Topic Proposal

Reddit is a platform for users to spontaneously discover and share content while other audiences can vote or comment to show their opinions on the submitted post. Categories of content on the site are known as "subreddits," which is like a community organized for those people who have similar interests, content on the subreddit includes news, video games, movies, music, books, fitness, food, and image sharing, among others (<https://nealschaffer.com/subreddit/>).

In our project, we are going to take posts and comments from two subreddits that are Apple and Windows to classify the related keywords difference between them, to predict the predominant sentiment comments on the submitted post with a pre-trained model, and to find out the relationship between sequences using transformer machine learning models [Interpretability and Explainability].

Dataset Description

The Reddit data can be extracted using Reddit API, sourced from . There are xxx submitted posts and xxx comments in date range would be used