

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

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

National

College of Ireland

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:	Priyanka			
Date:	29 January 2022			
DI FASE DEAD THE FOLLOWING INSTRUCTIONS AND CHECKLIST.				

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

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 UK¹ and France² 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 Figure 1. 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

¹ https://s3.amazonaws.com/amazon-reviews-pds/tsv/amazon_reviews_multilingual_UK_v1_00.tsv.gz

 $^{^{2}\} https://s3.amazonaws.com/amazon-reviews-pds/tsv/amazon_reviews_multilingual_FR_v1_00.tsv.gz$

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 website³ 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

³ 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 ⁴ 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

	Partners + Blog Company + Get Started
Individual Edition	
Your data science	
toolkit	Anaconda Individual Edition
toolkit With over 25 million users worldwide, the open-source Individual	
	Download 🛤
With over 25 million users worldwide, the open-source Individual	

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.

Onineer	DA NAVIGATOR					O Upgrade fo	
ft flore	Applications on 2404 (1960	+ Chennela					12 ⁵ 22 ⁵
Courses to the test the test to the test to the test test to the test test test test test test test	•	•	•	•		•	1
🖨 Learning	0	PL S	Ö.	lab	Jupyter	0	
A Connunts	CMD.eve Prompt	Datalore	IBM Watson Studio Cloud	Appletab P 711	Notebook	Pouershell Prompt	
	Ban a cendulose tar ennal with your carrient, environment Prom Navigator schladed	Online Data Analysis Tool with smart coding assistance by artificana, Bill and run	IBM Mattern Stadio David provides you the Looks to sharyon and vession date, to	An exclamable environment for interactive and reproducible computing, based on the	Web based, representation strengting relations an uncoverent, Bill and run	Run of Power cheft terrorised with your current environment from tewastor	
		year typhan notationals in the cloud and share them with your team.	clasmas and phase data, to press and train much me loar ring models. Prepare cats and build models, using toen source data	Jugging freemon and trafficulture.	human readable dras while days fields the data analysis	activered	
	Land Land	Laures	science table or elast modeling.	(and)	LANCE .	LAUNCE	
		Annual I			-		
	•	0	• •	•	e e	<pre> ° </pre>	
	1 P (y):	28 C		00	PC 1		
	Qt Censole	Soyder C + 12	G)(ex/G 18.5	Orange 3 3243	PyCharm Professional	#Studio 11.439	
O MINISTER	Ayor Old that subjects in the Figures, proper multitue adding with system reporting, protect outputs, and mice	Scientific Enthern Development Bauffornizers, Presented Pythen IDE with advanced editing, interactive cesting,	Multidevantional data visualization across New Reprint relationships within and among related decosits	Connective based data mining Remework. Data encodential and esta contains for routes and expert. Interactive workflows.	A full-fielded OE by certinality for both forms Fit and West Pyrton development, foreports 47545, 35 and 905	A set of respected tools designed to help you be ensure productive solth 8 includes 8 appendix and conduction.	
Back up your	- Herdening, granter strongs, and early	debugging and introduction features	Tricit International	with a large builder.	suggestion of the state of	The case and receivers.	
Plantenes for types		LAUNCH				T-MARK .	
Caulty best on part, and							
raiding any antimatment							
Decamentation							
Andendu Biog							

Figure 5. The launch page of the Navigator

⁴ 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≻condaversion 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 in 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 colaboratory⁵ need access of the google drive folders⁶ 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.

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

⁶ https://drive.google.com/drive/folders/1DtDFFrZSPUl4A3HZ1ASIIsre5Wfzml3o?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.

💭 Jupyter	bi_lstm Last	Checkpoint: a mi	nute ago (unsav	ved changes)
File Edit	View Insert	Cell Kerr	nel Widgets	Help
🖹 🕇 🎉 4	2 🗈 🛧 🔸	► Run 🔳	C Dode	~
In [1]:	<pre>import numpy import tensor import pandas import sklear import matplo</pre>	flow as tf as pd n	as plt	

Figure 9.Python on local machine

a	•		ranformer Edit View			ols Help <u>Last saved</u>	at December 15			
:=	+ Code + Text									
Q	#load the pandas dataframe data_uk = pd.read_csv('amazon_reviews_multilingual_UK_v1_00.tsv.gz',delimiter = '\t', quot data_france = pd.read_csv('amazon_reviews_multilingual_FR_v1_00.tsv.gz',delimiter = '\t',									
< >	> [] data_uk.head(2)									
$\{x\}$			marketp	lace	customer_id	review_id	product_id	product_parent	product_title	
			0	UK	10349	R2YVNBBMXD8KVJ	BOOMWK7BWG	307651059	My Favourite Faded Fantasy	
			1	UK	10629	R2K4BOL8MN1TTY	B006CHML4	835010224	Seiko 5 Men's Automatic Watch with Black Dial	
			•							
		Ε.]	data_france	.head	1(2)					
			marketp	lace	customer_id	review_id	product_id	product_parent	product_title	
			0	FR	14952	R32VYUWDIB5LKE	0552774294	362925721	The God Delusion	
			1	FR	14952	R3CCMP4EV6HAVL	B004GJXQ20	268067011	A Game of Thrones (A Song of Ice and	

Figure 10.Python file on cloud

4.1 Text Pre-processing

The public GitHub repository⁷ has the scripts which can be viewed online by login on the git hub website as in Figure 11.

⁷ https://github.com/Priyanka-code-sys/Thesis_Priyanka.git

	ch or jump to / Pul	l requests Issues Marketplace Explore	
Priyank	a-code-sys / Thesis_Priyanka	Public	
> Code	⊙ Issues 11 Pull requests ⊙ A	Actions 🖳 Projects 1 🔲 Wiki 🕐 Security	r 🗠 Insights 🕸 Settings
	🕼 main - 😵 1 branch 💿 0	tags	Go to file Add file * Code
	Priyanka-code-sys Create REA	DME.md	c37f757 11 minutes ago 🛛 5 commit
	gitattributes	Initial commit	14 hours age
	C README.md	Create README.md	11 minutes age
	🗅 bi_lstm.ipynb	Created using Colaboratory	22 minutes age

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.

$\label{eq:linear} \begin{tabular}{lllllllllllllllllllllllllllllllllll$	SHQOXQAME#scrollTo=ddbsx1Sm1fZA
	A
G Sign in with Google	
Choose an account	
to continue to Google Drive for desktop	
to continue to boogle prive for desixep	
PRIYANKA KHATANA	
priyankahello27@gmail.com	
② Use another account	
To continue, Google will share your name, email address,	
language preference, and profile picture with Google Drive	
for desktop. Before using this app, you can review Google	
Drive for desktop's privacy policy and terms of service.	
TINDOLE SKTGALI	*
<pre>import matplotlib.pyplot as plt</pre>	
() #set the memory growth for the gpu	
<pre>gpus = tf.config.experimental.list_physical_devices('GPU')</pre>	
for gpu in gpus:	

Figure 13 Allow access whenever script is running

	csv				
		0	: []]		50
				Ō	51
	csv			\times	
	Details		Activity		ø
					+
Upl	oading 3 items			~	\times
2 hr	20 min left. <mark>.</mark> .			CAN	CEL
	bi_lstm.ipynb			\bigcirc	
	cnn_1d.ipynb			\bigcirc	
	cnn_lstm.ipynb			\bigcirc	
	dataset.pkl			\bigcirc	

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.





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 Figure 12.

#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 <u>https://www.analyticsvidhya.com/blog/2021/06/amazon-product-review-sentiment-analysis-using-bert/</u>

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.