Telecom Subscriber Churn Prediction using Deep Learning : Examining the Influence of Customer Characteristics and Service Offerings

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

[123]Computer Science and Engineering Department,

Rajalakshmi Engineering College Chennai

[1]anandhi.s@rajalakshmi.edu.in

[2]200701309@rajalakshmi.edu.in

[3]200701237@rajalakshmi.edu.in

ABSTRACT **-** Telecom companies are facing a persistent challenge called customer churn, where subscribers terminate their services and switch to competing carriers. This phenomenon directly impacts revenue and long-term sustainability. To address this challenge, telecom companies are increasingly turning to predictive analytics and machine learning techniques. These methods entail creating prediction models that project client attrition by leveraging enormous archives of past customer data. By analyzing these models, telecom companies can proactively address churn dynamics and enhance customer retention efforts. To build robust churn prediction models, telecom companies explore a variety of machine learning algorithms and techniques. Traditional approaches like Gradient Boosting and Random Forest remain popular, but there is a growing interest in more advanced methodologies. Models like LightGBM offer promising avenues for enhancing predictive accuracy and performance. Additionally, the incorporation of Convolutional Neural Networks (CNNs), Long Short-Term Memory (LSTM) networks, and hybrid architectures like LSTM-CNNs and Wide and Deep Neural Networks present novel opportunities for capturing intricate patterns in customer data and improving churn prediction capabilities. The evaluation and selection of these models entail a rigorous assessment process, where performance metrics such as F1-score, accuracy, recall, precision, and ROC-AUC are meticulously scrutinized. By leveraging these state-of-the-art predictive analytics tools, telecom companies aim to not only identify customers at risk of churn but also to tailor targeted retention strategies that resonate with individual preferences and behaviors. Ultimately, the goal is to empower telecom providers with actionable insights that enable them to proactively address customer churn and fortify their competitive positioning in an increasingly volatile market landscape.

Keywords - Churn, Customer, Deep learning techniques, LightGBM, LSTM-CNNs, Wide and Deep Neural Networks, at-risk customers

1. Introduction

Customer churn prediction stands as a crucial instrument for telecom firms in curbing revenue losses and fortifying customer retention. The telecom sector confronts significant hurdles associated with customer churn, placing it at the forefront of priorities for leading corporations. By tapping into historical customer data encompassing consumption patterns, billing details, and consumer behavior, telecom operators can forecast potential churners and implement proactive measures to retain vulnerable customers. Machine learning algorithms play a pivotal role in discerning vital indicators and trends that hint at potential churn, thereby enabling the delivery of personalized offers and marketing campaigns to mitigate churn. For example, the Percentage of Transactions to/from other Operators emerges as a pivotal feature, with higher values often linked to churners, suggesting the influence of peer networks on churn decisions and the strong presence of rival companies in specific regions or communities. Moreover, exploratory data analysis (EDA) unveils invaluable insights such as gender distribution and the ratio of Senior Citizens to Non-Senior Citizens, providing telecom companies with deeper insights into their customer base. This approach proves pivotal for telecom companies as it aids in reducing customer acquisition costs and fostering strong client relationships, addressing a major concern within the industry. The integration of customer churn prediction into retention efforts is imperative for telecom companies to adapt to evolving customer needs and preferences, ultimately resulting in decreased churn rates and heightened customer satisfaction. In crafting the project solution, a multifaceted approach incorporating Wide and Deep Neural Networks, Light, Temporal Convolutional Neural Networks (CNN), and a fusion of Long Short-Term Memory (LSTM) with CNN models will be applied. These models will be harnessed to capture intricate patterns and temporal dependencies prevalent in telecom customer data, thereby enhancing the accuracy and efficacy of churn prediction and retention strategies. The Wide and Deep Neural Network will encapsulate both linear and non-linear relationships among features, while LightGBM will excel in handling extensive datasets and extracting feature importance. The Temporal CNN will adeptly capture temporal dependencies in sequential data such as consumption patterns and transaction histories. Lastly, the LSTM-CNN combination will effectively capture long-term dependencies and spatial features, furnishing a comprehensive understanding of customer churn behavior in the telecom sector. By amalgamating these advanced modeling techniques, the project seeks to formulate a robust and scalable solution for telecom companies to forecast and mitigate customer churn, ultimately fostering sustainable growth and customer contentment within the industry.

2. APPROACH

2.1 Temporal Convolutional Neural Network

Employing a Temporal Convolutional Neural Network (TCNN) for customer churn prediction offers a powerful approach rooted in its ability to capture sequential patterns effectively. TCNNs leverage convolutional operations to learn hierarchical representations of temporal data, making them adept at discerning complex patterns over different time scales. In the context of customer churn, the temporal nature of user interactions can be efficiently modeled by organizing the input data as a sequence of relevant features. The TCNN architecture involves stacking multiple one-dimensional convolutional layers, each followed by non-linear activation functions and pooling operations. These layers enable the model to automatically learn and extract temporal patterns at various levels of abstraction. Mathematically, the output of a TCNN layer i can be expressed as $h_i = ReLU\ (W_i * x + b_i\ )$, where W_i is the convolutional kernel, x is the input sequence, b_i is the bias term, and ReLU is the rectified linear unit activation function. The final prediction is derived by combining the outputs of these layers. The model is trained using a binary cross-entropy loss function that quantifies the dissimilarity between predicted probabilities and actual churn labels. The TCNN's effectiveness in capturing temporal dependencies within sequential data makes it well-suited for customer churn prediction, allowing for accurate identification of patterns that may precede or indicate customer churn.

2.2 LSTM + CNN

The LSTM+CNN hybrid architecture is a potent approach for customer churn prediction, leveraging the strengths of both Long Short-Term Memory (LSTM) networks and Convolutional Neural Networks (CNN). The LSTM component excels at capturing temporal dependencies within sequential customer data, crucial for understanding patterns that evolve. Each LSTM unit processes input sequences, updating hidden and cell states using intricate gating mechanisms that enable the network to selectively retain or discard information. The CNN component complements this temporal modeling by capturing spatial patterns and discerning relevant features across different positions in the input sequence. The convolutional layer applies filters to identify local patterns in the sequential data, enhancing the network's ability to extract meaningful spatial information. The combination of LSTM and CNN outputs creates a comprehensive representation that incorporates both temporal and spatial aspects of customer behavior. This joint representation is then fed into fully connected layers for final prediction. The model is trained using a loss function that typically involves a combination of binary cross-entropy and regularization terms, striking a balance between accurate prediction and prevention of overfitting. The LSTM+CNN architecture thus provides a robust solution for customer churn prediction, effectively capturing the complex interplay of temporal and spatial dynamics in customer data.

2.3 Wide and Deep Neural Network

In customer churn prediction, a common and effective approach is to employ a wide and deep neural network architecture. The wide component allows the model to capture intricate patterns and relationships in the data through a large number of cross-product feature interactions. This is achieved by including a broad set of sparse features, such as categorical variables, in the input layer. The deep component, on the other hand, leverages a series of hidden layers to learn hierarchical representations of the data, enabling the model to grasp more abstract and complex patterns. The combination of both wide and deep components allows the neural network to balance the memorization of specific instances with generalization to new data. Mathematically, the wide component involves the computation of cross-product feature interactions, often represented by the dot product or concatenation of embeddings. The deep component, on the other hand, utilizes activation functions like ReLU and employs backpropagation for iterative parameter updates. The overall loss function typically involves a combination of binary cross-entropy and regularization terms to prevent overfitting.

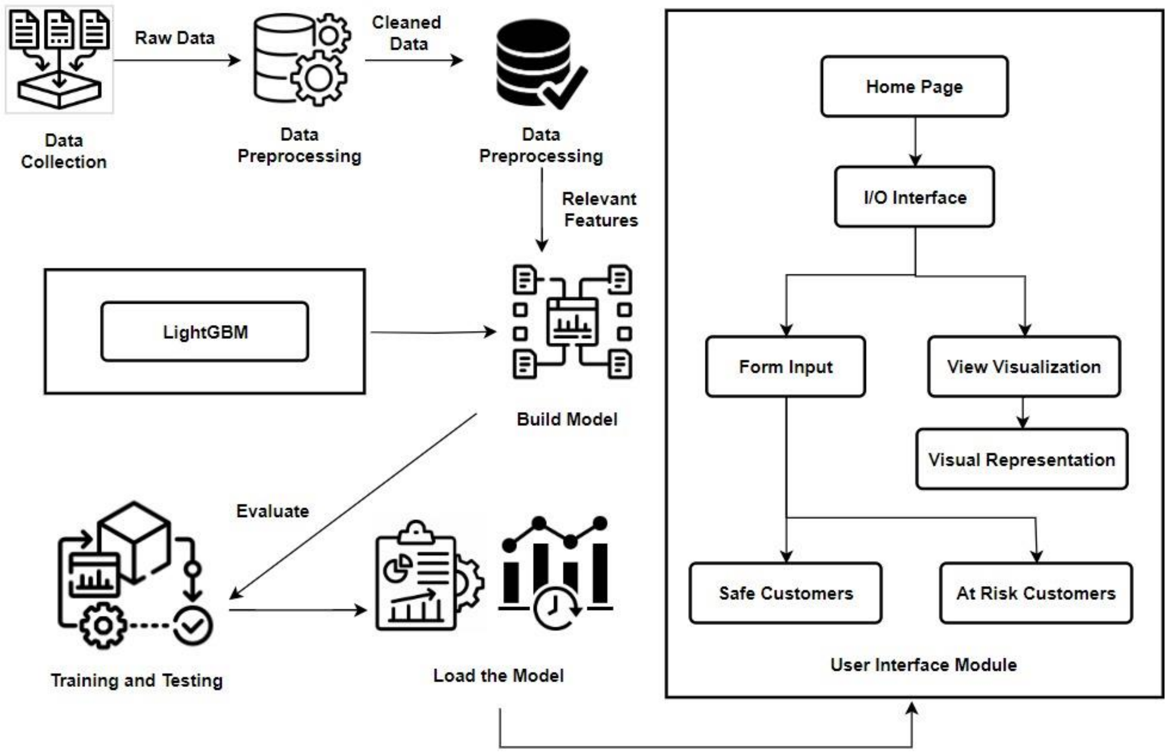

Fig.1. Architecture Diagram

The customer churn prediction process involves a comprehensive architecture that seamlessly integrates various components to derive meaningful insights. Initially, raw data, encompassing customer information and interactions, is collected. Following this, the dataset undergoes meticulous preprocessing, including noise removal and feature selection, to refine the data for further analysis. Subsequently, a Temporal Convolution Neural Network is employed, featuring Convolutional Layers for spatial feature extraction, MaxPooling Layers for down-sampling, and Flatten Layers to prepare the data for subsequent processing. Additionally, the hybrid model LSTM + CNN introduces a concatenation layer, batch normalization layer, activation layer, and dense layer, combining the strengths of both LSTM and CNN components. This intricate process flow ensures the effective extraction and utilization of relevant features from the data. Another pivotal component is LightGBM, a gradient boosting framework configured with parameters such as the number of estimators, objective function, and learning rate. LightGBM processes the data and generates predictions. The architecture diagram also involves a Wide and Deep Neural Network, integrating features processed by both LSTM-CNN and LightGBM. This architecture employs dense layers for intricate pattern recognition, aiming to harness the strengths of both shallow and deep learning components. The evaluation of the model's performance, utilizing metrics such as accuracy, precision, recall, and F1-score, ensures robust predictive capabilities before deploying the model for practical use. The final predicted output categorizes customers into two groups: those likely to have a positive experience and customers at high risk, requiring attention due to potential issues.

3. Proposed Work

The proposed approach for customer churn prediction using LightGBM offers a comprehensive and well-structured methodology that balances predictive accuracy with efficiency. By integrating the advanced capabilities of LightGBM, a gradient-boosting framework, the approach aims to deliver exceptional results in identifying potential customer churn. Leveraging LightGBM's efficiency in handling large datasets and categorical features, the model can capture complex patterns in customer behavior, making it well-suited for the dynamic nature of churn prediction in various industries [5][7]. Despite the sophistication of the underlying machine learning algorithm, the approach maintains accessibility through a user-friendly interface and clear instructions, ensuring that businesses of varying technical expertise can benefit from its predictive power. This technique balances efficiency and precision, making it appropriate for practical uses in customer relationship management.

The proposed approach begins by allowing users to input relevant customer data, such as transaction history, customer interactions, and demographics. Once the input data is identified, the methodology proceeds to a preprocessing stage where feature scaling and format adaptation ensure that the data is optimized for subsequent operations. The core steps of the prediction process include feature engineering, model training with LightGBM, and hyperparameter tuning. Feature engineering involves selecting and transforming relevant customer features to enhance the model's ability to capture patterns indicative of churn. LightGBM's gradient boosting algorithm is then employed to iteratively build decision trees and optimize predictive performance. Hyperparameter tuning allows users to customize the model according to their business requirements, providing flexibility and control over the predictive process.

Postprocessing steps involve evaluating model performance, interpreting feature importance, and generating predictions. This methodology ensures a high level of adaptability across various industries, allowing businesses to effectively identify and mitigate customer churn. The user-friendly interface and comprehensive instructions enhance accessibility and usability, making the approach suitable for a broad audience within the business and data analytics domains. The dedication to facilitating user understanding and interaction reflects a thoughtful integration of advanced machine learning techniques, showcasing the effectiveness and automatic functionality of the proposed customer churn prediction approach using LightGBM.

4. System Process Flow:

Telecom Customer Churn prediction is a comprehensive procedure that entails a structured approach to acquiring, cleaning, and preparing the dataset to facilitate analysis and model construction. This process is crucial for developing accurate predictive models that can effectively identify and address customer churn, thereby enabling telecom companies to implement targeted retention strategies and mitigate revenue loss.

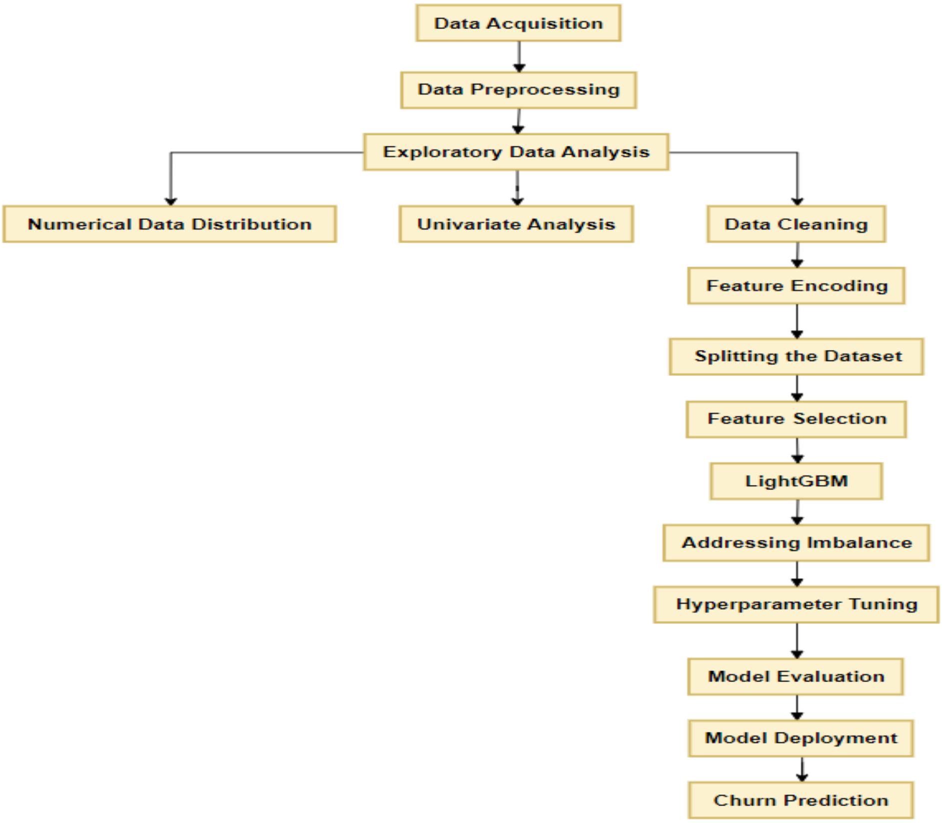

Fig.2. Process Flow

Data acquisition in telecom customer churn prediction is a critical initial step aimed at assembling comprehensive data encompassing various aspects of customer interactions and characteristics. This includes customer attributes, service subscriptions, account details, and demographic information. The overarching goal is to gather a holistic dataset that enables the prediction of customer behavior, specifically focusing on churn, to devise effective retention strategies. Preprocessing of this data involves several key stages to ensure its suitability for analysis and modeling. One fundamental aspect is the segregation of features into numerical and categorical categories, allowing for tailored preprocessing strategies for each type. This division facilitates focused data cleaning and transformation efforts, optimizing the dataset for subsequent analysis. Exploratory Data Analysis (EDA) emerges as a crucial phase during preprocessing, providing insights into the distribution and characteristics of the data. Through analysis of churn rates, it becomes evident that approximately 26.5% of customers have churned, signaling an imbalance in the dataset. Additionally, an estimated financial loss of $2,862,927 attributed to churn underscores the significance of addressing this imbalance for effective model training and prediction. Examination of the numerical data distribution reveals no outliers, indicating a relatively clean dataset suitable for further analysis without the need for outlier removal, streamlining the preprocessing pipeline. Further preprocessing steps include univariate analysis, which delves into the behavior and distribution of individual features, providing valuable insights for subsequent modeling efforts. The absence of missing values in the dataset alleviates the need for explicit handling of missing data, allowing the focus to shift towards feature encodingThe dataset is then split into training and testing sets to help with the model's evaluation and validation. The most pertinent features for model training are then found using feature selection approaches. The top 10 correlated features, particularly those with a significant association with the target variable (Churn), are selected for further analysis and modeling. Various machine learning models, including LSTM-CNN, LightGBM, and TCN, are fitted to the dataset to explore their performance in predicting customer churn. Imbalance in the dataset is addressed through techniques like SMOTEENN, which combines over-sampling and down-sampling with edited nearest neighbors to balance the dataset effectively. Hyperparameter tuning, particularly for the LightGBM model, further enhances model performance. Evaluation metrics such as accuracy, precision, recall, and F1-score are utilized to assess model effectiveness, with LightGBM demonstrating superior performance. Finally, the trained LightGBM model is saved using the pickle library for future deployment in churn prediction tasks, ensuring a structured approach to data acquisition and preprocessing that facilitates accurate predictions and informs effective customer retention strategies.

5. Experimental Result

The evaluation of test outcomes in predicting customer attrition surfaces unique performance patterns among the different models used in this research. Especially recognizable is the superior performance of the LightGBM model, which manages to attain an admirable accuracy level of 95%. This remarkable singular precision plays a significant role in attaining a

total accuracy of 96%. Conversely, the Temporal Convolutional Neural Network (TCNN) manifests a resilient performance, achieving a 92% accuracy rate that trails closely behind the LightGBM model. Demonstrating its capability to decipher temporal and spatial patterns within the customer's data, the LSTM-CNN combined model delivers a respectable accuracy of 90%. In comparison, the Wide and Deep Neural Network attains a relatively lower accuracy rate of just 68%. These results underscore the variability in model performance, with LightGBM leading in predictive accuracy and the Wide and Deep Neural Network exhibiting a lower accuracy rate.

Dataset	**Model**	**Accuracy**	**Precision**	**Recall**	**F1-Score**
Telco-Customer-Churn.csv	LightGBM	95	96	96	96
	Temporal Convolutional Neural Network	93	95	95	95
	LSTM - CNN	89	86	99	92
	Wide and Deep Neural Network	67	66	82	80
telecom_customer_churn.csv	LightGBM	93	93	96	94
	Temporal Convolutional Neural Network	91	89	92	91
	LSTM - CNN	88	87	90	88
	Wide and Deep Neural Network	66	64	69	66
WA_Fn-UseC_-Telco-Customer-Churn.csv	LightGBM	92	90	90	91
	Temporal Convolutional Neural Network	90	89	88	88
	LSTM - CNN	87	86	87	86
	Wide and Deep Neural Network	64	65	65	65

Table.1. Comparison Result

Precision

$$Precision = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}}$$

Recall

$$\text{Recall} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}}$$

F1 – Score

$$\text{F1} - \text{Score} = 2 \text{ x } \frac{\text{Precision} * \text{Recall}}{Precision + Recall}$$

In the Fig 3, In the context of customer churn prediction, the model is trying to predict whether a customer is likely to churn (cancel their service) or not. The rows of the confusion matrix represent the actual churn labels, and the columns represent the predicted churn labels. So, for example, the cell in the top-left corner of the matrix shows the number of customers who actually did not churn (the actual churn label is "no churn") and were also predicted not to churn (the predicted churn label is "no churn"). The diagonal cells of the confusion matrix show the number of correct predictions. In the image you sent, the diagonal cells are all colored light blue, which suggests that the model is making a good number of correct predictions. However, the off-diagonal cells are also not empty, which means the model is also making some incorrect predictions.

In the Fig 4, Customers with a tenure of less than 50 months have a higher churn rate than customers with a tenure of more than 50 months. Customers with higher monthly charges tend to have a lower churn rate than customers with lower monthly charges. Customers with very high TotalCharges might be less likely to churn because they're heavily invested in the service. On the other hand, very high charges due to a short tenure could indicate dissatisfaction and a higher churn risk which leads to correct prediction.

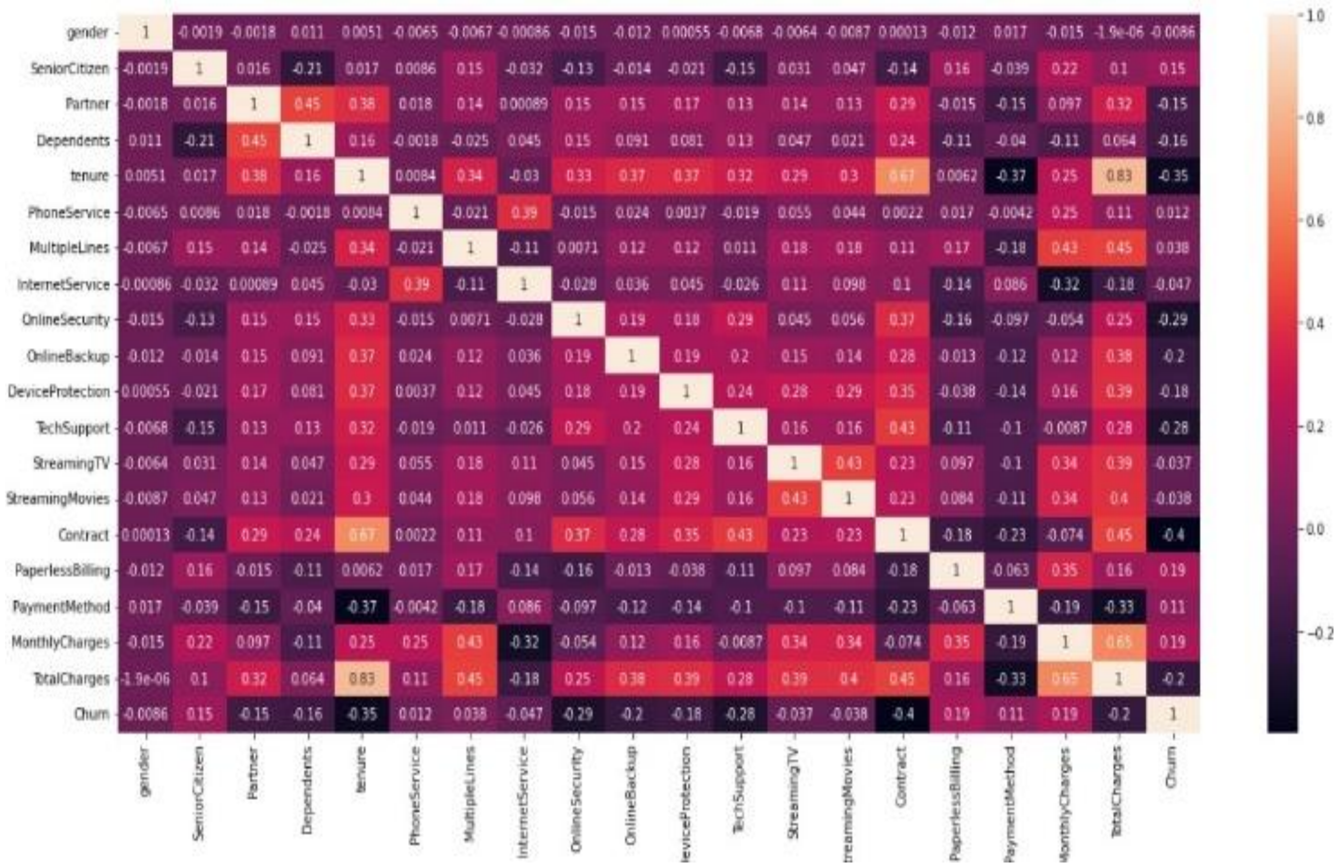

Fig.3. Confusion Matrix

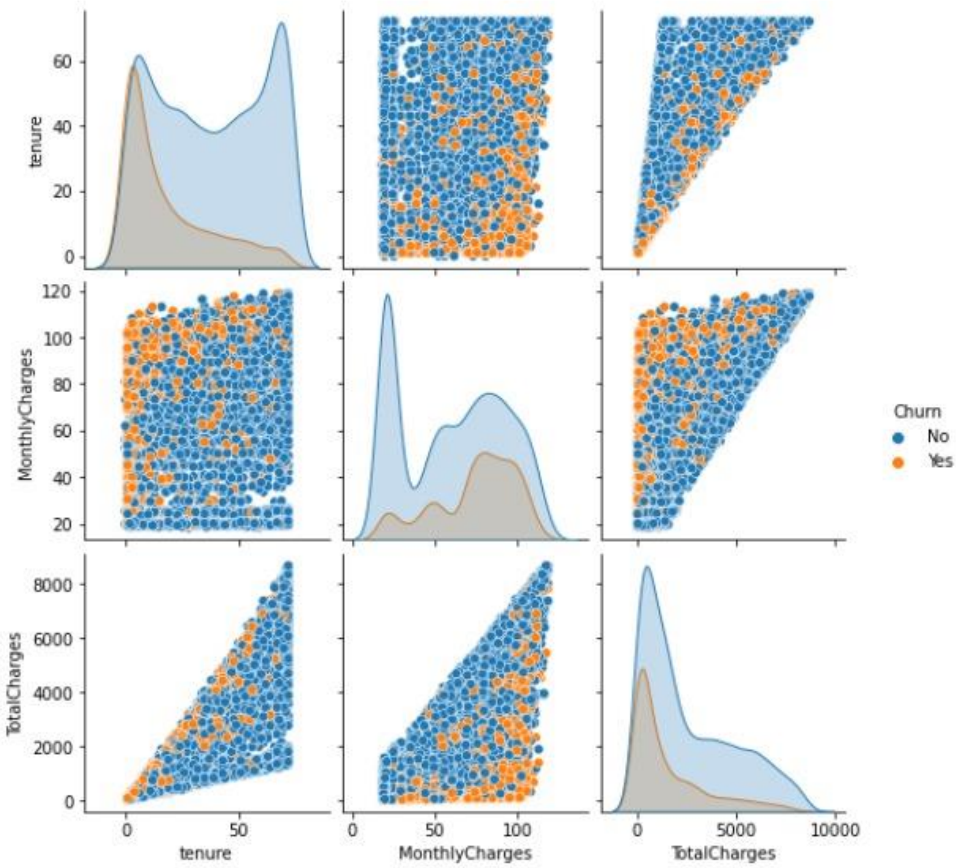

Fig.4. Scatter Plot

6. Output

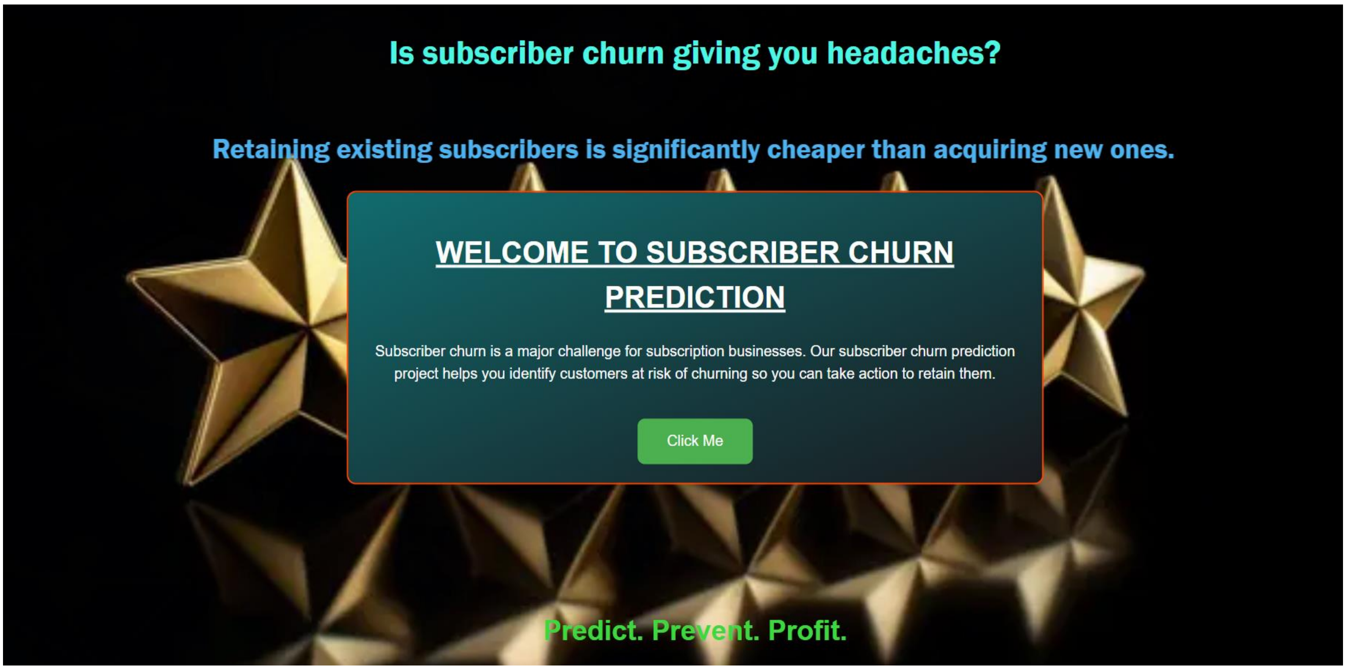

Fig.5. Home Page

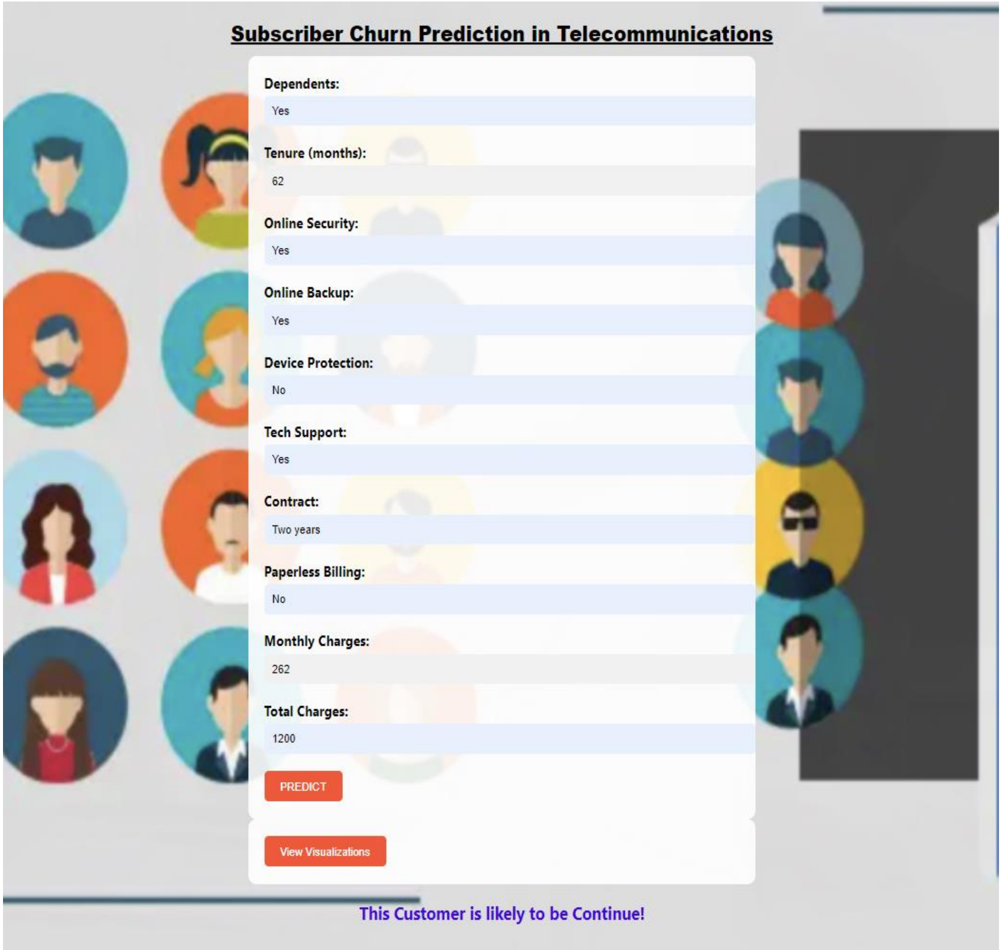

Fig.6. User Interface

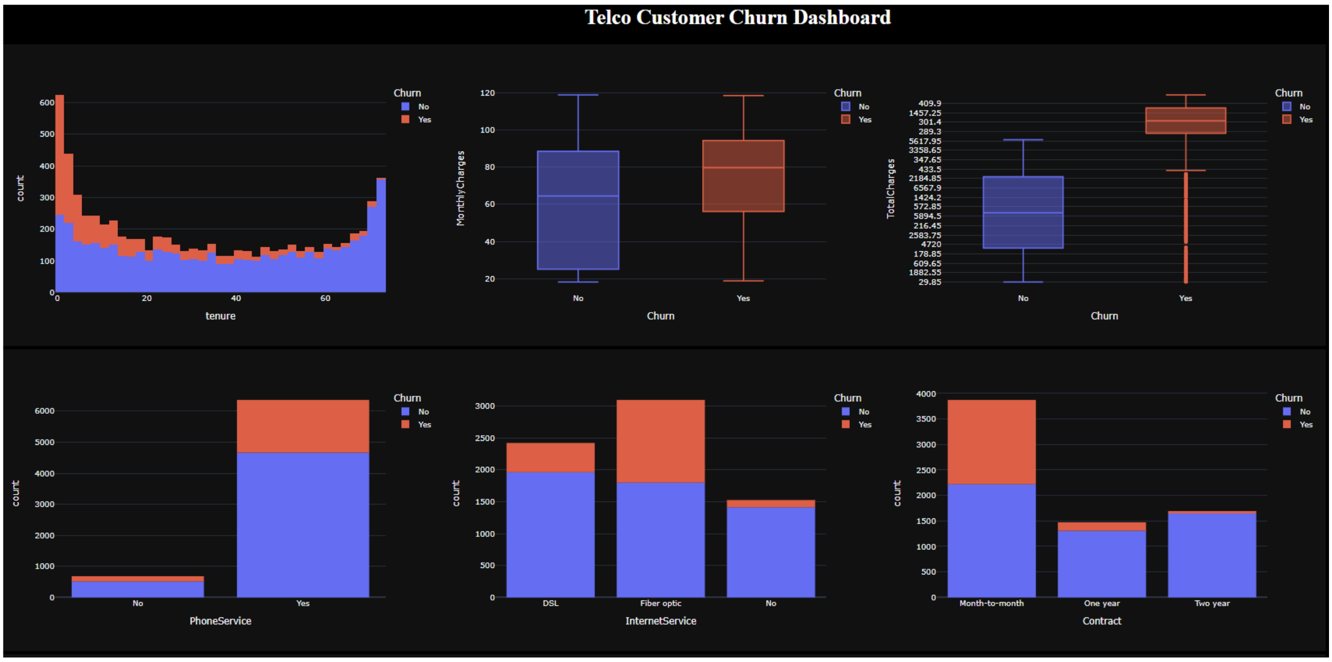

Fig.7. Data Visualizations

7. Conclusion

In conclusion, various deep learning and machine learning models have been effectively used in this work. When predicting customer attrition in the telecom industry, LightGBM is performing better than the others. The use of sophisticated models, such as Temporal Convolutional Neural Network (CNN), LSTM-CNN, Wide and Deep Neural Networks, and others, has made it possible to comprehend and predict customer churn dynamics completely. These models have shown differing degrees of success in capturing various aspects of customer behavior and temporal relationships present in sequential data through the comprehensive examination of historical customer data. In particular, the application of LightGBM has produced outstanding outcomes, demonstrating its ability to manage big datasets and accurately identify patterns suggestive of possible churn. The study emphasizes the need for feature engineering, careful model evaluation, and hyperparameter adjustment for maximizing predictive performance across different models.

8. Future Work

Real-Time Prediction: Enhance the system to support real-time churn prediction capabilities, allowing businesses to receive immediate insights into customer behavior and take proactive retention actions. Automated Reporting: Create RPA bots to generate automated reports on model performance and business KPIs. These reports can be scheduled to run at regular intervals and provide stakeholders with up-to-date insights into the effectiveness of the churn prediction system.

References

1. Aayushi Gandhi, Neha Katre, Parth Meswani, Pushkar Bhuse, Riya Muni. (2020) ‘Machine Learning Based Telecom-Customer Churn Prediction’, 3rd International Conference on Intelligent Sustainable Systems (ICISS), pp. 1297-1301.

2. Abdelrahim Kasem Ahmad, Assef Jafar, Kadan Aljoumaa. (2019) ‘Customer Churn Prediction in Telecom using Machine Learning in Big Data Platform’, Published as a research paper, but no specific journal mentioned in the provided document, 6(1), 1-24.

3. Abhishek Gaur, Ratnesh Dubey. (2018) ‘Predicting Customer Churn Prediction in Telecom Sector Using Various Machine Learning Techniques’, Published in “IEEE Access” Journal, pp. 1-5.

4. Aditya Das, Amit Gaikwad, Sanket Agrawal, Sudhir Dhage. (2018) ‘Customer Churn Prediction Modelling Based on Behavioural Patterns Analysis using Deep Learning’, International Conference on Smart Computing and Electronic Enterprise, pp. 1-6.

5. Ammar A Ahmed, Dr. D. Maheswari linen. (2018) ‘Customer Churn Prediction in the Telecommunication Industry: A Systematic Review and Meta-Analysis’, International Journal of Database Theory and Application, 45(2), 111-147.

6. Amol Chole, Rushikesh Chavan, Shradha Mandan, Trupti Mate. (2023) ‘Customers Churn Prediction using Machine Learning and Deep Learning’, Journal of Emerging Technologies and Innovative Research, ISSN-2349-5162.

7. Amril Syalim, Ketut Gde Manik Karvana, Petrus Mursanto, Setiadi Yazid. (2019) ‘Customer Churn Analysis and Prediction Using Data Mining Models in Banking Industry’, International Workshop on Big Data and Information Security, pp. 33-38

8. Anvita Gupta, Komanduri Srikar, Sreenitya Mandav. (2022) ‘Customer Churn Prediction in Banking Sector - A Hybrid Approach’, International Journal for Research in Applied Science & Engineering Technology (IJRASET), pp. 1-6.

9. Chin-Kuan Ho, Maw Maw, Su-Cheng Haw. (2019) ‘Customer Churn Prediction in Telecommunication: An Analysis on Issues, Techniques and Future Trends’, Published in "IEEE Access" journal, pp. 148-153.

10. Dang Xing – hua, Zhao Jing. (2008) ‘Bank Customer Churn Prediction Based on Support

Vector Machine: Taking a Commercial Bank's VIP Customer Churn as the Example', Published in "2008 IEEE" journal, pp. 1-4.

11. Dr. ALN Rao, Dr. Durgaprasad Gangodkar, Shwetkranti Taware, V K Srivastav, Suman Avdhesh Yadav, Varsha Agarwal. (2022) 'Customer - Churn Prediction Using Machine Learning', Expert Systems with Applications and IEEE Transactions on Industrial Informatics, pp. 893-899.

12. Dr. E. N. Ganesh, R. Sudharsan. (2022) 'Swish RNN Based Customer Churn Prediction in Telecommunication Industry', Published by Informa UKLimited, Trading as Taylor and Francis Group, 34(1), 1855-1876.

13. Jiang Yang, Jinlong Hu, Lei Lei, Minjie Huang, Runchao Zhu, Shoubin Dong, Yi Zhuang. (2018) 'pRNN: A Recurrent Neural Network based Approach for Customer Churn Prediction in Telecommunication Sector', 2018 IEEE International Conference on Big Data, pp. 4081-4085.

14. Kumar, P., Kumar, S.V. (2023). DDoS Attack Prediction System Using Machine Learning Algorithms. In: Tuba, M., Akashe, S., Joshi, A. (eds) ICT Systems and Sustainability. ICT4SD 2023. Lecture Notes in Networks and Systems, vol 765. Springer, Singapore.

15. Kumar, P., Vinodh Kumar, S., Priya, L. (2023). An Approach for Energy-Efficient Resource Allocation Through Early Planning of Virtual Machines to Servers in Cloud Server Farms. In: Doriya, R., Soni, B., Shukla, A., Gao, XZ. (eds) Machine Learning, Image Processing, Network Security and Data Sciences. Lecture Notes in Electrical Engineering, vol 946. Springer, Singapore.

16. Liu Juan, Luo Bin, Shao Peiji. (2007) 'Comparing to Techniques used in Customer Churn Analysis', Published in "2007 IEEE" Journal, 4(1), 30-38.

17. Moaiad Ahmad Khder, Samah Wael Fujo, Suresh Subramanian. (2022) 'Customer Churn Prediction in Telecommunication Industry Using Deep Learning', IEEE Access, 11(1), 24.

18. Mohammad Ridwan Ismail, Mohd Khalid Awang, Mokhairi Makhtar, M Nordin A Rahman. (2015) 'A Multi-Layer Perceptron Approach for Customer Churn Prediction', International Journal of Multimedia and Ubiquitous Engineering, 10(7), 213-222.

19. Mr. Anurag Bhatnagar, Dr. Sumit Srivastava. (2019) 'A Robust Model for Churn Prediction using Supervised Machine Learning', 9th International Conference on Advanced Computing (IACC), pp. 45-49.

20. Ozer Celik, Usame O. Osmanoglu. (2019) 'Comparing to Techniques used in Customer Churn Analysis', Journal of Targeting, Measurement, and Analysis for Marketing, 4(1), 30-38.

21. Ricardo F. Ramos, Sérgio Moro, Tianyuan Zhang. (2022) 'Telecom Customer Churn Prediction Using Hybrid Models', IEEE Access, 14(3):94.

22. V Kumar, Manas Rahman. (2020) 'Machine Learning Based Customer Churn Prediction In Banking', Procedia Computer Science, pp. 1196-1201.