Journal of Computing Theories and Applications ISSN:3024-9104

Research Article

An Integrated Framework for Optimizing Customer Retention Budget using Clustering, Classification, and Mathematical Optimization

Amirthanathan Prashanthan *

Department of Applied Data and AI Research, Data Insighty Private Limited, Colombo 00600, Sri Lanka; e-mail : prashanthana@hotmail.com
* Corresponding Author : Amirthanathan Prashanthan

Abstract: The study presents a comprehensive framework for optimizing customer retention budget by integrating clustering, classification, and mathematical optimization techniques. The study begins with the IBM Telco dataset, which is prepared through data cleansing, encoding, and scaling. In the preliminary phase, customer segmentation is performed using K-Means clustering, with K = 3 and K = 4 identified as optimal based on the elbow method and Silhouette score. The configurations produced three (Premium, Standard, Low) and four (Premium, Standard Plus, Standard, Low) customer segments based on purchase preferences, which served as input features for churn prediction. In the second phase, the dataset was divided into training and test sets in an 80:20 ratio, followed by data balancing using the Synthetic Minority Over-sampling Technique (SMOTE) and Edited Nearest Neighbors (ENN). Multiple classification algorithms were evaluated, including Naive Bayes (NB), Random Forest (RF), Categorical Boosting (CatBoost), Light Gradient Boosting Machine (LightGBM), Extreme Gradient Boosting (XGBoost), Gradient Boosting (GB), Support Vector Machine (SVM), Logistic Regression (LR), K-Nearest Neighbors (KNN), and Multi-Layer Perceptron (MLP) using F1-score as the performance metric. CatBoost and LightGBM, with k values of 3 and 4, respectively, were the highest-performing classification models, with only minimal differences in performance. Ultimately, customer segmentation established customer prioritization, whereas churn prediction assessed customer churn likelihood. Four distinct configurations were assessed utilizing mixed-integer linear programming (MILP) to optimise retention budget allocation within uniform budget constraints, discount amounts, and churn thresholds. In both the k=3 and k=4 scenarios, CatBoost surpassed LightGBM, with CatBoost at K=3 effectively discounting 66% of at-risk consumers across all three segments, hence improving the intervention's efficacy and budget allocation, making it the ideal choice for maximizing customer retention. The results demonstrate the importance of segmentation in enhancing retention budgeting and budget optimization, particularly concerning parameter sensitivity.

Keywords: Budget optimization; Churn prediction; Classification; Clustering; Customer segmentation; Machine learning; Mathematical optimization; Mixed-Integer Linear programming.

Received: June, 23rd 2025
Revised: July, 13th 2025
Accepted: July, 17th 2025
Published: July, 20th 2025

1. Introduction

Telecommunications service providers are currently facing challenges in improving profit margins due to high licensing fees, expenses related to spectrum allocations, and rising customer expenditures, while also reducing prices. Furthermore, acquiring new customers is significantly more costly than retaining existing ones [1]. Although many studies have utilized machine learning to predict customer churn [2]–[10], in the telecommunications industry, the majority concentrate exclusively on enhancing predictive accuracy, neglecting the practical implementation of intervention strategies within budgetary limitations. Furthermore, current research often overlooks the cost-effectiveness of retention strategies, failing to correlate predictive insights with tangible business outcomes. Customers ought to be segmented based on their profitability for effective churn management. Identifying profitable customer segments enables us to retain our most significant customers through appropriate promotions and campaigns [11]. Furthermore, while studies exist that integrate clustering with classification to enhance churn prediction accuracy [12]–[14], there is a paucity of research examining the combined use of these methods to customize interventions for distinct customer segments. Research comparing various classifiers in optimized churn management is limited, especially regarding their real-world impact as measured by budget utilization, customer coverage, and intervention efficiency.

The primary objective of this research is to develop and assess a budget-conscious churn intervention framework that combines clustering-based customer segmentation with machine learning classification to identify and prioritize customers for retention. The study aims to compare various configurations—specifically different classifiers with varying cluster sizes—to determine the most cost-effective and impactful intervention strategy, ensuring optimal budget utilization while maximizing customer retention outcomes. The primary contribution of this study is as follows:

- This study presents a pragmatic, data-informed framework that integrates customer segmentation via clustering with churn prediction using machine learning, while considering real-world budget constraints. It links predictive insights to actionable interventions through cost-effective decision-making.
- The study compares various machine learning classifiers under different clustering configurations (K values), providing insights into accuracy, customer reach, intervention cost, and budget utilization — aspects often overlooked in churn analysis.
- It analyses how different customer segments (Low, Standard, Premium) respond to retention strategies. This enables more targeted, effective, and scalable churn reduction tailored to customer value and cost sensitivity.

The paper is organized as follows. Section 2 provides the literature review, followed by the methodology outlined in Section 3. Section 4 outlines the results and discussion. Section 5 delineates the strengths, limitations, and future directions, whereas Section 6 offers the conclusion

2. Literature Review

2.1. Clustering

Clustering utilizes the inherent structure of data distributions to establish criteria for grouping data points with analogous characteristics [15]. This clustering process partitions a dataset according to clustering criteria without necessitating prior knowledge of the data. In an optimal clustering situation, each cluster comprises data instances that exhibit high similarity among themselves while being dissimilar to instances in other clusters. The degree of dissimilarity depends on the characteristics of the data and the objectives of the clustering algorithm. Clustering is fundamental to numerous data-driven applications and is regarded as a significant and extensively researched task in machine learning. It also plays a crucial role in associated disciplines such as statistics, pattern recognition, computational geometry, bioinformatics, optimization, image processing, and others [16]–[18]. Table 1 below outlines the frequently employed clustering techniques used for segmenting customers within the telecom sector.

Table 1. Clustering techniques.

Clustering Technique	References
K-means Clustering	[12], [13], [19]–[22]
K-medoids Clustering	[12], [13], [21], [22]
X-means Clustering	[13]
DBSCAN	[11]
Random Clustering	[12], [13]

2.2. Classification

Predictive analytics has become an effective tool for reducing customer churn in the telecommunications industry [5]. Numerous studies in the telecommunications sector demonstrate how predictive analytics utilizes machine learning and artificial neural network classification techniques to predict churn [3], [23], [32]–[36], [24]–[31]. The Table 2 below shows the classification techniques employed for churn prediction in the telecoms industry, especially utilizing the same public dataset used in this study.

Table 2. Classification techniques.

Classification Technique	References
Naive Bayes	[37]–[42]
Logistic Regression	[38], [39], [41]–[45]
Decision Tree	[39], [40], [46]–[50]
K-Nearest Neighbours	[37], [38], [40], [41], [51], [52]
Support Vector Machines	[38], [40]–[42], [53]
Random Forest	[37]–[39], [42], [48], [54], [55]
Adaptive Boosting	[39]–[41], [56]
Gradient Boosting	[57], [58]
Extreme Gradient Boosting	[10], [39], [40], [59]
Light Gradient Boosting Machine	[45], [60]
Category Boosting	[58], [60]
Linear Discriminant Analysis	[40], [61], [62]
Stochastic Gradient Descent Classifier	[4]
Multi-layer Perceptron Neural Network	[2], [41], [45], [59], [63]
Convolutional Neural Network	[2], [24], [41], [64]
Long Short-Term Memory – Recurrent Neural Network	[41], [42], [58], [65], [66]

2.3. Mathematical Optimization

Mathematical optimization has garnered significant attention from both scientific and practical perspectives to achieve optimal solutions across multiple objectives. The principal objective is often regarded as minimizing operational expenses [67]. A diverse array of industries employs mathematical optimization to enhance budgeting, including education [68], transportation [69], telecommunications [70], manufacturing [71], health [72], human resources [73], and construction [74]. Optimization problems can be categorized based on the characteristics of the optimization variables and the analytical properties of the objective and constraint functions, such as linear versus nonlinear, unconstrained versus constrained, smooth versus non-smooth, convex versus non-convex, stochastic versus deterministic, and integer versus continuous [75]. Numerous optimization techniques are employed in optimization problems; Table 3 below illustrates several widely utilized techniques in mathematical optimization.

Table 3. Optimization techniques.

Clustering Technique	References
Integer linear programming	[76]–[79]
Nonlinear programming	[80]–[82]
Stochastic optimization	[73]
Adaptive robust optimization	[68]
Bayesian optimization	[83], [84]

3. Proposed Method

This section outlines the methodological techniques of the study. Figure 1 succinctly outlines each of the main stages of the methodology.

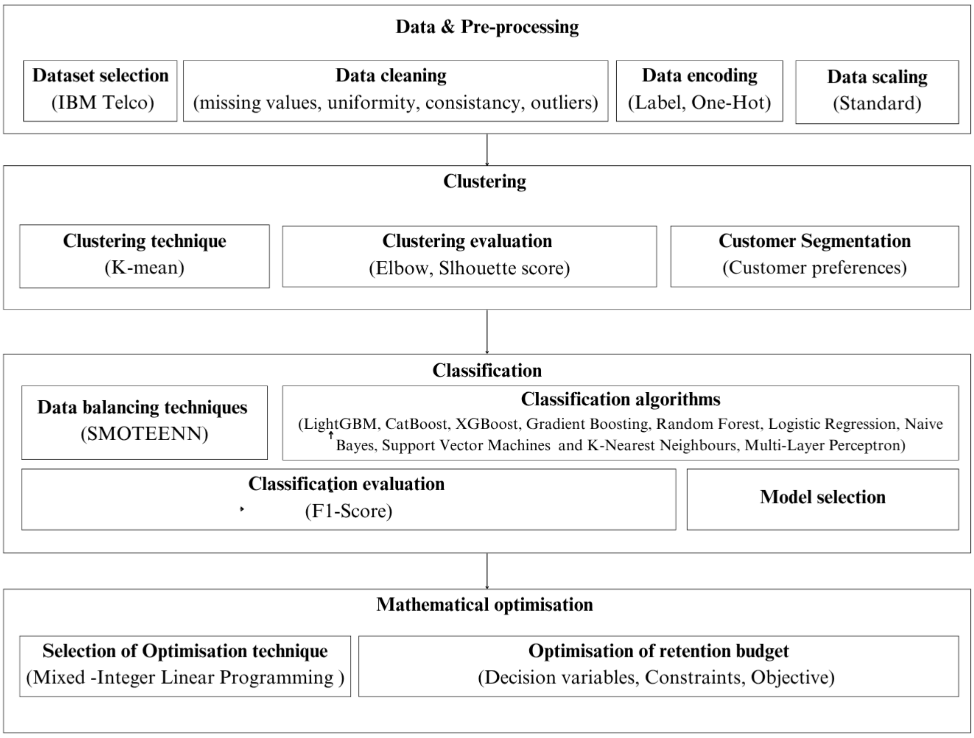


Figure 1. Main stages of the methodology of research

3.1. Data and pre-processing

3.1.1. Dataset selection

The study employs the publicly available IBM Telco Dataset [85]. The study dataset has 7,043 instances and 21 attributes, including 20 unique features and one target variable. The dataset comprises 26.54% of samples labelled as churn. The IBM Telco dataset comprises both categorical characteristics (e.g., service subscriptions, payment methods, contract types) and numerical variables (e.g., monthly charges, total charges), rendering it a diversified and valuable resource for analyzing customer behaviour. K-means, hierarchical clustering, and DBSCAN are suitable clustering algorithms, as they effectively manage both categorical and numerical data, particularly after the encoding of categorical variables. The dataset includes a binary target variable, churn, making it suitable for classification methods such as Logistic Regression (LR), Decision Trees, and Random Forests (RF), as well as ensemble learning algorithms, which are effective for predicting binary outcomes. The dataset's moderate size and diversity in consumer characteristics facilitate the application of machine learning techniques to discern trends in customer behaviour, including churn prediction and customer segmentation based on analogous preferences. By correlating method selection with data features, the study guarantees that the selected procedures can effectively manage the dataset's complexity, yielding significant insights into customer churn and retention strategies.

3.1.2. Data cleaning

Data cleaning involved refining the dataset by eliminating extraneous columns, standardizing column names for consistency, validating data types, and addressing missing values and outliers. Thus, box plots were utilised to assess the presence of outliers in the numerical variables of tenure, monthly charges, and total charges, revealing that none of the variables exhibited outliers in the overall distribution.

3.1.3. Data encoding

The study employed label encoding and one-hot encoding techniques to convert the identified nominal data into a numerical format.

3.1.4. Data scaling

The numerical columns were standardized via StandardScaler from Scikit-learn. This change ensured that each variable had a mean of 0 and a standard deviation of 1, thereby avoiding factors with broader ranges from significantly affecting the clustering outcomes.

3.2. Clustering analytics

3.2.1. Rationale for selecting K-means

K-means is a rapid and clearly comprehensible clustering technique, optimal for datasets characterized by distinct, spherical clusters; nonetheless, it encounters difficulties with noise, outliers, and sensitivity to initialization [86], [87]. Agglomerative Hierarchical Clustering, in comparison, provides flexibility by accommodating any distance metric and presents a visual dendrogram for cluster determination; however, it is computationally demanding, and early merging errors are irreversible [86], [88]. DBSCAN is notable for its resilience to outliers and capacity to identify clusters of arbitrary shapes without requiring a predetermined quantity; yet, it is significantly affected by parameter configurations and exhibits suboptimal performance in high-dimensional environments or datasets with heterogeneous densities [86], [89], [90].

The IBM Telco Customer Churn dataset is amenable to K-means clustering, owing to its efficiency, scalability, and interpretability, particularly when the data undergoes appropriate preprocessing, such as encoding categorical variables and normalizing features. Considering the dataset's relative structure and minimal noise, K-means effectively categorizes consumers according to analogous usage patterns, contract kinds, or churn behaviour.

3.2.2. K-mean clustering

K-means clustering is recognized as an unsupervised learning technique employed to address clustering-related issues [91], [92]. The k-means clustering algorithm is considered one of the most effective and widely utilized clustering techniques within the research community [14]. K-means clustering is a method for categorizing a dataset into a specified number of clusters, each characterized by k centroids. The k centers must be strategically positioned, as different locations yield distinct outcomes. The outcome will improve if each cluster is maximally distant from the others. The optimal number of clusters (k) that maximizes the distance calculable from the dataset [14], [93]. The k-means algorithm relies on the predetermined value of k, which must be specified to conduct any clustering analysis. Employing various k values in clustering will ultimately yield distinct outcomes [14]. This study utilizes optimal k values to delineate clusters for customer segmentation prior to churn prediction through classification.

3.2.3. Clustering evaluation

Determining the optimal number of clusters (k) is a crucial step in k-means clustering. The elbow method [94] and silhouette score [95] were utilized to determine the optimal number of clusters for customer segmentation, ensuring that the clustering method effectively delineates unique customer segments.

3.2.4. Customer segmentation

This research employed an unsupervised machine learning methodology to categorize customers based on their shopping behavior. A compilation of features relevant to customer purchase behavior was identified, including service subscriptions, billing preferences, payment methods, internet and contract types, monetary variables, and churn status. The following 22 factors were explicitly selected as demonstrated in Table 4.

Table 4. Optimization techniques.

Category	Features
Service subscriptions	phone_service, online_security, online_backup, device_protection, tech_support, streaming_tv, streaming_movies, multiple_lines_yes
Billing and payment preferences	paperless_billing, payment_method_Bank transfer (automatic), payment_method_Credit card (automatic), payment_method_Electronic check, payment_method_Mailed check
Internet and contract types	internet_service_DSL, internet_service_Fiber optic, contract_Month-to-month, contract_One year, contract_Two year
Monetary variables	monthly_charges, total_charges
Churn status	churn

K-Means clustering was used to categorize customers into uniform segments based on their standard purchase preferences. The number of clusters was predetermined as three (k = 3) and four (k = 4) based on cluster evaluation analysis and domain expertise. The algorithm was initialized with a predetermined random state (random_state=42) to guarantee the reproducibility of findings. Subsequently, each consumer was assigned to one of the three designated "preference clusters." Preference clusters utilized for customer segmentation based on purchase behaviour have been incorporated as a feature for churn prediction.

3.3. Classification analytics

3.3.1. Data balancing

Due to the uneven characteristics of churn datasets, SMOTEENN was utilized on the training data [96], [97]. Before that, the dataset was divided into training (80%) and testing (20%) sets by stratified sampling to maintain class distribution.

3.3.2. Selection of algorithms

This study employed a variety of classification algorithms to encompass a wide range of classification techniques, thereby facilitating a thorough assessment of performance across different data attributes. Ensemble methods like LightGBM, CatBoost, XGBoost, Gradient Boosting [98], and RF [54] were selected for their superior predictive capability, resilience to overfitting, and proficiency in managing intricate, nonlinear interactions [99]. Ensemble learning models are ideally suited for the IBM Telco Customer Churn dataset because of its heterogeneous data types, class imbalance, and nonlinear feature interactions. These models inherently accommodate categorical variables, obviating the need for intricate encoding and preserving critical feature information. They also automatically capture complex interactions and threshold effects prevalent in customer behaviour patterns, which linear models cannot adequately represent without manual feature engineering.

In contrast to neural networks, which typically necessitate substantial datasets and considerable parameter optimization, these models are more efficient, precise, and suitable for tabular datasets such as this [100], [101]. Logistic Regression and Naive Bayes [102] were selected for their simplicity, interpretability, and robust performance on linearly separable and high-dimensional datasets, respectively. Support Vector Machines [103] and K-Nearest Neighbours [104]were incorporated to exemplify geometric and instance-based learning techniques, providing insights into the performance of models predicated on margin maximization and local structure. The Multi-Layer Perceptron [63] neural network possesses deep learning functionalities, enabling it to identify intricate patterns that conventional models may overlook. These algorithms collectively provide a comprehensive framework for evaluating classification performance and selecting the most suitable model for a specific challenge.

3.3.3. Performance evaluation

The effectiveness of a classification model can be evaluated using various performance metrics. The performance metrics Table A1 include accuracy (1), precision (2), recall (3), and F-measure (4). The metrics' values can be calculated from the true positive (TP), false positive (FP), false negative (FN), and true negative (TN) values obtained from the confusion matrix [105].

3.4. Mathematical optimisation

3.4.1. Selection of MILP

MILP is a recognized, precise optimization method that ensures globally optimal solutions for linear problems within constraints [106]. Metaheuristic approaches are frequently employed in budget and resource allocation tasks, particularly for nonlinear issues or those with complex search spaces; nonetheless, they do not ensure optimal solutions and generally necessitate considerable tuning [107]–[109]. Conversely, the issue presented is formulated to permit a precise resolution via MILP, rendering it more suitable for guaranteeing optimal and consistent judgments within specified resource constraints. Moreover, MILP solvers are well-developed, scalable, and effective for medium-sized problems, which corresponds appropriately with the scope and magnitude of the dataset and decision variables pertinent to this investigation. Consequently, MILP was selected not only for its alignment with the problem structure but also for its capacity to provide interpretable and demonstrably optimal results—attributes that metaheuristics cannot guarantee. Table A2 delineates the nature of each variable in the integer linear programming formulation, affirming that it constitutes a MILP problem.

3.4.2. Optimization

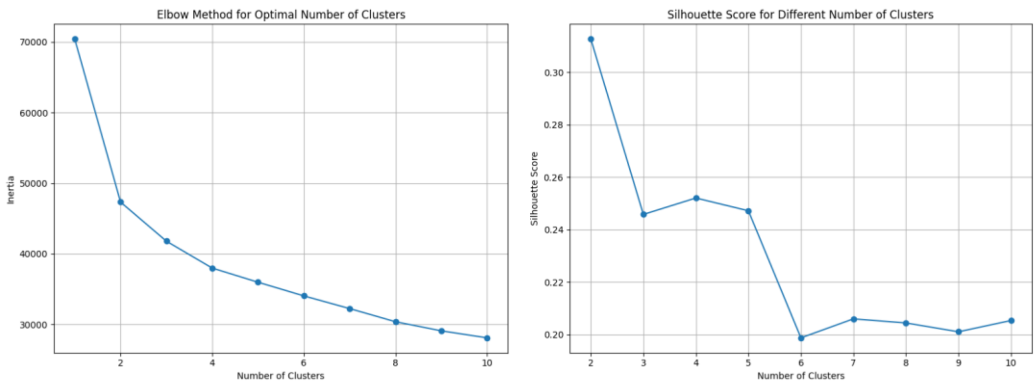
The study formulated an optimization model using MILP to ascertain the optimal number of at-risk consumers eligible for discounts within a limited budget. This strategy emphasises the high-value customers identified in the previous customer segmentation step, while adhering to budget limitations. Table A3 summarizes the components and details of this MILP optimization model.

- The objective function (5) aims to maximize the quantity of consumers whose churn probability falls below the threshold.
- The budget constraint (6) mandates that the cumulative intervention cost for all chosen consumers shall not surpass the allocated budget.
- The churn reduction constraint (7) stipulates that the decline in churn for every customer must not surpass their baseline churn probability.
- The churn below threshold constraint (8) ensures that the churn probability for each customer after the intervention remains beneath the specified threshold.
- The intervention consistency constraint (9) ensures congruence between the intervention and the reduction in churn.
- This constraint of prioritization (10) ensures that customers selected for intervention are prioritised based on customer segmentation.

4. Results and Discussion

4.1. Clustering analysis

The inflection point, often termed the “elbow,” was identified at approximately three or four clusters, suggesting that further clusters beyond this threshold yield only negligible improvements in compactness. Additionally, there is a minimal variation in the Silhouette Score between three and four clusters. Based on these observations, the selection of k = 3 or k = 4 is deemed reasonable. Figure 2 illustrates the outcomes of both evaluation methods.



(a) (b)

Figure 2. (**a**) Elbow method graph for optimal number of clusters; (**b**) Silhouette score graph for different number of clusters

The classification study assessed both k = 3 and k = 4 configurations to identify the optimal model for predicting churn by discerning various patterns.

4.2. Classification analysis

In telecommunication churn prediction, a false positive occurs when the model predicts that a customer will churn, but the customer remains. This may result in unnecessary expenditures on retention initiatives for consumers who were not at risk of leaving. A false negative occurs when the model predicts customer retention, but the customer ultimately churns, resulting in lost opportunities for proactive retention of important customers. Given that all kinds of errors incur significant, although distinct costs, it is imperative to employ a metric that equilibrates these inaccuracies. The F1-score is frequently regarded as the optimal metric for this purpose, as it incorporates precision (the accuracy of positive churn predictions) and recall (the capacity to identify actual churners) into a singular value. In contrast to accuracy, which can be deceptive in imbalanced datasets common in churn prediction, the F1-score provides a fair assessment by penalizing both false positives and false negatives, making it optimal for evaluating churn models where the cost of each inaccuracy is significant. The F1-score was employed as the ranking parameter to select the optimal classifier model. Tables 5 and 6 exhibit the classifier configurations for k = 3 and k = 4, respectively, arranged in a descending order according to their F1-scores.

Table 5. Classifier ranked by F1-score for k = 3.

Rank	Classifier	Accuracy	Precision	Recall	F1 Score
1	CatBoost	0.755855	0.526042	0.810160	0.637895
2	LightGBM	0.752307	0.521664	0.804813	0.633018
3	XGBoost	0.754436	0.524823	0.791444	0.631130
4	RF	0.748048	0.516239	0.807487	0.629823
5	GB	0.735273	0.500810	0.826203	0.623613
6	MLP	0.733854	0.499165	0.799465	0.614594
7	SVM	0.720369	0.484026	0.810160	0.606000
8	LR	0.696948	0.462089	0.863636	0.602050
9	NB	0.692690	0.456554	0.828877	0.588794
10	KNN	0.674237	0.440225	0.836898	0.576959

Table 6. Classifier ranked by F1-score for k = 4.

Rank	Classifier	Accuracy	Precision	Recall	F1 Score
1	LightGBM	0.755855	0.526316	0.802139	0.635593
2	CatBoost	0.752307	0.521739	0.802139	0.632244
3	GB	0.742370	0.509031	0.828877	0.630722
4	RF	0.747339	0.515571	0.796791	0.626050
5	XGBoost	0.746629	0.514991	0.780749	0.620616
6	LR	0.706884	0.471449	0.860963	0.609272
7	MLP	0.726757	0.490909	0.794118	0.606742
8	NB	0.716111	0.479624	0.818182	0.604743
9	SVM	0.717530	0.481013	0.812834	0.604374
10	KNN	0.675656	0.441467	0.836898	0.578024

Both CatBoost (k=3) and LightGBM (k=4) attained equivalent accuracy of 0.755855, indicating that both classifiers exhibited identical predictive performance. Nonetheless, upon comparing their additional metrics, nuanced distinctions arise. CatBoost exhibited marginally superior recall, attaining a value of 0.810160, which signifies that it accurately recognized 81% of the genuine positive cases. Conversely, LightGBM exhibited a recall of 0.802139, marginally lower by 0.008. This indicates that CatBoost exhibited marginally greater efficacy in identifying positive instances. LightGBM exhibited a slight advantage in precision, recording a value of 0.526316, whereas CatBoost achieved a value of 0.526042. This indicates that LightGBM exhibited a marginally superior rate of accurate positive predictions when it rendered a positive classification. In conclusion, the F1 score, which balances precision and recall, was 0.637895 for CatBoost and 0.635593 for LightGBM, indicating a marginal advantage for CatBoost in achieving this balance.

4.3. Optimization analysis

This study next investigates the same k value across both models (CatBoost, LightGBM) to assess the impact of model selection and k value on the optimization result. It also evaluates retention budget optimization for both k=3 and k=4 using the optimal models CatBoost and LightGBM, respectively.

4.3.1. Comparison of intervention analyses for k=3

Table 7 presents intervention analyses for customer segments using k=3 segmentation with CatBoost and LightGBM models, respectively.

Table 7. Intervention analysis for k=3.

Customer Type	Model	Total Intervention Cost	Average Intervention Cost	Total Customers
Low	CatBoost	4,752.72	30.47	156
	LightGBM	3,941.84	26.63	148
Standard	CatBoost	12,490.86	56.52	221
	LightGBM	13,407.11	59.32	226
Premium	CatBoost	7,720.90	514.73	15
	LightGBM	7,571.54	582.42	13

The comparative analysis of intervention strategies using CatBoost and LightGBM across selected customer segments reveals notable distinctions. Within the Low customer segment, CatBoost engaged a marginally greater number of customers, accompanied by a slightly higher total and average intervention cost relative to LightGBM. For the Standard segment, LightGBM demonstrated higher total and average intervention expenditures, resulting in a marginally larger customer base compared to CatBoost. Notably, in the Premium segment, CatBoost targeted a greater number of customers while incurring a lower average intervention cost, indicating enhanced cost efficiency within this high-value cohort. Collectively, these findings suggest that CatBoost may offer a more balanced approach in terms of cost-effectiveness and customer reach, particularly in the Low and Premium segments. In contrast, LightGBM appears to prioritize more intensive interventions, as reflected in its higher per-customer costs, especially within the Premium category.

4.3.2. Comparison of intervention analyses for k=4

Table 8 displays the intervention analyses for customer segments utilizing k=4 segmentation, employing LightGBM and CatBoost models, respectively.

Table 8. Intervention analysis for k=4.

Customer Type	Model	Total Intervention Cost	Average Intervention Cost	Total Customers
Premium	CatBoost	9,202.88	541.34	17
	LightGBM	10,167.25	564.85	18
Standard Plus	CatBoost	15,626.75	65.65	238
	LightGBM	14,700.74	63.37	232

The intervention analysis for k = 4, focusing on selected customers, reveals distinct patterns across customer segments and models. Within the Premium segment, both CatBoost and LightGBM demonstrate comparable total intervention costs and average intervention costs, with LightGBM incurring a marginally higher expenditure per customer. The total number of customers targeted by both models is also similar, indicating consistent coverage in this high-value segment. In the Standard Plus segment, CatBoost exhibits a slightly higher total intervention cost and average cost per customer compared to LightGBM, while reaching a marginally greater number of customers. These results suggest that CatBoost may adopt a more resource-intensive strategy in the Standard Plus group, potentially reflecting a prioritization of broader customer engagement. Conversely, LightGBM appears to allocate resources more conservatively in this segment, maintaining costs at a marginally lower level. Overall, the findings suggest that while both models exhibit similar performance in terms of intervention scope and expenditure in the Premium segment, they diverge somewhat in their approach within the Standard Plus segment.

4.3.3. Comparison of optimization efficiency for different models

According to Table 9, the analysis of intervention strategies across varying cluster sizes and machine learning models reveals that employing k = 3 with CatBoost provides the most effective optimization for customer churn management within the allocated budget. This configuration achieves a high intervention efficiency of 66.1%, closely rivaling only the LightGBM model at the same cluster size; yet, it distinguishes itself by selecting a greater number of customers for intervention (392 compared to 387). Moreover, the broader coverage of customer segments—including Premium, Standard, and Low tiers—ensures a more comprehensive approach to churn mitigation. Despite similar budget utilization across all scenarios, the superior efficiency and expanded customer reach under the k=3 CatBoost model indicate enhanced operational effectiveness. In contrast, models employing four clusters (k=4) demonstrate markedly lower intervention efficiencies, around 43.5%, and restrict their focus to fewer customer segments. Consequently, the evidence supports the conclusion that k=3, with CatBoost and MILP optimally balancing intervention success, resource allocation, and customer coverage, makes it the preferred strategy for maximizing churn reduction.

Table 9. Analysis of intervention strategies.

Factors	k = 3 with CatBoost	k = 3 with LightGBM	k = 4 with CatBoost	k = 4 with LightGBM
Available Budget	$25000.00	$25000.00	$25000.00	$25000.00
Discounted Total charges	10%	10%	10%	10%
Discounted Monthly charges	10%	10%	10%	10%
Threshold churn probability	0.3	0.3	0.3	0.3
Total "Likely to churn" customer	593	577	586	570
Min Intervention Cost	$3.86	$3.86	$3.86	$3.86
Max Intervention Cost	$710.71	$749.12	$723.04	$749.12
Customers Selected for Intervention	392	387	255	250
Intervention efficiency-customer	66.1%	67.1%	43.5%	43.9%
Total Budget Utilized	$24964.48	$24920.48	$24829.63	$24867.99
Selected customer segments	Premium, Standard and Low	Premium, Standard and Low	Premium and Standard plus	Premium and Standard plus

4.4. Optimal framework configuration

The optimal framework configuration, as indicated by the preceding table and analysis, comprises K-Means clustering (k = 3), a CatBoost classifier, and Mixed-Integer Linear Programming (MILP) for optimization. Figure 3 illustrates the flowchart of the optimal framework configuration.

K-Means with k = 3 distinctly segments the customer base into three categories: Premium, Standard, and Low. These clusters offer a more targeted approach to customer involvement, ensuring that each customer segment is addressed according to its unique requirements and propensity for churn. Due to the high dimensionality of the original features, Principal Component Analysis (PCA) was used to reduce the data to two principal components. This dimensionality reduction enabled the visualisation of clustering results in a two-dimensional space with minimal information loss, offering a clearer understanding of cluster separation and cohesion. Figure 4 illustrates the visualization of K-Means clusters in 2D PCA space for k = 3.

A subset of features closely associated with product and service preferences was chosen to reveal distinct customer purchasing behaviour patterns. The following items were included: Utilization of services (e.g., online security, technical support, streaming television), Contractual and billing alternatives (e.g., contract_*, payment_method_*, paperless_billing), and Financial metrics (monthly charges, total charges). The mean values for every feature within each cluster were calculated. This yielded a quantitative profile of the preferences and inclinations within each group. A heatmap (Figure 5) was employed to illustrate the feature-specific averages for each cluster, aiding in the recognition of patterns and the interpretation of behavioral segmentation.

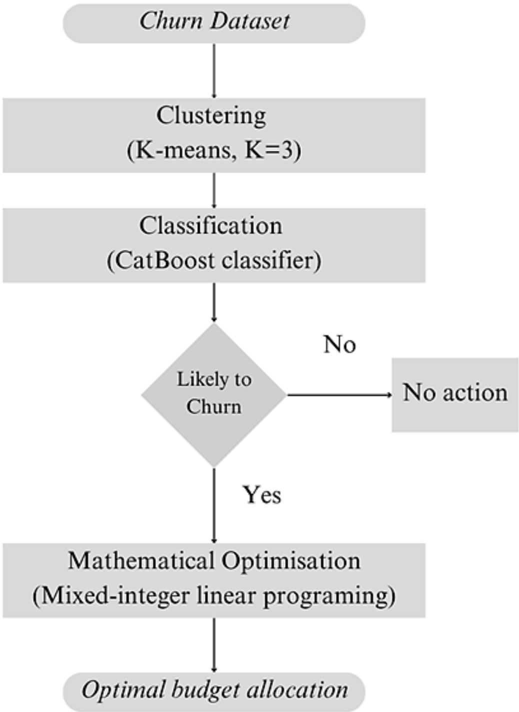


Figure 3. Flow diagram of optimal framework configuration.

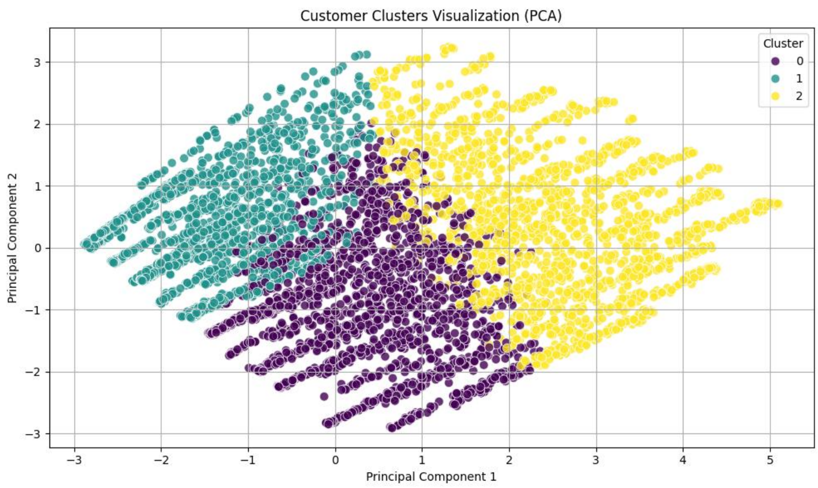


Figure 4. Visualization of K-Means clusters in 2D PCA space for K=3

Figure 5 illustrates the mean value of each feature across the three clusters, providing insights into customer preferences and behaviors for k = 3. Cluster 2, identified as the "Premium" segment, is characterised by elevated total and monthly charges, a preference for paperless billing, a predominance of automatic payment methods, and long-term contracts. This is succeeded by Cluster 1, designated as the "Standard" segment, while Cluster 0 is classified as the "Low" segment, representing the lowest priority segment.

The primary attributes utilised for clustering were total charges and tenure, denoting the overall expenditure by the customer and the duration of their subscription, respectively. The dataset was divided into separate preference clusters utilising a K-means algorithm with K set to 3. Figure 6 reveals patterns and distinctions among customer groups based on their total charges and tenure with the company, potentially guiding targeted churn retention strategies.

Figure 5. Heatmap of customer purchasing preferences by cluster for K=3

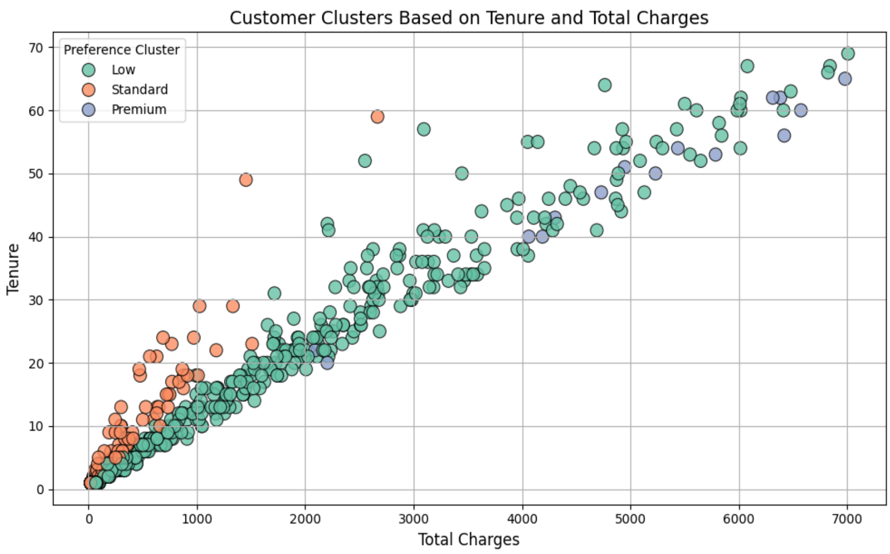


Figure 6. Customer preference clusters based on total changes and tenure for k = 3

CatBoost excels in predictive problems with a combination of numerical and categorical variables, rendering it ideal for customer churn prediction. In the k = 3 setup with CatBoost, the model identifies 593 customers as "likely to churn," offering an effective and precise prediction technique. CatBoost with k=3 demonstrated a marginal advantage, particularly in its ability to recall a greater number of positive instances, rendering it the preferable choice when the emphasis is on maximising positive captures. Figure 7 illustrates the confusion matrix of the CatBoost classifier for k = 3.

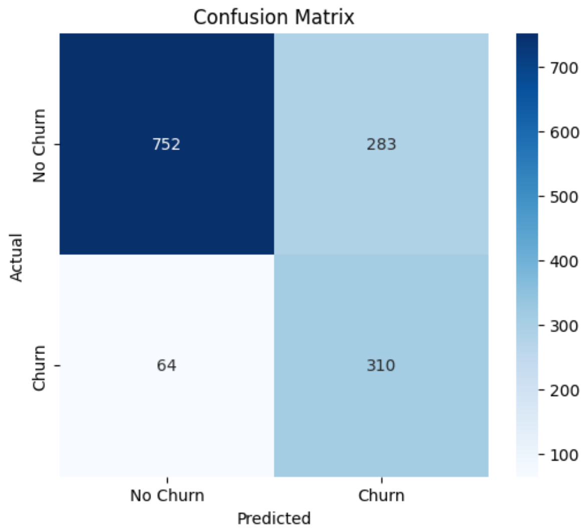


Figure 7. Confusion matrix for k=3 with CatBoost classifier.

MILP is essential for optimizing the comprehensive intervention strategy. It assists in identifying the most effective distribution of resources, including budget and intervention initiatives, while conforming to limitations such as intervention expenses and budgetary restrictions. MILP ensures efficient resource utilization by identifying the most suitable customers for action based on their likelihood of churn and potential impact. This study encompasses a comprehensive investigation of parameter sensitivity, illustrating how fluctuations in parameters affect the outcomes and showcasing the resilience of the MILP solution. Table 10 presents persuasive evidence that an optimal retention budget, sensitive to various parameter conditions, when integrated with clustering, classification, and MILP approaches, can yield tangible and strategically feasible business outcomes.

Table 10. Comparison of parameter sensitivity.

Available Retention Budget	Discount of "Total Charges"	Discount of "Monthly Charges"	Churn threshold value	Total Likely to Churn	Total eligible for Discount	Min intervention cost	Max intervention cost	Utilized Retention Budget
$25,000.00	10%	10%	0.3	593	392	$3.86	$710.71	$24,964.48
$20,000.00	20%	10%	0.4	593	151	$5.79	$1,411.47	$19,923.12
$15,000.00	15%	20%	0.3	593	142	$6.76	$1,071.03	$14,922.28
$10,000.00	20%	10%	0.35	593	104	$5.79	$1,411.47	$9,971.62
$7,500.00	15%	15%	0.5	593	87	$5.79	$1,066.06	$7,350.74

The integration of CatBoost and MILP yields efficient and comprehensible retention techniques across many budgetary and modelling scenarios. CatBoost, due to its ability to discern intricate nonlinear patterns and inherently manage categorical features, consistently recognized a greater number of potential churn customers than LightGBM. This expanded identification allowed the MILP model to distribute the intervention budget more effectively. The parameter sensitivity of MILP in optimization and constraint management is demonstrated by the accurate allocation of the budget across various scenarios. With a budget of $25,000, the model nearly depleted the allocation across all configurations, leaving negligible unused resources, thereby affirming optimal utilisation (e.g., $24,964.48 with CatBoost at K = 3).

Furthermore, as the budget diminished, MILP adeptly modified the number of qualifying customers while complying with the churn probability threshold and cost restrictions, for instance, selecting 62 customers within a $5,000 budget using a 10% discount method. The decision-making process was interpretable and aligned with real-world restrictions across all configurations. The collaboration between CatBoost's accurate churn forecasts and MILP's resource-conscious optimization confirms the methodological coherence and reinforces the efficacy of the suggested strategy in data-driven retention strategies.

To further demonstrate the validity of the proposed methodology, Table 11 presents a comparison analysis with past studies on churn retention budget optimization. This comparison highlights significant methodological distinctions, positioning this integrated approach as a competitive and comprehensive option within the existing landscape.

Table 11. Past studies on churn retention budget optimization

Study	Methodology/Algorithms
[70]	Support vector machine (SVM), fuzzy rule-based clustering, linear integer programming, and explainable AI
[104]	Profit- and AUC-focused prescriptive analytics method (PAM)
[105]	Fuzzy rule-based systems (Mamdani and Sugeno models)
This study	k-means clustering, CatBoost classifier, and Mixed-integer linear programming

5. Strengths, Limitations, and Future Work

This research demonstrates several significant strengths that augment its academic rigour and practical relevance. It presents a comprehensive, data-driven framework that adeptly combines unsupervised learning via clustering with supervised machine learning classification to enhance churn intervention strategies. A principal strength resides in its authentic integration of budget limitations and intervention expenses, which anchors the analysis in practical business contexts and augments its relevance. The study presents an optimization comparative analysis of various clustering and classification configurations, providing significant insights into the trade-offs among customer coverage, model efficacy, and cost efficiency. Furthermore, the analysis provides segment-specific insights, facilitating more targeted and strategic decision-making among various customer categories, including Low, Standard, and Premium. The research emphasizes operational impact and resource optimization by concentrating on model accuracy, customer reach, and average intervention cost. The framework and findings offer a judicious blend of analytical rigor and practical business value, constituting a significant contribution to churn management and customer retention.

Notwithstanding the encouraging outcomes derived from the analysis, several limitations must be recognized. The intervention strategy primarily relies on cluster-based segmentation, which may oversimplify the complexities and dynamics of customer behaviors over time. The clustering process, guided by unsupervised learning, may fail to represent the fundamental churn dynamics or shifting customer preferences accurately. Secondly, the estimations of intervention costs are presumed to be precise and consistent across each segment, which may not accurately represent the real-world variability in actual operational expenses. Third, the assessment is conducted on a static dataset; no temporal validation was executed to evaluate the performance of interventions over time. The analysis was restricted to classifiers, excluding optional configurations by applying different scaling, data balancing, feature selection, hyperparameter tuning, and cross-validation, thereby limiting the investigation of other potentially superior models.

Future research may mitigate these limitations by integrating dynamic and real-time churn prediction models that adjust in response to evolving customer behaviour. Investigating more detailed or novel clustering methodologies may yield more tailored intervention strategies. Moreover, incorporating reinforcement learning or optimization methodologies may facilitate the automatic adjustment of the intervention strategy in response to customer feedback and cost considerations. Future research should also explore multi-objective optimization frameworks that equilibrate cost, customer value, and long-term retention effects. The incorporation of causal inference methods to assess the true efficacy of interventions on churn reduction would strengthen the approach's robustness. Conducting tests across various datasets and industries would enhance generalizability. Future research may involve an ablation study to gain insights into the relative importance of the system's components.

6. Conclusions

This study demonstrates the effectiveness of combining clustering-based segmentation with machine learning classifiers and mathematical optimization to enhance customer churn retention budgets within specified restrictions. The research establishes an effective integrated framework for retention budget optimization by utilizing K-means clustering, the CatBoost classifier, and Mixed-Integer Linear Programming approaches. The k = 3 model, combined with CatBoost and Mixed-Integer Linear Programming methodology, exemplifies a balanced strategy by safeguarding high-value consumers while optimizing total retention across all three customer segments. This technique accurately identifies the maximum number of at-risk consumers and selects 392 individuals for intervention, resulting in a 66.1% intervention efficiency. The proposed framework demonstrates significant potential for enhancing data-driven decision-making in churn management and requires further refinement and comprehensive validation to realize its practical implications fully.

Funding: This research received no external funding.

Conflicts of Interest: The author declares no conflict of interest.

Appendix

Table A1. Performance metrics.

Metric	Description	Formula	
Accuracy	$\frac{\text{Accurate predictions}}{\text{Total number of predictions}}$	$\frac{TP + TN}{TP + TN + FP + FN}$	(1)
Precision	$\frac{\text{Correctly classified actual positives}}{\text{Everything classified as positive}}$	$\frac{TP}{TP + FP}$	(2)
Recall	$\frac{\text{Correctly classified actual positives}}{\text{All actual positives}}$	$\frac{TP}{TP + FN}$	(3)
F1-score	$\mathbf{2} * \frac{\text{Precison} \ * \ \text{Recall}}{\text{Precisison} \ + \ \text{Recall}}$	$\frac{2 \times TP}{2 \times TP + FP + FN}$	(4)

Table A2. Nature of variables.

Metric	Type	Nature	Reason
x[i]	Integer	Binary (0 or 1)	Represents whether customer i is selected for intervention.
$y[i]$	Integer	Binary (0 or 1)	Represents whether customer i's churn probability is reduced below threshold probability.
churn_reduction_var[i]:	Continuous	Continuous	Represents the amount by which customer i's churn probability is reduced.

Table A3. Components and details of MILP.

Components	Details	
Objective function	$\text{Maximize} \sum_{i=1}^{N} y[i]$	(5)
Budget constraints	$\sum_{i=1}^{N} \text{Intervention Cost}[i] \times x[i] \leq \text{Budge}$	(6)
Churn reduction constraint	$\text{Churn Reduction}[i] \leq \text{Initial Churn Probability}[i] \ \forall i$	(7)
Churn below threshold constrain	$\text{Initial Churn Probability}[i] - \text{Churn Reduction}[i] \leq threshold + (1 - y[i]) \times 1000 \ \forall i$	(8)

Components	Details	
Intervention consistency constraint	$x[i] \leq y[i] \quad \forall i$	(9)
Constraint of prioritization	$x[i] \geq x[j]$ if $\text{priority}[i] < \text{priority}[j]\ \forall i, j$	(10)

.

References

[1] S. Saleh and S. Saha, "Customer retention and churn prediction in the telecommunication industry: a case study on a Danish university," *SN Appl. Sci.*, vol. 5, no. 7, p. 173, Jul. 2023, doi: 10.1007/s42452-023-05389-6.

[2] Somya, S. Kaushik, S. Saini, S. Singhal, and S. Kumar, "Telecom Churn Prediction Using Data Science," *Educ. Adm. Theory Pract.*, vol. 30, no. 5, pp. 11026–11034, 2024, doi: 10.53555/kuey.v30i5.4880.

[3] S. Saha, C. Saha, M. M. Haque, M. G. R. Alam, and A. Talukder, "ChurnNet: Deep Learning Enhanced Customer Churn Prediction in Telecommunication Industry," *IEEE Access*, vol. 12, pp. 4471–4484, 2024, doi: 10.1109/ACCESS.2024.3349950.

[4] B. Prabadevi, R. Shalini, and B. R. Kavitha, "Customer churning analysis using machine learning algorithms," *Int. J. Intell. Networks*, vol. 4, pp. 145–154, 2023, doi: 10.1016/j.ijin.2023.05.005.

[5] M. R. Mohaimin *et al.*, "Predictive Analytics for Telecom Customer Churn: Enhancing Retention Strategies in the US Market," *J. Comput. Sci. Technol. Stud.*, vol. 7, no. 1, pp. 30–45, Jan. 2025, doi: 10.32996/jcsts.2025.7.1.3.

[6] A. Ben, "Enhanced Churn Prediction in the Telecommunication Industry," *SSRN Electron. J.*, 2020, doi: 10.2139/ssrn.3577712.

[7] R. Srinivasan, D. Rajeswari, and G. Elangovan, "Customer Churn Prediction Using Machine Learning Approaches," in *2023 International Conference on Artificial Intelligence and Knowledge Discovery in Concurrent Engineering (ICECONF)*, Jan. 2023, pp. 1–6. doi: 10.1109/ICECONF57129.2023.10083813.

[8] T. J. Shen, A. Samad, and B. Shibghatullah, "Customer Churn Prediction Model for Telecommunication Industry," *J. Adv. Artif. Life Robot.* , vol. 3, no. 2, pp. 85–91, 2022, [Online]. Available: https://alife-robotics.org/jallr.html

[9] S. Höppner, E. Stripling, B. Baesens, S. vanden Broucke, and T. Verdonck, "Profit driven decision trees for churn prediction," *Eur. J. Oper. Res.*, vol. 284, no. 3, pp. 920–933, Aug. 2020, doi: 10.1016/j.ejor.2018.11.072.

[10] G. Sam, P. Asuquo, and B. Stephen, "Customer Churn Prediction using Machine Learning Models," *J. Eng. Res. Reports*, vol. 26, no. 2, pp. 181–193, Feb. 2024, doi: 10.9734/jerr/2024/v26i21081.

[11] A. Karahoca and A. Kara, "Comparing clustering techniques for telecom churn management," in *The 5th WSEAS International Conference on Telecommunications and Informatics*, 2006, no. July, pp. 520–145. [Online]. Available: https://www.wseas.us/e-library/conferences/2006istanbul/papers/520-145.pdf

[12] R. Liu *et al.*, "An Intelligent Hybrid Scheme for Customer Churn Prediction Integrating Clustering and Classification Algorithms," *Appl. Sci.*, vol. 12, no. 18, p. 9355, Sep. 2022, doi: 10.3390/app12189355.

[13] S. Fakhar Bilal, A. Ali Almazroi, S. Bashir, F. Hassan Khan, and A. Ali Almazroi, "An ensemble based approach using a combination of clustering and classification algorithms to enhance customer churn prediction in telecom industry," *PeerJ Comput. Sci.*, vol. 8, p. e854, Feb. 2022, doi: 10.7717/peerj-cs.854.

[14] M. Ahmed, R. Seraj, and S. M. S. Islam, "The k-means Algorithm: A Comprehensive Survey and Performance Evaluation," *Electronics*, vol. 9, no. 8, p. 1295, Aug. 2020, doi: 10.3390/electronics9081295.

[15] A. K. Jain, M. N. Murty, and P. J. Flynn, "Data clustering," *ACM Comput. Surv.*, vol. 31, no. 3, pp. 264–323, Sep. 1999, doi: 10.1145/331499.331504.

[16] S. Tondini, C. Castellan, M. A. Medina, and L. Pavesi, "Automatic Initialization Methods for Photonic Components on a Silicon-Based Optical Switch," *Appl. Sci.*, vol. 9, no. 9, p. 1843, May 2019, doi: 10.3390/app9091843.

[17] M. Ahmed, "Collective Anomaly Detection Techniques for Network Traffic Analysis," *Ann. Data Sci.*, vol. 5, no. 4, pp. 497–512, Dec. 2018, doi: 10.1007/s40745-018-0149-0.

[18] M. Ahmed, "An Unsupervised Approach of Knowledge Discovery from Big Data in Social Network," *ICST Trans. Scalable Inf. Syst.*, vol. 4, no. 14, p. 153148, Sep. 2017, doi: 10.4108/eai.25-9-2017.153148.

[19] Y. Qiu, P. Chen, Z. Lin, Y. Yang, L. Zeng, and Y. Fan, "Clustering Analysis for Silent Telecom Customers Based on K-means++," in *2020 IEEE 4th Information Technology, Networking, Electronic and Automation Control Conference (ITNEC)*, Jun. 2020, pp. 1023–1027. doi: 10.1109/ITNEC48623.2020.9084976.

[20] M. Alkhayrat, M. Aljnidi, and K. Aljoumaa, "A comparative dimensionality reduction study in telecom customer segmentation using deep learning and PCA," *J. Big Data*, vol. 7, no. 1, p. 9, Dec. 2020, doi: 10.1186/s40537-020-0286-0.

[21] J. Kristian Vieri, T. A. Munandar, and D. B. Srisulistiowati, "Exclusive Clustering Technique for Customer Segmentation in National Telecommunications Companies," *Int. J. Inf. Technol. Comput. Sci. Appl.*, vol. 1, no. 1, pp. 51–57, Jan. 2023, doi: 10.58776/ijitcsa.v1i1.19.

[22] T. Velmurugan, "A State of Art Analysis of Telecommunication Data by k-Means and k-Medoids Clustering Algorithms," *J. Comput. Commun.*, vol. 06, no. 01, pp. 190–202, 2018, doi: 10.4236/jcc.2018.61019.

[23] S. W. Fujo, S. Subramanian, and M. A. Khder, "Customer Churn Prediction in Telecommunication Industry Using Deep Learning," *Inf. Sci. Lett.*, vol. 11, no. 1, pp. 185–198, Jan. 2022, doi: 10.18576/isl/110120.

[24] M. D. Gabhane, A. Suriya, and S. B. Kishor, "Churn Prediction in Telecommunication Business using CNN and ANN," *J. Posit. Sch. Psychol.*, vol. 2022, no. 4, pp. 4672–4680, 2021, [Online]. Available: http://journalppw.com

[25] A. Keramati, R. Jafari-Marandi, M. Aliannejadi, I. Ahmadian, M. Mozaffari, and U. Abbasi, "Improved churn prediction in telecommunication industry using data mining techniques," *Appl. Soft Comput.*, vol. 24, pp. 994–1012, Nov. 2014, doi: 10.1016/j.asoc.2014.08.041.

[26] H. Ribeiro, B. Barbosa, A. C. Moreira, and R. G. Rodrigues, “Determinants of churn in telecommunication services: a systematic literature review,” *Manag. Rev. Q.*, vol. 74, no. 3, pp. 1327–1364, Sep. 2024, doi: 10.1007/s11301-023-00335-7.
[27] N. Almufadi and A. Mustafa Qamar, “Deep Convolutional Neural Network Based Churn Prediction for Telecommunication Industry,” *Comput. Syst. Sci. Eng.*, vol. 43, no. 3, pp. 1255–1270, 2022, doi: 10.32604/csse.2022.025029.
[28] N. N. Y., T. Van Ly, and D. V. T. Son, “Churn prediction in telecommunication industry using kernel Support Vector Machines,” *PLoS One*, vol. 17, no. 5, p. e0267935, May 2022, doi: 10.1371/journal.pone.0267935.
[29] Y. Fareniuk, T. Zatonatska, O. Dluhopolskyi, and O. Kovalenko, “Customer churn prediction model: a case of the telecommunication market,” *Economics*, vol. 10, no. 2, pp. 109–130, Dec. 2022, doi: 10.2478/eoik-2022-0021.
[30] N. Mustafa, L. Sook Ling, and S. F. Abdul Razak, “Customer churn prediction for telecommunication industry: A Malaysian Case Study,” *F1000Research*, vol. 10, p. 1274, Dec. 2021, doi: 10.12688/f1000research.73597.1.
[31] N. I. Mohammad, S. A. Ismail, M. N. Kama, O. M. Yusop, and A. Azmi, “Customer Churn Prediction In Telecommunication Industry Using Machine Learning Classifiers,” in *Proceedings of the 3rd International Conference on Vision, Image and Signal Processing*, Aug. 2019, pp. 1–7. doi: 10.1145/3387168.3387219.
[32] A. Amin *et al.*, “Just-in-time customer churn prediction in the telecommunication sector,” *J. Supercomput.*, vol. 76, no. 6, pp. 3924–3948, Jun. 2020, doi: 10.1007/s11227-017-2149-9.
[33] D. M. Melian, A. Dumitrache, S. Stancu, and A. Nastu, “Customer Churn Prediction in Telecommunication Industry. A Data Analysis Techniques Approach,” *Postmod. Openings*, vol. 13, no. 1 Sup1, pp. 78–104, Mar. 2022, doi: 10.18662/po/13.1Sup1/415.
[34] V. Chang, K. Hall, Q. Xu, F. Amao, M. Ganatra, and V. Benson, “Prediction of Customer Churn Behavior in the Telecommunication Industry Using Machine Learning Models,” *Algorithms*, vol. 17, no. 6, p. 231, May 2024, doi: 10.3390/a17060231.
[35] C.-P. Wei and I.-T. Chiu, “Turning telecommunications call details to churn prediction: a data mining approach,” *Expert Syst. Appl.*, vol. 23, no. 2, pp. 103–112, Aug. 2002, doi: 10.1016/S0957-4174(02)00030-1.
[36] A. Amin *et al.*, “Cross-company customer churn prediction in telecommunication: A comparison of data transformation methods,” *Int. J. Inf. Manage.*, vol. 46, pp. 304–319, Jun. 2019, doi: 10.1016/j.ijinfomgt.2018.08.015.
[37] I. H. Sarker, “Machine Learning: Algorithms, Real-World Applications and Research Directions,” *SN Comput. Sci.*, vol. 2, no. 3, p. 160, May 2021, doi: 10.1007/s42979-021-00592-x.
[38] M. Badawy, N. Ramadan, and H. A. Hefny, “Healthcare predictive analytics using machine learning and deep learning techniques: a survey,” *J. Electr. Syst. Inf. Technol.*, vol. 10, no. 1, p. 40, Aug. 2023, doi: 10.1186/s43067-023-00108-y.
[39] A. Manzoor, M. Atif Qureshi, E. Kidney, and L. Longo, “A Review on Machine Learning Methods for Customer Churn Prediction and Recommendations for Business Practitioners,” *IEEE Access*, vol. 12, pp. 70434–70463, 2024, doi: 10.1109/ACCESS.2024.3402092.
[40] F. Pedregosa *et al.*, “Scikit-learn: Machine Learning in Python,” *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, 2011, doi: 10.1289/EHP4713.
[41] I. H. Sarker, “Deep Cybersecurity: A Comprehensive Overview from Neural Network and Deep Learning Perspective,” *SN Comput. Sci.*, vol. 2, no. 3, p. 154, May 2021, doi: 10.1007/s42979-021-00535-6.
[42] I. H. Sarker, “A machine learning based robust prediction model for real-life mobile phone data,” *Internet of Things*, vol. 5, pp. 180–193, Mar. 2019, doi: 10.1016/j.iot.2019.01.007.
[43] S. Le Cessie and J. C. Van Houwelingen, “Ridge Estimators in Logistic Regression,” *Appl. Stat.*, vol. 41, no. 1, p. 191, 1992, doi: 10.2307/2347628.
[44] C. Zhang, C. Liu, X. Zhang, and G. Almpanidis, “An up-to-date comparison of state-of-the-art classification algorithms,” *Expert Syst. Appl.*, vol. 82, pp. 128–150, Oct. 2017, doi: 10.1016/j.eswa.2017.04.003.
[45] Y. Wang *et al.*, “User Relationship Discovery Based on Telecom Data,” in *2023 IEEE 22nd International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom)*, Nov. 2023, pp. 2076–2081. doi: 10.1109/TrustCom60117.2023.00287.
[46] J. R. Quinlan, “Induction of decision trees,” *Mach. Learn.*, vol. 1, no. 1, pp. 81–106, Mar. 1986, doi: 10.1007/BF00116251.
[47] S. L. Salzberg, “C4.5: Programs for Machine Learning by J. Ross Quinlan. Morgan Kaufmann Publishers, Inc., 1993,” *Mach. Learn.*, vol. 16, no. 3, pp. 235–240, Sep. 1994, doi: 10.1007/BF00993309.
[48] L. Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone, *Classification And Regression Trees*. Routledge, 2017. doi: 10.1201/9781315139470.
[49] I. H. Sarker, Y. B. Abushark, and A. I. Khan, “ContextPCA: Predicting Context-Aware Smartphone Apps Usage Based On Machine Learning Techniques,” *Symmetry (Basel).*, vol. 12, no. 4, p. 499, Apr. 2020, doi: 10.3390/sym12040499.
[50] I. H. Sarker, A. Colman, J. Han, A. I. Khan, Y. B. Abushark, and K. Salah, “BehavDT: A Behavioral Decision Tree Learning to Build User-Centric Context-Aware Predictive Model,” *Mob. Networks Appl.*, vol. 25, no. 3, pp. 1151–1161, Jun. 2020, doi: 10.1007/s11036-019-01443-z.
[51] D. W. Aha, D. Kibler, and M. K. Albert, “Instance-based learning algorithms,” *Mach. Learn.*, vol. 6, no. 1, pp. 37–66, Jan. 1991, doi: 10.1007/BF00153759.
[52] F. Sanei, A. Harifi, and S. Golzari, “Improving the precision of KNN classifier using nonlinear weighting method based on the spline interpolation,” in *2017 7th International Conference on Computer and Knowledge Engineering (ICCKE)*, Oct. 2017, pp. 289–292. doi: 10.1109/ICCKE.2017.8167893.
[53] S. S. Keerthi, S. K. Shevade, C. Bhattacharyya, and K. R. K. Murthy, “Improvements to Platt’s SMO Algorithm for SVM Classifier Design,” *Neural Comput.*, vol. 13, no. 3, pp. 637–649, Mar. 2001, doi: 10.1162/089976601300014493.
[54] L. Breiman, “Random Forests,” *Mach. Learn.*, vol. 45, pp. 5–32, 2001, doi: 10.1023/A:1010933404324.
[55] Y. Amit and D. Geman, “Shape Quantization and Recognition with Randomized Trees,” *Neural Comput.*, vol. 9, no. 7, pp. 1545–1588, Oct. 1997, doi: 10.1162/neco.1997.9.7.1545.
[56] Y. Freund and R. E. Schapire, “Experiments with a New Boosting Algorithm,” in *Proceedings of the 13th International Conference on Machine Learning*, 1996, pp. 148–156. doi: 10.1.1.133.1040.

[57] P. Yang, "Data Visualization and Prediction for Telecom Customer Churn," *Highlights Sci. Eng. Technol.*, vol. 39, pp. 1080–1085, Apr. 2023, doi: 10.54097/hset.v39i.6711.
[58] S. H. Iranmanesh, M. Hamid, M. Bastan, G. Hamed Shakouri, and M. M. Nasiri, "Customer churn prediction using artificial neural network: An analytical CRM application," *Proc. Int. Conf. Ind. Eng. Oper. Manag.*, no. July, pp. 2214–2226, 2019.
[59] A. Sikri, R. Jameel, S. M. Idrees, and H. Kaur, "Enhancing customer retention in telecom industry with machine learning driven churn prediction," *Sci. Rep.*, vol. 14, no. 1, p. 13097, Jun. 2024, doi: 10.1038/s41598-024-63750-0.
[60] T. Kimura, "Customer churn prediction with hybrid resampling and ensemble learning," *J. Manag. Inf. Decis. Sci.*, vol. 25, no. 1, pp. 1–23, 2022, [Online]. Available: https://www.abacademies.org/articles/customer-churn-prediction-with-hybrid-resampling-and-ensemble-learning-13867.html
[61] I. H. Sarker, A. S. M. Kayes, and P. Watters, "Effectiveness analysis of machine learning classification models for predicting personalized context-aware smartphone usage," *J. Big Data*, vol. 6, no. 1, p. 57, Dec. 2019, doi: 10.1186/s40537-019-0219-y.
[62] G. H. John and P. Langley, "Estimating Continuous Distributions in Bayesian Classifiers," *arXiv*. Feb. 20, 2013. [Online]. Available: http://arxiv.org/abs/1302.4964
[63] J. Singh and R. Banerjee, "A Study on Single and Multi-layer Perceptron Neural Network," in *2019 3rd International Conference on Computing Methodologies and Communication (ICCMC)*, Mar. 2019, pp. 35–40. doi: 10.1109/ICCMC.2019.8819775.
[64] M. U. Tariq, M. Babar, M. Poulin, and A. S. Khattak, "Distributed model for customer churn prediction using convolutional neural network," *J. Model. Manag.*, vol. 17, no. 3, pp. 853–863, Aug. 2022, doi: 10.1108/JM2-01-2021-0032.
[65] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. London, England: MIT Press, 2016.
[66] S. Bakhvalov, E. Osadchy, I. Bogdanova, R. Shichiyakh, and E. L. Lydia, "Intelligent System for Customer Churn Prediction using Dipper Throat Optimization with Deep Learning on Telecom Industries," *Fusion Pract. Appl.*, vol. 14, no. 2, pp. 172–185, 2024, doi: 10.54216/FPA.140214.
[67] A. Achenbach and S. Spinler, "Prescriptive analytics in airline operations: Arrival time prediction and cost index optimization for short-haul flights," *Oper. Res. Perspect.*, vol. 5, pp. 265–279, 2018, doi: 10.1016/j.orp.2018.08.004.
[68] M. Mehdi Nasrabadi, "Robust Optimization for Performance-Based Budget Allocation at Payam Noor University," *Am. J. Appl. Math.*, vol. 4, no. 6, p. 310, 2016, doi: 10.11648/j.ajam.20160406.17.
[69] Y. Shi, Y. Xiang, H. Xiao, and L. Xing, "Joint optimization of budget allocation and maintenance planning of multi-facility transportation infrastructure systems," *Eur. J. Oper. Res.*, vol. 288, no. 2, pp. 382–393, Jan. 2021, doi: 10.1016/j.ejor.2020.05.050.
[70] A. Prashanthan, R. Roshan, and M. Maduranga, "RetenNet: A Deployable Machine Learning Pipeline with Explainable AI and Prescriptive Optimization for Customer Churn Management," *J. Futur. Artif. Intell. Technol.*, vol. 2, no. 2, pp. 182–201, Jun. 2025, doi: 10.62411/faith.3048-3719-110.
[71] R. N. Anderson, "'Petroleum Analytics Learning Machine' for optimizing the Internet of Things of today's digital oil field-to-refinery petroleum system," in *2017 IEEE International Conference on Big Data (Big Data)*, Dec. 2017, pp. 4542–4545. doi: 10.1109/BigData.2017.8258496.
[72] H. Harikumar, S. Rana, S. Gupta, T. Nguyen, R. Kaimal, and S. Venkatesh, "Differentially Private Prescriptive Analytics," in *2018 IEEE International Conference on Data Mining (ICDM)*, Nov. 2018, pp. 995–1000. doi: 10.1109/ICDM.2018.00124.
[73] L. Berk, D. Bertsimas, A. M. Weinstein, and J. Yan, "Prescriptive analytics for human resource planning in the professional services industry," *Eur. J. Oper. Res.*, vol. 272, no. 2, pp. 636–641, Jan. 2019, doi: 10.1016/j.ejor.2018.06.035.
[74] S. Malathi *et al.*, "A mathematical model for cost budget optimization in the early stage of house construction budget analysis," *J. Interdiscip. Math.*, vol. 25, no. 3, pp. 839–849, Apr. 2022, doi: 10.1080/09720502.2021.2017596.
[75] Y.-F. Liu *et al.*, "A Survey of Recent Advances in Optimization Methods for Wireless Communications," *IEEE J. Sel. Areas Commun.*, vol. 42, no. 11, pp. 2992–3031, Nov. 2024, doi: 10.1109/JSAC.2024.3443759.
[76] J. K. von Bischhoffshausen, M. Paatsch, M. Reuter, G. Satzger, and H. Fromm, "An Information System for Sales Team Assignments Utilizing Predictive and Prescriptive Analytics," in *2015 IEEE 17th Conference on Business Informatics*, Jul. 2015, pp. 68–76. doi: 10.1109/CBI.2015.38.
[77] M. Aref *et al.*, "Design and Implementation of the LogicBlox System," in *Proceedings of the 2015 ACM SIGMOD International Conference on Management of Data*, May 2015, pp. 1371–1382. doi: 10.1145/2723372.2742796.
[78] A. Ghoniem, A. I. Ali, M. Al-Salem, and W. Khallouli, "Prescriptive analytics for FIFA World Cup lodging capacity planning," *J. Oper. Res. Soc.*, vol. 68, no. 10, pp. 1183–1194, Oct. 2017, doi: 10.1057/s41274-016-0143-x.
[79] A. Baur, R. Klein, and C. Steinhardt, "Model-based decision support for optimal brochure pricing: applying advanced analytics in the tour operating industry," *OR Spectr.*, vol. 36, no. 3, pp. 557–584, Jul. 2014, doi: 10.1007/s00291-013-0338-3.
[80] A. F. Zakaria, S. C. J. Lim, and M. Aamir, "A pricing optimization modelling for assisted decision making in telecommunication product-service bundling," *Int. J. Inf. Manag. Data Insights*, vol. 4, no. 1, p. 100212, Apr. 2024, doi: 10.1016/j.jjimei.2024.100212.
[81] S. Ito and R. Fujimaki, "Optimization Beyond Prediction," in *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Aug. 2017, pp. 1833–1841. doi: 10.1145/3097983.3098188.
[82] T. Huang, D. Bergman, and R. Gopal, "Predictive and Prescriptive Analytics for Location Selection of Add‐on Retail Products," *Prod. Oper. Manag.*, vol. 28, no. 7, pp. 1858–1877, Jul. 2019, doi: 10.1111/poms.13018.
[83] J. R. Gardner, M. J. Kusner, Z. Xu, K. Q. Weinberger, and J. P. Cunningham, "Bayesian optimization with inequality constraints," in *Proceedings of the 31st International Conference on International Conference on Machine Learning - Volume 32*, 2014, pp. II–937–II–945.
[84] H. Harikumar, S. Rana, S. Gupta, T. Nguyen, R. Kaimal, and S. Venkatesh, "Prescriptive Analytics Through Constrained Bayesian Optimization," 2018, pp. 335–347. doi: 10.1007/978-3-319-93034-3_27.
[85] Blastchar, "Telco Customer Churn," *Kaggle.com*, 2018. https://www.kaggle.com/datasets/blastchar/telco-customer-churn (accessed May 15, 2025).
[86] C. X. Gao *et al.*, "An overview of clustering methods with guidelines for application in mental health research," *Psychiatry Res.*, vol. 327, p. 115265, Sep. 2023, doi: 10.1016/j.psychres.2023.115265.

[87] R. Agrawal, J. Gehrke, D. Gunopulos, and P. Raghavan, "Automatic subspace clustering of high dimensional data for data mining applications," in *Proceedings of the 1998 ACM SIGMOD international conference on Management of data*, Jun. 1998, pp. 94–105. doi: 10.1145/276304.276314.

[88] D. Müllner, "fastcluster : Fast Hierarchical, Agglomerative Clustering Routines for R and Python," *J. Stat. Softw.*, vol. 53, no. 9, 2013, doi: 10.18637/jss.v053.i09.

[89] M. Ester, H.-P. Kriegel, J. Sander, and X. Xu, "A density-based algorithm for discovering clusters in large spatial databases with noise," in *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining*, 1996, pp. 226–231.

[90] I. Daniel, L. Akinyemi, and O. Udekwu, "Identifying Landslide Hotspots Using Unsupervised Clustering: A Case Study," *J. Futur. Artif. Intell. Technol.*, vol. 1, no. 3, pp. 249–268, Nov. 2024, doi: 10.62411/faith.3048-3719-37.

[91] K. P. Sinaga and M.-S. Yang, "Unsupervised K-Means Clustering Algorithm," *IEEE Access*, vol. 8, pp. 80716–80727, 2020, doi: 10.1109/ACCESS.2020.2988796.

[92] J. Lwin, "Enhancing Cloud Task Scheduling with Multi-Objective Optimization Using K-Means Clustering and Dynamic Resource Allocation," *J. Comput. Theor. Appl.*, vol. 2, no. 2, pp. 202–211, Oct. 2024, doi: 10.62411/jcta.11337.

[93] K. Tabianan, S. Velu, and V. Ravi, "K-Means Clustering Approach for Intelligent Customer Segmentation Using Customer Purchase Behavior Data," *Sustainability*, vol. 14, no. 12, p. 7243, Jun. 2022, doi: 10.3390/su14127243.

[94] E. Umargono, J. E. Suseno, and S. . Vincensius Gunawan, "K-Means Clustering Optimization Using the Elbow Method and Early Centroid Determination Based on Mean and Median Formula," in *Proceedings of the 2nd International Seminar on Science and Technology (ISSTEC 2019)*, 2020. doi: 10.2991/assehr.k.201010.019.

[95] Y. Januzaj, E. Beqiri, and A. Luma, "Determining the Optimal Number of Clusters using Silhouette Score as a Data Mining Technique," *Int. J. Online Biomed. Eng.*, vol. 19, no. 04, pp. 174–182, Apr. 2023, doi: 10.3991/ijoe.v19i04.37059.

[96] M. Lamari *et al.*, "SMOTE–ENN-Based Data Sampling and Improved Dynamic Ensemble Selection for Imbalanced Medical Data Classification," 2021, pp. 37–49. doi: 10.1007/978-981-15-6048-4_4.

[97] R. E. Ako *et al.*, "Effects of Data Resampling on Predicting Customer Churn via a Comparative Tree-based Random Forest and XGBoost," *J. Comput. Theor. Appl.*, vol. 2, no. 1, pp. 86–101, Jun. 2024, doi: 10.62411/jcta.10562.

[98] I. D. Mienye and Y. Sun, "A Survey of Ensemble Learning: Concepts, Algorithms, Applications, and Prospects," *IEEE Access*, vol. 10, no. August, pp. 99129–99149, 2022, doi: 10.1109/ACCESS.2022.3207287.

[99] J. M. Ahn, J. Kim, and K. Kim, "Ensemble Machine Learning of Gradient Boosting (XGBoost, LightGBM, CatBoost) and Attention-Based CNN-LSTM for Harmful Algal Blooms Forecasting," *Toxins (Basel).*, vol. 15, no. 10, p. 608, Oct. 2023, doi: 10.3390/toxins15100608.

[100] K. Ileri, "Comparative analysis of CatBoost, LightGBM, XGBoost, RF, and DT methods optimised with PSO to estimate the number of k-barriers for intrusion detection in wireless sensor networks," *Int. J. Mach. Learn. Cybern.*, May 2025, doi: 10.1007/s13042-025-02654-5.

[101] A. Yavuz Ozalp, H. Akinci, and M. Zeybek, "Comparative Analysis of Tree-Based Ensemble Learning Algorithms for Landslide Susceptibility Mapping: A Case Study in Rize, Turkey," *Water*, vol. 15, no. 14, p. 2661, Jul. 2023, doi: 10.3390/w15142661.

[102] O. Peretz, M. Koren, and O. Koren, "Naive Bayes classifier – An ensemble procedure for recall and precision enrichment," *Eng. Appl. Artif. Intell.*, vol. 136, p. 108972, Oct. 2024, doi: 10.1016/j.engappai.2024.108972.

[103] M. A. Hearst, S. T. Dumais, E. Osuna, J. Platt, and B. Scholkopf, "Support vector machines," *IEEE Intell. Syst. their Appl.*, vol. 13, no. 4, pp. 18–28, Jul. 1998, doi: 10.1109/5254.708428.

[104] Z. Zhang, "Introduction to machine learning: k-nearest neighbors," *Ann. Transl. Med.*, vol. 4, no. 11, pp. 218–218, Jun. 2016, doi: 10.21037/atm.2016.03.37.

[105] K. M. Ting, "Confusion Matrix," in *Encyclopedia of Machine Learning*, Boston, MA: Springer US, 2011, pp. 209–209. doi: 10.1007/978-0-387-30164-8_157.

[106] F. Clautiaux and I. Ljubić, "Last fifty years of integer linear programming: A focus on recent practical advances," *Eur. J. Oper. Res.*, vol. 324, no. 3, pp. 707–731, Aug. 2025, doi: 10.1016/j.ejor.2024.11.018.

[107] Y. D. Sergeyev, D. E. Kvasov, and M. S. Mukhametzhanov, "On the efficiency of nature-inspired metaheuristics in expensive global optimization with limited budget," *Sci. Rep.*, vol. 8, no. 1, p. 453, Jan. 2018, doi: 10.1038/s41598-017-18940-4.

[108] N. S. Kumar Gandikota, M. H. Hasan, and J. Jaafar, "An adaptive metaheuristic approach for risk-budgeted portfolio optimization," *IAES Int. J. Artif. Intell.*, vol. 12, no. 1, p. 305, Mar. 2023, doi: 10.11591/ijai.v12.i1.pp305-314.

[109] R. Martín-Santamaría, M. López-Ibáñez, T. Stützle, and J. M. Colmenar, "On the automatic generation of metaheuristic algorithms for combinatorial optimization problems," *Eur. J. Oper. Res.*, vol. 318, no. 3, pp. 740–751, Nov. 2024, doi: 10.1016/j.ejor.2024.06.001.