AI-SHOPPING RECOMMENDATION SYSTEM

A Course Project report submitted

in partial fulfillment of requirement for the award of degree

BACHELOR OF TECHNOLOGY

in

ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING

by

B. MANIKANTA SAAKETH (2103A52076)

D. SRINIDHI (2103A52079)

V. SHIVANI (2103A52188)

Under the guidance of

Mr. D. RAMESH

Assistant Professor, Department of CSE.

Department of Computer Science and Artificial Intelligence

SR UNIVERSITY

Department of Computer Science and Artificial Intelligence

CERTIFICATE

This is to certify that project entitled **"AI-SHOPPING RECOMMENDATION SYSTEM"** is the bonafied work carried out by **B. MANIKANTA SAAKETH, D.SRINIDHI, V.SHIVANI** as a Course Project for the partial fulfillment to award the degree **BACHELOR OF TECHNOLOGY** in **ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING** during the academic year 2022-2023 under our guidance and Supervision.

Mr. D. RAMESH	**Dr. M. Sheshikala**
Asst. Professor,	Assoc. Prof .& HOD (CSE)
S R University,	S R University,
Ananthasagar ,Warangal	Ananthasagar ,Warangal

ACKNOWLEDGEMENT

We express our thanks to Course co-coordinator **Mr. D. Ramesh, Asst. Prof.** for guiding us from the beginning through the end of the Course Project. We express our gratitude to Head of the department CS&AI, **Dr. M. Sheshikala, Associate Professor** for encouragement, support and insightful suggestions. We truly value their consistent feedback on our progress, which was always constructive and encouraging and ultimately drove us to the right direction.

We wish to take this opportunity to express our sincere gratitude and deep sense of respect to our beloved Dean, School of Computer Science and Artificial Intelligence, **Dr C. V. Guru Rao**, for his continuous support and guidance to complete this project in the institute.

Finally, we express our thanks to all the teaching and non-teaching staff of the department for their suggestions and timely support.

ABSTRACT

A fashion recommendation system is an automated software solution that can suggest fashion items based on a user's preferences and style. This system relies on image analysis and machine learning algorithms to suggest personalized fashion choices for users. The fashion recommendation system uses computer vision and deep learning techniques to identify patterns and styles from images, and match them with users' preferences. This abstract highlights the key features of a fashion recommendation system and its potential benefits, including increased customer satisfaction and loyalty, improved sales and revenue, and reduced inventory waste. The image recommendation system can be used in various applications, such as social media platforms, e-commerce websites, and online search engines. It provides personalized and engaging content to users, which can improve user engagement and retention. Image-based fashion recommendation systems have at-traced a huge amount of attention from fast fashion retailers as they provide a personalized shop-ping experience to consumers. With the technological advancements, this branch of artificial intelligence exhibits a tremendous amount of potential in image processing, parsing and segmentation.

Table of Contents

INTRODUCTION

1.1 OVERVIEW

An AI project that recommends fashion products is like having a personal shopping assistant that suggests clothes and accessories that you might like. This project uses computer algorithms to group similar fashion items together based on their colours, styles, materials, and sizes. It then recommends items to you based on what you've bought before and what you like. The goal is to make shopping easier by suggesting things that match your taste. The image recommendation system can be used in various applications, such as social media platforms to suggest relevant content to users, e-commerce websites to recommend products based on user preferences, and search engines to provide personalized image search results. The system can provide numerous benefits, including improved user engagement and retention, increased conversion rates, and higher customer satisfaction. In the recommendation generation stage, the system uses the trained machine learning models to make recommendations based on users' interests and behaviours. The system can suggest related or similar images based on the content of the image, as well as recommend other content that users with similar interests have engaged with.

1.2. PROBLEM STATEMENT

The fashion industry is facing a significant challenge in providing personalized and relevant fashion choices to customers. With the growing demand for fast fashion and the ever-changing trends, customers find it difficult to navigate through the vast options and make informed decisions. With the increasing availability of technology, the problem of efficiently and accurately retrieving relevant images from large datasets has become more significant. Traditional methods of searching for products using keywords and tags are often inadequate, and users are increasingly seeking more sophisticated and personalized ways of finding images.

1.3. EXISTING SYSTEM

There are several existing fashion recommendation systems that use AI and machine learning to provide personalized fashion choices to customers. Personal Shopping apps like Amazon Personal Shopper. This AI-powered fashion recommendation system that suggests personalized outfits and accessories to customers based on their style and preferences. The system uses computer vision to analyze images and provide accurate recommendations. The existing fashion recommendation systems provide personalized and accurate fashion choices to customers, improving customer satisfaction and loyalty. However, there is still room for improvement in terms of accuracy, speed, and personalization.

1.4. PROPOSED SYSTEM

A proposed fashion recommendation system using clustering would utilize unsupervised machine learning techniques to provide personalized and relevant fashion choices to customers. The system would operate in the following stages:

Data Collection: The system would collect vast amounts of fashion data from various sources, such as social media platforms, e-commerce websites, and fashion blogs. The system would also collect customer data, such as past purchases, preferences.

Feature Extraction: The system would extract relevant features from the fashion data, such as color, texture, pattern, and style.

Clustering: The system would use clustering algorithms, such as k-means, to group fashion items and customers with similar characteristics into clusters. The system would also consider factors such as body type, skin tone, and personal style to ensure accurate clustering.

Personalized Recommendations: The system would provide personalized fashion choices to customers based on their cluster membership. The system would suggest outfits and accessories that match the customer's style.

1.5. OBJECTIVES

The main objectives of a fashion recommendation system are:

Personalization: To provide personalized fashion choices to customers based on their preferences, behavior, and body type, increasing customer satisfaction and loyalty.

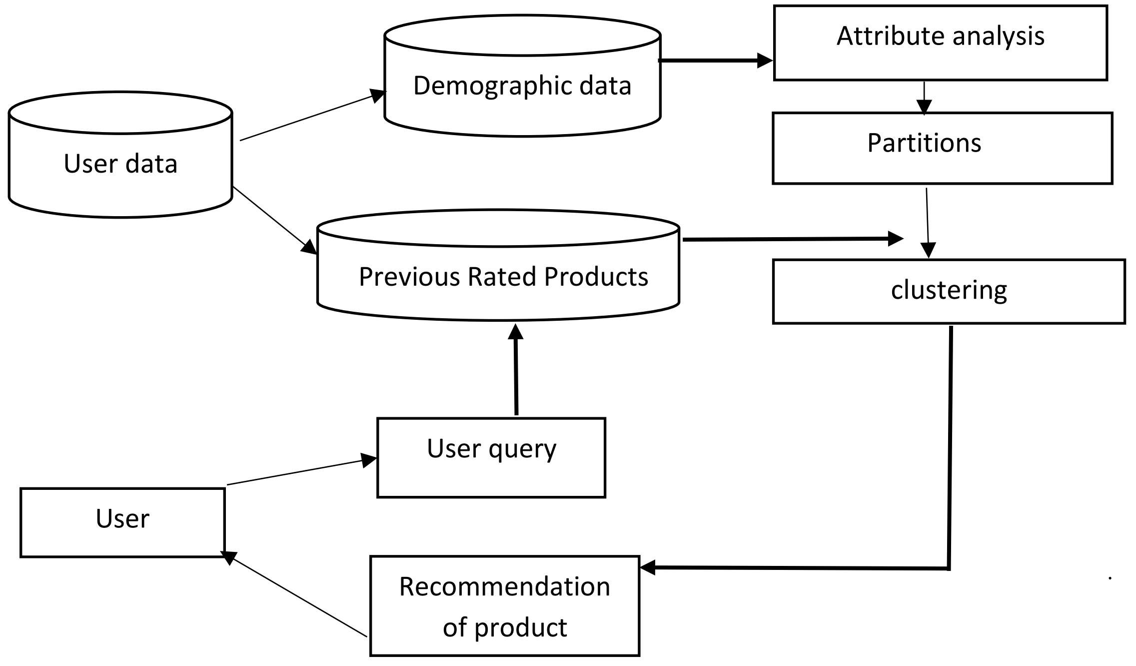
Relevance: To provide relevant fashion choices to customers based on current fashion trends and personal style, improving the chances of purchase and customer retention.

Efficiency: To provide fast and accurate recommendations, reducing the time and effort customers spend searching for the right fashion items.

Brand Loyalty: To increase brand loyalty by providing customers with a unique and personalized shopping experience, building a long-term relationship between customers and brands.

The objectives of a fashion recommendation system aim to provide personalized and relevant fashion choices to customers, increasing customer satisfaction and loyalty, and increasing sales and revenue for fashion brands and retailers.

1.6.ARCHITECTURE



2.1.1 LITERATURE SURVEY

Paper title	Authors	Years	Methodology	Data set	Key findings
"A Clustering-Based Recommendation System for Online Shopping"	Xiaolin Li, Xiaoping Fan	2016	Clustering, Collaborative Filtering	Taobao online shopping platform data	Proposed system outperformed traditional collaborative filtering and content-based recommendation systems in terms of accuracy and coverage.
"A novel Clustering Approach for Online Shopping Recommendation System"	Shuying Han, Kai Song	2018	Clustering, Collaborative Filtering, Association Rule Mining	JD.com online shopping platform data	Proposed system achieved higher accuracy and diversity than traditional collaborative filtering and content-based recommendation systems
"A Hybrid Recommendation System Based on Collaborative Filtering and Clustering for Online Shopping"	Weigang Zhang, Hui Peng	2019	Clustering, Collaborative Filtering, Singular Value Decomposition	Tmall online shopping platform data	Proposed system achieved better performance in terms of accuracy and coverage compared to traditional collaborative filtering and content-based recommendation systems.
"A User-Centric Clustering-Based E-Commerce Recommendation System"	Yunfeng Liu, Jie Liu	2019	Clustering, User Profiling	Amazon online shopping platform data	Proposed system provided personalized recommendations based on user clustering and achieved higher accuracy than traditional collaborative filtering.

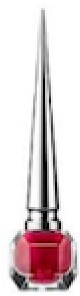
3.DATA PRE-PROCESSING

The data pre-processing steps in this project involve several key components. First, the images are loaded and converted to grayscale to simplify processing. Then, HOG (Histogram of Oriented Gradients) features are extracted from the images to capture important visual information. The resulting feature vectors are then standardized to ensure they have similar scales and are not biased towards any specific feature. Finally, K-means clustering algorithm is used to group the standardized feature vectors into clusters, allowing for easier analysis and understanding of the data. Overall, these pre-processing steps are critical in preparing the data for analysis and ensuring that the resulting insights are accurate and meaningful.

3.1.1 DATASET DESCRIPTION

This dataset consists of images of various fashion products such as shoes, lipsticks, necklaces, rings, handbags, nail polish, watches, earrings, bracelets. Each image has 150 x 150 dimensions.



3.2 DATA CONVERSION

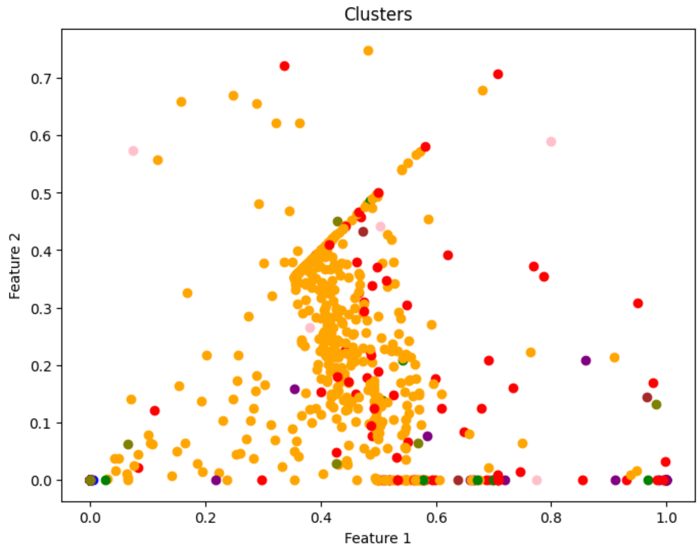
Data conversion from image data to an array is an important preprocessing step in many machine learning and computer vision applications. When working with images, we typically represent them as matrices of pixel values. However, machine learning algorithms require numerical data in the form of arrays or tensors to be able to process the data effectively. One common method of converting image data to an array is by using OpenCV, an open-source computer vision library. OpenCV provides a range of functions to load, process, and manipulate image data in a variety of formats. To convert an image to an array using OpenCV, we first load the image using the imread() function. This function reads the image and returns a numpy array representation of the image in BGR format. We display the image by using pyplot() module we use imshow function from it.

3.3 DATA AUGUMENTATION

Data augmentation is a set of techniques to artificially increase the amount of data by generat ng new data points from existing data. This includes making small changes to data or using d eep learning models to generate new data points, Machine learning applications especially in the deep learning domain continue to diversify and increase rapidly. Data-centric approaches to model development such as data augmentation techniques can be a good tool against challenges which the artificial intelligence world faces.

Data augmentation is useful to improve performance and outcomes of machine learning models by forming new and different examples to train datasets. If the dataset in a machine learning model is rich and sufficient, the model performs better and more accurately.

3.4 DATA VISUALISATION



The x-axis and y-axis of the graph represent two features of the data points. In this case, the features are extracted using the Histogram of Oriented Gradients (HOG) algorithm, which is a technique for object recognition in computer vision. The HOG algorithm calculates the distribution of gradient orientations in the image, which can be used to describe the shape and texture of the object.

Each cluster is represented by a different color. The color coding allows us to easily identify which data points belong to which cluster. The graph can be useful for visualizing the clustering results and identifying patterns or trends in the data. It can also help to identify any outliers or anomalies that do not belong to any of the clusters.

4. METHODOLOGY

4.1 PROCEDURE TO SOLVE THE GIVEN PROBLEM

Machine learning algorithms can be trained on large datasets. These algorithms can then be used to recommends fashion products based on the products that we purchased. It then recommends items to you based on what you like. Clustering algorithm is used to group the similar products together that have similar features.

CLUSTERING:

Clustering is a type of unsupervised machine learning technique that involves grouping similar data points together based on their features or characteristics. Clustering algorithms aim to identify patterns and structures in data without any prior knowledge of the groupings or labels.

The goal of clustering is to divide the data into clusters, where data points within each cluster are similar to each other, and data points across different clusters are dissimilar. Clustering can be used in various applications, such as image recognition, market segmentation, and recommendation systems.

There are several clustering algorithms, including:

K-Means Clustering: This algorithm partitions the data into K clusters by minimizing the sum of the squared distances between each data point and its assigned centroid.

Hierarchical Clustering: This algorithm creates a tree-like structure of nested clusters, where clusters at lower levels are merged to form larger clusters at higher levels.

Density-Based Spatial Clustering of Applications with Noise (DBSCAN): This algorithm groups data points together based on their density and connectivity.

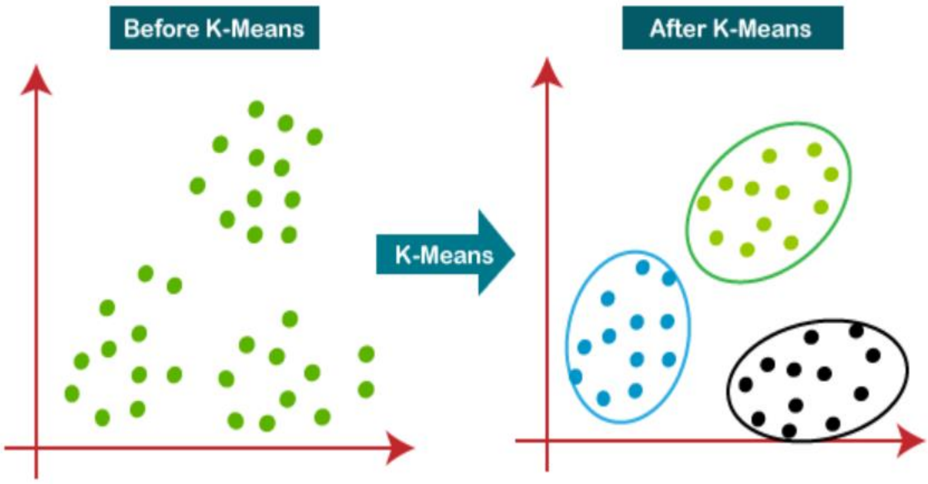
Mean Shift Clustering: This algorithm identifies clusters by locating high-density regions of the data and moving towards the nearest peak.

Clustering algorithms can be evaluated using metrics such as Silhouette score, Calinski-Harabasz index, and Davies-Bouldin index. The choice of clustering algorithm depends on the nature of the data and the problem at hand.

K-MEANS CLUSTERING:

K-means clustering is a popular unsupervised machine learning algorithm used to group data points into clusters based on their similarity. Here is a brief overview of the k-means clustering algorithm:

- Select the number of clusters (k) you want to create.
- Initialize k centroids randomly from the data points.
- Assign each data point to the nearest centroid based on their Euclidean distance.
- Recalculate the centroids of each cluster by taking the mean of all data points assigned to that cluster.
- Repeat steps 3 and 4 until the centroids no longer move or the maximum number of iterations is reached.



The k-means clustering algorithm aims to minimize the sum of squared distances between each data point and its assigned centroid. The algorithm is simple to implement and computationally efficient, making it suitable for large datasets.

However, k-means clustering has some limitations. It assumes that the clusters are spherical and have equal variance, which may not be true for all datasets. It also requires the number of clusters to be specified a priori, which can be challenging for some applications.

To overcome some of the limitations of k-means clustering, variations of the algorithm have been developed, such as hierarchical k-means, fuzzy k-means, and k-medoids clustering. These variations can handle non-spherical clusters, different cluster shapes, and varying cluster sizes.

The k means clustering algorithm includes:

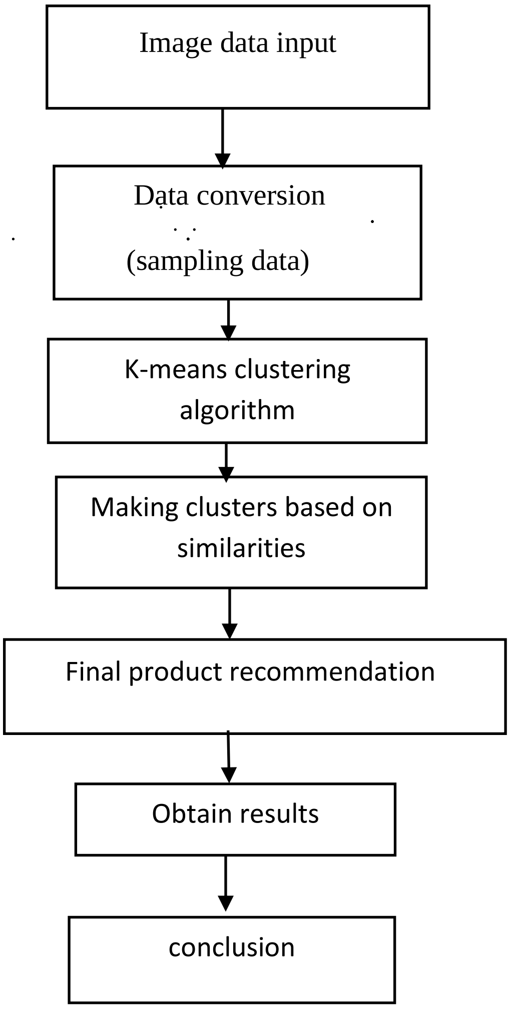
- **Initialization:** The initial positions of the k centroids can significantly impact the clustering results. Random initialization can result in different clustering results each time the algorithm is run. To mitigate this issue, k-means clustering can be run multiple times with different random initializations, and the clustering result with the lowest sum of squared distances can be selected.

- **Choosing the number of clusters:** One of the main challenges in k-means clustering is determining the appropriate number of clusters. A common approach is to use the elbow method, which involves plotting the sum of squared distances as a function of the number of clusters and selecting the number of clusters at the elbow point, where the rate of decrease in the sum of squared distances starts to level off.

- **Convergence:** The k-means clustering algorithm may not always converge to the optimal solution. In some cases, the algorithm may get stuck in a local minimum, which can result in suboptimal clustering results. To address this issue, various extensions of k-means clustering have been developed, such as k-means++, which uses a smarter initialization method to improve convergence.

- **Scaling:** K-means clustering is sensitive to the scaling of the input features. Features with larger scales can dominate the clustering results, leading to biased results. To mitigate this issue, it's often a good practice to normalize or standardize the input features before running k-means clustering.

4.2 MODEL ARCHITECTURE



4.3 SOFTWARE DESCRIPTION

Software requirements:

Operating system: Windows

Platform: google colab

Programing language: python

5. RESULTS

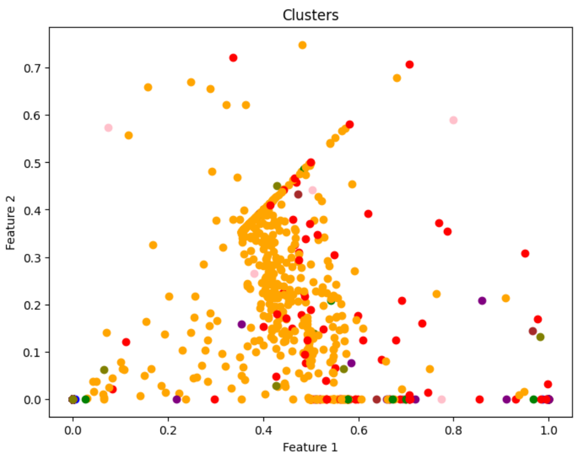
CODE:

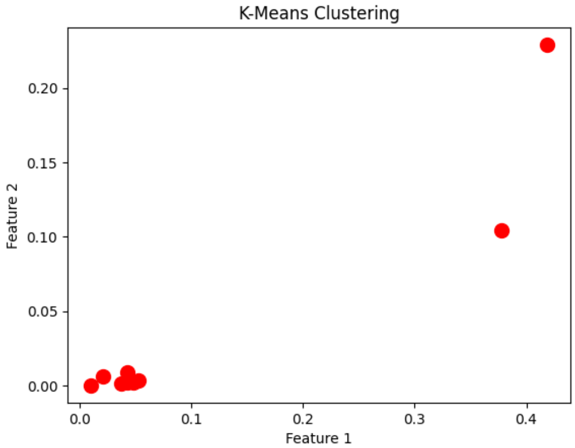
```
from google.colab.patches import cv2_imshow
from skimage.feature import hog
import csv
import cv2
import os
import numpy as np
import matplotlib.pyplot as plt
from sklearn.cluster import KMeans
with open('/content/drive/MyDrive/style[1]/style.csv', 'r') as file:
    csv_reader = csv.reader(file)
    all=[]
    for row in csv_reader:
        fname=row[4]
        all.append(fname)
folder_path="/content/drive/MyDrive/style[1]"
images=[]
features=[]
folder_list=all
labels=[]
print(folder_list)
img_sizeb=100
img_size=100
for fn in folder_list:
```

```
    fp = os.path.join(folder_path, fn)

    if os.path.isfile(fp) and fn.endswith(".png"):

        image = cv2.imread(fp)

        r_arr=cv2.resize(image,(img_sizeb,img_size))

        gray = cv2.cvtColor(image, cv2.COLOR_BGR2GRAY)

        fd,hog_image= hog(gray, orientations=8, pixels_per_cell=(16, 16), cells_per_block=(1,
1), visualize=True, multichannel=False)

        features.append(fd)

        labels.append(fn)

        images.append(r_arr)

max_length = max(len(fd) for fd in features)

features = [np.pad(fd, (0, max_length - len(fd)), mode='constant') for fd in features]

max_length = max(len(fd) for fd in features)

features = [np.pad(fd, (0, max_length - len(fd)), mode='constant') for fd in features]

features = np.array(features)

features = features.reshape(features.shape[0], -1)

kmeans = KMeans(n_clusters=9, random_state=0).fit(features)

import matplotlib.pyplot as plt

fig, ax = plt.subplots(figsize=(8, 6))

colors = ['red', 'blue', 'green', 'orange', 'purple', 'brown', 'pink', 'gray', 'olive']

for i, label in enumerate(kmeans.labels_):

    x = features[i, 0]

    y = features[i, 1]

    ax.scatter(x, y, color=colors[label])

ax.set_title("Clusters")

ax.set_xlabel("Feature 1")

ax.set_ylabel("Feature 2")
```

```
plt.show()


import matplotlib.pyplot as plt
centroids = kmeans.cluster_centers_
plt.scatter(centroids[:, 0], centroids[:, 1], s=100, c='red')
plt.title('K-Means Clustering')
plt.xlabel('Feature 1')
plt.ylabel('Feature 2')
plt.show()


for cluster_id in range(9):
  print(f"Cluster {cluster_id}:")
  for i, label in enumerate(kmeans.labels_):
    if label == cluster_id:
      cv2_imshow(images[i])
      print('\n')
```

OUTPUT:





Clusters **Centroids**

6. CONCLUSION AND FUTURE SCOPE

In conclusion, shopping recommendation systems have become an essential tool for e-commerce websites to provide personalized product recommendations to their users. Clustering-based recommendation systems have emerged as a promising approach to provide accurate and diverse recommendations by grouping users or items into clusters based on their similarities.

Based on the literature survey, it can be observed that clustering-based recommendation systems outperform traditional collaborative filtering and content-based recommendation systems in terms of accuracy and coverage. Moreover, hybrid approaches that combine clustering with other recommendation techniques such as collaborative filtering and association rule mining have been proposed to further improve the performance of the recommendation systems.

For future scope, some potential areas of research and development in shopping recommendation systems based on clustering could include, Integrating contextual information such as location, time, and user behaviour to provide more personalized and relevant recommendations. Developing more efficient clustering algorithms that can handle large and dynamic datasets. Incorporating deep learning techniques such as neural networks to improve the accuracy and diversity of the recommendation systems

7.REFERENCES

[1] Bart Goethals, Efficient Frequent Pattern Mining – Ph.D. Thesis, Universiteit Limburg, pp.21-29, 2002

[2] Jiawei Han, Micheline Kamber, Data mining concepts and techniques 2nd edition, Morgan Kaufmann Publishing pp.234-239, 2005

[3] Teuvo Kohonen, Self-Organizing Maps 3rd edition, Springer Publishing, 2000

[4] Ian H. Witten, Eibe Frank, Data Mining: Practical Machine Learning Tools and Techniques, 2nd Edition

[5] Angel F. Kuri-Morales , Automatic Clustering with Self-Organizing Maps and Genetic Algorithms II: an Improved Approach, Proceedings of the 5th WSEAS International Conference on Neural Networks and Applications, Udine, 2004

[6] Angel F. Kuri-Morales, Automatic Clustering with Self-Organizing Maps and Genetic Algorithms", Recent Advances in Simulation, Computational Methods and Soft Computing, WSEAS Press, 2002

[7] Francesco Maiorana, Performance improvements of a Kohonen self organizing classification algorithm on sparse data sets, Proceedings of the 4th WSEAS/IASME International Conference on Educational Technologies, Corfu, 2008

[8] Agrawal, R., Imielinski, T., Swami, A. Mining associations between sets of items in large databases, Proceedings of ACM SIGMOD International Conference on Management of Data, Washinton D.C., 1993

[9] Agrawal, R., Srikant, R. Fast Algorithms for mining association rules in large databases. Proceedings of 20th International Conference on Very Large Databases, Santiago de Chile, 1994

[10] Charalampos Vassiliou, Dimitris Stamoulis Drakoulis Martakos, A Recommender System Framework combining Neural Networks & Collaborative Filtering Proceedings of the 5th WSEAS International Conference on Instrumentation, Measurement, Circuits and Systems, Hangzhou, 2006

[11] Xuejun Zhang, John Edwards, Jenny Harding, Personalised online sales using web usage data mining, Computers in Industry 58, Elsevier, 2007

[12] Susanne Still, William Bialek, How Many Clusters? An Information-Theoretic Perspective, Neural Computation, Volume 16, Issue 12 2004

[13] Zhiyong Zhang ,Olfa Nasraoui Mining search engine query logs for social filtering-based query recommendation, Applied Soft Computing ,Volume 8, Issue 4, Elsevier September 2008

[14] Olfa Nasraoui, Fabio A. González,Cesar Cardona,Carlos Rojas,Dipankar Dasgupta: A Scalable Artificial Immune System Model for Dynamic Unsupervised Learning, GECCO 2003 [15] Olfa Nasraoui,Raghu Krishnapuram Robust Multi-Resolution Web Usage Mining with Genetic Niche Clustering (2008)