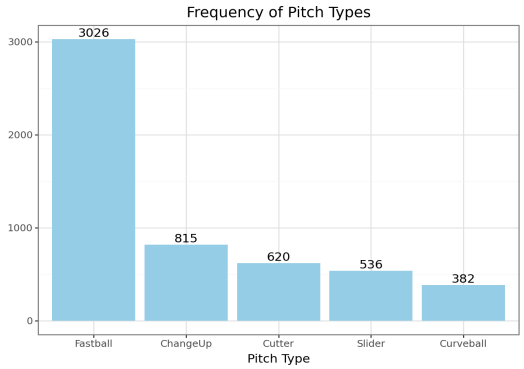
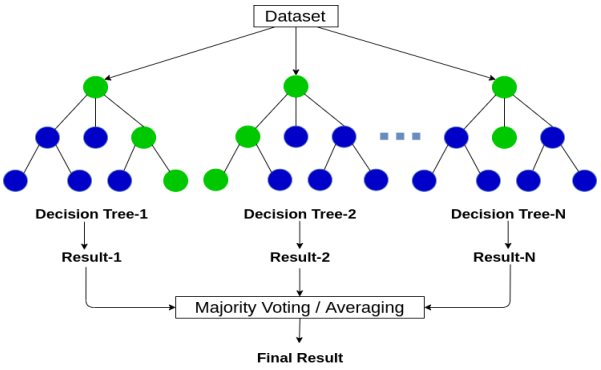
CSC 466 Final Project Results - Brandon Kim, James Irwin, Gabe Riedel

In baseball, there are many different types of pitches that are thrown. Each pitch has a unique combination of metrics such as velocity, spin rate, vertical movement, horizontal movement, and more that comprise the pitch's "shape." Using data from the Cal Poly baseball team with pitch shape metrics and type classifications, we wanted to see if we could predict the type of pitch thrown from the pitch shape metrics.





To solve this problem, we built a random forest classification model with 5 numeric features: velocity, spin rate, induced vertical break, horizontal break, and tilt. Since we are using numerical features, we used gini gain as our metric to determine how the trees in our forest split the data. We fit our trees on bootstrapped samples and a subset of our features.

After performing hyperparameter tuning on various combinations of number of trees in the forest, depth of the trees, and minimum number of splits in the trees, we found the optimal model, by accuracy, to have 100 trees, a depth of 15, and a minimum of 2 splits.

Trees	Depth	Min. Split	Accuracy
10	5	2	.9079
10	5	5	.9126
10	10	2	.9265
10	10	5	.9321
10	15	2	.9284
10	15	5	.9330
50	5	2	.9172
50	5	5	.9079
50	10	2	.9349

Trees	Depth	Min. Split	Accuracy
50	10	5	.9321
50	15	2	.9312
50	15	5	.9312
100	5	2	.9079
100	5	5	.9153
100	10	2	.9321
100	10	5	.9312
100	**15**	**2**	**.9349**
100	15	5	.9321

Actual/Predicted	Fastball	Cutter	Changeup	Curveball	Slider
Fastball	598	5	3	0	0
Cutter	5	101	1	0	17
Changeup	5	1	157	0	0
Curveball	0	0	0	74	3
Slider	2	20	0	1	85

We then ran the model on a test set of data, and produced this confusion matrix with incredible results; all metrics of accuracy, macro precision, macro recall, and macro f1 score were in the low-to-mid 90s.

Test Set Accuracy: 0.9415
Test Set Macro-Precision: 0.9025
Test Set Macro-Recall: 0.9094
Test Set Macro-F1 Score: 0.9036

Lastly, we performed a feature importance analysis, and found that pitch velocity and induced vertical break were the most relevant features of the pitch's shape in predicting the pitch's classification.