CHURN PREDICTION AND RECOVERY: A BLOCKCHAIN BASED APPROACH TO CUSTOMER RETENTION

PROJECT REPORT

21AD1513- INNOVATION PRACTICES LAB

JOTHIMALAR M Reg. No. 211422243132

AARTHI S Reg. No. 211422243004

AKSHAYA S Reg. No. 211422243020

in partial fulfillment of the requirements for the award of degree

of

BACHELOR OF TECHNOLOGY

in

ARTIFICIAL INTELLIGENCE AND DATA SCIENCE

PANIMALAR ENGINEERING COLLEGE, CHENNAI-600123

ANNA UNIVERSITY: CHENNAI-600 025

November, 2024

BONAFIDE CERTIFICATE

Certified that this project report titled "**CHURN PREDICTION AND RECOVERY: A BLOCKCHAIN BASED APPROACH TO CUSTOMER RETENTION**" is the bonafide work of **JOTHIMALAR M (211422243132), AARTHI S (211422243004), AKSHAYA S (211422243020)** who carried out the project work under my supervision. Certified further, that to the best of my knowledge the work reported herein does not form part of any other project report or dissertation on the basis of which a degree or award was conferred on an earlier occasion on this or any other candidate.

INTERNAL GUIDE

Dr. A. JOSHI M.E.,Ph.D
Professor,
Department of AI &DS,
Panimalar Engineering College,
Chennai-600123.

HEAD OF THE DEPARTMENT

Dr. S. MALATHI M.E., Ph.D
Professor and Head,
Department of AI & DS,
Panimalar Engineering College,
Chennai-600123.

Certified that the candidate was examined in the Viva-Voce Examination held on ………………………

INTERNAL EXAMINER **EXTERNAL EXAMINER**

ACKNOWLEDGEMENT

I also take this opportunity to thank all the Faculty and Non-Teaching Staff Members of Department of Artificial Intelligence and Data Science for their constant support. Finally I thank each and every one who helped me to complete this project. At the outset we would like to express our gratitude to our beloved respected Chairman, **Dr.Jeppiaar M.A.,Ph.D,** Our beloved correspondent and Secretary **Mr.P.Chinnadurai M.A., M.Phil., Ph.D.,** andour esteemed director for their support.

We would like to express thanks to our Principal, **Dr. K. Mani M.E., Ph.D.,** for having extended his guidance and cooperation.

We would also like to thank our Head of the Department, **Dr.S.Malathi M,E.,Ph.D.,** of Artificial Intelligence and Data Science for her encouragement.

Personally we thank **Dr.A.JOSHI M.E.,Ph.D**, Department of Artificial Intelligence and Data Science for the persistent motivation and support for this project, who at all times was the mentor of germination of the project from a small idea.

We express our thanks to the project coordinator **Ms.K.CHARULATHA M.E.,** Assistant Professor in Department of Artificial Intelligence and Data Science for their Valuable suggestions from time to time at every stage of our project.

Finally, we would like to take this opportunity to thank our family members, friends, and well-wishers who have helped us for the successful completion of our project.

We also take the opportunity to thank all faculty and non-teaching staff members in our department for their timely guidance in completing our project.

JOTHIMALAR M **(211422243132)**

AARTHI S **(211422243004)**

AKSHAYA S **(211422243020)**

ABSTRACT

Abstract. Customer churn prediction is a vital component of customer relationship management, enabling organizations to identify at-risk customers and implement strategies to enhance retention. This project explores the integration of blockchain technology with advanced machine learning algorithms to develop a robust churn prediction model. Blockchain offers enhanced data security, transparency, and integrity, ensuring that customer data is reliable and tamper-proof. In conjunction, machine learning algorithms—such as Logistic Regression, Random Forest, Support Vector Machines, and Neural Networks—analyze customer behavior patterns to forecast churn likelihood. The proposed system architecture includes modules for data collection, preprocessing, context-aware analysis, model training, and deployment, facilitating comprehensive churn analysis. This integration not only improves predictive accuracy but also instills trust through transparent data handling. Challenges such as scalability, regulatory compliance, and technical complexity are acknowledged, emphasizing the need for ongoing innovation. Overall, this project aims to provide a strategic framework for businesses to mitigate churn effectively, fostering increased customer loyalty and reduced operational costs.

Keywords: Root cause analysis, actionable insights, blockchain, machine learning, customer retention, churn prediction, customer churn, transparency, and product enhancement.

Using: RFM, Logistic regression, Decision Tree, NN

LIST OF FIGURES

LIST OF ABBREVIATIONS

ABBREVIATIONS	MEANING
AI	ARTIFICIAL INTELLIGENCE
API	APPLICATION PROGRAMMING INTERFACE
BI	BUSINESS INTELLIGENCE
CRM	CUSTOMER RELATIONSHIP MANAGEMENT
CSV	COMMA-SEPARATED VALUES
DNN	DEEP NEURAL NETWORK
IoT	INTERNET OF THINGS
KPI	KEY PERFORMANCE INDICATOR
ML	MACHINE LEARNING
NLP	NATURAL LANGUAGE PROCESSING
SVD	SINGULAR VALUE DECOMPOSITION
SLA	SERVICE LEVEL AGREEMENT
SQL	STRUCTURED QUERY LANGUAGE
TB	TELECOMMUNICATION BUSINESS
UAV	UNMANNED AERIAL VEHICLE
VPN	VIRTUAL PRIVATE NETWORK

XGBoost	EXTREME GRADIENT BOOSTING
QoS	QUALITY OF SERVICE
ROC	RECEIVER OPERATING CHARACTERISTICS
AUC	AREA UNDER THE CURVE
KNN	K-NEAREST NEIGHBORS

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

1.1 CHURN PREDICTION

A data-driven technique called churn prediction is used to determine which customers are most likely to discontinue using a good or service. Churn prediction algorithms identify early indicators of possible disengagement by examining previous customer contacts, behavioral patterns, and engagement levels. This enables companies to take proactive measures to resolve problems and enhance customer retention. In the end, this technique helps businesses reduce attrition and cultivate long-term loyalty by giving them information into churn-causing reasons including low product consumption, unmet needs, or dissatisfied service. It also helps them customize strategies to keep key consumers.

As companies strive to leverage data for competitive advantage, the integration of innovative technologies, such as blockchain, is also emerging. Blockchain offers enhanced security and transparency in data management, ensuring the integrity of customer information used for predictive analytics. Overall, customer churn prediction is not only a vital operational strategy but also an important component of enhancing customer experience and loyalty. By effectively predicting churn, organizations can take proactive measures to foster stronger relationships with their customers, ultimately driving revenue growth and sustainability in an increasingly dynamic marketplace.

1.2 BLOCKCHAIN TECHNOLOGY

The churn prediction system in this project uses blockchain technology to handle data in a transparent and safe manner. Blockchain technology provides a tamper-proof record that promotes accountability and trust by guaranteeing that all churn data, projections, and solution recommendations are reliably shared across networks. Product owners and teams can make well-informed decisions based on precise and easily accessible insights thanks to this extra layer of protection, which also secures sensitive customer data and permits transparent data exchange with stakeholders. The project's use of blockchain not only improves the accuracy of churn data but also fosters trust in the remediation and prediction processes, opening the door for long-term client retention tactics.

1.3 SCANNING AND SOLUTION SUGGESTION

In order to improve churn management by locating and resolving the root causes of customer disengagement, this project employs a scanning and solution suggestion system. When analyzing churned products or services, the scanning system looks for trends or flaws that might have led to customers leaving, like poor security, usability issues, or a lack of engagement features. Product owners can successfully enhance, redesign, or relaunch their offerings thanks to the solution suggestion module's automatic generation of practical suggestions based on this research. By turning churn insights into tangible solutions, this proactive strategy helps keep consumers and restores the value and marketability of churned items.

1.4 ARCHITECTURE DIAGRAM

A Churn Prediction and Recovery System intended to detect and hold onto at-risk clients is depicted in this architectural diagram. The first step involves gathering data from many sources, such as CRM systems, use analytics, transaction history, and consumer feedback. Preprocessing cleans and organizes the data after collection, preparing it for additional analysis. To make sure the dataset is correct and consistent, this stage eliminates unnecessary information and deals with missing values.

The processed data then moves on to the Model Selection and Training stage, where several machine learning methods (such Decision Trees and Logistic Regression) are assessed for their precision in churn prediction. After being chosen based on performance measures, the top model is trained to categorize clients. Either likely to keep or churn. As fresh information and user insights are incorporated into the system, this model is continuously improved, guaranteeing its correctness over time.

During the last phases, Prediction and Recovery, consumers who are at a high risk of churning are identified using the trained model. The Solution Recommendation component creates customized retention tactics for these clients, including exclusive deals or loyalty plans, to promote ongoing involvement. By combining information and suggestions into a dashboard for simple visualization, this layer enables companies to keep an eye on the wellbeing of their customers and take preventative measures to increase retention.

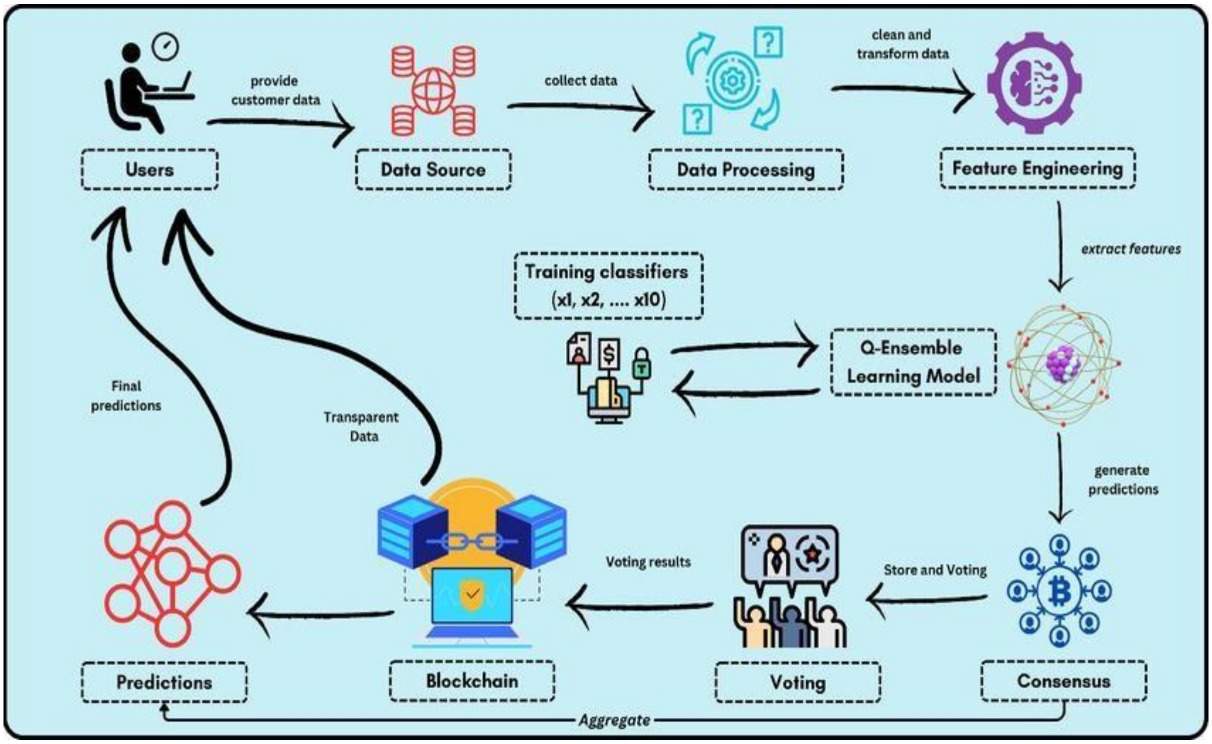

Fig 1.4: Architecture diagram of churn prediction

Transactions on a public blockchain are visible to all participants, promoting accountability and trust among users. Participants can independently verify transactions, ensuring a shared understanding of the system's state.

Blockchain utilizes a variety of cryptographic techniques to secure data. Each block contains a hash of the previous block, making it exceedingly difficult for unauthorized parties to alter the transaction history without detection.

1.5 REAL TIME APPLICATION OF CHURN PREDICTION

1. Telecommunications
2. Banking and Financial Services
3. Subscription-based services
4. Retail and E-commerce
5. Gaming Industry
6. Healthcare and Insurance
7. Hospitality and Travel
8. Online Education and E-learning platforms
9. Automotive Industry
10. Logistics and Delivery Services

1.6 CHALLENGES

Although churn prediction has many advantages, there are a number of issues that companies must resolve if they want precise and useful data. The following are some typical churn prediction problems:

1. ***Data Quality and Availability Challenge***: Accurate churn prediction requires high-quality data, but customer data may be imprecise, out-of-date, or incomplete. Furthermore, it's possible that pertinent data won't always be accessible for gathering.
2. ***Data Privacy and Security Challenge:*** Gathering, storing, and processing sensitive customer data might be difficult due to stringent adherence to privacy laws like the CCPA or GDPR.

3. ***Complexity of Customer Behavior Challenge:*** It might be challenging to pinpoint accurate churn indicators because customer behavior is impacted by a variety of intricate, perhaps invisible factors (such as individual preferences and external market conditions).

4. ***The Challenge of Model Interpretability:*** A lot of machine learning models, particularly intricate ones like neural networks, behave as "black boxes," making it challenging to understand why a consumer is expected to leave.

5. ***High Cost of Data Processing and Storage Challenge:*** Complex algorithms and managing massive data sets demand a lot of computer power, which can be expensive.

6. ***Changing Customer tastes and Market Trends Challenge:*** Models based on historical data may rapidly become out of date due to changing customer tastes and behaviors, which lowers forecast accuracy.

7. ***Feature Engineering and Selection Challenge:*** It might take a lot of time and effort to find and produce useful features from raw data, particularly for churn indicators that are not immediately apparent.

8. ***Unbalanced Data Challenge:*** Often, a small portion of the data is made up of churned consumers. Causing an imbalance of classes, which may lead to subpar model performance.

9. ***Scalability Challenge:*** It can be computationally demanding and resource-intensive to scale churn prediction systems to manage huge customer bases in real-time.

10. ***Integration with Business Processes Challenge:*** Data science and business teams must work together to convert churn prediction insights into workable business strategies.

11. ***Customer Retention Strategy Effectiveness Challenge:*** It can be challenging to implement successful retention tactics, even when turnover estimates are accurate.

Not all at-risk clients will respond well to interventions, and they may have unforeseen consequences.

12. ***Cost-Benefit Analysis Challenge:*** Businesses must be sure that the advantages of implementing retention tactics and churn prediction systems outweigh the costs.

The value gained from churn prediction programs can be significantly increased by successfully overcoming these obstacles, which call for a combination of technological know-how, strategic planning, and cautious resource allocation.

CHAPTER 2

LITERATURE REVIEW

By utilizing a decentralized blockchain framework, the system ensures secure, transparent, and tamper-proof storage of customer interaction data, which is crucial for effective analysis. This data is then processed using advanced machine learning algorithms, such as logistic regression, decision trees, and neural networks, to identify patterns and behaviors indicative of potential churn. The integration of blockchain not only mitigates data integrity concerns but also facilitates smart contracts that automate customer engagement strategies based on churn predictions. Overall, this innovative approach significantly improves the organization's ability to proactively manage customer relationships and reduce attrition rates, ultimately leading to enhanced customer loyalty and retention.

2.1 A Comparison of Machine Learning Algorithms for Customer Churn Prediction

Customer churn is a major problem in today's competitive environment for businesses in all sectors, and they are adopting machine learning (ML) to forecast and reduce revenue loss. ML algorithms Decision Tree, Random Forest, AdaBoost, and XGBoost are examined in this paper as the best methods for churn prediction in the banking, e-commerce, and telecom industries. Using these results in a churn prediction application is also covered.

Parth Pulkundwar, Krishna Rudani, Omkar Rane, Chintan Shah and Dr. Shyamal Virnodkar

2.2 Predicting Customer Churn Prediction In Telecom Sector Using Various Machine Learning Techniques

Since it is more expensive to acquire new consumers, telecom businesses must predict customer churn in order to keep their current subscribers. By examining consumer behavior, machine learning methods like as Logistic Regression, SVM, Random Forest, and Gradient Boosting are utilized to create classification models that aid in churn prediction. The purpose of this study is to compare different models in order to identify the best method for churn prediction. Abhishek Gaur, Ratnesh Kumar Dubey.

2.3 Research on a Customer Churn Combination Prediction Model Based on Decision Tree and Neural Network

For businesses, customer churn is a big problem because growth depends on customer retention. In order to increase prediction accuracy, this study creates a mixed prediction model utilizing decision trees and neural networks. The combined model performs better than individual models, providing more precise and perceptive churn predictions, according to the results. Xin Hu; Yanfei Yang; Lan hua Chen and Siru

2.4 Customer Churn Prediction Based on Interpretable Machine Learning Algorithms in Telecom Industry

For cost-effective tactics, it is essential to interpret feature importance in customer churn prediction. Using telecom data and Random Forest, Decision Tree, and Extra Trees models, this study finds that tenure, consumption, and recommendation are the most important churn indicators. Random Forest and Extra Trees produce the most dependable results, although other models function admirably. Liwen Ou

2.5 Customer Churn Prediction in the Telecom Sector

In the telecom sector, customer attrition is a major problem, therefore revenue preservation and retention depend on precise forecasting. The limits of several machine learning and deep learning approaches, such as Random Forest, SVM, and Logistic Regression, are examined in this work. A Hybrid Churn Prediction (HCP) model is suggested as a solution to these problems, combining many techniques for increased robustness and accuracy. According to experimental results on two telecom datasets, the HCP model routinely predicts customer attrition better than current techniques. Pallav Aggarwal; Vaidehi Vijayakumar

2.6 Evaluative study of cluster based customer churn prediction against conventional RFM based churn model

In the retail sector, where retention tactics are crucial to success, churn prediction is crucial. Conventional RFM analysis may take a while to produce useful insights, which could cause delays in reacting to shifts in consumer behavior. In comparison to real RFM features, this study assesses how well machine learning models forecast retail customer turnover based on k-means clusters. It also evaluates how reliable it is to combine the output of one predictive model with that of another for improved decision-making. Harish A. S., Malathy C.,

2.7 Predicting Customer Churn Prediction In Telecom Sector Using Various Machine Learning Techniques

In the telecom industry, customer churn analysis is essential since it helps forecast subscription cancellations by analyzing consumer behavior. Data mining techniques are crucial since keeping current consumers is more cost-effective than finding new ones. In order to forecast churn, this study evaluates the performance of several machine learning models, such as Logistic Regression, SVM, Random Forest, and Gradient Boosting. Abhishek Gaur, Ratnesh Kumar Dubey

2.8 Customer Churn Prediction Using Machine Learning Approaches

Since it directly affects revenue, customer turnover is a major concern for businesses, particularly in the telecom industry. In order to create a churn prediction model for telecom operators, this study looks at a number of machine learning methods. According to experimental data, Random Forest in conjunction with SMOTE-ENN produces the best outcomes, surpassing other methods and reaching a maximum Fl-score of 95%. Pallav Aggarwal; Vaidehi Vijayakumar

2.9 Efficacy of Customer Churn Prediction System

Businesses have a lot of difficulties because of customer attrition, which has led to research into machine learning (ML) applications for churn prediction in a variety of industries. Using criteria including precision, accuracy, recall, and F1 score, this study assesses seven popular machine learning algorithms: Logistic Regression, Decision Trees, Random Forests, Naive Bayes, KNN, AdaBoost, and XGBoost. By highlighting each model's advantages and disadvantages, the study helps organizations choose the best strategy for their requirements. It also explores how deep learning and machine learning are improving prediction accuracy and customization, providing insightful information for practitioners and future studies. Ritik Raj; Siddhant Gupta; Ritik Mishra; Medha Malik

2.10 Exploratory Data Analysis and Customer Churn Prediction for the Telecommunication Industry

Customer attrition costs the telecom sector a lot of money, therefore churn control techniques like focused advertising and promotions are essential. The asymmetric nature of telecom datasets presents a challenge for many machine learning algorithms; however, researchers used a variety of models, such as XGBoost, which achieved an accuracy of 82.80% on a native dataset. These outcomes show how well the model works to solve actual telecom data problems. Kiran Deep Singh; Prabh Deep Singh; Ankit Bansal; Gaganpreet Kaur; Vikas Khullar; Vikas Tripathi

2.11 An Improved Machine Learning Based Customer Churn Prediction for Insight and Recommendation in E-commerce

Keeping current customers is more cost-effective than finding new ones, lowering customer churn is essential in the cutthroat B2C e-commerce industry. The five components of the customer churn forecasting system presented in this paper—exploratory data analysis, data preparation, model tuning, model comparison, and insights—are used to evaluate transaction data and forecast attrition using artificial intelligence. The framework contributed to the research on churn prediction techniques by achieving high accuracy, with Cat Boost outperforming the others at 100% accuracy and F1-score. It also featured recursive feature elimination for feature ranking to improve comprehension and suggestions.

2.12 Customer Churn Prediction: Leveraging Data Analysis and Machine Learning Approaches

Because of the competitive environment and the desire to keep clients, predicting customer attrition is becoming more and more important in the banking sector. Using data from the banking industry, this study applies a variety of machine learning algorithms, highlighting the need of examining consumer behavior to improve retention initiatives. The findings show that the Light GBM and Random Forest algorithms are the most effective in forecasting client attrition, enabling banks to put focused customer loyalty initiatives into place.

2.13 Telecommunication Customers Churn Prediction using Machine Learning

Since keeping current customers is more cost-effective than finding new ones, customer churn presents a serious problem for telecom firms. This study assesses many machine learning models for churn prediction based on usage patterns, including decision trees, SVM, KNN, random forest, and linear regression. With an accuracy of 95.5% on the dataset, the results show that the random forest model fared better than the others. Nur Idora Abdul Razak; Muhammad Hazim Wahid

2.14 Customer Churn Prediction Using Machine Learning: Commercial Bank of Ethiopia

Business growth depends on identifying churned clients since it enables firms to comprehend the reasons behind churn and adjust their strategies accordingly. This study uses a dataset of 204,161 records and a variety of techniques, such as Logistic Regression, Random Forest, SVM, KNN, and Deep Neural Network (DNN), to create a machine learning model to forecast churn for the Commercial Bank of Ethiopia. With an accuracy of 79.32%, precision of 85.08%, and recall of 78.19%, the DNN produced the greatest results, proving its efficacy in forecasting customer attrition. Muhamed Hassen Seid; Michael Melese Woldeyohannis

CHAPTER 3

SYSTEM DESIGN

3.1 SYSTEM ARCHITECTURE:

A Churn Prediction and Recovery System intended to detect and hold onto at-risk clients is depicted in this architectural diagram. The first step involves gathering data from many sources, such as CRM systems, use analytics, transaction history, and consumer feedback. Preprocessing cleans and organizes the data after collection, preparing it for additional analysis. To make sure the dataset is correct and consistent, this stage eliminates unnecessary information and deals with missing values.

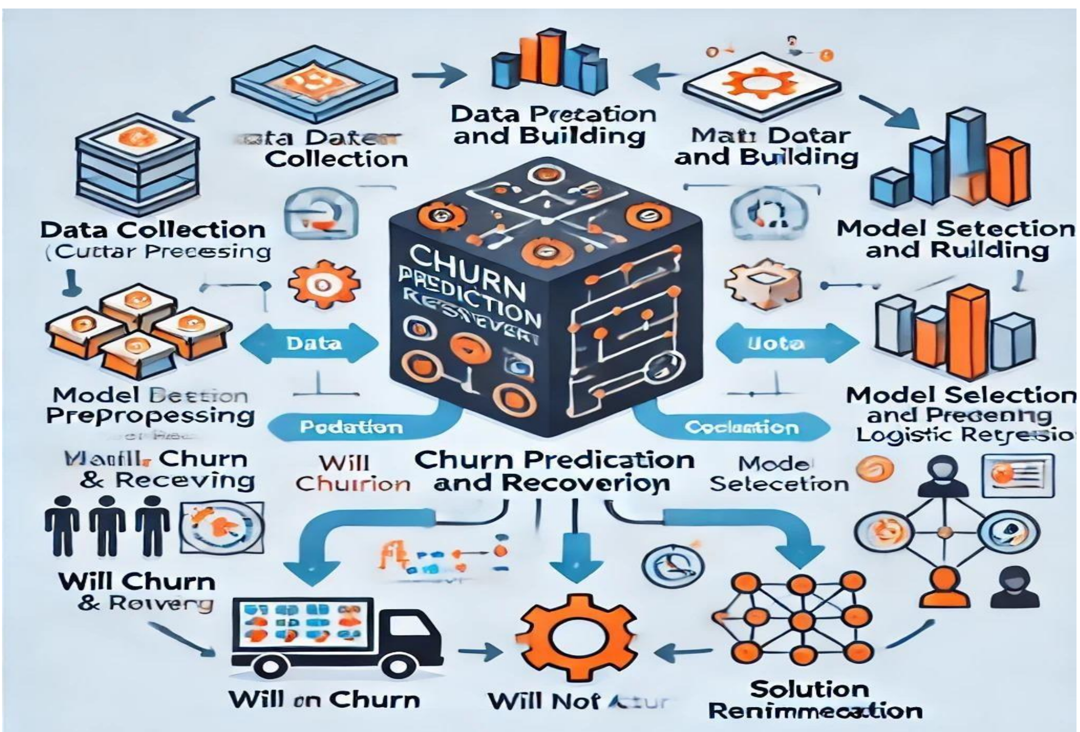

Fig 3.1: System architecture diagram for churn prediction

The processed data then moves on to the Model Selection and Training stage, where several machine learning methods (such Decision Trees and Logistic Regression) are assessed for their precision in churn prediction. After being chosen based on performance measures, the top model is trained to categorize clients. Either likely to keep or churn. As fresh information and user insights are incorporated into the system, this model is continuously improved, guaranteeing its correctness over time.

During the last phases, Prediction and Recovery, consumers who are at a high risk of churning are identified using the trained model. The Solution Recommendation component creates customized retention tactics for these clients, including exclusive deals or loyalty plans, to promote ongoing involvement. By combining information and suggestions into a dashboard for simple visualization, this layer enables companies to keep an eye on the wellbeing of their customers and take preventative measures to increase retention.

3.2 DATA FLOW DIAGRAM

We can divide the entire data flow into four separate segments, each of which will concentrate on a different facet of your project, in order to develop a more modular approach. Here's one way to organize it:

3.2.1 DFD-1

Your churn prediction project's fundamental framework is established by the Data Collection and Pre-Processing Flow. The first step is customer data collection, which involves gathering raw data from a variety of sources, including customer contacts, transactions, and feedback. Web forms, APIs, or customer relationship management (CRM) systems may be used to get this data. After being gathered, the data moves on to the Data Pre-Processing phase, where it is cleaned and transformed to guarantee its quality and usability. This entails dealing with missing values, standardizing data formats, and eliminating duplicates, finally producing a ready-to-use, wellstructured dataset, for further examination.

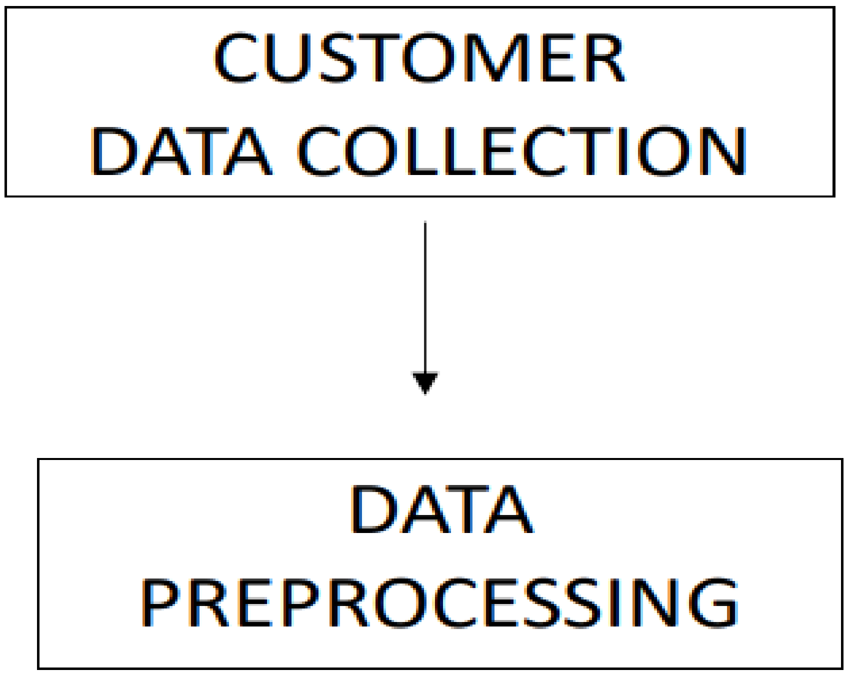

Fig 3.2.1: DFD-1 diagram

3.2.2 DFD-2

Analyzing the processed data to accurately forecast customer attrition is the focus of the following section, the Churn Prediction Flow. To find patterns in customer behavior that point to possible churn risks, the Churn Prediction Model uses machine learning methods including logistic regression, random forest, and neural networks. Each customer's anticipated churn probability is produced using this model. The flow also incorporates the Churn Reasons Identification phase, which examines these forecasts to identify certain churn-causing factors. Product owners can concentrate their efforts on resolving the root causes of disengagement by using this focused study.

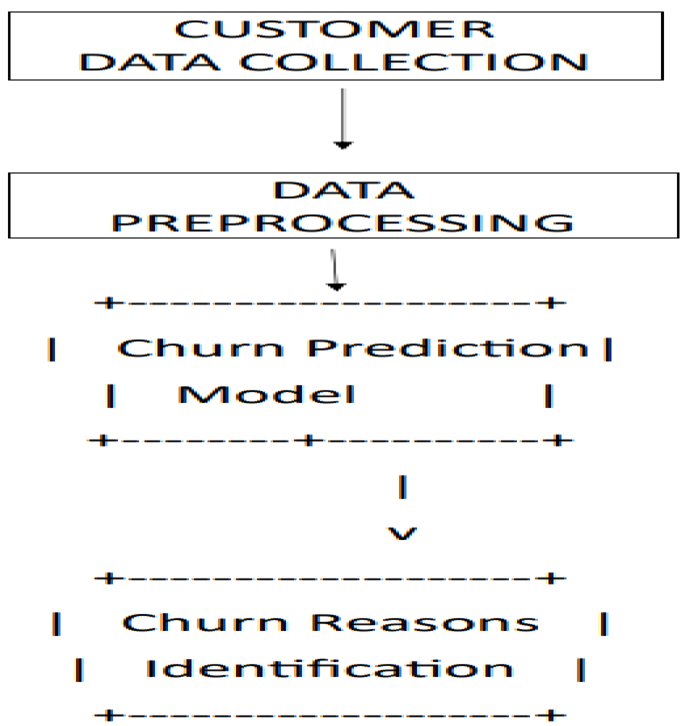

Fig 3.2.2: DFD-2 diagram

3.2.3 DFD -3

The Solution Suggestion Flow then uses the determined churn reasons to produce practical suggestions for product owners. The Solution Suggestion Module uses industry best practices and insights from prior successful interventions to create customized solutions. This leads to a list of specific steps that product owners can take to successfully reduce churn. The produced recommendations are kept in a Suggestions Repository for convenient access and long-term monitoring of their efficacy. As they strive to improve their product offers and customer retention tactics, product owners can benefit greatly from this repository.

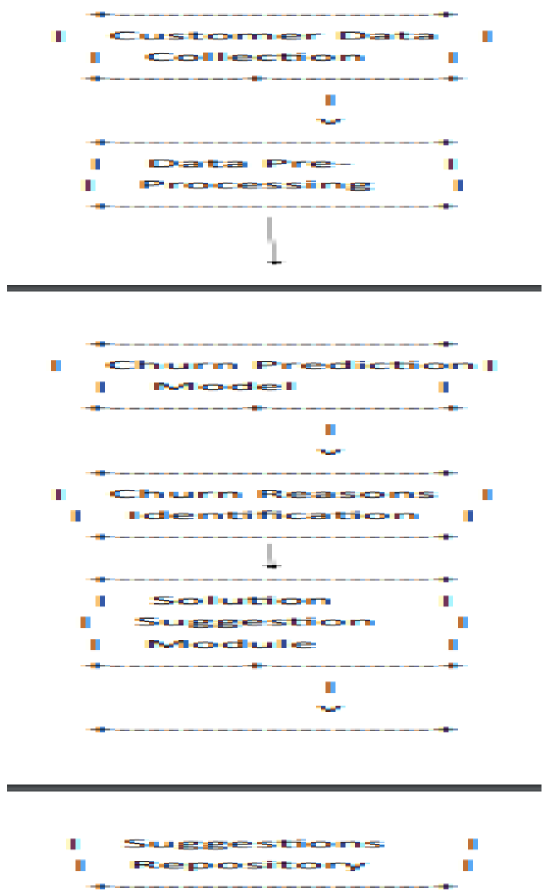

Fig 3.2.3: DFD-3 diagram

3.2.4 DFD -4

Finally, safe communication and user interaction with the churn prediction system are the main objectives of the Blockchain and User Interaction Flow. By safely and openly preserving all churn-related data and recommended fixes, the Blockchain Ledger plays a crucial part in building stakeholder trust. Product owners can evaluate churn estimates, investigate potential solutions, and access detailed data produced by the Reporting Module thanks to an easy-to-use user interface. Furthermore, the User Feedback & Monitoring component collects user feedback on how well the solutions were working, establishing an ongoing feedback loop that helps to improve the churn prediction model and the user experience in general. This methodical strategy, which is broken down into separate but related activities, guarantees a thorough system for anticipating and resolving client attrition while optimizing value of the product and client pleasure.

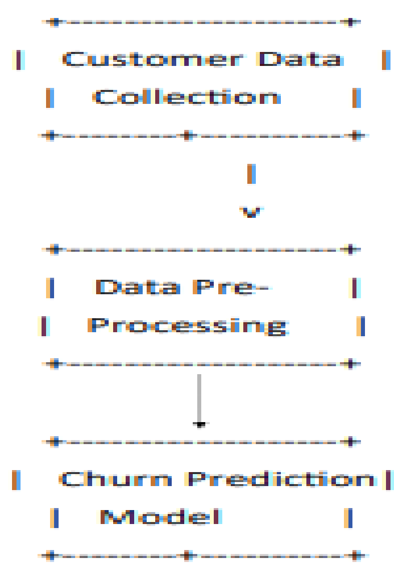

Fig 3.2.3: DFD-4 diagram

CHAPTER 4

PROJECT MODULES

4.1 MODULES

The project consists of Four modules. They are as follows,

1. Data Collection and Preprocessing
2. Model Architecture
3. Model Training and Validation
4. Prediction

4.2 DATA COLLECTION AND PREPROCESSING

A crucial phase in the churn prediction project is data collection, which aims to compile thorough information from multiple sources to produce a solid dataset that accurately represents consumer interactions and behaviors. This entails gathering account details including subscription kinds and payment histories in addition to client data like age, gender, and region. Furthermore, behavioral data—which includes transaction history, engagement indicators, and usage patterns—offers insights into how users engage with the product or service. External data sources, such as consumer input and market trends, are also used to enhance the dataset. Database extraction from current systems, APIs for real-time insights, surveys for direct client feedback, and web scraping for pertinent online information are some ways to gather data. Using these several approaches, the project guarantees a well-rounded data that will facilitate efficient churn analysis and forecasting.

4.2.1 Preprocessing

In order to convert raw data into a format that can be used for analysis and modeling, data preparation is essential. To guarantee data integrity and quality, a number of procedures must be followed. The main subtopics related to data preparation are listed below:

1. ***Cleaning of Data:***

Eliminating Duplicates: To prevent bias in analysis, find and remove duplicate entries from the dataset. Python's Pandas package contains methods like drop_duplicates() that can be used to accomplish this.

Managing Missing Values: Model performance may suffer from missing data. Some strategies are:

Removal: If a tiny percentage of the dataset consists of records with missing values, they should be eliminated.

Imputation: Using techniques like the mean, median, or mode for numerical data, or the most frequent value for categorical data, to fill in missing values.

Making an indicator variable to show whether a value was missing is known as flagging.

Finding and addressing outliers that could distort results is known as outlier detection. Among the methods are:

Standardizing values and determining which ones have Z-scores above or below a predetermined threshold (usually ±3) is known as Z-score.

Finding points that are outside of the permissible range (for example, below Q1

- 1.5IQR or above Q3 + 1.5IQR) involves calculating the interquartile range (IQR).

2. Transformation of Data:

Scaling numerical features to make sure they contribute evenly to the model is known as normalization or standardization. Techniques consist of:

Min-Max Rescaling the data to a range of 0 to 1 is known as scaling.

Standardizing data to have a mean of 0 and a standard deviation of 1 is known as Z-score standardization.

Encoding Categorical Variables: To ensure model compatibility, encode categorical data in a numerical format. Among the methods are:

Label Encoding: Giving every category a distinct integer.

One-Hot Encoding: To prevent the introduction of ordinal associations, binary variables are created for every category.

3. Engineering Features:

Developing New Features: Adding new features that could offer more insightful information. Among the examples are:

Tenure: Deducting the account creation date from the current date to determine how long a customer's subscription will last.

Features that show possible churn concerns, such the volume of customer support calls, are known as churn indicators.

Aggregating features is the process of compiling data to produce higher-level insights, such as figuring out the average monthly cost or the overall consumption across predetermined time periods.

5. ***Data Splitting Training and Testing Sets:***

To assess model performance, separate the dataset into training and testing subsets. A typical allocation is 30% for testing and 70% for training.

Stratified Sampling: To prevent bias in model evaluation, make sure the training and testing sets preserve a comparable distribution of the target variable (such as churned vs. non-churned consumers).

5. Verification of Data:

Verifying Data Types: confirming that the data type of each column matches the expected type (numerical values as floats, dates as datetime objects, etc.).

Checking Imputed Values: evaluating the effects of imputation techniques to make sure the underlying distribution is not distorted or biased.

6. Preprocessing Steps for Documentation Logs:

Keep thorough records of all preprocessing operations, including feature engineering choices, data sources, cleaning techniques, and transformations. Reproducibility and comprehension of the data preparation procedure depend on this transparency.

4.3 MODEL ARCHITECTURE

The structure and design of the machine learning models used to forecast customer attrition are referred to as model architecture. This includes the choice of algorithms, how they are set up, and how the data moves through the model as a whole. The main subtopics related to model architecture are listed below:

4.3.1 Algorithm Selection

Selecting Algorithms: Choose appropriate machine learning algorithms according to the task and the type of data. Typical churn prediction algorithms are as follows:

Based on input features, logistic regression is a statistical technique for binary classification that forecasts the likelihood of a consumer leaving.

Decision Trees: An interpretable model that divides data into branches for decision-making based on feature values.

Random Forests: An ensemble technique that reduces overfitting and increases accuracy by constructing many decision trees and combining their output.

SVMs, or support vector machines: a method of categorization that determines the hyperplane with the largest margin between classes.

Neural Networks: Specifically, deep learning models that use several layers of neurons to identify intricate patterns in big datasets.

4.3.2 Preparing Data for Model Input at the Feature Input Layer:

Get the dataset ready to be used as the model's input. This entails making certain that categorical variables are encoded, features are scaled appropriately, and any missing values are taken care of.

Configure the input layer's shape by taking into account the quantity of features in the dataset. For instance, the input layer would have 10 features if there were 10nodes.

4.3.3 Neurons and Hidden Layers in Neural Networks Configuration of the Layer

Find out how many hidden levels there are and how many neurons there are in each layer. The model's capacity to pick up intricate patterns is greatly impacted by this architecture.

Activation Functions: Select suitable activation functions for the output layer, such as sigmoid or softmax, and ReLU (Rectified Linear Unit) for hidden layers.

Methods of Regularization: Use strategies like L2 regularization or dropout to avoid overfitting, particularly in deeper neural networks.

4.3.4 Layer of Output

Output Shape: Depending on the classification strategy, the output layer for churn prediction usually consists of one node (for binary classification) or several nodes (for multi-class problems).

Use an appropriate activation function for the output layer, such as a sigmoid for binary data. categorization, which will produce odds that show how likely churn is.

4.3.5 Model Assessment and Adjustment

Method of Training: Describe the training process, including optimization methods (like Adam and SGD), evaluation measures (like accuracy, precision, recall, and F1 score), and loss functions (like binary cross-entropy for binary classification).

Cross-Checking: Use methods like as k-fold cross-validation to reduce overfitting and make sure the model performs effectively when applied to new data.

Tuning Hyperparameters: To maximize hyperparameters like learning rate, the number of trees in a random forest, or neural network architecture configurations, use strategies like grid search or random search.

4.4 MODEL TRAINING AND VALIDATION

Building a successful churn prediction model requires both model validation and training. The model runs through several epochs throughout the training phase, with each epoch denoting a full run through the training dataset. In order to reduce the loss, the model adjusts its weights using optimization methods after calculating the loss in each epoch by comparing its predictions to the actual results. The model's performance gradually improves as a result of this iterative process, which enables it to discover patterns in the data. A validation set is used to make sure the model performs well when applied to unknown data. Performance parameters including as accuracy, precision, recall, and F1 score are used to assess the model on this validation set at the end of each session. Techniques for early halting can be used to avoid overfitting by stopping training after a predetermined number of epochs if validation loss does not improve.

Several crucial procedures are usually included in the training steps: The dataset is first separated into sets for training and validation. The model architecture is then established, encompassing the selection of layers, activation functions, and algorithms. The model is then assembled using the proper optimizer and loss function. The model is trained on the training set and validated on the validation set at the end of each epoch as the training process progresses through several epochs. Hyperparameter tuning may be done to further optimize the model's performance, and performance indicators are monitored to track advancements. Lastly, the model's overall efficacy in forecasting customer attrition can be assessed by testing it on a different test set after training is finished.

Steps in Training

i. Data Preparation: Create training and validation sets from the dataset.
Define the architecture of the model, including its layers, nodes, and activation functions.

ii. Compilation: Choose an optimizer and loss function to compile the model.

iii. Training: Modify weights according to loss as you train the model over a predetermined period of epochs.

iv. Validation: Following each epoch, assess the model's performance on the validation set.

v. Monitoring: Keep an eye on performance indicators and, if necessary, use early stopping.

vi. The hyperparameter Tuning: Use methods like grid search or random search to optimize the model's parameters.

vii. Testing: To gauge generalization, evaluate the finished model's performance on a different test set.

4.5 PREDICTION

4.5.1 LOGISTIC REGRESSION

Logistic regression is an effective approach for predicting the likelihood that a client will stop using a product or service in the context of your churn prediction project. Using a logistic function that yields values between 0 and 1, logistic regression calculates the likelihood of churn by examining past customer data, including demographics, usage trends, and engagement levels. By using the log odds to express this relationship, the model lets you see how different attributes affect churn risk. Maximum likelihood estimation is used to estimate parameters, guaranteeing that the model accurately depicts the underlying patterns in the data. By comprehending the primary causes of client attrition and using data to inform their decision-making, firms may proactively handle possible churn and improve their customer retention tactics.

```
Model Accuracy: 0.86
Confusion Matrix:
 [[10  1]
 [ 1  2]]
Classification Report:
              precision    recall  f1-score   support

           0       0.91      0.91      0.91        11
           1       0.67      0.67      0.67         3

    accuracy                           0.86        14
   macro avg       0.79      0.79      0.79        14
weighted avg       0.86      0.86      0.86        14

Churn prediction data recorded on blockchain.
Automated Solution Suggestions:
- High support tickets detected. Suggest improving customer support.
- High support tickets detected. Suggest improving customer support.
```

Fig 4.5.1:Logistic Regression

4.5.2 RANDOM FOREST

By combining the output of several decision trees, the random forest algorithm is used as an ensemble learning technique in your churn prediction project to improve the precision and resilience of churn predictions. The model can identify a variety of patterns and interactions in the consumer behavior data since each decision tree is constructed using a random subset of the training data and characteristics. By using majority voting to combine the predictions from each individual tree for predicting churn, the random forest lowers the possibility of overfitting and enhances overall model performance. In addition to offering a more accurate estimate of the likelihood of churn, this method emphasizes the significance of certain features, enabling you to pinpoint the main causes of customer attrition. As a result, the random forest model helps companies.in making well-informed choices on interventions and enhancements to successfully retain clients.

```
Accuracy: 0.9047619047619048
Classification Report:
              precision    recall  f1-score   support

           0       0.93      0.93      0.93        15
           1       0.83      0.83      0.83         6

    accuracy                           0.90        21
   macro avg       0.88      0.88      0.88        21
weighted avg       0.90      0.90      0.90        21

Prediction for new customer: No Churn
Suggested Solution: Assist customer with payment options or address billing issues.

C:\ProgramData\anaconda3\Lib\site-packages\sklearn\base.py:464: UserWarning: X does not have valid feature names, but StandardS
caler was fitted with feature names
  warnings.warn(
```

Fig 4.5.2: Random Forest

4.5.3 DECISION TREE

Decision trees are used in your churn prediction project as a simple yet powerful way to categorize clients according to their propensity to leave. By recursively dividing the data into subsets according to the input feature values, the decision tree algorithm produces a structure resembling a tree, with each internal node denoting a feature decision, each branch denoting an outcome of that decision, and each leaf node denoting a final prediction (such as churn or no churn). Until the tree meets a stopping criterion—such as a minimum number of samples per leaf or a maximum depth—this process keeps going. Decision trees' clear design makes it simple to understand how various customer attributes affect churn risk, giving businesses the ability to see crucial elements and put specific tactics into place to improve client retention and deal with certain problems that cause turnover.

```
Accuracy: 0.9047619047619048
Classification Report:
              precision    recall  f1-score   support

           0       0.93      0.93      0.93        15
           1       0.83      0.83      0.83         6

    accuracy                           0.90        21
   macro avg       0.88      0.88      0.88        21
weighted avg       0.90      0.90      0.90        21

Prediction for new customer: No Churn
Suggested Solution: Assist customer with payment options or address billing issues.
C:\ProgramData\anaconda3\Lib\site-packages\sklearn\base.py:464: UserWarning: X does not have valid feature names, but StandardS
caler was fitted with feature names
  warnings.warn(
```

Fig 4.5.3: Random Forest

4.5.4 NEURAL NETWORK

Neural networks are used in your churn prediction project to represent intricate links in consumer behavior data that conventional algorithms can find difficult to identify. An linked layer of nodes, or neurons, makes up a neural network. Each neuron applies an activation function, processes inputs through weighted connections, and then sends the output to the layer above. Neural networks can capture non-linear interactions among several parameters that lead to customer churn by employing numerous layers to develop hierarchical patterns and representations. In order to maximize its capacity to forecast churn probabilities, the model uses backpropagation to modify its weights during training in response to the prediction error. This feature gives companies a strong tool for identifying at-risk clients and enables them to unearth complex insights on customer retention dynamics. Creating specialized interventions to increase client loyalty.

```
Epoch 1/50
4/4 ━━━━━━━━━━━━━━━━━━━━ 5s 198ms/step - accuracy: 0.4356 - loss: 0.6870 - val_accuracy: 0.8000 - val_loss: 0.5422
Epoch 2/50
4/4 ━━━━━━━━━━━━━━━━━━━━ 0s 35ms/step - accuracy: 0.4256 - loss: 0.6462 - val_accuracy: 0.8000 - val_loss: 0.4772
Epoch 3/50
4/4 ━━━━━━━━━━━━━━━━━━━━ 0s 30ms/step - accuracy: 0.5089 - loss: 0.5493 - val_accuracy: 0.8000 - val_loss: 0.4231
Epoch 4/50
4/4 ━━━━━━━━━━━━━━━━━━━━ 0s 31ms/step - accuracy: 0.5056 - loss: 0.5430 - val_accuracy: 0.8000 - val_loss: 0.3789
Epoch 5/50
4/4 ━━━━━━━━━━━━━━━━━━━━ 0s 32ms/step - accuracy: 0.6044 - loss: 0.4884 - val_accuracy: 0.9000 - val_loss: 0.3442
Epoch 6/50
4/4 ━━━━━━━━━━━━━━━━━━━━ 0s 33ms/step - accuracy: 0.8778 - loss: 0.4474 - val_accuracy: 0.9000 - val_loss: 0.3190
Epoch 7/50
4/4 ━━━━━━━━━━━━━━━━━━━━ 0s 30ms/step - accuracy: 0.9089 - loss: 0.4292 - val_accuracy: 0.9000 - val_loss: 0.2985
Epoch 8/50
4/4 ━━━━━━━━━━━━━━━━━━━━ 0s 32ms/step - accuracy: 0.9256 - loss: 0.4412 - val_accuracy: 0.9000 - val_loss: 0.2809
Epoch 9/50
4/4 ━━━━━━━━━━━━━━━━━━━━ 0s 34ms/step - accuracy: 0.8956 - loss: 0.4479 - val_accuracy: 0.9000 - val_loss: 0.2672
Epoch 10/50
4/4 ━━━━━━━━━━━━━━━━━━━━ 0s 25ms/step - accuracy: 0.8589 - loss: 0.4413 - val_accuracy: 0.9000 - val_loss: 0.2559
```

```
Epoch 11/50
4/4 ━━━━━━━━━━━━━━━━━━━━ 0s 21ms/step - accuracy: 0.8889 - loss: 0.4077 - val_accuracy: 0.9000 - val_loss: 0.2462
Epoch 12/50
4/4 ━━━━━━━━━━━━━━━━━━━━ 0s 21ms/step - accuracy: 0.9222 - loss: 0.3984 - val_accuracy: 0.9000 - val_loss: 0.2385
Epoch 13/50
4/4 ━━━━━━━━━━━━━━━━━━━━ 0s 20ms/step - accuracy: 0.8589 - loss: 0.4147 - val_accuracy: 0.9000 - val_loss: 0.2317
Epoch 14/50
4/4 ━━━━━━━━━━━━━━━━━━━━ 0s 23ms/step - accuracy: 0.9156 - loss: 0.3952 - val_accuracy: 0.9000 - val_loss: 0.2228
Epoch 15/50
4/4 ━━━━━━━━━━━━━━━━━━━━ 0s 23ms/step - accuracy: 0.8956 - loss: 0.3909 - val_accuracy: 0.9000 - val_loss: 0.2140
Epoch 16/50
4/4 ━━━━━━━━━━━━━━━━━━━━ 0s 23ms/step - accuracy: 0.9089 - loss: 0.3715 - val_accuracy: 0.9000 - val_loss: 0.2061
Epoch 17/50
4/4 ━━━━━━━━━━━━━━━━━━━━ 0s 22ms/step - accuracy: 0.8489 - loss: 0.4148 - val_accuracy: 0.9000 - val_loss: 0.1995
Epoch 18/50
4/4 ━━━━━━━━━━━━━━━━━━━━ 0s 20ms/step - accuracy: 0.8889 - loss: 0.3445 - val_accuracy: 0.9000 - val_loss: 0.1930
Epoch 19/50
4/4 ━━━━━━━━━━━━━━━━━━━━ 0s 21ms/step - accuracy: 0.8789 - loss: 0.3822 - val_accuracy: 0.9000 - val_loss: 0.1874
Epoch 20/50
4/4 ━━━━━━━━━━━━━━━━━━━━ 0s 26ms/step - accuracy: 0.8589 - loss: 0.3844 - val_accuracy: 0.9000 - val_loss: 0.1829
```

```
Accuracy: 0.9047619047619048
Classification Report:
              precision    recall  f1-score   support

           0       0.93      0.93      0.93        15
           1       0.83      0.83      0.83         6

    accuracy                           0.90        21
   macro avg       0.88      0.88      0.88        21
weighted avg       0.90      0.90      0.90        21
```

```
1/1 ━━━━━━━━━━━━━━━━━━━━ 0s 123ms/step
Prediction for new customer: Churn
Suggested Solution: Assist customer with payment options or address billing issues.
```

Fig ***4.5.5*** : Neural Network

CHAPTER 5

SYSTEM REQUIREMENT

5.1 INTRODUCTION

The software requirements for your churn prediction project, along with explanations for each component

5.2 REQUIREMENTS

5.2.1 Software Requirements

- Windows 10 or Later / Linux (Ubuntu 20.04 or Later)
- Python (Version 3.6 or Higher)
- Pandas, NumPy
- Scikit-learn, TensorFlow or PyTorch
- Matplotlib, Seaborn
- Jupyter Notebook, PyCharm or Visual Studio Code (VSCode)
- Git

5.3 Technology Used

I. Machine Learning Algorithms:

- Logistic Regression
- Random Forest
- Decision Trees
- Neural Networks

II. Python

III. Data Manipulation and Analysis Libraries:

- Pandas
- NumPy

IV. Machine Learning Frameworks:
   - Scikit-learn
   - TensorFlow (or PyTorch)

V. Git

VI. Data Visualization Tools:
   - Matplotlib
   - Seaborn

VII. Integrated Development Environment (IDE):
   - Jupyter Notebook
   - PyCharm (or Visual Studio Code)

VIII. Deployment Technologies:
   - Flask (or Django)
   - Docker

5.3.1 Software description

5.3.1.1 Linux (Ubuntu 20.04 or Later) or Windows 10 or Later

The operating system you choose will rely on your level of experience and the software tools you intend to employ. Linux is favored in data science due to its stability, package management, and support for open-source tools, but Windows is popular and easy to use.

5.3.1.2 Python (Version 3.6 or Higher)

Because of its community support, large library, and ease of reading, Python is the main programming language used in data science and machine learning. It offers a strong framework for managing data processing operations and putting machine learning algorithms into practice.

Python is considered platform-independent primarily due to its design and the way it is executed,

1. ***Interpreted Language:*** Python is an interpreted language, which means that it is executed line-by-line by the Python interpreter. Unlike compiled languages like C or C++, where source code is compiled into machine code for a specific platform, Python code is executed directly by the Python interpreter. This interpreter is available for various platforms, allowing Python code to run on different operating systems without modification.

2. ***Abstraction of Low-Level Details:*** Python abstracts many low-level details of the underlying hardware and operating system. It provides a high-level interface to common operations like file I/O, networking, and process management. Python's standard library contains platform-independent modules that handle these operations, ensuring consistent behaviour across different platforms.

5.3.1.3 Pandas

This package is necessary for analyzing and manipulating data. It offers data structures like DataFrames that facilitate time series data management, data cleaning, and structured data handling and analysis.

Large multi-dimensional arrays and matrices are supported by NumPy, a core Python library for numerical computations that also offers a number of mathematical functions for working with these data structures.

A popular library for putting machine learning algorithms into practice, such as logistic regression, decision trees, random forests, and model evaluation methods, is called Scikit-learn. It offers user-friendly tools for data preprocessing, model training, and prediction.

Two frameworks for creating and training neural networks are TensorFlow and PyTorch. While PyTorch is preferred for its dynamic computation graph and user-friendliness, especially in research contexts, TensorFlow is renowned for its scalability and production readiness. Depending on the requirements of your particular project, you can select one.

5.3.1.4 Matplotlib

A fundamental Python charting toolkit that enables the creation of interactive, animated, and static displays. Plotting data distributions, trends, and model performance is one of its uses.

Seaborn: Seaborn, which is based on Matplotlib, offers a high-level interface for creating visually appealing statistical graphics, simplifying the visualization of intricate data distributions and relationships.

5.3.1.5 Jupyter Notebook

An interactive online workspace that lets you create documents with rich text, code execution, and visuals. It's very helpful for communicating results in an understandable style and for data exploration.

Visual Studio Code (VSCode) or PyCharm: These IDEs are appropriate for larger projects and more structured development workflows since they provide robust Python development capabilities including code completion, debugging, and integrated terminal support.

5.3.1.6 Git

A version control system that facilitates teamwork and allows you to monitor changes made to your codebase. It is crucial for maintaining code organization and recoverability as well as for managing project versions

5.3.2 Technology used

5.3.2.1 Algorithms for Machine Learning Utilized by Technologies

Based on past data, logistic regression is a statistical technique for binary categorization that aids in forecasting the probability of client attrition. It simulates how input properties and the likelihood of an event (churn) happening are related.

Random Forest: To increase accuracy and robustness, this ensemble learning method constructs several decision trees and averages their predictions. Random forest lowers the chance of overfitting and is good at identifying intricate patterns in consumer data.

Decision trees are tree-based models that divide data according to feature values in a recursive manner in order to generate predictions. Decision trees are helpful for determining the main elements affecting churn and offer clear visual representations.

Neural Networks: Neural networks are deep learning models that can identify non-linear relationships in data. They are made up of several layers of interconnected nodes. They are especially helpful for managing complicated patterns and huge datasets that conventional algorithms could miss.

5.3.2.2 Language of Programming

Python is the main programming language used for implementing models, data analysis, and machine learning. Python can handle many parts of your project, from data pretreatment to model evaluation, thanks to its vast libraries and frameworks.

5.3.2.3 Libraries for Data Manipulation and Analysis

Pandas: An effective data manipulation and analysis library that lets you manage time series data, clean data, and work with structured data.

NumPy: Facilitates mathematical processes necessary for data processing by supporting numerical computations, especially with big arrays and matrices.

5.3.2.4 Frameworks for Machine Learning

Scikit-learn: A popular toolkit for putting machine learning methods like logistic regression, decision trees, and random forests into practice, as well as tools for validating and assessing models.

TensorFlow and PyTorch are neural network building and training frameworks that let you use sophisticated deep learning methods to increase the accuracy of churn predictions.

5.3.2.5 Tools for Data Visualization

Matplotlib is a basic charting package that helps you comprehend data distributions and model performance by enabling you to generate a variety of static and interactive representations.

Seaborn: Based on Matplotlib, Seaborn makes it easy to create visually appealing and educational statistical graphics, which facilitates the visualization of intricate relationships within data.

5.3.2.6 Environment for Integrated Development (IDE)

For exploratory data analysis and result presentation, Jupyter Notebook is an interactive web-based environment that combines code execution with rich text and graphics.

Visual Studio Code/PyCharm: With features like code completion, debugging, and integrated terminal support, these IDEs facilitate organized Python development.

5.3.2.7 Version Management

Git: A version control system that makes it possible to trace modifications made to your codebase, promoting teamwork and guaranteeing code organization and recovery.

5.3.2.8 Technologies of Deployment

Flask/Django: Frameworks such as Flask or Django can be used to develop an intuitive user interface for engaging with the churn prediction model if you intend to publish your model as a web application.

Docker: You can use Docker to bundle your application and its dependencies for containerization and to guarantee consistent environments between development and production.

CHAPTER 6

CONCLUSION & REMARK

1.1 CONCLUSION

To sum up, the churn prediction project uses cutting-edge machine learning methods to tackle the serious problem of customer attrition. The project efficiently analyzes consumer behavior data to forecast the chance of churn by utilizing methods including logistic regression, random forests, decision trees, and neural networks. Businesses can use this predictive capability to pinpoint at-risk clients and put focused measures in place to increase customer retention and enhance service offerings. Additionally, churn-related data can be communicated securely and transparently thanks to the incorporation of blockchain technology, which builds stakeholder trust. The project offers practical solutions for customer happiness and product enhancement in addition to insightful information about customer dynamics thanks to a strong framework for data manipulation, visualization, and model deployment. All things considered, this all-encompassing strategy puts companies in a position to Proactively responding to consumer requirements will eventually boost revenue and foster more client loyalty.

REFERENCES

[1]. Parth Pulkundwar, Krishna Rudani, Omkar Rane, Chintan Shah and Dr. Shyamal Virnodkar, - A Comparison of Machine Learning Algorithms for Customer Churn Prediction

[2]. Xin Hu; Yanfei Yang; Lanhua Chen and Siru Zhu - Research on a Customer Churn Combination Prediction Model Based on Decision Tree and Neural Network

[3]. Liwen Ou - Customer Churn Prediction Based on Interpretable Machine Learning Algorithms in Telecom Industry

[4]. Harish A. S., Malathy C., - Evaluative study of cluster based customer churn prediction against conventional RFM based churn model

[5]. Abhishek Gaur, Ratnesh Kumar Dubey - Predicting Customer Churn Prediction In Telecom Sector Using Various Machine Learning Techniques

[6]. Pallav Aggarwal; Vaidehi Vijayakumar - Customer Churn Prediction in the Telecom Sector

[7]. R. Srinivasan; D. Rajeswari; G. Elangovan - Customer Churn Prediction Using Machine Learning Approaches

[8]. Ritik Raj; Siddhant Gupta; Ritik Mishra; Medha Malik - Efficacy of Customer Churn Prediction System

[9]. Kiran Deep Singh; Prabh Deep Singh; Ankit Bansal; Gaganpreet Kaur; Vikas Khullar; Vikas Tripathi - Exploratory Data Analysis and Customer Churn Prediction for the Telecommunication Industry

[10]. Ishrat Jahan; Tahsina Farah Sanam - An Improved Machine Learning Based Customer Churn Prediction for Insight and Recommendation in E-commerce

[11]. Arsalan Mardi; Hananeh Ghorbani - Customer Churn Prediction: Leveraging Data Analysis and Machine Learning Approaches

[12]. Nur Idora Abdul Razak; Muhammad Hazim Wahid - Telecommunication Customers Churn Prediction using Machine Learning

[13]. Aryan Raj; D Vetrithangam - Machine Learning and Deep Learning technique used in Customer Churn Prediction: - A Review

[14]. Muhamed Hassen Seid; Michael Melese Woldeyohannis - Customer Churn Prediction Using Machine Learning: Commercial Bank of Ethiopia

[15]. Utku Yabas; Hakki Candan Cankaya; Turker Ince - Customer Churn Prediction for Telecom Services