**Aluno: Joao Paulo Pires Dantas**

O autor do artigo trata da análise de dados de transações de cartões de crédito para classificar quais delas são transações fraudulentas. Para essa tarefa, são utilizados algoritmos de Machine Learning KNN, Random Forest e Gradient Boosting.

A identificação do melhor algoritmo é realizada pela comparação da precisão, revocação e F-score. A precisão apresenta a razão entre as transações fraudulentas e as transações que o algoritmo previu fradulenta. A revocação ou sensibilidade apresenta a razão entre a previsão de transações fraudulentas e as transações efetivamente fraudulentas. Por fim o F-score mede a acurácia dos modelos com base no cálculo da média harmônica da precisão e revocação.

A base de dados escolhida é composta por 28 classes anônimas, as classes Time, Amount e a classe Fraude. Para ajudar a reduzir a dimensionalidade dessa base é utilizado o método não-linear t-SNE que visualmente ajuda a escolher a melhor composição de classes para modelagem dos dados pelo algoritmos. Ainda na redução dos atributos é utilizado o mapa de calor para identificação das correlações entre as classes. Esses métodos ajudaram a definir uma base completa e outra apenas com treze classes com correlação maior que 0,5 e menos que -0,5.

Para os algoritmos de treinamento e predição foram apurados que apontaram o Random Forest como melhor para classificar as transações fraudulentas. O KNN teve uma precisão de 0,86, porém a um custo de identificar poucas fraudes, ou seja, uma baixa sensibilidade e F-score, ambos, de apenas 0,10, em ambas as bases. Por sua vez, o Random Forest conseguiu identificar praticamente 100% das fraudes e classificou corretamente 72% delas. Em segundo lugar ficou o Gradient Boosting com uma sensibilidade de 0,77 e precisão de 0,58.

Esse comparativo mostra que existem algoritmos com pontos fortes e pontos fracos que devem ser utilizados conforme a necessidade. Em especial o Gradient Boost que não se destaca em nenhum aspecto é relevante informar que ele possui mais de 20 parâmetros para configuração no sklearn e estes foram utilizados os padrões da biblioteca, ou seja, há possibilidade de melhoria.

Referência ABNT:

Paul, A. d. C. B. Avaliação de métodos de machine learning na detecção de fraude em dados transacionais de cartão de crédito.