MAUN Fen Bil. Dergi., 12:1, 39-50 Araştırma Makalesi/ Research Article
MAUN J. of Sci., 12:1, 39-50 DOI: 10.18586/msufbd.1466246

# Neural Network Based a Comparative Analysis for Customer Churn Prediction

Anıl UTKU[1]🖉, M. Ali AKCAYOL[2]
[1] Computer Engineering, Faculty of Engineering, Munzur University, Tunceli, Turkey
[2] Computer Engineering, Faculty of Engineering, Gazi University, Ankara, Turkey

✉: anilutku@munzur.edu.tr [1] 0000-0002-7240-8713 [2] 0000-0002-6615-1237

Received (Geliş): 06.04.2024 Revision (Düzeltme):17.05.2024 Accepted (Kabul): 20.05.2024

## ABSTRACT

Customer churn refers to a customer's disconnection from a business. The expense associated with customer churn encompasses both the forfeited revenue and the marketing expenditures required to acquire new customers. Mitigating customer churn stands as the foremost objective for every business. Customer churn prediction will contribute to developing strategies enabling businesses to retain these customers by identifying customers with a high risk of loss. The importance of developing customer churn prediction models is increasing daily in the digital world. In this study, an MLP-based artificial neural network model was developed for customer churn prediction using customer data from an anonymous telecommunications company. The developed model was compared with kNN, LR, NB, RF, and SVM. The prediction results of the applied models were discussed, and the experimental results showed that all the models compared had over 70% accuracy. Experimental results showed that the developed MLP-based artificial neural network model has the most successful classification performance compared to other models, with approximately 94% accuracy.

**Keywords**: Artificial Neural Network, MLP, Customer Churn Prediction

# Müşteri Kayıp Tahmini için Sinir Ağı Tabanlı Karşılaştırmalı Analiz

## ÖZ

Müşteri kaybı, müşterinin bir işletmeyle bağlantısının kesilmesi anlamına gelir. Müşteri kaybıyla ilgili gider, hem kaybedilen geliri hem de yeni müşteriler kazanmak için gereken pazarlama harcamalarını kapsar. Müşteri kaybının azaltılması her işletmenin en önemli hedefidir. Müşteri kayıp tahmini, işletmelerin yüksek kayıp riski olan müşterileri belirleyerek bu müşterileri ellerinde tutmalarını sağlayan stratejiler geliştirmelerine katkıda bulunacaktır. Dijital dünyada müşteri kayıp tahmini modellerinin geliştirilmesinin önemi her geçen gün artmaktadır. Bu çalışmada, anonim bir telekomünikasyon şirketinden elde edilen müşteri verileri kullanılarak müşteri kayıp tahmini için MLP tabanlı yapay sinir ağı modeli geliştirilmiştir. Geliştirilen model kNN, LR, NB, RF ve SVM ile karşılaştırılmıştır. Uygulanan modellerin tahmin sonuçları tartışılmış ve deneysel sonuçlar, karşılaştırılan tüm modellerin %70'in üzerinde doğruluğa sahip olduğunu göstermiştir. Deneysel sonuçlar, geliştirilen MLP tabanlı yapay sinir ağı modelinin diğer modellere göre yaklaşık %95 doğrulukla en başarılı sınıflandırma performansına sahip olduğunu göstermiştir.

**Anahtar Kelimeler**: Yapay Sinir Ağları, Yapay Sinir Ağları, MLP, Müşteri Kayıp Tahmini

## INTRODUCTION

Customer churn means that a customer unsubscribes from a service they are using. The customer churn prediction also determines which customers are most likely to unsubscribe [1]. This information is essential for companies to retain their current customers. Therefore, the insights derived from the churn prediction help focus more on customers at high risk of leaving. For businesses, retaining existing customers is more accessible than acquiring new customers. Also, revenue from existing customers is often higher than revenue from new customers. The cost of acquiring customers can be even higher in a competitive industry where competitors are plentiful. Therefore, predicting customer churn before customers leave is essential for businesses to retain their customers [2].

There can be many reasons for customer churn. The presence of a new competitor in the market offering better prices or unsatisfactory service may result in customer churn. For reasons like these, there is no correct answer as to why the customer would want to give up. Although there are many factors for customer churn, it is usually simple to avoid. This is because the company makes its customers feel special and provides a customized experience to entice them to stay. Churn prediction is one of the most critical commercial sector data science applications. The fact that understanding its effects is more concrete and plays an essential factor in

the total profit earned by the business has made customer churn prediction a popular research area [3].
Many studies use machine learning and deep learning methods for customer churn prediction. This section continues by examining these studies and presenting their prominent features in a table.
Khodabandehlou and Rahman [4] conducted a machine learning-based study using the Iranian dataset for customer churn analysis. They selected a set of five variables and compared the model created with Artificial Neural Network (ANN), Support Vector Machine (SVM), and Decision Tree (DT). The results were impressive, with the proposed model achieving a 97.92% accuracy rate. This success underscores the effectiveness of machine learning in customer churn prediction.
Asthana et al. [5] presented a comparative analysis of machine learning methods using the UCL customer churn dataset. The algorithms of ANN, SVM, DT, Naïve Bayes (NB), and Linear Regression (LR) have been examined comparatively. 94% accuracy rate was obtained with ANN and DT algorithms.
Agrawal et al. [6] developed a deep learning-based customer churn prediction model using the Telco dataset. A multilayer neural network was developed to establish a non-linear classification model. The deep learning model obtained 80.03% accuracy.
Using the Telco dataset, Gaur and Dubey [7] conducted a machine learning-based comparative study for customer churn prediction. In the study, LR, SVM, Random Forest (RF), and Gradient Boosting algorithms have been compared. Experimental results showed that the Gradient Boosting algorithm has more successful with 0.845 Area Under Curve (AUC).
Halibas et al. [8] studied machine learning-based customer churn prediction using the Telco dataset. They compared NB, generalized linear models (GLM), LR, DT, RF, Gradient Boosting, and deep neural networks. Experimental results showed that gradient-enhanced trees are the best classifiers.
Kavitha et al. [9] conducted a machine learning-based customer churn prediction study using the Telco dataset. DT, RF, and eXtreme Gradient Boosting (XGBoost) algorithms have been compared in the study. Experimental results showed that RF is more successful than other models compared with 80% accuracy.
Lalwani et al. [10] conducted a machine learning-based customer churn prediction study using the Telco dataset. They used LR, DT, Adaboost classifiers, k Nearest Neighbour (kNN), RF, NB, SVM, XGBoost, and CatBoost classifiers. The CatBoost and AdaBoost classifiers obtained an accuracy rate of close to 82%.
Chabumba et al. [11] developed a customer churn prediction model using the Telco dataset. The model uses machine learning methods and a new feature selection method on the big data platform. Its AUC is 84%. The model has been compared with the LR, RF, SVM, and XGBoost algorithms. Experimental results showed that RF was the most successful model, with an 80% accuracy rate.

The reviewed studies in the literature are summarized in Table 1.

**Table 1.** Summary of studies in the literature

| Reference | Date | Methods | Dataset | Success rate |
|---|---|---|---|---|
| 4 | 2017 | ANN, SVM, DT | Iranian | %97.92 accuracy |
| 5 | 2018 | ANN, SVM, DT, NB, LR | UCL | %94 accuracy |
| 6 | 2018 | MLP | Telco | %80.03 accuracy |
| 7 | 2018 | LR, SVM, RF, Gradient Boosting | Telco | 0.845 AUC |
| 8 | 2019 | NB, GLM, LR, DT, RF, Gradient Boosting, MLP | Telco | %79.1 accuracy |
| 9 | 2020 | DT, RF, XGBoost | Telco | %80 accuracy |
| 10 | 2021 | LR, DT, Adaboost, kNN, RF, NB, SVM, XGBoost, CatBoost | Telco | %82 accuracy |
| 11 | 2021 | LR, RF, SVM, XGBoost | Telco | %80 accuracy |

Table 1 summarizes the studies in the literature examined according to their characteristics such as reference, publication date, methods used, dataset used, and success rate. The studies examined in the literature generally use machine learning methods. The algorithms used are generally LR, RF, DT, and SVM. Mainly, studies in the literature using the Telco dataset have been examined. The success rate in studies using the Telco dataset is approximately 80%.
In this study, a Multilayer Perceptron (MLP) based artificial neural network model was developed and compared with kNN, LR, NB, RF, and SVM. IBM Telco dataset, which consists of customer data of an anonymous telecommunications company, which is public access on Kaggle, was used as the dataset [12]. This study obtained a higher accuracy rate than the studies in the literature using the same dataset as the artificial neural network model developed. The contributions of this study to the literature can be summarized as follows:

- MLP-based artificial neural network model was created for customer churn prediction. In the existing literature, studies comparing MLP with other common models are limited. This study makes a significant contribution to the literature on performance evaluations in this field by examining in detail the performance of MLP in customer churn prediction.

- The developed MLP model was compared with commonly used models such as kNN, LR, NB, RF and SVM, and experimental results showed that the MLP model performed better with a higher accuracy rate (94%) compared to other models.
- The use of customer data obtained from an anonymous telecommunications company shows that the study is based on real-world data and is aimed at practical applications.
- In the study, anonymizing customer data and emphasizing ethical principles makes a significant contribution to the literature on how data privacy and ethical issues should be addressed in customer churn prediction models.

## CUSTOMER CHURN PREDICTION

Customer churn is a financial concept that pertains to a customer's discontinuation of engagement with a company or business. Likewise, the customer churn rate represents the pace at which customers depart from a business within a specified timeframe. Churn rate above a certain threshold can affect the business's success. Businesses want to retain as many customers as possible [13].

The churn rate serves as an indicator of customer contentment. A low churn rate signifies pleased customers, whereas a high churn rate indicates dissatisfied customers who have distanced themselves from the company [14]. Acquiring new customers will require much more effort and cost than retaining existing customers. Customer churn prediction is an indicator of growth potential for businesses. Churn rates represent the rate of lost customers, while growth rates represent newly acquired customers. Analysing these metrics gives information about the growth status of businesses. If the growth rate is higher than the loss rate, it can be said that the business is growing; otherwise, the business is shrinking [15].

The churn of customers applies in many contexts but generally relates to the business situation of customers who have stopped purchasing. Innovative customer retention strategies are more important for subscription-based and subscription-based business models. Analyzing growth in this area may involve monitoring parameters such as revenues and the proportion of new customers and performing customer analysis [16]. The churn rate quantifies the proportion of customer subscriptions or purchases that cease within a specified time frame for a particular service or product. Cancellation of customer subscriptions will naturally result in a loss of revenue. For this reason, examining the churn rate can help know customers and, for subscription-oriented businesses, effective marketing strategies to retain their customers [17].

Customer churn analysis can be evaluated as the churn of customers and revenue or as voluntary churn [18]. Customer churn represents the frequency at which customers discontinue their subscriptions. The churn of income refers to the loss in the monthly income of the enterprise. The churn of customers and loss of revenue may only sometimes be balanced [19]. The business may not lose customers, but there may be a loss of revenue due to the subscription tariffs that customers will change [20]. Damaging loss only applies to lost revenue. The generated revenue from existing customers surpasses the revenue lost due to cancellations and changes in subscription fees. Voluntary loss occurs when customers proactively terminate their service and take the necessary actions to discontinue their association [21]. This could result from customer discontentment or needing to realize the anticipated value. The churn of customers may occur due to poor customer service, financial issues, changes in customer needs, the service offered not meeting customers' expectations, or customers preferring rival companies [22].

## NEURAL NETWORK BASED CUSTOMER CHURN PREDICTION

Accurate customer churn prediction is paramount for telecommunications companies to retain their customer base proficiently. Obtaining new customers incurs higher costs compared to maintaining existing ones. Hence, prominent telecommunications companies endeavor to construct models that forecast which customers are at a higher risk of attrition and devise strategies based on these developed models. With the developments in data science and machine learning methods, the problem of identifying potential customers who may stop doing business with them soon comes to the fore.

This study developed an MLP-based model to predict how likely customers will abandon their business by analysing their demographic, account, and service information. The developed model aims to obtain a data-oriented solution that will reduce customer churn rates and increase customer satisfaction and operating income. The developed model was compared with kNN, LR, NB, RF, and SVM.

### Dataset

This study used the IBM Telco dataset consisting of customer data of an anonymous telecommunications company, which is public access on Kaggle. The dataset comprises 7043 rows and 21 columns, each corresponding to a customer in the dataset. The columns represent the individual attributes of each customer, which are utilized to predict the churn behavior of that specific customer. Of the features in the dataset, 17 features are categorical and 3 are numerical. Finally, there is the Churn attribute, expressed as Yes/no, in which it is predicted whether the customer will churn. The Churn column indicates the customer's departure status within the past month. "No" denotes customers who have not discontinued their association with the company during the last month, while "Yes" signifies

customers who have chosen to terminate their affiliation with the company.
The categorical attributes in the dataset can be described as follows:
• CustomerID: A unique customer ID number for each customer.
• gender: Indicates the gender of the customer as Female/Male.
• SeniorCitizen: Indicates whether the customer is a senior citizen or not, denoted as 1/0.
• Partner: Indicates Yes/No whether the customer is a business partner or not.
• Dependent: Indicates whether the customer is dependent or not, as Yes/No.
• PhoneService: Indicates whether the customer has phone service or not, denoted as Yes/No.
• MultipeLines: Indicates whether the customer has multiple phone lines or not, denoted as Yes/No/No phone service.
• InternetService: Indicates the customer's internet service provider type as DSL/Fiber optic/No.
• OnlineSecurity: Indicates whether the customer has online security or not, denoted as Yes/No/No internet service.
• OnlineBackup: Indicates whether the customer has online backup or not, denoted as Yes/No/No internet service.
• DeviceProtection: Indicates whether the customer has device protection or not, denoted as Yes/No/No internet service.
• TechSupport: Indicates whether the customer has technical support or not, denoted as Yes/No/No internet service.
• StreamingTV: Indicates whether the customer has streaming TV or not, denoted as Yes/No/No internet service.
• StreamingMovies: Indicates whether the customer has streaming movies or not, denoted as Yes/No/No internet service.
• Contract: Refers to the contract period of the customer, categorized as monthly, one-year, or two-year.
• PaperlessBilling: Refers to the contract period (monthly, one-year, two-years) on the customer's invoice.
• PaymentMethod: Indicates the customer's preferred payment method, categorized as mailed check, electronic check, credit card or bank transfer.
Numerical attributes can be described as follows:
• Tenure: Refers to the number of months the customer has been with the company.
• MonthlyCharges: It represents the monthly amount collected from the customer.
• TotalCharges: It represents the total amount collected from the customer.

The services subscribed by each customer are stored in the attributes: MultipleLines, PhoneService, OnlineSecurity, InternetService, DeviceProtection, OnlineBackup, TechSupport, StreamingMovies, StreamingTV. In Figure 1, customer churn rates for the attributes of each customer's registered service are shown.
Demographic customer information is kept in the fields of gender, SeniorCitizen, Partner, Dependents. Figure 2 shows customer churn rates for demographic attributes.

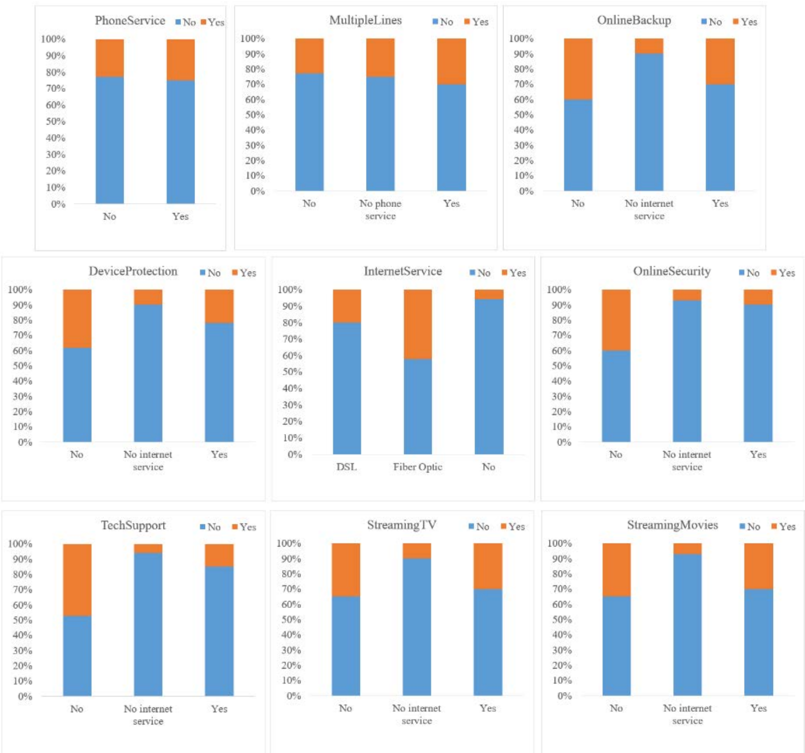


**Figure 1.** Customer churn rates for attributes of each customer's registered service

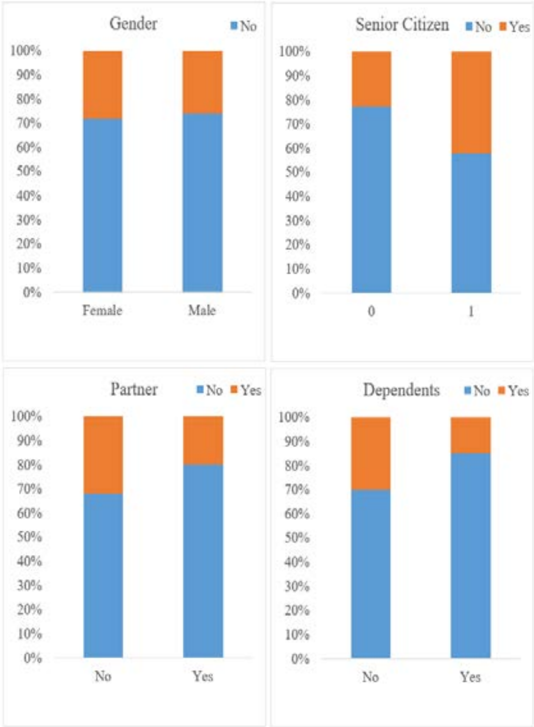


**Figure 2.** Customer churn rates for demographic attributes

Customer account information is kept in Contract, PaperlessBilling and PaymentMethod attributes. In Figure 3, the customer churn rates for the attributes of each customer account information are shown.

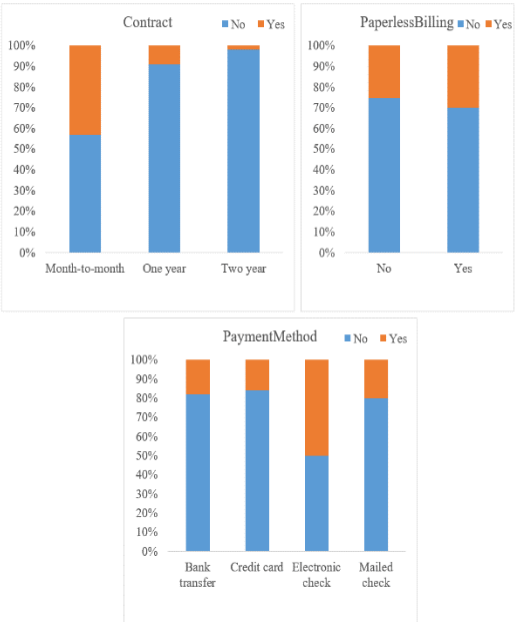


**Figure 3.** Customer churn rates for attributes of customer account information

According to the Churn attribute in the dataset, the Yes/No type customer churn rate is shown in Figure 1. Customer churn rates by churn attribute are shown in Figure 4.

As seen in Figure 4, the churn rate value for customers classified as 'No' in the Churn attribute is 0.734, while the churn rate value for customers classified as 'Yes' is 0.265.

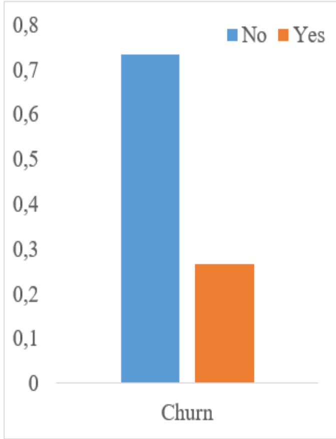


**Figure 4.** Customer churn rates by Churn attribute

## Data Pre-processing

The utilized dataset comprises 7043 rows and 21 columns. Among the features within the dataset, 17 are categorical, and 3 are numerical. The final column encompasses the dependent attribute, Churn. As seen in Table 2, each column is filled with non-empty values and has a data type.

**Table 2.** Attributes in the dataset

| # | Column | Non-Null Count | Dtype |
|---|---|---|---|
| 0 | customerID | 7043 non-null | object |
| 1 | Gender | 7043 non-null | object |
| 2 | SeniorCitizen | 7043 non-null | Int64 |
| 3 | Partner | 7043 non-null | object |
| 4 | Dependents | 7043 non-null | object |
| 5 | tenure | 7043 non-null | Int64 |
| 6 | PhoneService | 7043 non-null | object |
| 7 | MultipleLines | 7043 non-null | object |
| 8 | InternetService | 7043 non-null | object |
| 9 | OnlineSecurity | 7043 non-null | object |
| 10 | OnlineBackup | 7043 non-null | object |
| 11 | DeviceProtection | 7043 non-null | object |
| 12 | TechSupport | 7043 non-null | object |
| 13 | StreamingTV | 7043 non-null | object |
| 14 | StreamingMovies | 7043 non-null | object |
| 15 | Contract | 7043 non-null | object |
| 16 | PaperlessBilling | 7043 non-null | object |
| 17 | PaymentMethod | 7043 non-null | object |
| 18 | MonthlyCharges | 7043 non-null | Float64 |
| 19 | TotalCharges | 7043 non-null | object |
| 20 | Churn | 7043 non-null | object |

To ensure optimum prediction performance of the developed models, handling missing or incorrect values by removing or replacing them with appropriate values is

very important. Only 11 NULL values have been detected in the TotalCharges column. In the TotalCharges column, which represents the total amount charged from the customer, any NULL value is populated with the average value obtained from the same column.
The dataset has no Internet service because some categorical attributes have more than two categorical values, such as Yes/No/No Internet service. Values such as have been replaced with No. For example, the No phone or Internet service values have been replaced with No.
Since customer churn prediction involves a classification problem, converting the categorical data within the dataset into a numerical format is essential. For this reason, Yes for categorical variables has been replaced with 1 and No with 0. Similarly, in the Gender column, which represents the customers' gender, Male has been replaced with 1 and Female with 0.
Feature selection enables the identification of key features that are important in predicting the target feature. Table 3 shows some of the essential values of the features in the dataset.

**Table 3.** Properties and importance values in the dataset

| Features | Feature importance |
|---|---|
| InternetService_Fiber optic | 0.3268 |
| Contract_Month-to-month | 0.3090 |
| PaperlessBilling | 0.1658 |
| StreamingTV_Yes | 0.1342 |
| StreamingMovies_Yes | 0.1316 |
| OnlineSecurity_No | 0.1189 |
| PaymentMethod_Electronic check | 0.1138 |
| TechSupport_No | 0.0958 |
| MultipleLines_Yes | 0.0905 |
| SeniorCitizen | 0.0792 |
| OnlineBackup_No | 0.0535 |
| DeviceProtection_Yes | 0.0490 |
| DeviceProtection_No | 0.0230 |
| OnlineBackup_Yes | 0.0170 |
| PhoneService | 0.0019 |
| Partner | -0.0082 |
| gender | -0.0411 |

Table 3 shows that InternetService_Fiber optic and Contract_Month-to-month attributes have higher importance values, while gender and Partner attributes have much lower importance values.
Categorical values have been converted to numerical values using Label Encoding. Label encoding encodes a value in the range 0 to (n-1) into each row of data with a categorical value. Here n is the number of tags that differ. If a tag repeats, it will get the value previously assigned. In this way, it is ensured that categorical columns are expressed as numerical values.
Large numerical data must be scaled before the model is built, which can affect the model's performance. Normalization ensures that each input variable is scaled within a specified range. In this way, columns with large numeric values are scaled between -1 and 1 using MinMaxScaler.

**Developed Neural Network Based Prediction Model**

After pre-processing, the datasets were split into training, validation, and testing sets. The dataset was split into 80% for training and 20% for testing. Within the training set, 10% was further allocated for validation purposes, aiding in optimizing the model parameters. The resulting training dataset, obtained after separating training and testing sets, comprises 5634 rows, while the test dataset consists of 1409.
The selection of hyper-parameters enables evaluating the model's performance across various combinations of hyper-parameter values. Additionally, hyper-parameter tuning encompasses choosing the metrics or methods that yield optimal results based on a selected metric and validation approach. Grid search was used for hyper-parameter optimization. The flowchart of the developed system is shown in Figure 5.

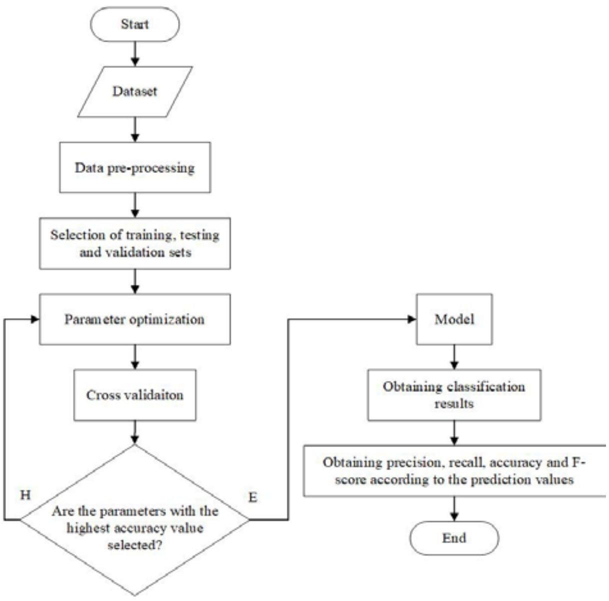


**Figure 5.** Flow chart of the developed system

The developed artificial neural network model takes customer data from the training dataset as input and generates an output indicating whether the customer is likely to churn or not for the customer data in the test dataset.
The developed MLP-based artificial neural network model consists of an input layer, hidden layers, and an output layer. The developed model has additional hidden layers compared to the traditional MLP model. Through the added layers, it is aimed that the model can better solve more complex and deep learning problems. Additionally, optimized ReLU activation functions were used for each layer. In order to prevent overfitting,

Dropout and L2 editing techniques are integrated. By using the dynamic learning rate, a faster and more stable optimization was achieved during the training. The connections between the layers represent the learned coefficients. The hidden layers are an intermediate processing step, combining weighted sums to derive the classification result. The developed model follows a sequential architecture with linear layers. The initial layer is the input layer, consisting of 64 input features and 64 output units. A Dropout layer is inserted between the input layer and the hidden layer. The output layer comprises a single unit that predicts the probability of customer churn. The ReLU activation function is applied to both the input and hidden layers, while the sigmoid activation function is employed in the output layer.
MLP is more successful than classical regression models because it calculates a weighted sum for each hidden unit when a nonlinear (ReLU) or hyperbolic (tanh) function is applied. The result of this function is then used to calculate the output. Learning nonlinear behaviours in data and learning models in real-time can be counted as advantages of MLP. However, the disadvantage of MLP is that a model with hidden layers has a loss function with more than one local minimum. The model requires parameter adjustment, such as the number of hidden neurons, iterations, and layers.

## EXPERIMENTAL RESULTS

This study compared kNN, LR, NB, RF, SVM, and the developed MLP-based artificial neural network model. The performance evaluation of the compared models was conducted using a confusion matrix. In this matrix, each column corresponds to the predicted classes. Each row represents the actual classes. TP, TN, FP, and FN values are obtained by using the confusion matrix. TP refers to samples predicted to be 1 and whose value is 0. TN denotes samples estimated to be 1 and whose value is 1. FP refers to samples predicted to be 1 and whose value is 0. FN represents the instances that are predicted as 0 but are 1. Accuracy, precision, recall, and F-score metrics were obtained by using TP, TN, FP, and FN values.
The accuracy calculates the proportion of correctly predicted values to the total test datasets. The accuracy metric is calculated using Eq.1.

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \tag{1}$$

The precision measures the proportion of positively predicted values that are truly positive. The precision metric is calculated using Eq. 2.

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

The recall metric determines how many of the values that should be predicted positively are predicted positively. The recall metric is calculated using Eq. 3.

$$Recall = \frac{TP}{TP + FN} \tag{3}$$

The F-score metric is determined by taking the harmonic mean of the precision and recall metrics. The F-score metric is calculated using Eq. 4.

$$F - score = 2 * \frac{Precision * Recall}{Precision + Recall} \tag{4}$$

The change in the accuracy value of the developed MLP-based artificial neural network model in the training steps called epoch is shown in Figure 6.

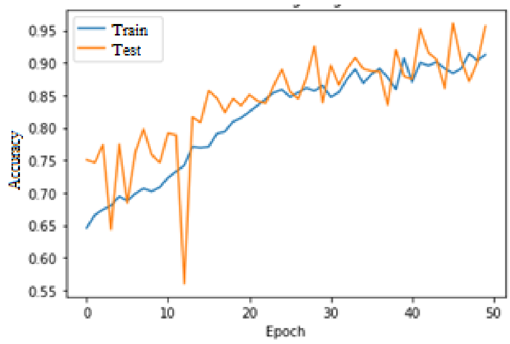


**Figure 6.** Variation of accuracy values according to epoch number

As seen in Figure 6, model accuracy varies according to the number of epochs. The model's performance is assessed by computing the average accuracy value during training.
Neural networks try to minimize the error. The function to be minimized or maximized is called the objective function or loss function, and the value calculated by the loss function is called loss. The Loss function determines how much the predictions the model produces differ from the true value. Figure 7 shows the variation in loss values as the number of epochs increases.

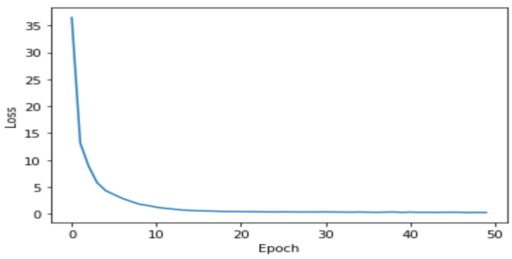


**Figure 7.** Change of loss values according to epoch number

kNN is a nonparametric method employed for both classification and regression tasks. Its fundamental principle involves identifying the neighboring data points, considering the test data point similar to these neighbors, and generating the output. In kNN, a specific number of k neighbors are searched, and predictions are made based on their characteristics.
The confusion matrix and experimental results for kNN are shown in Table 4 and Table 5.

**Table 4.** Confusion matrix for kNN

| | | Real values | |
|---|---|---|---|
| | | No | Yes |
| Predicted values | No | 204 | 144 |
| | Yes | 170 | 891 |

As seen in Table 4, the TP for kNN is 204, the FP is 144, the FN is 170, and the TN is 891.

**Table 5.** Accuracy, precision, recall and F-score values

| Accuracy | Precision | Recall | F-score |
|---|---|---|---|
| %77 | %58 | %54 | %56 |

As indicated in Table 5, the accuracy score for kNN is %77, the precision score is %58, the recall score is %54, and the F-score is %56.
LR is a statistical model that employs a logistic function to model a binary dependent variable. It is specifically designed for situations where the dependent variable is categorical. LR is particularly valuable in classification problems where the objective is to assess whether a new sample is most suitable for a particular category.
The confusion matrix and experimental results for LR are shown in Table 6 and Table 7.

**Table 6.** Confusion matrix for LR

| | | Real values | |
|---|---|---|---|
| | | No | Yes |
| Predicted values | No | 206 | 112 |
| | Yes | 168 | 923 |

As seen in Table 6, the TP value for LR is 206, the FP is 112, the FN is 168, and the TN is 923.

**Table 7**. Accuracy, precision, recall and F-score values

| Accuracy | Precision | Recall | F-score |
|---|---|---|---|
| %80 | %64 | %55 | %59 |

As indicated in Table 7, the accuracy score for LR is %80, the precision score is %64, the recall score is %55, and the F-score is %59.
The superior performance of LR compared to kNN can be attributed to their inherent characteristics as models. kNN is a non-parametric model that can handle non-linear solutions, whereas LR is a parametric model that primarily supports linear solutions. Additionally, LR can generate confidence levels for its predictions, whereas kNN provides only the output label without any associated confidence measure.
The NB algorithm is a classification technique based on Bayes Theorem. The NB classifier operates on the assumption that the presence of a specific attribute in a class is independent of the presence of any other attribute. NB is beneficial for massive datasets.
The confusion matrix and experimental results for NB are shown in Table 8 and Table 9.

**Table 8.** Confusion matrix for NB

| | | Real values | |
|---|---|---|---|
| | | No | Yes |
| Predicted values | No | 313 | 352 |
| | Yes | 61 | 683 |

As seen in Table 8, the TP for NB is 313, the FP is 352, the FN is 61, and the TN is 683.

**Table 9.** Accuracy, precision, recall and F-score values

| Accuracy | Precision | Recall | F-score |
|---|---|---|---|
| %70 | %47 | %83 | %60 |

As indicated in Table 9, the accuracy score for NB is %70, the precision score is %47, the recall score is %83, and the F-score is %60.
The fact that LR has better experimental results than NB can be interpreted as LR having better classification performance than NB in large datasets. NB works better on small datasets. LR outperforms NB on linearity as NB expects all features to be independent.
The critical distinction between kNN and NB lies in their classification approaches. kNN is a discriminative classifier, whereas NB is a generative classifier. NB assumes conditional independence among features and utilizes a maximum likelihood hypothesis. The superior classification performance of kNN compared to NB can be attributed to its non-parametric nature, whereas NB is considered parametric.
RF algorithm is an ensemble learning method that generates many decision trees during the training phase. In classification tasks, the RF outputs the class chosen by the majority of the trees. RF is a classifier that operates on multiple subsets of a dataset, employing a set of decision trees, and combines their results to enhance prediction accuracy. Unlike relying on a single decision tree, RF aggregates the predictions from each tree to produce the final output.
The confusion matrix and experimental results for RF are shown in Table 10 and Table 11.

**Table 10.** Confusion matrix for RF

| | | Real values | |
|---|---|---|---|
| | | No | Yes |
| Predicted values | No | 175 | 105 |
| | Yes | 199 | 930 |

According to Table 10, the RF classifier has a TP value of 175, an FP value of 105, an FN value of 199, and a TN value of 930.

**Table 11.** Accuracy, precision, recall and F-score values

| Accuracy | Precision | Recall | F-score |
|---|---|---|---|
| %78 | %62 | %46 | %53 |

As indicated in Table 11, the accuracy score for RF is %78, the precision score is %62, the recall score is %46, and the F-score is %53.
The reason why RF has more successful results than kNN is the values of the features in the dataset. RF assumes local similarities, and very similar samples are classified similarly. kNN can only select the most similar samples based on distance.
The superior performance of LR compared to RF can generally be interpreted as LR performing better when the number of noise variables is equal to or less than the number of explanatory variables. As the number of explanatory variables increases in a dataset, the ratio of TP to FP tends to increase for RF. The fact that RF has better experimental results than NB can be interpreted as RF being a distinctive model and NB being a productive model. Tree pruning in RF ensures that some features in the training data are neglected, thereby increasing the prediction accuracy.
SVM is a supervised learning model used for classification and regression analysis. It examines data and assigns new examples to specific categories based on training examples. SVM maps the training samples to space points to maximize the separation between the two categories. Subsequently, new samples are mapped to the same space and categorized based on which side of the separation they fall on.
The confusion matrix and experimental results for SVM are shown in Table 12 and Table 13.

**Table 12.** Confusion matrix for SVM

| | | Real values | |
|---|---|---|---|
| | | No | Yes |
| Predicted values | No | 184 | 101 |
| | Yes | 190 | 934 |

The SVM model achieved TP value of 184, FP value of 101, FN value of 190, and TN value of 934, as shown in Table 12.

**Table 13.** Accuracy, precision, recall and F-score values

| Accuracy | Precision | Recall | F-score |
|---|---|---|---|
| %79 | %64 | %49 | %55 |

As indicated in Table 13, the accuracy score for SVM is %79, the precision score is %64, the recall score is %49, and the F-score is %55.
Through kernel techniques, SVM offers support for both linear and non-linear solutions. In cases where training data is limited, SVM tends to handle outliers more effectively than LR. However, LR showed better classification performance than SVM due to the excess training data and the high number of features in the dataset.
In a classification problem, RF provides the probability of belonging to a particular class, whereas SVM provides the distance to the decision boundary. This characteristic often leads to SVM's superior performance compared to RF. SVM identifies support vectors in each class; the data points closest to the decision boundary separate the classes.
The fact that SVM has a better classification performance than kNN can be interpreted as the fact that SVM is more sensitive to outliers. When the number of training data samples dramatically exceeds the number of features, kNN can be more effective than SVM. However, in scenarios with numerous features and limited training data, SVM outperforms kNN.
MLP (Multi-Layer Perceptron) is an artificial neural network architecture consisting of an input layer, an output layer, and one hidden layer containing numerous interconnected neurons. MLP is a feed-forward model. The inputs are combined into a weighted sum with their initial weights and subjected to the activation function. Each layer feeds the next with the result of its calculations, an internal representation of the data. This process involves propagating information from the hidden layers to the output layer. Backpropagation is the learning mechanism that enables the MLP to iteratively adjust the network's weights in order to minimize the cost function. During each iteration, the gradient of the accuracy is calculated across all input and output pairs, following the transmission of weighted sums throughout the layers. Subsequently, the first hidden layer weights are updated using the gradient value for backpropagation. This propagation of weights continues until reaching the starting point of the neural network.
he confusion matrix and experimental results for MLP are shown in Table 14 and Table 15.

**Table 14.** Confusion matrix for MLP

| | | Real values | |
|---|---|---|---|
| | | No | Yes |
| Predicted values | No | 305 | 6 |
| | Yes | 69 | 1029 |

As seen in Table 14, the TP value for MLP is 309, the FP is 2, the FN is 74, and the TN is 1024.

Table 15. Accuracy, precision, recall and F-score values

| Accuracy | Precision | Recall | F-score |
|---|---|---|---|
| %94 | %99 | %80 | %89 |

As indicated in Table 15, the accuracy score for MLP is %79, the precision score is %65, the recall score is %49, and the F-score is %56.

MLP is a deep neural network architecture. MLP is a feedforward neural network architecture. It does not form loops like iterative neural networks. MLP uses backpropagation to train the network. In MLP, each new layer is a nonlinear function of the weighted sum of all outputs from the previous one. The fact that the experimental results of MLP are more successful than the other models can be interpreted as the fact that MLP is a feedforward network. Compared to SVM and RF, MLP requires many input data. The more data fed into the network, the better the network will generalize and make accurate predictions with fewer errors. On the other hand, SVM and RF require much less input data. For this reason, MLP was more successful than SVM and RF in the dataset used.

Table 16 and Figure 8 present the comparative experimental results for kNN, LR, NB, RF, SVM, and the MLP-based artificial neural network model based on accuracy, precision, recall, and F-score values.

**Table 16.** Comparative experimental results

| Model | kNN | LR | NB | RF | SVM | MLP |
|---|---|---|---|---|---|---|
| Accuracy | %77 | %80 | %70 | %78 | %79 | %94 |
| Precision | %58 | %64 | %47 | %62 | %64 | %99 |
| Recall | %54 | %55 | %83 | %47 | %49 | %80 |
| F-score | %56 | %59 | %60 | %53 | %55 | %89 |

Table 16 and Figure 8 illustrate the comparative results of different models, where the MLP-based artificial neural network model exhibits superior performance compared to other models. The MLP achieves 0.946 accuracy, 0.993 precision, 0.806 recall, and 0.890 F-score.

Following MLP, LR, SVM, RF, and kNN demonstrate relatively successful results. LR achieves 0.801 accuracy, 0.647 precision, 0.550 recall, and 0.595 F-score. SVM achieves 0.793 accuracy, 0.645 precision, 0.491 recall, and 0.558 F-score. RF achieves 0.784 accuracy, 0.625 precision, 0.467 recall, and 0.535 F-score. kNN achieves 0.777 accuracy, 0.586 precision, 0.545 recall, and 0.565 F-score. NB achieves 0.706 accuracy, 0.470 precision, 0.836 recall, and 0.602 F-score.

a

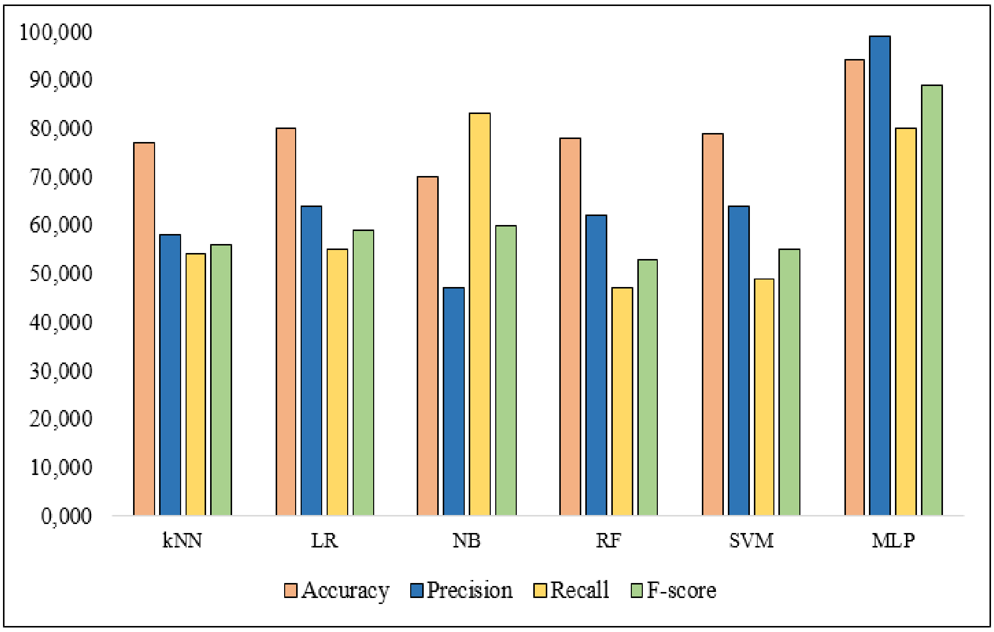


**Figure 8.** Comparative experimental results

As seen in Table 16 and Figure 8, MLP showed a better classification performance in customer churn prediction compared to other models. LR, SVM, RF and kNN are the models with the most successful results after MLP. NB has the worst classification performance among the compared models.

## CONCLUSIONS

Customer churn prediction means determining which customers are likely to leave a particular service or cancel a service subscription. After identifying customers at risk of canceling their subscription, a marketing strategy can be determined to maximize the customer's chances of staying subscribed. The churn of customers is a significant problem for businesses in most industries. For businesses to grow, they need to invest in gaining new customers. Every lost customer means a significant lost investment. For this reason, predicting when customers will leave and offering incentives can significantly save businesses.

The present study applied exploratory data analysis and feature extraction techniques to anonymous customer data obtained from a telecommunications company. The study compared the prediction performance of various machine learning and artificial neural network classifiers. The results revealed that all the models achieved exceeding 70% accuracy. The MLP-based artificial neural network model exhibited the highest classification performance among the compared models. The developed model achieved %94 accuracy, %99 precision, %80 recall, and %89 F-score. LR achieved %80 accuracy, %64 precision, %55 recall, and %59 F-score. SVM achieved %79 accuracy, %64 precision, %49 recall, and %55 F-score. RF achieved %78 accuracy, %62 precision, %46 recall, and %53 F-score. kNN achieved %77 accuracy, %58 precision, %54 recall, and %56 F-score. NB achieved %70 accuracy, %47 precision, %83 recall, and %60 F-score.

The experimental findings demonstrated that the developed model performed better than other models across all evaluation metrics. Moreover, all classifiers demonstrated significant performance enhancements when the oversampling technique was employed. The results show that the developed model can be preferred for customer churn prediction.

Using customer data to predict customer churn can have significant ethical implications. In this study, great importance was given to data privacy and respect for customer rights. Various verification and auditing processes have been applied to ensure that the predictions of the model developed using anonymized customer data are free of bias and fair. By adhering to ethical principles, we aimed to prevent the misuse of customers' data and improve service quality.

## REFERENCES

[1] Pondel, M., Wuczyński, M., Gryncewicz, W., Łysik, Ł., Hernes, M., Rot, A., Kozina, A. Deep learning for customer churn prediction in e-commerce decision support. In Business Information Systems. 3-12, 2021.

[2] Cenggoro, T. W., Wirastari, R. A., Rudianto, E., Mohadi, M. I., Ratj, D., Pardamean, B. Deep Learning as a Vector Embedding Model for Customer Churn. Procedia Computer Science. 179, 624-631, 2021.

[3] Pustokhina, I. V., Pustokhin, D. A., Nguyen, P. T., Elhoseny, M., Shankar, K. Multi-objective rain optimization algorithm with WELM model for customer churn prediction in telecommunication sector. Complex & Intelligent Systems. 1-13, 2021.

[4] Khodabandehlou, S., Rahman, M. Z. Comparison of supervised machine learning techniques for customer churn prediction based on analysis of customer behavior. Journal of Systems and Information Technology, 2017.

[5] Asthana, P. A comparison of machine learning techniques for customer churn prediction. International Journal of Pure and Applied Mathematics. 119(10), 1149-1169, 2018.

[6] Agrawal, S., Das, A., Gaikwad, A., Dhage, S. Customer churn prediction modelling based on behavioural patterns analysis using deep learning. In 2018 International conference on smart computing and electronic enterprise (ICSCEE). 1-6, 2018.

[7] Gaur, A., Dubey, R. Predicting Customer Churn Prediction In Telecom Sector Using Various Machine Learning Techniques. In 2018 International Conference on Advanced Computation and Telecommunication (ICACAT). 1-5, 2018.

[8] Halibas, A. S., Matthew, A. C., Pillai, I. G., Reazol, J. H., Delvo, E. G., Reazol, L. B. Determining the intervening effects of exploratory data analysis and feature engineering in telecoms customer churn modelling. In 2019 4th MEC International Conference on Big Data and Smart City (ICBDSC). 1-7, 2019.

[9] Kavitha, V., Kumar, G. H., Kumar, S. M., Harish, M. Churn prediction of customer in telecom industry using machine learning algorithms. International Journal of Engineering Research and Technology (IJERT). 9(5), 181-184, 2020.

[10] Lalwani, P., Mishra, M. K., Chadha, J. S., Sethi, P. Customer churn prediction system: a machine learning approach. Computing, 1-24, 2021.

[11] Chabumba, D. R., Jadhav, A., Ajoodha, R. Predicting telecommunication customer churn using machine learning techniques. In Interdisciplinary Research in Technology and Management. 625-636, 2021.

[12] Telco Customer Churn Dataset, 2018, Available: https://www.kaggle.com/blastchar/telco-customer-churn.

[13] Domingos, E., Ojeme, B., Daramola, O. Experimental Analysis of Hyperparameters for Deep Learning-Based Churn Prediction in the Banking Sector. Computation. 9(3), 34, 2021.

[14] De, S., Prabu, P., Paulose, J. Effective ML Techniques to Predict Customer Churn. In 2021 Third International Conference on Inventive Research in Computing Applications (ICIRCA). 895-902, 2021.

[15] Vo, N. N., Liu, S., Li, X., Xu, G. Leveraging unstructured call log data for customer churn prediction. Knowledge-Based Systems, 212, 106586, 2021.

[16] Karamollaoğlu, H., Yücedağ, İ., Doğru, İ. A. Customer Churn Prediction Using Machine Learning Methods: A Comparative Analysis. In 2021 6th International Conference on Computer Science and Engineering (UBMK). 139-144, 2021.

[17] Veningston, K., Rao, P. V., Selvan, C., Ronalda, M. (2022). Investigation on Customer Churn Prediction Using Machine Learning Techniques. In Proceedings of International Conference on Data Science and Applications. 109-119, 2022.

[18] Dingli, A., Marmara, V., Fournier, N. S. Comparison of deep learning algorithms to predict customer churn within a local retail industry. International journal of machine learning and computing. 7(5), 128-132, 2017.

[19] Khattak, A., Mehak, Z., Ahmad, H., Asghar, M. U., Asghar, M. Z., Khan, A. Customer churn prediction using composite deep learning technique. Scientific Reports. 13(1), 17294, 2023.

[20] Saha, L., Tripathy, H. K., Gaber, T., El-Gohary, H., El-kenawy, E. S. M. Deep churn prediction method for

telecommunication industry. Sustainability. 15(5), 4543, 2023.

[21] Amin, A., Adnan, A., Anwar, S. An adaptive learning approach for customer churn prediction in the telecommunication industry using evolutionary computation and Naïve Bayes. Applied Soft Computing. 137, 110103, 2023.

[22] Prabadevi, B., Shalini, R., Kavitha, B. R. Customer churning analysis using machine learning algorithms. International Journal of Intelligent Networks. 4, 145-154, 2023.