Drug Recommendation System Documentation

# Introduction

The Drug Recommendation System is designed to assist healthcare providers in prescribing appropriate medications to patients based on their medical history, symptoms, and other relevant factors. By leveraging machine learning (ML) techniques, the system analyzes patient data to generate personalized drug recommendations, aiming to improve treatment outcomes, minimize adverse reactions, and enhance patient safety.

# Data Source

The dataset used for the Drug Recommendation System is sourced from the MIMIC-IV database version 2.2, available on PhysioNet. This dataset contains de-identified electronic health records from patients admitted to the Beth Israel Deaconess Medical Center from 2011 to 2019. After downloading the dataset, it is loaded into a Redshift database for further analysis and processing.

# Preprocessing Steps

1. *Merge Admission Data with Diagnosis Data*: Admission records are merged with diagnosis data to associate each admission with the corresponding medical conditions diagnosed during the stay.

2*. Merge Admission Data with Chart Events*: Admission data is further enriched by merging it with chart events data, providing detailed clinical observations recorded during the patient's stay.

3*. Group Patient Label Data with Hospital Admissions*: Patient label data, such as readmission status, is grouped with admission records to create a labeled dataset for predictive modeling.

4. *Recategorize Admission Type and Race*: Admission types and race categories are recategorized into broader groups for simplification and improved model interpretability.

# Rules Applied

1. Recategorize Admission Type: "EW EMER.", "URGENT", "DIRECT EMER." are categorized as "Emergency".
2. "OBSERVATION ADMIT", "EU OBSERVATION", "DIRECT OBSERVATION", "AMBULATORY OBSERVATION" are categorized as "Observation".
3. "SURGICAL SAME DAY ADMISSION" is categorized as "Surgical Same Day Admission".
4. "ELECTIVE" is categorized as "Elective".
5. Recategorize Race: Various subgroups are grouped under broader categories such as "ASIAN", "BLACK/AFRICAN", "HISPANIC OR LATINO", etc.
6. Mapping of ICD Codes to Categories:

|  |  |
| --- | --- |
| **ICD Range** | **Chapter Block Title** |
| 001–139 | Certain infectious and parasitic diseases |
| 140–239 | Neoplasms |
| 240–279 | Endocrine, nutritional and metabolic diseases and immunity disorders |
| 280–289 | Diseases of the blood and blood-forming organs |
| 290–319 | Mental disorders |
| 320–389 | Diseases of the nervous system and sense organs |
| 390–459 | Diseases of the circulatory system |
| 460–519 | Diseases of the respiratory system |
| 520–579 | Diseases of the digestive system |
| 580–629 | Diseases of the genitourinary system |
| 630–679 | Complications of pregnancy, childbirth, and the puerperium |
| 680–709 | Diseases of the skin and subcutaneous tissue |
| 710–739 | Diseases of the musculoskeletal system and connective tissue |
| 740–759 | Congenital anomalies |
| 760–779 | Certain conditions originating in the perinatal period |
| 780–799 | Symptoms, signs, and ill-defined conditions |
| 800–999 | Injury and poisoning |
| E and V | External causes of injury and supplemental classification (E codes); (V codes) |
| XXI | Factors influencing health status and contact with health services |
| XXII | Codes for special purposes |

# Model Building Approaches

The Drug Recommendation System employs multiple approaches for generating drug recommendations based on patient diagnosis data. These approaches are outlined below:

1. Type 1 Model (with Admit Date): This approach incorporates admission date information along with patient diagnosis data to generate drug recommendations. It considers the temporal aspect of diagnoses and their relevance to drug prescriptions over time.

2. Type 2 Model (without Admit Date): Unlike the Type 1 model, this approach excludes admission date information and focuses solely on patient diagnosis data. It disregards temporal aspects and treats all diagnoses equally in generating drug recommendations.

3. K-Nearest Neighbors (KNN): The KNN approach calculates the similarity between patients based on their diagnosis profiles and recommends drugs based on the nearest neighbors' prescriptions. It considers the closeness of patient diagnosis patterns in recommending drugs.

4. Group Diagnosis and Unique Drugs Approach: This approach groups all diagnoses together and identifies unique drugs associated with each diagnosis. It analyzes the co-occurrence of diagnoses and drugs to generate drug recommendations based on diagnosis patterns.

5. Co-Occurrence Matrix: The system constructs a co-occurrence matrix to capture the frequency of diagnosis pairs and drug pairs occurring together. This matrix provides insights into the associations between diagnoses and drugs, guiding the recommendation process.

6. Co-Occurrence and Correlation for Diagnosis: Building upon the co-occurrence matrix, the system calculates correlations between diagnosis pairs to identify patterns of comorbidity. It uses these correlations to refine drug recommendations based on diagnosis combinations.

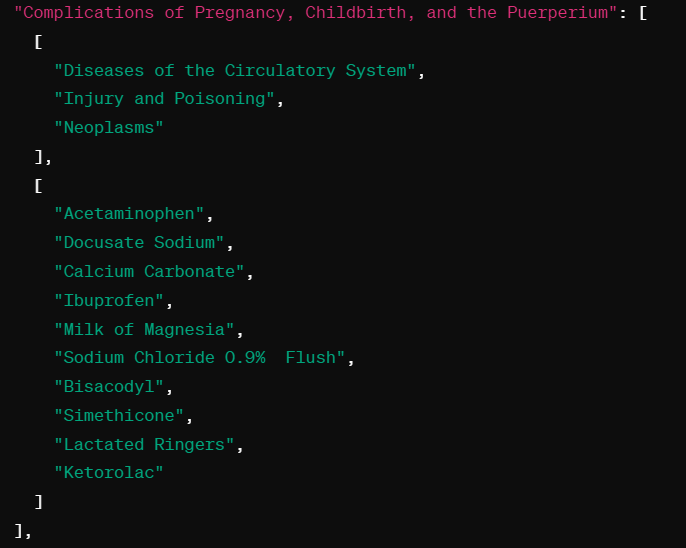
7. Convert Co-Occurrence Matrix to Correlation Matrix: The co-occurrence matrix is transformed into a correlation matrix to quantify the strength and direction of associations between diagnoses and drugs. This correlation matrix serves as the basis for filtering drugs and generating final recommendations.

# Final Recommendation Process

In the final recommendation process, the system predicts the nearest three diagnoses based on a given diagnosis and recommends drugs associated with these diagnoses. Since prescription records often lack specific diagnosis information, this approach helps bridge the gap by leveraging diagnosis patterns and comorbidity information for drug recommendations.

# Output

Diagnosis, followed by the nearest diagnosis possible. It also consists of all the nearest drugs for the diagnosis.



# Conclusion

The Drug Recommendation System employs various modeling approaches to generate personalized drug recommendations based on patient diagnosis data. By considering different aspects of diagnosis patterns and their associations with drug prescriptions, the system aims to enhance medication management, improve treatment outcomes, and optimize patient care in clinical settings.