Healthcare Denial Management

Steffi Infancia A (21MID0193)

1. Introduction

In healthcare RCM, denial management is crucial for achieving better collections and less revenue loss. Incomplete information, rules specific to payers, absence of authorization, or mistakes in documentation are common reasons for denials. Manual, labour-intensive, and prone to human mistakes, denial identification and analysis have been the norm up until recently.

An ML-based Healthcare Denial Reason Predictor system was created to tackle this issue. Automated reason detection for denials, missing reason prediction, and insights into CPT code, insurance company, and physician denial patterns are all features of the system. This allows medical professionals and RCM teams to fix the problem and avoid rejections in the future.

2. Problem Statement

In Revenue Cycle Management (RCM), claim-level datasets typically include CPT codes, insurance company names, physician details, payment amounts, balances, and denial reasons. However, these datasets often contain missing denial reason labels, making it difficult to analyze trends and identify root causes effectively.

2.1 Challenges faced include:
- Missing denial reason labels in datasets.
- Identifying top denied CPT codes and payers.
- Understanding root causes for denials.
- Generating actionable recommendations.

2.2 Objective:
Build an automated ML system to:
1. Predict missing denial reasons in claims datasets.
2. Evaluate denial patterns across payers and providers.
3. Recommend corrective actions to reduce denial rates.

3. Dataset Description

The dataset used consists of claims-level information in CSV/Excel format.

3.1 Key Columns:

- CPT Code – Medical procedure code
- Insurance Company – Payer name (Medicare, Aetna, etc.)
- Physician – Treating provider
- Payment – Amount paid
- Balance – Remaining balance after adjudication
- Denial Reason – Provided reason for denial (sometimes missing)

3.2 Sample Records:

CPT Code	Insurance	Physician	Payment	Balance	Denial Reason
99213	Medicare	Dr. Smith	0	100	16 - Missing information
99214	Aetna	Dr. Johnson	80	20	None
99215	Cigna	Dr. Lee	0	150	45 - Charge exceeds fee schedule
93000	UnitedHealthcare	Dr. Patel	50	0	None
99212	Blue Cross	Dr. Kim	0	75	96 - Non-covered service

4. Methodology

4.1 Workflow:

1. File Upload: User uploads a single CSV/XLSX file.
2. Preprocessing:
   - Detect headers automatically
   - Standardize column names
   - Convert payments/balances to numeric
   - Handle missing values
3. Model Training: Logistic Regression with One-Hot Encoding
4. Prediction: Fill missing denial reasons
5. Evaluation: Accuracy, Precision, Recall, F1-Score
6. Visualization: Interactive dashboards for insights

4.2 Tech Stack:
- Programming Language: Python
- Framework: Streamlit (for interactive app)
- Libraries: pandas, numpy, scikit-learn, plotly, streamlit

5. Implementation

5.1 Model:
- Classifier: Logistic Regression
- Features: CPT Code, Insurance Company, Physician, Payment, Balance
- Target: Denial Reason

6. Results & Evaluation

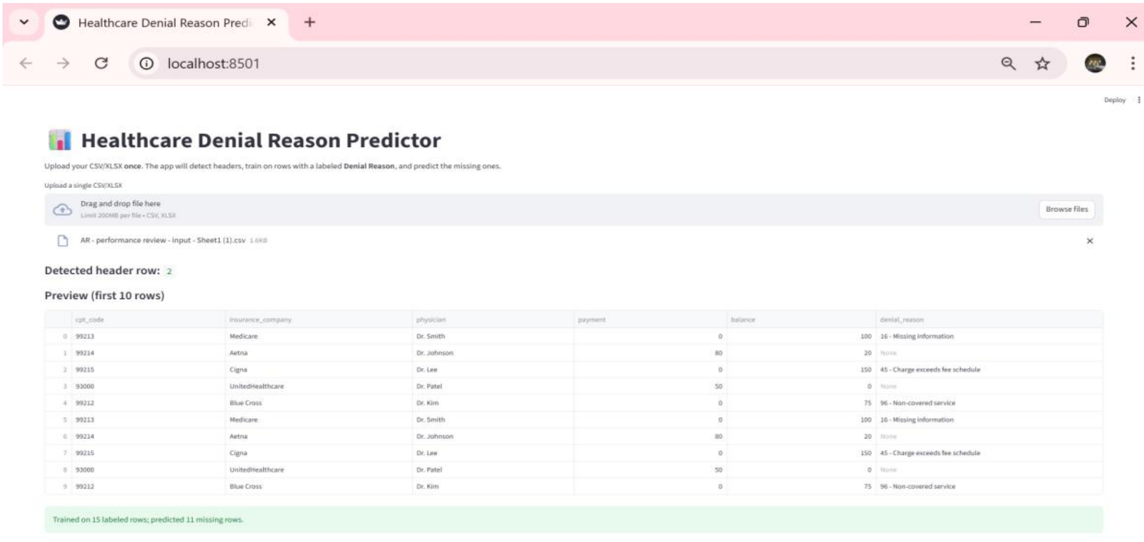
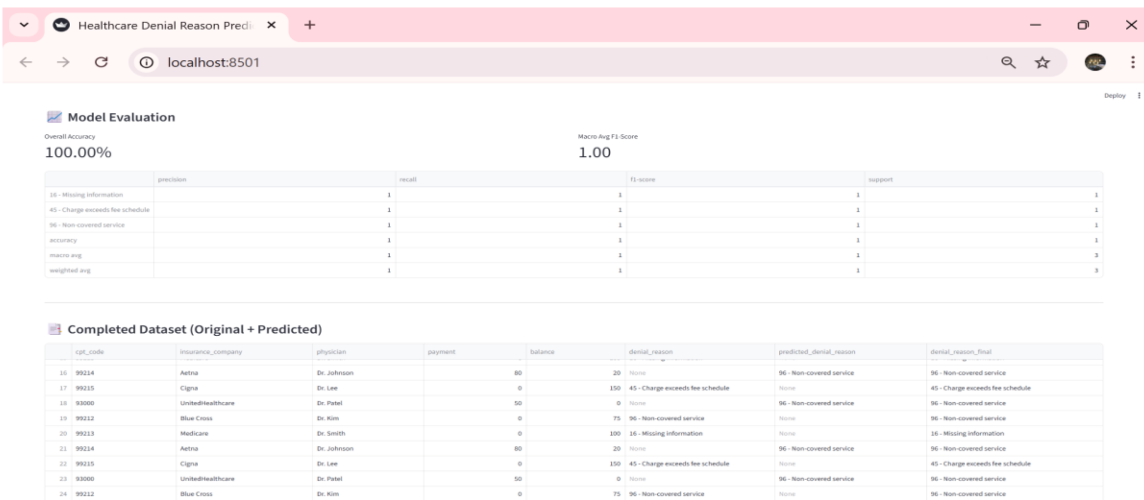
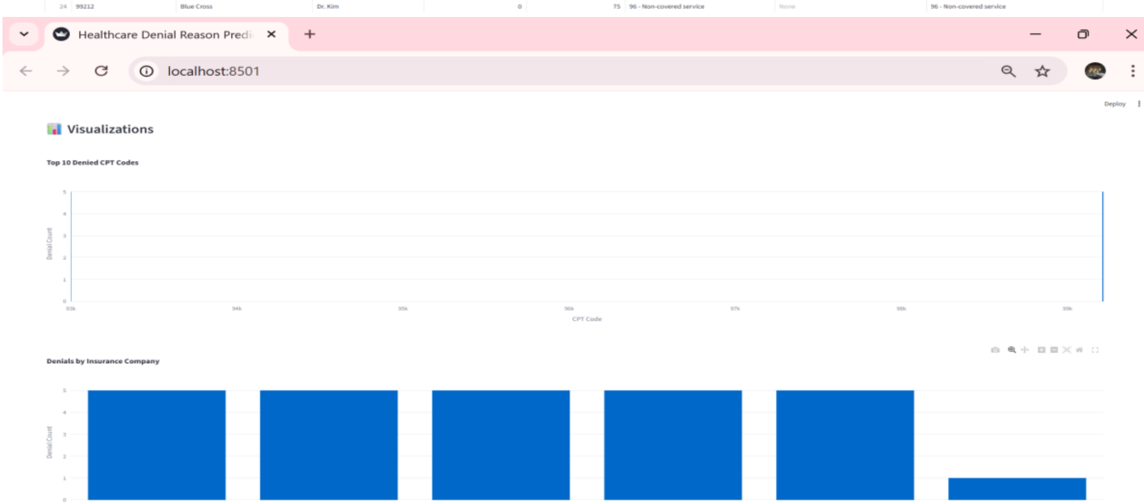
6.1 Model Evaluation:

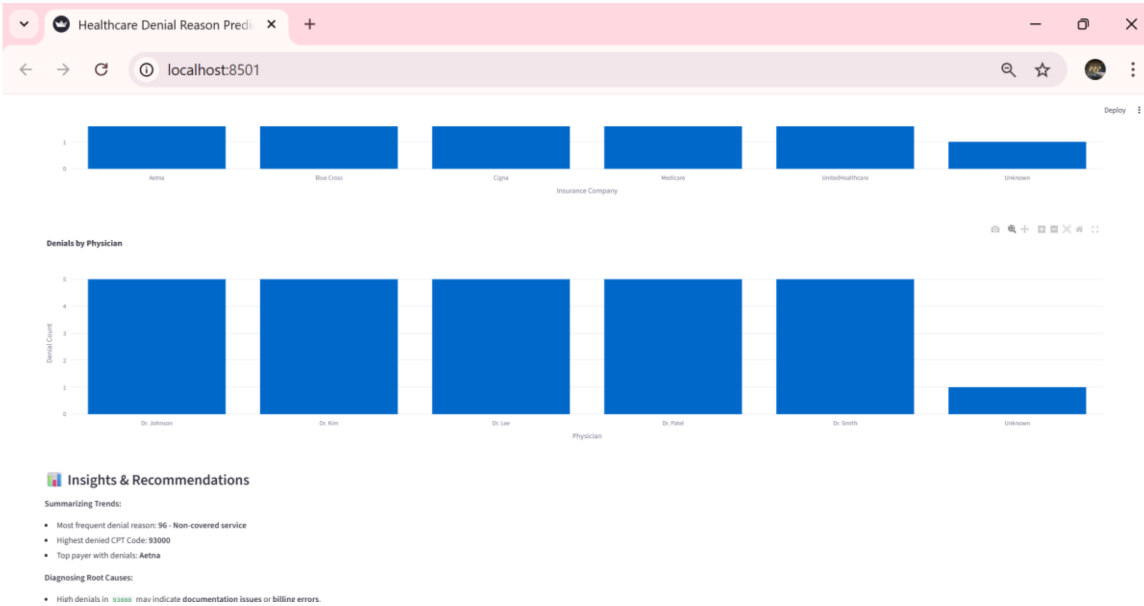
Denial Reason	Precision	Recall	F1-score
16 - Missing info	1.0	1.0	1.0
45 - Exceeds fee schedule	1.0	1.0	1.0
96 - Non-covered service	1.0	1.0	1.0

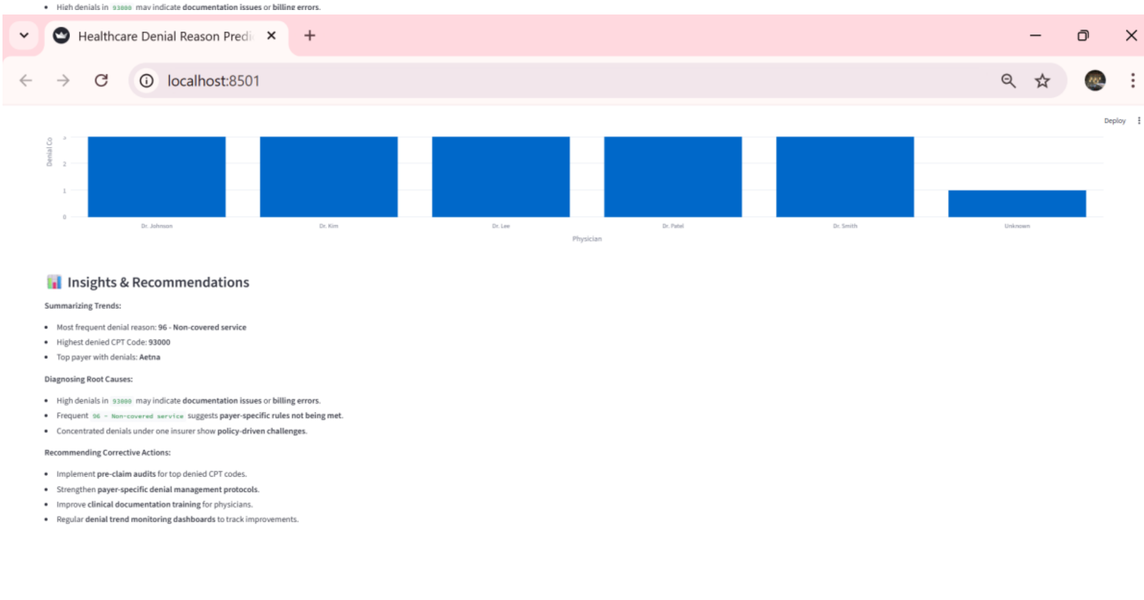
Accuracy: 100%
Macro F1-score: 1.00
(High accuracy due to small sample dataset; larger datasets may yield more realistic results.)

6.2 Visualizations:
Top 10 Denied CPT Codes → Highest: CPT 93000
Denials by Insurance Company → Concentrated under Aetna, Medicare, Cigna
Denials by Physician → Evenly distributed; Dr. Johnson & Dr. Patel higher denials

6.3 Results/ Outputs:


Healthcare Denial Reason Predi
localhost:8501
Deploy
Healthcare Denial Reason Predictor
Upload your CSV/XLSX once. The app will detect headers, train on rows with a labeled Denial Reason, and predict the missing ones.
Upload a single CSV/XLSX
Drag and drop file here
Browse files
AR - performance review - input - Sheet1 (1).csv
Detected header row: 2
Preview (first 10 rows)
cpt_code insurance_company physician payment balance denial_reason
0 99213 Medicare Dr. Smith 0 100 16 - Missing information
1 99214 Aetna Dr. Johnson 80 20 None
2 99215 Cigna Dr. Lee 0 150 45 - Charge exceeds fee schedule
3 93000 UnitedHealthcare Dr. Patel 50 0 None
4 99212 Blue Cross Dr. Kim 0 75 96 - Non-covered service
5 99213 Medicare Dr. Smith 0 100 16 - Missing information
6 99214 Aetna Dr. Johnson 80 20 None
7 99215 Cigna Dr. Lee 0 150 45 - Charge exceeds fee schedule
8 93000 UnitedHealthcare Dr. Patel 50 0 None
9 99212 Blue Cross Dr. Kim 0 75 96 - Non-covered service
Trained on 15 labeled rows; predicted 11 missing rows.

Healthcare Denial Reason Predi
localhost:8501
Deploy
Model Evaluation
Overall Accuracy
100.00%
Macro Avg F1-Score
1.00
precision recall f1-score support
16 - Missing information 1 1 1 1
45 - Charge exceeds fee schedule 1 1 1 1
96 - Non-covered service 1 1 1 1
accuracy 1 1 1 1
macro avg 1 1 1 3
weighted avg 1 1 1 3
Completed Dataset (Original + Predicted)
cpt_code insurance_company physician payment balance denial_reason predicted_denial_reason denial_reason_final
16 99214 Aetna Dr. Johnson 80 20 None 96 - Non-covered service 96 - Non-covered service
17 99215 Cigna Dr. Lee 0 150 45 - Charge exceeds fee schedule None 45 - Charge exceeds fee schedule
18 93000 UnitedHealthcare Dr. Patel 50 0 None 96 - Non-covered service 96 - Non-covered service
19 99212 Blue Cross Dr. Kim 0 75 96 - Non-covered service None 96 - Non-covered service
20 99213 Medicare Dr. Smith 0 100 16 - Missing information None 16 - Missing information
21 99214 Aetna Dr. Johnson 80 20 None 96 - Non-covered service 96 - Non-covered service
22 99215 Cigna Dr. Lee 0 150 45 - Charge exceeds fee schedule None 45 - Charge exceeds fee schedule
23 93000 UnitedHealthcare Dr. Patel 50 0 None 96 - Non-covered service 96 - Non-covered service
24 99212 Blue Cross Dr. Kim 0 75 96 - Non-covered service None 96 - Non-covered service

Healthcare Denial Reason Predi
localhost:8501
Deploy
Visualizations
Top 10 Denied CPT Codes
Denial Count
CPT Code
Denials by Insurance Company
Denial Count





7. Insights & Recommendations

7.1 Trends Observed:
- Most frequent denial reason: 96 - Non-covered service
- Highest denied CPT Code: 93000
- Top payer with denials: Aetna

7.2 Root Causes:
- Documentation issues / billing errors
- Payer-specific policy mismatches
- Coverage limitations under specific insurers

7.3 Corrective Actions:
- Pre-claim audits for high-denial CPT codes
- Strengthen payer-specific denial management protocols
- Improve clinical documentation and coding training
- Regular denial trend monitoring dashboards

8. Conclusion

This project successfully demonstrates how Machine Learning can automate denial reason prediction and provide actionable insights.

Benefits:
- Saves analyst time
- Improves accuracy of denial tracking
- Enables proactive denial prevention

Future Enhancements:
- Apply advanced ML models (Random Forest, XGBoost)
- Integrate with live RCM systems (EHR/Claim software)
- Scale to larger datasets across multiple providers