
ReadMe

Book Price Prediction

By: Batoul Alosaimi ,Norah Alqahtani and Shroaq Almutairi

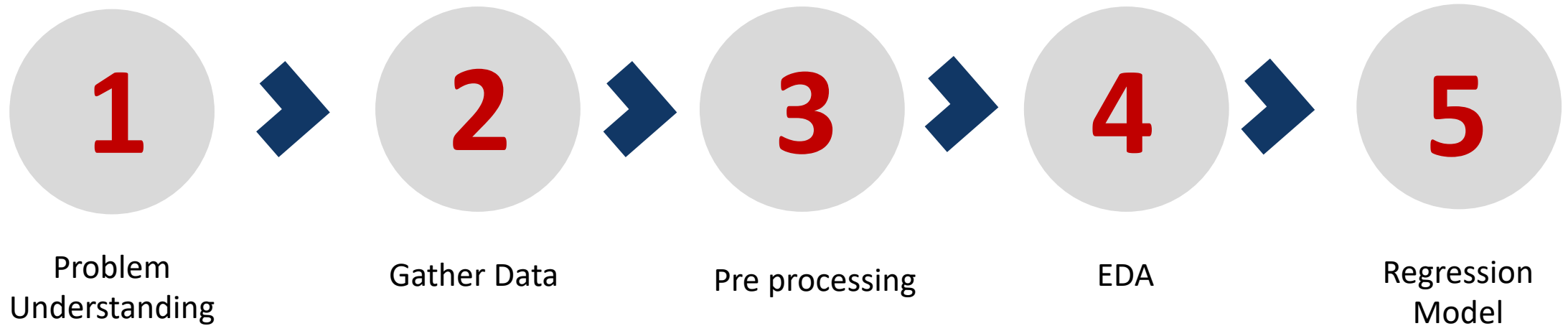


ReadME

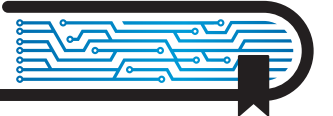
ReadMe which is an online web site that is selling books.



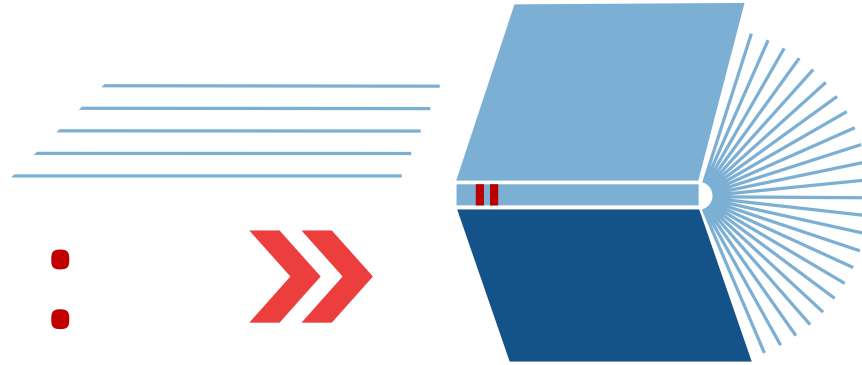
Methodology



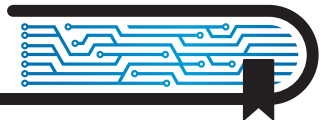
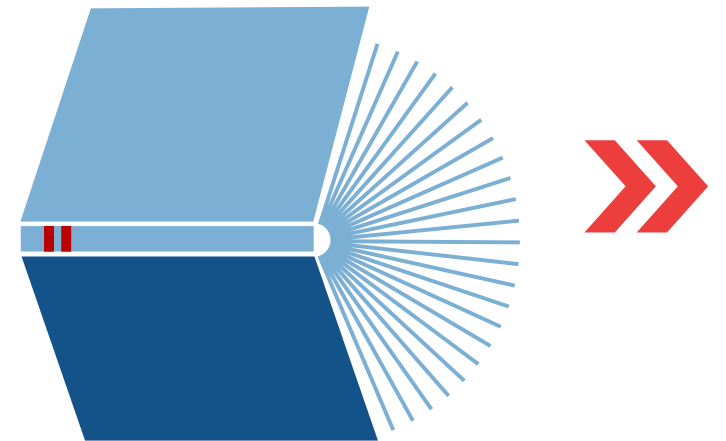
READ ME



Data Source : >>



1-kaggle(CSV file)
GoodReads dataset ,
11128 books with 11
features .

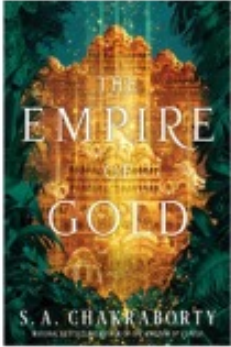


Book Finder :

Web Scraping
(Selenium)
1352 books
with 1 feature.

Search Results

Search Author, Title or ISBN



ISBN is **9780062678164** 0062678167

The Empire of Gold: A Novel (The Daevabad Trilogy, 3)
by **Chakraborty, S. A.**

4.51 avg rating • 23,112 ratings by GoodReads
Publisher: Harper Voyager, 2020
Edition: Hardcover
Language: English

Prices **INCLUDE** standard shipping to Saudi Arabia
[Show prices without shipping](#)
[Change shipping destination/currency.](#)

Shipping prices may be approximate. Please verify cost before checkout.

About the book:
No series since George R.R. Martin's palace intrigue and the way that tribal most. Paste Magazine

The final chapter in the bestselling, critically acclaimed series about a con-woman and an idealistic djinn prince in a devastating civil war.

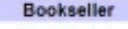

Daevabad has fallen.

After a brutal conquest stripped the city of its resurrected commander, Dara, must try to reclaim the city from fractious, warring people.




[Show more](#)

Click on the price to find out more about a book.

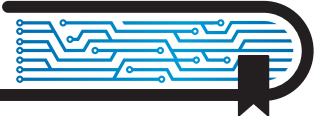
New books: 1 - 50 of 57

#	Bookseller	Notes	Price
1.	 amazon.com United States	Hardcover, ISBN 9780062678164 Publisher: Harper Voyager, 2020 Hardcover. Edition: Illustrated. 784 pages. Usually ships within 2 to 3 days. Fulfillment by Amazon.	\$31.85
2.	GreatBookPrices via  alibris.com United States	Hardcover, ISBN 9780062678164 Publisher: Harper Voyager, 2020 New: Hard cover. Sewn binding. Cloth over boards. 784 p. Contains: Illustrations. Daevabad Trilogy, 3. In Stock. 100% Money Back Guarantee. Brand New, Perfect Condition, allow 4-14 business days for standard shipping. To Alaska, Hawaii, U.S. protectorate, P.O. box, and APO/FPO addresses allow 4-28 business days for Standard shipping. No expedited shipping. All orders placed with expedited shipping will be cancelled. Over 3,000,000 happy customers.	\$35.49

Used books: 1 - 43 of 43

1.	World of Books Ltd via  BIBLIO.com United Kingdom	Softcover, ISBN 9780062678164 Publisher: Harper Voyager, 2020 Used - Like New: Paper
2.	World of Books Ltd via  BIBLIO.co.uk United Kingdom	Softcover, ISBN 9780062678164 Publisher: Harper Voyager, 2020 Used - Like New: Paper
3.	Seattle Goodwill via  Goodwill.com	Hardcover, ISBN 9780062678164 Publisher: Harper Voyager, 2020 Used - Good: Hardcover

READ ME

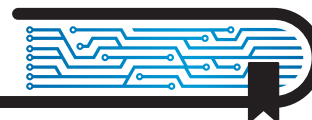


Book dataset :

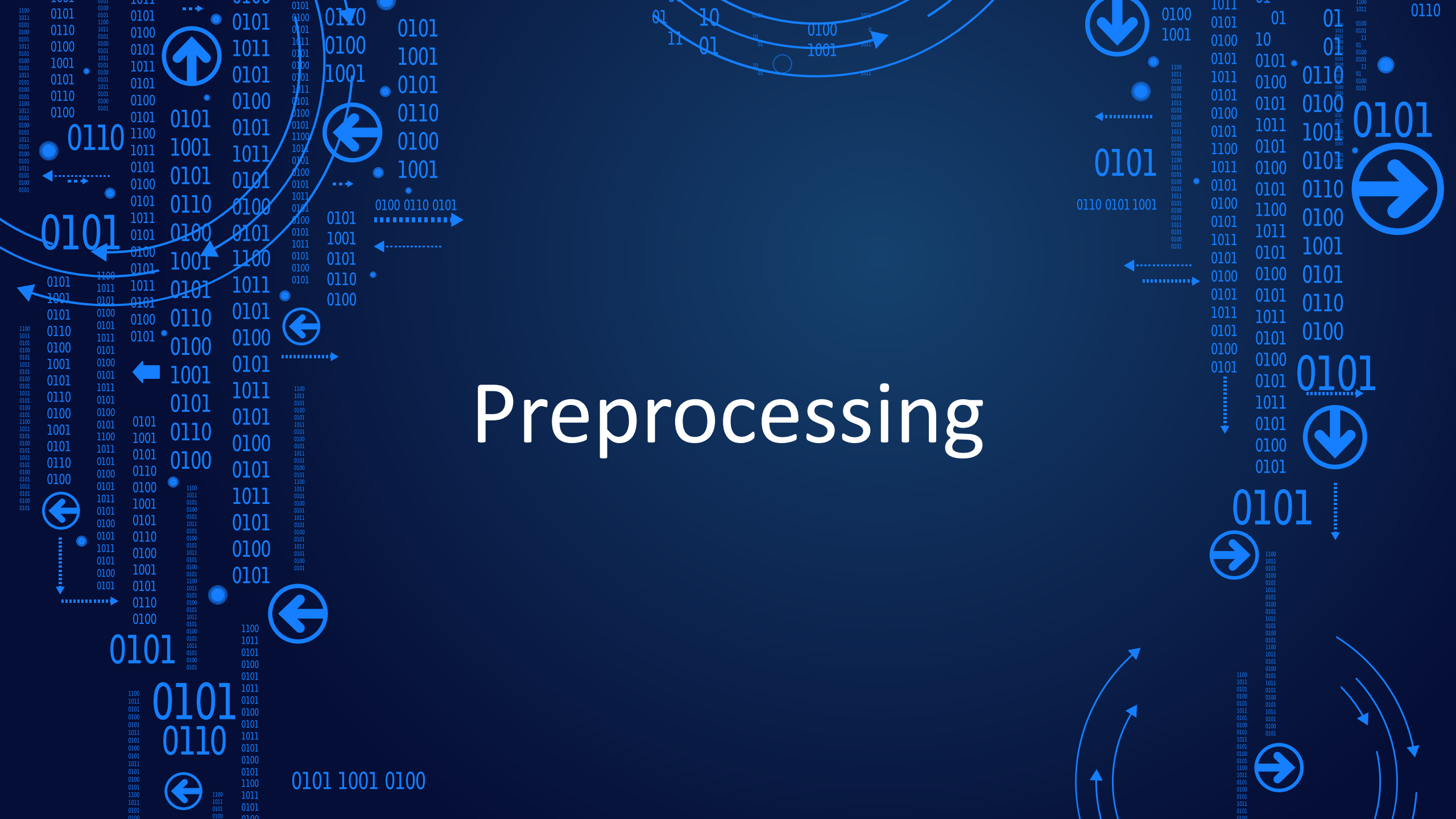
- 1298 books 12 features.

	title	authors	average_rating	isbn	isbn13	language_code	num_pages	ratings_count	text_reviews_count	publication_date	publisher	Price
bookID												
663	For the New Intellectual: The Philosophy of Ay...	Ayn Rand	3.68	451163087	9.780450e+12	eng	224	2750	108	12/1/1963	Signet Book	SAR 48.675
2326	Dirk Gently's Holistic Detective Agency (Dirk ...	Douglas Adams	3.98	1597770078	9.781600e+12	eng	6	58	15	9/30/2005	Phoenix Audio	Not found
3869	A Brief History of Time	Stephen Hawking	4.17	553380168	9.780550e+12	eng	212	239652	5860	9/1/1998	Bantam Books	SAR 48.675
2045	Later Novels and Other Writings: The Lady in t...	Raymond Chandler/Frank MacShane	4.47	1883011086	9.781880e+12	eng	1076	1107	48	10/1/1995	Library of America	SAR 173.732876712329

READ ME



Preprocessing

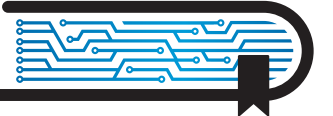


Pre processing:

Removing nulls
value.

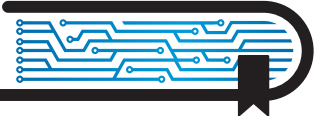
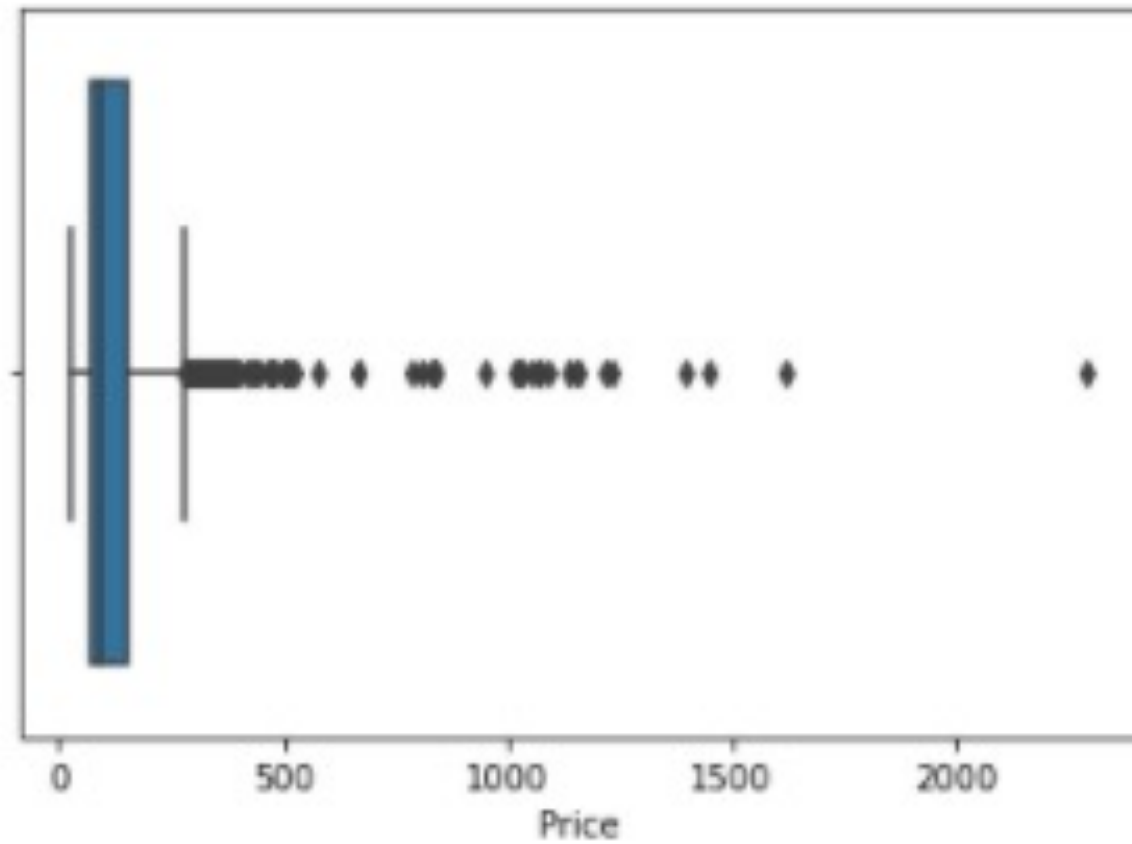
Removing
Duplicates.

Clean Price
column



Pre processing:

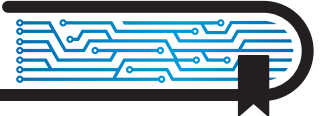
Removing outlier in
the price



Exploratory Data Analysis

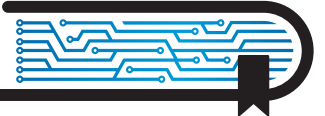
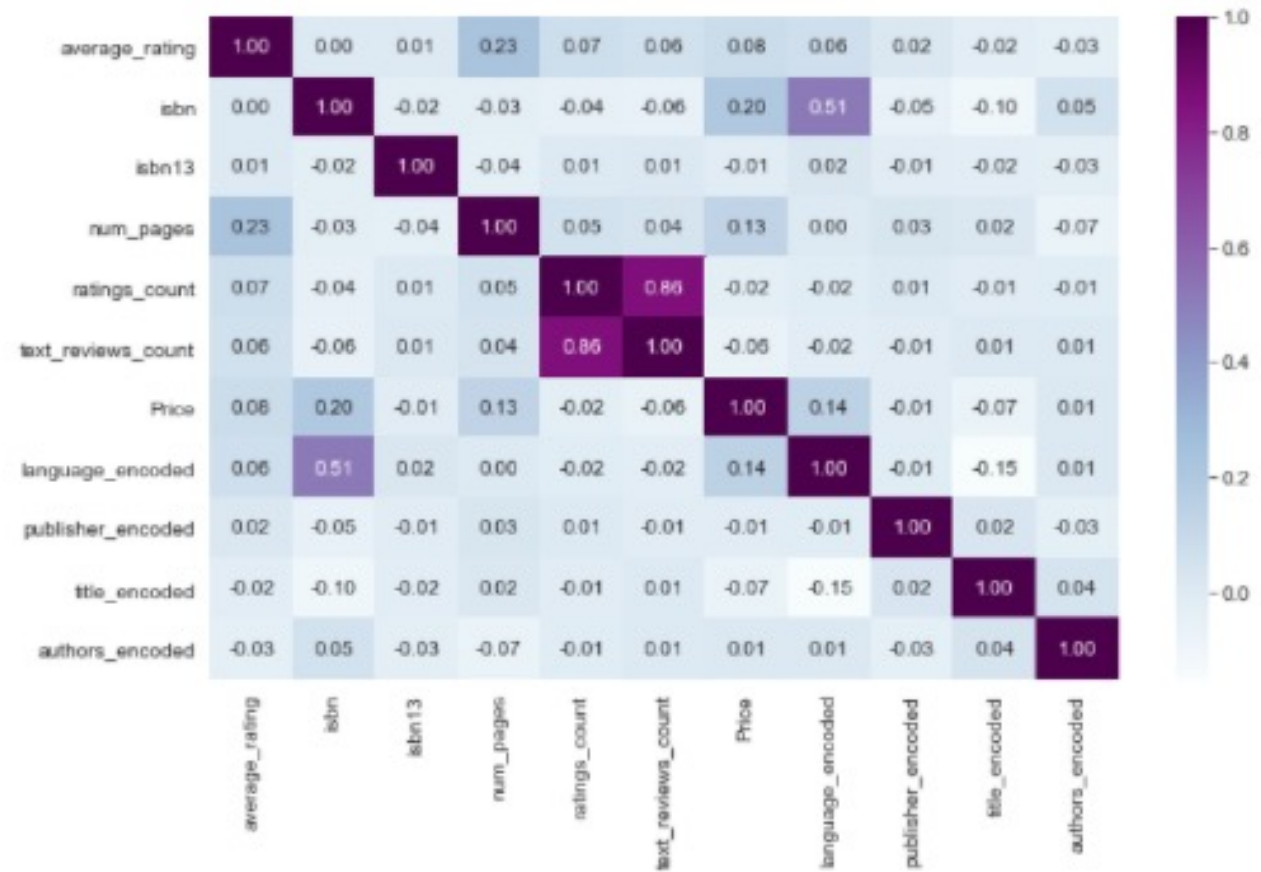
EDA:

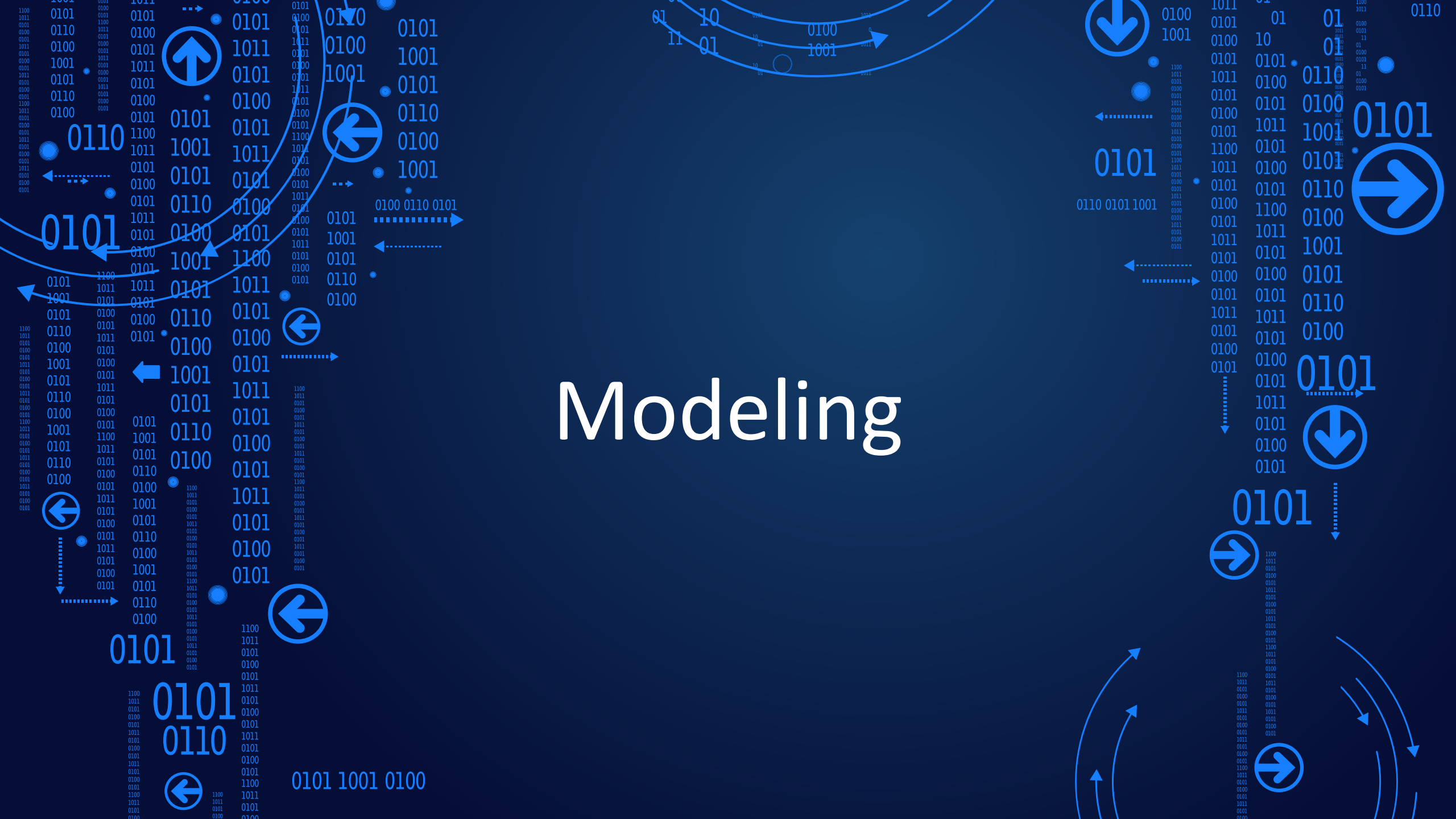
Avg price (SAR)
of each
language



EDA:

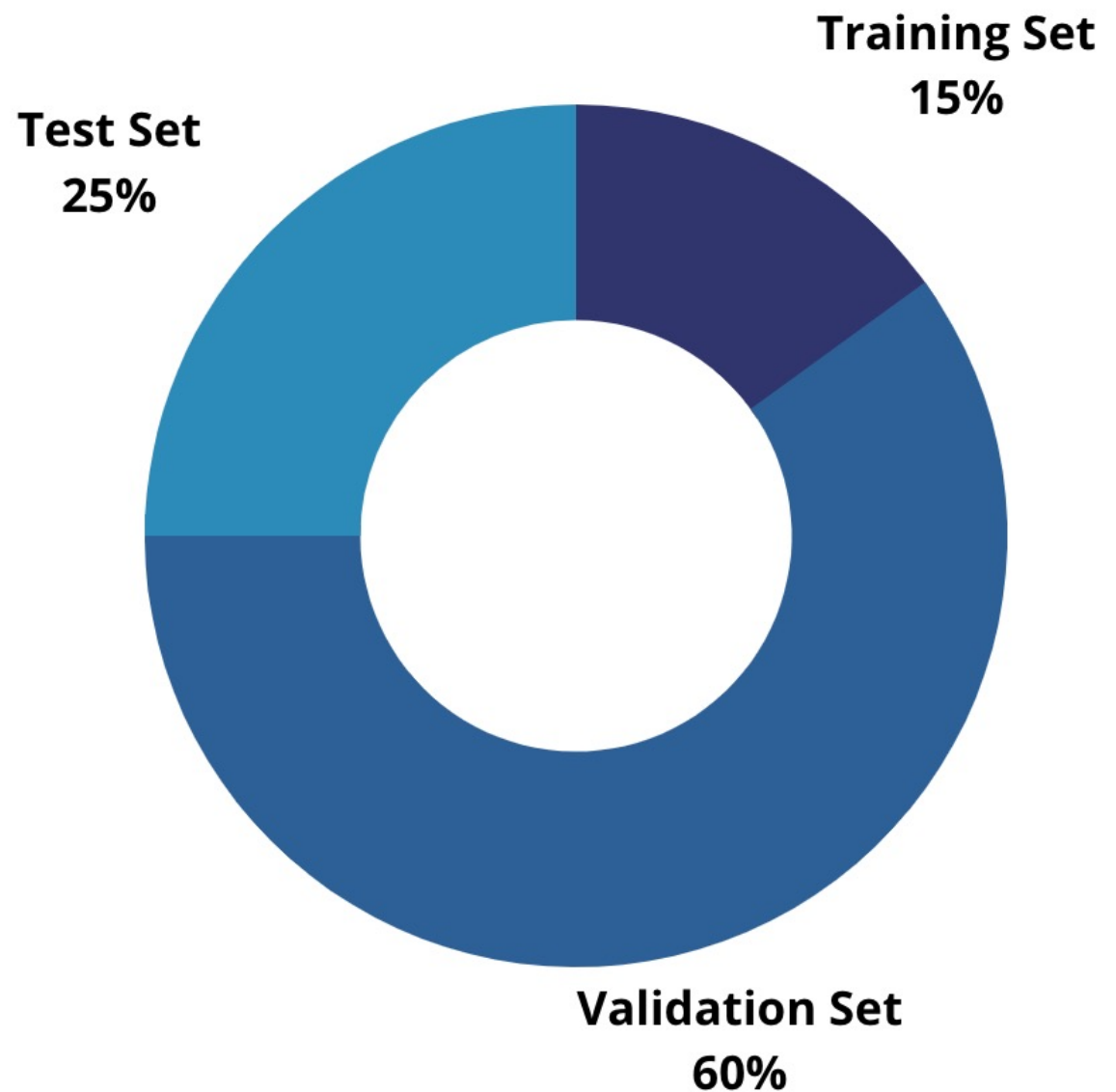
correlation
between feature:



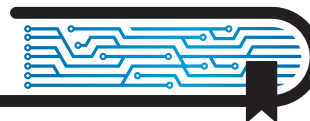


Modeling

Splitting the data
to Test (25%)
and Train-
Validation (75%)
sets



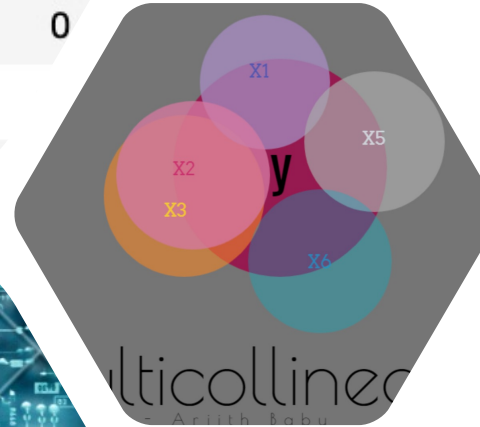
Modeling :



en-CA en-GB en-US eng enm fre ger grc jpn mul spr

bookID

1	0	0	0	1	0	0	0	0	0	0
2	0	0	0	1	0	0	0	0	0	0
5	0	0	0	1	0	0	0	0	0	0
8	0	0	0	1	0	0	0	0	0	0
9	0	0	1	0	0	0	0	0	0	0



Feature Engineering



Cross Validation

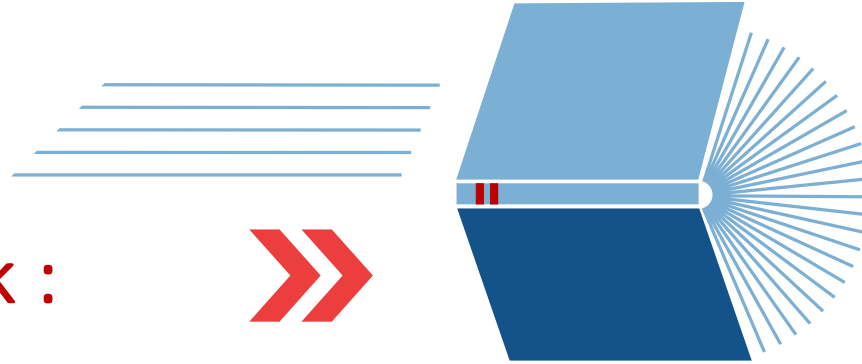
Regression Algorithm	Linear Regression	Ridge Regression	Polynomial Regression (degree = 2)	Lasso Regression
Training Score	0.075843	0.044664	0.068357	0.066383
Validation Score	0.029533	0.00066	0.006979	0.002216



Modeling After feature Engineering :

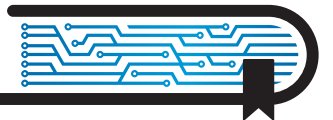
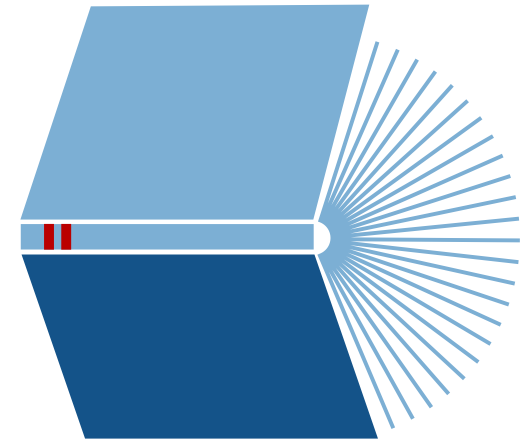
- Linear Regression Score Result:
- $R^2 = 0.119$

Conclusion & future work :



1. Data obtained via scraping websites are not reliable enough to validate a model's performance.

2. In future work we will co more effort to obtain more related features to reduce Bias



thank you

