INFORMATICS INSTITUTE OF TECHNOLOGY

In Collaboration with

UNIVERSITY OF WESTMINSTER

Classifying Twitter Posts for Movie Success Prediction using a BERT based Approach

A Dissertation by

Mr. Shanuka Fernando

W1761348 / 2019431

Supervised by

Ms. Sachinthani Perera

Submitted in partial fulfilment of the requirements for the BEng in Software Engineering degree at the University of Westminster.

July 2022

Abstract

Predicting a movie's success is a difficult undertaking because the movie industry is expanding exponentially. Many stakeholders rely on the movie's revenue, and marketing techniques can be affected by the outcome of forecasts to increase audience reach and increase revenue. Sentiment analysis, a branch of natural language processing (NLP) that employs NLP methods to extract emotions from text, has been used in several studies to forecast movie success. Several studies have employed sentiment analysis, a subset of natural language processing (NLP) that employs NLP techniques to extract emotions from text, to predict the box office success of films. The majority of studies that used sentiment analysis to predict the performance of movies used conventional research techniques to conduct their research. It has been noted that recent developments in sentiment analysis, such as transformer-based language models, which have the power to drastically alter outcomes in other domains, have not yet been introduced and applied to the domain of predicting movie success. Only a few attempts have been made to increase sentiment analysis's accuracy level.

As a result, this study used Twitter posts to introduce a new transformer-based sentiment analysis model to the domain of predicting movie success. With an F1 Score of 89.68 in multi-class classification, a multi-class categorization of tweets as negative, neutral, and positive was able to outperform the existing approaches.

Keywords – *Sentiment Analysis, BERT, Deep Learning, Language Models*