**Monolingual Sentiment Analysis on Pharmaceuticals Drug Reviews to Recommend Pain Medication to Patients**

Sharon Wanjiru Ndichu

118461

ICS 4 C

Supervisor Name

Tiberius Tabulu

Submitted in Partial Fulfilment of the Requirements of the Bachelor of Science in Informatics and Computer Science at Strathmore University

School of Computing and Engineering Science

Strathmore University

Nairobi, Kenya

July 2022

Declaration and Approval

I declare that this is work produced from my own research and understanding on this project. It is not subject to theft of any publication or writing of other people. Where information from other sources is used, it is well referenced in the research document.

Student Name: Sharon Wanjiru Ndichu

Admission Number: 118461

Student Signature: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Date: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

The Proposal of **Sharon Wanjiru Ndichu** has been reviewed and approved by **Mr. Tiberius Tabulu**

Supervisor Signature: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Date: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Acknowledgement

I would like to express my gratitude to everyone who assisted me in one way or the other with the completion of my proposal. First and foremost, I express my appreciation to Strathmore University for offering resources such as library materials to assist me with my research. In addition, I appreciate Mr. Tiberius Tabulu, my project supervisor, for his guidance throughout the project proposal. I am also thankful to my parents for providing support and resources needed for my project. Lastly, I am grateful to God for giving me good health and strength to complete this project.

Abstract

As the world's population grows, fewer health facilities and professionals are available to treat each individual. Due to a lack of resources, it is difficult for everyone to visit a health facility for medication. Also, there have been instances where individuals have died or experienced negative side effects because of the drugs they were given. As a result, a person's understanding of proper medication is critical. It will reduce errors and aid in self-treatment for those who do not have access to medical care. Dataset from Kaggle on patients’ pharmaceutical drug review will be analysed using NLP techniques to help identify the best medicine for pain. This can greatly reduce errors and assist patients who self-treat. This recommender system will use linear SVC with TF-IDF vectorization for sentiment analysis. The methodology to be used is design thinking which is solution-based, and the model tested using black box and white box testing. Additionally, accuracy testing and test cases will be used to test the model. The sentiment analysis of the reviews should give a list of medications for pain with the analysis mean, which a patient may use to choose a medication based on which medications received the best reviews.

Table of Contents

[Declaration and Approval ii](#_Toc108623367)

[Acknowledgement iii](#_Toc108623368)

[Abstract iv](#_Toc108623369)

[Table of Contents v](#_Toc108623370)

[List of Figures viii](#_Toc108623371)

[List of Abbreviations ix](#_Toc108623372)

[Chapter 1: Introduction 1](#_Toc108623373)

[1.1 Background Information 1](#_Toc108623374)

[1.2 Problem Statement 2](#_Toc108623375)

[1.3 Objectives 2](#_Toc108623376)

[1.3.1 General Aims 2](#_Toc108623377)

[1.3.2 Specific Objectives 2](#_Toc108623378)

[1.4 Research Questions 3](#_Toc108623379)

[1.5 Justification 3](#_Toc108623380)

[1.6 Scope and Delimitations 3](#_Toc108623381)

[1.7 Limitations 4](#_Toc108623382)

[Chapter 2: Literature Review 5](#_Toc108623383)

[2.1 Introduction 5](#_Toc108623384)

[2.2 Parameters Used in Drug Recommendations 5](#_Toc108623385)

[2.2.1 Composition of Drugs 5](#_Toc108623386)

[2.2.2 Age 5](#_Toc108623387)

[2.2.3 Side effects 5](#_Toc108623388)

[2.2.4 Addictive components 5](#_Toc108623389)

[2.3 Drug Recommendation Avenues 6](#_Toc108623390)

[2.3.1 Over the Counter Recommendation 6](#_Toc108623391)

[2.3.2 Recommendation from Information Sought from the Internet 6](#_Toc108623392)

[2.4 Challenges Associated with the Current Drug Recommendation 6](#_Toc108623393)

[2.5 Related Works 6](#_Toc108623394)

[2.5.1 Ontology-Based Drug Recommendation 6](#_Toc108623395)

[2.5.2 Drug Recommendation System by Implicit Feed Back and Crossing Recommendation 7](#_Toc108623396)

[2.5.3 Drug Recommendation based on Tensor Decomposition 8](#_Toc108623397)

[2.6 Gaps in Related Works 8](#_Toc108623398)

[2.7 Conceptual Framework 9](#_Toc108623399)

[Chapter 3: Methodology 10](#_Toc108623400)

[3.1 Introduction 10](#_Toc108623401)

[3.2 Methodology 10](#_Toc108623402)

[3.2.1 Empathize 10](#_Toc108623403)

[3.2.2 Define 11](#_Toc108623404)

[3.2.3 Ideate 11](#_Toc108623405)

[3.2.4 Prototype 11](#_Toc108623406)

[3.2.5 Test 11](#_Toc108623407)

[3.3 Analysis Diagrams 12](#_Toc108623408)

[3.3.1 Use Case Diagram 12](#_Toc108623409)

[3.3.2 Sequence Diagram 12](#_Toc108623410)

[3.3.3 System Sequence Diagram 12](#_Toc108623411)

[3.3.4 ERD 12](#_Toc108623412)

[3.3.5 Context Diagram 13](#_Toc108623413)

[3.3.6 Data Flow Diagrams, Level 1 13](#_Toc108623414)

[3.4 Design Diagrams 13](#_Toc108623415)

[3.4.1 Dataset Schema 13](#_Toc108623416)

[3.4.2 Wireframe 13](#_Toc108623417)

[3.4.3 System Architecture 13](#_Toc108623418)

[3.5 Deliverables 13](#_Toc108623419)

[3.5.1 Proposal 13](#_Toc108623420)

[3.5.2 Model 14](#_Toc108623421)

[3.5.3 Application Programming Interface 14](#_Toc108623422)

[3.6 Tools and Techniques 14](#_Toc108623423)

[3.6.1 Google Collaboratory 14](#_Toc108623424)

[3.6.2 FastAPI 14](#_Toc108623425)

[3.6.3 Python 14](#_Toc108623426)

[3.6.4 Kaggle 14](#_Toc108623427)

[3.6.5 GitHub Repository 15](#_Toc108623428)

[3.6.6 Sentiment Analysis 15](#_Toc108623429)

[References 16](#_Toc108623430)

[Appendix 19](#_Toc108623431)

[Appendix 1: Gantt Chart 19](#_Toc108623432)

List of Figures

[Figure 2.1 Cross Recommendation Model 7](#_Toc115597299)

[Figure 2.2 “User-Item-Tag” three tuple tensor 8](#_Toc115597300)

[Figure 2.3 Conceptual Diagram 9](#_Toc115597301)

[Figure 3.1 Design Thinking Methodology 10](#_Toc115597302)

[Figure 4.1 Use case Diagram 17](#_Toc115597303)

[Figure 4.2 Sequence Diagram 18](#_Toc115597304)

[Figure 4.3 System Sequence Diagram 18](#_Toc115597305)

[Figure 4.4 Entity Relationship Diagram 19](#_Toc115597306)

[Figure 4.5 Context Diagram 20](#_Toc115597307)

[Figure 4.6 Level 1 Diagram 20](#_Toc115597308)

[Figure 4.7 Database Schema 21](#_Toc115597309)

[Figure 4.8 Login Page 21](#_Toc115597310)

[Figure 4.9 Search Input 22](#_Toc115597311)

[Figure 4.10 Recommendation List 22](#_Toc115597312)

[Figure 4.11 System Architecture 23](#_Toc115597313)

List of Abbreviations

ANN- Artificial Neural Network

API – Application Programming Interface

EMRs- Electronic Medical Records

ERD- Entity Relationship Diagrams

ICD- International Classification of Disease

IFCR- Implicit Feedback and Crossing Recommendation

NLP- Natural Language Processing

OTC – Over the Counter

SDG- Standard Development Goal

SSAD- Structured systems analysis and design method

SVC- Support Vector Classifier

UNII- Unique Ingredient identifier

TF-IDF- Term Frequency-Inverse Document Frequency

# Introduction

## Background Information

Technology has transformed the world into a global village in which individuals exchange their thoughts and knowledge, which is readily available when required. Health information is one of the most frequently searched topics on the internet According to research, 60% of people searched the internet for health-related topics, with 35% of them searching for diagnosis (Rao et al., 2020). To avoid negative effects including death, the health information available must be accurate. Accurate drug recommendation will be done by processing the data available on the internet to more useful information using natural language processing language such as sentiment analysis on pharmaceutical drug review.

Currently, the doctor-to-patient ratio is limited. Since it takes 6 to 12 years to become a doctor, the number of doctors does not increase at the same rate as the population (Rao et al., 2020). As a result, there aren't enough doctors to treat each patient and the patients mostly self-treat for medication through methods such as over the counter medication and seeking information from the internet to self-medicate due to the low ratio. To provide information to patients who are unable to contact a specialist owing to this scarcity, an aid for drug recommendation is necessary.

There are also clinical mistakes which arise from misdiagnosis of a patient. About 40% medical practitioners make prescription mistakes, because they reference to only what they know which is limited (Wittich et al., 2014) . The number of medical facilities and resources available are not enough for the entire population, hence some of the diseases such as pain can be self-diagnosed. Having a medicine recommendation model from drug reviews will help to solve this gap. The drug reviews are feedbacks given by previous patients on drugs and the information can be analyzed using sentiment analysis which is a natural language processing on textual data to get the attitude of a user

Additionally, new discoveries and test on drugs are being made every day. There are more drugs that are added to the existing ones and some being banned for usage. It becomes progressively difficult for doctors to make a recommendation from only treatment or medication (Rao et al., 2020). Therefore, there is a need of a recommender model on drugs from the patients who have tried these medications to boost the already existing knowledge.

With increased use of the internet, there is a need to make use of the varied information that people have given. Some pharmaceuticals have included review sections to their products as one of the methods to measure performance and obtain feedback from customers. Many individuals rely on reviews before purchasing a product because they show the level of trust that other customers have in it. Monolingual sentiment analysis is identifying and categorizing opinions expressed in a single language text to determine the attitude towards a product (L. Zhang & Liu, 2017). Sentiment analysis can be used in drug recommendation to know the attitude of users towards a medication and drugs recommended using other user satisfaction.

A drug review sentiment analysis model can aid in providing individuals with additional medical recommendations. It can improve recommendation because other patients can use other patients’ opinions to choose the drugs with the best feedback. Also, it can help specialist know the feedback of patients on drugs they administer and recommend the best medication to their users. SDG goal 3 of good health and well-being will be attained because it reduces the hiccups experienced in medicine administering by a specialist or self.

## Problem Statement

Currently, most of the health informatics available on the internet is not credible, just raw data from users. Most of the patients depend on recommendation from people, pharmacists or opinions on the internet. This prevents people from accessing good quality information (Goyal et al., 2020). The problem that will be addressed in this model is lack of credible data from the internet for self-treatment.

The proposed solution aims at using the reviews given on pharmaceutical drugs to know the feedback from users and the solution to be used for suggestion to other patients. Reviews contain sentiments from different users and will be used as a measure of satisfaction. It will use vectorization algorithms such SVC to data mine and analyze the data to provide an accurate analysis for medicines.

## Objectives

### General Aims

The general objective is to do sentiment analysis on pharmaceuticals drug reviews to recommend drugs for pain to other patients.

### Specific Objectives

1. To investigate parameters considered for drug recommendation to patients
2. To review the ways of recommending drug to patients.
3. To investigate the challenges associated with current drug recommendation systems
4. To review the solutions that recommend drugs to a patient.
5. To develop a drug suggestion model using sentiment analysis on pharmaceuticals drug reviews.
6. To validate the model developed using FastAPI.

## Research Questions

1. What are the parameters considered for drug recommendation to patients?
2. What are the ways of recommending drugs to patients?
3. What are the challenges associated with current recommendation of systems?
4. What are the solutions that recommend drugs to a patient?
5. How will the proposed solution be developed?
6. How will the solution be validated using FastAPI?

## Justification

There are many challenges facing drug recommendation for patients. Some of the challenges are low ratio of a specialist per patient(Rao et al., 2020), the absence of enough health facilities and resources for each patient (Wittich et al., 2014) and clinical mistakes from specialists(Rao et al., 2020).With drug recommendation model , patients do not need to see a specialist unless symptoms persist or is critically ill hence reducing the hiccup of low ratio of specialist per patient and lack of enough facilities and resources for patients. Moreover, clinical mistakes can be reduced by pharmacist knowing which medications has the best feedback from patients and which ones do not have positive feedback.

The solution will use sentiment analysis on drug review from patients, to able to quantify their sentiments as either good or bad. With this analysis, the model will then give a list of the drugs and their analysis for patients or specialist to boost the already existing recommendation system.

## Scope and Delimitations

The project will only focus on drugs to treat pain because it is a common ailment that patients experience and the sentiment analysis of its medications. The proposed drug recommender model will give a list of medications associated with pain and the results from sentiment analysis on pharmaceutical drug reviews.

## Limitations

This project is only limited to giving a list of medications and their analysis which is the score from sentiment analysis but the actual prescription needs the consultation of a pharmacist. This is because patients require recommendations based on other characteristics such as age, which the recommender model will not provide.

# Literature Review

## Introduction

This chapter discusses the parameters used in drug recommendation, drug recommendation avenues, solutions for recommending drugs to patients, as well as the challenges associated with these drug recommendations avenues and solutions. It also explores related works in drug recommendation systems and gaps in related works. Conceptual framework is also illustrated in this chapter.

## Parameters Used in Drug Recommendations

### Composition of Drugs

The substances used in drugs is important to know the side effects or allergic reactions that may occur when recommending a drug (Rainsford, 2009). Additionally, knowing the constituents of drugs is important to prevent extreme cases such as fatalities which may arise from the adverse effects of a drug.

### Age

Age is a crucial factor when recommending medicines. Different age groups take different medications because the age can affect how drugs are absorbed and broken down by the body(Mangoni & Jackson, 2004). Some medications may be either too strong or too weak for an individual based on age, hence age is a factor considered in medicine recommendation.

### Side effects

Side effects of drugs is an important factor to consider before administering of drugs. The side effects may be contributed by factors such as other medication being taken and age of a patient. Medicine may cause adverse side effects if they are not considered such as dizziness, nausea failure of organs or even death (O’Donovan et al., 2019). Therefore, side effects a very important factor to consider in medicine recommendation.

### Addictive components

Addictive components is a key factor to consider when administering drugs to patients. Addiction occurs when a drug causes pleasure caused by changes to the brain and liking to an effect of a drug (Jakovljevic et al., 2015). In addition, craving of drugs hence the addictive components of a drug are important to consider in recommendation of drugs.

## Drug Recommendation Avenues

### Over the Counter Recommendation

Over the counter medicines are common and mostly use factors such as age and regular medicines to treat an ailment, with the most common over the counter medication including: cough medications, painkillers, codeine-based medicines, sedatives, antihistamines and decongestants(Cooper, 2013). The danger of using this method include adverse side effects such as dizziness and nausea from medicine because of the composition of the drugs. Also, allergic reactions such as rashes or itching may occur due to the constituents of a medicine (Chautrakarn et al., 2021). With such kind of recommendation, effects are not clearly considered and if the drug is worth to buy.

### Recommendation from Information Sought from the Internet

Most consumers utilize online health information to self-treat. Health information is widely available on the internet. According to Chautrakarn (Chautrakarn et al., 2021), most of the medication information is incorrect hence it can become a source of poisoning which can cause death. In addition, there might be side effects and allergic reactions because the composition of medicines is not considered.

## Challenges Associated with the Current Drug Recommendation

The current methods of drug recommendation do not consider insights into those affected. There isn't any qualitative research on the medications' effects. Medicines compositions are not considered hence there is probability of side effects and allergic reactions. A study by Cooper (Cooper, 2013), has shown OTC use proxy, self-report and non-OTC specific data. Moreover, there is a challenge identifying credible data from information sought from the internet(Cline & Haynes, 2001), which makes the current recommendation method not suitable.

## Related Works

### Ontology-Based Drug Recommendation

A research by Doulaverakis (Doulaverakis et al., 2012),used a semantic-enabled online service that can provide drug-disease interaction information. Medical data is converted to ontological form and compared to medical knowledge using international standards such as ICD-10 and UNII. It makes a drug recommendation based on the infection, sensitivity, and drug interactions of the patient. However, this system utilizes a lot of system resources which is undesirable when implementing. It also requires a lot of medical expertise from a developer hence may not solve the problem of the evolving drugs which may not capture the effects on a patient.

### Drug Recommendation System by Implicit Feed Back and Crossing Recommendation

According to chen (Chen et al., 2018) , this recommendation is done using the Electric Medical Records. The EMRs are designed to implement collecting, searching, statistical analysis and drug diagnosis. IFCR carries out deep analysis to achieve the most effective drug. IFCR involves three steps, the first is raw representation from EMRs, then model the representations by a non-negative matrix factorization for a set of robust features. Finally, the extracted features are used for the recommendation. The IFCR uses the cross recommendation where there are several symptoms. Figure 2.1 below is a sample pseudocode for cross recommendation:

A picture containing text

Description automatically generated

Figure . Cross Recommendation Model

However, the IFCR does not incorporate the patient’s emotions on the drugs, it is only based on the EMRs which do not include reviews. It therefore can recommend a drug but the recommendation is not based on the view of other patients.

### Drug Recommendation based on Tensor Decomposition

A research by Zhang (Y. Zhang et al., 2014) , used an algorithm based on tensor decomposition. It uses the ‘User-Item-Tag’ three tuple to model a tensor. Personalized recommendation can be received by patients based on extracted important tensor according to a drug predicted rating. Evaluation index used in this model is accuracy and recall. Finally, Top-N drug recommendation list is gotten from each user using tensor decomposition. It is then checked if the drug is in the recommendation list with the patient and tags. Figure 2.2 shows ‘User-Item-Tag’ three tuple tensor:

A picture containing white

Description automatically generated

Figure . “User-Item-Tag” three tuple tensor

However, this analysis has shortage of collaborative filtering when dealing with big and sparse data. It makes use of information such as the name, description, and rating. This approach does not allow for the analysis of user sentiment about a medication.

## Gaps in Related Works

Most of the works are focused on prescribing drugs based on patient records, ratings, and symptoms. They also use the expertise on drug such as the ontology-based recommendation system. However, these works do not include patients' sentiments, which can be utilized to determine a patient's emotional response to a drug, as well as whether it is successful and if they are satisfied with it.

## Conceptual Framework

The conceptual diagram below shows how the Kaggle drug review dataset will pass through data cleaning and preprocessing. In the data cleaning and preprocessing there will be checking of null values, removal of unnecessary values such as punctuations and checking duplicate values. Text tokenization is also done in the data preprocessing. The cleaned data is will then be split into 80% training dataset and 20% testing dataset. The training set is used in building of the model and it needs a lot of data to learn hence a higher ratio. The test dataset uses unseen data to gauge if the model is working as required. After the data cleaning and preprocessing, sentiment analysis is done on the data which involves extracting the data as either positive or negative. The sentiment analysis will use n-gram to classify the sentiments. The model will then built to cluster the medication for pain and deployed on FastAPI where a user can see the list of the recommended drugs to treat pain.

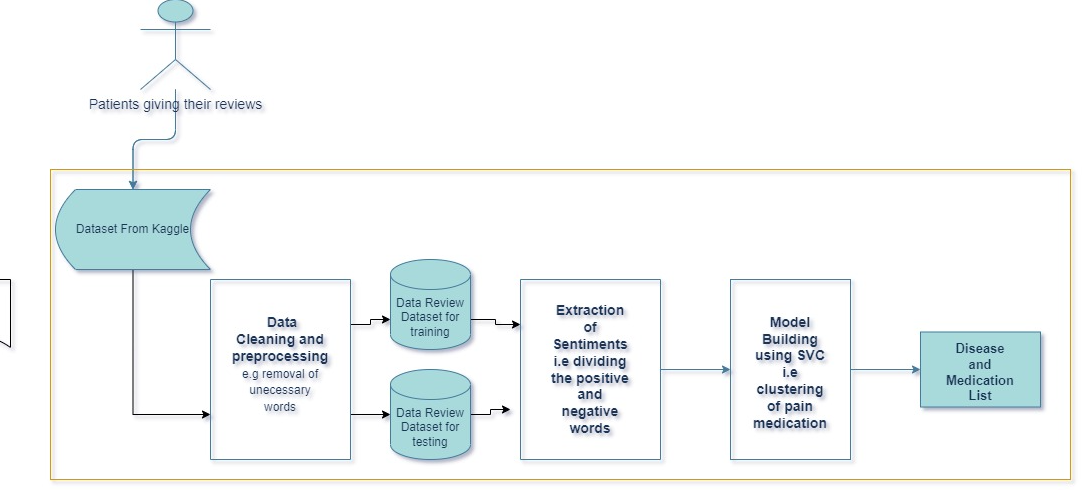


Figure . Conceptual Diagram

# Methodology

## Introduction

This chapter will cover the methodology approach that will be used in this application development and the steps involved in the development approach. Moreover, the chapter will cover analysis and design diagrams.

## Methodology

The methodology that will be used in the model development is design thinking. It is a method of problem-solving that puts the requirements of the users first. Solution-based approach is used to solve problems in design thinking, where it places more of an emphasis on finding solutions to problems unlike a problem-based approach that looks for limitations on why a problem exist (Raju, 2021). Figure 3.1 shows the design thinking methodology.

Diagram

Description automatically generated

Figure . Design Thinking Methodology

### Empathize

Empathy is the first step in the design thinking methodology because it helps understand the problem that is trying to be addressed. A designer observes or studies with empathy how users are affected by a situation. This stage is important for knowing the user's requirements besides the problem to provide a more personalized solution.

### Define

In the second step, findings from the empathize stage are used to define the problem that is trying to be solved. Considered factors include the challenges users face, the issues they encounter frequently, and how a problem affects them. This step is important because problems can be defined once they have been synthesized.

### Ideate

In this step, brainstorming of how to solve the problems defined takes place. This stage tries to create new ideas, whereby there may more be one than more idea. Possible problems from the user environment are considered when suggesting a proposed solution. Analysis diagrams using SSAD design paradigm in section 3.3 which include the use case diagrams, sequence diagram, system sequence diagram, ERD, context diagram and the dataflow diagram will be used to try and ideate how a user and the model will interact. In addition, design diagrams in section 3.4 which includes the Dataset schema, wireframe and system architecture will be drawn in this stage to show the visualization of how the project is supposed to be. This step is important because the best solution for a problem is sought.

### Prototype

This step involves actual development of a model. Effectiveness of a solution is tested in this stage. All the possible solution may be tested in this step and the effectiveness checked. This stage is important because the less effective option can be dropped and the best solution adopted for use. Tools and techniques discussed in section 3.6 will be used in the development of the prototype. The tools and techniques that will be used include the FastAPI, Google Collaboratory, sentiment analysis, python language, Kaggle dataset and the GitHub repository. These tools and techniques are important for the development of the prototype because they provide the platform, language and storage for the project.

### Test

This final stage tests the best solution from the prototype. The end results of this stage are refined because it is an iterative methodology. Blackbox testing and white box testing will be used to test the model. The black box testing is done against specification to discover faults because of not having completely fulfilled specification whereas white box testing is done against implementation and discovers faults in the implementation. Accuracy testing will be done on the model and results given in percentage to know if a model is overfitting or underfitting, this is done by dividing the number of correctly classified samples with total number of samples in the drug review dataset. Test cases will also be drawn, the first test case is where the results of drugs are expected when the ailment is pain and a test case where no results is shown when the ailment is not pain.

## Analysis Diagrams

The analysis methods that will be used are based on the SSAD methodology. The analysis diagrams to be used in this are use case, sequence diagram, system sequence diagram, ERD, context diagram and the data flow diagrams level 0 and 1.

### Use Case Diagram

A use case diagram describes a system and how the users of a model will use it but not the actual workings of the model. The requirements of the entire application or a specific portion are described. It displays the system's interactions with both internal and external actors(Waykar, 2015).It will be applied in this application to display user-system interactions.

### Sequence Diagram

A sequence diagram describes the events in a project and the specific order the project follows. It shows how these processes occur simultaneously. Horizontal lines with messages between them in the sequential sequence of their occurrence are placed between these activities (Al-Fedaghi, 2021). This  sequence diagram will be used in this application to show the activities in the model and how they follow each other.

### System Sequence Diagram

A system sequence diagram shows the whole system sequence diagram. This will include the use case, external actors, and internal events of the system. It generally shows how the whole system will work not only the internal processes of the application. It will be used in the application to show how the application will interact with the external environment.

### ERD

It is a relationship model that illustrates the entities to a system. It shows the relationship between the entities such as the people, objects, places and events within the system. The ERD in this model will be used to show the relationship between various entities in the model (Li & Chen, 2009).

### Context Diagram

A context diagram is used to show the entire system as one process. It illustrates the information flow between the system and external entities. A context diagram will be used for a clear visualisation of the whole application and how it will work (Ibrahim & Yen, 2010).

### Data Flow Diagrams, Level 1

In the data flow diagram level 1 the whole application is represented as a single process but sub-processes are added in the level 1 process.It is more descriptive and includes the processes of the context diagram (“Levels in Data Flow Diagrams (DFD),” 2019). It will be used in this application to show a more detailed system process.

## Design Diagrams

This is the process for defining the model, architecture, and their components. The system design satisfies specific requirements. The system design diagrams that will be used are the database schema, wireframes, system architecture (Odhiambo, 2018).

### Dataset Schema

It refers to the visual representation of the drug review dataset. The dataset schema shows the entities that will be used in the drug review sentiment analysis.

### Wireframe

A wireframe is a framework that shows the design and functionality of a user interface. The project's wireframes will show the interface elements that will be present on the relevant page.

### System Architecture

System architecture defines the behavior, the structure, the interactions, and the views of a system.  It addresses the properties, concepts, architectural principles and characteristics of the system.The system architecture describes the non-functional decisions and the functional decisions of a system. It acts as the blueprint and shows the coordination and communication in the system.

## Deliverables

### Proposal

The proposal for the project will be delivered, which provides an outline of how the project is expected to be. It consists of the abstract, chapter 1, chapter 2 and chapter 3. The abstract will give an overview of what is expected in the project. Chapter 1 consists of the background information of the project, problem statement, objectives, research questions, justification, scope, delimitations and limitations. Moreover, it covers chapter 2 which discusses parameters used in drug recommendation, drug recommendation avenues, related works and the gaps in this related works. This chapter also illustrates the conceptual diagram. Lastly, chapter 3 covers the methodology which includes the methodology, design paradigm that consists of the analysis diagrams, design diagrams, deliverables, the tools and techniques used in the project building. The Gantt chart is also presented to show the timeline of the project activities.

### Model

The recommender model will be delivered. The model is expected to do sentiment analysis using the support vector machine on drug reviews and give a list of recommended drugs for pain.

### Application Programming Interface

The API will allow the model to accessed by the user. The sentiment analysis model will be interfaced on the FastAPI, then a user will be able to use the model and test it. The user will key in the disease and a list of the medication and sentiment analysis displayed.

## Tools and Techniques

The tools and techniques discussed are the FastAPI, Google Collaboratory, sentiment analysis technique, python, GitHub repository and Kaggle dataset.

### Google Collaboratory

Google Collaboratory is an environment hosted on google drive to provide the tool for coding and building the project model. It is suitable for machine learning projects and data analysis. Google Collaboratory allows the project to be hosted on cloud making it suitable for use.

### FastAPI

FastAPI is a web framework for building restful APIs with python. The FastAPI will be used to deploy the model and allow the users to interact with the model.

### Python

Python programming language is a high-level language used in development. It is suitable to use because it is open source, scalable and easy to learn. It works well for natural language processing because of the rich processing features and simple syntax.

### Kaggle

Kaggle is a community hosted online for data scientists and machine learning engineers. Kaggle provides dataset to be used in model training and testing.

### GitHub Repository

GitHub repository is used to store a project’s development and collaboration. The GitHub repository can also have README file to give an insight of a project.

### Sentiment Analysis

Sentiment analysis is a technique that tries to get emotions and opinions from a text. The sentiment analysis can either be positive or negative. The technique can be used in a sentiment analysis project.

# System Analysis and Design

## Introduction

In this chapter, system requirements which include the functional and non-functional requirements are discussed. Moreover, the system analysis and design diagrams in Chapter 3 are drawn and discussed.

## System Requirements

System requirements are the specifications that are required by the system to make it functional and satisfy the user needs and make the system work. These requirements are both functional and nonfunctional. Some of the system requirements reviewed in the project include.

### Functional Requirements

The functional requirements covers what the system is supposed to do. These functional requirements include:

1. **Authentication Module**

The authentication module is used to verify a user. The login and the registration page were used in the authentication. The emails, username and passwords are collected. An email can only be used once. The password used password hashing function to hash the password.

1. **User Interface**

The user interface is used for a user interaction with the system. The user searches for drugs using a user interface hosted on the web and receive the recommendation list.

1. **Pain Medication Recommendation Model**

The model recommends the most suitable medication to treat pain from users’ sentiments. The model uses dataset from Kaggle to make recommendations. The model goes through processes before it can finally give a recommendation list. The first process includes data cleaning and preprocessing and then splitting of the data into training and testing dataset. Extraction of sentiments is then done to the data. A model is then built using SVC to cluster the pain medications.

### Non-Functional Requirements

Nonfunctional requirements is used to define how the application will behave and the limits of its functionality.The non-functional requirements of this model includes:

1. **Security**

The authentication was used to make the model secure and less prone to hacking or illegal access. Password hashing was one of the method used to convert the entered password to a hash and when logging in the password is compared to the hashed password using the password verify functionality. Also for security purposes on the authentication module, an email can only be used once.

1. **Model Accuracy**

Accuracy is used to measure the correctness of the prediction made by the model. The model needs to have minimum errors for it to have high accuracy levels. This ensures that the users are not mislead by the predictions.

1. **Data Integrity**

Data integrity ensures that the data used is accurate, consistent and reliable. This was done at the data cleaning and preprocessing stage of the dataset which includes removal of duplicate data, elimination of unnecessary words and removal of missing values. This ensures that the data used in making the model is of high quality.

1. **Performance**

Performance was used to assess if the model accurately achieves the task of recommending the highest rated pain medication. Moreover, it checked how the model performed when it was deployed to a web application through the FastAPI.

## System Analysis Diagrams

### Use Case Diagram

The use case in Figure 4.1 shows how actors of the system will interact with the model. The actor is any user of the system who interacts with the model to get the recommendation by the system.

Diagram

Description automatically generated

Figure . Use case Diagram

### Sequence Diagram

A sequence diagram illustrates the sequence of messages between objects in an interaction. It describes how and in what order a group of objects works together.This is shown in Figure 4.2:

Diagram

Description automatically generated

Figure . Sequence Diagram

### System Sequence Diagram

System sequence diagram shows the interaction of the user with the system showing the input and output events. The Figure 2.1 shows the system sequence diagram:

A picture containing diagram

Description automatically generated

Figure . System Sequence Diagram

### Entity Relationship Diagram

An Entity Relationship Diagram shows the relationship between each entity. It was used to show a model of the final system and attributes. The Figure 4.4 shows the ERD:

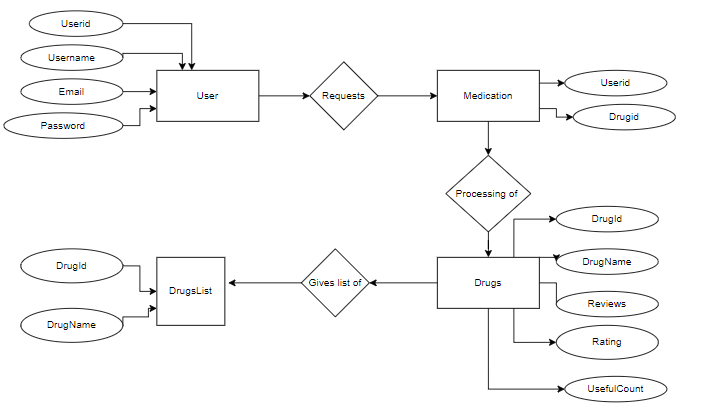


Figure . Entity Relationship Diagram

### Context Diagram, Level 0

A context diagram was used to display the system as a whole. It shows all the external entities and how they interact with the system. The application was put in the middle and the external entities that surround the system without going deep into the system.The Figure 4.5 shows the context diagram:

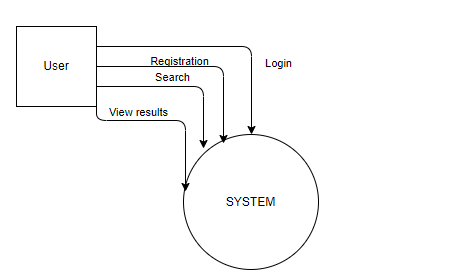


Figure . Context Diagram

### Level 1 Diagram

The Level 1 one diagram shows how the entities of the application interact with each other and how they interact with the application.

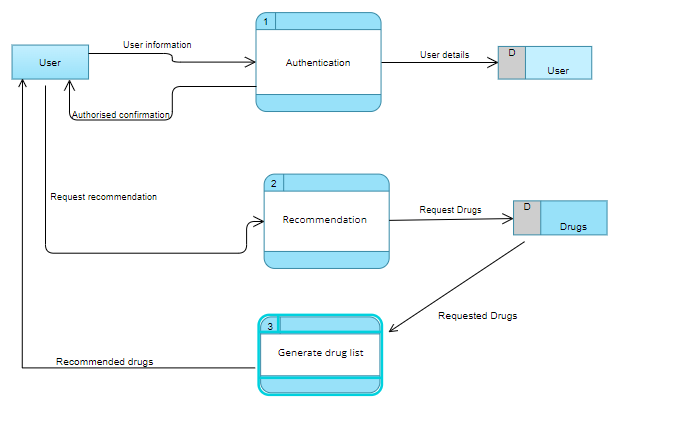


Figure . Level 1 Diagram

## Design Diagrams

### Dataset Schema

A database schema **w**as used to show a blueprint of how the dataset was constructed. It was used to show the relationship between the datasets.

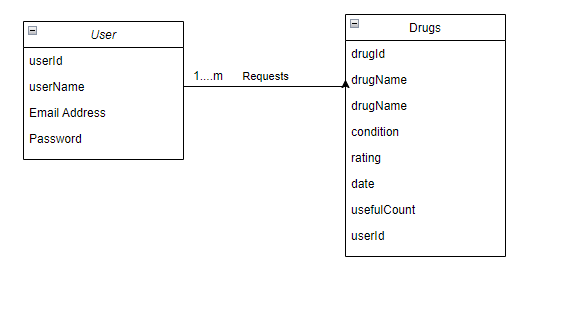


Figure . Database Schema

### Wireframe

Wire frames shows different interfaces. It shows the models how different interfaces will look like. Below are some of the wireframes.

Graphical user interface

Description automatically generated

Figure . Login Page

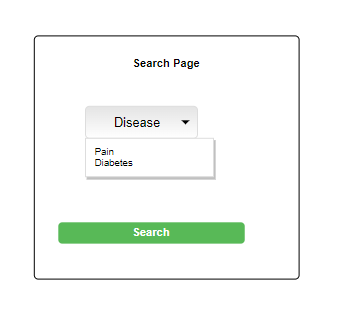


Figure . Search Input

Application

Description automatically generated

Figure . Recommendation List

### System Architecture

The system architecture is used to show the blueprint of the application and the coordination of the application.

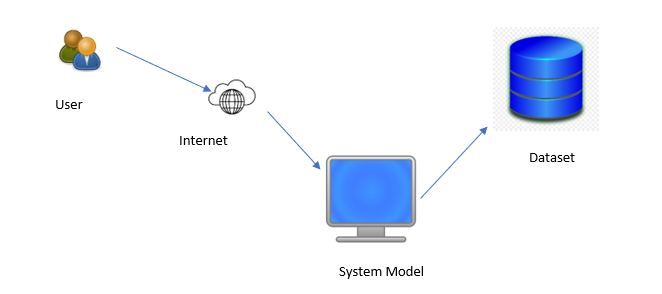


Figure . System Architecture

References

Al-Fedaghi, S. (2021). UML Sequence Diagram: An Alternative Model. *International Journal of Advanced Computer Science and Applications (IJACSA)*. https://hal.archives-ouvertes.fr/hal-03243870

Chautrakarn, S., Khumros, W., & Phutrakool, P. (2021). Self-Medication With Over-the-counter Medicines Among the Working Age Population in Metropolitan Areas of Thailand. *Frontiers in Pharmacology*, *12*. https://www.frontiersin.org/articles/10.3389/fphar.2021.726643

Chen, C., Zhang, L., Fan, X., Wang, Y., Xu, C., & Liu, R. (2018). A Epilepsy Drug Recommendation System by Implicit Feedback and Crossing Recommendation. *2018 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computing, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI)*, 1134–1139. https://doi.org/10.1109/SmartWorld.2018.00197

Cline, R. J. W., & Haynes, K. M. (2001). Consumer health information seeking on the Internet: The state of the art. *Health Education Research*, *16*(6), 671–692. https://doi.org/10.1093/her/16.6.671

Cooper, R. J. (2013). Over-the-counter medicine abuse – a review of the literature. *Journal of Substance Use*, *18*(2), 82–107. https://doi.org/10.3109/14659891.2011.615002

Doulaverakis, C., Nikolaidis, G., Kleontas, A., & Kompatsiaris, I. (2012). GalenOWL: Ontology-based drug recommendations discovery. *Journal of Biomedical Semantics*, *3*(1), 14. https://doi.org/10.1186/2041-1480-3-14

Goyal, V. A., Parmar, D. J., Joshi, N. I., & Champanerkar, K. (2020). *Medicine Recommendation System*. *07*(03), 5.

Ibrahim, R., & Yen, S. Y. (2010). Formalization of the Data Flow Diagram Rules for Consistency Check. *International Journal of Software Engineering & Applications*, *1*(4), 95–111. https://doi.org/10.5121/ijsea.2010.1406

Jakovljevic, M. (Michael), Lazarevic, M., Jurisevic, M., & Jovanovic, M. R. (2015). When cure becomes an illness—Abuse of addictive prescription medicines. *Frontiers in Pharmacology*, *6*. https://doi.org/10.3389/fphar.2015.00193

Levels in Data Flow Diagrams (DFD). (2019, March 18). *GeeksforGeeks*. https://www.geeksforgeeks.org/levels-in-data-flow-diagrams-dfd/

Li, Q., & Chen, Y.-L. (2009). Entity-Relationship Diagram. In Q. Li & Y.-L. Chen (Eds.), *Modeling and Analysis of Enterprise and Information Systems: From Requirements to Realization* (pp. 125–139). Springer. https://doi.org/10.1007/978-3-540-89556-5\_6

Mangoni, A. A., & Jackson, S. H. D. (2004). Age-related changes in pharmacokinetics and pharmacodynamics: Basic principles and practical applications. *British Journal of Clinical Pharmacology*, *57*(1), 6–14. https://doi.org/10.1046/j.1365-2125.2003.02007.x

O’Donovan, B., Rodgers, R. M., Cox, A. R., & Krska, J. (2019). Use of information sources regarding medicine side effects among the general population: A cross-sectional survey. *Primary Health Care Research & Development*, *20*, e153. https://doi.org/10.1017/S1463423619000574

Rainsford, K. D. (2009). Importance of pharmaceutical composition and evidence from clinical trials and pharmacological studies in determining effectiveness of chondroitin sulphate and other glycosaminoglycans: A critique. *The Journal of Pharmacy and Pharmacology*, *61*(10), 1263–1270. https://doi.org/10.1211/jpp/61.10.0001

Raju, K. S. (2021, April 23). *What Is Design Thinking And Why Is It Important?* GreatLearning Blog: Free Resources What Matters to Shape Your Career! https://www.mygreatlearning.com/blog/importance-of-design-thinking/

Rao, T. V. N., Unnisa, A., & Sreni, K. (2020). *Medicine Recommendation System Based On Patient Reviews*. *9*(02), 5.

Waykar, Y. (2015). Role of use case diagram in software development. *International Journal of Management and Economics*.

Wittich, C. M., Burkle, C. M., & Lanier, W. L. (2014). Medication Errors: An Overview for Clinicians. *Mayo Clinic Proceedings*, *89*(8), 1116–1125. https://doi.org/10.1016/j.mayocp.2014.05.007

Zhang, L., & Liu, B. (2017). Sentiment Analysis and Opinion Mining. In C. Sammut & G. I. Webb (Eds.), *Encyclopedia of Machine Learning and Data Mining* (pp. 1152–1161). Springer US. https://doi.org/10.1007/978-1-4899-7687-1\_907

Zhang, Y., Zhang, D., Hassan, M., Alamri, A., & Peng, L. (2014). CADRE: Cloud-Assisted Drug REcommendation Service for Online Pharmacies. *Mobile Networks and Applications*, *20*, 348–355. https://doi.org/10.1007/s11036-014-0537-4

Appendix

1. Gantt Chart

Chart

Description automatically generated

Appendix : Gantt Chart