Project Title

Cirrhosis Prognosis Pro: AI Driven Drug Prediction

Team Members:

- Nayera Ammar Imam (Team Leader)
- Farah Attallah
- Mohamed Mohsen

Supervised by:

Instructor: Dr. Moshira Ibrahim Ghaleb
Position: Professor of Data Science
Institution: Faculty of Computer Science, [University Name]
Department: AI & Data Science

Acknowledgment

We express our deepest appreciation to Dr. Moshira Ibrahim Ghaleb for her guidance and support throughout the project. We also thank [University Name] for providing the necessary resources to complete this research.

Table of Contents

1. Abstract

This project develops an AI-Driven Drug Recommendation System for cirrhosis patients. It uses advanced Machine Learning models to predict the most effective drug for treating cirrhosis based on patient-specific clinical data. The project integrates Data Engineering for data warehousing and processing with AI to develop predictive models that support clinicians in making data-driven decisions about liver disease treatment.

2. Introduction

2.1 Project Name:

Cirrhosis Prognosis Pro: AI-Driven Drug Recommendation for Liver Disease

2.2 Problem Statement:

Cirrhosis is a complex liver disease, and predicting the optimal drug treatment is challenging due to the heterogeneity in patient responses. Current methodologies lack precise, data-driven approaches to forecast which drug will work best for a given patient. A more accurate AI system is necessary to help healthcare providers recommend the best drug treatment for cirrhosis.

2.3 Project Idea and Objective:

The project aims to develop a system that uses advanced data engineering techniques to store and process patient data, and machine learning algorithms to predict the most suitable drug for cirrhosis patients based on their clinical profiles. This AI-driven system provides real-time insights to clinicians, enabling more personalized and effective treatment decisions.

3. Project Overview

3.1 Data Engineering Section

3.1.1 Tools and Technologies

- Tools: Python, SQL, Power BI
- Technologies: Pandas, NumPy, Matplotlib, Seaborn

3.1.2 Data Pipeline

- Ingestion of patient data
- Cleaning and preprocessing steps
- Storage in a structured database

3.1.3 Data Warehousing

- Centralized data storage for efficient access

3.1.4 Data Visualization

- Development of interactive dashboards in Power BI

3.2 Machine Learning Section

3.2.1 Tools and Technologies

- Tools: Scikit-learn, XG Boost, TensorFlow
- **Libraries:** Pandas, NumPy
- Technologies: Gradient Boosting, Support Vector Machines, Neural Networks

3.2.2 Workflow

- Data Collection
- Feature Engineering
- Model Training
- Evaluation

3.2.3 Model Selection and Training

- Comparison of various models
- Selection based on performance metrics

3.2.4 Model Evaluation and Deployment

- Rigorous testing
- Deployment in a real-time environment

4. Implementation

4.1 Data Engineering Implementation

- Detailed steps on data ingestion and preprocessing

4.2 Machine Learning Implementation

- Step-by-step guide on model training and evaluation

5. Results

5.1 Data Insights and Visualization

The Power BI dashboard provides a range of interactive insights into patient demographics, drug effectiveness, and treatment outcomes for cirrhosis patients. This dashboard helps healthcare providers visualize patterns in drug responses and evaluate treatment success rates among different patient profiles.

Visualizations Included:

- Patient Demographics: Breakdown by age, gender, and medical history.
- Drug Effectiveness: Comparative performance of different drugs.
- Treatment Success Rate: Success rates segmented by disease severity and patient group.

5.2 Model Performance and Evaluation

The machine learning models, particularly XGBoost and Random Forest, were evaluated based on accuracy, F1-score, and recall. The XGBoost model achieved the highest accuracy, indicating its effectiveness in predicting the best drug for cirrhosis treatment based on complex patient data.

Evaluation Metrics:

- Accuracy: 71% & 85.26

5.3 User Interface (UI) for Clinician Interaction

The User Interface (UI) was designed with the goal of enabling clinicians to interact seamlessly with the AI system. This interface allows clinicians to input patient data and view model recommendations in real time.

UI Features:

- Data Input Form: Clinicians can enter key patient details, such as age, gender, clinical stage, and laboratory values.
- Prediction Output: Once the data is entered, the model recommends the most effective drug with a confidence score.
- Data Visualization Links: Direct links to relevant visualizations in Power BI, helping clinicians assess treatment patterns and outcomes across similar patient profiles.

6. Conclusion

Summary of key findings and their implications.

7. Future Work

Future Work:

1. **Model Enhancements:** Further optimization of the machine learning models using larger and more diverse datasets to improve accuracy and robustness.
2. **Incorporation of Genomic Data:** Integrate genomic data to enhance personalized treatment plans for patients with genetic predispositions.
3. **Real-World Testing:** Collaborate with healthcare institutions to validate model predictions in real-world clinical settings.
4. **Expand Drug Library:** Include a wider range of drugs to cover more treatment

options and provide more comprehensive recommendations.

5. **Patient Monitoring System:** Develop a real-time patient monitoring system to update treatment recommendations based on
   ongoing clinical data.

8. Team Members

Nayera Ammar Imam: Nayera led the machine learning efforts and conceptualized the project. She collected and cleaned the dataset, worked with Python, and developed machine slearning models.

Farah Attallah: Farah focused on data warehousing and SQL queries. Her expertise ensured a robust and efficient data pipeline and storage solution.

Mohamed Mohsen: Mohamed managed the user interface, connecting the data engineering work with the machine learning models. He designed the interactive Power BI dashboard to provide clinicians with valuable insights.

9. References

- Dataset: Obtained from Kaggle and Google Calab.
- Research Literature: Sourced from Google Scholar.