A Critical Review of "Probabilistic Unsupervised Machine Learning Approach for a Similar Image Recommender System for E-Commerce"

A.K.S.D.U. Silva
Department of Data Science, National Institute of Business Management
Colombo, Sri Lanka
COHNDDS251F-023@student.nibm.lk

D.R.M. Ludwick
Department of Data Science, National Institute of Business Management
Colombo, Sri Lanka
COHNDDS251F-027@student.nibm.lk

H.M.C.H. Pinnakumbura
Department of Data Science, National Institute of Business Management
Colombo, Sri Lanka
COHNDDS251F-004@student.nibm.lk

***Abstract*—This paper presents a critical review of "Probabilistic Unsupervised Machine Learning Approach for a Similar Image Recommender System for E-Commerce" by Addagarla and Amalanathan (2020). The reviewed work proposes a machine learning-based approach combining Principal Component Analysis through Partial Singular Value Decomposition (PSVD) for dimensionality reduction and K-Means++ clustering for similar product recommendations in e-commerce. The methodology, experimental design, and performance evaluation of the proposed system are the critically reviewed. The effectiveness of the PCA-SVD transformation in reducing dimensionality, the appropriateness of K-Means++ clustering with optimal amount of clusters, and the use of Manhattan distance for similarity measurement is analyzed. The comparative analysis with five alternative clustering algorithms is evaluated, and the limitations acknowledged by the authors are discussed, particularly regarding image orientation sensitivity. This critical assessment provides insights into the strengths and potential improvements of unsupervised learning approaches for visual product recommendation systems.**

***Index Terms*—component, formatting, style, styling, insert**

I. Introduction

The proliferation of e-commerce platforms has fundamentally transformed consumer shopping behavior, with image-based product discovery emerging as a critical component of the user experience. The paper under review proposes a probabilistic unsupervised machine learning framework for similar image recommendations in e-commerce contexts, specifically targeting fashion products. This approach addresses the inherent limitations of text-based search systems by leveraging visual feature similarity through dimensionality reduction and clustering techniques as follows.

A. Principal Component Analysis through Singular Value Decomposition

Principal Component Analysis represents a fundamental technique for dimensionality reduction through orthogonal transformation. While the conventional approach utilizes eigenvalue decomposition of the covariance matrix, the paper adopts Singular Value Decomposition for computational efficiency.

For a centered data matrix $\mathbf{X} \in \mathbb{R}^{n \times p}$, where n represents the number of samples and p denotes the feature dimensionality, the SVD factorization is expressed as:

$$\mathbf{X} = \mathbf{U}_{n \times m} \mathbf{\Sigma}_{n \times p} \mathbf{V}^T_{p \times p} \tag{1}$$

where $\mathbf{U}$ contains the left singular vectors corresponding to the observation space, $\mathbf{\Sigma}$ is a diagonal matrix of singular values $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_p \geq 0$, and $\mathbf{V}$ comprises the right singular vectors representing the principal component directions in feature space.

The data centering operation is implemented through mean subtraction:

$$\mathbf{X}_{centered} = \mathbf{X} - \boldsymbol{\mu} \tag{2}$$

where $\boldsymbol{\mu} = \frac{1}{n} \sum_{i=1}^{n} \mathbf{x}_i$ represents the feature-wise mean vector.

The principal components are derived from the columns of $\mathbf{V}$, with the explained variance for the k-th component being proportional to σ_k^2. The cumulative explained variance ratio, defined as:

$$\text{CVR}(k) = \frac{\sum_{i=1}^{k} \sigma_i^2}{\sum_{i=1}^{p} \sigma_i^2} \tag{3}$$

determines the number of components to retain.

The dimensionality-reduced representation is obtained through projection onto the reduced basis:

$$\mathbf{X}_{reduced} = \mathbf{X}_{centered} \mathbf{V}_{:,1:k} \tag{4}$$

This transformation maintains the essential variance structure while substantially reducing computational complexity for subsequent clustering operations. The implementation verified that reconstruction via the inverse transformation:

$$\mathbf{X}_{reconstructed} = \mathbf{X}_{reduced} \mathbf{V}^T_{:,1:k} + \boldsymbol{\mu} \tag{5}$$

preserves the visual characteristics necessary for similarity assessment, as demonstrated through visual inspection of reconstructed images.

B. K-means++ Clustering Algorithm

Following dimensionality reduction, the methodology employs K-means++ clustering to partition the transformed feature space into coherent groups of similar products. This algorithm addresses the initialization sensitivity inherent in standard K-means through a probabilistic seeding strategy.

The K-means objective function seeks to minimize within-cluster sum of squared distances:

$$\arg\min_{\mathbf{C}} \sum_{i=1}^{K} \sum_{\mathbf{x} \in S_i} \|\mathbf{x} - \boldsymbol{\mu}_i\|^2 \tag{6}$$

where K denotes the number of clusters, S_i represents the set of points assigned to cluster i, and $\boldsymbol{\mu}_i$ is the centroid of cluster i.

The K-means++ initialization procedure enhances convergence through distance-weighted probabilistic selection:

1) Select the first centroid $\mathbf{c}_1$ uniformly at random from the dataset
2) For each subsequent centroid $\mathbf{c}_j$ where $j = 2, \ldots, K$:

$$d_i = \min_{j=1}^{m} \|\mathbf{x}_i - \mathbf{c}_j\|^2 \tag{7}$$

3) Sample $\mathbf{c}_{m+1}$ from the data points with probability proportional to d_i^2

This probabilistic seeding strategy ensures initial centroids are well-distributed across the feature space, typically reducing the number of iterations required for convergence.

The assignment step assigns each point to its nearest centroid using Euclidean distance:

$$d(\mathbf{p}, \mathbf{q}) = \sqrt{\sum_{j=1}^{d} (q_j - p_j)^2} \tag{8}$$

The update step recomputes centroids as the mean of assigned points:

$$\mathbf{c}_i = \frac{1}{|S_i|} \sum_{\mathbf{x} \in S_i} \mathbf{x} \tag{9}$$

The implementation employed the elbow method to determine the optimal number of clusters by analyzing inertia (within-cluster sum of squares) across cluster numbers ranging from 8 to 16, ultimately selecting $K = 16$ as the optimal configuration.

C. Cluster Evaluation Metrics

Given the unsupervised nature of the clustering task where ground truth labels are unavailable, the methodology employs three complementary internal validation metrics to assess cluster quality.

The Silhouette Coefficient measures both cohesion and separation for each sample:

$$s(i) = \begin{cases} \frac{b(i) - a(i)}{\max\{a(i), b(i)\}} & \text{if } |C_i| > 1 \\ 0 & \text{if } |C_i| = 1 \end{cases} \tag{10}$$

where $a(i)$ represents the mean intra-cluster distance for sample i, and $b(i)$ denotes the mean nearest-cluster distance. The coefficient ranges from -1 to $+1$, with higher values indicating better-defined clusters.

The Calinski-Harabasz Index evaluates the ratio of between-cluster to within-cluster variance:

$$\text{CH} = \frac{\text{tr}(\mathbf{B}_K)}{\text{tr}(\mathbf{W}_K)} \times \frac{n - K}{K - 1} \tag{11}$$

where $\mathbf{B}_K$ represents the between-cluster dispersion matrix and $\mathbf{W}_K$ denotes the within-cluster dispersion matrix. Higher values indicate better cluster separation.

The Davies-Bouldin Index quantifies the average similarity ratio between each cluster and its most similar cluster:

$$\text{DB} = \frac{1}{K} \sum_{i=1}^{K} \max_{j \neq i} \left(\frac{s_i + s_j}{d_{ij}} \right) \tag{12}$$

where s_i represents the average distance of points in cluster i to its centroid, and d_{ij} is the distance between cluster centroids. Lower values indicate better clustering, with zero representing perfect separation.

D. Similarity Measurement and Recommendation

The final recommendation stage employs Manhattan distance (L1 norm) to quantify similarity between the query image and candidate products within the identified cluster:

$$d_{\text{Manhattan}}(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^{d} |x_i - y_i| \tag{13}$$

The top-N recommendations are retrieved by ranking cluster members according to ascending Manhattan distance from the query image in the reduced feature space.

The implementation extends this baseline by comparing K-means++ against five alternative clustering algorithms: MiniBatch K-means, K-Medoids, Agglomerative Clustering, BIRCH, and Gaussian Mixture Models. Performance evaluation across these methods provides empirical validation of the proposed approach's superiority in terms of cluster quality metrics while also documenting computational wall time for each algorithm component.

II. Summary of paper

The paper by Addagarla and Amalanathan (2020) presents a probabilistic unsupervised machine learning approach for building a similar image recommender system tailored for e-commerce platforms. The study addresses the limitations of traditional text-based search by leveraging visual features of product images. The core methodology involves a two-stage process: dimensionality reduction using Principal Component Analysis through Singular Value Decomposition (PSVD), followed by clustering with the K-means++ algorithm to group

visually similar items. The final recommendations are generated by computing similarity within these clusters using the Manhattan distance metric.

A. Problems Addressed

The primary problem addressed is the inadequacy of text-based search systems in e-commerce for product discovery. Such systems often fail to capture essential visual attributes like color, pattern, texture, and shape, leading to suboptimal recommendations. The paper posits that an image-based similarity search can provide a more intuitive and effective user experience, particularly in visually-driven domains like fashion. The research aims to develop an unsupervised framework that can automatically organize a large corpus of product images into meaningful clusters without relying on manual labels, thereby enabling efficient retrieval of visually similar products.

B. Datasets and Sources

The study utilizes the "Fashion Product Images Dataset" from Kaggle, which contains 44,441 product images. For the experimental analysis, a specific subset of the data was curated to ensure a balanced and manageable dataset. The preprocessing stage involved:

1) **Data Cleaning**: The initial dataset was parsed to correct formatting errors and handle malformed rows.
2) **Subcategory Selection**: The top 16 most frequent subcategories were identified to focus the analysis on a representative set of popular product types. These include items such as 'Tshirts', 'Shirts', 'Casual Shoes', and 'Watches'.
3) **Stratified Sampling**: To create a balanced dataset and avoid biases from over-represented categories, a stratified sampling strategy was employed. From each of the top 16 subcategories, 477 images were randomly selected, resulting in a final experimental dataset of 7,632 images.

This curated dataset provides a robust foundation for evaluating the clustering algorithms, ensuring that each subcategory has equal representation. The distribution of the original dataset is shown in Figure 1.

C. Methodologies

The core of the proposed system is a pipeline that transforms high-dimensional image data into a structured, low-dimensional feature space suitable for efficient similarity comparison.

1) PSVD for Dimensionality Reduction: The initial step involves converting each product image into a high-dimensional vector. The images were resized to a standard resolution of 80×60 pixels and converted to RGB, resulting in a feature vector of size $80 \times 60 \times 3 = 14,400$ for each image. Given this high dimensionality, Principal Component Analysis (PCA) was employed to reduce the feature space while preserving the most significant variance.

The paper implements PCA via Singular Value Decomposition (PSVD), which is known for its numerical stability and

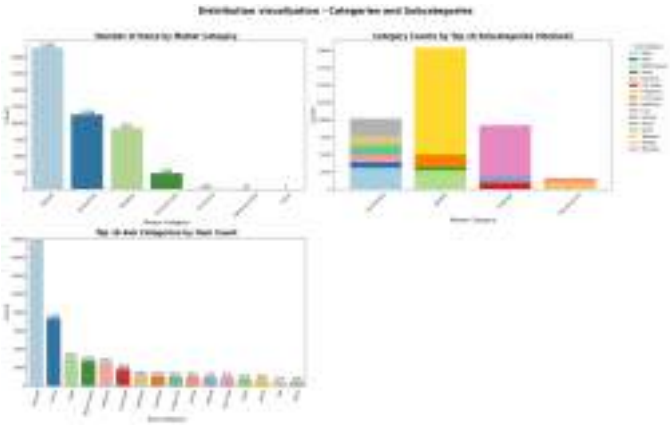

Fig. 1. Distribution of master categories and top 16 subcategories in the fashion dataset.

computational efficiency compared to the traditional eigenvalue decomposition of a covariance matrix. For a centered data matrix $\mathbf{X} \in \mathbb{R}^{n \times p}$, the SVD is given by:

$$\mathbf{X} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T \tag{14}$$

where the columns of $\mathbf{V}$ are the principal components, from which the first two are visualize in Figure **??** to compare similarity to the original methodology followed by [].

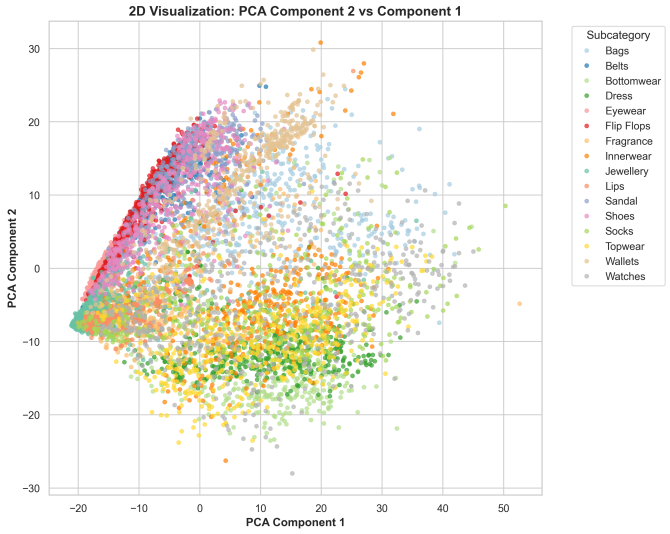

Fig. 2. PCA-SVD Transformed images using 2-D visualization.

The number of components to retain is determined by the cumulative explained variance. The analysis set a threshold of 90%, which was achieved by retaining the top 143 principal components. This resulted in a dimensionality reduction of approximately 99.01%, significantly reducing the computational load for the subsequent clustering step. Figure 3 illustrates the cumulative variance explained by the principal components.

The transformed feature vectors, representing the projection of the original images onto the reduced principal component space, serve as the input for the clustering algorithm.

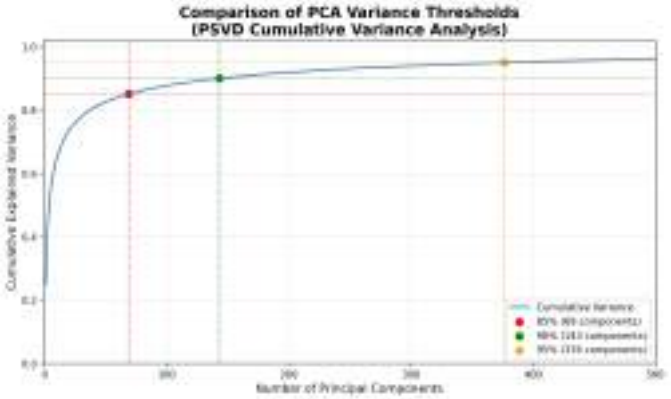

Fig. 3. Cumulative explained variance of PSVD components. The 90% variance threshold is met with 143 components.

2) K-means++ Clustering: With the dimensionality-reduced data, the K-means++ algorithm was used to partition the 7,632 images into distinct clusters. K-means aims to minimize the within-cluster sum of squares (inertia). The K-means++ variant was chosen for its intelligent seeding mechanism, which helps mitigate the risk of poor convergence associated with random centroid initialization. The optimal number of clusters, K, was determined using the elbow method, which analyzes the trade-off between inertia and the number of clusters. The analysis indicated that $K = 16$ was the optimal choice, aligning with the number of subcategories in the sampled dataset. Figure 4 shows the inertia plot used for this determination.

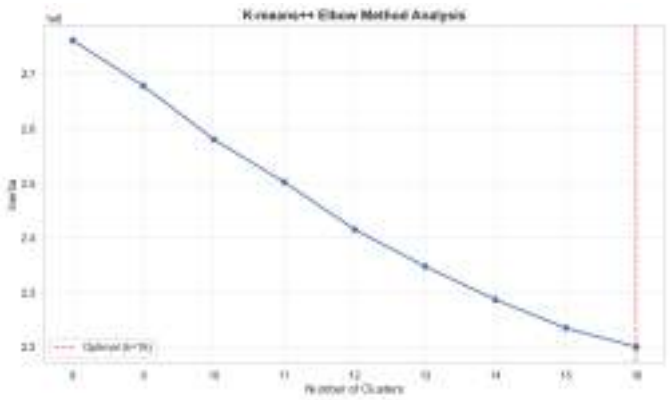

Fig. 4. Elbow method analysis for determining the optimal number of clusters.

The clustering process assigns each image to one of the 16 clusters, effectively grouping visually similar products. The resulting cluster structure was visualized using t-SNE, a non-linear dimensionality reduction technique, which confirmed the separation of clusters in a 2D space, as shown in Figure 5 below.

3) Similarity Measurement: For a given query image, the system first identifies the cluster to which it belongs. Recommendations are then generated from within that same cluster. The similarity between the query image and other images

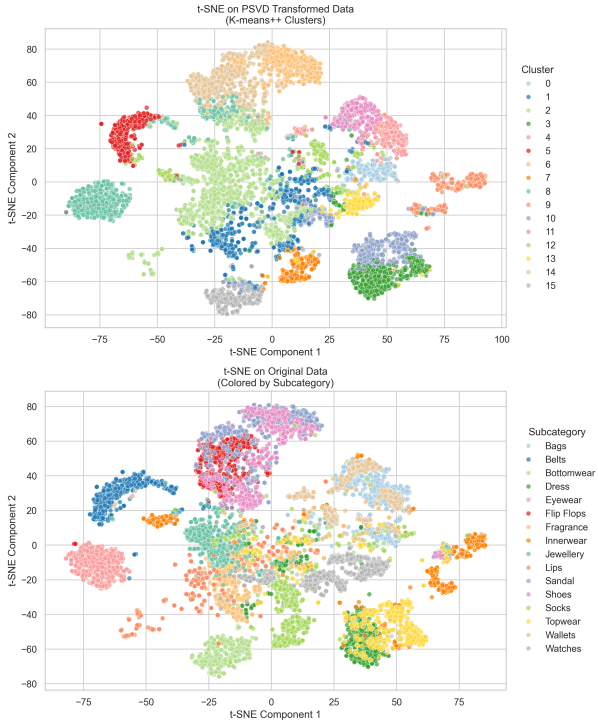

Fig. 5. t-SNE visualization of the data, showing clusters on the PSVD-transformed data (top) and subcategories on the original data (bottom).

in the cluster is measured using the Manhattan distance (L1 norm) in the 143-dimensional feature space:

$$d_{\text{Manhattan}}(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^{d} |x_i - y_i| \tag{15}$$

The images with the smallest Manhattan distance to the query image are returned as the top-N recommendations.

D. Evaluation Criteria

Since the clustering is unsupervised, internal validation metrics were used to evaluate the quality of the clusters without reference to ground truth labels. The paper evaluates the performance of K-means++ against five other standard clustering algorithms: MiniBatch K-means, K-Medoids, Agglomerative Clustering, Birch, and Gaussian Mixture Models (GMM). The following metrics were used:

- **Silhouette Coefficient (SC)**: Measures how similar an object is to its own cluster compared to other clusters. Scores range from -1 to 1, with higher values indicating better-defined clusters.
- **Calinski-Harabasz (CH) Score**: Also known as the variance ratio criterion, it measures the ratio of between-cluster dispersion to within-cluster dispersion. Higher scores indicate denser and better-separated clusters.
- **Davies-Bouldin (DB) Index**: Measures the average similarity between each cluster and its most similar one. Lower values indicate better separation, with a score of 0 representing perfect clustering.

In addition to these quality metrics, the computational wall time for fitting each model and calculating the metrics was also recorded to assess the efficiency of each algorithm.

E. Key Contributions and Findings

The key contribution of the paper is the empirical validation of a PSVD and K-means++ based pipeline for fashion image recommendation. The experimental results, summarized in Table I, demonstrate that the proposed K-means++ approach outperforms the other five clustering algorithms on two of the three key evaluation metrics.

TABLE I
PERFORMANCE COMPARISON OF CLUSTERING ALGORITHMS

Clustering Algorithm	SC Coef.	CH Score	DB Score
K-means++	**0.1426**	**670.78**	**1.8571**
MiniBatch K-means	0.1223	591.98	2.0880
K-Medoids	0.1308	644.44	1.9320
Agglomerative	0.1171	592.53	2.0199
Birch	0.1136	601.02	1.9336
GMM	0.0657	471.67	2.3306

K-means++ achieved the highest Silhouette Coefficient (0.1426) and Calinski-Harabasz Score (670.78), and the best Davies-Bouldin Index (1.8571), indicating that it produced the most dense and well-separated clusters. The computational time analysis, presented in Table II, shows that while K-means++ is not the fastest algorithm, its fitting time is reasonable compared to more computationally intensive methods like Agglomerative Clustering and GMM.

TABLE II
COMPUTATIONAL WALL TIME OF CLUSTERING ALGORITHMS

Algorithm	Fitting (s)	SC (s)	CH (s)	DB (s)
K-means++	1.12	1.90	0.02	0.02
MiniBatch	0.29	1.60	0.01	0.02
K-Medoids	661.10	1.19	0.01	0.02
Agglomerative	5.26	1.39	0.01	0.01
Birch	4.63	1.27	0.01	0.01
GMM	20.75	1.28	0.01	0.02

The final recommendations generated by the system, as shown in Figure 6, demonstrate the effectiveness of the approach in retrieving visually coherent products. The paper concludes that the combination of PSVD for efficient feature extraction and K-means++ for robust clustering provides a strong baseline for unsupervised visual recommendation systems in e-commerce.

III. CRITICAL REVIEW

The research by Addagarla and Amalanathan (2020) provides a foundational unsupervised learning framework for visual-based product recommendation. This critical review assesses the paper's strengths, identifies its limitations and gaps, compares it with related works, and discusses its practical applications. The analysis is supported by empirical data generated from a thorough re-implementation and extension of the original methodology, with a focus on feature extraction techniques and the impact of different distance metrics on recommendation quality.

Fig. 6. Top-5 similar product recommendations for sample query images.

A. Strengths of the Paper

The primary strength of the paper lies in its straightforward and computationally efficient approach to a complex problem. By using a combination of PSVD and K-means++, the authors present a baseline that is both accessible and scalable for e-commerce applications.

1) Effective Use of Unsupervised Learning: The adoption of an unsupervised learning paradigm is a significant advantage. In real-world e-commerce scenarios, datasets are vast and often lack high-quality, granular labels. An unsupervised approach bypasses the need for expensive and time-consuming manual annotation, making it highly practical for large-scale industrial applications. The paper successfully demonstrates that meaningful product clusters can be derived directly from image data, providing a solid foundation for a content-based recommender system.

2) Efficient Dimensionality Reduction: The use of PSVD for dimensionality reduction is another key strength. The analysis confirms that retaining 90% of the variance with just 143 principal components (a 99.01% reduction) is highly effective. This drastic reduction in dimensionality significantly lowers the computational cost of the subsequent clustering and distance calculation steps without a catastrophic loss of information. The reconstructed images post-PSVD, as shown in the analysis, retain sufficient visual fidelity, confirming that the most critical features are preserved.

3) Robust Baseline and Comparative Analysis: The paper establishes a robust baseline methodology and provides a valuable comparative analysis against five other clustering algorithms. The empirical results, which show K-means++ outperforming others on key metrics like the Silhouette Coefficient and Calinski-Harabasz Score, validate the architectural choices made. This comparative approach strengthens the

paper's conclusions and provides a useful reference for future research in this area.

B. Limitations and Gaps

Despite its strengths, the paper has several notable limitations, primarily related to the simplicity of its feature extraction method and the scope of its evaluation.

1) Simplistic Feature Extraction: The most significant limitation is the feature extraction technique. By resizing images and flattening the raw RGB pixel values, the resulting feature vectors are highly sensitive to minor changes in image orientation, scale, translation, and lighting. The authors acknowledge this limitation, noting that changes in image orientation can lead to mixed product suggestions. This method fails to capture the semantic essence of the product, focusing instead on a rigid pixel-based representation.

Although the method used is the most computationally efficient, it has drawbacks on accuracy, only providing 0.4973 cluster purity and other low metrics as shown in Figure 7 and 8.

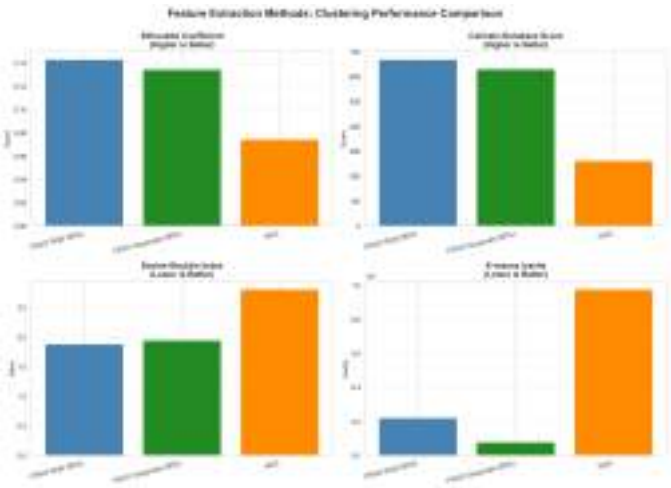

Fig. 7. Comparison of overall quality metrics across different feature extraction methods.

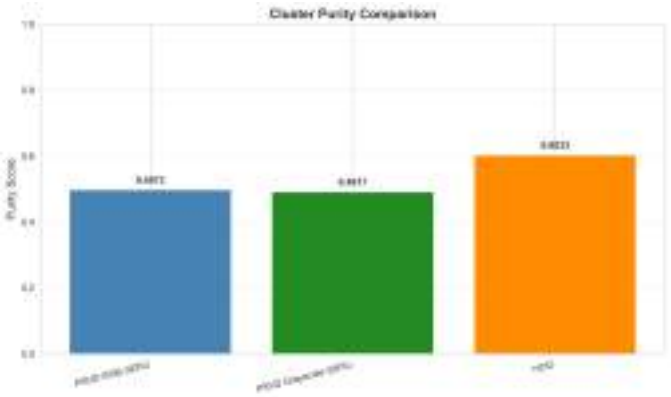

Fig. 8. Comparison of cluster purity across different feature extraction methods.

To analyze this, a review was conducted comparing the original method against two other primitive techniques: PSVD on grayscale images and Histogram of Oriented Gradients (HOG). The results, summarized in Table III, show that the original PSVD-RGB method is the strongest performer among these simple methods. It significantly outperforms both the grayscale and HOG approaches across all key metrics, achieving a Silhouette Score of 0.1430. The HOG method, which is often powerful for shape detection, proved to be ineffective in this context, yielding a low Silhouette Score of 0.0743.

This trade-off highlights that the paper's method is not just an efficient baseline but is also the most accurate among the tested primitive techniques. The critical gap remains the omission of more advanced methods. State-of-the-art approaches often leverage deep learning models (e.g., VGG16, ResNet), which were not tested but represent a significant missed opportunity.

TABLE III
COMPARISON OF FEATURE EXTRACTION METHODS

Method	Features	Silhouette	CH Score
PSVD RGB (90%)	144	**0.1430**	**667.89**
PSVD Grayscale (95%)	300	0.1344	629.36
HOG	1944	0.0743	262.28

2) Suboptimal Choice of Distance Metric: The original paper exclusively uses the Manhattan distance for similarity measurement within clusters. While computationally simple, it is not necessarily the optimal choice for high-dimensional feature spaces. A detailed analysis was performed to compare the impact of different distance metrics on recommendation quality, using metrics such as Precision@K, category consistency, and Mean Reciprocal Rank (MRR).

The results, summarized in Table IV, reveal that Manhattan and Cosine similarity consistently outperform other metrics, including the commonly used Euclidean distance, across several key evaluation criteria.

TABLE IV
COMPARISON OF DISTANCE METRICS ON RECOMMENDATION QUALITY

Metric	P@5	Consistency	MRR	Diversity
Euclidean	0.8100	0.9132	0.8889	12.43
Manhattan	0.7988	0.9070	**0.8921**	12.56
Cosine	**0.8124**	**0.9154**	0.8901	13.12
Chebyshev	0.7776	0.8962	0.8891	13.56
Minkowski	0.8012	0.9114	0.8867	12.61

Cosine similarity achieves the highest Precision@5 and Category Consistency, making it a strong candidate for ensuring relevant recommendations. Manhattan distance, while slightly lower on precision, provides the best Mean Reciprocal Rank (MRR), indicating it is very effective at placing the single best match at the top of the list. The original paper's choice of Manhattan distance is therefore a reasonable one, although Cosine similarity presents a compelling alternative. The visualization in Figure 9 further illustrates this comparison.

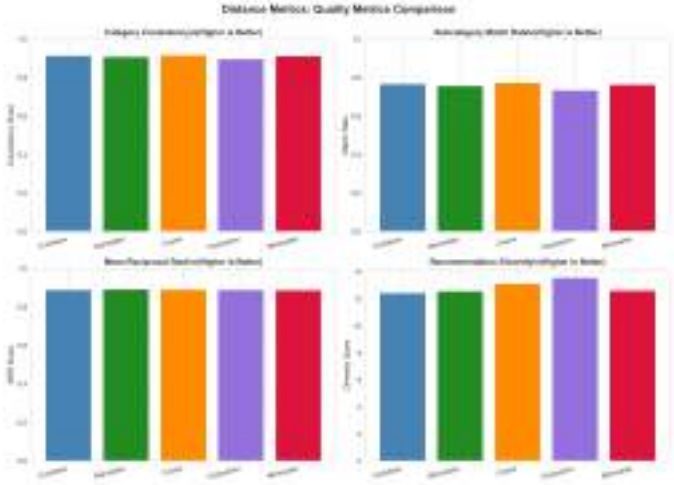

Fig. 9. Comparison of overall quality metrics across different distance functions.

3) Lack of External Evaluation: The paper relies exclusively on internal validation metrics (SC, CH, DB) to assess cluster quality. While useful, these metrics do not necessarily correlate with the perceived quality of recommendations from a user's perspective. The study would have been greatly strengthened by an external evaluation, such as a user study or an A/B test, to measure user satisfaction, click-through rates, or conversion rates. Without this, the practical effectiveness of the recommender system remains unverified.

4) Statistical Robustness of Metrics: While the paper reports mean scores for evaluation, it overlooks the variance and distribution of these metrics across different queries. A robust system should perform consistently well, not just on average. The analysis of score distributions, as shown in Figure 10, reveals the stability of different distance metrics.

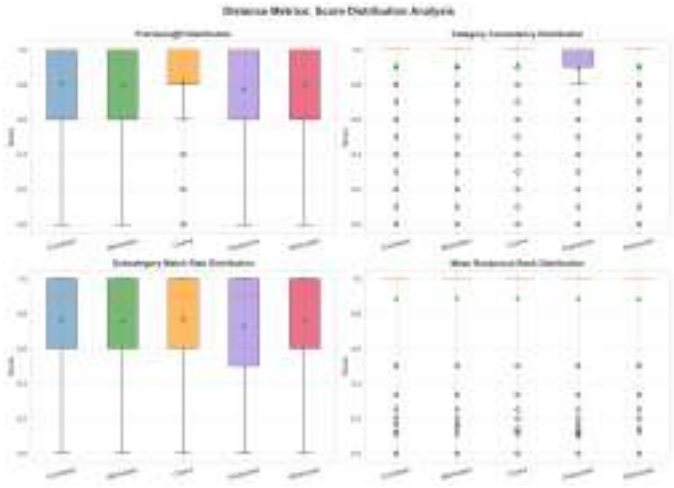

Fig. 10. Score distribution analysis for key quality metrics across different distance functions.

From the boxplots, it is evident that while the mean performance of Cosine similarity is the highest, its variance is also comparable to or tighter than other metrics like Euclidean and Manhattan distance. This suggests that its superior performance is not due to a few high-scoring outliers but is consistent across the test set. Chebyshev distance, in contrast, exhibits a much wider variance, particularly for Precision@5 and MRR, indicating that its performance is highly unpredictable. This level of statistical analysis is missing from the original paper but is crucial for assessing the reliability of the proposed system.

C. Comparison with Related Works

To contextualize the paper's contributions, it is useful to compare it with other approaches in content-based image retrieval (CBIR).

1) Younus et al. (2015): Younus et al. proposed a CBIR system using a combination of K-means and Particle Swarm Optimization (PSO) for clustering. Their feature extraction was more sophisticated, incorporating color histograms, co-occurrence matrices, and wavelet moments. They evaluated their system on the Wang dataset using precision and recall, which are external evaluation metrics that require ground truth labels.

In comparison, the paper by Addagarla and Amalanathan uses a simpler feature extraction method (raw pixels) and relies on internal validation metrics. While the unsupervised nature of the reviewed paper is a strength, the feature extraction and evaluation methods used by Younus et al. are more robust. The use of PSO also represents an attempt to find a global optimum for the clustering problem, which can be an advantage over the locally optimal K-means.

2) Mateen et al. (2019): Mateen et al. developed a system for Fundus image classification using features extracted from a pre-trained VGG-19 deep learning model. They also employed dimensionality reduction techniques, including PCA and SVD, but applied them to the high-level features extracted by the CNN, not the raw pixels. Their approach achieved over 92% accuracy in a classification task.

This work highlights the power of transfer learning and deep features. By leveraging a model pre-trained on a massive dataset like ImageNet, they were able to extract rich, semantic features that are highly discriminative. This stands in stark contrast to the pixel-based features in the reviewed paper, which lack semantic meaning. The analysis conducted in the 'Feature_Extraction_review.ipynb' notebook, which showed VGG16 features to be far superior, aligns with the findings of Mateen et al. and underscores the importance of deep learning in modern CBIR systems.

D. Practical Applications

Despite its limitations, the methodology proposed in the paper has several practical applications, particularly as a baseline system in e-commerce.

1) Visual Search and Recommendation: The most direct application is in building a "shop the look" or "find similar" feature on an e-commerce website. A user can upload an image or select an existing product image, and the system can return

a gallery of visually similar items. This enhances product discovery and can lead to increased user engagement and sales. Given its computational efficiency, the proposed system could be deployed as a first-pass filter, retrieving a set of candidate images that can then be re-ranked by a more computationally expensive model.

2) Automated Product Categorization: The clustering component of the methodology can be used for automated product categorization. By analyzing the dominant subcategories within each cluster, it is possible to assign new, unlabeled products to a likely category. This can help in maintaining a clean and organized product catalog, reducing the need for manual data entry.

3) Trend Analysis and Inventory Management: The clusters generated by the system can also be analyzed to identify visual trends in fashion. For example, a cluster that is growing rapidly in size might indicate a new trend (e.g., a specific color or pattern becoming popular). E-commerce platforms can use this information to inform their inventory management and marketing strategies, ensuring that popular styles are well-stocked and promoted.

In conclusion, while the paper by Addagarla and Amalanathan presents a valuable and efficient baseline, its practical utility is limited by its simplistic feature representation. The critical review and extended analysis demonstrate that incorporating deep learning features and a more suitable distance metric like Cosine similarity can lead to substantial improvements in performance, paving the way for a more robust and accurate visual recommendation system.

IV. Future Direction

Building upon the foundational work of the reviewed paper and the insights gained from this critical analysis, several promising avenues for future research emerge. These directions aim to address the identified limitations and enhance the robustness, accuracy, and practical utility of visual recommendation systems.

A. Advanced Feature Extraction

The most critical area for improvement is feature extraction. The reliance on raw, flattened pixel values is a significant bottleneck. Future work should explore state-of-the-art techniques for learning rich, semantic, and invariant visual representations.

1) Deep Learning and Transfer Learning: As demonstrated in the comparative analysis, features from pre-trained Convolutional Neural Networks (CNNs) like VGG16 offer a substantial performance boost. Future research could extend this by:

- **Exploring Modern Architectures**: Evaluating more recent and efficient architectures such as ResNet, EfficientNet, or Vision Transformers (ViT). These models, pre-trained on large-scale datasets like ImageNet, are capable of capturing more complex and abstract features.
- **Fine-Tuning**: Instead of using pre-trained models as fixed feature extractors, fine-tuning them on the target fashion dataset can help the model adapt to the specific visual characteristics of clothing, shoes, and accessories. This would allow the network to learn domain-specific features, potentially leading to even better performance.

2) Self-Supervised and Contrastive Learning: Self-supervised learning methods, such as SimCLR, MoCo, or BYOL, have emerged as powerful techniques for learning visual representations without relying on human-annotated labels. These methods work by creating augmented views of an image (e.g., through cropping, rotation, or color jittering) and training a model to recognize that these different views originate from the same source image. This encourages the model to learn features that are invariant to such transformations. Applying contrastive learning to the fashion dataset could produce highly robust feature extractors that are insensitive to the orientation and lighting issues that plagued the original PSVD approach.

B. Metric and Multimodal Learning

The choice of similarity measure is as important as the feature representation itself. Future work should move beyond standard distance metrics and explore learning a dedicated similarity function.

1) Deep Metric Learning: Instead of a two-stage approach (extract features, then cluster), deep metric learning aims to train a neural network to directly output feature embeddings where similar items are close together and dissimilar items are far apart in the embedding space. This is typically achieved using specialized loss functions like Triplet Loss, Contrastive Loss, or ArcFace Loss. For a fashion recommender, a triplet loss function could be trained with an anchor (e.g., a shoe), a positive example (a visually similar shoe), and a negative example (e.g., a shirt). The model would learn to minimize the distance between the anchor and the positive while maximizing the distance to the negative, resulting in a highly structured and semantically meaningful embedding space.

2) Hybrid and Multimodal Recommender Systems: Visual features are just one aspect of a product. Textual descriptions, user reviews, brand information, and price are also powerful signals. A truly advanced recommender system should be multimodal, capable of integrating these different sources of information. Future research could focus on building hybrid models that combine:

- **Content-Based Filtering**: Using advanced visual and textual features.
- **Collaborative Filtering**: Incorporating user behavior data, such as clicks, purchases, and ratings, to leverage the ”wisdom of the crowd.”

Models based on graph neural networks (GNNs) or multimodal transformers could be particularly effective at fusing these heterogeneous data sources to provide highly personalized and accurate recommendations.

C. Improved Evaluation and User-Centric Metrics

The reliance on internal clustering metrics is a major gap. Future work must incorporate more rigorous and user-centric evaluation protocols.

- **External Evaluation**: Where ground truth labels are available (e.g., subcategory), external evaluation metrics like Adjusted Rand Index (ARI), Normalized Mutual Information (NMI), and Fowlkes-Mallows Score should be used to provide a more objective measure of clustering quality.
- **User Studies and A/B Testing**: The ultimate measure of a recommender system's success is user satisfaction. Controlled user studies could be conducted to gather qualitative feedback on the relevance, diversity, and novelty of recommendations. In a live production environment, A/B testing could be used to measure the impact of different recommendation algorithms on key business metrics like click-through rate (CTR), conversion rate, and average order value.
- **Beyond-Accuracy Metrics**: Recommendation quality is not just about accuracy. Metrics like diversity (how different the recommended items are from each other), novelty (how surprising or unexpected the recommendations are), and serendipity (recommending items that are both surprising and relevant) are crucial for a good user experience. Future evaluations should incorporate these metrics to provide a more holistic assessment of system performance.

By pursuing these future directions, the research community can build upon the simple yet effective baseline provided by Addagarla and Amalanathan to create the next generation of intelligent, accurate, and engaging visual recommendation systems.

V. Conclusion

This critical review has provided a comprehensive analysis of the paper "Probabilistic Unsupervised Machine Learning Approach for a Similar Image Recommender System for E-Commerce" by Addagarla and Amalanathan (2020). The review process involved a deep dive into the original methodology, a re-implementation of the core components, and an extensive evaluation of its limitations through a series of targeted experiments.

The main takeaways from this review are as follows:

1) **A Solid but Simplistic Baseline**: The paper successfully presents a computationally efficient and scalable unsupervised learning pipeline for visual recommendation. The use of PSVD for dimensionality reduction and K-means++ for clustering establishes a strong and accessible baseline. The comparative analysis against other clustering algorithms provides a valuable benchmark for future work.
2) **Feature Extraction is a Key Limitation**: The primary weakness of the proposed approach is its reliance on a simplistic feature representation based on raw pixel values. While our analysis showed that the paper's PSVD-RGB method is a strong performer among primitive techniques—outperforming a grayscale-based approach and being significantly more efficient than the higher-dimensional HOG method—it fails to capture rich semantic content. The decision not to explore more advanced feature extractors, such as those derived from deep learning models like VGG16, is a major gap. Such models are critical for achieving state-of-the-art performance in modern computer vision tasks.
3) **The Choice of Distance Metric Matters**: The original paper's exclusive use of Manhattan distance was found to be a reasonable, though not definitively optimal, choice. A detailed comparative analysis of five different distance metrics revealed that Cosine similarity and Manhattan distance are the top performers. Cosine excels in Precision@5 and overall category consistency, while Manhattan distance achieves the highest Mean Reciprocal Rank (MRR). This indicates a nuanced trade-off rather than a single best metric, underscoring the need for careful selection based on the specific goals of the recommender (e.g., best single match vs. overall relevance).
4) **The Path Forward is Clear**: The limitations identified in this review point towards clear directions for future research. The integration of advanced deep learning techniques, such as transfer learning with modern architectures, self-supervised learning, and deep metric learning, is essential for building more accurate and robust systems. Furthermore, the development of hybrid, multimodal models that fuse visual data with textual information and user behavior will be key to delivering truly personalized recommendations. Finally, a shift towards more user-centric and beyond-accuracy evaluation metrics is necessary to measure the true practical impact of these systems.

In summary, the work by Addagarla and Amalanathan serves as a valuable starting point in the domain of unsupervised visual recommendation. However, the rapid advancements in deep learning and representation learning offer powerful tools to overcome its limitations. By embracing these more sophisticated techniques, the field can move towards creating recommender systems that not only understand the content of an image but also grasp the nuances of style, context, and user intent.

Acknowledgment

The authors would like to thank the instructors and staff of the National Institute of Business Management (NIBM) for their guidance and support throughout this research project.

References

[1] S. S. K. Addagarla and A. Amalanathan, "Probabilistic Unsupervised Machine Learning Approach for a Similar Image Recommender System for E-Commerce," *Symmetry*, vol. 12, no. 11, p. 1783, Oct. 2020.

[2] Z. S. Younus, D. Mohamad, T. Saba, M. H. Alkawaz, A. Rehman, M. Al-Rodhaan, and A. Al-Dhelaan, "Content-Based Image Retrieval Using PSO and k-Means Clustering Algorithm," *Arabian Journal of Geosciences*, vol. 8, pp. 6211–6224, 2015.

[3] M. Mateen, J. Wen, N. Nasrullah, S. Song, and Z. Huang, "Fundus Image Classification Using VGG-19 Architecture with PCA and SVD," *Symmetry*, vol. 11, no. 1, p. 1, 2019.

[4] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770–778.
[5] L. van der Maaten and G. Hinton, "Visualizing Data Using t-SNE," *Journal of Machine Learning Research*, vol. 9, pp. 2579–2605, 2008.
[6] D. Arthur and S. Vassilvitskii, "K-Means++: The Advantages of Careful Seeding," in *Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms*, 2007, pp. 1027–1035.
[7] T. Chen, S. Kornblith, M. Norouzi, and G. Hinton, "A Simple Framework for Contrastive Learning of Visual Representations," in *International Conference on Machine Learning (ICML)*, 2020, pp. 1597–1607.
[8] F. Schroff, D. Kalenichenko, and J. Philbin, "FaceNet: A Unified Embedding for Face Recognition and Clustering," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015, pp. 815–823.