# TRI-AD
Data Scientist Challenge

Ankush Khullar
May 7, 2020

Part 3: Reporting

A random forest model was selected as the best model to predict whether a student will pass the test using age, sex, dojo_class, and hours_studied as predictors. The chart on the left shows the overall accuracy of the model and the chart on the right breakdowns the proportion of observations that were correctly and incorrectly classified.
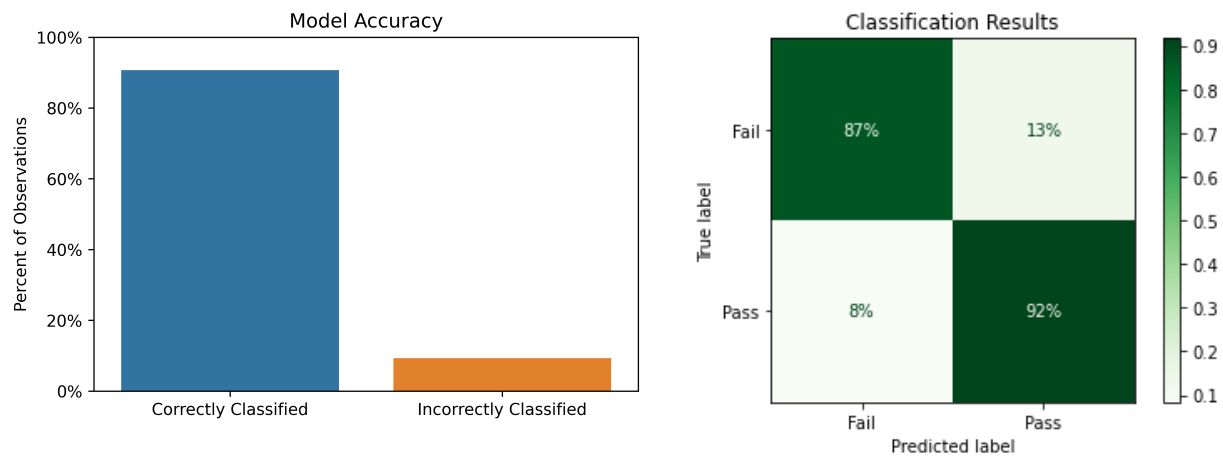
The following can be inferred from the classification results:

- 90.6% of observations were correctly classified and 9.4% were misclassified;
- 87% of students that failed the test were correctly predicted to fail the test;
- 92% of students that passed the test were correctly predicted to pass the test;
- 13% of students that failed the test were incorrectly predicted to pass the test; and,
- 8% of students that passed the test were incorrectly predicted to fail the test.

Although the model has high accuracy, additional features could lead to better predictions. For example, although it was shown that taking the Dojo class increases the likelihood of passing the test, a student who takes the class and is highly engaged likely has a better chance of passing than a student who takes the class and is not engaged. Additional features that could be used to improve the model include:

1. Attendance – the number or percent of days the student was absent from the Dojo class;
2. Engagement – a measure of how engaged the student is during the Dojo class (e.g., how many times the student raises his or her hand to ask a question);
3. Class grades – grades on assignments given in the Dojo class;
4. Practice test scores – scores of any practice exams given during the Dojo class;

5. Student surveys – ratings given by the student on how effective they think the class is;
6. Time of day – the time of day the test was administered;
7. Day of week – the day of the week the test was administered; and,
8. Attempt number – the number of times the student took the test (if students are allowed to retake the test after failing).

The results of the analysis show that it is possible to predict whether a student passes the test with reasonably high accuracy using only a few features. Adding additional metrics will likely improve the model as well as provide additional insights into how the Dojo class can be improved to better prepare students to pass the test.