Detection of fraudulent financial transactions using AI

Lydie Rosenkrancová feat Václav Šmídl

Presentation outline

- Introduction into the problematics of Fraud Detection

- The dataset

- Experiments with following models:

  - SVM

  - Random Forest

  - MLP

# The problematics of fraud detection:

- Fraud detection is, given a set of credit card transactions, the process of identifying if a new authorised transaction belongs to the class of fraudulent or genuine transactions

- Challenges of fraud detection:

  ‣ Highly unbalanced classes distributions

  ‣ Non-stationary distribution of the data

  ‣ Availability of few transactions labeled by fraud investigators

 ‣ Public data is scarcely available due to confidentiality issues
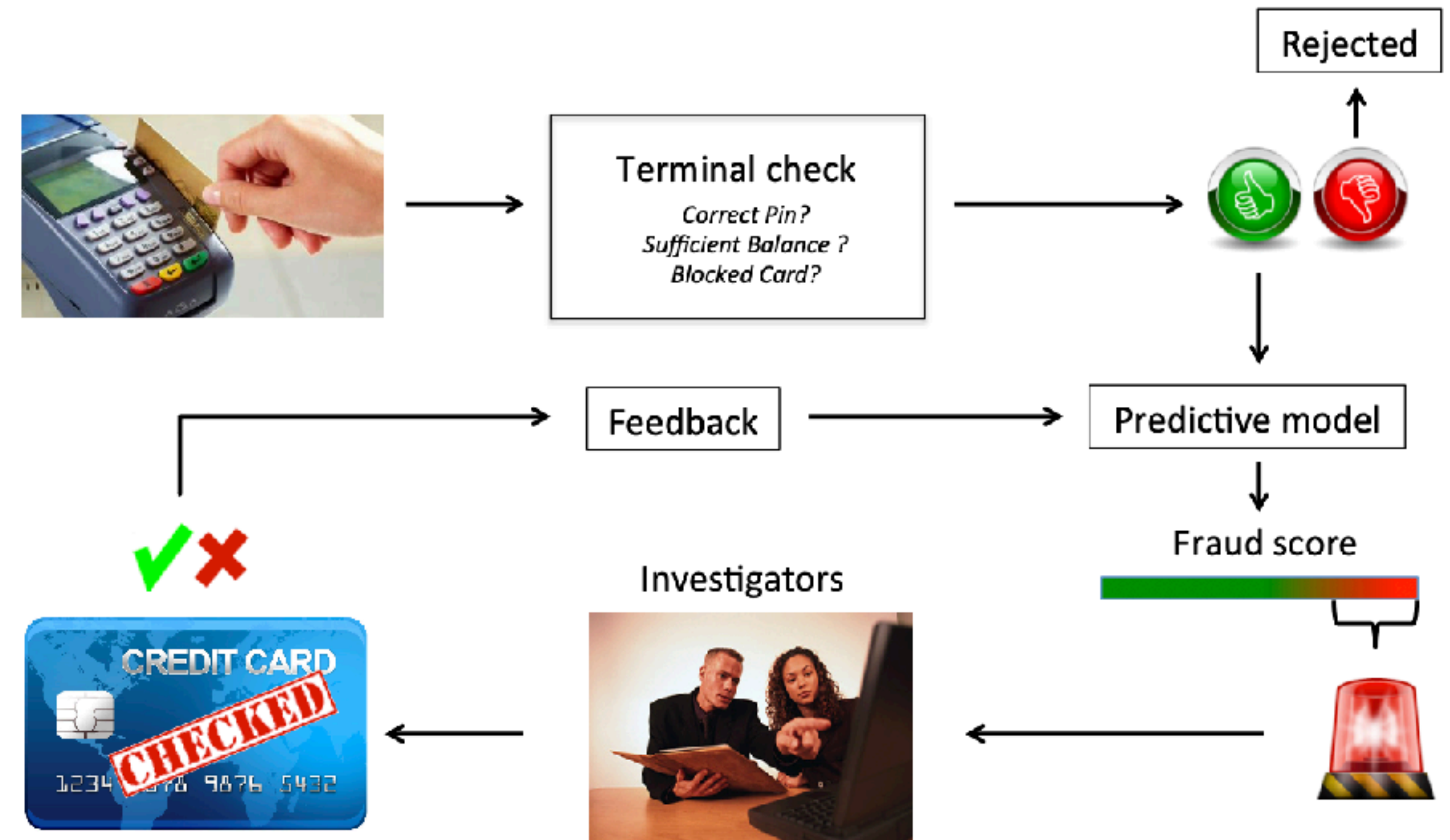


The Credit Card Fraud Detection process.

The dataset:

- A real dataset comprising of financial transactions an the information linked to them is not available as finance-related data is extremely sensitive and protected

- The dataset is artificially generated by creating a virtual world of consumers, merchants and fraudsters and is consistent with relevant real world characteristics such as wealth distribution in the United States
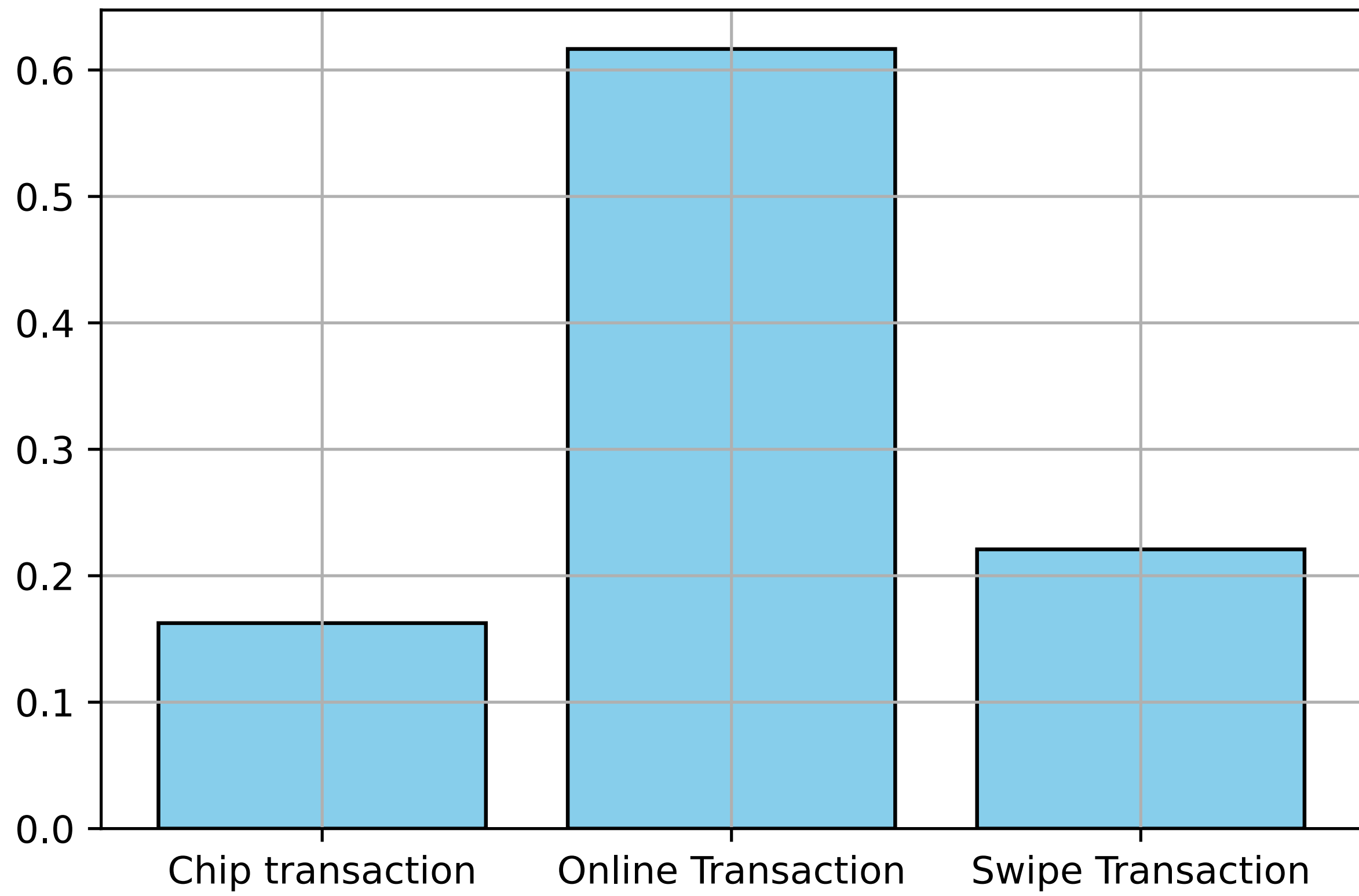
.

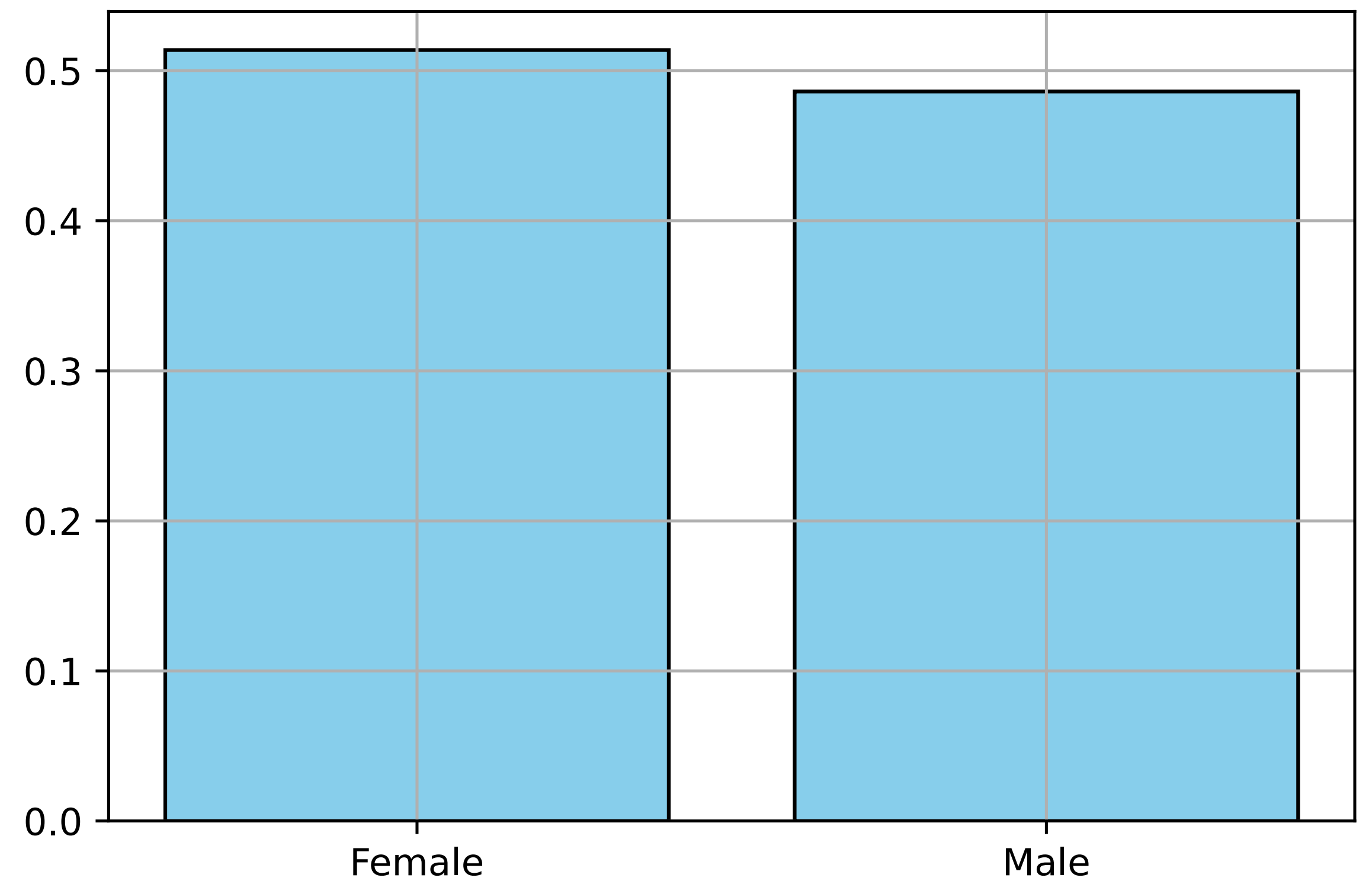Normalized counts of fraudulent transactions with various payment methods
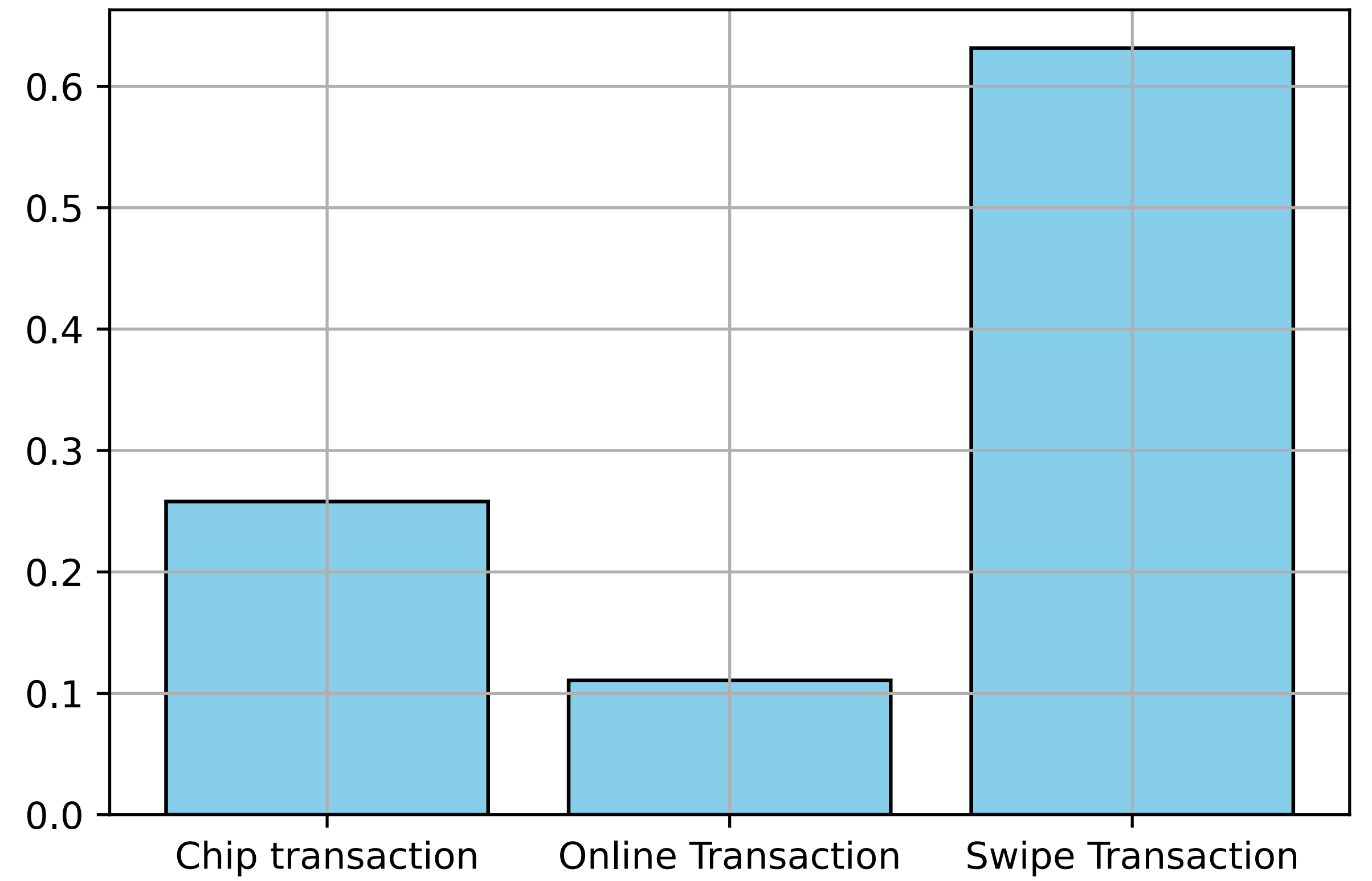
Normalized counts of genuine transactions with various payment methods
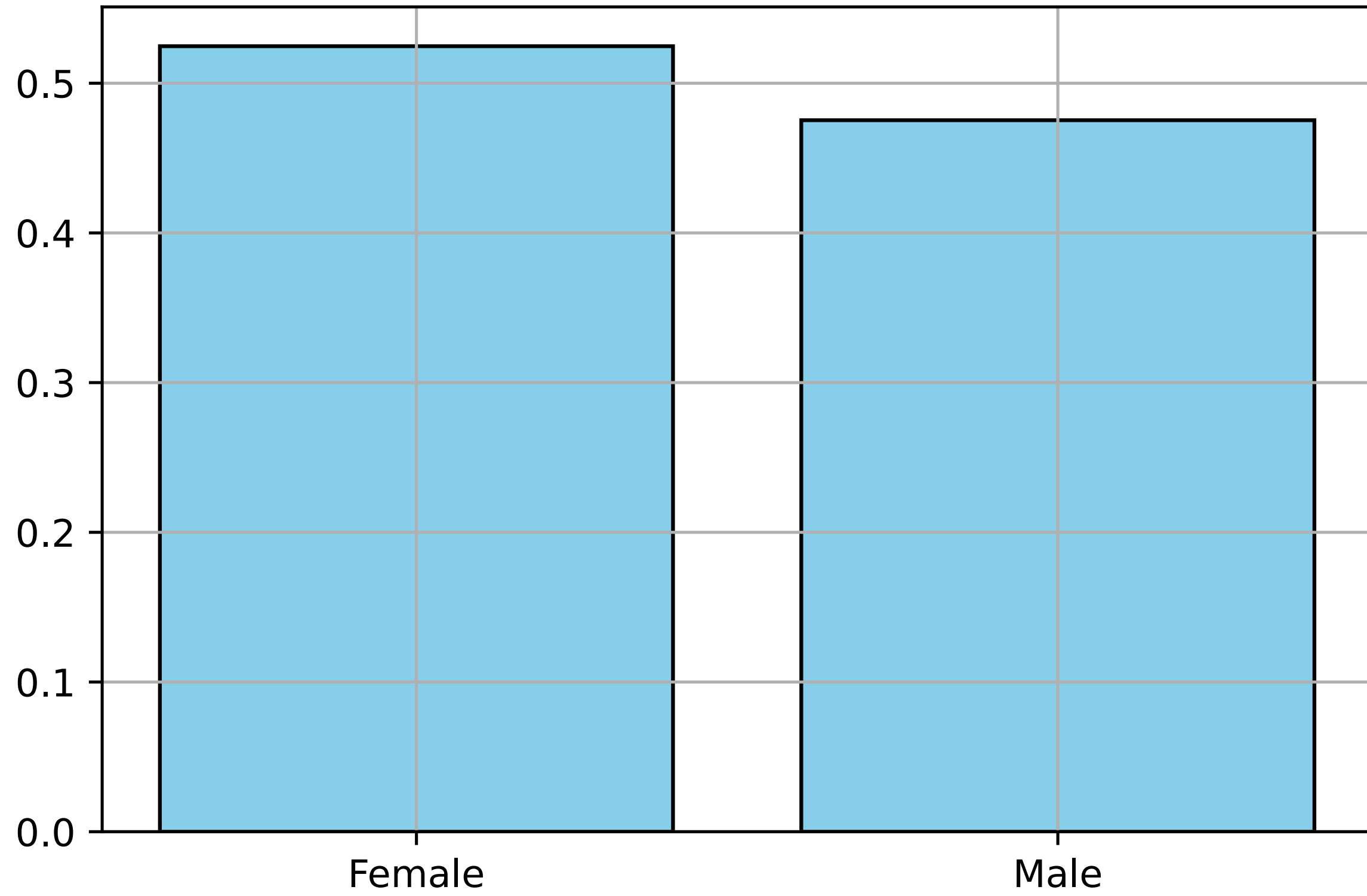
Normalized counts of genuine transactions

with respect to gender

Normalized counts of fraudulent transactions

with respect to gender

Results of SVM

SVM Performance

	Training	Validation	Test
Accuracy	0,85	0,81	0,81
Recall	0,85	0,83	0,82
F1	0,85	0,81	0,80

Results of Random Forest

Random Forest Performance

	Training	Validation	Test
Accuracy	0,86	0,81	0,81
Recall	0,86	0,83	0,83
F1	0,87	0,80	0,81

Results of the neural network:

First attempt performance of the Neural Network

	Training	Validation	Test
Accuracy	0,99	0,99	0,99
Recall	0,01	0,012	0,01
F1	0,002	0,001	0,005

Current Performance of the Neural Network

	Training	Validation	Test
Accuracy	0,98	0,98	0,96
Recall	0,75	0,63	0,81
F1	0,08	0,13	0,05

Results comparison

Comparison of performance on the test data

	SVM	Random Forest	Neural Network
Accuracy	0,81	0,81	0,96
Recall	0,81	0,83	0,81
F1	0,80	0,81	0,05