

# Capstone Project Submission

## Instructions:

- i) Please fill in all the required information.
- ii) Avoid grammatical errors.

Team Member's Name, Email and Contribution:
<p><b>Name:</b> Shivani Thakur</p> <p><b>Email Id:</b> <a href="mailto:tshivani344@gmail.com">tshivani344@gmail.com</a></p> <p><b>Contribution:</b></p> <ul style="list-style-type: none"><li>• Data Wrangling</li><li>• Data description and summary</li><li>• Problem Statement</li><li>• Knowing the data</li><li>• Data pre-processing</li><li>• Outliers Detection</li><li>• Exploratory data analysis</li><li>• EDA Conclusion</li><li>• Recommender System</li><li>• Model Evaluation</li><li>• Conclusion</li></ul>
<p><b>Please paste the GitHub Repo link:</b> <a href="https://github.com/ShivaniThakur-19/Book-Recommendation-System-ML-Capstone-III">https://github.com/ShivaniThakur-19/Book-Recommendation-System-ML-Capstone-III</a></p>
<p><b>Please write a short summary of your Capstone project and its components. Describe the problem statement, your approaches and your conclusions. (200-400 words)</b></p>

## **Introduction:**

A book recommendation system is a type of recommendation system where we have to recommend similar books to the reader based on his interest. The books recommendation system is used by online websites which provide ebooks like google play books, open library, goodRead's, etc. In a very general way, recommender systems are algorithms aimed at suggesting relevant items to users (items being movies to watch, text to read, products to buy, or anything else depending on industries). Recommendation systems are used in hundreds of different services everywhere from online shopping to music to movies. Recommender systems are really critical in some industries as they can generate a huge amount of income when they are efficient or also be a way to stand out significantly from competitors.

The main objective of our project is to create book recommendation systems for users on various approaches.

## **Data preparation and Cleaning:**

- We began by importing all the necessary libraries to help us explore the problem statement and perform EDA to draw conclusions.
- Data wrangling is the first step we need to take.
- There are no duplicate values were found.
- It has some null objects.
- Checking Outlier.
- Replaced null values and treated outliers.
- There are some incorrect entries in data which was replaced with correct values
- After all the processing of the data
- The data we collected was analyzed.

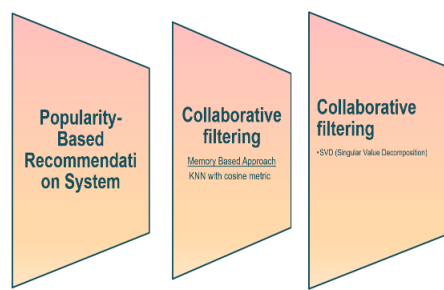
## **Exploratory Analysis and Visualization:**

In this section, we will explore the data to get insights about it. A summary statistic or graphical representation is used to detect patterns and anomalies in data, test hypotheses, and confirm assumptions during exploratory data analysis

## **Exploratory Analysis and Visualization(EDA) CONCLUSIONS:**

- ❖ The Lovely Bones: A Novel and Wild Animus are the two most read books.
- ❖ Most popular book author based on the number of ratings is Stephan King.
- ❖ Ballantine Books and Pocket are the top publishers based on the number of ratings that their books have received.
- ❖ The majority of readers are between the ages of 20 and 40.
- ❖ The majority of readers who have given the books ratings are from the United States and Canada.
- ❖ Regardless of the age group, The Lovely Bones and Wild animus appear on lists of the top-rated books.

## **Recommender System:**



## **Model Evaluation:**

	hits@5_count	hits@10_count	interacted_count	recall@5	recall@10	_person_id
36	35	89	545	0.064220	0.163303	11676
202	62	81	139	0.446043	0.582734	98391
271	27	35	93	0.290323	0.376344	153662
60	22	36	88	0.250000	0.409091	16795
474	16	28	73	0.219178	0.383562	95359
485	52	60	72	0.722222	0.833333	114368
390	32	32	61	0.524590	0.524590	104636
456	17	24	54	0.314815	0.444444	158295
660	34	43	54	0.629630	0.796296	123883
659	7	13	53	0.132075	0.245283	35859

## **Conclusion:**

- ❖ In EDA, the Top-10 most rated books were essentially novels.
- ❖ Books like The Lovely Bone and The Secret Life of Bees were very well perceived.
- ❖ Majority of the readers were of the age bracket 20-35 and most of them came from North American and European countries namely USA, Canada, UK, Germany and Spain.
- ❖ If we look at the ratings distribution, most of the books have high ratings with maximum books being rated 8.
- ❖ Ratings below 5 are few in number.
- ❖ Author with the most books was Agatha Christie, William Shakespeare and Stephen King.
- ❖ We evaluated the performance of Singular Value Decomposition based recommender and obtained a Global Recall@5 of 30 % and recall@10 of 41%

I.