**Functional Specification**. The document should have the following sections:

1. Background. The problem being addressed.

Provider Fraud is one of the biggest problems facing Medicare. According to the government, the total Medicare spending increased exponentially due to frauds in Medicare claims. Healthcare fraud is an organized crime which involves peers of providers, physicians, beneficiaries acting together to make fraud claims.

Rigorous analysis of Medicare data has yielded many physicians who indulge in fraud. They adopt ways in which an ambiguous diagnosis code is used to adopt costliest procedures and drugs. Insurance companies are the most vulnerable institutions impacted due to these bad practices. Due to this reason, insurance companies increased their insurance premiums and as result healthcare is becoming a costly matter day by day.

Healthcare fraud and abuse take many forms. Some of the most common types of frauds by providers are:

a) Billing for services that were not provided.

b) Duplicate submission of a claim for the same service.

c) Misrepresenting the service provided.

d) Charging for a more complex or expensive service than was actually provided.

e) Billing for a covered service when the service actually provided was not covered.

Problem Statement

The goal of this project is to "predict the potentially fraudulent providers" based on the claims filed by them along with this, we will also discover important variables helpful in detecting the behavior of potentially fraud providers. Further, we will study fraudulent patterns in the provider's claims to understand the future behavior of providers.

We are trying to build an infrastructure for the Government to identify fraudulent providers and also get insights into frauds in specific facilities, by specific providers and in accordance with specific physicians.

1. User profile. Who uses the system. What they know about the domain and computing (e.g., can browse the web, can program in Python)

1) Government Medicare Agency:

This project will potentially be used to create a machine learning model that predicts the likelihood of a claim being fraudulent. The government medicare agency responsible for accepting/rejecting the claims could then use this model to flag potentially fraudulent claims for further investigation, saving time and resources that would otherwise be spent manually reviewing all claims.

Assume the model identifies a claim with a high likelihood of fraud. In that case, the government may prioritize the review of that claim, potentially preventing fraudulent payments. The model could also be used to assign each claim a fraud score, which represents the likelihood of fraudulent activity. This score could be used to prioritize claims for review or to inform payment processing decisions.

Thus the government could take targeted actions to prevent fraudulent claims from being paid out by analyzing claims data and identifying patterns of behavior that may indicate fraudulent activity.

Assume the analysis uncovers a group of providers who routinely bill for costly procedures or drugs that are not medically necessary. In that case, the government could launch an investigation into these providers, potentially resulting in the recovery of millions of dollars in fraudulent claims.

Input:

Our machine learning model, which will be used to predict potentially fraudulent claims. The input will include data columns related to the claim, such as:

Output:

For each claim, the model will produce a probability score or a binary classification (fraudulent or not fraudulent) (based on supervised or unsupervised algorithm). Additional insights, such as which variables were most important in determining the likelihood of fraud for a given claim, may be included in the model.

2) Developer:

The developer is the one who is responsible to maintain the model and ensure its proper functionality. The developer is technically proficient.

1. The developer regularly updates the fraud detection model incase of the availability of new features.

2. The developer updates the visualization based on updated stakeholder requirements.

1. Data sources. What data you will use and how it is structured.  
     
     
     
   <https://www.kaggle.com/code/rohitrox/medical-provider-fraud-detection/data?select=Train_Outpatientdata-1542865627584.csv>   
     
   <https://www.kaggle.com/code/rohitrox/medical-provider-fraud-detection/data?select=Train_Inpatientdata-1542865627584.csv>   
     
   <https://www.kaggle.com/code/rohitrox/medical-provider-fraud-detection/data?select=Train_Beneficiarydata-1542865627584.csv>   
     
   <https://www.kaggle.com/code/rohitrox/medical-provider-fraud-detection/data?select=Train-1542865627584.csv>
2. Use cases. Describing at least two use cases. For each, describe: (a) the objective of the user interaction (e.g., withdraw money from an ATM); and (b) the expected interactions between the user and your system. Additional use cases are recommended, if any!

Use Case 1: Identifying High-Risk Claims for Review

Objective: The objective of this use case is to identify claims that are at a high risk of being fraudulent and flag them for further review and investigation.

Expected Interactions: The patient would input claim data into the provider portal/claim form and then this claim form will be then submitted to the government medicare agency for checking the claim. The government medical agency can call our API where our machine learning model is hosted to check the probability of a claim being fraudent. The model would analyze the data and generate a fraud score or probability for each claim. The government medicare agency would then review the claims with the highest fraud scores and prioritize them for further investigation, potentially leading to the prevention of fraudulent payments.

Use Case 2: Claims Prioritization for Payment Processing

The goal of this use case is to prioritize claims for payment processing based on their likelihood of fraud, thereby lowering the risk of fraudulent payments.

Interactions to be Expected: he patient would input claim data into the provider portal/claim form and then this claim form will be then submitted to the government medicare agency for checking the claim. The government medical agency can call our API where our machine learning model is hosted to check the probability of a claim being fraudent. For each claim, the model would analyze the data and generate a fraud score or probability. The user would then prioritize claims with low fraud scores for payment processing, potentially lowering the risk of fraudulent payments and saving taxpayer money.

Use case 3:

Checking Data Visualization for Medicare Claims as an Example

The goal of this use case is to examine the visualization of data related to Medicare claims in order to identify trends, patterns, and outliers that may indicate fraudulent activity.

Interactions to be Expected: The user (government) will access a dashboard or web-based interface that displays various visualization tools, such as charts, graphs, and maps, that display data related to Medicare claims. The user would interact with the visualization tools, choosing different data points, filters, and parameters to view the data from various angles. The user would be able to drill down into specific areas of interest and thoroughly examine the data to identify trends and patterns that could indicate fraudulent activity. The visualization tools could also be used to compare different data sets, such as claims from different providers or regions, to identify disparities or anomalies that need to be investigated further.Finally, the visualizations could be downloaded or exported for further analysis and reporting. Overall, the goal of this use case is to provide the government with a user-friendly and interactive way to visually explore and analyze Medicare claims data, allowing them to identify potential fraud and abuse more efficiently.