National College of Ireland

Configuration Manual

MSc Research Project
Programme Name

Yashwanth Muddanna Hanumantharaya
Student ID: 23284986

School of Computing
National College of Ireland

Supervisor: Professor Paul Stynes

National College of Ireland

MSc Project Submission Sheet

School of Computing

Student Name:	Yashwanth Muddanna Hanumantharaya		
Student ID:	23284986		
Programme:	MSc in AI	**Year:**	2024-25
Module:	MSc (Research) Practicum		
Lecturer:	Professor Paul Stynes		
Submission Due Date:	11 Aug 2025		
Project Title:	An Enhanced Hybrid Classification Model Using Distance Metric to Detect Financial Fraud		
Word Count:	519	**Page Count: 6**	

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.
ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature: Yashwanth Muddanna Hanumantharaya

Date: 11 Aug 2024

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST

Attach a completed copy of this sheet to each project (including multiple copies)	□
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	□
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	□

Assignments that are submitted to the Programme Coordinator Office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Configuration Manual

Yashwanth Muddanna Hanumantharaya
Student ID: 23284986

1 System Configuration

1.1 Hardware Requirements

The application has been optimized to use as normal laptop or desktop computer applications without fail. The product features minimum recommended specifications as follows:

- Memory (RAM):8 GB or above is required; 16 GB suggested to run operations of SMOTE and PCA run smoothly.
- Storage: 5 GB or more of free disk space to store training datasets, processed data, model artifacts and logs.
- CPU: Dual-core CPU (Intel i5 or better suggested) to be used in training and forecasting.
- GPU: Optional - GPU (4 GB+ VRAM) that supports CUDA can optionally be used to speed-up training of neural network (MLP).

1.2 Reference System Specification

The final version of codes was run in the following system specifications:

Component	Specification
Processor	AMD Ryzen 5 7235HS (3.20 Hz)
Installed RAM	16 GB
GPU	NVIDIA GeForce RTX 3050 (6 GB VRAM)
Storage	512 GB SSD

1.3 Software Prerequisites

- Operating System: Windows 10/11, macOS, Linux
- Python Version: 3.10 or above
- VS Code software

1.4 Core Python Dependencies

All the required libraries were listed in requirement.txt as listed below:

- Pandas
- numpy
- scikit-learn
- matplotlib
- joblib
- streamlit

The requirement.txt was run using the following command on git-bash in VS Code:

pip install -r requirements.txt

2 Experiment Setup and Dataset Configuration

2.1 Project Structure

A folder was created as NCI under that the following folders are setup as follows:

```
project_root/
├── src/
│   ├── data_loader.py
│   ├── preprocessing.py
│   ├── train.py
│   ├── evaluate.py
│   └── eda.py
├── data/
│   ├── raw/
│   └── processed/
├── artifacts/
├── eda_outputs/
├── logs/
├── main.py
├── streamlit_app.py
└── requirements.txt
```

2.2 Dataset Specifications

The dataset consists of 284,807 transactions and 492 of these have been identified as fraud (Class = 1). Features include:

- Time: Number of seconds elapsed since first transaction in dataset.
- V1 to V28: PCA features to secure information.
- Amount: The amount of transaction.
- Class: Consists of 1 & 0, fraud and non-fraud transactions.

3 Data Processing

3.1 Data Loading:

It reads data by csv file provided as data/raw/creditcard.csv, divides it into training and test sets with stratified sampling method.

3.2 Scaling:

StandardScaler() standardizes feature to calculate distance fairly in KNN.

3.3 SMOTE balancing:

Synthetic Minority Over-sampling Technique (SMOTE) creates some more samples of fraud so that the balance of the dataset is possible.

3.4 PCA (Dimensionality Reduction):

The SMOTE is followed by PCA to preserve 95 percent of the variance and to minimise features.

3.5 Saving Artifacts:

Scaler, PCA transformer and trained models are stored in the artifacts/ directory to reuse later.

4 Model Training and Evaluation

Model Training:
KNN - Tuned with n_neighbors=3, weights='distance', metric='euclidean'
MLP - Tuned with random_state=42, max_iter=300, activation='relu',
hidden_layer_sizes=(64, 32), learning_rate_init=0.001
Metrics used: Accuracy, Precision, Recall, F-1 Score, AUC-ROC, Confusion Metrix

5 Runing Application

5.1 Training the model

Runn the main.py using python main.py in git-bash to perform the following steps:

- Data Preprocessing (scaling, SMOTE, PCA)
- Model Training and saving
- Evaluation and EDA plots

5.2 Streamlit Web Application (optional)

Streamlit helps to classify the data into fraud and non-fraud using same features of different time stamp data. This is also generating the classified report, where we can download and view the classified transactions to take further actions.

6 Output Files

After running the main.py, the hybrid classification model generates outputs and saves the following:

- Logs
- Output Images
- Pickle Files