International Conference on Recent Trends in Data Science and its Applications
DOI: rp-9788770040723.159

Improved Performance Of Product Recommendation System

R. Jayaram
Data Science and Business Systems.
SRM Institute of Science and Technology
Chennai, India

Dr. P. Rajasekar
Data Science and Business Systems.
SRM Institute of Science and Technology
Chennai, India

***Abstract*—E-commerce, also known as electronic commerce or internet commerce, refers to the exchange of moneyanddataforconductingbusinessonline.Althoughitcanrefer toany type of business transaction facilitated by the internet, the term commonly describes the online selling of physical goods. In contrast, e-business encompasses a wider range of activities, including online marketing, customer service, and payment processing. Product suggestion is a filtering system used in e- commerce to predict and offer the items that a user may be interested in purchasing. Even if it isn't fully accurate, if it reveals what you want to see, it has succeeded in its purpose. Nowadays, with the abundance of data available, businesses like Amazon use their massive data sets to recommend products to customers. Systems can estimate the rating of new things based on similarities between items. In order to forecast how other users would feel about a specific item, recommender systems employ user, item, and rating information. The need to have a thorough grasp of recommender systems serves as the driving force behind this effort. In this research, a model has been created that employs a variety of recommendation techniques, including association mining with the Apriori algorithm and frequent itemset.**

I. Introduction

A system for recommending products is designed to come up with suggestions for products or information that a par- ticular user might like to use or purchase. Using machine learning algorithms and a large amount of data on both particular products and unique customers, the system creates a complex network of complicated linkages between those things and individuals. A product recommendation system is a software application designed to generate personalized recommendations for products or information that a user may be interested in purchasing or interacting with. This system creates a complex network of connections between products and users using machine learning algorithms and extensive data on individual consumers and specific products. By analyzing patterns in consumer behavior and preferences, the system can offer tailored suggestions to users, improving their overall shopping experience and increasing the likelihood of sales.

II. Ease Of Use

A. *Motivation*

Beforee-commerce,thingswereonlysoldinphysicalstores.Storeinventorywaslimitedtothephysicalspaceofthestore,andslow-sellingitemswereunprofitable.Fixedinventory motivated retailers to sell only the most popularmainstreamproducts.Inthemid-1990s,theintroductionofon-linemarketplacesrevolutionizedretailing.Unlimitedinventoryis now possible in this new type of digital marketplace. Thismeant that merchants could expand their product offerings toinclude niche items rather than mainstream items. The massmarket is evolving into a mass of niches, as Chris Andersonwrites in The Long Tail. Niche products can outperform bestsellers by overcoming inventory constraints. As a result, e-commerce businesses are more interested than ever in specialisedproducts.

III. Existing System

Existing projects have proposed approaches to use the userratings to improve the performance of recommender systems.TheAmazonproductdataset,whichcomprisesofproductratings and reviews, is the subject of experiments. Comparingthetraditionalrating-basedandtheproposedrecommendersystem, we can see that the call score and root mean squared(RMSE)scoreoftherecommendersystemisdecreased.

IV. Proposed Methodology

The performance of recommender systems can be enhancedbyusingtheuserratings,accordingtoexistinginitiatives.The Experiments focus on the Amazon product dataset, whichincludes product ratings and reviews. The call score and rootmean squared (RMSE) score of the suggested recommendersystem are lower when compared to the conventional rating-basedsystem.

A. *AssociationMining*

Associationruleminingisapowerfulmethodusedtodiscoverpatterns,correlations,andassociationsthatfrequentlyoccur in a wide range of databases, including relational andtransactional databases. The process of developing associationrules involves analyzing data and identifying recurring if/thenpatterns.Thestrengthoftheassociationsisdeterminedbytwokeyparameters:support,whichindicatesthefrequencyofthe if/then relationship in the database, and confidence, whichreflects how often these associations have been proven to betrue. Association rule mining is applied to diverse data sets toidentifycommonpatterns,correlations,relationships,orcausalstructures.

B. *AprioriAlgorithm*

It is a technique used to uncover the relationships betweendifferent items. For example, in a supermarket, customers maypurchase a variety of products and there is

DOI: rp-9788770040723.159

often a pattern totheir purchases. For instance, mothers with young childrentend to buy items such as milk and diapers, while bachelorsmaybuybeerandchips,andwomenoftenbuycosmetics.By identifying the connections between items purchased indifferenttransactions,businessescanincreasetheirprofits.

Clustering

Cluster analysis or clustering is a machine learning technique that groups an unlabeled dataset into multiple clusters,where each cluster contains data points that share similarities.This technique identifies common patterns in the dataset, suchas shape, size, color, behavior, etc., and categorizes the databased on the presence or absence of these patterns. By doingso, items with potential resemblances are grouped together,whiledatapointswithlittleornosimilaritytoeachothere re placed into different clusters. Clustering is an effectivemethod for discovering structure in large datasets and can beused in various applications such as customer segmentation,imagesegmentation,andanomalydetection.

C. *TypesofClusteringMethods*

There are two main types of clustering techniques in ma-chinelearning:HardClustering,whereeachdatapointbelongsto a single group, and Soft Clustering, where data points canbelongtomultiplegroups.However,thereareseveraldifferentclustering techniques available, including: 1) Clustering andPartitioning, 2) Density-based Clustering, 3) Clustering basedonaDistributionModel,4)HierarchicalClusterAnalysis,and5)FuzzyClusterAnalysis.Thesetechniquesareusedinunsupervisedlearning,wherethealgorithmworkswithan unlabeled dataset and does not receive any supervision.Each method has its own strengths and weaknesses, and theselection of the clustering technique depends on the specificcharacteristicsofthedataandtheobjectivesoftheanalysis.

V. LITERATURE STUDY

[1] TITLE: Product Recommendation Based on Content-based Filtering Using XGBoost Classifier AUTHORS: ZeinabShahbazi, Yung-CheolByun - 2019 DESCRIPTION: [1] Akeycomponentofthemachinelearningprocessistheusageof recommendation systems to provide the user with relatedideasbasedontheirrequests.Whenadoptingthecontent-basedfiltering(CBF)techniquetopromoteanitemtoitsusers, many online shopping websites that have appropriateratinginformationhavedifficulties.Usersarenotsatisfiedwiththesearchresultswhentransitorypurchasepatternsfromsequentialpatternanalysis(SPA)areapplied.Thegoalofthis study is to recommend products using XGBoost-basedtechnology using records from the Jeju online shopping malldataset. We compare the result with the performance of otherresearch outputs based on the output of the XGBoost method.A superior rating than other individual ones is successfullydemonstrated by the proposed CBF recommendation and SPAresults.[2]TITLE:ContextualSentimentBasedRecommenderSystemtoProvideRecommendationintheElectronicProductsDomain AUTHORS: N. A. Osman, S. A. M. Noah, and M.Darwich-2019

DESCRIPTION: Sometimes people are in a rush to get thenewest products that they don’t fully consider. As a result,recommender services are growing more popular. It is crucialto separate out the most pertinent information for consumerelectronics before buying their products by looking at markettrends, speaking with a large number of influential industrystakeholders,andusingpubliclyavailabledata.Inthisstudy,asentimentanalysis-basedelectronicproductrecommendationsystem is introduced. The majority of the time, recommenda-tion algorithms predict goods based on user ratings. By usinguser comments and preferences to generate recommendations,we provide a contextual information sentiment-based modelfor recommender systems. This method’s goal is to preventterm ambiguity, a problem in recommendations known as the“domain sensitivity problem.” Utilizing the results of RMSEandMAEmeasurements,thesuggestedcontextualinformationsentiment-based model compares favourably to the traditionalcollaborativefilteringstrategywhenitcomestoelectronicproductsuggestion.

[3] TITLE: A Hybrid Collaborative Filtering Model UsingCustomer Search Keyword Data for Product Recommenda-tion AUTHORS: Ha-Ram Won, Yunju Lee; Jae-Seung Shim,HyunchulAhn - 2019 DESCRIPTION: [3] A recommendersystemisatoolthatusesmachinelearningorstatisticalmethods to suggest goods or services based on the interests ofeach individual consumer. The most often used algorithm forcreating recommender systems is collaborative filtering (CF).Althoughtherearealotofclient-provideddataavailable,ithas typically just used purchase history or customer ratings.Customerswhoshoponlinetypicallyusethesearchfeaturetosiftthroughtheenormousselectionofproductsavailabletolocatetheonestheyareinterestedin.Suchinformationon search terms might be a goldmine for modelling customerpreferences.Yet,recommendationengineshardlyeveruseitasadatasource.Inthisstudy,weintroduceadistinctivehybrid Doc2Vec CF model using search phrases and purchasehistorydatafromcustomersofonlineshoppingmalls.To verify the recommended model’s applicability, we empiricallyexamined its performance using information from a genuineKoreanonlineshoppingmall.Asaresult,wefoundthatsearch phrase information may efficiently reflect consumerpreferences and help traditional CF advance. [4] TITLE: Sen-timent Analysis for Product Recommendation Using RandomForest AUTHORS: GayatriKhanvilkar, Prof. DeepaliVora - 2018DESCRIPTION:Thetechniqueoflookingatspokenlanguage and figuring out the emotions that people innatelytransmitisknownasanalysisoffeelings.Sentimentanalysisistodeterminethepolarityofanauthor’stextualviewpoint.Itisusefultousesentimentanalysistosuggestproducts.Based on the user’s reviews, the products might be recom-mendedtoanotheruser.Topproductwebsitesutilisesentimentan

International Conference on Recent Trends in Data Science and its Applications
DOI: rp-9788770040723.159

alysistounderstandcustomersatisfactionandproblemswiththeproduct.Theprimarytypeofsentimentanalysisinvolvesdifficultiesofpositiveandnegativeclassification.Ordinalclassificationaidsintheclarificationofsentimentsin sentiment analysis. The proposed method classifies userreviews according to their polarity using ordinal classification.The system will employ SVM and Random Forest as twomachinelearningtechniquestoprovidepolarity.

[5]TITLE:RecommendationandSentimentAnalysisBasedonConsumerReviewandRatingAUTHORS:PinNi,YumingLi and Victor, Chang - 2020 DESCRIPTION: Making productsuggestions based on online reviews and rating data requiresprecise targeting of the appropriate customer segments. Salesoftheproductsmayriseasaresult.Inthisstudy,arec-ommendationandsentimentclassificationmodelisutilizedtoanalyzebeerproductdataandimprovetheperformanceof the recommendation model to cater to diverse consumerneeds. The model is developed based on online beer reviewsandratingsofbeerproducts.Thesentimentanalysisoftextis performed using ten categorization techniques, includingboth deep learning and conventional machine learning models.The beer recommendation is one of the techniques employed.Combiningthetwoevaluationscanincreasethecredibilityof the suggested beer and potentially boost beer sales. Thisapproach filters products in the market with more negativereviews in the recommendation algorithm and improves useracceptance,asdemonstratedintheexperiment.

VI. Results And Discussions

The packages used in this model are shown in Fig. 1.

```python
# for basic operations
import numpy as np
import pandas as pd

# for visualizations
import matplotlib.pyplot as plt
import squarify
import seaborn as sns

# for defining path
import os
import warnings

# for market basket analysis
from mlxtend.frequent_patterns import apriori
from mlxtend.frequent_patterns import association_rules
```

Fig. 1. List of all imported packages.

Random rows of Dataset are shown in Fig. 2.

```python
# checking the random entries in the data
data.sample(10)
```

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1133	vegetables mix	antioxydant juice	french fries	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
7714	grated cheese	frozen vegetables	ground beef	olive oil	butter	salmon	blueberries	oil	chocolate	light mayo	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
4204	cookies	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
18096	gums	cottage cheese	brownies	protein bar	low fat yogurt	green tea	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
7344	frozen vegetables	mineral water	pancakes	low fat yogurt	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
22767	frozen vegetables	mineral water	low fat yogurt	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
2543	grated cheese	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
10240	red wine	avocado	french fries	frozen smoothie	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN

Fig. 2. Dataset Sample Entries.

Frequently purchased items are shown in Fig. 3 using WordCloud Library.

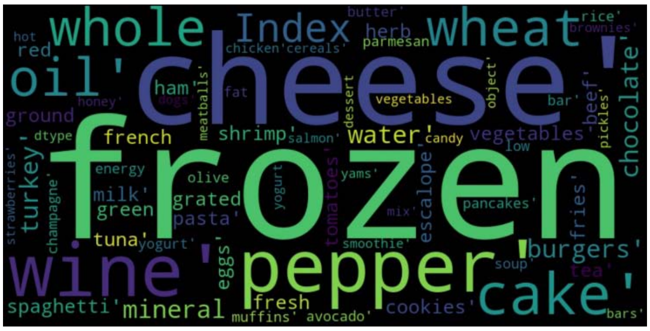

Fig. 3. Frequent Itemset.

Correlation of Items are shown in Fig. 4.

	mineral water	burgers	turkey	chocolate	frozen vegetables	spaghetti	shrimp	grated cheese	eggs	cookies	french fries	herb & pepper
mineral water	1.000000	0.029195	0.044564	0.085058	0.101125	0.115121	0.054780	0.052325	0.049572	-0.101147	-0.041823	0.056621
burgers	0.029195	1.000000	0.073111	0.027399	0.028579	0.058601	0.036692	0.012063	0.124943	-0.016765	0.065064	-0.002956
turkey	0.044564	0.073111	1.000000	0.015533	0.042563	0.065654	0.033767	0.037577	0.082848	-0.036945	0.004164	0.037118
chocolate	0.085058	0.027399	0.015533	1.000000	0.065558	0.075032	0.063963	0.031690	0.027374	-0.028905	0.049067	0.009348
frozen vegetables	0.101125	0.028579	0.042563	0.065558	1.000000	0.098305	0.127327	0.031296	0.039693	-0.062435	0.029599	0.016785
spaghetti	0.115121	0.058601	0.065654	0.075032	0.098305	1.000000	0.089324	0.088778	0.039060	-0.060039	-0.019386	0.094666
shrimp	0.054780	0.036692	0.033767	0.063963	0.127327	0.089324	1.000000	0.010307	0.004396	-0.027272	-0.021572	0.062354
grated cheese	0.052325	0.012063	0.037577	0.031690	0.031296	0.088778	0.010307	1.000000	-0.002864	-0.024666	0.012786	0.077629
eggs	0.049572	0.124943	0.082848	0.027374	0.039693	0.039060	0.004396	-0.002864	1.000000	-0.038556	0.034309	0.030511
cookies	-0.101147	-0.016765	-0.036945	-0.028905	-0.062435	-0.060039	-0.027272	-0.024666	-0.038556	1.000000	-0.009457	-0.025873
french fries	-0.041823	0.065064	0.004164	0.049067	0.029599	-0.019386	-0.021572	0.012786	0.034309	-0.009457	1.000000	-0.012959
herb & pepper	0.056621	-0.002956	0.037118	0.009348	0.016785	0.094666	0.062354	0.077629	0.030511	-0.025873	-0.012959	1.000000

Fig. 4. Correlation.

Fig. 5. shows the Frequent Patterns

	support	itemsets
0	0.237481	(mineral water)
1	0.087291	(burgers)
2	0.061765	(turkey)
3	0.160080	(chocolate)
4	0.095540	(frozen vegetables)
...	...	...
207	0.010398	(ground beef, eggs, mineral water)
208	0.012978	(milk, eggs, mineral water)
209	0.010633	(ground beef, milk, mineral water)
210	0.010164	(spaghetti, eggs, chocolate)
211	0.010477	(spaghetti, milk, chocolate)

Fig. 5. Frequent Patterns.

Fig. 6. shows the Clustering Output

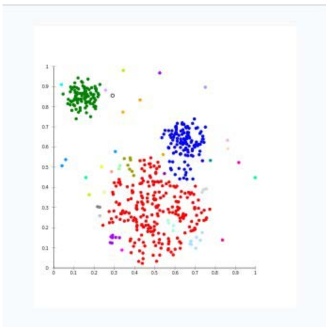

Fig. 6. Clustering.

DOI: rp-9788770040723.159

Future Enhancement

From the implementation perspective, Thus the, 1. Implementation of the Association mining algorithm 2. Implementation of the apriori algorithm. 3. Identifying the relationshipbetween the different items. 4. Finding the correlation amongthe shopping products. 5. Finding the frequently bought itemsusing apriori algorithm. 6. Prepared a model to analyse theassociations and relations. 7. Predicted the product recommendations for the customer has been implemented. The futureenhancementistoboosttheaccuracywiththeseveralboostingtechniques.

Conclusion

In this way, we have successfully implemented a recommendation system. The set of frequent items that includes association rules and the apriori algorithm was found to be the best, as the accuracy in this case was higher compared to the other methods. Using an algorithm to create a web-based recommender system was one method for analysing massive datasets. This is comparable to the algorithm that Netflix employs to suggest movies to users of his website. It was difficult to implement a web-based recommendation system with this much data. There are numerous recommender systems. They are used by people to find companions for relationships as well as books, music, news, and smartphones. There are suggestions for almost every good, service, or piece of information to assist consumers in selecting the best option from a wide range of options. A thriving research commu- nity with innovative interaction ideas, potent new algorithms,and meticulous experiments is supporting these commercial applications.

References

[1] Z.Shahbazi,andY.C.Byun,“Productrecommendationbasedon content-based filtering using XGBoost classifier,” Int. J. Adv. Sci.Technol,vol29,pp.6979-6988, 2019.

[2] Gomathy, V., Janarthanan, K., Al-Turjman, F., Sitharthan, R., Rajesh, M., Vengatesan, K., &Reshma, T. P. (2021). Investigating the spread of coronavirus disease via edge-AI and air pollution correlation. ACM Transactions on Internet Technology, 21(4), 1-10.

[3] V.Shatskykh,I.Kohut,O.Petruchenko,L.Dzyubyk,V.Bobrivetc,V.Panasyuk,S.Sachenko,M.Komar,V. Lytvyn,and V.Vysotska,“Design of a recommendation system based on CollaborativeFilteringandmachinelearningconsideringpersonalneedsoftheuser,” vol. 4,no. 2-100,pp.6-28, 2019.

[4] G. Khanvilkar,and D.Vora,“Sentiment analysis for product recommendation using random forest,” International Journal of EngineeringTechnology,vol. 7, no. 3,pp.87-89, 2018.

[5] P.Ni,Y. Li,and V.Chang,“Recommendation and sentimentanalysis based on consumer review and rating,” In Research Anthologyon Implementing Sentiment Analysis Across Multiple Disciplines,IGIGlobal, pp. 1633-1649, 2022.

[6] S. Gavhane, J.Patil, H.Kadwe, P.Thakhre, and S. Manna, “ProductRecommendationusingMachineLearningAlgorithm-ABetterAppoarch”,

[7] H. Tuinhof,C.Pirker,andM. Haltmeier,“Image-basedfashion product recommendation with deep learning,” In Internationalconference on machine learning, optimization, and data science,Springer,Cham.,pp. 472-481, September, 2018.

[8] Z.Shahbazi,D.Hazra,S.Park,andY.C.Byun,“Towardimproving the prediction accuracy of product recommendation systemusing extreme gradient boosting and encoding approaches,” Symmetry,vol. 12, no. 9,p.1566, 2020.

[9] Dhanabalan, S. S., Sitharthan, R., Madurakavi, K., Thirumurugan, A., Rajesh, M., Avaninathan, S. R., & Carrasco, M. F. (2022). Flexible compact system for wearable health monitoring applications.Computers and Electrical Engineering, 102, 108130.

[10] Y.Huang,N.N.Wang,H. Zhang,and J.Wang,“A novel productrecommendation model consolidating price, trust and online reviews,”Kybernetes,vol. 48, no. 6,pp.1355-1372, 2018.