ISSN: 0974-5823 Vol. 7 No. 2 February, 2022

International Journal of Mechanical Engineering

Product Recommendation to Promote Business Activities using Machine Learning Technique

R. Sheeba Mary Ananthi[1], Dr. V. Joseph Peter[2]

Research Scholar[1], Associate Professor[2]

[1,2]Research Department of Computer Science, Kamaraj College,

Affiliated to Manonmaniam Sundaranar University, Thoothukudi, Tamilnadu, India.

Abstract-To promote business activities, a product recommendation system is designed to develop and deliver recommendations for things or content that a certain user might want to buy or engage with.The ability to make recommendations,many types of user-preference and user-requirement data is provided by a recommendation system. The recommendation system faces many challenges, one of which is the requirement for huge amounts of data to make effective recommendations. The main issues raised are changing data, changing user preferences, and unpredictability. To overwhelm this, High Average-Utility Mining is used in the proposed work. The actual dataset is transformed into a high average utility itemset and then the dataset is normalized during preprocessing. The preprocessed items are getting into Content-Based Filtering model which is one of the machine learning techniques. The predicted products and the recommended products are classified in this work, which yields satisfactory results.

Keywords-content-based filtering, product prediction, product recommendation, high average utility itemset, cosine similarity

I INTRODUCTION

Several data mining applications necessitate the identification of objects, patterns, properties, and events that are similar or distinct in the data. In plenty of other words, a rigorous approach to assessing data object similarity is necessary. The computing of similarity is required in almost all data mining issues, including clustering, outlier identification, and classification. A formal formulation of the similarities or distance quantification issue[6].

The purpose of a recommender system, also known as just an information collecting system, is to create an algorithm that accounts for a wide range of demands and competency levels. It gives improved options during the project development cycle's requirement and design phases. The recommender system was used to increase product sales in the social networking and e-commerce marketplaces by offering exact results. Based on the knowledge obtained throughout the software engineering process, it suggests either a service or a product.

Transactional databases have grown highly popular in recent years for mining high utility itemsets (HUI). A frequent form of the HUI mining issue is finding the High Average Utility Itemsets (HAUIs), in which an alternative metric termed average-utility has been used to evaluate the value of itemsets by examining their lengths. Despite the fact that HAUI mining has already been thoroughly explored, existing methods usually require a huge amount of data and execute slowly due to the vast search space including the use of flexible upper limits to compute the average utility of itemsets. [9] [10].

In this proposed work, the actual data is transformed into high average utility itemset sequence database and then is transformed into normalized data. The normalized data is fed into Content-Based Filtering technique to predict and recommend products from a massive database. Fig. 1 shows the block diagram of the proposed work.

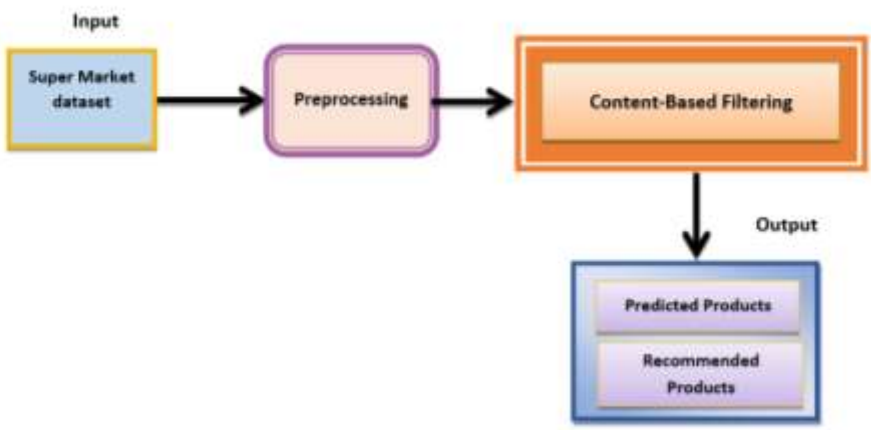

Fig. 1. Blockdiagram of the proposed work

II LITERATURE REVIEW

[1]proposes a new Content-Based Filtering model using a network of multi-attribute that reflect more attributes effectively while calculating the correlations for a recommendation of items to the users. To absorb the reciprocal ties among items and define the cluster patterns of the interactions, this research uses centrality and cluster approaches. Furthermore, our approach assures that a diverse range of products are offered to the user, as well as fixing the issue of sparse and over-specialization that plagues recommender systems.

The author of [3] examines the influence of temporal dynamics in client buying patterns on the production of individualised suggestions that are relevant. As a consequence, a bi-directional Recurrent Neural Network (RNN) containing an long short - term memory is developed for modelling temporal features in customer-product interaction data. Sequence modelling of consumers' buying behaviours is important for understanding shifting desires and contributes to the presentation of reinforced suggestions, according to the findings of the experiment. The empirical study shows that bidirectional and attention processes in consumers are beneficial.

In [4] a recommendation model generates to discover solutions based on student marks inside the college campus. The proposed work is done using Content-Based Filtering to recommend placement of employees to companies according to their requirements in a short time. This paper aims to satisfy the requirements of the employer and the employee.

In[5] a model wasproposed to providecustomized recommendationsof product information.Python performs morphological analysis on unstructured data after crawling and reviewing the material. When purchasing a product, the consumer searches for it and selects the most significant purchase criterion. Extraction as well as analysis of reviews include the purchasing criteria used by the consumer to choose a product from a pool of user reviews. The positively and negatively ratings found in the mined brand to review the data are totaled, and the value of average is used to excerpt the top ten goods, which may then be classified and suggested to consumers.

For online shopping, a smart recommendation search engine was proposedin [7].In this work,the evident comprehension images are used as input to identify the products is it in any figure.Neural networks are used to classify the data according to their class and supportively points out the similar score of image pairs. Jaccard similarity is used to calculate the training data's similarity score. The dataset includes images, class labels, etc., from Amazon. At last, the products recommend with higher similarity.

In [8] a large number of unique items and transactions using genetic algorithms (GA) are handled. GA proved that the search techniques is a robust purpose. In utility mining, there are two important challenges,they are exponential search space and database-dependent which has minimum utility threshold.GA was proposed to resolve the issue and obtain high utility itemset (HUI). Thereby a satisfactory result was incurred.

III PREPROCESSING

The steps involved in preprocessing [2] are the raw data is converted into high average utility sequence database, then it is transformed into Long Format DataFrame which converts from wide format into long format finally the data gets normalized by min-max scalar. The preprocessing block diagram is shown in Fig. 2.

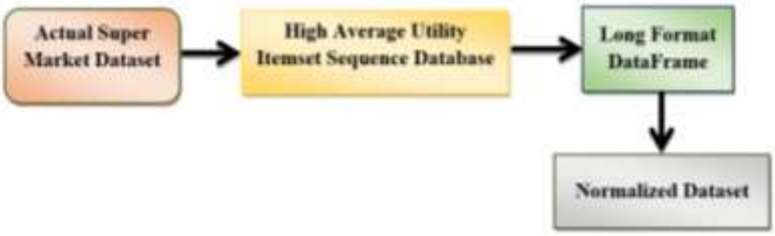

Fig. 2 Preprocessing block diagram

3.1 DataFrame

DataFrame is made up of three major components which include data, columns and rows. DataFrame can be created from the lists, dictionary, and a list of dictionary, etc. The dataset is transformed into LFDF for evaluation in this work. This ensures that the instances are reshaped.

3.2 Data Normalization

The technique of translating real-valued numeric properties into the 0 to 1 range is known as normalisation. It is also used to making model training less responsive to feature scale, allowing the machine to converge for stronger weights and therefore become more accurate.The min-max normalization process is one of the normalization methods. The min-max normalization can be done using the .min () and .max () methods which is presented in the following Eq. 1.

$$v' = \frac{v - min_A}{max_A - min_A}(new\_max_A - new\_min_A) + new\_min_A \qquad (1)$$

where v represents the respective value of the attribute, min and maxrepresent the maximum value of the attribute.

IV DATA MODELLING

4.1 Content-Based Filtering

One of the machine learning methods used to recommend and forecast items is Content-Based Filtering (CBF). CBF leverages the item's attributes to suggest further things that are related to what the user loves, based on their previous actions or obvious indication. The CBF model must suggest items that are useful to the user. To do so, one needs choose a similarity metric first (dot product).The system should be set up for each candidate item to score as per the similarity metric. Fig. 3 shows the process of the CBF model.

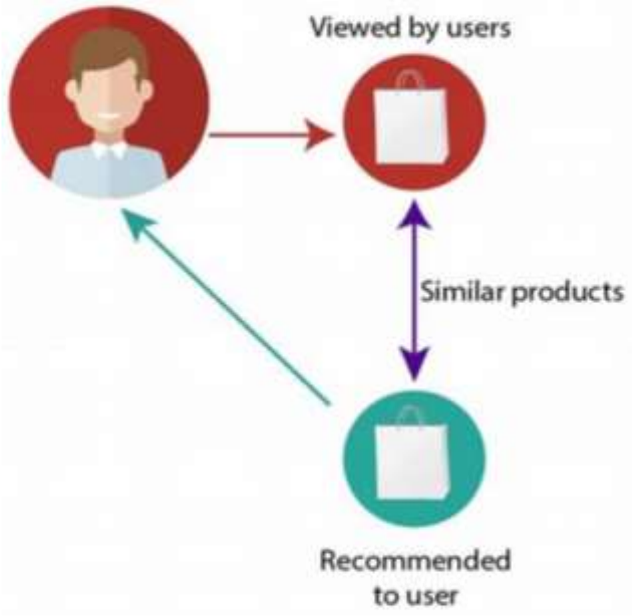

Fig. 3 Content-Based Filtering Process

4.2 Cosine Similarity

Cosine similarity (CS) is a metric used in CBF to recommend and predict items, that can be used to determine how similar data objects are regardless of size. In CS data objects in a dataset are treated as vectors. The CS algorithm is used to determine the similarity of two objects. CS is computed using the following Eq. 2.

$$Similarity\ (A,B) = \frac{A.B}{||A|| \times ||B||} = \frac{\sum_{i=1}^{n} A_i \times B_i}{\sqrt{\sum_{i=1}^{n} A_i^2} \times \sqrt{\sum_{i=1}^{n} B_i^2}} \qquad (2)$$

where A dot B is indeed the product of two vectors'A' and'B', ||A|| & ||B|| are just the lengths of both the two vectors'A' and'B', and ||A||x||B|| has been the cross value of the total vectors'A' and'B'.The value of ranges from -1 to 1, with representing -1 is perfect dissimilarity and the value of 1 representing the perfect similarity. When calculating the similarity of a small number of sets, the function is most useful.

V EXPERIMENTAL RESULTS

5.1 Dataset

This dataset named Groceries were collected from Kaggle. It contains all the purchases withinaperiod of22 months. Many of the customers are wholesalers. The numbers of instances used are 38766. The number of attributes that are used is 3. They are Member_number, Date anditem_description. The total number of products used is 167 and3898. These instances are used to work where 2728 are used for traing and 1170 are used for testing.

Table 1 Groceries dataset attributes

No.	Attribute Name	Description
1	Member_number	4-digit unique number for each transaction
2	Date	Date format of each purchase
3	Item_description	Product Name

5.2 Performance Measures

In this proposed work, the performance measures are calculated by Content-Based Filtering techniquenamely Mean Squared Error (MSE), Mean Absolute Error (MAE),Root Mean Squared Error (RMSE), Mean Average Precision (MAP) and Mean Average Recall (MAR). The formulas for the performance measures are given in Table 2.

Methods	Formulas
MSE	$\frac{\sum(u,j) \in E^{e^2}uj}{\lvert E \rvert}$
$RMSE$	$\sqrt{\frac{\sum(u,j) \in E^{e^2}uj}{\lvert E \rvert}}$
MAE	$\frac{\sum_{i=1}^{n} \lvert \hat{r}_{ui} - r_{ui} \rvert}{n}$
$MAP @ K$	$\frac{\sum AP^{(u)} @ \mathrm{K}}{\lvert U \rvert}$
$AP @ K$	$\frac{1}{N_u}\sum_{i}\sum_{r=1}^{K}\left(P_i(r)\right)x\, rel_i(r))$
$MAR @ K$	$\frac{\sum AR^{(u)} @ \mathrm{K}}{\lvert U \rvert}$
$AR @ K$	$\frac{1}{N_u}\sum_{i}\sum_{r=1}^{K} rel_i(r))$

Table 3 shows the performance measures of MSE, RMSE, MAE, MAP and MAR of the proposed work CBF.

Table 3 Performance measures of the proposed work

Methods	CBF
MSE	0.1085
RMSE	0.3293
MAE	0.2376
MAP	1.0
MAR	0.0602

VI CONCLUSION

In this proposed work the machine learning algorithm CBF is used to predict and recommend the products from the Groceries dataset.CBF technique is a powerful technique to predict and recommend products. The proposed work gives satisfactory results which are based on mean absolute error. In thefuture, other various machine learning techniques with various datasetscan be used for evaluation.

VII REFERENCES

[1] Jieun Son and SeoungBum Kim, Content-based filtering for recommendation systems using multiattribute networks, Expert Systems with Applications, 89 (2017), pp. 404–412.

[2] Wei Jianping, Research on Data Preprocessing in Supermarket Customers Data Mining, 978-1-4244-7941-2/10, 2010 IEEE.

[3] Latchoumi, T. P., & Parthiban, L. (2022). Quasi oppositional dragonfly algorithm for load balancing in cloud computing environment. Wireless Personal Communications, 122(3), 2639-2656.

[4] Naresh Nelaturi, Golagani Lavanya Devi, A Product Recommendation Model Based on Recurrent Neural Network, Journal Europeen des Systemes Automatises, Vol. 52, No. 5, pp. 501-507, 2019.

[5] Karnan, B., Kuppusamy, A., Latchoumi, T. P., Banerjee, A., Sinha, A., Biswas, A., & Subramanian, A. K. (2022). Multi-response Optimization of Turning Parameters for Cryogenically Treated and Tempered WC–Co Inserts. Journal of The Institution of Engineers (India): Series D, 1-12.

[6] N. Karbhari, A. Deshmukh and V. D. Shinde, Recommendation system using content filtering: A case study for college campus placement, 2017 International Conference on Energy, Communication, Data Analytics and Soft Computing (ICECDS), 2017, pp. 963-965, doi: 10.1109/ICECDS.2017.8389579.

[7] Jinyoung Kim, Doyeun Hwang, Hoekyung Jung, Product recommendation system based user purchase criteria and product reviews, International Journal of Electrical and Computer Engineering (IJECE), Vol. 9, No. 6, December 2019, pp. 5454-5462 ISSN: 2088-8708.

[8] Charu C. Agarwal, Data Mining: The Text Book, Springer, 2015.

[9] Ranjeeth, S., Latchoumi, T. P., & Paul, P. V. (2021). Optimal stochastic gradient descent with multilayer perceptron based student's academic performance prediction model. Recent Advances in Computer Science and Communications (Formerly: Recent Patents on Computer Science), 14(6), 1728-1741.

[10] Luyang Chen, Fan Yang and Heqing Yang, (2017), Image-based Product Recommendation System with Convolutional Neural Networks.

[11] S. Kannimuthu and Dr. K.Premalatha, Discovery of High Utility Itemsets UsingGenetic Algorithm, International Journal of Engineering and Technology (IJET),Vol 5 No 6 Dec 2013-Jan 2014, pp. 4866-4880.

[12] Lin, Chun-Wei and Ren, Shifeng & Fournier Viger, Philippe & Hong, Tzung-Pei & Su, Ja-Hwung and Vo, Bay., (2017), A fast algorithm for mining high average-utility itemsets. Applied Intelligence.

[13] Lin, J. C. W., Fournier-Viger, T. Li, P., Hong, T. P., Zhan, J., and Voznak, M., (2016), An Efficient Algorithm to Mine High Average-Utility Itemsets, Advanced Engineering Informatics, vol. 30, no. 2, pp. 233-243.

[14] Garikapati, P., Balamurugan, K., Latchoumi, T. P., & Malkapuram, R. (2021). A Cluster-Profile Comparative Study on Machining AlSi7/63% of SiC Hybrid Composite Using Agglomerative Hierarchical Clustering and K-Means. Silicon, 13(4), 961-972.