



MENTROSHIP PROGRAM

AI AND MACHINE LEARNING

College of Engineering and Computer Science | Prince Sattam
bin Abdulaziz University

student:
Rand Albej

supervisor:
Dr.Haya AlAskar

ABOUT AI AND MACHINE LEARNING

MACHINE LEARNING AND AI ARE TWO RELATED FIELDS IN COMPUTER SCIENCE THAT INVOLVE DEVELOPING SYSTEMS THAT CAN LEARN FROM DATA, MAKE PREDICTIONS OR DECISIONS, AND PERFORM TASKS THAT NORMALLY REQUIRE HUMAN INTELLIGENCE. MACHINE LEARNING ALGORITHMS ALLOW COMPUTERS TO IDENTIFY PATTERNS IN LARGE DATASETS AND USE THOSE PATTERNS TO MAKE PREDICTIONS OR TAKE ACTIONS WITHOUT BEING EXPLICITLY PROGRAMMED. ARTIFICIAL INTELLIGENCE (AI) REFERS TO THE BROADER CONCEPT OF MACHINES THAT CAN PERFORM TASKS THAT WOULD TYPICALLY REQUIRE HUMAN INTELLIGENCE, SUCH AS RECOGNIZING IMAGES OR SPEECH, UNDERSTANDING NATURAL LANGUAGE, AND EVEN MAKING DECISIONS. MACHINE LEARNING IS OFTEN A CRUCIAL COMPONENT OF AI SYSTEMS.

WHY RECOMMENDATION SYSTEM?

ARE WE BEING WATCHED? OR DO DEVICES AND SOCIAL MEDIA SITES HAVE EARS TO HEAR US? OR THEY MUST BE READING OUR MINDS!

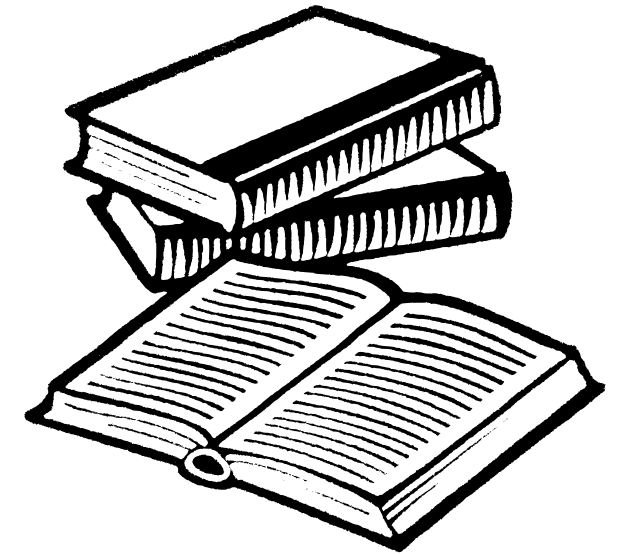
I OFTEN HAD THESE THOUGHTS EVERY TIME AN ADVERTISEMENT APPEARED FOR THE PERFUME OR WATCH THAT I TOLD MY SISTER THAT I INTENDED TO BUY, OR THAT I SEARCHED FOR ON A SPECIFIC SITE, AND NOW IT APPEARS ON ALL SITES, IN ADDITION TO OTHER SIMILAR PRODUCTS.

AND AS SOON AS I WATCH A CERTAIN MOVIE OR SERIES, MY PROGRAMS SUGGESTIONS ARE FILLED WITH SIMILAR MOVIES AND SERIES. THEREFORE, I DECIDED TO RESEARCH HOW THIS ALGORITHM WORKS AND HOW TO DESIGN A SIMILAR ALGORITHM USING MACHINE LEARNING. SO I TRY TO UNDERSTAND AND APPLY AN ALGORITHM TO DEVELOP A CONTENT RECOMMENDATION ENGINE THAT AIMS TO ANALYZE USER'S CHOICES AND SUGGEST CHOICES SIMILAR TO THEIR PREFERENCES.

BOOK RECOMMENDATION SYSTEM USING MACHINE LEARNING

construction stages

- DATA COLLECTION
- DATA LOADING
- DATA ANALYSIS
- MODEL BUILDING
- TRAINING THE MODEL
- EVALUATION



1. DATA COLLECTION

COLLECT DATA FROM VARIOUS SOURCES, SUCH AS AMAZON, GOODREADS, OR ANY OTHER ONLINE BOOKSTORES. THE DATA SHOULD INCLUDE INFORMATION ABOUT THE BOOKS, SUCH AS AUTHOR, GENRE, PUBLICATION DATE, RATINGS, AND REVIEWS.

in my project I used a data set from kaggle

2. DATA LOADING

```
[ ] books=pd.read_csv('/content/book_data/Books.csv')
```

```
[ ] books.head(2)
```

	ISBN	Book-Title	Book-Author	Year-Of-Publication	Publisher	Image-URL-S	Image-URL-M
0	0195153448	Classical Mythology	Mark P. O. Morford	2002	Oxford University Press	http://images.amazon.com/images/P/0195153448.0...	http://images.amazon.com/images/P/0195153448.0...
1	0002005018	Clara Callan	Richard Bruce Wright	2001	HarperFlamingo Canada	http://images.amazon.com/images/P/0002005018.0...	http://images.amazon.com/images/P/0002005018.0...

```
[ ] users=pd.read_csv('/content/book_data/Users.csv')
```

```
users.head()
```

	User-ID	Location	Age
0	1	nyc, new york, usa	NaN
1	2	stockton, california, usa	18.0
2	3	moscow, yukon territory, russia	NaN
3	4	porto, v.n.gaia, portugal	17.0
4	5	farnborough, hants, united kingdom	NaN

3.DATA ANALYSIS

- INVOLVES EXAMINING AND INTERPRETING DATA IN ORDER TO DRAW MEANINGFUL INSIGHTS FROM IT, LIKE CLEANING AND PREPROCESSING THE DATA BY REMOVING DUPLICATES, MISSING VALUES, AND IRRELEVANT COLUMNS.

```
[ ] num_rating=ratings_with_books.groupby('title')['rating'].count().reset_index()
```

```
[ ] num_rating.rename(columns={"rating":"num_of_ratings"},inplace=True)
```

```
[ ] final_rating=ratings_with_books.merge(num_rating, on='title')
```

```
[ ] final_rating.head()
```

	user_id	ISBN	rating	title	author	year	Publisher	img_url	num_of_ratings
0	277427	002542730X	10	Politically Correct Bedtime Stories: Modern Ta...	James Finn Garner	1994	John Wiley & Sons Inc	http://images.amazon.com/images/P/002542730X.0...	82
1	3363	002542730X	0	Politically Correct Bedtime Stories: Modern Ta...	James Finn Garner	1994	John Wiley & Sons Inc	http://images.amazon.com/images/P/002542730X.0...	82
2	11676	002542730X	6	Politically Correct Bedtime Stories: Modern Ta...	James Finn Garner	1994	John Wiley & Sons Inc	http://images.amazon.com/images/P/002542730X.0...	82
3	12538	002542730X	10	Politically Correct Bedtime Stories: Modern Ta...	James Finn Garner	1994	John Wiley & Sons Inc	http://images.amazon.com/images/P/002542730X.0...	82
4	13552	002542730X	0	Politically Correct Bedtime	James Finn	1994	John Wiley &	http://images.amazon.com/images/P/002542730X.0...	82

3.DATA ANALYSIS

- HERE I USED PIVOT TABLE WHICH ENABLE YOU TO SELECT THE DATA THAT YOU WANT TO ANALYZE, CHOOSE THE CATEGORIES YOU WANT TO GROUP IT BY, AND THEN APPLY CALCULATIONS SUCH AS SUM, AVERAGE, COUNT, OR OTHER FUNCTIONS TO GENERATE A SUMMARY REPORT.

3-pivot table

```
[ ] book_pivot=final_rating.pivot_table(columns='user_id',index='title',values='rating')
```

```
[ ] book_pivot
```

	user_id	254	2276	2766	2977	3363	3757	4017	4385	6242	6251	...	274004	274061	2
	title														
	1984	9.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	
	1st to Die: A Novel	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	
	2nd Chance	NaN	10.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	
	4 Blondes	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	0.0	...	NaN	NaN	
	84 Charing Cross Road	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	

4.MODEL BUILDING

- THIS STEP INVOLVES CHOOSING AN APPROPRIATE MACHINE LEARNING ALGORITHM BASED ON THE NATURE OF THE PROBLEM AND THE CHARACTERISTICS OF THE DATA.

4.MODEL BUILDING

I USED THE NEAREST NEIGHBORS ALGORITHM WHICH IS A MACHINE LEARNING ALGORITHM THAT FALLS UNDER THE CATEGORY OF UNSUPERVISED LEARNING. IT IS USED FOR FINDING SIMILAR ITEMS OR OBJECTS BASED ON THEIR FEATURES OR CHARACTERISTICS. THE ALGORITHM WORKS BY IDENTIFYING THE CLOSEST DATA POINT(S) TO A GIVEN QUERY POINT, AND THEN USING THESE NEAREST NEIGHBORS TO MAKE PREDICTIONS OR RECOMMENDATIONS.

```
[ ] from scipy.sparse import csr_matrix
```

```
[ ] book_sparse=csr_matrix(book_pivot)
```

```
▶ from sklearn.neighbors import NearestNeighbors  
model=NearestNeighbors(algorithm='brute')
```

5. TRAINING THE MODEL

*FITTING THE MODEL ON THE
TRAINING SET*

```
] model.fit(book_sparse)
```

▼ NearestNeighbors
NearestNeighbors(algorithm='brute')

```
] distance, suggestion=model.kneighbors(book_pivot.iloc[237,:].values.reshape(1,-1),n_neighbors=6)
```

6.EVALUATING

TO TEST THE MODEL I ENTER A BOOK NAME SO
THE MODEL SUGGEST ANY SIMILAR BOOKS TO THE
ONE I CHOOSE

```
[ ] book_name='A Bend in the Road'  
    recommend_book(book_name)
```

```
A Bend in the Road  
No Safe Place  
Exclusive  
The Cradle Will Fall  
Family Album  
Lake Wobegon days
```

6.EVALUATING

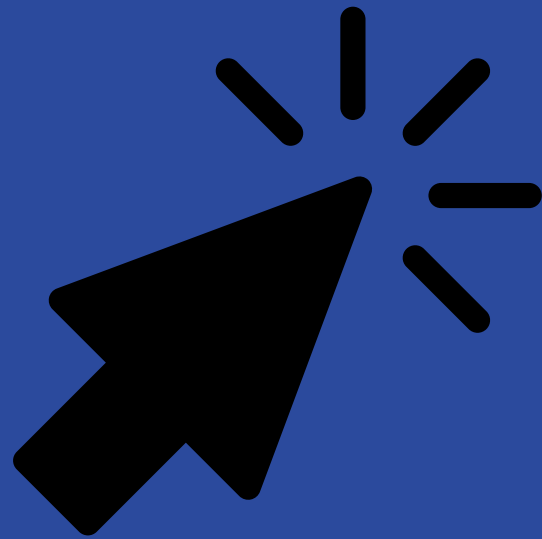
FOR EVALUATING THE MODEL I LOAD A NEW BOOK DATA CONSISTING OF TWO BOOKS WITH THEIR TITLES, AUTHORS, AND YEARS OF PUBLICATION. WE THEN USE THE PREDICT METHOD OF THE TRAINED KNEIGHBORSCLASSIFIER OBJECT TO MAKE PREDICTIONS ON THE NEW DATA. FINALLY, PRINT THE PREDICTED GENRES FOR THE TWO BOOKS.

```
# Loading a new book data for testing
new_book = pd.DataFrame({'title': ['The Da Vinci Code', 'Harry Potter and the Goblet of Fire'],
                        'author': ['Dan Brown', 'J.K. Rowling'],
                        'year': [2003, 2000]})

# making predictions on the new data using the trained model
new_predictions = knn.predict(new_book)

# printing the predicted genres
print(new_predictions)
```

The code link



THANKS AND APPRECIATION

I WOULD LIKE TO THANK THE GIFTED
CARE UNIT FOR THIS INFORMATIVE
AND INCREDIBLE PROGRAM AND
GIVING ME THE OPPORTUNITY TO BE
PART OF IT, AND MANY THANKS FOR
DR.HAYA ALASKAR FOR HER SUPPORT
AND KNOWLEDGE DURING THE
PROJECT, AND THANKS TO EVERYONE
WHO CONTRIBUTED TO PROVIDING
THIS OPPORTUNITY FOR THE
STUDENTS.