

Configuration Manual

MSc Research Project Data Analytics

Surabhi Singh Student ID: X21148325

School of Computing National College of Ireland

Supervisor: Noe

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.

<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:	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 office must be placed into the assignment box located outside the office.

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 config., software required, and hardware utilized, and list of libraries to be imported for successfully running the project code.

2 System Config.

The configuration 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-processing	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

💭 Jupyter	Quit Logout
Files Running Clusters	
Select items to perform actions on them.	Upload New - 2
🔲 0 💌 🖿 / Desktop / podcast-recommendation-engine-master	Name Last Modified File size
۵	seconds ago
🗆 🗀 data	3 years ago
embeddings	8 hours ago
images	8 hours ago
notebooks	9 hours ago
□ □ scripts	9 hours ago

Figure 1: Project Folder

The Figure 2 in the screenshot of JSON file 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.



Figure 2: Dataset

🗂 Jupyter	Quit Logout
Files Running Clusters	
elect items to perform actions on them.	Upload New - 2
0 V Desktop / podcast-recommendation-engine-master / scripts	Name Last Modified File size
۵	seconds ago
Spotify Podcasts.lpynb	9 hours ago 5.13 kB
Clean_podcast_df.py	3 years ago 1.13 kB

Figure 3: DatasetClean

🗂 jupyter	Quit	Logout
Files Running Clusters		
elect items to perform actions on them.	Upload Ne	ew 🕶 🖸
0 V Desktop / podcast-recommendation-engine-master / notebooks	Name Last Modified F	ile size
۵.	seconds ago	
Ip_podcast_analysis.ipynb	9 hours ago	248 kB
Dependence de la companya	2 days ago	33.3 kB
B w2v_visualization.ipynb	9 hours ago	962 kB

Figure 4: Project Notebooks

jupyter	Quit Logo	out
Files Running Clusters		
lect items to perform actions on them.	Upload New -	C
0 V Desktop / podcast-recommendation-engine-master / images	Name 🔶 Last Modified File si	ze
۵	seconds ago	
Genre wise median podcast.png	3 years ago 46	6 kB
Genre wise podcast count.png	3 years ago 42.1	kB
¹ similar_words.png	3 years ago 623	kB
C Lastas.png	3 years ago 100	kB
¹ w2v.png	3 years ago 1.17	мв

Figure 5: Images