Artificial Neural Network (ANN) Model Architecture for Customer Churn Prediction

1. Overview

This document outlines the architecture and configuration of the Artificial Neural Network (ANN) model developed to predict customer churn. The model was designed to achieve strong predictive performance while maintaining transparency and reproducibility. The model was built using the TensorFlow and Keras libraries in Python.

2. Input Layer

The input layer receives a set of standardised numeric and categorical features derived from the prepared training dataset. The input dimension is equal to the number of features after preprocessing and scaling. All input variables were standardised using the StandardScaler method to ensure consistent feature range.

3. Hidden Layers

The network includes three hidden layers optimised for predictive accuracy and generalisation performance. Each layer uses the Rectified Linear Unit (ReLU) activation function to enable non-linear learning.

• First hidden layer: 128 neurons, ReLU activation, followed by a 20% dropout layer to reduce overfitting.

• Second hidden layer: 64 neurons, ReLU activation.

• Third hidden layer: 32 neurons, ReLU activation.

4. Output Layer

The output layer contains a single neuron with a Sigmoid activation function. This configuration outputs a probability value between 0 and 1, representing the likelihood of a customer churning. A decision threshold is applied to classify customers as 'likely to churn' or 'not likely to churn'.

5. Optimisation and Training

The model was optimised using the Adam optimiser with a learning rate of 0.0001. Binary Cross-Entropy was used as the loss function, which is appropriate for binary classification problems. The model was trained for up to 50 epochs with a batch size of 32, using early stopping and learning rate reduction strategies to avoid overfitting and improve performance.

6. Model Performance and Evaluation

The model was evaluated using accuracy, precision, recall, F1-score, and ROC-AUC metrics. Threshold optimisation was conducted to identify the probability threshold that maximised

the F1-score. Performance results and visual outputs (such as loss and accuracy plots, ROC curve, and confusion matrices) were generated to support interpretability and transparency.

7. Summary of Key Model Parameters

• Optimiser: Adam (learning rate = 0.0001)

• Loss Function: Binary Cross-Entropy

• Activation Functions: ReLU (hidden layers), Sigmoid (output layer)

• Dropout Rate: 20% after first hidden layer

• Epochs: 50 (with early stopping)

• Batch Size: 32

• Evaluation Metrics: Accuracy, Precision, Recall, F1-score, ROC-AUC

8. Compliance and Reproducibility

This model was developed following open and transparent research practices. All preprocessing, training, and evaluation steps have been documented and saved as artefacts. This ensures that the model can be reproduced and audited according to good data science and government standards.