Detailed Project Proposal

Project Topic: Sentiment analytical system to enhance patient satisfaction on drugs using a fine-tuned Python libraries-based model.

Module Code: 7COM1039

Nneoma Linda Nnorom

Student ID: 21073212

MSc. Artificial Intelligence and Robotics

TABLE OF CONTENT

TABLE OF FIGURES

1.0 THE PROJECT AIM:

In recent years, there is need for a more sufficient way of analysing patient reviews because the amount of patient reviews data increase as time unfolds (Smith, 2017). This project aims at developing a sentiment analytical system to enhance patient satisfaction on drugs using a fine-tuned model-based python libraries. Also, this project aims at investigating and recommending the best model, and python libraries to use in ensuring the enhancement of patient satisfaction.

Sentiment analysis is a subfield of Natural Language Processing (NLP) that understands people's emotions, attitudes, appraisals, and opinions in unstructured text about topics, issues, entities, events, and products. Sentiment analysis will enable us to interpret these numerous drug reviews data and the insights gotten from the information will aid in improving the health care system (Nasrullah et al., 2024). Considering the various fields of Natural Language Processing (NLP), sentiment analysis can be seen to have evolved into a promising domain, which covers various sectors such as economics, politics, and healthcare, etc. Particularly in the pharmaceutical sector, Sentiment analysis can play a vital role because of the ability to access large volumes of user-generated content which gathers detailed information of drug effectiveness and side effects (Uddin et al., 2022).

The application of sentiment analysis for the enhancement of patient satisfaction on drugs will detect and understand patient's emotions and reviews, classify these sentiments, identify specific aspects of drugs, evaluate the result, and finally compare it with the result of other algorithms. This will be achieved through the use of sentiment analysis python libraries which includes: Textblob, Spacy, Naïve Bayes, Scikit-learn This aims at ensuring that at high performance, and high accuracy, patient satisfaction are being met.

2.0 RESEARCH QUESTIONS:

    i. To what extent does leveraging on Spacy and Textblob for preprocessing and sentiment analysis of the data increase the algorithm's performance and accuracy and reduce the complexity of the model.

    ii. To what extent was the aspect-based sentiment analysis using machine learning able to classify and identify the specific aspects of the drug reviews.

    iii. To what extent does comparing the results of the other sentiment analysis models with the result of the proposed system's algorithm enhance the patient satisfaction on drugs.

3.0 PROBLEM STATEMENT:

    Understanding and interpreting patient's reviews on drugs reveals that some drugs prescribed for patients reacts negatively on them thereby making patients not to be satisfied with the treatment administered to them. Leveraging sentiment analysis on patient reviews on drugs have increased the researcher's attention in recent years thereby making them to explore methodologies and approaches to interpret, understand and further enhance patient satisfaction on drugs. However, existing models and algorithms fall short of accurately

classifying these patients review, as the algorithms are too complex to perform, and also have low performance when applied on patient reviews.

4.0 THE PROJECT OBJECTIVES:

The specific objectives of this project are:

1. To leverage sentiment analysis in the development of a model that utilizes the power of Natural language processing to enhance patient satisfaction of drugs.
2. To design and implement an algorithm that uses python libraries Spacy to preprocess data and understand the data with the purpose of obtaining accurate analysable results.
3. To design and implement an algorithm that will apply sentiment analysis on patient reviews on drugs to categorize the review into positive, negative, and neutral using the Textblob.
4. To split the dataset into train and test using the Sklearn.
5. To design an algorithm that will train the naïve bayes classifier.
6. To classify the specific aspects of the patient's reviews using machine learning algorithm.
7. To investigate and evaluate the accuracy, precision and performance of the proposed model.
8. Compare the performance and efficiency of the proposed model with that of other models.
9. To design a User Interface for the model.

5.0 PROJECT PLAN AND TIMELINE:

- **Week 1:** Project setup, Project development environment and GitHub repositories setup
- **Week 2:** Define the project scope, objectives and review existing literature on sentiment analysis of patient satisfaction on drugs. Data Collection of patient reviews on drugs.
- **Week 3:** Review existing literature continuation. Gathering, completion and submission of project proposal.
- **Week 4:** Data cleaning and preprocessing using spacy to ensure accuracy and good performance of the system.
- **Week 5:** Perform sentiment analysis using TextBlob library to classify the patient review into Positive, Negative and Neutral.
- **Week 6:** Exploratory data analysis to visualize the data distribution using matplotlib and seaborn.
- **Week 7:** Split the dataset and perform advanced sentiment analysis to train the Naive Bayes classifier.
- **Week 8:** Implement aspect-based sentiment analysis to identify specific areas that would help in identifying sentiments related to specific aspects of drugs.
- **Week 9:** Data visualization to gain insights. Improve the model based on feedback and translate the findings into insights with recommendations to enhance patient satisfaction.
- **Week 10-11:** Design a User-Friendly Interface environment, validate and test the reliability of the recommendations, and start project report documentation.

- **Week 12:** Final testing, complete project Report and submit with the code repository, and presentation.

6.0 SHORT DESCRIPTION OF IDEA:

This model will leverage on Natural Language Processing Technique to apply sentiment analysis on patient reviews on drugs in the effect to enhance patient satisfaction.The proposed model will be using sentiment analysis and leverage on python libraries. The model's algorithm will be written with Spacy to preprocess and clean the dataset to remove noise, interpret and analyse the sentiments which returns the polarity and subjectivity using Textblob. These polarities range from -1 to 1, which can be further classified as Positive, Negative, or Neutral. Visualization and exploration of the features of the dataset (polarity and sentiments) to understand the positive and negative feedback using matplotlib and Seaborn. To understand the specific aspect of the drug reviews, an aspect-based extraction and classification will be designed using machine learning.

Furthermore, the dataset will be split into Train and Test datasets, the train dataset will be trained on Naïve Bayes classifier which will in turn be evaluated and investigated. The efficiency of the result of this model will be compared with the efficiency of the result of another existing model and recommendations will be made based on the best method that enhances patients' satisfaction. In conclusion, insights can be gained from the results of the model which pharmaceutical companies and healthcare providers can implement in order to improve the drugs that are prescribed and provided to patients.

7.0 METHODOLOGIES:

The methodology that will be used to carry out this project is the CRISP-DM, and this has many phases which includes:

i)   Business Understanding: In this project, I will be focusing on leveraging Natural language technique (Sentiment Analysis) to find the best means to ensure that patient's satisfaction on drugs are enhanced. This will be achieved by comparing and evaluating the results of the existing models with my proposed model.

ii) Data Understanding: This is the second stage of the CRISP-DM where I understood the type of Data for this project. Having understood the project, I searched for datasets and discovered data from an open source (Kaggle) which I will use for this project. This data contains the drug name, the condition, patient's reviews, unique ID, etc.

iii) Data Preparation: I will carefully prepare the data by cleaning the dataset and ensuring that words that are not needed are erased. I will remove stop words, carry out word tokenization and finally explore the data to ensure that it is properly cleaned and prepared because I will be training the dataset with libraries that will give a perfect result when the dataset is properly cleaned.

iv) Modelling: This is the fourth stage of the CRISP-DM. At this stage, I will apply sentiment analysis on the dataset using Textblob and this returns the polarity which would in turn be classified based on positive, negative, and neutral. I will also design an aspect-based algorithm which will classify the various aspects of these patients reviews.

v)   Evaluation: This is the fifth stage of CRISP-DM Methodology. I will evaluate the performance of the model based on its accuracy and precision.

vi)   Deployment: At this stage, I will design a user interface and implement the model on it. This will enable users to easily work with the model.
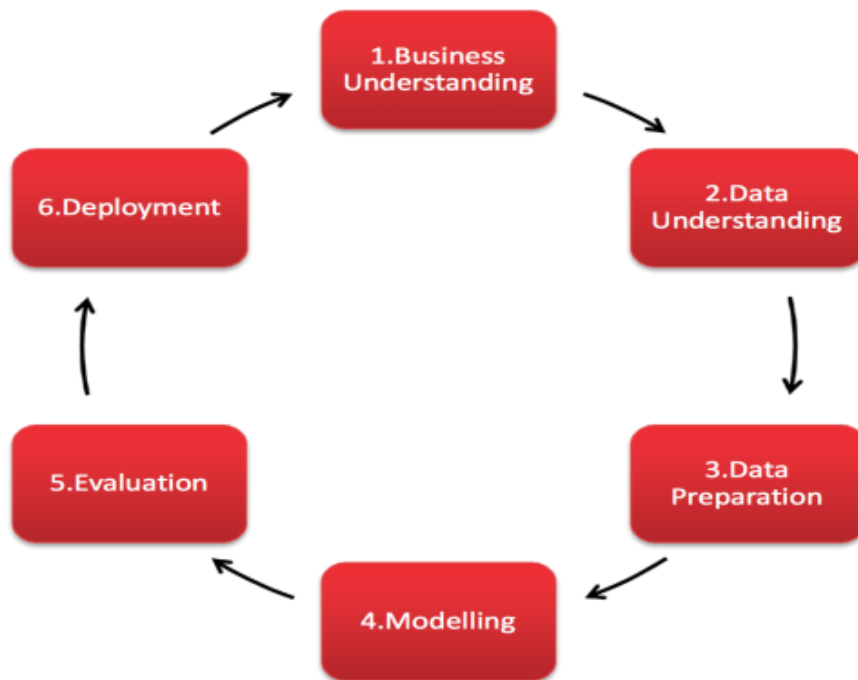

Figure 1: CRISP-DM Methodology (Source: https://www.sv-europe.com/crisp-dm-methodology/)

8.0 REFERENCES

1. Nasrullah Makhdom, H N Verma, Arun Kumar Yadav (2024) Sentiment Analysis of Patient Review in Drug. Journal of Emerging Technologies and Innovative Research(JETIR),

2. Uddin, M. N., Hafiz, M. F. B., Hossain, S., & Islam, S. M. M. (2022). Drug sentiment analysis using machine learning classifiers. International Journal of Advanced Computer Science and Applications, 13(1), 92-100.).

3. Smith, P. (2017). *Sentiment analysis of patient feedback* (Doctoral dissertation, University of Birmingham).

4. Available at: *https://www.sv-europe.com/crisp-dm-methodology/* *[Accessed on 04/06/2024]*

9.0 PROJECT GANTT CHART:

Tasks	Week 1	Week 2	Week 3	Week 4	Week 5	Week 6	Week 7	Week 8	Week 9	Week 10	Week 11	Week 12
Project setup, define the project scope, objectives and review existing literature on sentiment analysis of patient satisfaction on drugs.	■											
Project development environment and GitHub repositories setup, Data Collection of patient reviews on drugs.		■										
Review existing literature continuation. Gathering, completion and submission of project proposal.			■									
Data cleaning and preprocessing using spacy to ensure accuracy and good performance of the system.				■								
Exploratory data analysis to visualize the data distribution using matplotlib.pyplot and seaborn.					■							
Perform sentiment analysis using TextBlob library to classify the patient review into Positive, Negative and Neutral.						■						
Split the dataset and perform advanced sentiment analysis to train the Naive Bayes classifier.							■					
Implement aspect-based sentiment analysis to identify specific areas that would help in identifying sentiments related to specific aspects of drugs.								■				
Data visualization to gain insights. Improve the model based on feedback and translate the findings into insights with recommendations to enhance patient satisfaction.									■			
Design a User-Friendly Interface environment, validate and test the reliability of the recommendations, and start project report documentation										■		
Design a User-Friendly Interface environment, validate and test the reliability of the recommendations, and start project report documentation											■	
Final testing, complete project report/submit with the code repository, and presentation of project.												■

LEGEND
Completed
WIP