**TEAM-4**

*Members:*

**Narendra Khadayat**

**Ravi Manisankar Reddy**

**Rahuman Khan**

**Problem Statement:**

**Sarcasm Detection in YouTube Comments**

Sarcasm detection in user-generated content is a significant challenge in natural language processing (NLP). Traditional sentiment analysis tools often misinterpret sarcastic comments, leading to incorrect sentiment classification. This is particularly problematic for platforms like YouTube, where user engagement and content moderation depend heavily on accurately understanding user sentiment. Misinterpretation can lead to poor user experiences and inadequate moderation responses.

**Why It Matters:**

1. **Improved User Experience:** Accurate sarcasm detection enhances content recommendations and user interactions.

2. **Enhanced Content Moderation:** Better detection filters harmful comments, maintaining a healthier community.

3. **Brand Reputation Management:** Helps in identifying and mitigating negative or sarcastic comments that could harm brand reputation.

**Use Cases:**

**Content Moderation:**

**Scenario:** A popular YouTuber uploads a video about climate change.

**Use Case:** Sarcasm detection filters out comments like "**Oh, sure, let's just ignore the melting ice caps!"** (which is actually sarcastic support for climate action).

Benefit: Keeps the comment section respectful and relevant.

**Brand Reputation Management:**

**Scenario:** A smartphone brand releases a new model.

**Use Case:** Detecting sarcastic comments like **"Wow, another groundbreaking feature: a slightly better camera!"** helps the brand understand user sentiments.

**Benefit:** Brands can address concerns and engage with users effectively.

**Sentiment Analysis Enhancement:**

**Scenario:** A music video receives mixed reactions.

**Use Case:** Identifying sarcasm in comments like **"This song cured my insomnia"** (when it's actually disliked) improves overall sentiment analysis.

**Benefit:** Creators get a more accurate picture of audience feelings.

**Solution Approach:**

We propose a deep learning-based approach to detect sarcasm in YouTube comments. Our method combines features extracted from a Convolutional Neural Network (CNN) with handcrafted contextual features to improve detection accuracy.

1. **Data Collection:** Gather a substantial dataset of YouTube comments, identifying sarcastic and non-sarcastic instances.

2. **Feature Extraction:**

**Deep Learning Features:** Utilize a CNN to automatically extract relevant features from the comments.

**Contextual Features:** Manually create features based on contextual clues (e.g., specific words or phrases, punctuation).

3. **Model Training:** Train a classification model using the combined feature set to optimize for accuracy, precision, recall, and F1-score.

4. **Evaluation:** Compare the performance of different models and feature combinations to determine the best approach.

**Scope of the Solution**

**Boundaries and Limitations:**

1. **Scope:**

-Focus on English language YouTube comments.

- Detection limited to text-based sarcasm, excluding other forms such as videos or images.

2. **Limitations:**

- Context limited to textual clues within individual comments and threads.

- Potential variability in sarcasm detection due to the diversity in user expression and cultural differences.

**Various Datasets Considered**

**Datasets Explored:**

1. **Twitter Sarcasm Corpus:**

- **Characteristics:** Contains tweets labelled with hashtags like #sarcasm.

**- Pros:** Large volume of data, varied user expressions.

- **Cons:** Limited context, as tweets are often standalone.

2. **Amazon Reviews:**

- **Characteristics:** Includes product reviews with a mix of genuine and sarcastic feedback.

- **Pros:** Diverse sentiments and product categories.

- **Cons:** Sarcasm is less frequent, and context is often limited.

3. **Reddit Comments:**

- **Characteristics:** Rich in context with threaded conversations.

- **Pros:** High sarcasm occurrence, varied topics.

**-Cons:** Requires extensive preprocessing to handle nested comments.

4. **YouTube Comments (Selected):**

- **Characteristics:** Comments on a wide range of videos, often with replies providing context.

- **Pros:** Rich contextual information, high frequency of sarcasm.

- **Cons:** Noise from off-topic comments and spam.

5. **Dialogues from TV Shows and Movies:**

- **Characteristics:** Scripted dialogues labelled for sarcasm.

- **Pros:** High-quality labels, clear examples of sarcasm.

- **Cons:** May not reflect natural user-generated content accurately.

**Final Dataset**: **YouTube Comments Dataset**

**Reason for Selection**: Although existing datasets like SARC are highly valuable, we determined that a custom dataset of YouTube comments would provide the most direct relevance to our business problem. By collecting and annotating YouTube comments ourselves, we ensure the data is directly applicable, capturing platform-specific nuances and linguistic styles. This tailored dataset will enhance model accuracy and relevance in real-world applications.

