RAJALAKSHMI ENGINEERING COLLEGE

Department of Artificial Intelligence and Data Science

Drug Effectiveness and Adverse Reaction Detection Using Big Data Analytics

Supervisor Name: K Subramanian
Team members:

Aishwarya M (231801003)

Angelin Mary R (231801010)

Darshini R (231801026)

Problem Statement and Motivation

- **Problem Statement:** Adverse Drug Reactions (ADRs) are a significant cause of illness and mortality worldwide. Traditional methods for detecting ADRs, like clinical trials and voluntary reporting, are often slow, expensive, and suffer from under-reporting. This delay in identifying harmful side effects can put patient populations at risk.

- **Motivation:** With the explosion of real-world data from electronic health records, patient forums, and social media, there's a massive, untapped resource for monitoring drug safety. The motivation for this project is to leverage a powerful big data platform, **Databricks**, to analyze this large-scale data. By doing so, we can create a proactive system to detect potential adverse reactions and assess drug effectiveness much faster than traditional methods, ultimately improving patient safety.

Existing System

Current Methods for ADR Detection:

- **Pre-market Clinical Trials:** Conducted in controlled environments with a limited number of participants. They often fail to detect rare or long-term side effects.
- **Spontaneous Reporting Systems (SRS):** Healthcare professionals and patients voluntarily report adverse events.

Limitations of Existing Systems:

- **Data Latency:** There is a significant time lag between when an adverse event occurs and when it is reported and analyzed.
- **Under-reporting:** It's estimated that over 90% of adverse drug reactions go unreported.
- **Scalability Issues:** Traditional databases and analytics tools cannot efficiently handle the massive volume and variety of modern healthcare data

Objectives

- To build a scalable **data ingestion** pipeline on the **Databricks** platform to collect healthcare-related datasets.
- To process and structure the raw data by creating optimized tables in the Databricks environment.
- To identify Key Performance Indicators (KPIs) for drug effectiveness and adverse reactions and use SQL queries (assisted by generative AI tools like ChatGPT) to analyze the data.

- To develop an interactive **dashboard** to **visualize** these KPIs, allowing for real-time monitoring and trend analysis.
- To demonstrate the system's capability to identify potential correlations between specific drugs and adverse outcomes.

Abstract

This project presents a big data approach for the real-time detection of drug effectiveness and adverse reactions. The current pharmacovigilance systems face challenges with data latency and under-reporting. To address this, our proposed system utilizes the **Databricks** platform to build a comprehensive data analytics pipeline. The process involves **ingesting** large volumes of data, creating structured tables, and executing advanced SQL queries to derive key insights. These insights and Key Performance Indicators (KPIs) will be visualized on an interactive **dashboard**. This will provide a powerful tool for healthcare analysts to proactively monitor drug safety, identify potential risks, and make data-driven decisions, ultimately enhancing patient care and public health.

Thank You