

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

Timo Seppa
Analytics consultant

Scouting or stats?

Why not both?

But in a different way than you might think...

Scouting reports can provide a valuable **textual data source** for an analytics approach



Over a half million words

Seven years
of scouting
reports were
used, for
2010-2016

Textual data
comprised of
over **583,000**
words

More can be
added easily

18 — 2016 BLACK BOOK

Abramov, Vitalii

LW – Gatineau Olympiques (QMJHL) 5'10", 170

HockeyProspect.com Ranking: 29

Vitalii entered his first full season in the QMJHL as he was selected in the CHL Import Draft in 2015 coming over from the Russian Jr. league. He did not look out of place at all where in his 1st game he scored 4 points and finished the year with 93 points tied for 5th in league scoring.

Vitalii is an undersized forward with great speed and high hockey IQ that creates a lot of offensive chances for himself and his teammates. He is a good skater and is really good on his edges and has good mobility to get out of tight areas, and has quick acceleration to separate himself from defenders. He has a quick and accurate release with his shot and is really good at coming off the half boards or off the rush and cutting into the slot and taking shot on net. Vitalii has very good stickhandling ability where he can handle the puck at top speed coming into the zone and can stickhandle in a phone booth. He has high hockey IQ, he is really good at anticipating how the play is going to develop in front of him, where his teammates are going to be on the ice, he is also very good at finding the open areas of the zone to get set up for an offensive chance. He hockey sense really shows on the power play where he can be very creative and seems to always find the open man, or taking the proper shot on net. Vitalii also can be very creative coming into the zone of the rush, whether it's stopping and finding the trailer or making a move on the D, he always seems to slow the game down. We would like to see him go hard to the net or to the dirty areas of the rink; he tends to do a lot of his work on the perimeter.

Vitalii defensive of game would be the thing he has to work on for the next level, he needs to get stronger along the boards, there were a lot of times in our viewings he was losing the puck along the boards for the break out or if the defenseman pinches down. He also tends to run around too much in his own zone, instead of focusing on his own man and staying with them. Once he does get the puck along the boards he's good at those touch passes to his center curling or touch pass of the boards. Vitalii has 1st round talent and hockey smarts, he just needs to focus on getting bigger and stronger and working on the little parts to the game away from the puck, he could be a very good hockey player at the next level and be very effective.

Quotable: "Vitalii is an undersized forward with good skating ability and high hockey IQ, he can create a lot of offensive chances, whether it's scanning down the wing and stopping to wait for the play to develop in front of him, or coming down the wing and cutting into the slot using his quick release and accurate shot." - HP Scout, Justin Spruille

Quotable: "He is a hard working forward that competes every shift he is on the ice, seldom takes a night off." - NHL Scout (April 2016)

Quotable: "Vitalii's vision on the ice is outstanding, he somehow finds the open man that nobody could find. I love his ability on the power play to control the puck and understand where and who the puck should go to." - HP Scout, Justin Spruille (April 2016)

Quotable: "One of our top interviews, he knows his game and gave great answers to our questions." - NHL Scout (Combine week)

Quotable: "(Laughing) He was good (combine interview), I liked the kid. His english was fine and he was group oriented, he spoke to everyone in the room. He said he knows he's 5'9" but it doesn't deter him and that he uses different ways to get to the net. He had some inner drive in him. Good kid. He knows what he is." NHL Scout (Combine week)

HOCKEYPROSPECT.com 2016

HL DRAFT BLACK BOOK



Massaging the text: exclusion lists, lemmatization

Keyword-in-Context

List: User defined Sort by: Case number Word: ABILITY Context delimiter: Paragraph

	FREQ	CASENO	KEYWORD	CODE	CATEGORY
SHOW ALL ITEMS		2	ns and even coaches. Etem's game is built around speed. He uses his top-notch skating	ability	to gain space on opposing defensemen, protect the puck and create opportunities. Althc
1731 - ability	9	2	at times on the defensive side of the puck. Etem certainly has the speed and offensive	ability	to be a productive NHL player. Whether or not he is a 3rd line center or 1st line will depe
349	8	2	pend on how well he develops his overall game. With his attitude and natural leadership	ability	, it is safe to say that Etem could slide into a 3rd line NHL Role, but his speed makes top-
307 - ability	7	4	forecheck, doesn't take very many shifts off and prides himself in his strong two-way	ability	. Rakell is particularly effective when on the penalty kill. Has good hands and controls th
513	7	4	we project him to play if he cracks an NHL lineup some day. We don't see top 6 scoring	ability	. In fact we don't think he is close as far as scoring and NHL shooting ability. Rakell shoi
1809	6	4	top 6 scoring ability. In fact we don't think he is close as far as scoring and NHL shooting	ability	. Rakell showed nicely at the World Juniors in Buffalo by banging players around and pl
1785 - ability	6	8	ctors mirroring forward Victor Rask's production. Karlsson showed off his great skating	ability	, combined with his physical play. Karlsson plays at a very high energy, but is also very
66	6	8	ysical play. Karlsson plays at a very high energy, but is also very creative, and has the	ability	to make high difficulty passes under pressure.
125	5	11	p for it, he just didn't need to in Lincoln. Roy is easily a draftable player. His goal scoring	ability	was something unlike the USHL has seen in years. In terms of projecting a round, with i
1863 - ability	5	11	is patient enough to wait the extra half-second to make the better play. His goal scoring	ability	is off the charts (scoring 54 goals in the USHL is unheard of and his goals were no fluk
1822 - ability	5	12	t's, 5'9.25" 170 Besse is a sleeper Minnesota hockey player who has great goal scoring	ability	. He played this season for Wayzata of Minnesota high school hockey. You just cant dei
1733 - ability	5	15	at shot and is a natural goal scorer. He's a good skater and you can see a ton of natural	ability	. He had some shifts that looked kind of lazy. He has good instincts and knows how to f
1716 - ability	5	17	he could get them on net more often. The biggest concern of Theodore is his defensive	abilities	. He has a tough time getting into proper position without the puck, and often gets caught
2037	5	17	as skilled with the puck as Shea Theodore in the upcoming NHL Entry Draft. He has the	ability	to go coast to coast and score a goal to change the momentum of the game, and providi
1924 - ability	5	17	ng reach to keep pucks away from opponents as he quickly picks up speed. He has the	ability	to quickly identify open lanes, and changes his direction in a hurry to drive to the net and
790 - ability	5	17	l the tools to become an NHL defenseman one day, but his play without the puck and his	ability	to think the game will determine just how successful he can be. Quotable: "I think his off
1204 - abilities	4	18	, aware, poised, and fundamentally sound, with pinches at the right times, puck-rushing	ability	, excellent passing, the ability to quarterback a power play and get pucks on net, and in
117	4	18	entally sound, with pinches at the right times, puck-rushing ability, excellent passing, the	ability	to quarterback a power play and get pucks on net, and in
1126	4	19	puck well and can carry the puck end to end seemingly at will. It was this puck rushing	ability	along with his hockey sense and quick st
143 - abilities	4	20	ar viewings winning some draws clean. Flashed a good wrister and some stickhandling	abilities	.
141 - ability	4	20	0" 165 Kase has shown us flashes of having great skills and excellent speed. Liked his	ability	to change speeds quickly, delaying and d
1358 - ability	4	23	ind anticipates developing plays quickly. He doesn't get enough credit for his playmaking	ability	, probably because he has had so little ta
1336	4	23	e is one of the most NHL ready prospects in this draft. He blends size, strength, scoring	ability	and a mean streak into his game. We exp
1849 - ability	4	25	arsson is a promising defenseman blessed with projectable size and impressive skating	ability	. At 17 he already received some icetime
1810 - ability	4	26	ongest player in corners and you don't see him dishing out any big hits. Nattinen has the	ability	to make his linemates better, he sees the
178	4	26	xup to be physical. I think his strength is defensive hockey. While he has shown me the	ability	to distribute the puck, I think he really exc
1768 - ability	4	27	all three forward positions, Sideroff is a particularly good fit on the wing because of his	ability	to drive wide with his speed. Indeed, Sider
1748 - ability	4	29	heavy accurate shot that he is able to keep low for deflections or rebounds. His shot and	ability	to make the right play is especially helpful
1743 - ability	4				
1721	4				
17	4				

score a goal to change the momentum of the game, and provide a strong offensive presence from the backend. Theodore is excellent at using his long reach to keep pucks aw to quickly identify open lanes, and changes his direction in a hurry to drive to the net and forces his opponents to be caught flat footed. He can control the puck so effortlessly his head up to identify any open teammates. Theodore can also turn on a dime to get away from defenders, and buy his team more time to get open. He often stays on the ice for the entire power play for the thunderous and does a very admirable job. He moves the puck quickly and accurately, but may give up a few turnovers throughout a game as he gambles and tries to connect on long stretch passes. Theodore is definitely a pass-first player who can improve on his shot from the point. His slap shot could have more velocity on it, while he could get them on net more often. The biggest concern of Theodore is his defensive **abilities**. He has a tough time getting into proper position without the puck, and

WordStat 7.1.4 -

Dictionaries Options Frequencies Extraction

Dictionaries

Preprocessing: 3-grams Edit New

Stemming: English (Porter) Edit New

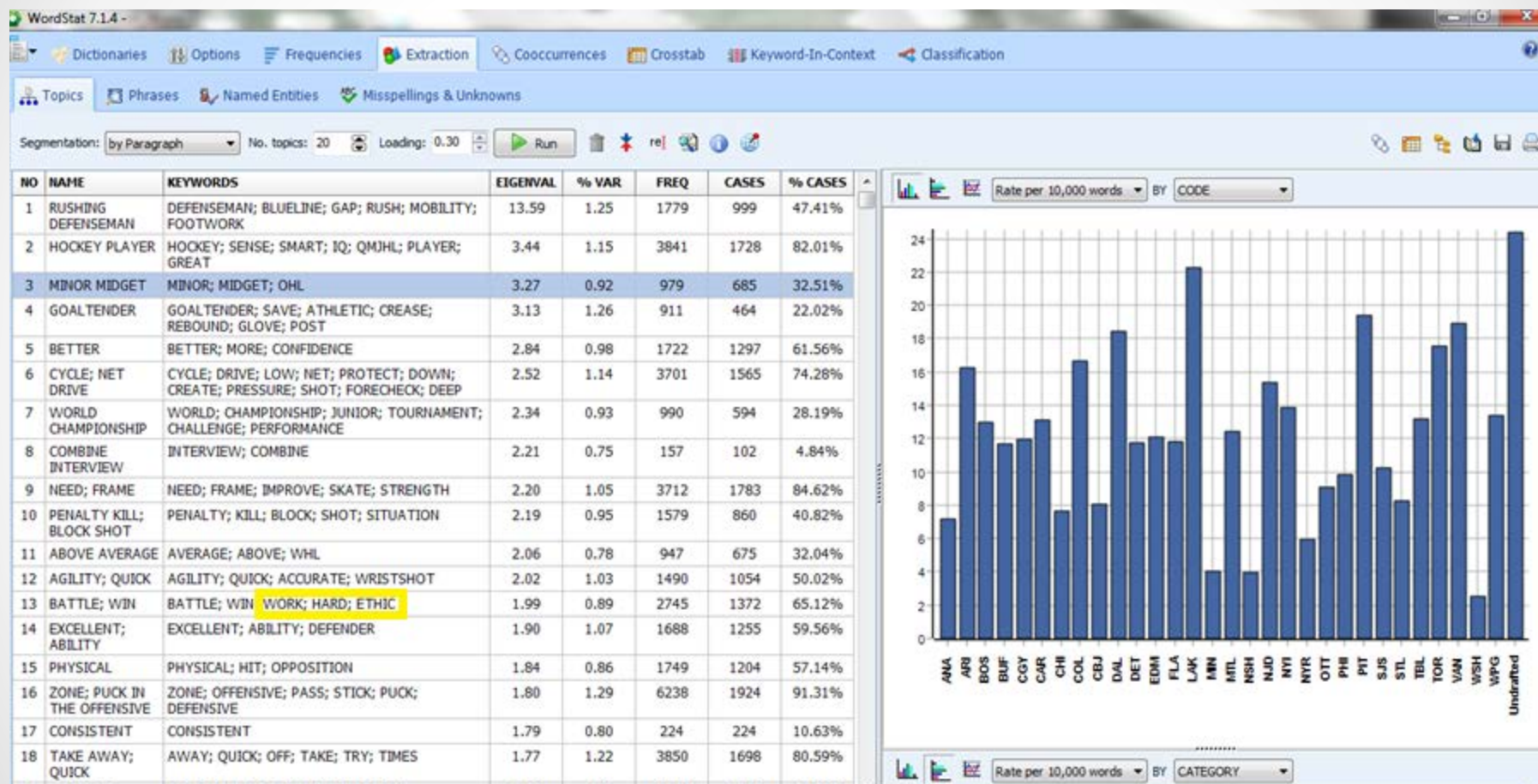
Substitution: English - Lemmatization Edit New

Exclusion: English.exc Open New

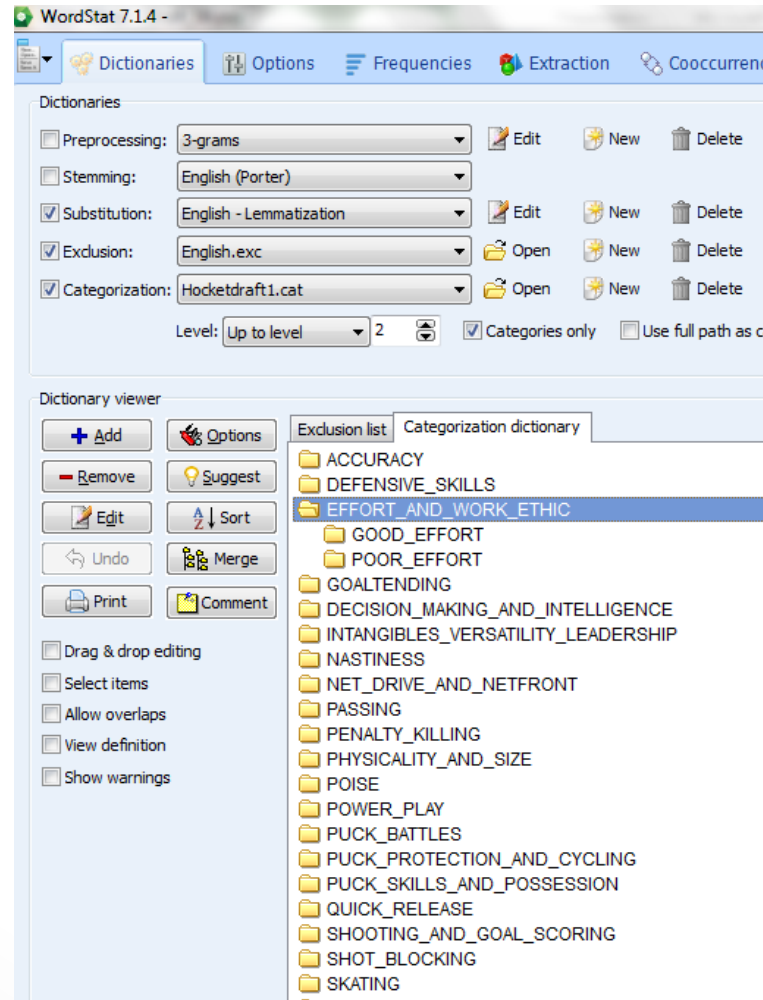
Categorization: Open New

Lemmatization: grouping together different inflected forms of a word for analysis

Exploring the data via topic extraction



Creating categories -> sentiment analysis



What constitutes “Poor effort”?

Exclusion list Categorization dictionary

- ACCURACY
- DEFENSIVE_SKILLS
- EFFORT_AND_WORK_ETHIC
 - GOOD_EFFORT
 - POOR_EFFORT
 - BIT_LAZY (1)
 - BRING_HIS_A_GAME (1)
 - COMPETE_HARDER (1)
 - CONSISTENCY_FROM_GAME_TO_GAME (1)
 - CONSISTENCY_FROM_SHIFT_TO_SHIFT (1)
 - CONSISTENCY_IN_HIS_GAME (1)
 - DROP_OFF_A_BIT (1)
 - ENOUGH_EFFORT (1)
 - LEFT_I_WANT_MORE (1)

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

Is “compete hard” = “compete harder”?

[Case #4]

He just needs to **compete harder**, pure and simple

[Case #10]

he ought to use that 6'1 frame to **compete harder**

[Case #49]

he'll have to **compete harder** to get the puck

[Case #52]

he is kid who could make us look like idiots...if he ever started to **compete harder**

[Case #60]

We'd like to see him block shots and **compete harder** in one-on-one battles

[Case #5]

I'm not sure he always **competes hard** enough

[Case #9]

He was able to win puck battles against men, **compete hard** along the boards

[Case #9]

He **competes hard** very consistently

[Case #12]

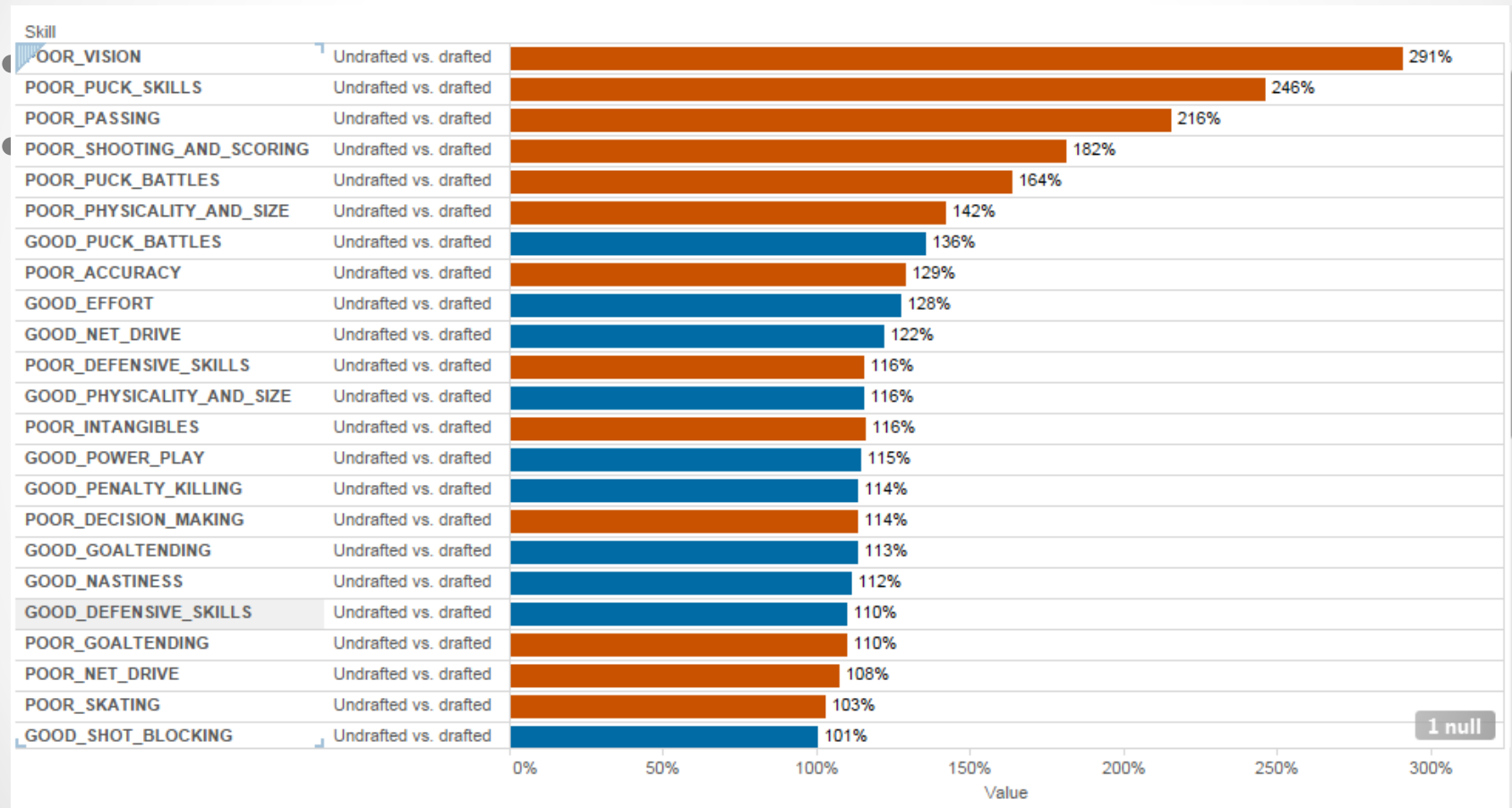
He **competes hard** and is a tough kid who plays a lunch pail blue collar game

[Case #13]

He shows good backcheck and **competes hard** defensively

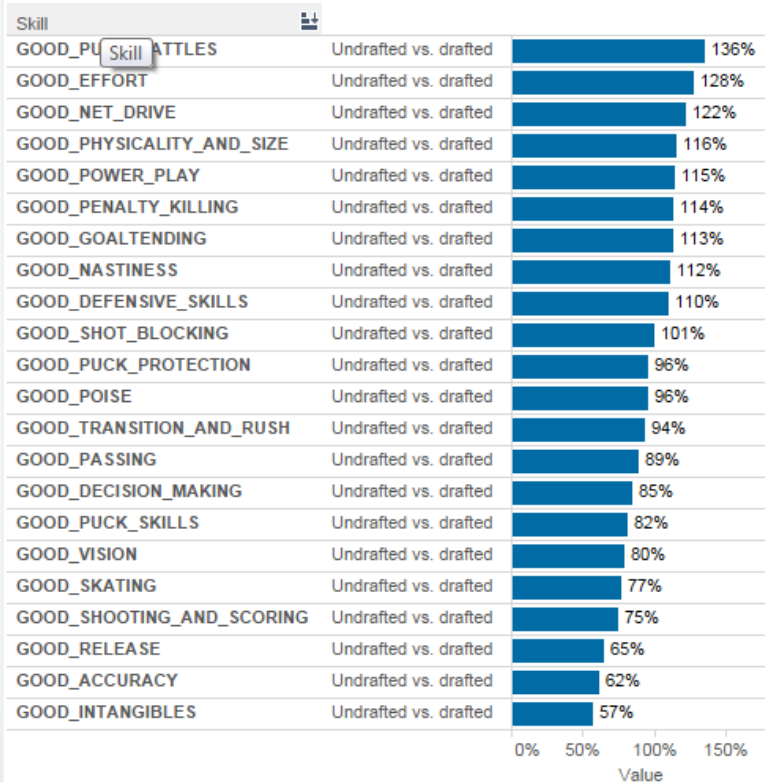
Conclusion: COMPETE_HARDER is categorized as POOR_EFFORT, while COMPETE_HARD is not

What can we say about undrafted players?

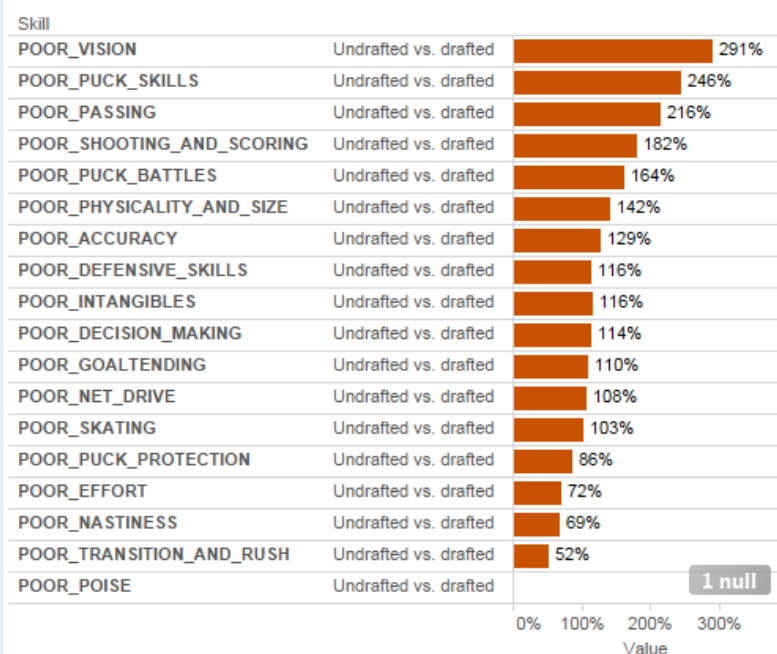


Undrafted players are better at...

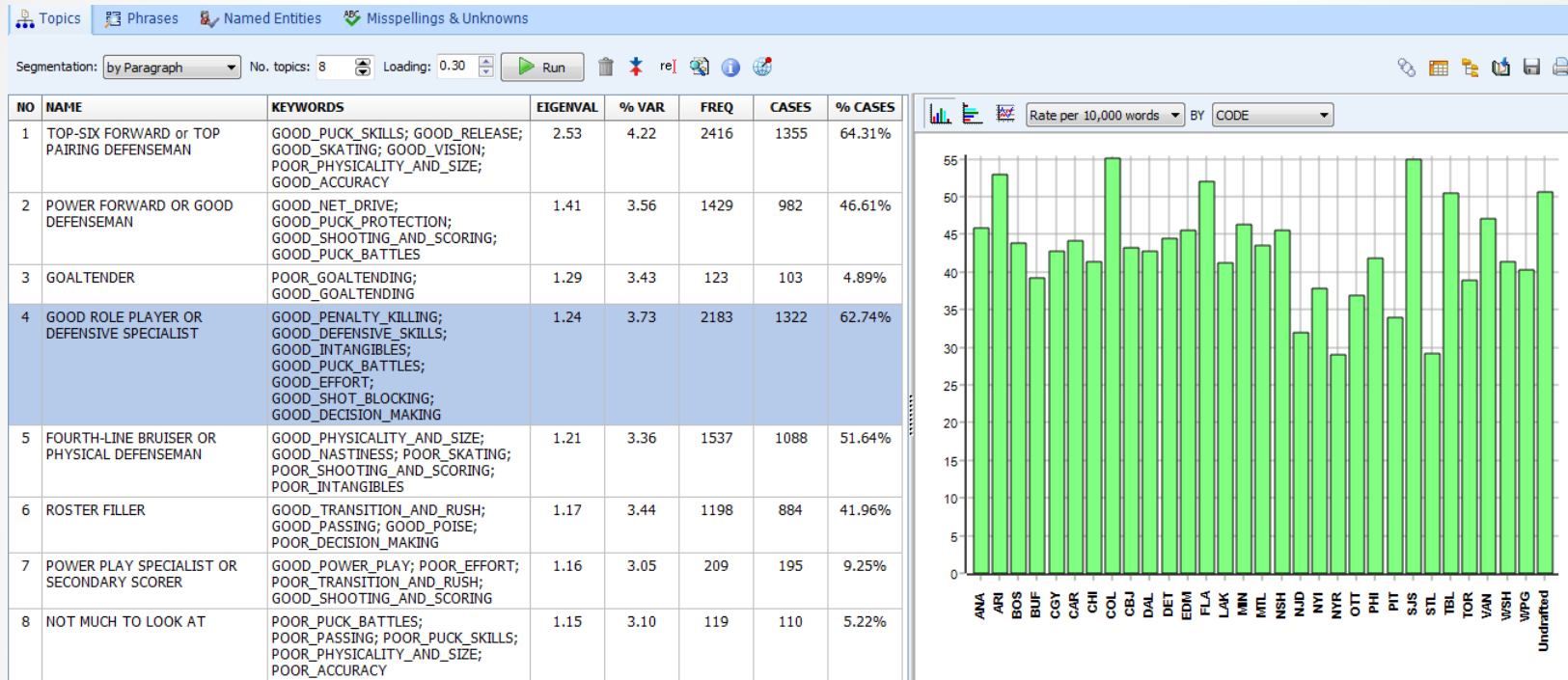
Undrafted - good skills



Undrafted - poor skills



Topic extraction (on categories)-> common skillsets



Topic extraction yields familiar hockey archetypes

Predicting ESG/60

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

Results for predicting
AHL even-strength goals per 60 minutes
This model used CHL ESG/60, GOOD_RELEASE, GOOD_EFFORT,
POWER_FORWARD (not all positive coefficients)

Predicting ESA1/60

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

Results for predicting

AHL even-strength primary assists per 60 minutes

This model used CHL ESA1/60, GOOD_ACCURACY, POOR_INTANGIBLES, GOOD_RELEASE, GOOD_VISION, GOOD_DEFENSIVE_SKILLS, POOR_PUCK_SKILLS (not all positive coefficients)

Example – Alex DeBrincat

- Erie Otters, 5'7" forward
- 51-50-101 in 60 GP in 2015-16
- 65-62-127 in 63 GP in 2016-17
- Selected 2nd round, 39th overall by Chicago

Our rank	Actual draft rank (of 40 CHL forwards)	Player (overall selection)	Predicted AHL ES primary pts/60
1	14	Alex DeBrincat (39)	1.53
2	40	Ty Ronning (201)	1.51
3	39	Dmitry Sokolov (196)	1.50
4	22	Brandon Gignac (80)	1.50
5	3	Alexander Nylander (8)	1.50
6	1	Pierre-Luc Dubois (3)	1.40

Conclusions

1. Draft analytics have outperformed team picks.
2. Surprisingly (?) scouting-based analytics performed better than stat-based analytics at predicting future (pro) scoring rates.
3. Using both together provide the best predictions.
4. Scouting remains vital for assessing skills **BUT**
5. Decision making on Draft Day needs to evolve to include draft analytics in a prominent role, including text-mined scouting data.

Ongoing work - version 2.0

- Projections for both forwards and defensemen
- Performance analytics for all four major junior leagues (OHL, QMJHL, WHL, USHL)
- Scouting-based projections for all prospects (European, all junior leagues, high school)
- Use both NHL and AHL target variables
- Add 2017 scouting reports to corpus
- Project future performance of 2017 draft class

Thanks!



This presentation was based on a white paper by **Timo Seppa**, **Michael Schuckers**, and **Mike Rovito**

<http://statsportsconsulting.com/2017/02/22/text-mining-and-improving-nhl-draft-analytics/>