The Slot-ification of Baseball

Creating increased gambling opportunities in America's Pastime

Joe Buzzelli

May 12, 2020

In DC, sports gambling is poised for rapid growth

- The Council of the District of Columbia's Act 22-594, the Sports Wagering Lottery Amendment Act, passed on January 23rd, 2019
- It sets the scene for Ted Leonsis augment his current, sports-centric business operations

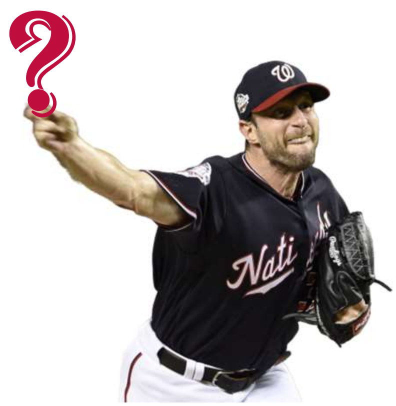
*"In **Ted Leonsis's entrepreneurial dreams**, Capital One Arena will transform into a year-round sanctuary because of **legalized sports gambling**. On any given day — regardless of whether the Capitals and Wizards are good, bad or just off for the night — the place will be buzzing and creating **fresh revenue streams**. Within this place, he will be able to marry two of his great passions — **sports and technology** — in an innovative, **bettor-friendly way** and design an unstoppable business model."* *

*https://www.washingtonpost.com/

The Slot-ification of Baseball adheres to Leonsis's current efforts

- My capstone project adopts the scenario of proposing a new sports gambling app to Ted Leonsis based on baseball

- The ultimate goal is to create a mobile app that allows a user to make a large volume of low-value wagers where they guess the type of the next pitch in a major league baseball (MLB) game



The Slot-ification of the MLB increases transactions by **24,980%**

	Traditional Gambling	**In-Running**	**Slot-ification**
Description	Based on the final result of the game (winner or total score)	Based on the outcome of each hitter's at bat	Based on the type of each pitch in a game
Transactions per MLB season	**162** games	**10,692** at bats per season (**33** per game per team)	**40,630** pitches per season (**3.8** pitches per at bat)
Percent increase from baseline	**-%**	**6,500%**	**24,980%**
Potential transactions (not to scale)			

The first step towards Slot-ification is setting the odds

- Prior to offering bets on the pitch type and location of every MLB pitch, a statistical analysis is required to assess the odds of pitch types per pitcher

- To explore the viability of MLB's Slot-ification, Max Scherzer is the focus of this analysis
  - Analysis uses data from every pitch he threw in the 2019 season
  - Includes the running performance metrics of each batter he faced

- **Goal**: Create a model predict pitches with at least **75% accuracy**

MLB's PITCHf/x provides exhaustive data for Slot-ification

- **Data elements sourced from Baseball Savant at baseballsavant.mlb.com**

Data Element	Transformation	Data Element	Transformation
Pitch type (target)	Ordinal	Inning	MinMax Scaling
Batter stance (left/right)	Categorical	Outs during at bat	MinMax Scaling
Infield alignment	Categorical	Batter's whiffs	MinMax Scaling
Outfield alignment	Categorial	Batter's swings	MinMax Scaling
Balls and strikes	Categorical	Batter's takes	MinMax Scaling
Runners on base	Categorical	Batter's strikeouts	MinMax Scaling
Nationals (home/away)	Categorical	Batter's walks	MinMax Scaling
Batter's batting average	Already standardized	Batter's singles	MinMax Scaling
Batter's slugging percentage	Already standardized	Batter's doubles	MinMax Scaling
Batter's isolated power	Already standardized	Batters triples	MinMax Scaling
Batter's BA on balls in play	Already standardized	Batters homeruns	MinMax Scaling
Pitch number (per at bat)	MinMax Scaling	Batter's contact types	MinMax Scaling
Pitch number (per batter per game)	MinMax Scaling	Batter's RBIs	MinMax Scaling
Pitch number (per batter per season)	MinMax Scaling	Batter's sacrifices	MinMax Scaling

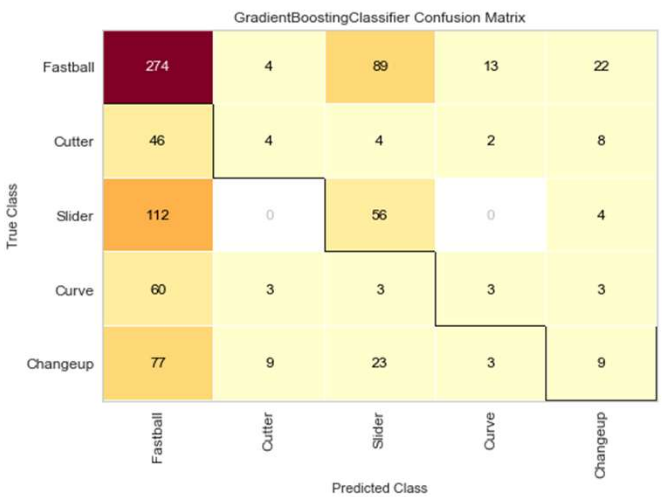
The Gradient Boosting Classifier performed best in initial testing

- The table below outlines the models assessed in this analysis to include default results, cross validation (CV) scores, and grid search/randomized (GSRS) search results

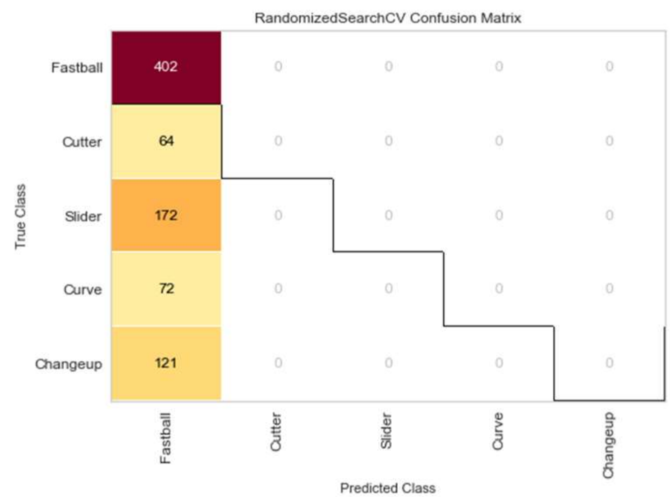
Model	Default Accuracy	CV Accuracy	GSRS
Dummy	48%	n/a	n/a
Linear Regression	tbd	tbd	tbd
Decision Tree	tbd	tbd	tbd
Random Forest	tbd	tbd	tbd
Gradient Boosting Classifier	69.0%	43.1%	48.4%
LSTM	tbd	tbd	tbd
XGBoost	tbd	tbd	tbd
MLP Classifier	tbd	tbd	tbd

Fine tuning the best model lead to troubling results

- With initial accuracy of 69%, the gradient boosting classifier showed the best performance while not overfitting
- Parameter tuning regressed to the dummy model

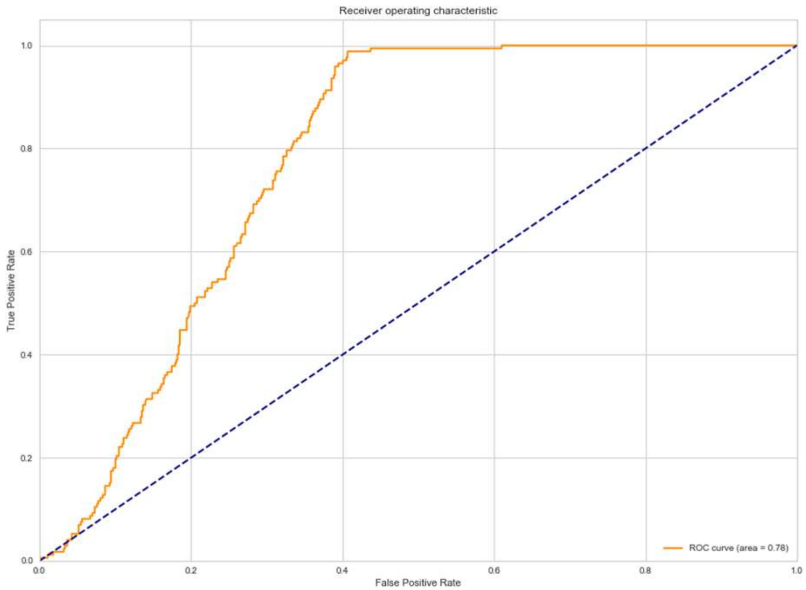


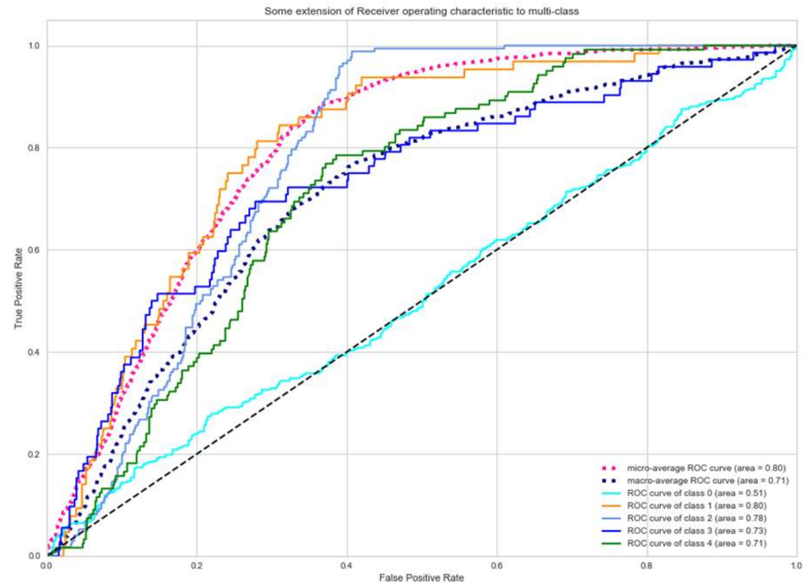
Tuning turns model into a dummy



ROC curves do not bode well for the efficacy of the MPV

- ROC curves for the default gradient boosting classifier do not look promising





Next steps

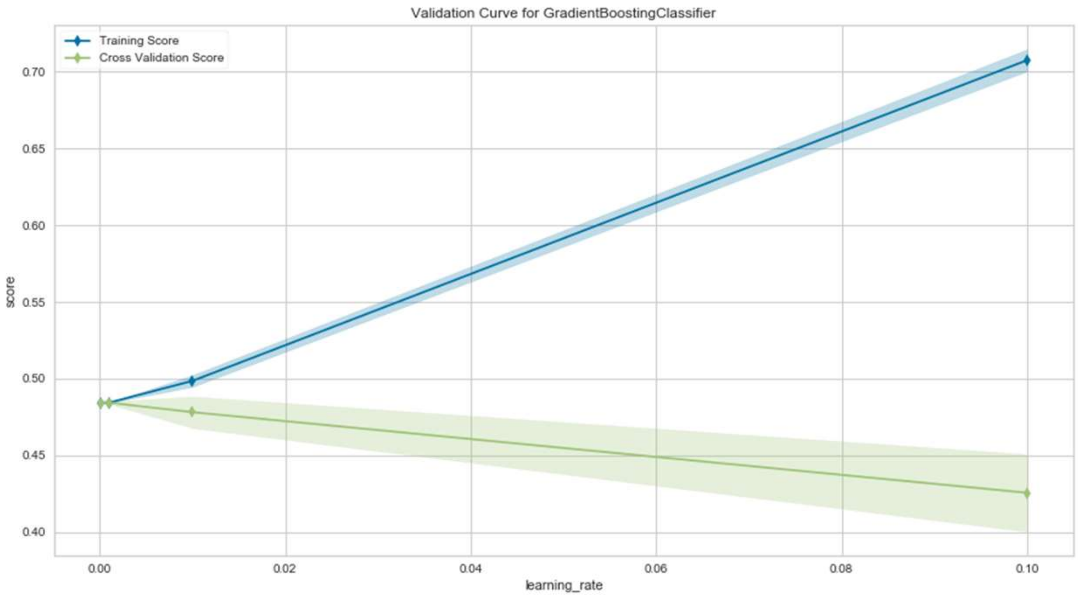
- My best model struck out after facing exhaustive tuning
- Initial efforts lead to the creation of a reusable template against which other models will be assessed
- Why is my MVP such an abject failure?
  - Max Scherzer is good and mixes pitches in an unpredictable fashion
  - Too many fastballs thrown, attempted fewer classes with no success
  - Used too small of a data set, should have used more than one pitcher

Backup Slides

//

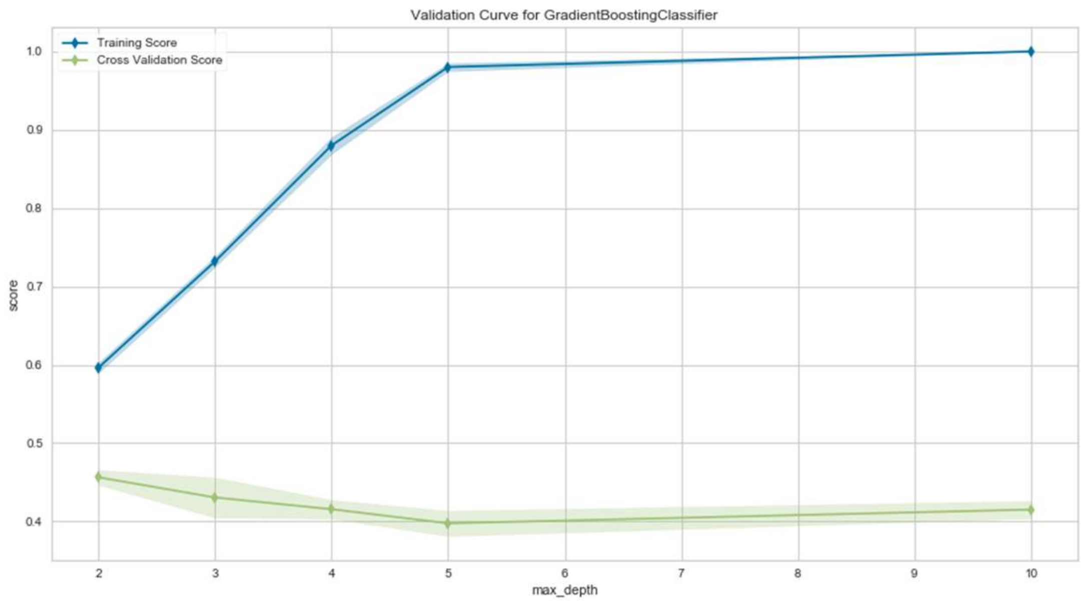
Gradient Boosting Classifier (GBC) Learning Rate Tuning

- Optimal learning rate equals .01



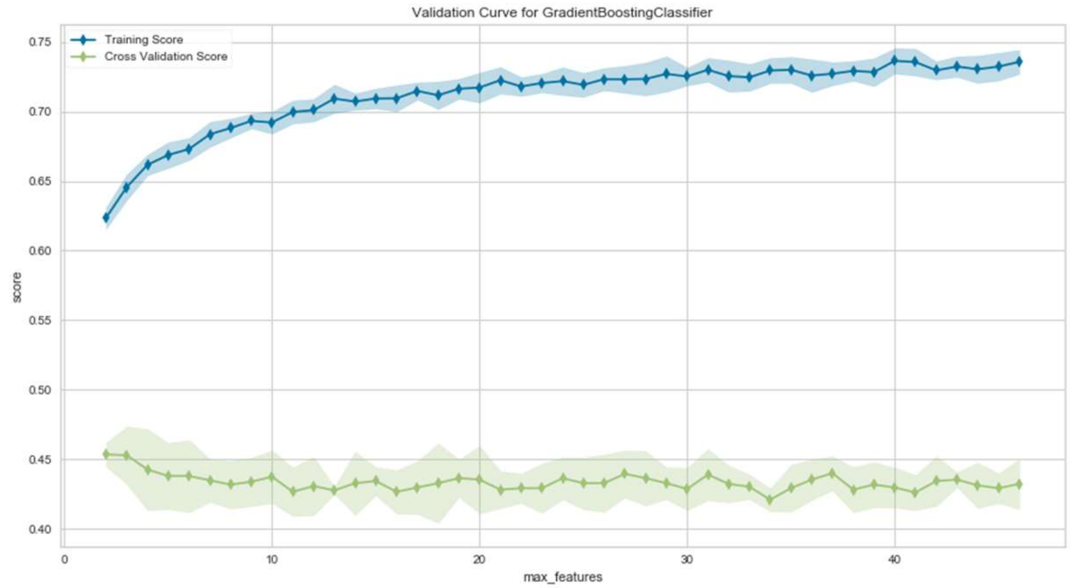
GBC Max Depth

- Optimal max depth equals 5



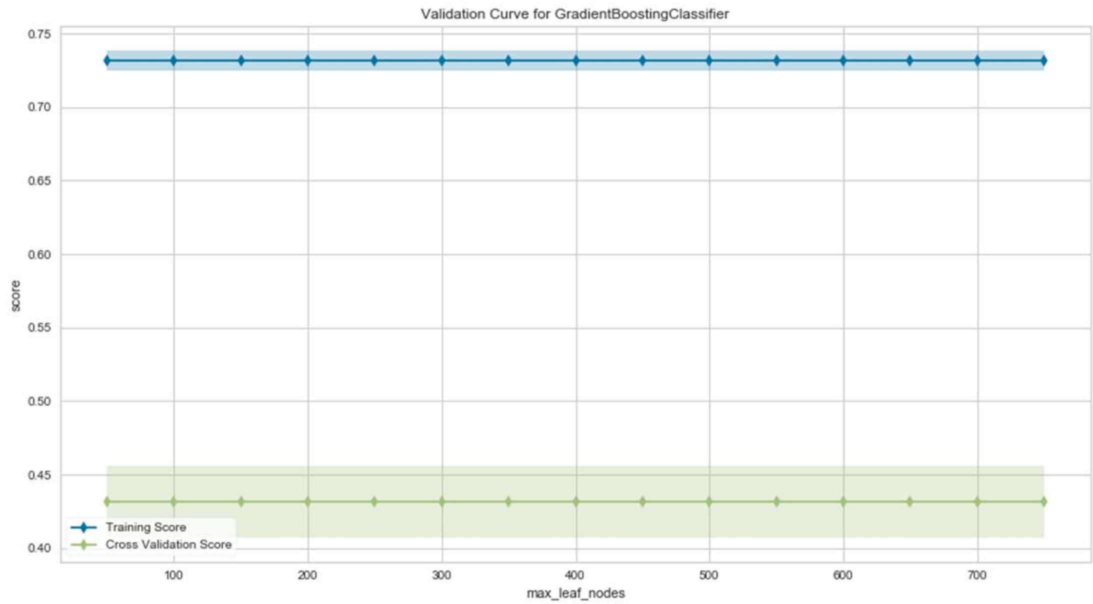
GBC Max Features

- Optimal max depth is between 15 and 20



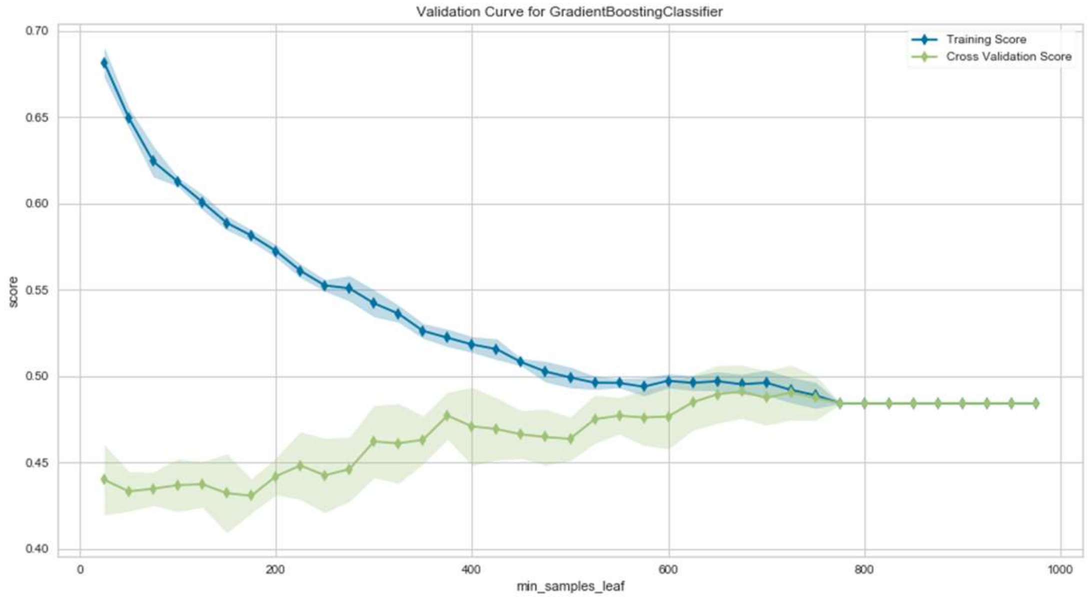
GBC Max Leaf Nodes

- Max leaf nodes do not seem to affect model performance



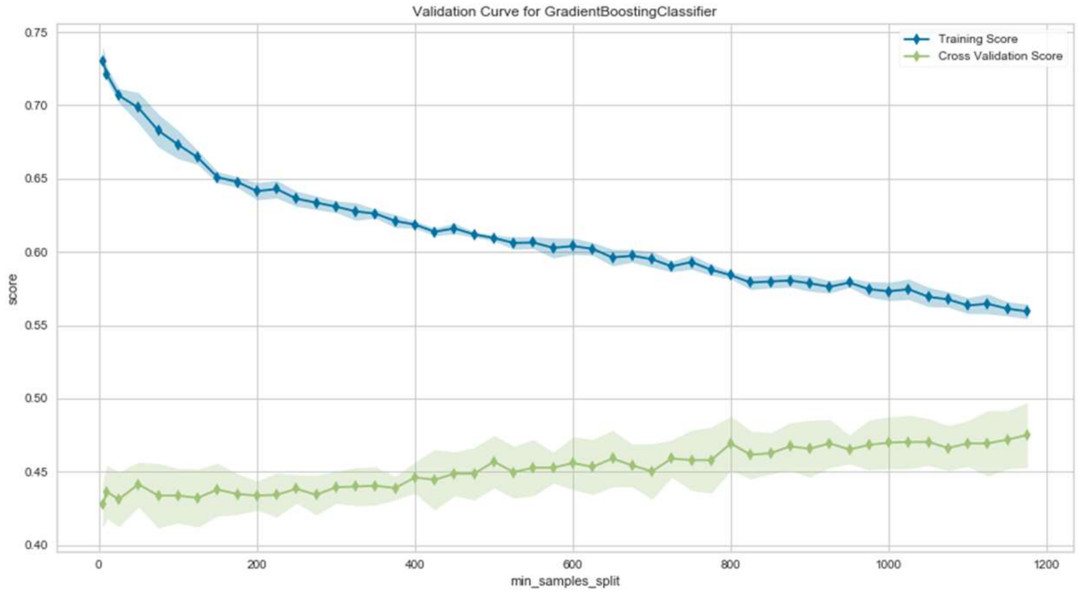
GBC Minimum Samples per Leaf

- Optimal minimum samples per leaf is 775



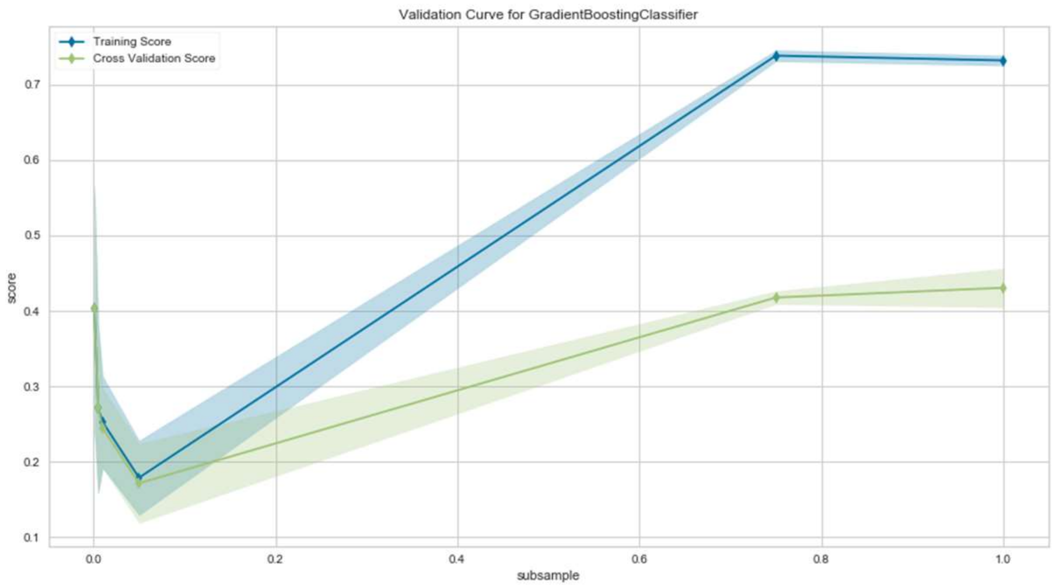
GBC Minimum Samples per Split

- Optimal minimum samples per split is 200



GBC Subsamples

- Optimal number of subsamples equals 1.0



The Slot-ification of the MLB increases transactions by **24,980%**

- Typical sports gambling relate to the final result of a game (over/under, or winner) with **162** games in the MLB season

- Although not as mainstream, in-running gambling exists in the MLB but this focuses on the action of the hitter (gets a hit or does not)
  - 10,692 at bats per season at an average of 33 at bats per game per team

- By focusing on the outcomes of each pitch, gambling transactions increase dramatically
  - Almost 40,630 transactions possible when gambling at a per pitch level given the average of 3.8 pitches per at bat in the MLB