Article

Deep Churn Prediction Method for Telecommunication Industry

Lewlisa Saha [1], Hrudaya Kumar Tripathy [1], Tarek Gaber [2,3,*], Hatem El-Gohary [4] and El-Sayed M. El-kenawy [5]

1   School of Computer Engineering, Kalinga Institute of Industrial Technology, Bhubaneswar 751024, India
2   Faculty of Computers and Informatics, Suez Canal University, Ismailia 41522, Egypt
3   School of Science, Engineering, and Environment, University of Salford, Salford M5 4WT, UK
4   College of Business and Economics, Qatar University, Doha 2713, Qatar
5   Department of Communications and Electronics, Delta Higher Institute of Engineering and Technology, Mansoura 35111, Egypt
*   Correspondence: t.m.a.gaber@salford.ac.uk

Abstract: Being able to predict the churn rate is the key to success for the telecommunication industry. It is also important for the telecommunication industry to obtain a high profit. Thus, the challenge is to predict the churn percentage of customers with higher accuracy without comprising the profit. In this study, various types of learning strategies are investigated to address this challenge and build a churn predication model. Ensemble learning techniques (Adaboost, random forest (RF), extreme randomized tree (ERT), xgboost (XGB), gradient boosting (GBM), and bagging and stacking), traditional classification techniques (logistic regression (LR), decision tree (DT), and k-nearest neighbor (kNN), and artificial neural network (ANN)), and the deep learning convolutional neural network (CNN) technique have been tested to select the best model for building a customer churn prediction model. The evaluation of the proposed models was conducted using two pubic datasets: Southeast Asian telecom industry, and American telecom market. On both of the datasets, CNN and ANN returned better results than the other techniques. The accuracy obtained on the first dataset using CNN was 99% and using ANN was 98%, and on the second dataset it was 98% and 99%, respectively.

1. Introduction

Churn rate is the number of customers leaving a company annually. This can pose a challenge for any business organization. The prediction of customers who may want to leave the company is then a crucial task for any business. With recent advances in data analytics, many forms of Customer Relationship Management (CRM) systems have been embedded as data analytical methods and these have became the focus of many studies and practices. Such analytical methods pay much attention to customer-centric approaches over product-centric approaches. As a consequence, customer–company interactions have changed in such a way that numerous new advertising openings have emerged. The most profitable marketing tactics that make the most of shareholder value include reducing churn and maintaining current clients [1].

Customers remain the most valuable entity in the telecommunications industry for a company to continue operating. A reduction in the number of customers is an unanticipated event for businesses. As a result, businesses must examine consumer profiles in order to undertake business segmentation and make more informed decisions [2]. In today's highly competitive sectors, such as the telecom sector, customer churn is among the most urgent concerns [3]. Due to the high price of enlisting new consumers, the telecom industry has turned its attention to maintaining existing customers. Compared to new consumers,

keeping existing customers leads to more sales and lower marketing costs [4]. Therefore, customer churn prediction has become an essential aspect of the telecommunication sector's strategic executive and planning procedure [5].

According to one report, the telecom industry's annual churn rate stretches from 20% to 40%, and the price of preserving current customers is 5–10 times cheaper than the price of acquiring new consumers [6]. Predicting customer turnover is 16 times less expensive than acquiring new customers. The profit increases from 25% to 85% when the churn rate is reduced by 5% [1]. This demonstrates the importance of predicting client attrition in the telecom industry. CRM is critical for telecom companies in retaining existing customers and reducing customer churn. Thus, the precision of the CRM analyzers' prediction systems is critical. No advertising can be run if analyzers are inaccurate in forecasting consumer attrition [7]. Data mining and machine learning technologies, which have seen recent advancements in data science, provide answers to consumer attrition.

In order for businesses to improve their customer relationships and obtain more devoted customers, personalized marketing strategies (which can be done through social media marketing) are very necessary. To accomplish this, it is necessary to give their customer support and sales people the ability to obtain information on the company's users and provide training so that they can properly interact with all of their clients. In the age of big data [4], such a task could be easily undertaken using AI techniques without putting much strain on the customer support and sales teams. It is essential to include AI in all business activities, including marketing/social marketing, CRM, and sales, among others, in order to successfully engage customers and earn their confidence.

Considering the important role that social media and other electronic marketing platforms play in the business world today, it is very vital to comprehend how to use, adapt, and implement such platforms effectively [8,9]. Deep learning and customer behaviour analysis can significantly affect social media marketing and other marketing activities of a company by allowing for more personalized and targeted marketing activities. By analyzing customer data, businesses can gain insights into what is likely to resonate with their audience. A business can use such information to create more effective social media and marketing campaigns. This will, in turn, lead to higher customer engagement and conversion rates. Additionally, deep learning algorithms can help companies to automate and optimize their advertising efforts, saving time and resources while improving overall performance.

As discussed in the next section, the existing works on churn prediction have been undertaken on different industry customers, e.g., employee churn [10]; however, the majority of them are on telecom data since this industry deals with the most severe losses from losing customers. Maximizing profit while attempting to predict the churn rate has been a point of concern for some researchers; however, for others, obtaining a better accuracy in predicting churn has been the focal point. However, the main issue that is not well-studied is that the target should be gaining higher prediction accuracy without compromising profit, which means less complicated techniques that can be used without a large amount of investment.

To address this issue, in this paper, various types of machine learning strategies are investigated in order to build a churn predication model. Traditional classifiers, ensemble learning, and deep learning have been investigated to build the model which was evaluated using two public datasets. The first dataset is from the Indian and Southeast Asian telecom industry [11] and the second dataset is from the American telecom market [12]. The benefactions of the proposed research model are as follows:

- Recommending a prediction model for churn with a high precision and accuracy.
- The proposed model is able to conquer the complication of lower quantity of churn in the datasets in comparison to the non-churn customers.
- The proposed model has been thoroughly evaluated using diverse performance metrics using two public datasets. These include accuracy, recall, precision, F1 score, and AUC- (Area Under the ROC Curve) ROC (Receiver Operating Characteristic). Also,

the model was compared with different related works on churn prediction rate, which showed that our model outperforms all of them.

- Further, the statistical analysis using ANOVA and Wilcoxon showed that the proposed models are statistically significant compared to other models.

In addition to the above section, the paper is structured as follows. Section 2 discusses the related work. The tools, dataset, and approach employed in the research are briefly described in Section 3. The research findings have been thoroughly described in Sections 4 and 5 wraps up the entire project.

2. Literature Survey

Consumer preferences and expectations have shifted as a result of evolving technology and the increasingly extensive accessibility of numerous services and products, resulting in a highly competitive environment across a number of customer service sectors, including the financial sector. The authors Shirazi and Mohammadi [2] discussed the effects of this situation on the Canadian banking industry. Their main goal is to combine structured archived data with unstructured factual material such as online web sites, the amount of website traffic, and phone call records to create a predicted churn model. They also studied how different customer habits affect churning decisions. Companies' success is largely determined by their ability to analyze existing data and extract useful information. A cloud-based ETL (Extract, Transform, and Load) architecture for data analysis and the combination of diverse sources was recommended by Zdravevski et al. [13]. In the churn prediction example, they showed that they could identify the specific cause of churn and detect over 98% of churners. As a result, the support and sales teams could implement focused retention initiatives. In their study, Vo et al. [14] provided a customer churn forecasting model based on unstructured data, such as verbal comments made during phone conversations. They conducted extensive testing on substantial call center data using calls from clients. Their results showed that, utilizing understandable machine learning through behavior and customer sections, their model can effectively estimate consumer churn prospects and generate useful insights.

An effective retention strategy must not only correctly identify potential leavers but also those who are most profitable to the business and, hence, are deserving of keeping. Therefore, the best churn prediction model can both accurately pinpoint churners and take into account the company's profit. This could be achieved by embedding the (EMPC) metric, which stands for "Expected Maximum Profit amount for customer Churn," into a machine learning-based churn prediction model [15]. In this model, a profit-based decision tree was suggested which was used to evaluate real-life datasets from several communications service providers. The results showed that the profit-based decision tree model allowed for a noteworthy profit boost above conventional accuracy-driven tree-based techniques. In another work on maximizing profit alongside churn detection, Stripling et al. [16] presented ProfLogit in which a machine learning classifier was integrated with a genetic algorithm to exploit the EMPC in the training stage.

The telecommunication industry experiences customer turnover at a very high rate because of the easier portability options which are available these days and the severe market competition. Hence being able to predict the churn rate of customers has become a vital part of the whole business process. Table 1 shows several works which have studied churn predictions done using telecommunication industry data.

Table 1. Literature Survey of Papers on Churn Prediction.

Author Name	Technique Used	Dataset Used	Accuracy (%)
Coussement et al. [5]	LR	A significant European mobile telecommunications provider's home database.	-
Mishra and Reddy [1]	Bagging, Boosting and RF	UCI Repository.	91.66
Caigny et al. [7]	Logit leaf model (Level 1: DT; Level 2: LR)	14 different churn datasets.	-
Arifin and Samopa [17]	SVM	Dataset from an Indonesian telecom company	-
Jain et al. [18]	LR and Logit boost	Dataset from an American telecom company named Orange	85.23
Amin et al. [19]	Naive Bayes, kNN, GBM, Single Rule Induction and Deep learner Neural network	2 publicly available dataset.	-
Ahmad et al. [4]	DT, RF, GBM, XGB	SyriaTel dataset	-
Amin et al. [20]	Naïve Bayes	2 open-source datasets	89.01
Amin et al. [21]	Rough Set Theory (RST) and Rule-based Decision-making.	Open-source dataset	98.1
Alboukaey et al. [22]	Long Short-Term Memory (LSTM) and CNN	Real customer data by MTN Operator	-
Karuppaiah and Palanisamy [23]	Heterogeneous ensemble stacking (Initial level: GBM and Naïve Bayes Secondary level: SVM)	UCI repository	89.0
Caigny et al. [24]	CNN	Data from a European financial services provider	-
Mitroví'c et al. [25]	LR and RF	2 real life datasets	-
Bock and Caigny [26]	14 real life datasets	Sparse-group lasso (SGL) regularized regression	-
Xu et al. [27]	Stacking Ensemble using Extreme gradient boosting, Logistic regression, Decision tree, and Naïve Bayes	Publicly available dataset	98.09
Óskarsdóttir et al. [28]	Similarity forests	3 distinct CDR datasets from European telcos	-

3. Proposed Methodology

3.1. Techniques Used

3.1.1. Ensemble Learning

An assortment of various learning tools or classifiers known as an ensemble work together to provide outcomes that are more precise and reliable by integrating all available techniques. In ensemble approaches, an inducer is a learner whose definite classification of the categorized training set is randomly well correlated [27,29]. A poor learner is, to some extent, related to the genuine classification. As a result of the ensemble learning framework, new approaches such as bagging, boosting, and random forests have arisen. One of the aggregation techniques used in ensemble learning to lower variation in prediction models is bagging. Two further techniques are boosting, which helps turn many poor learners into one aggregated strong model, and stacking, which seeks to lessen prediction bias. Several ensemble learning methods have been implemented in this study to forecast the

customer turnover rate in the telecom sector, including RF, ERT, GBM, XGB, adaboost, and stacking. The different types of ensemble models used in this research work have been discussed below.

RF is a sort of ensemble study in which classification and regression are performed using numerous decision-making trees. In the RF classifier, there is some randomness in the selection of subsets and features for the nodes of each tree. One of the factors used to partition the data into random forests is the Gini index. The Gini index is a potent indicator of the randomness, impurity, or entropy in a dataset's values. It seeks to reduce the contaminants in a decision tree model from the root nodes to the leaf nodes. Variables that help not just the creation of an accurate model but also the prediction are vital for the random forest technique [29,30]. Figure 1 shows the basis representation RF.
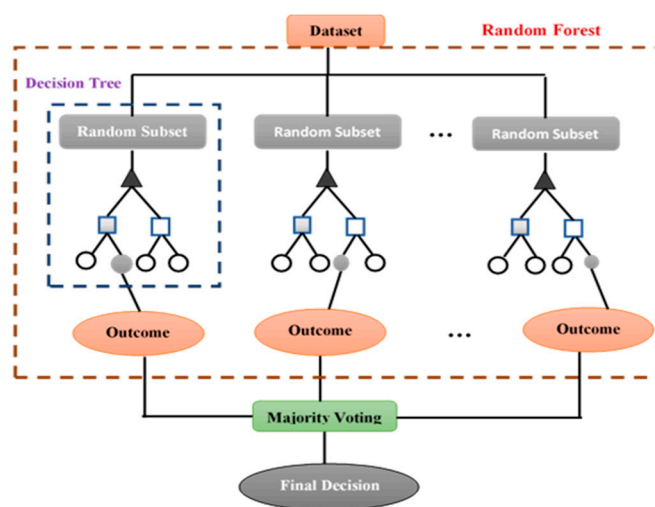

Figure 1. Tree representation of random forest.

In essence, the Extreme Randomized Tree (ERT) technique involves extensively randomizing the selection of trait and cut-points while dividing a tree node. In the worst case, it generates entirely random trees, whose designs are unaffected by the output of the learning model. The Extra-Trees or ERT technique produces an ensemble of unpruned results or regression trees in accordance with the conventional top-down method. It splits nodes randomly and generates trees using the complete learning sample, which are its two main distinctions from other tree-based ensemble techniques [31].

GBM is frequently applied to problems which involve classification and regression. It can be applied to a variety of real-world situations with great benefit. This is a method of numerical optimization that seeks recognition of an additive model with the lowest possible loss [32,33].

Regression tree XGBoost adheres to the Decision Tree concept, to which it is similar. It supports both classification and regression. This gradient booster (GBM) version is widely applied in machine learning and in its applications. It is scalable and efficient. The training is based on an "additive strategy": k additive functions are employed to expect each tree ensemble model when a molecule i and a descriptor x_i vector are given [34,35].

Adaptive boosting is abbreviated as AdaBoost. In order to satisfy the classification requirements of datasets, it is a kind of dichotomous classification algorithm that develops and fuses a number of base classifiers. AdaBoost increases the weight of a test sample that the previous base classifier incorrectly classified, while decreasing the weight of the samples that were appropriately classified by the subsequent weak learner [36–38].

Bagging is a technique for enhancing the precision of predictions made by other learning algorithms. It is a method for merging multiple compound models and then majority voting (in classification) or averaging (in regression) their outputs to generate a more powerful prediction model [30,39].

Stacking, or stacked generalization, is an ensemble machine learning algorithm. Using a meta-learning strategy, it learns how to aggregate estimates from two or more fundamental machine learning algorithms. On either a classification or regression job, stacking has the advantage of combining the abilities of numerous high-performing models to produce predictions that are superior to any specific model within the ensemble [27].

3.1.2. Artificial Neural Network (ANN)

ANN is a mathematical description of the human nervous system that is heavily dependent on its functional and structural elements. In order to provide the needed scalar output, the final layer of the network employs a function, which is referred to as an activation or a transfer function [40]. An artificial neuron network has been mathematically depicted below:

$$O(t) = f\left(\sum_{i=1}^{n} v_i(t) \cdot I_i(t) + c\right) \tag{1}$$

where $O(t)$ = output at a given time, f = transfer function, c = bias, $I_i(t)$ = inputs, and $v_i(t)$ = weights [29,41].

The mathematical representation of a feed-forward NN consisting of one input layer, one hidden layer, and one output layer, as shown in Figure 2, is given through the equations below:

$$m_1 = f_1(v_1 I_1 + c_1) \tag{2}$$

$$m_2 = f_1(v_1 I_1 + c_1) \tag{3}$$

$$m_3 = f_2(v_2 I_2 + c_2) \tag{4}$$

$$m_4 = f_2(v_2 I_2 + c_2) \tag{5}$$

$$m_5 = f_3(v_3 I_3 + c_3) \tag{6}$$

$$m_6 = f_3(v_3 I_3 + c_3) \tag{7}$$

$$q_1 = f_4(p_1 m_1 + p_3 m_3 + p_5 m_5 + c_4) \tag{8}$$

$$q_2 = f_5(p_2 m_2 + p_4 m_4 + p_6 m_6 + c_5) \tag{9}$$

$$O = f_6(s_1 q_1 + s_2 q_2 + c_6) \tag{10}$$

$$O = f_6\left[\begin{array}{l} s_1\left(f_4\left[\begin{array}{l} p_1 f_1[v_1 I_1 + c_1] + p_3 f_2[v_2 I_2 + c_2] \\ + p_5 f_3[v_3 I_3 + c_3] + c_4 \end{array}\right]\right) \\ + s_2\left(f_5\left[\begin{array}{l} p_2 f_1[v_1 I_1 + c_1] + p_4 f_2[v_2 I_2 + c_2] \\ + p_6 f_3[v_3 I_3 + c_3] + c_5 \end{array}\right]\right) \\ + c_6 \end{array}\right] \tag{11}$$

where m_i, q_i = output of the preceding layer; p_i, s_i = weight of the current layer.
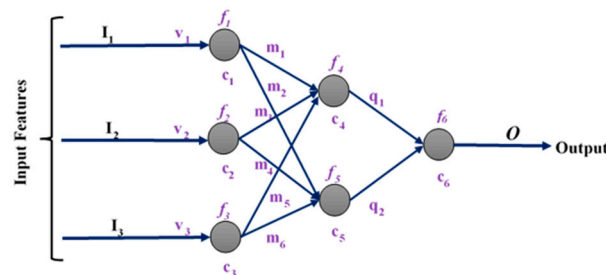

Figure 2. Artificial Neural Network.

3.1.3. Decision Tree (DT)

One of the several extensively used summative evaluations for categorizing data or identifying the hidden pattern in a batch of data is the decision tree. The first node in a decision tree is called the root node, while the second and third nodes are called the

internal and leaf nodes. The leaf nodes that make up the decision tree's final layer each have a specified class goal value. Separating the nodes on each level in accordance with the splitting criteria forms the basis for a decision tree. Up until a pausing criterion is met, this splitting and expanding phase continues [29]. The various standards can be embodied as follows:

$$InformationGain(b_i, R) = Entropy(z, R) - \sum\nolimits_{u_{i,j} \in dom(b_i)} \frac{\left| \sigma_{b_i = u_{i,j}} R \right|}{|R|} \cdot Entropy\left(z, \sigma_{b_i = u_{i,j}} R\right) \tag{12}$$

where,

$$EEntropy(z, R) = \sum\nolimits_{d_j \in dom(z)} - \frac{\left| \sigma_{z = d_j} R \right|}{|R|} \cdot \log_2 \frac{\left| \sigma_{z = d_j} R \right|}{|R|} \tag{13}$$

$$ini(z, R) = 1 - \sum\nolimits_{d_j \in dom(z)} \left(\frac{\left| \sigma_{b = u_{i,j}} R \right|}{|R|} \right)^2 \tag{14}$$

The assessment criteria are defined as,

$$GiniGain(b_i, R) = Gini(z, R) - \sum\nolimits_{u_{i,j} \in dom(b_i)} \frac{\left| \sigma_{b_i = u_{i,j}} R \right|}{|R|} \cdot Gini\left(z, \sigma_{b_i = u_{i,j}} R\right) \tag{15}$$

$$Gain\ Ratio(b_i, R) = \frac{InformationGain(b_i, R)}{Entropy(b_i, R)} \tag{16}$$

where R = a training set; b_i = a discrete attribute; z = target attribute; $u_{i,j}$ = values.

3.1.4. k Nearest Neighbor

In order to classify a new standard or invasive process, the kNN classifier compares the new structure to the training procedure instances. It then predicts the new process class using the closest k number of classes. The procedure takes for granted that, in the vector space, processes from the same class are clustered together. The value of k determines how many neighbors are required to describe the data class. The closest neighbors are chosen by a majority vote. The distance can be calculated using the Euclidean distance metric [30,42].

3.1.5. Logistic Regression (LR)

A statistically probabilistic method of categorization is called logistic regression. A subcategory variable that is impacted by one or more of the predictive elements can also be forecasted using this method (such as client characteristics). This method was utilized in our case after the original dataset had undergone considerable data preparation [43].

3.1.6. Convolutional Neural Network (CNN)

CNN is a method built on learning depiction, where the outline researches and identifies the properties necessary for uncovering from the several layers processing input information. It has historically been used in image processing applications; however, more recently, it has been used to forecast customer churn rate in the telecom industry using one-dimensional frameworks. CNN is a neural network (NN) with a multilayer architecture that consists of multiple fully connected and convolutional layers. The hierarchy of the convolution layers serves as the network's basic component of building. Further, 1D CNNs are naturally suited to processing data from consumer profiles [44].

The core architecture of the CNN method has been modified in this study to support the analysis of 1D churn data. The 1D design is also quicker than the 2D structure since it is more straightforward and has fewer parameters. The representation of a 1D convolutional method is:

$$x_j^L = f_a\left(\sum\nolimits_{i=1}^{C} x_i^{L-1} \cdot ke_{ij}^L + bi_j^L\right) \tag{17}$$

where L = convolutional layer, f_a = activation or transfer function, b_i = bias, k_e = quantity of convolutional layers, and C = quantity of input channels [45,46].

3.2. Dataset Used

In this work, two publicly available churn prediction datasets of the telecommunication industry have been used for comparison of the performances of the machine learning models. In the first dataset, the initial number of instances was 99,999 and the number of features was set at 226. After the filtering of high value customers was been performed, the number of instances was brought down to 30,001 and after the completion of the whole cleaning procedure, the numbers of features was brought down to 57. The final cleaned dataset was composed of 91.9% non-churn customers and 8.1% churn customers. For the second dataset, filtering was not required as a part of the preprocessing; hence, only a basic cleaning procedure was performed on the dataset. The number of instances and features in the final cleaned dataset were 3333 and 20, respectively. This dataset was composed of 85.5% and 14.5% non-churn and churn customers. Figure 3 shows the churn distribution of both of the datasets.
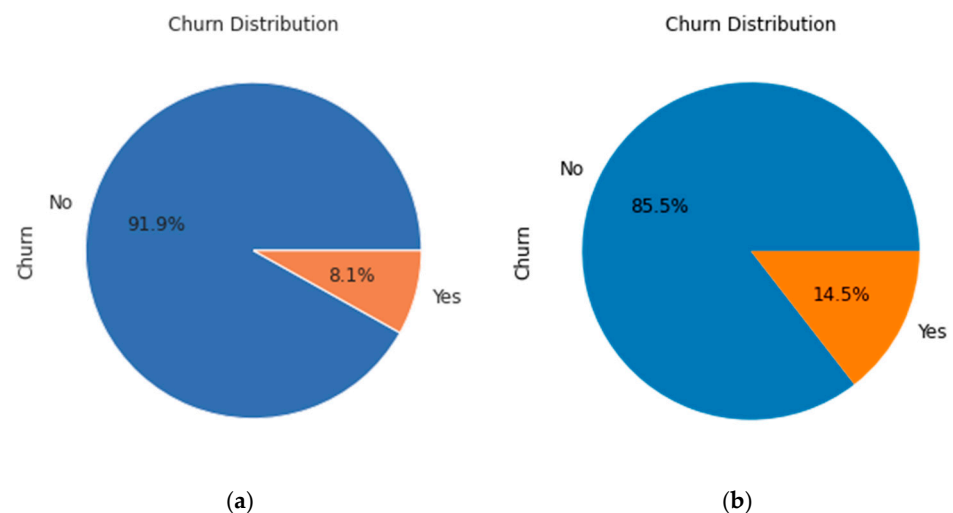

| (**a**) | (**b**) |

Figure 3. (**a**) Churn distribution of dataset 1. (**b**) Churn distribution of dataset 2.

3.3. Research Model

This work recommends a research design which forecasts customers' churn rate, i.e., whether a subscriber is going to leave or not by analyzing their behavioral pattern. In the first dataset, the churn is predicted only for the high value customers, which is determined by the amount of recharge over a certain period of time. A customer does not decide to churn instantly, it is a decision made over a certain period of time. This period is divided into three phases: the good phase, where the customer is happy and behaves in the usual way; the action phase, where the customers experience starts to sore; and the churn phase, where the customer has churned. The dataset used in this work spans over a window of four months: June, July, August, and September. Here, June and July are the good phase and, by analyzing the recharge amount spent in these two months, the high value customers are determined. Here, also, the amount used to indicate high value was being above 70th percentile of the average recharge amount. As shown in Figure 4, after filtering out the high value customers, the churn tag was assigned to the instances of churn, and then the new feature extracting was performed. In the second dataset, these preprocessing steps were not been used as necessary steps and it was instead cleaned by removing the unnecessary and redundant data.
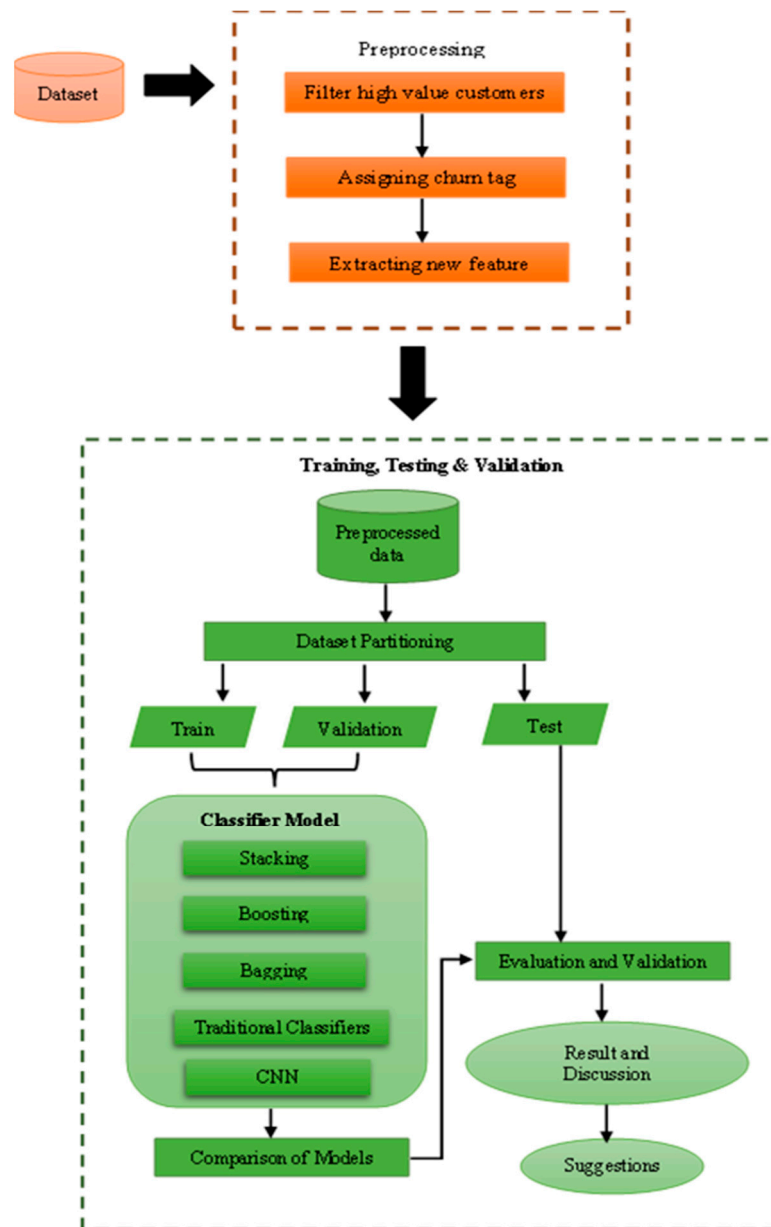
Figure 4. The proposed research model.

Several machine learning approaches were applied to the dataset after cleaning, with the purpose of making predictions. One of the categorization strategies for the predictive system was a stacking ensemble model. The stacking model had two levels, as shown in Figure 5. DT, RF, LR, kNN, and SVM made up level 0, whereas LR made up level 1. The final output was predicted using a soft voting technique. RF, ERT, AdaBoost, GBM, XGB, and Bagging are some of the other ensemble models which were used. Other than ensemble learning, the conventional classifiers utilized here are DT, kNN, LR, and ANN, and the deep learning technique used is CNN. The split for the training and testing was 80% and 20%, respectively, and a 10-fold cross validation was used to validate the data. The simulations have been performed on a system with processor 11th Gen Intel(R) Core (TM) i7-1195G7 @ 2.90GHz 2.92 GHz, Installed RAM 16.0 GB (15.8 GB usable), System type 64-bit operating system, x64-based processor, Windows 11 Home Single Language Version 22H2. The complete specifications of the various classifiers utilized in the study, which have been determined through trial and error, are presented in Tables 2 and 3.
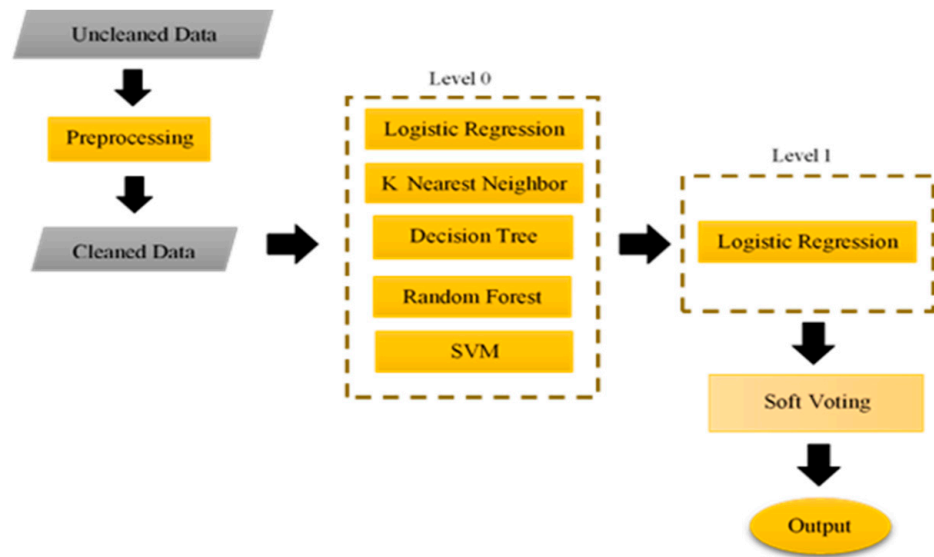
Figure 5. The proposed stacking model.

Table 2. Model Specification of Dataset 1.

Technique Used	Specifications
Random Forest	Estimators = 8000, Max Depth = 30, Min Samples Split = 10, Criterion = Entropy
ERT	Criterion = Gini, Min Samples Split = 10, Estimators = 5000
AdaBoost	Estimators = 1000, Learning-Rate = 0.1, Random State = 10
XGM	Random State = 10, Learning-Rate = 0.1, Estimators = 5000, Max-Depth = 3
GBM	Learning Rate = 0.01, Random State = 1, Estimators = 6000
Bagging	Estimators = 8000, Random State = 10
Stacking	Cv = 10, Estimators = [LR, K NN, DT, RF, SVC], Final Estimator = LR
DT	Max Depth = 2, Criterion = Gini, Random State = 99
kNN	Neighbors = 38
LR	Solver = Liblinear, Max Iter = 10,000
ANN	Dense_Layer = 32, Trans_Function = ReLU Dense_Layer = 16, Trans_Function = ReLU Dense_Layer = 8, Trans_Function = ReLU Dense_Layer = 1, Trans_Function = Sigmoid Optimizers = Adam, Learning Rate = 0.001 Loss = Binary Crossentropy
CNN	Conv1d = 128, Activation = ReLU Conv1d = 128, Activation = ReLU Dense_Layer = 64, Trans_Function = Tanh Dense_Layer = 32, Trans_Function = Tanh Dense_Layer = 16, Trans_Function = ReLU Dense_Layer = 1, Trans_Function = Sigmoid

Table 3. Model Specification of Dataset 2.

Technique Used	Specifications
Random Forest	Estimators = 5000, Max Depth = 25, Min Samples Split = 5, Criterion = Entropy
ERT	Criterion = Gini, Min Samples Split = 10, Estimators = 5000
AdaBoost	Estimators = 2000, Learning-Rate = 0.01, Random State = 10
XGM	Random State = 1, Learning-Rate = 0.01, Estimators = 6000, Max-Depth = 15
GBM	Learning Rate = 0.001, Random State = 1, Estimators = 7000
Bagging	Estimators = 500, Random State = 10
Stacking	Cv = 10, Estimators = [LR, KNN, DT, RF, SVC], Final Estimator = LR
DT	Max Depth = 2, Criterion = Gini, Random State = 99
kNN	Neighbors = 38
LR	Solver = Liblinear, Max Iter = 10,000
ANN	Dense_Layer = 32, Trans_Function = ReLU Dense_Layer = 16, Trans_Function = ReLU Dense_Layer = 8, Trans_Function = ReLU Dense_Layer = 1, Trans_Function = Sigmoid Optimizers = Adam, Learning rate = 0.001 Loss = Binary Crossentropy
CNN	Conv1d = 128, Activation = ReLU Conv1d = 128, Activation = ReLU Dense_Layer = 64, Trans_Function = Tanh Dense_Layer = 32, Trans_Function = Tanh Dense_Layer = 16, Trans_Function = ReLU Dense_Layer = 1, Trans_Function = Sigmoid

3.4. Performance Measures

Different performance measures such as accuracy, sensitivity, precision, F1 score, specificity, and AUC-ROC value of the suggested predictive system for consumer behavior were assessed in this study.

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

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

$$Sensitivity = True\ Positive\ Rate(TPR) = \frac{TP}{TP + FN}$$

$$Specificity = \frac{TN}{TN + FP}$$

$$F1\ score = 2 \times \frac{Precision \times Sensitivity}{Precision + Sensivity}$$

where TP = true positive, FP = false positive, TN = true negative, and FN = false negative. The accuracy rate is the percentage of correctly foretold positive outcomes.

The receiver operating characteristic (ROC) is used to describe the diagnostic capability of a classifier which is represented by a curve. Plotting the true positive rate (TPR) versus the false positive rate (FPR) at several settings results in the curve.

$$FPR = \frac{FP}{N} = \frac{FP}{(FP + TN)} = 1 - TNR$$

4. Results and Discussion

In order to predict customer churn, various supervised learning approaches have been used in this study on behavioral data from consumers. The data were preprocessed to remove the high value, as was covered in the section before this one. The churn was labeled for classification after filtering. Here, CNN, as a deep learning technique, some ensemble learning methods, and a few classical classifications have all been used. The research model was verified using k-fold cross validation, where k = 10 for both of the models. Tables 4 and 5 show that CNN and ANN returned the best accuracy, among all of the techniques which were applied, for both of the datasets: 99% and 98% on the first dataset, and 98% and 99% on second dataset, respectively. The AUC score and AUC-ROC curve were used to visualize the performance of the classifiers. CNN and ANN revealed the most efficient outcome among all of the machine learning techniques which were applied, with an AUC score of 0.99 using both the techniques on the first dataset and 0.99 and 0.96 on the second. On the first dataset, all of the models exhibited an accuracy over 90% and the AUC scores for all of the ensemble models were valued between 0.70 and 0.76. On the second dataset, all of the ensemble models except for AdaBoost exhibited an accuracy over 90% and the AUC value for all of the ensemble models except for AdaBoost were between 0.80 and 0.89. The traditional classifiers, except for ANN, exhibited an accuracy over 90% on the first dataset; however, when applied on the second dataset, the accuracy was below 90%. kNN and LR exhibited lowest AUC scores on both of the datasets.

Table 4. Performance of the Classifier Models for Dataset 1.

	Classifiers	Accuracy (%)	Precision	Sensitivity	Specificity	F1-Score	AUC
	RF	94	0.98	0.95	0.74	0.97	0.72
	ERT	94	0.93	0.94	0.76	0.93	0.70
	AdaBoost	94	0.98	0.96	0.72	0.97	0.75
ELT	XGM	94	0.98	0.95	0.67	0.97	0.73
	GBM	94	0.98	0.96	0.70	0.97	0.74
	Bagging	94	0.98	0.95	0.71	0.97	0.74
	Stacking	95	0.98	0.95	0.71	0.97	0.73
	DT	94	0.98	0.95	0.75	0.97	0.70
TC	k-NN	92	0.99	0.93	0.60	0.96	0.60
	LR	93	0.97	0.93	0.67	0.96	0.61
	ANN	98	0.92	0.92	0.86	0.92	0.99
DLT	CNN	99	0.99	0.99	0.99	0.99	0.99

Table 5. Performance Of The Classifier Models for Dataset 2.

	Classifiers	Accuracy (%)	Precision	Sensitivity	Specificity	F1-Score	AUC
	RF	95	0.99	0.95	0.93	0.97	0.84
	ERT	95	0.95	0.95	0.86	0.95	0.87
	AdaBoost	88	0.98	0.89	0.66	0.93	0.61
ELT	XGM	96	0.99	0.96	0.91	0.97	0.87
	GBM	95	0.99	0.95	0.94	0.97	0.84
	Bagging	94	0.98	0.95	0.86	0.96	0.85
	Stacking	95	0.97	0.96	0.84	0.97	0.88
	DT	87	0.93	0.91	0.53	0.92	0.70
TC	k-NN	86	0.91	0.86	0.71	0.93	0.52
	LR	85	0.96	0.87	0.45	0.91	0.58
	ANN	99	0.98	0.96	0.98	0.97	0.96
DLT	CNN	98	0.99	0.99	0.97	0.99	0.99

Figures 6 and 7 show the accuracy, using a bar graph plot and line graph plot, of the AUC score measures for all of the classifiers on both the datasets used in this work. In both of the graphs, the distinct difference in the values for both of the datasets can be visualized. The accuracy plot helps to visualize that Adaboost, DT, kNN, and LR achieved better efficacy on the first dataset than the second; whereas bagging and stacking performed the same on both datasets. The AUC graph shows that all of the ensemble models, except Adaboost, exhibited a better score on the second dataset, and all of the other models exhibited similar scores on both datasets.
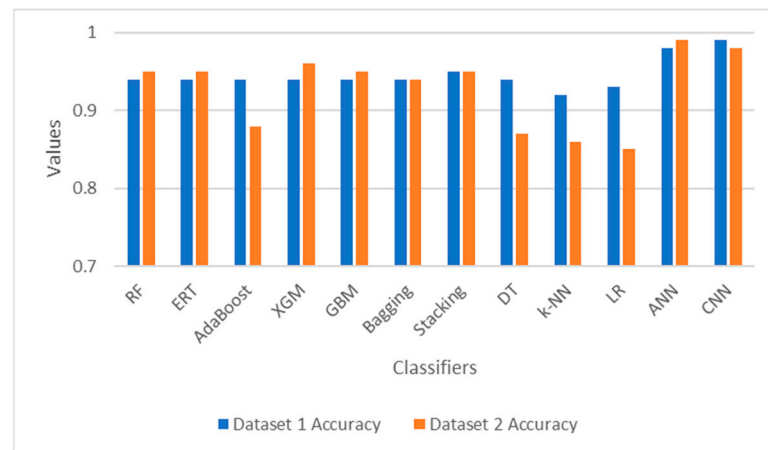

Figure 6. Accuracy plot of different classifiers on both datasets.
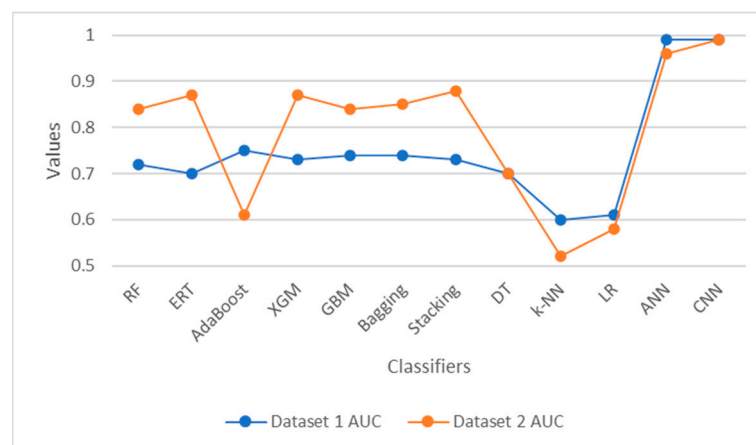
Figure 7. AUC score plot of different classifiers on both datasets.

The line graph plot for sensitivity and F1 scores for all of the classifiers for both datasets are shown in Figure 8. The F1 score is biased toward the genuine positive rate because it depends on the accuracy and sensitivity values. Therefore, it may be deduced from the F1 score and sensitivity that there were fewer false positive samples.

The accuracy of each classifier is displayed as a bar plot in Figure 9, along with the matching AUC ratings. This graph shows that, despite having a high accuracy value, some models showed quite a low AUC score.

Figures 10 and 11 show the bar plot of the precision and specificity for both of the datasets. While the precision plot shows a high true positive rate, the specificity graph shows a comparatively lower true negative rate, which is more definitive in respect to the efficacy of the ensemble models on the first dataset and on some of the traditional classifier's performances on the second dataset.

Figure 8. F1 score and sensitivity plot for both datasets.
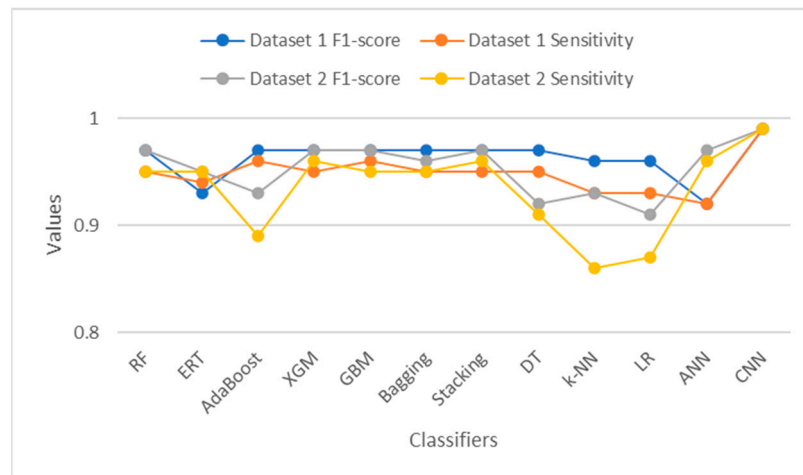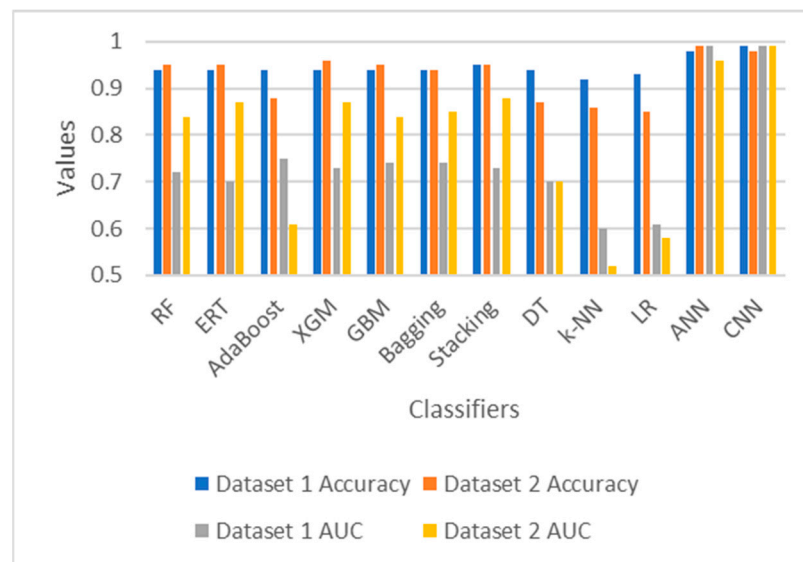

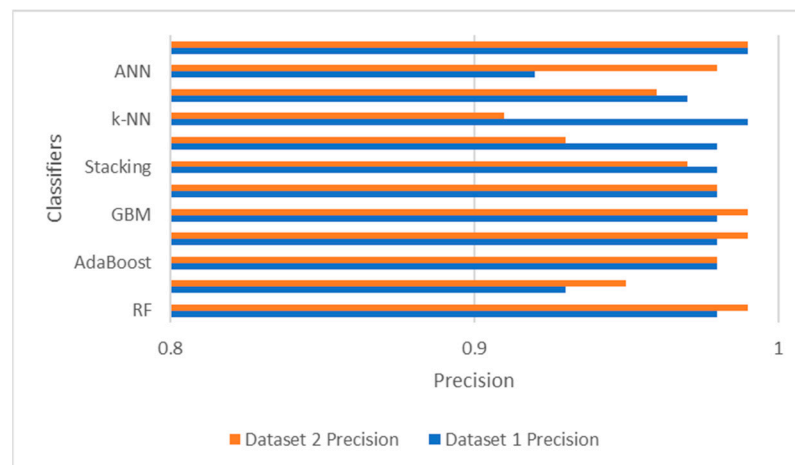**Figure 9.** Accuracy and AUC plot for both datasets.

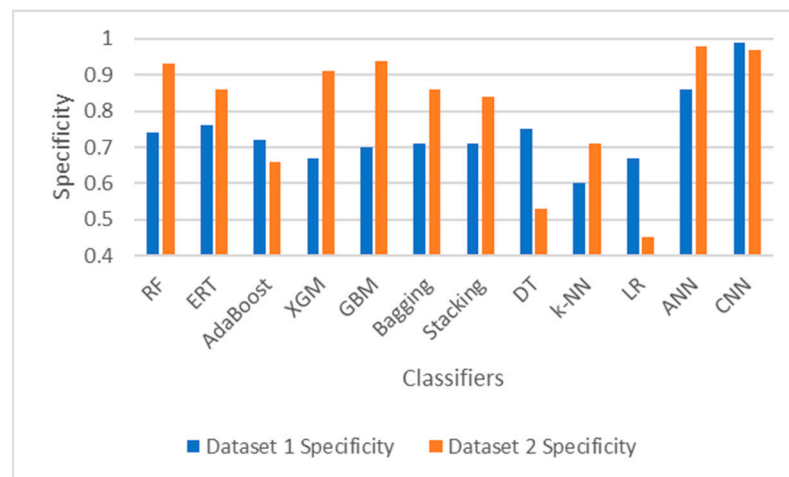**Figure 10.** Precision plot for both datasets.

Figure 11. Specificity plot for both datasets.

Figure 12 shows the AUC-ROC curve plot for the XGB, stacking ensemble model, ANN, and CNN on both datasets, respectively. These four models exhibited the best AUC scores among all of the techniques which were applied to both datasets. Here, we can see the almost perfect curve for both ANN and CNN because of the high AUC value which was obtained. It can be seen that XGB and stacking have performed better on the second dataset compared to the first dataset.

In Figure 13, a bar graph is shown which compares the existing studies on churn prediction with this proposed work. The graph shows that this work has achieved better results than most of the other existing works, predicting the churn customers with much higher accuracy.

From the results above, it could be argued that AI-driven customer churn prediction, as well as machine learning and smart data, could assist businesses in knowing their customers and understanding the demands of those customers. Such an understanding can help to retain existing customers, which is cost-effective, as this is 5–10 times cheaper than the cost of acquiring new consumers [6]. Meanwhile, mMachine learning models can significantly influence social media marketing in several ways, among which are:

- Targeted Advertising: machine learning models can analyze customer data such as demographics, interests, and online behavior to identify potential target audiences for specific products and services. This allows for more effective and efficient targeting, leading to higher conversion rates.
- Content Optimization: machine learning algorithms can help businesses determine the best times to post content and the types of content that perform best, leading to increased customers engagement and reach.
- Advertising Optimization: machine learning models can be used to automate and optimize advert placement and bid prices, saving time and resources while improving overall campaign performance.
- Sentiment Analysis: machine learning models can be used to analyze customer sentiment and feedback, providing valuable insights into customer opinions and preferences. Such information can then be used to improve the company marketing strategies and campaigns.

Figure 12. *Cont.*

(**g**)

(**h**)
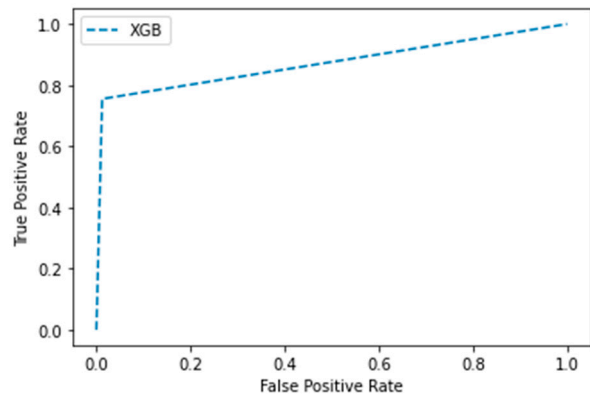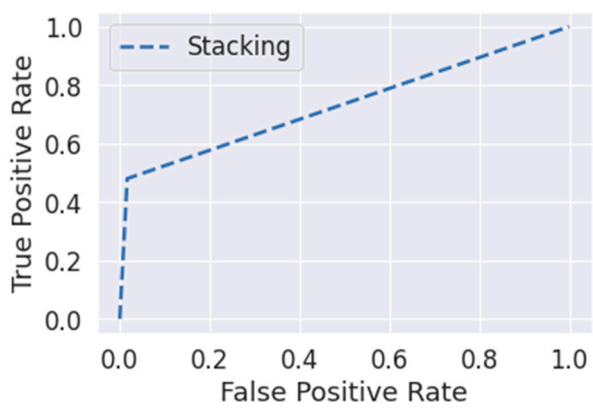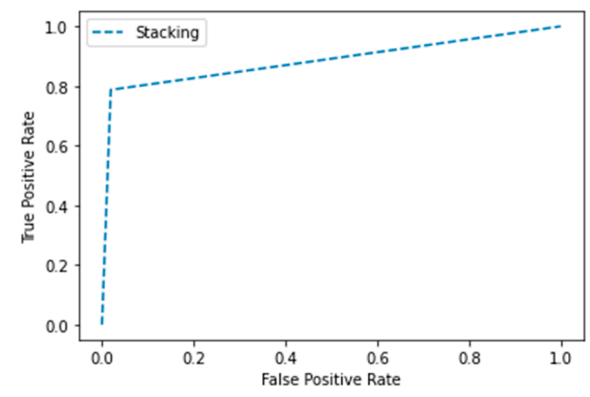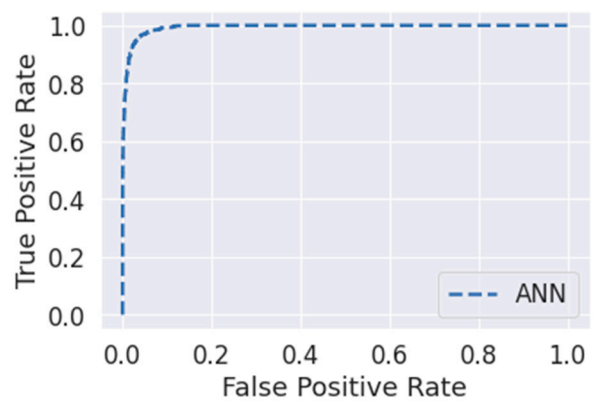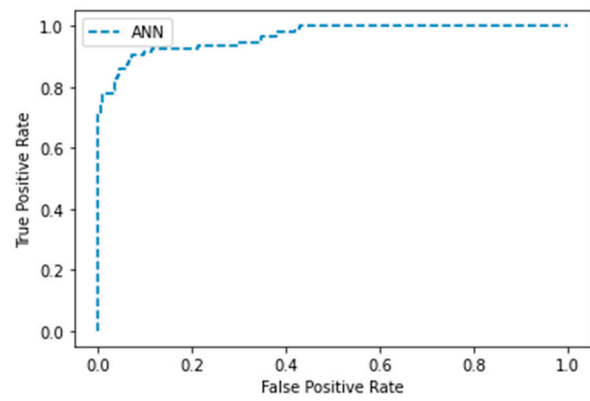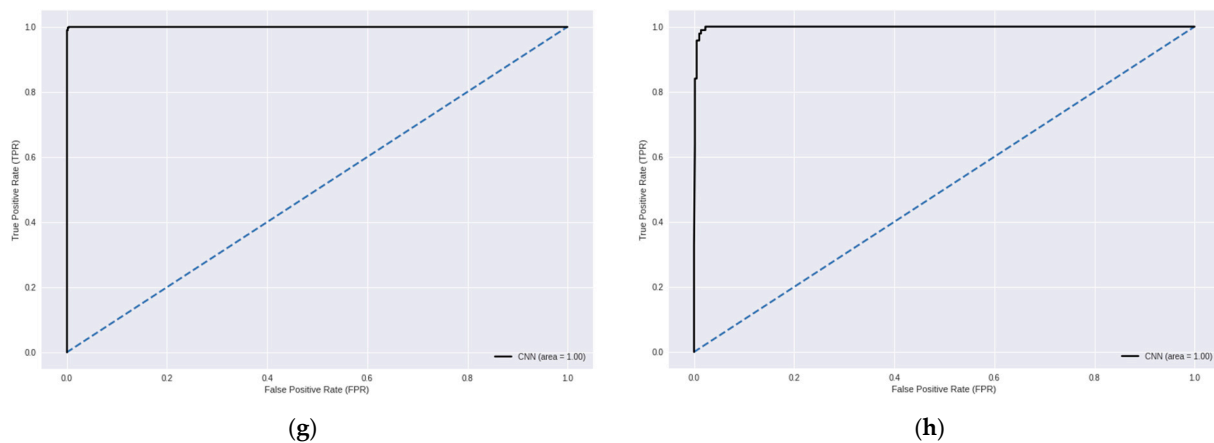
Figure 12. AUC-ROC curve of (**a**) XGB for dataset 1; (**b**) XGB for dataset 2; (**c**) stacking for dataset 1; (**d**) stacking for dataset 2; (**e**) ANN for dataset 1; (**f**) ANN for dataset 2; (**g**) CNN for dataset 1; (**h**) CNN for dataset 2.
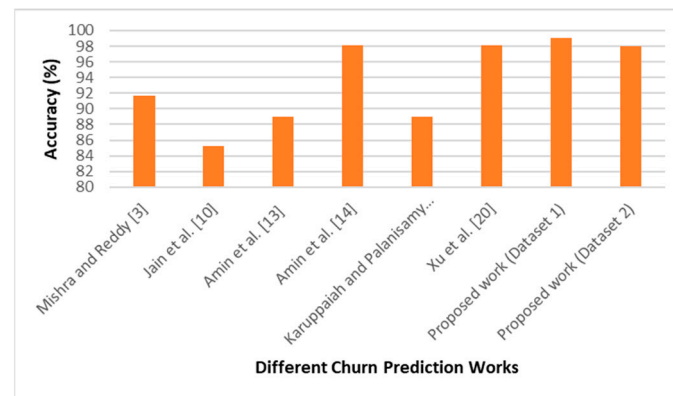
Figure 13. Bar graph for comparison based on accuracy of different published works on churn prediction.

As such, machine learning models can significantly improve the effectiveness and efficiency of social media marketing efforts.

Further, such machine learning-based tools are not only easy to use and quick to operate but are also capable of learning from their own data, which contributes to a positive experience for clients and encourages repeat business. Machine-learning-based churn prediction systems have the capability to learn from their preceding errors and enhance their results through collecting historical data [47]. Because of this, we are able to determine the value of each individual consumer, forecast future expenses and income, and determine the areas in which the majority of marketing efforts regarding social media or other platforms should be focused.

Statistical Analysis

ANOVA and Wilcoxon rank-sum tests were applied to statistically assess the quality of the algorithms as applied to the two public datasets: Southeast Asian telecom industry, and American telecom market (as shown in Tables 6–9). Tables 6 and 7 show that the *p*-values are less than 0.05, which indicates that there is a statistically significant difference between the groups. In Tables 8 and 9, the obtained *p*-value > 0.05 shows that the results have no significant difference.

Table 6. ANOVA for Dataset 1.

ANOVA Table	SS	DF	MS	F (DFn, DFd)	*p* Value
Treatment (between columns)	0.04342	11	0.003947	F (11, 108) = 424.4	*p* < 0.0001
Residual (within columns)	0.001004	108	9.3×10^{-6}		
Total	0.04442	119			

Table 7. ANOVA for Dataset 2.

ANOVA Table	SS	DF	MS	F (DFn, DFd)	*p* Value
Treatment (between columns)	0.2648	11	0.02407	F (11, 108) = 700.7	*p* < 0.0001
Residual (within columns)	0.00371	108	3.44×10^{-5}		
Total	0.2685	119			

Table 8. Wilcoxon Signed Rank Test Dataset 1.

	RF	ERT	AdaBoost	XGM	GBM	Bagging	Stacking	DT	k-NN	LR	ANN	CNN
Theoretical median	0	0	0	0	0	0	0	0	0	0	0	0
Actual median	0.94	0.94	0.94	0.94	0.94	0.94	0.95	0.94	0.92	0.93	0.98	0.99
Number of values	10	10	10	10	10	10	10	10	10	10	10	10
Wilcoxon Signed Rank Test												
Sum of signed ranks (W)	55	55	55	55	55	55	55	55	55	55	55	55
Sum of positive ranks	55	55	55	55	55	55	55	55	55	55	55	55
Sum of negative ranks	0	0	0	0	0	0	0	0	0	0	0	0
p value (two tailed)	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.002
Exact or estimate?	Exact	Exact	Exact	Exact	Exact	Exact	Exact	Exact	Exact	Exact	Exact	Exact
p value summary	-	-	-	-	-	-	-	-	-	-	-	-
Significant (alpha = 0.05)?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Table 9. Wilcoxon Signed Rank Test Dataset 2.

	RF	ERT	AdaBoost	XGM	GBM	Bagging	Stacking	DT	k-NN	LR	ANN	CNN
Theoretical median	0	0	0	0	0	0	0	0	0	0	0	0
Actual median	0.95	0.95	0.88	0.96	0.95	0.94	0.95	0.87	0.86	0.85	0.99	0.98
Number of values	10	10	10	10	10	10	10	10	10	10	10	10
Wilcoxon Signed Rank Test												
Sum of signed ranks (W)	55	55	55	55	55	55	55	55	55	55	55	55
Sum of positive ranks	55	55	55	55	55	55	55	55	55	55	55	55
Sum of negative ranks	0	0	0	0	0	0	0	0	0	0	0	0
p value (two tailed)	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.002
Exact or estimate?	Exact	Exact	Exact	Exact	Exact	Exact	Exact	Exact	Exact	Exact	Exact	Exact
p value summary	-	-	-	-	-	-	-	-	-	-	-	-
Significant (alpha = 0.05)?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

5. Conclusions

The telecom industry has become highly competitive over the last two decades. Hence, having a highly efficient churn predictive system is very important. Therefore, in this work several supervised machine learning techniques have been tested on two churn datasets from two different countries' telecom markets and with different features. In both of the scenarios, ANN and CNN returned better results. The other techniques which were used are RF, ERT, AdaBoost, XGB, GBM, bagging, DT, kNN, and LR. All of these techniques performed quite well on both of the datasets, with accuracies over 90% for all of them except for AdaBoost, DT, kNN, and LR, which did not perform well on the second dataset even though they returned a satisfactory result for the first dataset. As shown in Section 3.2, the churn distribution is drastically imbalanced, which is very obvious considering that the number of churning customers should always be lower than the non-churning customers. However, this imbalance in the data strongly disturbs the performance of the applied model. As inferred from the specificity values, which is also known as the true negative rate, these models have a higher detection rate of the non-churn customers compared to the churners [48]. This problem was solved using CNN, which exhibited a very high true negative prediction rate, which is the churn customers in this scenario. The efficiency of the other models can be upgraded by using data balancing; however, when the percentage of the churn customers is too low, even data balancing may not improve the performance. Hence, using datasets containing a higher number of churn customers to train the models is the best-case scenario. Within the same line of benefits, deep churn prediction is a very useful tool in dealing with social media marketing activities for any company, as it helps to identify those customers who are at risk of leaving a business, or "churn." Through predicting churn, a business can take proactive steps and marketing activities to retain their customers and prevent them from leaving, leading to increased customer loyalty, revenue, and preventing the loss of valuable customers.

Thus, in the context of social media marketing, deep churn prediction models can analyze customer behavior, engagement, and other factors to pinpoint customers who are at risk of leaving. Such information can then be used to target these customers with personalized marketing campaigns and incentives to retain them. For example, a deep churn prediction model may identify that a customer who has stopped engaging with a company's social media accounts is at high risk of churning. The company can then target this customer with a personalized communication, offer, or special promotion to encourage them to remain a customer and reengage with the company. In conclusion, deep churn prediction is a valuable tool for social media marketers.

Any research on data can be improved by enhancing the quality and the quantity of the same [49]. Churn prediction is no exception to this condition. A higher quality of customer data can return a higher accuracy in prediction, as well as increase company profit.

This research can also be enhanced by including data from both structured and unstructured sources, which can also help with increasing the quantity of the data. The customers' behavioral and demographic data should also be analyzed to obtain an improved understanding of their requirements, which will help to reduce the churn rate [50,51]. Non-traditional sources of data such as social media for product sentiments, call centers for number of complaints, offers from competitors, and so on should be taken into consideration for better customer retention.

Author Contributions: Conceptualization, L.S. and H.K.T.; methodology, L.S., H.K.T. and T.G.; software, L.S.; validation, T.G., H.K.T. and H.E.-G.; investigation, H.K.T. and E.-S.M.E.-k.; resources, H.K.T. and H.E.-G.; data curation, L.S. and E.-S.M.E.-k.; writing—original draft preparation, L.S. and H.K.T.; writing—review and editing, T.G., H.E.-G. and E.-S.M.E.-k.; visualization, L.S.; supervision, H.E.-G. and T.G. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Two datasets were used. The first dataset is from the Indian and Southeast Asian telecom industry [40] and the second dataset is from American telecom market [41].

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Mishra, A.; Reddy, U.S. A comparative study of customer churn prediction in telecom industry using ensemble-based classifiers. In Proceedings of the 2017 International Conference on Inventive Computing and Informatics (ICICI), Coimbatore, India, 23–24 November 2017; IEEE: Piscataway, NJ, USA, 2017; pp. 721–725.
2. Shirazi, F.; Mohammadi, M. A big data analytics model for customer churn prediction in the retiree segment. *Int. J. Inf. Manag.* **2019**, *48*, 238–253. [CrossRef]
3. Bhattacharyya, J.; Dash, M.K. Investigation of customer churn insights and intelligence from social media: A netnographic research. *Online Inf. Rev.* **2020**, *45*, 174–206. [CrossRef]
4. Ahmad, A.K.; Jafar, A.; Aljoumaa, K. Customer churn prediction in telecom using machine learning in big data platform. *J. Big Data* **2019**, *6*, 28. [CrossRef]
5. Coussement, K.; Lessmann, S.; Verstraeten, G. A comparative analysis of data preparation algorithms for customer churn prediction: A case study in the telecommunication industry. *Decis. Support Syst.* **2017**, *95*, 27–36. [CrossRef]
6. Ly, T.V.; Son, D.V.T. Churn prediction in telecommunication industry using kernel Support Vector Machines. *PLoS ONE* **2022**, *17*, e0267935.
7. De Caigny, A.; Coussement, K.; De Bock, K.W. A new hybrid classification algorithm for customer churn prediction based on logistic regression and decision trees. *Eur. J. Oper. Res.* **2018**, *269*, 760–772. [CrossRef]
8. El-Gohary, H.; Trueman, M.; Fukukawa, K. Understanding the factors affecting the adoption of E-Marketing by small business enterprises. In *E-Commerce Adoption and Small Business in the Global Marketplace*; Thomas, B., Simmons, G., Eds.; IGI Global: Hershey, PA, USA, 2009; pp. 237–258.
9. El-Gohary, H. E-Marketing: Towards a conceptualization of a new marketing philosophy e book chapter. In *E-Business Issues, Challenges and Opportunities for SMEs: Driving Competitiveness*; IGI Global: Hershey, PA, USA, 2010.
10. Jain, N.; Tomar, A.; Jana, P.K. A novel scheme for employee churn problem using multi-attribute decision making approach and machine learning. *J. Intell. Inf. Syst.* **2021**, *56*, 279–302. [CrossRef]
11. Indian and Southeast Asian Telecom Industry Dataset Which Is. Available online: https://www.kaggle.com/datasets/priyankanavgire/telecom-churn (accessed on 22 March 2021).
12. American Telecom Market Dataset. Available online: https://www.kaggle.com/datasets/mnassrib/telecom-churn-datasets (accessed on 18 February 2020).
13. Zdravevski, E.; Lameski, P.; Apanowicz, C.; Ślęzak, D. From Big Data to business analytics: The case study of churn prediction. *Appl. Soft Comput.* **2020**, *90*, 106164. [CrossRef]
14. Vo, N.N.; Liu, S.; Li, X.; Xu, G. Leveraging unstructured call log data for customer churn prediction. *Knowl.-Based Syst.* **2021**, *212*, 106586. [CrossRef]
15. Höppner, S.; Stripling, E.; Baesens, B.; vanden Broucke, S.; Verdonck, T. Profit driven decision trees for churn prediction. *Eur. J. Oper. Res.* **2020**, *284*, 920–933. [CrossRef]
16. Stripling, E.; vanden Broucke, S.; Antonio, K.; Baesens, B.; Snoeck, M. Profit maximizing logistic regression modeling for customer churn prediction. In Proceedings of the 2015 IEEE International Conference on Data Science and Advanced Analytics (DSAA), Paris, France, 19–21 October 2015; IEEE: Piscataway, NJ, USA, 2015; pp. 1–10.
17. Arifin, S.; Samopa, F. Analysis of Churn Rate Significantly Factors in Telecommunication Industry Using Support Vector Machines Method. In *Journal of Physics: Conference Series*; IOP Publishing: Bristol, UK, 2018; Volume 1108, p. 012018.
18. Jain, H.; Khunteta, A.; Srivastava, S. Churn prediction in telecommunication using logistic regression and logit boost. *Procedia Comput. Sci.* **2020**, *167*, 101–112. [CrossRef]
19. Amin, A.; Shah, B.; Khattak, A.M.; Moreira, F.J.L.; Ali, G.; Rocha, Á.; Anwar, S. Cross-company customer churn prediction in telecommunication: A comparison of data transformation methods. *Int. J. Inf. Manag.* **2019**, *46*, 304–319. [CrossRef]
20. Amin, A.; Al-Obeidat, F.; Shah, B.; Adnan, A.; Loo, J.; Anwar, S. Customer churn prediction in telecommunication industry using data certainty. *J. Bus. Res.* **2019**, *94*, 290–301. [CrossRef]
21. Amin, A.; Anwar, S.; Adnan, A.; Nawaz, M.; Alawfi, K.; Hussain, A.; Huang, K. Customer churn prediction in the telecommunication sector using a rough set approach. *Neurocomputing* **2017**, *237*, 242–254. [CrossRef]
22. Alboukaey, N.; Joukhadar, A.; Ghneim, N. Dynamic behavior based churn prediction in mobile telecom. *Expert Syst. Appl.* **2020**, *162*, 113779. [CrossRef]
23. Karuppaiah, K.S.; Palanisamy, N.G. Heterogeneous ensemble stacking with minority upliftment (HESMU) for churn prediction on imbalanced telecom data. In *Materials Today: Proceedings*; Elsevier: London, UK, 2021.
24. De Caigny, A.; Coussement, K.; De Bock, K.W.; Lessmann, S. Incorporating textual information in customer churn prediction models based on a convolutional neural network. *Int. J. Forecast.* **2020**, *36*, 1563–1578. [CrossRef]

25. Mitrović, S.; Baesens, B.; Lemahieu, W.; De Weerdt, J. On the operational efficiency of different feature types for telco Churn prediction. *Eur. J. Oper. Res.* **2018**, *267*, 1141–1155. [CrossRef]

26. De Bock, K.W.; De Caigny, A. Spline-rule ensemble classifiers with structured sparsity regularization for interpretable customer churn modeling. *Decis. Support Syst.* **2021**, *150*, 113523. [CrossRef]

27. Xu, T.; Ma, Y.; Kim, K. Telecom Churn Prediction System Based on Ensemble Learning Using Feature Grouping. *Appl. Sci.* **2021**, *11*, 4742. [CrossRef]

28. Óskarsdóttir, M.; Van Calster, T.; Baesens, B.; Lemahieu, W.; Vanthienen, J. Time series for early churn detection: Using similarity based classification for dynamic networks. *Expert Syst. Appl.* **2018**, *106*, 55–65. [CrossRef]

29. Chakrabarti, S.; Swetapadma, A.; Pattnaik, P.K. A channel independent generalized seizure detection method for pediatric epileptic seizures. *Comput. Methods Programs Biomed.* **2021**, *209*, 106335. [CrossRef] [PubMed]

30. Alsouda, Y.; Pllana, S.; Kurti, A. Iot-based urban noise identification using machine learning: Performance of SVM, KNN, bagging, and random forest. In Proceedings of the International Conference on Omni-Layer Intelligent Systems, Heraklion, Crete, Greece, 5–7 May 2019; pp. 62–67.

31. Geurts, P.; Ernst, D.; Wehenkel, L. Extremely randomized trees. *Mach. Learn.* **2006**, *63*, 3–42. [CrossRef]

32. Touzani, S.; Granderson, J.; Fernandes, S. Gradient boosting machine for modeling the energy consumption of commercial buildings. *Energy Build.* **2018**, *158*, 1533–1543. [CrossRef]

33. Beygelzimer, A.; Hazan, E.; Kale, S.; Luo, H. Online gradient boosting. *arXiv* **2015**, arXiv:1506.04820.

34. Ma, B.; Meng, F.; Yan, G.; Yan, H.; Chai, B.; Song, F. Diagnostic classification of cancers using extreme gradient boosting algorithm and multi-omics data. *Comput. Biol. Med.* **2020**, *121*, 103761. [CrossRef]

35. Sheridan, R.P.; Wang, W.M.; Liaw, A.; Ma, J.; Gifford, E.M. Extreme gradient boosting as a method for quantitative structure–activity relationships. *J. Chem. Inf. Model.* **2016**, *56*, 2353–2360. [CrossRef]

36. Wang, F.; Jiang, D.; Wen, H.; Song, H. Adaboost-based security level classification of mobile intelligent terminals. *J. Supercomput.* **2019**, *75*, 7460–7478. [CrossRef]

37. Freund, Y.; Schapire, R.E. Experiments with a new boosting algorithm. In *ICML*; ACM Digital Library: New York City, NY, USA, 1996; Volume 96, pp. 148–156.

38. Drucker, H.; Schapire, R.; Simard, P. Improving performance in neural networks using a boosting algorithm. *Adv. Neural Inf. Process. Syst.* **1993**, *5*, 42–49.

39. Sreng, S.; Maneerat, N.; Hamamoto, K.; Panjaphongse, R. Automated diabetic retinopathy screening system using hybrid simulated annealing and ensemble bagging classifier. *Appl. Sci.* **2018**, *8*, 1198. [CrossRef]

40. Savkovic, B.; Kovac, P.; Dudic, B.; Gregus, M.; Rodic, D.; Strbac, B.; Ducic, N. Comparative Characteristics of Ductile Iron and Austempered Ductile Iron Modeled by Neural Network. *Materials* **2019**, *12*, 2864. [CrossRef]

41. Chakrabarti, S.; Swetapadma, A.; Ranjan, A.; Pattnaik, P.K. Time domain implementation of pediatric epileptic seizure detection system for enhancing the performance of detection and easy monitoring of pediatric patients. *Biomed. Signal Process. Control* **2020**, *59*, 101930. [CrossRef]

42. Liao, Y.; Vemuri, V.R. Use of k-nearest neighbor classifier for intrusion detection. *Comput. Secur.* **2002**, *21*, 439–448. [CrossRef]

43. Saha, L.; Tripathy, H.K.; Nayak, S.R.; Bhoi, A.K.; Barsocchi, P. Amalgamation of Customer Relationship Management and Data Analytics in Different Business Sectors—A Systematic Literature Review. *Sustainability* **2021**, *13*, 5279. [CrossRef]

44. Ghasemi Darehnaei, Z.; Shokouhifar, M.; Yazdanjouei, H.; Rastegar Fatemi, S.M.J. SI-EDTL: Swarm intelligence ensemble deep transfer learning for multiple vehicle detection in UAV images. *Concurr. Comput. Pract. Exp.* **2022**, *34*, e6726. [CrossRef]

45. Saha, L.; Tripathy, H.K.; Sahoo, L. Business Intelligence Influenced Customer Relationship Management in Telecommunication Industry and Its Security Challenges. In *Privacy and Security Issues in Big Data*; Springer: Singapore, 2021; pp. 175–188.

46. Abiyev, R.; Arslan, M.; Bush Idoko, J.; Sekeroglu, B.; Ilhan, A. Identification of epileptic EEG signals using convolutional neural networks. *Appl. Sci.* **2020**, *10*, 4089. [CrossRef]

47. Fujo, S.W.; Subramanian, S.; Khder, M.A. Customer Churn Prediction in Telecommunication Industry Using Deep Learning. *Inf. Sci. Lett.* **2022**, *11*, 24.

48. Sudharsan, R.; Ganesh, E.N. A Swish RNN based customer churn prediction for the telecom industry with a novel feature selection strategy. *Connect. Sci.* **2022**, *34*, 1855–1876. [CrossRef]

49. Saha, L.; Tripathy, H.K.; Masmoudi, F.; Gaber, T. A Machine Learning Model for Personalized Tariff Plan based on Customer's Behavior in the Telecom Industry. *Int. J. Adv. Comput. Sci. Appl. (IJACSA)* **2022**, *13*, 2022. [CrossRef]

50. Sana, J.K.; Abedin, M.Z.; Rahman, M.S.; Rahman, M.S. A novel customer churn prediction model for the telecommunication industry using data transformation methods and feature selection. *PLoS ONE* **2022**, *17*, e0278095. [CrossRef]

51. Adhikary, D.D.; Gupta, D. Applying over 100 classifiers for churn prediction in telecom companies. *Multimed. Tools Appl.* **2021**, *80*, 35123–35144. [CrossRef]