Fraud Modeling

Using Almerys dataset

Interim Results 21/May/2021



The problem we are solving



The problem we are solving



Our technical approach

Pinsight combines unsupervised and supervised learning techniques together with Explainable AI to fully automatically create state-of-the-art AI models on **raw data**.



Overview of work done

• Technical approach:

- Unsupervised learning:
 - We used Local Outlier Factor approach to detect anomalous behavior and rank opticians based on LOF score
 - We used clustering algorithms such as k-means and agglomerative clustering to get segments and compare them based on fraud rate
- Supervised learning:
 - We used Logistic Regression, Support Vector Machines, Decision Trees, Random Forests and different versions of Gradient boosting machines.
 - We used decision trees and decile separation for supervised clustering.

• Model evaluation

• Apart from ROC AUC, Recall and other performance measures, **Lift & Gain** scores are used to evaluate models. The decision is made on the fact that it captures how well records are ranked.

• Results

- Supervised learning dominates in terms of both predictive power and segmentation
- The winner model exhibits competitive performance for predicting fraudulency

Model Performance: overall

- The model exhibits 95% accuracy of prediction in total.
- After the model ranks opticians in decreasing fraud score order, the top decile (aka the top part of the ranking) covers 73% of opticians who are flagged as fraudulent.
- The winner model is LGBM and results are provided on unseen test data.



Insight #0: Key factors influencing fraud score



Relative Importance (%)

Insight #1: Patient age - negative effect

INSIGHT: It is less likely to observe fraud among older patients. Typically, Fraud probability is large if patients are comparable younger.



Insight #2: Price of the frame - positive effect

INSIGHT: The higher the "price of the frame" the higher the probability of fraud.

It is quite interesting that the probability of fraud sharply increases when the price of the frame is higher than ~140.



Insight #3: Gross price of frame - negative effect

INSIGHT: When "Gross price of frame" is lower, it is more likely to observe fraudulency, while higher values indicate less probability of fraud. The likelihood of fraud strictly decreases when the gross price is around 0-3.2.



Insight #4: Gross price of glasses - negative effect

INSIGHT: "Gross price of glasses" exhibits the same behavior as the prices of frames. In other words, it also has negative effect on fraud score.



Insight #5: Number of reimbursement requests by prescriber - positive effect

INSIGHT: The higher "the number of reimbursement requests by prescriber" the higher the probability of fraud. Probability of fraud sharply increases when the number of reimbursement requests increases to ~4.



Further work

• Within segment analysis

- Use unsupervised clustering techniques (such as k-means or hierarchical clustering) to create clusters of opticians
- Capture anomalous opticians within segment
- Reasoning: opticians within segments are assumed to have similar behaviors. Hence, anomalies between them are flagged as potentially fraudulent

• Raw data analysis

- Current analysis is built upon aggregated data
- Aggregated data neglects timestamp information
- When available, raw data will allow Pinsight to calculate trend, seasonality and other time series components that can increase the value of insights.

How it Works?

Step 1 -----Connect data



weicome,			×		
We are so excited to have					
create a project and let Pir		Project Information			
		,			
		Create a project to make predictions and analize the results. Please. fill the required information.			
PROJECTS					<
		* Project Name			
		Fraud analysis			
		Name your project to easily identify it.			
		C * Data			
		historical_data.csv -			
		Select the data on which you want to analyse.			
		* Target IDs			
		Almerys_fraudulent_optician_ID.csv -			
		Select target IDs.			
DATA CENTER					
Name				vtics	
historical_data.csv				ore 🗈 🖬	
Almerys_fraudulent_opticia	← васк		CREATE PROJECT	ore 🚡	
* Here you can find your date					

Step 2 Create Project







Accuracy

Great You have a Great model. Continue analysing it to get more insights.

Name	Created	Average fraud score	Accuracy
Very high fraud score	19.05.2021	0.89	80%
High fraud score	19.05.2021	0.72	85%
Neutral	19.05.2021	0.5	87%
Low fraud score	19.05.2021	0.35	82%
Very low fraud score	19.05.2021	0.12	89%

* We have identified 5 segments for you. Check segments page to find out more.





MODEL HEALTH

OPTICIANS ACCORDING TO THEIR DECREASING PROBABILITY OF FRAUD

Download

ID	Probability of fraud	Rank
dcdd15e23162877	92.7%	1
bbe27a40e899143	91.35%	2
8ac2cb18a35380f	90.02%	3
c252448903abe48	89.3%	4
dc0025ca591774f	88.02%	5

 \mathbf{Q}

□ ~ 0

Features and Benefits



Features and Benefits



How it Works?

Step 1 -----Connect data



weicome,			×		
We are so excited to have					
create a project and let Pir		Project Information			
		,			
		Create a project to make predictions and analize the results. Please. fill the required information.			
PROJECTS					<
		* Project Name			
		Fraud analysis			
		Name your project to easily identify it.			
		C * Data			
		historical_data.csv -			
		Select the data on which you want to analyse.			
		* Target IDs			
		Almerys_fraudulent_optician_ID.csv -			
		Select target IDs.			
DATA CENTER					
Name				vtics	
historical_data.csv				ore 🗈 🖬	
Almerys_fraudulent_opticia	← васк		CREATE PROJECT	ore 🚡	
* Here you can find your date					

Step 2 Create Project







Accuracy

Great You have a Great model. Continue analysing it to get more insights.

Name	Created	Average fraud score	Accuracy
Very high fraud score	19.05.2021	0.89	80%
High fraud score	19.05.2021	0.72	85%
Neutral	19.05.2021	0.5	87%
Low fraud score	19.05.2021	0.35	82%
Very low fraud score	19.05.2021	0.12	89%

* We have identified 5 segments for you. Check segments page to find out more.





MODEL HEALTH

OPTICIANS ACCORDING TO THEIR DECREASING PROBABILITY OF FRAUD

Download

ID	Probability of fraud	Rank
dcdd15e23162877	92.7%	1
bbe27a40e899143	91.35%	2
8ac2cb18a35380f	90.02%	3
c252448903abe48	89.3%	4
dc0025ca591774f	88.02%	5

 \mathbf{Q}

□ ~ 0

Features and Benefits



Features and Benefits



Enabling Enterprise AI Transformation

Pinsight

https://metric.am/ https://fb.com/metric.am https://github.com/Metricam info@metric.am #actmeasurable +37499 02-06-62

Unsupervised learning

Approach

- 1. Train Local Outlier Factor model
- 2. Assign LOF score to each optician
- 3. Rank opticians based on LOF (from highest to lowest)
- 4. Based on ranking create deciles (e.g. top 10%, top 20% etc.)

Results

- 1. LOF was able to come up with fair ranking.
- 2. The Gain at the top decile (aka % of fraudulent records in the first decile among total) is 16%.

Approach

- 1. Build clustering model using hierarchical clustering
- 2. Calculate fraud rate among total in each cluster
- 3. Calculate within cluster fraud among total fraud

Results

- Clustering was able to identify specific optician segments (around 500 opticians) who do not commit fraud.
- 2. Clustering failed to find a fraudulent segment.