



أكاديمية سدايا
SDAIA Academy

Children's Books

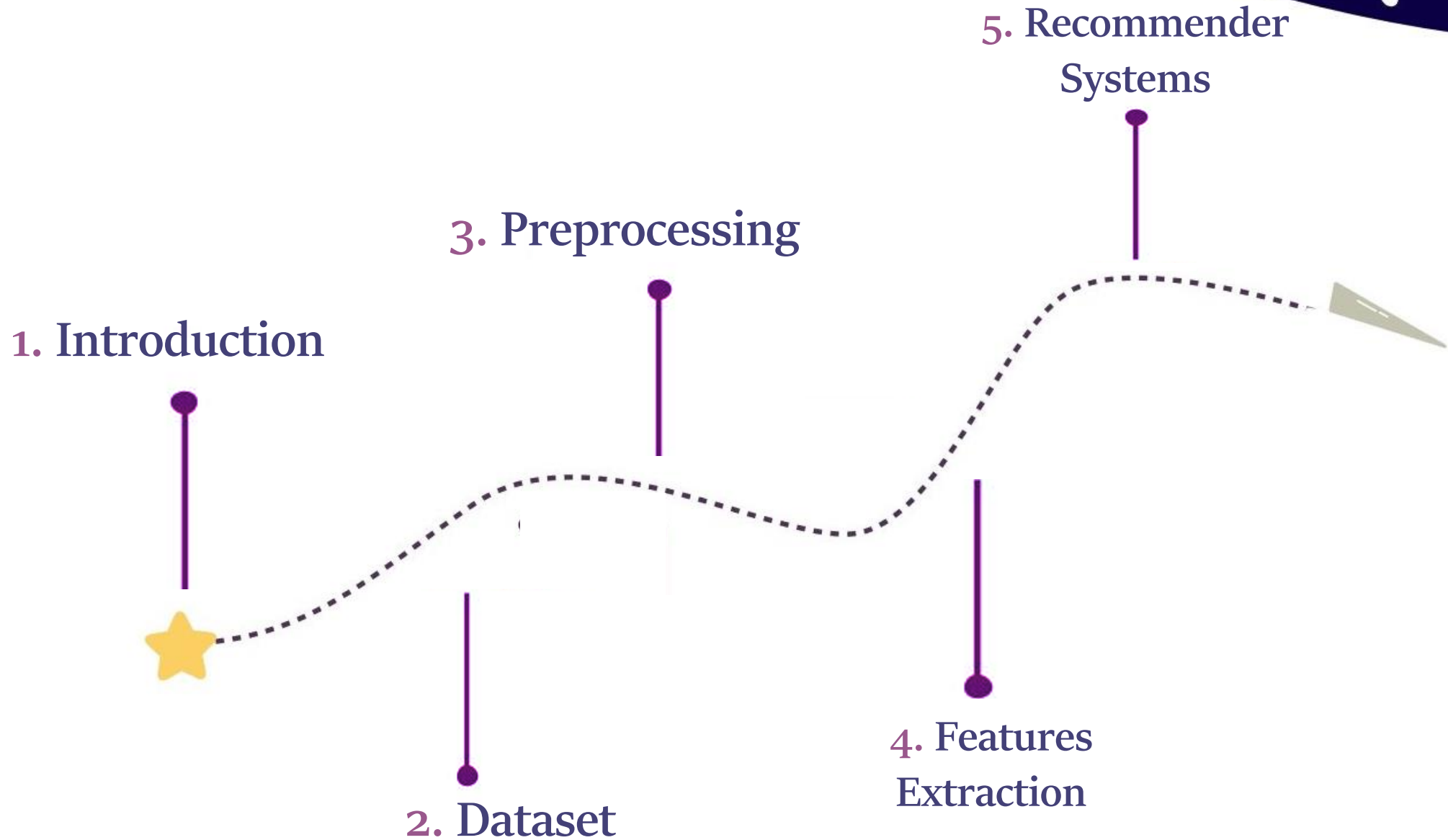
Natural Language Processing
&
Unsupervised Learning

Abdumajeed Alnfaie

Nouf Alshabani

Ahmad Hakami

Agenda





Introduction

The importance of reading to children cannot be underestimated. Reading for pleasure can benefit a child's education, social and cognitive development, well-being, and mental health. The effect of reading on a child's development is significant. Research by **BookTrust** has discovered the profound benefits of reading for child development, facilitating social interaction between adults and children, and encouraging children to interact with the world around them.

Inspiration

Teachers and parents are in an important position to ensure that reading is an essential part of children's daily routine. We 'll give them a system that recommends books based on the child's interests and age.

Dataset

Source

Highly Rated Children Books And Stories

It was uploaded in Kaggle.com

Features

3269 Books

Columns

5 Columns

*Title, Author, Desc, Inerest_age, Reading_age, They are all **OBJECT!***

Target

Building a system to recommend books by age

Dataset

Title	Author	Desc	Inerest_age	Reading_age
The Girl Who Drank the Moon	Kelly Barnhill	Every year, the evil Protectorate offers a bab...	10-14	10-14
Time Between Us	Tamara Ireland Stone	Sixteen-year-old Anna is struggling to underst...	13+	12+
Girl Out of Water	Nat Luurtsema	Lou Brown's swimming ambitions sank without tr...	13-18	10+
Captive	A J Grainger	Robyn is scared. Ever since the attempted assa...	13+	13
The School of Music	Rachel Bowen and Meurig Bowen	Illustrator: Dan...	10+	10+
...
Spy for the Queen of Scots	Theresa Breslin	Beautiful young aristocrat Ginette, known as J...	12+	12+
Jimmy Corrigan:	Chris Ware	This extraordinary graphic novel is so obvious...	14+	14+
The Martian Girl	NaN	Following on from the brilliant Lost on Mars, ...	10+	10+
Buffalo Soldier	Tanya Landman	Plantation slave Charley Smith is eleven when ...	13+	11+
Witchstruck	Victoria Lamb	The future Elizabeth I has been imprisoned by ...	12+	12+

Preprocessing

Tokenize

Breaks the sentence text into words.

Lemmatization

The processes of grouping together the different inflected forms of a word so they can be analyzed as a single item.

Stopwords

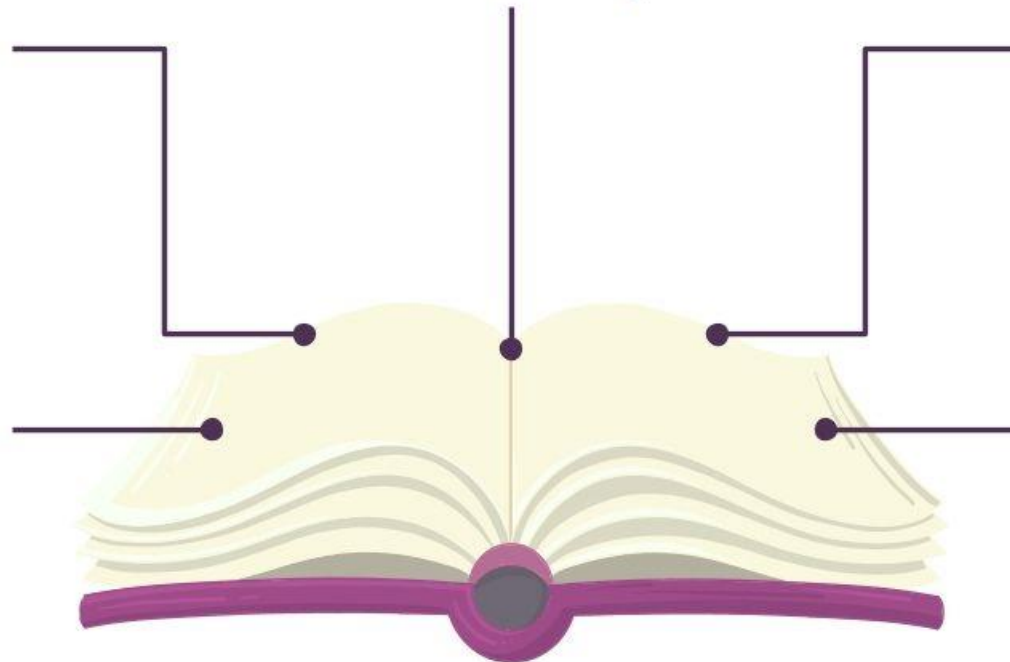
Remove stopwords are common words in the language.

Punctuation

To remove the punctuations from the text data.

Lower case

To convert the input text to lower case.



Feature Extraction

Countvectorizer

is a way to convert a given set of strings into a frequency representation.

TF-IDF

focuses on the frequency of words present in the corpus and provides the importance of the words.



Unsupervised models

Latent Dirichlet Allocation, Countvectorizer

THE TOP 5 WORDS FOR LDA CountVectorizer #0

```
['love', 'plot', 'daughter', 'historical', 'high', 'character', 'charley', 'shakespeare', 'play', 'twylla']
```

THE TOP 5 WORDS FOR LDA CountVectorizer #1

```
['baby', 'witch', 'friend', 'ben', 'roger', 'family', 'xan', 'mary', 'amy', 'london']
```

THE TOP 5 WORDS FOR LDA CountVectorizer #2

```
['make', 'sister', 'old', 'live', 'feel', 'teen', 'year', 'butter', 'family', 'novel']
```

THE TOP 5 WORDS FOR LDA CountVectorizer #3

```
['want', 'reader', 'landscape', 'make', 'mother', 'people', 'strange', 'feeling', 'boy', 'father']
```


Unsupervised models

Latent Dirichlet Allocation, Countvectorizer

THE TOP 5 WORDS FOR LDA CountVectorizer #0

['love', 'plot', 'daughter', 'historical', 'high', 'character', 'charley', 'shakespeare', 'play', 'twylla']

THE TOP 5 WORDS FOR LDA CountVectorizer #1

['baby', 'witch', 'friend', 'ben', 'roger', 'family', 'xan', 'mary', 'amy', 'london']

THE TOP 5 WORDS FOR LDA CountVectorizer #2

['make', 'sister', 'old', 'live', 'feel', 'teen', 'year', 'butter', 'family', 'novel']

THE TOP 5 WORDS FOR LDA CountVectorizer #3

['want', 'reader', 'landscape', 'make', 'mother', 'people', 'strange', 'feeling', 'boy', 'father']

Unsupervised models

K-means Clustering, TF-IDF

cluster 0:

red, right, man, london, dad, mare, suzy, charley, robyn, people,

cluster 1:

father, mother, amy, butter, ben, caz, music, play, shakespeare, novel,

cluster 2:

sister, twylla, lennie, life, girl, dory, meg, feeling, cole, appeal,

cluster 3:

frankenstein, mary, human, friend, xan, monster, strange, nick, frank, tale,

Unsupervised models

Latent Dirichlet Allocation, TF-IDF

THE TOP 4 WORDS FOR TOPIC, LDA With TF-IDF #0

['high', 'character', 'amber', 'right', 'robyn', 'charley', 'meg', 'play', 'shakespeare', 'twylla']

THE TOP 4 WORDS FOR TOPIC, LDA With TF-IDF #1

['jenny', 'scot', 'baby', 'rory', 'mary', 'xan', 'roger', 'london', 'ben', 'amy']

THE TOP 4 WORDS FOR TOPIC, LDA With TF-IDF #2

['live', 'family', 'face', 'music', 'sister', 'cassandra', 'novel', 'teen', 'dory', 'butter']

THE TOP 4 WORDS FOR TOPIC, LDA With TF-IDF #3

['fall', 'people', 'landscape', 'cole', 'suzy', 'father', 'boy', 'feeling', 'strange', 'caz']

Unsupervised models

Latent Semantic Analysis, TF-IDF

THE TOP 4 WORDS FOR TOPIC, LSA With TF-IDF #0

['friendship', 'father', 'novel', 'grief', 'feel', 'family', 'people', 'feeling', 'friend', 'year']

THE TOP 4 WORDS FOR TOPIC, LSA With TF-IDF #1

['romance', 'fan', 'grief', 'caz', 'queen', 'bennett', 'fall', 'mary', 'year', 'meg']

THE TOP 4 WORDS FOR TOPIC, LSA With TF-IDF #2

['victor', 'monstrous', 'man', 'creature', 'mary', 'dramatic', 'version', 'human', 'monster', 'frankenstein']

THE TOP 4 WORDS FOR TOPIC, LSA With TF-IDF #3

['robyn', 'northward', 'amber', 'involved', 'rory', 'man', 'away', 'brother', 'london', 'cole']

Unsupervised models

✓ Non-Negative Matrix Factorization, TF-IDF

THE TOP 4 WORDS FOR TOPIC, NMF With TF-IDF #0

['amy', 'landscape', 'father', 'feel', 'fall', 'amber', 'best', 'feeling', 'grief', 'year']

THE TOP 4 WORDS FOR TOPIC, NMF With TF-IDF #1

['terrifies', 'man', 'creature', 'monstrous', 'victor', 'dramatic', 'version', 'human', 'monster', 'frankenstein']

THE TOP 4 WORDS FOR TOPIC, NMF With TF-IDF #2

['start', 'fun', 'frank', 'nick', 'butter', 'bullied', 'make', 'michael', 'strange', 'people']

THE TOP 4 WORDS FOR TOPIC, NMF With TF-IDF #3

['french', 'witch', 'historical', 'twylla', 'scot', 'jenny', 'court', 'queen', 'meg', 'mary']

Dataset After Preprocessing

	book_id	Title	Author	Reading_age	Desc_Vectorizer	NMF	Image
0	0	The Girl Who Drank the Moon	Kelly Barnhill	10	every year evil protectorate offer baby evil w...	Fantasy	http://images.amazon.com/images/P/0195153448.0...
1	1	Time Between Us	Tamara Ireland Stone	12	sixteenyearold anna struggling understand benn...	Romantic	http://images.amazon.com/images/P/0002005018.0...
2	2	Girl Out of Water	Nat Luurtsema	10	lou brown swimming ambition sank without trace...	Romantic	http://images.amazon.com/images/P/0060973129.0...
3	3	Captive	A J Grainger	13	robyn scared ever since attempted assassinatio...	Science fiction	http://images.amazon.com/images/P/0374157065.0...
4	4	The School of Music	Rachel Bowen and Meurig Bowen Illustrator: Dan...	10	welcome music charge maestro sergio trunk thre...	Science fiction	http://images.amazon.com/images/P/0393045218.0...

NMF: Fantasy, Romantic, Science fiction, Science

Recommender Systems

Let's see the web-site !

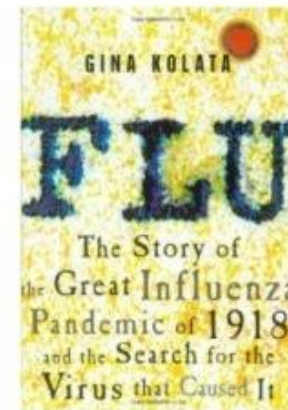
Top 4 recommended books

Type or select a Title from the dropdown

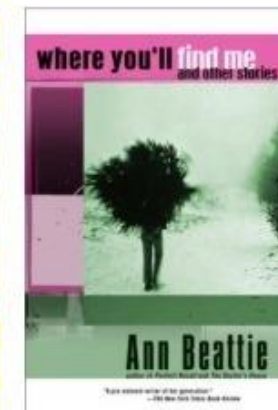
The Song from Somewhere Else

Recommend

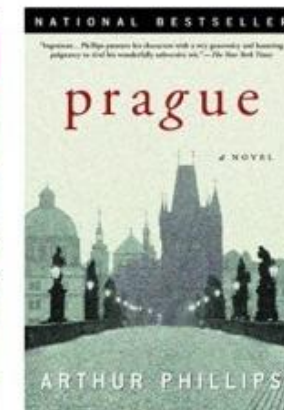
Captive



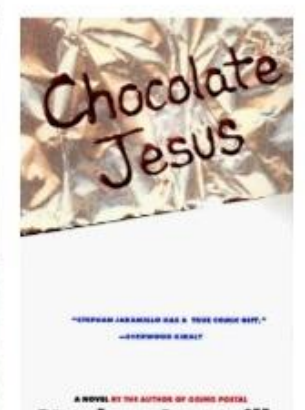
Butter



Invictus



The Name of the Star



TOOLS



Plotly



Jupyter Notebook



Scikit learn



Python



Pandas



PyCharm



Streamlit



Conclusion

To make reading fun and educational for children, it is necessary to focus on their interests and ages to choose the best books and help them develop their imagination in the best way.



Thanks!

Do you have any questions?

Abdulmajeed

Github
[@AbdulmajeedAlnefaie](#)

Nouf

Github
[@NoufAlshabani](#)

Ahmad

Github
[@AhmadHakami](#)