**Sentiment Analysis of Social Media Presence**

## A PROJECT REPORT

***Submitted by,***

**B Sai Kumar – 20211CBD0009**

**K Ganesh – 20211CBD0005**

**P Srinivas – 20211CBD0019**

**Y Sai Pallavi – 20211CBD0055**

### *Under the guidance of,*

**Dr.Medikonda Swapna-Asso.Prof**

***in partial fulfillment for the award of the degree of***

**BACHELOR OF TECHNOLOGY**

**IN**

**COMPUTER SCIENCE AND TECHNOLOGY**

**At**



**PRESIDENCY UNIVERSITY**

**BENGALURU**

**MAY 2025**

**PRESIDENCY UNIVERSITY**

**SCHOOL OF COMPUTER SCIENCE ENGINEERING**

**CERTIFICATE**

This is to certify that the Project report **“TITLE OF THE PROJECT”** being submitted by “STUDENTS NAMES” bearing roll number(s) “STUDENTS ROLL NUMBERS” in partial fulfillment of the requirement for the award of the degree of Bachelor of Technology in Computer Science and Engineering is a bonafide work carried out under my supervision.

|  |  |
| --- | --- |
| **Dr/Mr/Ms. <SUPERVISOR NAME>**  DESIGNATION  School of CSE&IS  Presidency University | **Dr. <HoD Name>**  DESIGNATION & HoD  School of CSE&IS  Presidency University |

|  |  |  |
| --- | --- | --- |
| **Dr. L. SHAKKEERA**  Associate Dean  School of CSE  Presidency University | **Dr. MYDHILI NAIR**  Associate Dean  School of CSE  Presidency University | **Dr. SAMEERUDDIN KHAN**  Pro-VC School of Engineering  Dean -School of CSE&IS  Presidency University |

**PRESIDENCY UNIVERSITY**

**SCHOOL OF COMPUTER SCIENCE ENGINEERING**

**DECLARATION**

We hereby declare that the work, which is being presented in the project report entitled **TITLE OF THE PROJECT** in partial fulfillment for the award of Degree of **Bachelor of Technology** in **Computer Science and Engineering**, is a record of our own investigations carried under the guidance of **SUPERVISOR NAME, DESIGNATION,** **School of Computer Science Engineering & Information Science, Presidency University, Bengaluru.**

We have not submitted the matter presented in this report anywhere for the award of any other Degree.

|  |  |
| --- | --- |
|  | **Name(s), Roll No(s) and Signature(s) of the Students** |

**ABSTRACT**

The rise of social media platforms has led to an unprecedented volume of user-generated content, making sentiment analysis a crucial tool for understanding public opinion, brand perception, and social trends. This paper explores various machine learning and natural language processing (NLP) techniques used in sentiment analysis, including lexicon-based methods, supervised and unsupervised learning models, and hybrid approaches. A comparative analysis of existing sentiment analysis models, their accuracy, and application areas is provided. The study also discusses challenges such as handling sarcasm, multilingual data, and contextual ambiguity. The findings highlight the importance of sentiment analysis in fields like politics, healthcare, business intelligence, and crisis management.

Sentiment analysis of social media presence has become a critical research area, driven by the rapid growth of user-generated content on platforms like Twitter, Facebook, and Instagram. This study explores various sentiment analysis techniques, including lexicon-based, machine learning, and hybrid models, to classify social media posts into different sentiment categories. Sentiment classification accuracy can be improved with the help of artificial intelligence and deep learning approaches like Naïve Bayes, SVM, and LSTM.

However, despite these advancements, challenges remain in areas such as sarcasm detection, multilingual text processing, and real-time sentiment analysis. The research has underlined the importance of sentiment analysis in real-world applications such as marketing, politics, finance, healthcare, and crisis management. Future developments in explainable AI, cross-lingual analysis, and advanced deep learning techniques will further enhance the capabilities of sentiment analysis, making it an indispensable tool for businesses, researchers, and policymakers.

**ACKNOWLEDGEMENT**

First of all, we indebted to the **GOD ALMIGHTY** for giving me an opportunity to excel in our efforts to complete this project on time.

We express our sincere thanks to our respected dean **Dr. Md. Sameeruddin Khan**, Pro-VC, School of Engineering and Dean, School of Computer Science Engineering & Information Science, Presidency University for getting us permission to undergo the project.

We express our heartfelt gratitude to our beloved Associate Deans **Dr. Shakkeera L and Dr. Mydhili Nair,** School of Computer Science Engineering & Information Science, Presidency University, and Dr. “NAME OF THE HOD”, Head of the Department, School of Computer Science Engineering & Information Science, Presidency University, for rendering timely help in completing this project successfully.

We are greatly indebted to our guide **Dr./Mr.Ms. Name, Designation** and Reviewer **Dr./Mr.Ms. Name, Designation**, School of Computer Science Engineering & Information Science, Presidency University for his/her inspirational guidance, and valuable suggestions and for providing us a chance to express our technical capabilities in every respect for the completion of the project work.

We would like to convey our gratitude and heartfelt thanks to the PIP2001 Capstone Project Coordinators **Dr. Sampath A K, Dr. Abdul Khadar and Mr. Md Zia Ur Rahman,** department Project Coordinators “NAME” and Git hub coordinator **Mr. Muthuraj.**

We thank our family and friends for the strong support and inspiration they have provided us in bringing out this project.

**Name of the Student (1)**

**Name of the Student (2)**

**Name of the Student (3)**

**Name of the Student (**

**LIST OF TABLES**

|  |  |  |  |
| --- | --- | --- | --- |
| **Sl. No.** | **Table Name** | **Table Caption** | **Page No.** |
| 1 | Table 1.1 | Software modules versus Reusable components | 5 |

**LIST OF FIGURES**

|  |  |  |  |
| --- | --- | --- | --- |
| **Sl. No.** | **Figure Name** | **Caption** | **Page No.** |
| 1 | Figure 1.1 | Software modules versus Reusable components | 5 |

**TABLE OF CONTENTS**

|  |  |  |
| --- | --- | --- |
| **CHAPTER NO.** | **TITLE** | **PAGE NO.** |
|  | **ABSTRACT ACKNOWLEDGMENT**  **…** | **i**  **ii**  **…** |
| **1.** | **INTRODUCTION** | 1 |
|  | 1.1 GENERAL | 1 |
|  | 1.2 . . . . . . . . . . . . | 2 |
|  | 1.2.1 General | 5 |
|  | 1.2.2.1 General | 8 |
|  | 1.2.2.2 . . . . . . . . | 10 |
|  | 1.2.2 | 12 |
|  | 1.3 . . .. . . . . . .. . . . . . . | 13 |
|  | 1.4 . . . . . . . . . . . . . . . . | 15 |
| **2.** | **LITERATURE REVIEW** | **16** |
|  | 2.1 GENERAL | 17 |
|  | 2.2 . . . . .. | 19 |
|  | 2.2. | 20 |

**CHAPTER-1**

**INTRODUCTION**

* 1. **Sub topic-1**
     1. **Sub topic-1 of Sub topic-1**

Include the content here with font family Times New Roman. Chapter Name font size should be 16. Sub topic font size should be 14. Further sub topics & Body text font size should be 12. Line spacing 1.5.

**1.2 Sub topic-2**

**1.3 Sub topic-3**

**CHAPTER-2**

**LITERATURE SURVEY**

To build up this model, we have read some earlier research papers.

Fuzzy Rule-based Unsupervised Sentiment Analysis from Social Media Posts, The paper [1]proposes a fuzzy logic-based unsupervised approach to sentiment classification, employing multiple lexicons and word sense disambiguation to classify posts as positive, negative, or neutral. While good with mixed datasets, its weaknesses include poor performance with short texts (tweets), no sarcasm detection, reliance on pre-defined lexicons, and poor support for multilingual data. For this, deep learning models like LSTM or BERT, hybrid lexicon-ML approaches, sarcasm detection approaches, and multilingual NLP approaches can be utilized to improve sentiment classification accuracy and contextual understanding.

Sentiment Analysis in Social Media and Its Application ,The paper [2] This is a systematic review of sentiment analysis methods, i.e., lexicon-based and opinion mining methods on Twitter data. It recognizes the extensive application of sentiment analysis in marketing, politics, and healthcare, but also recognizes the challenges of contextual ambiguity, detection of sarcasm, and the inability to adapt in real-time. To mitigate such challenges, deep learning models such as transformers (BERT, GPT), sentiment classification hybrid methods, and real-time streaming data processing can be used.

Artificial Intelligence for Social Media Safety and Security,

The paper[3] This article describes the application of AI to detect threats, disinformation, and hate speech on social media by machine learning and deep learning-based sentiment analysis. While AI enhances automated moderation and threat detection, challenges are bias in AI algorithms, privacy, and ethics in content filtering. For these, explainable AI (XAI), unbiased dataset selection, and ethical AI frameworks must be integrated to ensure fair and transparent sentiment analysis.

Impact of Social Media in Security and Crisis Management,

The paper[4] This paper explains how social media sentiment analysis can be used to aid crisis management by monitoring public emotions and reactions during crises. The research quotes the application of big data analytics and machine learning to derive social media insights but mentions challenges like the spread of misinformation, posting of fake news, and processing data in real-time. Solutions are fact-checking algorithms, real-time NLP models, and the application of geospatial analysis in combination with sentiment detection to improve crisis response plans.

A Systematic Review of Social Media-Based Sentiment Analysis, Emerging Trends and Challenges,The paper[5] This paper briefly discusses some of the techniques used in sentiment analysis, grouping them as lexicon-based, machine learning, and hybrid approaches and enumerating key issues in handling multilingual data, class imbalance, and real-time processing. It suggests that combining deep learning architectures with transfer learning (BERT, RoBERTa), data augmentation methods, and improved feature engineering can be used to enhance sentiment classification on diverse datasets.

A Review on Sentiment Analysis from Social Media Platforms,The paper [6] This paper briefly discusses some of the techniques used in sentiment analysis, grouping them as lexicon-based, machine learning, and hybrid approaches and enumerating key issues in handling multilingual data, class imbalance, and real-time processing. It suggests that combining deep learning architectures with transfer learning (BERT, RoBERTa), data augmentation methods, and improved feature engineering can be used to enhance sentiment classification on diverse datasets.

Beyond Positive or Negative: Qualitative Sentiment Analysis of Social Media Reactions to Unexpected Stressful Events,The paper [7] This work proposes a qualitative sentiment analysis with a contextual and affective interpretation rather than positive-negative classification alone. It proposes coping mechanism categories but is not real-time, automatic, or scalable. To these, hybrid qualitative-quantitative models, psychological NLP models, and automatic sentiment tagging systems can be employed to enhance sentiment comprehension.

A Model for Sentiment and Emotion Analysis of Unstructured Social Media Text,The paper [8] This book emphasizes machine learning and lexicon-based methods for sentiment and emotion extraction from social media. Naïve Bayes, SVM, and TF-IDF feature extraction are robust methods but are not context-sensitive, do not recognize sarcasm, and are not sensitive to complex sentence structures. Deep learning architectures such as CNNs, LSTMs, and attention-based transformers (BERT, GPT) can be used to increase context-sensitivity and sentiment accuracy.

Investigating Sentimental Relation Between Social Media Presence and Academic Success of Turkish Universities,

The paper[9] This study explores the connection between academic performance and social media sentiment using statistical sentiment analysis methods to analyze the reputation of universities. It lacks depth learning-based sentiment tracking, real-time fine-tuning, and contextual sentiment analysis. The following can be improved by incorporating AI-based sentiment prediction, multi-source data analysis, and longitudinal sentiment studies to provide more accurate academic insights.

Sentiment Analysis on Social Media, The paper [10] This work proposes a low-resource sentiment analysis approach based on basic NLP and machine learning techniques like Naïve Bayes and SVM for sentiment analysis on Twitter. While sufficient for basic polarity classification, it lacks state-of-the-art context awareness, sarcasm identification, and multilinguality. The integration of deep learning models, sentiment-aware embeddings like Word2Vec, FastText, and transformer-based sentiment models can be a game-changer in terms of performance and real-world applicability.

**CHAPTER-3**

**RESEARCH GAPS OF EXISTING METHODS**

Sentiment analysis of social media data has gained significant attention due to its applications in various domains, including business intelligence, crisis management, and public opinion analysis. While numerous techniques—ranging from lexicon-based approaches to advanced deep learning models—have been developed, several key research gaps remain, limiting their effectiveness.

**Sarcasm and Contextual Ambiguity**

One of the biggest challenges in sentiment analysis is detecting sarcasm and handling contextual ambiguity. Existing machine learning and lexicon-based approaches struggle to differentiate between literal and sarcastic statements, often leading to misclassification of sentiments.

**Multilingual and Code-Mixed Text Processing**

Most sentiment analysis models are trained on English datasets and do not generalize well to other languages, especially in code-mixed scenarios where multiple languages are used in a single text. The lack of large, annotated datasets for underrepresented languages further exacerbates this issue.

**Real-Time Sentiment Analysis**

While deep learning models like LSTM, BERT, and GPT perform well in offline settings, their computational complexity makes real-time processing of large-scale social media data challenging. Efficient real-time sentiment analysis models that balance speed and accuracy are still an open research problem.

**Bias and Ethical Concerns in AI Models**

Many sentiment analysis models exhibit biases due to the datasets they are trained on. These biases can lead to unfair or inaccurate sentiment classifications, particularly for sensitive topics such as political discourse or social issues. Explainable AI (XAI) and fairness-aware training methodologies need further development.

**Sentiment Analysis Beyond Polarity Classification**

Traditional sentiment analysis models primarily focus on classifying text as positive, negative, or neutral. However, emotions are more complex and nuanced. Existing methods do not sufficiently capture emotions such as fear, anger, or joy, which are crucial for a deeper understanding of user sentiment.

**Misinformation Detection and Crisis Management**

Sentiment analysis is increasingly used in crisis management to track public opinion and misinformation. However, current models lack mechanisms to filter out fake news, manipulated content, or bot-generated sentiments, which can skew the analysis results.

**CHAPTER-4**

**PROPOSED MOTHODOLOGY**

This research aims to develop an efficient sentiment analysis system for social media data using machine learning and natural language processing (NLP) techniques. The methodology consists of multiple phases, including data collection, preprocessing, model training, and deployment.

**1. Data Collection**

* **Data Sources**: Twitter (via Tweepy API), social media websites (via Scrapy, if applicable).
* **Dataset**: Pre-collected datasets from Kaggle will be used for training and validation.
* **Data Format**: Social media text, including tweets and posts, with sentiment labels (positive, negative, neutral).

**2. Data Preprocessing**

Before applying machine learning models, raw text data undergoes preprocessing to improve accuracy:

* **Tokenization**: Splitting text into individual words.
* **Stopword Removal**: Eliminating common words (e.g., "is," "the") using NLTK.
* **Lemmatization/Stemming**: Converting words to their base forms.
* **Handling Emojis and Special Characters**: Replacing emojis with their sentiment labels and removing unnecessary symbols.
* **Removing URLs and Mentions**: Eliminating links and user mentions to retain only relevant content.

**3. Feature Extraction**

* **TF-IDF (Term Frequency-Inverse Document Frequency)**: To quantify text importance.
* **Word Embeddings (Optional)**: Pre-trained embeddings like Word2Vec or FastText can be used for improved representation.

**4. Sentiment Classification Model**

Two machine learning models will be implemented and compared:

* **Random Forest Classification**: A robust ensemble learning method that builds multiple decision trees for sentiment classification.
* **Logistic Regression**: A statistical method effective for binary classification tasks like positive vs. negative sentiment.

**5. Model Training and Evaluation**

* **Training Data**: The collected dataset will be split into training and testing sets (e.g., 80% training, 20% testing).
* **Evaluation Metrics**: Accuracy, Precision, Recall, and F1-score will be used to measure model performance.

**6. Backend and Frontend Development**

* **Backend**: Python-based Flask/Django API to serve predictions.
* **Frontend**: A minimal web interface for users to input text and receive sentiment classification results.

**7. Deployment and Real-Time Sentiment Analysis**

* **Integration with Twitter API**: The trained model will be deployed to analyze live tweets.
* **Real-Time Processing**: The system will continuously fetch, analyze, and display sentiment trends.
* **Cloud Deployment (Optional)**: The final application may be hosted on cloud platforms like AWS or Heroku.

By implementing this methodology, the project will provide an efficient and scalable sentiment analysis system that can be used for social media analytics, opinion mining, and trend prediction.

**CHAPTER-5**

**OBJECTIVES**

1. Develop a Robust Sentiment Analysis Model

* Implement machine learning-based sentiment classification using Random Forest and Logistic Regression.
* Utilize NLP techniques for feature extraction and text preprocessing to improve classification accuracy.
* Optimize model performance using evaluation metrics like accuracy, precision, recall, and F1-score.

2. Handle Language and Contextual Challenges

* Address challenges related to sarcasm, contextual ambiguity, and slang commonly found in social media text.
* Implement preprocessing techniques such as stopword removal, lemmatization, and emoji handling.
* Explore word embeddings or hybrid models for better contextual understanding.

3. Perform Real-Time Sentiment Tracking

* Integrate with the Twitter API to fetch and analyze live tweets.
* Develop a real-time sentiment monitoring system that processes and classifies social media data dynamically.
* Ensure efficient handling of high-volume streaming data with optimized computational techniques.

4. Provide Visualization and Reporting Tools

* Design a web-based interface to display sentiment analysis results interactively.
* Implement graphical representations such as sentiment trend graphs and word clouds.
* Enable users to track sentiment changes over time for data-driven decision-making.
* By achieving these objectives, the proposed system will provide an accurate, scalable, and real-time solution for sentiment analysis on social media platforms.

**CHAPTER-6**

**SYSTEM DESIGN & IMPLEMENTATION**

1. System Design

The system is designed as a web-based sentiment analysis platform that processes social media data to determine sentiment polarity (positive, negative, or neutral). The architecture follows a modular approach, ensuring scalability and efficiency.

1.1 System Components

1. User Interface (Frontend)
   * A web-based interface where users can input text or analyze social media sentiment.
   * Displays sentiment analysis results with visualizations (charts, word clouds).
2. Backend (Flask/Django API)
   * Handles user requests and processes sentiment analysis.
   * Connects with the machine learning model and data sources.
3. Data Collection Module
   * Tweepy: Fetches tweets for real-time sentiment tracking.
   * Scrapy: (Optional) Scrapes social media content from websites.
   * Kaggle Datasets: Used for model training and evaluation.
4. Preprocessing & Feature Extraction
   * Uses NLTK for text cleaning, tokenization, and stopword removal.
   * TF-IDF and Word Embeddings help in converting text into numerical features.
5. Machine Learning Models
   * Random Forest Classification and Logistic Regression for sentiment classification.
   * Evaluated using accuracy, precision, recall, and F1-score.
6. Database (Optional)
   * Stores processed sentiment data for historical analysis.
7. Visualization & Reporting
   * Provides sentiment trend analysis using graphs and reports.

2. Implementation

The implementation follows an iterative development approach, ensuring continuous improvements.

2.1 Technology Stack

* Programming Language: Python (for ML and backend processing)
* Frameworks & Libraries:
  + NLP & ML: NLTK, Scikit-learn, TensorFlow/Keras (if needed)
  + Data Handling: Pandas, NumPy
  + Web Framework: Flask/Django for backend API
  + Frontend: HTML, CSS, JavaScript (React.js or simple UI)
  + Data Collection: Tweepy (Twitter API), Scrapy (Web Scraping)

2.2 Development Phases

1. Data Collection & Preprocessing
   * Fetch social media data using APIs (Twitter, Scrapy).
   * Clean text (removing stopwords, URLs, special characters).
   * Convert text into feature vectors using TF-IDF/Word Embeddings.
2. Model Training & Evaluation
   * Train Random Forest and Logistic Regression on Kaggle datasets.
   * Evaluate models and optimize hyperparameters.
3. API Development & Integration
   * Develop Flask/Django API to serve model predictions.
   * Integrate frontend for user interactions.
4. Visualization & Deployment
   * Create real-time dashboards with sentiment trends.
   * Deploy the system on AWS/Heroku for accessibility.

3. Expected Outcomes

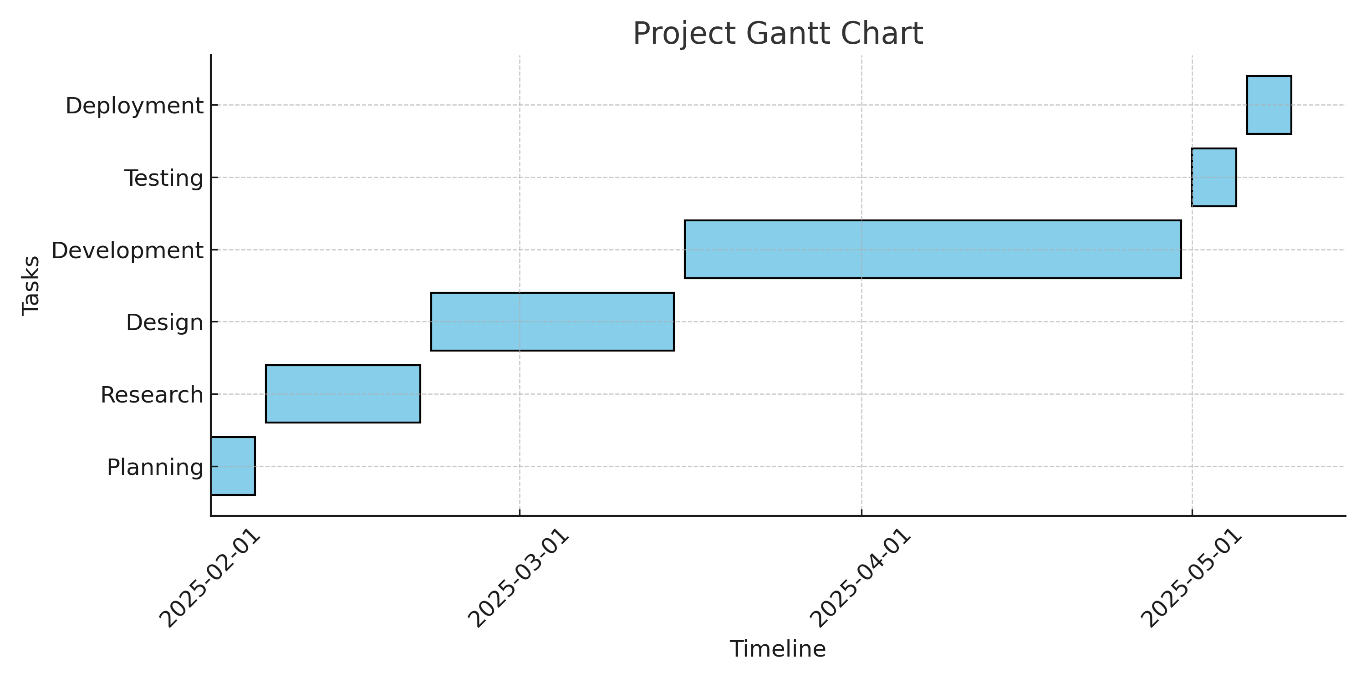
* A real-time sentiment analysis system that can process social media data dynamically.
* Improved accuracy using hybrid NLP and machine learning models.
* User-friendly frontend with sentiment visualization.
* Potential future integration with explainable AI (XAI) and deep learning models.

This section provides a detailed breakdown of system design and implementation. Let me know if you need any modifications or additions!

**CHAPTER-7**

**TIMELINE FOR EXECUTION OF PROJECT**

**(GANTT CHART)**

****

**CHAPTER-8**

**OUTCOMES**

**CHAPTER-9**

**RESULTS AND DISCUSSIONS**

**CHAPTER-10**

**CONCLUSION**

**REFERENCES**

**APPENDIX-A**

**PSUEDOCODE**

**APPENDIX-B**

**SCREENSHOTS**

**APPENDIX-C**

**ENCLOSURES**

**1. Journal publication/Conference Paper Presented Certificates of all students.**

**2. Include certificate(s) of any Achievement/Award won in any project-related event.**

**3. Similarity Index / Plagiarism Check report clearly showing the Percentage (%). No need for a page-wise explanation.**

**4.** **Details of mapping the project with the Sustainable Development Goals (SDGs).**