

DeepSphere.AI
Enterprise AI and IIoT for Analytics

HEALTH CARE CLAIM DETECTION

MACHINE LEARNING MODEL IMPLEMENTATION



Machine Learning
Applied in
Healthcare

DeepSphere. AI

DeepSphere.AI Is the Most Simple, Easy to Use, and Applied Artificial Intelligence Educational Platform. Our Platform Offers Two Unique Products to Reinvent Education to the Next Level AI Learning and Digital Education Transformation.

DeepSphere.AI Educational Platform

Powered By



Google Cloud

AI Lab Infrastructure

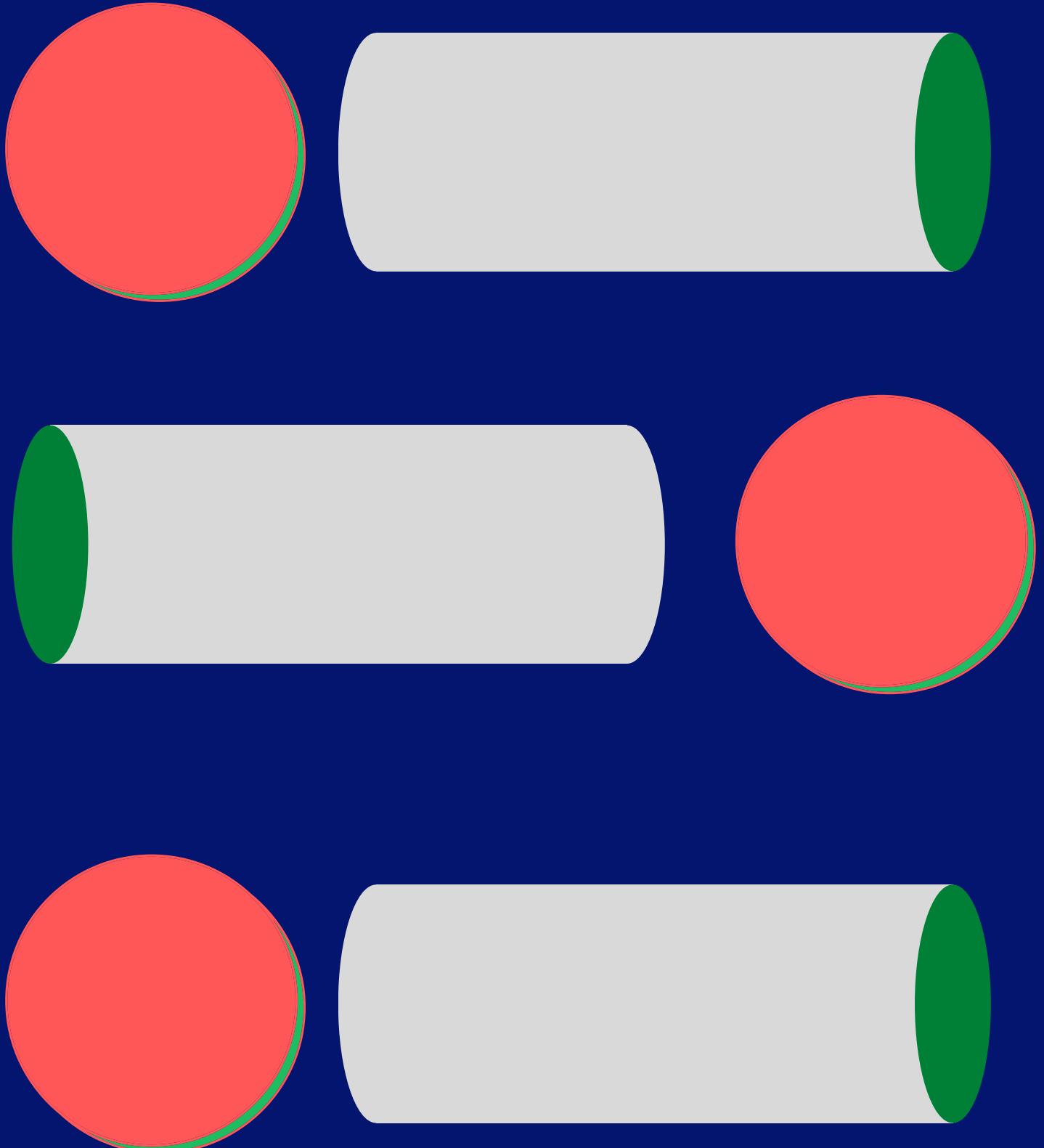
iLMS and Industrial Curriculum



Cloud Computing



Measurable Values of Data Science and Intelligent Automation



CONTENTS

- 1** Disclaimer
- 2** Executive Summary
- 3** Problem Statement
- 4** Business Context
- 5** High-level solution Flow
- 6** Model Building Blocks
- 7** GCP Components
- 8** Distributed Training Components
- 9** Model implementation steps
- 10** Visualization



DISCLAIMER

All software and hardware used or referenced in this guide belong to their respective vendor. We developed this guide based on our development infrastructure and this guide may or may not work on other systems and technical infrastructure. We are not liable for any direct or indirect problems caused by users using this guide.



EXECUTIVE SUMMARY

The purpose of this document is to provide information to the users about the machine learning model. in order to find the solution for this problem used many techniques to identify whether a Patient is fraudulent or not.

"Our ability to manufacture fraud now exceeds our ability to detect it."





PROBLEM STATEMENT

- Health Care Fraud Detection Problem includes modeling past transactions and data with the knowledge of the Patients that turned out to be fraud. This model is then used to identify whether a new Patient is fraudulent or not.
- Our aim here is to detect 100% of the fraudulent Patients while minimizing the incorrect fraud classifications.
- Calculated probabilities, weightage, and confidence level for different features.
- Visualization for graphic representation of output data.

BUSINESS CONTEXT

BUSINESS CHALLENGES

- The detection of fraud would improve strive for an optimal effort per situation.
- One of the major challenges in healthcare fraud is the lack of knowledge by the general public on this specific issue.
- some of the general public believe it is too costly to detect healthcare fraud and would
- rather just accept the amount of fraud as a loss in the process of doing everyday business.

BUSINESS BENEFITS

- If education is provided in this area of fraud, it should lead to increased enforcement of rules in the medical environment and establish adequate punishment of the guilty parties.
- opportunity and rationalization are the main causes of fraud they can become more educated in ways to prevent it.
- With the combination of implementation, understandability, awareness, and education; Healthcare fraud can be stopped

Machine Learning Technologies in Scope

There are several deep learning technical and functional components involved in implementing this model. Here are the critical building blocks to implement the deep learning model.



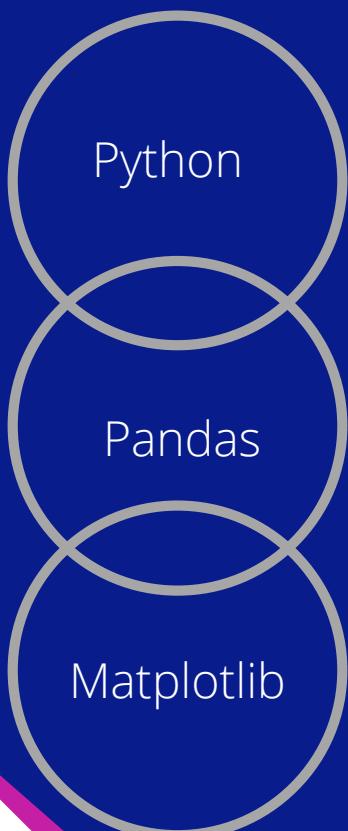
Google Cloud Platform



Kubeflow



PySpark



sklearn



Scipy

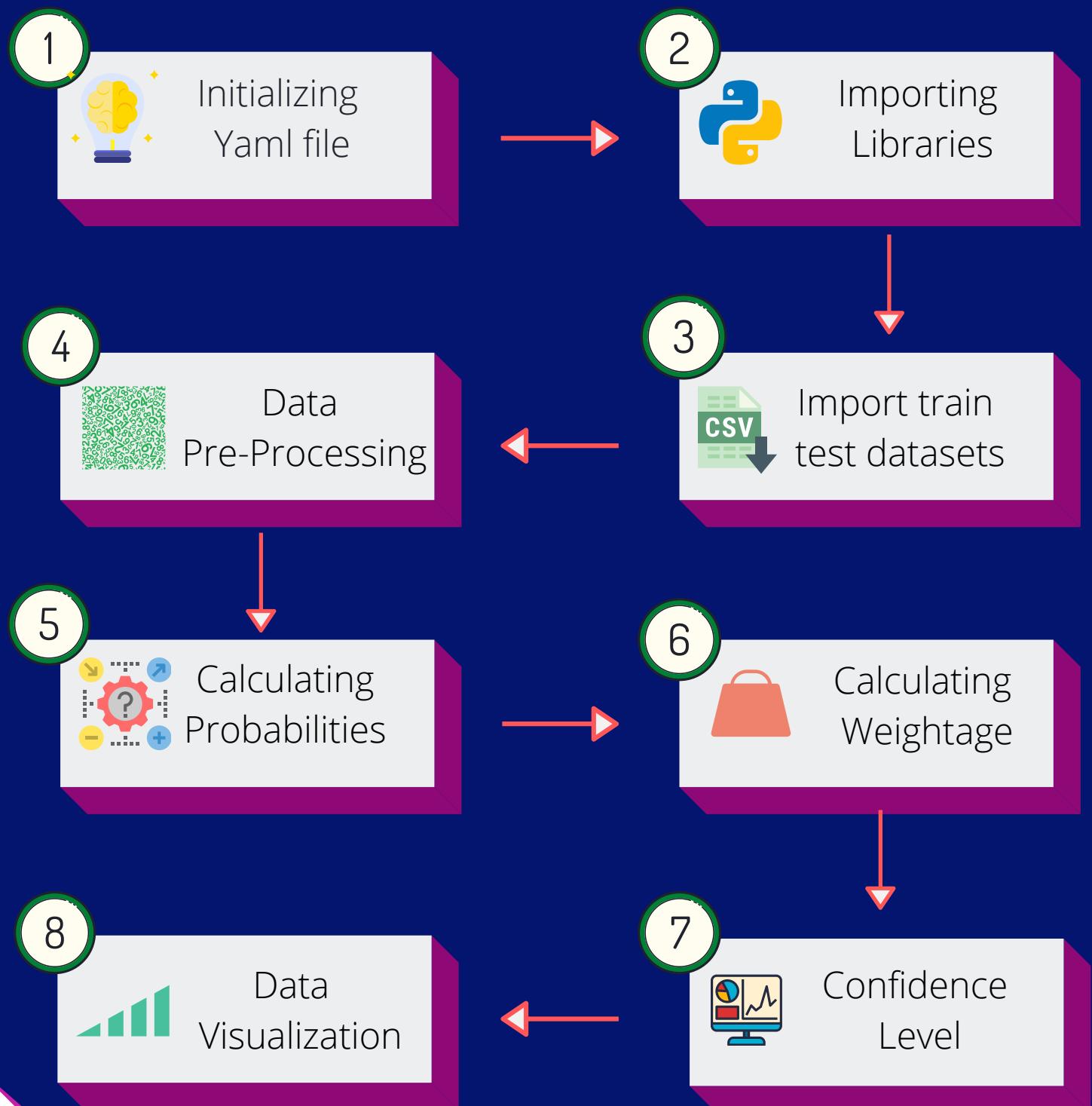


NumPy

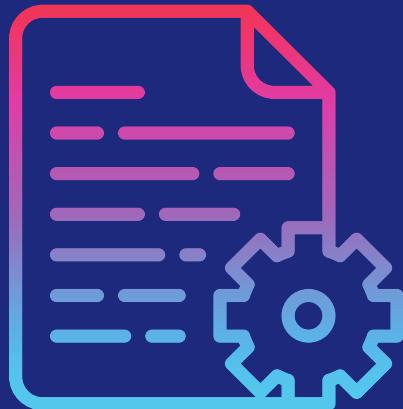


HIGH LEVEL SOLUTION FLOW

Machine learning model implementation is used to solve a problem involving several steps. Here are the key steps that we can learn from this guide.



INITIALIZING YAML FILE



```
import yaml
import os
yaml_file_path='C:\\\\AI\\\\FRAUD_DETECTION\\\\ML\\\\INI\\\\config.yaml'
with open (yaml_file_path) as file:
    config=yaml.safe_load(file)
```

- YAML is a human-readable data serialization standard that can be used in conjunction with all programming languages and is often used to write configuration files.
- Here, We took the train and test datasets for this model and answer, ans_data and charts are used for the output.
- Here Yaml is the application that needs to store the data or transmits the data, where os is used to perform this operation in any Windows, Ubuntu, Linux and Mac.

IMPORTING LIBRARIES



Pandas

pandas is a software library written for the Python programming language for data manipulation and analysis.

NumPy arrays are stored at one continuous place in memory unlike lists, so processes can access and manipulate them very efficiently.



Numpy



Matplotlib

Matplotlib is a plotting library in python. We can generate bar graphs, histograms and pie charts, etc and interactive visualizations.



Sklearn



Scipy

Scipy is a open-source python library used for scientific and technical computing. It contains modules such as linear algebra, and optimization.

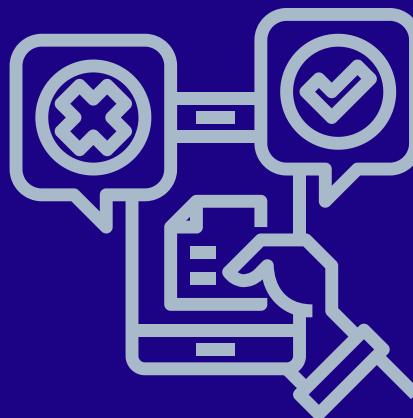


DeepSphere.AI
Enterprise AI and IIoT for Analytics

IMPORTING DATASETS

CSV (comma-separated value) files are a common file format for transferring and storing data. The ability to read, manipulate, and write data to and from CSV files using Python is a key skill to master for any data scientist or business analysis. we are importing our train data Using Pandas

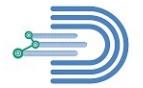
PROVIDER	PATIENT_ID	DISEASE	SYMPTOMS	MEDICATION	CLAIM_NO	BILL_AMOUNT	HEALTH_CARE_CARD_ID	C
ZERIC121	1202Sara	Dengue	Haemorrhage (severe bleeding), blood plasma le...	Acetaminophen	1	30000	LL3379	
XHO121	1395Sada	Dengue	Haemorrhage (severe bleeding), blood plasma le...	Acetaminophen	1	30000	LL3401	
DHARMA121	1822Varm	Nearsightedness	Blurry vision while looking at distant objects...	Corrective lense	1	6000	LL3081	
XHO121	1430Hars	Dengue	Haemorrhage (severe bleeding), blood plasma le...	Acetaminophen	1	30000	LL3574	
HEALTHY121	1106Anag	Dengue	Haemorrhage (severe bleeding), blood plasma le...	Acetaminophen	1	30000	LL3948	
ZERIC121	1585Gabo	Nearsightedness	Blurry vision while looking at distant objects...	Corrective lense	1	6000	LL3681	
SSS121	1549Same	Dengue	Haemorrhage (severe bleeding), blood plasma le...	Acetaminophen	1	30000	LL3631	
DHARMA121	1435Mall	Dengue	Haemorrhage (severe bleeding), blood plasma le...	Acetaminophen	1	30000	LL3588	
XHO121	1374Shar	Dengue	Haemorrhage (severe bleeding), blood plasma le...	Acetaminophen	1	30000	LL3659	
ZERIC121	1287Swar	Dengue	Haemorrhage (severe bleeding), blood plasma le...	Acetaminophen	1	30000	LL3499	
SWASTH121	1974Lila	Dengue	Haemorrhage (severe bleeding), blood plasma le...	Acetaminophen	1	30000	LL3922	
SWASTH121	1466Gama	Migraine	Feeling nauseous and queasy;Discomfort in brig...	Ibuprofen	1	12000	LL3020	



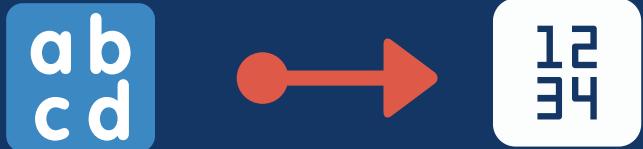
```
input_data["PROFESSION"] = input_data["PROFESSION"].astype('category')
input_data["ADDRESS"] = input_data["ADDRESS"].astype('category')
input_data["PAYER"] = input_data["PAYER"].astype('category')
input_data["MEDICATION"] = input_data["MEDICATION"].astype('category')
input_data["SYMPTOMS"] = input_data["SYMPTOMS"].astype('category')
input_data["CLAIMED_MEDICATION"] = input_data["CLAIMED_MEDICATION"].astype('category')
input_data["DISEASE"] = input_data["DISEASE"].astype('category')
input_data.dtypes
```

```
input_data["PROFESSION"] = input_data["PROFESSION"].cat.codes
input_data["ADDRESS"] = input_data["ADDRESS"].cat.codes
input_data["PAYER"] = input_data["PAYER"].cat.codes
input_data["MEDICATION"] = input_data["MEDICATION"].cat.codes
input_data["CLAIMED_MEDICATION"] = input_data["CLAIMED_MEDICATION"].cat.codes
input_data["SYMPTOMS"] = input_data["SYMPTOMS"].cat.codes
input_data["DISEASE"] = input_data["DISEASE"].cat.codes
input_data
```

- Categorical variables are known to hide and mask lots of interesting information in a data set. It's crucial to learn the methods of dealing with such variables.
- A categorical variable has too many levels. This pulls down the performance level of the model.
- some ML libraries do not take categorical variables as input. Thus, we convert them into numerical variables.



	Provider	Patient_ID	Disease	Symptoms	Medication	Claim_No	Bill_Amount	Health_Care_Card_Id	Gender
0	ZERIC121	1202Sara	Dengue	Haemorrhage (severe bleeding), blood plasma le...	Acetaminophen	1	30000	LL3379	Male
1	XHO121	1395Sada	Dengue	Haemorrhage (severe bleeding), blood plasma le...	Acetaminophen	1	30000	LL3401	Male
2	DHARMA121	1822Varm	Nearsightedness	Blurry vision while looking at distant objects...	Corrective lense	1	6000	LL3081	Female
3	XHO121	1430Hars	Dengue	Haemorrhage (severe bleeding), blood plasma le...	Acetaminophen	1	30000	LL3574	Female
4	HEALTHY121	1106Anag	Dengue	Haemorrhage (severe bleeding), blood plasma le...	Acetaminophen	1	30000	LL3948	Female
5	ZERIC121	1585Gabo	Nearsightedness	Blurry vision while looking at distant objects...	Corrective lense	1	6000	LL3681	Male
6	SSS121	1549Same	Dengue	Haemorrhage (severe bleeding), blood plasma le...	Acetaminophen	1	30000	LL3631	Female
7	DHARMA121	1435Mall	Dengue	Haemorrhage (severe bleeding), blood plasma le...	Acetaminophen	1	30000	LL3588	Male
8	XHO121	1374Shar	Dengue	Haemorrhage (severe bleeding), blood plasma le...	Acetaminophen	1	30000	LL3659	Male
9	ZERIC121	1287Swar	Dengue	Haemorrhage (severe bleeding), blood plasma le...	Acetaminophen	1	30000	LL3499	Female
10	SWASTH121	1974Lila	Dengue	Haemorrhage (severe bleeding), blood plasma le...	Acetaminophen	1	30000	LL3922	Male
11	SWASTH121	1466Gama	Migraine	Feeling nauseous and queasy, Discomfort in brig...	Ibuprofen	1	12000	LL3020	Male



Provider	Patient_ID	Disease	Symptoms	Medication	Claim_No	Bill_Amount	Health_Care_Card_Id	Gender	Insurance_Id	Join_Da
ZERIC121	1773Azar	10	1	3	1	6000	LL3470	Male	HEALTH2513	15-10-201
SSS121	1941Meha	10	1	3	1	6000	LL3885	Female	HEALTH2832	25-06-20
LMC121	1236Hari	6	5	2	1	35000	LL3364	Male	HEALTH2203	05-08-201
XHO121	1192Alak	3	3	5	1	40000	LL3511	Male	HEALTH2247	07-12-20
SWASTH121	1694Eela	4	7	0	1	30000	LL3064	Female	HEALTH2237	23-10-20
XHO121	1143Chan	4	7	0	1	30000	LL3052	Male	HEALTH2696	02-08-20
SSS121	1330Spoo	4	7	0	1	30000	LL3405	Female	HEALTH2785	02-03-202
SSS121	1502Tria	4	7	0	1	30000	LL3668	Male	HEALTH2035	09-07-201
SSS121	1987Joth	10	1	3	1	6000	LL3346	Female	HEALTH2635	03-06-20
LMC121	1022Amog	6	5	2	1	35000	LL3108	Male	HEALTH2667	16-07-20
SSS121	1126Hatt	4	7	0	1	30000	LL3679	Male	HEALTH2884	01-03-202
ZERIC121	1641Aato	4	7	0	1	30000	LL3228	Male	HEALTH2561	29-04-202
DHARMA121	1235Soll	4	7	0	1	30000	LL3598	Male	HEALTH2085	11-11

- Probability is a field of mathematics that quantifies uncertainty.
- Framing the problem as a prediction of class membership simplifies the modeling problem and makes it easier for a model to learn. It allows the model to capture ambiguity in the data, which allows a process downstream, such as the user to interpret the probabilities in the context of the domain.
- here, we detected the probability of Fraud and Non-Fraud for each feature such as DISEASE, SYMPTOMS,_MEDICATION, ADDRESS, BILL_AMOUNT, PROFESSION and PAYER separately

1.PROBABILITY OF DISEASE

DETECT_FRAUD	Probability of NON- DETECT_FRAUD	Probability of DETECT_FRAUD
1	0.414721	0.585279
1	0.425905	0.574095
0	0.701389	0.298611
0	0.835648	0.164352
0	0.655499	0.344501
0	0.748889	0.251111
0	0.744588	0.255412

- Detecting the probability of fraud and Non-Fraud with the help of DETECT_FRAUD .
- Here, we are detecting the probability of DISEASE.

2. PROBABILITY OF SYMPTOMS

DETECT_FRAUD	Probability of NON- DETECT_FRAUD	Probability of DETECT_FRAUD
0	0.726474	0.273526
0	0.782667	0.217333
0	0.771176	0.228824
0	0.885428	0.114572
0	0.580282	0.419718
0	0.762083	0.237917
0	0.853847	0.146153

- Detecting the probability of fraud and Non-Fraud with the help of DETECT_FRAUD.
- Here, we are detecting the probability of SYMPTOMS.

3. PROBABILITY OF BILL AMOUNT

DETECT_FRAUD	Probability of NON- DETECT_FRAUD	Probability of DETECT_FRAUD
0	0.769765	0.230235
0	0.814681	0.185319
0	0.743089	0.256911
0	0.886786	0.113214
0	0.636979	0.363021
0	0.740125	0.259875
0	0.849709	0.150291

- Detecting the probability of fraud and Non-Fraud with the help of DETECT_FRAUD.
- Here, we are detecting the probability of BILL AMOUNT.



4. PROBABILITY OF ADDRESS

DETECT_FRAUD	Probability of NON- DETECT_FRAUD	Probability of DETECT_FRAUD
0	0.762452	0.237548
0	0.871187	0.128813
0	0.751233	0.248767
0	0.814599	0.185401
0	0.664410	0.335590
0	0.753976	0.246024
0	0.846550	0.153450

- Detecting the probability of fraud and Non-Fraud with the help of DETECT_FRAUD.
- Here, we are detecting the probability of ADDRESS.

5. PROBABILITY OF MEDICATION

DETECT_FRAUD	Probability of NON-DETECT_FRAUD	Probability of DETECT_FRAUD
0	0.725253	0.274747
0	0.754685	0.245315
0	0.692726	0.307274
0	0.893522	0.106478
0	0.588117	0.411883
0	0.678492	0.321508
0	0.852684	0.147316

- Detecting the probability of fraud and Non-Fraud with the help of DETECT_FRAUD.
- Here, we are detecting the probability of MEDICATION.



6. PROBABILITY OF PROFESSION

DETECT_FRAUD	Probability of NON- DETECT_FRAUD	Probability of DETECT_FRAUD
0	0.765338	0.234662
0	0.866799	0.133201
0	0.760112	0.239888
0	0.829792	0.170208
0	0.640615	0.359385
0	0.765338	0.234662
0	0.837795	0.162205

- Detecting the probability of fraud and Non-Fraud with the help of DETECT_FRAUD.
- Here, we are detecting the probability of PROFESSION.



7. PROBABILITY OF PAYER

DETTECT_FRAUD	Probability of NON- DETTECT_FRAUD	Probability of DETTECT_FRAUD
0	0.756288	0.243712
0	0.850239	0.149761
0	0.756288	0.243712
0	0.850239	0.149761
0	0.629109	0.370891
0	0.756288	0.243712
0	0.850239	0.149761

- Detecting the probability of fraud and Non-Fraud with the help of DETECT_FRAUD.
- Here, we are detecting the probability of PAYER.

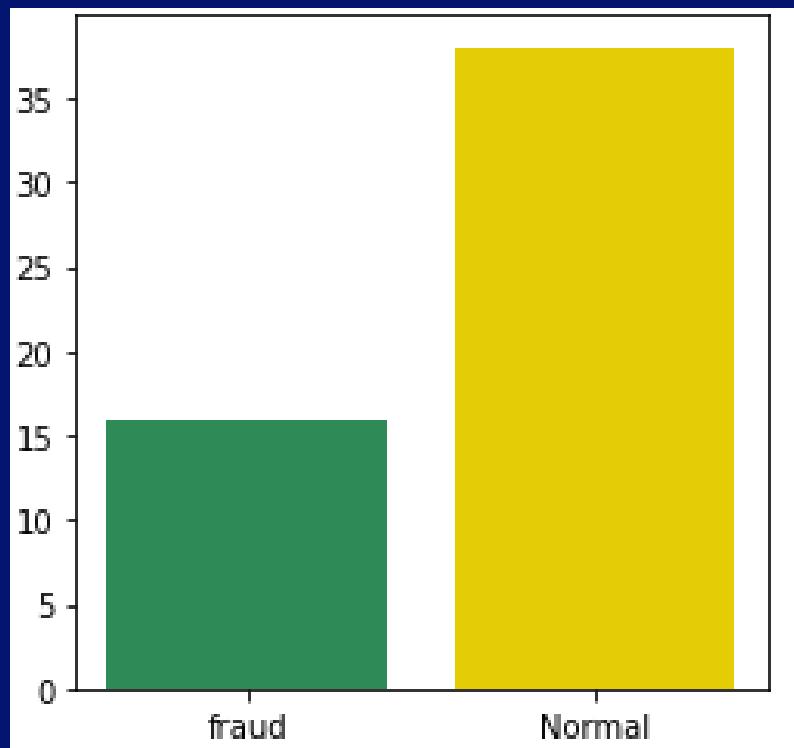


CALCULATING THE WEIGHTAGE

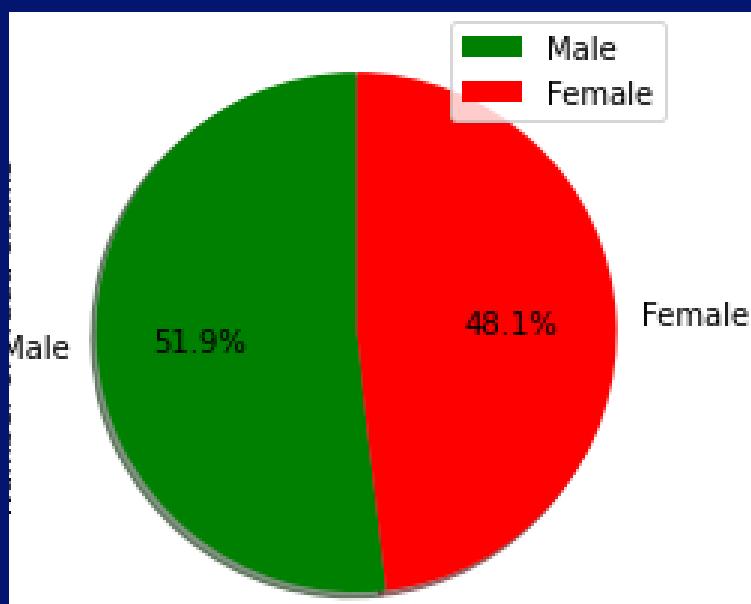
In fraud detection, false negatives tend to cost more than false positives. we can't go all the way to the extreme, because flagging every transaction as fraudulent is all clearly bad business decisions. But the point is, certain kinds of misclassification are more expensive than other kinds of misclassification, so we might want to influence our model to make decisions that are wrong more often, but still less expensive overall. Using weighting is a reasonable tool to help solve this problem.

Weightage of BILL_AMOUNT B_NDF	Weightage of ADDRESS A_NDF	Weightage of PROFESSION P_NDF	Weightage of DISEASE D_DF	Weightage of SYMPTOMS S_DF	Weightage of MEDICATION M_DF	Weightage of BILL_AMOUNT B_DF	Weightage of ADDRESS A_DF	Weightage of PROFESSION P_DF
0.007314	-0.002887	0.009050	0.482762	0.001221	-0.044512	0.007314	-0.002887	
-0.056505	0.004387	0.016561	0.632906	0.027982	-0.059996	-0.056505	0.004387	
-0.008145	-0.008878	0.003824	0.527464	0.078450	-0.050363	-0.008145	-0.008878	
0.072187	-0.015193	-0.020447	0.735666	-0.008094	0.006736	0.072187	-0.015193	
-0.027431	0.023795	0.011506	0.209390	-0.007836	-0.048862	-0.027431	0.023795	
-0.013851	-0.011362	0.009050	0.518371	0.083591	-0.061633	-0.013851	-0.011362	
0.003159	0.008755	-0.012444	0.704086	0.001163	0.002975	0.003159	0.008755	
0.040560	0.002692	-0.012328	0.525501	-0.010973	-0.007025	0.040560	0.002692	
0.058649	-0.015869	-0.012444	0.722977	-0.005002	-0.002834	0.058649	-0.015869	
-0.040907	0.020689	0.003824	0.471488	0.005668	-0.030361	-0.040907	0.020689	
0.150320	-0.000132	-0.023621	0.029041	-0.058150	-0.001571	0.150320	-0.000132	
-0.003243	0.018739	0.005159	0.057599	0.029844	0.044904	-0.003243	0.018739	
-0.007447	0.003043	0.009050	0.476824	0.003303	-0.043701	-0.007447	0.003043	
0.011257	0.009053	-0.016407	0.711738	-0.001190	0.008547	0.011257	0.009053	
0.051383	-0.013894	-0.017868	0.572295	0.052357	-0.012259	0.051383	-0.013894	
-0.066110	0.019258	0.037648	0.313069	0.081719	-0.017663	-0.066110	0.019258	

Fraud and Normal

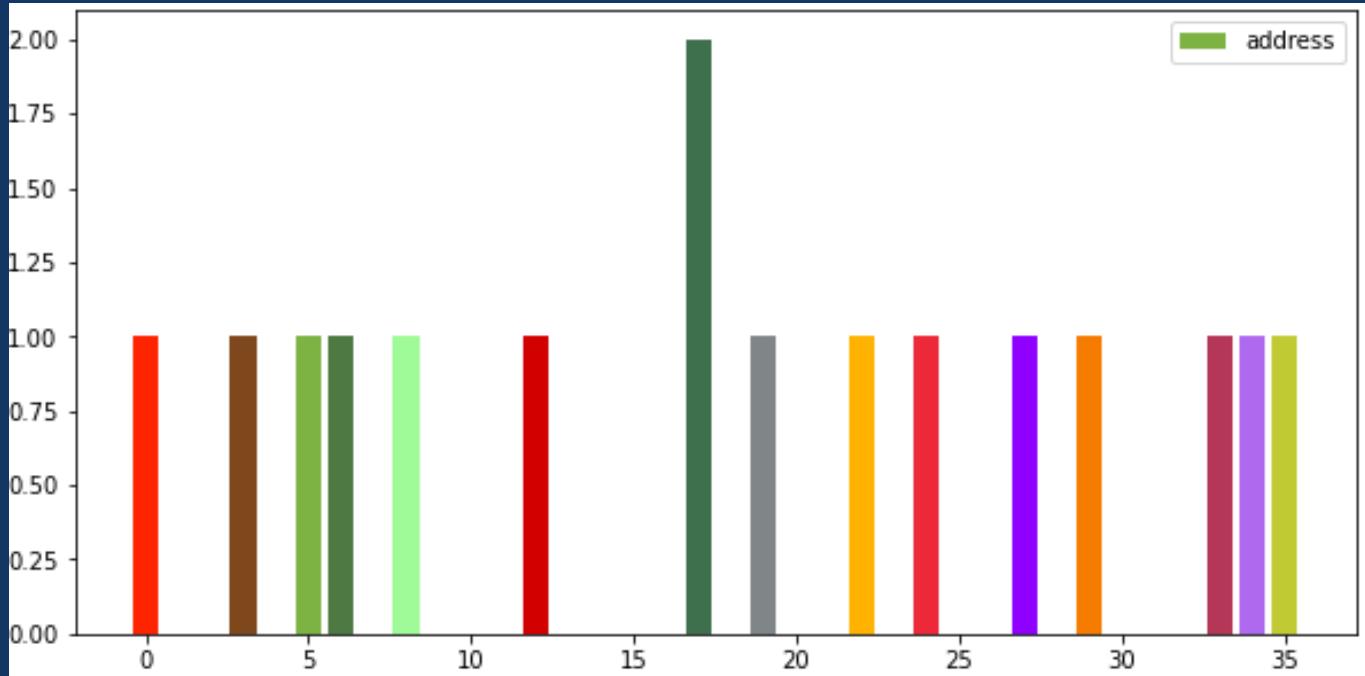


Male and Female

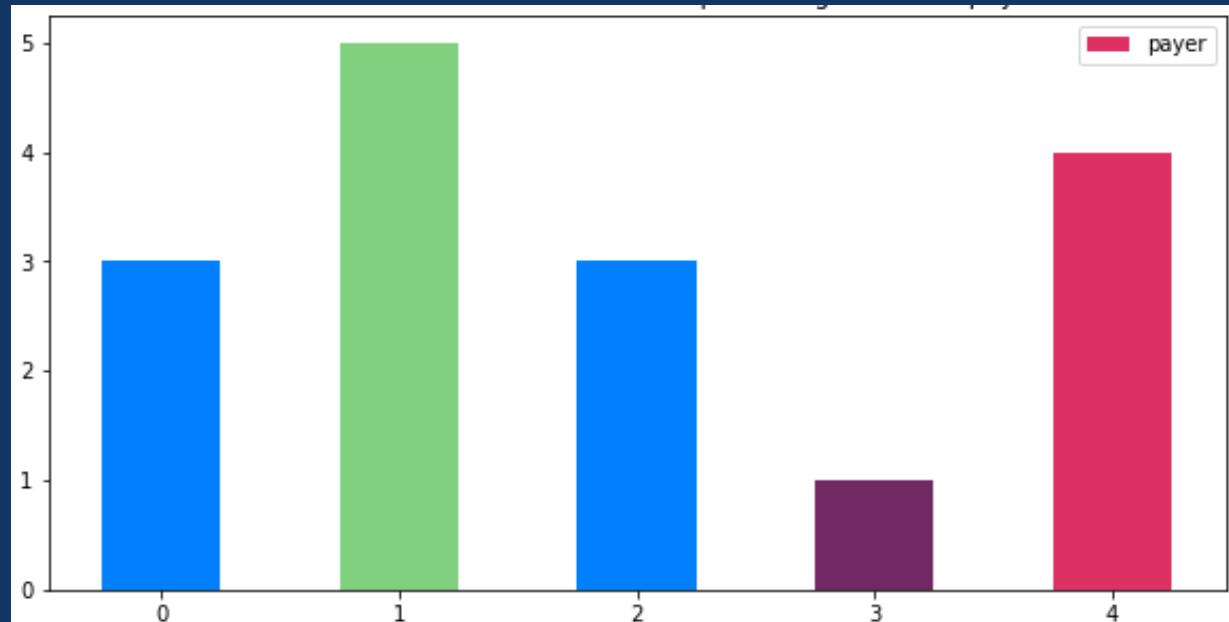




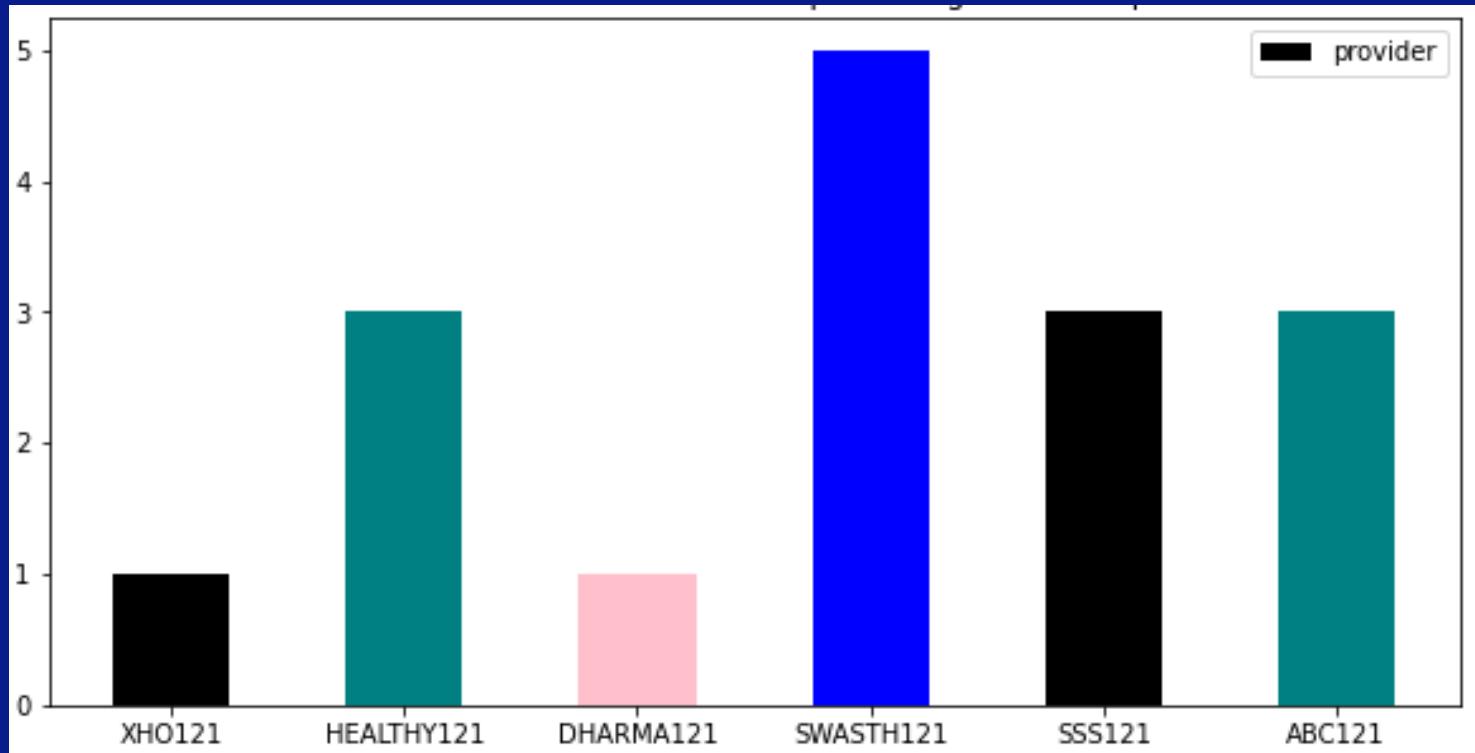
Detected Fraud in Address



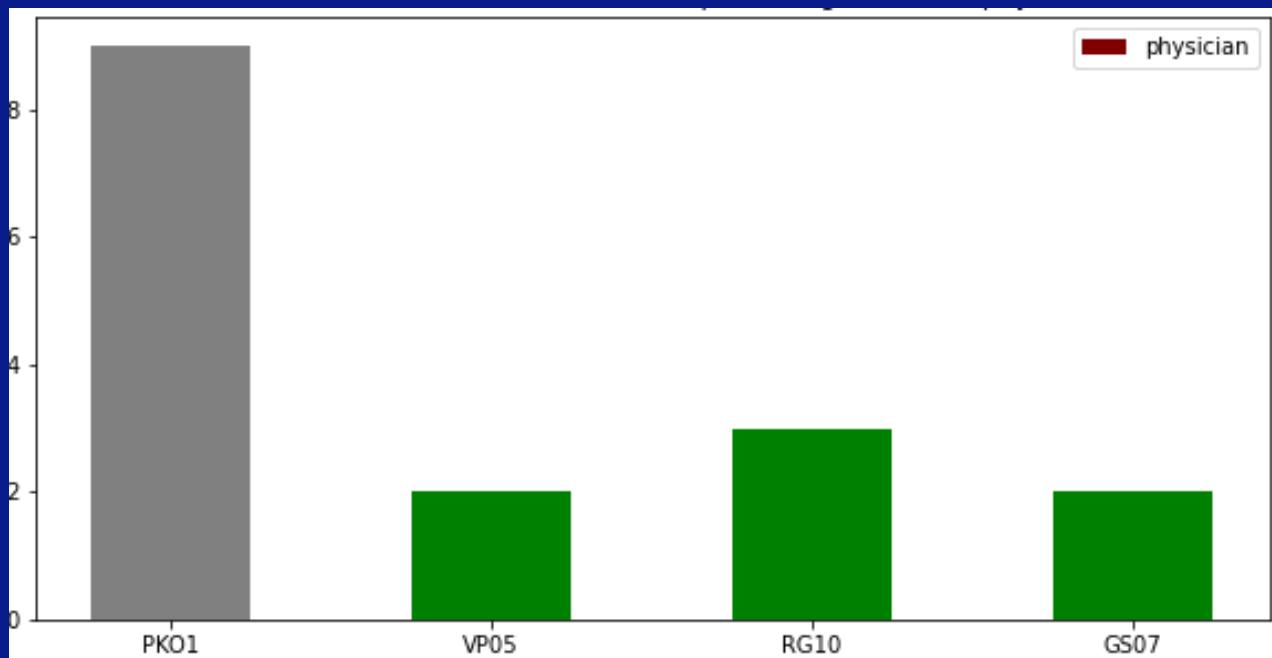
Detected Fraud in Payer



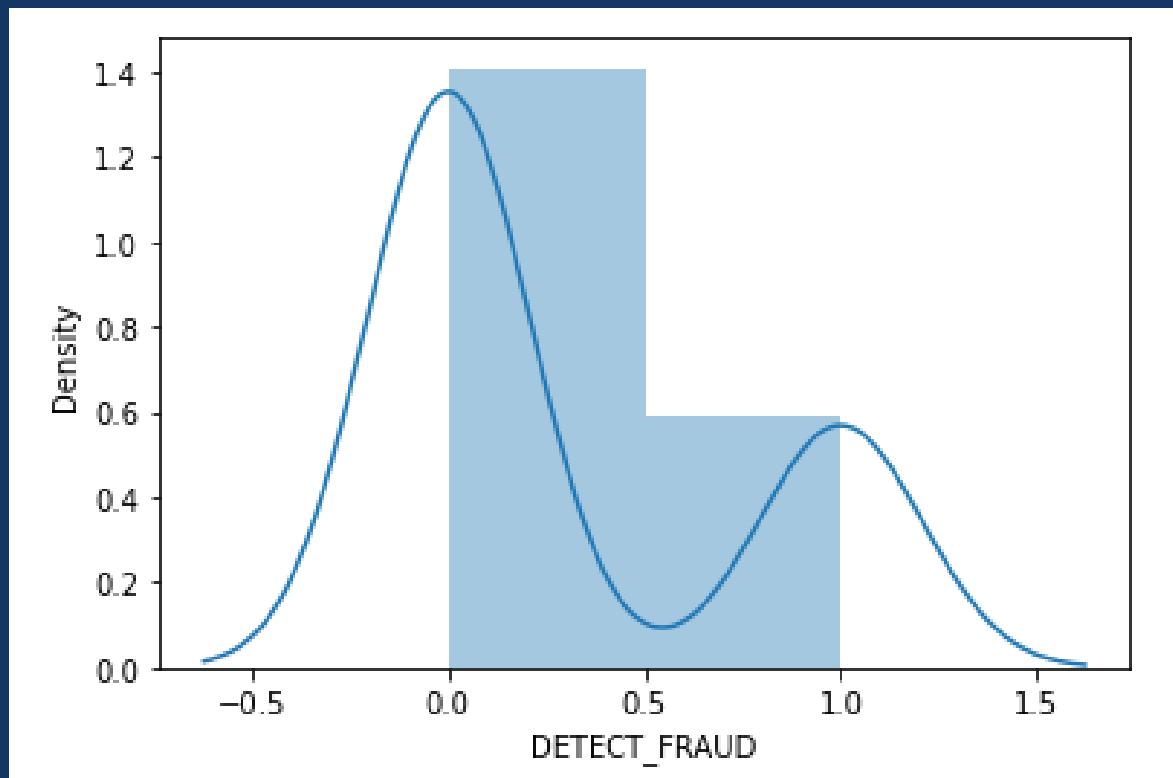
Detected Fraud in Physician Id's



Detected Fraud in Provider Id's



- In statistics, a confidence interval is a type of estimate computed from the statistics of the observed data. This proposes a range of plausible values for an unknown parameter. The interval has an associated confidence level that the true parameter is in the proposed range.
- In this the confidence level got the values lower and upper in lower the value is 0.14573483695919795 and in upper the value is 0.4468577556333946.



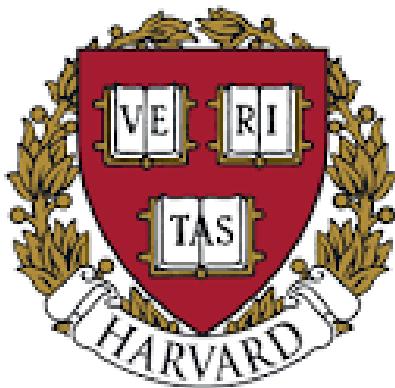
Who We Are

The DeepSphere.AI team comprises MIT learning facilitators, University of California instructors, Harvard PhDs, Stanford alumni, industry leaders, and proven entrepreneurs. The group collectively brings business and technology together for in-depth, hands-on AI learning and a risk-free implementation and AI adoption.

Our Team



**Massachusetts
Institute of
Technology**



**Stanford
University**



"Jothi...I am honored to learn from your comments and messages in the MIT Sloan&CSAIL course. I am deeply and impressed inspired by your ideas which make a great impact on my learning path... "

SinTing (Adele) Lui
MIT CSAIL AI Student



DeepSphere.AI
Enterprise AI and IoT for Analytics

Our Educational Platform

Google Cloud Platform(GCP)

Elementary | Middle | High Schools | Universities

Academic
School
Curriculum

After
School
Curriculum

Summer
School
Curriculum

Applied AI Learning
Curriculum

Digital Educational
Transformation

**DeepSphere.AI Educational Platform
for Next Level Learning**



Next Steps



Contact

DeepSphere.AI, Inc.
2100 Geng Road, Suite 210
Palo Alto, CA 94303
USA
Info@ DeepSphere.AI

DeepSphere.AI is the first and the ONLY
subscription-based comprehensive
applied artificial intelligence platform for
learning artificial intelligence at the
deepest level



DeepSphere.AI
Enterprise AI and IIoT for Analytics