

Unsupervised Learning for the Analysis and Detection of Fraud in the Insurance Industry

José A. Álvarez-Jareño, José Manuel Pavía

Applied Economic Department, University of Valencia, Spain.

Abstract

Analysis and detection of fraud in the insurance sector has traditionally been carried out through supervised learning. The main problem is that the data presents a strong imbalance and techniques are used to balance the variable. Unsupervised learning is an alternative to consider, especially anomaly detection methodologies. If the fraud variable has a significant imbalance, then it can be treated as an anomaly. That is, the behavior of the fraudsters must be different from the rest of the insured.

The main methodologies used are Isolation Forest, Attribute wise learning for scoring outliers (ALSO), Trimmed K-means, Autoencoders (neural networks) and Principal Component Analysis. The objective is through dimensionality reduction techniques to obtain a model with which to make predictions. The instances that present greater differences between the real values and the values estimated with this methodology will be considered anomalies and analyzed as if they were fraud.

The results obtained show that these methodologies can be used as a complement to supervised learning. The assembly of models will allow the integration of both methodologies and improve the detection of fraud in the insurance sector.

Keywords: *Insurance; Fraud, Unsupervised Learning, Isolation Forest, ALSO, Autoencoders.*
