Comparative Evaluation of Anomaly Detection Methods for Fraud Detection in Online Credit Card Payments

Hugo Thimonier[1], Fabrice Popineau[1], Arpad Rimmel[1], Bich-Liên Doan[1] and Fabrice Daniel[2]

[1] Université Paris-Saclay, CNRS, CentraleSupélec, Gif-sur-Yvette

[2]LUSIS AI, Paris

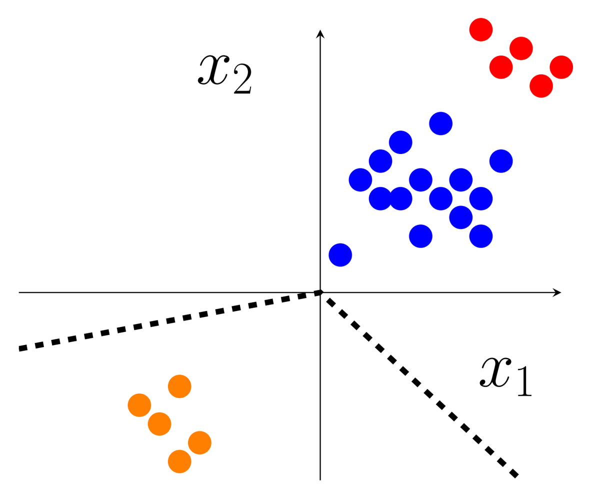
Paper

Context

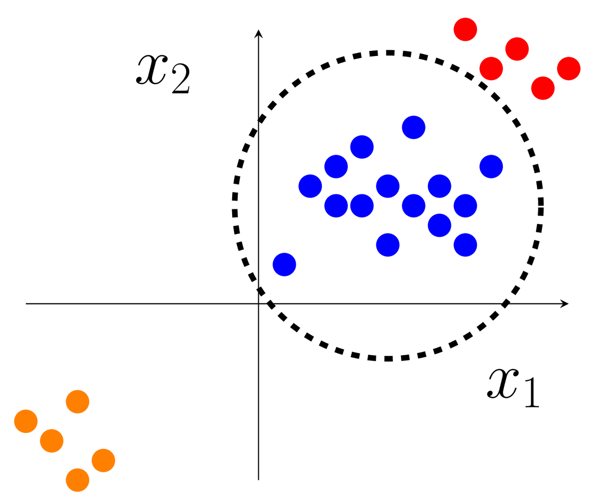
- Credit card fraud cost **$28.58 billion** in 2021 (Nilson Report).
- Traditional **rule-based fraud detection is costly** and requires continuous expert updates.
- **Gradient Boosted Decision Trees** (GBDT) are the top-performing models for tabular data.
- **Anomaly detection** methods emerge as a distinct class of algorithms designed to address the challenge of fraud detection.

Fraud Approach: Supervised or Anomaly Detection?

Supervised Learning
Construct a decision frontier using both normal samples and anomalies



Anomaly Detection:
Characterization of the *normal* distribution regardless of anomalies



●Normal sample ●Known anomaly ●Unseen anomaly

Comparative Evaluation

Let's compare LightGBM, a supervised learning approach, against anomaly detection methods.

Distribution shift

- Due to the Covid-19 pandemic, **consumption and payment behaviors changed** between the pre and post-Covid era.
- Hence, our dataset displays a **distribution shift** between the 2018-2019 and 2020-2021 periods

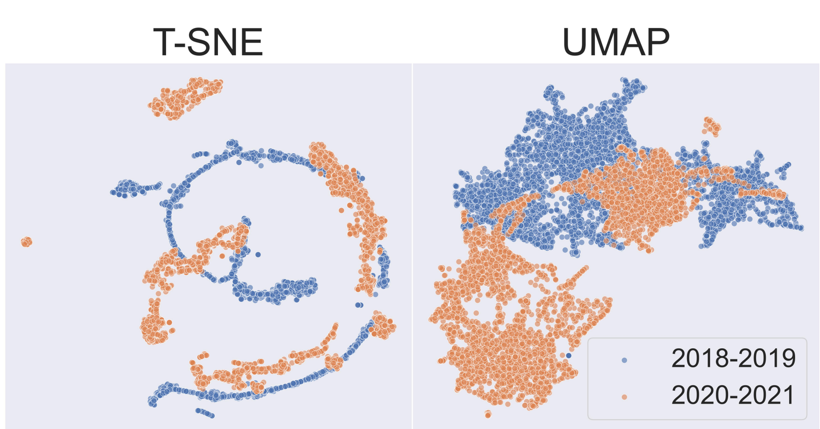


Figure: Country A

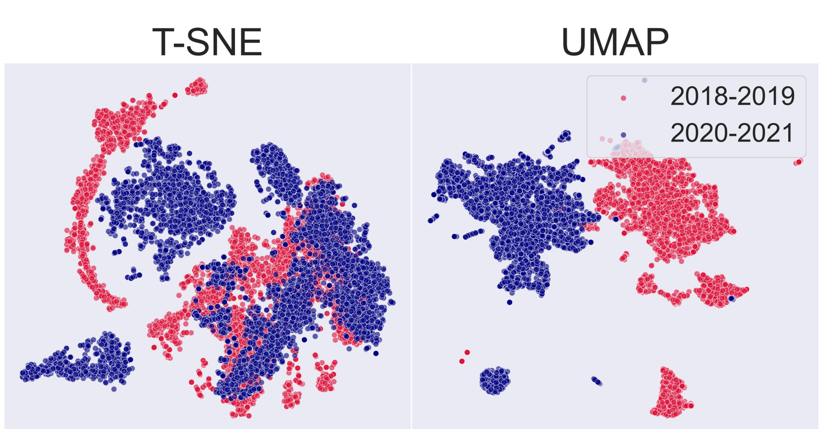
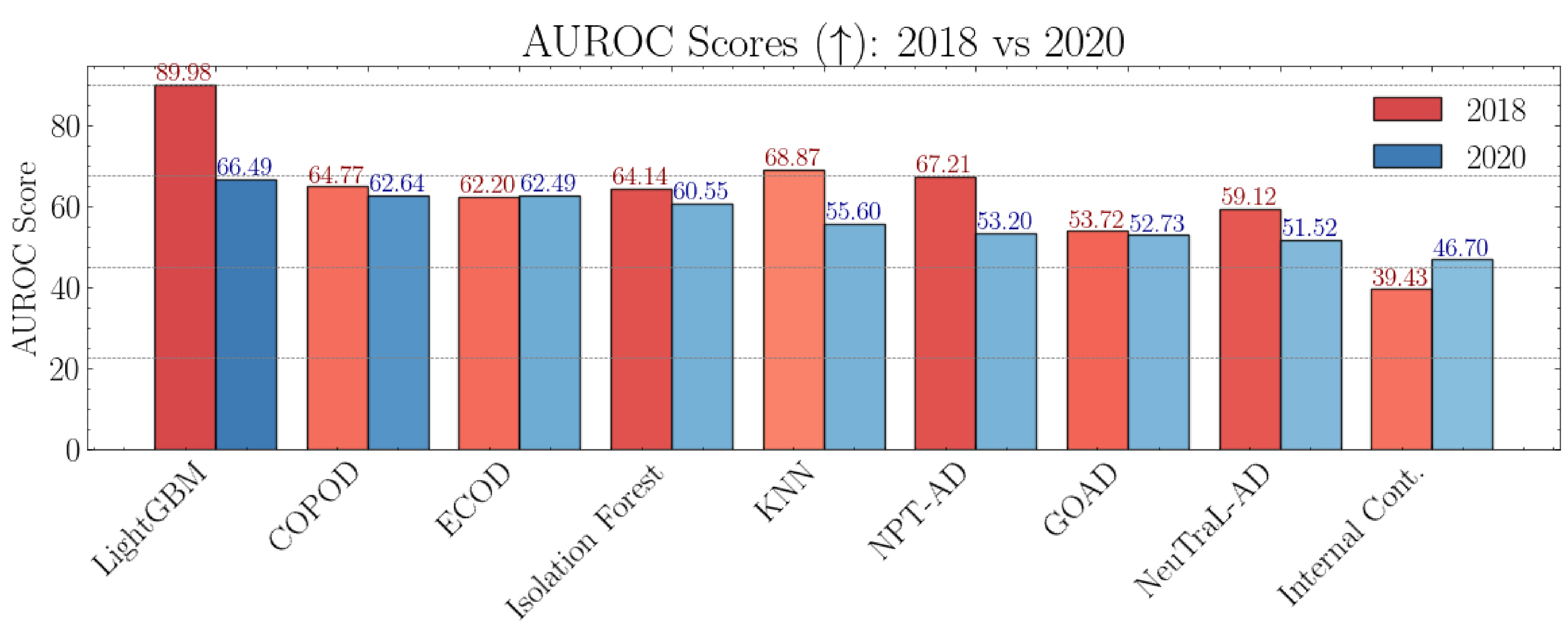


Figure: Country B



What is anomaly detection ?

Vanilla binary classification case:

- Training set composed of **samples belonging to both classes,** $y_i = 0$ **and** $y_i = 1$:
$$D_n^{train} = \{(x_i, y_i), x_i \in \mathcal{X}, y_i \in \{0, 1\}\}_{i=1}^n$$
- The goal is to directly learn a classifier using the training set
$$f : \mathcal{X} \to \{0, 1\}$$

Standard approaches to AD:

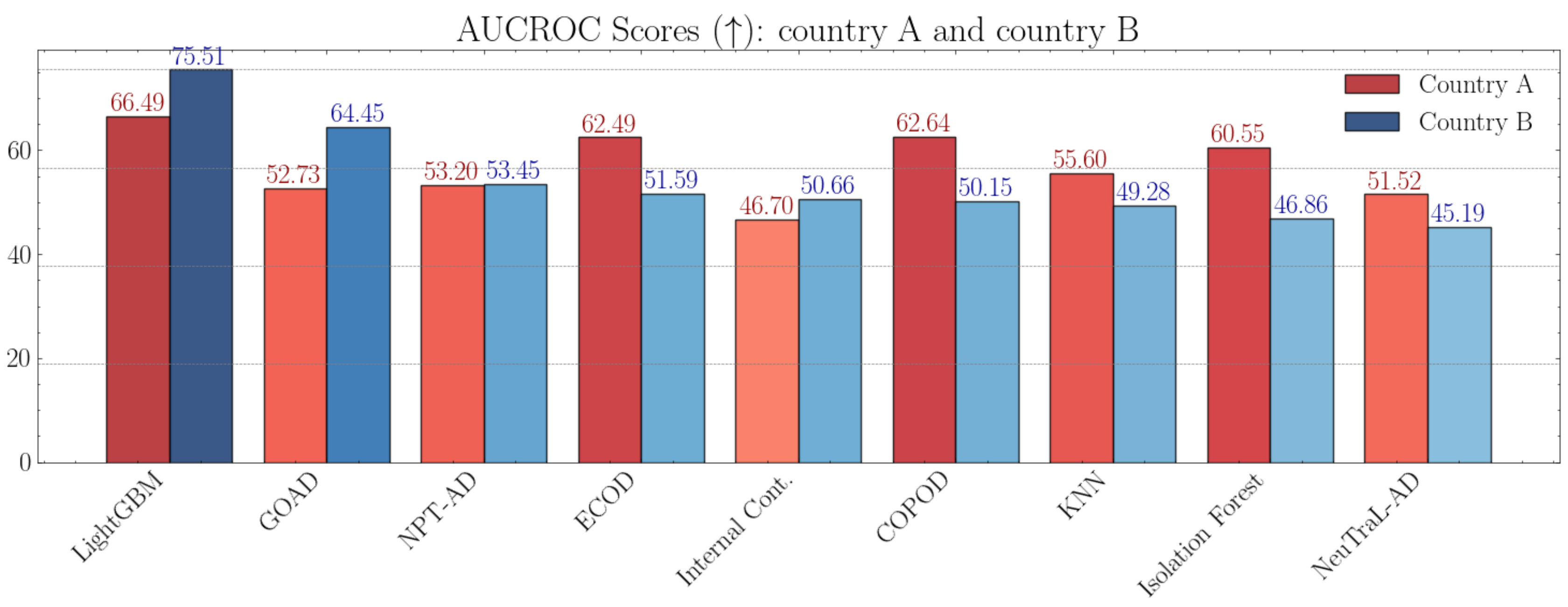
- Training set D_n^{train} solely composed of *normal* samples:
$$D_n^{train} = \{(x_i, y_i), y_i = 0\}_{i=1}^n$$
where $x_i \in \mathcal{X} \subseteq \mathbb{R}^d$, $y_i \in \mathcal{Y} = \{0, 1\}$.
- Most AD methods aim at **characterizing the distribution of the *normal* samples** ($y = 0$), $\mathbb{P}_{y=0}$.

Anomaly detection

Learns the distribution of **normal transactions** rather than explicitly classifying fraud cases.

Experiment

- **Real-life credit card payment** dataset made available to by a large french bank
- Frauds represents less than 1% of total **480 million transactions**
- We restrict our analysis to **two countries** (Country A and B) in which payments were made.



Conclusion

- While AD methods appear as good alternatives to standard supervised classification methods, when confronted with real-life settings, **all tested AD methods perform poorly**
- We do observe a severe degradation of performance between both period: **distribution shift does hinder the performance**

Supervised or Anomaly Detection?

For **real-world datasets**, supervised learning approaches, such as LightGBM, continue to outperform anomaly detection methods.