Configuration Manual

MSc Research Project
MSCDAD_B

Renjitha Methanath
Student ID: x23308532

School of Computing
National College of Ireland

Supervisor: Bharat Agarwal

National College of Ireland

MSc Project Submission Sheet

School of Computing

Student Name:	…….Renjitha Methanath………………………………………………………………
Student ID:	………x23308532…………………………………………………………………………
Programme:	…………MSCDAD_B……………………………… **Year:** 2024-25
Module:	………MSc Research Practicum………………………………………………………
Lecturer:	………Bharat Agarwal…………………………………………………………………
Submission Due Date:	…………11-08-2025………………………………………………………………….…
Project Title:	…Leveraging BERT-GRU for Sentiment-Driven Demand Forecasting and Product Insight in E-Commerce…
Word Count:	…………………300…………………… **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.

Signature: ………Renjitha Methanath……………………………………………………

Date: …………10-08-2025………………………………………………………

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 Report

Renjitha Methanath
x23308532
MSCDAD_B
National College of Ireland

This report outlines the required software tools, libraries, and settings necessary to successfully replicate the experimental environment used in the sentiment analysis and trend forecasting project.

1 Operating System

The experiments were conducted on a system compatible with Python and PyTorch environments.

Recommended OS:

- Windows 10/11 (64-bit)
- Ubuntu 20.04+ or any Linux distribution with Python support
- macOS (for development, not recommended for GPU training with CUDA)

2 Hardware Requirements

- **CPU:** Intel i5/i7 or AMD equivalent
- **RAM:** Minimum 8 GB (16 GB recommended)
- **GPU:** NVIDIA GPU with CUDA support (e.g., GTX 1650, RTX 2060 or higher)
- **Storage:** At least 5 GB free space for datasets and model checkpoints

3 Programming Language

Python Version: 3.8 or higher (tested with Python 3.9)

4 Dataset Configuration

The dataset file must be named **Review_Dataset.csv** and placed in the working directory.

Required columns:

- asins – Product identifier
- reviews.text – Review text
- reviews.rating – Numerical rating (1-5)
- reviews.doRecommend – Boolean recommendation field
- reviews.date – Review date for time-series analysis

5 Required Python Libraries

Install all dependencies via pip:

```
pip install pandas numpy matplotlib seaborn scikit-learn nltk wordcloud statsmodels torch
    torchvision transformers
```

List of Required Libraries

Library	Version	Usage in Project
pandas	1.5.3	Data manipulation, cleaning, and handling datasets
matplotlib	3.7.1	Visualization of results and time series plots
seaborn	0.12.2	Enhanced statistical data visualization
scikit-learn (`sklearn`)	1.3.0	Model training (Logistic Regression), feature extraction (TF-IDF), evaluation metrics, label encoding, train-test splitting
nltk	3.8.1	Natural Language Processing tasks like tokenization, stemming, lemmatization, stopword removal
re (regex)	Built-in (Python 3.10)	Text cleaning and preprocessing (removing special characters, URLs, etc.)
numpy	1.24.3	Numerical computations, array operations, and ROC/AUC calculations
transformers (HuggingFace)	4.30.0	BERT Tokenization, loading BERT model, fine-tuning for sentiment classification
torch (PyTorch)	2.0.1	Building, training, and optimizing neural network models
wordcloud	1.9.2	Generating visual word clouds from review text
statsmodels	0.14.0	Time series analysis, ARIMA/SARIMA modeling, and statistical testing (ADF test)
time	Built-in (Python 3.10)	Measuring execution/training time

6 NLTK Resource Downloads

The following NLTK resources must be downloaded before running the script:

```
import nltk
nltk.download('stopwords')
nltk.download('wordnet')
```

7 Model Configurations

7.1 Traditional ML Models

- Vectorization: TF-IDF with max_features=5000
- Logistic Regression:
  - max_iter=1000
  - Evaluation metrics: Accuracy, Precision, Recall, F1-score, Cohen's Kappa

7.2 Deep Learning Model (BERT + GRU)

- Tokenizer: bert-base-uncased (from HuggingFace Transformers)
- GRU Hidden Dimension: 128
- Dropout: 0.3
- Batch Size: 16
- Optimizer: AdamW with lr=2e-5

- Loss Function: CrossEntropy with class weights
- Epochs: 3
- Fine-tuning: Two variants were trained (Frozen BERT vs. Fine-tuned BERT)

8 Time Series Forecasting Configuration

Models Used: ARIMA, SARIMAX
Orders Tested:

- ARIMA: (1,0,1), (1,0,2), (2,0,1), (2,0,2), (2,1,1), (2,1,2), (3,0,1), (3,1,2) and (4,0,1)
- SARIMAX: (1,1,1), (2,1,1), (2,1,2), (3,1,2), (1,0,1), (2,0,2), (4,1,2), (4,1,3) and (4,1,4)

Exogenous Variables for SARIMAX: recommend_rate and sentiment_score

9 Visualization Settings

- Dark background theme applied to all plots.
- Custom color palette for sentiment distribution:
  - Positive: Red (#FF0000)
  - Neutral: Green (#00FF00)
  - Negative: White (#FFFFFF)
- All generated plots (e.g., confusion matrices, word clouds, ARIMA/SARIMAX forecasts) are saved as .png files in the working directory.

10 Additional Notes

- Ensure CUDA is enabled for faster training if using a GPU.
- If running on CPU, expect slower BERT training times.
- Ensure all random seeds (`random_state=42`) are set to maintain reproducibility.

Table of Contents for Reserach Report

References

Ahmed, M., Chen, Q., & Li, Z. (2020). Constructing domain-dependent sentiment dictionary for sentiment analysis. *Neural Computing and Applications*, 32, 14719–14732. https://doi.org/10.1007/s00521-020-04824-8

Akhtar, M. S., Ekbal, A., Bhattacharyya, A., et al. (2017). A multilayer perceptron based ensemble technique for fine-grained financial sentiment analysis. *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, 540–546. https://doi.org/10.18653/v1/D17-1057

Alamdari, P. M., Navimipour, N. J., Hosseinzadeh, M., Safaei, A. A., & Darwesh, A. A. (2020). Systematic study on the recommender systems in the e-commerce. *IEEE Access*, 8, 115694–115716. https://doi.org/10.1109/ACCESS.2020.3002803

AlBadani, B., Shi, R., & Dong, J. (2022). A novel machine learning approach for sentiment analysis on twitter incorporating the universal language model fine-tuning and svm. *Applied System Innovation*, 5(1), 13. https://www.mdpi.com/2571-5577/5/1/13

Alzahrani, M. E., Aldhyani, T. H., Alsubari, S. N., Althobaiti, M. M., & Fahad, A. (2022). Developing an intelligent system with deep learning algorithms for sentiment analysis of e-commerce product reviews. *Computational Intelligence and Neuroscience*, 1–10. https://doi.org/10.1155/2022/3840071

Antony, C., et al. (2025). A systematic review of sentiment analysis approaches and techniques. *2025 6th International Conference on Mobile Computing and Sustainable Informatics (ICMCSI).* https://doi.org/10.1109/ICMCSI64620.2025.10883219

Assiri, A., & Gumaei, A. (2024). Deberta-gru: Sentiment analysis for large language model. *Computers, Materials & Continua.* https://doi.org/10.32604/cmc.2024.050781

Bansal, B., & Srivastava, S. (2019). Hybrid attribute based sentiment classification of online reviews for consumer intelligence. *Applied Intelligence*, 49(1), 137–149. https://link.springer.com/article/10.1007/s10489-018-1299-7

Birjali, M., Kasri, M., & Beni-Hssane, A. (2021). A comprehensive survey on sentiment analysis: Approaches, challenges and trends. *Knowledge-Based Systems*, 226, 107134. https://doi.org/10.1016/j.knosys.2021.107134

Cambria, E., Das, D., Bandyopadhyay, S., & Feraco, A. (2017). Affective computing and sentiment analysis. In *A practical guide to sentiment analysis: Socio-affective computing* (pp. 1–10). Springer. https://doi.org/10.1007/978-3-319-55394-8_1

Chen, Y. (2024). Sentiment analysis based on multiscale convolutional neural network and bidirectional long short-term memory. *Membrane Technology.* https://doi.org/10.52710/mt.97

Chintalapudi, N., Battineni, G., Canio, M. D., Sagaro, G., & Amenta, F. (2021). Text mining with sentiment analysis on seafarers' medical documents. *International Journal of Information Management Data Insights*, 1(1), 100005. https://doi.org/10.1016/j.jjimei.2020.100005

Darraz, N. (2023). Using sentiment analysis to spot trending products. *2023 Sixth International Conference on Vocational Education and Electrical Engineering (ICVEE).* https://doi.org/10.1109/ICVEE59738.2023.10348252

Geetha, M. P., & Renuka, D. K. (2021). Improving the performance of aspect based sentiment analysis using fine-tuned bert base uncased model. *International Journal of Intelligent Networks*, 2, 64–69. https://doi.org/10.1016/j.ijin.2021.06.005

Hussein, D. M. (2018). A survey on sentiment analysis challenges. *Journal of King Saud University-Engineering Sciences*, 30(4), 330–338. https://doi.org/10.1016/j.jksues.2016.04.002

Iddrisu, A.-M., Mensah, S., Boafo, F., Yeluripati, G. R., & Kudjo, P. A. (2023). A sentiment analysis framework to classify instances of sarcastic sentiments within the aviation sector. *International Journal of Information Management Data Insights*, 3(2), 100180. https://doi.org/10.1016/j.jjimei.2023.100180

Jayakody, J., et al. (2021). Sentiment analysis on product reviews on twitter using machine learning approaches. *2021 International Conference on Decision Aid Sciences and Application (DASA)*. https://ieeexplore.ieee.org/document/9682291

Nimmi, K., et al. (2022). Pre-trained ensemble model for identification of emotion during covid-19 based on emergency response support system dataset. *Applied Soft Computing*, 122, 108842. https://pubmed.ncbi.nlm.nih.gov/35465357/

Obiedat, R., Al-Darras, D., Alzaghoul, E., & Harfoushi, O. (2021). Arabic aspect-based sentiment analysis: A systematic literature review. *IEEE Access*, 9, 152628–152645. https://doi.org/10.1109/ACCESS.2021.3127140

Pang, B., Lee, L., & Vaithyanathan, S. (2002). Thumbs up? sentiment classification using machine learning techniques. https://aclanthology.org/W02-1011/

Sanyal, S., & Wamique, M. (2019). Factors affecting customer satisfaction with e-commerce websites– an omani perspective. *2019 International Conference on Digitization (ICD)*. https://doi.org/10.1109/ICD47981.2019.9105780

Savci, P., & Das, B. (2023). Prediction of the customers' interests using sentiment analysis in e-commerce data for comparison of arabic, english, and turkish languages. *Journal of King Saud University- Computer and Information Sciences*, 35(3), 227–237. https://doi.org/10.1016/j.jksuci.2023.02.017

Singh, R., et al. (2024). Ai-driven multi-modal demand forecasting: Combining social media sentiment with economic indicators and market trends. *Journal of Informatics Education and Research*, 4(3). https://www.researchgate.net/publication/387075739

Sirisha, U., & Bolem, S. C. (2022). Aspect-based sentiment & emotion analysis with roberta, lstm. *International Journal of Advanced Computer Science and Applications (IJACSA)*. https://doi.org/10.14569/IJACSA.2022.0131189

Tubishat, M., Idris, N., & Abushariah, M. (2021). Explicit aspects extraction in sentiment analysis using optimal rules combination. *Future Generation Computer Systems*, 114, 448–480. https://doi.org/10.1016/j.future.2020.08.019

Zhang, J., Zhang, A., Liu, D., & Bian, Y. (2021). Customer preferences extraction for air purifiers based on fine-grained sentiment analysis of online reviews. *Knowledge-Based Systems*, 228, 107259. https://doi.org/10.1016/j.knosys.2021.107259

Zhang, Y., et al. (2024). A comparative study of sentiment analysis on customer reviews using machine learning and deep learning. *Computers*, 13(12), 340. https://doi.org/10.3390/computers13120340