Implementation of Machine Learning Algorithms for Customer Churn Prediction

Manal Loukili[1*], Fayçal Messaoudi[2], Raouya El Youbi[1]

[1].National School of Applied Sciences, Sidi Mohamed Ben Abdellah University, Fez, Morocco
[2].National School of Business and Management, Sidi Mohamed Ben Abdellah University, Fez, Morocco

Abstract

Churn prediction is one of the most critical issues in the telecommunications industry. The possibilities of predicting churn have increased considerably due to the remarkable progress made in the field of machine learning and artificial intelligence. In this context, we propose the following process which consists of six stages. The first phase consists of data pre-processing, followed by feature analysis. In the third phase, the selection of features. Then the data was divided into two parts: the training set and the test set. In the prediction process, the most popular predictive models were adopted, namely random forest, k-nearest neighbor, and support vector machine. In addition, we used cross-validation on the training set for hyperparameter tuning and to avoid model overfitting. Then, the results obtained on the test set were evaluated using the confusion matrix and the AUC curve. Finally, we found that the models used gave high accuracy values (over 79%). The highest AUC score, 84%, is achieved by the SVM and bagging classifiers as an ensemble method which surpasses them.

Keywords: Machine Learning; Churn Prediction; Bagging SVM; k-NN; Random Forest.

1- Introduction

The exponential growth in the number of operators in the market, due to globalization and advances in the telecommunications industry, is increasing competition. In this competitive era, it has become imperative to maximize profits regularly, for this reason, various approaches have been proposed, especially acquiring new customers, up-selling existing customers and increasing the retention period of current customers. Retaining existing customers is the simplest and least expensive strategy compared to the others. In order to adopt this strategy, companies need to reduce the eventual churn of customers. The main reason for this loss of customers is dissatisfaction with the services provided to consumers and the support mechanism. To solve this problem, the solution is to predict which customers are likely to churn [1]. Predicting churn is a crucial objective that helps in establishing customer retention and loyalty strategies. Along with the increasing competition in service delivery markets, the risk of customer churn is also increasing explosively. As a result, it has become imperative to implement strategies to keep track of loyal customers (non-churners). Churn models aim to identify early signals of churn and attempt to predict which customers leave voluntarily. Thus, many companies have realized that their existing database is one of their most valuable assets [2] and according to Abbasdimehr [3], churn prediction is a very useful tool to predict at-risk customers.

This article is organized as follows: the next section describes the problematic. Section 3 summarizes some related work. Section 4 represents a brief review of the selected classification techniques used for this study. The different steps of our methodology are discussed in Section 4. The results are presented in section 5. And section 6 to present the results and analyze the performance of each model. Finally, section 7 concludes the article.

2- Problem Description

To overcome the above problem, the company must correctly predict the customer's behavior. Churn

✉ **Manal Loukili**
manal.loukili@usmba.ac.ma

management can be done in two ways: reactive and proactive.

The reactive strategy is to wait for the cancellation request received from the customer and then to offer interesting plans to retain the customer. The proactive strategy, on the other hand, prevents the customer from unsubscribing. This is because the possibility of unsubscribing is anticipated, and plans are offered to customers accordingly. This results in a binary classification problem where churners are distinguished from non-churners.

To deal with this problem, machine learning algorithms have emerged as a very powerful technique for predicting information on the previously captured database [4], including linear regression, support vector machines, naive bayes, decision trees, random forests, etc. In machine learning models, the information is then used to predict the churn.

In machine learning models, after preprocessing, feature selection plays a major role in increasing the accuracy of classification. Researchers have developed a large number of approaches for feature selection that reduce dimensionality, overfitting, and computational complexity. For churn prediction, these features are taken from the given input vector and used for churn prediction. In this paper, to solve this problem, we used the following machine learning algorithms: Support vector machine, K-Nearest Neighbors, and Random Forest.

Support vector machines: This algorithm can be used in cases where there are two classes of objects (e.g., churners and non-churners). It can also be used when there are more than two classes of objects (e.g., churners, potential churners, and non-churners).

K-nearest neighbor: This algorithm is suitable for time series data, categorical data, and sparse datasets.

Random forest: This is a supervised machine learning algorithm which is based on classification and regression trees. It is designed to handle both categorical and continuous data types.

3- Related Work

This section briefly summarizes some related work proposed by leading researchers for the prediction of churn in the telecommunications sector.

The authors in [5] analyzed the variables that impact customer churn. They also conducted a comparative study of three machine learning models such as regression, neural network, and regression trees. The results showed that the decision tree performs better than the others because of its rule-based architecture. However, the accuracy obtained can be further improved by using one of the existing feature selection methods.

The authors in [6] adopted three machine learning approaches, namely support vector machine, neural network, and Bayesian networks for attrition rate prediction. A principal component analysis is taken into account in the feature selection process which allows for a reduction in data dimensions. However, the feature selection process can be improved to increase the accuracy of the classification by applying an optimization algorithm. The gain measure and the ROC curve were used to evaluate the performance.

In another study [7], the authors tried to solve the customer loss prediction problem using a support vector machine, a random forest and logistic regression. The performance of the SVM was approximately equal to that of the logistic regression and the random forest, but once optimal parameter selection was considered, the SVM outperformed the logistic regression and the random forest in terms of PCC and AUC.

Paper [8] presents an overview of all the machine learning models considered, as well as a detailed analysis of the feature selection techniques in use. The authors found that in the prediction models the decision tree had a higher efficiency than the others. In feature selection, optimization techniques also play an essential role that improves the prediction techniques. After a comparative analysis of existing methods.

In [9], the authors adopted two machine learning models, decision tree and logistic regression on a churn prediction dataset. They used the WEKA tool for experimentation. But the customer churn problem can be solved more effectively by using other machine learning methods.

4- A Brief Review of the Machine Learning Classification Algorithms Used

4-1- Bagging Support Vector Machine

Support Vector Machine or SVM is a supervised learning technique that aims to analyze data to detect patterns. There are two types of support vector machines: linear and nonlinear [10]. If the data domain can be divided linearly (e.g., straight line or hyperplane) to separate the classes in the original domain, it is referred to as a linear support vector machine. Nonlinear support vector machine is used when the data domain cannot be split linearly and can be translated to a space called the feature space where the data domain can be divided linearly to distinguish the classes [11]. On the basis of a set of training data, SVM attempts to determine the optimal separating hyperplanes between examples of distinct classes by representing observations as points in a high dimensional space. New instances are represented in the same space and assigned to a class depending on their closeness to the dividing gap [12].

Bagging, also called Bootstrap aggregating, is an ensemble learning approach that helps in the improvement of

performance and accuracy of a machine learning algorithm. It is mainly used to minimize a prediction model's variance and to deal with bias-variance tradeoffs. In [13] Various simulated results for IRIS data categorization and hand-written digit identification demonstrate that the proposed SVM ensembles with bagging significantly outperform a single SVM in terms of classification accuracy. When it comes to the customer churn issue, ensemble-learning techniques have been used as shown in [14], [15], [16].

4-2- K-Nearest Neighbors

When there is little or no prior knowledge about the distribution of the data, K-Nearest Neighbors (k-NN) classification is one of the most fundamental and straightforward classification procedures and should be one of the initial options for classification research [17]. The k-NN classification arose from the necessity to do discriminant analysis when valid parametric estimates of probability densities are unknown or impossible to calculate.

The k-NN method predicts the values of new data points based on "feature similarity", which implies that the new data point will be assigned a value depending on how closely it resembles the points in the training set. k-NN does not attempt to build an internal model, and no calculations are done until classification time. k-NN merely stores instances of the training data in the features space, and an instance's class is chosen by the majority vote of its neighbors. The class most prevalent among its neighbors is assigned to the instance. k-NN finds neighbors based on distance utilizing Euclidian, Manhattan, or Murkowski distance measures for continuous variables and hamming distance measures for categorical data [18].

4-3- Random Forest

Random forests, also known as random choice forests, are an ensemble learning approach for classification, regression, and other tasks that work by building a large number of decision trees during training. It contains multiple decision trees, each reflecting a unique instance of the random forest's classification of data input. The random forest approach examines each case independently, selecting the one with the most votes as the chosen prediction. The classification findings in [18] suggest that Random Forest outperforms Decision Tree (J48) for the same number of characteristics and big data sets, i.e., with a higher number of instances, but Decision Tree (J48) is useful for small data sets (less number of instances). In addition to that the study in [19] shows that, the best classification model out of naïve Bayes, decision tree, and random forest is random forest because of its high

accuracy of 97.5 % when compared to the classification model of decision tree, which has an accuracy of 88.7 %.

5- Methodology

The steps and advantages of the proposed technique are as follows (Fig.1):

The gravitational search algorithm was used to select the features and reduce the dimensions of the data set, unlike the above existing approaches where the prediction accuracy is low due to inadequate feature selection.

After data preprocessing, some of the most important machine learning techniques used for predictions, including SVM, were applied. To avoid overfitting, cross-validation was performed, unlike other techniques where the overfitting prevention mechanism is not considered.

The power of ensemble learning was then utilized to optimize the algorithms and obtain better results, unlike the previously mentioned techniques where the performance of ensemble learning is not taken into account, which explains the low accuracies obtained.

The algorithms were then evaluated on the test set using the confusion matrix and the AUC curve to compare the best performing algorithm for the given data set.
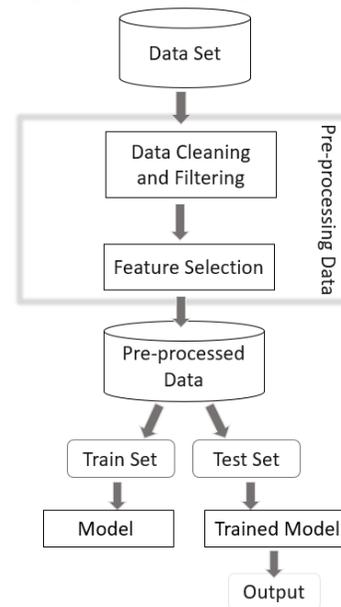

Fig. 1 Proposed system architecture

5-1- Presentation of the Data Set Used

The data set that we used in our experiments is "Telco Customer Churn" which is available on the Kaggle site, which contains a data set of 7043 customers. The data set includes information about:

• Customers who have left the company in the last month - the column is called Churn.

• The services to which each customer has subscribed: telephone, multiple lines, internet, online security, online backup, device protection, technical support and streaming TV and movies.

• Customer account information: how long they have been a customer, type of contract, method of payment, electronic billing, monthly charges, and total charges.

• Demographic information about customers: gender, age range, and whether they have partners and dependents.

The database consists of 21 attributes including a target value called Churn. The data set of the customer attributes alongside with their description is presented in Table. 1, Table. 2 presents the type of these attributes.

Table 1: Margin specifications

Attribute Name	Type	Description
CustomerID	Unique key	A code specific to each customer
Gender	Categorical	The gender of the customer
SeniorCitizen	Categorical	Whether the client is young or old
Partner	Categorical	Whether the client is married or not
Dependents	Categorical	If the client has someone dependent on him
Tenure	Integer	The number of months during which the customer is loyal
PhoneService	Categorical	Whether the customer has a telephone service or not
MultipleLines	Categorical	Whether the customer has a multitude of lines or not
InternetService	Categorical	If the customer has an internet service
OnlineSecurity	Categorical	If the customer has online security
OnlineBackup	Categorical	If the client has an online backup
DeviceProtection	Categorical	If the customer has device security
TechSupport	Categorical	If the customer has technical support
StreamingTV	Categorical	If the customer has on-demand television
StreamingMovie	Categorical	If the customer has the movies on demand
Contract	Categorical	The contract renewal period
PaperlessBilling	Categorical	Whether the customer has paperless or non-paperless billing
PaymentMethod	Categorical	The payment method of the customer
MonthlyCharges	Integer	The monthly charge of the client
TotalCharges	Integer	The total charge of the client
Churn	Categorical	If the customer cancels his contract or not

Table 2: Type of attributes

	Index
	Demographic Information
	Services
	Account information
	Target

5-2- Data Analysis

After importing data from the ".csv" file, the df.info() command was executed to display information about the database.

```
<class 'pandas.core.frame.DataFrame'>
Index: 7043 entries, 7590-VHVEG to 3186-AJIEK
Data columns (total 20 columns):
 #   Column            Non-Null Count   Dtype
---  ------            --------------   -----
 0   gender            7043 non-null    object
 1   SeniorCitizen     7043 non-null    int64
 2   Partner           7043 non-null    object
 3   Dependents        7043 non-null    object
 4   tenure            7043 non-null    int64
 5   PhoneService      7043 non-null    object
 6   MultipleLines     7043 non-null    object
 7   InternetService   7043 non-null    object
 8   OnlineSecurity    7043 non-null    object
 9   OnlineBackup      7043 non-null    object
 10  DeviceProtection  7043 non-null    object
 11  TechSupport       7043 non-null    object
 12  StreamingTV       7043 non-null    object
 13  StreamingMovies   7043 non-null    object
 14  Contract          7043 non-null    object
 15  PaperlessBilling  7043 non-null    object
 16  PaymentMethod     7043 non-null    object
 17  MonthlyCharges    7043 non-null    float64
 18  TotalCharges      7043 non-null    object
 19  Churn             7043 non-null    object
dtypes: float64(1), int64(2), object(17)
memory usage: 1.1+ MB
```

It has been noticed that there is a problem with the types of the attributes "SeniorCitizen" and "TotalCharges" in which "SeniorCitizen" needs to be converted to a string type, and "TotalCharges" needs to be converted to an integer type. The conversion of these attributes to their appropriate types will be carried out as the first step.

```
df["TotalCharges"] =
pd.to_numeric(df["TotalCharges"],
errors='coerce')
df['SeniorCitizen'].replace(to_replace=1,
value='Yes', inplace=True)
df['SeniorCitizen'].replace(to_replace=0,
value='No', inplace=True)
```

After the conversion, it was observed that the "TotalCharges" column has 11 missing values. It is known that the "TotalCharges" variable can be calculated by multiplying the two variables "Tenure" and "MonthlyCharges". However, for all the entries in the "TotalCharges" column, the corresponding "Tenure" value is 0, indicating that these customers are in their first month. Therefore, the value of "MonthlyCharges" will be directly assigned to them as their "TotalCharges" value.

```
df_na =
df[df['TotalCharges'].isna()].copy()
df_na['TotalCharges'] =
df_na['MonthlyCharges']
df[df['TotalCharges'].isna()] = df_na
```

The database has been cleaned and is now prepared for visualization.

5-3- Data Visualization

a) Qualitative Variables

The qualitative variables were visualized using Python, as depicted in the figures below Fig. 2-Fig. 18:
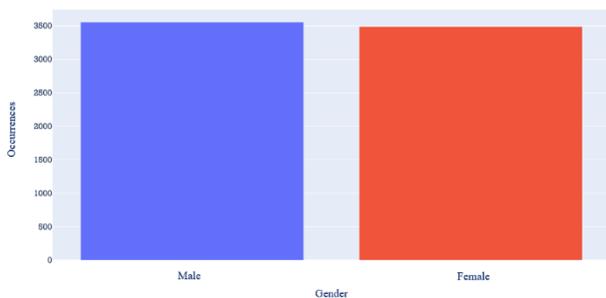
Fig. 2 Male/female distribution
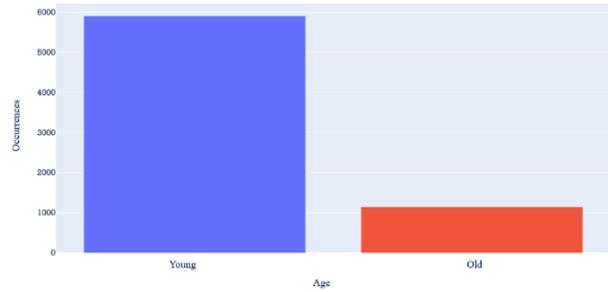
Fig. 3 Young/old people distribution
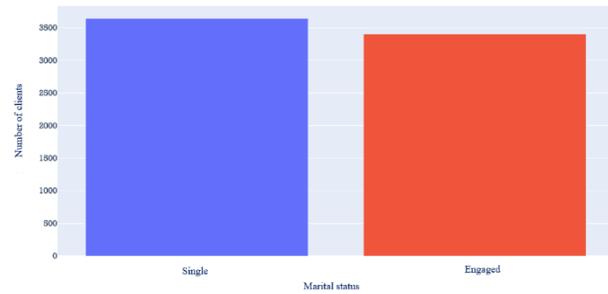
Fig. 4 Single/engaged people distribution
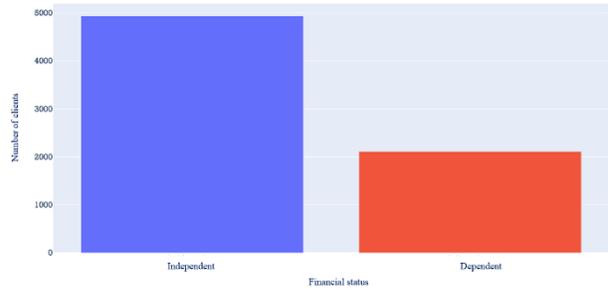
Fig. 5 Independent/ dependent people distribution
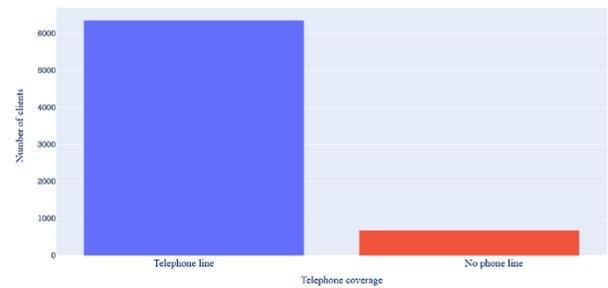
Fig. 6 Distribution of the customers having a telephone line at disposition
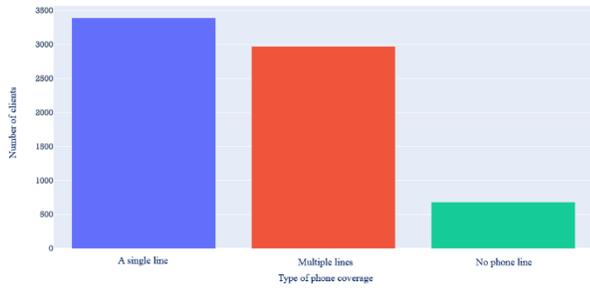
Fig. 7 Distribution of customers with several telephone lines available

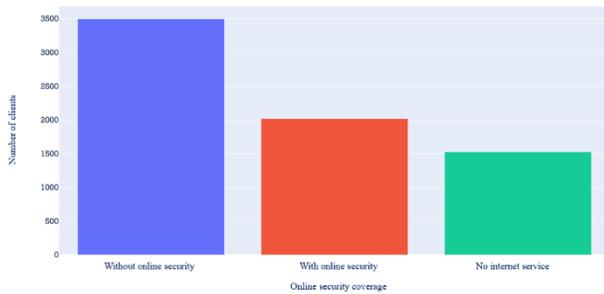Fig. 8 Type of the customer's internet service provider

Fig. 9 Distribution of customers with online security
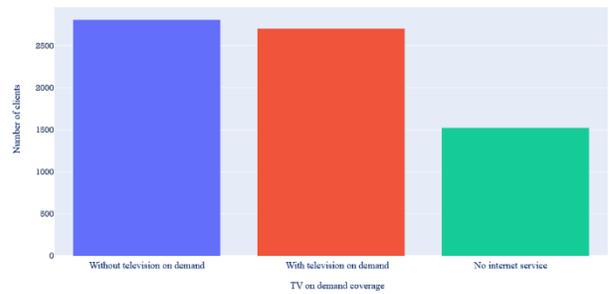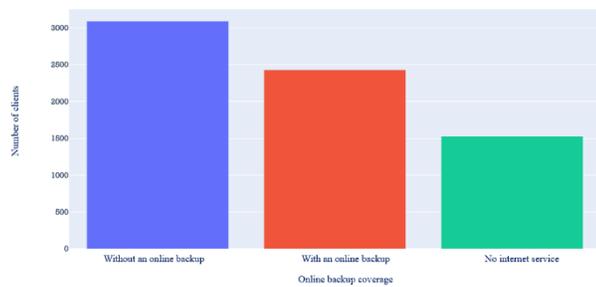

Fig. 10 Distribution of customers with an online backup available

Fig. 11 Distribution of customers with a protective device

Fig. 12 Distribution of customers with technical support

Fig. 13 Distribution of customers with on-demand television available
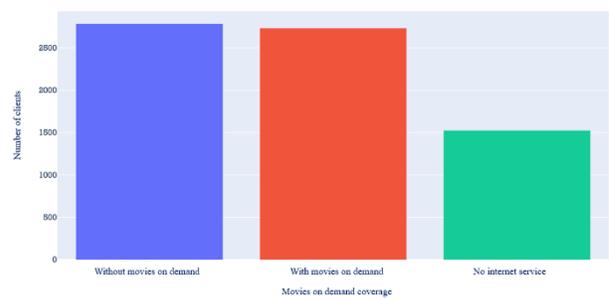
Fig. 14 Distribution of customers with on-demand movies available
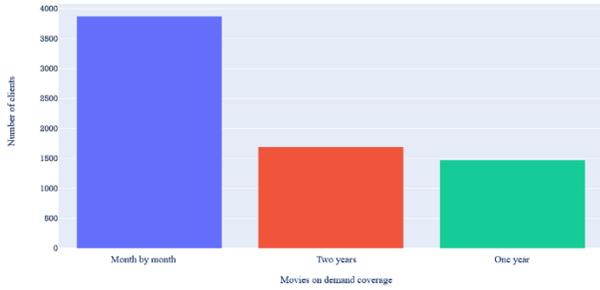
Fig. 15 Distribution of customers according to the duration of their contract

Fig. 16 Distribution of customers according to their billing

Fig. 17 Distribution of customers according to their type of payment

Fig. 18 Distribution of customers according to their loyalty

b)   Quantitative Variables

The quantitative variables were visualized as shown in the figures below Fig. 19-Fig. 21:

Fig. 19 Distribution number of months of loyalty

Fig. 20 Distribution of the amount invoiced to the customer each month

Fig. 21 Distribution of the amount invoiced to the customer in total

5-4- Study of Correlation

The figures that follow represent the correlation between each of the variables and the target variable Fig. 22-Fig. 36:
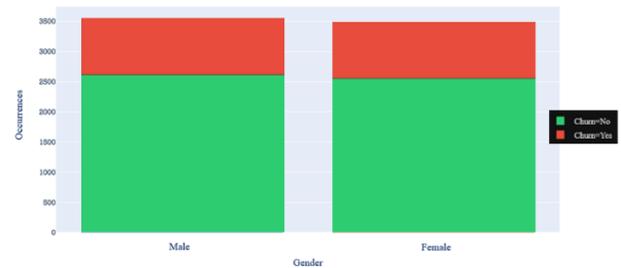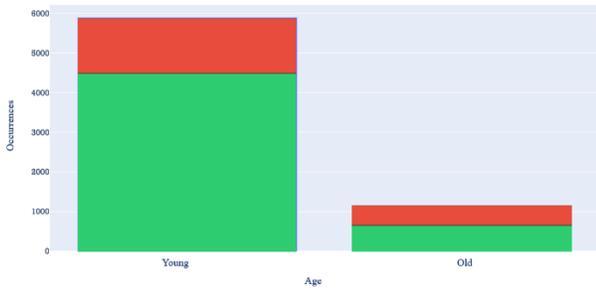

Fig. 22 Gender and churn correlation

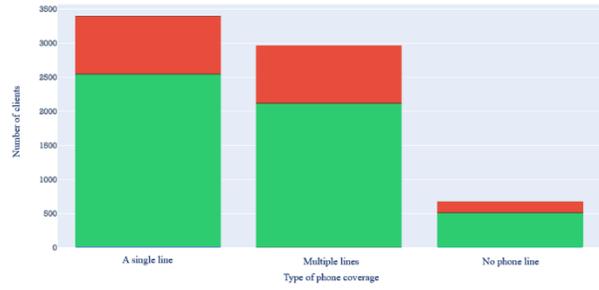Fig. 23 Age and churn correlation

Fig. 27 Type of phone coverage and churn correlation
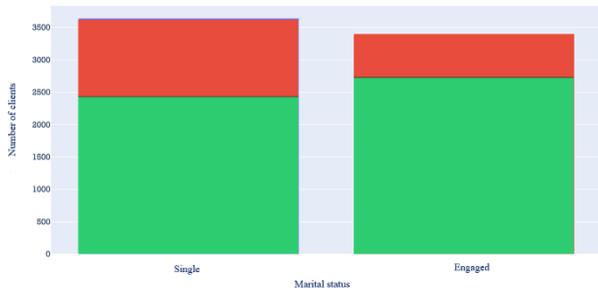

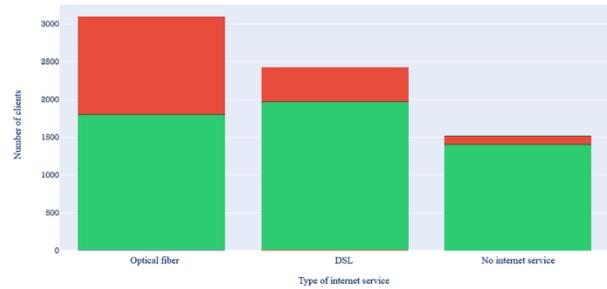Fig. 24 Marital status and churn correlation

Fig. 28 Type of internet service and churn correlation
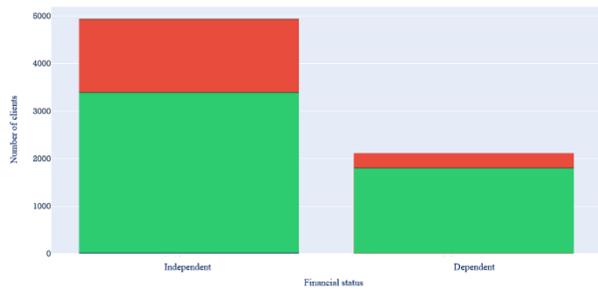

Fig. 25 Financial status and churn correlation
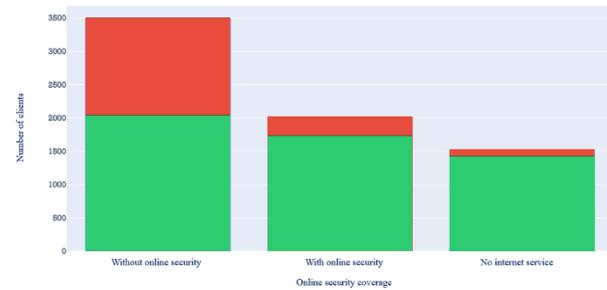

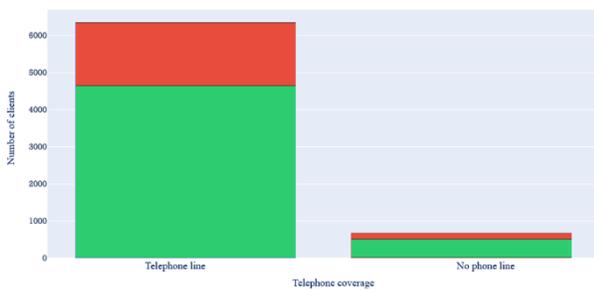Fig. 29 Online security coverage and churn correlation

Fig. 26 Telephone coverage and churn correlation
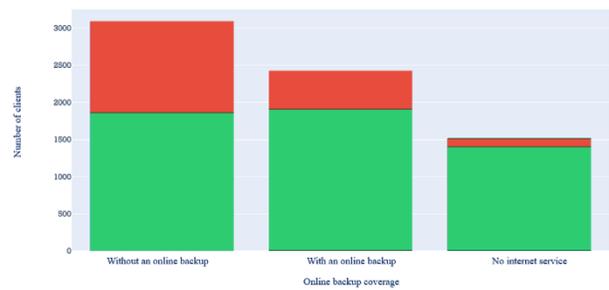
Fig. 30 Online backup coverage and churn correlation
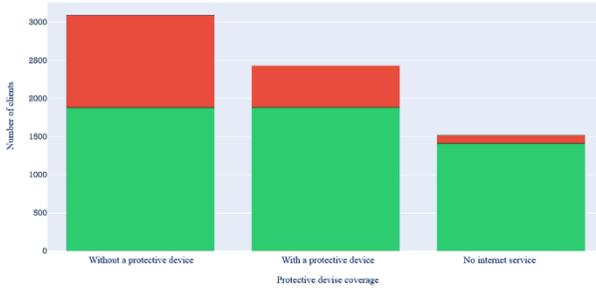
Fig. 31 Protective devise coverage and churn correlation
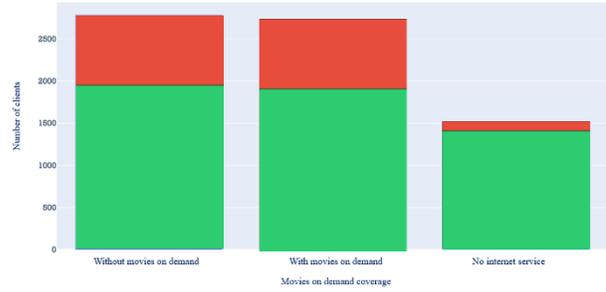

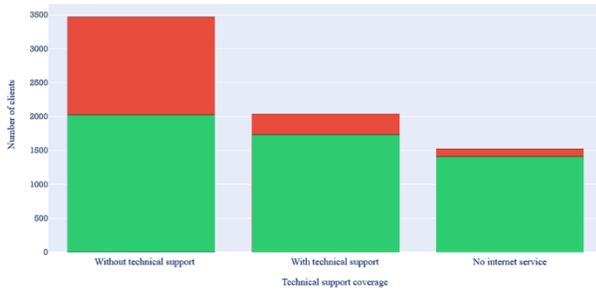Fig. 34 Movies on demand coverage and churn correlation

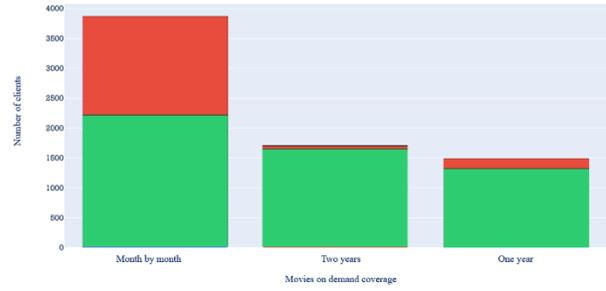Fig. 32 Technical support coverage and churn correlation

Fig. 35 Movies on demand's duration of contract and churn correlation
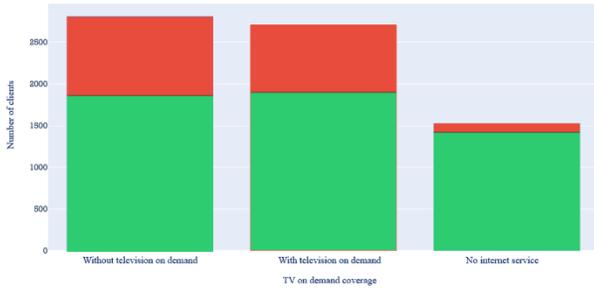

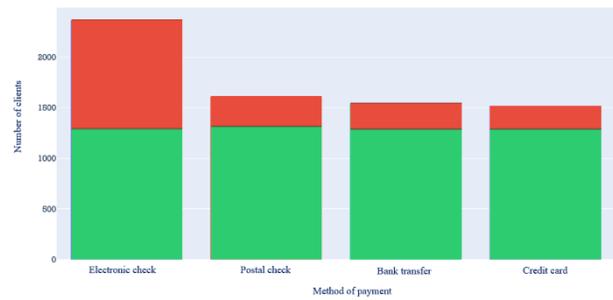Fig. 33 TV on demand coverage and churn correlation

Fig. 36 Method of payment and churn correlation

The Fig. 37 shows the correlation between the churn variable and the other variables:
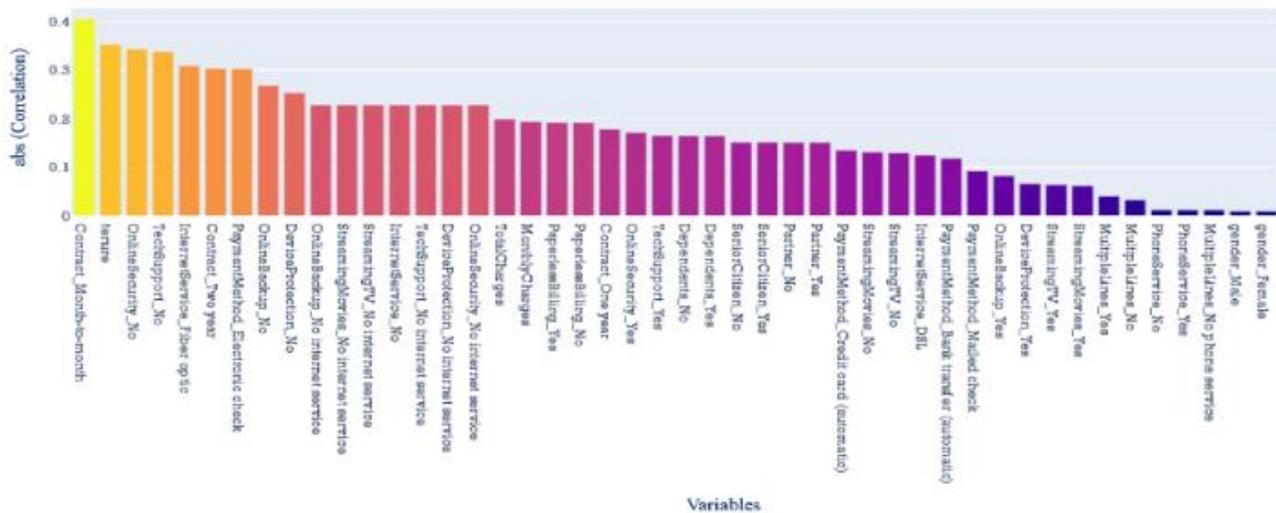
Fig. 37 The correlation between churn and other variables

Based on the correlation diagram and the results obtained from the visualization of the independence of each variable with the target variable, it has been observed that there are variables with weak correlation. However, these variables will be retained as the objective is not to generalize the model, but rather to specifically fit it to this particular company. This approach may result in the model being overfitted to match the company's specific characteristics.

5-5- The Learning Phase

The learning in this study is based on three machine learning algorithms: Bagging SVM, Random Forest and the k-NN algorithm.
The first step in the learning is to convert all qualitative variables into quantitative variables.

```python
df['Churn'].replace(to_replace='Yes',
 value=1, inplace=True)
df['Churn'].replace(to_replace='No',
value=0, inplace=True)
df_dum = pd.get_dummies(df)
```
.
After the execution, the "df_dum" DataFrame now consisted solely of quantitative variables.
Subsequently, the various methods and algorithms available in the "Scikit-Learn" library were implemented using Python language.

```python
from sklearn.preprocessing import
StandardScaler
from sklearn.pipeline import
make_pipeline
from sklearn.svm import SVC
from sklearn.ensemble import
BaggingClassifier,RandomForestClassifie
r
from sklearn.neighbors import
KNeighborsClassifier
from sklearn.model_selection import
train_test_split
```

5-6- Algorithms Implementation

Bagging SVM

```python
model = None
score = 0
for i in range(10):
  X_train, X_test, y_train, y_test = train_te
st_split(X, y, test_size=0.30)
  mdl = make_pipeline(StandardScaler(),
BaggingClassifier(base_estimator=SVC(gamma="a
uto"), n_estimators=10))
  mdl.fit(X_train, y_train)
  a = mdl.score(X_test, y_test)
  if a > score:     model = mdl
                score = a
```

K-Nearest Neighbors

```python
model = None
score = 0
for i in range(10):
  X_train, X_test, y_train, y_test = train_tes
t_split(X, y, test_size=0.30)
  mdl = KNeighborsClassifier(n_neighbors=10)
  mdl.fit(X_train, y_train)
  a = mdl.score(X_test, y_test)
  if a > score : model = mdl
                score = a
```

Random Forest

```python
model = None
score = 0
for i in range(10):
  X_train, X_test, y_train, y_test = train_t
est_split(X, y, test_size=0.30)
  mdl = RandomForestClassifier(max_depth=5)
  mdl.fit(X_train, y_train)
  a = mdl.score(X_test, y_test)
  if a > score :     model = mdl
                score = a
```

6- Results and Performance Analysis

6-1- Confusion Matrix

In order to evaluate the performance of the applied models and the prediction rate of customer churn on the test set, various metrics were utilized, including precision, recall, accuracy, and F-score..They measure the ability of the predictive models to correctly predict the customer churn. The four indicators previously mentioned are calculated from the information captured using the confusion matrix and are presented in Table 3. The representation of the confusion matrix is shown in Fig. 38. True positives and false positives are denoted Tp and Fp, while false negatives and true negatives are denoted Fn and Tn [20].

- True positive (TP): The number of clients who are in the churn class and whom the predictive model correctly predicted.

- True negative (TN): The number of clients who are in the non-churners class and that the predictive model correctly predicted.

- False positive (FP): The number of customers who are not churners but that the predictive algorithm identified as churners.

- False Negative (FN): The number of customers who are churners but that the predictive model identified as non-churners.
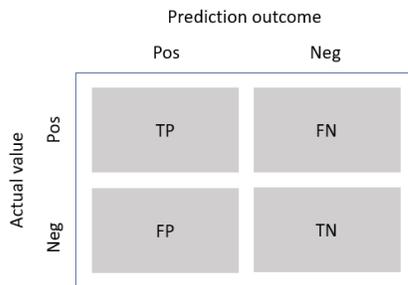

Fig. 38 Confusion matrix

6-2- Performance Indicators

• Recall:
The recall is the ratio of true churners or true positives (TP), and is calculated as follows:
Recall = TP/(TP+FN)

• Precision:
The precision is the ratio of predicted correct churners, its formula is as follows:
Precision = TP/(TP+FP)

• Accuracy:
The accuracy is the ration of the number of all correct predictions and is written as:
Accuracy = (TP+TN)/(TP+FP+TN+FN)

• F-score:
The F-score is the harmonic mean of precision and recall and is written as follows:
F-score = (2* Precision*Recall)/(Precision+Recall)

A value closer to 1 implies a better combination of precision and recall achieved by the classifier.

6-3- AUC Curve

To evaluate the performance of the models on the positive and negative classes of the test set, we employed the AUC curve. A high value of the AUC score indicates that the model performs better on the positive and negative classes. The AUC scores achieved for the three predictive models used to predict the target variable Churn are shown in Table 3 and Fig. 39, graphically represents the AUC scores obtained for Bagging SVM, k-NN and Random Forest. According to the AUC scores, all the selected models perform well on the test set. However, the most adequate classifier is Bagging SVM with an AUC score of 84%.

Table 3: Comparison of the model used

Model	Accuracy %	Recall %	Precision %	F-score %	AUC Score %
Bagging SVM	80.26	80.63	79.71	80.16	84
KNN	79.03	76.30	76.30	76.92	80
Random Forest	79.47	79.44	78.91	79.14	82
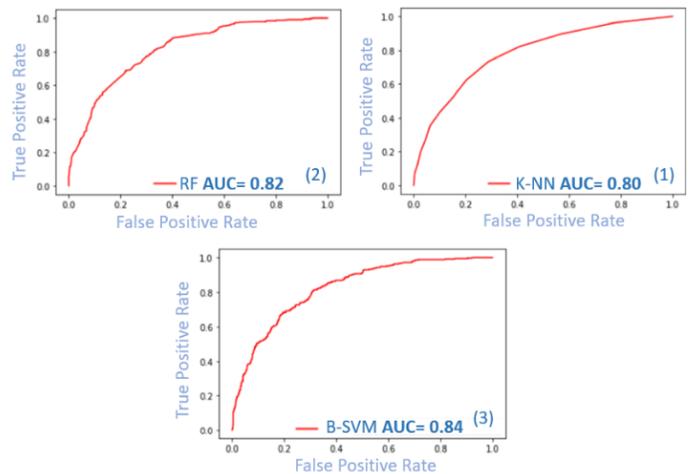
Fig. 39 Models AUC curve. (1): K-Nearest Neighbors. (2): Random Forest. (3): Bagging SVM.

Accuracy of predictive models

Recall of predictive models

Precision of predictive models
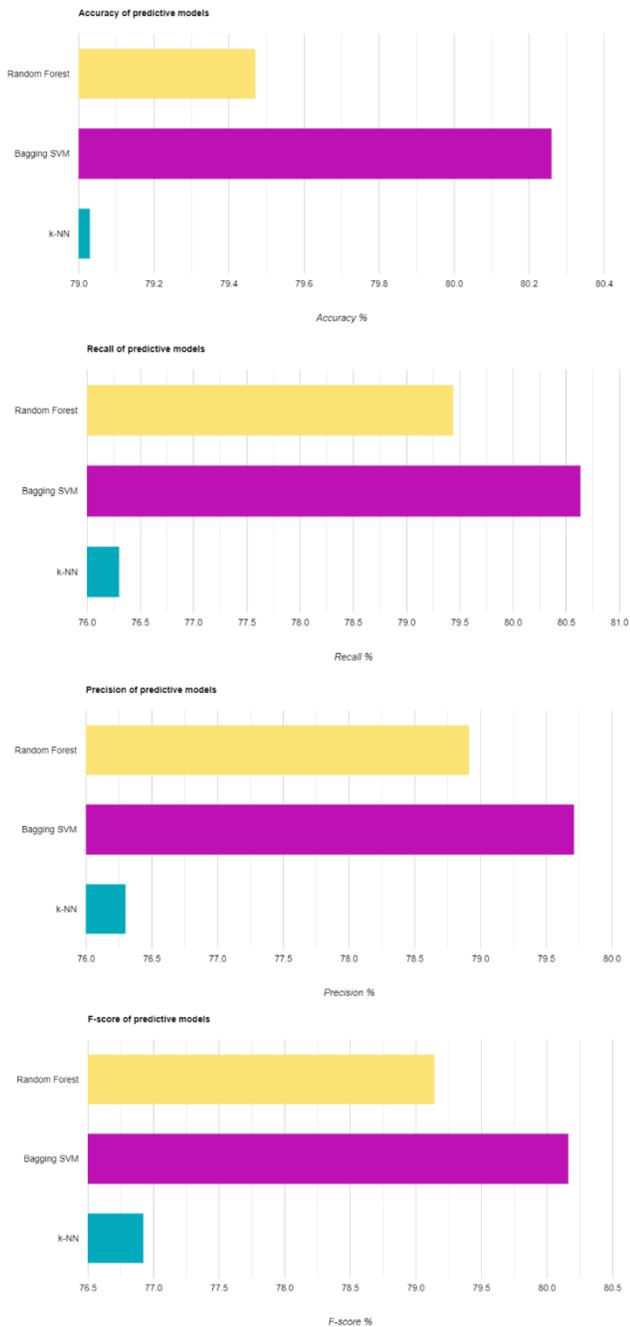

F-score of predictive models

Fig. 40 Predictive models performance indicators: Accuracy, Recall, Precision, F-score

7- Conclusion

In the telecommunications sector, churn prediction is an issue that has attracted the interest of various researchers in recent years. It is becoming one of the sources of revenue for companies and helps to prevent customers from terminating their contracts, it opens the possibility of renegotiation in order to retain the customer by implementing retention strategies.

In this paper, we provide a comparative study of churn prediction in the telecom sector using well-known machine learning techniques such as random forest, k-nearest neighbor, and support vector machine. The experimental results show that all three machine learning techniques give high accuracy for churn prediction. The SVM and bagging classifiers as an ensemble method outperformed the other algorithms in terms of all performance measures such as accuracy, precision, F-measure, recall and AUC score. Predicting churn for a business is proving to be a very tedious task so there is stiff competition in the market to retain customers by providing services that are beneficial to both parties. In futurity, with the upcoming concepts and frameworks in the field of reinforcement learning and deep learning, machine learning proves to be one of the most used and efficient ways to overcome the problems like churn prediction with better accuracy and precision.

References

[1] M. Loukili, F. Messaoudi, and M. El Ghazi, "Supervised Learning Algorithms for Predicting Customer Churn with Hyperparameter Optimization", International Journal of Advances in Soft Computing & Its Applications, Vol. 14, No. 3, 2022, pp. 49-63. doi: 10.15849/IJASCA.221128.04

[2] K. Matuszelański, and K. Kopczewska, "Customer Churn in Retail E-Commerce Business: Spatial and Machine Learning Approach". J. Theor. Appl. Electron. Commer. Res. 2022, 17, pp. 165-198. https://doi.org/10.3390/jtaer17010009.

[3] H. Abbasimehr, M Setak, and M Tarokh, "A neuro-fuzzy classifier for customer churn prediction", International Journal of Computer Applications, Vol. 19, No. 8, 2011, pp. 35-41.

[4] A. K. Ahmad, A. Jafar, and K. Aljoumaa, "Customer churn prediction in telecom using machine learning in big data platform". Journal of Big Data, Vol. 6, No. 1, 2019, pp. 28

[5] J. Hadden, A. Tiwari, R. Roy, and D. Ruta, "Churn prediction : Does technology matter", International Journal of Intelligent Technology, Vol. 1, No. 2, 2006, pp. 104-110

[6] I. Brându₃soiu, G. Toderean, and H. Beleiu, "Methods for churn prediction in the pre-paid mobile telecommunications industry", in 2016 International conference on communications (COMM), IEEE, 2016, pp. 97-100.

[7] K. Coussement, and D. Van den Poel, "Churn prediction in subscription services: An application of support vector machines while comparing two parameter-selection techniques", Expert systems with applications, Vol. 34, No. 1, pp. 313-327

[8] J. Hadden, A. Tiwari, R. Roy, and D. Ruta, "Computer assisted customer churn management: State-of-the-art and future trends", Computers & Operations Research Vol. 34, No. 10, 2007, pp. 02-29.

[9] K. Dahiya, and S. Bhatia, "Customer churn analysis in telecom industry", in 2015 4th International Conference on Reliability, Infocom Technologies and Optimization (ICRITO), Trends and Future Directions, 2015, pp. 1-6.

[10] L. Bottou, "Large-scale machine learning with stochastic gradient descent", in Proceedings of COMPSTAT'2010, 2010, Physica-Verlag HD, pp. 177-186.

[11] S. Suthaharan, "Support Vector Machine in Machine learning Models and Algorithms for Big Data Classification", Integrated Series in Information Systems, Springer, New York, Vol. 36, 2016, pp. 207-235.

[12] S. F. Sabbeh, "Machine-learning techniques for customer retention: A comparative study", International Journal of Advanced Computer Science and Applications, Vol. 9, No. 2, 2018.

[13] H. C. Kim, S. Pang, H. M. Je, D. Kim, and S. Y. Bang, "Support vector machine ensemble with bagging", Berlin, Heidelberg, Springer, 2002, pp. 397-408.

[14] H. Abbasimehr, M. Setak, and M. J. Tarokh, "A Comparative Assessment of the Performance of Ensemble Learning in Customer Churn Prediction", Int. Arab J. Inf. Technol, Vol. 11, No. 6, 2014, pp. 599-606.

[15] S. Tavassoli, and H. Koosha, "Hybrid Ensemble Learning Approaches to Customer Churn Prediction", Kybernetes, 2021.

[16] A. Mishra, and U. S. Reddy, "A comparative study of customer churn prediction in telecom industry using ensemble-based classifiers", in 2017 International Conference on Inventive Computing and Informatics (ICICI), 2017, IEEE, pp. 721-725.

[17] N. Ali, D. Neagu, and P. Trundle, "Evaluation of k-nearest neighbour classifier performance for heterogeneous data sets", SN Applied Sciences, Vol. 1, 2019, pp. 1-15.

[18] J. Ali, R. Khan, N. Ahmad, and I. Maqsood, "Random Forests and Decision Trees, International", Journal of Computer Science Issues, Vol. 9, No. 5, 2012, pp. 272-275.

[19] A. Alamsyah, and N. Salma, "A Comparative Study of Employee Churn Prediction Model", in 2018 4th International Conference on Science and Technology, IEEE, 2018, pp. 1-4.

[20] M. Loukili, F. Messaoudi, and M. El Ghazi, "Sentiment Analysis of Product Reviews for E-Commerce Recommendation based on Machine Learning", International Journal of Advances in Soft Computing & Its Applications, Vol. 15, No. 1, 2023, pp. 1-13.