Machine Learning Model for Data Analysis

(Customer Churn Prediction using CNN, LSTM, and BiLSTM)

Abstract

Customer churn prediction plays a crucial role in retaining customers and improving business profitability, especially in sectors such as banking, telecommunications, and insurance. This project focuses on applying deep learning techniques to predict customer churn using the Churn Modelling dataset. Three neural network architectures—Convolutional Neural Network (CNN), Long Short-Term Memory (LSTM), and Bidirectional LSTM (BiLSTM)—are implemented and compared based on performance metrics such as accuracy, precision, recall, and F1-score. The results indicate that while CNN and LSTM offer competitive performance, the BiLSTM model achieves the highest accuracy and precision. Additionally, an ensemble approach is discussed to enhance prediction robustness. This study highlights the effectiveness of deep learning models in capturing complex customer behaviour patterns and supporting data-driven retention strategies.

---

1. Introduction

Customer churn refers to the loss of customers when they discontinue a company's services. Predicting churn is essential for organizations because acquiring new customers is significantly more expensive than retaining existing ones. Traditional statistical methods often fail to capture complex, nonlinear relationships present in modern customer datasets.

With the advancement of machine learning and deep learning, more accurate churn prediction models have emerged. Deep learning models such as CNN, LSTM, and BiLSTM can automatically learn hidden patterns from large datasets and handle sequential dependencies effectively.

The objective of this project is to design and compare three deep learning models—CNN, LSTM, and BiLSTM—for predicting customer churn using a real-world banking dataset. The study evaluates these models based on predictive accuracy, precision, recall, and computational efficiency to identify the most suitable approach for churn prediction systems.

---

2. Dataset and Data Preprocessing

2.1 Dataset Description

The project uses the **Churn Modelling dataset**, which contains approximately **10,000 customer records** with demographic, financial, and behavioural attributes. Key features include customer age, gender, credit score, balance, tenure, and transaction behaviour. The target variable is binary, indicating whether a customer has exited (churned) or not.

2.2 Data Preprocessing

To ensure high-quality input data, several preprocessing steps were applied:

- **Handling missing values** and checking data consistency
- **Encoding categorical variables** using Label Encoding and One-Hot Encoding
- **Feature scaling** using Min-Max normalization to improve training stability

- **Class imbalance handling**, where necessary, using resampling techniques

The dataset was then split into training, validation, and test sets to evaluate model generalization.

---

3. Methodology and Model Architecture

3.1 Convolutional Neural Network (CNN)

Although CNNs are traditionally used for image data, they are effective for structured and sequential data through 1D convolutions. In this project, CNN is used to capture local feature interactions between customer attributes. The CNN model is computationally efficient and achieves faster training times.

3.2 Long Short-Term Memory (LSTM)

LSTM networks are designed to model long-term dependencies in sequential data. They are well-suited for capturing temporal behavior patterns in customer activity. The LSTM model uses memory cells and gating mechanisms to retain relevant information across sequences.

3.3 Bidirectional LSTM (BiLSTM)

BiLSTM extends LSTM by processing data in both forward and backward directions. This allows the model to capture both past and future contextual information, improving prediction accuracy. In churn prediction, BiLSTM is particularly useful for understanding complex behavioral patterns.

All models were trained using the same preprocessed dataset to ensure fair comparison.

---

4. Results and Discussion

4.1 Model Performance

The performance of CNN, LSTM, and BiLSTM models was evaluated using accuracy, precision, recall, F1-score, and training time.

- **CNN** achieved good recall and faster training, making it suitable for time-sensitive applications.
- **LSTM** provided balanced performance across all metrics and handled sequential dependencies effectively.
- **BiLSTM** achieved the **highest accuracy (86.20%) and precision**, indicating superior predictive capability.

However, BiLSTM required more training time due to its bidirectional structure. The results demonstrate that model selection should depend on business priorities such as precision, recall, or computational efficiency.

4.2 Ensemble Learning

To further improve performance, an ensemble approach using weighted soft voting was discussed. By combining CNN, LSTM, and BiLSTM, the ensemble model leveraged the strengths of all architectures, resulting in improved robustness and overall prediction accuracy.

5. Conclusion

This project successfully implemented and compared CNN, LSTM, and BiLSTM models for customer churn prediction. Each model demonstrated unique strengths: CNN offered speed and recall, LSTM provided balanced performance, and BiLSTM delivered the highest accuracy and precision. The ensemble approach further enhanced prediction reliability by combining multiple models.

The findings confirm that deep learning techniques are effective for churn prediction and can significantly support proactive customer retention strategies. Future work may include incorporating additional behavioural features, applying advanced ensemble techniques, and deploying the model in real-time business environments.