ATTRIBUTES SUBSET SELECTION BASED IMAGE RETRIEVAL USING DATA MINING TECHNIQUE

MRS.DEEPIKA.S[1] , JANANY SREE P[2]

MRS.S. DEEPIKA

ASSISTANT PROFESSOR, DEPARTMENT OF B.COM (BUSINESS ANALYTICS),

PSGR KRISHNAMMAL COLLEGE FOR WOMEN, COIMBATORE, INDIA.

deepika@psgrkcw.ac.in

P. JANANY SREE

UG SCHOLAR, B.COM (BUSINESS ANALYTICS),

PSGR KRISHNAMMAL COLLEGE FOR WOMEN, COIMBATORE, INDIA

jananysree22@gmail.com

ABSTRACT:

The main objective or aim in this paper is to retrieve the images based on their similarity of searches made before. Surfing things is common in this regular world and the instant results about our searches is very important. So taking this as a main objective taking a data set created by ourselves with the attribute that relate to the image size and the image size , image quality , search visits etc. With these attributes images are retrieved according to the similar searches before.

KEY WORDS: Similarity based retrieval , vectors, search visits.

I. INTRODUCTION

This paper is mainly about image retrieval based on the searches we have made . We search for bat in the browser and there is two options relating to that , it may be a sports bat or may the bird. If you are a sports person and surfed mostly about the bat's then the result would be based only on sports bats and nothing more than that**.** For this we open spider and launch it in order to use . First we import the data set and then check whether all the attributes are displayed and then we drop attributes and then find every possible and then we use seaborn graphs. Seaborn is a Python data visualization library based on matplotlib. It provides a high-level interface for drawing attractive and informative statistical graphics.[3] . We use prediction variable and predict things accordingly. Scatter plot is created for every attribute.Variables are created in order to retrieve images based on the searches made before , so we take search visits and rank and some more attributes to retrieve images accordingly

II. REVIEW OF LITERATURE

In Content-based Image Retrieval (CBIR), accurately ranking the returned images is of paramount importance, since users consider mostly the topmost results. The typical ranking strategy used by many CBIR systems is to employ image content descriptors, so that returned images that are most similar to the query image are placed higher in the rank. While this strategy is well accepted and widely used, improved results may be obtained by combining multiple image descriptors. [1]Image retrieval applying deep convolutional features has achieved the most advanced performance in most standard benchmark tests. In image retrieval, deep metric learning (DML) plays a key role and aims to capture semantic similarity information carried by data points. First, when learning the similarity of negative examples, current methods separate negative pairs into equal distance in the embedding space. Thus, the intraclass data distribution might be missed. Second, given a query, either a fraction of data points, or all of them, are incorporated to build up the similarity structure, which makes it rather complex to calculate similarity or to choose example pairs. [2] Semantic based video/image retrieval is a most critical issue in the multimedia search enginerelated research. Multimedia content annotation improves the accuracy of semantic based content retrieval system where multimedia content annotation can be done in two ways: One is content based annotation and another is context based annotation. The content based annotation of images and videos consist of both low level and high level features which would be derived from the detailed pixel intensity information of images whereas context information would provide the semantic details about the images/videos [5]the widespread of smart devices along with the exponential growth of virtual societies yield big digital image databases. These databases can be counter-productive if they are not coupled with efficient Content-Based Image Retrieval (CBIR) tools. The last decade has witnessed the introduction of promising CBIR systems and promoted applications in various fields. In this article, a survey on state of the art content based image retrieval including empirical and theoretical work is proposed. This work also includes publications that cover research aspects relevant to CBIR area.[4]In various application domains such as entertainment, biomedicine, commerce, education, and crime prevention, the volume of digital data archives is growing rapidly. The very large repository of digital information raises challenging problems in retrieval and various other information manipulation tasks. Content-based image retrieval (CBIR) is aimed at efficient retrieval of relevant images from large image databases based on automatically derived imagery features. However, images with high feature similarities to the query image may be very different from the query in terms of semantics (6)

III. METHODOLOGY

A. Seaborn library :

Seaborn may be a library for creating applied mathematics graphics in Python. It builds on high of matplotlib and integrates closely with pandas information structures.

Seaborn helps you explore and perceive your information. Its plotting functions operate data frames and arrays containing whole datasets and internally perform the mandatory linguistics mapping and applied mathematics aggregation to provide informative plots. Its dataset-oriented, declarative API helps you to specialize in what the various parts of your plots mean, instead of on the small print of the way to draw them.

B. Scatter plot:

A scatter plot (also known as a scatterplot, scatter graph, scatter chart, scatter gram, or scatter diagram)is a kind of plot or mathematical diagram victimization philosopher coordinates to show values for generally 2 variables for a collection of knowledge. If the points area unit coded (colour/shape/size), one extra variable are often displayed. the info area unit displayed as a set of points, every having price| the worth} of 1 variable crucial the position on the horizontal axis and also the value of the opposite variable crucial the position on the vertical axis.

C. Prediction variable:

Predictor variable is that the name given to AN variable utilized in regression analyses. The variable provides info on AN associated variable quantity relating to a specific outcome. The term variable arises from a vicinity of applied mathematics that uses applied math to estimate future occurrences of an incident supported collected quantitative proof.
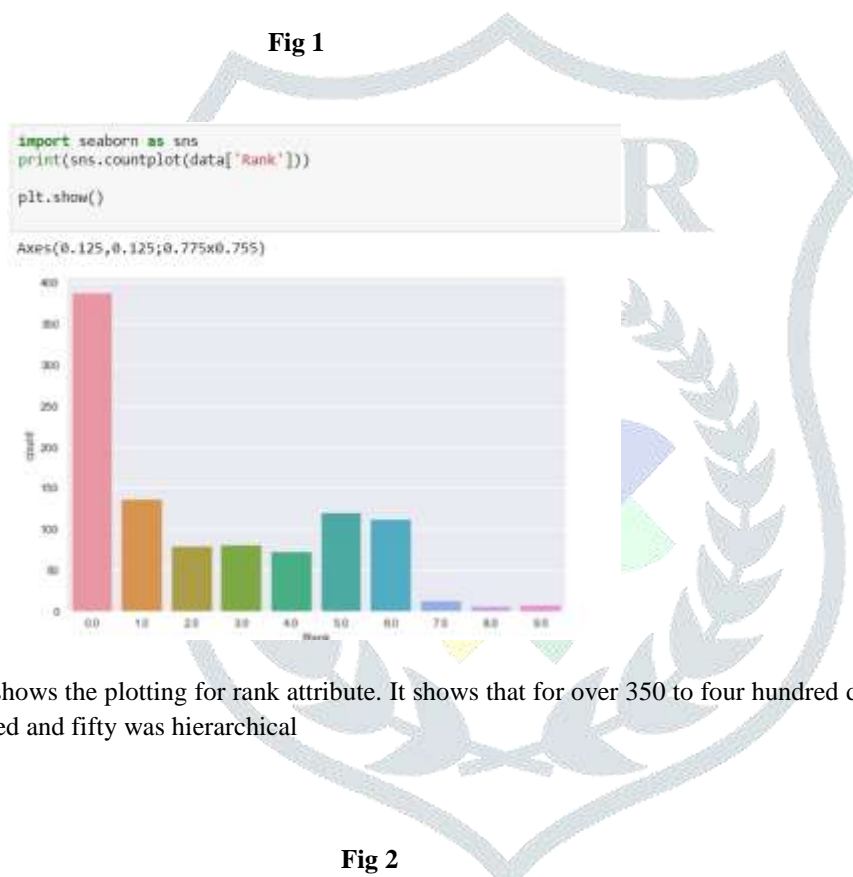
D. Creating vectors:

Create vectors in order to retrieve the images based on similarity using search visits and related attributes. Here we retrieve the images based on their searches and how many times they have searched for a particular product again gain

.

III.RESULT

✚ **Seaborn graph**

Fig 1


```
import seaborn as sns
print(sns.countplot(data['Rank']))

plt.show()

Axes(0.125,0.125;0.775x0.755)
```
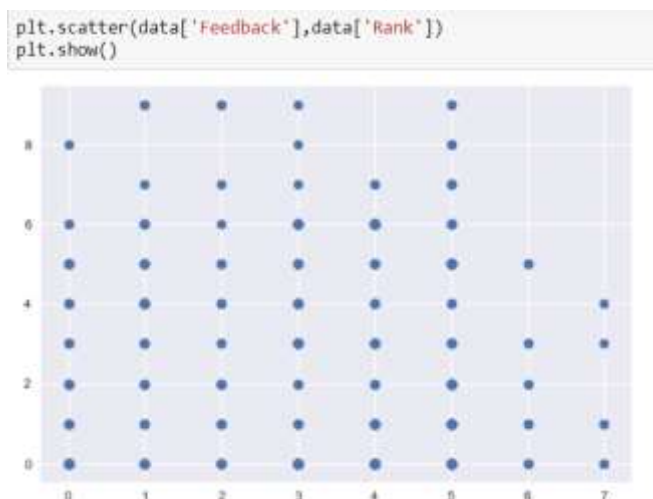
The higher than graph shows the plotting for rank attribute. It shows that for over 350 to four hundred dataset has been hierarchical zero and one hundred to a hundred and fifty was hierarchical
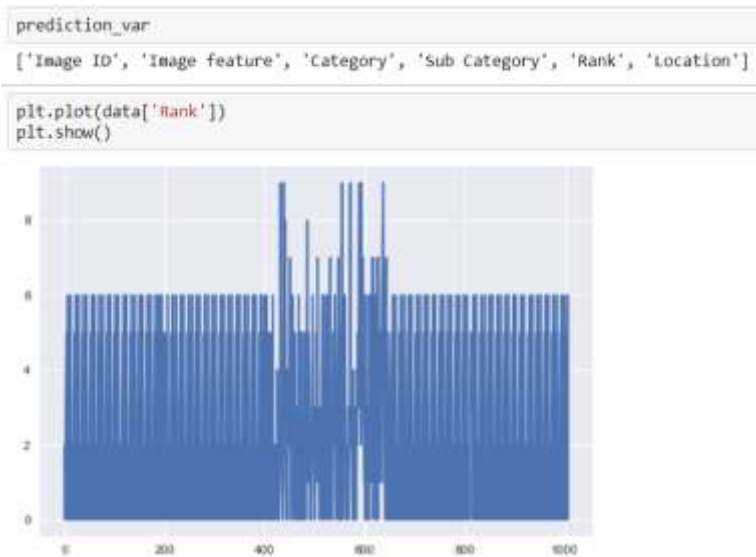
✚ **Scatter plot**

Fig 2


```
plt.scatter(data['Feedback'],data['Rank'])
plt.show()
```

A scatter plot may be used either once one continuous variable that's underneath the management of the experimenter and also the different depends thereon or once each continuous variables square measure freelance. Here the plot is plotted for Feedback and Rank.
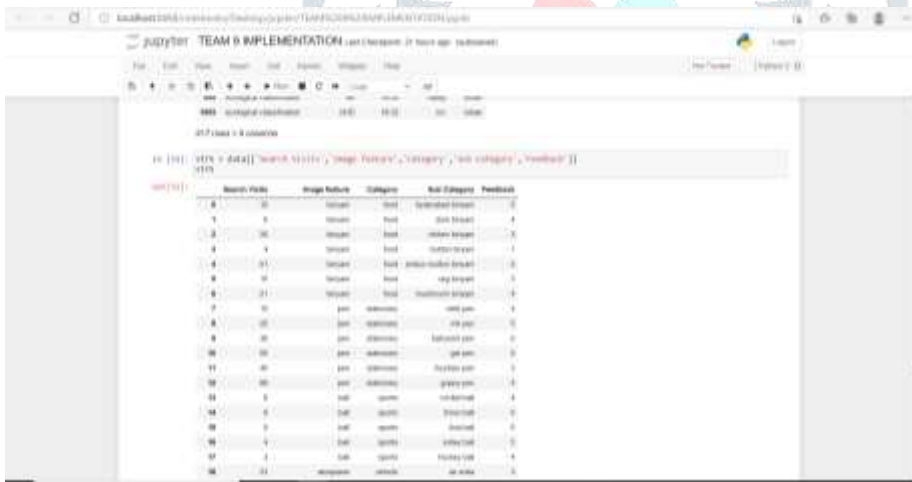
Prediction variable:

Fig 3

```
prediction_var

['Image ID', 'Image feature', 'Category', 'Sub Category', 'Rank', 'Location']
```

```
plt.plot(data['Rank'])
plt.show()
```


This prediction variable is used to select the attributes in which the operations are done so these are the variables **which are used for retrieving.**

Creating vectars:

Fig 4

Based on the image features, category, search visits, sub category and feedback we find the similarity based retrieval. According to the search visits and their search category their search visits the results are given.

Fig 5

Based on the search visits of skin care the products are displayed. The person has searched mostly about skin products so the result is displayed accordingly. This is how the results are displayed according to similarity.

FINDINGS:

- Seaborn graph The higher than graph shows the plotting for rank attribute. It shows that for over 350 to four hundred dataset has been hierarchical zero and one hundred to a hundred and fifty was hierarchical
- Scatter plot is plotted for rank and feedback
- Prediction variable are taken to map the attributes to which the process is done .
- Creating the vector and retrieving the attributes according to similarity.

V. CONCLUSION

The main objective or aim during this paper is to retrieve the photographs supported their similarity of searches created before. surfriding things is common during this regular world and therefore the instant results regarding our searches is extremely vital. therefore taking this as a main objective taking an information set created by ourselves with the attribute that relate to the image size and therefore the image size , image quality , search visits etc. With these attributes pictures area unit retrieved in line with the similar searches before**.**

FURTHER WORK:
It is advised that the strategy of answer will be any extended to K-means cluster algorithmic program for grouping the attributes supported similarity for attributes like search visits and image feature conjointly, it's advised in supplying regression

REFERENCES :

[1] Fabio F. Faria - Institute of Computing University of Campinas Campinas, SP Brazil

[2] Lily Fan - College of Computer Science and Technology, Jilin University, Changchun 130012, China Key Laboratory of Symbolic Computation and Knowledge Engineering of Ministry of Education, Jilin University, Changchun 130012, China

[3] Michael waskom

[4] Mohamed Maher Ben Ismail - College of Computer and Information Sciences, King Saud University, Riyadh, KSA

[5] Tamil Priya - M.E degree in Computer Science and Engineering from Anna University, India. She is Assistant Professor in the Department of Information Technology, VIT University, India. Her research interests include image processing, data mining, computer vision and multimedia, semantic Indexing, and Machine learning

[6] Yixin Chen - The Pennsylvania State University, The Graduate School, Department of Computer Science and Engineering