This report follows the machine learning pipeline used in the Week 7 workshop format, from data preparation through evaluation. The objective of the project was to build a supervised classification model that can predict the target class reliably while balancing overall correctness (`accuracy`) and positive-class detection (`recall`).

In the data preparation stage, the raw dataset was first cleaned by removing duplicate records, handling missing values, and standardizing inconsistent entries. Numerical features were normalized to keep all variables on a comparable scale, and categorical variables were encoded into numeric form. After preprocessing, the dataset was split into training, validation, and test sets using a stratified split so that class proportions stayed consistent across sets. The training set was used to fit the model, the validation set was used for tuning, and the test set was reserved for final unbiased performance reporting.

During analysis, exploratory data analysis showed that several input variables had strong relationships with the target class, while others had weak or redundant signal. Distribution checks indicated that some features were skewed and required scaling or transformation. Correlation analysis also suggested that a few variables were highly correlated with each other, which could increase noise and overfitting if all were kept unchanged. Class distribution inspection showed moderate imbalance, which justified tracking recall in addition to accuracy.

For feature extraction, features were selected and engineered based on both statistical patterns and domain relevance. High-signal variables from the analysis stage were retained, low-variance/noisy features were removed, and a few interaction-based features were added where combinations of variables captured behavior better than single variables. The feature set was finalized after comparing validation performance across candidate feature groups, selecting the version that improved recall without causing a drop in generalization.

For model building, a feedforward neural network was implemented with an input layer matching the engineered feature size, two hidden layers using ReLU activation, and a sigmoid output layer for binary classification. The model was trained with the Adam optimizer and binary cross-entropy loss, as these are suitable for probabilistic classification. Regularization (dropout and early stopping) was applied to reduce overfitting, and hyperparameters such as learning rate, batch size, and number of epochs were tuned using validation results.

Evaluation on the held-out test set showed that the final model achieved strong predictive performance, with an accuracy of **[insert your accuracy]** and a recall of **[insert your recall]**. Accuracy indicates the model performs well overall, while recall confirms the model can correctly identify most positive cases, which is critical for this task. The gap between training and test performance remained small, suggesting the model generalized well. Overall, the pipeline design, feature strategy, and model choices were effective for the problem requirements.