Data Management Pitch

API-Driven: Drug Data Management and Integration System

Presented by:
Abdul Sami (939542)
Davina David (946851)
Shiraz Ahmed (948290)

Introduction

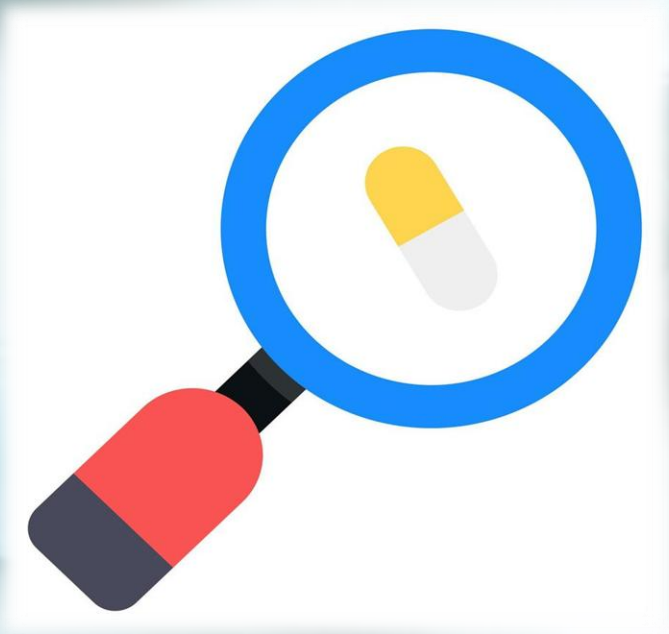
- **Healthcare data is rapidly digitized**, creating opportunities but also challenges in management and analysis.
- **Accurate and complete drug information** is essential for safe prescriptions, informed decisions, and patient care.
- **Project Focus:** Acquire, integrate, and analyze drug-related data using **RxNorm and OpenFDA APIs**.
- **Research Question:** *"How accurate is the RxNorm API extraction in providing RxCUI identifiers when combined with OpenFDA data?"*
- **Approach:** Data is sourced from multiple platforms, integrated into a unified structure, pre-processed, and validated for accuracy and completeness.

Objectives

- Evaluate the accuracy of **RxCUI** extraction from the RxNorm API.
- Integrate **RxNorm** data with **OpenFDA** information to create a comprehensive drug dataset using (rxcuis) unique identifier.
- Develop a reliable data management pipeline using Python for acquisition, cleaning, and integration.
- Store structured data in SQLite database for efficient access and to maintain data consistency.
- Ensure the dataset is suitable for healthcare applications through rigorous validation and quality checks.

Scope and Limitation

The project is focused on integrating and validating RxNorm and OpenFDA drug data to produce accurate, consistent, and usable datasets for healthcare data management, without extending into additional APIs or predictive analytics



Software Architecture

- Programming:

Python

- APIs:

RxNorm, openFDA Drug API

- Database:

SQLiteLibraries: requests, sqlite3

The overall architecture follows a **modular ETL-like approach**, where data is **Extracted** from external APIs, **Transformed** through preprocessing and integration, and **Loaded** into a structured SQLite database.

Process Workflow

1. Data Acquisition:

Two web-apis are extracted for drug data management

-RxNorm API: having information of standardized drug identifiers and nomenclature.

-openFDA drug API: having regulatory details and safety warnings.

2. Data Integration:

RxNorm and openFDA drug API are integrated using RxCUI (unique identifier).

3. Database Pre-Processing:

The integrated api is then parsed and cleaned before storing in database.

3. Database Storage:

This pre-processed data is then stored in Relational database using SQLite.

4. Data Quality Assessment:

Data is then validated by removing duplicates and handling missing values.

5. Exploratory Analysis:

Further data is evaluated using Data Metrics like consistency and accuracy.

Exploratory Analysis Results

– Data Accuracy

- RxNorm provides highly accurate RxCUIs for queried drugs.
- openFDA adds regulatory and safety validation.
- Simulated test data slightly reduces accuracy.

- Data Consistency

- Standardized identifiers maintain uniformity across APIs.
- Scripts enforce consistent data types and formats.
- Duplicate entries removed during post-storage validation.

Conclusion

- RxNorm API is reliable for accurate Rxcui extraction and basic drug identification.
- High Accuracy and completeness of core fields support healthcare data integration.

Future Work

-Integrate additional APIs for enriched profiles.

-Apply advanced data quality metrics (consistency, timeliness).

-Enhance visualization and analysis for clinical applications.

References

- RxNorm API Documentation: https://rxnav.nlm.nih.gov/RxNormAPIs.html
- openFDA Drug API Documentation:

https://open.fda.gov/apis/drug/

- SQLite Documentation:

https://sqlite.org/docs.html

THANKYOU!