International Journal of Advances in Engineering and Management (IJAEM)
Volume 3, Issue 3 Mar. 2021, pp: 1150-1155 www.ijaem.net ISSN: 2395-5252

Rating Based Product Recommendation on Hybrid Filtering Algorithm in Online Application

M.Shobana[1], N.Anuja[2], V.Jenifar[3], M.Jersini[4], R.Keerthana[5]
[1] *Assistant Professor, Department of computer science , Dhanalakshmisrinivasan engineering college*
[2,3,4,5] *Department of computer science , Dhanalakshmisrinivasan engineering college*

Submitted: 10-03-2021 Revised: 30-03-2021 Accepted: 01-04-2021

ABSTRACT- E-Commerce has been known as a rapidly growing commercial enterprise, and even though on line purchasing has no longer accompanied those identical boom patterns within the beyond, it's miles now being diagnosed for its capability. Sentiment evaluation is one of the current research subjects in the subject of textual content mining. Opinions and sentiments mining from natural language are very difficult task. Sentiment analysis is the best solution. This gives important information for decision making in various domains. Various sentiment detection methods are available which affect the quality of result. In this project we are finding the sentiments of people related to the services of E-shopping websites. The sentiments include reviews, ratings and emoticons. The main goal is to recommend the products to users which are posted in E-shopping website and analyzing which one is the best. For this we use hybrid learning algorithm which analyze various feedbacks related to the services. Text mining algorithm is used to find scores of each word. Then sentiments are classified as negative, positive and neutral. It has been observed that the pre-processing of the data is greatly affecting the quality of detected sentiments. Finally analysis takes place based on classification. To find out fake review in the website can be analyzed. This device will discover fake critiques made via posting fake remarks about a product via figuring out the MAC deal with in conjunction with assessment posting styles. User will login to the device using his consumer identification and password and could view various merchandise and will give assessment approximately the product. To discover the evaluation is fake or authentic, system will find out the MAC address of the consumer if the machine observes fake assessment send by way of the identical MAC Address many a times it'll inform the admin to do away with that overview from the device. This gadget uses information mining technique. This machine allows the user to find out accurate overview of the product.

Index Terms- Hybrid Filtering Algorithm, recommender system,E-commerce, Sentiment analysis,E-shopping, user preferences.

I.INTRODUCTION

Big Data is a collection of data that is huge in volume, yet growing exponentially with time. It is a data with so large size and complexity that none of traditional data management tools can store it or process it efficiently. Big data is also a data but with huge size.Following are the types of Big Data: Structured,Unstructured,Semi-structured,The
Structured Any data that can be stored, accessed and processed in the form of fixed format is termed as a 'structured' data. Any data with unknown form or the structure is classified as unstructured data. In addition to the size being huge, un-structured data poses multiple challenges in terms of its processing for deriving value out of it. Semi-structured data can contain both the forms of data. We can see semi-structured data as a structured in form but it is actually not defined with e.g. a table definition in relational DBMS.

II. RELEVANT WORK

Big data can be described by the following characteristics: Volume,Variety,Velocity,Variability
(i) Volume – The name Big Data itself is related to a size which is enormous. Size of data plays a very crucial role in determining value out of data. Also, whether a particular data can actually be considered as a Big Data or not, is dependent upon the volume of data. Hence, **'Volume'** is one characteristic which needs to be considered while dealing with Big Data.
(ii) Variety – The next aspect of Big Data is its **variety**.
Variety refers to heterogeneous sources and the nature of data, both structured and unstructured. During earlier days, spreadsheets and databases

International Journal of Advances in Engineering and Management (IJAEM)
Volume 3, Issue 3 Mar. 2021, pp: 1150-1155 www.ijaem.net ISSN: 2395-5252

were the only sources of data considered by most of the applications. Nowadays, data in the form of emails, photos, videos, monitoring devices, PDFs, audio, etc. are also being considered in the analysis applications. This variety of unstructured data poses certain issues for storage, mining and analyzing data.

(iii) Velocity – The term **'velocity'** refers to the speed of generation of data. How fast the data is generated and processed to meet the demands, determines real potential in the data.Big Data Velocity deals with the speed at which data flows in from sources like business processes, application logs, networks, and social media sites, sensors, Mobile devices, etc. The flow of data is massive and continuous.

(iv) Variability – This refers to the inconsistency which can be shown by the data at times, thus hampering the process of being able to handle and manage the data effectively.

Application of Big Data:

1) **Government:**The use and adoption of big data within governmental processes allows efficiencies in terms of cost, productivity, and innovation, but does not come without its flaws. Data analysis often requires multiple parts of government (central and local) to work in collaboration and create new and innovative processes to deliver the desired outcome. A common government organization that makes use of big data is the National Security Administration (NSA), who monitor the activities of the Internet constantly in search for potential patterns of suspicious or illegal activities their system may pick up.Civil registration and vital statistics (CRVS) collects all certificates status from birth to death. CRVS is a source of big data for governments.
2) **International Development:**Research on the effective usage of information and communication technologies for development (also known as "ICT4D") suggests that big data technology can make important contributions but also present unique challenges to international development. Advancements in big data analysis offer cost-effective opportunities to improve decision-making in critical development areas such as health care, employment, economic productivity, crime, security, and natural disaster and resource management. Additionally, user-generated data offers new opportunities to give the unheard a voice. However, longstanding challenges for developing regions such as inadequate technological infrastructure and economic and human resource scarcity exacerbate existing concerns with big data such as privacy, imperfect methodology, and interoperability issues. The challenge of "big data for development"is currently evolving toward the application of this data through machine learning, known as "artificial intelligence for development (AI4D).
3) A major practical application of big data for development has been "fighting poverty with data". In 2015, Blumenstock and colleagues estimated predicted poverty and wealth from mobile phone metadata and in 2016 Jean and colleagues combined satellite imagery and machine learning to predict poverty. Using digital trace data to study the labor market and the digital economy in Latin America, Hilbert and colleagues argue that digital trace data has several benefits such as:

- Thematic coverage: including areas that were previously difficult or impossible to measure
- Geographical coverage: our international sources provided sizable and comparable data for almost all countries, including many small countries that usually are not included in international inventories
- Level of detail: providing fine-grained data with many interrelated variables, and new aspects, like network connections
- Timeliness and timeseries: graphs can be produced within days of being collected

4) **Education:** A McKinsey Global Institute study found a shortage of 1.5 million highly trained data professionals and managers and a number of universities including University of Tennessee and UC Berkeley, have created masters programs to meet this demand.
5) **Media:**To understand how the media uses big data, it is first necessary to provide some context into the mechanism used for media process. It has been suggested by Nick Couldry and Joseph Turow that practitioners in media and advertising approach big data as many actionable points of information about millions of individuals.
6) **Insurance:**Health insurance providers are collecting data on social "determinants of health" such as food and TV consumption, marital status, clothing size and purchasing habits, from which they make predictions on health costs, in order to spot health issues in their clients. It is controversial
7) **IOT:**Big data and the IoT work in conjunction. Data extracted from IoT devices provides a

International Journal of Advances in Engineering and Management (IJAEM)
Volume 3, Issue 3 Mar. 2021, pp: 1150-1155 www.ijaem.net ISSN: 2395-5252

mapping of device inter-connectivity. Such mappings have been used by the media industry, companies and governments to more accurately target their audience and increase media efficiency. The IoT is also increasingly adopted as a means of gathering sensory data, and this sensory data has been used in medical, manufacturing and transportationcontexts.

8) **Information Technology:** Especially since 2015, big data has come to prominence within business operations as a tool to help employees work more efficiently and streamline the collection and distribution of information technology (IT). The use of big data to resolve IT and data collection issues within an enterprise is called IT operations analytics (ITOA).

III.RECOMMENDER SYSTEMS

Recommender systems or recommendation systems are a subclass ofinformation filtering system that seek to predict the "rating" or "preference" that a user would give to an item. Recommendation System are information filtering system that deal with the problem of information overload by filtering vital information fragment out of large amount of dynamically generated information according to user's preferences, interest (or) observed behaviour about item. Recommender System has the ability to predict whether a particular user would prefer an item (or) not based on user's profile. Recommender Systems are beneficial to both service provider and user. They reduce transaction costs of finding and selecting items in an online shopping environment. Recommender systems have become increasingly popular in recent years, and are utilized in a variety of areas including movies, music, news, books, research articles, search queries, social tags, and products in general. Recommendation System has also proved to improve decision making process and quality. In e-commerce setting, recommender system enhances revenues, for the fact that they are effective means of selling more products. In scientific libraries, recommender system support users by allowing them to move beyond catalog searches. Therefore, the need to use efficient and accurate recommendation techniques within a system that will provide relevant and dependable recommendations for cannot be over-emphasized. In general, Recommender systems are classified as Collaborative Filtering (CF), Content Based and Hybrid recommender systems. CF is widely used in RS, and this recommendation can be divided into User-Based and Item-Based.

Recommender System consists of two types, they are

1. content based filtering
2. collaborative based filtering

Content-based filtering also referred to as cognitive filtering, recommends items based on a comparison between the content of the items and a user profile. The content of each item is represented as a set of terms. The user profile is represented with the same terms and built up by analyzing the content of items which have been seen by the user. This method provides recommendations based on features of users or items, however it is difficult to extract in some circumstances. Several issues have to be considered when implementing a content-based filtering system. First, terms can either be assigned automatically or manually. When terms are assigned automatically a method has to be chosen that can extract these terms from items. Second, the terms have to be represented such that both the user profile and the items can be compared in a meaningful way. Third, a learning algorithm has to be chosen that is able to learn the user profile based on seen items and can make recommendations based on this user profile.

Collaborative filtering also referred to as social filtering, filters information by using the recommendations of other people. It is based on the idea that people who agreed in their evaluation of certain items in the past are likely to agree again in the future. A person who wants to see a movie for example, might ask for recommendations from friends. The recommendations of some friends who have similar interests are trusted more than recommendations from others. This information is used in the decision on which product to see. This method is used to overcome the difficulty in content based filtering. They just use known ratings of item made by users to predict ratings of new user-item pairs. The problem arises in using a collaborative filtering are scalability and sparseness.

STEPS :

Online E-Commerce Framework:

E-commerce framework is used to buy the products in online to easy retrieval the mobile products. This module is used to create android and web site for recommending best mobiles in specific area. Admin is the responsibility for maintaining the all details in server and server can be design in server. There are two accounts such as admin and user account. Admin can login to the system and post item details with expiry dates. User can login to

International Journal of Advances in Engineering and Management (IJAEM)
Volume 3, Issue 3 Mar. 2021, pp: 1150-1155 www.ijaem.net ISSN: 2395-5252

the mobile to choose the language and area. Then view the products with specified filter. This module is used to create web site buy or post products for users. Admin can login to the system and post products with features. User can login to the system to view product details.

Reviews Collection:

Admin collect reviews and have various types of reviews. Reviews may be rating reviews, text reviews and smileys reviews. All reviews are stored in database for future evaluation. Ratings, reviews and emoticons are stored in database. Rating, Reviews and Emoticons are the evaluation or assessment of something, in terms of quality (as with a critic rating a novel), quantity or some combination of both.

Sentiment Analysis:

Sentiment analysis refers to the use of natural language processing, text analysis, computational linguistics to systematically identify, extract, quantify, and study affective states and subjective information. Sentiment analysis is widely applied to voice of the customer materials such as reviews and ratings for applications that range from marketing to customer service to buy the products efficiently. Admin can analyze whether the product is positive or negative. In star rating, we can calculate star count values. In text reviews, extract keywords and matched with database. Then smileys reviews are calculated based positive and negative symbols.

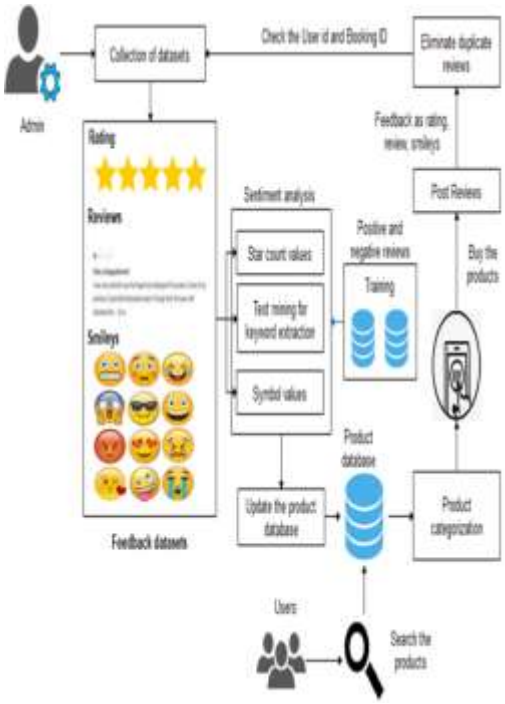
Recommendation System:

Recommender systems are a subclass of information filtering system that seek to predict the "rating" or "preference" that a user would give to an item. User can search the product in search bar. And view the list of products based on price and review details. Implement the stochastic learning algorithm to classify the products such as positive or negative. Positive products are display in recommendation panel based on ratings and reviews. If the product has negative review means, automatically the positive products in recommendation panel.

Fake Reviews Monitoring:

In this module, fake reviews are analyzed by admin. A media access control address (MAC address) of a computer is a unique identifier assigned to network interfaces for communications at the data link layer of a network segment. Admin can get user account details, Mobile address and Order id details. So user can post one reviews that will be genuine reviews.

IV. SYSTEMDESIGN:



V. SOFTWARE TESTING

Software testing is a method of assessing the functionality of a software program. There are many different types of software testing but the two main categories are dynamic testing and static testing. Dynamic testing is an assessment that is conducted while the program is executed; static testing, on the other hand, is an examination of the program's code and associated documentation. Dynamic and static methods are often used together.

Testing is a set activity that can be planned and conducted systematically. Testing begins at the module level and work towards the integration of entire computers based system. Nothing is complete without testing, as it is vital success of the system.

Testing Objectives:

There are several rules that can serve as testing objectives, they are

1. Testing is a process of executing a program with the intent of finding an error
2. A good test case is one that has high probability of finding an undiscovered error.
3. A successful test is one that uncovers an undiscovered error.

If testing is conducted successfully according to the objectives as stated above, it would uncover errors in the software. Also testing demonstrates that software functions appear to the

International Journal of Advances in Engineering and Management (IJAEM)
Volume 3, Issue 3 Mar. 2021, pp: 1150-1155 www.ijaem.net ISSN: 2395-5252

working according to the specification, that performance requirements appear to have been met. There are three ways to test a program

1. For Correctness
2. For Implementation efficiency
3. For Computational Complexity.

Tests for correctness are supposed to verify that a program does exactly what it was designed to do. This is much more difficult than it may at first appear, especially for large programs.

Tests used for implementation efficiency attempt to find ways to make a correct program faster or use less storage. It is a code-refining process, which reexamines the implementation phase of algorithm development. Tests for computational complexity amount to an experimental analysis of the complexity of an algorithm or an experimental comparison of two or more algorithms, which solve the same problem.

The data is entered in all forms separately and whenever an error occurred, it is corrected immediately. A quality team deputed by the management verified all the necessary documents and tested the Software while entering the data at all levels. The development process involves various types of testing. Each test type addresses a specific testing requirement. The most common types of testing involved in the development process are:

- Unit Test
- Functional Test
- Integration Test

Unit Testing

The first test in the development process is the unit test. The source code is normally divided into modules, which in turn are divided into smaller units called units. These units have specific behavior. The test done on these units of code is called unit test. Unit test depends upon the language on which the project is developed. Unit tests ensure that each unique path of the project performs accurately to the documented specifications and contains clearly defined inputs and expected results.

Functional Testing

Functional test can be defined as testing two or more modules together with the intent of finding defects, demonstrating that defects are not present, verifying that the module performs its intended functions as stated in the specification and establishing confidence that a program does what it is supposed to do.

Integration Testing

In integration testing modules are combined and tested as a group. Modules are typically code modules, individual applications, source and destination applications on a network, etc. Integration Testing follows unit testing and precedes system testing. Testing after the product is code complete. Betas are often widely distributed or even distributed to the public at large in hopes that they will buy the final product when it is released.

CONCLUSION

In this project, we have presented a novel implementation of a product recommendation system based on hybrid recommendation algorithm. The main advantages of our method are a visual organization of the data based on the underlying structure, and a significant reduction in the size of the search space per result output. And user can easily search the products anywhere and anytime. Ratings, reviews and emoticons are analyzed and categorized as positive and negative sentiments. Search the products based on price based filtering and reviews based filtering. MAC based filtering approach can be used to avoid fake reviews. Supermarket can benefits because easy buying, easy transactions and to get more customers. Our method was evaluated against real user data collected through an online website, by using a subset of the movies liked by each user as input to the system. The current results are notably better than random approach. However, we feel that with a better dataset and a number of improvements to our method, we may achieve better results. Hybrid Recommendations is one of the main modules of the system which helps overcome the drawbacks of the traditional Collaborative and Content Based Recommendations. We have obtained promising results using our current model.

We can extend the work with number of directions our work can potentially take in the future. Challenging task in opinion mining is each user expresses their feedback in different language such as English, Arabic, and French etc. To analyse aspect keyword from the user feedback is quite difficult for many language.

REFERENCES

[1]. Koo, Dong-Mo. "Impact of tie strength and experience on the effectiveness of online service recommendations." Electronic Commerce Research and Applications 15 (2016): 38-51.
[2]. Bobadilla, Jesús, et al. "Recommender systems survey." Knowledge-based systems 46 (2013): 109-132.

[3]. Moradi, Parham, and SajadAhmadian. "A reliability-based recommendation method to improve trust-aware recommender systems." Expert Systems with Applications 42.21 (2015): 7386-7398.
[4]. Wang, Gang, et al. "Sentiment classification: The contribution of ensemble learning." Decision support systems 57 (2014): 77-93.
[5]. Polatidis, Nikolaos, and Christos K. Georgiadis. "Recommender systems: The Importance of personalization in E-business environments." International Journal of E-Entrepreneurship and Innovation (IJEEI) 4.4 (2013): 32-46.
[6]. Anand, Deepa, and Kamal K. Bharadwaj. "Utilizing various sparsity measures for enhancing accuracy of collaborative recommender systems based on local and global similarities." Expert systems with applications 38.5 (2011): 5101-5109.
[7]. Zhao, Xin, et al. "Jointly modeling aspects and opinions with a MaxEnt-LDA hybrid." ACL, 2010.
[8]. Lin, Chenghua, et al. "Weakly supervised joint sentiment-topic detection from text." IEEE Transactions on Knowledge and Data engineering 24.6 (2011): 1134-1145.
[9]. Nguyen, Thien Hai, and KiyoakiShirai. "Topic modeling based sentiment analysis on social media for stock market prediction." Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers). 2015.
[10]. Bobadilla, JesúS, et al. "A collaborative filtering approach to mitigate the new user cold start problem." Knowledge-based systems 26 (2012): 225-238.

AUTHORS

First Author – M.Shobana, M.E, Dhanalakshmi Srinivasan Engineering college.
Second Author –N.Anuja, Department of cse, Dhanalakshmi Srinivasan Engineering College.
Third Author – V.Jenifar, Department of cse, Dhanalakshmi Srinivasan Engineering College.
Fourth Author – M.Jersini, Department of cse, Dhanalakshmi Srinivasan Engineering College.
Fifth Author – R.Keerthana, Department of cse, Dhanalakshmi Srinivasan Engineering College.