

---

---

# Fraud Detection

## - Data Exploration -

— **Y. Maruccia**, S. Cavuoti, G. Riccio —  
OAC - 07/10/2024

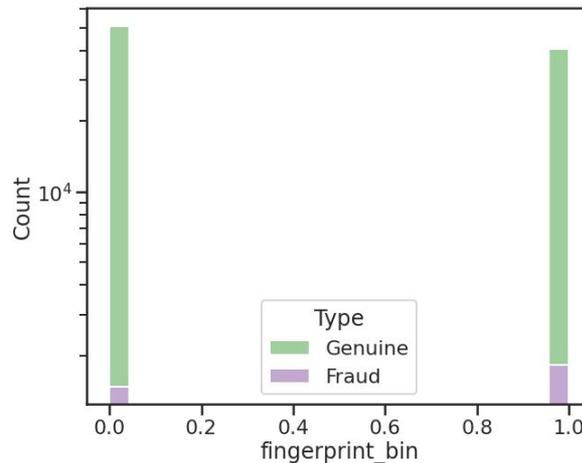
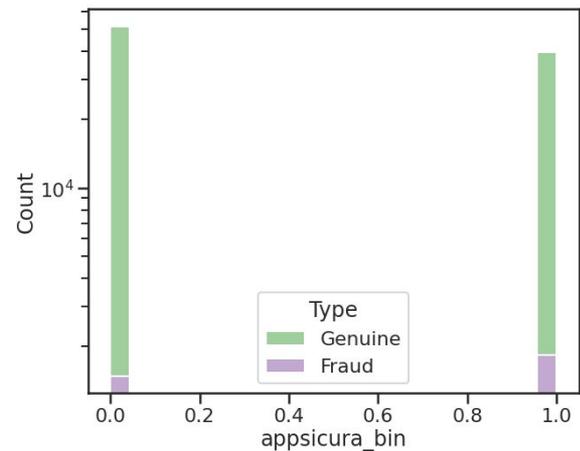
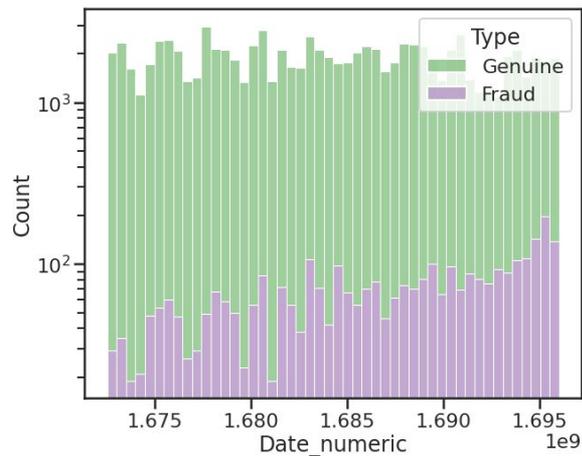
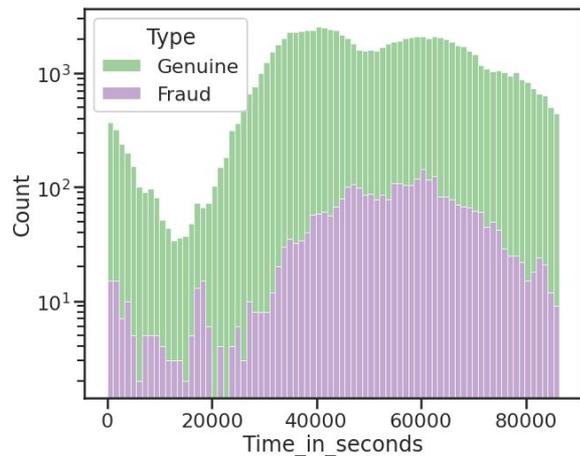
---

---

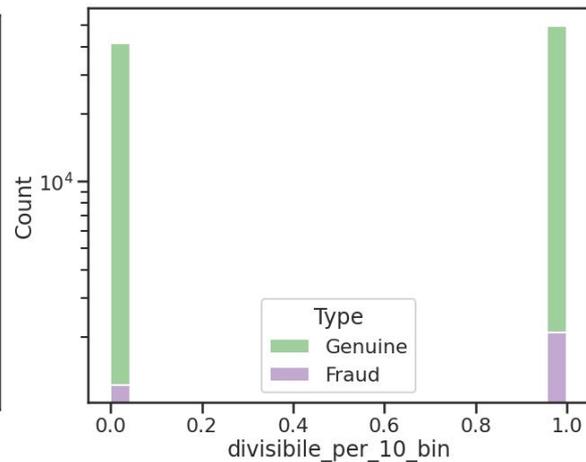
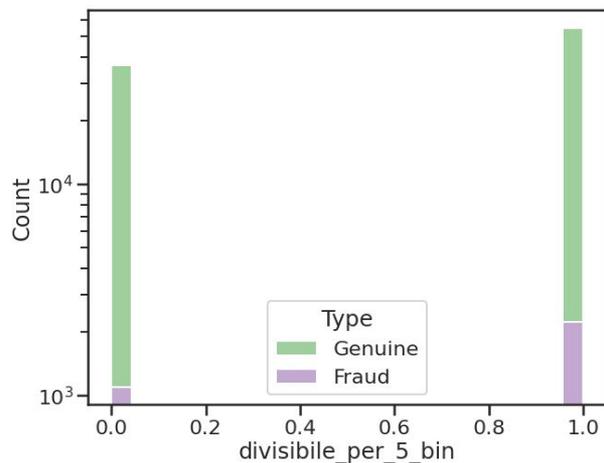
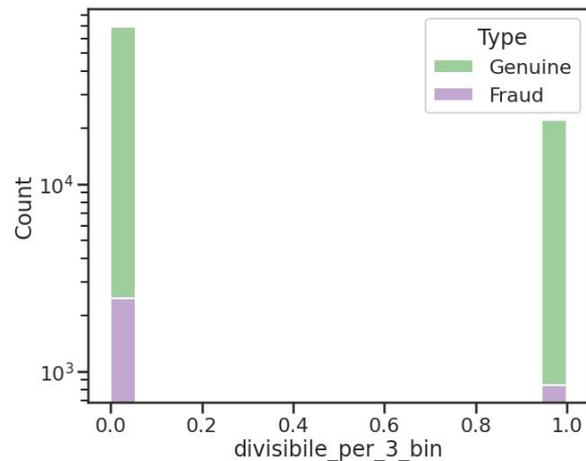
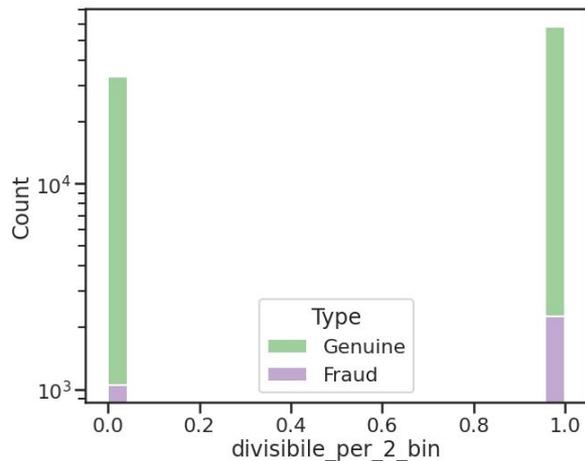
# Overview

- Starting point:
  - cartella *punto\_2*, contenente tutte le transazioni dei prodotti e tutti gli eventi da Gennaio a Settembre 2023 relative a BT che hanno avuto almeno una frode di tipo bonifico.
- Esplorazione delle features
- Esperimento con Random Forest

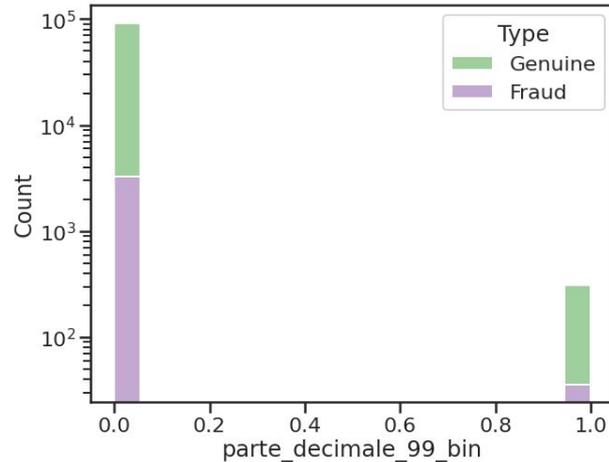
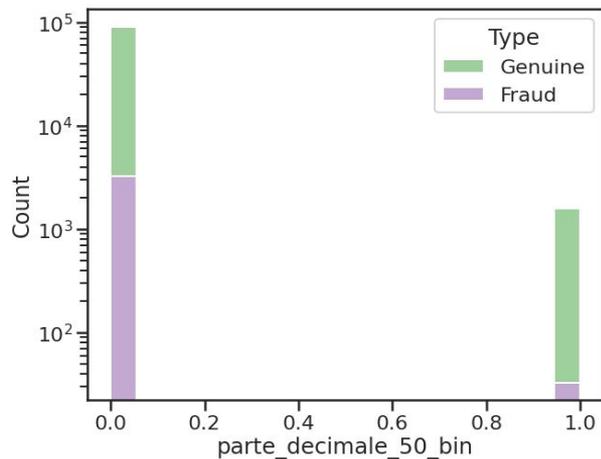
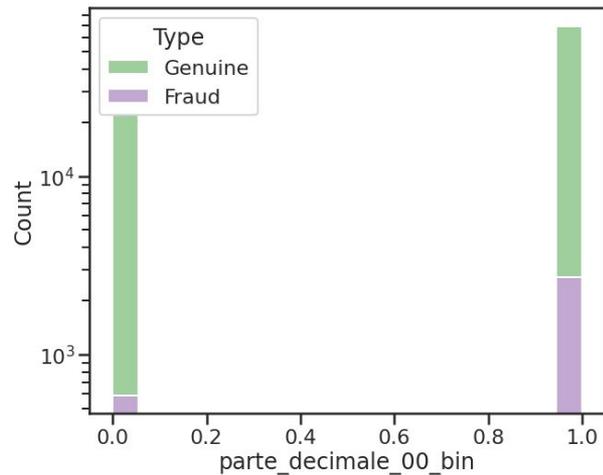
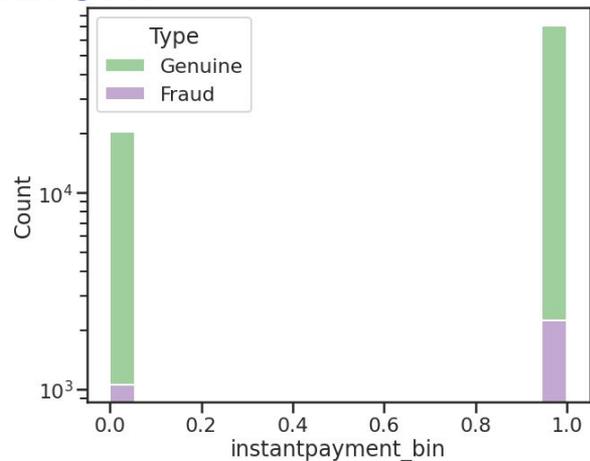
# Distribuzioni



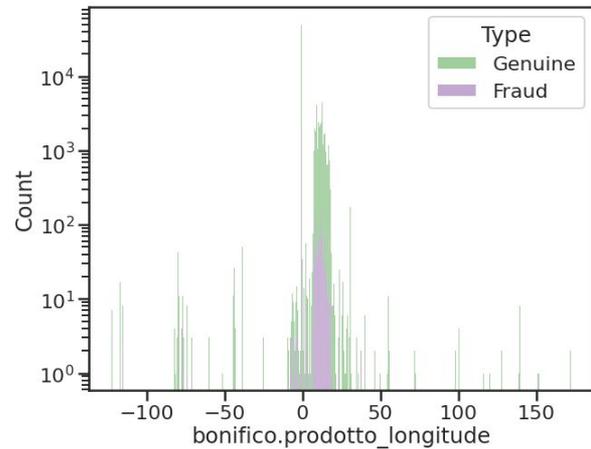
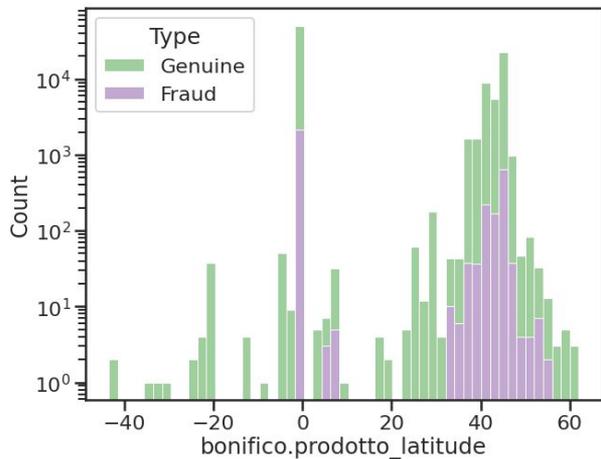
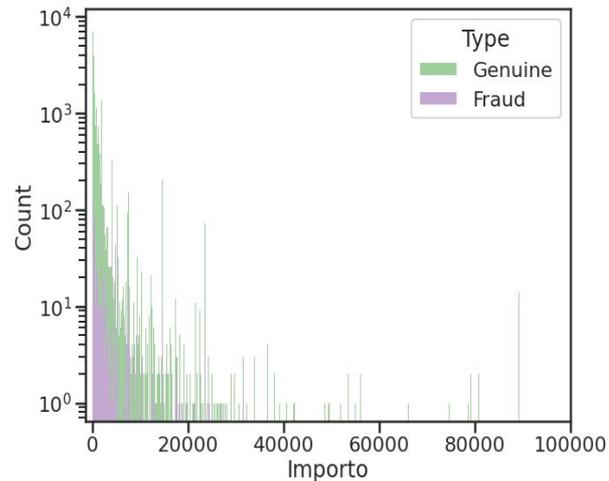
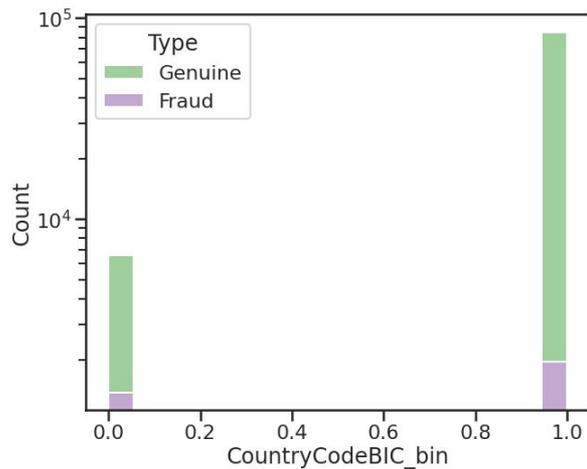
# Distribuzioni



# Distribuzioni



# Distribuzioni



# Random Forest

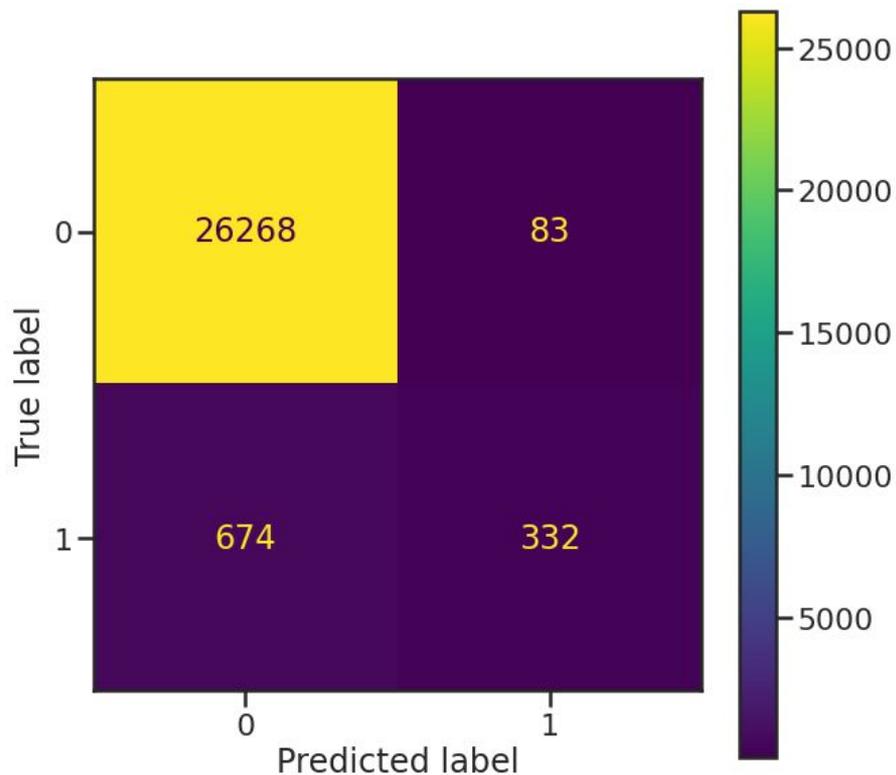
## Features

- Date\_numeric
- Time\_in\_seconds
- appsicura\_bin
- fingerprint\_bin
- bonifico.prodotto\_latitude
- bonifico.prodotto\_longitude
- CountryCodeBIC\_bin
- instantpayment\_bin
- bonifico.prodotto\_importooperazione
- divisibile\_per\_2\_bin
- divisibile\_per\_3\_bin
- divisibile\_per\_5\_bin
- divisibile\_per\_10\_bin
- parte\_decimale\_00\_bin
- parte\_decimale\_50\_bin
- parte\_decimale\_99\_bin

## Target

- bonifico.last\_status  
(**0**→ Genuine, **1**→ Fraud)

# Random Forest - 1

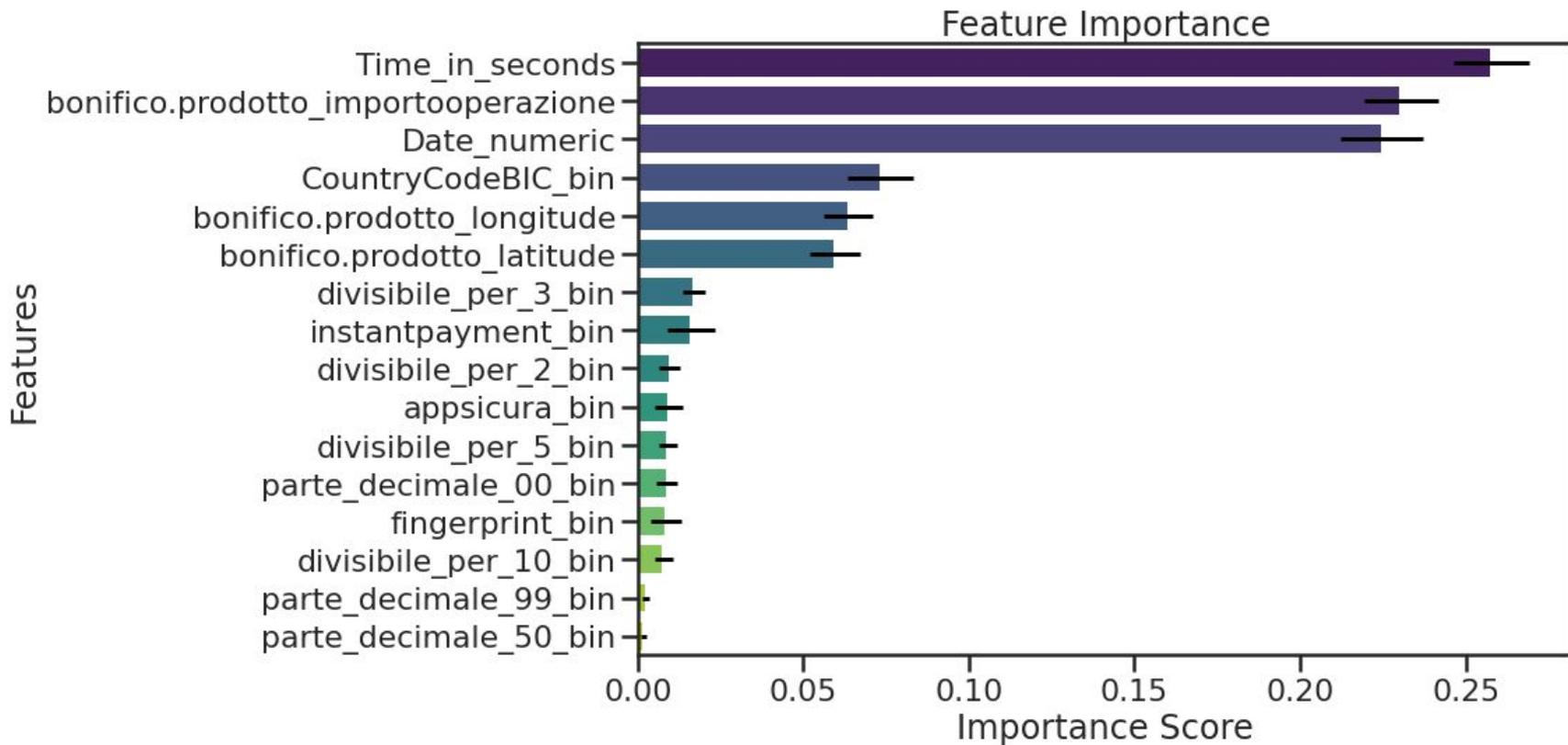


Class weight: None

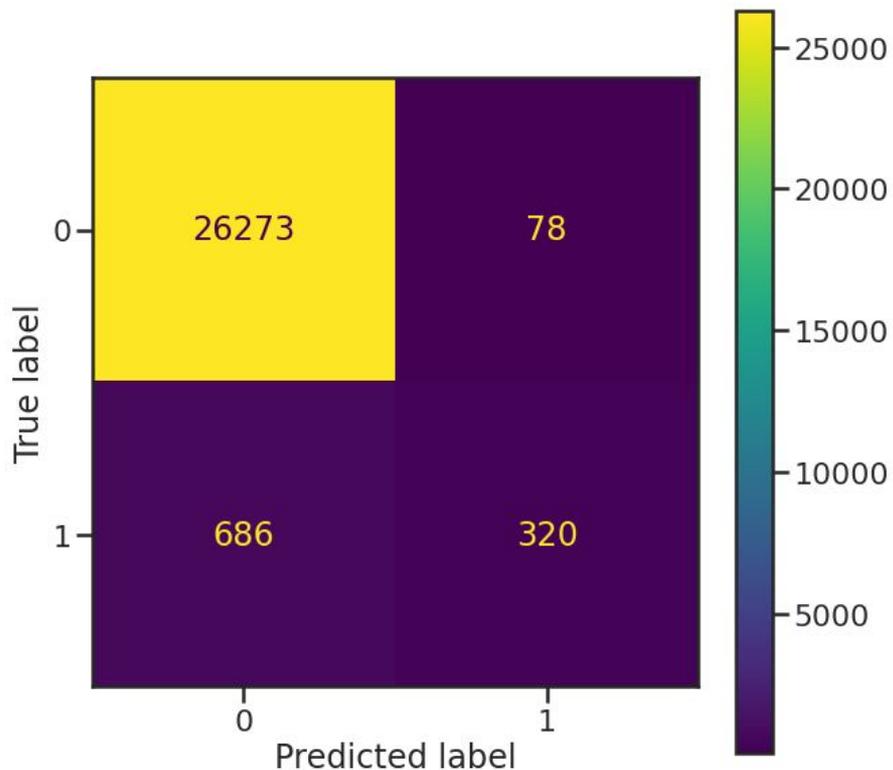
All classes are supposed to have weight one

	precision	recall	f1-score	support
0	0.97	1.00	0.99	26351
1	0.80	0.33	0.47	1006
accuracy			0.97	27357
macro avg	0.89	0.66	0.73	27357
weighted avg	0.97	0.97	0.97	27357

# Random Forest - 1



# Random Forest - 2

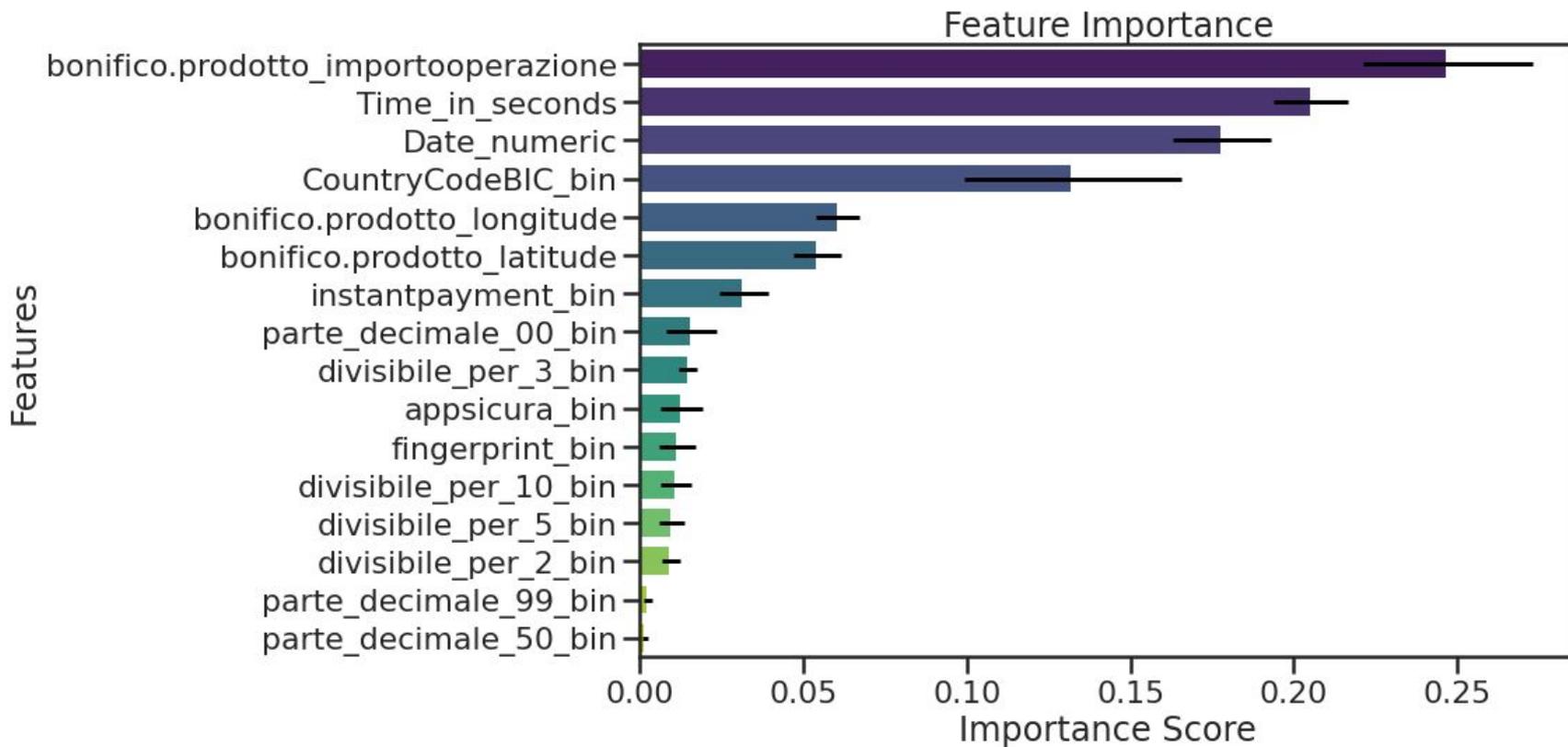


## Class weight: balanced

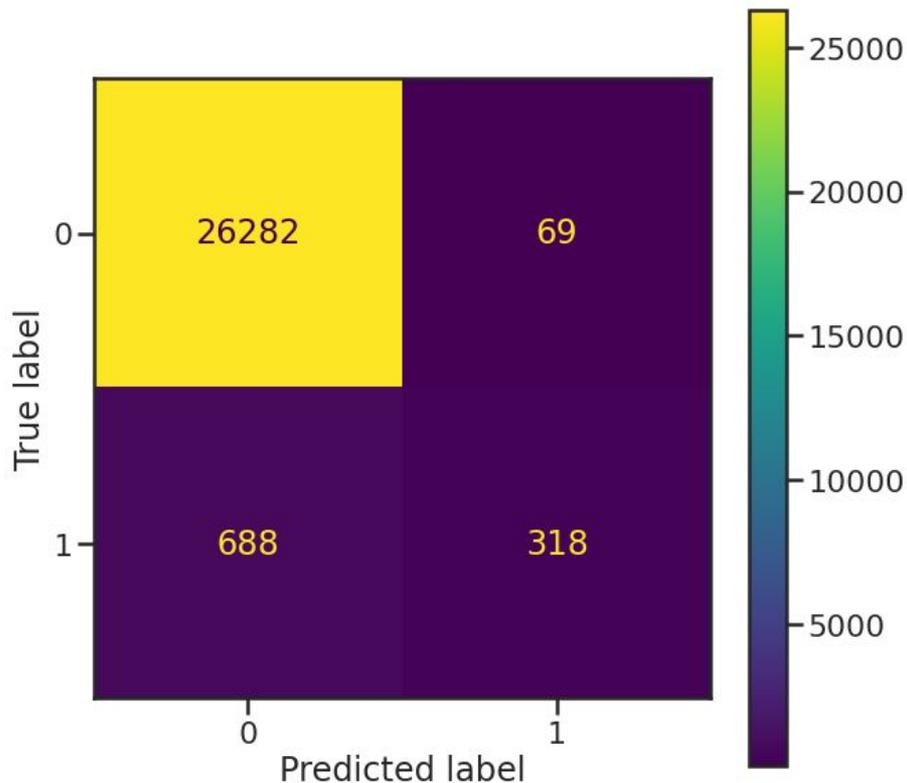
The “balanced” mode uses the values of  $y$  to automatically adjust weights inversely proportional to class frequencies in the input data as  $n\_samples / (n\_classes * np.bincount(y))$

	precision	recall	f1-score	support
0	0.97	1.00	0.99	26351
1	0.80	0.32	0.46	1006
accuracy			0.97	27357
macro avg	0.89	0.66	0.72	27357
weighted avg	0.97	0.97	0.97	27357

# Random Forest - 2



# Random Forest - 3

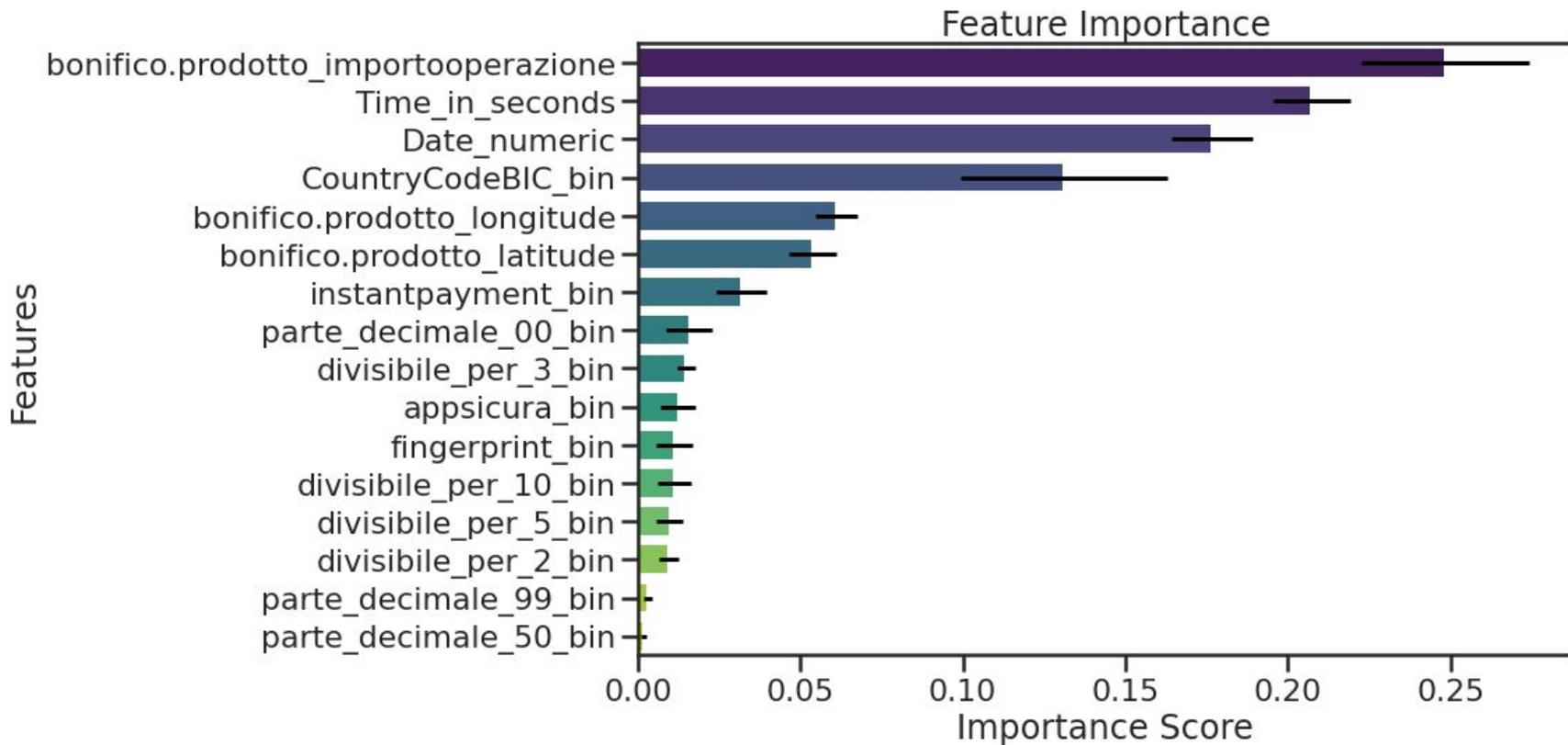


## Class weight: `balanced_subsample`

The “`balanced_subsample`” mode is the same as “`balanced`” except that weights are computed based on the bootstrap sample for every tree grown.

	precision	recall	f1-score	support
0	0.97	1.00	0.99	26351
1	0.82	0.32	0.46	1006
accuracy			0.97	27357
macro avg	0.90	0.66	0.72	27357
weighted avg	0.97	0.97	0.97	27357

# Random Forest - 3



# Preliminary results and future works

- Utilizzando uno qualsiasi dei tre metodi, è possibile recuperare fino al 33% delle transazioni fraudolente, con un incremento di appena lo 0,3% nei falsi positivi.
- In questa fase preliminare, alcune features non sono ancora state utilizzate, mentre altre sono state categorizzate.
  - Per esempio, *CountryCodeBIC* è 1 se assume il valore più frequente, 0 altrimenti.
- Step futuri:
  - Estensione dell'insieme di features utilizzate e del dataset.
  - Adozione di approcci alternativi.
  - Inserimento di features di profilo dell'utente.