Bank Churn Prediction Business Presentation

Name: Arjun Dubewar

Date: 11/22/2024

Contents / Agenda

- Business Problem Overview and Solution Approach
- EDA Results
- Data Preprocessing
- Model Performance Summary
- Executive Summary

Business Problem Overview and Solution Approach

Banks face a significant challenge of customer **churn**, where customers leave to join competing service providers. This not only impacts revenue but also increases acquisition costs for new customers. It is important to understand which aspects of the service influence a customer's decision in this regard. Understanding which factors drive customer decisions to leave is critical for management to prioritize improvements in service quality and customer retention strategies. Our approach to solving this problem is to develop a **neural network-based classifier** to predict whether a bank customer is likely to churn within the next six months, enabling proactive interventions to retain valuable customers.

EDA Results

The dataset that we utilized in order to create our models contained 10,000 datapoints and contained data, on the 13 features noted below, on each customer. Our analysis also indicated that there were no missing values in the dataset and the unique entries in each category are also listed below:

CustomerId: Unique ID which is assigned to each customer

Surname: Last name of the customer

CreditScore: It defines the credit history of the customer.

Geography: A customer's location

Gender: It defines the Gender of the customer

Age: Age of the customer

Tenure: Number of years for which the customer has been with the bank

NumOfProducts: It refers to the number of products that a customer has purchased through the bank.

Balance: Account balance

HasCrCard: It is a categorical variable that decides whether the customer has a credit card or not.

EstimatedSalary: Estimated salary

isActiveMember: It is a categorical variable that decides whether the customer is an active member of the bank or not (Active member in the sense, using bank products regularly, making transactions, etc)

Exited: It is a categorical variable that decides whether the customer left the bank within six months or not. It can take two values

0=No (Customer did not leave the bank)

1=Yes (Customer left the bank)

RowNumber	0
CustomerId	0
Surname	0
CreditScore	0
Geography	0
Gender	0
Age	0
Tenure	0
Balance	0
NumOfProducts	0
HasCrCard	0
IsActiveMember	0
EstimatedSalary	0
Exited	0

RowNumber	10000
CustomerId	10000
Surname	2932
CreditScore	460
Geography	3
Gender	2
Age	70
Tenure	11
Balance	6382
NumOfProducts	4
HasCrCard	2
IsActiveMember	2
EstimatedSalary	9999
Exited	2

Data Preprocessing

As mentioned earlier, in our analysis of the data we pointed out that there were no missing values in the dataset and we once again display all the unique values in each feature of the data below. We preprocessed the data by dividing the dataset into three separate categories: training set, testing set, and validity set. The training set contains 7,000 data entries while the testing and validity set contain 1,500 entries each. We decided on these numbers because it is standard procedure to assign a bulk of the data in datasets to training the model so that it is able to provide more accurate results.

RowNumber	10000
CustomerId	10000
Surname	2932
CreditScore	460
Geography	3
Gender	2
Age	70
Tenure	11
Balance	6382
NumOfProducts	4
HasCrCard	2
IsActiveMember	2
EstimatedSalary	9999
Exited	2

Model 0: Neural Network with SGD Optimizer

Model Structure

- **Input Layer**:
  - 64 neurons with ReLU activation.
- **Hidden Layer**:
  - 32 neurons with ReLU activation.
- **Output Layer**:
  - 1 neuron with Sigmoid activation (binary classification).

Optimizer

- **SGD (Stochastic Gradient Descent)**: A simple and effective optimization algorithm for neural networks.

Strengths:

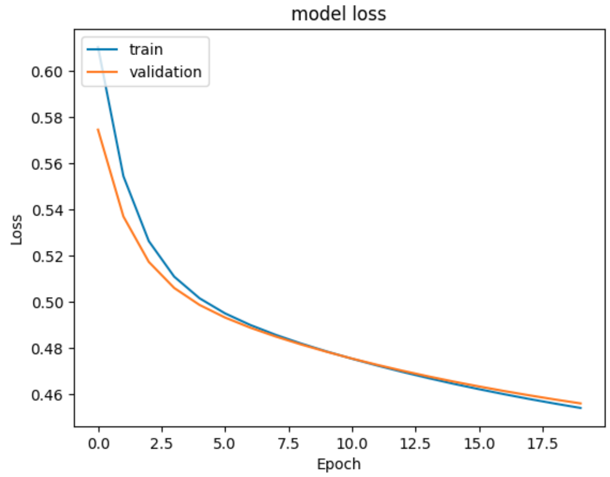
- Proper convergence with no major signs of overfitting.
- Efficient learning of the majority class (Class 0).

Weaknesses:

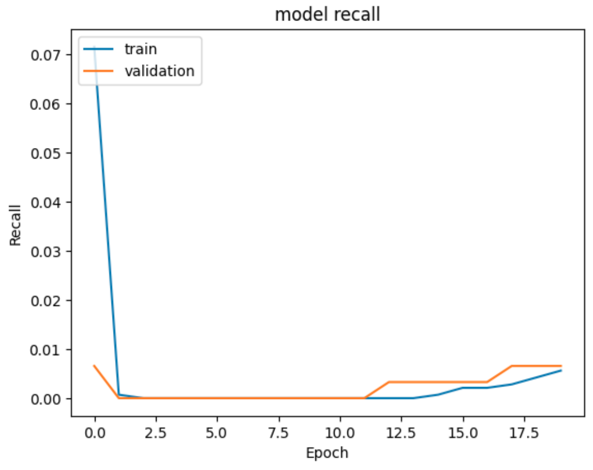
- The model fails to predict churners (Class 1), resulting in **extremely poor recall** for the minority class.
- Accuracy is misleading due to the class imbalance.

Conclusion:

- **Model 0 is inadequate for the problem**, as it does not handle the class imbalance effectively.



cor



	precision	recall	f1-score	support
0	0.80	1.00	0.89	5574
1	0.69	0.01	0.01	1426
accuracy			0.80	7000
macro avg	0.74	0.50	0.45	7000
weighted avg	0.78	0.80	0.71	7000

	precision	recall	f1-score	support
0	0.80	1.00	0.89	1195
1	0.67	0.01	0.01	305
accuracy			0.80	1500
macro avg	0.73	0.50	0.45	1500
weighted avg	0.77	0.80	0.71	1500

Model 1: Neural Network with Adam Optimizer

Model Structure

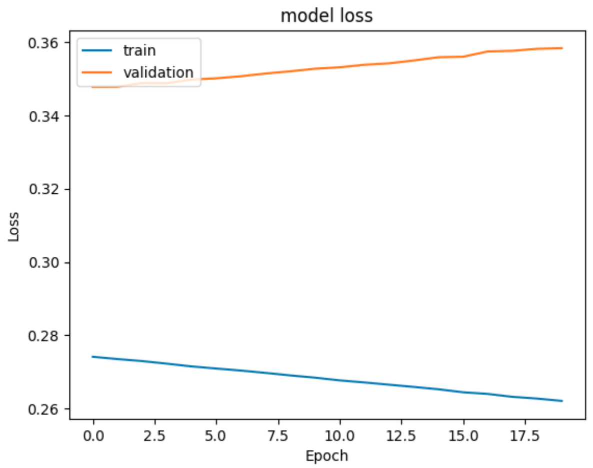
- **Input Layer**:
  - 64 neurons with ReLU activation.
- **Hidden Layer**:
  - 32 neurons with ReLU activation.
- **Output Layer**:
  - 1 neuron with Sigmoid activation for binary classification.

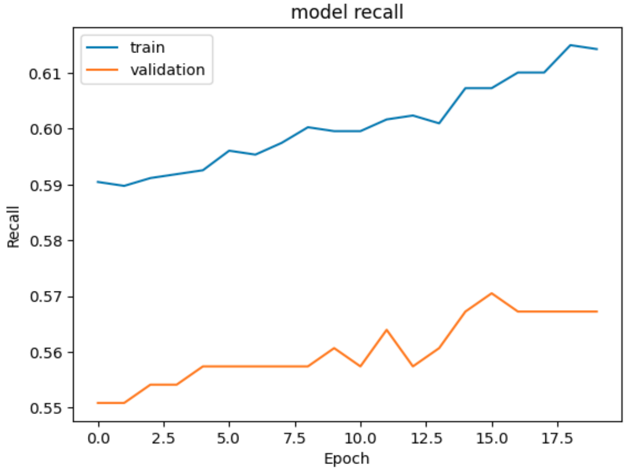
Optimizer

- **Adam**: An advanced optimization algorithm combining the benefits of Momentum RMSprop, known for adaptive learning rates and better convergence.

Overall Analysis

- **Strengths:**
  - Improved performance over Model 0 in terms of recall for churners (Class 1).
  - Adam optimizer helps the model converge faster with better recall.
- **Weaknesses:**
  - Significant overfitting observed as the gap between training and validation metrics widens.
  - Validation recall for churners (0.57) still needs improvement for real-world applicability.
- **Conclusion:**
  - Model 1 shows significant improvement in identifying churners compared to Model 0. However, there is a much bigger problem of overfitting that is seen in the graphs.





	precision	recall	f1-score	support
0	0.92	0.96	0.94	5574
1	0.80	0.66	0.72	1426
accuracy			0.90	7000
macro avg	0.86	0.81	0.83	7000
weighted avg	0.89	0.90	0.89	7000

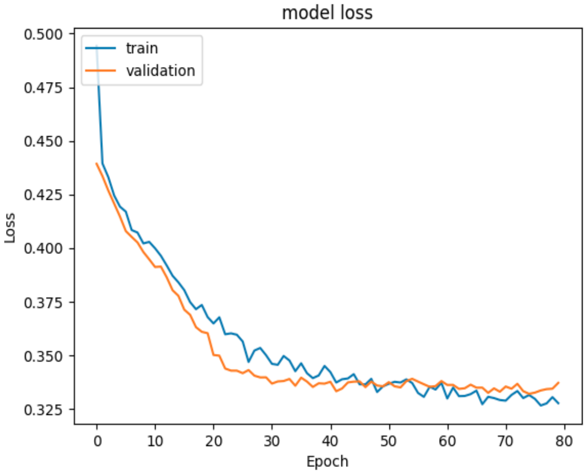
	precision	recall	f1-score	support
0	0.89	0.93	0.91	1195
1	0.68	0.57	0.62	305
accuracy			0.86	1500
macro avg	0.79	0.75	0.77	1500
weighted avg	0.85	0.86	0.85	1500

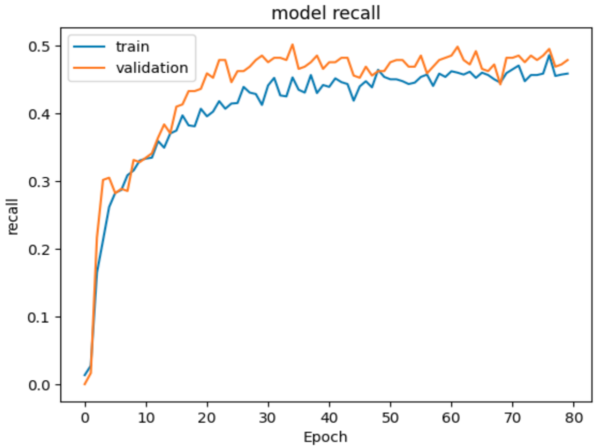
Model 2: Neural Network with Adam Optimizer and Dropout

Model Structure

- **Input Layer**:
  - 32 neurons with ReLU activation.
- **Hidden Layers**:
  - **First Layer**: 16 neurons with ReLU activation and a dropout of 0.2.
  - **Second Layer**: 8 neurons with ReLU activation and a dropout of 0.1.
  - **Third Layer**: 4 neurons with ReLU activation.
- **Output Layer**:
  - 1 neuron with Sigmoid activation for binary classification.

ADAM Optimizer





Overall Analysis

- **Strengths:**
  - **Use of dropout layers effectively reduces overfitting, as observed in the alignment of training and validation curves.**
  - **Significant improvement in recall for churners (Class 1) compared to Model 0 and Model 1.**
  - **Balanced performance for both training and validation sets.**
- **Weaknesses:**
  - **Recall for churners (Class 1) remains moderate at 0.48, leaving room for improvement.**
  - **The model shows limitations in handling class imbalance effectively, despite better generalization.**
- **Conclusion:**
  - **Model 2 demonstrates improved generalization and reduced overfitting compared to earlier models. However, further enhancements are necessary to improve recall for churners.**

	precision	recall	f1-score	support
0	0.88	0.97	0.92	5574
1	0.83	0.48	0.60	1426
accuracy			0.87	7000
macro avg	0.85	0.73	0.76	7000
weighted avg	0.87	0.87	0.86	7000

	precision	recall	f1-score	support
0	0.88	0.97	0.92	1195
1	0.80	0.48	0.60	305
accuracy			0.87	1500
macro avg	0.84	0.72	0.76	1500
weighted avg	0.86	0.87	0.86	1500

Model 3:Neural Network with SMOTE Applied Data and SGD Optimizer

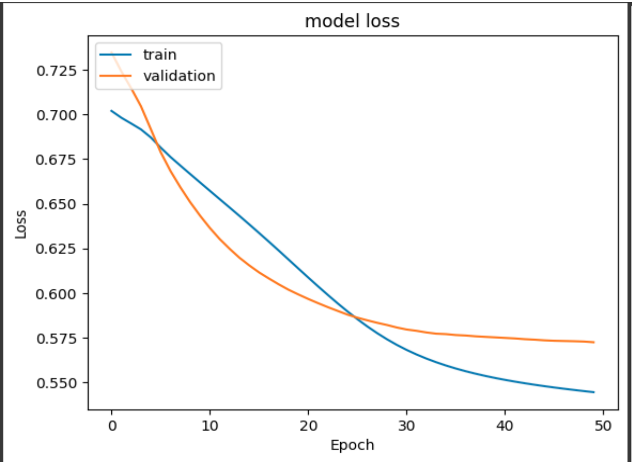
Model Structure:

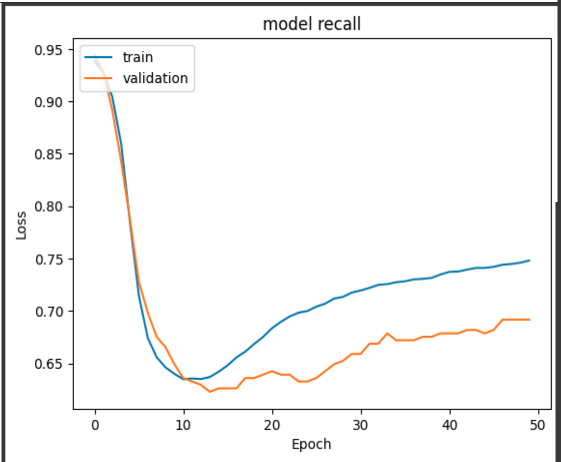
- Input Layer: 32 neurons with ReLU activation.
- Hidden Layers:
  - 16 neurons with ReLU activation.
  - 8 neurons with ReLU activation.
- Output Layer: 1 neuron with Sigmoid activation for binary classification.
- The data was balanced using SMOTE to address class imbalance.

- **Strengths:**
  - SMOTE successfully addresses the class imbalance in the training set, leading to relatively balanced recall and precision between classes on the training data.
  - The recall for churners (`Class 1`) on the validation set improves significantly compared to earlier models without SMOTE.
- **Weaknesses:**
  - Validation recall for churners is still lower than the training recall, showing that the model has not generalized as well as desired to unseen data.
  - Precision for churners (`Class 1`) on the validation set is low, suggesting misclassification of non-churners as churners.

Overall Performance:

- The use of SMOTE improves recall for the minority class (`Class 1`), making the model better at identifying churners.
- The SGD optimizer performs adequately, but the model struggles with precision and generalization to the validation set. Needs more tuning in order to improve generalization and overfitting.





	precision	recall	f1-score	support
0	0.74	0.73	0.73	5574
1	0.73	0.75	0.74	5574
accuracy			0.74	11148
macro avg	0.74	0.74	0.74	11148
weighted avg	0.74	0.74	0.74	11148

	precision	recall	f1-score	support
0	0.90	0.71	0.80	1195
1	0.38	0.69	0.49	305
accuracy			0.71	1500
macro avg	0.64	0.70	0.64	1500
weighted avg	0.79	0.71	0.73	1500

Model 4:Neural Network with SMOTE Applied Data and Adam Optimizer

Model Structure:

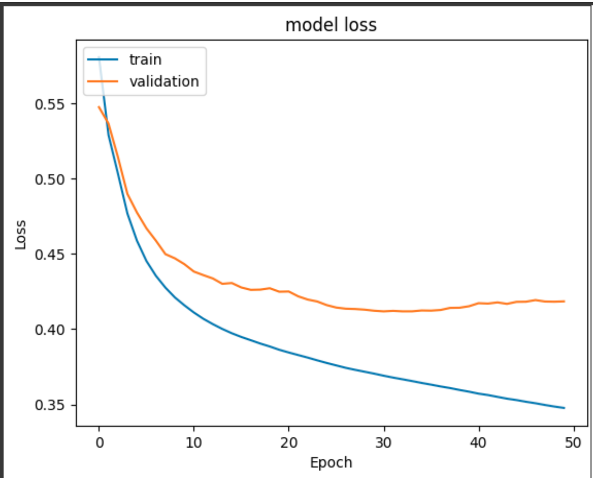
- **Input Layer**: 32 neurons with ReLU activation.
- **Hidden Layers**:
    - 1st Layer: 16 neurons, ReLU activation.
    - 2nd Layer: 8 neurons, ReLU activation.
- **Output Layer**: 1 neuron with sigmoid activation for binary classification.
- **The data was balanced using SMOTE to address class imbalance and utilizes ADAM optimizer**

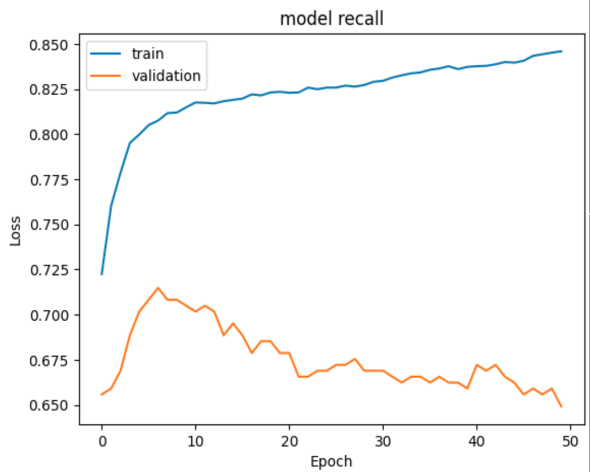
Strengths:

- The model performs well on the training data, achieving high precision and recall for both classes.
- SMOTE balances the dataset, improving the model's ability to detect churners compared to earlier models without balancing.

Weaknesses:

- Overfitting is evident, as validation recall and F1-scores are significantly lower compared to the training set as well as the gap which is seen in the chart.
- Validation metrics for churners are suboptimal, which could lead to missed churn predictions in real-world scenarios.





	precision	recall	f1-score	support
0	0.84	0.87	0.86	5574
1	0.87	0.83	0.85	5574
accuracy			0.85	11148
macro avg	0.85	0.85	0.85	11148
weighted avg	0.85	0.85	0.85	11148

	precision	recall	f1-score	support
0	0.91	0.86	0.88	1195
1	0.54	0.65	0.59	305
accuracy			0.82	1500
macro avg	0.72	0.75	0.74	1500
weighted avg	0.83	0.82	0.82	1500

Model 5:Neural Network with SMOTE Applied Data, Adam Optimizer, and DropOut

Great Learning
POWER AHEAD

Model Structure

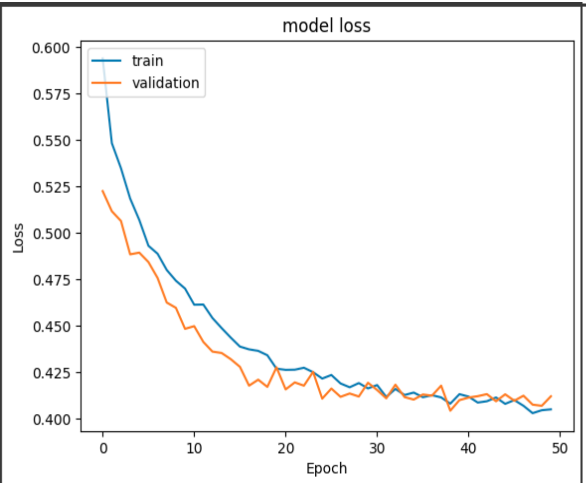
- **Input Layer:** 32 neurons with ReLU activation.
- **Dropout Layers:** Dropout rate of 0.2 after the input layer and 0.1 after subsequent hidden layers to reduce overfitting.
- **Hidden Layers:**
  - First layer: 16 neurons with ReLU activation.
  - Second layer: 8 neurons with ReLU activation.
- **Output Layer:** 1 neuron with Sigmoid activation to predict binary outcomes (churn vs. no churn).
- **SMOTE:** Synthetic Minority Oversampling Technique applied to balance the dataset, addressing class imbalance.
- **Optimizer:** Adam optimizer for efficient gradient descent.

Strengths

1. **Balanced Performance: SMOTE ensures balanced performance across both classes, addressing the original class imbalance.**
2. **Dropouts: Effectively reduce overfitting, as reflected in comparable training and validation performance.**
3. **Validation Recall: Achieves one of the highest recalls (72%) among all models, crucial for identifying potential churners.**

Weaknesses

1. **Precision for Class 1: Slightly lower precision for identifying churners (52%), which may result in more false positives.**
2. **Validation Loss Stability: Fluctuations in validation loss indicate room for further hyperparameter optimization.**



	precision	recall	f1-score	support
0	0.83	0.84	0.83	5574
1	0.84	0.82	0.83	5574
accuracy			0.83	11148
macro avg	0.83	0.83	0.83	11148
weighted avg	0.83	0.83	0.83	11148

	precision	recall	f1-score	support
0	0.92	0.83	0.87	1195
1	0.52	0.71	0.60	305
accuracy			0.81	1500
macro avg	0.72	0.77	0.74	1500
weighted avg	0.84	0.81	0.82	1500

Executive Summary

- Based on the Training Performance and the Validation Performace Comparisons, our recommendation is Model 5: NN with SMOTE, Adam & Dropout in order to identify Customer Churn for banks. . With a validation recall of 0.708, it outperforms all other models in capturing churners (minority class) on unseen data while maintaining robustness against overfitting.

Training performance comparison

	recall
NN with Adam & Dropout	0.476157
NN with SMOTE & SGD	0.747219
NN with SMOTE & Adam	0.831898
NN with SMOTE,Adam & Dropout	0.823466

Validation set performance comparison

	recall
NN with Adam & Dropout	0.478689
NN with SMOTE & SGD	0.691803
NN with SMOTE & Adam	0.649180
NN with SMOTE,Adam & Dropout	0.708197