
TWITTER SENTIMENT ANALYSIS USING MULTIPLE LANGUAGES

AIML/CCV/DA/CSF/IIOT Group No.:

Project Group Members:

1. Aastha Dubey (Sec-M 201550001)
2. Nitya Agarwal (Sec- M 201550095)
3. Anushka Saxena (Sec-M 201550034)
4. Parth Goyal (Sec-N 201550096)

Project Supervisor: Mr. Rahul Pradhan, Assistant Professor /Associate Professor

About the Project:

In the age of digital communication and social media, Twitter stands as a global platform that encapsulates diverse voices and opinions. While it fosters meaningful conversations and connections, it also serves as a breeding ground for offensive language and harmful content. Understanding the prevalence and sentiment associated with foul language in tweets is of paramount importance in shaping a healthier and more respectful online environment. This major project embarks on an ambitious journey to analyse and interpret Twitter data containing foul language in a multilingual context. By examining tweets in languages including English, Hindi, Bengali, and others, we aim to unravel the intricate tapestry of sentiments woven into these expressions. The project strives to delve deep into the dynamics of online discourse, offering valuable insights into the emotional undercurrents that accompany offensive language across linguistic boundaries.

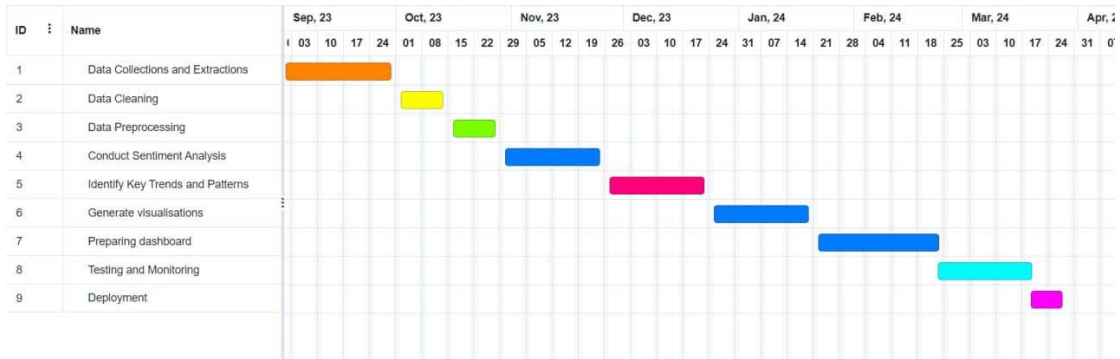
Motivation:

The motivation behind this project stems from the critical need to foster a respectful and inclusive online environment. As the internet continues to serve as a primary medium for communication and information dissemination, it is imperative that platforms like Twitter actively address the challenges posed by offensive language. By gaining a deep insight into the nature of foul language and its sentiment in multilingual tweets, we aim to contribute to the creation of safer and more enjoyable online spaces for users of diverse linguistic backgrounds.

Project Planning:

The project aims to develop a multilingual sentiment analysis tool for Twitter data, encompassing languages such as English, Spanish, French, and German. The plan involves a phased approach, beginning with the collection and preprocessing of diverse language data, followed by the training of natural language processing (NLP) models tailored for each language. The subsequent development of a user-friendly interface and thorough testing across languages will ensure the tool's accuracy and usability. Potential risks include language-specific nuances in sentiment expression and limited labelled data for certain languages. Success will be measured by achieving an accuracy above 85% for sentiment classification in each language, coupled with positive user feedback on the tool's

performance. The project timeline spans eight weeks, with ongoing adaptability to address language-specific challenges and ensure optimal outcomes.



Tools required:

➤ Hardware Requirements:

- Laptops / Computers with Minimum 8 GB RAM

➤ Software Requirements:

- Windows Operating System
- Google Colab
- Python 3.10.4
- Jupyter Lab
- Jupyter Notebook

Signature of Project Supervisor: _____