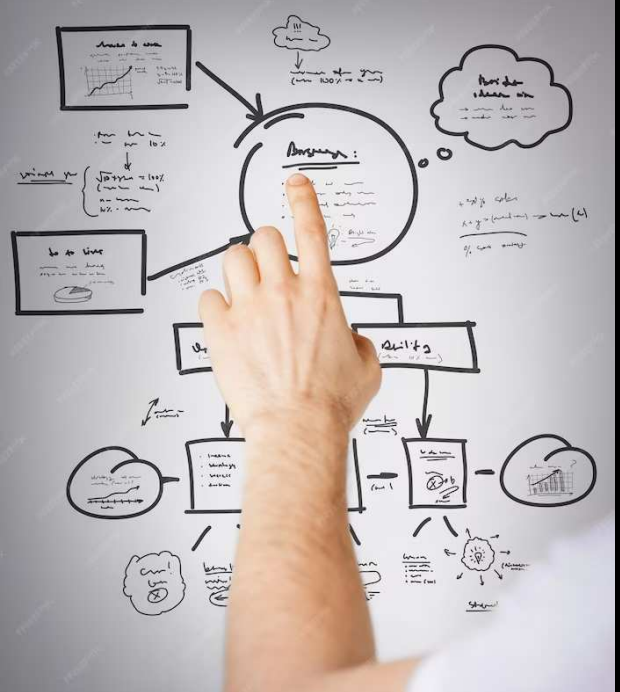
Hinglish Sentiment Analysis: A Methodology for Creating Labeled Datasets and Predicting User Sentiments

Introduction

This presentation explores *Enhancing Hinglish Sentiment Analysis* through a novel methodology for creating labeled datasets and predicting user sentiments. The study aims to improve sentiment analysis accuracy for Hinglish text, which is a blend of **Hindi** and **English**.



Methodology for Labeled Dataset Creation

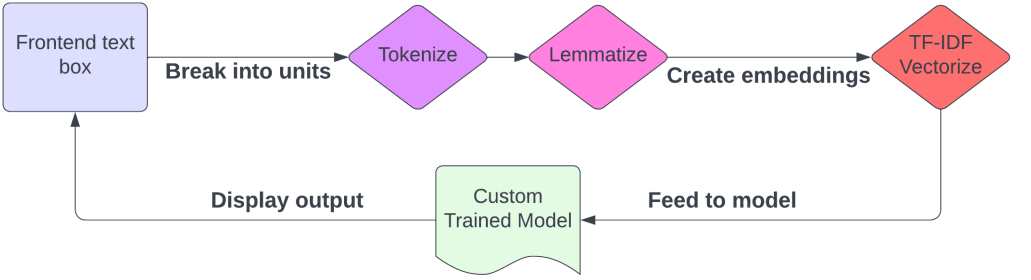
Our methodology leverages **machine learning** algorithms for automatic sentiment labeling, supplemented by manual verification. We also integrate **linguistic analysis** to capture the nuances of Hinglish expressions and emotions.

Brief Details of The Approach

Our method involves manually cleaning a part of a dataset having sentiment labels applied by the pre-trained models.

The so cleaned part of the dataset is used to train a custom machine learning model and iteratively increase the size of the "Initially correct data" and finally making a functional model to predict sentiments of any hinglish sentence.

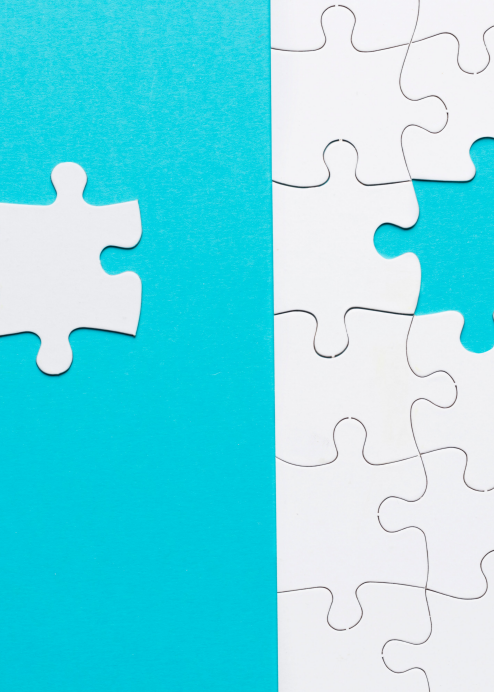
Flow of the model



What we achieved

The dataset had about 18000 sentences

We achieved a test accuracy of 94% on an average of five custom trained ML models.



Challenges in Hinglish Sentiment Analysis

The complexities of **Hinglish** pose challenges for sentiment analysis, including code-switching, transliteration, and informal language. Additionally, existing datasets lack sufficient **labeled data**, hindering accurate sentiment prediction.

Conclusion

The proposed methodology offers a promising approach to enhance sentiment analysis for Hinglish text. By addressing the challenges and leveraging advanced techniques, we can achieve more accurate predictions of user sentiments in this unique linguistic context.

Thanks!

Do you have any questions?

Project by: Viraj Bhagiya, Smit Sutaria, Saurav Navdhare and Aditya Jethani