

Configuration Manual

MSc Research Project
MSc Data Analytics

Mary Ann Antony Kizhakechethipuza
Student ID: 23360895

School of Computing
National College of Ireland

Supervisor: Furqan Rustam

National College of Ireland
MSc Project Submission Sheet
School of Computing



Student Name: Mary Ann Antony Kizhakechethipuzha

Student ID: 23360895

Programme: MSc in Data Analytics

Year: 24-2025

Module: MSc Research Project

Lecturer: Furqan Rustam

Submission

Due Date: August 11th, 2025

Project Title: Evaluation of Sustainable Development Business Strategies of Higher Education Institutions using Data Mining and Machine Learning Techniques.

Word

Count: 673 **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.

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: Mary Ann Antony Kizhakechethipuzha

Date: August 11th, 2025

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST

Attach a completed copy of this sheet to each project (including multiple copies)	<input type="checkbox"/>
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	<input type="checkbox"/>
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.	<input type="checkbox"/>

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

Mary Ann Antony Kizhakechethipuza
Student ID: 23360895

1 Mac OS

All analysis was done using mac OS Sequoia 15.5. No special configuration was needed to run the analysis through the tools chosen to do the analysis

2 RStudio, Bibliometrix and Biblioshiny

This software was downloaded from the internet using the link <https://www.bibliometrix.org/home/index.php/download>. Bibliometrix is a library that can be initiated through RStudio console and Biblioshiny is a web application that can initiated through the same RStudio console once Bibliometrix library has been initiated. Default system settings were used as which comes when these software were downloaded. No special configuration was required.

3 Jupyter Notebook and Machine Learning algorithms

Jupyter Notebook was the application used to run the machine leaning algorithms. Default system configuration was used to open the jupyter notebook from Anaconda Navigator. Anaconda was downloaded from the website anaconda.org. The Machine learning algorithms used with their design in included below as well as in the report.

Algorithm	Design Parameters of finalized model	Dataset
KMeans Clustering	Programming language: Python Application: Jupyter Notebook Data preprocessing: Normalization using MinMaxScaler Train-test split: 80/20 Random_state: 23360895 k =n_clusters = 6 n_init='auto'	2019 THE-IR
KMeans Clustering	Programming language: Python Application: Jupyter Notebook Data preprocessing: Normalization using MinMaxScaler Train-test split: 80/20 Random_state: 0 k =n_clusters = 3	2025 THE-IR

	n_init='auto'	
Random Forest Classifier	Programming language: Python Application: Jupyter Notebook Data preprocessing: Label encoding Train-test split: 80/20 Random_state: 23360895 Features included for training set: columns Name, country, region, Sdg 3,4,5,8,9,10,11,12,13,16,17, Target Best hyperparameters: {'max_depth': 17, 'n_estimators': 234} Default values were used for all other hyperparameters	2019 THE-IR
Logistic Regression	Programming language: Python Application: Jupyter Notebook Data preprocessing: Label encoding Train-test split: 80/20 Random_state: 16 max_iter: 1000 Features included for training set: columns Name, country, region, EnvironmentD, SocialD, EconomicD, Target All other hyperparameters like penalty, dual, fit_intercept etc. were used with default values.	2025 THE-IR

In addition, in built libraries like Pandas, NumPy, Sklearn, seaborn matplotlib was used at different stages of the data analysis.

4 Microsoft Excel and Adobe pdf

These tools were used in data extraction and in thematic analysis. These software were already installed in the system previously and default software settings were used with no special configuration requirement.

References

Bibliometrixdownload. Bibliometrix.org. <https://www.bibliometrix.org/home/index.php/download>

ImpactRankings. timeshighereducation.com. <https://www.timeshighereducation.com/impactrankings>

How food connects all the sdgs(2016 June 14).stockholmresilience.org
<https://www.stockholmresilience.org/research/research-news/2016-06-14-how-food-connects-all-the-sdgs.html>

Anaconda Navigator. anaconda.org
<https://anaconda.org/anaconda/anaconda-navigator>