Project Proposal

Group Number: *19*

January 21, 2025

Contents

1 Overview

This project, titled **Sarcastic Text Detector**, aims to classify whether a given text is sarcastic or not. The task involves analyzing short text samples (up to 300 characters) and assigning one of two labels: *Sarcastic* or *Not Sarcastic*.

The primary goal is to develop a robust sarcasm detection system that can handle the subtleties and nuances of sarcastic language. The project is a classification problem, and the expected outcomes include:

- A well-trained model capable of sarcasm detection.

- Insights into linguistic patterns and features that contribute to sarcasm.

- Recommendations for improving sarcasm detection methodologies.

2 Dataset Description

2.1 Dataset Source

The dataset comprises 1000 or more short text samples collected from publicly available sources, including:

- Social media APIs such as Twitter .

- Pre-existing datasets from platforms like Hugging Face or Kaggle.

2.2 Dataset Features

- **Number of Entries:** 1000+ samples of text.

- **Text Length:** Maximum of 300 characters.

- **Labels:** Each sample is classified as either *Sarcastic* or *Not Sarcastic*.

- **Annotation:** Each text will ideally be reviewed by at least three annotators. A majority vote among annotators will decide the label. If disagreements persist, a group discussion or additional annotator will resolve conflicts

2.3 Samples

Below are some examples:

- Source: "Oh great, another Monday. Just what I needed to ruin my week!"
  Classification: *Sarcastic*

- Source: "Looking forward to the weekend with my family. It's been a long week!"
  Classification: *Not Sarcastic*

- Source: "Wow, this traffic jam is exactly what I needed to make my day perfect!"
  Classification: *Sarcastic*

- Source: "Oh, fantastic! Another software update that breaks everything. Exactly what I needed."
  Classification: *Sarcastic*

- Source: "I appreciate how quickly you responded to my email. Thank you!"
  Classification: *Not Sarcastic*

3 Team Contract

3.1 Team Expectations

- Clear communication : Members should convey their thoughts clearly to not create confusion later on.

- Equitable workload distribution : The work load will be distributed in three equal parts and is expected to be completed before the deadline set by the team.

3.2 Conflict Resolution

- Open discussions to identify and resolve issues , voting among three people if necessary.

- If the issue is not resolved consult Dr Charles Welch for help.

3.3 Signatures

Person 1: Devansh Singh

Signature:

Person 2: Ankur Pandey

Signature:

Person 3: Yu Xi

Signature: