LIBRARY BOOK RECOMMENDATION SYSTEM

ACKNOWLEDGEMENT

We would like to acknowledge with sincere gratitude and appreciation towards all those who helped/guided us for this mini project-2 report, such as Dr. Priyanka Sharma’s support and criticism towards our work. Also we would like to show appreciation for the faculties who helped us with areas that we were facing difficulties in.

ABSTRACT/ Outline

The Book Recommendation system as the name suggests is an application where we perform classification,segregation,etc and suggest books to users from their past and other books that are relevant. This report touches the basic introduction to an Recommendation system and later goes into explaining various methods etc and then explaining K-means algorithm and other working of the project.

CONTENTS

Introduction

1.1 General

Recommendation systems are algorithms aimed toward suggesting relevant things to users (which can be anything from being movies to observe, text to scan, product to shop for or the rest betting on industries and in this case suggesting books).

1.2 Scope of Study

Our project is meant to understand the patterns among readers and user that issue books and recommend them books from similar categories
and predict what else they might like

Literature Survey

2.1 General

The Literature Survey covered documents such as papers, authentic articles that introduced Recommendation systems, Various Machine learning techniques, and faults in them. The papers and articles are chosen in such a way that topics that were described in scope are covered.

About System

3.1 Core Ideas

3.1.1 **Base:** Recommend books based on previously issued books.

3.1.2 **Survey:** Ask a few questions to understand the interests of student from the book issued

3.1.3 **Rewards:** Reward students with points that'll motivate them to issue more books which in return will give us more data to analyze

- **False Positive:** Algorithm may register false positives when student issues some books which he/she may not want to be recommended.
- **Groups:** Creating groups based on similar interests so students can also see what books/materials are viewed by students in the same group.
- **Recommend Faculty:** Associate faculties with technology/subject that they have expertise or understanding on a deeper level.

3.2 Types of recommendation techniques

Machine learning algorithms in in these systems are usually classified into two categories — content based and collaborative filtering methods although modern recommenders combine both approaches. Content based methods are based on similarity of item attributes and collaborative methods calculate similarity from interactions.

Content Based

This Filtering approach takes makes use of a user's profile which is constructed based on his previous ratings.

Collaborative based filtering

In this filtering, recommendation for a user is decided by other users' profiles. Here recommendation occurs based on the ratings of other users who have similar interests as the user under consideration.

Hybrid Content-Collaborative Based Filtering

Combination of both and possess advantages of both the methods and is better than both of them.

K-mean Clustering

In the k-mean clustering, the similarity between the objects is calculated by the means of various distance measures such as Euclidean distance, Pearson Correlation, etc.

Naive Bayes

It is a conditional probability based
classifier. The prior knowledge of the classifier assists learning.
The naive assumption is that the features are conditionally
Independent .

Implementation

4.1 Technology to be used

-Jupyter Notebook

-Python (pandas,numpy,sklearn,matplotlib, etc)

-HTML,CSS,JS,Php,SQL

4.2 Jupyter Notebook

- It is a free and open source web based application that allows us to create documents that contain live code, equations, visualizations and narrative text. What we used is, statistical modeling, data visualization, machine learning, and much more.

4.3 Python

- Python is an interpreted, high-level, general-purpose programming language.
- We used various libraries and frameworks such as pandas for dataframe manipulation and visualization
- numpy for all statistical operations that was performed on the data.
- Sklearn was used for implementation of K means algorithm.
- Matplotlib was used to visualize the data.

4.4 HTML,CSS,Js,Php,SQL

- We used these technologies for developing the web based application after performing the operations for the users to view their recommendations

Pre-processing

5.1 Clean text & handling missing values

The dataset that was provided to us had several inconsistencies and missing values which we with the help of panda library cleaned and resolved the missing value issues

5.2 Generating tags

We needed to classify the books with tags for example a book about machine learning should have tags such as Machine Learning , Statistics,Artificial Intelligence etc.

For that we created a web scraper that would search and retrieve tags from a website.

Working of System

6.1 Walkthrough

After doing the preprocessing we started experimenting with different algorithms and found K means was much simpler and efficient to implement for the dataset that was provided

6.2 K Means

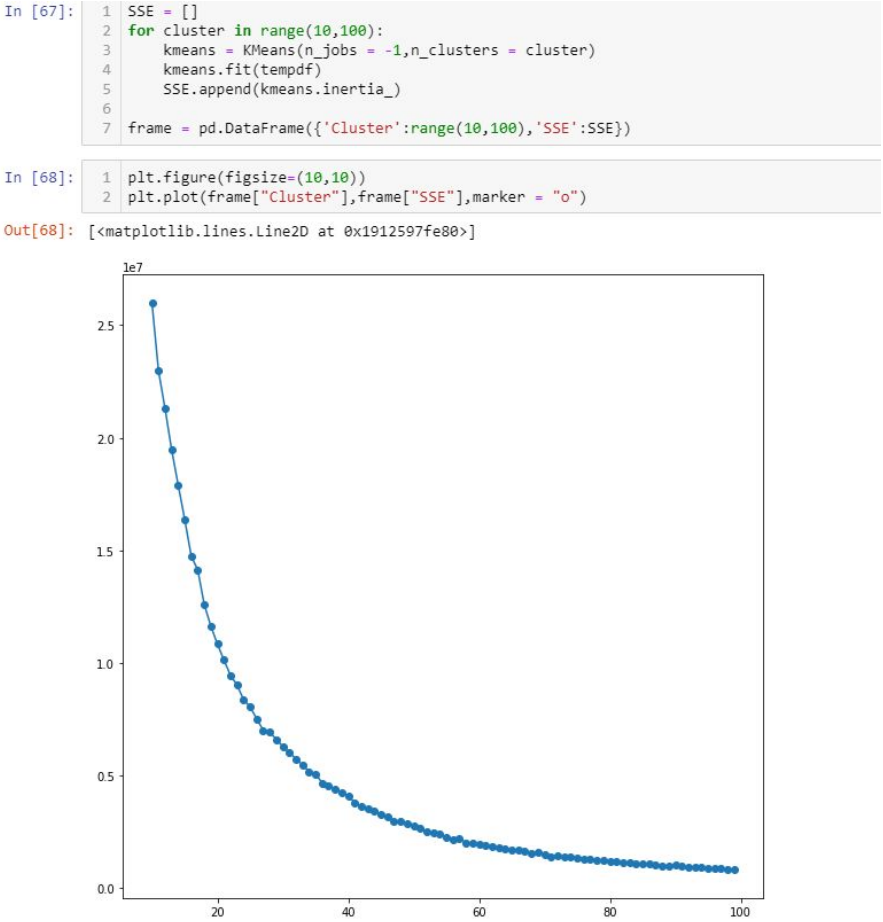
This is an iterative algorithm that does partition in the dataset into K already defined unique non-overlapping clusters where every point belongs to a single group. It will try to make the inter-cluster points as similar to other inter-clusters as possible while keeping them as far as possible. Then it will assign data points to a cluster such that the sum of the squared distance between all points and the cluster's centroid is at the least. The lesser the variation within clusters, the more similar the data points are inside the same cluster.

These are the steps when k means is performed.

- Initially we mention number of clusters that is K.
- Then we initialize the centroids by first randomizing the dataset and then selecting K data points for the centroids without any replacement.
- Keep iterating while change in centroids is occuring.

- Calculate sum of squared distance between points and all their centroids.
- Give each data point to the nearest centroid (clusters)
- Find the centroids for the clusters by averaging of all points that belong to every cluster.

We tried implementing K means but it requires numerical values so we then needed to convert text to numbers we did that using label encoders.



```
In [67]: SSE = []
         for cluster in range(10,100):
             kmeans = KMeans(n_jobs = -1,n_clusters = cluster)
             kmeans.fit(tempdf)
             SSE.append(kmeans.inertia_)

         frame = pd.DataFrame({'Cluster':range(10,100),'SSE':SSE})

In [68]: plt.figure(figsize=(10,10))
         plt.plot(frame["Cluster"],frame["SSE"],marker = "o")

Out[68]: [<matplotlib.lines.Line2D at 0x1912597fe80>]
```

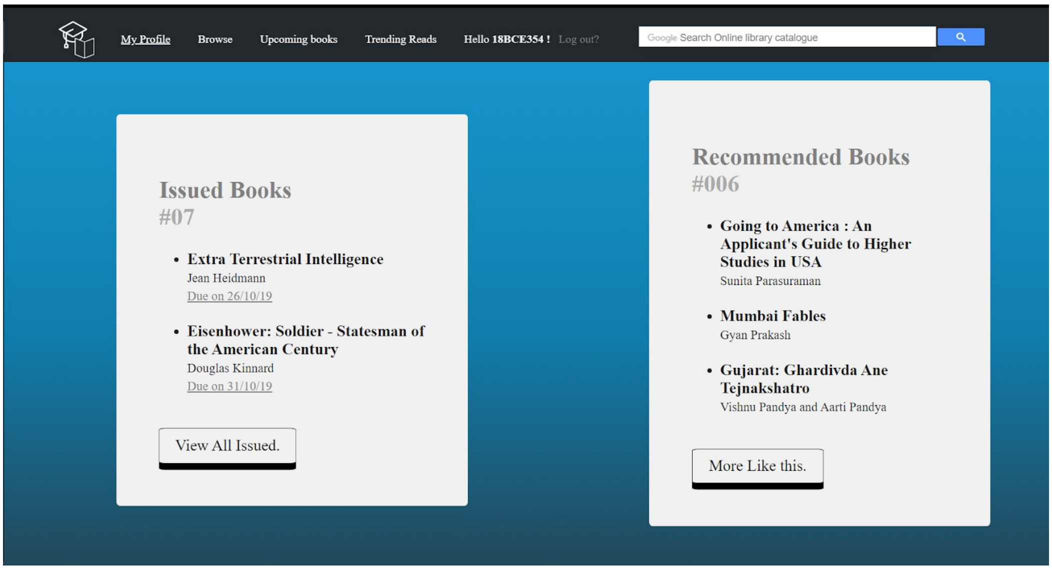
fig 6.1

After doing so we were able to implement K Means but we were getting an average score which was not really helpful so we tried and tested various other combinations and later found out that combining a column to identify student and a column to identify books was the way to go.

6.3 Website

Then we performed KMeans on it and later converted the values into scalar form and formed clusters again.

After doing so we were getting even better results which we exported as a table and used to show users what books are recommended to them and so on.


My Profile
Browse
Upcoming books
Trending Reads
Hello 18BCE354 ! Log out?
Google Search Online library catalogue

Issued Books
#07

- **Extra Terrestrial Intelligence**
  Jean Heidmann
  Due on 26/10/19
- **Eisenhower: Soldier - Statesman of the American Century**
  Douglas Kinnard
  Due on 31/10/19

View All Issued.

Recommended Books
#006

- **Going to America : An Applicant's Guide to Higher Studies in USA**
  Sunita Parasuraman
- **Mumbai Fables**
  Gyan Prakash
- **Gujarat: Ghardivda Ane Tejnakshatro**
  Vishnu Pandya and Aarti Pandya

More Like this.

fig 6.2 Dashboard

Summary & Conclusion

7.1 Summary

In this report we have discussed the basics of BOOK RECOMMENDATION SYSTEM and also covered what are concepts behind all of the functionality that is happening in the system/network.With which any beginner would find helpful for understanding and implementing similar recommendation system.

7.2 Conclusion

We can understand and realize the gravity and wide array of applications for recommendation systems and data science as there are no limits to the application,it can be used for something as simple as learning patterns to recommend books and learning patterns to catch bank fraud in advance.

References

1 Machine Learning Algorithms for Recommender System - a comparative analysis [Satya Prakash Sahu], [Anand Nautiyal],[Mahendra Prasad]

2 A Survey of Imperfection of Existing Recommender Systems [Nikhat Akhtar]

Appendix A - List of Useful Websites

Complete Methodology:-
https://www.analyticsvidhya.com/blog/2018/06/comprehensive-guide-recommendation-engine-python/
Video Lecture and ppt:-
http://technocalifornia.blogspot.com/2014/08/introduction-to-recommender-systems-4.html