

# TrafficML



UNIVERSITÀ  
DEGLI STUDI DI BARI  
ALDO MORO

---

RILEVAMENTO DI  
ATTACCHI SULLA RETE  
CON MACHINE LEARNING

A cura di: Mongelli Antonio

Docenti: Casalino Gabriella  
Zaza Gianluca



# Introduzione

Obiettivo: rilevare automaticamente attacchi informatici tramite ML.

Strumenti utilizzati:

- Dataset principale: CIC-IDS2017.
- Strumento di raccolta traffico: Suricata (IDS).
- Piattaforma da monitorare: T-pot (Honeypot).



# Contesto

- Necessità di automatizzare la rilevazione di minacce.
- Superare i limiti dei sistemi basati su firme.
- Applicabilità anche su traffico cifrato o variabile.



# Obiettivi del Sistema

- Rilevamento in tempo reale.
- Spiegabilità delle predizioni (SHAP).
- Automazione delle risposte (logging, allarmi, blocchi IP).

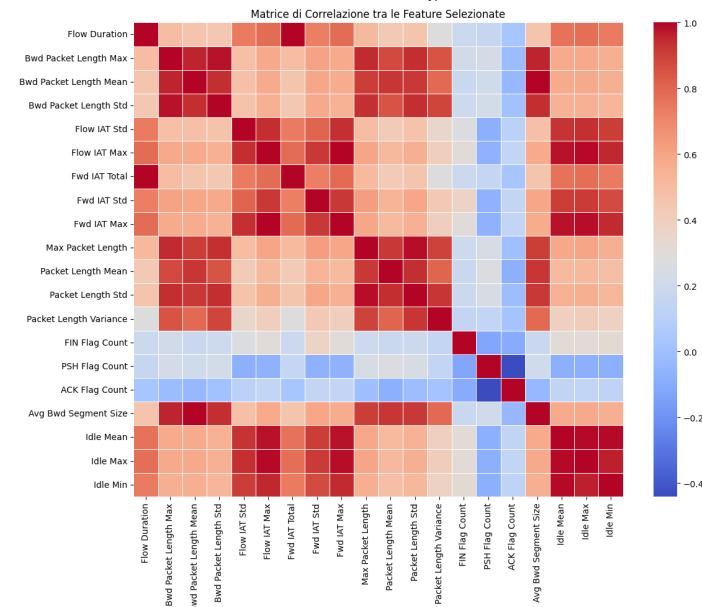
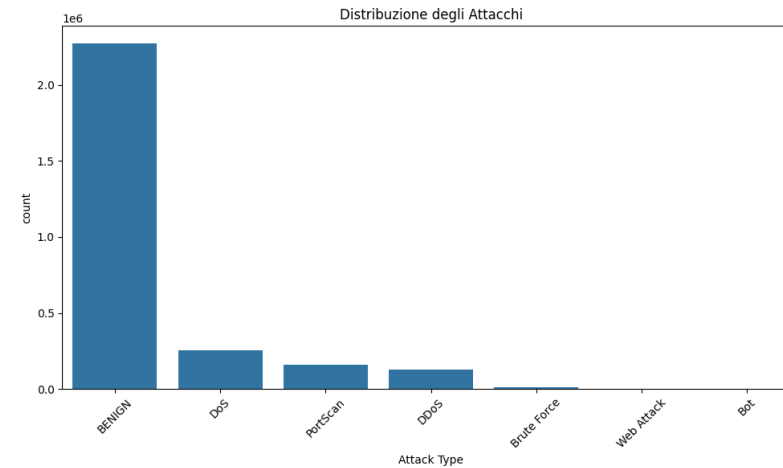


# Scelta dell'Algoritmo

- Approccio supervisionato.
- Classificazione vs clustering.
- Vantaggi: precisione, spiegabilità, integrazione operativa.

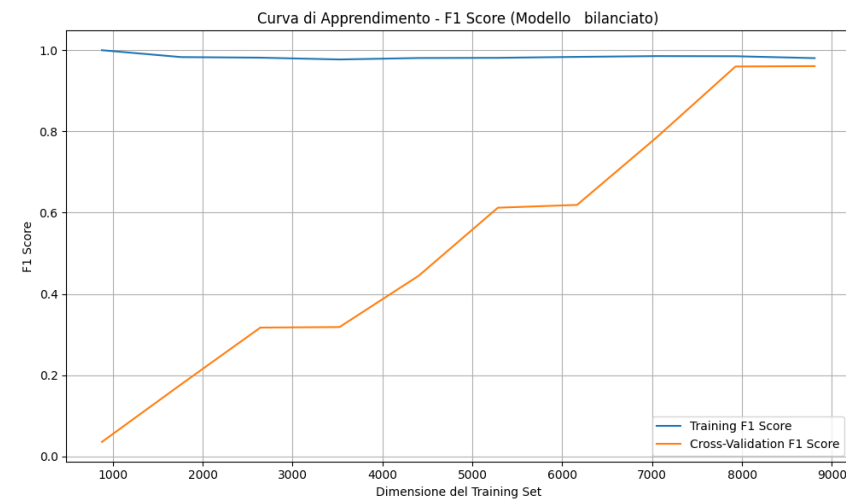
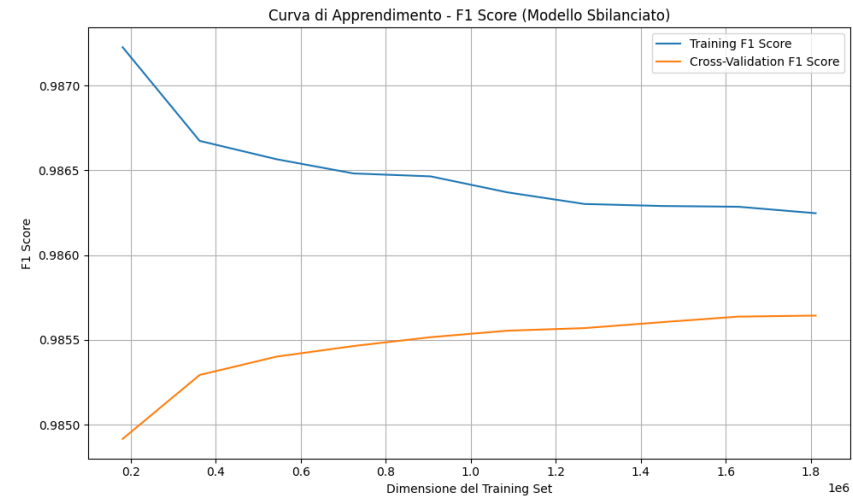
# Data Preparation

- Preparazione del dataset
- Pulizia: rimozione duplicati, normalizzazione etichette.
- Feature engineering: encoding, imputazione, normalizzazione, selezione feature.

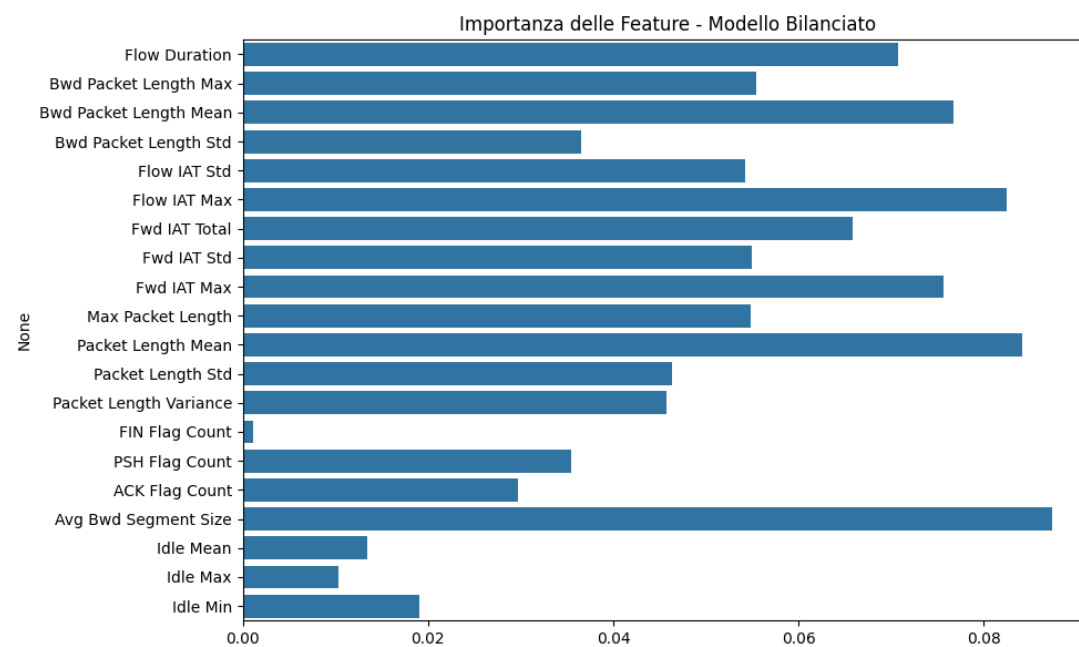
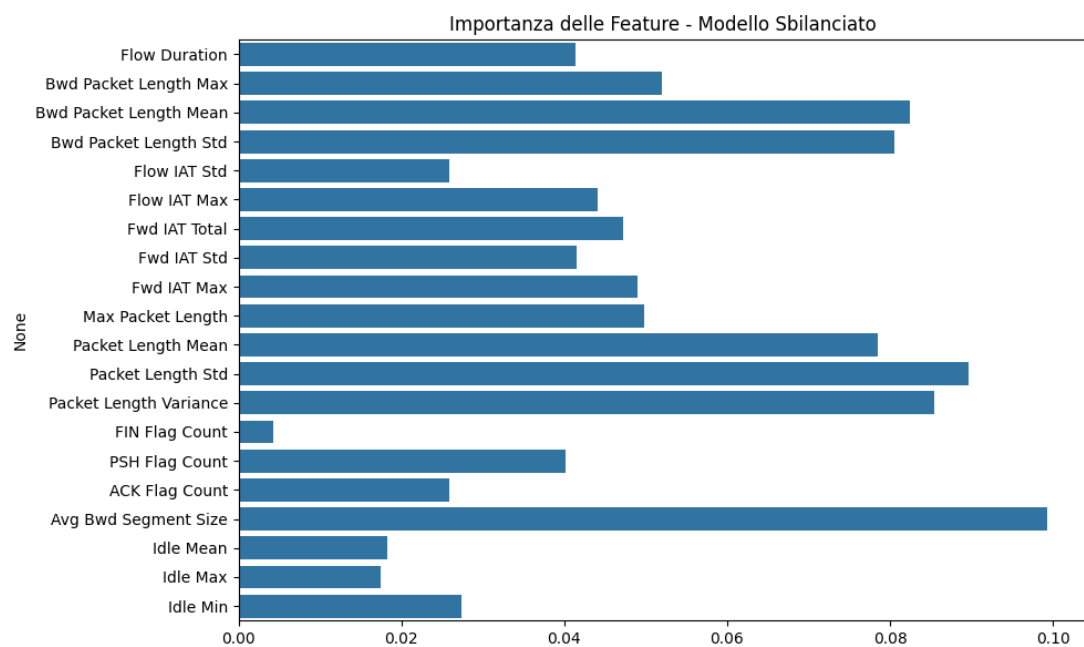


# Addestramento del modello

- Bilanciamento con RandomUnderSampler.
- Divisione dei dati: 20% test set
- Algoritmo: Random Forest (100 alberi, max depth 20).
- Addestramento su dati originali e bilanciati.



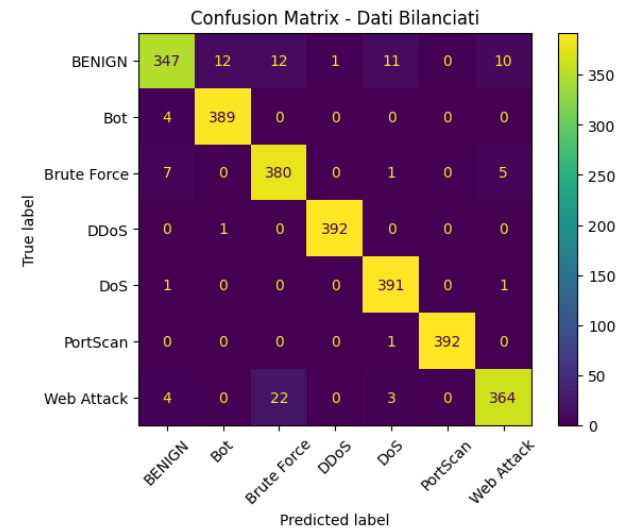
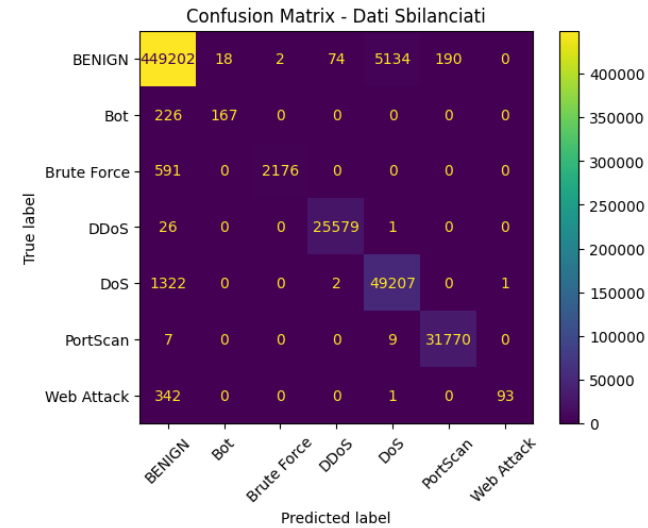
# Importanza delle Features





# Testing e Metriche

- Due strategie: test su set sbilanciato e bilanciato.
- Metriche: Accuracy, Precision, Recall, F1 Score.
- Matrice di confusione per analisi dettagliata.



# Testing e Metriche

## TEST SU TEST-SET SBILANCIATO

--- Test Set (Sbilanciato) ---

Accuracy: 0.9859646023951673

F1 Score: 0.9858148822759902

Precision: 0.9865337275516963

Recall: 0.9859646023951673

	precision	recall	f1-score	support
0	0.99	0.99	0.99	454620
1	0.90	0.42	0.58	393
2	1.00	0.79	0.88	2767
3	1.00	1.00	1.00	25606
4	0.91	0.97	0.94	50532
5	0.99	1.00	1.00	31786
6	0.99	0.21	0.35	436
accuracy			0.99	566140
macro avg	0.97	0.77	0.82	566140
weighted avg	0.99	0.99	0.99	566140

## TEST SU TEST-SET BILANCIATO

--- Test Set (Bilanciato) ---

Accuracy: 0.965103598691385

F1 Score: 0.9648331730445453

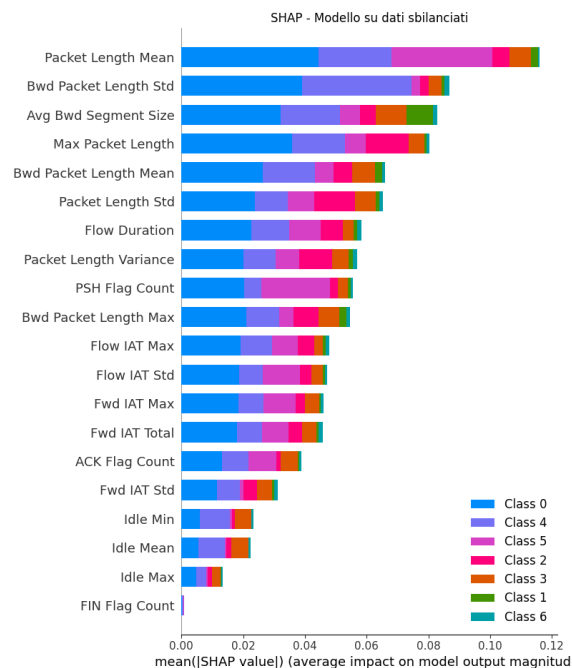
Precision: 0.9653567315656155

Recall: 0.965103598691385

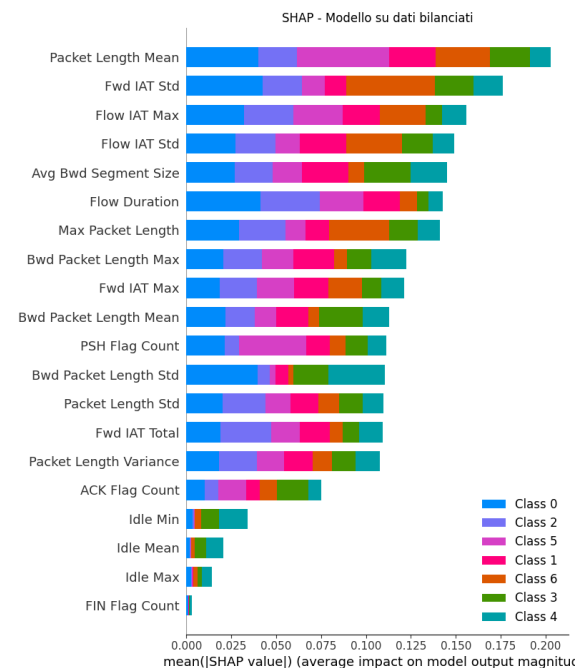
	precision	recall	f1-score	support
0	0.96	0.88	0.92	393
1	0.97	0.99	0.98	393
2	0.92	0.97	0.94	393
3	1.00	1.00	1.00	393
4	0.96	0.99	0.98	393
5	1.00	1.00	1.00	393
6	0.96	0.93	0.94	393
accuracy			0.97	2751
macro avg	0.97	0.97	0.96	2751
weighted avg	0.97	0.97	0.96	2751


# Spiegabilità con SHAP

## CONTRIBUZIONE DELLE FEATURE SULLA PREDIZIONE (DATI SBILANCIATI)



## CONTRIBUZIONE DELLE FEATURE SULLA PREDIZIONE (DATI BILANCIATI)





# Salvataggio del Modello

- Serializzazione con `joblib`.
- Oggetti salvati: modello, scaler, encoder, imputer, feature selezionate.
- Vantaggi: riutilizzabilità, portabilità, coerenza.



# Conclusioni e Futuri Sviluppi

- Sistema robusto, interpretabile, pronto per l'uso reale.
- Miglioramenti futuri:
  - Notifiche mobile.
  - Dashboard interattiva con SHAP in tempo reale.