Abstract:

Due to the extensive use of mobile devices and high-speed internet, social media has connected people at the tip of their fingers in the booming internet age. Social Media has become a major source of sharing different points of view and discussions on different topics, i.e., politics, business, sports, etc. People are associated together via different social media platforms. Twitter is a social media platform allowing users to express their thoughts regarding any event or person, giving the reader diverse insights into popular opinions and sentiments. Such opinions can be in various languages and dialects. This research serves the purpose of text processing on a multilingual dataset, including English, Urdu, and Roman Urdu, exploring machine learning approaches for sentiment analysis, training, and testing textual data collected from Twitter. Our study dataset consisted of tweets in three major languages English, Urdu, and Roman Urdu. However, limited research is being carried out on sentiment analysis on multilingual tweets. We explored six Machine Learning classification algorithms for sentiment analysis and, eventually, the comparison of the ensemble techniques moving forward with acquired results.

Keywords: Machine Learning, Sentiment Analysis, NLP, NLTK, Multilingual Tweets, Artificial Intelligence