DOI: 10.15849/IJASCA.221128.04

Int. J. Advance Soft Compu. Appl, Vol. 14, No. 3, November 2022
Print ISSN: 2710-1274, Online ISSN: 2074-8523

Supervised Learning Algorithms for Predicting Customer Churn with Hyperparameter Optimization

Manal Loukili, Fayçal Messaoudi, and Mohammed El Ghazi

National School of Applied Sciences,
Sidi Mohamed Ben Abdellah University, Fez, Morocco.
manal.loukili@usmba.ac.ma
National School of Business and Management,
Sidi Mohamed Ben Abdellah University, Fez, Morocco.
faycal.messaoudi@usmba.ac.ma
Superior School of Technology,
Sidi Mohamed Ben Abdellah University, Fez, Morocco.
mohammed.elghazi@usmba.ac.ma

Abstract

Churn risk is one of the most worrying issues in the telecommunications industry. The methods for predicting churn have been improved to a great extent by the remarkable developments in the word of artificial intelligence and machine learning. In this context, a comparative study of four machine learning models was conducted. The first phase consists of data preprocessing, followed by feature analysis. In the third phase, feature selection. Then, the data is split into the training set and the test set. During the prediction phase, some of the commonly used predictive models were adopted, namely k-nearest neighbor, logistic regression, random forest, and support vector machine. Furthermore, we used cross-validation on the training set for hyperparameter adjustment and for avoiding model overfitting. Next, the hyperparameters were adjusted to increase the models' performance. The results obtained on the test set were evaluated using the feature weights, confusion matrix, accuracy score, precision, recall, error rate, and f1 score. Finally, it was found that the support vector machine model outperformed the other prediction models with an accuracy equal to 96.92%.

Keywords: *Churn Prediction, Classification Algorithms, Hyperparameter Optimization, Machine Learning, Telecommunications.*

1 Introduction

Due to globalization and advances in the telecommunications industry, the exponential growth in the number of operators in the market is leading to increased competition. In this competitive environment, it has become essential to continuously maximize profits. Therefore, different approaches have been developed, such as acquisition of new customers, incentive selling to existing customers, and strengthening the loyalty period of existing customers [1]. However, the strategy of retaining existing customers is the simplest, as well as being less expensive compared to other strategies and more profitable. Companies that implement this strategy are driven to minimize the rate of customer churning which is mainly due to dissatisfaction with the services provided to consumers and the support mechanism. To overcome this problem, it is necessary to predict which customers are likely to churn. Churn prediction models aim to detect early warning signals of churn. Machine learning has made it possible to effectively combat churn through the predictive approach [2]. By being able to anticipate user behaviors, companies can measure the customer relationship and act beforehand to avoid churn. In this way, companies can engage in marketing actions that are more profitable than acquiring new customers. From a machine learning perspective, churn prediction is a supervised classification learning problem. This prediction can be done based on several machine learning classification algorithms.

In this paper, we set up a machine learning process based on a dataset from Kaggle (Telco Customer Churn). Then we compared the performances of the different models used, namely k-nearest neighbor, logistic regression, random forest, and support vector machine. The pipeline used consists of the following steps:

- Problem Definition
- Data Collection
- Exploratory Data Analysis and Data Cleaning
- Data visualization
- Feature Engineering
- Train set /Test set Split
- Training, Prediction, and Assessment
- Hyperparameters Tunning
- Performance Metrics comparison

This paper is structured as follows. Section 2 provides a literature review, section 3 describes the problem and the dataset used, section 4 describes the machine learning models employed, section 5 describes the methodology, and the experimental results and analysis are shown in section 6. The conclusion is stated in the last section.

2 Literature Review

This literature review examines various studies by several researchers in the telecommunications industry to make a prediction of customer churn behavior.

Huang, F et al. [3] (2015) highlights the 3Vs which are volume, variety, and velocity. Prediction can be achieved efficiently by using a large amount of training data from a large number of features and raising the velocity of data processing. The result of this method yielded an accuracy rate of 0.96.

Kumar, N et al. [4] (2017) compared three machine learning algorithms for churn prediction namely logistic regression, random forest, and balanced random forest. They used PCA for feature selection. As a result, it was found that the logistic regression model had the greatest AUC score of 0.86 compared to the other models.

Zhang, Z et al. [5] (2018) mentioned that customer churn analysis is very useful for telecommunications companies for retain weight clients. The customer churn prediction model with more accuracy is highly weighted for the decision of customer retention. In this study, SVM model is employed because it is better for precision. As it can solve the samples in a low dimensional space which is linear inspirable in a two-dimensional space. However, the limitation of the proposed model is that it is very complex to measure the unsubscribed customers.

Bharat, A (2019) [6] adopted a model that focuses on the customer's activity scheme. It is built on the customer's activity by searching for the mean duration of idle time and idle frequencies. This method can also be used in other areas for the prediction of customer churn.

Andrews, R et al. [7] (2019) used a dataset of 10,000 customer records from a telecommunications company. They used deep learning models and 10-overlap cross-approval methods to assess prediction accuracy. The results yielded an AUC score of 0.89.

Calzada, L et al. [8] (2020) conducted a comparative study of two methods for predicting churn, namely Time-Order-Graph and Aggregated-Static-Graph, and a forest classifier based on three threshold metrics to assess the quality of prediction of the similarity forest classifier against each centrality metric.

Wang, X et al. [9] (2020) presented a comparative analysis, comparing popular classification approaches used to overcome the customer churning in the telecommunications industry. Based on the performance of the models, they selected the models with higher performance, namely Light GBM, XGBoost, Random Forest and Decision Tree. Then they developed an ensemble model using the soft voting technique, which obtained the highest AUC score.

Nam, N [10] (2021) also compared two prediction methodes, namely SMOTE and Deep Belief Network (DBN), with cost-sensitive learning approaches, such as Focal Loss and Weighted Loss. This study revealed higher overall performance of Focal Loss and Weighted Loss in comparison to SMOTE and DBN.

Siddika, A. et al. [11] (2021) used machine learning and deep learning techniques to determine which customers drop out and the salient factors that cause them to drop out. The Random Forest model outperformed the other algorithms, followed by the CNN and MLP deep learning models. The results of the experiment highlighted the prediction models that identify likely churners with optimal accuracy and the major factors that influence customer churn.

Zadoo, A. et al. [12] (2022), conducted a review of customer churn in the telecommunications sector, using a combination of supervised learning techniques: for prediction of customer attrition, and unsupervised learning techniques of clustering: for customer segmentation.

Vakeel, A. et al. [13] (2022), proposed a machine learning model that includes boosting algorithms to identify the churn rate and a Gaussian mixture technique to cluster the churned customers. The obtained results of the Gaussian mixture model adopted for customer clustering provided a silhouette score of 0.36.

This study is a comparative analysis of some commonly used classification methods in supervised machine learning to mitigate the risk of losing clients in the telecommunications industry. It aims to evaluate and compare the performance of the proposed models.

3 Materials

3.1 Problem Description

Predicting customer churn is crucial for telecommunications organizations to effectively retain their customers. However, it costs more to attract new customers than to retain existing ones. For this reason, large telecom companies are looking to develop models to forecast which customers are most susceptible to churn and then act accordingly. In this context, we will implement machine learning models to predict the churn probability of a customer based on the analysis of his characteristics, namely his demographic information, his account information, and the information about the services he subscribes to. Next, we will compare these models in terms of different parameters such as accuracy, precision, recall, f1-score, and ROC curve. Then select the best performing model that is likely to reduce churn and, consequently, increase customer satisfaction and company revenues.

3.2 Data Set Description

The dataset used in this article is available on the Kaggle website (the dataset is an adapted version of the original IBM data). It comprises twenty columns of variables that indicate the characteristics of a telecommunications company's customers plus the target variable churn as shown in the figure 1. The Churn column, which represents the target variable, indicates whether the customer has churned in the last month or not. The No class contains non-churned customers, while the Yes class contains churned customers (who have terminated their contracts with the company). The purpose of this analysis is to find the relationship between the customer characteristics and the churn variable.

	customerID	gender	SeniorCitizen	Partner	Dependents	tenure	PhoneService	MultipleLines	InternetService	OnlineSecurity	...	DeviceProtection	TechSupp
0	7590-VHVEG	Female	0	Yes	No	1	No	No phone service	DSL	No	...	No	
1	5575-GNVDE	Male	0	No	No	34	Yes	No	DSL	Yes	...	Yes	
2	3668-QPYBK	Male	0	No	No	2	Yes	No	DSL	Yes	...	No	
3	7795-CFOCW	Male	0	No	No	45	No	No phone service	DSL	Yes	...	Yes	
4	9237-HQITU	Female	0	No	No	2	Yes	No	Fiber optic	No	...	No	
5	9305-CDSKC	Female	0	No	No	8	Yes	Yes	Fiber optic	No	...	Yes	
6	1452-KIOVK	Male	0	No	Yes	22	Yes	Yes	Fiber optic	No	...	No	
7	6713-OKOMC	Female	0	No	No	10	No	No phone service	DSL	Yes	...	No	
8	7892-POOKP	Female	0	Yes	No	28	Yes	Yes	Fiber optic	No	...	Yes	
9	6388-TABGU	Male	0	No	Yes	62	Yes	No	DSL	Yes	...	No	

10 rows × 21 columns

Fig. 1: The churn data set

Table 1: List of features

Features	Type
PhoneService	Customer services
MultipleLines	
InternetServices	
DeviceProtection	
OnlineBackup	
OnlineSecurity	
TechSupport	
StreamingMovies	
StreamingTV	
Tenure	Customer account information
Contract	
PaperlessBilling	
PaymentMethod	
MonthlyCharges	
TotalCharges	
CustomerID	Customer demographic information
Gender	
SeniorCitizen	
Partner	
Dependents	
Churn	Response variable

4 Machine Learning Classification Models Used

There are several machine learning classification algorithms for predicting customer churn in the telecommunications industry. In the following, we briefly present four classification techniques used in this paper for the prediction of customer attrition.

4.1 k-Nearest Neighbor

The k-nearest neighbor (k-NN) is a supervised learning classification algorithm used in various applications including customer churn prediction. It aims at classifying target points or the unknown class, according to their distances from points constituting a training sample whose class is known a priori. The new instance's class is defined according to its neighbors' plurality votes. It is a simple and non-parametric algorithm requiring only the definition of an entry value k that represents the number of neighbors that participate to the classification [14].

4.2 Logistic Regression

Logistic regression (Log. R) is a probabilistic statistical classification algorithm that performs best in binary classification problems. It is employed to interpret the relationship of a binary dependent variable and a set of independent variables [15]. To predict customer churn, the weighted independent variables are fed into the logistic function and output as a probability interval from zero to one. That is, if a customer's churn probability exceeds 0.5, that customer is forecasted as a churner.

4.3 Random Forest

Random Forest (RF) is a leading ensemble classification algorithm based on the decision tree concept, in which ensemble modeling is based on the use of two or more different but related analytical models. To generate new data sets, RF applies the bagging approach and adjusts a decision tree model to each of the new data sets [16]. Concretely, the Random Forest model divides the initial data set in random subgroups of samples, followed by generating a decision tree on each subgroup during the model learning phase. Next, in the prediction phase, each decision tree produces its forecast. The result of the prediction corresponds to the majority prediction value.

4.4 Support Vector Machine

The support vector machine (SVM) is a supervised learning algorithm that performs data analysis to identify patterns. It is used to solve classification, regression, or anomaly detection problems [17]. For customer churn prediction, this model considers every observation as points in a high-dimensional space and identifies the optimal separating hyperplanes among instances of different classes in the training data set. In the prediction step, it represents new within the given space and classifies them into one class depending on their neighborhood to the separation space. The kernel trick is used to move from linear to nonlinear classification. Kernel functions allow non-linear representation of input data in high-dimensional feature spaces in a computationally efficient way.

5 Methodology

The following figure (Fig.2) illustrates the pipeline of this study.

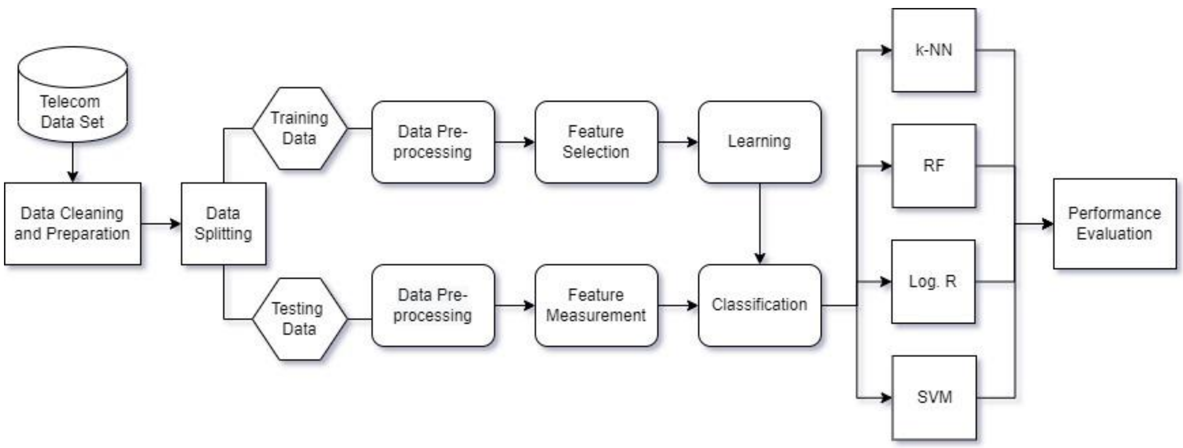

Fig. 2: The study's pipeline

5.1 Data Pre-processing

- Data Cleaning

This step consists of deleting the missing data. The figure 3 shows that the data set contains11 missing values for the variable "TotalCharges". For simplicity, all rows of missing data have been removed.

```
customerID          0
gender              0
SeniorCitizen       0
Partner             0
Dependents          0
tenure              0
PhoneService        0
MultipleLines       0
InternetService     0
OnlineSecurity      0
OnlineBackup        0
DeviceProtection    0
TechSupport         0
StreamingTV         0
StreamingMovies     0
Contract            0
PaperlessBilling    0
PaymentMethod       0
MonthlyCharges      0
TotalCharges       11
Churn               0
dtype: int64
```

Fig. 3: Checking missing data

- Feature Engineering

The following steps, based on the data types, are taken to pre-process the features:
Removing irrelevant variables: the "CustomerID" column has been removed as it is not meaningful.
Encoding of the labels: the following characteristics: "gender", "Partner", "Dependents", "PaperlessBilling", "PhoneService", "Churn", are categorical and have two values each (yes, no). Therefore, they are turned into binary integers.
One-Hot Encoding: the following characteristics "InternetService", "MultipleLines", "OnlineSecurity", "OnlineBackup", "TechSupport", "DeviceProtection", "StreamingMovies", "StreamingTV", "PaymentMethod", "Contract", are categorical, but not ordinal, and contain more than two values. A new variable is generated for each value, with a binary integer denoting whether the value appeared in a data entry (1,0).
Data Normalization: the following characteristics "MonthlyCharges", "TotalCharges", "Tenure", are numeric. They are rescaled between 0 and 1 via the min-max scaler, which aims at scaling the values about an average of 0 and a standard deviation of 1.
Transforming all categorical characteristics available in the dataset into numerical ones.
Creating a new variable "N_services" by summing up the number of services the customer has subscribed to ("OnlineSecurity", "OnlineBackup", "DeviceProtection", "StreamingMovies", "StreamingTv", "TechSupport").

- Data Visualisation

The following figure (figure 4) shows a bar chart of the percentage of observations corresponding to each class of the response variable "Churn": no and yes. As shown in the figure below, the dataset contains 73.4% of non-churners and 26.6% of customers who churned to that service provider.

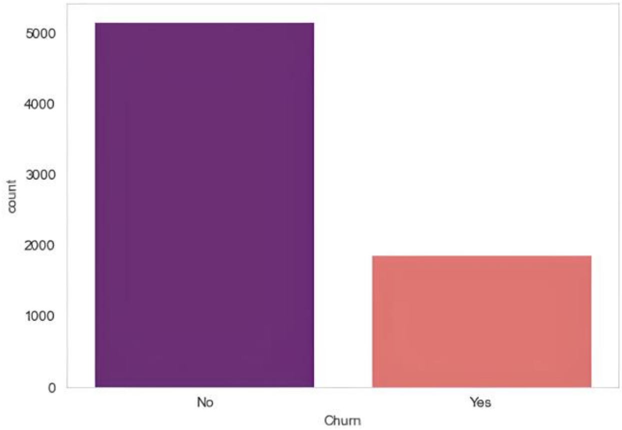

Fig. 4: A bar chart of the percentage of churn and non-churn customers

Based on the plot of the "TotalCharges", "MonthlyCharges", and "tenure" variables against the "Churn" variable as shown in the figure 5. It can be seen that the "TotalCharges" variable, which is the summation of "MonthlyCharges", provides more accurate information about churn on an individual basis. Additionally, churners have higher "MonthlyCharges", with a median of about $80 and a much lower interquartile range than non-churners with a median of about $65. As well, based on the plot of the "tenure" variable, churners have a very short lifetime with a median of about 10 months, while the median for non-churners is about 38 months.

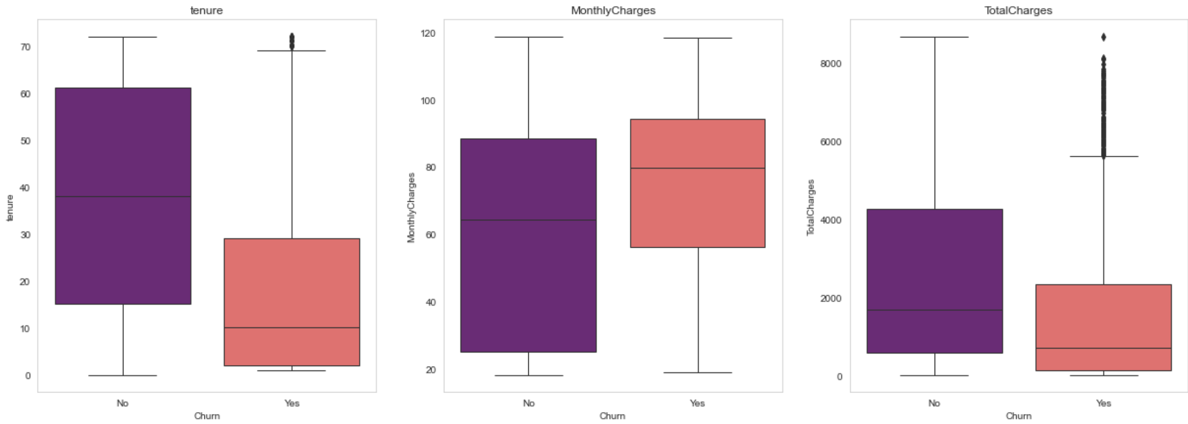

Fig. 5: The plot of the "TotalCharges", "MonthlyCharges", and "tenure" variables against the "Churn"

Based on the plots of the variables "Partner", "Dependents", "SeniorCitizen" and "Contract" as shown in the figure 6. The following observations can be made:
Customers without partners have a higher churn rate.
Independent customers have a higher churn rate compared to dependent customers.
Senior citizens have a much higher churn rate than non-senior citizens.
Customers with month-to-month contracts tend to churn compared to customers with one- and two-year contracts.

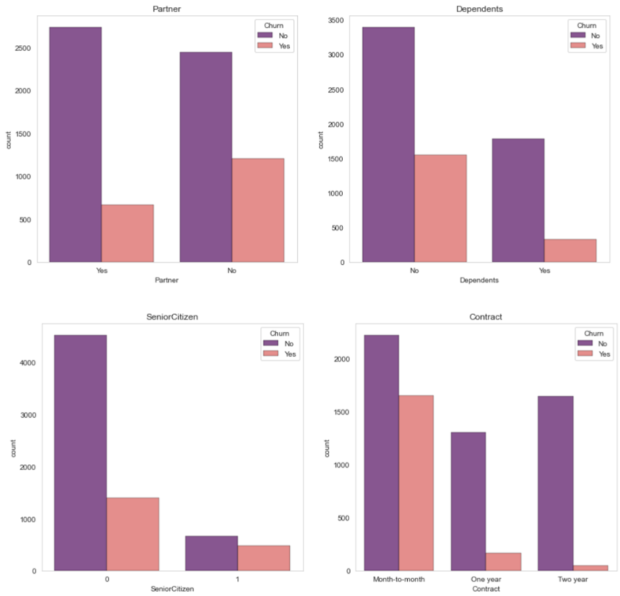

Fig. 6: The plot of the "Partner", "Dependents", "SeniorCitizen", and "Contract" variables against the "Churn"

The figure 7 is the plot of the variable "N_services" that is the number of services to which the customer is subscribed. It can be seen that as the number of services is higher, the churn rate is lower.

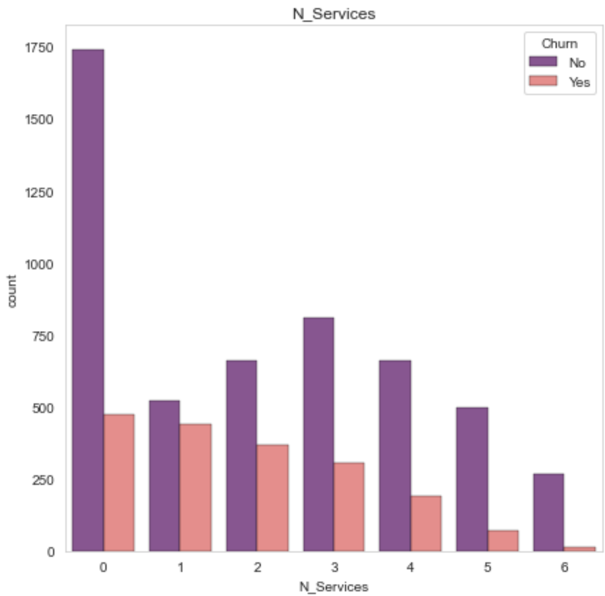

Fig. 7: The plot of the variable number of services against the "Churn"

- Correlation

The figure 8 displays the correlation diagram to analyze the correlation of churn with each of the other characteristics.

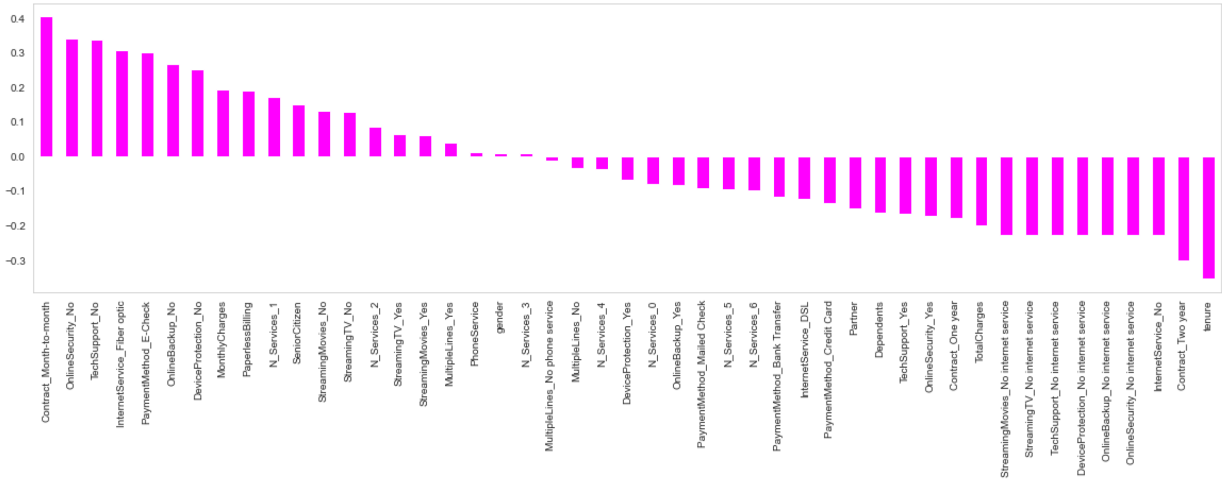

Fig. 8: The correlation diagram

5.2 Training and Testing Sets

After the data preprocessing steps, the churn dataset is then trained and tested. The data is divided into two parts, one for training and one for testing. The split factor parameter was set to 70: 30% for training and testing respectively for better performance. This implies that 4922 of the data set will be the training data and the other 2110 will be the test data. This method avoids reusing the training set for the test and prevents overfitting by applying the classifiers.

5.3 Model Performance Metrics

The purpose of using the performance measures is to explore the models employed and to preselect the optimal one. All candidate models listed above were trained, using the following performance metrics: accuracy, error-rate, precision, recall, and f1-score.

Confusion Matrix

The following indicators: precision, recall, accuracy, error-rate, and F-score are calculated from the captured information using the confusion matrix shown in the table 2. True positives and false positives are denoted TP and FP, and false negatives and true negatives are denoted FN and TN respectively.

The performance metrics are calculated as follows:

$$Precision = TP/(TP+FP) \tag{1}$$

$$Recall = TP/(TP+FN) \tag{2}$$

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

$$Error\text{-}rate = 1\text{-}\ Accuracy \tag{4}$$

$$F1\text{-}score = (2*\ Precision*Recall)/(Precision+Recall) \tag{5}$$

Table 2: Confusion matrix

		Prediction outcome	
		Positive	Negative
Actual value	Positive	TP	FN
	Negative	FP	TN

5.4 Hyperparameter Optimization

The holdout method is a simple method that generally involves training models from the training set and evaluating them from the test set [18]. However, these default predefined hyper-parameters are unlikely to be adequate. For this reason, a method has been developed based on the holdout method, called k-fold cross-validation, which is an alternative method that is used during the setting of hyper-parameters with "Grid Search" and "Randomized Search". It consists in dividing the data set into k subsets, and the holdout method is repeated k times. This method is frequently used when the data are unbalanced, as it ensures that each subset contains nearly the equal percentage of samples from each class (churn and non-churn) as the whole set. The result of the cross-validation is k values for all metrics on the k-fold CV. That is, in the k-NN Model, for a better performance, GridSearchCV is needed to determine the optimal number of neighbors. In the Random Forest model, RandomizedSearchCV serves to optimize the hyperparameters max_features, n_estimators, max_depth, bootstrap, and criterion. In the logistic regression model, GridSearchCV is employed to apply different regularization values to reduce the impact of insignificant features to zero or to simplify the model by relativizing strong patterns that are identified in the training, in order to determine the best model. In the SVM model, GridSearchCV is also applied to determine the optimal margin value around the support vector.

6 Model Evaluation and Comparison

Table 3 shows the parameters of the confusion matrices (TP, TN, FP, FN) of each of the proposed models.

Table 3: Confusion matrices of the proposed models

Classifiers	TP	TN	FP	FN
k-NN	1036	345	126	603
Log R	1294	331	140	345
RF	1122	352	120	516
SVM	1441	320	152	197

From the confusion matrix of the models used, performance metrics: precision, recall, and f1-score are calculated. Tables 4, 5, 6, and 7 shows the performance metrics for each proposed model before hyperparameter tuning.

Table 4: SVM model performance metrics

	precision	recall	f1-score
Churn = No	0.90	0.88	0.89
Churn = Yes	0.62	0.68	0.65
macro avg	0.76	0.78	0.77
accuracy			83.46 %
error-rate			16.54 %

Table 5: k-NN model performance metrics

	precision	recall	f1-score

Churn = No	0.89	0.63	0.74
Churn = Yes	0.36	0.73	0.49
macro avg	0.62	0.68	0.61
accuracy			65.45 %
error-rate			34.55 %

Table 6: Log. R model performance metrics

	precision	recall	f1-score
Churn = No	0.90	0.79	0.84
Churn = Yes	0.49	0.70	0.58
macro avg	0.69	0.74	0.71
accuracy			77.01 %
error-rate			22.99 %

Table 7: RF model performance metrics

	precision	recall	f1-score
Churn = No	0.90	0.68	0.78
Churn = Yes	0.41	0.75	0.53
macro avg	0.65	0.71	0.65
accuracy			69.86 %
error-rate			30.14 %

Table 8 summarizes the performance metrics of all the proposed models before the adjustment of hyperparameters.

Table 8: Performance metrics of the proposed models before the adjustment of hyperparameters

Model	accuracy	error-rate	M. precision	M. recall	M. f1-score
SVM	83.46 %	16.54 %	0.76	0.78	0.77
KNN	65.45 %	34.55 %	0.62	0.68	0.61
Log. R	77.01 %	22.99 %	0.69	0.74	0.71
RF	69.86 %	30.14 %	0.65	0.71	0.65

From the above table we can see that the performance parameters are low before the hyperparameters tuning. Therefore, the hyperparameters were adjusted in order to increase the performance of the models used. Table 9 summarizes the performance metrics of all the proposed models after the adjustment of hyperparameters.

Table 9: Performance metrics of the proposed models after the adjustment of hyperparameters

Model	accuracy	error-rate	M. precision	M. recall	M. f1-score
SVM	96.92 %	3.08 %	0.94	0.97	0.95
KNN	79.05 %	20.95 %	0.72	0.78	0.73
Log. R	91.37 %	8.63 %	0.86	0.91	0.88
RF	88.82 %	11.18 %	0.83	0.90	0.85

Based on the above tables, we can notice a large increase in the performance metrics after the hyperparameters adjustment. Thus, according to the performance metrics, we can see that the SVM model outperformed the other models with an accuracy of 96.92%, a precision of 0.94, a recall of 0.97, an f-score of 0.95 and an error rate of 3.08%.

7 Conclusion

Predicting churn is a major issue for telecommunications companies. Indeed, identifying customers dissatisfied with the services supplied enables companies to work on their pricing plans, offers and promotions, as well as customer preferences, in order to reduce the possibility of customer churn. There are many techniques to predict customer churn, this study focused on comparing different machine learning classification models to find the best performing model for churn prediction based on performance metrics. The results showed that the proposed models obtained a better accuracy by adjusting the hyperparameters of the different models via cross-validation. Thus, the SVM classifier outperformed the other models. This study can be improved by utilizing big data analysis. Indeed, social network analysis is highly recommended to determine the satisfaction rate of customers towards the services of a telecom organization, then this data can be exploited to minimize the churn rate of customers. Also, utilizing other data sets can increase the trust level of the results.

References

[1] Patil, K., Patil, S., Danve, R., & Patil, R. (2022). Machine Learning and Neural Network Models for Customer Churn Prediction in Banking and Telecom Sectors. Proceedings of Second International Conference on Advances in Computer Engineering and Communication Systems. Algorithms for Intelligent Systems. Springer, Singapore. https://doi.org/10.1007/978-981-16-7389-4_23

[2] Matuszelański, K., & Kopczewska, K. (2022). Customer Churn in Retail E-Commerce Business: Spatial and Machine Learning Approach. J. Theor. Appl. Electron. Commer, 165-198. https://doi.org/10.3390/jtaer17010009

[3] Huang, F., Zhu, M., Yuan, K., & Deng, E.O. (2015). Telco churn prediction with big data. ACM SIGMOD international conference on management of data, 607-618.

[4] Kumar, N., & Naik, C. (2017). Comparative Analysis of Machine Learning Algorithms for Their Effectiveness in Churn Prediction in the Telecom Industry. International Research Journal of Engineering and Technology, 4, 485-489.

[5] Zhang, X., Zhang, Z., Liang D., & Jin, H. (2018). A Novel Decision Tree Based on Profit Variance Maximization Criterion for Customer Churn Problem. 2018 10th International Conference on Intelligent Human-Machine Systems and Cybernetics (IHMSC), 20-23, doi: 10.1109/IHMSC.2018.00013.

[6] Bharat, A. (2018). Consumer Engagement Pattern Analysis Leading to Improved Churn Analytics: An Approach for Telecom Industry. Advances in Intelligent Systems and Computing, 203-211. doi:10.1007/978-981-13-1274-8_16

[7] Andrews, R., Zacharias, R., Antony, S., & James, M. M. (2019). Churn Prediction in Telecom Sector Using Machine Learning. International Journal of Information

Systems and Computer Sciences, 8, 132-134. https://doi.org/10.30534/ijiscs/2019/31822019

[8] Calzada-Infante, L., Óskarsdóttir, M., & Baesens, B. (2020). Evaluation of customer behavior with temporal centrality metrics for churn prediction of prepaid contracts. Expert Systems with Applications, 113553. doi:10.1016/j.eswa.2020.113553

[9] Wang, X., Nguyen, K., & Nguyen, B. P. (2020). Churn Prediction using Ensemble Learning. In *Proceedings of the 4th International Conference on Machine Learning and Soft Computing*. doi:10.1145/3380688.3380710

[10] Nam, N., & Duong, T. (2021). Comparison of Two Main Approaches for Handling Imbalanced Data in Churn Prediction Problem. Journal of Advances in Information Technology, 12(1), 29-35. doi: 10.12720/jait.12.1.29-35

[11] Siddika, A., Faruque, A., & Masum, A. K. M. (2021). Comparative Analysis of Churn Predictive Models and Factor Identification in Telecom Industry. In *2021 24th International Conference on Computer and Information Technology*. *ICCIT* (pp. 1-6). doi: 10.1109/ICCIT54785.2021.9689881

[12] Zadoo, A., Jagtap, T., Khule, N., Kedari, A., & Khedkar, S. (2022). A review on Churn Prediction and Customer Segmentation using Machine Learning. In *2022 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing*. *COM-IT-CON* (pp. 174-178). doi: 10.1109/COM-IT-CON54601.2022.9850924

[13] Vakeel, A., Vantari, N. R., Reddy, S. N., Muthyapu, R., & Chavan, A. (2022). Machine Learning Models for Predicting and Clustering Customer Churn Based on Boosting Algorithms and Gaussian Mixture Model. In *2022 International Conference for Advancement in Technology*. *ICONAT* (pp. 1-5). doi: 10.1109/ICONAT53423.2022.9725957

[14] Merabet, M., & Kourtiche, A. (2021). Edited Nearest Neighbors with SVM for Customer Churn Prediction in Spark. In *International Conference on Artificial Intelligence in Renewable Energetic Systems* (pp. 682-687). Springer, Cham.

[15] Lalwani, P., Mishra, M. K., Chadha, J. S., & Sethi, P. (2022). Customer churn prediction system: a machine learning approach. Computing, 104(2), 271-294.

[16] Kiguchi, M., Saeed, W., & Medi, I. (2022). Churn prediction in digital game-based learning using data mining techniques: Logistic regression, decision tree, and random forest. Applied Soft Computing, 118, 108491.

[17] Xiahou, X., & Harada, Y. (2022). B2C E-Commerce Customer Churn Prediction Based on K-Means and SVM. Journal of Theoretical and Applied Electronic Commerce Research, 17(2), 458-475.

[18] Turan, S. C., & Cengiz, M. A. (2022). Ensemble Learning Algorithms. Journal of Science and Arts, 22(2), 459-470.

Notes on contributors

Manal Loukili is an IT engineer and PhD student at the University of Sidi Mohamed Ben Abdellah in Fez, Morocco. She is a member of the laboratory: Artificial Intelligence, Data Science and Emerging Systems (IASSE). Her principal research areas are Machine Learning and E-Marketing.

Fayçal Messaoudi is an accredited professor at the National School of Business and Management in Fez, Morocco. He is a member of the Research Laboratory in Management, Finance and Audit of Organizations, and Artificial Intelligence, Data Science and Emerging Systems laboratory. His main teaching and research interests concern Artificial Intelligence, Data Analysis, Database Management, and E-Marketing.

Mohammed El Ghazi is an accredited professor at the Superior School of Technology in Fez, Morocco. He is a member of the Artificial Intelligence, Data Science and Emerging Systems laboratory. His major teaching and research focus involve Artificial Intelligence and Machine Learning, Networks and Telecommunications, and Computer Science.