CUSTOMER CHURN PREDICTION USING DEEP LEARNING

PHASE I REPORT

Submitted by

SHRUTHI K 200701237

ILAKIA S R 200701309

in partial fulfillment for the award of the degree

of

BACHELOR OF ENGINEERING

IN

COMPUTER SCIENCE AND ENGINEERING

RAJALAKSHMI ENGINEERING COLLEGE

ANNA UNIVERSITY, CHENNAI – 600 025

DECEMBER 2023

RAJALAKSHMI ENGINEERING COLLEGE

ANNA UNIVERSITY, CHENNAI – 600 025

BONAFIDE CERTIFICATE

Certified that this Report titled “**CUSTOMER CHURN PREDICTION USING DEEP LEARNING**” is the bonafide work of **SHRUTHI K (200701237), ILAKIA S R (200701309)** who carried out the work under my supervision. Certified further that to the best of my knowledge, the work reported here in does not form part of any other thesis or dissertation based on which a degree or award was conferred on an earlier occasion on this or any other candidate.

SIGNATURE

Dr. P. KUMAR, M.E., Ph.D.,
HEAD OF THE DEPARTMENT
Dept. of Computer Science and Engineering,
Rajalakshmi Engineering College,
Rajalakshmi Nagar Thandalam,
Chennai - 602 105.

SIGNATURE

Mrs. S. ANANDHI, M.E., (Ph.D)
SUPERVISOR
Assistant Professor
Dept. of Computer Science and Engineering,
Rajalakshmi Engineering College,
Rajalakshmi Nagar Thandalam,
Chennai - 602 105.

Submitted to Project Viva-Voce Examination held on ________________________

Internal Examiner **External Examiner**

ACKNOWLEDGEMENT

Initially, we thank the Almighty for being with us through every walk of our life and showering his blessings through the endeavor to put forth this report. Our sincere thanks to our Chairman **Mr. S. MEGANATHAN, B.E, F.I.E.**, our respected Chairperson **Dr. (Mrs.) THANGAM MEGANATHAN**, **Ph.D.,** and our Vice Chairman **Mr. ABHAY SHANKAR MEGANATHAN, B.E., M.S.,** for providing us with the requisite infrastructure and sincere endeavoring to educate us in their premier institution. Our sincere thanks to **Dr. S.N. MURUGESAN, M.E., Ph.D.,** our beloved Principal for his kind support and facilities provided to complete our work in time. We express our sincere thanks to **Dr. P. KUMAR, M.E., Ph.D.**, Professor and Head of the Department of Computer Science and Engineering for his guidance and encouragement throughout the project work. We convey our sincere and deepest gratitude to our internal guide, **Mrs. S. ANANDHI, M.E., (Ph.D)** Assistant Professor**,** Department of Computer Science and Engineering, Rajalakshmi Engineering College, for her valuable guidance throughout the project. We are very glad to thank our Project Coordinator, **Dr. M. RAKESH KUMAR**, **M.E., Ph.D.,** Assistant Professor Department of Computer Science and Engineering for his valuable suggestion during our review to build our project.

SHRUTHI K - 200701237

ILAKIA S R - 200701309

ABSTRACT

The telecom industry has rapidly grown and changed. In this dynamic market, telecom service providers are more concerned about customer attrition. In addition to costing money, customer churn, or the losing of subscribers to rival carriers, affects the overall survival and growth of telecom companies. As a remedy to this issue, predictive analytics and machine learning techniques have gained popularity for their ability to predict client attrition. Customer churn, or the loss of subscribers to other providers, has an impact on telecom firm overall viability and expansion in addition to causing revenue loss. Predictive analytics and machine learning approaches have become popular for predicting client attrition as a solution to this problem. Because it's usually less expensive to keep current clients than to get new ones, churn prediction is an essential component of business strategy. This study delves into crafting predictive models for customer attrition by analyzing an array of historical customer data facets such as demographics, transaction records, customer relations, and usage habits. Our methodology encompasses several key stages, beginning with comprehensive data preparation, followed by intricate feature engineering, meticulous model selection, and rigorous evaluation processes to derive meaningful insights. Data preparation, feature engineering, model selection, and evaluation are the steps in our methodology. To create precise churn prediction models, major machine learning methods including Gradient Boosting, Random Forest, and Neural Networks are used. Metrics like F1-score, accuracy, recall, precision, and ROC-AUC are used to evaluate the performance of the model and provide a thorough understanding of its prediction skills.

Keywords – Churn; Customer; Demographics; Data preprocessing; Feature Engineering; Model selection; Evaluation; Feature selection; Input features

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

SVM	-	Support Vector Machine
KNN	-	K-Nearest Neighbors
CNN	-	Convolutional Neural Network
DEEP-BP-ANN	-	Deep Back Propagation Artificial Neural Network
ROS	-	Random Over-Sampling
LFNN	-	Locally Feedforward Neural Network
ANN	-	Artificial Neural Network
S-RNN	-	Stacked Recurrent Neural Network
LGBM	-	Light Gradient Boosting Machine
XGBOOST	-	Extreme Gradient Boosting
CRM	-	Customer Relationship Management
LSTM	-	Long Short-Term Memory
CLARA	-	Clustering Large Applications
BP – BOA	-	Back Propagation - Bayesian Optimization Algorithm
CCP	-	Convolutional Compressive Sensing
PNN1	-	Probabilistic Neural Network 1
PRNN	-	Probabilistic Recurrent Neural Network
PLSTM	-	Probabilistic Long Short-Term Memory

LIST OF ABBREVIATIONS

LSTMNN	-	Long Short-Term Memory Neural Network
AUC	-	Area Under The Curve
SOM	-	Self-Organizing Map
PSO	-	Particle Swarm Optimization
RAM	-	Random Access Memory
GPU	-	Graphics Processing Unit
GB	-	Gradient Boosting
ROC	-	Receiver Operating Characteristic

CHAPTER 1

INTRODUCTION

1.1 OVERVIEW

Customer churn prediction is a critical tool for telecom companies to mitigate revenue loss and enhance customer retention. The telecom sector faces significant challenges related to customer churn, making it a top priority for large companies. By leveraging historical customer data, including consumption patterns, billing information, and consumer behavior, telecom operators can predict potential churners (Bhuse et al., 2020) and take proactive measures to retain at-risk customers. Machine learning algorithms play a pivotal role in identifying key indicators and trends that signal potential churn, enabling companies to offer customized offers and marketing campaigns to mitigate churn. For instance, the Percentage of Transactions to/from other Operator has been identified as a crucial feature, as it tends to be larger for churners, potentially indicating the influence of friends on churn decisions and the strong presence of competing companies in specific regions or communities. Additionally, exploratory data analysis (EDA) has revealed valuable insights, such as the distribution of gender and the ratio of Senior Citizens to Non-Senior Citizens, providing telecom companies with a deeper understanding of their customer base. This approach is crucial for telecom companies as it helps reduce customer acquisition costs (Rahman et al., 2020) and maintain strong client relationships, addressing one of the most significant concerns in the industry. The integration of customer churn prediction into customer retention efforts is essential for telecom companies to adapt to evolving customer needs and preferences, ultimately leading to reduced churn rates and increased customer satisfaction.

1.2 OBJECTIVE

The main objective of a customer churn prediction project is to identify customers who are about to churn (stop using the product or service) in the near future.

1. By leveraging machine learning techniques, we can construct a predictive model

that accurately identifies customers at risk of churn, enabling proactive retention strategies.

2. Through machine learning algorithms, we can pinpoint specific churn risk factors that contribute to customer attrition, allowing businesses to implement targeted countermeasures.

3. By utilizing machine learning, we are able to develop a predictive system that can identify consumers who are likely to leave, allowing for prompt intervention and retention initiatives.

4. By dividing consumers into discrete categories according to their likelihood of churn, we may perform segmentation analysis using machine learning approaches, allowing us to develop customized retention tactics for each group.

1.3 EXISTING SYSTEM

An emphasized analysis of historical customer data (Bhatnagar et al., 2019). The systematic approach involves a meticulous examination of call logs, usage patterns, and billing information. Employing simple statistical models such as logistic regression or decision trees, the algorithm evaluates specific features to estimate the likelihood of churn. However, inherent limitations hinder its effectiveness. Notably, the manual creation of customer categories using predefined criteria might overlook subtle churn indicators that fall outside these established classifications. Also, the system's reliance on static thresholds to identify at-risk users fails to adapt to the dynamic shifts in customer behavior, potentially resulting in missed opportunities to detect evolving churn patterns. The absence of real-time capabilities further exacerbates these shortcomings, limiting the system's agility to swiftly respond to emerging churn signals. Consequently, the system faces challenges in delivering highly tailored offers or recommendations tailored to individual customer preferences (Gaur et al., 2018), impeding its ability to enhance customer retention strategies.

1.3.1 DRAWBACKS IN EXISITING SYSTEM

1. **Reliance on historical data:** To estimate the chance of churn, the current approach mostly uses past customer data. Nuanced churn indications that result from dynamic changes in customer behavior might not always be captured by this method.

2. **Manually defined customer segments:** Based on preset criteria, customer categories are manually defined. This method might not always be able to identify subtle churn indications that result from sudden changes in consumer behavior.

3. **Static thresholds:** To categorize clients who are at danger, the system often uses static thresholds. This method might overlook dynamic changes in behavior that point to possible churn.

4. **Lack of real-time capabilities:** Due to its lack of real-time capabilities, the current system might find it difficult to react quickly to new churn signals.

5. **Challenges in providing personalized recommendations:** It could be difficult for the current system to provide highly customized offers or recommendations based on each customer's unique tastes.

1.4 PROPOSED SYSTEM

Customer churn prediction is a crucial aspect of customer retention for telecom companies. To proactively identify and retain at-risk customers, a sophisticated system that uses state-of-the-art machine learning (Ahmad et al., 2019) and deep learning approaches is suggested. The proposed system will gather and preprocess existing customer data, including demographics, transaction histories, and customer interactions. A comparison of various machine learning algorithms, such as logistic regression and deep learning algorithms, will be used to estimate customer attrition behavior. Feature engineering and selection will be utilized to extract pertinent information. The all-inclusive, data-driven approach to customer churn prediction seeks to assist businesses in lowering attrition, boosting customer loyalty, and eventually increasing profitability. By adjusting various hyperparameters, the models' overall accuracy can be improved. The use of machine learning models can help telecom companies predict customers who are likely to churn and take (Fujo et al., 2022) preemptive measures to retain them. The models can also help identify the factors behind customer churn and propose a churn prediction framework currently lacking in the telecommunications industry. The integration of customer churn prediction into customer retention efforts is essential for telecom companies to adapt to evolving customer needs and preferences, ultimately leading to reduced churn rates and increased customer satisfaction.

1.4.1 ADVANTAGES IN PROPOSED SYSTEM

1. Proactive customer retention: Anticipating churn allows telecom firms to intervene proactively, tailoring retention strategies like targeted offers or enhanced services. This proactive approach prevents issues from escalating, ultimately boosting customer satisfaction and ensuring long-term loyalty within the industry.
2. Data-Driven Insights: Utilizing data empowers informed decision-making, identifying patterns and behaviors contributing to churn for targeted interventions.
3. Sophisticated Algorithm Comparison: By comparing a variety of machine learning algorithms such as logistic regression and deep learning approaches, the system ensures the selection of the most suitable models for predicting customer attrition behavior. This comprehensive approach increases the chances of accurately identifying
   potential churners.
4. Feature Engineering and Selection: Optimized feature selection enhances model accuracy, capturing relevant indicators of potential customer churn.
5. Diverse Modeling Approaches: Employing varied models increases predictive accuracy, improving the ability to forecast customer churn across different scenarios.
6. Hyperparameter Tuning: Fine-tuning model parameters enhances predictive performance, ensuring more accurate churn predictions for proactive action.

CHAPTER 2
LITERATURE SURVEY

(Agarwal et al., 2022) customer churn, or the progressive drop in repeat business, is a major worry for organizations, particularly in the banking industry. Early switcher detection supports proactive retention initiatives. In order to predict the possibility of customer churn, this article uses machine learning, specifically Logistic Regression (LR) and Naive Bayes (NB), along with information about the client's age, geography, credit history, and balance. NB is revealed to be the best model. The paper promotes enhanced churn prediction methodologies, highlighting the combination of methods, such as LGBM-Classifiers and boosting procedures, for improved accuracy and performance. These developments show promise for future churn computation, offering priceless support for customer retention initiatives.

(Agrawal et al., 2018) In Customer churn, or when a group of customers stop using a company's services, has an effect on a company's profitability and stability, hence it is important to foresee it. Deep Learning stands out because it can find important features in complicated, unstructured datasets like those from the telecom industry. Analyzing the major factors significantly determining churn rates is necessary for understanding and predicting churn. The study uses deep learning to predict churn with an accuracy of 80.03%, showing characteristics that are important for client retention efforts. This study enables businesses to focus on important factors for customer retention, reducing losses to rivals.

(Ahmad et al., 2019) this project presents the essential issue of customer turnover is addressed in this study, particularly in the telecom industry where it has a direct impact on revenues. A churn prediction model is created using machine learning techniques on a big data platform along with cutting-edge feature engineering and selection approaches. The area under the curve (AUC) score for the model is an amazing 93.3%. Notably, adding Social Network Analysis (SNA) elements improves performance even more, increasing AUC from 84% to 93.3%. The model surpasses the XGBOOST method for classification when tested on a sizable dataset from the telecom firm SyriaTel. This study is helpful in guiding profitability and reducing customer churn for telecom firms like SyriaTel.

(Ahmed et al., 2017) the Customer turnover, a crucial component of the telecom industry, frequently results from discontent or better deals from rivals. Churn prediction and prevention are crucial, driving businesses to use various data mining and machine learning techniques. Customized products and effective retention techniques are required due to the intense competition. Hybrid methods are the most accurate, however existing techniques such as meta-heuristics and machine learning show effective churn prediction. This study highlights not just precise churn prediction but also examines churn causes and method shortcomings, laying the groundwork for possible hybrid model development in subsequent studies.

(Amol Chole et al., 2023) In Large businesses, especially those in the telecom sector, face a substantial difficulty as a result of customer turnover, necessitating the development of reliable prediction techniques. By creating a churn prediction model using machine learning and deep learning techniques on a sizable dataset derived from GitHub, this study makes a contribution. The model performed better than expected when tested using algorithms including Random Forest, SVM, KNN, and CNN. The Random Forest approach produced an accuracy of 83.11%. In order to improve churn prediction, future research will concentrate on fine-tuning hyperparameters and investigating various machine learning techniques for feature selection and resampling data.

(Anvita Gupta et al., 2022) predicting customer churn is crucial for banks to proactively engage with at-risk customers and prevent attrition. Early intervention alone can reduce churn by 11%. Utilizing past customer data through machine learning and data science techniques offers a solution. This study compares various churn prediction models used by financial organizations, ultimately advocating for a hybrid approach. Results indicate that this hybrid method outperforms existing models and voting classifiers, showcasing its superior accuracy. This underscores the importance of feature impact assessment and dataset clustering for tailored prediction. Future work may involve further sub-clustering and employing additional classification algorithms for enhanced accuracy and outlier mitigation.

(Bhatnagar et al., 2019) customer churn is a challenge for businesses because of fierce competition and a wide range of telecommunication services. Potential churners must be early identified for retention strategies to succeed and be profitable. Churn categorization encompasses both voluntary and involuntary churn, with an emphasis on anticipating purposeful churn. This forecast, a task for supervised classification, aids companies in retaining customers and lowering customer acquisition expenses. Machine learning classifiers like Logistic Regression, Support Vector Machine and Decision Tree are frequently employed in churn prediction models. This paper evaluates the state-of-the-art in churn prediction research, highlighting issues and suggesting future research directions while providing sage advice for young researchers. However, the research is limited in time and focuses on consumer-initiated turnover.

(Bhuse et al., 2020) delivers that the customers have many options in today's competitive market, making client turnover a critical concern for banks. In order to retain engagement, this article uses machine learning approaches to forecast client attrition in the banking industry. The study examines consumer behaviour by classifying data using KNN, SVM, Decision Tree, and Random Forest classifiers, as well as feature selection techniques. Following oversampling, experimental results on a Kaggle churn modelling dataset favoured the Random Forest model, displaying improved accuracy. The study stresses the significance of early-stage churn prediction in the banking industry and offers insights for larger-scale applications while using a very small, unbalanced dataset. The results also show how important oversampling is for resolving data imbalances, especially when applied to SVM classifiers.

(Bin et al., 2007) customer attrition prediction is essential for profitability in the cutthroat Chinese telecom market. Even with insufficient customer data, it's crucial to improve attrition models. The paper recommends decision tree-based experimentation for efficient churn prediction. The churn model's recall rate, precision rate, and F-measure increased as a result of changing the sub-periods of training data, misclassification costs, and sample techniques. With the use of this technique, China Telecom can successfully predict and control customer churn, increasing customer retention in a cutthroat industry. To further improve prediction churn models in related scenarios, future research could investigate alternative data mining techniques.

(Celik et al., 2019) minimizing expenses is essential in today's fiercely competitive environment. Research shows that keeping existing customers costs ten times less than recruiting new ones, highlighting the importance of customer churn monitoring. In the context of customer churn analysis, this paper examines a number of machine learning methods, including ANN, decision tree, SVM, naive bayes, knn, and XG Boost. Machine learning algorithms are considered to be trustworthy substitutes for time-related event estimates, such as customer turnover, even though deep learning approaches exhibit greater performance in complex circumstances. While deep learning approaches excel in complex structures, the Cox regression model efficiently analyses independent variables influencing temporal variables and risk groups. Deep learning techniques are expected to continue to progress and produce even higher success rates over time.

(Fujo et al., 2022) this study addresses the pressing issue of customer churn in the telecom industry by implementing a Deep-BP-ANN model, bolstered by feature selection methods and overfitting prevention techniques. The model outperforms traditional ML techniques like KNN, Logistic Regression, XG Boost, and Naïve Bayes on real datasets (IBM Telco and Cell2cell) with an accuracy exceeding 88%. Lasso regression proves pivotal for feature selection, particularly in datasets with numerous attributes. The ROS technique effectively balances the datasets, and activity regularization aids in mitigating overfitting. Fine-tuning parameters, such as neuron count and epoch number, significantly enhance performance. This model sets a new benchmark, outclassing previous deep learning techniques like CNN, LFNN, and ANN.

(Gaur et al., 2018) says that in Churn research, which makes use of data mining, forecasts client attrition, which is essential in today's cutthroat marketplaces. Predicting customer loss improves marketing, customer loyalty, and communication, which has an effect on profitability. To efficiently retain customers, businesses, particularly telecom providers, concentrate on identifying customer churn factors. Gradient Boosting emerges as the most efficient, followed by Logistic Regression and Random Forest, with SVM performing somewhat less well. These machine learning models include Logistic Regression, SVM, Random Forest, and Gradient Boosted Tree.

(Hu et al., 2018) client churn prediction uses a variety of machine learning classifiers and is essential for client retention and current CRM. Time series customer data analysis is now possible because to recent advancements in data technology, improving accuracy. A pRNN model with LSTM units and product operations has been proposed, and it exhibits great accuracy in predicting churn in the telecom industry. The article covers potential future research topics and emphasizes the importance of recurrent neural networks in processing sequential input. To validate findings and investigate long-term prediction views, additional diversified real-world datasets are required.

(Ismail et al., 2015) the customer management is essential in the telecommunications sector to prevent churn. The large expenses involved with adopting it across the whole customer base can be avoided with targeted retention initiatives for likely churn clients. Utilizing historical churn data and predictive factors, churn management focuses on prediction. While long-term success is assured by keeping existing clients, traditional marketing places greater emphasis on obtaining new ones. With a prediction accuracy of 91.28%, Multilayer Perceptron Artificial Neural Network outperforms conventional statistical models in predicting customer attrition. These information should be used in customer retention initiatives to effectively lower churn rates.

(Karvana et al., 2019) the customer attrition in banking may be accurately predicted by data mining. Recall rates are highly influenced by sample size and inter-class comparisons, favoring a 50:50 data ratio with a 70% recall. Each class has roughly 7,975 samples out of approximately 15,949 data samples. The 50:50 SVM sampling model is the most effective one, which identifies important characteristics like vintage, EDC transaction volume and amount, average balance, and age and generates a large profit of 456 billion. This is consistent with the research, which highlights SVM's accuracy while highlighting Logistic Regression's ability to reduce losses.

(Kumar, P et al., 2023) distributed denial-of-service (DDoS) attacks pose a significant threat to the confidentiality and integrity of computer networks, disrupting web traffic to target servers and impeding authorized user access to services. Detection of DDoS attacks can be challenging, requiring robust mitigation strategies due to the diverse methods used to flood networks or servers. These assaults leverage resource limitations, impacting the functionality of the targeted organization's website infrastructure. Analyzing the most recent datasets is crucial for identifying and understanding the evolving landscape of DDoS attacks, assessing their varied techniques, and evaluating their efficacy. Clients accessing network services are consistently exposed to this pervasive and severe threat, necessitating ongoing vigilance and proactive security measures.

(Kumar, P. et al., 2023) delves into the pressing issue of increasing energy consumption within cloud server farms, highlighting their substantial contribution to environmental pollution resulting from heightened power usage. This study accentuates the complexities associated with mitigating power consumption while upholding agreements concerning service quality. To address this challenge, the paper proposes a solution centered on optimizing resource allocation. This involves a strategic approach that limits the operation of dynamic servers, thereby aiming to curtail energy usage while simultaneously meeting the demands of clients and ensuring efficient task performance. To validate the efficacy of their proposed algorithms, the researchers leverage Cloud Sim, a simulation tool, utilizing real-world data obtained from a significant pool of over 1000 Planet Lab virtual machines. The study underscores the pivotal role played by server farms in this evolving technological landscape, emphasizing the critical need to strike a balance between energy conservation and maintaining high-quality service provision.

(Maw et al., 2019) the companies have a problem from customer churn, which is a result of severe competition and a variety of telecommunication services. For retention initiatives to be successful and profitable, potential churners must be identified quickly. With an emphasis on foreseeing intentional churn, churn categorization includes both voluntary and involuntary churn. This forecast, a supervised categorization task, helps businesses keep clients and cut acquisition costs. In churn prediction models, machine learning classifiers like Support Vector Machine, Logistic Regression, and Decision Trees are widely used. This paper examines current churn prediction research, noting problems and potential areas for future research while offering insightful advice for up-and-coming scientists. The research, though, is time-bound and concentrates on churn that is instigated by the consumer.

(Rahman et al., 2020) the customers have many options in today's competitive market, making client turnover a critical concern for banks. In order to retain engagement, this article uses machine learning approaches to forecast client attrition in the banking industry. The study examines consumer behaviour by classifying data using KNN, SVM, Decision Tree, and Random Forest classifiers, as well as feature selection techniques. Following oversampling, experimental results on a Kaggle churn modelling dataset favoured the Random Forest model, displaying improved accuracy. The study stresses the significance of early-stage churn prediction in the banking industry and offers insights for larger-scale applications while using a very small, unbalanced dataset. The results also show how important oversampling is for resolving data imbalances, especially when applied to SVM classifiers.

(Sudharsan et al., 2022) in the cutthroat and fast-paced telecom sector, client turnover is a significant problem that needs to be addressed. S-RN0N is a unique framework for precise churn prediction. The model divides clients into churners and non-churners, and if churn is anticipated, it prompts further study for retention tactics. Data collection, preprocessing, filtering, grouping, feature engineering, and classification are all included in the suggested methodology. The S-RNN model performs admirably in experimental analysis, attaining outstanding metrics like 98.27% sensitivity, 92.31% specificity, and 95.99% accuracy. The suggested method also performs better in terms of resilience and reliability than current methods. Future studies might examine changing consumer behaviour patterns utilizing cutting-edge forecasting techniques and trend analysis.

(Zhang et al., 2022) the telecom companies face a pressing challenge with client churn, impacting profits in a saturated global market. Although attracting new clients is expensive, keeping the ones you already have is more cost-effective. Predicting and preventing customer churn has become a top priority for telecom companies. This study introduces discriminant and logistic regression models using customer segmentation data from major Chinese telecom firms. The findings empower managers to accurately predict customer behaviour, enhance retention strategies, and optimize budgets. Notably, this research fills a gap in telecom customer churn studies by employing Fisher discriminant and logistic regression analyses, offering valuable insights for industry improvement.

(Zhao et al., 2008) the customer loyalty has a bigger impact on bank profits than things like growth and market share. client churn lowers sales and new client acquisition. Data mining provides for accurate churn prediction and customized marketing tactics. With excellent accuracy and practical considerations, a support vector machine (SVM) model beat other classifiers in predicting bank customer attrition. SVM is a reliable method for churn prediction because of its straightforward classification surface, good generalization, and fitting accuracy.

Table 2.1 Synoptic Overview

Author / Year	**Approach**	**Merits**	**Demerits**
(Agarwal et al., 2022)	It primarily focuses on SVM, Naive Bayes	The study demonstrates the successful implementation of machine learning algorithms in predicting customer churn, with Naive Bayes showing notably high accuracy.	However, it could benefit from discussing challenges, limitations, or areas for future improvement in the methodologies applied.
(Agrawal et al., 2018)	It utilizes Artificial Neural Network	The paper demonstrates the effectiveness of Deep Learning techniques, particularly ANNs, in	The proposed model may not scale effectively for long-term user interactions and faces challenges in

		mobile network churn prediction. It highlights the stability of the model across different months and emphasizes the improvement achieved by incorporating location data.	handling textual parameters during data preprocessing, potentially limiting its applicability over extended periods.
(Ahmad et al., 2019)	XGBOOST, Random Forest, GBM, Decision Trees are utilised	The utilization of XGBOOST led to an impressive accomplishment, achieving a notably high AUC value of 93.301%. This outcome notably surpassed the performance of various other algorithms that were put to the test. The integration of Social Network Analysis features played a pivotal role in elevating the accuracy.	Data unavailability to the public due to restrictions by the telecom company, limiting broader research access. The specific journal of publication is not mentioned.
(Ahmed et al., 2017)	It goes through techniques such as SVM, Neural Networks, PSO, Anti Miner+	Hybrid models combining SVM, ANN, and SOM exhibit notably high accuracy levels while maintaining lower complexity, showcasing their efficiency. Rule induction techniques and PSO (Particle	Models have limitations like false positives, time complexity for rule-set generation, and issues with imbalanced data; certain approaches overlook demographic factors and misclassification costs

		Swarm Optimization) prove to be effective methodologies for predicting churn with promising outcomes.	
(Amol Chole et al., 2023)	It involves Random Forest, K - Nearest Neighbours and Support Vector Machine.	The model achieves a notable accuracy in predicting customer churn, which is crucial for revenue retention in the highly competitive telecom industry.	It's essential to consider factors like scalability and real-world implementation challenges in future applications of the model.
(Anvita Gupta et al., 2022)	It utilizes Decision Tree Classifier, Logistic Regression, Stochastic Gradient Descent, Support Vector Machine, K-Nearest Neighbours, Voting Classifier, Random Forest, Naive Bayes.	It emphasizes identifying impactful features and clustering data to apply diverse models for precise predictions. The method showcases potential to extend to multiple sub-clusters for enhanced accuracy.	Limitations could include complexity in determining optimal clusters and potential challenges in scaling the approach for very large datasets without clearly defined clusters.
(Bhatnagar et al., 2019)	It involves KNN, Logistic Regression	The study compares KNN and Logistic Regression for customer churn prediction, favoring KNN with 2.0% higher accuracy.	The study lacks exploration of feature importance and actionable strategies to mitigate customer churn
(Bhuse et al., 2020)	Random Forest, SVM, XG Boost, Ridge classifier, KNN, Deep Neural Network.	The paper provides a comprehensive analysis of different techniques for customer churn prediction in the telecom sector, offering insights into	The study does not delve into specific challenges or limitations faced during the implementation of the models. Additionally, the paper does not address potential

		the effectiveness of various algorithms. The Random Forest model demonstrated the highest accuracy, indicating its potential for real-world applications.	scalability issues that may arise in large-scale telecom operations.
(Bin et al., 2007)	It employs Decision tree	Effective utilization of decision tree for diverse applications in different domains, showcasing its versatility and usefulness in decision-making processes.	Specific limitations or drawbacks of the decision tree method are not detailed in the provided document.
(Celik et al., 2019)	It includes Cox Regression, ANN, KNN, Decision Tree, SVM, Logistic Regression, Naïve Bayes, XG Boost	Accurate prediction of event times for customer churn (Martinsson, 2016) Improved accuracy in predicting customer churn compared to traditional methods	Limitations could include complexity in determining optimal clusters and potential challenges in scaling the approach for very large datasets without clearly defined clusters.
(Fujo et al., 2022)	It is Deep-BP-ANN, CNN, LFNN, Transfer learning, ensemble – classifiers.	Implemented Deep-BP-ANN model outperformed other machine learning techniques, achieving higher accuracy and efficiency.	Improving accuracy in existing models that use deep learning techniquesfor customer churn prediction.
(Gaur et al., 2018)	Logistic Regression, SVM, Gradient Boosting, Random Forest	The study highlights gradient boosting's effectiveness in predicting customer churn for telecom, emphasizing its role in	The document lacks author names, journal details, and dataset preprocessing.

		customer retention strategies.	
(Hu et al., 2018)	Logistic Regression, Random Forest, PNN1, LSTM, pRNN, pLSTM, LSTMNN	pRNN demonstrates superior performance in customer churn prediction, offering valuable insights for customer relationship management and marketing strategies	The study suggests validating findings with diverse datasets and highlights future research directions such as long-term churn prediction and addressing accuracy-scalability trade-offs.
(Ismail et al., 2015)	Neural Networks, Logistic Regression, SVM, Bayes Network, Rough Set Theory, K-Means, Time Series, Regression Forests, Association Rules	Neural network, specifically the MLP model, demonstrated superior prediction accuracy (91.28%) compared to traditional statistical models like regression analysis.	The high computational complexity and potential overfitting are challenges associated with using neural network models for churn prediction.
(Karvana et al., 2019)	Neural Network, Decision Tree, SVM, Logistic Regression, Naïve Bayes	This project serves as a demonstration of data mining techniques applied specifically to churn prediction, underscoring the crucial significance of sample size in this predictive analysis.	limited data source detail, lack of thorough comparative analysis, insufficient exploration of alternate models, inadequate future research recommendations
(Kumar, P et al., 2023)	The approach analyzes real-time datasets to detect and mitigate DDoS attacks, evaluating various attack vectors for effective defense strategies.	Enhances network resilience by utilizing real-time datasets for accurate DDoS attack analysis. Enables the development of effective defense strategies by evaluating various	DDoS attack early detection difficulties could make it more difficult to put mitigation measures in place in a timely manner. Updating defense strategies requires

		attack vectors, ensuring uninterrupted service for authorized users.	constant adaptation, which makes it difficult to keep defenses in place.
(Kumar, P et al., 2023)	The approach optimizes cloud server farm resource allocation to cut energy consumption, validated using CloudSim	It addresses escalating carbon emissions in cloud computing, striving for efficiency by reducing environmental impact while maintaining service quality.	Challenges may arise in implementing limitations on dynamic servers without compromising performance, and the methodology's effectiveness may vary based on the specific characteristics of the cloud environment.
(Maw et al., 2019)	CCP, SVM, Random Forest, Neural Network, Decision Tree	Comprehensive analysis of recent literature in churn prediction, revealing emerging research opportunities, and providing insights into the challenges faced in the telecom industry.	Limited to a specific time frame, potential exclusion of important studies, potentially missing broader perspectives.
(Rahman et al., 2020)	It involves SVM, Random Forest, Decision Tree, K - Nearest Neighbours	The SVM model, especially with integrated random sampling, improved predictive accuracy and precision in churn prediction compared to other models. It addressed data imbalance issues	The SVM model struggled with highly imbalanced data, affecting its predictive capability. The study highlighted challenges in achieving high accuracy with imbalanced datasets using SVM.

		through resampling techniques.	
(Sudharsan et al., 2022)	CLARA, BM-BOA, Feature selection, Classification, retention process, Swish RNN.	Achieves high accuracy and sensitivity .Efficient clustering and feature selection methodologies improve performance. Utilizes Swish RNN, a novel approach, for churn prediction.	limited discussion on recent data due to the document's age.
(Zhang et al., 2022)	This employs Decision tree, logistic regression, Cluster analysis.	The attainment of a substantial accuracy rate of 93.94% in forecasting customer churn holds significant implications for telecom companies, offering invaluable insights to fine-tune operational costs..	Lack of information about the specific journal where the paper was published, and limited discussion on recent data due to the document's age.
(Zhao et al., 2008)	SVM Logistic, Regression Naïve, Bayes ANN	Support Vector Machines (SVM) have demonstrated superior performance over an array of classifiers, showcasing remarkable effectiveness specific in customer churn prediction.	SVMs can be sensitive to parameter choice, computationally intensive for large datasets, and less interpretable compared to some other models

From the above Table 2.1 provides a comprehensive overview of various image cartoon methods proposed by different authors across multiple years. Each approach is described in terms of its methodology, merits, and demerits, covering aspects such as image quality improvement, computational complexity, and applicability to diverse domains.

CHAPTER 3

SYSTEM DESIGN

3.1. GENERAL

A project's design is an important aspect that displays the objective of the model that will be constructed. The process of translating the requirements into a representation of the software is known as software design. The field of design is where excellence is produced. The process of specifically translating customer needs into final products is called design.

3.2. DEVELOPMENT ENVIRONMENT

3.2.1 HARDWARE REQUIREMENTS

The hardware specifications outlined in Table 3.1 serve as the comprehensive system implementation prerequisites, forming the fundamental basis upon which software engineers build their system designs. Rather than delineating the operational procedures, these specifications provide a detailed overview of the system's functionalities and the essential hardware components necessary for its successful deployment. They offer a blueprint guiding engineers in configuring and structuring the system architecture, ensuring compatibility and optimal performance based on the outlined hardware prerequisites.

Table 3.1 Hardware Requirements

COMPONENTS	SPECIFICATION REQUIREMENTS
CPU(Processor)	Inter i5/Ryzen 2000 Series
Clock Speed	525MHz
GPU	Integrated Graphics
STORAGE(HDD/SSD)	2GB
RAM	8GB RAM

3.2.2 SOFTWARE REQUIREMENTS

The system specification is found in the software requirements paper. It ought to have a description and a list of prerequisites. Rather than focusing on how the system should operate, it is a list of what it should perform. The foundation for developing the software requirements specification is provided by the software requirements. It is helpful for tracking the team's progress during the development activity, organizing team activities, carrying out tasks, and predicting costs.

- Python 3.10
- Numpy

3.3 ARCHITECTURE DIAGRAM

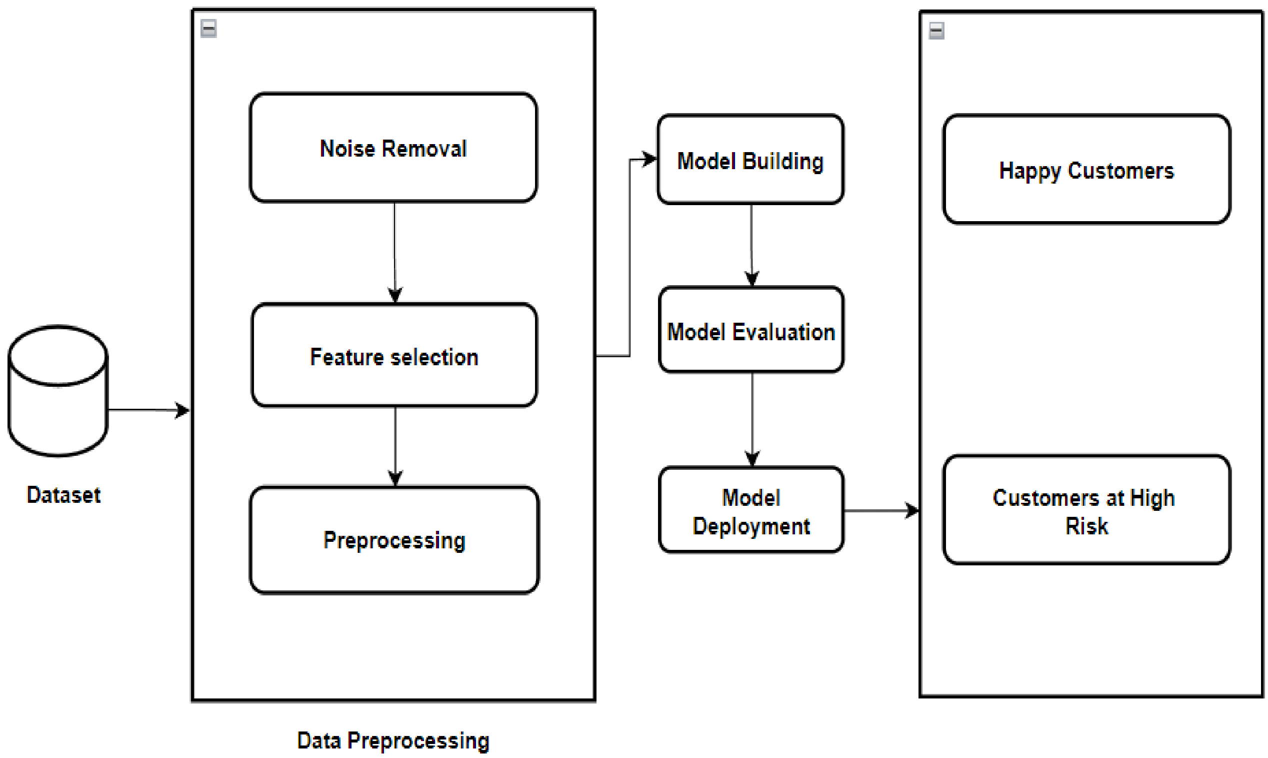

Figure 3.1 Architecture Diagram

In the Fig 3.1, An architecture diagram is a visual representation illustrating the structure, components, interactions, and relationships within a system or application. It serves as a blueprint, providing a high-level overview of the system's design, including its various elements, their functionalities, and how they interact with one another. Typically, these diagrams employ symbols, shapes, and annotations to depict different components, such as servers, databases, modules, interfaces, and their connections or communication pathways. In the The first step involved in Customer Churn Prediction is to import the dataset. After importing the dataset feature processing is done. The next stage is where Model Building and Model Evaluation is done. Finally the output is predicted whether the customers are happy customers or the customers who are at high risk.

Customer churn prediction begins with the acquisition and loading of a comprehensive dataset that encapsulates historical customer interactions, behavior, and other relevant attributes. This dataset typically comprises information like customer demographics, purchase history, service usage patterns, feedback, and any other data points deemed important for churn analysis. Using Python libraries such as pandas, the dataset is imported into a structured format for further analysis and processing.

The loaded dataset might contain noise in the form of missing values, outliers, or inconsistent data, which can adversely impact the predictive model's accuracy. Addressing these issues is crucial in ensuring the quality and reliability of the dataset. Techniques like imputation for missing values, outlier detection and treatment (e.g., removal or transformation), and handling inconsistencies play a pivotal role in preparing a clean dataset for subsequent analysis.

Preprocessing involves a series of steps to transform and prepare the dataset for effective modelling. Categorical variables are encoded into numerical representations using methods like one-hot encoding or label encoding. Feature engineering techniques might be applied to extract more informative features or derive insights from existing ones. Additionally, scaling, normalization, or standardization of numerical features ensures uniformity in the data and prevents any bias due to varying scales. The preprocessed dataset is then used to build predictive models aimed at forecasting customer churn. Several machine learning algorithms, including but not limited to logistic regression, decision trees, random forests, support vector machines, gradient boosting, and neural networks, are employed for this purpose. The dataset is split into training and testing sets to train the model on historical data and evaluate its performance on unseen data. Hyperparameter tuning and cross-validation techniques are used to optimize model parameters and improve predictive accuracy.

The performance of the trained models is rigorously evaluated using various evaluation metrics tailored to churn prediction tasks. Metrics such as accuracy, precision, recall, F1-score, area under the ROC curve (AUC-ROC), or lift curves are calculated using the test dataset. These metrics provide insights into how well the model identifies customers at risk of churn while minimizing false predictions. Model evaluation aids in selecting the most effective model for deployment based on its predictive capabilities.

The final step involves interpreting the model's predictions and presenting actionable insights. Visualizations like confusion matrices, ROC curves, or precision-recall curves illustrate the model's performance in a comprehensible manner. Insights derived from the model, such as significant features influencing churn prediction, assist decision-makers in devising targeted strategies for customer retention. These results and actionable insights are communicated to stakeholders or business decision-makers, enabling them to implement proactive measures to mitigate churn and improve customer retention rates.

Each customer is assigned a probability score, delineating the model's confidence in its prediction. This output facilitates the identification of high-risk customers, enabling proactive retention strategies.

In a telecom customer churn prediction, an architecture diagram plays a pivotal role by visually presenting the system's design and technical components. It provides a concise and clear representation of the data sources utilized (such as call logs, usage trends), the analytics methods employed (like statistical models), and the overall predictive model's framework. This diagram serves as a crucial reference point, aiding in understanding the system's architecture for all stakeholders involved, including developers, project managers, and decision-makers. Visual representations of data flows, processing stages, and algorithm integration offer a streamlined way to communicate the intricate technicalities of a project. This approach aids in simplifying complex concepts, allowing for clearer comprehension among stakeholders.

3.4 DESIGN OF THE SYSTEM

3.4.1 USE CASE DIAGRAM

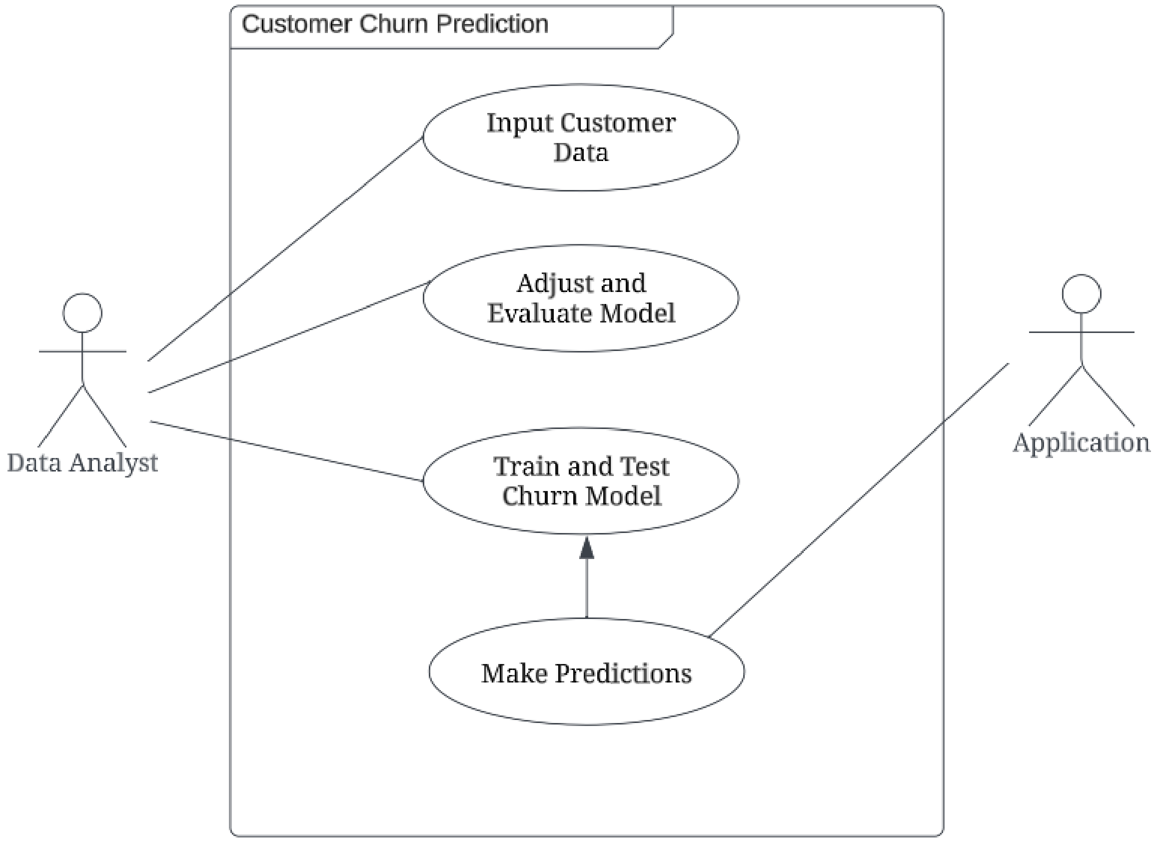

Figure 3.2 Use Case Diagram

In the Fig 3.2, A use case diagram is a visual representation that illustrates the various interactions and relationships between actors and the functionalities or services provided by a system. It provides a high-level view of how users interact with a system and helps in understanding the system's behavior from an external perspective. The outlined use case diagram depicts the fundamental interactions and relationships between various actors and the system components involved in a customer churn prediction system. The "Data Analyst" actor engages with the system through pivotal functionalities: "Input Customer Data," where the analyst provides historical customer data and churn-related variables, "Adjust and Evaluate Model," initiating the process to develop a churn prediction model and enabling the assessment of model performance, and "Train and Test Model," allowing the analyst to improve the model to get trained. These functionalities represent the critical phases in model development, validation, and refinement. The "Application" actor operates independently in predicting customer churn based on the provided input data through the "Make Predictions" use case. This feature underscores the system's core functionality in autonomously processing data to generate churn predictions. This use case diagram provides a structured overview of the system's functionalities and interactions, delineating the roles of each actor in the churn prediction process. It showcases a cyclical flow where data ingestion, model development, prediction, customer engagement, and model refinement converge to address the dynamic nature of customer churn management.

3.4.2 SEQUENCE DIAGRAM

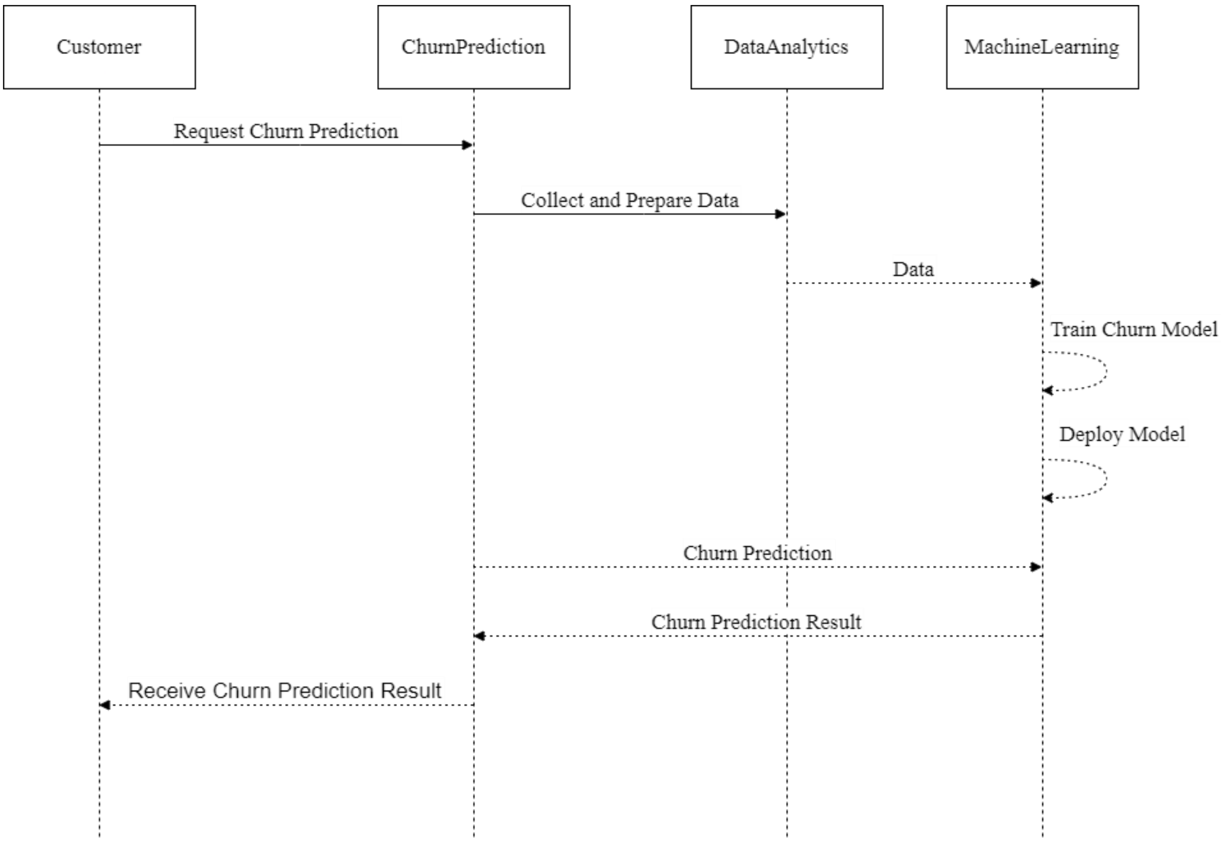

Figure 3.3 Sequence Diagram

In the Fig 3.3, A sequence diagram is a visual representation illustrating the interactions and flow of messages or actions between different components, objects, or entities within a system. In the context of customer churn prediction in the telecom industry, a sequence diagram can depict the step-by-step processes and communication pathways between various elements involved in the churn prediction system. It showcases how data is gathered from diverse sources like call logs, usage patterns, or customer profiles, how it is processed through analytical models or algorithms, and how the predictive outcomes are derived and utilized. The sequence diagram visually encapsulates the intricate flow of interactions and data exchange among the diverse components involved in the comprehensive process of customer churn prediction. It commences with the pivotal role of the "Customer," who initiates the churn prediction process by submitting a request for information. This action triggers the subsequent steps orchestrated by the "Churn Prediction" system, which functions as the central coordinator for the entire predictive pipeline. Upon receipt of the customer's request, the "Churn Prediction" component seamlessly interfaces with the "Data Analytics" subsystem, prompting it to commence the collection and meticulous preparation of the necessary data. This data aggregation phase is fundamental, providing the essential groundwork for predicting customer churn accurately.

Once prepared, the data is relayed from the "Data Analytics" module to the specialized "Machine Learning" component. Within the confines of the "Machine Learning" module, the received data serves as the core material for training a sophisticated churn prediction model, representing a pivotal step in forecasting potential customer churn. The model undergoes an intensive training process, leveraging the dataset to comprehend patterns and behaviors critical for accurate churn prediction. Upon successful training, the churn prediction model undergoes deployment within the "Machine Learning" component, becoming operational for churn prediction purposes. The "Churn Prediction" system then interfaces with this deployed model, engaging it to generate churn predictions based on the incoming data. The model utilizes its acquired insights and learned patterns to forecast potential churn instances. The churn predictions generated by the model are then communicated back to the "Churn Prediction" component, marking a critical phase in the predictive loop. Finally, the "Churn Prediction" component consolidates the churn prediction results and delivers them to the original requester, the "Customer," fulfilling the initial information request for churn predictions.

CHAPTER 4

PROJECT DESCRIPTION

4.1 MODULE DESCRIPTION

Customer Churn Prediction is divided into five modules:

1. Data Collection and Preparation
2. Feature Engineering
3. Model Selection and Training
4. Model Evaluation and Deployment
5. Churn propensity scoring

4.1.1 DATA COLLECTION AND PREPARATION

In the Fig 4.1, Customer churn prediction initiates with a fundamental phase centered on data collection. During this stage, businesses actively source pertinent information from a diverse array of channels such as customer databases, transaction records, user engagements, and feedback surveys. This comprehensive data compilation comprises a spectrum of valuable insights, encompassing demographic particulars, usage patterns, purchasing behaviors, and interactions with customer support services. Subsequently, the amassed data undergoes a critical phase termed data cleaning. This pivotal step involves meticulous handling of various aspects including addressing missing values, identifying and eliminating outliers, and ensuring the overall consistency and accuracy of the data. Through this thorough refinement process, businesses establish a robust and reliable foundation upon which to base more precise churn predictions. This refined dataset serves as a cornerstone for the development of sophisticated models and algorithms, significantly enhancing the accuracy of forecasts regarding customer attrition. Ultimately, the refined data empowers businesses to formulate proactive retention strategies, thereby bolstering customer retention efforts and overall business success.

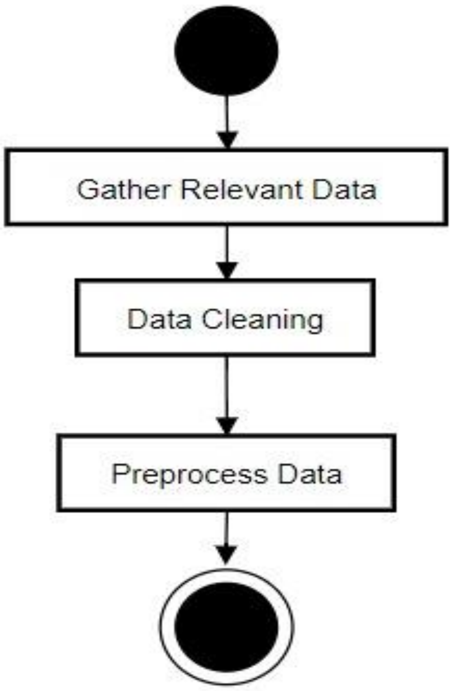

Figure 4.1 Data Collection and Preparation

4.1.2 FEATURE ENGINEERING

In the Fig 4.2, In the process of customer churn prediction, feature engineering plays a pivotal role in refining the predictive model. The initial step involves feature selection, where the most pertinent variables for predicting churn are identified. Following this, feature transformation becomes imperative, necessitating the transformation and preprocessing of selected features. This includes actions such as one-hot encoding categorical variables, scaling numerical features for uniformity, and crafting new features capable of encapsulating customer behaviors or evolving trends. Additionally, for businesses dealing with time-dependent data, a specialized approach is required, often involving time-series analysis. This analysis focuses on understanding patterns within time-related customer activities, enabling the model to account for temporal trends and dependencies. Together, these steps in feature selection, transformation, and time-series analysis contribute significantly to the construction of a robust churn prediction model, ensuring a more comprehensive and nuanced understanding of customer behavior to facilitate informed decision-making for retention strategies.

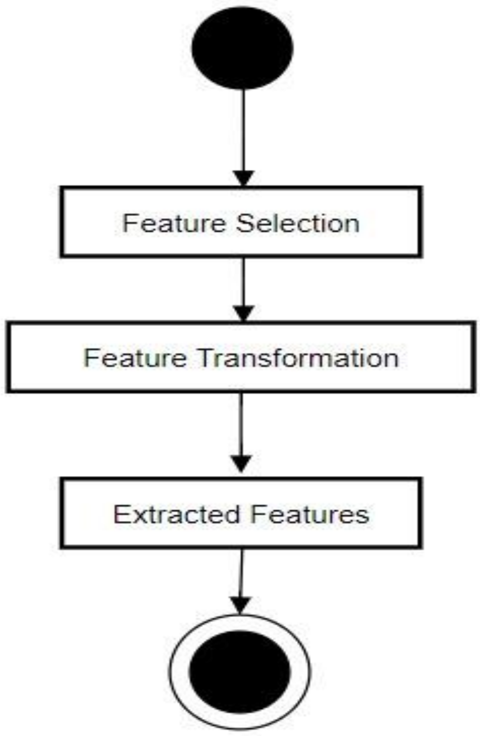

Figure 4.2 Feature Engineering

4.1.3 MODEL SELECTION AND TRAINING

In the Fig 4.3, The selection of appropriate machine learning models serves as a critical step. Various models, such as random forests, logistic regression, decision trees, gradient boosting, and neural networks, stand as common choices owing to their efficacy in predictive analytics. Once chosen, evaluating the model's performance becomes pivotal, necessitating the division of data into distinct training and test sets. To ensure the model's robustness and reliability, cross-validation techniques can be employed, allowing for comprehensive assessment across different subsets of the data. With the data properly partitioned, the selected models undergo training using the training dataset. During this phase, fine-tuning of hyperparameters takes place, aiming to optimize the model's performance. This meticulous process of model selection, evaluation, and training stands crucial in developing a churn prediction framework that not only accurately identifies potential churn but also aids in devising proactive retention strategies, thereby empowering businesses to mitigate customer attrition effectively.

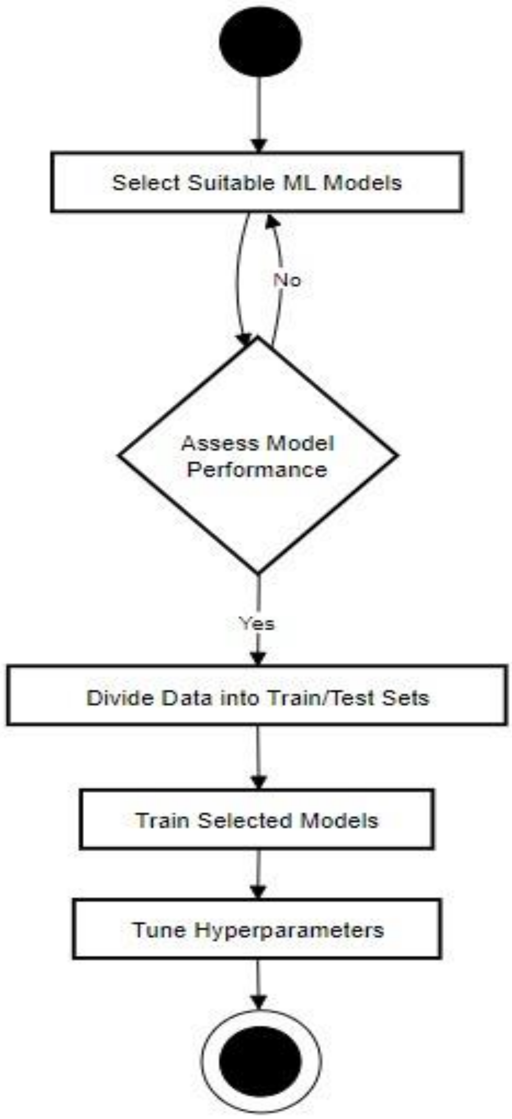

Figure 4.3 Model Selection and Training

4.1.4 MODEL EVALUATION AND DEPLOYMENT

In the Fig 4.4, The culmination of the churn prediction process involves evaluating the model's performance using a spectrum of metrics tailored to the business's objectives, such as recall, accuracy, precision, F1 score, and ROC AUC. The selection of these metrics is crucial as they provide insights into the model's effectiveness in predicting churn. It's imperative to align these metrics with the overarching business goals to ascertain their relevance and significance. Once the best-performing model is identified, it's deployed into a production environment, enabling real-time predictions for new customers. Continuous monitoring of the deployed model becomes paramount to ensure its sustained performance over time. Periodic retraining of the model with new data allows for adaptation to evolving trends and patterns, ensuring its efficacy in making accurate churn predictions. This ongoing cycle of evaluation, deployment, monitoring, and retraining forms a crucial iterative process, facilitating the creation of a dynamic and resilient churn prediction system that aids businesses in proactively addressing customer attrition and enhancing customer retention strategies.

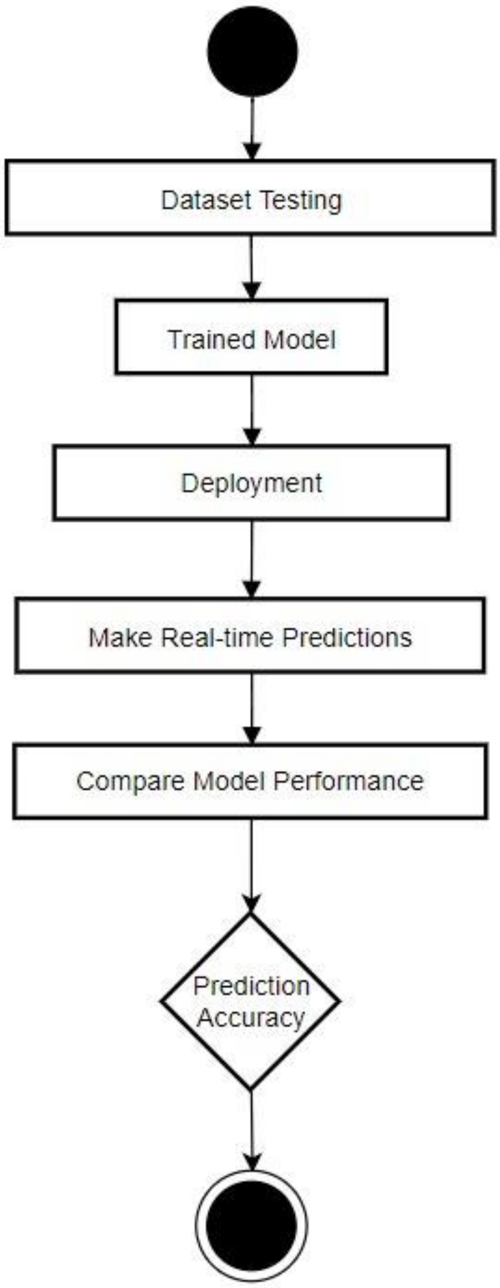

Figure 4.4 Model Evaluation and Deployment

4.1.5 CHURN PROPENSITY SCORING

In the Fig 4.5, The essence of churn prediction lies in computing a churn propensity score tailored to individual customers. This score acts as a predictive indicator, reflecting the likelihood of a customer churning in the foreseeable future based on insights gleaned from the model's predictions. By assigning a churn propensity score to each customer, businesses gain a nuanced understanding of the potential attrition risks associated with their clientele. Customers exhibiting high propensity scores are flagged as being at elevated risk of churning. Identifying these high-risk customers becomes pivotal as it prompts the implementation of targeted retention strategies. Through focused and tailored efforts aimed at high propensity score customers, businesses can proactively engage and address the underlying factors contributing to potential churn, thereby bolstering their retention initiatives and fostering stronger customer relationships.

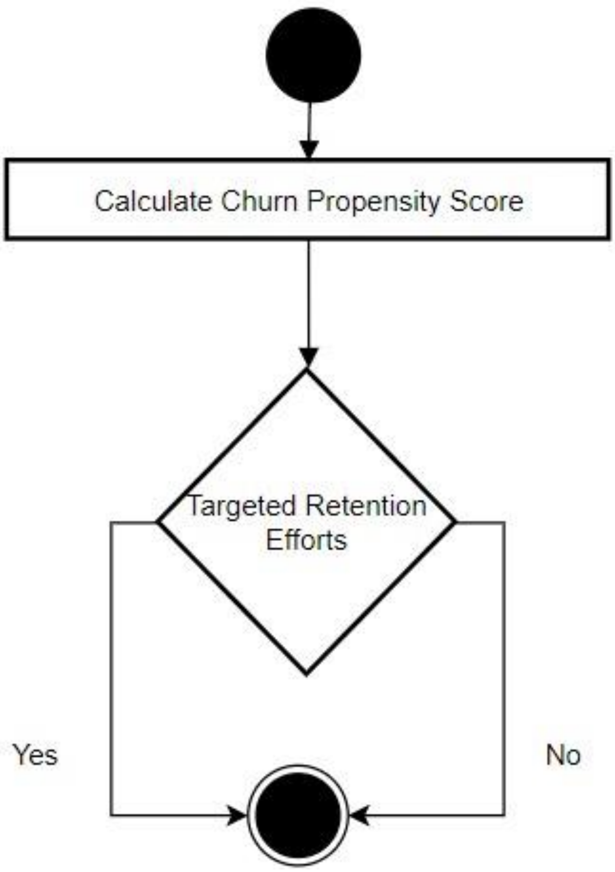

Figure 4.5 Churn Propensity Scoring

CHAPTER 5

IMPLEMENTATION AND RESULT

5.1 DATASET AND INPUT

In the Fig 5.1, The telecom industry, inputs for customer churn prediction systems typically include a combination of demographic details such as age, location, and income, alongside usage patterns like call frequency, data consumption, and service plan features. Billing information, payment history, and customer service interactions, including the number of calls and reasons for contact, are also crucial. These features collectively enable machine learning models to discern patterns and forecast the likelihood of a customer churning, aiding telecom companies in proactively managing customer retention strategies.

	customerID	gender	SeniorCitizen	Partner	Dependents	tenure	PhoneService	MultipleLines	InternetService	OnlineSecurity	...
0	7590-VHVEG	Female	0	Yes	No	1	No	No phone service	DSL	No	...
1	5575-GNVDE	Male	0	No	No	34	Yes	No	DSL	Yes	...
2	3668-QPYBK	Male	0	No	No	2	Yes	No	DSL	Yes	...
3	7795-CFOCW	Male	0	No	No	45	No	No phone service	DSL	Yes	...
4	9237-HQITU	Female	0	No	No	2	Yes	No	Fiber optic	No	...
...	...	...	...	...	...	...	...	...	...	...	...
7038	6840-RESVB	Male	0	Yes	Yes	24	Yes	Yes	DSL	Yes	...
7039	2234-XADUH	Female	0	Yes	Yes	72	Yes	Yes	Fiber optic	No	...
7040	4801-JZAZL	Female	0	Yes	Yes	11	No	No phone service	DSL	Yes	...
7041	8361-LTMKD	Male	1	Yes	No	4	Yes	Yes	Fiber optic	No	...
7042	3186-AJIEK	Male	0	No	No	66	Yes	No	Fiber optic	Yes	...

7043 rows × 21 columns

Figure 5.1 Dataset

5.2 CHURNED CUSTOMERS

Within Figure 5.2, the output of a customer churn prediction system offers a crucial overview of anticipated customer behavior in the telecom industry. This output, often in binary classification, segments customers into those likely to continue their services and those at risk of churning. This division serves as a valuable tool for businesses to identify and prioritize customer retention efforts. By focusing on those predicted to churn, companies can proactively intervene with tailored strategies, such as personalized offers or targeted communication, to prevent customer loss. Simultaneously, it allows the allocation of resources towards retaining valuable clientele, enhancing overall satisfaction, and fostering long-term loyalty. In the Fig 5.3, the predictive output acts as a strategic guidepost, enabling businesses to optimize their retention strategies, mitigate potential churn, and ultimately strengthen customer relationships for sustained business growth within the telecom sector.

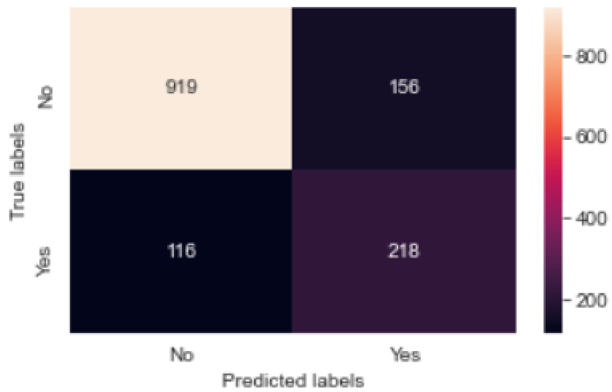

Figure 5.2 Churned Customers

	precision	recall	f1-score	support
No	0.85	0.89	0.87	1035
Yes	0.65	0.58	0.62	374
accuracy			0.81	1409
macro avg	0.75	0.74	0.74	1409
weighted avg	0.80	0.81	0.80	1409

Figure 5.3 Accuracy for churned customers

CHAPTER 6
CONCLUSION AND FUTURE WORK

6.1 CONCLUSION

This study highlights the application of deep learning techniques in predicting client retention, emphasizing their effectiveness in capturing intricate customer behaviour patterns. The telecom sector, in particular, stands to benefit significantly from this approach. By harnessing the power of deep neural networks, telecom companies can analyse extensive customer data to proactively address churn-causing factors. This predictive strategy enables the implementation of targeted retention initiatives, such as personalized offers and engagement programs, ultimately leading to reduced churn rates and heightened customer satisfaction. It is crucial to note that successful implementation of deep learning models requires high-quality data and ongoing refinement. As the telecom industry evolves, adopting deep learning for churn prediction will be imperative for maintaining competitiveness and securing customer loyalty

6.2 FUTURE WORK

In Phase I of the project, the initial step is to meticulously identify and scrutinize the diverse inputs required for various machine learning models. This involves a thorough examination of numerous papers and e-journals to derive essential insights and propose goals for the project. Building upon this research, a comprehensive system architecture and workflow are meticulously formulated to guide the project's direction and implementation. As the project transitions into Phase II, the focus shifts towards the practical execution of the proposed work. This involves the actual implementation of the suggested strategies and methodologies, followed by a detailed analysis of the obtained results. To further enhance the precision and effectiveness of forecasting, the project will advance into real-time execution, harnessing the power of cutting-edge deep learning and machine learning models. This strategic integration aims to leverage the capabilities of these advanced technologies for more accurate predictions and insights.

APPENDIX

CODE

Importing Libraries

```
from slearn.preprocessing import MinMaxScaler,StandardScaler
from sklearn.model_selection import GridSearchCV, RandomizedSearchCV,
train_test_split, StratifiedKFold, cross_val_score
from sklearn.ensemble import RandomForestClassifier
from sklearn.linear_model import LogisticRegression
from sklearn.svm import SVC
from sklearn import metrics
from imblearn.over_sampling import SMOTE
from imblearn.under_sampling import NearMiss,CondensedNearestNeighbour
from sklearn.metrics import classification_report, confusion_matrix,
ConfusionMatrixDisplay
import xgboost as xgb
from xgboost import XGBRegressor
from sklearn.neural_network import MLPClassifier
import warnings
warnings.filterwarnings('ignore')
sns.set_style("darkgrid")
df = pd.read_csv('WA_Fn-UseC_-Telco-Customer-Churn (1).csv')
df
```

Separate 20% of the data for testing

```
train_df, test_df = train_test_split(df, test_size=0.2, random_state=42, stratify=df['Churn'])
train_df.reset_index(drop=True, inplace=True)
test_df.reset_index(drop=True, inplace=True)
train_df
display(train_df.info())
```

Data Cleaning

```
train_df['TotalCharges'] = pd.to_numeric(train_df['TotalCharges'], errors='coerce')
train_df.isna().sum()
train_df['TotalCharges'].fillna((train_df['TotalCharges'].mean()), inplace=True)
train_df.duplicated().sum()
```

Exploaratory Data Analysis

```
print(train_df['Churn'].value_counts())

_ = sns.countplot(x='Churn', data=train_df, palette='crest')

cat_cols =
['gender','SeniorCitizen','Partner','Dependents','PhoneService','MultipleLines','InternetServic
e',
            'OnlineSecurity', 'OnlineBackup', 'DeviceProtection', 'TechSupport', 'StreamingTV',
'StreamingMovies',
             'Contract', 'PaperlessBilling']

plt.figure(figsize=(15, 15))

for n, variable in enumerate(cat_cols):

    ax = plt.subplot(5, 4, n + 1)

    g=sns.countplot(data=train_df, x=train_df[variable], ax=ax, palette='crest')


plt.show()

plt.figure(figsize=(11,3))

_ = sns.countplot(x= 'PaymentMethod', hue='Churn', data=train_df, palette='crest')

plt.show()

plt.tight_layout()

plt.figure(figsize=(10,10))

plt.subplot(3, 2, 1)

_ = sns.boxplot(y= train_df['tenure'])

plt.subplot(3, 2, 2)

_ = sns.histplot(x='tenure', data=train_df)

plt.subplot(3, 2, 3)
```

```
_ = sns.boxplot(y= train_df['MonthlyCharges'])
plt.subplot(3, 2, 4)
_ = sns.histplot(x='MonthlyCharges', data=train_df)
plt.subplot(3, 2, 5)
_ = sns.boxplot(y= train_df['TotalCharges'])
plt.subplot(3, 2, 6)
_ = sns.histplot(x='TotalCharges', data=train_df)
plt.figure(figsize=(10,4))
sns.scatterplot(data=train_df, x='MonthlyCharges', y='TotalCharges', hue='Churn')
train_df.replace(['No internet service','No phone service'], 'No', inplace=True)
```

Feature Engineering

```
condition = [((train_df.tenure >= 0)&(train_df.tenure <= 12)), ((train_df.tenure > 12)&(train_df.tenure <= 24)),
        ((train_df.tenure > 24)&(train_df.tenure <= 36)),((train_df.tenure > 36)&(train_df.tenure <= 48)),
        ((train_df.tenure > 48)&(train_df.tenure <= 60)), (train_df.tenure > 60)]

#choice = ['0-1year','1-2years', '2-3years', '3-4years','4-5years','more than 5 years']
choice = [0,1, 2, 3, 4, 5]
train_df['tenure_range'] = np.select(condition, choice)
_ = sns.countplot(x= 'tenure_range', hue='Churn', data=train_df, palette='crest', order=choice)
plt.tight_layout()
train_df['MonthlyCharges']=np.log1p(train_df['MonthlyCharges'])
train_df['TotalCharges']=np.log1p(train_df['TotalCharges'])
plt.figure(figsize=(15,2))
plt.subplot(1, 3, 2)
_ = sns.histplot(x='MonthlyCharges', data=train_df)
plt.subplot(1, 3, 3)
```

```
_ = sns.histplot(x='TotalCharges', data=train_df)
```

Training and Testing

```
def test_prep (test_df):
    ### Data cleaning
    #Converting 'TotalCharges' column to numeric
    test_df['TotalCharges'] = pd.to_numeric(test_df['TotalCharges'], errors='coerce')

    #Replacing 'No internet service' and 'No phone service' with 'No'
    test_df.replace(['No internet service','No phone service'], 'No', inplace=True)
    # if there is null values in the continous features --> fill with the mean of columns in training set (mapping)

    for col in test_df.columns:
        if test_df[col].isna().sum() > 0:
            test_df[col] = test_df[col].fillna(train_df[col].map(np.mean))
    ### Categorical features encoding
    test_df = pd.concat([test_df, pd.get_dummies(test_df[cat_cols])], axis='columns')
    test_df = test_df.drop(columns=cat_cols)
    test_df['Churn'] = np.where(test_df['Churn'] == 'Yes', 1, 0)

    ### Feature engineering
    #Binning 'tenure' feature into 6 ranges
    condition = [((test_df.tenure >= 0)&(test_df.tenure <= 12)), ((test_df.tenure > 12)&(test_df.tenure <= 24)),
                 ((test_df.tenure > 24)&(test_df.tenure <= 36)),((test_df.tenure > 36)&(test_df.tenure <= 48)),
                 ((test_df.tenure > 48)&(test_df.tenure <= 60)), (test_df.tenure > 60)]
    #choice = ['0-1year','1-2years', '2-3years', '3-4years','4-5years','more than 5 years']
    choice = [0,1, 2, 3, 4, 5]
```

```
    test_df['tenure_range'] = np.select(condition, choice)

    ### Feature Scaling
    test_df['MonthlyCharges']=np.log1p(test_df['MonthlyCharges'])
    test_df['TotalCharges']=np.log1p(test_df['TotalCharges'])
    return test_df

test_df = test_prep(test_df)
test_df
X_test = test_df.drop(columns=['customerID','Churn'])
y_test = test_df['Churn']
model_logReg = LogisticRegression(C=200, max_iter=1000)
scores = cross_val_score(model_logReg, X_train, y_train, cv=10, scoring="f1")
print(scores)
model_logReg.fit(X_train, y_train)
log_pred = model_logReg.predict(X_test)
cm = confusion_matrix(log_pred, y_test)
f, ax= plt.subplots(1,1,figsize=(5,3))
sns.heatmap(cm, annot=True, fmt='g', ax=ax)

ax.set_xlabel('Predicted labels'); ax.set_ylabel('True labels') ; ax.set_title('Confusion Matrix')
ax.xaxis.set_ticklabels(['No', 'Yes']) ; ax.yaxis.set_ticklabels(['No', 'Yes'])
print(classification_report(y_test, log_pred, target_names=['No', 'Yes']))
```

PUBLICATION STATUS

ANANDHI. S, ILAKIA S R, SHRUTHI K. (2023, October). CUSTOMER CHURN PREDICTION USING DEEP LEARNING. Scopus-Journal.

TITLE	**:**	CUSTOMER CHURN PREDICTION USING DEEP LEARNING
AUTHOR	**:**	ANANDHI. S, ILAKIA S R, SHRUTHI K
CONFERENCE NAME	**:**	International Conference on Intelligent Computing and Smart Communication Systems (ICICSCS'23)
JOURNAL	**:**	Scopus Indexed -Journal
MODEL OF PUBLICATION	**:**	ONLINE
STATUS	**:**	ACCEPTED

REFERENCES

1. Agarwal, V., Taware, S., Yadav, S. A., Gangodkar, D., Rao, A. L. N., and Srivastav, V. K. (2022, October). Customer-Churn Prediction Using Machine Learning. In 2022 2nd International Conference on Technological Advancements in Computational Sciences (ICTACS) (pp. 893-899). IEEE.
2. Agrawal, S., Das, A., Gaikwad, A., and Dhage, S. (2018, July). Customer churn prediction modelling based on behavioural patterns analysis using deep learning. In 2018 International conference on smart computing and electronic enterprise (ICSCEE) (pp. 1-6). IEEE.
3. Ahmad, A. K., Jafar, A., and Aljoumaa, K. (2019). Customer churn prediction in telecom using machine learning in big data platform. Journal of Big Data, 6(1), 1-24.
4. Ahmed, Ammar and Maheswari, D. (2017). A review and analysis of churn prediction methods for customer retention in telecom industries. 1-7.
5. Amol Chole, Rushikesh Chavan, Shradha Mandan and Trupti Mate. (2023). Customers Churn Prediction using Machine Learning and Deep Learning. In 2023 Journal of Emerging Technologies and Innovative Research, ISSN-2349-5162.
6. Anvita Gupta, Komanduri Srikar and Sreenitya Mandav. (2022) 'Customer Churn Prediction in Banking Sector - A Hybrid Approach', International Journal for Research in Applied Science & Engineering Technology (IJRASET), pp. 1-6.
7. Bhatnagar, A., and Srivastava, S. (2019, December). A robust model for churn prediction using supervised machine learning. In 2019 IEEE 9th international conference on advanced computing (IACC) (pp. 45-49). IEEE.
8. Bhuse, P., Gandhi, A., Meswani, P., Muni, R., and Katre, N. (2020, December). Machine learning based telecom-customer churn prediction. In 2020 3rd International Conference on Intelligent Sustainable Systems (ICISS) (pp. 1297-1301). IEEE.
9. Bin, Luo & Peiji, Shao and Juan, Liu. (2007). Customer Churn Prediction Based on the Decision Tree in Personal Handyphone System Service. Proceedings - ICSSSM'07: 2007 International Conference on Service Systems and Service Management. 1 - 5.
10. Çelik, O., and Osmanoglu, U. O. (2019). Comparing to techniques used in customer chum analysis. Journal of Multidisciplinary Developments, 4(1), 30-38.
11. Fujo, S. W., Subramanian, S., and Khder, M. A. (2022). Customer churn prediction in

telecommunication industry using deep learning. Information Sciences Letters, 11(1), 24.

12. Gaur, A., and Dubey, R. (2018, December). Predicting customer churn prediction in telecom sector using various machine learning techniques. In 2018 International Conference on Advanced Computation and Telecommunication (ICACAT) (pp. 1-5). Ieee.
13. Hu, J., Zhuang, Y., Yang, J., Lei, L., Huang, M., Zhu, R., and Dong, S. (2018, December). pRNN: A recurrent neural network based approach for customer churn prediction in telecommunication sector. In 2018 IEEE International Conference on Big Data (Big Data) (pp. 4081-4085). IEEE.
14. Ismail, M. R., Awang, M. K., Rahman, M. N. A., and Makhtar, M. (2015). A multi-layer perceptron approach for customer churn prediction. International Journal of Multimedia and Ubiquitous Engineering, 10(7), 213-222.
15. Karvana, K. G. M., Yazid, S., Syalim, A., and Mursanto, P. (2019, October). Customer churn analysis and prediction using data mining models in banking industry. In 2019 international workshop on big data and information security (IWBIS) (pp. 33-38). IEEE.
16. Kumar, P., and Kumar, S.V. (2023). DDoS Attack Prediction System Using Machine Learning Algorithms. In: Tuba, M., Akashe, S., Joshi, A. (eds) ICT Systems and Sustainability. ICT4SD 2023. Lecture Notes in Networks and Systems, vol 765. Springer, Singapore.
17. Kumar, P., Vinodh Kumar, S., and Priya, L. (2023). An Approach for Energy-Efficient Resource Allocation Through Early Planning of Virtual Machines to Servers in Cloud Server Farms. In: Doriya, R., Soni, B., Shukla, A., Gao, XZ. (eds) Machine Learning, Image Processing, Network Security and Data Sciences. Lecture Notes in Electrical Engineering, vol 946. Springer, Singapore.
18. Maw, M., Haw, S. C., and Ho, C. K. (2019, November). Customer Churn Prediction in Telecommunication: An Analysis on Issues, Techniques and Future Trends. In 2019 IEEE Conference on Sustainable Utilization and Development in Engineering and Technologies (CSUDET) (pp. 148-153). IEEE.
19. Rahman, M., and Kumar, V. (2020, November). Machine learning based customer churn prediction in banking. In 2020 4th international conference on electronics, communication and aerospace technology (ICECA) (pp. 1196-1201). IEEE.
20. Sudharsan, R., and Ganesh, E. N. (2022). A Swish RNN based customer churn prediction

for the telecom industry with a novel feature selection strategy. Connection Science, 34(1), 1855-1876.

21. Zhang, T., Moro, S., and Ramos, R. F. (2022). A data-driven approach to improve customer churn prediction based on telecom customer segmentation. Future Internet, 14(3), 94.

22. Zhao, J., Dang, X. H. (2008, October). Bank customer churn prediction based on support vector machine: Taking a commercial bank's VIP customer churn as the example. In 2008 4th International Conference on Wireless Communications, Networking and Mobile Computing (pp. 1-4). IEEE.

Customer Churn Prediction

ORIGINALITY REPORT

8%	5%	2%	3%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to National Institute Of Technical Teachers' Training & Research Student Paper	1%
2	www.mdpi.com Internet Source	1%
3	dspace.daffodilvarsity.edu.bd:8080 Internet Source	<1%
4	link.springer.com Internet Source	<1%
5	eprints.utar.edu.my Internet Source	<1%
6	www.ijraset.com Internet Source	<1%
7	Jing Zhao. "Bank Customer Churn Prediction Based on Support Vector Machine: Taking a Commercial Bank's VIP Customer Churn as the Example", 2008 4th International Conference on Wireless Communications Networking and Mobile Computing, 10/2008 Publication	<1%

Customer Churn Prediction using Deep Learning

S Anandhi[1], Ilakia S R[2], Shruthi K[3]

[123]*Computer Science and Engineering Department Department,*
Rajalakshmi Engineering College Chennai
[1]*anandhi.s@rajalakshmi.edu.in*
[2]*200701309@rajalakshmi.edu.in*
[3]*200701237@rajalakshmi.edu.in*

***Abstract*— The issue of customers ending their association with a business, known as customer churn, poses a substantial challenge to businesses in a variety of industries. Because it's usually less expensive to keep current clients than to get new ones, churn prediction is an essential component of business strategy. In order to create predictive models for customer attrition, we investigate in this study the use of historical customer data, including demographics, transaction history, customer interactions, and usage habits. Data preparation, feature engineering, model selection, and evaluation are the steps in our methodology. We talk about how crucial feature engineering and selection are to producing useful input features for predictive models. To create precise churn prediction models, major machine learning methods including Random Forest, Gradient Boosting and Neural Networks are used. Metrics are part of the model performance review process. The evaluation of model performance includes metrics such as precision , F1-score, accuracy, ROC-AUC, and recall offering a comprehensive view of predictive capabilities.**

Keywords - Churn, Customer, Demographics, Data preprocessing, Feature engineering, Model selection, Evaluation, Feature selection, Input features

I INTRODUCTION

In today's fiercely competitive landscape, customer churn remains a pivotal concern for industries reliant on subscription-based services, particularly in the domains of telecommunications and banking. The cost-effectiveness of retaining existing clientele over acquiring new customers has propelled businesses to invest heavily in predictive models aimed at pre-empting customer attrition. This survey paper delves into the extensive body of research surrounding customer churn prediction, focusing on its application in the dynamic realms of telecommunications and banking sectors. The telecommunications sector, driven by rapid technological innovations, has witnessed a surge in the adoption of triple play services across European markets. However, this unprecedented growth has been paralleled by escalating pressure on providers to deliver seamless voice, data, and video services. Customer churn, defined as the migration of subscribers to competitor services, poses a significant threat to industry profitability, necessitating the establishment of effective churn prediction models. Existing churn prediction models have grappled with a multitude of challenges, including the oversight of temporal customer behaviour, disregard for class rarity, and the inability to conclusively identify underlying churn causes. To address these limitations, various machine learning algorithms have been employed, providing a robust foundation for enhancing customer retention strategies. The banking sector, likewise, confronts the imperative of accurately forecasting customer churn to optimize resource allocation. This study delves into a comparative analysis of classification models, encompassing Logistic Regression, KNN , Decision Tree Classifier, Voting Classifier, Stochastic Gradient Descent, Naive Bayes, SVM and Random

Forest. The aim is to discern the most suitable model for predicting churn, thereby curtailing resource wastage. Customer relationship management (CRM) assumes paramount importance in customer retention efforts, making accurate churn prediction an indispensable tool for marketers. This survey underscores the spectrum of supervised machine learning classifiers, including Neural Networks, Random Forests, and SVM. Additionally, it spotlights a novel approach leveraging deep learning through RNN models equipped with LSTM units to discern patterns in customer data. The authors present an innovative deep learning framework for customer churn prediction in the telecommunications sector, harnessing the power of RNN with LSTM units. This model showcases exceptional accuracy in churn prediction, validated through metrics such as F1-score, AUC, Maximum Profit measure, and Expected Maximum Profit measure on real-world telecommunications datasets. By synthesizing a wealth of research findings, this survey paper offers a comprehensive overview of the evolving landscape of customer churn prediction models. The convergence of technological advancements and analytical methodologies in this domain presents an invaluable resource for industry stakeholders seeking to fortify their customer retention strategies in an increasingly competitive market.

II RELATED WORK

(Agarwal et al., 2022) customer churn, or the progressive drop in repeat business, is a major worry for organizations, particularly in the banking industry. Early switcher detection supports proactive retention initiatives. In order to predict the possibility of customer churn, this article uses machine learning, specifically Logistic Regression (LR) and Naive Bayes (NB), along with information about the client's age, geography, credit history, and balance. NB is revealed to be the best model. The paper promotes enhanced churn prediction methodologies, highlighting the combination of methods, such as LGBM-Classifiers and boosting procedures, for improved accuracy and performance. These developments show promise for future churn computation, offering priceless support for customer retention initiatives.

(Agrawal et al., 2018) In Customer churn, or when a group of customers stop using a company's services, has an effect on a company's profitability and stability, hence it is important to foresee it. Deep Learning stands out because it can find important features in complicated, unstructured datasets like those from the telecom industry. Analyzing the major factors significantly determining churn rates is necessary for understanding and predicting churn. The study uses deep learning to predict churn with an accuracy of 80.03%, showing characteristics that are important for client retention efforts. This study enables businesses to focus on important factors for customer retention, reducing losses to rivals.

(Ahmad et al., 2019) this project presents the essential issue of customer turnover is addressed in this study, particularly in the telecom industry where it has a direct impact on revenues. A churn prediction model is created using machine learning techniques on a big data platform along with cutting-edge feature engineering and selection approaches. The area under the curve (AUC) score for the model is an amazing 93.3%. Notably, adding Social Network Analysis (SNA) elements improves performance even more, increasing AUC from 84% to 93.3%. The model surpasses the XGBOOST method for classification when tested on a sizable dataset from the telecom firm SyriaTel. This study is helpful in guiding profitability and reducing customer churn for telecom firms like SyriaTel.

(Ahmed et al., 2017) the Customer turnover, a crucial component of the telecom industry, frequently results from discontent or better deals from rivals. Churn prediction and prevention are crucial, driving businesses to use various data mining and machine learning techniques. Customized products and effective retention techniques are required due to the intense competition. Hybrid methods are the most accurate, however existing techniques such as meta-heuristics and machine learning show effective churn prediction. This study highlights not just precise churn prediction but also examines churn causes and method shortcomings, laying the groundwork for possible hybrid model development in subsequent studies.

(Amol Chole et al., 2023) In Large businesses, especially those in the telecom sector, face a substantial difficulty as a result of customer turnover, necessitating the development of reliable prediction techniques. By creating a churn prediction model using machine learning and deep learning techniques on a sizable dataset derived from GitHub, this study makes a contribution. The model performed better than expected when tested using algorithms including Random Forest, SVM, KNN, and CNN. The Random Forest approach produced an accuracy of 83.11%. In order to improve churn prediction, future research will concentrate on fine-tuning hyperparameters and investigating various machine learning techniques for feature selection and resampling data.

(Anvita Gupta et al., 2022) predicting customer churn is crucial for banks to proactively engage with at-risk customers and prevent attrition. Early intervention alone can reduce churn by 11%. Utilizing past customer data through machine learning and data science techniques offers a solution. This study compares various churn prediction models used by financial organizations, ultimately advocating for a hybrid approach. Results indicate that this hybrid method outperforms existing models and voting classifiers, showcasing its superior accuracy. This underscores the importance of feature impact assessment and dataset clustering for tailored prediction. Future work may involve further sub-clustering and employing additional classification algorithms for enhanced accuracy and outlier mitigation.

(Bhatnagar et al., 2019) customer churn is a challenge for businesses because of fierce competition and a wide range of telecommunication services. Potential churners must be early identified for retention strategies to succeed and be profitable. Churn categorization encompasses both voluntary and involuntary churn, with an emphasis on anticipating purposeful churn. This forecast, a task for supervised classification, aids companies in retaining customers and lowering customer acquisition expenses. Machine learning classifiers like Logistic Regression, Support Vector Machine and Decision Tree are frequently employed in churn prediction models. This paper evaluates the state-of-the-art in churn prediction research, highlighting issues and suggesting future research directions while providing sage advice for young researchers. However, the research is limited in time and focuses on consumer-initiated turnover.

(Bhuse et al., 2020) delivers that the customers have many options in today's competitive market, making client turnover a critical concern for banks. In order to retain engagement, this article uses machine learning approaches to forecast client attrition in the banking industry. The study examines consumer behaviour by classifying data using KNN, SVM, Decision Tree, and Random Forest classifiers, as well as feature selection techniques. Following oversampling, experimental results on a Kaggle churn modelling dataset favoured the Random Forest model, displaying improved accuracy. The study stresses the significance of early-stage churn prediction in the banking industry and offers insights for larger-scale applications while using a very small, unbalanced dataset. The results also show how important oversampling is for resolving data imbalances, especially when applied to SVM classifiers.

(Bin et al., 2007) customer attrition prediction is essential for profitability in the cutthroat Chinese telecom market. Even with insufficient customer data, it's crucial to improve attrition models. The paper recommends decision tree-based experimentation for efficient churn prediction. The churn model's recall rate, precision rate, and F-measure increased as a result of changing the sub-periods of training data, misclassification costs, and sample techniques. With the use of this technique, China Telecom can successfully predict and control customer churn, increasing customer retention in a cutthroat industry. To further improve prediction churn models in related scenarios, future research could investigate alternative data mining techniques.

(Celik et al., 2019) minimizing expenses is essential in today's fiercely competitive environment.

Research shows that keeping existing customers costs ten times less than recruiting new ones, highlighting the importance of customer churn monitoring. In the context of customer churn analysis, this paper examines a number of machine learning methods, including ANN, decision tree, SVM, naive bayes, knn, and XG Boost. Machine learning algorithms are considered to be trustworthy substitutes for time-related event estimates, such as customer turnover, even though deep learning approaches exhibit greater performance in complex circumstances. While deep learning approaches excel in complex structures, the Cox regression model efficiently analyses independent variables influencing temporal variables and risk groups. Deep learning techniques are expected to continue to progress and produce even higher success rates over time.

(Fujo et al., 2022) this study addresses the pressing issue of customer churn in the telecom industry by implementing a Deep-BP-ANN model, bolstered by feature selection methods and overfitting prevention techniques. The model outperforms traditional ML techniques like KNN, Logistic Regression, XG Boost, and Naïve Bayes on real datasets (IBM Telco and Cell2cell) with an accuracy exceeding 88%. Lasso regression proves pivotal for feature selection, particularly in datasets with numerous attributes. The ROS technique effectively balances the datasets, and activity regularization aids in mitigating overfitting. Fine-tuning parameters, such as neuron count and epoch number, significantly enhance performance. This model sets a new benchmark, outclassing previous deep learning techniques like CNN, LFNN, and ANN.

(Gaur et al., 2018) says that in Churn research, which makes use of data mining, forecasts client attrition, which is essential in today's cutthroat marketplaces. Predicting customer loss improves marketing, customer loyalty, and communication, which has an effect on profitability. To efficiently retain customers, businesses, particularly telecom providers, concentrate on identifying customer churn factors. Gradient Boosting emerges as the most efficient, followed by Logistic Regression and Random Forest, with SVM performing somewhat less well. These machine learning models include Logistic Regression, SVM, Random Forest, and Gradient Boosted Tree.

(Hu et al., 2018) client churn prediction uses a variety of machine learning classifiers and is essential for client retention and current CRM. Time series customer data analysis is now possible because to recent advancements in data technology, improving accuracy. A pRNN model with LSTM units and product operations has been proposed, and it exhibits great accuracy in predicting churn in the telecom industry. The article covers potential future research topics and emphasizes the importance of recurrent neural networks in processing sequential input. To validate findings and investigate long-term prediction views, additional diversified real-world datasets are required.

(Ismail et al., 2015) the customer management is essential in the telecommunications sector to prevent churn. The large expenses involved with adopting it across the whole customer base can be avoided with targeted retention initiatives for likely churn clients. Utilizing historical churn data and predictive factors, churn management focuses on prediction. While long-term success is assured by keeping existing clients, traditional marketing places greater emphasis on obtaining new ones. With a prediction accuracy of 91.28%, Multilayer Perceptron Artificial Neural Network outperforms conventional statistical models in predicting customer attrition. These information should be used in customer retention initiatives to effectively lower churn rates.

(Karvana et al., 2019) the customer attrition in banking may be accurately predicted by data mining. Recall rates are highly influenced by sample size and inter-class comparisons, favoring a 50:50 data ratio with a 70% recall. Each class has roughly 7,975 samples out of approximately 15,949 data samples. The 50:50 SVM sampling model is the most effective one, which identifies important characteristics like vintage, EDC transaction volume and amount, average balance, and age and generates a large profit of 456 billion. This is consistent with the research, which highlights SVM's accuracy while highlighting Logistic Regression's ability to reduce losses.

(Kumar, P et al., 2023) distributed denial-of-service (DDoS) attacks pose a significant threat to the confidentiality and integrity of computer networks, disrupting web traffic to target servers and impeding authorized user access to services. Detection of DDoS attacks can be challenging, requiring robust mitigation strategies due to the diverse methods used to flood networks or servers. These assaults leverage resource limitations, impacting the functionality of the targeted organization's website infrastructure. Analyzing the most recent datasets is crucial for identifying and understanding the evolving landscape of DDoS attacks, assessing their varied techniques, and evaluating their efficacy. Clients accessing network services are consistently exposed to this pervasive and severe threat, necessitating ongoing vigilance and proactive security measures.

(Kumar, P. et al., 2023) delves into the pressing issue of increasing energy consumption within cloud server farms, highlighting their substantial contribution to environmental pollution resulting from heightened power usage. This study accentuates the complexities associated with mitigating power consumption while upholding agreements concerning service quality. To address this challenge, the paper proposes a solution centered on optimizing resource allocation. This involves a strategic approach that limits the operation of dynamic servers, thereby aiming to curtail energy usage while simultaneously meeting the demands of clients and ensuring efficient task performance. To validate the efficacy of their proposed algorithms, the researchers leverage Cloud Sim, a simulation tool, utilizing real-world data obtained from a significant pool of over 1000 Planet Lab virtual machines. The study underscores the pivotal role played by server farms in this evolving technological landscape, emphasizing the critical need to strike a balance between energy conservation and maintaining high-quality service provision.

(Maw et al., 2019) the companies have a problem from customer churn, which is a result of severe competition and a variety of telecommunication services. For retention initiatives to be successful and profitable, potential churners must be identified quickly. With an emphasis on foreseeing intentional churn, churn categorization includes both voluntary and involuntary churn. This forecast, a supervised categorization task, helps businesses keep clients and cut acquisition costs. In churn prediction models, machine learning classifiers like Support Vector Machine, Logistic Regression, and Decision Trees are widely used. This paper examines current churn prediction research, noting problems and potential areas for future research while offering insightful advice for up-and-coming scientists. The research, though, is time-bound and concentrates on churn that is instigated by the consumer.

(Rahman et al., 2020) the customers have many options in today's competitive market, making client turnover a critical concern for banks. In order to retain engagement, this article uses machine learning approaches to forecast client attrition in the banking industry. The study examines consumer behaviour by classifying data using KNN, SVM, Decision Tree, and Random Forest classifiers, as well as feature selection techniques. Following oversampling, experimental results on a Kaggle churn modelling dataset favoured the Random Forest model, displaying improved accuracy. The study stresses the significance of early-stage churn prediction in the banking industry and offers insights for larger-scale applications while using a very small, unbalanced dataset. The results also show how important oversampling is for resolving data imbalances, especially when applied to SVM classifiers.

(Sudharsan et al., 2022) in the cutthroat and fast-paced telecom sector, client turnover is a significant problem that needs to be addressed. S-RN0N is a unique framework for precise churn prediction. The model divides clients into churners and non-churners, and if churn is anticipated, it prompts further study for retention tactics. Data collection, preprocessing, filtering, grouping, feature engineering, and classification are all included in the suggested methodology. The S-RNN model performs admirably in experimental analysis, attaining outstanding metrics like 98.27% sensitivity, 92.31% specificity, and 95.99% accuracy. The suggested method also performs better in terms of resilience and reliability than current methods. Future studies might examine changing consumer behaviour patterns utilizing cutting-edge forecasting techniques and trend analysis.

(Zhang et al., 2022) the telecom companies face a pressing challenge with client churn, impacting profits in a saturated global market. Although attracting new clients is expensive, keeping the ones you already have is more cost-effective. Predicting and preventing customer churn has become a top priority for telecom companies. This study introduces discriminant and logistic regression models using customer segmentation data from major Chinese telecom firms. The findings empower managers to accurately predict customer behaviour, enhance retention strategies, and optimize budgets. Notably, this research fills a gap in telecom customer churn studies by employing Fisher discriminant and logistic regression analyses, offering valuable insights for industry improvement.

(Zhao et al., 2008) the customer loyalty has a bigger impact on bank profits than things like growth and market share. client churn lowers sales and new client acquisition. Data mining provides for accurate churn prediction and customized marketing tactics. With excellent accuracy and practical considerations, a support vector machine (SVM) model beat other classifiers in predicting bank customer attrition. SVM is a reliable method for churn prediction because of its straightforward classification surface, good generalization, and fitting accuracy.

III SYNOPTIC OVERVIEW

Author / Year	Approach	Merits
(Agarwal et al., 2022)	It primarily focuses on SVM, Naive Bayes	The study demonstrates the successful implementation of machine learning algorithms in predicting customer churn, with Naive Bayes showing notably high accuracy.
(Agrawal et al., 2018)	It utilizes Artificial Neural Network	The paper demonstrates the effectiveness of Deep Learning techniques, particularly ANNs, in mobile network churn prediction. It highlights the stability of the model across different months and emphasizes the improvement achieved by incorporating location data.
(Ahmad et al., 2019)	XGBOOST, Random Forest, GBM, Decision Trees are utilised	The utilization of XGBOOST led to an impressive accomplishment, achieving a notably high AUC value of 93.301%. This outcome notably surpassed the performance of various other algorithms that were put to the test. The integration of Social Network Analysis features played a pivotal role in elevating the accuracy.
(Ahmed et al., 2017)	It goes through techniques such as SVM, Neural Networks, PSO, Anti Miner+	Hybrid models combining SVM, ANN, and SOM exhibit notably high accuracy levels while maintaining lower complexity, showcasing their efficiency. Rule induction techniques and PSO (Particle Swarm Optimization) prove to be effective methodologies for predicting churn with promising outcomes.

(Amol Chole et al., 2023)	It involves Random Forest, K - Nearest Neighbours and Support Vector Machine.	The model achieves a notable accuracy in predicting customer churn, which is crucial for revenue retention in the highly competitive telecom industry.
(Anvita Gupta et al., 2022)	It utilizes Decision Tree Classifier, Logistic Regression, Stochastic Gradient Descent, Support Vector Machine, K-Nearest Neighbours, Voting Classifier, Random Forest, Naive Bayes.	It emphasizes identifying impactful features and clustering data to apply diverse models for precise predictions. The method showcases potential to extend to multiple sub-clusters for enhanced accuracy.
(Bhatnagar et al., 2019)	It involves KNN, Logistic Regression	The study compares KNN and Logistic Regression for customer churn prediction, favoring KNN with 2.0% higher accuracy.
(Bhuse et al., 2020)	Random Forest, SVM, XG Boost, Ridge classifier, KNN, Deep Neural Network.	The paper provides a comprehensive analysis of different techniques for customer churn prediction in the telecom sector, offering insights into the effectiveness of various algorithms. The Random Forest model demonstrated the highest accuracy, indicating its potential for real-world applications.
(Bin et al., 2007)	It employs Decision tree	Effective utilization of decision tree for diverse applications in different domains, showcasing its versatility and usefulness in decision-making processes.
(Celik et al., 2019)	It includes Cox Regression, ANN, KNN, Decision Tree, SVM, Logistic Regression, Naïve Bayes, XG Boost	Accurate prediction of event times for customer churn (Martinsson, 2016) Improved accuracy in predicting customer churn compared to traditional methods
(Fujo et al., 2022)	It is Deep-BP-ANN, CNN, LFNN, Transfer learning, ensemble – classifiers.	Implemented Deep-BP-ANN model outperformed other machine learning techniques, achieving higher accuracy and efficiency.
(Gaur et al., 2018)	Logistic Regression, SVM, Gradient Boosting, Random Forest	The study highlights gradient boosting's effectiveness in predicting customer churn for telecom, emphasizing its role in customer retention strategies.
(Hu et al., 2018)	Logistic Regression, Random Forest, PNN1, LSTM, pRNN, pLSTM, LSTMNN	pRNN demonstrates superior performance in customer churn prediction, offering valuable insights for customer relationship management and marketing strategies
(Ismail et al., 2015)	Neural Networks, Logistic Regression, SVM, Bayes Network, Rough Set Theory, K-Means, Time	Neural network, specifically the MLP model, demonstrated superior prediction accuracy (91.28%) compared to traditional statistical models like regression analysis.

	Series, Regression Forests, Association Rules	
(Karvana et al., 2019)	Neural Network, Decision Tree, SVM, Logistic Regression, Naïve Bayes	This project serves as a demonstration of data mining techniques applied specifically to churn prediction, underscoring the crucial significance of sample size in this predictive analysis.
(Kumar, P et al., 2023)	The approach analyzes real-time datasets to detect and mitigate DDoS attacks, evaluating various attack vectors for effective defense strategies.	Enhances network resilience by utilizing real-time datasets for accurate DDoS attack analysis. Enables the development of effective defense strategies by evaluating various attack vectors, ensuring uninterrupted service for authorized users.
(Kumar, P et al., 2023)	The approach optimizes cloud server farm resource allocation to cut energy consumption, validated using CloudSim	It addresses escalating carbon emissions in cloud computing, striving for efficiency by reducing environmental impact while maintaining service quality.
(Maw et al., 2019)	CCP, SVM, Random Forest, Neural Network, Decision Tree	Comprehensive analysis of recent literature in churn prediction, revealing emerging research opportunities, and providing insights into the challenges faced in the telecom industry.
(Rahman et al., 2020)	It involves SVM, Random Forest, Decision Tree, K - Nearest Neighbours	The SVM model, especially with integrated random sampling, improved predictive accuracy and precision in churn prediction compared to other models. It addressed data imbalance issues through resampling techniques.
(Sudharsan et al., 2022)	CLARA, BM-BOA, Feature selection, Classification, retention process, Swish RNN.	Achieves high accuracy and sensitivity .Efficient clustering and feature selection methodologies improve performance. Utilizes Swish RNN, a novel approach, for churn prediction.
(Zhang et al., 2022)	This employs Decision tree, logistic regression, Cluster analysis.	The attainment of a substantial accuracy rate of 93.94% in forecasting customer churn holds significant implications for telecom companies, offering invaluable insights to fine-tune operational costs..
(Zhao et al., 2008)	SVM Logistic, Regression Naïve, Bayes ANN	Support Vector Machines (SVM) have demonstrated superior performance over an array of classifiers, showcasing remarkable effectiveness specific in customer churn prediction.

IV METHODOLOGY

1) Logistic Regression: Whenever the dependent variable is binary, logistic regression regression analysis is useful (binary). At the nominal, ordinal, interval, or ratio levels, data can be described and the relationship about one dependent binary variable and one or maybe more independent variables can be demonstrated using logistic regression. Logistic regression is used to determine the probability ratio of a group of explanatory factors. The response variable is binomial, but other than that, the process is comparable to multiple linear regressions. The outcome is how each variable affects the overall rate of the pertinent occurrence.

2) Linear Discriminant Analysis: Using linear discriminant analysis is one way to lower dimensionality. It serves as an initial step in applications involving pattern recognition and machine learning. To minimise energy and dimension costs and prevent the curse of dimensionality, the LDA distributes the functions into a smaller dimensional area inside a greater environment.

3) K-Nearest Neighbours Classifier: Through using nearest neighbours as an illustration of a question and those neighbours to determine the query's class, the Nearest Neighbour Classifier carries out classification. This categorization approach is highly intriguing because present data centres do not solve common run-time efficiency issues.

4) Support Vector Machines: Regression is another name for support vector machines (SVMs), which are supervised learning methods for classification. In order to maximise predicted accuracy and avoid overfitting the training set, Support Vector Machines (SVM) employ regression modelling and classifiers that are already part of machine learning theory. Typically, SVMs are conceptualised as systems that employ functions inside a high-dimensional feature space, and are trained by a statistical learning strategy based on optimisation theory.

5) Random Forest: Eighty percent of the dataset is used for training with one hundred estimators, meaning that one hundred trees are built and the average of the predictions is then calculated. For every decision tree, the maximum depth is assumed to be 10. Using this method, significant traits are ranked and forecast according to decision tree voting. Thus, Random Forest provides us with excellent accuracy.

6) XG Boost: One of the classifiers is the Python implementation of the XGBoost package. Extreme Gradient Boosting, or XGBoost for short, is a boosting and gradient descent hybrid. A supervised ensemble learning approach called "boosting" operates by allocating distinct weights to the training data distribution for every iteration. To determine the ideal split, the XGBoost algorithm employs the exact greedy algorithm. Due to its high cache optimization, the XG boost classifier out performs in the prediction model; nevertheless, the iteration method necessitates more training time. The model operates with good performance and speed of execution.

V PROPOSED SYSTEM

Proposed System for Customer Churn Prediction: In this project paper, we propose the development of an advanced customer churn prediction system that leverages cutting-edge machine learning and deep learning techniques to help businesses proactively identify and retain at-risk customers. The system will involve the collection and preprocessing of historical customer data, including demographic information, transaction history, and customer interactions. Feature engineering and selection will be employed to extract relevant information, and a diverse set of machine learning algorithms, such as logistic regression, decision trees, neural networks, and Convolution Neural Networks will be implemented to model customer churn behaviour. By offering a comprehensive, data-driven approach to customer churn prediction, this system aims to help organizations reduce customer attrition, enhance customer loyalty, and ultimately improve their bottom line. To enhance the accuracy of the models we fine tune this with different hyperparameters and increase the overall accuracy of these models.

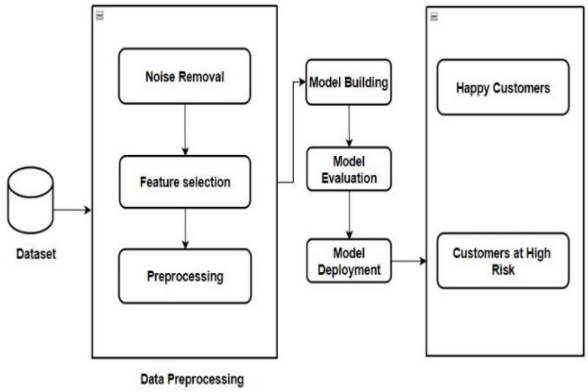

Figure 1.1 Architecture Diagram

VI CONCLUSION

This study highlights the application of deep learning techniques in predicting client retention, emphasizing their effectiveness in capturing intricate customer behaviour patterns. The telecom sector, in particular, stands to benefit significantly from this approach. By harnessing the power of deep neural networks, telecom companies can analyse extensive customer data to proactively address churn-causing factors. This predictive strategy enables the implementation of targeted retention initiatives, such as personalized offers and engagement programs, ultimately leading to reduced churn rates and heightened customer satisfaction. It is crucial to note that successful implementation of deep learning models requires high-quality data and ongoing refinement. To overcome churn, a multifaceted strategy to overcome churn. This entails enhancing customer support, tailoring communications, putting loyalty plans into place, examining data trends, and applying predictive algorithms to spot at-risk clients early. Reducing churn also involves competitive pricing, personnel training, targeted retention marketing, communication, customer feedback, and retention campaigns. By combining these strategies, businesses can increase profitability, strengthen customer loyalty, and gain competitive advantages in the fast-paced market of today.

VII REFERENCES

1. Agarwal, V., Taware, S., Yadav, S. A., Gangodkar, D., Rao, A. L. N., and Srivastav, V. K. (2022, October). Customer-Churn Prediction Using Machine Learning. In 2022 2nd International Conference on Technological Advancements in Computational Sciences (ICTACS) (pp. 893-899). IEEE.
2. Agrawal, S., Das, A., Gaikwad, A., and Dhage, S. (2018, July). Customer churn prediction modelling based on behavioural patterns analysis using deep learning. In 2018 International conference on smart computing and electronic enterprise (ICSCEE) (pp. 1-6). IEEE.
3. Ahmad, A. K., Jafar, A., and Aljoumaa, K. (2019). Customer churn prediction in telecom using machine learning in big data platform. Journal of Big Data, 6(1), 1-24.
4. Ahmed, Ammar and Maheswari, D. (2017). A review and analysis of churn prediction methods for customer retention in telecom industries. 1-7.
5. Amol Chole, Rushikesh Chavan, Shradha Mandan and Trupti Mate. (2023). Customers Churn Prediction using Machine Learning and Deep Learning. In 2023 Journal of Emerging Technologies and Innovative Research, ISSN-2349-5162.
6. Anvita Gupta, Komanduri Srikar and Sreenitya Mandav. (2022) ‘Customer Churn Prediction in Banking Sector - A Hybrid Approach’, International Journal for Research in Applied Science & Engineering Technology (IJRASET), pp. 1-6.
7. Bhatnagar, A., and Srivastava, S. (2019, December). A robust model for churn prediction

using supervised machine learning. In 2019 IEEE 9th international conference on advanced computing (IACC) (pp. 45-49). IEEE.
8. Bhuse, P., Gandhi, A., Meswani, P., Muni, R., and Katre, N. (2020, December). Machine learning based telecom-customer churn prediction. In 2020 3rd International Conference on Intelligent Sustainable Systems (ICISS) (pp. 1297-1301). IEEE.
9. Bin, Luo & Peiji, Shao and Juan, Liu. (2007). Customer Churn Prediction Based on the Decision Tree in Personal Handyphone System Service. Proceedings - ICSSSM'07: 2007 International Conference on Service Systems and Service Management. 1 - 5.
10. Çelik, O., and Osmanoglu, U. O. (2019). Comparing to techniques used in customer churn analysis. Journal of Multidisciplinary Developments, 4(1), 30-38.
11. Fujo, S. W., Subramanian, S., and Khder, M. A. (2022). Customer churn prediction in telecommunication industry using deep learning. Information Sciences Letters, 11(1), 24.
12. Gaur, A., and Dubey, R. (2018, December). Predicting customer churn prediction in telecom sector using various machine learning techniques. In 2018 International Conference on Advanced Computation and Telecommunication (ICACAT) (pp. 1-5). Ieee.
13. Hu, J., Zhuang, Y., Yang, J., Lei, L., Huang, M., Zhu, R., and Dong, S. (2018, December). pRNN: A recurrent neural network based approach for customer churn prediction in telecommunication sector. In 2018 IEEE International Conference on Big Data (Big Data) (pp. 4081-4085). IEEE.
14. Ismail, M. R., Awang, M. K., Rahman, M. N. A., and Makhtar, M. (2015). A multi-layer perceptron approach for customer churn prediction. International Journal of Multimedia and Ubiquitous Engineering, 10(7), 213-222.
15. Karvana, K. G. M., Yazid, S., Syalim, A., and Mursanto, P. (2019, October). Customer churn analysis and prediction using data mining models in banking industry. In 2019 international workshop on big data and information security (IWBIS) (pp. 33-38). IEEE.
16. Kumar, P., and Kumar, S.V. (2023). DDoS Attack Prediction System Using Machine Learning Algorithms. In: Tuba, M., Akashe, S., Joshi, A. (eds) ICT Systems and Sustainability. ICT4SD 2023. Lecture Notes in Networks and Systems, vol 765. Springer, Singapore.
17. Kumar, P., Vinodh Kumar, S., and Priya, L. (2023). An Approach for Energy-Efficient Resource Allocation Through Early Planning of Virtual Machines to Servers in Cloud Server Farms. In: Doriya, R., Soni, B., Shukla, A., Gao, XZ. (eds) Machine Learning, Image Processing, Network Security and Data Sciences. Lecture Notes in Electrical Engineering, vol 946. Springer, Singapore.
18. Maw, M., Haw, S. C., and Ho, C. K. (2019, November). Customer Churn Prediction in Telecommunication: An Analysis on Issues, Techniques and Future Trends. In 2019 IEEE Conference on Sustainable Utilization and Development in Engineering and Technologies (CSUDET) (pp. 148-153). IEEE.
19. Rahman, M., and Kumar, V. (2020, November). Machine learning based customer churn prediction in banking. In 2020 4th international conference on electronics, communication and aerospace technology (ICECA) (pp. 1196-1201). IEEE.
20. Sudharsan, R., and Ganesh, E. N. (2022). A Swish RNN based customer churn prediction for the telecom industry with a novel feature selection strategy. Connection Science, 34(1), 1855-1876.
21. Zhang, T., Moro, S., and Ramos, R. F. (2022). A data-driven approach to improve customer churn prediction based on telecom customer segmentation. Future Internet, 14(3), 94.
22. Zhao, J., Dang, X. H. (2008, October). Bank customer churn prediction based on support vector machine: Taking a commercial bank's VIP customer churn as the example. In 2008 4th International Conference on Wireless Communications, Networking and Mobile Computing (pp. 1-4). IEEE.

Customer Churn Prediction using Deep Learning

ORIGINALITY REPORT

9%	6%	5%	%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	researchspace.ukzn.ac.za Internet Source	1%
2	Nurulhuda Mustafa, Lew Sook Ling, Siti Fatimah Abdul Razak. "Customerchurnprediction fortelecommunicationindustry: A Malaysian Case Study", F1000Research, 2021 Publication	1%
3	www.ijraset.com Internet Source	1%
4	Jinlong Hu, Yi Zhuang, Jiang Yang, Lei Lei, Minjie Huang, Runchao Zhu, Shoubin Dong. "pRNN: A Recurrent Neural Network based Approach for Customer Churn Prediction in Telecommunication Sector", 2018 IEEE International Conference on Big Data (Big Data), 2018 Publication	1%
5	www.mdpi.com Internet Source	1%

CO-PO Mapping

PROJECT WORK COURSE OUTCOME (COs):

CO1: On completion the students capable of execute the proposed plan and become aware of and overcome the bottlenecks throughout every stage.

CO2: On completion of the project work students could be in a role to take in any difficult sensible issues and locate answer through formulating right methodology.

CO3: Students will attain a hands-on revel in in changing a small novel idea / method right into an operating model / prototype related to multi-disciplinary abilities and / or understanding and operating in at team.

CO4: Students will be able to interpret the outcome of their project. Students will take on the challenges of teamwork, prepare a presentation in a professional manner, and document all aspects of design work.

CO5: Students will be able to publish or release the project to society.

PROGRAM OUTCOMES (POs)

PO1: Engineering knowledge: Apply the knowledge of mathematics, science, engineering fundamentals, and an engineering specialization to the solution of complex engineering problems.

PO2: Problem analysis: Identify, formulate, review research literature, and analyze complex engineering problems reaching substantiated conclusions using first principles of mathematics, natural sciences, and engineering sciences.

PO3: Design/development of solutions: Design solutions for complex engineering problems and design system components or processes that meet the specified needs with appropriate consideration for the public health and safety, and the cultural, societal, and environmental considerations.

PO4: Conduct investigations of complex problems: Use research-based knowledge and research methods including design of experiments, analysis and interpretation of data, and synthesis of the information to provide valid conclusions.

PO5: Modern tool usage: Create, select, and apply appropriate techniques, resources, and modern engineering and IT tools including prediction and modeling to complex engineering activities with an understanding of the limitations.

PO6: The engineer and society: Apply reasoning informed by the contextual knowledge to assess societal, health, safety, legal and cultural issues and the consequent responsibilities relevant to the professional engineering practice.

PO7: Environment and sustainability: Understand the impact of the professional engineering solutions in societal and environmental contexts, and demonstrate the knowledge of, and need for sustainable development.

PO8: Ethics: Apply ethical principles and commit to professional ethics and responsibilities and norms of the engineering practice.

PO9: Individual and team work: Function effectively as an individual, and as a member or leader in diverse teams, and in multidisciplinary settings.

PO10: Communication: Communicate effectively on complex engineering activities with the engineering community and with society at large, such as, being able to comprehend and write effective reports and design documentation, make effective presentations, and give and receive clear instructions.

PO11: Project management and finance: Demonstrate knowledge and understanding of the engineering and management principles and apply these to one's own work, as a member and leader in a team, to manage projects and in multidisciplinary environments.

PO12: Life-long learning: Recognize the need for, and have the preparation and ability to engage in independent and life-long learning in the broadest context of technological change.

PROGRAM SPECIFIC OUTCOMES (PSOs)

PSO1: Foundation Skills: Ability to understand, analyze and develop computer programs in the areas related to algorithms, system software, web design, machine learning, data analytics, and networking for efficient design of computer-based systems of varying complexity. Familiarity and practical competence with a broad range of programming language and open-source platforms.

PSO2: Problem-Solving Skills: Ability to apply mathematical methodologies to solve computational task, model real world problem using appropriate data structure and suitable algorithm. To understand the Standard practices and strategies in software project development using open-ended programming environments to deliver a quality product.

PSO3: Successful Progression: Ability to apply knowledge in various domains to identify research gaps and to provide solution to new ideas, inculcatepassion towards higher studies, creating innovative career paths to be an entrepreneur and evolve as an ethically socially responsible computer science professional.

PO/PSOCO	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12	PSO1	PSO2	PSO3
CO 1	3	3	3	3	2	3	3	3	3	3	2	2	2	2	2
CO 2	3	3	3	3	3	2	2	1	3	3	2	2	3	3	3
CO 3	3	3	3	3	3	2	2	1	3	3	2	1	3	3	3
CO 4	3	3	3	2	2	1	1	1	3	3	1	-	2	2	2
CO 5	3	3	3	3	3	2	2	2	3	3	2	3	3	3	3
Average	3	3	3	5.8	2.6	2	2	1.6	3	3	1.8	2	2.6	2.6	2.6