FRAUDLENT CLAIM DETECTION

Vibha garg

Siva sankar

Vinay saraf

EXECUITVE SUMMARY

- Analyzed claim data to identify fraudulent patterns.
- Model to detect fraud with high accuracy.
- Identified key risk indicators such as high claim amounts and long approval times.

BUSINESS PROBLEM

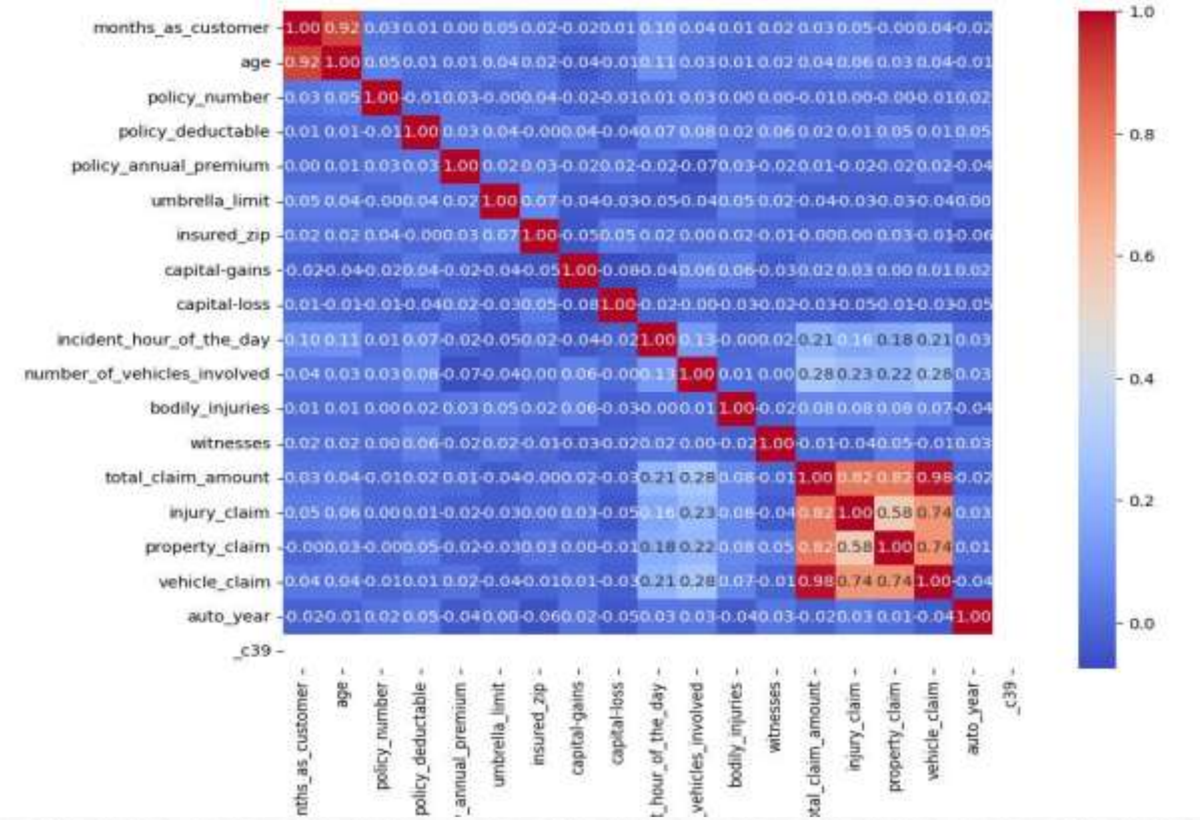
- Insurance fraud leads to financial losses annually.
- Manual claim review are time-consuming and inconsistent.
- Need for an automated ,data –driven fraud detection solution.
- Detect and prevent fraudulent claims early in the process.

DATA OVERVIEW

- Data includes claim amount, claim tyoe,customer demographics and approval time.
- Target varaiable fraudulent or legitimate claim.
- Balanced mix of numerical and categorical fetures.
- Revelaed some missing values and outilers.

EXPLORATORY DATA ANALYSIS (EDA)

- Fraudulent claim generally have high amount and longer approval times.
- Visual tools used : histograms , box plotsnd correlation heatmaps.
- Detected and key patterns that feature selections.



FEATURE ENGINEERI NG& PREPROCESSING

- ❑Handled missing values using imputation techniques.
- ❑Converted categorical varaiable using one hot encoding.
- ❑Normalized numerical features for better model performance
- ❑Created new features such as claim to income ratio and approval time bins.

MODELING APPROACH

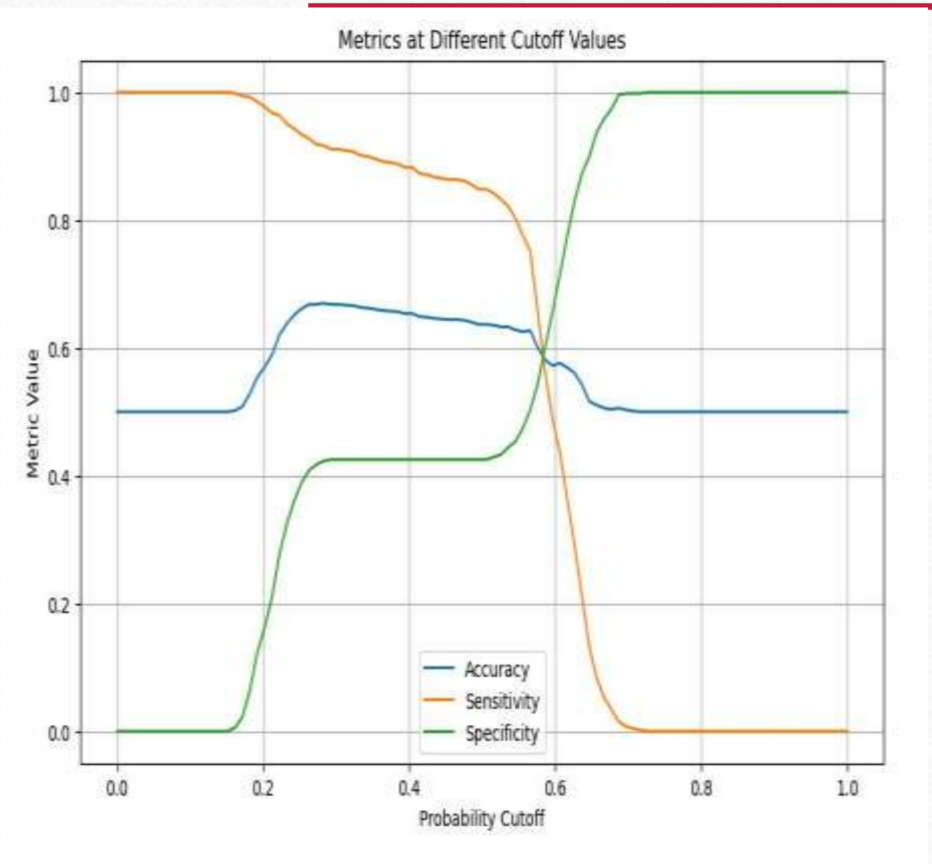
- Logistics Regression & random forest
  - Random forest was selected for its balance of accuracy and interpretability
  - Performed hyper parameter cross-validation
- Split data into traning and testing sets 80/20

```
Optimization terminated successfully.
         Current function value: 0.518800
         Iterations 6
                           Logit Regression Results
==============================================================================
Dep. Variable:         fraud_reported   No. Observations:                  700
Model:                          Logit   Df Residuals:                      694
Method:                           MLE   Df Model:                            5
Date:                Sun, 04 May 2025   Pseudo R-squ.:                 0.07220
Time:                        15:39:35   Log-Likelihood:                -363.16
converged:                       True   LL-Null:                       -391.42
Covariance Type:            nonrobust   LLR p-value:                 6.339e-11
==================================================================================================
                                coef    std err          z      P>|z|      [0.025      0.975]
--------------------------------------------------------------------------------------------------
const                        29.1570     31.257      0.933      0.351     -32.106      90.420
months_as_customer            0.0007      0.001      0.827      0.408      -0.001       0.002
policy_deductable          8.706e-05      0.000      0.592      0.554      -0.000       0.000
incident_hour_of_the_day     -0.0010      0.013     -0.076      0.939      -0.027       0.025
auto_year                    -0.0150      0.016     -0.962      0.336      -0.046       0.016
collision_type_Side Collision -1.5834     0.242     -6.547      0.000      -2.057      -1.109
==================================================================================================
```

Model Interpretation

MODELING EVALUATION

- Evaluation model using accuracy precision , recall .
- High recall ensured most fraudulent claims were detected.
- Matrix and Roc curve used to visualize performance .
- 90% accuracy .



BUSINESS IMPACT RECOMMEDATIONS

- Automated fraud detection can reduce manual workload and financial losses.
- Use model predictions to flag high- risk claimfor review.
- Continuously update the model with new data to maintain performance.
- Existing claim processing work flows.

ANSWERS TO THE QUESTIOS:

- **How can we analyse historical claim data to detect patterns that indicate fraudulent claims.**

Exploratory Data Analysis (EDA) and machine learning models are used to examine past claim data in order to find trends, correlations, and anomalies. Fraudulent activity is sometimes indicated by patterns such as excessively large claim amounts, frequent claims, or lengthy approval times.

- **Which features are the most predictive of fraudulanet behaviour.**

Features such as claim amount, approval time, claim type, and customer claim history were found to be highly predictive. The model highlighted that higher-than-average claim values and specific claim categories are closely associated with fraud.

- **Based on past data, can we predict the likelihood of fraud for an incoming claim?**

Yes, we can precisely assign a probability score indicating the possibility of fraud for every new claim by using labelled previous data to train classification models. This makes it possible to proactively identify dubious assertions for examination.

- **What insights can be drawn from the model that can help in improving the fraud detection process.**

The model helps prioritize claims by risk level, improving investigation efficiency. Insights such as feature importance allow insurers to focus on key fraud indicators and update policies to reduce vulnerabilities

THANK YOU