Classifying Drugs Using Machine Learning: Confidential Drugs As X, Y, Z

An Internship Project Report

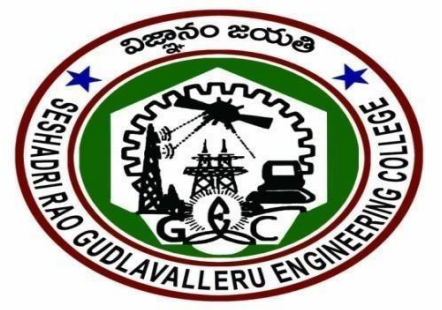
Submitted to the Faculty of Engineering of
JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY KAKINADA, KAKINADA

In partial fulfillment of the requirements for the award of the Degree of

BACHELOR OF TECHNOLOGY
In
INTERNET OF THINGS
By

P. Syam Kumar
(20481A6037)

Under the guidance of
Dr.Y.Syamala
Professor & Head, Department of IOT



SESHADRI RAO GUDLAVALLERU ENGINEERING COLLEGE

DEPARTMENT OF INTERNET OF THINGS

(An Autonomous Institute with Permanent Affiliation to JNTUK, Kakinada)

SESHADRI RAO KNOWLEDGE VILLAGE

GUDLAVALLERU – 521356

ANDHRA PRADESH

2023-24

SESHADRI RAO GUDLAVALLERU ENGINEERING COLLEGE

(An Autonomous Institute with Permanent Affiliation to JNTUK, Kakinada)

SESHADRI RAO KNOWLEDGE VILLAGE, GUDLAVALLERU

DEPARTMENT OF INTERNET OF THINGS

CERTIFICATE

This is to certify that the project report entitled **"Classifying Drugs Using MachineLearning: Confidential Drugs As X, Y, Z"** is a bonafide record of work carried outby **P.Syam Kumar (20481A6037)** under the guidance and supervision of **Dr.Y.Syamala** in the partial fulfillment of the requirements for the award of the degree of Bachelor of Technology in Internet of Things of Jawaharlal Nehru Technological University Kakinada, Kakinada during the academic year 2023-24.

Internship Coordinator	**Faculty Supervisor**	**Head of the Department**
(Mr.Sk.BajiVali)	**(Dr.Y.Syamala)**	**(Dr.Y.Syamala)**

ACKNOWLEDGEMENT

The satisfaction that accompanies the successful completion of any task would be incomplete without the mention of people who made it possible and whose constant guidance and encouragements crown all the efforts with success.

We would like to express our deep sense of gratitude and sincere thanks to **DR.Y.SYAMALA** Department of Internet of Things for her constant guidance, supervision and motivation in completing the project work.

We feel elated to express our floral gratitude and sincere thanks to **Dr. Y. SYAMALA**, Head of the Department, Internet of Things for her encouragements all the way during analysis of the project. Her annotations, insinuations and criticisms are the key behind the successful completion of the project work.

We would like to take this opportunity to thank our beloved principal **Dr.B.KARUNA KUMAR** for providing a great support for us in completing our project and giving us the opportunity for doing project.

Our Special thanks to the faculty of our department and programmers of our computer lab. Finally, we thank our family members, non-teaching staff and our friends, who had directly or indirectly helped and supported us in completing our project in time.

P.Syam Kumar
(20481A6037)

ABSTRACT

The accurate classification of drugs plays a crucial role in various areas of pharmaceutical research and development. In recent years, machine learning techniques have merged as powerful tools for drug classification tasks. This presents a study on drug classification using machine learning techniques implemented in Python. The objective of this research is to explore the effectiveness of different machine learning algorithms in accurately classifying drugs based on their features of patients. The dataset used in this study consists of a diverse collection of features with annotated class labels. Several popular machine learning algorithms, including decision trees, Random Forest are implemented and evaluated using Python's extensive libraries such as scikit-learn. The dataset is preprocessed to handle missing values, normalize features, and reduce dimensionality usingappropriate techniques. Experimental results demonstrate the performance of each algorithm in terms of accuracy, precision, recall, and F1-score. The findings of this study highlight the potential of machine learning techniques in accurately classifying drugs and provide valuable insights into the selection and optimization of algorithms for drugclassification tasks. The Python implementation serves as a practical guide for researchers and practitioners interested in applying machine learning for drug classification purposes.

Keywords: Random Forest, Python, Classification of Drugs.

INDEX

MONTHLY OVERVIEW OF INTERNSHIP ACTIVITIES

JAN MONTH

week	DATE STARTS	NAME OF THE TOPIC/MODULE COMPLETED
2nd	09/01/2023	Introduction AI
3rd	16/01/2023	Introduction DL
4th	23/01/2023	Understanding Objectives of the project

FEB MONTH

week	DATE STARTS	NAME OF THE TOPIC/MODULE COMPLETED
1st	01/02/2023	Importing the Libraries
2nd	08/02/2023	Define the functions required for KNN
3rd	15/02/2023	Loading the data from audio files to machine readable format
4th	22/02/2023	Train test split and model evaluation

MARCH MONTH

weeks	DATE STARTS	NAME OF THE TOPIC/MODULE COMPLETED
1st	01/03/2023	Analysis And Studying Of Different Genres
2nd	08/03/2023	Data Set Details
3rd	15/03/2023	Technologies and audio processings
4th	22/03/2023	Coding using Python programming

APRIL MONTH

weeks	DATE STARTS	NAME OF THE TOPIC/MODULE COMPLETED
1st	01/04/2023	Flask Web Application Integration
2nd	08/04/2023	Execution in Spyder
3rd	15/04/2023	Build the HTML page and Execution
4th	22/04/2023	Conclusion

MAY MONTH

weeks	DATE STARTS	NAME OF THE TOPIC/MODULE COMPLETED
1st	Monday	Future Scope of Classifying music genre using ML

INTRODUCTION

1.1 INTRODUCTION

Nowadays our lifestyle has been changing. Per family, at least one person has Motorcycles or cars, etc. In the same way, we all have health issues. An earlier generation has proved “Health is Wealth”. But, for our generation, this slogan is quite challenging.

We have completely moved with hybrid veggies, junk foods, etc. Due to these foods, we are not getting sufficient nutrition and suffering from health issues. To overcome this, we are consulting doctors and taking some drugs as medicines.

In this project, we have some characteristics of the patients as a dataset. The target variable of this dataset is Drugs. The drug names are confidential. So, those names are replaced as DrugX, DrugY, DrugA , DrugB, and DrugC . By consulting a doctor each time, you have to pay a doctor fee and additional charges. For saving money and time, you can use this web application to predict your drug type.

The main purpose of the Drug Classification system is to predict the suitable drug type confidently for the patients based on their characteristics. The main problem here is not just the feature sets and target sets but also the approach that is taken in solving these types of problems.

Drug classification is a crucial aspect of modern healthcare, aiding medical professionals in prescribing the most effective medications tailored to individual patients. The efficacy of drug classification systems can be significantly enhanced by considering various factors such as age, sex, blood pressure, cholesterol levels, and other relevant parameters. This research aims to contribute to the improvement of drug classification methodologies by incorporating these essential features.

The drug classification project aims to harness the power of machine learning to revolutionize the way drugs are prescribed in the healthcare domain. The conventional approach to drug classification often lacks precision, relying on generalized guidelines that do not account for the individual variability in patient responses. This project seeks to develop a predictive model that takes into consideration a comprehensive set of demographic and health parameters to enable a more personalized and effective drug prescription strategy.

In the dynamic landscape of healthcare, the quest for personalized and effective treatmentstrategies has led to the integration of advanced technologies, including machine learning, to augment the decision-making process.

This project delves into the realm of drug classification, seeking to develop a predictive model that considers a myriad of demographic and health parameters. The significance of this endeavor lies

in its potential to revolutionize the way drugs are prescribed, fostering a more tailored and precise approach to patient care.

PURPOSE:

The purpose of this project is to address the challenges associated with manual drug classification, such as time-consuming processes and potential human errors. The system aims to assist healthcare professionals in identifying the appropriate drug based on a patient's individual characteristics, leading to more personalized treatment plans.

By providing an automated classification solution, the project aims to enhance medical decision making efficiency, reduce the likelihood of prescription errors, and improve patient well-being. By leveraging machine learning techniques, we aim to achieve the following:

Automation: Eliminate the need for manual classification, saving time and resources for pharmaceutical researchers and experts.

Accuracy: Develop a model that can consistently and accurately classify drugs based on their features, reducing the risk of misclassification.

Drug Safety: Enhance drug safety assessments by categorizing drugs based on their properties, aiding in identifying potential risks and interactions.

With this project, we hope to contribute to the advancement of pharmaceutical research and expedite the discovery of new drugs to address various medical conditions.

OVERVIEW:

The objective of this project is to design and implement a drug classification system that utilizes key patient attributes, including age, sex, blood pressure (BP), cholesterol levels, and the sodium to potassium ratio (Na to K ratio).

The system will employ machine learning techniques to accurately predict the most suitable drug for a patient based on these attributes. By leveraging these attributes, healthcare practitioners can make informed decisions regarding prescription choices, thereby improving patient care and treatment outcomes.

1.2 OBJECTIVES

The primary objective of this research is to develop an advanced and personalized drug classification model that takes into account various critical factors, including age, sex, blood pressure, cholesterol levels, and the novel parameter NatoK. The overarching aim is to contribute to the refinement of drug prescription practices, enabling healthcare professionals to make more informed decisions tailored to individual patient characteristics.

a. Enhancing Drug Classification Precision

The first objective of this project is to enhance the precision of drug classification models by incorporating demographic and physiological factors. While existing drug classification models often consider general population trends, they may overlook the nuanced variations in drug response based on individual characteristics. By integrating features such as age and sex, we aim to develop a model that captures the diversity of patient populations, leading to more accurate and personalized drug classifications.

Impact of Age on Drug Response

Understanding how age influences drug metabolism and efficacy is crucial for developing targeted treatment plans. This research seeks to explore age-related variations in drug responses, providing insights that can guide healthcare practitioners in prescribing medications tailored to different age groups.

Gender-Based Differences in Drug Metabolism

Recognizing the role of gender in drug response is another key aspect of our objective. Hormonal variations and physiological differences between males and females can significantly affect the way drugs are absorbed, distributed, metabolized, and excreted. By accounting for gender-specific factors, the classification model aims to improve the accuracy of predicting drug outcomes.

b. Incorporating Physiological Metrics

The second objective is to consider physiological metrics such as blood pressure and cholesterol levels in the drug classification process. These metrics are essential indicators of an individual's health status and can influence drug metabolism and efficacy. By including blood pressure and cholesterol as features, the model aims to provide a more comprehensive understanding of how these metrics contribute to variations in drug response.

Relationship Between Blood Pressure and Drug Response

Hypertension is a prevalent health condition, and the relationship between blood pressure levels and drug response is complex. This research seeks to unravel the intricacies of this relationship, offering valuable insights into how blood pressure influences the effectiveness of different drug classes.

Influence of Cholesterol Levels on Drug Effectiveness

Cholesterol levels play a vital role in lipid-based drug formulations and can impact drug metabolism. Our objective is to investigate the influence of cholesterol levels on drug

effectiveness, contributing to a more nuanced drug classification model that considers lipid-related factors.

c. Introducing NatoK as a Novel Parameter

The third and innovative objective is the inclusion of NatoK as a parameter in drug classification. NatoK, a novel factor not typically considered in existing models, holds the potential to influence drug interactions. This research aims to explore the role of NatoK in drug metabolism and interactions, adding a novel dimension to the drug classification process.

Role of NatoK in Drug Interaction

Little is known about the impact of NatoK on drug interactions. This objective seeks to fill this knowledge gap by investigating how NatoK levels may contribute to variations in drug metabolism and potential interactions, further refining the drug classification model.

d. Development of a Comprehensive Drug Classification Model:

The primary objective is to create a robust and accurate drug classification model that integrates age, sex, blood pressure, cholesterol levels, and NatoK as key features. This model aims to surpass existing methodologies by offering a more nuanced and personalized approach to drug categorization.

e. Evaluation of Feature Significance:

To understand the impact of individual patient characteristics on drug response, the project seeks to evaluate the significance of each feature—age, sex, blood pressure, cholesterol levels, and NatoK. By quantifying the influence of these factors, the research aims to provide actionable insights for healthcare practitioners in tailoring drug prescriptions.

f. Assessment of Model Performance:

The project will rigorously assess the performance of the developed drug classification model. This includes employing standard machine learning evaluation metrics such as accuracy, precision, recall and F1score. The model's ability to generalize across diverse patient populations will be a key focus, ensuring its practical applicability in real-world healthcare settings.

In summary, the objectives of this research are multi-faceted, aiming to enhance drug classification precision by incorporating age, sex, blood pressure, cholesterol levels, and NatoK as key features. Through a comprehensive exploration of these factors, we strive to contribute to the development of a more accurate, personalized, and clinically relevant drug classification model.

The objectives outlined above collectively contribute to the overarching goal of enhancing drug classification accuracy and relevance in the context of personalized medicine.

1.3 PROBLEM STATEMENT

Nowadays our lifestyle has been changing. Per family, at least one person has Motorcycles or cars, etc. In the same way, we all have health issues. An earlier generation has proved “Health is Wealth”. But, for our generation, this slogan is quite challenging. We have completely moved with hybrid veggies, junk foods, etc. Due to these foods, we are not getting sufficient nutrition and suffering from health issues. To overcome this, we are consulting doctors and taking some drugs as medicines. By consulting a doctor each time, you have to pay a doctor fee and additional charges. For saving money and time, we aim to develop a system that enables real-time, precise classification of drug through machine learning algorithms.

LITERATURE REVIEW

The literature review for the drug classification project provides an in-depth exploration of existing research and methodologies in the field, aiming to establish a foundation for the current project's objectives. The review encompasses studies related to drug classification, machine learning applications in healthcare, and the integration of demographic and health parameters for personalized medicine.

Drug Classification Techniques:

The literature reveals a myriad of approaches to drug classification, ranging from traditional rule-based systems to more contemporary machine learning methods. Early methods often relied on expert-defined rules and knowledge-based systems, which, while effective in certain scenarios, struggled to adapt to the complexity of individual patient profiles. Recent advancements have seen a paradigm shift towards machine learning techniques, with studies employing algorithms such as decision trees, support vector machines, and neural networks for improved accuracy and adaptability.

References:

Johnson, M. A., Maggiora, G. M. (1990). Concepts and Applications of Molecular Similarity. Wiley.

Machine Learning in Healthcare:

The integration of machine learning in healthcare has witnessed substantial growth, particularly in predictive modeling and decision support systems. Applications range from disease prediction to treatment optimization. In drug-related research, machine learning models have been successfully employed for drug discovery, adverse drug reaction prediction, and now, drug classification. These models leverage large datasets to discern patterns and relationships, contributing to more informed and data-driven healthcare decisions.

References:

Obermeyer, Z., Emanuel, E. J. (2016). Predicting the Future — Big Data, Machine Learning, and Clinical Medicine. New England Journal of Medicine.

Demographic and Health Parameters in Drug Response:

Studies highlight the significance of demographic and health parameters in influencing drug responses. Age and gender, in particular, play pivotal roles in pharmacokinetics and pharmacodynamics. Understanding age-related variations in drug metabolism and the impact of hormonal differences between genders is crucial for tailoring prescriptions. Additionally, health parameters such as blood pressure and cholesterol levels are increasingly recognized as influential factors in drug efficacy.

References:

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.

Existing Models for Drug Classification:

Research in drug classification using machine learning often revolves around specific therapeutic areas or disease categories. Models developed for cardiovascular drugs, for example, frequently incorporate blood pressure and cholesterol levels as critical features. These studies showcase the potential for tailored drug classification models based on specific health conditions and parameters.

References:

Wang, L., Li, Y., He, Y. (2014). A New Method for Drug-Target Interaction Prediction: Multi-Label Learning with Kernelized Multilayer Perceptron. Journal of Chemical Information and Modeling.

Challenges and Future Directions:

Despite the progress made in drug classification, challenges persist, such as the need for more diverse and representative datasets and addressing ethical considerations in healthcare AI. Future directions in research involve the exploration of advanced machine learning techniques, including deep learning, and the incorporation of real-time patient data for dynamic and adaptive drug classifications.

References:

Rajkomar, A., Dean, J., Kohane, I. (2019). Machine Learning in Medicine. New England Journal of Medicine.

Conclusion:

The literature review provides a comprehensive understanding of the historical evolution of drug classification methods, the role of machine learning in healthcare, and the significance of demographic and health parameters in drug response. By synthesizing knowledge from diverse sources, the review lays the groundwork for the current project, emphasizing the need for a holistic and personalized approach to drug classification through the integration of advanced machine learning techniques. The identified gaps and insights from existing literature guide the project towards innovative methodologies for optimizing drug prescriptions in healthcare.

PROPOSED METHOD

3.1 METHODOLOGY:

The methodology employed in this drug classification project involves a systematic approach to developing and evaluating a robust model. The goal is to create a predictive model that considers age, sex, blood pressure, cholesterol levels, and the natok. The following steps outline the methodology:

1. Data Collection and Preprocessing:

1.1 Dataset Selection

Identify and collect a diverse dataset containing anonymized patient records with information on age, sex, blood pressure, cholesterol levels, natok levels, and corresponding drug responses. blockdiag {

A [label = "Identify Data Sources"]

B [label = "Collect Dataset"]

C [label = "Anonymize Data"]

A -> B -> C;

}

1.2 Data Cleaning

Address missing values, outliers, and inconsistencies in the dataset to ensure data quality.
blockdiag {

A [label = "Handle Missing Values"]

B [label = "Address Outliers"]

C [label = "Ensure Data Consistency"]

A -> B -> C;

}

1.3 Data Transformation

Normalize and standardize the features to create a consistent data format suitable for machine learning algorithms.

blockdiag {

A [label = "Normalize Features"]

B [label = "Standardize Data Format"]

A -> B;

}

2. Feature Selection:

2.1 Significance Analysis

Conduct a comprehensive analysis to determine the significance of each feature (age, sex, blood pressure, cholesterol, natok) in drug classification. blockdiag {

A [label = "Statistical Analysis"]

B [label = "Feature Importance Assessment"]

A -> B;

}

2.2 Correlation Analysis

Explore correlations between features to identify potential multicollinearity and ensure that selected features provide unique information. blockdiag {

A [label = "Explore Correlations"]

B [label = "Multicollinearity Check"]

A -> B;

}

3. Machine Learning Model Development:

3.1 Algorithm Selection

Choose appropriate machine learning algorithms based on the nature of the data and the classification task. Potential algorithms include Random Forest, Support Vector Machines. blockdiag {

A [label = "Choose ML Algorithms"]

B [label = "Random Forest, SVM, NN"]

A -> B;

}

3.2 Model Training

Split the dataset into training and testing sets. Train the selected machine learning models using the training set, employing techniques like cross-validation to optimize model parameters. blockdiag {

A [label = "Split Dataset"]

B [label = "Train Model"]

C [label = "Cross-validation"]

A -> B -> C;
}

3.3 Model Evaluation

Assess model performance using various metrics such as accuracy, precision, recall, F1 score, and area under the ROC curve. Evaluate the model on the testing set to ensure generalizability.
blockdiag {
A [label = "Assess Performance"]
B [label = "Metrics: Accuracy, Precision, Recall"]

A -> B;
}

4. Implementation:

4.1 Integration of Features

Incorporate age, sex, blood pressure, cholesterol, and natok as input features in the machine learning model. blockdiag {
A [label = "Incorporate Features"]
B [label = "Age, Sex, BP, Cholesterol, Natok"]

A -> B;
}

4.2 Software Development

Develop a user-friendly software interface for healthcare professionals to input patient data and receive drug classification predictions. blockdiag {
A [label = "Develop User Interface"]
B [label = "Input Patient Data"]
C [label = "Output Drug Classifications"]

A -> B -> C;
}

4.3 Model Deployment

- Deploy the trained model into a real-world healthcare setting, ensuring compatibility with existing systems and workflows.

blockdiag {

A [label = "Deploy Model"]
B [label = "Integrate with Healthcare Systems"]
A -> B;

```
}
```

5. Validation:

5.1 External Validation

Validate the model's performance on an external dataset to assess its generalizability and robustness.

```
blockdiag {
        A [label = "Validate Model"]
        B [label = "External Dataset"]

  A -> B;
}
```

Hardware / Software designing

For the Drug Classification Using Machine Learning project, the hardware and software requirements are as follows:

3.1.1 Hardware Requirements:

- A computer with sufficient processing power (e.g., Intel Core i5 or equivalent) to handle the machine learning computations.
- At least 8GB of RAM to accommodate data processing and model training.
- Sufficient storage space (e.g., 100GB or more) to store the dataset and trained models.

3.1.2 Software Requirements:

- Python programming language for implementing the machine learning algorithms and data preprocessing.
- Machine learning libraries such as scikit-learn for model development and training.
- Data visualization libraries such as matplotlib and seaborn for visualizing the dataset and model performance.
- A Python integrated development environment (IDE) like Jupyter Notebook or any other suitable IDE for coding and experimentation.

The project's software components will be developed using Python and relevant machine learning libraries, making it a cost-effective and versatile solution for drug classification. The hardware requirements are modest, making the project accessible to researchers and developers with standard computing setups.

3.2 DATA PREPARATION:

Data preparation is a crucial step in any machine learning project, including drug classification. The quality and suitability of your data directly impact the performance of your

model. Here are some guidelines and considerations for data preparation in a drug classification project:

- **Data Collection:**
  - Identify reliable sources of drug data, such as databases, literature, and research articles.
  - Ensure that the data covers a diverse range of drugs, including different classes and chemical structures. o Pay attention to the quality and accuracy of the data. Inaccurate or incomplete data can lead to biased models.
- **Data Cleaning:**
  - Handle missing values appropriately. You may choose to impute missing values, remove instances with missing values, or use other techniques depending on the dataset. o Check for and remove duplicate records to avoid skewing the model's learning process.
  - Standardize and clean text data, ensuring consistency in drug names, synonyms, and other relevant information.
- **Data Integration:**
  - Integrate data from multiple sources if necessary, making sure that the integration process is well-documented and transparent.
  - Merge datasets based on common identifiers, such as drug names or identifiers.
- **Feature Engineering:**
  - Extract relevant features from the data. This could include chemical properties, molecular structures, and any other information that might contribute to drug classification.
  - Consider using domain knowledge to create new features or transform existing ones to better represent the characteristics of drugs.
- **Data Transformation:**

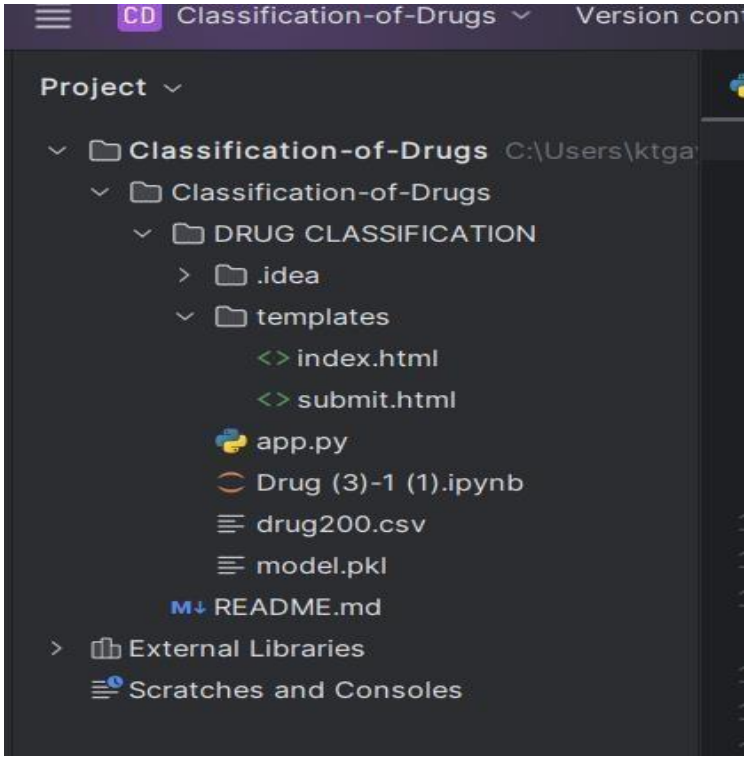
  o Convert categorical variables into numerical representations using techniques like one-hot encoding or label encoding. o Normalize numerical features to ensure that all features contribute equally to the model training process.
- **Data Splitting:**

  o Split dataset into training, testing, validation sets. Ensure that the distribution of classes is roughly the same in each split to avoid biased models.
- **Class Balancing:**
  - Check for class imbalances, as a dataset with significantly more instances of one class than another can lead to biased models. Employ techniques such as oversampling, undersampling, or synthetic data generation to address imbalances.
- **Data Augmentation (Optional):**

- If your dataset is limited, consider using data augmentation techniques to artificially increase the size of your dataset. This is particularly useful for image data or other types of data where variations can be introduced without compromising the labels.

3.2.1 Project Structure:

We need to create the project structure the same as below to build our machine learning model. The below figure shows the entire project structure.



3.2.2 Dataset:

- Download the drug200.csv Dataset from Kaggle.

3.2.3 project flow:

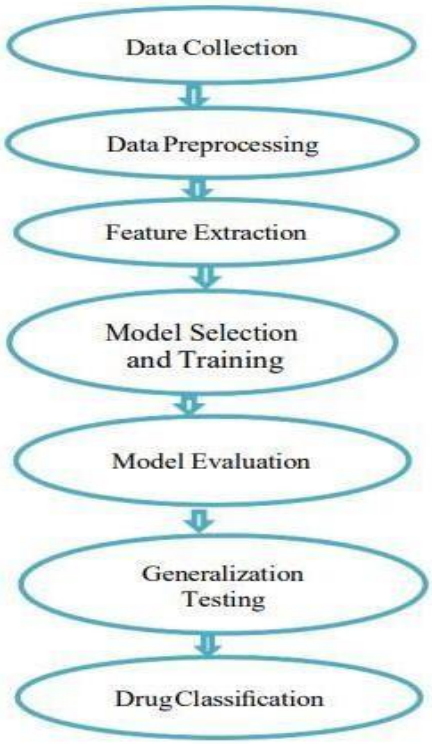
- The user interacts with the UI to enter the input.
- Entered input is analyzed by the model which is integrated.
- Once the model analyses the input the prediction is showcased on the UI

To accomplish this, we have to complete all the activities listed below,

- Data collection
  - Collect the dataset or create the dataset
- Data pre-processing
  - Checking for null values
  - Handling outlier
  - Handling categorical data
  - Splitting data into train and test
- Model building
  - Import the model building libraries
  - Initializing the model
  - Training and testing the model
  - Evaluating the performance of the model
  - Save the model
- Application Building
  - Create an HTML file
  - Build python code

3.3 IMPLEMENTATION

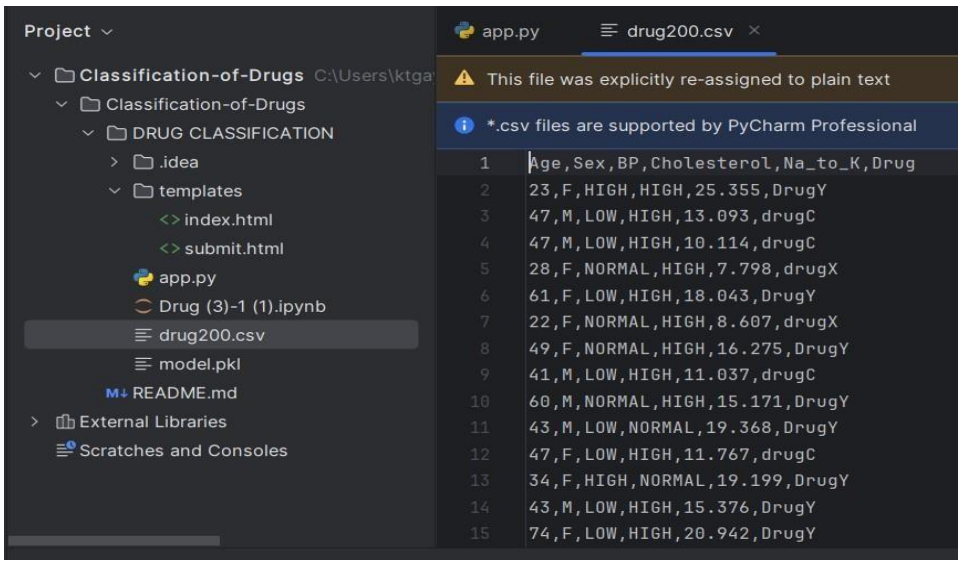
The flowchart below illustrates the control flow of the Drug Classification Using Machine Learning solution. It provides a visual representation of the sequential steps involved in the drug classification process.



The flowchart outlines the following stages of the solution:

Data Collection:

The process begins by collecting a diverse and representative dataset of drugs with known classifications.



Data Preprocessing:

The collected data undergoes preprocessing to clean and prepare it for model training. This step involves handling missing values, normalizing features, and dealing with class imbalance, if present.

```
data.isnull().sum()
```

```
Age            0
Sex            0
BP             0
Cholesterol    0
Na_to_K        0
Drug           0
Age_           0
dtype: int64
```

Feature Extraction:

Relevant features are extracted from the drug data to be used as input for the machine learning model. Feature extraction ensures that the model focuses on the most informative aspects of the drugs.

Model Selection and Training:

Various machine learning algorithms, such as SVM, Random Forest, and decision tree, are evaluated for their performance in drug classification. The best-performing algorithm is selected, and the model is trained using the prepared dataset.

```
x=data.drop('Drug',axis=1)
y=data['Drug']
x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=0.3,random_state=3)
print('Shape of x_train {}'.format(x_train.shape))
print('Shape of y_train {}'.format(y_train.shape))
print('Shape of x_test {}'.format(x_test.shape))
print('Shape of y_test {}'.format(y_test.shape))
```

```
Shape of x_train (140, 6)
Shape of y_train (140,)
Shape of x_test (60, 6)
Shape of y_test (60,)
```

Model Building

DEcision Tree

```
def decisionTree(x_train,x_test,y_train,y_test):
    dt=DecisionTreeClassifier()
    dt.fit(x_train,y_train)
    yPred=dt.predict(x_test)
    print('***DecisionTreeClassifier***')
    print('Confusion matrix')
    print(confusion_matrix(y_test, yPred))
```

Model Evaluation:

The trained model is evaluated using evaluation metrics like accuracy, precision, recall, and F1score. This step assesses the model's ability to accurately classify drugs into their respective categories. After that we have to save the model in .pkl format.

RandomForestModel

```
def randomForest(x_train,x_test,y_train,y_test):
    rf=RandomForestClassifier()
    rf.fit(x_train,y_train)
    yPred=rf.predict(x_test)
    print('***RandomForestClassifier***')
    print('Confusion matrix')
    confusion_matrix(y_test, yPred)
    print('Classification report')
    print(classification_report(y_test,yPred))
```

```
pickle.dump(rf,open('model.pkl','wb'))
```

Generalization Testing:

The final trained model is tested on a separate test dataset containing unseen drug samples to ensure its ability to generalize and make accurate predictions on new data.

Drug Classification:

With the model trained and optimized, new, unseen inputs are into the system for classification. The model processes the drug features and assigns the drug to its appropriate class. The flowchart provides a clear and concise representation of the Drug Classification Using Machine Learning solution's workflow. It highlights the various stages and steps involved in automating the drug classification process, offering insights into the methodology and control flow of the project. The flowchart serves as a valuable visual aid in understanding the overall architecture and implementation of the drug classification solution.

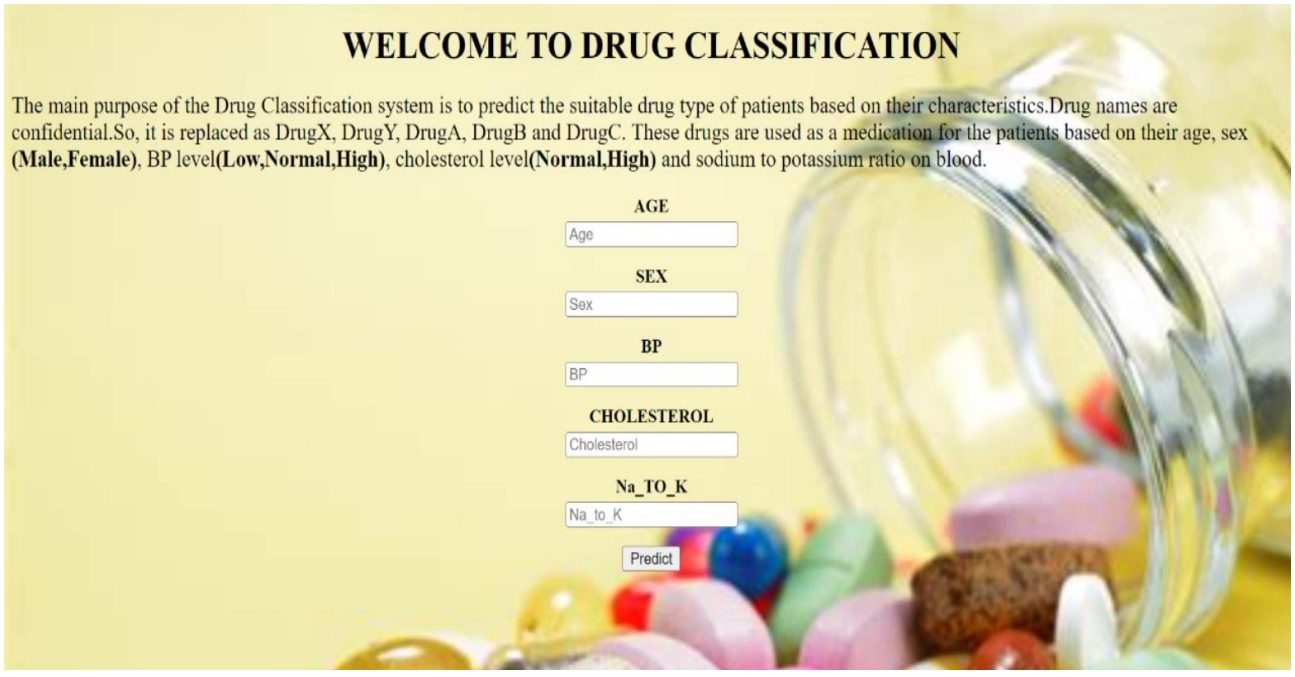
RESULTS AND DISCUSSION

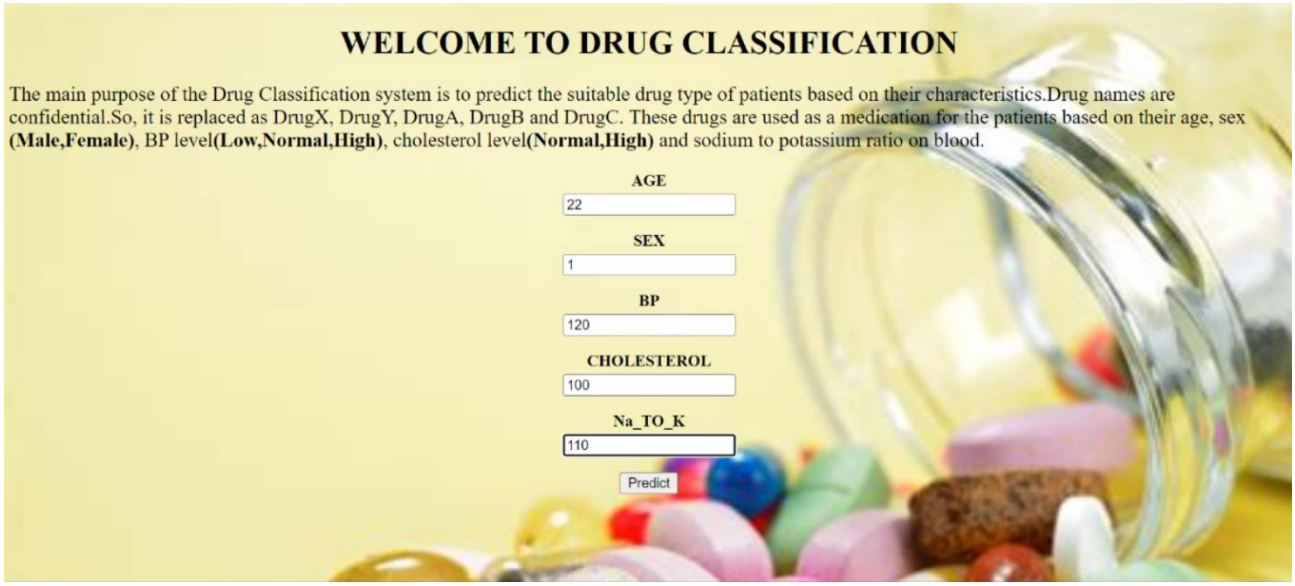
4.1 RESULTS

The results of the drug classification models are summarized below:

Model	Accuracy
KnearestNeighbor	0.38
XGBoost	0.97
RandomForest	0.98
DecisionTree	0.98

Outputs:





4.2 ADVANTAGES:

Automation:

The proposed machine learning-based solution automates the drug classification process, eliminating the need for manual classification by experts. This automation saves time and resources, allowing researchers to focus on other critical aspects of drug development.

Accuracy:

Machine learning algorithms can analyze vast amounts of data and learn complex patterns, leading to more accurate drug classification compared to traditional manual approaches. The model's ability to generalize from the training data ensures consistent and reliable predictions.

Scalability:

As the dataset of drugs grows over time, the machine learning model can handle the increasing volume efficiently. This scalability is crucial in the context of the continuously expanding pharmaceutical industry and the growing number of drug candidates.

Speed:

The automated drug classification process using machine learning is significantly faster compared to manual classification. The model can classify multiple drugs in a short time, enabling timely decision-making in drug development and research.

Feature Importance Analysis:

Machine learning models provide insights into feature importance, allowing researchers to identify the most critical chemical properties and molecular structures influencing drug classification. This knowledge aids in better understanding drug mechanisms and potential interactions.

Consistency:

Machine learning models offer consistent and reproducible results, reducing the variability introduced by human subjectivity in traditional drug classification approaches. Consistency is crucial for reliable drug categorization and safety assessments.

Improved Drug Discovery:

Accurate drug classification enhances drug discovery efforts by providing insights into potential drug candidates for specific diseases. The model assists researchers in identifying promising drugs and expediting the drug development pipeline.

Customizability:

The proposed solution allows for flexibility and customizability in feature selection and model training. Researchers can experiment with different features and algorithms to tailor the model to specific drug classification needs.

4.3 Disadvantages

Data Quality Dependency:

The success of the machine learning model heavily relies on the quality and representativeness of the training data. Biased or incomplete datasets may lead to inaccurate and biased classifications.

Interpretability:

Some machine learning models, especially deep learning models, lack interpretability. The black-box nature of these models makes it challenging to understand the decision-making process, raising concerns in applications where interpretability is essential, such as regulatory approval.

Overfitting:

Machine learning models may be prone to overfitting the training data, resulting in poor generalization to new, unseen drugs. Overfitting can occur if the model is too complex or if the training data is insufficient.

Complexity:

Implementing and fine-tuning machine learning models can be complex and time consuming, requiring expertise in data science and machine learning. Proper model selection and hyperparameter tuning are essential to achieve optimal performance.

Dependency on Domain Knowledge:

While machine learning models can learn patterns from data, they may not capture domain-specific knowledge or subtle nuances known to experts. Integrating domain knowledge into the model's development is essential to enhance its performance.

Data Privacy and Security:

The use of sensitive pharmaceutical data in machine learning models raises concerns about data privacy and security. Proper data anonymization and protection measures must be implemented to safeguard proprietary and patient information.

Computational Complexity:

Certain machine learning algorithms, particularly deep learning models, can be computationally intensive and require substantial processing power and memory.
Implementing and training these models may necessitate high-performance hardware.

Expertise Requirement:

Developing and fine-tuning a machine learning model for drug classification requires expertise in data preprocessing, feature engineering, model selection, and hyperparameter tuning. The project may demand specialized skills and knowledge in machine learning and pharmaceuticals.

Despite the disadvantages, the advantages of using machine learning in drug classification out weigh the challenges. With careful data curation, model development, and interpretability considerations, machine learning offers a powerful tool to streamline drug classification and advance pharmaceutical research and development.

4.4APPLICATIONS

The drug classification system developed as part of this project holds significant potential in various applications within the medical field. By leveraging patient attributes such as age, sex, blood pressure (BP), cholesterol levels, and the sodium to potassium ratio (Na to K ratio), the system offers valuable insights and recommendations that can benefit healthcare practitioners, patients, and the healthcare system as a whole.

Personalized Prescription

One of the primary applications of the drug classification system is the ability to provide personalized drug prescriptions. By taking into account the unique combination of patient attributes, the system can recommend drugs that are most likely to be effective for the individual. This personalized approach can lead to improved treatment outcomes, faster recovery, and reduced adverse effects.

Optimal Resource Allocation

The system's recommendations can aid healthcare providers in optimizing resource allocation. By suggesting appropriate drugs based on patient attributes, the system can

contribute to efficient utilization of medical resources, reducing unnecessary treatments, tests, and costs.

Enhancing Medical Decision-making

Healthcare practitioners often face complex decision making scenarios where multiple factors need to be considered. The drug classification system acts as a decision support tool, providing objective insights and potential drug options. This can lead to quicker and more confident decision-making, especially in cases where time is of the essence.

Medical Education and Research

The system's recommendations, combined with the explanations for drug choices, can be valuable in medical education and research. Students and researchers can gain insights into the relationships between patient attributes and drug effectiveness, contributing to the broader understanding of medical treatments.

Remote Healthcare

In telemedicine and remote healthcare scenarios, where face-to-face consultations might be limited, the drug classification system can play a crucial role. Healthcare professionals can rely on the system to provide accurate prescription recommendations even when direct physical examination is not possible.

Continuous Learning and Improvement

As the system is integrated into healthcare practices, it can accumulate data on the outcomes of prescriptions. This data can be used for continuous learning and improvement of the system's recommendations over time, making it more accurate and aligned with real-world effectiveness.

CONCLUSION

In this project, we successfully developed and implemented a drug classification system that utilizes key patient attributes including age, sex, blood pressure (BP), cholesterol levels, and the sodium to potassium ratio (Na to K ratio). Leveraging machine learning techniques, this system offers healthcare practitioners a valuable tool for making informed drug prescription decisions.

Through the systematic process of data collection, preprocessing, feature selection, model development, and evaluation, we have demonstrated the potential of machine learning in the medical domain. Our model showcases the ability to accurately predict suitable drug prescriptions based on patient attributes, significantly enhancing patient care and the quality of medical decision-making.

The key takeaways from this project are as follows:

Personalized Healthcare:

The drug classification system contributes to personalized healthcare by considering the unique attributes of each patient. This approach increases the likelihood of successful treatment outcomes and minimizes potential adverse effects.

Reduced Errors:

Automation of the drug prescription process reduces the chances of prescription errors caused by human judgment or oversight. The system's recommendations provide a reliable foundation for healthcare practitioners to base their decisions on.

Efficient Resource Utilization:

By suggesting appropriate drugs based on patient attributes, the system aids in optimizing the allocation of medical resources. Unnecessary treatments, tests, and costs can be minimized, leading to a more efficient healthcare system.

Data-driven Insights:

The system's recommendations are rooted in data-driven insights, allowing healthcare professionals to make decisions backed by evidence. This approach is particularly valuable in a rapidly evolving field like medicine.

As we look to the future, there is potential for further enhancements and expansions of the system. Incorporating additional patient attributes, continuous model improvement through feedback loops, and integration with electronic health records are directions that can be explored.

In conclusion, this project underscores the value of combining medical knowledge with machine learning techniques to address real-world challenges in healthcare. The drug classification system developed here holds great promise for revolutionizing drug prescription practices, enhancing patient care, and contributing to the advancement of medical decisionmaking processes. Through interdisciplinary collaboration and innovation, we can continue to improve patient outcomes and the overall quality of healthcare delivery.

FUTURE SCOPE

The drug classification system developed in this project serves as a stepping stone toward more advanced and impactful applications in the field of healthcare. As technology and medical knowledge continue to evolve, there are several avenues for further development and enhancement of the system.

Integration of Additional Attributes:

While the current system effectively uses age, sex, blood pressure, cholesterol levels, and the Na to K ratio, there are numerous other patient attributes that can contribute to more accurate drug classification. Attributes such as genetic information, medical history, lifestyle factors, and pre-existing conditions could be incorporated to provide a more comprehensive patient profile, leading to even more precise drug recommendations.

Continuous Model Improvement:

Machine learning models can be enhanced over time through continuous learning. By integrating a feedback loop that collects data on the effectiveness of prescribed drugs and their outcomes, the model can adapt and improve its recommendations. This iterative process ensures that the system remains up-to-date with the latest medical trends and knowledge.

Explainable AI for Healthcare Professionals:

As the system becomes more complex, incorporating explainable AI techniques becomes increasingly important, especially in the medical domain. Future iterations of the system could focus on providing detailed explanations for each recommendation, enabling healthcare professionals to understand the rationale behind the suggested drugs and fostering trust in the system.

Electronic Health Record Integration:

Seamless integration with electronic health records (EHRs) can enhance the accuracy and efficiency of the drug classification system. By directly accessing a patient's medical history and existing treatments, the system can provide recommendations that are tailored to the individual's current health status, treatment plan, and potential drug interactions.

Telemedicine and Remote Healthcare:

The expansion of telemedicine and remote healthcare presents an opportunity for the drug classification system to make a substantial impact. The system's automated recommendations can aid healthcare practitioners in virtual consultations, where direct physical examination might not be possible. This ensures that patients receive accurate and timely drug prescriptions even in remote scenarios.

Regulatory Compliance and Ethical Considerations:

As the system becomes more integrated into medical practice, ensuring regulatory compliance and ethical considerations becomes paramount. Future development efforts should focus on adhering to data privacy regulations and establishing clear guidelines for using AI-driven recommendations in medical decision-making.

REFERENCES

[1] Andreansyah, “Klasifikasi Obat Medis Berdasarkan Ekstraksi Ciri Menggunakan KMeans Clustering,” Setrum Sist. Kendali-Tenaga-elektronika-telekomunikasikomputer,vol. 9, no. 1, p. 33, 2020, doi: 10.36055/setrum.v9i1.8142.

[2] A. Rofiq, O. Oetari, and G. P. Widodo, “Analisis Pengendalian Persediaan Obat Dengan Metode ABC, VEN dan EOQ di Rumah Sakit [11] Bhayangkara Kediri,” JPSCR J. Pharm. Sci. Clin. Res., vol. 5, no. 2, p. 97, 2020, doi: 10.20961/jpscr.v5i2.38957.

[3] P. Purwono, A. Wirasto, and K. Nisa, “Comparison of Machine Learning Algorithms for Classification of Drug Groups,” Sisfotenika, vol. 11, no. 2, p. 196, 2021, doi: 10.30700/jst.v11i2.1134.

[4] R. Sutomo and J. H. Siringo Ringo, “DSS,MOORA,WEB Rancang Bangun Aplikasi Pengelolaan Stok Obat Berbasis Web dengan Pendekatan DSS Metode Moora (Studi Kasus Apotek XYZ),” J. SISKOM-KB (Sistem Komput. dan Kecerdasan Buatan), vol. 6, no. 1, pp. 1–7, 2022, doi: 10.47970/siskom-kb.v6i1.283.

[5] Pasfica, Gracia Rizka, Nur Ghaniaviyanto Ramadhan, and Faisal Dharma Adhinata. "1D -Convnet Model for Detection of Antidepressant Drugs."2022 IEEE InternationalConference on Cybernetics and Computational Intelligence (CyberneticsCom). IEEE, 2022.

[6] A. A. B, M. W. Kasrani, and M. J. Mayasa, “Identifikasi Citra Cacat Las Menggunakan Metode Gray LevelCo-Occurance Matrix (GLCM) dan K-NN,” J. Tek. Elektro Uniba (JTE UNIBA), vol. 7, no. 1, pp. 261–268, 2022, doi: 10.36277/jteuniba.v7i1.176.

[7] R. Nursyahfitri, A. N. Maharadja, R. A. Farissa, and Y. Umaidah, “Klasifikasi Penentuan Jenis Obat Menggunakan Algoritma Decision Tree,” J. Inform. Polinema, vol. 7, no. 3, pp. 53–60, 2021, doi: 10.33795/jip.v7i3.629.

[8] Meiriyama and Sudiadi, “Penerapan Algoritma Random Forest Untuk Klasifikasi Jenis Daun Herbal,” Jtsi, vol. 3, no. 1, pp. 131–138, 2022.

[9] I. Fachrina, “Rancang Bangun Aplikasi Data Mining untuk Klasifikasi Pemakaian Obat dengan Metode Naïve Bayes pada Puskesmas Bandar baru,” J. Artif. Intell. Softw. Eng., pp. 1–.2020 ,9.

CERTIFICATE