SOCIAL MEDIA

PROJECT

SARCASM DETECTION OF SOCIAL MEDIA POSTS

BY ~SUMIT KUMAR

SECTION: DS-1 UNIVERSITY ROLL NO. 2023936

GRAPHIC ERA
DEEMED TO BE UNIVERSITY

GUIDED BY DR. GULSHAN DHASMANA

INTRODUCTION TO SARCASM DETECTION

In today's digital age, **sarcasm** has become a prevalent form of communication on social media. This presentation will explore **advanced techniques** for detecting sarcasm, enhancing our understanding of user interactions and improving systems designed to interpret online conversations.

PROBLEM STATEMENT

Despite advancements in natural language processing (NLP), sarcasm detection remains a challenging task due to its subjective nature, dependency on context, and cultural nuances. Existing sentiment analysis systems often fail to detect sarcasm, leading to incorrect interpretations and insights.

OBJECTIVE

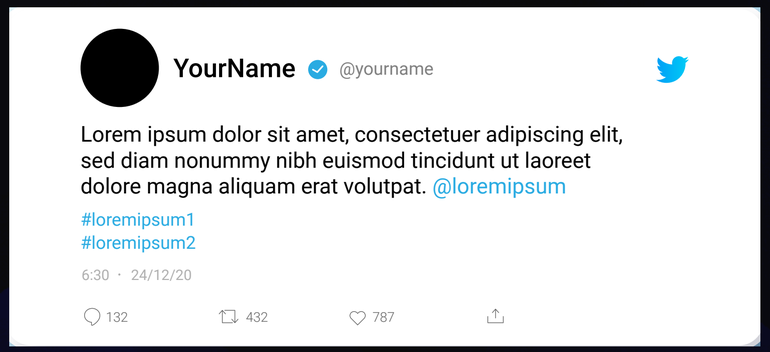
The goal is to develop a robust sarcasm detection system using Natural Language Processing (NLP) techniques to address these complexities and accurately identify sarcasm in text.

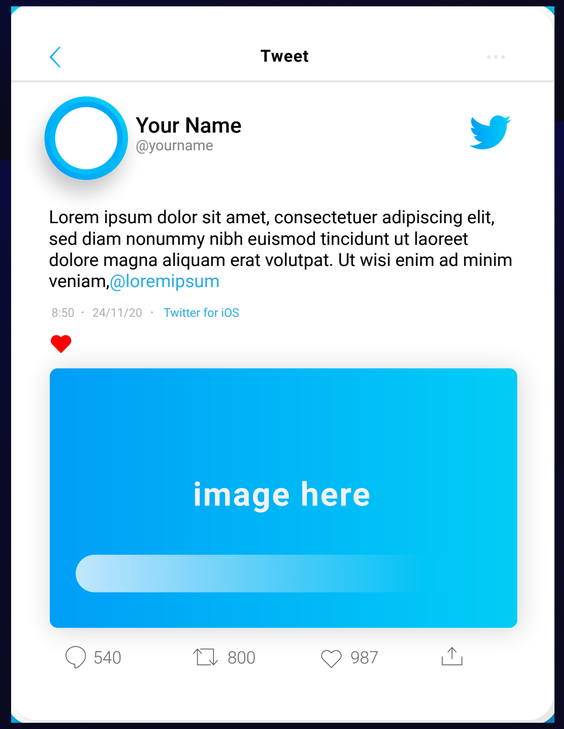
MACHINE LEARNING APPROACHES

Recent advancements in **machine learning** have led to the development of algorithms capable of detecting sarcasm.
Techniques such as **natural language processing** (NLP) and sentiment analysis are pivotal in training models to recognize sarcastic cues in text.

DATASET

The dataset for the project includes text from sources such as Tweets, Reddit comments, or public datasets like the Sarcasm Headlines dataset. These sources provide diverse examples of sarcastic and non-sarcastic text for training and evaluation.





PREPROCESSING

To prepare the data, preprocessing steps include removing URLs, mentions, hashtags, and special characters. Text is tokenized into meaningful components, emojis are converted into descriptive words (e.g., 😀 to "happy"), and words are lemmatized for consistency.

MODEL SELECTION AND TRAINING

For model training, Logistic Regression is used as a simple yet effective classifier, and pipelining is applied to streamline the workflow. The pipeline integrates preprocessing steps like tokenization and vectorization with model training, ensuring a smooth and efficient process for sarcasm detection.

WORDCLOUD

A **word cloud** is a visual representation of text data where the frequency of words is displayed with varying font sizes. More frequent words appear larger, making it easier to identify key terms and themes in the dataset. For sarcasm detection, a word cloud can highlight the most common words associated with sarcastic and non-sarcastic statements, offering insights into language patterns that are useful for model training.

CHALLENGES

Despite advancements, sarcasm detection still faces significant challenges.

Factors such as cultural differences, evolving language, and the use of emoji can complicate the interpretation of sarcasm, requiring continuous adaptation of detection models.

FUTURE DIRECTIONS

In the future, the model can be expanded for multilingual sarcasm detection and improved by incorporating multimedia elements such as images, GIFs, and videos. Additionally, deploying it as a real-time plugin for social media platforms offers practical and impactful applications.

CONCLUSION

The sarcasm detection system achieves good accuracy and enhances NLP tasks like sentiment analysis by addressing the complexities of sarcasm in text.

“Sarcasm is the body’s natural defense against stupidity.”

—UNKNOWN