İSTANBUL UNIVERSITY PRESS

Acta Infologica,
ACIN 2025, 9 (1): 55–73

https://doi.org/10.26650/acin.1584030

Submitted: 29.11.2024
Revision Requested: 31.12.2024
Last Revision Received: 07.01.2025
Accepted: 21.02.2025
Published Online 27.02.2025

Acta Infologica

Research Article

Open Access

A Novel and Robust LSTM Model for Customer Churn Analysis Using Deep, Machine Learning, and Ensemble Learning: A Telecommunications Case

Muhammet Sinan Başarslan [1], Aslıhan Ünal [2] & Fatih Kayaalp [3]

[1] İstanbul Medeniyet University, Faculty of Engineering and Natural Sciences, Computer Engineering, İstanbul, Türkiye
[2] İstanbul Gelişim University, Faculty of Economics, Administrative and Social Sciences, Department of Business Administration, İstanbul, Türkiye
[3] Düzce University, Faculty of Engineering, Computer Engineering, Düzce, Türkiye

Abstract Customer churn is an important issue in increasing both the long- and short-term revenues. If companies identify customers' churn behavior, they can prevent churn, ensure customer loyalty, and, in turn, gain better financial returns. The telecommunications sector is a customer-oriented sector that requires customer retention to survive in the market. In this sector, customer churn is observed at a high level. In recent years, artificial intelligence-based customer churn analysis has been widely used to predict customer churn behavior. In this study, a customer churn analysis was conducted using publicly shared Telco telecommunications data. Predictive models were constructed using machine learning (LR, KNN, SVM, DT, RF, ANN), ensemble learning (XGBoost, Majority Voting), and deep learning (LSTM) methods. In addition, a 3-layered LSTM model was proposed. Accuracy (Acc), F1-score (F1), Precision (Prec), and Recall (Rec) rates were used to evaluate the models. As a result, the novel 3-layered LSTM model achieved 91.90% Acc, 91.49% Prec, 92.31% Rec, and 91.90% F1 values. The proposed model is competitive with the existing models.

Keywords Customer Churn Analysis · Ensemble Learning · Machine Learning · Deep Learning · Telecommunication

Citation: Başarslan, M. S., Ünal, A. & Kayaalp, F. (2025). A novel and robust LSTM model for customer churn analysis using deep, machine learning, and ensemble learning: A telecommunications case. *Acta Infologica*, *9*(1), 55-73. https://doi.org/10.26650/acin.1584030

Corresponding author: Muhammet Sinan Başarslan muhammet.basarslan@medeniyet.edu.tr

Introduction

Incremental developments in Information Technologies (IT) and the Internet have led to radical changes in consumer behaviors, organizational structures, and business processes. The new age is based on a knowledge-based view (KBV), and in this age, transforming data into information for decision-making is a vital strategic tool. In data-based decision-making, executives use the outcomes of predictive data analysis to make predictions and provide recommendations for current and future activities (Kozak et al., 2021).

The rise of e-commerce in the mid-1990s triggered fierce competition. In such an environment, retaining customers has become even more difficult for companies. The emergence of online businesses has led to new revenue models. Web catalogs, digital content, advertising-supported, advertising-subscription mixed, and fee-based models are new sources of income for online businesses (Gary Schneider, 2016). E-commerce companies require appropriate methods for predicting consumer behavior and decreasing customer churn rates to survive or gain a competitive advantage in the market. Customer churn analysis is a widely employed method in the sectors that adopt subscription-based income models as telecommunication, insurance, and banking (Osmanoglu, 2019).

In today's world, speed, easiness, user-friendly applications, and websites have a positive impact on consumer experience by providing access to information in the right place at the right time. These positive user experiences play an important role in consumers' buying decisions (PwC consumer Experience Series, n. d.). Customers can react in three dimensions when their expectations are not satisfied or they cannot experience the user experience they desire: Exit, voice, and loyalty. Customer churn is the behavior a customer engages in to end his/her relationship with the company (e.g., unsubscribe, transfer, or suspend the subscription), and when the customer leaves the company and starts to buy another company's product, it is called customer churn (Fu, 2022). In the case of customer churn, the customer can be convinced to stay with the company by implementing customized marketing strategies (Cao et al., 2019).

The saturation of markets due to globalization gave rise to a highly competitive environment in the business world. As a result, the conversion costs of customers decreased, and the customer acquisition costs of companies increased (Ahn et al., 2020). Because customer loyalty directly affects company revenues, it is one of the primary goals of customer relationship management (CRM). As acquiring new customers is more costly than retaining existing customers, companies conduct customer churn prediction analysis (Karvana et al., 2019).

Customer churn analysis is used widely in various sectors to increase the level of predictability (Karvana et al., 2019). and plays an important role in the sectors that adopt a subscription-based revenue model as telecommunication, banking, and insurance (Osmanoglu, 2019). The telecommunications sector witnesses high churn rates in its highly competitive market environment (Su et al., 2022), and identifying churn risk is required for retaining existing customers and gaining a competitive advantage (Edwine et al., 2022).

Customer churn in the telecommunications sector can be defined as customers who tend to unsubscribe from a company's service. Telecommunication companies need to know the reasons for churns, and this information can be extracted from the data of the customers (Gaur & Dubey, 2018). Customer churn behavior in the telecommunication sector came into prominence in the 2000s in the literature, and "churn prediction and modeling" have become one of the most examined areas (Bhattacharyya & Dash, 2022). Research in this field includes building classification models using traditional machine learning methods (eg. Decision Tree-DT, Random Forest- RF, Naïve Bayes-NB, Logistic Regression-LR, Support Vector Machine-SVM) and/or deep learning algorithms (e.g.. Artificial Neural Network-ANN, Convolution Neural Network (CNN), and determining the Acc levels of the models using various evaluation metrics (Mishra & Reddy, 2017), (Indulkar & Patil, 2021). Because of developments in big data technology, models based on deep learning algorithms increase prediction efficiency by reducing processing time and providing higher accuracy (Mishra & Reddy, 2017).

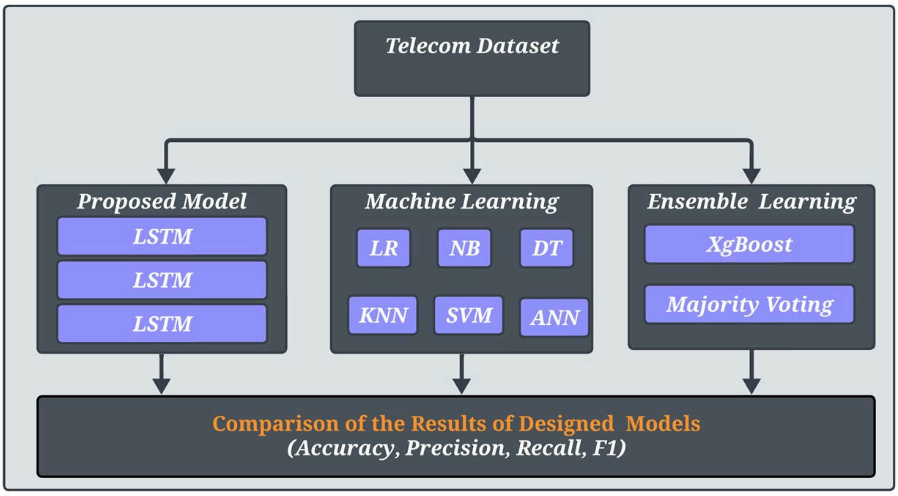
The performance of predictive models based on traditional machine learning algorithms, deep learning algorithms, and hybrid algorithms has played an important role in the literature. For example, Gaur and Dubey compared the performance of several classification models (LR, SVM, RF, and gradient-boosted tree) (Gaur & Dubey, 2018). Findings show that the Gradient boosting model is the best model with a performance of 84.57% AUC. In addition, in Zatonatska et al.'s research, all four models were defined as "high quality" with Acc rates of over 80%. have Acc rates (Zatonatska et al., 2023). Jeyaprakaash and Sashirekha compared predictive models based on Adaboost and K-Nearest Neighbor (KNN) algorithms and as a result, the Adaboost model with an accuracy level of 90% was defined as the best predictive model. (Jeyaprakaash & Sashirekha, 2022). Bayrak et al. compared two deep learning models based on CNN and ANN algorithms, and as a result, the CNN model with a 97.62% accuracy rate was defined as the best predictive model. Bayrak et al. compared a hybrid model based on CNN and long short-term memory (LSTM) algorithms to an RNN model and found that the hybrid model demonstrated the best performance model (Bayrak et al., 2021).

If churn customers can be predicted as much as possible as earlier and accurately, it will be possible to plan new marketing offers, for whom customers who are considering leaving can be predicted in advance and accurately, making it possible to plan and implement new marketing campaigns and offers to increase the loyalty of these customers. Therefore, the classification performance of churn customer analysis is a crucial issue. Various factors affect model success. The types of attributes in the dataset, missing/incorrect data, conformity of data types to the algorithms used, and parameter optimizations of the algorithms are some of these factors. Hence, the processes related to these factors should be examined through various experiments, and appropriate models should be constructed. To objectively evaluate the success of the models, it would be useful to compare them with the findings of similar research in the literature.

In this study, experiments were conducted on an open-source Telco dataset to compare the performances of the widely used machine learning, deep learning, and ensemble learning methods for predicting customer churn. The experimental results were evaluated according to the Acc, Prec, and F1 metrics. As a result, we observed that the LSTM deep learning method, which is an emerging technology, provides better results than traditional machine learning methods. The flow diagram of the study is shown in Figure 1.

Figure 1

Flow diagram of the study



Customer Churn Management

Customer churn management (CCM) aims to predict the customers who most likely intend to transfer to competing service providers and convince them to stay with the company by providing customized offers to these customers. (Hadden et al., 2007). There are two kinds of churn customers for companies (Wanchai, 2017): voluntary and involuntary. Involuntary churns include cases in which customers are unwilling to quit. Death, relocation, or termination of subscription due to nonpayment of invoices are examples of involuntary churns. Voluntary churn is the churn of customers because they voluntarily give up the company's goods and services (Ahn et al., 2020). Voluntary customers are also called "unhappy customers" (Kurt et al., 2019).

Research shows that the cost of acquiring a new customer is higher than retaining existing ones. Therefore, customer churn management focuses on convincing voluntary customers to continue to use the company's services. Involuntary churns are excluded from consideration because they are out of the company's control (Wanchai, 2017). Customer churn prediction is performed in various businesses to increase the revenue of the company by increasing the retention rate of customers (Ahn et al., 2020). Telecommunication, banking, insurance, gaming, music streaming, newspaper subscriptions, social media-based services, public relations, and energy are examples of industries where customer churn analysis is utilized (Ahn et al., 2020). Particularly, customer churn prediction has a vital role in industries where value is directly related to active customer amount and requires customer retention in telecommunication, banking, and insurance (Sadi Evren Şeker, 2016). Hence, companies struggle to prevent customer churn by using data mining methods when developing strategies to increase customer loyalty (Guo & Qin, 2017).

The success of CCM is related to predicting voluntary churns and understanding the underlying reasons (Routh et al., 2021). To achieve this goal, companies use data science and machine learning methods to define customer satisfaction, reveal the reasons for dissatisfaction, and predict customer churn (Zatonatska et al., 2023). Predictive models play a key role in developing customer churn strategies and sustaining customer loyalty (Lalwani et al., 2022).

Customer Churn Analysis

Customer churn analysis (CCA), a subfield of customer analytics (Seymen et al., 2023), aims to predict customer churn based on historical data and behavioral patterns and to convince customers by providing customized offerings (Karvana et al., 2019).

Data mining methods and machine learning algorithms are traditionally used in CCA. Developments in big data technology have made it possible to build predictive models based on deep learning and obtain real-time warnings (S. Cao et al., 2019). Data mining is the process of exploring relationships in the data without relying on assumptions or premises. This feature distinguishes it from traditional data analysis (Guo & Qin, 2017). Data mining can be used to classify data into series, patterns, and trends (descriptive mining) and to predict by processing data with various algorithms (predictive mining) (Wanchai, 2017). The behavior of churn customers can be extracted from the data analyzed using data mining methods (Ullah et al., 2019).

A predictive model for churn analysis follows five steps (Wanchai, 2017). The first step involves collecting appropriate data for analysis. Examples of data used in CCA include demographic data, customer account information, call data in the telecommunications sector (Wanchai, 2017), and sales data in the retail sector (Bayrak et al., 2021). In the second step, data are organized and analyzed to explore patterns and relationships. First, a threshold value should be defined to distinguish "churn customer" from "ordinary customer" (Bayrak et al., 2021). The time frame when several customers are inactive is also referred to as the "time window". Customers who exceed this threshold are identified as "churn customers". The time window period can change due to the type of services provided by the company (Ahn et al., 2020). Then, training data are prepared for data mining. (Bayrak et al., 2021). The training data can be used as input to a predictive model (Karvana et al., 2019). In the third step, an appropriate data mining method is selected for the given data. Several machine learning methods frequently used in this process are LR, SVM, RF, gradient boosted tree (GBT), KNN algorithms, DT, ANN, rough set approach, linear discriminant analysis, market basket analysis, and sequential pattern mining, NB (Gaur & Dubey, 2018) (Kurt et al., 2019). Data mining can be performed in two ways: supervised and unsupervised. In CCA, the supervised learning method is preferred to classify churning and nonchurning customers (Wanchai, 2017). In the final step, a predictive model is built based on the preferred data mining method, and the results are evaluated and verified in terms of usability and accuracy.

Literature

Essentially, customer churn analysis is a classification problem. Hence, there are many classification-methods-based CCA research in the literature. in the literature. Ullah et al. analyzed two samples of telecom customer data using the Bayes method and found that the test results ranged from 51% to 82.11% depending on the type of sample and iteration level (Amin et al., 2019). Arifin and Samopa conducted customer churn classification research in Indonesia using the SVM method. As a result, the predictive model achieved an 80% accuracy rate. (Arifin & Samopa, 2018). In addition, De Caigny et al. combined LR and DT methods and proposed a hybrid model. Acc rates for individual classifiers were defined as 81.6% for LR and 79.6% and 79.6% for RF, respectively, and with the hybrid model, the result increased to 86% (De Caigny et al., 2018). Coussement et al. built classification models based on various machine learning algorithms such as RF, SVM, DT, SGB, and NB, and they analyzed the customer data of a large telecom service provider in Europe. The Acc levels of the models were found as follows. NB-62%, RF = 63%, SVM = 53%, and SGB = 61%. (Coussement et al., 2017). Zhang et al. The LR, DT, and ANN methods were analyzed to classify churn customers in the

South Asia telecom sector, and as a result, the ANN model achieved a 75.53% Acc level (Zhang et al., 2010). In Azeem et al.'s study, a model based on fuzzy classifiers was applied to a dataset of 600.000 subscribers with 84 attributes, and as a result, the model provided a 57% Acc level (Azeem et al., 2017). Qureshi et al analyzed telecom customer data with LR-model prediction and predicted customer churn with a 78% Acc level (Qureshi et al., 2013). Varan and Bahara's NB model achieved an acc of 68%.

Wanchai analyzed the customer data of a telecommunication company in Thailand using the ANN model and obtained an acc level of 86.13% (Wanchai, 2017). Zatonatska et al. built classification models with LR, RF, SVM, and Extreme Gradient Boosting (XGBoost) and analyzed Telco Telecom customer data using a Python program. As a result, the performance of the models for Acc rates was defined as follows: SVM-82%, RF-80.9%, LR-80.8%, and XGBoost-83% (Zatonatska et al., 2023). Lalwani et al. proposed a six-layered customer churn model to analyze telecommunications customer data. The authors performed preprocessing and attribute selection in the first three layers and selected attributes using a gravitational search algorithm in the third layer. They separated the data into training and test data using the data holdout method (80% and 20%) in the fourth layer. In the last layer, the models were evaluated using the AdaBoost and XGBoost methods. As a result, the AdaBoost model performed at 81.71% acc rate, and the XGBoost model at 80.8% acc rate (Lalwani et al., 2022).

Gaur and Dubey built LR, SVM, RF, and GBT classification models for churn prediction. The data holdout method was applied to the data with rates of 75% and 25%, and as a result, the Acc rates of the models were defined as follows: LR-82.86%; SVM-79.75%; RF-81.26%, and GBT-84.59% (Gaur & Dubey, 2018). Seymen et al. We performed customer churn analysis in the telecommunications industry using an RF model and achieved an Acc rate of 88.63% (Ullah et al., 2019).

The RNN and LSTM deep learning architectures are widely used in many fields due to their high performance and compatibility. For example; energy forecasting of wind turbines (Çelebi & Fidan, 2024), sales forecasting in e-commerce (Ecevit, et.al, 2023), and energy forecasting generated by wind power in SCADA systems (Çelebi & Karaman, 2023). Although LSTM is widely used in classification problems, it has not been applied in the abovementioned churn analysis studies. In order to investigate the performance of LSTM in this domain, we build an LSTM-based model using the proposed method.

Material and Method

This section provides information about the dataset used in this study and the machine learning, deep learning, and ensemble learning methods applied to the dataset.

Dataset

The experiments were conducted on the Telco dataset (Zhuang, 2022), which is an open-source dataset of customer data from a telecommunications company in California in the second quarter of 2022. The Telco customer churn dataset contains data about a fictional telecommunications company operating in California during the third quarter. This dataset comprises 7043 customer records, including customer retention, departure, and sign-ups for services. This dataset contains 21 attributes. Table 1 lists the attributes and their descriptions (Zhuang, 2022).

Table 1

Attributes and their descriptions

Attribute Name	Description
Customer Id	Customer ID
Gender	Gender of the customer
SeniorCitizen	Customers' senior citizenship status
Partner	Does the client have a partner?
Addicts	The presence of dependents
Mission time	The month that he stayed with the company
PhoneService	Telephone service subscription
MultipleLines	Multiple lines
Internet Service	Internet service provider
OnlineSecurity	Online safety status
OnlineBackup	Online backup status
Device Protection	Device protection ownership status
Tech Support	Availability of technical support
StreamingTV	TV subscription status
Streaming Movies	Movie subscription status
Agreement	Contract duration
Paperless Billing	E-billing status
payment method	Customers' payment methods
MonthlyFees	Amount to be paid by the customer per month
TotalCharges	Total fees paid by the customer
Churn	Customer churn status

The dataset was preprocessed for missing data and outliers. Missing data are only present in the Total Charges attribute. This missing data was completed by assigning the median value of the relevant attribute, as is commonly done in the literature (Hanif et al 2017, Mokhairi et al 2016).

Machine Learning

Machine learning is a branch of artificial intelligence that enables computer systems to learn from data, recognize patterns, and make decisions. Machine learning is used in many fields today due to its ability to understand and utilize various data. Machine learning is a revolutionary approach to improve the data analysis and learning capabilities of computers. Machine learning has made solving complex problems more accessible and effective (Öztürk et al., 2022). The machine learning methods used in this study are described below.

Random Forest

RF was developed by Leo Breiman and combines the decisions of multiple trees trained on different training datasets rather than building a single DT (Breiman, 2001).. By creating multiple DTs, each trained on a different dataset, the decisions of these trees are combined, resulting in a more powerful and stable classifier. RF is a popular algorithm that provides successful results for various machine learning problems, such as classification and regression. It is also favored because of its ability to work effectively on high-dimensional and large datasets. RF is widely used in data analysis because it has various applications and is generally resistant to overfitting (Breiman, 2001).

Support vector machine

SVM can be defined as a vector space-based machine learning method that finds a decision boundary between two classes and places this decision boundary at the furthest point from any point in the training data. SVM is a structural risk minimization approach in statistical learning theory. It is noteworthy that SVM is a structural risk minimization approach in statistical learning theory. One of the basic assumptions of the SVM is that all samples in the training set are independent of each other and are distributed similarly (Ahmed et al., 2023a). This property makes SVM an effective solution for linear and nonlinear data classification and regression problems. Due to its high performance and generalizability on low-dimensional and high-dimensional data sets, SVM has a wide range of applications and is a powerful tool for machine learning and data analysis (Ahmed et al., 2023a).

Decision tree

DT is an important tool in data analysis and machine learning. Used in several data mining processes, decision trees facilitate information extraction with the ability to transform complex data structures into simple decision rules. DTs provide an effective way to make data-driven decisions by analyzing features in a dataset (Uzun et al., 2012). The DT building process is a fundamental part of data mining processes. In the first step, the data are preprocessed, and feature selection is performed. Then, the tree-building process begins, where nodes are created to identify the best discriminative decisions based on the features in the dataset. However, the tree was pruned to prevent overfitting and ensure generalizability. Dts have yielded successful results in many fields, such as disease diagnosis in medicine, risk assessment in finance, and customer segmentation in marketing. As a result, DTs are an important tool that provides valuable information in data-driven decision-making processes and constitute a cornerstone of data analysis (Uzun et al., 2012).

Artificial Neural Networks

ANNs are artificial model systems inspired by the biological nervous system and provide impressive results in the fields of machine learning and data analysis. These artificial models can be used in various tasks due to their data processing and pattern recognition capabilities (Ustebay et al., 2019).

LR is a classification technique that is widely used in statistical analysis and machine learning. The proposed method, which is generally used in binary classification problems, attempts to estimate the probability of a data point belonging to a particular class. This estimation is realized by determining a threshold value of the probability values and classifying them. LR derives its name from the "logit" function. This function maps probabilities to an unlimited range of real numbers. The primary purpose of the model is to estimate the probability of a given event occurring based on the given input characteristics. Therefore, the output value is a probability value and lies between 0 and 1 (Baltsou et al., 2022).

At the heart of LR is the idea of finding a line or hyperplane (this line or plane is called the "decision boundary") that best separates the dataset. This line or plane attempts to best capture the separation between classes. The training process involves adapting the parameters (weights) of the model to the data. At the end of training, the model can be used to predict new data points (LaValley, 2008).

With a subfield called deep learning, deep structured networks can produce outstanding results by extracting features from large and complex datasets. ANNs play an important role in processing large amounts of data and understanding complex relationships (Öztürk et al., 2022).

Logistic Regression

LR is a classification technique that is widely used in statistical analysis and machine learning. The proposed method, which is generally used in binary classification problems, attempts to estimate the probability of a data point belonging to a particular class. This estimation is realized by determining a threshold value of the probability values and classifying them. LR derives its name from the "logit" function. This function maps probabilities to an unlimited range of real numbers. The primary purpose of the model is to estimate the probability of a given event occurring based on the given input characteristics. Therefore, the output value is a probability value and lies between 0 and 1 (Baltsou et al., 2022).

At the heart of LR is the idea of finding a line or hyperplane (this line or plane is called the "decision boundary") that best separates the dataset. This line or plane attempts to best capture the separation between classes. The training process involves adapting the parameters (weights) of the model to the data. At the end of training, the model can be used to predict new data points (LaValley, 2008).

K-Nearest Neighbor

The KNN algorithm is widely used for pattern recognition, classification, and regression. The proposed KNN algorithm has a simple structure. When a data point needs to be classified or predicted, the k nearest neighbors around it are determined. The class labels or values of these neighbors are examined. In the classification task, the class to which the largest number of neighbors belongs is selected. In the case of regression, the values of the neighbors are averaged, and prediction is performed (Ahmed et al., 2023b).

Deep Learning

Deep learning is a branch of machine learning that aims to automatically learn data analysis, pattern recognition, and predictive capabilities using complex model structures, such as artificial neural networks. By processing large datasets, the proposed method can identify more complex and high-level features. This can lead to more accurate results and better understanding (Zavrak & Iskefiyeli, 2023). Deep learning is characterized by making the layers of artificial neural networks wider and deeper. These neural networks use many intermediate layers to process the data, allowing more complex features to be discovered. The learning process involves automatically adjusting weights and features using large amounts of data. Deep learning attempts to extract features directly from the data. This makes it possible to learn more information and achieve better results. The application areas of deep learning are wide-ranging. It has been successfully used in many fields, such as image recognition, audio processing, natural language processing, medical diagnosis, and autonomous vehicles. For example, deep learning models can identify objects in images, understand text, and diagnose diseases (Zavrak & Yilmaz, 2023). The LSTM deep learning method is described in this section.

Long Short-Term Memory

LSTM is a powerful recurrent neural network developed for deep learning, especially for processing sequential data (Hochreiter & Schmidhuber,1997). LSTM has achieved great success in areas such as natural language processing, time series analysis, and video processing. The proposed model is designed to overcome the traditional methods for learning and understanding sequential data (Tan et al., 2022). The key feature of LSTM is its ability to understand long-term dependencies in sequential data. This determines the superiority of LSTM over regular Recurrent Neural Network (RNN) (Malik et al., 2022).

The LSTM cell has three components; Forget gate, Input gate, and Output gate. The forget gate controls the length of the forgotten cell state. The input gate sets the insertion of new information. The exit gate sets the time at which the updated cell state is used.

Ensemble Learning

Ensemble learning is used in many fields such as social networks, communication networks, and biological networks. With this method, groups within a network and their relationships with each other can be better understood. Ensemble learning algorithms identify communities by identifying similar nodes or grouping nodes according to the strength of their connections to each other. The ensemble learning methods used in this study are described below (Başarslan & Kayaalp, 2023).

Xgboost

XGBoost is an ensemble learning algorithm used in machine learning. It demonstrates high performance, especially in classification and regression problems. XGBoost offers a more powerful and faster solution than previous Gradient Boosting methods.

XGBoost works as a tree-based algorithm. The model combines multiple weak learners (usually DTs) to construct a robust prediction model. The algorithm performs the learning process with an approach that focuses on the sequential addition of trees and corrects the errors of previous trees (Chen & Guestrin, 2016).

Voting

Voting is a machine learning and statistical forecasting method. In this method, the predictions of different prediction models or classifiers are combined, and the result is the prediction of the majority vote. Voting was used to compensate for the weaknesses of the models used alone and to obtain more reliable predictions. There are two types: majority voting and probabilistic voting (J. Cao et al. (2012).

Majority voting (Hard Voting): This is a method in which most predictions produced by different models are taken. For example, we consider three different classifiers: A, B, and C. If models A and B predict an instance as Class 1, while model C predicts it as Class 2, the Majority Voting method predicts it as Class 1 (Mohammadifar et al., 2022).

Probabilistic Voting (Soft Voting): This method takes the average or weighted average of the forecast probabilities of different models. The proposed method can obtain more accurate predictions by considering the reliability levels of the models. The prediction probabilities of each model are combined, and the class with the highest probability is predicted (J. Cao et al., 2012; Mohammadifar et al. (2022)

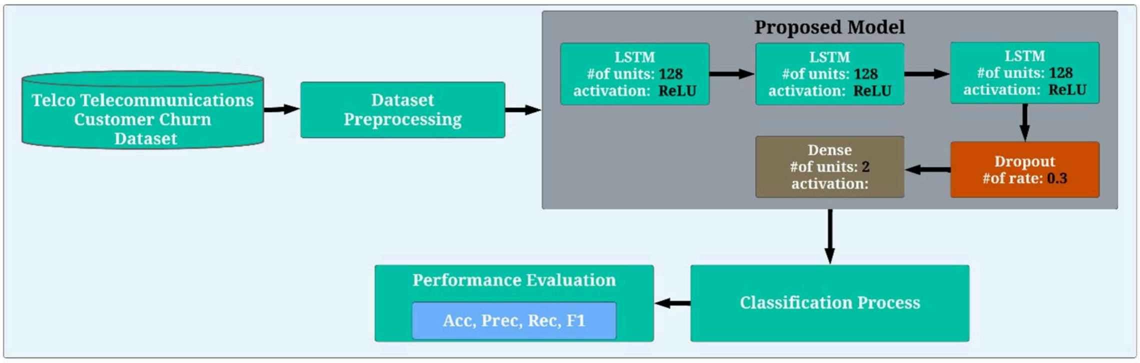
Voting ensemble learning is used to combine the predictions of different prediction models or classifiers to obtain more robust and reliable results. In this study, all machine learning methods were used with majority voting.

Proposed Model

In the proposed model, three LSTM neural networks were used. A graphical representation of this model is shown in Figure 2.

Figure 2

Proposed model



As shown in Figure 2, the Telco dataset was preprocessed for missing and outlier data, and then scaled and prepared for the classification process. The prepared dataset was then modeled with a 3-layer LSTM, dropout layer, and dense layer. The model is then evaluated using performance metrics. Experiments on the model were conducted on Google Colaboratory (Colab) using the TensorFlow 2.9.0 and Keras 2.9 libraries and Python version 3.9.13. The Colab Pro version was used for faster and uninterrupted results of the experiments. In addition, various callbacks were used to create a model with better results. These are ReduceLROnPlateau and Early stopping algorithms. In this study, hyperparameter optimization was performed using KerasTuner. Parameters such as epoch, batch size, activation function, optimization function, number of neural network nodes, and dropout rate are given in Table 2.

Table 2

Hyperparameters of proposed model

Parameters	**Values**
Batch Size	32, **64**,128
Epoch	**10**, 20, and 30
Activation Functions	**ReLU**, Sigmoid, Tanh,
Optimizaton Functions	**Adam**, RMSprop, and SGD
Number of Units	64, **128**, and 256
Learning Rate Starting Value	**0.001** for all optimizers
Loss Function	Binary cross-entropy
Dropout rate	**0.3**; 0.4; 0.5; 0.6

The parameters shown in bold in Table 2 represent the most optimum parameters with the KerasTuner model. The results of the proposed model with these parameters are given under the experimental results section. The comparison of this model with the results obtained using other machine and ensemble learning methods is also given under the same heading.

Experimental Results and Evaluation

In this section, we describe the metrics used to compare model performance and obtained results.

Performance Evaluation

The performance evaluation of the constituent models was performed using the complexity matrix. The matrix from which Acc and other metrics are obtained is given in Table 3 (Şentürk & Şentürk, 2016).

Table 3
Confusion matrix

		Actual value	
		Positive	**Negative**
Predicted Value	Positive	TP	FP
	Negative	FN	TN

The Acc of the model is expressed as the ratio of correctly classified samples to the total number of samples. High Acc indicates that the model has a high ability to predict correctly, whereas low Acc indicates that improvements in the model's performance are needed. The Acc value is given in Equation (1) (Şentürk & Şentürk, 2016).

$$\text{Acc} = \frac{T_P + T_N}{T_P + F_P + F_N + T_N} \tag{1}$$

Prec denotes the proportion of samples predicted by the model as positive to true positives. A high Prec indicates that the number of false positive predictions is low, and most of the samples classified as positive are positive. Prec is calculated as in Equation (2)(Sinan Basarslan & Kayaalp, 2021).

$$\text{Prec} = \frac{T_P}{T_P + F_P} \tag{2}$$

The Rec value is a performance measure used in classification problems and refers to the rate at which all true positive examples are predicted as positive. Prec is an important measure to reduce the number of false negatives and minimize the number of missing true positive examples. Rec is calculated using Equation (3) (Sinan Basarslan & Kayaalp, 2021)

$$\text{Rec} = \frac{T_P}{T_P + F_N} \tag{3}$$

Mathematically, the harmonic mean of Rec and Prec defines F1. Equation (4) provides the formula for F1 (Şentürk & Şentürk, 2016).

$$\text{F1} = \frac{2^* \text{ Recall } * \text{ Precision}}{\text{Recall} + \text{Precision}} \tag{4}$$

In this study, the Acc performance criterion was used in the classification models created for the customer churn. A comparison of the literature on similar studies on the same dataset using the Acc criterion is provided in the conclusion and discussion section.

Experimental Results

Experimental results from a 5-fold cross-validation decomposition on the Telco dataset are presented in this section. Within the scope of this research, the results of experiments on models created using machine learning, deep learning, and ensemble learning in the Google Colaboratory environment are given in Table 4.

Table 4

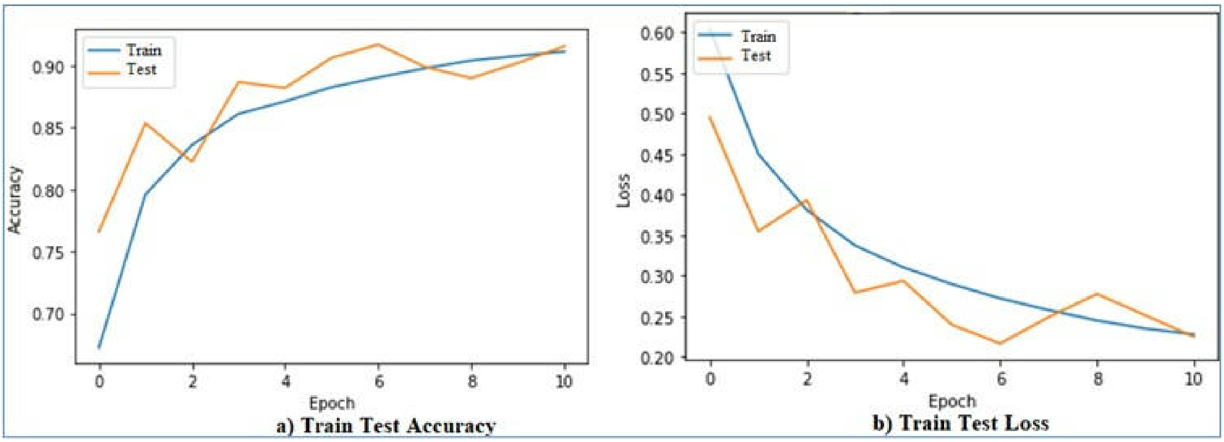
Experimental results of models on Telco Telecom dataset

Category	Classifiers	Acc	Prec	Rec	F1
Machine Learning	LR	80.48	88.48	78.55	83.20
	KNN	77.71	85.86	76.36	80.83
	SVM	80.62	88.70	78.60	83.34
	DT	80.70	88.89	78.43	83.33
	RF	80.84	89.15	78.48	83.48
	ANN	84.95	90.33	74.24	81.50
Ensemble Learning	XGBoost	82.97	86.77	72.97	79.27
	Majority Voting	80.98	81.52	72.41	76.70
Deep Learning	LSTM	89.03	94.32	81.91	87.68
	Proposed Model	**91.90**	**91.49**	**92.31**	**91.90**

According to Table 4, the Proposed Model achieved the highest accuracy (91.90%) among all methods applied on the Telco Telecom dataset in training-test separation with 5-fold cross-validation. In the deep learning category, the LSTM model achieved an accuracy of 89.03%, which is lower than the Proposed Model but higher than all other methods. Among the machine learning methods, RF obtained the highest accuracy (80.84%), followed closely by DT (80.70%), SVM (80.62%), and LR (80.48%). However, the proposed KNN method lagged other methods with an accuracy of 77.71%. Ensemble learning methods demonstrated better performance than machine learning methods, with XGBoost achieving an accuracy of 82.97% and the Majority Voting classifier achieving 80.98%. The training-validation accuracy and loss graph for the proposed model are shown in Figure 3.

Figure 3

Acc and loss graphics during training and testing.

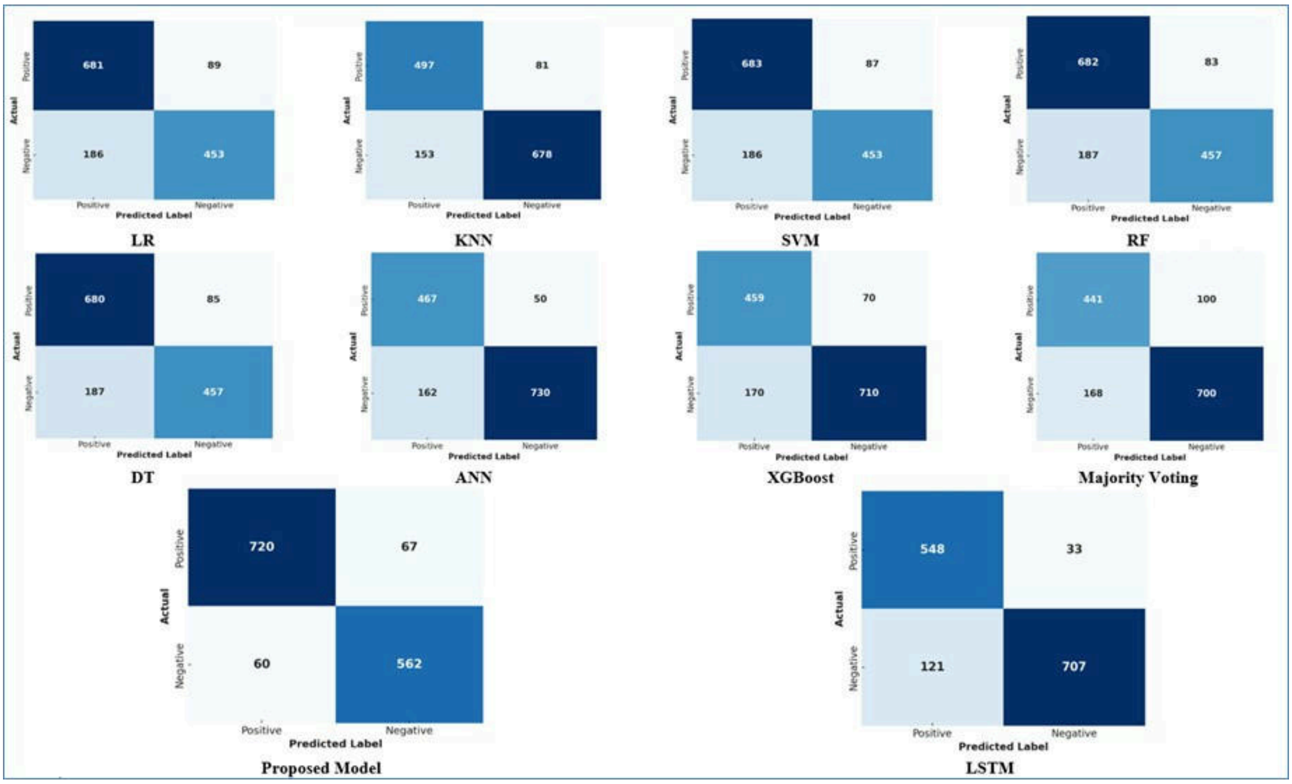


In Figure 3, we present the training and test accuracy and loss graphs over 10 epochs to illustrate the learning dynamics of the proposed model. In the training and test accuracy graph (Figure 3 a), both curves show a steady increase, with the test accuracy exceeding the training accuracy in the early epochs (e.g., 2-4). This demonstrates that the model generalizes well from the outset, possibly due to effective regularization techniques By epoch 10, the training and test accuracy converged, demonstrating minimal overfitting and strong generalization to the test data. In the training and test loss plots (Figure 3 b), both losses decreased significantly during the first epochs, indicating that the model effectively optimized the error function. The test loss stabilized after epoch 6, indicating the consistency of the model in handling unseen data. More importantly, there was no discernible deviation between the training and test loss, which further supported the robustness and stability of the model across the datasets.

Taken together, the two plots in Figure 3 (3a and 3b) confirm that the proposed model learns effectively, avoids overfitting, and generalizes well, making it reliable for predictions on unseen data. The confusion matrices of the models that were created in the study are shown in Figure 4.

Figure 4

Results of the confusion matrix of the models



Limitations of The Study

The dataset itself is a limitation because the performance of the model may vary depending on the characteristics of the dataset. Another limitation is that the domain of the dataset is defined as communications because customer churn is also important for other sectors, such as banking or subscription-based domains. At the same time, the selected deep learning and machine learning methods can also be considered limitations.

Conclusion and Discussion

In today's market conditions, the value of existing customers for companies is quite high. Therefore, companies attach great importance to customer disruption analysis. There are many studies in this field in the literature. In this study, churn analysis experiments were conducted on the Telco Telecommunication open dataset with models based on some of the abovementioned machine learning, deep learning, and ensemble learning methods. The Acc, Prec, Rec, and F1 performance metrics were used to evaluate the experimental results.

The results of the experiments performed on the Telco dataset using the above methods are presented in Table 4. When Table 4 is analyzed, it can be seen that the method that gives the best Acc result is the three LSTM models, which consist of a hybrid of deep learning methods with 91.90%. The comparison of the best results obtained from the models developed in this study with those from other studies in the literature on the same dataset is presented in Table 5.

Table 5

Comparison with literature

Model	Acc (%)
(Azeem et al., 2017)	57.00
(Coussement et al., 2017)	63.00
(Varan & Behara, 2014)	68.00
(Zhang et al., 2010)	75.53
(Qureshi et al., 2013)	78.00
(Arifin & Samopa, 2018)	80.00
(De Caigny et al., 2018)	81.60
(Lalwani et al., 2022)	81.71
(Amin et al., 2019)	82.11
(Zatonatska et al., 2023)	83.00
(Gaur & Dubey, 2018)	84.59
(Wanchai, 2017)	86.13
(Ullah et al., 2019)	88.63
Proposed Model	**91.90**

It can be seen that this study competes with the literature by examining the values given in Table 5. Future studies are planned to evaluate the performance of other deep learning methods, which are cutting-edge technology, in customer churn analysis studies.

Although churn analysis is focused on the telecom sector, which is a subscription-based industry, it is also relevant to industries such as banking, insurance, and healthcare. In addition, it is expected that different deep learning methods such as LSTM can be applied as single or ensemble learning. The proposed model stands out for its high accuracy and generalizability. Tested with 5-fold cross-validation, the model demonstrated consistent performance for different data partitions and reduced error rates for critical business problems, such as customer churn prediction. This demonstrates both the analytical accuracy and operational flexibility of the proposed model. As a result, companies will be able to predict churns with greater accuracy and reduce losses as much as possible. These studies provide companies with an advantage in all types of actions, including marketing strategies.

Investigating the effects of Bidirectional Long Short-Term Memory Networks (BiLSTM) and transfer learning methods on enhancing model performance are planned as future work.

Peer Review: Externally peer-reviewed.
Author Contributions: Conception/Design of Study- M.S.B., A.Ü., F.K.; Data Analysis/Interpretation- A.Ü., M.S.B., F.K.; Drafting Manuscript- A.Ü., M.S.B., F.K.; Critical Revision of Manuscript- A.Ü., F.K., M.S.B.; Final Approval and Accountability- M.S.B., A.Ü., F.K.
Conflict of Interest: The authors have no conflict of interest to declare.
Grant Support: The authors declared that this study has received no financial support.

Author Details

Muhammet Sinan Başarslan
[1] İstanbul Medeniyet University, Faculty of Engineering and Natural Sciences, Computer Engineering, İstanbul, Türkiye
0000-0002-7996-9169 muhammet.basarslan@medeniyet.edu.tr

Aslıhan Ünal
[2] İstanbul Gelişim University, Faculty of Economics, Administrative and Social Sciences, Department of Business Administration, İstanbul, Türkiye
0000-0001-5896-8880

Fatih Kayaalp
[3] Düzce University, Faculty of Engineering, Computer Engineering, Düzce, Türkiye
0000-0002-8752-3335

References

Ahmed, R., Bibi, M., & Syed, S. (2023a). Improving Heart Disease Prediction Accuracy Using a Hybrid Machine Learning Approach: A Comparative study of SVM and KNN Algorithms. *International Journal of Computations, Information and Manufacturing (IJCIM)*, *3*(1), 49–54. https://doi.org/10.54489/ijcim.v3i1.223

Ahmed, R., Bibi, M., & Syed, S. (2023b). Improving Heart Disease Prediction Accuracy Using a Hybrid Machine Learning Approach: A Comparative study of SVM and KNN Algorithms. *International Journal of Computations, Information and Manufacturing (IJCIM)*, *3*(1), 49–54. https://doi.org/10.54489/ijcim.v3i1.223

Ahn, J., Hwang, J., Kim, D., Choi, H., & Kang, S. (2020). A Survey on Churn Analysis in Various Business Domains. *IEEE Access*, *8*, 220816–220839. https://doi.org/10.1109/ACCESS.2020.3042657

Amin, A., Al-Obeidat, F., Shah, B., Adnan, A., Loo, J., & Anwar, S. (2019). Customer churn prediction in telecommunication industry using data certainty. *Journal of Business Research*, *94*, 290–301. https://doi.org/10.1016/j.jbusres.2018.03.003

Arifin, S., & Samopa, F. (2018). Analysis of Churn Rate Significantly Factors in Telecommunication Industry Using Support Vector Machines Method. *Journal of Physics: Conference Series*, *1108*, 012018. https://doi.org/10.1088/1742-6596/1108/1/012018

Azeem, M., Usman, M., & Fong, A. C. M. (2017). A churn prediction model for prepaid customers in telecom using fuzzy classifiers. *Telecommunication Systems*, *66*(4), 603–614. https://doi.org/10.1007/s11235-017-0310-7

Baltsou, G., Tsichlas, K., & Vakali, A. (2022). Local community detection with hints. *Applied Intelligence*, *52*(9), 9599–9620. https://doi.org/10.1007/s10489-021-02946-7

Başarslan, M. S., & Kayaalp, F. (2023). Sentiment analysis with ensemble and machine learning methods in multi-domain datasets. *Turkish Journal of Engineering*, *7*(2), 141–148. https://doi.org/10.31127/tuje.1079698

Bayrak, A. T., Aktas, A. A., Tunali, O., Susuz, O., & Abbak, N. (2021). *Personalized Customer Churn Analysis with Long Short-Term Memory*. 2021 IEEE International Conference on Big Data and Smart Computing (BigComp), 79–82. https://doi.org/10.1109/BigComp51126.2021.00024

Bhattacharyya, J., & Dash, M. K. (2022). What Do We Know About Customer Churn Behaviour in the Telecommunication Industry? A Bibliometric Analysis of Research Trends, 1985–2019. *FIIB Business Review*, *11*(3), 280–302. https://doi.org/10.1177/23197145211062687

Breiman, L. (2001). Random forests. *Machine Learning*, *45*(1), 5–32. https://doi.org/10.1023/A:1010933404324

Cao, J., Lin, Z., Huang, G.-B., & Liu, N. (2012). Voting based extreme learning machine. *Information Sciences*, *185*(1), 66–77. https://doi.org/10.1016/j.ins.2011.09.015

Cao, S., Liu, W., Chen, Y., & Zhu, X. (2019). *Deep Learning Based Customer Churn Analysis*. 2019 11th International Conference on Wireless Communications and Signal Processing (WCSP), 1–6. https://doi.org/10.1109/WCSP.2019.8927877

Chen, T., & Guestrin, C. (2016). *XGBoost*. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 785–794. https://doi.org/10.1145/2939672.2939785

Coussement, K., Lessmann, S., & Verstraeten, G. (2017). A comparative analysis of data preparation algorithms for customer churn prediction: A case study in the telecommunication industry. *Decision Support Systems*, *95*, 27–36. https://doi.org/10.1016/j.dss.2016.11.007

Çelebi, S. B., & Fidan, Ş. (2024). RNN-Based Time Series Analysis for Wind Turbine Energy Forecasting. *International Journal of Engineering and Innovative Research*, *6*(1), 15-28. https://doi.org/10.47933/ijeir.1387314

Çelebi, S. B., & Karaman, Ö. A. (2023). Multilayer LSTM Model for Wind Power Estimation in the Scada System. *European Journal of Technique (EJT)*, *13*(2), 116-122. https://doi.org/10.36222/ejt.1382837

De Caigny, A., Coussement, K., & De Bock, K. W. (2018). A new hybrid classification algorithm for customer churn prediction based on logistic regression and decision trees. *European Journal of Operational Research*, *269*(2), 760–772. https://doi.org/10.1016/j.ejor.2018.02.009

Ecevit, A., Öztürk, İ., Dağ, M., & Özcan, T. (2023). Short-Term Sales Forecasting Using LSTM and Prophet Based Models in E-Commerce. *Acta Infologica*, *7*(1), 59-70, 2023. https://doi.org/10.26650/acin.1259067

Edwine, N., Wang, W., Song, W., & Ssebuggwawo, D. (2022). *Detecting the Risk of Customer Churn in Telecom Sector: A Comparative Study*. Mathematical Problems in Engineering, 2022, 1–16. https://doi.org/10.1155/2022/8534739

Fu, W. (2022). Research on the construction of early warning model of customer churn on an e-commerce platform. *Applied Mathematics and Nonlinear Sciences*, 0(0). https://doi.org/10.2478/amns.2022.1.00016

Gary Schneider. (2016). *Electronic Commerce (12th Edition)*. Cengage Learning.

Gaur, A., & Dubey, R. (2018). *Predicting Customer Churn Prediction In Telecom Sector Using Various Machine Learning Techniques*. 2018 International Conference on Advanced Computation and Telecommunication (ICACAT), 1–5. https://doi.org/10.1109/ICACAT.2018.8933783

Guo, F., & Qin, H. (2017). Data Mining Techniques for Customer Relationship Management. *Journal of Physics: Conference Series*, *910*, 012021. https://doi.org/10.1088/1742-6596/910/1/012021

Hadden, J., Tiwari, A., Roy, R., & Ruta, D. (2007). Computer assisted customer churn management: State-of-the-art and future trends. *Computers and Operations Research*, *34*(10), 2902–2917. https://doi.org/10.1016/j.cor.2005.11.007

Hanif, A., & Azhar, N. (2017, December). *Resolving class imbalance and feature selection in customer churn dataset*. In 2017 International Conference on Frontiers of Information Technology (FIT) (pp. 82-86). IEEE.

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, *9*(8), 1735–1780.

Indulkar, Y., & Patil, A. (2021). *Comparative Study of Machine Learning Algorithms for Twitter Sentiment Analysis*. 2021 International Conference on Emerging Smart Computing and Informatics (ESCI), 295–299. https://doi.org/10.1109/ESCI50559.2021.9396925

Jeyaprakaash, P., & Sashirekha, K. (2022). Accuracy Measure of Customer Churn Prediction in Telecom Industry using Adaboost–Decision Tree Algorithm. *Journal of Pharmaceutical Negative Results*, *13*(SO4). https://doi.org/10.47750/pnr.2022.13.S04.179

Karvana, K. G. M., Yazid, S., Syalim, A., & Mursanto, P. (2019). *Customer Churn Analysis and Prediction Using Data Mining Models in Banking Industry*. 2019 International Workshop on Big Data and Information Security (IWBIS), 33–38. https://doi.org/10.1109/IWBIS.2019.8935884

Kozak, J., Kania, K., Juszczuk, P., & Mitręga, M. (2021). Swarm intelligence goal-oriented approach to data-driven innovation for customer churn management. *International Journal of Information Management*, *60*, 102357. https://doi.org/10.1016/j.ijinfomgt.2021.102357

Kurt, K. K., Atay, H. T., Cicek, M. A., Karaca, S. B., & Turkeli, S. (2019). *The Importance of Multidisciplinary in Data Science: Application of the Method in Health Sector to Telecommunication Sector*. 2019 International Artificial Intelligence and Data Processing Symposium (IDAP), 1–4. https://doi.org/10.1109/IDAP.2019.8875917

Lalwani, P., Mishra, M. K., Chadha, J. S., & Sethi, P. (2022). Customer churn prediction system: a machine learning approach. *Computing*, *104*(2), 271–294. https://doi.org/10.1007/s00607-021-00908-y

LaValley, M. P. (2008). Logistic Regression. *Circulation*, *117*(18), 2395–2399, doi:10.1161/CIRCULATIONAHA.106.682658

Malik, S., Bansal, P., Sharma, P., Jain, R., & Vashisht, A. (2022). *Image Retrieval Using Multilayer Bi-LSTM* (pp. 745–755). https://doi.org/10.1007/978-981-16-2597-8_64

Mishra, A., & Reddy, U. S. (2017). *A Novel Approach for Churn Prediction Using Deep Learning*. 2017 IEEE International Conference on Computational Intelligence and Computing Research (ICCIC), 1–4. https://doi.org/10.1109/ICCIC.2017.8524551

Mohammadifar, A., Gholami, H., & Golzari, S. (2022). Stacking- and voting-based ensemble deep learning models (SEDL and VEDL) and active learning (AL) for mapping land subsidence. *Environmental Science and Pollution Research*, *30*(10), 26580–26595. https://doi.org/10.1007/s11356-022-24065-7

Mokhairi, M., Nordin, A.R. & Khalid, A. (2016). Predictive modeling for telco customer churn using rough set theory. *ARPN Journal of Engineering and Applied Sciences*, *11*(5), 3203-3207.

Osmanoglu, O. O. (2019). *Comparing to Techniques Used in Customer Churn Analysis*. In Journal of Multidisciplinary Developments (Vol. 4, Issue 1).

PwC consumer Experience Series. (n.d.). https://www.pwc.de/de/consulting/pwc-consumer-intelligence-series-customer-experience.pdf. https://www.pwc.de/de/consulting/pwc-consumer-intelligence-series-customer-experience.pdf

Qureshi, S. A., A. S. Rehman, A. M., & Qamar, A. (2013). *Telecommunication subscribers' churn prediction model using machine learning*. Eighth International Conference on Digital Information Management (ICDIM 2013), 131–136. https://doi.org/10.1109/ICDIM.2013.6693977

Routh, P., Roy, A., & Meyer, J. (2021). Estimating customer churn under competing risks. *Journal of the Operational Research Society*, *72*(5), 1138–1155. https://doi.org/10.1080/01605682.2020.1776166

Seymen, O. F., Ölmez, E., Doğan, O., Er, O., & Hızıroğlu, K. (2023). Customer Churn Prediction Using Ordinary Artificial Neural Network and Convolutional Neural Network Algorithms: A Comparative Performance Assessment. *Gazi University Journal of Science*, *36*(2), 720–733. https://doi.org/10.35378/gujs.992738

Shi Long Zhuang. (n.d.). *Telecom Churn Dataset*. 2022. https://www.kaggle.com/datasets/shilongzhuang/telecom-customer-churn-by-maven-analytics

Sinan Basarslan, M., & Kayaalp, F. (2021). *Performance Evaluation of Classification Algorithms on Diagnosis of Breast Cancer and Skin Disease* (pp. 27–35). https://doi.org/10.1007/978-981-15-6321-8_2

Su, C., Wei, L., & Xie, X. (2022). *Churn Prediction in Telecommunications Industry Based on Conditional Wasserstein GAN*. 2022 IEEE 29th International Conference on High Performance Computing, Data, and Analytics (HiPC), 186–191. https://doi.org/10.1109/HiPC56025.2022.00034

Şeker, S. E. (2016). Müşteri Kayıp Analizi (Customer Churn Analysis). *YBS Ansiklopedi*, 3(1).

Şentürk, A., & Şentürk, Z. K. (2016). Yapay Sinir Ağları İle Göğüs Kanseri Tahmini. *El-Cezeri Fen ve Mühendislik Dergisi*, *3*(2). https://doi.org/10.31202/ecjse.264199

Tan, K. L., Lee, C. P., Anbananthen, K. S. M., & Lim, K. M. (2022). *RoBERTa-LSTM: A Hybrid Model for Sentiment Analysis With Transformer and Recurrent Neural Network*. IEEE Access, 10, 21517–21525. https://doi.org/10.1109/ACCESS.2022.3152828

Ullah, I., Raza, B., Malik, A. K., Imran, M., Islam, S. U., & Kim, S. W. (2019). A Churn Prediction Model Using Random Forest: Analysis of Machine Learning Techniques for Churn Prediction and Factor Identification in Telecom Sector. *IEEE Access*, *7*, 60134–60149. https://doi.org/10.1109/ACCESS.2019.2914999

Ustebay, S., Turgut, Z., & Aydin, M. A. (2019). *Cyber Attack Detection by Using Neural Network Approaches: Shallow Neural Network, Deep Neural Network and AutoEncoder* (pp. 144–155). https://doi.org/10.1007/978-3-030-21952-9_11

Uzun, E., Agun, H. V., & Yerlikaya, T. (2012). *Web content extraction by using decision tree learning*. 2012 20th Signal Processing and Communications Applications Conference (SIU), 1–4. https://doi.org/10.1109/SIU.2012.6204476

Varan, S., & Behara, R. S. (2014). *Customer churn analysis in the wireless industry: A data mining approach*. https://www.researchgate.net/publication/228411974

Wanchai, P. (2017). *Customer churn analysis: A case study on the telecommunication industry of Thailand*. 2017 12th International Conference for Internet Technology and Secured Transactions (ICITST), 325–331. https://doi.org/10.23919/ICITST.2017.8356410

Zatonatska, T., Fareniuk, Y., & Shpyrko, V. (2023). Churn Rate Modeling for Telecommunication Operators Using Data Science Methods. *Marketing and Management of Innovations*, *14*(2), 163–173. https://doi.org/10.21272/mmi.2023.2-15

Zavrak, S., & Iskefiyeli, M. (2023). Flow-based intrusion detection on software-defined networks: a multivariate time series anomaly detection approach. *Neural Computing and Applications*, *35*(16), 12175–12193. https://doi.org/10.1007/s00521-023-08376-5

Zavrak, S., & Yilmaz, S. (2023). Email spam detection using hierarchical attention hybrid deep learning method. *Expert Systems with Applications*, *233*, 120977. https://doi.org/10.1016/j.eswa.2023.120977

Zhang, X., Liu, Z., Yang, X., Shi, W., & Wang, Q. (2010). *Predicting customer churn by integrating the effect of the customer contact network*. Proceedings of 2010 IEEE International Conference on Service Operations and Logistics, and Informatics, 392–397. https://doi.org/10.1109/SOLI.2010.5551545