INFORMATICS INSTITUTE OF TECHNOLOGY

In Collaboration With

UNIVERSITY OF WESTMINSTER

Rankrr - Sentiment Analysis Based Product Ranking With Emphasized Informal Words & Emojis Interpretation

A Final Report by

Mr. E.M.D.R. Bandara

Supervised By:

Ms. Divya Premanantha

May 2023

This document is submitted in partial fulfilment of the requirements for the BEng (Hons) Software Engineering degree at the University of Westminster.

ABSTRACT

When it comes to recommendation systems which rank products through sentiment analysis, the sentiment scores of emojis are not considered. Also, emphasized informal words such as ‘liiiike’, ‘puuuuurfct’, or ‘yasssss’ are very common in user generated text contents in various online platforms. However, these words cannot be preprocessed properly with existing methods and will lead to poor correctness on tasks such as sentiment analysis which is a part of the NLP (Natural Language Processing) domain. These words are also not considered in sentiment analysis based recommendation systems which will affect in poor correctness on product rankings.

With the use of machine learning, the problem of emphasized informal words has been addressed by training a model based on a supervised classification algorithm with a synthesized dataset of emphasized words so that this model can predict the actual word which is meant by the emphasized counterpart. With this model as well as another model which considers emojis when analyzing sentiment, the correctness of sentiment based ranking process has been improved.

Tests on emphasized informal words model yielded good results on standard evaluation metrics. It also performs well on words with intentional misspellings with repeated letters while the ranking process was also improved according to the benchmarks.

Keywords - Recommendation Systems, Product Ranking, NLP, Sentiment Analysis, Machine Learning, Informal Text Preprocessing, Data Science