Share Stories Using Collaborative Filtering

1st Shishir A
Computer Science And Engineering
New Horizon College of Engineering
Bangalore, India
shishirashok17@gmail.com

2nd Rahul B R
Computer Science And Engineering
New Horizon College of Engineering
Bangalore, India
rahulbargavram@gmail.com

3rd Pruthvi S
Computer Science And Engineering
New Horizon College of Engineering
Bangalore, India
pruthvisrinivas2119@gmail.com

Guided By: DR. S. J. Subhashini
Associate Professor
Computer Science And Engineering
New Horizon College of Engineering
Bangalore, India

***Abstract*—As recommender systems have such a broad scope, we'll focus on a key technique called collaborative filtering, which filters information by leveraging the system's interactions and data from other users. Here, Share stories use collaborative filtering to filter information based on the interactions and data gathered from other users. It's built on the notion thatpeople who agree on a given item's value are more likely toagree on it again in the future. The idea is simple: when we're looking for a new movie to watch, we frequently turn to our friends for recommendations. Naturally, we put more faith in recommendations from friends who have similar likes to us. The bulk of collaborative filtering systems employ the so-called similarity index-based methodology. A group of users who are similar to the current user is chosen using the neighborhood-based strategy. Hence in share stories,users can post, like, read and comment on blogs.**

***Index Terms*—Recommender system, content-based filtering, Collaborative filtering**

I. Introduction

In order to get relevant information, an average person spends about half of his time every single day.In information filtering systems, recommendation systems play a critical role in determining how to best offer things or information that are relevant to the user. A recommendation system, also known as a recommender system, can be used in a variety of situations. A recommender system is a system that filters information for its users to recommend those items, it filters the information and recommends the items. These systems or engines are commonly used in movie lenses, e-commerce sites, book-crossing and in the recent time, it is used in corporate firms to present information on products and items that are likely to be of interest to the clients and customers. The interest of the consumer in the past is seen and evaluated for some product suggestion. The recommendation system uses knowledge of the account of the registered user's profile, actions, interests and activities of their entire user community when providing the recommendations, and compares the information to presentthe recommendations. It is heavily reliant on similarity esti- mates.

The process of leveraging other people's opinions to filter or evaluate items is known as collaborative filtering. Although the term collaborative filtering (CF) has only been around for a little over a decade, CF draws its roots from something that people have been doing for centuries -sharing opinions with others. People have been standing over the back fence or in the office break room for years and debating booksthey've read, restaurants they've tried,and films they've seen then using these conversations to shape opinions. For instance, when enough of Amy's colleagues say they liked the latest Hollywood release, she might decide that she should see it as well. Similarly, she might decide to spend her money elsewhere if many of them considered it a disaster. Better still, Amy should remember that Matt recommends the kinds of films she finds fun, Paul has a history of recommending films she despises, and all seems to be recommended by Margaret. She learns over time what views she can listento and how these views can be applied to help her assessan item's quality.Computers and the internet allow us to step beyond easy word-of-mouth. The Internet helps us to considerthe views of thousands, instead of restricting ourselves to tens or hundreds of people. Computer speed enables us to process these views in real time and not only decide what a much wideraudience thinks of an object, but also to build a genuinely customized view of that item using the views that are most suitable for a particular user or group of users.

II. Related Work

A. Collaborative Filtering

Collaborative filtering suggests products to the user based on other users' ratings. The primary idea is to automate word-of-mouth marketing. The users give the items ratings first and the system compares these ratings with other users, and the user is recommended based on similar taste items. The downside of collaborative filtering is that the user is not adequately conscious to recommend things and it suffers from problems such as data sparsity and cold start issues. This technique is mainly used by e-commerce websites that recommend products based on ratings from other users. Quality items that are missing from results generated automatically by the system are therefore recommended.

Types of Collaborative Filtering

Collaborative filtering is mainly based on 2 Types of techniques :

- Memory-Based or User Based Collaborative Filtering
- Model- Based or Item Based Collaborative Filtering

B. Memory-Based or User Based Collaborative Filtering:

The memory-based collaborative filtering generates recommendations for users based on the full user-item database. The K-Nearest Neighbor algorithm is a popular method.. In neighborhood-based algorithms, the initial step is to select a subset of clients based on their proximity to the dynamic client, and then use a weighted combination of their appraisals to build item expectations for the dynamic client. This methodology is popular and well known for its straight forwardness and its productivity and also has been very flourishing in the past, but it has some difficulties like Scalability and Data Sparsity.

C. Model-Based or Item Based Collaborative Filtering:

The model based collaborative filtering uses the ratings provided to the items to recommend them to the users. Ratings of the items are given preferences. Clustering and rule based techniques are some of the well known techniques which are used. When dealing with sparse data, this particular typeof collaborative filtering techniques like clustering algorithms allow for more accurate predictions.

D. Hybrid Recommender Systems:

This recommender system employs a hybrid of content-based and collaborative filtering techniques. Hybrid techniques can be applied in a variety of ways, including creating separate content-based and collaborative-based forecasts and then integrating them; and adding content-based capabilities to a collaborative-based method (and vice versa) Hybrid methods may offer suggestions that are more detailed than pure approaches. In proposing systems such as cold start and the sparsity dilemma, these approaches can also be used to solve some of the common issues.

III. METHODOLOGY

Methodology is used to define the structure of execution for a project. A project with well-defined methodology will guarantee the proper flow of execution and will give the project a definitive timeline and goals to be reached. The methodology also specifies what steps are to be carried out in what order. This is crucial if the project is split among people and if one job is dependent on the completion of another. In this project the initial step is define the requirement specification which involves functional requirements, non-functional requirements, hardware and software requirements and the domain and UI specific requirements. The functional and non-functional requirements take all the stakeholders in picture.The next step would be to decide a user story that would benefit the end- users and developers appropriately and then move to the next step which is to create a design for the project.The design can be done by extensively studying similar products and reading papers that talks about work done in this domain. The main focus in this project is to work on Collaborative filtering applications and several papers provide inspiration and techniques to make this project a reality.The next step isto make use of all the information gathered and following the design to implement the project by deciding on which tools, frameworks and programming languages will be used to make the project a reality.The penultimate step would be to develop each part of the application in a modular fashion and realize the design step by step.The final step would be to test each partof the project through unit testing and then finally integrate each of the modules and perform an integrated test. On the final test if the project meets all the functional requirements and the non functional requirements then it can be deemed a successful project.

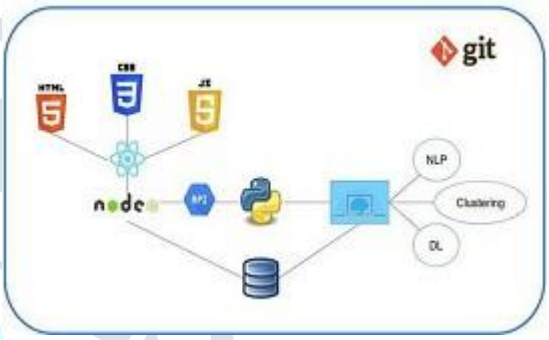

Fig. 1. System Architecture

- React : The main idea of using react is to having a dynamic component on the website and making the website more dynamic and making the website a single page application Respond is a segment based JavaScript library that powers you to think. This reasoning modelfits well with UIs.
- Node.js : It is a JavaScript runtime environment open-source and cross-platform. A Node.js software runs in a single process rather than establishing a new thread for each request.Node.js provides in its standard library a set of asynchronous input/output node libraries and primitives that prevent JavaScript code from stalling. Because js is typically built in non-blocking paradigms, blocking action is the exception rather than the rule.
- REST : REST API is a quick and versatile way to access web services without any coding.REST technology is generally preferred to the more robust Simple Object Access Protocol (SOAP) technology since REST usesless bandwidth,simplicity, and flexibility to make it more suitable for use on the Internet.

With all this in place the data transfer would look something like this:

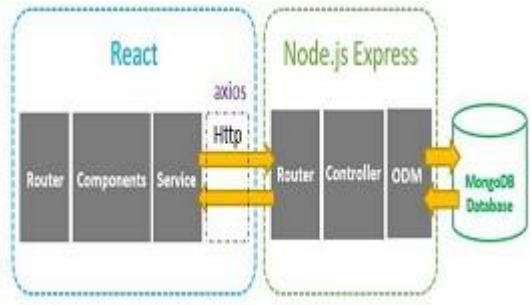

Fig. 2. Transfer of Data

A. Algorithm

The Kmeans algorithm is an iterative technique that attempts to split a dataset into K separate non-overlapping subgroups (clusters), each of which contains only one data point. This tries to make intra-cluster data points as comparable as possi- ble while still maintaining clusters as distinct, that is, as far as possible. This distributes data points to clusters in a manner that the sum of the squared distances between them and the cluster's centroid, which is the arithmetic mean of all the data points in that cluster, is as small as possible. Inside clusters, the less variance there is, the more homogenous the data pointsare.

The following is how the kmeans algorithm works:

- K is the number of clusters to specify.
- Initialize the centroids by shuffling the dataset and then picking K data points at random for the centroids without replacing them.
- Continue iterating until the centroids do not change. i.e. the clustering of data points does not change.
- Calculate the total of all data points' squared distances from all centroids.
- Provide each data point a cluster that is closest to it. This is the centroid.
- Calculate the cluster centroids by averaging all of the data points that correspond to each cluster.

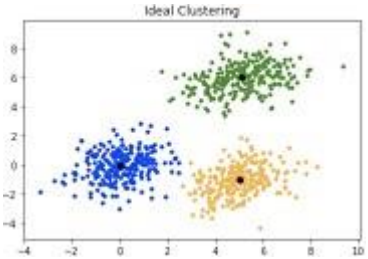

Fig. 3. K-means clustering

B. Authentication

For authenticating with the server application we have used JWT token for security purposes. We have used third party packages to build the jwt token and this is sent to the user and this jwt token is stored in the user local storage which is accessed every time the user needs information from the user. JSON Web Token (JWT) is an open standard that specifies a compact and self-contained method for securely communicating information as a JSON object among the two parties. Because it is digitally signed, this information can be checked and trusted. JWTs can be signed using a secret or an entrypted key. Although JWTs can be encrypted to guarantee party-to-party confidentiality, we will concentrate on signed tokens. Signed tokens can be used to validate the validity of the claims they contain, whilst encrypted tokens keep those claims hidden from third parties. When public/private key pairings are used to sign tokens, the signature additionally verifies that only the person with the private key signed it. JSON Web Tokens are made up of three pieces that are separated by dots (.). They are:

- Signature
- Header
- Payload

As a result, a JWT often looks like this: xxxxx.yyyyy.zzzzz

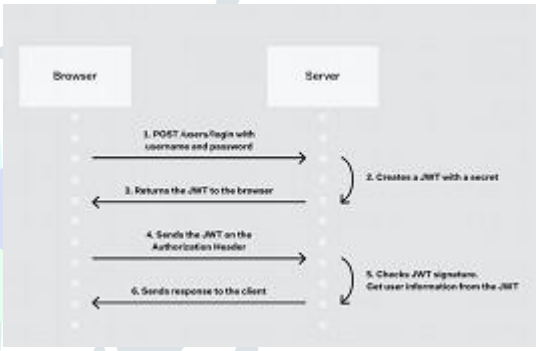

Fig. 4. JWT handshakes

We used the SHA256 technique to encrypt the token. We normally transmit the user id as the payload, and the jwt secret word is put in the env file so that no one else can see it.

IV. Future Work and Conclusions

Establish a centralized platform for users. Spam detection done using machine learning. Making the application accessi- ble to all kinds of people. Keeping the UI extremely simple in order to enhance the experience of the user. Using Machine Learning models to provide accurate as well as the best content for the user. Having the platform safe from unwanted information by using the concepts of sentimental analysis and filtering. This application demonstrates the integrationof machine learning into the social media domain for the improvement of the social sector. We implement numerous machine learning concepts like clustering, natural language processing, classification and Image processing to get the most accurate and realistic model of an actual recommender system. We use a machine learning methodology that is based on a

large number of attributes derived from social media data. The project also takes care of spam with its integrated spam detection using the NLP.

REFERENCES

[1] AlZu'bi, S., Hawashin, B., ElBes, M., and Al-Ayyoub, M. (2018). A novel recommender system based on apriori algorithm for requirements engineering. Fifth International Conference on Social Networks Analysis, Management and Security (SNAMS), (pp. 323-327) Irbid, Jordan. IEEE.

[2] Ramos, F. B. A., Costa, A. A. M., Perkusich, M.Almeida, H. O., and Perkusich, A. (2018). A Non-Functional Requirements Recommendation System for Scrum-based Projects.

[3] Palomares, C., Franch, X., and Fucci, D. (2018, March). Personal recommendations in requirements engineering: the OpenReq approach. In International Working Conference on Requirements Engineering: Foundation for Software Quality. Springer, Cham.

[4] Matyokurehwa, K., Mavetera, N., and Jokonya, O. (2017). Requirements Engineering Techniques: A Systematic Literature Review. International Journal of Soft Computing and Engineering

[5] Anand Shanker Tewari ; Abhay Kumar ; Asim Gopal Barman, 2014, "Book recommendation system based on combined features of content based filtering, collaborative filtering and association rule mining", 2014 IEEE International Advance Computing Conference (IACC), PP: 500-503.

[6] Turgu, Ertan & Komuscu, Ali Umran. (2019). Using K-Means Methodology for Reclassifying Rainfall Regions of Turkey.

[7] Pasquale Lops, Marco de Gemmis and Giovanni Semeraro, "Content-based Recommender Systems: State of the Art and Trends" Springer Science+Business Media, LLC 2011.

[8] Gilbert Badaro, Hazem Hajj, Wassim El-Hajj and Lama Nachman, "A Hybrid approach with collaborative filtering for recommender systems",9th International Wireless Communications and Mobile Computing Conference (IWCMC), 2013.

[9] Anil Poriya, Neev Patel, Tanvi Bhagat and Rekha Sharma, Ph. D, "Non-Personalized Recommender Systems and User-based Collaborative Recommender Systems",International Journal of Applied Information Systems (IJAIS) – ISSN : 2249-0868 Foundation of Computer Science FCS, New York, USA Volume 6– No. 9, March 2014.

[10] Luo Zhenghua, "Realization of Individualized Recommendation System on Books Sale", IEEE 2012 International Conference on Management of e-Commerce and e-Government.