

DOCUMENTAZIONE PROGETTO

INGEGNERIA DELLA CONOSCENZA 2021-2022

Progetto realizzato in solitaria da **Raffaele Di Anna**, matricola **725811**.



UNIVERSITÀ
DEGLI STUDI DI BARI
ALDO MORO

Email istituzionale: r.dianna@studenti.uniba.it

Link progetto: <https://github.com/Raffaele470/RDsAudioBookManager>

INTRODUZIONE:

Il progetto da me realizzato è chiamato **RD's AudioBook Manager**.

Si tratta di un'applicazione che ha come scopo quello di aiutare l'utente nel prevedere e consigliare il prezzo a cui vendere un AudioLibro e di suggerirne uno in base ai dati forniti in input.

In più è possibile effettuare ulteriori query sulla base di conoscenza.

Ho utilizzato un dataset pre-esistente in rete, contenente le informazioni riguardanti degli AudioLibri, in modo tale da poter effettuare le task preventivate.

Al dataset sono state apportate delle modifiche, aggiungendo alcuni valori, al fine di aver una migliore gestione, come verrà spiegato successivamente.

Inoltre, è stata realizzata una piccola base di conoscenza per effettuare delle interrogazioni su di essa.

REQUISITI:

Il progetto è stato realizzato in **Python**, quindi per il suo funzionamento è necessario avere a disposizione un programma che permetta la sua esecuzione. (ad. Esempio Visual Studio Code).

Inoltre, sono state importate delle **librerie**, per questo è necessario installarle da terminale:

scikit-plot → usando il comando ***"pip install scikit-plot"***

fsspec → usando il comando ***"pip install fsspec"***

sklearn → usando il comando ***"pip install sklearn"***

numpy → usando il comando ***"pip install numpy"***

matplotlib → usando il comando ***"pip install -U matplotlib"***

pytholog → usando il comando ***"pip install pytholog"***

pandas → usando il comando ***"pip install pandas"***

Una volta installate queste librerie, sarà possibile avviare l'applicativo correttamente.

GUIDA

Per avviare il programma basterà aprire il file “**Starting_Screen.py**”.

Una volta aperto, sarà visualizzata la **confusion matrix**.

Una volta chiusa quella finestra sarà visualizzato il menù, in cui si avrà la possibilità di selezionare una tra le opzioni presenti.

Per selezionare l’opzione relativa basterà digitare il numero corrispondente e premere invio.

```
-----
RD-AudibleManager
-----
Digita il numero corrispondente per effettuare la scelta
-----
1) Previsione prezzo
2) Suggerisci Audiolibro
3) Accuratezza
4) Interrogazioni sugli Audiolibri
5) Exit
-----
Inserisci un'operazione valida -->
```

1) PREVISIONE PREZZO

Dopo aver effettuato questa selezione, qui si avrà la possibilità di indicare dei dati in input al fine da poter ricevere un suggerimento sul prezzo a cui vendere un audiolibro.

Verrà chiesto di indicare la durata dell’audiolibro da vendere, il rating e il numero di recensioni.

Dopo averli inseriti si otterrà il prezzo consigliato.

```
Inserisci la durata in ore dell'audiolibro (per esempio: 1 equivale a 1 ora, 0.12 a 12 minuti, 0.02 a 2 minuto ) --> 3
Inserisci il rating dell'audiolibro (numero da 1 a 5) --> 4
Inserisci il numero di recensioni (numero da 1 a 242323)--> 20000
Il prezzo previsto per l'audiolibro e' pari a --> 11 euro
```

2) SUGGERISCI AUDIOLIBRO

Dopo aver effettuato questa selezione, qui si avrà la possibilità di indicare dei dati in input al fine da poter ricevere un consiglio su quale Audiolibro potrebbe piacere all’utente.

Verrà chiesto di inserire la categoria del libro e la sua durata. Dopo averli inseriti si otterrà l’Audiolibro consigliato.

```

----- CATEGORIA AUDIOLIBRO -----
----- 1) Arts & Entertainment
----- 2) Biographies & Memoirs
----- 3) Business & Careers
----- 4) Children's Audiobooks
----- 5) Computers & Technology
----- 6) Education & Learning
----- 7) Erotica
----- 8) Health & Wellness
----- 9) History
----- 10) Home & Garden
----- 11) LGBTQ+
----- 12) Literature & Fiction
----- 13) Money & Finance
----- 14) Mystery, Thriller & Suspense
----- 15) Politics & Social Sciences
----- 16) Relationships, Parenting & Personal Development
----- 17) Religion & Spirituality
----- 18) Romance
----- 19) Science & Engineering
----- 20) Science Fiction & Fantasy
----- 21) Teen
----- 22) Travel & Tourism
-----
Indica la categoria dell'audiolibro che preferisci --> 8
----- 1) fino 30 minuti
----- 2) fino a 4 ore
----- 3) fino a 15 ore
----- 4) fino a 24 ore
Indica la durata che vuoi che abbia l'audiolibro --> 3
L'audiobook consigliato per te e' --> ['A Court of Frost and Starlight']

```

3) ACCURATEZZA

Dopo aver effettuato questa selezione, qui si avrà la possibilità di visualizzare alcune caratteristiche in merito alla **Precision**, **Recall**, **Accuratezza** e **Mae** della previsione.

4) INTERROGAZIONI SUGLI AUDIOLIBRI

Dopo aver effettuato questa selezione, qui si avrà la possibilità di effettuare delle interrogazioni per la base di conoscenza sfruttando **Prolog**.

5) EXIT

Dopo aver effettuato questa selezione, qui si avrà la possibilità di **uscire** dall'applicazione.

PROGETTAZIONE APPLICAZIONE

Il **dataset** utilizzato contiene circa **2000** audiolibri. Originariamente esso era formato da:

Title, Book Author, Book Narrator, audioRuntime, Categories, average_rating, ratings_count e price.

Per una migliore gestione si sono aggiunte successivamente le informazioni:

audioRuntime_converted, average_rating_converted, ratings_count_converted, category_converted.

audioRuntime_converted → converte l'*audioRuntime*(la durata) in valori da 1 a 4.

average_rating_converted → converte l'*average_rating* in valori da 1 a 5.

ratings_count_converted → converte il *ratings_count* in valori da 1 a 5.

category_converted → converte la *categoria* scritta in formato stringa in valori da 1 a 22.

È stato implementato un **DecisionTreeClassifier** per la previsione del prezzo a cui vendere il l'Audiolibro. La decisione del tipo del classificatore è stata fatta confrontando l'accuratezza.

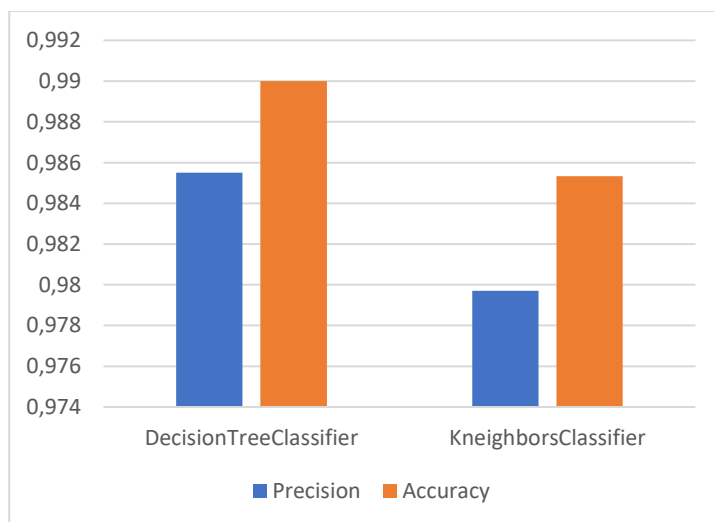
È stata effettuata una comparazione tra il *DecisionTreeClassifier* e il *KNeighborsClassifier*, andando a poi a preferire il primo.

L'accuratezza è stata misurata con la **K-fold Cross Validation**.

Si allega una tabella contenente i rispettivi dati:

CLASSIFICATORE	PRECISION	ACCURACY
DecisionTreeClassifier	0.9855072463768116	0.99
KNeighborsClassifier	0.9797101449275363	0.9853267973856208

Da questa tabella si evince la scelta del classificatore per la previsione del prezzo.



Come detto precedentemente è possibile visualizzare le informazioni riguardanti il sistema come:

- **Precision & recall** attraverso la funzione nel menù.

- **Mae** così da tener conto dell'errore assoluto tra feature predette e quelle reali che il sistema presenta sia in fase di test che in fase di training.

```

----> Precision : 0.9797101449275363
----> Recall : 1.0
----> Accuratezza : 0.9853267973856208
----> MAE - train set : 0
----> MAE - test set : 0

```

Per consigliare un AudioLibro è stato implementato invece, un **K-nearest_neighbors**. In base ai valori ricevuti in input, rappresentanti le caratteristiche dell'Audiolibro da voler trovare, ne verrà consigliato uno dal dataset.

È stato implementato anche **Prolog**, realizzando una piccola base di conoscenza per effettuare delle semplici query. Questa base di conoscenza, come è mostrato anche in foto, è composta da informazioni riguardanti gli Audiolibri, come il loro nome, categoria e durata.

Una volta selezionata la rispettiva voce all'interno del menù iniziale, verrà visualizzato un nuovo menù che consiglierà le possibili query.

```

1  :- disjointuous(i_book/3).
2
3  %-----
4  %Regole Book
5
6  %Mostra durata libro in base a titolo
7  audioruntime(Book,AudioRuntime) :- i_book(Book,_,AudioRuntime).
8
9  %Mostra nome categoria in base a titolo
10 book(Title, I_book1) :- i_book(Title,I_book1,_).
11
12
13 %-----
14
15 %DATI
16
17 %Immissione book
18
19 i_book(travels_with_charley_in_search_of_america,travel_and_tourism,07.58).
20 i_book(the_100,teen,07.17).
21 i_book(anarchy,romance,09.58).
22 i_book(profound_good,religion_and_spirituality,04.25).
23 i_book(evicted,politics_and_social_sciences,11.03).
24 i_book(100_baggers,money_and_finance,06.19).
25 i_book(brushfire,literature_and_fiction,19.18).
26 i_book(alone,lgbtq+,09.52).
27 i_book(meth_head_syndrome,home_and_garden,0.01).
28 i_book(heroes,history,15.01).
29 i_book(beyond_order,health_and_wellness,13.11).
30 i_book(self_discipline,education_and_learning,0.16).
31 i_book(the_hobbit,childrens_audiobooks,11.05).
32 i_book(winning,business_and_careers,06.44).
33 i_book(becoming,biographies_and_memoirs,19.03).
34 i_book(broken_horses,arts_and_entertainment,10.03).
35 i_book(igen,computers_and_technology,09.52).
36 i_book(cosmic_journey,science_and_engineering,0.10).
37 i_book(bloodline,science_fiction_and_fantasy,11.12).
38

```

Come si può notare abbiamo 2 regole e varie informazioni che possiamo sfruttare per eseguire 4 principali query:

- 1) **Trovare la durata dell'audiolibro** → inserendo il nome dell'audiolibro sarà possibile trovare la sua durata.
- 2) **Trovare la categoria dell'audiolibro** → inserendo il nome dell'audiolibro sarà possibile trovare la sua categoria.
- 3) **Trovare l'audiolibro di una durata specifica** → inserendo la durata sarà possibile trovare l'audiolibro con quella specifica durata.
- 4) **Trovare l'audiolibro di una categoria specifica** → inserendo una categoria sarà possibile trovare l'audiolibro con quella specifica categoria.

DATASET

Il dataset contiene circa 2000 Audiobook, di cui sono presenti diverse informazioni.

	A	B	C	D	E	F	G	H	I	J	K	L
1	Title	Book Author	Book Narrator	audioRuntime	Categories	average_rating	ratings_count	price	audioRuntime_converted	average_rating_converted	ratings_count_converted	category_converted
2	Travels with Charley in Search of America	John Steinbeck	Gary Sinise	07.58	Travel & Tourism	4.6	4254	10	3	4	3	22
3	In Charge of Moonlight	Dina Gregory	Matt Addis	0.05	Teen	4	79	6	1	4	1	21
4	Panic: Ghosts and Legends	Lauren Oliver	Camron Jones	03.13	Teen	4	40	7	2	4	1	21
5	We Were Liars	E. Lockhart	Ariadne Meyers	06.27	Teen	4	4759	10	3	4	3	21
6	Beautiful Creatures	Kami Garcia	Kevin T. Collins	17.33	Teen	4.1	5361	12	4	4	4	21
7	Lord of the Flies	William Golding	William Golding	06.35	Teen	4.2	7688	11	3	4	4	21
8	The 100	Kass Morgan	Justin Torres	07.17	Teen	4.2	1065	10	3	4	3	21
9	The Giver	Lois Lowry	Ron Rifkin	04.49	Teen	4.4	11683	11	3	4	4	21
10	A Court of Frost and Starlight	Sarah J. Maas	Amanda Leigh Cobb	06.23	Teen	4.4	5110	11	3	4	4	21
11	Fable	Adrienne Young	Emma Lysy	08.54	Teen	4.4	533	9	3	4	2	21
12	Shadow and Bone	Leigh Bardugo	Lauren Fortgang	09.21	Teen	4.4	8194	11	3	4	4	21
13	Divergent	Veronica Roth	Emma Galvin	11.11	Teen	4.4	35049	12	3	4	5	21
14	They Both Die at the End	Adam Silvera	Michael Crouch	08.30	Teen	4.5	4199	10	3	4	3	21
15	Tweet Cute	Emma Lord	Dan Bittner	10.08	Teen	4.5	235	9	3	4	2	21

Come si può visualizzare dalla foto sono presenti le colonne:

Title → titolo dell'audiolibro

Book Author → autore dell'audiolibro

Book Narrator → narratore dell'audiolibro

audioRuntime → durata dell'audiolibro

Categories → categoria dell'audiolibro

average_rating → voto medio dell'audiolibro

ratings_count → numero di voti dell'audiolibro

price → Prezzo dell'audiolibro

audioRuntime_converted → audioRunTime convertito per motivi progettuali.

average_rating_converted → average_rating convertito per motivi progettuali.

ratings_count_converted → ratings_count convertito per motivi progettuali.

category_converted → category convertito per motivi progettuali.

PREVISIONE PREZZO

Come detto precedentemente, è stata implementata una funzionalità per la previsione del prezzo a cui vendere un Audiolibro dopo aver inserito dei dati di input. Questa funzione sfrutta un **DecisionTreeClassifier** andando ad addestrare grazie al file CSV.

I dati di input chiesti all'utente sono:

- **audioRuntime** → durata dell'audiolibro
- **average_rating** → voto medio dell'audiolibro
- **ratings_count** → numero di voti dell'audiolibro

Ma come detto prima, questi sono stati convertiti:

- da **audioRuntime** in **audioRuntime_converted** sulla base di:

*Se **audioRuntime** ≤ 0.30 → **audioRuntime_converted** = 1*

*Se **audioRuntime** ≤ 4 → **audioRuntime_converted** = 2*

*Se **audioRuntime** ≤ 15 → **audioRuntime_converted** = 3*

*Se **audioRuntime** > 15 → **audioRuntime_converted** = 4*

- da **average_rating** in **average_rating_converted** sulla base di:

*se **average_rating** ≥ 5 → **average_rating_converted** = 5*

*se **average_rating** ≥ 4 → **average_rating_converted** = 4*

*se **average_rating** ≥ 3 → **average_rating_converted** = 3*

*se **average_rating** ≥ 2 → **average_rating_converted** = 2*

*se **average_rating** < 2 → **average_rating_converted** = 1*

- **ratings_count** in **ratings_count_converted** sulla base di:

*se **ratings_count** ≥ 20000 → **ratings_count_converted** = 5*

*se **ratings_count** > 5000 → **ratings_count_converted** = 4*

*se **ratings_count** > 1000 → **ratings_count_converted** = 3*

*se **ratings_count** > 100 → **ratings_count_converted** = 2*

*se **ratings_count** ≤ 100 → **ratings_count_converted** = 1*

Inoltre, anche **Categories** è stato convertito in **category_converted** sulla base di:

Categories	Category_converted
Arts & Entertainment	1
Biographies & Memoirs	2
Business & Careers	3
Children's Audiobooks	4
Computers & Technology	5
Education & Learning	6
Erotica	7
Health & Wellness	8
History	9
Home & Garden	10
LGBTQ+	11
Literature & Fiction	12
Money & Finance	13
Mystery, Thriller & Suspense	14
Politics & Social Sciences	15
Relationships, Parenting & Personal	16
Religion & Spirituality	17
Romance	18
Science & Engineering	19
Science Fiction & Fantasy	20
Teen	21
Travel & Tourism	22

Sfruttando questi dati otterremo come risultato il prezzo del libro previsto.

```
Il prezzo previsto per l'audiolibro e' pari a --> 11 euro
```

SUGGERISCI AUDIOLIBRO

È stata implementata una funzionalità anche per il suggerimento di un libro, sulla base dei valori inseriti in input, sfruttando il **K-Nearest_Neighbors**.

Attraverso questa funzionalità, verranno chiesti in input alcuni valori:

- **category_converted** → numero indicante la categoria dell'audiolibro
- **audioRuntime_converted** → numero indicante la durata dell'audiolibro

Infine, quindi, si otterrà un audiolibro consigliato.

```
L'audiobook consigliato per te e' --> ['A Court of Frost and Starlight']
```

ESEMPI PRATICI

- Previsione prezzo:

```
Inserisci la durata in ore dell'audiolibro (per esempio: 1 equivale a 1 ora, 0.12 a 12 minuti, 0.02 a 2 minuto ) --> 3
Inserisci il rating dell'audiolibro (numero da 1 a 5) --> 4
Inserisci il numero di recensioni (numero da 1 a 242323)--> 20000
Il prezzo previsto per l'audiolibro e' pari a --> 11 euro
```

- Suggerisci Audiolibro:

```
Inserisci un'operazione valida --> 2
|----- CATEGORIA AUDIOLIBRO -----|
|---- 1) Arts & Entertainment          |
|---- 2) Biographies & Memoirs         |
|---- 3) Business & Careers            |
|---- 4) Children's Audiobooks         |
|---- 5) Computers & Technology        |
|---- 6) Education & Learning          |
|---- 7) Erotica                      |
|---- 8) Health & Wellness             |
|---- 9) History                      |
|---- 10) Home & Garden                |
|---- 11) LGBTQ+                      |
|---- 12) Literature & Fiction         |
|---- 13) Money & Finance              |
|---- 14) Mystery, Thriller & Suspense |
|---- 15) Politics & Social Sciences   |
|---- 16) Relationships, Parenting & Personal Development |
|---- 17) Religion & Spirituality      |
|---- 18) Romance                     |
|---- 19) Science & Engineering        |
|---- 20) Science Fiction & Fantasy    |
|---- 21) Teen                        |
|---- 22) Travel & Tourism             |
|-----|
Indica la categoria dell'audiolibro che preferisci --> 9
----- 1) fino 30 minuti
----- 2) fino a 4 ore
----- 3) fino a 15 ore
----- 4) fino a 24 ore
Indica la durata che vuoi che abbia l'audiolibro --> 3
L'audiobook consigliato per te e' --> ['102 Minutes']
```

- Interrogazioni sugli Audiolibri (prolog):

```
Indica l'operazione da effettuare --> 1
Indica il nome dell' audiolibro --> the_100
{'D': '07.17'}
|-----|
|----- 1) Trova la durata dell'audiolibro
|----- 2) Trova la categoria dell'audiolibro
|----- 3) Trova audiolibro di una durata specifica
|----- 4) Trova audiolibro di una categoria specifica
|----- 5) Ritorna al menu' principale
|-----|
Indica l'operazione da effettuare --> 2
Indica il nome dell'audiolibro --> the_100
{'T': 'teen'}
|-----|
|----- 1) Trova la durata dell'audiolibro
|----- 2) Trova la categoria dell'audiolibro
|----- 3) Trova audiolibro di una durata specifica
|----- 4) Trova audiolibro di una categoria specifica
|----- 5) Ritorna al menu' principale
|-----|
Indica l'operazione da effettuare --> 3
Indica durata dell'audiolibro -->07.17
{'F': 'the_100'}
|-----|
|----- 1) Trova la durata dell'audiolibro
|----- 2) Trova la categoria dell'audiolibro
|----- 3) Trova audiolibro di una durata specifica
|----- 4) Trova audiolibro di una categoria specifica
|----- 5) Ritorna al menu' principale
|-----|
Indica l'operazione da effettuare --> 4
Indica il genere dell'audiolibro --> history
{'F': 'heroes'}
|-----|
|----- 1) Trova la durata dell'audiolibro
|----- 2) Trova la categoria dell'audiolibro
|----- 3) Trova audiolibro di una durata specifica
|----- 4) Trova audiolibro di una categoria specifica
|----- 5) Ritorna al menu' principale
|-----|
```

CONCLUSIONI

Il progetto è stato realizzato al fine dello svolgimento dell'esame di **Ingegneria della Conoscenza**, frequentata presso **l'Università di Bari Aldo Moro**.

Personalmente, penso che questo progetto sia stato molto utile per mettere in pratica le conoscenze teoriche acquisite durante il corso e di aver potuto interagire per la prima volta con l'ambiente Python, nonostante le difficoltà riscontrate durante lo sviluppo, soprattutto ritrovandomi a svolgerlo in solitaria.

Raffaele Di Anna, 725811