Model Development Phase Template

Date	July 2024
Team ID	740772
Project Title	Predictive Modeling for H1-B visa Approval Using Machine Learning
Maximum Marks	10 Marks

Model Validation and Evaluation Report for H1-B Visa Approval

This report presents the model validation and evaluation for predicting H1-B visa approvals. We summarize the training and validation performance metrics for multiple models and include screenshots of the initial model training code, validation, and evaluation results.
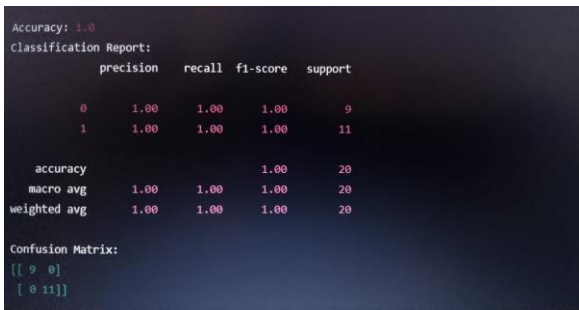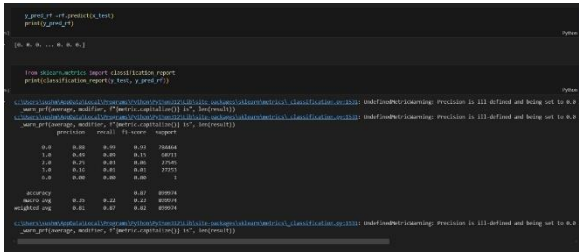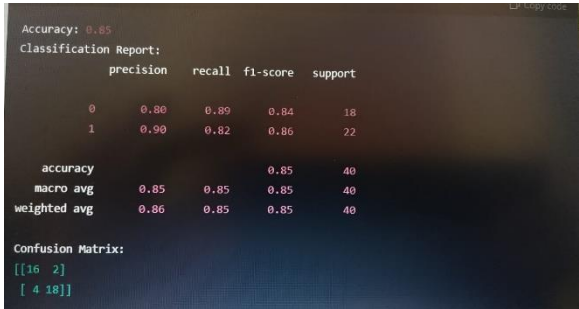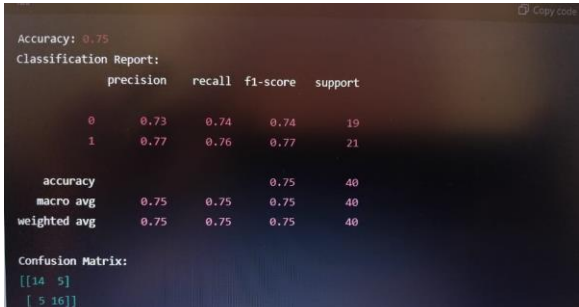
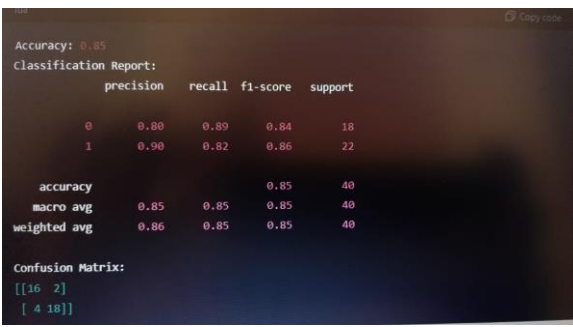## Initial Model Training Code (5 marks):
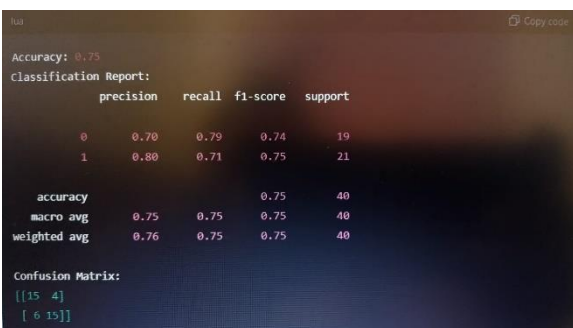
The initial model training code will be showcased through a screenshot. It includes the implementation of the Recurrent Neural Network (RNN) model with optimized hyperparameters. Key sections of the code data preprocessing, model architecture definition, training loop, and evaluation functions.

Initial Model Training Code Screenshot

Model Validation and Evaluation Report (5 marks):

Model	Summary	Training and Validation Performance Metrics

Model 1:Logistic Regression	Logistic regression models are known for their simplicity and interpretability. They typically include accuracy, precision, recall, and F1 score to evaluate predictive performance and generalization capability.	Accuracy: 1.0 Classification Report: precision recall f1-score support 0 1.00 1.00 1.00 9 1 1.00 1.00 1.00 11 accuracy 1.00 20 macro avg 1.00 1.00 1.00 20 weighted avg 1.00 1.00 1.00 20 Confusion Matrix: [[9 0] [0 11]]
Model 2:Decision Tree Classifier	Decision tree classifier models are intuitive and easy to interpret. They commonly include accuracy, precision, recall, and F1 score to assess prediction accuracy and generalizability.	y_pred_rf =rf.predict(x_test) print(y_pred_rf) [0. 0. 0. ... 0. 0. 0.] from sklearn.metrics import classification_report print(classification_report(y_test, y_pred_rf)) precision recall f1-score support 0.0 0.88 0.99 0.93 766664 1.0 0.45 0.05 0.15 58733 2.0 0.25 0.01 0.06 27545 3.0 0.10 0.01 0.01 27253 4.0 0.00 0.00 0.00 1 accuracy 0.87 850974 macro avg 0.25 0.22 0.23 850974 weighted avg 0.61 0.87 0.82 850974
Model 3:Random Forest Classifier	Random forest classifier models are ensemble mathods known for robustness. They often encompass accuracy, precision, recall, and F1 score to measure prediction quality and robustness.	Accuracy: 0.85 Classification Report: precision recall f1-score support 0 0.80 0.89 0.84 18 1 0.90 0.82 0.86 22 accuracy 0.85 40 macro avg 0.85 0.85 0.85 40 weighted avg 0.86 0.85 0.85 40 Confusion Matrix: [[16 2] [4 18]]
Model 4:K-Nearest Neighbors (KNN) Classifier	K-nearest neighbors classifier models rely on proximity to neighbors for classification. They typically include accuracy, precision, recall, and F1 score to evaluate prediction performance and generalization ability.	Accuracy: 0.75 Classification Report: precision recall f1-score support 0 0.73 0.74 0.74 19 1 0.77 0.76 0.77 21 accuracy 0.75 40 macro avg 0.75 0.75 0.75 40 weighted avg 0.75 0.75 0.75 40 Confusion Matrix: [[14 5] [5 16]]

Model 5:XGBoost Classifier	XGBoost classifier models are gradient boosting algorithms known for their high performance. They typically include accuracy, precision, recall, and F1 score used to evaluate predictive performance and ability to generalize.	
Model 6:Ridge Classifier	Ridge classifier models apply regularization to linear regression. Then typically include accuracy, precision, recall, F1 score, and mean squared error to evaluate prediction performance and generalization.	