Università degli
Studi di Salerno

CCF Detector

Using Machine Learning against Credit Card Frauds

Marco Santoriello
Mat. 0512114100

Corso di Fondamenti di Intelligenza Artificiale
A.A. 23/24

INTRODUZIONE

Negli ultimi anni, il mondo delle transazioni finanziarie è stato, ed è tuttora, interessato da una crescita nell’utilizzo dei pagamenti elettronici.

- **COMODITA’**
- **AFFIDABILITA’**
- **TRACCIABILITA’**

LA COSTANTE MINACCIA DEI TRUFFATORI

SECONDO LA FEDERAL TRADE COMMISSION, IL TIPO PIÙ DIFFUSO DI FURTO DI IDENTITÀ NEGLI USA, RIGUARDA LE CARTE DI CREDITO

Appropriazione fisica	**Appropriazione dei dati**
I criminali rubano fisicamente la carta di credito oppure utilizzando dispositivi contactless per rubare denaro alle vittime.	I criminali entrano in possesso dei dati relativi ad una carta di credito, in modo da poter effettuare ed autorizzare transazioni fraudolente.
• Appropriazione illecita • Clonazione • Utilizzo di dispositivi POS	• Invio di particolari messaggi • Induzione della vittima a fornire i dati • Effettuazione di chiamate alla vittima

IL PROGETTO

01. Realizzazione di un modello di Machine Learning per l'individuazione di transazioni fraudolente

02. Studio delle features più rilevanti per le transazioni fraudolente

03. Esplorazione dei dati per rilevare pattern nascosti alla base del modo di agire dei truffatori

IL DATASET

RICERCA:

DATI PARTICOLARMENTE SENSIBILI:

- SCARSA DISPONIBILITA’ DI DATASET REALI
- DATASET CON FEATURES NASCOSTE PER PRIVACY

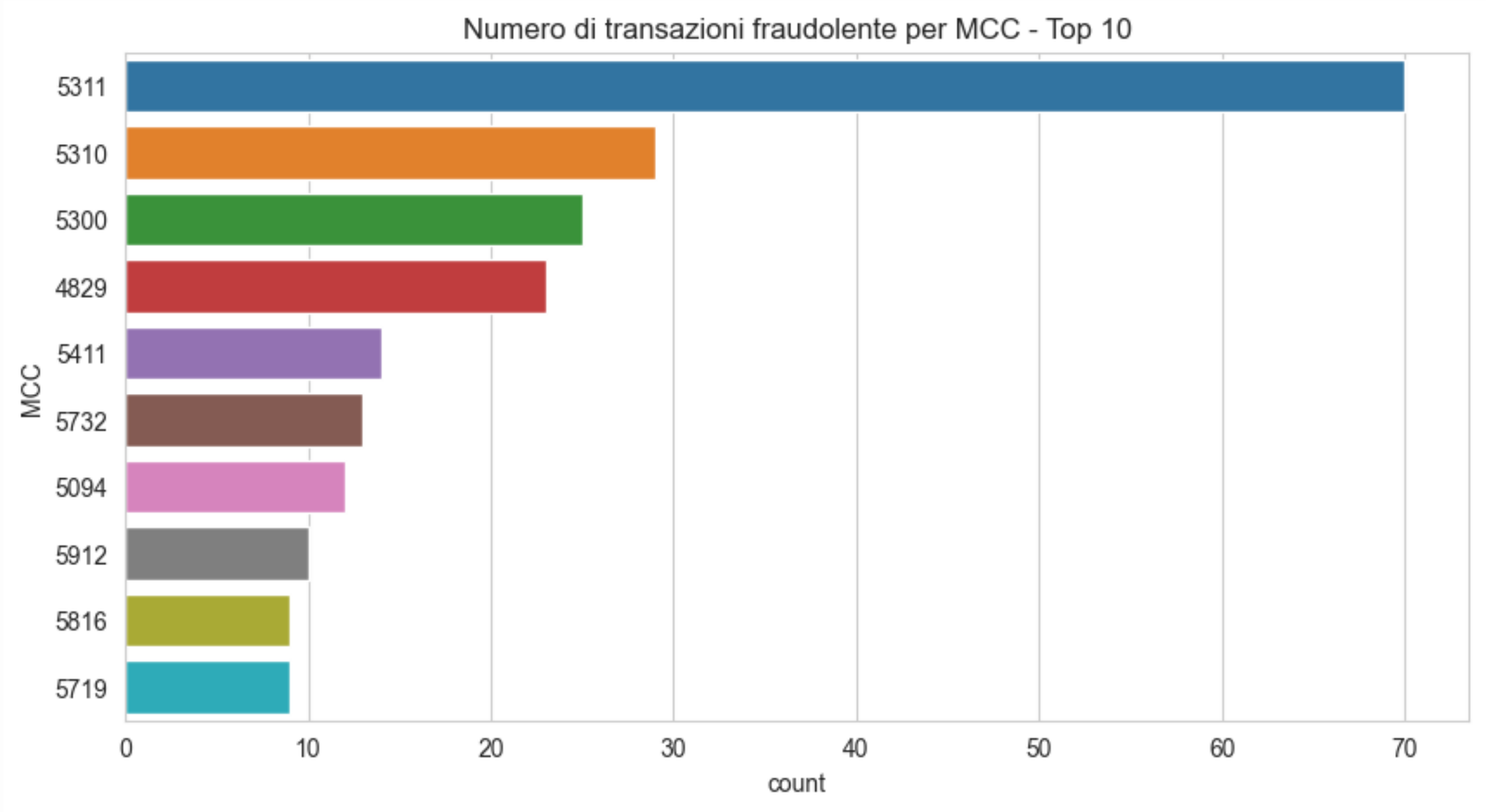
SOLUZIONE: DATASET SIMULATI

IL DATASET INDIVIDUATO:

- OLTRE 24 MILIONI DI TRANSAZIONI
- DATI SIMULATI
- FEATURES IN CHIARO E DESCRITTE

ESPLORAZIONE DEI DATI

MCC - MERCHANT CATEGORY CODE

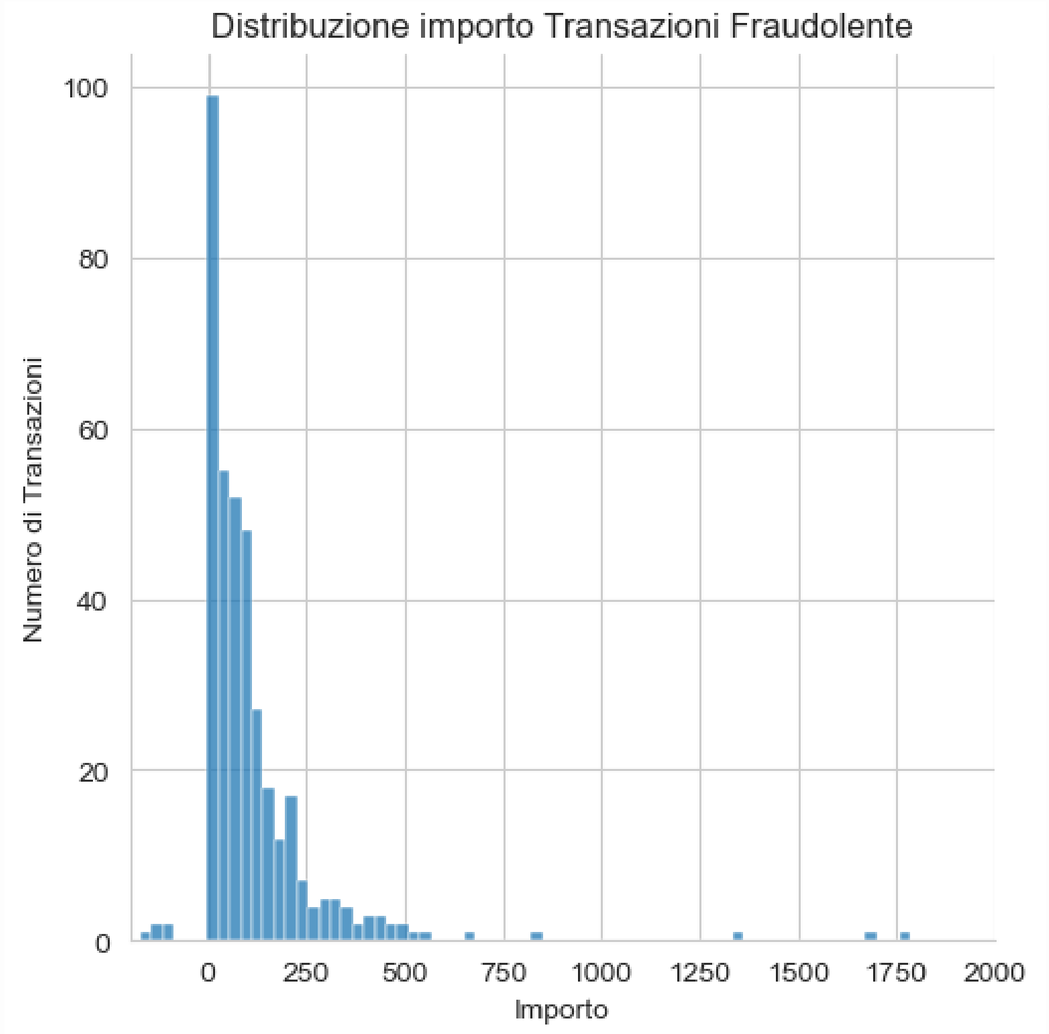


Categorie maggiormente colpite:

- **Grandi magazzini**
- **Discounts**
- **Ingrosso**

ESPLORAZIONE DEI DATI

IMPORTI TRANSAZIONI FRAUDOLENTE

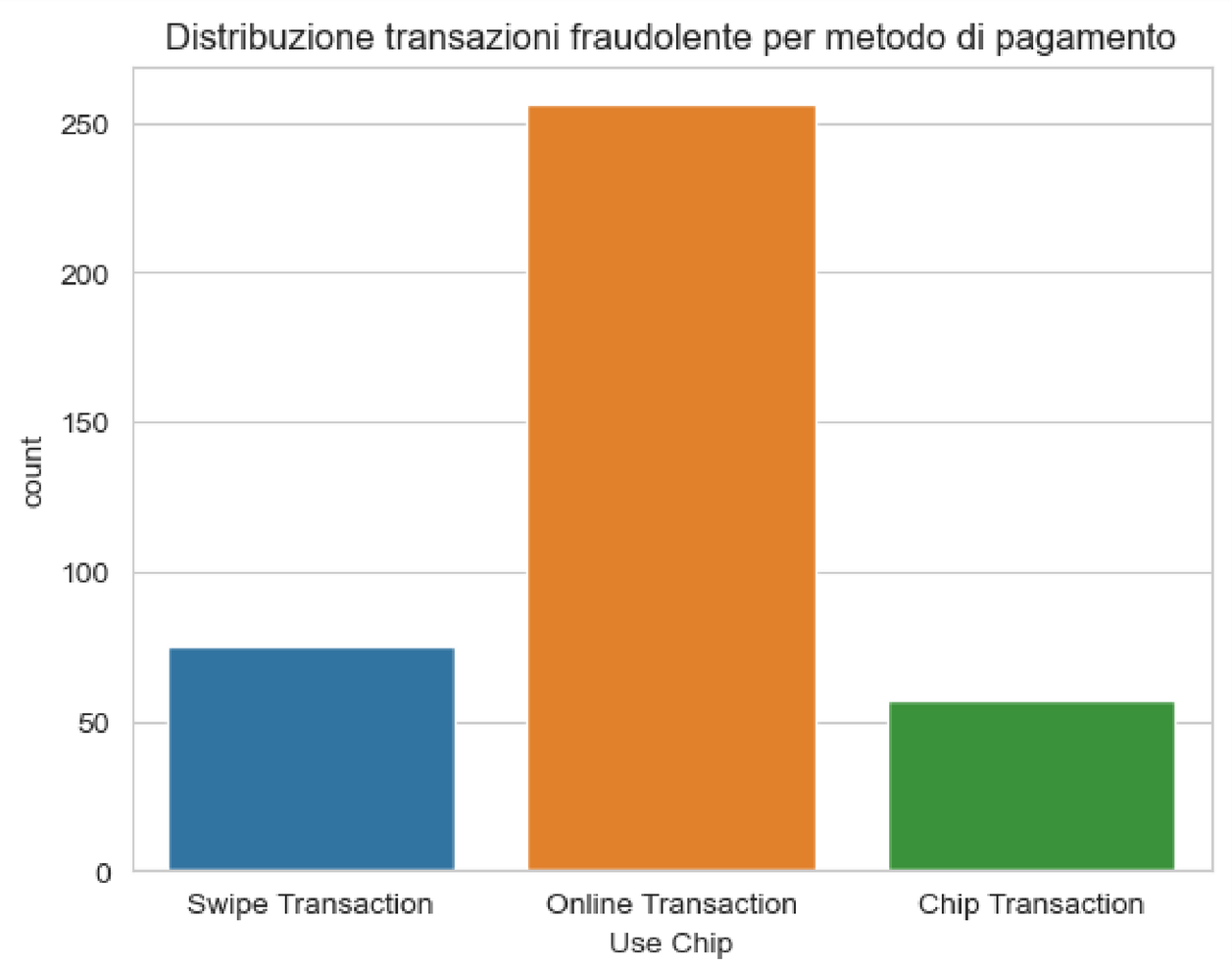


Prevalenza di frodi con importi bassi:

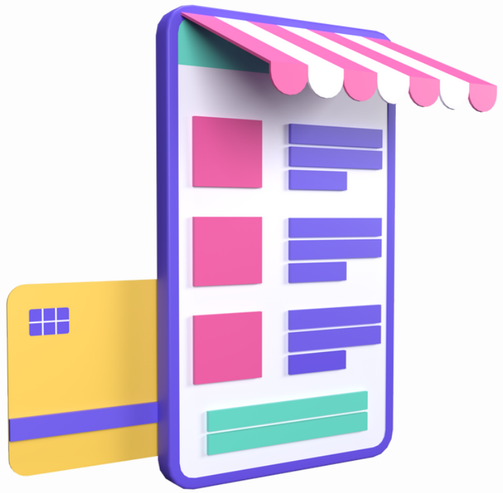
- **da $0 a $250**

ESPLORAZIONE DEI DATI

METODI DI PAGAMENTO

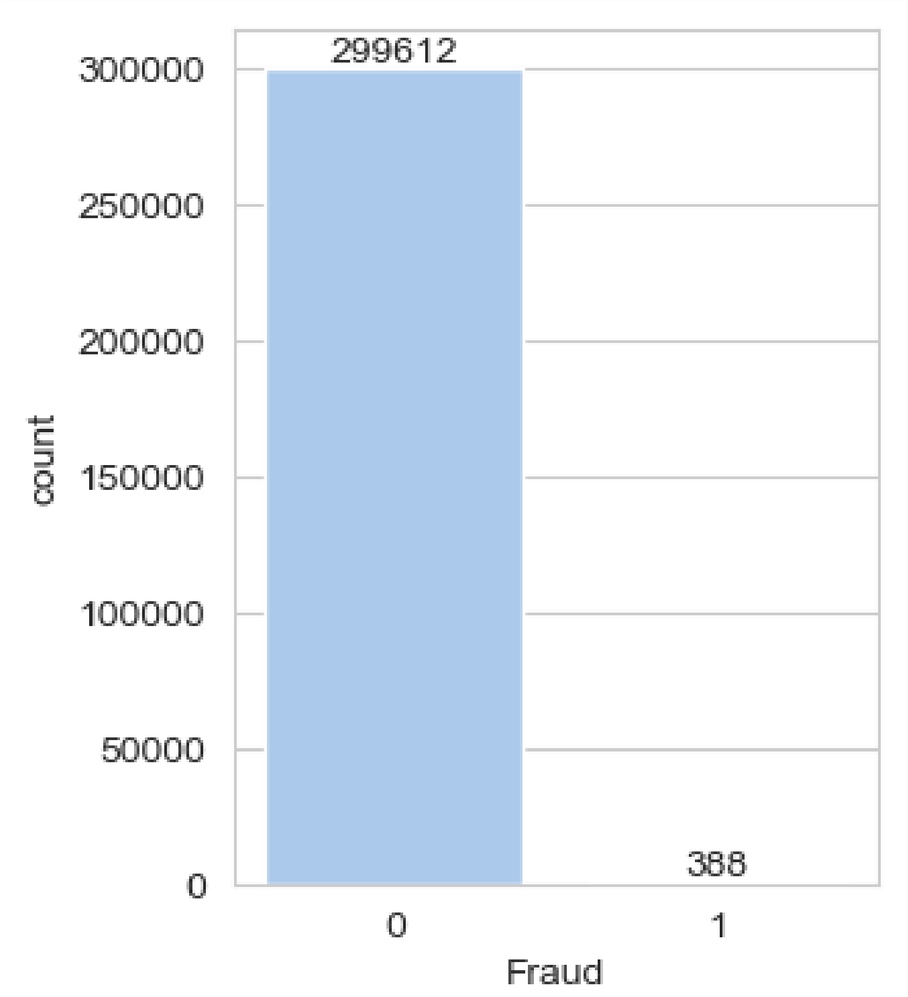


La maggior parte delle frodi avviene con pagamenti online



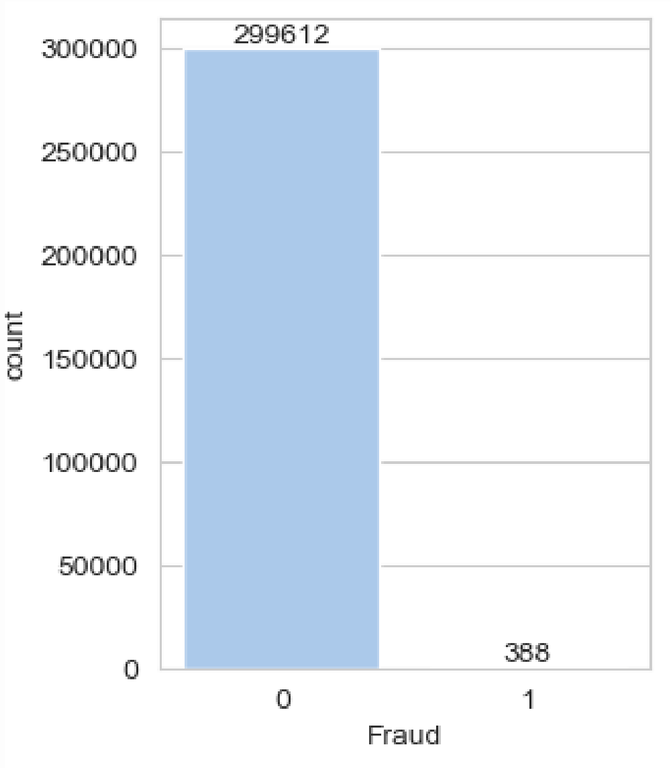
ESPLORAZIONE DEI DATI

PROBLEMA: DATASET ESTREMAMANTE SBILANCIATO

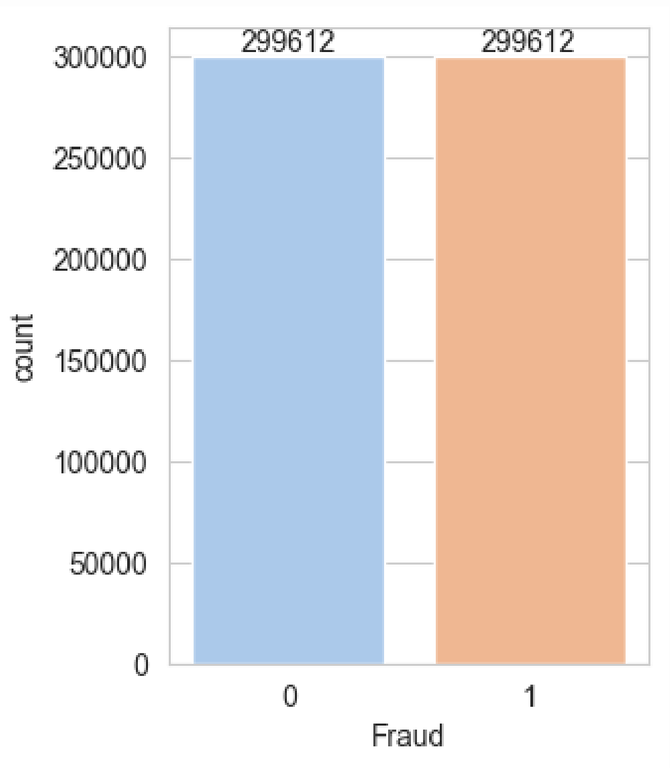


ESPLORAZIONE DEI DATI

SOLUZIONE

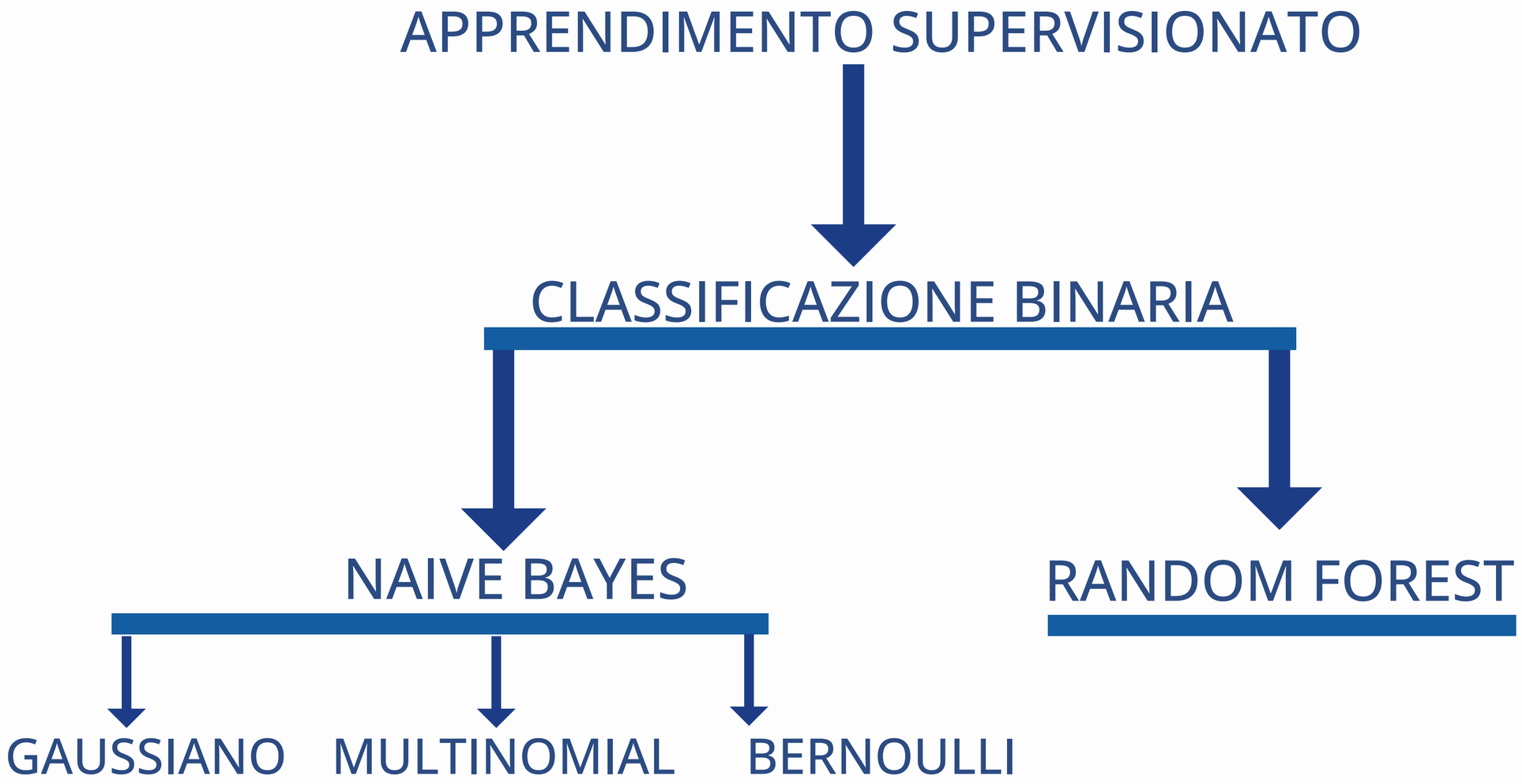


Prima di SMOTE



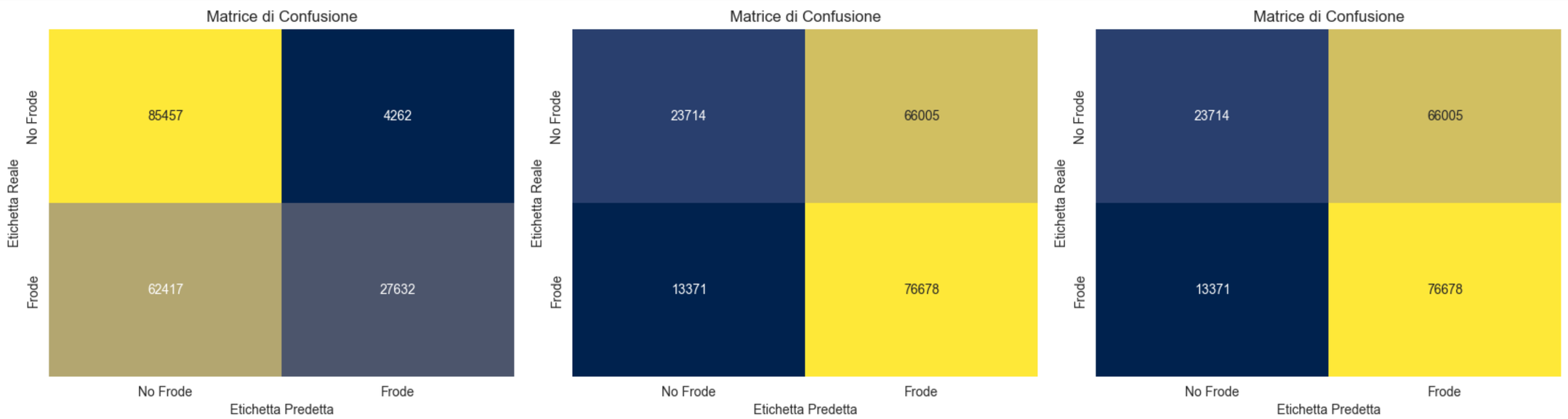
Dopo SMOTE

SCELTA DELL'ALGORITMO



CONFRONTO MODELLI

VARIANTI NAIVE BAYES



GAUSSIAN:

- ACCURACY MEDIOCRE (0.6)
- ALTA PRECISION (0.86)
- BASSO RECALL (0.3)

BERNOULLI:

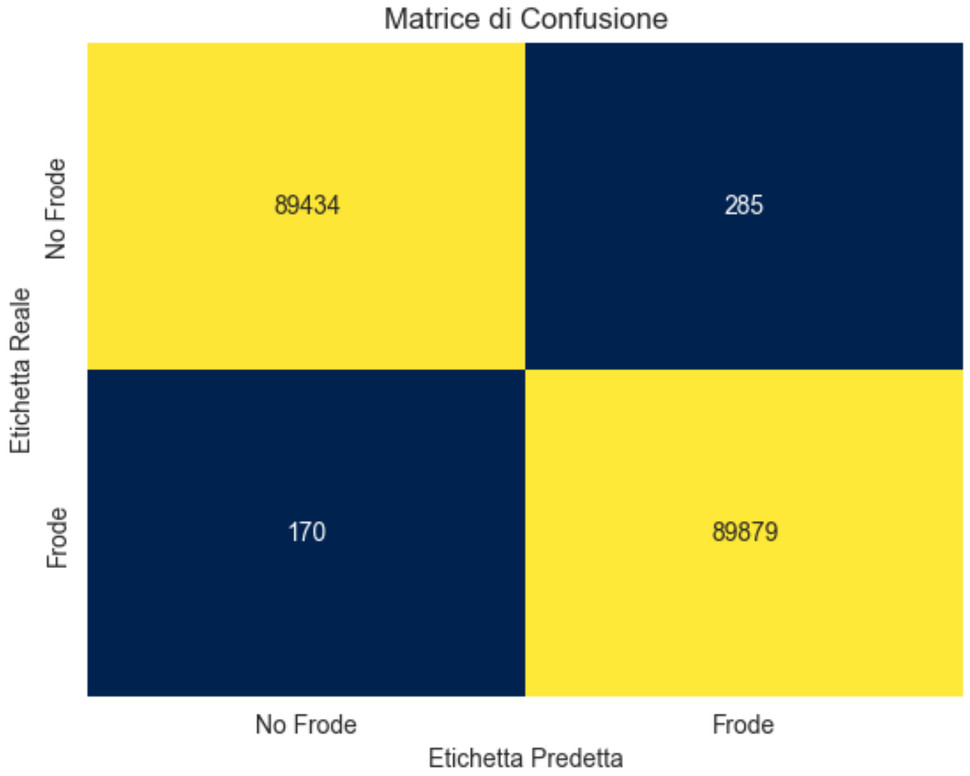
- BASSA ACCURACY (0.5)
- BASSA PRECISION (0.5)
- ALTO RECALL (0.8)

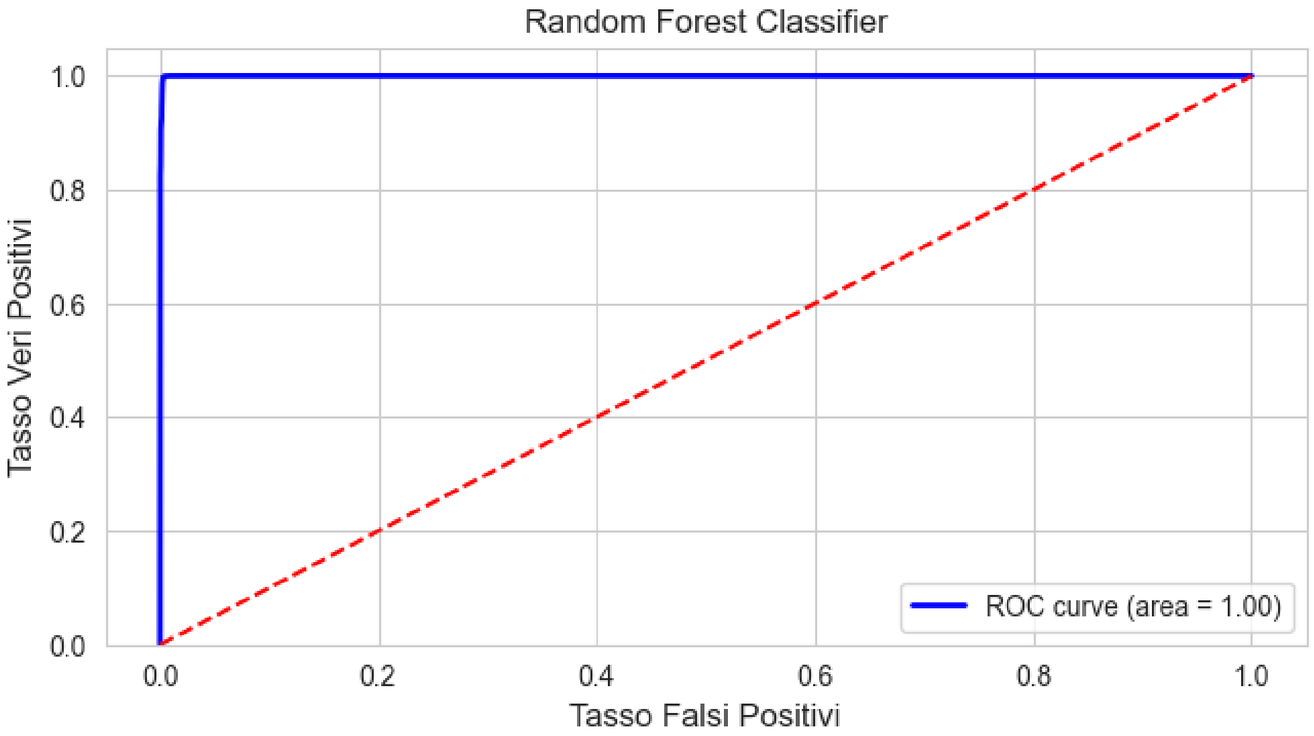
MULTINOMIAL:

- BASSA ACCURACY (0.5)
- BASSA PRECISION (0.5)
- BASSO RECALL (0.8)

CONFRONTO MODELLI

RANDOM FOREST



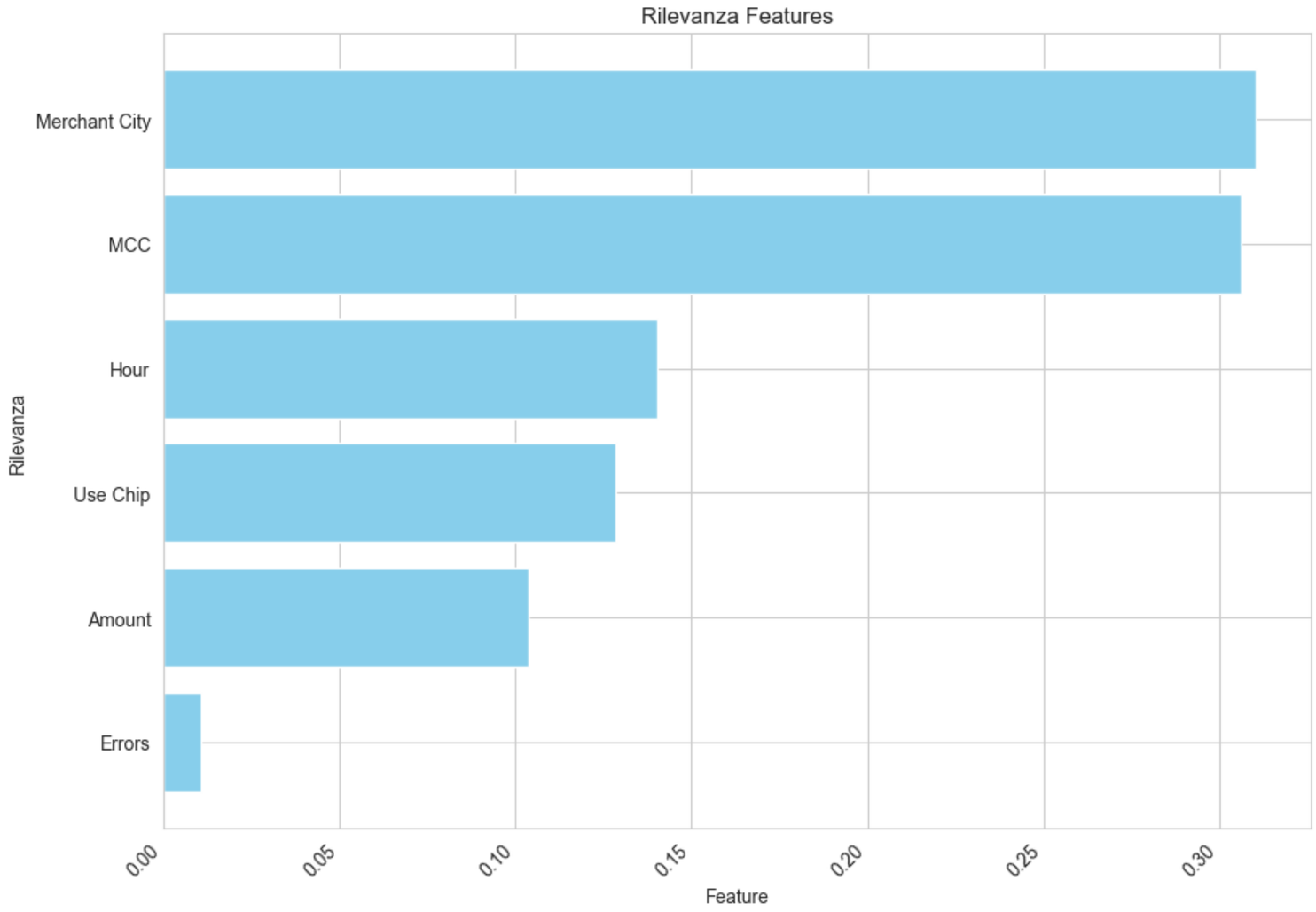


- ALTA ACCURACY(0.99)
- ALTA PRECISION (0.99)
- ALTO RECALL (0.99)

- OTTIMO PUNTEGGIO AUC: CLASSIFICATORE DISTINGUE NETTAMENTE TRANSAZIONI FRAUDOLENTE DA TRANSAZIONI LECITE

CARATTERISTICHE RILEVANTI

1. **POSIZIONE DEL VENDITORE (ONLINE COMPRESO)**
2. **MCC**
3. **ORARIO DI EFFETTUAZIONE**
4. **MODALITA’ DI PAGAMENTO**
5. **IMPORTO TRANSAZIONE**



CONCLUSIONI

- RISULTATI DEL MODELLO SODDISFACENTI
- DA INTEGRARE IN SISTEMA DI CONTROLLO REAL-TIME
- MONITORARE MAGGIORMENTE DETERMINATE TIPOLOGIE DI BUSINESS
- FORNIRE AUTORIZZAZIONE ANCHE PER TRANSAZIONICON IMPORTI BASSI
- SENSIBILIZZARE GLI UTENTI ALLA VERIFICA DI PAGINE DI ACQUISTO, SMS DI ALLERTA…
- STRUIRLI AI RISCHI LEGATI AI PAGAMENTI ELETTRONICI

FINE

GRAZIE PER L’ATTENZIONE!