**SYNOPSIS**

## Report on

**SENTIMENT ANALYSIS On GOVERNMENT SCHEMES**

**by**

Tanishka Gupta 2400290140216

Shipra Upadhyay 2400290140194

Sonam Dobriyal 2400290140208

Tanisha Jain 2400290140215

## Session:2025-2026 (III Semester)

Under the supervision of

## Dr. Amit Kumar Gupta (Professor)

**KIET Group of Institutions, Delhi-NCR, Ghaziabad**

****

**DEPARTMENT OF COMPUTER APPLICATIONS KIET GROUP OF INSTITUTIONS, DELHI-NCR,**

**GHAZIABAD-201206**

(2024 - 2025)

# ABSTRACT

This project focuses on the development of a Sentiment Analysis System that analyzes and classifies user opinions from textual data. With the rise of social media platforms, customer feedback forums, and online reviews, sentiment analysis has become an essential tool for understanding public opinion and improving decision-making. The project applies Natural Language Processing (NLP) and Machine Learning (ML) techniques to extract insights from text, enabling automatic classification of sentiments into categories such as positive, negative, or neutral.

The study involves data collection, preprocessing, model training, and evaluation using suitable algorithms and tools. By leveraging open-source libraries such as requests, Scikit-learn, and TensorFlow, the system is designed to achieve efficiency and accuracy. The outcome of this project is a robust sentiment analysis framework that can be applied to domains like business intelligence, social media monitoring, product feedback analysis, and customer service enhancement. This work contributes to bridging the gap between raw data and meaningful insights.

**Key Highlights:**

* Focuses on automatic sentiment classification of text data.
* Uses NLP and ML techniques for opinion mining.
* Data preprocessing, model training, and evaluation form the methodology.
* Tools: TextBlob, Scikit-learn, TensorFlow, Keras.
* Applications: business, social media, product reviews, and customer feedback.
* Provides accurate and efficient sentiment analysis system.

# TABLE OF CONTENTS

Page Number

1. Introduction 4 - 5
2. Literature Review 6 - 7
3. Research Objective 8 - 9
4. Hardware and Software Requirements 10
5. Research Methodology 11
6. Research Outcome 12
7. Proposed Time Duration 13

[References/ Bibliography 1](#_TOC_250000)4

# INTRODUCTION

In the modern digital landscape, social media platforms have emerged as a primary source of communication and expression for millions of citizens. Among these, Twitter has become a popular platform where people voice their opinions, concerns, and feedback about government schemes and policies. However, the diversity of languages used in India poses a significant challenge for analyzing this feedback effectively. News are often written in Hindi, English, or a mix of regional dialects, making sentiment analysis a complex task. This project addresses the challenge by leveraging **machine translation tools** to standardize multilingual text and applying advanced **Natural Language Processing (NLP) models like TEXTBLOB** to identify sentiments accurately. By analyzing the polarity of opinions—positive, negative, or neutral—the project contributes toward understanding the effectiveness of government initiatives, thereby aiding in better policy formulation, improved governance, and enhanced citizen engagement.

Key Points:

1. **Role of Social Media in Governance**
   * Social media provides a real-time platform where citizens freely express their views about government schemes.
   * Headlines collected from news reflects genuine sentiments, which are valuable for policymakers.

**2. Multilingual Data Challenge**

* News often appear in multiple languages (Hindi, English, and regional dialects).
* Direct analysis becomes difficult without a translation mechanism.

**3. Importance of Sentiment Analysis**

* Classifies feedback into positive, negative, or neutral categories.
* Helps measure overall public satisfaction or dissatisfaction.

**4. Use of Translation Tools**

* Translation bridges the language gap by converting regional text into English.
* Ensures consistency in processing data before feeding it into the model.

**5. Adoption of TEXTBLOB Model**

* TEXTBLOB is a library which detects polarity and performs Sentiment Analysis.
* Provides higher accuracy in understanding contextual meaning compared to traditional models.

**6. Project Purpose**

* To integrate translation + sentiment analysis for evaluating news on government schemes.
* Generates structured insights from unstructured text data.

**7. Impact on Governance**

* Real-time public feedback allows policymakers to assess the success of schemes.
* Enhances transparency, citizen trust, and promotes evidence-based decision-making.

# LITERATURE REVIEW

Sentiment analysis of social media has been a growing research area, particularly in the domain of public opinion mining. Many researchers have explored techniques to classify news and posts to understand user emotions, with applications in politics, marketing, and governance. However, one of the persistent challenges has been the multilingual nature of social media data, especially in countries like India. Existing studies show that while traditional **machine learning models like Naïve Bayes, Support Vector Machines (SVM), and Random Forests** were effective in English text analysis, their performance declined in multilingual or code-mixed scenarios. To address this, scholars introduced translation tools and deep learning approaches. With the advancement of NLP, models like **Word2Vec, LSTMs, and more recently TEXTBLOB and its variants** have proven to be highly efficient in capturing contextual meaning. Previous work demonstrates that multilingual translation combined with transformer-based architectures significantly improves sentiment classification accuracy. Thus, this project builds upon prior research by applying machine translation and TEXTBLOB-based sentiment analysis to evaluate public opinion on government schemes through Twitter data.

**Key Points:**

**1. Traditional Approaches**

* Earlier methods used **Naïve Bayes, SVM, and Logistic Regression** for sentiment analysis.
* Suitable for English-only datasets but limited in handling mixed-language inputs.

**2. Challenges in Multilingual Sentiment Analysis**

* News often contain **code-mixing** (e.g., Hindi-English "Hinglish").
* Lack of uniform language datasets reduces model accuracy.

**3. Use of Translation Tools in Research**

* Prior studies used **Google Translate, Indic Transliteration, and other APIs** to normalize text.
* Translation improves consistency before classification.

**4. Deep Learning Innovations**

* Neural networks like **CNNs and LSTMs** improved sentiment detection by learning contextual relations.
* However, they struggled with long sentences and mixed-lingual content.

**5. Transformer-based Models**

* TEXTBLOB and its variants (mTEXTBLOB, DistilTEXTBLOB, RoTEXTBLOBa) brought a major breakthrough.
* They provide **context-aware embeddings** for accurate sentiment classification.

**6. Sentiment Analysis in Governance**

* Research indicates that public opinion mining helps governments **measure citizen satisfaction**.
* Case studies in healthcare, education, and transportation show improved decision-making using sentiment data.

**7. Gap in Existing Research**

* Few studies combine machine translation + TEXTBLOB models for Indian multilingual data.
* This project fills that gap by specifically analyzing news on government schemes.

# RESEARCH OBJECTIVE

The primary objective of this project is to analyze public sentiment on social media platforms, specifically Twitter, regarding government schemes. Social media serves as a dynamic space where citizens share their thoughts, feedback, and criticisms. By studying these opinions, governments and policymakers can gain insights into the effectiveness of their initiatives and identify areas for improvement. This project focuses on bridging the gap between multilingual communication and sentiment analysis by leveraging machine translation tools along with advanced NLP models like TEXTBLOB. The research aims to achieve accurate sentiment classification in English as well as regional languages that are translated into English for uniform processing.

Furthermore, the project is designed not only to test the accuracy of sentiment analysis models but also to provide a **framework for future governance-related studies**. By evaluating the positive, negative, or neutral sentiments of citizens, the system can offer policymakers actionable intelligence. This analysis can lead to better decision-making, improved public engagement, and transparent governance. Ultimately, the project contributes toward building **data-driven mechanism** for monitoring and enhancing government schemes.

Key Objectives:

**1. Sentiment Extraction**

* Collect and preprocess news related to government schemes.
* Identify emotional tone (positive, negative, neutral) from multilingual data.

**2. Multilingual Data Handling**

* Translate regional news into English using reliable APIs.
* Overcome challenges of **code-mixed and transliterated text.**

**3. Model Implementation**

* Apply machine learning and transformer-based models (SVM, LSTM, TEXTBLOB).
* Compare performance to determine the most effective model.

**4. Accuracy Evaluation**

* Use metrics such as **precision, recall, F1-score, and accuracy** to measure performance.
* Ensure robustness in handling noisy social media text.

**5. Visualization of Results**

* Represent outcomes using graphs, charts, and cluster visualizations.
* Provide policymakers with clear and interpretable insights.

**6. Impact on Governance**

* Assist government bodies in understanding public perception.
* Enable data-driven policy revisions and better communication strategies.

**7. Research Contribution**

* Develop a generalized framework for multilingual sentiment analysis.
* Contribute to the growing field of AI for governance and public opinion mining.

# HARDWARE AND SOFTWARE REQUIREMENTS

### Hardware Requirements:

**1. Processor (CPU):** Minimum Intel i5 or AMD equivalent (for basic tasks); Intel i7/i9 or AMD Ryzen (for advanced tasks).

**2. Graphics Processing Unit (GPU):** NVIDIA GPU with CUDA support (e.g., Tesla T4, RTX 3060+) for model training.

**3. RAM:** At least 8 GB (basic) to 16/32 GB (recommended for TEXTBLOB-based models).

**4. Storage:** 512 GB SSD or higher for dataset storage and faster processing.

**5. Cloud Resources:** Google Colab, AWS, or Azure for high-performance training.

**6. Internet Connectivity:** Stable and high-speed for real-time news collection and API integration.

### Software Requirements:

**1. Operating System:** Windows 10/11, Linux (Ubuntu preferred), or macOS.

**2. Programming Language:** Python 3.8+ (primary language for AI/ML development).

**3. NLP Libraries:** NLTK, SpaCy for preprocessing tasks.

**4. Machine Learning Frameworks:** Scikit-learn for classical ML models.

**5. Deep Learning Frameworks:** TensorFlow, Keras, PyTorch for advanced model training.

**6. Transformer Libraries:** Hugging Face Transformers for TEXTBLOB implementation.

**7. Translation Tools:** Google Translate API or similar multilingual tools.

**8. Database Support:** MySQL or MongoDB for structured storage.

**9. Visualization Tools:** Matplotlib, Seaborn, Plotly for graphical representation.

**10. Version Control:** Git/GitHub for project collaboration and tracking.

# RESEARCH METHODOLOGY

The research methodology for this project follows a **structured and step-by-step approach** to ensure accuracy and reliability in sentiment analysis. It begins with **data collection**, where Twitter API is used to gather news related to various government schemes. Since social media data is often noisy, the next step is **data preprocessing,** which includes cleaning unwanted characters, removing stop words, handling hashtags, tokenizing text, and translating multilingual news into English for consistency.

Once the data is clean, the next stage involves **feature extraction,** where techniques like TF-IDF, Word2Vec, and advanced embeddings such as TEXTBLOB are used to capture semantic meaning from text. These features are then used to train **machine learning and deep learning models** such as SVM, LSTM, and TEXTBLOB-based classifiers.

The trained models categorize sentiments into **Positive, Negative, or Neutral.** To ensure reliability, performance is evaluated using metrics such as accuracy, precision, recall, and F1-score. Finally, results are **visualized through graphs and charts**, making it easier to interpret and derive meaningful insights for decision-making.

1. **Data Collection:** News extracted using Twitter API related to government schemes.
2. **Data Preprocessing:** Cleaning, tokenization, stop-word removal, stemming/lemmatization, translation.
3. **Feature Extraction:** Use TF-IDF, Word2Vec, and TEXTBLOB embeddings.
4. **Model Training:** Implement ML/DL models (SVM, LSTM, TEXTBLOB).
5. **Sentiment Classification:** Classify into Positive, Negative, Neutral.
6. **Evaluation:** Accuracy, Precision, Recall, F1-score.
7. **Visualization & Reporting:** Graphs, charts, and dashboards for results.

# RESEARCH OUTCOME

### The project on sentiment analysis of government schemes using Twitter data is expected to produce meaningful and practical outcomes. The primary result will be the development of a working sentiment analysis model capable of classifying public opinion into positive, negative, or neutral sentiments with a high degree of accuracy. This model will provide a real-time monitoring system for policymakers, enabling them to evaluate how citizens perceive different government initiatives. Additionally, the project will create a visual representation of sentiment trends through graphs and dashboards, making insights more understandable for non-technical stakeholders.

### Another significant outcome is the identification of key issues and areas of dissatisfaction expressed by the public, helping the government modify or improve its policies. Academically, the project will contribute to the growing field of social media analytics and natural language processing (NLP), serving as a research reference for future scholars. Furthermore, the work has the potential for publication in reputed journals or conferences, adding value to both the academic and professional community.

* **Working Sentiment Classifier:** Categorizes news as positive, negative, or neutral.
* **Real-time Monitoring System:** Tracks live feedback on government schemes.
* **Visualization Dashboards:** Graphs, charts, and trend analysis for easy interpretation.
* **Policy Improvement Insights:** Identifies problem areas in schemes through negative sentiments.
* **Academic Contribution:** Adds value to research in NLP and sentiment analysis.
* **Publication Potential:** Scope for research papers, reports, or conference presentations

# PROPOSED TIME DURATION

|  |  |
| --- | --- |
| **PHASE** | **TIME DURATION** |
| Requirement Analysis | 2 Weeks |
| Design Phase | 4 Weeks |
| Development Phase | 7 Weeks |
| Testing phase | 3 Weeks |
| Deployment Phase | 2 Weeks |
| Total Duration | 18 Weeks |

## REFERENCES/ Bibliography

* Bing Liu, Sentiment Analysis and Opinion Mining, Morgan & Claypool Publishers, 2012.
* Pang, B., & Lee, L., Opinion Mining and Sentiment Analysis, Foundations and Trends in Information Retrieval, 2008.
* Cambria, E., Schuller, B., Xia, Y., & Havasi, C. (2013). New Avenues in Opinion Mining and Sentiment Analysis. IEEE Intelligent Systems.
* Twitter Developer Documentation: (<https://developer.twitter.com> )
* Scikit-learn Documentation: ( <https://scikit-learn.org> )
* TensorFlow Documentation: (<https://www.tensorflow.org> )
* Kaggle Datasets: <https://www.kaggle.com>
* GitHub Repositories for Sentiment Analysis Implementations.