Driveline OpenBiomechanics Project Independent Study

In this study, we are looking for ways to improve the already excellent biomechanical features put together by Driveline Baseball to predict throwing speed. Driveline compiled data using markers on the body for pitchers and hitters in their throwing and swinging motions, respectively. We place our focus on the pitching side of this project for this study. The appropriate model for this data seems like a linear regression, due to the predictive nature of the project, the desire for weights and statistical significance of each feature, and the transparency of the results. After splitting the 411 data examples into a training, validation, and test set, the initial model formed by Driveline performed with a linear regression score of 0.658. We have added interaction terms between the features with high correlations, and we have applied a focus to measuring timing and pacing metrics using biomechanical features and timestamps. Through this, we found two significant timing features: time from peak knee height to foot plant and time from peak knee height to ball release. Time from foot plant to ball release was also added to the model but was not statistically significant. These added features have increased the validation score to 0.727, a good improvement with only the addition of a few features. The next step in this process will be to look for features that have an effect on these timings, study those effects, and possibly add them to the model. Another future development will be looking into other ways to perform the linear regression, such as through boosted forests like XGBoost.