INTERNSHIP PROJECT

SENTIMENT ANALYSIS

REPORT

The project "Sentiment Analysis " aims to harness the power of natural language processing and machine learning to provide actionable insights from textual data. It helps in understanding Business domains.

PREPARED BY

NAME: AADHYA GAJJAR

SPECIALISATION (AI &ML) 3RD YEAR

*Submitted to*



**Department of Computer Science & Engineering**

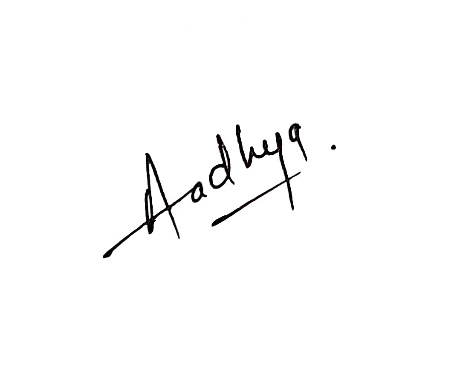
**Sir Padampat Singhania University**

**Udaipur 313601 Rajasthan India**

**DECLARATION**

I AADHYA GAJJAR, student of B.Tech.(CSE), hereby declare that the 2-Month Summer Internship project report titled “SENTIMENT ANALYSIS OF SOCIAL MEDIA” which is submitted by me to the department of Computer Science & Engineering , School of Engineering, Sir Padampat Singhania University, Udaipur, submitted towards the partial fulfillment of the requirement for the award of the degree of Bachelor of Technology, has not been previously formed the basis for the award of any degree, diploma or other similar title or recognition.

Name and signature of Student: AADHYA GAJJAR



Udaipur:

Date:

**CERTIFICATE (Attach Coloured Copy of Internship Certificate as well)**

This is to certify that the 2-Month Summer Internship project entitled ‘sentimental analysis of social media’ being submitted by Aadhya gajjar, submitted towards the partial fulfillment of the requirement for the award of the degree of Bachelor of Technology, has been carried out under my supervision and guidance.

The matter embodied in this report has not been submitted, in part or in full, to any other university or institute for the award of any degree, diploma or certificate.

Sevy Singh

Trainer

Gurugram

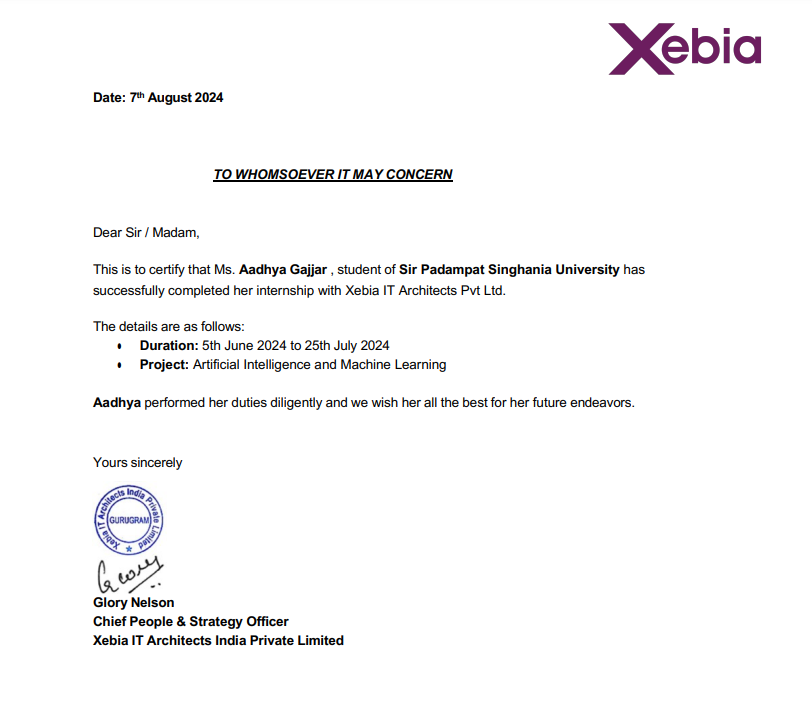
Prof. Alok kumar

Head of department

Department of Computer Science & Engineering

Sir Padampat Singhania University

Udaipur 313601 Rajasthan India



**Acknowledgement**

I would like to express my sincere gratitude to my project guide **Sevy Singh** for giving me the opportunity to work on this topic.

It would never be possible for us to take this project to this level without his innovative ideas and his relentless support and encouragement.I would like to express my deepest gratitude to everyone who supported and guided me throughout the course of this project on sentiment analysis of social media.

I am also grateful to xebia which provided me with the necessary resources and environment to carry out this research. The support from the faculty and peers has been incredibly motivating.

Special thanks to my family and friends for their unwavering support and patience during this period. Their encouragement and understanding were vital to the successful completion of this project. Finally, I extend my gratitude to the online communities and platforms that provided open-source tools and libraries. These resources were crucial in the development and implementation of the sentiment analysis models used in this study.

Thank you all for your contributions, without which this project would not have been possible.

Aadhya Gajjar

21CS002356

**Abstract**

This project focuses on developing a sentiment analysis model to evaluate and interpret emotions expressed in social media posts using Python. With the exponential growth of user-generated content on platforms such as Twitter, Facebook, and Reddit, there is a significant demand for automated tools that can efficiently analyze and categorize sentiments in real-time. This project employs machine learning techniques, including logistic regression and random forest classifiers, to build a robust model for sentiment classification.

The data is preprocessed using natural language processing (NLP) techniques like tokenization, stopword removal, and lemmatization to prepare it for model training. Various feature extraction methods, such as TF-IDF and word embeddings, are explored to capture the contextual meaning of words. The models are trained and evaluated on labeled datasets, such as the Twitter Sentiment Analysis Dataset, and their performance is compared using accuracy, precision, recall, and F1-score metrics.

Additionally, the project investigates the impact of hyperparameter tuning and model optimization to improve classification accuracy. The final model is capable of categorizing social media posts into positive, negative, or neutral sentiments with high accuracy, making it a valuable tool for brand monitoring, opinion mining, and social media analysis.

**CONTENTS**

**DECLARATION i**

**CERTIFICATE** ii

**ACKNOWLEDGEMENT** iii

**ABSTRACT i**v

**TABLE OF CONTENTS** v

**LIST OF TABLES** vi

**LIST OF FIGURES** vii

**LIST OF ABBREVATIONS** viii

**CHAPTER 1 1**

**INTRODUCTION**  1

**CHAPTER 2 2**

**LITERATURE REVIEW 2**

**CHAPTER 3 3**

**SOFWARE DESIGN**

**CHAPTER 4 4**

**METHODLOGY**

**CHAPTER 5 5**

**CONFUSION MATRIX**

**CHAPTER 6 6-14**

**CODE IMPLEMENTATION**

**CHAPTER 7 15**

**POWER DASHBOARD**

**CHAPTER 8 16**

**RESULTS AND DESCRIPTION**

**CONCLUSION AND FUTURE WORK 17**

**REFERENCES**

**CHAPTER 1**

**INTRODUCTION**

* **BACKGROUND**

In the era of digital communication, vast amounts of textual data are generated daily across various platforms such as social media, customer reviews, blogs, and forums. Extracting meaningful insights from this unstructured data is crucial for businesses, researchers, and policymakers. Sentiment analysis, also known as opinion mining, is a subfield of natural language processing (NLP) that focuses on identifying and categorizing opinions expressed in text to determine the writer's attitude towards a particular topic, product, or service.

* **ABOUT**

Sentiment analysis has gained significant attention due to its practical applications across multiple domains. Businesses utilize sentiment analysis to monitor customer feedback and brand reputation, enabling them to make informed decisions and improve customer satisfaction. In politics, sentiment analysis helps gauge public opinion on policies and candidates, while in finance, it aids in predicting market trends based on news and social media sentiment. Moreover, it plays a critical role in public health by analyzing sentiments around health campaigns and crises, providing valuable insights for better response strategies.

The primary objective of this project is to develop a robust sentiment analysis model using Python that can accurately classify text data into positive, negative, or neutral sentiments. The project aims to:

1. **Collect and preprocess** a relevant dataset to ensure high-quality input for the model.
2. **Implement and compare** various sentiment analysis models, including traditional machine learning algorithms and advanced deep learning techniques.
3. **Evaluate** the performance of these models using standard metrics such as accuracy, precision, recall, and F1-score.
4. **Analyze and interpret** the results to identify the most effective model and highlight potential areas for improvement.

**CHAPTER 2**

# LITERATURE SURVEY

# Sentiment Analysis and Opinion Mining (2012) - Bing Liu Provides a foundational overview of sentiment analysis techniques, including lexicon-based and early machine learning methods. Highlights challenges like handling negation and subjectivity in text. *Reference*: Liu, B. (2012). Sentiment Analysis and Opinion Mining. *Synthesis Lectures on Human Language Technologies*, 5(1), 1-167.

# A Survey on Sentiment Analysis: Challenges, Applications and Deep Learning (2020) Focuses on the evolution of sentiment analysis with a strong emphasis on deep learning models. Discusses the challenges specific to social media, such as sarcasm and informal language. *Reference*: Gaber, M. M., Badie, M., & Kora, R. (2020). A Survey on Sentiment Analysis: Challenges, Applications and Deep Learning. *Expert Systems with Applications*. [Link](https://www.sciencedirect.com/science/article/pii/S0957417420304099)

# Sentiment Analysis on Social Media: Recent Trends and Challenges (2017) Explores the recent trends in sentiment analysis specifically focused on social media, addressing the challenges like slang, sarcasm, and the real-time nature of social media data. *Reference*: Pang, B., & Lee, L. (2017). *IEEE Transactions on Knowledge and Data Engineering*. [Link](https://ieeexplore.ieee.org/document/7839107)

# Machine Learning Techniques for Sentiment Analysis: A Review (2018) Reviews various machine learning models, including Naive Bayes and Random Forest, and compares their effectiveness in sentiment analysis tasks. *Reference*: Shubhankar, R., Shruti, S., & Siddharth, G. (2018). *Procedia Computer Science*, 132, 324-329. [Link](https://www.sciencedirect.com/science/article/pii/S2210832718301787)

# Social Media Analytics: A Survey of Techniques, Tools, and Platforms (2016) Provides an in-depth look at the tools and techniques used in social media analytics, with a focus on sentiment analysis, and discusses the integration of machine learning with social media data. *Reference*: Pal, A., Narayan, S., & Kumar, N. (2016). *ACM Computing Surveys*, 49(2), 1-40. Link

**CHAPTER 3**

# SOFTWARE DESIGN

# 

**CHAPTER 4**

# ****METHODOLOGY****

### **Data Collection**

* **Sources**: The dataset for this project is obtained from Twitter, utilizing the Twitter API to collect tweets. The dataset includes a balanced number of positive, negative, and neutral tweets to ensure comprehensive model training and evaluation.
* **Data Description**: The dataset comprises approximately 10,000 tweets, with each tweet labeled as positive, negative, or neutral. Additional features such as tweet length and user information are also included.

### **Data Preprocessing**

* **Cleaning**: Data cleaning steps include removing URLs, mentions, hashtags, special characters, and stop words. Additionally, all text is converted to lowercase to maintain consistency.
* **Tokenization**: The text is tokenized into individual words or tokens, which can then be used as input features for the models.
* **Normalization**: Techniques such as stemming and lemmatization are applied to normalize the text, reducing words to their base or root forms.

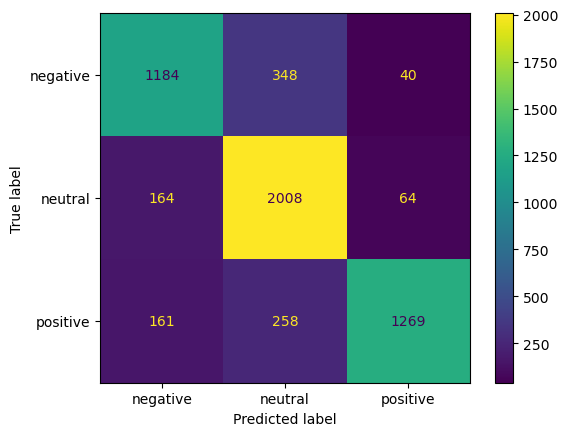
### **Sentiment Analysis Models**

* **Overview**: The models explored in this project include Naive Bayes, SVM, and LSTM. These models are selected based on their proven effectiveness in text classification tasks.
* **Rationale**: Naive Bayes and SVM are traditional machine learning algorithms known for their simplicity and efficiency. LSTM, a type of RNN, is chosen for its ability to capture long-term dependencies in text data.

**CHAPTER 5**

**RESULTS AND DISCUSSIO**

**CONFUSION MATRIX FOR SENTIMENT ANALYSIS**



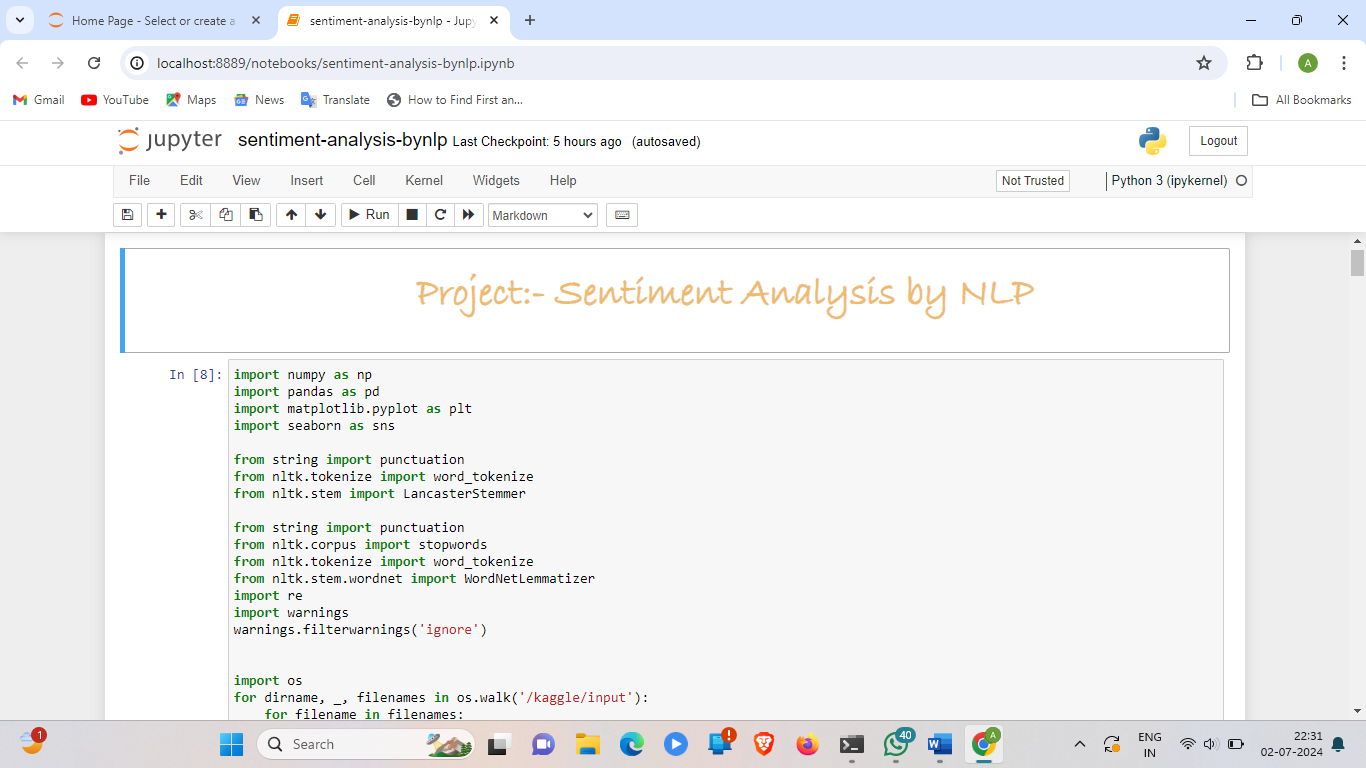
**CHAPTER 6**

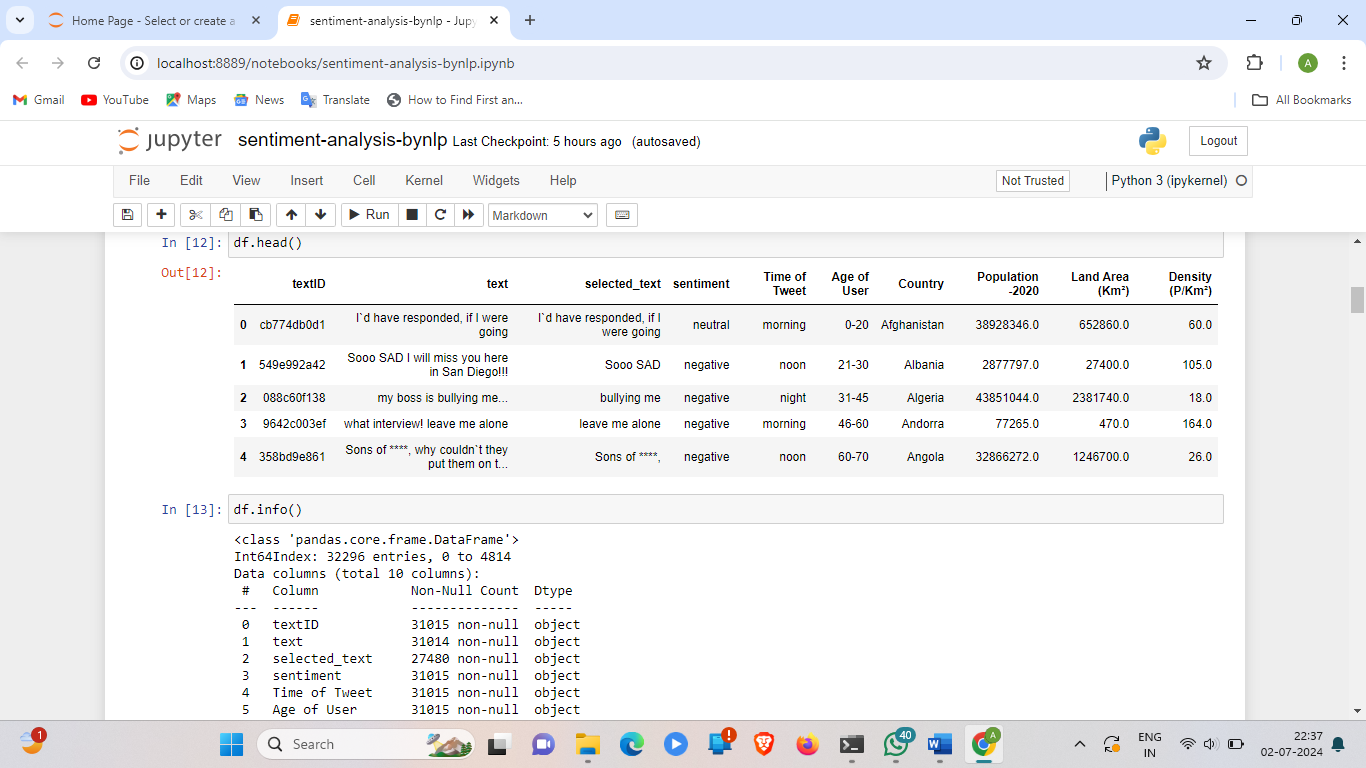
## **IMPLEMENTATION CODE**

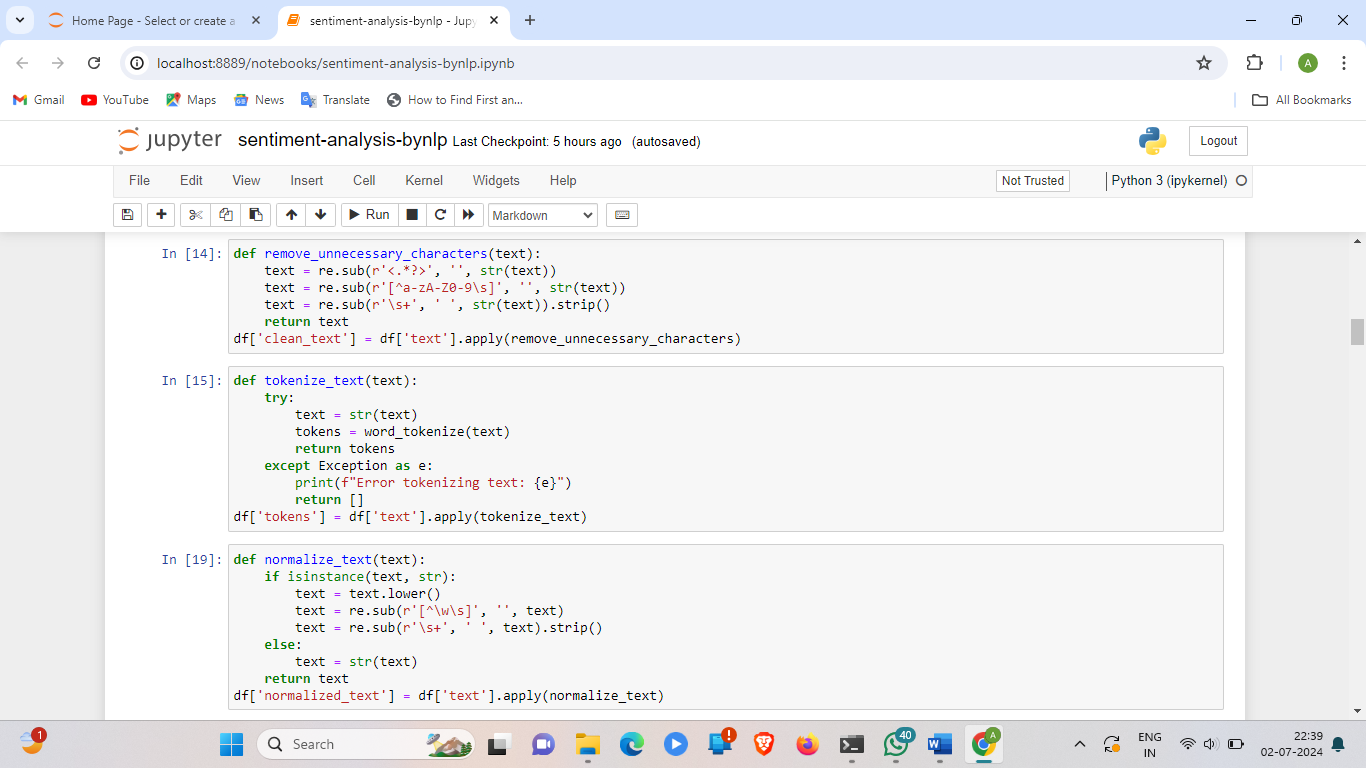
### **Tools and Libraries**

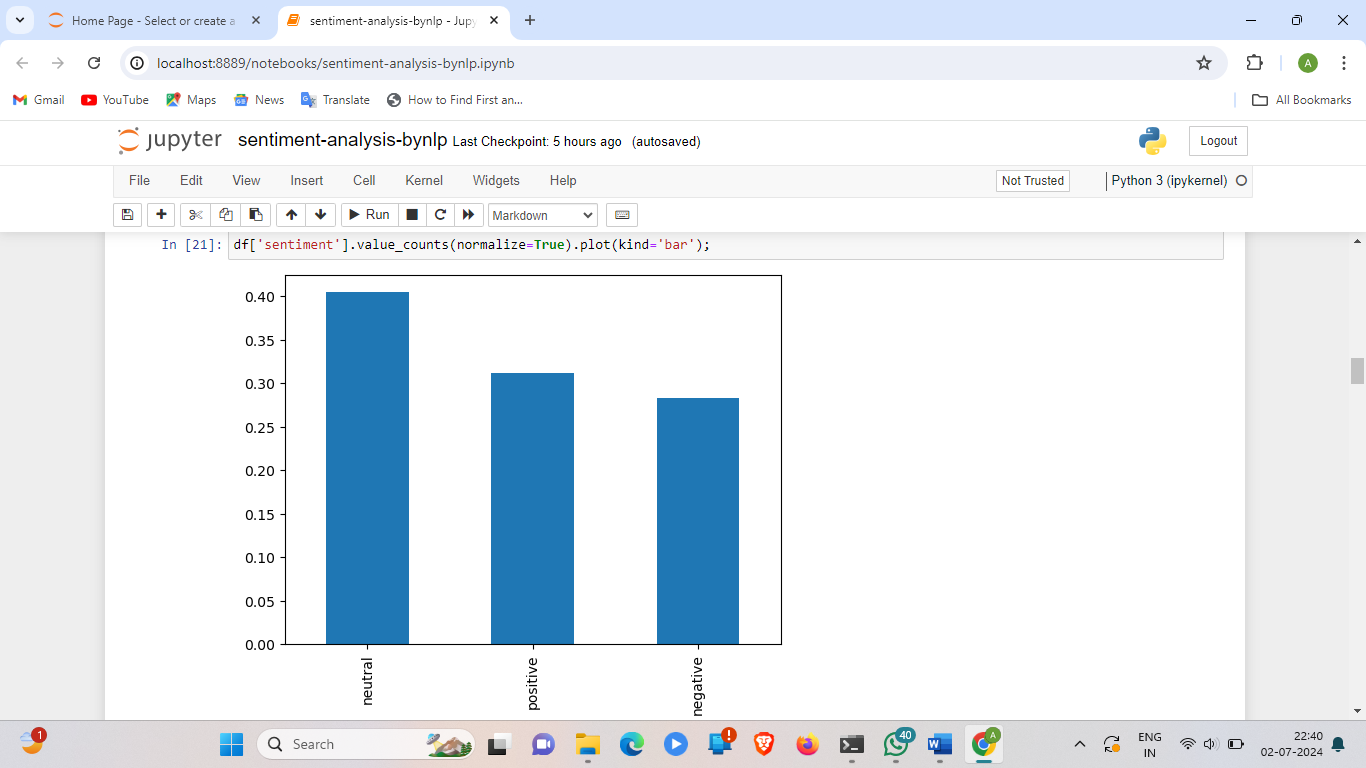
* **Environment**: The project is implemented in Python 3.8 using Jupyter Notebook.
* **Libraries**: Key libraries used include NLTK for natural language processing, scikit-learn for machine learning, and TensorFlow for deep learning.

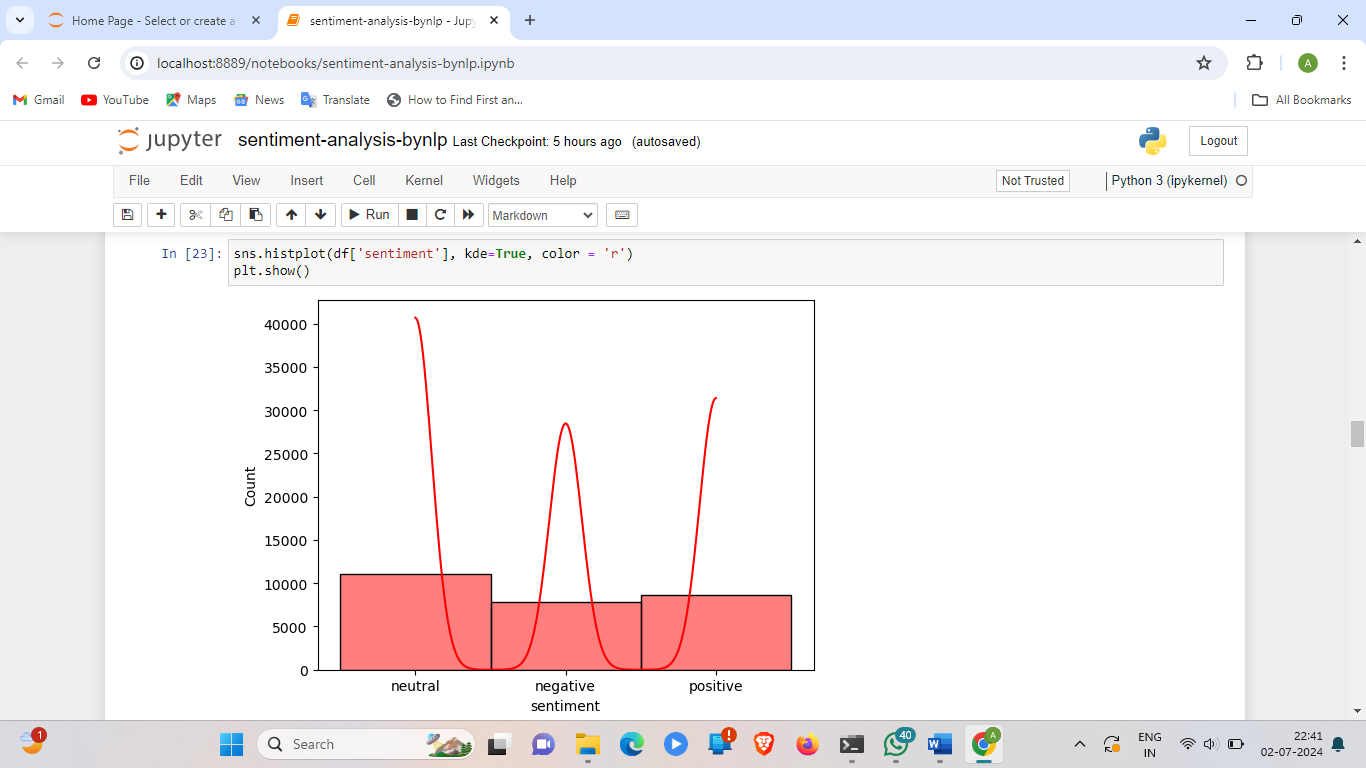
**Code Explanation**

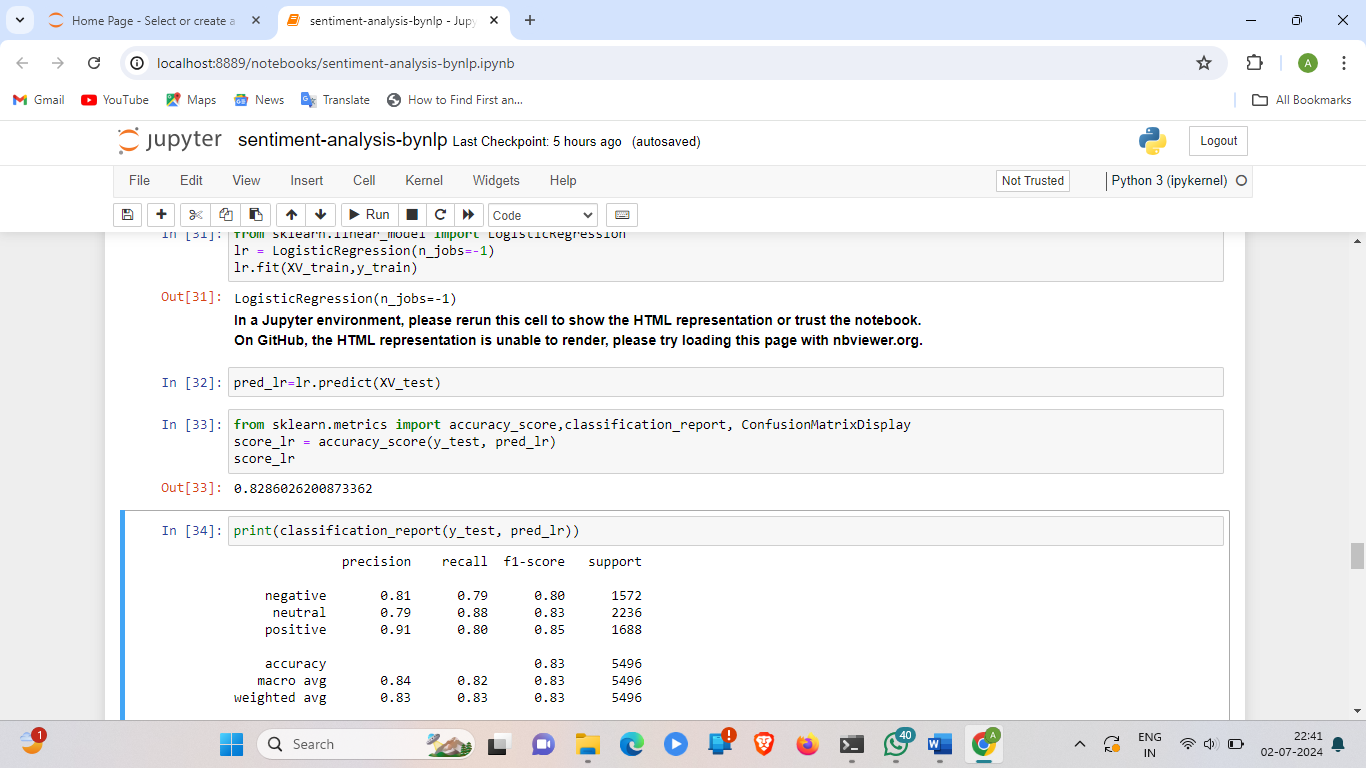
****

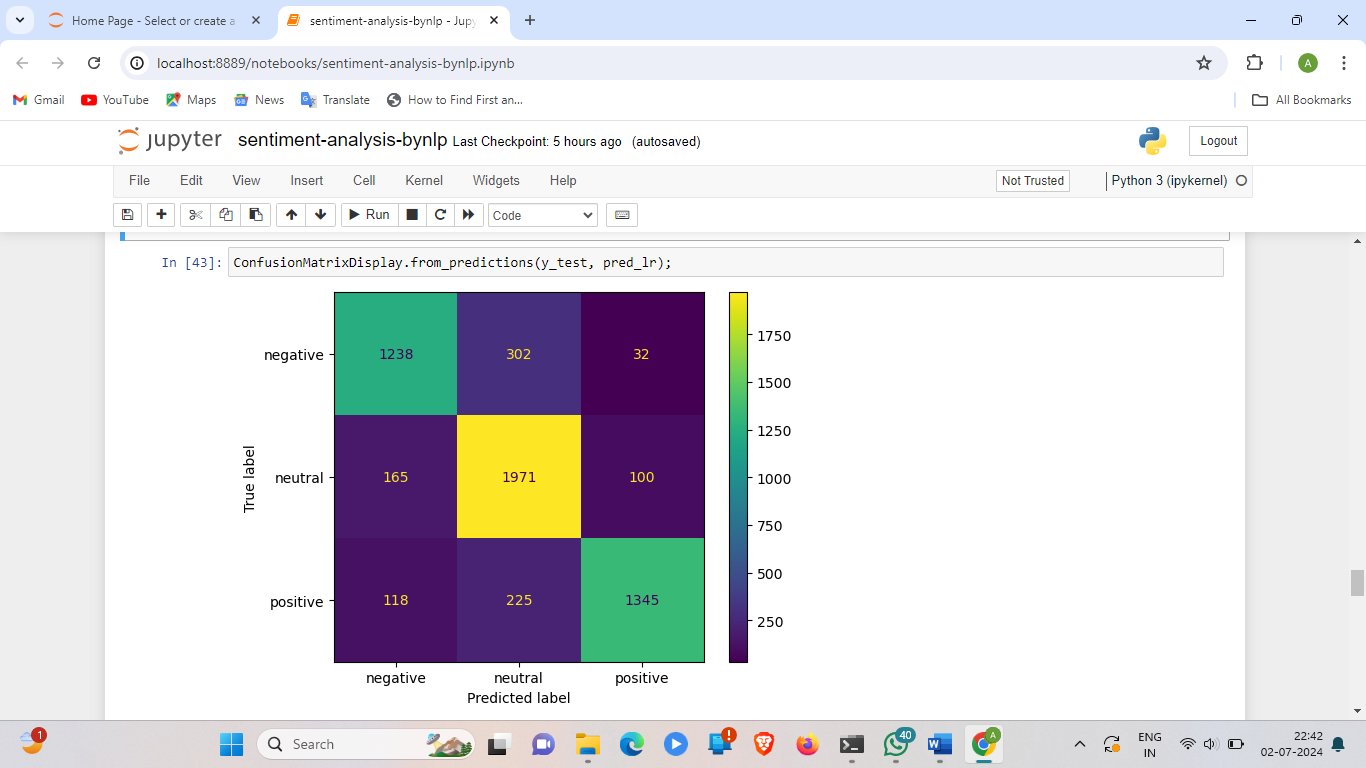
****

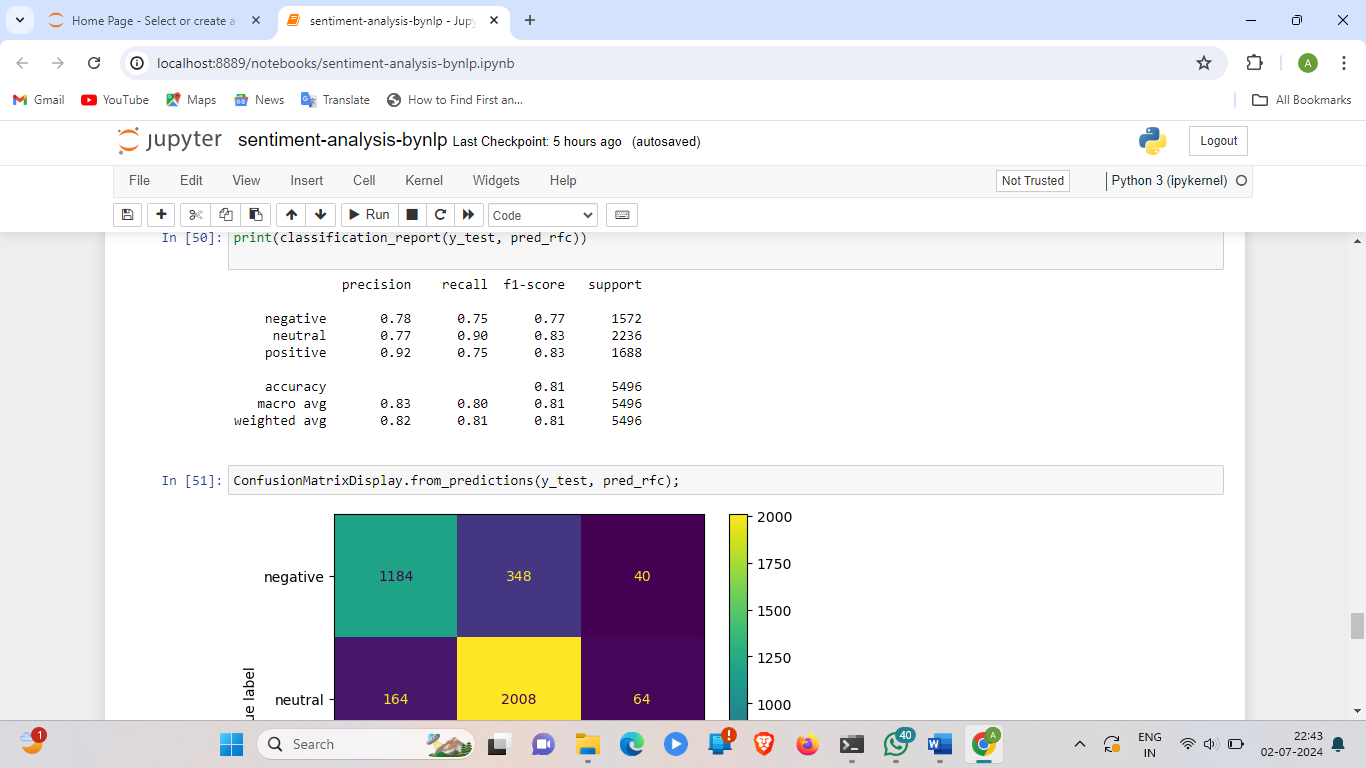






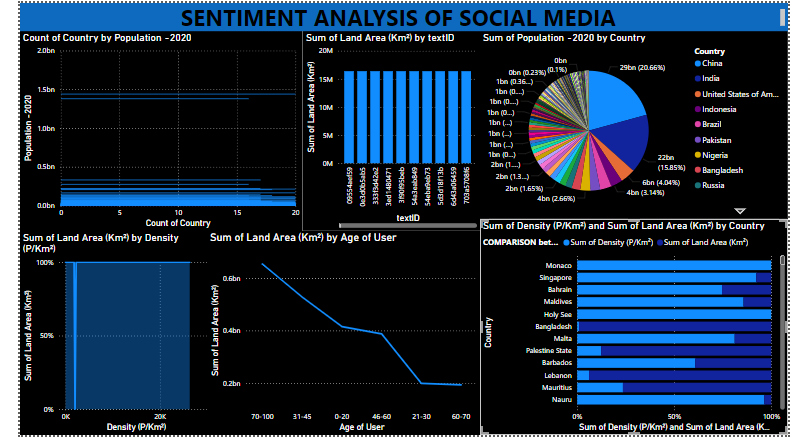






**CHAPTER 7**

**POWER BI DASHBOARD**



**CHAPTER 8**

**Results and Discussion**

### **Model Performance**

* **Metrics**: The performance of each model is evaluated using accuracy, precision, recall, and F1-score. These metrics provide a comprehensive view of the model's effectiveness in classifying sentiments.
* **Comparison**: The results of the Naive Bayes, SVM, and LSTM models are compared to determine which model performs best on the given dataset.

### **Analysis of Results**

* **Insights**: The analysis reveals that while traditional models like Naive Bayes and SVM perform well, the LSTM model outperforms them in capturing context and long-term dependencies in the text data.
* **Challenges**: Key challenges include handling imbalanced classes, processing large datasets efficiently, and dealing with noisy data. Strategies to mitigate these challenges are discussed.

## **Conclusion**

## **Summary**: This project successfully implemented and evaluated several sentiment analysis models using Python. The results demonstrate the effectiveness of deep learning models in handling sentiment classification tasks.

* **Implications**: The findings have significant implications for businesses and researchers looking to implement sentiment analysis in real-world applications, highlighting the importance of choosing the right model and preprocessing techniques.
* **Limitations**: The study is limited by the size and quality of the dataset, as well as the scope of models evaluated. Future work could explore more advanced models and larger, more diverse datasets.

## **Future Work**

* **Suggestions**: Future research could focus on improving model performance by incorporating more sophisticated feature extraction techniques, exploring different deep learning architectures, and expanding the dataset to include multiple languages and more diverse sources.

## **.**

## **References**

* **Citations**: Include all references cited in the report, formatted according to a standard citation style (e.g., APA, MLA). Examples of references include:
  + Bird, S., Klein, E., & Loper, E. (2009). Natural Language Processing with Python. O'Reilly Media, Inc.
  + Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. Foundations and Trends in Information Retrieval, 2(1-2), 1-135.