JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

SIMILAR PRODUCT RECOMMENDATION SYSTEM USING KNN

[1] **Mrs. Tulasi Miriyala**
Assistant Professor , Department of Computer Science & Applications,
AQJ Centre for PG Studies, Gudilova, Anandapuram, Visakhapatnam-531173.

[2] **Mr. Gangadhar Doma**
Associate Professor , Department of Computer Science & Applications,
AQJ Centre for PG Studies, Gudilova, Anandapuram, Visakhapatnam-531173.

ABSTRACT:

People in the modern years choose similar product recommendations since they are too busy browsing through the whole CatLog as they perceive it as a challenging endeavour. However, the existing approach does not provide reliable and precise results. In this case, we are employing the KNN artificial intelligence algorithm along with multiple techniques for extracting features such as VGG16, VGG19, and Resnet50 and reduction of dimensionality approaches as PCA, LDA, and K-Means. Customers will benefit from this since it offers very precise and precise outcomes according to user requirements, as well as product advice for many different buying patterns.

Key Words: KNN, PCA, LDA, VGG16, VGG19, Resnet50

I. INTRODUCTION

In the world of e-commerce, systems that recommend products based on a customer's past purchases, searches, and other behavioral data are important applications of artificial intelligence. The technology uses the latest machine learning techniques to analyze both data sets in order to identify similarities between user data and item data and provide recommendations to consumers. We examine the efficiency of different machine learning methods for providing recommendations to users, including collaborative filtering, content-based filtering, and hybrid filtering. Content-based filtering examines product content to make recommendations, while collaborative filtering incorporates data from other users with similar tastes. Combining the two techniques, hybrid filtering improves the accuracy of suggestions. Product suggestion algorithms are becoming an important part of many online shopping sites, helping customers find relevant products and drive purchases. Based on users' browsing and purchasing habits, these platforms use artificial intelligence algorithms to analyze information and make specific suggestions. Collaboration is a popular way of proposing products filtering, which relies on the principle that users who have similar preferences are likely to enjoy similar products. In this method, the computer examines user data to find things that are commonly bought or viewed together and suggests these items to other users who have showed interest in one of them.

Content-based filtering using product properties such as title, definition, group, cost, and other aspects is an alternative strategy. To find products with equivalent attributes, the system looks at product characteristics and generates a similarity metric. Algorithms then recommend products based on the user's desire for specific properties. For more accurate and diverse recommendations, a hybrid recommender system integrates collaboration and content-based filtering. These devices take full advantage of both technologies to avoid their shortcomings and provide consumers with better and more personalized offers. To improve customer satisfaction and increase sales, a website that sells products needs product suggestions. These engines use artificial intelligence techniques such as collaborative filtering, content-based filtering, and hybrid recommendation systems to make personalized suggestions based on user information and product characteristics.

In other words, a proposal scheme is the process of presenting a product acceptable to the customer. For example, Netflix's recommendation system analyzes the user's previous preferences, such as "Which movie to watch?", "Which product to buy," and "Which book to read?" to display matching products. 1]. A user's favorite product attributes or tags are referred to herein as content. This type of technology assigns keywords to products, searches databases to determine consumer needs, and attempts to offer a range of products to meet those needs. For illustration purposes, consider a movie recommendation system. Each movie is given a genre under this system, called tags or features in the example above. Suppose user A logs into the system for the first time, but no information about the user is available. Early on, the system will try to recommend popular movies to viewers or solicit responses to surveys designed to gather information about movies. After a while, a user may have rated a particular movie. For example, [3] an animated movie might get bad reviews, but an action movie might get good reviews. Therefore, our approach recommends action movies to consumers. But in this situation it is impossible to justify the person's contempt for animation. B. Because of the acting and the plot, but I really like anime movies and want more information.

Recommendation systems have a variety of uses, but the most well-known are opening playlists for audio and video products and services, product suggestions for online stores, material suggestions for social networking platforms, and online content recommendations. increase. These platforms can be used to access a single input such as music, or multiple inputs from one platform to another such as news, books, search, etc. In addition, the recommendation system for topics such as online dating and gastronomy is also popular. Research articles, experts, and other technologies are also produced to support financial services. Recommender systems are widely used to suggest products and services on the Internet. These techniques are typical of e-commerce websites [5].

These systems perform two important functions. Provide product suggestions and other information to customers to combat knowledge overload. Selling additional products contributes to a company's revenue generation. In this project, we propose a comparable product recommendation system for e-commerce applications using the ANN algorithm. Users receive personalized product recommendations from the suggested system based on their historical data and interests. To find patterns and similarities, the system is designed to analyze user behavioral data such as searches, browsing history and recent purchases. To find product similarities, the algorithm also looks at product information such as features, ratings, and reviews.

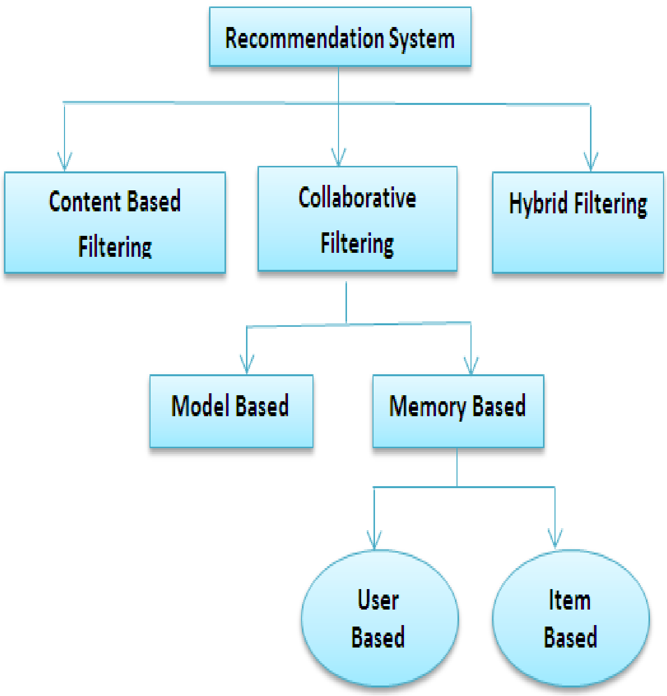

Figure 1: Recommendation System Classification

II. RELATED WORK

These systems perform two important functions. Provide product suggestions and other information to customers to combat knowledge overload. Selling additional products contributes to a company's revenue generation.

In this project, we propose a comparable product recommendation system for e-commerce applications using the ANN algorithm. Users receive personalized product recommendations from the suggested system based on their historical data and interests. To find patterns and similarities, the system is designed to analyze user behavioral data such as searches, browsing history and recent purchases. To find product similarities, the algorithm also looks at product information such as features, ratings, and reviews.

The author of [9] suggested using GANS to obtain features and determine how an image differs from all other images. The best N-like images were found and suggestions were made. UTZap 50K was used as a subject in this study. [10] describes how the authors used the CNN method to extract features from two different datasets of fashion photography.

Farhan Ullah and others.Mohammed FadhelAljunid, ManjaiahDoddaghattaHuchaiah Informatics, Mangalore University, Mangalore, India Please contact ngm505@yahoo.com. [6] As a result of the large amount of explicit input such as browsing and clicking, many researchers have developed recommendation systems (RS) that use implicit feedback. Typing it aloud is too difficult, but very useful for developing recommender systems. Standard population filtering technique. B. Matrix factorization suffers from cold start and information economy issues due to poor learning ability.

Farhan Ullah and others.Mohammed FadhelAljunid, ManjaiahDoddaghattaHuchaiah Informatics, Mangalore University, Mangalore, India Please contact ngm505@yahoo.com. [6] As a result of the large amount of explicit input such as browsing and clicking, many researchers have developed recommendation systems (RS) that use implicit feedback. Typing it aloud is too difficult, but very useful for developing recommender systems. Standard population filtering technique. B. Matrix factorization suffers from cold start and information economy issues due to poor learning ability.

Limitations:

1. All models are working with less accuracy.
2. The existing system is showing some irrelevant products.
3. We can't search the outside product from the website.

To avoid the reported limitations, we proposed a new model i.e VGG 19, LDA with KNN and also K-Means, Resnet50 and KNN.

III. EXISTING SYSTEM

In the existing system the following algorithms are implemented such as PCA, VGG 16 and KNN were used when developing the current system. As a result, you will get inaccurate results. Hence there is no accurate results generated in the existing system for similarity detection.

IV.PROPOSED SYSTEM & ITS METHODOLOGY

The suggested system was improved to address shortcomings of previous systems. The dimensionality reduction methods LDA, K-Means, and KNN algorithms are combined with VGG 19, Resnet 50, and these techniques. The proposed approach creates a fast and reliable way to find similar products.

Modules:

1. Collection of Data
2. Data processing
3. Training to the Model
4. Model Optimization
5. Evaluation of Model

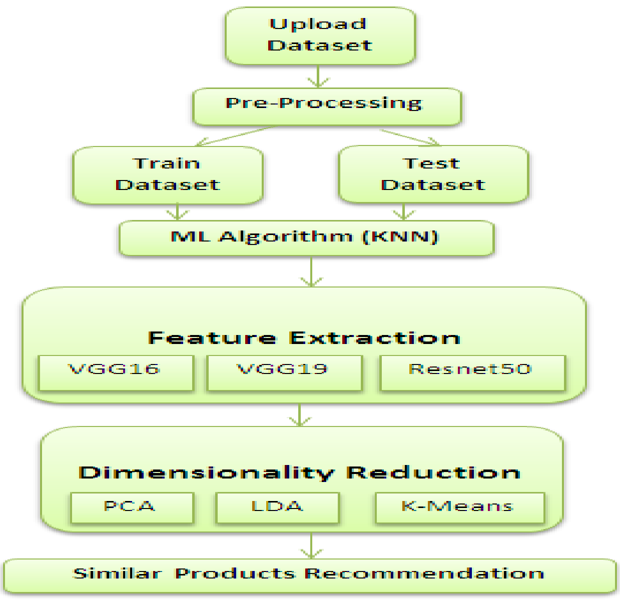

Figure 2: Proposed Model

1. **Collection of Data**

Data is collected and loaded from kaggle from fashion product images dataset and all the coding part is done in Jupiter notebook by importing necessary python libraries.

2. **Data preparation:**
   a. Null value handling
   b. Categorical handling
   c. Scaling

The specified data set for this data preparation module is missing some values, which may lead to inaccurate predictions. Remove missing values using Python. Also, to get correct results, the replicated data should be removed from the data set. Normalization is the process of scaling data to a specific range.

3. **Model Training:**

Models are built using previously processed data and multivariate analysis is performed using learning algorithms. During the training phase, the machine learning system needs to be fed data to find and learn good values for all relevant properties. Supervised learning model KNN is used.

4. **Model Optimization:**

In this model is dimensionality reduction and feature extraction is done with the help of VGG16, VGG19, and Resnet50 along with PCA, LDA, and K-means.

5. **Model Evaluation:**

Each model is evaluated using accuracy and similar products are recommended.

This work's main objective is to present a novel aggregation approach in demand to make forecasts that are more accurate. We offered the ensemble three KNN, such as KNN Model-1, KNN Model-2, and KNN Model-3, and the specifics of every one of them are described below.

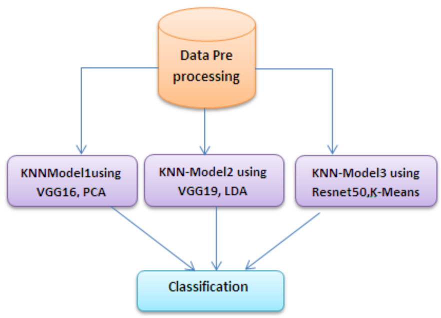

Figure 3: Proposed ensumble Model

KNN model-1:

According to this model the requirements of the recommendation task are KNN, PCA, and VGG16 can be applied in various ways for developing recommendation systems.

Collaborative filtering and recommendation systems can be created using KNN. In this method, user or item similarity is assessed based on their conversations or evaluations. KNN can be used to find the target user or item's k-nearest neighbors by applying it to the similarity matrix. Then, based on the choices of the closest neighbor's, recommendations are made. KNN's performance and scalability can be increased by combining it with other methods like dimensionality reduction and clustering.

VGG16 is available for developing image-based suggestion systems. In this process, photo features relevant to the rating task are extracted with VGG16. High-dimensional spatial features allow you to use features to represent imagery. Suggestions are then made based on similarity of photos in component space.

PCA can be used in content-driven recommender systems for feature extraction and dimensionality reduction. This method extracts features relevant to the recommendation task from the content of items such as: B. Text, Photo, or Audio. Using PCA, we can reduce the dimensionality of the feature space and determine the most important features. Suggestions are then made based on how well the characteristics of the item in question match the characteristics of the items in the database.

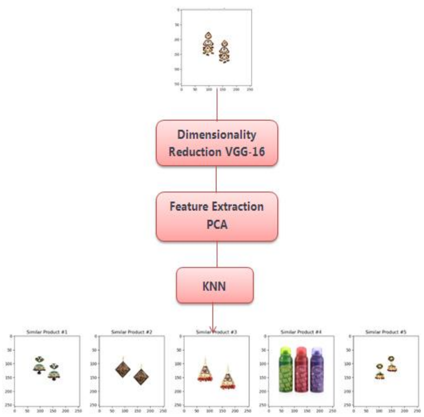

Figure 4: KNN Model-1 Flowchart

KNN Model-2:

This model combines KNN, LDA, and VGG19 to create a recommendation system that maximizes the benefits of each approach.A very small (3x3) convolution filter is a crucial feature of the VGG-19 structure. As a result, networks can go deeper while maintaining a manageable set of parameters. A max pooling layer is also used to minimize the geographic dimensions that make up the mapping features. An ImageNet information set containing over 1 million photos from 1,000 different categories was used to train the VGG-19 model. This model has shown excellent results in many computer vision-related tasks, such as object recognition and image classification.

Natural language processing programs can use LDA to extract features. By extracting features from unstructured text input, machine learning algorithms can use the data numerically. For LDA, the output of the model is a set of themes, each characterized by a probability distribution over the set of terms and phrases. The entered text information can be characterized by these topics.

Here is a possible flowchart for building a recommendation system using KNN, LDA, and VGG19:

Data Preprocessing: Clean and preprocess the data, such as user-item interaction data, item content data, or image data.

Feature Extraction: Extract the features that are relevant to the recommendation task. For example, in a content-based recommendation system, features can be extracted from item content using LDA. In an image-based recommendation system, features can be extracted from images using VGG19.

Dimensionality Reduction: Reduce the dimensionality of the feature space using LDA to identify the most important topics or themes in the item content.

Similarity Calculation: Calculate the similarity between the target item and the items in the database using KNN based on their features or interactions.

Neighborhood Selection: Select the k-nearest neighbors of the target item based on their similarity scores using KNN.

Recommendation Generation: Generate recommendations based on the preferences of the nearest neighbors or the similarity between the features of the target item and the items in the database.

Evaluation: Evaluate the performance of the recommendation system using appropriate metrics, such as accuracy, recall, and precision.

Iteration: Iterate and refine the recommendation system based on the evaluation results and feedback from users.

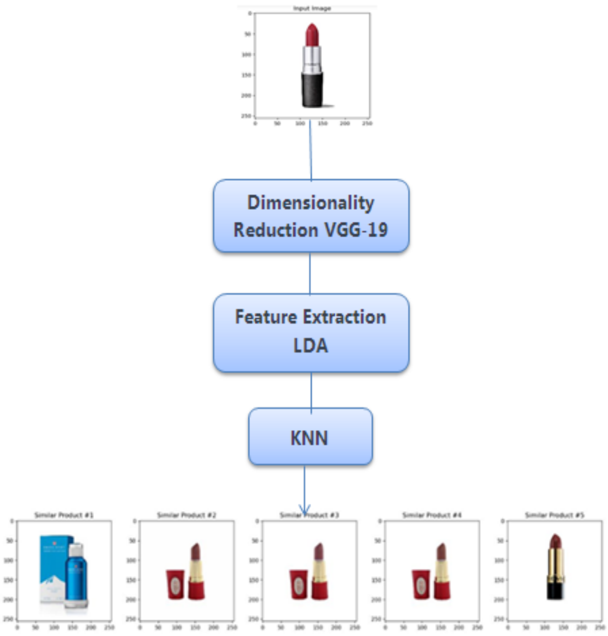

Figure 5: KNN Model-2 Flowchart

KNN Model-3:

Here we are developing the model using Resnet 50,K-Means and KNN.

The ResNet-50 architecture contains 50 layers including residual blocks, batch normalization layers, and convolution layers. The rest of the blocks are an important part of the architecture, allowing the network to be very deep without encountering the vanishing gradient problem.

A remainder blocks passes an input to two or more convolutional layers and then adds the output of the second convolutional layer to the original input. This phenomenon, called residual connectivity, can cause the network to learn the residual function instead of the original function. This provides a direct path for gradients to flow backwards through the network and helps solve the vanishing gradient problem.

A set of data points is divided into k groups. Where k is a user-defined parameter, using a clustering technique known as k-means. The purpose of this algorithm is to reduce the sum of diagonal distances between information points and their assigned cluster centroids. Here's a high-level overview of how the K-means algorithm works:

1. **Initialization:** First, k initial cluster centroids are randomly chosen from the data points.
2. **Assignment:** Each data point is assigned to the cluster with the nearest centroid, based on the Euclidean distance between the data point and the centroid.
3. **Update:** The centroids of each cluster are updated to be the mean of the data points assigned to that cluster.
4. **Repeat:** Steps 2 and 3 are repeated until convergence is reached. Convergence is typically defined as a small change in the position of the centroids between iterations or a maximum number of iterations.

The K-means method produces a set of k clusters, each represented by its centroid. These clusters have a wide range of applications, including customer segmentation, picture segmentation, and identifying anomalies.

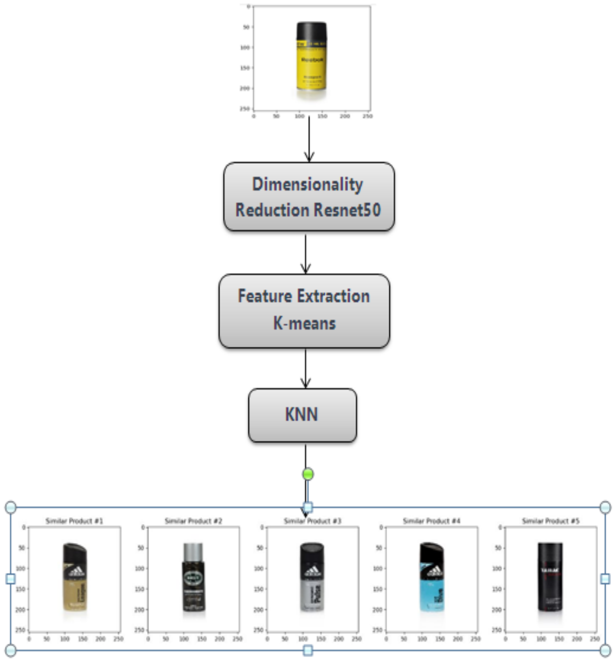

Figure 6: KNN Model-3 Flowchart

V. MACHINE LEARNING ALGORITHMS

K-Nearest Neighbor Classifier:

- K-Nearest Neighbor, one of the most basic machine learning algorithms, is based on supervised learning approaches.
- The K-NN method places new cases in groups that closest resemble the available categories, based on the assumption that the new instances and data are comparable to the already available examples.
- The K-NN algorithm identifies new points of data based on similarities to previously collected and stored data. As new data is generated, the the K-NN algorithm look at can quickly classify it according to essential groups.
  K-NN is a nonparametric algorithm. K-NN is a nonlinear algorithm. H. There are no assumptions about what underlies the data. They are also called delayed learning algorithms because they do not immediately integrate what they know into new data. But save the record and take action when new data is categorized.

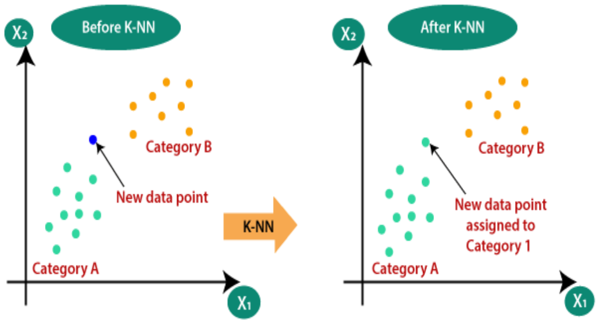

Figure 7: KNN Algorithm

VI. RESULTS

Results KNN Model-3:

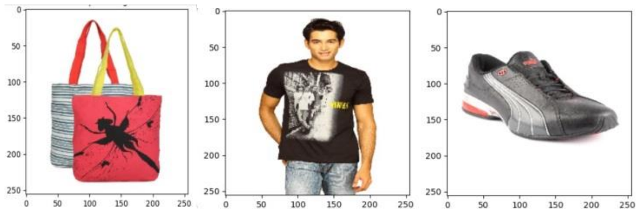

Figure 8: Test Images for KNN Model-3Recommendation

Figure 9: Recommendation Items for KNN Model-3

Results KNN Model-1:

Figure 10: Test Images for KNN Model-1 Recommendation

9Figure 11: Recommendation Items for KNN Model-1

Results KNN Model-2:

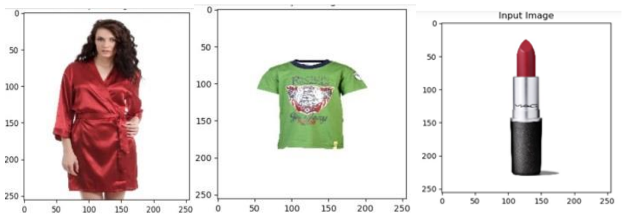

Figure 12: Test Images for KNN Model-2 Recommendation

Figure 13: Recommendation Items for KNN Model-2

VII. CONCLUSION

This work involves data preprocessing, feature collection, and application of KNN models to the data using various dimensionality reduction and feature extraction strategies. Accuracy improves significantly when more factors are used. This is very important for forecasting systems. A better understanding of how subcategories work can increase the effectiveness of your system. Model evaluation is an important step in creating an effective artificial intelligence model. Therefore, it is important to develop a model and use it to create metric suggestions. Despite the fact that approaches such as deep studies and neural network modeling have low processing costs. It serves an important purpose in areas where resources are scarce and spending is high. As a result, the demand for products from crossword puzzles, supermarkets and small businesses is increasing. This approach can be further improved by using expensive methods such as DNN and Tensor Flow.

VIII. REFERENCES

1 Mohammed FadhelAljunid and Manjaiah D.H "a survey on recommendation mechanisms for social networks using big data analytics" Special Issue (SACAIM 2017), International Journal of Latest Trends in Engineering and Technology, pp.48-58.

2. S. Taneja, C. Gupta, D. Gureja, and K. Goyal, "K Nearest-Neighbor Techniques for Data Classification-AReview," Proceedings of the International Conference on Computing, Informatics, and Networks (ICCIN2K14), January 2014, pp. 69-73.

3. "An Improved KNN Algorithm Based on Adaptive Cluster Distance Bounding for High Dimensions Indexing," H. Hong, G. Juan, and W. Ben. Conference Publishing Services, 3rd Global Congress on Intelligent Systems (GCIS), Nov. 2012, pp. 213-217, doi:10.1109/GCIS.2012.86.

4. "A New Distance-weighted k-nearest Neighbour Classifier," J. Gou, L. Du, Y. Zhang, and T. Xiong. Journal of Information and Computational Science, pp. 1429-1436, June 2012.

5. "Learning Semantic Similarity for Very Short Texts," C. De Boom, S. Van Canneyt, S. Bohez, T. Demeester, and B. Dhoedt, 2015.

6. A. Pea-Ayala, "Educational data mining: A survey and data mining-based analysis of recent works," Expert Syst. Appl., vol. 41, no. 4, pp. 1432-1462, March 2014.

7. G. Adomavicius, "Towards the generation that follows of Recommender Systems: An Overview of the State-of-the-Art and Possible Extensions", IEEE Transactions on Learning and Data Engineering, vol 17, no. 6, 2005.

8. W. Yang, Z. Wang, and M. You, "An upgraded collaborative filtering technique for recommendation generation," Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics, 2004.

9. B. Ay, G. Aydn, Z. Koyun, and M. Demir, "A Visual Similarities Recommendation System Using The term generative Adversarial Networks," 2019 International Conference on Machine Learning and Deep Learning in Developing Applications (Deep-ML), pp. 44-48, doi:10.1109/Deep-ML.2019.00017. N. YarahmadiGharaei, C. Dadkhah, and L. Daryoush, "Content-based Garment.

10. Recommender System Using Deep Recurrent Network," 2021 26th International Computer Conference, Computer Society of Iran (CSICC), 2021, pp.1- 6,doi:10.1109/CSICC52343.2021.9420544.