

SAS - Fraud Detection Using Descriptive, Predictive, and Social Network Analytics

| | |
|----------------|-----------------------------|
| Code: | BFRSUSN |
| Length: | 2 days |
| URL: | View Online |

A typical organization loses an estimated 5% of its yearly revenue to fraud. This course shows how learning fraud patterns from historical data can be used to fight fraud. The course discusses the use of supervised learning (using a labeled data set), unsupervised learning (using an unlabeled data set), and social network learning (using a networked data set). The techniques can be applied across a wide variety of fraud applications, such as insurance fraud, credit card fraud, anti-money laundering, healthcare fraud, telecommunications fraud, click fraud, tax evasion, and counterfeiting. The course provides a mix of both theoretical and technical insights, as well as practical implementation details. During the course, the instructor reports extensively on his recent research insights about the topic. Various real-life case studies and examples are presented for further clarification.

Skills Gained

- preprocess data for fraud detection (sampling, missing values, outliers, categorization, and so on)
- build fraud detection models using supervised analytics (logistic regression, decision trees, neural networks, ensemble models, and so on)
- build fraud detection models using unsupervised analytics (hierarchical clustering, non-hierarchical clustering, k-means, self organizing maps, and so on)
- build fraud detection models using social network analytics (homophily, featurization, egonets, PageRank, bigraphs, and so on).

Who Can Benefit

- Fraud analysts, data miners, and data scientists; consultants working in fraud detection; validators auditing fraud models; and researchers in financial services companies, banks, insurance companies, government institutions, healthcare institutions, and consulting firms

Prerequisites

- Before attending this course, you should have a basic knowledge of statistics, including descriptive statistics, confidence intervals, and hypothesis testing.

Course Details

Introduction

Fraud Detection

- the importance of fraud detection
- defining fraud
- anomalous behavior
- fraud cycle
- types of fraud
- examples of insurance fraud and credit card fraud
- key characteristics of successful fraud analytics models
- fraud detection challenges
- approaches to fraud detection

Data Preprocessing

- motivation
- types of variables
- sampling
- visual data exploration
- missing values
- outlier detection and treatment
- standardizing data
- transforming data
- coarse classification and grouping of attributes
- recoding categorical variables
- segmentation
- variable selection

Supervised Methods for Fraud Detection

- target definition
- linear regression
- logistic regression
- decision trees
- ensemble methods: bagging, boosting, random forests
- neural networks
- dealing with skewed class distributions
- evaluating fraud detection models

Unsupervised Methods for Fraud Detection

- unsupervised learning
- clustering approaches: hierarchical clustering, k-means clustering, self-organizing maps
- peer group analysis
- break point analysis

Social Networks for Fraud Detection

- social networks and applications
- is fraud a social phenomenon?
- social network components
- visualizing social networks
- social network metrics
- community mining
- social network based inference (network classifiers and collective inference)
- from unipartite toward bipartite graphs
- featurizing a bigraph
- fraud propagation
- case study

Fraud Analytics: Putting It All to Work

- quantitative monitoring: backtesting, benchmarking
- qualitative monitoring: data quality, model design, documentation, corporate governance

Schedule (as of 3)

| Date | Location |
|-------------|-----------------|
|-------------|-----------------|
