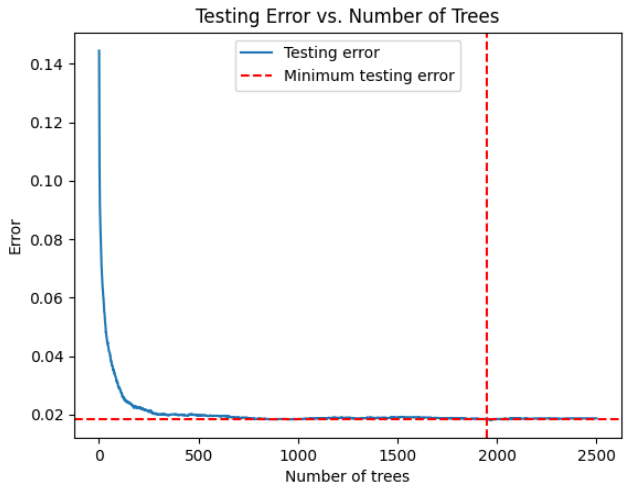
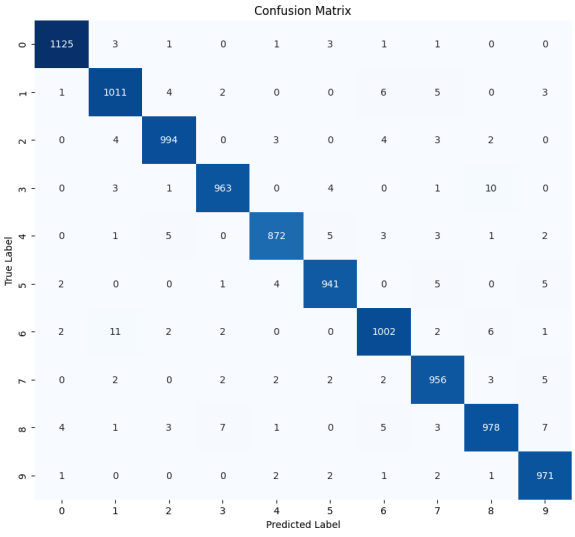
Part 2: MNIST

This part of the report discusses our results for an XGBoost classifier trained on the MNIST dataset of handwritten digits. Before classifying with XGBoost, the labels had to be re-encoded to prevent a mismatched label error thrown by the model, and later inverted after training was complete. By first using `RandomSearchCV` to get a sense for the range of good parameters then maximizing the test set accuracy with `GridSearchCV`, we found the following parameters to return the best classifier:

- **Max depth:** 6. This depth is enough to capture a medium amount of complexity and interactions between features such as slight differences in number pairs like 0 and 8, 1 and 7, or 4 and 9, but not too much as to overfit.
- **Learning rate:** 0.1. This rate affects the contribution of each tree, a lower amount increasing robustness towards noise at the cost of more training time for the same results.We decided on 0.1 because the slight increases in accuracy came at exponentially longer training times and wasn't feasible for the time frame of this project.
- **N Estimators:** 1912. The number of trees was determined by the first index of minimum testing error, meaning the test error was no longer improving past 1912 trees.

The data was plotted in several ways we found meaningful, particularly confusion matrix and test error as a function of trees. Somewhat meaningful were the plot of top 20 important features, the heatmap of pixel importance, and the precision / recall table. We did not find the correlation heatmap matrix meaningful. Our best forest had a test set accuracy of 98.15%.





Part 2: MNIST-LeNet5

This part of our report discusses our results for an XGBoost classifier trained on the MNIST dataset, after it has been pre-trained through LeNet5 convolutional neural net. The methods for finding the best parameters and meaningfulness of plots were significantly the same, except for the heatmap of pixel importance since the features no longer correlate to original pixels. We found the following parameters to return the best classifier:

- **Max depth:** 9. This depth is enough to capture the features complexity and interactions between features, with less risk of overfitting as the data has already been trained through a neural net.
- **Learning rate:** 0.1. Same as MNIST.
- **N Estimators:** 2358. The number of trees was determined by the first index of minimum testing error, meaning the test error was no longer improving past 2358 trees.

We found the classifier trained using this pretrained MNIST set to be far more accurate than training on the MNIST set itself, which was expected.