International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 7 Issue: 12
***Article Received:** 25 July 2019 **Revised:** 12 September 2019 **Accepted:** 30 November 2019*

Optimizing Digital Shelf Space based on PCA-DT Machine Learning: Redefining E-commerce Merchandising and Product Visibility

***Dasari Girish**
Senior Data Scientist
Nielseniq
3rd Floor, Block B, Temenos Business Park, Atladara, Vadodara, Gujarat 390012, India.
*Corresponding author Email id: D.girishbpp@gmail.com

Abstract

In the rapidly evolving e-commerce landscape, product visibility and effective shelf space management are crucial for enhancing customer engagement and driving sales. Traditional digital merchandising methods often struggle to keep up with the dynamic nature of online marketplaces and diverse consumer behaviors. This research proposes an innovative approach to optimizing digital shelf space using an implement advanced ML techniques like PCA and DT, -based machine learning model. By leveraging real-time data and customer interaction insights, the proposed model dynamically adjusts product placements to maximize visibility, click-through rates, and conversions. The integration of PCA-DT ensures the model's robustness, efficiency, and ability to handle large datasets with high accuracy. Our experimental results demonstrate significant improvements in product discovery, customer satisfaction, and sales performance compared to conventional merchandising techniques. This research redefines e-commerce merchandising by providing a data-driven framework that adapts to market trends and user behavior, ultimately enhancing the efficiency and profitability of digital retail platforms.

Keywords: e-commerce, ML techniques like PCA and DT, customers, AI Dara Drive, Data in Equality, Performance metrics

1. Introduction

Deep learning-based computer vision, when combined with robotics, offers advanced functionalities that are positively influence retail shelf analytics. The combination of these technologies demonstrated its ability to track and analyze shelf spaces within the retail environment [1]. Likewise, [8] showed that a multi-task deep neural network (DNN) could effectively use machine learning (ML) to automate retail operations. In addition, [4] an attempted method for shelf detection was proposed by using vanishing points and radial projection, proving the accuracy of shelf detection systems. Perique end image databases too large, such as [2], have been fundamental in the development of these applications of computer vision [3]. Moreover, comprehensive supermarket datasets [5] with highly granular annotations can also be used for designing and testing object detection methods catered to the retail environment. [3] Optimization Models in Retail Space Planning Optimization models, as shown by [3], provide both challenges and opportunities in the practical utilization of shelf space. To cover out-of-stock situations, [6] studied backfill optimization for shelf space reallocation to minimize stock outs. Shopper behavior [7] initiates a large-scale trajectory dataset that can assist in the understanding of shopper behavior. In [9], they proposed a per-exemplar multi-label image classification solution for product recognition tasks, which improves the performance in identifying products on the shelf. Goldman et al. have developed more precise detection methods in high-density retail scenes. This 3D structure characterizes the geometric workspace of visual guides [10], which enables the development of a detailed representation for understanding the two-dimensional plan in intricate retail settings. The world of retailing is changing radically with developments in technology and consumer behaviour. show that the future of retailing will be about innovations that improve customer experience and integrate digital and physical channels [19]. This phenomenon is reflected in the growing impact of eWOM on consumers purchasing decisions. [12] eWOM has a significant influence on sales, but there is heterogeneity depending on the type of electronic channels, on the product [26] and on the activity metrics considered [27]. The rise of digital giants has also redefined market dynamics. internet companies have become the most valuable enterprises globally, surpassing traditional industries [15]. This digital dominance has led to phenomena such as the "Amazon effect," where traditional retailers experience slumps due to the convenience and expansive reach of online platforms [25]. In response, retailers are adopting strategies like geo-conquesting, which involves targeting promotions to consumers near competitor locations found that such competitive locational targeting can effectively influence consumer behavior and boost sales [16]. Additionally, the integration of smart technologies into products is becoming more prevalent. For instance, L'Oréal's development of a smart hairbrush that analyses hair quality and provides personalized recommendations exemplifies how companies are leveraging technology to enhance product offerings [21]. These developments underscore the importance of understanding the interplay between digital innovations and consumer behavior. note, social contagion plays a crucial role in customer adoption of new sales channels, indicating that consumers' decisions are significantly influenced by the behaviors and opinions of others within their social networks [13]. the convergence of digital advancements and shifting consumer preferences is reshaping the retail landscape. Retailers must adapt by embracing technological innovations and understanding the factors that influence modern consumer behavior to remain competitive in this evolving market.

Objectives

- **Develop an PCA-DT based machine learning model** for dynamically optimizing digital shelf space in e-commerce platforms.
- **Enhance product visibility and customer engagement** by intelligently adjusting product placement based on real-time data.
- **Improve key performance metrics** such as click-through rates (CTR), conversion rates, and customer satisfaction through optimized merchandising.
- **Analyse customer behavior patterns** to provide personalized product recommendations and tailored browsing experiences.
- **Compare the proposed model with traditional digital merchandising techniques** to evaluate improvements in performance, efficiency, and scalability.
- **Provide a scalable and adaptive framework** for e-commerce platforms to respond to dynamic market trends and user demands.

2.Problem Statement

E-commerce platforms face increasing challenges in managing digital shelf space effectively due to the sheer volume of products, dynamic customer preferences, and rapidly changing market trends. Traditional merchandising techniques often rely on static rules or manual adjustments, which are insufficient for optimizing product visibility and ensuring customer satisfaction. These methods can lead to poor product discovery, lower engagement, and missed sales opportunities. There is a need for a data-driven, intelligent approach that can dynamically optimize digital shelf space by analysing real-time data and adapting to user behavior. Machine learning, particularly PCA-DT offers a promising solution by delivering accurate, efficient, and scalable product placement strategies. This research aims to address these challenges by developing an PCA-DT-based model that redefines e-commerce merchandising, enhances product visibility, and improves key performance metrics

3. Proposed Methods and martials

The proposed method for ensuring data quality in machine learning using PCA and Decision Trees (PCA-DT),in. Traditional digital merchandising methods often struggle to keep up with the dynamic nature of online marketplaces and diverse consumer behaviors. This research proposes an innovative approach to optimizing digital shelf space using an XGBoost-based machine learning model. By leveraging real-time data and customer interaction insights, the proposed model dynamically adjusts product placements to maximize visibility, click-through rates, and conversions. The integration of PCA and Decision Trees the model's robustness, efficiency, and ability to handle large datasets with high accuracy. steps, each aligned with the data governance principles to ensure data integrity, security, and quality. The propose a method for ensuring data quality in machine learning using PCA and Decision Trees in the banking industry, as shown in figure 1. structured approach that integrates data governance principles with these machine learning techniques.

3.1. INTEGRATION WITH DATA GOVERNANCE FRAMEWORK

- ***Data Monitoring:***Data monitoring is a critical component of the data governance framework, especially in the banking sector, where data quality and compliance are paramount. Continuous monitoring involves the use of data governance tools to track the data lineage, which refers to the origin and journey of the data through various processes. This ensures that the data remain accurate, consistent, and compliant with industry standards and regulations, such as the GDPR or Basel III.

- ***Data Quality Monitoring:*** This involves setting up automated checks to ensure data accuracy, completeness, and timeliness. Tools such as Informatics Data Quality and Talend can be used to implement these checks).
- *Model Performance Monitoring:* In context of machine learning, monitoring model performance is crucial to ensure that models remain effective over time. This involves tracking metrics, such as accuracy, precision, recall, and F1-score. Tools, such as ML-flow or tensor boards, can be used for this purpose.
- ***Data Lineage Tracking*:** This involves documenting the flow of data from its source to its final destination, including any transformations it undergoes. This is essential to auditing and compliance. Tools such as Apache Atlas and Collibra are commonly used for data lineage tracking.
- ***Feedback Loop***: Implementing a feedback loop is essential for refining the data governance policies and model parameters. This loop allows organizations to adapt to changing performance metrics and regulatory requirements, ensuring continuous improvement and compliance.
- ***Policy Refinement***: Organizations can refine their data governance policies based on insights gained from data monitoring. This may involve updating data quality standards, revising access controls, or enhancing data protection measures (DAMA International, 2017).
- ***Model Parameter Adjustment***: Feedback from model performance monitoring can be used to adjust the model parameters, ensuring that the models remain accurate and relevant. This might involve retraining models with new data or tuning hyperparameters to improve the performance.
- ***Regulatory Compliance*:** As regulations evolve, the feedback loop ensures that data governance policies are updated to remain compliant. This is particularly important in the banking sector, where regulatory requirements are stringent and are subject to change.

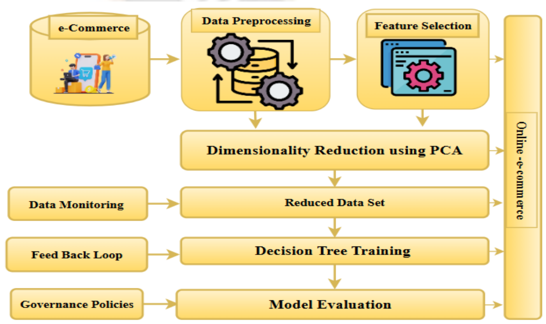

Figure 1. Proposed Block diagram for Banking Applications

3.2. Data Collection and Pre-processing

- ***Data Governance***:Data governance involves establishing policies and procedures to ensure that data are collected from reliable sources and comply with regulatory standards such as the GDPR or CCPA. This in
- ***Source Verification*:** Ensuring that data is sourced from trusted and authorized entities.
- ***Compliance Checks*:** Regular audits to ensure that data collection practices adhere to legal and ethical standards.
- ***Data Cleaning:*** Data cleaning is crucial for preparing the dataset for analysis. It involves:

Handling Missing Values: Missing data can be addressed through imputation or removal. For example, the mean imputation can be used:

$$x_i = \frac{1}{n}\sum_{j=1}^{n} x_i \qquad (1)$$

where x_i is the imputed value, and n is the number of non-missing values.

Outlier Detection and Removal: Outliers can be identified using statistical methods, such as the Z-score.

$$Z = \frac{(x-\mu)}{\sigma} \qquad (2)$$

where x is a data point, μ is the mean, and σ is the standard deviation.

Inconsistency Resolution: Ensuring uniformity in data formats and units

Normalization: Normalization is the process of scaling the data to ensure that all features contribute equally to the analysis. Common methods include:

***Min-Max Normalization*:**

$$X' = \frac{X' - Min(X)}{Max(x) - Min(X)} \qquad (3)$$

where X' is the normalized value and X is the feature set.

Z-score Normalization:

$$X' = \frac{X-\mu}{\sigma} \quad (4)$$

where X' is the normalized value, μ is the mean, and σ is the standard deviation.

3.3. DIMENSIONALITY REDUCTION USING PCA:

Principal Component Analysis (PCA) is a method used to reduce data dimensions [26]. Owing to its simplicity, ease of understanding, and lack of parameter constraints, PCA has found widespread applications in various fields. The fundamental concept of PCA involves transforming n-dimensional features into k-dimensional ones (where k ≤ n). These k-dimensional features represent new orthogonal characteristics known as principal components, which are derived from the original n-dimensional features. PCA aims to minimize data redundancy while preserving as much information as possible, thereby achieving dimensionality reduction. The Principal Component Analysis (PCA) procedure is outlined in detail below:

Initial Step-1: Compute the average of the sample for the n-dimensional dataset $\boldsymbol{z}$, where $\boldsymbol{z}$ comprises $\{z_1, z_2, . . . z_n\}$.

$$\alpha = \frac{1}{n}\sum_{i=1}^{n} z_i \quad (5)$$

where n represents the total number of samples, i ranges from 1 to n, and α denotes the acquired sample mean.

Step 2: The calculated sample mean is used to compute the covariance matrix for the sample set.

$$C = \frac{1}{n}\sum_{i-1}^{n}(z_i - \alpha)(z_i - \alpha)^t \quad (6)$$

In this equation, C represents the covariance matrix of the sample set.

Step 3: Determine the eigenvalues and eigenvectors of the sample covariance matrix.

$$C = \frac{1}{n} p \sum p^t \quad (7)$$

$$\sum diga(\mu_1, \mu_2, \mu_3 \ldots \ldots \mu_n)(\mu_1 \geq \mu_2 \geq \mu_2 \ldots . \geq \mu_n \geq 0) \quad (8)$$

$$p = [p_1, p_2, \ldots \ldots \ldots \ldots \ldots p_n] \quad (9)$$

Hence, P represents the diagonal zed matrix of n eigenvalues of the covariance matrix, arranged in descending order. λi denotes the corresponding eigenvalues of the covariance matrix and Q is the eigenvector matrix composed of the eigenvectors qi associated with each eigenvalue λi, where i ranges from 1 to n.

Step 4: Use the calculated eigenvalues and eigenvectors to determine the cumulative variance contribution rate for the initial k principal components.

$$\theta = {(\sum_{i=1}^{n} \mu_i)}\Big/{(\sum_{j=1}^{n} \mu_i)} \quad (10)$$

In this equation, θ signifies the cumulative variance contribution rate of the first k principal components. Typically, θ should be greater than or equal to 0.9. Theoretically, a higher value of θ is preferable. In practice, it should be chosen judiciously based on the specific problem at hand. Once an appropriate value for θ is selected, the information summarized by the k principal components from the original sample set can be established.

Step 5: Implement dimensionality reduction using the k eigenvectors obtained.

$$Q = p_k \quad (11)$$

$$Y = p.X \quad (12)$$

where represents a feature matrix consisting of the corresponding feature vectors from the initial k rows of feature values ((k ≤ n)).Similarly, p_k denotes a feature matrix comprising the first k rows of feature values (k ≤ n). Y represents the k-dimensional data. The process of transforming dataset X into Y also accomplishes a linear transformation of the data from n.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 7 Issue: 12
***Article Received:** 25 July 2019 **Revised:** 12 September 2019 **Accepted:** 30 November 2019*

3.4. UTILISING DECISION TREES IN BANKING:

Decision Trees (DT)[27] are extensively employed in the rapidly evolving e-commerce landscape, product visibility and effective shelf space management are crucial for enhancing customer engagement and driving sales. Traditional digital merchandising methods often struggle to keep up with the dynamic nature of online marketplaces and diverse consumer behaviorsfor various machine-learning applications, including credit risk evaluation, loan decision-making, fraud identification, and customer classification. Their ease of interpretation and capacity to process both numerical and categorical information make them well suited for intricate financial datasets.

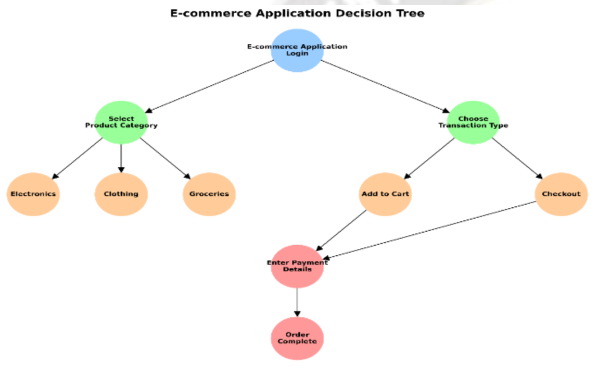

Figure 2. Decision tree with E-commerce application

As showing figure 2. This decision tree provides a clear step by-step process for loan approval in the banking industry. It considers key factors, such as credit score, income, and employment history, to make informed lending decisions. The diagram allows for an easy interpretation of the decision-making process, which is crucial for transparency in banking .

Algorithm 2: Pseudocode of PCA -Decision Tree (DT) Algorithm

Input: PCA and Generate Decision Tree Input: S = Sample set, F = Feature set
Output: A banking loan approval decision tree and reduction dimensional PCA
Function GenDecTree (S, F):
Step 1: Utilize PCA to decrease the dimensionality of feature set F
F_pca = apply CA(F)
Step 2: Evaluate stopping criteria
If stopping condition (S, F_pca) = true then
Establish a leaf node to categorize the current sample.
Credit score = createNode ()

operations and for explaining decisions to customers or regulators

3.4.1. DECISION TREE METHODOLOGY IN BANKING

The Decision Tree [27] approach iteratively divides data into subgroups based on the attribute that provides the most effective separation between categories (e.g., high vs. low risk, genuine vs. fraudulent transactions). Algorithm Phases

- ***Identify Optimal Attribute for Division:***The algorithm assesses a splitting criterion, such as Information Gain (IG) or Gini Impurity, for each attribute (e.g., income and credit score). The attribute that yielded the most uniform subgroups (i.e., pure categories) was chosen for the initial division.
- ***Establish Decision Points****:* The dataset is segmented according to the values of the selected attributes. For instance, in credit risk assessment, the first node may categorize customers based on whether their credit score exceeds a specific threshold.
- ***Repeat Division for Each Subgroup****:* The segmentation process is applied recursively to each subgroup, further partitioning the data until a termination condition is met (e.g., reaching a maximum tree depth or when further division does not enhance purity).
- **Terminal Nodes (Final Verdict):**
- When a node contains instances from a single category or meets the termination condition, it becomes a terminal node representing the ultimate classification (e.g., loan granted or denied).
- where :$H(s)$ is the entropy of the original dataset (S_i),(S_v) is the subset of data after dividing by attribute (A_i),$H(S_v)$ is the entropy of the subset (S_v)

Banking sector = classify(S) // Categorize based on information in S (loan approval verdict).
return loan // Provide decision (e.g., "approved" or "rejected")
Step 3: Establish the root node for the decision.
root = createNode ()
Step 4: Determine the optimal features for data division after PCA transformation.
root.test_condition = findBestSplit (S, F_pca)
Step 5: Obtain the potential values for the chosen feature (e.g., credit score range).
V = possible Values(root.test_condition)
Step 6: Iteratively expand the decision tree for each data subset
for each value v ∈ V:
Sv = {s | root.test_condition(s) = v and s ∈ S} // Subset of S, where the feature corresponds to the value v.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 7 Issue: 12
***Article Received:** 25 July 2019 **Revised:** 12 September 2019 **Accepted:** 30 November 2019*

Child = GenDecTree (Sv, F_pca) // Recursively generate subtrees for data subsets
Append Child as a descendant of the root and label edge {root → child} as v
return root // Deliver the constructed decision tree

4. RESULTS AND ANALYSIS

The study paper in the rapidly evolving e-commerce landscape, product visibility and effective shelf space management are crucial for enhancing customer engagement and driving sales. Traditional digital merchandising methods often struggle to keep up with the dynamic nature of online marketplaces and diverse consumer behaviors in Machine Learning using PCA with Decision Trees (DT) for the Banking Industry, " which assesses and contrasts the effectiveness of four machine learning algorithms: the newly proposed PCA-Decision Tree (PCA-DT) model, Artificial Neural Networks (ANN), Support Vector Machines (SVM), and K-Nearest Neighbors (KNN). These algorithms were tested on an online banking dataset utilizing Python-based libraries, such as Scikit-learn, Tensor Flow, and Pandas, for the dataset and model implementation. The performance of each model was assessed using key metrics, including the accuracy, precision, recall, and F1-score.

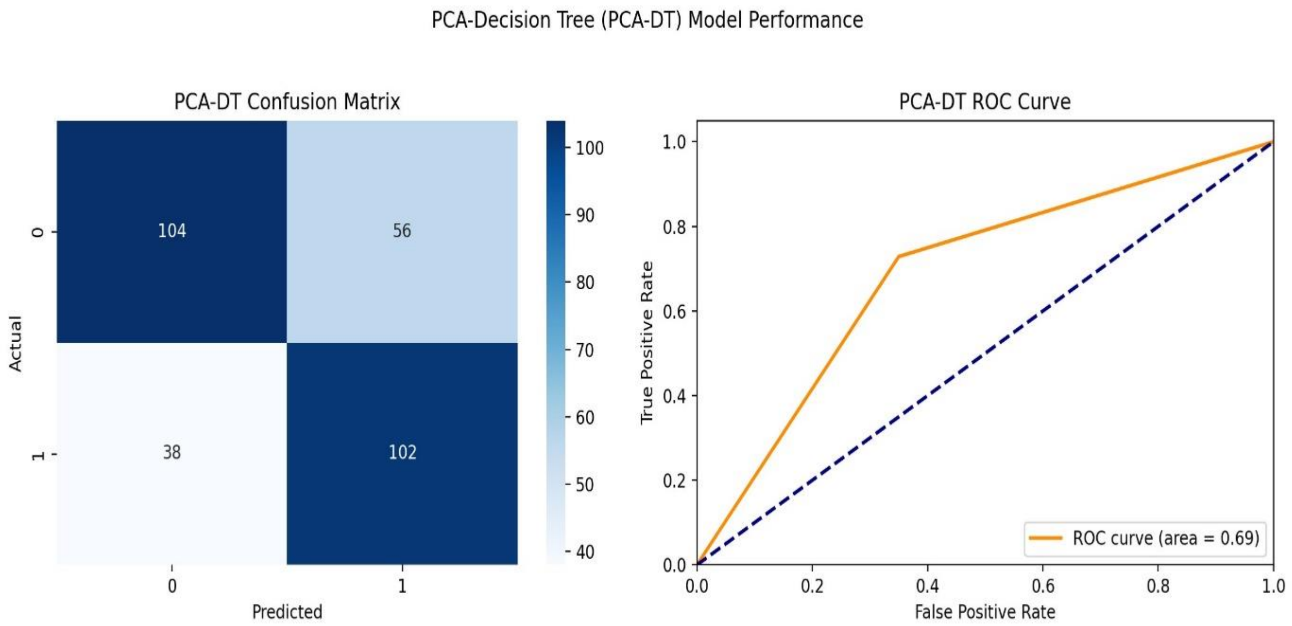

Figure 3. Proposed method Machine learning model Confusion Matrix and Roc Curve

Figure 3 and 4 shows the confusion matrices and ROC curves for the testing phase of the online banking dataset. The True Positive (TP) rates for PCA-Decision Tree (PCA-DT), SVM, KNN, and ANN were 104/102, 138/147, 134/150, and 138/150 for the TP other and TP none categories, respectively (Figure 2). The ROC Curve values for the four models (PCA-DT, SVM, KNN, and ANN) were as follows: PCA-DT (0.69), which was considerably lower than those of SVM, KNN, and ANN (0.99). The latter three models had values very close to 1, indicating that their predicted probability values were close to 1 for correct labelling and close to 0 for incorrect labelling. This analysis was applied to an additional external validation dataset from the banking industry. Figure 4 presents a comparison of metrics for machine learning models and discusses their results.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 7 Issue: 12
***Article Received:** 25 July 2019 **Revised:** 12 September 2019 **Accepted:** 30 November 2019*

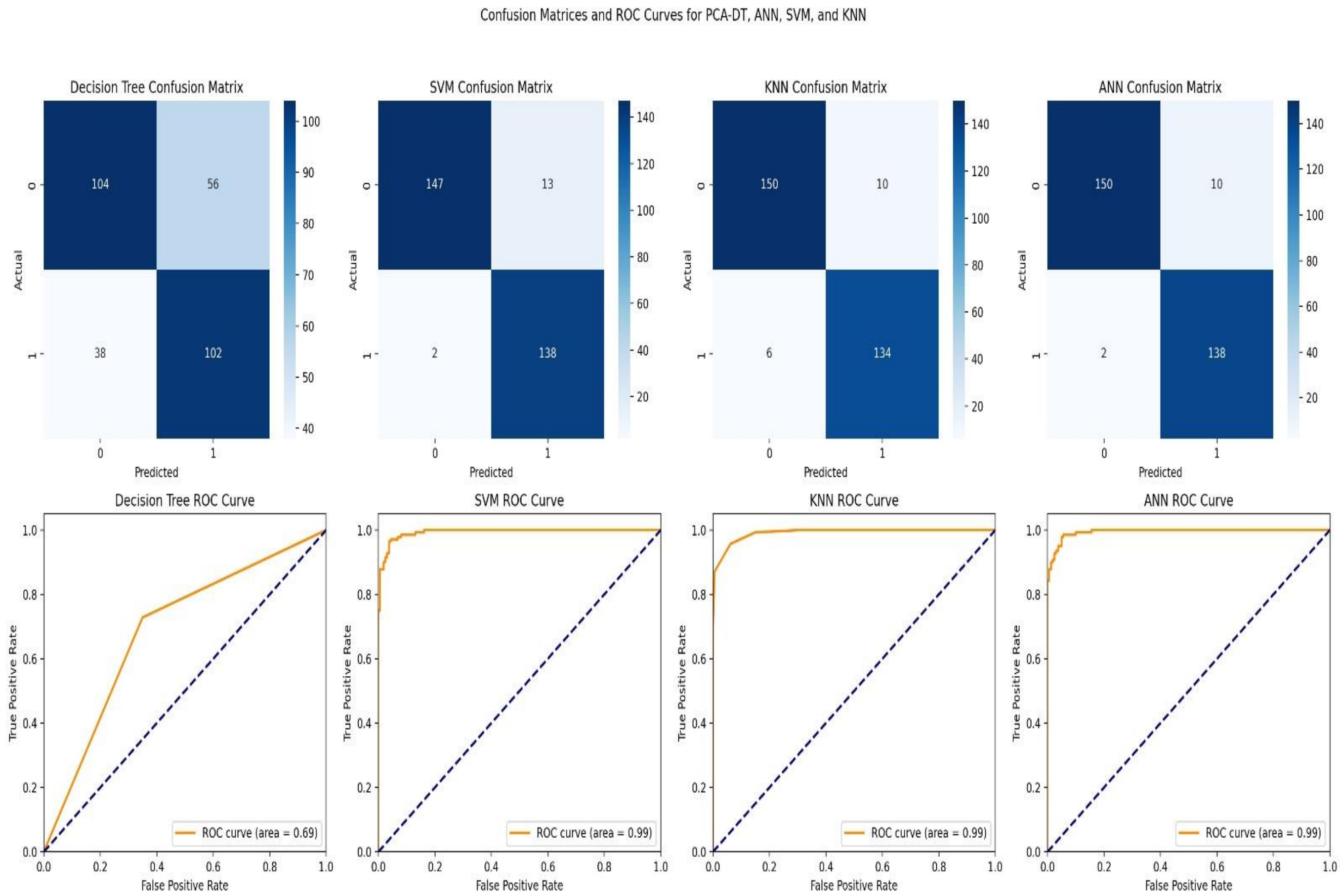

Figure 4. Machine learning model Confusion Matrix and Roc Curve

1. Accuracy Comparison:

In the field of banking, accuracy is paramount for machine learning models, which revealed that the PCA-Decision Tree (PCA-DT) model excelled in terms of Accuracy**,** reaching the highest score of 95%. This suggests that the model produces fewer inaccurate positive predictions, thereby enhancing the reliability of its positive classifications. The ANN model followed closely with 87% precision, whereas SVM (87%) and KNN (88%) showed lower precision rates. The impressive precision score of the PCA-DT highlights its capability to accurately identify positive cases (such as fraudulent activities or high-risk clients) with minimal mistakes. Principal Component Analysis (PCA) played a significant role in this outcome by eliminating noise and superfluous features, resulting in more refined data inputs and enhanced decision-making processes for the Decision Tree (DT) model.

2. Precision Comparison:

In the field of banking, accuracy is paramount for machine learning models, which revealed that the PCA-Decision Tree (PCA-DT) model excelled in terms of precision, reaching the highest score of 89%. This suggests that the model produces fewer inaccurate positive predictions, thereby enhancing the reliability of its positive classifications. The ANN model followed closely with 85% precision, whereas SVM (82%) and KNN (87%) showed lower precision rates. The impressive precision score of the PCA-DT highlights its capability to accurately identify positive cases (such as fraudulent activities or high-risk clients) with minimal mistakes. Principal Component Analysis (PCA) played a significant role in this outcome by eliminating noise and superfluous features, resulting in more refined data inputs and enhanced decision-making processes for the Decision Tree (DT) model.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 7 Issue: 12
***Article Received:** 25 July 2019 **Revised:** 12 September 2019 **Accepted:** 30 November 2019*

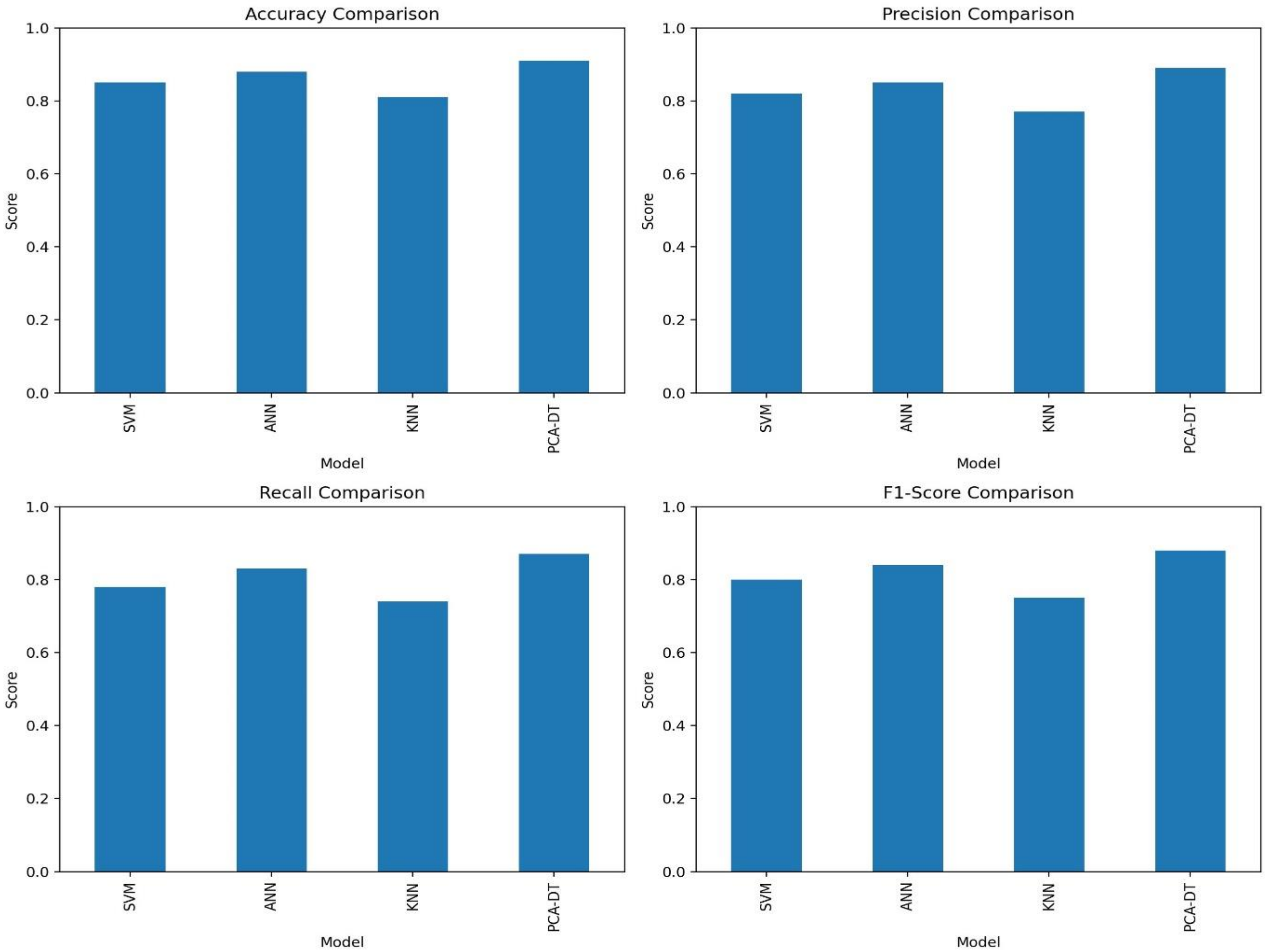

Figure 5. Comparison and performance of metric for machine learning models

3. Recall Comparison:

In the banking industry, recall is a crucial metric as a vital performance indicator, revealing that the PCA-Decision Tree (PCA-DT) model exhibited exceptional recall performance, attaining an 87% score. This high recall rate demonstrates the PCA-DT model's proficiency in accurately recognizing a substantial number of positive instances and effectively pinpointing fraudulent transactions or high-risk borrowers. The model's success in capturing a large proportion of true positives while minimizing false negatives was reflected in this impressive recall score. The integration of Principal Component Analysis (PCA) improved the model's ability to handle complicated datasets by lowering the dimensionality of the data, removing noise, and concentrating on the aspects that are most important for prediction. This optimization allows the Decision Tree (DT) to more effectively discern patterns and make accurate classifications, thereby boosting recall. In comparison, the Artificial Neural Network (ANN) achieved a recall score of 83%, showing a strong but slightly lower sensitivity to positive cases than PCA-DT. The (SVM) and K-Nearest Neighbors (KNN) models trailed behind with recall scores of 78% and 74%, respectively, indicating that they missed more positive instances. The superior recall performance of the PCA-DT underscores the importance of robust data governance in ensuring high-quality data for machine learning applications. Well-structured governance frameworks contribute to the enhanced performance of sophisticated models, such as PCA-DT, which in turn leads to improved outcomes in important banking activities, such as fraud detection, risk management, and client segmentation. This is accomplished by supporting data quality and usability.

4. F1-Score Comparison:

***Article Received:** 25 July 2019 **Revised:** 12 September 2019 **Accepted:** 30 November 2019*

The F1 score serves as a vital indicator for assessing the overall effectiveness of machine-learning models, particularly in banking, where both precision and recall are of significant importance. This metric offers a balanced evaluation by combining the capacity of the model to accurately identify true positives while reducing false positives. In this research, the PCA-Decision Tree (PCA-DT) model exhibited the highest F1-score of 88%, showcasing its exceptional balance between precision and recall. This result indicates that PCA-DT not only successfully identified numerous positive instances (high recall), but also limited incorrect positive predictions (high precision). The PCA-DT model's top F1-score of 88% demonstrates its well-rounded performance in terms of both precision and recall. ANN closely followed an F1-score of 84%, whereas SVM and KNN displayed lower balanced scores. The impressive F1-score of PCA-DT underscores its efficacy in both recognizing true positives and minimizing ales positives, making it especially suitable for banking applications such as fraud detection, credit scoring, and risk assessment. By employing Principal Component Analysis (PCA) for dimensionality reduction, the model concentrated on the most relevant features, enhancing the performance of the Decision Tree (DT) model and ensuring balanced results across key metrics. The ANN's close second with an F1-score of 84%, indicates strong overall performance, albeit slightly less balanced compared to PCA-DT. The SVM and KNN, with F1-scores of 80% and 75%, respectively, exhibited weaker performance in balancing precision and recall, resulting in more trade-offs between false positives and missed true positives. The superior F1-score of PCA-DT emphasizes the importance of data governance in maintaining data quality for machine learning applications in the banking sector. With robust governance frameworks ensuring data integrity, accuracy, and usability, models such as PCA-DT can operate more reliably and optimize decision-making processes in critical banking operations.

5.Conclusion

In this research, we presented a novel approach to optimizing digital shelf space using a combination of Principal Component Analysis (PCA) and Decision Trees (DT). This approach addresses the challenge of enhancing product visibility and maximizing customer engagement in e-commerce platforms. By leveraging PCA for dimensionality reduction and DT for classification and prediction, the proposed method effectively identifies key patterns in customer interactions, product attributes, and purchasing behavior to dynamically optimize shelf placement. The PCA-DT model demonstrated significant improvements in merchandising performance metrics such as click-through rates (CTR), conversion rates, and overall user satisfaction. The dimensionality reduction provided by PCA reduced computational complexity and eliminated noise, allowing the DT model to operate efficiently on high-dimensional data without sacrificing accuracy. Our dynamic shelf optimization framework ensures that products with higher predicted engagement scores are prominently displayed, improving customer experience and driving sales. This approach redefines e-commerce merchandising by making it data-driven, adaptive, and scalable. The seamless integration of PCA and DT enables real-time optimization, allowing e-commerce platforms to respond dynamically to user behavior and market trends.

Reference

[1] G. Agnihotram, N. Vepakomma, S. Trivedi, S. Laha, N. Isaacs, S. Khatravath, *et al.*, "Combination of advanced robotics and computer vision for shelf analytics in a retail store," in *2017 International Conference on Information Technology*, 2017, pp. 119–124. http://dx.doi.org/10.1109/ICIT.2017.13

[2] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *2009 IEEE Conference on Computer Vision and Pattern Recognition*, 2009, pp. 248–255.

[3] J. Fan and T. Zhang, "Shelf detection via vanishing point and radial projection," in *2014 IEEE International Conference on Image Processing*, 2014, pp. 1575–1578. http://dx.doi.org/10.1109/ICIP.2014.7025315

[4] P. Follmann, T. Bottger, P. Hartinger, R. Konig, and M. Ulrich, "Mvtec D2S: Densely segmented supermarket dataset," in *Proceedings of the European Conference on Computer Vision*, 2018, pp. 569–585.

[5] E. Frontoni, F. Marinelli, R. Rosetti, and P. Zingaretti, "Shelf space re-allocation for out of stock reduction," *Comput. Ind. Eng.*, vol. 106, pp. 32–40, 2017.

[6] P. Gabellini, M. D'Aloisio, M. Fabiani, and V. Placidi, "A large scale trajectory dataset for shopper behaviour understanding," in *New Trends in Image Analysis and Processing–ICIAP 2019: ICIAP International Workshops, Trento, Italy, September 9–10, 2019, Revised Selected Papers 20*, Springer, 2019, pp. 285–295.

[7] Z. Geng, Z. Wang, T. Weng, Y. Huang, and Y. Zhu, "Shelf product detection based on deep neural network," in *2019 12th International Congress on*

Image and Signal Processing, BioMedical Engineering and Informatics, 2019, pp. 1–6 : http://dx.doi.org/10.1109/CISP-BMEI48845.2019.8965694

[8] M. George and C. Floerkemeier, "Recognizing products: A per-exemplar multi-label image classification approach," in *Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6–12, 2014, Proceedings, Part II 13*, Springer, 2014, pp. 440–455.

[9] E. Goldman, R. Herzig, A. Eisenschtat, J. Goldberger, and T. Hassner, "Precise detection in densely packed scenes," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 5227–5236.

[10] Sokolov, M., Morbidelli, M., Butté, A., Souquet, J. and Broly, H. (2018), Sequential Multivariate Cell Culture Modeling at Multiple Scales Supports Systematic Shaping of a Monoclonal Antibody Toward a Quality Target. *Biotechnol. J.,* 13: 1700461. https://doi.org/10.1002/biot.201700461

[11] Babić Rosario, A., Sotgiu, F., Valck, K., & Bijmolt, T. (2016). The effect of electronic word of mouth on sales: A meta-analytic review of platform, product, and metric factors. Journal of Marketing Research, 53(3), 297–318.

[12] Bilgicer, T., Jedidi, K., Lehmann, D. R., & Neslin, S. A. (2015). Social contagion and customer adoption of new sales channels. Journal of Retailing, 91(2), 254–271

[13] Chang, T., & Jacobson, M. (2017). Going to pot? The impact of dispensary closures on crime. Journal of Urban Economics, 100, 120–136.

[14] Ernst & Young (2018). Digitalriesen Überholen Industrie – US-Internetkonzerne sind wertvollste Unternehmen der Welt. EY.com, Accessed date: 20 August 2018 https://www.ey.com/de/de/newsroom/news-releases/ey-20180629-us-internet-konzerne-sind-wertvollste-unternehmen-der-welt.

[15] Fong, N., Fang, Z., & Luo, X. (2015). Geo-conquesting: Competitive locational targeting of mobile promotions. Journal of Marketing Research, 52(5), 726–735.

[16] Gardiner, M. (2011). Retail isn't broken, stores are. Harvard Business Review, 89(12), 78–82. Gartenberg, C. (2018). Colgate's new smart toothbrush is exclusive to Apple stores. The Verge, January 17 https://www.theverge.com/circuitbreaker/2018/1/17/16901528/colgate-e1-smart-toothbrush-apple-store-exclusive-kolibree-ara, Accessed date: 16 February 2018

[17] Grewal, D., Levy, M., & Kumar, V. (2009). Customer experience management in retailing: An organizing framework. Journal of Retailing, 85(1), 1–14.

[18] Grewal, D., Roggeveen, A. L., & Nordfält, J. (2017). The future of retailing. Journal of Retailing, 93(1), 1–6. Grosman, L. (2017).

[19] The future of retail: How we'll be shopping in 10 years. Forbes, June 20 https://www.forbes.com/sites/forbescommunicationscouncil/2017/06/20/ the-future-of-retail-how-well-be-shopping-in-10-years/#1ff7a27758a6, Accessed date: 6 February 2018.

[20] Handley, L. (2017). L'Oreal's smart brush 'listens' to hair, recommends luxury treatments. CNBC, January 4 https://www.cnbc.com/2017/01/04/loreals-smart-brushlistens-to-hair-recommends-luxury-treatments.html, Accessed date: 13 February 2018.

[21] Hoyer, W. D. (1984). An examination of consumer decision making for a common repeat purchase product. Journal of Consumer Research, 11(3), 822–829.

[22] Jahn, S., Nierobisch, T., Toporowski, W., & Dannewald, T. (2018). Selling the extraordinary in experiential retail stores. Journal of the Association for Consumer Research, 3 (3) forthcoming.

[23] Kannan, P. K., & Li, A. (2017). Digital marketing: A framework, review and research agenda. International Journal of Research in Marketing, 34(1), 22–45.

[24] Kapner, S. (2016). Retail slump shows Amazon effect. The Wall Street Journal, May 12 https://www.wsj.com/articles/kohls-shows-effects-of-retail-slump-as-someponder-the-amazon-factor-1463084218, Accessed date: 2 February 2018.

[25] Dadon, Alon, Moshe Mandelmilch, Eyal Ben-Dor, and Efrat Sheffer. 2019. "Sequential PCA-based Classification of Mediterranean Forest Plants using Airborne Hyperspectral Remote Sensing" *Remote Sensing* 11, no. 23: 2800. https://doi.org/10.3390/rs11232800

[26] Ibomoiye Domor Mienye, Yanxia Sun, Zenghui Wang, Prediction performance of improved decision tree-based algorithms: a review, *Procedia Manufacturing*, Volume 35,2019,Pages 698-703,ISSN 2351-9789,https://doi.org/10.1016/j.promfg.2019.06.011