Sentiment Analysis on Social Media Data

Govind Gupta, Harsh Sharma, Kanchan Sagar, Kartikey Raghuvanshi

ABSTRACT

In today's digital landscape, social media and online platforms serve as primary channels for expressing opinions and emotions. Understanding these sentiments is crucial for businesses, policymakers, and researchers to analyze public perception and make data-driven decisions. This project focuses on developing an advanced **Sentiment Analysis** system using **Natural Language Processing (NLP)** and **machine learning techniques** to classify sentiments in textual data.

The system leverages **both English and Hindi datasets**, incorporating diverse sources such as **tweets, product reviews, and social media comments**. By applying **preprocessing techniques**—including tokenization, stop word removal, stemming, and lemmatization—the model ensures high-quality text input. Feature extraction methods like **TF-IDF (Term Frequency-Inverse Document Frequency)** and **word embeddings (Word2Vec, GloVe)** further enhance data representation.

To achieve accurate sentiment classification, various **machine learning algorithms** such as **Naïve Bayes, Logistic Regression, and Support Vector Machines (SVM)** are implemented and evaluated. The performance of these models is measured using **accuracy, precision, recall, and F1-score** to ensure reliable predictions. Additionally, deep learning techniques using **Recurrent Neural Networks (RNN) and Long Short-Term Memory (LSTM)** networks are explored for enhanced sentiment understanding.

The project also introduces sentiment analysis for **sarcasm detection in Hindi tweets**, addressing challenges in multilingual NLP. The use of **real-time data visualization** through word clouds and interactive dashboards provides meaningful insights into sentiment trends.

By integrating **machine learning and NLP techniques**, this sentiment analysis system offers a powerful tool for businesses and researchers to monitor public sentiment, improve customer engagement, and drive informed decision-making. The solution provides a scalable, efficient, and multilingual approach to sentiment analysis, advancing the capabilities of natural language understanding in the digital era.

ACKNOWLEDGEMENTS

Success in life is never attained single-handedly. My deepest gratitude goes to my project supervisor, **Mr. Rabi Narayan Panda** for his/ her guidance, help, and encouragement throughout my project work. Their enlightening ideas, comments, and suggestions.

Words are not enough to express my gratitude to Dr. Akash Rajak, Professor and Head, Department of Computer Applications, for his insightful comments and administrative help on various occasions.

Fortunately, I have many understanding friends, who have helped me a lot on many critical conditions.

Finally, my sincere thanks go to my family members and all those who have directly and indirectly provided me with moral support and other kind of help. Without their support, completion of this work would not have been possible in time. They keep my life filled with enjoyment and happiness.

Govind Gupta

Harsh Sharma

Kanchan Sagar

Kartikey Raghuvanshi

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES