Prediction of tiers in the ranking of ice hockey players - Extended version

Timmy Lehmus Persson   Haris Kozlica   Niklas Carlsson   Patrick Lambrix

Linköping University, Sweden

Abstract. Many teams in the NHL utilize data analysis and employ data analysts. An important question for these analysts is to identify attributes and skills that may help predict the success of individual players. This study uses detailed player statistics from four seasons, player rankings from EA's NHL video games, and six machine learning algorithms to find predictive models that can be used to identify and predict players' ranking tier (top 10%, 25% and 50%). We also compare and contrast which attributes and skills best predict a player's success, while accounting for differences in player positions (goalkeepers, defenders and forwards). When comparing the resulting models, the Bayesian classifiers performed best and had the best sensitivity. The tree-based models had the highest specificity, but had trouble classifying the top 10% tier players. In general, the models were best at classifying forwards, highlighting that many of the official metrics are focused on the offensive measures and that it is harder to use official performance metrics alone to differentiate between top tier players.

1 Introduction

The success of a sports team depends a lot on the individual players making up that team. However, not all positions on a team are the same. In ice hockey there are three main types of players: goalkeepers, defenders and forwards. While evaluating players it is therefore important to take into account these types.

In this paper, we compare and contrast which attributes and skills best predict the success of individual ice hockey players in different positions. First, using the method in [13] we investigate which performance features were important for the three main position types in the National Hockey League (NHL) for four different seasons. For the data processing, feature selection and analysis we used R 3.6.3 and packages dplyr 0.8.3, ggplot2 3.0.0, gridExtra 2.3 and caret 6.0 as well as Weka 3.8.4 [6]. Our work (including [13] for football) distinguishes itself from other work on player valuation or player performance, by working with tiers of players, i.e., the top 10%, 25% and 50% players in different positions (in contrast to individual ratings). An exact ranking of players may not always be available, and for several tasks, e.g., scouting, an exact ranking of players is not necessary. In these cases using tiers is a useful approximation. Further, we deal with many skills.

Second, we evaluate different techniques for generating prediction models for players belonging to the different top tiers of players. We used Weka 3.8.4 for estimation of the models. We found that the two Bayesian classifiers performed best and that, in general, the models were best at classifying forwards.

The remainder of the paper is organized as follows. Sect. 2 presents related work. Sect. 3 discusses the data sets and the data preparation. Sect. 4 and 5 present the feature selection and prediction methods, respectively, and show and discuss the corresponding results. Finally, the paper concludes in Sect. 6.

2 Related work

In many sports work has started on measuring player performance. For the sake of brevity, we address the related work in ice hockey.

Many of the models for evaluating player performance in ice hockey define a particular stat or evaluation measure that assigns values based on particular types of actions in the game. For instance, the well-known goal measure, assist measure, and the more recent Fenwick and Corsi measures[1] attribute a value to goal-scoring actions, to passes that lead to goals and to different types of shots, respectively. To deal with some of the weaknesses of traditional measures new approaches have been proposed, including regression models replacing the +/- measure (e.g., [11, 12, 3]). One main recognized weakness is the lack of influence of the context in which the actions are performed. This is the basis of the work on added goal value [14] that attributes value to goals, but the value of the goal is dependent on the situation in which it is scored.

Recent works often take several kinds of actions into account for defining a measure. For instance, in [4] principal component analysis was performed based on 18 traditional measures and a performance measure based on the four most important components was proposed. Further, many of these approaches also take some context into account. For instance, event impacts for different kinds of actions in [18] are based on the probability that the event leads to a goal (for or against) in the next 20 seconds. Several works model the dynamics of an ice hockey game using Markov games (e.g., [21, 7]). In [15, 19, 20, 8] action-value Q-functions are learned with respect to different targets. The proposed measure in [8] showed the highest correlation to 12 out of 14 traditional measures compared to measures such as +/-, goal-above-replacement, win-above-replacement and expected goals. In [16] the action-value Q-functions are used to define variants of these player impact measures. In [10] action-value Q-functions are used to define measures for pairs of players. Player rankings used for the NHL draft are presented in [17, 9].

---

[1] See, e.g., `https://en.wikipedia.org/wiki/Analytics_(ice_hockey)`.

Table 1: Attributes for field players and goalkeepers. Attributes in italics were removed during data preparation.

Position	Attributes
Field players	Player, Age, Team, POS (position), GP (games played), *G* (goals), *A* (assists), *PTS* (points), +/-, PIM (penalty minutes), PS (point shares), EVG (even strength goals), PPG (powerplay goals), SHG (shorthanded goals), GWG (game-winning goals), EVA (even strength assists), PPA (powerplay assists) , SHA (shorthanded assistss), S (shots on goal), *S%* (shots on goal percentage), TOI (time on ice), *TOI/60*, BLK (blocks), HIT (hits), FWON (face-offs won), FOL (face-offs lost), *FO%* (face-off percentage), CF (Corsi For), CA (Corsi Against), *CF%* (Corsi For percentage), CF%Rel (Corsi For percentage relative), FF (Fenwick For), FA (Fenwick Against), *FF%* (Fenwick For percentage), FF%Rel (Fenwick For percentage relative), oiSH% (on ice shooting percentage), oiSV% (on-ice save percentage) *PDO*, oZS% (offensive zone start percentage), dZS% (defensive zone start percentage), TOI(EV) (time on ice even strength), TK (takeaways), GV (giveaways), E+/- (expected +/-), SAtt. (shot attempts), Thru% (through percentage), SHFT (shift length), EVTOI (even strength time on ice), GF/60 (even strength Goals For per 60 minutes), GA/60 (even strength Goals Against per 60 minutes), PPTOI (powerplay time on ice), *PPCF%Rel* (powerplay Corsi For percentage relative), *PPGF/60* (powerplay goals for per 60 minutes), *PPGA/60* (powerplay goals against per 60 minutes), SHTOI (shorthanded time on ice), *SHCF%Rel* (shorthanded Corsi For percentage relative), *SHGF/60* (shorthanded Goals For per 60 minutes), *SHGA/60* (shorthanded Goals Against per 60 minutes)
Goalkeepers	Player, Age, Team, GP (games played), GS (game starts), W (wins), L (losses), OTL (overtime losses), GA (goals against), SA (shots against), *SV* (saves), *SV%* (save percentage), *GAA* (goals against average), SO (shutouts), GPS (goalkeeper point shares), MIN (minutes), QS (quality starts), *QS%* (quality starts percentage), RBS (really bad starts), *GA%* (goals against percentage), *GSAA* (goals saved above average), *G* (goals), A (assists), PTS (points), PIM (penalty minutes)

3 Data collection and preparation

3.1 Data collection

The data regarding players was taken from Hockey Reference[2] for the seasons 2015/16 to 2018/19. Different attributes were gathered for goalkeepers and field players. The lists of attributes are given in Table 1. The descriptions of the attributes are given in Appendix G.

The ranking used as a response variable was directly taken from Electronic Arts NHL games between 2016 and 2019 (NHL17, NHL18, NHL19 and NHL20).

[2] `https://www.hockey-reference.com/`

Table 2: Number of players per position with ratings. In parentheses we show the number of players without ratings that were removed from the data set.

Season	Forwards	Defenders	Goalkeepers
2015/16	582 (10)	297 (9)	91 (1)
2016/17	572 (17)	287 (12)	90 (5)
2017/18	555 (28)	297 (10)	93 (2)
2018/19	545 (35)	302 (24)	87 (8)

We use the player rating value that is supposed to be a summary of a player's individual attributes[3]. The range for this value is between 1 and 99.

3.2 Data preparation

The data was then split using player position: goalkeepers, defenders, and forwards[4], resulting in 12 data sets (3 player positions $\times$ 4 seasons). As some of the players did not have a rating in the NHL games, data about these players was removed. Table 2 shows the number of retained players per position and the number of removed players.

For each of the data sets, attributes that were combinations of other attributes were removed. For field players these are G, A, PTS, S%, TOI/60, FO%, CF%, FF%, and PDO. For goalkeepers these are SV, SV%, GAA and QS%. Further, G was removed for goalkeepers as no goalkeeper scored those seasons. For other attributes data was missing and it was decided to impute the value 0 (Thru%, oiSH%, oiSV%, oZS%, dZS%) or remove the attribute (PPCF%Rel, SHCF%Rel, PPGF/60, PPGA/60, SHGF/60, SHGA/60, GA%, GSAA). All temporal attributes were rewritten into seconds. The value for Team was set to the team for which the player played the most games or in case of a tie to the team in which the player ended the season. Numerical data was normalized using the min-max-method to values between 0 and 1.

The rating was used to create the top 10%, 25% and 50% tiers. However, as several players had the same rating it was not always possible to take a tier without having players with the same rating in the tier and outside the tier. Therefore, we decided to use a cutoff such that the actual percentages are less than or equal to the desired percentage for the tier. Using this strategy the actual percentages for the top 10%, 25% and 50% tiers for the different position and seasons were between 6.5% and 9.3%, 19.5% and 25%, and 39.6% and 49.3%, respectively. The exact numbers for each data set are given in Appendix A.

For each of the data sets resulting from the steps above, we made an 80%-20% split where the 80% is used in the feature selection (Sect. 4) and as training set in the prediction (Sect. 5) while the 20% is used as test set in the prediction.

[3] `https://www.ea.com/games/nhl/nhl-20/ratings`

[4] In the original data the forwards were categorized as left wing, right wing, center and wing.

4 Feature selection

4.1 Filter method

Filter methods for feature selection examine data using statistical methods to determine which attributes are relevant. They often use relatively simple calculations and are often relatively fast. We used correlation-based feature selection (CFS) which aims to identify sets of attributes that are highly correlated to the classification, but not correlated with each other [5]. Essentially, CFS computes the Pearson correlation coefficient where all attributes have been standardized and uses this as a measure of merit for the attribute subsets. Further, we used 10-fold cross validation. This results in different subsets for the different runs. We retained the attributes that appeared in at least two of these subsets.

4.2 Wrapper method

Wrapper methods try to identify which subsets of attributes give the best results when used in a model by testing combinations of attributes. Wrapper methods employ a supervised learning method to compute the merit of each subset and are thus dependent on the chosen learning method.

We used the machine learning methods Logistic Regression (LR), Naïve Bayes (NB), Bayesian Network (BN) with $\alpha = 0.1$ and u = 1, Decision Tree (DT) with C = 0.25 and M = 2, k-Nearest Neighbor (KNN) with k = 3 and Random Forest (RF) with I = 100. For the Bayesian methods the attributes should be of nominal type and therefore the values of all numeric-type attributes were discretized by creating ten intervals with a width of 0.1 and ranging from 0 to 1 [2].

We used the Weka settings $\epsilon = 0.01$ and k = 5. This means that we started from the empty set and used best-first search with backtracking after five consecutive non-improving nodes in the search tree. As measure for merit we used AUC. Each algorithm was run over 10 folds and for each attribute and each algorithm the number of folds that contained the attribute was registered. Then for each attribute the mean over this number for the different algorithms was computed and if this mean was larger than 2 the attribute was retained.

4.3 Results and discussion

Table 3 shows the number of attributes that were retained per position, tier and season for both the filter and wrapper methods. Table 4 shows the most common attributes per position for the filter and wrapper methods. The full list of attributes for each data set is given in Appendix B for the filter method and Appendix C for the wrapper method.

For goalkeepers W and QS were common for several tiers in the same season for both methods, while GPS was also common for the filter method. QS was important for all tiers over all seasons for the filter method. For the wrapper method SO was important over all seasons for the 25% and 50% tiers. For defenders PS and PPA were important for all tiers and all seasons for the filter

Table 3: Number of retained attributes for the filter and wrapper methods, respectively. (filter/wrapper).

Season	Tier	Goalkeepers	Defenders	Forwards
	Top 10%	3/7	9/13	10/11
2015/16	Top 25%	2/5	17/8	12/13
	Top 50%	5/5	21/11	22/14
	Top 10%	5/5	11/11	16/14
2016/17	Top 25%	5/5	13/12	14/11
	Top 50%	9/6	17/13	23/9
	Top 10%	2/3	6/9	8/9
2017/18	Top 25%	7/7	11/8	13/8
	Top 50%	5/7	18/12	11/11
	Top 10%	4/6	15/10	11/11
2018/19	Top 25%	6/3	13/18	13/10
	Top 50%	9/6	18/11	20/11

Table 4: Most common attributes per position for filter and wrapper methods.

G-filter	D-filter	F-filter	G-wrapper	D-wrapper	F-wrapper
QS(11)	PS(12)	PS(12)	SO(9)	PS(9)	PS(11)
W(10)	TOI(EV)(12)	PPA(12)	W(9)	TOI(EV) (8)	PPA(10)
GPS(8)	PPA(12)	TOI(EV)(12)	QS(8)	PPA(7)	TOI(EV)(9)
GP(6)	EVA(9)	SHFT(11)	GPS(6)	oiSH%(7)	EVTOI(9)
SO(6)	S(9)	EVTOI, PPTOI(9)	SA, GS(5)	SHG, EVTOI, GA/60, PPTOI(7)	PPTOI(9)

method, while TOI(EV) and S appeared often. For the wrapper method PPTOI and TOI(EV) appeared in all tears for several seasons. For the top 10% tier GA/60 was important for all seasons for the wrapper method, while PPA was important for the top 25% tier. For forwards PS and PPA were important for the filter and wrapper methods and TOI(EV) for the filter method. SHFT was an important attribute for the filter method for forwards, but not so much for defenders. In general, S is more common for top 50% tier players, while PPA is most common for top 25% tier players. Interestingly, PPA is selected more often than EVA. Further, in contrast to the wrapper method, for the filter method it is more common that attributes for a particular tier are selected in different seasons. Season 2017/18 was different in two senses. First, more attributes were selected for defenders and forwards than for the other seasons. Secondly, PPTOI and EVTOI were often selected in other seasons, but not in 2017/18.

We note that many of the selected attributes for field players are measures related to offense (e.g., related to assists, goals and shots) or neutral (e.g., related to time on ice), but the most often occurring measure (PS) relates to both offense and defense. For defenders, there are additionally measures related to goals against. This may reflect the kinds of stats that are collected for players.

In the data preparation step we removed attributes that are combinations of other attributes and these included much used metrics (e.g., goals and assists), which hockey professionals would want to use. Therefore, we investigated whether these metrics 'appeared' in the results, meaning that the attributes on which they depend were selected. Two of the removed combined attributes for goalkeepers appeared often. QS% (as a combination of QS and GS) appeared often both in the filter (5 data sets) and wrapper (8 data sets) methods. SV and SV% (as combinations of SA and GA) appeared in 3 filter data sets. We also note that the removed GA% depends on GA and a league average. Regarding field players, the removed A (as a combination of EVA, PPA and SHA) appeared in 2 filter data sets for defenders. However, as SHA does not appear that often, it is also interesting to check the co-occurrence of EVA and PPA. This appeared in 8 filter data sets for defenders and 8 for forwards as well as in 3 wrapper data sets for forwards. The wrapper data sets contained the less common combination of PPA and SHA (twice for defenders and 3 times for forwards) and EVA and SHA (once for defenders). Similarly, for the removed G (as a combination of EVG, PPG and SHG), EVG and PPG appeared in filter data sets (1 for defenders and 2 for forwards), and a wrapper data set for forwards, while the more unusual combination of PPG and SHG appeared in 2 wrapper data sets for defenders and one for forwards. S% is a combination of EVG, PPG, SHG and S. This combination did not occur. However, the combination EVG, PPG and S occurred in 1 filter and 1 data set for forwards as did the combination EVG and S. The removed CF% (as combination of CA and CF) and FF% (as combination of FA and FF) appeared in filter data sets (2 for defenders and 1 for forwards), and FF% in 1 wrapper data set for defenders and 2 filter data sets for forwards. Further, the removed FO% (as combination of FOW and FOL) occurred in 2 wrapper data sets for defenders and 2 filter data sets for forwards. PDO (as combination of oiSH% amd oiSV%) occurred in filter (1 for forwards) and wrapper (3 for defenders and 1 for forwards) data sets.

5 Prediction

5.1 Methods

For each data set that was used in the feature selection step, we then created two new data sets, one where we used the attributes selected by the filter method and one with the attributes selected by the wrapper method. For the top 10% and top 25% tier data sets we used SMOTE [1] to overcome the class imbalance. This oversampling technique synthetically determines copies of the instances of the minority class to be added to the data set to match the quantity of instances of the majority class.

5.2 Results and discussion

A detailed performance of all algorithms on all data sets is given in Appendix D for the filter data sets and Appendix E for the wrapper data sets. Fig. 1 shows specificity, AUC, F1, sensitivity and accuracy for different seasons, positions, tiers, filter/wrapper and machine learning algorithms. The largest variation among the measures was for F1. Fig. 2 shows F1 for different positions and tiers with respect to season, filter/wrapper and machine learning algorithm.

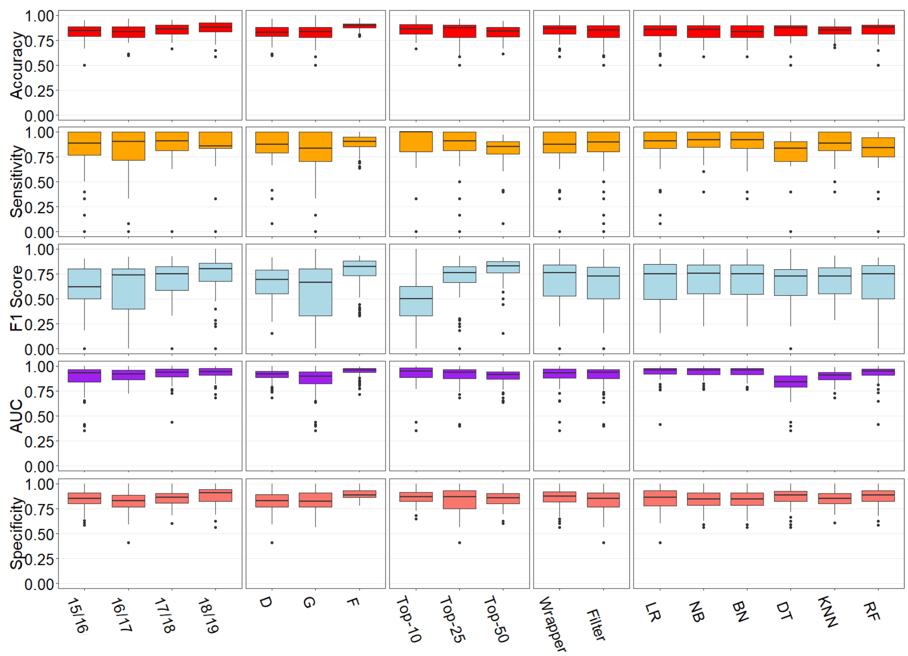

Fig. 1. Specificity, AUC, F1, sensitivity and accuracy for different seasons, positions, tiers, filter/wrapper and machine learning algorithms.

Overall, the choice between the filter and wrapper methods for different measures is not that important (Fig. 1), although for particular tiers and positions there may be a difference (e.g., goalkeepers top 10% and 25%, Fig. 2).

When comparing the resulting models, the two Bayesian classifiers were top performers for most data sets and evaluation measures and performed evenly across all combinations of comparisons. This is in line with the study in [13] regarding football. The tree-based models had the highest specificity, but had a lower sensitivity. They seemed to prioritize the majority class which resulted in lower performance when classifying the top 10% tier, and especially for the smaller data sets (e.g., goalkeepers). Overall, the models achieved high sensitivity, although for small data sets the tree-based models did not do well. In general, the models were best at classifying forwards, highlighting that many of the official metrics are focused on the offensive measures. This suggests that more work is needed to develop equally good defensive metrics. The models also achieved higher F1 for the top 50% highlighting that it is harder to differentiate between the highest rank top tier players using official performance metrics alone.

There is variation over the seasons, reflecting, among others, that different attributes were selected for different seasons.

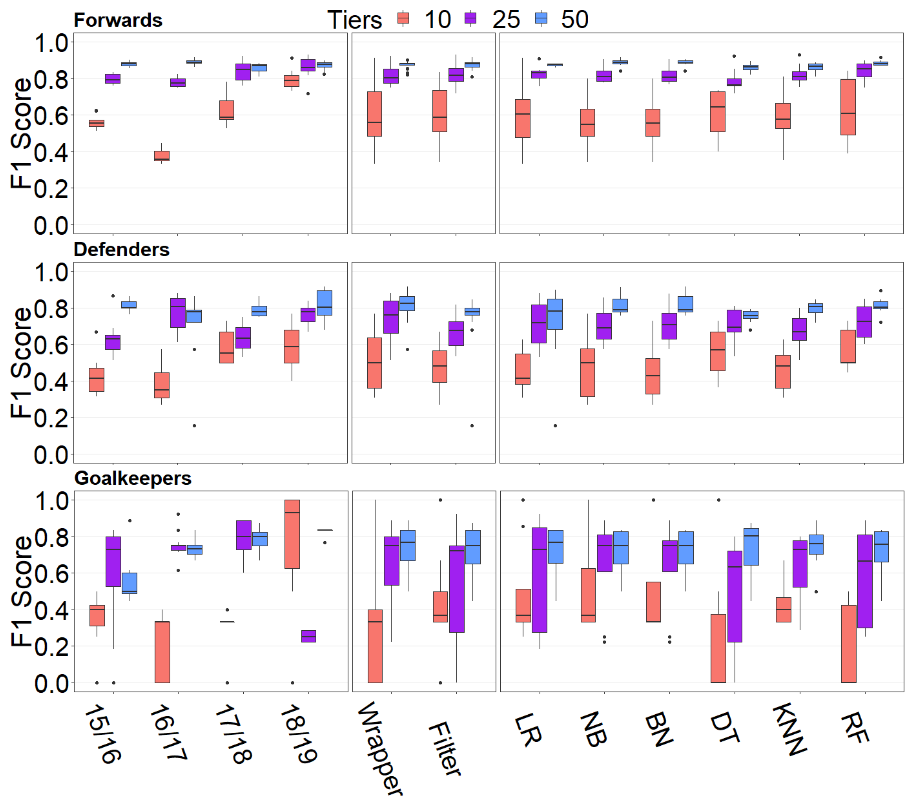

Fig. 2. F1 for different positions and tiers with respect to season, filter/wrapper and machine learning algorithm.

A closer look at the misclassified players explains why the above problems are so hard. For example, of the top 10% forwards of the 2018/19 season, 19 players were misclassified by at least one out of 12 (2×6) combined models and the best model (BN) misclassified 8 players with the filter method and 10 with the wrapper method. However, some of these players either had weaker than normal years and therefore may have been classified lower than they normally would have by some models (e.g., Taylor Hall 4/12 wrong, Gabriel Landeskog 2/12 wrong, Joe Pavelski 1/12 wrong, and Auston Matthews 1/12 wrong, Patrice Bergeron 1/12 wrong), was a Rookie (Elias Pettersson 1/12 wrong) or were players outside the top 10% tier that were classified into this top tier at least once. For the first set we note that the most frequent player that should be in the set but sometimes is classified outside is Taylor Hall. He is a former Hart Trophy (league MVP) winner (2017/18 season) that had an injury plagued 2018/19 were he only played 33 out of 82 games. Similarly, the misclassification of the two most frequently misclassified players of the last set can also be explained. Teuvo Teravainen is an upcoming star who ranked 29th in the scoring race when the 2019/20 season shut down for a covid-19 break, and Evgenii Dadonov had a career year (scoring

72 points 2018/19) playing on a line with Aleksander Barkov and Jonathan Huberdeau (which both finished with over 90 points). The lists of misclassified players for all data sets are given in Appendix F.

A limitation of the study is that for the algorithms with many parameters, we did not perform experiments to find the optimal parameter setting, but usually used the default values. An area for future work is, therefore, to experiment with optimal settings as well as other algorithms. Further, there are some choices in the experiments that may have an influence on the results. For instance, the choice of the number of occurrences in the feature selection step influences which attributes to retain and thus the data sets on which the machine learning algorithms are evaluated. It would be interesting to investigate these choices in a systematic way. Another track for future work is to use player performance methods for ranking instead of the EA player rating and to compare the results of the different methods.

6 Conclusion

In this paper we used 6 different machine learning methods (Logistic regression, k-Nearest neighbour, Decision tree, Random forest, Naïve Bayes and Bayesian network) and 2 different feature selection methods (filter and wrapper) to predict players' ranking tier (top 10%, 25% and 50%) for 3 player positions (forwards, defenders, and goalkeepers), looking at 4 seasons (2015/16 - 2018/19). The study highlights key performance metrics for the different player categories and provides insights into the difference in the complexity of identifying the key attributes and skills that may help predict the success of individual players.

When comparing the resulting models, the two Bayesian classifiers performed best and had the best sensitivity. The tree-based models had the highest specificity, but had trouble classifying the top 10% tier players. In general, the models were best at classifying forwards, highlighting that many of the official metrics are focused on the offensive measures. The development of equally good defensive metrics still remains an open problem.

References

1. Chawla, N.V., Bowyer, K.W., Hall, L.O., Kegelmeyer, W.P.: SMOTE: Synthetic Minority Over-sampling Technique. Journal of Artificial Intelligence Research **16**, 321–357 (2002), doi: 10.1613/jair.953
2. Dougherty, J., Kohavi, R., Sahami, M.: Supervised and unsupervised discretization of continuous features. In: Prieditis, A., Russell, S.J. (eds.) Machine Learning, Proceedings of the Twelfth International Conference on Machine Learning. pp. 194–202 (1995), doi: 10.1016/b978-1-55860-377-6.50032-3
3. Gramacy, R.B., Jensen, S.T., Taddy, M.: Estimating player contribution in hockey with regularized logistic regression. Journal of Quantitative Analysis in Sports **9**, 97–111 (2013), doi: 10.1515/jqas-2012-0001

4. Gu, W., Foster, K., Shang, J., Wei, L.: A game-predicting expert system using big data and machine learning. Expert Systems with Applications **130**, 293–305 (2019), doi: 10.1016/j.eswa.2019.04.025
5. Hall, M.: Correlation-based Feature Selection for Machine Learning. Ph.D. thesis, The University of Waikato, New Zealand (1999)
6. Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., Witten, I.H.: The WEKA Data Mining Software: An Update. SIGKDD Explorations **11**(1), 10–18 (2009), doi: 10.1145/1656274.1656278
7. Kaplan, E.H., Mongeon, K., Ryan, J.T.: A Markov Model for Hockey: Manpower Differential and Win Probability Added. INFOR Information Systems and Operational Research **52**(2), 39–50 (2014), doi: 10.3138/infor.52.2.39
8. Liu, G., Schulte, O.: Deep reinforcement learning in ice hockey for context-aware player evaluation. In: Lang, J. (ed.) Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence. pp. 3442–3448 (2018), doi: 10.24963/ijcai.2018/478
9. Liu, Y., Schulte, O., Li, C.: Model Trees for Identifying Exceptional Players in the NHL and NBA Drafts. In: Brefeld, U., Davis, J., Van Haaren, J., Zimmermann, A. (eds.) Machine Learning and Data Mining for Sports Analytics. MLSA 2018. Lecture Notes in Computer Science, vol. 11330, pp. 93–105 (2019), doi: 10.1007/978-3-030-17274-9_8
10. Ljung, D., Carlsson, N., Lambrix, P.: Player pairs valuation in ice hockey. In: Brefeld, U., Davis, J., Van Haaren, J., Zimmermann, A. (eds.) Machine Learning and Data Mining for Sports Analytics. MLSA 2018. Lecture Notes in Computer Science, vol. 11330, pp. 82–92 (2019), doi: 10.1007/978-3-030-17274-9_7
11. Macdonald, B.: A Regression-Based Adjusted Plus-Minus Statistic for NHL Players. Journal of Quantitative Analysis in Sports **7**(3) (2011), doi: 10.2202/1559-0410.1284
12. Macdonald, B.: An Improved Adjusted Plus-Minus Statistic for NHL Players. In: MIT Sloan Sports Analytics Conference (2011)
13. Nsolo, E., Carlsson, N., Lambrix, P.: Player valuation in European football. In: Brefeld, U., Davis, J., Van Haaren, J., Zimmermann, A. (eds.) Machine Learning and Data Mining for Sports Analytics. MLSA 2018. Lecture Notes in Computer Science, vol. 11330, pp. 42–54 (2019), doi: 10.1007/978-3-030-17274-9_4
14. Pettigrew, S.: Assessing the offensive productivity of NHL players using in-game win probabilities. In: MIT Sloan Sports Analytics Conference (2015)
15. Routley, K., Schulte, O.: A Markov Game Model for Valuing Player Actions in Ice Hockey. In: Meila, M., Heskes, T. (eds.) Uncertainty in Artificial Intelligence. pp. 782–791 (2015)
16. Sans Fuentes, C., Carlsson, N., Lambrix, P.: Player impact measures for scoring in ice hockey. In: Karlis, D., Ntzoufras, I., Drikos, S. (eds.) MathSport International 2019 Conference. pp. 307–317 (2019)
17. Schuckers, M.: Draft by Numbers: Using Data and Analytics to Improve National Hockey League (NHL) Player Selection. In: MIT Sloan Sports Analytics Conference (2016)
18. Schuckers, M., Curro, J.: Total Hockey Rating (THoR): A comprehensive statistical rating of National Hockey League forwards and defensemen based upon all on-ice events. In: MIT Sloan Sports Analytics Conference (2013)
19. Schulte, O., Khademi, M., Gholami, S., Zhao, Z., Javan, M., Desaulniers, P.: A Markov Game model for valuing actions, locations, and team performance in ice hockey. Data Mining and Knowledge Discovery **31**(6), 1735–1757 (2017), doi: 10.1007/s10618-017-0496-z

20. Schulte, O., Zhao, Z., Javan, M., Desaulniers, P.: Apples-to-apples: Clustering and Ranking NHL Players Using Location Information and Scoring Impact. In: MIT Sloan Sports Analytics Conference (2017)
21. Thomas, A., Ventura, S.L., Jensen, S., Ma, S.: Competing Process Hazard Function Models for Player Ratings in Ice Hockey. The Annals of Applied Statistics **7**(3), 1497–1524 (2013), doi: 10.1214/13-AOAS646

Appendix A: Classes for approximating the tiers

The table in this appendix shows the data sets that approximate the tiers. The desired percentage is 10, 25 or 50 which leads to a desired number of players in a tier. The actual percentage and number refers to the actual data set that is an approximation of the tier. These numbers are always lower than or equal to the desired percentage and number. Split refers to the lowest rating value in the tier.

Table 5: Classes for approximation of top 10%, 25% and 50% tiers of players.

Season	Position	Desired percentage	Desired number	Actual percentage	Actual number	Split
15/16	Defenders	10	29	6.7	20	89
	Defenders	25	74	20.2	60	86
	Defenders	50	148	44.8	133	83
	Forwards	10	58	8.9	52	88
	Forwards	25	145	23.7	138	85
	Forwards	50	291	49.3	287	81
	Goalkeepers	10	9	7.7	7	90
	Goalkeepers	25	22	23.1	21	86
	Goalkeepers	50	45	39.6	36	83
16/17	Defenders	10	28	7.3	21	86
	Defenders	25	71	23.7	68	83
	Defenders	50	143	45.7	131	80
	Forwards	10	57	8.7	50	85
	Forwards	25	143	20.3	116	83
	Forwards	50	286	44.1	252	79
	Goalkeepers	10	9	8.9	8	88
	Goalkeepers	25	22	23.3	21	85
	Goalkeepers	50	45	47.8	43	81
17/18	Defenders	10	29	8.1	24	86
	Defenders	25	74	24.2	72	83
	Defenders	50	148	46.8	139	80
	Forwards	10	55	9.2	51	86
	Forwards	25	138	24.3	135	83
	Forwards	50	277	41.1	230	80
	Goalkeepers	10	9	6.5	6	89
	Goalkeepers	25	23	22.6	21	84
	Goalkeepers	50	46	39.8	37	82
18/19	Defenders	10	30	9.3	28	86
	Defenders	25	75	19.5	59	84
	Defenders	50	151	44.7	135	80
	Forwards	10	54	8.6	47	87
	Forwards	25	136	25.0	136	83
	Forwards	50	272	43.7	238	80
	Goalkeepers	10	8	9.2	8	89
	Goalkeepers	25	21	21.8	19	86
	Goalkeepers	50	43	41.4	36	82

Appendix B: Selected attributes for the filter method

The following tables show the selected attributes using the filter method. Each table refers to a specific position. The data sets are for a specific tier and season for the position. The numbers in parentheses refer to the numbers of results for the different runs for different folds the attribute occurred in.

Table 6: Selected attributes for goalkeepers for the different tiers in different seasons by the filter method.

Data set	Attributes
$G10_{15/16}$	Team(2), GP(7), GS(2), QS(9), PIM(5)
$G25_{15/16}$	Team(9), W(2), QS(9)
$G50_{15/16}$	GP(3), W(4), GA(2), SA(2), SO(10), GPS(5), QS(9), PIM(2)
$G10_{16/17}$	W(10), SO(3), GPS(10), MIN(3), QS(5)
$G25_{16/17}$	W(9), OTL(4), GA(7), GPS(10), QS(4)
$G50_{16/17}$	Team(5), GP(3), GS(8), W(4), L(2), GA(2), SA(4), SO(8), GPS(4), QS(10), PIM(6)
$G10_{17/18}$	GS(5), W(10), QS(2)
$G25_{17/18}$	GP(6), W(10), L(7), SA(3), SO(6), GPS(9), QS(5)
$G50_{17/18}$	GP(9), W(10), L(6), GPS(10), MIN(2), QS(6)
$G10_{18/19}$	W(8), OTL(4), GA(2), SA(6), GPS(2), QS(6)
$G25_{18/19}$	GP(2), GS(2), W(5), SA(3), SO(6), GPS(10), MIN(9), QS(8), A(2)
$G50_{18/19}$	GP(10), GS(8), W(10), L(9), SA(3), SO(5), GPS(9), MIN(2), QS(10), A(5)

Table 7: Selected attributes for defenders for the different tiers in different seasons by the filter method.

Data set	Attributes
$D10_{15/16}$	Team(2), PIM(9), PS(10), EVA(2), PPA(3), SHA(9), S(10), TOI(3), TOI(EV)(5), TK(5), E+/-(2), SAtt.(2), SHFT(10), EVTOI(2)
$D25_{15/16}$	+/-(8), PS(6), PPG(3), EVA(5), PPA(6), SHA(9), S(3), TOI(10), CA(7), TOI(EV)(10), GV(9), E+/-(4), SAtt.(2), Thru%(8), EVTOI(10), GF/60(4), GA/60(2), PPTOI(10), SHTOI(10)
$D50_{15/16}$	GP(7), PIM(4), PS(10), PPG(8), GWG(9), EVA(9), PPA(8), BLK(10), HIT(2), CF(3), CA(7), FF(8), FA(4), oiSV%(9), oZS%(5), TOI(EV)(10), TK(6), GV(5), E+/-(2), SAtt.(10), Thru%(2), EVTOI(10), GF/60(2), PPTOI(6), SHTOI(10)
$D10_{16/17}$	PS(8), PPA(4), S(3), TOI(10), CA(6), TOI(EV)(7), GV(4), E+/-(3), SAtt.(4), EVTOI(3), PPTOI(9)
$D25_{16/17}$	+/-(5), PS(10), GWG(5), EVA(6), PPA(10), S(10), TOI(2), FF(3), FF%Rel(8), TOI(EV)(7), E+/-(10), SAtt.(8), Thru%(2), EVTOI(3), PPTOI(10)
$D50_{16/17}$	GP(5), +/-(8), PS(9), PPG(2), EVA(6), PPA(10), TOI(5), BLK(10), CF(3), CA(3), FF(7), FA(7), oiSH%(8), oZS%(4), TOI(6), GV(10), E+/-(4), EVTOI(9)
$D10_{17/18}$	PIM(5), PS(5), PPA(9), S(10), TOI(10), TOI(EV)(9), TK(2), SAtt.(2)
$D25_{17/18}$	+/-(8), PS(10), EVA(10), PPA(10), S(7), TOI(5), CF%Rel(5), FF%Rel(3), TOI(EV)(10), TK(5), E+/-(2), SAtt.(6), GF/60(2)
$D50_{17/18}$	+/-(6), PIM(3), PS(10), PPG(7), GWG(8), EVA(6), PPA(8), S(3), TOI(5), BLK(9), CF(4), CF%Rel(3), FF(10), oiSV%(7), oZS%(2), TOI(EV)(10), TK(4), E+/-(7), SAtt.(8)
$D10_{18/19}$	PS(10), EVG(3), PPG(4), GWG(2), EVA(4), PPA(6), SHA(5), S(3), HIT(3), CF%Rel(3), FF%Rel(7), oZS%(8), TOI(EV)(8), SAtt.(8), SHFT(2), GA/60(3), PPTOI(10)
$D25_{18/19}$	PS(10), EVG(2), PPG(8), EVA(6), PPA(7), S(2), FF(10), FF%Rel(8), TOI(EV)(9), E+/-(5), SAtt.(3), Thru%(7), SHFT(4), EVTOI(6), GF/60(2), PPTOI(10)
$D50_{18/19}$	+/-(2), PS(4), SHG(4), EVA(7), PPA(9), S(8), TOI(3), CF(8), FA(6), FF%Rel(3), oiSV%(6), TOI(EV)(10), TK(5), GV(2), E+/-(7), SAtt.(9), Thru%(7), EVTOI(10), EVCF%Rel(4), GF/60(2), PPTOI(5)

Table 8: Selected attributes for forwards for the different tiers in different seasons by the filter method.

Data set	Attributes
$F10_{15/16}$	+/-(5), PS(9), PPA(10), TOI(6), FF(6), TOI(EV)(10), E+/-(6), SHFT(6), EVTOI(10), PPTOI(9)
$F25_{15/16}$	Age(8), +/-(2), PS(10), PPA(9), FOW(9), CF%Rel(2), FF(6), FF%Rel(5), oZS%(9), TOI(EV)(8), E+/-(2), SAtt.(9), SHFT(6), EVTOI(10), PPTOI(10)
$F50_{15/16}$	Age(3), GP(7), PIM(10), PS(10), EVG(5), PPG(10), GWG(8), EVA(8), PPA(10), S(9), TOI(3), CF(8), FF(8), FA(2), oiSH%(2), oZS%(7), TOI(EV)(3), TK(8), E+/-(3), SAtt.(5), Thru%(10), EVTOI(7), GF/60(4), PPTOI(3)
$F10_{16/17}$	PS(4), PPG(3), GWG(5), EVA(3), PPA(3), S(4), TOI(4), FOL(7), CF(4), FF(5), TOI(EV)(7), GV(5), E+/-(2), SAtt.(10), SHFT(5), EVTOI(6), PPTOI(10)
$F25_{16/17}$	+/-(2), PS(10), EVG(2), PPG(5), GWG(2), PPA(8), TOI(9), FOW(3), CF(2), CF%Rel(4), FF(2), FF%Rel(5), oiSH%(2), oiSV%(9), oZS%(9), TOI(EV)(9), GV(8), SHFT(5), EVTOI(9), PPTOI(10)
$F50_{16/17}$	Age(9), PS(10), PPG(10), EVA(9), PPA(9), S(9), TOI(6), FOW(8), FOL(2), CF(3), CA(7), CF%Rel(2), FA(4), oiSH%(4), oiSV%(2), oZS%(8), TOI(EV)(10), TK(3), GV(8), E+/-(10), SAtt.(10), Thru%(2), SHFT(3), EVTOI(9), EVCF%Rel(7), GF/60(3), PPTOI(10)
$F10_{17/18}$	PS(10), PPG(3), EVA(2), PPA(9), S(6), FOW(5), CF%Rel(5), oZS%(2), TOI(EV)(10), GV(2), SAtt.(2), SHFT(6)
$F25_{17/18}$	PS(10), PPG(9), PPA(10), TOI(6), FOW(3), FOL(7), CF(8), FF(2), oiSV%(4), oZS%(3), TOI(EV)(10), E+/-(6), SAtt.(5), Thru%(2), SHFT(10)
$F50_{17/18}$	PS(10), PPG(10), EVA(6), PPA(10), TOI(2), BLK(2), CF(8), FF(2), FF%Rel(3), TOI(EV)(10), E+/-(6), SAtt.(3), Thru%(4), SHFT(10)
$F10_{18/19}$	PS(8), EVA(8), PPA(9), S(2), FOW(6), CA(9), FA(10), oZS%(2), TOI(EV)(8), GV(9), SHFT(5), EVTOI(10), PPTOI(10)
$F25_{18/19}$	PS(10), PPG(3), EVA(8), PPA(10), S(9), CF(2), CF%Rel(3), oiSV%(10), oZS%(10), TOI(EV)(10), E+/-(7), SHFT(7), EVTOI(10), EVCF%Rel(2), PPTOI(10)
$F50_{18/19}$	+/-(9), PS(10), EVG(7), PPG(10), EVA(3), PPA(10), S(8), TOI(7), FOL(3), CA(2), CF%Rel(8), FF(4), FA(2), FF%Rel(4), oiSV%(5), oZS%(3), TOI(EV)(10), E+/-(2), SAtt.(9), EVCF%Rel(2), GA/60(3), PPTOI(10)

Appendix C: Selected attributes for the wrapper method

The following tables show the selected attributes using the wrapper method. Each table refers to a specific position. The data sets are for a specific tier and season for the position.

Table 9: Selected attributes for goalkeepers for the different tiers in different seasons by the wrapper method.

Data set	Attributes
$G10_{15/16}$	Age, Team, GP, GS, SA, MIN, QS
$G25_{15/16}$	GP, W, L, SO, QS
$G50_{15/16}$	GS, OTL, SO, QS, PIM
$G10_{16/17}$	GS, W, OTL, SO, GPS
$G25_{16/17}$	W, OTL, GA, SO, GPS
$G50_{16/17}$	Age, W, SA, SO, GPS, PIM
$G10_{17/18}$	Age, W, SA
$G25_{17/18}$	W, L, SA, SO, GPS, QS, PIM
$G50_{17/18}$	Team, GS, W, SO, MIN, QS, RBS
$G10_{18/19}$	Age, W, SA, GPS, QS, PIM
$G25_{18/19}$	SO, GPS, QS
$G50_{18/19}$	GS, W, L, SO, QS, A

Table 10: Selected attributes for defenders for the different tiers in different seasons by the wrapper method.

Data set	Attributes
$D10_{15/16}$	Age, GP, PIM, PS, EVG, SHA, S, FOL, oZS%, SHFT, EVTOI, GA/60, SHTOI
$D25_{15/16}$	PPA, SHA, TOI, FOL, Thru%, GF/60, GA/60, SHTOI
$D50_{15/16}$	PS, GWG, EVA, SHA, TOI, FOW, oiSH%, TOI(EV), SAtt., EVTOI, PPTOI
$D10_{16/17}$	PS, SHG, PPA, FOW, FOL, oiSH%, Thru%, EVTOI, GF/60, GA/60, PPTOI
$D25_{16/17}$	+/-, PS, PPG, SHG, PPA, FF%Rel, oiSH%, oZS%, TOI(EV), SAtt., EVTOI, PPTOI
$D50_{16/17}$	Age, GP, PS, SHG, TOI, BLK, FOW, FF, oiSH%, E+/-, EVTOI, PPTOI, SHTOI
$D10_{17/18}$	Age, PPA, SHA, S, HIT, FF%Rel, TOI(EV), TK, GA/60
$D25_{17/18}$	PS, SHG, PPA, CF%Rel, oiSH%, oiSV%, TOI(EV), SHFT, GF/60
$D50_{17/18}$	+/-, PS, PPG, SHG, EVA, CF, FF, oZS%, dZS%, TOI(EV), SAtt., SHTOI
$D10_{18/19}$	PS, GWG, SHA, HIT, oiSH%, oiSV%, TOI(EV), GF/60, GA/60, PPTOI
$D25_{18/19}$	PS, SHG, PPA, FOL, CF%Rel, FF, oiSH%, oiSV%, oZS%, dZS%, TOI(EV), TK, GV, SHFT, EVTOI, GA/60, PPTOI, SHTOI
$D50_{18/19}$	+/-, SHG, PPA, FOW, FOL, oiSV%, TOI(EV), SAtt., EVTOI, GA/60, PPTOI

Table 11: Selected attributes for forwards for the different tiers in different seasons by the wrapper method.

Data set	Attributes
$F10_{15/16}$	PIM, PS, SHG, GWG, PPA, SHA, BLK, E+/-, EVTOI, PPTOI, SHTOI
$F25_{15/16}$	Age, +/-, PS, EVG, PPA, S, FOW, FF%Rel, oZS%, TOI(EV), SHFT, EVTOI, PPTOI
$F50_{15/16}$	Age, PS, EVG, PPG, GWG, EVA, PPA, S, HIT, FF, dZS%, TOI(EV), EVTOI, PPTOI
$F10_{16/17}$	PS, PPG, SHG, EVA, PPA, FOW, oiSH%, oiSV%, TOI(EV), TK, Thru%, SHFT, EVTOI, PPTOI
$F25_{16/17}$	PS, EVG, EVA, PPA, FOL, oiSH%, EVTOI, GF/60, GA/60, PPTOI, SHTOI
$F50_{16/17}$	Age, PS, SHG, EVA, oZS%, TOI(EV), SAtt., EVTOI, PPTOI
$F10_{17/18}$	GP, PS, PPG, PPA, SHA, FOW, TOI(EV), SAtt., SHFT
$F25_{17/18}$	PS, PPG, PPA, FOW, oZS%, TOI(EV), Thru%, SHFT
$F50_{17/18}$	Age, PS, PPG, GWG, PPA, FF%Rel, oiSH%, TOI(EV), Thru%, EVCF%Rel, SHTOI
$F10_{18/19}$	Age, PS, PPA, HIT, oiSH%, oZS%, GV, Thru%, EVTOI, GA/60, PPTOI
$F25_{18/19}$	PPA, SHA, FF%Rel, oiSV%, oZS%, dZS%, TOI(EV), SHFT, EVTOI, PPTOI
$F50_{18/19}$	PS, PPG, S, FOL, CF%Rel, oiSV%, TOI(EV), Thru%, EVTOI, EVCF%Rel, PPTOI

Appendix D: Performance of machine learning algorithms on the filter data sets

The following tables show the performance of logistic regression (LR), Naïve Bayes (NB), Bayesian network (BN), decision trees (DT), k-Nearest neighbour (KNN) and random forest (RF) regarding the measures accuracy, AUC, F1, specificity and sensitivity on the filter data sets.

Table 12: Accuracy of the algorithms for the filter data sets for different positions, tiers and seasons.

Data set	LR	NB	BN	DT	KNN	RF
$G10_{15/16}$	0.8333	0.8889	0.7778	0.8889	0.8333	0.8889
$G25_{15/16}$	0.5000	0.8333	0.8333	0.5000	0.7778	0.5000
$G50_{15/16}$	0.7222	0.7778	0.7778	0.7222	0.7778	0.7222
$G10_{16/17}$	0.7778	0.7778	0.7778	0.7778	0.7778	0.7222
$G25_{16/17}$	0.9444	0.7778	0.7778	0.7778	0.8333	0.7778
$G50_{16/17}$	0.8333	0.7222	0.7222	0.7222	0.7778	0.7778
$G10_{17/18}$	0.7778	0.7778	0.7778	0.7778	0.7778	0.8333
$G25_{17/18}$	0.8333	0.9444	0.9444	0.7778	0.8333	0.8889
$G50_{17/18}$	0.8333	0.7778	0.7778	0.8889	0.8333	0.8333
$G10_{18/19}$	1.0000	1.0000	1.0000	1.0000	0.8824	0.8824
$G25_{18/19}$	0.7059	0.6471	0.6471	0.5882	0.7059	0.6471
$G50_{18/19}$	0.8824	0.8824	0.8824	0.8824	0.8824	0.8824
$D10_{15/16}$	0.8475	0.7797	0.7797	0.8814	0.8814	0.8983
$D25_{15/16}$	0.7966	0.6949	0.6949	0.8305	0.6780	0.7288
$D50_{15/16}$	0.8644	0.8136	0.8136	0.7797	0.8136	0.8136
$D10_{16/17}$	0.8421	0.8070	0.8070	0.9123	0.8772	0.9123
$D25_{16/17}$	0.5965	0.7193	0.7193	0.7193	0.8246	0.8596
$D50_{16/17}$	0.6140	0.7895	0.7895	0.7719	0.7895	0.8070
$D10_{17/18}$	0.8644	0.8305	0.8305	0.9322	0.8983	0.9322
$D25_{17/18}$	0.7961	0.7797	0.7797	0.7627	0.8305	0.8136
$D50_{17/18}$	0.8136	0.7966	0.7966	0.7966	0.7797	0.8644
$D10_{18/19}$	0.7500	0.8833	0.8833	0.9333	0.8333	0.8667
$D25_{18/19}$	0.9167	0.8167	0.8333	0.9000	0.8500	0.8833
$D50_{18/19}$	0.7667	0.8333	0.8333	0.7167	0.8667	0.7833
$F10_{15/16}$	0.9052	0.8621	0.8621	0.8793	0.8966	0.8879
$F25_{15/16}$	0.9224	0.9052	0.9138	0.8966	0.8966	0.9224
$F50_{15/16}$	0.8534	0.8879	0.8879	0.8966	0.8534	0.8707
$F10_{16/17}$	0.8158	0.7982	0.7982	0.8584	0.8509	0.8509
$F25_{16/17}$	0.9123	0.9035	0.8947	0.8947	0.9211	0.8860
$F50_{16/17}$	0.8860	0.9211	0.9123	0.8772	0.8947	0.9211
$F10_{17/18}$	0.9099	0.8739	0.8739	0.9279	0.8829	0.9550
$F25_{17/18}$	0.9279	0.8919	0.8919	0.8739	0.9459	0.9459
$F50_{17/18}$	0.9099	0.8739	0.8739	0.9009	0.8559	0.9009
$F10_{18/19}$	0.9083	0.9083	0.9083	0.9266	0.9358	0.9450
$F25_{18/19}$	0.8991	0.9541	0.9541	0.8624	0.9633	0.9450
$F50_{18/19}$	0.8807	0.8991	0.9083	0.8716	0.8899	0.8991

Table 13: AUC of the algorithms for the filter data sets for different positions, tiers and seasons.

Data set	LR	NB	BN	DT	KNN	RF
$G10_{15/16}$	0.941	1.000	1.000	0.941	0.941	1.000
$G25_{15/16}$	0.417	0.778	0.792	0.396	0.681	0.417
$G50_{15/16}$	0.769	0.877	0.877	0.638	0.762	0.731
$G10_{16/17}$	1.000	0.882	0.824	0.794	0.912	0.765
$G25_{16/17}$	0.958	0.847	0.875	0.833	0.826	0.868
$G50_{16/17}$	0.760	0.896	0.896	0.721	0.779	0.831
$G10_{17/18}$	0.765	0.853	0.853	0.882	0.941	0.941
$G25_{17/18}$	1.000	0.982	0.982	0.839	0.813	0.964
$G50_{17/18}$	0.925	0.938	0.938	0.888	0.869	0.938
$G10_{18/19}$	1.000	1.000	1.000	1.000	0.786	1.000
$G25_{18/19}$	0.938	0.938	0.938	0.781	0.938	0.938
$G50_{18/19}$	0.909	0.924	0.924	0.826	0.894	0.894
$D10_{15/16}$	0.964	0.982	0.982	0.938	0.955	0.988
$D25_{15/16}$	0.893	0.881	0.888	0.856	0.824	0.865
$D50_{15/16}$	0.942	0.938	0.936	0.808	0.863	0.923
$D10_{16/17}$	0.938	0.907	0.914	0.796	0.775	0.836
$D25_{16/17}$	0.939	0.957	0.959	0.739	0.942	0.969
$D50_{16/17}$	0.788	0.920	0.919	0.767	0.867	0.908
$D10_{17/18}$	0.940	0.936	0.932	0.937	0.879	0.953
$D25_{17/18}$	0.973	0.973	0.980	0.757	0.929	0.933
$D50_{17/18}$	0.898	0.903	0.902	0.850	0.843	0.909
$D10_{18/19}$	0.978	0.960	0.967	0.955	0.927	0.975
$D25_{18/19}$	0.961	0.936	0.941	0.866	0.888	0.940
$D50_{18/19}$	0.920	0.941	0.943	0.684	0.924	0.916
$F10_{15/16}$	0.983	0.980	0.980	0.828	0.957	0.962
$F25_{15/16}$	0.973	0.965	0.966	0.773	0.893	0.954
$F50_{15/16}$	0.958	0.953	0.949	0.891	0.929	0.948
$F10_{16/17}$	0.992	0.962	0.965	0.907	0.940	0.958
$F25_{16/17}$	0.971	0.960	0.963	0.808	0.927	0.962
$F50_{16/17}$	0.958	0.974	0.971	0.878	0.943	0.965
$F10_{17/18}$	0.970	0.975	0.975	0.928	0.879	0.976
$F25_{17/18}$	0.984	0.960	0.963	0.914	0.936	0.974
$F50_{17/18}$	0.980	0.969	0.972	0.890	0.943	0.972
$F10_{18/19}$	0.979	0.983	0.984	0.855	0.942	0.982
$F25_{18/19}$	0.996	0.990	0.991	0.716	0.961	0.984
$F50_{18/19}$	0.967	0.970	0.971	0.820	0.916	0.955

Table 14: F1 of the algorithms for the filter data sets for different positions, tiers and seasons.

Data set	LR	NB	BN	DT	KNN	RF
$G10_{15/16}$	0.400	0.500	0.333	0.500	0.400	0.500
$G25_{15/16}$	0.182	0.727	0.727	0.000	0.600	0.303
$G50_{15/16}$	0.444	0.600	0.600	0.444	0.500	0.444
$G10_{16/17}$	0.333	0.333	0.333	0.000	0.333	0.000
$G25_{16/17}$	0.923	0.750	0.750	0.750	0.769	0.714
$G50_{16/17}$	0.769	0.667	0.667	0.667	0.714	0.714
$G10_{17/18}$	0.333	0.333	0.333	0.333	0.333	0.400
$G25_{17/18}$	0.727	0.889	0.889	0.600	0.727	0.800
$G50_{17/18}$	0.769	0.750	0.750	0.875	0.800	0.800
$G10_{18/19}$	1.000	1.000	1.000	1.000	0.667	0.500
$G25_{18/19}$	0.286	0.250	0.250	0.222	0.286	0.250
$G50_{18/19}$	0.833	0.833	0.833	0.833	0.833	0.833
$D10_{15/16}$	0.400	0.316	0.316	0.462	0.462	0.500
$D25_{15/16}$	0.667	0.571	0.571	0.688	0.558	0.600
$D50_{15/16}$	0.846	0.800	0.800	0.764	0.800	0.800
$D10_{16/17}$	0.308	0.267	0.267	0.444	0.364	0.444
$D25_{16/17}$	0.610	0.692	0.692	0.692	0.783	0.818
$D50_{16/17}$	0.154	0.778	0.778	0.723	0.778	0.784
$D10_{17/18}$	0.556	0.500	0.500	0.667	0.625	0.667
$D25_{17/18}$	0.600	0.581	0.581	0.533	0.643	0.621
$D50_{17/18}$	0.776	0.778	0.778	0.750	0.764	0.846
$D10_{18/19}$	0.400	0.588	0.588	0.667	0.500	0.500
$D25_{18/19}$	0.800	0.686	0.722	0.800	0.727	0.759
$D50_{18/19}$	0.720	0.773	0.773	0.679	0.818	0.723
$F10_{15/16}$	0.621	0.556	0.556	0.533	0.625	0.552
$F25_{15/16}$	0.836	0.807	0.821	0.760	0.800	0.836
$F50_{15/16}$	0.860	0.885	0.885	0.895	0.855	0.874
$F10_{16/17}$	0.364	0.343	0.343	0.444	0.414	0.414
$F25_{16/17}$	0.815	0.814	0.793	0.769	0.824	0.755
$F50_{16/17}$	0.879	0.916	0.906	0.863	0.887	0.916
$F10_{17/18}$	0.667	0.588	0.588	0.714	0.581	0.783
$F25_{17/18}$	0.846	0.786	0.786	0.759	0.880	0.880
$F50_{17/18}$	0.881	0.841	0.841	0.867	0.810	0.871
$F10_{18/19}$	0.737	0.762	0.762	0.733	0.811	0.833
$F25_{18/19}$	0.841	0.906	0.906	0.717	0.929	0.885
$F50_{18/19}$	0.860	0.889	0.898	0.863	0.880	0.887

Table 15: Specificity of the algorithms for the filter data sets for different positions, tiers and seasons.

Data set	LR	NB	BN	DT	KNN	RF
$G10_{15/16}$	0.824	0.882	0.765	0.882	0.824	0.882
$G25_{15/16}$	0.667	0.917	0.917	0.750	0.917	0.583
$G50_{15/16}$	0.846	0.846	0.846	0.846	0.923	0.846
$G10_{16/17}$	0.765	0.765	0.765	0.824	0.765	0.765
$G25_{16/17}$	0.917	0.667	0.667	0.667	0.833	0.750
$G50_{16/17}$	0.909	0.727	0.727	0.727	0.818	0.818
$G10_{17/18}$	0.765	0.765	0.765	0.765	0.765	0.824
$G25_{17/18}$	0.786	0.929	0.929	0.786	0.786	0.857
$G50_{17/18}$	1.000	0.800	0.800	0.900	0.900	0.900
$G10_{18/19}$	1.000	1.000	1.000	1.000	0.929	1.000
$G25_{18/19}$	0.688	0.625	0.625	0.563	0.688	0.625
$G50_{18/19}$	0.909	0.909	0.909	0.909	0.909	0.909
$D10_{15/16}$	0.839	0.768	0.768	0.875	0.875	0.893
$D25_{15/16}$	0.761	0.630	0.630	0.826	0.609	0.674
$D50_{15/16}$	0.829	0.743	0.743	0.714	0.743	0.743
$D10_{16/17}$	0.852	0.815	0.815	0.926	0.889	0.926
$D25_{16/17}$	0.410	0.590	0.590	0.590	0.744	0.795
$D50_{16/17}$	1.000	0.727	0.727	0.818	0.727	0.788
$D10_{17/18}$	0.868	0.830	0.830	0.962	0.906	0.962
$D25_{17/18}$	0.760	0.740	0.740	0.740	0.800	0.780
$D50_{17/18}$	0.906	0.813	0.813	0.906	0.781	0.906
$D10_{18/19}$	0.727	0.873	0.873	0.945	0.818	0.873
$D25_{18/19}$	0.978	0.804	0.804	0.913	0.848	0.913
$D50_{18/19}$	0.700	0.825	0.825	0.625	0.850	0.750
$F10_{15/16}$	0.906	0.849	0.849	0.887	0.887	0.896
$F25_{15/16}$	0.944	0.921	0.933	0.955	0.889	0.944
$F50_{15/16}$	0.797	0.898	0.898	0.898	0.831	0.831
$F10_{16/17}$	0.806	0.787	0.787	0.861	0.843	0.843
$F25_{16/17}$	0.921	0.888	0.888	0.921	0.944	0.910
$F50_{16/17}$	0.857	0.889	0.889	0.889	0.873	0.889
$F10_{17/18}$	0.910	0.870	0.870	0.930	0.890	0.970
$F25_{17/18}$	0.931	0.885	0.885	0.862	0.954	0.954
$F50_{17/18}$	0.889	0.833	0.833	0.889	0.847	0.875
$F10_{18/19}$	0.914	0.892	0.892	0.968	0.935	0.946
$F25_{18/19}$	0.863	1.000	1.000	0.938	0.988	1.000
$F50_{18/19}$	0.949	0.915	0.932	0.864	0.898	0.932

Table 16: Sensitivity of the algorithms for the filter data sets for different positions, tiers and seasons.

Data set	LR	NB	BN	DT	KNN	RF
$G10_{15/16}$	1.000	1.000	1.000	1.000	1.000	1.000
$G25_{15/16}$	0.167	0.667	0.667	0.000	0.500	0.333
$G50_{15/16}$	0.400	0.600	0.600	0.400	0.400	0.400
$G10_{16/17}$	1.000	1.000	1.000	0.000	1.000	0.000
$G25_{16/17}$	1.000	1.000	1.000	1.000	0.833	0.833
$G50_{16/17}$	0.714	0.714	0.714	0.714	0.714	0.714
$G10_{17/18}$	1.000	1.000	1.000	1.000	1.000	1.000
$G25_{17/18}$	1.000	1.000	1.000	0.750	1.000	1.000
$G50_{17/18}$	0.625	0.750	0.750	0.875	0.750	0.750
$G10_{18/19}$	1.000	1.000	1.000	1.000	0.667	0.333
$G25_{18/19}$	1.000	1.000	1.000	1.000	1.000	1.000
$G50_{18/19}$	0.833	0.833	0.833	0.833	0.833	0.833
$D10_{15/16}$	1.000	1.000	1.000	1.000	1.000	1.000
$D25_{15/16}$	0.923	0.923	0.923	0.846	0.923	0.923
$D50_{15/16}$	0.917	0.917	0.917	0.875	0.917	0.917
$D10_{16/17}$	0.667	0.667	0.667	0.667	0.667	0.667
$D25_{16/17}$	1.000	1.000	1.000	1.000	1.000	1.000
$D50_{16/17}$	0.083	0.875	0.875	0.708	0.875	0.833
$D10_{17/18}$	0.833	0.833	0.833	0.667	0.833	0.667
$D25_{17/18}$	1.000	1.000	1.000	0.889	1.000	1.000
$D50_{17/18}$	0.704	0.778	0.778	0.667	0.778	0.815
$D10_{18/19}$	1.000	1.000	1.000	0.800	1.000	0.800
$D25_{18/19}$	0.714	0.857	0.929	0.857	0.857	0.786
$D50_{18/19}$	0.900	0.850	0.850	0.900	0.900	0.850
$F10_{15/16}$	0.900	1.000	1.000	0.800	1.000	0.800
$F25_{15/16}$	0.852	0.852	0.852	0.704	0.889	0.852
$F50_{15/16}$	0.912	0.877	0.877	0.895	0.877	0.912
$F10_{16/17}$	1.000	1.000	1.000	1.000	1.000	1.000
$F25_{16/17}$	0.880	0.960	0.920	0.800	0.840	0.800
$F50_{16/17}$	0.922	0.961	0.941	0.863	0.922	0.961
$F10_{17/18}$	0.909	0.909	0.909	0.909	0.818	0.818
$F25_{17/18}$	0.917	0.917	0.917	0.917	0.917	0.917
$F50_{17/18}$	0.949	0.949	0.949	0.923	0.872	0.949
$F10_{18/19}$	0.875	1.000	1.000	0.688	0.938	0.938
$F25_{18/19}$	1.000	0.828	0.828	0.655	0.897	0.793
$F50_{18/19}$	0.800	0.880	0.880	0.880	0.880	0.860

Appendix E: Performance of machine learning algorithms on the wrapper data sets

The following tables show the performance of logistic regression (LR), Naïve Bayes (NB), Bayesian network (BN), decision trees (DT), k-Nearest neighbour (KNN) and random forest (RF) regarding the measures accuracy, AUC, F1, specificity and sensitivity on the wrapper data sets.

Table 17: Accuracy of the algorithms for the wrapper data sets for different positions, tiers and seasons.

Data set	LR	NB	BN	DT	KNN	RF
$G10_{15/16}$	0.6667	0.8333	0.8333	0.8333	0.7778	0.8333
$G25_{15/16}$	0.8333	0.8889	0.8333	0.8889	0.8889	0.8889
$G50_{15/16}$	0.7222	0.7778	0.7778	0.8333	0.9444	0.7778
$G10_{16/17}$	0.8333	0.7778	0.7778	0.9444	0.8333	0.7778
$G25_{16/17}$	0.8889	0.7778	0.7778	0.7778	0.8333	0.7222
$G50_{16/17}$	0.8889	0.7778	0.7778	0.8333	0.7778	0.7778
$G10_{17/18}$	0.7778	0.7778	0.7778	0.8333	0.8333	0.7778
$G25_{17/18}$	0.9444	0.8888	0.8888	0.8333	0.8889	0.9444
$G50_{17/18}$	0.6667	0.8333	0.8333	0.8889	0.7222	0.8333
$G10_{18/19}$	0.9412	1.0000	1.0000	0.8235	0.8824	0.8235
$G25_{18/19}$	0.6471	0.5882	0.5882	0.5882	0.7059	0.7059
$G50_{18/19}$	0.8824	0.8824	0.8824	0.8824	0.8235	0.8824
$D10_{15/16}$	0.8644	0.7797	0.8136	0.9492	0.8136	0.8983
$D25_{15/16}$	0.7966	0.8136	0.8136	0.7627	0.7119	0.8136
$D50_{15/16}$	0.7966	0.8644	0.8814	0.8305	0.8305	0.8475
$D10_{16/17}$	0.7895	0.9474	0.9298	0.8772	0.8421	0.9649
$D25_{16/17}$	0.9298	0.8947	0.9123	0.8772	0.8596	0.8947
$D50_{16/17}$	0.7368	0.8772	0.8772	0.7895	0.7368	0.8246
$D10_{17/18}$	0.8305	0.8305	0.8305	0.9492	0.8475	0.9492
$D25_{17/18}$	0.7288	0.8983	0.8983	0.8814	0.8475	0.8644
$D50_{17/18}$	0.8814	0.7797	0.7797	0.7966	0.8644	0.8305
$D10_{18/19}$	0.9000	0.9500	0.9500	0.8500	0.8833	0.9333
$D25_{18/19}$	0.9000	0.9167	0.9167	0.9000	0.8167	0.9000
$D50_{18/19}$	0.8167	0.8000	0.8000	0.7833	0.7833	0.7833
$F10_{15/16}$	0.8534	0.8534	0.8621	0.8966	0.8707	0.8707
$F25_{15/16}$	0.8966	0.8966	0.8966	0.8966	0.8879	0.9224
$F50_{15/16}$	0.8793	0.9052	0.9052	0.8621	0.8793	0.8793
$F10_{16/17}$	0.7895	0.8070	0.8070	0.8421	0.8070	0.8333
$F25_{16/17}$	0.8596	0.8860	0.8772	0.8860	0.8684	0.8772
$F50_{16/17}$	0.8860	0.8947	0.8947	0.8947	0.8947	0.9035
$F10_{17/18}$	0.8739	0.8378	0.8378	0.9459	0.8739	0.9369
$F25_{17/18}$	0.9189	0.9279	0.9279	0.9189	0.9001	0.9550
$F50_{17/18}$	0.9099	0.9099	0.9099	0.8559	0.8919	0.9099
$F10_{18/19}$	0.9725	0.9266	0.9266	0.9083	0.9174	0.9450
$F25_{18/19}$	0.9450	0.9266	0.9266	0.9358	0.9174	0.9266
$F50_{18/19}$	0.8899	0.9083	0.8899	0.8349	0.8532	0.8899

Table 18: AUC of the algorithms for the wrapper data sets for different positions, tiers and seasons.

Data set	LR	NB	BN	DT	KNN	RF
$G10_{15/16}$	1.000	1.000	1.000	0.353	0.912	0.882
$G25_{15/16}$	0.819	0.972	0.972	0.833	0.986	0.979
$G50_{15/16}$	0.877	0.800	0.785	0.654	0.838	0.646
$G10_{16/17}$	1.000	0.765	0.765	0.882	0.941	0.765
$G25_{16/17}$	0.972	0.861	0.861	0.826	0.889	0.813
$G50_{16/17}$	0.948	0.922	0.922	0.916	0.864	0.909
$G10_{17/18}$	0.765	0.824	0.824	0.441	0.824	0.824
$G25_{17/18}$	0.954	0.982	0.982	0.929	0.911	0.929
$G50_{17/18}$	0.794	0.900	0.888	0.888	0.725	0.900
$G10_{18/19}$	0.976	1.000	1.000	1.000	0.798	1.000
$G25_{18/19}$	0.938	0.938	0.938	0.781	0.938	0.969
$G50_{18/19}$	0.909	0.924	0.924	0.826	0.879	0.955
$D10_{15/16}$	0.929	0.994	0.988	0.973	0.896	0.991
$D25_{15/16}$	0.866	0.883	0.875	0.768	0.833	0.906
$D50_{15/16}$	0.920	0.927	0.925	0.839	0.888	0.929
$D10_{16/17}$	0.926	0.914	0.914	0.781	0.775	0.870
$D25_{16/17}$	0.967	0.980	0.984	0.919	0.937	0.980
$D50_{16/17}$	0.865	0.936	0.932	0.841	0.879	0.908
$D10_{17/18}$	0.943	0.918	0.915	0.956	0.884	0.939
$D25_{17/18}$	0.971	0.956	0.962	0.940	0.940	0.938
$D50_{17/18}$	0.914	0.897	0.898	0.839	0.917	0.916
$D10_{18/19}$	1.000	0.985	0.985	0.884	0.945	0.982
$D25_{18/19}$	0.935	0.967	0.966	0.866	0.838	0.895
$D50_{18/19}$	0.899	0.913	0.915	0.788	0.844	0.895
$F10_{15/16}$	0.981	0.981	0.983	0.840	0.939	0.961
$F25_{15/16}$	0.960	0.969	0.969	0.779	0.914	0.960
$F50_{15/16}$	0.965	0.958	0.958	0.844	0.954	0.938
$F10_{16/17}$	0.977	0.975	0.977	0.904	0.935	0.972
$F25_{16/17}$	0.967	0.951	0.949	0.917	0.933	0.949
$F50_{16/17}$	0.960	0.968	0.969	0.905	0.944	0.959
$F10_{17/18}$	0.978	0.967	0.967	0.841	0.911	0.963
$F25_{17/18}$	0.980	0.975	0.977	0.908	0.940	0.977
$F50_{17/18}$	0.986	0.975	0.975	0.857	0.965	0.979
$F10_{18/19}$	0.983	0.978	0.978	0.830	0.956	0.981
$F25_{18/19}$	0.995	0.981	0.981	0.895	0.927	0.962
$F50_{18/19}$	0.962	0.974	0.973	0.859	0.902	0.963

Table 19: F1 of the algorithms for the wrapper data sets for different positions, tiers and seasons.

Data set	LR	NB	BN	DT	KNN	RF
$G10_{15/16}$	0.250	0.400	0.400	0.000	0.333	0.000
$G25_{15/16}$	0.727	0.833	0.769	0.800	0.800	0.833
$G50_{15/16}$	0.615	0.500	0.500	0.571	0.889	0.500
$G10_{16/17}$	0.400	0.333	0.333	?	0.400	0.000
$G25_{16/17}$	0.833	0.750	0.750	0.714	0.727	0.615
$G50_{16/17}$	0.833	0.750	0.750	0.769	0.750	0.714
$G10_{17/18}$	0.333	0.333	0.333	0.000	0.400	0.000
$G25_{17/18}$	0.889	0.800	0.800	0.667	0.800	0.889
$G50_{17/18}$	0.667	0.824	0.824	0.875	0.667	0.824
$G10_{18/19}$	0.857	1.000	1.000	?	0.667	?
$G25_{18/19}$	0.250	0.222	0.222	0.222	0.286	0.286
$G50_{18/19}$	0.833	0.833	0.833	0.833	0.769	0.833
$D10_{15/16}$	0.429	0.316	0.353	0.667	0.353	0.500
$D25_{15/16}$	0.866	0.645	0.645	0.611	0.514	0.645
$D50_{15/16}$	0.786	0.846	0.863	0.792	0.815	0.830
$D10_{16/17}$	0.333	0.571	0.333	0.364	0.308	0.500
$D25_{16/17}$	0.882	0.857	0.878	0.811	0.800	0.850
$D50_{16/17}$	0.571	0.863	0.863	0.778	0.717	0.808
$D10_{17/18}$	0.545	0.500	0.500	0.727	0.526	0.727
$D25_{17/18}$	0.529	0.750	0.750	0.696	0.667	0.692
$D50_{17/18}$	0.863	0.755	0.755	0.750	0.846	0.800
$D10_{18/19}$	0.625	0.769	0.727	0.471	0.588	0.714
$D25_{18/19}$	0.769	0.839	0.839	0.786	0.667	0.800
$D50_{18/19}$	0.899	0.913	0.915	0.788	0.844	0.895
$F10_{15/16}$	0.514	0.541	0.556	0.571	0.571	0.516
$F25_{15/16}$	0.760	0.778	0.778	0.760	0.787	0.830
$F50_{15/16}$	0.881	0.903	0.903	0.857	0.879	0.879
$F10_{16/17}$	0.333	0.353	0.353	0.400	0.353	0.387
$F25_{16/17}$	0.758	0.787	0.767	0.780	0.754	0.750
$F50_{16/17}$	0.876	0.885	0.885	0.887	0.885	0.897
$F10_{17/18}$	0.588	0.526	0.526	0.727	0.563	0.667
$F25_{17/18}$	0.824	0.852	0.852	0.924	0.792	0.898
$F50_{17/18}$	0.878	0.874	0.884	0.822	0.854	0.881
$F10_{18/19}$	0.914	0.800	0.800	0.737	0.780	0.842
$F25_{18/19}$	0.893	0.840	0.840	0.868	0.816	0.846
$F50_{18/19}$	0.875	0.898	0.878	0.824	0.830	0.875

Table 20: Specificity of the algorithms for the wrapper data sets for different positions, tiers and seasons.

Data set	LR	NB	BN	DT	KNN	RF
$G10_{15/16}$	0.647	0.824	0.824	0.882	0.765	0.882
$G25_{15/16}$	0.917	0.917	0.833	1.000	1.000	0.917
$G50_{15/16}$	0.692	0.923	0.923	1.000	1.000	0.923
$G10_{16/17}$	0.824	0.765	0.765	1.000	0.824	0.824
$G25_{16/17}$	0.917	0.667	0.667	0.750	0.917	0.750
$G50_{16/17}$	1.000	0.727	0.727	0.909	0.727	0.818
$G10_{17/18}$	0.765	0.765	0.765	0.882	0.824	0.824
$G25_{17/18}$	0.929	0.857	0.857	0.857	0.857	0.929
$G50_{17/18}$	0.600	0.800	0.800	0.900	0.800	0.800
$G10_{18/19}$	0.929	1.000	1.000	1.000	0.929	1.000
$G25_{18/19}$	0.625	0.563	0.563	0.563	0.688	0.688
$G50_{18/19}$	0.909	0.909	0.909	0.909	0.818	0.909
$D10_{15/16}$	0.857	0.768	0.804	0.946	0.804	0.893
$D25_{15/16}$	0.783	0.826	0.826	0.739	0.717	0.826
$D50_{15/16}$	0.714	0.829	0.857	0.857	0.771	0.800
$D10_{16/17}$	0.778	0.963	0.963	0.889	0.852	1.000
$D25_{16/17}$	0.974	0.846	0.872	0.897	0.846	0.872
$D50_{16/17}$	0.970	0.848	0.848	0.727	0.697	0.788
$D10_{17/18}$	0.680	0.830	0.830	0.981	0.849	0.981
$D25_{17/18}$	0.680	0.880	0.880	0.880	0.820	0.840
$D50_{17/18}$	0.938	0.813	0.813	0.906	0.906	0.906
$D10_{18/19}$	0.891	0.945	0.964	0.850	0.883	0.933
$D25_{18/19}$	0.957	0.913	0.913	0.935	0.817	0.913
$D50_{18/19}$	0.800	0.775	0.775	0.750	0.750	0.750
$F10_{15/16}$	0.849	0.840	0.849	0.906	0.858	0.877
$F25_{15/16}$	0.955	0.933	0.933	0.955	0.888	0.955
$F50_{15/16}$	0.847	0.915	0.915	0.881	0.864	0.864
$F10_{16/17}$	0.778	0.796	0.796	0.833	0.796	0.824
$F25_{16/17}$	0.820	0.865	0.865	0.876	0.854	0.888
$F50_{16/17}$	0.873	0.889	0.889	0.873	0.889	0.873
$F10_{17/18}$	0.870	0.830	0.830	0.970	0.880	0.970
$F25_{17/18}$	0.931	0.920	0.920	0.931	0.908	0.966
$F50_{17/18}$	0.903	0.861	0.875	0.806	0.889	0.889
$F10_{18/19}$	0.968	0.914	0.914	0.914	0.903	0.935
$F25_{18/19}$	0.975	1.000	1.000	0.988	1.000	0.988
$F50_{18/19}$	0.932	0.932	0.915	0.831	0.915	0.932

Table 21: Sensitivity of the algorithms for the wrapper data sets for different positions, tiers and seasons.

Data set	LR	NB	BN	DT	KNN	RF
$G10_{15/16}$	1.000	1.000	1.000	0.000	1.000	0.000
$G25_{15/16}$	0.667	0.833	0.833	0.667	0.667	0.833
$G50_{15/16}$	0.800	0.400	0.400	0.400	0.800	0.400
$G10_{16/17}$	1.000	1.000	1.000	0.000	1.000	0.000
$G25_{16/17}$	0.833	1.000	1.000	0.833	0.667	0.667
$G50_{16/17}$	0.714	0.857	0.857	0.714	0.857	0.714
$G10_{17/18}$	1.000	1.000	1.000	0.000	1.000	0.000
$G25_{17/18}$	1.000	1.000	1.000	0.750	1.000	1.000
$G50_{17/18}$	0.750	0.875	0.875	0.875	0.625	0.875
$G10_{18/19}$	1.000	1.000	1.000	?	0.667	0.000
$G25_{18/19}$	1.000	1.000	1.000	1.000	1.000	1.000
$G50_{18/19}$	0.833	0.833	0.833	0.833	0.833	0.833
$D10_{15/16}$	1.000	1.000	1.000	1.000	1.000	1.000
$D25_{15/16}$	0.846	0.769	0.769	0.846	0.692	0.769
$D50_{15/16}$	0.917	0.917	0.917	0.792	0.917	0.917
$D10_{16/17}$	1.000	0.667	0.333	0.667	0.667	0.333
$D25_{16/17}$	0.833	1.000	1.000	0.833	0.889	0.944
$D50_{16/17}$	0.417	0.917	0.917	0.875	0.792	0.875
$D10_{17/18}$	1.000	0.833	0.833	0.667	0.833	0.667
$D25_{17/18}$	1.000	1.000	1.000	1.000	1.000	0.889
$D50_{17/18}$	0.815	0.741	0.741	0.667	0.815	0.741
$D10_{18/19}$	1.000	1.000	0.800	0.800	1.000	1.000
$D25_{18/19}$	0.714	0.929	0.929	0.786	0.786	0.857
$D50_{18/19}$	0.850	0.850	0.850	0.850	0.850	0.850
$F10_{15/16}$	0.900	1.000	1.000	0.800	1.000	0.800
$F25_{15/16}$	0.704	0.778	0.778	0.704	0.889	0.815
$F50_{15/16}$	0.912	0.895	0.895	0.842	0.895	0.895
$F10_{16/17}$	1.000	1.000	1.000	1.000	1.000	1.000
$F25_{16/17}$	1.000	0.960	0.920	0.920	0.920	0.840
$F50_{16/17}$	0.902	0.902	0.902	0.922	0.902	0.941
$F10_{17/18}$	0.909	0.909	0.909	0.727	0.818	0.636
$F25_{17/18}$	0.875	0.958	0.958	0.875	0.875	0.917
$F50_{17/18}$	0.923	0.974	0.974	0.949	0.897	0.949
$F10_{18/19}$	1.000	1.000	1.000	0.875	1.000	1.000
$F25_{18/19}$	0.862	0.724	0.724	0.793	0.690	0.759
$F50_{18/19}$	0.840	0.880	0.860	0.840	0.780	0.840

Appendix F: Wrongly classified players

In this section we present the wrongly classified players from the 20% test data. Each table refers to a specific tier for a specific position in a specific season. The column *C* refers to the correct class. When the value is 1 it means that the player belongs to the tier, while a value 0 means the player does not belong to the tier. The columns *Correct* and *Wrong* refer to the number of machine learning algorithms that identified the correct class correctly and wrongly, respectively. The *BN* columns mark whether the best algorithm over all data (Bayesian Networks) classified the player wrongly.

Table 22: Wrongly classified top 10% goalkeepers in the 2015/16 season.

		Filter			Wrapper		
Player	C	Correct	Wrong	BN	Correct	Wrong	BN
Niklas Bäckström	0	5	1		5	1	
Devan Dubnyk	0	0	6	*	1	5	*
Jhonas Enroth	0	6	0		5	1	
Chad Johnson	0	6	0		5	1	
Anton Khudobin	0	6	0		5	1	
Keith Kinkaid	0	5	1	*	2	4	
Darcy Kuemper	0	5	1	*	6	0	
Henrik Lundqvist	1	6	0		4	2	
Al Montoya	0	5	1		6	0	
Karri Rämö	0	6	0		5	1	
Cam Talbot	0	6	0		3	3	*
Semyon Varlamov	0	0	6	*	3	3	*
				4			3

Table 23: Wrongly classified top 25% goalkeepers in the 2015/16 season.

		Filter			Wrapper		
Player	C	Correct	Wrong	BN	Correct	Wrong	BN
Sergei Bobrovsky	1	0	6	*	0	6	*
Peter Budaj	0	3	3		6	0	
Mike Condon	0	4	2		4	2	*
Brian Elliot	0	6	0		3	3	*
Jhonas Enroth	0	3	3		6	0	
Jonas Gustavsson	0	0	6	*	6	0	
Braden Holtby	1	4	2		6	0	
Martin Jones	1	4	2		6	0	
Pekka Rinne	1	2	4		6	0	
Cory Schneider	1	4	2		6	0	
Semyon Varlamaov	1	0	6	*	3	3	
Scott Wedgewood	0	5	1		6	0	
				3			3

Table 24: Wrongly classified top 50% goalkeepers in the 2015/16 season.

		Filter			Wrapper		
Player	C	Correct	Wrong	BN	Correct	Wrong	BN
Sergei Bobrovsky	1	2	4		2	4	*
Scott Darling	0	4	2	*	6	0	
Andrew Hammond	0	6	0		5	1	
Carter Hutton	0	3	3		5	1	
Eddie Läck	0	0	6	*	2	4	*
Jacob Markström	1	0	6	*	2	4	*
Matt Murray	1	0	6	*	0	6	*
Joni Ortio	0	6	0		5	1	
				4			4

Table 25: Wrongly classified top 10% goalkeepers in the 2016/17 season.

		Filter			Wrapper		
Player	C	Correct	Wrong	BN	Correct	Wrong	BN
Jake Allen	0	1	5	*	2	4	*
Craig Anderson	0	0	6	*	1	5	*
John Gibson	0	0	6	*	3	3	*
Thomas Greiss	0	0	6	*	1	5	*
Carey Price	1	4	2		4	2	
				4			4

Table 26: Wrongly classified top 25% goalkeepers in the 2016/17 season.

		Filter			Wrapper		
Player	C	Correct	Wrong	BN	Correct	Wrong	BN
Jake Allen	0	0	6	*	0	6	*
Craig Anderson	1	6	0		3	3	
Jared Coreau	0	6	0		4	2	
Connor Hellebuyck	0	3	3	*	4	2	*
Steve Mason	1	4	2		3	3	
James Reimer	0	1	5	*	4	2	*
Mike Smith	0	2	4	*	2	4	*
				4			4

Table 27: Wrongly classified top 50% goalkeepers in the 2016/17 season.

		Filter			Wrapper		
Player	C	Correct	Wrong	BN	Correct	Wrong	BN
Aaron Dell	0	3	3	*	3	3	*
Jimmy Howard	1	6	0		4	2	
Michael Hutchinson	1	0	6	*	4	2	
Carter Hutton	0	1	5	*	2	4	*
Chad Johnson	0	1	5	*	0	6	*
Darcy Kuemper	1	0	6	*	3	3	*
Anthony Stolarz	0	5	1		6	0	
Cam Ward	1	6	0		5	1	
				5			4

Table 28: Wrongly classified top 10% goalkeepers in the 2017/18 season.

		Filter			Wrapper		
Player	C	Correct	Wrong	BN	Correct	Wrong	BN
Cam Talbot	0	0	6	*	0	6	*
Devan Dubnyk	0	0	6	*	0	6	*
Frederik Andersen	0	0	6	*	1	5	*
Connor Hellebuyck	0	1	5	*	3	3	*
Braden Holtby	1	6	0		4	2	
				4			4

Table 29: Wrongly classified top 25% goalkeepers in the 2017/18 season.

		Filter			Wrapper		
Player	C	Correct	Wrong	BN	Correct	Wrong	BN
Scott Darling	0	3	3		6	0	
Jonathan Bernier	0	0	6	*	0	6	*
Carter Hutton	0	3	3		2	4	*
Curtis McElhinney	0	5	1		6	0	
Braden Holtby	1	5	1		5	1	
				1			2

Table 30: Wrongly classified top 50% goalkeepers in the 2017/18 season.

		Filter			Wrapper		
Player	C	Correct	Wrong	BN	Correct	Wrong	BN
Ryan Miller	1	0	6	*	0	6	*
Anton Khudobin	1	2	4	*	6	0	
Scott Darling	0	4	2	*	5	1	*
Cam Ward	0	1	5	*	0	6	*
Corey Crawford	1	5	1		4	2	
Carey Price	1	5	1		5	1	
Mike Condon	0	6	0		5	1	
Michal Neuvirth	0	6	0		5	1	
				4			3

Table 31: Wrongly classified top 10% goalkeepers in the 2018/19 season.

		Filter			Wrapper		
Player	C	Correct	Wrong	BN	Correct	Wrong	BN
Sergei Bobrovsky	1	4	2		4	2	
Devan Dubnyk	1	6	0		4	2	
Robin Lehner	0	5	1		4	2	
Andrei Vasilevskiy	1	5	1		4	2	
				0			0

Table 32: Wrongly classified top 25% goalkeepers in the 2018/19 season.

		Filter			Wrapper		
Player	C	Correct	Wrong	BN	Correct	Wrong	BN
Craig Anderson	0	0	6	*	2	4	*
Jack Campbell	0	5	1		1	5	*
Jimmy Howard	0	1	5	*	0	6	*
Anton Khudobin	0	0	6	*	0	6	*
Darcy Kuemper	0	0	6	*	0	6	*
Petr Mrazek	0	0	6	*	1	5	*
Linus Ullmark	0	1	5	*	1	5	*
				6			7

Table 33: Wrongly classified top 50% goalkeepers in the 2018/19 season.

		Filter			Wrapper		
Player	C	Correct	Wrong	BN	Correct	Wrong	BN
Jonathan Bernier	0	5	1		4	2	
Anders Nilsson	0	1	5	*	1	5	*
Antti Raanta	1	0	6	*	0	6	*
				2			2

Table 34: Wrongly classified top 10% defenders in the 2015/16 season.

		Filter			Wrapper		
Player	C	Correct	Wrong	BN	Correct	Wrong	BN
Karl Alzner	0	2	4	*	3	3	*
Jay Bouwmeester	0	6	0		5	1	
Johnny Boychuk	0	1	5	*	0	6	*
Cody Ceci	0	3	3	*	3	3	*
Ryan Ellis	0	1	5	*	1	5	*
Niklas Kronwall	0	5	1		6	0	
Dmitry Kulikov	0	4	2	*	6	0	
Adam Larsson	0	0	6	*	3	3	
Josh Manson	0	5	1		3	3	*
Marc Methot	0	6	0		4	2	*
John Moore	0	4	2	*	5	1	
Jake Muzzin	0	0	6	*	0	6	*
Darnell Nurse	0	2	4	*	3	3	*
Andrej Sekera	0	2	4	*	2	4	*
Anton Strålman	0	0	6	*	0	6	*
Mark Streit	0	4	2	*	5	1	
Jacob Trouba	0	2	4	*	2	4	*
				13			11

Table 35: Wrongly classified top 25% defenders in the 2015/16 season.

		Filter			Wrapper		
Player	C	Correct	Wrong	BN	Correct	Wrong	BN
Francois Beauchemin	0	0	6	*	0	6	*
Jordie Benn	0	5	1		5	1	
Jonas Brodin	1	0	6	*	0	6	*
Jason Demers	0	1	5	*	4	2	
Brenden Dillon	0	2	4	*	5	1	
Brian Dumoulin	0	1	5	*	6	0	
Alexander Edler	1	5	1		3	3	*
Aaron Ekblad	1	6	0		1	5	*
Mattias Ekholm	0	1	5	*	2	4	*
Alexei Emelin	0	2	4	*	5	1	
Justin Faulk	1	6	0		4	2	
Mark Fayne	0	5	1		4	2	
Shayne Gostisbehere	1	6	0		5	1	
Andy Greene	0	0	6	*	1	5	*
Nicklas Grossmann	0	5	1		5	1	
Adam Larsson	0	0	6	*	0	6	*
John-Michael Liles	0	2	4	*	6	0	
Andrei Markov	0	0	6	*	0	6	*
Kevan Miller	0	0	6	*	3	3	
Connor Murphy	0	0	6	*	3	3	
Brett Pesce	0	4	2	*	5	1	
David Schlemko	0	3	3	*	2	4	*
Nate Schmidt	0	2	4	*	5	1	
Michael Stone	0	1	5	*	0	6	*
Keith Yandle	0	0	6	*	0	6	*
				18			11

Table 36: Wrongly classified top 50% defenders in the 2015/16 season.

		Filter			Wrapper		
Player	C	Correct	Wrong	BN	Correct	Wrong	BN
Zach Bogosian	1	5	1		6	0	
Ben Chiarot	0	2	4	*	5	1	
Dylan DeMelo	0	4	2		6	0	
Simon Despres	1	1	5	*	0	6	*
Alexander Edler	1	6	0		5	1	
Joel Edmundson	0	3	3	*	5	1	
Mark Fayne	0	1	5	*	4	2	
Matt Greene	1	0	6	*	0	6	*
Nicklas Grossmann	0	0	6	*	4	2	
Ron Hainsey	0	0	6	*	0	6	*
Jyrki Jokipakka	0	1	5	*	0	6	*
Steven Kampfer	0	6	0		4	2	
Darnell Nurse	0	2	4	*	1	5	*
Alexander Petrovic	0	0	6	*	0	6	*
Jeff Petry	1	5	1		5	1	
Michal Rozsival	0	3	3	*	4	2	
Jaccob Slavin	0	0	6	*	0	6	*
Brian Strait	0	4	2		6	0	
Dennis Wideman	1	6	0		5	1	
				12			7

Table 37: Wrongly classified top 10% defenders in the 2016/17 season.

		Filter			Wrapper		
Player	C	Correct	Wrong	BN	Correct	Wrong	BN
Karl Alzner	0	6	0		5	1	
Francois Beauchemin	0	4	2	*	6	0	
Zdeno Chara	0	3	3	*	5	1	
Mattias Ekholm	0	3	3	*	4	2	
Justin Faulk	0	0	6	*	1	5	*
Niklas Hjalmarsson	1	0	6	*	1	5	*
Torey Krug	0	0	6	*	3	3	
Esa Lindell	0	4	2	*	5	1	
Dion Phaneuf	0	0	6	*	3	3	
Kyle Quincey	0	6	0		5	1	
Brady Skjei	0	5	1		3	3	
Anton Strålman	0	1	5	*	3	3	
Jacob Trouba	0	2	4	*	4	2	
Shea Weber	1	6	0		4	2	*
Zach Werenski	0	2	4	*	1	5	*
				11			4

Table 38: Wrongly classified top 25% defenders in the 2016/17 season.

		Filter			Wrapper		
Player	C	Correct	Wrong	BN	Correct	Wrong	BN
Karl Alzner	1	6	0		3	3	
Jay Bouwmeester	1	6	0		4	2	
Justin Braun	1	6	0		2	4	
Connor Carrick	0	1	5	*	5	1	
Fredrik Claesson	0	5	1		6	0	
Adam Clendening	0	4	2		6	0	
Calvin de Haan	0	0	6	*	2	4	*
Anthony DeAngelo	0	2	4	*	4	2	*
Dylan DeMelo	0	3	3		6	0	
Michael Del Zotto	0	1	5	*	5	1	
Brenden Dillon	0	5	1	*	6	0	
Jason Garrison	0	3	3	*	6	0	
Andy Greene	0	1	5	*	1	5	*
Erik Gudbranson	0	5	1		6	0	
Matt Irwin	0	3	3	*	6	0	
Nick Jensen	0	5	1		6	0	
Paul Martin	0	0	6	*	6	0	
John Moore	0	1	5	*	1	5	*
Darnell Nurse	0	4	2		6	0	
Alexander Petrovic	0	5	1		6	0	
Mark Pysyk	0	2	4	*	6	0	
Luca Sbisa	0	2	4	*	6	0	
David Schlemko	0	0	6	*	2	4	
Nate Schmidt	0	3	3	*	6	0	
Shea Theodore	0	2	4		6	0	
Trevor van Riemsdyk	0	0	6	*	2	4	*
Nikita Zadorov	0	3	3	*	5	1	
				16			5

Table 39: Wrongly classified top 50% defenders in the 2016/17 season.

		Filter			Wrapper		
Player	C	Correct	Wrong	BN	Correct	Wrong	BN
Nathan Beaulieu	1	5	1		5	1	
Matt Benning	0	1	5	*	2	4	*
Johnny Boychuk	1	5	1		5	1	
Brandon Carlo	1	5	1		6	0	
John Carlson	1	5	1		6	0	
Ben Chiarot	0	3	3	*	4	2	
Trevor Daley	1	5	1		5	1	
Jonathan Ericsson	0	1	5	*	4	2	
Mark Giordano	1	5	1		6	0	
Dan Girardi	1	5	1		5	1	
Shayne Gostisbehere	1	5	1		6	0	
Ron Hainsey	0	1	5	*	1	5	*
Ben Hutton	1	3	3		5	1	
Nick Jensen	0	3	3	*	3	3	
Roman Josi	1	5	1		6	0	
Seth Jones	1	5	1		6	0	
John Klingberg	1	5	1		6	0	
Niklas Kronwall	1	2	4		2	4	
Kris Letang	1	5	1		5	1	
Jonathon Merrill	1	3	3	*	5	1	
Josh Morrissey	1	5	1		5	1	
Ryan Murray	1	4	2		4	2	
Darnell Nurse	1	1	5	*	0	6	*
Johnny Oduya	0	2	4	*	2	4	*
Alexander Petrovic	0	2	4	*	3	3	
Kyle Quincey	0	1	5	*	1	5	*
Kris Russell	0	1	5	*	0	6	*
Andrej Sustr	1	4	2		5	1	
Patrick Wiercioch	0	5	1		4	2	
Tyler Myers	1	0	6	*	0	6	*
Nikita Zadorov	1	3	3		4	2	
				12			7

Table 40: Wrongly classified top 10% defenders in the 2017/18 season.

		Filter			Wrapper		
Player	C	Correct	Wrong	BN	Correct	Wrong	BN
T.J. Brodie	0	0	6	*	2	4	*
Erik Johnson	0	2	4	*	2	4	*
Mike Matheson	0	2	4	*	2	4	*
Charlie McAvoy	0	3	3	*	5	1	
Jake Muzzin	0	3	3	*	3	3	*
Jordan Oesterle	0	6	0		5	1	
Dmitry Orlov	0	2	4	*	2	4	*
Nate Schmidt	0	3	3	*	2	4	*
Shea Theodore	0	4	2	*	2	4	*
Jacob Trouba	0	6	0		3	3	*
Sami Vatanen	0	1	5	*	0	6	*
Marc-Edouard Vlasic	1	4	2		4	2	
Shea Weber	1	0	6	*	1	5	*
				10			10

Table 41: Wrongly classified top 25% defenders in the 2017/18 season.

		Filter			Wrapper		
Player	C	Correct	Wrong	BN	Correct	Wrong	BN
Karl Alzner	0	3	3	*	5	1	
Kyle Capobianco	0	6	0		5	1	
Brandon Carlo	0	2	4	*	5	1	
Calvin de Haan	0	5	1		5	1	
Michael Del Zotto	0	0	6	*	0	6	*
Toby Enström	0	5	1		6	0	
Derek Forbort	0	0	6	*	0	6	*
Thomas Hickey	0	0	6	*	2	4	*
Nick Leddy	1	5	1		6	0	
Andrew MacDonald	0	0	6	*	2	4	
Ryan Murphy	0	4	2		5	1	
Darnell Nurse	0	0	6	*	0	6	*
Brooks Orpik	0	6	0		5	1	
Mark Pysyk	0	1	5	*	5	1	
Kris Russell	0	0	6	*	5	1	
Joakim Ryan	0	4	2	*	4	2	
Brady Skjei	1	6	0		5	1	
Sami Vatanen	0	0	6	*	0	6	*
Nikita Zadorov	0	0	6	*	0	6	*
Nikita Zaitsev	0	0	6	*	3	3	
				13			6

Table 42: Wrongly classified top 50% defenders in the 2017/18 season.

		Filter			Wrapper		
Player	C	Correct	Wrong	BN	Correct	Wrong	BN
Sebastian Aho	0	5	1		6	0	
Ethan Bear	0	5	1		5	1	
Jordie Benn	0	0	6	*	0	6	*
Robert Bortuzzo	0	3	3	*	4	2	*
Brandon Carlo	1	5	1		6	0	
Connor Carrick	1	0	6	*	1	5	*
Ian Cole	1	5	1		4	2	
Dylan DeMelo	0	3	3	*	5	1	
Jake Dotchin	0	5	1		3	3	*
Toby Enstrom	1	0	6	*	1	5	*
Andy Greene	0	0	6	*	2	4	*
Radko Gudas	0	2	4	*	2	4	*
Ben Hutton	1	3	3		0	6	*
Robert Hägg	0	3	3	*	4	2	*
Jack Johnson	1	6	0		5	1	
Adam McQuaid	1	0	6	*	1	5	*
Marc Methot	1	1	5	*	1	5	*
Andrej Sustr	1	0	6	*	0	6	*
Chris Tanev	1	2	4	*	2	4	*
Nikita Zaitsev	1	4	2		5	1	
				12			13

Table 43: Wrongly classified top 10% defenders in the 2018/19 season.

		Filter			Wrapper		
Player	C	Correct	Wrong	BN	Correct	Wrong	BN
Cody Ceci	0	5	1		5	1	
Dennis Cholowski	0	4	2		6	0	
Brian Dumoulin	0	4	2		4	2	
Anders Englund	0	5	1		6	0	
Samuel Girard	0	0	6	*	4	2	
Alex Goligoski	0	1	5	*	5	1	
Mike Green	0	0	6	*	3	3	
Henri Jokiharju	0	5	1		6	0	
Hampus Lindholm	0	1	5	*	3	3	
Darnell Nurse	0	1	5	*	0	6	*
Adam Pelech	0	6	0		5	1	
Jeff Petry	0	0	6	*	0	6	*
Justin Schultz	0	5	1		5	1	
Damon Severson	0	1	5	*	2	4	
Devon Toews	0	5	1		6	0	
Marc-Edouard Vlasic	0	4	2		6	0	
Shea Weber	1	4	2		4	2	*
				7			3

Table 44: Wrongly classified top 25% defenders in the 2018/19 season.

		Filter			Wrapper		
Player	C	Correct	Wrong	BN	Correct	Wrong	BN
Jonas Brodin	0	3	3	*	5	1	
Thomas Chabot	1	6	0		5	1	
Samuel Girard	0	2	4	*	2	4	*
Miro Heiskanen	0	0	6	*	2	4	*
Niklas Hjalmarsson	1	0	6	*	0	6	*
Filip Hronek	0	4	2	*	6	0	
Ryan McDonagh	1	5	1		5	1	
Colin Miller	0	1	5	*	0	6	*
Ryan Murray	1	4	2		4	2	
Brett Pesce	0	4	2	*	5	1	
Neal Pionk	0	3	3	*	2	4	*
Kevin Shattenkirk	0	1	5	*	3	3	
Troy Stetcher	0	3	3	*	5	1	
Sami Vatanen	0	5	1		5	1	
Marc-Edouard Vlasic	1	5	1		4	2	
				10			5

Table 45: Wrongly classified top 50% defenders in the 2018/19 season.

		Filter			Wrapper		
Player	C	Correct	Wrong	BN	Correct	Wrong	BN
Robert Bortuzzo	0	3	3		6	0	
Ben Chiarot	0	1	5	*	0	6	*
Connor Clifton	0	3	3		3	3	
Ian Cole	1	4	2		2	4	*
Kevin Connauton	0	4	2		5	1	
Carl Dahlström	0	4	2		2	4	*
Radko Gudas	0	1	5	*	2	4	
Ron Hainsey	0	0	6	*	0	6	*
Caleb Jones	0	3	3		1	5	*
Brett Kulak	0	0	6	*	0	6	*
Dmitry Kulikov	0	5	1		6	0	
Scott Mayfield	0	0	6	*	0	6	*
Mirco Mueller	0	5	1		6	0	
Patrik Nemeth	1	0	6	*	0	6	*
Jamie Oleksiak	1	1	5	*	0	6	*
Xavier Ouellet	0	5	1		6	0	
Mike Reilly	0	0	6	*	0	6	*
Jimmy Schuldt	0	3	3		0	6	*
Justin Schultz	1	4	2	*	6	0	
Marc Staal	0	2	4	*	3	3	*
Chris Tanev	1	6	0		4	2	
				10			12

Table 46: Wrongly classified top 10% forwards in the 2015/16 season.

		Filter			Wrapper		
Player	C	Correct	Wrong	BN	Correct	Wrong	BN
Max Domi	0	2	4	*	1	5	*
Brandon Dubinsky	0	2	4	*	0	6	*
Alex Galchenyuk	0	3	3	*	1	5	*
Patric Hörnqvist	0	0	6	*	0	6	*
Jaromir Jagr	0	0	6	*	1	5	*
Mikko Koivu	0	0	6	*	0	6	*
Anders Lee	0	6	0		4	2	*
Bryan Little	0	0	6	*	0	6	*
Brad Marchand	0	2	4	*	2	4	*
Ondrej Palat	0	2	4	*	1	5	*
Kyle Palmieri	0	0	6	*	0	6	*
Victor Rask	0	1	5	*	0	6	*
Sam Reinhart	0	5	1		4	2	
Brayden Schenn	0	0	6	*	0	6	*
Jordan Staal	0	2	4	*	0	6	*
Derek Stepan	1	3	3		3	3	
Jonathan Toews	1	4	2		4	2	
James van Riemsdyk	0	0	6	*	3	3	
Kris Versteeg	0	3	3	*	3	3	*
Mika Zibanejad	0	3	3	*	1	5	*
				16			16

Table 47: Wrongly classified top 25% forwards in the 2015/16 season.

		Filter			Wrapper		
Player	C	Correct	Wrong	BN	Correct	Wrong	BN
Ryan Callahan	0	5	1		5	1	
Sean Couturier	1	3	3	*	2	4	*
Shane Doan	0	0	6	*	0	6	*
Michael Frolik	1	0	6	*	0	6	*
Alex Galchenyuk	1	6	0		5	1	
Brendan Gallagher	1	5	1		4	2	
Evander Kane	1	5	1		6	0	
Ryan Kesler	1	5	1		5	1	
Alex Killorn	0	1	5	*	1	5	*
Chris Kunitz	0	0	6	*	0	6	*
Mark Letestu	0	6	0		5	1	
Brock Nelson	1	1	5	*	0	6	*
Ondrej Palat	1	5	1		5	1	
P.A. Parenteau	0	1	5	*	2	4	*
David Perron	0	4	2		5	1	
Mike Ribeiro	0	0	6	*	0	6	*
Brandon Saad	1	6	0		5	1	
Brandon Sutter	1	0	6	*	0	6	*
Alex Tanguay	0	5	1		5	1	
Tomas Tatar	1	4	2		1	5	*
Kris Versteeg	0	2	4	*	3	3	*
Justin Williams	1	5	1		6	0	
Travis Zajac	1	6	0		3	3	*
				10			12

Table 48: Wrongly classified top 50% forwards in the 2015/16 season.

		Filter			Wrapper		
Player	C	Correct	Wrong	BN	Correct	Wrong	BN
Oliver Björkstrand	0	5	1		6	0	
Joseph Blandisi	0	4	2		4	2	*
Lance Bouma	1	0	6	*	0	6	*
J.T. Brown	0	1	5	*	1	5	*
Adam Cracknell	0	6	0		2	4	
Phillip Danault	0	5	1		6	0	
David Desharnais	1	6	0		5	1	
Derek Dorsett	0	3	3		4	2	
Turner Elson	0	5	1		6	0	
Micheal Ferland	0	1	5	*	3	3	
Mike Fisher	1	6	0		5	1	
Sam Gagner	1	5	1		5	1	
Matt Hendricks	1	0	6	*	2	4	*
Mattias Janmark	0	0	6	*	0	6	*
Jacob Josefson	0	3	3		4	2	
Joffrey Lupul	1	3	3	*	5	1	
Clarke MacArthur	1	0	6	*	0	6	*
Jonathan Marchessault	0	4	2		0	6	*
Jared McCann	0	2	4	*	5	1	
Jay McClement	0	4	2		6	0	
Cody McLeod	0	4	2		4	2	
Chris Neil	0	5	1		5	1	
Cal O'Reilly	0	6	0		5	1	
Sam Reinhart	0	0	6	*	0	6	*
Colton Sceviour	0	1	5	*	3	3	
Jaden Schwartz	1	4	2		6	0	
Jack Skille	0	5	1		5	1	
Nick Spaling	1	1	5	*	1	5	*
Chris Thorburn	1	2	4	*	1	5	*
Viktor Tikhonov	0	6	0		5	1	
Jordin Tootoo	1	2	4	*	0	6	*
				13			11

Table 49: Wrongly classified top 10% forwards in the 2016/17 season.

		Filter			Wrapper		
Player	C	Correct	Wrong	BN	Correct	Wrong	BN
Sven Baertschi	0	6	0		5	1	
Connor Brown	0	4	2	*	6	0	
Jonathan Drouin	0	0	6	*	0	6	*
Sam Gagner	0	3	3	*	2	4	*
Brendan Gallagher	0	5	1		6	0	
Jake Guentzel	0	6	0		5	1	
Mike Hoffman	0	0	6	*	0	6	*
Jaromir Jagr	0	0	6	*	0	6	*
Marcus Johansson	0	0	6	*	0	6	*
Anders Lee	0	2	4	*	0	6	*
Mark Letestu	0	5	1		5	1	
Bryan Little	0	2	4	*	0	6	*
Patrick Marleau	0	1	5	*	0	6	*
Mitch Marner	0	0	6	*	0	6	*
J.T. Miller	0	3	3	*	2	4	*
Ondrej Palat	0	0	6	*	0	6	*
P.A. Parenteau	0	6	0		5	1	
Mathieu Perreault	0	1	5	*	1	5	*
Brayden Point	0	0	6	*	0	6	*
Rickard Rakell	0	0	6	*	1	5	*
Victor Rask	0	1	5	*	0	6	*
Bobby Ryan	0	6	0		4	2	
Conor Sheary	0	3	3	*	3	3	*
Jakob Silfverberg	0	0	6	*	1	5	*
Jeff Skinner	0	0	6	*	0	6	*
Jason Spezza	0	1	5	*	0	6	*
Alex Steen	0	1	5	*	1	5	*
Dylan Strome	0	6	0		5	1	
Joe Thornton	0	2	4	*	1	5	*
Mats Zuccarello	0	0	6	*	0	6	*
				23			22

Table 50: Wrongly classified top 25% forwards in the 2016/17 season.

		Filter			Wrapper		
Player	C	Correct	Wrong	BN	Correct	Wrong	BN
Sebastian Aho	0	0	6	*	0	6	*
David Backes	0	2	4	*	0	6	*
Josh Bailey	0	0	6	*	0	6	*
Tyler Bozak	0	0	6	*	0	6	*
Matt Duchene	1	4	2		6	0	
Brian Gionta	0	6	0		4	2	*
Markus Granlund	0	4	2	*	2	4	*
Joshua Ho-Sang	0	6	0		4	2	
Evander Kane	1	5	1		6	0	
Alex Killorn	0	3	3	*	0	6	*
Leo Komarov	0	6	0		4	2	
Dylan Larkin	1	0	6	*	2	4	*
J.T. Miller	1	1	5	*	5	1	
Frans Nielsen	1	4	2		4	2	
Brayden Schenn	1	6	0		5	1	*
Andrew Shaw	0	6	0		4	2	
C.J. Smith	0	6	0		5	1	
Reilly Smith	0	2	4	*	0	6	*
Paul Stastny	0	0	6	*	0	6	*
Alex Steen	1	5	1		6	0	
Brandon Sutter	0	2	4	*	0	6	*
Matthew Tkachuk	0	6	0		0	6	*
Tyler Toffoli	1	3	3		3	3	
Radim Vrbata	0	0	6	*	0	6	*
Jason Zucker	0	6	0		5	1	
				12			14

Table 51: Wrongly classified top 50% forwards in the 2016/17 season.

		Filter			Wrapper		
Player	C	Correct	Wrong	BN	Correct	Wrong	BN
Oliver Björkstrand	0	5	1		1	5	*
Jay Beagle	1	5	1		6	0	
Brian Boyle	1	6	0		5	1	
Dustin Brown	0	0	6	*	0	6	*
Drake Caggiula	0	1	5	*	4	2	*
Casey Cizikas	1	3	3	*	2	4	*
Kyle Clifford	0	5	1		6	0	
Cal Clutterbuck	0	0	6	*	1	5	*
J.T. Compher	0	6	0		5	1	
Matt Cullen	0	1	5	*	0	6	*
Jake Guentzel	1	5	1		6	0	
Chris Kelly	0	4	2	*	6	0	
Nikolai Kulemin	1	4	2		3	3	*
Artturi Lehkonen	0	0	6	*	0	6	*
Trevor Lewis	1	6	0		5	1	
Shawn Matthias	0	6	0		5	1	
Benoit Pouliot	0	1	5	*	2	4	
Michael Raffl	1	1	5	*	1	5	*
Devin Setoguchi	0	5	1		5	1	
Riley Sheahan	1	5	1		6	0	
Drew Stafford	1	6	0		5	1	
Joe Thornton	1	5	1		6	0	
Frank Vatrano	0	2	4		0	6	*
Antoine Vermette	1	5	1		6	0	
Jordan Weal	0	3	3		4	2	
Tom Wilson	1	5	1		0	6	*
Tommy Wingels	1	0	6	*	0	6	*
				10			12

Table 52: Wrongly classified top 10% forwards in the 2017/18 season.

		Filter			Wrapper		
Player	C	Correct	Wrong	BN	Correct	Wrong	BN
Josh Anderson	0	6	0		5	1	
Cam Atkinson	0	4	2	*	2	4	*
Jeff Carter	1	0	6	*	0	6	*
Sean Couturier	1	6	0		4	2	
Christian Dvorak	0	6	0		5	1	
Jordan Eberle	0	3	3	*	4	2	*
Sam Gagner	0	5	1		5	1	
Alex Galchenyuk	0	1	5	*	3	3	*
Yanni Gourde	0	6	0		4	2	*
Erik Haula	0	2	4	*	1	5	*
Kevin Hayes	0	6	0		3	3	*
Tomas Hertl	0	3	3	*	2	4	*
Patric Hörnqvist	0	2	4	*	2	4	*
Evander Kane	0	2	4	*	2	4	*
David Krejci	0	1	5	*	3	3	*
Bryan Little	0	6	0		5	1	
Auston Matthews	1	6	0		5	1	
J.T. Miller	0	2	4	*	3	3	*
Sam Reinhart	0	1	5	*	2	4	*
Jaden Schwartz	1	4	2		3	3	
Tomas Tatar	0	6	0		4	2	*
Alex Tuch	0	5	1		5	1	
James van Riemsdyk	0	1	5	*	0	6	*
Justin Williams	0	5	1		3	3	*
Mika Zibanejad	0	2	4	*	2	4	*
Jason Zucker	0	1	5	*	2	4	*
				14			18

Table 53: Wrongly classified top 25% forwards in the 2017/18 season.

		Filter			Wrapper		
Player	C	Correct	Wrong	BN	Correct	Wrong	BN
Jesper Bratt	0	5	1		6	0	
Alex DeBrincat	1	5	1		5	1	
Jake Debrusk	0	4	2		4	2	
Jonathan Drouin	1	6	0		5	1	
Micheal Ferland	0	4	2	*	2	4	*
Nick Foligno	0	2	4	*	2	4	*
Sam Gagner	0	2	4	*	3	3	*
Erik Haula	0	0	6	*	0	6	*
Zach Hyman	0	2	4	*	5	1	
Mattias Janmark	0	4	2	*	6	0	
Chris Kreider	1	6	0		5	1	
Ondrej Palat	1	1	5	*	2	4	
Zach Parise	1	0	6	*	0	6	*
Mathieu Perreault	0	2	4	*	3	3	*
Victor Rask	0	1	5	*	2	4	*
Devin Shore	0	5	1		6	0	
Colton Sissons	0	5	1		5	1	
Craig Smith	0	0	6	*	0	6	*
Ryan Spooner	0	5	1		4	2	
Carl Söderberg	0	4	2	*	6	0	
Colin Wilson	0	5	1		6	0	
Travis Zajac	0	5	1		6	0	
				12			8

Table 54: Wrongly classified top 50% forwards in the 2017/18 season.

		Filter			Wrapper		
Player	C	Correct	Wrong	BN	Correct	Wrong	BN
Sam Bennett	0	3	3	*	5	1	
Tyler Bertuzzi	0	2	4	*	3	3	*
Connor Brown	1	4	2		5	1	
Blake Coleman	0	1	5	*	2	4	*
J.T. Compher	0	2	4	*	2	4	*
Phillip Danault	1	6	0		5	1	
Ryan Donato	0	6	0		5	1	
Marcus Foligno	0	5	1		6	0	
Michael Frolik	1	3	3		2	4	
Ryan Hartman	0	0	6	*	2	4	
Vinnie Hinostroza	0	0	6	*	0	6	*
Marcus Johansson	1	0	6	*	1	5	*
Milan Lucic	1	5	1		5	1	
Jared McCann	0	0	6	*	1	5	*
Sonny Milano	0	5	1		3	3	
Zach Parise	1	3	3	*	5	1	
Tyler Pitlick	0	2	4	*	1	5	*
Tom Pyatt	0	6	0		5	1	
Tobias Rieder	0	2	4	*	4	2	
Evan Rodrigues	0	1	5	*	0	6	*
Derek Ryan	0	0	6	*	0	6	*
Nikita Scherbak	0	6	0		5	1	
Andrew Shaw	0	3	3	*	1	5	*
Conor Sheary	1	5	1		6	0	
Drew Stafford	0	6	0		5	1	
				14			10

Table 55: Wrongly classified top 10% forwards in the 2018/19 season.

		Filter			Wrapper		
Player	C	Correct	Wrong	BN	Correct	Wrong	BN
Patrice Bergeron	1	4	2		6	0	
Brock Boeser	0	2	4	*	0	6	*
Logan Couture	0	0	6	*	1	5	*
Evgenii Dadonov	0	0	6	*	0	6	*
Jonathan Drouin	0	1	5	*	1	5	*
Ryan Dzingel	0	3	3	*	6	0	
Taylor Hall	1	2	4		6	0	
Nazem Kadri	0	4	2	*	6	0	
William Karlsson	0	1	5	*	3	3	*
Gabriel Landeskog	1	5	1		5	1	
Anthony Mantha	0	6	0		2	4	*
Auston Matthews	1	6	0		5	1	
Joe Pavelski	1	5	1		6	0	
Elias Pettersson	1	5	1		6	0	
Reilly Smith	0	3	3	*	1	5	*
Derek Stepan	0	6	0		5	1	
Teuvo Teravainen	0	0	6	*	0	6	*
James van Riemsdyk	0	6	0		5	1	
Jason Zucker	0	4	2	*	6	0	
				10			8

Table 56: Wrongly classified top 25% forwards in the 2018/19 season.

		Filter			Wrapper		
Player	C	Correct	Wrong	BN	Correct	Wrong	BN
Pontus Åberg	0	4	2		6	0	
Pavel Buchnevich	0	3	3		6	0	
J.T. Compher	0	3	3		5	1	
Evgenii Dadonov	1	5	1		6	0	
Jake Debrusk	1	3	3	*	1	5	*
Pierre-Luc Dubois	1	5	1		5	1	
Filip Forsberg	1	6	0		2	4	*
Erik Haula	0	4	2		5	1	
Jonathan Huberdeau	1	5	1		6	0	
Alex Iafallo	0	4	2		5	1	
Tyler Johnson	1	2	4	*	0	6	*
Nazem Kadri	1	5	1		2	4	*
Evander Kane	1	6	0		5	1	
Kasperi Kapanen	0	4	2		6	0	
William Karlsson	1	5	1		6	0	
Alexander Kerfoot	0	3	3		5	1	
Phil Kessel	1	5	1		4	2	*
James Neal	0	4	2		6	0	
Kyle Okposo	0	5	1		6	0	
Ondrej Palat	1	2	4	*	0	6	*
Richard Panik	0	4	2		6	0	
Nick Ritchie	0	4	2		6	0	
Brandon Saad	1	2	4	*	0	6	*
Paul Stastny	1	5	1		5	1	
Kyle Turris	1	2	4	*	0	6	*
Thomas Vanek	0	3	3		6	0	
				5			8

Table 57: Wrongly classified top 50% forwards in the 2018/19 season.

		Filter			Wrapper		
Player	C	Correct	Wrong	BN	Correct	Wrong	BN
David Backes	0	2	4		5	1	
Henrik Borgström	0	6	0		5	1	
Jesper Bratt	1	6	0		4	2	
Paul Byron	1	4	2		1	5	*
Andrew Copp	0	5	1		6	0	
Phillip Danault	1	6	0		5	1	
Joonas Donskoi	1	4	2		5	1	
Jesper Fast	1	0	6	*	1	5	*
Sam Gagner	0	6	0		4	2	
Conor Garland	0	3	3		2	4	
Michael Grabner	1	0	6	*	0	6	*
Martin Hanzal	0	6	0		5	1	
Danton Heinen	1	5	1		6	0	
Kasperi Kapanen	1	6	0		5	1	
Leo Komarov	1	1	5	*	0	6	*
Luke Kunin	0	5	1		4	2	*
Andrew Ladd	1	0	6	*	0	6	*
Artturi Lehkonen	0	4	2	*	4	2	*
Oskar Lindblom	0	0	6	*	0	6	*
Adam Lowry	1	0	6	*	0	6	*
Vladislav Namestnikov	1	6	0		5	1	
Alexander Nylander	0	6	0		5	1	
Victor Olofsson	0	5	1		5	1	
Corey Perry	1	1	5	*	2	4	*
Michael Rasmussen	0	5	1		4	2	
Evan Rodrigues	0	0	6	*	0	6	*
Jack Roslovic	0	5	1		6	0	
Sam Steel	0	6	0		5	1	
Ryan Strome	1	6	0		5	1	
Brandon Tanev	0	3	3	*	5	1	*
Robert Thomas	1	5	1		4	2	
Tage Thompson	0	5	1		5	1	
Chris Tierney	1	6	0		5	1	
Jake Virtanen	1	5	1		6	0	
				10			12

Appendix G: Attributes and their meaning

The following tables list the attributes with a name, description and scale. The scale can be continuous (C), discrete (D) or nominal (N).

Table 58: Attributes - 1.

Attribute	Name	Description	Scale
+/-	Plus-Minus	Goals for the team minus goals against the team when player is on the ice and player's team has equal or more players on the ice than the opposing team	D
A	Assists	Number of assists	D
Age	Age	Player's age	N
BLK	Blocks	Number of shots the player blocked	D
CA	Corsi Against	Even strength number of shots of opponent when player is on the ice	D
CF	Corsi For	Even strength number of shots of player's team when player is on the ice	D
CF%	Corsi For percentage	CF/(CF + CA)	C
CFoff%	Corsi For off-ice percentage	Percentage of all even strength number of shots that player's team performed when player is off the ice	C
CF% rel	Corsi For percentage relative	CF% - CFoff%	C
dZS%	Defensive zone start percentage	DZ Faceoffs / (OZ Faceoffs + DZ Faceoffs), that took place while player is on the ice	C
E+/-	Expected plus-minus	Expected even strength plus-minus given where shots came from while player is on the ice	C
EVA	Even strength assists	Number of assists at even strength	D
EVG	Even strength goals	Number of goals scored at even strength	D
EVTOI	Even strength time on ice	Number of seconds that the player played in even strength	D
FA	Fenwick Against	Even strength number of unblocked shots of opponent team when player is on the ice	D
FF	Fenwick For	Even strength number of unblocked shots of player's team when player is on the ice	D
FF%	Fenwick For percentage	FF/(FF + FA)	
FFoff%	Fenwick For off-ice percentage	Percentage of all even strength unblocked shots that player's team performed when player is off the ice	C
FF%rel	Fenwick For percentage relative	FF% - FFoff%	C
FO%	Face-off percentage	Percentage of won face-offs	C
FOL	Face-offs lost	Number of lost face-offs	D
FWON	Face-offs won	Number of won face-offs	D

Table 59: Attributes - 2.

Attribute	Name	Description	Scale
G	Goals	Number of goals	D
GA	Goals Against	Number of Goals Against	D
GA/60	Even strength Goals Against per 60 minutes	Number of Goals Against per 60 minutes when player is on ice	C
GA%	Goals Against percentage	Goals Against for goalkeeper relative to league average	C
GAA	Goals against average	Average number of Goals Against per 60 minutes	C
GF/60	Even strength Goals For per 60 minutes	Number of Goals For per 60 minutes when player is on ice	C
GP	Games played	Number of games the player played during the season	D
GPS	Goalkeeper point shares	Number of points for team that goalkeeper is considered to be responsible for	C
GS	Game starts	Number of games a goalkeeper started	D
GSAA	Goals saved above average	Saves percentage relative to league average	C
GV	Giveaways	Number of giveaways	D
GWG	Game winning goals	Number of game winning goals	D
HIT	Hits	Number of hits the player performed	D
L	Losses	Number of losses for goalkeepers	D
MIN	Minutes	Number of minutes played for goalkeepers	D
oiSH%	On-ice shooting percentage	Even strength shooting percentage when player is on ice	C
oiSV%	On-ice save percentage	Even strength save percentage when player is on ice	C
OTL	Overtime losses	Number of overtime losses for a goalkeeper	D
OVR	Rating	Number between 1 and 99 with higher values representing better players	D
oZS%	Offensive zone start percentage	OZ Faceoffs / (OZ Faceoffs + DZ Faceoffs) that took place while player is on the ice	C
PDO	PDO	oiSH% + oiSV%	C
PIM	Penalty Minutes	Number of penalty minutes	D
Player	Player	Player's name	N
POS	Position	Player position	N
PPA	Powerplay assists	Number of assists in powerplay	D
PPCF%rel	Powerplay Corsi For percentage relative	Similar to CF%rel, but for powerplay	C
PPG	Powerplay goals	Number of goals scored in powerplay	D
PPGA/60	Powerplay Goals Against per 60 minutes	Number of Goals Against per 60 minutes in powerplay	C
PPGF/60	Powerplay Goals For per 60 minutes	Number of Goals For per 60 minutes in powerplay	C
PPTOI	Powerplay time on ice	Number of seconds that the player played in powerplay	D

Table 60: Attributes - 3.

Attribute	Name	Description	Scale
PS	Point Shares	Number of points for team that player is considered to be responsible for	D
PTS	Points	G + A	D
QS	Quality starts	Number of starts for goalkeepers where the save percentage is higher than the average of the league during the season or higher than 88.5% in games where shots against was at most 20	D
QS%	Quality start percentage	QS/GS	C
RBS	Really bad starts	Number of starts for goalkeepers with save percentage lower than 85%	D
S	Shots on goal	Number of shots on goal	D
S%	Shots on goal percentage	Percentage of shots on goal that resulted in goals	D
SA	Shots against	Number of shots against for goalkeepers	D
SAtt.	Shot attempts	Number of shots attempted	D
SHA	Shorthanded assists	Number of assists in boxplay	D
SHCF%rel	Shorthanded Corsi For percentage relative	Similar to CF%rel, but for boxplay	C
SHFT	Shift	Shift length in seconds	D
SHG	Shorthanded goals	Number of goals scored in boxplay	D
SHGA/60	Shorthanded Goals Against per 60 minutes	Number of Goals Against per 60 minutes in boxplay	C
SHGF/60	Shorthanded Goals For per 60 minutes	Number of Goals For per 60 minutes in boxplay	C
SHTOI	Shorthanded time on ice	Number of seconds that the player played in boxplay	D
SO	Shutouts	Number of full games the goalkeeper did not concede a goal	D
SV	Saves	Number of saves for goalkeepers	D
SV%	Saves percentage	SV/SA	C
Thru%	Through percentage	Percentage of shots taken that go on net	C
TK	Takeaways	Number of takeaways	D
Tm	Team	Player's team	N
TOI	Time on ice	Number of seconds that the player played	D
TOI/60	Time on ice per 60 minutes	TOI/60	D
TOI(EV)	Time on ice per 60 minutes even strength	TOI/60 at even strength	D
W	Wins	Number of wins for goalkeepers	D