

Assignment 2: Integrating Structured and Unstructured Data in Healthcare Analytics

Background

Healthcare organizations generate a variety of data. Patient records (structured data) contain fields such as patient demographics, diagnoses, medications, and lab results. Medical imaging (unstructured data), such as X-rays, MRIs, or CT scans, holds rich diagnostic information that is not readily captured in tables.

This assignment challenges students to combine structured and unstructured data to derive meaningful insights and propose clinical decision-support applications.

Datasets Provided (Sample for Case Study)

Structured Data: Patient Records (CSV/Tabular format)

| Patient_ID | Age | Gender | Diagnosis | Medication | Lab_Result | Outcome |
|------------|-----|--------|--------------|------------|----------------|-----------|
| P001 | 56 | M | Diabetes | Metformin | HbA1c=8.1 | Stable |
| P002 | 44 | F | Hypertension | Lisinopril | BP=150/95 | Unstable |
| P003 | 68 | M | Stroke | Aspirin | CT=Lesion | Recovered |
| P004 | 39 | F | Asthma | Salbutamol | Spirometry=60% | Stable |

Unstructured Data: Medical Imaging (Descriptions / Metadata)

P001 – Chest X-ray: Mild cardiac enlargement, no pulmonary edema.

P002 – MRI Brain: Small ischemic changes, possibly linked to hypertension.

P003 – CT Brain Scan: Left hemisphere ischemic lesion, consistent with stroke.

P004 – Spirometry Graph: Reduced lung capacity, matches asthma symptoms.

Assignment Tasks

Data Understanding

- Differentiate structured (patient records) vs. unstructured (imaging data) formats.
- Identify the challenges of analyzing each type of data.

Data Integration

- Link structured records with corresponding imaging findings (by Patient_ID).
- Propose a framework to combine both datasets for clinical decision support.

Analysis

- From structured data: Identify correlations (e.g., lab results vs. outcomes).
- From unstructured data: Extract diagnostic insights (e.g., lesions in CT scans).
- Discuss how combining both could improve diagnostic accuracy.