Ethan Pirso

260863065

MGSC 673

April 14, 2024

Analyzing Bank Customer Churn Using Feed-Forward Neural Networks

In the project documented in the Jupyter notebook titled "Bank_Customer_Churn_Prediction," the primary objective was to harness the capabilities of feed-forward neural networks to predict customer churn based on a comprehensive dataset gathered from banking customers. This project is situated within the broader context of customer relationship management, where understanding and predicting customer churn—defined as the phenomenon where customers discontinue their relationship with a business—is crucial for maintaining financial stability and fostering customer loyalty.

Churn prediction is particularly important in the banking sector because acquiring new customers can be significantly more costly than retaining existing ones. By accurately predicting which customers are likely to leave, banks can implement targeted intervention strategies aimed at improving customer satisfaction and retention rates, thereby mitigating financial losses and enhancing customer engagement.

Feed-forward neural networks are particularly well-suited for this task due to their ability to model complex non-linear relationships and interactions among variables. These networks propagate data forward from the input to the output layer (possibly passing through multiple hidden layers) and are capable of learning from large amounts of data through their adaptable structure of neurons and synapses, which adjust as the model learns from the data presented.

Dataset and Preprocessing

The dataset utilized in this study, referred to as "Customer-Churn-Records.csv," comprises a rich collection of features that are instrumental in predicting customer churn. These features include demographic, account, and service-level attributes that provide insights into customer behavior and preferences, which are vital for understanding the determinants of customer retention or churn.

Features Overview:

1. Demographics and Account Information:

- Credit Score: Ranges from 350 to 850, impacting the customer's likelihood of churn. Higher credit scores typically indicate more financial stability and possibly lower churn rates.

- Geography: Includes three unique country data points, which could affect churn due to varying economic conditions and service experiences in different regions.

- Gender: Distributed as 55% male and 45% female in the dataset, potentially influencing churn patterns.

- Age: Ranging from 18 to 92 years, with a concentration of customers between 32 and 47 years, highlighting significant age-related banking needs and churn risks.

- Tenure: Shows the number of years customers have been with the bank, ranging from 0 to 10 years, which correlates with loyalty and churn rates.

2. Financial Behaviors and Products:

- Balance: Varies widely, with some customers holding balances as high as 250,898.09, indicating varying levels of engagement and financial activity, which are critical for predicting churn.

- Number of Products: Customers own between 1 and 4 products, where the diversity in product holdings could suggest different levels of satisfaction or engagement.

- Has Credit Card: Reveals if customers possess a bank-issued credit card, a factor that often ties a customer more closely to the bank.

- Is Active Member: Active membership status can be a strong indicator of customer engagement and potentially lower churn.

3. Feedback and Interaction Metrics:

- Complain: Indicates whether a customer has registered complaints, with around 20% having done so, directly affecting their likelihood to churn.

- Satisfaction Score: Ranges from 1 to 5, directly derived from customer feedback on complaint handling, reflecting overall customer satisfaction and propensity to churn.

- Card Type and Points Earned: Various card types (e.g., DIAMOND, GOLD) and points earned through transactions can also impact customer retention, as these elements reflect usage intensity and reward engagement.

Preprocessing Steps:

The initial preprocessing involved discarding non-informative variables such as 'RowNumber,' 'CustomerId,' and 'Surname.' These fields were considered noise since they do not contribute to the analysis. The remaining variables were subjected to further preprocessing to prepare the dataset for modeling:

- Encoding Categorical Variables: Transforming categorical data into a format that can be easily integrated into predictive models. This includes creating dummy variables for categorical attributes like geography and gender.

- Feature Scaling: Standardizing features to a uniform scale is crucial in neural network models to ensure that no single feature dominates the learning process due to its scale.

By systematically organizing and preprocessing these data, the study establishes a robust foundation for applying feed-forward neural networks to accurately predict customer churn, leveraging the intrinsic patterns and correlations present within the dataset. This comprehensive data description provides the necessary context for understanding the subsequent modeling steps and their implications in the predictive analysis of customer churn.

Model Construction and Training

In this project, three distinct architectures of feed-forward neural networks were meticulously designed to tackle the problem of predicting customer churn, each progressively more complex to capture different aspects and relationships within the data. Below is a detailed description of the models' architectures and their respective training methodologies:

1. SimpleNN1 (`model1`): This is the most basic model in the lineup, featuring a single linear layer. It was designed to serve as a baseline for performance comparison. The architecture is straightforward, with:

   - A fully connected layer (`fc`) that maps the input features directly to two outputs, corresponding to the churn classes (churned or not churned).

   - The model's simplicity allows for rapid training and testing of initial hypotheses about the data.

2. SimpleNN2 (`model2`): An enhanced model that includes three hidden layers and utilizes the ReLU activation function to introduce non-linearity, enabling the model to learn more complex patterns in the data. The layers are configured as follows:

- First hidden layer with 64 units.

- Second hidden layer with 32 units, following the first.

- A final linear layer that outputs to the two churn classes.

- ReLU activations are applied after the first and second hidden layers to add non-linearity.

3. ComplexNN (`model3`): This model is the most advanced among the three, incorporating four hidden layers and dropout regularization to prevent overfitting—a common challenge in deep learning models. The specifics of this architecture include:

- Four hidden layers with decreasing units: 128, 64, 32, and 16, respectively.

- A dropout layer with a drop rate of 50% after each of the first three ReLU activations to reduce overfitting by randomly omitting units during training.

- The final output layer again maps to the two churn classes.

Each model was trained using the Adam optimizer with a learning rate of 0.001 across 40 epochs to optimize the cross-entropy loss. The dataset was split into training, validation, and testing sets using a stratified approach to ensure that each set is representative of the overall dataset. Specifically, the data was first split into a training set (70%) and a temporary set (30%). The temporary set was then equally split into validation and test sets. This partitioning allows for effective training and hyperparameter tuning on the training set, performance tuning and early stopping using the validation set, and unbiased evaluation on the test set (training and validation loss curves shown in Figure 1).

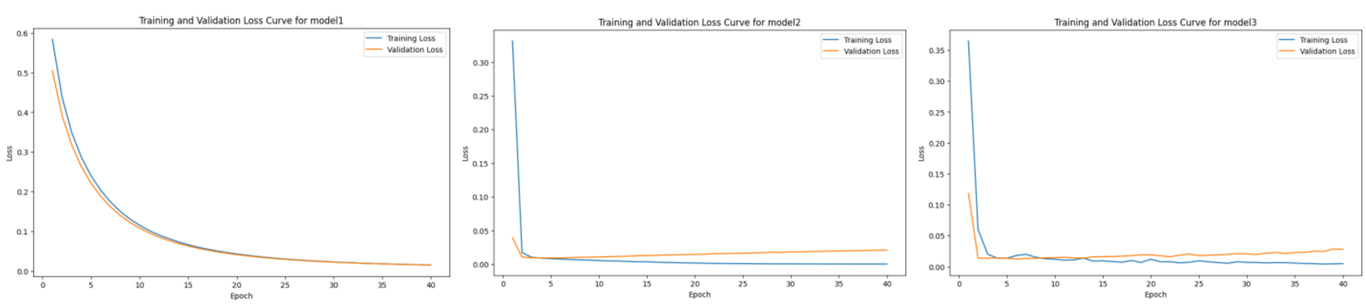

Figure 1: The training and validation loss curves for each of the model architectures described above.

These models provide a structured approach to understanding and predicting customer behavior, each adding a layer of complexity and robustness to the analysis, allowing for in-depth learning from the data provided.

Evaluation and Metrics

The models were evaluated based on accuracy, precision, recall, and F1 score using a separate test dataset to ensure unbiased evaluation. The results demonstrated very high performance across all models, with slight variations (Figure 2). For instance, Model 1 achieved an accuracy of approximately 99.87%, with similar precision and F1 scores, as shown in the confusion matrices (Figure 3). These matrices provided a detailed view of true versus predicted classifications, affirming the models' robustness in accurately predicting churn.

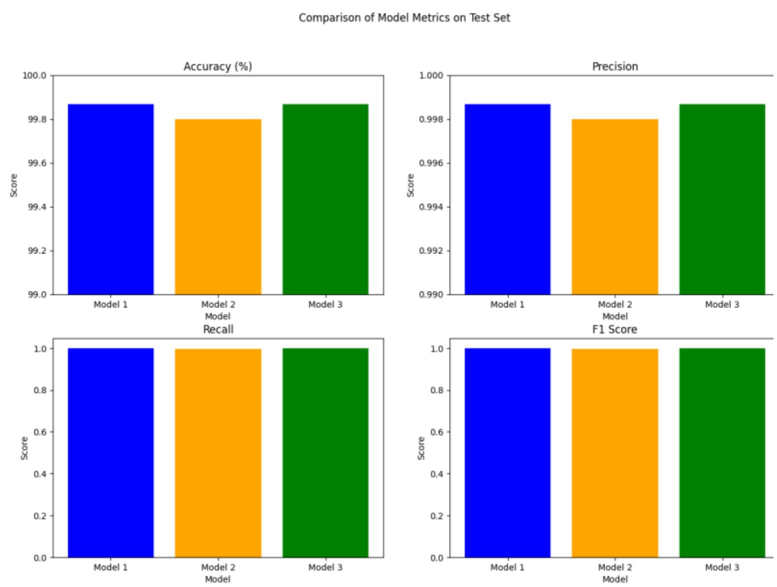

Figure 2: Accuracy, Precision, Recall, and F1 scores.

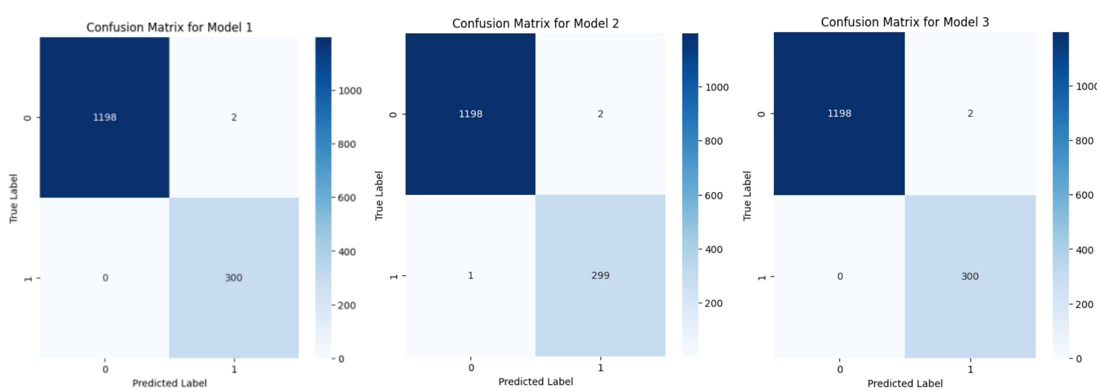

Figure 3: Confusion Matrices.

Discussion

The use of different architectures and the integration of dropout in the ComplexNN model underscored the effects of network depth and regularization on performance. Notably, the dropout layers helped manage overfitting effectively, as evidenced by the consistent validation loss alongside the training loss. The results across different metrics confirmed that even the simplest model performed remarkably well, which might be attributed to the dataset's feature-rich nature, allowing even basic models to achieve high accuracy.

In conclusion, this exercise not only highlighted the capabilities of neural networks in handling predictive analytics tasks such as churn prediction but also showcased the importance of proper data preprocessing and the impact of architectural choices on the model's performance. The detailed figures included in the report, like the loss curves and confusion matrices, provide a clear visual representation of the models' training dynamics and their capabilities in accurately classifying the churn predictions.