

# Configuration Manual

MSc. Research Project  
MSc. in Data Analytics

Priyanka--  
Student ID: x20192037

School of Computing  
National College of Ireland

Supervisor: Prof. Aalok Anant

**National College of Ireland  
Project Submission Sheet  
School of Computing**



<b>Student Name:</b>	Priyanka--
<b>Student ID:</b>	x20192037
<b>Programme:</b>	MSc in Data Analytics
<b>Year:</b>	2021-22
<b>Module:</b>	MSc Research Project
<b>Supervisor:</b>	Prof. Aalok Anant
<b>Submission Due Date:</b>	31-01-22
<b>Project Title:</b>	Configuration Manual
<b>Word Count:</b>	980
<b>Page Count:</b>	9

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.

<b>Signature:</b>	Priyanka--
<b>Date:</b>	29 January 2022

**PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST:**

Attach a completed copy of this sheet to each project (including multiple copies).	
<b>Attach a Moodle submission receipt of the online project submission</b> , to each project (including multiple copies).	
<b>You must ensure that you retain a HARD COPY of the project</b> , 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.

<b>Office Use Only</b>	
Signature:	
Date:	
Penalty Applied (if applicable):	

# Configuration Manual

Priyanka--  
x20192037

## 1 Introduction

This setup guide for current research on: “Prediction of Ratings from Amazon Reviews and Polarity Analysis using Deep Learning Techniques” helps to execute the developed tasks to do the analysis using multiple customized model ("Sentiment Analysis using BERT | Amazon Review Sentiment Analysis", 2021). This would verify that now the solution works error-free it went seamlessly. This section includes knowledge related to basic device configuration of the system over which the tasks were executed and recommended similar setup can be used to easily reproduce it for future data analysis by other researcher without difficulty.

### 1.1 Data Source

This study is based on public dataset of customer reviews available on Amazon ratings and reviews datasets of English and French language from UK <sup>1</sup> and France <sup>2</sup> including marketplace, customer id, product id, helpful votes, both of which are freely available online. Since several items online and customer tend to get confused with choices. The research uses a subset of "reviews". download page

## 2 System Configuration

### 2.1 Hardware Configuration

The Operating system used in the project is Windows and its specification are displayed in Figure1. With the details of processor and RAM.

- Device name: LAPTOP-NN9581CJ
- Processor: Intel(R) Core (TM) i5-10300H CPU @ 2.50GHz 2.50 GHz
- Installed RAM: 16.0 GB
- System type: 64-bit operating system, x64-based processor

---

<sup>1</sup> [https://s3.amazonaws.com/amazon-reviews-pds/tsv/amazon\\_reviews\\_multilingual\\_UK\\_v1\\_00.tsv.gz](https://s3.amazonaws.com/amazon-reviews-pds/tsv/amazon_reviews_multilingual_UK_v1_00.tsv.gz)

<sup>2</sup> [https://s3.amazonaws.com/amazon-reviews-pds/tsv/amazon\\_reviews\\_multilingual\\_FR\\_v1\\_00.tsv.gz](https://s3.amazonaws.com/amazon-reviews-pds/tsv/amazon_reviews_multilingual_FR_v1_00.tsv.gz)

Windows specifications	
Edition	Windows 10 Home
Version	21H1
Installed on	3/2/2021
OS build	19043.1348
Serial number	PF24F2ZC
Experience	Windows Feature Experience Pack 120.2212.3920.0

*Figure 1 Display the OS Configuration.*

## 2.2 Software Configuration

The code can be executed on the local machine using Anaconda jupyter notebook or on the google colaboratory.

- ❖ Programming Language: Python
- ❖ Integrated Development Environment (IDE): Google Colaboratory

Using Python-based Jupyter notebook contained within Anaconda bundle the code can be run accomplish the work by installing the necessary libraries. The research is done using the google colaboratory.

## 3 Downloads and Installation

### Python

Python is used to carry out here proposed study. This includes a large and notable collection of Machine Learning and Deep Learning compatible models. This even includes a variety of libraries and packages which aid in seamless pre-processing stage, and deployment. As a result, installation of Python is the first prerequisite for writing the program on a machine. This could be done by going to python website<sup>3</sup> download section and installing the program installers for the chosen version based on the windows device. Figure2. Python site, from which the latest version may be downloaded. Once downloaded, this application should be executed by using the installation guide.



*Figure 2. Download page of python website*

<sup>3</sup> <https://www.python.org/downloads/>

## Anaconda

One the famous tool for data analysis in the field of research. It can be downloaded from its website <sup>4</sup> by following steps shown in Figure 3. The download page of the navigator can be seen in below Figure 4 for reference. As it's an open source no license is required to install this package in your system. It is easy and user-friendly package.

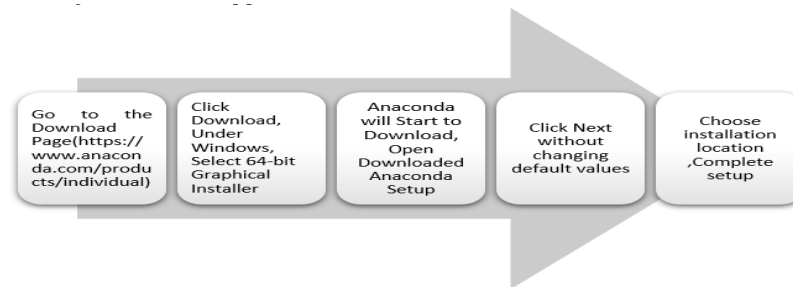


Figure 3. Downloading steps for Anaconda

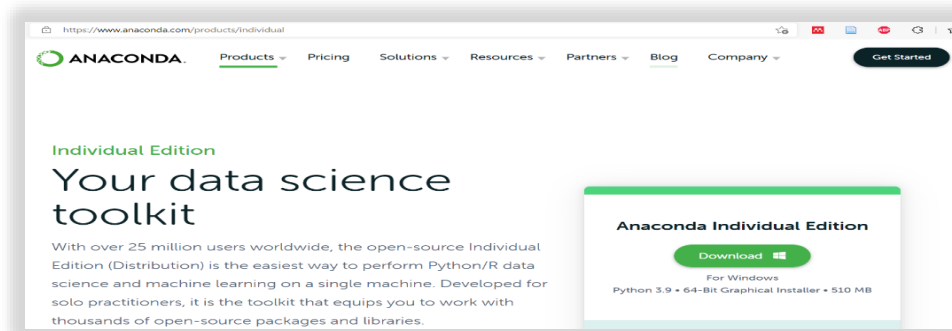


Figure 4. Download page of Anaconda website

Platforms supported are Windows OS and Linux. The submitted paper can run on the same for data analysis on Windows platform but it will need the additional libraries by clicking on Launch Jupyter Figure 6 icon out of various IDE's shown in Figure 5. Downloading the scripts from the submitted code and start running the project scripts.

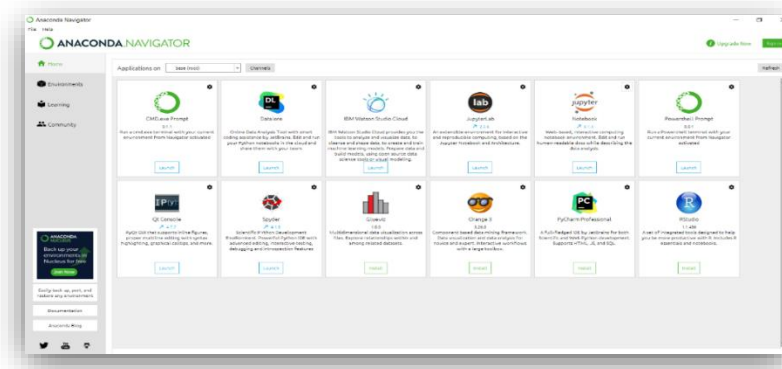


Figure 5. The launch page of the Navigator

<sup>4</sup> <https://www.anaconda.com/products/individual>

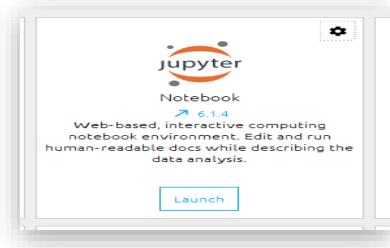


Figure 6. Jupyter launch button in Anaconda

Verification of the installed software version can easily be done by either using the jupyter notebook or the windows command prompt as shown below in Figure 7 and 8.

```

Anaconda Prompt (anaconda3)

(base) C:\Users\Priya>conda --version
conda 4.10.3

(base) C:\Users\Priya>

```

Figure 7. Check Anaconda version

```

Anaconda Prompt (anaconda3)

(base) C:\Users\Priya>Python -V
Python 3.7.6

```

Figure 8. Check Python version

- **Google Colaboratory** is open Jupyter notebook platform which operates fully with internet & needs no configuration. This is available on Cloud Servers and therefore is administered by Google for such advantage of Python programmers to execute and verify their programs online. The google colaboratory<sup>5</sup> need access of the google drive folders<sup>6</sup> that contain scripts from project owner to run the code that will run the program without much need of installing all the libraries.

- Operating System: windows 10
- Random Access Memory (RAM): 12GB (Provided by Colab)
- Hard disk: 15GB (Provided by Colab)
- Languages: Python
- Cloud Platform: Google Colab
- Python libraries: numpy, Pandas, matplotlib, tensorflow, numpy, seaborn and keras.

<sup>5</sup> <https://colab.research.google.com/notebooks/intro.ipynb>

<sup>6</sup> <https://drive.google.com/drive/folders/1DtDFFrZSPUI4A3HZ1ASIIIsre5Wfzml3o?usp=sharing>

## 4 Project development

- Anaconda: If the project is running on local machine.
- Google colaboratory: If the project is running on cloud

As shown in Figure 9, Jupyter Notebook can be used to run the scripts on local machine by uploading the necessary libraries shown in Figure 10. Can be launched from the navigator.

Click on File > New Notebook > Python3

The format of the python form is. ipynb.

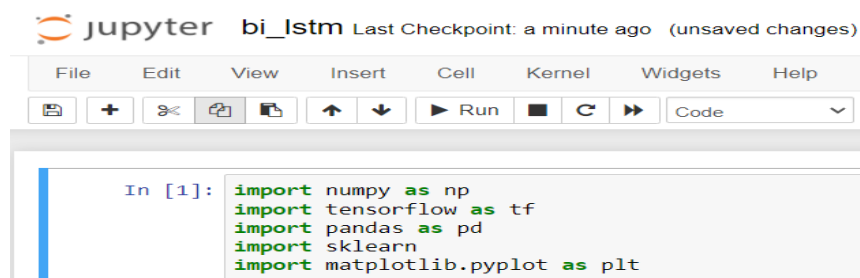


Figure 9. Python on local machine

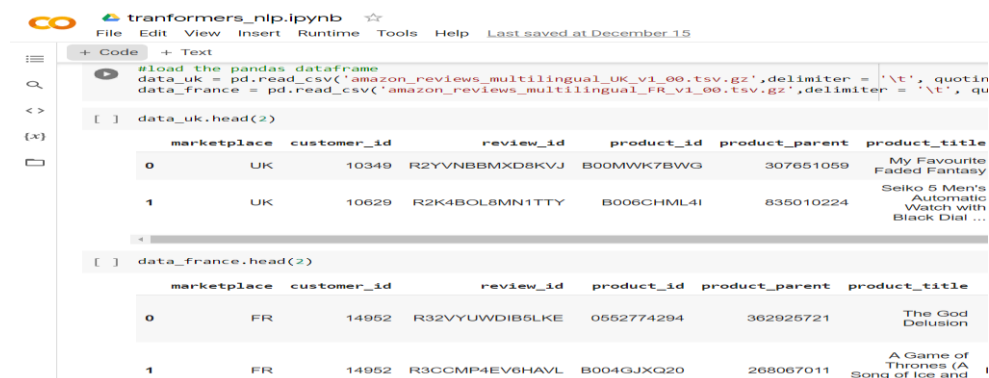


Figure 10. Python file on cloud

### 4.1 Text Pre-processing

The public GitHub repository<sup>7</sup> has the scripts which can be viewed online by login on the git hub website as in Figure 11.

<sup>7</sup> [https://github.com/Priyanka-code-sys/Thesis\\_Priyanka.git](https://github.com/Priyanka-code-sys/Thesis_Priyanka.git)

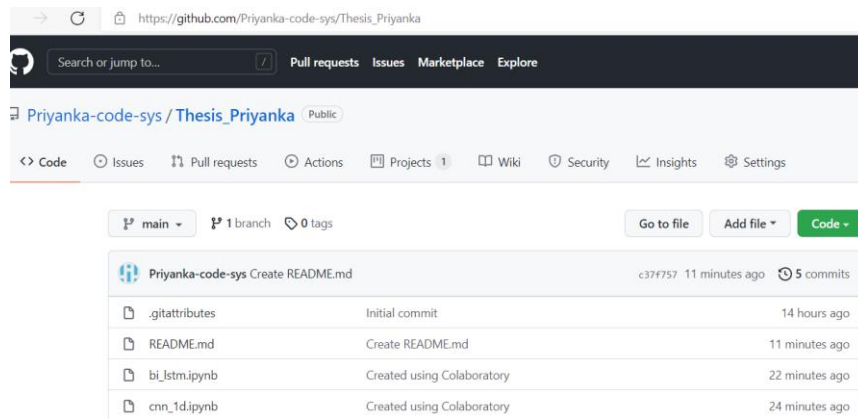


Figure 11. GitHub page website

Request access on google drive for running the script or check git repository location to verify result following below steps

- a) Request access on google drive folder 'csv'.

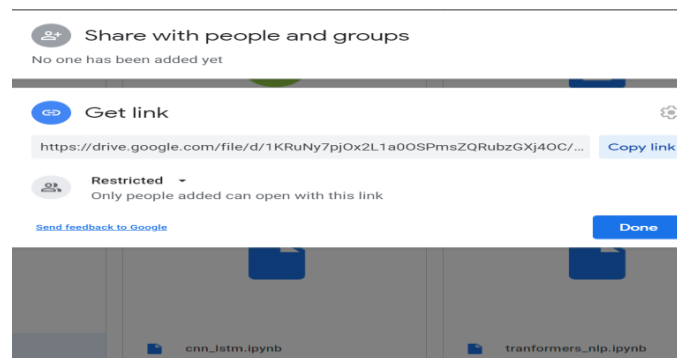


Figure 12 request access using the link in footer.



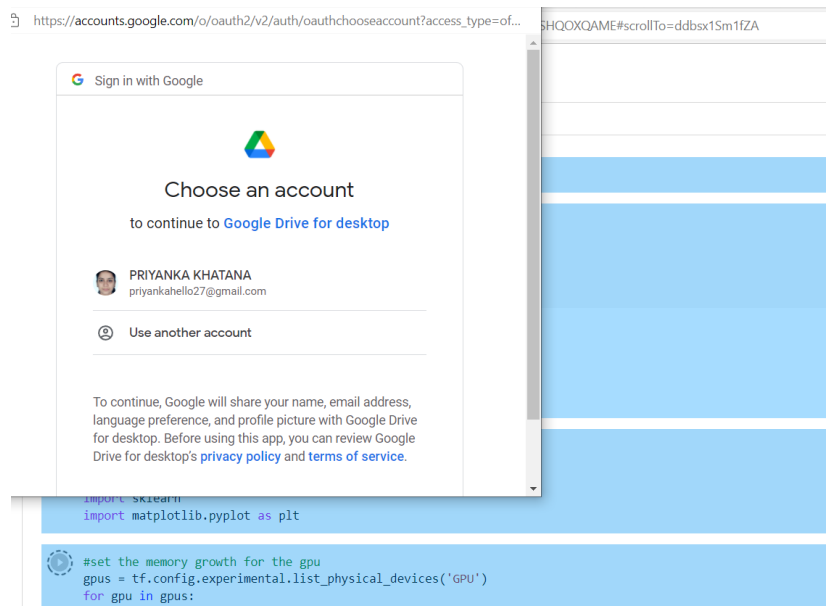


Figure 13 Allow access whenever script is running

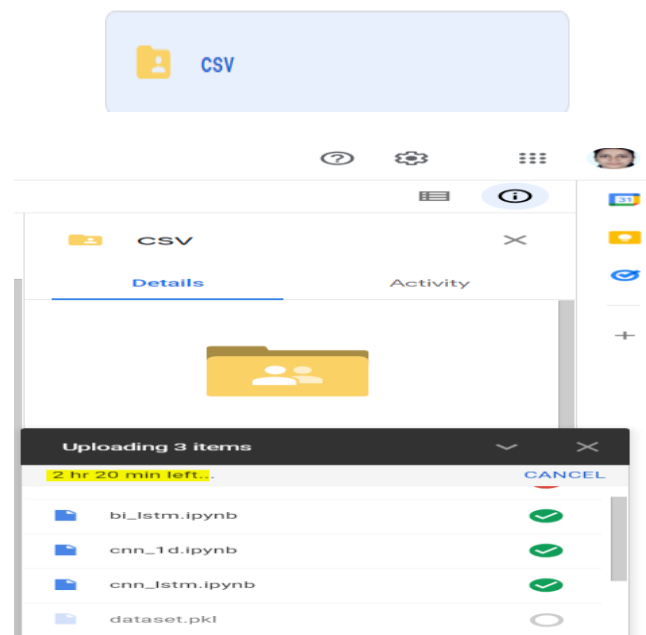


Figure 14 access csv folder on google colab if running on cloud

If access is already given open the folder as shown in Figure 12.

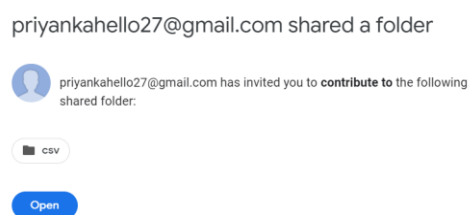


Figure 15 open folder after access granted

b) Run the .ipynb files by double clicking on the files

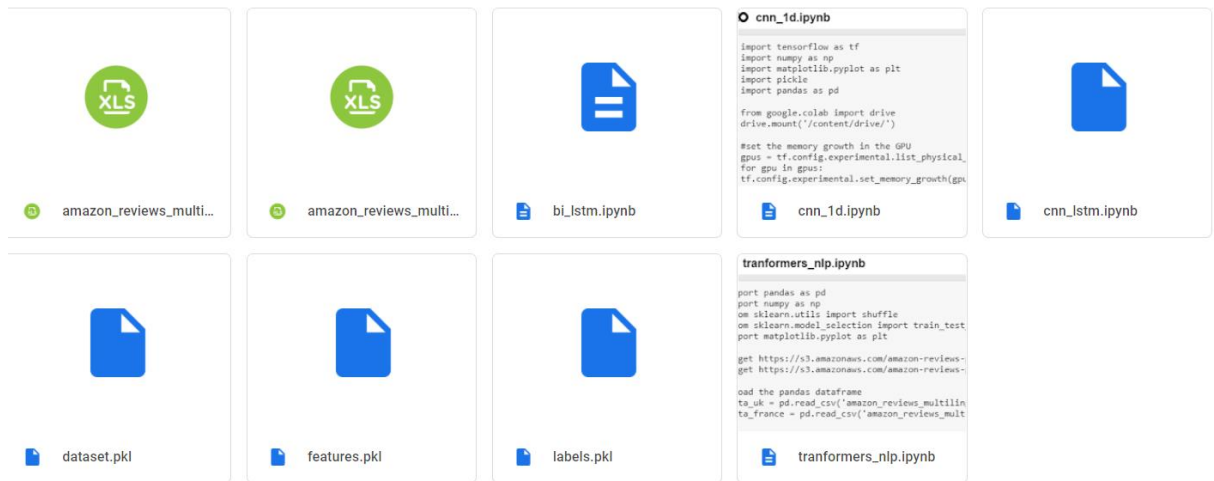
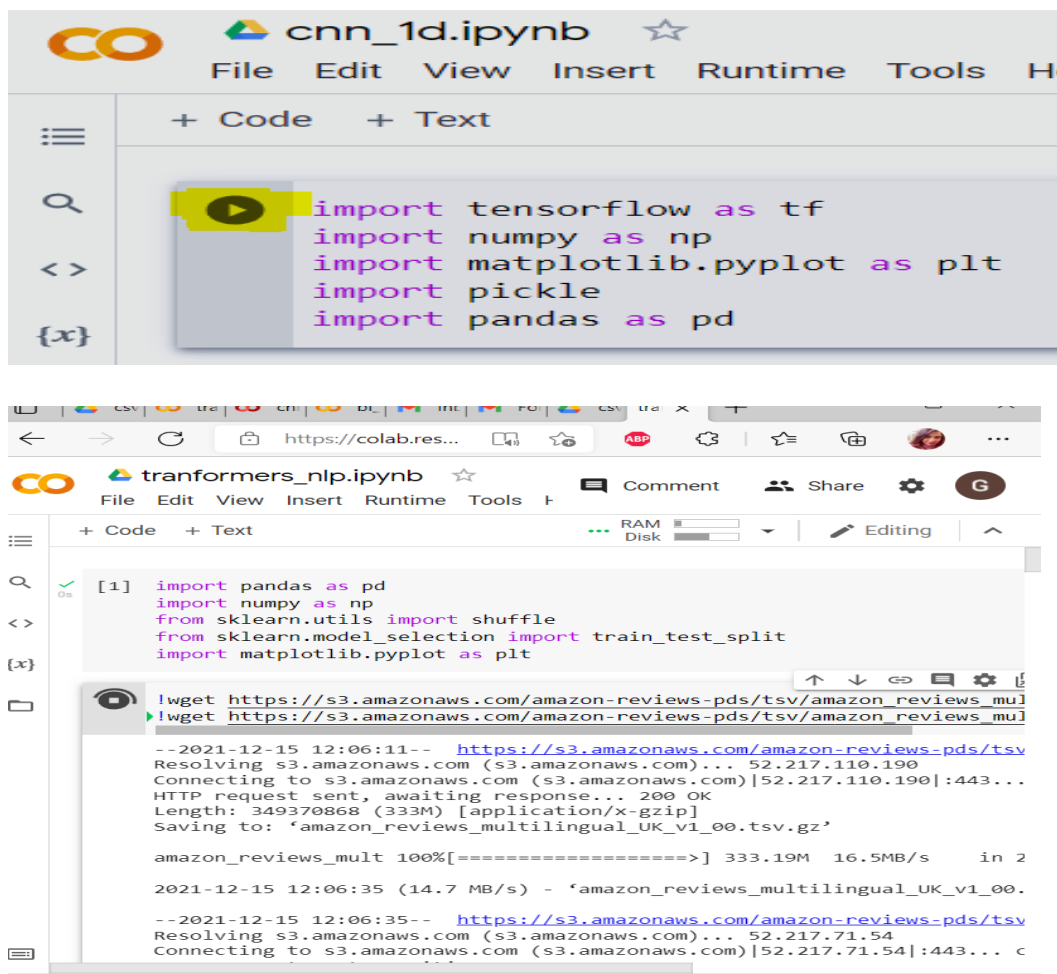


Figure 16 python scripts within csv folder

- c) Use the highlighted button to start executing the code to apply multiple models on the datasets.



```

values = X_test[:3]
print(values)
preds = rnn_model.predict(values)
print("the number of stars in the process = ",np.argmax(preds,axis=1)+1

['awesome this book is really good  although the tutorials are on you
'five stars great sturdy books  ideal for my 1 year old granddaughter
'j adore ce film mais le dvd est de mauvaise qualité car amputé de tou
the number of stars in the process =  [5 5 4]

```

Figure 17 python code

The research is carried out in the google colab. The. ipynb files are executed error free. First the libraries and dataset are called then using the panda library the unstructured data was converted into the data frames as shown in Figure12.

```
#check the dataset
df.head(10)
```

	review	star_rating	word_list	word_len
1059547	disappointing i had this as a sample then acc...	0.2	[disappointing, i, had, this, as, a, sample, t...	28
1220190	origin of the dark knight this film could have...	0.4	[origin, of, the, dark, knight, this, film, co...	199
866978	absolutely fantastic best book i ve read in ...	1.0	[absolutely, fantastic, best, book, i, ve, rea...	55
119482	batterie hs au bout de 5 mois je suis très déç...	0.2	[batterie, hs, au, bout, de, 5, mois, je, suis...	34
120960	rihanna christina alicia et les autres rihann...	0.2	[rihanna, christina, alicia, et, les, autres, ...	249
10010	du pur bonheur dommage du format 4 3 pour le ...	0.8	[du, pur, bonheur, dommage, du, format, 4, 3, ...	30
119452	grosse deception ne pas acheter cet album pou...	0.4	[grosse, deception, ne, pas, acheter, cet, alb...	364
100091	bon film d action un peu dans la lignée de cer...	0.8	[bon, film, d, action, un, peu, dans, la, lign...	50
752237	one star not very good	0.2	[one, star, not, very, good]	5
204613	on passe devant un film sans grand interet un ...	0.4	[on, passe, devant, un, film, sans, grand, int...	44

Figure 18 data shown in data frame using panda's library

## References

Sentiment Analysis using BERT | Amazon Review Sentiment Analysis. (2021). Retrieved 15 December 2021, from <https://www.analyticsvidhya.com/blog/2021/06/amazon-product-review-sentiment-analysis-using-bert/>

Best Python Libraries for Sentiment Analysis. (n.d.). Retrieved December 15, 2021, from <https://thecleverprogrammer.com/2021/06/26/best-python-libraries-for-sentiment-analysis>.