Churn Prediction Using Machine Learning and Recommendations Plans for Telecoms

Khulood Ebrah[1*], Selma Elnasir[2]

[1]Department of Information Technology, International University of Africa, Khartoum, Sudan
[2]Department of Computer Science, International University of Africa, Khartoum, Sudan
Email: *Khulood.ebrah@gmail.com, Selmanasir@hotmail.com
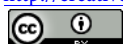
Abstract

Keeping customers satisfied is truly essential for saying that business is successful especially in the telecom. Many companies experience different techniques that can predict churn rates and help in designing effective plans for customer retention since the cost of acquiring a new customer is much higher than the cost of retaining the existing one. In this paper, three machine learning algorithms have been used to predict churn namely, Naïve Bayes, SVM and decision trees using two benchmark datasets IBM Watson dataset, which consist of 7033 observations, 21 attributes and cell2cell dataset that contains 71,047 observations and 57 attributes. The models' performance has been measured by the area under the curve (AUC) and they scored 0.82, 0.87, 0.77 respectively for IBM dataset and 0.98, 0.99, 0.98 respectively for cell2cell dataset. The proposed models also obtained better accuracy than the previous studies using the same datasets.

1. Introduction

Customers retaining is the most important asset for any business as it is stated that "the cost of acquiring a new customer can be higher than that of retaining a customer by as much as 700%; increasing customer retention rates by a mere 5% could increase profits by 25% to 95%" [1]. So one of the best solution to retain the customers is to reduce churn rate, where "churn" means moving the customer from service provider to another one, or stopping using specific services over specific periods for many reasons that can be detected previously if the

company analyzes its data records and uses machine learning technology which enables the companies to predict the customers who are likely to churn. A lot of studies approved its efficiency to this situation [2] [3] [4] so the company can respond quickly to the behavioral changes in the customer's minds. Telco's today is refining & optimizing the customer experience which is the key to sustaining a market differentiation and reducing churn [5], where retaining an existing customer costs much lower than acquiring a new one. This research studies the machine learning algorithms and recommended the best solutions for telecoms. In the competitive telecom sector, customers can easily switch from one provider to another, which lets the telecom providers worried about their customers and how to retain them but they can predict the customers who will move to another provider previously by analyzing their behavior. They can retain them by providing offers and their preferred services according to their historical records so the aim of this study is to predict churn previously and detect the main factors that may let the user move to another provider in telecoms.

2. Related Work

Many studies are available for churn problem from different viewpoints with different datasets, algorithm and for different industries where churn analysis is one of the world wide used to analyze the customer behaviors and predict the customers who are about to leave the service agreement from a company. Studies revealed that gaining new customers is 5 to 10 times costlier than keeping existing customers happy and loyal in today's competitive conditions, and that an average company loses 10 to 30 percent of customers annually [6] [7]. Most of the literature focused more on data mining algorithms, but only a few of them focused on distinguishing the important input variables for churn prediction and on enhancing the data samples through efficient pre-processing to be used for data mining algorithms implementation [8] [9]. Amin, A., *et al.* [10] presented a novel churn prediction approach based on the classifier's certainty estimation using distance factor where they grouped the dataset into different zones based on the distance which are then divided into two categories with high and low certainty, they used 4 datasets with different samples and they have been discretized by size, the values that exists in each attribute of the dataset, and then assigned certain labels and at the end produced specific list of values in different number of groups of an attribute. They used Naïve Bayes as classifier and it obtained high accuracy in the zone with greater distance factor's value (*i.e.,* customer churn and non-churn with high certainty) than those placed in the zone with smaller distance factor's value (*i.e.,* customer churn and non-churn with low certainty). Accuracy in the last tenth iteration was (82.91% & 84.30%, 70.60% & 74.80%, 70.00% & 89.01%, 57.00% & 56.00%) for the (UDT & LDT) on the 4 datasets used. Andrews, R., *et al.* [3] used dataset of 10,000 client records from telecom each with 21 attribute, in which 2900 are churners from customers of a Telecom Company in Belgium. They applied profound learning models and they used 10-overlap cross approval methods to check the predic-

tion exactness and the area under curve score is 0.89. Ahmad, A. K., Jafar, A. and Aljoumaa, K. [2] developed machine learning techniques on big data platform for analyzing data from SyriaTel telecom contained all customers' information over 9 months. The model experimented four algorithms: Decision Tree, Random Forest, Gradient Boosted Machine Tree "GBM" and Extreme Gradient Boosting "XGBOOST". The AUC for the four models were 83, 87.76, 90.89 and 93.3. The best results were obtained by applying XGBOOST and it obtained 93.3% where it used (SNA) features, which enhanced the performance of the model from 84% to 93.3%. The model was prepared and tested through Spark environment. Saraswat, S. and Tiwari, A. [11] described a framework that was proposed to conduct for the churn prediction model using Naïve Bayes algorithm for classification task and then apply Elephant Herding Optimization algorithm for solving optimization task used the dataset which was obtained from https://www.kaggle.com and it contains 21 attributes and 3333 instances. Data contains 483 churn' customer where predicted 244 correctly as churner customer using naïve equation and after applying Elephant Herding Optimization Algorithm 199 churner, model accuracy is 87%. Different algorithms are used by Ahmed, A.A. and D. Maheswari [12], which are Firefly algorithm and the Hybrid Firefly algorithm on Orange Dataset which contains 50,000 samples and 230 attributes. The dataset was segregated with 90% data for training and 10% for testing. The search space was populated with 20 fireflies and classification was carried out with a *maxgen* of 1000. The ACC obtained is (86.36%, 86.38%). Some researchers compared between different models as Kumar, N. and C. Naik [13] who used three models Logistic regression, random forest and balanced random forest on dataset contains from 25,000 samples and 110 attributes and used PCA for feature selection and partitions used 70% & 30% for training and testing. The result presented that Logistic regression model has the highest area under the curve where the ACC of the three models (0.861, 0.83, 0.83).

3. The Research Strategy

The method used in this paper has been summarized in Figure 1 and it has been explained in detail in the next paragraphs.

3.1. Datasets Visualization

There are two datasets used in this study. The first dataset consists of 7034 samples and 20 attributes while the second dataset contains 71,047 samples and 57 attributes. Datasets details are as shown in Table 1. Both datasets have been visualized using Orange.

In Figure 2 & Figure 3 the churn class histogram for both datasets were illustrated. The 0's value refers to the non-churned customers and shown in blue color and the 1's value refers to the churned customers and shown in orange color.

The samples from IBM dataset shown in Table 2 are the features which have been used in prediction models.

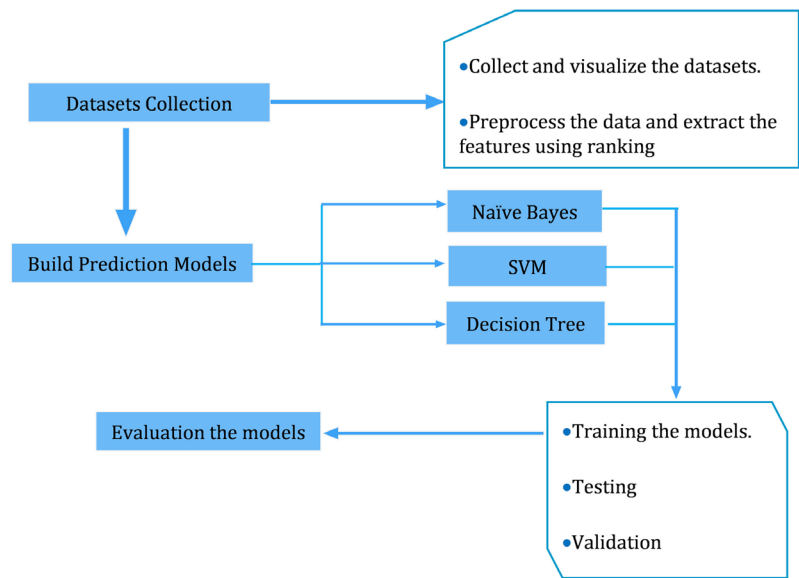And Table 3 includes the samples with the features of cell2cell dataset.
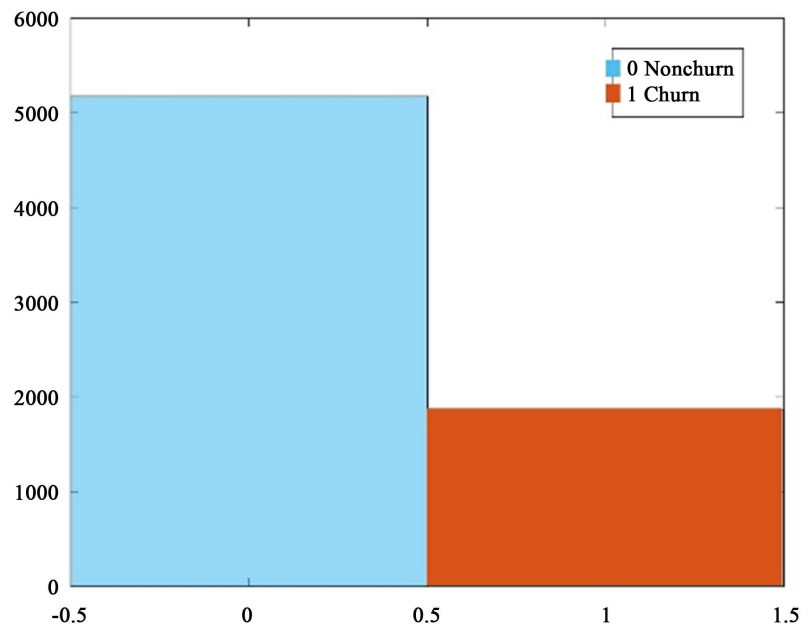
Figure 1. The research strategy.

Figure 2. Churn on IBM dataset.

Table 1. Datasets used.

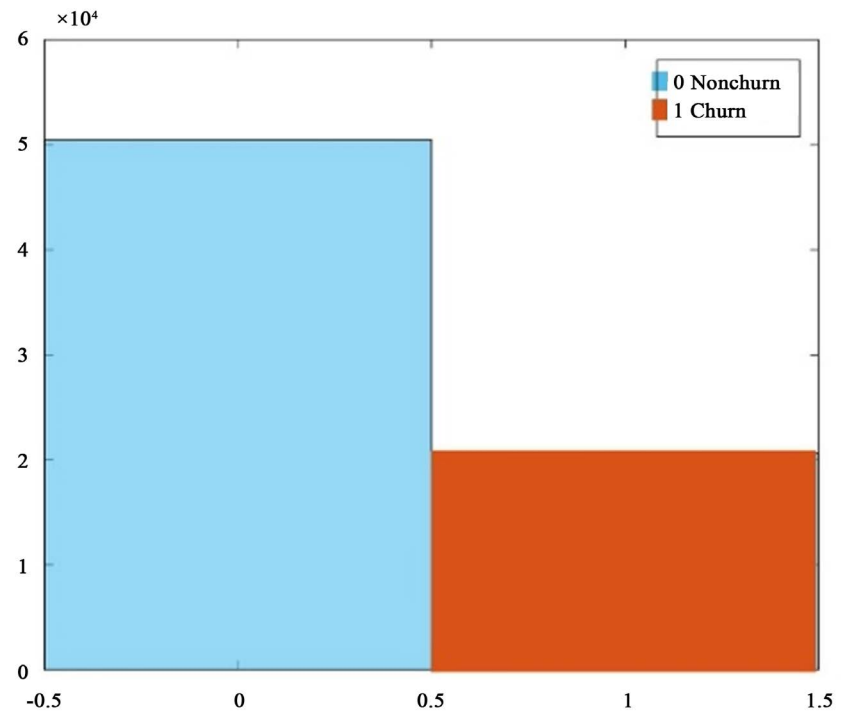| Dataset	Dataset 1	Dataset 2
Samples	7034	71,047
Features	20	57
Classes	2	2
Missing values %	0.0%	0.7%
negative samples	1869 (73.46%)	20,609 (29.01%)
Data sources	IBM Watson [14] [15]	Cell2cell [16]

Figure 3. Churn on cell2cell dataset.

Table 2. IBM dataset samples.

Churn	Dependents	Tenure	Phone Service	Multiple Lines	Internet Service	Online Backup	Device Protection	Contract	Paperless Billing	Payment Method	Monthly Charges	Total Charges
0	0	1	0	2	1	1	0	1	1	1	29.9	29.85
0	0	34	1	0	1	0	1	2	0	2	57	1890
1	0	2	1	0	1	1	0	1	1	2	53.9	108.2
0	0	45	0	2	1	0	1	2	0	3	42.3	1841
1	0	2	1	0	2	0	0	1	1	1	70.7	151.7
1	0	8	1	1	2	0	1	1	1	1	99.7	820.5
0	1	22	1	1	2	1	0	1	1	4	89.1	1949

Table 3. Cell2cell dataset samples.

churn	revenue	mou	recchrge	changem	custcare	mourec	months	phones	models	eqpdays	creditaa	prizmtwn	webcap
0	−6.2	0	−6	0	0	0	7	1	1	203	0	0	1
0	−5.9	0	−5	0	0	0	15	1	1	452	0	0	1
0	−2.5	211	0.5	NA	1	1.69	18	2	2	281	0	0	1
1	0	2	0	NA	0	0	27	2	2	597	1	0	1
1	0	55	0	NA	5	7.06	26	3	3	371	0	0	1
0	0	0	0	0	0	0	7	1	1	199	0	0	1
0	0	76	30	0	1	11.2	30	1	1	883	0	0	1
0	0.2	12	0	0	0	7.24	31	3	3	263	0	0	1

3.1.1. IBM Dataset Visualization and Preprocessing

The dataset is for customers who left within the last month. The column is called Churn where it contains the below attributes [14]:

- Services that each customer has signed up, internet, online security, online backup, device protection, tech support, and streaming TV and movies;
- Customer account information how long they've been a customer, contract, payment method, paperless billing, monthly charges, and total charges;
- Demographic info about customers—gender, age range, and if they have partners and dependents.

Figures 4-12 show the attributes and their distributions according to the churn class where orange color indicates the churn customers and the blue for non-churn. As noticed that:

- Most churned customers have internet service type Fiber optic;
- They use paperless billing;
- Most of the customers were dependents;
- Their payment method was electronic check;
- They don't use "device protection" or "online backup" services, rather they use phone service;
- Their tenure was less than 14 months.

Therefore, the predictor attributes have been selected according to this analysis.

Figure 10 and Figure 11 show the correlation between Total charges, Monthly Charge and Tenure.

3.1.2. Cell2cell Dataset Visualization and Preprocessing

Cell2cell is the 6th largest wireless company in the US, Cell2cell dataset consists of 71,047 signifying whether the customer had left the company two months after observation and 57 attributes [17]. The histograms Figures 13-17 show the attributes and their distributions according to the churn in similar way as done with IBM dataset visualization. What has been noticed on cell2cell dataset that:

- Churned customers have average (mean) monthly minutes of use which is less than 530' minute;
- They have service for only 11 - 15 months;
- The numbers of days for their equipment were between 300 & 361 day;
- The numbers of models issues are less than 2;
- Their prizm code refer to town;
- Their handsets have web capability.

The excluded data according to the churn class has been illustrated in Figure 18 and Figure 19. Where Figure 18 plots the Churn attribute vs Total Revenue and Total Charge. Figure 19 plots the surface fitting for Total charges, Change in Miute Use and Change in Revenues. There are 238 outliers samples have been removed and they have been marked by red color.
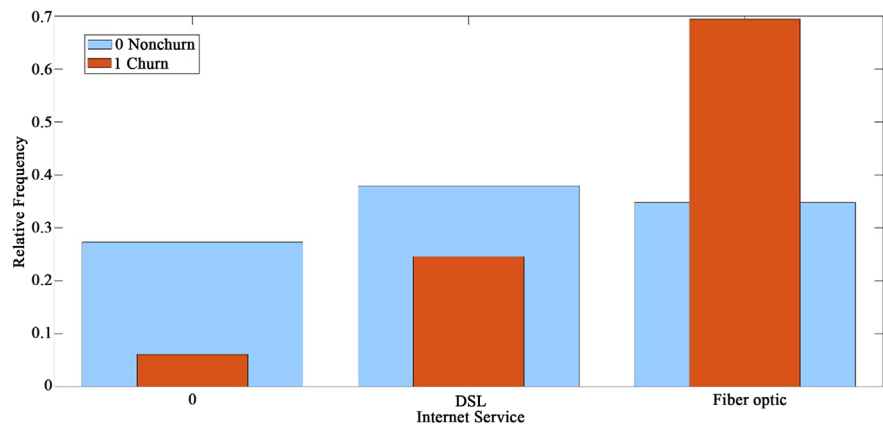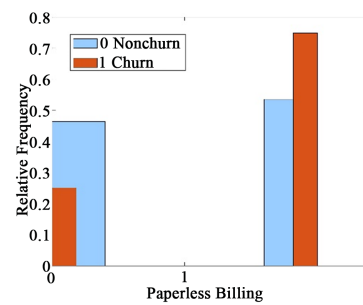
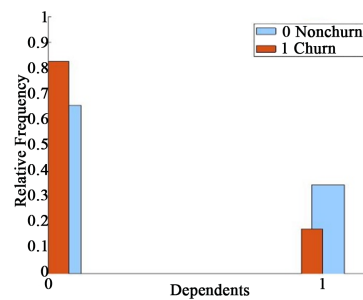**Figure 4.** Internet service.
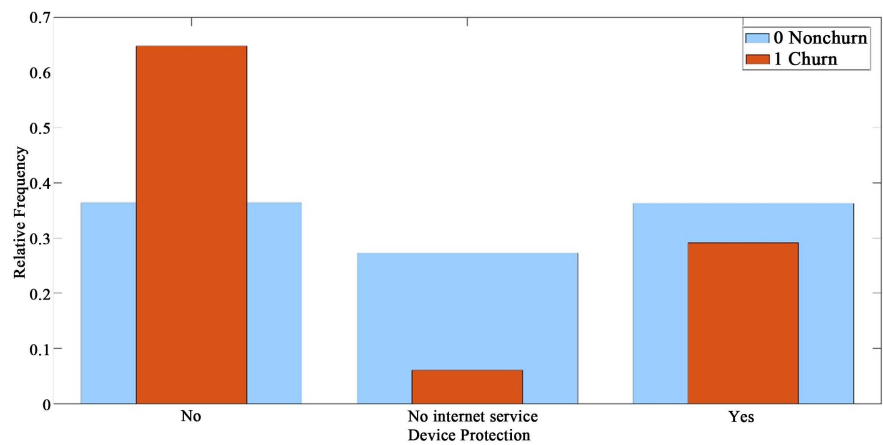

Figure 5. Paperless billing.
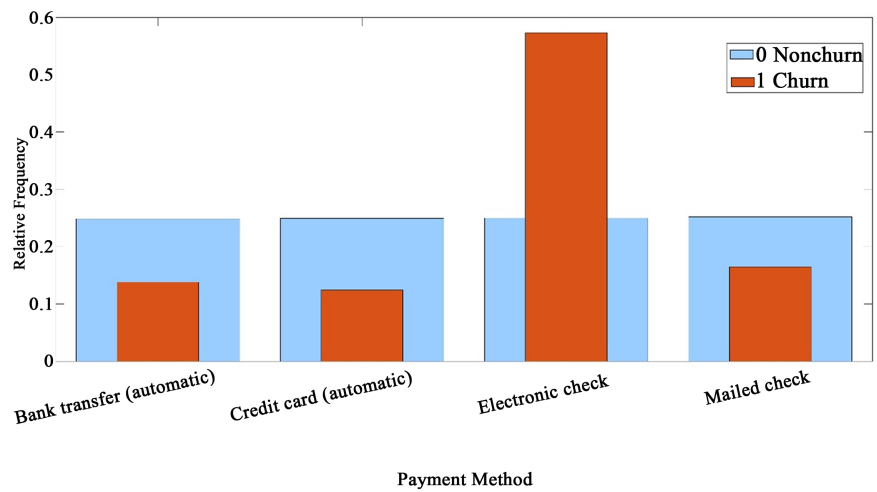

Figure 6. Dependents.

Figure 7. Device protection.
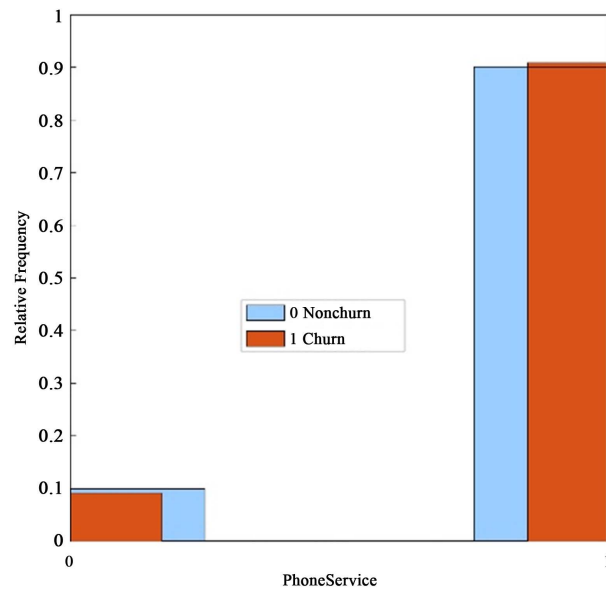
Figure 8. Payment method.
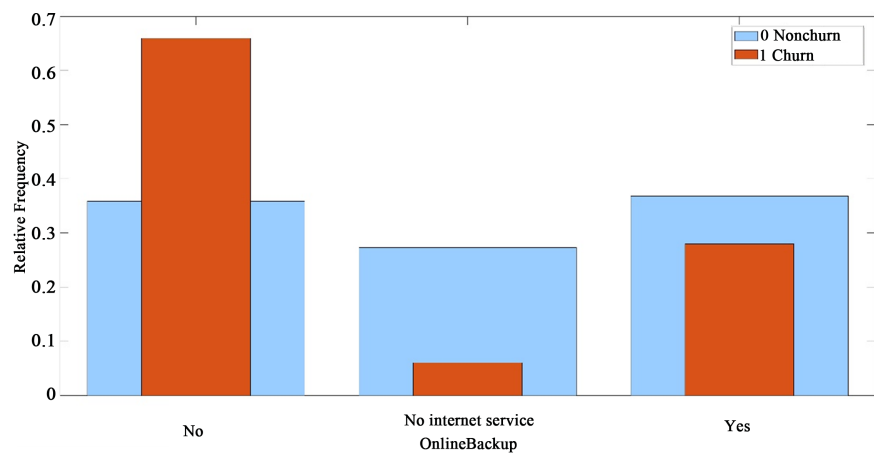

Figure 9. Phone service.
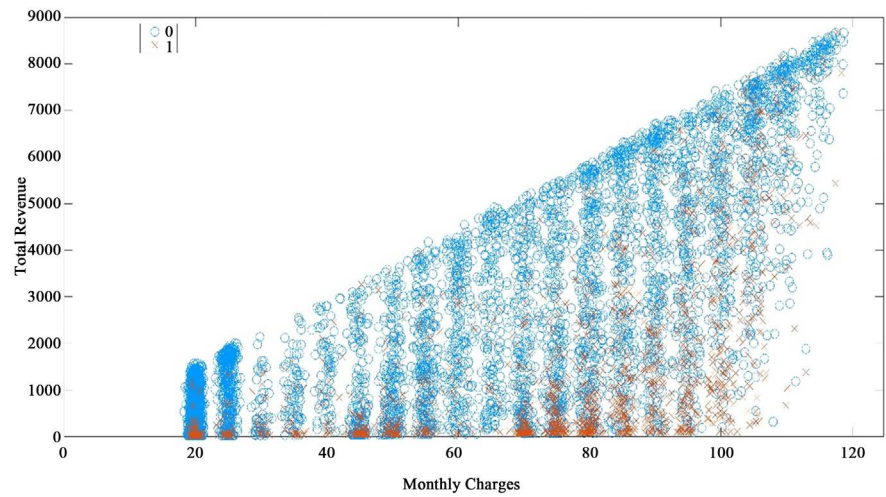

Figure 10. Online backup.

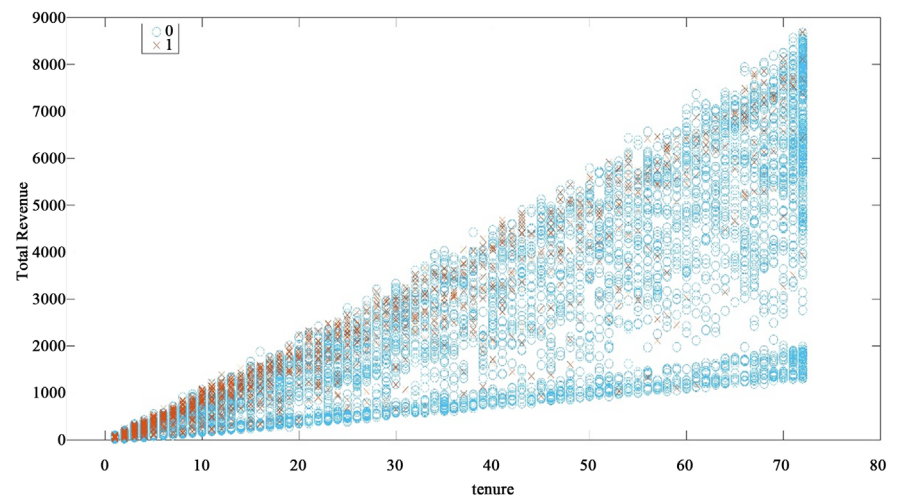**Figure 11.** Monthly charges & total charges.

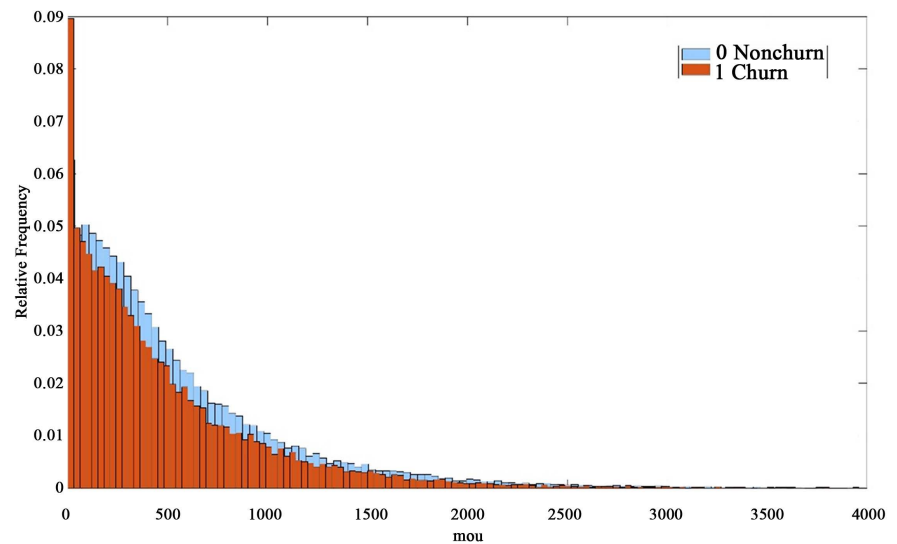**Figure 12.** Tenure & total charges.

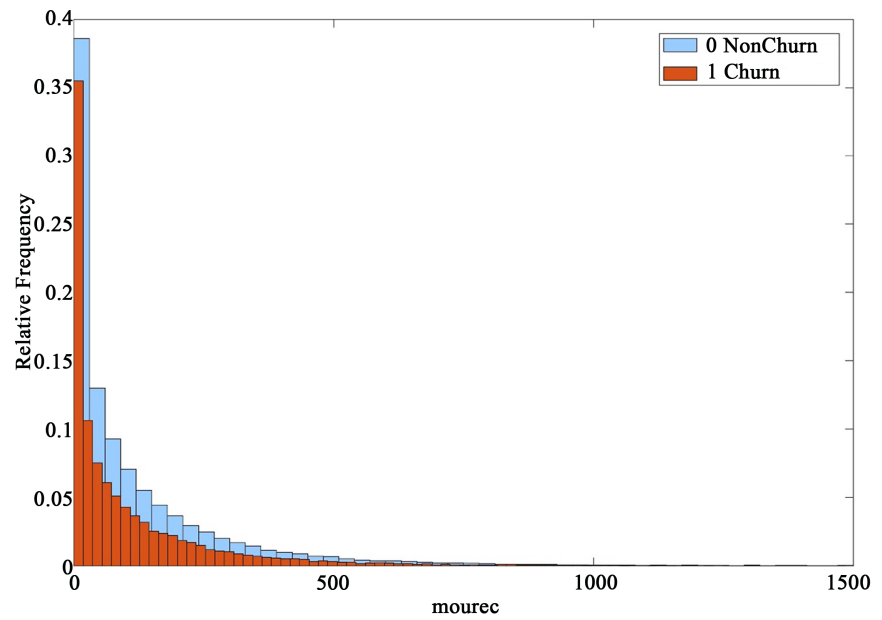**Figure 13.** Number of months in service.

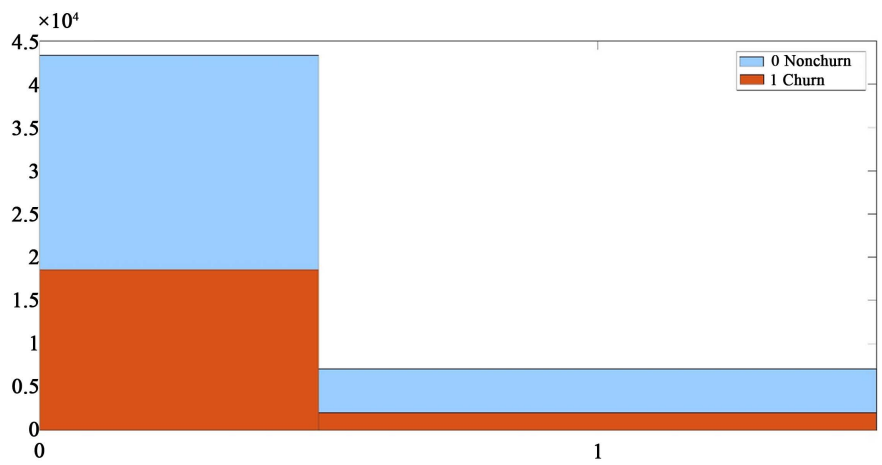**Figure 14.** Mean monthly minutes of use.
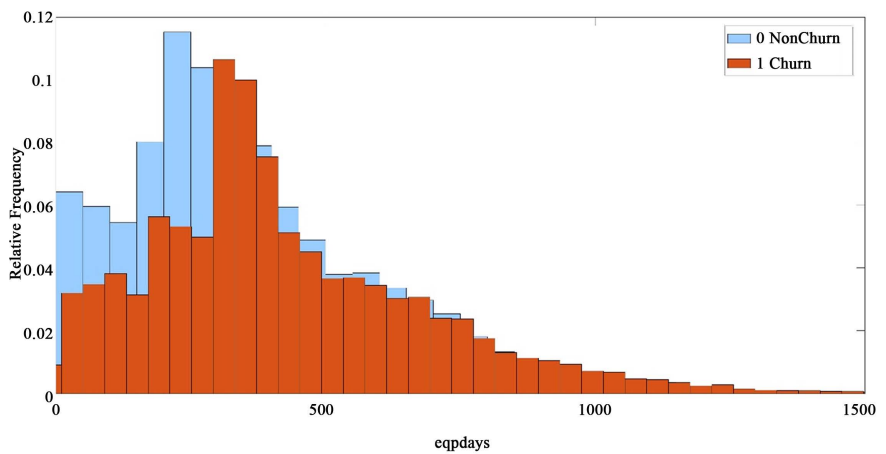

Figure 15. Creditaa.

Figure 16. Number of days of the current equipment.
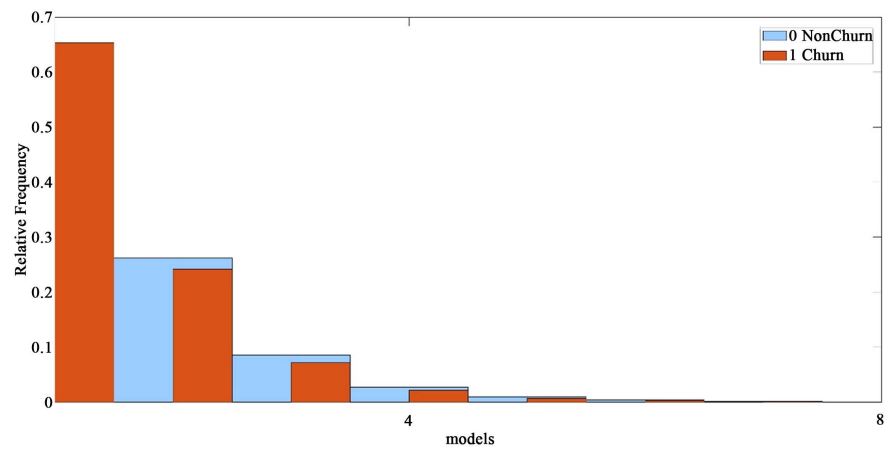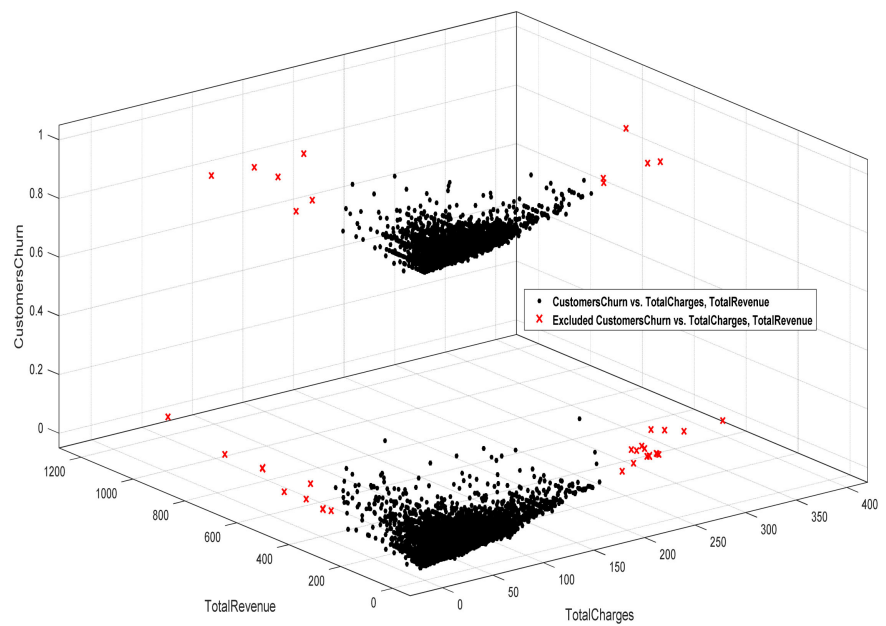
Figure 17. Models issued.

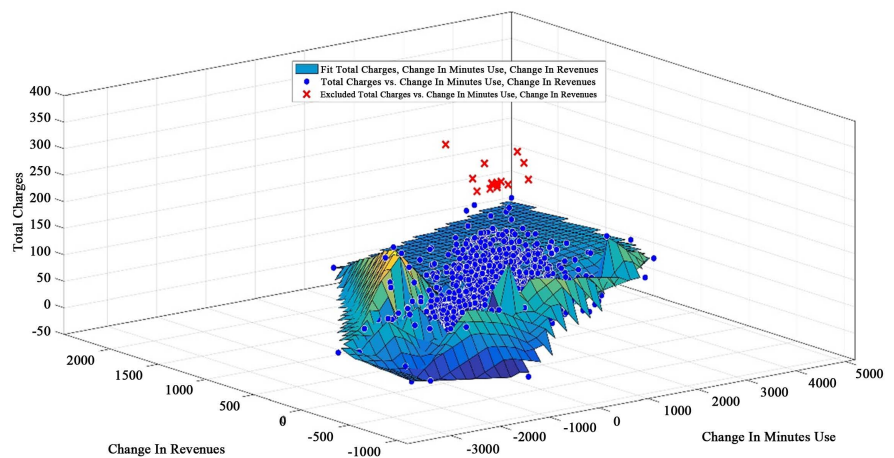**Figure 18.** Churn vs. total revenue, total charge.

Figure 19. Total charges vs. change in miute use, change in revenues.

3.2. Naïve Bayes Algorithm

The Naive Bayes algorithm is a classification algorithm based on Bayes rule and a set of conditional independence assumptions [18]. To predict the class label of X, $P(X|C_i)P(C_i)$ is evaluated for each class C_i. The classifier predicts that the class label of tuple X is the class C_i if and only if

$$P(X|C_i)P(C_i) > P(X|C_j)P(C_j) \text{ for } 1 \le j \le m, j \ne i \qquad (1)$$

In other words, the predicted class label is the class C_i for which $P(X|C_i)P(C_i)$ is the maximum [19]. Models posterior probabilities according to Bayes rule. That is, for all $k = 1, \cdots, K$,

$$\hat{P}(Y = k \mid X_1, \cdots, X_p) = \frac{\pi(Y = k)\prod_{j=1}^{P} P(X_j \mid Y = k)}{\sum_{k=1}^{K} \pi(Y = k)\prod_{j=1}^{P} P(X_j \mid Y = k)} \qquad (2)$$

where:

Y is the random variable corresponding to the churn class index of an observation.

$X_1, \cdots, X_p$ are the predictors of an observation.

$\pi(Y = k)$ is the prior probability that a class index is k.

The model use mean and standard deviation to distrubite the predictors within each class.

Naive Bayes classification classify data into the training data, the method estimates the parameters of a probability distribution, assuming predictors are conditionally independent given the class. Prediction step: For any unseen test data, the method computes the posterior probability of that sample belonging to each class. The method then classifies the test data according the largest posterior probability.

3.3. Support Vector Machine Algorithm

SVM algorithm for the classification of both linear and nonlinear data. It transforms the original data into a higher dimension, from where it can find a hyperplane for data separation using essential training tuples called support vectors [19]. The SVM binary classification algorithm searches for an optimal hyperplane that separates the data into two classes. For separable classes, the optimal hyperplane maximizes a margin (space that does not contain any observations) surrounding itself, which creates boundaries for the positive and negative classes. The data for training is a set of points (vectors) x_j along with their categories y_j. For some dimension d, the $x_j \in R^d$, and the $y_j = \pm 1$. The equation of a hyperplane is [20]

$$f(x) = x'\beta + b = 0 \qquad (3)$$

where $\beta \in R^d$ and b is a real number.

As the data used is not allow for a separating hyperplane, the SVM used a *soft margin*, meaning a hyperplane that separates many, but not all data points.

There are two standard formulations of soft margins. Both involve adding slack variables $\xi = (\xi_1, \xi_2, \cdots, \xi_N)$ and a penalty parameter C.

- The L^1-norm problem is:

$$\min \beta, b, \xi (12\beta'\beta + C \sum j\xi j) \tag{4}$$

such that

$$yjf(xj) \geq 1 - \xi j\xi j \geq 0 \tag{5}$$

- The L^2-norm problem is:

$$\min \beta, b, \xi (12\beta'\beta + C \sum j\xi j2) \tag{6}$$

In these formulations, it can be used C places more weight on the slack variables ξj, meaning the optimization attempts to make a stricter separation between classes. Equivalently, reducing C towards 0 makes misclassification less important.

The propsed SVM model standardizes the predictors using their corresponding weighted means and weighted standard deviations. Means it standardizes predictor $j(x_j)$ using

$$xj* = xj - \mu j * \sigma j \tag{7}$$

$$\mu j* = 1 \sum kwk * \sum kwk * xjk \tag{8}$$

X_{jk} is observation k (row) of predictor j (column).

$$(\sigma j*)2 = v1v12 - v2 \sum kwk * (xjk - \mu j*)2 \tag{9}$$

$$v1 = \sum jwj* \tag{10}$$

$$v2 = \sum j(wj*)2 \tag{11}$$

3.4. Decision Tree Algorithm

Decision tree induction is the learning of decision trees from class-labeled training tuples. A decision tree is a flowchart-like tree structure, where each internal node (nonleaf node) denotes a test on an attribute, each branch represents an outcome of the test, and each leaf node (or terminal node) holds a class label. The topmost node in a tree is the root node [19]. The Classification Tree splits nodes based on either impurity or node error. Impurity means one of several things, depending on the Split Criterion name-value pair argument:

- Gini's Diversity Index (gdi)—the Gini index of a node is

$$\text{Gini}(D) = 1 - \sum_{i=1}^{m} P_i^2, \tag{12}$$

where the sum is over the classes i at the node, and $p(i)$ is the observed fraction of classes with class i that reach the node. A node with just one class (a pure node) has Gini index 0; otherwise the Gini index is positive. So the Gini index is a measure of node impurity.

- Deviance ("deviance")—with $p(i)$ defined the same as for the Gini index, the deviance of a node is

$$\sum_i p(i)\log_2 p(i) \tag{13}$$

A pure node has deviance 0; otherwise, the deviance is positive.

3.5. Models Evaluation's Methods

The models have been evaluted using the holdout method and *k*-fold cross-validation. In the holdout partition method, the given data are randomly partitioned into two independent sets, a training set and a test set. [19]. And in this partition type, a scalar parameter (let's say "*p*") which randomly selects approximately $p*$ n observations for the test set. The *p* value used here is 0.3 which divided datasets into 70% for training and 30% for testing. In *k*-fold cross-validation, the initial data are randomly partitioned into *k* mutually exclusive subsets or "folds", D_1, D_2, ... D_k, each of approximately equal size. Training and testing is performed *k* times [19]. The datasets here have been divided into 10 folds.

4. Experiments and Results

The three models trained on IBM and cell2cell datasets and have been divided into training and test sets using cross validation with partition types "hold-out" 30% and "*k*-fold" where the *k* value used is 10. The training and testing error shown in Table 4 and it shows the best result obtained from training and testing. The models have been trained from 4 to 5 times for each dataset and they didn't give better accuracy.

The ROC curve for IBM dataset shown in Figures 20-22 according to Table 2 for each model output. Whereas Figures 23-25 show the ROC for cell2cell dataset according Table 4 too. ROC curve for each of the three models shows the trade-off between the true positive rate (TPR) and the false positive rate (FPR). Given a test set and a model, TPR is the proportion of positive (churned) tuples that are correctly labeled by the model; FPR is the proportion of negative (nochurn) tuples that are mislabeled as positive [19].

In the following experiment the models were evaluted with *k*-fold value of 10, as shown in Table 5 for both datasets respectvely. There are small variances between error rates within *k*-fold cross-validation experiment. The best result obtained from SVM model in fold number 8 on IBM dataset and in fold number 1 on cell2cell dataset.

Table 4. Training with holdout 30%.

Dataset	Model	Training Error	Testing Error	AUC	ACC
	Naïve Bayes	0.23829	0.25334	0.81721	76%
IBM Waston	SVM	0.20564	0.20335	0.83683	80%
	Decision Tree	0.086392	0.23684	0.76198	76.3%
	Naïve Bayes	0.021036	0.019907	0.98149	90%
Cell2cell	SVM	0.0082883	0.0089536	0.99212	98.2%
	Decision Tree	0.12620	0.012142	0.9855	98.8%

Table 5. Training with *k*-fold with 10 value for IBM and cell2cell datasets.

| Model	Fold	IBM Dataset				Cell2cell Dataset			
		Training Error	Testing Error	AUC	ACC	Training Error	Testing Error	AUC	ACC
Naïve Bayes	1	0.2436	0.2270	0.8135	77.30%	0.0309	0.0305	0.9665	97
	2	0.2435	0.2482	0.8149	75.20%	0.0295	0.0302	0.9671	97
	3	0.2394	0.2639	0.8156	73.60%	0.0314	0.0254	0.9667	97
	4	0.2403	0.2596	0.8181	74.00%	0.0295	0.0270	0.9681	97.3
	5	0.2423	0.2386	0.8117	76.10%	0.0312	0.0328	0.9663	96.7
	6	0.2390	0.2613	0.8166	73.90%	0.0300	0.0321	0.9691	96.8
	7	0.2440	0.2400	0.8125	76.00%	0.0318	0.0343	0.9647	96.6
	8	0.2425	0.2457	0.8123	75.40%	0.0304	0.0316	0.9679	96.8
	9	0.2447	0.2272	0.8133	77.30%	0.0312	0.0330	0.9652	96.7
	10	0.2448	0.2187	0.8106	79.10%	0.0309	0.0326	0.9677	96.7
SVM	1	0.2048	0.2033	0.8300	79.90%	0.0087	0.0072	0.9927	99
	2	0.2046	0.1836	0.8262	81.70%	0.0086	0.0084	0.9908	98.9
	3	0.2048	0.1838	0.8498	81.70%	0.0087	0.0075	0.9935	99
	4	0.1993	0.2330	0.8064	76.70%	0.0085	0.0085	0.9920	98.9
	5	0.2037	0.1925	0.8447	80.80%	0.0086	0.0082	0.9924	98.8
	6	0.2018	0.2150	0.8138	78.60%	0.0083	0.0109	0.9882	98.7
	7	0.2042	0.1945	0.8239	80.50%	0.0086	0.0079	0.9917	99
	8	0.2062	0.1831	0.8655	81.70%	0.0085	0.0092	0.9926	98.8
	9	0.2002	0.2272	0.8089	77.30%	0.0086	0.0081	0.9926	99
	10	0.2033	0.2118	0.8224	78.80%	0.0085	0.0091	0.9913	98.8
Decision Tree	1	0.0858	0.2499	0.7240	75%	0.0001	0.0120	0.9853	98.8
	2	0.0841	0.2469	0.7667	75.30%	0.0001	0.0130	0.9859	98.7
	3	0.0896	0.2298	0.7447	77%	0.0002	0.0123	0.9826	98.8
	4	0.0882	0.2795	0.7168	72.10%	0.0001	0.0124	0.9846	98.9
	5	0.0880	0.2618	0.7264	73.90%	0.0002	0.0109	0.9887	98.9
	6	0.0882	0.2684	0.6994	73.20%	0.0001	0.0110	0.9853	98.8
	7	0.0868	0.2542	0.7125	74.60%	0.0000	0.0136	0.9831	98.6
	8	0.0841	0.2641	0.7321	73.60%	0.0001	0.0117	0.9854	98.8
	9	0.0844	0.2400	0.7079	76%	0.0002	0.0127	0.9863	98.7
	10	0.0899	0.2854	0.7020	71.40%	0.0001	0.0120	0.9847	98.8
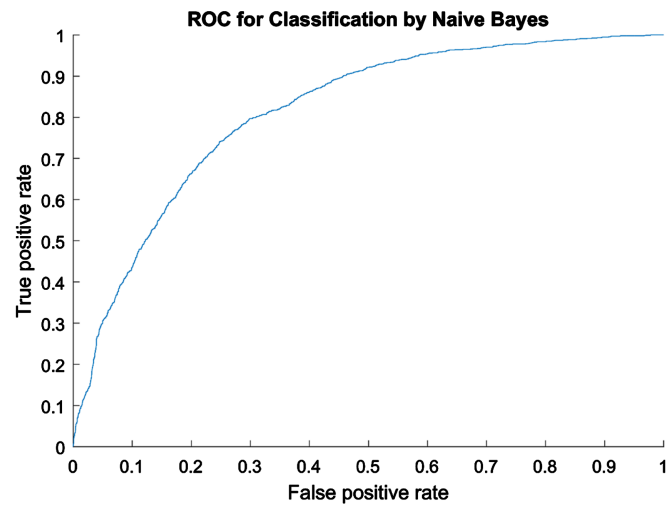
Figure 20. Naïve Bayes (IBM data).
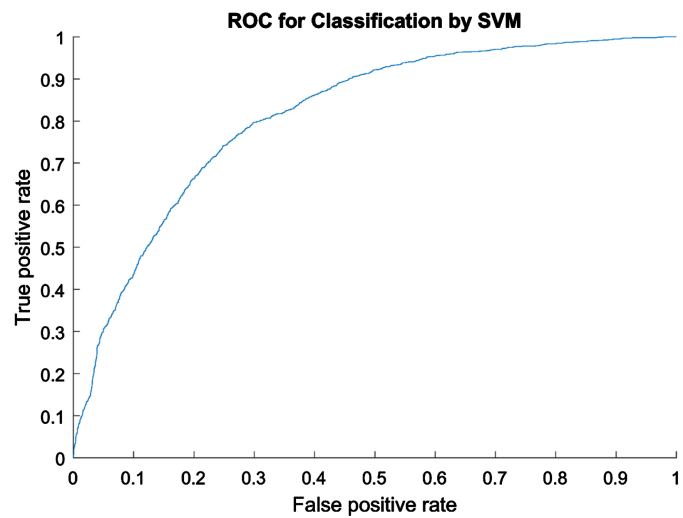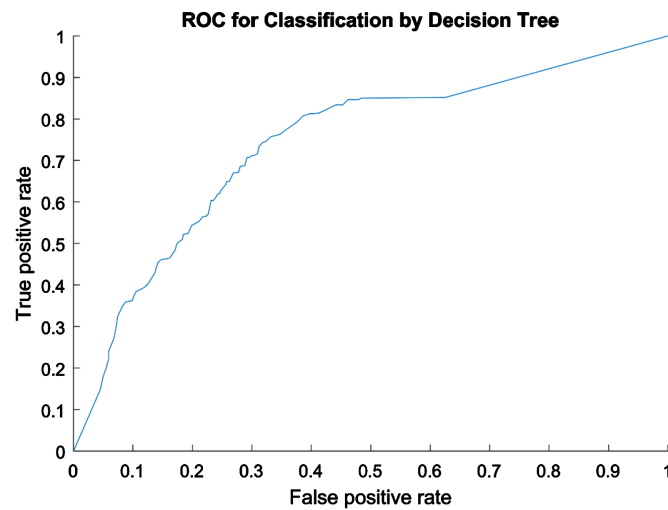

Figure 21. SVM (IBM data).

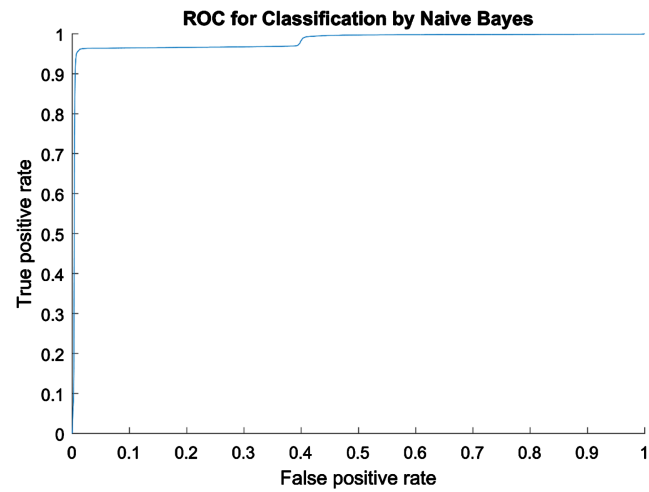**Figure 22.** Decision tree ((IBM data).
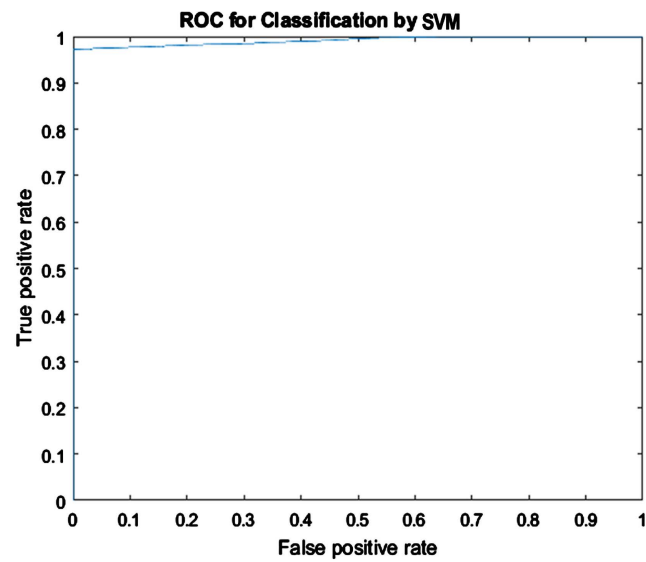
Figure 23. Naïve Bayes (Cell2cell).
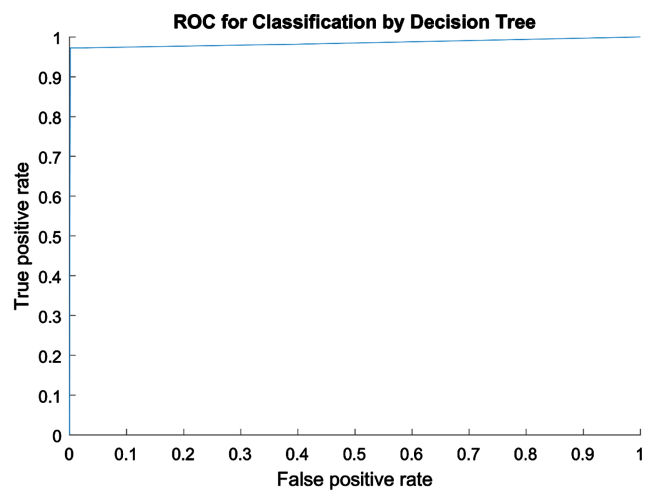

Figure 24. SVM (Cell2cell).

Figure 25. Decision tree (Cell2cell).

In order to check the models, they have been compared with previous papers which used similar datasets. The result approved that the model is more accurate as shown in Table 4.

ApurvaSree, G., *et al.* [4] and Induja, S. & D. V. P. Eswaramurthy [21] used IBM Waston dataset with different algorithms including SVM for the first paper & Naïve Bayes for the second mentioned paper, both results are similar to the results obtained from this paper. However, our proposed method obtained higher accuracy by using SVM model on IBM dataset with *k*-fold partition, *k* value = 10 to produce an area under curve reached to 0.86548. As for cell2cell dataset, the papers in [21] [22] [23] [24] also used different algorithms including SVM where the best accuracy for previous studies was 94.13 for AUC whereas the AUC in the proposed model using SVM is 0.99 as shown in Table 6.

The AUC' values have been plotted for the three models as shown in Figure 26 & Figure 27, which shown that best result obrained using algorithm from SVM for both datasets.
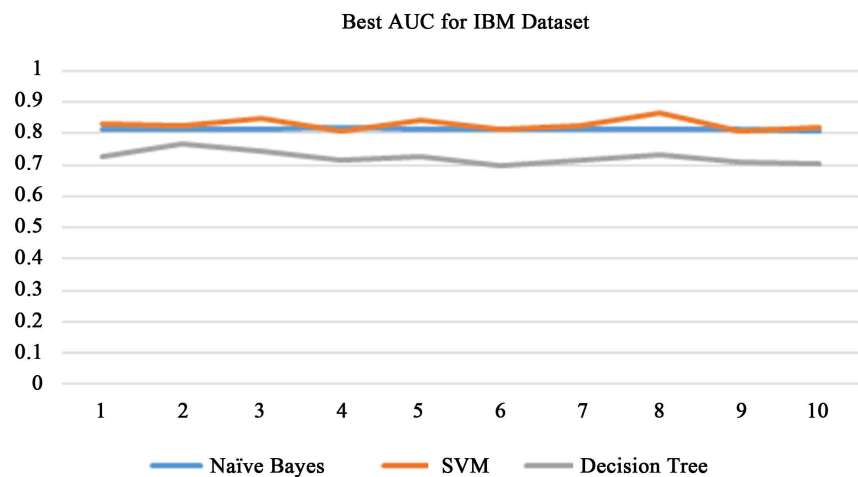

Figure 26. Comparing the models results for IBM dataset.
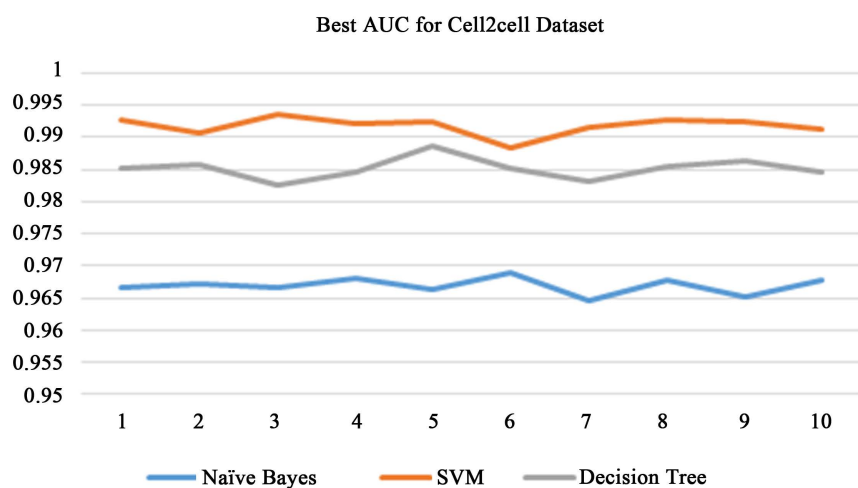

Figure 27. Comparing the models results for Cell2cell dataset.

Table 6. Comparison with previous papers which used the same datasets.

Paper	Algorithms	Result	Dataset	Proposed Model Result
[4]	Random forest, Logistic regression, **SVM**	Accuracy (80.75%, 80.88%, 82%)	IBM Waston	AUC for **SVM** AUC (87%)
[21]	**Naive Bayes**	AUC (83%)	IBM Waston	AUC for **Naive Bayes** same (83%)
[22]	**Decision trees**, Logistic regression, Neural networks and **SVM**	Accuracy (62.98, 61.65, 61.40 and 61.78)	Cell2cell	ACC for **decision trees** (98%) and **SVM** (99%)
[23]	GP-AdaBoost	AUC (0.91)	Cell2cell	The best AUC for **SVM** (99%) AdaBoost not used
[24]	C4.5 decision tree	AUC (63.04)	Cell2cell	AUC **decision trees** (98%)
[25]	**SVM**	AUC (94.13)	Cell2cell	AUC for **SVM** (99%)

5. Conclusion

This paper analyzed two datasets, IBM Watson dataset consists of 7033 observations, 21 attribute and cell2cell dataset consists of 71,047 observations and 57 attribute where they have been visualized using orange software. The three predictive models "Naïve Bayes, SVM and decision tree" have been implemented in Matlab. The paper aims to find the best accurate model for churn prediction in telecom and selecting the most important reasons that let customers churn. The models performance has been measured by area under curve where the best AUCs are (0.82, 0.87, 0.78) for IBM dataset & (0.98, 0.99, 0.98) for cell2cell dataset. The AUC, which obtained using SVM algorithm, is better compared with the previous papers. As noticed that the churned customers have some similar services, which means that any telecom company can detect the predictors and retain their customers. The paper concluded that telecom operators can get best predictive models if they analyzed their whole records and tracked the customers' behavior so they can build different marketing approaches to retain the churners based on the predictors which can be detected when analyzing the historical customer's records. All churn prediction models in this paper can be used in other customer response models as well, such as cross-selling, up-selling, or customer acquisition.

Acknowledgements

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

[1] John, T., *et al.* (2018) Telecom Churn.

[2] Ahmad, A.K., Jafar, A. and Aljoumaa, K. (2019) Customer Churn Prediction in Telecom Using Machine Learning in Big Data Platform. *Journal of Big Data*, **6**, 28. https://doi.org/10.1186/s40537-019-0191-6

[3] Andrews, R., *et al.* (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

[4] ApurvaSree, G., *et al.* (2019) Churn Prediction in Telecom Using Classification Algorithms. *International Journal of Scientific Research and Engineering Development*, **5**, 19-28.

[5] Tata Tele Business Services (2018) Big Data and the Telecom Industry.

[6] Kayaalp, F. (2017) Review of Customer Churn Analysis Studies in Telecommunications Industry. *Karaelmas Science Engineering Journal*, **7**, 696-705.

[7] Umayaparvathi, V. and Iyakutti, K. (2016) A Survey on Customer Churn Prediction in Telecom Industry: Datasets, Methods and Metrics. *International Research Journal of Engineering and Technology*, **3**, 1065-1070.

[8] Kaur, S. (2017) Literature Review of Data Mining Techniques in Customer Churn Prediction for Telecommunications Industry. *Journal of Applied Technology and Innovation*, **1**, 28-40.

[9] Ahmed, A. and Linen, D.M. (2017) A Review and Analysis of Churn Prediction Methods for Customer Retention in Telecom Industries. *4th International Conference on Advanced Computing and Communication Systems*, Coimbatore, 6-7 January 2017, 1-7. https://doi.org/10.1109/ICACCS.2017.8014605

[10] Amin, A., *et al.* (2019) Customer Churn Prediction in Telecommunication Industry Using Data Certainty. *Journal of Business Research*, **94**, 290-301. https://doi.org/10.1016/j.jbusres.2018.03.003

[11] Saraswat, S. and Tiwari, A. (2018) A New Approach for Customer Churn Prediction in Telecom Industry. *International Journal of Computer Applications*, **181**, 40-46. https://doi.org/10.5120/ijca2018917698

[12] Ahmed, A.A. and Maheswari, D. (2017) Churn Prediction on Huge Telecom Data Using Hybrid Firefly Based Classification. *Egyptian Informatics Journal*, **18**, 215-220. https://doi.org/10.1016/j.eij.2017.02.002

[13] Kumar, N. and 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.

[14] IBM Waston Dataset 2018-11-29. https://www.kaggle.com/jpacse/datasets-for-churn-telecom

[15] IBM Data. https://www.ibm.com/communities/analytics/watson-analytics-blog/predictive-insights-in-the-telco-customer-churn-data-set

[16] Cell2cell Dataset. https://www.kaggle.com/jpacse/telecom-churn-new-cell2cell-dataset

[17] Business F.S.O. (2002) Cell2cell: The Churn Game. ((A) 8/26/02).

[18] Mitchell, T.M. (2015) Generative and Discriminative Classifiers: Naive Bayes and Logistic Regression.

[19] Han, J., Pei, J. and Kamber, M. (2011) Data Mining: Concepts and Techniques. El-

sevier, Amsterdam.

[20] Hastie, T., Tibshirani, R. and Friedman, J. (2008) The Elements of Statistical Learning: Data Mining, Inference and Prediction. Springer, Berlin.

[21] Induja, S. and Eswaramurthy, D.V.P. (2016) Customers Churn Prediction and Attribute Selection in Telecom Industry Using Kernelized Extreme Learning Machine and Bat Algorithms. *International Journal of Science and Research*, **5**, 258-265.

[22] Gajowniczek, K., Orłowski, A. and Ząbkowski, T. (2016) Entropy Based Trees to Support Decision Making for Customer Churn Management. *Acta Physica Polonica A*, **129**, 971-979. https://doi.org/10.12693/APhysPolA.129.971

[23] Idris, A., Iftikhar, A. and ur Rehman, Z.J.C.C. (2017) Intelligent Churn Prediction for Telecom Using GP-AdaBoost Learning and PSO Undersampling. Springer Science + Business Media, Berlin, 1-15. https://doi.org/10.1007/s10586-017-1154-3

[24] Gajowniczek, K., Ząbkowski, T. and Orłowski, A. (2015) Comparison of Decision Trees with Rényi and Tsallis Entropy Applied for Imbalanced Churn Dataset. *Federated Conference on Computer Science and Information Systems*, Lodz, 13-16 September 2015, 39-44. https://doi.org/10.15439/2015F121

[25] Maldonado, S., *et al.* (2015) Profit-Based Feature Selection Using Support Vector Machines—General Framework and an Application for Customer Retention. *Applied Soft Computing*, **35**, 740-748. https://doi.org/10.1016/j.asoc.2015.05.058