Major Project Report on

A Comparative Study on Machine Learning Algorithms for Customer Churn Analytics with Power BI Dashboard

Submitted By:

Sadgi Sharma

2K18/MBA/921

Under the Guidance of:

Dr. Gaganmeet Kaur Awal

Assistant Professor, USME, DTU

UNIVERSITY SCHOOL OF MANAGEMENT

& ENTREPRENEURSHIP

Delhi Technological University

MAY 2020

CERTIFICATE

This is to certify that the project report titled **"A Comparative Study on Machine Learning Algorithms for Customer Churn Analytics with Power BI Dashboard"** is a genuine work carried out by **Ms. Sadgi Sharma** of **MBA Business Analytics, 2018-20** submitted to University School of Management and Entrepreneurship, Delhi Technological University, Bawana Road, Delhi for partial fulfillment of the requirement for the award of the Degree of Masters of Business Administration.

Signature of Guide   Signature of Head (USME)

Place: U.S.M.E, DTU   Seal Head

Date: 22nd May 2020

DECLARATION

I, **SADGI SHARMA**, student of **MBA Business Analytics, University School of Management and Entrepreneurship, Delhi Technological University** hereby declare that the major project report titled "**A Comparative Study on Machine Learning Algorithms for Customer Churn Analytics with Power BI Dashboard**" that is submitted by me to **University School of Management and Entrepreneurship, Delhi Technological University** in partial fulfillment of requirement for the award of the degree of Master of Business Administration in 2018-20, has not been previously formed the basis for the award of any degree, diploma or other similar title or recognition.

NAME – SADGI SHARMA
SIGNATURE

ACKNOWLEDGEMENT

I would like to extend my gratitude to my guide Dr. Gaganmeet Kaur Awal, Assistant Professor at the esteemed Management Department of USME. The project report became possible because of her sheer support and guidance throughout the project duration.

I present to you my report on the topic “Comparative Study on Machine Learning Algorithms for Customer Churn Analytics with Power BI Dashboard”. I also take this opportunity to express regards to my guide for her guidance, monitoring, and constant encouragement throughout the course of his project. The timely guidance given by her has been very valuable. It gives me immense pleasure putting together this set of facts and information, studying in detail about the concerned topic and chalking it down in the form of this report.

SADGI SHARMA

2K18/MBA/921

ABSTRACT

With the increase in competition and technological advancements in the Telecommunications Industry, being able to assess and predict the customer attribution ahead of time is of extreme importance to a company. This Research project describes the application of Machine Learning and Neural Network techniques in R and Python to predict customer churn.

Customer churn is considered to be the major problem in the telecommunications industry. Numerous studies have shown that attracting new customers in a telecommunications industry is much more expensive than retaining existing ones. Therefore, companies are focusing on developing accurate and reliable predictive models to identify potential customers that will churn in the near future. The aim of this project is investigating the main reasons for churn telecommunications sector. This major project report provides a comparative study on different machine learning techniques used for predicting customer churn. The study also covers different phases including business scenario, data analysis, data pre-processing, and implementation of different algorithms for classification.

The resulted models have outperformed the evaluation metrics of different research papers published in the similar domain. For the businesses that offer subscription-based services, it is important to predict customer churn as well as elucidate the parameters related to attrition. The predictive analytics-based techniques like logistic regression can be not as much of precise than new techniques such as machine learning and ensemble models. This report discusses using Artificial Neural Network and Ensemble Models in R and Python to predict the customer churn.

TABLE OF CONTENTS

Certificate i

Declaration ii

Acknowledgement iii

Abstract iv

Table of Contents v

List of Figures and Tables vi

CHAPTER 1 Introduction 8

1.1 Statement of the Problem 9
1.2 Purpose of the Study 11
1.3 Structure of the Study 13

CHAPTER 2 Literature Review 14

2.1 Customer Churn 14
2.2 Prediction Techniques 15

CHAPTER 3 Research Methodology 16

3.1 Data Description 16
3.2 Data Analysis 23
3.3 Data Pre-Processing 24
3.3.1 Pruning 24
3.3.2 Train-Test Split 24
3.4 Data Transformation 25
3.4.1 Feature Discretization 25
3.4.2 One-Hot Encoding 25
3.4.3 Feature Scaling 25
3.4.4 SMOTE 25
3.5 Model Development 26
3.6 Power BI Dashboard 29

CHAPTER 4 Results and Recommendations 38

4.1 Machine Learning Approaches 38
4.2 Evaluation Metrics 41
4.3 Exploratory Data Analysis 45
4.4 Recommendations 50

CHAPTER 5 Limitations and Conclusion 52

References 54
Plagiarism Report 56

LIST OF FIGURES AND TABLES

Sr. No.	Description	Page Number
Fig 1.1	Telecommunications Industry	9
Fig 1.2	Customer Churn	12
Fig 2.1	Types of Classifiers with Cluster Combination	16
Fig 2.2	Self-Organizing Maps	16
Fig 3.1	Framework for Experiment Conducted	22
Fig 3.2	Data View in Python	22
Fig 3.3	CSV File	23
Fig 3.4	Summary in R	23
Fig 3.5	Column Values in Python	24
Fig 3.6	One Hot Encoding of Categorical Variables	25
Fig 3.7	Feature Scaling in Python	25
Fig 3.8	SMOTE Implementation	26
Fig 3.9	ANN Model	26
Fig 3.10	MLP Model	27
Fig 3.11	Mathematical Depiction of ANN	27
Table 3.1	Power BI Term Description	30
Fig 3.12	Power BI Architecture	30
Fig 3.13	Gender Distribution	31
Fig 3.14	Senior Citizen Pie Chart	32
Fig. 3.15	Distribution of customers with Dependents and Partners	32
Fig 3.16	Churn Rate	33
Fig 3.17	Churn by Contract Type	34
Fig 3.18	Churn by Phone service and MultipleLines	35
Fig 3.19	Churn Count by Payment Mode	35
Fig 3.20	Churn Vs Tenure Group in Months	36
Fig 3.21	Power BI Dashboard for other features	36

Fig 3.22	Customer Churn Web Dashboard	37
Fig. 4.1	Decision Tree	39
Fig. 4.2	Rank-wise importance of features for RF Classifier	40
Fig. 4.3	ROC curve for ANN model	42
Fig 4.4	TP, FP, TN, FN	42
Table 4.1	Confusion Matrix	43
Table 4.2	Comparison of Evaluation Metrics for Different Models	44
Fig 4.5	Correlation among variables in R	46
Fig 4.6	Correlation Visualization in Python	47
Fig 4.7	Exploratory Data Analysis	49

CHAPTER 1

INTRODUCTION

Business Analytics is the thorough analysis of technologies, skills, practices and continuous evaluation of the business performance to come up with new improvements and provide insights on how to proceed resourcefully. Business analysis attempts to interpret the business performance based on data and statistical methods. The future decisions regarding the businesses come after a successful business analysis. Business analytics can be sub-divided into two components namely business intelligence and statistical analysis. Both are the underlying concepts that are used to determine customer churn for an online business.

1. Business Intelligence covers the research on past data to evaluate the performance score of the management team, staff, and the business department. This helps the startups/established companies to relate the efforts of the company with its outcomes. This further carries out the area of focus and improvement. It does not include heavy predictive modelling or forecasting.
2. Statistical Analysis involves the use of predictive analysis using statistical methods. It is useful in figuring out the future performance of the product. Customer Churn Analysis also uses statistical analysis techniques based on cluster analysis in order to develop group of customers on the basis of similar characteristics so that the marketing campaigns can be aimed at achieving set goals. Customer Churn is also known as Customer Attrition and refers to a situation when a client or a subscriber ceases his or her affiliation with the business. Customer may vary from industry to industry for example for a Gaming Industry, customer is a player, for a Telecom Industry, customer is a subscriber, for a Product company, it may be a user. It has been found that in most Online Businesses, typically a customer drops using a service once a considerable amount of time has passed. This costs company and it becomes important to replace the churned customer by acquiring new customers. Minimizing the customer churn is the key business goal of every company specially a company that mostly does its business online.

The cost incurred to get the customer on board initially is also booked as a loss in revenue, once the company loses the customer later. Furthermore, acquiring a customer is rather more difficult and expensive than it is to retain a paying customer. Since the boom in online business, this has become a major problem that needs a solution immediately. Here, in this project the data considered for Analysis is based on IBM Watson Telco Customer Churn dataset which is retrieved on 15th April 2020. The report discusses using Machine Learning and Neural Network Techniques along with Power BI Visualization to understand and find the customers that have high probability of churning in near future. It discusses one of the major market drivers of revenue.

Customers are considered to be the fuel that controls a business. A loss in customer will impact sales for a company. Furthermore, as mentioned earlier it is much more costly and time consuming acquire new customers in opposition to holding the existing ones. Thus, the need of the hour for organizations is to emphasize on plummeting customer churn.

1.1 Statement of the Problem

The telecommunication industry has been one of the top performing industry in terms of revenue. During recent times, telecom companies have faced multiple challenges that generate from frequent technological changes and customer demands. Moreover, the customers keep switching to different operators based on their geographical location. Most of the companies provide its customers with single connectivity and unified user experience.

Telecom is the fifth major industry globally and it also plays a crucial role in the world economy. As per TRAI's telecom services performance indicator report for 2018, the Adjusted Gross Revenue received by companies by providing telecom services plummeted over 10 per cent to Rs 1,44,446 crore during 2018 from Rs 1,60,814 crore in 2017.

Fig 1.1 Telecommunications Industry (Source The Financial Express)[1]

The Telecom industry is one of the many industries that consists of multiple subsectors. No single company can bear the costs for every installation, service, or customer experience. It allows users to communicate intra as well as across borders. This makes the industry as one of the most essential services to organizations and humans globally. Telecommunication is the

[1] https://www.financialexpress.com/industry/supreme-court-order-on-telecom-revenue-definition-deals-disastrous-blow-to-industry-coai/1745097/

transmission of information over substantial distances in order to affiliate communication. The telecommunications sector is delicate to even smallest modifications in regulations, economic factors, technology, etc. thus, faces specific trials budding from these fluctuations. A Telecommunication company provides multiple services including Telephone, mobile phones, radio and television, internet, LAN and WAN.

The Telecommunications Industry majorly invests in Technological Innovation by developing newest solutions. With the increasing growth rate, new value-added products are responsible for igniting the consumer spending behavior. Multiple cellular providers spend millions of dollars to upgrade the infrastructure from 3G to 4G. Some developed countries like Finland, Japan, Korea have moved towards 5G technology.

Every competitor in Telecommunications Industry wants to emerge as a successful player by maximizing the revenue and minimizing the cost. The major key driver in the business is changing customer preferences, operators are forced to evaluate and respond to the ever-changing customer demands and expectations. From using multiple devices to number porting, in case of a customer being unhappy by the service provided by the telecom service provider, he/she feels prompted to switch to a different business.

The data used for analysis is provided by the IBM Watson Telco Customer Churn Dataset retrieved on 15th April 2020. The idea is to Predict customer churn behavior. The platform allows the analysis of customer data in order to create specific retention programs for customers. IBM Watson allows data scientists and developers to run machine learning models and access its cloud platform for AI projects. Features of IBM Watson include deployment of machine learning models through IBM Watson Studio and other open source tools, dynamic retaining of models, generating automatic APIs to build AI- based applications, model management platform IBM Watson Open scale, and streamline model management and end to end deployment platform. Some additional features of the platform include:

- Automated AI Lifecycle
- Easy Deployment
- Model Operations
- End-to-End Integrated User Interface
- Model Scaling
- Dynamic Retaining

Since a specific organization has not been defined by the source of data, it is crucial to understand the revenue metric for a Telecommunication Service Provider.

A telecom service provider uses key revenue metric as ARPU also known as average revenue per user or average revenue per unit. It orders to calculate ARPU, a valid time frame is to be defined. Most of the telecom service providers calculate on a monthly basis to determine the total revenue generated through paying subscribers as well as communications devices. Also, because the number of units can vary from one day to another, total revenue is divided by the number of units.

$$\text{ARPU} = \frac{\text{Total Revenue}}{\text{Number of Users}}$$

Equation 1.1 ARPU

Another metric used is ARPPU that stands for Average Revenue Per Paying Customer. ARPPU is calculated by dividing the revenue among the users who paid. The figure obtained from ARPPU is larger than ARPU. High ARPU is not an indicator of high profitability since it does not take cost of providing services in account. Thus, to take the cost in account, companies use another revenue metric, AMPU (Average Profit Margin Per User). AMPU is more reliable and closer to actual number as compared to ARPU.

1.2 Purpose of the Study

A telecom company has been affected by the increasing number of customers subscribing to the services of a competitor. It is costlier to attract prime customer than to retain an old customer. At the same time, spending too much on or spending on the wrong factor for retaining customer who has no intention to leave (or who was not leaving for that factor which was addressed) could be a waste of money. Therefore, it is important to identify the customer who has high probability of leaving and zero down on the reason for it. An analysis of the past records of the customers can give great insights on who might leave and what is the cause. The telecom company already has this data available and data scientist need not collect the data in this case.

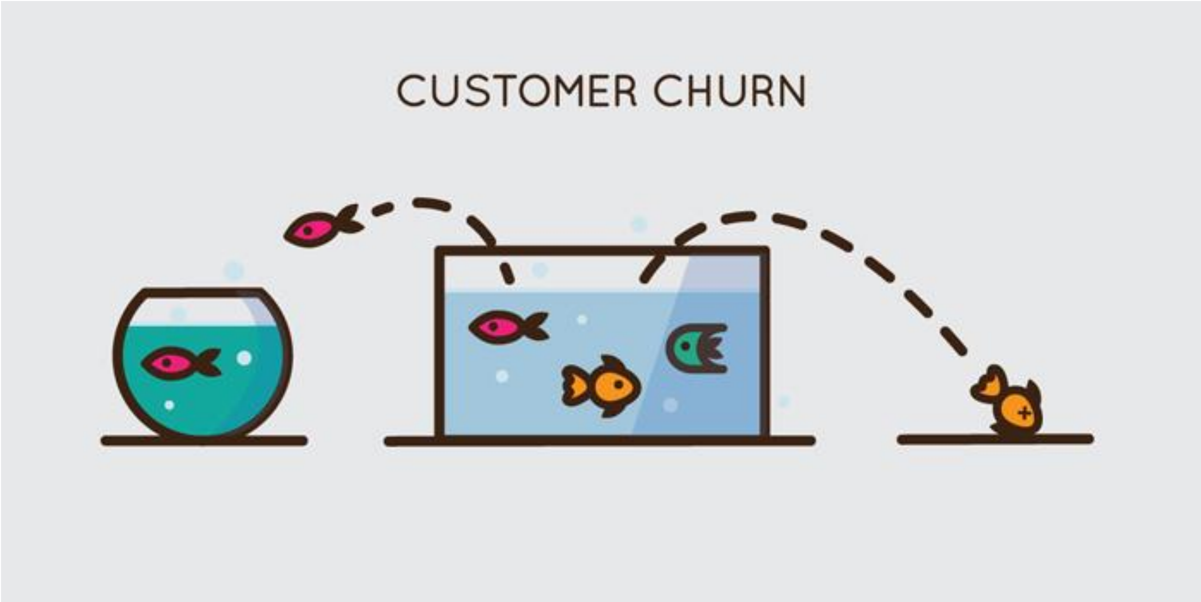


Fig 1.2 Customer Churn (Source Displayr.com)[2]

The prime objective of the study is to calculate the Customer Attrition Rate for the Telecommunications Company. Customer Churn is mathematically defined as the (Number. of Users at the beginning of the period subtracted to the Number of Users at the end of the period) divided by the total number of users at the beginning of the period.

$$\frac{\text{USERS AT BEGINNING OF PERIOD} - \text{USERS AT END OF PERIOD}}{\text{USERS AT BEGINNING OF PERIOD}} = \text{CHURN RATE}$$

Equation 1.2 Churn Rate

In a real scenario, we would get additional information from the business owner on relationship between charges and remaining factors. Still we will have to verify them from the data, because their assumptions on how they run business and the reality can differ.

Hence,

The Questions to which we seek answers:

1. Are monthly charges different for different contract types?
2. Are monthly charges solely dependent on the number/type of the services?
3. Are there any discounts given to loyal customers?

[2] https://www.displayr.com/predict-customer-churn-gradient-boosting/

4. Is there a correlation between monthly charges and churn?
5. Is there a correlation between tenure and churn?
6. Is there a correlation between certain type of services and churn?
7. Is there any person specific trends in churn?
8. Can we make predictions on likelihood of a customer churn given the predictor variables?

1.2 Structure of the Study

The major research project starts with a brief Introduction in Chapter 1 which discusses the purpose of the study and the rationale behind predicting customer churn in Telecommunications Industry. Chapter 2 is the Literature Review that has been done in order to provide a detailed explanation with the help of summarization of different research papers related to the dissertation topic. This chapter also discusses the customer churn and prediction techniques developed and studied by renowned authors in similar industry. Chapter 3 is Experiments Conducted which describes the Dataset along with its analysis and the pre-processing and data transformation techniques used to clean the data. Further, it also discusses the types of models developed. Additionally, the chapter illustrates various Power BI Dashboards and Visualizations that are created to provide a better understanding of the business scenario. Chapter 4 is the Results section which provides a brief description of the Machine Learning algorithms used along with the evaluation metrics used to compare the findings. This chapter also highlights the findings and recommendations for a company that aims to solve the problem of customer churn. Limitations of the Study and Conclusion is provided in Chapter 5. The chapter also reveals the overall understanding of the research and how along with its future scope.

CHAPTER 2

LITERATURE REVIEW

Customer is the backbone of any business and with the increasing competition, a customer is offered with multiple choices for a service or a product. This seems to be very evident in the Telecommunications Industry. There can be multiple reasons for a customer to drop a business and move on to the better option in the market. The major reason is the poor customer service. When customers do not receive the desired service, they move on to the competitors and share the negative feedback on social media channels. This directly hurts the revenue stream of a company. Other reasons for customer churn include lack of brand loyalty, lack of value, and low-quality service. As per recent survey it was found that companies spend seven times more on acquiring customers than losing a customer and the average global value of losing a customer is $243. The best way to reduce the customer churn is to predict it beforehand and address the problem before it becomes severe. By getting insights from customer patterns using Big Data platforms, companies can classify the subscribers who are at the verge of leaving the company. Since, multitude of companies already have the data required for predictive analysis of a customer's risk behaviour, this report uses one such dataset from IBM to evaluate the risk profile of each customer.

2.1 Customer Churn

Customer churn is the situation when a customer stops using services of a company. The Major reason for customer churn is divided into two classes, one is intentional and the other is accidental. Intentional churn occurs when customers intentionally switch from one company to another company that provides similar services whereas Accidental churn occurs when a customer is not able to use the services in future due to some unforeseen situation. For example, a service that becomes economically too costly for a customer. (J. Hadden et. Al, 2006)

It is important for a company to understand that not every customer is worth retaining and therefore, the churn management efforts must not focus across entire customer base. It should also be noted that customer retention costs money and an attempt to retain all the customers may lead to a wastage of resource, time, and money. (J. Hadden et. al, 2007)

As per the industry, it is believed that the best future marketing strategy is to retain existing customers and avoid losing customers to competitors. (Kim, Park, & Jeong, 2004; Kim & Yoon, 2004). Burez and Van den Poel (2007) indicated that Customer Churn can be avoided using two target techniques, one is reactive approach while the other is proactive approach. In the reactive approach, a company waits till the customer explicitly asks to cancel the service whereas in Proactive approach, a company attempts to identify customers that are most likely to churn. Here, the company pushes marketing incentives and programs to change the customer's mind. Regardless of the method used, it is important to predict the high-risk customer as accurately as possible because if the incentives are pushed to a customer who had no intention of leaving, this will lead to high marketing cost and low profits. (Van den Poel & Larivie‘re, 2004; Burez & Van den Poel, 2007).

Because Journals and Research papers are considered to be the most reliable source of information (Nord & Nord, 1995), some renowned online journals were explored from different publications including Elsevier, IEEE Xplore, Academia.edu, and SpringerLink to review the different techniques used for predicting customer attrition.

2.2 Prediction Techniques

To ensure an accurate customer churn prediction model is made, multiple techniques have been used by researchers. The techniques range from as simple as Logistic Regression to as complex as Deep Learning. The main aim is to perform the exploratory data analysis of the dataset and perform accurate predictions.

Chih-Fong Tsai & Yu-Hsin Lu, 2009 discusses hybrid data mining techniques to predict customer churn. Two combination methods of techniques are used to form hybrid models. One is based on uniting classification and clustering techniques i.e. SOM and ANN. SOM stands for Self-Organizing Maps, a competitive learning-based technique that clusters using neighbourhood function and preserves the topological features of the input space. The other method used by the paper is combining ANN+ANN models. Kohonen, 1987 demonstrated SOM that works well in cases of data that has high dimensionality and complexity. It is used to find associations in the data followed by clustering according similarity of data. The Figure 3 given below shoes an example of 4*4 SOM. The baseline model is a single ANN model that is used to compare with the hybrid models. Dataset used by the paper is a CRM dataset of American telecom companies. Data includes 34,741 churning customers and 16,545 non churning customers from July 2001 to January 2002. The objective was to find if any subscriber left the company during 31-60 days after the initial original sampling. It considers five-fold cross validation technique with 80% data as training set and 20% as test set. The model snippet is given in the figure below. First figure is a classifier combines with a cluster whereas other is a cluster combined with a classifier. As shown in Fig 2.1, To evaluate the models, Type 1 and Type 2 errors are considered that are measured by a confusion matrix.

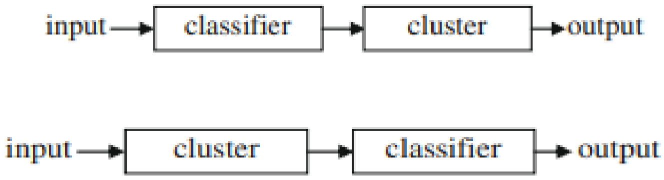


Fig 2.1 Types of Classifiers with Cluster Combination (Tsai, et. al., 2007)[3]

The results indicated that hybrid models perform better than the single neural network model in terms of accuracy. Out of the two hybrid models considered, ANN+ANN model outperforms ANN+SOM model. The paper also suggests to use other prediction techniques including Support vector machines, GA (Genetic Algorithms), etc.

Bucknix and Van den Poel, 2005 and Hung et. al., 2006 uses retail industry dataset to predict customer churn. The techniques used include neural networks and logistic regression. The paper does not perform any data pre-processing or cross-validation. Initially, the original dataset is trained using MLP. Kim and Yoon, 2004 discusses customer churn prediction in five mobile carriers in Korea using logistic regression.

Coussement and Van den Poel, 2008 uses a newspaper subscriber database to predict customer churn using support vector machines, random forests, and logistic regression. The performance of SVM is found to be better than logistic regression and random forests.

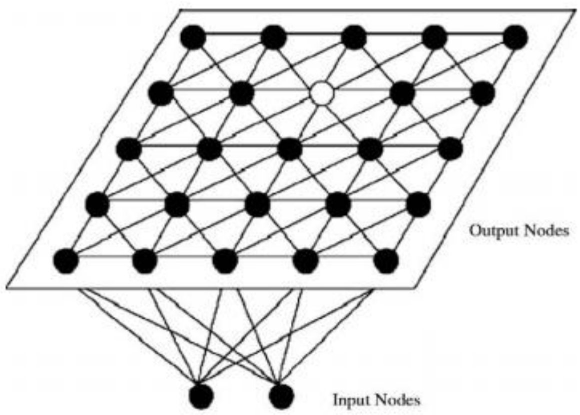


Fig 2.2 Self Organizing Maps (Source Kohonen, 1987)[4]

[3] [4]Tsai, C.-F., & Lu, Y.-H. (2009). *Customer churn prediction by hybrid neural networks. Expert Systems with Applications, 36(10), 12547–12553.* doi:10.1016/j.eswa.2009.05.032

Machine learning techniques can be used to predict and recognize customers who are at a high risk of churning before their attrition. The prediction and recognition are usually done on a month-by-month basis so that it can be later used to generate a specific selling and marketing strategy in order to stop such customers from churning. Major companies buy such products as Software as a service in order to deploy them in their day to day business operations and create progressive retention strategies. Further research has depicted that the performance of most of the machine learning algorithms is sternly affected by the representation used to describe the data being processed (features) (Y. Bengio, 2013). The reason (also known as curse of dimensionality) to this is the entwinement of features with other features and which in turn fleeces some descriptive factors that could have described the disparity in the input.

Decision Trees have also been used for churn prediction throughout the years. It is a tree like depiction for a classification problem. Nodes in a decision tree are called non-leaf's and these nodes represent explanatory variables or features. This gives a hierarchical representation of features that are linked to the target variable. Decision Trees are based on multiple algorithms including CART and C4.5. Also, it has been found that single decision trees underperform as compared to other methods. To solve this problem, bagging or ensemble methods were developed. Random Forest is also a type of ensemble method and it addresses the poor performance of decision trees. Random Forest Algorithm randomly splits the dataset into subset of samples and generates decision trees based on each subset. Although Random Forest Algorithm improves the performance and avoids overfitting but it does not perform well on extremely imbalanced dataset. Churn Prediction dataset is also type of such dataset. Thus, it is important to perform under sampling, oversampling, or SMOTE in order to balance the dataset. This flaw motivated (Y. Xie et. Al, 2009) to develop two algorithms namely Weighted Forests and Balanced Forests. The latter algorithm samples a dataset by matching the samples for class distribution of every tree. The process is recurrent until every tree has considered the majority class. The former random forest algorithm assigns weights to each class in such a way that weight of the majority class has a lesser weight than the minority class. This helps in penalizing misclassification accordingly. The proposed algorithm was evaluated on a Chinese Bank dataset. The dataset was also trained and tested on other algorithms including ANN, Random Forest, ADA Boost, and XG Boost. Results show that the proposed algorithms Balanced Trees and Weighted Trees performed better than other algorithms.

Vapnik introduced Support Vector Machines in 1995. SVM maps labelled data points in linear feature space such that the separation between classed is maximized. A hard margin has less probability of misclassifying as compare to a soft margin. The separation margin is maximized through an optimization algorithm. (B. Boser et. Al, 1992). In practice the data is not clearly linearly separable. To solve this, the input feature space is mapped non-linearly using kernel trick. Kernel Trick allows SVM to generalize non linearly separable data based on the kernel function used. Recently SVM was explored in churn prediction of subscription services. (Van De Poel et. al., 2008) Here, SVM model was applied to real data collected from Belgian Newspaper. The results showed that SVM performs better on noisy dataset as compared to Random Forests and Logistic Regression. Using SVM technique has one disadvantage that it took more time to train the model as compared to the time taken by Random Forest and Logistic Regression. This is considered to be the major drawback of SVM, as companies mostly deal with huge, high-dimensional datasets. Although the work done in the paper was a based on real data, Support Vector Machines can be unscalable in big data world.

Multiple Deep Learning models have also been developed. Deep learning models can be sub divided into 3 types of deep architecture with multiple applications. The types of deep architectures include:

- Hybrid Architectures: These architectures are considered to have high accuracy and are used when the objective is to perform classification. Such architectures are reinforced by consequences of Generative architecture.
- Discriminative Architecture: This type of architecture has application in classifying patterns that are described by preceding observations mentioned by experimental data.
- Generative Deep Architecture: The data with high correlation is more suitable for this architecture. It combines statistical techniques to perform pattern analysis.

Although Deep Learning has many applications but there are not many research papers that are applicable to churn prediction problems. (F. Castanedo et. al, 2005) is identified as one published paper that describes this scenario. The paper discusses the use of deep learning in predicting customer attrition for a dataset generated from a telecommunications industry. It recommends a discriminative deep architecture that uses a binary classifier which consists of four-layer feedforward neural network. This network uses call patterns of customers in order to distinguish between churner customers and non-churners based on the call patterns of users. The key inspiration behind the use of such architectures is to explore the likelihood of dodging feature engineering process that consumes. The architectures also assist in getting better predictions. Due to the high fundamental intricacies of customer call interactions, the need was to introduce a data representation architecture that describes customer performance across multiple layers whereas keeping the detailed representation intact. The activation function used by the architecture was a sigmoid function in hidden layers. The model was deployed on the training dataset whereas the results were checked on the testing dataset. According to the results published in the paper, it has been seen that the model does not overfit the data and take a broad view of the data. Additionally, the business was able to pointedly improve its estimate accuracy from 73.2% to 77.9% Area Under Curve (AUC).

CHAPTER 3

RESEARCH METHODOLOGY

The data used for analysis has been retrieved from IBM Watson Telco Customer Dataset. Dataset includes the information of customers based on churn (the customers who left last month), account information, services including payment method, monthly chargers, total charges, how long they have been a customer, paperless billing, etc. along with demographic information including age, gender, dependents, etc. Along with Data Description, the chapter discusses data analysis and pre-processing techniques used in order to develop efficient models. The pre-processing techniques performed on the raw data include Pruning, Feature Scaling, and One-Hot Encoding. Later classification models like Decision Trees, Logistic Regression, and Artificial Neural Network model constructed using Keras are developed to compare the accuracy with the Ensemble models like ADA Boost and XG Boost. Consistent data has been used throughout to evaluate model performance. This chapter also showcases the dashboards and visualizations created using Power BI in order to summarize a customer overview. The visualizations display customer churn risk and recommends ways to improve customer experience. Power BI Dashboard provides a visual representation for customer profile and revenue allocation. It delivers churn risk with respect to each customer.

3.1 Data Description

In order to perform the comparative study, the dataset considered for the analysis has been retrieved from IBM Watson Telco Dataset on 15th April 2020. According to IBM, the business challenge is:

"It is important for a company in the Telecommunications Industry to find the subscribers that are leaving as this results in huge losses in revenue for the company and business".

The dataset consists of 7043 (customers) entries. Each entry has 21 features and a column stating if the customer has churned in the past. Churn is the target variable here. The columns include following information about a customer:

a) **Customer_ID:** Identification ID specific to a subscriber

b) **Gender:** The customer's gender: Male, Female

c) **Senior_Citizen:** (Categorical Value) if a customer is 65+ then yes, else marked as no

d) **Partner**: Married subscriber: Yes, Unmarried subscriber: No

e) **Dependents:** Subscriber living with parents, children, grandparents(dependents): Yes, else No

f) **Tenure (Months):** Number of months for which the subscriber has had a relationship with the company

g) **Phone_Service:** Yes/No

Subscribed to home phone service:

h) **Multiple_Lines:** Yes/No

Subscriber has multiple telephonic connections:

i) **Internet_Service:** Mentions if a subscriber receives internet services from the company, Fiber optic, cable, DSL, or no such service

j) **Online_Security:** Yes/No

If a subscriber gets additional online security service from the company,

k) **Online_Backup:** Yes/No

If a subscriber gets additional online backup service from the company.

l) **Device_Protection_Plan:** Yes, No

Plan to protect device

m) **Technical_Support:** Yes, No

Technical Support plan with less turn-around time.

n) **Streaming_TV:** Yes/No

Television streaming through internet from a third-party service provider

o) **Streaming_Movies:** Yes/No

Movie streaming through internet from a third-party service provider

p) **Contract:** Month-to-Month/ One Year/ Two Year Contract Type

q) **Paperless_Billing:** Yes, No

If the subscriber has chosen e-billing option

r) **Payment_Method:** Mailed Check/ Credit Card/Bank Withdrawal

Bill payment method

s) Monthly_Charge: Numeric value for the charge for each month for a customer

t) Total_Charges: Numeric value for charges at the end of 3 months

u) **Churn:** Subscriber leaves the company: Yes, Subscriber stays with the company: No

The framework followed throughout the project is given in Fig 3.1 below. It gives the step by step flowchart that has been followed to perform necessary experiment during the course of this project.

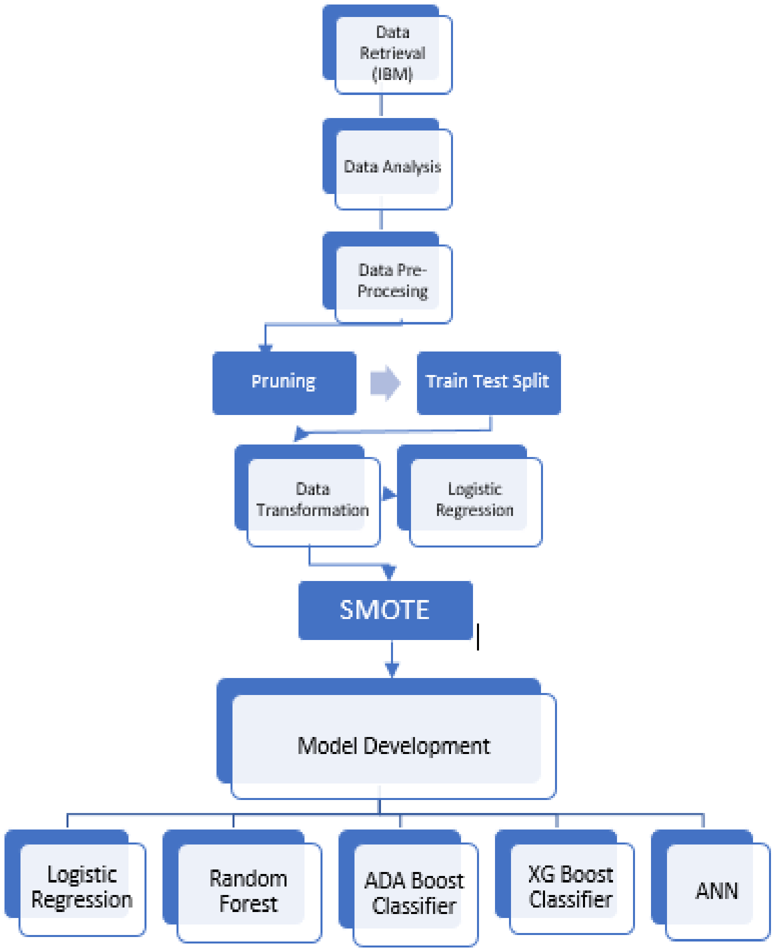


Fig 3.1 Framework for Experiment Conducted

Data view when loaded in Python and CSV File view is given in figure 3.2 and 3.3.

```
churn.head()
```

	customerID	gender	SeniorCitizen	Partner	Dependents	tenure	PhoneService	MultipleLines	InternetService	OnlineSecurity	...	DeviceProtection	TechSup
0	7590-VHVEG	Female	0	Yes	No	1	No	No phone service	DSL	No	...	No	
1	5575-GNVDE	Male	0	No	No	34	Yes	No	DSL	Yes	...	Yes	
2	3668-QPYBK	Male	0	No	No	2	Yes	No	DSL	Yes	...	No	
3	7795-CFOCW	Male	0	No	No	45	No	No phone service	DSL	Yes	...	Yes	
4	9237-HQITU	Female	0	No	No	2	Yes	No	Fiber optic	No	...	No	

5 rows × 21 columns

Fig 3.2 Data View in Python

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
1	customerI	gender	SeniorCiti	Partner	Depender	tenure	PhoneSer	MultipleLi	InternetSe	OnlineSec	OnlineBac	DevicePro	TechSupp	Streaming	Streaming	Contract	Paperless	PaymentM	MonthlyC	TotalChar	Churn
2	7590-VHV	Female	0	Yes	No	1	No	No phone	DSL	No	Yes	No	No	No	No	Month-to-	Yes	Electronic	29.85	29.85	No
3	5575-GNV	Male	0	No	No	34	Yes	No	DSL	Yes	No	Yes	No	No	No	One year	No	Mailed ch	56.95	1889.5	No
4	3668-QPYI	Male	0	No	No	2	Yes	No	DSL	Yes	Yes	No	No	No	No	Month-to-	Yes	Mailed ch	53.85	108.15	Yes
5	7795-CFO	Male	0	No	No	45	No	No phone	DSL	Yes	No	Yes	Yes	No	No	One year	No	Bank trans	42.3	1840.75	No
6	9237-HQIT	Female	0	No	No	2	Yes	No	Fiber opti	No	No	No	No	No	No	Month-to-	Yes	Electronic	70.7	151.65	Yes
7	9305-CDSI	Female	0	No	No	8	Yes	Yes	Fiber opti	No	No	Yes	No	Yes	Yes	Month-to-	Yes	Electronic	99.65	820.5	Yes
8	1452-KIOV	Male	0	No	Yes	22	Yes	Yes	Fiber opti	No	Yes	No	No	Yes	No	Month-to-	Yes	Credit car	89.1	1949.4	No
9	6713-OKO	Female	0	No	No	10	No	No phone	DSL	Yes	No	No	No	No	No	Month-to-	No	Mailed ch	29.75	301.9	No
10	7892-POO	Female	0	Yes	No	28	Yes	Yes	Fiber opti	No	No	Yes	Yes	Yes	Yes	Month-to-	Yes	Electronic	104.8	3046.05	Yes
11	6388-TAB	Male	0	No	Yes	62	Yes	No	DSL	Yes	Yes	No	No	No	No	One year	No	Bank trans	56.15	3487.95	No
12	9763-GRSI	Male	0	Yes	Yes	13	Yes	No	DSL	Yes	No	No	No	No	No	Month-to-	Yes	Mailed ch	49.95	587.45	No
13	7469-LKBC	Male	0	No	No	16	Yes	No	No	No intern	No intern	No intern	No intern	No intern	No intern	Two year	No	Credit car	18.95	326.8	No
14	8091-TTV	Male	0	Yes	No	58	Yes	Yes	Fiber opti	No	No	Yes	No	Yes	Yes	One year	No	Credit car	100.35	5681.1	No
15	0280-XJGE	Male	0	No	No	49	Yes	Yes	Fiber opti	No	Yes	Yes	No	Yes	Yes	Month-to-	Yes	Bank trans	103.7	5036.3	Yes
16	5129-JLPIS	Male	0	No	No	25	Yes	No	Fiber opti	Yes	No	Yes	Yes	Yes	Yes	Month-to-	Yes	Electronic	105.5	2686.05	No
17	3655-SNQ	Female	0	Yes	Yes	69	Yes	Yes	Fiber opti	Yes	Yes	Yes	Yes	Yes	Yes	Two year	No	Credit car	113.25	7895.15	No
18	8191-XWS	Female	0	No	No	52	Yes	No	No	No intern	No intern	No intern	No intern	No intern	No intern	One year	No	Mailed ch	20.65	1022.95	No
19	9959-WOF	Male	0	No	Yes	71	Yes	Yes	Fiber opti	Yes	No	Yes	No	Yes	Yes	Two year	No	Bank trans	106.7	7382.25	No
20	4190-MFLI	Female	0	Yes	Yes	10	Yes	No	DSL	No	No	Yes	Yes	No	No	Month-to-	No	Credit car	55.2	528.35	Yes
21	4183-MYFI	Female	0	No	No	21	Yes	No	Fiber opti	No	Yes	Yes	No	No	Yes	Month-to-	Yes	Electronic	90.05	1862.9	No
22	8779-QRD	Male	1	No	No	1	No	No phone	DSL	No	No	Yes	No	No	Yes	Month-to-	Yes	Electronic	39.65	39.65	Yes
23	1680-VDC	Male	0	Yes	No	12	Yes	No	No	No intern	No intern	No intern	No intern	No intern	No intern	One year	No	Bank trans	19.8	202.25	No

Fig 3.3 CSV File

3.2 Data Analysis

Data Analysis depicts the structure and summary of the data. It is important to know if there is any imbalance in the data as it may lead to skewed results. Key takeaways from Analysis include:

- Out of 21 features, 17 features are categorical features as shown in Fig 3.5.
- The dataset is imbalanced as it has very few minority cases as compared to the majority cases for the target variable (73.4% majority cases; Churn=No whereas 26.6% minority cases; Churn=Yes)
- Variable TotalCharges has 11 missing values out of 7043 rows as shown in Fig 3.4.
- We can drop CustomerID column since it has no part in analysis of the dataset
- Tenure is negatively correlated to churn whereas Contract month-to-month is positively correlated to the target variable.

```
> summary(churn_data)
      customerID       gender      SeniorCitizen    Partner    Dependents      tenure       PhoneService          MultipleLines     InternetService
 0002-ORFBO:   1   Female:3488   Min.   :0.0000   No :3641   No :4933   Min.   : 0.00   No : 682     No              :3390   DSL        :2421
 0003-MKNFE:   1   Male  :3555   1st Qu.:0.0000   Yes:3402   Yes:2110   1st Qu.: 9.00   Yes:6361     No phone service: 682   Fiber optic:3096
 0004-TLHLJ:   1                 Median :0.0000                         Median :29.00                Yes             :2971   No         :1526
 0011-IGKFF:   1                 Mean   :0.1621                         Mean   :32.37
 0013-EXCHZ:   1                 3rd Qu.:0.0000                         3rd Qu.:55.00
 0013-MHZWF:   1                 Max.   :1.0000                         Max.   :72.00
 (Other)   :7037
             OnlineSecurity              OnlineBackup             DeviceProtection              TechSupport                StreamingTV
 No                 :3498   No                 :3088   No                 :3095   No                 :3473   No                 :2810
 No internet service:1526   No internet service:1526   No internet service:1526   No internet service:1526   No internet service:1526
 Yes                :2019   Yes                :2429   Yes                :2422   Yes                :2044   Yes                :2707



             StreamingMovies            Contract     PaperlessBilling                   PaymentMethod   MonthlyCharges    TotalCharges      Churn
 No                 :2785   Month-to-month:3875   No :2872          Bank transfer (automatic):1544   Min.   : 18.25   Min.   :  18.8   No :5174
 No internet service:1526   One year      :1473   Yes:4171          Credit card (automatic)  :1522   1st Qu.: 35.50   1st Qu.: 401.4   Yes:1869
 Yes                :2732   Two year      :1695                     Electronic check         :2365   Median : 70.35   Median :1397.5
                                                                    Mailed check             :1612   Mean   : 64.76   Mean   :2283.3
                                                                                                     3rd Qu.: 89.85   3rd Qu.:3794.7
                                                                                                     Max.   :118.75   Max.   :8684.8
                                                                                                                      NA's   :11
> |
```

Fig 3.4 Summary in R

```
churn.columns.values
```

```
array(['customerID', 'gender', 'SeniorCitizen', 'Partner', 'Dependents',
       'tenure', 'PhoneService', 'MultipleLines', 'InternetService',
       'OnlineSecurity', 'OnlineBackup', 'DeviceProtection',
       'TechSupport', 'StreamingTV', 'StreamingMovies', 'Contract',
       'PaperlessBilling', 'PaymentMethod', 'MonthlyCharges',
       'TotalCharges', 'Churn'], dtype=object)
```

Fig. 3.5 Column Values in Python

3.3 Data Pre-Processing

Initial step to process the data is Pruning. Pruning refers to removing the unnecessary columns and rows.

3.3.1 Pruning

The data has some columns that were removed before applying the deep learning model. The "customer_ID" is a specific identification that is specific to every customer and is not needed for data modelling hence we can remove this column. Further, Column "Total Charges" has total 11 NA values, since this is just 0.2% of the total cases available. We can drop the rows that have 0 value in this column.

3.3.2 Train-Test Split

Once the pruning is done, we will divide the data into test and training sets using train-test-split in Python. Here 70% data is training set whereas 30% is test set. Post sampling, it was found that out of total 7032 cases, 4922 cases are is training set while 2110 cases are in test set. Ensemble models are run on both train and test set.

3.4 Data Transformation

After pre-processing of data, some data transformation steps are used to prepare the data so that the algorithm can be run effectively.

3.4.1 Feature Discretization

Features depicting length of time like years and age can be generalized into a cohort. The "tenure" feature is also an example of such independent variable and has numeric values that can be categorized into groups. By dividing tenure in 6 parts it will become easy for a Machine Learning algorithm to predict if a part is at high risk for churning.

3.4.2 One Hot Encoding

One Hot Encoding refers to the method of changing categorical variables into binary vectors. This requires mapping of categorical values to its integer values where the formed matrix is also called design matrix. Forming dummy variables is easy in case of Yes and No since we just have to convert 1's and 0's, it is easy here. Here we have four similar type of features including Contract, Internet Service, Payment Method, and Multiple Lines. One hot encoding in Python is given in Figure 3.6.

```
df_dummies=pd.get_dummies(df)
df_dummies.head()
```

	SeniorCitizen	tenure	MonthlyCharges	TotalCharges	Churn	gender_Female	gender_Male	Partner_Nc
0	0	1	29.85	29.85	0	1	0	(
1	0	34	56.95	1889.50	0	0	1	1
2	0	2	53.85	108.15	1	0	1	1
3	0	45	42.30	1840.75	0	0	1	1
4	0	2	70.70	151.65	1	1	0	1

5 rows × 46 columns

Fig 3.6 One of Encoding of Categorical Variables

3.4.3 Feature Scaling

Since we are using Artificial Neural Network to train our model, it is considered that ANN performs faster when the features are normalized or scaled. This process is also called standardizing or scaling as shown in Fig 3.7. Feature scaling is an important step for modelling multiple types of algorithms including SVM, Neural Networks, KNN, etc.

```
from sklearn.preprocessing import MinMaxScaler
features=X.columns.values
scaler=MinMaxScaler(feature_range=(0,1))
scaler.fit(X)
X=pd.DataFrame(scaler.transform(X))
X.columns=features
```

Fig 3.7 Feature Scaling in Python

3.4.4 SMOTE

On analysis it was found that the dataset is imbalanced as there are few cases of customers who churned in proportion to the number of cases where the churn did not occur. Since, out business problem is to find the number of cases that may churn, it is important to use a balanced dataset. In order to balance the dataset, SMOTE (Synthetic Minority Over-sampling technique) is used as shown in Fig 3.8. It is a statistical technique that is used to increase the number of cases in the dataset in a balanced manner. This technique generates new cases from existing minority cases that have been provided as Input. Here, SMOTE will synthetically increase the number of customers who have churned so as to interpret the data better and move on to get accurate predictions. We will also see how using SMOTE improves Precision and Recall Score of a model.

```
os = SMOTE(random_state=41)

os_data_X,os_data_y=os.fit_sample(X_train,y_train)
os_data_X = pd.DataFrame(data=os_data_X,columns=columns )
os_data_y= pd.DataFrame(data=os_data_y,columns=["Churn"])
```

Fig 3.8 SMOTE Implementation

3.5 Model Development

In an attempt to use Neural Network here, we use ANN model which is multi-layered intricately connected set of neural nets as shown in Fig 3.9 and Fig 3.10. The components of ANN include input layer, hidden layers and an output layer. The network is deep if there are multiple number of hidden layers.

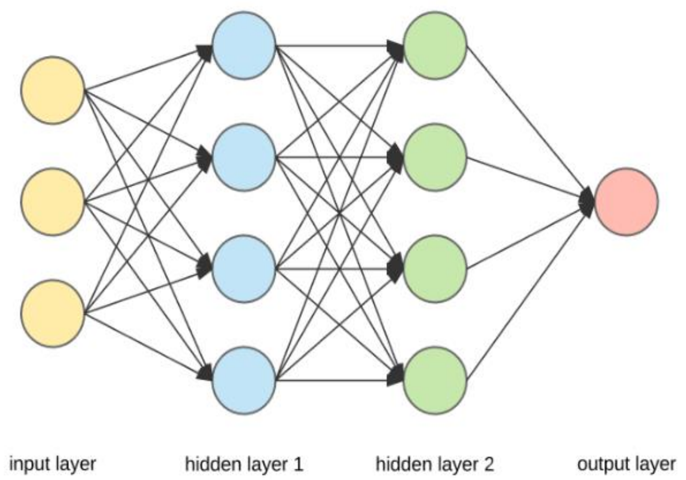


Fig 3.9 ANN Model (Source Towards Data Science)[5]

[5] https://towardsdatascience.com/applied-deep-learning-part-1-artificial-neural-networks-d7834f67a4f6

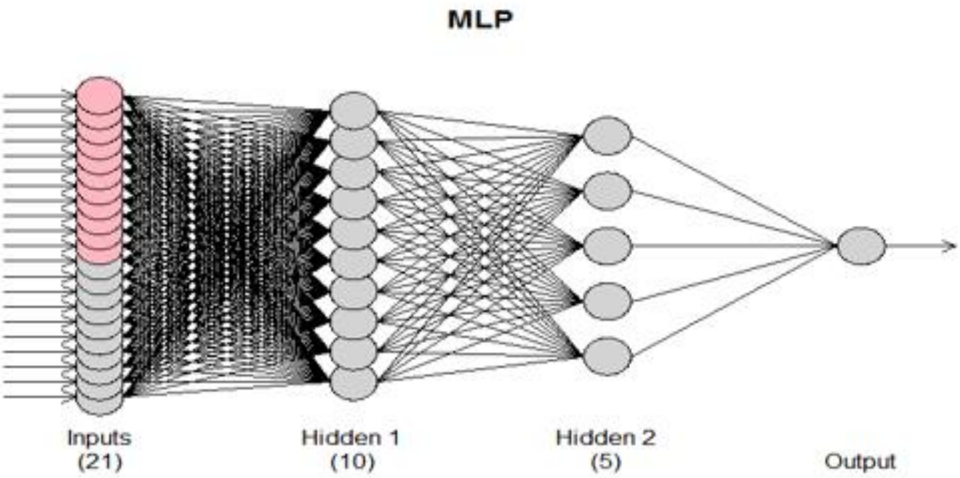


Fig 3.10 MLP

Sum of weighted inputs is given as input to a node. The inputs are then passed to a non-linear activation function. The output of this layer becomes the input of next layer and so on. The flow of signal takes place from left to right. The activation function and weights are shown in Figure 3.11. Any ANN model with more than 3 layers is considered as Deep Learning.

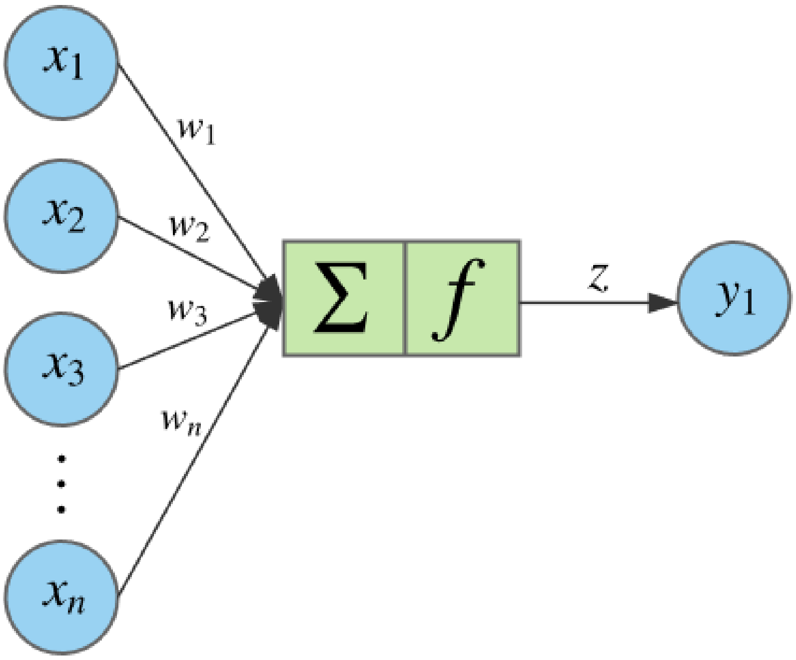


Fig 3.11 Mathematical Depiction (Source Towards Data Science)[6]

[6] https://towardsdatascience.com/applied-deep-learning-part-1-artificial-neural-networks-d7834f67a4f6

There are many packages available in R for creating a Deep Learning model. Some of the packages include:

a) MXNetR: FFNN, CNN

b) darch: Deep belief network

c) deepnet: Stacked autoencoders, FFNN, DBN

d) deepr: Simplification of some packages from H2O and deepnet

e) H2O: Deep autoencoders and feed forward neural network.

Here, for the customer churn problem, we use ANN model called Multi-Layer Perceptron. We use Keras in R to form the required model for churn prediction in our problem. These are the most basic form of Deep Learning but they are considered to be precise and perform better with such data. Once we have built a sequential model, we apply layers to that model. The layers consist of output, hidden, and input layers.

Once we are done with model development. We make predictions on the test dataset which was not known during modelling. The functions used for generating predictions are predict_proba() and predict_classes(). Unfortunately, on further investigation it was found that the Deep Learning Model does not perform well as compared to other models. This may be due to small size of the data of due to many categorical variables in the dataset.

To progress the precision and recall of the model, few Ensemble and Boosting algorithms like ADA Boost and XGBoost are implemented in Python. XGBoost is an ensemble model that is based on decision tree implementation and it uses the gradient boosting framework. This algorithm is expected to outperform other algorithms for unstructured algorithms. This algorithm was developed by Tianqui Chen et. al., 2006. The model has multiple features including:

- Core computing
- Avoids overfitting
- Regularization
- Tree pruning using depth-first approach
- Efficient handling of missing data
- Parallelized tree building
- In built cross validation capability

On implementing these models, it was found that the accuracy has increased to 86.6% whereas the precision and recall has also increased to 85.6 and 87.7 % respectively. In a customer churn model, accuracy of 87% with such precision can be considered a better model. Not many research papers have explored the possibility of using such boosting models on churn prediction problem.

3.6 Power BI Dashboard

Microsoft offers a visualization platform that assists a user to gather data from multiple sources and publish them into BI reports and user engaging dashboards. This platform is a Business Intelligence tool also known as Power BI. Various offerings of Power BI versions include:

- Service-based (SaaS)
- Desktop
- Mobile Power BI

Features of Power BI include built-in real-time dashboard, reliable connection for data sources. Hybrid configuration, quick deployment, natural language query-based data exploration, updated community-based dashboard visualisation. Different versions and offerings of Power Business Intelligence software include:

Power BI Data Gateway: Gateway is a link that is installed by administrator and it connects service and data sources that are present on the servers of a company. It allows numerous operations including Import and Live Query.

Power BI service: A SaaS platform that allows users to build reports and data models.

Power BI Desktop: It allows developers to generate models and create reports. This is the most used tool of Power BI.

Power BI Mobile Application: Mobile application is used to view and amend reports and dashboards that have been created by others. The application is available on Android, Windows and iOS devices.

Power BI Report Server: The server is updated by the IT team in every 4 months. It allows users to upload reports and dashboards to be uploaded on the server. Once the reports are uploaded, these can be viewed by the users.

Table 3.1 Power BI Term Description

Power BI Term	**Meaning**
Visualization	Visualization reports and dashboards that are used to showcase information hidden in the data
Datasets	Imported dataset that can be renamed, refreshed, and explored.
Dashboards	Collection of widgets and tiles that can be used to visualize the subsets of underlying datasets.
Reports	One or multiple pages visualizations created using datasets.
Tiles	Single visualization dashboard.

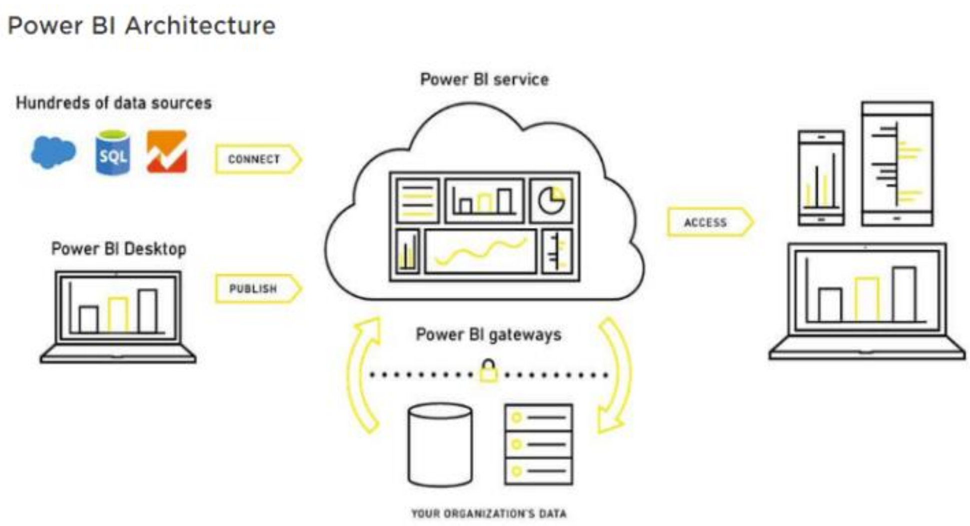


Fig 3.12 Power BI Architecture (Source Microsoft)[7]

[7] https://docs.microsoft.com/en-us/power-bi/admin/service-admin-power-bi-security

Power BI term description is given in Table 3.1. As per Power BI Architecture shown in Fig 3.12, it has three components as explained below:

Data Integration: Data can be acquired through multiple resources and can be in different file formats. Data can be structured, unstructured, and semi-structured. Once extracted from different sources, the data can be integrated into a standard form for further processing and analysis. Data integration is done at a common area.

Data Processing: After integration of data, the data is pre-processed. Pre-processing includes removing redundant values, missing values, omitting columns, etc. This helps in providing clean analysis.

Data Presentation: After the data processing, data is loaded into Power BI server for visualizations. The data can be used to build reports and dashboards which allows a user to present the data more intuitively. It also helps in taking effective business decisions that are based on the visualizations.

The data visualization for IBM Telecom dataset is mentioned below. Power BI offers effective colours and visualization options that pass on the information in a systematic manner.

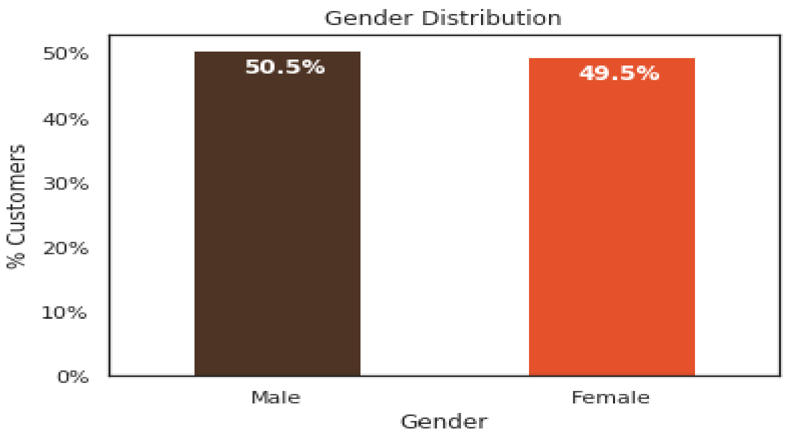


Fig. 3.13 Gender Distribution

Gender Distribution - About half of the customers in our data set are male while the other half are female as shown in Fig 3.13.

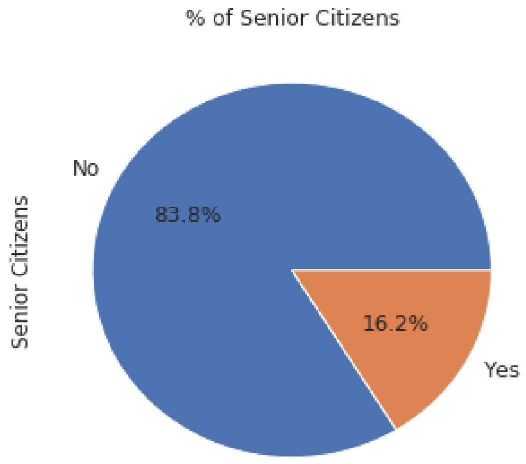


Fig. 3.14 Senior Citizen Distribution Pie Chart

Senior Citizen: There are only 16% of the customers who are senior citizens. Thus, most of our customers in the data are younger people as shown in Fig 3.14.

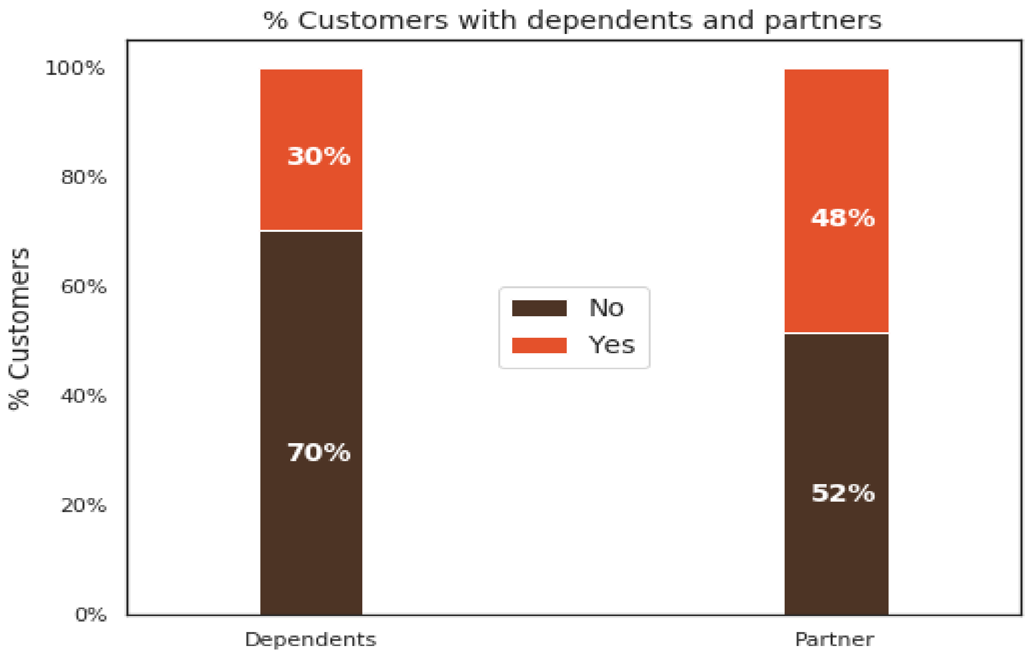


Fig. 3.15 Distribution of customers with Dependents and Partners

As shown in Fig 3.15, Partner and dependent status - About 50% of the customers have a partner, while only 30% of the total customers have dependents.

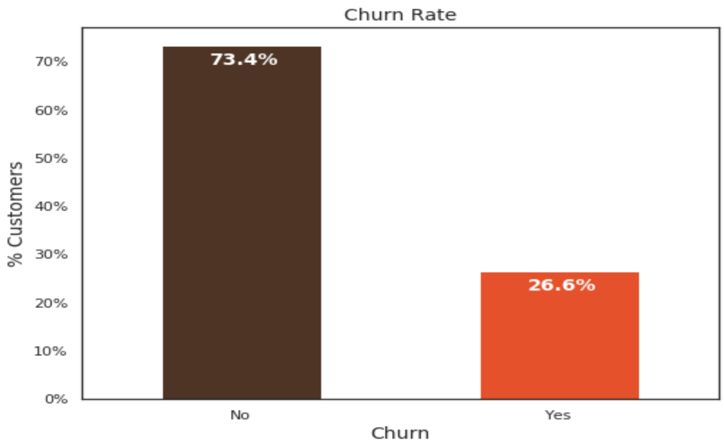


Fig 3.16 Churn Rate

As shown in Fig 3.16, In our data, 74% of the customers do not churn. Clearly the data is skewed as we would expect a large majority of the customers to not churn. This is important to keep in mind for our modelling as skewness could lead to a lot of false negatives. We have seen in the modelling section on how to avoid skewness in the data. We used SMOTE (Synthetic Minority Oversampling Technique) to balance the data. This also improved the precision and recall score of the model. The detailed score and comparison are mentioned in the Evaluation Metric section of the research paper.

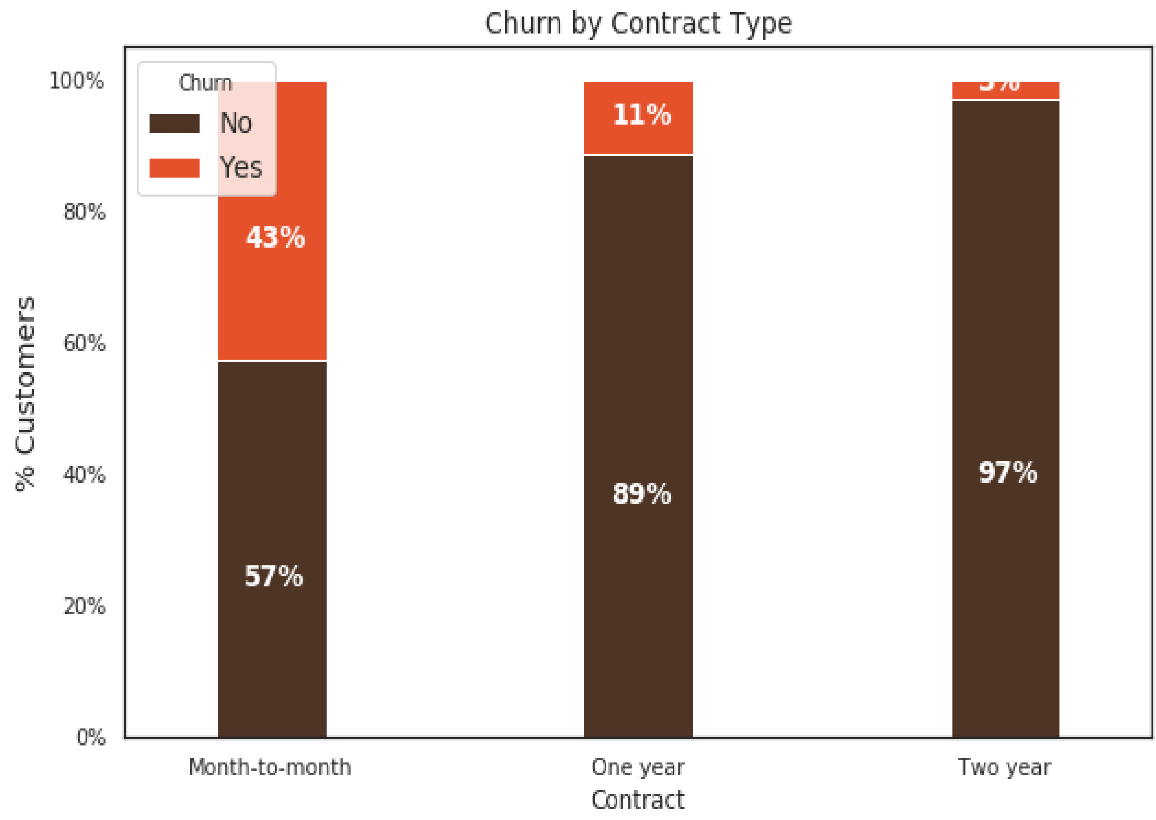


Fig 3.17 Customer Churn Vs Contract Mode

Customer Churn by Contract Mode: The subscribers with month to month contract have a very high churn proportion. This can also be seen in the correlation plot as Month-to-Month contracts are highly positively correlated with the target variable churn whereas the customers who opt for 2 year subscription are very less likely churn as per the given data as shown in Fig 3.17.

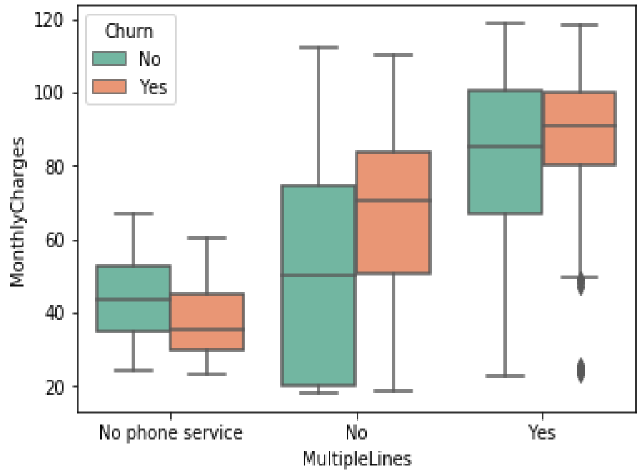


Fig 3.18 Churn by Phone service and MultipleLines

Customers that have phone service and multiple lines will churn less as compare to the subscribers who do not avail such services as shown in Fig. 3.18.

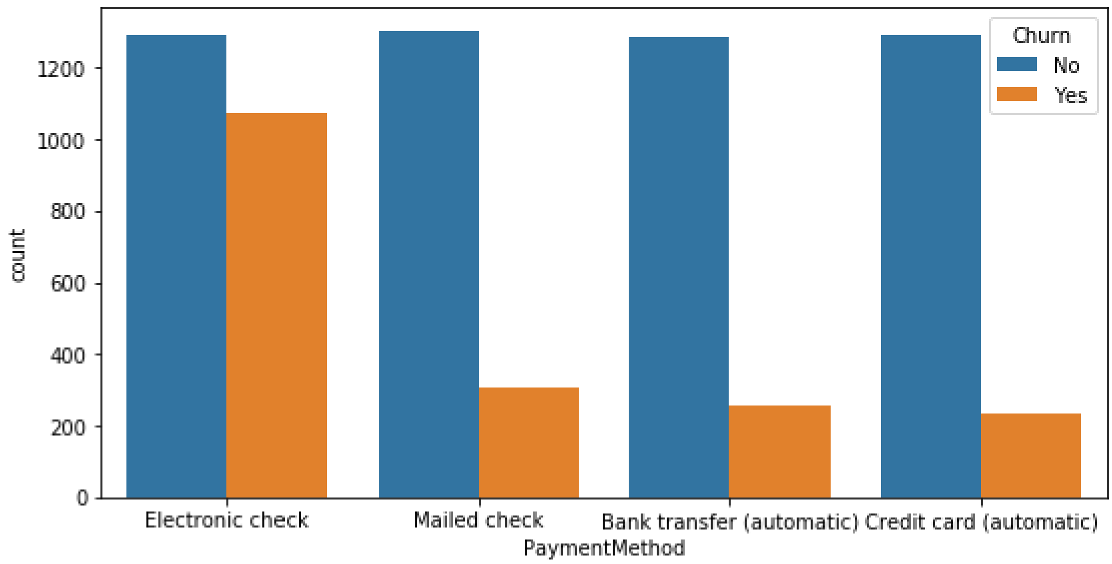


Fig 3.19 Churn Count by Payment Mode

As per figure 3.19, it has been found that the customers that paid through electronic mode are more susceptible to churning whereas the customers who paid through credit card have less churn rate.

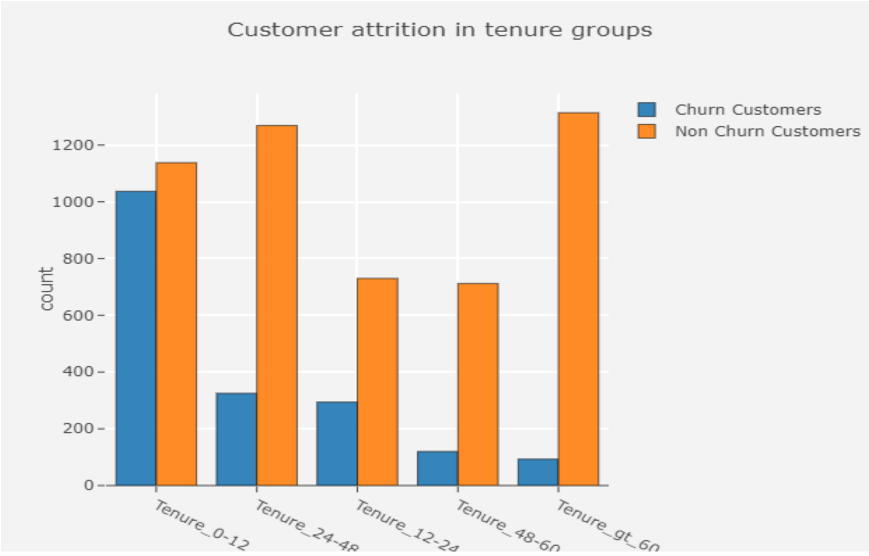


Fig 3.20 Churn Vs Tenure Group in Months

Customers in Tenure group more than 60 months are non-churners whereas customers withing Tenure group of 0-12 months have highest risk of churning as shown in Fig 3.20.

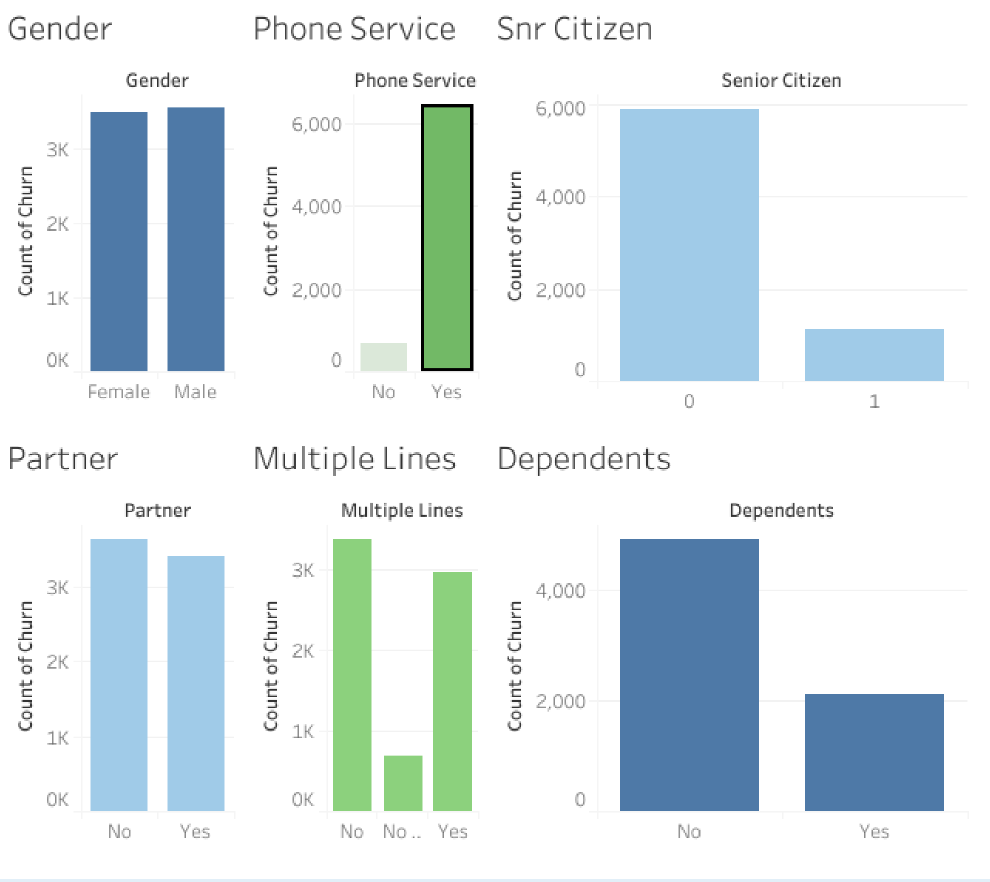


Fig 3.21 Power BI Dashboard for other features

Shiny is a package in R language that allows users to build interactive web applications directly from the R program interface. The package allows users to host singular apps on a webpage or insert them in R markdown documents. Additionally, users can put the shiny app on web by using RStudio's hosting service. In this project as an additional feature, we have created a Shiny application that summarizes information of each customer through a web-based platform as shown in Fig 3.21 and Fig 3.22.

As a Data Analyst or Machine Learning Engineer, it is important to communicate insights generated from data mining techniques to the management of the company. Many higher-level management officials in administration play a vital role in making essential business decisions on an everyday basis.

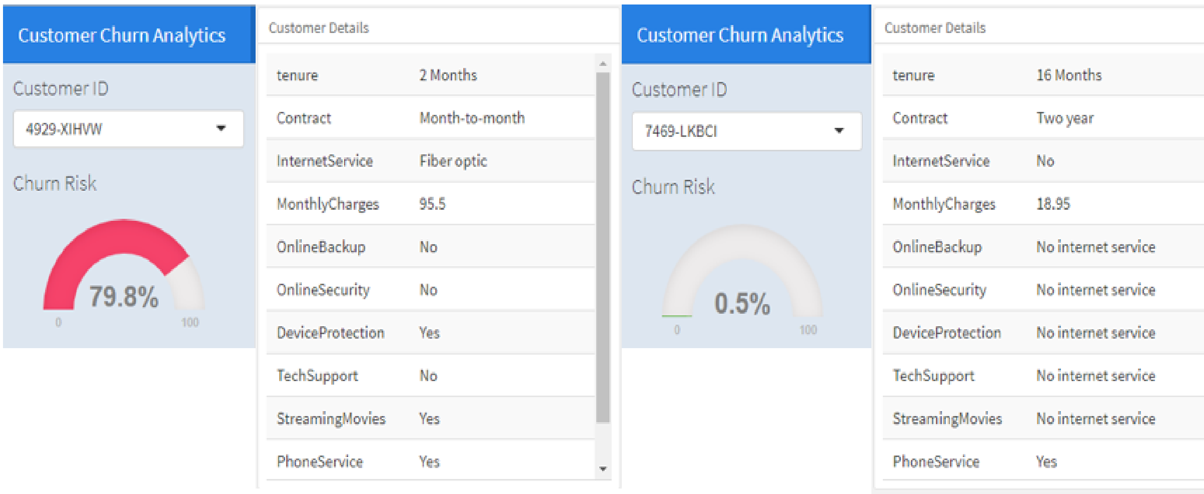


Fig 3.22 Customer Churn Web Dashboard

CHAPTER 4
RESULTS AND RECOMMENDATIONS

The obtained results will be of much value to the marketing team of a telecom company since it helps in better aligning the marketing resources and strategies with respect to the high-risk customers. The basic rule of marketing states that it is difficult to gain a new customer as compared to retaining a current customer, analysis of customer churn will also help a company in analyzing the reasons behind low customer retention. The Revenue model of a telecommunication company is directly related to its number of subscribers. Therefore, finding factors that may help a company in reducing its customer churn will lead to more revenue. The model accuracy has been measured through multiple formats including Accuracy, AUC (Area Under Curve), Precision and Recall, and F1 Score. ANN model provides an accuracy of 72% whereas Precision and Recall values for it are 0.65 and 0.62 respectively. Since ANN and Logistic Regression model did not provide the desired accuracy, ensemble models were used like XGBoost and ADABoost. Accuracy of these models reached up to 87% with a better precision and recall score. This project report also abridges the churn prediction techniques to provide an exhaustive understanding of the customer attribution along with proving the hypothesis that most precise churn prediction is given by the ensemble models rather than simple algorithmic models or an ANN model. It assists in making the Telecom industries aware of the needs of high-risk subscribers and improve their services to upturn the churn decision.

4.1 Machine Learning Approaches

a) Logistic Regression: Logistic Regression is more interpretable, faster, gives better overall accuracy than Random Forest, with slight degradation in recall of Churn group.

Confusion Matrix for Logistic Regression

	FALSE	TRUE
0	1392	156
1	273	287

Unbalanced Dataset:

SMOTE (Synthetic Minority Oversampling Technique):

Oversampled data: 7228

Yes: 3614

No: 3614

Churn yes data after SMOTE is 0.5

Churn no data after SMOTE is 0.5

b) Random Forests: There is actually no need to drop correlated features for Random Forest. It selects best features at every node of every tree by itself. But from a business perspective, TotalCharges is related to monthly and tenure, we intentionally drop it to get the correct picture (better interpretation) on feature importance. Phone service is also dropped.

Regarding the variance importance it is found that logistic regression model and decision tree based random forest model are only somewhat different. Both models have Streaming_Movies and Partner , Monthly_Charges, Tenure, Payment_Method, Contract and od as significant predictors and have gender, Streaming_TV, as insignificant predictors. However, in the logistic regression model, Online_Backup, Phone_Service, Paperless_Billing, and show substantial impact on the churn rate, whereas in the Random Forest model, such features have much less importance. The basic Decision tree model is given in Fig 4.1.

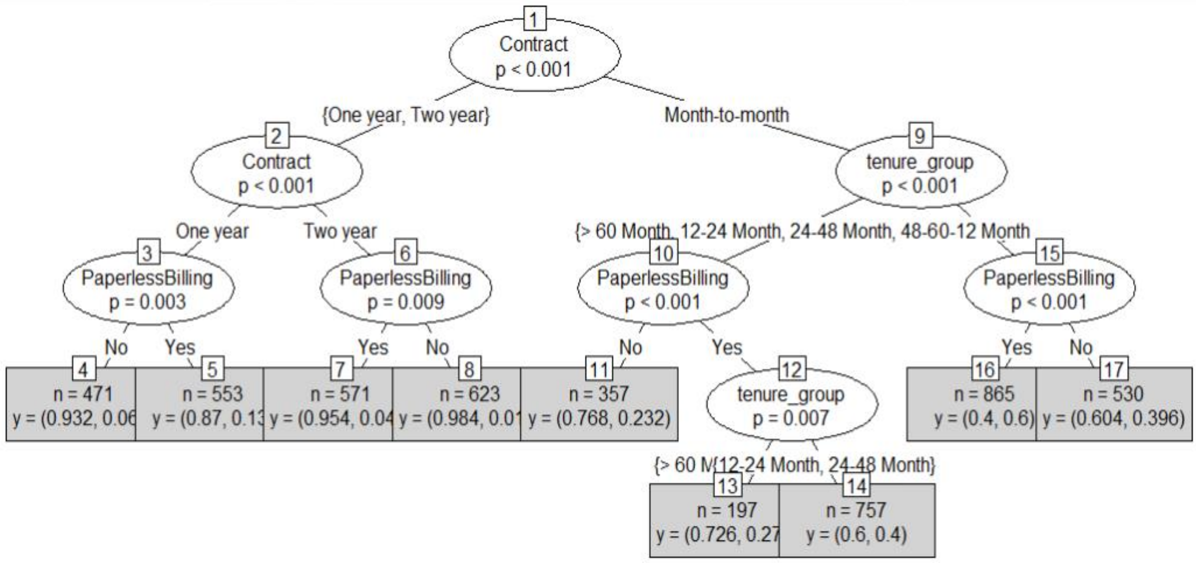


Fig. 4.1 Decision Tree

Confusion Matrix for Decision Tree

Actual

Predicted No Yes

No 1395 346

Yes 153 214

Decision Tree Accuracy 0.763282732447818

Confusion matrix for Random Forest

No Yes class.error

No 3245 370 0.1023513

Yes 647 662 0.4942704

Random Forest Accuracy 0.7898878676878

Rank-wise importance of features for RF Classifier is given in Fig 4.2.

	importance
Contract_Month-to-month	0.231273
PaymentMethod_Electronic check	0.151775
TotalCharges	0.087196
tenure	0.085803
MonthlyCharges	0.081890
PaperlessBilling_Yes	0.036800
DeviceProtection_No	0.033297
Contract_Two year	0.030802
OnlineSecurity_Yes	0.025759
Contract_One year	0.023299
OnlineBackup_Yes	0.022433
InternetService_Fiber optic	0.021337
Dependents_Yes	0.020315
gender_Male	0.020308
TechSupport_Yes	0.018715
StreamingMovies_Yes	0.016825
Partner_Yes	0.013548
MultipleLines_Yes	0.013324
StreamingTV_Yes	0.012898
SeniorCitizen_0	0.012122
PaymentMethod_Credit card (automatic)	0.010465
DeviceProtection_Yes	0.008161

Fig. 4.2 Rank-wise importance of features for RF Classifier

c) ANN Model: Artificial Neural Network based Multi-Layer Perceptron model performs the best out of all the machine learning algorithms developed through the course of this paper. With multiple hidden layers the accuracy of the model increased as compared to the previous models: `Truth`

```
Prediction  no yes

       no  950 161

       yes  99 196

accuracy

      <dbl>

1 0.840782
```

d) ADA Boost: For binary classification with weak learners we use ADA Boost algorithm. It is based on boosting and bagging techniques. In ADA Boost, weighted training data is used to train the weak models. ADA Boost classifier gives an accuracy of 84.7% with F1 score of 85%.
e) XG Boost: XG Boost is a custom tree building algorithm that is based on eXtreame Gradient Boosting technology. It can be used for problems related to classification, regression, and ranking with custom loss functions. For the churn prediction model, XG Boost gives the highest accuracy of 87% with precision and recall of 87.2 and 86.28 respectively.

4.2 Evaluation Metrics

To calculate evaluation metrics, confusion matrix is used, as shown in Table 4.1. Following are the evaluation metrics that were used to evaluate the performance of the models: a) Accuracy: Accuracy is defined as the fraction of number of accurate predictions to the total number of input values. It performs well if the dataset is balanced and has identical number of samples that belong to each case as shown in Fig 4.4. In case the dataset has 97% of samples in class A and only 3% in class B, accuracy is not the best way to evaluate a model developed using this dataset.
The accuracy of ANN model developed was 75% whereas precision and recall score was 0.66 and 0.67 respectively.

b) AUC: AUC stands for Area Under ROC Curve. It is used for dual classification and its value lies between 0.5 and 1. ROC stands for Receiver Operating Characteristic which is a graph between true positive rate on x-axis and true negative rate on the y axis. As given in the graph below. AUC value of an ANN model for the Telecom industry dataset is 0.72.

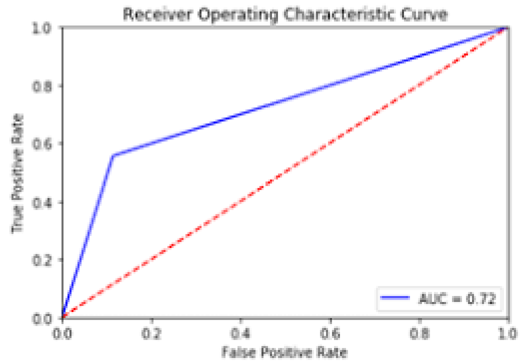


Fig. 4.3 ROC curve for ANN model

```
estimates_keras_tbl %>% roc_auc(truth, class_prob)
```

```
0.723951
```

c) Precision and Recall: Precision in the proportion of related outcomes whereas Recall is the proportion of total related outcomes given by model. Mathematically, precision is the fraction of true positive cases to the actual results that are false positive, true positive and false positive cases combined. Recall is the ratio of true positive to the predicted results that is false negative and true positive cases combined. The values of true positive, false negative, true negative, false positive, false negative, and true negative are given by the confusion matrix. Recall is also known as sensitivity and precision is also called positive predictive value. Both are used to understand and evaluate relevance. With reference to classification tasks, precision and recall are given by confusion matrix as given in figure below. The formula for precision and recall is given in Equation 4.1 and 4.2.

The figure below gives a visual representation of precision and recall values in a given dataset.

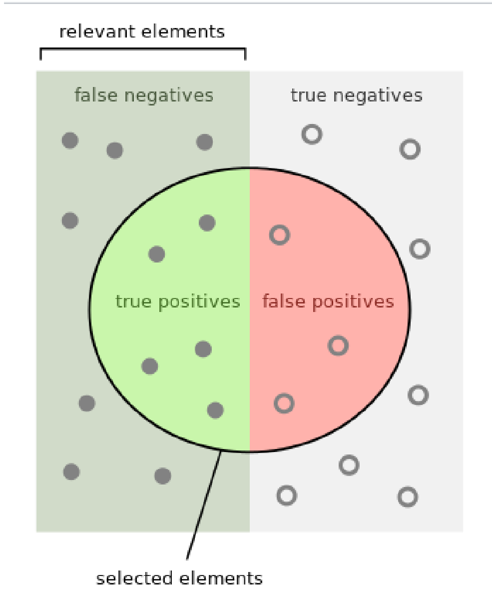


Fig 4.4 TP, FP, TN, FN (Source Wikipedia)[8]

Table 4.1 Confusion Matrix

		Predicted	
		Negative	**Positive**
Actual	**Negative**	True Negative	False Positive
	Positive	False Negative	True Positive

$$\text{Precision} = \frac{True\ Positive}{True\ Positive + False\ Positive}$$

Equation 4.1 Precision

$$\text{Recall} = \frac{True\ Positive}{True\ Positive + False\ Negative}$$

Equation 4.2 Recall

[8] https://en.wikipedia.org/wiki/Precision_and_recall

d) F1-Score: It is used to convey the stability between precision and recall. It is the HM (Harmonic Mean) of precision and recall. It has application in information retrieval systems, query classification, search and document classification systems. F-score is also used in machine learning and natural language processing systems such as in evaluating word segmentation and entity recognition. Its counterpart is G-measure which is the geometric mean of precision and recall. F-score is also used to evaluate multiclass classification problems. If F1 score is one, it is considered to be the best value. Formula for G mean is given in Equation 4.3.

$$F1 = 2 \times \frac{Precision * Recall}{Precision + Recall}$$

Equation 4.3 F1 Score

The comparison of final evaluation metrics have been summarised in Table 4.2:

Table 4.2 Comparison of Evaluation Metrics for Different ML Models

Algorithm	Accuracy Metric	F-1 Score	Precision Score	Recall Score	AUC
Logistic Regression (Unbalanced Dataset)	0.75	0.60	0.54	0.68	0.72
Logistic Regression (Balanced Dataset)	0.76	0.64	0.79	0.53	0.77
ANN	0.75	0.65	0.66	0.67	0.72
Random Forest	0.81	0.58	0.50	0.70	0.71
ADA Boost Classifier	0.847	0.85	0.851	0.848	0.847

XG Boost Classifier	0.87	0.867	0.862	0.872	0.866

As per the comparison table given in 4.2, it can be said that accuracy of the model improves after balancing the data using SMOTE and it further increases on using Boosting Ensemble methods like ADA Boost and XG Boost.

4.3 Exploratory Data Analysis:

EDA is an approach to analyse datasets and summarize its main findings through visualizations. It is an initial method to assess the data before data modelling or hypothesis testing. Multiple graphical techniques are used to perform EDA. Exploratory Data Analysis also involves grouping of data to get a holistic view of the data and extract information out of it. For example, if a dataset contains information of 1000 migrant workers along with the names of their hometown and the government has to arrange buses for these workers. As a whole the data will be very abruptly structed and will not provide a proper view of the areas associated with these workers. This is an effective way to visualize and summarize the data using less time consuming and easily interpretable techniques.

- Hypothesis testing
- Group By- Summarization
- Correlation
- Data Visualization (Boxplots)
- Histograms and Heat Maps

Correlation is a statistical method that gives the value of correlation i.e if the increase in value of one variable affects the increase in value of the other variable. A target variable can be positively or negatively correlated with some or all other variables. In case of no correlation the, correlation coefficient is 0 and it ranges from -1 to +1. Figure 5.1 gives the churn correlation analysis of target variable, churn with other features. It can be said that if two variables are correlated it does not mean that they are also causing each other’s effect. If the customer subscribing to a channel has increased with increasing rain in the area. It does not imply that the one factor is causing the occurrence of the other.

Correlation amongst variables is given below. It can be concluded that the variables are not highly correlated and thus PCA will not be performed in order to capture maximum variance in the data.

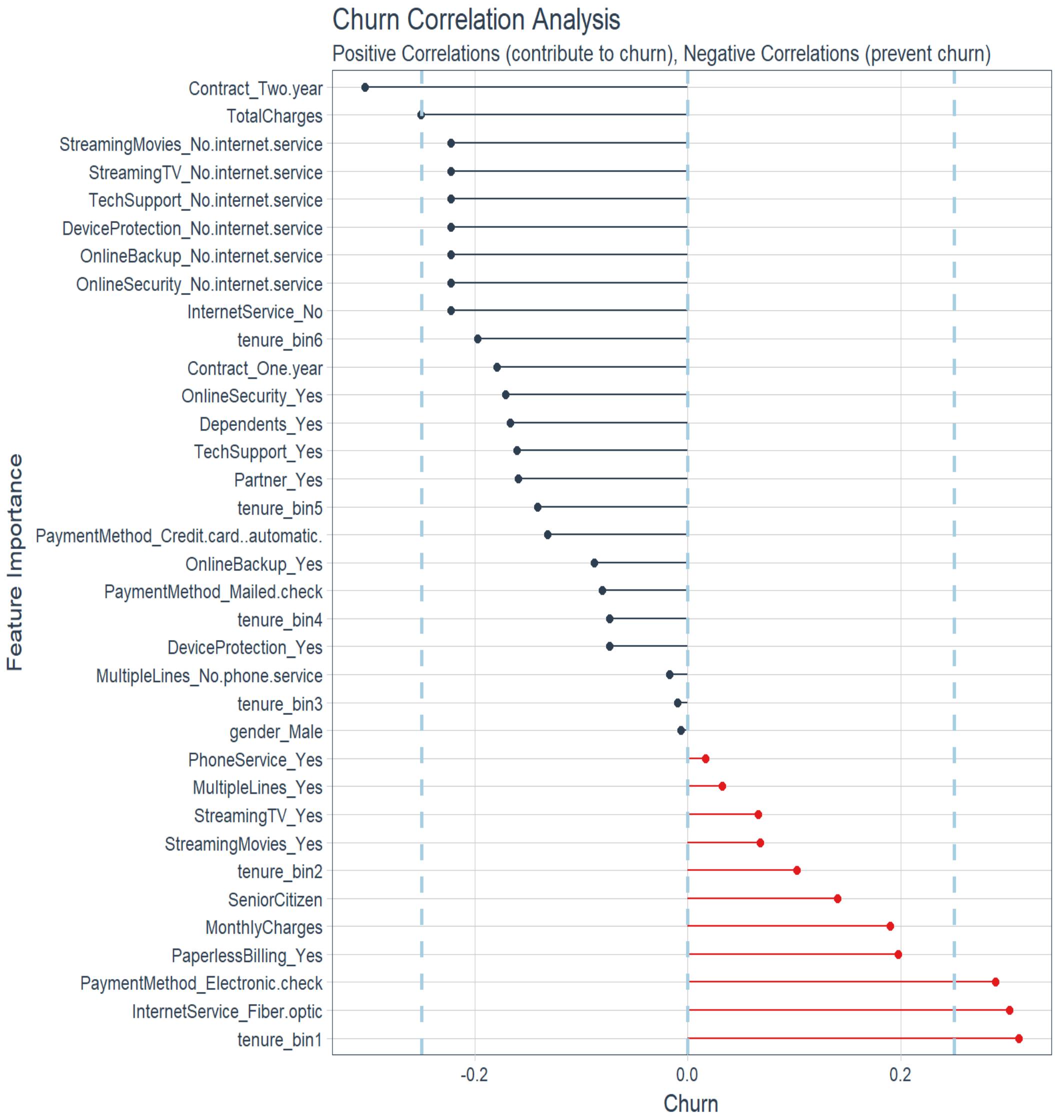


Figure 4.5 Correlation among variables in R

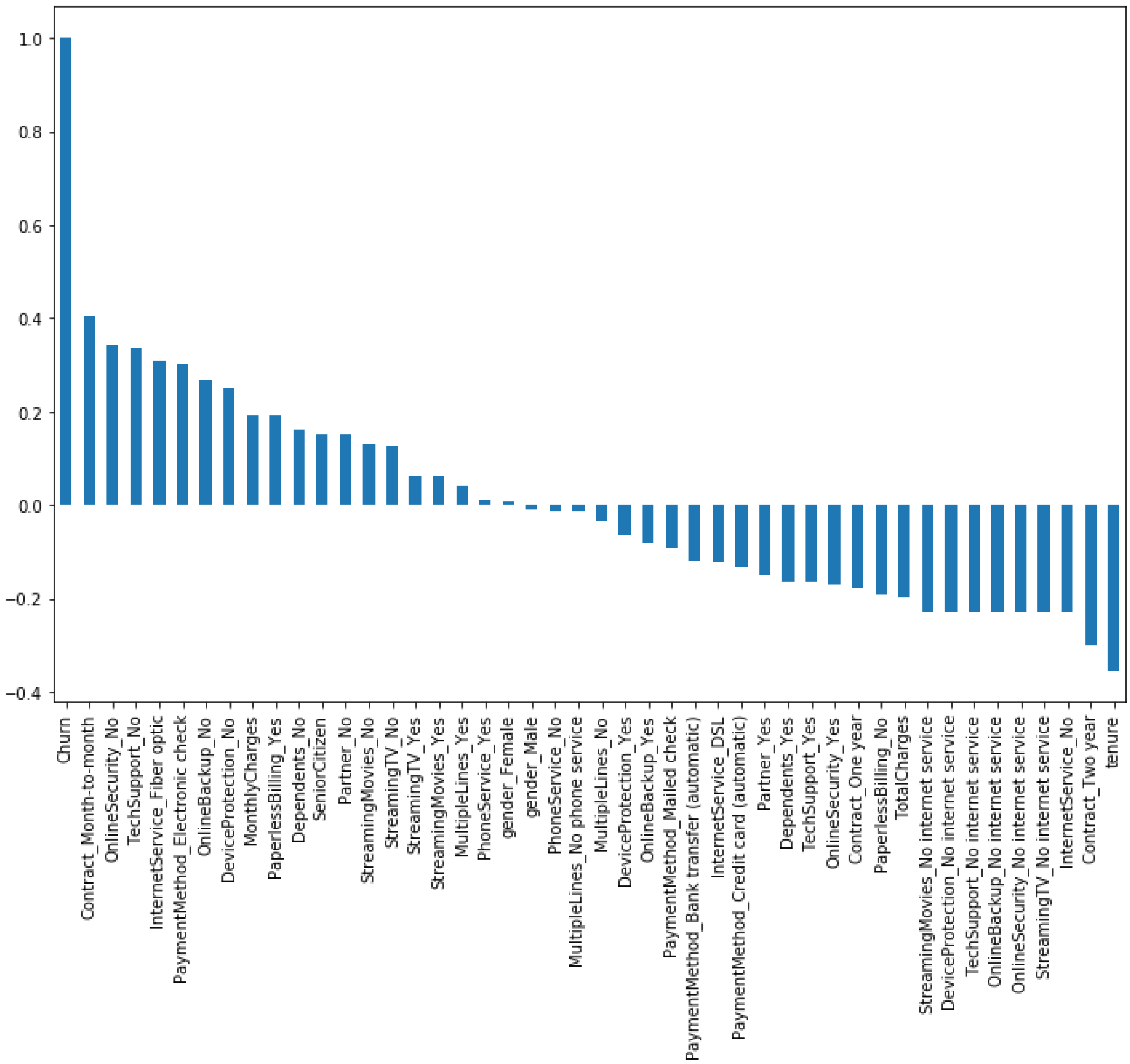
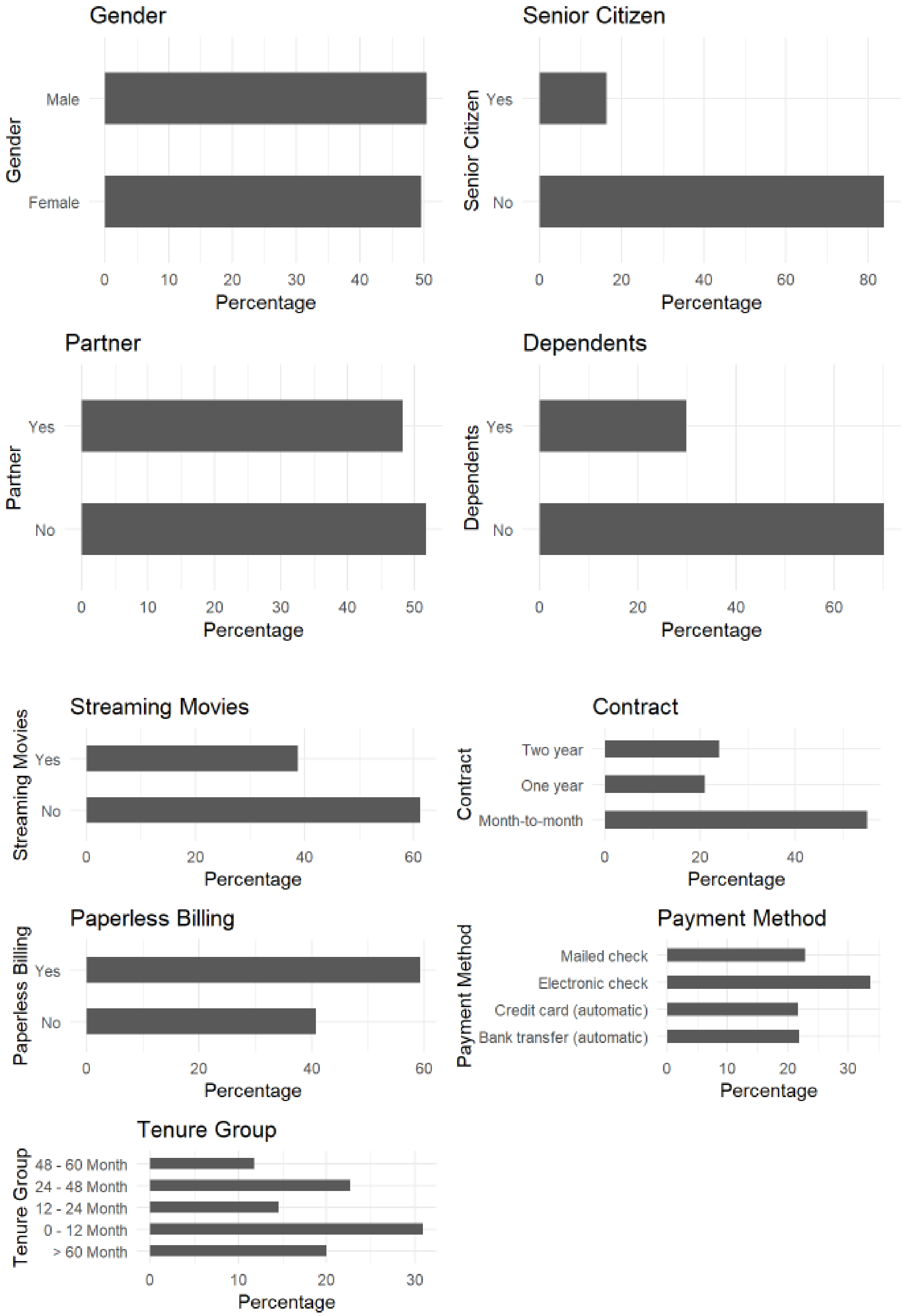


Fig 4.6 Correlation Visualization in Python

As illustrated in Fig 4.5 and 4.6, here, features like tenure, contract, Billing, Monthly Charges and Internet Service are important factors that contribute to customer attrition. As per the correlation analysis, the relation between gender and the target variable was found to be null whereas the correlation with other independent variables like tenure, PaymentMethod, and InternetService are positively correlated with churn. Customers with a contract for month-to-month basis and a tenure of less than a year are more susceptible to churning. Subscribers with longer contracts more than a year and tenure of more than twelve months are less likely to be susceptible to churn.

Some of the thorough findings of the analysis are given below:

Major tools include box plot, histogram, scatter plot, odds ratio, pareto chart, etc. Below given is the box plot for each categorical feature. Since all the categorical values have been found to have extensive categorization thus, they were considered for detailed analysis and data modelling. Most of the findings have been given by EDA techniques and correlation analysis.



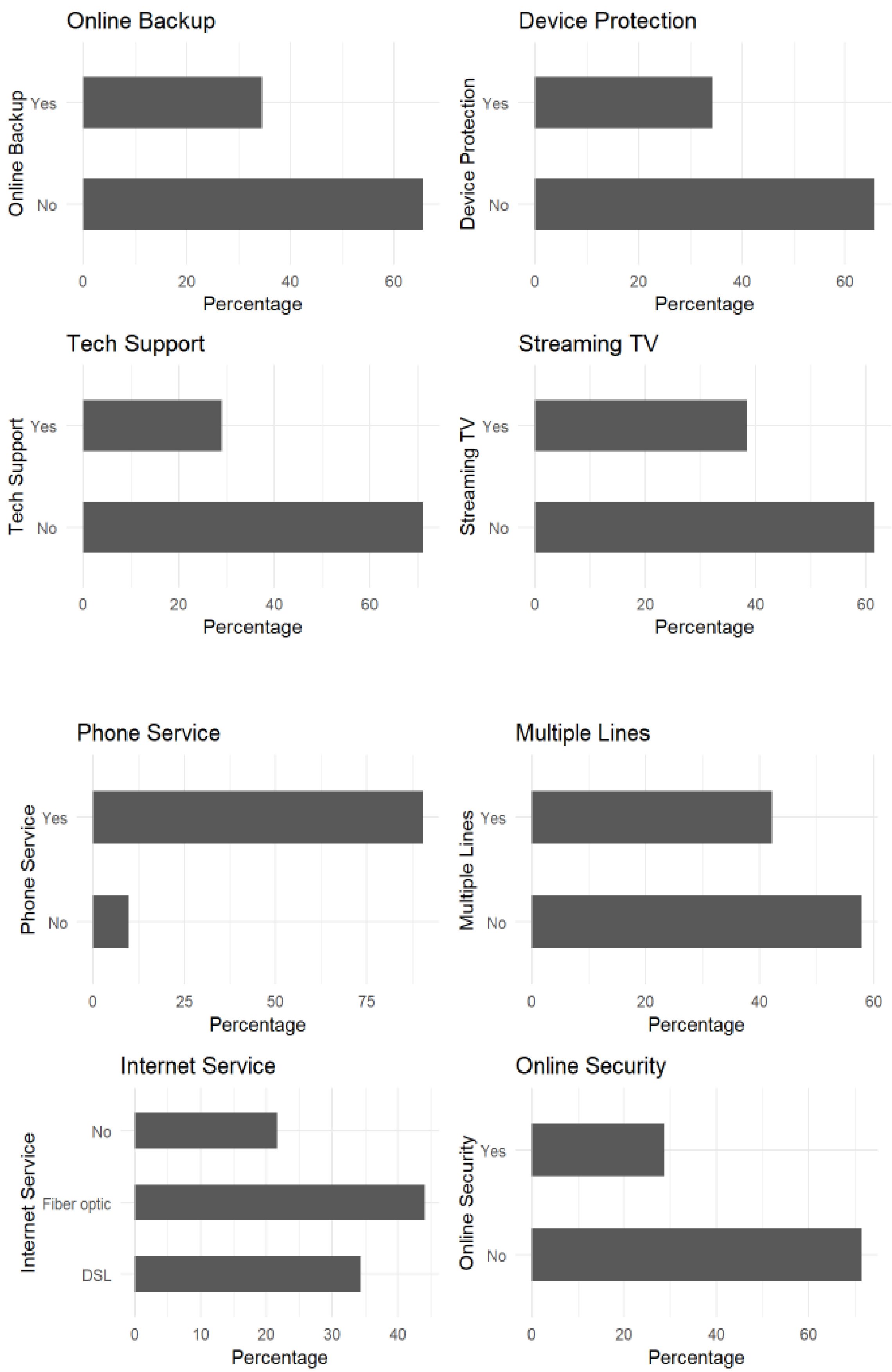


Fig 4.7 Exploratory Data Analysis

4.4 Recommendations:

1. According to the exploratory analysis in Fig 4.7, it was found that many customers with phone service churned. It can be inferenced that maybe these customers didn't really use the phone service. In order to retain such customers, it would help to move them to a plan without phone service so that they can save some money.
2. Customers with fibre optic internet were at the high risk of churning as compared to the subscribers who opted for no internet or DSL. Maybe moving some of those customers to DSL or eliminating their internet service would be an option. A promotion for continuing the fibre optic plan may help the customers to stay.
3. Further, it was also found that some customers without online security, backup, device protection did churn equally recurrently. It can be concluded that their devices may have crashed which has caused them to lose important files. They may have also experienced fraud or identity theft that has left them very unhappy. Moving these customers to some of these services may help safeguard their systems.
4. Similarly it was found that customers who opted for online backup and security, those without device protection tended to churn more than those that subscribed for the service.
5. It was also found that the customers who did not avail technical support were at high risk of leaving the company as compared to the customers who opted for technical support services. If these customers can be convinced to move to technical support accounts, it may be a possible way to prevent further churning.

A predictive model is given that provides ranks to the subscribers on the basis of their probability to drop the services and the revenue that they bring. The company can use this model to prioritize whose concerns to be addressed first. Sometimes it might be case by case basis.

Following actions can be taken by the company:

a) Try striking a longer contract with new customers: two year or one year in that order of preference.

b) Leverage the time to provide better service including the high cost ones like Fiber optic.

c) Improve on the Technical support on all services like streaming, phone connection and internet. Be up-to-date with current technology.

d) Collect customer feedback and act on it immediately to prevent new customer churn.

e) One could collect more data through surveys, analyse them using Natural Language Processing (NLP) techniques and take more measures.

f) There is also a scope to collect historical data on company customers over a few decades, and find out clear reasons for customer drop that happened 70 years ago.

CHAPTER 5
LIMITATIONS AND CONCLUSION

From Business perspective, it will be helpful to understand why churn started 5.5 years ago thus more historical data could have been given for analysis. From model perspective, the ANN model performed fine, better than Logistic Regression and Random Forest, but the performance could have been better. Some techniques including Hyper Parameter Tuning along with developing hybrid models could be used to improve the performance.

a) Hyper Parameter Tuning: A parameter is called hyper-parameter if its value is mentioned initially prior to the learning procedure. Multiple tuning strategies can be used for the same such as Grid Search (specific search across the given values) and Random Search (searches subset of hyperparameters randomly). It can be used to get the parameter values at which the model provides a better evaluation metric. One can avoid manually trying each parameter by using this technique. Initially, there was not much research done in this area but since the computing power of machines have improved, this can also be further explored.

b) Hybrid Models: Combining two ANN models or some other model has provided better results as given in some research papers. This can be explored along with parameter tuning.

It can be concluded that the objective of the project has been accomplished. The project analyzed IBM Telecom Dataset and developed a neural network-based algorithm along with ensemble models to envisage the customer attrition for the future time period. Also, it was found that with an accuracy of 85 percent, the Artificial Neural Network model performed better than basic models like Logistic regression and random forest. It can be said that the model performance highly depends on the quality of data that is being passed through it. Thus, data from some other sources can also be analyzed to get to a conclusion whether ANN performs better in all the situations. Some of the research paper also included clustering technique to cluster the customers on the basis of their risk of attrition, but due to time crunch, this could not be done during the course of this paper. The project report also discussed integration of Data Visualization tools like Power BI along with the results achieved. Being new to Deep Learning, only initial techniques could be used that did not involve hyper-parameter tuning and hybrid model technique. Initially Convolution Neural Network and Recurrent Neural Network were planned to be implemented on the dataset but due to less data availability, computing power and limited reliability of system, it was not possible given the time frame. Future prospects of the experiment will include feature engineering and using advanced deep learning concepts to find the customers who are at the verge of churning.

Some algorithms with deep architecture like generative and hybrid deep architecture can be used to improve the estimated accuracy of such problem sets. Additionally, Deep belief network and convolutional neural networks can provide better accuracy in terms of predicting the customer churn. Both of the above-mentioned algorithms can be explored for further analysis and future work.

REFERENCES

1.https://m.economictimes.com/industry/telecom/telecom-news/telecom-sector-revenue-slips-7-per-cent-in-2018-trai/articleshow/71299512.cms

2.https://www.technofunc.com/index.php/domain-knowledge/telecom-industry/item/overview-of-telecom-sector

3. J. Hadden, et al. "Churn prediction: Does technology matter." International Journal of Intelligent Technology, vol. 1, no. 2, pp. 104–110, 2006.

4. Nord, J. H., and Nord, G. D., "MIS research: Journal status and analysis", Information and Management, 1995, 29, 29–42

5. J. Hadden, et al. “Churn prediction: Does technology matter” International Journal of Intelligent Technology, voL 1, no. 2, pp. 104-110-2006

6. John Hadden et. al. “Computer Assisted Customer Churn Management: State -Of-The-Art and Future Trends”, Computers & Operations Research v34(10) October 2007, pp2902-2917.

7. Kim, M., Park, M., & Jeong, D. (2004). The effects of customer satisfaction and switching barrier on customer loyalty in Korean mobile telecommunication services. Telecommunications Policy, 28, 145–159

8. Burez, J., & Van den Poel, D. (2007). Crm at a pay-TV company: Using analytical models to reduce customer attrition by targeted marketing for subscription services. Expert Systems with Applications, 32, 277–288.

9. Tsai, C.-F., & Lu, Y.-H. (2009). *Customer churn prediction by hybrid neural networks. Expert Systems with Applications, 36(10), 12547–12553.* doi:10.1016/j.eswa.2009.05.032

10. https://towardsdatascience.com/predict-customer-churn-with-r-9e62357d47b4

11. https://towardsdatascience.com/hyperparameter-tuning-c5619e7e6624

12. Kohonen, T. (1987). Adaptive, associative, and self-organizing functions in neural computing. Applied Optics, 26(23), 4910–4918.

13. Coussement, K., & Van den Poel, D. (2008). Churn prediction in subscription services: An application of support vector machines while comparing two parameter-selection techniques. Expert Systems with Applications, 34, 313–327

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

15. Kim, H. S., & Yoon, C. H. (2004). Determinants of subscriber churn and customer loyalty in the Korean mobile telephony market. Telecommunications Policy, 28, 751–765

16. Amin, A., Al-Obeidat, F., Shah, B., Adnan, A., Loo, J., & Anwar, S. (2018). Customer churn prediction in telecommunication industry using data certainty. Journal of Business Research.

17. Y. Bengio, Deep Learning of Representations: Looking Forward, Statistical Language and Speech Processing, Berlin, Springer, 2013

18. Y. Bengio, A. Courville and P. Vincent, Representation Learning: A Review and New Percpectives, IEEE Transactions on Pattern Analysis & Machine Intelligence, vol. 35, no. 08, pp. 1798-1828, 2013.

19. Y. Xie, X. Li, E. Ngai and W. Ying, Customer churn prediction using improved balanced random forests, Expert Systems with Applications, vol. 36, no. 3, pp. 5445-5449, 2009.

20. B. Boser, I. Guyon and V. Vapnik, A training algorithm for optimal margin classifiers, Fifth Annual Workshop on Computational Learning Theory, New York, 1992.

21. L. Deng and R. Microsoft, A Tutorial Survey of Architectures, Algorithms, and Applications for Deep Learning, ASIPA Transactions on Signal and Information Processing, 2012.

22. F. Castanedo, G. Valverde, J. Zaratiegui and A. Vazquez, Using Deep Learning to Predict Customer Churn in a Mobile Telecommunication Networks, Wise Athena LLC, 2014. [Online]. Available: http://wiseathena. com/pdf/wa dl.pdf.

PLAGIARISM REPORT

Comparative Study on Machine Learning Algorithms for Customer Churn Analytics with Power BI Dashboard

ORIGINALITY REPORT

5%	2%	2%	4%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	arxiv.org Internet Source	1%
2	Tsai, C.F.. "Customer churn prediction by hybrid neural networks", Expert Systems With Applications, 200912 Publication	1%
3	pdfs.semanticscholar.org Internet Source	<1%
4	Submitted to Institute of Management Technology Student Paper	<1%
5	Submitted to University of Lancaster Student Paper	<1%
6	Submitted to De Montfort University Student Paper	<1%
7	m.economictimes.com Internet Source	<1%
8	Submitted to The University of Manchester Student Paper	<1%