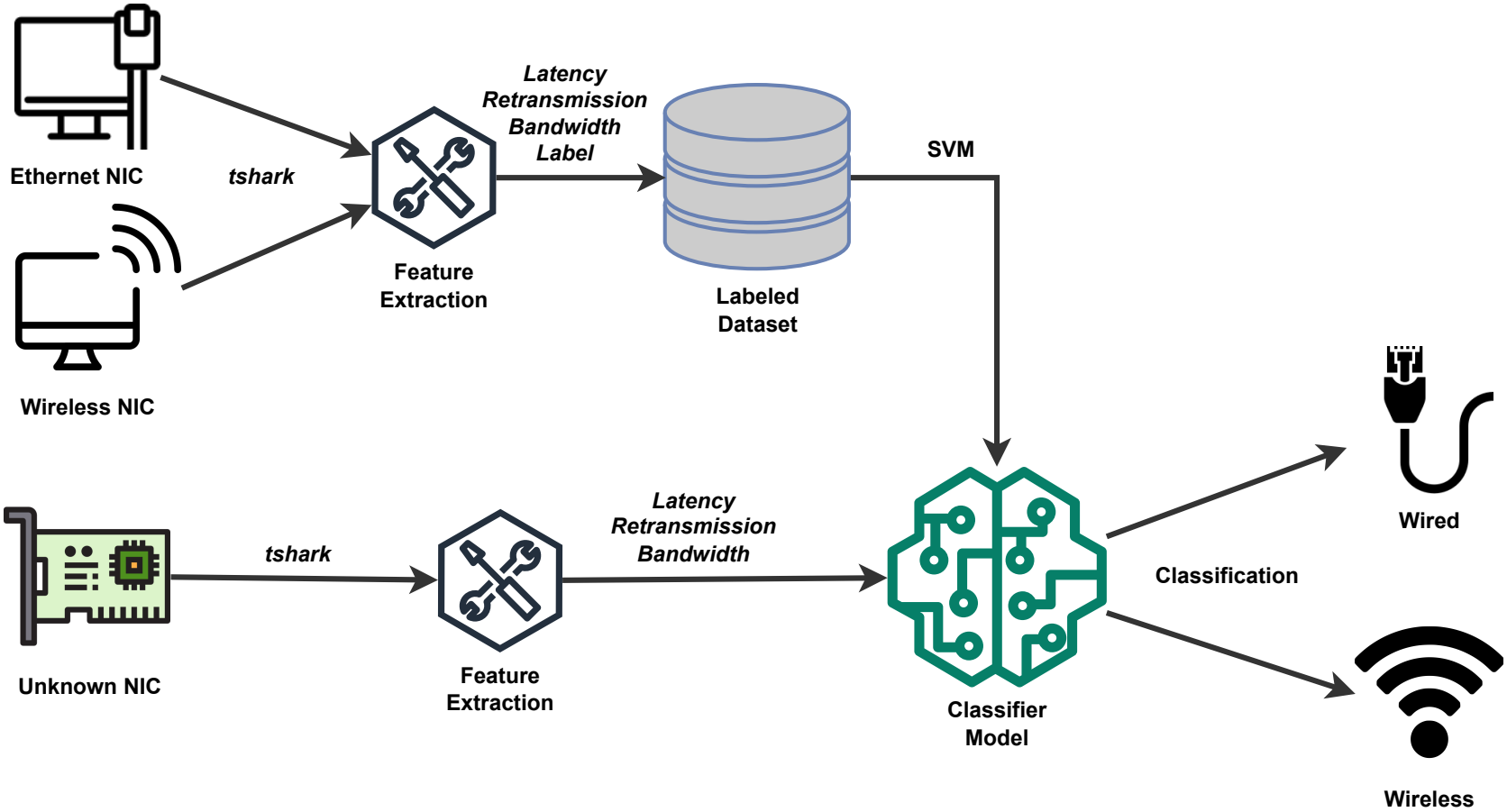
Preparazione ed analisi di un dataset di traffico TCP da terminali wireless/wired

Alex Ardelean

Dipartimento di Ingegneria
Università degli Studi di Perugia

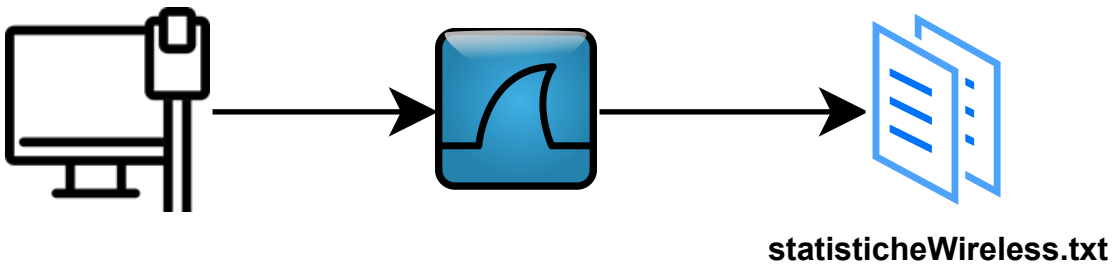
2023

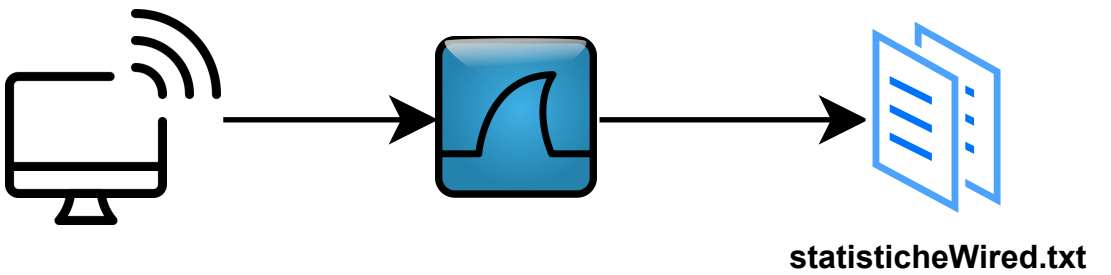
Approccio


Ethernet NIC
tshark
Wireless NIC
Feature Extraction
Latency
Retransmission
Bandwidth
Label
Labeled Dataset
SVM
Unknown NIC
tshark
Feature Extraction
Latency
Retransmission
Bandwidth
Classifier Model
Classification
Wired
Wireless

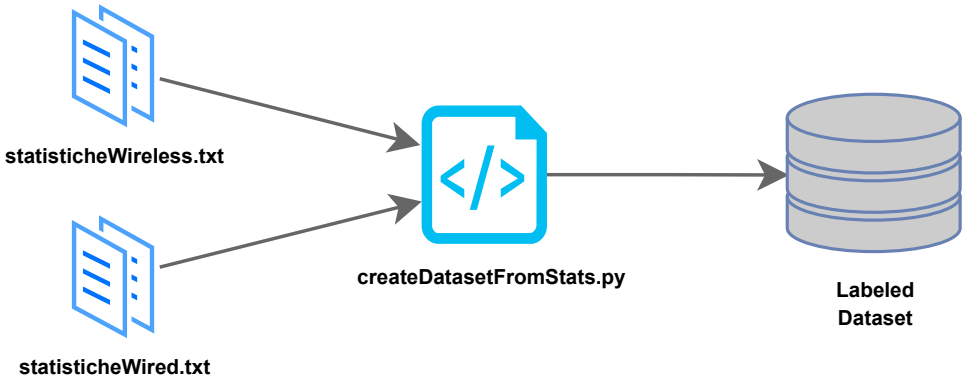
Cattura del traffico

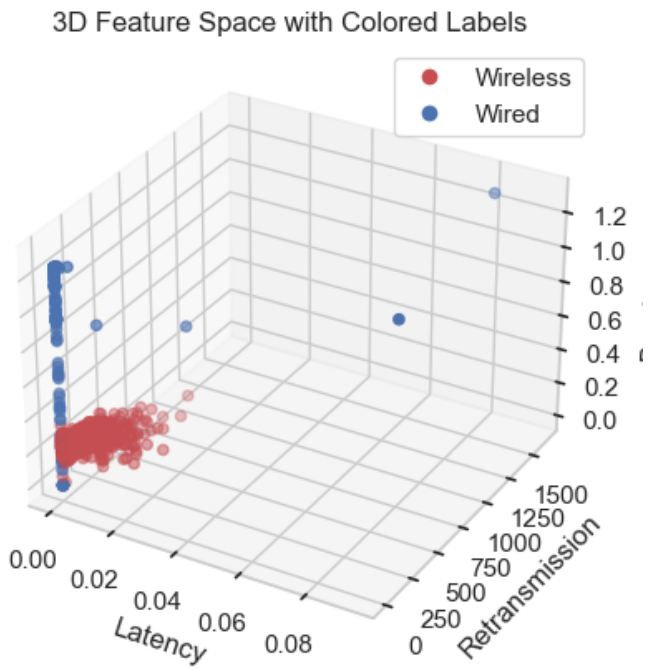
```
tshark -i <interface> -a duration:600 -q -z
"io,stat,1,
AVG(tcp.analysis.ack_rtt)tcp.analysis.ack_rtt,
COUNT(tcp.analysis.retransmission)tcp.analysis.retransmission,
BYTES"
> nomeFile
```



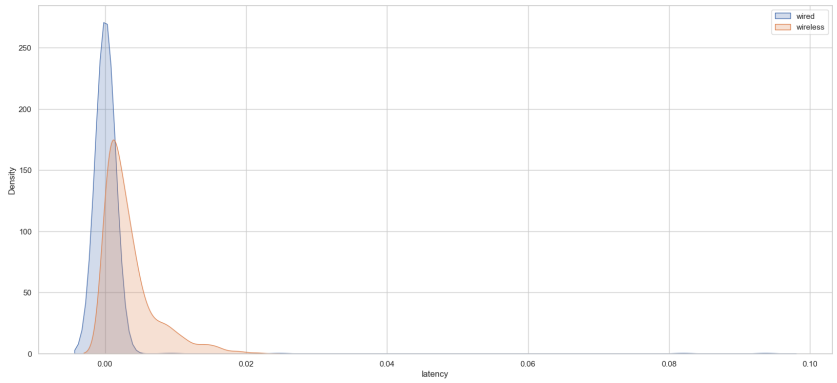


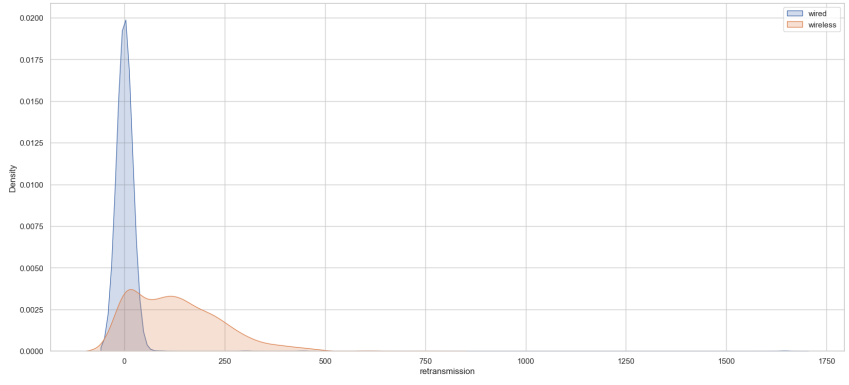
Estrazione delle features

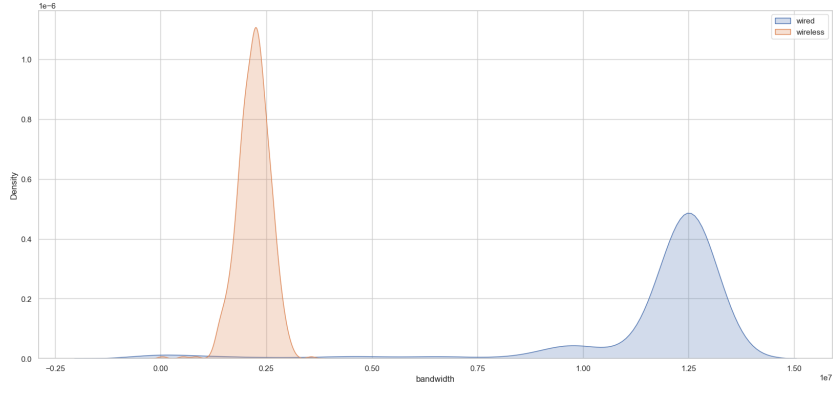




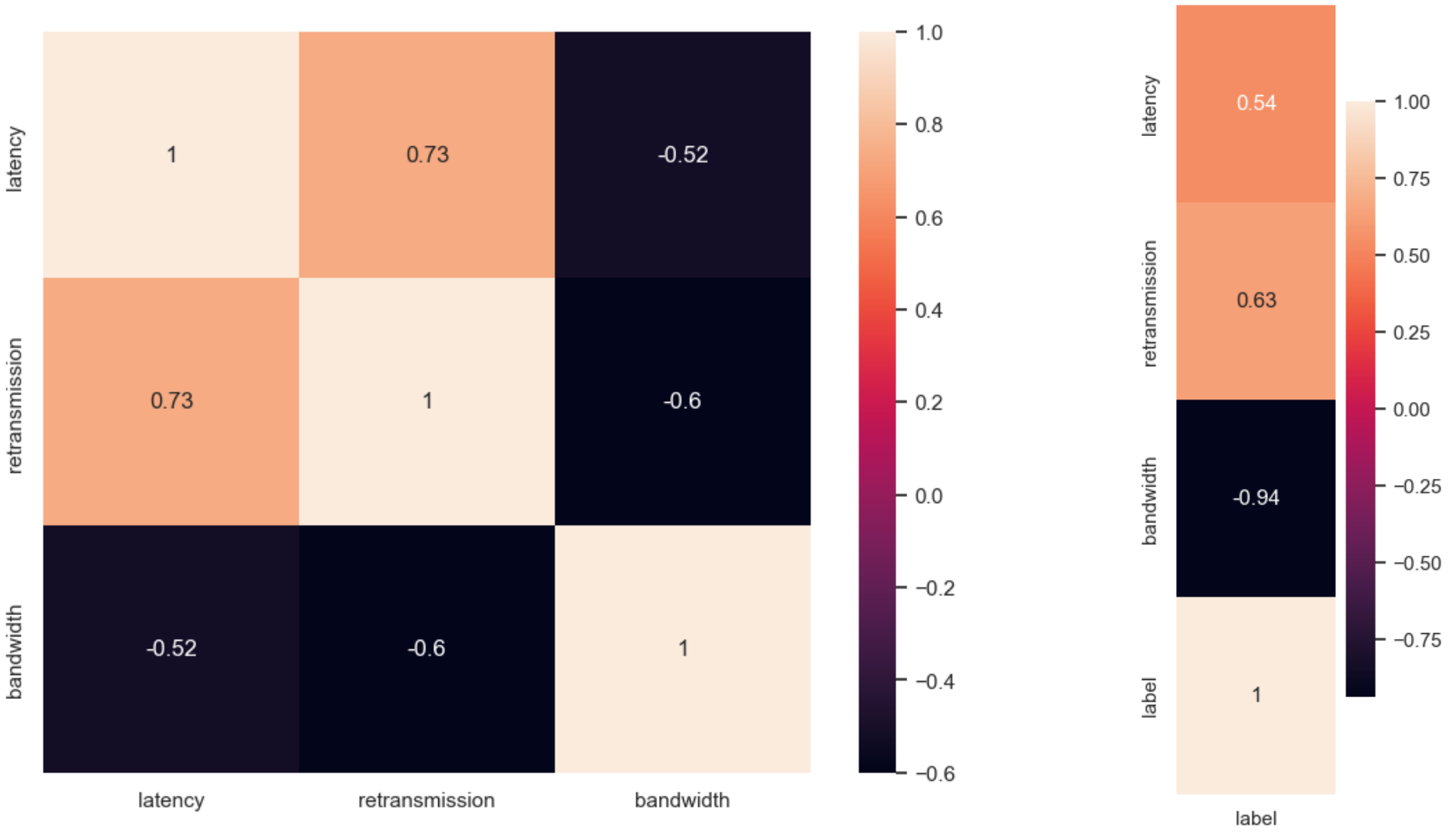
Distribuzione delle features rispetto alla classe



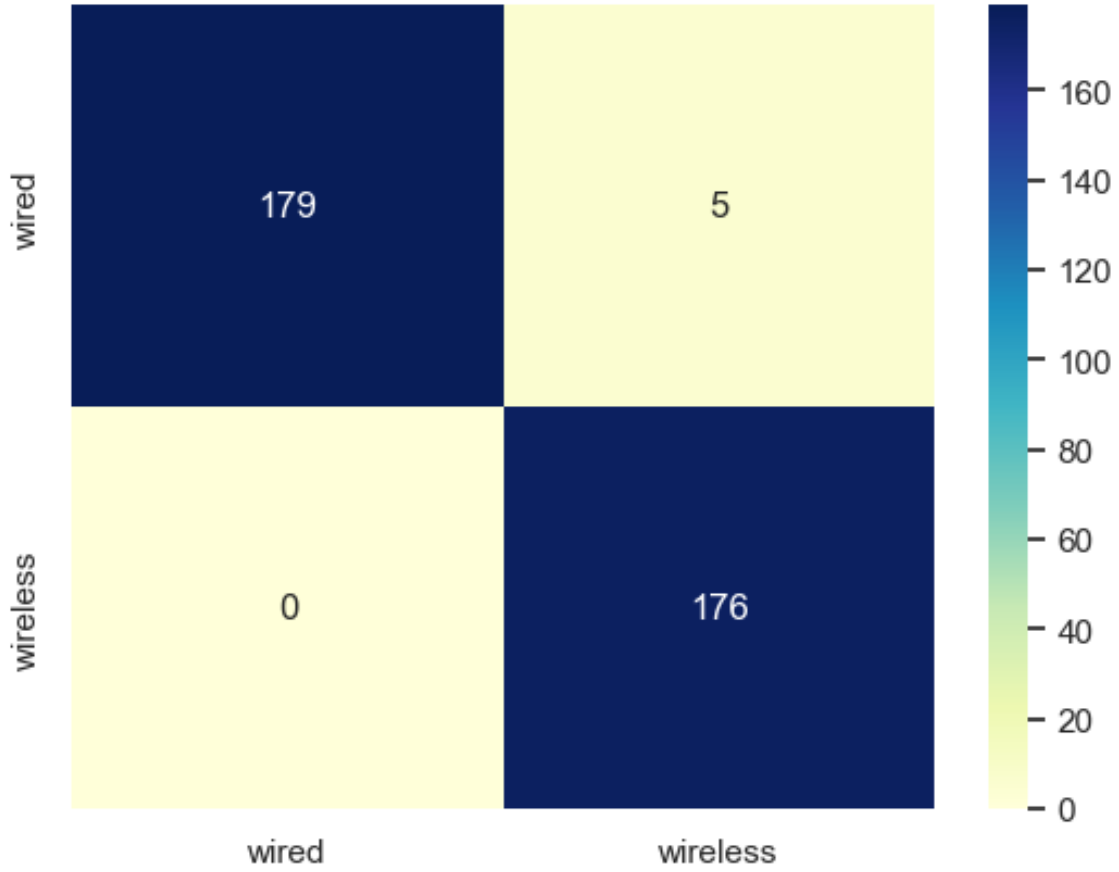




Correlazione

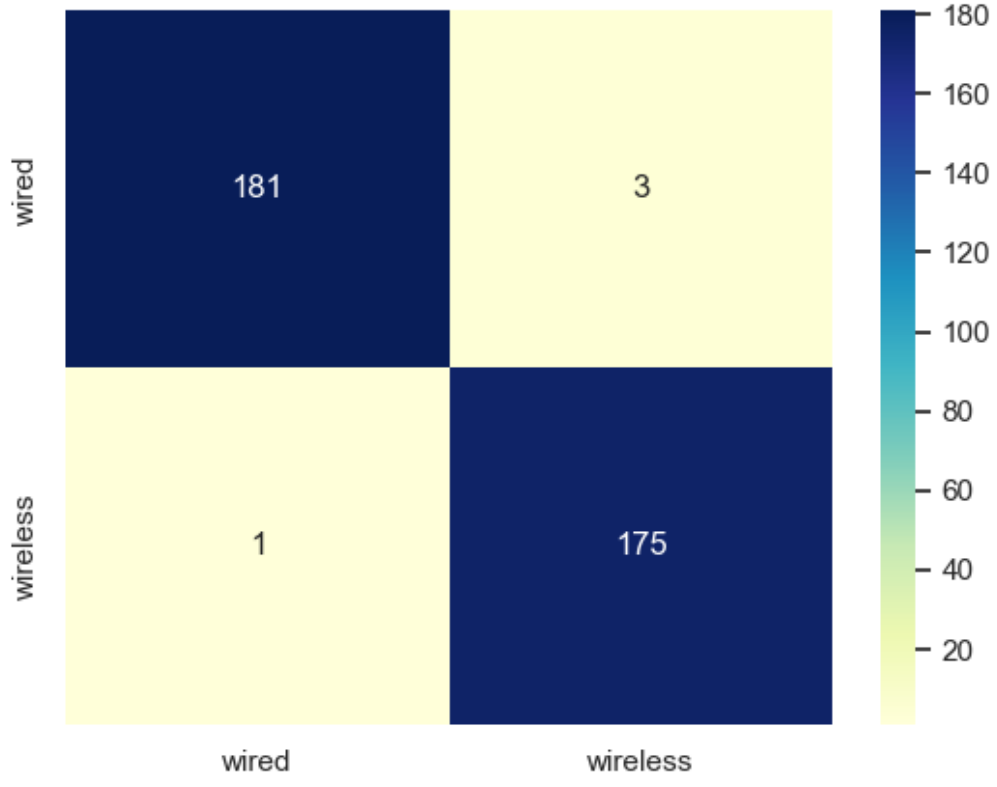


Logistic Regression



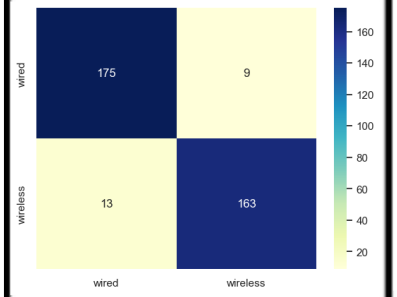
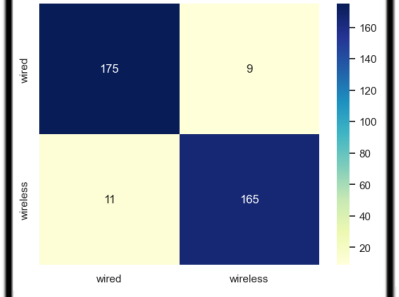
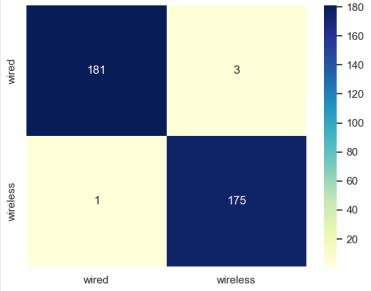
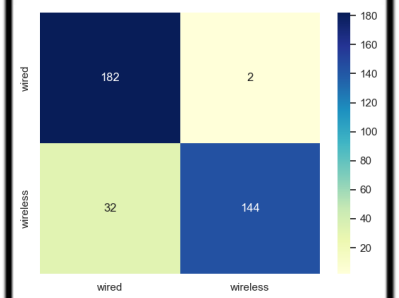
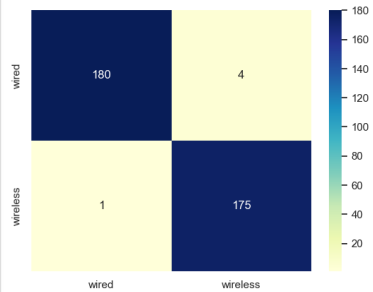
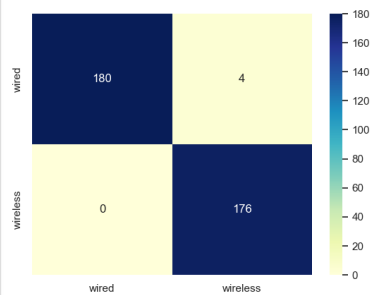
- Accuracy on the test set: 0.986
- F1 on the test set: 0.985

SVM

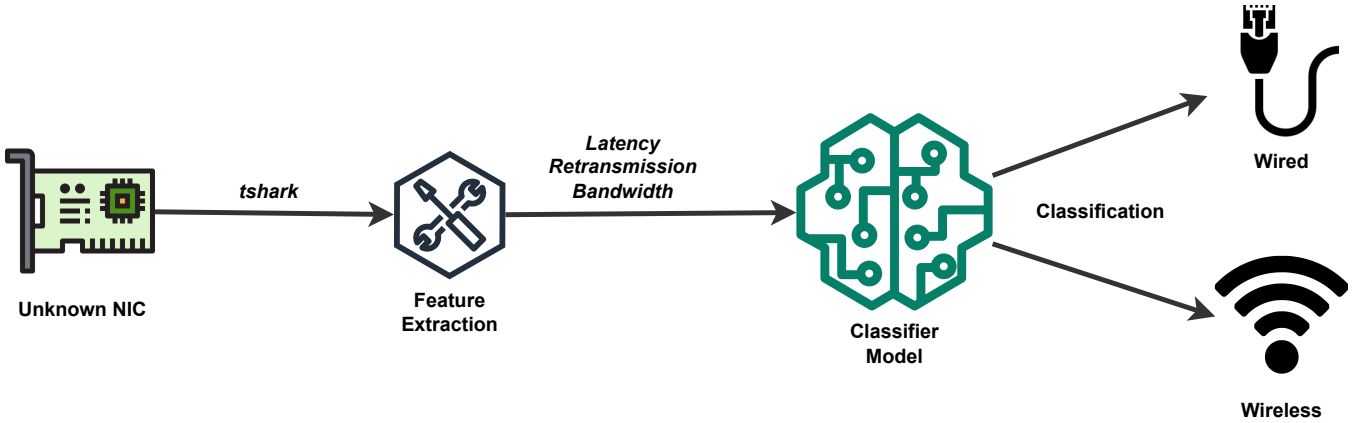


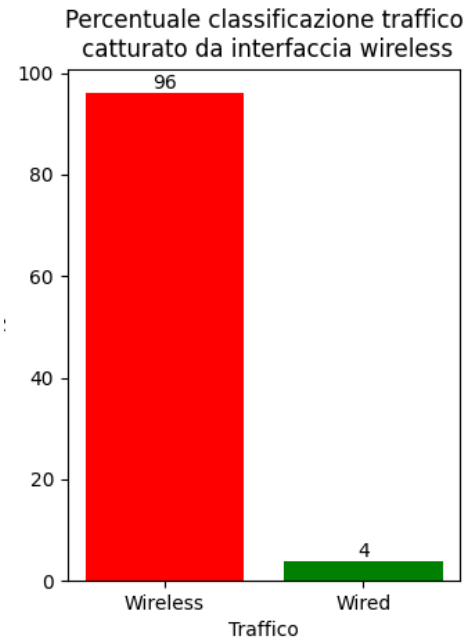

```
GridSearch CV best score : 0.9893


Parameters that give the best results :

 {'C': 100, 'gamma': 0.4, 'kernel': 'rbf'}


Estimator that was chosen by the search :

 SVC(C=100, gamma=0.4)
Model classification report with GridSearch CV:
              precision    recall  f1-score   support

           0       0.99      0.98      0.99       184
           1       0.98      0.99      0.99       176

    accuracy                           0.99       360
   macro avg       0.99      0.99      0.99       360
weighted avg       0.99      0.99      0.99       360
```

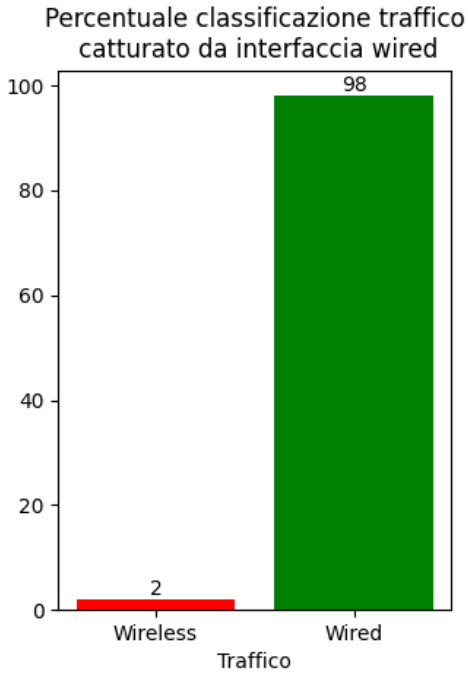
Impatto delle singole features

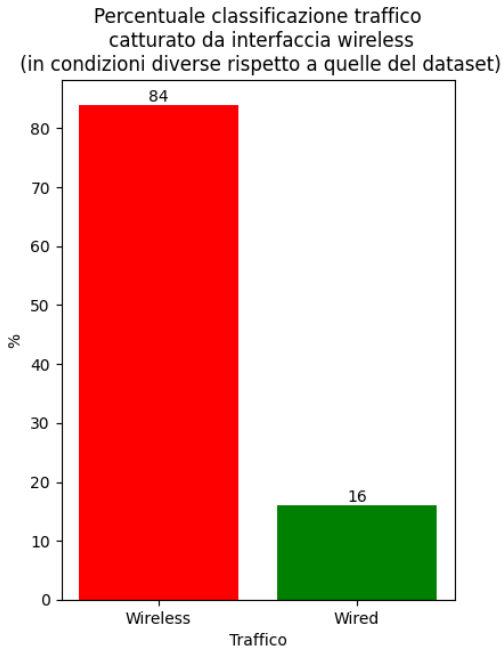
	Latenza	Ritrasmissioni	Banda
Latenza			
Ritrasmissioni			
Banda			

Test su traffico live









Impatto sulle prestazioni