporated are the following:

- Days with signal over the past 30 days
- Days with signal over the past 60 to 90 days
- Month
- Total active records
- Average motor speed
- Latitude
- Longitude

These are used to predict new default status.

Results

The model achieves an AUC of 0.72.

The predictions enable SunCulture to take early action by categorizing customers into three risk tiers:

- **High Risk**: High likelihood of default.
- **Medium Risk**: Moderate likelihood of default.
- **Low Risk**: Low likelihood of default.

The model provides a list of customers sorted by risk, prioritizing those most likely to default. This ranking helps SunCulture allocate resources efficiently and focus outreach efforts on the highest-risk customers.

The thresholds for the risk tiers are easily adjustable, allowing the company to tailor the model to changing business needs or customer engagement strategies. For the initial implementation, we suggested thresholds based on discussions with the client, emphasizing a strategy to maximize true positives.