Journal of Theoretical and Applied Information Technology
31st March 2024. Vol.102. No 6

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

RESEARCH INTELLIGENT PRECISION MARKETING OF INSURANCE BASED ON EXPLAINABLE MACHINE LEARNING: A CASE STUDY OF AN INSURANCE COMPANY

NOUHAILA EL KOUFI[1], ABDESSAMAD BELANGOUR[1]

[1]Laboratory of Information Technology and Modeling (LTIM), Faculty of Sciences Ben M'sik, Hassan II University in Casablanca, Morocco

Corresponding author E-mail: [1]elkoufi.nouhaila1@gmail.com

ABSTRACT

Today, being a marketer is not an easy task, as it requires guiding relevant interactions with customers and driving business success. This is particularly challenging in the realm of traditional marketing. Over the past few years, marketers have observed that they are spending a significant amount of money on advertising their brands or services without any assurance of a response from the customers who receive their direct mail. This lack of knowledge about their audience makes it difficult to identify the interactors from the non-interactors, leaving marketers feeling like they are marketing blindly. They operate without knowing if they are reaching the right audience at the right time, which further complicates the issue and prolongs the process of creating engagements and building an audience for their brands or services. The primary goal of any marketer is to reduce costs and increase revenues. With the widespread digitalization of services and communication technology in different domains, like the insurance sector, online platforms are producing a huge amount of data every day about customer behaviors. Thanks to this source of information, and driven by new challenges in the market, realizing a more accurate and intelligent marketing approach becomes an increasing necessity among researchers and companies. This study presents an intelligent system based on the combination of advanced features engineering approaches and machine learning techniques. The aim of the suggested precision-making system is to assist managers in discerning customer categories based on potential characteristics. Firstly, a comprehensive customer persona was developed by extracting a range of data features, including basic attributes and consumption attributes. Then, we evaluated the effectiveness of various algorithms, such as CatBoost, XGBoost, random forest (RF), k-nearest neighbor (K-NN), nave Bayes (NB), and support vector machine (SVM) methods, for predicting the response of existing customers to the next offer. Various feature selection techniques were employed to determine the most significant features. Furthermore, the performance of the models used was assessed and compared. The results showed that CatBoost had higher accuracy, kappa, precision, Fmeasure and AUC values of 0.871, 0.711, 0.94, 0.822, and 0.85, respectively, outperforming the other models. To illustrate the advantages of our proposed precision-making system, we used a real-world dataset from an American insurance company as a case study.

Keywords: *Precision Marketing, Machine Learning, Features Engineering, Big Data Analysis, Customer Persona, Decision-making System*

1. INTRODUCTION

The insurance sector plays a crucial role in the global economy by offering financial protection to both individuals and businesses against various risks and uncertainties. Insurance companies generate revenue by charging premiums to assume the risk of potential losses, including property damage, accidents, and illness. The insurance sector is divided into two primary categories: life insurance, which covers risks related to human life, and non-life insurance, which includes coverage for property damage and liability. This paper focuses on non-life insurance products, specifically car insurance.

The insurance industry is currently facing three significant challenges: embracing advanced digital applications, integrating AI and big data technology, and improving their marketing strategies. These challenges arise due to the rapid pace of technological development and the surge in the number of clients. Insurance companies have started to recognize the importance of digitalization and marketing in their operations. The utilization of digital technologies such as mobile apps and online portals can enable insurance companies to offer more convenient and efficient services to their customers. Digitalization can also help them collect and analyze customer data, which can be utilized to enhance marketing strategies and provide personalized offers. By adopting innovative marketing techniques, insurance companies can expand their reach and distinguish themselves from their competitors. In today's digital era, insurance companies that prioritize digitalization and marketing are better positioned to succeed and stay ahead of the curve, resulting in a significant impact on their overall business performance.

Understanding customer loyalty is crucial for any business, and this is especially true for insurance companies. Insurance customers are often making a long-term commitment to a provider, and customer loyalty can make all the difference when it comes to retaining clients and building a successful business. However, achieving customer loyalty requires more than just offering competitive prices or good customer service. Precision marketing is key to understanding the unique needs and preferences of each customer, and tailoring your products and services to meet those needs [1]. By utilizing customer data and analytics, insurance companies can gain insight into the factors that drive customer loyalty, and develop targeted marketing strategies to build long-term relationships with their clients. Ultimately, the success of an insurance company depends on its ability to understand and meet the needs of its customers, and precision marketing is an essential tool for achieving this goal.

The aim of this study is to investigate how the combination of machine learning methods and feature engineering can be used to analyze customer data and create accurate and multidimensional user portraits. Furthermore, it aims to explore how the precise customer persona developed through this process can serve as a robust foundation for building a precision marketing model. In order to reach this goal, this study interested in the following points: (1) Proposing a precision-making system to assist managers in discerning customer categories based on potential characteristics. (2) Building a multidimensional user portrait based on the extraction of a set of important features. (3) Proposing and applying a predictive model based on machine learning and feature engineering in a real-life scenario to assist an insurance company in forecasting loyal customers who are likely to renew their insurance policy. The objective of this proposed framework is to track and analyze customer behavior and develop an appropriate precision marketing scheme.

Amidst the escalating competition in the market, enterprises face the pressing challenge of adopting effective strategies to achieve precision marketing. To stay competitive and guarantee long-term development, companies are increasingly required to implement precision marketing model. Analyzing and investigating clients' behavior is a long-standing issue that has drawn the interest of researchers and scholars in the business sector. Every company aims to retain their customers for an extended period of time and stay competitive [2]. With the emergence of the big data age and the advancement of artificial intelligence techniques, it has become possible to track customer behavior using these techniques and provide customers with appropriate and precise marketing strategies. In [3], Zhang et al. proposed a predictive model based on the combination of logistic regression and neural network to predict potential luxury car buyers. The researchers validated their proposed model using a real-world dataset consisting of information on both telecom users and automobile proprietors. This data was obtained from telecom operators and the traffic management department in China. In [4], researchers proposed an accurate marketing optimization scheme by combining fuzzy methods and neural network modeling. They also introduced a Logistics Warehousing Center model as a solution to address the problems faced by the distribution system of e-commerce logistics. To address the issue of poor correlation between data models and spatial redundancy in initial marketing data, Su Ying Liu proposed an accurate precision marketing decision system in [5]. This system is based on spatio-temporal data, the k-means method, and neural network modeling. In [6], Zhang et al. proposed a precision marketing framework that combines machine learning methods, including K-Nearest-Neighbor (K-NN), support vector machine (SVM), and on-line learning programming, to optimize resource allocation. To validate their proposed system, a case study was conducted on a loan agency in China, where data on customers who are small business proprietors was collected. The comparison of selected classifiers revealed that K-NN achieved good results with an accuracy of 99.1%. In order to analyze and extract user characteristics based on their purchase history, Li et al. [7] employed machine learning techniques, namely decision trees, cluster analysis, and naive Bayes. The study found that decision trees outperformed clustering analysis and naive Bayesian algorithms in terms of prediction

Journal of Theoretical and Applied Information Technology
31st March 2024. Vol.102. No 6

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

accuracy, promotion degree, and precision. In the work of Chiu et al. [8], an omni-channel chatbot was introduced that offers customized services and targeted marketing with high accuracy based on convolutional neural networks. A case study of a Chinese shared kitchen was conducted to demonstrate how to implement the proposed chatbot. The results showed that the proposed solution is effective. In [9], Ze Gao conducted an analysis of sales data for agricultural products from a Chinese e-commerce platform in order to achieve precision marketing for this sector. He proposed an improved k-nearest neighbor (K-NN) algorithm for classifying users based on their personal information. The prediction accuracy of the K-NN algorithm was measured at varying K thresholds, and the results showed that it provided good results when the size of K was chosen to be 10. Hongping Liu [10] has adopted neural network modeling to address issues such as unscientific algorithms and data pollution. The study focuses on user churn prediction and value enhancement.

The rest of the paper organized in the following way: In Section 2, we present, descript, and discuss the proposed system. Section 3 includes the findings, and the experimentation phase, and a discussion regarding the results of the implementation of the proposed system. Finally in Section 4, we present the conclusion and the future direction of this proposed study

2. MATERIALS AND METHOD

2.1 System Architecture

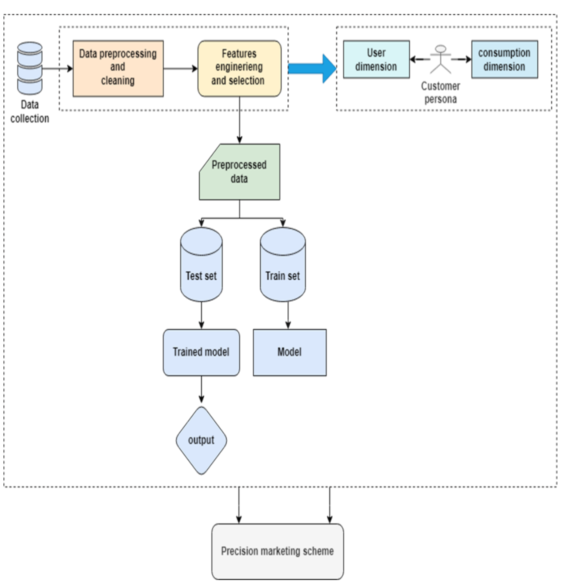

Figure 1: Proposed Framework Architecture

Figure 1 provides a graphical representation of the proposed model architecture. The proposed model in this work primarily involves four steps, which are: (1) Data collection, (2) Data preprocessing, (3) Oversampling, (4) Customer persona analysis, (5) Modeling.

2.2 Description of the Proposed System

2.2.1 Data Collection

In this study, we utilized real-world data from an American insurance company, which included 9416 records and 26 variables. Each row within the dataset contains valuable personal information about an individual customer, including basic details such as gender, age, education level, work status, income, urbanity of residence, and marital status, as well as information on customer behavior such as well as information on customer behavior coverage type, premium, months since last sinister, date of status contract, type of intermediary, policy type, contract level, sinister open, date of effect, vehicle category, and vehicle size. The attributes of the selected data are presented in Table 1.

Table 1: Attributes Description.

Attributes	Possible Values
Policy number	-
Gender	Female Male
Marital status	Married Divorced Single
Age	-
Education level	High school or below Bachelor degree Master degree PhD College
Work status	Employed Retired Unemployed Medical leave Infirm
Income (I)	<14283 14283<= I < 28566 28566 <= I < 42849 42849<= I < 57132 57132<= I < 71415 71415 <= I < 85698 85698 <= I < 99980
Urbanity of residence	Urban Semi urban Rural
CLV	-
Policy status	E R
Branch	Auto insurance
Coverage type	Basic

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

	Extended Premium
Premium (P)	<146 146 <= P < 192 192 <= P < 238 238 <= P < 284
Months since last Sinister	0 <=LS<=35
Date of Status Contract	-
Type of intermediary	Online from the Website Agent Branch Call center
Intermediary code	-
Policy Type	Personal Corporate Special
Contract level	Level 1 Level 2 Level 3
Sinister open (SO)	0<=SO<=5
Cause of sinister	-
Category sinister	-
Date of effect	-
Number of contracts NC	1<=NC<=9
Vehicle category	Normal car Sport utility vehicle (SUV) Luxury car Luxury sport utility vehicle (LSUV)
Auto size	Large Midsize Small
Regalement	-
Reserve	-
S/P features	-
Sinister number amount (SA)	312<=SA<=997
Renewal propriety	Yes No

2.2.2 Data Preprocessing

Preprocessing data is a critical step that encompasses all the necessary actions taken prior to the modeling phase. It involves cleaning, transforming, and preparing raw data to make it suitable for training a machine learning model. This process is essential because the quality of data used for training directly affects the accuracy and performance of the model. The data preprocessing phase includes various techniques, such as handling missing values, removing outliers, scaling or normalizing data, and selecting or extracting features. These techniques ensure that the data used for training is accurate, complete, and appropriate for the intended modeling task. Effective data preprocessing is crucial for achieving optimal results in the modeling phase. By investing time and effort into data preprocessing, the quality and reliability of the ML model can be improved, leading to better outcomes and insights.

Irrelevancy: Data that is irrelevant or unnecessary can have a negative impact on the performance of the model. Therefore, it is important to identify and remove such attributes to enhance the accuracy of prediction. After a thorough analysis and understanding of the provided data, we have excluded the irrelevant features, which include intermediary code, policy number, and branch. The data transformation is also considered.

Transformation: Transforming data is the process of converting data from one form to another. Validating and structuring data correctly can improve its quality and protect applications from issues like null values, duplicate entries, incorrect indexing, and incompatible formats. In this study, we have performed the following data transformations: Gender (Female ->0, male ->1), marital status (married ->0, divorced ->1, single ->2), age ([18, 25]− > 0, [26, 30]− > 1, [31, 40]− > 2, [41, 50]− > 3, [51, 60]− > 4, [61, 79] − − > 5), education level (high school or below − > 0, bachelor degree − > 1, master degree − > 2, PhD − > 3, college− > 4), work status (employed − > 0, retired − > 1, unemployed − > 2, medical leave − > 3, infirm− > 4), income ([0, 14282]− > 1, [14283, 28565]− > 2, [28566, 42848]− > 3, [42849, 57131]− > 4, [57132, 71414]− > 5, [71415, 85697]− > 6, [85698, 99980]− > 7), urbanity of residence (urban− > 0, semi urban − > 1, rural− > 2), policy status (E− > 0,R− > 1), coverage type (basic->0, extended ->1, premium->2), premium ([0,145]- >0, [146,191]- >1, [192,237]- >2, [238,284]- >3), type of intermediary (website->0,agent->1,branch->2,call center->3), policy type (personal->0, corporate->1, special->2), contract level (level 1->0, level 2->1, level 3->2), vehicle category (normal car ->0, sport utility vehicle->1, luxury car ->2, LSUV ->3), auto size (large ->0, midsize->1, small->2), renewal propriety (Yes->1, No->0). Furthermore, we removed features with a significant number of missing values, such as date of status contract, cause of sinister, category sinister. After excluding these features, the proportion of missing values in the data decreased.

Unbalanced data: In other to classify the provided data, the target column was divided into two classes

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

(1, 0), with 1 representing customers interested in renewing their insurance and 0 representing those who are not. The distribution of the two classes for the output column is shown in Figure 2, which demonstrates that the data is imbalanced, with a non-uniform distribution of classes.

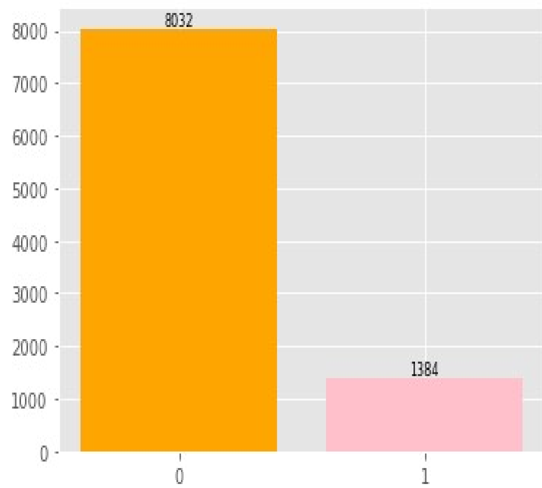

Figure 2: The distribution of the two classes for the output column

As depicted in Figure 2, class 0 has a significantly higher number of observations (8032), accounting for 87% of the total, while class 1 has only a small proportion of observations, making up just 14,69% (1384). In other words, there are fewer customers interested in renewing their insurance than those who are not. The issue of imbalanced data is commonly encountered when working with real-world data. Training a model with an imbalanced data can lead to an imbalanced training set, causing overfitting, and negatively affecting the accuracy of classification results. Therefore, we have decided to employ random oversampling to increase the representation of minority classes. To perform the oversampling, we utilized the ROSE function from the ROSE library, a package available in the R environment, which allowed us to balance the provided data

2.2.3 Features Selection

After cleaning the data, it is time to choose the most relevant features suitable for model building. The selection of the most important features plays a critical role in improving the model performance, reducing the dimensionality of the data and training time. In this study, we utilized the Pearson correlation coefficient (PCC) to select the most useful features. The correlation between input variables is calculated as follows (See Eq 1):

$$\rho_{X,Y} = \frac{cov(X,Y)}{\sigma_X \sigma_Y} = \frac{E(XY)-E(X)E(Y)}{\sqrt{E(X^2)-E^2(X)}\sqrt{E(Y^2)-E^2(Y)}} \quad (1)$$

According to results founded, we notice that reserve, regalement, S/P features demonstrated no correlation with the target variable. Thus we drop this features. To identify the hyper-parameters that would yield the best performance, we utilized the Gridsearch method.

2.2.4 Customer Persona Analysis

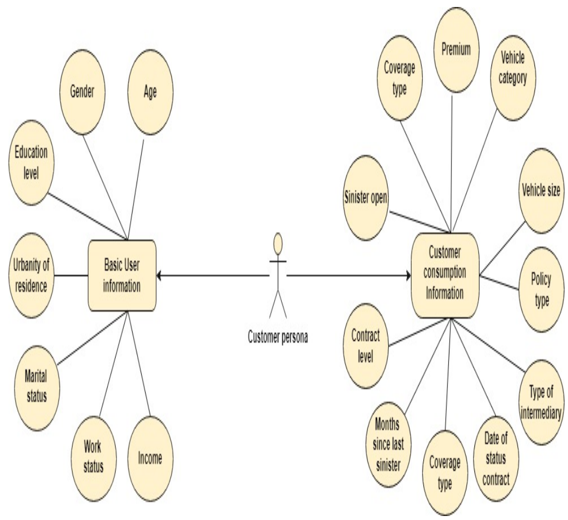

Figure 3: Customer Persona

A Customer Persona serves as a means to outline the characteristics of a customer and depict how they make purchases. In simpler terms, it's a made-up portrayal that represents a real customer. By gathering and examining pertinent data about the individual, such as spending habits, demand situations, customer actions, general traits, and other aspects, we can uncover valuable insights that mirror the entirety of user information. This technique is a fundamental tool for tailored recommendations, data-guided operations, data analysis, and focused marketing. Moreover, customer personas empower us to grasp the buying requirements of every section of our audience and adapt our approach accordingly. This, in turn, aids in constructing a more captivating brand experience and forging stronger, lasting connections with our clientele. In this study, customer portrait is conducted based on two dimension: demographic (basic) information and customer behavior information. Basic information provides basic social information about users, while customer behavior information covers consumption behavior.

Together, they allow for the development of a concise user label that can be used to represent and categorize consumers. Figure 3 represents an example of a user persona.

2.2.5 Modeling

After the data has been prepared, oversampled, and split into training and testing sets, it is important to select suitable models for training. As this is a classification problem, appropriate classifiers must be chosen for conducting the classification. In order to build the proposed predictive model, a variety of classification algorithms were employed, including CatBoost, XGBoost, Random Forest (RF), KNearest Neighbors, and Support Vector Machine (SVM), and Nave bays (NB). Following the selecting of the appropriate models for training, the training data is utilized to train the models using the relevant features that have been selected through feature selection techniques. This ensures that the trained models are optimized and capable of making accurate predictions.

2.3 Performance metrics

In classification problems in machine learning, the choice to compare performance using metrics such as accuracy, AUC (Area Under the ROC Curve), kappa, recall, f-score, and precision stems from the need for a comprehensive evaluation that considers different aspects of the model's predictive capability. Accuracy provides an overall measure of correct predictions, AUC evaluates the model's ability to discriminate between classes, kappa accounts for agreement beyond chance, recall emphasizes the ability to capture true positives, precision focuses on the accuracy of positive predictions, and the F-score strikes a balance between precision and recall. Together, these metrics offer a multifaceted view of the model's performance, accounting for factors like class imbalance, false positives, false negatives, and overall predictive accuracy, enabling a more informed assessment of its effectiveness in various aspects of classification.

2.3.1 Precision

The precision metric is utilized to evaluate the accuracy of the classification of a class and its correctness in being assigned to the appropriate category. The Eq. (2) is used to calculate the precision metric.

$$Precision = \frac{TP}{TP+FP} \tag{2}$$

2.3.2 Accuracy

The accuracy metric is used to measure how well a classifier model predicts the outcomes of a given dataset. To calculate the accuracy metric, Eq. (3) is used.

$$Accuracy = \frac{TP+TN}{(TP+FP+TN+FN)} \tag{3}$$

2.3.3 Recall

The recall metric is used to evaluate the accuracy of correctly identifying the positive class fraction, which indicates how effectively the model can identify a specific class. To calculate the recall metric, Eq. (4) is used

$$Recall = \frac{TP}{TP+FN} \tag{4}$$

2.3.4 F-measure

The F-measure is an evaluation metric that combines precision and recall and is also known as the F-score. The following equation (Eq. (5)) is used to calculate the F-measure metric.

$$F-measure = \frac{(2 \times Precision \times Recall)}{(Recall+Precision)} \tag{5}$$

2.3.5 Kappa

Kappa statistics, also known as Cohen's Kappa, is a crucial measure that goes beyond accuracy by taking into account the probability of correct predictions in both classes. This measure is especially relevant for datasets with imbalanced classes. To calculate the recall metric, Eq. (6)) is used

$$Kappa = \frac{Pr_0 - Pr_e}{1+Pr_e} \tag{6}$$

Pr_0: represents the degree of agreement observed among raters.

Pr_e: signifies the hypothetical probability of chance agreement.

2.3.6 Confusion matrix

The calculation of each selected metric depend on information extracted using the confusion matrix [1]. A confusion matrix is important method for determining the accuracy of predicted class outputs. Tabl. 2 displays the confusion matrix, where the rows indicate the predicted class and the columns indicate the actual class. The values TP and TN correspond to the number of correctly predicted positive and negative instances, respectively, while FP and FN correspond to the number of incorrectly predicted positive and negative instances, respectively

Table 2: Confusion Matrix.

	Actual positive	Actual negative
Forecasted positive	TP	FN
Forecasted negative	FP	TN

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

2.3.7 AUC curve metric

The performance of the model on the positive and negative classes of the test set has been measured using the AUC curve. A higher score indicates better performance. The AUC metric is preferred over accuracy, precision, and recall as they are not robust to changes in class distribution. The AUC metric is a widely used ranking evaluation technique, also known as ROC or the global classifier performance metric, as it can measure different classification schemes to compare overall performance. The other metrics might not perform well if the test set changes its distribution of positive and negative instances. However, the ROC curve is insensitive to changes in the proportion of positive and negative instances and class distribution.

3. RESULTS AND DISCUSSION

In this study, we utilized a real-world dataset collected from an American insurance company. This dataset contains comprehensive information about customer consumption habits, basic demographics, and more. This data was employed to validate the proposed decision-making framework, which is elaborated in detail in Section 2. The dataset comprises 9416 records and 29 variables. After collecting, preprocessing, and oversampling the data, as well as constructing customer personas, we proceeded to evaluate performance by comparing six predictive models. This section presents the results obtained from the experimental analysis. Specifically, Table 3, Figure 4, and Figure 5 showcase the performance comparison of the six machine learning methods we selected: CatBoost, XGBoost, K-NN, random forest, naïve Bayes, and SVM. The choice of supervised learning was appropriate given that we possess labeled data consisting of pre-existing records with corresponding target values. These target values can lead to two potential outcomes: 'yes' or 'no,' representing customer responses.
The assessment of method performance involved various metrics such as Accuracy, Precision, Recall, F-measure, AUC score, Kappa, and the confusion matrix. These metrics collectively contribute to a comprehensive evaluation of the selected methods.

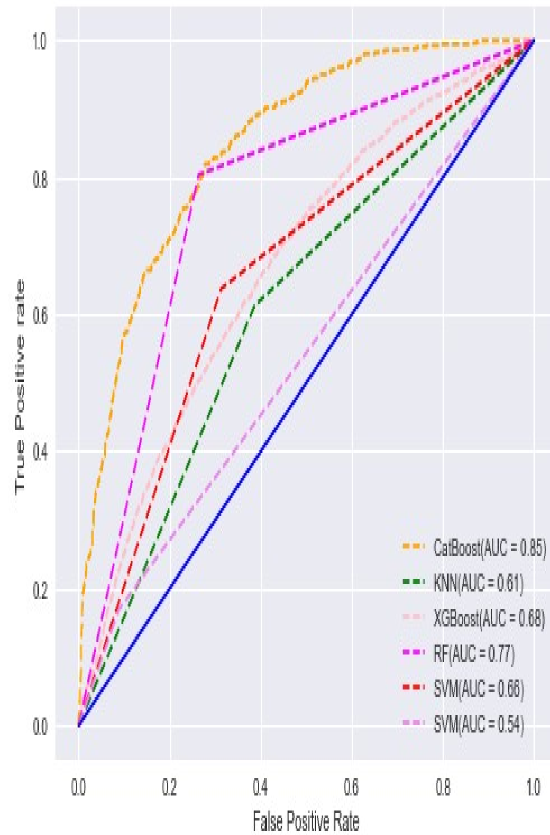

Figure 4: Models AUC curve

Figure 4 illustrates the ROC-AUC curves of the six models utilized in this study. According to the AUC comparison of the models, CatBoost outperformed the other models with an AUC score of 0.85, indicating the best performance. It was closely followed by RF with a score of 0.77, XGBoost with a score of 0.68, and SVM with a score of 0.61. These results imply that CatBoost, RF, and XGBoost exhibited satisfactory performance. On the other hand, nave Bayes demonstrated poor performance.

In Figure 5, a simulation is presented to illustrate the comparison between six models based on various performance measures. The results show that CatBoost performed the best in terms of accuracy, precision, Kappa, and f-measure. Additionally, RF achieved good results in terms of accuracy, precision, and f-measure. On the other hand, the K-NN and NB models performed the worst. Overall, these findings demonstrate that CatBoost exhibited the strongest performance. Thus, we pick up the CatBoost as the central algorithm of our proposed system.

Journal of Theoretical and Applied Information Technology
31st March 2024. Vol.102. No 6

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Table 3: Comparison of models performances.

Methods (%)	Precision (%)	Accuracy (%)	Recall (%)	Kappa (%)	F-measure (%)	AUC score (%)
K-NN	62.4	67.1	41.1	26	49.4	61
CatBoost	94	87.1	72	71.12	82.2	85
XGBoost	68.1	71.3	50.1	36.1	58.3	67.1
RF	78	79.1	68.3	56.2	72.3	77.8
SVM	6 2.5	70.4	64.3	38.2	63.1	66
NB	56	61.1	14.4	7.9	23	54

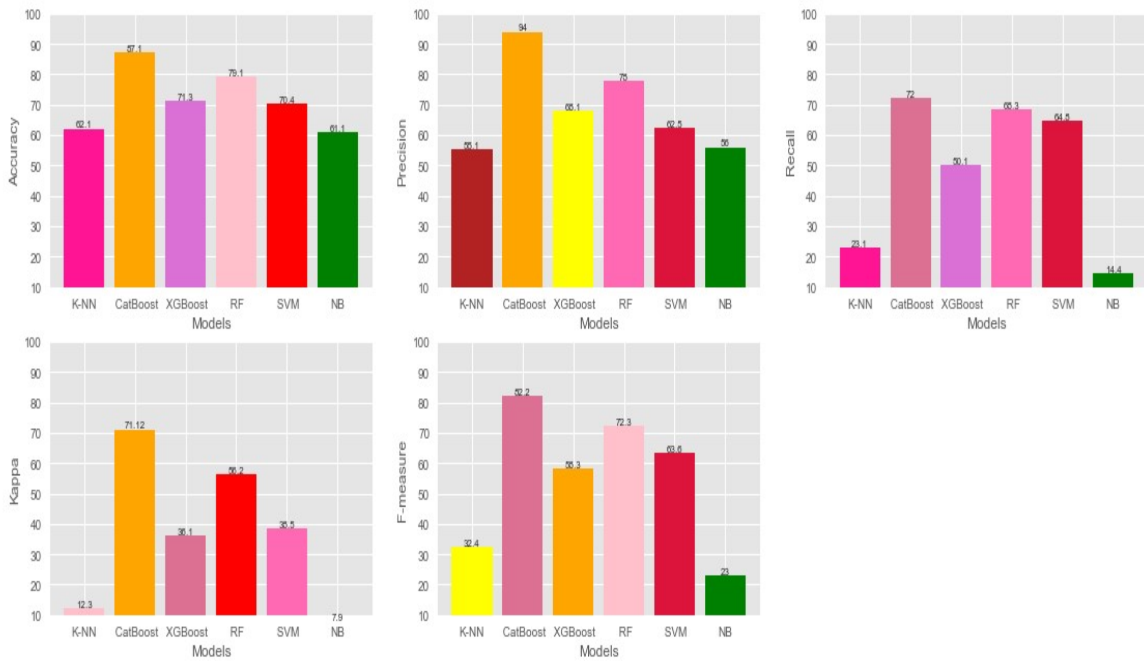

Figure 5: The performance of the selected models comparison

Table 4: Comparison of models performances from the literature.

Models (%)	Accuracy (%)
[11]	82
[12]	84.20
[13]	85.1
Proposed model	87.1

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Table 4 illustrates a comparison between our proposed model and previously suggested models known for delivering favorable results in the literature, particularly in terms of accuracy. Our model demonstrates robust performance, achieving an accuracy of 87.1%.

3. CONCLUSION

In the insurance sector, the majority of companies leverage artificial intelligence for detecting claims or crises. However, there has been limited progress in deploying it for the refinement of more precise marketing strategies. Many still rely on traditional marketing approaches, centered around mass marketing, which tends to be more expensive and time-consuming. For this reason, in this paper, we present an intelligent precision marketing framework that integrates machine learning, feature engineering, and customer persona analysis. A case study of an American insurance company was conducted to validate the effectiveness of this framework. The system offers the potential to enhance marketing services by analyzing customer preferences, leading to the delivery of accurate and timely services and products. By combining machine learning, feature engineering, and customer persona analysis. The advantage of the proposed precision marketing system for insurance companies, allowing them to remain competitive, gain a better understanding of their customers, and achieve long-term development while saving time and increasing customer satisfaction. Existing relationship between customer lifetime value and customer response by utilizing hybrid classification methods. The limitation of this work can be addressed by including more model comparisons. As a future direction for this work, we plan to enhance the functionalities of our suggested system by incorporating additional variables and integrating a broader range of data sources. The incorporation of a more expansive dataset will facilitate the investigation of advanced methodologies. We also intend to work with data from various areas.

REFERENCES:

[1] El Koufi N., Belangour A., and Sadiq M., “Machine learning application in precision marketing: a systematic literature review and comparative study”, *The International Conference on Artificial Intelligence and Smart Environment, Springer*, Vol. 635, No. 60, 2023, pp. 601-607.

[2] El Koufi N., Belangour A., and Sadiq M., “Research on precision marketing based on big data analysis and machine learning: case study of Morocco”, *International Journal of Advanced Computer Science and Applications*, Vol. 13, No. 10, 2022, pp. 58-63.

[3] Zhang H., Zhang L., Cheng L., and Chen W., “A novel precision marketing model based on telecom big data analysis for luxury car”, *16th International Symposium on Communications and Information Technologies (ISCIT)*, Vol. 59, No. 10, 2016, pp. 307-311.

[4] Xiao K., and Hu X., “Study on maritime logistics warehousing center model and precision marketing strategy optimization based on fuzzy method and neural network model”, *International Journal of Advanced Computer Science and Applications*, Vol. 13, No. 10, 2017, pp. 58-63.

[5] Liu S. Y., “Precision marketing scheme based on integrating spatio-temporal data clustering and neural network”, *Journal of Physics: Conference Series*, Vol. 1087, No. 3, 2017, pp. 1-7.

[6] Zhang S., Liao P.,Ye H. Q., and Zhou Z., “Multiple resource allocation for precision marketing”, *Journal of Physics: Conference Series,* Vol. 31592, No. 1, 2020, pp. 58-63.

[7] Li J., Pan S., and Huang L., “A machine learning based method for customer behavior prediction”, *Tehnicki vjesnik*, Vol. 26, No. 6, 2019, pp. 1670-1676.

[8] Chiu M. C., Chuang K. H., “Applying transfer learning to achieve precision marketing in an omni-channel system–a case study of a sharing kitchen platform”, *International Journal of Production Research*, Vol. 59, No. 24, 2021, pp. 7594-7609.

[9] Gao Z., “Precision marketing mode of agricultural products e-commerce based on KNN algorithm”, *The International Conference on Innovative Computing, Springer*, Vol. 791, No. 90, 2021, pp. 1151-1158.

[10] Liu H., “Big data precision marketing and consumer behavior analysis based on fuzzy clustering and PCA model”, *Journal of Intelligent and Fuzzy Systems*, Vol. 40, No. 4, 2021, pp. 6529-6539.

[11] Abakouy, R., En-Naimi, E. M., El Haddadi, A., Elaachak, “Machine Learning as an Efficient Tool to Support Marketing Decision-Making”, International Conference on Smart City Applications, Springer, Vol 4, 2020, pp. 244-258

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

[12] Zheng, Y., Li, L., Zhang, J., Xie, Q., & Zhong, L., "Using sentiment representation learning to enhance gender classification for user profiling", International Conference on Web and Big Data, Springer, Vol 55, 2019, pp. 3-11.

[13] Wang, Q., Cai, R., Zhao, M., "E-commerce brand marketing based on FPGA and machine learning", Microprocessors and Microsystems. Vol 2, 2020, pp. 1–17.