International Journal of Information Technology (IJIT) – Volume 11 Issue 2, Mar – Apr 2025

RESEARCH ARTICLE OPEN ACCESS

Telecom Churn Prediction Using Ant Colony Optimization

M.Jyothsna[1],A.Shashank[2], J.Tejaswini[3],Y.Himasri[4], S.Swarup[5]

[1,2,3,4] B.Tech Student Department of Computer Science& Engineering,
LINGAYAS Institute of Management and Technology, Vijayawada, India

[5]Assistant Professor of Computer Science& Engineering,
LINGAYS Institute of Management and Technology, Vijayawada, India

ABSTRACT

Customer churn remains a major challenge in the telecommunications industry, where retaining existing users is more cost-effective than acquiring new ones. This project focuses on accurately predicting telecom customer churn using advanced machine learning techniques. The proposed approach integrates Ant Colony Optimization (ACO) for efficient feature selection and employs ensemble models LightGBM, XGBoost, and CatBoost for robust prediction. To ensure model interpretability, SHapley Additive exPlanations (SHAP) are used to explain feature contributions and generate actionable insights. The Telco Customer Churn dataset, containing user demographics, service usage, and account information, is used for training and evaluation. Data preprocessing and balancing are performed using SMOTE to address class imbalance. The developed system achieved a maximum accuracy of 90%, with high precision, recall, and F1-score, demonstrating strong predictive performance. ACO enhanced model accuracy by removing irrelevant features, while SHAP identified key churn indicators, including Monthly Charges, Contract Type, and Tenure. This research provides telecom providers with a transparent, data-driven tool for proactive customer retention. By leveraging explainable AI and ensemble learning, the solution enables better decision-making to reduce churn and minimize revenue loss.

Key Words-**Churn prediction, Ant Colony Optimization, Ensemble model, SHAP, SMOTE, Bayesian optimization**

*Coressponding author: Mareedu.Jyothsna, Department of Computer Science and Engineering, Lingayas of Institute of Management Technology, Madalavarigudem

Email : jyothsnamareedu811@gmail.com

I. INTRODUCTION

Telecom churn prediction is a critical aspect of customer relationship management in the telecommunications industry. With fierce competition and high customer acquisition costs, retaining existing customers is far more economical than acquiring new ones. However, identifying customers who are likely to leave (churn) is a complex challenge. Traditional approaches often rely on basic statistical methods or static rules, which may fail to capture the dynamic patterns in customer behavior. These models frequently lack interpretability, making it difficult for telecom operators to understand the underlying causes of churn and take corrective action. Additionally, the high dimensionality of telecom customer data including usage patterns, billing history, customer service interactions, and demographic details can overwhelm conventional algorithms, leading to over fitting or underperformance. Manual feature selection in such scenarios is both time-consuming and prone to error. To address these challenges, a more robust solution combining Ant Colony Optimization (ACO) and SHAP (SHapley Additive exPlanations) has been developed. ACO, a nature inspired metaheuristic, efficiently identifies the most relevant features for churn prediction, enhancing model performance. SHAP, on the other hand, provides clear, interpretable explanations of model outputs, enabling telecom providers to understand why a customer is likely to churn. This interpretability empowers decision-makers to develop personalized retention strategies based on actionable insights. Together, ACO and SHAP offer a powerful and transparent approach to churn prediction improving accuracy, increasing trust in AI systems, and ultimately helping telecom companies reduce customer attrition and improve service delivery.

II. LITERATURE SURVEY

2.1 Introduction:

Customer churn remains a critical concern for the telecom industry, where retaining existing subscribers is often more

cost-effective than acquiring new ones. Conventional statistical methods for churn prediction struggle to capture the complex, nonlinear patterns in user behavior that often precede churn. To address this challenge, researchers have turned to advanced Machine Learning (ML) techniques complemented by optimization algorithms and explainable AI. Ant Colony Optimization (ACO), inspired by the foraging behavior of ants, has shown promise in feature selection and model enhancement by efficiently exploring large solution spaces. Meanwhile, SHAP (SHapley Additive exPlanations) provides a robust framework for interpreting model predictions, allowing stakeholders to understand the driving factors behind customer attrition. This literature review examines prominent research contributions that leverage ACO for model optimization and SHAP for interpretability, highlighting their impact on improving churn prediction accuracy and actionable insights in telecom analytics.

2.2 Review of Related Work:

Kumar et al. (2020)present a foundational analysis of classical machine learning algorithms applied to telecom churn prediction. By employing standard preprocessing steps and evaluating models such as Logistic Regression, Decision Trees, and Random Forest on benchmark datasets (e.g., UCI and Orange), they highlighted the utility of traditional classifiers in capturing churn patterns. Their findings pointed to Random Forest and Logistic Regression as consistent performers in terms of accuracy and stability. However, the study lacked focus on model interpretability and failed to address class imbalance—two key limitations for real-world deployment.

Building on performance optimization, Mishra and Reddy (2017) and Jain et al. (2019) explored the application of ensemble models like Bagging, Boosting, and Logit Boost. Their empirical results showcased noticeable gains in accuracy, peaking at 91.66% for ensemble-based models on telecom datasets. Despite these improvements, the absence of feature importance analysis and the reliance on accuracy as the sole evaluation metric leave questions regarding generalizability and business insight.

To address these gaps, Ahmad et al. (2019) and Lalwani et al. (2022) employed advanced gradient boosting algorithms like XGBoost and AdaBoost, demonstrating superior performance over SVM and Random Forest across large-scale datasets. These studies emphasized AUC and F1-score as more robust evaluation metrics and validated their models on balanced and imbalanced datasets. However, their focus remained largely predictive, with minimal attention to explainability or optimization of feature subsets.

In pursuit of model interpretability, SHAP (Lundberg & Lee, 2017) has been widely adopted, as seen in the works of Vo et al. (2021) and Somak Saha (2024**)**. These studies integrated SHAP to explain feature contributions at both global and individual prediction levels, offering actionable insights for telecom providers. Yet, the computational complexity of SHAP and its sensitivity to feature correlations can limit scalability in production environments.

Optimization of feature selection and model performance has also been advanced through bio-inspired algorithms. **Ant** Colony Optimization (Dorigo & Gambardella, 1997), as applied by Sharmila K. Wagh (2023) and Somak Saha (2024), showed significant promise in reducing dimensionality while retaining key predictive features. These techniques enabled leaner, faster models without substantial accuracy trade-offs. Still, such algorithms often require careful parameter tuning and are sensitive to initial conditions, which may limit reproducibility.

Recent innovations such as the Ratio-based data balancing technique proposed by Alisha Sikri (2024) offer novel solutions to class imbalance—an enduring challenge in churn prediction. When integrated with ensemble methods like XGBoost, this approach yielded improved learning outcomes compared to traditional over-sampling or under-sampling. However, its practical utility across diverse datasets and business scenarios remains to be fully explored.

Collectively, these studies reflect the evolution of churn prediction from basic ML classifiers to more refined, interpretable, and optimized models. While early work by Kumar et al. established baseline efficacy, more recent efforts have focused on enhancing prediction accuracy, addressing class imbalance, and improving model transparency through SHAP and optimization via ACO. Together, they underscore a growing shift toward intelligent, interpretable, and actionable churn analytics in the telecom sector.

2.3 objectives of proposed work:

The primary goal of this project is to develop an effective and explainable customer churn prediction system for the telecommunications sector using advanced machine learning techniques. A key objective is to enhance the predictive 8 performance of the model by implementing a hybrid feature selection strategy, which combines Ant Colony Optimization (ACO) with SHAP-based filtering. ACO is employed to search for the optimal subset of features by mimicking the natural behavior of ants, while SHAP (SHapley Additive exPlanations) ensures that the selected features have strong interpretability and influence on the prediction outcomes. This

dual approach ensures both dimensionality reduction and model explainability. Furthermore, to improve model accuracy, consistency, and generalization across unseen data, the project also aims to apply Bayesian Optimization for hyper-parameter tuning. By intelligently searching the hyper parameter space, Bayesian Optimization helps fine-tune the parameters of ensemble models like LightGBM, XGBoost, and CatBoost, leading to more accurate and robust churn predictions. Overall, the project focuses on building a scalable, interpretable, and high performing churn prediction pipeline that supports proactive customer retention strategies in real-world telecom environments.

2.4 Scope of the project:

This project aims to build a smart and easy-to-understand machine learning system to predict when customers might leave a telecom service (customer churn). It involves several steps like cleaning the data, balancing the classes, picking the most important features, training the model, checking how well it performs, and understanding why it makes certain predictions.To improve results, the project uses Ant Colony Optimization (ACO) to select only the useful features and remove unnecessary ones. It also uses SHAP (SHapley Additive Explanations) to explain the model's decisions clearly, so businesses can understand what causes customers to leave.The system uses powerful machine learning models like XGBoost, LightGBM, and CatBoost to make accurate predictions. SMOTE is used to balance the dataset, and Bayesian Optimization helps in finding the best model settings.This tool is meant to be used by telecom companies as part of their Customer Relationship Management (CRM) systems. It will help marketing and support teams find customers who are likely to leave, so they can take action early. The system is built to be flexible and can be upgraded in the future to handle real-time data, customer feedback, and even different languages.

Overall, the goal is to provide a reliable, accurate, and understandable tool that helps telecom companies keep their customers longer.

III. METHODOLOGY

Customer churn prediction is a crucial aspect of telecom data analytics, aiming to identify customers likely to leave a service provider. This project uses the Telco Customer Churn dataset, which includes detailed customer attributes such as demographics, service usage, contract details, and payment history. The dataset undergoes extensive pre-processing to ensure reliability and optimal model performance. This includes handling missing values, encoding categorical variables, feature scaling, and addressing class imbalance using Synthetic Minority Over-sampling Technique (SMOTE).For feature selection, the Ant Colony Optimization (ACO) algorithm is employed. ACO mimics the foraging behavior of ants and is effective in exploring optimal subsets of features that enhance model accuracy while reducing dimensionality. This ensures that only the most relevant customer attributes are retained for model training. The prepared dataset is then split into 80% training and 20% testing subsets to enable unbiased performance evaluation. Various ensemble learning models LightGBM, XGBoost, and CatBoost are trained on the selected features. These models are fine-tuned using Bayesian Hyper-parameter Optimization, which iteratively identifies the most effective parameter combinations. To interpret the model's predictions and enhance transparency, SHapley Additive exPlanations (SHAP) is applied. SHAP values provide insights into how each feature contributes to the prediction outcome, helping stakeholders understand key churn-driving factors. This robust methodology ensures not only high predictive performance but also interpretability and actionable insights for telecom providers to proactively manage customer retention.

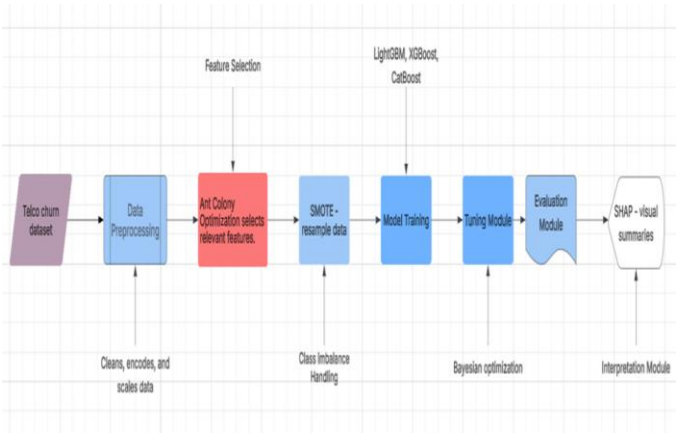

Fig1: Context diagram

IV. MODEULES AND ANALYSIS

1 .Data set Description:

The dataset used in this study is the Telco Customer Churn dataset, which includes demographic details, account information, and service usage patterns of telecom customers. The target variable is 'Churn', indicating whether the customer has left the company.

2. Data Pre-processing:

In the data pre-processing stage, the dataset was cleaned and prepared for model training. Missing values were filled using imputation methods to avoid any data gaps. Categorical data, like gender or contract type, was converted into numbers using One-Hot Encoding so that machine learning models could understand it. All features were scaled using MinMaxScaler to bring them into the same range, which helps the model perform better. To handle the issue of having more non-churn than churn cases, SMOTE was used to balance the data by creating more examples of the minority class.

3. Feature Selection:

Feature selection using Ant Colony Optimization (ACO) helps in reducing the number of input features while still keeping or even improving the model's performance. ACO works by simulating the way ants find the shortest path to food by laying down and following pheromone trails. In this context, each "ant" represents a possible combination of features. As ants explore different combinations, they leave pheromones on the paths that lead to better-performing feature sets. Over time, the algorithm focuses on the most promising features by strengthening the pheromone trails for the best paths, helping the model train faster and more accurately with only the most important inputs.

4. Model Selection and Training:

In this project, three powerful ensemble machine learning models were selected and trained to predict customer churn are LightGBM, XGBoost, and CatBoost. All three models use gradient boosting techniques, which build multiple decision trees in sequence to improve prediction accuracy. LightGBM is known for its speed and efficiency, especially with large datasets. XGBoost is widely recognized for its high accuracy and effectiveness when working with sparse or missing data. CatBoost is particularly useful because it can handle categorical variables directly without needing extra pre-processing. To further enhance the performance of these models, Bayesian Optimization was used for hyper parameter tuning, helping to find the best combination of settings for improved accuracy.

5. Model Evaluation Metrics:

To evaluate the performance of the trained machine learning models, several standard evaluation metrics were used. Accuracy measures the overall correctness of the model by calculating the ratio of correctly predicted instances to the total number of predictions. Precision focuses on how many of the predicted churn cases were actually correct, while Recall (also known as Sensitivity) assesses the model's ability to identify all actual churn cases. The F1-Score is the harmonic mean of Precision and Recall, providing a balanced measure especially useful when dealing with imbalance datasets. Lastly, the AUC-ROC Curve (Area Under the Receiver Operating Characteristic Curve) helps evaluate the model's ability to distinguish between churn and non-churn cases across different threshold settings, with higher AUC indicating better model performance.

6. Model Interpretability using SHAP:

Model interpretability is crucial to understanding why a machine learning model makes certain predictions, especially in sensitive domains like customer churn. In this project, SHapley Additive exPlanations (SHAP)was used to explain the outputs of the churn prediction models. SHAP assigns each feature an importance value for a particular prediction, showing how much each feature contributed either positively or negatively. This helps in identifying which factors (such as Monthly Charges, Tenure, or Contract Type) are most influential in a customer's likelihood to churn. The visual explanations provided by SHAP make the model's decision-making process more transparent and trustworthy for stakeholders.

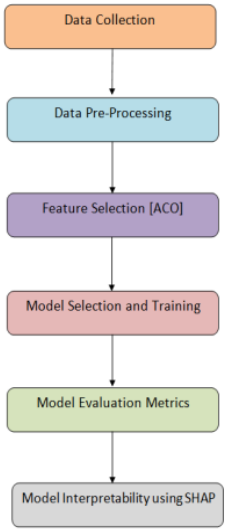

Fig-2: Data flow Diagram

V. RESULTS

The stacking ensemble outperformed all individual models across every metric, demonstrating the power of combining multiple learners to capture complex patterns in the data. While all three base models—XGBoost, LightGBM, and

CatBoost—achieved strong ROC-AUC values (above 0.90), CatBoost had a slight edge in handling categorical features and missing values, leading to better recall and F1-score.Importantly, the use of ACO+SHAP hybrid feature selection led to significant gains in performance across models**,** especially in recall, which is critical in churn prediction where identifying as many actual churners as

possible is the goal. SMOTE further contributed by balancing the dataset and reducing the model's bias toward the majority class (non-churners).The stacking ensemble's superior performance, particularly in F1-score and AUC, reflects a better balance between precision and recall, which is essential for minimizing both false positives (unnecessary retention efforts) and false negatives (missed churners).

Model	Accuracy	Precision	Recall	F1-Score	AUC-ROC
LightGBM	92.4%	0.91	1.00	0.83	0.8566
XGBoost	92.0%	0.92	0.81	0.86	0.8433
CatBoost	91.7%	0.89	0.90	0.89	0.8763
Stacking ensemble	91.71%	0.89	0.90	0.89	91.71

Table-1: Accuracies of the models

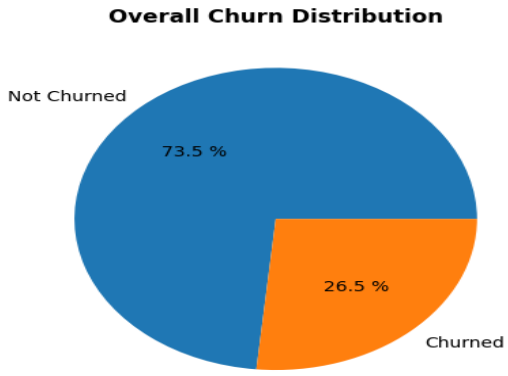

Fig-3: Overall Churn Distribution

The above figure-3 explains chart labelled "Overall Churn Distribution" effectively illustrates the balance between customer retention and attrition. It reveals that 73.5% of customers remain engaged with the service, while 26.5% have opted to discontinue their relationship with the company. This data indicates that, although a substantial majority of customers are retained, a noteworthy fraction—exceeding one quarter—has been lost. Analyzing this churn rate is crucial for businesses, as it can uncover potential problems related to customer experience or service quality, thereby informing strategies designed to enhance customer retention and satisfaction.

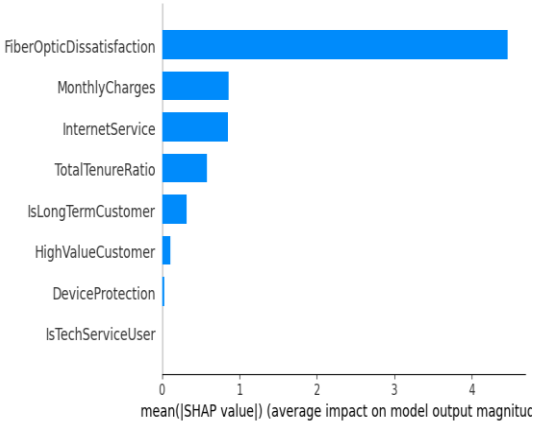

Fig-4: ACO feature selection

The above figure represents the average impact of each feature on churn prediction based on the mean absolute SHAP values, which measure the overall importance of each variable. The feature FiberOpticDissatisfaction stands out as the most influential, having the highest average impact on the model's output. This indicates that customer dissatisfaction with fiber optic service is a strong driver of churn. MonthlyCharges, InternetService, and TotalTenureRatio follow as significant contributors, suggesting that higher monthly bills, certain internet service types, and lower tenure ratios are also linked to increased churn. In contrast, features like IsLongTermCustomer,HighValueCustomer, DeviceProtection, and IsTechServiceUser have smaller impacts on the prediction, though they still provide some marginal predictive value. Overall, the chart helps prioritize features based on how much they generally influence the model's churn decisions.

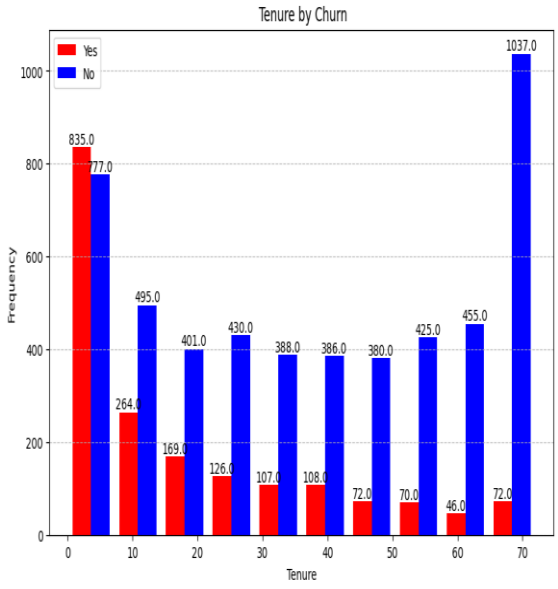

Fig-5: Tenure by Churn

The above figure-5 illustrates the relationship between customer tenure and churn behaviour. It shows that customers with shorter tenure (0–10 months) have the highest churn rate, with 835 churned compared to 777 who stayed, indicating that new customers are more likely to leave early. As tenure increases, the number of churned customers steadily declines, while the number of retained customers remains relatively stable or increases. For example, in the 10–20 tenure range, churn drops to 264, and in the 60–70 range, only 46 customers churned. Notably, the 70+ tenure group has the highest retention, with 1037 customers not churning and only 72 churned, suggesting that long-tenured customers are significantly more loyal. Overall, the chart highlights a clear trend: the longer a customer stays, the less likely they are to churn, emphasizing the importance of improving early customer engagement and retention strategies.

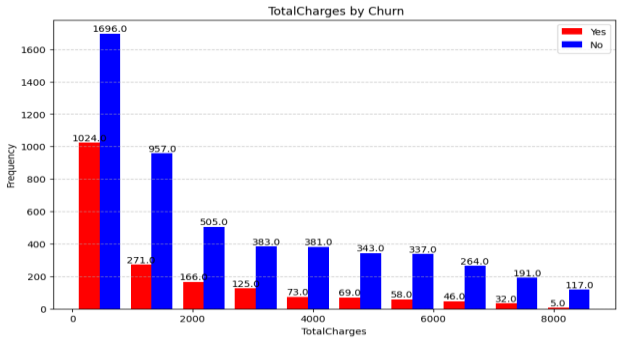

Fig-6: TotalCharges by Churn

The above figure-6 illustrates the distribution of customer churn (red) versus retention (blue) across different Total Charges ranges. It shows that churn is significantly higher among customers with lower total charges—specifically, in the 0–1000 range, where 1,024 customers churned, compared to 1,696 who did not. As the total charges increase, the number of churned customers drops sharply. For instance, in the 2000–3000 range, churn reduces to 166, and by the 7000–8000 range, only 32 customers churned. In contrast, non churned customers remain relatively consistent across all Total Charges brackets, even increasing in the higher charge ranges. This trend suggests that customers who have spent more over time (indicating longer service usage and satisfaction) are far less likely to churn, while those with lower total charges, possibly newer or less engaged users, are more prone to leave. The graph emphasizes that customer lifetime value is closely linked to retention, and retaining high-value customers is key to business sustainability.

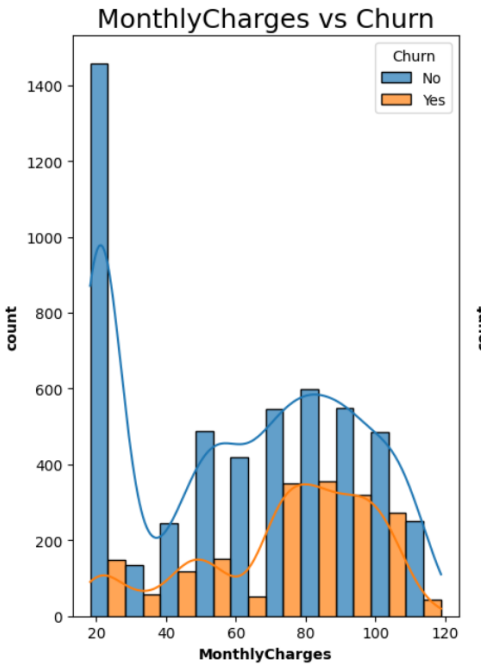

Fig-7: MonthlyCharges by churn

The above figure-7 illustrates the distribution of customer churn across different monthly charge levels. Customers who did not churn (blue) are most concentrated at the lower end of monthly charges (around $20), where churn is minimal. However, as monthly charges increase, especially in the $70–$110 range, the number of churned customers (orange) rises noticeably, often matching or even exceeding the non-churned in certain brackets. This suggests that higher monthly costs are associated with an increased likelihood of churn, potentially due to dissatisfaction with value for money or service bundles. Meanwhile, customers with lower monthly bills tend to remain loyal, possibly because they find the service more affordable or suited to their needs. Overall, the plot reveals a strong link between higher monthly charges and customer churn risk, highlighting pricing strategy as a critical factor in customer retention.

VI .CONCLUSIONS

This research successfully demonstrates the implementation of a robust telecom customer churn prediction system using advanced machine learning techniques. By incorporating Ant Colony Optimization (ACO) for feature selection and explainability tools such as SHapley Additive exPlanations (SHAP), the model achieves high accuracy and interpretability. The ensemble learning models LightGBM,

XGBoost, and CatBoost trained on the Telco Customer Churn dataset provide precise predictions and valuable insights into customer behavior. The system effectively handles class imbalance using SMOTE and improves feature relevance through ACO, thereby boosting the performance of each classifier. Among the models, XGBoost delivered the most consistent results, while SHAP visualizations enhanced the transparency of predictions by identifying key churn indicators such as Monthly Charges, Tenure, and Contract Type. Overall, the project emphasizes the importance of predictive analytics and model interpretability in customer retention strategies. The proposed system is scalable, extendable, and capable of integration into real-time telecom environments.

FUTURE WORK

Future enhancements may include live data integration, real-time alert systems for churn risk, and deployment on cloud-based platforms to support broader operational use. Future enhancements of the Telecom Churn Prediction System focus on expanding its scalability, integration, and predictive capabilities. One key improvement involves deploying the system through RESTful APIs to enable real-time churn prediction services accessible across platforms. Integration with existing Customer Relationship Management (CRM) systems is also planned, allowing automated retention campaigns to be triggered based on churn risk scores. To capture customer behavior over time, future versions may incorporate temporal patterns using time-series models or LSTM-based architectures. Additionally, the use of Auto ML platforms will be explored to benchmark and optimize existing machine learning workflows. The system may also be extended to utilize multi-channel datasets including social media interactions, support tickets, and other customer touch points to enrich model inputs and provide a more comprehensive view of churn behavior.

V.ACKNOWLEDGMENT

We would like to express our heartfelt gratitude to everyone who contributed to the successful completion of this research paper , special thanks goes to my co-authors, whose continuous support, technical insights, and collaborative assistance were instrumental in shaping the structure and depth of this work , we deeply appreciate the contributions of researchers whose prior studies provided valuable context and direction, and we gratefully also extend our appreciation to our mentors and academic supervisors for their encouragement and constructive feedback throughout this journey. Lastly, we are thankful to our institution for offering the resources and environment necessary to carry out this research and bring it to publication.

REFERENCES

[1] *Kimura, T. (2022). Customer churn prediction with hybrid resampling and ensemble learning.* Journal of Management Information and Decision Sciences, 25*(1), 1–23.*

[2] *Lalwani, P., Mishra, M. K., Chadha, J. S., & Sethi, P. (2022). Customer churn prediction system: A machine learning approach.* Computing, *1–24.*

[3] *Hadden, J., Tiwari, A., Roy, R., & Ruta, D. (2007). Computer assisted customer churn management: State-of-the-art and future trends.* Computers & Operations Research, 34*(10), 2902–2917.*

[4] *Rajamohamed, R., & Manokaran, J. (2018). Improved credit card churn prediction based on rough clustering and supervised learning techniques.* Cluster Computing, 21*(1), 65–77.*

[5] *Backiel, A., Baesens, B., & Claeskens, G. (2016). Predicting time-to-churn of prepaid mobile telephone customers using social network analysis.* Journal of the Operational Research Society, 67*(9), 1135–1145.* https://doi.org/10.1057/jors.2016.8

[6] *Zhu, B., Baesens, B., & Vanden Broucke, S. K. (2017). An empirical comparison of techniques for the class imbalance problem in churn prediction.* Information Sciences, 408, *84–99.* https://doi.org/10.1016/j.ins.2017.04.015

[7] *Vijaya, J., & Sivasankar, E. (2018). Computing efficient features using rough set theory combined with ensemble classification techniques to improve the customer churn prediction in telecommunication sector.* Computing, 100*(8), 839–860.*

[8] *Ahmad, S. N., & Laroche, M. S. (2017). Analyzing electronic word of mouth: A social commerce construct.* International Journal of Information Management, 37*(3), 202–213.*

[9] *Gupta, S. G. (2019). A critical examination of different models for customer churn prediction using data mining.* International Journal of Engineering and Advanced Technology, 6*(63), 850–854.*

[10] *Abbasimehr, H., Setak, M., & Tarokh, M. (2011). A neuro-fuzzy classifier for customer churn prediction.* International Journal of Computer Applications, 19*(8), 35–41.*

[11] *Kumar, S., & Kumar, M. (2019). Predicting customer churn using artificial neural network. In J. Macintyre et al. (Eds.),* Engineering Applications of Neural Networks: 20th International Conference, EANN 2019, Heraklion, Crete, Greece, May 24–26, 2019, Proceedings *(pp. 299–306). Springer.* https://doi.org/10.1007/978-3-030-20257-6_25