Dynamic Churn Prediction Using Ann- Predict Your Customer Behavior

A. Amulya[1], Afifa Butool[2], Sedmaki Aishwarya[3]
[1]*Professor, Mahatma Gandhi Institute of Technology*
[2,3]*UG Student, Mahatma Gandhi Institute of Technology*

Abstract- **In today's competitive landscape, customers prioritize high-quality services, compelling organizations to focus on technologically advanced solutions to meet customer expectations. Amid this growing competition, effective customer relationship management (CRM) systems have become essential for organizations to attract new customers, strengthen existing relationships, and enhance customer retention. This, in turn, contributes to business profitability and sustainability. Leveraging machine learning models such as Artificial Neural Networks (ANN), Support Vector Machines and Random Forest Classifier techniques offers significant potential in understanding and predicting customer behavior. These advanced models enable organizations to identify churn risks, make data-driven decisions, and develop proactive strategies to improve customer satisfaction and loyalty. By integrating ANN, businesses can gain deeper insights into customer behavior, ensuring transparency and fostering trust, ultimately leading to improved retention and sustained growth.**

Keywords: **Churn analysis, Data Processing, Artificial Neural Networks, machine learning, Deep Learning, Support Vector Machines, Random Forest, KNN, XG Boost, SHAP, Explainable AI, CRM.**

I. INTRODUCTION

In the competitive business landscape, understanding customer behavior and addressing their needs has become a cornerstone for organizational success. Customers today demand high-quality services and personalized experiences, pushing organizations to innovate and adopt advanced technologies to stay ahead. As a result, Customer Relationship Management (CRM) systems have emerged as indispensable tools for managing customer interactions, improving retention rates, and driving business growth. However, traditional CRM systems often fall short in predicting customer churn is a critical factor impacting profitability and long-term sustainability. This project, titled Dynamic Churn Prediction Using Deep Learning - Predict Your Customer Through Customer Behavior, leverages state-of-the-art machine learning techniques to address this challenge. By integrating Artificial Neural Networks (ANN), the project aims to create a robust,transparent framework for identifying customers at risk of churn. ANN provides the computational power to analyze complex behavioral patterns, while XAI offers interpretability, enabling businesses to understand the underlying factors influencing customer decisions.

A. Problem Statement.

In an increasingly competitive market, retaining customers has become a significant challenge for businesses striving to maintain profitability and sustainability. Despite advancements in Customer Relationship Management (CRM) systems, many organizations struggle to effectively analyze customer behavior and identify churn risks. Traditional methods lack the precision and transparency required to make informed, proactive decisions. The absence of reliable, explainable models hinders businesses from understanding customer needs and improving loyalty. This project addresses the need for an innovative solution by utilizing Artificial Neural Networks (ANN), Support Vector Machines and Random Forest Classifier to predict customer churn, foster trust, and enhance retention strategies.

B. Existing System

Traditional models like logistic regression and decision trees are widely used for customer churn analysis due to their simplicity and interpretability. Logistic regression predicts binary outcomes by mapping probabilities using a logistic function, making it a common choice for churn prediction. Decision trees, on the other hand, employ a flowchart-like structure to model decisions based on specific features, with final predictions represented at the leaf nodes. While these models are effective for straightforward tasks, they face limitations when dealing with complex, non-linear, or large-scale customer data. Their inability to effectively capture intricate feature interactions reduces their predictive power in scenarios where understanding variable relationships is critical. Furthermore, traditional models often struggle with overfitting,

learning the noise and outliers in training data too well, which hampers their ability to generalize to unseen data. Random forests, an ensemble approach of multiple decision trees, mitigate some of these issues by improving accuracy and reducing overfitting through aggregated predictions. However, the inherent challenges of data complexity and limited feature interactions remain significant drawbacks for traditional models in accurately predicting customer churn.

II. PROPOSED SYSTEM

A. Architecture of Proposed System.

The proposed system focuses on developing a robust churn prediction model by integrating Artificial Neural Networks (ANN), Support Vector Machines and Random Forest Classifier. This system aims to analyze customer behavior patterns to accurately identify potential

churn risks. By leveraging ANN, the model achieves high prediction accuracy. Organizations can utilize this system to make informed, data-driven strategies, enabling proactive measures to enhance customer retention and satisfaction. The system is designed to foster trust through transparency, ultimately driving business growth and maintaining competitive advantage in a rapidly evolving market.

B. Advantages of Proposed System.

- *Enhanced Prediction Accuracy*
- *Interpretability and Transparency*
- *Data-Driven Decision Making*
- *Proactive Customer Retention*
- *Competitive Edge*

III. LITERATURE SURVEY

This research investigates the Supply chain evolution over diverse industries and forecasts a telecom churn using a publicly available dataset, our study offers a thorough technique for assessing and projecting customer attrition The procedure includes applying different machine learning models to predict customer turnover, meticulously pre-processing the data, and conducting exploratory data analysis (EDA)through informative statistics and graphics, Exploratory Data Analysis revealed important elements like rival offerings and the kind of internet service, delivering insightful information about what causes churn. The study then used various machine learning models, such as Decision Tree, Random Forest, K-Neighbours, and XG Boost classifiers. With an accuracy of 98.25 percent, Random Forest beat the Decision Tree model, which had 98.02 percent accuracy. These models are tested using accuracy, precision, recall, F1-score, and AUC-ROC. This approach emphasizes how important predictive analytics is to comprehend the dynamics of customer turnover and lays the groundwork for tactical interventions meant to improve customer happiness and retention in the telecom industry [1].

Customer churn prediction is a key challenge for businesses aiming to retain customers and optimize revenue. This paper presents a comparative analysis of several machine learning algorithms for predicting customer churn across various industries. We evaluate Logistic Regression (LR), Random Forest (RF), Support Vector Machines (SVM), Gradient Boosting Machines (GBM), and K-Nearest Neighbors (KNN) on a real-world customer dataset, focusing on performance metrics such as accuracy, precision, recall, F1-score, and AUC-ROC. We also explore various feature engineering techniques, including customer demographic data, transaction history, and behavioral data to improve model performance. Our results show that ensemble methods such as Random Forest and Gradient Boosting outperform other models in terms of overall accuracy and robustness against class imbalance. Moreover, we analyze the interpretability of the models, providing insights into the key features driving churn predictions [2].

Customer churn prediction is a critical issue in many service-oriented industries, where understanding the temporal behavior of customers is key to developing effective retention strategies. In this paper, we propose a Time Series-based Churn Prediction Model using Long Short-Term Memory (LSTM) networks, which can capture the sequential dependencies in customer behavior over time. Traditional churn prediction methods often overlook the importance of temporal patterns, leading to suboptimal performance in predicting churn. Our model leverages time-series data including customer transaction history, service usage patterns, and engagement metrics, to predict the likelihood of churn at different time points.The proposed LSTM model is evaluated against conventional machine learning algorithms, such as logistic regression, decision trees, and random forests, on real-world datasets. Experimental results show that the LSTM-based model significantly outperforms traditional methods, achieving higher accuracy, precision, recall, and F1-score in predicting customer churn [3].

Customer churn prediction is a critical challenge for e- commerce platforms aiming to retain users and maintain revenue growth. In this study, we propose a Hybrid Deep Learning Model that combines the strengths of Convolutional Neural Networks (CNN) **and** Long Short-Term Memory (LSTM) networks to effectively predict customer churn. The model leverages historical customer behavior, including purchase frequency, transaction amounts, session durations, and product categories, to identify patterns leading to churn. To enhance the predictive capability, the hybrid model uses CNN layers to extract spatial features from customer data and LSTM layers to capture sequential temporal dependencies. The proposed method is evaluated on real-world e- commerce datasets and compared with traditional machine learning models such as Logistic Regression, Random Forest, and Support Vector Machines (SVM). The results demonstrate that the hybrid deep learning model achieves superior performance, with significant improvements in accuracy, precision, recall, and F1- score [5].

Deep neural networks were also used in these efforts to extract features without taking into account the sequence information. In view of these issues, the current study provides an effective method for predicting customer churn based on a hybrid deep learning model termed BiLSTM-CNN. The goal is to effectively estimate customer churn using benchmark data and increase the churn prediction process's accuracy. The experimental results show that when trained, tested, and validated on the benchmark dataset, the proposed BiLSTM-CNN model attained a remarkable accuracy of 81% [6].

IV. SYSTEM DESIGN AND METHODOLOGY

A. Design

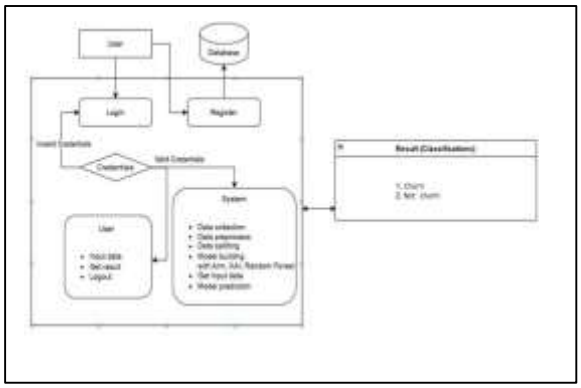

Figure1: System Design

The diagram represents a customer churn prediction system that involves user authentication, data processing, and machine learning-based classification. Users can either log in or register, with credentials verified against a database. If login credentials are invalid, access is denied; otherwise, users can input customer data, obtain churn predictions, and log out. The system handles multiple stages, including data collection, preprocessing, and splitting for training and testing. It employs machine learning models such as Artificial Neural Networks (ANN), Explainable AI (XAI), and Random Forest to analyse customer data and predict whether a customer is likely to churn or not. The final classification results—either "Churn" or "Not Churn"—are displayed to the user, facilitating decision-making for customer retention strategies.

B. Methodology

The system proposes to integrate advanced machine learning techniques that enhance Churn Prediction. By integrating Artificial Neural Networks (ANN) with Explainable AI (XAI) techniques, the system aims to provide actionable insights into customer churn risks, enabling businesses to make informed, data-driven decisions. This approach will enhance transparency, foster trust, and support the creation of proactive strategies for improving customer satisfaction, loyalty, and retention. Ultimately, the project seeks to empower organizations to strengthen customer relationships, boost profitability, and ensure sustainable growth in a competitive market.

User Login/Registration:

- The user can either log in to an existing account or register a new one.
- The system validates the user's credentials.
- If the credentials are valid, the user proceeds to the next step.
- If the credentials are invalid, the user is prompted to try again.

Data Collection:

- The system collects data on customer behavior, which might include usage patterns, purchase history, and other relevant information.

Data Preprocessing:

- The collected data is preprocessed to prepare it for analysis. This may involve cleaning, normalization, and feature engineering.

Data Splitting:

- The preprocessed data is split into training and testing sets.

Model Building:

- The training set is used to train a machine learning model, which could be an Artificial Neural Network (ANN), XAI (Explainable AI), or Random Forest.

Model Prediction:

- The trained model is used to predict the churn probability for new or existing customers.

Result Display:

- The predicted churn probability is displayed to the user, categorized as either "Churn" or "Not Churn."
- In preprocessing steps, the system works to impute any disorders in the data set and extract the features.

2.3 Training:
In the training phase system generates the model from the dataset by using python modules.

2.4 Generate Results:
System generates the detection results from the model whether the customer continues with the same network or discontinue with the network.

3. XAI Techniques

3.1 Feature Prediction:
Users can check which feature causes churn.

VI. ALGORITHMS

1. ARTIFICIAL NEURAL NETWORKS
1. User:
V. *Upload:*

MODULES
Users can upload the dataset for the model building.

1.1 Model Training:
Users should select the machine learning model for training.

1.2 Prediction:
Users need to enter input to detect the desired output

1.3 View Results:
Users can view both results and graph generated by the system.

2. System

2.1 Take the dataset:
The system works with the dataset provided to it for model building.

2.2 Preprocessing:

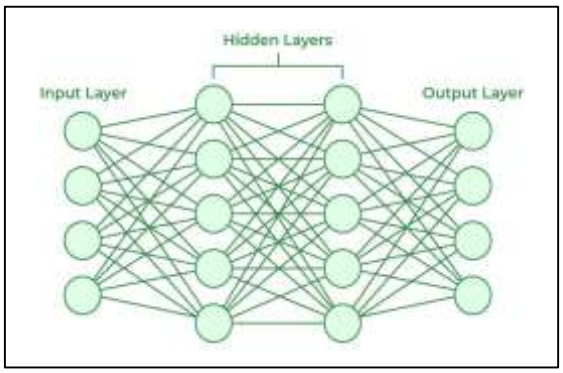

Figure 2: Neural network model

An Artificial Neural Network (ANN) is a computational model inspired by the human brain, consisting of interconnected neurons organized into layers: an input layer that receives data, hidden layers that process information, and an output layer that generates predictions. The ANN operates by receiving input data, propagating it through the hidden layers where weighted sums and activation functions introduce non-linearity, and finally producing an output, such as a classification or regression value. Training an ANN involves

designing its architecture, including the number of layers, neurons per layer, and activation functions, often implemented using libraries like TensorFlow or PyTorch. The training process includes forward propagation, where input data is passed through the network to generate predictions, and backpropagation, where the error between predicted and actual values is calculated and propagated backward to adjust weights and biases.

Optimization techniques like gradient descent iteratively update these parameters to minimize error, while validation data helps monitor performance and prevent overfitting, ensuring the model generalizes well to unseen data.

2. *RANDOM FORESTS*

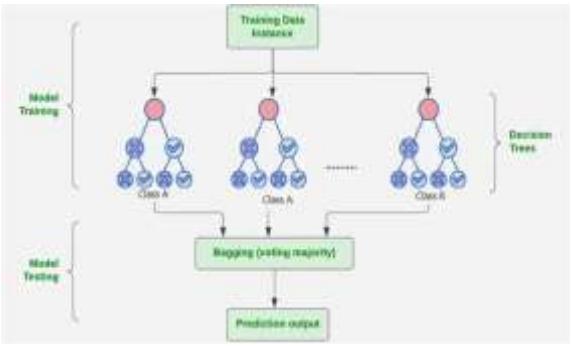

Figure 3: Random Forests Model

A Random Forest is an ensemble learning technique that enhances prediction accuracy and robustness by combining multiple decision trees. It operates through bootstrapping, where random subsets of training data are created with replacement, and each subset is used to train an independent decision tree. During tree construction, at each node, a random subset of features is selected for splitting, reducing overfitting and improving generalization. When making predictions, each tree in the forest votes, and the most frequent prediction is chosen as the final output, making Random Forests highly reliable for classification and regression tasks. Training a Random Forest model involves creating a classifier using machine learning libraries like scikit-learn, training it on the dataset, and tuning hyperparameters such as the number of trees and maximum depth. Techniques like grid search or cross- validation help optimize these parameters for better performance. A key advantage of Random Forests is their ability to determine feature importance, providing insights into how much each feature contributes to the model's predictions. These importance scores can be visualized using bar charts, helping analysts understand which factors most influence outcomes.

Model evaluation is performed using test data and metrics such as accuracy, precision, recall, and F1- score to measure effectiveness. Additionally, comparisons can be made between Random Forest and Artificial Neural Networks (ANNs) to determine the most suitable model for a given task. Random Forests offer several advantages, including high accuracy, robustness against overfitting, and the ability to handle missing or imbalanced data without requiring complex preprocessing. Their interpretability and effectiveness in various scenarios make them a powerful tool for predictive analytics.

VII. RESULTS

In the data analysis of the customer churn prediction, all the variables and their relations to one another are used as important features for prediction. The thorough Exploratory Data Analysis (EDA) provides important new information about the dynamics of customer turnover in the telecommunication industry. Important factors impacting churn are identified using descriptive statistics and visualizations like count plots and histograms. Most notably, the 'Churn Reason' histogram emphasizes number of service calls as the main motivator. The impact of category factors on customer retention can be revealed through analysis, such as "Type of Offer" and "Internet Service". Moreover, month-to-month contracts and the absence of premium tech assistance are shown to be major causes of churn in numerical feature histograms. These results highlight how crucial accommodating and helpful service options are raising client retention and satisfaction. When used in conjunction with careful pre-processing, the EDA guarantees strong data quality and provides useful insights. Through comprehension of these dynamics, interested parties can develop focused churn mitigation methods, which in turn can facilitate well-informed decision making and strategic interventions meant to reinforce customer satisfaction and loyalty in the cut-throat telecommunication market.

In the future, the scope of this project can be expanded by integrating real-time data pipelines and customer sentiment analysis from sources like social media and customer reviews. Incorporating advanced deep learning architectures such as LSTM or transformer models can help capture temporal patterns in customer interactions. Additionally, using unsupervised learning for customer segmentation, applying explainable AI (XAI) for transparency, and deploying the model in a cloud-based environment for scalability can significantly enhance the system's performance and applicability in real-world business environments.

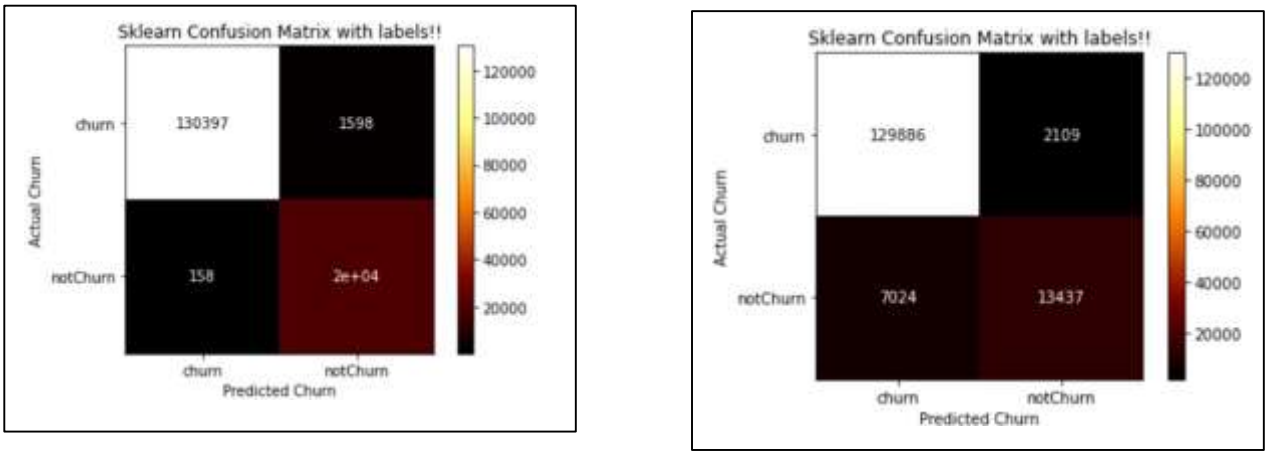

Fig 4: Confusion matrices of ANN and Random Forests

Results show that the ANN model has the accuracy of 86.20% and Random Forest classifier has an accuracy of 96.85%. due to a number of significant benefits, the random forest classifier here frequently performs better than other classifier models in use. First of all, because it is non-parametric, it can handle both skewed and non-ordinal data's underlying distribution. Because the random forest uses an ensemble learning approach, which aggregates the predictions of numerous decision trees, it can also handle many independent predictor variables without over-fitting. When compared to other classifiers, this strong methodology typically yields better predicted accuracy and generalization, particularly in complex datasets with high dimensionality and intricate feature interactions.

VIII. CONCLUSION & FUTURE SCOPE

The project on Customer Churn Analysis using Artificial Neural Networks (ANN) and Random Forests highlights the potential of machine learning in accurately predicting customer attrition. While Random Forests offer high interpretability and robustness, ANNs excel at identifying complex, non- linear patterns in customer behaviour. The combination of these models enables businesses to make informed decisions and proactively implement retention strategies.

REFERENCES

[1] Cross-Sector Application of Machine Learning in Telecommunications: Enhancing Customer Retention Through Comparative Analysis of Ensemble Methods by M. Afzal, S. Rahman, D. Singh and A. Imran. January 2024 IEEE Access.

[2] Customer Churn Prediction Using Machine Learning Methods: A Comparative Analysis by Konda Avinash, Varala Aditya Ram, Seguru Ritvik Reddy, Bharath Kumar Nangunuri. 2024 Oct IEEE Xplore.

[3] Time Series-based Churn Prediction Model Using LSTM by Zhang, W., Zhou, J. & Zhang, X. International Journal of Data Science and Analytics 2024 June.

[4] Hybrid Deep Learning Model for Customer Churn Prediction in E-commerce by Lin, X., Chen, Y., & Ren, Y. Electronic Commerce Research 2023 Dec.

[5] Enhancing Customer Churn Prediction: Addressing Disparities and Imbalance in Machine Learning Models. By J Emmanuel Chai, Kennedy Khadullo and Kevin Tole The International Journal of Engineering and Science (IJES) 2023 April.

[6] Customer churn prediction using composite deep learning technique by Khattak, A., Mehak, Z., Ahmad. Springer 12 Oct 2023.