12.12.25

DABN14 – Project 1: Credit on balance

- Predicting credit card fraud on an imbalanced dataset

Erik Broms & William Nordansjö

1. Introduction, Data & Research Question
2. Imbalanced Data
3. Models & Theory
4. Results
5. Conclusion

1. **Introduction, Data & Research Question**
2. Imbalanced Data
3. Models & Theory
4. Results
5. Conclusion

“Fraud is the category of crime that has increased the most in 2023 according to the Swedish National Council for crime prevention.”

Credit Card Fraud Detection - Machine Learning Group of ULB (Université Libre de Bruxelles)

- 284,807 transactions of European customer over a 2-day period.
- Variables anonymised through Principal Components (28 PC's)
- Other variables:
  - Pre-labeled "class"
  - Amount
  - Time (from first transaction, not used)
- Heavily imbalanced: 0.172% fraudulent transactions

“Can we achieve meaningful performance on an imbalanced dataset by predicting fraudulent credit card transactions using machine learning models?”

1. Defining the Problem
2. **Imbalanced Data**
3. Models & Theory
4. Results
5. Conclusion

Imbalanced Data

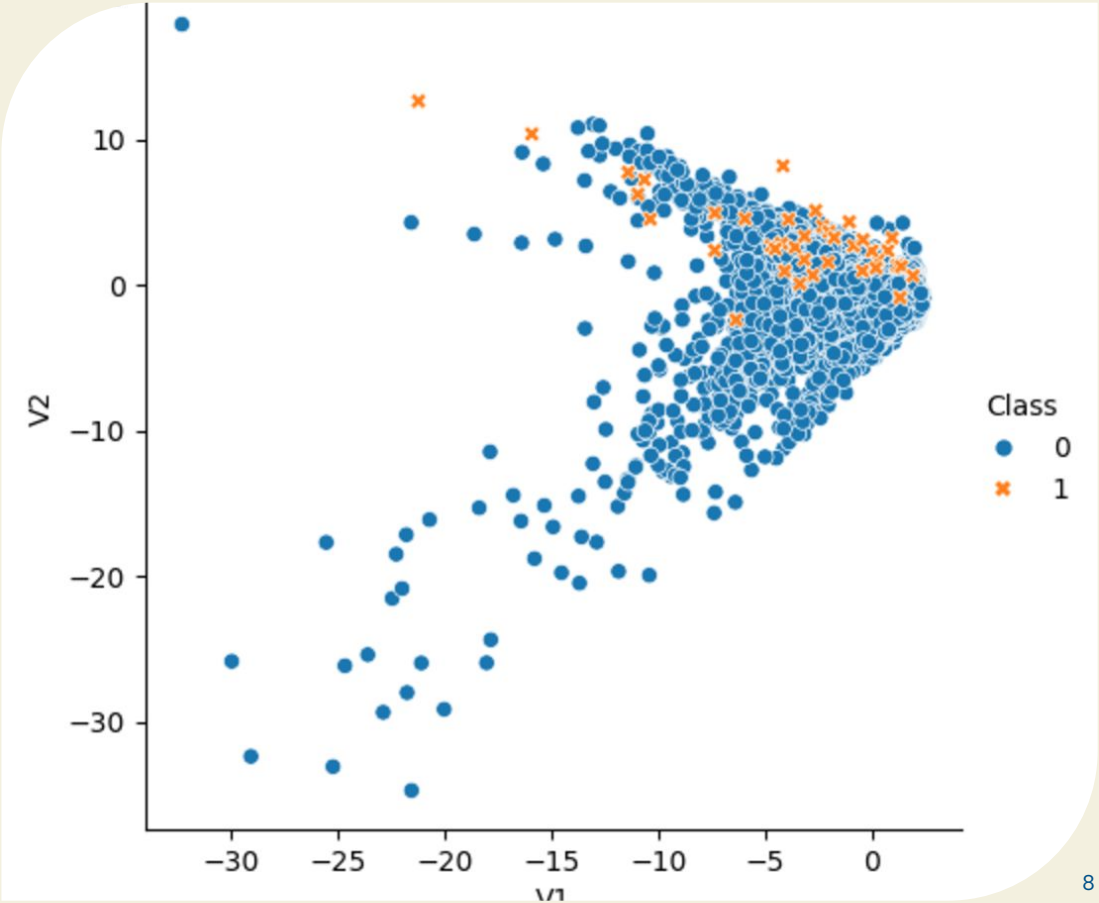
The first two principal components plotted with 10% of the data, with “class 1” (fraud) plotted above “class 0” (legitimate).

Problems:

- Training (Dastidar 2024, Lucas & Jurgovsky 2020)
- Evaluation (Saito & Rhemsmeier 2015)
  - Accuracy & AUC not good enough

Solution:

- SMOTE for training (Chawla et. al 2002)
- AUPRC, F_β-score for evaluation (Saito & Rehmsmeier 2015; Lindholm et. al 2022 (draft))



1. Defining the Problem
2. Imbalanced Data
3. **Models & Theory**
4. Results
5. Conclusion

“XGBoost is widely used by data scientists to achieve state of the art results.”

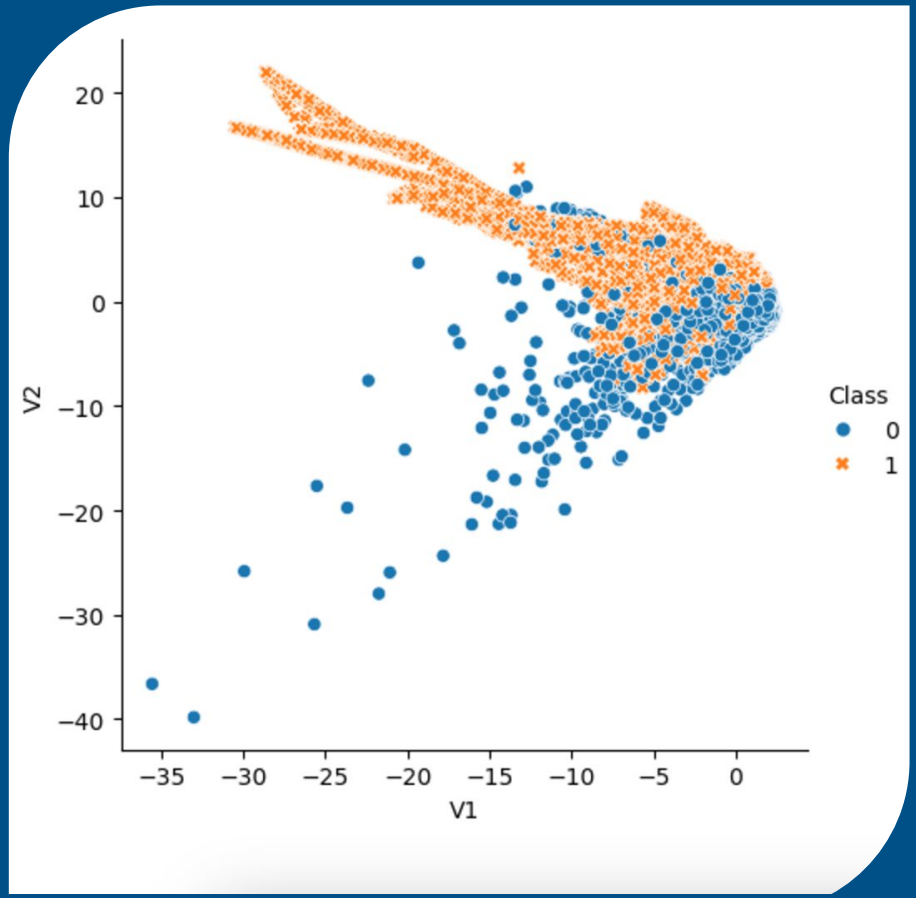
eXtreme Gradient Boosting (XGB)

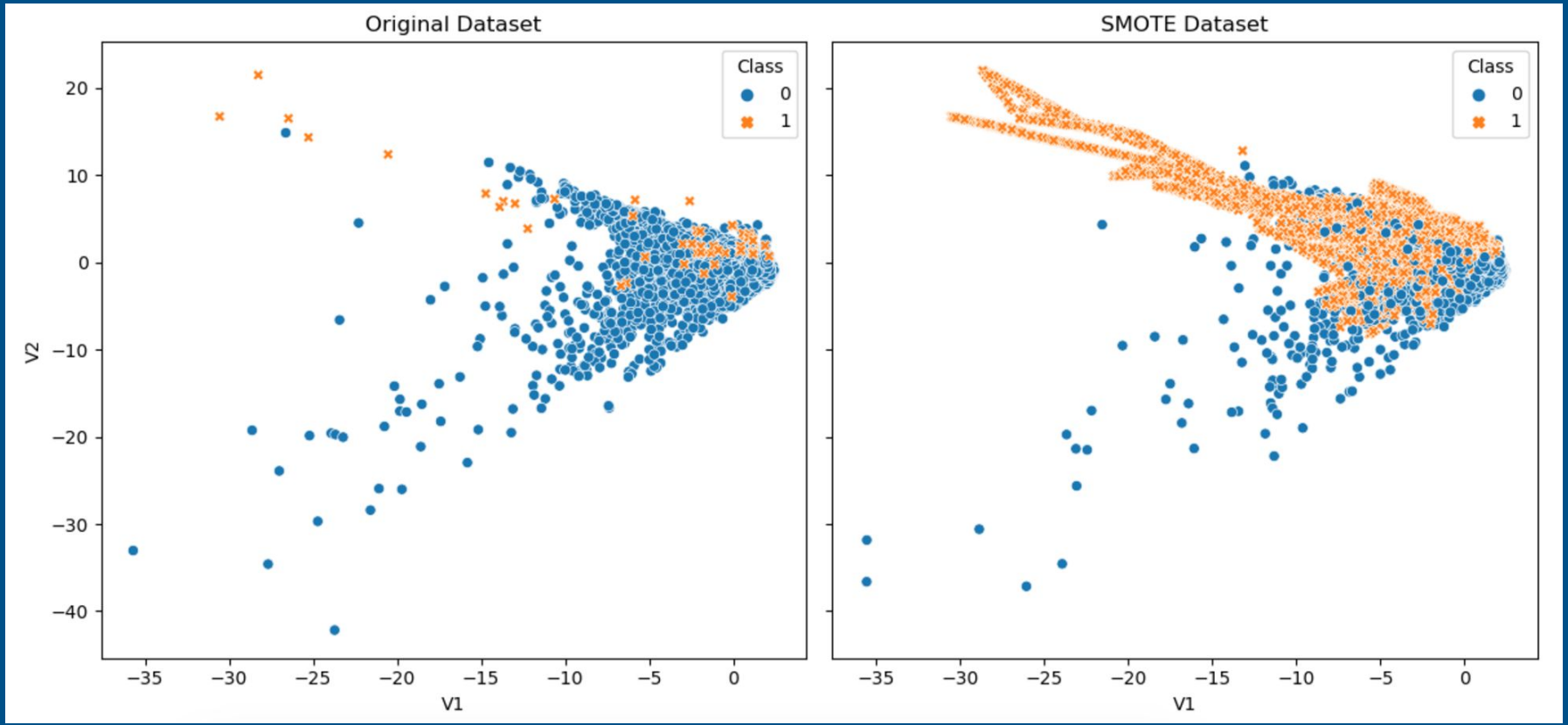
- Chen & Guestrin (2016)
- Builds upon the general boosting system
- Widely recognized performance in several machine learning and data mining challenges
- Scaleable, multiple tweakable hyperparameters
- Will be compared to a standard and simple Decision Tree model

Synthetic Minority Oversampling TechniquE (SMOTE)

The “not majority” class was oversampled until they became equalized. The plot below shows the result of oversampling on the first two principal components using 10% of the data.

- Synthetically create a new datapoint based on one of K=5 *Nearest Neighbours*
- Chawla et. al (2002)





Note: The 10% of the data shown is randomly chosen and thus differs slightly in the plots

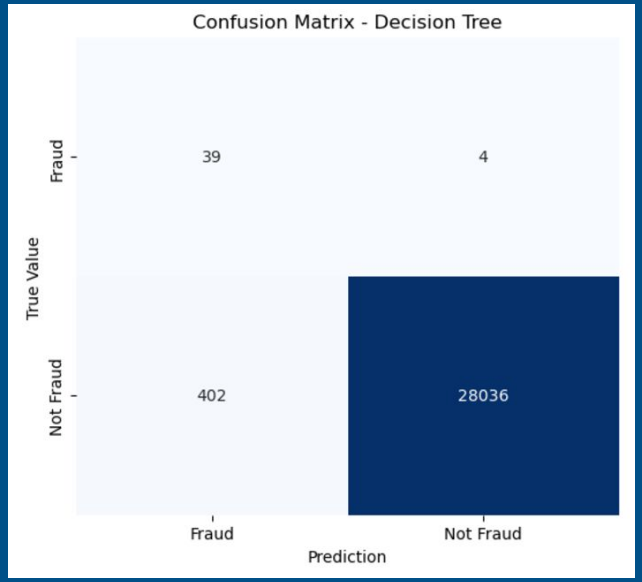
1. Introduction, Data & Research Question
2. Imbalanced Data
3. Models & Theory
4. **Results**
5. Conclusion

Confusion Matrix - DT

Highlights:

- Accuracy: 99%

For class 1
- Precision: 0.09
- Recall: 0.91
- F_1-score: 0.16



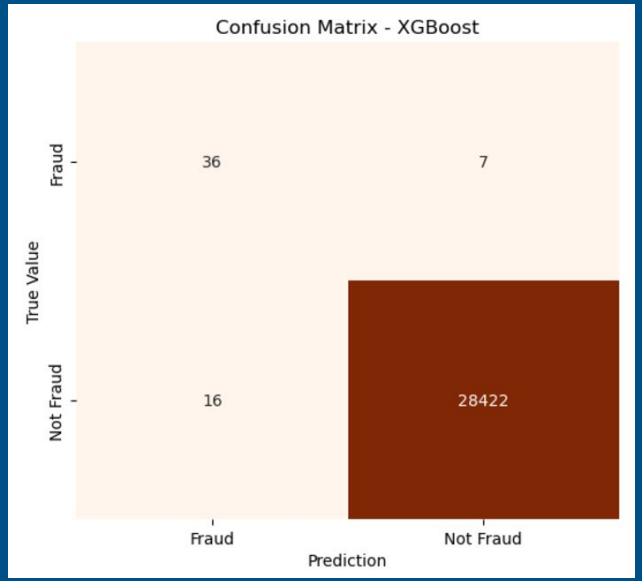
	precision	recall	f1-score	support
0	1.00	0.99	0.99	28438
1	0.09	0.91	0.16	43
accuracy			0.99	28481
macro avg	0.54	0.95	0.58	28481
weighted avg	1.00	0.99	0.99	28481

Confusion Matrix - XGB

Highlights:

- Accuracy: 100%

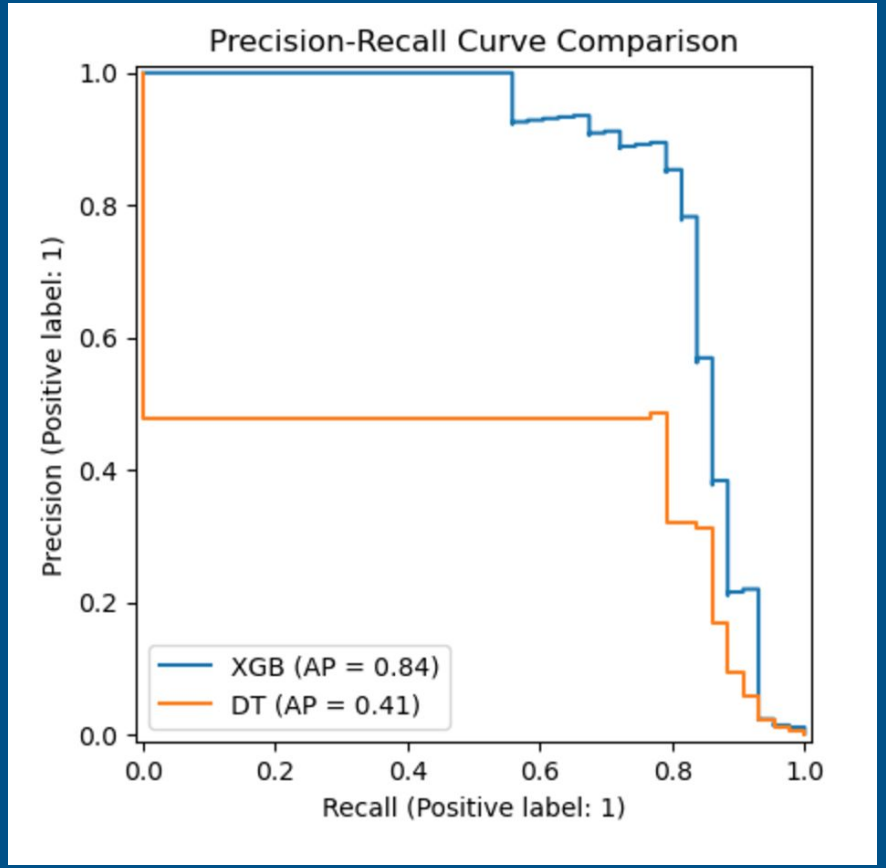
For class 1
- Precision: 0.69
- Recall: 0.84
- F_1-score: 0.76



	precision	recall	f1-score	support
0	1.00	1.00	1.00	28438
1	0.69	0.84	0.76	43
accuracy			1.00	28481
macro avg	0.85	0.92	0.88	28481
weighted avg	1.00	1.00	1.00	28481

Precision -Recall curve

- Plots Precision and Recall across all hypothetical decision thresholds
- AUPRC ≈ AP
  - XGB AP = 0.84
  - DT AP = 0.41
- XGB model outperformed the Decision Tree model

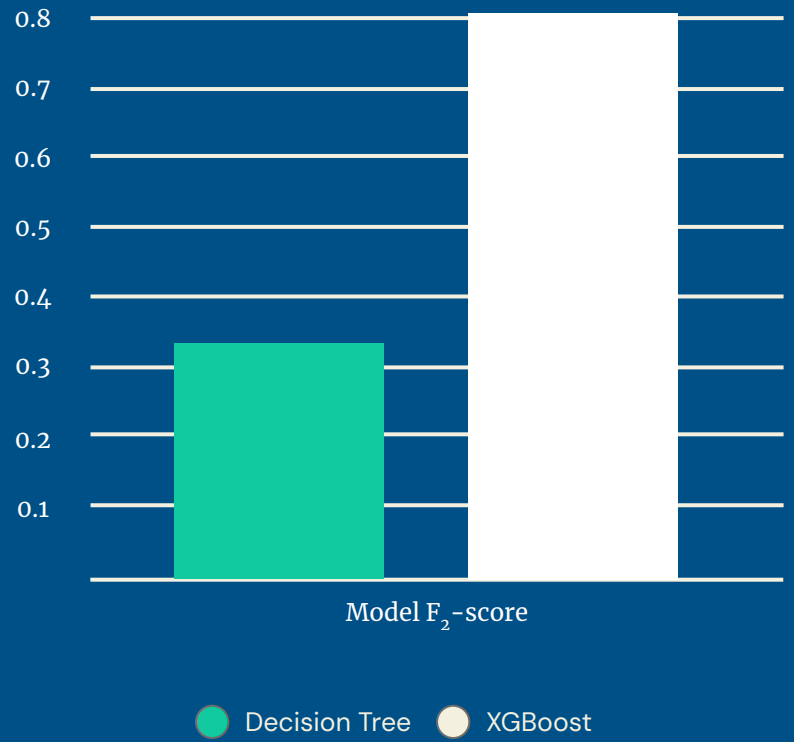


$$F_\beta = \frac{(1+\beta^2)precision * recall}{\beta^2 * precision + recall},$$

F_β-score

Where β = 2, meaning that we value Recall twice as high as Precision

- DT model F_2- score: 0.318
- XGB model F_2- score: 0.804



1. Introduction, Data & Research Question
2. Imbalanced Data
3. Models & Theory
4. Results
5. **Conclusion**

Conclusion

Meaningful performance

- Accuracy is high with both models
- Imbalanced data requires special consideration
  - Training
  - Evaluation
- XGB model superior because it balances Precision and Recall while still prioritising Recall