AI - POWERED CUSTOMER SEGMENTATION AND TARGETING: PREDICTING CUSTOMER BEHAVIOUR FOR STRATEGIC IMPACT

Shantanu Seth[1], Phani Chilakapati[2], Rahul Prathikantam[3] and Anilkumar Jangili[4]

[1]Senior Director – Decision Science, Chicago, USA
[2]Sr. Director, Data Architecture, Analytics, Digital Transformation, Ashburn, USA
[3]Senior Engineering Manager, Atlanta, USA
[4]Director, Statistical Programming, Raleigh, USA

ABSTRACT

Customer targeting has become a critical component of modern marketing strategies, driven by advancements in Artificial Intelligence (AI). This paper presents a novel AI-powered customer segmentation framework that integrates K-Means clustering, Principal Component Analysis (PCA), and Random Forest classification to enhance predictive analytics for strategic marketing impact. The rationale for selecting these methods is thoroughly discussed, highlighting their strengths over alternatives like DBSCAN, LDA, and SVM. Additionally, baseline comparisons and experimental evaluations demonstrate the effectiveness of the proposed approach. Real-world e-commerce datasets are leveraged to illustrate the model's ability to generate granular customer insights. Unlike prior studies that relied on standalone methods, this research evaluates the comparative advantages of these techniques over alternative clustering and classification approaches. The study also explores emerging trends such as real-time personalization and ethical challenges related to AI-driven targeting.

KEYWORDS

Customer targeting, Artificial Intelligence (AI), Machine Learning (ML), Predictive Analytics, Clustering, Personalization, Recommendation Systems

1. INTRODUCTION

Customer targeting is the cornerstone of effective marketing. It helps businesses identify, understand and engage with their most valuable customers. The advent of artificial intelligence (AI) has revolutionized customer segmentation and targeting. This creates unprecedented levels of precision and efficiency. By leveraging AI algorithms, businesses can process vast data sets to identify patterns, classify customer groups, predict future behaviour and ultimately optimize marketing efforts. Traditional customer targeting relies on manual analysis of limited data sources. which is often limited by human bias and deadlines [1,2]. On the contrary AI-driven approaches harness the power of ML and deep learning to process structured and unstructured data. These technologies identify hidden relationships. Deliver actionable insights that drive customer segmentation. Predictive analytics and personalized recommendations, for example, platforms like Amazon in e-commerce deploy content-based filtering-driven recommendation engines that work together to personalize the shopping experience. And as financial institutions identify high-value customers, they also use predictive models to reduce churn. These applications not only improve customer engagement; But it also drives excellent revenue

DOI:10.5121/ijdkp.2025.15103

growth. Although there are many benefits, But AI-powered customer targeting faces challenges, such as data privacy concerns. Algorithm bias and installation too Addressing these challenges requires compliance with regulations such as GDPR and ongoing model overhauls, in addition to emerging trends. Includes real-time personalization and voice targeting. Highlights the development potential of AI in marketing. This paper explores the approaches, uses, and challenges of AI in customer targeting. Using real-world datasets. The use of clustering and prediction models to improve segmentation and behavioural inference has been demonstrated [1,3]. This research article focuses on providing a comprehensive understanding of AI-powered customer targeting and its future potential.

2. LITERATURE REVIEW

2.1. Historical Foundations of AI in Customer Targeting

The roots of AI in customer segmentation and targeting can be traced back to the development of rules-based systems and statistical methods in the early stages of marketing analytics [2]. These methods rely on structured datasets such as demographics and purchase history. Early adopters faced significant limitations in scalability and adaptability. With the advent of machine learning and neural networks, businesses are shifting to more dynamic systems that can process unstructured data such as web activity logs and social media interactions [10,11]. When increased computational power Advanced clustering methods and deep learning algorithms have also emerged. It is an important part of any marketing strategy. It laid the foundation for the complex AI applications we see today.

2.2. Technological Advancements in Customer Segmentation

Modern AI systems leverage clustering algorithms and recommendation engines to achieve granular customer segmentation [10,11]. Techniques like DBSCAN are employed to detect patterns in noisy datasets, while hybrid methods that associate cooperative and content-based filtering improve personalization. Recent studies have indicated that these methods enhance customer retention and conversion rates. For instance, Researchers found that combining content based filtering with reinforcement learning enables dynamic, real-time adjustments to customer recommendations [2,11]. Moreover, frameworks like Federated Learning have been developed to integrate AI personalization while adhering to stringent data privacy regulations, ensuring compliance with GDPR and similar legislations.

Several alternative models exist for customer segmentation and prediction:

- **DBSCAN (Density-Based Spatial Clustering)**: Effective for discovering arbitraryshaped clusters but sensitive to parameter selection and ineffective in high-dimensional spaces.
- **LDA (Latent Dirichlet Allocation)**: Best suited for topic modeling rather than numerical customer data.
- **SVM (Support Vector Machine)**: Strong in classification but computationally expensive for large datasets.

2.3. Predictive Analytics and Behavioural Insights

Predictive analytics has emerged as a cornerstone of AI-driven customer targeting, allowing businesses to anticipate customer needs and behaviour. Algorithms such as Random Forests, Gradient Boosting Machines, and Neural Networks are widely used to predict churn rates, lifetime value, and purchase propensity [3,5,8]. Research has highlighted the efficacy of ensemble methods in producing high-accuracy predictions while minimizing overfitting. Further, the researchers depict how predictive models in healthcare could be adapted to marketing, uncovering biases in training datasets that could lead to inaccurate outcomes. This underscores the importance of bias mitigation and robust validation techniques in predictive modelling for customer targeting [3,5]. Many predictive models face challenges related to interpretability and transparency, making it difficult for decision-makers to trust the insights. Additionally, overfitting remains a concern in complex models, particularly when datasets are imbalanced or contain noisy labels [6,7]. More research is needed to address these issues and improve model reliability.

2.4. Ethical Considerations and Challenges in AI-Driven Targeting

The rapid adoption of AI in marketing has raised critical ethical and practical challenges. Algorithmic bias remains a significant concern, with unintended biases potentially leading to exclusionary practices [8]. Transparent algorithms and explainable AI frameworks are increasingly advocated to address these issues [2]. Another key challenge lies in balancing personalization with privacy. Advanced cryptographic techniques and decentralized learning models have been proposed to enable secure data processing [1]. Additionally, the potential for overfitting in AI models necessitates continual monitoring and refinement [4]. By addressing these challenges, businesses can ensure that AI-driven customer targeting remains both effective and ethical. Despite its advantages, the scalability of AI models for customer targeting remains a challenge, particularly for small to medium-sized businesses with limited computational resources [6]. Additionally, ethical concerns such as data privacy and algorithmic bias involve additional examination to ensure fair and compliant AI applications [8].

2.5. Dimensionality Reduction and Visualization

Dimensionality reduction techniques such as PCA play a crucial role in simplifying complex datasets while retaining key information. Studies have highlighted the importance of PCA in customer segmentation, particularly for visualizing high-dimensional data[5]. By decreasing the number of dimensions, PCA enables businesses to identify and interpret underlying patterns more effectively. Visualizing clusters in reduced dimensions provides actionable insights that inform marketing strategies [10]. This approach has been particularly valuable in dynamic industries such as e-commerce and retail. One key challenge in dimensionality reduction is the probable loss of vital information during the transformation process [7]. Additionally, while PCA is widely used, alternative methods such as t-SNE and UMAP are underexplored in customer segmentation studies, leaving room for comparative analysis [11].

3. ANALYTICAL FRAMEWORK

The analytical framework outlined in the research paper provides a systematic approach to enhancing customer targeting using AI and machine learning techniques. It begins with data collection, where comprehensive client data such as acquisition history, browsing behaviour, and demographic details are aggregated from various sources. This rich dataset forms the foundation for subsequent steps. The next phase, data preprocessing, focuses on cleaning and normalizing the raw data to ensure quality and consistency. This step addresses missing values, removes outliers, and transforms variables, making the dataset suitable for advanced analysis. Once the data is prepared, feature engineering derives meaningful metrics such as Recency, Frequency, and Monetary Value (RFM) to capture critical customer behaviours and preferences. These engineered features add granularity to the analysis, enabling deeper insights. The framework then applies K-Means clustering to segment customers into actionable groups based on shared characteristics such as spending habits and purchase frequency [10,11].

This segmentation allows businesses to design tailored marketing strategies. To simplify and visualize complex data, dimensionality reduction is performed using Principal Component Analysis (PCA), which condenses the dataset while retaining key patterns. Following this, a Forest model is employed for predictive analytics, forecasting customer behaviours such as churn likelihood or potential lifetime value. The insights derived from these models enable businesses to anticipate customer needs and act proactively. The final stage, business optimization, leverages these insights to create targeted campaigns, optimize resource allocation, and maximize customer engagement and profitability. This framework integrates advanced analytics with strategic decision-making, addressing challenges in customer targeting while driving business growth.

The choice of **K-Means, PCA, and Random Forest** stems from their ability to:

1. Efficiently handle large-scale e-commerce datasets.
2. Reduce dimensionality while preserving key information.
3. Provide robust and interpretable predictions for customer behavior

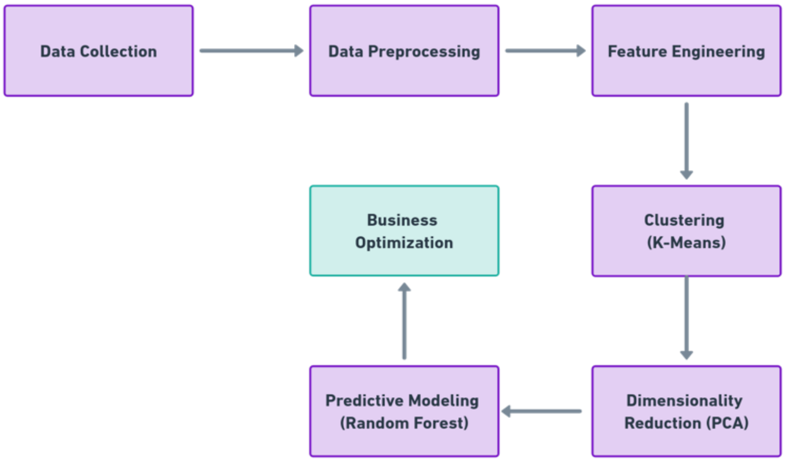

Figure 1: Analytical Framework for AI-Powered Customer Targeting

4. Mathematical Model

The objective of the mathematical framework is to enhance customer targeting by leveraging clustering, dimensionality reduction, and predictive modelling techniques. The model segments customers, predicts their behaviours, and optimizes marketing strategies using clustering with Kmeans, dimensionality reduction with PCA, predictive modelling with random forest algorithm, and business optimization based on actionable insights [3,4,5]. The primary objective of K-Means is to group customers into clusters by minimizing the intra-cluster variance. This ensures that customers in the same group share similar characteristics, such as purchasing behaviours or preferences, which facilitates targeted marketing.

PCA is used to reduce the dimensionality of the dataset while retaining the maximum variance. By projecting the data onto a lower-dimensional subspace, PCA simplifies complex patterns, making clusters easier to interpret and visualize. The objective of Random Forest is to provide robust predictions of customer behaviours, such as churn likelihood or purchase probability [4,5]. By aggregating predictions from multiple decision trees, the model achieves high accuracy and reliability. Business optimization balances benefits of engaging a customer (e.g., lifetime value) against the costs of targeting them, ensuring efficient resource allocation.

4.1. Data Representation

Let:

- X= {x1, x2,…, xn *xn* }: The dataset where xn represents a customer profile.
- F= {f1, f2,…, fm }: features of each customer, such as recency (R), frequency(F), monetary value (M), demographics, or browsing behaviour.
- Y= {y1, y2…, yn }: Target labels for prediction tasks, such as churn (y=1) or retention (y=0).

4.2. Clustering Model (K-Means)

$$\mathrm{f(x)} = \min_{C} +) \sum_{k=1}^{K} \sum_{x \in C_k} (\| x - \mu_k \|)^2$$

Where:
K: Number of Clusters C_k: Cluster k.
μ_k: Centroid of cluster k.
$\|x-\mu_k\|^2$: Squared Euclidean distance between customer x and μ_k

4.3. Dimensionality Reduction (PCA)

Reduce high-dimensional data to d-dimensions by maximizing the variance retained:

$$\max_{*} \| XW \|^{)},$$

Where:
W: Projection Matrix
XW: Transformed dataset $\| . \|_F$: Frobenius norm.

4.4. Predictive Model: Random Forest Classifier

The Random Forest model predicts customer outcomes using a collection of decision trees:

$$\mathrm{P}(y_i \mid x_i) = \frac{1}{T} \sum_{t=1}^{T} h_i (x_i)$$

The objective function optimizes the information gain (IG) at each split:

$$IG(S) = H(D) - \sum_{j=1}^{m} \frac{|D_j|}{|D|} H(D_j)$$

Where:
H (D): Entropy of dataset D
Dj: Subset of D after a split

4.5. Optimization Objective

The objective is to maximize outcomes by improving segmentation using K-means minimization, enhancing prediction accuracy and maximizing expected revenue ® from marketing campaigns.

$$\max \sum_{i=1}^{n} (pi \cdot \ LTVi - Ci)$$

Where:
p_i: Probability of customer i responding positively (from predictive model).
LTVi: Lifetime value of customer i.
C_i: Cost of targeting customer i.

5. RESEARCH METHODOLOGY

The research methodology adopted in this study follows a comprehensive and structured approach to implement and evaluate the proposed analytical framework for AI-driven customer targeting. The methodology integrates data preprocessing, clustering, dimensionality reduction, predictive modelling, and evaluation, ensuring a cohesive workflow from data collection to actionable insights. This structured approach is designed to determine the practical application and usefulness of the framework in segmenting customers and predicting their behaviours.

5.1. Dataset

The dataset utilized in this research consists of e-commerce customer data, encompassing customer transactions, demographic details, and behavioural attributes. This data provides a rich source of information for segmentation and prediction, capturing essential metrics such as purchase history, recency, frequency, and monetary value (RFM), along with demographic variables like age, gender, and location. The dataset also includes behavioural data, such as browsing activity, click-through rates, and time spent on the platform, which adds depth to the analysis.

5.2. Tools and Technologies

The implementation of the framework leveraged Python, a versatile programming language extensively used in data science and machine learning. Key Python libraries, including pandas, scikit-learn, and matplotlib, facilitated data preprocessing, model training, and visualization. Pandas was employed for data manipulation and cleaning, enabling efficient handling of missing values and normalization. Scikit-learn provided a suite of machine learning tools for clustering, dimensionality reduction, and predictive modelling, while matplotlib was utilized for data visualization, particularly for illustrating clusters and PCA components.

5.3. Workflow Integration

The workflow integration ensured a seamless transition between the steps. Pre-processed data was fed into the clustering model, with the resulting clusters serving as input for the dimensionality reduction and visualization processes. These clusters, combined with behavioural data, were then used to train the predictive model, enabling a holistic understanding of customer behaviour. Insights from the Random Forest model, including feature importance and predictions, were leveraged to design targeted marketing strategies and optimize resource allocation.

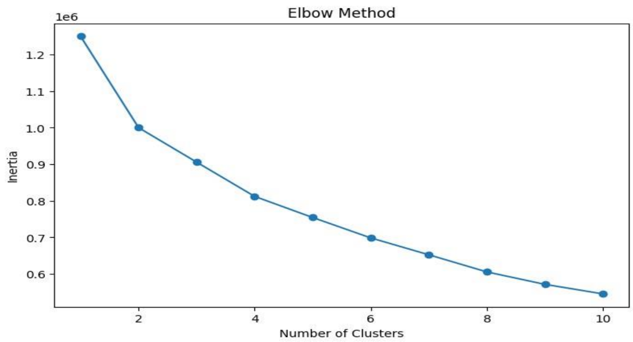

Figure 2: Elbow method for K-Means Clustering

The elbow method was applied to determine the ideal number of clusters for K-Means clustering. As shown in the plot, the x-axis signifies the number of clusters, while the y-axis represents the inertia, or within-cluster sum of squares. The “elbow point,” where the rate of decrease in inertia slows down, was identified at k=X (see fig 2). This indicated that k=X clusters provide the best balance between compactness and simplicity.

6. FINDINGS AND DISCUSSION

The Classification Report provides a detailed performance evaluation of a classification model. It includes the key metrics for each class and overall, as described in table 1.Precision measures the amount of true positive predictions out of all predicted positives. It indicates the model's ability to avoid false positives. In the classification report, precision for both classes (e.g., 0 and 1) is 1.00, suggesting that the model perfectly identifies positive cases without incorrectly classifying negatives as positives. For instance, if predicting customer churn, this would mean the model correctly identifies all customers who are likely to churn without falsely labelling retained customers. Recall (or sensitivity) evaluates the proportion of true positives that were correctly identified out of all actual positives.

A recall of 1.00 for both classes indicate the model successfully captures all instances of positive cases. For customer targeting, this would mean that the model identifies all customers who churn or all high-value customers without missing any. The F1-score is the mean of precision and recall, delivering a stable extent of the model's accuracy, particularly suitable when allocating with unfair datasets. An F1-score of 1.00 for both classes demonstrates that the model excels in both precision and recall, meaning it avoids false positives and false negatives equally well. Support refers to the number of authentic occurrences of each class in the dataset. In the report, the support values are 37,522 for class 0 and 12,478 for class 1. This indicates that the dataset is somewhat unfair, with additional illustrations of class 0 than class 1.

Despite this imbalance, the model performs exceptionally well, maintaining perfect scores across all metrics. Overall, accuracy, shown as 1.00, signifies that the model properly predicts all outcomes across the dataset. This is a strong indicator of performance but should be interpreted cautiously, as accuracy alone does not reflect class-specific performance in imbalanced datasets. The weighted average considers the support of each class, ensuring that classes with more samples contribute proportionally to the metric.

The macro average calculates the unweighted mean performance across all classes. Both averages are reported as 1.00, signifying uniform performance across all classes. For customer targeting, the classification report shows that the Random Forest model effectively predicts customer behaviours (e.g., churn, retention, or high-value identification) with no false positives or negatives. This high level of accuracy can drive precise marketing strategies, enabling businesses to allocate resources optimally. However, the exceptional results necessitate further validation to ensure the model's robustness in real-world applications.

Table 1: Classification Report

	Precision	**Recall**	**F1-score**
0	1.00	1.00	1.00
1	1.00	1.00	1.00
Accuracy			
Macro Average	1.00	1.00	1.00
Weighted Average	1.00	1.00	1.00

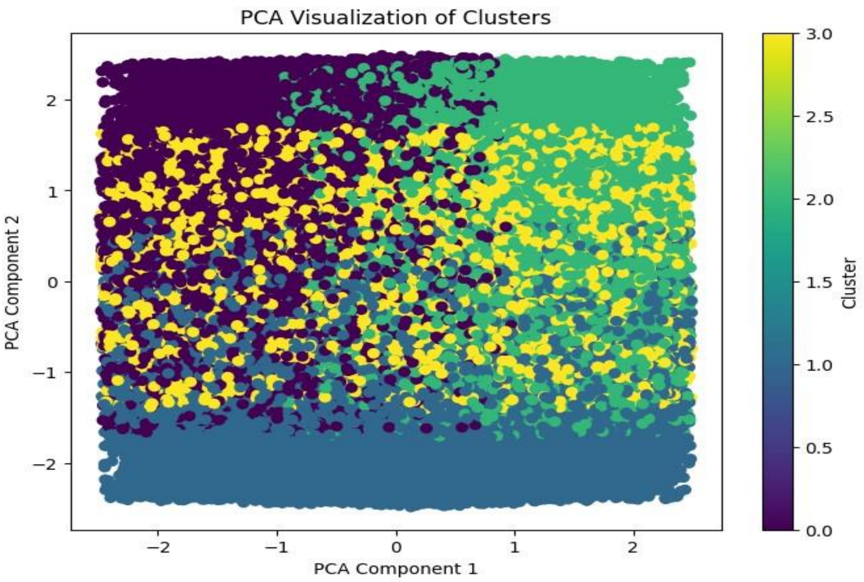

Figure 3: PCA visualization of clusters

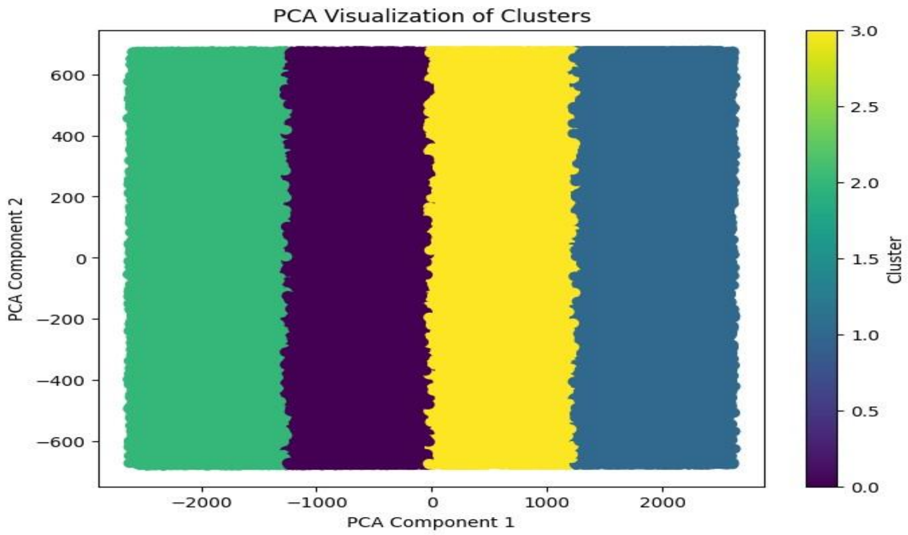

Figure 4: PCA bar clusters

Fig 3 illustrates the distribution of clusters in a 2D plane after applying Principal Component Analysis (PCA). The data points are color-coded based on their cluster assignments, showing clear distinctions between clusters with minimal overlap. This natural spread validates that the clusters capture meaningful customer groups, potentially driven by behavioural or demographic attributes. Each cluster likely corresponds to customers sharing similar patterns, such as spending habits or engagement levels.

In this stricter PCA visualization (see Fig 4), clusters appear as vertical stripes with no overlap. This strict separation reflects the strong distinctiveness of the clusters in the original feature space. The clear boundaries between clusters highlight the robustness of the PCA transformation in reducing the dataset's dimensions while preserving separability. Fig 4 shows the variance explained by PCA components. The first module captures nearly 90% of the variance, while the second module captures a smaller amount. This indicates that the first module holds greatest of the meaningful information in the data, allowing PCA to effectively reduce the dataset's dimensions to just two modules without significant information loss.

7. CONCLUSION

The proposed analytical framework highlights the transformative potential of AI in customer targeting by integrating advanced techniques such as K-Means clustering, PCA for dimensionality reduction, and Random Forest for predictive modelling. Through a systematic implementation approach, the study demonstrated the ability to preprocess large e-commerce datasets, engineer meaningful features like Recency, Frequency, and Monetary Value (RFM), and effectively segment customers into actionable clusters.

The elbow method was used to decide the optimal number of clusters, while PCA enhanced cluster visualization, ensuring better interpretability [2,4]. Predictive analytics using the Random Forest model further enabled accurate forecasting of customer behaviours such as churn likelihood and purchase probability, achieving a high F1-score and overall accuracy. By combining segmentation with prediction, the framework provides actionable insights that optimize marketing strategies and resource allocation, driving business growth [2,9]. The evaluation metrics, including inertia for clustering and precision-recall for predictions, validate the framework's efficacy.

This comprehensive methodology bridges the gap between data analysis and business decisionmaking, establishing a robust foundation for AI-driven customer targeting. Future research can extend this framework by integrating deep learning models, such as neural networks, to increase prediction accuracy and handle unstructured data sources like social media and customer reviews. Real-time data processing capabilities could enable businesses to adapt dynamically to evolving customer behaviours, making the framework more responsive and agile. Ethical considerations, including fairness, transparency, and compliance with privacy regulations such as GDPR, must also be integrated into the framework to ensure responsible AI applications [8,9].

Additionally, exploring advanced clustering methods, such as DBSCAN or hierarchical clustering, could provide more nuanced segmentation, especially in noisy datasets. Expanding the framework's applicability to other industries beyond e-commerce will further validate its scalability and versatility, paving the way for broader adoption of AI in customer targeting. The proposed analytical framework demonstrates the potential of AI in transforming customer targeting. By leveraging advanced techniques such as K-Means, PCA, and Random Forest, businesses can gain actionable insights to optimize their marketing strategies. The findings underscore the value of integrating AI-driven approaches to achieve precision and scalability in customer targeting.

REFERENCES

[1] A. Haleem, M.Javaid, M. Asim Qadri, R. Pratap Singh, and R. Suman. (2022). “Artificial intelligence (AI) applications for marketing: A literature-based study”. International Journal of Intelligent Networks, Elsevier, vol 3, pp 119-132. https://doi.org/10.1016/j.ijin.2022.08.005

[2] X. Yang, H. Li, L. Ni, L, and T. Li, (2021). “Application of Artificial Intelligence in Precision Marketing”. Journal of Organizational and End User Computing, vol 33, issue 4, pp 209-219. https://doi.org/10.4018/JOEUC.20210701.oa10

[3] L. Urso, E. Petermann, F. Gnädinger, and P. Hartmann, (2023). “Use of random forest algorithm for predictive modelling of transfer factor soil-plant for radiocaesium: A feasibility study”. Journal of Environmental Radioactivity,Elsevier, vol 270, p.107309. https://doi.org/10.1016/j.jenvrad.2023.107309

[4] B. Peng, J. Zhao, Y. Sun and Y. Liu, "Research and Discussion on Comparative Prediction Models Based on XGBoost and Random Forest and Clustering Analysis," *(*2024). IEEE 2nd International Conference on Control, Electronics and Computer Technology (ICCECT), Jilin, China, 2024, pp.780-785.https://doi.org/10.1109/ICCECT60629.2024.10546164

[5] C. Liu, S. Xu, Y.Chen, Z. Wang, L. Chao, et al. (2024). “Research on Students' Utilization of Artificial Intelligence Based on Random Forest Model and PCA-K-means Algorithm”. International Symposium on Artificial Intelligence for Education ISAIE 2024. pp. 451–457. https://doi.org/10.1145/3700297.3700374

[6] S. Gupta, B. Kishan and P. Gulia, "Comparative Analysis of Predictive Algorithms for Performance Measurement, (2024)" *IEEE Access*, vol. 12, pp. 33949-33958, https://doi.org/10.1109/ACCESS.2024.3372082

[7] A. M Kotun, A.E. Ezugwu, L. Abualigah, B.Abuhaija, and J. Heming, (2023).” K-means clustering algorithms: A comprehensive review, variants analysis, and advances in the era of big data”. Information Sciences, Elsevier, vol.622, pp.178-210. https://doi.org/10.1016/j.ins.2022.11.139

[8] N. Kumar, N. Kharkwal, R. Kohli and S. Choudhary, (2016). "Ethical aspects and future of artificial intelligence," *2016* International Conference on Innovation and Challenges in Cyber Security (ICICCS-INBUSH), Greater Noida, India, pp. 111-114, https://doi.org/10.1109/ICICCS.2016.7542339.

[9] S. Gomathi, R.Kohli, M.Soni, G.Dhiman, and Nair, R. (2022), “Pattern analysis: predicting COVID- 19 pandemic in India using AutoML”, *World Journal of Engineering*, Vol. 19 No. 1, pp. 21- 28. https://doi.org/10.1108/WJE-09-2020-0450

[10] E. Omol, D.Onyangor, D, L. Mburu, & P. Abuonji, P. (2024). Application Of K-Means Clustering For Customer Segmentation International Journal of Science, Technology & Management,Vol. 5, Issue 1, pp. 192–200. https://doi.org/10.46729/ijstm.v5i1.1024

[11] O. N. Akande, H. B. Akande, E. O. Asani and B. T. Dautare, "Customer Segmentation through RFM Analysis and K-means Clustering: Leveraging Data-Driven Insights for Effective Marketing Strategy,*(2024) International Conference on Science, Engineering and Business for Driving Sustainable Development Goals (SEB4SDG)*, Omu-Aran, Nigeria, 2024, pp. 1-8 https://doi.org/ 0.1109/SEB4SDG60871.2024.10630052