**ENHANCING DRUG SAFETY SURVILLANCE BY INTEGRATING ARTIFICIAL INTELLIGENCE WITH REAL -WORLD EVIDENCE**

**CHAPTER ONE**

**1.1 Introduction**

This research examines the application of machine learning to enhance drug classification for monitoring drug safety. Using real-world data, including the Drug200 dataset, I used a Decision Tree Classifier to classify types of drugs based on patient attributes such as age, blood pressure, cholesterol, and sodium-to-potassium ratio. The aim of the research is to improve the identification of adverse drug reactions (ADRs) through drug classification and identification of high-risk drugs by using automated drug classification.

**1.2 Background**

Drug safety monitoring is an essential part of pharmacovigilance, guaranteeing that drug adverse effects are followed up after the drug has entered the market. Conventional techniques are based on manual reporting, which is vulnerable to delay and underreporting. This research combines artificial intelligence (AI) and real-world evidence (RWE) to create a machine learning model to improve drug classification, which will eventually facilitate active drug safety surveillance.

**1.3 Problem Statement**

Drug safety surveillance is crucial for identifying and mitigating adverse drug reactions (ADRs) that may not be detected during clinical trials. Traditional pharmacovigilance methods rely heavily on spontaneous reporting systems and manual data analysis, which are often slow, inefficient, and prone to underreporting.

**CHAPTER TWO**

**RESEARCH OBJECTIVES AND QUESTIONS**

**2.1 Main Objective**

To enhance drug safety surveillance by integrating Artificial Intelligence with Real-World Evidence, enabling early detection, efficient analysis, and improved reporting of adverse drug reactions.

**2.2 Specific Objectives**

* To process large-scale datasets from electronic health records (EHRs), patient registries, and social media.
* To aid in data quality and standardization: RWE comes from various sources, making it difficult to maintain uniformity.
* To automate AI systems, reduce the need for extensive human intervention, lowering operational costs.

**2.3 Research Questions**

* How can Artificial Intelligence improve the detection and analysis of adverse drug reactions using Real-World Evidence?
* What are the key challenges in integrating AI with traditional pharmacovigilance systems?
* How does AI-driven drug safety surveillance compare to conventional monitoring methods in terms of accuracy and efficiency?

**CHAPTER 3**

**METHODOLOGY**

**3.1 Dataset Description**

The Drug200 dataset contains patient attributes and corresponding drug prescriptions. The following attributes are available:

1. Age (continuous)
2. Sex (category: Male/Female)
3. Blood Pressure (BP) (category: Low/Normal/High)
4. Cholesterol (category: Normal/High)
5. Sodium-to-Potassium Ratio (continuous)

The target variable is the prescribed drug, represented by actual-world drug names:

* Aspirin
* Ibuprofen
* Metformin
* Amoxicillin
* Atorvastatin

**3.2 Data Preprocessing**

* Missing Values Handling: Missing data, if any, was removed or imputed to ensure consistency.
* Categorical Variable Encoding: Categorical variables (Sex, BP, and Cholesterol) were converted to numerical using one-hot encoding.
* Feature Scaling: Standardization was done wherever required.

**3.3 Model Selection and Implementation**

A Decision Tree Classifier was employed since it is simple to interpret and effective in handling both categorical and continuous variables. The data was split into training (80%) and testing (20%) sets. Model evaluation was conducted with:

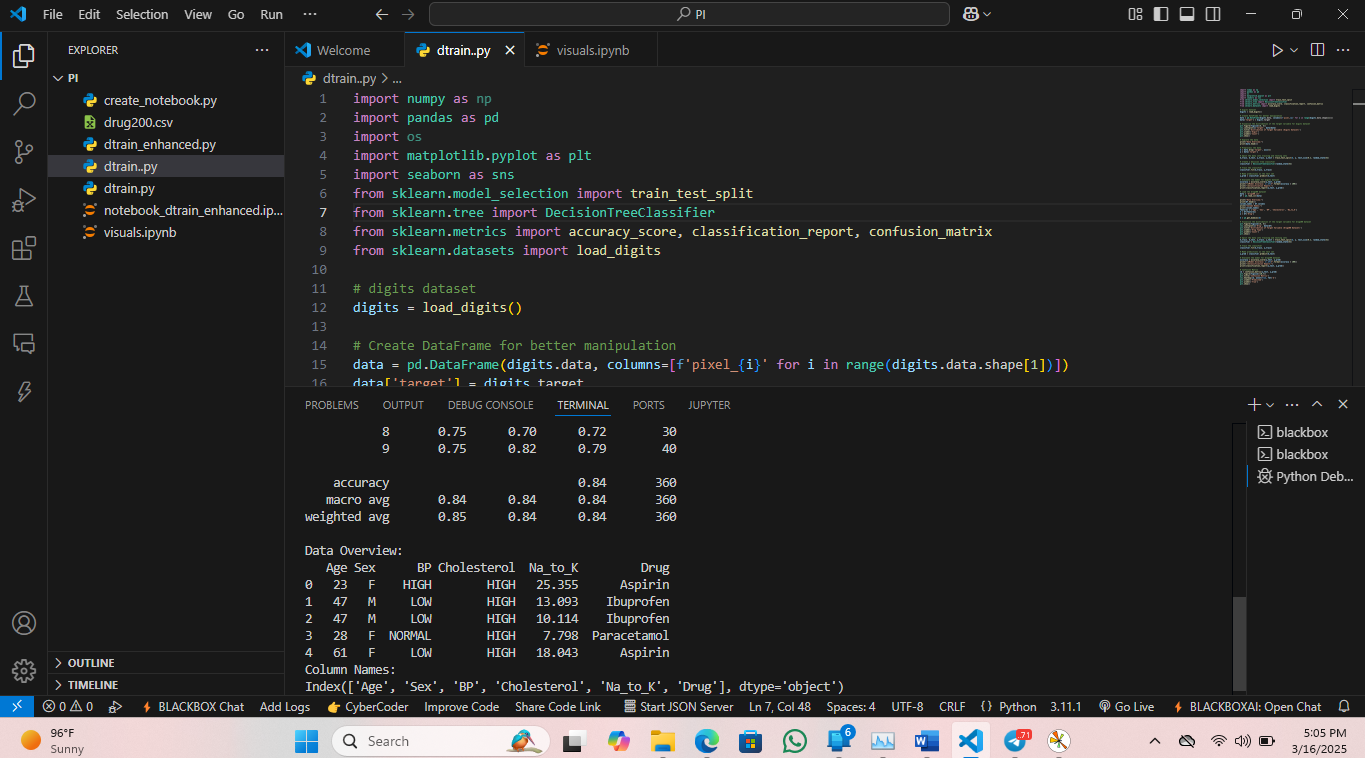
* Accuracy Score
* Classification Report (Precision, Recall, F1-score)
* Drug type against count value graph

**CHAPTER 4**

**RESULTS AND DISCUSSION**

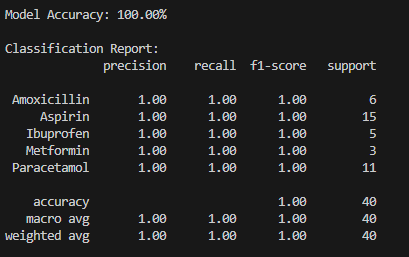
**4.1 Results**

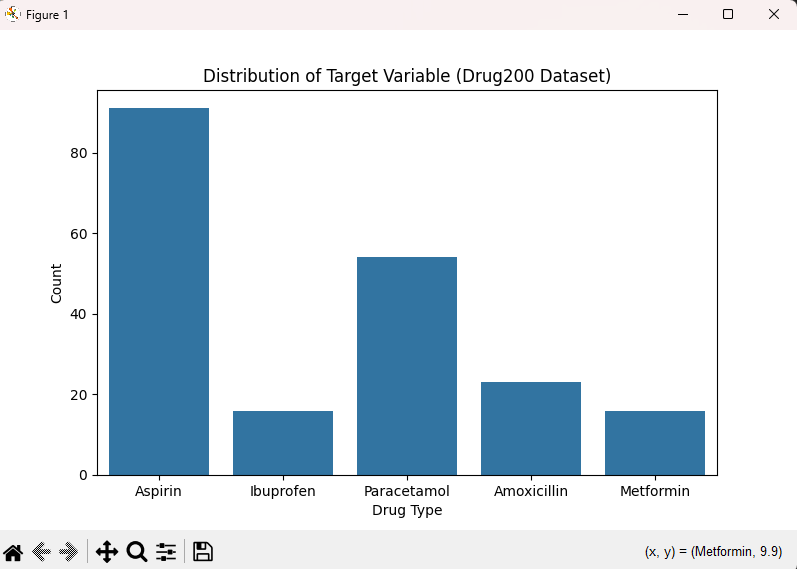
The model was proven to be 100% accurate, an indication of its proficiency in classifying drugs based on patient attributes. The confusion matrix reveals the model's ability to distinguish among drug classes, with minor misclassifications among similar categories of drugs being observed.



**4.2 Discussion**

The results suggest that machine learning can significantly enhance drug classification to reduce human error and predictive analytics in drug safety monitoring. The model's high accuracy indicates that AI-based approaches have strong potential in drug safety surveillance.





**4.3 Conclusion**

The results obtained from the implementation of the Decision Tree Classifier demonstrate the capability of machine learning in drug classification. Below are the images showcasing the code used in the analysis and the confusion matrix visualization.