

## Supplementary Figures

Part 1 Resampling: Figure 1 – 24

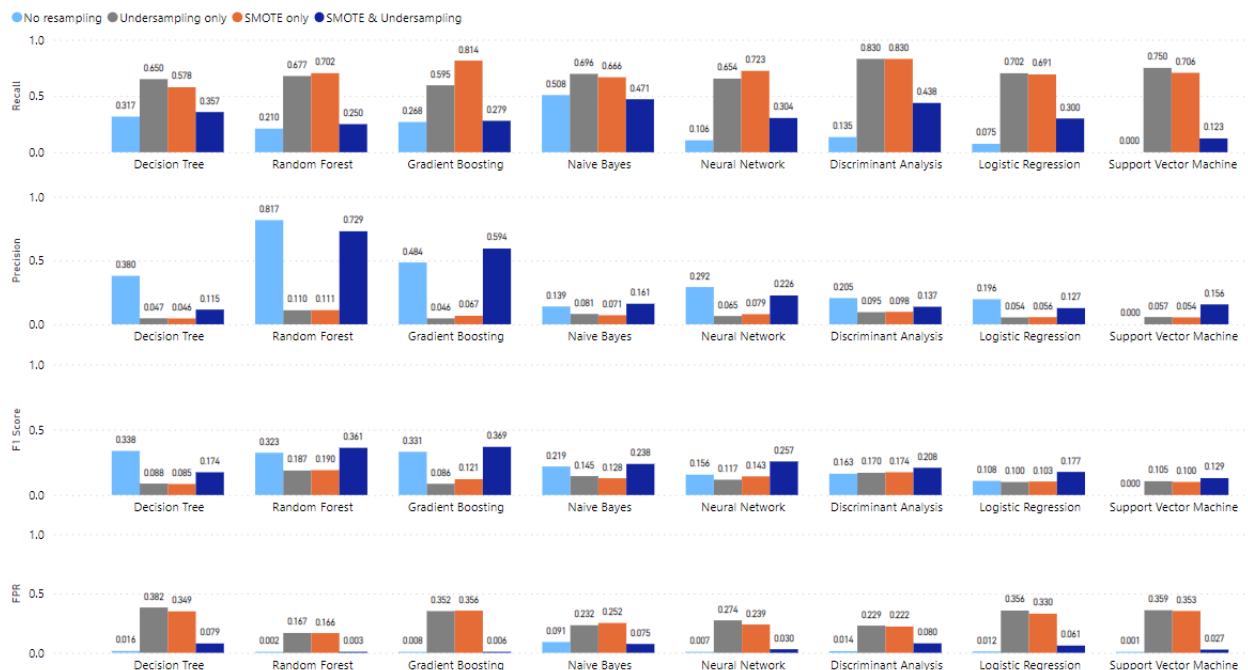
Part 2 Feature Sets: Figure 25 - 37

Part 3 Eight base models: Figure 38 – 52

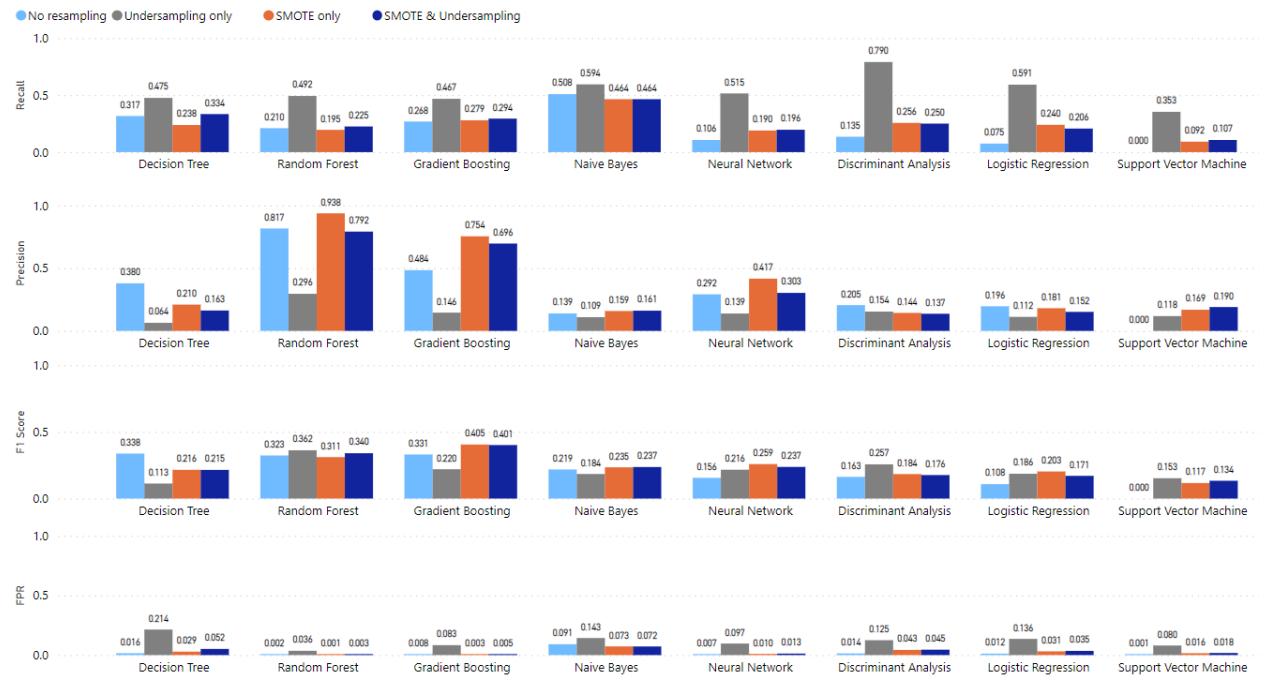
Part 4. Stacking Only (1-component models): Figure 53 – 61

Part 5. 1-component meta-model (with top-2 base models) compared with 8 base models: Figure 61 - 66

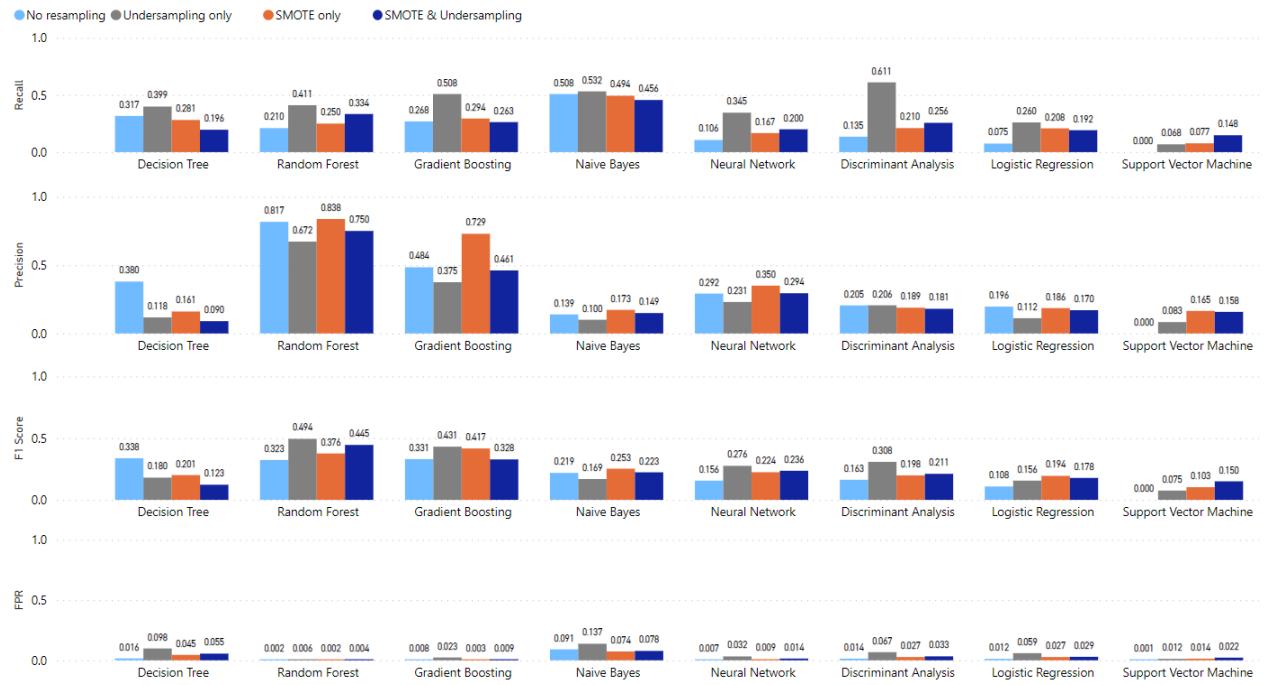
### Part 1 - resampling



**Figure 1.** Model performance comparison with and without resampling based on item responses, response time and summative statistics with a ratio of 1:1 between the non-cheater and cheater classes.



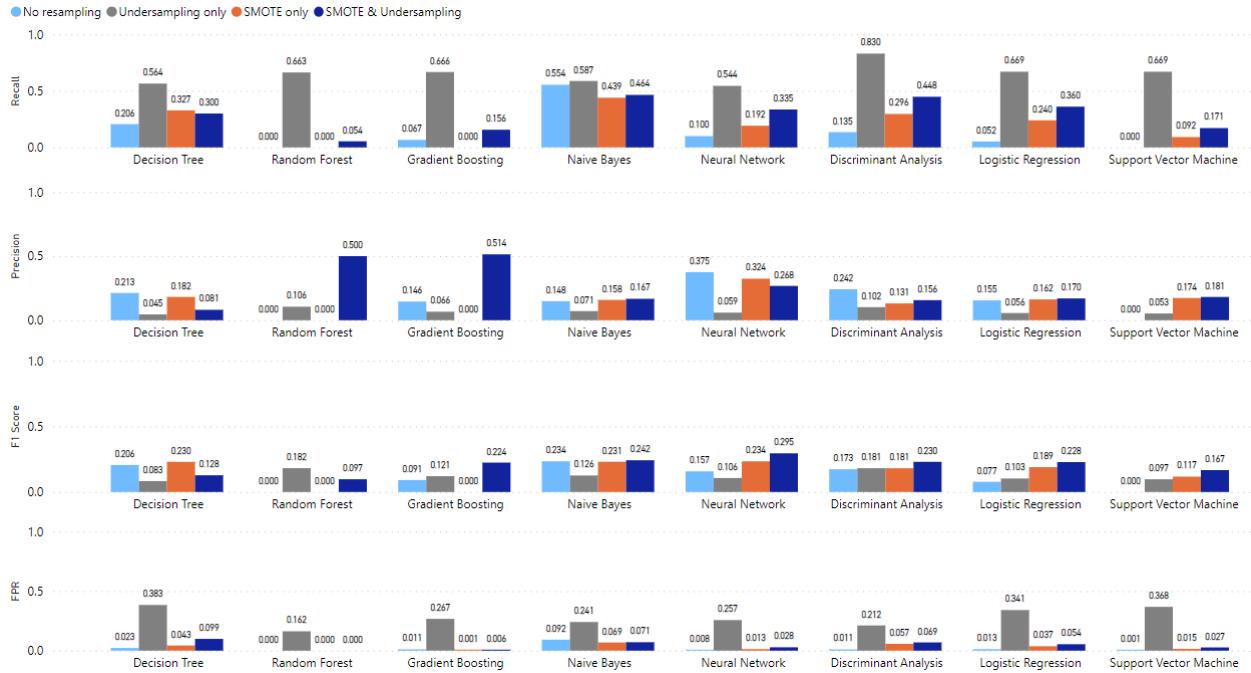
**Figure 2.** Model performance comparison with and without resampling based on item responses, response time and summative statistics with a ratio of 2:1 between the non-cheater and cheater classes.



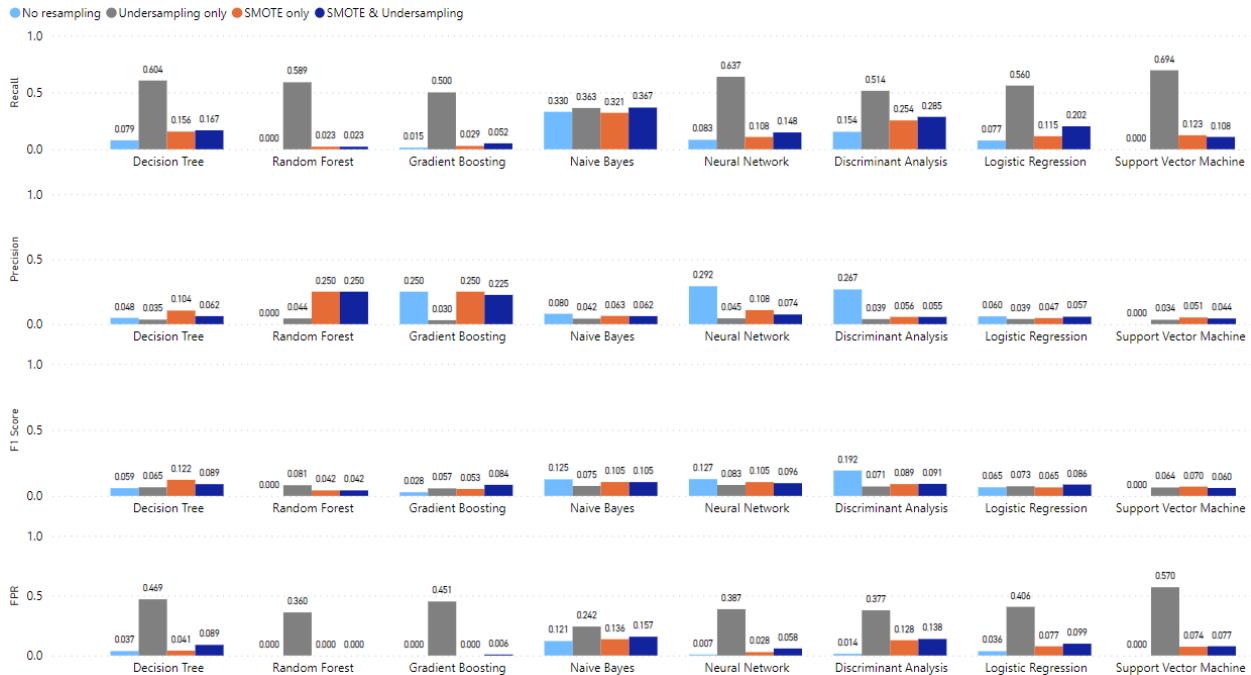
**Figure 3.** Model performance comparison with and without resampling based on item responses, response time and summative statistics with a ratio of 5:1 between the non-cheater and cheater classes.



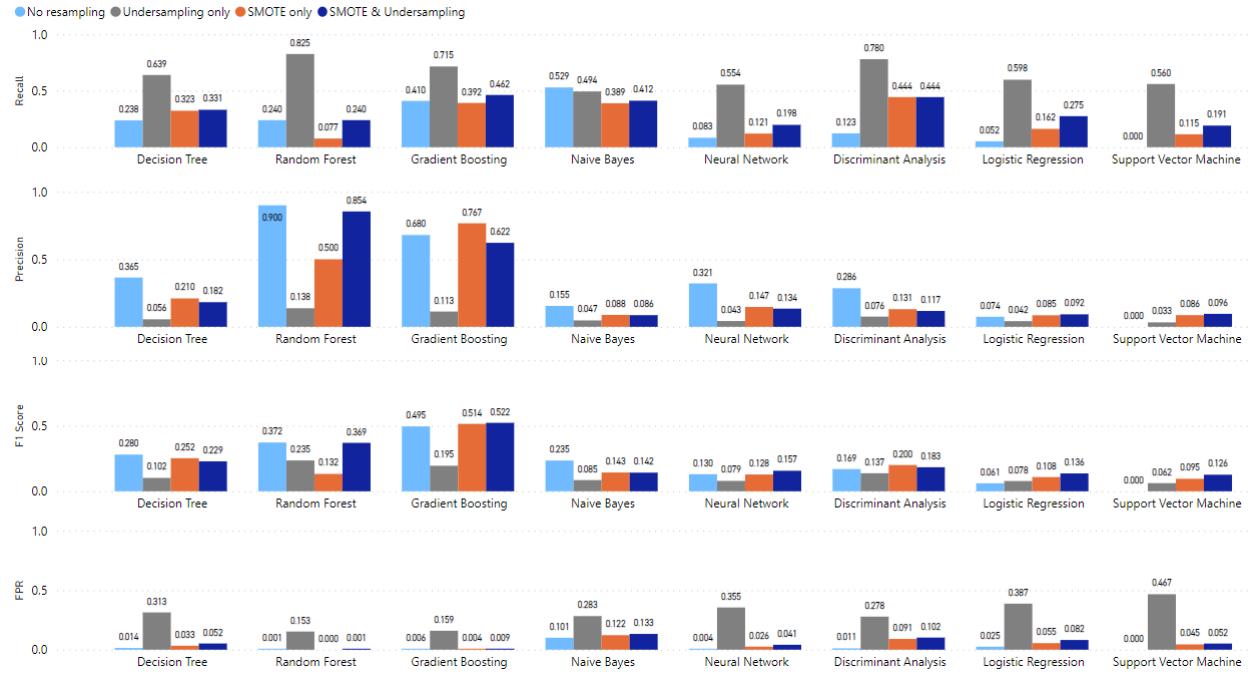
**Figure 4.** Model performance comparison with and without resampling based on item responses, response time and summative statistics with a ratio of 10:1 between the non-cheater and cheater classes.



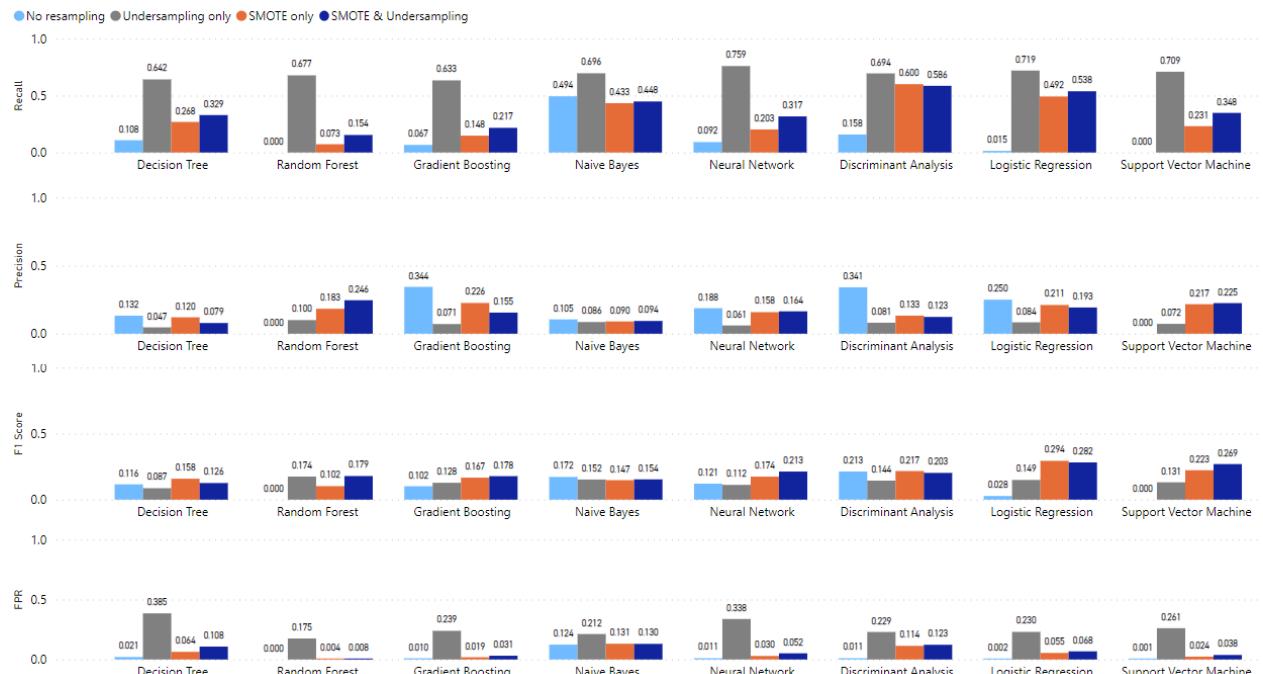
**Figure 5.** Performance comparison of models with and without resampling based on item response, response time with a ratio of 1:1 between the non-cheater and cheater classes.



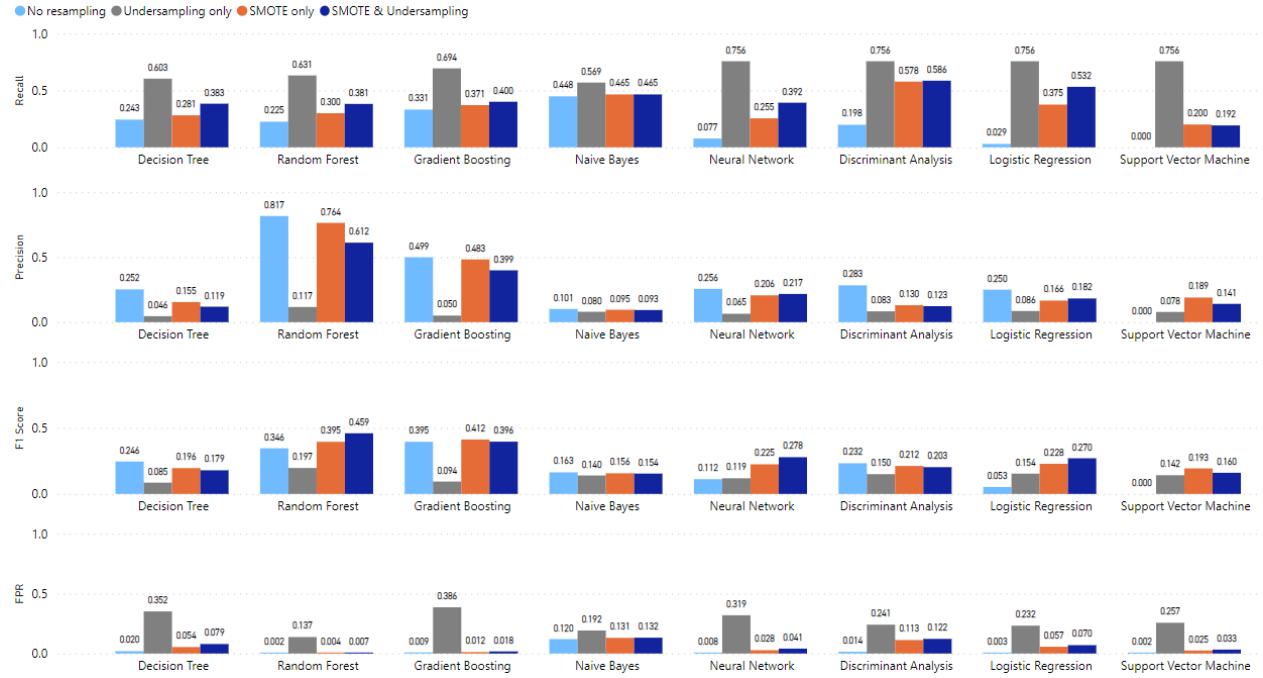
**Figure 6.** Performance comparison of models with and without resampling based on item response with a ratio of 1:1 between the non-cheater and cheater classes.



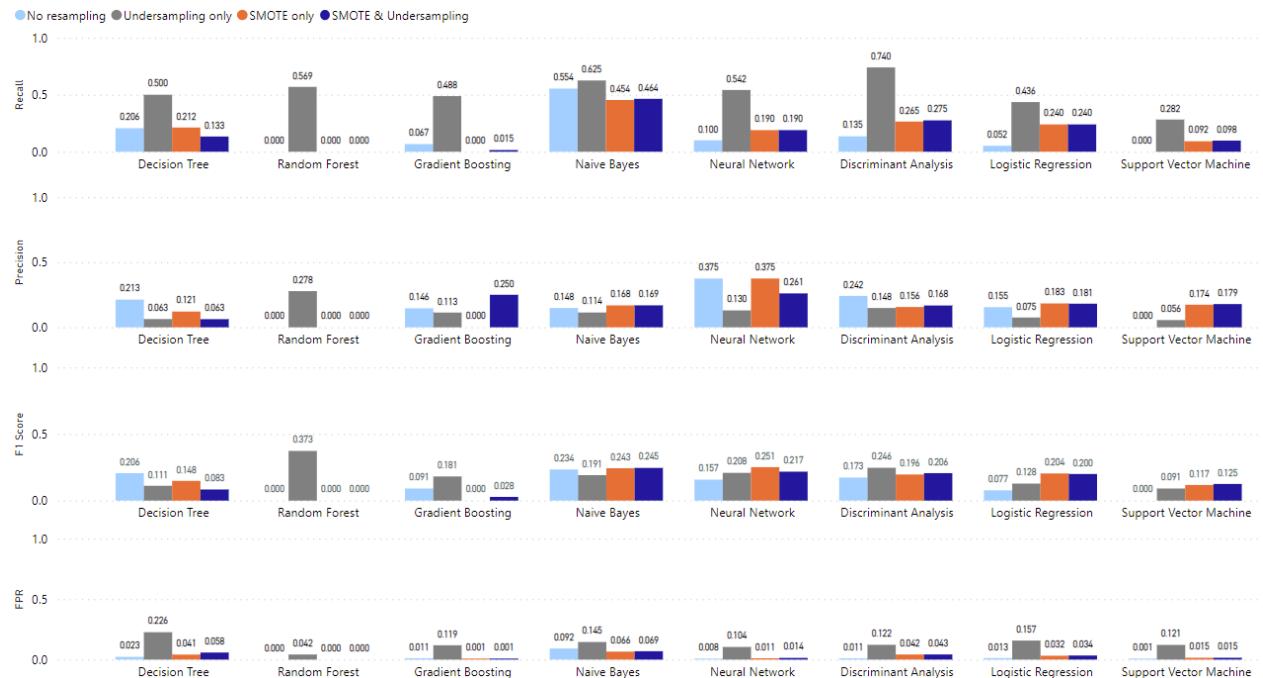
**Figure 7.** Performance comparison of models with and without resampling based on item response and summative statistics with a ratio of 1:1 between the non-cheater and cheater classes.



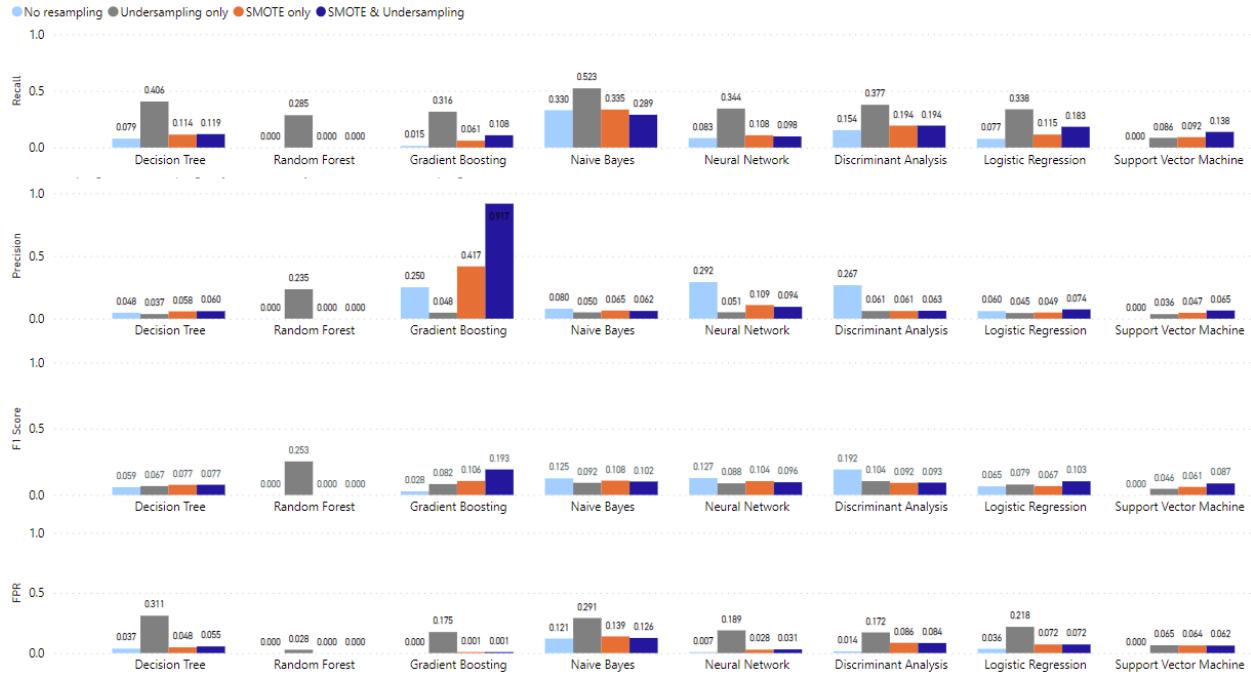
**Figure 8.** Performance comparison of models with and without resampling based on response time with a ratio of 1:1 between the non-cheater and cheater classes.



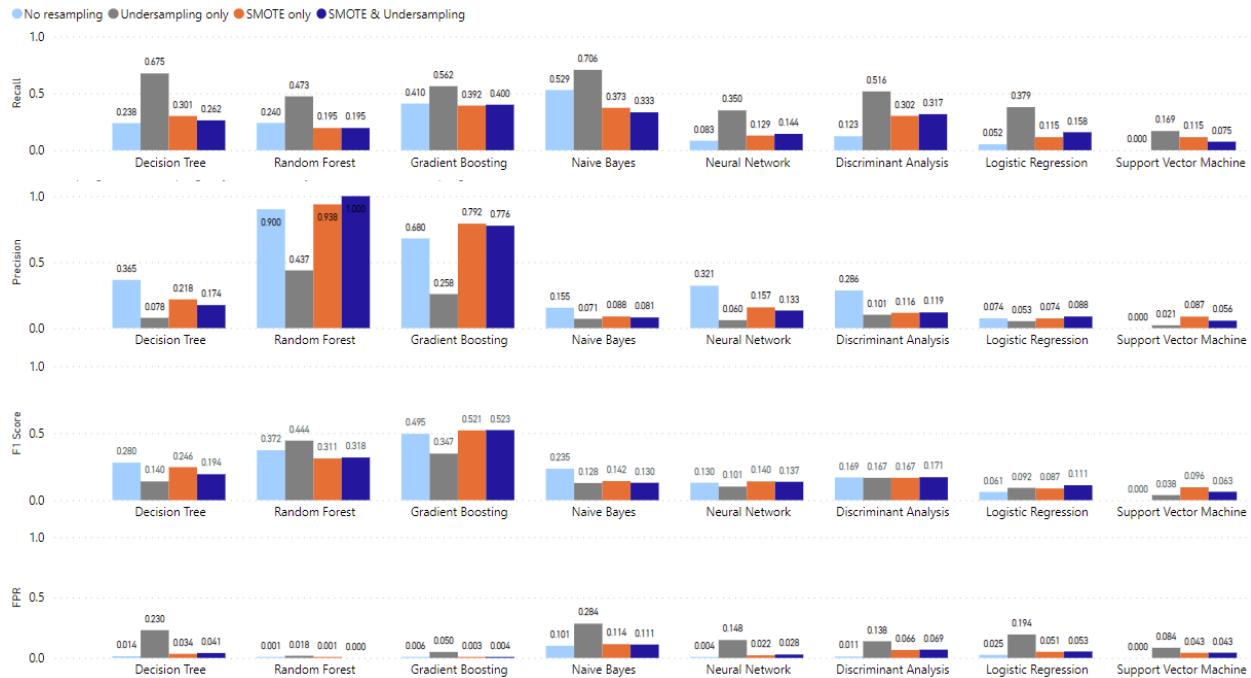
**Figure 9.** Performance comparison of models with and without resampling based on response time and summative statistics with a ratio of 1:1 between the non-cheater and cheater classes.



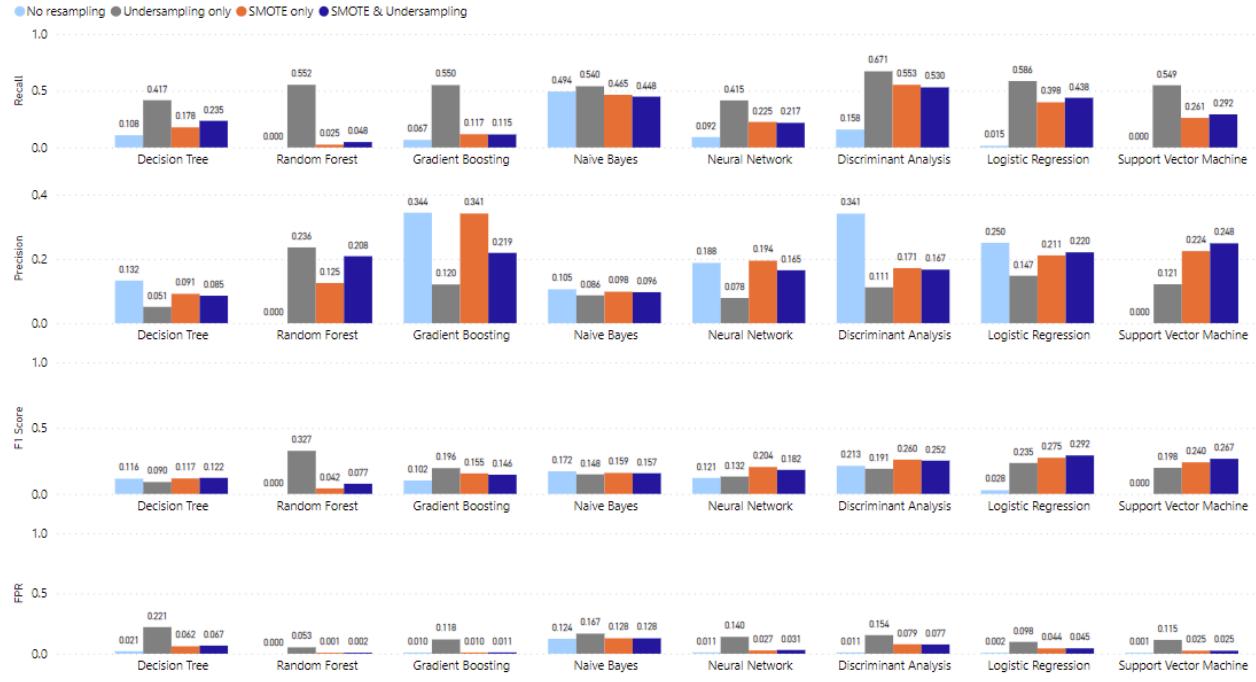
**Figure 10.** Performance comparison of models with and without resampling based on item response, response time with a ratio of 2:1 between the non-cheater and cheater classes.



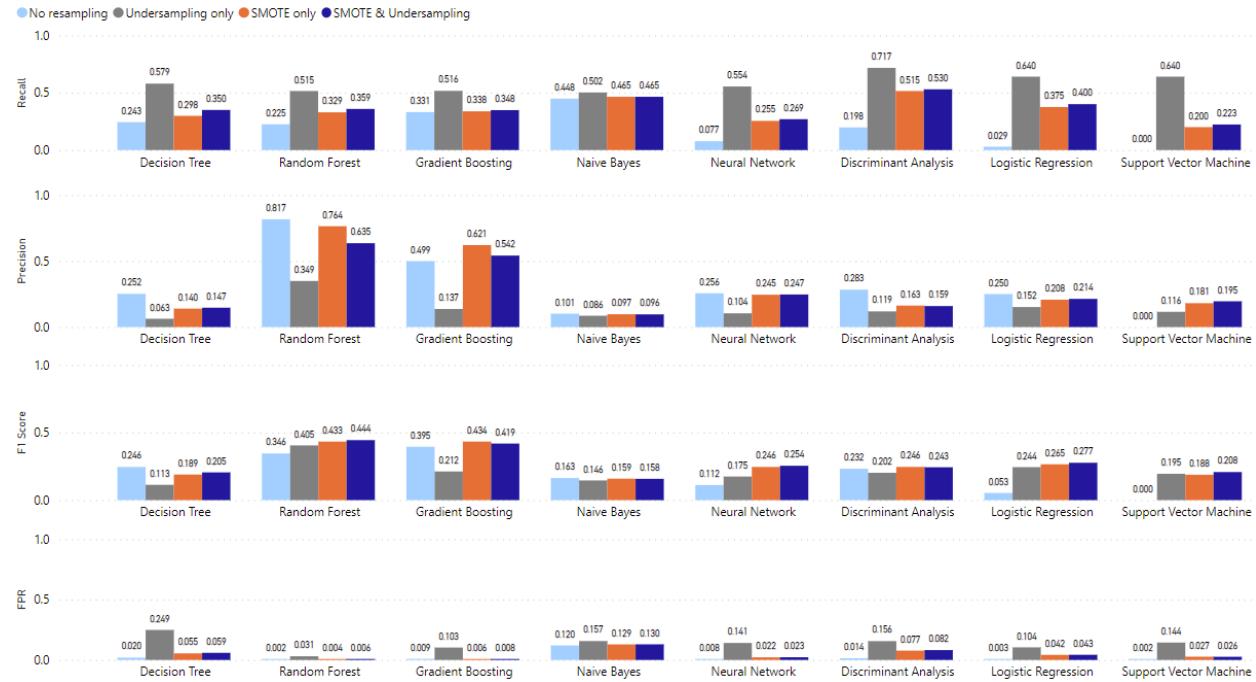
**Figure 11.** Performance comparison of models with and without resampling based on item response with a ratio of 2:1 between the non-cheater and cheater classes.



**Figure 12.** Performance comparison of models with and without resampling based on item response and summative statistics with a ratio of 2:1 between the non-cheater and cheater classes.



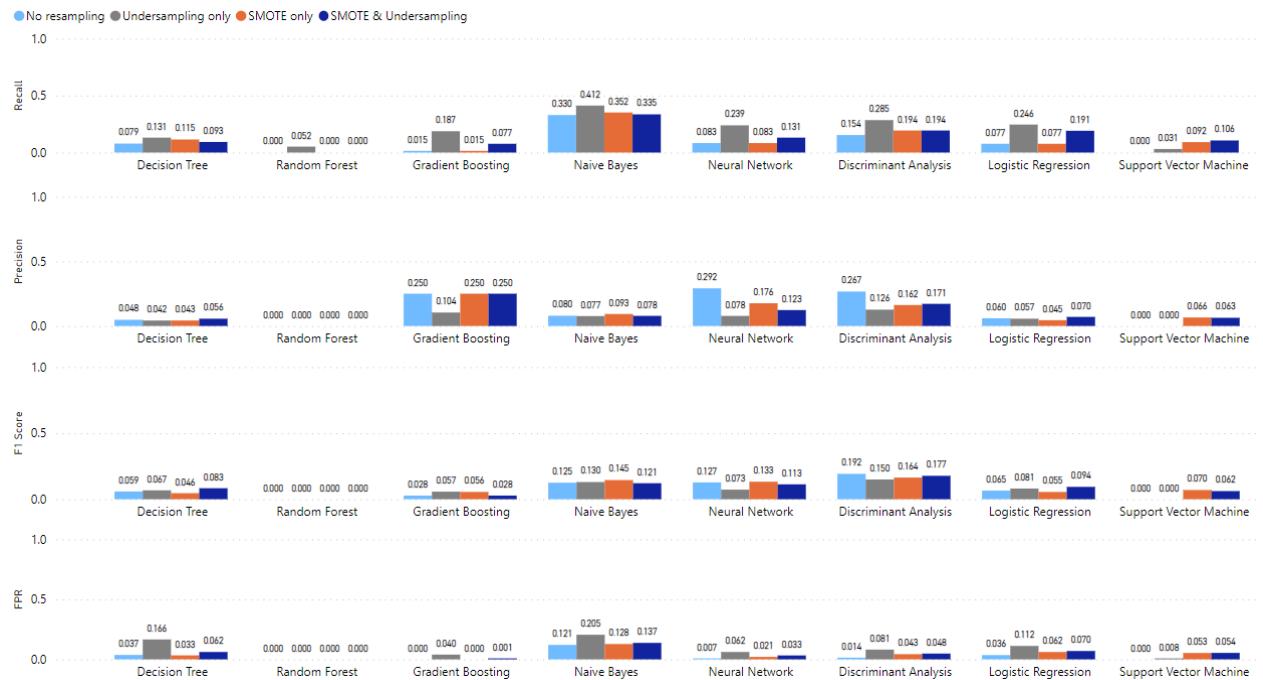
**Figure 13.** Performance comparison of models with and without resampling based on response time with a ratio of 2:1 between the non-cheater and cheater classes.



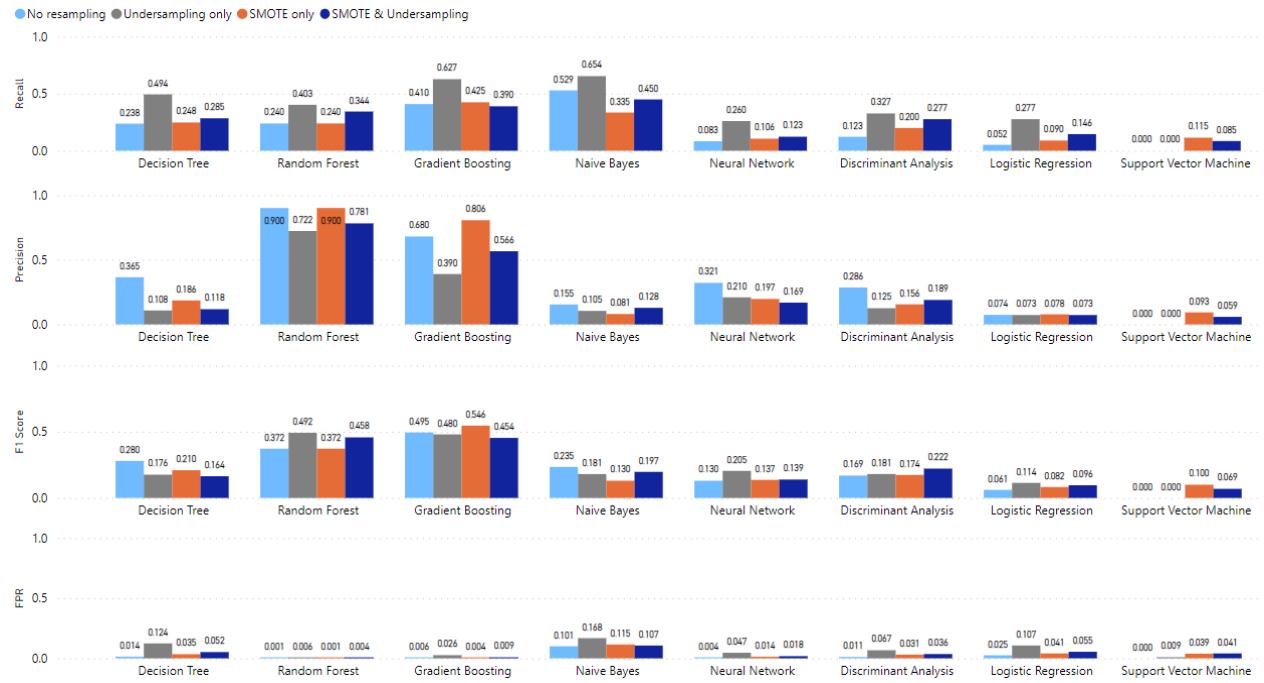
**Figure 14.** Performance comparison of models with and without resampling based on response time and summative statistics with a ratio of 2:1 between the non-cheater and cheater classes.



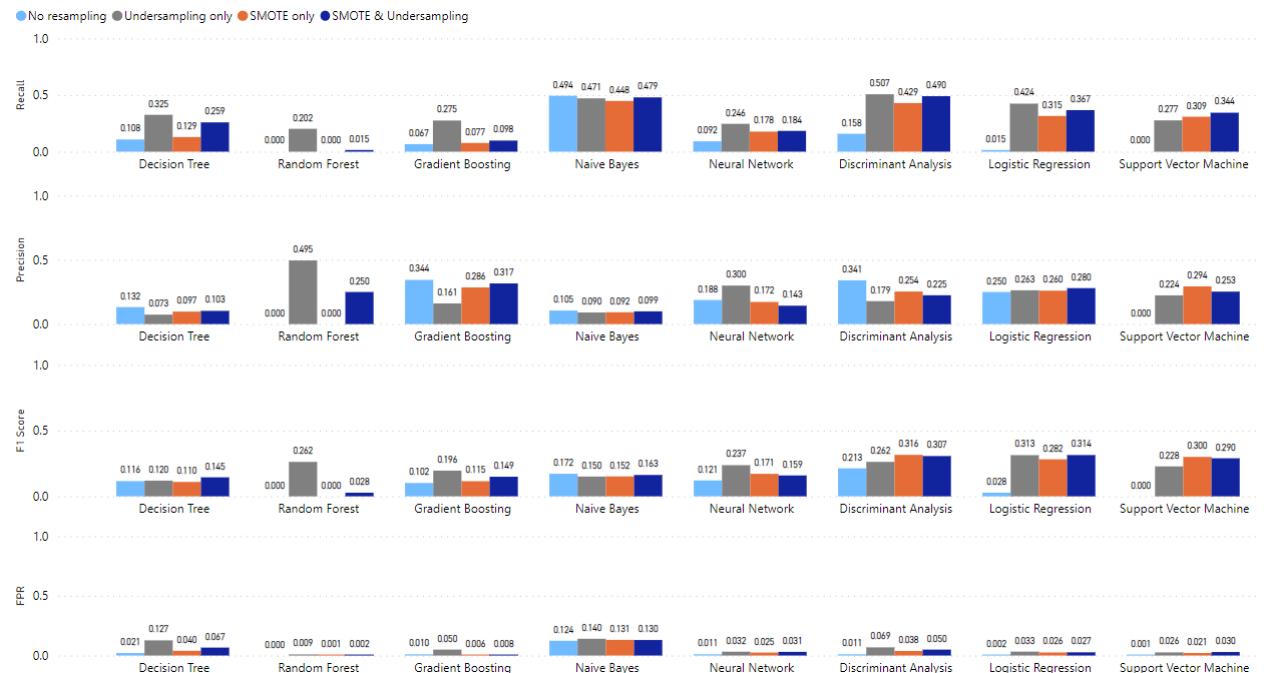
**Figure 15.** Performance comparison of models with and without resampling based on item response, response time with a ratio of 5:1 between the non-cheater and cheater classes.



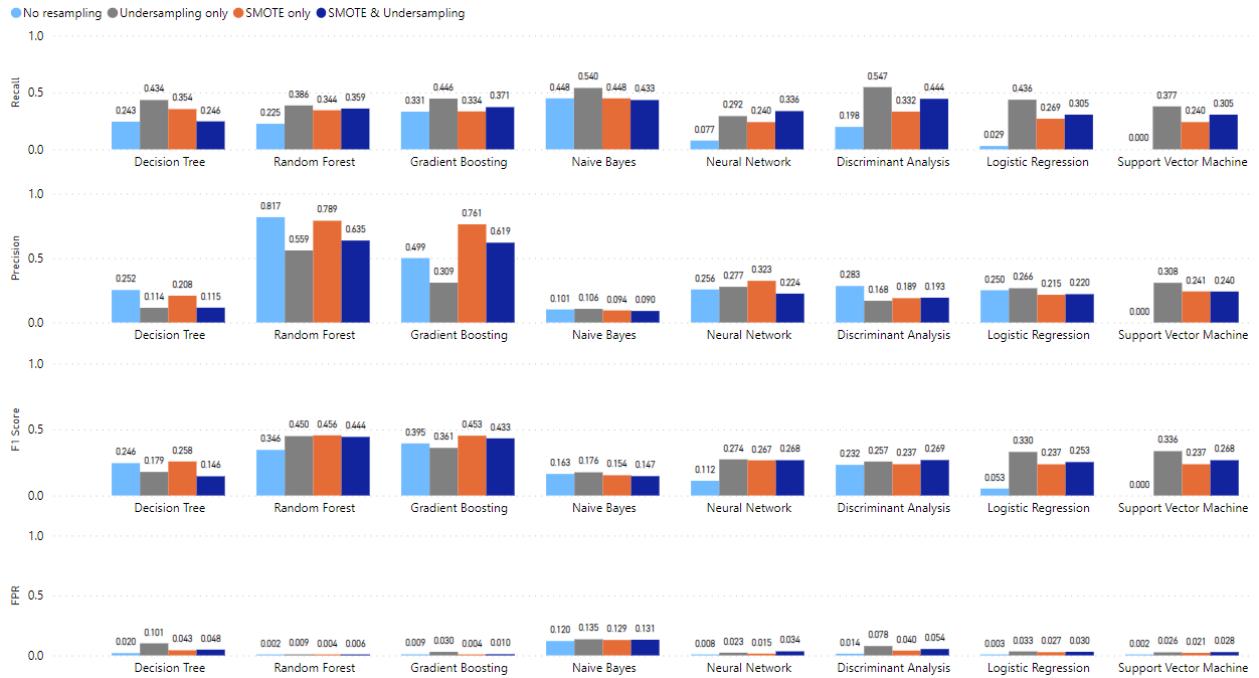
**Figure 16.** Performance comparison of models with and without resampling based on item response with a ratio of 5:1 between the non-cheater and cheater classes.



**Figure 17.** Performance comparison of models with and without resampling based on item response and summative statistics with a ratio of 5:1 between the non-cheater and cheater classes.



**Figure 18.** Performance comparison of models with and without resampling based on response time with a ratio of 5:1 between the non-cheater and cheater classes.



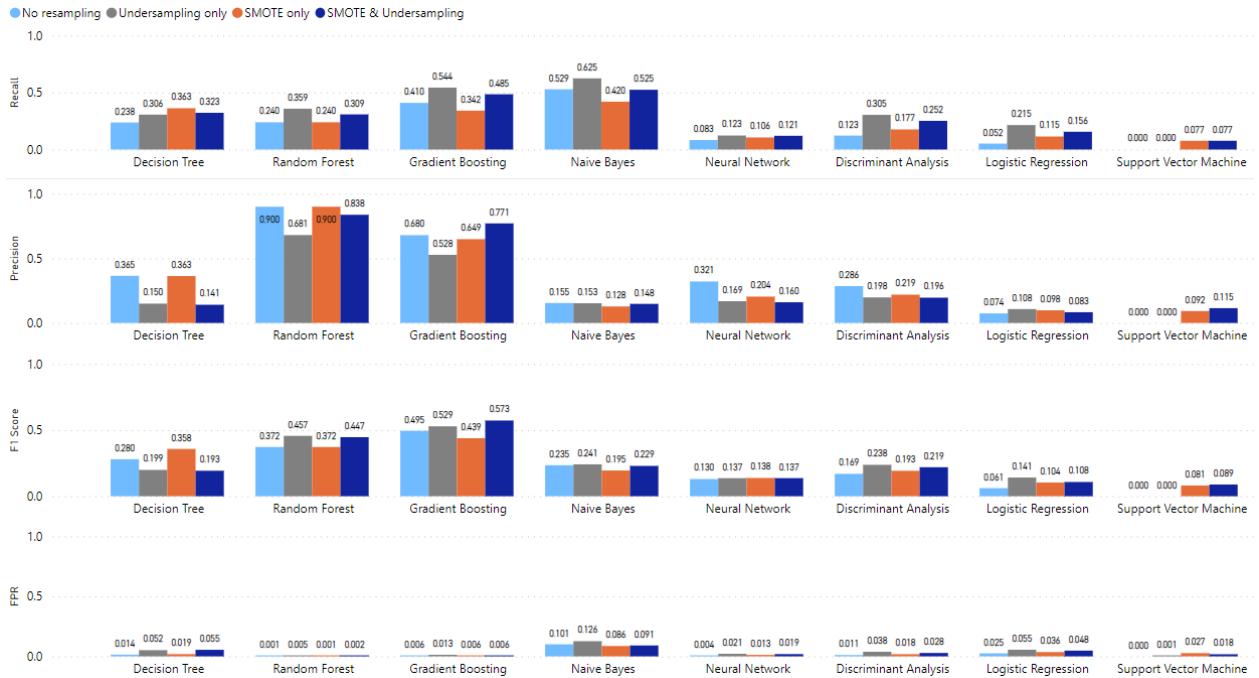
**Figure 19.** Performance comparison of models with and without resampling based on response time and summative statistics with a ratio of 5:1 between the non-cheater and cheater classes.



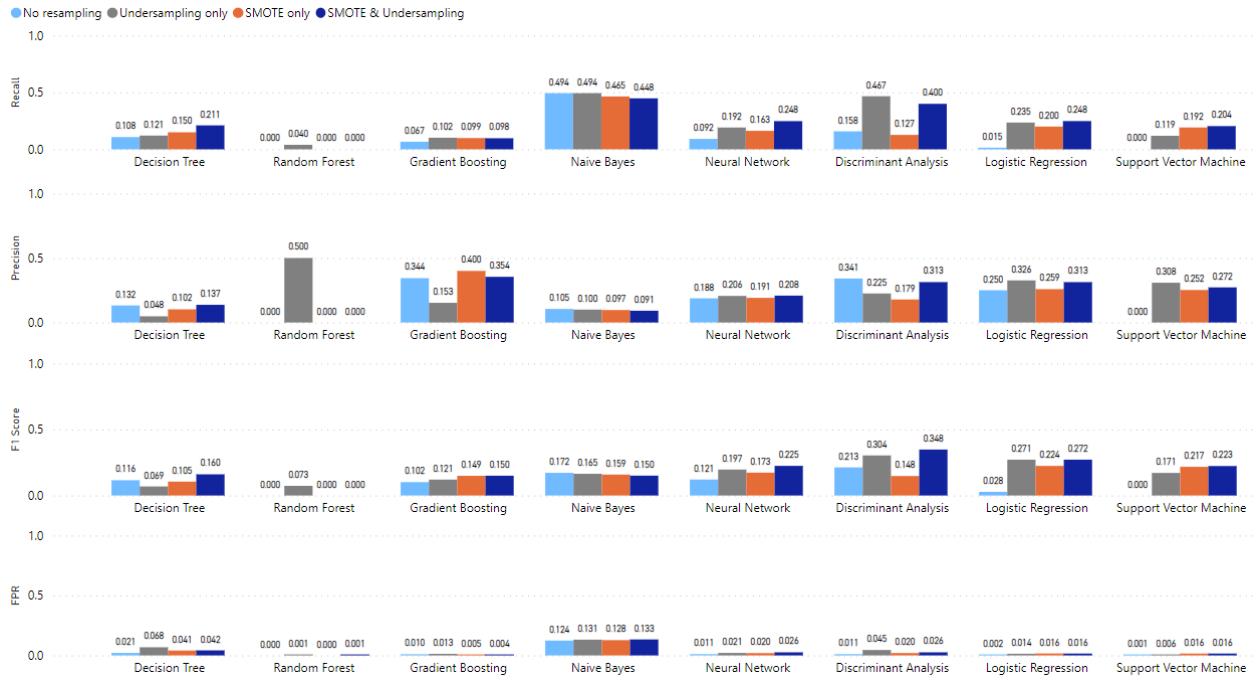
**Figure 20.** Performance comparison of models with and without resampling based on item response, response time with a ratio of 10:1 between the non-cheater and cheater classes.



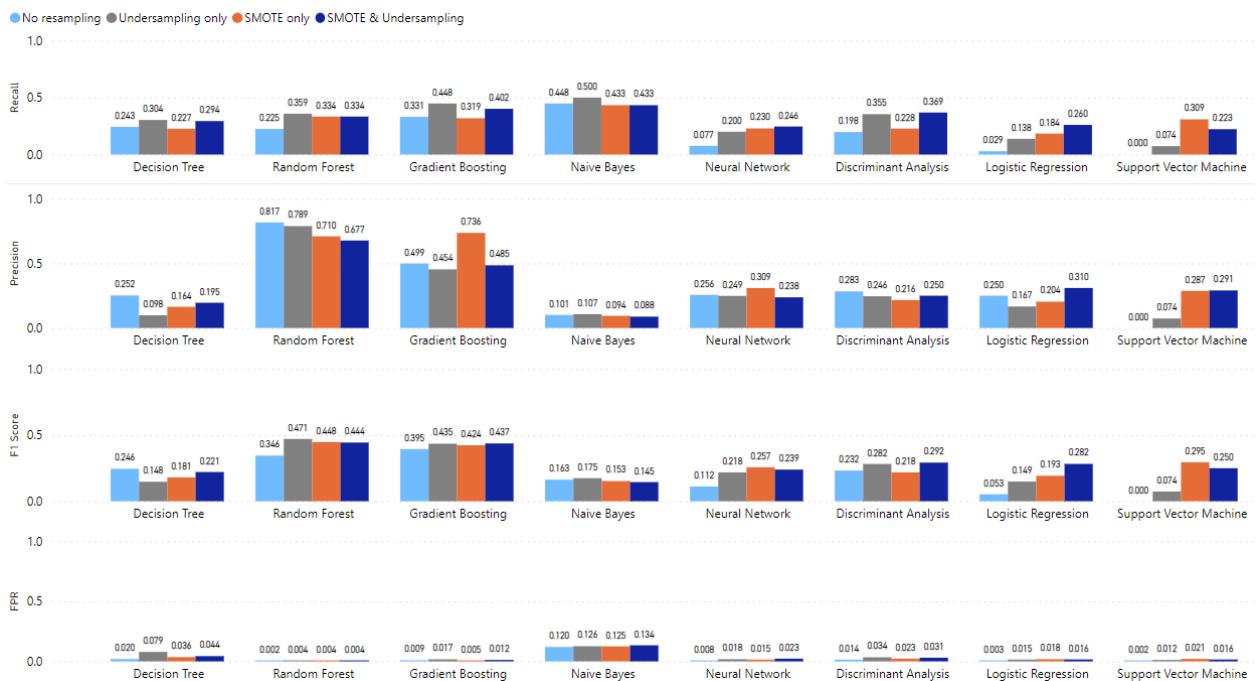
**Figure 21.** Performance comparison of models with and without resampling based on item response with a ratio of 10:1 between the non-cheater and cheater classes.



**Figure 22.** Performance comparison of models with and without resampling based on item response and summative statistics with a ratio of 10:1 between the non-cheater and cheater classes.



**Figure 23.** Performance comparison of models with and without resampling based on response time with a ratio of 10:1 between the non-cheater and cheater classes.

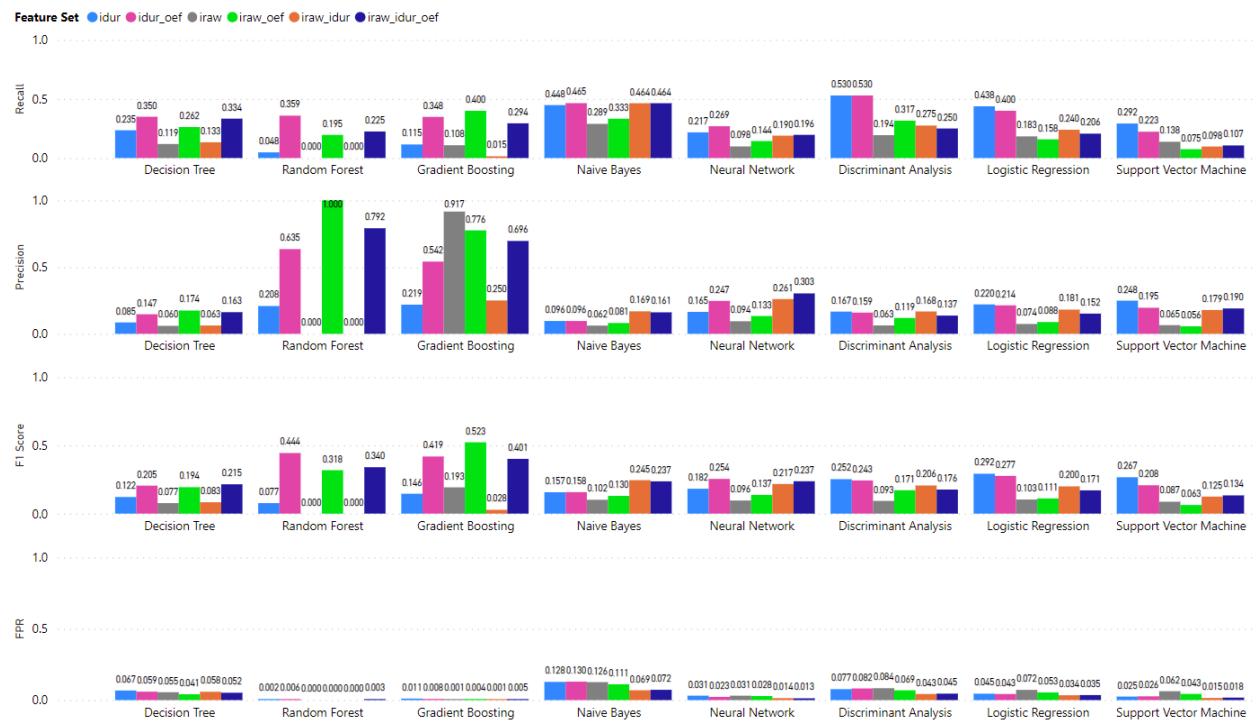


**Figure 24.** Performance comparison of models with and without resampling based on response time and summative statistics with a ratio of 10:1 between the non-cheater and cheater classes.

## Part 2 features



**Figure 25.** Model performance comparison with six different input features for dual resampling with a ratio of 1:1 between the non-cheater and cheater classes.



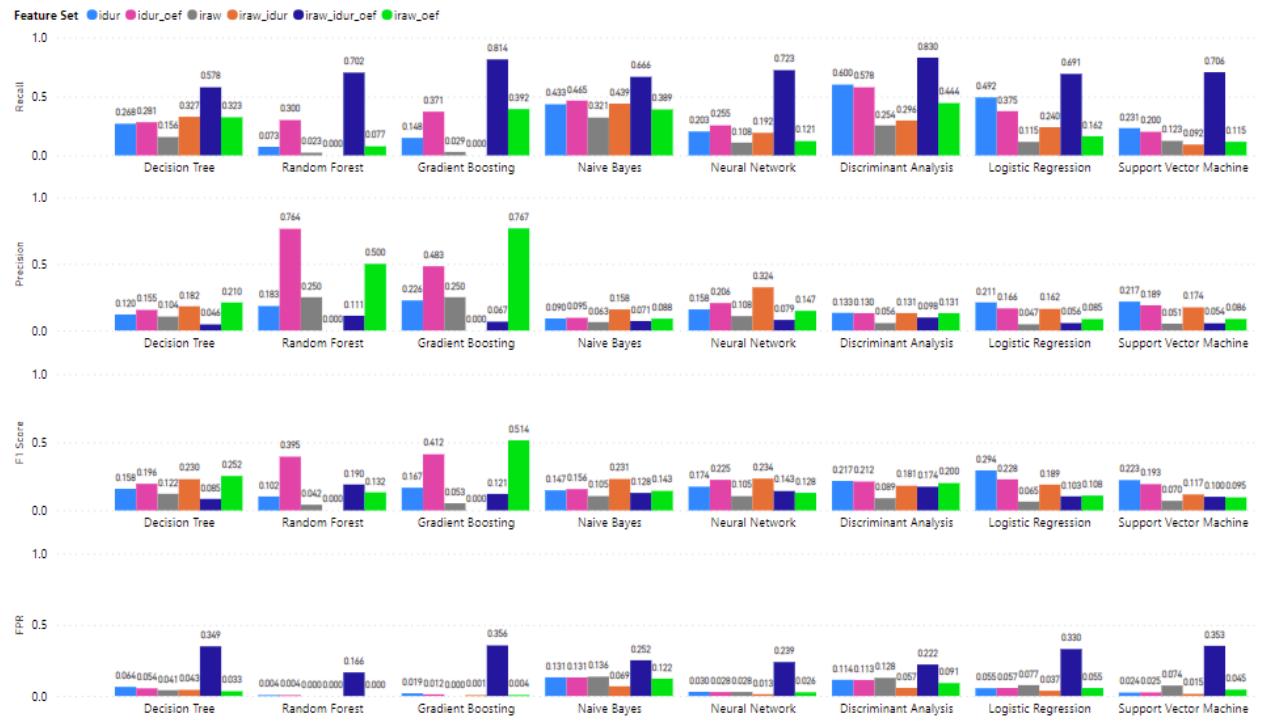
**Figure 26.** Model performance comparison with six different input features for dual resampling with a ratio of 2:1 between the non-cheater and cheater classes.



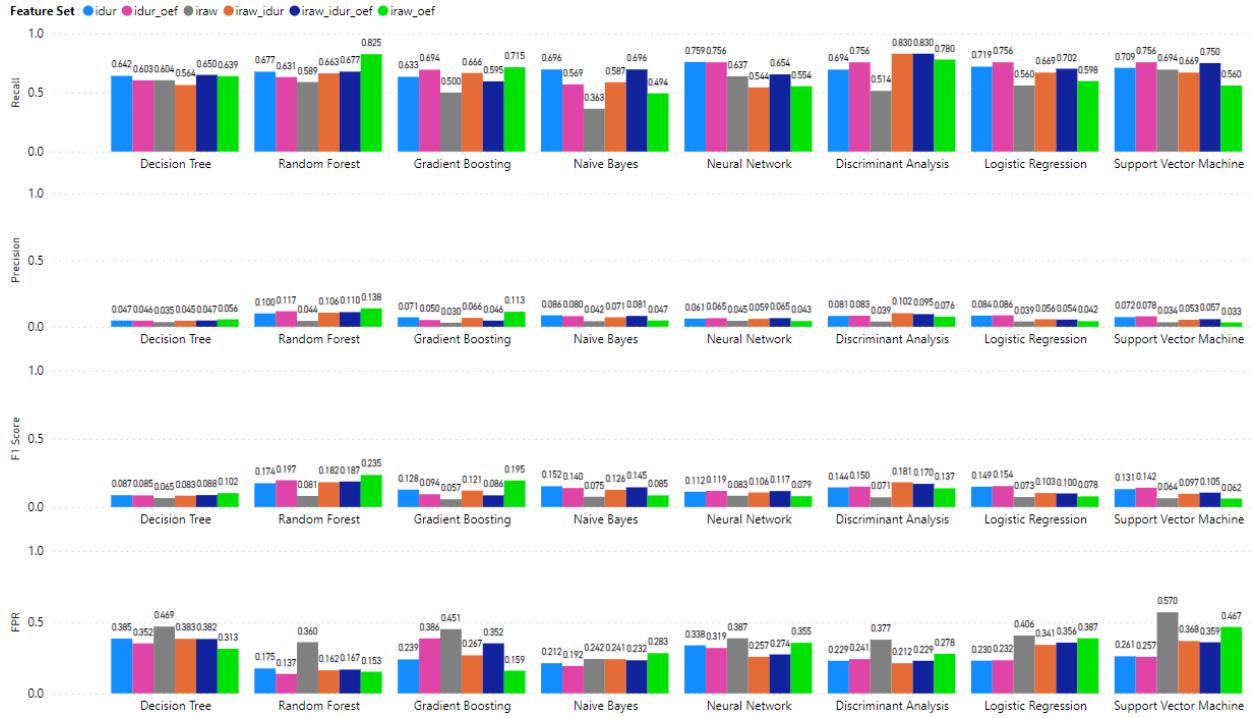
**Figure 27.** Model performance comparison with six different input features without resampling methods applied.



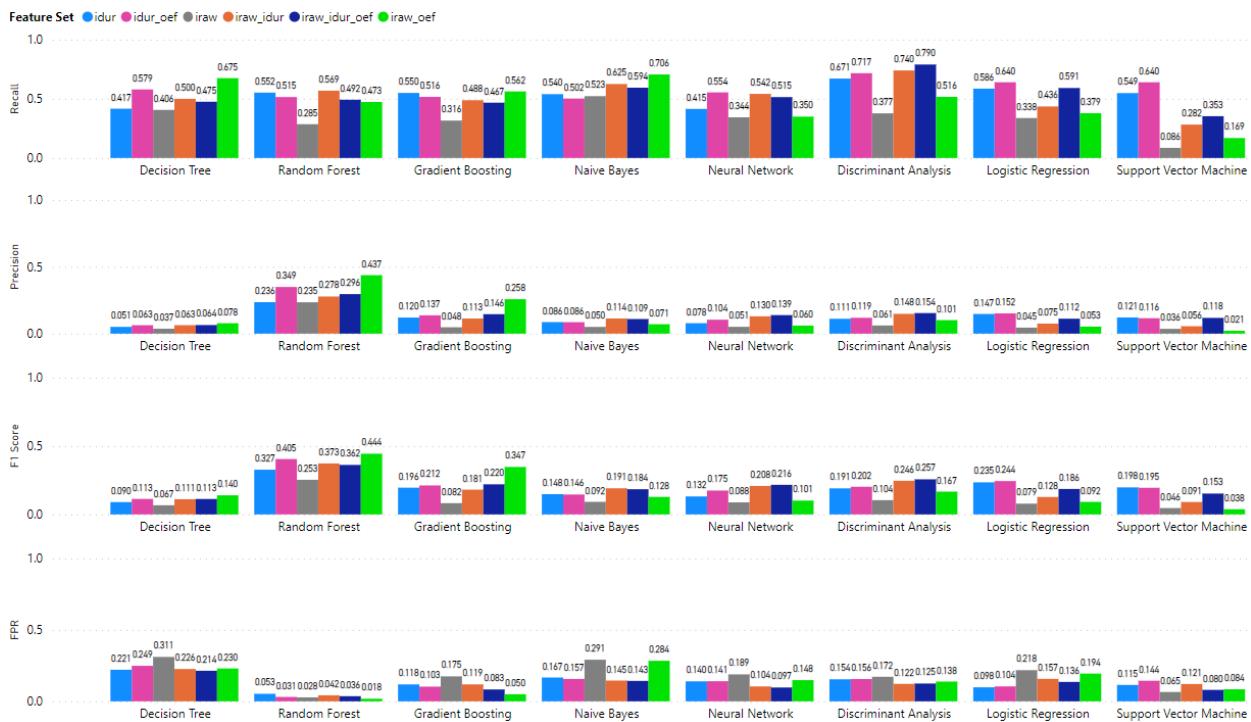
**Figure 28.** Model performance comparison with six different input features for oversampling SMOTE only with a ratio of 2:1 between the non-cheater and cheater classes.



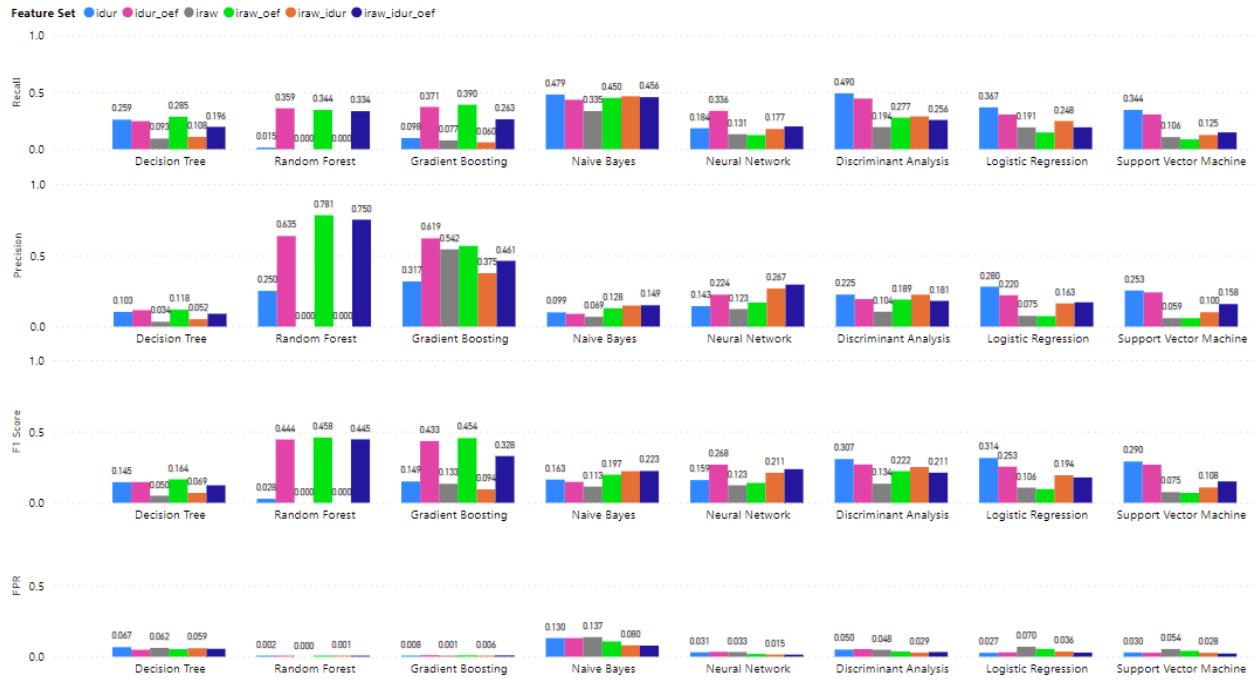
**Figure 29.** Model performance comparison with six different input features for oversampling SMOTE only with a ratio of 1:1 between the non-cheater and cheater classes.



**Figure 30.** Model performance comparison with six different input features for under-sampling only with a ratio of 1:1 between the non-cheater and cheater classes.



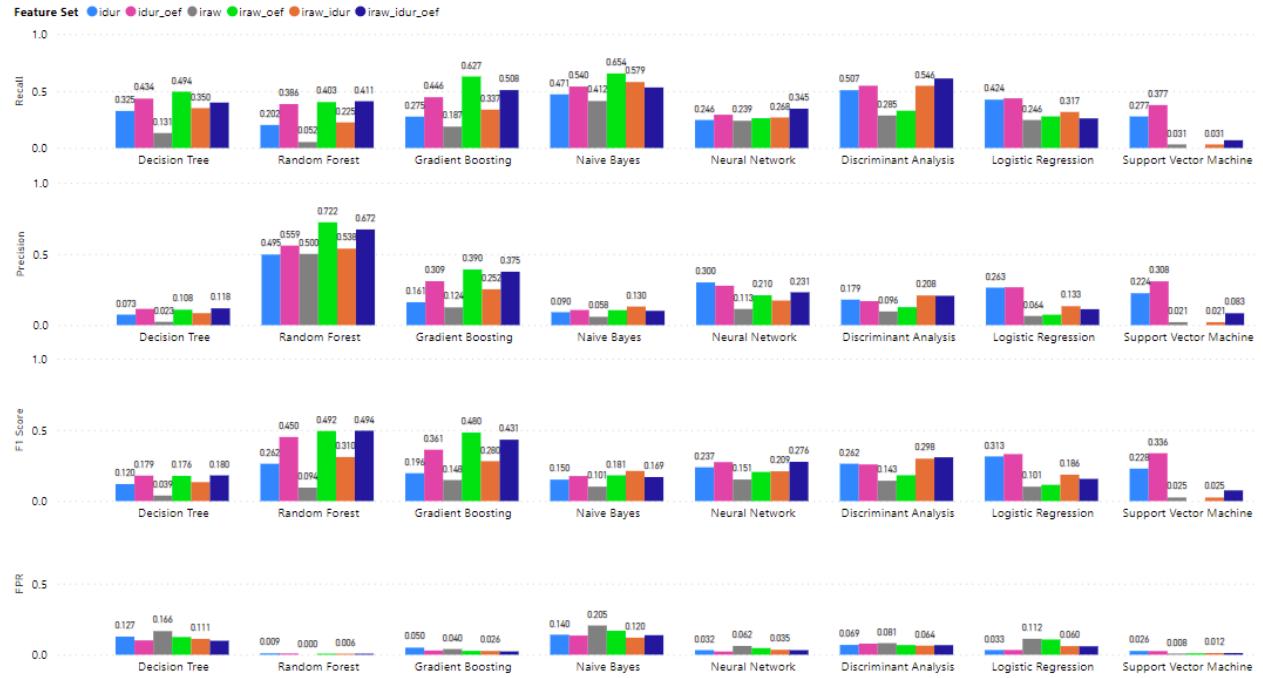
**Figure 31.** Model performance comparison with six different input features for under-sampling only with a ratio of 2:1 between the non-cheater and cheater classes.



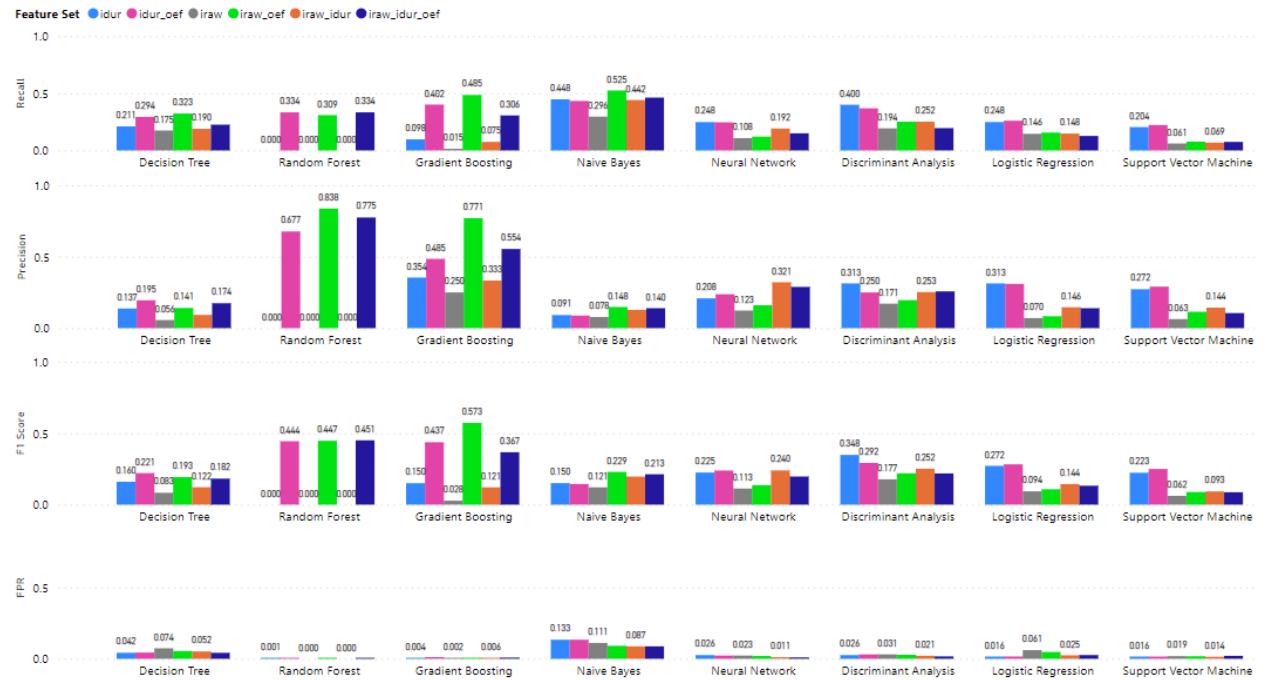
**Figure 32.** Model performance comparison with six different input features for a dual resampling with a ratio of 5:1 between the non-cheater and cheater classes.



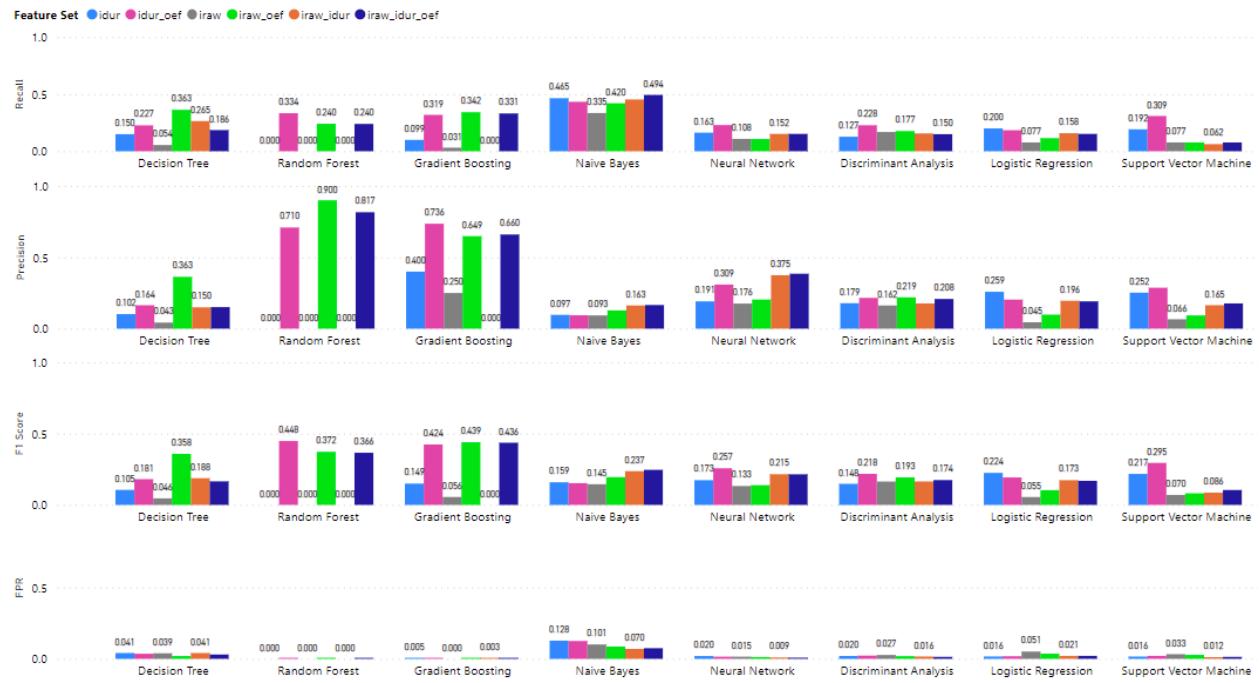
**Figure 33.** Model performance comparison with six different input features for oversampling SMOTE only with a ratio of 5:1 between the non-cheater and cheater classes.



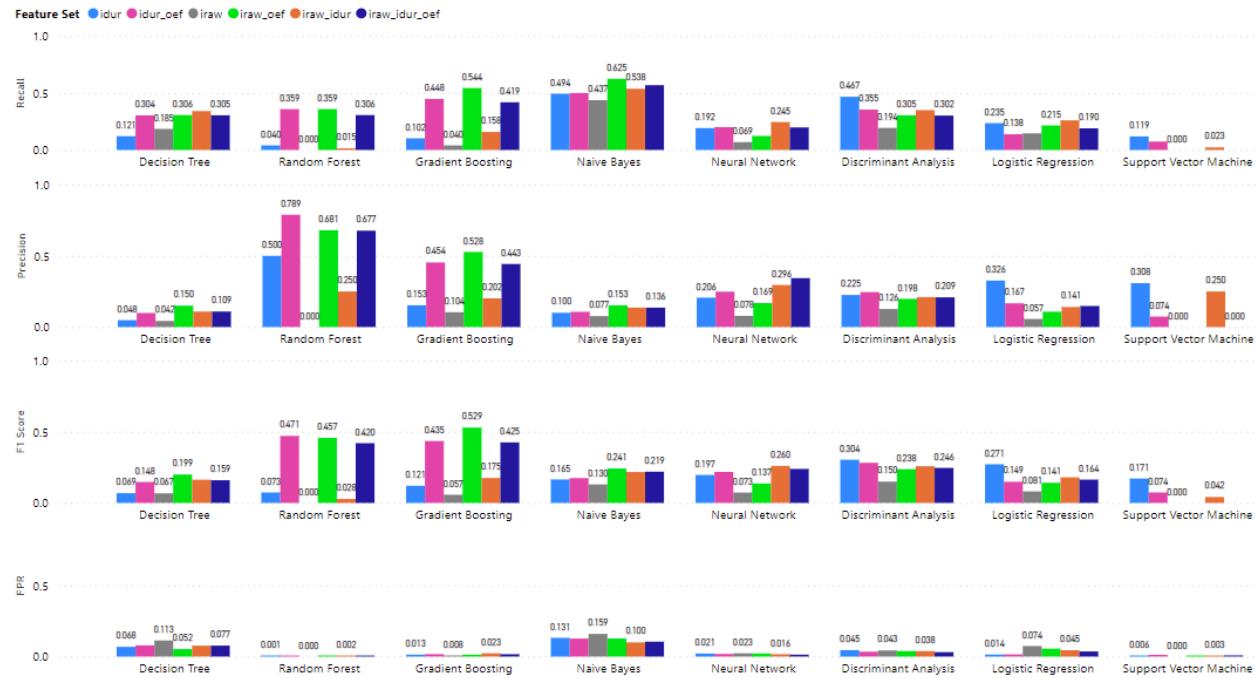
**Figure 34.** Model performance comparison with six different input features for under-sampling only with a ratio of 5:1 between the non-cheater and cheater classes.



**Figure 35.** Model performance comparison with six different input features for a dual resampling with a ratio of 10:1 between the non-cheater and cheater classes.

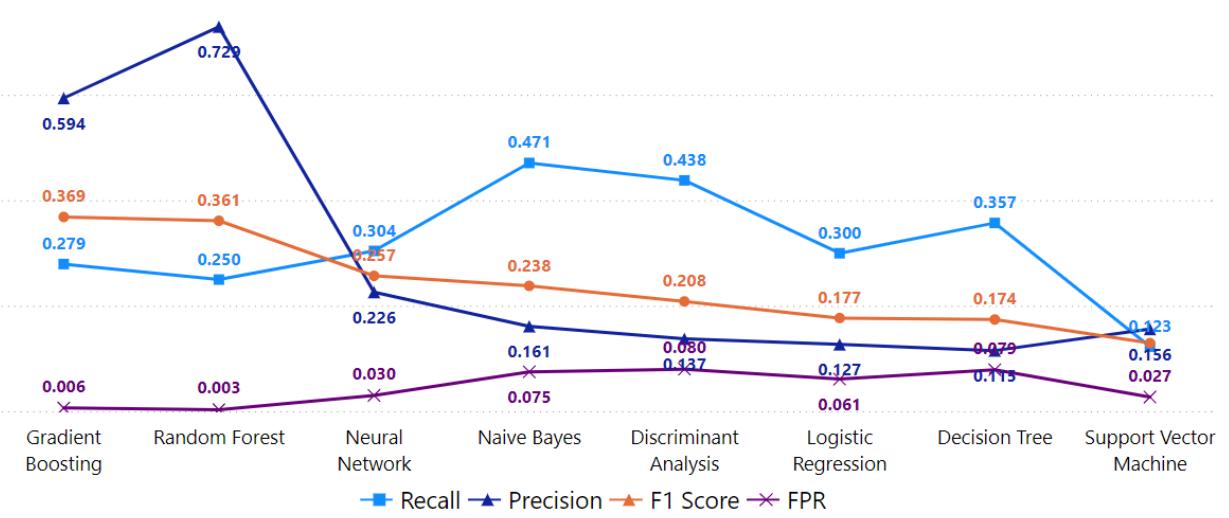


**Figure 36.** Model performance comparison with six different input features for oversampling SMOTE only with a ratio of 10:1 between the non-cheater and cheater classes.

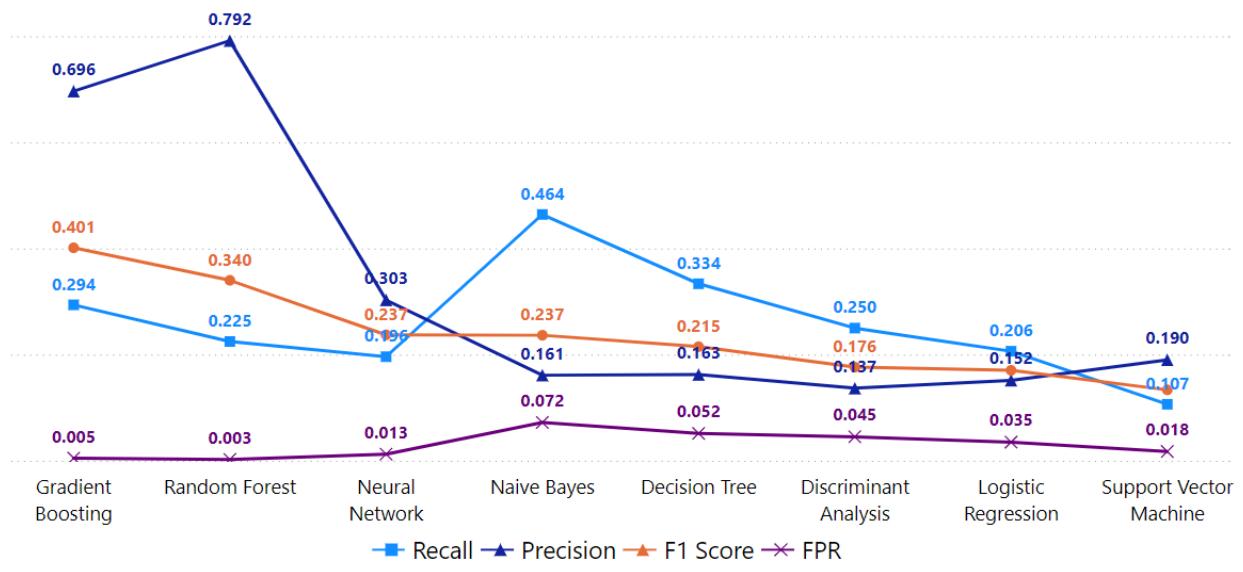


**Figure 37.** Model performance comparison with six different input features for under-sampling only with a ratio of 10:1 between the non-cheater and cheater classes.

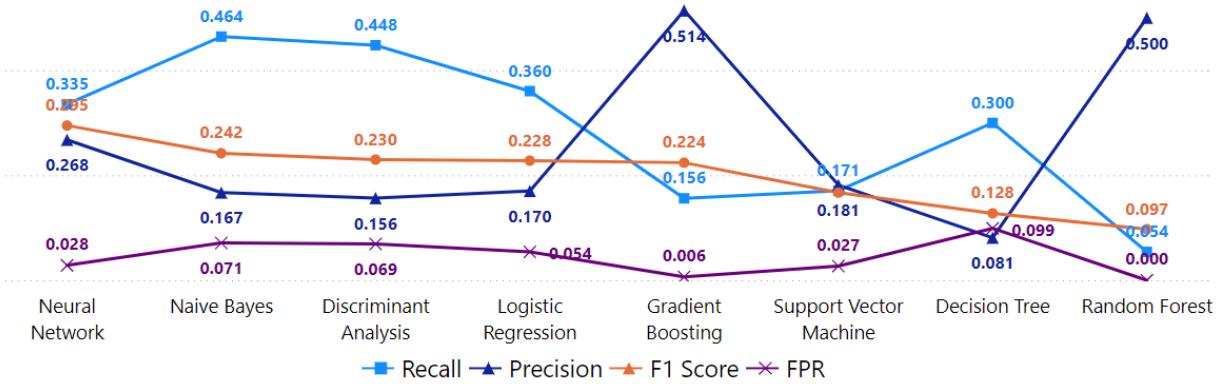
### Part 3 – eight base models



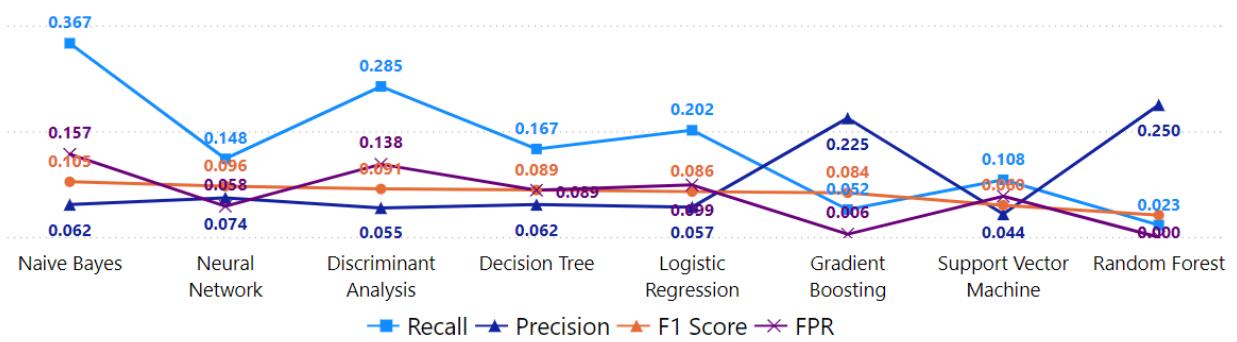
**Figure 38.** Base model comparison based on all features with a dual resampling ratio of 1:1 between the non-cheater and cheater classes.



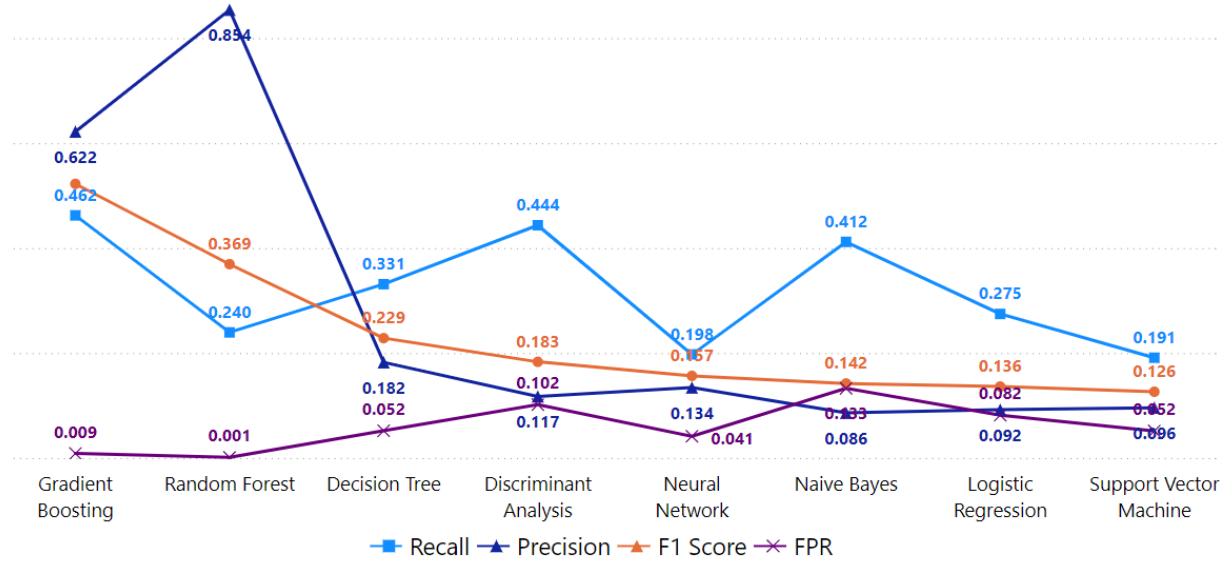
**Figure 39.** Base model comparison based on all features with a dual resampling ratio of 2:1 between the non-cheater and cheater classes.



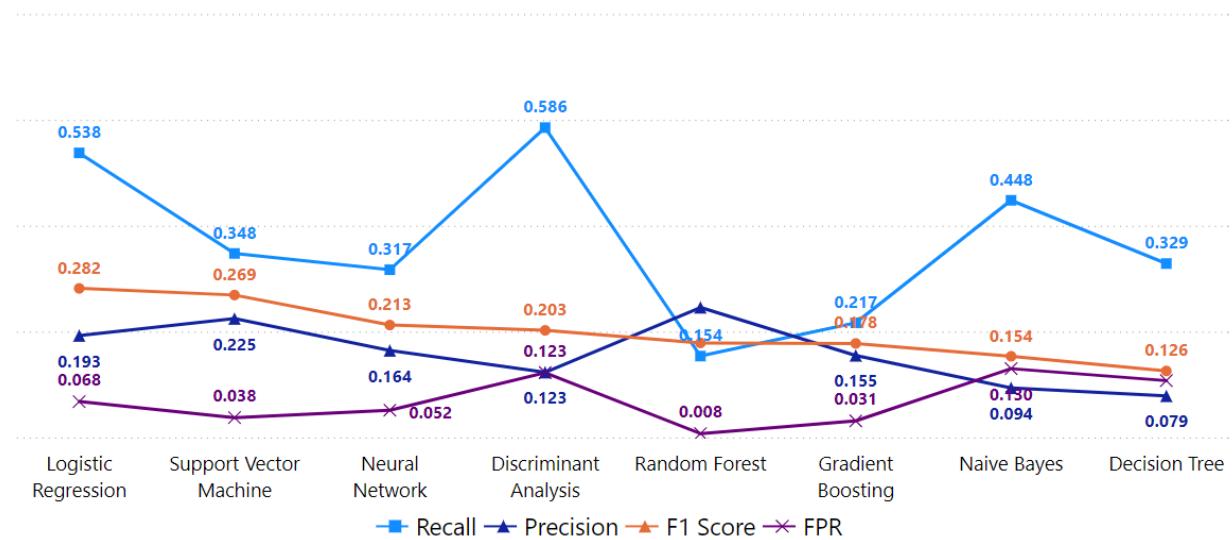
**Figure 40.** Base model comparison based on item response and response time with a dual resampling ratio of 1:1 between the non-cheater and cheater classes.



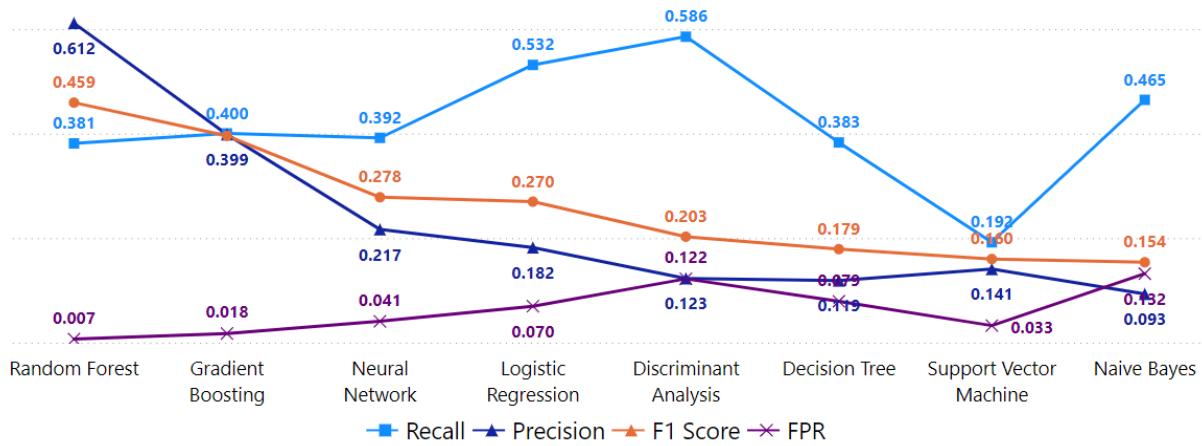
**Figure 41.** Base model comparison based on item response with a dual resampling ratio of 1:1 between the non-cheater and cheater classes.



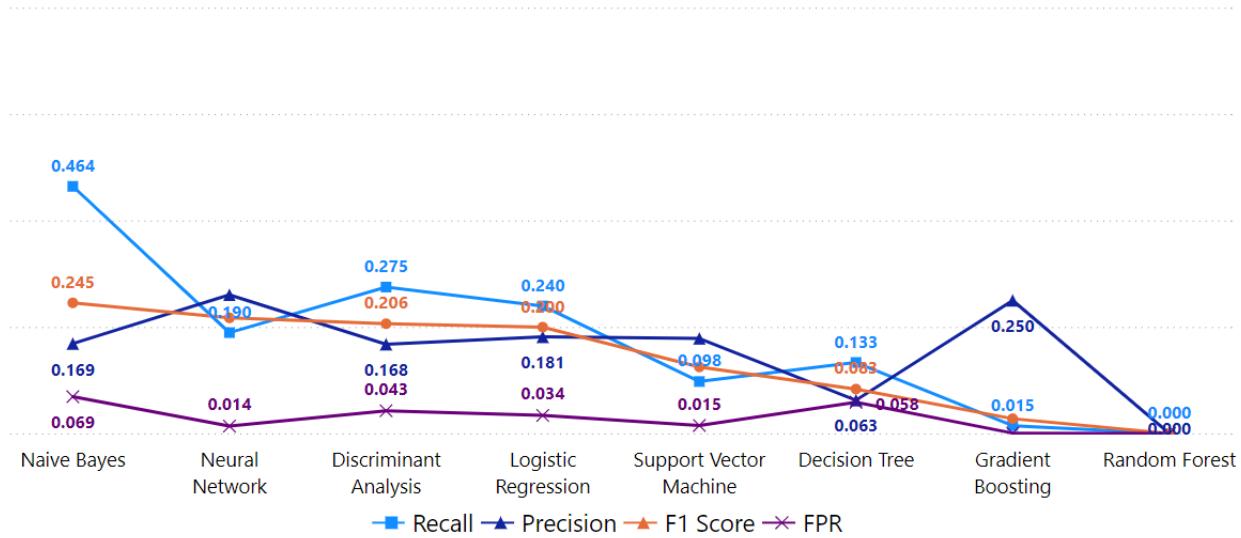
**Figure 42.** Base model comparison based on item response and summative statistics with a dual resampling ratio of 1:1 between the non-cheater and cheater classes.



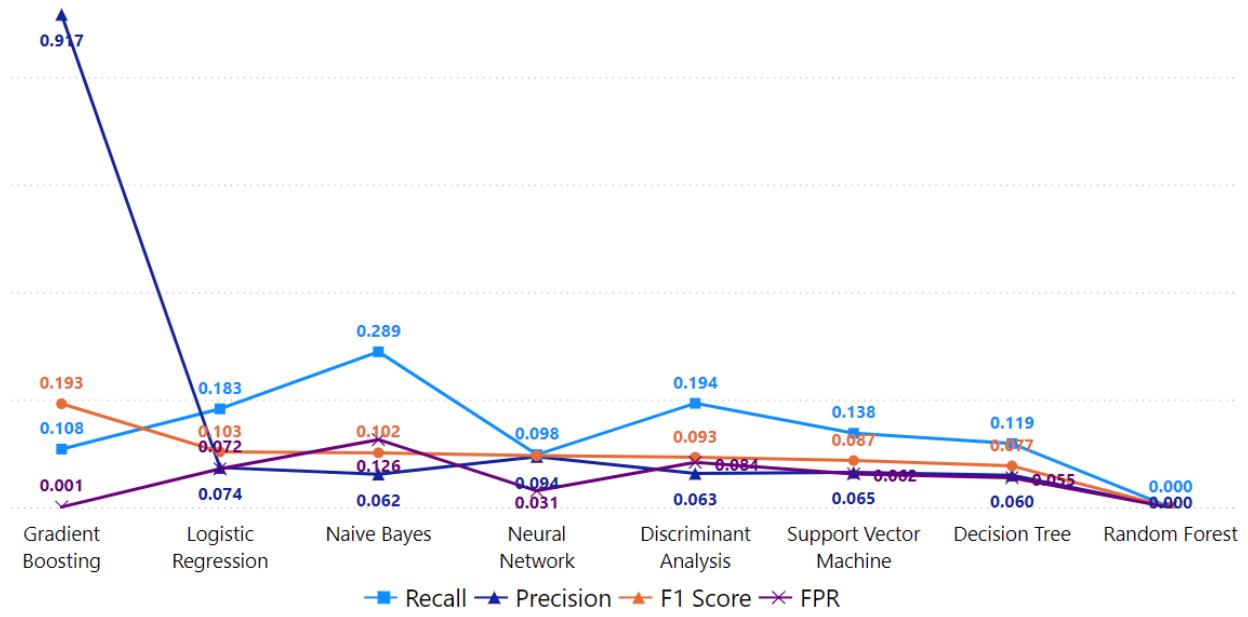
**Figure 43.** Base model comparison based on response time with a dual resampling ratio of 1:1 between the non-cheater and cheater classes.



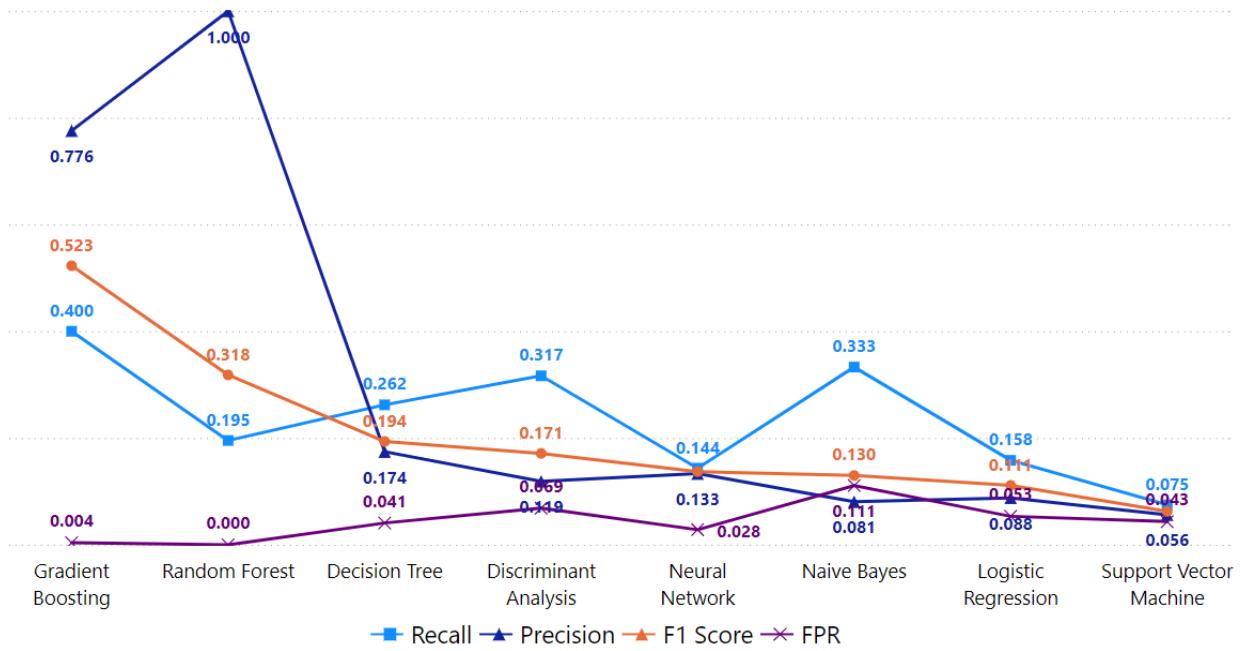
**Figure 44** Base model comparison based on response time and summative statistics with a dual resampling ratio of 1:1 between the non-cheater and cheater classes.



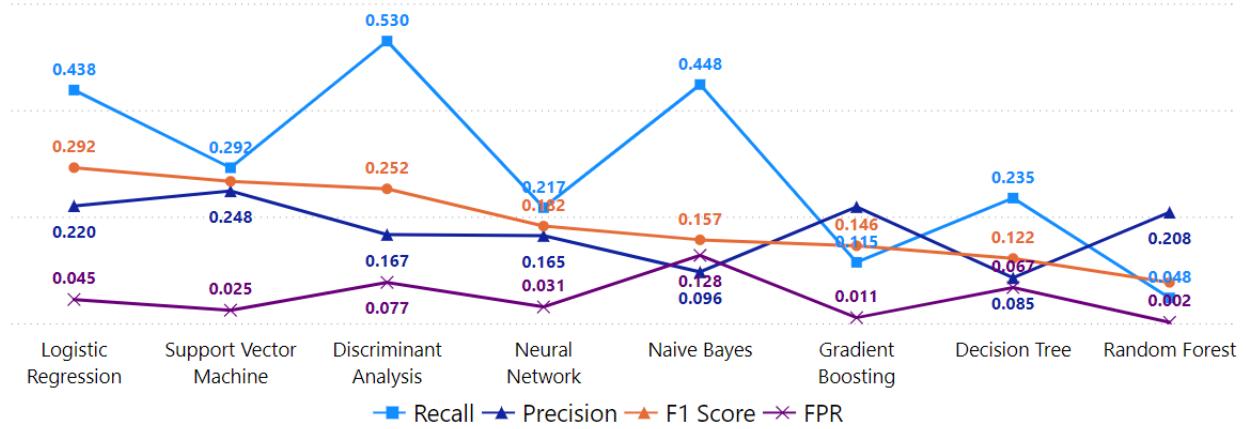
**Figure 45.** Base model comparison based on item response and response time with a dual resampling ratio of 2:1 between the non-cheater and cheater classes.



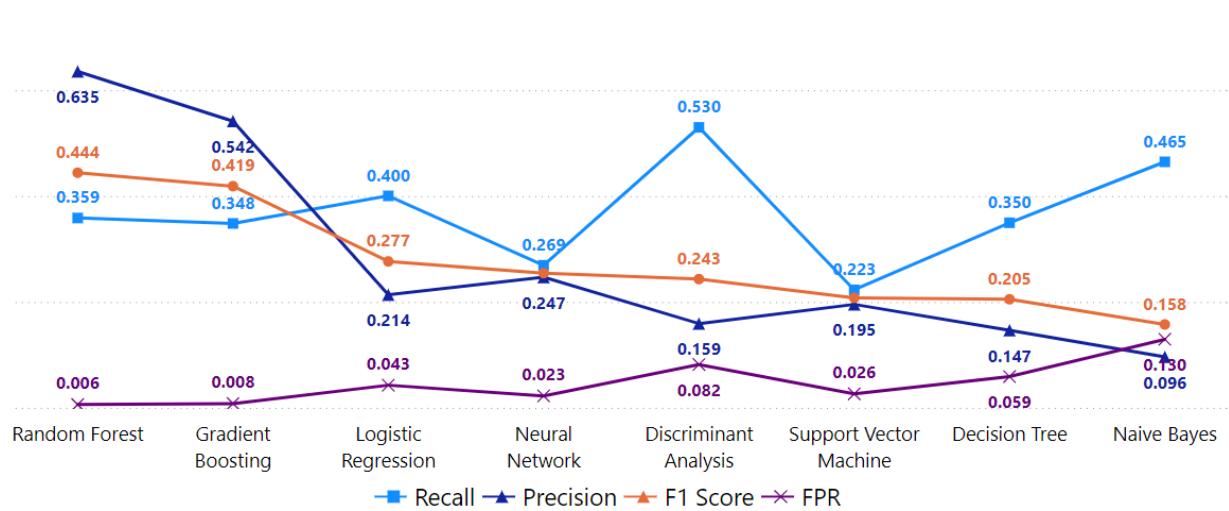
**Figure 46.** Base model comparison based on item response with a dual resampling ratio of 2:1 between the non-cheater and cheater classes.



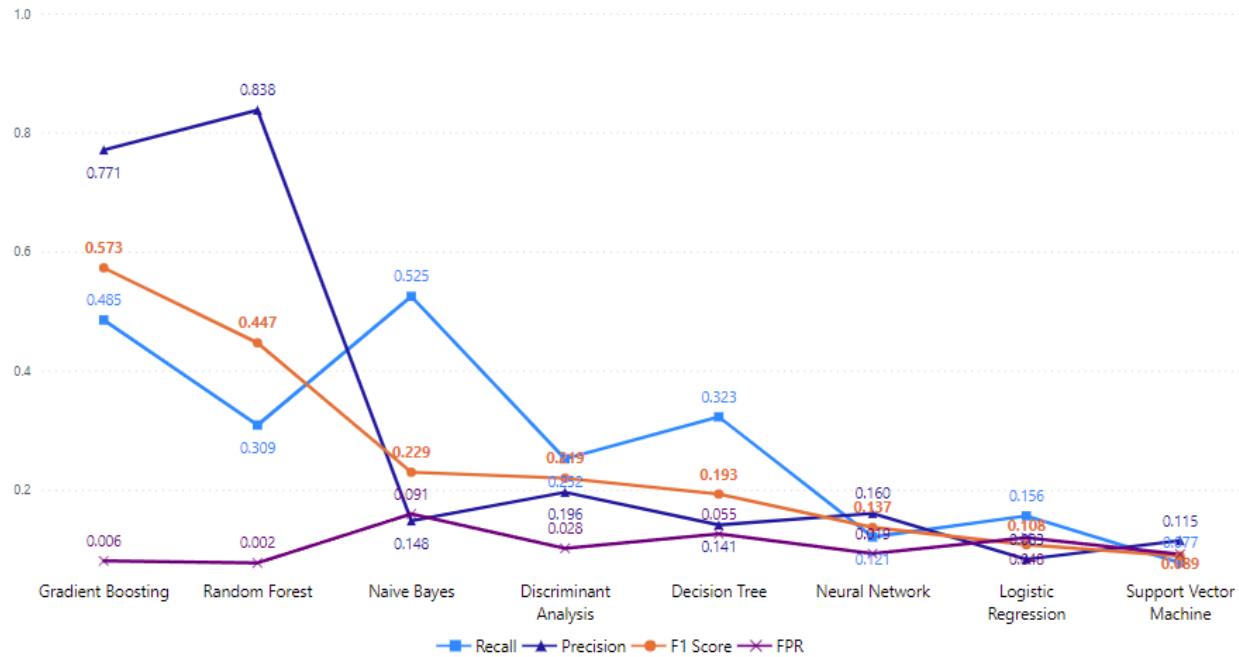
**Figure 47.** Base model comparison based on item response and summative statistics with a dual resampling ratio of 2:1 between the non-cheater and cheater classes.



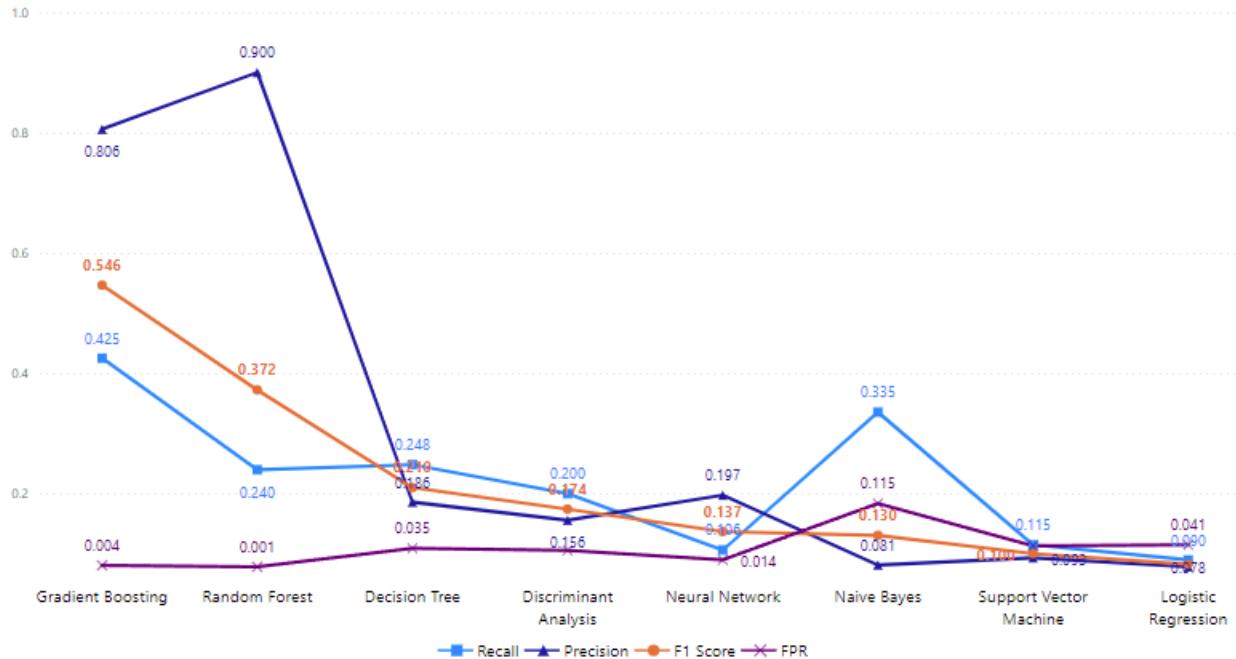
**Figure 48.** Base model comparison based on response time with a dual resampling ratio of 2:1 between the non-cheater and cheater classes.



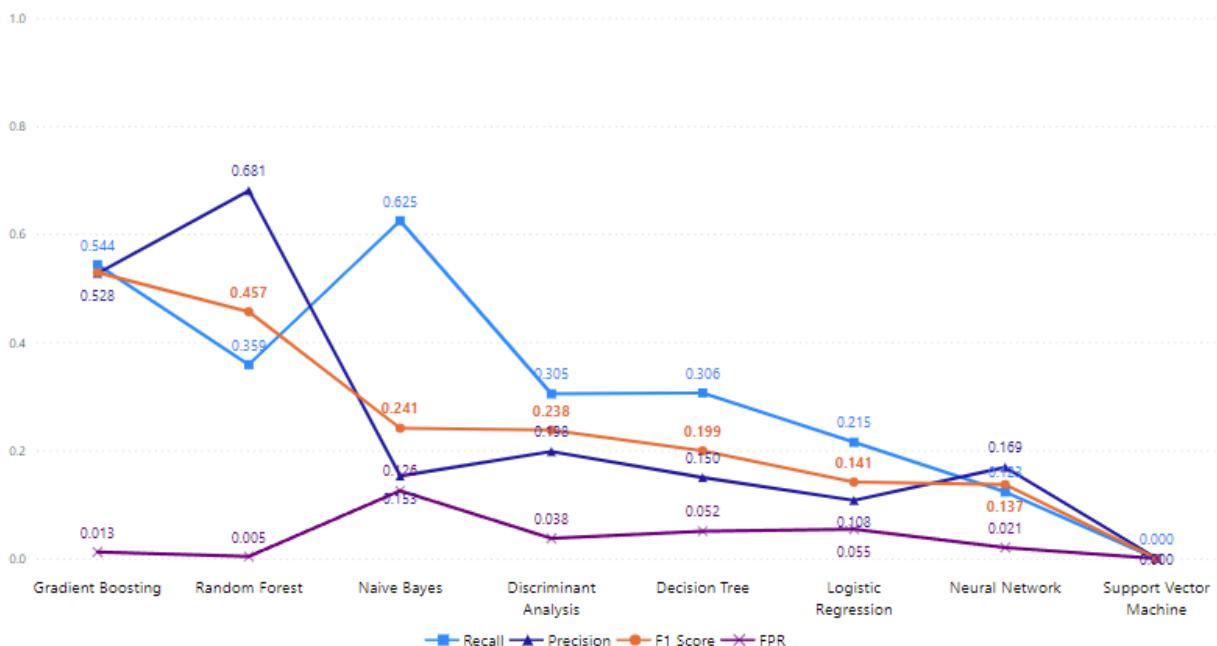
**Figure 49.** Base model comparison based on response time and summative statistics with a dual resampling ratio of 2:1 between the non-cheater and cheater classes.



**Figure 50.** Base model comparison based on item response and summative statistics with a dual resampling ratio of 10:1 between the non-cheater and cheater classes.

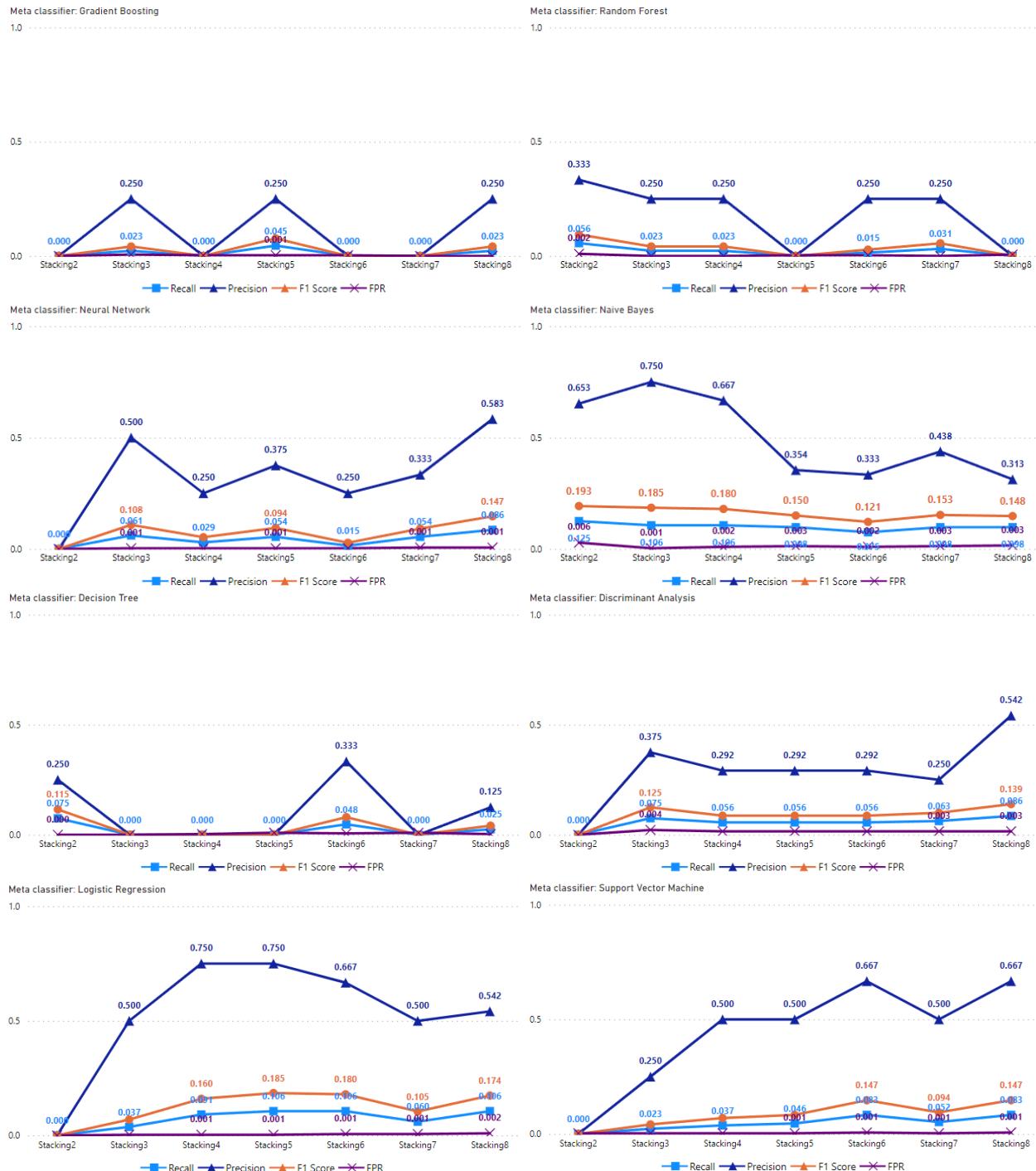


**Figure 51.** Base model comparison based on item response and summative statistics for oversampling SMOTE only with a ratio of 5:1 between the non-cheater and cheater classes.

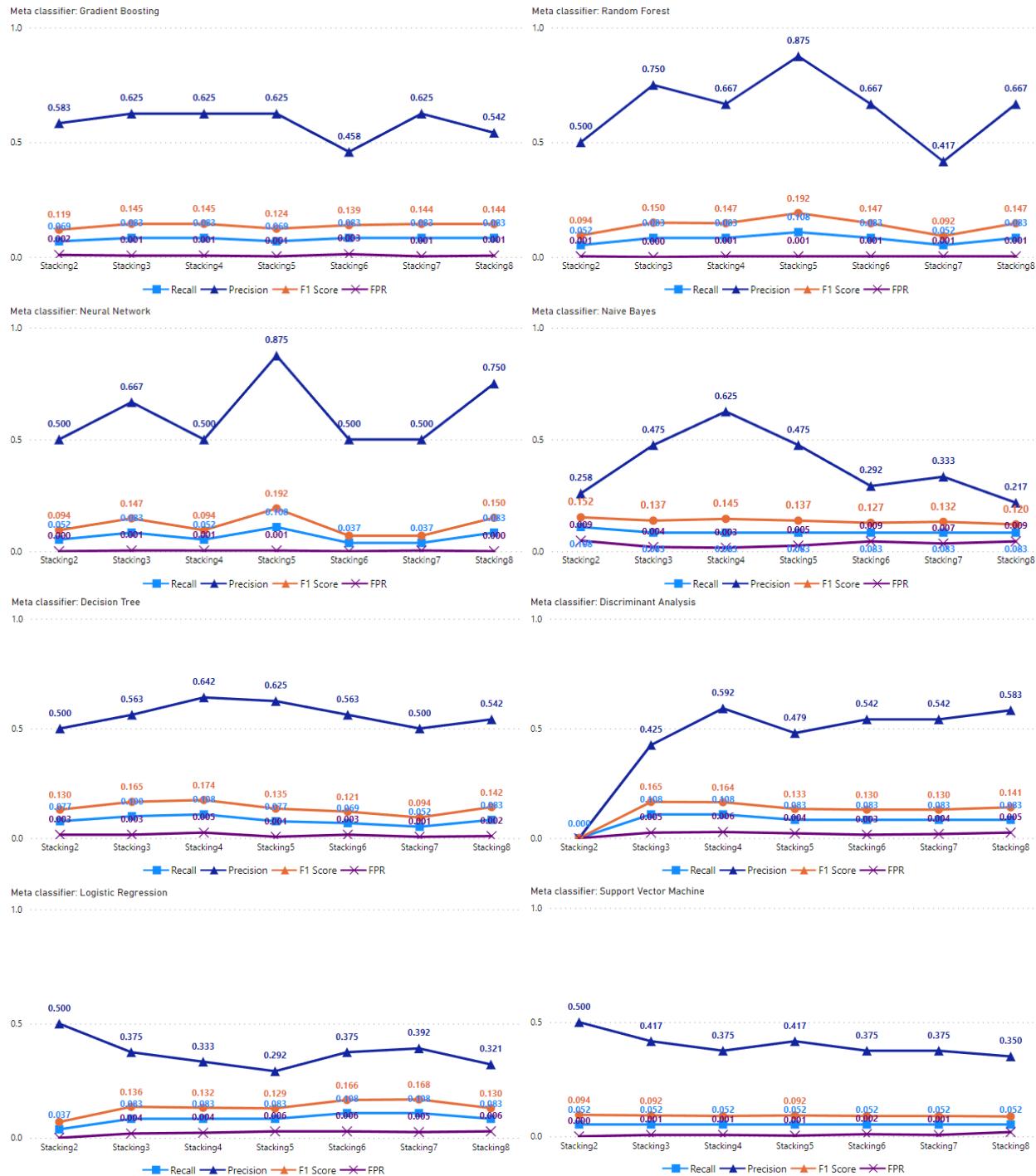


**Figure 52.** Base model comparison based on item response and summative statistics for undersampling only with a ratio of 10:1 between the non-cheater and cheater classes.

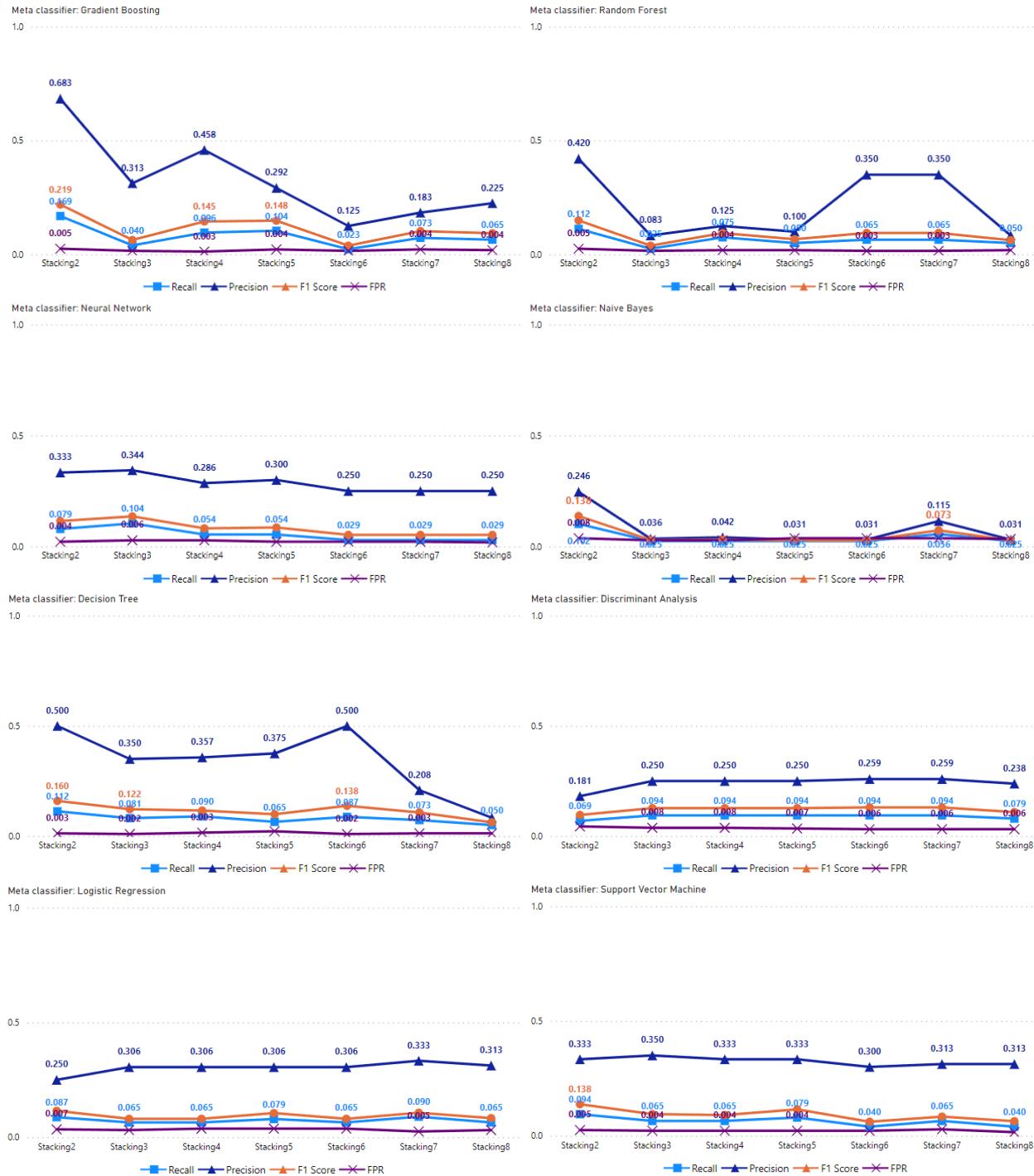
## Part 4. Stacking Only (1-component models)



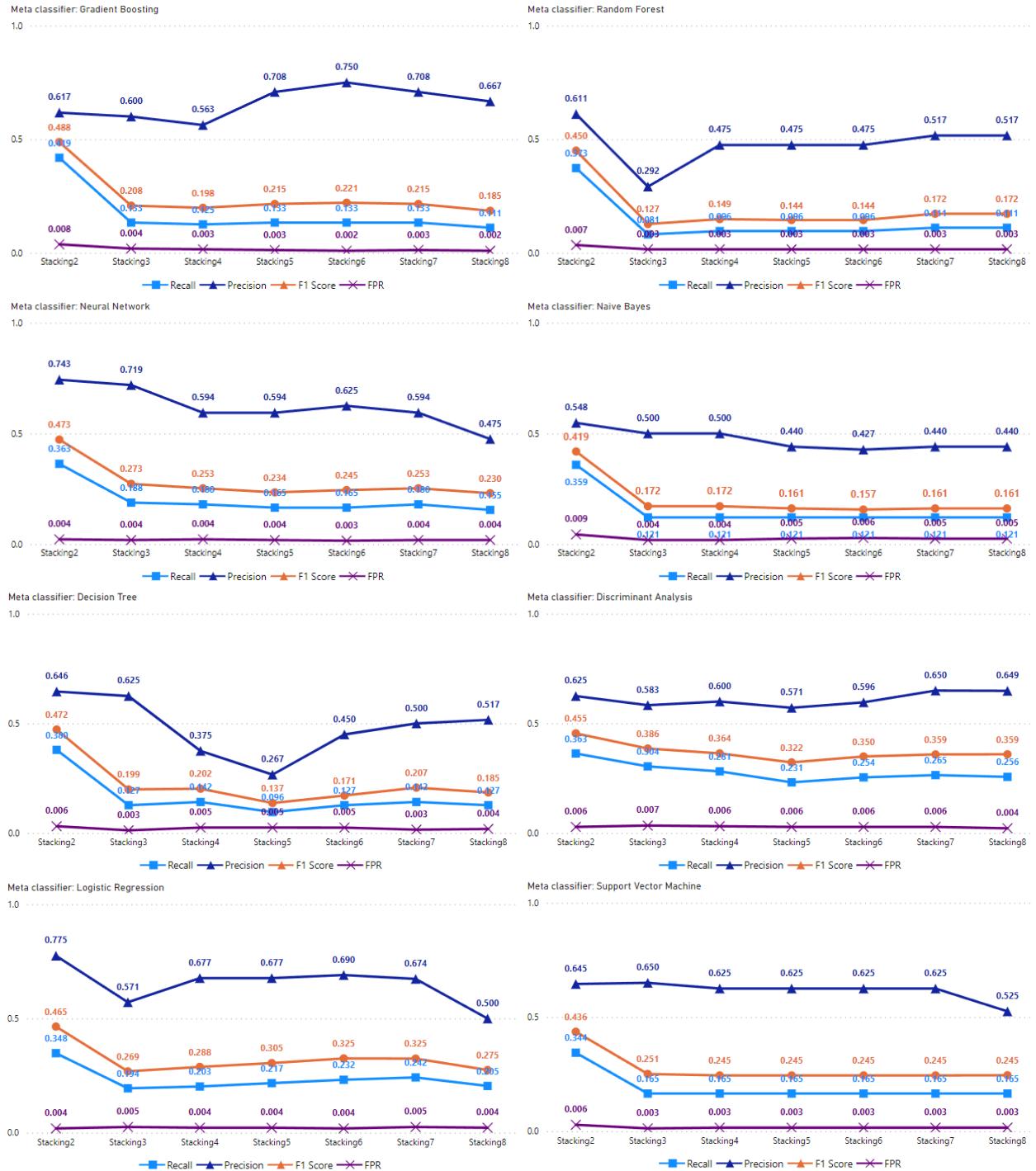
**Figure 53.** Stacking with seven different sets of rank-ordered base models based on item response and response time with a dual resampling ratio of 2:1 between the non-cheater and cheater classes.



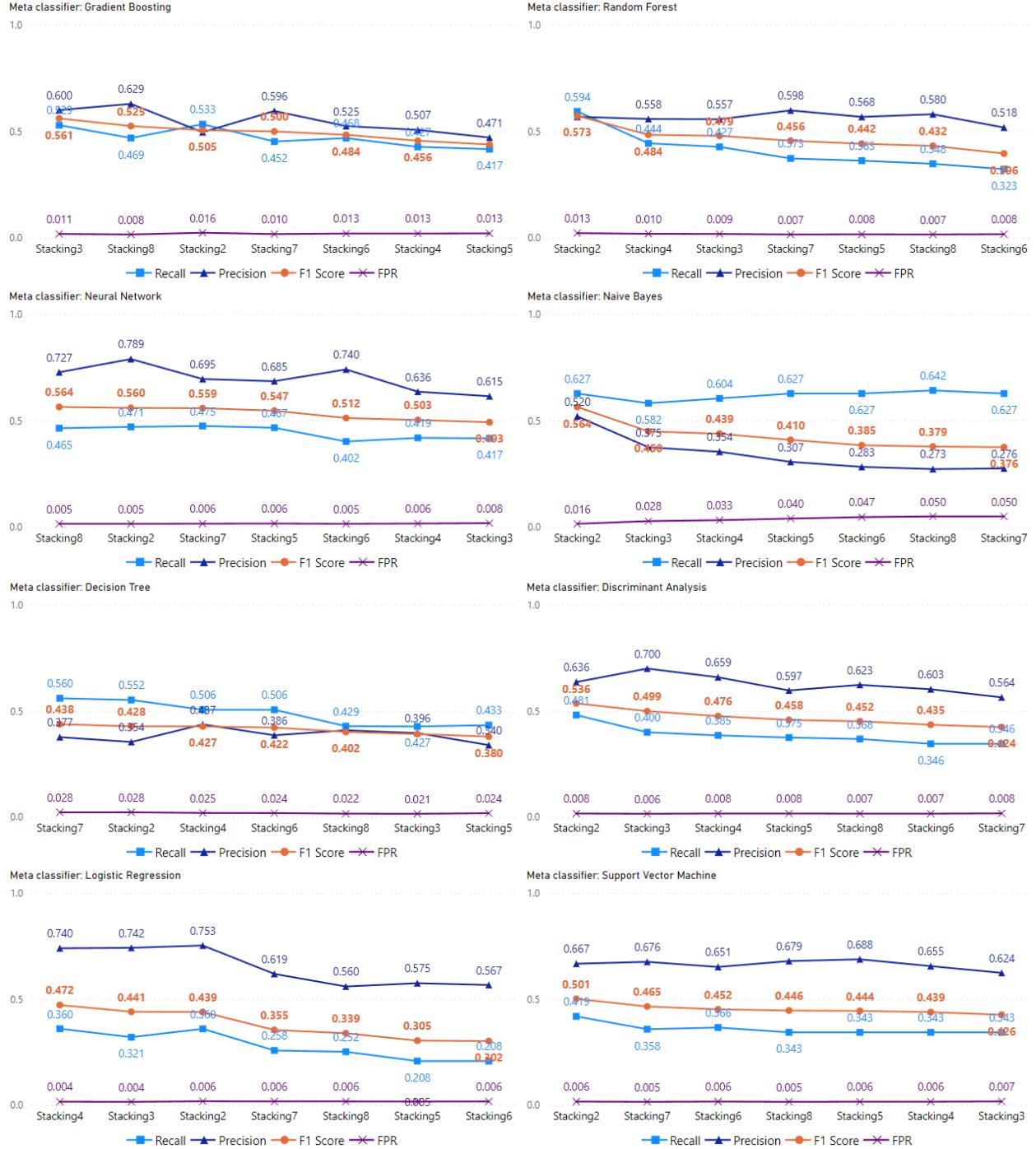
**Figure 54** Stacking with seven different sets of rank-ordered base models based on item response with a dual resampling ratio of 2:1 between the non-cheater and cheater classes.



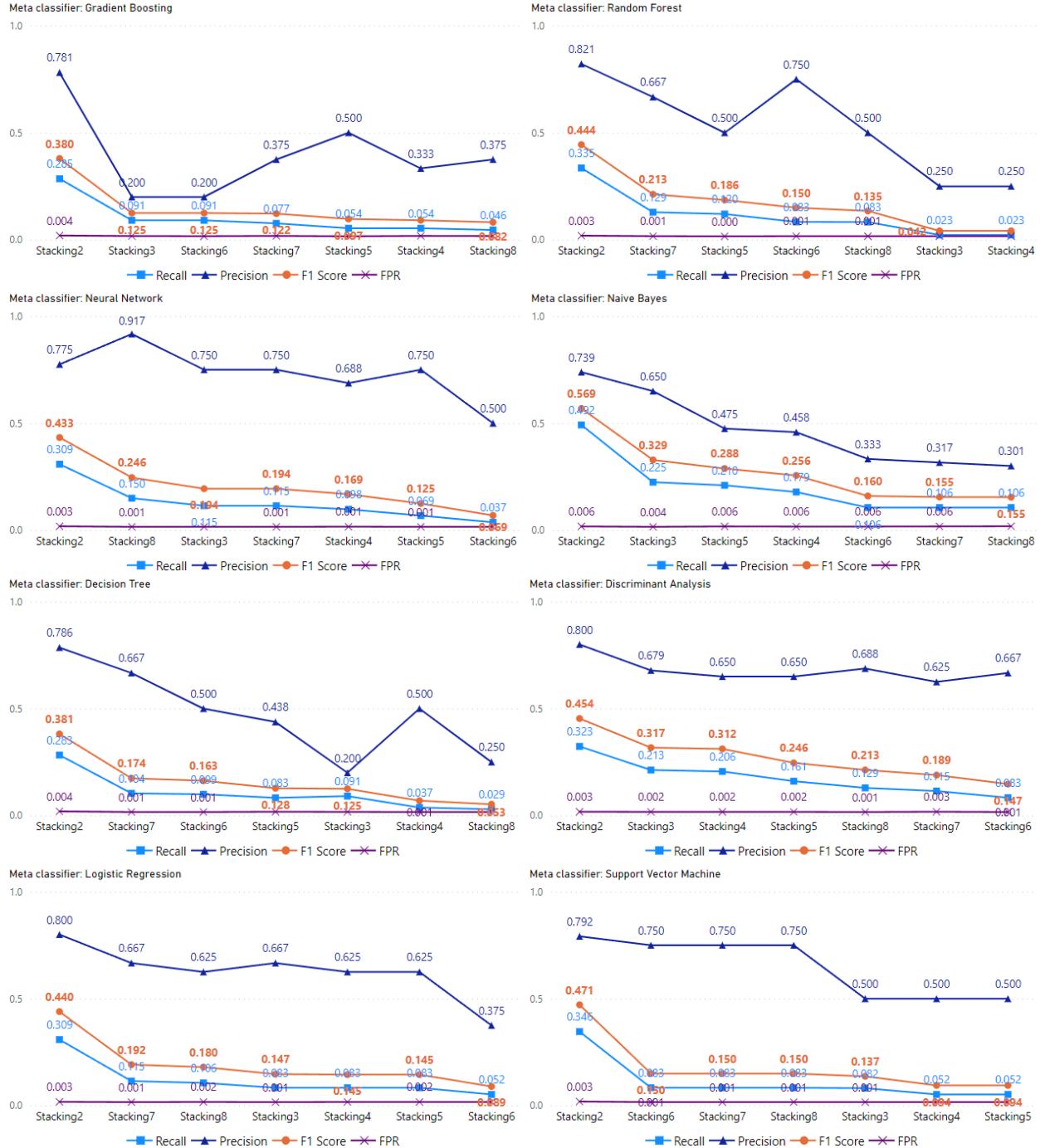
**Figure 55.** Stacking with seven different sets of rank-ordered base models based on response time with a dual resampling ratio of 2:1 between the non-cheater and cheater classes.



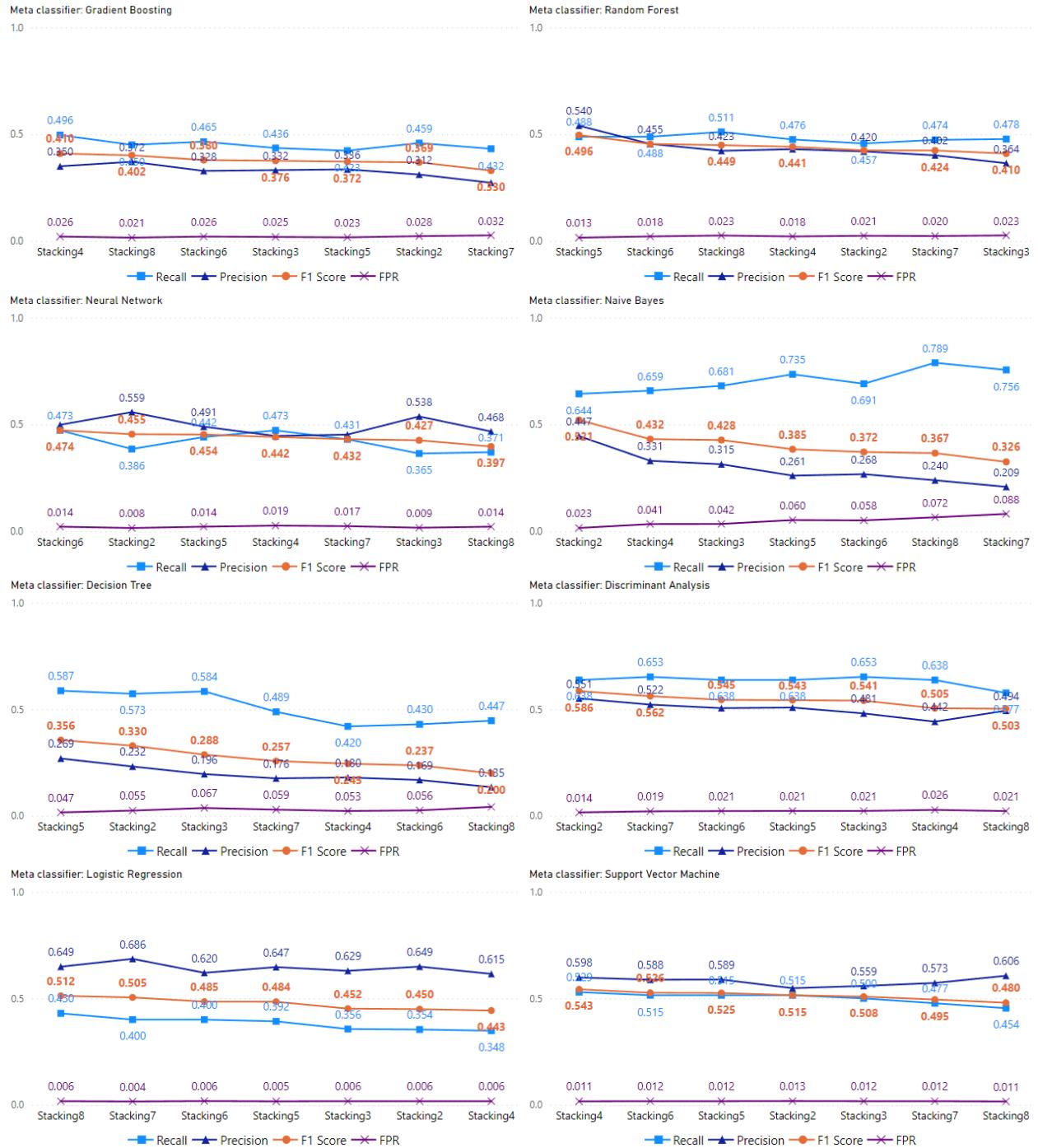
**Figure 56.** Stacking with seven different sets of rank-ordered base models based on response time and summative statistics with a dual resampling ratio of 2:1 between the non-cheater and cheater classes.



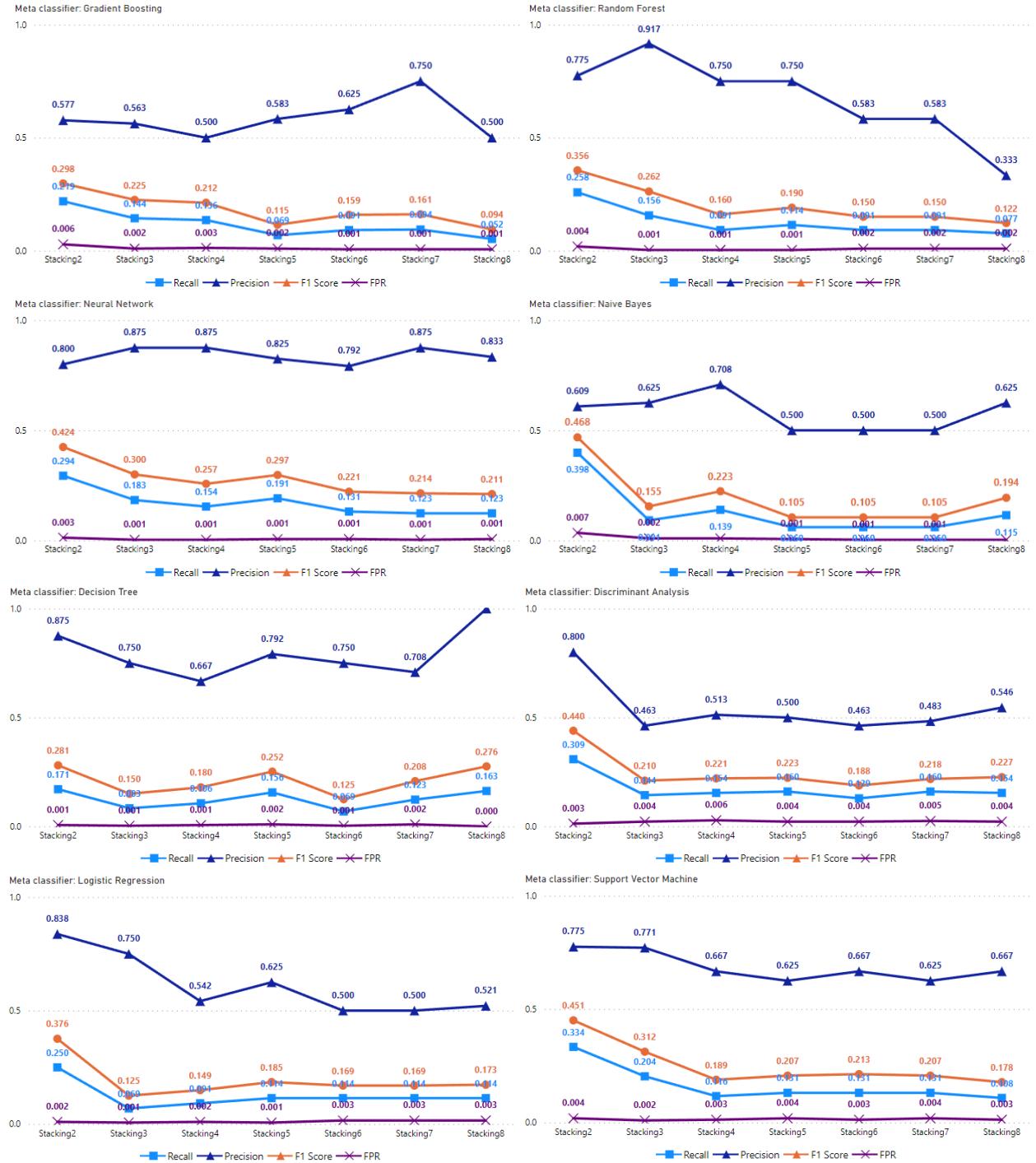
**Figure 57.** Stacking using different meta-models with seven different sets of rank-ordered base models based on item response and summative statistics with a dual resampling ratio of 10:1 between the non-cheater and cheater classes.



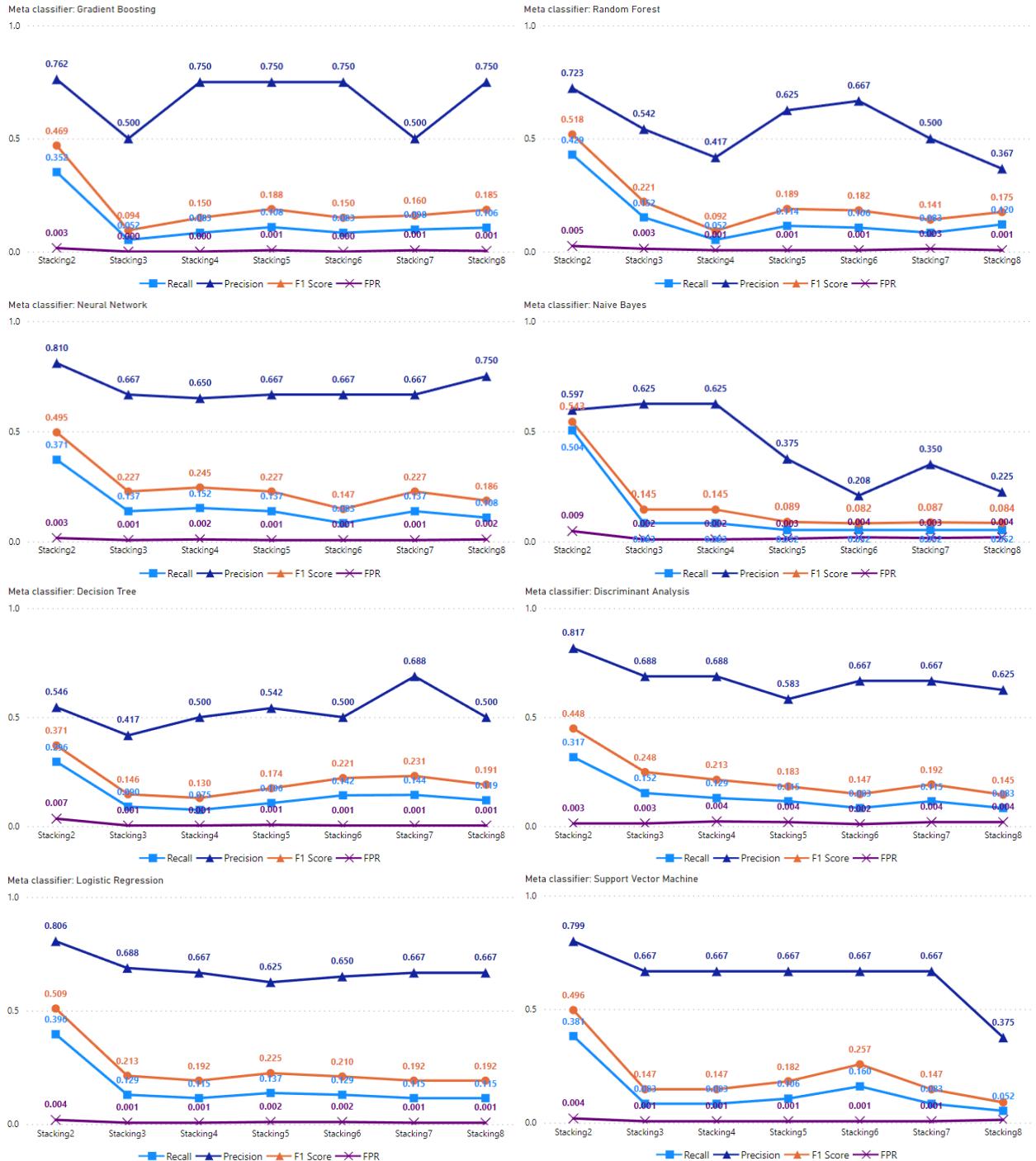
**Figure 58.** Stacking using different meta-models with seven different sets of rank-ordered base models based on item response and summative statistics for oversampling SMOTE only with a ratio of 5:1 between the non-cheater and cheater classes.



**Figure 59.** Stacking using different meta-models with seven different sets of rank-ordered base models based on item response and summative statistics for under-sampling only with a ratio of 10:1 between the non-cheater and cheater classes.

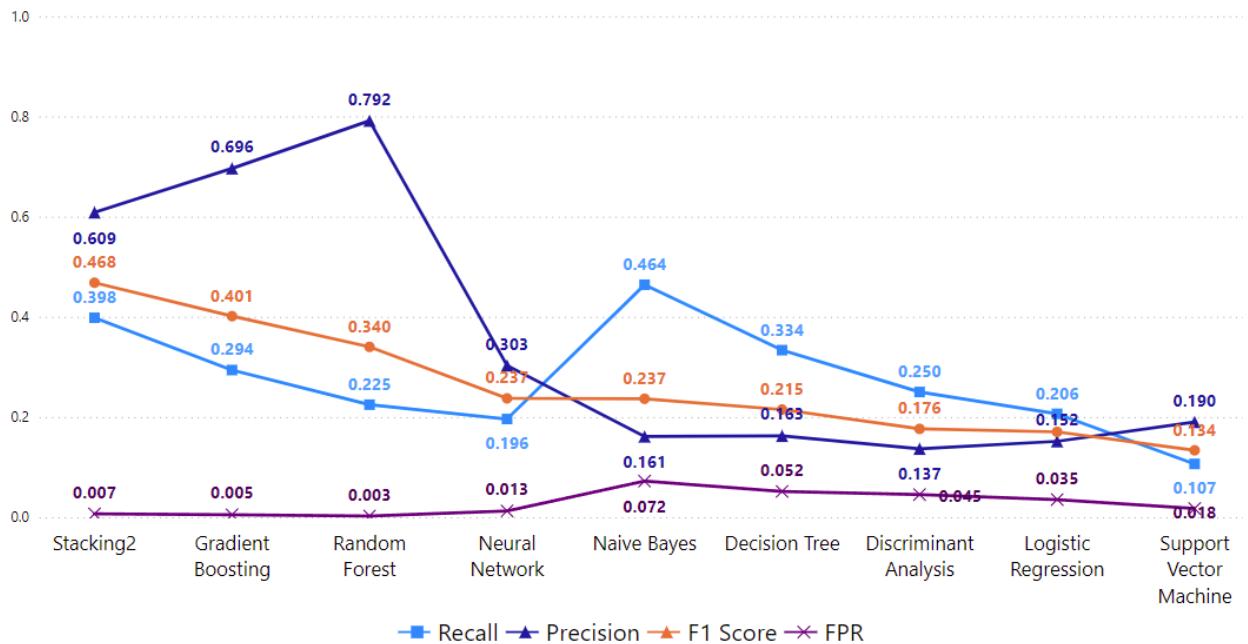


**Figure 60.** Stacking using different meta-models with seven different sets of rank-ordered base models based on item response, response time and summative statistics with a dual resampling ratio of 2:1 between the non-cheater and cheater classes.

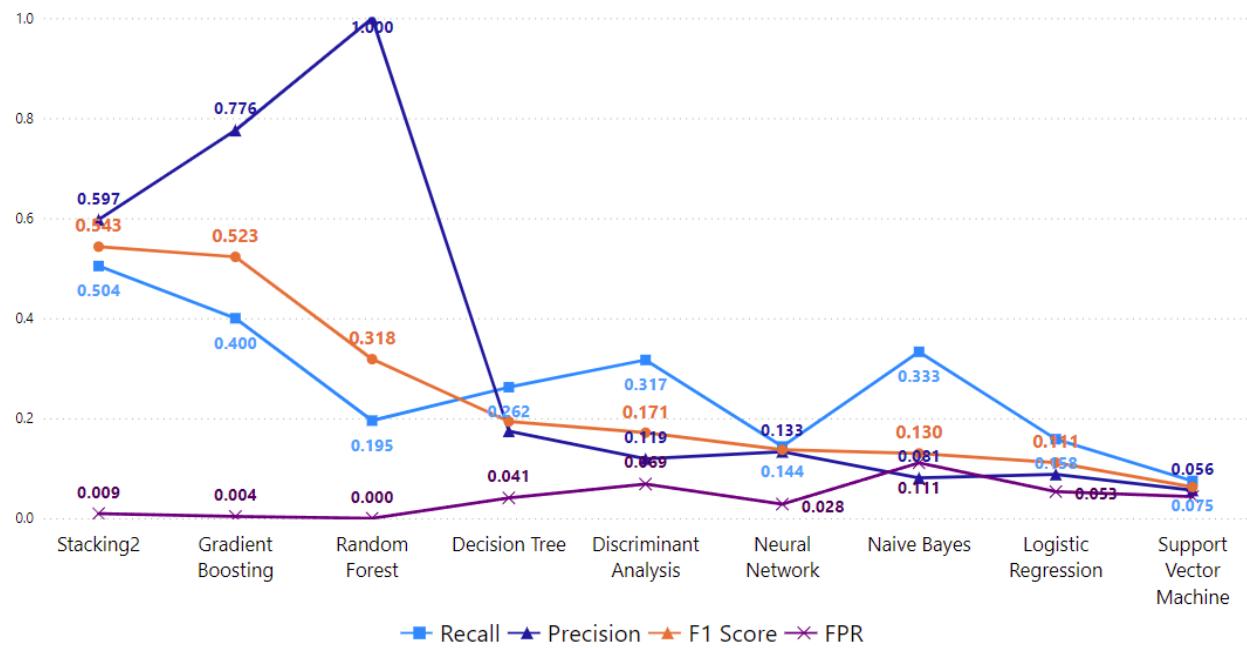


**Figure 61.** Stacking using different meta-models with seven different sets of rank-ordered base models based on item responses and summative statistics with a dual resampling ratio of 2:1 between the non-cheater and cheater classes.

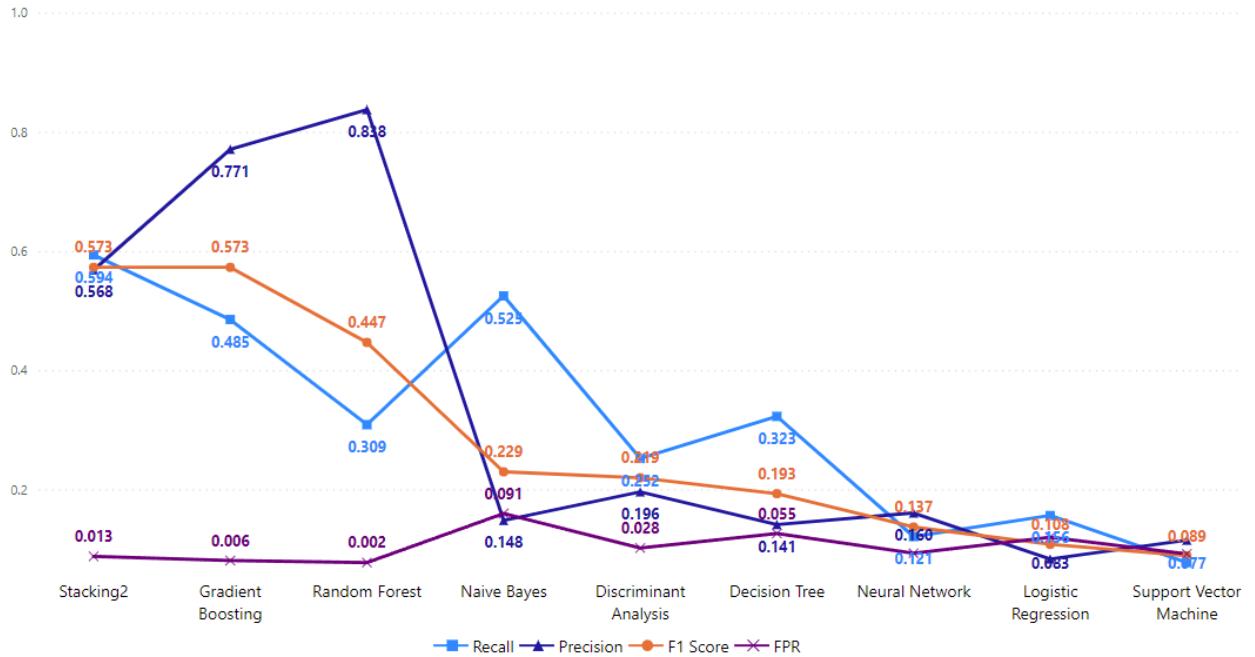
## Part 5. Stacking2 model (1-component Stacking Only) and 8 base models



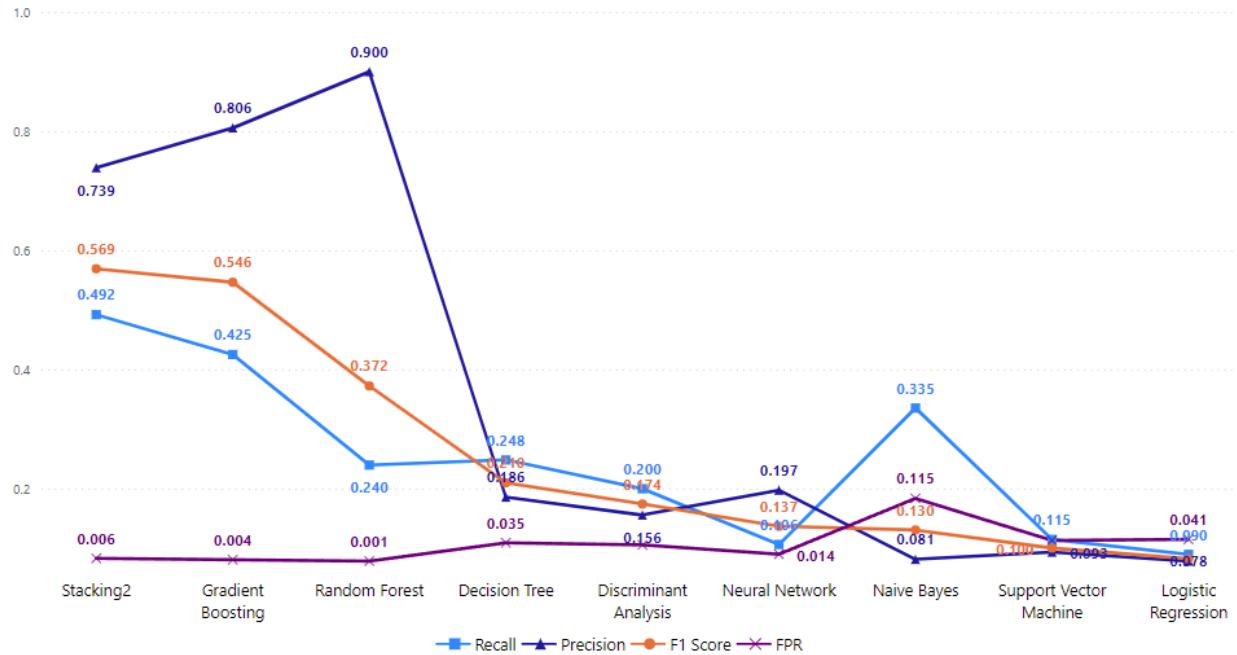
**Figure 62.** Comparison between the base models and the meta-model built upon stacking using Naive Bayes based on item responses, response time and summative statistics with a ratio of 2:1 between the non-cheater and cheater classes.



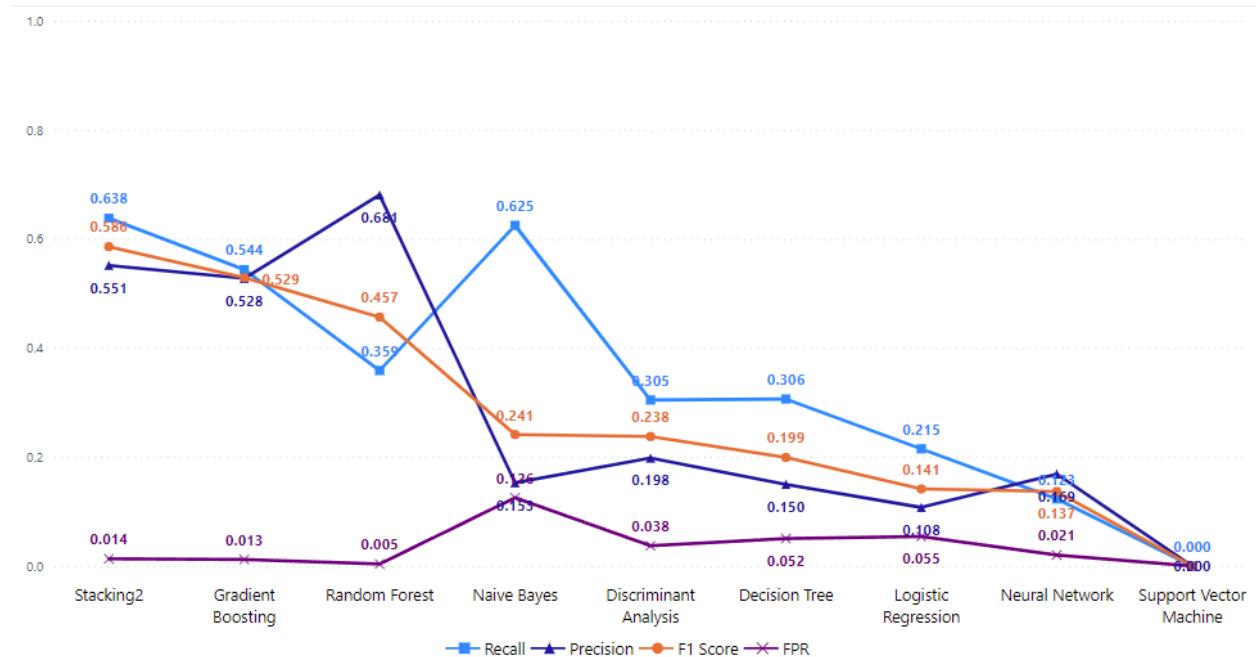
**Figure 63.** Comparison between the base models and the meta-model built upon stacking using Naive Bayes based on item responses and the summative statistics with a ratio of 2:1 between the non-cheater and cheater classes.



**Figure 64.** Comparison between the base models and the meta-model built upon stacking using Random Forest based on item response and summative statistics with a dual resampling ratio of 10:1 between the non-cheater and cheater classes.



**Figure 65.** Comparison between the base models and the meta-model built upon stacking using Naive Bayes based on item response and summative statistics for oversampling SMOTE only with a ratio of 5:1 between the non-cheater and cheater classes.



**Figure 66.** Comparison between the base models and the meta-model built upon stacking using Discriminant Analysis based on item response and summative statistics for under-sampling only with a ratio of 10:1 between the non-cheater and cheater classes.