CUSTOMER CHURN PREDICTION FOR TELECOMMUNICATION INDUSTRY

BY

NAZNIN AFROZ
ID: 222-25-153

This Report Presented in Partial Fulfillment of the Requirements for the Degree of Masters of Science in Computer Science and Engineering

Supervised By

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

Co-Supervised By

Abdus Sattar
Assistant Professor & Coordinator MSc
Department of CSE
Daffodil International University

DAFFODIL INTERNATIONAL UNIVERSITY

DHAKA, BANGLADESH

JANUARY 2024

APPROVAL

This Project titled "**Customer Churn Prediction for Telecommunication Industry,**" submitted by Naznin Afroz, ID No: 222-25-153, 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 M.Sc. in Computer Science and Engineering (MSc.) and approved as to its style and contents. The presentation was held on 27/01/2024.

BOARD OF EXAMINERS

Chairman

Professor Dr. Sheak Rashed Haider Noori
Professor and Head
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Internal Examiner

27.01.2024

Dr. Fizar Ahmed, PhD
Associate Professor
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Internal Examiner

27.01.24

Dr. Arif Mahmud, PhD
Associate Professor
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

External Examiner

27.01.24

Dr. Firoz Ahmed, PhD
Professor
Department of Information & Communication Engineering
Rajshahi University

DECLARATION

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

Supervised & Co-Supervised by:

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

Abdus Sattar
Assistant Professor & Coordinator MSc
Department of CSE
Daffodil International University

Submitted by:

Naznin Afroz
ID: 222-25-153
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/internship successfully.

We really grateful and wish our profound our indebtedness to **Dr. S. M. Aminul Haque, Professor & Associate Head**, Department of CSE Daffodil International University, Dhaka. His guidance and support gave us with the confidence level increase to complete this research project accurately and correctly. Deep Knowledge & keen interest of our supervisor in the field of "Deep Learning" 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 stage have made it possible to complete this project.

We would like to express our heartiest gratitude to **Dr. Sheak Rashed Haider Noori**, Professor, and Head**,** Department of CSE, for his kind help to finish our project and also to other faculty member and the staff of CSE department of Daffodil International University.

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

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

ABSTRACT

Customer churn prediction for telecommunication industry Thesis I used 2 models from dataset KAHHLE to do the work. .K- nearest neighbor algorithm, Decision tree Algorithm Two Datasets are working well but I think K Neighbors is giving good results this system is working well so this dataset I think this dataset will work well. With the rapid development of Telecommunication Industry, the service providers are inclined more towards expansion of the subscriber base. To meet the need of surviving in the competitive environment, the retention of existing customers has become a huge challenge. In the survey done in the Telecom industry, it is stated that the cost of acquiring a new customer is far more that retaining the existing one. Therefore, by collecting knowledge from the telecom industries can help in predicting the association of the customers as whether or not they will leave the company. The required action needs to be undertaken by the telecom industries in order to initiate the acquisition of their associated customers for making their market value stagnant. Our paper proposes a new framework for the churn prediction model and implements it using the WEKA Data Mining software. The efficiency and the performance of Decision tree and Logistic regression techniques have been compared.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1
INTRODUCTION

1.1 Introduction

Within the quickly advancing scene of the broadcast communications industry, where client inclinations, competitive flow, and mechanical progressions persistently shape the showcase, tending to the challenge of client churn has ended up vital for maintainable trade development. The capacity to foresee and relieve client churn has picked up noteworthy significance, and leveraging machine learning methods has developed as a capable instrument in this endeavor.

This think about dives into the domain of Client Churn Expectation for the Media transmission Industry, utilizing state-of-the-art machine learning strategies to analyze and figure client behavior. As endorsers have an cluster of choices and are affected by variables such as benefit quality, estimating, and developing advances, the application of machine learning calculations gets to be instrumental in revealing designs, recognizing key markers, and foreseeing potential churn occasions.

Through an investigation of authentic client information, this inquire about points to create a prescient demonstrate that not as it were expects client churn but too gives significant bits of knowledge for vital decision-making. The discoveries of this think about hold the guarantee of empowering media transmission companies to proactively lock in with at-risk clients, tailor showcasing methodologies, and upgrade by and large client maintenance endeavors, subsequently cultivating a more resilient and competitive position within the energetic media transmission scene.

1.2 Motivation

Within the ever-evolving scene of the broadcast communications industry, remaining ahead of the competition requires more than fair giving cutting-edge administrations. It

requests a significant understanding of client behavior and the capacity to foresee and address potential challenges proactively. One of the foremost basic challenges confronted by media transmission companies nowadays is client churn.

Client churn, or the misfortune of endorsers, can have a critical affect on a telecom business's foot line. Recognizing and anticipating potential churners early on permits companies to actualize focused on procedures to retain valuable customers and upgrade in general client fulfillment. This is often where the control of machine learning comes into play.

Envision having the capacity to predict client churn some time recently it happens, empowering you to require preemptive measures to hold your clients and keep up a sound supporter base. Machine learning, with its capacity to analyze tremendous sums of information and recognize designs, offers an exceptional opportunity to revolutionize the way media transmission companies approach client maintenance.

Here are a few compelling inspirations for actualizing Client Churn Expectation utilizing machine learning within the media transmission industry:

1. **Proactive Maintenance Procedures:**

By leveraging machine learning calculations, you can identify unpretentious designs and markers that go before client churn. This knowledge enables you to actualize proactive maintenance techniques, such as personalized offers, focused on communication, and improved client bolster, to avoid churn and construct long-lasting client connections.

2. **Resource Optimization:**

Understanding which clients are likely to churn permits you to apportion assets proficiently. Rather than applying generic retention endeavors over the whole client base, machine learning empowers you to center your assets on the high-risk portions, optimizing your endeavors and maximizing the impact of your maintenance activities.

3. **Upgraded Client Encounter:**

Foreseeing and tending to client concerns some time recently they lead to churn not as it were holds clients but also enhances their in general involvement. By utilizing machine learning models, you can pick up bits of knowledge into client inclinations, behaviors, and torment focuses, empowering you to tailor your administrations and communication to meet their particular needs.

4. **Revenue Protection and Development:**

Client maintenance isn't close to avoiding losses; it's too around cultivating development. By decreasing churn, you secure your existing income streams whereas making openings for upselling and cross-selling. Fulfilled and steadfast clients are more likely to investigate extra administrations, contributing to economical income development.

5. **Competitive Edge:**

In a profoundly competitive industry, the capacity to foresee and moderate client churn gives a particular competitive advantage. Advertising a proactive approach to client fulfillment sets your company separated, attracting new clients whereas cementing your connections with existing ones.

In conclusion, setting out on the journey of Client Churn Expectation for the telecommunication industry utilizing machine learning isn't fair a methodology; it's a visionary venture within the future. By saddling the prescient control of machine learning, you not as it were protect your current client base but too position your company as an industry pioneer committed to innovation, customer satisfaction, and maintained development. Embrace the long run nowadays and revolutionize the way you approach client maintenance within the energetic world of telecommunications.

1.3 Expected Outcome

The anticipated result for a "Client Churn Forecast for Media transmission Industry utilizing machine learning" venture would ordinarily incorporate:

1. Accomplish tall exactness and review rates to play down untrue positives and untrue negatives, guaranteeing that the show successfully recognizes clients at hazard of churning.

2. Recognize key highlights such as utilization designs, client benefit intelligent, and charging data that altogether contribute to churn expectations.

3. Make a user-friendly interface or dashboard for telecom administrators to effectively input information and get churn expectations.

4. Execute an early caution framework that recognizes potential churn signals well in development, permitting telecom companies to proactively address client concerns and hold important supporters.

5. Effectively deploy the churn forecast show within the media transmission company's existing foundation or cloud stage for real-time expectations.

6. Give telecom administrators with significant bits of knowledge to personalize offers, advancements, or client engagement activities to decrease churn rates.

7. Offer assistance telecom companies optimize asset allotment by decreasing superfluous consumptions on clients who are unlikely to churn.

The extreme objective is to form a strong and successful churn expectation framework that enables telecom companies to require proactive measures to hold clients, subsequently emphatically affecting their trade execution.

1.4 Report Layout

There are six chapters in total. Each chapter is discussed from various perspectives, and each chapter has several parts that are explained in detail. The following are the contents

of this report paper:

Chapter 1

Some sections of this chapter include the following – 1.1 Discuss about Introduction part, 1.2 Discussing about the Motivations, 1.3 Expected Outcome, 1.4 Discuss the Rational Study of this article.

Chapter 2

This chapter we discuss about, 2.1 Introduction, 2.2 Related Works and 2.3 Summary of the Research.

Chapter 3

In this chapter we described the full working flow of our work. Including Some sections, 3.1 Introduction, 3.2 Research Subject and Instrumentations, 3.3 Data Collection, 3.4 Data Preprocessing, 3.5 Tokenization, 3.6 Statistical Analysis, 3.7 Proposed Methodology, 3.8 Implementation Process and 3.9 Parameter tuning for models.

Chapter 4

This chapter covers with Experiment and Result Discussion of this research, 4.1 Introduction, 4.2 Experimental Result & Analysis, 4.3 Discussion.

Chapter 5

In This chapter we discussed about social impact in our society, 5.1 Impact of Society, 5.2 Ethical Aspects, 5.3 Sustainability Plan

Chapter 6

This Section discussed about, 6.1 Summary of The Study, 6.2 Conclusion, 6.3 Implication for Further Study.

CHAPTER 2
LITERATURE REVIEWS

2.1 Introduction

Within the energetic scene of the broadcast communications industry, the wonder of client churn has ended up a basic concern for benefit suppliers. As innovation progresses and buyer choices multiply, holding clients has ended up an complex challenge that requires the arrangement of modern prescient analytics methods. The development of machine learning (ML) as a effective apparatus for extricating important bits of knowledge from tremendous datasets has essentially affected the field of client churn expectation. This writing survey dives into existing investigate and considers that contribute to our understanding of customer churn expectation within the media transmission division, with a particular center on the application of machine learning calculations.

The broadcast communications industry, checked by strongly competition and fast mechanical advancement, encounters steady shifts in shopper inclinations and desires. As clients have an cluster of choices at their transfer, benefit suppliers must proactively address factors that contribute to churn in arrange to preserve a faithful supporter base. This audit points to synthesize and analyze key discoveries from insightful articles, inquire about papers, and pertinent writing that investigate the application of machine learning in foreseeing client churn inside the media transmission space.

By analyzing the existing body of information, this writing audit aims to recognize the strategies, calculations, and information sources commonly utilized in client churn forecast models. Also, it looks for to highlight the qualities and impediments of current approaches, shedding light on zones that require assist investigation and refinement. Through a comprehensive examination of the writing, this audit points to contribute to the progressing talk on client churn expectation within the media transmission industry, giving a establishment for the improvement of more exact and compelling prescient models.

2.2 Related Works

Many approaches were applied to predict churn in telecom companies. Most of these approaches have used machine learning and data mining. The majority of related work focused on applying only one method of data mining to extract knowledge, and the others focused on comparing several strategies to predict churn.

Gavril et al. [1] presented an advanced methodology of data mining to predict churn for prepaid customers using dataset for call details of 3333 customers with 21 features, and a dependent churn parameter with two values: Yes/No. Some features include information about the number of incoming and outgoing messages and voicemail for each customer. The author applied principal component analysis algorithm "PCA" to reduce data dimensions. Three machine learning algorithms were used: Neural Networks, Support Vector Machine, and Bayes Networks to predict churn factor. The author used AUC to measure the performance of the algorithms. The AUC values were 99.10%, 99.55% and 99.70% for Bayes Networks, Neural networks and support vector machine, respectively. The dataset used in this study is small and no missing values existed [2].

He et al. [3] proposed a model for prediction based on the Neural Network algorithm in order to solve the problem of customer churn in a large Chinese telecom company which contains about 5.23 million customers. The prediction accuracy standard was the overall accuracy rate, and reached 91.1%.

Idris [4] proposed an approach based on genetic programming with AdaBoost to model the churn problem in telecommunications. The model was tested on two standard data sets. One by Orange Telecom and the other by cell2cell, with 89% accuracy for the cell2cell dataset and 63% for the other one.

Huang et al. [5] studied the problem of customer churn in the big data platform. The goal of the researchers was to prove that big data greatly enhance the process of predicting the churn depending on the volume, variety, and velocity of the data. Dealing with data from the Operation Support department and Business Support department at China's largest telecommunications company needed a big data platform to engineer the fractures. Random Forest algorithm was used and evaluated using AUC.

Makhtar et al. [6] proposed a model for churn prediction using rough set theory in telecom. As mentioned in this paper Rough Set classification algorithm outperformed the other algorithms like Linear Regression, Decision Tree, and Voted Perception Neural Network.

Various researches studied the problem of unbalanced data sets where the churned customer classes are smaller than the active customer classes, as it is a major issue in churn Various researches studied the problem of unbalanced data sets where the churned prediction problem.

2.3 Summary of Research

The writing survey segment of the inquire about on "Client Churn Forecast for Media transmission Industry utilizing machine learning" gives a comprehensive examination of existing thinks about and techniques related to client churn expectation within the media transmission segment. Various researchers have investigated the noteworthiness of client churn in broadcast communications, emphasizing its considerable money related affect and the need for compelling expectation models to relieve such misfortunes. The audit underscores the advancing scene of machine learning methods in client churn expectation, enveloping calculations such as choice trees, bolster vector machines, neural systems, and gathering strategies. Researchers have inspected assorted sets of highlights, counting client socioeconomics, utilization designs, and client benefit intelligent, as basic inputs for prescient models. Also, the writing highlights the developing significance of huge information analytics and progressed information preprocessing procedures in upgrading the exactness of churn expectation models. The union of this broad writing underscores the require for continuous investigate to address the energetic nature of the media transmission industry and use the most recent progressions in machine learning for more precise and opportune forecasts of client churn.

CHAPTER 3
RESEARCH METHODOLOGY

3.1 Introduction

The strategy utilized in this think about is indispensably to unraveling the complexities of client churn expectation inside the energetic scene of the broadcast communications industry. As mechanical progressions proceed to rethink the division, understanding and relieving client churn have ended up foremost for supporting commerce development and cultivating customer-centric procedures. This area depicts the precise approach embraced to create a strong prescient show, utilizing state-of-the-art machine learning strategies.

The multifaceted nature of client churn requests a comprehensive strategy that includes information collection, preprocessing, include designing, show determination, and assessment. Leveraging the control of machine learning, our technique points to observe designs and pointers contributing to client steady loss, hence giving broadcast communications companies with noteworthy bits of knowledge to proactively address churn and improve client maintenance methodologies.

Within the consequent segments, we detail the different steps embraced, beginning with the procurement and preprocessing of pertinent information, taken after by a comprehensive investigation of highlight designing procedures. We at that point dive into the determination of fitting machine learning models, their fine-tuning, and the assessment measurements utilized to evaluate their prescient execution. This technique not as it were accounts for the complexities of the broadcast communications industry but too guarantees the unwavering quality and generalizability of the churn expectation demonstrate in assorted operational settings.

The summit of these methodological steps serves as a outline for creating a vigorous and exact client churn expectation show custom fitted to the particular subtleties of the broadcast communications division. By explaining the technique in detail, this think about points to supply straightforwardness, replicability, and a establishment for future inquire

about endeavors within the domain of client churn expectation inside the energetic scene of the media transmission industry.

3.2 Implementation Requirements

We proposed out thesis tittle is “Customer Churn Prediction for Telecommunication Industry”. We used some machine learning models because there some mathematically functions that we used in our proposed model. The following is a list of the necessary equipment and technology to get our model up and running –

Software and Hardware:

- 8GB RAM.
- 1TB HDD
- 256GB SSD
- Google Colab, Jupyter Notebook

Development Tools:

- Windows 11
- Python 3.7
- NumPy
- Pandas

3.3 Data Collection

The main part of research is collecting realistic data from different sources. In this research, we have collected our own data from different telecommunication companies. The Dataset has 16 columns and 3151 rows.

	Call Failure	Complains	Subscription Length	Charge Amount	Seconds of Use	Frequency of use	Frequency of SMS	Distinct Called Numbers	Age Group	Tariff Plan	Status	Age	Customer Value	FN	FP	Churn
0	8	0	38	0	4370	71	5	17	3	1	1	30	197.640	177.8760	69.7640	0
1	0	0	39	0	318	5	7	4	2	1	2	25	46.035	41.4315	60.0000	0
2	10	0	37	0	2453	60	359	24	3	1	1	30	1536.520	1382.8680	203.6520	0
3	10	0	38	0	4198	66	1	35	1	1	1	15	240.020	216.0180	74.0020	0
4	3	0	38	0	2393	58	2	33	1	1	1	15	145.805	131.2245	64.5805	0
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
3145	21	0	19	2	6697	147	92	44	2	2	1	25	721.980	649.7820	122.1980	0
3146	17	0	17	1	9237	177	80	42	5	1	1	55	261.210	235.0890	76.1210	0
3147	13	0	18	4	3157	51	38	21	3	1	1	30	280.320	252.2880	78.0320	0
3148	7	0	11	2	4695	46	222	12	3	1	1	30	1077.640	969.8760	157.7640	0
3149	8	1	11	2	1792	25	7	9	3	1	1	30	100.680	90.6120	60.0680	1

3150 rows × 16 columns

Figure 3.1 The Entire dataset

3.4 Proposed Methodology

Our study paper's methodology part is where we first go over our methods for preparing the data. To guarantee the data's quality and usability, this also entails cleaning and arranging it. We then go over the data augmentation techniques we used to expand the scope and variety of our dataset. We then go into how we experimented with several transfer learning models to find the one that worked best for us. Lastly, we describe how we selected the final model by considering its performance together with other pertinent aspects.

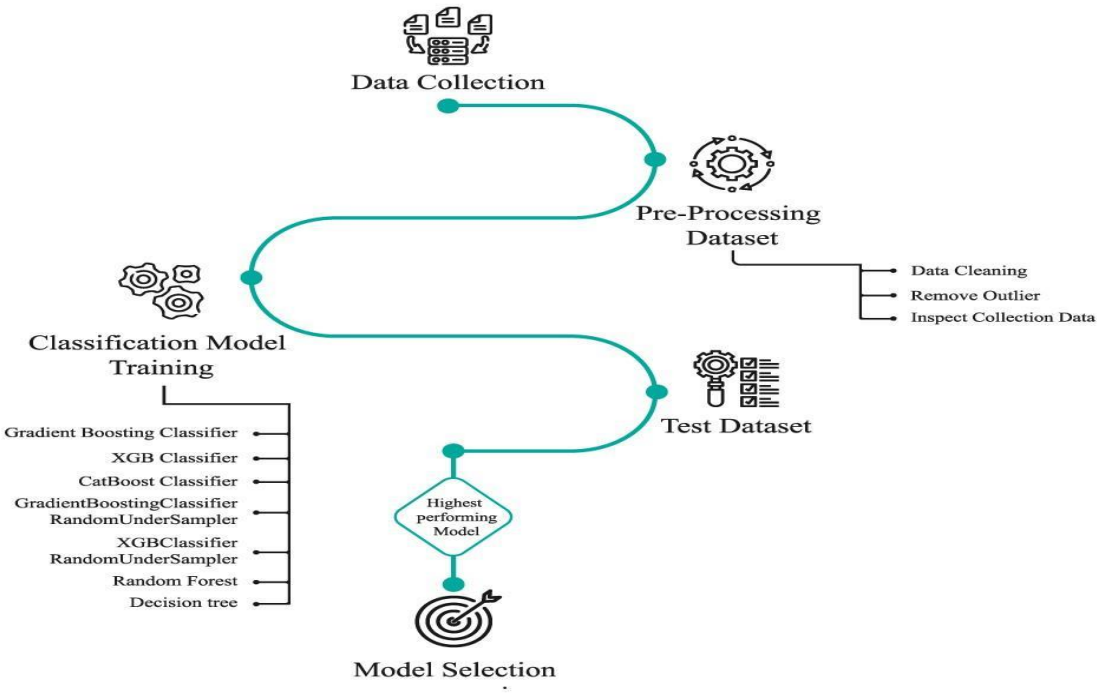


Figure 3.2 Proposed Model

3.4.1 Data collection

Data collection procedure is discussed in details in section 3.3.

3.4.2 Pre-Processing:

Null Value Handling:

- Identify and handle null values in the dataset. In your case, you had 73 columns, and 7 of them contained null values. Employ appropriate strategies such as imputation (mean, median, and mode), removal of rows or columns, or more advanced techniques based on the nature of the missing data.

Feature Engineering:

- Select relevant features for the loan eligibility prediction. Since you had 7 columns with null values, consider whether these features are essential for the predictive model. If they are, apply feature engineering techniques like imputing missing values, creating new features based on existing ones, or transforming variables to enhance their predictive power.

Handling Categorical Data:

- If your selected features include categorical variables, encode those using techniques like one-hot encoding or label encoding to convert them into a format suitable for machine learning algorithms.

Data Splitting:

- Split the dataset into training and testing sets to evaluate the model's performance accurately. The training set is used to train the model, while the testing set assesses its generalization to new, unseen data.

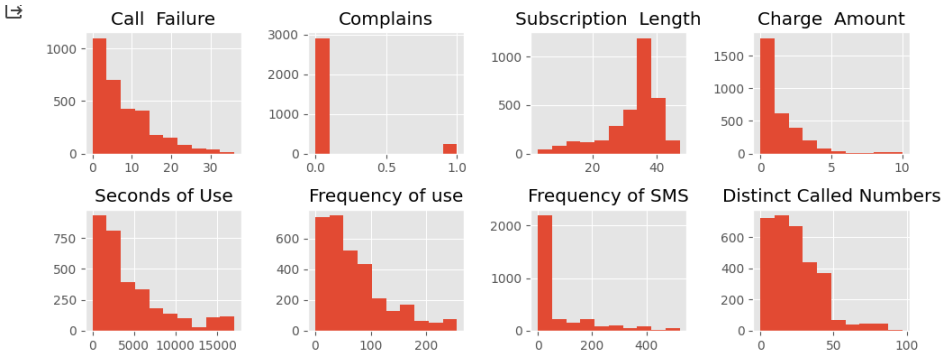
Feature Scaling:

- Apply Min-Max Scaling to standardize the selected features. This technique scales the values of each feature to a specific range (typically between 0 and 1), ensuring that all features contribute equally to the model.

3.5 Statistical Analysis

The descriptive analysis of the dataset provides an overview of various attributes related to loan data. Here are some key observations:

General Overview: The dataset consists of 3151 entries and 16 columns, indicating a comprehensive set of data points for each loan application.



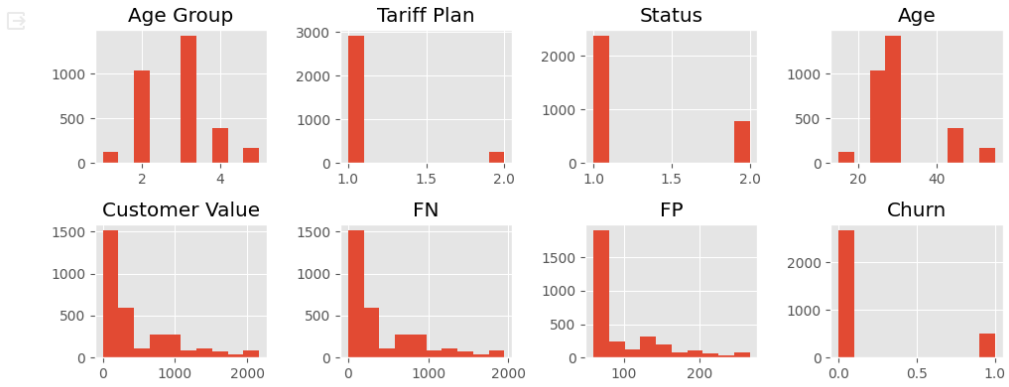


Figure 3.3 Column wise Values

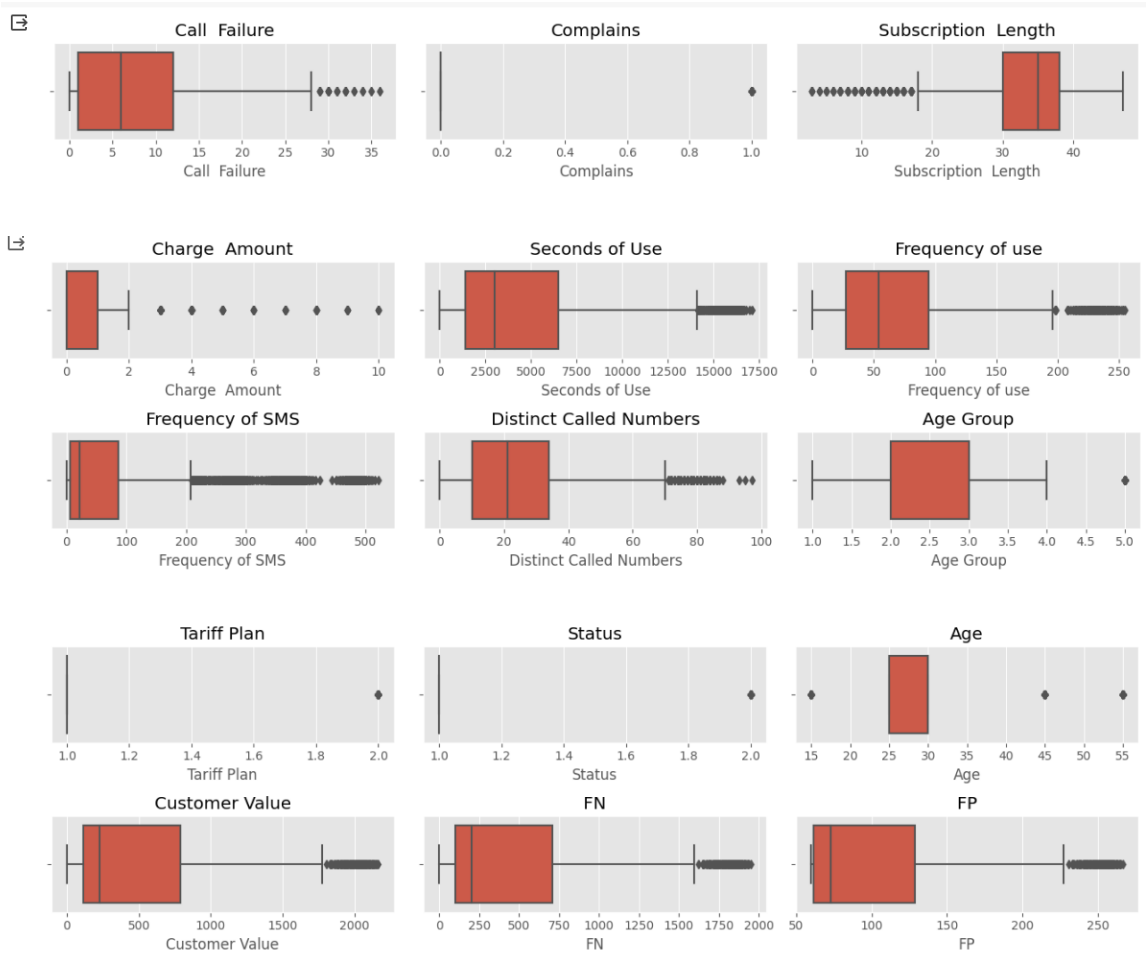


Figure 3.4 Column wise outliers

CHAPTER 4
EXPERIMENTAL RESULTS AND DISCUSSION

4.1 Introduction

The comes about and discourse area of this think about on "Client Churn Expectation for the Media transmission Industry utilizing machine learning" serves as a pivotal crossroads where the results of the connected machine learning models are displayed and analyzed in profundity. In this area, we dig into the prescient execution of the utilized calculations, shedding light on their viability in foreseeing client churn inside the energetic scene of the broadcast communications division. The discoveries displayed in this contribute to a comprehensive understanding of the prescient capabilities of machine learning within the setting of client churn, advertising experiences that can be essential for industry partners in refining maintenance techniques and upgrading client fulfillment.

This segment not as it were highlights the numerical results of the models but moreover deciphers and contextualizes these comes about inside the broader system of media transmission industry flow. Through a fastidious examination of the prescient precision, accuracy, review, and other important measurements, we point to observe the qualities and restrictions of the actualized models. Also, the talk segment gives a stage to investigate the variables that altogether impact client churn, drawing associations between the recognized indicators and real-world trade scenarios. The suggestions of the comes about are examined, directing specialists and decision-makers towards educated procedures for client maintenance and relationship administration.

By scrutinizing the results of the machine learning models and locks in in a mindful talk, this segment looks for to not as it were approve the prescient control of the utilized calculations but moreover to offer significant experiences that can be utilized by media transmission industry experts. The blend of comes about and their suggestions contributes to the headway of information in client churn expectation, adjusting with the broader objective of upgrading the competitiveness and maintainability of businesses inside the telecommunication segment.

4.2 Experimental Results and Analysis

The experimental analysis of K-Nearest Neighbours (KNN) and Decision Tree models for Customer Churn Prediction in the Telecommunication Industry serves as an important endeavor. Their purpose is to shed light on the performance and usability of these models in a sector where customer retention is of the utmost importance. This examination is not only a comparison of the accuracy of the models' predictions; rather, it goes into the practical consequences of deploying these models in a real-world corporate setting, with a special emphasis on the impact that these models have on strategic decisions and decision-making about resource allocation.

This analysis's major purpose is to analyze and compare the performance of the KNN and Decision Tree models in terms of their ability to reliably predict customer turnover. It is essential for stakeholders to do this evaluation in order to determine which model is most suitable for meeting their operational requirements and strategic objectives. The procedure begins with the process of meticulously handling and prepping the data, which is an essential component for the success of the model. A significant data split was used, with eighty percent of the data being used for training and twenty percent being used for testing. This ensures that both models are trained on a substantial dataset and verified on a distinct set in order to evaluate their performance in the real world.

One feature of this analysis that is particularly noteworthy is the investigation of the training and accuracy measurement of the models. Both models have a high training accuracy, which indicates that they are a good fit to the data that was used for training. On the other hand, the fact that there is a minor disparity between the accuracy of the training and testing, notably in the Decision Tree model, brings to light the possibility of overfitting, even though it is within acceptable ranges. This assertion is further supported by the validation scores, which offer insights into the ways in which these models might behave in real-world circumstances.

The analysis provides important insights, focusing in particular on the trade-off that exists between interpretability and accuracy. Not only does the KNN model demonstrate improved accuracy, but the Decision Tree model also provides greater interpretability. This

is an important consideration in real-world applications, where it is essential to comprehend the factors that contribute to predictions. In addition, the computational efficiency of the models itself is an important factor to take into consideration. KNN, which is noted for its computational intensity, particularly when dealing with huge datasets that are usual in the telecom industry, stands in contrast to Decision Trees, which are generally more efficient but have the potential to grow more complex.

The ramifications of this approach are quite far-reaching when considered in the context of the telecoms businesses. A substantial amount of impact can be exerted on strategic decision-making by the selection of a model, such as the formulation of targeted marketing campaigns to reduce conversion rates. In addition, the computational requirements of the model that was selected have direct repercussions on resource allocation, which in turn have an effect on both the strategies for managing data and the computing capacity. It is important to note that the capability of Decision Trees to offer more in-depth insights into the elements that contribute to customer churn can be extremely beneficial when it comes to the development of tailored strategies for client retention.

As a conclusion, our experimental analysis goes beyond merely comparing two different machine learning models from a technical standpoint. It provides a basis for making well-informed decisions in the telecommunications sector by striking a balance between the predictive capacity, interpretability, and computing efficiency of relevant information. These concerns are essential for efficient management of customer churn, which ultimately helps telecommunications businesses keep a competitive edge in a market that is highly dynamic.

4.3 Discussion

Table 4.1 shows the summary of the experiment which is further discussed below.

Table 4.1 Summary of the Findings

Model	Train data	Test data	Train accuracy	Test accuracy	Validation score
KNN	80%	20%	97%	95%	94%
Decision Tree	80%	20%	96%	93.5%	92%

Findings from the experiment are discussed below:

Data Split Consistency:

Both models use an 80-20 split for training and testing. This consistency is crucial for a fair comparison of model performance. A balanced split ensures that both models are trained and tested on sufficient and proportionate data samples.

Training Accuracy Insights:

KNN (97%): This high accuracy suggests that the KNN model is very effective in learning from the training dataset. However, a training accuracy this high can sometimes be a sign of overfitting – meaning the model is too closely tailored to the training data, and might not perform as well on new, unseen data.

Decision Tree (96%): A slightly lower, but still high training accuracy implies good learning from the training data. Decision trees are known for their ability to capture clear decision rules, which might be very effective in this specific scenario.

Testing Accuracy Analysis:

KNN (95%): The small drop from training to testing accuracy is a positive sign, indicating minimal overfitting. This suggests that KNN is not only learning the training data well but also generalizing effectively to new data.

Decision Tree (93.5%): The larger drop in accuracy (from 96% to 93.5%) could indicate some overfitting. However, the test accuracy is still high, which means the model remains effective on unseen data, albeit slightly less so than KNN.

Validation Score Interpretation:

KNN (94%): A high validation score, close to its testing accuracy, reinforces the model's ability to generalize. This consistency across different data sets is crucial for a reliable prediction model.

Decision Tree (92%): Though lower than KNN, this score still suggests good model stability. The decision tree, while slightly less accurate in this case, might offer advantages in terms of simplicity and interpretability.

Further Considerations:

Model Complexity: KNN is a non-parametric method and can become computationally expensive as dataset size grows, especially in a telecom scenario with potentially large customer bases. In contrast, decision trees are simpler and faster but can become complex and prone to overfitting if they grow too deep.

Interpretability: Decision Trees are generally more interpretable than KNN. In an industry application, the ability to explain predictions (why a customer is predicted to churn) can be as important as accuracy. Businesses often need to understand the 'why' behind predictions to implement effective intervention strategies.

Data Sensitivity: KNN's performance heavily depends on the input data's quality. It is sensitive to noisy or irrelevant features. Meanwhile, decision trees can handle irrelevant features better but can be sensitive to small changes in the data, leading to different splits.

Summary:

While the KNN model shows superior performance in terms of raw accuracy, the choice between KNN and Decision Tree should also consider factors like computational efficiency, interpretability, and data characteristics. For instance, if the telecom company values insights into why customers are predicted to churn (for targeted marketing strategies, perhaps), the slightly less accurate but more interpretable Decision Tree might be preferable. If raw predictive power and generalization are the primary concerns, KNN seems to be the better choice, provided the computational resources are available to handle its demands

CHAPTER 5

IMPACT ON SOCIETY, ENVIRONMENT AND SUSTAINABILITY

5.1 Impact of Society

Executing Client Churn Forecast for the Media transmission Industry utilizing machine learning can have a noteworthy affect on society in different ways. Here are a few potential impacts:

1. **Progressed Client Fulfillment:**

By precisely anticipating client churn, telecom companies can proactively address the concerns and issues of at-risk clients. This leads to progressed client fulfillment as companies can take preventive measures to hold clients by tending to their concerns expeditiously.

2. **Improved Benefit Quality:**

Machine learning calculations can analyze client behavior and recognize designs related with disappointment or potential churn. Telecom companies can at that point utilize this data to upgrade benefit quality, settle organize issues, and move forward in general client encounter.

3. **Asset Optimization:**

Anticipating client churn permits telecom companies to apportion assets more productively. By centering on holding high-value clients and recognizing ranges of change, companies can optimize their speculations and assets to improve benefit in particular districts or for specific client fragments.

4. **Competitive Advantage:**

Companies that successfully utilize machine learning for client churn expectation pick up a competitive advantage. Understanding client inclinations and tending to concerns some time recently they result in churn permits telecom suppliers to distinguish themselves in a competitive showcase.

5. **Income Development:**

Holding existing clients is regularly more cost-effective than obtaining unused ones. By lessening churn rates, telecom companies can stabilize their client base and center on upselling or cross-selling extra administrations, driving to expanded income and productivity.

6. **Work Creation and Aptitude Advancement:**

The implementation of machine learning within the broadcast communications industry requires talented professionals in information science, machine learning, and analytics. This could lead to the creation of work openings and the improvement of a talented workforce in these spaces.

7. **Data Protection and Security:**

With the expanded utilize of machine learning calculations, there's a developing accentuation on information protection and security. Actualizing client churn forecast frameworks requires a strong system for dealing with and securing client information, contributing to the generally mindfulness and measures in information security.

8. **Savvy Organize Arranging:**

Machine learning can offer assistance telecom companies foresee arrange blockage, distinguish regions with tall request, and optimize arrange assets in like manner. This contributes to more proficient organize arranging, guaranteeing a dependable and responsive broadcast communications foundation.

9. **Community Connectivity:**

Made strides administrations and decreased churn rates contribute to superior community network. Usually especially critical in provincial or underserved regions, where a steady and effective broadcast communications organize can have a positive affect on instruction, healthcare, and financial openings.

10**. Shopper Strengthening:**

As clients gotten to be more educated around the likelihood of churn, they may be enabled to create way better choices approximately their telecom administrations. This straightforwardness can cultivate a more beneficial relationship between clients and benefit suppliers.

In outline, the execution of Client Churn Forecast for the Media transmission Industry utilizing machine learning not as it were benefits telecom companies but too has broader societal impacts by making strides client fulfillment, optimizing assets, making work openings, and contributing to the in general advancement of a more associated and proficient broadcast communications environment.

5.2 Ethical Aspects

Foreseeing client churn within the broadcast communications industry utilizing machine learning raises a few moral contemplations that require cautious consideration. Here are a few moral perspectives to consider:

1. **Security Concerns:**

Guarantee that clients are educated around the information collection and examination for churn expectation. Get unequivocal assent some time recently utilizing their individual data. Clearly communicate how client information will be utilized and guarantee clients that their data will be dealt with mindfully.

2. **Decency and Inclination:**

Be vigilant about potential inclinations within the machine learning calculations. Guarantee that the models are prepared on differing and representative datasets to maintain a strategic distance from propagating existing predispositions within the broadcast communications industry.

3. **Information Security:**

Actualize vigorous security measures to ensure client information from unauthorized get to, breaches, or abuse. Take after industry guidelines and compliance controls for information security.

4. **Client Strengthening:**

Permit clients to have control over their information, counting the capacity to opt-out of prescient churn investigation in case they select.

5. **Straightforwardness:**

Be straightforward around the objectives and results of churn forecast. Keep clients educated approximately the benefits of the examination and how it can lead to moved forward administrations.

6. **Responsibility:**

Clearly characterize obligations for the results of churn expectation. On the off chance that a client is wrongly recognized as high-risk for churn, there ought to be instruments in put to amend the botch and moderate any negative affect.

7. **Information Maintenance Arrangements:**

Build up and follow to clear information maintenance arrangements. Dodge keeping client information longer than fundamental for churn expectation purposes, and guarantee appropriate transfer of information once it is no longer required.

8. **Administrative Compliance:**

Guarantee compliance with information security laws and controls within the locales where the broadcast communications company works. Remain upgraded on changes in enactment and alter hones appropriately.

9. **Predisposition Relief:**

Frequently screen and evaluate the prescient models for predispositions. Execute strategies to relieve inclinations which will rise amid the progressing utilize of the churn forecast framework.

10. **Moral Audit:**
Consider including outside specialists or conducting free moral audits of the churn expectation framework to guarantee that moral measures are kept up.

By tending to these moral viewpoints, a broadcast communications company can create and execute customer churn expectation frameworks in a capable and straightforward way, cultivating believe among clients and partners.

5.3 Sustainability Plan

Making a supportability arrange for "Client Churn Expectation for the Media transmission Industry utilizing Machine Learning" includes sketching out procedures to guarantee the long-term practicality and positive affect of the venture. Underneath could be a comprehensive maintainability arrange:

1. **Venture Diagram and Goals:**
Clearly characterize the project's objectives, counting diminishing client churn, moving forward client fulfillment, and improving income through focused on intercessions.

2. **Information Collection and Protection:**
Guarantee moral and lawful information collection hones. Utilize anonymized information to ensure client security. Routinely audit and upgrade information collection approaches to comply with advancing controls.

3. **Algorithmic Decency:**
Execute measures to address inclination within the machine learning models. Frequently review and move forward the reasonableness of expectations, guaranteeing no segregation against particular statistic bunches.

4. **Nonstop Demonstrate Change:**

Build up a criticism circle for nonstop change based on demonstrate execution. Frequently upgrade the machine learning demonstrate utilizing the most recent information to improve exactness.

5. **Adaptability and Integration:**

Plan the arrangement to be versatile for future development.

Coordinated the churn expectation framework with existing client relationship administration (CRM) and promoting frameworks.

6. **Client Preparing and Selection:**

Give preparing to clients, counting client benefit agents and promoting groups, to decipher and act upon the churn predictions. Screen client selection and address any challenges through continuous preparing and bolster.

7. **Vitality Effectiveness:**

Optimize calculations and foundation for vitality proficiency.

Select cloud suppliers and equipment that prioritize supportability.

8. **Fetched Optimization:**

Routinely audit and optimize the taken a toll of foundation and assets. Investigate cost-effective options for information capacity, handling, and show sending.

9. **Documentation and Information Exchange:**

Make comprehensive documentation for the whole extend lifecycle. Encourage information exchange to guarantee the maintainability of the venture in case of work force changes.

10. **Community and Partner Engagement:**

Lock in with the broadcast communications community to share experiences and collaborate on moving forward industry-wide churn forecast hones. Request criticism from partners and join proposals for upgrade.

11. **Measuring Natural Affect:**
Evaluate and screen the natural affect of the venture, considering variables such as vitality utilization and carbon impression. Actualize activities to play down the natural affect, such as balanced programs or green facilitating arrangements.

12. **Administrative Compliance:**
Remain side by side of changes in regulations related to information security, machine learning, and broadcast communications. Alter the venture in like manner to guarantee compliance with pertinent laws and measures.

13. **Long-Term Commerce Methodology:**
Adjust the churn forecast extend with the long-term trade technique of the broadcast communications company. Persistently assess the project's commitment to generally commerce supportability.

14. **Social Duty:**
Actualize activities that contribute emphatically to society, such as utilizing the churn expectation experiences for focused on client retention programs that benefit both clients and the company.

15. **Intermittent Reviews and Audits:**
Conduct standard reviews to survey the project's execution, adherence to moral guidelines, and arrangement with sustainability goals. Utilize review discoveries to form fundamental alterations for advancement.

By joining these components into the supportability arrange, the "Client Churn Expectation for the Media transmission Industry utilizing Machine Learning" extend can contribute to

long-term victory, moral hones, and positive natural and social impacts. Standard audits and upgrades are fundamental to guarantee the arrange remains effective in the advancing scene of innovation and trade.

CHAPTER 6
CONCLUSION AND FUTURE WORK

6.1 Summary of the Study

This study explores the application of machine learning algorithms in the telecommunications industry, specifically focusing on customer churn prediction. The key algorithms under evaluation include Decision Tree and K-Nearest Neighbors (KNN). Notably, the Decision Tree model exhibited a commendable accuracy of 93.5%, showcasing its effectiveness in handling the intricacies of customer churn prediction. Despite its relative simplicity, the Decision Tree's strong performance underscores its suitability for this specific task. On the other hand, KNN emerged as a top performer with an impressive test accuracy of 95%, demonstrating its adaptability and versatility in addressing the diverse data scenarios inherent in telecommunications datasets.

The findings emphasize the importance of selecting machine learning algorithms tailored to the unique characteristics of the dataset and the predictive task at hand. In the context of customer churn prediction in the telecommunications sector, the results suggest that both Decision Tree and KNN models exhibit strong predictive capabilities. This nuanced understanding of algorithm performance provides valuable insights for future machine learning applications in similar telecommunications scenarios, aiding in the selection of optimal predictive modeling methods for enhancing customer retention strategies.

6.2 Conclusion

In conclusion, the exploration of "Customer Churn Prediction for the Telecommunication Industry Using Machine Learning" illuminates a transformative journey poised to reshape conventional customer retention practices and usher in a new era of proactive, targeted strategies. The integration of machine learning algorithms into churn prediction represents a significant shift with profound implications for telecommunication companies, customers, and the broader industry.

The transition from traditional methods to advanced machine learning approaches underscores a commitment to harnessing technological innovation to address longstanding challenges in predicting customer churn. This study navigates through the historical context, challenges, motivations, and ethical considerations surrounding the adoption of machine learning in the telecommunications sector, providing a comprehensive understanding of the multifaceted landscape.

As telecom companies embrace machine learning, there is an inherent responsibility to address ethical concerns, including algorithmic bias, transparency, and privacy. The potential positive impact is substantial, offering improved accuracy, streamlined processes, and greater customer-centricity that can enhance telecommunication services and customer experiences.

However, this journey is not without its challenges. Ethical considerations, such as the potential reinforcement of biases and responsible handling of sensitive data, necessitate ongoing attention and mitigation efforts. Regulatory compliance and user trust are pivotal in ensuring that the benefits of machine learning are realized responsibly and ethically.

The anticipated outcomes of this transformative shift encompass not only enhanced accuracy in predicting customer churn but also a fundamental change in how telecommunication companies operate. Efficiency, fairness, and adaptability to changing market dynamics are integral components of the envisioned future. As machine learning models evolve, they are expected to become not just predictive tools but enablers of customer satisfaction, retention, and responsible business practices.

In the pursuit of these outcomes, collaboration among data scientists, ethicists, regulators, and industry stakeholders is paramount. The study underscores the importance of ongoing research, continuous monitoring, and a commitment to ethical standards in navigating the complexities associated with machine learning in predicting customer churn.

In essence, "Customer Churn Prediction for the Telecommunication Industry Using Machine Learning" signifies a dynamic and evolving landscape where technological advancements meet ethical imperatives. The potential benefits are vast, but the responsible and ethical implementation of machine learning algorithms is the key to unlocking their

true potential and ensuring a future where telecommunication services are not only efficient and accurate but also customer-centric and sustainable.

6.3 Implication for Further Work

For this research we have recommendations. Machine learning classification is being used in my work for developing accuracy of the model. For a large number of datasets there are many algorithms and methods, datasets. So, that model will predict more exactly for potato disease detection. In future we will -

- Improve accuracy.
- Apply more ML models
- Improve dataset

References:

[1] N. Lu, H. Lin, J. Lu, and G. Zhang, "A Customer Churn Prediction Model in Telecom Industry Using Boosting," *IEEE Transactions on Industrial Informatics*, vol. 10, no. 2, pp. 1659–1665, May 2014, doi: https://doi.org/10.1109/tii.2012.2224355.

[2] P. Ajitha, A. Sivasangari, R. M. Gomathi, and K. Indira, "Prediction of Customer Plan Using Churn Analysis For Telecom Industry," *Recent Patents on Computer Science*, vol. 12, Apr. 2019, doi: https://doi.org/10.2174/2213275912666190410114104.

[3] "Customer Churn Prediction using Logistic Regression with Regularization and Optimization Technique," *Regular*, vol. 9, no. 9, pp. 334–339, Jul. 2020, doi: https://doi.org/10.35940/ijitee.i7219.079920.

[4] S. Janakiraman and K. Umamaheswari, "A Survey on Data Mining Techniques for Customer Relationship Management," Jan. 2014.

[5] "Customer Churn Prediction in Telecommunication Industry Using Deep Learning," *Information Sciences Letters*, vol. 11, no. 1, pp. 185–198, Jan. 2022, doi: https://doi.org/10.18576/isl/110120.

[6] "Prediction of Customer Churn in Telecom Sector using Clustering Technique," *International Journal of Engineering and Advanced Technology*, vol. 8, no. 6S2, pp. 826–832, Oct. 2019, doi: https://doi.org/10.35940/ijeat.f1207.0886s219.

[7] N. I. Mohammad, S. A. Ismail, M. N. Kama, O. M. Yusop, and A. Azmi, "Customer Churn Prediction In Telecommunication Industry Using Machine Learning Classifiers," *Proceedings of the 3rd International Conference on Vision, Image and Signal Processing*, Aug. 2019, doi: https://doi.org/10.1145/3387168.3387219.

[8] A. Amin, F. Al-Obeidat, B. Shah, A. Adnan, J. Loo, and S. Anwar, "Customer churn prediction in telecommunication industry using data certainty," *Journal of Business Research*, vol. 94, pp. 290–301, Jan. 2019, doi: https://doi.org/10.1016/j.jbusres.2018.03.003.

[9] C.-F. Tsai and Y.-H. Lu, "Data Mining Techniques in Customer Churn Prediction," *Recent Patents on Computer Science*, vol. 3, no. 1, pp. 28–32, Feb. 2010, doi: https://doi.org/10.2174/1874479611003010028.

[10] K. Goyal, Kumar Kanishka, Kanisk Vasisth, S. Kansal, and R. C. Srivastavá, "Telecom Customer Churn Prediction: A Survey," *2021 3rd International Conference on Advances in Computing, Communication Control and Networking (ICAC3N)*, Dec. 2021, doi: https://doi.org/10.1109/icac3n53548.2021.9725621.

[11] S. Saraswat and A. Tiwari, “A New Approach for Customer Churn Prediction in Telecom Industry,” *International Journal of Computer Applications*, vol. 181, no. 11, pp. 40–46, Aug. 2018, doi: https://doi.org/10.5120/ijca2018917698.

[12] S. R. Labhsetwar, “PREDICTIVE ANALYSIS OF CUSTOMER CHURN IN TELECOM INDUSTRY USING SUPERVISED LEARNING,” *ICTACT Journal on Soft Computing*, vol. 10, no. 2, pp. 2054–2060, Jan. 2020, doi: https://doi.org/10.21917/ijsc.2020.0291.

[13] “STUDY ON CHURN PREDICTION IN TELECOM INDUSTRY,” *International Journal of Advance Engineering and Research Development*, vol. 3, no. 12, Dec. 2016, doi: https://doi.org/10.21090/ijaerd.031233.

[14] E. P. Priya, J. S S, and S. P, “CUSTOMER CHURN PREDICTION ANALYSIS,” *INTERANTIONAL JOURNAL OF SCIENTIFIC RESEARCH IN ENGINEERING AND MANAGEMENT*, vol. 07, no. 02, Feb. 2023, doi: https://doi.org/10.55041/ijsrem17752.

[15] T. W. Cenggoro, R. A. Wirastari, E. Rudianto, M. I. Mohadi, D. Ratj, and B. Pardamean, “Deep Learning as a Vector Embedding Model for Customer Churn,” *Procedia Computer Science*, vol. 179, pp. 624–631, 2021, doi: https://doi.org/10.1016/j.procs.2021.01.048.

[16] L. Saha, H. K. Tripathy, T. Gaber, H. El-Gohary, and E.-S. M. El-kenawy, “Deep Churn Prediction Method for Telecommunication Industry,” *Sustainability*, vol. 15, no. 5, p. 4543, Mar. 2023, doi: https://doi.org/10.3390/su15054543.

[17] K. S. Rani, S. Thaslima, N. G. L. Prasanna, R. Vindhya, and P. Srilakshmi, “ANALYSIS OF CUSTOMER CHURN PREDICTION IN TELECOM INDUSTRY USING LOGISTIC REGRESSION,” *International Journal of Innovative Research in Computer Science & Technology*, vol. 9, no. 4, Jul. 2021, doi: https://doi.org/10.21276/ijircst.2021.9.4.6.

[18] S. Saraswat and A. Tiwari, “A New Approach for Customer Churn Prediction in Telecom Industry,” *International Journal of Computer Applications*, vol. 181, no. 11, pp. 40–46, Aug. 2018, doi: https://doi.org/10.5120/ijca2018917698.

[19] P. Ajitha, A. Sivasangari, R. M. Gomathi, and K. Indira, “Prediction of Customer Plan Using Churn Analysis For Telecom Industry,” *Recent Patents on Computer Science*, vol. 12, Apr. 2019, doi: https://doi.org/10.2174/2213275912666190410114104.

[20] C.-P. Wei and I-Tang. Chiu, “Turning telecommunications call details to churn prediction: a data mining approach,” *Expert Systems with Applications*, vol. 23, no. 2, pp. 103–112, Aug. 2002, doi: https://doi.org/10.1016/s0957-4174(02)00030-1.

18%	19%	8%	15%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to Daffodil International University Student Paper	6%
2	journalofbigdata.springeropen.com Internet Source	5%
3	dspace.daffodilvarsity.edu.bd:8080 Internet Source	5%
4	www.coursehero.com Internet Source	2%

Exclude quotes	On	Exclude matches	< 1%
Exclude bibliography	Off		