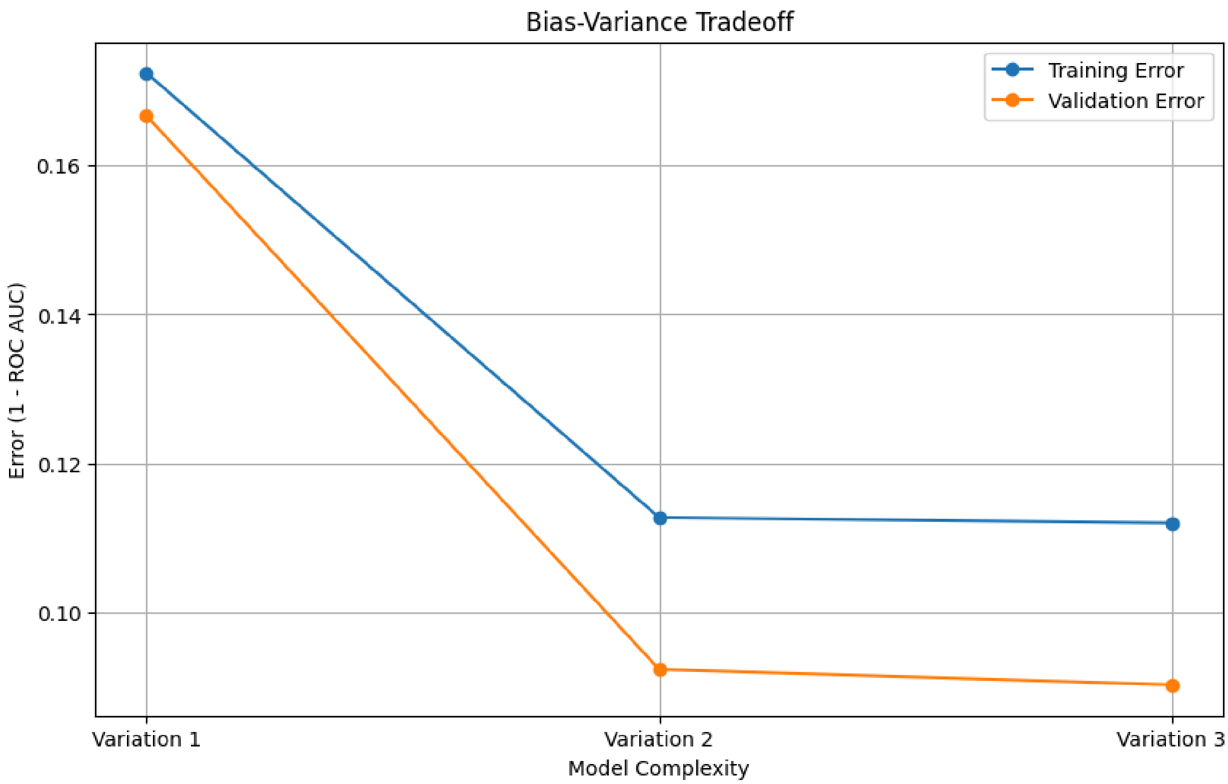
Week 9 – Select The Winning Model

Aritra Ray

Maria Alice Fagundes Vieira

- The validation error for all models (9 models in total)

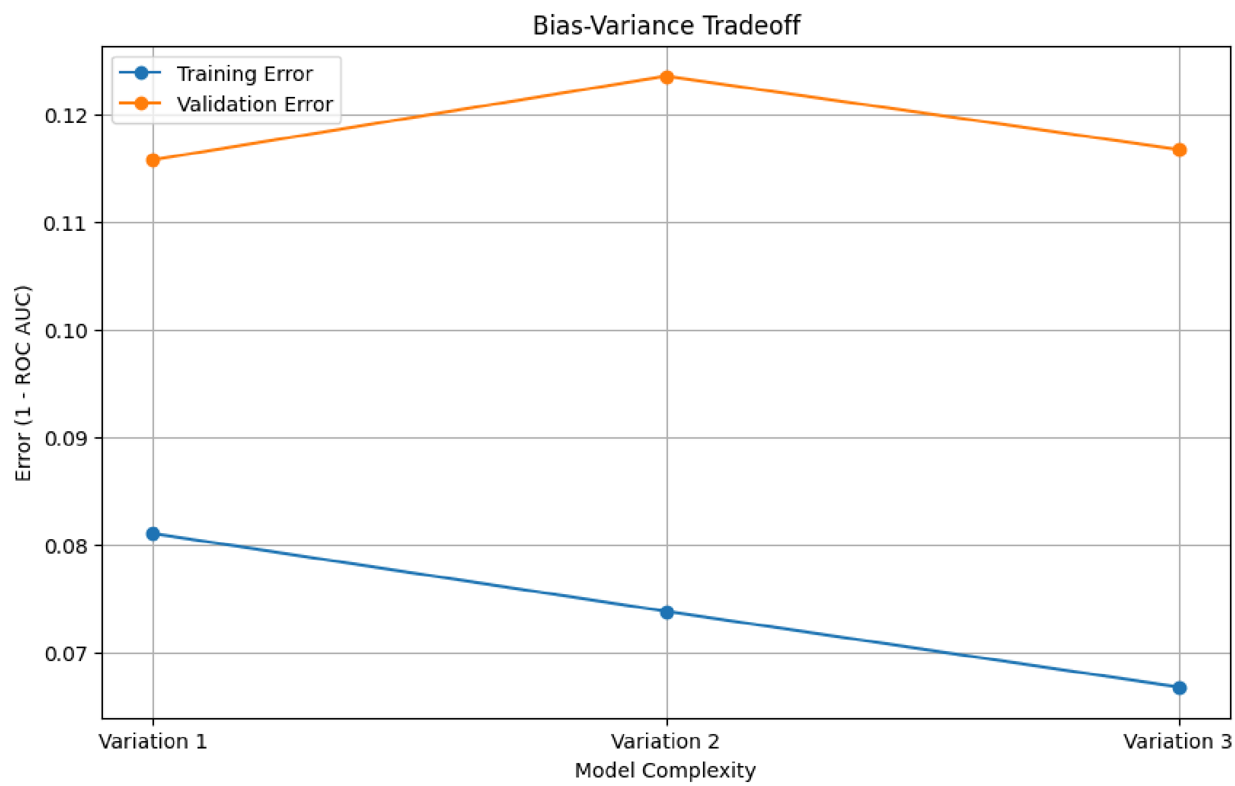
Validation error on first model – Collaborative filtering




```
Summary Table:
  Variation Dataset  Avg ROC AUC  Avg F1 Score
Variation 1   train     0.827547      0.075518
Variation 1     val     0.833304      0.085716
Variation 2   train     0.887300      0.110169
Variation 2     val     0.907706      0.119564
Variation 3   train     0.888038      0.110746
Variation 3     val     0.909773      0.118087

Best Model For This Week: Variation 3
```

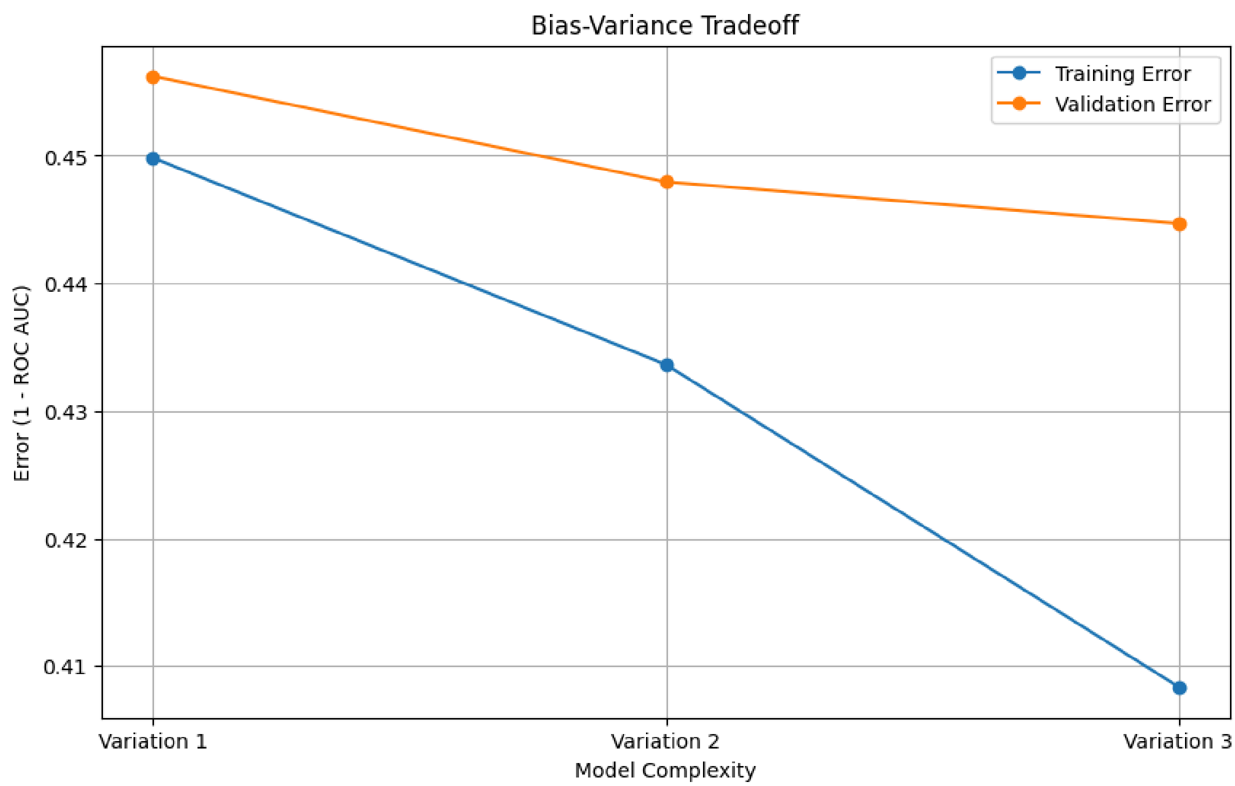
Validation error on second model – XGBoost



Variation	Train ROC AUC	Train F1 Score	Val ROC AUC	Val F1 Score
Variation 1	0.918963	0.145609	0.884238	0.139814
Variation 2	0.926188	0.146433	0.876465	0.137302
Variation 3	0.933224	0.134320	0.883285	0.116577

The best model for this week is Variation 1

Validation error on third model – Random Forest




```
Results Table:
  Variation  Train ROC AUC  Train F1 Score  Train Accuracy  Val ROC AUC  Val F1 Score  Val Accuracy
Variation 1       0.550161        0.138603        0.959607     0.543766      0.114174      0.971915
Variation 2       0.566368        0.186514        0.961059     0.552054      0.141665      0.971660
Variation 3       0.591654        0.261473        0.964706     0.555317      0.152253      0.971202

 The best model for this week is: Variation 3
```

Based on the validation error on the validation dataset, the winning model was the variation 3 of the Collaborative filtering model. The variation 3 of the collaborative filtering model had the best average ROC AUC score. The model that presented the best F1 scores was the XGBoost model, however, the ROC AUC score was lower than the other models.

This is the model from week 6 and our simplest model. I believe it was the best model because it was able to capture the relationships between clients behavior and predict well on unseen data. This is an approach frequently used on product recommendation systems as it compares the clients behaviors with previous clients. It takes advantage of customer interactions with products and similar customer's behavior, to help predict the possibility that a customer will buy a product. This approach is effective because it doesn't require detailed product features, and instead relies on the collective preferences of users, which in this case, was enough to identify the patterns in the dataset.

- Calculate model performance metrics for the final selected model using the test dataset

For our test dataset, the model performance metrics were as below:

- Average ROC AUC: 0.8831
- Average F1 Score: 0.2014

This means that the ROC AUC was slightly below the validation set, but the F1 score was considerably higher than the ones seen on validation and training sets, which makes us believe the model is good enough for our prediction task

	Avg. ROC AUC	Avg F1 Score
Train	0.8880	0.1107
Validation	0.9098	0.1181
Test	0.8831	0.2014

The ROC AUC remains fairly high across all datasets, indicating that the model has a good ability to distinguish between classes. The slight improvement in the Validation set suggests that the model generalizes well during validation, but there is a small drop in performance on the Test set.

The F1 Score is quite low on both the Train and Validation sets, but it improves substantially on the Test set. This suggests that while the model may have struggled with class imbalance or precision/recall trade-offs during training and validation, it performed better on unseen test data.