

# Digital Signal & Image Management

---

Audio classification

Laurea Magistrale in  
**DATA**SCIENCE

Laura Nembrini Matr. 819059  
Gabriele Strano Matr. 866563  
Davide Mancino Matr. 847700

# DATASET

---



## 3 SPEAKER



## 1800 TRACCE VOCALI

È stata effettuata data augmentation su ogni singola traccia generando ulteriori

3 tracce:

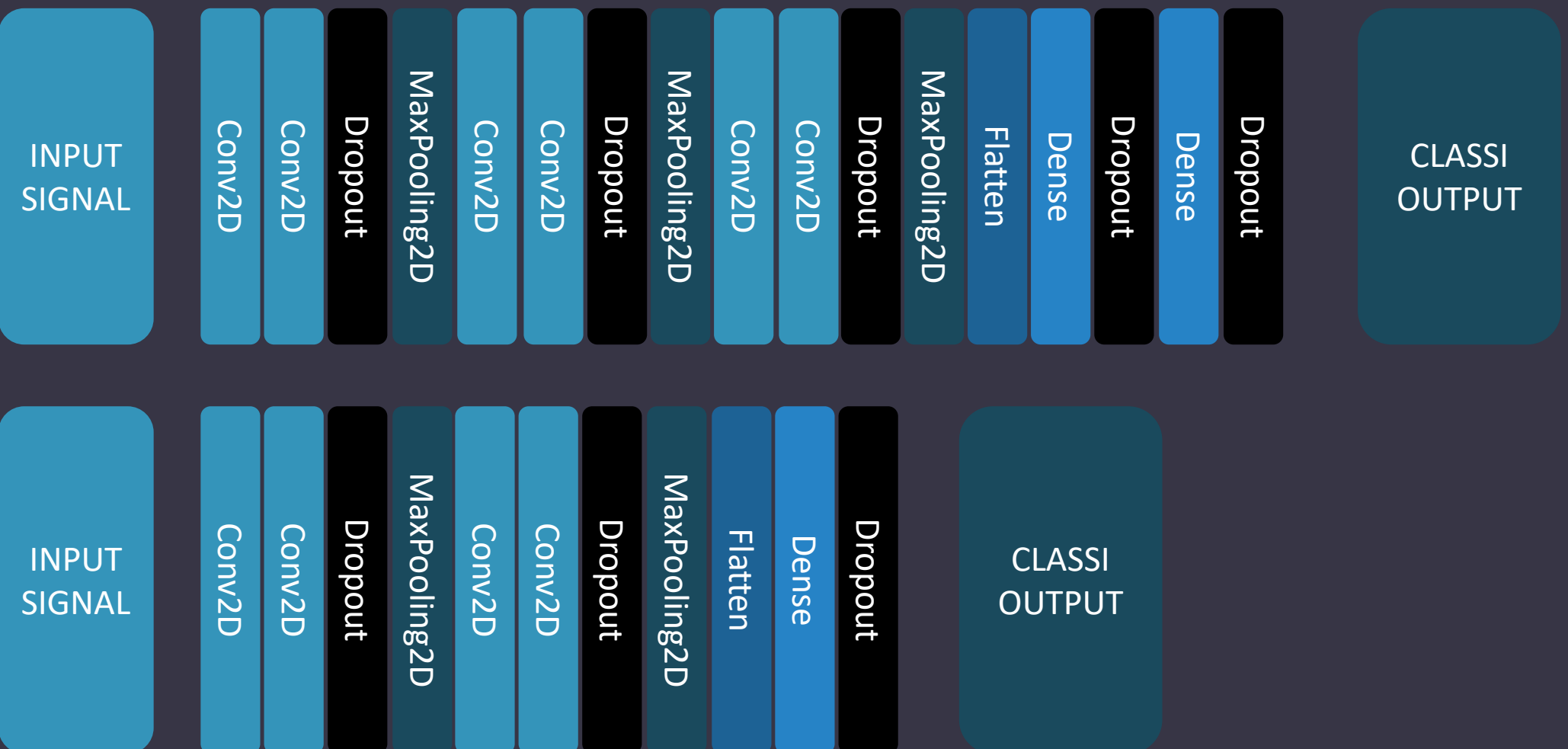
- rumore
- tono
- velocità



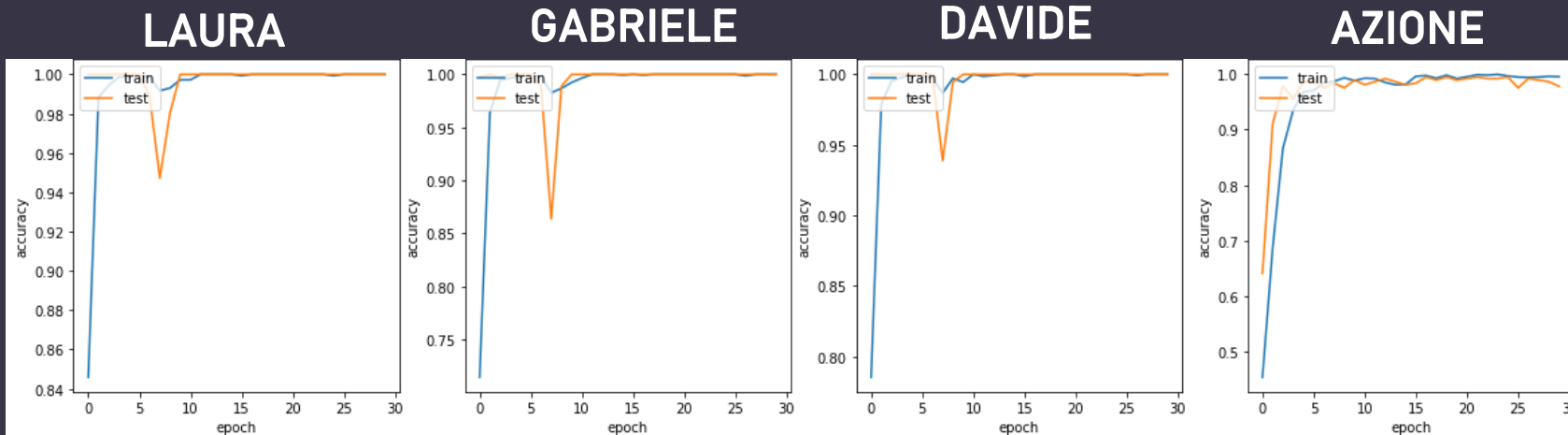
## TASKS

- Riconoscimento speaker
- Riconoscimento azione

# MODELLO

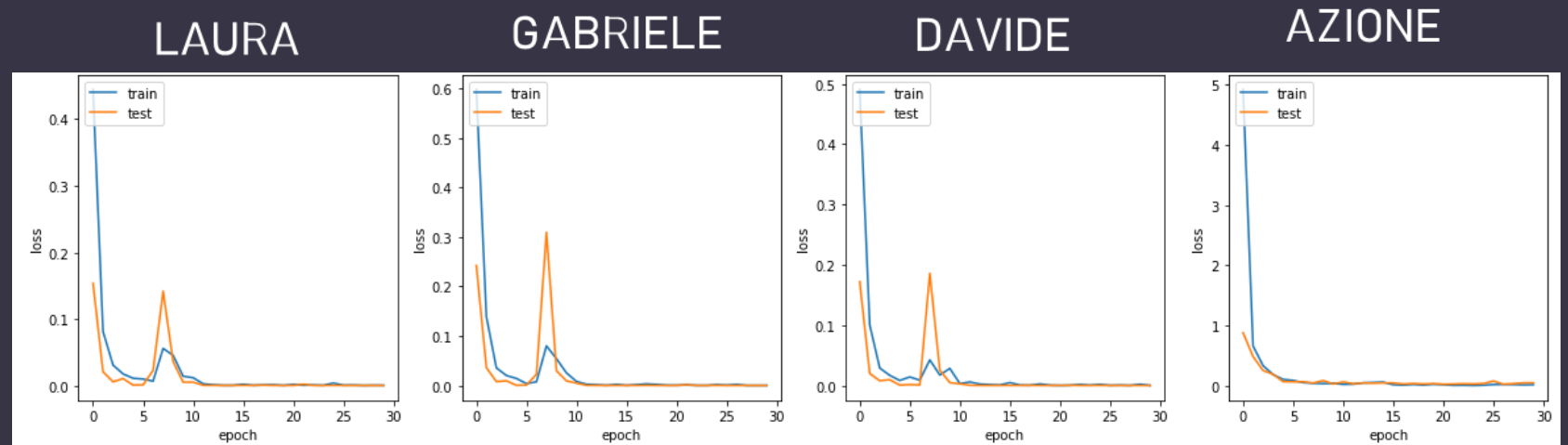


# RISULTATI



MODEL  
ACCURACY

MODEL  
LOSS



# DEMO

---



# Digital Signal & Image Management

---

Image classification

# DATASET

---



## 3 PERSONE

- Due volti maschili e un volto femminile
- Acquisizione di 30 foto per ciascun individuo



## CONDIZIONE DI ACQUISIZIONE

- Luminosità variabile
- Angolazioni differenti
- Background distinti
- Qualità differente dei dispositivi di acquisizione

# MODELLO

---

## FACE DETECTION

Utilizzo di *Haar cascades*  
fornita dalla libreria  
OpenCV

## PRE-PROCESSING

- Adattamento immagine a formato richiesto
- Trasformazione labels in one-hot encoding

## TRAIN e TEST SPLIT

Il dataset è stato diviso  
tra Train (70%) e Test  
(30%) con divisione equa  
tra i soggetti

## MOBILENETV2

- Caricamento rete pre-addestrata
- Data augmentation

## FINE TUNING

- 2 hidden layer fully-connected, funzione di attivazione reLU.
- output 3 neuroni e funzione di attivazione softmax.

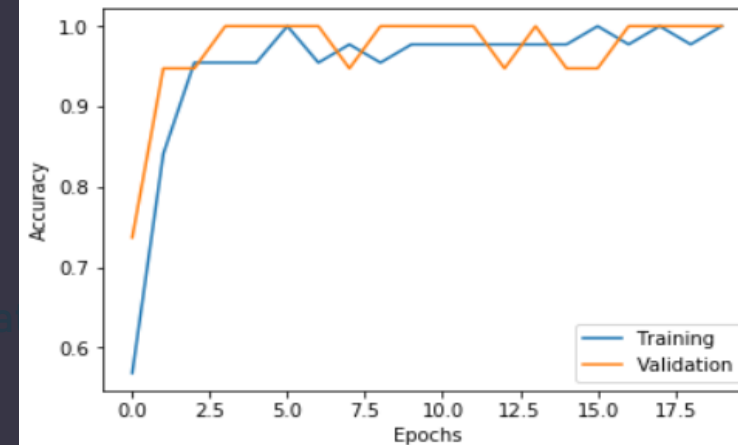
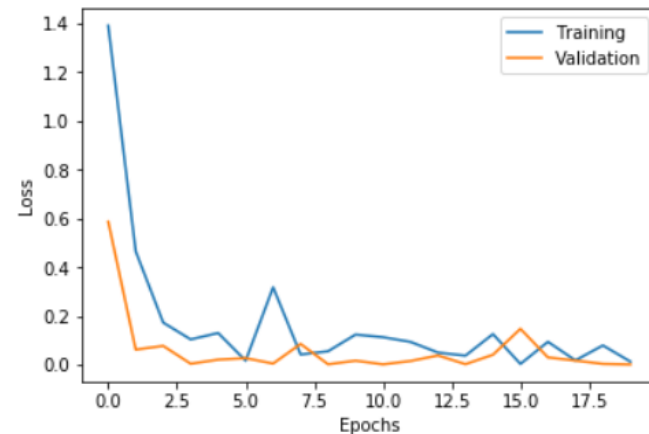
## TRAINING MODELLO

Addestramento modello  
con batch size= 4 e  
numero di epoche pari  
a 20



# RISULTATI

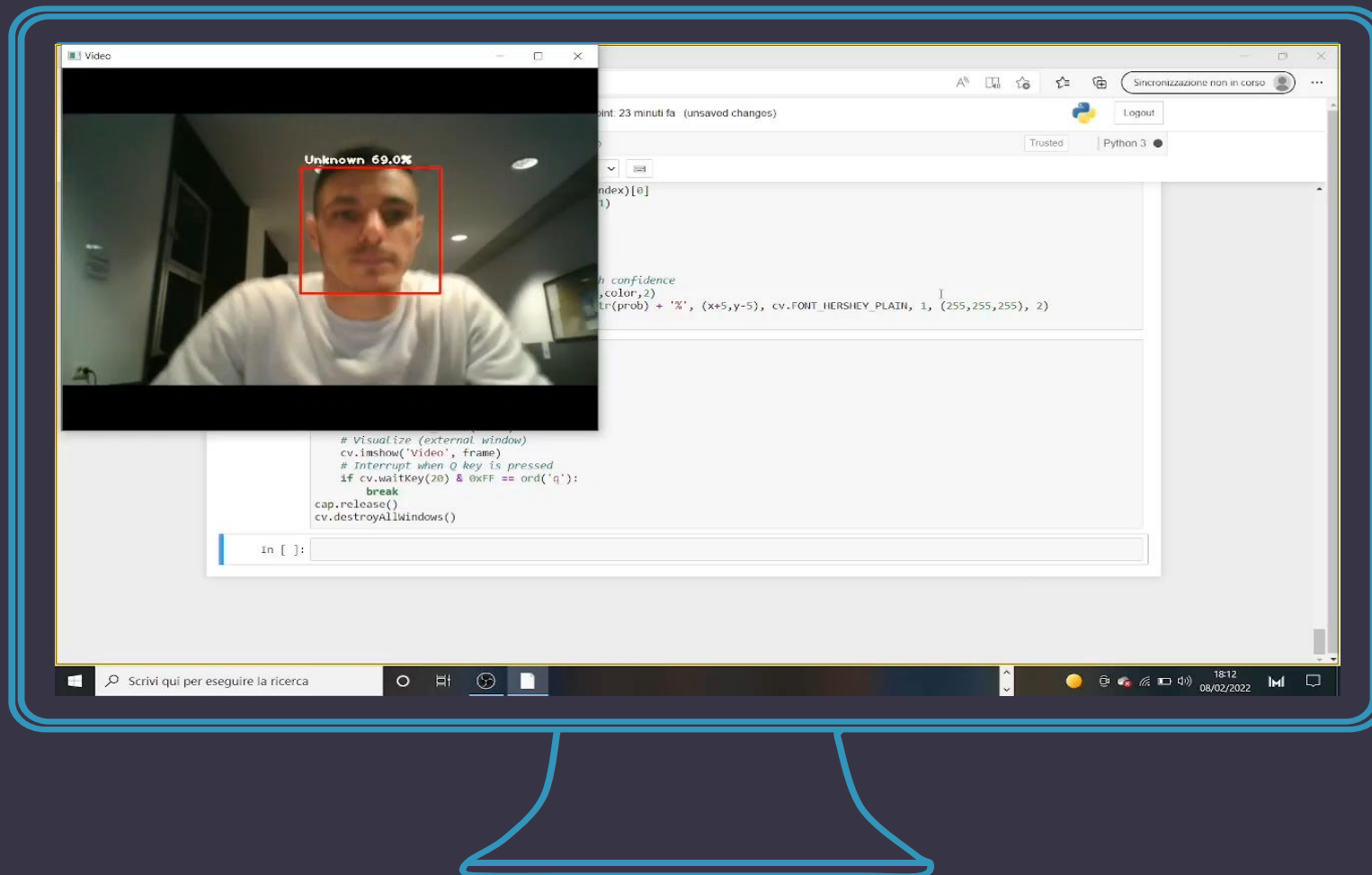
## TRAIN/VALIDATION HISTORY



è deciso di utilizza

	precision	recall	f1-score	support
DAVIDE	1.00	1.00	1.00	9
GABRIELE	1.00	1.00	1.00	9
LAURA	1.00	1.00	1.00	9
micro avg	1.00	1.00	1.00	27
macro avg	1.00	1.00	1.00	27
weighted avg	1.00	1.00	1.00	27

# DEMO



# **Digital Signal & Image Management**

---

**Image retrieval**

# DATASET

---



Il dataset di partenza, utilizzato per la scelta dell'individuo predefinito, è il medesimo utilizzato nel modello 2D



L'altro è un set di dati formato da circa 1580 personaggi famosi, ritratti in diversi momenti della loro vita. Il subset da noi utilizzato comprende 10 foto per ogni vip presente nel dataset per un totale di 15800 immagini.

# MODELLO PARTE 1

---

- L'obiettivo è quello di trovare dieci volti somiglianti alla foto di un individuo predefinito, analizzando un Dataset contenente volti di diverse persone.
- La rete scelta è una rete pre-addestrata VGGFace, basata sul modello ResNet50.
- La modalità di pooling è settata su «Max», sono stati esclusi i top layers e come Input Shape è stato scelto (512,512,3).
- Successivamente le features vengono estratte tramite la rete VGGFace.

# MODELLO PARTE 2

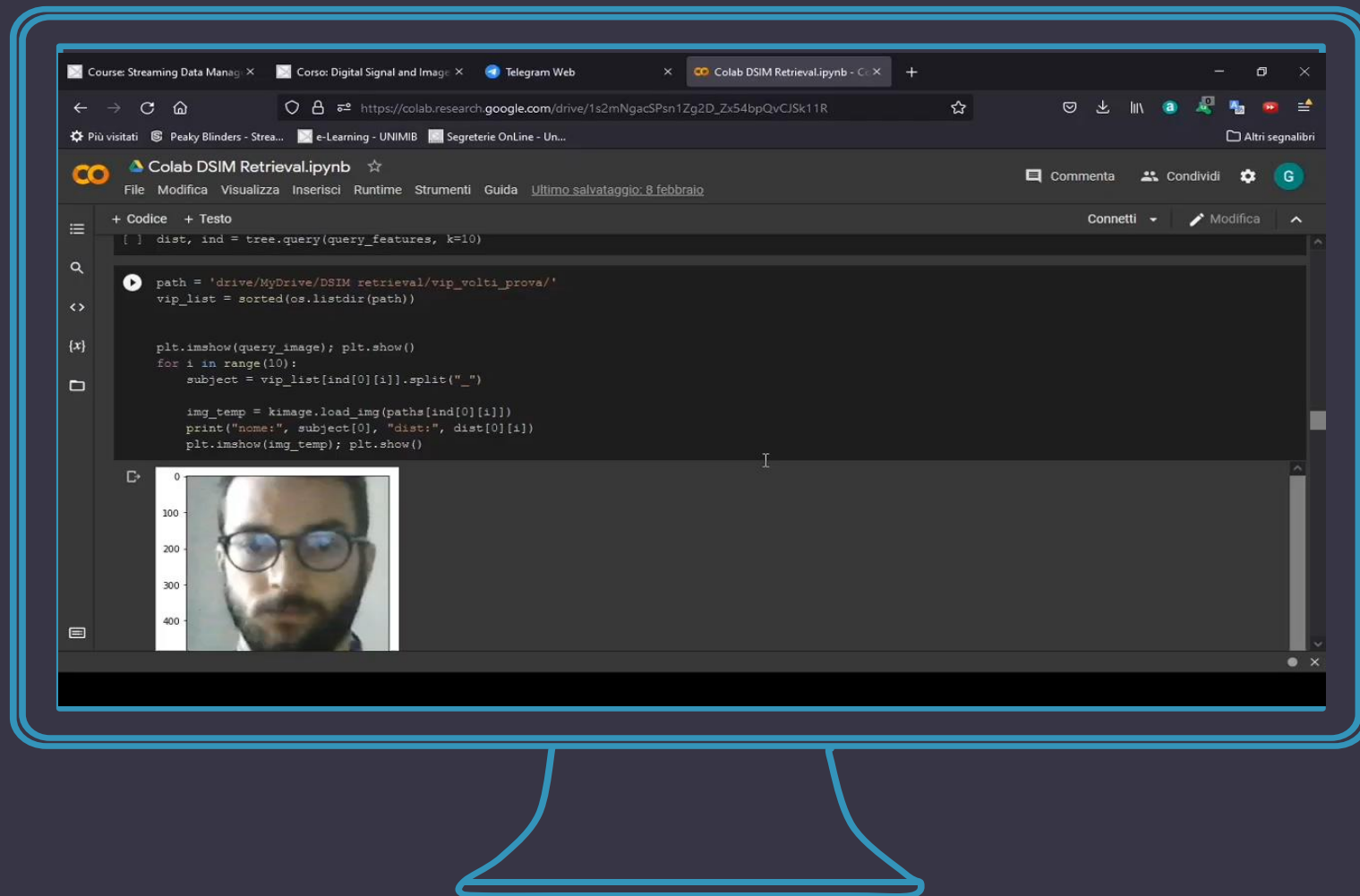
---

- Al fine di svolgere al meglio l'analisi, come modello per il retrieval delle immagini è stato usato l'algoritmo KDTree, ovvero un albero binario di ricerca multidimensionale che partiziona lo spazio per organizzare i punti in uno spazio k-dimensionale.

Si è deciso di utilizzare TF-IDF

- La metrica di riferimento scelta è la distanza euclidea. Infine, è stata utilizzato un ciclo for che data l'immagine di partenza, mostra i volti simili e la distanza.

# DEMO



The background is a dark navy blue. On the left side, there are several overlapping squares in various shades of blue and grey, some rotated 45 degrees. On the right side, a large white diamond is outlined by a thin grey border. Inside this diamond, the text "GRAZIE PER L'ATTENZIONE" is written in a bold, black, sans-serif font.

**GRAZIE PER  
L'ATTENZIONE**