

Fraud Detection – Problem Statement

Develop a model for predicting fraudulent transactions for a financial company and use insights from the model to develop an actionable plan. Data for the case is available in CSV format having 6362620 rows and 11 columns.

You have flexibility in choosing the machine learning approach for this task. Adhere to standard model development practices, including training on calibration data and evaluating performance on validation data. This project demands a combination of statistical expertise and innovative thinking.

The task is to execute the process for proactive fraud detection while answering the following questions:

1. Data cleaning including missing values, outliers and multi-collinearity.
2. Describe your fraud detection model in elaboration.
3. How did you select variables to be included in the model?
4. Demonstrate the performance of the model by using best set of tools.
5. What are the key factors that predict fraudulent customer?
6. Do these factors make sense? If yes, How? If not, How not?
7. What kind of prevention should be adopted while company update its infrastructure?
8. Assuming these actions have been implemented, how would you determine if they work?