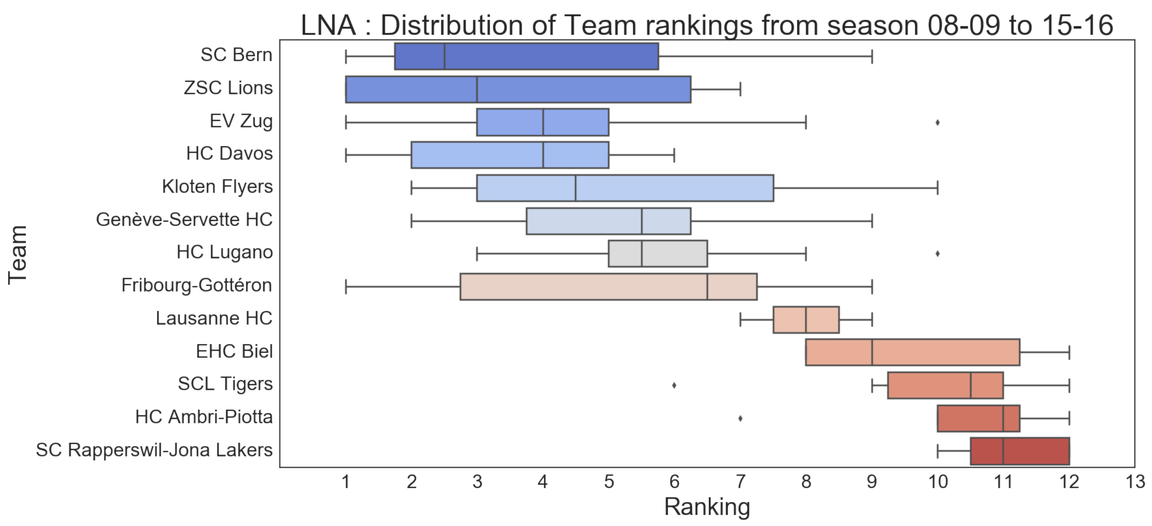
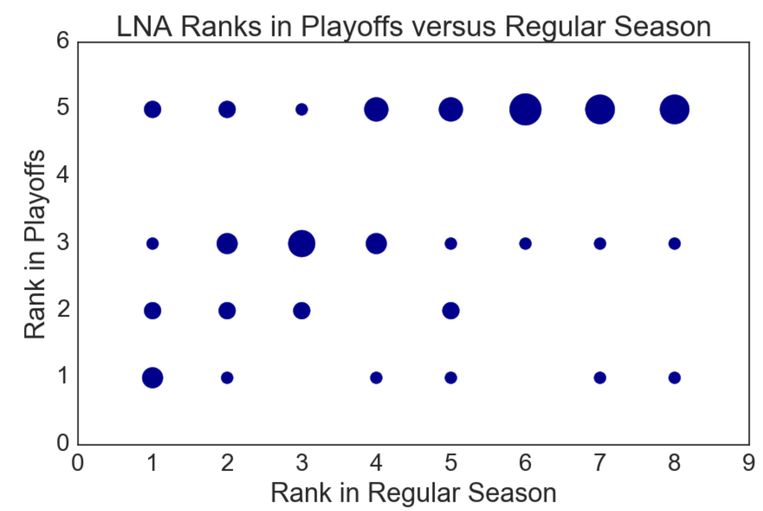
Predicting Swiss Ice Hockey Playoffs using Machine Learning Models

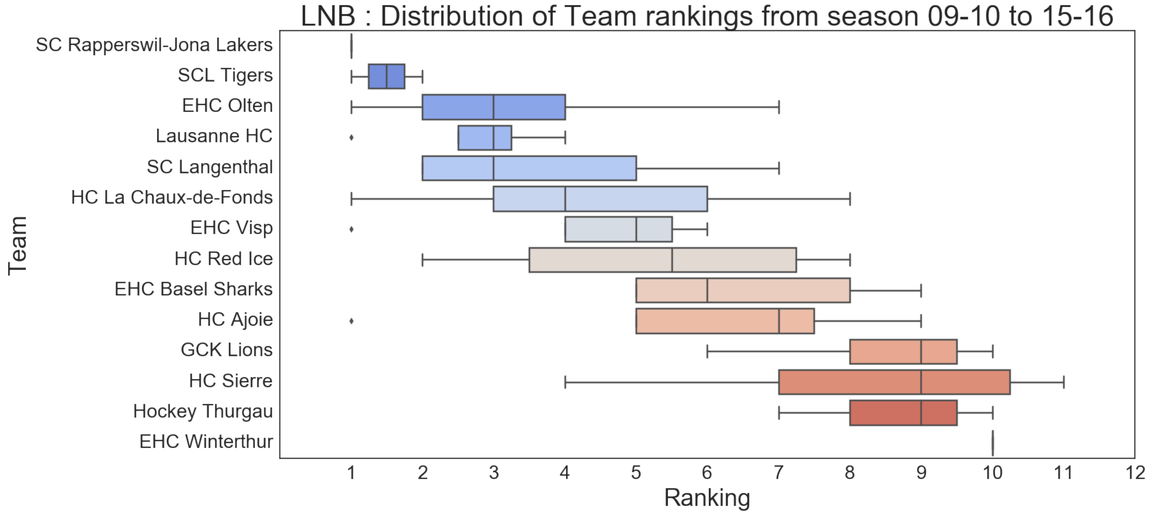
Sport is a very interesting domain when it comes to Data Analysis. We wanted to use a clean Dataset, in order to extract real informations and being able to make accurate predictions. Because of the few choices of Sports with large Dataset in Switzerland, we went with Ice Hockey.

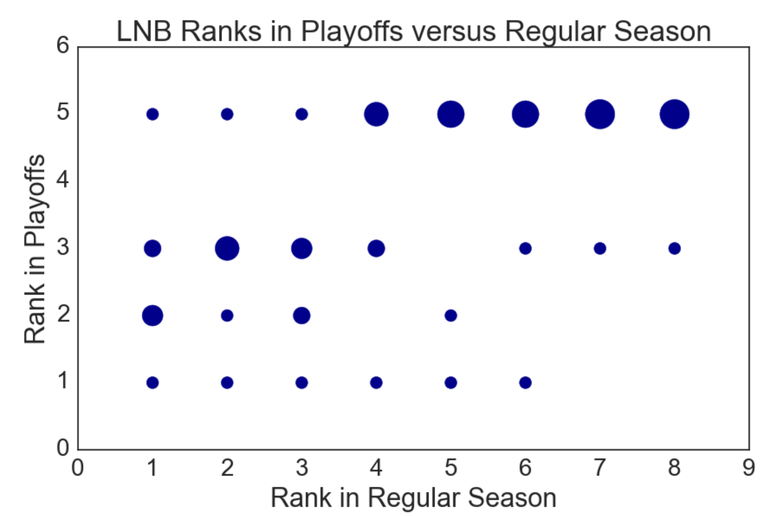
We extracted statistics from the regular season in the Swiss Ice Hockey League in order to predict the outcomes of the playoffs. We used different models, such as SVM, Neural Network, Random Forest, Naive Bayes Classifier and Logistic Regression.

Is using regular season data relevant for playoffs ?



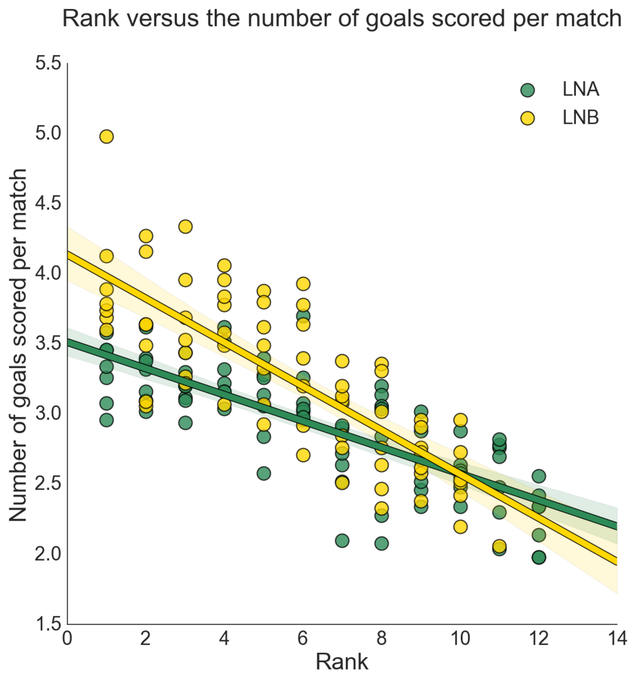


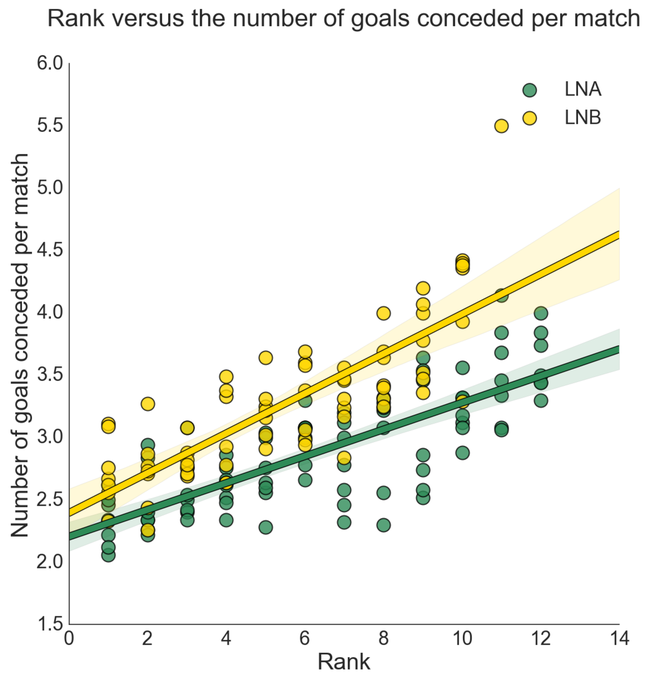


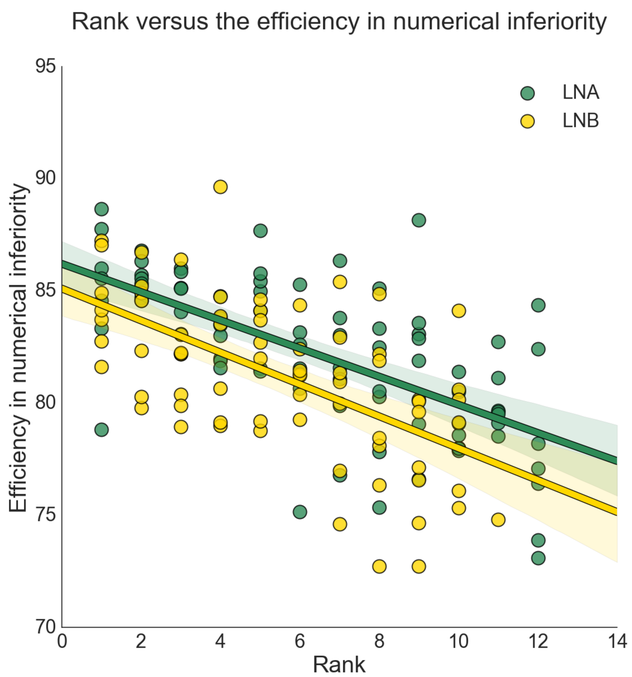


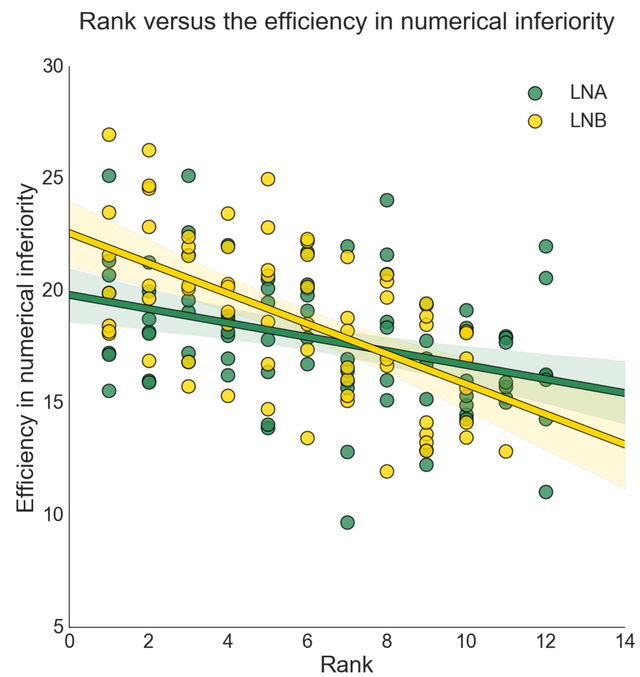
The hierarchy is stable within both leagues

There is correlation between ranking in playoffs and in regular season, for both leagues









The dataset contains 34 usable features, distributed in 4 categories:

- Goals
- Penalties
- Power Play
- Power Kill

Here are some features highly correlated with the ranking.

Data Processing Pipeline

Playoffs games

	Home	Away	Resultat	Year	Result	Series_ID	Opponent 0	Opponent 1
0	HC Lugano	SC Bern	5:4	1516	1	1516_HC Lugano SC Bern	HC Lugano	SC Bern
1	SC Bern	HC Lugano	1:0	1516	1	1516_HC Lugano SC Bern	HC Lugano	SC Bern
2	HC Lugano	SC Bern	2:3	1516	1	1516_HC Lugano SC Bern	HC Lugano	SC Bern

Regular Season Statistics

		GF	PK SI	PK SI/GP	PK GA	PK GA/GP	PK%	PKT	PK time/GP
Team	**Year**								
EHC Biel	**0809**	126	306	6.12	55	1.10	82.03	27694	553
	0910	134	266	5.32	45	0.90	83.08	24336	486
	1011	135	201	4.02	47	0.94	76.62	19409	388
	1112	114	200	4.00	35	0.70	82.50	19787	395
	1213	160	198	3.96	33	0.66	83.33	19053	381

Aggregation per series
(first to win 4 games)

Compute the difference
between the two teams

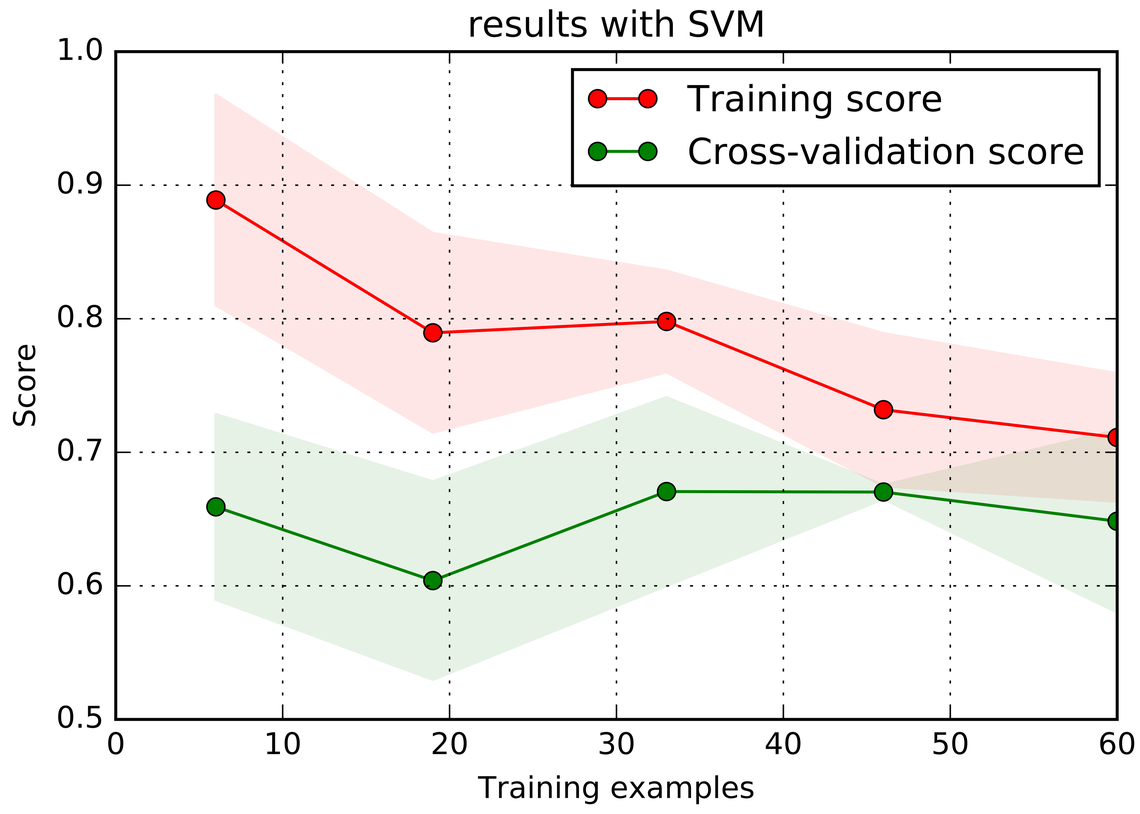
	Winner	Year	Opponent 1	Opponent 0	GF	PK SI	PK SI/GP	PK GA	PK GA/GP
Series_ID									
0809_EV Zug Kloten Flyers	1	0809	Kloten Flyers	EV Zug	-3.0	18.0	0.36	4.0	0.08
0809_EV Zug SC Bern	0	0809	SC Bern	EV Zug	-16.0	27.0	0.54	-5.0	-0.10
0809_Fribourg-Gottéron HC Davos	1	0809	HC Davos	Fribourg-Gottéron	-30.0	-25.0	-0.50	11.0	0.22
0809_Fribourg-Gottéron ZSC Lions	0	0809	ZSC Lions	Fribourg-Gottéron	-22.0	31.0	0.62	17.0	0.34
0809_Genève-Servette HC Kloten Flyers	1	0809	Kloten Flyers	Genève-Servette HC	-17.0	26.0	0.52	0.0	0.00

Final dataset

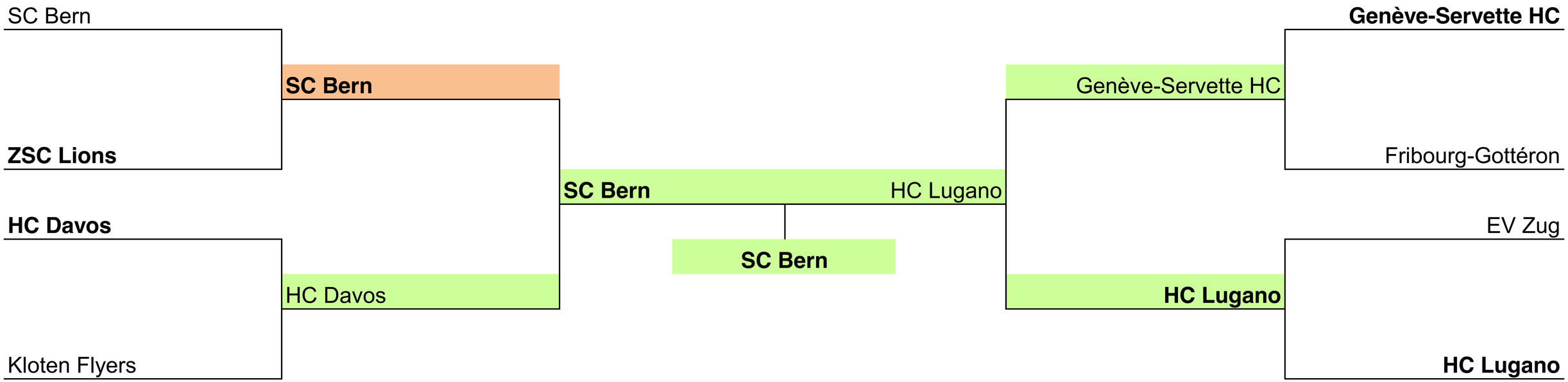
Models

Classifier	Mean Validation Accuracy (%)	std
Baseline	**50**	**0**
Random Forest	**57**	**0.08**
Logistic Regression	**61**	**0.14**
Neural Network	**66**	**0.07**
SVM	**71**	**0.1**

- Models were trained on the final dataset
- Parameters were selected using grid search and 5-fold cross-validation
- SVM was chosen as our final model



Predictions of the 15-16 season playoffs



The final model correctly predicted 6 out of 7 playoffs games in the 2015-2016 Season.

Nevertheless, games outcomes in Ice Hockey are more difficult to predict than basketball for example, as total score range is very small. This is the main reason why our model predicts directly the winner to the series and not the outcome of each game.

Issues

- Since statistics are available only since 2008, the size of dataset is limited
- To cope with the above mentioned issue, we included data from LNB, although it was not as reliable as the original data from LNA
- The statistics by player are available, unfortunately they are not complete enough to be used efficiently. Therefore we had to stick to team statistics.

Future Outlook

- It could be interesting to use more advanced statistics (not available for the Swiss League)
- Using the win rate of the matchup in the regular season as a feature for prediction
- We could apply our method to other sports with similar playoffs format