Put the Biscuit in the Basket:

A data mining approach to predicting NHL free agent salaries

SYS 6018 Final Project Report

Jack Prominski, Court Haworth, Tyler Lewris

9 December 2017

Identifying the Problem

General Managers (GMs) in the National Hockey League (NHL) are charged with a difficult task: they must allocate limited resources in a competitive and uncertain environment, all while having their performance and decisions be the subject of relentless public scrutiny. The increasingly widespread adoption of advanced analytical techniques in the hockey community has both increased this scrutiny and made the market more competitive. As more teams embrace data, they will be armed with the information to make better decisions. Despite the increased analytical adoption, irrational actors and market forces dictate that the market is still inefficient, and that there are opportunities for a shrewd GM to exploit.

As background, NHL GMs, broadly, are responsible for putting a team on the ice. They are intimately concerned with how they can optimize their spending on player contracts to maximize their team's success, subject to the constraint of the salary cap. Particularly, they would like to avoid both low-balling free agent contract offers, at the risk of the player going to another team, as well as overpaying players beyond their market value. Players and their agents are also interested in this information. They would like to know what they stand to earn in the open market and can use that information as leverage at the negotiating table or to evaluate a contract offer. Beyond contract evaluation, GMs would also like to identify bargains in the open market. We believe that this aspect can deliver the highest impact. If GMs can more accurately predict a player's value on the open market, they will be able to exploit market inefficiencies and assemble the best team possible, thus leading to a higher chance of winning more games, and ultimately, the Stanley Cup.

Objectives & Metrics

We propose to explore several modeling techniques to accurately predict what a free agent will get paid in the open market. We will also determine what his value should be, as determined by his past performance. We will be predicting the salary cap hit (henceforth just referred to as salary) of 280 free agent signings in the 2017 offseason. We will join this dataset, which is our response variable, with another dataset, which includes NHL performance metrics and salaries over the past three seasons as well as other demographic information. As forwards and defensemen have different jobs and their values are determined differently, we have also split up our dataset by position and will model these separately. At the end of this report we have included a glossary that provides definitions of the variable names that we discuss. We have measured the accuracy of our models using a cross validated Mean Absolute Error (MAE) and R^2. The data mining techniques we used include multiple linear regression, Random Forest, and KNN.

Understanding the State of the Art

There is an active hockey blogging community, and we have found cases of bloggers using linear regression and KNN to tackle this problem.[1][2] These models have been somewhat effective, with the regression model delivering an R^2 of .76 and the KNN model delivering an MAE of roughly $650,000.

Many teams likely have their own models to value players, but these techniques are proprietary and not publicly known. This problem is especially difficult for several reasons. One is that the hockey environment changes relatively quickly. Styles of play go in and out of fashion, and thus the market value of players with particular skill sets fluctuate. Supply of players is also not constant. A player in a relatively weak free agent class may command more in salary than he would in another year. Player performance also fluctuates from year to year and is difficult to predict. Salary cap rules and negotiating rights stipulated by the league also affect how a player is paid. Player potential is also not reflected in any of these models. A GM may be willing to pay a player more when he thinks the player will have a better performance than in previous years. Another factor that makes this problem difficult is that GMs sometimes make irrational decisions, which introduces more noise to the data. Many of these problems are not solvable given our current dataset and approach. However, we hope that the analytical rigor of our modeling techniques can match or beat the performance of other publically available models.

Hypothesis and Approach

We began by developing two hypotheses to explore in our analysis. Our first hypothesis is that the market is inefficient. We believe an efficient market is one in which player performance alone determines a player's salary. And that the better the player has performed, the higher his salary will be. To test this hypothesis, we will build models with variables that only measure performance and models based on our full dataset, which incorporates both performance and additional information. This additional information includes the player's salary history, draft information, height, weight, and nationality. If our hypothesis is correct, then our Full models will outperform our reduced, performance-only (Perf) models. This will indicate that there is more to predicting a player's salary than simply his performance. Perhaps it is the case that GMs are biased towards tall players, or players from Canada/USA, or are unduly influenced by the player's salary last year. These are all examples of biases that we believe may exist in the market that create a departure between true player value and player salary.

There also may be variables that our dataset does not include, which may be predictive of player salary. For instance, a player's star-factor could inflate his salary. Perhaps he draws large crowds, and so the team generates more revenue from ticket sales and is willing to pay him more. Income inequality in the market shows that there are a few players who earn very large contracts, and many players who are at or around the league minimum of $650,000. It is possible that this inequality will show up in our analysis and show that many players who are paid the league minimum are actually worth more. If our hypothesis that the market is inefficient is true, GMs empowered by our models will be able to exploit market inefficiencies. Accurate modeling of player value will allow GMs to identify undervalued and overvalued players, and act accordingly. They will be able to sign contracts with better players who won't command as much in salary and avoid those players whose salary demands are too high relative to their value.

Our second hypothesis is that a linear regression model will perform better than both the KNN and Random Forest models. "As a general rule, parametric models tend to outperform non-parametric approaches when there are a small number of observations per predictor."[3] Our data exhibit a "small n,

large p" problem, which will make all modeling challenging. Given that we only have 190 observations in our forward dataset, 90 in our defensive dataset, along with 90 predictor variables, a linear model performing better would make logical sense. This hypothesis is easily testable as we can compare summary statistics, MAEs, and compare plots of predicted values. We also feel linear regression will give us more insight about the role each predictor plays. If it turns out a KNN or Random Forest model performs better, we may still draw more meaningful conclusions from a linear regression model. This is because of the bias-variance tradeoff. Even though we would be sacrificing prediction accuracy and accepting additional bias, we would significantly improve the interpretability of the model which helps us better understand the role of each predictor.

Additionally, we anticipate a conventional KNN model not to perform as well as the other techniques. KNN in particular performs poorly in a high dimensional space. We expect the curse of dimensionality to be at play here because we are spreading 90 and 190 observations over 90 dimensions. Thus, there is a chance some observations may not have any nearby neighbors. However, we are taking a novel approach to KNN, as described below, that will help us address these concerns.

Executed Approach and Results

Linear Regression Approach:
We have identified a clear hypothesis and are looking to determine whether or not a linear regression approach can uncover more insight about the predictive variables associated with hockey performance metrics versus the predictive variables associated with the full hockey dataset. The objective is to create four of the best linear models: one from the defense full dataset, one from the offense full dataset, one from the defense performance metrics dataset, and one from the offense performance metrics dataset. Once these models are identified, we want to compare the MAEs, the Adjusted R^2 values, the summary statistics, cross validation results and the specific variables included in the model to test against our hypotheses and ultimately determine if a hockey player's performance is not correlated with his compensation.

First, we read in each of the four datasets (full_D, full_F, perfonly_D, perfonly_F) and appropriately classified each column as either factor or numeric. One interesting observation was that the majority of hockey players in our dataset were from either Canada or USA, with a few observations in other countries such as Sweden and Germany. This factor level imbalance will cause issues later down the road and the added levels most likely will not give us any additional information about the model or our hypotheses. Thus, we converted the nationality column to a binary classification: 1 for Hockey players from either USA or Canada, 0 for any other nationality. We then removed the original nationality column as it is no longer required.

After extensive and thorough data cleaning and exploration, we created a basic linear model to include all of the variables and analyzed the results. With nearly 90 regressor variables, multicollinearity was clearly going to be an issue and this basic linear model confirmed our assumption. Utilizing the VIF (Variance

Inflation Factors) function to try and identify influential or collinear variables, the results were incredibly high.

The next step was to identify an alternative approach to addressing multicollinearity. Lasso is a regression analysis method that performs both variable selection and regularization in order to enhance the prediction accuracy and interpretability of the statistical model it produces. This is exactly the approach we needed to take; to identify the variables in our dataset that were significant and also reduce multicollinearity. We then used lasso regression on each of our four datasets to identify the best models. After identifying the best models, we cross validated each to prevent overfitting and obtain cross validated MAEs. The resulting models are below (all p-values are significant at the 0.05 level):

Lasso Regression	Defense		Forwards	
	Full Model	Perf Metrics Only	Full Model	Perf Metrics Only
Mean Absolute Error	$644,558	$673,349	$603,837	$603,837
Predicted R^2	72.5%	71.9%	84.4%	84.4%
Selected Variables	*xGF + ly_salary + ev_A1 + ev_ixG + ev_iSCF + ev_iTKA + pp_G*	*xGF + ev_A1 + ev_ixG + ev_iTKA + pp_G + pk_A + ev_Prev3PTS + stars*	*ev_PTS + ev_iFOW + pp_A + pp_PTS + stars*	*ev_PTS + ev_iFOW + pp_A + pp_PTS + stars*

This approach worked well as it significantly reduced the number of predictors in our model and removed any multicollinearity that was present in our previous models. However, we wanted to implement another approach to generate additional models and compare against our lasso models. Stepwise regression is an appropriate alternative method as the choice of predictive variables is carried out by an automatic procedure. A variable is considered for addition to or subtraction from the set of explanatory variables based on AIC in each step. We performed forward selection, backward selection, and stepwise selection. We ultimately chose the models resulting from stepwise selection as they take into account both forward and backward and had the lowest resulting AIC. From there, we once again utilized the VIF function and removed the variables with values greater than 10. This completely eliminated any multicollinearity from the models. The resulting models are below (all p-values are significant at the 0.05 level):

Stepwise Regression	Defense		Forwards	
	Full Model	Perf Metrics Only	Full Model	Perf Metrics Only
Mean Absolute Error	$744,523	$757,496	$679,883	$690,240
Predicted R^2	78.8%	78.2%	78.9%	78.4%
Selected Variables	*ev_iTKA + pp_Prev3PPG + ev_Prev3G + ev_G + ev_iHA + pp_Pct. + OTG + GWG + stars + DftYr + ev_Prev3A + CA_U*	*pp_G + ev_iTKA + pk_Prev3TOI + pp_Prev3PPG + NPD + ev_iDS + ev_G + ev_Prev3Corsi + pp_Prev3PPA + pp_A1 + Grit + OTG*	*ev_iFOW + AGE + pp_G + pp_TOI.GP + ev_iTKA + Grit + pp_A1 + pk_RelPct. + Eplusminus + pk_A + pk_G*	*ev_iFOW + Eplusminus + pk_RelPct. + pp_G + ev_iHF + ev_iTKA + pp_A1 + pp_Prev3PPA + ev_iRB + pk_A + pk_TOI*

This approach also worked well as we were confidently able to reduce the number of predictors and remove any issues with multicollinearity from our models. Similar to our lasso, we cross validated our stepwise models. We performed 3 different K-fold cross validations settings K = 3, 5, and 10. For the purpose of this project, K=5 made the most sense both computationally and logically. We utilized the cross validation predictions, which are the predictions generated from training the model on the observations not in the K-fold, to calculate our MAEs. Given we do not have a testing dataset, the cross validation predictions are informative to use as they better simulate the model's performance on unseen data.

We now have eight models, four generated from lasso regression and four generated from stepwise regression. We used anova tables, performed residual analysis, and compared adjusted R^2 and MAE values to ultimately select the best models. Although there were fluctuations among the R^2 values, the lasso regression models were typically more attractive.

After selecting the four models generated through lasso regression, we performed an in-depth analysis of residual plots, influence measures, and qq plots. The residual plots all had a similar pattern: slight fanning out but distributed around 0. There were no significant influential points in any of the models. The qq plots also had very similar findings: departure towards the tails. The residual and qq plots indicate that there may be an issue with our assumption of normality in the data. Perhaps a linear model may not be the best fit to these datasets or a transformation of variables may be necessary. We attempted a few transformations on the variables that showed patterns. Specifically, we square rooted *pp_A1* and *pk_A* because of their inward fan patterns and took the log of *xGF* because of its outward fan pattern. After transforming the variables and applying them to our models, they did not significantly improve our residual plots or our summary statistics. Thus, we did not follow through with any transformations.

Linear Regression Results:
Our best models are below:

Best Models	Defense		Forwards	
	Full Model	Perf Metrics Only	Full Model	Perf Metrics Only
Mean Absolute Error	$644,558	$673,349	$603,837	$603,837
Predicted R^2	72.5%	71.9%	84.4%	84.4%
Most Significant Variables	*ly_salary, ev_iTKA, pp_G, xGF*	*ev_iTKA, pp_G, ev_Prev3PTS*	*ev_iFOW, pp_A, star*	*ev_iFOW, pp_A, star*

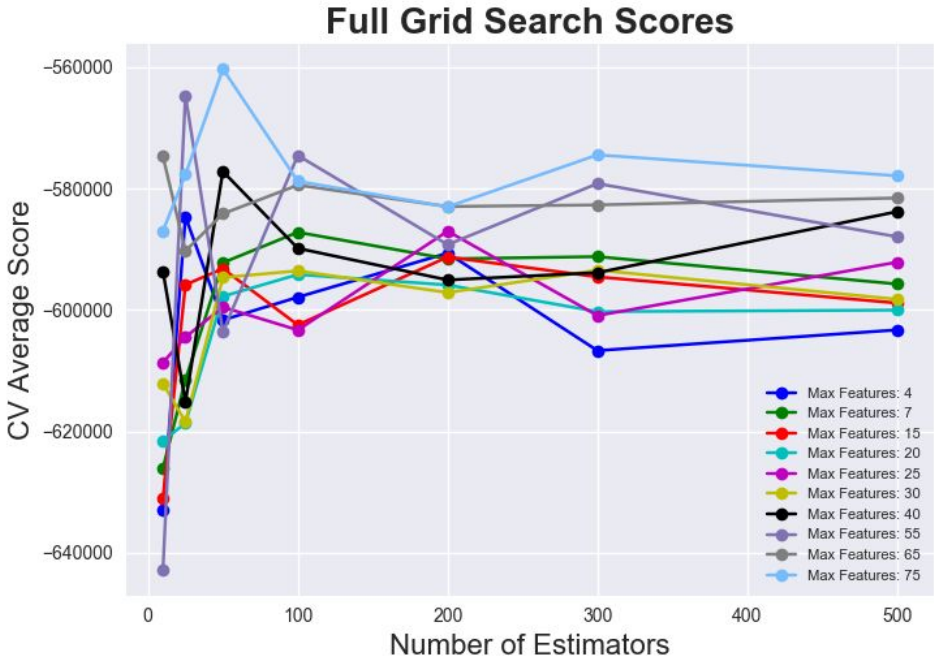
The results of our linear regression models do not prove our hypothesis. Creating a model using only performance metrics does not significantly differ from models created using all possible regressors. When comparing the defense models, we see that the adjusted R^2 is only marginally higher when using the full dataset. However, the MAE is slightly lower, by roughly $30,000. This may indicate that the NHL salary of defensive players can be better predicted with variables outside of only performance, but given the similar adjusted R^2 values the MAE alone is not significant enough to verify this assumption. Additionally, both models had two of the same statistically significant variables: *ev_iTKA* and *ev_Prev3PTS*. When comparing the offense models, they are identical models with the same adjusted R^2 and MAE values. The models also have the same significant variables which appear in the performance only dataset: *ev_iFOW*, *pp_A*, and *stars*. This may indicate that the NHL salary of offensive players can be better predicted using only performance metrics as the resulting models are the same. However, with an MAE of roughly $600,000 making this assumption would be a reach. Thus, using linear regression methods does not seem to provide us with concrete evidence that NHL salaries are not correlated with performance metrics.

Random Forest Approach:
After tackling a linear model approach to predicting salary, we decided to tackle alternative methods, the first of which being a Random Forest model. There are a number of advantages to using a Random Forest approach. One advantage of Random Forests is that they do not assume linearity of variables. Often times the relationship between predictors and response variables is not a linear one and Random Forest's make no assumptions about these relationships. Another advantage is that Random Forests are an ensemble of many decision trees combined together into an approach that does well to avoid overfitting while simultaneously achieving strong predictive performance. The other side of the coin is that Random Forests sacrifice interpretability for performance, making it difficult to understand the importance of each regressor.

Due to the nature of the Random Forest algorithm, there was no need to separate the defenseman and the forwards in our datasets. As such, the first task was to combine the data into two big datasets, the Full dataset and the dataset containing Performance Only statistics. A variable was added corresponding to

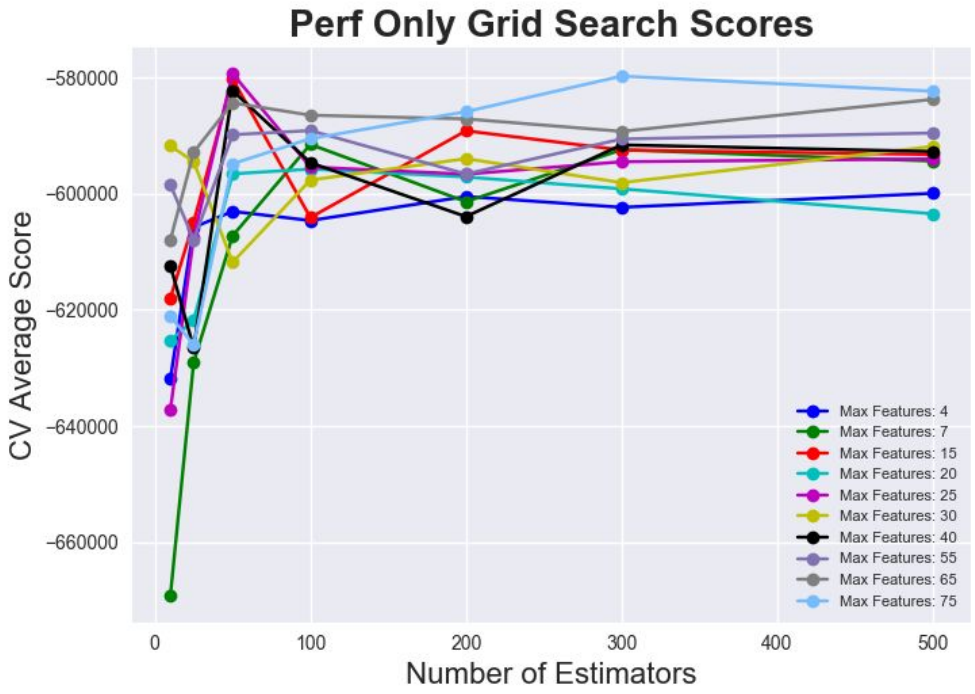
position (either Forward or Defense) so the trees could split on this if necessary. The next step was to perform grid search cross validation in order to optimize the hyperparameters. The two parameters that were deemed most necessary were the number of estimators used and the max number of features considered for each decision tree in the random forest. Seven possible values of n_estimators were used, ranging from 10 to 500, and ten values of max_features were used ranging from 4 to 75. MAE was used to choose the best performance. Three fold cross validation was used, to determine the optimal set of parameters. The parameter choices that had the lowest MAE across the three folds were chosen to be used in the final model.

Random Forest Results:
First we consider the results of the model built on the full range of variables. Below is a graph of the cross validated average score for each combination of number of estimators and max features considered. One note of importance is that, by default, the cross validation method used considers negative MAE. As a result, the maximum values on this graph is what we are interested in as it corresponds to the lowest MAE.



As you can see, the model that performed best was 75 features, with 50 estimators. This model achieved a cross validated average score of $560,319. When compared to our linear models, this MAE is significantly lower, contrary to our hypothesis that the linear model would perform better. Another note of importance is that once a certain number of estimators is hit, around 100, any increase in estimators does not significantly improve error. This result is counterintuitive as conventional knowledge suggests that considering fewer features tends to lead to better Random Forest performance. In this case, a large number of features has led to our best results. This may be due to the small sample size or another issue worth investigating in the future.

Next, we consider the results of a series of Random Forests fit on the performance statistics only.



For this data, the model that performed best was 25 features with 50 estimators. This model achieved a cross validated average score of $579,321. This, again, indicates our Random Forest model performs better than our linear regression models when comparing MAEs.

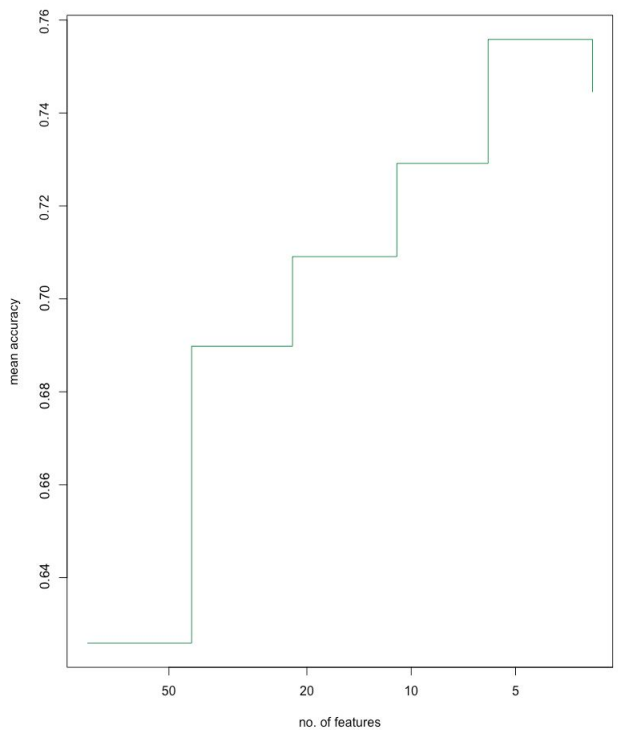
When comparing the two models, there is no noticeable difference in model effectiveness with models generated using performance only metrics versus all possible regressors. These findings do not help prove our hypothesis that we can better predict salary with full models as compared to performance-only models. While the full model did perform better, it was not by a large margin, only a 3% difference. The full model also had more variables to choose select which could explain its improved performance. Consequently, it does not seem as though there is a significant difference in predictive power for all variables versus performance only statistics.

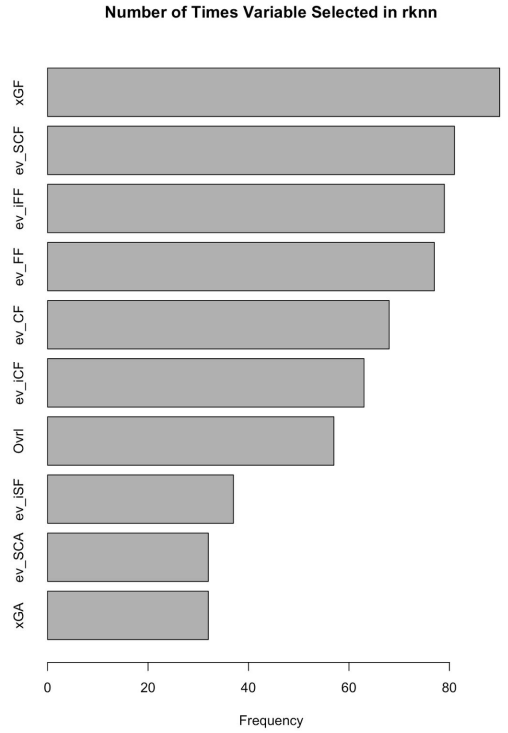
KNN Approach:

The final data mining technique we explored was K-Nearest Neighbors regression. KNN tends to perform poorly with high dimensional data, and especially with a small sample size, so we needed to investigate variable selection strategies to dramatically reduce dimensionality. Fortunately, there exists an R package designed for exactly this scenario. The rknn package was developed to address "small n, large p problems."[4] Its Random KNN technique, similar to Random Forest, is an ensemble of single KNN models constructed from a random sampling of variables. Unlike Random Forest though, rknn can also be used for feature selection through its process of ranking feature importance and recursive backward model selection. One advantage of rknn over Random Forest is that rknn's feature selection technique outputs a single "best" KNN model, which is more interpretable than Random Forest's ensemble of trees. I used

rknn's feature selection tool, and then ran a base cross validated KNN regression model using the FNN package, using the optimal variables derived from rknn.

There are several parameters for rknn, including *k*, the number of nearest neighbors to use for prediction; *r*, the number of iterations of KNN to run; *mtry*, the number of variables to select for each iteration; and *stopat*, the minimum number of variables to keep in each model. I tuned these parameters, selecting the ones that resulted in models with the highest R^2, although model performance did not seem very sensitive to changes in these parameters. I eventually decided on 5 for *k*, 500 for *r*, the square root of *p* (8 for perf and 9 for full datasets) for *mtry*, and 2 for *stopat*. Below you can see two visualizations derived from rknn's output. The figure on the left demonstrates the process that rknn takes when optimizing for the number of variables in a model. This informed the selected values of *mtry* and *stopat.* The figure on the right displays, for the Full Defensive model, the frequency of a variable being selected during rknn's model selection.





We experienced a high degree of variance in the predicted R^2 when running the same models multiple times, even with a sufficiently large *r* in the rknn step, so we decided to run the feature selection and model building steps 100 times and average the outputs to reduce this variance.

KNN Results:

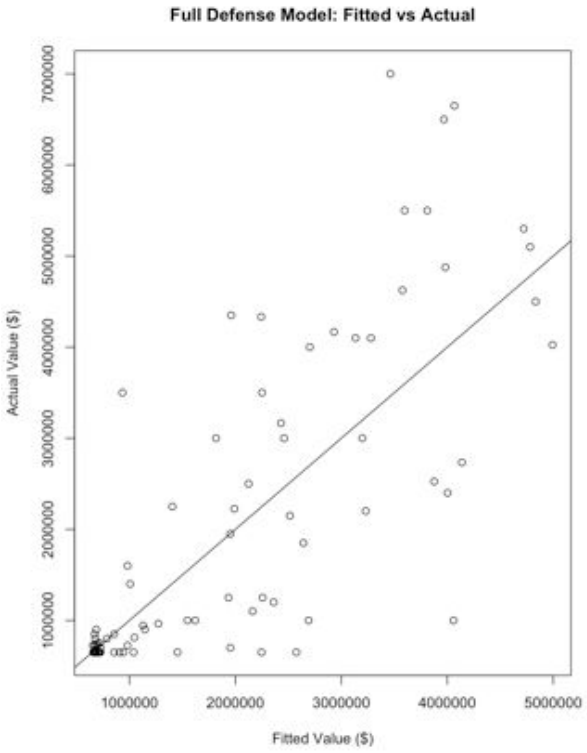
The results were in line with the other techniques. Full models were not able to significantly outperform our reduced, performance-metric only models. The variables selected in each model, in fact, were very

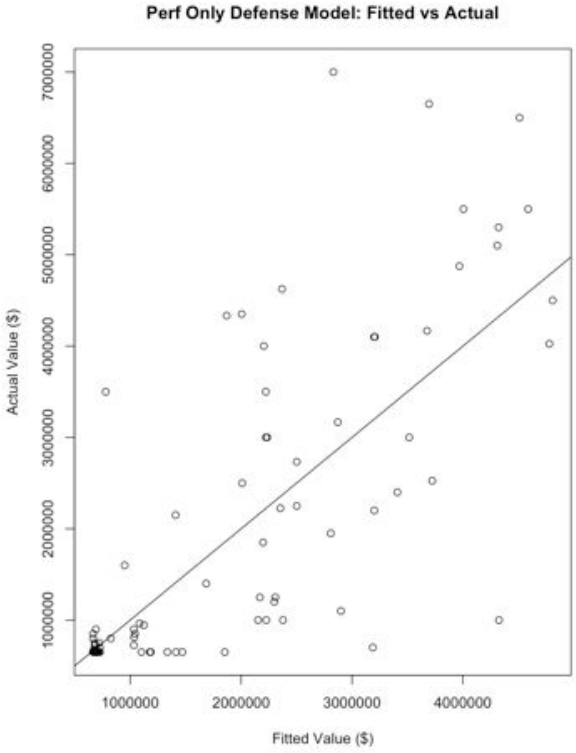
similar between the full and performance models, so the additional variables available in the full model were not even incorporated. The results are below:

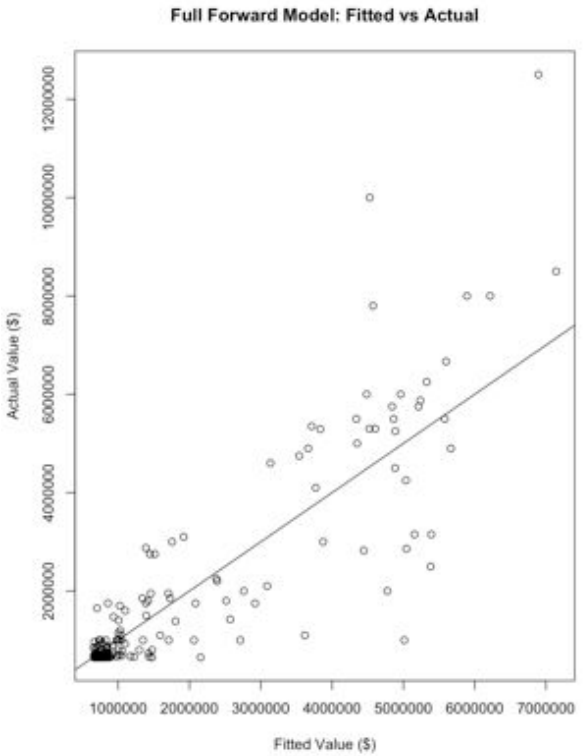
	Defense		Forwards	
	Full Model	Perf Metrics Only	Full Model	Perf Metrics Only
Mean Absolute Error	$691,130	$666,516	$527,185	$519,263
Predicted R^2	53.9%	59.6%	71.7%	73.3%
Most commonly selected variables	*xGF, ev_iFF, ev_FF, ev_SCF, ev_iCF*	*xGF, ev_iFF, ev_FF, ev_SCF,ev_iCF*	*xGF, ev_PTS, pp_PTS, pp_TOI, pp_A*	*xGF, pp_TOI, ev_FF, ev_PTS, pp_PTS*

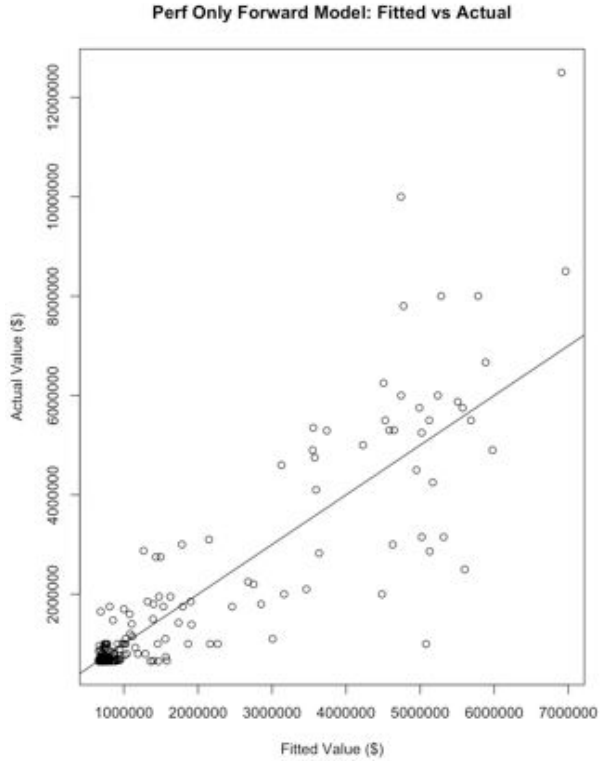
We were surprised that *ly_salary* was not selected in the Full_D or Full_F models, given how important the variable was to the linear regression model. I wondered if scaling this predictor would result in it being chosen, however a scaled *ly_salary* was still not selected by rknn. KNN performance can often be improved by normalizing the variables involved, as KNN uses Euclidean distance to determine the nearest neighbors. Without scaling or applying another normalization technique, a variable with a larger scale can result in that variable having an outsized importance in the model and hurting model performance.

As you can see visualized below and in the table above, the defense models see much higher variance between the predicted and actual values and do not perform as well. Since the defense data set was only 90 observations, while the forwards data set was 190, we believe that the small sample size may be affecting the strength of our defensive model. Additionally, as mentioned earlier, you can visually see below how similar the two forward models are and how similar the two defense models are.


Full Defense Model: Fitted vs Actual
Actual Value ($)
Fitted Value ($)
1000000
2000000
3000000
4000000
5000000
6000000
7000000


Perf Only Defense Model: Fitted vs Actual
Actual Value ($)
Fitted Value ($)
1000000
2000000
3000000
4000000
5000000
6000000
7000000


Full Forward Model: Fitted vs Actual
Actual Value ($)
Fitted Value ($)
2000000
4000000
6000000
8000000
10000000
12000000
1000000
2000000
3000000
4000000
5000000
6000000
7000000


Perf Only Forward Model: Fitted vs Actual
Actual Value ($)
Fitted Value ($)
2000000
4000000
6000000
8000000
10000000
12000000
1000000
2000000
3000000
4000000
5000000
6000000
7000000

Conclusion

Through this analysis, we have concluded that, based on our R^2 and MAE values, the market is somewhat inefficient. Player performance does not completely predict salary, as we believe it would in an efficient market. Our hypothesis that we would be able to better predict salary with full models as compared to performance-only models was incorrect. Metrics indicated very little difference between the models built on the two sets of variables. There are several additional considerations that may explain some of our results. It is possible that available statistical measures do not accurately describe true player value. Though advanced statistics have recently been developed and are supposed to more accurately quantify a player's value to his team, it is possible that they still fall short. Or, perhaps a full model is actually more predictive of salary, and our dataset is missing key variables. If either of these things are true, then we cannot accurately test this hypothesis.

Our second hypothesis was that linear regression will outperform the other techniques, given our dataset contains very few observations per predictor. After further investigation, it seems as though this hypothesis is also invalid. The Random Forest algorithm achieved a significantly lower cross validated MAE indicating it is a better modeling technique to tackle our dataset. The KNN offense models, too, outperforms the linear regression model in terms of MAE. Even though we have a small sample size and large number of predictors, the more flexible models in general performed slightly better. One possible explanation for this is that we underwent significant dimensionality reduction in our model building process. Though we were still working with a small sample size, it's possible that the number of observations per predictor dropped to a point where an inflexible model no longer provided advantages.

While our models did not perform as well as we had hoped, we believe that our framing of the problem and our approach to solving it are still valid. General Managers are always looking for an advantage over their competition, and we believe that superior analytics can deliver that. Our approach could also be generalized to other sports with more predictive data, or frankly, any situation where prices differ from underlying value. This is the basis for many aspects of the global economy. We raised many issues in the Understanding the State of the Art section of this report that still remain unsolved and make this problem especially challenging. The game changes quickly, market forces are out of our control, and maybe most of all, we are dealing with humans who change, grow, and have biases.

Variable Glossary

pp: powerplay
ev: even strength
pk: penalty kill

A: Assists by this individual
FF: The team's unblocked shot attempts (Fenwick, USAT) while this player was on the ice
G: Goals scored by this individual
iCF: Shot attempts (Corsi, SAT) taken by this individual
iFF: Unblocked shot attempts (Fenwick, USAT) taken by this individual
iFOW: Faceoffs won by this individual
iTKA: Takeaways by this individual
ly_salary: The individuals 2016-2017 salary
Prev3PTS: Total points scored by this individual between 2013-2016
PTS: Goals plus all assists scored by this individual
SCF: The team's scoring chances while this player was on the ice
Stars: Total number of times this individual was awarded the 1st, 2nd, or 3rd star of a game
TOI: Time this individual spent on the ice
xGF: The team's expected goals (weighted shots) while this player was on the ice

References

[1] “Predicting Free Agent Salaries,” puck , 28-Jun-2015. [Online]. Available: https://puckplusplus.com/2015/06/28/predicting-free-agent-salaries/. [Accessed: 09-Dec-2017].

[2] A. E. Perry, “Hockey and Euclid: Predicting AAV With K-Nearest Neighbours,” Corsica. [Online]. Available: http://www.corsica.hockey/blog/2016/10/31/hockey-and-euclid-predicting-aav-with-k-nearest-neighbours/. [Accessed: 09-Dec-2017].

[3] G. James, D. Witten, T. J. Hastie, and R. J. Tibshirani, An Introduction To Statistical Learning: with applications in R, 109. New York: Springer, 2017.

[4] Li, S., Harner, E. J., & Adjeroh, D. A. (2011). Random KNN feature selection - a fast and stable alternative to Random Forests. BMC Bioinformatics, 12(1), 450. https://doi.org/10.1186/1471-2105-12-450

[5] Li, S. (2009). *Random KNN modeling and variable selection for high dimensional data* (Order No. 3381197). Available from ProQuest Dissertations & Theses Global. (305031503).

[6] Liaw, Andy & Wiener, Matthew. (2001). Classification and Regression by RandomForest. Forest. 23. .

[7] C. Gaines, “The 25 highest-paid players in the NHL,” Business Insider, 08-Oct-2017. [Online]. Available: http://www.businessinsider.com/nhl-highest-paid-players-2017-10/#25-patrick-marleau-8500000-1. [Accessed: 09-Dec-2017].

[8] A. Krieckhaus, “Show Me the Money! Are NHL Players Paid to Play or to Win?,” The Hockey Writers, 10-Jun-2017. [Online]. Available: https://thehockeywriters.com/show-me-the-money-are-nhl-players-paid-to-play-or-to-win/. [Accessed: 09-Dec-2017].