Project Design Phase-II
Data Flow Diagram & User Stories

Date	27 February 3035
Team ID	SWTID-2026-5418
Project Name	Insurance Fraud Detection Using Maching Learning
Maximum Marks	4 Marks

Project Documentation: Insurance Fraud Detection Systeml. Data Flow Diagrams (DFD)

A. DFD Level-0 (Context Diagram)

The **Insurance Fraud Detection System** interacts with one external entity: the **Claims Officer (User)**.

External Entity	Data Flow (Input)	Main Process	Data Flow (Output)	External Entity
Claims Officer	**Claim Details**	Insurance Fraud Detection System	**Fraud / Not Fraud Result + Confidence Score**	Claims Officer

B. DFD Level-1 (Detailed Flow)

This diagram illustrates the step-by-step processing of a claim within the system.

1. External Entity: Claims Officer

2. Processes:

- **P1:** Input Validation
- **P2:** Data Preprocessing
- **P3:** Fraud Prediction Model
- **P4:** Generate Confidence Score
- **P5:** Display Results

3. Data Stores (Artifacts):

- **D1:** Insurance Claims Dataset (CSV)
- **D2:** Trained Model (.pkl files)
- **D3:** Scaler & Feature List

Detailed Flow:

1. **Claims Officer** submits data.
2. Data undergoes **(P1) Input Validation**.
3. Validated data proceeds to **(P2) Data Preprocessing**.
4. Preprocessed data is fed to **(P3) Fraud Prediction**, which accesses **(D2) ML Model** and **(D3) Scaler**.
5. Prediction output is used for **(P4) Confidence Score Calculation**.
6. The final result is presented via **(P5) Result Display (Dashboard)** to the Claims Officer.

II. User Stories

The following table outlines the system's functional and non-functional requirements from the perspective of the users.

No.	User Type	Requirement (Epic)	User Story	Acceptance Criteria	Priority	Release
USN-1	Claims Officer (Web User)	Claim Submission	As a claims officer, I can enter insurance claim details through a web form.	All required fields must be validated before submission.	High	Sprint-1
USN-2	Claims Officer	Fraud Prediction	As a claims officer, I want the system to predict whether a claim is fraudulent.	System returns a clear Fraud / Not Fraud result.	High	Sprint-1

USN-3	Claims Officer	Confidence Scoring	As a claims officer, I want to see the confidence score of the prediction.	System displays the prediction probability percentage.	High	Sprint-1
USN-7	System	Performance	As a user, I want results in less than 1 second.	Prediction response time must be less than 1 second.	High	Sprint-1
USN-4	Claims Officer	Explanation	As a claims officer, I want to understand why a claim was flagged.	System displays feature-level insights (explainability).	Medium	Sprint-2
USN-5	Administrator	Model Management	As an admin, I want to retrain the model with updated data.	New model artifacts are saved and loaded successfully for deployment.	Medium	Sprint-2
USN-6	System User	API Access	As a system user, I want prediction results via API.	JSON input correctly returns prediction and scores in JSON output.	Medium	Sprint-2

III. Summary of System Documentation

The documentation effectively details the system's architecture and requirements:

- **DFD Clarity:** It clearly maps the flow of **Claim Details** through Validation, Preprocessing, and the Machine Learning Model to produce the **Fraud Prediction** and **Confidence Score**. It also identifies key artifacts like the Trained Model and Scaler.
- **User Story Definition:** The requirements are well-defined, specifying **who** (User Type), **what** (User Story), and **how success is measured** (Acceptance Criteria), along with planning attributes (Priority, Release).