

**Project Design Phase-II**  
**Solution Requirements (Functional & Non-functional)**

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

## Project: Insurance Fraud Detection System

### Solution Requirements Overview

#### Functional Capabilities

Epic Requirement	Key Functionality (Stories / Sub-Tasks)
<b>Data Input (FR-1)</b>	Users can submit insurance claim details via a web form. Input fields are validated (e.g., numeric ranges), and categorical features utilize dropdown selections.
<b>Prediction Core (FR-2)</b>	The system processes the submitted claim using the underlying Machine Learning model, applying the necessary preprocessing pipeline (encoding, scaling) to generate a binary fraud/not-fraud prediction.
<b>Risk Scoring (FR-3)</b>	A confidence score (0–100%) indicating the probability of fraud is displayed, often represented visually via a risk meter.
<b>Model Explainability (FR-4)</b>	Feature-based insights are provided to justify the prediction, including value-level interpretations (e.g., low, medium, high influence).
<b>Model Management (FR-5)</b>	Trained models and scalers are reliably loaded from storage, and feature consistency between training and prediction is strictly maintained.
<b>API Interface (FR-6)</b>	A standard REST API endpoint ( <code>/api/predict</code> ) is available to accept and return claim data in JSON format for external integration.
<b>Data Pipeline (FR-7)</b>	Handles the loading, preprocessing, feature engineering, and feature selection for the insurance dataset used in model

	development.
<b>Model Evaluation (FR-8)</b>	Computes and reports standard metrics (Accuracy, Precision, Recall, F1-score, ROC-AUC), including the generation of a confusion matrix and test report.

## Non-Functional Qualities

Quality Attribute	Description
<b>Usability (NFR-1)</b>	Requires a simple, responsive web dashboard with guided input ranges and the option to load sample data for testing.
<b>Security (NFR-2)</b>	Ensures robust input validation, secure structured API handling, and prevention of invalid or malicious requests.
<b>Reliability (NFR-3)</b>	The consistent preprocessing pipeline prevents data leakage and guarantees repeatable, trustworthy predictions.
<b>Performance (NFR-4)</b>	The optimized Random Forest model must deliver prediction responses in less than 1 second.
<b>Availability (NFR-5)</b>	The system runs on a local server and is capable of cloud deployment to ensure continuous 24/7 access.
<b>Scalability (NFR-6)</b>	Utilizes a modular 3-tier architecture, allowing independent scaling of the backend application or the ML model.

<b>Maintainability (NFR-7)</b>	The code employs a clean, modular structure, logically separating preprocessing, training, evaluation, and the application layer.
<b>Accuracy (NFR-8)</b>	The model must achieve a high fraud detection rate, maintained through optimized feature selection techniques.

## Summary

This solution is designed to be a comprehensive fraud detection tool that:

- ✓ Accepts structured insurance claim inputs.
- ✓ Processes data through a robust Machine Learning pipeline.
- ✓ Predicts fraud risk and provides a confidence score.
- ✓ Delivers explainable, feature-based insights.
- ✓ Guarantees high performance, reliability, and scalability.