

01

PROGETTO MACHINE E DEEP LEARNING



Studente:
Mattia Gatto, 216649

Docenti:
Fabrizio Angiulli, Fabio Fassetti,
Luca Ferragina



02

PROBLEM

Classificazione e anomaly detection di immagini

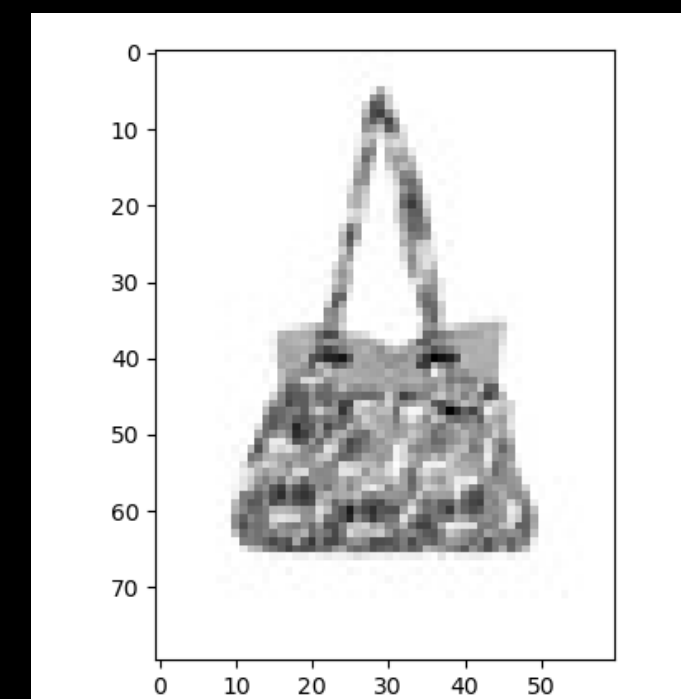
VISUALIZZAZIONE

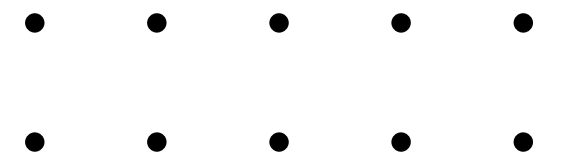
Fase di analisi dei dati, per valutare l'insieme migliore di tecniche da definire per la fase di pre-processing.

Il formato dati è una cartella che dispone di 5477 immagini a colori, di dimensione 80*60, in formato jpg.

Vi sono 4 tipi di immagini:

- “handbags”
- “sports-shoes”
- “tops”
- “trousers”

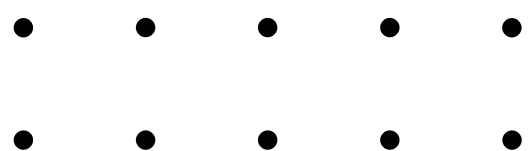




PREPROCESSING

Le operazioni effettuate sul dataset di immagini sono state:

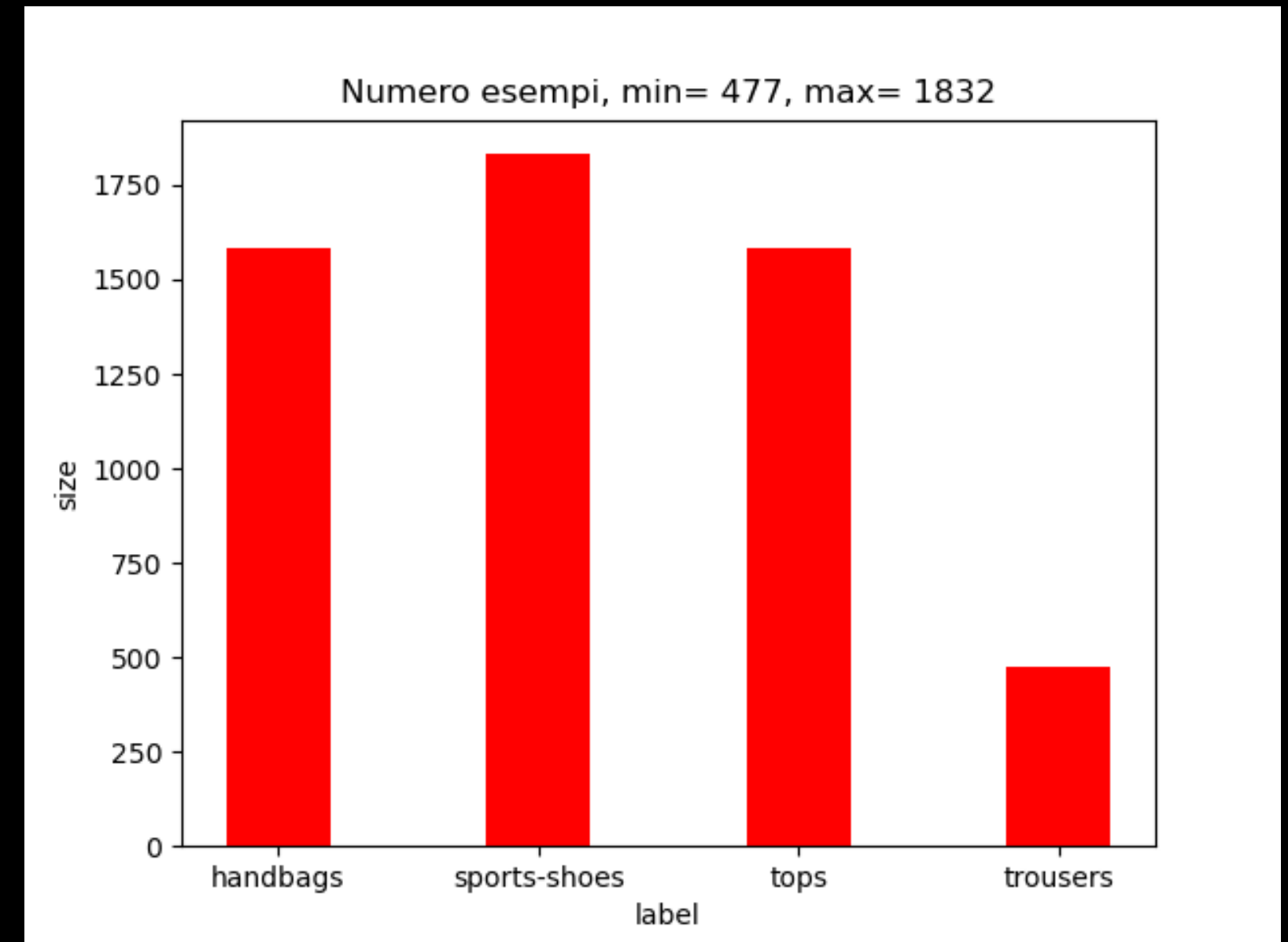
- Trasformazione tramite la libreria “opencv” delle immagini in matrici;
- Normalizzazione delle intensità dei pixel secondo un formato float che va da 0 a 1;
- trasformazione delle immagini in scala di grigi.



LABEL

Per quanto riguarda le label ho definito un'indicatore per ogni tipo :

- 0 per “handbags”
- 1 per “sports-shoes”
- 2 per “tops”
- 3 per “trousers”

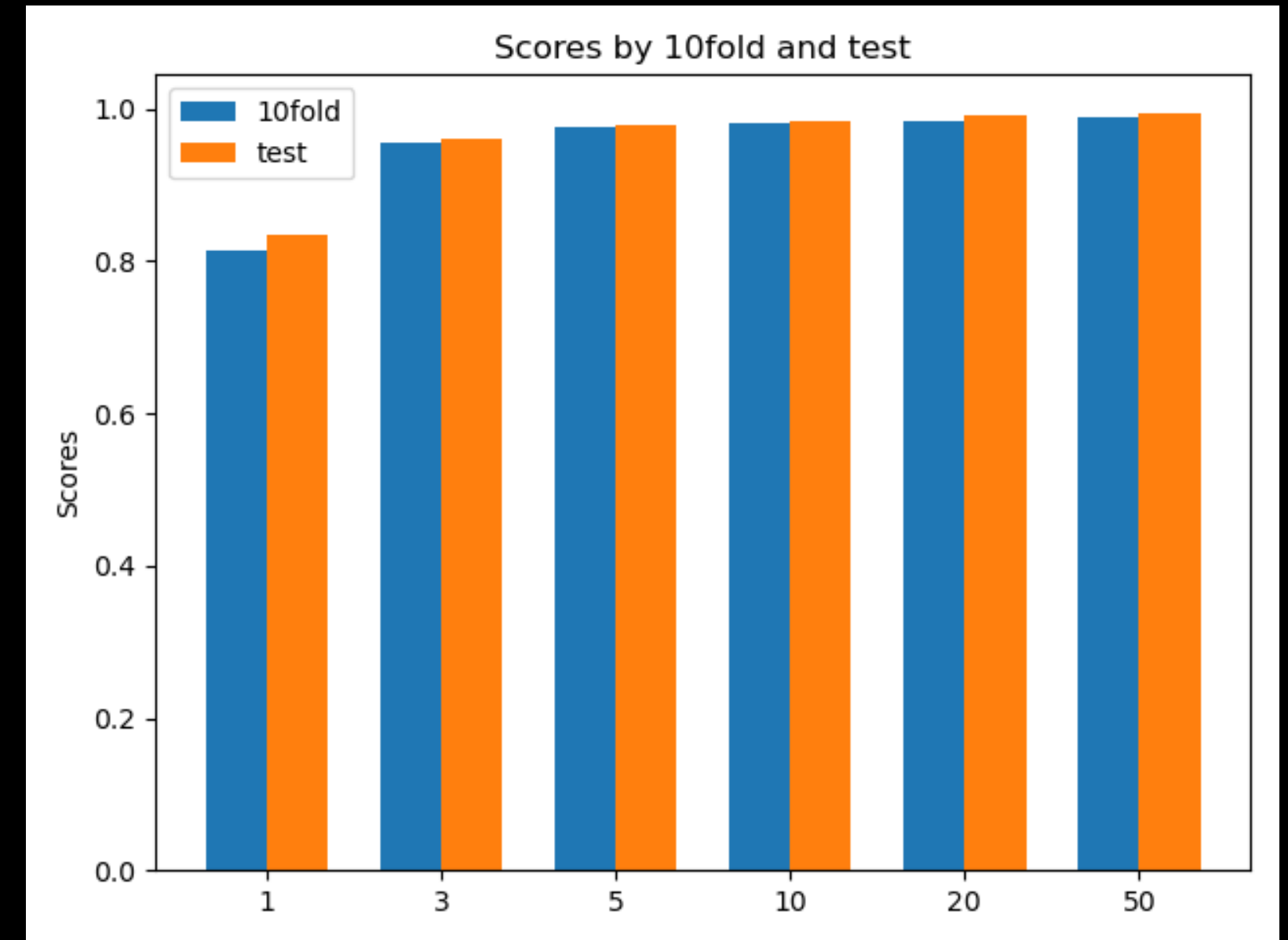
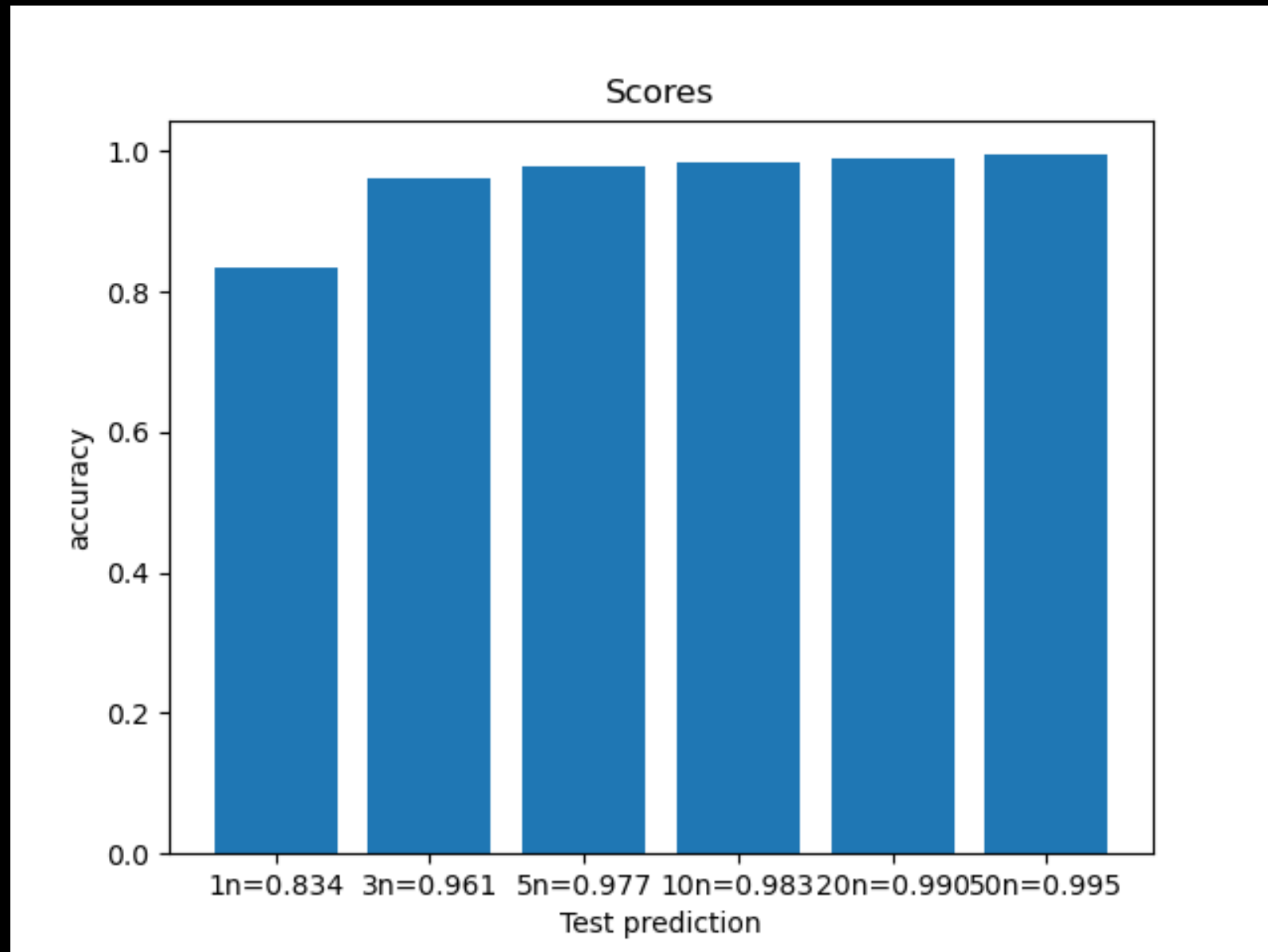


Per ogni tecnica di classificazione, al termine dell'addestramento ho effettuato :

- la valutazione dell'accuratezza con il test set;
- la valutazione delle prestazioni sfruttando kfold validation;
- confrontato i risultati ottenuti tra kfold validation e testvalidation;

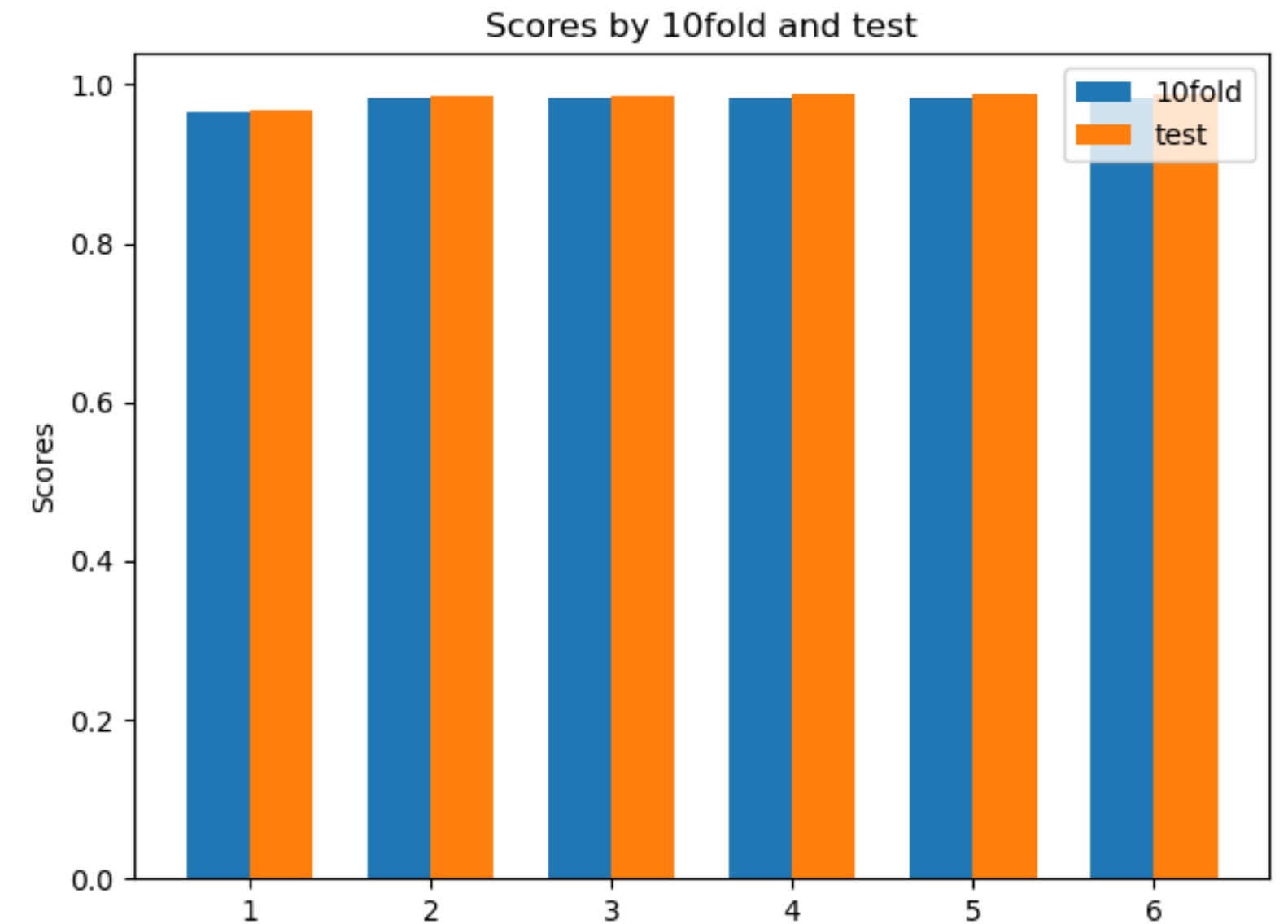
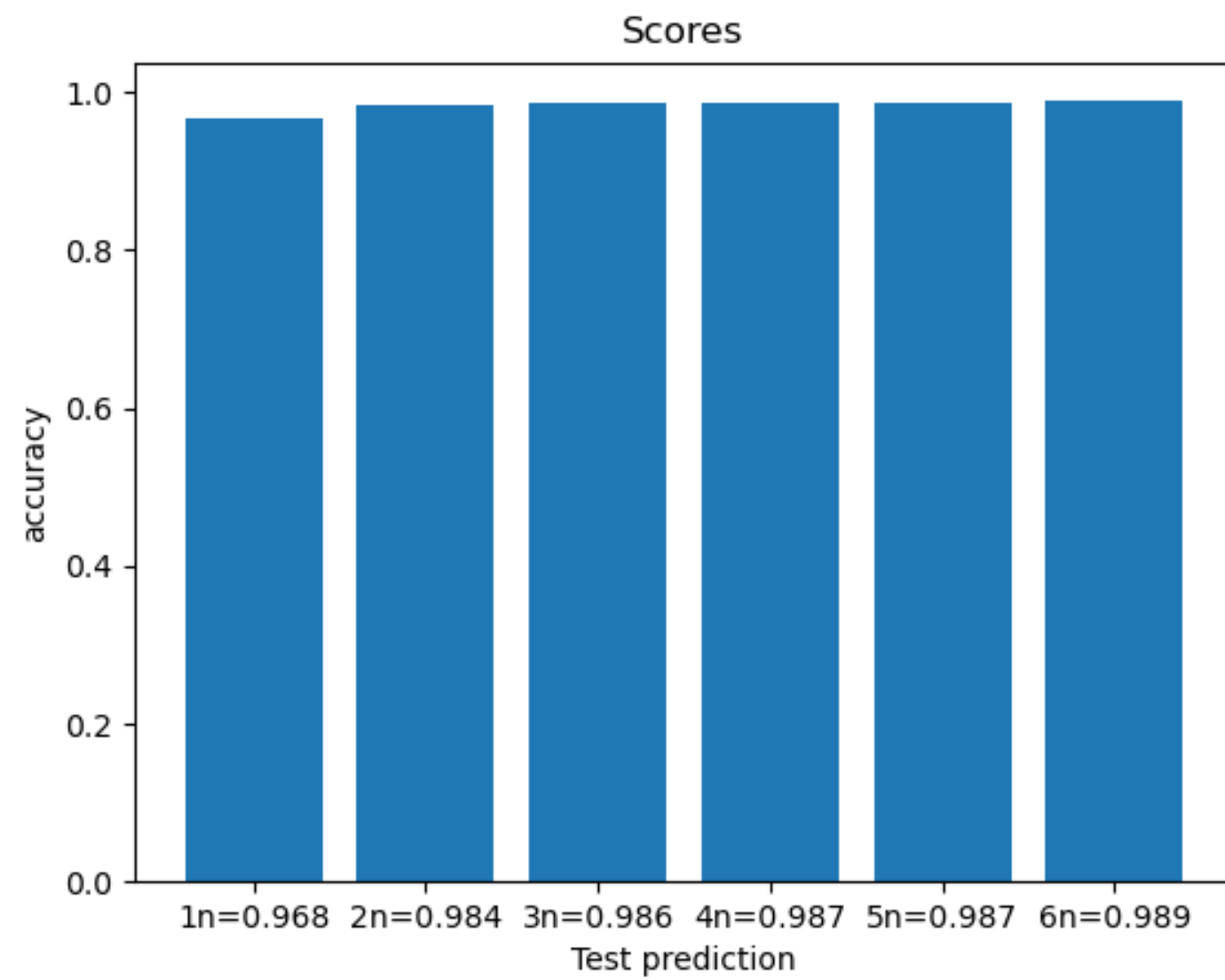
ADABOOST

06



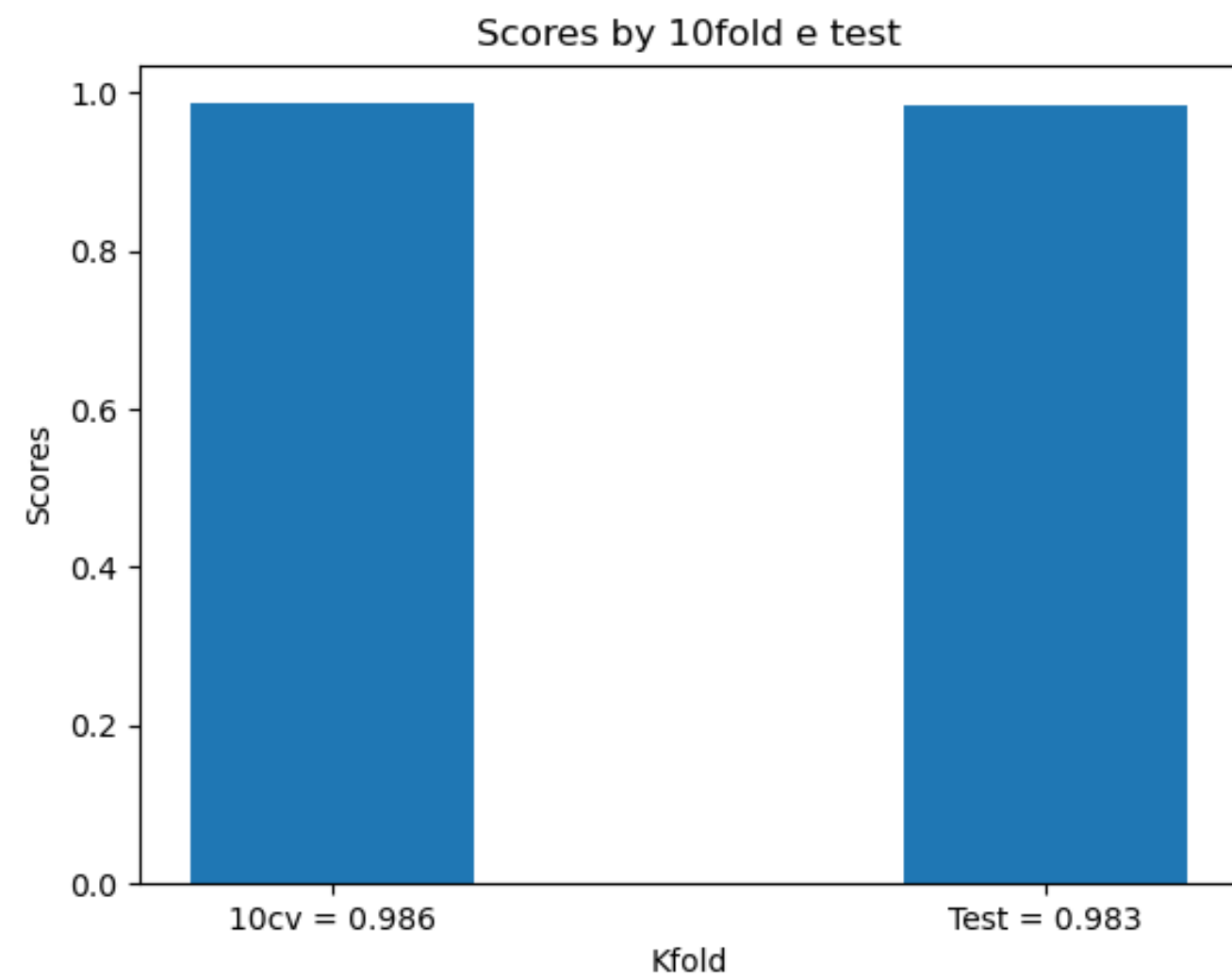
SVM

07



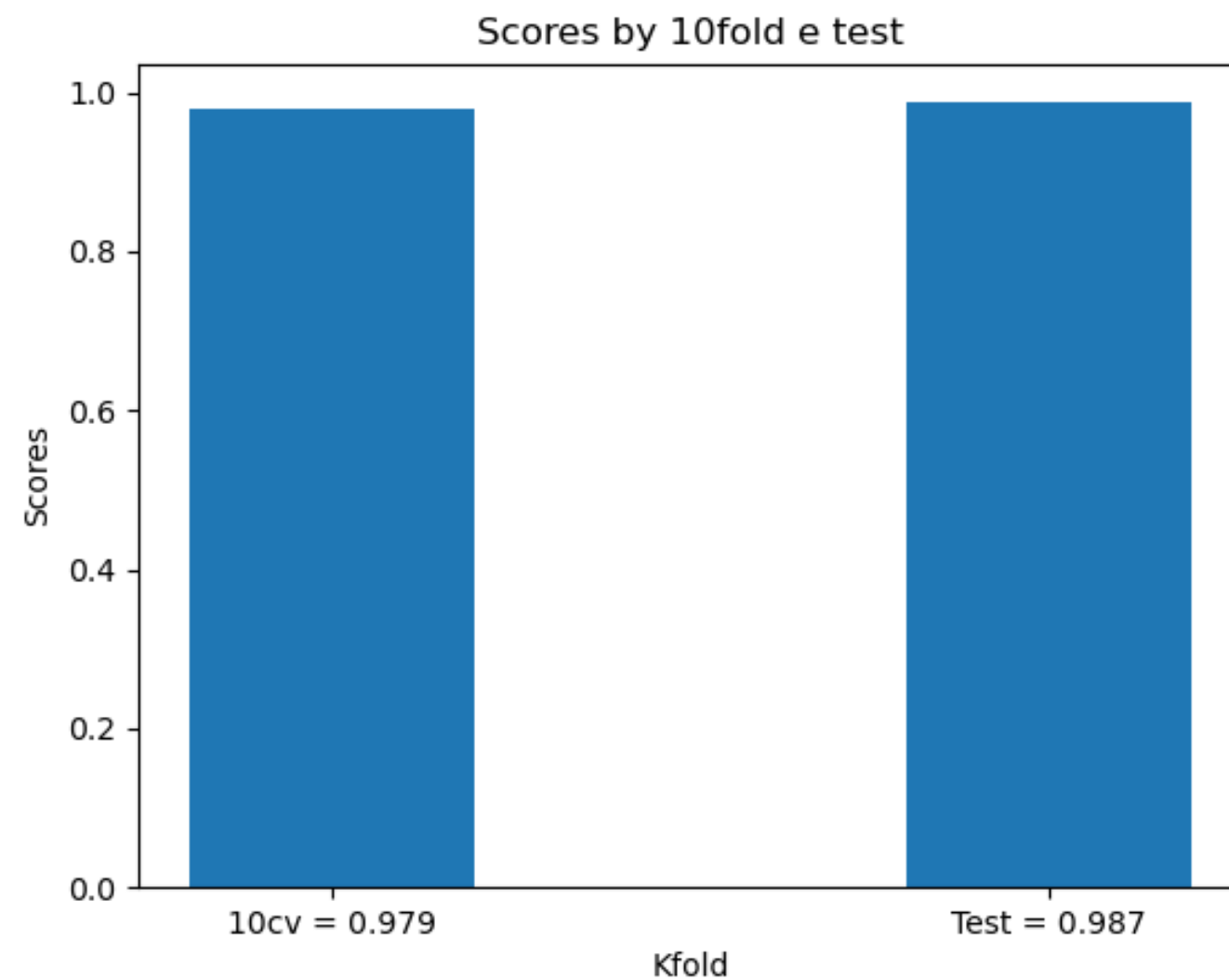
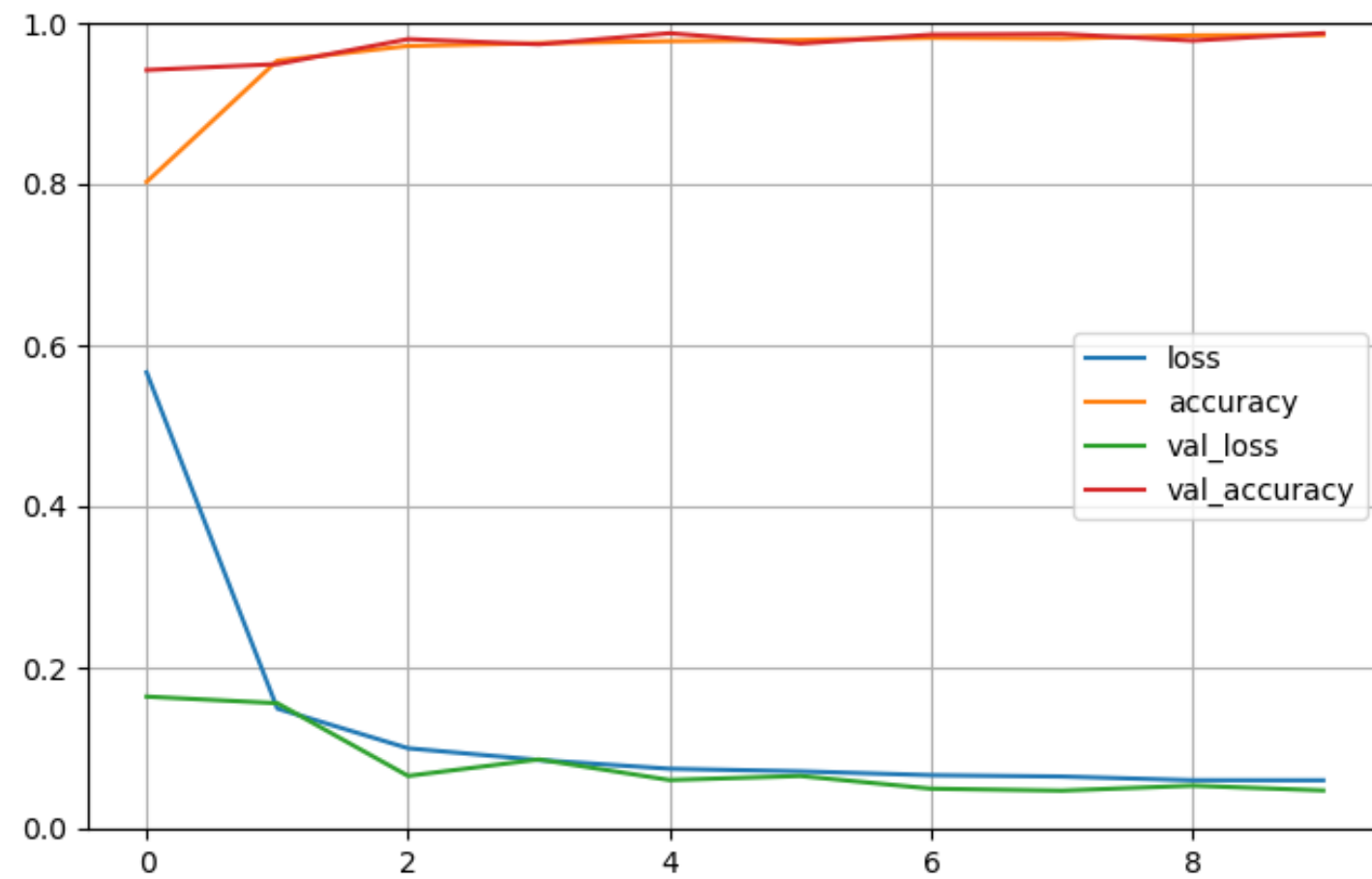
RETI NEURALI

- RETE DENSA



RETI NEURALI

- RETE
CONVOLUZIONALE



STIME DI DENSITÀ

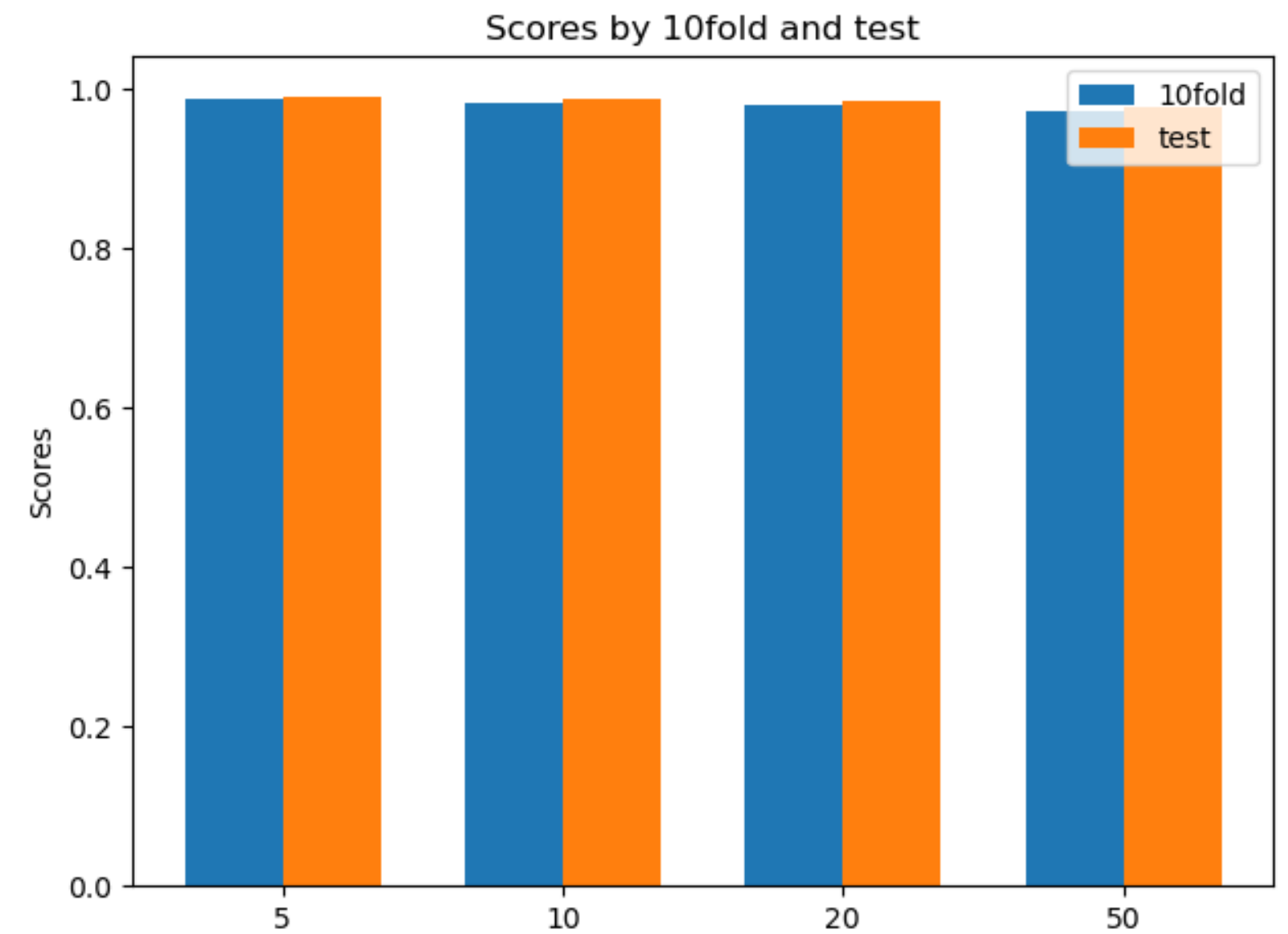
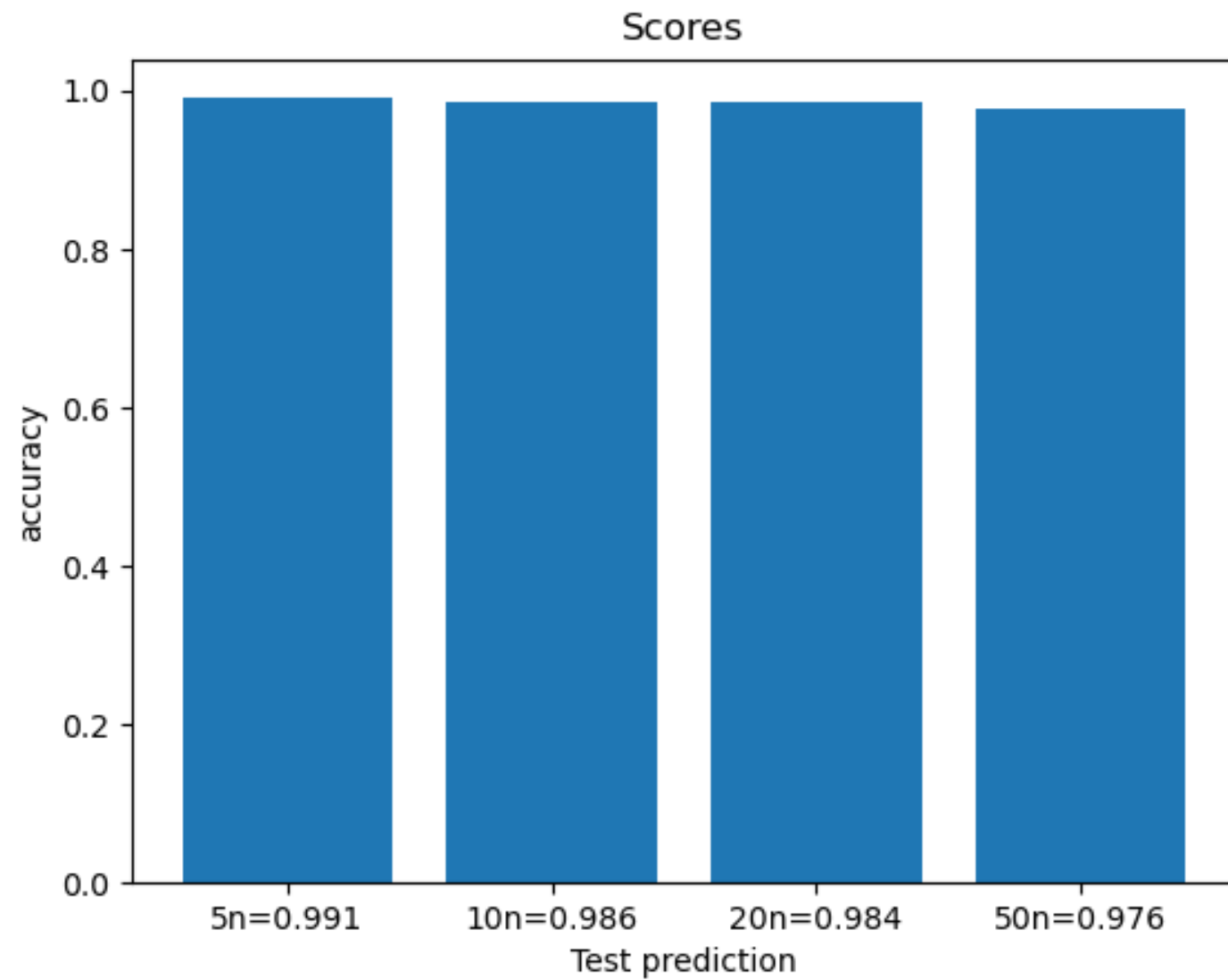
10

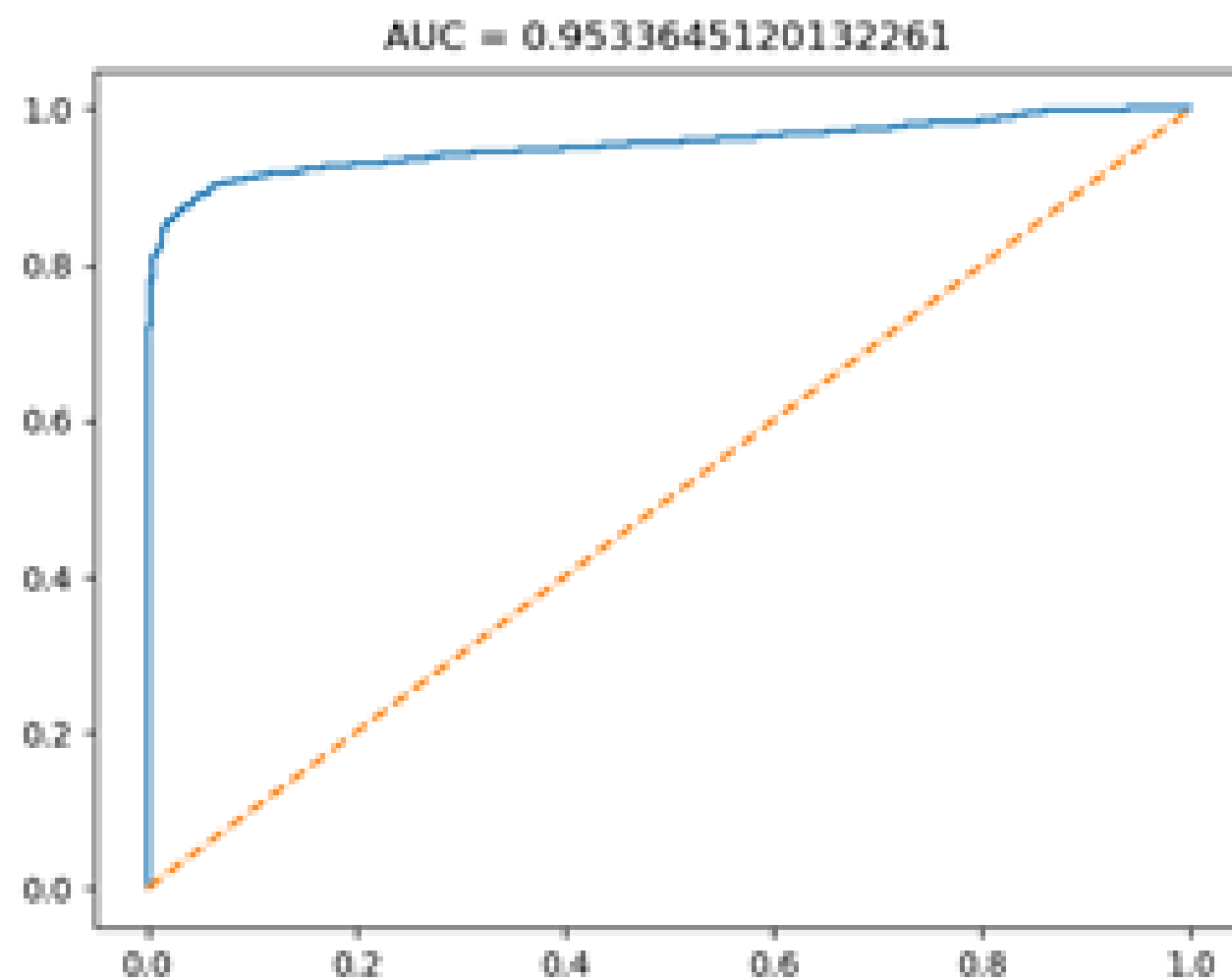
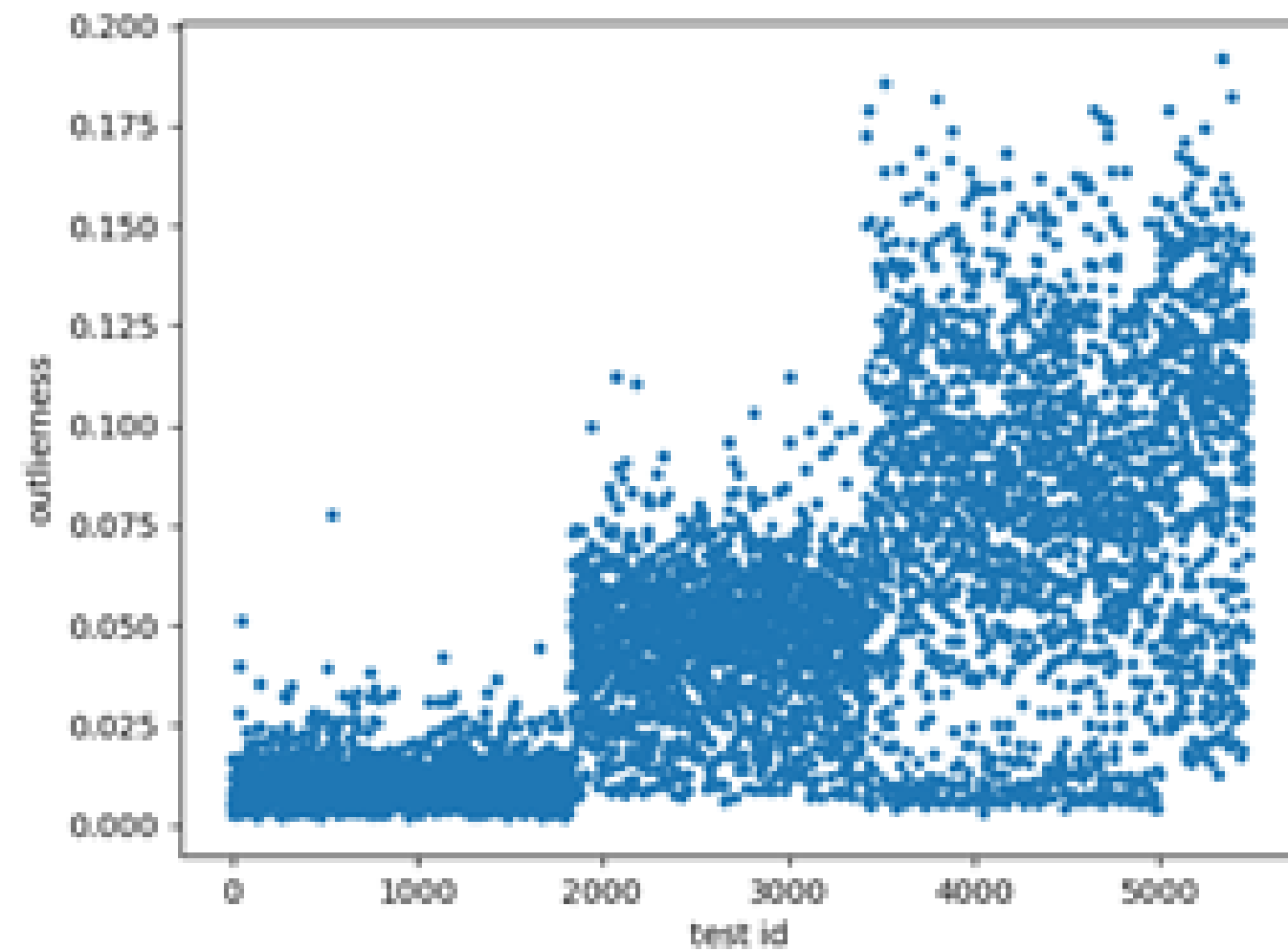
- GAUSSIAN NAIVE BAYES



11

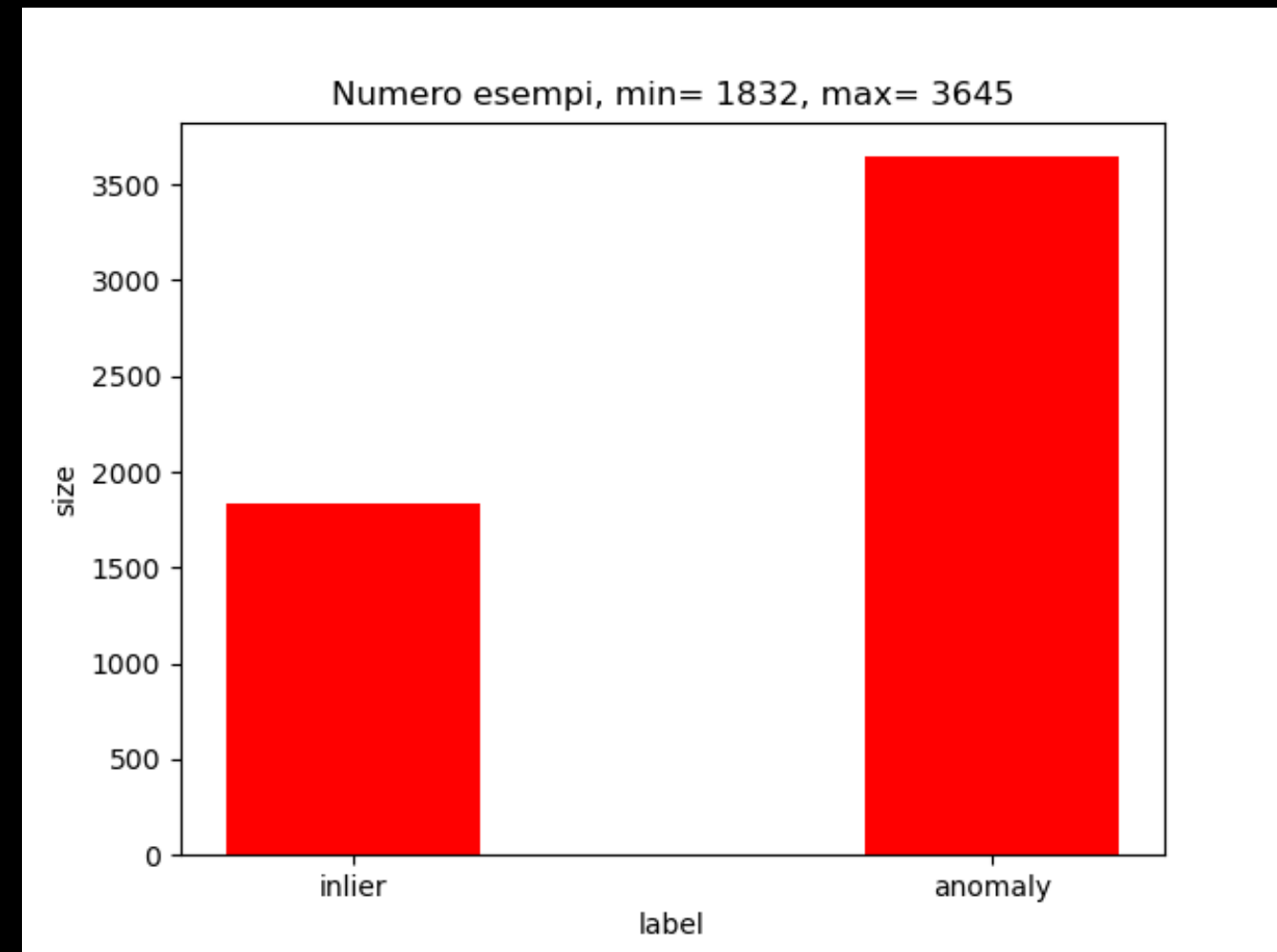
- K NEAREST NEIGHBOR





ANOMALY DETECTION

- RETE DENSA

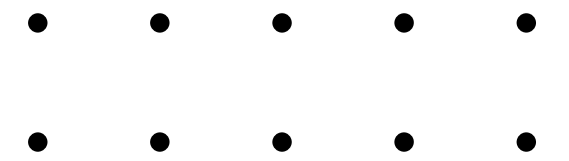


VISUALIZZAZIONE

Fase di analisi per valutare l'insieme migliore di tecniche da definire per la fase di pre-processing.

Il formato dati è un file excell che dispone di 10262 recensioni, ognuna delle quali ha una valutazione da 1 a 5.

PREPROCESSING

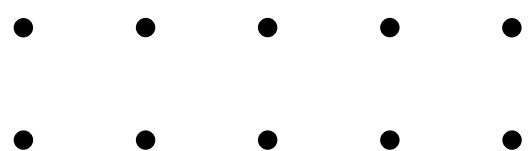


Attraverso l'utilizzo della libreria "nltk" al fine di riuscire elaborare ogni frase, infatti, per ognuna di esse vado ad effettuare:

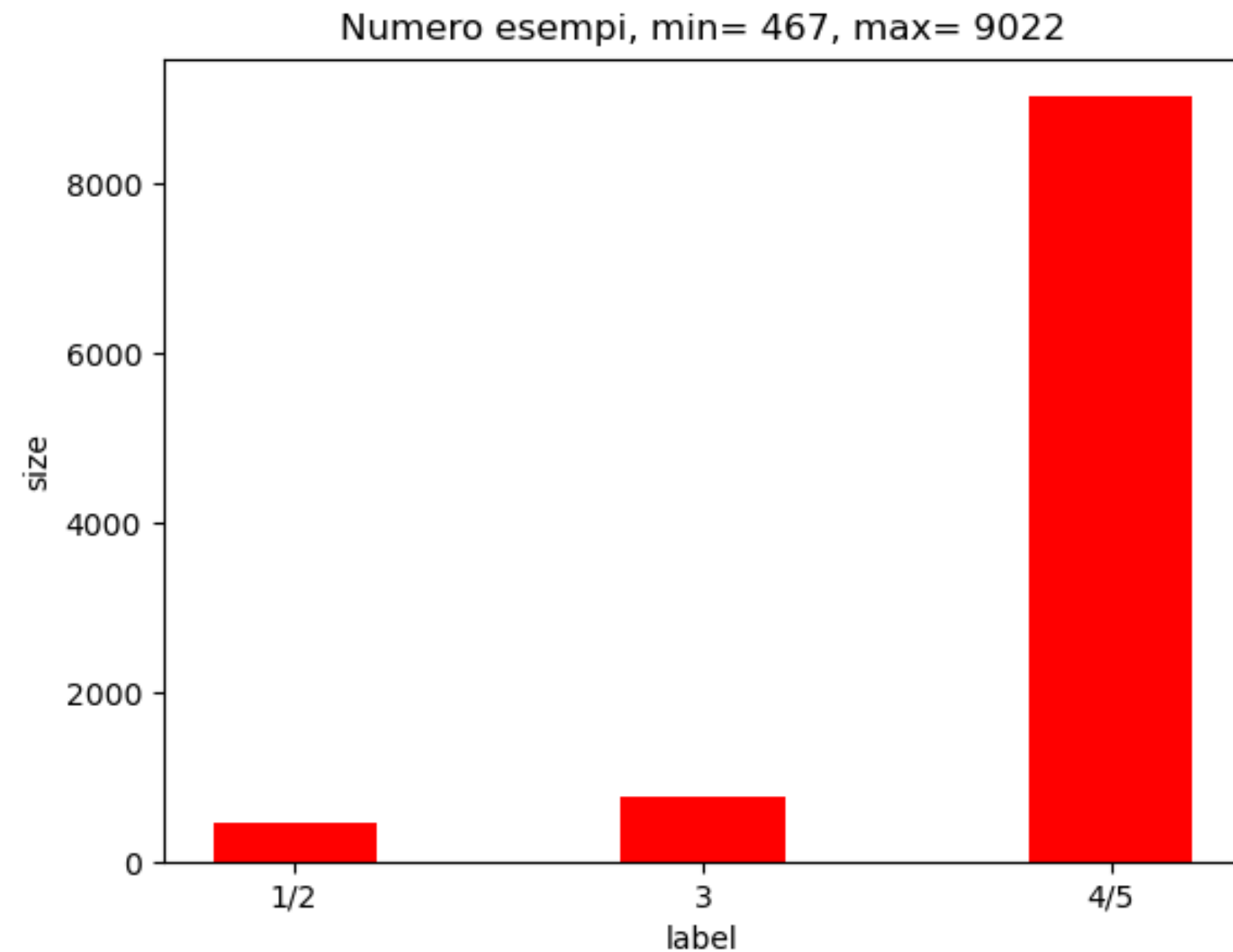
- Tokenizatione;
- Processo di rimozione della punteggiatura;
- Stopping;
- Stemming;
- Feature_extraction: nello specifico TfidfVectorizer.

Per quanto riguarda le label le ho definite secondo questo formato:

- 0 indica le label 1 e 2 cioè valutazione di tipo pessima.
- 1 indica la label 3 cioè valutazione di tipo normale.
- 2 indica le label 4 e 5 cioè valutazione di tipo ottima.



LABEL

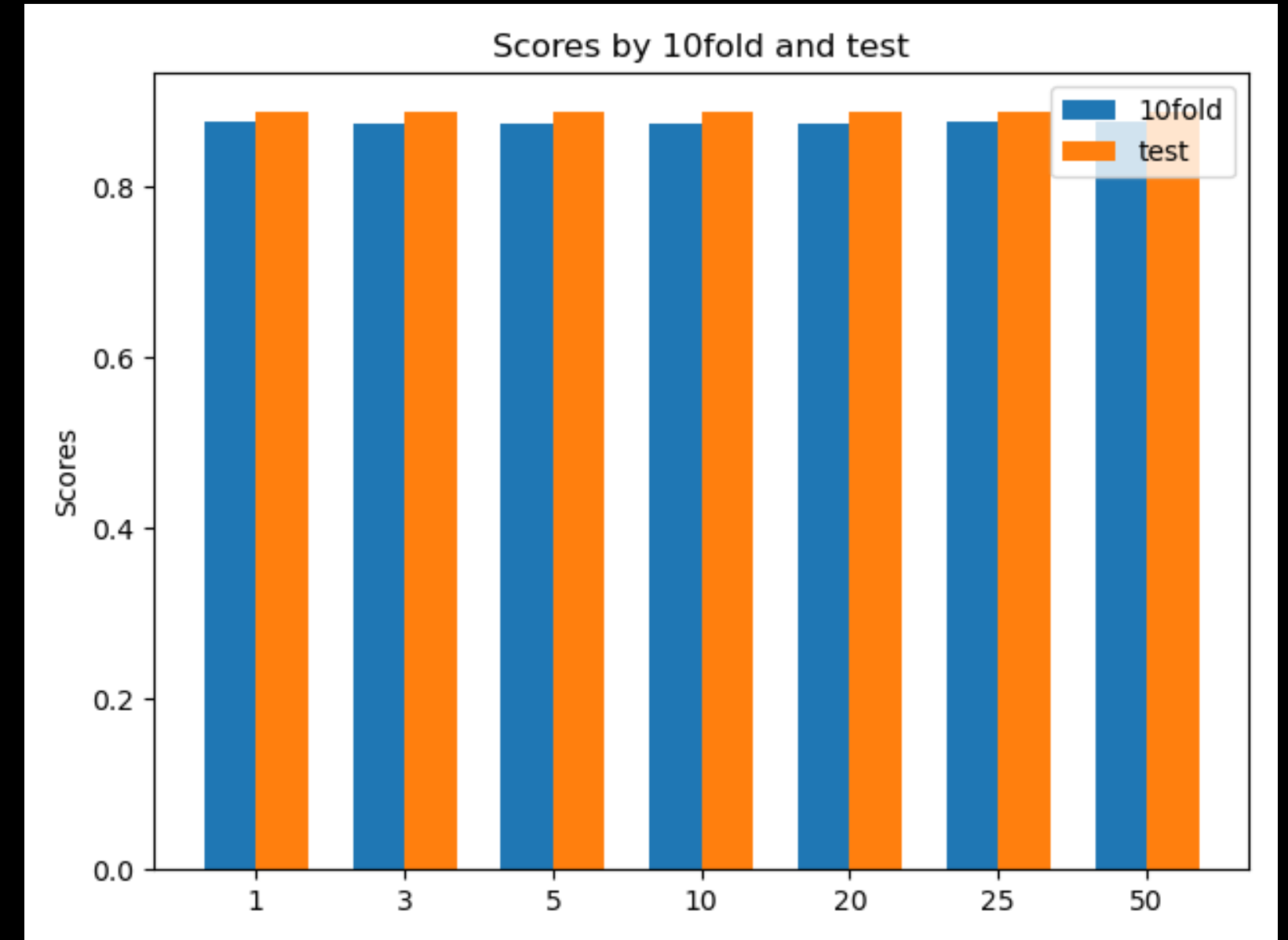
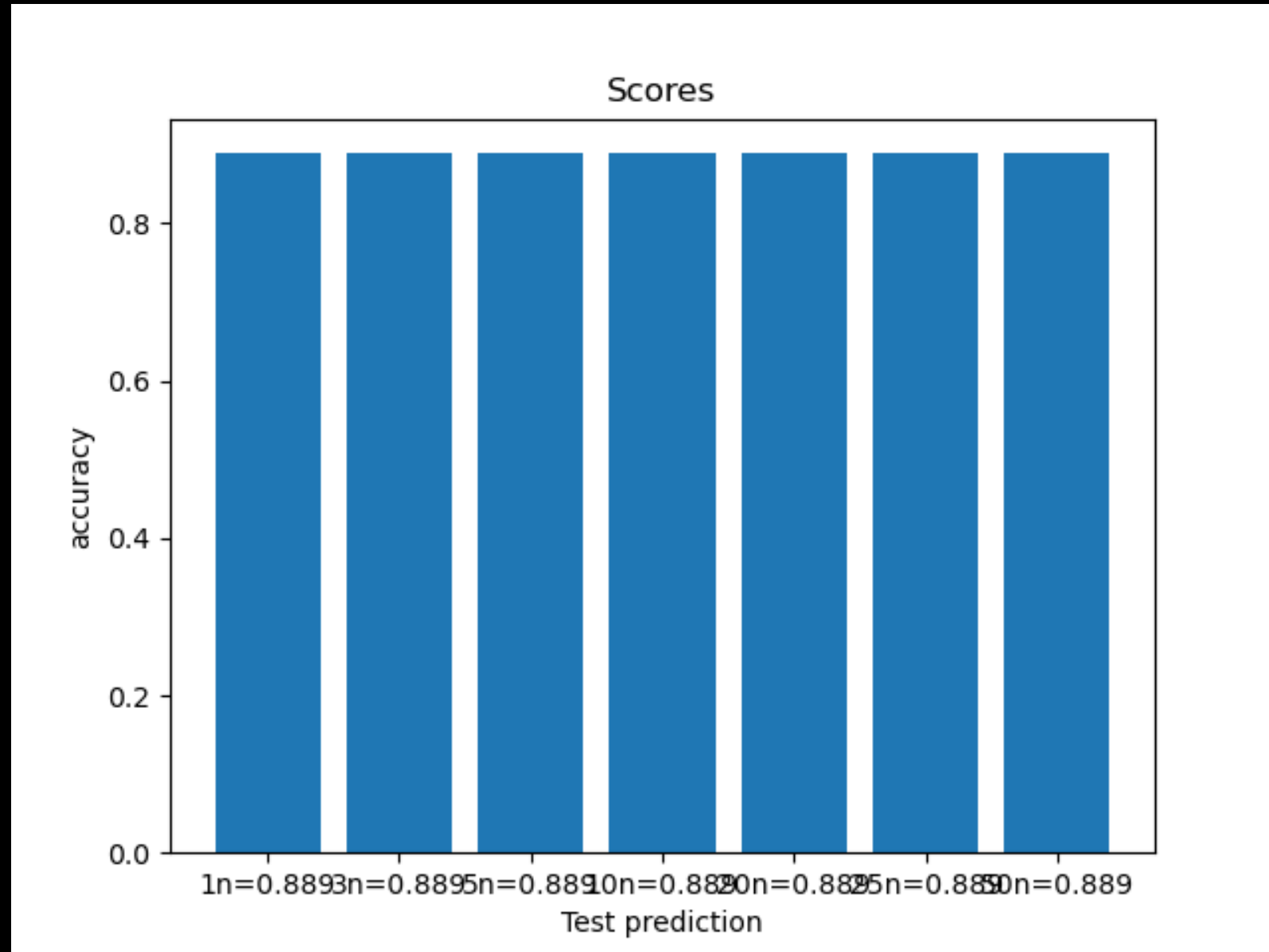


Per quanto riguarda le label le ho definite secondo questo formato:

- 0 indica le label 1 e 2 cioè valutazione di tipo pessima.
- 1 indica la label 3 cioè valutazione di tipo normale.
- 2 indica le label 4 e 5 cioè valutazione di tipo ottima.

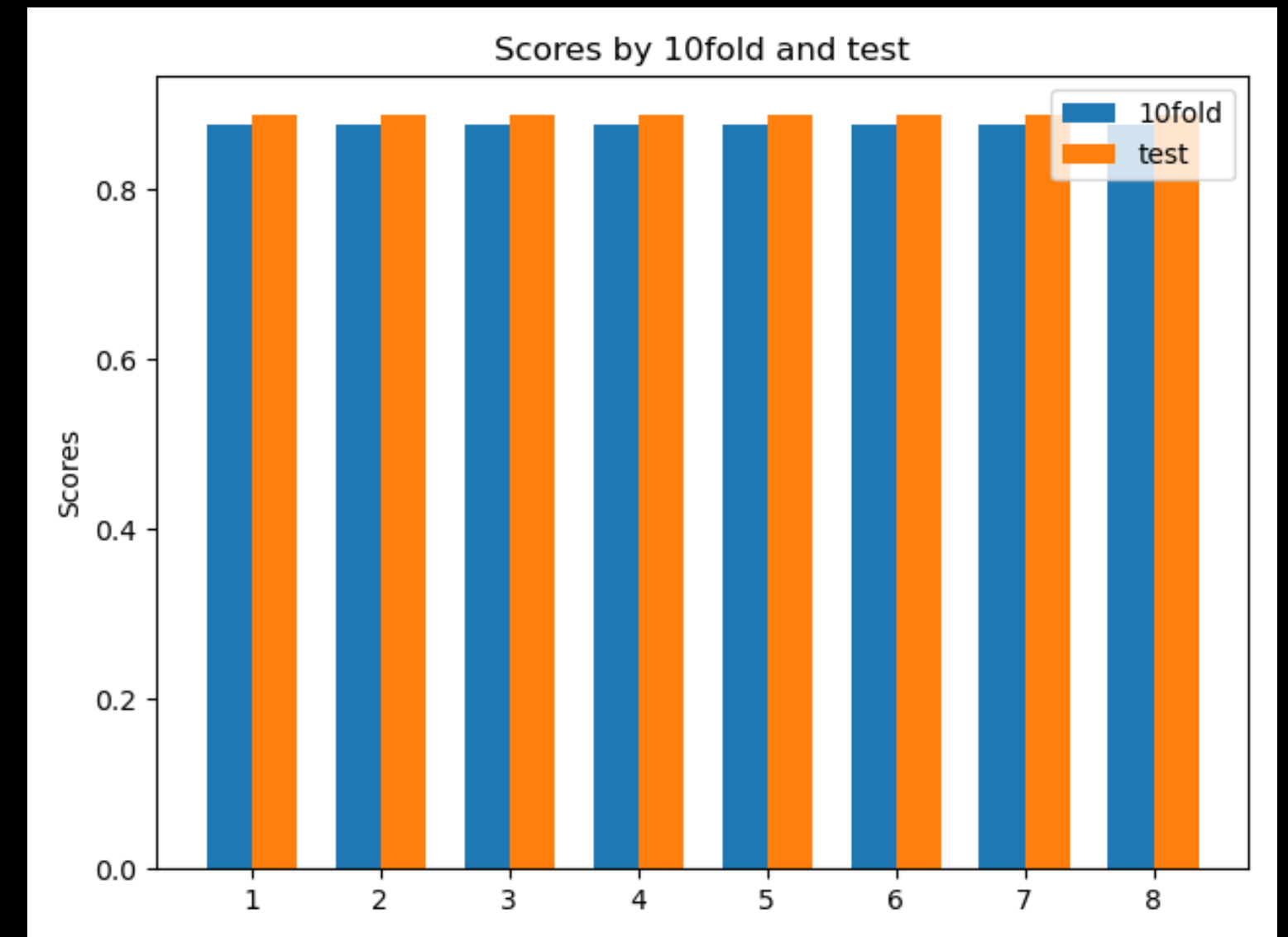
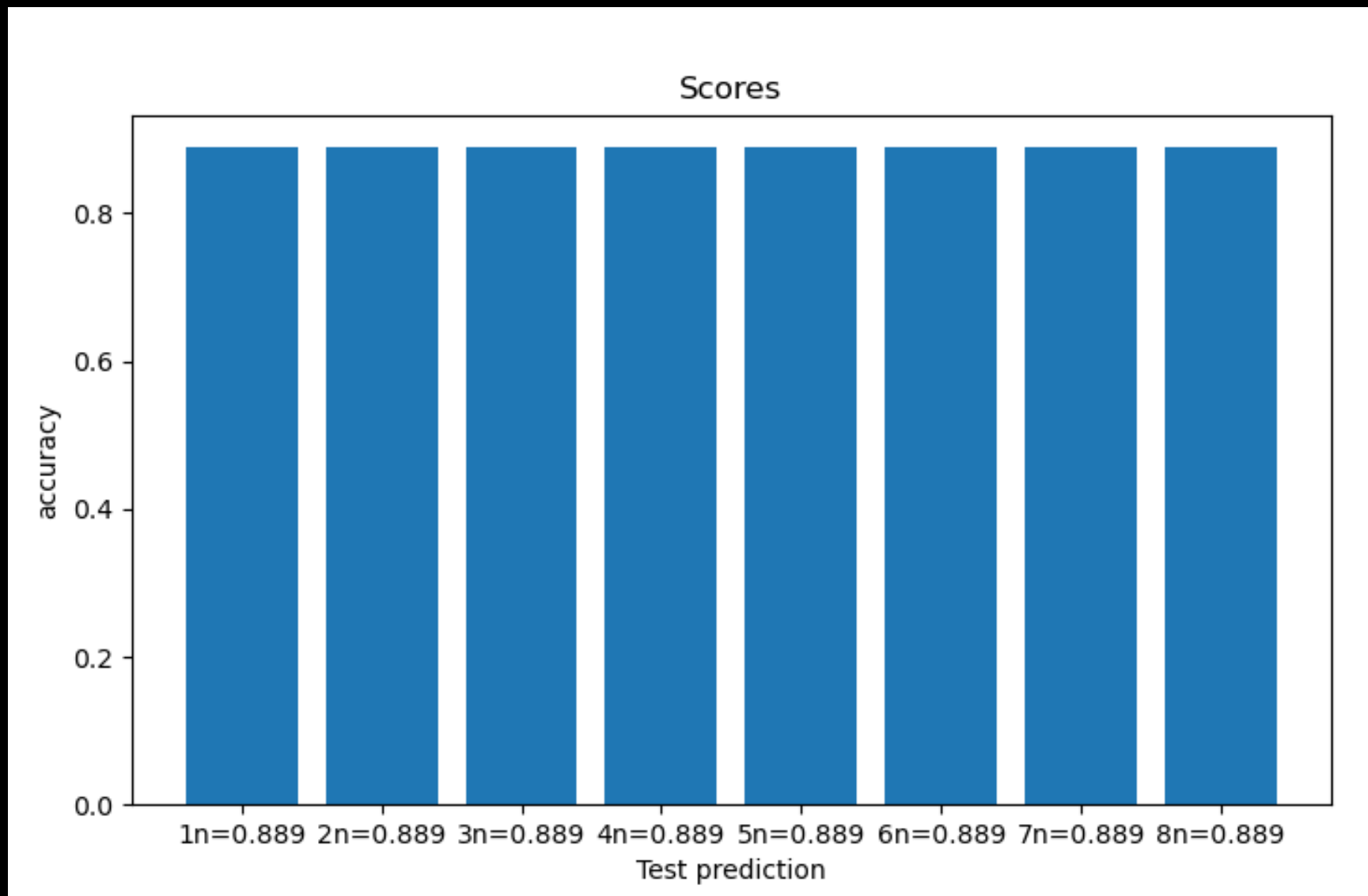
ADABOOST

16



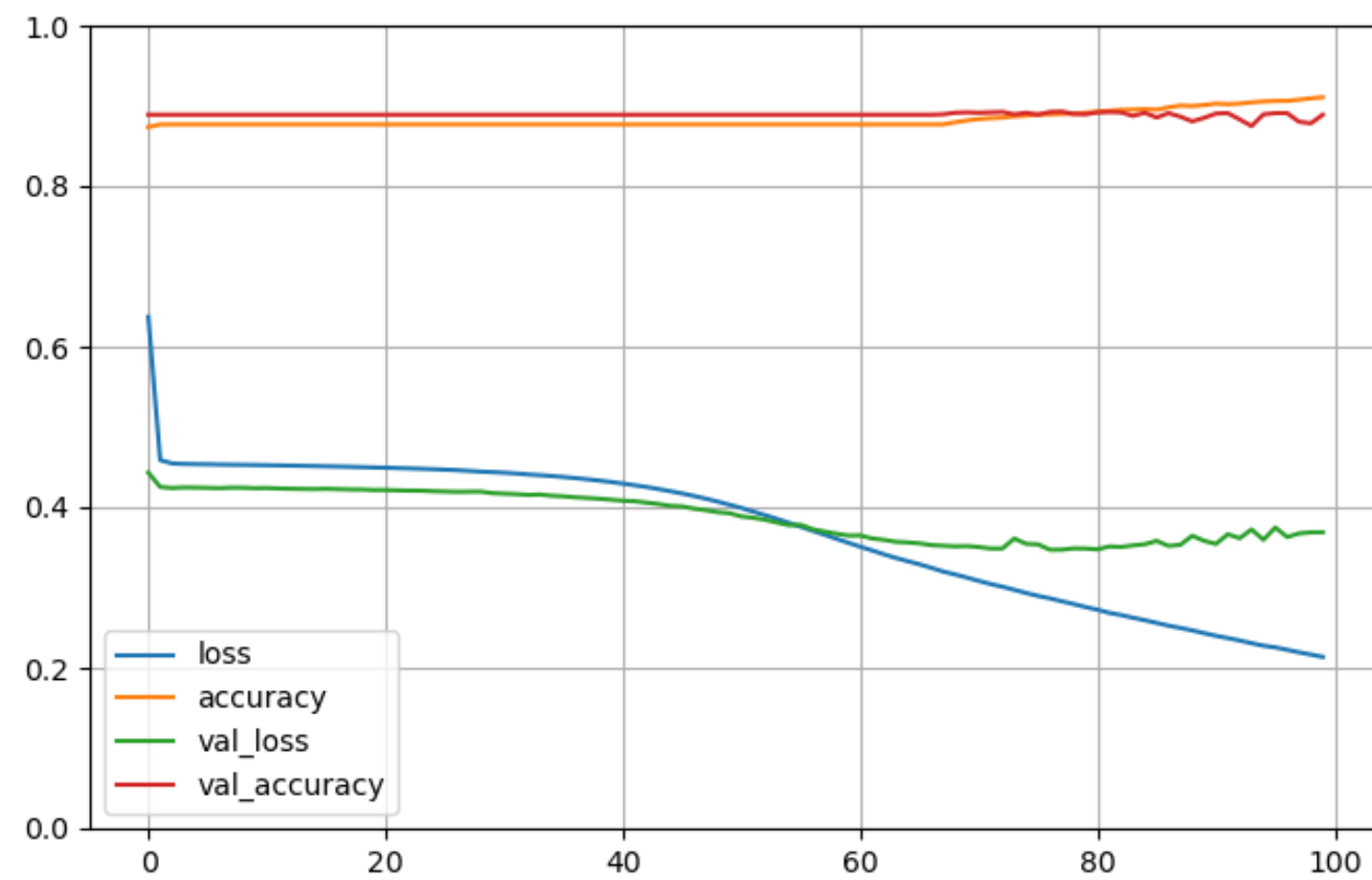
SVM

17



RETI NEURALI

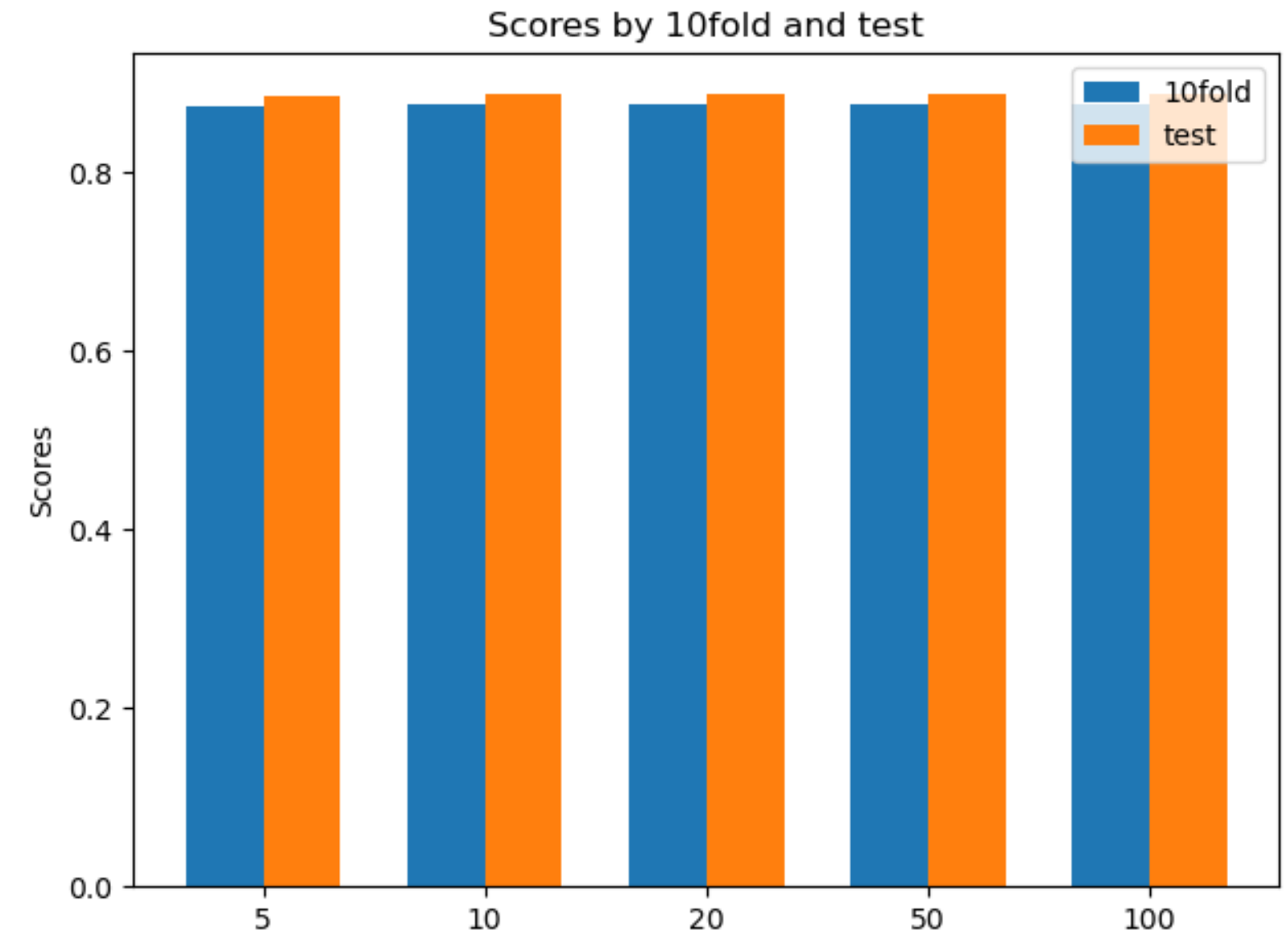
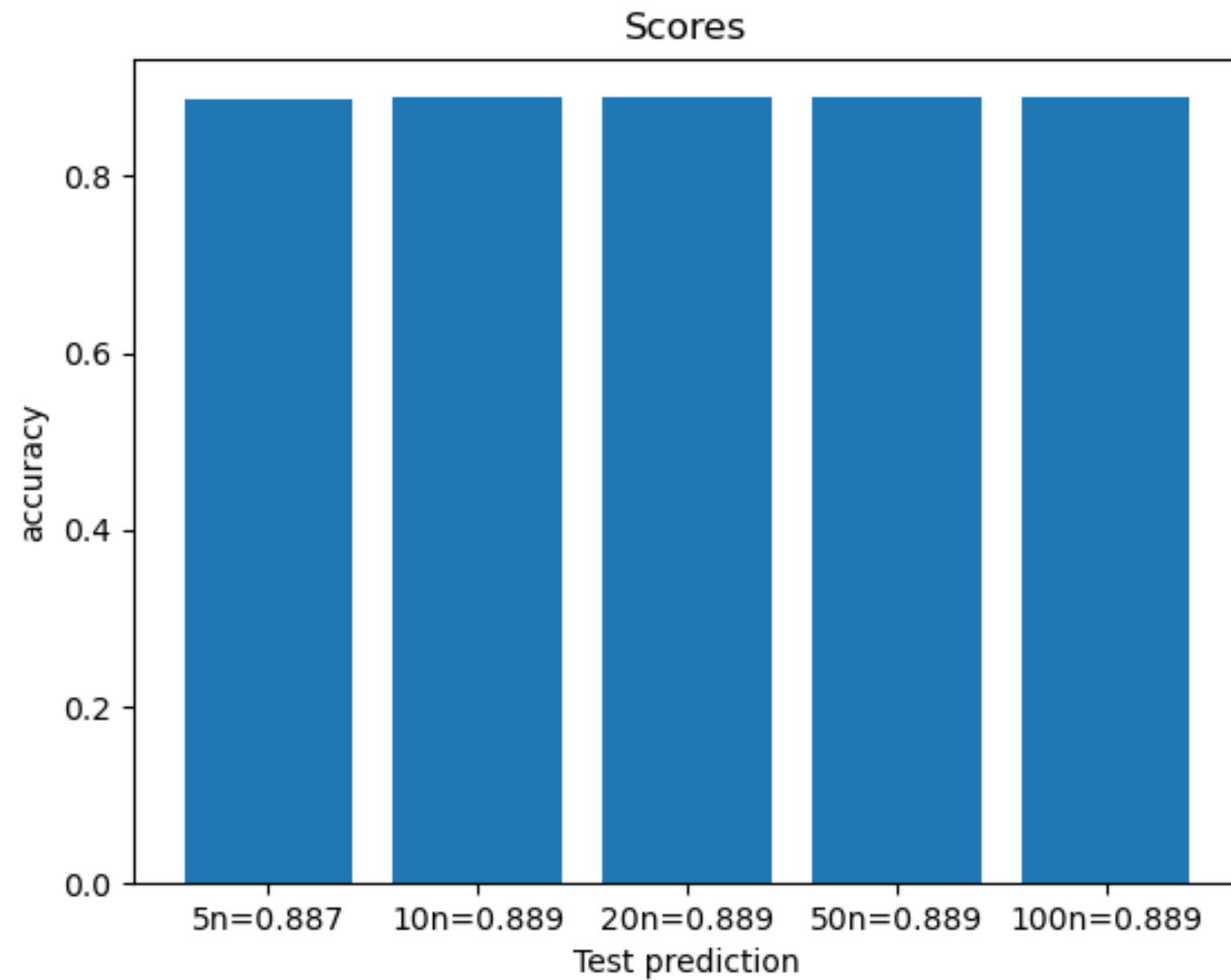
- RETE DENSA

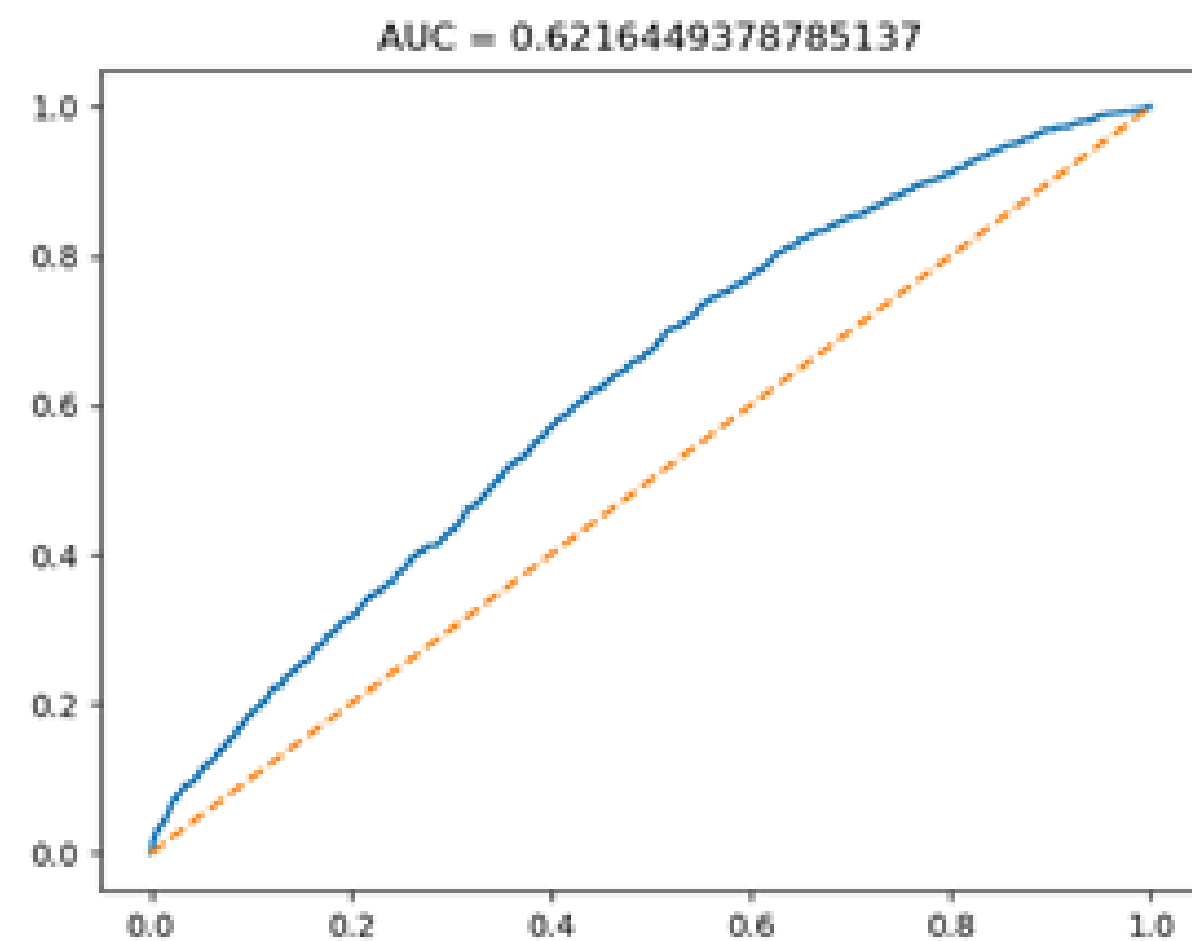
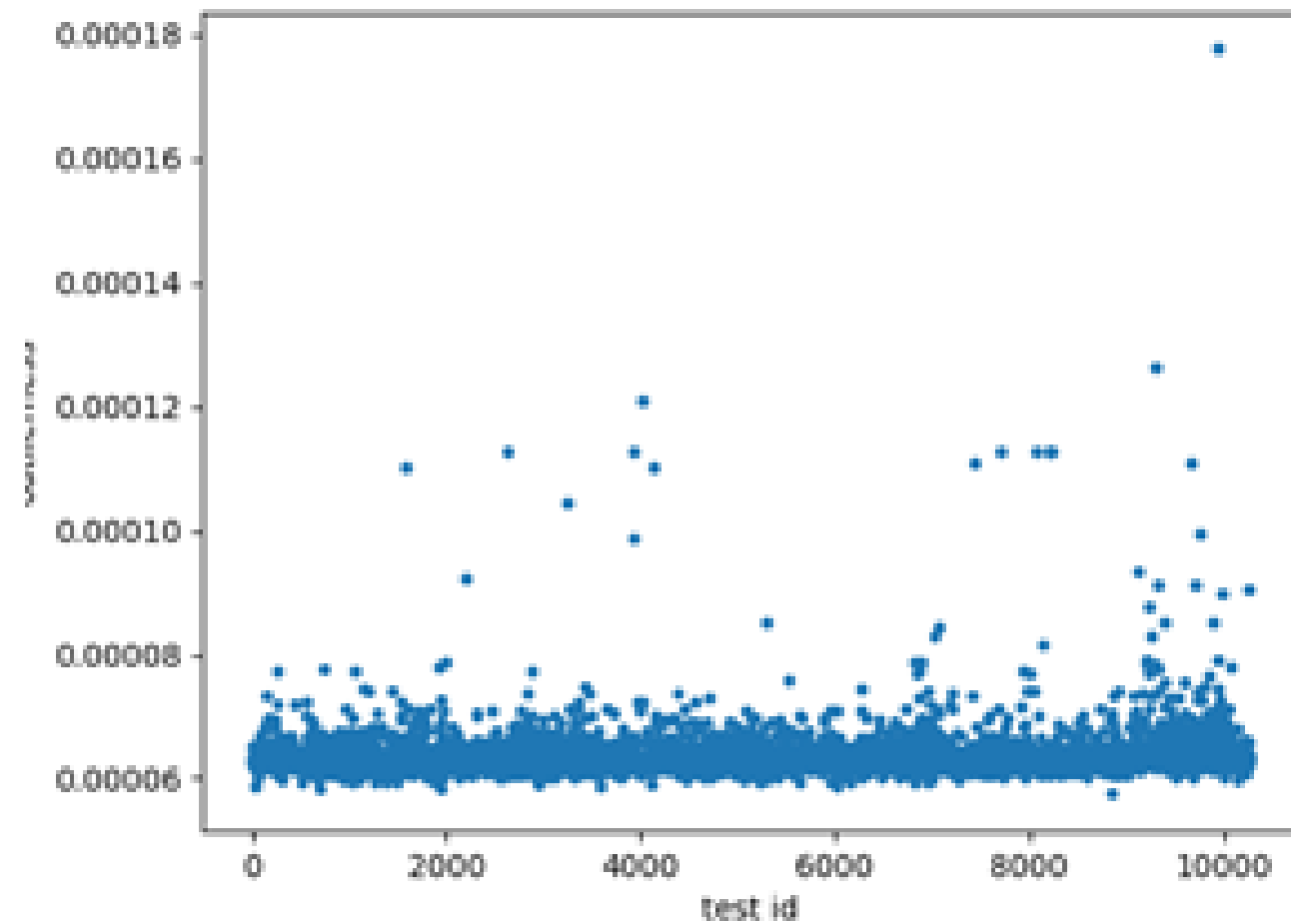


STIME DI DENSITÀ

19

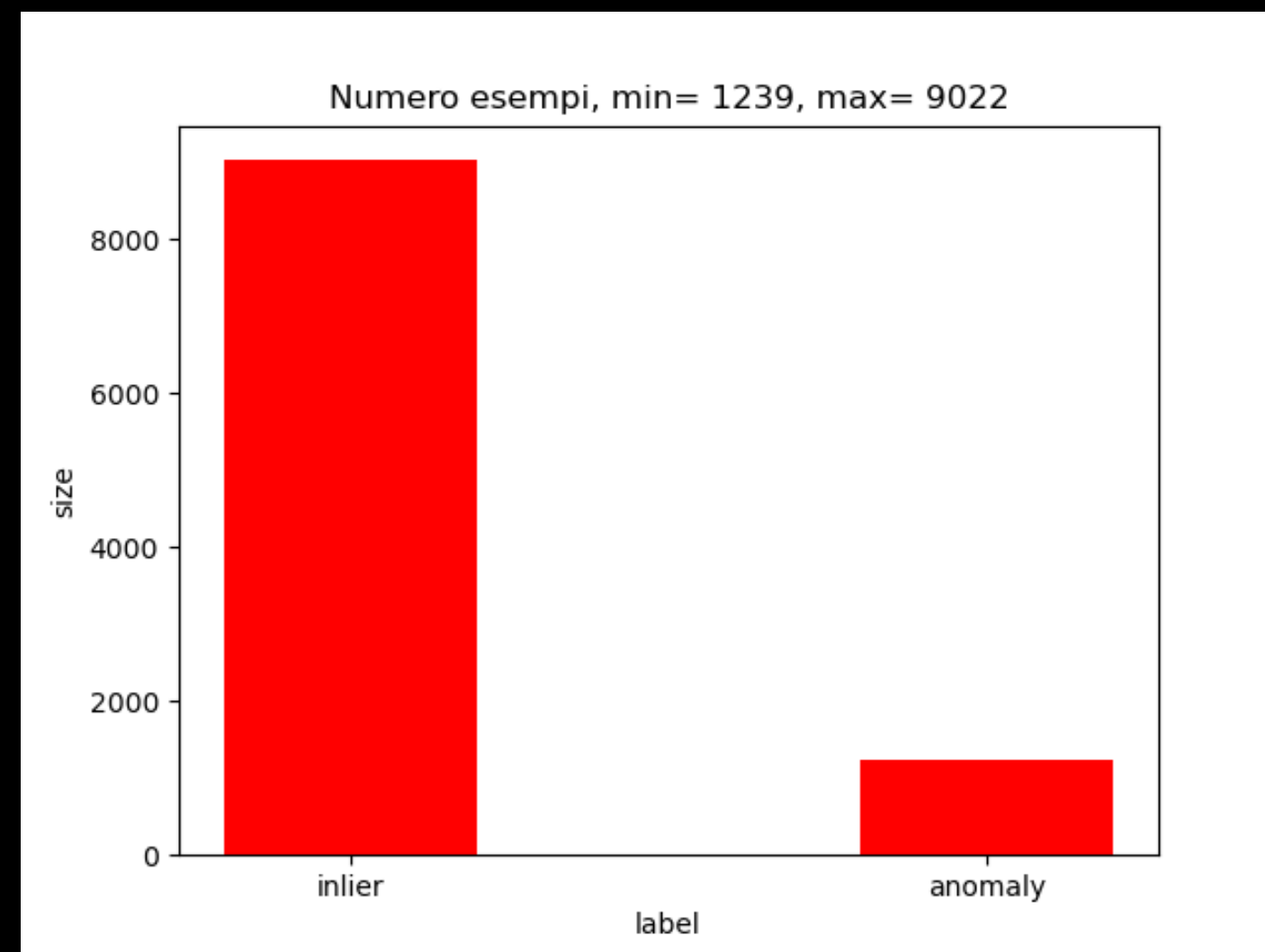
- K NEAREST NEIGHBOR





ANOMALY DETECTION

- RETE DENSA



ESEMPIO DI SCRIPT

21

- **python MAIN.py -i immagini-3 :**

```
Hai scelto image!

Che algoritmo vuoi lanciare?
Classificazione:
    * (1) AdaBoost
    * (2) SVM
    * Reti neurali:
        + (3) Neural network
        + (4) Convolutional Neural network
    * Stime di Densità:
        + (5) Gaussian Naive Bayes
        + (6) Nearest Neighbor

Anomaly Detection:
    * (7) Density autoencoder
```

- **python MAIN.py -t test-3.xlsx :**

```
Hai scelto text!

Che algoritmo vuoi lanciare?
Classificazione:
    * (1) AdaBoost
    * (2) SVM
    * Reti neurali:
        + (3) Neural network
    * Stime di Densità:
        + (4) Nearest Neighbor

Anomaly Detection:
    * (5) Density autoencoder
```

GRAZIE

PER

L'ATTENZIONE