# Research on colour matching recommendation for clothing users based on the DBSCAN clustering algorithm

**DOI: 10.35530/IT.077.02.202510**

REN XIANGFANG
FAN RU
SHEN LEI

**ABSTRACT – REZUMAT**

**Research on colour matching recommendation for clothing users based on the DBSCAN clustering algorithm**

*Colour plays an important role in promoting the sales relationship between clothing brands and users. This article is based on publicly available e-commerce image data from Tmall and the brand's official website. Taking the Taiping Bird brand as an example, a large amount of image data is preprocessed using the Structural Similarity SSIM index, which is used to evaluate image similarity. The image data is vectorised using a non-linear conversion formula from RGB format to HSV format, and K-means clustering analysis is performed to obtain the main colours of the image. Finally, the BDSCAN clustering algorithm is used to cluster the large amount of colour data, continuously modifying the neighbourhood threshold eps and minimum sample threshold minPts parameters until a sufficient number of clusters are obtained and a relatively low noise rate is achieved. Finally, for each category, a greedy algorithm is used to solve for the minimum subset of samples that can represent each cluster. The results are presented in the form of web pages, divided into two-page entrances for users and brand owners. It provides competitive brand colour recommendations and real-time sales to brand owners, and personalised colour preference recommendations and direct links to recommended products to users.*

***Keywords:*** *BDSCAN clustering algorithm, greedy algorithm, computer vision, clothing, colour matching recommendation*

**Cercetare privind recomandările de asortare a culorilor pentru utilizatorii de îmbrăcăminte, pe baza algoritmului de grupare DBSCAN**

*Culoarea joacă un rol important în promovarea relației comerciale dintre brandurile de îmbrăcăminte și utilizatori. Acest articol se bazează pe date imagistice din comerțul electronic, disponibile public, provenite de pe Tmall și de pe site-ul oficial al brandului. Luând ca exemplu brandul Taiping Bird, o cantitate mare de date imagistice este preprocesată folosind indicele de similitudine structurală SSIM, utilizat pentru evaluarea similitudinii imaginilor. Datele de imagine sunt vectorizate folosind o formulă de conversie neliniară din formatul RGB în formatul HSV, iar analiza de grupare K-means este efectuată pentru a obține culorile principale ale imaginii. În cele din urmă, algoritmul de grupare BDSCAN este utilizat pentru a grupa cantitatea mare de date de culoare, modificând continuu parametrii pragului de vecinătate „eps" și pragul minim de eșantionare „minPts" până când se obține un număr suficient de grupuri și se atinge o rată de zgomot relativ scăzută. În cele din urmă, pentru fiecare categorie, se utilizează un algoritm de tip „greedy" pentru a determina subansamblul minim de eșantioane care poate reprezenta fiecare cluster. Rezultatele sunt prezentate sub formă de pagini web, împărțite în intrări de două pagini pentru utilizatori și proprietari de mărci. Recomandări competitive privind culorile mărcii și vânzări în timp real sunt oferite proprietarilor de mărci, precum și recomandări personalizate privind preferințele de culoare și linkuri directe către produsele recomandate utilizatorilor.*

***Cuvinte-cheie****: algoritmul de grupare BDSCAN, algoritm de tip „greedy", viziune computerizată, îmbrăcăminte, recomandări privind asortarea culorilor*

## INTRODUCTION

The style, colour, and pattern of clothing are collectively known as the three key elements of fashion. These elements are not only crucial criteria for consumers when making purchasing decisions but also vital considerations for brands during the design and production process. Against the background of the rapid development of the Internet, the clothing industry, as an important part of the traditional manufacturing industry, has also faced huge development challenges [1]. Consumers' colour preferences for clothing vary across different seasons and times. For example, in summer, consumers tend to prefer light colours with low saturation that evoke a sense of calm, while in autumn, they favour warm and stable colours like warm browns or black. For brands, understanding the current colour preferences of consumers is a crucial task. For consumers, while following trends, maintaining their own colour preferences without blindly following the crowd is an important consideration in life. To meet these needs, marketing channels are diversified, including vertical chain, franchise, direct sales, commission sales, and horizontal online businesses, stores and webchat business, etc., which need different business methods to gradually innovate [2].

The so-called colour marketing is that enterprises use consumers' emotional demands to integrate colours, realize the organic unity of items, colours and people, to attract consumers' attention, promote consumer consumption, strengthen consumers' impression of the brand, and make consumers form a chain reaction of thinking on the colour, to realize the purpose of long-term stable development of enterprises and maximize the efficiency of commodity marketing [3]. Brand designers need to consider consumers' age, gender, psychological needs, cultural background and other differences, as well as consumers' experience of colour [4].

With the rapid development of computer technology and major e-commerce platforms, both brands and customers have more ways and methods to analyse market data and make choices. For brands, understanding the marketing market and the colour needs and preferences of consumers is essential for increasing sales. During the marketing process, colour is used to convey the company's image, complete product packaging design, advertising, and to effectively connect with the market [5]. For consumers who follow market trends and pursue individuality, appropriate colour references are also an indispensable part. With the widespread use of social media and its deep integration into everyday life, consumer engagement has reached unprecedented levels. Consumers are more inclined to use social media and online reviews to gather product information, which significantly influences their purchasing decisions [6].

The combination of computer vision and neural network algorithms has become a popular research method. For instance, Shuguang et al. utilised an improved VGGNet neural network combined with YOLOv3, Faster R-CNN, and SSD object detection algorithms to achieve pattern recognition and localisation [7]. Liang developed a computer colour matching system design based on machine learning, which, by continuously collecting user feedback, imitates the decision-making patterns of human designers in the colour matching process, thereby optimising colour matching schemes and providing efficient and personalised colour matching suggestions to better meet consumer needs [8]. Yani, based on deep learning knowledge, proposed a colour matching board generation network (SPCPN) that integrates user emotions with popular colours and colour matching rules, and an interactive fabric image colour matching network (IFCN). She built an intelligent fabric colour matching system based on Pyqt5, realising an interactive intelligent fabric colour matching system primarily aimed at designers [9].

It is not difficult to see that the data information contained in the open network can be used to identify the interests, needs and behaviour habits of the target audience through data analysis technology, so as to realise the push of personalised content and the accurate dissemination of marketing information. This personalised marketing strategy not only improves the information arrival rate, but also enhances the user's sense of participation and satisfaction [10]. Therefore, enterprises urgently need to use digital means to improve the quality of products and services. Relying on advanced digital technologies, companies can precisely monitor market dynamics and consumer preferences [11].

In this study, a hierarchical colour analysis framework is proposed, which combines K-means++ with the DBSCAN density clustering algorithm and is applied to the field of clothing colour recommendation. In the implementation of the traditional K-means algorithm, the sample data is assumed to be spherical clusters. However, clothing colour often presents complex distribution in HSV space, such as multimodal characteristics of young women's clothing. DBSCAN automatically identifies arbitrary shape clusters and separates noise through density threshold, and combines K-means++ for dominant colour extraction, which significantly improves the robustness of feature representation.

And, existing tools like Huemint and Palettemaker only provide static colour schemes. The platform integrates consumer demand and click volume, dynamically displays the seasonal colour migration trajectory, and supports real-time screening of user preferences. Compared with commercial software, this platform realises cross-category adaptive recommendation, such as the independent modelling of the colour scheme of young women's clothing and the monochromatic business style of men's clothing, which provides ideas for the colour matching of subsequent clothing colours in a single category.

## RESEARCH PRINCIPLES

The data used in this study were sourced from the Tmall e-commerce platform and the official websites of various brands.

The dataset plays a crucial role in the outcomes of machine learning, and web data obtained via web scraping tools may contain duplicates, damage, or low-resolution issues. This paper employs the Structural Similarity Index (SSIM) to perform data cleaning on a large set of image data. Samples with SSIM $\geq$ 0.9, low pixel count, or damage are removed before selecting the primary colours of the images. The RGB format is converted to HSV format using a linear transformation formula, which avoids the issues of high correlation between indicators and human visual errors inherent in the RGB colour space. Subsequently, K-means clustering analysis is conducted to extract the cluster centres as the main colours of these images. After obtaining the main colours of the images, the BDSCAN clustering algorithm is applied to cluster the colour data, continuously adjusting the neighbourhood threshold eps and the minimum sample threshold *minPts* parameters until a suitable number of clusters is obtained with a relatively low noise rate. Finally, a greedy algorithm is used for each category to find the smallest sample subset that best represents each cluster (figure 1).

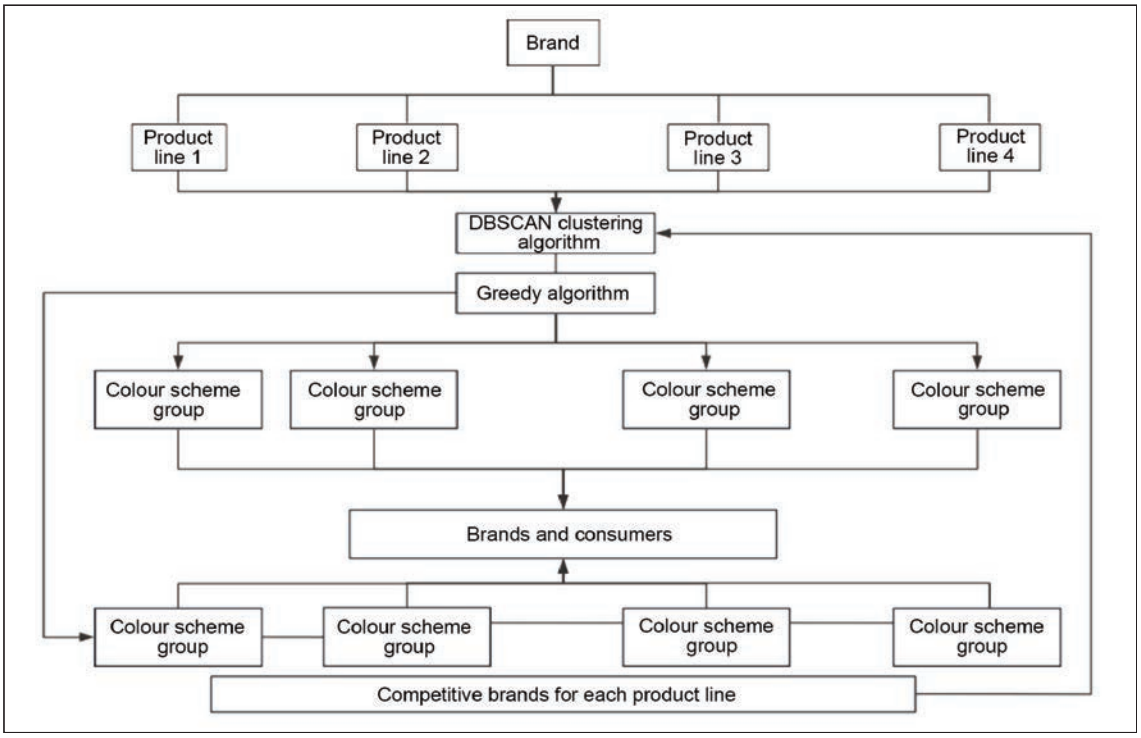


Fig. 1. Schematic diagram of the research process

## Data cleansing

The selection of data is crucial to the effectiveness of algorithm models. Common image processing methods based on web crawler technology include up and down flipping, mirror flipping, adding salt and pepper noise and other preprocessing and data enhancement operations [12]. Considering the redundancy, duplication, and low-resolution images on websites, this paper adopts the Structural Similarity Index (SSIM). By calculating the average luminance of the images ($\sigma_x$, $\sigma_y$), the standard deviation of image luminance ($\mu_x$, $\mu_y$), and the luminance covariance between two images ($\sigma_{xy}$), and combining these with the simplified SSIM (formula 1) results, the similarity between images is evaluated. Images with a similarity above 0.9 are removed.

$$SSIM(x,y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \tag{1}$$

where $C_1 = (k_1L)^2$, $C_2 = (k_2L)^2$,

and $L = 2^B - 1$ (image pixel range), $k_1 = 0.01$, $k_2 = 0.03$

Additionally, images with a resolution lower than 10*10 pixels and damaged images are removed to ensure the stability and accuracy of the model. The number of images after cleaning is shown in table 1, and the cleaned data are used as the input samples for machine learning.

Table 1

DATA CLEANING RESULTS (PART 1)

| Brand | Before washing | After washing |
|---|---|---|
| PEACEBIRD Women | 2944 | 2848 |
| PEACEBIRD Men | 2605 | 2450 |
| PEACEBIRD Kids | 2092 | 2003 |
| Ledin | 2960 | 2753 |

Table 2

DATA CLEANING RESULTS (PART 2)

| Brand | Before washing | After washing |
|---|---|---|
| Zara & Ochirly | 2528 | 2311 |
| Youngor & GXG | 8276 | 7252 |
| Balabala & Annil | 6809 | 5206 |
| Elf Sack & H&M | 2685 | 2144 |

## Basic algorithms

*RGB model and HSV model*

The RGB model (figure 2) is a commonly used colour representation model composed of three channels of data, red, green, and blue, the three primary colours of light, which are combined to form visible colours. The RGB model is widely used for colour storage and computation within computers, with the advantage of being intuitive and easy to understand. However, the three variables that make up the model are highly correlated, meaning that a change in one channel can potentially affect the values of the other two channels. The difference between colours cannot be measured by the distance between RGB values, and the same RGB values can display differently on different devices, making it unsuitable for standardised representation and inter-device communication.

On the other hand, the HSV model (figure 3) defines colour by hue, saturation, and brightness, which not

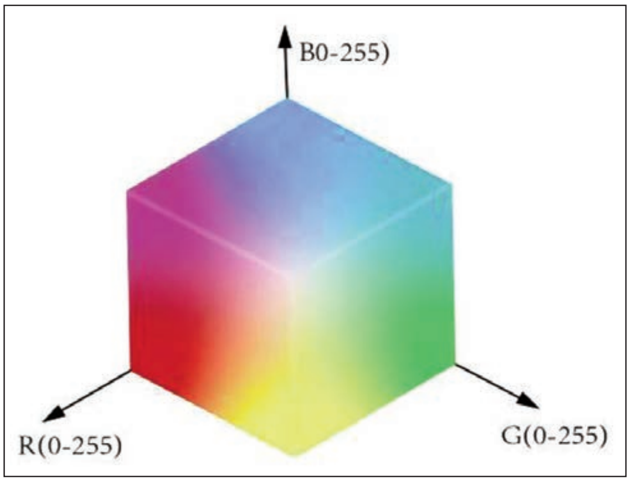


Fig. 2. RGB colour space model

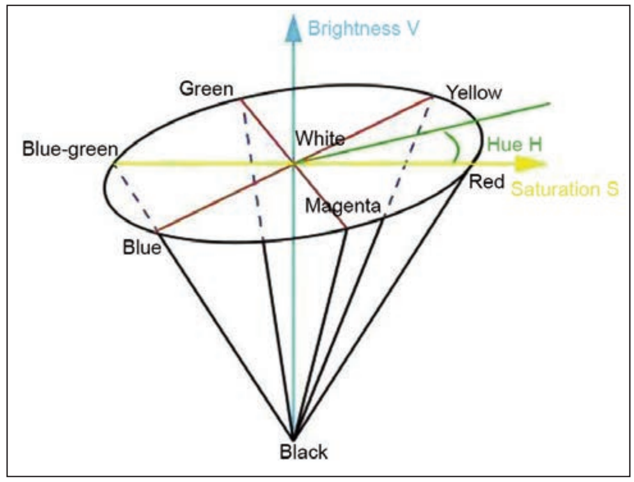


Fig. 3. HSV colour space model

only avoids the issue of high correlation in RGB channels but also makes colour representation closer to what the human eye perceives. Hue represents the type of colour perceived by the human eye, with the entire hue circle being a central angle of 360◦, where different angles correspond to different colours perceived by the human eye; brightness represents the lightness or darkness of the colour; and saturation represents the purity of the colour [13]. After transformation, the RGB data stored within the computer can be non-linearly mapped into the HSV colour space. The specific steps for this are:

Normalisation of RGB (formula 2)

$$R' = \frac{255}{R},\ G' = \frac{255}{G},\ B' = \frac{255}{B} \tag{2}$$

where $R$, $G$, and $B$ represent the corresponding three-channel data in the RGB colour model of the image.

Calculation of Brightness $V$ (formula 3)

$$V = \max(R', G', B') \tag{3}$$

3. Calculation of Saturation $S$ (formula 4)

$$S = \frac{V - \min(R', G', B')}{V} \tag{4}$$

4. Calculation of Hue $H$ (formula 5)

$$H = \begin{cases} 60° \times \left(\frac{G' - B'}{V - \min(R', G', B')}\right) \bmod 6 & \text{if } V = R' \\ 60° \times \left(\frac{B' - R'}{V - \min(R', G', B')} - 2\right) & \text{if } V = G' \\ 60° \times \left(\frac{R' - G'}{V - \min(R', G', B')} + 4\right) & \text{if } V = B' \end{cases} \tag{5}$$

## METHODS

### Colour extraction in computer vision

This paper opts to perform computer vision analysis in the HSV colour space. Using a nonlinear transformation formula from RGB to HSV format, the image data is vectorised, and a two-dimensional array is created for each image [14, 15]. The vectorised image data is then subjected to K-means clustering analysis, with the cluster centres formed by this unsupervised algorithm extracted as the main colours of the image. The primary steps of the K-means algorithm are as follows:

1. Random initialisation: randomly select K data points from the dataset as initial cluster centres, where K represents the target number of clusters.
2. Assignment: assign each data point to the nearest category based on distance $D_i$, using the Euclidean distance (formula 6), where $x_i$ and $y_i$ are corresponding indicators of the same dimension in different two-dimensional arrays:

$$D_i(x,y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2} \tag{6}$$

3. Recalculation: recalculate the mean of all data points within each category, using the obtained mean as the new cluster centre for that category.
4. Iteration: repeat steps 2 and 3 until the cluster centres no longer change significantly or the preset number of iterations is reached. The preset number of iterations is 1000, and the preset threshold for cluster centre change is 0.01.

The within-cluster sum of squares (SSE) formula for K-means clustering is as follows:

$$SSE = \sum_{K=1}^{k} \sum_{X \in c_k} ||x - \mu_k||^2 \tag{7}$$

The clustering results are as follows, but considering the influence of the background colour of the image, $K=5$ is selected to retain 90% of the main colour information while avoiding the recommendation redundancy caused by excessive segmentation. In practical application, the height can be selected appropriately according to the extraction situation of the dominant colour (figure 4).

The K-means clustering method does not separate noise data, making it sensitive to the number of clusters and noise data. The number of clusters can be selected based on rules such as the elbow method. However, subjective parameter selection can result in significant differences in clustering results. Therefore, this paper defaults the number of clusters to 5,

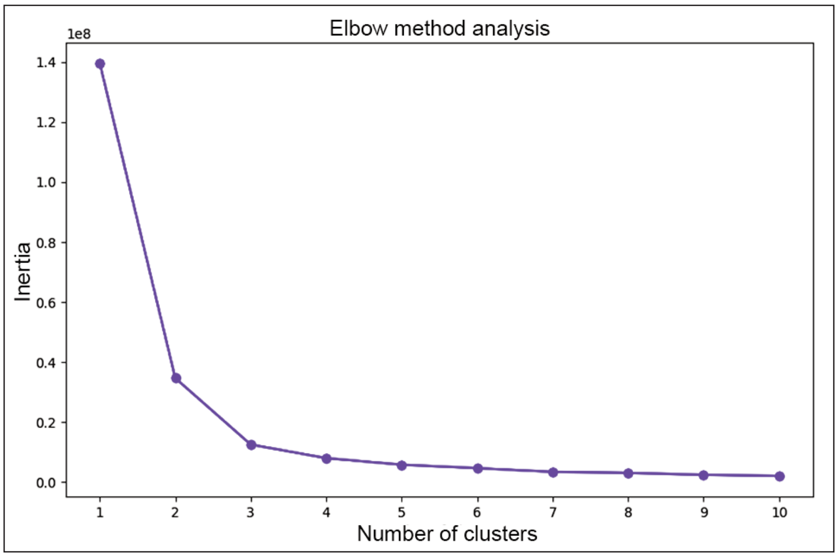

Fig. 4. SSE for cluster number selection

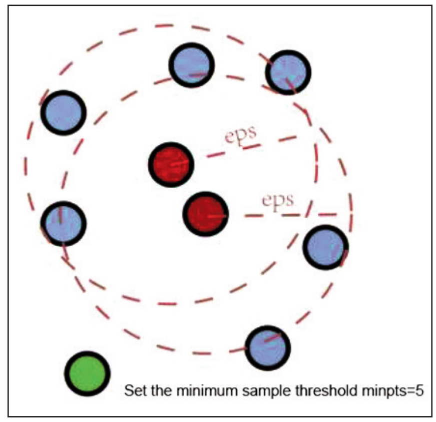

Fig. 5. Sample display example

assuming that the cluster centres of 5 clusters represent the five primary colours of the image.

**DBSCAN clustering**

DBSCAN (Density-Based Spatial Clustering of Applications with Noise) is a density-based spatial clustering algorithm that divides high-density regions into clusters based on the density differences inherent in the sample set. It defines clusters as the largest set of density-connected points. Unlike the K-means algorithm, which requires the user to specify the number of clusters, the DBSCAN algorithm achieves automatic clustering based on the density of data points under determined parameters, thus avoiding the influence of subjective cluster number selection on results and enabling the separation of noise data from the sample set. The core idea of DBSCAN is to divide the data points to be clustered into core points, edge points, and noise points based on the specified neighbourhood threshold $\epsilon$ and the minimum sample threshold *minPts*. The steps of the algorithm are as follows:

1. Initialisation: mark all points as unvisited.
2. Iteration: randomly select an unvisited point, calculate the distance from the other points to this point, and count the number of points within the neighbourhood threshold $\epsilon$. If the number is greater than or equal to the minimum sample threshold *minPts*, the point is considered a core point, and a new cluster is created, with the point added to this cluster. If the number is less than the minimum sample threshold *minPts*, the point is considered a noise or edge point, and it is skipped, continuing to the next unprocessed data point.

As shown in the figure 5, if the number of points is greater than the set neighbourhood threshold $\epsilon$, the point is marked as a red core point, while points within the neighbourhood threshold $\epsilon$ are marked as blue edge points. The above steps are repeated until all points have been traversed.

3. Formation of clusters: for each core point within a cluster, find all points within the neighbourhood threshold $\epsilon$. If any of these points are unprocessed core points, add them to the cluster and continue searching for points within the neighbourhood threshold $\epsilon$ of these newly added points until no more core points can be added. If any of these points are edge points, add them to the cluster as well.
4. Determination of noise points: repeat steps (2) and (3) until all points have been processed. At this stage, any data points that have not been added to any cluster are considered noise points.

Through these steps, the data can be divided into different clusters, and noise data can be separated. In this algorithm, the choice of neighbourhood threshold $\epsilon$ and minimum sample threshold *minPts* parameters is crucial and needs to be determined through multiple trials or based on experience in the target domain.

In the parameter selection of the DBSCAN algorithm, this study fixed *minPts* as the empirical value range (5–15), and generated candidate parameter combinations in the eps $\in$ [0.1, 0.3] interval with a step size of 0.01. The initial value is set based on the data density distribution. For example, the data distribution of men's clothing is sparse, and the initial eps is 0.18. Young women's clothing is data-intensive with initial eps = 0.15. Then, the largest silhouette coefficient is selected to measure the closeness between the sample and its cluster and the separation between the sample and other clusters. The formula is as follows:

$$\bar{s} = \frac{1}{N}\sum_{i=1}^{N}\frac{b(i) - a(i)}{\max\{a(i), b(i)\}} \tag{8}$$

where $a(i)$ is the average distance from the sample to other samples in the same cluster, and $b(i)$ – the average distance from the sample to the nearest different cluster sample. The contour coefficient and noise rate were used for manual verification, and the second derivative method was used to identify the critical point where the NR decline rate was significantly slowed down, and the parameters with a large contour coefficient were confirmed to confirm the rationality of the cluster structure.

### Greedy algorithm

The core idea of the greedy algorithm is to break the entire problem into identical subproblems. First, make the optimal solution for the smallest subproblem based on the limited information available, and then gradually improve the information to obtain the optimal solution for the entire problem. After obtaining the clustering results from DBSCAN, this paper extracts the most representative *k* colours from each cluster of each brand, which are the most distinct primary colours. To find the most representative sample subset, this paper considers the *k* representative colours as a dynamically changing sample subset *T*. When the number of nodes in the dynamically changing sample subset reaches NUM, the iteration is performed based on the existing sample subset. The smallest Euclidean distance *Di* from the points in the remaining set NOT to the optimal sample subset is calculated, and the sample point with the largest minimum Euclidean distance is added to the optimal sample subset. The process is as follows:

1. Data grouping: group the data based on the clustering categories of each colour obtained from the DBSCAN algorithm.
2. Iterative step: initially, set the number of nodes in the dynamically changing sample subset NUM = 0, and select the first sample point to be added to the sample subset *T*. For each sample in the selected sample subset *T*, calculate the minimum distance between it and the unselected samples NOT, and select the sample that is “farthest” from *T* (i.e., the sample point in NOT with the largest minimum distance from *T*).
3. Repeat step (2): continue until *k* samples are selected or all samples have been considered.

## RESULTS AND ANALYSIS

### DBSCAN clustering results

The neighbourhood threshold ϵ defines the distance threshold within which data points are considered “close” to each other. A too large ϵ might lead to too many points being grouped into the same cluster, or even all points being classified into a single cluster, thus losing the local structural information of the data. Conversely, a too small ϵ might result in most points being considered as noise, leading to too many clusters or clusters with too few sample points. The minimum sample threshold *minPts* determines the minimum number of points required to form a cluster within a given neighbourhood. A higher *minPts* value can reduce the number of noise points, but also makes it harder to form clusters, possibly causing some small but meaningful clusters to be overlooked. Conversely, a lower *minPts* value might capture more small clusters but also introduce more noise points.

Therefore, selecting appropriate ϵ and *minPts* parameters is crucial for clustering effectiveness. The parameter tuning results, as shown in table 3, illustrate the outcomes of selecting these two core parameters and the corresponding noise rates for different clusters. This reflects the varying sensitivity of the algorithm to different data distribution characteristics and also reveals some of the internal structural features of the data.

For example, the colour clustering parameters for Lativ are ϵ = 0.154 and *minPts* = 10, with a noise rate of 13.77%. Similarly, for Peacebird Men, the parameters are ϵ = 0.152 and *minPts* = 10, with a noise rate of 13.36%, indicating that the density of data points within these two brands is comparable [16].

In contrast, although Elf Sack and H&M share the same colour clustering parameters (ϵ = 0.15 and *minPts* = 10), the noise rate for Elf Sack is 22.83%, which suggests a greater variability in data point density for Elf Sack compared to H&M. This contrast underscores the differing internal structures within the datasets for these brands, demonstrating how DBSCAN responds to the inherent characteristics of each data distribution.

In the results shown in table 3, the data points identified as noise by the DBSCAN algorithm, those that do not meet the criteria for forming clusters, often represent rare or non-mainstream colour choices in market applications. Although these colours have lower usage frequencies, they may symbolise innovation, uniqueness, or the preferences of specific target markets for brands. By analysing these noise data points as separate clusters, brands can understand which colour preferences are held by a small but loyal group of consumers [17]. This insight allows brands to design products or services that cater to this demographic or to use the colours found in the noise data as design inspiration to create novel and distinctive brand images or product lines that attract consumers seeking personalisation and differentiation.

Table 3

DBSCAN PARAMETER SELECTION

| Brand | Brand | | Noise rate (%) | Brand | Comparison brand | | Noise rate (%) |
|---|---|---|---|---|---|---|---|
| | ϵ | *minPts* | | | ϵ | *minPts* | |
| Peacebird Women | 0.17 | 10 | 10.96 | Zara & Ochirly | 0.165 | 12 | 14.86 |
| Peacebird Men | 0.152 | 10 | 13.36 | Youngor & GXG | 0.28 | 12 | 1.63 |
| Peacebird Kids | 0.15 | 10 | 16.61 | Balabala & Annil | 0.2 | 12 | 6.94 |
| Ledin | 0.154 | 10 | 13.77 | Elf Sack & H&M | 0.15 | 10 | 22.83 |

Therefore, although noise data in the DBSCAN algorithm is traditionally considered useless or data points that need to be excluded, in the context of market applications, it contains valuable information and potential significance. By appropriately adjusting the $\epsilon$ and *minPts* parameters in DBSCAN and fully leveraging both the clustered data and the noise data, brands can gain a more comprehensive understanding of market demand. This approach can help brands stand out in a highly competitive market and achieve differentiated development.

**Greedy algorithm results**

After obtaining the clustering results for each colour, this study aims to construct an optimal sample subset such that each subset can represent a cluster category. Based on the cluster to which each colour belongs, the first element from each cluster is initially selected to be included in the optimal sample subset. Then, the minimum Euclidean distance from the remaining data points to the points in the optimal sample subset is calculated. The data point with the maximum minimum Euclidean distance is then added to the optimal sample subset. The target number of samples in the subset is set to $k$ = 10.

Using the strategy described above, a representative optimal sample subset can be constructed, which effectively reflects the main characteristics and distribution patterns of the colours in each cluster from the original dataset. This has significant practical implications for subsequent applications, such as visualisation analysis and market research.

**Result output analysis**

This paper aims to use an innovative and user-friendly approach to bridge the gap between consumers and brands regarding colour trends and applications through a carefully designed colour-selection website. This platform not only provides users with more ways to incorporate their preferences into fashion choices but also offers brands valuable insights into real-time changes in the consumer market. The website serves not only as a colour exploration platform but also as a tool for enhancing market insights and optimising brand strategies.


Fig. 6. Homepage of the colour-selection website

The homepage of the colour-selection website (figure 6) is designed to be simple yet visually appealing, striking a fine balance between aesthetic appeal and functional information delivery. The homepage offers clearly distinguished entry points for different types of users (general consumers and brands), guiding them quickly to the areas of content they are most interested in through intuitive icons and brief descriptions. This design approach not only enhances the user experience but also reflects a deep understanding of the needs of the target user groups. Additionally, the site includes sections for database sources and user suggestions, providing explanations about the sources of the public website data and creating a channel for real-time communication between developers and users.

Upon entering the consumer interface (figure 6), users can effortlessly select different clothing categories, such as women's wear, men's wear, children's wear, or accessories. As users make their selections, the interface dynamically displays the colour usage across various brands within the chosen category. Users can click on colours of interest or use filters to refine their preferences, allowing them to browse colour choices and clothing styles presented in the form of intuitive colour blocks and thumbnails. This makes the relationship between colours and clothing styles clear at a glance. When a user clicks on a specific colour, the page redirects to a detailed display of garments using that colour, complete with direct links to e-commerce platforms, fully satisfying both aesthetic and shopping needs.

After registering through the website or mobile terminal, users can fill in the basic information in the personal centre and select the initial colour preference (such as cold tone, high saturation, etc.) through the interactive colour tray. The system will dynamically update the preference model according to this information and the update of the e-commerce platform. Based on the DBSCAN clustering results and the 10 representative colours generated by the greedy algorithm, combined with the user's historical behaviour data, the clothing scheme that conforms to their aesthetic preferences will be recommended.

For brands, the colour-selection website serves as a powerful tool for market analysis and strategic planning. The brand interface (figure 7) focuses on analysis, offering precise matching and visual comparison of similar colour schemes between the brand's products and those of competitors in the market. This feature helps brands quickly identify their strengths and weaknesses in colour usage. Additionally, brand representatives can view product images from competing brands on e-commerce platforms and access direct links to sales data on platforms like Tmall and Taobao. This data provides robust support for refining market positioning, product optimisation, and marketing strategy adjustments.


Fig. 7. Consumer interface of the colour-selection website, colour comparison, and competitor

In addition, enterprises can view the category colour clustering results in real time, including the mainstream colour distribution, noise point characteristics and time series trends. At the same time, based on the clustering parameters and noise rate, enterprises can identify the market segments and adjust the product line design.

For example, limited edition colour series are launched for high noise rate categories. In terms of product promotion, companies can use colour sensitivity combined with geolocation data to deliver customised ads on social media platforms, such as high-brightness colour ads in tropical regions. High saturation of colour is often accompanied by decorative and symbolic performance, making the dress full of vitality [18].

For the prediction and verification of the fashion industry, users can carry out real-time trend capture and verification based on this platform. By continuously crawling the latest clothing data of the e-commerce platform, combined with DBSCAN clustering results, it can help experts verify the accuracy of fashion colour prediction. For example, if Pantone predicts that “mocamus” will become mainstream in 2025, the platform can provide empirical support by analysing the growth of searches for this colour in the young women’s clothing category. Moreover, the platform supports the comparison and visualisation of multi-category colour distribution, and experts can quickly identify common trends, which provides a basis for cross-domain colour scheme design. The application can continue to expand to home furnishing, textile and other categories.

In summary, the colour-selection website is a powerful tool for market analysis and strategic planning. Brands can use it as a practical tool focused on colour comparison and competitor analysis, allowing them to precisely match and visually compare similar colour schemes between their products and those of competitors in the market. This helps brands quickly identify their strengths and weaknesses in colour usage and provides more timely references for observing market trends. Meanwhile, users can see it as a convenient platform for exploring fashion colours and discovering their preferred clothing.

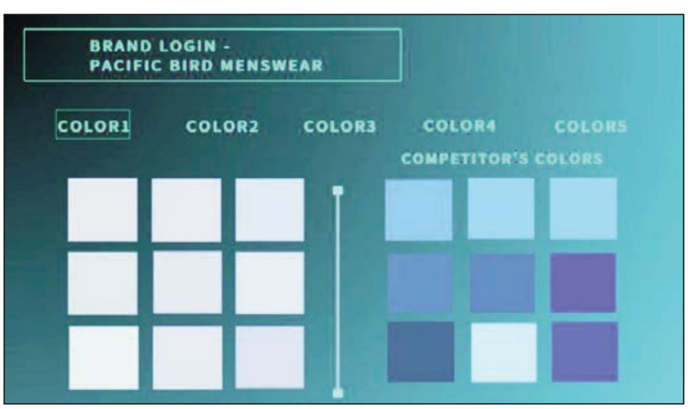


Fig. 8. Brand interface of the colour-selection website

## CONCLUSION

At present, the model focuses on clothing colour analysis, but does not integrate the key design elements such as style cut, fabric texture and pattern. In future research, multi-scale Gabor filters can be introduced to extract fabric texture features (such as the gloss of silk and the roughness of wool), and the HSV colour space can be combined to construct cross-modal feature vectors. The attention mechanism (such as Transformer) is used to dynamically allocate weights to achieve collaborative recommendation of materials and colours. At the same time, the object detection technology can be introduced to segment the main body of the clothing and the background to improve the efficiency and accuracy of colour recognition.

In addition, we can continue to explore the possibility of market applications and design a reinforcement learning framework based on Deep Q-Network (DQN) to build a reward function with user behaviour sequences (click, like, purchase, return), so as to better realise market applications. Moreover, in the subsequent research, the cultural semantics and long-tail requirements of noise data in DBSCAN algorithm can be continued to be mined, and the BERT model can be used to analyse the semantic of clothing description texts (such as “Chinese style” and “Bohemian”), establish the “colour-cultural symbol” mapping relationship, and support the generation of colour scheme driven by cultural narrative.

For brands, the results of market colour analysis can not only inspire new product designs and accessory pairings but also provide deep insights into the evolution of colour trends in the present and near future. These insights can reveal which colours will lead the trends and which may gradually fade from view. Brands can leverage this information to cleverly incorporate popular colour elements into new designs and colour pairings, making their products not only functional but also visually appealing and emotionally resonant, thus increasing their market share. For example, by understanding that spring colour trends lean towards fresh and bright tones, a brand could design a vibrant and energetic spring clothing line or launch matching accessories to attract fashion-conscious and individualistic consumers. They could also integrate their unique brand colours and elements into new designs, thereby better conveying and reinforcing the brand image.

For consumers, observing market colour trends enables a better understanding of the popular colours. By focusing on colour analysis results, consumers can sharply capture how seasonal changes, cultural shifts, and societal moods influence colour choices. This insight allows them to foresee which colours will become the latest trends, enabling them to make more forward-thinking shopping decisions. Whether purchasing clothing, home decor, or personal accessories, consumers can ensure that their choices are both in line with current aesthetic trends and reflective of their individuality. These insights can not only guide them in selecting products that suit them best but also help them maintain a unique style while following trends. This approach fosters a more rational view of trends, continuously enhancing their taste and aesthetic sensibilities.

## ACKNOWLEDGEMENT

This research was supported by Youth Fund for Humanities and Social Sciences Research of the Ministry of Education in 2024 (Project No.24YJCZH238); 2024 "Light of the Taihu Lake Lake" Science and Technology Research (Basic Research) Project (Project No.K20241030); 2023 Undergraduate Teaching Reform Project of Jiangnan University (Project No.JG2302047); 2023 National Social Science Foundation Art Major Project (Project No.23ZD13). The authors acknowledge the above financial support.

## REFERENCES

[1] Jingzhao, Z., *Flood image recognition method based on the deep learning research*, In: Northwest Agriculture and Forestry University of Science and Technology, 2024

[2] Yilan, W., *Research on the spatial colour composition of Dunhuang Grottots,* In: Colour, 2025, 01, 159–161

[3] Mei, X., Li, G., *Chinese snack brand communication based on colour marketing concept discussion,* In: Journal of Cooperation in Economy and Technology, 2023, 20, 62–63

[4] Shengyan, Z., Qiulei, D., *Research on the strategy of integrating colour marketing into brand image design*, In: Grand View, 2023, 05, 39–41

[5] Jie, J., *The application strategy of clothing colour matching in colour marketing,* In: Marketing Strategy, 2018, 29, 10, 61–62

[6] Mei, X., Li, G., *Discussion on Chinese style Dim sum brand communication based on colour marketing concept*, In: Cooperative Economy and Technology, 2023, 20, 10, 62–63

[7] Shuguang, H., Kaiwen, J., Liyan, Z., *Clothing three element recognition based on deep learning*, In: Journal of Fashion, 2022, 7, 5, 399–407

[8] Liang, L., *Design of Computer Intelligent Colour Matching System Based on Machine Learning*, In: Dyeing and Finishing Technology, 2024, 46, 5, 53–55

[9] Yani, Z., *Interactive intelligent fabric colour matching that combines emotions and popular colours*, Master's Thesis, Donghua University, 2023, 1–40

[10] Shujian, Y., Zhizhong, Q., *New media era clothing brand marketing strategy study*, In: Journal of Fashion Designers, 2025, 02, 99–103

[11] Aimei, Z., *Analysis of clothing brand marketing strategy under the background of digital economy,* In: Brand Marketing of Time-famous Brands, 2025, 03,13–15

[12] Chang, L., Jun, L., Xiaoming, L., *Research on Image Region of Interest Extraction Method Based on HSV Colour*, In: Space Software Engineering, 2024, 27, 7, 1–5

[13] Wang, Y., The colour matching design method of cross-border e-commerce website under the guidance of cognitive psychology, In: Psychiatria Danubina, 2021, 33, suppl 6, 107–108

[14] Behura, A., *The cluster analysis and feature selection: Perspective of machine learning and image processing*, In: Data Analytics in Bioinformatics: A Machine Learning Perspective, 2021, 249–280

[15] Qiang, W., Pei, T., Bo, Y., et al., *A Multi-colour space method for vascular enhancement of endoscopic images*, In: Opto-electronic Engineering, 2020, 47, 01, 48–53

[16] Huimin, W., *Research on the Implementation Strategy of Clothing Enterprise Product Network Marketing*, In: Textile Report, 2020, 44, 03, 36–38

[17] Xiaomin, G., *Research on New Mode of Clothing Marketing under the Background of Mobile Internet,* In: Chemical Fiber & Textile Technology, 2020, 54, 03, 94–96

[18] Nagavalli, S.P., Srivatava, A., *AI-Driven Product Recommendations in eCommerce: Enhancing User Engagement and Sales*, In: International Journal of AI, BigData, Computational and Management Studies, 2024, 5, 2, 38–47

**Authors:**

REN XIANGFANG*, FAN RU*, SHEN LEI

School of Digital Technology and Creative Design, Jiangnan University, Wuxi, China
e-mail: 513141421@qq.com, rxf1991@outlook.com

*Ren Xiangfang and Fan Ru contributed equally to this work

**Corresponding author:**

REN XIANGFANG
e-mail: 942372345@qq.com