Eric Lim, Klass van Kempen, Jude Moukarzel

Georgetown University |

Project Scope Statement

Credit Card Fraud dataset

Table of Contents

[Project # 1](#_Toc86276921)

[Project Description 1](#_Toc86276922)

[Date Submitted 1](#_Toc86276923)

[Project Priority 1](#_Toc86276924)

[Step 1. Project Deliverables 1](#_Toc86276925)

[Step 2. List of Project Tasks 1](#_Toc86276926)

[Step 3. Out of Scope 1](#_Toc86276927)

[Step 4. Project Assumptions 1](#_Toc86276928)

[Step 5. Project Constraints 2](#_Toc86276929)

[Step 6. Updated Estimates 2](#_Toc86276930)

[Step7. Approvals 2](#_Toc86276931)

[Introduction 3](#_Toc86276932)

[Analysis of the dataset and Trained Model 3](#_Toc86276933)

[Exploratory Analysis and Visualization 3](#_Toc86276934)

[Baseline Model 4](#_Toc86276935)

[Model Selection 5](#_Toc86276936)

|  |  |  |  |
| --- | --- | --- | --- |
| Project # | Project Description | Date Submitted | Project Priority |
| 1 | GT Bank is a nationwide bank located in Washington D.C. Every day, thousands of credit card transactions flow through their systems, and they would like to predict which ones are fraudulent. We are commissioned to build a machine learning application that the bank can use to flag all fraudulent transactions so that they can take action and protect their customers. |  | Priority 01 |

## Step 1. Project Deliverables

Please list *all project deliverables* listed in the Project Charter and, if necessary, elaborate on them. *Do not list dates*. Add more rows as necessary.

|  |  |
| --- | --- |
| **Deliverable ID#** | **Description** |
| 1 | Problem Identification |
| 2 | Method Evaluations |
| 3 | Solution by Hyperparameter Optimization |
| 4 | Deployment Pipeline/Platform |
| 5 | Project Demo In class |

## Step 2. List of Project Tasks

Please list ***all project tasks*** to be completed, based on the “Deliverables” specified in the Project Charter. *Do not list dates*. Add more rows as necessary. Optional: you may substitute a work breakdown structure (WBS) or mind-map in lieu of Step 2. Please attach WBS or mind-map to document.

|  |  |  |  |
| --- | --- | --- | --- |
| **Task ID#** | **Task to be completed** | **Delivery Date** | **For Deliverable #** |
| 1 | Submit Project Charter / Problem Identification | 09/07/2023 | 1 |
| 2 | Submit Method Evaluations | 09/28/2023 | 2 |
| 3 | Submit Solution by Hyperparameter Optimization | 10/12/2023 | 3 |
| 4 | Submit Deployment Pipeline/Platform | 10/26/2023 | 4 |
| 5 | Project Demo in Class | 11/16/2023 | 5 |

## Step 3. Out of Scope

|  |  |
| --- | --- |
| This project **will NOT accomplish or include** the following: | This project will not include any post-detection decision making steps taken by the bank to deal with the fraudulent transactions identified. |

## Step 4. Project Assumptions

Please list any project factors that will be considered to be true, real, or certain. Assumptions generally involve a certain degree of risk.

|  |  |
| --- | --- |
| **#** | **Assumption** |
| 1 |  |
| 2 |  |
| 3 |  |

## Step 5. Project Constraints

|  |  |
| --- | --- |
| Project Start Date | 09/4/2023 |
| Launch/Go-Live Date | 12/05/2023 |
| Project End Date | 12/05/2023 |
| List any hard deadline(s) | 09/17/2023  09/28/2023  10/12/2023  10/26/2023  11/16/2023 |
| List other dates/descriptions of key milestones | None |
| Budget constraints Enter information about project budget limitations | Total (maximum) project budget  Limited funding for Cloud computing |
| Quality or performance constraints Enter any other requirements for the functionality, performance, or quality of the project | Software must load in 10 seconds or less;  System must provide 99.9% uptime;  Machine Learning algorithm must be trained on local machine and on subset of dataset due to no access to Cloud technology;  Fraudulent transactions detected must be accurate at 90% accuracy or more. |
| Equipment/personnel Constraints Enter any constraints regarding equipment or people that will impact the project | Cloud technology won’t be available until December 2023;  Local machines are used to train/test the ML model;  Employees Eric, Klass & Jude are the only employees available to complete this work; |
| Regulatory constraints  Enter any legal, policy or other regulatory constraints | Software must comply with HIPAA;  Website must comply with CSU accessibility policy;  Database must comply with campus Information Security policy. |
| Dataset Constraints | Missing data elements such as:   * Location of transaction * Location of Customer * Time between transactions (which may be resolved with feature engineering using the customer ID) |

## Step 6. Updated Estimates

|  |  |  |  |
| --- | --- | --- | --- |
| Estimate T&C hours required to complete project | N/A | If charge-back project, list total estimated T&C cost | N/A |

## Step7. Approvals

|  |  |  |  |
| --- | --- | --- | --- |
| Required For Project Class… | Role of Approver | Submitted for Approval on: | Approval Received on: |
| All classes | 1. Client + Client Supervisor | Nakul R. Padalkar |  |
| All classes | 2. T&C Supervising Manager | Nakul R. Padalkar |  |
| Class 3 + 4 only | 4. VP for Technology & Communication | Nakul R. Padalkar |  |
| Class 3 + 4 only | 5. Project Review Board | Nakul R. Padalkar |  |

Attach any additional documentation.

Office Use Only:

# Introduction

General Overview of topic

## Project Background and Description

1. What is so interesting about this problem?
   1. Main focus on financial industry
   2. Credit cards + mobile payments are already common and will continue to become more integrated in day-to-day transactions
2. What are the proposed benefits of the solution (why do we need to solve this)?
   1. User security + customer experience
      1. Need to manage balance of actually catching fraud and false positives ruining customer experience
   2. Mitigate financial loss
3. What problem type are we evaluating (classification? regression, or something else?)
   1. Binary Classification
4. What are the success criteria?
   1. Model works 🙂
   2. High accuracy
   3. Good precision + recall measures
   4. Identify most important features

## Proposed Methods

1. How?
   1. What methods are we evaluating? (generative, nongenerative, trees? hyperparameters?)
      1. Hyperparameter tuning for all methods
      2. Generative
         1. Boosting
      3. Nongenerative
         1. Bagging
      4. Ensemble
         1. Stacking
   2. What is the deliverable?
      1. Well-performing model
   3. What is out of scope?
      1. Post-detection decision making

# Analysis of the dataset and Trained Model

## Exploratory Analysis and Visualization

## Baseline Model

# **Model** Selection

## Model Performance Evaluation

# Initial Deployment

## Screen 1

## Screen 2

Figure 7 Predicted vector

## Heroku Application

### The Heroku Platform

### Application path

### Deployment, debugging, and updates

# Conclusion