

Societe Generale Approach

Team Members:

1. Sakshi Suryawanshi - C22021772803 (ENTC)
2. Sakshi Chaudhari-C22021222303(COMP)

Theme: Cybersecurity (ML based healthcare fraud detection)

Healthcare fraud, particularly Provider Fraud, has emerged as a critical issue affecting Medicare, leading to exponential increases in total Medicare spending. This organized crime involves collusion among providers, physicians, and beneficiaries to submit fraudulent claims, impacting the healthcare system profoundly. Rigorous analysis of Medicare data has exposed physicians engaged in fraudulent practices, such as manipulating diagnosis codes to bill for costlier procedures and drugs. Insurance companies, being vulnerable to these malpractices, respond by raising premiums, making healthcare increasingly expensive.

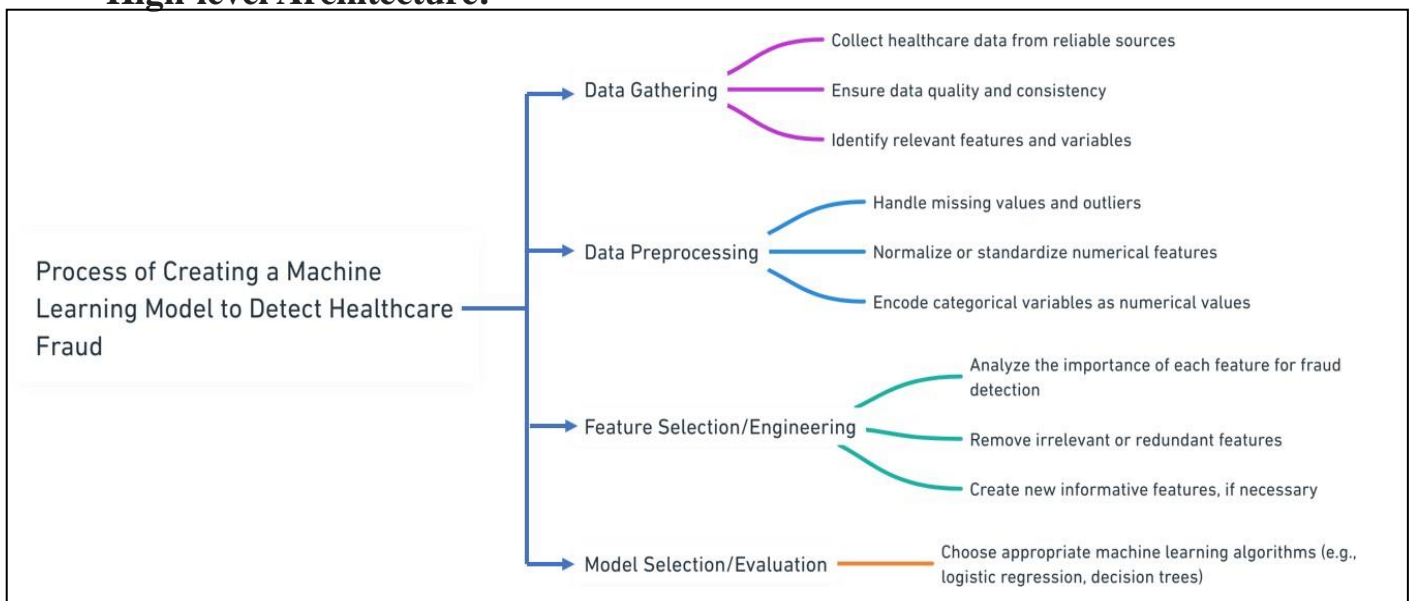
This project aims to predict potentially fraudulent providers based on their claims and identify essential variables for detecting such behaviour. Additionally, the study seeks to analyze fraudulent patterns in providers' claims to anticipate future fraudulent activities.

Dataset:

The dataset for this project comprises Inpatient claims, Outpatient claims, and Beneficiary details for each provider.

- The Inpatient data reveals insights into claims filed for admitted patients, including admission and discharge dates and diagnosis codes.
- The Outpatient data provides information on claims filed for non-admitted patients.
- The Beneficiary Details data contains KYC information, such as health conditions and regional affiliation.

High-level Architecture:



Thought process behind the solution:-

During the Covid-19 pandemic, we have seen cases of fraud in healthcare sector. Recognizing the prevalence of healthcare fraud and its detrimental impact on patients and resources, we were compelled to find a solution. The potential to leverage technology for fraud detection, improve resource allocation, and enhance patient care became our driving motivation. With a vision of a more transparent and efficient healthcare system, we embarked on the journey to develop a predictive model that would contribute to a healthier and more trustworthy medical environment.

Proof of concept :- <https://www.william-russell.com/blog/insurance-fraud-covid-creates-new-claims/>

Outline/Approach:

1. Firstly, we merged all the datasets together and then we removed all the null values from the dataset and dropped the columns with missing values. After that, we explored the relations between the features and the target "Possible Fraud" variable.
2. Next, we split the data into train and test and then used Logistic regression & Random forest classification to classify the trained model on the basic or potential fraud.
3. We cross-checked the predicted values against the tested "Possible Fraud" variable and evaluated the accuracy of the model using plots.

Methodologies used:

Logistic regression:- We have selected this because of its efficiency of detecting frauds based on its ability to isolate the data belonging to different binary classes. It also provides easily interpretable results, it gives coefficients for each feature indicating strength and direction of relationship between the independent variables and probability of fraud.

Random forest:- It is another powerful supervised algorithm that uses decision trees for classification. We used this to obtain confusion matrix on which its performance is evaluated.

Future Scope:-

1. We can improve the accuracy of the model by using more complex machine learning techniques like Support Vector Machine (SVM), Naive Bayes, etc.
2. We can build a model which would predict values based on real-time data which would help provide a broader scope.
3. Integration with Claims Processing Systems: Integrate your model with healthcare claims processing systems. This integration could automatically flag potential fraud cases and provide recommendations for further investigation.
4. We can also implement our model in already existing fraud detection software which would further enhance their performance.
5. Continuous benchmarking of the model against other fraud detection systems in order to constantly keep track of its efficiency.