Group 1 Project Proposal: Sentiment Analysis of Urban Dictionary Terms

This project proposes the development of a Natural Language Processing (NLP) model that categorizes words and phrases from Urban Dictionary as either positive or negative based on their definitions and usage examples. This analysis will involve creating an extensive dataset harvested from Urban Dictionary and applying sentiment analysis techniques to classify the sentiment polarity of the words. Beyond just binary classification, the model will attempt to understand the depth of sentiment in these dynamically evolving terms.

The initial phase of the project will involve data collection, where we will use the following dataset available on Kaggle: <https://www.kaggle.com/datasets/therohk/urban-dictionary-words-dataset>. Following this, we will employ text preprocessing methods to clean and organize the data. Using sentiment analysis algorithms, possibly coupled with deep learning techniques such as recurrent neural networks (RNN) or convolutional neural networks (CNN), the model will be trained to classify the terms into positive or negative categories. As a stretch goal, the project may explore the feasibility of identifying neutral terms. Through this project, we aim to develop a tool that not only aids in understanding the sentiment encapsulated in urban language but also serves as a resource for linguistic research.