National College of Ireland

Configuration Manual

MSc Research Project
Data Analytics

Surabhi Singh
Student ID: X21148325

School of Computing
National College of Ireland

Supervisor: Noel Cograve

National College of Ireland
Project Submission Sheet
School of Computing

Student Name:	Surabhi Singh
Student ID:	X21148325
Programme:	Data Analytics
Year:	2023
Module:	MSc Research Project
Supervisor:	Noel Cograve
Submission Due Date:	31/01/2023
Project Title:	Configuration Manual
Word Count:	289
Page Count:	3

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 speci ed in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature:	Surabhi Singh
Date:	31st January 2023

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 o ce must be placed into the assignment box located outside the o ce.

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

Configuration Manual

Surabhi Singh
X21148325

1 Introduction

Implementation of a framework design that combines word embedding based model with cosine similarity to recommend top 9 podcasts based on input podcast to provide suggestion based on user's interest. The presented manual includes details on system con g., software required, and hardware utilized, and list of libraries to be imported for successfully running the project code.

2 System Config.

The con guration details used to conduct the research project is as per the given tables.

2.1 Hardware Used

Details of hardware required

OS	macOS13.0.1
System Processor	Mac M1 Chip
RAM	8.00GB
Storage	256GB

2.2 Software Required

Details of Software required

Programming Language Python	3.9.15 and above
Tools	Jupyter Notebook

3 Project Artefacts

As part of project development, some important libraries were imported that are shared in the below table

Process Flow	Libraries imported
Data Mining	spotipy.util, urllib.parse, pandas, numpy, json
Pre-proccessing	nltk.tokenize for RegexpTokenizer, stopWords nltk.stem.porter for PorterStemmer, nltk.stem for wordNetLemmatizer
Model Building	sklearn.featureextaction.text for CountVectorizer and TF-IDF, gensim.models 4.3.0 for Word2Vec and KeyedVectors(GloVe), sklearn.decomposition for PCA,matplotlib, sklearn.metric.pairwise for cosine similarity

3.1 Important Libraries

3.2 Project Code Artifacts

The Figure 1 shows the project folder when downloaded from zip folder

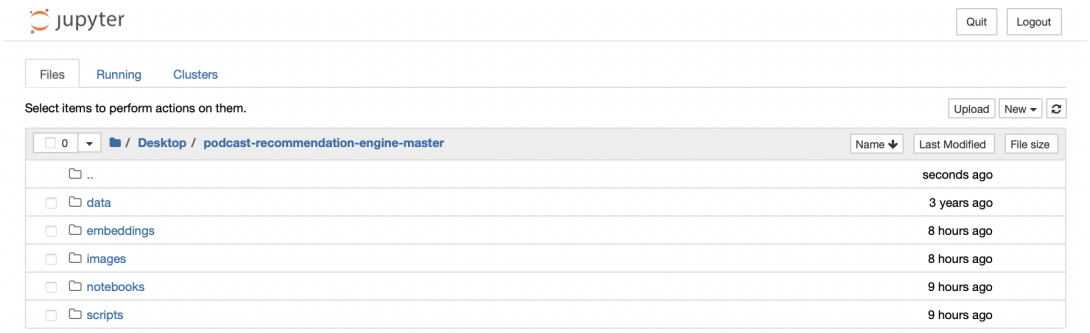

Figure 1: Project Folder

The Figure 2 in the screenshot of JSON le created when dataset is collected from Spotify Web API. The Figure 3 shows the juypter notebooks to perform dataset collection query and cleaning of data. The Figure 4 shows the juypter notebooks to perform data exploration, pre-processing, recommendation model design and visualization fro result. The Figure 5 all the images created in the research project.