CREDIT CARD FRAUD DETECTION

NEHA MURALIDHARAN

Credit card is one of the most commodious financial products that can be used for big ticket items or even everyday purchases. Credit cards are extremely advantageous beyond convenience, ergo prone to fraud. As the name suggests credit card fraud is when your credit card or credit account is used to make an unauthorized payment or transaction. In recent years credit card fraud has become a growing issue. Machine learning is considered as one of the most successful technique to identify the fraud. The performance of fraud detecting in credit card transactions is greatly affected by the sampling approach on data-set, selection of variables and detection techniques used. The dataset used in this project (from Kaggle) is highly imbalanced holding a total of 2,84,808 observations. The overall dataset contains 2,84,315 true transactions (class 0) and 492 false transactions (class 1) (99.83% and 0.17% respectively). The main objectives of this project are: To Analyze and understand the chosen dataset; to clean the chosen dataset in order to predict accurately i.e., fixing or removing incorrect, corrupted, incorrectly formatted, duplicate, or incomplete data within the dataset if any; To analyze the effect of handling the imbalanced data in three ways namely random over sampling, random under sampling and smote; To Predict the Presence of fraud using Classification techniques.

Key words: Fraud detection, Credit card, random over sampling, random under sampling, smote.