Random Forest Implementation and Results

Preprocessing

```
rec_up <- recipe(days_since_semester_start ~ grit_overall +
    matrices_hit_rate +
                 neuroticism_score + extraversion_score +
                     conscientiousness_score +
                 openness_score + agreeableness_score + adhd_total
                     +
                 growth_mindset_score + semester, data = data_trn)
                     |>
  step_impute_median(all_numeric_predictors())
```

Hyperparameter Tuning

Grid search was performed to tune the hyperparameters `trees`, `mtry`, and `min_n`:

```
grid_rf <- expand_grid(
  trees = c(100, 500, 750, 1000),
  mtry = c(3, 5, 8, 10),
  min_n = c(5, 8, 10, 12)
)

fits_rf_up <- tune_grid(
  rand_forest(trees = tune(), mtry = tune(), min_n = tune()) |>
    set_engine("ranger") |>
    set_mode("regression"),
  preprocessor = rec_up,
  resamples = splits_boot,
  grid = grid_rf,
  metrics = metric_set(rmse)
)
```

Fitting the Final Model

The best parameters (`trees = 1000`, `mtry = 8`, `min_n = 10`) were used to fit the final model:

```
fit_rf <-
  rand_forest(trees = 1000, mtry = 8, min_n = 10) |>
  set_engine("ranger", importance = "impurity", seed = 20010429) |>
  set_mode("regression") |>
  fit(days_since_semester_start ~ ., data = feat_trn_rf)
```

Evaluation on Test Data

The model was evaluated on test data using RMSE:

```r
rmse_test <- rmse_vec(
  truth = feat_test_rf$days_since_semester_start,
  estimate = predict(fit_rf, feat_test_rf)$.pred
)
```

Permutation Feature Importance

Permutation feature importance analysis was conducted. The results are summarized in Table 1:

Table 1: Permutation Feature Importance Results

Variable	**Mean Dropout Loss (RMSE)**
full_model	15.24604
conscientiousness_score	19.90218
extraversion_score	19.64418
openness_score	19.13487
matrices_hit_rate	18.85664
growth_mindset_score	18.53905
neuroticism_score	18.49867
agreeableness_score	18.46523
adhd_total	18.19964
grit_overall	17.99741
semester	15.24604
baseline (all permuted)	27.50965