Hao Zhou, Hong-feng Chai, Mao-lin Qiu, 2018. Fraud detection within bankcard enrollment on mobile device based payment using machine learning. *Frontiers of Information Technology & Electronic Engineering*, 19(12):1537-1545. https://doi.org/10.1631/FITEE.1800580

Fraud detection within bankcard enrollment on mobile device based payment using machine learning

Key words: Fraud detection; Mobile payment; Bankcard enrollment; Mobile device based; GBDT; XGBoost

Corresponding author: Hao Zhou
E-mail: zhouhaounionpay@sjtu.edu.cn
ORCID: http://orcid.org/ 0000-0002-0869-6570

Motivations

1. When training a bankcard fraud detection model, it is important to use features that enable good classification.

2. Bankcard enrollment on mobile devices has become

the primary target of fraud attempts.

3. Existing rule-based expert system solutions are inadequate to face the challenges of data loss and social engineering.

4. The improved gradient boosting decision tree (GBDT) algorithm is used to detect bankcard fraud during the enrollment stage.

Main ideas

1. We introduced machine learning algorithms into a real dataset of mobile device based bankcard enrollment.

2. Device data is help recognize an unique device and analyse the mapping relationship between device and bankcard.

3. We aim to design a fraud scoring framework to reflect the possibility of fraud of bankcard enrollment with precision.

Methods

1. We compared three algorithms: random forest (RF), logistic regression (LR), and gradient boosting decision tree (GBDT).

2. We used XGBoost as our software library for implementation of GBDT.

3. We adopted a dataset provided by a worldwide payment processor. The dataset includes fraudulent and legitimate Chinese bankcard enrollment during the year of 2017.

Major results

GBDT-based model implemented by XGBoost significantly outperforms the other methods.

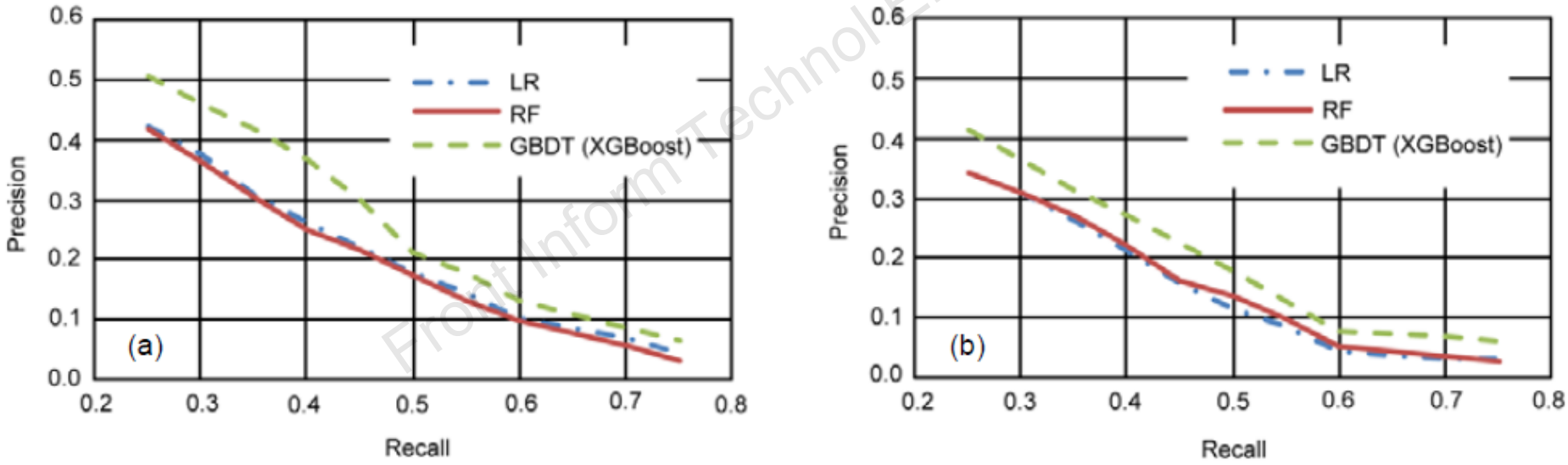

Fig. 2 Precision–recall curves in the dataset: (a) Apple Pay; (b) non-Apple pay

Conclusions

1. Machine learning performs better than expert rules, especially in differentiating high and low risk.

2. However, LR, RF, and GBDT have different performances. GBDT, implemented by XGBoost, is the most efficient.

3. We have implemented the GBDT-based model and scored framework in a real system.