

# Executive Summary of Propensity to Pay Project

## Team All Stars

**Problem Description:** Whether or not customers pay their debts on time to a company is crucial to its financial well being. With the strength of companies being tied closely to their capital, the ability to foresee cash flows from accounts receivable would be an invaluable tool for suggesting proactive policy decisions. The problem is to identify what factors influence customer propensity to pay and how long it takes defaulters to settle.

**Objective:** Given the accounts receivable data of invoices, the goal is to develop a dynamic classification model using machine learning to predict whether a customer will default and in what timeframe. The timeframe element allows for additional customer segmentation based on whether they will default for 0-30, 30-60, or 60+ days.

**Methodology:** The project lifecycle consisted of the following:

- Exploratory data analysis - developed a preliminary understanding of columns and distributions
- Data processing - cleaned the dataset by identifying and fixing issues such as duplicates, outliers, data corruption, and data type transformation
- Feature engineering - generated variables to enhance the dataset like public holiday and weekend indicators as well as year/month/date variables
- Modeling - trained/tested random forest and LightGBM model while noting their performance and conducting hyperparameter tuning by gridsearch
- Evaluation - developed visualizations in Tableau facilitating analysis of results to develop business insights and strategies

## Observations and Recommendations:

- Most customers who defaulted on their payment had an expected clearing date on Sunday. A solution to this is to have weekday clearing dates exclusively. This was confirmed later by our model as well since Sunday was third in variable importance.
- Third party payments and no payment methods accounted for the highest concentration of defaulters (more than 70%). We advise that prospective debtors that request these payment methods be vetted with more scrutiny or in some cases excluded completely.
- Customers who default more than 60 days have the highest loan amounts on average. Cash flows are more significantly affected when these high value customers default. Besides reviewing these customers more closely, an additional clause should be added to their contracts that incurs heavy fines for defaulting for 60+ days.:
- Post-modeling analysis displays that the large majority of defaulters settled their balance within 30 days after the due date. An approach to mitigate these instances, our model would deem quite useful. After a prospective client is categorized as a potential defaulter, it is our recommendation that their loan period be extended by 30 days to increase the likelihood of timely settlement.
- Though region was not initially deemed much bearing, modeling showed it was a highly important variable in prediction. Given the high relevance of regions, it is our recommendation that problematic regions go through inspection of their lending process to identify possible causes for their high default rates. However, specific policy advice requires identification of these regions to allow for finer analysis.