**RECOMMENDATION SYSTEMS IN E-COMMERCE WEBSITES**

By Radhika Bailurkar (2019430003)

Faculty Incharge: Prof. Nataasha Raul

**ABSTRACT**

As the amount of data is increasing, showing a user the information they are seeking becomes important. This has led to the increasing popularity of recommendation systems. The paper presents an overview of the field of recommender systems and describes the current generation of recommendation methods that are usually classified into the following three main categories: content-based, collaborative, and hybrid recommendation approaches. Web recommender systems anticipate the needs of web users and provide them with recommendations to personalize their navigation. Such systems had been expected to have a bright future, especially in ecommerce and e-learning environments. The paper also describes various limitations of current recommendation methods.

**PROBLEM STATEMENT**

We use the recommender system almost everyday in our lives. Recommendations can be either based on your actions in the past or by pattern of items or by patterns between you and other users of a particular system. Given dataset input the system should be able to give recommendations to the user. We need to create a system which recommends user based on collaboorative filtering approach.

**OBJECTIVES**

* To give comparative analysis about various methodologies and performance metrics used.
* To study and explore about customer behaviour based on the data and do basic customer behavior exploration
* To implement model using suitable methodologies

**INTRODUCTION**

In our day to day lives we get suggestions in every aspect of our lives. Ideally, suggested items are as relevant to the us as, so that the user can engage with those items: YouTube videos, news articles, online products, and so on. A recommendation system is a computer program that helps a user discover products and content by predicting the user’s rating of each item and showing them the items that they would rate highly.[11] Behind the scenes, these systems are powered by a recommender function. A recommender function takes in information about the user and predicts the rating the user would give the product. That means you can show the user only the things they would like the best and not waste their time with products they won’t care about.[7]

**Types of Recommendations**

* Content-based recommendations: The user is recommended items similar to the ones the user preferred in the past. For example, in a movie recommendation application, in order to recommend movies to user, the content-based recommender system tries to understand the commonalities among the movies user c has rated highly in the past (specific actors, directors, genres, subject matter, etc.). Then, only the movies that have a high degree of similarity to whatever user’s preferences are would get recommended.
* Collaborative recommendations: The user is recommended items that people with similar tastes and preferences liked in the past. For example, in a movie recommendation application, in order to recommend movies to user, the collaborative recommender system tries to find the “peers” of user, i.e., other users that have similar tastes in movies (rate the same movies similarly). Then, only the movies that are most liked by the “peers” of user c would get recommended.
* Hybrid approaches: These methods combine collaborative and content-based methods.

1. Implementing collaborative and content-based methods separately and combining their predictions.
2. Incorporating some content-based characteristics into a collaborative approach.
3. Incorporating some collaborative characteristics into a content-based approach.
4. Constructing a general unifying model that incorporates both content-based and collaborative characteristics

**LITERATURE SURVEY**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Sr no. | Year | Name of the paper | Author | Content |
| 1 | 2016 | Implementation of a Recommendation System using Association Rules and Collaborative Filtering | JinHyun Joo, SangWon Bang, GeunDuk Park | Implicit Customer information collection and prediction through association rule analysis and collaborative filtering by using NFC. |
| 2 | 2018 | An Improved Collaborative Filtering Based Recommender System using Bat Algorithm | Sambhav Yadav,Sushama Nagpal | Heuristic Technique to compute the features to find better neighbourhood for active user |
| 3 | 2016 | Research on E-Commerce Platform-Based Personalized Recommendation Algorithm | Zhijun Zhang, Gongwen Xu, and Pengfei Zhang | Improved Collaborative Filtering Algorithm, NewRec |
| 4 | 2013 | A social recommender mechanism for e-commerce: Combining similarity, trust and relationship | Yung-Ming Li, Chun-Te Wu, Cheng-Yang Lai | This study proposes a social recommender system that can generate personalized product recommendations based on preference similarity, recommendation trust, and social relations |
| 5 | 2007 | Google News Personalization: Scalable Online Collaborative Filtering | Abhinandan Das, Mayur Datar, Ashutosh Garg | Generating personalized recommendations for Google News User |
| 6 | 2018 | DeepHCF: A Deep Learning Based Hybrid Collaborative Filtering Approach for Recommendation Systems | Meshal Alfarhood, Jianlin Cheng | DeepHCF uses two sources of data, ratings matrix and item reviews, to train two deep models via joint training. |
| 7 | 2018 | Parameter Based Survey of Recommendation Systems | Janani Balasubramanian, Soumya Koppaka, Chinmay Rane, and Nataasha Raul | Survey paper which explains the concept of recommendation system, its types, uses and limitations. |
| 8 | 2018 | Market basket analysis of heterogeneous data sources for recommendation system improvement | Kutuzova Tatiana, Melnik Mikhail | Propose a study on the integration of heterogeneous data sources from a grocery supermarket based on Market basket analysis |
| 9 | 2004 | Combining Usage, Content, and Structure Data to Improve Web Site Recommendation | Jia Li and Osmar R Zaiane | A novel web recommender system,  which combines usage data, content data, and structure data in a  web site to generate user navigational models |
| 10 | 2007 | Building Personalized Recommendation System in E-Commerce using Association Rule-Based Mining and Classification | Xi-Zheng Zhang | A personalized recommendation system using association rule  Mining and classification in e-commerce |

**COMPARATIVE ANALYSIS**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Sr No | Year | Method | Performance Metric | Features | Gaps Identified |
| 1 | 2016 | KMeans , Apriori, kNN | Estimation Values | Business Type Code,Business Name, Estimation Value | Sparsity will result if one customer gives ratings to business that are not frequented by other customers |
| 2 | 2018 | BAT Algorithm(BA, ABC Algos), five fold cross validation | MAE,Precision, Recall, F1 Score, RMSE,  6.9%BA>ABC | Swarm Size, Iteration and Run | In this work, the authors  present a different scheme to conquer collaboration problems by giving weights to items whilst evaluating the similarity between users. The weights are assigned and iteratively improved using the algorithms mentioned. |
| 3 | 2016 | NewRec, TimeRec, HF,CF | MAE and RMSE | Data Sparsity, Influence of Time, Influence of no. of Users, Comparison among algorithms | Influences of context and user interaction behaviour |
| 4 | 2013 | Resnick model, Average model,SR,TR,ST,STR Model | Precision,MAE,RMSE | Preference similarity, recommendation trust, and social relation | Experiment was performed on a small scale. Although the social networking service is increasingly promising, the offered features of social commerce are still limited to prestigious e-commerce sites. |
| 5 | 2007 | MinHash Clustering, PLSI, covisitation counts | Precision-Recall Curve | User click history, Clicked story on News Feed | Increased Cost, Sometimes while analyzing live traffic results may differ acc to Algorithms |
| 6 | 2018 | MLP,CNN | MAE,RMSE | Rating Matrix and Item Review |  |
| 7 | 2018 | PCA, Karhunen-loeve, Discrete Fourier transform | Precision, Recall,  Speedup/Time, F1 metrics, MAE,RMSE | Attributes of user, or of the items or both. | Relevant Variable Selection, Periodicity and Seasonality |
| 8 | 2018 | K-means Algorithm,  Apriori Algorithm | Quality Metrics – confidence distance CD  and recommendation conformity RC | Order ID, Product ID, Product Name | Results were different from expected because of heterogeneous dataset. |
| 9 | 2004 | DC Tree Algorithm, HITS | Recommendation Accuracy, Shortcut Gain | No of Clicks on the website | Incomplete or Limited Information Problem,Incorrect Information Problem, Persistence Problem |
| 10 | 2007 | CBA-CB | Associative Rules, Decision Tree Metrics | Requirement Phrases such as: Stylish, Colorful screen, light,etc | Data fragmentation. Doesnt work well with continuous data. |

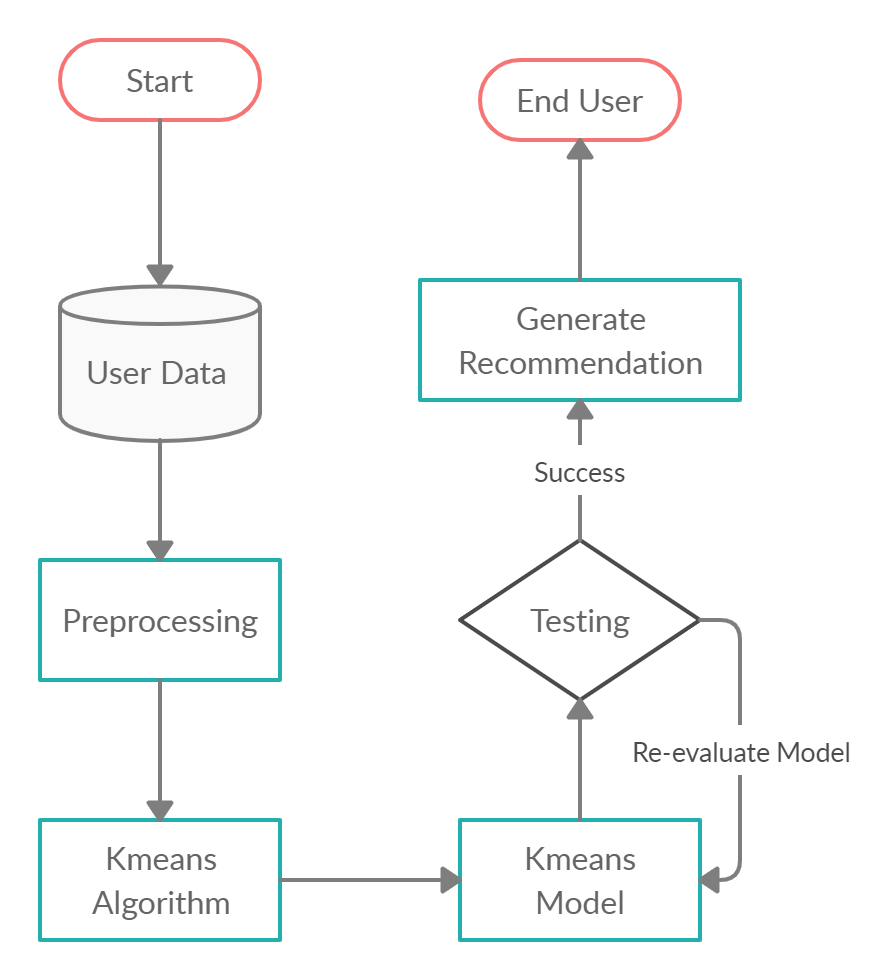
**GAPS IDENTIFIED**

1. **Cold Start Problem**: This happens due to lack of information on users or items, which results in an inability to draw inferences to recommend new items to users. Three kinds of cold start problems[2]

* New User problem
* New Item problem
* New System problem

1. **Over-specialization:** When the system can only recommend items that score highly against a user’s profile, the user is limited to being recommended items similar to those already rated.
2. **Sparsity:** In practice, many commercial recommender systems are based on large datasets. As a result, the user-item matrix used for collaborative filtering could be extremely large and sparse, which brings about the challenges in the performances of the recommendation.[1]
3. **Scalability**: As the numbers of users and items grow, traditional CF algorithms will suffer serious scalability problems. For example, with tens of millions of customers,O(M) and millions of items,O(N) , a CF algorithm with the complexity of N is already too large.
4. **Periodicity** : The periodic nature of user behaviour varying with trends can affect the quality of recommendations.[7]
5. **Changing trends in the market** : Seasonal changes in the market may lead to shift in a cluster of customers from one main topic to another.[7]

**METHODOLOGY/WORKFLOW/SYSTEM**



We take the dataset available, these datasets may have a lot of noise, we may have to filter out the noise. Then we explore the files to find relationships between the entities. We apply the preprocessed files in model. We divide the data into train data and test data. Finally we generate results through evaluation parameters and display them with the help of classification report.

**IMPLEMENTATION**

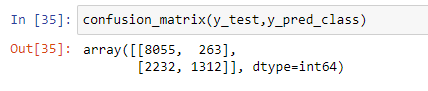
**A]Logistic Regression**

Logistic Regression is a Machine Learning algorithm which is used for the classification problems, it is a predictive analysis algorithm and based on the concept of probability. I have separated the features and the target values which are taken as train and test data for the model.

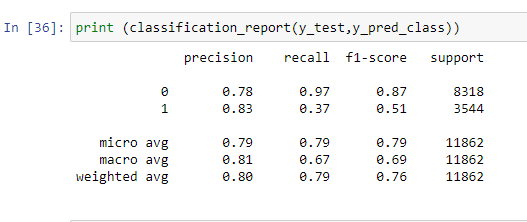
Accuracy: 78%

Metrics:

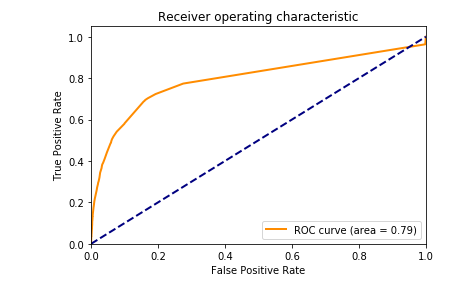
1. Confusion Matrix



1. Classification Report



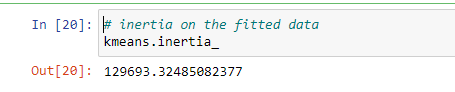
1. MAE: 0.21
2. RMSE: 0.45
3. ROC Graph



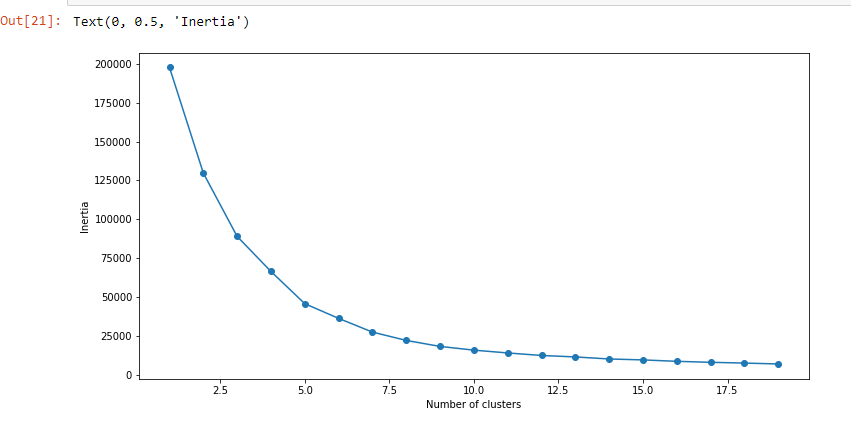
**B] KMeans Algorithm**

K Means algorithm is an iterative algorithm that tries to partition the dataset into *K*pre-defined distinct non-overlapping subgroups (clusters) where each data point belongs to only one group. First I have scaled the data files, since it is a distance based algorithm. Then created a kmeans function and fit it on the data

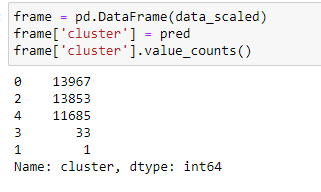
1. Inertia: Inertia is the sum of squared error for each cluster. Therefore the smaller the inertia the denser the cluster, closer together all the points are



# Elbow Method for optimal value of k in KMeans



1. Let’s look at the value count of points in each of the above-formed clusters



**RESULTS AND OBSERVATIONS**

1. In this study, it was observed that data in the files was numerical as well as categorical. Various statistics and graphs explored the files to find meaningful insights.
2. Some of the insights gained are( Visualized with the helps of graphs from seaborn, matplotlib libraries, such as barplot, piechart, pairplot,etc)

* Unique no of customers/visitors
* Comparison between view, add-to-cart and transaction behavior of customer
* No of customers who visited the website
* Checking missing values, if any,etc

1. Models Description: I have evaluated the datasets into two algorithms. Logistic Regression was chosen because I thought the data was linearly scaled, accuracy was calculated. Kmeans was used to find optimum clusters in the datasets

**CONCLUSION**

Recommender systems made significant progress over the last decade when numerous content based,collaborative and hybrid methods were proposed and several systems have been developed. We reviewed various research papers with their limitations of the current recommendation methods and discussed implementation that can provide better recommendation capabilities. It was observed that lack of importance was given to Sparity developed in larger databases, leading to cold start problems, overspecialization. Algorithms such as Logistic Regression and K means were used for analysis of the dataset gave results according to the defined metrics. It was observed that it was necessary to set the clusters to get optimum clusters. Classification algorithm gave us the accuracy metric. In this way, we can conclude that recommendation system helps the user to make valued recommendations for future.

**REFERENCES**

1. JinHyun Joo, SangWon Bang, GeunDuk Park, ”Implementation of a Recommendation System using Association Rules and Collaborative Filtering”, *Information Technology and Quantitative Management (ITQM 2016)*, 2016
2. Sambhav Yadav, Sushama Nagpal, “An Improved Collaborative Filtering Based Recommender System using Bat Algorithm”*, International Conference on Computational Intelligence and Data Science (ICCIDS 2018)*,2018
3. Zhijun Zhang, Gongwen Xu, and Pengfei Zhang, ”Research on E-Commerce Platform-Based Personalized Recommendation Algorithm”, *Applied Computational Intelligence and So Computing*,2016
4. Yung-Ming Li, Chun-Te Wu, Cheng-Yang Lai, ”A social recommender mechanism for e-commerce: Combining similarity, trust and relationship”, *Decision Support Systems*,2013
5. Abhinandan Das, Mayur Datar, Ashutosh Garg, ”Google News Personalization: Scalable Online Collaborative Filtering”, *Track: Industrial Practice and Experience*,2007
6. Meshal Alfarhood, Jianlin Cheng, ”DeepHCF: A Deep Learning Based Hybrid Collaborative Filtering Approach for Recommendation Systems”, *17th IEEE International Conference on Machine Learning and Applications*, 2018
7. Janani Balasubramanian, Soumya Koppaka, Chinmay Rane, and Nataasha Raul, ”Parameter Based Survey of Recommendation Systems”, 2018
8. Kutuzova Tatiana, Melnik Mikhail, “Market basket analysis of heterogeneous data sources for recommendation system improvement”, *7th International Young Scientist Conference on Computational Sciences*, 2018
9. Jia Li and Osmar R Zaiane, ”Combining Usage, Content, and Structure Data to Improve Web Site Recommendation”, 2004
10. Xi-Zheng Zhang, "Building Personalized Recommendation System in E-Commerce using Association Rule-Based Mining and Classification," *2007 International Conference on Machine Learning and Cybernetics, Hong Kong*, 2007, pp. 4113-4118.
11. Gediminas Adomavicius, and Alexander Tuzhili, ”Towards the Next Generation of Recommender Systems:A Survey of the State-of-the-Art and Possible Extensions”
12. Neha Verma, Devanand, Bhavna Arora, ”Experimental Analysis of Recommendation System in e-Commerce”, *International Journal of Innovative Technology and Exploring Engineering (IJITEE),ISSN: 2278-3075, Volume-8 Issue-8S3*, June 2019
13. J. Ben Schafer, Joseph A. Konstan, John Riedl,”E-Commerce Recommendation Applications”,*E-Commerce Recommender Applications*
14. Francisco, G. S., Rafael,”An integrated approach for developing e-commerce applications”,  *Expert Systems with Applications*, 2005
15. Xiaojun, Liu, “An improved clustering based collaborative filtering recommendation algorithm”,*Cluster Computing*,2017
16. Guibing Guo, Jie Zhang, and Neil Yorke-Smith,”A Novel Recommendation Model Regularized with User Trust and Item Ratings”,*IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING,*2015
17. Y. S. Cho, S. C. Moon, S. C. Noh and K. H. Ryu, "Implementation of personalized recommendation system using k-means clustering of item category based on RFM," 2012 IEEE International Conference on Management of Innovation & Technology (ICMIT), Sanur Bali, 2012, pp. 378-383, doi: 10.1109/ICMIT.2012.6225835.