Amazon product recommendation system using machine learning model based approach

Gowtham Paluri
700741225
dept.Computer Science
University of Central Missouri
gxp12250@ucmo.edu

Ashok Sai Sudireddy
700734963
dept.Computer Science
University of Central Missouri
axs49630@ucmo.edu

Rithviksena Reddy Malepally
700737712
dept.Computer Science
University of Central Missouri
rxm77120@ucmo.edu

[1] ***Abstract*—Recommendation system estimates the user preferences based on the items purchased, users implicit rating and explicit rating. A good recommender system on an e-commerce site helps the users to find the best products according to their preferences automatically. In this project we are building a product recommendation system using Amazon electronics products dataset. To implement recommendation systems we have multiple techniques like memory based and model based. Model based approaches are easy to evaluate the results.To recommend the products we have used two techniques: Popularity based recommendation—common recommendation User - Item collaborative filtering method—-personalized recommendation. User- item collaborative filtering is implemented using the Matrix factorisation method ie. model based collaborative filtering SVD(Singular Value Decomposition) method.Experimental analysis is conducted on Amazon electronics products.The original source of the data is Recommender Systems and Personalization Datasets repository. The dataset contains 7 Million data points but due to memory constraints we have conducted the experiments on the subset of the data i.e. 5000 data points.**

[2] ***Index Terms*—Memory based recommendation system, model based recommendation system, collaborative filtering, hybrid filtering, SVD(Singular Value Decomposition) and Matrix factorisation.**

I. INTRODUCTION

A recommendation engine is a filtering system that examines data from many sources that belong to various users and generates solutions to predict their interests and offer the appropriate products to the appropriate users in accordance. The recommendation system, on the other hand, uses a machine learning algorithm to suggest things to consumers based on their prior preferences.

Recommendation systems are gaining popularity to recommend the products on e-commerce sites. In this paper we are reviewing various recommendation systems to recommend the products. Recommending the products purely on reviews gives irrelevant searches to the users, they are not inclined to user preferences. So we are reviewing other techniques to filter the results collaborative filtering methods using model based algorithms to analyze the predictive modeling results .

With the increase of online business everyday buying the products are just a click away. So recommending the products is becoming automated. But which method is effective is a challenging task. In this paper we are proposing a collaborative filtering model based technique using KNN algorithm is used.

In collaborative filtering one of the effective types of filtering is item-item based filtering. This method recommends the products to the user based on the similar users preferences. This experimental analysis was conducted on an Amazon product dataset. Unlike traditional algorithms this method gives real-time recommendations, high quality recommendations.

Online grocery shopping has grown significantly in popularity in recent years, and services like Instacart, Amazon Fresh, Shipt, and Walmart Grocery have drawn millions of users. It is crucial to offer pertinent, individualised advice and make the purchasing process simple for customers in order to meet their needs. In this article, we offer a tensor-based approach that makes use of a three-mode tensor to represent product-to-product interactions for users and using tensor decomposition techniques to simultaneously learn user and product embeddings that can be used to infer recommendations for products within baskets. Co-purchased goods are modelled as tensors when they are part of a single transaction. The latent components that indicate the innate user and product interactions are then captured using the RESCAL tensor decomposition technique. Our suggested tensor-based approach on the Instacart dataset.

Collaborative recommendation systems compile user reviews or suggestions of things, identify user relationships based on their reviews, and create new recommendations by comparing user reviews. In most cases, it is based on gathering and examining data regarding the user's actions, interests, or preferences in order to forecast what they will like based on similarities with other users. Comparative user and/or item analysis is done by collaborative systems.

A good example of the application of hybrid recommender systems is Netflix. The website offers films that share qualities with films that a user has rated highly (content-based filtering) as well as by analysing the viewing and searching habits of comparable users (collaborative filtering).

The filtering method is based on the interactions and preferences of a single user. It is founded on metadata obtained from user interactions and history. Based on the characteristics of the things the user has evaluated, a content-based suggestion

[1]https://github.com/lxk89990/FinalProject
[2]https://github.com/lxk89990/FinalProject

builds a profile of the new user's interests. The algorithms are set up such that they propose related products to users who have previously or are presently seeing items they liked.

II. Motivation

Recommendation system estimates the user preferences based on the items purchased, users implicit rating and explicit rating. A good recommender system on an e-commerce site helps the users to find the best products according to their preferences automatically. In this project we are building a product recommendation system using Amazon electronics products dataset. To implement recommendation systems we have multiple techniques like memory based and model based. Model based approaches are easy to evaluate the results.

III. Objectives

Content based filtering similarity between the products. Based on user history he searched the products he gets the recommendations using the similarity.But main challenges in this system are lack of novelty the system continuously recommends similar items. This method is not scalable because everytime a new product gets added to the products list it needs to be tagged. The other challenge is incorrect attributes information leads to incorrect recommendations. So we are opting for collaborative filtering in this paper. Memory based recommendation systems can not recommend the real-time recommendations and these systems are not fast and scalable. To overcome these challenges we are implementing a model based recommendation system.

To recommend the products we have used two techniques: Popularity based recommendation—common recommendation User - Item collaborative filtering method—-personalized recommendation User- item collaborative filtering is implemented using the Matrix factorisation method ie. model based collaborative filtering SVD(Singular Value Decomposition) method.

IV. Related work

In the study, recommendation systems (RS) have gained a lot of attention since they aim to help customers find products online by providing options that closely match their preferences. Customers might be recommended an irrelevant product if recommendations are made only based on quantitative reviews. Online retailers like Amazon and Flipkart utilise a variety of recommendation algorithms to present different options to various customers. Item-to-item collaborative filtering, used by Amazon, spreads to huge data sets and generates high-quality real-time suggestions. A user-friendly list of suggestions is created using the results of this type of filtering, which compares the items that customers have purchased and rated against related ones [15].

Every day, new items are released on the market as a result of the development of technology. The increase in items has sparked rivalry amongst various firms. Industries employ a variety of tactics to improve the consumer appeal of their products and maintain market share. The success of the top competitive items on the market that have comparable features must thus be examined before introducing any new product. For the new product's success, this information may be helpful. An strategy called variance-based product recommendation (VPR), which seeks to identify top rivals for every recently introduced product with a comparable description, is suggested in light of this problem. VPR is based on a product's ratings [10].

When producing product suggestions, the majority of conventional recommendation algorithms rely on user-shared grading criteria. The recommendation effect is poor and the data are quite scarce. As a result, a better algorithm for collaborative filtering recommendations is suggested. The system builds a user product network based on user purchase records, extracts the low-dimensional embedded semantic relationship between user and product nodes, and calculates the cosine similarity to determine the semantic similarity between items. The topic characteristics of the products are then produced using the hidden Dirichlet topic distribution model, and the cosine similarity is utilised to determine how similar the topic features are between the products. The data sparse problem is effectively resolved and enhanced by the use of the linear fusion approach [7].

When recommending products, the majority of conventional recommendation algorithms use user-level score criteria. As a result of the extremely scant data, there are poor recommendations. This leads to the proposal of an enhanced collaborative filtering recommendation algorithm. The algorithm creates a user product network from user purchase records using a representation learning method. It then determines the low-dimensional embedded semantic relationship between user and product nodes and utilises cosine similarity to determine how semantically similar the goods are to one another. The topic features of the products are then produced in accordance with the hidden Dirichlet topic distribution model, and the topic feature similarity between the goods is determined using the cosine similarity formula [2].

The rapid expansion of the e-commerce sector has increased the importance of recommendation systems. Utilising user feedback, recommendation systems make product recommendations that may be helpful to the user and assist in finding long-tail products. Users' ratings are the foundation of conventional recommendation systems. However, thanks to improvements in data collection, the majority of e-commerce websites today also collect additional beneficial feedback, such as reviews and reviews' helpfulness. The method suggested in this study uses user reviews to enhance the performance of recommendation systems. The tests are conducted using the product ratings and reviews from Amazon as the data source. The recall and root mean square error (RMSE) scores of the proposed recommendation system have improved when compared to the conventional rating-based recommendation system [19] [5].

Many e-commerce companies today allow customers to sign in using their social network accounts, such as Twitter or Facebook. The main goal of this is to increase social networks' acceptance of e-commerce. This resulted in the present recommendation systems on e-commerce websites failing to suggest products to consumers who sign in using social network accounts. The term ”cross website cold start product recommendation” refers to this issue. In a cold start circumstance, either the product or the consumer is new, and the recommendation system has no prior knowledge of them, hence it is unable to make recommendations. Our method uses microblogging reviews of content gathered from the social media network Twitter to address this issue. Text reviews in Hindi and English as well as audio or video reviews are taken into consideration while making recommendations [9].

The study discusses the various recommendation approaches used in the present e-commerce website and discusses the lack of a semantic component in recommendation systems. Three sorts of recommendation systems can be broadly distinguished: content-based, collaborative, and hybrid approaches. Content-based systems view the goods' qualities as recommendations. For instance, if an Amazon customer has purchased a lot of romantic novels, the database's content-based recommendation system will suggest other novels as belonging to the ”romantic” genre. Items are recommended by collaborative filtering systems based on measurements of similarity between like-minded users and/or items. The products that are suggested to a user are ones that similar users prefer. In order to accurately propose products, this study also emphasises the necessity of semantics in the existing recommendation system [16].

Deep learning-based image identification has made successful attempts to address Computer Vision issues. Tasks requiring face detection, facial expression recognition, age, and gender determination have benefited most from deep learning algorithms. The researchers have assessed the algorithms' performances, and they are now making them available as cloud services. Application developers can use pre-tested algorithms in their work and pay for the used service by gaining access to the cloud. Web service Amazon Rekognition offers highly accurate facial analysis that is better suited for facial recognition. Any image file saved in Amazon S3 can be rapidly analysed using the straightforward, user-friendly API that comes with Amazon Rekognition [18] [20].

Since many e-commerce applications are heavily populated with the textual description of the product, text-based product similarity through text vectorization technique is very helpful in performing content-based product recommendation and recommending the similar item to the user. A better suggestion of the product results from the combination of bag of words and TF-IDF, which creates an n-dimensional vector with a different textual description of the product. The experimental findings and analysis section explains in detail how the suggested model for text-based items determines which products are similar to the product being searched for and outputs the best-recommended product [17] [8].

The recommendation problem is typically solved using one of three methods: classic collaborative filtering, cluster models, or search-based techniques. We contrast these approaches with our algorithm, which we refer to as item-to-item collaborative filtering, in this instance. Our algorithm's online computing scales regardless of the number of consumers and the number of products in the product catalogue, in contrast to conventional collaborative filtering. Our algorithm delivers high-quality recommendations, scales to enormous data sets, and produces recommendations in real-time [11] [14].

In order to innovate searching queries following product recommendations, this study additionally incorporates the FP Intersect clustering technique. Java technology and the Hadoop environment have been used to create the proposed solution, which also takes into account the Amazon dataset for experimental analysis and provides a big data infrastructure. To assess the consistency of the suggested solution, the entire solution was approximated based on computation time and executed for two separate datasets. Regardless of the increase in data size, a considerable improvement has been seen in multi node cluster solutions compared to single node cluster configuration [13] [1].

Amazon provides highly accessible, scalable cloud computing services with a variety of features, including data storage and processing engines. In this work, we used Apache Spark running on an Amazon Web Services (AWS) Elastic Map Reduce (EMR) cluster to apply the ALS algorithm for making product recommendations with good accuracy and improved scalability [3].

J-Recs is a mechanism that creates numerous reasons based on various forms of product and user data (such as purchase history and product qualities), regardless of the recommendation model used. Designing a principled graph-based technique for justification generation addresses the difficulty of jointly processing multiple forms of input. We provide a thorough review of synthetic and real-world data in addition to theoretical analysis. Our findings demonstrate that J-Recs effectively generates effective reasons, satisfying desirable justification features, and matches user preferences up to 20 percent more precisely than baselines [12] [4].

A recommendation system is an information filtering tool that predicts user ratings by taking into account their prior evaluations. These systems make up a sizable portion of the majority of cutting-edge applications. These systems heavily rely on the readily available data found on the internet. One of the methods for creating a recommendation system that, when built with the ALS algorithm, promises amazing outcomes is collaborative filtering. This is due to ALS's two-step iterative matrix factorization methodology. By supporting a master node and a small number of slave nodes, this recommendation system was developed utilising Amazon food product reviews on Apache Spark. The ALS model was built using PySpark's ml package, and the enormous amount of data was handled using RDD. assembling the model with more than 50 components [6].

V. Data Description

The original source of the data is Recommender Systems and Personalization Datasets repository. The dataset contains 7 Million data points but due to memory constraints we have conducted the experiments on the subset of the data i.e. 5000 data points.

S.No.	Attribute	Description
1	User ID	Every user identified with a unique id
2	Product ID	Every product identified with a unique id
3	Rating	Rating of the corresponding product by the corresponding user
4	Time Stamp:	Time of the rating

Figure 1. dataset description

	userId	productId	rating	timestamp
0	AKM1MP6P0OYPR	0132793040	5.0	1365811200
1	A2CX7LUOHB2NDG	0321732944	5.0	1341100800
2	A2NWSAGRHCP8N5	0439886341	1.0	1367193600
3	A2WNBOD3WNDNKT	0439886341	3.0	1374451200
4	A1GI0U4ZRJA8WN	0439886341	1.0	1334707200

Figure 2. dataset sample view

VI. Proposed Framework

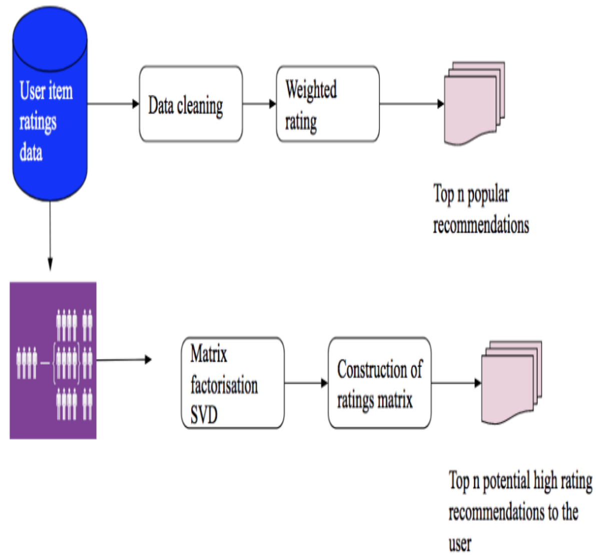

Figure 3. Workflow

- The implementation of the project is divided into following steps:
- Data is collected from open source repository Kaggle.
- Data is cleaned by removing null values, duplicate values
- After the preprocessing of data recommendation systems are built using model based methods.
- In this paperp we are implementing two types of recommendation systems:Popularity based recommendation system and User-item based recommendation system
- In the popularity based recommendation system weighted rating is used to build the model
- In the user-item collaborative filtering SVD (Singular Value Decomposition) method.
- In the user - item collaborative filtering method cross validation technique is used
- to check the accuracy of the model. 5 fold cross validation method is used to check the validation accuracy of the model.
- To evaluate validation accuracy RMSE and MAE metrics are used

A. Methods

Popularity based recommendation: Popularity based recommendation is the most common recommendation system in every system. The results of this system are Top 10 products of this month,top 10 movies list etc.,.These recommendations reflect the current trends. And also for the new user when there is no history of the user behavior in the database. Limitations of this technique is these recommendations will be the same for every user that means there is no personalisation in the recommendations.

Weighted rating: In some situations we find a number of items similar to the user. It is difficult to filter the top 3 or 4 from the list. In that case we apply a weighted ratings technique to find the most similar items.

Steps to calculate weighted ratings: r :is the rating of the product v: total number of votes for the product Min v:minimum number of votes required to top the list avg r :average rating of the product considering the whole dataset

Weighted ratings of the product = (v ÷ (v+Minv)) × r + (Minv ÷ (v+Minv)) × Minv Popular when the weightedratings of the product ¿ percentile 80 of total votes

Figure 4. Block diagram of Popularity based recommendation

Collaborative filtering (User-Item recommendation): The main idea of collaborative filtering is recommending the products to the user 1 based on the preferences of similar users. This is implemented using model based collaborative filtering that SVD.We use to reconstruct the ratings matrix.To find the relationship between the user and item We split the individual matrices into factors of small matrices that is matrix factorization. In matrix factorization we have multiple techniques, one of them Singular value decomposition.

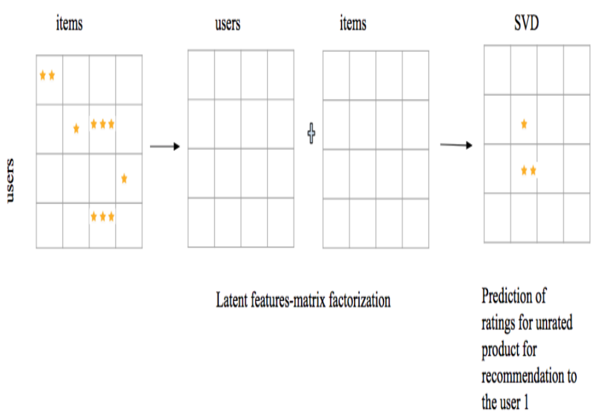

Figure 5. Block diagram of matrix factorization SVD

VII. Results Analysis

A. Exploratory Data Analysis

Exploratory data analysis is a mandatory step in machine learning pipeline. To understand the underlying structure of the data.Here we have summarize the data patterns using different visualizations barplots, density plots and count plots. The visualizations are represented using seaborn and matplotlib.

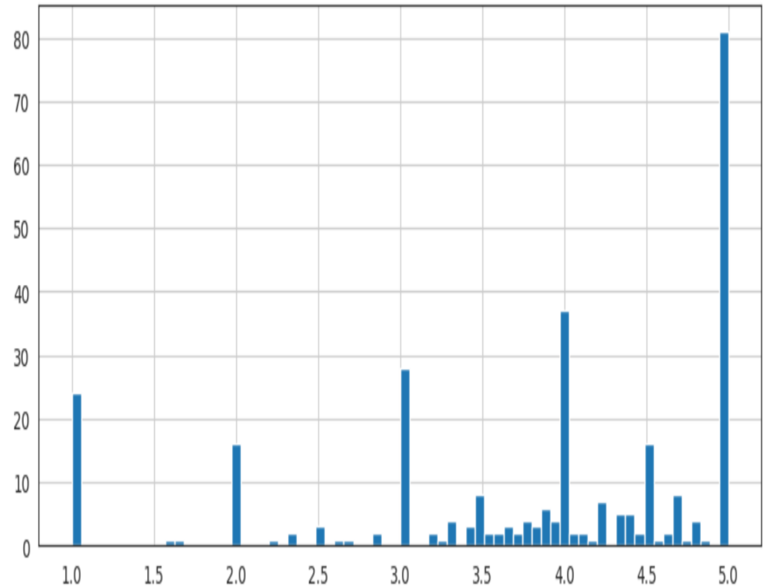

Figure 6. ratings distribution

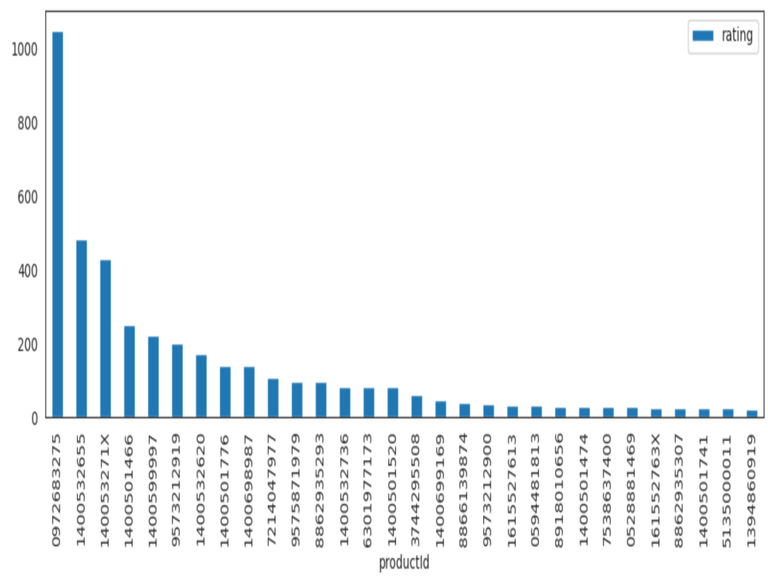

Figure 7. productwise ratings distribution

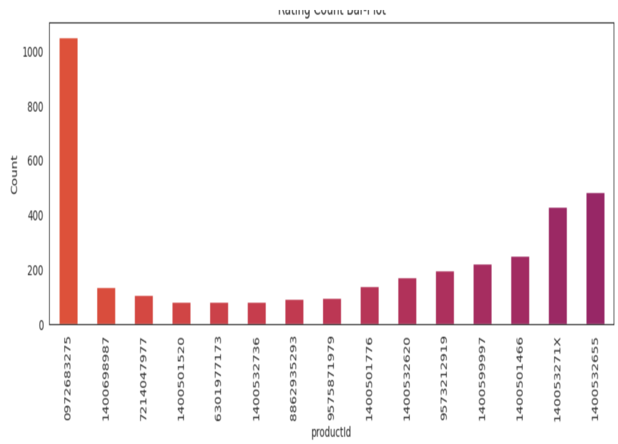

Figure 8. Productwise total ratings

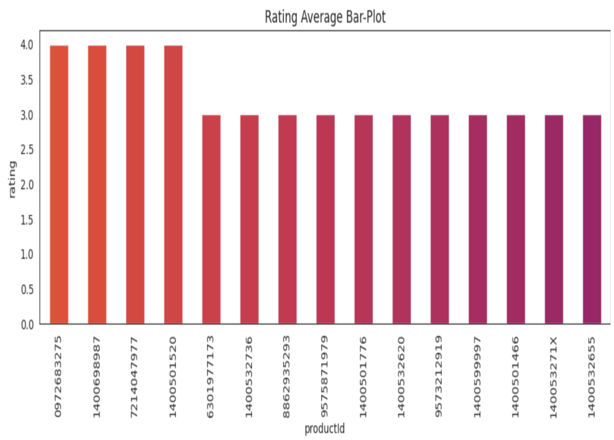

Figure 9. productwise average ratings

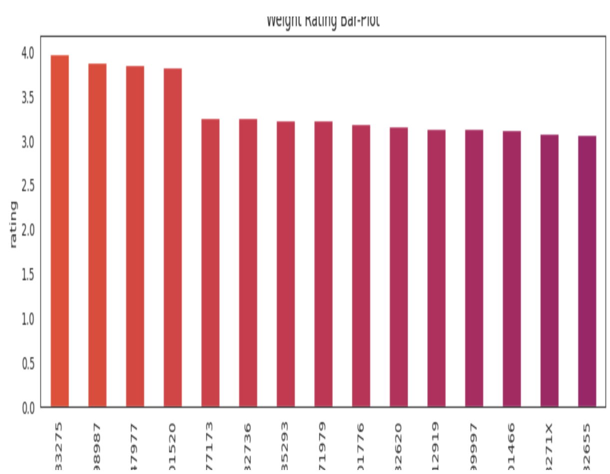

Figure 10. Productwise weighted ratings

B. Singular Value Decomposition

A matrix's Singular Value Decomposition (SVD) is a factorization into three different matrices. It communicates significant geometrical and theoretical insights regarding linear transformations and has several intriguing algebraic characteristics. It also has a few significant uses in data science.

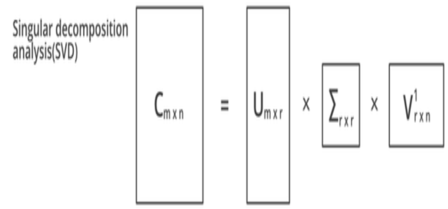

Figure 11. SVD block diagram

C. Cross validation:

Since this is a model based approach so in the training phase we have applied the cross validation that is 5 fold cross validation and recorded the RMSE and MAE scores of the model for each fold.

	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Mean	Std
RMSE (testset)	1.3226	1.3965	1.3882	1.3494	1.3471	1.3608	0.0276
MAE (testset)	1.0693	1.1167	1.1244	1.1111	1.0952	1.1033	0.0195
Fit time	0.18	0.51	0.28	0.28	0.22	0.29	0.12
Test time	0.03	0.02	0.04	0.02	0.01	0.02	0.01

Figure 12. Cross validation scores

D. Predictions

In this project we have implemented the SVD using the suprise module we have trained the model using training set and predicted the potential products from the model.

References

[1] Nur Syadhila Bt Che Lah, Ab Razak Bin Che Hussin, and Halina Mohamed Dahlan. A concept-level approach in analyzing review readership for e-commerce persuasive recommendation. In *2017 International Conference on Research and Innovation in Information Systems (ICRIIS)*, pages 1–5, 2017.

[2] Rahul Kumar Chaurasiya and Utkarsh Sahu. Improving performance of product recommendations using user reviews. In *2018 3rd International Conference and Workshops on Recent Advances and Innovations in Engineering (ICRAIE)*, pages 1–4, 2018.

[3] Lin Chen, Rui Li, Yige Liu, Ruixuan Zhang, and Diane Myung-kyung Woodbridge. Machine learning-based product recommendation using apache spark. In *2017 IEEE SmartWorld, Ubiquitous Intelligence Computing, Advanced Trusted Computed, Scalable Computing Communications, Cloud Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI)*, pages 1–6, 2017.

[4] Utpal Chandra De, Shobhan Banerjee, Manas Kumar Rath, Tanmaya Swain, and Tapaswini Samant. Content based apparel recommendation for e-commerce stores. In *2022 3rd International Conference for Emerging Technology (INCET)*, pages 1–6, 2022.

[5] Negin Entezari, Evangelos E. Papalexakis, Haixun Wang, Sharath Rao, and Shishir Kumar Prasad. Tensor-based complementary product recommendation. In *2021 IEEE International Conference on Big Data (Big Data)*, pages 409–415, 2021.

[6] Yashvanth Kumar Guntupalli, Vemula Sai Saketh, S Amudheswaran, and Devashish S Vaishnav. High-scale food recommendation built on apache spark using alternating least squares. In *2020 5th IEEE International Conference on Recent Advances and Innovations in Engineering (ICRAIE)*, pages 1–5, 2020.

[7] Wen Hu and Changshun Ge. Product recommendation algorithm combining network structure and text attributes. In *2020 IEEE Conference on Telecommunications, Optics and Computer Science (TOCS)*, pages 18–22, 2020.

[8] Sandeep Kone, Sk Meenaz Farheen, B. Lokesh, and T. Sai Pavani. A novel approach to recommend products in e-commerce. In *2021 IEEE International Conference on Intelligent Systems, Smart and Green Technologies (ICISSGT)*, pages 17–21, 2021.

[9] Nilesh Kumbhar and Krushnadeo Belerao. Microblogging reviews based cross-lingual sentimental classification for cold-start product recommendation. In *2017 International Conference on Computing, Communication, Control and Automation (ICCUBEA)*, pages 1–4, 2017.

[10] Venkatanareshbabu Kuppili, Deepak Kumar, Gayatri Pradip Kudchadker, and Ankush Arora. Variance based product recommendation using clustering and sentiment analysis. In *2015 IEEE Workshop on Computational Intelligence: Theories, Applications and Future Directions (WCI)*, pages 1–5, 2015.

[11] G. Linden, B. Smith, and J. York. Amazon.com recommendations: item-to-item collaborative filtering. *IEEE Internet Computing*, 7(1):76–80, 2003.

[12] Namyong Park, Andrey Kan, Christos Faloutsos, and Xin Luna Dong. J-recs: Principled and scalable recommendation justification. In *2020 IEEE International Conference on Data Mining (ICDM)*, pages 1208–1213, 2020.

[13] Ritu Patidar and Sachin Patel. Design implementation of product recommendation solution using sentiment analysis. In *2022 International Conference on Edge Computing and Applications (ICECAA)*, pages 233–239, 2022.

[14] Rajeev Rastogi. Machine learning @ amazon. In *2017 IEEE 24th International Conference on High Performance Computing (HiPC)*, pages 182–182, 2017.

[15] Nikita Mariam Santhosh, Jo Cheriyan, and M Sindhu. An intelligent exploratory approach for product recommendation using collaborative filtering. In *2021 2nd International Conference on Advances in Computing, Communication, Embedded and Secure Systems (ACCESS)*, pages 232–237, 2021.

[16] Shakila Shaikh, Sheetal Rathi, and Prachi Janrao. Recommendation system in e-commerce websites: A graph based approached. In *2017 IEEE 7th International Advance Computing Conference (IACC)*, pages 931–934, 2017.

[17] Rahul Shrivastava and Dilip Singh Sisodia. Product recommendations using textual similarity based learning models. In *2019 International Conference on Computer Communication and Informatics (ICCCI)*, pages 1–7, 2019.

[18] R. Suguna, M. Shyamala Devi, Akash Kushwaha, and Puja Gupta. An efficient real time product recommendation using facial sentiment analysis. In *2019 IEEE International Conference on Electrical, Computer and Communication Technologies (ICECCT)*, pages 1–6, 2019.

[19] Michalis Vafopoulos, Thomas Theodoridis, and Dimitris Kontokostas. Inter-viewing the amazon web salespersons: Trends, complementarities and competition. In *2011 15th Panhellenic Conference on Informatics*, pages 299–303, 2011.

[20] Zhefu Wu, Paul Agyemang, Minyu Chan, Hongxu Zhou, and Yun Xiang. Improved one-class collaborative filtering for online recommendation. In *2017 International Workshop on Complex Systems and Networks (IWCSN)*, pages 205–209, 2017.