Predicting Customer Churn for Subscription Services

Jackson Baxter

Goal:

Creating a deep learning model that forecasts user attrition for subscription-based businesses is the aim of this research. Customer churn is the term used to describe when consumers discontinue using a service. Businesses can use churn prediction to put retention tactics into place, enhance customer satisfaction, and minimize revenue loss. The model will use past behavioral data, interaction history, and subscription habits to determine whether or not a customer is likely to churn.

Approach:

In order to do this, the project will either create a simulated dataset that represents customer behaviors or use an existing public customer turnover dataset. Features including account age, usage frequency, customer service interactions, payment history, and subscription renewal habits would all be included in the data. The project will concentrate on developing a deep neural network (DNN) to forecast if a customer is likely to churn within a specified time frame after data cleaning and preprocessing.

Key Steps:

1. **Data Preprocessing:** Handle missing values, normalize numerical characteristics, and encode categorical variables (such as subscription type and payment method) to clean up and preprocess the dataset. To identify useful patterns from client data, including the duration of the subscription and the time since the previous interaction, feature engineering will be done.
2. **Model Development:** A DNN model will be implemented, using architectures such as fully connected layers with ReLU activations and dropout layers to reduce overfitting. The model will be trained with binary cross-entropy loss, optimizing for binary classification.
3. **Training and Optimization:** The Adam optimizer will be used to train the model and adjust hyperparameters such as layer depth, learning rate, and batch size. To prevent overfitting, strategies like early halting and regularization will be employed.
4. **Evaluation:** The model's capacity to recognize churners accurately will be assessed using accuracy, precision, recall, and F1-score. In addition, the model performance will be evaluated across several thresholds using the ROC curve and area under the curve (AUC).

Measures of Success:

The following measures will be used to determine whether the project is successful:

Accuracy: On the test set, the model's ability to predict customer attrition should be at least 85% accurate.

Precision and Recall: In order to ensure a high true positive rate (which captures churners) and a low false positive rate (which minimizes incorrect churn predictions), the model should have precision and recall values of at least 0.80.

AUC: A higher than 0.85 target AUC indicates good performance in differentiating between churners and non-churners.

Business Impact: The model should yield useful information that can guide retention tactics, such as identifying clients who are likely to leave within the next 30 days.

Resources:

Software: The PyTorch framework will be used to build and train the model, with libraries like pandas for data preprocessing and scikit-learn for evaluation metrics.

Dataset: A publicly available customer churn dataset (e.g., the Telco Customer Churn dataset) will be used, or a similiar dataset will be used to capture typical subscription behaviors and patterns.