STUDYING MACHINE LEARNING ALGORITHMS FOR CUSTOMER CHURN PREDICTION

PREPARED BY

MD. FORHAD HOSSAIN
ID: 161-15-846

MD. UMAID HASAN
ID: 161-15-921

MD. FAHAD HOSEN
ID: 161-15-942

This Report Demonstrated in some Fruition of the Capital Requirements for the Bachelor of Science Degree in Computer Science and Engineering.

Supervised By
Dr. S. M. Aminul Haque
Associate Professor
Department of CSE
Daffodil International University

Co-Supervised By
Ms. Amatul Bushra Akhi
Lecturer
Department of CSE
Daffodil International University

DAFFODIL INTERNATIONAL UNIVERSITY

Dhaka, Bangladesh

December 2019

APPROVAL

This Project titled "**Studying Machine Learning Algorithms for Customer Churn Prediction**", submitted by **Md. Forhad Hossain**, Id: 161-15-846, **Md. Umaid Hasan**, Id: 161-15-921, **Md. Fahad Hosen**, Id: 161-15-942 to the Department of Computer Science and Engineering, Daffodil International University, has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of B.Sc. in Computer Science and Engineering and approved as to its style and contents. The presentation has been held on 26 November 2019.

BOARD OF EXAMINERS

Dr. Syed Akhter Hossain — **Chairman**
Professor and Head
Department of CSE
Faculty of Science & Information Technology
Daffodil International University

Dr. S. M. Aminul Haque — **Internal Examiner**
Associate Professor
Department of CSE
Faculty of Science & Information Technology
Daffodil International University

Saif Mahmud Parvez — **Internal Examiner**
Lecturer
Department of CSE
Faculty of Science & Information Technology
Daffodil International University

(Name) — **External Examiner**
Designation
Department of -------
Jahangirnagar University

DECLARATION

We hereby declare that this project has been done by us under the supervision of **Dr. S M. Aminul Haque, Associate Professor, Department of CSE** Daffodil International University. We also declare that neither this project nor any part of this project has been submitted elsewhere for the award of any degree or diploma.

Supervised by:

____

Dr. S. M. Aminul Haque
Associate Professor
Department of CSE
Daffodil International University

Co-Supervised by:

____

Ms. Amatul Bushra Akhi
Lecturer
Department of CSE
Daffodil International University

Submitted by:

____

Md. Forhad Hossain
ID: 161-15-846
Department of CSE
Daffodil International University

____

Md. Umaid Hasan
ID: 161-15-921
Department of CSE
Daffodil International University

____

Md. Fahad Hosen
ID: 161-15-942
Department of CSE
Daffodil International University

ACKNOWLEDGEMENT

First, we express our heartiest thanks and gratefulness to Almighty God for His divine blessing makes us possible to complete the final year project successfully.

We really grateful and wish our profound our indebtedness to **Dr. S. M. Aminul Haque**, **Associate Professor**, Department of CSE Daffodil International University, Dhaka. Deep Knowledge & keen interest of our supervisor in the field of "*Data Mining*" to carry out this project. His endless patience, scholarly guidance, continual encouragement, constant and energetic supervision, constructive criticism, valuable advice, reading many inferior drafts and correcting them at all stages have made it possible to complete this project.

We would like to express our heartiest gratitude to, **Dr. Syed Akhter Hossain**, **Professor and Head** along with **Ms. Amatul Bushra Akhi, Lecturer, Mehedi Hasan Abid, Lecturer, Mohd. Saifuzzaman, Lecturer** for their kind help to finish our project and also to other faculty members and the staff of CSE department of Daffodil International University.

We would like to thank our entire course mate at Daffodil International University, who took part in this discussion while completing the course work.

Finally, we must acknowledge with due respect the constant support and patience of our parents.

ABSTRACT

In this new era, customer relationship management is a challenging deed in the telecommunications industry because this is a profoundly competitive sector and continually challenged by customer churn. For predicting out the customer churn accurately, this article represents a comparative analysis among the most prevalent machine learning techniques. The first step to handle the challenging issue of a customer churn prediction is the uses of Data Mining and Machine Learning tools. Feature Engineering along with widely utilized classification methods such as (DT) Decision Tree, ANN (Artificial Neural Network) and SVM (Support Vector Machine), is implemented on a public domain telecoms dataset. After the main phase, this analysis finds out the best overall classifier using Accuracy, Precision, Support, Recall, F-measure, which is determined from the substance of the Confusion Matrix.

Keywords: Customer Churn Analysis, Data Mining, Data Analysis, Feature Engineering, Machine Learning Techniques (SVM, ANN, DT).

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

1.1 Introduction

We know that during this era Customer Relationship, management is a troublesome deed and anticipating out the clarification of Customer Churn is a major part of Data Mining. Given the significance of customers because of the most beneficial assets of organizations, client retention looks to be a basic demand for any organization. Then a matter might arise in mind what is Customer Churn?

Customer churn is the first imperative measurements for a growing business to determine. This is a malignant measure because of providing the arduous truth concerning its client retention to a company. Customer churn is the proportion of shoppers that stopped victimization a company's product or service throughout a precise period. Churn models find out churning sign and acknowledge customers with a raised possibility to depart willfully.

Therefore, the fundamental destinations are customer retention because the cost for customer securing is route bigger than the cost of customer retention. A few firms endure the substantial downside of customer abandonment, thanks to low-level client satisfaction, dynamic market conditions, aggressive competitive methods, fierce competition ensuing from saturated markets, new merchandise, rules and nonstop presentations of recent competitive contributions. Emphasizing the upper prices related to attracting new customers compared with holding existing customers, and consequently, the certain reality that long clients will, in general, provide additional profits [1], asserts that client retention will increase profitability. Several competitive organizations have obtained that a key strategy for survival at intervals the business is to retain existing customers. Churn rate is calculated by dividing the number of consumers lost throughout that point amount by the number of consumers had at the start of that point amount. For instance, if a company begins with 500 customers and finishes with 475 customers churn rate is 5% because of lost 5% of customers. A company ought to aim for a churn rate that is as close to 0% is attainable. It has emerged a curiosity why it is a necessity to calculate churn rate. Normally, 5% does not sound appalling, right? An ascent in client retention of just 5% will deliver at least a 25% increment in profit. Accordingly, a company will pay less on the operative prices of getting to amass new customers.

1.2 Motivation

Managing customer churn and prediction is a vital concern for the powerfully competitive telecommunication company. A telecom company needs to retain its endorsers has to be able to predict that of them could also be at risk of changing services and can create those subscribers the main focus of customer maintenance endeavors [2]. In addition, the literature proposes that little alteration within the retention rate may end up an important impact on business [3]. For dealing with customer churn, it is basic to make an increasingly reasonable and correct client churn prediction model. Machine Learning techniques are used for predicting churn. This analysis underpins the previous errand, which infers that it intends unexpectedly an approach to deal with data processing techniques to assists telecommunication churn management and this article presents a comparative study on the foremost efficient machine learning techniques among ANN, SVM and Decision Tree for the difficult downside of a client churning prediction within the telecommunications company and evaluate the accurate model from keeping Customer's information for predicting churn and to stint the customer's turnover. It will offer the method a way to scale back customer churn. In this way, the telecommunication operators will improve the competitive edge.

1.3 Research Questions

- What is the main problem in the Business Sector?
- What is Customer Churn?
- How to measure customer churn?
- Is customer churn prediction taking a significant place in data mining?
- Can we solve the churn rate prediction by machine learning technique?
- Which machine-learning algorithm is providing the best prediction accuracy?

1.4 Expected Outcome

Customer Churning is the main problem in the business sector, customer retention is the main issue in this new business era. Therefore, every enterprise looks after the best prediction method by following the machine-learning algorithm. Customer churn prediction can be solved by using a machine-learning algorithm. Our utilized algorithms are Artificial Neural Network, Decision tree, Support Vector Machine and we use the confusion matrix component to predict the best classifier.

1.5 Report Layout

Before start anything we need introduction and motivation and our paper chapter 1 we discuss about introduction and our motivation on this topic which is customer churn prediction and why we chose it. Chapter 2 covers the related working area. When we complete our introduction, we focused on those worked what is related with our topics collect their valuable information. Basically, our 2 chapter normally discussed about related work. Consequence In chapter 3 we discuss about implement algorithm, we chose famous 3 algorithm's (DT, ANN, SVM) and applied those algorithm's in our dataset to searching best one and it's our methodology part. Chapter 4 we talk about all of our algorithm's outcome and evaluation. Finally, Chapter 5 we got our distinguish result which can say the comparative best one. It's our conclusion part. At last, we show references from where we collect our information.

CHAPTER 2

LITERATURE REVIEW

2.1 Introduction

In the course of the most recent decade, there has been expanding enthusiasm for pertinent investigations in areas together with telecommunication[4, 5], banking[6], insurance enterprises [7], the diversion [8] and so on. A few, extremely popular in analysis community Machine Learning Models are anticipated to tackle the churning prediction downside. The SVM approach has been numerous business applications as of late, primarily within the space of your time series expectation and classification[9], promoting[10], bankruptcy prediction[11], credit rating analysis[12].

2.2 Related Works

Prediction and management of Customer churn is a concerning issue for industries; however, it is significantly intense within the strongly focused and currently extensively changed versatile telecommunications trade. For predicting out churn in medium corporations, a couple of approaches were applied. Machine learning and Data Mining is mostly used approaches. For extracting knowledge, the agglomeration of associated work fixated on applying only one methodology of data mining, and for this reason, the others centered to examine many ways to predict churn. To grasp however related work develops their Prediction Models and this article surveys a few of the present related works appeared within the table:

Table 2.2.1. Related studies on Customer Churn

Reference	Used Dataset	Prediction Method and Accuracy
Hung et al [13]	Telecommunication Data	(DT, NN) Clustering (Means)
Bucking & Poel[14]	Retailing Dataset	NN, logistic regression

Burez & Poel[15]	Pay-Tv Company	Random Forests and LR
Gavril Toderean, Horia Beleiu [16]	Call Details Record (21 Features with 3333 clients)	Data Mining, NN (99.55%), SVM (99.70%), and Bayes Networks (99.10%)
Wei et al. [2]	Taiwan Mobile Company	Decision Tree
He et al. [17]	China Telecom and China Netcom (2007)	Radial-Basis Function (RBF), Neural Network (91.1%)
Kim et al. [18]	Korean's Mobile Company	Logistic Regression
Huang et al. [19]	Business and Operation Support department at China's largest Telecom Company.	Random Forest, Evaluated by AUC.
Makhtar et al. [20]	Dataset of Local Telecommunication Company's billing period	Rough Set Theory, Experiments were conducted using different split factors and reduction methods using Rough Set Technical Analysis Toolkit.
Idris et al. [21]	Tested on two data sets (Orange Telecom and cell2cell)	Genetic Programming (GP) based approach, Adaboost. cell2cell dataset (89% accuracy) and Orange Telecom (63% Accuracy)
Burez and Van den Poel [22]	Six real-life proprietary European churn modeling data sets.	Weighted Random Forests, Random Sampling, Gradient Boosting Model.
Sharma et al. [23]	Churn Dataset (UCI Repository)	NN Based approach with an accuracy of more than 92%.

Dalvi et al. [24]	Telecommunication Industry	Decision Trees, Logistic Regression
Vafeiadis et al. [25]	Public Domain Dataset.	SVM-POLY with AdaBoost (Accuracy 97%)
Backiel A et al. [26]	A Telecom Dataset of 1.4 million customers with millions of calls record each month.	RST-based feature reduction algorithm

2.3 Research Summary

As per following the literature review concerning customer churn, the greater part of the associated work centers around victimization only one information processing technique like classification or cluster for extracting out the information of customer retention. We did not notice any analysis inquisitive about the downside that provides a comparative analysis of three Machine Learning algorithms in any telecommunication enterprise. The vast majority of the previous research article did not perform this entire machine learning algorithmic program for a particular dataset. In this paper, the methodology and the result section are taken into thought to make a comparison of three machine-learning algorithms. We tend to prepare the information and compared the results of three machine learning algorithms.

This article proposes an Artificial Neural Network (ANN) primarily based approach to predict client churn in subscription of telecom services.

2.4 Scope & Challenges

After getting this data set, firstly we check manually if there is any missing data or irrelevant data. We found some missing data rows and this is our first challenge to fill-up that data rows. Then we convert our dataset into numerical format because of SVM, ANN needs numerical values. That was our challenge. Our main scope is no one works with these three algorithms at the same dataset.

CHAPTER 3

METHODOLOGY

3.1 Introduction

In research work, the main part is the analytical part and it is totally dependent on evaluation or developing an algorithm. As our work is with a dataset of a business sector and the dataset comes from a wireless mobile company, we have to follow some procedure for finishing out the developing or implementing process.

3.2 Research Subject and Instrumentation

Our research work is Predicting out the customer churn of a telecommunication company. For this work, we have to implement machine learning techniques like SVM, DT or ANN. Most of the Machine-learning algorithms work efficiently with data, and for taking a great measurement, we also fed our algorithm an imbalanced dataset on Customer Churning of a telecommunication company.

3.3 Working Procedure Flowchart

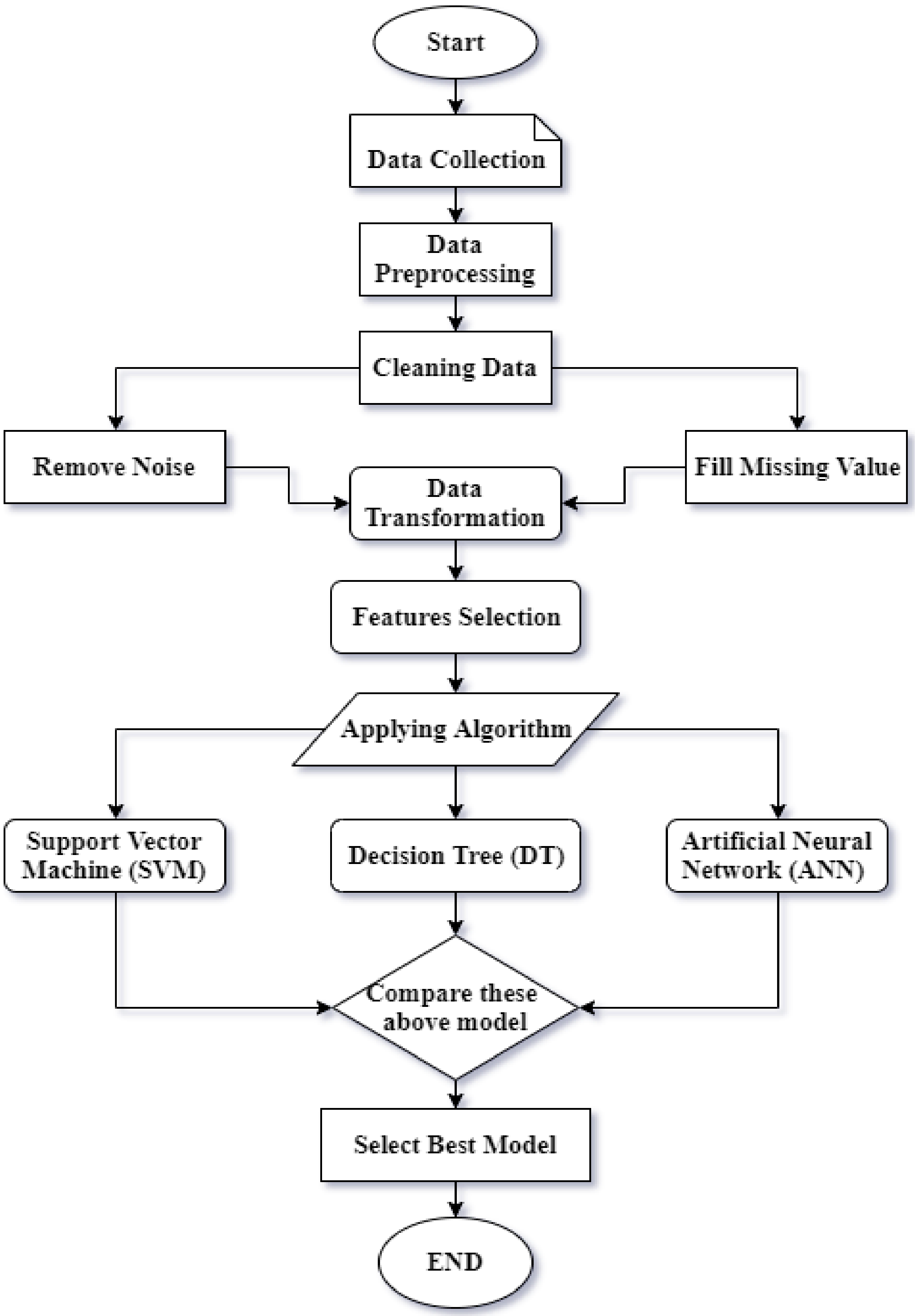

Fig. 3.3.1 Working procedure flowchart

3.4 Data source and collection

Korean Telecom Company provides its data about customer for protecting customer churn. This Dataset provides information about 7043 customers with 21 features.

3.5 Data preprocessing

Data preparation is the main process of data mining to get a piece of accurate knowledge. Noise, inconsistent values, missing values may be contained by a dataset. For removing this, data preprocessing is done by following data cleaning, data integrations, data transformation and data reduction process. There are some missing values in the TotalCharges column and we handled this by the MEAN method.

After that, we convert the data set from string to numeric, as we know that SVM and ANN can't handle string value. We have set the 'Gender' column's attribute 0 and 1 for Male and Female respectively, all yes and no are represented as 1 and 0 gradually, Internet Service column has two attributes DSL and Fiber optic which we have represented as 0 and 1 respectively, Contract column has 3 attribute Month-to-month, one year, two years which we converted as 1,2,3 sequentially. Payment method column has 4 at-tributes 'Electronic Check', 'Bank Transfer', 'Mailed Check', 'Credit Card' which we represented as 1, 2, 3,4 respectively.

3.5.1 Feature Engineering

After having fitted our dataset, we can get the classifier's coefficients using the trained model. Feature selection is done after differentiating the size of these coefficients to each other. Decreasing the number of features from tremendous datasets in Machine Learning is a noteworthy activity. This can be told to avoid overfitting, speed up training and ultimately provide better classification results. By looking at the coefficients, it is, therefore, possible to identify the main features used in classification and discard which hold less contrast. From the dataset, we find these features from the 21 features and the target class is "Churn", which is holding "yes" (means customer churned) or "no" (means, the customer did not churn).

3.5.2 Criteria for selection

Correlation refers to having a linear relationship between two variables. The correlation coefficient is evaluated within -1 to 1. Values present correlation like 0 represent weaker (exact 0 implies no correlation), near to 1 and value near at -1 implies respectively positive and negative correlation. High correlation represent linearly dependency between features along with providing same effect on the dependent variable. We can remove one of the features after finding two features which have high correlation.

```
[0.38784438 0.02393747 0.01870484 0.0243714  0.02063373 0.35036036
 0.00346798 0.02390097 0.06919206 0.0775868 ]
```

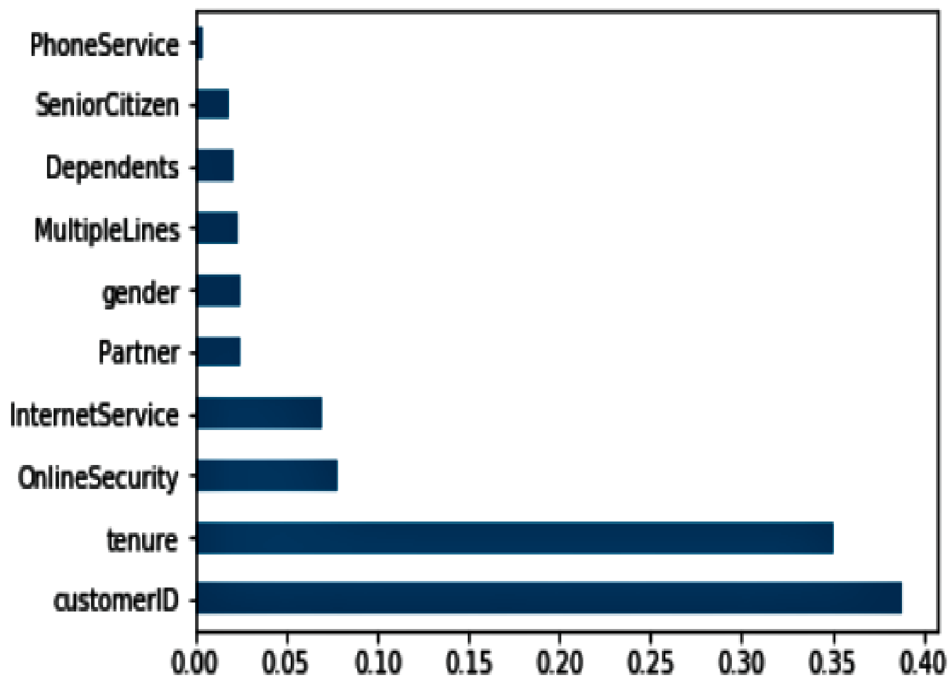

Fig. 3.5.2.1 Features Importance (constructed by ExtraTreesClassifier)

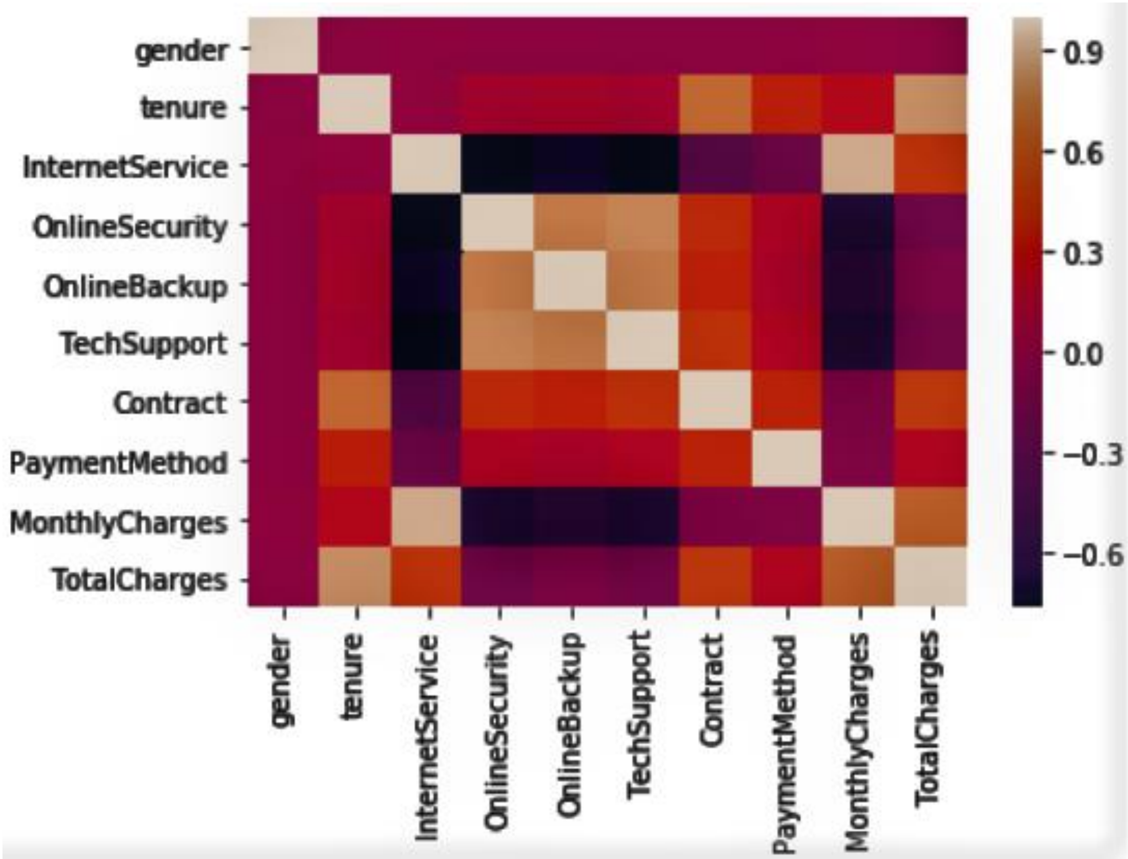

Fig. 3.5.2.2 Correlation Heatmap for the Important Features

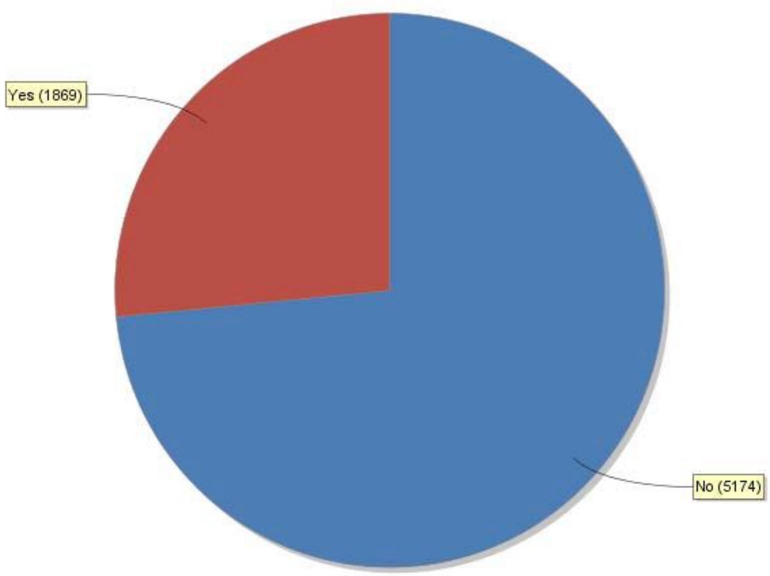

Fig.3.5.2.3 Churners Vs Non-Churners (Blue for Non-Churners)

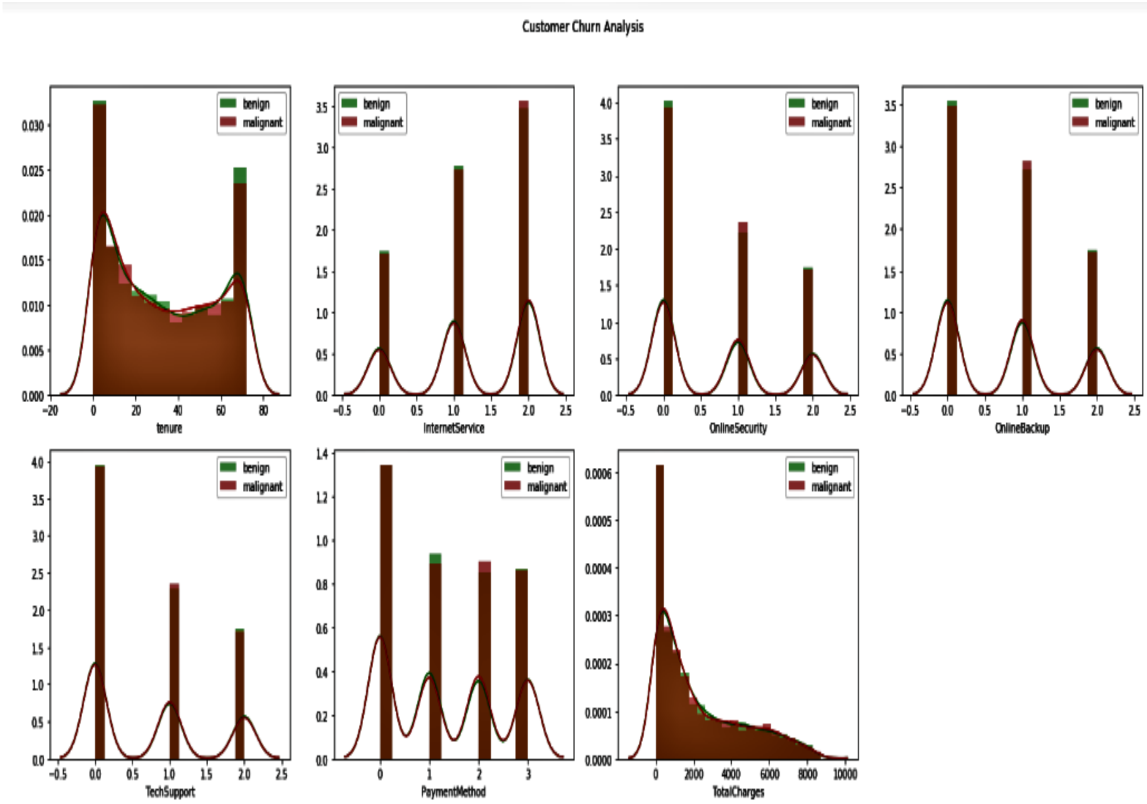

Fig. 3.5.2.4 Distribution plot for the selected features

This figure shows the co-relation with each other according to the class label.

3.6 Implemented Algorithm

This article used a supervised learning technique, which performs based on training and testing. The training dataset constructs the classification model. This constructed model is then implemented on the testing dataset to get the result. The succeeding explanations will quickly depict three machine-learning algorithms.

3.6.1 Artificial Neural Networks (ANN)

For managing intricate issues ANN is an outstanding method, as an instance, the churn prediction issue. NN can be programming based absolutely (computer models) or hard-ware-based totally (neurons are represented by physical segments) and can use an arrangement of topologies and become increasingly acquainted with figuring's. ANN is a coordinated acing framework that worked on an extensive number of primary components, referred to as perceptions or neurons. Every neuron can fit on fundamental decisions and feeds them one's decisions to specific neurons, made in interconnected layers. The neural system can copy essentially any limit, and answer any question, given enough getting tests and accumulate power. Three layers of neurons are:

- ✓ **Input layer** acknowledges the inputs of the model.
- ✓ **Hidden layer**
- ✓ **Output layer** predictions generator.

To train the NN inside the input layer NN are normally deals with some real values and each value is passed through neurons to neurons are considered as a set of inputs for getting the outputs. Neurons in the output layer generate each output value. The input passing system is done through some Common activation features like as:

Sigmoid: input is real-valued and ranges within 0 and 1

TanH: input is real-valued and range [-1, 1].

ReLu: Represents 'Rectified Linear Units'. Input is real-valued and thresholds it to 0 (Negative values replace with 0)

Output values are generated by the Activation features within an adequate assortment. A numeric number is given to every neuron named weight. The weights, collectively with the activation characteristic, outline every neuron's output besides a forward pass passed off. How far the real output of the current model is from the accurate output is defined by 'Blunders Function'. To discover an optimum weight for the neurons, we carry out a backward pass known as backpropagation that is executed with the aid of a mathematical process called gradient descent.

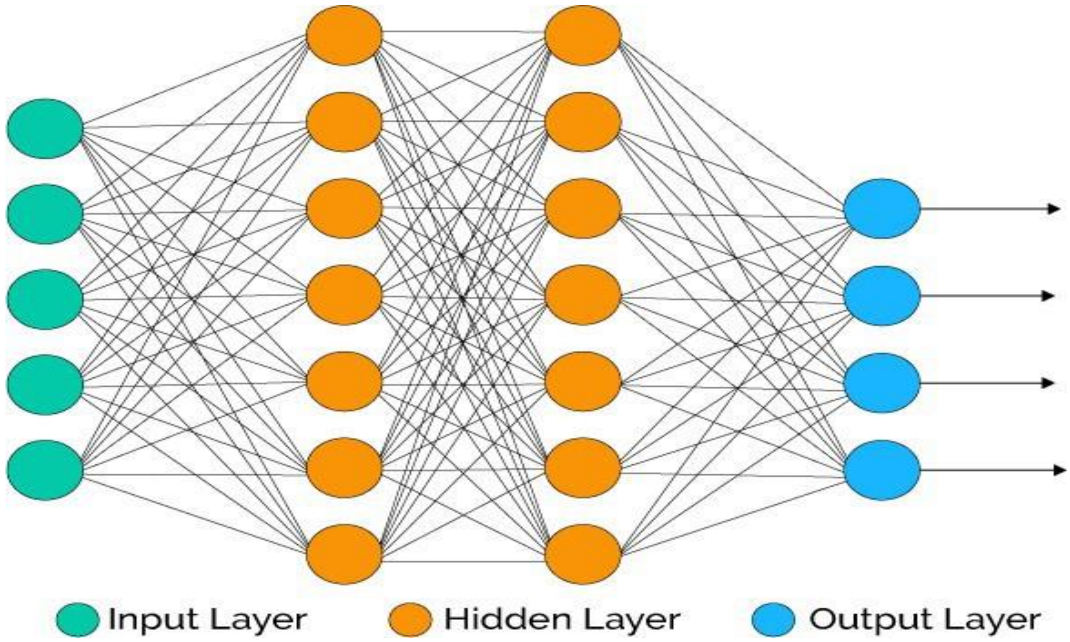

Fig.3.6.1.1. Neural Network

3.6.2 Support Vector Machines (SVM)

SVM are supervised and associated learning models that analyze patterns and information for classification and analysis which is based on structural risk minimization and Kernel functions have been utilized for improving efficiency [27].

The nonlinear relationship among the features is permitted by the Kernel while keeping the calculations in the original input space. Linear Kernel is utilized in any case when the data is linearly distinguishable and it tends to be disengaged using a singular line. It is one of the most widely recognized kernels to be utilized. It is generally utilized when there are a huge number of Features in a specific Data Set. Training an SVM with a Linear Kernel is faster than with any other Kernel.

SVM outperforms ANN and DT, depending mostly on the types of data and transformation of data that takes place among them in the churn forecasting issues [28, 29].

3.6.3 Decision trees (DT)

DT represents sets of decisions which is capable of creating classification rules for a specific dataset [30], where leaves and branches represent class labels and features respectively. Tree-based knowledge like Information Gain, Entropy is utilized from DT. Entropy measures the disorder. High entropy means low-level purity. Entropy is estimated somewhere in the range of 0 and 1.

A ‘DT’ gathers decisions on the best order of features for splitting on and decides when to stop at the time it trains itself on a particular dataset. DT uses Information Gain (ig) and Entropy (e) to decide what feature to part their hubs/nodes on as they are being prepared on a particular dataset. DT has no incredible execution on capturing unpredictable and nonlinear associations between the attributes. However, in the customers' churn issue, the precision of a DT can be contingent and high depending on the information type [31].

CHAPTER 4

EXPERIMENTAL RESULTS AND DISCUSS

4.1 Evaluation of Decision Tree

We construct this tree based on the entropy criteria and we set the max-depth =3 for better visualization. The more the depth will be it will be over fitted and failed to capture useful patterns and testing error will arise a lot.

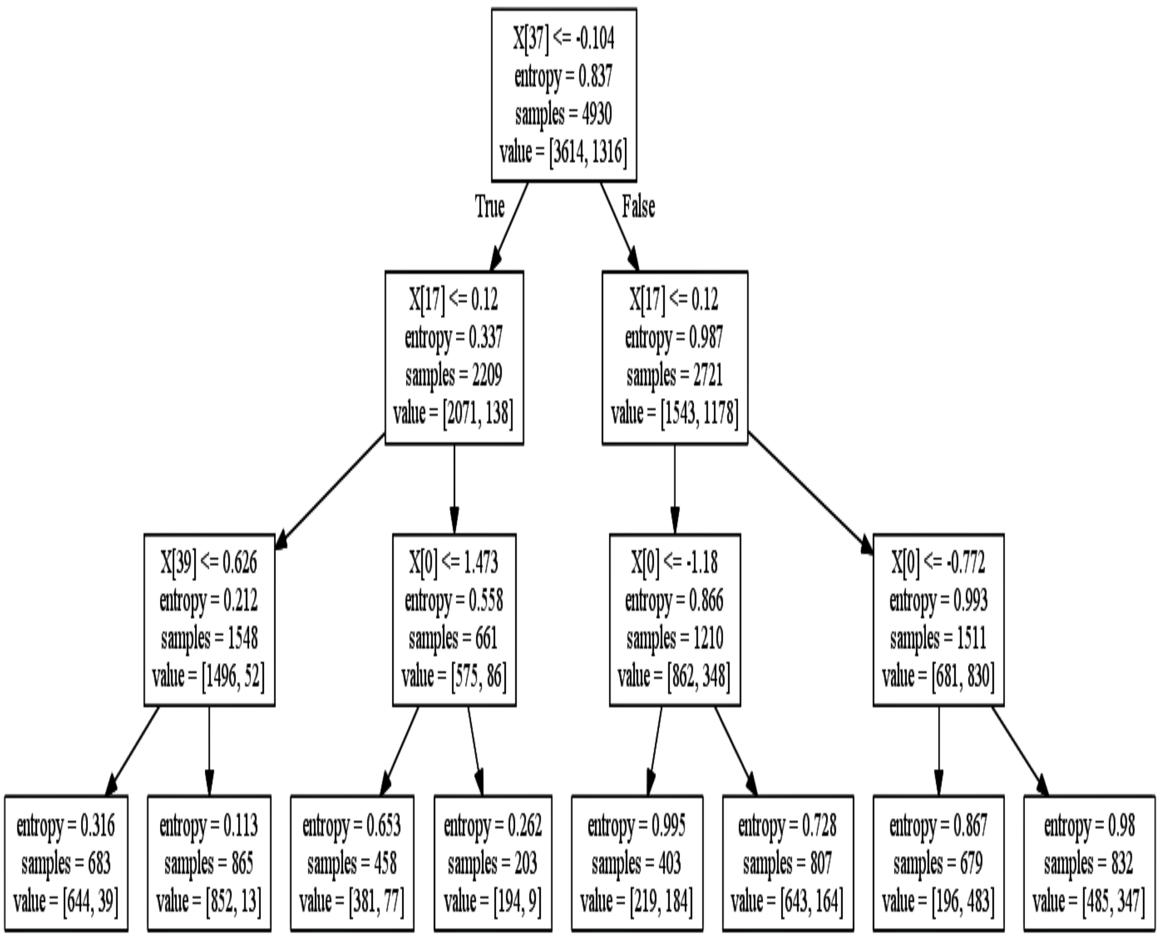

Fig. 4.1.1. Decision tree generated from our dataset (max depth set 3)

4.1.1 Confusion Matrix of Decision Tree

```
[[1470   90]
 [ 363  190]]
              precision    recall  f1-score   support

           0       0.80      0.94      0.87      1560
           1       0.68      0.34      0.46       553

    accuracy                           0.79      2113
   macro avg       0.74      0.64      0.66      2113
weighted avg       0.77      0.79      0.76      2113
```

Fig. 4.1.1.1. Confusion Matrix of Decision Tree

4.2 Evaluation of SVM

Support Vector machine is a supervised machine learning technique and this is broadly evaluated by confusion matrix.

4.2.1 Confusion Matrix of Support Vector Machine

```
[[1404  181]
 [ 228  300]]
              precision    recall  f1-score   support

           0       0.86      0.89      0.87      1585
           1       0.62      0.57      0.59       528

    accuracy                           0.81      2113
   macro avg       0.74      0.73      0.73      2113
weighted avg       0.80      0.81      0.80      2113
```

Fig. 4.2.1.1. Confusion Matrix of Support Vector Machine (SVM)

4.3 Evaluation of ANN

Artificial neural network is evaluated by the training loss and validation on for determining overfitting or underfitting.

4.3.1 Visualizations of Training Loss and Validation Loss

To determine the performance with a human brain, the Visualization of any machine learning model give a sense that whether the data being trained out of the model properly. Visualizing the Overfitting (Validation loss is greater than Training Loss) or Underfitting (Training Loss is greater than Validation loss) is the best way to evaluate the model whether it is sufficiently trained or not. Data is split into Training Set, Validation Set, and Test Set. The model loss represents that the model has comparable (training loss ~ validation loss) performance.

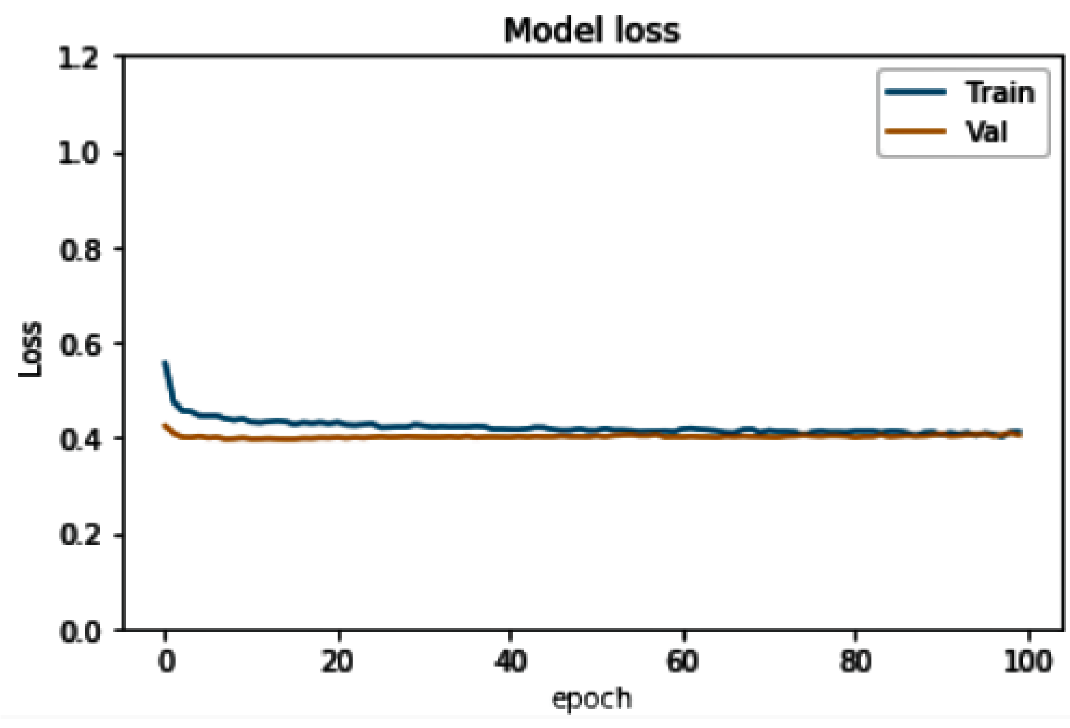

Fig.4.3.1.1. Visualizations of Training Loss and Validation Loss

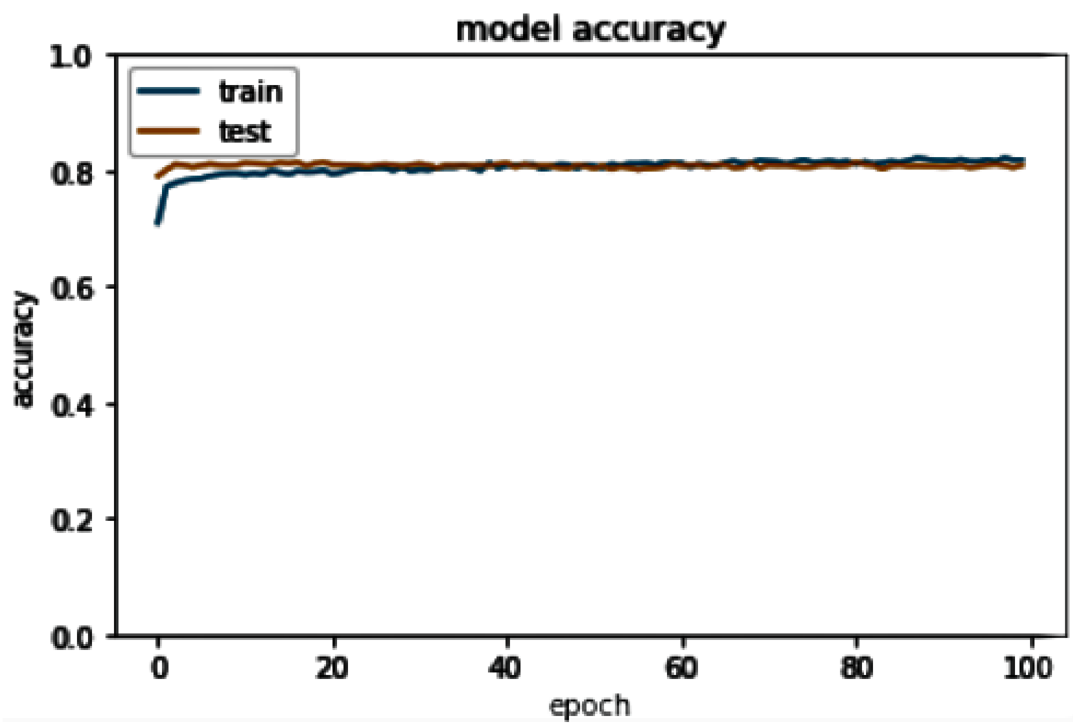

Fig. 4.3.1.2. Visualizations of model accuracy

4.3.2 Confusion Matrix of Artificial Neural Network

```
[[1392  193]
 [ 208  320]]
              precision    recall  f1-score   support

           0       0.87      0.88      0.87      1585
           1       0.62      0.61      0.61       528

    accuracy                           0.81      2113
   macro avg       0.75      0.74      0.74      2113
weighted avg       0.81      0.81      0.81      2113
```

Fig. 4.3.2.1. Confusion Matrix of Artificial Neural Network

4.4 Evaluation measures

To evaluate the performance of the classifiers in churn prediction for various schemes with their appropriate parameters, the proportions of recall, support, accuracy, precision and F1-score which are derived from the substance of the confusion matrix. Cases are denoted as False-Positive (FP) and True-Positive (TP) while False-Negative (FN) and True-Negative (TN)

- Recall = $TP/(TP + FN)$ [Proportion of positive cases]
- Precision = $TP/(TP + FP)$ [Proportion of the anticipated positive cases which is determined from the condition]
- Accuracy = $(TP + TN)/(TP + FP + TN + FN)$ [proportion of the total number of predictions]
- F- Measure = $2 \times Recall \times Precision/(Precision + Recall)$ [Measures the balance between precision and recall]
- We also calculate the cross validation for 10-fold, result is given in Table 2.
- Now considering a table from 2113 trainee data:

Table 4.4.1. Comparison of Model Performance According to Confusion Matrix for 30% Trainee data

Confusion Matrix	Model Name			
	Yes/No	Artificial Neural Network	Support Vector Machine	Decision Tree
precision %	Yes	64%	62%	68%
	No	86%	86%	80%
recall %	Yes	55%	57%	34%
	No	90%	89%	94%
f1-Score%	Yes	59%	59%	46%
	No	88%	87%	87%
support	Yes	528	528	553
	No	1585	1585	1560

accuracy	80.93	80.64	78.56
Accuracy%	81%	81%	79%
Cross Validation (Standard Deviation)	0.20	0.14	0.03
Total Data	7043		
Train Data	4930 (70%)		
Test Data	2113(30%)		

The above table represent that ANN outperformed the other classifier SVM and Decision tree.

CHAPTER 5

CONCLUSION AND FUTURE WORK

5.1 Conclusion and Future Work

For improving the performance of classifiers for customer, churn prediction on Telecommunication Dataset Feature Engineering and Data analysis is very significant. Because of making an accurate churn prediction we need to know which algorithms give, us better performance. In this research paper, we try to discuss about three effective machine-learning techniques (SVM, DT, ANN) and find out the best method. Where we get 79% accuracy for the Decision Tree, 81% accuracy for ANN and SVM also. The experimental evaluation demonstrated that Artificial Neural Network outperformed the other classifiers. Besides SVM is also provides better accuracy very close to ANN. We hope for more progressively refined data mining techniques will be developed as business multifaceted nature increments. In addition, upcoming researchers will also work with churn prediction. This discovering provides some new point of view on the progressing searches for the best classification performance that chase for new, further developed algorithms.

REFERENCES

1. Verbeke, W., et al., Building comprehensible customer churn prediction models with advanced rule induction techniques. Expert systems with applications, 2011. **38**(3): p. 2354-2364.
2. Wei, C.-P. and I.-T. Chiu, Turning telecommunications call details to churn prediction: a data mining approach. Expert systems with applications, 2002. **23**(2): p. 103-112.
3. Van den Poel, D. and B. Lariviere, Customer attrition analysis for financial services using proportional hazard models. European journal of operational research, 2004. **157**(1): p. 196-217.
4. Jadhav, R.J. and U.T. Pawar, Churn prediction in telecommunication using data mining technology. International Journal of Advanced Computer Science and Applications, 2011. **2**(2).
5. Radosavljevik, D., P. van der Putten, and K.K. Larsen, The impact of experimental setup in prepaid churn prediction for mobile telecommunications: What to predict, for whom and does the customer experience matter? Trans. MLDM, 2010. **3**(2): p. 80-99.
6. Eichinger, F., D.D. Nauck, and F. Klawonn. Sequence mining for customer behaviour predictions in telecommunications. in Proceedings of the Workshop on Practical Data Mining at ECML/PKDD. 2006.
7. Günther, C.-C., et al., Modelling and predicting customer churn from an insurance company. Scandinavian Actuarial Journal, 2014. **2014**(1): p. 58-71.
8. Kawale, J., A. Pal, and J. Srivastava. Churn prediction in MMORPGs: A social influence based approach. in 2009 International Conference on Computational Science and Engineering. 2009. IEEE.
9. Huang, W., Y. Nakamori, and S.-Y. Wang, Forecasting stock market movement direction with support vector machine. Computers & operations research, 2005. **32**(10): p. 2513-2522.
10. Ben-David, S. and M. Lindenbaum, Learning distributions by their density levels: A paradigm for learning without a teacher. Journal of Computer and System Sciences,

1997. **55**(1): p. 171-182.

11. Shin, K.-S., T.S. Lee, and H.-j. Kim, An application of support vector machines in bankruptcy prediction model. Expert systems with applications, 2005. **28**(1): p. 127-135.

12. Huang, Z., et al., Credit rating analysis with support vector machines and neural networks: a market comparative study. Decision support systems, 2004. **37**(4): p. 543-558.

13. Hung, S.-Y., D.C. Yen, and H.-Y. Wang, Applying data mining to telecom churn management. Expert Systems with Applications, 2006. **31**(3): p. 515-524.

14. Buckinx, W. and D. Van den Poel, Customer base analysis: partial defection of behaviourally loyal clients in a non-contractual FMCG retail setting. European Journal of Operational Research, 2005. **164**(1): p. 252-268.

15. Burez, J. and D. Van den Poel, CRM at a pay-TV company: Using analytical models to reduce customer attrition by targeted marketing for subscription services. Expert Systems with Applications, 2007. **32**(2): p. 277-288.

16. Brânduşoiu, I., G. Toderean, and H. Beleiu. Methods for churn prediction in the prepaid mobile telecommunications industry. in 2016 International conference on communications (COMM). 2016. IEEE.

17. He, Y., Z. He, and D. Zhang. A study on prediction of customer churn in fixed communication network based on data mining. in 2009 Sixth International Conference on Fuzzy Systems and Knowledge Discovery. 2009. IEEE.

18. Kim, H.-S. and C.-H. Yoon, Determinants of subscriber churn and customer loyalty in the Korean mobile telephony market. Telecommunications policy, 2004. **28**(9-10): p. 751-765.

19. Huang, Y., et al. Telco churn prediction with big data. in Proceedings of the 2015 ACM SIGMOD international conference on management of data. 2015. ACM.

20. Makhtar, M., et al., Churn classification model for local telecommunication company based on rough set theory. Journal of Fundamental and Applied Sciences, 2017. **9**(6S): p. 854-868.

21. Idris, A., A. Khan, and Y.S. Lee. Genetic programming and adaboosting based churn prediction for telecom. in 2012 IEEE International Conference on Systems, Man, and Cybernetics (SMC). 2012. IEEE.

22. Burez, J. and D. Van den Poel, Handling class imbalance in customer churn prediction. Expert Systems with Applications, 2009. **36**(3): p. 4626-4636.

23. Sharma, A., D. Panigrahi, and P. Kumar, A neural network based approach for predicting customer churn in cellular network services. arXiv preprint arXiv:1309.3945, 2013.

24. Dalvi, P.K., et al. Analysis of customer churn prediction in telecom industry using decision trees and logistic regression. in 2016 Symposium on Colossal Data Analysis and Networking (CDAN). 2016. IEEE.

25. Vafeiadis, T., et al., A comparison of machine learning techniques for customer churn prediction. Simulation Modelling Practice and Theory, 2015. **55**: p. 1-9.

26. Backiel, A., B. Baesens, and G. Claeskens, Predicting time-to-churn of prepaid mobile telephone customers using social network analysis. Journal of the Operational Research Society, 2016. **67**(9): p. 1135-1145.

27. Kirui, C., et al., Predicting customer churn in mobile telephony industry using probabilistic classifiers in data mining. International Journal of Computer Science Issues (IJCSI), 2013. **10**(2 Part 1): p. 165.

28. Hur, Y. and S. Lim. Customer churning prediction using support vector machines in online auto insurance service. in International Symposium on Neural Networks. 2005. Springer.

29. Shaaban, E., et al., A proposed churn prediction model. 2012.

30. Jun Lee, S. and K. Siau, A review of data mining techniques. Industrial Management & Data Systems, 2001. **101**(1): p. 41-46.

31. Hadden, J., et al., Churn prediction: Does technology matter. International Journal of Intelligent Technology, 2006. **1**(2): p. 104-110.