Anomaly Detection

Dataset:- [Healthcare Providers dataset](#)

Provider Fraud is one of the biggest problems facing Medicare. Healthcare fraud is an organized crime which involves peers of providers, physicians acting together to make fraud claims.

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:

Objective: The goal of this project is to " predict the potentially fraudulent providers " based on the given features in the dataset. Along with this, we will also discover important variables helpful in detecting the behaviour of potentially fraud providers. further, we will study fraudulent patterns in the provider's claims to understand which are the normal transaction and fraudulent transaction i.e. we will detect anomalies in the dataset.

About Dataset

There are 100k rows and 27 columns in the dataset.
Try out various unsupervised techniques to find the anomalies in the data.

Detailed Data File:

The following variables are included in the detailed Physician and Other Supplier data file (see Appendix A for a condensed version of variables included)).

npi – National Provider Identifier (NPI) for the performing provider on the claim. The provider NPI is the numeric identifier registered in NPPES.

nppes_provider_last_org_name – When the provider is registered in NPPES as an individual (entity type code='I'), this is the provider's last name. When the provider is registered as an organization (entity type code = 'O'), this is the organization's name.

nppes_provider_first_name – When the provider is registered in NPPES as an individual (entity type code='I'), this is the provider's first name. When the provider is registered as an organization (entity type code = 'O'), this will be blank.

nppes_provider_mi – When the provider is registered in NPPES as an individual (entity type code='I'), this is the provider's middle initial. When the provider is registered as an organization (entity type code= 'O'), this will be blank.

nppes_credentials – When the provider is registered in NPPES as an individual (entity type code='I'), these are the provider's credentials. When the provider is registered as an organization (entity type code = 'O'), this will be blank.

nppes_provider_gender – When the provider is registered in NPPES as an individual (entity type code='I'), this is the provider's gender. When the provider is registered as an organization (entity type code = 'O'), this will be blank.

nppes_entity_code – Type of entity reported in NPPES. An entity code of 'I' identifies providers registered as individuals and an entity type code of 'O' identifies providers registered as organizations.

nppes_provider_street1 – The first line of the provider's street address, as reported in NPPES.

nppes_provider_street – The second line of the provider's street address, as reported in NPPES.

nppes_provider_city – The city where the provider is located, as reported in NPPES.

nppes_provider_zip – The provider's zip code, as reported in NPPES.

nppes_provider_state – The state where the provider is located, as reported in NPPES. The fifty U.S. states and the District of Columbia are reported by the state postal abbreviation. The following values are used for all other areas:

'XX' = 'Unknown'
'AA' = 'Armed Forces Central/South America'
'AE' = 'Armed Forces Europe'
'AP' = 'Armed Forces Pacific'
'AS' = 'American Samoa'
'GU' = 'Guam'

'MP' = 'North Mariana Islands'

'PR' = 'Puerto Rico'

'VI' = 'Virgin Islands'

'ZZ' = 'Foreign Country'

nppes_provider_country – The country where the provider is located, as reported in NPPES. The country code will be 'US' for any state or U.S. possession. For foreign countries (i.e., state values of 'ZZ'), the provider country values include the following:

AE=United Arab Emirates IT=Italy

AG=Antigua JO= Jordan

AR=Argentina JP=Japan

AU=Australia KR=Korea

BO=Bolivia KW=Kuwait

BR=Brazil KY=Cayman Islands

CA=Canada LB=Lebanon

CH=Switzerland MX=Mexico

CN=China NL=Netherlands

CO=Colombia NO=Norway

DE= Germany NZ=New Zealand

ES= Spain PA=Panama

FR=France PK=Pakistan

GB=Great Britain RW=Rwanda

GR=Greece SA=Saudi Arabia

HU= Hungary SY=Syria

IL= Israel TH=Thailand

IN=India TR=Turkey

IS= Iceland VE=Venezuela

provider_type – Derived from the provider specialty code reported on the claim.

medicare_participation_indicator – Identifies whether the provider participates in Medicare and/or accepts the assigned assignment of Medicare allowed amounts.

place_of_service – Identifies whether the place of service submitted on the claims is a facility (value of 'F') or non-facility (value of 'O'). Non-facility is generally an office setting; however other entities are included in non-facility.

hcpcs_code – HCPCS code used to identify the specific medical service furnished by the provider.

hcpcs_description – Description of the HCPCS code for the specific medical service furnished by the provider.

hcpcs_drug_indicator –Identifies whether the HCPCS code for the specific service furnished by the provider is an HCPCS listed on the Medicare Part B Drug Average Sales Price (ASP) File.

line_srvc_cnt – Number of services provided; note that the metrics used to count the number provided can vary from service to service.

bene_unique_cnt – Number of distinct Medicare beneficiaries receiving the service.

bene_day_srvc_cnt – Number of distinct Medicare beneficiary/per day services.

average_Medicare_allowed_amt – Average of the Medicare allowed amount for the service.

stdev_Medicare_allowed_amt – Standard deviation of the Medicare allowed amounts.

average_submitted_chrg_amt – Average of the charges that the provider submitted for the service.

stdev_submitted_chrg_amt – Standard deviation of the charge amounts submitted by the provider.

average_Medicare_payment_amt – Average amount that Medicare paid after deductible and coinsurance amounts have been deducted for the line item service.

Milestones:

1: Data collection and exploration
2: EDA(Univariate, bivariate analysis)
3: Visualization
4: Machine learning algorithms
5: Deep learning approach
6: Presentation and docs