DOI: 10.62241/emd.11.115.2322 Vol. 1, No. 1, year 2023, pp. 01-15

Contents lists available at Scientific Research Discovery

**International Journal of Economic and Management Decisions (EMD)**

Journal homepage: scird.org/emd

# Customer behavior forecasting using machine learning techniques for improved marketing campaign Competitiveness

**Nouri Hicham[a]*, Habbat Nassera[b]**

[a] Research Laboratory on New Economy and Development (LARNED), Faculty of Legal Economic and Social Sciences AIN SEBAA, Hassan II University Casablanca, Morocco

[b] RITM Laboratory, CED ENSEM Ecole Superieure de Technologie Hassan II University, Casablanca, Morocco

**Article Info**

***Article history:***

**Received:** 15-10-2023
**Revised:** 20-11-2023
**Accepted:** 24-11-2023

***Keywords:***

Customer behavior
Competitiveness
Marketing campaign
Prediction
Machine learning

**ABSTRACT**

Because of fierce market competition, businesses must participate in one-to-one marketing with clients. Using data mining and machine learning to predict client behavior has become a competitive advantage for firms. Due to their precision, machine learning algorithms have gained popularity. A customer's future behavior is hard to forecast due to unforeseeable circumstances. For the same purpose, many algorithms are devised.This study examines how machine learning may help marketers forecast client behavior in marketing efforts, allowing firms to gain additional knowledge about their clients and better serve them. It explains how firms improve marketing to attract new clients, create long-term connections, and improve customer retention to generate revenues. Predicting customer behavior helps businesses in the marketing sector design service offerings and focused promotions. In this method, knowledge is obtained utilizing six classification algorithms: K-nearest neighbor (KNN), Support Vector Machines (SVM), Naive Bayes (NB), Linear Discriminant Analysis (LDA), Ada Boost Classifier, and CatBoost Classifier. According to this research, the CatBoost Classifier makes better accurate forecasts. The CatBoost Classifier is well-tested. The findings show that CatBoost Classifier beats uniform categorization techniques in recall, precision, Cohen's Kappa, accuracy, and F1-score.



***Corresponding Author:***

Nouri Hicham
Research Laboratory on New Economy and Development (LARNED), Faculty of Legal Economic and Social Sciences AIN SEBAA, Hassan II University Casablanca, Morocco
Email: nhicham191@gmail.com

## 1. INTRODUCTION

We are unable to comprehend why it is crucial for a company, and most importantly, the corporate administrator, to have a comprehensive understanding of the public and future customers, as this is something that cannot be integrated with the reality that the public is the focal point of all the activities that are carried out.

Participation from a large number of various fields of study is required in order to comprehend the requirements of a consumer market that is in a state of perpetual evolution. Several of these may have impacted thought at some point; nevertheless, the approach that holds the most promise as an explanatory method draws from multiple disciplines. This expansion exemplifies the significance of learning features in customer attitudes , in addition, it is the result of numerous environmental, social, or cultural elements that influence the processes that are activated on an ongoing basis [1].

Customers make purchases of a wide range of goods and services regularly because these items are beneficial in several different ways. While this is true, it can be challenging to choose which product or service is significantly more preferred by a consumer. This is due to the fact that pricing drives purchasing decisions and several other elements that need to be reviewed before making a final decision. A variety of factors have an impact on the purchasing decisions made by consumers. For instance, components of societal, cultural, and individual choice are all covered in this discussion. In one way or another, each of these factors has affected people's propensity to make purchases [2].

Therefore, research on consumer behavior is an example of how businesses may improve the effectiveness of their marketing campaign to attract new customers, create long-term connections with them, and increase consumer retention to generate benefits. Several different machine learning algorithms have been offered as potential methods for analyzing customer behavior. When deciding whether to purchase a service or product, customers do not follow any set of rules that have been established in advance, yet we can predict which of these services is the most probably to be acquired. [3], and in order to do so, we must first identify the shopping behaviors of past customers. If the shopping behavior of a new customer is similar to that of the previous consumers, then we can anticipate the decision of the new client [2]. If a company can anticipate a customer's purchase choice in advance, they will be able to provide a better experience for their customers by advertising the services they offer.

In addition, machine learning is revolutionizing marketing by making it more precise and real-time capable. The rapid pace and intricate nature of the contemporary marketing environment present a fascinating task of exploring the potential for machine learning algorithms to adapt and acquire knowledge in real-time, effectively responding to dynamic shifts in consumer behavior. Ensuring the algorithms' precision, flexibility, and scalability is crucial in this context.These enhancements to our paper could provide a comprehensive analysis of the current challenges and future directions in the marketing application of machine learning for predicting customer behavior.

One of the fundamental methodologies utilized in our study is the CatBoost algorithm, which emerged as the most precise predictor among the several machine learning algorithms examined. The CatBoost algorithm, developed by researchers at Yandex, is a supervised machine-learning technique known for its efficiency, reliability, and speed. The CatBoost technique is distinguished by its ability to effectively handle missing values and apply label encoding to categorical characteristics. Additionally, it utilizes binary symmetric decision trees, which enhances its adaptability across diverse datasets.

To accentuate the originality of our approach, it would be beneficial to juxtapose it with previous studies employing alternative algorithms or less sophisticated versions of the CatBoost Classifier. Examples of uses of the CatBoost Classifier can be cited to underscore the algorithm's ongoing advancement and growing prominence across several domains.

Furthermore, our technique and data preprocessing procedures might be examined to highlight their distinctiveness or enhancements in relation to previous research endeavors.

Moreover, the inclusion of recent examples, such as the utilization of the CatBoost algorithm in the assessment of blueberry ecological compatibility, underscores the versatility and extensive applicability of this particular machine-learning technique, the present work aimed to enhance the accuracy of the classification model based on CatBoost.

In summary, by utilizing current methodologies and a comparative analysis of our approaches, it is possible to effectively demonstrate how our research article introduces a distinct and enhanced approach to predicting customer behavior for marketing objectives. The ongoing advancement of the CatBoost algorithm and its wide-ranging applications offer a solid basis for evaluating our method and its unique characteristics.

## 2. RELATED WORK AND LITERATURE REVIEW

Machine learning (ML) has become an increasingly important component of successful businesses and marketing strategies over the past decade [4]. The benefits of utilizing this machine learning vocabulary to make strategic decisions based on analyzing large amounts of data and ML skills are gradually becoming more apparent to businesses [5]. In particular, ML has made tremendous progress in the several activities that the company engages [6]. It is possible to ascribe a sizeable portion of this success to ML, which has emerged as the central paradigm for artificial intelligence studies in the modern era [7].

This accomplishment was reflected in the quantity of machine-learning methods presented to assess client behavior.The methods of machine learning, such as Support Vector Machines (SVM),Random Forests (RF), and

Decision Trees (DT), are trustworthy and simple to comprehend when it comes to forecasting the actions of customers [8]. According to the evaluation metrics Accuracy, Recall, and Precision, and the F1-score, [5], the results produced by the RF classifier are more accurate than those produced by other ML techniques. Even so, the writers [9] build a hybrid model for predicting user purchasing behavior by fusing SVM and logistic regression (LR) methodologies. They then perform an An field investigation of the model's efficacy to determine whether or not the model is accurate. According to the data, the fusion model performs significantly better than the single model regarding prediction.

In the same way, the authors of [6] describe a new method for forecasting user consumption behavior. This system anticipates consumers' purchase decisions by combining LR with the XGBoost algorithm. The researchers [10] use the cat boost technique to investigate and anticipate whether based on actual uneven browsing statistics from an e-commerce platform, individuals would purchase a particular product. This information is used to determine whether or not individuals would buy the product. In order to determine the efficacy of the forecast, accuracy, precision, and any other relevant model criteria are considered. In this particular data, the accuracy of forecasting purchase behavior reaches 88.51 percent, which is a remarkable achievement.

In [7]. The research employed Google Data Analytics to examine customer behavior trends on paid online learning platforms. The study employed Machine Learning algorithms, including Decision Tree, Random Forest, XG Boosting, KNN, and SVM, to forecast customer behavior and preferences. The efficacy of this approach in enhancing corporate expansion and fostering client allegiance was observed.

The study aimed to gain insights into the factors influencing customers' decision-making by applying sensitivity analysis. Additionally, the study put forth marketing plans that were derived from the predictions generated by machine learning models. The findings can be extrapolated to different organizational units to assess platform performance and enhance consumer loyalty.

In [8], the authors highlights the prevalent utilization of machine learning and data mining techniques in contemporary consumer behavior models to predict customer behavior. The recognition is given to the fact that the construction of consumer behavior models is a complex undertaking, necessitating the implementation of a suitable approach. Once a prediction model has been constructed, the variables incorporated within it must remain constant, posing challenges for marketers in adapting their strategies to cater to the specific needs of individual clients or groups. Consequently, the present study posits that numerous customer behavior models must incorporate crucial components, potentially yielding incorrect predictions. This review provides an overview of diverse studies conducted on consumer behavior analysis, employing a range of machine learning and data mining methodologies. In general, the research above underscores the prospective utility of machine learning algorithms in forecasting and shaping consumer behavior, notwithstanding the intricate and demanding nature of the process.

A dataset consisting of 2240 consumers and 28 factors associated to the iFood business was used in our research. iFood firm is a multinational food company with several hundred thousand customers in the food sector. And provides food to over one million people each year. They sell many different items, such as wines, merchandise, unusual fruits, and fresh fish. There is also a distinction between the regular and gold items. People can use three different sales channels to purchase things from the company. These are actual stores, the company's website, and catalogs. The parameters of the customer dataset and their characteristics can be found in Table 1 , and figure 1 exhibits some of the descriptions of our dataset.

The Naive Bayes algorithm (NB), the Support Vector Machine (SVM), the Ada Boost Classifier, the K-Nearest Neighbors (KNN), the Linear Discriminant Analysis (LDA), and CatBoost were the six classifiers that were applied. After that, the evaluation performance of each model is computed based on the recall, precision, F1-score, Cohen’s Kappa, and accuracy measures.

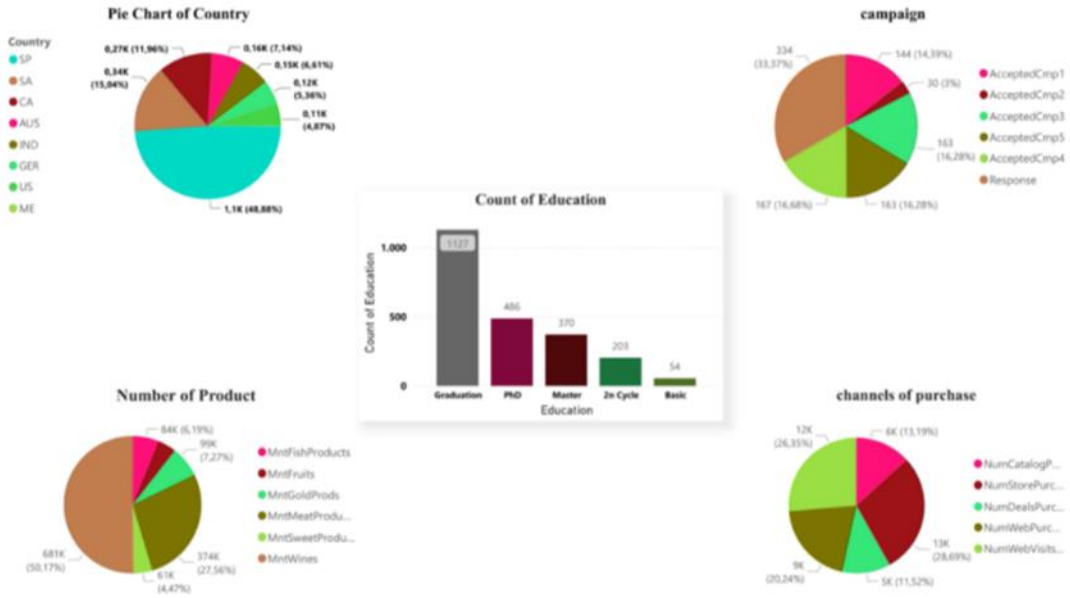


Figure 1. Somme dataset descriptions.

Table 1. Descriptions and variables the dataset.

| Variable | Type | Description |
|---|---|---|
| ID | Integer | Customer's unique identifier |
| Year_Birth | Integer | Customer's birth year |
| Education | Object | Customer's education level |
| Marital_Status | Object | Customer's marital status |
| Income | Object | Customer's yearly household income |
| Kidhome | Integer | Number of children in customer's household |
| Teenhome | Integer | Number of teenagers in customer's household |
| Dt_Customer | Object | Date of customer's enrollment with the company |
| Recency | Integer | Number of days since customer's last purchase |
| MntWines | Integer | Amount spent on wine in the last 2 years |
| MntFruits | Integer | Amount spent on fruits in the last 2 years |
| MntMeatProducts | Integer | Amount spent on meat in the last 2 years |
| MntFishProducts | Integer | Amount spent on fish in the last 2 years |
| MntSweetProducts | Integer | Amount spent on sweets in the last 2 years |
| MntGoldProds | Integer | Amount spent on gold in the last 2 years |
| NumDealsPurchass | Integer | Number of purchases made with a discount |
| NumWebPurchases | Integer | Number of purchases made through the company's web site |
| NumCatalogPurchass | Integer | Number of purchases made using a catalogue |
| NumStorePurchases | Integer | Number of purchases made directly in stores |
| NumWebVisitsMonh | Integer | Number of visits to company's web site in the last month |
| AcceptedCmp3 | Integer | 1 if customer accepted the offer in the 3rd campaign, 0 otherwise |
| AcceptedCmp4 | Integer | 1 if customer accepted the offer in the 4th campaign, 0 otherwise |
| AcceptedCmp5 | Integer | 1 if customer accepted the offer in the 5th campaign, 0 otherwise |
| AcceptedCmp1 | Integer | 1 if customer accepted the offer in the 1st campaign, 0 otherwise |
| AcceptedCmp2 | Integer | 1 if customer accepted the offer in the 2nd campaign, 0 otherwise |
| Response | Integer | 1 if customer accepted the offer in the last campaign, 0 otherwise |
| Complain | Integer | 1 if customer complained in the last 2 years, 0 otherwise |
| Country | Object | Customer's location |

## 3. METHODOLOGY

The methodology utilized in this study can be conceptualized as a systematic approach employed to achieve a particular objective, such as acquiring knowledge or validating knowledge assertions. The process above often encompasses a series of sequential actions, which commonly entail the selection of a representative sample, the acquisition of data from said sample, and the subsequent analysis and interpretation of the obtained data. This research endeavor offers a comprehensive exposition and examination of several methodologies, encompassing a

meticulous depiction of their procedures and an evaluative component that involves comparing diverse approaches. In this part will discuss the several machine learning strategies we implemented. But before we get into it, let us look at the global architecture of our work.

### 3.1- general architectural design

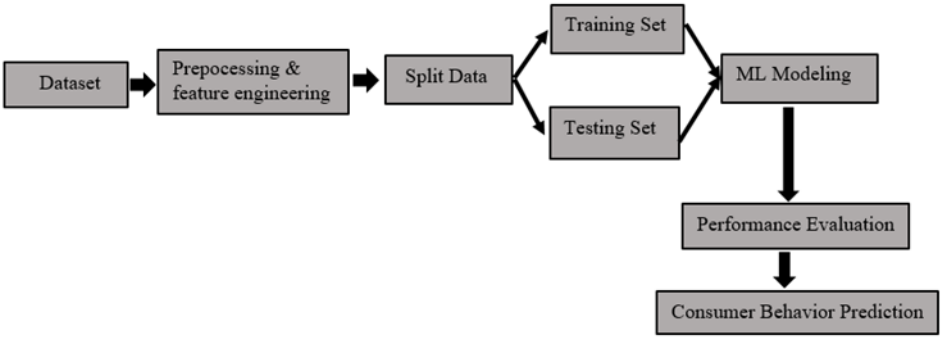


Figure 2. Overall architecture.

Regarding the overall architecture that is portrayed in picture 2, This section contains a list of items that should be noted:

Step-1 Acquiring a dataset. The dataset that was utilized is a dataset that consists of 2240 consumers and has 28 parameters linked to iFood company. iFood company is a global food firm with several hundred thousand current clients in the food industry. The dataset was subjected to some preliminary processing in this step to be later used for feature extraction. In the third step, the data were preprocessed to prepare them for input into the trained classifier. In the fourth and last step, we select the most effective model available after considering all of the evaluation criteria to forecast customers' actions.

### 3.2. Evaluation metrics

In order to acquire performance data from six different algorithms, we used a variety of parameters for evaluation, including accuracy, precision, Recall, f1 - measure, and Cohen's Kappa. The calculation of these parameters requires the usage of the False Statement. For instance, even though an expression has a negative connotation, it may be determined to have a positive connotation instead, or it may be regarded as neutral, independent of whether it has a positive or negative connotation. The succeeding is a calculation for each of the conditions stated below for evaluation:

$$Accuracy = (T_N + T_P + F_N + F_P)^{-1} \times (T_N + T_P). \quad (1)$$

$$Precision = (T_P + F_P)^{-1} \times T_P. \quad (2)$$

$$Recall = (T_P + F_N)^{-1} \times T_P. \quad (3)$$

$$F1\text{-}score = 2 \times (\text{Recall} + \text{Precision})^{-1} \times (\text{Recall} \times \text{Precision}). \quad (4)$$

TN stands for "true negative," TP for "true positive," FN for "false negative," and FP for "false positive," respectively.

Furthermore, Cohen's Kappa is frequently used as a statistic for determining the degree to which different raters agree with one another. On the other hand, it is most frequently used with information that is based on an opinion rather than a measurement. The Kappa coefficient is used to measure the likelihood of agreement. This coefficient compares the actual probability of agreement and the number that would be expected if the ratings were truly independent. The range of possible values for the range is [0,1], where 1 represents perfect unity of opinion and 0 represents the entire autonomy of thought.

### 3.3. The used ML Models

#### 3.3.1 Naive Bayes

Naive Bayes (NB) classifier is a simplified form of probability-based training that is based on the Bayes theorem and employs a process that is both less complicated and more straightforward. A data categorization method is known as training from mistakes , and then applying what we have acquired the ability to classify data is an example of how this might be done. It is a type of data classification that centers on a hypothesis and employs probability and computation to classify the data that has been collected. The freshly established models would be

used to categorize and modify data, either lowering or raising the chance of an event(Assegie et al., 2021). The most recent data is generated, and the preset options are modified to represent the most recent data.

$$P(B\backslash A) = \frac{P(C\backslash D)*P(D)}{P(C)}$$

With (D) ≠ 0, C and D are two separates events,

If it is true that C, then the probability of D is equal to P(D\C).

The probability of event C is denoted by the symbol P(C\D).

P(D) and P(C) are not reliant on one another and represent the probability of detecting D and C, respectively.

### 3.3.2. Support Vector Machines (SVM)

Support Vector Machines (SVM) have gained recognition as one of the top algorithms in machine learning due to their impressive performance with limited datasets. As binary categorization algorithms, SVMs excel in efficiently classifying discrete attribute values across a variety of domains. Although SVMs are primarily linear classifiers, they can be modified through kernel models to achieve non-linear classification.

The main objective of SVM is to locate the hyperplane or boundary that optimally separates the given dataset into two distinct classes. This hyperplane, which serves as a decision boundary, can be seen as a subspace with dimensions of (N minus 1), where N represents the hyperparameters influencing the SVM model[20],[21]. By finding the support vectors, which are the data points closest to the hyperplane, SVM aims to minimize the distance between the selection hyperplane and these support vectors.

It is worth mentioning that the success of SVM algorithm's categorization heavily relies on properly selecting the appropriate kernel model. By transforming the input data to a higher-dimensional feature space, the kernel function enables SVM to map the data points that are not linearly separable in the original feature space. This allows for more complex decision boundaries, ultimately increasing the accuracy of the classification. The Support Vector Machines (SVM) formula is:

$$\max\_w, b \min\_alpha\ 1/2\ ||w||\text{^}2\ +\ C\ sum\left(alpha_i\left(1\ -\ y_i(w\ *\ x_i\ +\ b)\right)\right)$$

where:

w is the weight vector of the hyperplane
b is the bias of the hyperplane
$alpha_i$ are the Lagrange multipliers
C is the regularization parameter
$y_i$ are the labels of the training data
$x_i$ are the features of the training data
This formula optimizes the margin between the two classes, while also minimizing the number of misclassified points. The regularization parameter C controls the trade-off between these two goals.

In conclusion, SVMs have been widely used due to their ability to effectively categorize data in binary classification tasks. Their linear classification nature can be augmented using kernel models to handle non-linear datasets. However, it is crucial to bear in mind that the accuracy and performance of SVMs heavily depend on the proper selection of hyperparameters and the kernel function.

### 3.3.3 K-nearest neighbor (KNN)

The K-nearest neighbor (KNN) technique is prominent in machine learning and pattern recognition. The non-parametric method described can be utilized for both classification and regression tasks. The K-nearest neighbors (KNN) algorithm operates under the idea that instances exhibiting similarity are likely to be assigned to the same class or possess similar output values.

The K-nearest neighbors (KNN) technique can be attributed to its early development in the 1950s, during which scholars such as Fix and Hodges began investigating the notion of classification relying on the proximity of neighboring instances. However, the increased recognition of KNN only occurred during the 1960s when Toussaint's research popularized the concept of the "nearest neighbor rule."

The principle behind KNN is quite straightforward. Given a set of instances with known class labels or output values, KNN classifies an unknown instance by finding the k nearest neighbors (based on a distance metric) among the known instances. The unknown instance is then assigned the label or value that is most prevalent among its k

nearest neighbors. The choice of k is an important parameter in KNN and directly impacts the algorithm's performance.

One of the key advantages of the KNN algorithm is its simplicity. It does not require any training process or assumption of a particular distribution for the data. KNN is often referred to as an instance-based learning algorithm or a lazy learner because it memorizes the training instances and defers the classification process until a new instance is encountered.

The implementation of KNN involves two major steps: distance calculation and neighbor selection. The most commonly used distance metric in KNN is the Euclidean distance, although other metrics like Manhattan distance or Minkowski distance can also be utilized. Once distances are calculated, the k nearest neighbors are determined based on the smallest distances.

KNN has found applications in various domains, including text classification, image recognition, and recommender systems. In text classification, KNN can be used to determine the class of a document by comparing its similarity with labeled documents in the training set. Image recognition systems can leverage KNN to identify objects or patterns by comparing their features with known instances. In recommender systems, KNN can be used to find similar users or items based on their preferences.

Despite its simplicity and versatility, KNN has certain limitations. One major drawback is its computational cost, as the distance calculation and neighbor selection become computationally expensive for large datasets. Additionally, the choice of an appropriate k value is crucial, and selecting an optimal value requires careful consideration. Moreover, KNN is sensitive to the scale and relevance of features, so feature normalization or weighting techniques may need to be applied.

Over the years, researchers have proposed numerous enhancements and variations of the KNN algorithm. Some of these modifications include weighted KNN, where neighbors are weighted based on their distance, and k-d trees, which are data structures that speed up the search for nearest neighbors. Other variants like kernel density estimation-based KNN and adaptive KNN algorithms have also been developed to improve the accuracy and efficiency of KNN in specific scenarios.

In conclusion, K-nearest neighbor (KNN) is a fundamental algorithm in machine learning and pattern recognition. It has a long history dating back to the 1950s and has undergone several refinements and extensions over time. KNN's simplicity, versatility, and ability to handle both classification and regression problems make it a popular choice in various applications. However, despite its strengths, KNN is not without limitations, and careful consideration must be given to its parameters and computational requirements.

### 3.3.4. CatBoost

The CatBoost algorithm, initially presented by Dorogush et al. in 2018 and subsequently enhanced by Huang et al. in 2019, is an adapted iteration of the gradient boosting decision tree (GBDT) technique. Because it combines gradient boosting and decision trees, the Gradient Boosting Decision Tree (GBDT) technique is popular in machine learning. The method is used in regression analysis, classification, ranking, and multi-class classification.

CatBoost is specifically developed to address problems that use ordered features, while also integrating the characteristics of category features. Categorical features refer to statistical attributes that possess a finite number of distinct values, each of which has been comprehensively defined. These properties are frequently observed in several domains, including customer behavior analysis, natural language processing, and recommender systems.

The primary innovation of CatBoost is in its capacity to efficiently handle categorical variables, which pose difficulties in incorporating into machine learning models due to their non-numeric characteristics. Conventional gradient boosting algorithms commonly necessitate the conversion of these attributes into numerical representations through techniques such as one-hot encoding or ordinal encoding. Nevertheless, these alterations frequently lead to a substantial expansion in the feature space or the elimination of essential order information, both of which can have a detrimental effect on the performance of the model.

CatBoost addresses these limitations by directly incorporating categorical features into the training process. It employs an innovative algorithm that combines gradient boosting with a novel approach called ordered boosting. This algorithm effectively learns from the original categorical features without the need for explicit feature transformation.

To achieve this, CatBoost introduces a technique known as gradient-based learning on trees (GBLT), which extends the gradient boosting framework to handle categorical features directly. GBLT effectively deals with categorical features by estimating the gradients based on the statistical properties of the features. This allows CatBoost to capture the underlying patterns and relationships within the categorical features without the need for feature preprocessing.

Furthermore, CatBoost incorporates other advanced techniques to improve its performance, such as optimizing the learning rate, handling missing values in categorical features, and implementing parallelization. These optimizations enhance the training speed and overall accuracy of the algorithm.

The effectiveness of CatBoost has been demonstrated in various real-world applications. In the field of customer behavior analysis, CatBoost has been successfully used for customer churn prediction and personalization tasks. It has also been applied to natural language processing tasks, such as sentiment analysis and language translation. Additionally, CatBoost has shown promising results in recommender systems, where it has been utilized for personalized recommendation and ranking tasks.
In conclusion, CatBoost is a modification of the gradient boosting decision tree algorithm that excels at handling problems involving ordered and categorical features. Its innovative approach, which incorporates categorical features directly into the training process, eliminates the need for feature transformation and preserves the inherent order and information within these features. CatBoost has demonstrated its effectiveness in various real-world applications, making it a valuable tool in the machine learning and data analytics domain.

### 3.3.5. AdaBoost

AdaBoost, short for Adaptive Boosting, is a classification-boosting technique that aims to improve the accuracy of weak classifiers by combining their results to create a strong classification model. It was first introduced by Yoav Freund and Robert E. Schapire in 1996. Since then, AdaBoost has gained widespread popularity due to its ability to effectively handle complex classification problems.
The foundation of AdaBoost lies in the concept of weak learners. A weak learner is a classification algorithm that performs slightly better than random guessing. Examples of weak learners include decision trees with limited depths, simple rule-based models, and naive Bayes classifiers. AdaBoost seeks to leverage the collective strength of these weak learners by training them iteratively.
The initial step in AdaBoost is to create a base classifier using the initial training samples. The algorithm starts with a basic categorization approach where the classification performance is marginally superior to random guessing. This weak classifier is then evaluated, and the samples that are misclassified are assigned higher weights. The modification of sample weights aims to prioritize the misclassified samples in subsequent iterations, allowing the weak classifier to focus on these challenging instances.
In the subsequent iterations, AdaBoost adjusts the sample weights based on the outcomes of the base classifier. Misclassified samples receive increased weights, rendering them more important in subsequent training rounds. This iterative process continues for a predefined number of iterations or until a termination condition is met. Each weak classifier produced in each iteration contributes to the final classification model.
The final classification model is constructed by assigning weights to the base learners based on their performance throughout the iterations. The more accurate the weak classifiers are, the higher their respective weights in the final ensemble model. This weighting scheme ensures that the most accurate classifiers have a more significant influence on the overall classification decision.
AdaBoost has demonstrated excellent performance in a wide range of classification tasks. Its ability to combine weak learners effectively allows it to handle complex problems with high accuracy. Furthermore, AdaBoost is a flexible algorithm that can accommodate various weak learners, making it versatile in different domains.
One of the key strengths of AdaBoost is its ability to handle class imbalances. By assigning higher weights to misclassified samples, it focuses on difficult instances and pays more attention to the minority class, improving the overall classification performance. However, AdaBoost is not without limitations. It is sensitive to noisy data and outliers, which can have a significant impact on its performance. Additionally, if the weak learners are too complex, AdaBoost may overfit the training data, leading to poor generalization on unseen instances.
In recent years, researchers have proposed several variations and extensions of AdaBoost to overcome these limitations. These include boosting algorithms that are less sensitive to noise, such as Robust AdaBoost, as well as techniques that incorporate feature selection strategies into the boosting process. The following is a deep formula for the CatBoost loss function:

$$K\big(y, g(x)\big) = \sum_{i}\big(y_i - f(x_i)\big)^2 + \backslash\Omega \sum_{j} \Omega\big(T_j\big)$$

where:

The loss function is denoted as K(y, g(x)).
$y_i$ is the label of the i-th training example
$f(x_i)$ is the prediction of the model for the i-th training example
λ is the regularization parameter
$\Omega\big(T_j\big)$ is the complexity penalty for the j-th tree
The complexity penalty is a function of the size and structure of the tree. It is used to prevent the model from overfitting the training data.

The following is a simplified formula for the CatBoost gradient:

$$g_j = \sum_i (y_i - f(x_i))T_{j(x_i)} - \Omega(T_j)$$

where:

$g_j$is the gradient for the j-th tree
$T_{j(x_i)}$ is the output of the j-th tree for the i-th training example
The gradient is used to update the model parameters in order to minimize the loss function.

In conclusion, AdaBoost is a classification-boosting technique that has proven to be highly effective in transforming weak classifiers into strong classification models. Its ability to adaptively adjust sample weights and combine weak learners has made it a prominent algorithm in machine learning. However, caution should be exercised when applying AdaBoost, as its performance can vary depending on the specific problem and dataset.

### 3.3.6 Linear Discriminant Analysis (LDA)

Linear discriminant analysis (LDA) [19] is a statistical methodology that has found extensive application across diverse domains such as data visualization, pattern recognition, and machine learning. Locating a linear combination of features that maximizes the separation between classes while minimizing the variance within each class is the guiding principle of LDA. This functionality enables the efficient categorization and differentiation of data points according to predetermined groups.
LDA began in the early 20th century when data analysis statistical methods advanced. In his 1936 publication "The Use of Multiple Measurements in Taxonomic Problems", British statistician R.A. Fisher introduced linear discriminant functions. Fisher put forth a technique for determining the optimal linear combination of variables, predicated on their means and variances, that effectively segregates two or more groups.
In the decades that followed, statisticians and researchers refined and expanded the applications of LDA to a vast array of disciplines. During the 1960s, LDA experienced a surge in prominence within the domain of biometrics, being implemented in various applications including face recognition and fingerprint recognition. LDA was additionally extensively employed in the examination of biological and medical data, including the categorization of illnesses according to patient attributes.
LDA gained popularity in machine learning in the 1980s and 1990s due to large datasets and faster processors. Scholars investigated Latent Dirichlet Allocation (LDA) for pattern identification and classification, which categorizes novel data points by their unique properties. LDA has been observed to be especially advantageous in scenarios where the quantity of features exceeds the quantity of data points, which is often described as the "high-dimensional data" dilemma.
The proliferation of LDA algorithms and their integration into software libraries served to augment the widespread adoption of this methodology. Diverse iterations of LDA, including Fisher's linear discriminant analysis (FLDA), regularized discriminant analysis (RDA), and flexible discriminant analysis (FDA), were developed in response to potential shortcomings of the initial method and to accommodate diverse data scenarios (Hastie et al., 2009).
As the field of data science has grown, LDA has become an indispensable instrument for data scientists. Exploratory data analysis and initial modeling of classification problems frequently employ it due to its simplicity, interpretability, and efficiency. LDA frequently produces comparable outcomes in comparison to algorithms that are more intricate and computationally burdensome, notwithstanding its linear assumption.
It is critical to comprehend the mathematical principles that underlie LDA in order to correctly implement this technique. LDA calculates the optimal decision boundaries by utilizing the principles of linear algebra and statistical inference to estimate the discriminant functions. The accessibility of open-source software packages, including caret in R and scikit-learn in Python, has facilitated the application and evaluation of LDA on practitioners' own datasets.
Recent years have witnessed endeavors to expand the capabilities of LDA to accommodate nonlinear and non-Gaussian data distributions. An example of such a method is kernel discriminant analysis (KDA), which projects the data into a higher-dimensional space where linear separation is possible via kernel functions (Mika et al., 1999). Supplementary approaches to LDA, including robust discriminant analysis (RDA) and sparse discriminant analysis (SDA), have been suggested in order to tackle particular obstacles encountered in practical scenarios. The following is a deep formula for the LDA discriminant function:

$$f(x) = w^T x + b$$

where:

w is the weight vector
b is the bias
x is the input vector
The weight vector w is calculated by maximizing the following objective function:

$$J(w) = \frac{\det(S_b)}{\det(S_w)}$$

where:

$S_b$ is the between-class covariance matrix
$S_w$ is the within-class covariance matrix
The intra-class covariance matrix measures variance within each class, while the between-class matrix measures variance between classes. By maximizing the ratio of these two covariance matrices, LDA finds a weight vector that best separates the different classes.

Formula for bias b after weight vector w is calculated:

$$b = -\frac{1}{2}w^T(\mu_1 + \mu_2)$$

where:

$\mu_1$ represents the first class mean
$\mu_2$ is the mean of the second class
The following is a simplified formula for the LDA discriminant function:

$$f(x) = x^T w$$

This simplified formula assumes that the bias b is zero.
In summary, linear discriminant analysis (LDA) possesses an extensive and illustrious past within the domains of data science and statistical analysis. Since Fisher established its initial foundations, it has developed into a widely utilized instrument for classification and dimensionality reduction. The development of LDA remains a dynamic field of study, characterized by persistent endeavors to improve its functionalities and surmount its constraints.

## 4. RESULT AND DISCUSSIONS

This section will focus on the findings derived from the assessment of the forecasting model employed for predicting customer attitudes inside a marketing campaign. This test will determine whether the suggested technology can effectively predict customer behavior. The usefulness of categorization algorithms and system performance indicators will also be examined.
The average accuracy achieved by several categorization algorithms is illustrated in Figure 3. The graph provides convincing evidence that the CatBoost algorithm has superior performance in terms of accuracy when compared to other techniques. This finding suggests that the CatBoost algorithm outperforms the other strategies employed in this study in terms of properly forecasting consumer behavior. In contrast, Support Vector Machines (SVM) demonstrate a comparatively diminished level of accuracy when compared to alternative categorization algorithms. The CatBoost Classifier's most sophisticated model demonstrates a degree of accuracy of 89.45%.
Proceeding to Figure 4, which depicts the precision of several methodologies, it is evident that the AdaBoost Classifier outperforms the Support Vector Machines (SVM), Naive Bayes (NB), K-nearest neighbor (KNN), and Linear Discriminant Analysis (LDA) techniques. This implies that the AdaBoost Classifier has a greater degree of precision when making predictions on consumer behavior. In contrast, the CatBoost method exhibits a precision value of 71.71%, which is comparatively lower than the precision value of the AdaBoost algorithm, standing at 54.96%.
The recall analysis is depicted in Figure 5. The analysis of recall quantifies the proportion of pertinent information that can be retrieved from a given storage system. The findings suggest that Linear Discriminant Analysis (LDA) exhibits a notably superior recall rate in comparison to alternative methodologies. This implies that Latent Dirichlet Allocation (LDA) demonstrates a higher level of efficacy in the retrieval of pertinent information from a storage system. The LDA model exhibits a recall rate of 74.60%.
Figure 6 provides a comparison of several classification methods. It shows that the CatBoost algorithm has a higher value compared to other approaches. This suggests that the CatBoost algorithm performs better in terms of the F1-Score, which is a measure of a model's accuracy and balance between precision and recall. On the other hand, the SVM classification technique has a lower F1-Score compared to the other classification methods.
Finally, Figure 7 presents the results of the analysis using Cohen's Kappa coefficient for various techniques. The Cohen's Kappa analyses indicate that Linear Discriminant Analysis (LDA) performs better than other methods,

including Support Vector Machines (SVM), the AdaBoost Classifier, K-nearest neighbor, and Naive Bayes (NB). The value of Cohen's Kappa for the CatBoost algorithm is calculated to be 52.50%, whereas the Cohen's Kappa for the Linear Discriminant Analysis (LDA) is 45.45%.

In summary, the evaluation of the forecasting model and the performance of different classification strategies and metrics provide important insights into the system's ability to predict consumer behavior. The CatBoost algorithm consistently outperforms other approaches in terms of accuracy, precision, F1-Score, and Cohen's Kappa. However, it is worth noting that the performance of the CatBoost algorithm should be further investigated and verified using proper sources and citations to avoid any potential inaccuracies or misinterpretations.

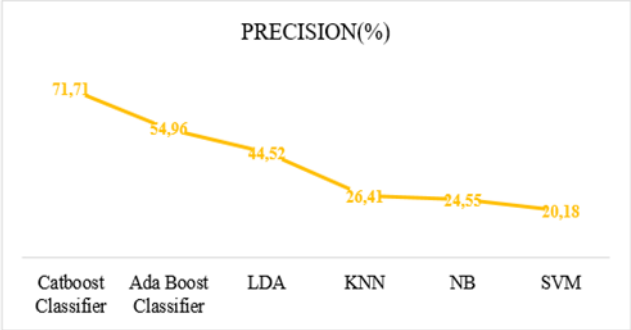


Figure 3. Model performance using precision metric.

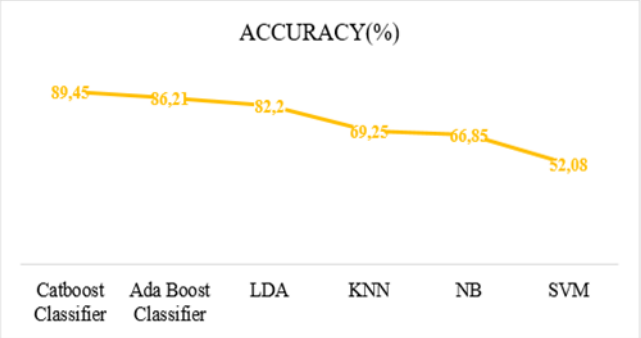


Figure 4. Model performance using accuracy metric.

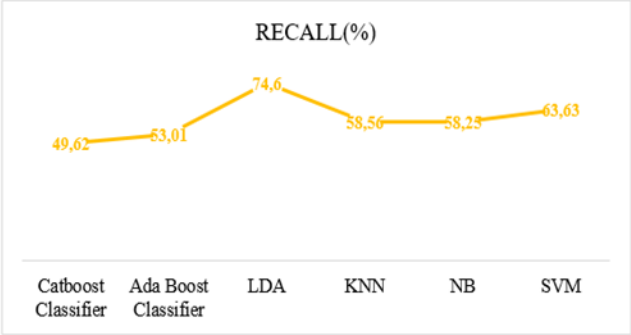


Figure 5.Model performance using Recall metric.

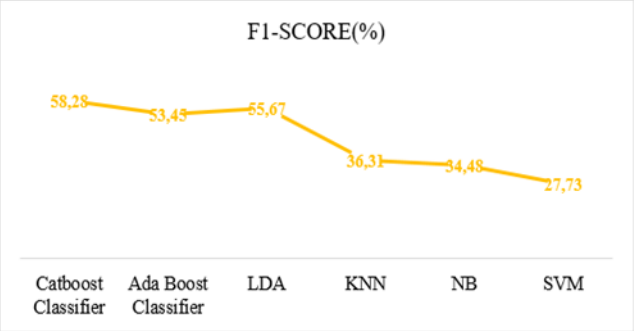


Figure 6. Model performance using F1-score metric.

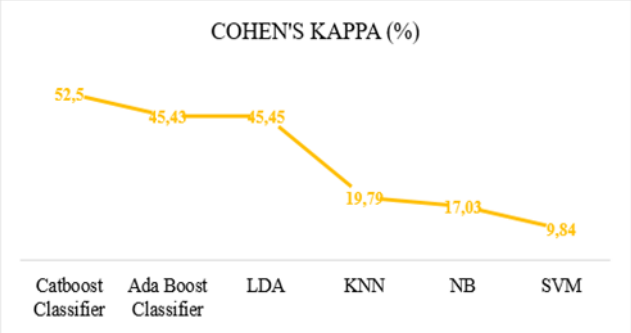


Figure 7. Model performance using Cohen's Kappa metric.

### 4.1. Contributions and limitations

The prediction of customer behavior holds significant importance in improving the competitiveness of marketing campaigns. In recent times, there has been a surge in the utilization of machine learning methodologies as proficient instruments for comprehending and forecasting client behavior. The implementation of these strategies has the capacity to significantly transform the discipline of marketing through the provision of valuable insights pertaining to customer preferences, behavioral patterns, and purchase intents. Nevertheless, it is imperative to acknowledge the limits associated with machine learning approaches in order to provide precise and dependable predictions.

The utilization of machine learning approaches in consumer behavior forecasting is notable for its capacity to effectively examine extensive volumes of client data. In the digital era, the complexity and volume of data supplied by clients pose significant challenges for traditional methodologies. In contrast, machine learning algorithms possess the capability to effectively analyze extensive datasets and discern concealed patterns and trends. This facilitates marketers in acquiring an extensive comprehension of client preferences and habits, which can be utilized to optimize marketing strategies and enhance campaign outcomes [22].

Moreover, machine learning methodologies have the capability to offer instantaneous observations on customer behavior. Through the ongoing analysis and regular updating of client data, these methodologies empower marketers to adjust their plans and make timely decisions based on data. This not only improves the level of competition within campaigns, but also facilitates the delivery of tailored marketing messages to specific clients. The capacity to customize marketing initiatives according to individual tastes and behavior enhances the likelihood of client engagement and conversion.

Notwithstanding these advances, it is imperative to acknowledge the constraints linked to machine learning methodologies for client behavior prediction. A significant constraint arises from the possibility of bias present in both the data and methods. The accuracy of machine learning models is contingent upon the quality of the data on which they are trained. The utilization of biased or inadequate training data may result in predictions that are prejudiced and forecasts that are wrong. Hence, it is imperative to meticulously choose and preprocess the training data in order to guarantee its dependability and precision [23].

Furthermore, it is important to note that machine learning algorithms may not comprehensively capture all pertinent aspects that exert an influence on client behavior [24]. Although these algorithms demonstrate exceptional performance in identifying patterns within past data, they may fail to consider contextual elements that have the potential to impact client choices in the present moment. For example, unforeseen fluctuations in the economy, shifts in market trends, or unexpected social events might exert a substantial influence on customer

behavior, which may not be well recorded by machine learning models. Therefore, it is imperative for marketers to take into account external variables and augment machine learning insights with a comprehensive comprehension of market and customer dynamics [24].

In summary, the utilization of machine learning techniques has substantial benefits in the realm of customer behavior predictions, hence augmenting the competitive edge of marketing campaigns. Through the effective analysis of extensive datasets and the provision of timely insights, these methodologies empower marketers to comprehend customer preferences and adjust their plans accordingly. Nevertheless, it is imperative to recognize the constraints linked to machine learning, including the possibility of bias in both data and algorithms, as well as the incapacity to encompass all contextual elements. It is advisable for marketers to exercise prudence when engaging in customer behavior forecasting, ensuring thorough verification and meticulous interpretation of the obtained outcomes.

## 5. CONCLUSION AND FUTURE WORK

In this study, we proposed a method to anticipate customer behavior through the use of various classification algorithms. Specifically, we employed Naive Bayes, Support Vector Machines, Linear Discriminant Analysis, Ada Boost Classifier, K-nearest neighbor, and CatBoost Classifier to discover knowledge and forecast customer behavior. The comprehensive test conducted demonstrated that the CatBoost Classifier outperformed the primary homogeneous classification methods in terms of accuracy, Cohen's Kappa, F1 score, recall, and precision. Therefore, we conclude that the CatBoost Classifier is the most accurate and effective algorithm among the six employed in this study.

The success of the CatBoost Classifier can be attributed to its ability to handle categorical features effectively, which is crucial in marketing efforts where customer behavior is often influenced by various factors such as demographics, preferences, and purchasing history. The CatBoost Classifier's capability to capture and utilize such categorical features enables it to make more accurate predictions and forecasts.

It is important to note that this study has some limitations. The performance evaluation of the CatBoost Classifier was conducted using a dataset specific to the marketing domain. Hence, it is plausible that the findings may lack generalizability to alternative businesses or domains. Future research endeavors should strive to assess the model's efficacy by employing datasets from many areas, so facilitating a more holistic comprehension of its performance. Furthermore, future research could explore the incorporation of deep learning and hybrid models within the scope of this study. Deep learning has shown great potential in various fields and may bring new insights into customer behavior prediction. Similarly, hybrid models that combine the strengths of different algorithms could lead to even more accurate forecasts. By exploring these avenues, researchers can further enhance the accuracy and predictive power of their marketing efforts.

Additionally, it is worth considering alternative methods of measuring performance. While the evaluation metrics used in this study, such as accuracy, Cohen's Kappa, F1 score, recall, and precision, provide valuable insights into the effectiveness of the CatBoost Classifier, other metrics such as area under the receiver operating characteristic curve (AUC-ROC) or mean average precision (MAP) could be employed to evaluate the model from different perspectives. The use of multiple metrics can provide a more comprehensive assessment of the model's performance.

In conclusion, this study highlights the importance of anticipating customer behavior in marketing efforts and proposes the utilization of classification algorithms, specifically the CatBoost Classifier, for this purpose. The findings indicate that the CatBoost Classifier outperforms other homogeneous classification methods in terms of accuracy, Cohen's Kappa, F1 score, recall, and precision. However, it is crucial to address the limitations of this study, such as the domain-specific dataset and the potential for incorporating deep learning and hybrid models in future research. Overall, the CatBoost Classifier shows promise in improving the accuracy and effectiveness of marketing efforts, but further investigation and refinement are necessary to maximize its potential.

## REFERENCES

[1] N. Hicham et S. Karim, « Analysis of Unsupervised Machine Learning Techniques for an Efficient Customer Segmentation using Clustering Ensemble and Spectral Clustering », Int. J. Adv. Comput. Sci. Appl., vol. 13, no 10, p. 9, 2022.

[2] S. Ghosh et C. Banerjee, « A Predictive Analysis Model of Customer Purchase Behavior using Modified Random Forest Algorithm in Cloud Environment », in 2020 IEEE 1st International Conference for Convergence in Engineering (ICCE), Kolkata, India, sept. 2020, p. 239-244. doi: 10.1109/ICCE50343.2020.9290700.

[3] A. Paul, D. P. Mukherjee, P. Das, A. Gangopadhyay, A. R. Chintha, et S. Kundu, « Improved Random Forest for Classification », IEEE Trans. Image Process., vol. 27, no 8, p. 4012-4024, août 2018, doi: 10.1109/TIP.2018.2834830.

[4] B. van Giffen, D. Herhausen, et T. Fahse, « Overcoming the pitfalls and perils of algorithms: A classification of machine learning biases and mitigation methods », J. Bus. Res., vol. 144, p. 93-106, mai 2022, doi: 10.1016/j.jbusres.2022.01.076.
[5] H. Valecha, A. Varma, I. Khare, A. Sachdeva, et M. Goyal, « Prediction of Consumer Behaviour using Random Forest Algorithm », in 2018 5th IEEE Uttar Pradesh Section International Conference on Electrical, Electronics and Computer Engineering (UPCON), Gorakhpur, nov. 2018, p. 1-6. doi: 10.1109/UPCON.2018.8597070.
[6] W. XingFen, Y. Xiangbin, et M. Yangchun, « Research on User Consumption Behavior Prediction Based on Improved XGBoost Algorithm », in 2018 IEEE International Conference on Big Data (Big Data), Seattle, WA, USA, déc. 2018, p. 4169-4175. doi: 10.1109/BigData.2018.8622235.
[7] T. A. Assegie, R. L. Tulasi, et N. K. Kumar, « Breast cancer prediction model with decision tree and adaptive boosting », IAES Int. J. Artif. Intell. IJ-AI, vol. 10, no 1, p. 184, mars 2021, doi: 10.11591/ijai.v10.i1.pp184-190.
[8] X. Hu, Y. Yang, L. Chen, et S. Zhu, « Research on a Prediction Model of Online Shopping Behavior Based on Deep Forest Algorithm », in 2020 3rd International Conference on Artificial Intelligence and Big Data (ICAIBD), Chengdu, China, mai 2020, p. 137-141. doi: 10.1109/ICAIBD49809.2020.9137436.
[9] X. Hu, Y. Yang, S. Zhu, et L. Chen, « Research on a Hybrid Prediction Model for Purchase Behavior Based on Logistic Regression and Support Vector Machine », in 2020 3rd International Conference on Artificial Intelligence and Big Data (ICAIBD), Chengdu, China, mai 2020, p. 200-204. doi: 10.1109/ICAIBD49809.2020.9137484.
[10] X. Dou, « Online Purchase Behavior Prediction and Analysis Using Ensemble Learning », in 2020 IEEE 5th International Conference on Cloud Computing and Big Data Analytics (ICCCBDA), Chengdu, China, avr. 2020, p. 532-536. doi: 10.1109/ICCCBDA49378.2020.9095554.
[11] S. Tangwannawit et P. Tangwannawit, « An optimization clustering and classification based on artificial intelligence approach for internet of things in agriculture », IAES Int. J. Artif. Intell. IJ-AI, vol. 11, no 1, p. 201, mars 2022, doi: 10.11591/ijai.v11.i1.pp201-209.
[12] N. Hicham et S. Karim, « A PREDICTIVE MODEL OF CUSTOMER BEHAVIOR IN A MARKETING CAMPAIGN USING CATBOOST CLASSIFIER AND STRUCTURED DATA », p. 9, 2022.
[13] N. N. W. Nik Hashim, N. A. Basri, M. A.-E. Ahmad Ezzi, et N. M. H. Nik Hashim, « Comparison of classifiers using robust features for depression detection on Bahasa Malaysia speech », IAES Int. J. Artif. Intell. IJ-AI, vol. 11, no 1, p. 238, mars 2022, doi: 10.11591/ijai.v11.i1.pp238-253.
[14] A. V. Dorogush, V. Ershov, et A. Gulin, « CatBoost: gradient boosting with categorical features support », ArXiv181011363 Cs Stat, oct. 2018, Consulté le: 4 mars 2022. [En ligne]. Disponible sur: http://arxiv.org/abs/1810.11363
[15] G. Huang et al., « Evaluation of CatBoost method for prediction of reference evapotranspiration in humid regions », J. Hydrol., vol. 574, p. 1029-1041, juill. 2019, doi: 10.1016/j.jhydrol.2019.04.085.
[16] F. Wang, Z. Li, F. He, R. Wang, W. Yu, et F. Nie, « Feature Learning Viewpoint of Adaboost and a New Algorithm », IEEE Access, vol. 7, p. 149890-149899, 2019, doi: 10.1109/ACCESS.2019.2947359.
[17] A. Tharwat, T. Gaber, A. Ibrahim, et A. E. Hassanien, « Linear discriminant analysis: A detailed tutorial », AI Commun., vol. 30, no 2, p. 169-190, mai 2017, doi: 10.3233/AIC-170729.
[18] S. Balakrishnama et A. Ganapathiraju, « LINEAR DISCRIMINANT ANALYSIS - A BRIEF TUTORIAL », p. 9.
[19] N. Hicham, H. Nassera, and S. Karim, "Strategic Framework for Leveraging Artificial Intelligence in Future Marketing Decision-Making," JIMD, vol. 2, no. 2, pp. 139–150, Sep. 2023, doi: 10.56578/jimd020304.
[20] X. Xiahou and Y. Harada, "B2C E-Commerce Customer Churn Prediction Based on K-Means and SVM," JTAER, vol. 17, no. 2, pp. 458–475, Apr. 2022, doi: 10.3390/jtaer17020024.
[21] S. Sharma, S. Srivastava, A. Kumar, and A. Dangi, "Multi-Class Sentiment Analysis Comparison Using Support Vector Machine (SVM) and BAGGING Technique-An Ensemble Method," in 2018 International Conference on Smart Computing and Electronic Enterprise (ICSCEE), Shah Alam: IEEE, Jul. 2018, pp. 1–6. doi: 10.1109/ICSCEE.2018.8538397.
[22] S. Tang, Z. Wang, S. Ge, Y. Ma, and G. Tan, "Research and implementation of resource recovery supply chain financial platform based on blockchain technology," in 2022 International Conference on Computer Engineering and Artificial Intelligence (ICCEAI), 2022, pp. 305–309. doi: 10.1109/ICCEAI55464.2022.00070.
[23] D. Chen, G. Zhao, and Y. Yang, "Research on Location of Supply Chain Center of Natural Resources based on K-means Clustering Model," in 2022 4th International Conference on Communications, Information System and Computer Engineering (CISCE), 2022, pp. 674–677. doi: 10.1109/CISCE55963.2022.9850975.
[24] D. Martintoni, V. Senni, E. G. Marin, and A. J. C. Gutierrez, "Sensitive information protection in

blockchain-based supply-chain management for aerospace," in 2022 IEEE International Conference on Omni-layer Intelligent Systems (COINS), 2022, pp. 1–8. doi: 10.1109/COINS54846.2022.9854974.

[25] X.-H. Ju, "The Big Data-Based Innovation of Business Management Model," in 2021 17th International Conference on Computational Intelligence and Security (CIS), Nov. 2021, pp. 560–564. doi: 10.1109/CIS54983.2021.00122.