Kocaeli University

Kocaeli Journal of Science and Engineering

http://dergipark.org.tr/kojose

Forecasting Customer Churn using Machine Learning and Deep Learning Approaches

Ceren AKSOY [1], Ayhan KÜÇÜKMANİSA [2,*], Zeynep Hilal KİLİMCİ [3]

[1] *Department of Electronics and Communication Engineering, Kocaeli University, Kocaeli, Türkiye*, ***ORCID:*** *0000-0002-7349-4967*
[2] *Department of Electronics and Communication Engineering, Kocaeli University, Kocaeli, Türkiye*, ***ORCID:*** *0000-0002-1886-1250*
[3] *Department of Information Systems Engineering, Kocaeli University, Kocaeli, Türkiye*, ***ORCID:*** *0000-0003-1497-305X*

Article Info

Research paper

Received : *August 1, 2024*
Accepted : *December 23, 2024*

Keywords

Customer churn prediction
Random forest
Deep learning
Convolutional neural networks

Abstract

Customer churn forecasting is a challenging task recommended for churn prevention for companies operating in various industries such as banking, telecommunications, and insurance. Forecasting customer churn is very important for many companies because gaining potential customers usually costs more than retaining present ones. That is why companies, analysts, and researchers are center on analyzing the dynamics behind customer churn behaviors. In this study, we present a comparative study for the purpose of forecasting customer churn employing publicly available datasets, namely, IBM Watson and Call-Detailed Record (CDR). For this purpose, logistic regression, random forest, decision tree, k-nearest neighbor, extreme gradient boosting, and naive Bayes techniques are evaluated as machine learning approaches while artificial neural networks and convolutional neural networks are assessed as deep learning models. Experiment results indicate that the random forest method exhibits superior performance with 79.94% accuracy for the IBM Watson dataset and 96.34% accuracy for the Call Detailed Report (CDR) dataset. To demonstrate the effectiveness of the suggested framework, a comparison with the state-of-the-art studies is performed.

1. Introduction

Customer churn forecast plays a significant role in determining whether existing customers will leave the service provider [1]. Many service providers find that keeping an existing customer is much less expensive than finding new ones. For this reason, many companies are making large investments in customer churn prediction studies today. Customer churn prediction studies are carried out in many different areas, especially in the telecommunication and banking sectors. According to research in the field of churn prediction, the telecommunication industry is a sector in which customer churn prediction is used the most with 60%. The telecommunication sector is followed by Banking and Finance with 9%, Combined (Telecom + Banking + Retail etc.) with 9%, and Media with 7%. Education, Online Games, Energy and Organization sectors are the least used application areas in customer churn prediction studies with a rate of 1% [2]. Machine learning is a sub-field of artificial intelligence that allows to predict the future situations and results by using and processing the data with the help of various algorithms. Many studies are carried out in different subjects such as disease detection, weather forecasting, stock prediction, speech recognition, object recognition, and so on by using machine learning methods. Logistic regression, random forest, support vector machines, k-nearest neighbour, and naïve Bayes are the widely used machine learning algorithms in many domains. Today, the most popular subfield of the machine learning approach is deep learning. It is possible to learn different level features with multiple processing layers with deep learning algorithms. The state-of-the-art in text classification, speech recognition, visual object recognition, object detection, and many other fields have been significantly enhanced by these techniques [3]. Deep learning methodology has become more popular in recent years as it exhibits higher performance results in large datasets when compared to conventional machine learning methods.

In this study, we present a comparative study to detect the churn of customers using two publicly available datasets. First, the relationship between the features in the

* Corresponding Author: ayhan.kucukmanisa@kocaeli.edu.tr

2667-484X © This paper published in Kocaeli Journal of Science and Engineering is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License

datasets is analyzed using the correlation matrix. Then, the pre-processing stage is carried out and the data is fed into the machine learning and deep learning methods. To demonstrate the effectiveness of the suggested model, various machine learning and deep learning methods are employed, and experiment results are compared with the state-of-the-art results. These are logistic regression (LR), k-nearest neighbor (k-NN), extreme gradient boosting (XGBoost), artificial neural networks (ANN), and convolutional neural networks (CNN). Extensive experiment results show that the random forest algorithm has the capability to detect customer churn for both datasets.

The major contributions of this study are as follows:

- A notable achievement lies in the adeptness at working across diverse domains, showcasing versatility rather than confining efforts to a single sector.
- The study emphasizes a meticulous approach to testing, ensuring comprehensive evaluations.
- The research attains commendable results by achieving high performance on two distinct datasets, highlighting the method's efficacy across varied data scenarios.

This paper is organized as follows. Section 2 provides a comprehensive review of recent churn prediction methods in the literature. Suggested methods and used datasets are detailed in Section 3. Experimental results and detailed performance evaluation are given in Section 4. Finally, Section. 5 provides the conclusion.

2. Literature Review

In this section, a brief of literature studies related to customer churn prediction is presented. [4] design a 5-layered ANN model to detect churn of customers utilizing the IBM Watson dataset. Using a 5-layered ANN model, 80.03% of accuracy is achieved. [5] develop a DNN model using the same dataset. The authors report that the usage of the DNN model provides 80.62% of accuracy and 95.35% of specificity. To show the efficiency of the proposed model, experiment results are compared with RF and XGBoost methods. Authors conclude the paper that using DNN model for predicting customer churn exhibits higher classification success. [6] conduct a study by employing both machine learning and deep learning models. Artificial Neural Networks (ANN), Decision Tree (DT), and Self-Regulating Maps (SOM) models are evaluated. Experiment results indicate that the proposed hybrid model by blending SOM, and ANN methods performs better and yields 79.5% of accuracy, 83.90% of precision, 89.40% of sensitivity, and 86.56% of F1-score.

[7] propose to analyze churn prediction using ANN models by diversifying the number of hidden layers and presenting the effect of hyperparameters on this task. The authors report that the proposed ANN model with 3 hidden layers and RMSprop optimizer configuration exhibits the best classification success obtaining 83.09% accuracy. To ensure a fair comparison, the performance of the proposed model is compared with RF, k-NN, and DT methods. [8] focuses on the hybrid model blending k-means and XGBoost algorithms for estimating the churn of customers. For this purpose, decision tree, logistic regression, k-means and XGBoost machine learning models are assessed. Experiment results show that the proposed hybrid model achieves 81% of accuracy, 75% of precision, and 90% of sensitivity. [9] present a comparative study to detect the churn customers utilizing various machine learning and deep learning methods on the Call-Detailed Record dataset. The authors report that the inclusion of the random forest model shows 94.66% of accuracy, 94% of precision, 83% of sensitivity, and 88% of F1-score. It is followed by the CNN model with 93.07% of accuracy, 91% of precision, 79% of sensitivity, and 84% of F1-score. [10] analyze the behaviors of customers for the aim of detecting churn customers on IBM Watson dataset. The authors conclude the study that logistic regression outperforms decision tree and random forest models with 97.8% of accuracy and 97.9% of precision. It is followed by a decision tree with 78.3% of accuracy and 77.7% of precision, and a random forest with 79.2% of accuracy and 78% of precision.

[11] center on estimating customer churn using the the XGBoost and the logistic regression algorithms. Authors report that utilization of XGBoost method demonstrates superior performance with 80.05% accuracy while the logistic regression model achieves 78.89% accuracy. [12] focus on the customer churn prediction in the telecommunication sector and present an efficient framework by comparing the success of different algorithms. Artificial neural network (ANN), decision tree, k-nearest neighbour, Gaussian naive Bayes, and support vector machine (SVM) models are employed. Experiment results show ANN model performs 79% accuracy, 67% precision, 55% recall and 60% F1-Score. Decision tree is capable of detecting customer churn with 70% of accuracy, 49% of precision, 49% of recall, and 49% of F1-score. The accuracy scores of the K-nearest neighbour, Gaussian Naive Bayes and Support vector machine methods are reported as 75%, 75%, and 78%, respectively. [13] investigate different machine learning techniques to estimate the customer of churn. These algorithms are the k-nearest neighbour, random forest and XGBoost. Experiment results demonstrate that k-nearest neighbour exhibits 75.4% of accuracy, and 49.5% of F-score. It is followed by random forest by ensuring 77.5% of accuracy and 50.6% of F-score. The best classification results are obtained with the usage of XGBoost by ensuring 79.8% of accuracy, and 58.2% of F-score. [14] focuses on only the logistic regression method

by obtaining 80.3% of recall, 76.7% of accuracy score and 76.7% of AUC value for estimating churn customers. A random forest classifier model is proposed in [15] using up-sampling and the Edited Nearest Neighbors (ENN) technique. The authors reported achieving an accuracy of 99.09%, but they did not specify their train-test procedure.

3. Materials and Methods

In this part, the suggested framework is introduced by giving details on dataset collection, pre-processing stages, model construction and experiment setup. After IBM Watson and Call Detailed Report (CDR) datasets are gathered, the relationship between features in each dataset is examined employing a data visualization platform. In the data visualization step, various relational graphs and correlation matrices are constructed. After that, the data pre-processing stage is proceeded applying standardization and a label encoder method. While standardization is performed on both datasets, label encoder is also employed for the purpose of converting object values to numerical values in the IBM Watson dataset. Next, the datasets are divided into two separate sets training and test data. The training data is utilized in the training of deep learning and machine learning models while the test data is used to test the performance of the models. The methods used in this work are logistic regression, k-nearest neighbor, decision tree, random forest, xgboost, naive bayes, artificial neural networks and convolutional neural networks. To measure the classification performance of each model, precision, recall, accuracy, F1-score, AUC values are assessed as evaluation metrics. The block diagram of the suggested method is given in Figure 1.

IBM Watson and Call Detailed Report (CDR) dataset churn ratios are given in Figure 2, and Figure 3. Features in the IBM Watson dataset are CustomerID, Gender, SeniorCitizen, Partner, Dependents, Tenure, Phone Service, MultipleLines, InternetService, OnlineSecurity, OnlineBackup, DeviceProtection, TechSupport, StreamingTV, StreamingMovies, Contract, Paperless Billing, Payment Method, Monthly Charges, Total Charges. The features of the CDR dataset are composed of Phone Number, Account Lenght is VMail Message, Day Mins, Day Calls, Day Charge Eve Mins, EveCalls, Eve Charge, Night Mins, Night Calls, Night Charge, Intl Mins, Intl Calls, Intl Charge, CustServ Calls, Churn label.

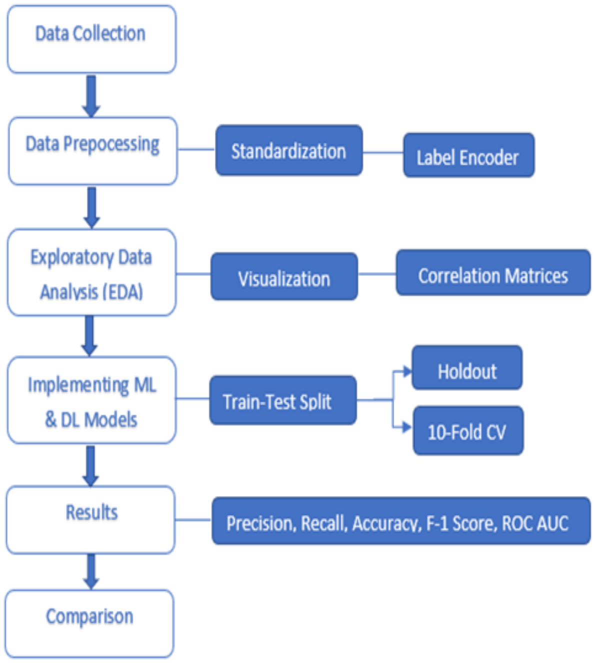


Figure 1. Block diagram of the proposed method

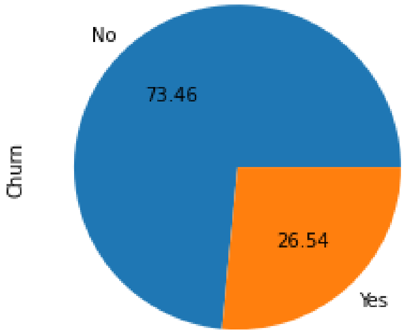


Figure 2. Customer churn ratio of IBM Watson dataset

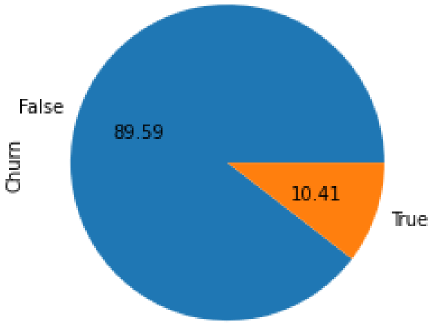


Figure 3. Customer churn ratio of CDR dataset

3.1. Data pre-processing

In the data preprocessing stage, whether there is a missing value or not is controlled. After that, data standardization and normalization methods are carried out. The relationship between the features in the datasets is examined and feature selection is performed.

1. Standardization: The data is converted to numeric data values in the [0,1] range using the min-max scaler.

2. Feature Selection: The relationship between features is observed by extracting correlation matrices and creating various relational graphs. At this stage, significant features affecting churn are observed. The feature selection step is only applied to the IBM Watson dataset, as better performance is achieved when the feature space is narrowed by feature selection in the IBM Watson dataset. After applying the feature selection procedure IBM Watson dataset is modelled by removing Gender, PhoneService, StreamingMovies, StreamingTV, MultipleLines, SeniorCitizen, TotalCharges features.

3. Label Encoding: For machine learning techniques, all input and output variables must be converted to numbers. The categorical data is transformed into numerical values for this purpose before moving on to the modeling stage. Hot encoding and integer encoding are the most widely used methods. Embedding learning, for example, can offer a neutral ground in the development of newer techniques that can process these two approaches [16]. Furthermore, a typical data pre-processing step is label encoding binary columns, which is carried out by standardizing the data. To simplify data processing and shorten processing times for machine learning models, data in the form of strings is converted to binary form (0 or 1) [5]. The label encoding task in this work is accomplished using integer coding.

3.2. Modeling for Classification

Different machine learning and deep learning methods are employed in this study. These are Logistic Regression (LR), Decision Tree (DT), Random Forest (RF), K-Nearest Neighbor (KNN), Extreme Gradient Boosting (XGBoost), Naïve Bayes (NB) Artificial Neural Networks (ANN), and Convolutional Neural Networks (CNN).

3.2.1. Logistic Regression (LR)

A common linear statistical model in the literature is logistic regression (LR). It also falls under the category of supervised learning, where each piece of data is assigned a label. One or more independent variables that explain the relationship between the dependent and independent variables make up generalized linear models, such as logistic regression (LR) [17].

3.2.2. Decision Tree (DT)

Decision trees (DT) are hierarchical and tree structure models. The main aim of DT is to build a model that predicts the value of a target variable by learning decision rules derived from the data features. A decision tree's logical rules are much simpler compared to a neural network's numerical weights. Each decision tree node helps to predict the target by dividing possible outcomes [18]. There are additional nodes that branch off from each of those outcomes, opening additional possibilities. This gives it a shape resembling a tree.

3.2.3. Random Forest (RF)

The main possible problem of using a single decision tree is facing an overfitting problem. As a solution to this problem Random Forest is suggested. A random forest's fundamental idea is to build numerous decision trees on random subsets of the initial training dataset [19].

3.2.4. K-Nearest Neigbour (KNN)

The Nearest Neighbor Classification method's underlying idea is quite simple: data space is grouped according to the class of their closest neighbors. The technique is more commonly known as k-Nearest Neighbour (k-NN) Classification because it is frequently useful to take into account more than one neighbor. K nearest neighbors are used to determine the class [20].

3.2.5. Extreme Gradient Boosting (XGBoost)

Extreme gradient boosting is known as XGBoost. Gradient-boosted decision trees are implemented using XG Boost, which is a quick and effective algorithm. It is a supervised and ensemble learning technique that combines trees to produce a more generalizable machine learning model. Boost classifiers are typically built with many different trees, averaging the results for better prediction. [21].

3.2.6. Naive Bayes (NB)

The Naive Bayes classifier is a probabilistic method that benefits from Bayes theorem. It makes predictions based on the probability of an object. To simplify computation, naive Bayes classifiers use the assumption known as "class conditional independence," which holds that each feature's value affects each class in an independent manner [22].

3.2.7. Artificial Neural Network (ANN)

ANNs are massively parallel computing systems made up of an enormous number of straightforward processors connected by many connections, and they were inspired by biological neural networks. The "organizational" principles that are allegedly used by humans are attempted to be used by ANN models. The nodes are viewed as "artificial neurons" in one type of network. Artificial neural networks (ANNs) are what these are. A computational model of an artificial neuron is one that was inspired by natural neurons [23]. For ANN models, a train-test split is used in this study, with 80% of the training dataset and 20% of the test dataset for both datasets. Adam optimizer is employed as the ANN model's optimizer for both datasets. The learning rate of the Adam optimizer is 0.00015. Algorithms for gradient descent are optimized using the Adam Optimizer (AO), a technique for adaptive moment estimation. The AO also uses a hybrid of the Root Mean sq\. Propagation (RMSP) and Momentum gradient descent methods [24]. The output layer in the ANN models built for both datasets uses the sigmoid activation function instead of the ReLU activation function in all other layers. The sigmoid and hyperbolic tangent activation functions cannot be derived after a certain threshold value during training, which causes the disappearing gradient problem [25]. The ReLU activation function is created to address this issue. One of the distinguishable nonlinear activation functions is the ReLU activation function [26]. One of the most popular activation functions (AFs) in artificial neural networks (ANNs) is the ReLU function, which is defined as max (0, x), where x is the input variable [27]. One of the most well-known functions used in feedforward neural networks is the sigmoid function, also known as the logistic sigmoid function. This is mainly because of its nonlinearity and the derivative's simplicity, which requires little computational effort. The AF sigmoid function is a bounded differentiable real function that has positive derivatives and is defined for all real input values [28]. The proposed ANN architecture is given in Figure 4.

3.2.8. Convolutional Neural Network (CNN)

CNN has neurons with trainable weights and biases. Deep learning, which is another name for CNN, can have multiple layers. Generally, more than one convolutional layer is used in CNN. Typically, they work as a feed-forward neural network. As a final layer [29], one or more fully connected layers come after CNNs. The CNN architecture of IBM Watson and Call Detailed Report (CDR) datasets are given in Figure 5.

In the proposed CNN model Adamax is used as the optimizer. The learning rate of the Adamax optimizer is 0.02. An upgraded version of the Adam optimizer built on the infinite norm is available as the Adaptive Max Pooling (AdaMax) Optimizer [30]. AdaMax has the significant advantage of giving a more dependable solution and being far less susceptible to hyper-parameter selection [31]. AdaGrad can be improved using the adaptive learning strategy known as root-mean-square prop (RMSprop) [32]. It uses an “exponential moving average” rather than the cumulative sum of squared gradients, similar to AdaGrad [33]. Batch size and epochs are hyperparameters used in deep learning models. Batch defines the number of samples to run on before updating model parameters. Epochs define how many times the algorithm will run across the entire training dataset. In this study, the batch size is set to 50, and the epoch is 300.

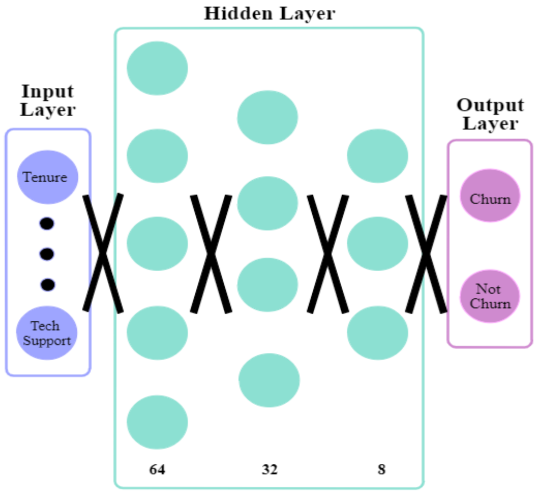


Figure 4. ANN architecture

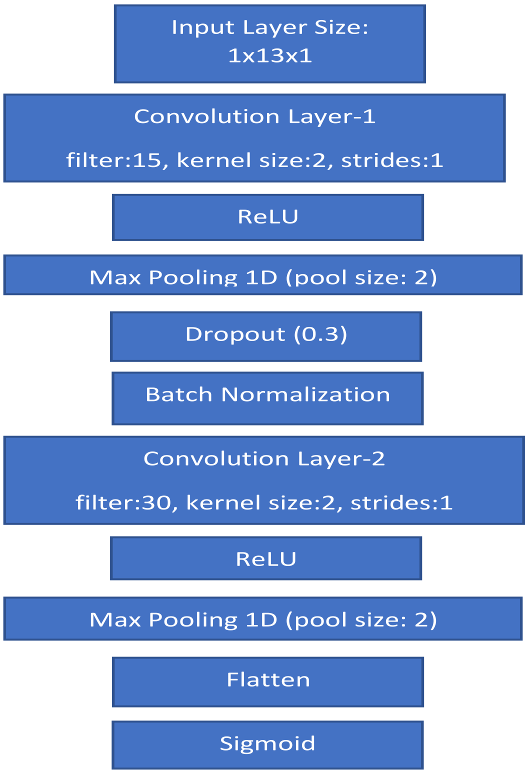


Figure 5. CNN architecture

4. Experimental Results

In this section, first, the datasets used in this study are analyzed in detail. Afterward, the performance of the suggested methods is examined and compared with the methods in the literature.

4.1. IBM Watson Dataset Data Visualization Results

As a result of the data visualization of the IBM Watson dataset, it has been observed that some features in the dataset are quite effective in the churn. The results of data visualization are as follows:

- As can be seen in Figure 6, as the usage/rental time (tenure) of the Users increases, the loss rate decreases significantly.
- Figure 7 shows that users who use fiber optic as an Internet Service have a higher tendency to abandon their subscribers than users who use DSL or do not have an Internet provider.
- As can be seen in Figure 8, users with monthly contracts have the highest loss rate. For users with a 1-year contract, this rate is much lower. For users with a 2-year contract, the loss rate is almost non-existent.
- Figure 9 indicates that users without technical support are more likely to churn than users with technical support or no internet service.
- As can be seen in Figure 10, a great churn is observed in those who made their payment via the Electronic Check method.
- A high rate of churn has been observed among those without Device Protection.
- The churn rate for those who do not have online security is much higher than those with online security and for those who do not have internet service.
- When the MonthlyCharges value is between 70-100, the most churn is experienced.

4.1. Results and Evaluation

Suggested methods in this work are evaluated using precision, sensitivity, F1-score, accuracy and ROC AUC metrics. IBM Watson and Call Detailed Report (CDR) datasets are used in the performance analysis. The hyperparameters used in the machine-learning based methods, ANN and CNN models of both datasets are given in Table 1, Table 2 and Table 3, respectively. Table 4 shows the 10-fold cross-validation performance results of the IBM Watson dataset, and Table 5 shows the 10-fold cross-validation performance results of the Call-Detailed Record dataset. The values presented in these tables are outcomes of an exhaustive analysis encompassing all hyperparameters of suggested methods. Table 1 also shows the parameter values that yielded the highest results throughout this comprehensive analysis. An example of parameter analysis, Table 6 presents a detailed parameter analysis for the suggested Random Forest method.

Accuracy can be defined as a percentage of correct classifications achieved by a machine learning model. This metric is preferred for the general evaluation of methods. As seen from Table 4 and Table 5, the suggested random forest model has the best accuracy value. Random Forest outperformed other methods likely due to its ensemble nature, which combines multiple decision trees to reduce overfitting and improve generalization. Its ability to handle complex and non-linear patterns in the data, along with feature randomness, allows it to effectively capture relationships in the datasets better than the other methods. This result also proves that machine learning methods generally have better performance than deep learning methods on smaller datasets. To investigate more capability of the suggested method, such as class confusion rate, ROC curve and confusion matrix of the Random Forest model are examined. These analysis results are shown in Figure 11 for IBM Watson dataset and Figure 12 for Call Detailed Report (CDR) dataset.

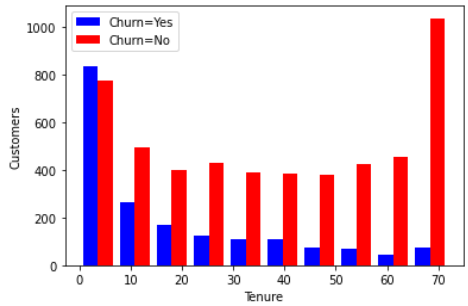


Figure 6. Tenure - Churn

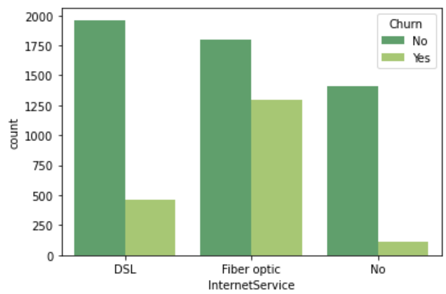


Figure 7. Internet Services – Churn

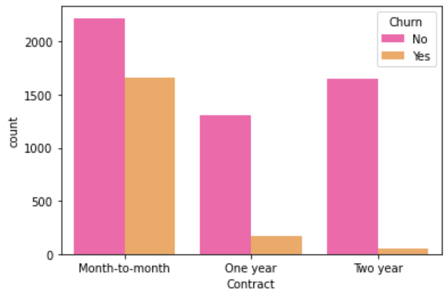


Figure 8. Contract - Churn

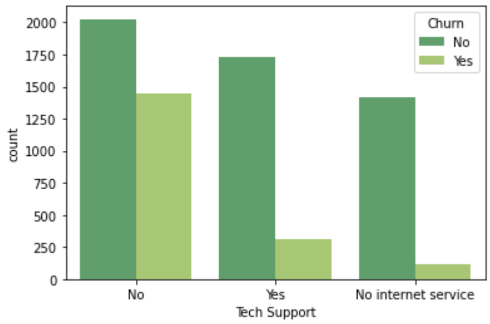


Figure 9. Tech Support – Churn

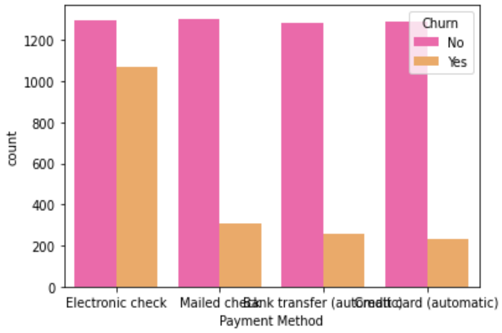


Figure 10. Payment Method- Churn

Table 1. Machine learning based methods parameters

LR	max_iter=100, solver='liblinear', penalty='l2'
k-NN	n_neighbors=5, weights='uniform', algorithm='auto', leaf_size=30, p=2
DT	criterion= 'gini', splitter='best', max_depth=None, min_samples_split=2,min_samples_leaf=1, max_features=None, random_state= None, max_leaf_nodes= None
RF	n_estimators=100, criterion='gini', min_samples_split=5, max_features='sqrt', max_leaf_nodes =None, class_weight= None, random_state=None
XGBoost	booster= 'gbtree', eta= 0.3, gamma = 0, max_depth = 6,min_child_weight=1, max_delta_step = 0, subsample= 1, colsample_bytree=1, seed = 0
NB	var_smooting = 1e-9

Table 2. ANN model parameters

Parameters	Value
Layers	[64,32,8,1]
Optimizer	Adam
Learning Rate	0.00015
Activation Functions	Relu + Sigmoid
Epochs	100
Batch Size	30

Table 3. CNN model parameters

Parameters	Value
Layers	[15,30]
Optimizer	Adamax
Learning Rate	0.02
Activation Functions	Relu + Sigmoid
Epochs	300
Batch Size	50

Table 4. Accuracy and performance results for IBM Watson dataset

Models	Precision	Sensitivity	F1-Score	Accuracy	ROC AUC
Logistic Regression	0.85	0.91	0.88	79%	0.856
K-Nearest Neighbor	0.83	0.88	0.86	77.44%	0.801
Decision Tree	0.82	0.83	0.82	72.80%	0.657
Random Forest	**0.84**	**0.92**	**0.88**	**79.94%**	**0.863**
XGBoost	0.84	0.91	0.88	79.87%	0.861
Naive Bayes	0.90	0.76	0.82	75.43%	0.844
ANN	0.84	0.92	0.88	79.33%	0.715
CNN	0.86	0.86	0.86	78.51%	0.749

Table 5. Accuracy and performance results for Call-Detailed Record (CDR) dataset

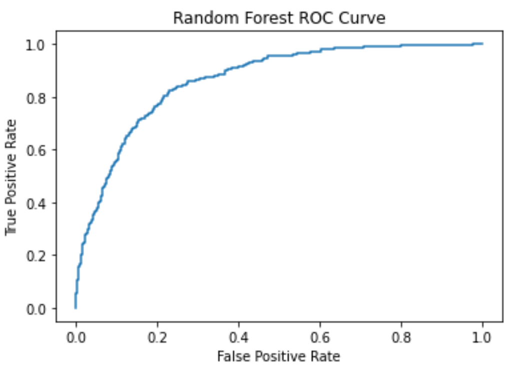
Models	Precision	Sensitivity	F1-Score	Accuracy	ROC AUC
Logistic Regression	0.90	1.00	0.95	90.00%	0.552
K-Nearest Neighbor	0.94	0.95	0.95	96.16%	0.904

Table 5. (Cont.) Accuracy and performance results for Call-Detailed Record (CDR) dataset

Decision Tree	0.96	0.81	0.86	94.86%	0.716
Random Forest	**0.91**	**0.99**	**0.95**	**96.34%**	**0.859**
XGBoost	0.95	0.71	0.81	91.48%	0.765
Naive Bayes	0.90	0.90	0.90	84.23%	0.554
ANN	0.90	1.00	0.95	89.55%	0.5
CNN	0.90	1.00	0.94	89.27%	0.5

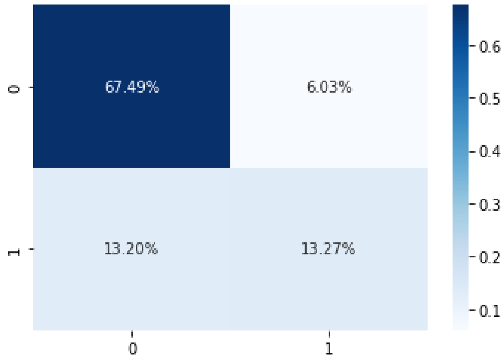
Table 6. Parameter analysis of the suggested Random Forest Method

Parameters	Accuracy
n_estimators = 100	78.57%
n_estimators = 50	78.46%
n_estimators = 200	78.62%
n_estimators = 250	78.48%
criterion = ‘gini’	78.57%
criterion= ‘entropy’	78.46%
min_samples_leaf=1	78.57%
min_samples_leaf=3	79.60%
min_samples_leaf=5	79.94%
min_samples_leaf=7	79.78%
min_samples_leaf=9	79.78%
max_features = ‘sqrt’	78.57%
max_features = ‘log2’	78.57%
max_features = ‘auto’	78.57%
class_weight= ‘balanced’	78.70%
class_weight= ‘balanced_subsample’	78.50%



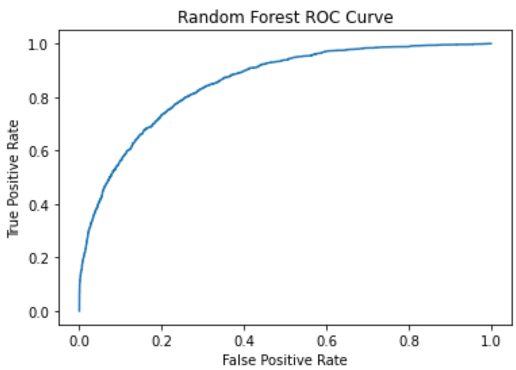
(a) ROC curve

Figure 11. ROC Curve and confusion matrix for IBM Watson Dataset of the suggested Random Forest model

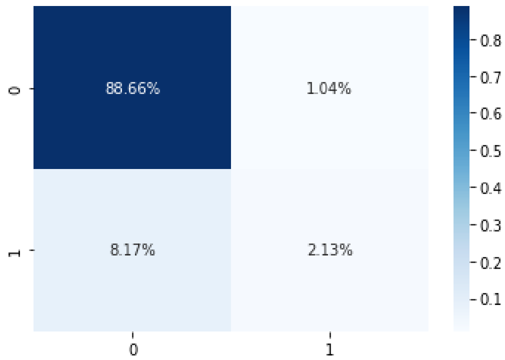


(b) Confusion matrix

Figure 11. (Cont.) ROC Curve and confusion matrix for IBM Watson Dataset of the suggested Random Forest model



(a) ROC curve



(b) Confusion matrix

Figure 12. ROC Curve and confusion matrix for Call-Detailed Record (CDR) Dataset of the suggested Random Forest model

The methods suggested in this study are compared with the methods in the literature. These evaluation results are provided in Table 7 for the IBM Watson dataset and Table 8 for the CDR dataset. In these tables model details of compared methods are given as well. When Table 7 is examined, it is seen that the highest accuracy result is the suggested random forest model. Even though [15] reported a result of 99.09%, it is not included in Table 7 for comparison due to the lack of details regarding their test procedure. There are not many methods in the literature that can be used for comparison for the Call Detailed Report (CDR) dataset. In Table 8, it is seen that we have presented a more successful result by giving higher results in 4 of the 6 similar methods applied in this dataset. It also scores higher than all recommended methods given in [11] with the highest performance value of 96.34%.

The processing time of the suggested methods is also analyzed. The analysis times are important as they demonstrate the speed at which the methods generate results on their respective platforms. Table 9 shows the processing times of the suggested methods. These times are obtained on a PC with a 2.80 Ghz CPU, 8 GB RAM and onboard GPU. These results show that the suggested Random Forest method has a good balance between performance and processing speed.

Table 7. Comparison of the suggested method and recent literature studies on IBM Watson dataset

Methods	Model	Accuracy
[17]	Logistic Regression	76.70%
[23]	XGBoost	79.80%
[23]	Random Forest	77.50%
[23]	KNN	75.40%
[23]	XGBoost	80.05%
[7]	Random Forest	77.87%
[7]	XGBoost	76.45%
[8]	Decision Tree	73.41%
[14]	Decision Tree	70.00%
[14]	KNN	75.00%
[14]	Naive Bayes	75.00%
[14]	ANN	79.00%
[8]	ANN + SOM	79.53%
[7]	DNN	80.60%
[6]	ANN	80.03%
Suggested method (10-fold CV)	Random Forest	79.94%
Suggested method (Holdout)	Random Forest	**80.76%**

Table 8. Comparison of the suggested method and recent literature studies on CDR dataset

Models	[11]	Suggested Methods
Naive Bayes	85.06%	84.23%
Random Forest	94.66%	96.34%
KNN	86.68%	96.16%
Decision Tree	91.96%	94.89%
Logistic Regression	77.38%	89.57%
ANN	92.44%	89.27%

Table 9. Processing times of the suggested methods (sec)

Models	On IBM Watson dataset	On CDR Dataset
Logistic Regression	0.002	0.003
K-Nearest Neighbor	0.107	1.108
Decision Tree	0.001	0.001
Random Forest	0.027	0.202
XGBoost	0.003	0.016
Naive Bayes	0.002	0.009
ANN	0.307	1.134
CNN	0.272	1.088

4. Conclusion

There are many reasons why customers churn in the telecommunications industry. Many factors, such as the length of contract that customers make with the company, the internet service method they use, or the payment method, are effective in the churn. The purpose of this study is to predict whether users who receive service from a service provider will leave that service provider before they leave. A customer churn prediction study is basically considered a binary classification problem. By considering the previous behaviours of the customers, it is tried to predict the churn that may be encountered in the future. Although both datasets belong to the same sector, factors such as the fact that they contain information about different features of different customers, and the total amount of data in the datasets is different, causing the same models to give different performance results in different datasets. Experimental results show that the suggested random forest method has the best results among to others. Also, the random forest model outperformed when it is compared with recently published methods in the literature.

Declaration of Ethical Standards

The authors of this article declare that the materials and methods used in this study do not require ethical committee permission and/or legal-special permission

Conflict of Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] Ahmad A. K., Jafar A., Aljoumaa K., 2019. Customer Churn Prediction in Telecom Using Machine Learning in Big Data Platform. Journal of Big Data, **6**(1).

[2] Koçoğlu F. Ö., Özcan T. Ş., Baray A., 2016. Veri madenciliğinde ayrılan müşteri üzerine literatür araştırması. Uluslarası Katılımlı 16. Üretim Araştırmaları Sempozyumu, İstanbul, Türkiye, 12-14 October.

[3] LeCun Y., Bengio Y., Hinton G., 2015. Deep learning. Nature, **521**(7553), pp. 436–444.

[4] Agrawal S., Das A., Gaikwad A., Dhage S., 2018. Customer churn prediction modelling based on behavioural patterns analysis using Deep Learning. 2018 International Conference on Smart Computing and Electronic Enterprise (ICSCEE), Selangor, Malaysia, 11-12 July.

[5] Putu O. H. N., Setyo A., 2020. Telecommunication service subscriber churn likelihood prediction analysis using diverse machine learning model. 2020 3rd International Conference on Mechanical, Electronics, Computer, and Industrial Technology (MECnIT), Medan, Indonesia, 25-27 June.

[6] Jay Shen T., Bin Shibghatullah A., 2022. Developing machine learning and deep learning models for customer churn prediction in telecommunication industry. Proceedings of International Conference on Artificial Life and Robotics, **27**, pp. 533–539.

[7] Nalatissifa H., Pardede H. F., 2021. Customer decision prediction using deep neural network on Telco Customer Churn Data. Jurnal Elektronika Dan Telekomunikasi, **21**(2), pp. 122.

[8] Tang P., 2020. Telecom customer churn prediction model combining K-means and XGBoost algorithm. 2020 5th International Conference on Mechanical, Control and Computer Engineering (ICMCCE), 25-27 December.

[9] Siddika A., Faruque A., Masum A. K., 2021. Comparative analysis of churn predictive models and factor identification in telecom industry. 2021 24th International Conference on Computer and Information Technology (ICCIT), Dhaka, Bangladesh, 18-20 December.

[10] Tamuka N., Sibanda K., 2020. Real time customer churn scoring model for the telecommunications Industry. 2020 2nd International Multidisciplinary Information Technology and Engineering Conference (IMITEC), Kimberley, South Africa, 25-27 November.

[11] Parmar P., Serasiya S., 2021. Telecom Churn Prediction Model using XgBoost Classifier and Logistic Regression Algorithm. International Research Journal of Engineering and Technology (IRJET*)*, **8**(5), pp. 1100-1105.

[12] Makruf M., Bramantoro A., Alyamani H. J., Alesawi S., Alturki R., 2021. Classification methods comparison for customer churn prediction in the telecommunication industry. International Journal of Advanced and Applied Sciences, **8**(12), pp. 1–8.

[13] Raja J., Pandian S., 2019. An Optimal Ensemble Classification for Predicting Churn in Telecommunication. Journal of Engineering Science and Technology Review, **13**(2), pp. 44-49.

[14] Hargreaves C. A., 2019. A machine learning algorithm for Churn Reduction & Revenue Maximization: An application in the telecommunication industry. International Journal of Future Computer and Communication, **8**(4), pp. 109–113.

[15] Wagh Sharmila K., Andhale Aishwarya A., Wagh Kishor S., Pansare Jayshree R., Ambadekar Sarita P., Gawande S.H., 2024. Customer churn prediction in telecom sector using machine learning techniques. Results in Control and Optimization Volume 14, March 2024, 100342.

[16] Brownlee J., 2023. 3 Ways to Encode Categorical Variables for Deep Learning, https://machinelearningmastery.com/how-to-prepare-categorical-data-for-deep-learning-in-python, 03.03.2023.

[17] Wibowo V. V. P., Rustam Z., Laeli A. R., Said A. A., 2021. Logistic regression and logistic regression-genetic algorithm for classification of Liver Cancer Data. 2021 International Conference on Decision Aid Sciences and Application (DASA), Sakheer, Bahrain, 07-08 December.

[18] Kotsiantis S. B., 2021. Decision trees: A recent overview. Artificial Intelligence Review, **39**(4), pp. 261–283.

[19] He S., Wu J., Wang D., He X., 2022. Predictive modeling of groundwater nitrate pollution and evaluating its main impact factors using random forest. Chemosphere, **290**, pp. 133388.

[20] Cunningham P., Delany S. J., 2021. K-nearest neighbour classifiers - a tutorial. ACM Computing Surveys, **54**(6), pp. 1–25.

[21] Pamina J., Raja B., SathyaBama S., Soundarya S., Sruthi M. S., Kiruthika S., Aiswaryadevi V. J., 2019. An Effective Classifier for Predicting Churn in Telecommunication. Jour of Adv Research in Dynamical & Control Systems, **11**, pp. 221-229.

[22] Dong T., Shang W., Zhu H., 2011. Naive bayesian classifier based on the improved feature weighting algorithm. Communications in Computer and Information Science, **152**, 142–147.

[23] Gupta N., 2013. Artificial Neural Network. Network and Complex Systems, **3**(1), pp. 24-28.

[24] Bock S., Goppold J., Weiß M., 2018. An improvement of the convergence proof of the ADAM-Optimizer. arXiv:1804.10587.

[25] Nair V., Hinton G. E., 2010. Rectified linear units improve restricted boltzmann machines. ICML, Haifa, Israel, 807-814.

[26] Adem K., 2022. Impact of activation functions and number of layers on detection of exudates using circular hough transform and convolutional neural networks. Expert Systems with Applications, **203**, pp. 117583.

[27] Surekcigil Pesch I., Bestelink E., de Sagazan O., Mehonic A., Sporea R. A., 2022. Multimodal transistors as ReLU activation functions in physical neural network classifiers. Scientific Reports, **12**(1), 670.

[28] Gustineli M., 2022. A survey on recently proposed activation functions for deep learning. arXiv:2204.02921.

[29] Abbas S., Alhwaiti Y., Fatima A., Khan M. A., Adnan Khan M., Ghazal T. M., Kanwal A., Ahmad M., Sabri Elmitwally, N., 2022. Convolutional neural network based intelligent handwritten document recognition, Computers, Materials & Continua, **70**(3), pp. 4563-4581.

[30] Prilianti K. R., Brotosudarmo T. H., Anam S., Suryanto A., 2019. Performance comparison of the convolutional neural network optimizer for photosynthetic pigments prediction on plant digital image. AIP Conference Proceedings, 020020, pp. 1-8.

[31] Haji S. H., Abdulazeez A. M., 2021. Comparison of Optimization Techniques Based on Gradient Descent Algorithm: A Review. Palarch’s Journal of Archaeology of Egypt/Egyptology, **18**(4), 2021.

[32] Štifanić J., Štifanić D., Zulijani A., Car Z., 2021. Multiclass Classification of Oral Squamous Cell Carcinoma.Student Scientific Conference RiSTEM 2021, Rijeka, Croatia, 10 June.

[33] Bansal K., Bathla R. K., Kumar Y., 2022. Deep transfer learning techniques with hybrid optimization in early prediction and diagnosis of different types of oral cancer. Soft Computing, **26**, pp. 11153-11184.