Configuration Manual

MSc Research Project

Programme Name: MSCFTD1

KOMATINENI VENU BABU

Student ID: X23106662

School of Computing

National College of Ireland

Supervisor: Faithful Onwuegbuche

National College of Ireland

MSc Project Submission Sheet

School of Computing

Student Name:	……Komatineni venu babu…………
Student ID:	…x23106662………………………………………………..……
Programme:	………MSCFTD1……………………… **Year:** ……2024……………..
Module:	**MSC Research project**………………………………………………………………
Lecturer:	………………: Faithful Onwuegbuche ……………………………
Submission Due Date:	……………………………16/09/2024………………………………………………
Project Title:	CREDIT RISK ASSESSMENT OF CONSUMER LOANS IN INDIA USING MACHINE LEARNING TECHNIQUES
Word Count:	………715…………… Page Count: ……………5………………….…………

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.
<u>ALL</u> 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:	……………K. VENU BABU…………………………………………
Date:	…………16/09/2024……………………………………………………

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):	

1   Introduction

This is the configuration manual of the research project for the title. This will provide detail instruction for setup the environment, software for code.

2 Hardware/Software Requirements
 2.1 Hardware Requirements

The system hardware configuration on which the research models are developed and successfully executed. These configurations are given below.
Hardware configuration

Device name   venuchowdary7
Processor   Intel(R) Core(TM) i5-1035G1 CPU @ 1.00GHz   1.20 GHz
Installed RAM   8.00 GB (7.74 GB usable)
Device ID   1D106EF5-9A66-40BC-B8EE-576C654FAACF
Product ID   00327-36310-15496-AAOEM
System type   64-bit operating system, x64-based processor
Pen and touch   No pen or touch input is available for this display

2.2 Software Requirement:

The software requirement for the research project to develop the machine learning and neural network algorithms are given below.

- Integrated Development Environment: Jupyter notebook
- Programming language: Python 3.7+
- Data storage: local server
- Other software: Excel /CSV, Text editor**.**

3 Environment setup:

Jupyter notebook: To setup jupyter notebook, Anaconda software must be downloaded first. After installing the anaconda, jupyter is the default software in anaconda.
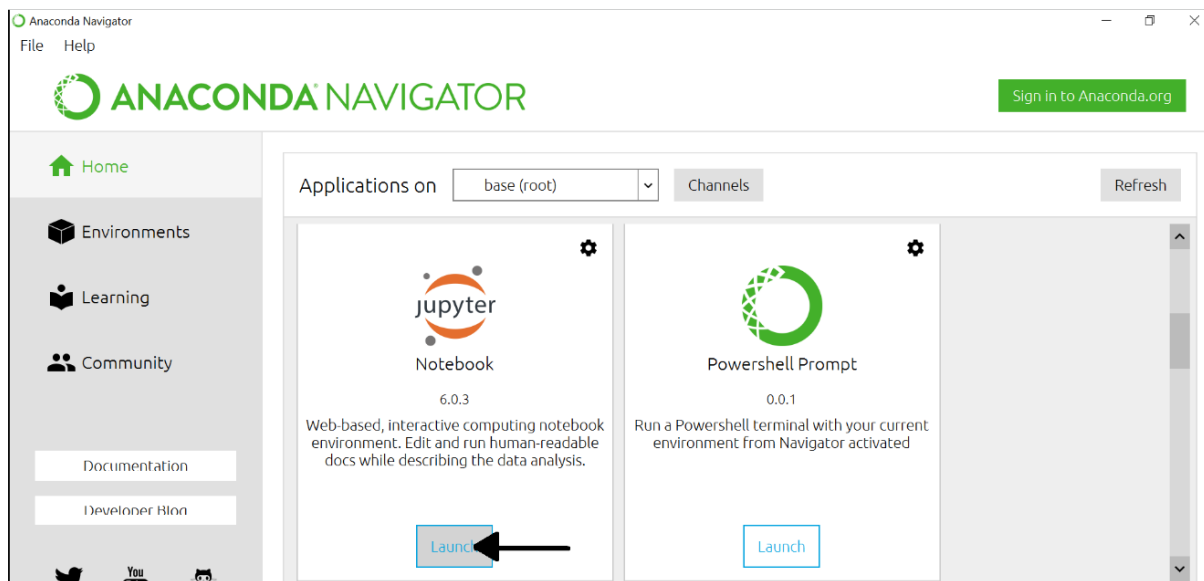
Fig1 : Anaconda application

5. Data Preparation
5.1 Load Dataset
1. Place your dataset (e.g., credit_data.csv) in the project directory.
2. Load the dataset in your Python script or Jupyter Notebook:

python

```
import pandas as pd

df = pd.read_csv('credit_data.csv')
```

5.2 Data Cleaning
1. **Handle Missing Values:**

```
Python df = df.dropna()  # Drop rows with missing values
# Or fill missing values
df.fillna(method='ffill', inplace=True)
```

2. **Encode Categorical Variables:**

```
Python df = pd.get_dummies(df, drop_first=True)
```

3. **Feature Scaling:**

```
Python from sklearn.preprocessing import StandardScaler
scaler = StandardScaler()
scaled_features = scaler.fit_transform(df)
```

4. **Split Dataset:**

```
Python from sklearn.model_selection import train_test_split
X = df.drop('target_variable', axis=1)
y = df['target_variable']
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
```

6. Model Building
6.1 Random Forest Classifier Setup
1. Import the Random Forest Classifier:

```
Python from sklearn.ensemble import RandomForestClassifier

model = RandomForestClassifier(n_estimators=100, random_state=42)
```

2. Train the model:

```
Python model.fit(X_train, y_train)
```

6.2 Model Evaluation
    1. **Performance Metrics:**

```python
Python from sklearn.metrics import accuracy_score, precision_score, recall_score, f1_score, confusion_matrix
y_pred = model.predict(X_test)
accuracy = accuracy_score(y_test, y_pred)
precision = precision_score(y_test, y_pred)
recall = recall_score(y_test, y_pred)
f1 = f1_score(y_test, y_pred)
conf_matrix = confusion_matrix(y_test, y_pred)

print(f'Accuracy: {accuracy}')
print(f'Precision: {precision}')
print(f'Recall: {recall}')
print(f'F1-Score: {f1}')
print(f'Confusion Matrix:\n{conf_matrix}')
```

7. Running the Project

7.1 Execute in Jupyter Notebook
    1. Launch Jupyter Notebook:

```
jupyter notebook
```

    2. Open your notebook file and run the cells sequentially.

7.2 Execute in Python Script
    1. Ensure your virtual environment is activated.
    2. Run your Python script:

```
python main.py
```

8. Conclusion

This configuration manual provides a step-by-step guide to set up and run your machine learning project for credit risk assessment using the Random Forest Classifier. Ensure that all dependencies are installed correctly and follow the steps in order for optimal performance.