**A Machine Learning Framework for Predicting Medical Claim Denial**

**Abstract**

Medical claim denials cause significant financial losses and delays for healthcare organizations. Denied claims often result from errors such as missing information, incorrect coding, or eligibility issues. Currently, many billing teams address denials after claims are rejected, which requires time-consuming appeals and increases administrative costs. This project explores the use of machine learning (ML) to help predict which claims are likely to be denied before they are submitted.

A synthetic dataset of healthcare claims was generated to simulate real-world scenarios while protecting patient privacy. Various machine learning models were trained to predict claim approval status and, when denied, to identify likely reasons for denial. The models included Logistic Regression, Random Forest, Decision Tree, and XGBoost. The trained models were deployed in a simple web application using Flask, allowing users to upload a batch of claims and receive predictions.

This project demonstrates that ML-based tools can assist healthcare billing teams by identifying high-risk claims early, helping to reduce denials and improve revenue cycle efficiency. While this proof-of-concept was built using synthetic data, future work can extend this approach to real-world data and integrate it with existing billing systems.

**Problem Statement**

Medical claim denials remain a major challenge in healthcare revenue cycle management. When insurance companies reject submitted claims, healthcare providers lose revenue, face delays in reimbursement, and incur additional administrative costs to process appeals. Denials occur for many reasons, including incomplete documentation, incorrect or missing coding, eligibility mismatches, and failure to meet insurance payer requirements. Currently, most denial management processes in healthcare organizations are reactive. Billing teams typically respond to denials only after claims have been rejected, which creates additional workload and leads to inefficient revenue cycle operations. Addressing denials after the fact can delay cash flow and negatively impact on the financial health of healthcare providers.

There is a clear need for proactive solutions that can help billing staff and revenue cycle teams identify potential claim denials before submission. Such tools would enable early detection of high-risk claims, giving staff an opportunity to correct errors and reduce denial rates. The goal of this project is to explore the use of machine learning (ML) as a practical solution to this problem, by developing a predictive framework that can analyze claim data and provide early warnings about claims likely to be denied.

**Rationale and Framework of Solutions**

There is growing interest in applying data-driven tools to improve healthcare revenue cycle management. Machine learning (ML) offers a promising approach because it can analyze large amounts of healthcare claim data and detect patterns that are difficult to identify through manual review or rules-based systems. ML models can be trained to predict which claims are likely to be denied, based on past claim outcomes and known risk factors. This allows billing staff to focus their attention on high-risk claims and correct potential issues before submission.

The rationale for using ML in this project is based on its ability to improve prediction accuracy over traditional methods, provide scalable solutions that can process large volumes of claims, and offer insights into the reasons for potential denials. Additionally, building an ML-based tool can support healthcare organizations of various sizes, including those that cannot afford expensive commercial denial management software.

The framework of this solution involves creating a machine learning pipeline that includes data preparation, model training, evaluation, and deployment. In this project, synthetic healthcare claim data was generated to simulate real-world scenarios while maintaining privacy. Multiple ML models, including Logistic Regression, Random Forest, Gradient Boosting, and XGBoost, were trained to predict claim approval status and potential denial reasons. The trained models were deployed through a simple web application built with Flask, allowing users to upload batches of claims and receive predictions. This system provides a practical example of how healthcare organizations can use ML to enhance denial management and reduce revenue loss.

**Background and Literature review**

Healthcare providers continue to face a persistent challenge with medical claim denials. When claims are submitted to payers—whether public (e.g., Medicare, Medicaid) or private (e.g., BlueCross BlueShield, Athena)—an estimated 10–15% are initially denied [1]. These denials are frequently due to missing documentation, incorrect coding, authorization issues, or questions of medical necessity. Although more than half of these denied claims are eventually overturned and paid, the cost of resubmission ranges between $25 and $48 per claim. As a result, providers spend billions of dollars annually on avoidable denials, which strain cash flow, increase administrative workload, and can delay patient care [1]. To mitigate these impacts, healthcare organizations are prioritizing early identification and correction of at-risk claims—addressing errors in demographics, coding, authorization, or documentation prior to submission to improve first-pass acceptance rates and reduce operational rework.

Historically, denial prevention strategies have relied on rule-based systems within billing software that flag incomplete claims or invalid codes. While useful, these systems are static and require constant manual updates, making them less effective in detecting complex claim issues or adapting to evolving payer policies. To overcome these limitations, researchers and healthcare leaders have begun leveraging advanced technologies such as Machine Learning (ML), Natural Language Processing (NLP), Robotic Process Automation (RPA), and blockchain to improve the accuracy and efficiency of claims processing.

Machine learning models, in particular, have shown strong potential in analyzing historical claims data to identify patterns that predict claim denials. Unlike rule-based systems, ML models learn from data and can adjust to changing payer rules, improving overall claim acceptance and payment rates [2]. For example, one study applying a gradient-boosted tree model to healthcare claims data achieved an AUC of 0.91 and reduced denial rates by 25% over six months [3]. Another recent study followed a Design Science Research (DSR) framework to evaluate six interpretable ML models. Among them, the AdaBoost algorithm performed best, achieving an AUC of 0.83 and demonstrating strong potential for lowering operational costs while enhancing claims management [4].

Natural Language Processing (NLP) is another emerging technology that transforms unstructured clinical data—such as physician notes and discharge summaries—into structured formats that are useable for claims adjudication. NLP has been shown to reduce coding errors by accurately classifying diagnoses and procedures [5]. Additionally, NLP-powered chatbots are being deployed to answer claims-related queries, easing administrative tasks for patients and staff [6]. Other NLP tools based on machine learning can detect inconsistencies in medical records, helping to prevent inaccurate or fraudulent claims [7]. However, these systems often require large, diverse datasets for training, and their effectiveness can be limited by the complexity and ambiguity of clinical language [8]. Despite these challenges, NLP continues to support faster approvals, lower denial rates, and improved documentation accuracy [9].

To further enhance automation, many healthcare organizations are adopting Robotic Process Automation (RPA). RPA excels at repetitive tasks such as data entry, claim validation, and payment processing. One study found that RPA significantly reduced the need for human involvement by automating common claim workflows, thereby improving processing speed [10]. When integrated with AI models, RPA has shown efficiency gains of over 40% in large-scale claims operations [11]. However, legacy systems in some healthcare organizations struggle to integrate with modern RPA frameworks, and implementation can be hindered by high setup costs and resistance to change among staff [12].

Blockchain technology also holds promise for transforming claims management. Smart contracts built on blockchain platforms can automate claim settlements with third-party payers, enabling real-time verification and reducing processing delays [13]. Decentralized claims frameworks using blockchain allow encrypted patient and claim data to be securely shared among stakeholders, eliminating redundant checks and enhancing transparency [14]. Technology’s immutability also strengthens data integrity. Nonetheless, challenges remain scalability, regulatory compliance, and high computational demands have limited widespread adoption across healthcare systems [15].

Although these technologies offer significant advantages, ongoing challenges must be addressed—such as interoperability between systems, data privacy concerns, and integration with legacy infrastructure. Researchers emphasize that standardizing data exchange protocols and implementing strong cybersecurity frameworks will be essential for overcoming these barriers and enabling large-scale adoption [16].

Even though many new technologies are helping with claims processing, most healthcare organizations still don’t have easy-to-use tools that can catch problems before a claim is sent. Many systems only work after a claim is denied, or they don’t clearly explain why the denial happened. This project aims to fill that gap by building a simple machine learning–based website that predicts whether a claim might be denied *before* it’s submitted. If a claim is likely to be denied, the system will also show the possible reason. Using synthetic data that closely reflects real-life situations, this tool can help hospitals and clinics fix errors early, reduce costs, and avoid delays in payments.

**Methods and Models**

**Data generation and preparation**: Since real patient and insurance claim data are often restricted due to privacy laws, synthetic data is used in this project to simulate real-world claim scenarios. The dataset was generated using a custom python script that included fields like patient demographics, diagnosis codes, procedure codes, provider information, insurance details, and denial reasons. A total of 10,000 claims are created.

**Data Preprocessing**: Before applying machine learning models, the data was cleaned and prepared. Missing values were handles appropriately. Categorical features like insurance provider, denial reason were encoded using techniques like label encoding and one hot encoding. The final dataset was split into features [X] and target variable [Y].

**Exploratory Data Analysis:** EDA was performed using pandas, Matplotlib, and seaborn to understand claim distribution, denial rates, and top denial reasons. EDA helps us uncover hidden patterns and gain insights into the data before building prediction models.

1. A blue circle with a pink circle with a number of text

   AI-generated content may be incorrect.Claims Approval Vs Denial Rate

This pie chart shows that 79.9% of medical claims were approved and only 20.1% were denied. So out of 10,000 claims 7,989 claims are approved and 2,011 claims are denied. This indicates there is a class imbalance which needs to be addressed during model training.

1. A pie chart with different colored circles

   AI-generated content may be incorrect.Breakdown of claims Denial by operational area

This pie chart presents the operational breakdown of claim denials across the healthcare revenue cycle, mapped to the Front, Middle, and Back Office functions. Most denials approximately 43.1%—originate in the Front Office, where processes like patient scheduling and registration, eligibility and benefit verification, financial counselling, and prior authorization are handled. Errors here often include incomplete registration, incorrect insurance details, or missing prior authorization. The Middle Office accounts for 31% of denials, commonly tied to clinical documentation, utilization management, medical coding, and charge capture, where issues such as insufficient documentation or coding inaccuracies can lead to claim rejection. Finally, 25.9% of denials stem from the Back Office, where processes like payment posting, denial management/A/R follow-up, patient collections, and reporting take place. Typical causes here include duplicate submissions, coordination of benefits errors, and missed filing deadlines. These insights emphasize the importance of strengthening workflows across all phases to minimize preventable denials and optimize revenue performance.

1. Distribution of Denial reasons:

A graph of a number of patients

AI-generated content may be incorrect.

This bar chart displays the distribution of specific denial reasons among all denied claims in the dataset. The most common reason for denial is Incomplete Registration Data, indicating that many claims fail due to missing or invalid patient information captured at the front desk. Other frequent causes include Eligibility Verification Failure, Inadequate Clinical Documentation, and Duplicate Claim Submission, highlighting issues that span both the front and middle office. Less frequent but still important are denials caused by Incorrect Procedure Codes, Missing Prior Authorization, and Exceeded Timely Filing Limits. These insights emphasize the need for improved processes in data entry, clinical documentation, and payer coordination to reduce claim denials and enhance revenue cycle performance.

1. Distribution of claims by medical specialty:

A graph of a number of claims

AI-generated content may be incorrect.

This bar chart illustrates the distribution of claims across medical specialties. Specialties like Plastic Surgery, Emergency Medicine, and Radiology account for a larger share of total submitted claims in this synthetic dataset. This reflects the frequency of procedures typically performed in these high-volume areas, which may be subject to higher denial scrutiny based on clinical complexity, coding accuracy, and payer policies. Understanding specialty-wise claim patterns helps prioritize targeted denial management efforts in high-volume departments.

**Machine learning models**:

The goal of this project was to develop machine learning models that could:

* 1. Predict whether a medical claim would be approved or denied (yes/no), and

1. If denied, identify the likely reason (such as missing documentation or eligibility error).

To do this, four machine learning models were trained and tested:

1. Logistic Regression: This is one of the simplest models used for classification. It works like a smart yes/no calculator that looks at input factors (like provider, procedure code, etc.) and calculates the probability of denial.
2. Decision Tree: This model makes decisions by asking a series of yes/no questions. It splits the data into branches based on features, like “Was prior authorization provided?” or “Is the claim amount above a certain value?”
3. Random Forest Classifier: This is a collection of many decision trees that work together. Each tree gives an answer, and the model takes a vote. This helps improve accuracy and reduce errors that might come from using just one tree.
4. XGBoost Classifier: This is a more advanced model that builds decision trees in sequence. Each new tree tries to fix the mistakes made by the previous ones. It is often used in real-world problems because it is fast and accurate.

These models were selected because they are well-suited for structured data like healthcare claims. The dataset was divided into two parts — 80% for training the models and 20% for testing how well they perform.

To handle the imbalance between approved and denied claims, SMOTE (Synthetic Minority Over-sampling Technique) was applied to the training dataset. This technique creates new, realistic examples of denied claims by using existing ones as a base, helping the model learn denial patterns more effectively. This improved the model’s ability to detect denials and led to better performance, especially in recall and F1-score.

**Model Evaluation**:

To measure how well the machine learning models performed, the F1-score was used as the main evaluation metric. The F1-score is especially useful when the data has an uneven number of approved and denied claims, which is often the case in real healthcare settings.

The F1-score balances two things:

* Precision: How many of the claims predicted as “denied” were actually denied
* Recall: How many of the actual denied claims were correctly identified

Using the F1-score helps ensure that the model is not only accurate but also fair and reliable in both detecting true denials and avoiding false alarms. This is important for billing teams who want to focus their efforts only on high-risk claims. Each model was trained on 80% of the data and tested on the remaining 20%. The model with the highest F1-score was chosen for deployment.

**Model Deployment**:

After evaluating all the machine learning models, the one with the highest F1-score was selected for real-world use. This model was then deployed into a web-based tool using Flask, which is a lightweight and beginner-friendly web framework built with Python.

The purpose of this tool is to make the model accessible to users who may not have any programming or technical background — such as hospital billing staff, claims managers, or healthcare administrators.

In the application, users can simply upload a CSV file containing a list of new insurance claims. The model will automatically analyze each claim and give the following results:

* Whether the claim is likely to be approved or denied
* If denied, the most likely reason for denial (for example, missing documentation, eligibility issues, or coding errors)

This tool acts as an early warning system. It helps users quickly identify which claims are at risk of being denied, so they can be reviewed and corrected before they are submitted to the insurance company. This proactive step can help reduce claim rejection rates, save time, and avoid financial losses for healthcare organizations.

By turning a complex machine learning model into a simple, easy-to-use web interface, the project shows how technology can support better decision-making in healthcare billing.

**Results**

To determine which machine learning model worked best for predicting medical claim denials, four models were trained and tested using synthetic healthcare claims data to predict whether a claim will be approved or denied. If denied, the reason for the denial. In this project the dataset contains an imbalance between Approved and Denied claims (Approximately 80% approved and 20% denied). Using accuracy alone would be misleading, as a model could appear highly accurate simply by predicting most claims as approved. To better evaluate model performance on both classes—especially the minority class (Denied)—the F1-Score was used as the primary metric. F1-Score balances Precision (how many predicted Denied were correct) and Recall (how many actual Denied claims were detected), providing a more reliable assessment for imbalanced classification problems such as medical claim denials. The goal is to help healthcare providers identify potential denials and address issues proactively.

We need two models one for predicting whether the claim will be approved or rejected and the other one is for predicting the denial reason if the claim is rejected.

**Model 1**: For Predicting the Claim status (Approved/Rejected)

In order to overcome the class imbalance, the technique SMOTE is used. Used different models for training and the best model based on F1 score was chosen. The following models were trained and evaluated:

* Random Forest
* Logistic Regression
* Decision Tree
* XGBoost

Below is the performance of each model

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Model** | **Accuracy** | **Precision** | **Recall** | **F1-Score** |
| Random Forest | 0.6575% | 0.6759% | 0.6575% | 0.6662% |
| Logistic Regression | 0.5490% | 0.6832% | 0.5490% | 0.5930% |
| Decision Tree | 0.6100% | 0.6836% | 0.6100% | 0.6396% |
| XGBoost | 0.6485% | 0.6708% | 0.6485% | 0.6590% |

The Random Forest Classifier demonstrated the best F1-score and overall balanced performance on the test data. It was therefore selected as the final model for predicting the claim status. This model was exported and deployed as part of the prediction web interface.

**Model 2:** For predicting the Denial reason

A separate model is used for predicting the denial reason, if the claim is predicted to be rejected. The following models are trained and evaluated and the best model based on F1 score is selected.

* Random Forest
* Logistic Regression
* Decision Tree
* XGBoost

Below is the performance of each model.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Model** | **Accuracy** | **Precision** | **Recall** | **F1-Score** |
| Random Forest | 0.1439% | 0.0999% | 0.1439% | 0.1111% |
| Logistic Regression | 0.2333% | 0.0671% | 0.2333% | 0.0948% |
| Decision Tree | 0.0844% | 0.0898% | 0.0844% | 0.0868% |
| XGBoost | 0.1241% | 0.0943% | 0.1241% | 0.1045% |

To improve the accuracy of denial reason prediction, the five most frequent denial reasons in the dataset were identified. These top five reasons were retained as individual categories, while all other less frequent reasons were grouped under a single category labeled “Other.” This grouping was implemented using a simple Python function that checked whether each denial reason was in the top five; if not, it was reassigned to the "Other" category. This strategy reduced the complexity of the classification task, allowing the model to focus on the most impactful denial patterns. As a result, model performance improved, showing higher F1-scores. The table below shows the performance of each model after applying this grouping strategy.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Model** | **Accuracy** | **Precision** | **Recall** | **F1-Score** |
| Random Forest | 0.3548% | 0.2447% | 0.3548% | 0.2676% |
| Logistic Regression | 0.4144% | 0.1717% | 0.4144% | 0.2428% |
| Decision Tree | 0.2382% | 0.2582% | 0.2382% | 0.2471% |
| XGBoost | 0.3251% | 0.2570% | 0.3251% | 0.2794% |

Based on the F1-Score, the XGBoost model outperformed the other classifiers for predicting grouped denial reasons. Therefore, XGBoost was selected as the final model for denial reason prediction.

Overall, the machine learning models demonstrated good potential in predicting both claim status and denial reasons using the synthetic claims dataset. The Random Forest model was selected for claim approval/denial prediction, while XGBoost was chosen for denial reason classification. These models have been integrated into the project’s web interface prediction, providing a practical decision-support tool to help healthcare providers anticipate and manage claim denials more effectively.

**Discussion:**

This project showed that machine learning can help healthcare organizations reduce the number of denied medical claims. By training models on claim data, the system was able to predict if a claim would be approved or denied and also suggest the possible reason for denial. This kind of prediction can help billing teams fix problems before sending claims to insurance companies, which can save time and reduce money loss.

The Random Forest model gave the best results for predicting whether a claim would be approved or denied. The XGBoost model worked best for predicting the denial reason, especially after grouping similar reasons together. Since there were fewer denied claims in the data, a technique called SMOTE was used to create more denied claim examples for training. This helped the models learn better and give more balanced results.

The final models were added to a simple web application where users can upload a file of new claims and get predictions right away. This tool is easy to use and does not require any technical knowledge. It helps healthcare staff quickly identify risky claims so they can correct them before submission.

However, the data used was synthetic, not real hospital data. This means the results may not fully match what happens in the real world. Also, the model was less accurate when predicting rare denial reasons because it didn’t have enough examples to learn from.

Even with these challenges, the project shows that machine learning can be a useful tool for preventing claim denials and improving how healthcare billing is managed.

**Conclusion and next steps**:

This project successfully developed a machine learning-based tool to predict medical claim denials using synthetic data. Two models were built: one to predict whether a claim would be approved or denied, and another to identify the likely reason for denial. These models were then integrated into a simple web application that allows users to upload claim files and receive instant predictions.

The project achieved its main goal — creating a working system that demonstrates how data science can support billing teams in reviewing claims before submission. While the models performed well on synthetic data, further work is needed to apply them in real healthcare settings.

Next Steps

To enhance and expand this work, the following steps are recommended:

* Use real-world claim data (under proper data privacy protections) to improve model reliability.
* Include more features, such as patient history, provider specialty, or payer policies, to improve accuracy.
* Improve denial reason prediction, especially for rare categories, by collecting more balanced data.
* Add single-claim prediction to the web app, so users can check individual claims in real time.
* Conduct testing with billing teams to gather user feedback and evaluate practical usefulness.

This work shows that machine learning can be a helpful tool in healthcare billing supporting faster, more accurate claim submission and reducing preventable denials.

**Technologies Used**:

The following tools and technologies were used throughout this project:

* Python – for data preprocessing, model training, evaluation, and deployment
* Pandas and NumPy – for data cleaning and manipulation
* Scikit-learn – for building machine learning models and applying SMOTE
* XGBoost – for training the gradient boosting model
* Matplotlib and Seaborn – for data visualization and EDA plots
* Flask – to build the web application for batch prediction
* Jupyter Notebook – for developing and testing the code
* VS Code – as the main development environment

**Bibliography:**

1. Alkire MJ. Unnecessary insurance claim denials compromise patient care and provider bottom lines. *Stat* [Internet]. 2024 May 1 [cited 2025 Jun 25].
2. Simbo.ai. The evolution of machine learning in medical billing: identifying trends & improving claim denial rates through historical data analysis [Internet]. 2025 Jun [cited 2025 Jun 25].
3. Chaturvedi R, Sharma S, Pandian PKGP, Sharma S. Leveraging machine learning to predict and reduce healthcare claim denials [Internet]. Zenodo; 2023 [cited 2025 Jun 25].
4. Johnson M, Albizri A, Harfouche A. Responsible artificial intelligence in healthcare: predicting and preventing insurance claim denials for economic and social wellbeing.
5. Singh A, Verma B. Automating medical claims adjudication using NLP. *IEEE Trans Comput Intell AI Healthc*. 2024
6. Adams L. Conversational AI in healthcare claims processing. In: *Proc IEEE Int Conf Artif Int Med*; 2023.
7. Hameed A. The growing role of NLP [Internet]. LinkedIn; [cited 2025 Jun 25].
8. Chen P. Impact of robotic process automation on medical claims processing. *IEEE Trans Autom Sci Eng*. 2024.
9. Sharma K, Das R. Enhancing claims processing efficiency using AI-integrated RPA. In: *Proc IEEE Int Conf Intell Syst Appl*; 2023.
10. Nguyen H, Lee S. Challenges in implementing RPA for legacy healthcare systems. In: *Proc IEEE Eng Med Biol Soc Conf*; 2023.
11. Johnson M, Brown D. Blockchain-enabled smart contracts for healthcare claim settlements. *IEEE Trans Blockchain*. 2024.
12. Wilson C. Decentralized claims processing: a blockchain framework. *IEEE Trans Emerg Top Comput*. 2024.
13. Robinson T, Kim F. Scalability issues in blockchain adoption for healthcare claims. *IEEE Trans Cloud Comput*. 2023.
14. Carter D, Zhao H. Standardizing data exchange protocols for automated claims processing. *IEEE Trans Health Inform*. 2024.

**Acknowledgement:**

I would like to sincerely thank all those who supported and guided me throughout this project.

First and foremost, I am deeply grateful to my capstone mentor, Mr. Krish from CitiusTech, for his valuable guidance, consistent encouragement, and insightful feedback throughout the project.

I would also like to express my sincere appreciation to the Professor. Beth Meyers, my project coordinator, for her support, coordination, and helpful input at each stage of this work.

Many thanks to the faculty of the University of Minnesota, Master of Health Informatics program, for equipping me with the skills and knowledge needed to complete this project.

I am especially thankful to my family for their constant encouragement and support during this journey.

Lastly, I would like to acknowledge the open-source community and the tools that made this project possible, including Python, Flask, Scikit-learn, and Jupyter Notebook.

**Appendices:**

Appendix A: Sample claim data record (CSV format)

A screenshot of a computer

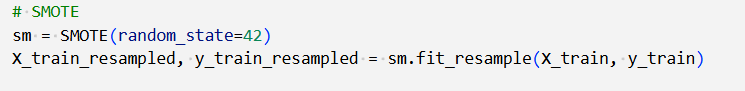
AI-generated content may be incorrect.

Appendix B: Features description

* **claim\_id**: Unique identifier for each claim
* **patient\_id**: Unique identifier for the patient
* **patient\_age**: Age of the patient at the time of service
* **patient\_gender**: Gender of the patient (e.g., Male, Female)
* **specialty**: Medical specialty of the provider (e.g., Cardiology, Orthopedics)
* **provider\_id**: Unique identifier for the healthcare provider
* **procedure\_code**: Code for the medical procedure performed (CPT code)
* **diagnosis\_code**: Code representing the medical diagnosis (ICD-10 code)
* **service\_date**: Date the healthcare service was provided
* **claim\_submission\_date**: Date the claim was submitted to the insurer
* **claim\_amount**: Billed amount for the service in USD
* **is\_denied**: Binary value indicating if the claim was denied (1 = denied, 0 = approved)
* **claim\_status**: Final status of the claim (e.g., Approved, Denied)
* **denial\_reason**: Text description of why the claim was denied (if applicable)
* **denial\_category**: Grouped reason for denial (e.g., Eligibility, Documentation, Coding)
* **insurance\_provider**: The insurance company processing the claim

Appendix C: Key code snippets

* 1. Handling class imbalance with SMOTE



* 1. Grouping less frequent denial reasons

A computer screen shot of a computer code

AI-generated content may be incorrect.

1. Saving the best models for both, predicting status and the denial reason

A screenshot of a computer code

AI-generated content may be incorrect.

A computer code with text

AI-generated content may be incorrect.

Appendix D: Web application Interface

A web application was planned for future deployment using Flask. Although deployment was not completed online, the application was tested locally using Flask in a development environment like Jupyter Notebook or VS Code. Screenshots are added below :

* Claim Upload Interface

A screenshot of a computer

AI-generated content may be incorrect.

We can choose the new csv file and click on upload and predict.

* Prediction Results Display

A screenshot of a computer screen

AI-generated content may be incorrect.

After uploading we get a link to download the CSV file with predicted results.

Appendix E: Glossary of Key Terms

* Claim: A request for payment submitted to an insurance company for services provided to a patient.
* Denial: Rejection of a claim by the insurer due to errors, missing info, or eligibility issues.
* SMOTE: Synthetic Minority Over-sampling Technique; a way to balance datasets where one class is much smaller.
* Precision: How many of the claims predicted as denied were actually denied.
* Recall: How many of the actual denied claims were correctly predicted.
* F1-Score: A balanced measure that considers both precision and recall.

Appendix F: Project GitHub Repository

It contains all the coding part for this project

<https://github.com/Loukyakilari/Predicting-Medical-claim-Denials>