

## Learning trends in customer churn with rule-based and kernel methods

Nahier Aldhafferi<sup>1</sup>, Abdullah Alqahtani<sup>1</sup>, Fatema Sabeen Shaikh<sup>1</sup>, Sunday Olusanya Olatunji<sup>2</sup>,  
Abdullah Almurayh<sup>3</sup>, Fahad A. Alghamdi<sup>4</sup>, Ghalib H. Alshammri<sup>5</sup>, Amani K. Samha<sup>6</sup>,  
Mutaseem Khalil Alsmadi<sup>4</sup>, Hayat Alfagham<sup>4</sup>, Abderrazak Ben Salah<sup>4</sup>

<sup>1</sup>Department of Computer Information Systems, College of Computer Science and Information Technology, Imam Abdulrahman Bin Faisal University, Dammam, Saudi Arabia

<sup>2</sup>Department of Computer Science, College of Computer Science and Information Technology, Imam Abdulrahman Bin Faisal University, Dammam, Saudi Arabia

<sup>3</sup>Department of Education Technologies, College of Education, Imam Abdulrahman Bin Faisal University, Dammam, Saudi Arabia

<sup>4</sup>Department of Management Information Systems, College of Applied Studies and Community Service, Imam Abdulrahman Bin Faisal University, Dammam, Saudi Arabia

<sup>5</sup>Department of Computer Science, Community College, King Saud University, Riyadh, Saudi Arabia

<sup>6</sup>Department of Management Information Systems, College of Business Administration, King Saud University, Riyadh, Saudi Arabia

### Article Info

#### Article history:

Received Oct 28, 2020

Revised Apr 14, 2022

Accepted May 11, 2022

#### Keywords:

Churn model

Data science

Decision tree

Intelligent system

Rough set theory

Rule-based method

Support vector machine

### ABSTRACT

In the present article an attempt has been made to predict the occurrences of customers leaving or ‘churning’ a business enterprise and explain the possible causes for the customer churning. Three different algorithms are used to predict churn, viz. decision tree, support vector machine and rough set theory. While two are rule-based learning methods which lead to more interpretable results that might help the marketing division to retain or hasten cross-sell of customers, one of them is a kernel-based classification that separates the customers on a feature hyperplane. The nature of predictions and rules obtained from them are able to provide a choice between a more focused or more extensive program the company may wish to implement as part of its customer retention program.

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



### Corresponding Author:

Nahier Aldhafferi

Department of Computer Information System, College of Computer Science and Information Technology,  
Imam Abdulrahman Bin Faisal University

P.O. Box 1982, Dammam 31433, Saudi Arabia

Email: naldhafferi@iau.edu.sa

## 1. INTRODUCTION

Customer relationship management is acknowledged as a popular and effective strategy for creating, controlling, and reinforcing strong long-term relationship between the business and the customer. The subject has been actively researched and extensively applied to different sectors such as finance, retail businesses or telecommunication. One of the main goals of effectively managing the relationship with the customer is customer retention. It is a well-established fact that the costs associated with securing a new customer is disproportionately more than that of holding back pre-existing customers [1], [2]. It is imperative to explore and research up-to-date and robust methods to develop suitable models that could help avoid customer churn. In a rapidly growing and transforming market environment, the act of a customer’s changing services has been generally observed to be associated with concomitant indicators like one’s transaction history, customer service calls, or interactions on the social media, among others [3].

Satisfaction of customers, customer care complains, launch of new products, or upcoming security or privacy regulations also affect the movement of customers from one seller to another. The objective of a machine learning model to combat churn is identifying early signals of customer's behavior that could probably lead one to snap a continuing business relation. Customer churn or 'attrition' seriously affects the business of a company, which is why companies strive to precisely identify those customers who are more prone to churn during the next month or quarter. There are numerous business models that aim to speculate the lifetime value of a customer from the corresponding lifetime period. These models were developed based on some criteria such as classifying circular sustainable or specific product lifetime extension (PLE) business models, or tapping inevitably into some PLE forms [4]. However, the factor that is filled with the highest level of uncertainty in these calculations is the probability of a customer to churn within a certain period. Depending through an accurate prediction of the churn probability of customers during a time window, companies are likely to gain foot by adopting the right strategy for customer retention programs by offering promotions or discounts, through upsell of stock items, or via cross-sell of most-likely-buying items in a single basket.

This article presents the comparative study of three algorithms (decision tree, support vector machine and rough set theory) in the prediction of churn events of customers in a telecom dataset and tries to find the rules that can generalize the churning behavior of customers. Results have been compared with one another, and the advantages of using either of the algorithms in prediction or understanding customer behavior to decide future marketing strategies are discussed. The study observed that decision tree (DT) and rough set theory (RST) record higher prediction accuracy, while support vector machine (SVM) perform better with respect to area under the receiver operating characteristic (ROC) curve.

## 2. RELATED WORK

Several research outcomes have been reported in literature on managing customer relationships and the churning pattern of customers. For example, Ascarza *et al.* [5] reviews the theoretical aspects of the problem while Santharam and Krishnan [6] reviewed the prediction techniques applied by researchers in the area of customer churn. Vafeiadis *et al.* [1] described a study on some popular machine learning (ML) methods as an application to predict customer churn. The comparative study is based on the dataset from the telecommunication industry. A series of Monte Carlo simulations for each model and a wide range of parameters are used which demonstrate the supremacy of the boosted versions against other versions. SVM-POLY using AdaBoost is one of the best classifiers resulting in almost 97% accuracy and over 84% F-measure. Comparative studies among five mostly used classification methods on customer churn indicate that their boosting versions have more predictive accuracy.

The paper authored by Yu *et al.* [7] worked with an algorithm known as extended support vector machine (ESVM) to forecast customer churn prediction. They investigated customers of e-commerce, one of the most unstable service sectors. The authors have proposed that ESVM is the best fit to work with imbalanced and non-linear nature of churn behavior. According to the paper working with four popularly used algorithms, ESVM shows superior results from different accuracy measures such as the rate of hits, coverage, or lift.

To retain old customers than to gain new ones is becoming the new vision of today's business world. From this context, this Xie *et al.* [8] tried to solve one of the challenging problems of churn prediction-imbalanced data distribution. They worked on the churn dataset of Chinese bank customers, applying an algorithm improved balanced random forests (IRBF) by combining balanced random forests and weighted random forests through assigning a larger weight to the minority class. To get better prediction, the paper proposed the addition of integrating methods for sampling and learning processes dependent on cost with standard random forests (SRF). They also reported that IRBF produces more accuracy than other algorithms like artificial neural network, decision trees, despite its superior results in other fields [9]–[21].

Neslin *et al.* [22] explained how methodology affects the accuracy of predictive model for customer churn. The paper analyzed a churn modeling tournament between researchers and practitioners. Both of them have to prepare a model for predicting churning customers on two validation databases. Different types of calculating approaches confirm that method is an important factor and staying power of the said model is another bold line which can be at least 3 months from calibration to compilation of a database. The authors have observed a notable difference in the approaches, indicating there is more to prediction than just the estimation techniques.

Burez and Poel [23] illustrated in their paper how to reduce the effect of class imbalance to get more real and accurate churn predictions. They have worked with both type of sampling-random and advanced sampling, using area under the curve (AUC) and lift evaluation metrics. To find out better results they compared two specific algorithms such as gradient boosting machine (GBM) and weighted random forests (WRF) (i.e., the gradient boosted model and random forests implementing differential weights) along with

some other well-known algorithms. The authors concluded that there is no effective improvement using advanced sampling but advised for WRF algorithm in order to get better results.

The paper of Ahn *et al.* [24] studied about determinants of churn. They worked on the dataset from telecommunication service of Korean market. In order to identify the partial or total defection, the authors have recognized a term customer's status. Uses of the service (frequent or medium or no-use) determine the status of customers. Study proposed that the relation between churn determinants and possibility of churn calculated through customers' status change.

Chen *et al.* [4] tried to explain customer churn prediction from another angle of reducing noise in data range. Databases of business houses are fully occupied with invariant customer data and only specific behavior-related information. In this paper the authors approach towards direct use of behavior related information, instead of transforming it into an invariant form to get more accuracy. For this innovation they developed the novel variant of the SVM with several kernels arranged in different hierarchies, for a three phases application. The paper suggests that it gets better prediction of churn following this new method.

### 3. MATERIALS AND METHODS

In the present sections a description of the dataset used in the study is provided, and the methods adopted in building the customer churn model that would help to: i) identify certain customers with maximum likelihood of churning in a particular month, ii) choose certain features that are more useful in predicting the churn behavior, iii) map the relationship between the relevant features and churn probability, and iv) draft suitable strategies to combat impending churn and retain valuable customers. The machine learning models are introduced in this section. These models will be applied to the dataset, and ensuing results obtained from the analysis are presented in the following section.

#### 3.1. The dataset

The data has been downloaded from Kaggle website BlastChar [25], and is one of the most popular datasets on customer churn. It contains particulars about 7043 customers, and depicts 20 different features of a customer, which fall under the following broad categories: demographic information about customers, customers' account-related information, Services they have signed up for, and the class variable *Churn*. Other than these, there is a *CustomerID*, which identifies each customer with a unique code. We ignore this in our analysis. The exhaustive list of features is given in Table 1, along with a short description and the categoric values or variable types.

#### 3.2. Data cleaning and preprocessing

On exploring the dataset, a few rows with missing values were found in the *TotalCharges* column. One-level deeper introspection revealed that all these customers had a *tenure* of 0 months, which means their total charges has not been calculated yet and should be 0. Consequently, all these values were coerced to '0'. Next a check was performed to ensure that all 'No' values in the *PhoneService* feature corresponded to 'No phone service' in the *MultipleLines* feature. Similarly, it was also checked that all *InternetService*='No' implied *OnlineSecurity*, *OnlineBackup*, *DeviceProtection*, *TechSupport*, *StreamingTV*, and *streamingMovies* features had a value of 'No internet service' (see features 7, 8 and 9 – 15 in Table 1). A training subset was constituted by taking 80% of the customers randomly from the original dataset, while the other 20% served as the test set. The efficacy of the models built from the training set is evaluated by predicting churn-status of customers in the test dataset. Methodologies outlined in the following subsections are used to build the predictive and introspective models from the data.

#### 3.3. Decision tree

Decision tree (DT) is a rather pragmatic approach of extracting rules from a dataset Safavian and Landgrebe [26]. The rules in fact generalize examples or observations (here 'customers') with known class values (occurrence or non-occurrence of 'Churn'). The most popular and simple procedure for constructing a decision tree is achieved by partitioning the set of observations along certain conditions or 'check's.

For example, if the set of all 7,043 customers in our dataset are split along the condition *SeniorCitizen* =1, we obtain two branches, with 1,142 customers who are senior citizens, and 5,901 customers who are not. This can continue recursively, with each branch being further sub-branched along a different condition, till a certain 'stopping criteria' is reached. An ideal stopping criterion could be when all customers in each terminal branch of the tree belongs to the same class. In this process, every node (starting from the root node) becomes a decision node with a specific condition, and every leaf can be identified with a specific class label (viz. Churn or Not-Churn). This greedy approach for constructing a tree maps a given set of examples to a particular tree, which ultimately classify each customer as having (or not having) churned.

Table 1. Variables in the dataset

No.	Variable	Description (Values/types)
1.	Customer ID	ID of customers (alpha-numeric)
2.	gender	Sex (Male, Female)
3.	Senior Citizen	Age group of customers (1, 0)
4.	Partner	If subscriber lives with partner (Yes, No)
5.	Dependents	If he/she has family dependents (Yes, No)
6.	tenure	Count of months he/she is a customer (integer)
7.	Phone Service	Whether the customer uses a phone (Yes, No)
8.	Multiple Lines	If many lines are used (Yes, No, No i phone i service)
9.	Internet Service	Type of Internet connection (DSL, Fiber i optic, No)
10.	Online Security	Has online security (Yes, No, No i internet i service)
11.	Online Backup	Has online backup (Yes, No, No i internet i service)
12.	Device Protection	Is the device protected (Yes, No, No i internet i service)
13.	Tech Support	Technical support used (Yes, No, No i internet i service)
14.	Streaming TV	Subscribed for TV (Yes, No, No i internet i service)
15.	Streaming Movies	Subscribed for movies (Yes, No, No i internet i service)
16.	Contract	Period of contract (Monthly, Yearly, Biyearly)
17.	Paperless Billing	Whether bills avoid paper print (Yes, No)
18.	Payment Method	Mode of paying (Electronic i check, mailed i check, Bank i transfer i automatic, Credit i card i automatic)
19.	Monthly Charges	Amount billed every month (floating point)
20.	Total Charges	Total amount billed (floating point)
21.	Churn	Has the customer left the company (Yes, No)

The method is termed as ‘greedy’ from the consideration that at every step the objective is to maximize separation between the classes. In other words, among all possible conditions available for branching, the one that achieves maximum class separation is chosen. This is done with the help of a suitable metric such as Gini Impurity or Entropy that purportedly measures the impurity inherent in a node, or lack of ‘class separation’ as (1) and (2):

$$\text{Gini Impurity} : I_G(p) = \sum_{i=1}^c p_i(1 - p_i) \quad (1)$$

$$\text{Entropy} : H(p) = - \sum_{i=1}^c p_i \log_2 p_i \quad (2)$$

where  $p$  is the probability of finding an observation of class  $i$  at any particular node or leaf, and  $c$  is the cardinality of the set of classes in the dataset. In the present scenario,  $c=2$  with just two classes (Churn and Not Churn).

### 3.4. Support vector machine

Support vector machine (SVM) is a mathematically oriented procedure to divide the space of objects belonging to different classes with suitable lines, planes, or hyperplanes. The procedure was introduced by Vapnik and Chervonenkis [27] as a method of statistical learning in the 1960’s. Finding the hyperplane  $w$  essentially involves solving the optimization problem.

$$\min_{w,b} \langle w \cdot w \rangle \quad \text{s. t.} \quad y^i (\langle w \cdot x^i \rangle + b) \geq 1 \quad (3)$$

Where  $x^i$  and  $y^i$  are the input and target values of the  $i$ th object, with the index of objects spanning the range  $i = 1, 2, \dots, N$ .

Implementation of the SVM classifier was achieved with the help of the SVC module of the scikitlearn Python library Nelli [28]. A polynomial kernel was used with a degree varying from 1 to 5. The results are presented in the results and discussion section.

### 3.5. Rough set theory

Rough set theory (RST) is a recent paradigm that is gaining popularity in data science and machine learning. It was proposed by Pawlak [29]. An object or observation is considered to belong to some concept or class, depending on incomplete information about the object. The information system  $\mathcal{J}$  is expressed as a crisscross of objects and input + target features (or attributes).

$$\mathcal{J} = \{U, A \cup \{d\}\} \quad (4)$$

Where  $U$  is the set of all observations,  $A$  is the aggregate of input features, and  $\{d\}$  is the target variable. Rules may be inferred from a given set of data using features from the *reduct*  $R \subset A$  which contains sufficient information to derive the class of objects with the same accuracy as  $A$ . The procedure of obtaining

rules is preceded by discretizing the continuous features into suitable intervals, and subsequently selecting a bare minimum set of features that classifies the dataset with the same accuracy as the entire set of attributes. Several algorithms exist in the R *RoughSets* library Riza *et al.* [30] to discretize features and select them into a reduct. The global discernibility heuristic and the quick reduct algorithms Shen and Chouchoulas [31] were used in the present analysis respectively for discretization and feature selection. Both are supervised methods, categorizing and selecting features keeping in view the final task of classification. For extracting rules, the *AQRules* method Michalski *et al.* [32] was used with the confidence threshold set to 0.8 and times covered as 1, to get only the most accurate rules with no oversampling.

### 3.6. Accuracy metrics

Certain metrics are customarily used to evaluate the performance of a prediction or classification model, particularly following a specific algorithm. They are discussed in the following paragraphs.

- Confusion matrix: It gives the overall picture of the correct and wrong predictions of each class. For a binary class problem, a confusion matrix shows the number of true negatives (TN), false negatives (FN), true positives (TP) and false positives (FP). The typical form of a confusion matrix for the case where there are just two classes, 0 and 1, is shown in Table 2.

Table 2. Confusion Matrix for two classes

Actual Class	Predicted Class	
	0	1
0	TN	FP
1	FN	TP

- Accuracy: It is the simplest metric for assessing the correctness of predictions across all classes. For a binary class problem, it takes the expression:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FN} + \text{FP} + \text{TN}} \quad (5)$$

- Specificity: Specificity is the ability of a classifier to reject false prediction of the actual negatives and predict all negative results correctly. It is the ratio of negative observations predicted correctly to the total number of negative observations in the dataset.

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}} \quad (6)$$

- Sensitivity: It measures how well a model senses positive values or the degree of response of an algorithm. It calculates the proportion of true positives to the total number of actually positive observations.

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (7)$$

- Area under the ROC curve (AUC-ROC): It is the area under the receiver operating characteristic (ROC) curve, which is generated by the graph of sensitivity, also called true positive rate (TPR) against the false positive rate (FPR), taken at different thresholds. TPR and FPR may be expressed as (8) and (9).

$$\text{TruePositiveRate} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (8)$$

$$\text{FalsePositiveRate} = \frac{\text{FP}}{\text{FP} + \text{TN}} \quad (9)$$

The above metrics are used in the following section to integrate the results obtained from the models developed, and measure their abilities in predicting churned customers correctly. The set of churned customers (Churn = Yes) is taken as the positive class.

- Accuracy and support of a rule: They are two important parameters that determine the significance of an if-then rule of the form:

$$\text{If } \phi \text{ then } \psi \quad (10)$$

The two quantities are defined as (11) and (12):

$$\text{Accuracy of rule} = \frac{||\phi|| \cap ||\psi||}{||\phi||} \tag{11}$$

$$\text{Support of rule} = ||\phi|| \cap ||\psi|| \tag{12}$$

where  $||\phi||$  is the set of objects obeying the condition  $\phi$ , and  $||\psi||$  is the set of objects that fall into the class  $\psi$ , and  $|\cdot|$  gives the number of elements a set has. Accuracy of any rule denotes the confidence with which the rule classifies a set of objects into a specific class, while coverage is the number of objects correctly predicted by it. These two qualifiers are provided at the end of each rule presented in the next section.

#### 4. RESULTS AND DISCUSSION

We first explore to visualize the percentage of subscribers who have churned in the entire dataset with the help of a pie-chart as shown in Figure 1. Subsequently the distribution of *tenure* and *Monthly Charges* are explored with the help of density plots as shown in Figure 2. The three different machine learning algorithms (DT, SVM and RST) elucidated in materials and methods section were trained on 80% of the dataset, and the trained models are used to predict the churn status on the remaining 20% of the customers.

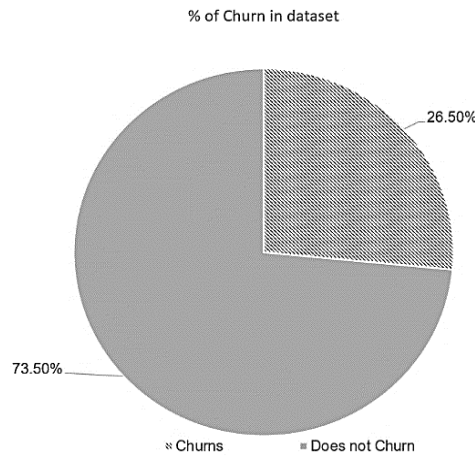


Figure 1. Percentage of customers who have [not] churned

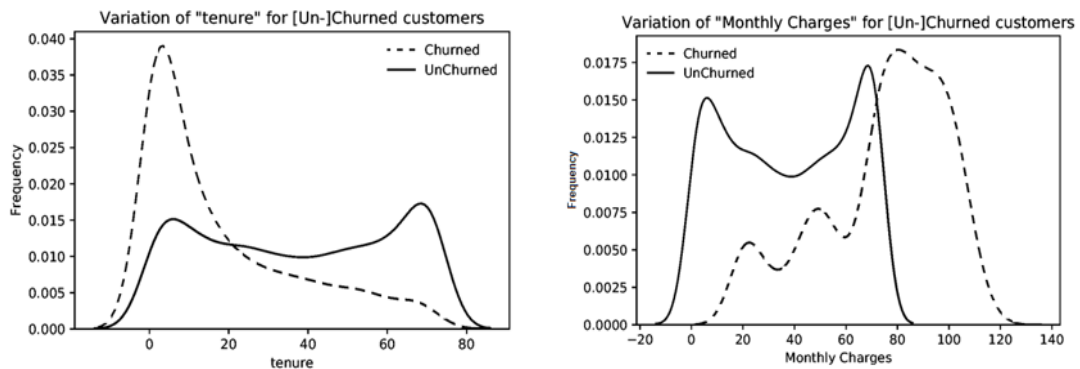


Figure 2. Distribution of [un] churned customers along the feature's tenure and monthly charges

The parameter tuning parameters of the three algorithms are explored across a wide range to get the optimal performance as: i) DT: depth of tree = 3 (tried 2, 3, ..., 9); ii) SVM: degree of polynomial = 2 (tried 1, 2, ..., 5); and iii) RST: confidence threshold = 0.8 (tried 0.7, 0.75, 0.8, 0.85, 0.9, 0.95).

The confusion matrix that provides the entire picture of the predicted classes vis-à-vis the actual classes of the customers both in the training and test sets are presented in the first two rows of Table 3. The confusion matrix that provides the entire picture of the predicted classes vis-à-vis the actual classes of the customers both in the training and test sets are presented in the first two rows of Table 3. In the next four rows the accuracy, sensitivity, specificity, and area under the ROC curve obtained from the prediction on the three algorithms on the training and test sets are also presented. The best accuracy is obtained from the decision tree and rough set theory, with nearly 80% accuracy on both the training and validation subsets.

Table 3. Confusion matrices and prediction accuracies from different machine learning algorithms

Actual Class	Training Set						Test Set					
	DT		SVM		RST		DT		SVM		RST	
Prediction:	0	1	0	1	0	1	0	1	0	1	0	1
0	3884	255	2760	1379	4096	51	971	64	690	345	993	34
1	930	565	431	1064	1081	406	222	152	88	286	287	95
Accuracy	79.0%		67.9%		79.9%		79.7%		69.3%		77.2%	
Sensitivity	93.8%		66.7%		98.8%		93.8%		66.7%		96.7%	
Specificity	37.8%		71.2%		27.3%		40.6%		76.5%		24.9%	
AUC-ROC	65.8%		68.9%		63.0%		67.2%		71.6%		60.8%	

However, in respect of AUC-ROC, support vector machine shows a slightly enhanced performance compared to the other two approaches by around 3 to 4%. The specificity and sensitivity of SVM is also more stabler than DT or RST which attain more than 90% sensitivities at the cost of specificities less than 40%. This tendency is the most marked in RST where specificity falls below 30% to get a 97% accuracy.

This makes SVM predictions suitable when a more extensive retention program is in place, as it is able to identify 286 out of 374 churning customers in the test set, at the same time wrongly bringing in 345 loyal customers under this umbrella. On the other hand, DT or RST prediction can be more useful when the company would not afford to waste its resources for a retention program on its loyal customers rather wishes to take a more focused churn-intensive retention program, not only because fewer (around 100 to 150) churns are identified here, but also since these two algorithms bring in much fewer misclassified (35 to 65) customers under this program.

No signs of overfitting are apparent in the results, as the test set metrics are very close to the training set metrics. However, in the different parametric tuning, it was observed that a decision tree of depth more than 7 or a kernel polynomial of degree more than 4 seems to set in overfitting to some extent. Finally, the most significant rules obtained from DT and RST for the Churn class is presented below. The figures in square brackets at the end of each rule denote the rule accuracy and coverage explained in (12). The rules generated through the Decision Tree model (of depths 3, 4 and 5) are as follows:

*If Contract = Monthly and Internet Service = Fiber optic and tenure ≤ 15.5 then Churn [69%, 565].*

*If Contract = Monthly and Internet Service = Fiber optic and tenure ≤ 15.5 and TotalCharges ≤ 120 then Churn [86%, 161].*

*If Contract = Monthly and Internet Service = Fiber optic and tenure ≤ 15.5 and TotalCharges > 120 and MonthlyCharges > 97.50 then Churn [83%, 69].*

*If Contract = Monthly and Internet Service ≠ Fiber optic and tenure ≤ 4.5 and MonthlyCharges ≤ 55.20 then Churn [57%, 156].*

The rules obtained from Rough Set analysis are as:

*If gender = Female and TotalCharges ∈ (309, 1180] and MonthlyCharges ∈ (90.2, 100) then Churn [88%, 38].*

*If tenure ≤ 3 and StreamingTV = Yes and SeniorCitizen = 1 then Churn [88%, 28].*

*If StreamingTV = Yes and TechSupport = No and tenure ∈ (3, 6.5) and StreamingMovies = Yes then Churn [84%, 31].*

*If PaymentMethod = Electronic check and OnlineBackup = No and MonthlyCharges ∈ (90.2, 100) and tenure ≤ 3 then Churn [83%, 29].*

*If InternetService = Fiber optic and tenure ∈ (3, 6.5) and gender = Female and StreamingMovies = Yes then Churn [88%, 28].*

If gender = Female and MonthlyCharges  $\in$  (73.6, 90.2) and TotalCharges  $\leq$  91.3 and StreamingMovies = No then Churn [90%, 28].  
 If tenure  $\in$  (3, 6.5) and MultipleLines = Yes and gender = Male and OnlineSecurity = No then Churn [81%, 25].

The rules from DT obtained are broader in scope, involve few features, but less accurate. On the other hand, the rules from RST are more specific, generally more accurate, however supported by fewer objects. Though this can be tuned by the tuning parameters of the two algorithms, the contrast and complimentary is intentionally maintained with specific business strategies in mind. This can be viewed as a typical advantage of pluralistic analysis with several different algorithms set to appropriate parameters, where results of different nature are obtained, and the company executive gets the choice of selecting the right kind of rules that would best fit its present purpose. For example, if the management decides to reach as many customers as possible to reduce churn (with an extensive program) rules from DT might appear more appropriate as each of the four rules presented above would help reach almost a thousand customers, but 25% of its resources on customer retention might go wasted, since the accuracy of the rules are low. However, if on some other occasion the management is obliged to reach only a few customers (maybe as a pilot study to see how they respond to discounts and offers or similar retention programs) the marketing executive is free to device their benefits targeting only a few chosen sections of the churning customers, but where the probability of churn is much higher, and only 10 to 15% of the resources may be wasted.

## 5. CONCLUSION

The problem of customers leaving a particular service and migrating to its competitors involve a challenging aspect of customer relationship management. Churn in customers can be prevented if a proper strategy is implemented to detect early chances of churn, followed by suitable retention programs. The paper compares the performance of three different algorithms on their ability to predict customer churn from a publicly available dataset. The respective parameters of each algorithm are properly tuned to achieve optimal prediction. It is observed that DT and RST record higher prediction accuracy, while SVM perform better with respect to area under the ROC curve. The specificity and sensitivity of SVM are more balanced, whereas DT and RST are able to score better sensitivities at the cost of lower specificity. Rules obtained from DT and RST are enlisted with accompanying accuracy and coverage values. The different nature of rules revealed by the two algorithms gives the management a chance to choose between extensive and intensive programs on different sections of its customers at appropriate implementation levels of customer retention strategies it might be interested in adopting.

## REFERENCES




- [1] T. Vafeiadis, K. I. Diamantaras, G. Sarigiannidis, and K. C. Chatzivasvas, "A comparison of machine learning techniques for customer churn prediction," *Simulation Modelling Practice and Theory*, vol. 55, pp. 1–9, Jun. 2015, doi: 10.1016/j.simpat.2015.03.003.
- [2] P. Asthana, "A comparison of machine learning techniques for customer churn prediction," *International Journal of Pure and Applied Mathematics*, vol. 119, pp. 1149–1169, 2018.
- [3] T. Zhang, S. Moro, and R. F. Ramos, "A data-driven approach to improve customer churn prediction based on telecom customer segmentation," *Future Internet*, vol. 14, no. 3, Mar. 2022, doi: 10.3390/fi14030094.
- [4] Z.-Y. Chen, Z.-P. Fan, and M. Sun, "A hierarchical multiple kernel support vector machine for customer churn prediction using longitudinal behavioral data," *European Journal of Operational Research*, vol. 223, no. 2, pp. 461–472, Dec. 2012, doi: 10.1016/j.ejor.2012.06.040.
- [5] E. Ascarza *et al.*, "In pursuit of enhanced customer retention management: Review, key issues, and future directions," *Customer Needs and Solutions*, vol. 5, no. 1–2, pp. 65–81, Mar. 2018, doi: 10.1007/s40547-017-0080-0.
- [6] A. Santharam and S. B. Krishnan, "Survey on customer churn prediction techniques," *International Research Journal of Engineering and Technology*, vol. 5, no. 11, pp. 131–137, 2018.
- [7] X. Yu, S. Guo, J. Guo, and X. Huang, "An extended support vector machine forecasting framework for customer churn in e-commerce," *Expert Systems with Applications*, vol. 38, no. 3, pp. 1425–1430, Mar. 2011, doi: 10.1016/j.eswa.2010.07.049.
- [8] Y. Xie, X. Li, E. W. T. Ngai, and W. Ying, "Customer churn prediction using improved balanced random forests," *Expert Systems with Applications*, vol. 36, no. 3, pp. 5445–5449, Apr. 2009, doi: 10.1016/j.eswa.2008.06.121.
- [9] M. Alzaqebah, N. Alrefai, E. A. E. Ahmed, S. Jawameh, and M. K. Alsmadi, "Neighborhood search methods with moth optimization algorithm as a wrapper method for feature selection problems," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 10, no. 4, pp. 3672–3684, Aug. 2020, doi: 10.11591/ijece.v10i4.pp3672-3684.
- [10] M. K. Alsmadi, M. Tayfour, R. A. Alkhasawneh, U. Badawi, I. Almarashdeh, and F. Haddad, "Robust features extraction for general fish classification," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 9, no. 6, pp. 5192–5204, Dec. 2019, doi: 10.11591/ijece.v9i6.pp5192-5204.
- [11] M. K. Al Smadi, "Hybrid genetic algorithm with tabu search with back-propagation algorithm for fish classification: determining the appropriate feature set," *International Journal of Applied Engineering Research*, vol. 14, no. 3, pp. 4387–4396, 2019.
- [12] M. K. Alsmadi, "An efficient similarity measure for content based image retrieval using memetic algorithm," *Egyptian Journal of Basic and Applied Sciences*, vol. 4, no. 2, pp. 112–122, Jun. 2017, doi: 10.1016/j.ejbas.2017.02.004.






- [13] S. Khodabandehlou and M. Z. Rahman, "Comparison of supervised machine learning techniques for customer churn prediction based on analysis of customer behavior," *Journal of Systems and Information Technology*, vol. 19, no. 1/2, pp. 65–93, Mar. 2017, doi: 10.1108/JSIT-10-2016-0061.
- [14] M. K. Al Smadi, "Facial expression recognition." ed: Google Patents, 2018.
- [15] M. K. Al Smadi, K. B. Omar, and S. A. Noah, "Proposed method to decide the appropriate feature set for fish classification tasks using artificial neural network and decision tree," *IJCSNS International Journal of Computer Science and Network Security*, vol. 9, no. 3, pp. 297–301, 2009.
- [16] B. A. Aldeeb *et al.*, "Hybrid intelligent water drops algorithm for examination timetabling problem," *Journal of King Saud University-Computer and Information Sciences*, Jun. 2021, doi: 10.1016/j.jksuci.2021.06.016.
- [17] M. Alzaqebah *et al.*, "Memory based cuckoo search algorithm for feature selection of gene expression dataset," *Informatics in Medicine Unlocked*, vol. 24, 2021, doi: 10.1016/j.imu.2021.100572.
- [18] M. Alzaqebah *et al.*, "Hybrid feature selection method based on particle swarm optimization and adaptive local search method," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 11, no. 3, pp. 2414–2422, Jun. 2021, doi: 10.11591/ijece.v11i3.pp2414-2422.
- [19] M. K. Alsmadi and I. Almarashdeh, "A survey on fish classification techniques," *Journal of King Saud University - Computer and Information Sciences*, vol. 34, no. 5, pp. 1625–1638, May 2022, doi: 10.1016/j.jksuci.2020.07.005.
- [20] M. K. Alsmadi, "Content-based image retrieval using color, shape and texture descriptors and features," *Arabian Journal for Science and Engineering*, vol. 45, no. 4, pp. 3317–3330, Apr. 2020, doi: 10.1007/s13369-020-04384-y.
- [21] Y. K. Qawqzeh *et al.*, "Applying the big bang-big crunch metaheuristic to large-sized operational problems," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 10, no. 3, pp. 2484–2502, Jun. 2020, doi: 10.11591/ijece.v10i3.pp2484-2502.
- [22] S. A. Neslin, S. Gupta, W. Kamakura, J. Lu, and C. H. Mason, "Defection detection: Measuring and understanding the predictive accuracy of customer churn models," *J. of Marketing Research*, vol. 43, no. 2, pp. 204–211, 2006, doi: 10.1509/jmkr.43.2.204.
- [23] J. Burez and D. Van den Poel, "Handling class imbalance in customer churn prediction," *Expert Systems with Applications*, vol. 36, no. 3, pp. 4626–4636, Apr. 2009, doi: 10.1016/j.eswa.2008.05.027.
- [24] J.-H. Ahn, S.-P. Han, and Y.-S. Lee, "Customer churn analysis: Churn determinants and mediation effects of partial defection in the Korean mobile telecommunications service industry," *Telecommunications Policy*, vol. 30, no. 10–11, pp. 552–568, Nov. 2006, doi: 10.1016/j.telpol.2006.09.006.
- [25] BlastChar, "Telco customer churn: Focused customer retention programs," Kaggle, 2017, Accessed: Nov. 18, 2019. [Online]. Available: <https://www.kaggle.com/datasets/blastchar/telco-customer-churn>.
- [26] S. R. Safavian and D. Landgrebe, "A survey of decision tree classifier methodology," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 21, no. 3, pp. 660–674, 1991, doi: 10.1109/21.97458.
- [27] V. Vapnik and A. Y. Chervonenkis, "A class of algorithms for pattern recognition learning," *Avtomat. i Telemekh.*, vol. 25, pp. 937–945, 1964.
- [28] F. Nelli, *Python data analytics*. Berkeley, CA: Apress, 2018.
- [29] Z. Pawlak, "Rough sets," *International Journal of Computer & Information Sciences*, vol. 11, no. 5, pp. 341–356, Oct. 1982, doi: 10.1007/BF01001956.
- [30] L. Riza and A. Janusz, "RoughSets: data analysis using rough set and fuzzy rough set theories," in *R package version 1.3-0*, 2015.
- [31] Q. Shen and A. Chouchoulas, "A modular approach to generating fuzzy rules with reduced attributes for the monitoring of complex systems," *Engineering Applications of Artificial Intelligence*, vol. 13, no. 3, pp. 263–278, Jun. 2000, doi: 10.1016/S0952-1976(00)00010-5.
- [32] R. S. Michalski, K. A. Kaufman, and J. Wnek, "The AQ family of learning programs: A review of recent developments and an exemplary application," Reports of the Machine Learning and Inference Laboratory, 1991.

## BIOGRAPHIES OF AUTHORS






**Nahier Aldhafferi**    is an associate professor at the College of Computer Sciences and Information Technology, Department of Computer Information Systems, Imam Abdurrahman Bin Faisal University in Saudi Arabia. He received his BS in 2005 from Dammam Teachers college, Saudi Arabia. He also received his master's degree in 2009 in Internet Technology from Wollongong University, Australia. Currently, he has PhD in Information Technology from New England University, Australia. He has published a lot of research in Data science and Information privacy and Artificial intelligence. He can be contacted at email: [naldhafferi@iau.edu.sa](mailto:naldhafferi@iau.edu.sa).






**Abdullah Alqahtani**    is an associate professor in the Computer Information Systems Department, College of Computer Science and Information Technology (CCSIT) at Imam Abdulrahman bin Faisal University (IAU), Saudi Arabia. He joined CCSIT in April 2004. Dr. Alqahtani received his Ph.D. degree in Software Engineering from University of Technology, Sydney, Australia in 2015. He received his M.S. degree in Information and Communication Technology from University of Wollongong, Australia in 2008 and B.S. degree in computer from Computer Department at College of Education in Dammam, Saudi Arabia in 2002. Dr. Alqahtani was the computer science teacher for two years (2002-2004). Recently, Dr. Alqahtani is the Vice Dean for Postgraduate Studies and Scientific Research at CCSIT. His current research interests are in the fields of data science, data mining, e-gov service integration, privacy, machine learning, and software engineering. He can be contacted at email: [aamqahtani@iau.edu.sa](mailto:aamqahtani@iau.edu.sa).






**Fatema Sabeen Shaikh**    received a PhD in the field of heterogeneous mobile and wireless networks from Middlesex University in London. Her specialization includes the development of intelligent, context aware solutions to facilitate the seamless functioning of multi-homed, multi-interfaced mobile clients in converged, heterogeneous networks. Her research interests also include AI in the field of Medicine and big data solutions for the development of software architectures for Self-Learning Health Systems to support personalized treatments and clinical trials. Fatema is also involved in the development of pedagogical solutions to facilitate curriculum internationalization. She is a Fellow of the Higher Education Academy (UK) and currently works as an assistant professor at Imam Abdulrahman bin Faisal University. She can be contacted at email: fsshaikh@iau.edu.sa.






**Sunday Olusanya Olatunji**    is currently an Associate Professor in the College of Computer Science & IT, Imam Abdulrahman Bin Faisal University, Dammam, Saudi Arabia. He obtained his PhD in Computer Science from the Universiti Teknologi Malaysia (UTM) in 2012, bagging the best graduating Student Award Medal (PhD Computer Science). He received M.Sc. Degree in Computer Science, University of Ibadan, Nigeria in 2003. He obtained another M.Sc. Degree in Information and Computer Science, King Fahd University of Petroleum and Minerals (KFUPM), Saudi Arabia in 2008 under the prestigious special Research Assistantship scholarship of KFUPM. He received the B.Sc. (Hons) Degree in Computer Science, Ondo State University (Now University of Ado Ekiti), Nigeria in 1999. He has authored well over 100 publications as journal articles, books, and peer reviewed conferences proceedings, with Journals articles mostly in peer reviewed highly rated impact reputable SCOPUS and ISI journals. He can be contacted at email: osunday@iau.edu.sa; oluolatumji.aadam@gmail.com.






**Abdullah Almurayh**    is a Computer Science Assistant Professor at Imam Abdulrahman Bin Faisal University, Saudi Arabia. He received his bachelor's degree in computer science in 2002 from Teaching college, Saudi Arabia, his master's degree in computer science in 2011 from University of Colorado Colorado Springs (UCCS), United States. He earned his PhD in computer Science from University of Colorado Colorado Springs (UCCS) in 2014, with emphasis of Human Computer Interaction (HCI). Dr. Almurayh has invented patents registered in US Patent office in computer science, artificial intelligence, and internet of things fields. He can be contacted at email: asalmurayh@iau.edu.sa, and can be reached on social media (Linkedin, Facebook, and Twitter) at @almurayh.






**Fahad Ali Alghamdi**    is currently an assistant at the Faculty of Applied Studies and Community Service, Department of Management of Information Systems, Imam Abdurrahman Bin Faisal University, Saudi Arabia. He received his BS degree in computer science in 2002 from Imam Abdurrahman bin Faisal University, Saudi Arabia, his MSc degree in computer science in 2009 from University Wollongong, Australia, and his PhD in Computer Science in 2017 from Aberystwyth University in UK. His research interests include network, applied computer software, cloud computing and online learning. He is currently the Vice Dean of Faculty of Applied Studies and Community Service, Imam Abdurrahman Bin Faisal University, Saudi Arabia. He can be contacted at email: faghamdi@iau.edu.sa.






**Ghalib H. Alshammri**    is an assistant professor in artificial intelligence and data science at King Saud University - Community College and Scientific Research Consultant at Saudi Electronic University. Furthermore, he is a Co-Founder and Board Member at Saudi Society for Data Science (SSDS). He received Ph.D. with particular interests in Artificial Intelligence and Nano Molecular Communication at Stevens Institute of technology, USA in May 2019. He completed MSc in Intelligent Systems from the University of Wollongong, Australia in 2009. His research interests also relate to artificial intelligence, robotics, machine learning, fuzzy techniques, and internet of bio-things. He can be contacted at email: galshammri@ksu.edu.sa.






**Amani K. Samha**    is an assistant professor in artificial intelligence and data science at King Saud University – College of Business Administration, Management Information Systems and Digital Transformation and Information Security Consultant at the Ministry of Education. Furthermore, she is consultant in the National Committee for Labour Committees in Saudi Arabia (Saudi Arabia L20). She was Vice Dean for Quality and Development in 2018. She received Ph.D. degree with particular interests in Data Mining at Queensland University of Technology – Information Technology Department, Australia in July 2016, which is under dissertation untitled “Aspect-Based Opinion Mining from Customer Reviews”. She completed a Master of Information technology Studies from the Australian National University, Australia, Canberra in July 2011. Her research interests also relate to Artificial Intelligence, Machine Learning, Fuzzy Techniques, and Internet of Things. She published many scientific papers with high citations. She can be contacted at email: [asamha@ksu.edu.sa](mailto:asamha@ksu.edu.sa).






**Mutasem Khalil Alsmadi**    is currently an associate professor at the Faculty of Applied Studies and Community Service, Department of Management of Information Systems, Imam Abdurrahman Bin Faisal University. He received his BS degree in Software engineering in 2006 from Philadelphia University, Jordan, his MSc degree in intelligent systems in 2007 from University Utara Malaysia, Malaysia, and his PhD in Computer Science from The National University of Malaysia. He has published more than one hundred papers in the image processing and Algorithm optimization areas. His research interests include Artificial intelligence, Pattern recognition, Algorithms optimization and Computer vision. He can be contacted at email: [mksalsmadi@gmail.com](mailto:mksalsmadi@gmail.com).



**Hayat Alfagham**    received her BS degree in Computer Information Systems from King Faisal University, Saudi Arabia, her MSc degree in intelligent systems in 2015 from Kings College London, United Kingdom, and her PhD in Computer Science in 2020 from Queen Mary University of London. She is currently an associate professor at the Faculty of Applied Studies and Community Service, Department of Management of Information Systems, Imam Abdurrahman Bin Faisal University. Her research interests include Link prediction, Big Data, Pattern recognition, Machine Learning, Network evolution. She can be contacted at email: [hmalfagham@iau.edu.sa](mailto:hmalfagham@iau.edu.sa).



**Abderrazak Ben Salah**    is currently an assistant professor at the Faculty of Applied Studies and Community Service, Department of Management of Information Systems, Imam Abdurrahman Bin Faisal University, Dammam KSA. He received his Engineer degree in Software engineering in 1991 from Wrocław University of Science and Technology, Poland, his MSc degree in Computers Networks in 1992 from Wrocław University of Science and Technology, Poland, and his PhD in Applied Computer Science in 1995 from Wrocław University of Economics and Business, Poland. He has published many papers in the Applied Computer Science areas. His research interests include Applied Computer Science, E Commerce Security, Networks and Artificial intelligence. He can be contacted at email: [asalah@iau.edu.sa](mailto:asalah@iau.edu.sa).