1) Problem Statement

Global Insure, a leading insurance company, processes thousands of claims annually. However, a significant percentage of these claims turn out to be fraudulent, resulting in considerable financial losses. The company's current process for identifying fraudulent claims involves manual inspections, which is time-consuming and inefficient. Fraudulent claims are often detected too late in the process, after the company has already paid out significant amounts. Global Insure wants to improve its fraud detection process using data-driven insights to classify claims as fraudulent or legitimate early in the approval process. This would minimize financial losses and optimize the overall claims handling process.

2) Business Objective

Global Insure wants to build a model to classify insurance claims as either fraudulent or legitimate based on historical claim details and customer profiles. By using features like claim amounts, customer profiles and claim types, the company aims to predict which claims are likely to be fraudulent before they are approved.

3) Experimental Evaluation

1. Data Preparation and Data Cleaning-
    - Manage the data having null values
    - Dropping the columns containing all null values
    - Replace junk values with meaning full values.
    - Fixing Datatype
2. Train Test Split
3. EDA-
    - To understand the patter of data
    - To understand the relation of dependent variables with independent variables.
    - Strong correlation among dependent variables
4. Feature Creation-
    - Get understanding of data from data and eliminated the columns which not not that much important for our purpose
    - Remapped the Categorical columns with low value counts
    - Creating dummy variables
5. Model chosen for the purpose
    - Logistic Regression

- Random Forest

4) Future Work

We found logistic Regression model is performing better in our case. Although definitely there is a scope where we can try by changing different parameters and applying various other methodology to see if the performance of the model is improving.

5) Conclusion

We are interested to get fraud cases properly. In our case logistic regression is showing more overall accuracy, precision, recall/sensitivity.