National College of Ireland

Configuration Manual

MSc Research Project
Data Analytics

Nurul Hanis Binti Ibrahim
Student ID: x20246862

School of Computing
National College of Ireland

Supervisor: Dr Catherine Mulwa

National College of Ireland
Project Submission Sheet
School of Computing

Student Name:	Nurul Hanis Binti Ibrahim
Student ID:	x20246862
Programme:	Data Analytics
Year:	2022
Module:	MSc Research Project
Supervisor:	Dr Catherine Mulwa
Submission Due Date:	15/12/2022
Project Title:	Configuration Manual
Word Count:	433
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:	Nurul Hanis Binti Ibrahim
Date:	30th 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

Nurul Hanis Binti Ibrahim
x20246862

1 Introduction

This configuration manual is for the research project of the Recommendation of Korean-Pop Bands using Topic Modelling Algorithm and Myers-Briggs Type Indicator. Section 2 explained the system configuration to implement the project. Whereas, in Section 3 outlined the data extraction process. Next, Section 4 entailed the required Python libraries for the implementation of the project. This is followed by the recommender demonstration in Section 5.

2 System Configuration

2.1 Hardware Requirements

The project is conducted at a local system with the specifications below:

1. Model Name: MacBook Air
2. System OS: MacOS Monterey (Version 12.6.1)
3. Processor: 1.6 GHz Dual-Core Intel Core i5
4. Memory: 8 GB

2.2 Software Requirements

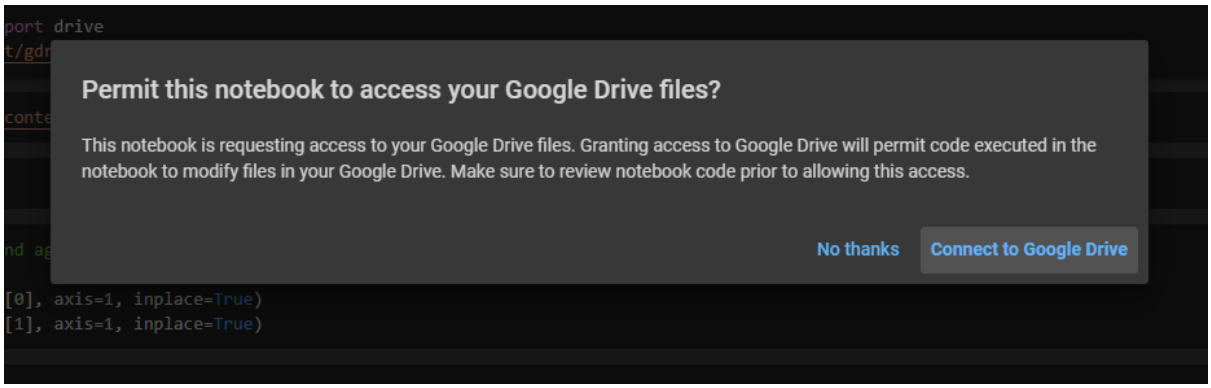
1. Python 3.8.16: The version (Zakka; 2020) of the programming language that the code is written in throughout the project.
2. Google Colaboratory: A Jupyter notebook environment that is free and runs entirely in the cloud. It helps to execute the code.
3. Google Drive: The location of the stored raw data.
4. Anvil: A drag-and-drop web application builder that is specifically Python-based that is used to deploy the recommender algorithm.

3 Data Extraction

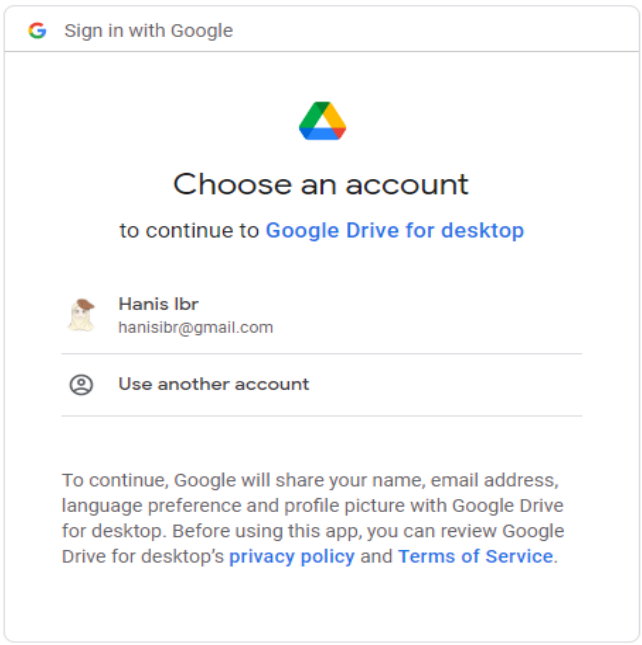
Firstly, the dataset must be located in Google Drive. Then, the extraction of data begins by executing these lines of code:

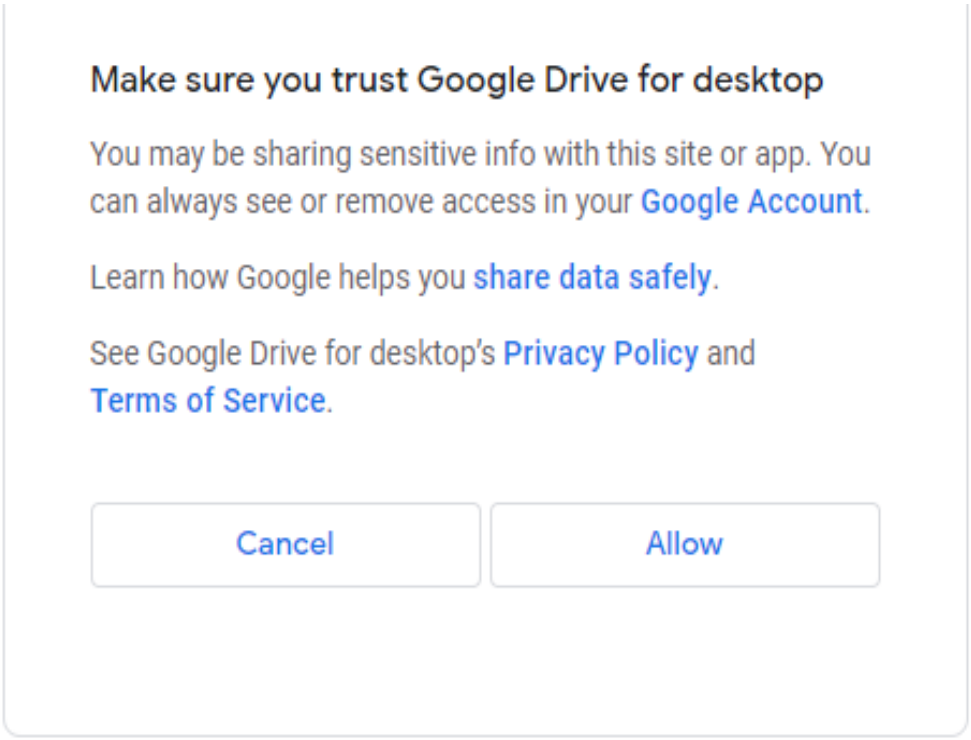
```
[2]  from google.colab import drive
     drive.mount('/content/gdrive')
```

By executing, a pop-up message will appear to permit the notebook in connecting to Google Drive:



Next, the user is redirected to a page to select an account to be connected to Google Drive and allowing the permission to be connected:





Upon allowing, the user is returned to the notebook environment. As an indication of a successful mount from Google Drive, the output is produced below. Thus, the dataset can be called to proceed with the subsequent step.

```
from google.colab import drive
drive.mount('/content/gdrive')
```

```
Mounted at /content/gdrive
```

```
df1 = pd.read_csv('/content/gdrive/MyDrive/rp/dataset.csv')
```

4 The Required Python Libraries for the Implementation of the Project

The Python Libraries utilised for the entire project implementation is illustrated below:

```
import pandas as pd
import re
import numpy as np

# LDA Model
import gensim
from gensim.utils import simple_preprocess
import gensim.corpora as corpora
from pprint import pprint
!pip install -U spacy
import spacy
import nltk
from nltk.tokenize import word_tokenize
from nltk.corpus import stopwords
nltk.download('stopwords')
from gensim import corpora, models, similarities

# Deploying the application
!pip install anvil-uplink
import anvil.server

# Evaluation of LDA model
from sklearn.metrics.pairwise import cosine_similarity
from collections import OrderedDict
```

5 Recommender Demonstration

The following code that is not commented is added to connect to Anvil:

```
# Connecting the script

anvil.server.connect("WZEQQX5P62VGZB7HOACJT5QG-PMZICLHDCMXLCP7E")

@anvil.server.callable
# def band_recommender(mbti):
#   recommendation_scores = []
#   for idx,row in vertical_stack.iterrows():
#     if row.name[0] == mbti:
#       for each in mbti:
#         sims = index[lda_vectors]
#         sims = list(enumerate(sims))
#         for sim in sims:
#           recommendation_score = [row.name[1], sim[1]]
#           recommendation_scores.append(recommendation_score)
#         recommendation = sorted(recommendation_scores, key=lambda x: x[1], reverse=True)
#       return 'Here are the band recommendations for {}:'.format(mbti) + " " + str(recommendation[0])

anvil.server.wait_forever()
```

References

Zakka, K. (2020). Cs231n python tutorial with google colab, `https://colab.research.google.com/github/cs231n/cs231n.github.io/blob/master/python-colab.ipynb`.