Detecting Sarcasm and Figurative Language in Tweets Using NLP Techniques

Sarcasm is a complex way of using words to mean something different from their literal interpretation, something that even the best of us can struggle with when we look at the world of communication. Social media platforms like Twitter are full of posts that rely on sarcasm, irony, or other forms of figurative expression, where tone and context completely change meaning.

Our goal in this project is to build a natural language processing (NLP) system that could effortlessly cut through the ambiguity of social media posts, specifically tweets, and classify tweets into four categories: Regular, Sarcasm, Irony, and Figurative (a mix of sarcasm and irony). We combined linguistic analysis with deep contextual embeddings. We will begin with an analysis of words that don't necessarily follow the normal rules, such as adverbs, intensifiers, commas, and emotional turnabouts. We will then apply well-established NLP models to evaluate which approach best captures this challenging form of language.

The modeling pipeline includes three main stages:

Baseline Model: Logistic Regression with **TF-IDF** vectorization to establish an interpretable benchmark.
Deep Learning Model: BiLSTM or **GRU** using pre-trained **GloVe** embeddings to capture sequential dependencies and contextual cues.
Transformer Model: Fine-tuning **DistilBERT** for deeper semantic understanding of sarcasm and irony.

Expected outcome: A classifier capable of accurately classifying tweets into Regular, Sarcasm, Irony, and Figurative categories while highlighting which linguistic or contextual cues drive the predictions. Model performance will be evaluated using **Accuracy**, **F1-Score (macro)**, and **Confusion Matrix** to ensure both interpretability and robustness.

About Dataset :

Link: Tweets with Sarcasm and Irony (Kaggle)
This dataset contains English tweets classified into four categories: **Regular**, **Sarcasm**, **Irony**, and **Figurative** (tweets combining elements of sarcasm and irony). It is based on the research paper *"An Empirical, Quantitative Analysis of the Differences between Sarcasm and Irony"* by Jennifer Ling and Roman Klinger (2016). This four-class structure makes it ideal for supervised multi-class text classification, allowing comparison of how different NLP models handle various figurative language types in social media text.

Tools and Packages :
Preprocessing: nltk, spacy, re, pandas
Modeling: scikit-learn, keras, torch, transformers
Visualization: matplotlib, seaborn, wordcloud

Team Members :
Darsini Lakshmiah - G31629191
Hussain Nathani - G36308827
Rasika Nilatkar - G28505126