

## **Evaluation of Three Data Mining Algorithms (Decision Tree, Naive Bayes, Logistic Regression) in Auto Insurance Fraud Detection**

**Atousa Goodarzi\***  
**Sadjad Janatbabaei\*\***

### **Abstract**

From an economic standpoint, financial fraud in the insurance industry is becoming an increasingly serious problem. Nowadays, data mining techniques are commonly used to develop detection models and fraud patterns. Data mining plays an important role in financial fraud detection, helping to uncover hidden facts from among very large quantities of data. Insurance companies can discover hidden patterns in the data and, with the use of well-designed models, make efforts to reduce fraud, manage risks and fulfill their commitments.

In this paper, current techniques for detecting auto insurance fraud (logistic regression, decision tree, Naive Bayes) are used. Data mining techniques are typically used to develop models that can spot patterns in fraudulent claims. The techniques introduced are applied to real data in order to also offer predictions.

Seven explanatory variables have been examined in the three models. These variables are: policy age, number of claims, delay, age, gender, police sketches, amount of loss. After obtaining the necessary permits, the data required was collected from one of Iran's largest insurance companies. The data was then divided into two sections, one to develop models and the other for classification. Research findings showed that the logistic regression model demonstrated greater accuracy compared to the other two models in predicting total claims (fraudulent and non-fraudulent).

**Keywords:** Data mining; Fraud; Auto insurance; Decision tree; Naive Bayes; Logistic regression

**JEL:** G22, C21, C11

---

\* Associate Professor, ECO College of Insurance, Allameh Tabataba'i University, (Corresponding Author) Atousagoodarzi@yahoo.com

\*\* M.Sc. in Actuarial Science, ECO College of Insurance, Allameh Tabataba'i University, S.janatbabaei@gmail.com