# Predicting NHL Team Points Proposal

Eric Foote

UNBSJ

September 20th 2018

"Hockey statistics is in its infancy"[1] from what started out in the earliest day as a Yahoo group[2] has grown to a few teams having an analytics department that infulences everything from personnal choices to player scouting. (e.g. Toronto)

"Over the past few years, though, analytics has been the buzzword around NHL circles. As teams like the Chicago Blackhawks and Los Angeles Kings, winners of four of the past five Stanley Cups, find success through the use of data, other teams will naturally have to follow suit if they want to compete." [3]

We felt that Kyle was a hockey man who understood the analytical part of the game, said Nonis about the hiring of Dubas. He wasnt an analyst. To me, there is a pretty big distinction. Kyle understands that these stats and trends are just part of what we do to make decisions. Were going to use as much as we can going forward. [3]


"Chayka uses analytics as a tool in his decision-making process and as a resource in dealing with players, coaches, agents and opposing GMs." [4]
"Chayka went on to co-found Stathletes, a hockey analytics company, in 2010 while earning a business degree. Stathletes' goal was to use video analysis to create statistics that better help measure the ability and value of players – and to make those numbers more accessible." [4]

How does this apply to me?
My project is going to predict the team points

| Team | Season | GP | W↓ | L | T | OT | P↓ |
|---|---|---|---|---|---|---|---|
| Search | | | | | | | |
| Nashville Predators | 2017-18 | 82 | 53 | 18 | 0 | 11 | 117 |
| Winnipeg Jets | 2017-18 | 82 | 52 | 20 | 0 | 10 | 114 |
| Tampa Bay Lightning | 2017-18 | 82 | 54 | 23 | 0 | 5 | 113 |
| Boston Bruins | 2017-18 | 82 | 50 | 20 | 0 | 12 | 112 |
| Vegas Golden Knights | 2017-18 | 82 | 51 | 24 | 0 | 7 | 109 |
| Toronto Maple Leafs | 2017-18 | 82 | 49 | 26 | 0 | 7 | 105 |
| Washington Capitals | 2017-18 | 82 | 49 | 26 | 0 | 7 | 105 |
| Minnesota Wild | 2017-18 | 82 | 45 | 26 | 0 | 11 | 101 |
| Anaheim Ducks | 2017-18 | 82 | 44 | 25 | 0 | 13 | 101 |

How does the NHL calculate points?

A win is 2 points

A overtime loss or shootout loss is 1 point

A loss is no points

There are 31 teams currently in the NHL
There are 2 confrences the East and West:
In the 16 team East there is the Atlantic Division and the Metropolitan
In the 15 team West there is the Central Division and the Pacific
For the datasets that I am considering there is only going to be 30 teams
For the predictor year there will be **31** teams

The teams that I am going to be predicting are:
Anaheim Ducks
Boston Bruins
Buffalo Sabres
Calgary Flames
Carolina Hurricanes
Chicago Blackhawks
Colorado Avalanche
Columbus Blue Jackets
Dallas Stars
Detroit Red Wings
Edmonton Oilers
Florida Panthers
Los Angeles Kings
Minnesota Wild
Montreal Canadiens
Nashville Predators
New Jersey Devils
New York Islanders

New York Rangers
Ottawa Senators
Philadelphia Flyers
Phoenix Coyotes
Pittsburgh Penguins
Saint Louis Blues
San Jose Sharks
Tampa Bay Lighting
Toronto Maple Leafs
Vancouver Canucks
**Vegas Golden Knights**
Washington Capitals
Winnipeg Jets

| | | |
|---|---|---|
| timeOnIce | character [1] | '667:12' |
| assists | double [1] | 5 |
| goals | double [1] | 7 |
| pim | double [1] | 15 |
| shots | double [1] | 45 |
| games | double [1] | 59 |
| hits | double [1] | 48 |
| powerPlayGoals | double [1] | 0 |
| powerPlayPoints | double [1] | 0 |
| powerPlayTimeOnIce | character [1] | '35:30' |
| evenTimeOnIce | character [1] | '530:04' |
| penaltyMinutes | character [1] | '15' |
| faceOffPct | double [1] | 50 |
| shotPct | double [1] | 15.6 |
| gameWinningGoals | double [1] | 1 |
| overTimeGoals | double [1] | 0 |
| shortHandedGoals | double [1] | 2 |
| shortHandedPoints | double [1] | 2 |
| shortHandedTimeOnIce | character [1] | '101:38' |
| blocked | double [1] | 16 |
| plusMinus | double [1] | 0 |
| points | double [1] | 12 |
| shifts | double [1] | 954 |
| timeOnIcePerGame | character [1] | '11:18' |
| evenTimeOnIcePerGa... | character [1] | '08:59' |
| shortHandedTimeOnI... | character [1] | '01:43' |
| powerPlayTimeOnIceP... | character [1] | '00:36' |

| | | |
|---|---|---|
| gamesPlayed | double [1] | 82 |
| wins | double [1] | 46 |
| losses | double [1] | 29 |
| ot | double [1] | 7 |
| pts | double [1] | 99 |
| ptPctg | character [1] | '60.4' |
| goalsPerGame | double [1] | 2.415 |
| goalsAgainstP... | double [1] | 2.354 |
| evGGARatio | double [1] | 1.0246 |
| powerPlayPerc... | character [1] | '15.6' |
| powerPlayGoals | double [1] | 50 |
| powerPlayGoa... | double [1] | 54 |
| powerPlayOp... | double [1] | 320 |
| penaltyKillPerc... | character [1] | '82.8' |
| shotsPerGame | double [1] | 28.8049 |
| shotsAllowed | double [1] | 27.5244 |
| winScoreFirst | double [1] | 0.725 |
| winOppScoreF... | double [1] | 0.405 |
| winLeadFirstPer | double [1] | 0.806 |
| winLeadSecon... | double [1] | 0.879 |
| winOutshootO... | double [1] | 0.609 |
| winOutshotBy... | double [1] | 0.514 |
| faceOffsTaken | double [1] | 4286 |
| faceOffsWon | double [1] | 2160 |
| faceOffsLost | double [1] | 2126 |
| faceOffWinPer... | character [1] | '50.4' |
| shootingPctg | double [1] | 8.4 |
| savePctg | double [1] | 0.914 |

```r
library(rjson)

Root.URL <- "https://statsapi.web.nhl.com"
URL <- NULL
Stats.Endpoint1 <- "/stats?stats=statsSingleSeason&season=20072008"
Stats.Endpoint2 <- "/stats?stats=statsSingleSeason&season=20082009"
Stats.Endpoint3 <- "/stats?stats=statsSingleSeason&season=20092010"
Stats.Endpoint4 <- "/stats?stats=statsSingleSeason&season=20102011"
Stats.Endpoint5 <- "/stats?stats=statsSingleSeason&season=20112012"
Stats.Endpoint6 <- "/stats?stats=statsSingleSeason&season=20122013"
Stats.Endpoint7 <- "/stats?stats=statsSingleSeason&season=20132014"
Stats.Endpoint8 <- "/stats?stats=statsSingleSeason&season=20142015"
Stats.Endpoint9 <- "/stats?stats=statsSingleSeason&season=20152016"
Stats.Endpoint10 <- "/stats?stats=statsSingleSeason&season=20162017"
Stats.Endpoint11 <- "/stats?stats=statsSingleSeason&season=20172018"
```

```
Roster.2017.2018 <- fromJSON(file = URL.2017.2018)
Roster.2016.2017 <- fromJSON(file = URL.2016.2017)
Roster.2015.2016 <- fromJSON(file = URL.2015.2016)
Roster.2014.2015 <- fromJSON(file = URL.2014.2015)
Roster.2013.2014 <- fromJSON(file = URL.2013.2014)
Roster.2012.2013 <- fromJSON(file = URL.2012.2013)
Roster.2011.2012 <- fromJSON(file = URL.2011.2012)
Roster.2010.2011 <- fromJSON(file = URL.2010.2011)
Roster.2009.2010 <- fromJSON(file = URL.2009.2010)
Roster.2008.2009 <- fromJSON(file = URL.2008.2009)
Roster.2007.2008 <- fromJSON(file = URL.2007.2008)


Roster.2017.2018 <- Roster.2017.2018$teams
Roster.2016.2017 <- Roster.2016.2017$teams
Roster.2015.2016 <- Roster.2015.2016$teams
Roster.2014.2015 <- Roster.2014.2015$teams
Roster.2013.2014 <- Roster.2013.2014$teams
Roster.2012.2013 <- Roster.2012.2013$teams
Roster.2011.2012 <- Roster.2011.2012$teams
Roster.2010.2011 <- Roster.2010.2011$teams
Roster.2009.2010 <- Roster.2009.2010$teams
Roster.2008.2009 <- Roster.2008.2009$teams
Roster.2007.2008 <- Roster.2007.2008$teams
```

```
NJ.2007.2008 <- Roster.2007.2008[[1]][["roster"]]
NJ.2008.2009 <- Roster.2008.2009[[1]][["roster"]]
NJ.2009.2010 <- Roster.2009.2010[[1]][["roster"]]
NJ.2010.2011 <- Roster.2010.2011[[1]][["roster"]]
NJ.2011.2012 <- Roster.2011.2012[[1]][["roster"]]
NJ.2012.2013 <- Roster.2012.2013[[1]][["roster"]]
NJ.2013.2014 <- Roster.2013.2014[[1]][["roster"]]
NJ.2014.2015 <- Roster.2014.2015[[1]][["roster"]]
NJ.2015.2016 <- Roster.2015.2016[[1]][["roster"]]
NJ.2016.2017 <- Roster.2016.2017[[1]][["roster"]]
NJ.2017.2018 <- Roster.2017.2018[[1]][["roster"]]
```

This block was replicated 30 times

The last line is used once more for Vegas

```
for(i in 1:length(ANA.2007.2008$roster)){
  Names.ANA.2007.2008[i] <- ANA.2007.2008$roster[[i]]$person$fullName
  Position.ANA.2007.2008[i] <- ANA.2007.2008$roster[[i]]$position$code
  Links.ANA.2007.2008[i] <- ANA.2007.2008$roster[[i]]$person$link
}
```

```
for(i in 1:length(Links.ANA.2007.2008)){
  URL <- paste(Root.URL, Links.ANA.2007.2008[[i]], Stats.Endpoint1, sep = "")
  Roster.ANA.2007.2008[[i]] <- fromJSON(file = URL)
}
```

```
for(i in 1:length(Roster.ANA.2008.2009)){
  Roster.ANA.2008.2009[[i]] <- Roster.ANA.2008.2009[[i]]$stats[[1]]$splits[[1]]$stat
}
```

**Methods**

Linear Regression

Logistic Regression

Gradient Boosted Decision Trees

*Other Trees* If time Permits

The plan currently is that I am going to use these three different methods to predict team points for the 2017-2018 season using some of these variables from the 2007-2008 to the 2016-2017 seasons.

TimeLine:

| Item | Rough Week Length |
|---|---|
| Sum of Squares | 1 |
| Logistic Regression | 2-3 |
| Gradient Boosted Tree | 2-3 |
| Other Regression Trees | 1-2? |
| Analysis | 1.5 |
| Report Writing | 1.5 |


Joshua Weissbock Forecasting Success in the National Hockey League using In-Game Statistics and Textual Data School of Electrical Engineering and Computer Science Faculty of Engineering University of Ottawa, Ottawa, Canada, 2014


Robert Vollman Aug 24, 2016
http://www.hockeyabstract.com/thoughts/abriefhistoryofhockeyanalytics


Mark J Burns Sep 25, 2014
https://www.forbes.com/sites/markjburns/2014/09/25/assistant-gm-kyle-dubas-injects-youth-smarts-into-maples-leafs-front-office/684caa5c188e


Josh Cooper October 10 2018
*http* : //*www.espn.com*/*nhl*/*story*//*id*/20922909/*nhl* − *arizona* − *coyotes* − *gm* − *john* − *chayka* − *more* − *just* − *numbers* − *nerd*


http://arizonasports.com/story/1346879/coyotes-chayka-happy-home/

https://torontosun.com/sports/hockey/nhl/toronto-maple-leafs/soo-roots-run-deep-for-new-maple-leafs-gm-kyle-dubas