Institute and Faculty of Actuaries

# IFoA India Conference 2024

29 November – 1 December, Andaz Hotel, New Delhi

Institute and Faculty of Actuaries

# Understanding AI

## eXplainable AI (XAI) techniques in practice

**Presenters:**
- Karol Gawlowski
- Satya Sai Mudigonda, SSSIHL
- Dr. Rohan Yashraj Gupta, SSSIHL & Accenture

**IFoA India Conference 2024**

# Agenda Items

- Current landscape
- XAI
- XAI – ICE/PDP
- XAI - SHAP
- AI/ML in Actuarial Work
- AI Case Studies
- Where can it go wrong?
- XAI Case Study

# Current AI landscape

- **35% of global companies report using AI** in their business
- **42% are exploring AI** adoption, and 50% plan to incorporate it as of recent reports
- The global AI in the insurance industry generated **$2.74 billion in 2021**, and is anticipated to generate **$45.74 billion by 2031**[1].
- **More than half** the insurance companies used ML-driven predictive analytics for insurance underwriting[2].
- Most respondents **(75%) believe AI and ML** can provide companies with a competitive advantage[3].
- In 2022, **North America held a dominant position** in the global AI market, accounting for 43%[4].

Source :
1 - mckinsey.com,
2 - risk.lesixnexis.com,
3 - analyticsvidhya.com,
4 - simplesolve.com

# Introduction – motivation

## Performance – interpretability trade-off

| | | Interpretability | |
|---|---|---|---|
| | | yes | no |
| Predictive power | + | | Machine Learning |
| | - | Statistical Modelling | |

XAI

- Interpretability – the degree to which a human can consistently predict the model's result
- Predictive power – degree to which a mathematical model is useful in determining results of some process

**Why XAI?**

- No more trade-off between models' power and explainability
- XAI is essential in debugging black box models
- Black box models can serve as benchmarks on parameter level, not just performance
- Modelers can assess fairness and bias
- Most popular XAI techniques are model agnostic

1. B. Kim et. al. "Examples are not enough, learn to criticize! Criticism for interpretability." Advances in Neural Information Processing Systems (2016)

# Introduction – a black box model

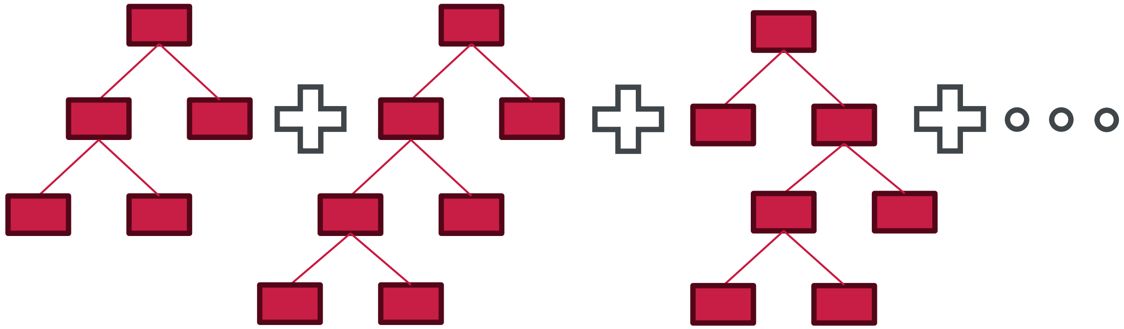
Boosted trees with `library(xgboost)` and `CASdatasets::data(freMTPL2freq)`

**The data**

- 677,991 Motor Third Party Liability policies
- Exposure and claim data
- 7 features:
  - Power (categorical)
  - VehAge (cont.)
  - DrivAge (cont.)
  - Brand (cont.)
  - Gas (Diesel/Regular)
  - Region (categorical)
  - Density (cont.)

**The model**

- Predict claim frequency (poisson)
- Gradient Boosting Machine
- Parameters – Grid search



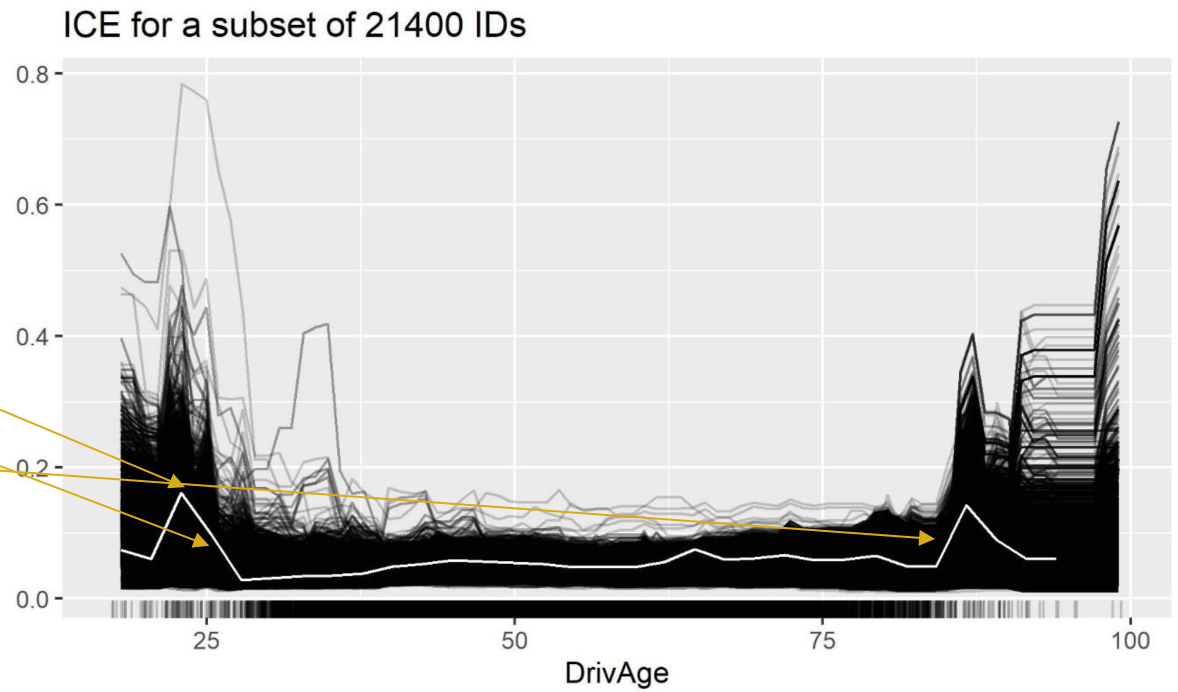
github.com/Karol-Gawlowski/GIRO_2022    The Actuary

# 1. ICE – Individual Conditional Expectation

| ID | Area | … | DrivAge | Pred |
|---|---|---|---|---|
| 123 | D | | 9 | 0.475 |

| ID' | Area | … | DrivAge | Pred |
|---|---|---|---|---|
| 123.1 | D | | 20 | 0.060 |
| 123.2 | D | | 23 | 0.161 |
| 123.3 | D | | 25 | 0.098 |
| … | … | … | … | … |
| 123.32 | D | | 84 | 0.049 |

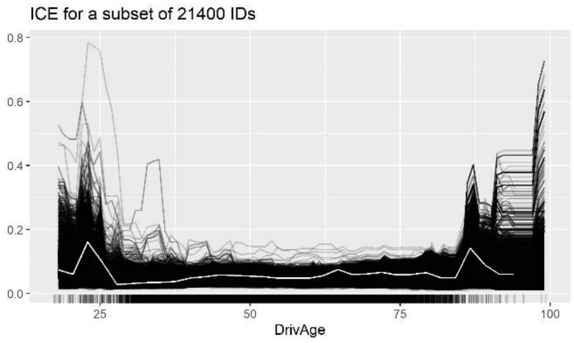
Repeat for a sample of policies

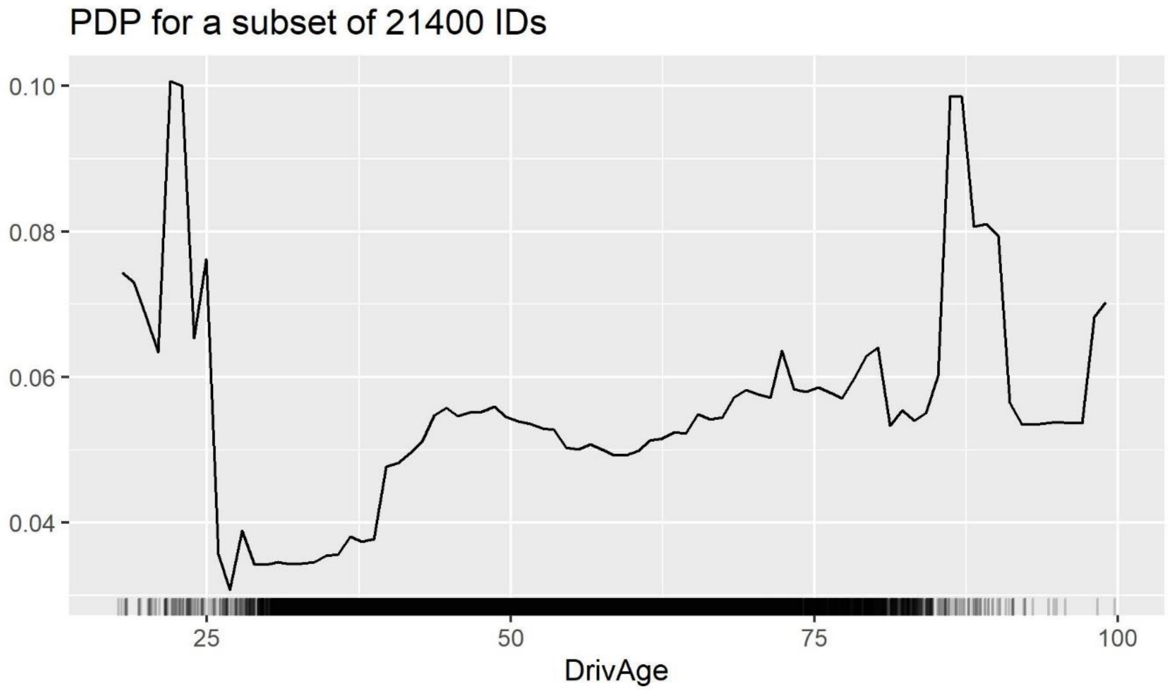



- Observation/variable level explanation
- Alter DrivAge keeping other variables constant and obtain the prediction for the artificial points

The IDs chosen are for drivers with lowest Bonus Malus score owning 5-year-old cars

# 2. PDP – Partial Dependence Plot




- Aggregated ICE output – model level explanation
- Shows a global influence/impact
- Other variants that handle correlation/dependence:
  - ALE plots / M-plots

The IDs chosen are for drivers with lowest Bonus Malus score and 5-year-old cars

# 3. SHAP – Shapley values

**The set up:**

1) A cooperative game involving two or more participants
2) Specified payout function – mapping (sub)sets of participants (called coalitions), to an expected numerical game outcome

**Shapley values:**

- Assign contributions to players in each coalition
- Distribute the obtained coalition payout as individual contributions among the players, so it sums up to the total amount

**Note:**

- Shapley values are the only way to distribute gains fairly among coalition members
- The formula below shows a contribution for player $i$ out of $N$, in a game with payout $v$ for possible coalitions $S$, including an empty coalition with payout $v(\emptyset) = 0$

$$\varphi_i(v) = \sum_{S \subseteq N \setminus \{i\}} \underbrace{\frac{|S|!\,(n - |S| - 1)!}{n!}}_{weight} (v(S \cup \{i\}) - v(S))$$

# 3.1 SHAP – Adapting Shapley values to XAI

## The set up

1) We treat input variables as participants in a game
2) The payout function is simply the black box model to be explained
3) As an explanation of models' prediction for a given observation, we obtain an additive decomposition of its output, distributed among the inputs

## The problem

If our model is built on a set $S$ of input variables, how to obtain an output for $v(S\backslash \mathrm{k})$ or even $v(\emptyset)$?

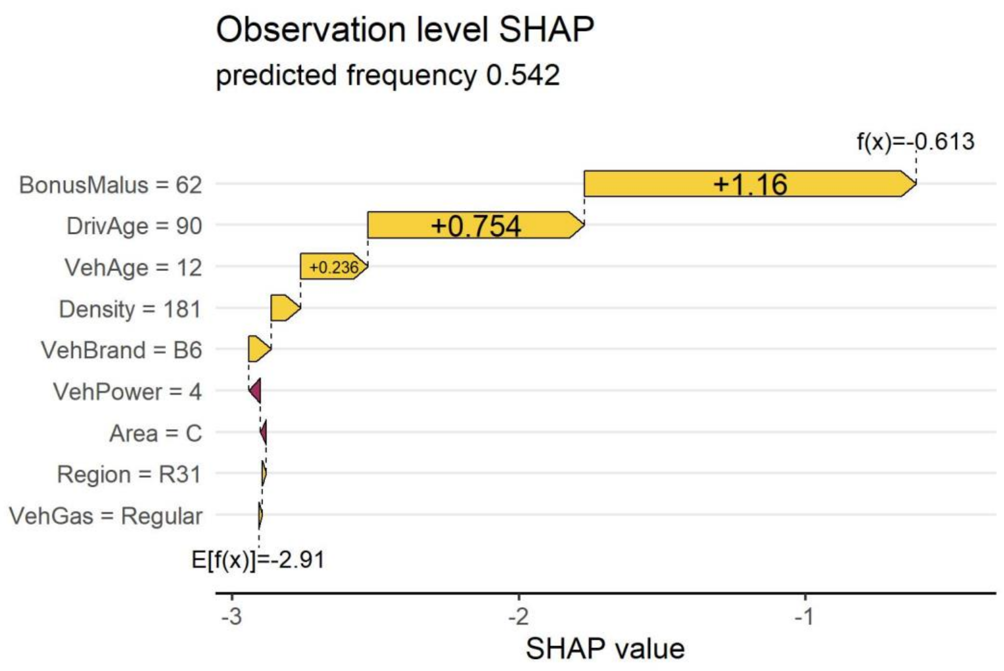
⇩

Remove the effect of feature $k$ by averaging through its range of possible values.

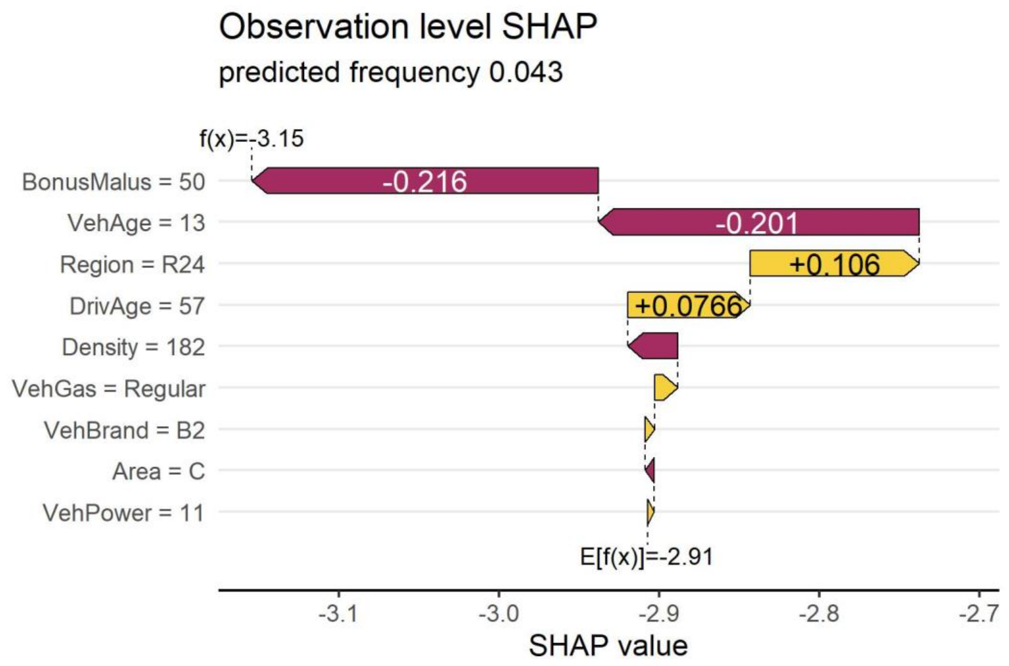
Scott M. Lundberg, Su-In Lee – A Unified Approach to Interpreting Model Predictions

# 3.2 SHAP – observation level explanation

**Example A**



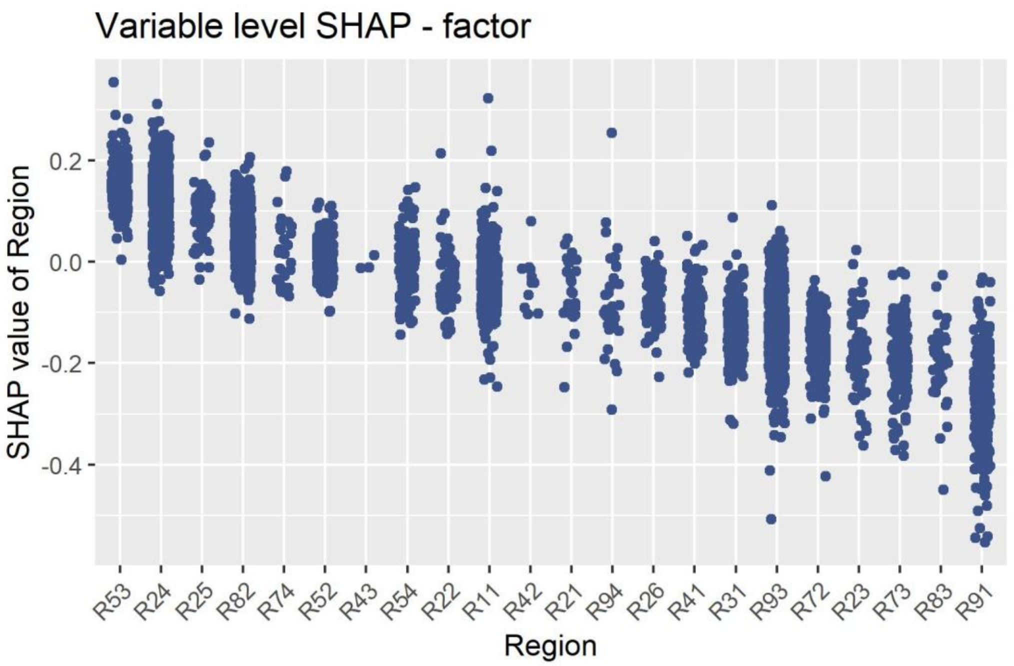

**Example B**



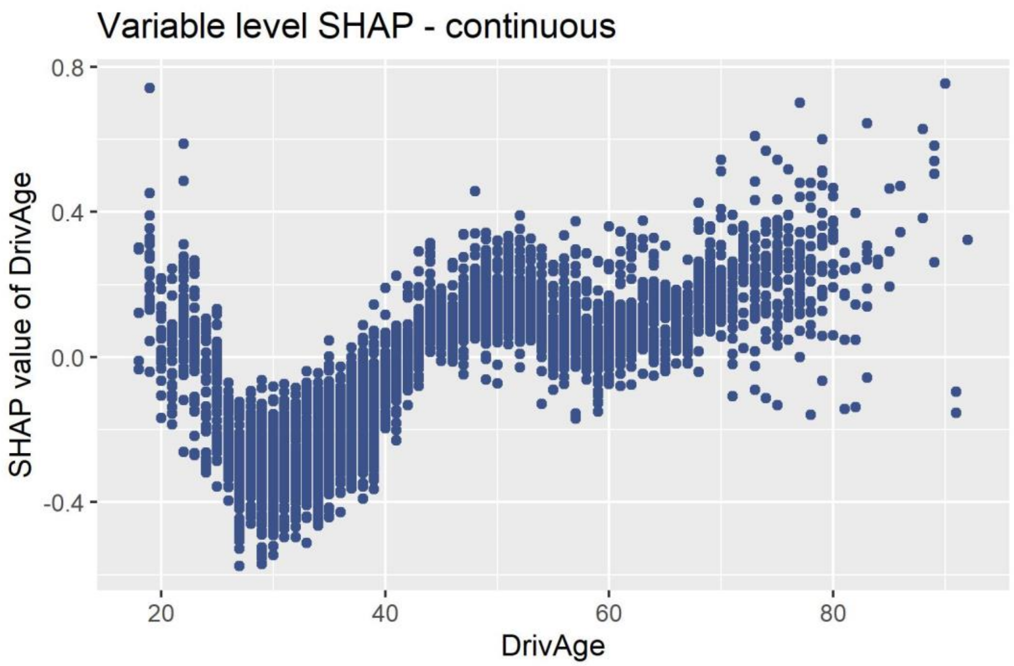

Note: The SHAP values shown are relativities, so the predicted frequency is $\exp(f(x))$

# 3.3 SHAP – variable level explanation

**Example A**




**Example B**



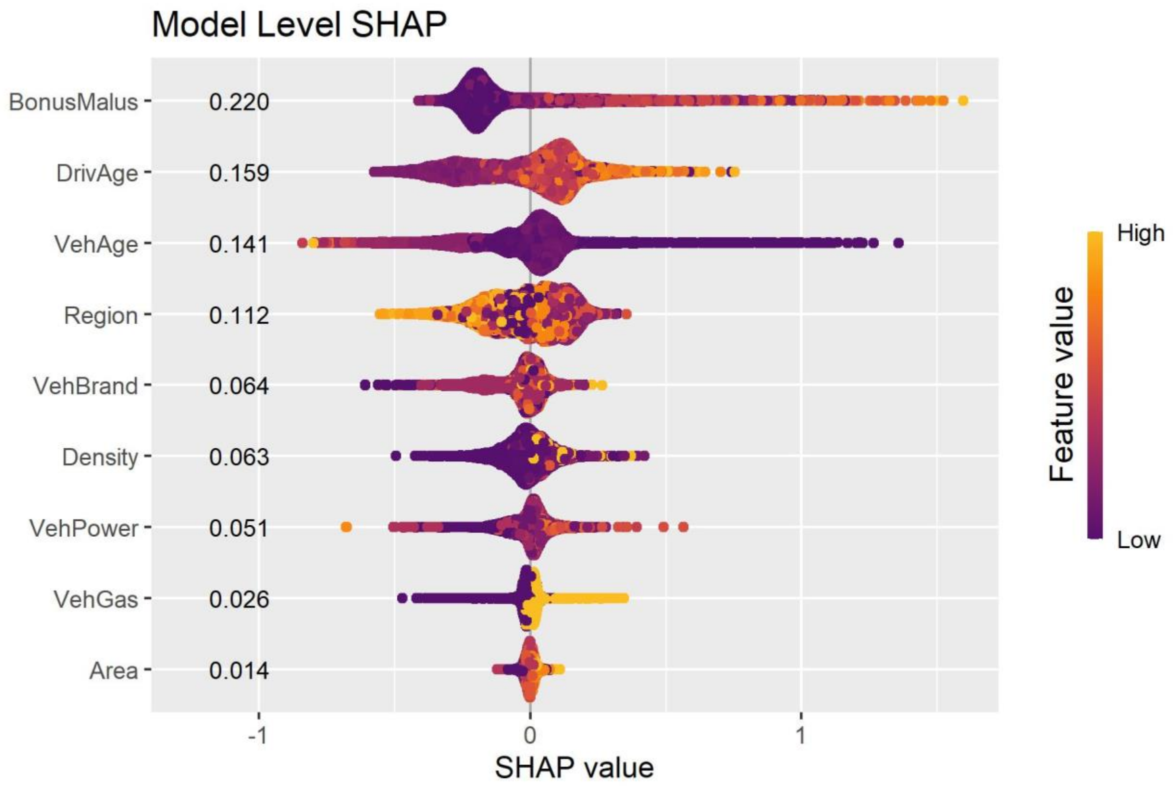

Note: The SHAP values shown are relativities, so the predicted frequency is $\exp(f(x))$

# 3.4 SHAP – model level explanation

- High/low feature value can both increase/decrease models' output
- SHAP is model agnostic (e.g. kernel-SHAP)
- There are model specific SHAP algorithms that significantly speed up the calculations (e.g. Tree-SHAP)
- Some SHAP algorithms account for correlated variables (e.g. XGBoost or shapr)

**More on kernel-SHAP:**

*The Actuary – All clear: How Shapley values make opaque models more transparent*




Note: The SHAP values shown are relativities, so the predicted frequency is $\exp(f(x))$

# 3.5 SHAP in R

Visualization with `library(shapviz)`

- A predictor object holding the model, data and a predictor function if needed
- shapviz produces ggplot objects
- Supports outputs of:
  - XGBoost
  - LightGBM
  - h2o
  - shapr
  - fastshap

```r
viz = shapviz(object = xgb,
              X = data,
              X_pred = data %>% encode())

sv_waterfall(viz, row_id = 1) +
ggtitle("Observation Level Explanation")

sv_dependence(viz,v = "VehAge") +
ggtitle("Variable Level Explanation")

sv_importance(viz, kind = "beeswarm") +
ggtitle("Model Level Explanation")
```

# AI/ML in Actuarial Work

| | | | |
|---|---|---|---|
| Automated Pricing Models | Dynamic Pricing | Anomaly Detection Models | Natural Language Processing (NLP) in Textual Data Analysis |
| Predictive Reserving Models | Inflation Adjustment with ML | Claims Settlement | Chatbots & Virtual Assistants, etc… |

# Case Study

Pricing

# Ayushman Bharat : In Numbers

**66.7 Cr +***
Health Accounts Created

**6.86 Cr +**
Hospital Admissions

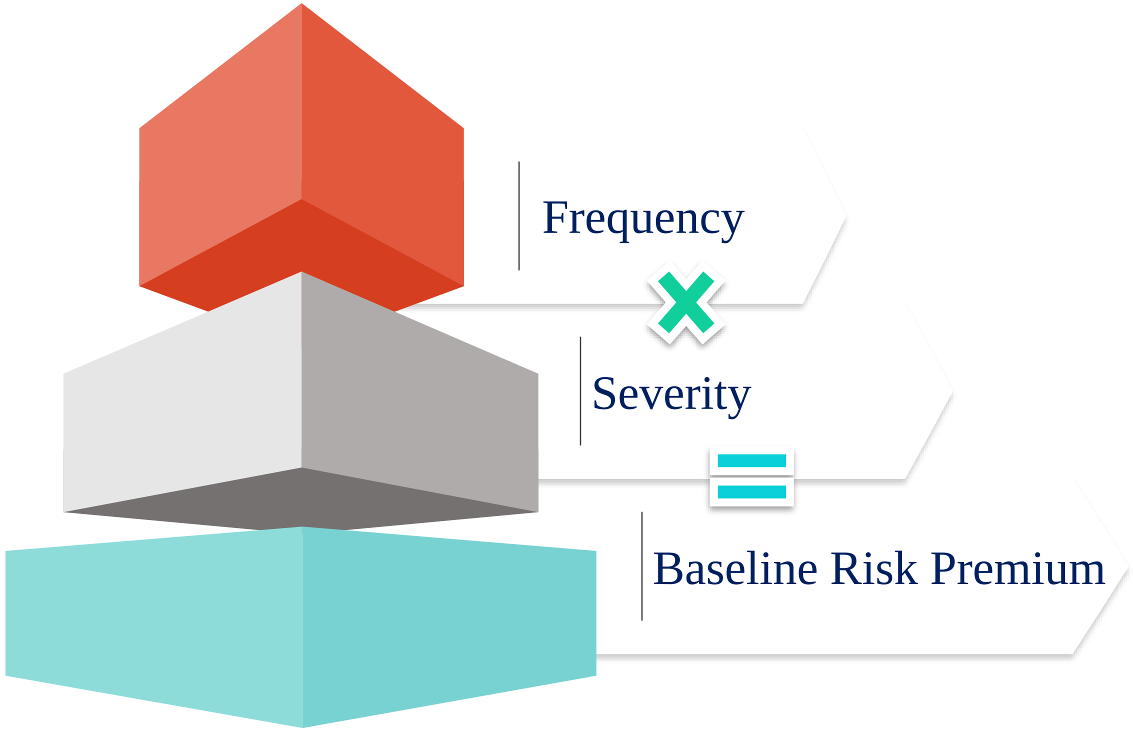
**90,204 Cr +**
Amount Paid in Claims

Hospitals empaneled
**30,671 +**

Packages
**1393 +**

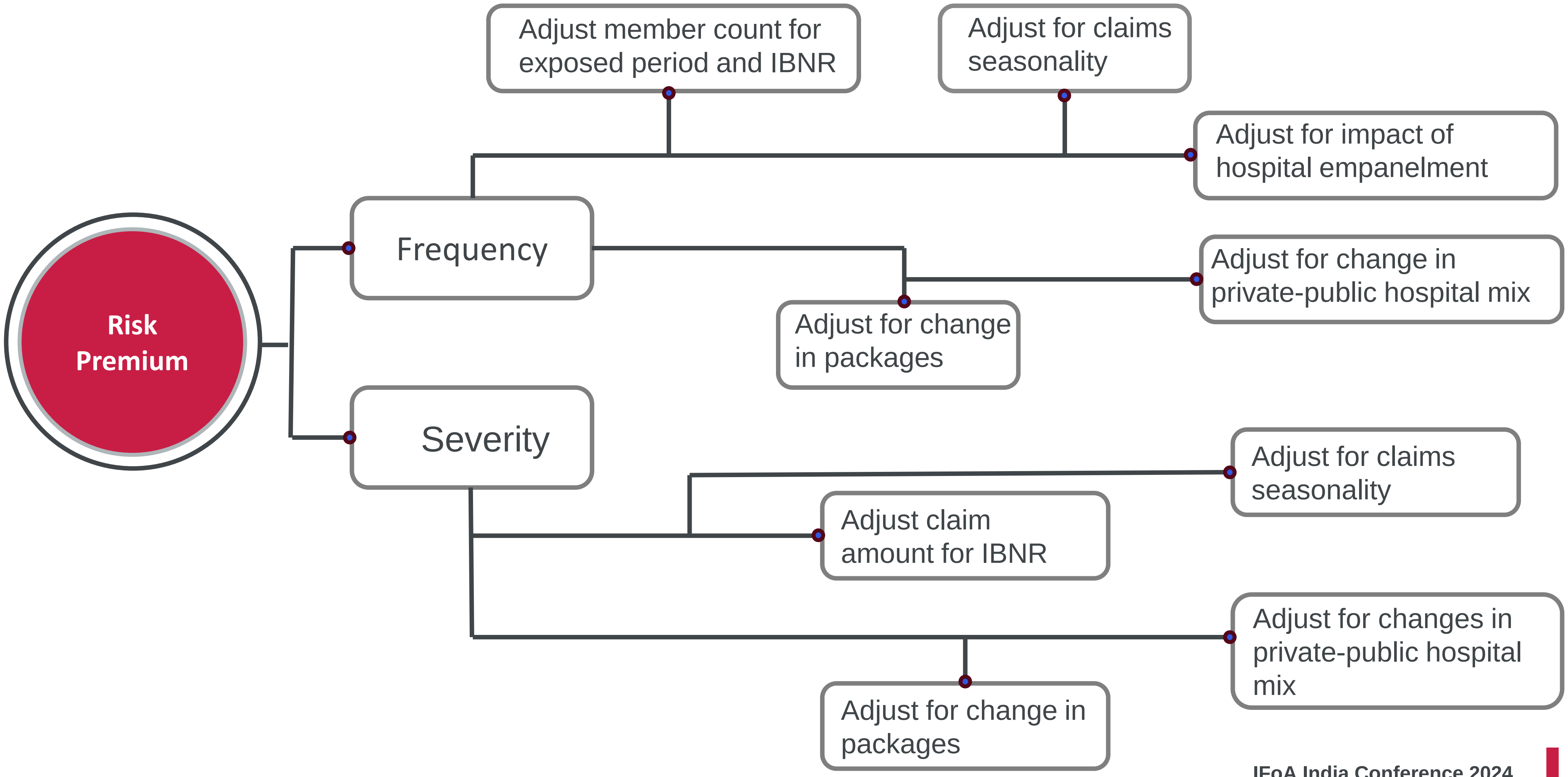
Press Note Details: Press Information Bureau
NHA | Setu Dashbaord
* 6 Cr. of 70+ senior citizens

# Arriving at adjusted risk Premium




- Data
- Assumptions
- Model
- Computation of Risk Premium
- Other Considerations

# Pricing




**Tools and Techniques**

**Actuarial/ Statistical**

- Burning cost approach
- GLM
- ILF benchmarking
- Chainladder
- Bornhuetter Ferguson
- Cape cod
- Time series
- Credibility theory
- Scenario testing
- Ruin Theory

**Data science**

- Boosting
- GAM
- Neural networks
- Support Vector Machines (SVM)
- Classification and Regression Trees (CART)
- Clustering

# Sample Risk Premium Calculation

***Select input for projecting risk premium***

Frequency

Lower | Average | Upper

Severity

Lower | Average | Upper

Frequency Trend

Lower | Average | Upper

Severity trend

Lower | Average | Upper

Hospital Empanelment Impact

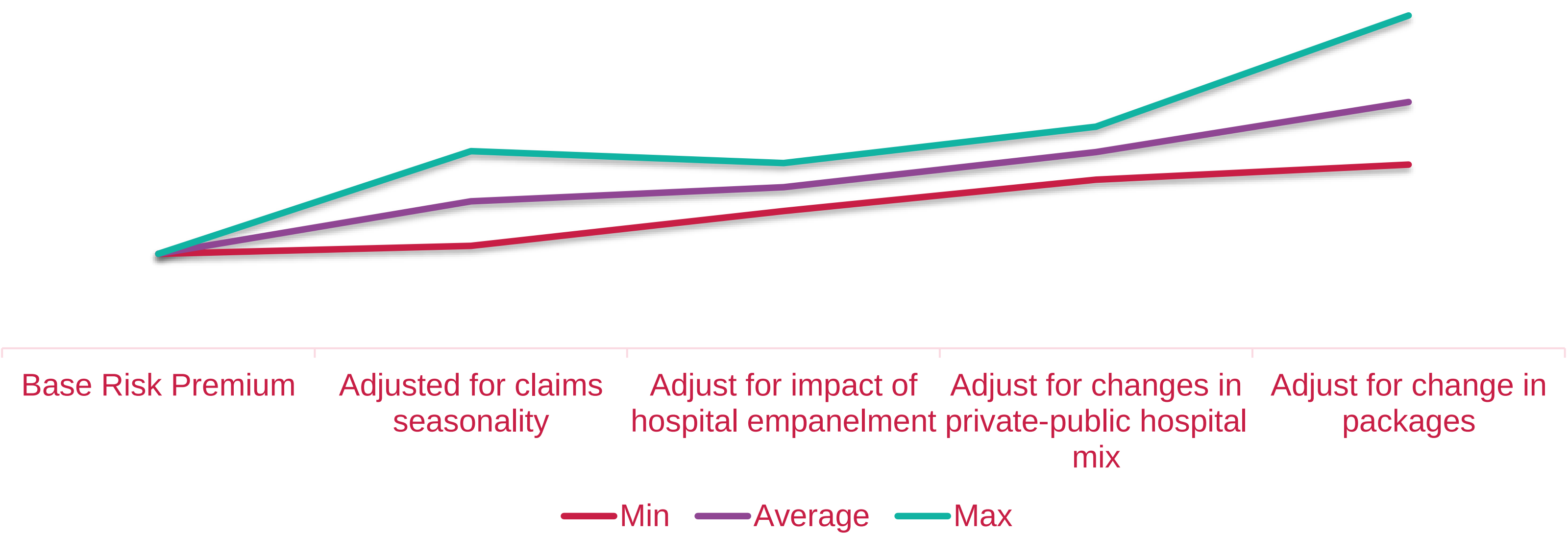
Lower | Average | Upper

Package Rate Changes

Lower | Average | Upper

Package Count Changes

Lower | Average | Upper




| | Min | Average | Max |
|---|---|---|---|
| Base Risk Premium | 1000 | 1000 | 1000 |
| Adjusted for claims seasonality | 1084 | 1554 | 2086 |
| Adjust for impact of hospital empanelment | 1452 | 1702 | 1960 |
| Adjust for changes in private-public hospital mix | 1782 | 2076 | 2344 |
| Adjust for change in packages | 1942 | 2606 | 3522 |

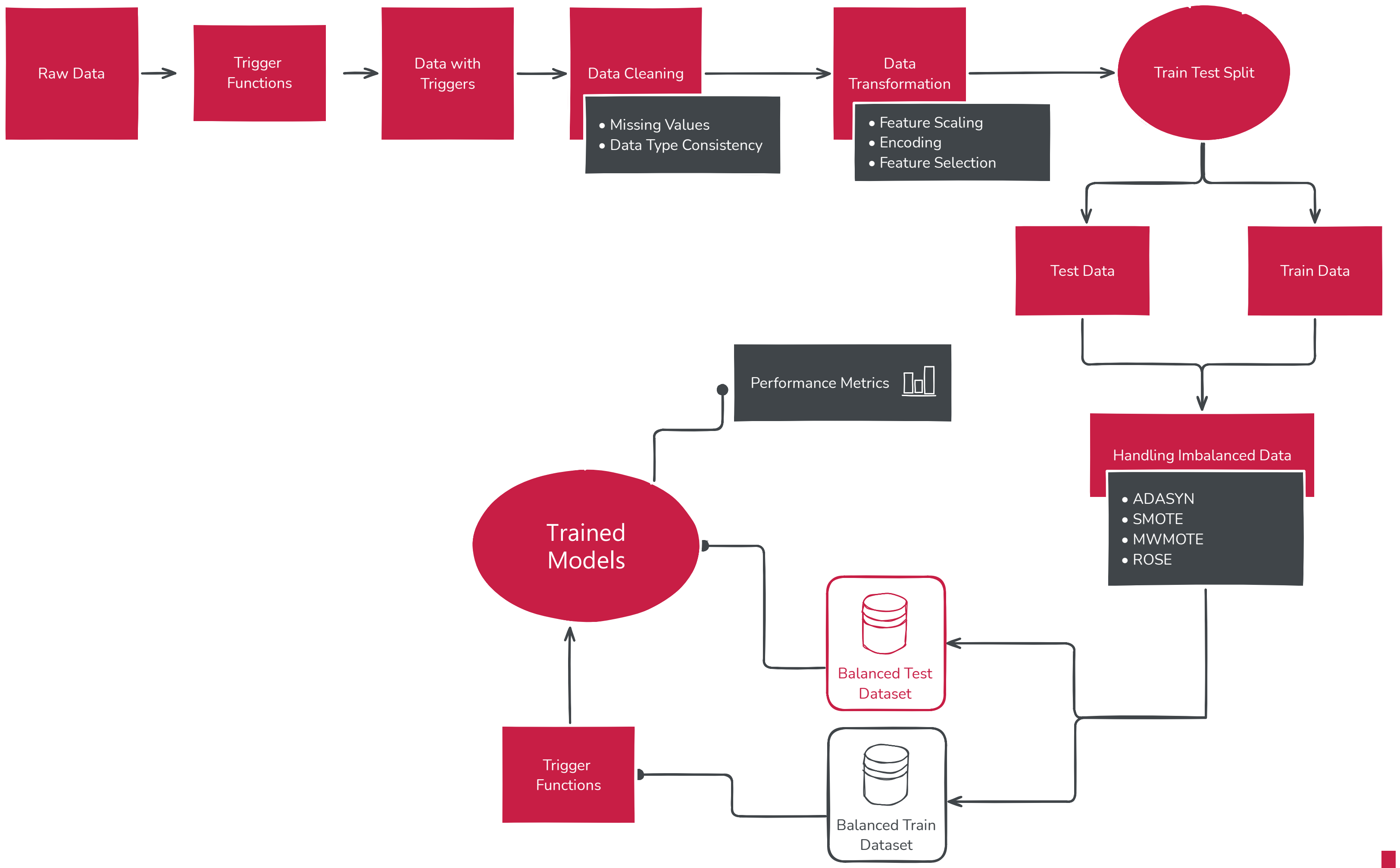
# Benefits of using ML in pricing

- Better and in-depth insights lead to better-informed pricing decisions.
- Improved risk assessment leads to more accurate underwriting decisions, reducing the likelihood of under-pricing or over-pricing policies.
- Those who leverage ML can more effectively compete with rivals in terms of both pricing and product innovation.
- ML enables insurers to analyze historical data and predict future pricing trends, allowing for proactive adjustments to pricing strategies.

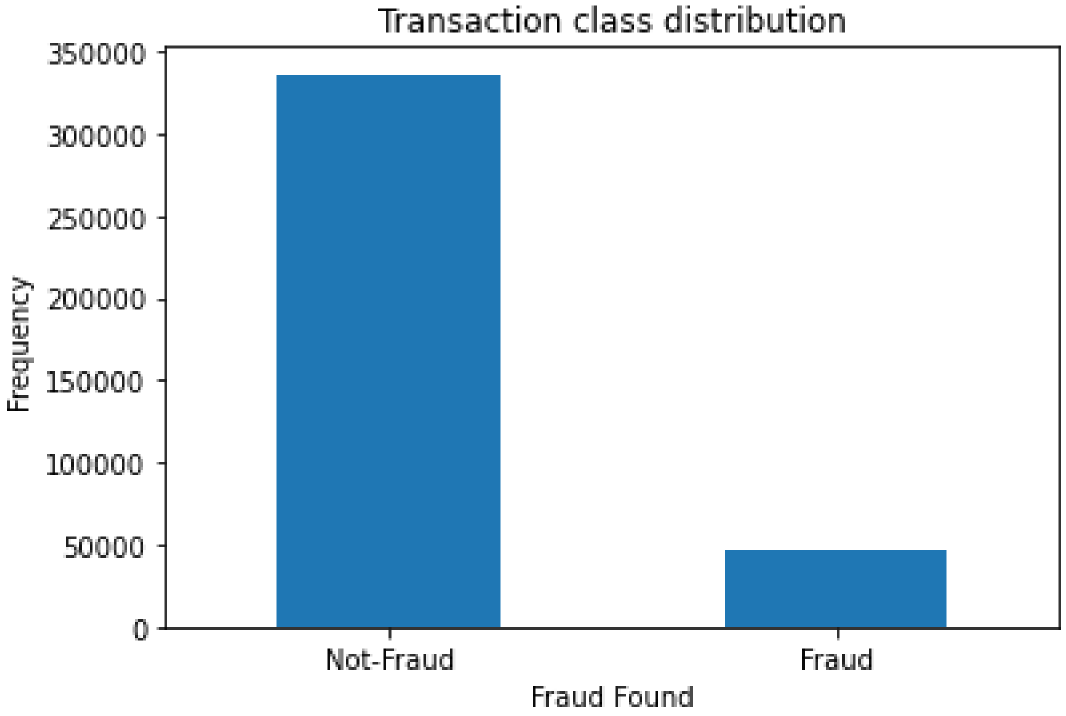
# Case Study

Fraud Detection

# Framework


Raw Data
Trigger Functions
Data with Triggers
Data Cleaning
• Missing Values
• Data Type Consistency
Data Transformation
• Feature Scaling
• Encoding
• Feature Selection
Train Test Split
Test Data
Train Data
Handling Imbalanced Data
• ADASYN
• SMOTE
• MWMOTE
• ROSE
Performance Metrics
Trained Models
Balanced Test Dataset
Trigger Functions
Balanced Train Dataset

# Class distribution – Data Imbalance




**Example**

Fraud = 46,400 claims out of 3,82,587

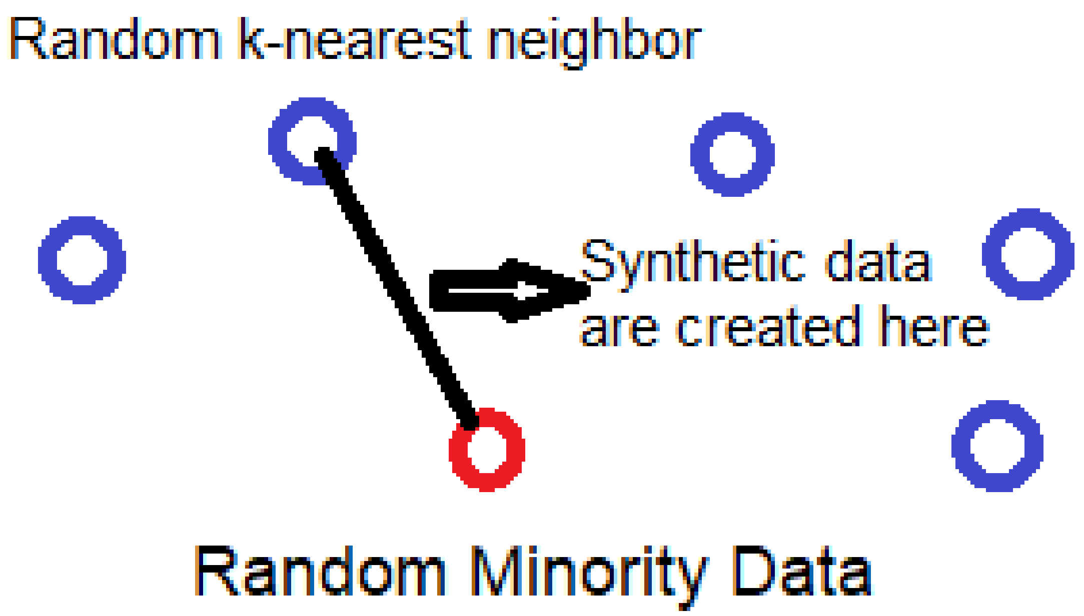
Train test split = 75:25

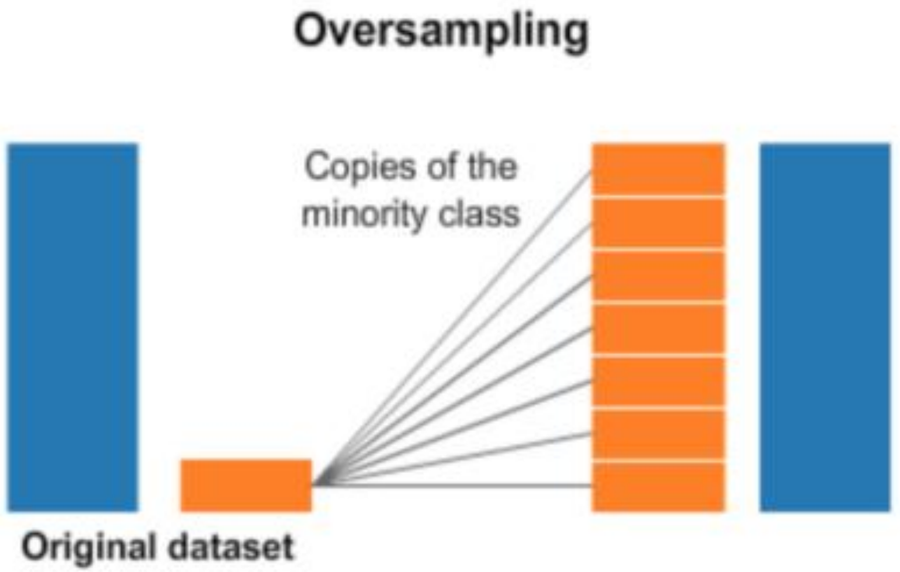
Train = 286940
Test = 95647

Train and the test dataset was divided such that the percentage of fraud and non-fraud cases were identical

https://cag.gov.in/en/audit-report/details/119060
This is the main source of information to identify different types of fraudulent claims, as described in Chapter 5, "Claims Management,"

# SMOTE




Source: Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: synthetic minority over-sampling technique. *Journal of artificial intelligence research*, *16*, 321-357.

# Benefits of using ML in Fraud Detection

- ML algorithms can identify patterns indicative of fraudulent activities, helping insurers prevent and detect fraudulent claims.
- ML models continuously learn from new data, adapting to emerging fraud tactics and improving their detection capabilities over time.
- ML algorithms can distinguish between normal and suspicious behavior with greater precision, minimizing false positives.
- Improved fraud detection leads to operational cost savings and allows organizations to allocate resources more efficiently.

# Where can it go wrong?

**Model selection**

- XGBoost
- Random Forest
- GBM
- GLM
- Neural Network
- Deep Learning

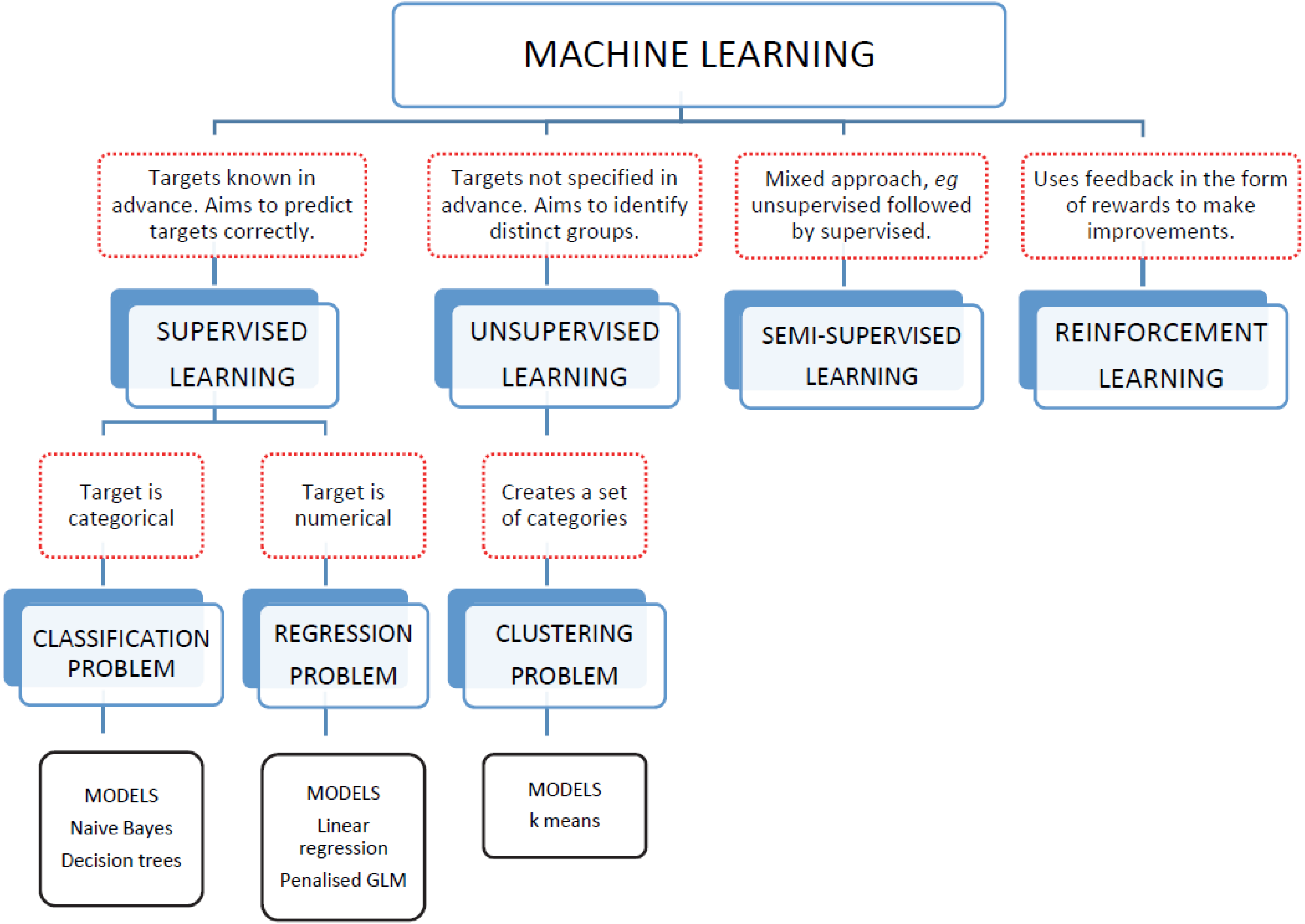
**Data imbalance**

- ADASYN
- SMOTE
- MWMOTE
- Generative AI models

**Interpretability**

- Partial Dependency Plots
- SHapley Additive exPlanations
- Permutation Feature Importance
- Local Interpretable Model-agnostic Explanations

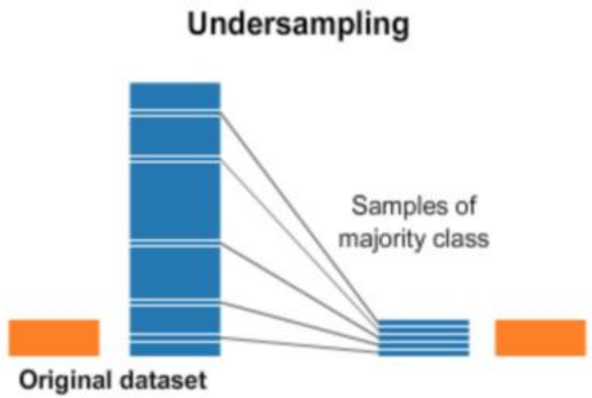
# Challenge 1 – Model Selection

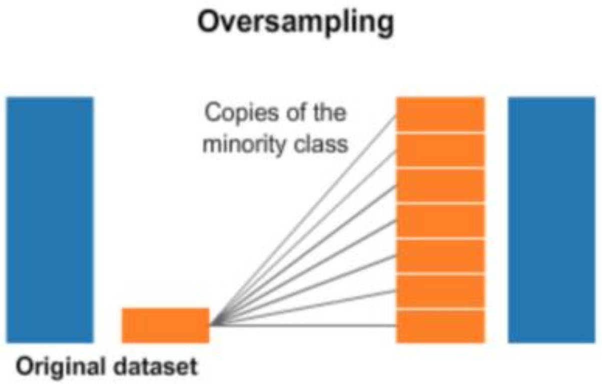



Source: IFoA Material

# Challenge 2 – Data Imbalance

- Skewed nature of the dataset results in the bias during training
- High accuracy could be deceptive





- Ali, A., Shamsuddin, S. M., & Ralescu, A. L. (2013). Classification with class imbalance problem. Int. J. Advance Soft Compu. Appl, 5(3).
- Mellor, A., Boukir, S., Haywood, A., & Jones, S. (2015). Exploring issues of training data imbalance and mislabelling on random forest performance for large area land cover classification using the ensemble margin. ISPRS Journal of Photogrammetry and Remote Sensing, 105, 155-168.

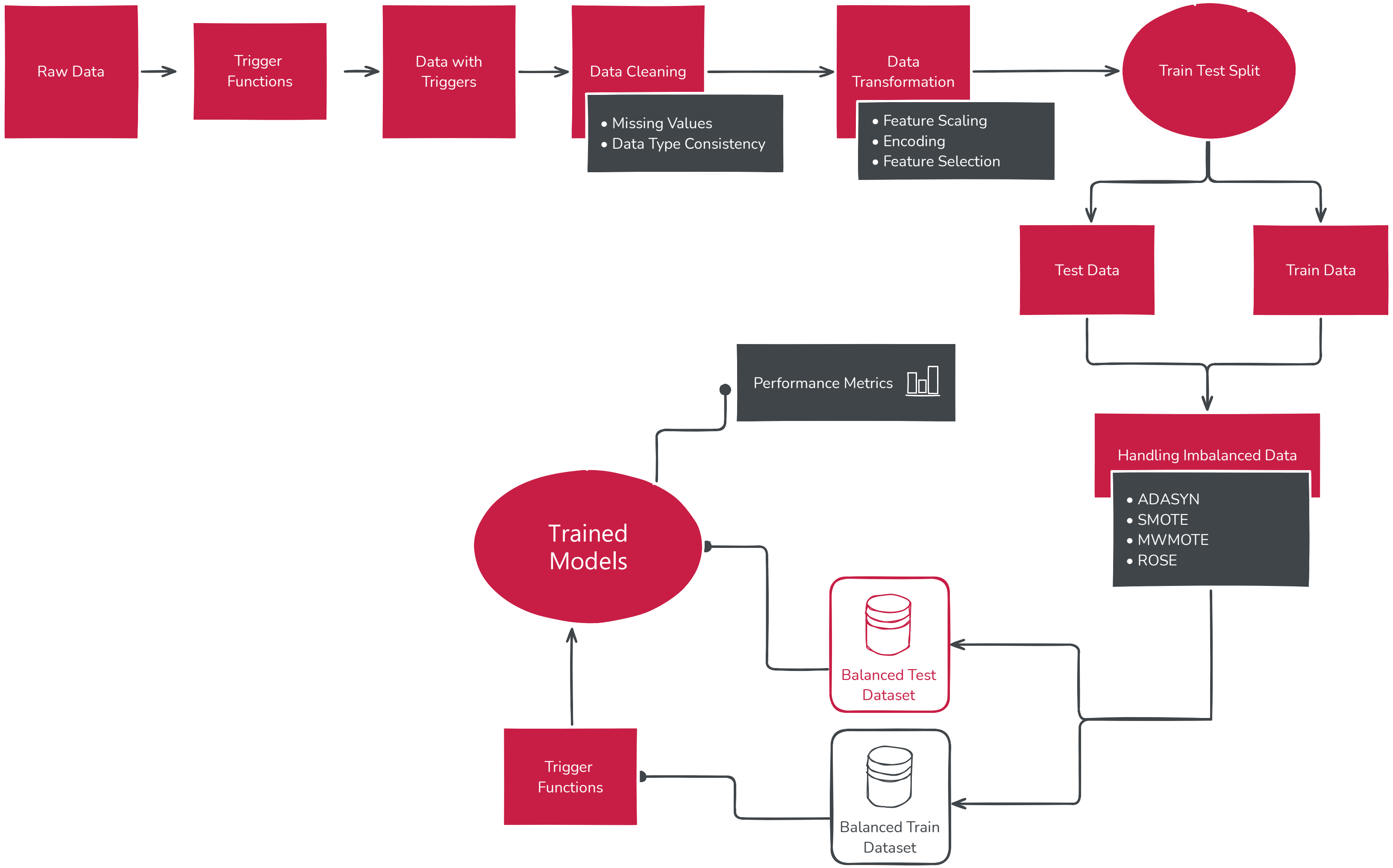
# Case Study

Fraud Detection

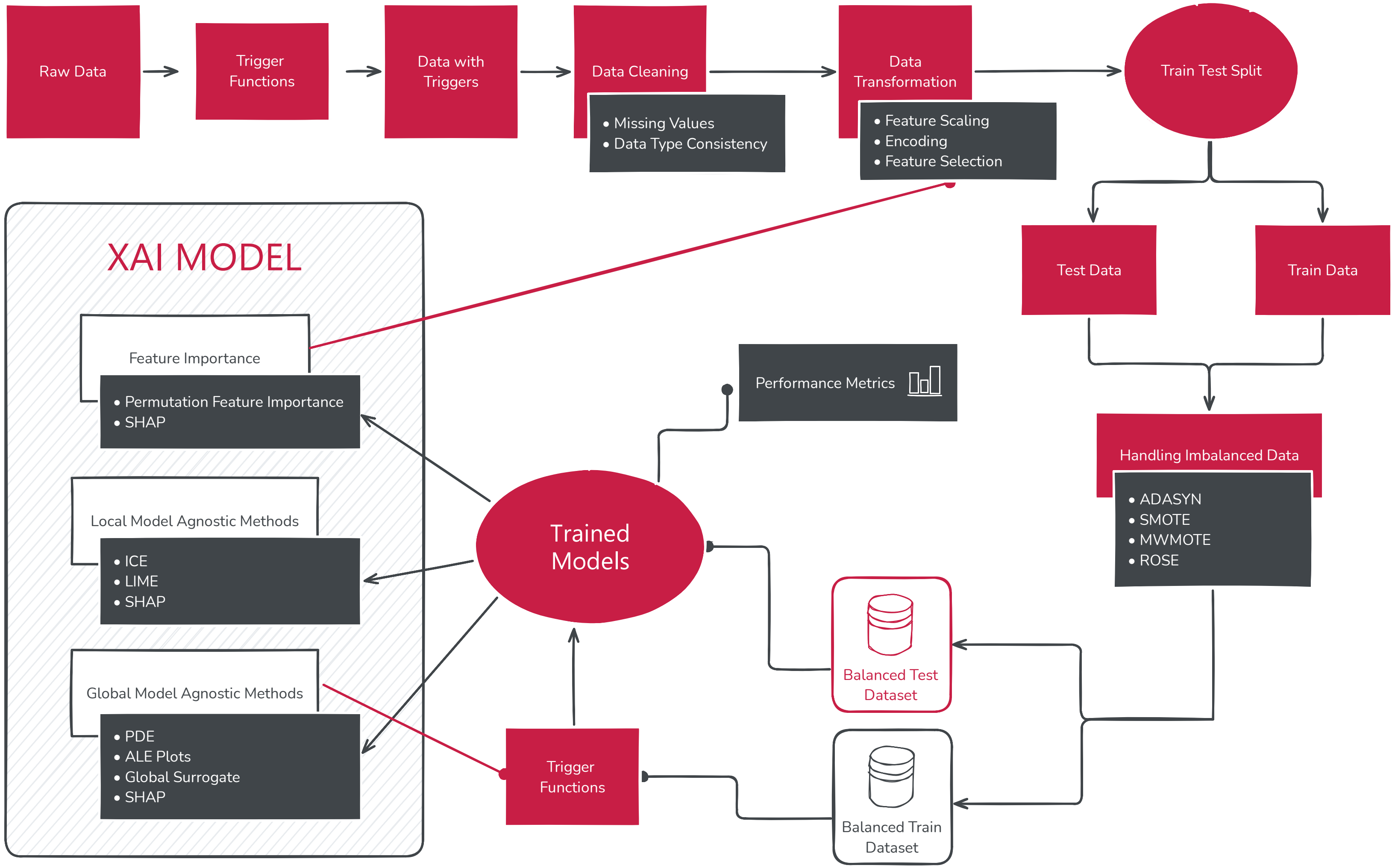
# XAI Techniques in Fraud Detection for Health Insurance

- Objective: Identify fraudulent claims in health insurance using AI models.
- Context: Health insurers face substantial fraud risk, with false or exaggerated claims costing millions annually.
- Challenge: Traditional models detect fraud but lack transparency, making it hard for actuaries to explain decisions.

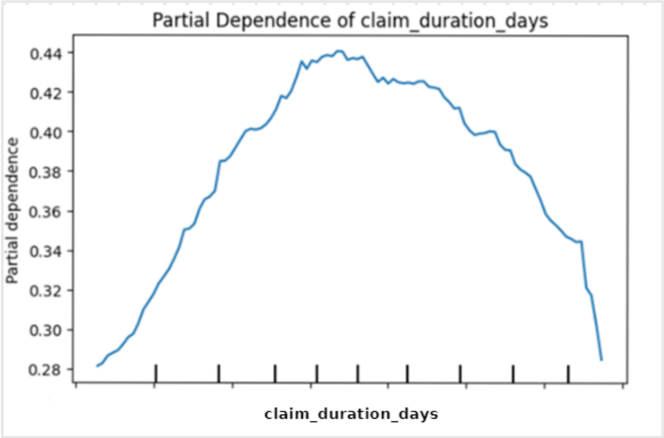
# Framework

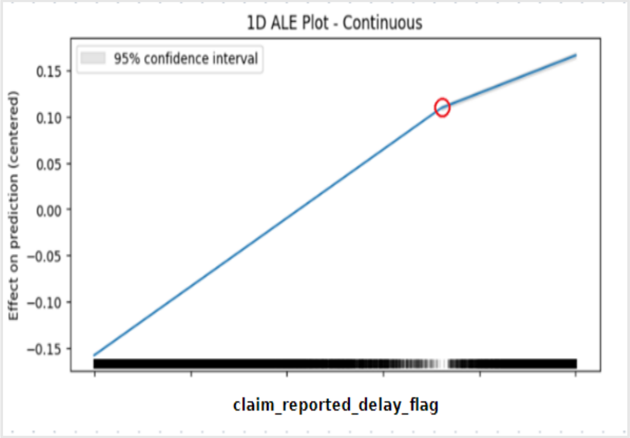

Raw Data
Trigger Functions
Data with Triggers
Data Cleaning
• Missing Values
• Data Type Consistency
Data Transformation
• Feature Scaling
• Encoding
• Feature Selection
Train Test Split
Test Data
Train Data
Handling Imbalanced Data
• ADASYN
• SMOTE
• MWMOTE
• ROSE
Performance Metrics
Trained Models
Balanced Test Dataset
Balanced Train Dataset
Trigger Functions

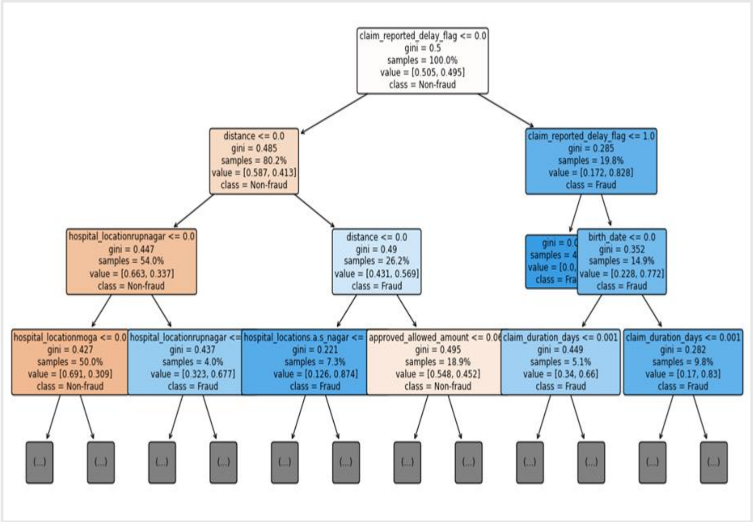
# Framework

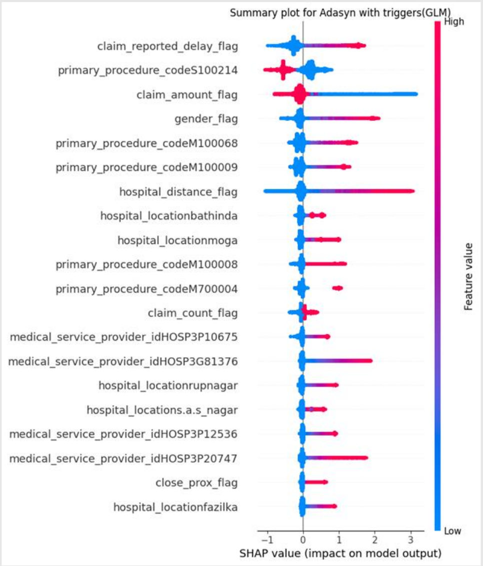

Raw Data
Trigger Functions
Data with Triggers
Data Cleaning
• Missing Values
• Data Type Consistency
Data Transformation
• Feature Scaling
• Encoding
• Feature Selection
Train Test Split
Test Data
Train Data
Handling Imbalanced Data
• ADASYN
• SMOTE
• MWMOTE
• ROSE
XAI MODEL
Feature Importance
• Permutation Feature Importance
• SHAP
Local Model Agnostic Methods
• ICE
• LIME
• SHAP
Global Model Agnostic Methods
• PDE
• ALE Plots
• Global Surrogate
• SHAP
Performance Metrics
Trained Models
Balanced Test Dataset
Balanced Train Dataset
Trigger Functions

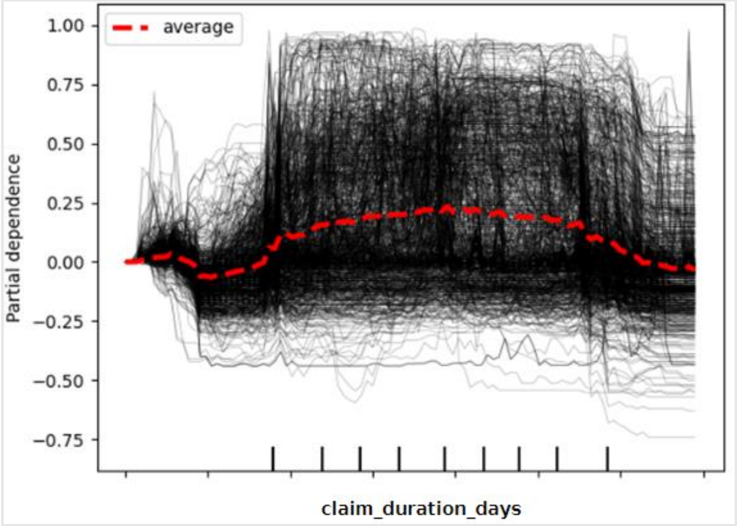
# Local & Global Agnostic Models- Snapshot

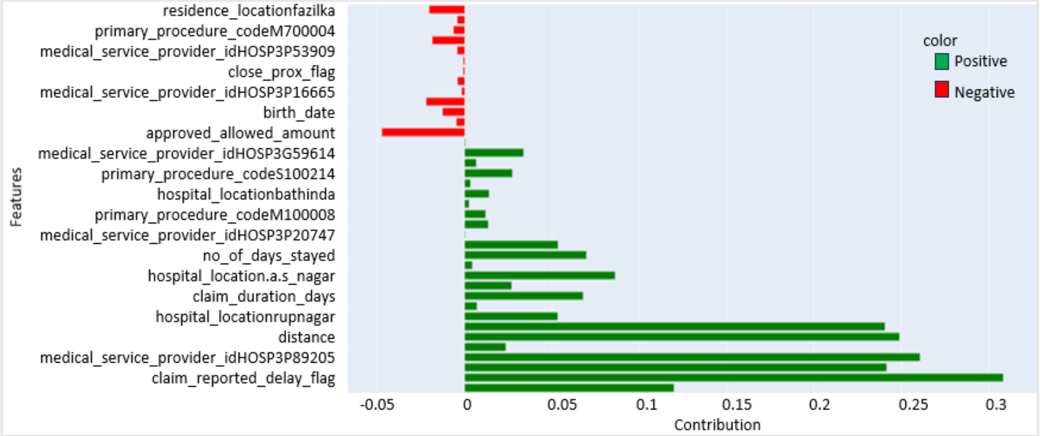



*Using Interpretable Machine Learning Methods: An Application to Health Insurance Fraud Detection | SOA

# Feature Importance Results

| | SHAP | | | PFI | | |
|---|---|---|---|---|---|---|
| Variable | Random Forest | XGBoost | GLM | Random Forest | XGBoost | GLM |
| claim_reported_delay_flag | 1 | 2 | 1 | 1 | 2 | 2 |
| distance | 2 | 4 | 11 | 3 | 3 | 11 |
| no_of_days_stayed | 9 | 1 | 11 | 2 | 1 | 11 |
| claim_duration_days | 3 | 3 | 11 | 5 | 5 | 11 |
| claim_count_flag | 7 | 5 | 8 | 4 | 4 | 11 |
| gender_flag | 11 | 9 | 4 | 6 | 11 | 3 |
| Primary_procedure_codeS100214 | 8 | 11 | 2 | 11 | 11 | 1 |
| claim_amount_flag | 11 | 11 | 3 | 11 | 11 | 5 |
| birth_date | 4 | 8 | 11 | 11 | 8 | 11 |
| Hospital_locationrupnagar | 5 | 11 | 11 | 11 | 7 | 9 |
| hospital_locationmoga | 11 | 11 | 11 | 7 | 10 | 4 |
| approved_allowed_amount | 6 | 6 | 11 | 11 | 11 | 11 |
| Medical_service_provider_idHOSP3G81376 | 11 | 11 | 11 | 11 | 6 | 6 |
| Hospital_locationbathinda | 10 | 7 | 7 | 11 | 11 | 11 |
| Primary_procedure_codeM100068 | 11 | 11 | 5 | 11 | 11 | 8 |
| Primary_procedure_codeM100009 | 11 | 11 | 6 | 11 | 11 | 11 |
| close_prox_flag | 11 | 10 | 11 | 8 | 11 | 11 |
| medical_service_provider_idHOSP3P12536 | 11 | 11 | 11 | 11 | 11 | 7 |
| hospital_distance_flag | 11 | 11 | 10 | 11 | 11 | 10 |
| Medical_service_provider_idHOSP3P10675 | 11 | 11 | 9 | 11 | 11 | 11 |
| hospital_locations.a.s_nagar | 11 | 11 | 11 | 9 | 11 | 11 |
| Medical_service_provider_idHOSP3P20747 | 11 | 11 | 11 | 11 | 9 | 11 |
| Medical_service_provider_idHOSP3G8137 | 11 | 11 | 11 | 10 | 11 | 11 |

SHAP – SHapley Additive exPlanations
PFI – Permutation Feature Importance

Institute
and Faculty
of Actuaries

https://ifoadatascienceresearch.github.io/
https://cads.sssihl.edu.in/projects/

github.com/karol-gawlowski
github.com/rohanyashraj

linkedin.com/in/karol-gawlowski/
linkedin.com/company/sssihl-cads/

karol.gawlowski@bayes.city.ac.uk
coordinator@cads.sssihl.edu.in

**IFoA India Conference 2024**

# Thank You!