**LITERATURE SURVEY**

1. **Statistical outlier detection method for identifying anomalies in transactional data**

**AUTHORS:** Bolton and Hand

The detection of fraudulent credit-card transactions has been an active area of research for several years, owing to the growing number of online payment systems and digital banking platforms. Numerous researchers have explored statistical, machine-learning, and hybrid techniques to improve accuracy and minimize financial losses caused by fraudulent activities. This section presents a review of related work and highlights the contributions that informed the proposed hybrid ensemble framework. Early studies focused primarily on rule-based expert systems and statistical modeling. These systems relied on pre-defined thresholds and manually crafted rules derived from domain experts. While effective for detecting known fraud patterns, such systems were unable to adapt to new or evolving attack behaviors. Bolton and Hand (2002) proposed a statistical outlier detection method for identifying anomalies in transactional data. Although it provided a foundation for fraud detection, the approach struggled with high-dimensional data and the imbalance between legitimate and fraudulent transactions.

**2. Logistic Regression (LR)**, **Decision Trees (DT), and Support Vector Machines (SVM)** **to improve model adaptability.**

**AUTHORS:** Brause et al., Patidar and Sharma

Subsequent research adopted machine-learning approaches such as Logistic Regression (LR), Decision Trees (DT), and Support Vector Machines (SVM) to improve model adaptability. Brause et al. (1999) used neural networks for early credit-card fraud detection, demonstrating improved learning capability over rule-based methods. Patidar and Sharma (2011) explored the use of SVM and Random Forest (RF) classifiers for real-time detection, showing that ensemble methods tend to outperform single-model approaches. However, these methods often faced the class-imbalance problem, which led to biased models that favored legitimate transactions.

**3. Ensemble and hybrid frameworks, combining multiple models to exploit their complementary strengths.**

**AUTHORS:** Sahin et al., Dal Pozzolo et al., and Carcillo et al.

Recent research has increasingly focused on ensemble and hybrid frameworks, combining multiple models to exploit their complementary strengths. Sahin et al. (2013) developed a hybrid model that integrated decision trees with neural networks, achieving better generalization and reduced false positives. Similarly, Dal Pozzolo et al. (2015) emphasized the importance of sampling techniques such as SMOTE and undersampling to balance the dataset before model training. They showed that balanced training data significantly enhances detection accuracy for minority (fraudulent) classes. More contemporary approaches employ deep learning and gradient-boosting algorithms to capture complex nonlinear relationships in data. XGBoost, LightGBM, and CatBoost have emerged as powerful tools for handling structured transactional datasets. Carcillo et al. (2019) demonstrated that hybrid ensembles combining tree-based learners and neural networks can adapt effectively to concept drift—the gradual change in fraud patterns over time. Hybrid frameworks therefore represent a promising direction for building scalable, high-accuracy fraud-detection systems. The reviewed literature indicates that while single classifiers can achieve moderate accuracy, hybrid and ensemble learning approaches consistently deliver higher detection rates, better precision–recall balance, and enhanced robustness. However, the key challenge remains designing an architecture that effectively integrates multiple classifiers without incurring excessive computational cost or overfitting. The proposed project builds upon these findings by introducing a **hy**brid ensemble architecture that integrates Logistic Regression, Random Forest, SVM, and XGBoost models within a unified framework. The ensemble leverages voting and stacking strategies to combine model predictions, supported by a resampling technique (SMOTE) to address data imbalance. This configuration aligns with contemporary research trends and demonstrates superior predictive performance when compared with traditional, single-model systems.