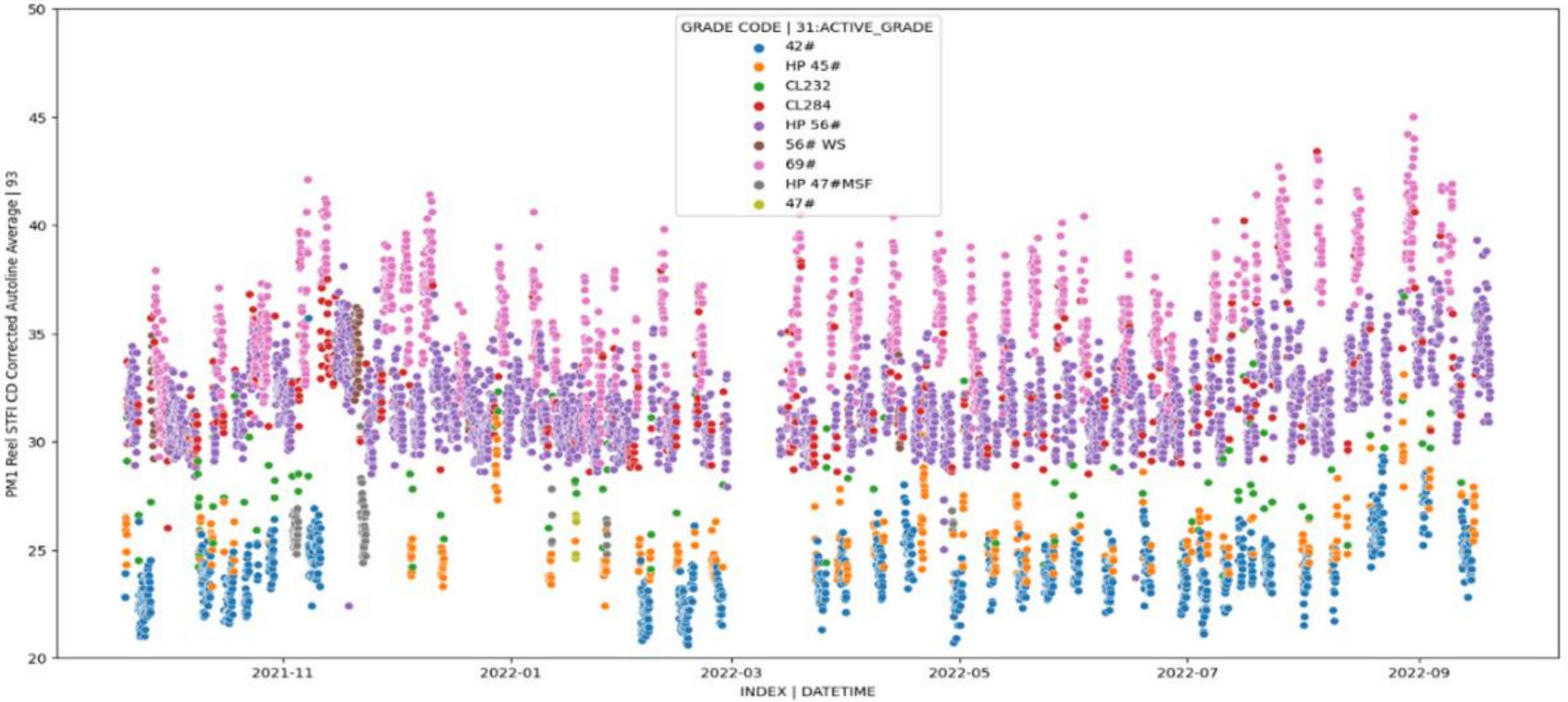
Final Progress Report

Team IP1: Parker Jamison, Barbara Remmers, Sophia Su

Georgia Tech

Project & Data Overview

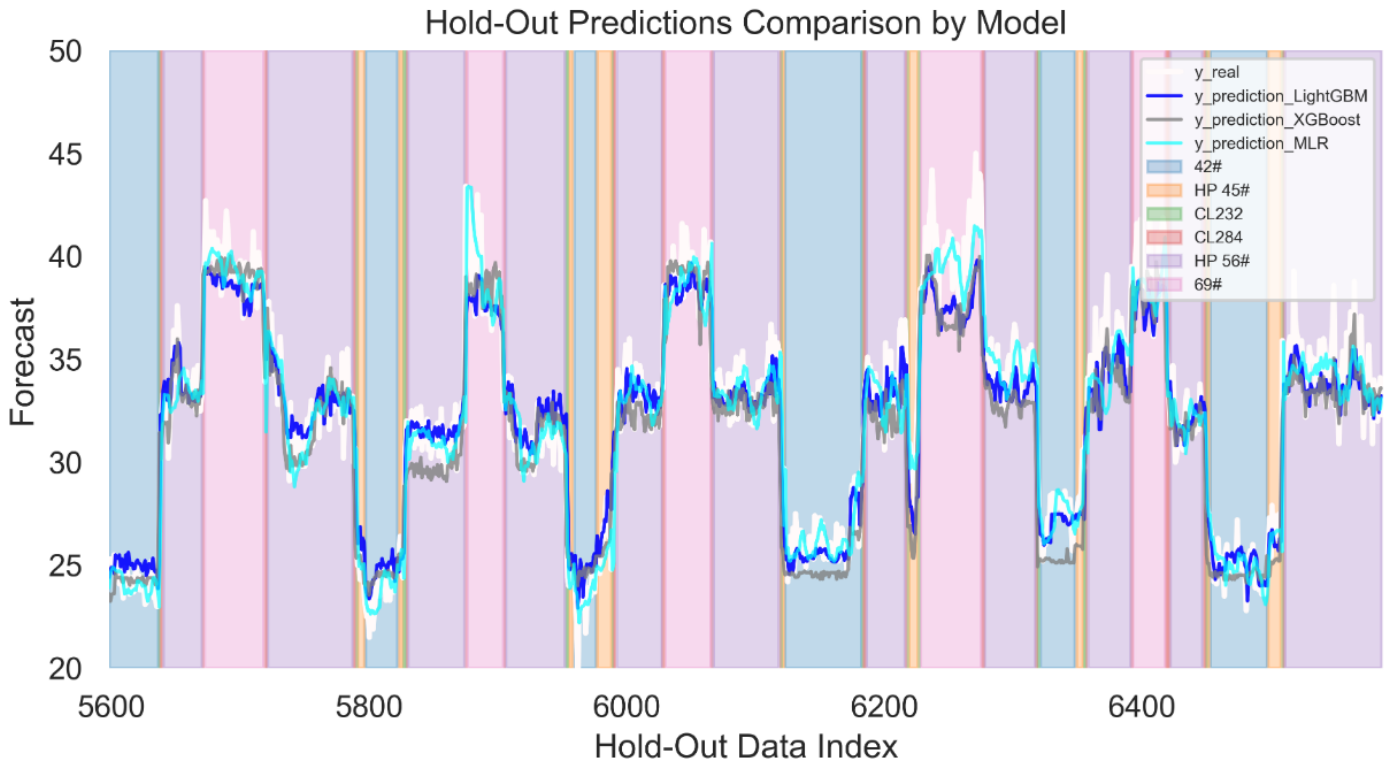
Build models to predict a quality test result that could be implemented during the manufacturing process.
Also use the models to identify which parts of the production process have predictive power.



- *Target variable data is highly variable: by grade, grade runs, and within a grade run.*
- *Potential predictor variables are plentiful and noisy.*

Georgia Tech

Project Results



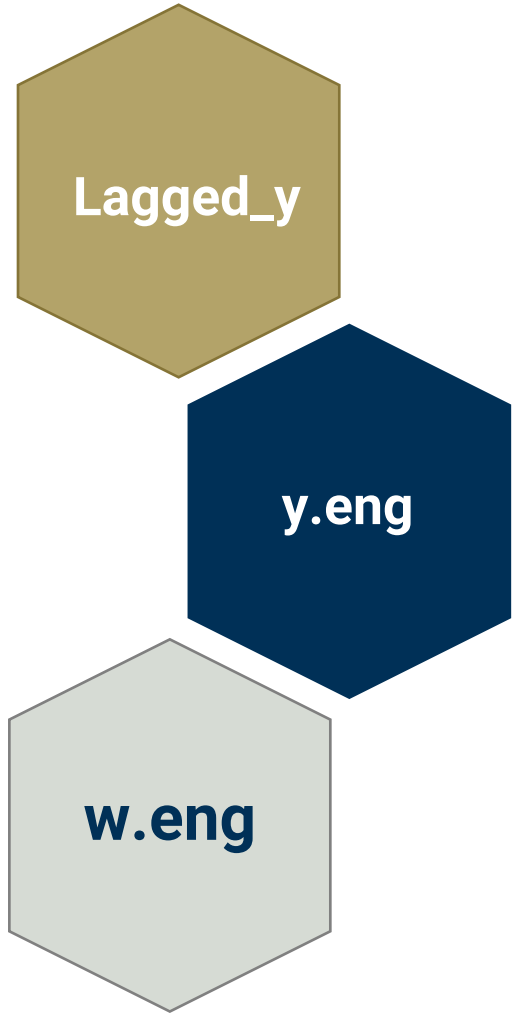
Model	Avg. RMSE Test	RMSE on Hold Out	Avg. MASE on Test	MASE on Hold Out
Naïve	1.57	1.70	1.00	1.00
MLR	1.49	1.59	0.93	0.93
MLR single grades	1.47	1.59	0.99	0.93
Random Forest	1.45	2.02	0.88	1.16
LightGBM	1.15	1.56	0.70	0.95
XGBoost	1.36	1.81	1.12	1.49

* MLR stands for Multiple Linear Regression

* See all-grade model results by grade and expanded table in Appendix

Georgia Tech

Feature Engineering



Lagged variables

- Other Autoline tests performed at same time as the STFI test (target variable)
- Cleaned weight and target variable

Engineered Feature using lagged targets

- Aims to estimate current target with less noise than lagged_y
- Uses prior targets in grade run
- For long enough runs, uses linear estimate of next target based on prior 12.
- First observation in a grade run uses lagged_y

Engineered Feature using lagged weights and current weight

- Quantifies how much current weight deviates from the mean of aver weights in the grade run
- Uncorrelated with y.eng, small but statistically stable effect

Georgia Tech

Feature Importance

- Engineered Features

Model	Predictors	Coefficients	RMSE Train	RMSE Test	MASE Train
Naïve	Lagged_y	1.00	1.57	1.57	1.00
OLS	y.eng	1.00	1.49	1.52	0.92
MLR	y.eng, w.eng	1.00, 0.33	1.46	1.49	0.91

- Sensor Features

Feature Name	Selected by Rnd. Forrest	Selected by LightGBM	Selected by XGBoost
REEL SPEED | 31:IMV005	X	X	X
PRI TOTAL HEAD | 31:IMV018	X	X	X
CW SCAN AVG | 31:F13SAN		X	X
PRI SLICE POSITION | 31:IMV022	X		X
SEC TOTAL HEAD | 31:IMV019	X		X
WIRE TO BINIP DRAW | 31:DRAW001	X	X	

Georgia Tech

Recommendations & Conclusions

- Focus on all-grade models
- Retrain tree-based models using more years of data and re-evaluate performance
- Keep the top sensors in good working order

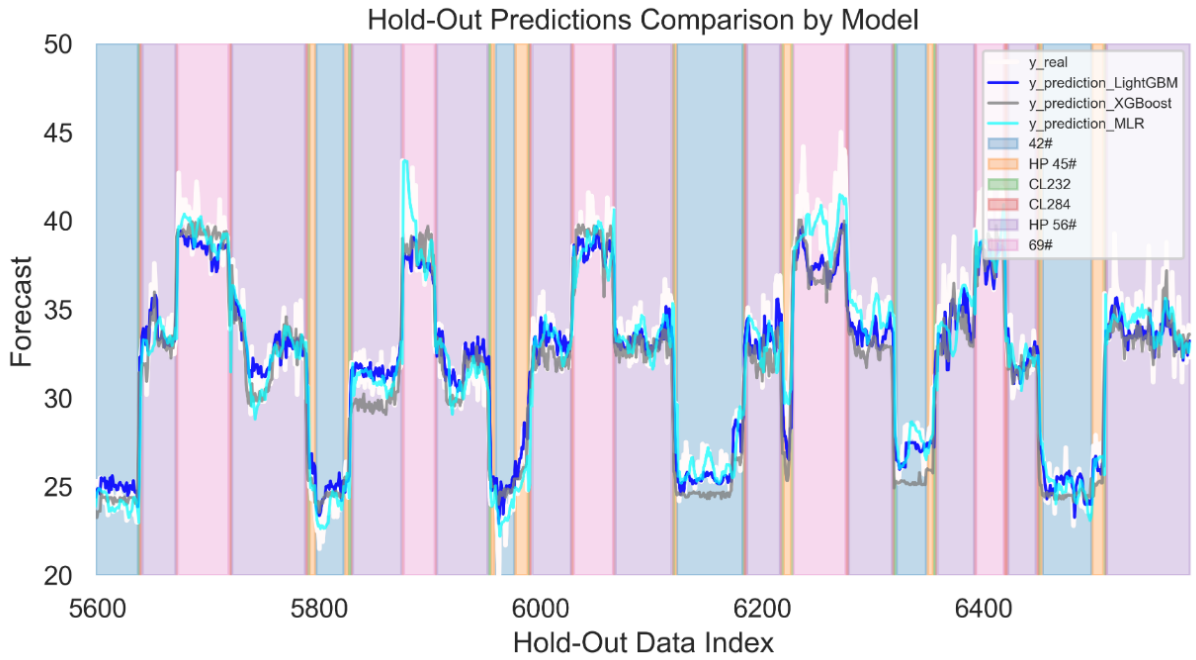
Feature Name	Selected by Rnd. Forrest	Selected by LightGBM	Selected by XGBoost
REEL SPEED | 31:IMV005	X	X	X
PRI TOTAL HEAD | 31:IMV018	X	X	X
CW SCAN AVG | 31:F13SAN		X	X
PRI SLICE POSITION | 31:IMV022	X		X
SEC TOTAL HEAD | 31:IMV019	X		X
WIRE TO BINIP DRAW | 31:DRAW001	X	X	

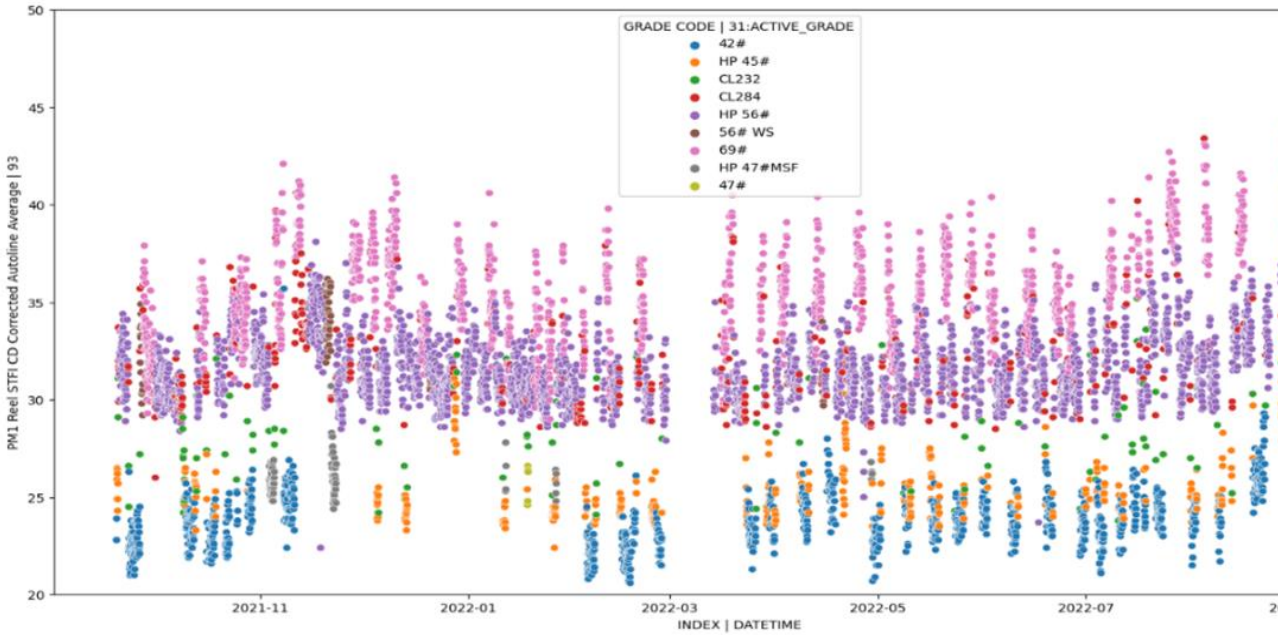
- Consider adding new sensors
- Test sensor-only models to see how they compare to the linear models

Georgia Tech

Discussion

Appendix: All-Grade Model Results, by Grade





STFI data values distribution graph

Show Hide

Appendix: All-Grade Model Performance Comparison

Model	Avg. RMSE Train	Avg. RMSE Test	RMSE on Hold Out	Avg. MASE on Train	Avg. MASE on Test	MASE on Hold Out
Naïve	1.57	1.57	1.70	1.00	1.00	1.00
Multiple Linear Regression (MLR)	1.46	1.49	1.59	0.91	0.93	0.93
MLR single grades	1.48	1.47	1.59	0.99	0.99	0.93
Random Forest	1.35	1.45	2.02	0.93	0.88	1.16
LightGBM	0.95	1.15	1.56	0.64	0.70	0.95
XGBoost	1.08	1.36	1.81	0.72	1.12	1.49

Georgia Tech