IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)
An International Open Access, Peer-reviewed, Refereed Journal

Predictive Analytics For Telecom Churn Prediction: A Multi-Model Approach

[1]Dr.Santosh Singh, [2]Poonam Jain, [3]Ashirvaad Bhat, [4]Rahul Menon
[1]HOD, [2]Assistant Professor, [3,4] PG Student
[1,2,3,4]Department of IT,
[1] Thakur College of Science and Commerce, Thakur Village,
Kandivali(East),Mumbai, Maharashtra, India

Abstract: This research investigates machine learning and deep learning techniques for predicting customer churn in telecom services. We employ various models, including Logistic Regression and Temporal Convolutional Network (TCN), to classify customers based on their likelihood of churning. The dataset undergoes preprocessing steps such as feature encoding, scaling, and transformation to ensure optimal model performance. Logistic Regression is used as a baseline, while TCN leverages sequential patterns in customer data for enhanced predictive accuracy. The results highlight the advantages of deep learning architectures in capturing temporal dependencies within telecom datasets, offering insights into customer retention strategies.

Index Terms **- machine learning, deep learning, telecom churn prediction, logistic regression, temporal convolutional network, customer retention.**

I. Introduction

Customer churn is a major challenge for telecom service providers, directly impacting revenue and customer retention strategies. With rising competition, minimizing churn is crucial, requiring accurate predictive models to identify at-risk customers. However, predicting churn is complex due to various influencing factors such as service usage, billing issues, and customer support interactions.

Traditional models like Logistic Regression and Decision Trees offer interpretability but struggle to capture intricate, nonlinear patterns in customer behavior. Telecom data is inherently sequential, necessitating advanced techniques that can model temporal dependencies. While deep learning approaches such as Recurrent Neural Networks (RNNs), Long Short-Term Memory (LSTM) networks, and Gated Recurrent Units (GRUs) have been explored, they suffer from vanishing gradients and high computational costs. A promising alternative is Temporal Convolutional Networks (TCN), which leverage causal and dilated convolutions to capture long-range dependencies without recurrent structures.

This study compares Logistic Regression, a widely used baseline, with TCN, a deep learning model designed for sequential data. Using datasets containing service usage, billing history, and complaint logs, we aim to develop predictive models for churn detection. Our findings will offer insights into the effectiveness of different techniques, guiding telecom providers in adopting optimal strategies for customer retention.

II. Literature review

Our study includes traditional models such as Logistic Regression, Decision Trees, and Support Vector Machines (SVM), which have been widely used for churn prediction. Logistic Regression offers interpretability but lacks the ability to capture nonlinear relationships, whereas Decision Trees provide better decision rules but may suffer from overfitting (Ahmed et al., 2020).[1]

Our study presents ensemble methods such as Random Forest and Gradient Boosting Machines (GBM) to improve predictive accuracy. These models combine multiple weak classifiers to enhance overall performance, effectively capturing important churn indicators like service usage patterns, customer tenure, and billing history (Chen & Guo, 2021).[2]

Our study includes deep learning models such as Recurrent Neural Networks (RNNs), Long Short-Term Memory (LSTM) networks, and Gated Recurrent Units (GRUs), which have been extensively investigated for their ability to model sequential dependencies in customer behavior. However, these methods are computationally expensive and prone to vanishing gradient problems (Hassan et al., 2022).[3]

An emerging alternative in our study is the Temporal Convolutional Network (TCN), which utilizes causal and dilated convolutions to process sequential data without recurrent connections. Studies indicate that TCN achieves superior performance in churn prediction by effectively capturing long-term dependencies while maintaining computational efficiency (Zhang et al., 2023). [4]

Our study presents feature selection and transformation as essential steps in building accurate churn prediction models. Techniques such as Recursive Feature Elimination (RFE) and Principal Component Analysis (PCA) are commonly employed to enhance model performance. Moreover, addressing class imbalance through Synthetic Minority Over-sampling Technique (SMOTE) or cost-sensitive learning is crucial for improving predictive accuracy (Singh & Patel, 2023).[5]

Our study includes comparative analyses revealing that deep learning techniques, particularly TCN, outperform traditional machine learning methods in handling complex, sequential telecom datasets. While models such as Logistic Regression and Decision Trees provide interpretability, deep learning models exhibit higher predictive accuracy and robustness (Lee et al., 2024).[6]

III. METHODOLOGY

Data Collection:

The dataset used in this study consists of customer records from a telecom provider. It includes features such as customer demographics, subscription history, service usage patterns, billing details, customer support interactions, and churn labels. The dataset is stored in a structured format (CSV) and undergoes initial exploration to understand feature distributions and missing values.

Data Preprocessing:

To ensure data quality and enhance model performance, various preprocessing steps are applied:

- **Handling Missing Values:** Missing entries in numerical columns are imputed using the mean, while categorical missing values are replaced with the most frequent category.
- **Feature Encoding:** Categorical variables, including 'telecom_partner,' 'state,' and 'subscription_type,' are transformed using one-hot encoding to facilitate their use in machine learning models.
- **Feature Scaling:** Continuous features, such as monthly charges and tenure, are normalized using **StandardScaler** to ensure consistency across different models.
- **Temporal Features:** The 'date_of_registration' column is converted to a datetime format, and a new feature, 'days_on_platform,' is computed to quantify customer tenure.

Feature Selection:

Relevant features, including customer demographics, usage metrics, and service details, are selected based on domain knowledge and correlation analysis. Although explicit feature selection techniques like Recursive Feature Elimination (RFE) are not applied, only features contributing significantly to churn prediction are retained.

Machine Learning Techniques:

Multiple machine learning and deep learning models are implemented to predict customer churn:

- **Logistic Regression:** A baseline model using L2 regularization to establish benchmark performance.
- **Temporal Convolutional Network (TCN):** A deep learning model leveraging convolutional layers to capture temporal dependencies in customer data.
- **Random Forest Classifier:** An ensemble learning model used for comparative analysis.

Model Training and Hyperparameter Tuning:

- Logistic Regression is trained with an L2 penalty, using the **lbfgs** solver with 1000 iterations.
- The TCN model is designed with **two Conv1D layers** (64 and 128 filters, kernel size = 3), **batch normalization**, **dropout (30%)**, and **fully connected layers** with a sigmoid activation function. The Adam optimizer is used, and the model is trained for **20 epochs** with a batch size of **64**.
- Random Forest hyperparameters, such as the number of trees and depth, are fine-tuned using grid search for optimal performance.

Model Evaluation:

Each model's performance is assessed using multiple evaluation metrics:

- **Accuracy Score:** Measures overall prediction correctness.
- **Precision, Recall, and F1-Score:** Used to analyze class-wise performance, especially given the imbalanced nature of churn data.
- **Confusion Matrix:** Visualized using Seaborn to understand classification distribution.
- **Feature Importance Analysis:** Applied to logistic regression and random forest models to identify key churn indicators.

Model Interpretation:

While accuracy metrics offer performance insights, model explainability is critical for practical implementation. **SHAP (Shapley Additive Explanations)** is applied to analyze feature contributions in customer churn predictions. Additionally, the impact of customer tenure, billing patterns, and service usage frequency is examined through data visualization.

Ethical Considerations:

Predictive analytics in customer churn must adhere to ethical guidelines:

- **Data Privacy:** Customer data is anonymized, and personally identifiable information is excluded.
- **Bias Mitigation:** The dataset is evaluated for potential biases related to gender, location, or service type to prevent discriminatory predictions.
- **Transparency:** Explainable AI techniques are used to ensure model predictions are interpretable and actionable for business decision-making.

Software and Tools:

The following Python libraries and frameworks are used for data handling, model development, and evaluation:

- **pandas, NumPy:** Data preprocessing and manipulation.
- **scikit-learn:** Machine learning model implementation and evaluation.
- **TensorFlow, Keras:** Deep learning model development, particularly for TCN.
- **Matplotlib, Seaborn:** Data visualization for insights and feature analysis.

Limitations:

Despite a robust methodology, certain limitations exist:

- **Imbalanced Data:** The dataset may have fewer churned customers, impacting model performance. Techniques like oversampling or SMOTE can be explored in future work.
- **Feature Selection:** While domain knowledge guides feature selection, automated methods like PCA or RFE can improve performance.
- **Computational Costs:** Deep learning models like TCN require more computational resources compared to traditional models like logistic regression.

IV. RESULTS

The analysis employed various machine learning and deep learning models to predict telecom customer churn based on behavioral and subscription data. Logistic Regression, Random Forest, and Temporal Convolutional Network (TCN) were evaluated, each providing unique insights into churn prediction.

- **Logistic Regression** achieved an accuracy of **0.82**, serving as a baseline model. It provided interpretable feature importance, highlighting key churn indicators such as customer tenure and billing history.
- **Random Forest** demonstrated an improved accuracy of **0.87**, effectively capturing nonlinear relationships in customer data. The model performed particularly well in identifying high-risk churn customers.
- **Temporal Convolutional Network (TCN)** outperformed traditional models with an accuracy of **0.89**, leveraging sequential dependencies in customer behavior to enhance predictions. It showed superior precision and recall, reducing false negatives and ensuring better retention strategies.

Feature importance analysis identified **customer tenure, monthly charges, and service usage frequency** as key churn indicators. The confusion matrix and classification reports revealed that deep learning approaches, particularly TCN, provided a more balanced prediction across churn and non-churn classes. These results demonstrate the advantage of deep learning models in capturing complex sequential patterns in customer data.

VII. CONCLUSION

The implementation of various machine learning and deep learning models, including Logistic Regression, Random Forest, and Temporal Convolutional Networks (TCN), proved effective in predicting customer churn in the telecom industry. Each model offered distinct strengths, with Logistic Regression providing interpretability, Random Forest capturing nonlinear relationships, and TCN excelling in identifying sequential patterns in customer behavior.

The findings highlight the potential of **deep learning-based approaches**, particularly TCN, in **enhancing churn prediction accuracy and improving customer retention strategies**. Businesses can leverage these insights to develop proactive engagement strategies, reduce customer attrition, and optimize marketing interventions.

Future research will focus on **hybrid models and ensemble techniques** to further enhance predictive performance. Additionally, integrating real-time customer interactions and external market trends could refine churn models, making them more robust and adaptive to dynamic customer behaviors. By leveraging predictive analytics, telecom providers can drive data-driven decision-making, improving overall business sustainability and customer satisfaction.

VIII. REFERENCES

1. Ahmed, M., Jones, L., & Kim, H. (2020). Predicting customer churn in telecom using logistic regression and decision trees. *International Journal of Data Science and Analytics*, 7(3), 112-126.
2. Chen, J., & Guo, W. (2021). An ensemble-based approach for telecom churn prediction using Random Forest and Gradient Boosting. *Journal of Machine Learning Research*, 22(1), 556-572.
3. Hassan, R., Sharma, P., & Kaur, M. (2022). Deep learning for customer churn prediction: A comparative study of LSTM and GRU. *IEEE Access*, 10, 23456-23468.
4. Zhang, L., Wang, Y., & Li, X. (2023). Temporal Convolutional Networks for customer churn prediction in telecom services. *Expert Systems with Applications*, 215, 119304.
5. Singh, A., & Patel, R. (2023). Feature selection techniques for improving churn prediction models in telecom. *Applied Soft Computing*, 129, 109567.
6. Lee, C., Park, H., & Kim, J. (2024). Machine learning vs. deep learning: A comparative analysis of churn prediction models. *Neural Computing and Applications*, 36(4), 678-695.