

A man in a trench coat and hat stands in front of a futuristic, glowing mechanical structure. The background is filled with intricate, glowing mechanical components and wires, creating a steampunk or cyberpunk aesthetic. The lighting is warm and orange, highlighting the man's face and the metallic parts behind him.

ANTONIO DI GIORGIO

MOVIE MATE - IA

400TM

46

400TM

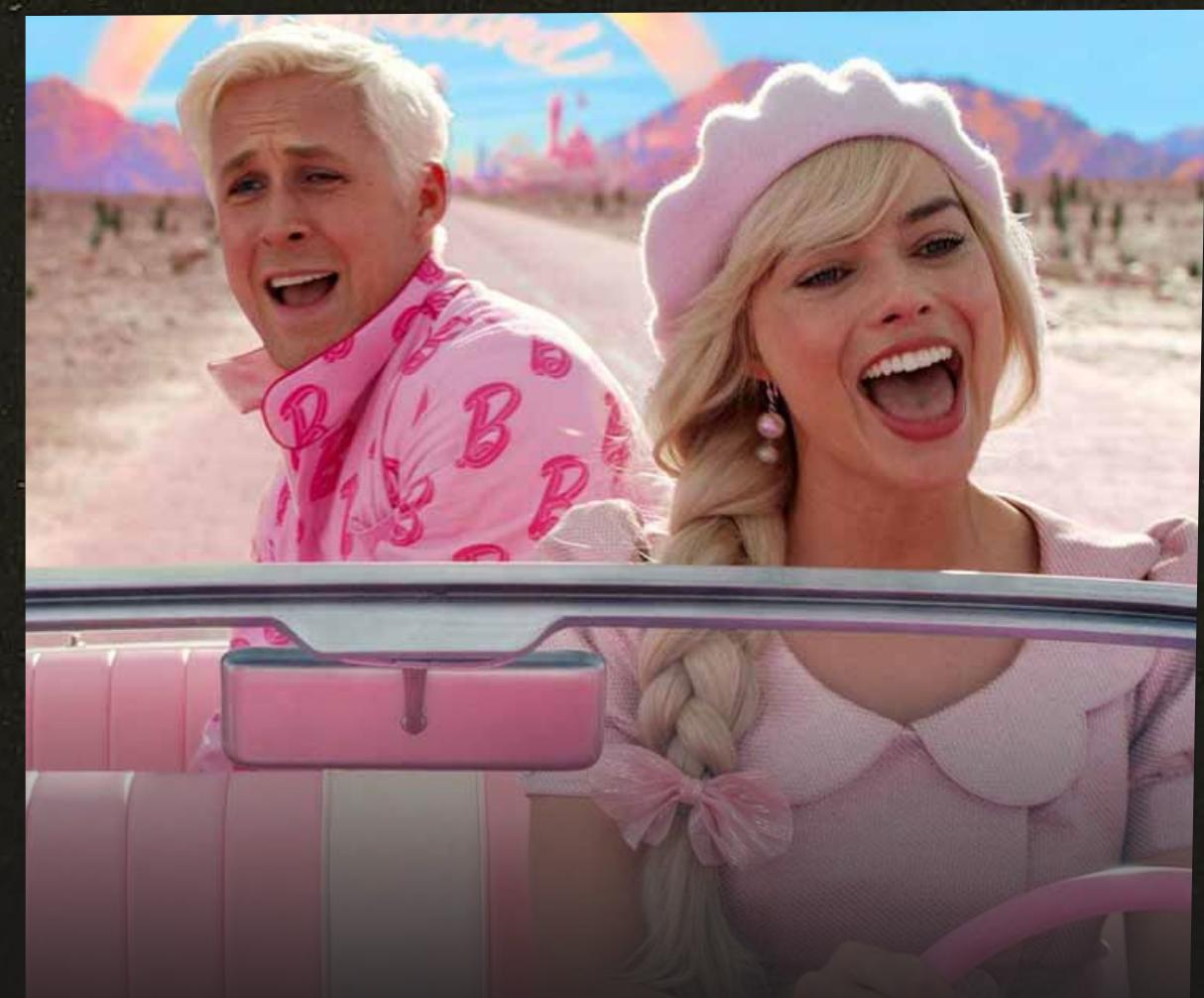
47

400TM

48

400TM

49



# OBIETTIVI

MovieMate-IA si pone l'obiettivo di realizzare un modulo di intelligenza artificiale in grado di prevedere una lista di film che possano interessare l'utente a seguito di determinate caratteristiche da lui inserite.

# SPECIFICA PEAS

---



## PERFORMANCE

La misura di prestazione si basa sui tempi di risposta dell'agente ma soprattutto sulla correttezza della lista di film proposti.



## ENVIRONMENT

L'ambiente in cui opera l'agente ` e costituito dai dataset con tutti i film e le rispettive statistiche estratte da essi, oltre che dall'interfaccia del chatbot utilizzata dall'utente per eseguire il software.



## ACTUATORS

L'agente agisce sull'ambiente fornendo all'utente una lista di film.



## SENSORS

L'agente percepisce l'ambiente tramite l'input testuale, da parte dell'utente, del genere, durata massima e rating del film oltre alle risposte ai temi associati al genere inserito



## ANALISI DEL PROBLEMA

---

Il problema preso in esame risulta essere un'istanza di apprendimento supervisionato, più nello specifico risulta essere un problema di classificazione.

Infatti, in base a degli input forniti dall'utente rappresentanti genere, durata, rating del film e preferenza o meno di determinati temi, l'agente esegue un'analisi delle risposte e fornisce una lista di film, composta da massimo 3 elementi, che rispecchiano le caratteristiche descritte dall'utente.

# MODELLO CRISP-DM

Per la realizzazione del progetto MovieMate-IA si è utilizzato il modello CRISP-DM, esso, infatti, rappresenta il ciclo di vita di tutti i progetti basati sull'intelligenza artificiale e il data science.

Il modello CRISP-DM si divide in:

BUSINESS UNDERSTANDING

DATA UNDERSTANDING

DATA PREPARATION

DATA MODELING

EVALUATION

DEPLOYMENT



# DATA UNDERSTANDING

Come dataset per il progetto MovieMate-IA è stato scelto da Kaggle il dataset “Letterboxd”.

Tale dataset è un’insieme di oltre 890 mila film e rispettive caratteristiche.

Risulta essere perfetto per lo scopo di MovieMate-IA



|                                     |                                    |
|-------------------------------------|------------------------------------|
| 1 Barbie                            | 2023 She's everything. He's just f |
| 2 Parasite                          | 2019 Act like you own the place.   |
| 3 Everything Everywhere All at Once | 2022 The universe is so much big   |
| 4 Fight Club                        | 1999 Mischief. Mayhem. Soap.       |
| 5 Interstellar                      | 2014 Mankind was born on Earth.    |
| 6 Joker                             | 2019 Put on a happy face.          |
| 7 Spider-Man: Into the Spider-Verse | 2018 More than one wears the m     |
| 8 Knives Out                        | 2019 Hell, any of them could have  |
| 9 La La Land                        | 2016 Here's to the fools who dre   |
| 10 Pulp Fiction                     | 1994 Just because you are a cha    |
| 11 The Batman                       | 2022 Unmask the truth.             |
| 12 Oppenheimer                      | 2023 The world forever changes.    |
| 13 Whiplash                         | 2014 The road to greatness can i   |
| 14 Get Out                          | 2017 Just because you're invited   |
| 15 Midsommar                        | 2019 Let the festivities begin.    |
| 16 The Dark Knight                  | 2008 Why So Serious?               |
| 17 Inception                        | 2010 Your mind is the scene of th  |

|     |                 |     |     |
|-----|-----------------|-----|-----|
| 002 | Comedy          | 0   | 0   |
| 002 | Thriller        | 0   | 0   |
| 002 | Drama           | 0   | 0   |
| 003 | Science Fiction | 0   | 1   |
| 003 | Action          | 0   | 1   |
| 003 | Adventure       | 0   | 1   |
| 004 | Drama           | 0   | 0   |
| 005 | Science Fiction | 0   | 1   |
| 005 | Drama           | 0   | 1   |
| 005 | Adventure       | 0   | 1   |
| 006 | Thriller        | 0   | 0   |
| ... | ...             | ... | ... |



# DATA PREPARATION

## CREAZIONE DI “NUOVO\_DATASET.CSV”

Il dataset originale è composto da 7 file, di cui non tutti utili, ed era complessa la sua gestione, si è optato per la realizzazione di un nuovo dataset che permettesse una visione e gestione più facile del dataset stesso.

L'operazione si è resa necessaria per due motivi:

1. Errori di elaborazione del dataset
2. Velocità di elaborazione

## DATA CLEANING

Una volta realizzato il nuovo dataset, su quest'ultimo viene effettuata una grande operazione di data cleaning.

Essa ci permette di eliminare tutti i dati duplicati, mancanti e/o inconsistenti e in generale elimina tutto ciò che può compromettere la qualità stessa dei dati.

## DECISION TREE

| Precision | Accuracy | Recall | F1-score |
|-----------|----------|--------|----------|
| 0.16      | 0.18     | 0.18   | 0.16     |

## RANDOM FOREST

| Precision | Accuracy | Recall | F1-score |
|-----------|----------|--------|----------|
| 0.17      | 0.19     | 0.19   | 0.17     |

## NAIVE BAYES

| Precision | Accuracy | Recall | F1-score |
|-----------|----------|--------|----------|
| 0.31      | 0.22     | 0.22   | 0.17     |

## GRADIENT BOOSTING MACHINE (GBM)

| Precision | Accuracy | Recall | F1-score |
|-----------|----------|--------|----------|
| 0.32      | 0.35     | 0.35   | 0.31     |

# DATA MODELING

Come abbiamo detto in precedenza, il problema in esame è un problema di classificazione.

Tra le tecniche e gli algoritmi analizzati per trovare la soluzione del problema, abbiamo:

1. Decision Tree
2. Random Forest
3. Naive Bayes
4. Gradient Boosting Machine (GBM)

Di ognuno di essi è stata analizzata la rispettiva Metrica di Valutazione e Matrice di Confusione, dalle quali si evince che il GBM risulta essere la tecnica migliore da utilizzare.



CANVA STORIES F20



CANVA STORIES F20

# EVALUATION

## CRITICITÀ

Come abbiamo potuto vedere nelle sezioni precedenti, il dataset è stato enormemente modificato affinché esso potesse rispondere alle esigenze del problema. Ciò significa che, prima di eseguire il sistema, deve essere realizzato il nuovo file “nuovo\_dataset.csv”, ad esempio ogni volta che il dataset originale viene aggiornato è necessaria tale operazione.

## SVILUPPI FUTURI

Un possibile sviluppo futuro è quello relativo all’eliminazione del file “nuovo\_dataset.csv” e l’utilizzazione integrale del dataset originale..



# DEPLOYMENT

---

Il progetto MovieMate-IA è attualmente disponibile ed utilizzabile da tutti, scaricabile dal GitHub dell'autore.

Link: <https://github.com/AntonioDG30/MovieMate-IA>



MARVEL

THANK YOU