1.7 Documentation

This section presents a complete and structured record of the workflow, design decisions, experiments, and reasoning applied throughout the development of the **Healthcare Fraud Detection System**. The documentation is intended to serve both as a technical audit trail and as a reproducible reference for future development, deployment, and review.

1.7.1 Project Workflow Overview

The project followed an end-to-end machine learning workflow structured around a standard data science pipeline:

1. **Data Understanding & Exploration**
   Raw datasets were profiled to identify schema, inconsistencies, missing values, distributions, and entity relations.
2. **Feature Engineering & Aggregation**
   Claim-level data was transformed into provider-level features using statistical aggregations.
3. **Class Imbalance Handling**
   Fraud providers were underrepresented and addressed using oversampling techniques.
4. **Modeling & Algorithm Comparison**
   Multiple candidate models were trained and evaluated using cross-validation.
5. **Evaluation & Error Analysis**
   Performance was assessed using predictive metrics and business-oriented cost modeling.
6. **Explainability & Interpretation**
   Local and global model explanations were introduced using SHAP.
7. **Operational Readiness & Deployment Design**
   A final decision layer was defined to support human-in-the-loop governance.

1.7.2 Data Exploration and Processing Decisions

Dataset Composition

Four datasets were integrated:

- Provider fraud labels
- Beneficiary demographics
- Inpatient claims
- Outpatient claims

Entity Relationships

Entity	Join Key	Role
Provider	Provider ID	Fraud label
Beneficiary	BeneID	Demographic link
Claims	ClaimID	Transaction identity

Data Quality Measures

Issues Detected:

- Missing beneficiary death dates (~99%)
- Extreme claim outliers
- Duplicate and inconsistent records
- Mixed date formats

Remediation Applied:

- Type casting and datetime normalization
- Log-transformation for skewed monetary values
- Numeric validation filtering
- Non-numeric feature elimination
- Aggregation-based imputation

1.7.3 Feature Engineering Strategy

Fraud patterns manifest at the provider-level, not the transaction-level.

Feature Class	Examples
Volume	Claim count
Monetary	Total & mean reimbursement
Diversity	Unique beneficiaries
Behavior ratios	Inpatient / outpatient
Risk proxies	Claims per patient
Temporal	Monthly claim density
Geographic	Top states served

Justification: Aggregated behavior signals are more predictive than individual claims.

1.7.4 Class Imbalance Handling Strategy

Fraud providers represent a small minority.

Solution:

- SMOTE oversampling was applied
- Stratified train–test split
- Recall and PR-AUC were prioritized over accuracy

Business Rationale:

False negatives carry substantially higher financial costs than false positives.

1.7.5 Model Selection Rationale

Evaluated Models

Model	Purpose
Logistic Regression	Baseline interpretability
Decision Tree	Pattern transparency
Random Forest	Feature importance
Gradient Boosting	Final model

Final Model: Gradient Boosting

Why selected:

- Highest PR-AUC score
- Consistent validation performance
- Robust to class imbalance
- Non-linear learning capability

1.7.6 Experimental Log & Trials

Experiment	Outcome
No SMOTE	Poor recall
SMOTE	Recall +20%
Scaling added	Logistic/SVM improved
Tree depth tuning	Reduced variance
PR-AUC optimization	Better fraud ranking

Feature Iterations:

- Removed identifiers
- Filtered low-variance columns

- Added behavior ratios

1.7.7 Error Analysis

False Positives

- Cause: Legitimate large providers resemble fraud volume patterns
- Cost: Moderate audit overhead
- Mitigation: Peer-group normalization

False Negatives

- Cause: Low-activity fraud blending into normal profiles
- Cost: High financial risk
- Mitigation: Temporal features and anomaly detection

1.7.8 Explainability and SHAP Analysis

SHAP Integration

Local explanations were generated for:

- Multiple False Positives
- Multiple False Negatives

Each case includes:

- Waterfall charts
- Feature contribution analysis
- Short interpretation narratives

Findings:

- High reimbursement drives false alarms
- Averaged patterns hide small fraud behavior
- Feature interactions matter

(Screenshots and interpretations provided in the appendix.)

1.7.9 Statistical Robustness & Confidence Intervals

To avoid performance overstatement:

- Bootstrapped confidence intervals were computed for:
    - PR-AUC
    - Recall
    - Precision
    - F1-score

This ensures conclusions reflect uncertainty.

1.7.10 Hyperparameter Configuration (Appendix)

Documented:

- Grid search ranges
- Best-performing values
- Comparison tables

(Full grid included in Appendix B.)

1.7.11 Operational Decision Framework

Decision Threshold

Threshold chosen based on PR curve to minimize cost:

$$\text{Expected Cost} = (FP \times 500) + (FN \times 10000)$$

Operational Outputs

Component	Value
Threshold	Optimized
Providers flagged/month	Estimated
Review load	Controlled
Expected gain	Positive

Deployment Pipeline

1. Monthly data ingestion
2. Provider feature aggregation
3. Model inference
4. Risk-based thresholding

5. Human audit
6. Feedback loop

1.7.12 Documentation Standards

The project adheres to:

✅ Reproducibility
✅ Audit traceability
✅ Feature transparency
✅ Business alignment
✅ Ethical reasoning
✅ Deployment readiness