

# SENTIMENT ANALYSIS REPORT

## EXECUTIVE SUMMARY

This report provides an overview of the sentiment analysis project conducted on the social media platform using a Random Forest classifier. The goal of the project was to analyze user-generated content and classify sentiments into categories such as positive, neutral, offensive, or hate speech. The Random Forest classifier demonstrated high accuracy in achieving this objective.

## PROJECT OVERVIEW

### Business Objective:

The primary objective of the project was to enhance user experience and moderating content on the social media platform by implementing a sentiment analysis system. This system aimed to identify and flag content containing hate speech, offensive language, or any form of harmful communication.

### Methodology:

#### Data Collection:

- A diverse dataset containing user-generated content, including posts and comments, was collected from the social media platform.

#### Data Preprocessing:

- Textual data underwent preprocessing, including tokenization, cleaning, and TF-IDF vectorization for feature extraction.

Model Training:

- A Random Forest classifier and naïve bayes were two machine learning algorithms trained using a labeled dataset, encompassing various sentiments.

Evaluation:

- The model's performance was evaluated using a separate test dataset, measuring accuracy, precision, recall, and F1 score.

Model Performance:

Random Forest Classifier:

	precision	recall	f1-score	support
0	0.48	0.16	0.23	290
1	0.93	0.96	0.94	3832
2	0.83	0.90	0.86	835
accuracy			0.90	4957
macro avg	0.75	0.67	0.68	4957
weighted avg	0.89	0.90	0.89	4957

Random Forest, Accuracy Score: 0.9017550938067379

The Random Forest classifier demonstrated robust performance across all evaluated metrics, indicating its effectiveness in accurately classifying sentiments in user-generated content.

CONCLUSION

The Random Forest classifier has proven to be a robust solution for sentiment analysis on the social media platform, providing high accuracy and effective identification of offensive and hate speech content. Ongoing efforts will focus on incorporating user feedback and continuous model refinement to ensure adaptability to evolving language patterns and improved user experience.

## **FUTURE WORK**

For the future work on sentiment analysis it is necessary to perform real time sentiment polarity assigning to the Twitter data. To do so, by preparing an outline to implement same data processing algorithm on cloud that increase the performance for sentiment analysis using Natural Language Processing (NLP) techniques. This can be done by creating nodes on cloud data platform like Hadoop that allow us to store the data on cloud using HDFS (Hadoop File System) and Map-reduce concept to distribute the data processing algorithm on cloud to load and process large size data set and real time sentiment analysis for the linguistic data. This will be contribution towards real time sentiment analysis in a cloud environment and will allow the business user to fetch real time sentiment analysis for the linguistic data.