



# DOCUMENTAZIONE FRAVIT

## INGEGNERIA DELLA CONOSCENZA 21-22

L'applicativo software è stato sviluppato dal gruppo composto da:

- Francesco Sasso ( <https://github.com/franklin2219> )
  - Matricola: 715742
  - Mail istituzionale: f.sasso16@studenti.uniba.it
- Vito Musco ( <https://github.com/VitoMusco> )
  - Matricola: 724569
  - Mail Istituzionale: v.musco1@studenti.uniba.it
- Repository progetto: <https://github.com/franklin2219/FRAVIT>

### Introduzione

L'assistente virtuale intelligente **\*\*FRAVIT\*\*** nasce con lo scopo di aiutare i negozi locali che si occupano della vendita al dettaglio di Film in DVD.

Esso aiuterà il negoziante nella scelta del giusto prezzo per la vendita di un Film, permetterà al negoziante di suggerire un determinato film (in base a dei criteri) ai clienti che ne fanno richiesta, gestirà la suddivisione dei film negli scaffali suddivisi per sezione (in base alle caratteristiche di ciascun Film) nel negozio.

Infine, utilizzando una base di conoscenza, FRAVIT permette di comprendere in quale sala sarà proiettato un determinato film, di comprendere il genere e la durata, e di visionare gli addetti alle proiezioni di quella determinata sala.

Per la gestione del negozio di Film abbiamo scelto di utilizzare un dataset preesistente in rete, in modo tale da poter utilizzare un ammontare di dati sufficienti per ottenere gli obiettivi prestabiliti.

Sono state effettuate delle modifiche a tale dataset come l'inserimento delle colonne "voto\_convertito", "durata\_convertita", "anno\_convertito", "genere\_convertito" utilizzate per fornire degli esempi di prezzi dei film presenti nel dataset e per effettuare il suggerimento del film al cliente.

Per la gestione delle proiezioni dei Film nelle sale abbiamo creato una piccola base di conoscenza contenente un piccolo numero di Film suddivisi per genere, durata e sala i quali verranno interrogati sfruttando le regole prestabilite in essa.

Il nome dell'applicativo è composto utilizzando le iniziali del nome dei membri del gruppo.

## Requisiti funzionali

Avendo progettato tale applicativo in Python si richiede un ambiente in grado di eseguire codice come Eclipse, IntelliJ, Visual Studio Code ...

Le librerie importate all'interno del progetto sono:

- **'numpy'** tramite il comando da terminale ```pip install numpy```;
- **'pandas'** tramite il comando da terminale ```pip install pandas```;
- **'sklearn'** tramite il comando da terminale ```pip install scikit-learn``` per implementare i classificatori utilizzati;
- **'matplotlib'** tramite il comando da terminale ```pip install -U matplotlib```;
- **'pytholog'** tramite il comando da terminale ```pip install pytholog```;

## Manuale Utente

*La prima volta che si utilizza FRAVIT la compilazione potrebbe richiedere qualche secondo in più.*

Quando il programma viene avviato si visualizzerà la schermata principale a linea di comando:

```

+++++
|                                |
|          GESTORE FILM E SERIE TV          |
|                                |
+++++
|          1) Predici Prezzo          |
|          2) Suggerisci film          |
|          3) Trova Sezione Negozio    |
|          4) Precision,Recall,Mae,Accuratezza    |
|          5) Effettua query sui film in proiezione    |
|          6) Esci                      |
|                                |
+++++

```

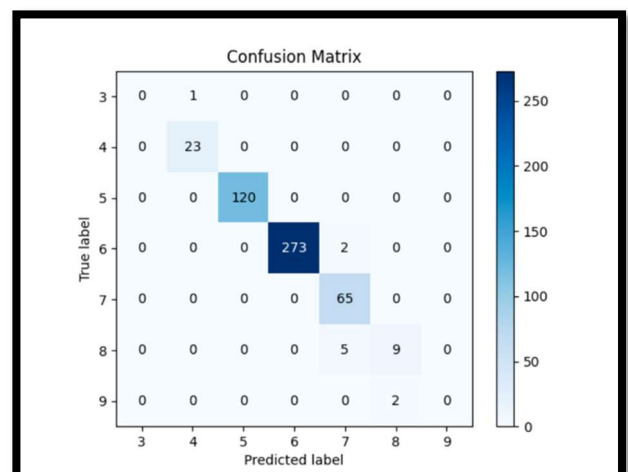
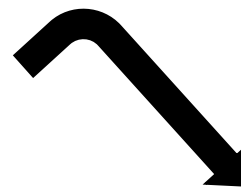
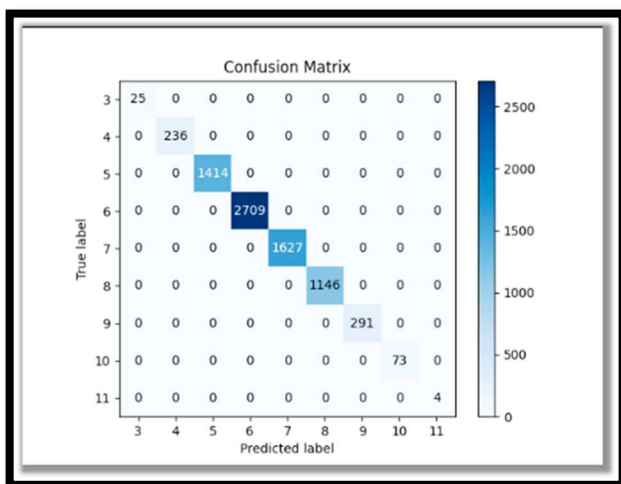
I comandi disponibili all'interno del menù sono i seguenti:

- **``Predici Prezzo``** → comando che viene suggerito dal banner iniziale, e che se invocato permette di inserire i dati di un nuovo Film che dovranno essere inseriti nel negozio in modo tale da comprendere quale sarà il loro prezzo.

- **``Suggerisci Film``** → comando che viene suggerito dal banner iniziale, e che se invocato permette di inserire i dati relativi ai gusti del cliente permettendo a FRAVIT di suggerire uno tra i film presenti nel dataset e che rispetteranno i gusti forniti.
- **``Trova Sezione Negozio``** → comando che viene suggerito dal banner iniziale, e che permette di inserire i dati relativi al Film di cui si vuole comprendere la sezione e si otterrà la sezione nella quale quel determinato Film deve essere posto.
- **``Precision, Recall, Mae, Accuratezza``** → comando che viene suggerito dal banner iniziale, e che permette di visionare i dati relativi alla precision, alla recall, al Mae ed all'accuratezza sul train/test effettuato nel sistema.
- **``Query su film in proiezione``** → comando che viene suggerito dal banner iniziale, e che permette di interrogare la KB per comprendere la sala dove verrà proiettato un determinato film, il suo genere e la durata, e quali sono i proiezionisti prestabiliti per una determinata sala.
- **``Esci``** → comando per terminare l'esecuzione dell'applicazione.

## Scelte Progettuali

Prima di passare direttamente all'apprendimento supervisionato abbiamo modificato il dataset scelto in precedenza effettuando una *feature selection* in modo tale da ottenere le *feature obiettivo* prestabilite rappresentando i cambiamenti mediante l'utilizzo di una *confusion matrix*.



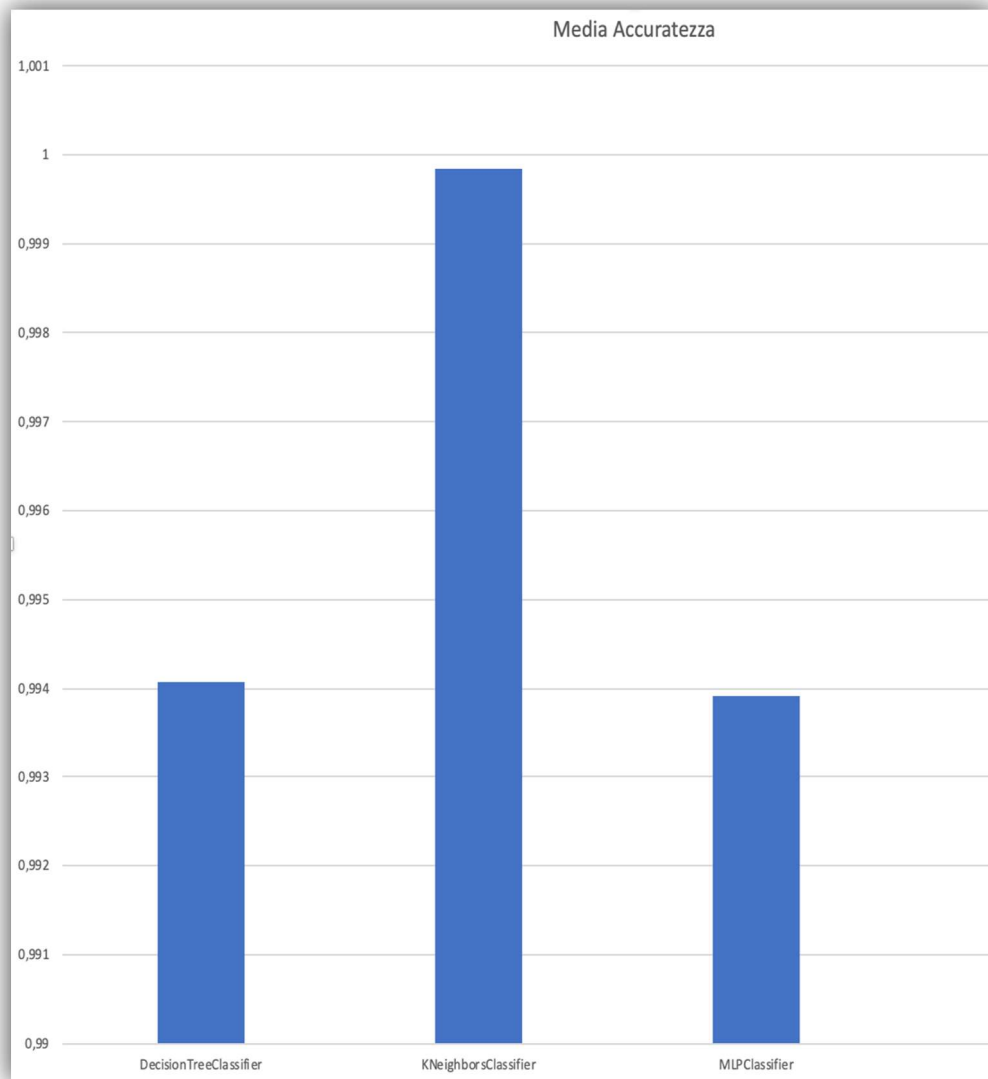
Calcolando l'accuratezza del sistema prima e dopo l'eliminazione delle features abbiamo notato una piccola variazione (circa 0.02), ma che abbiamo pensato fosse accettabile.

- **Dataset** → contiene più di 3000 Film alla quale però sono state aggiunte informazioni necessarie per l'utilizzo dell'applicazione come:
  - `genere\_convertito` utilizzato per la conversione del genere di un Film in valori numerici utili agli scopi progettuali (da 1 a 27).
  - `voto\_convertito` utilizzato per rappresentare il voto fornito dall'utente in un valore numerico utile agli scopi progettuali (da 1 a 4).
  - `durata\_convertita` utilizzata per rappresentare la durata del Film con dei valori numerici utili agli scopi progettuali (da 1 a 3).
  - `anno\_convertito` utilizzato per rappresentare l'anno di pubblicazione del Film con dei valori numerici utili agli scopi progettuali (da 1 a 4).
- **Un K-nearest\_neighbors** → In grado di prevedere il prezzo di un nuovo film immesso nel sistema in base a degli esempi di film contenuti nel dataset utilizzato.  
 Per decidere quale classificatore utilizzare per la predizione della feature obiettivo non ci siamo affidati al caso ma abbiamo visionato l'accuratezza riscontrata in 5 esecuzioni del sistema, ognuna delle quali utilizzando una K-Fold cross validation, mostrando la media riscontrata tra le precisioni calcolate.

	CLASSIFICATORE	PRECISION	ACCURATEZZA
MEDIA	MLPClassifier	0,9956	0,9939
	KNeighborsClassifier	1	0,9998
	DecisionTreeClassifier	1	0,9940

Abbiamo deciso di utilizzare il KNN come classificatore per la predizione della feature obiettivo. Vista anche l'accuratezza riscontrata nelle suddette prove che possiamo visionare nell'immagine precedente.

Possiamo visionare le prestazioni dei classificatori testati per la feature nel modello che segue.

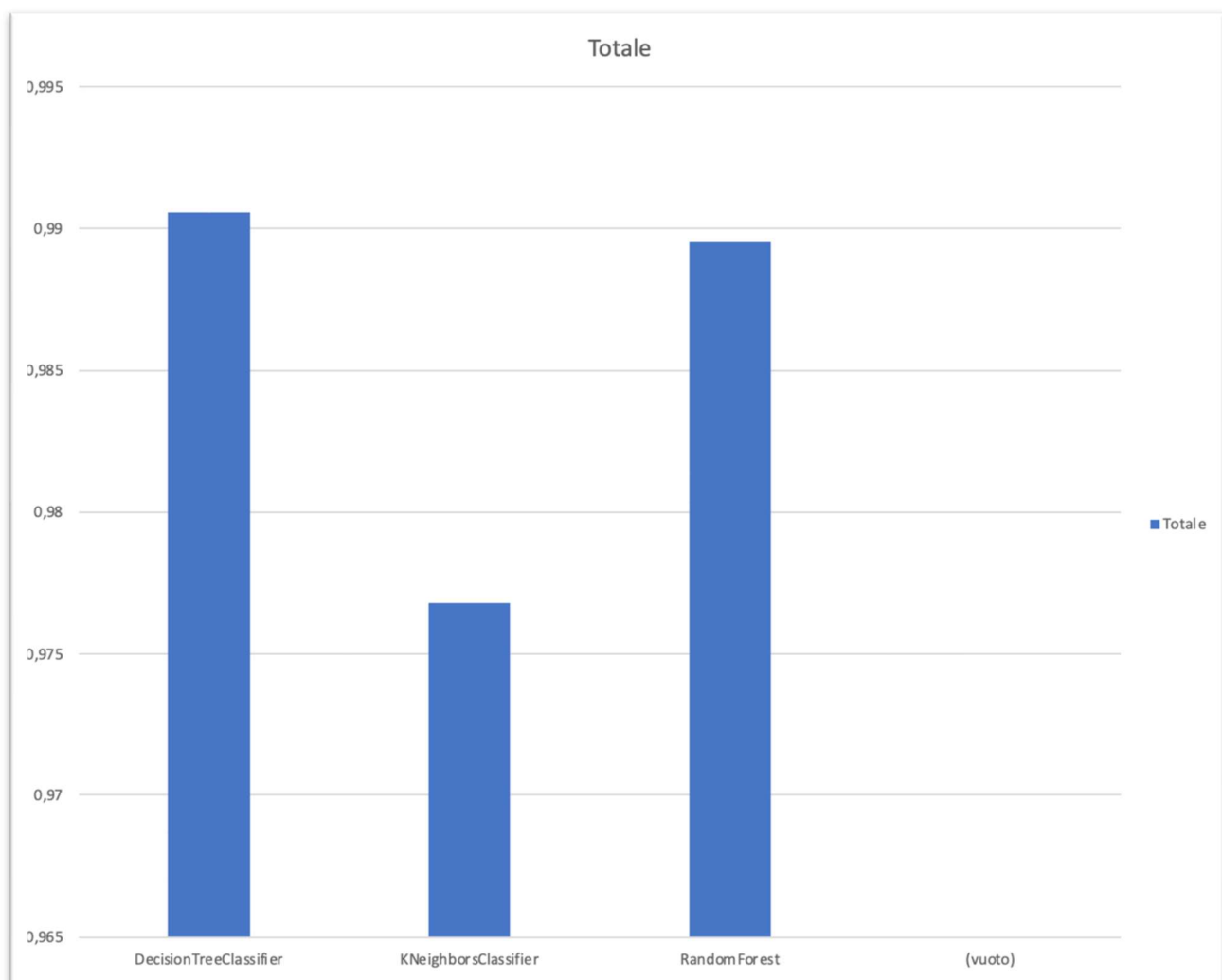


- **K-fold cross validation** → Come strumento per misurare l'accuratezza del sistema, ci siamo avvalsi della K-Fold cross-validation, in particolare usavamo 10 fold.
- **Precision e Recall** → Abbiamo implementato anche una funzione che permette di visionare la precisione del sistema, nell'esecuzione corrente.
- **Mae** → così come per l'accuratezza, abbiamo pensato potesse tornare utile tenere conto dell'errore che il sistema presenta, quindi sempre mediante un'apposita funzione, è possibile osservare il "Mean Absolute Error", cioè l'errore assoluto riscontrato tra le features predette e quelle reali sia in fase di training che di test.

```
· MAE sul test set : 0.003501750875437719
· MAE sul train set : 0.002
```

- **Un K-nearest\_neighbors** → in grado di suggerire un determinato film o una serie TV al cliente in base alle informazioni fornite in input che rappresenteranno le proprietà del film che si vorrà vedere.  
Verrà richiesto il genere di Film che si intende vedere, il periodo cinematografico che si preferisce ed i livelli di humor, tensione, ritmo e impegno richiesti.
- **Un albero di decisione** → in grado di far comprendere in quale sezione del negozio dovrà essere posto un determinato Film in base alle caratteristiche inserite in input.
  - sezione 1 → Film con maggior humor.
  - sezione 2 → Film con maggior ritmo.
  - sezione 3 → Film con maggior impegno.
  - sezione 4 → Film con maggior tensione.

MEDIA	CLASSIFICATORE	PRECISION	ACCURATEZZA
	Random Forest	0,98	0,9895
	KNeighborsClassifier	0,98	0,9767
	DecisionTreeClassifier	0,986	0,9905



Per decidere quale classificatore utilizzare per la predizione della feature obiettivo abbiamo visionato l'accuratezza riscontrata in 5 esecuzioni del sistema, ognuna delle quali utilizzando una K-Fold cross validation, mostrando la media riscontrata tra le precisioni calcolate.

Come possiamo vedere seppur con una piccola differenza, risulta con un'accuratezza superiore il DecisionTreeClassifier.

- **Prolog** → Ci siamo voluti cimentare con il Prolog, effettuando una semplice KB composta da asserzioni che specificano le caratteristiche di un determinato Film come la sala di proiezione, il genere e la durata, ed altre che descrivono informazioni inerenti agli addetti alla proiezione.  
Come possiamo evincere dalla foto successiva:

```
:- disjointuous(i_film/4).
:- disjointuous (i_sala/3).
:- disjointuous (i_personale/3).
%-----
%Regole Film

%Mostra il titolo del film in base al genere e viceversa
film(Titolo, I_film1) :- i_film(Titolo,I_film1,_,_).
%Mostra la durata del film in base al Titolo e viceversa
durata(Film,Durata) :- i_film(Film,_,_,Durata).
%Mostra il titolo del film in base alla sala selezionata e viceversa
sala(Titolo,Sala) :- i_film(Titolo,_,Sala,_).
%Mostra l'addetto alla proiezione di una determinata sala
addetto(Sala,Addetto) :- i_sala(Sala,Addetto,_).

%-----
%Regole Dipendenti Sale
|
%Mostriamo i proiezionisti per quella sala
%-----

%DATI

%Immissione film

i_film(dumbo,animazione,sala1,90).
i_film(abissi,avventura,sala2,180).
i_film(assassination,azione,sala1,100).
i_film(ben_Hur,biblico,sala3,90).
i_film(bird,biografico,sala4,60).
i_film(la_crociera_del_terrore,catastrofico,sala1,130).
i_film(abbronzatissimi,commedia,sala3,90).
i_film(cineocchio,documentario,sala2,180).
i_film(accattone, drammatico,sala3,90).
i_film(alien, fantasy,sala4,100).
i_film(borsalino, gangster,sala3,60).
i_film(amarcord, grottesco,sala2,90).
i_film(bataan, guerra,sala1,180).
i_film(dracula, horror,sala4,90).
i_film(gli_argonauti, mitologico,sala3,70).

%Informazioni Addetti Proiezioni film

i_sala(sala1,vito,proiezionista).
i_personale(sala1,vito,altamura,30-09-2000).
i_sala(sala2,michele,proiezionista).
i_personale(michele,gravina,19-05-1998).
```

Il sistema permette di scegliere tra le seguenti opzioni:

```
print("+++++")
print("|      1) Trova genere film      |")
print("|      2) Trova durata film      |")
print("|      3) Mostra film in proiezione in una sala      |")
print("|      4) Mostra addetto alla proiezione di una sala      |")
print("|      5) Trova film di un determinato genere      |")
print("|      6) Trova film di una determinata durata      |")
print("|      7) Mostra in quale sala è proiettato un film      |")
print("|      8) Mostra in quale sala è assegnato un addetto alla proiezione      |")
print("|      9) Torna indietro      |")
print("+++++")
```

## **Dataset**

Dataset contenente informazioni relative ai film usciti durante gli anni.

Contenente le seguenti colonne:

- filmtv\_id;
- titolo\_originale;
- titolo\_italiano;
- anno;
- genere;
- durata;
- voto;
- humor;
- ritmo;
- impegno;
- tensione;
- prezzo;
- anno\_convertito;
- durata\_convertita;
- voto\_convertito;
- genere\_convertito;
- sezione.



dataset_completo																
filmtv_id	titolo_originale	titolo_italiano	anno	genere	durata	voto	humor	ritmo	impegno	tensione	prezzo	anno_convertito	durata_convertita	voto_convertito	genere_convertito	sezione
2	Bugs Bunny's Third Movie: 1001 Rabbit Tales	Le 1001 favole di Bugs Bunny	1982	Animazione	76	7,7	3	3	0	0	5	1	1	3	1	1
3	18 anni tra una settimana	18 anni tra una settimana	1991	Drammatico	98	6,5	0	2	0	2	6	1	2	3	2	1
17	Ride a Wild Pony	A cavallo di un pony selvaggio	1978	Sentimentale	91	5,6	1	2	1	0	6	1	2	3	3	2
18	Diner	A cena con gli amici	1982	Commedia	95	7	2	2	0	1	6	1	2	3	4	1
20	A che servono questi quattrini?	A che servono questi quattrini?	1942	Commedia	85	5,9	3	1	1	0	5	1	1	3	4	1
22	A ciascuno il suo	A ciascuno il suo	1967	Drammatico	93	7,7	0	2	3	3	6	1	2	3	2	3
23	Dead-Bang	Dead Bang - A colpo sicuro	1989	Poliziesco	109	6	0	2	0	2	6	1	2	3	5	1
24	A... come assassino	A... come assassino	1966	Thriller	80	5,5	1	2	0	1	5	1	1	3	6	2
26	At Close Range	A distanza ravvicinata	1986	Drammatico	115	7,4	0	2	1	2	6	1	2	3	2	2
30	A Ghentar si muore facile	A Ghentar si muore facile	1968	Avventura	101	4,8	0	2	0	1	5	1	2	2	7	2
32	Sleeping with the Enemy	A letto con il nemico	1990	Drammatico	96	5,1	1	3	2	3	6	1	2	3	2	2
34	In Bed With Madonna	A letto con Madonna	1990	Documentario	111	5,3	1	3	0	0	6	1	2	3	8	2
36	Bowery at Midnight	A mezzanotte come il terrore	1942	Horror	62	5,3	0	2	1	3	5	1	1	3	9	4
37	A mezzanotte va la ronda del piacere	A mezzanotte va la ronda del piacere	1975	Commedia	100	5,9	2	2	0	0	6	1	2	3	4	1
38	Mr. Majestyk	A muso duro	1974	Azione	105	6,2	0	3	2	2	6	1	2	3	10	2
45	Warning Sign	A pochi secondi dalla fine	1985	Catastrofico	99	4,8	0	2	1	1	5	1	2	2	11	2
47	About Last Night	A proposito della notte scorsa...	1986	Commedia	113	5,9	2	3	1	1	6	1	2	3	4	2
49	Fail-Safe	A prova di errore	1964	Drammatico	110	8,2	0	3	3	3	7	1	2	4	2	2
51	Some Like It Hot	A qualcuno piace caldo	1959	Commedia	120	9,1	4	3	2	2	8	1	3	4	4	1
53	A qualsiasi prezzo	A qualsiasi prezzo	1968	Avventura	94	3,7	1	2	0	1	5	1	2	2	7	2
54	Johnny O'Clock	A sangue freddo	1947	Thriller	95	7,7	1	3	1	3	6	1	2	3	6	1
55	Back to School	A scuola con papà	1986	Commedia	100	6,5	3	2	1	0	6	1	2	3	4	1
58	Driving Miss Daisy	A spasso con Daisy	1989	Commedia	100	7	2	2	3	1	6	1	2	3	4	3
59	Enid Is Sleeping	A spasso con la morta	1989	Commedia	102	7	3	2	0	1	6	1	2	3	4	1
61	The Appaloosa	A Sud Ovest di Sonora	1966	Western	98	6,9	0	2	1	3	6	1	2	3	12	4
63	...a tutte le auto della polizia...	...a tutte le auto della polizia...	1975	Poliziesco	100	5,4	0	2	0	2	6	1	2	3	5	1
67	You'll Like My Mother	A un passo dalla morte	1973	Thriller	94	7,8	0	3	1	3	6	1	2	3	6	2
68	Abbasso la miseria!	Abbasso la miseria!	1945	Commedia	90	6	2	2	0	0	6	1	2	3	4	1
69	Abbasso la ricchezza!	Abbasso la ricchezza!	1946	Commedia	93	5,6	2	2	1	0	6	1	2	3	4	1
71	Charlie Muffin	L'abbraccio dell'orso	1979	Spionaggio	109	6,2	1	2	1	3	6	1	2	3	13	4
73	Abbronzatissimi	Abbronzatissimi	1991	Commedia	114	3,2	2	1	0	0	5	1	2	2	4	1
74	The Deep	Abissi	1977	Avventura	130	5,2	1	2	1	2	7	1	3	3	7	1

## Previsione Prezzo

Il programma è stato dotato di una funzione che sfrutta un classificatore KNN per determinare il prezzo a cui apparterrà un determinato Film inserito in input. Abbiamo utilizzato un file CSV per addestrare il nostro classificatore. Le informazioni che l'utente deve inserire sono:

- **Anno:** ovvero l'anno di pubblicazione del Film;
- **Durata:** ovvero la durata del Film espressa in minuti.
- **Voto:** ovvero il voto della critica a tale Film.

Ad ogni campo sono stati associati i seguenti valori:

- Per il titolo originale ed il titolo in italiano, una stringa;
- Per l'anno, un intero successivamente modificato in:

anno >= 2021	anno >= 2012	anno >= 2002	anno < 2002
4	3	2	1

- Per il genere, una stringa successivamente modificata in:

Animazione	Drammatico	Sentimentale	Commedia	Poliziesco
1	2	3	4	5

Thriller	Avventura	Documentario	Horror	Azione
6	7	8	9	10

Catastrofico	Western	Spionaggio	Biografico	Musicale
11	12	13	14	15

Fantasy	Guerra	Grottesco	Gangster	Mitologico
16	17	18	19	20

Storico	Noir	Supereroi	Biblico	Sportivo
21	22	23	24	25

Sperimentale	Cortometraggio
26	27

- Per il voto, un range di valori interi da 1 a 10 successivamente modificato in:

voto $\geq 8$	voto $\geq 5$	voto $\geq 3$	voto $< 3$
4	3	2	1

- Per la durata, un intero successivamente modificato in:

durata $< 90$	durata $< 120$	durata $< 140$	durata $> 140$
1	2	3	4

Il risultato della funzione **Predici Prezzo** sarà il prezzo alla quale quel determinato film o serie TV potrà esser venduto.

#### Esempio:

Il prezzo per il film e' di : 11 euro

### K-Nearest Neighbors – Suggerimento film

Nel nostro applicativo è stata implementata una seconda funzione che sfrutta il K-Nearest\_Neighbors.

Questo è in grado di suggerire un Film presente all'interno del dataset.

L'assistente, infatti, ponendo delle domande all'utente otterrà abbastanza dati per poter effettuare il suggerimento. In base ai dati:

- **Genere** del Film che si desidera vedere;

- **Anno**, periodo cinematografico del Film che si desidera guardare;
- Livello di **Humor** del Film che si desidera guardare;
- Livello di **Ritmo** del Film che si desidera guardare;
- Livello di **Impegno** del Film che si desidera guardare;
- Livello di **Tensione** del Film che si desidera guardare;

Il risultato della funzione 'Suggerisci nuovo film' sarà un determinato Film in base agli input forniti.

#### **Esempio suggerimento Film:**

```
Il film che ti suggeriamo e': ['Ladri di cadaveri. Burke & Hare']
```

### **Albero di decisione**

Nel nostro applicativo è stata implementata una funzione che sfrutta il DecisionTreeClassifier.

Questo è in grado di assegnare un film ad una determinata sezione presente all'interno del negozio, in base alle caratteristiche inserite in input.

I dati che vengono utilizzati sono:

- Livello di Humor del Film;
- Livello di Ritmo del Film;
- Livello di Impegno del Film;
- Livello di Tensione del Film;

#### **Esempio Sezione Film:**

```
Il film dovra' essere riposto nella sezione : impegno
```

### **Esempi di iterazione**

#### **1. Iterazione Predici Prezzo:**

```
Inserisci l'anno di pubblicazione del film:2013  
Definisci la durata del film:125  
Inserisci il voto (un numero da 1 a 10) :6
```

#### **2. Iterazione Suggerisci Film:**

```
+++++
+          1)Animazione          +
+          2)Drammatico          +
+          3)Sentimentale        +
+          4)Commedia            +
+          5)Poliziesco          +
+          6)Thriller            +
+          7)Avventura            +
+          8)Documentario        +
+          9)Horror              +
+          10)Azione             +
+          11)Catastrofico       +
+          12)Western            +
+          13)Spionaggio         +
+          14)Biografico         +
+          15)Musicale           +
+          16)Fantasy            +
+          17)Guerra             +
+          18)Grottesco          +
+          19)Gangster           +
+          20)Mitologico         +
+          21)Storico            +
+          22)Noir               +
+          23)Supereroi          +
+          24)Biblico            +
+          25)Sportivo           +
+          26)Sperimentale       +
+          27)Cortometraggio     +
+++++
Inserisci il genere di Film che ti piacerebbe vedere :1
Inserisci il valore il livello di humor tra 0 a 5 :3
Inserisci il valore il livello di ritmo tra 0 a 5 :4
Inserisci il valore il livello di impegno tra 0 a 5 :0
Inserisci il valore il livello di tensione tra 0 a 5 :0
Inserisci il periodo cinematografico che preferisci:
1: piu' di 20 anni fa
2: dal 2002 al 2012
3: dal 2012 al 2021
4: dal 2021 in poi:
--> 4
Il film che ti suggeriamo e': ['Alla ricerca della valle incantata']
```

### 3. Iterazione Scelta Sezione:

```
Inserisci il numero dell'operazione da effettuare :3
Inserisci il valore il livello di humor tra 0 a 5 :3
Inserisci il valore il livello di ritmo tra 0 a 5 :4
Inserisci il valore il livello di impegno tra 0 a 5 :0
Inserisci il valore il livello di tensione tra 0 a 5 :0
Il film dovra' essere riposto nella sezione : ritmo
```

### Implementazioni Future

In futuro, alcune feature che potrebbero essere implementate sono:

1. Inserimento GUI;
2. Inserimento gestione noleggio film;
3. Inserimento predizione mora su mancata restituzione di film;

### Processo di Sviluppo e Organizzazione del Lavoro

Il progetto è stato sviluppato a partire da inizio maggio 2022 fino a metà giugno 2022, **\*completamente in remoto\***.

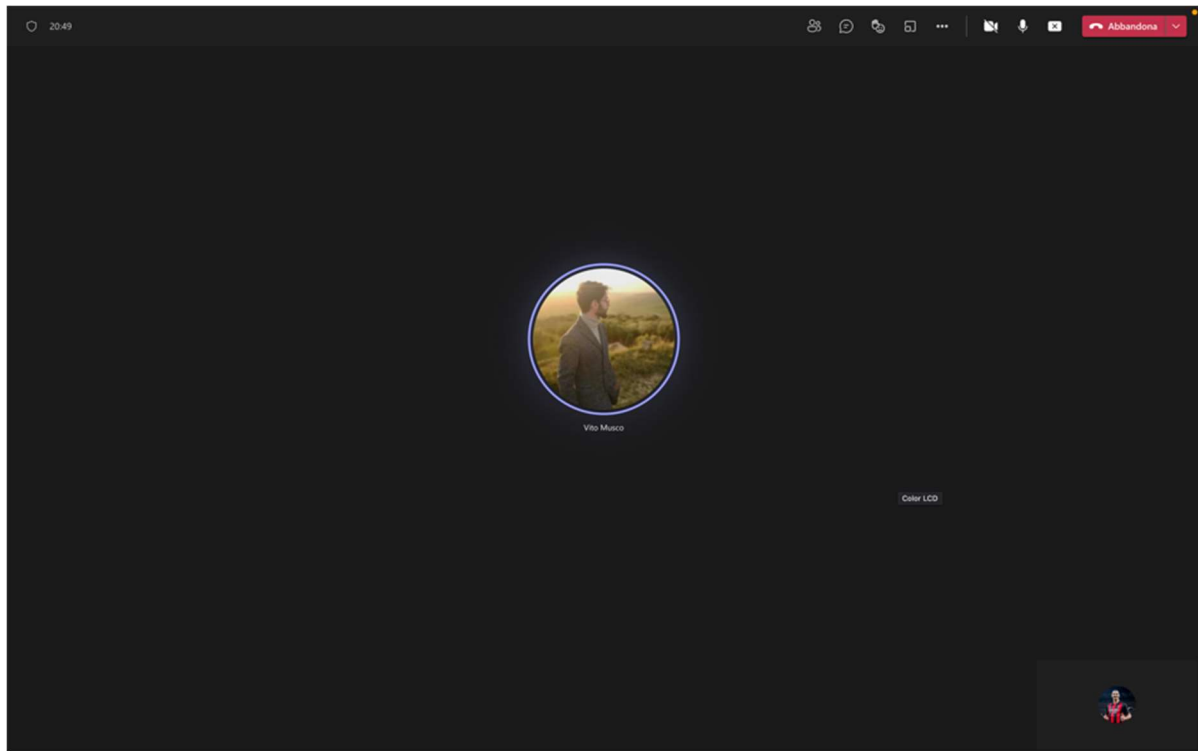
### Piattaforme di comunicazione

Per la comunicazione, il nostro gruppo, ha adottato due piattaforme:

- **Microsoft Teams**
- **Whatsapp**

La prima è stata scelta poiché è una piattaforma a tutti i membri del gruppo familiare, anche a causa del periodo pandemico attraversato, la quale permetteva di organizzare videoconferenze e di condividere lo schermo.

Ciò è stato molto utile nel momento in cui sorgevano difficoltà in quanto si poteva risolvere il problema tutti insieme.



La seconda è stata scelta poiché era il mezzo di comunicazione più immediato. Tramite questa piattaforma, è stato possibile confrontarsi durante lo sviluppo, decidere i giorni e gli orari per le nostre chiamate e per tenerci contatto.

## Conclusioni

Riteniamo che questo progetto sia stato un buon banco di prova. Nonostante le difficoltà riscontrate siamo riusciti comunque a rispettare gli obiettivi stabiliti e a trarre il meglio da questa esperienza formativa.

Ringraziamo per l'attenzione.

Gli studenti: *Francesco Sasso, Vito Musco*  
Applicazione: [FRAVIT]