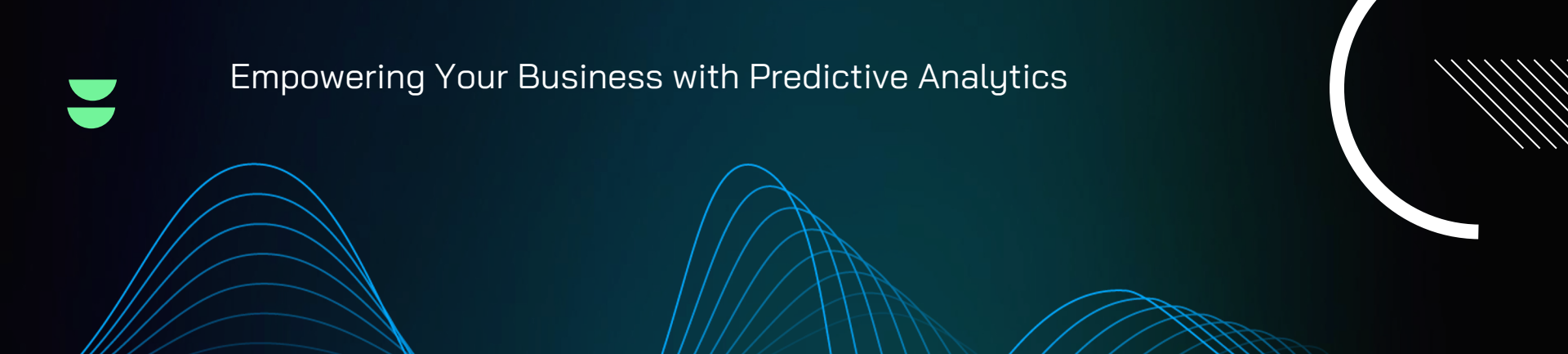




Identification of High Value Prospects with a Data-Driven Model

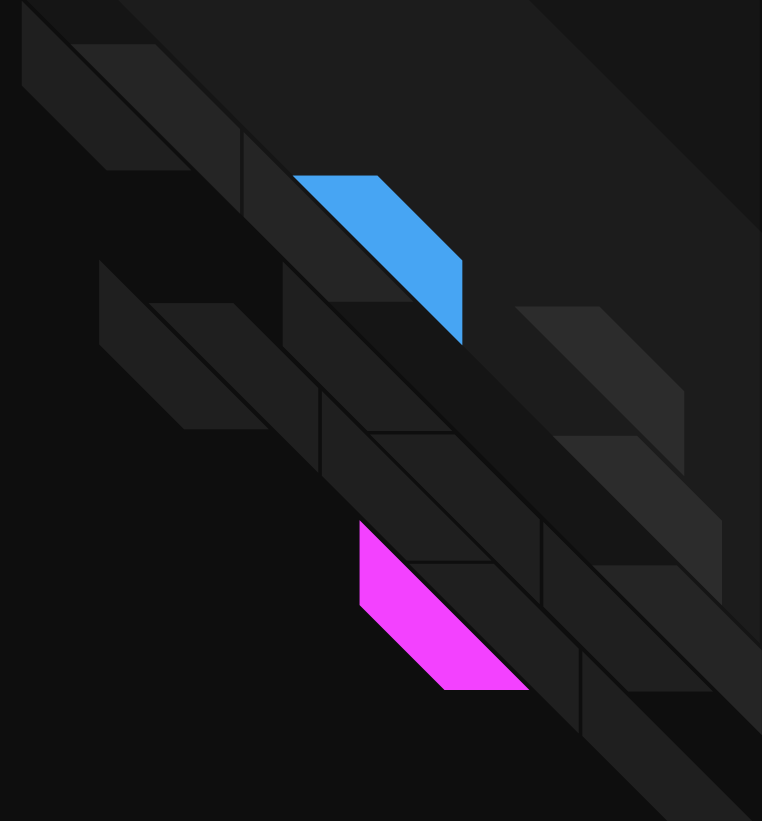
Branata Kurniawan (34534388)
Ng Zhi Hui (34557319)



Empowering Your Business with Predictive Analytics

AGENDA

1. Understanding the problems
2. Data and Knowledge Acquisition
3. Machine Learning Algorithms
4. System Development Journey
5. Revision and Evaluation of the System
6. Integration and Maintenance Plan



Understanding the problems



01

The COVID-19 pandemic has significantly impacted consumer behaviour, with many customers shifting their purchasing habits and priorities

- ❑ Creating new opportunities and challenges for businesses looking to engage with their customers effectively

02

Top spenders are generally determined by business organisations through **domain expert knowledge** of the salesman

- ❑ Introduce bias into the selection process



Data and Knowledge Acquisition

Here are the data that we are using to build our model:

- Customer Acquisition (Demographic Information)
 - Age
 - City
 - Income
- Customer's Spending (Transaction Information)
 - Spending Period
 - Purchase Type
 - Amount
- Customer's Repayment
 - Repayment Period
 - Repayment Amount

Data Preparation

Data Cleaning

- Identify missing, duplicate, outliers data & standardization

Data Processing

- Creating new variables
 - Average Spending
 - Average Repayment
 - Repayment Rate
 - Credit Spending Ratio (Spent : Credit)
- Data Encoding

Machine Learning Algorithms

Based on the data that we acquired, we will be building the model using machine learning algorithms, which are:

- Logistic Regression algorithm
- Decision Tree algorithm
- Random Forest algorithm



System Development Journey



1. Prototype

This include choosing the programming language, pre-process the data and building the working prototype.

2. Construction

Using the prototype as the base, we explored different ways to improve the model. (e.g. introduce more performance metric, pruning, and also estimation performance boost).



4. Revision & Evaluation

Lastly, using the output generated, we choose the best model and evaluate what can be better improve.

3. Testing

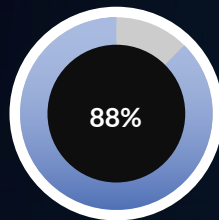
In this phase, we tested the complete system to make sure that it meet the requirements.



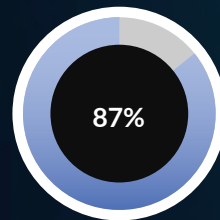
System Evaluation and Enhancement

During the evaluation, we concluded that Random Forest model by considering:

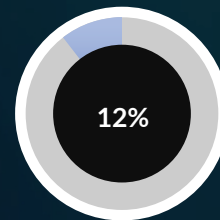
- 1) Accuracy
- 2) Area Under Curve (AUC)
- 3) Minimum Squared Error (MSE)
- 4) Precision Rate
- 5) Recall Rate



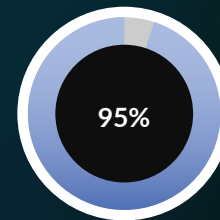
Accuracy Score



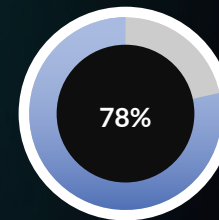
AUC Score



MSE Score



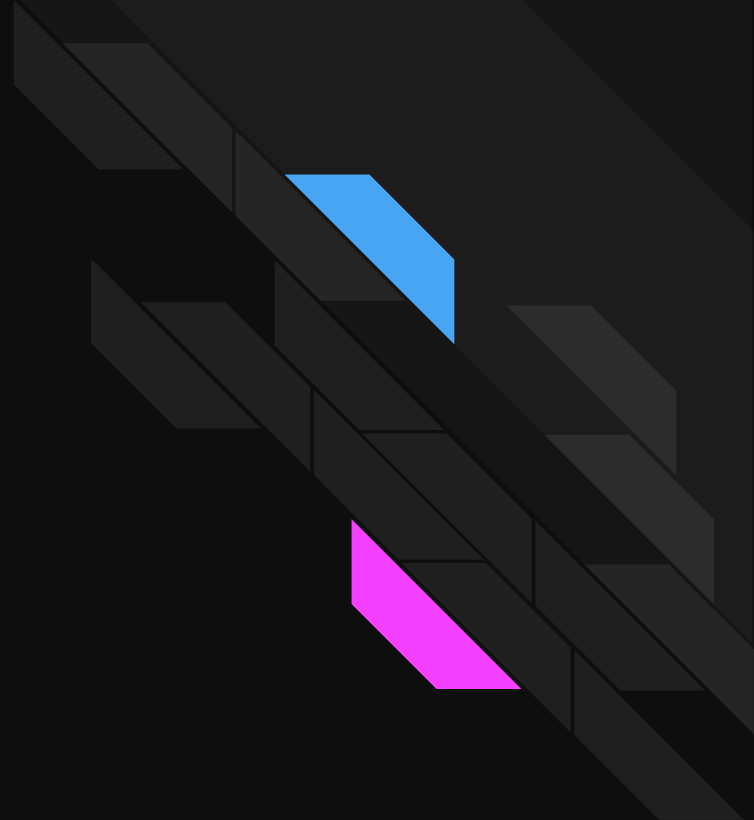
Precision Rate



Recall Rate



Integration and Maintenance Plan



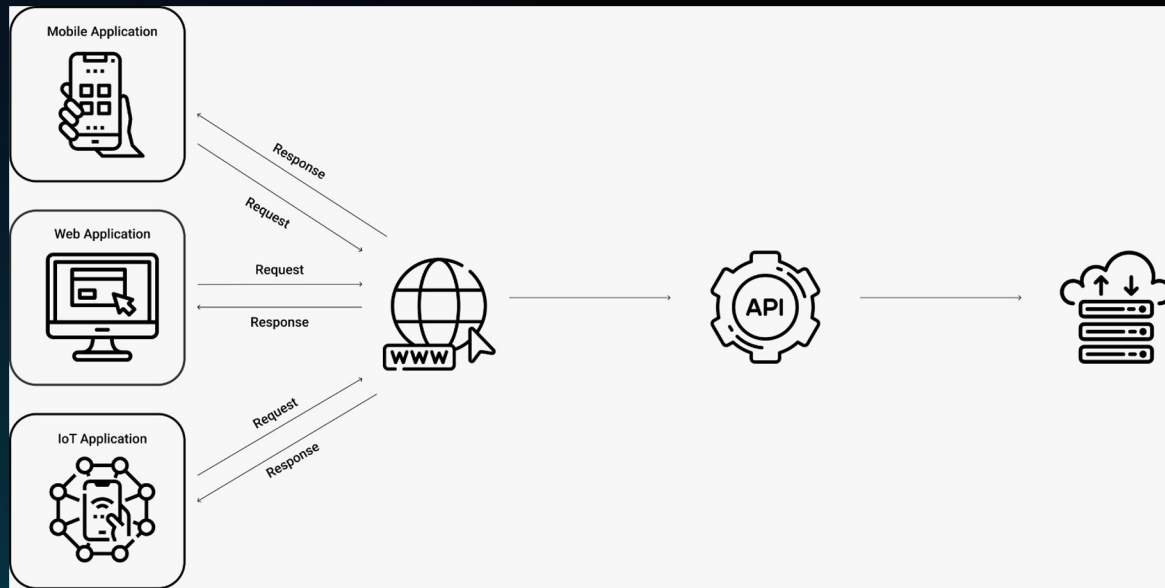
Integration Plan

Using the generated model, here are the our plans for integration

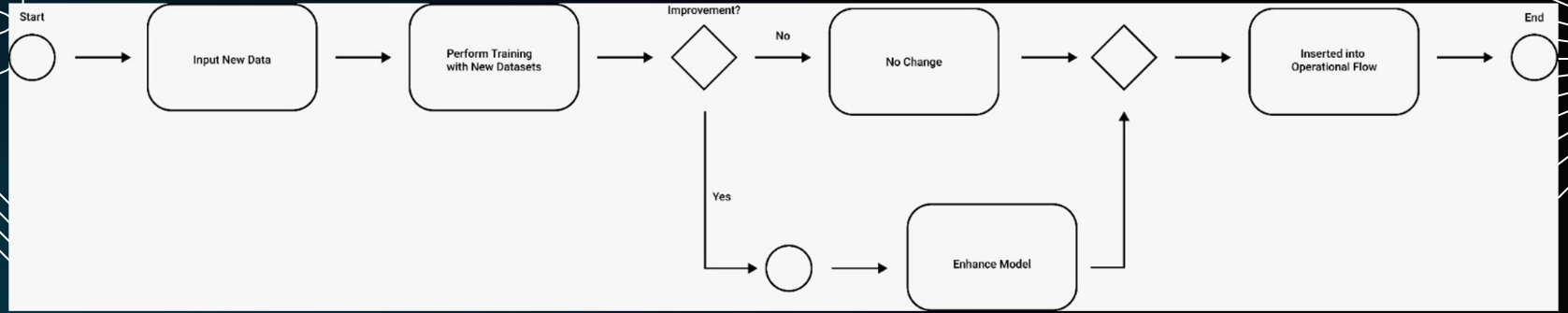
- **[Recommended]**
To integrate the model through an API (Application Programming Interface)
- Integrate the model into operation process by creating an automated workflow



API Integration



Workflow Automation



Maintenance Plan



01 Set up an alerting system to monitor the model's performance

02 Retrain the model regularly with new data

03 Track the model's performance over time





Thank you for your attention

Q & A