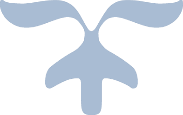


HIGH LEVEL DESIGN (HLD)

Credit Card Default Prediction



#### VIKRAM TAVITI

iNeuron.ai

# Document Version Control

|  |  |  |  |
| --- | --- | --- | --- |
| **Date Issued** | **Version** | **Description** | **Author** |
| **07.09.2022** | V1.0 | Initial HLD- V1.0 | VIKRAM TAVITI |

Contents

[Document Version Control 1](#_bookmark0)

[Abstract 3](#_bookmark1)

* 1. [Introduction 3](#_bookmark2)
  2. [Why this High-Level Design Document? 3](#_bookmark3)
  3. [Scope 4](#_bookmark4)
  4. [Definitions 4](#_bookmark5)
  5. [General Description 5](#_bookmark6)
  6. [Product Perspective 5](#_bookmark7)
  7. [Problem Statement 5](#_bookmark8)
  8. [Proposed Solution 5](#_bookmark9)
  9. [Further Improvements 5](#_bookmark10)
  10. [Technical Requirements 6](#_bookmark11)
  11. [Data Requirements 6](#_bookmark12)
  12. [Tools Used 6](#_bookmark13)
  13. [Constraints 8](#_bookmark14)
  14. [Assumptions 8](#_bookmark15)
  15. [Design Details 8](#_bookmark16)
  16. [Process Flow 8](#_bookmark17)
  17. [Event Log 9](#_bookmark18)
  18. [Performance 9](#_bookmark19)
  19. [Reusability 9](#_bookmark20)
  20. [Application Compatibility 9](#_bookmark21)
  21. [Deployment 10](#_bookmark22)

[5.0 Dashboards 10](#_bookmark23)

[6.0 Conclusion 11](#_bookmark24)

# Abstract

Financial threats are displaying a trend about the credit risk of commercial banks as the incredible improvement in the financial industry has arisen. In this way, one of the biggest threats faces by commercial banks is the risk prediction of credit clients. The dataset I have used to solve the problem is Credit card default dataset of Taiwan from April 2005 to September 2006.

I The data analysis of this dataset revealed various insights how their credit usage affect their risk of being default. Data preprocessing and data transformation is done prior to training. The dataset is divided into train dataset and test dataset and various classification algorithms are used to build models.

All the models are tested and evaluated on various metrics and the model built using random forest algorithm has performed well. The Random forest model is further gone through hyper parameter tuning for the optimization of the model.

A web application is built using the model for predictions and deployed on Heroku platform .

# Introduction

## Why this High-Level Design Document?

The purpose of this High-Level document is to add necessary details to current project description to represent a suitable model for coding. This document is used as a reference manual for how the model interact at a high-level.

4 High Level Design (HLD)

### The HLD will

* + - Presents all design aspects and define them in detail.
    - Describe the user interface being implemented.
    - Describe the hardware and software interfaces.
    - Describe the performance requirements.
    - Include design feature and the architecture of the project.

## Scope

The HLD document presents the structure of the system, such as the database architecture, application architecture, and technology architecture. The HLD uses non-technical to middle-technical terms which should be understandable to the administrators of the system.

## Definitions

**Term Description**

Database IDE API KPI

VS Code EDA

Collection of all the information

Integrated Development Environment Application Programming Interface Key Performance Indicator

Visual Studio Code Exploratory Data Analysis

* 1. General Description

## Product Perspective

The credit card default is a machine learning based predictive model which determines whether a person is going to default or not provided his/her credit details.

## Problem Statement

To develop an API interface to credit card default payment and analyzing the following:

* + - To predict the probability of credit default based on credit card owner's characteristics and payment history.
    - To create API interface to determine the credit card defaulter

## Proposed Solution

The solution proposed to predict the default of credit card clients is to build a model based on the Taiwan credit card clients data. This is done first by analyzing the dataset by visualization to get insights and perform some data transformations, train the model using the best algorithm that fit for our problem, testing and optimization is done.

An API interface is built using the model for predictions. The web application is deployed on Heroku platform and it ask some inputs to enter credit details of a user , after submitting the details it predicts and show the result.

* 1. Technical Requirements

The solution is a cloud-based application hosted on Heroku platform For accessing this application below are the minimum requirements:

* + - Good internet connection.
    - Web Browser.
    - Deployed link

For training model, the system requirements are as follows:

• +4 GB RAM preferred

* + - Operation System: Windows/ Linux/ Mac
    - Pycharm/Visual Studio
    - Jupyter notebook/Google colab
  1. Data Requirements

Data requirements completely depends on our problem statement.

* + - Comma separated values (CSV) file.
    - Input file feature/field names and its sequence should be followed as per decided.
  1. Tools Used

Python programming language and frameworks such as NumPy, Pandas, Scikit-learn, Seaborne are used to build the whole model.













* + - Pandas is an open-source Python package that is widely used for data analysis and machine learning tasks.
    - NumPy is most commonly used package for scientific computing in Python.
    - Seaborne is an open-source data visualization library used to create interactive and quality charts/graphs.
    - Scikit-learn is used for a machine learning.
    - Pycharm is used as IDE (Integrated Development Environment)
    - GitHub is used as version control system.
    - Front end development is done using HTML/CSS.
    - Heroku is used for deployment of the model.
  1. Constraints

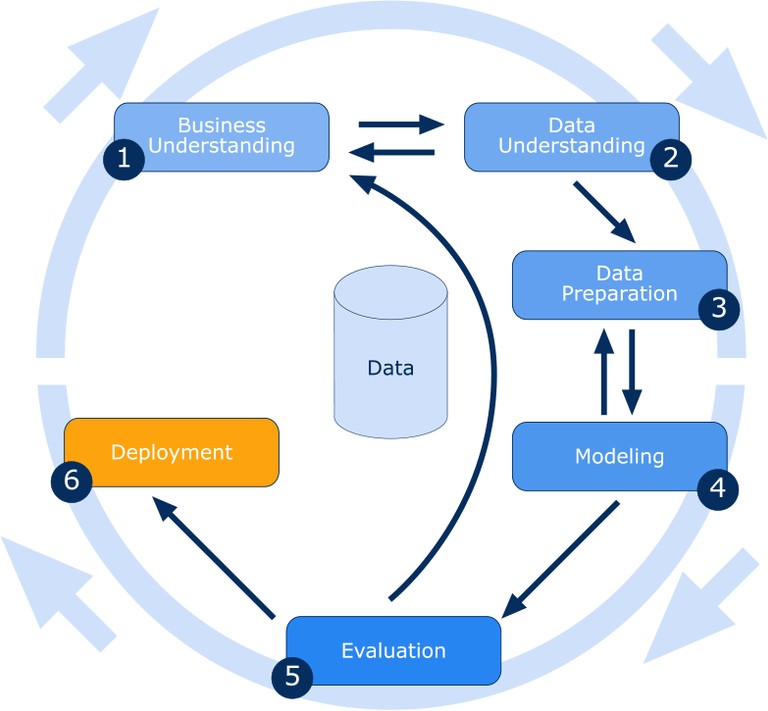
This model is user friendly, as automated as possible and users are not be required to know any of the workings.

* 1. Assumptions

The main objective of the project is to develop an API to predict the probability of credit default based on credit card owner’s characteristics and payment history.

* 1. Design Details

## Process Flow



## Event Log

The system logs every event so that the user knows what process is running internally.

**Initial Step-By-Step Description:**

* + - The system identifies at what step logging required.
    - The system should be able to log each and every system flow.
    - Developer can choose logging method. You can choose database logging.

System should not hang out even after using so many loggings.

# Performance

## Reusability

The entire solution is done in modular fashion and it is API oriented. So, in the case of the scaling the application, the components are completely reusable.

## Application Compatibility

The interaction with the application is done through the designed user interface, which the end user can access through any web browser.

## Deployment



# Dashboards

A dashboard is a data visualization and analysis tool that displays on one screen the status of key performance indicators (KPIs) and other important business metrics.



As a high-level reporting mechanism, dashboards provide fast ‘big picture’ answer to critical business questions and assist and benefit decision making in several ways:

* + - Communicating how premium is varies with BMI value.
    - Visualizing relationship of gender with premium in easy-to understand way.

6.0 Conclusion

This system shows risk of credit card clients getting default . The CCDP model is used to make the accurate prediction for the credit card default, this model can be used by various bank organizations to get to know their credit card clients and solve the biggest threats face by the commercial banks.