Introduction:

The main intention of this project is to help user to choose relevant product.

This system makes use of KNN Algorithm. By considering various parameters such as category, review count, price and overall reviews.

Background Information:

In this contemporary era of beauty centric market, choosing an appropriate product is highly necessary to avoid deleterious effects such as skin allergies. Huge number of choices necessitates a procedure to ease options based on attributes which users specify.

In this project I have considered Nykaa dataset[1] which has 3664 rows.

This project gives at most priority to user experience by recommending products according to chosen category, review count, overall reviews, minimum price and maximum price.

To achieve this KNN algorithm is ideal tool to classify. The distance between products can be calculated using Euclidean distance method which takes user ratings(out of 5) and review counts and the range values(minimum price and maximum price)

Learning Algorithm used:

The K Nearest Neighbours algorithm is an ideal choice for segregating beauty products.

This algorithm differentiates products by recognizing the nearest neighbours based on 4 attributes(in this project) review count, ratings, minimum price and maximum price.

The distance between the products is calculated using a simpler formula which makes use user ratings and review count. This formula is represented as shown below.[2]

d=

In this case the x2 is user’s reviews, y2 is user’s ratings. x 1 is user’s input for review and y1 is user’s input for rating. z2 is user’s input for minimum price and maximum price and z1 is the price range present in the file.

Here the nearest 5 neighbours will be considered by default.

Results and Discussion:

A screenshot of a computer

Description automatically generated

Figure 1

A computer screen with a black background

Description automatically generated

Figure 2

In this project users will enter the product category like Shampoo, Soap or Lipstick, after entering this user will type the rating count as prompted by the command prompt as in **Figure 1** after entering rating count they will enter the rating like 100,1000 or above.

After this the algorithm will recommend that particular product to the user as in **Figure 2** if the product is available otherwise it will display an error message.

Error Handling Mechanisms

* If the user enters the values which are out of range then the appropriate message is displayed and ask users to re-enter the values.

The Instruction for execution is present in Execution Instructions Documentation.

Conclusion:

KNN classification is more suitable for classifying beauty products based on user’s reviews and review count, but if more attributes are considered KNN May not be suitable algorithm because of curse of dimensionality, and it does not have a memory.

And according to Authors Nikolaos Kouiroukiddis and Georgios Evangelidis that in high dimensional places distance between nearest and highest points become almost equal.

( <https://ieeexplore.ieee.org/document/6065061>)[3]

So, I plan to include weighted KNN algorithm or Neural Networking for more accurate prediction next time.

Sometimes Companies pay extra amount to fake reviewers to increase their profit as a result there might be fluctuations in the review section, so next stage of this project will be to collect only the verified reviews and proceed with classification.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |

References:

* [1] https://www.kaggle.com/datasets/jithinanievarghese/nykaa-popular-brands-cosmetics-beauty-products
* [2] <https://www.expii.com/t/distance-formula-4560>
* [3] https://ieeexplore.ieee.org/document/6065061