National College of Ireland

Configuration Manual

MSc Research Project
Msc Fintech

Ibrahim Okunlola
Student ID: X22210016

School of Computing
National College of Ireland

Supervisor: Brian Byrne

National College of Ireland

MSc Project Submission Sheet

School of Computing

Student Name: Okunlola Ibrahim Olajide

Student ID: X22210016

Programme: Msc Fintech **Year:** 2024

Module: Msc Research Project

Lecturer: Brian Byrne

Submission Due Date: 12th August 2024

Project Title: FINTECH'S ROLE IN RESHAPING TRADITIONAL BANKING MODELS CASE STUDY: Fintech Innovation's Effect on Financial Inclusion – Moniepoint in Nigeria

Word Count: 1060 **Page Count:** 7

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: Ibrahim Okunlola

Date: 12th August 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

Ibrahim Okunlola
Student ID: X22210016

1 Section 1

This paper implements the system setup and other technique setup used in Research Project: "**FINTECH'S ROLE IN RESHAPING TRADITIONAL BANKING MODELS *CASE STUDY: Fintech Innovation's Effect on Financial Inclusion – Moniepoint in Nigeria***".

Phyton was used and Google Collab is the virtual environment of this Research Project. All the data source and the results are then uploaded to Google collab Drive: https://colab.research.google.com/drive/1DfsZljh9qaC2cZSjE2Fd6VWLd0KxqDW?usp=sharing

2 Section 2

Device Specification
Device name LAPTOP-1T7T8RBL
Processor 13th Gen Intel(R) Core(TM) i5-1335U 1.30 GHz
Installed RAM 8.00 GB (7.72 GB usable)
Device ID 083383B8-5919-48A3-81C7-E1E43BF57AED
Product ID 00342-21183-90526-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 specification
Gmail account with the accession to Google Drive and Google Colab (Okunsolajide@gmail.com)

3 Research Project:

Dataset is obtained through primary source from 80 correspondent and analysed, I have attached a video of the step-to-step Analysis:

Analysis Video.mp4

Below is a detailed explanation of this video in text.

Summary of the Code and its Components Introduction
This code is a comprehensive data analysis and machine learning pipeline applied to a dataset stored in a CSV file named "PRC 00000473." The pipeline covers data loading, preprocessing, exploratory data analysis, encoding categorical variables, model training, evaluation, and comparison of various machine learning models.

1. Libraries and Tools Used

- **Pandas**: For data loading and manipulation.
- **NumPy**: For numerical operations.
- **Matplotlib, Pyplot, Seaborn**: For data visualization.
- **Random**: To set a seed for reproducibility of results.

2. Data Loading

- **Loading Data**: The dataset is loaded into a pandas DataFrame.
- **Previewing Data**: The first and last few rows of the data are previewed to understand its structure and content.
- **Inspecting Column Names**: Column names are printed to identify relevant and irrelevant columns.

3. Data Cleaning

- **Dropping Unnecessary Columns**: Columns that are not relevant, such as timestamps, are removed to streamline the data.
- **Renaming Columns**: Columns are renamed for clarity and ease of analysis.

4. Handling Missing and Duplicate Values

- **Checking Missing Values**: Identifies any missing data that needs to be handled, either through imputation or removal.
- **Checking Duplicate Values**: Detects duplicate entries that could affect analysis results.

5. Creating Pivot Tables for Categorical Variables

- **Pivot Tables**: For each categorical variable, a pivot table is created to count the occurrences of each category, helping to understand the distribution of responses.

6. Data Visualization

- **Bar Plots and Pie Charts**: Various plots are used to visualize the distribution of variables, aiding in gaining insights.
- **Annotations and Label Rotation**: Counts are annotated on plots for better interpretability, and labels are rotated if necessary for readability.

7. Data Encoding

- **Label Encoding**: Categorical variables are converted into numerical formats required for machine learning algorithms.

8. Preparing Data for Machine Learning

- **Feature Selection**: Certain variables that are not predictive or are the target itself are excluded.
- **Train-Test Split**: The data is divided into training and testing sets to evaluate model performance on unseen data.

9. Training and Evaluating Machine Learning Models

- **Models Used**: Random Forest, Decision Tree, K-Nearest Neighbors (KNN), and XGBoost classifiers.
- **Process**:
    - **Initialization**: Each model is instantiated with appropriate parameters. ○ **Training**: Models are trained on the training data. ○ **Prediction**: Models make predictions on the test data.
    - **Evaluation**: Models are evaluated using accuracy scores, classification reports, and confusion matrices.
    - **Visualization**: Confusion matrices are plotted for better interpretation of model predictions.

10. Model Comparison

- **Comparing Models**: The accuracy scores of all four models are compared using bar plots.
- **Best Model**: Random Forest is identified as the best-performing model, achieving an accuracy of 68.75%.

Conclusion

The code provides a complete pipeline from data loading to model evaluation, offering insights into the dataset and assessing the predictive capabilities of different machine learning models. Random Forest emerged as the best model for predicting financial inclusion based on the provided data.

References

References should be formatted using APA or Harvard style as detailed in NCI Library Referencing Guide available at https://libguides.ncirl.ie/referencing
You can use a reference management system such as Zotero or Mendeley to cite in MS Word.

1. Géron, A. (2019) Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow. 2nd edn. Sebastopol, CA: O'Reilly Media.

2. Chollet, F. (2018) Deep Learning with Python. Shelter Island, NY: Manning Publications.

3. Goodfellow, I., Bengio, Y., & Courville, A. (2016) Deep Learning. Cambridge, MA: MIT Press.

4. Brownlee, J. (2017) Machine Learning Mastery With Python: Understand Your Data, Create Accurate Models, and Work Projects End-To-End. Vermont: Machine Learning Mastery.