ILLINOIS INSTITUTE OF TECHNOLOGY
COLLEGE OF COMPUTING

PROJECT PROPOSAL

By:
Zohreh Zzaeimzadeh (A20478411)
Tommy Vadakumchery (A20442320)
Somanshu Gupta (A20499077)

CS-584-01, MACHINE LEARNING

Dr. Yan Yan

1st October, 2022

Project Overview:

Sentiments helps businesses to understand the emotion and sentiments of their customers.Sentiment analysis is used to examine raw text in order to produce objective, quantitative results. In this project, we built and compare different classification models that uses patient reviews to predict the sentiment , which can help recommend the top drug for a given disease.

Problem Statement:

Since covid-19 pandemic has shown up, shortage of healthcare workers, lack of medicines is at its peak. Due to unavailability , individuals started taking medicines without appropriate consultation, making entire medical fraternity in distress. We aim to answer following questions.

- How to use Sentiment Analysis to recommend the drugs?
- To discover the most influential age group and gender of the target audience for a chosen drug?
- What is the emotional inclination of users towards a chosen drug?

Proposed Work:

The efficacy and safety of drugs heavily relies on controlled trials. Current methodologies consisted of the use of statistical measures that involves looking at query logs or other categorical variables for the identification of events such as side effects but there are limited number of studies in the field of drug proposal utilizing sentimental analysis on the basis of medication reviews. As reviews available for drugs are different than those for general products. The drug reviews are much more complicated as they contain medical terminologies like chemical used in drugs, medical health conditions etc. In this project, we will be evaluating the feasibility of leveraging machine learning and natural language processing to classify user ratings based on their textual review to identify the locations of contingency and model will also be used to identify overtly positive or negative scores. These scores are user ratings which was incorrectly classified by the model to identify some generalizations. We will also try to use computational technique to highlight the phrases in the text that have positive or negative impact on the classification results.

Action Plan:

PHASE	DESCRIPTION	EXPECTED END DATE
Phase 1	Dataset Preparation	16th October 2022
Phase 2	Modelling	30th October 2022
Phase 3	Evaluation	7th November 2022
Phase 4	Recommendation	15th November 2022

References

[1] Yassine Al Amrani, Mohamed Lazaar, and Kamal Eddine El Kadiri. Random forest and support vector machine based hybrid approach to sentiment analysis. Procedia Computer Science, 127:511–520, 2018.

[2] P. Karthika, R. Murugeswari, and R. Manoranjithem. Sentiment analysis of social media network using random forest algorithm. In 2019 IEEE International Conference on Intelligent Techniques in Control, Optimization and Signal Processing (INCOS), pages 1–5, 2019.

[3] Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. Roberta: A robustly optimized bert pretraining approach. arXiv preprint arXiv:1907.11692, 2019.

[4] Akhil Shiju and Zhe He. Classifying drug ratings using user reviews with transformer-based language models. medRxiv, 2021.

[5] Kian Long Tan, Chin Poo Lee, Kalaiarasi Sonai Muthu Anbananthen, and Kian Ming Lim. Roberta-lstm: A hybrid model for sentiment analysis with transformer and recurrent neural network. IEEE Access, 10:21517–21525, 2022.

[6] Sairamvinay Vijayaraghavan and Debraj Basu. Sentiment analysis in drug reviews using supervised machine learning algorithms. arXiv preprint arXiv:2003.11643, 2020.

[7] Nurulhuda Zainuddin and Ali Selamat. Sentiment analysis using support vector machine. In 2014 International Conference on Computer, Communications, and Control Technology (I4CT), pages 333–337, 2014.