Text Mining of Scouting Reports as a Novel Data Source for Improving NHL Draft Analytics

Timo Seppa, Michael E. Schuckers, Mike Rovito

Abstract

Combining performance data ("stats") and scouting information is the Holy Grail of sports analytics. In this paper, we develop a methodology to combine scouting report information with performance metrics to improve the evaluation of players eligible for the NHL Entry Draft. In this new approach, text-mined data from scouting reports was used to develop variables for out-of-sample prediction. We demonstrate that by adding these variables to performance metrics, we can substantially improve the prediction of future performance.

1. Introduction

The Collective Bargaining Agreement between the National Hockey League (NHL) and the National Hockey League Players Association (NHLPA) places substantial restrictions on drafted players up to age 27 (or seven accrued seasons played), when they reach unrestricted free agent status [1]. Until that time, players remain under substantial control – regarding movement, and in effect, salary – of the team that drafted them. These factors make the contracts of younger players particularly favorable to the drafting teams. Consequently, teams that succeed in drafting the best talent have a significant advantage over teams that draft poorly, as they can retain the abilities of good-to-elite level players at a fraction of the cost[1] of a team needing to procure the same talents in free agency or via trade. A key related consideration is that peak performance occurs within this age span. In particular, peak scoring for forwards has been shown to occur at age 24-25 (various sources, including [2]).

Prior to the annual NHL Entry Draft ("Draft") held in late June, each team employs a small army of amateur scouts – typically 10 to 15 scouts – to watch thousands of junior, college, high school, and European games, plus international tournaments, in an attempt to assess the best draft-eligible amateur talent. Draft eligibility requires a player to be 18 on or before the following September 15[2]. In short, scouting departments must assess the future performance of still-developing 17-year-olds on hundreds of teams in a variety of leagues, based on a handful of viewings each. High school prospects, in particular, are especially hard to project given widely varying levels of competition. Further complicating matters, the skills of defensemen and goaltenders are considered to develop later than the skills of forwards, requiring projections further into the future. Amateur scouting has been an inexact science – and really, more art than science – requiring franchise-altering judgments to be made based on a limited number of subjective observations.

---

[1] Salary, and more importantly, cap hit
[2] Undrafted prospects remain eligible if they will be younger than 21 years old on December 31.

But, in recent years, after decades of the Draft being the exclusive realm of old hockey men with traditional hockey sensibilities – and unchallenged biases – "hockey analytics" has begun to pose new questions, crunch data, and look for inefficiencies in the scouting and drafting processes. Analysts have tackled the issue objectively, with numbers, seeking to uncover overrated and underrated draft prospects by the use of analytics techniques, advanced stats, or even simply through the intelligent application of conventional stats.

However, with the limited statistics publicly tracked in most "feeder leagues", a novel source of data was required to open new doors to analysis. To tap an additional, existing data set, this project set out to combine the old with the new: to use traditional scouting inputs – pre-Draft scouting reports on amateur prospects – as the source of data for a text-mining approach.

1.1. Prior research

Compared to other areas within the field of hockey analytics, draft analytics lay relatively untouched until the past few years, both due to the complexity in projecting a prospect's future pro performance and the relative lack of stats tracked in leagues outside of the NHL. Notable early work on "league equivalencies" was done by Desjardins [3], predicting player production when transferring from other leagues to the NHL, accomplished by comparing the average points per game of players while in the other league versus their NHL points per game in the following season, to obtain a translation factor. Vollman refined this process, while rebranding league translations as "NHL equivalencies". An excellent discussion and bibliography of related contributions and findings can be found in [4].

Fyffe wrote a series of articles (including [5], [6]) about projecting draft prospects' future career value, hinting at relevant factors correlating to future success, such as a prospect's exact age, proportion of goals to assists, penalty minutes, and the relative strength of a prospect's team.

More recently, Jessop demonstrated [7] that a simple algorithm would have easily outperformed Vancouver's selections in the 14 drafts from 2000-2013. The algorithm only selected forwards from the three CHL leagues (OHL, WHL, QMJHL), based solely on points in their first draft-eligible season, yet would have nearly doubled the number of NHL games played over the team's draft picks. Extending the work, Jessop and Weissbock found that 19 of the 30 NHL teams would have fared better using the simplistic algorithm than by their actual selections [8].

With Lawrence, Weissbock developed a cohort-based approach [9], [10], looking to predict NHL performance based on the junior performance of players with similar characteristics, while adjusting for league strength using NHL equivalencies. Attributes showing a strong correlation to NHL success were points per game, age, and height. For the 12 teams analyzed from 2005-2009, the approach outperformed team scouting by points scored, but performed similarly by games played.

Schuckers [11] combined scouting rankings from the NHL's Central Scouting Service with amateur performance and demographic variables – for prospects worldwide – to create an ordering of players that yielded better NHL performance than the actual Draft order. Future performance was measured in terms of NHL games played and time on ice in a prospect's first seven post-draft years.

Initial work on text mining of scouting data for draft analytics was begun by Seppa [12], and it forms the basis of this project. In this paper, our contribution is to utilize direct scouting information from text mining of scouting reports – combining it with advanced junior-league statistics – to improve the prediction of which players will perform best in the pros.

2. Scouting Data and Performance Metrics

This analysis uses traditional scouting reports – subjective scouting observations, albeit generated by "subject matter experts" – as the textual data source for applying objective analytics. As the scouting reports of the 30 NHL teams are proprietary, commercially-available scouting reports [13]-[19], created by an independent scouting service, were used as a proxy for how NHL scouting departments assessed available prospects prior to each of the last seven NHL Entry Drafts. In all, over 583,000 words from scouting reports for 2010-2016 were analyzed, covering 1,020 drafted prospects and 1,053 undrafted prospects.

In the selection of the textual data source, it was vital to have uniform information from year to year – Draft to Draft – consistent in content and style. The 2010 NHL Entry Draft was as far back as we could go with a publicly available source that covered a large number of prospects, similarly and in detail. Other sources cover fewer years and typically only about 100 prospects per Draft[3].

2.1. Text preparation: Tagging, exclusion lists, and lemmatization

The textual data required a significant amount of preparation before it could be utilized effectively for analytics. However, once the text-processing rules were in place, the inclusion of more scouting reports – from other sources, or in future years – would require only modest additional effort.

As a first step, individual scouting reports within the text corpus were coded[4] with player name, drafting team name (or "Undrafted"), draft year of the scouting report, and name of the data source (i.e. scouting service). Importantly, the tagging allows text-mining results to be reported by player, team, or team plus draft year. The second step, the creation of a custom exclusion list, enables the text-mining software to ignore phrases beginning or ending with unimportant words (e.g. "around", "from", "of", "the", "you"), and proper names of teams (e.g. "Acadie-Bathurst Titan") and players (e.g. "Aaron").

The third step was creating custom lemmatization rules appropriate for the text being studied – in this case, hockey scouting. Lemmatization is the grouping together of different inflected forms of a word so they can be analyzed together. For example, "ability" and "abilities" are candidates to group together as a single term with lemmatization. However, it is important to ensure that the various inflected forms occur similarly within phrases of the corpus. In particular, one inflected form should not convey a more positive or negative sentiment than another form. Therefore, before any custom lemmatization rule was finalized, we drilled down into each form's usage within the corpus. Positive or negative usage of the inflected forms had to agree at least 90% of the time for a custom lemmatization rule to be put in place.

2.2. Text mining: Categorization, sentiment analysis, topic extraction

Initial exploration indicated the need for a final step in processing the text before further analysis, creating custom categories. Fortunately, a few dozen player skills or traits are commonly cited by

---

[3] Subsequently, scouting reports from another commercially-available source [20]-[26] were added to the text corpus. Though spanning fewer players than our original source, it served to provide a "second opinion" on more-touted prospects. In all, over 739,000 words were available for analysis.

[4] The corpus was tagged ("coded") using a key feature of the selected text-mining software [27].

the hockey scouting community, such as defense, physicality, puck skills, skating, and work ethic. From these options, 22 categories were selected. Further, "Good" and "Poor" versions of these skills were created as subcategories, producing 44 possible subcategories, although categorizable phrases did not occur for all of the Poor subcategories within the corpus. For instance, it would be unusual for a scouting report to state directly that a player was poor at the power play.

The most commonly occurring 2-, 3-, 4-, and 5-word phrases within the corpus were then categorized, a lengthy process, but necessary towards the goal of accurately providing sentiment analysis on a player or team level. Phrases that could refer ambiguously to different categories were left uncategorized even if of like sentiment. Similarly, phrases that were not overwhelmingly positive or negative were left uncategorized. It was deemed better to exclude some potentially useful phrases (false negatives) than to introduce misleading output data for certain players or teams (false positives).

As shown in Figure 1, examples of phrases that were found to indicate POOR_EFFORT – carefully checked with keyword-in-context drilldown – included "BIT_LAZY", [needs to] "BRING_HIS_A_GAME", [needs to] "COMPETE_HARDER", [doesn't bring] "ENOUGH_EFFORT", and "LEFT_[me]_WANT[ing]_MORE" (lemmatization leads to the display of "LEFT_I_WANT_MORE"). Clearly, those phrasings are unlikely to be used to convey positive sentiment regarding EFFORT_AND_WORK_ETHIC.
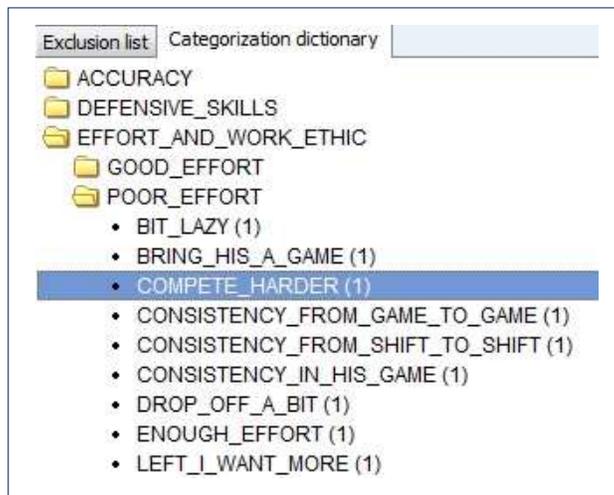

Figure 1. Some of the 2-, 3-, 4-, and 5- word phrases categorized as POOR_EFFORT

The phrase COMPETE_HARDER is a good illustration of the lemmatization choices that were made. The inflected forms COMPETE, COMPETES, and COMPETING were lemmatized, making each form equivalent for the text analytics, and therefore, all considered together for all phrases. However, that was not the case for HARD and HARDER. Upon close examination in keyword-in-context drilldown, the terms were not considered equivalent, which is why only COMPETE_HARDER was categorized under POOR_EFFORT. In Figure 2, the differences between phrases including COMPETE_HARDER and COMPETE_HARD are readily apparent, with COMPETE_HARDER indicating negative sentiment (red font) towards the prospect in every case, and COMPETE_HARD indicating positive sentiment (green font) towards the prospect in nearly every case.
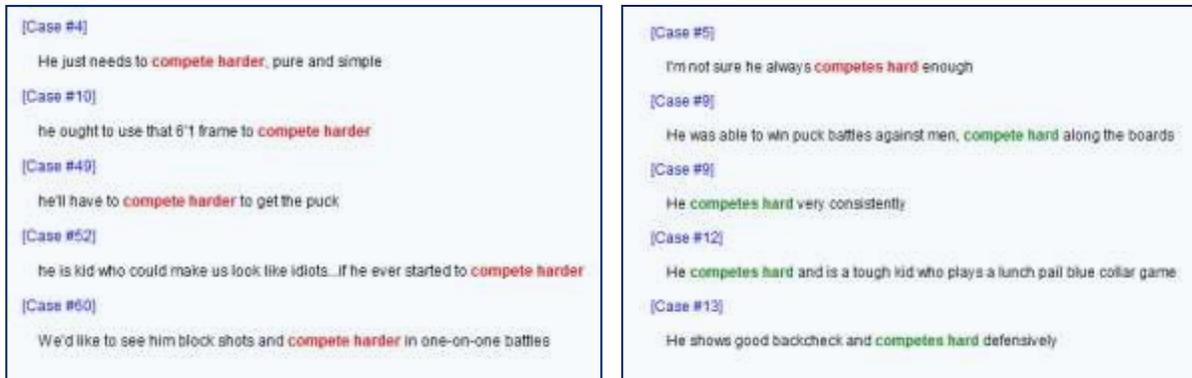
Figure 2. COMPETE_HARDER is categorized as POOR_EFFORT, while COMPETE_HARD is not.

2.3. Text mining: Topic extraction

After processing hundreds of terms for custom exclusion and lemmatization, and hundreds of phrases for custom categorization, the text corpus was ready for exploration utilizing topic extraction[5], a relative of cluster analysis. However, unlike the hierarchical clustering of k-means cluster analysis, topic extraction allows words to occur in more than one topic/cluster.

As shown in Figure 3, eight topics were generated, each with associated skill subcategories. Intuitively, the skillsets made sense together, corresponding to player subtypes or roles within hockey: top-six forward or top-pairing defenseman, power forward, goaltender, role player or defensive specialist, fourth liner or physical defenseman, marginal player ("roster filler"), power-play specialist or secondary scorer, and non-prospect players ("not much to look at"). In section 3, we will return to these role-oriented groupings of skillsets as being applicable to some of our models.
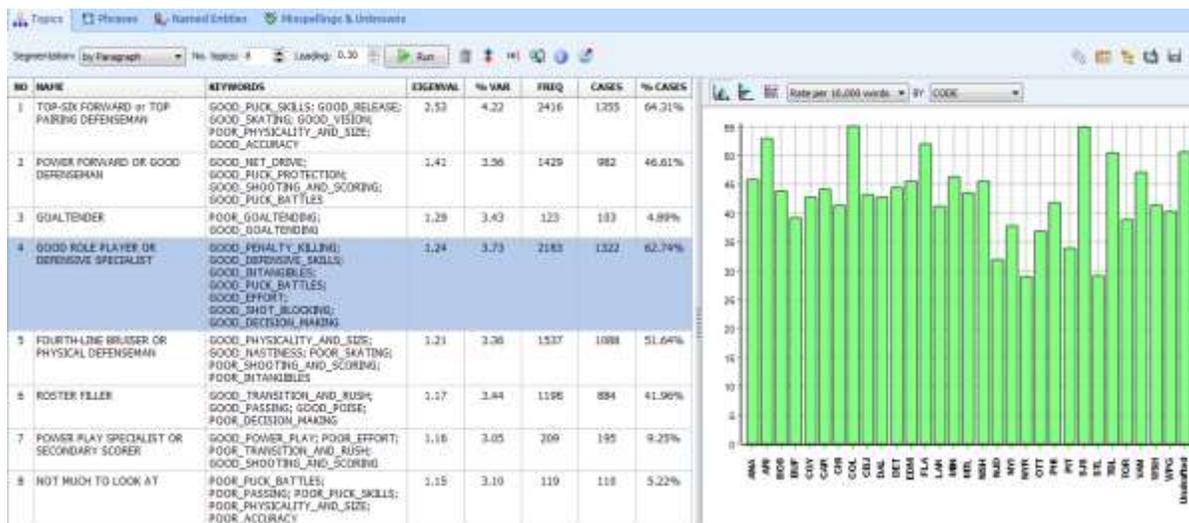

Figure 3. Topic extraction yields familiar hockey skillsets and roles.

---

[5] The topic extraction function performed by the text-mining software is described as "topic modeling" using "factor analysis with Varimax rotation", which "more realistically represents the polysemous nature of some words as well as the multiplicity of context of word usages". [28]

2.4. Performance metrics and target variables

Draft analytics have most frequently measured a prospect's success by NHL games played, time on ice, or points. However, the simplicity of these measures also limits the usefulness of information they provide. Our approach looked to measure more than mere existence at the pro level – more than just games played or time on ice – and to use a more discerning metric than total points.

For all of our performance variables, we chose to look at even-strength scoring to remove masking effects of players receiving or not receiving power-play opportunities. Further, we looked at even-strength scoring rates, as opposed to raw totals, to compare players on a per-60-minute basis. Keep in mind that power-play ice-time and overall ice-time are not simply functions of a player earning or not earning playing time – they are affected by management and coaching philosophies, as well as the depth of talent on a team and in an organization. Some teams famously prefer giving their prospects long AHL apprenticeships. This should not be counted against a player, or vice versa.

Specifically, we looked at even-strength goal-scoring rates (ESG/60) and even-strength primary assist rates (ESA1/60), as secondary assist rates are far less repeatable and predictive. Similarly, goals and primary assists produced in empty-net situations were not included. While we explored analyzing NHL even-strength scoring rates, our primary focus was to predict AHL even-strength scoring rates. Though an organization's ultimate goal is to draft the best future NHL players, there are many more prospects that make it to the highest tier of the minor leagues, the American Hockey League (AHL). This gave us more data points to compare[6].

For this study, we concentrated our analysis on 133 forwards from the three major junior leagues of Canadian Hockey League (CHL) meeting the following criteria: first eligible to be drafted between 2010-2015, scouting report available, 200+ minutes of even-strength time on ice (ESTOI) in their age-17 CHL season, 200+ cumulative minutes of ESTOI in the AHL from 2011-12 through 2015-16.

As CHL and AHL even-strength time on ice are not publicly tracked, they were estimated by a method suggested by Fyffe [29] and later utilized by Awad [30]. Team rates of even-strength goals for and goals against per ESTOI are calculated. Then, based on how many even-strength goals for and goals against a player was on the ice for, his ESTOI is estimated by multiplying by the team rate.

For CHL rates – the performance metrics – the first draft-eligible season ESG/60 and ESA1/60 rates were utilized. For the target variables AHL ESG/60 and AHL ESA1/60, average rates over the past five seasons, 2011-12 to 2015-16 were utilized. In the case of mature AHL players, use of multiple seasons increases the sample size for a more accurate representation of a player's actual skill.

3. Statistical Modeling and Results

The goal of the statistical analysis was to predict a prospect's future AHL performance given information known about that prospect when they were drafted. We broke our collection of variables into two types: scouting and performance. With over 100 variables, we took a multi-step approach to model building. The first step was to use random forests, regression trees, and elastic nets for dimension reduction [31]. An example of a regression tree used for variable selection is shown in Figure 4.

---

[6] Although a handful of top prospects never play in the AHL after heading straight to the NHL, many more players plateau at the AHL level – with a negligible number of career NHL games, if any.
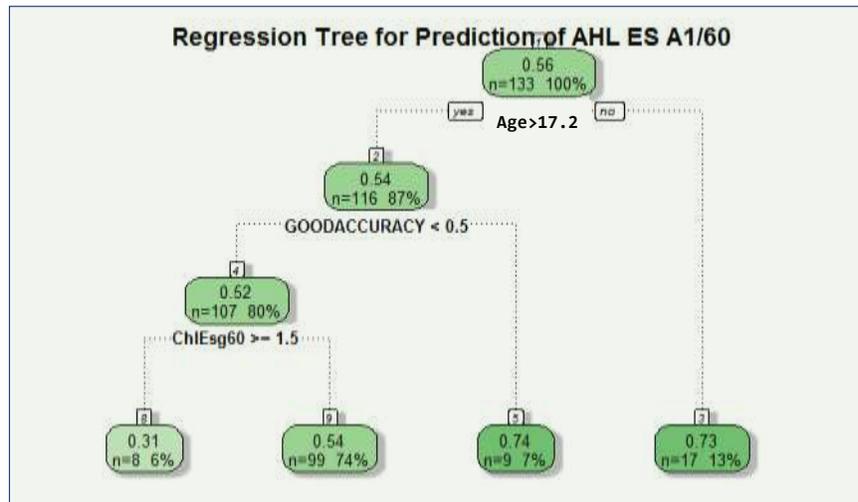
Figure 4. Example of a regression tree for combined scouting- and performance-variable prediction.

The second step was to consider an "all subsets" regression approach to the variables that were deemed important from the previous step. Then, we determined the final model for prediction in our third step by assessing how each candidate model performed in a 10-fold cross-validation. While we considered both random forests and regression trees for our final models, both of these approaches were outperformed by general linear models. In Table 1, we report and compare the predictive capabilities of our models for AHL ESG/60.

Table 1. Results for predicting AHL even-strength goals per 60 minutes

Set of predictors	Adjusted r^2	Average cross-validation mean squared error
Actual NHL team picks	2.9%	0.0573
Performance Only	12.1%	0.0521
Scouting Only	23.7%	0.0495
Performance & Scouting	25.0%	0.0445

To illustrate the utility of including the scouting information, we used three sets of predictors: Performance Only, Scouting Only, and Performance & Scouting. For each group, we followed the model-building approach given above. Results of these analyses are found in Table 1 and Table 2.

We have measured the outcome of our final prediction models by using the adjusted r^2 and the average mean squared error from our 10-fold cross-validation. In Table 1, the results from prediction of AHL ESG/60 show that use of both scouting and performance variables nearly doubled the penalized percent of variation explained, adjusted r^2, by the models, while the average mean squared error was about 15% smaller for the final model with both performance and scouting variables. Using only scouting variables did nearly as well as using both sets of predictors by adjusted r^2, but using the combined set improved the cross-validated error. The variables

included in the final model were CHL ESG/60, GOOD_RELEASE, GOOD_EFFORT, and POWER_FORWARD (per the topic extraction discussed in Section 2.3)[7].

Table 2. Results for predicting AHL even-strength primary assists per 60 minutes

Set of predictors	Adjusted r^2	Average cross-validation mean squared error
Actual NHL team picks	0.4%	0.0498
Performance Only	5.1%	0.0485
Scouting Only	14.8%	0.0450
Performance & Scouting	18.8%	0.0425

Using the same metrics for predictive ability, we present the results of the final models for AHL ESA1/60 in Table 2. As with goal scoring, we discovered substantial gains in predictive capacity by adding the text-mined scouting data to our models. Scouting only improved the adjusted r^2 by about 10% while performance plus scouting improved that metric by an additional 4%. The variables in the combined model included CHL ESA1/60, GOOD_ACCURACY, POOR_INTANGIBLES, GOOD_RELEASE, GOOD_VISION, GOOD_DEFENSIVE_SKILLS, POOR_PUCK_SKILLS, and Age relative to draft class[8]. As we saw with AHL ESG/60, the text-mined scouting variables did a better job of prediction than the performance variables alone, but the best performance for the cross-validated mean squared error was from having both sets of variables in our model.

An example from the 2016 NHL Entry Draft of a CHL prospect with a very good AHL projection is Alex DeBrincat, a five-foot-seven winger of the OHL's Erie Otters. DeBrincat was selected in the second round, 39th overall, by Chicago, making him the 14th of 40 CHL forwards selected. While we project those 40 CHL forwards to average 1.27 AHL even-strength goals plus primary assists per 60 minutes, the Performance & Scouting model has DeBrincat as the top projected CHL forward, at 1.53 even-strength primary points per 60 minutes. Why does he rank so high? An outstanding 1.51 CHL ESG/60, tied with third-overall pick Pierre-Luc Dubois, achieved without a Power Forward skillset. To date, he is averaging over two points per game in his 2016-17 OHL campaign.

4. Future Work

Our analysis focused on CHL forwards due to the large number of CHL players drafted each year[9] as well as the comparability of OHL, QMJHL, and WHL stats. However, with some success in predicting AHL scoring rates simply through the analysis of key scouting parameters, non-CHL prospects – in leagues of varying strengths – could be evaluated by a Scouting Only model while CHL prospects were projected by the Performance & Scouting model. This way, all scouted prospects could be evaluated by analytics.

---

[7] Not all coefficients are positive. Some traits have negative correlations.
[8] Not all coefficients are positive. Some traits have negative correlations.
[9] In 2016, CHL players comprised 96 of 211 draft picks (45%), and 15 of 30 first round picks (50%).

The inclusion of additional demographics, scouting reports, and performance metrics should be investigated in ongoing efforts of model improvement. And naturally, similar models could be created for CHL defensemen.

5. Conclusions

The central question of the best-selling novel *Moneyball* could have been phrased "Stats or scouting?" This is the question that hockey analytics has posed regarding the NHL Entry Draft as well. As in *Moneyball*, stats have been crowned the winner – at least as far as much of the hockey analytics community is concerned. Yet, how should we weigh in on that question now, given the results of this study?

The answer is nuanced and should not be misinterpreted.

The answer is not simply stats, or scouting, or both.

We have seen from prior work that draft analytics have outperformed the results of NHL teams at the Draft. Make no mistake: that is damning criticism of how teams have drafted, if "an unpaid nerd in his basement" outperforms a well-staffed hockey operations department. Organizations that have not re-evaluated how these vital selections are made – from scouting to analytics to Draft Day decisions – are either not paying attention or stubbornly ignoring reality. Such teams will fall behind the brave few that pay attention and take steps to reinvent how they operate.

Interestingly, though, our results have pointed toward the significant value of scouting information. Not only did text mining of scouting reports improve the performance of junior-league stats in evaluating the best pro prospects, but surprisingly – or not, depending on your point of view – the text-mined scouting data was more predictive on its own than the performance-based analytics (at least for this data set and these variables). If you think about it, this makes intuitive sense. While performance-based analytics may outperform traditional scouting biases, we see great value in identifying and properly weighting skillsets that do or do not translate to performance on the pro level.

Therefore, our analysis ultimately points towards good value, even in a brave new hockey analytics world, for scouting – specifically as a source of data for assessing the skillsets of prospects. However, as we have shown, not only should analytics get the nod over scouting, but importantly, the role of scouting should be skill assessment – only – not Draft Day decisions.

In summary, amateur scouting departments have value, providing expert assessments on prospect skillsets. Hockey analytics departments, with a significant focus on the vital field of draft analytics, should proliferate and grow in influence. Draft Day decisions should ideally rest in the hands of front office personnel who understand and can incorporate all of the insights of performance-based and scouting-based analytics. At the end of the day, it is all about making the best decisions towards building a perennial winner.

References

[1] NHLPA. (2013). *Collective bargaining agreement*. Retrieved July 8, 2016, from National Hockey League Players' Association, http://www.nhlpa.com/inside-nhlpa/collective-bargaining-agreement

[2] Seppa, T. (2011). Core Age and the Strategic Direction of NHL Teams. In T. Seppa (Ed.), *Hockey Prospectus 2011-12* (pp. 404–410). Createspace.

[3] Desjardins, G. (2004). *Hockey analysis and statistics*. Retrieved July 8, 2016, from Behind the Net, http://www.behindthenet.ca/projecting_to_nhl.php

[4] Vollman, R. (2013). Translating Data from Other Leagues. In *Rob Vollman's Hockey Abstract* (pp. 159–182)

[5] Fyffe, I. (2009, July 9). *Up and coming: Predicting NHL Success*. Retrieved July 8, 2016, from Hockey Prospectus, http://www.hockeyprospectus.com/puck/article.php?articleid=172

[6] Fyffe, I. (2009, July 16). *Up and coming: Refining the estimate*. Retrieved July 8, 2016, from Hockey Prospectus, http://www.hockeyprospectus.com/puck/article.php?articleid=184

[7] Jessop, R. (2014, May 20). *We think the Vancouver Canucks may have a scouting problem(!!!!)*. Retrieved July 8, 2016, from Canucks Army, http://canucksarmy.com/2014/5/20/we-think-the-vancouver-canucks-may-have-a-scouting-problem

[8] Weissbock, J. (2015, January 12). *The NHL has a scouting problem*. Retrieved July 8, 2016, from The Hockey Writers, http://thehockeywriters.com/the-nhl-has-a-scouting-problem/

[9] Lawrence, C. (2015, April 11). *The Draft Files: The Historical Cohort Based Approach Gets Its Sham On*. Retrieved July 8, 2016, from Canucks Army, http://canucksarmy.com/2015/4/11/the-draft-files-the-historical-cohort-based-approach-gets-its-sham-on

[10] Weissbock, J. (2015, May 26). *Draft Analytics: Unveiling the prospect cohort success model*. Retrieved July 8, 2016, from Canucks Army, http://canucksarmy.com/2015/5/26/draft-analytics-unveiling-the-prospect-cohort-success-model

[11] Schuckers, M. (2016). *Draft by numbers: Using data and analytics to improve National Hockey League (NHL) player selection*, MIT Sloan Sports Analytics Conference, Boston, 2016. Boston: MIT.

[12] Seppa, T. (2016, July 8). *Text mining of scouting reports for NHL Entry Draft insights.* Unpublished master's capstone final report, Quinnipiac University, Hamden, Connecticut.

[13] HockeyProspect.com. (2010). *2010 NHL Draft Guide*. United States: Hockey Press.

[14] HockeyProspect.com. (2011). *2011 NHL Draft Guide*. United States: Hockey Press.

[15] HockeyProspect.com. (2012). *2012 NHL Draft Black Book*. United States: Hockey Press.

[16] HockeyProspect.com. (2013). *2013 NHL Draft Black Book*. United States: Hockey Press.

[17] HockeyProspect.com. (2014). *2014 NHL Draft Black Book*. United States: Hockey Press.

[18] HockeyProspect.com. (2015). *2015 NHL Draft Black Book*. United States: Hockey Press.

[19] HockeyProspect.com. (2016). *2016 NHL Draft Black Book*. United States: Hockey Press.

[20] Red Line Report. (2010). *The Red Line Report: 2010 Draft Guide*. United States: Red Line Report.

[21] Red Line Report. (2011). *The Red Line Report: 2011 Draft Guide*. United States: Red Line Report.

[22] Red Line Report. (2012). *The Red Line Report: 2012 Draft Guide*. United States: Red Line Report.

[23] Red Line Report. (2013). *The Red Line Report: 2013 Draft Guide*. United States: Red Line Report.

[24] Red Line Report. (2014). *The Red Line Report: 2014 Draft Guide*. United States: Red Line Report.

[25] Red Line Report. (2015). *The Red Line Report: 2015 Draft Guide*. United States: Red Line Report.

[26] Red Line Report. (2016). *The Red Line Report: 2016 Draft Guide*. United States: Red Line Report.

[27] Provalis Research. (2014). QDAMiner 4: User's Guide. Montreal: Provalis Research.

[28] Provalis Research. (2014). WordStat7: User's Guide. Montreal: Provalis Research.

[29] Fyffe, I. (2001, March 14). *Puckerings archive: Estimating ice time*. Retrieved November 30, 2016, from Hockey Historysis, http://hockeyhistorysis.blogspot.com/2014/06/puckerings-archive-estimating-ice-time.html

[30] Awad, T. (2009, August 3). *Numbers on Ice: Understanding GVT, Part 2*. Retrieved November 30, 2016, from Hockey Prospectus, http://www.hockeyprospectus.com/puck/article.php?articleid=235

[31] Hastie, T., Tibshirani, R. and Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, Second Edition. New York: Springer.

Biographies

Timo Seppa is the former Editor-in-Chief of Hockey Prospectus, where he was editor and co-author of all six of their annual books. Over the past five seasons, he has been analytics consultant for multiple NHL and NCAA hockey teams. He recently earned his Master of Science in Business Analytics at Quinnipiac University.

Michael Schuckers is the Rutherford Professor of Statistics at St. Lawrence University. He is the co-founder of Statistical Sports Consulting, LLC and has consulted for organizations in baseball, hockey, and football.

Mike Rovito is a programmer who has worked on draft analytics for the past five seasons. He received his Master of Business Administration from Rutgers Business School in 2016.

For more information, contact Timo Seppa at 860-946-3187 or thetimoseppa@gmail.com.