Drug Recommendation System Report

Title: Drug Recommendation System Based on Side Effects and Medical Conditions

Platform Used: Jupyter lab

Presented By: Kumbha Phani Sohan

1. Introduction

This project focuses on analysing the side effects of various drugs and identifying common medical conditions associated with them. The ultimate goal is to support better and safer prescription decisions using a data-driven approach.

Objective:

- To analyse drug-related data to find links between drugs, side effects, and medical conditions.
- To build a system that recommends drugs based on effectiveness and side-effect minimization.

Why This Topic?

- Highly relevant to healthcare and patient safety.
- Supports evidence-based prescription decisions.
- Uses data to uncover patterns and improve treatment outcomes.

2. Dataset Overview

- The dataset includes:
- Drug names
- Side effects
- Medical conditions
- Drug classes
- Characteristics:
- Multiple entries per drug.
- Side effects and classes are stored as text lists.

- Actions taken:
- Required splitting, cleaning, and extraction of features for modeling.

3. Project Workflow

The project followed a structured data science workflow:

1. Data Cleaning & Preprocessing
2. Categorical Data Handling
3. Feature Engineering
4. Data Visualization & Analysis
5. Drug Recommendation Systems

4. Data Cleaning & Preprocessing

- Checked and handled missing values.
- Transformed categorical text values into structured formats:
- Split lists by semicolon (;).
- Removed unnecessary whitespaces.
- Converted to a clean DataFrame and saved as a CSV for further analysis.

5. Categorical Data Handling

- Used Label Encoding and One-Hot Encoding to convert text data into numerical format.
- Extracted drug classes and split them using commas.
- Normalized drug-related fields and stored them in individual columns to prepare for machine learning tasks.

6. Feature Engineering

Created custom Boolean columns to flag important characteristics such as:

- has_dizziness
- has_hives
- is_painkiller
- has_fever, etc.

These features enabled:

- Advanced filtering
- Grouping
- Detailed analysis of symptoms and drug categories

7. Data Visualization & Relationship Analysis

- Visualized data to understand patterns in drug usage and their side effects.
- Explored relationships between conditions, side effects, and drug classes.

8. Drug Recommendation System

Two systems were developed:

System 1: Name-Based Drug Recommendation

- Recommends drugs based on partial or full name match.
- Technique: Text similarity and string matching.
- Input: Drug name
- Output: Top-N similar drugs
- Purpose: Suggest alternative medications with similar efficacy.

System 2: Health Issue-Based Recommendation

- Focuses on minimizing side effects while treating a condition.
- Input: Specific health issue
- Output: Safer drug alternatives
- Purpose: Safety-oriented recommendations using symptom-based filtering.

9. Project Novelty

- Designed two distinct drug recommendation systems.
- Applied Boolean-based feature engineering to enhance data analysis.
- Combined text cleaning, NLP techniques, and recommender logic.
- Unique approach for a student project, rarely implemented at this level.

10. Future Scope

- Integrate with real-time APIs (e.g., openFDA, DrugBank).
- Develop a user-friendly web interface or chatbot for live recommendations.
- Use machine learning models to predict side effects for new drug-condition pairs.
- Expand to include drug interactions and contraindications.

11. Conclusion

- Successfully analysed healthcare data to draw meaningful conclusions.
- Built two functional and distinct drug recommendation systems:
- One for similar drug recommendations
- One for side-effect-aware suggestions
- Demonstrated the impact of data-driven decision making in healthcare.
- Showcased potential for practical applications in clinical decision support tools.