Department of Computer Science Course: CS-505-740
Semester: Fall 2022
Under the Guidance Of: Dr.Juefei Yuan

South East Missouri State University

Group Name: Team_Coders

Team Member	SEMO ID
Sai Gajjala	S02035282
Usha Devulapally	S02035040
Kolli Pradeep Kumar	S02033678
Phanindra Kumar Perumalla	S02036172
Nachi Ketan Reddy	S02034793
Sharath Kumar Polisetty	S02036169

Outline

Abstract

Sports and Math do they fit together

- **Goal:**
  Check whether emerging technologies from machine learning and data mining can be used to get any meaningful predicts. That data and statistics play a huge role in the industry.
- **Why:**
  From historical data and fundamental scorekeeping to algorithmic performance forecasting and extremely specific player statistics, big data is the industrys most valuable player.

Introduction

Where to find statistics data

Recently National Hockey League started they API service with stats data from 2010 year until present. API can be accessed publicly for free at:
statsapi.web.nhl.com/api/v1/endpoint_name

This API operates on HTTP/S protocol and returns results in JSON format. Among others following statistics can be obtained from it:

- Game
- Team
- Player

What statistics can be obtained

Following endpoints are available:

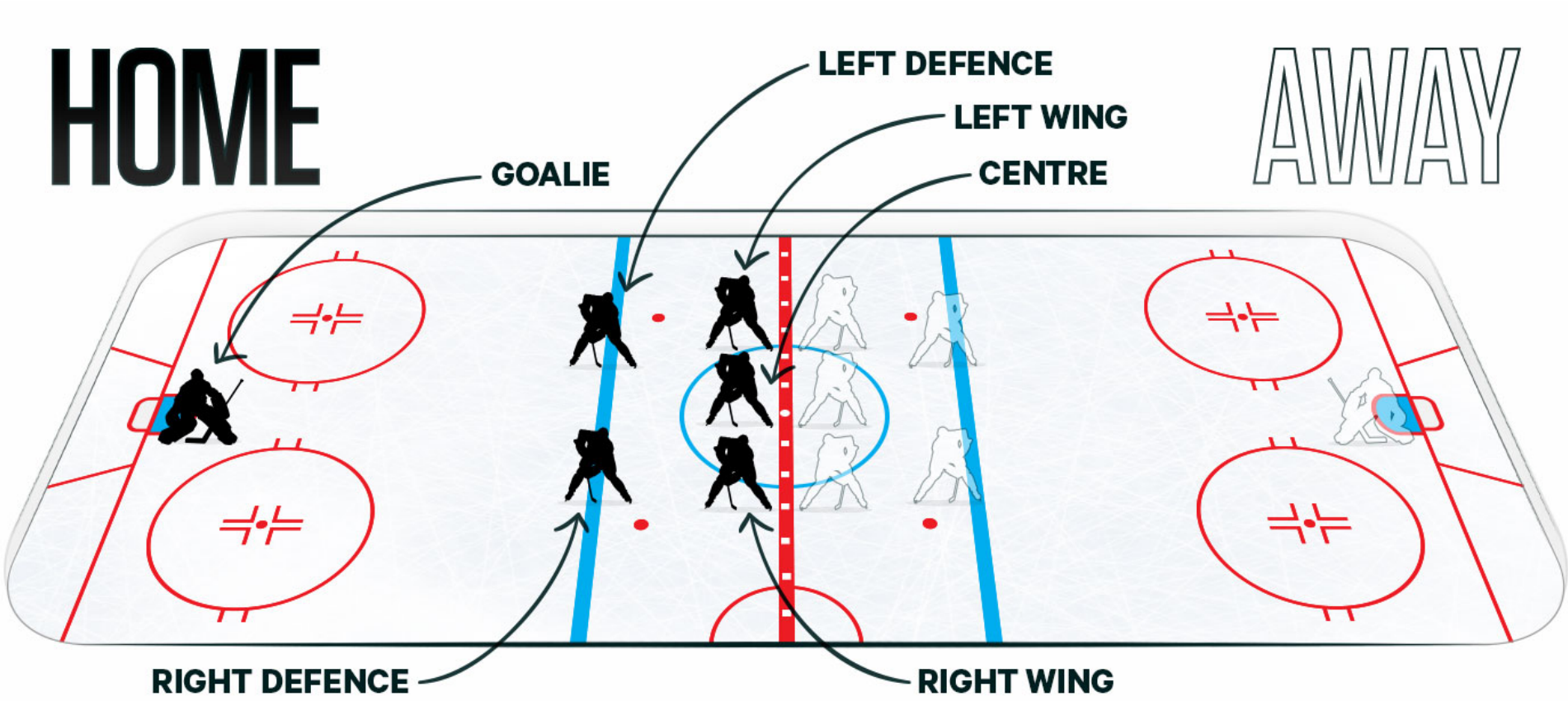
/conferences	Get all current NHL conferences,
/divisions	Get all current NHL divisions,
/draft	Get round-by-round data for current year's NHL Entry Draft,
/game	Get stats for an NHL game,
/people	Get an NHL player,
/schedule	Get the NHL game schedule,
/standings	Get NHL division standings,
/standingsTypes	Get all available NHL standing types,
/statTypes	Get all available NHL statistic types,
/teams	Get all NHL teams.

Introduction

How does NHL game look

- 5 vs 5 game, 4 skaters, 1 goaltender
- 3 periods of 20 minutes
- game that ends with a tie goes into 5 minute sudden-death overtime
- shootouts happen after overtime in 3 round if tie remains, shootout continues until one team fails attempt of other

What are player positions



Introduction

What are player positions

- goalie
- left defence
- right defence
- left wing
- centre
- right wing

Related Work

Other sports Machine Learning approaches

- Matsuzawa, Takehiro. 2017. Using Machine Learning to Predict Future Points in the NHL
- Silver, N. Were predicting the career of every nba player. heres how.

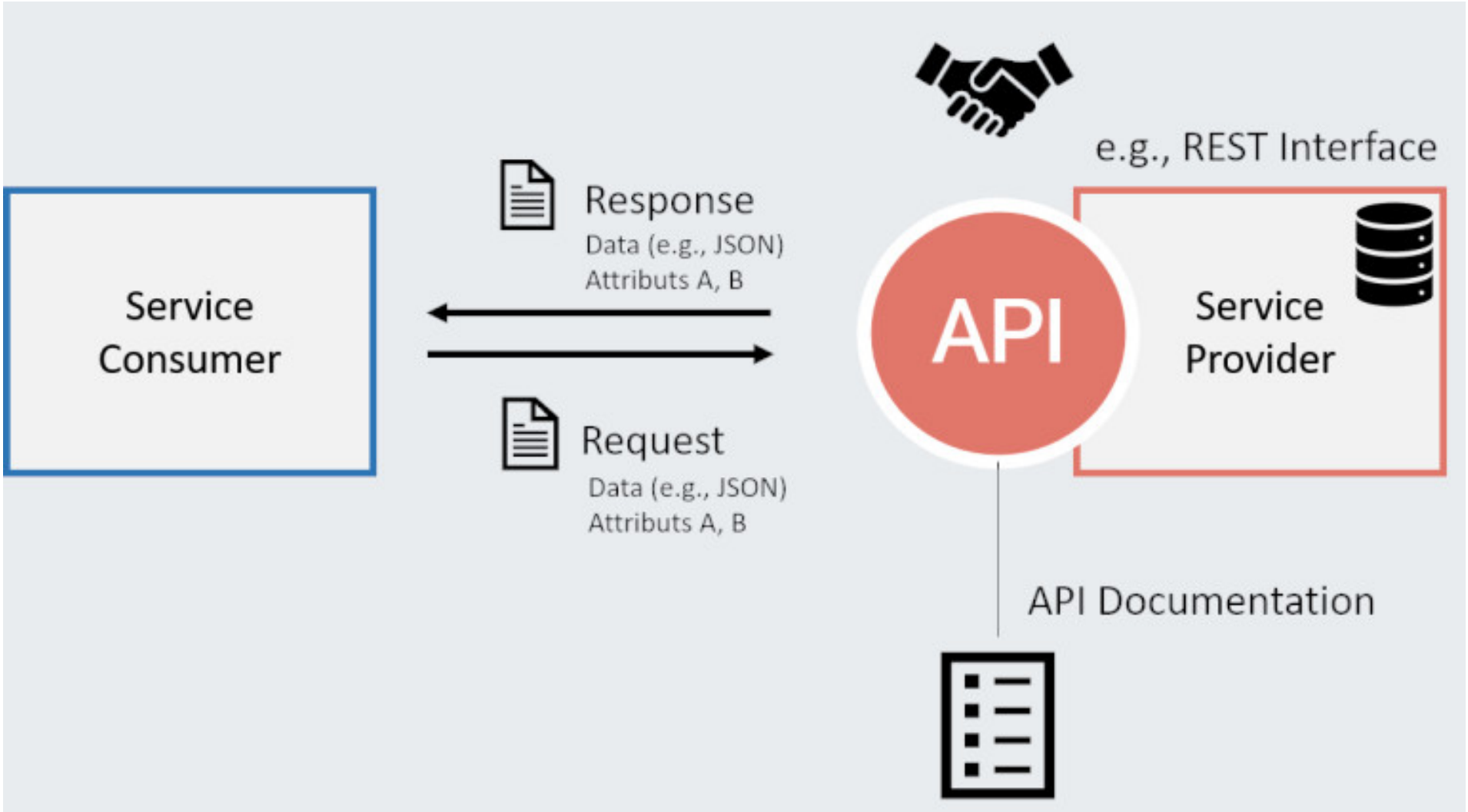
Mining data from NHL stats API

Since API is accessible through Web API for optimization purposes and for easy shaping data set it must be downloaded. For this purpose two python libraries were chosen:

python-requests The requests library is the de facto standard for making HTTP requests in Python. It abstracts the complexities of making requests behind a beautiful, simple API.

python-grequests The GRequests library allows you to use Requests with Gevent to make asynchronous HTTP Requests easily.

Materials and Methods



Storing data set

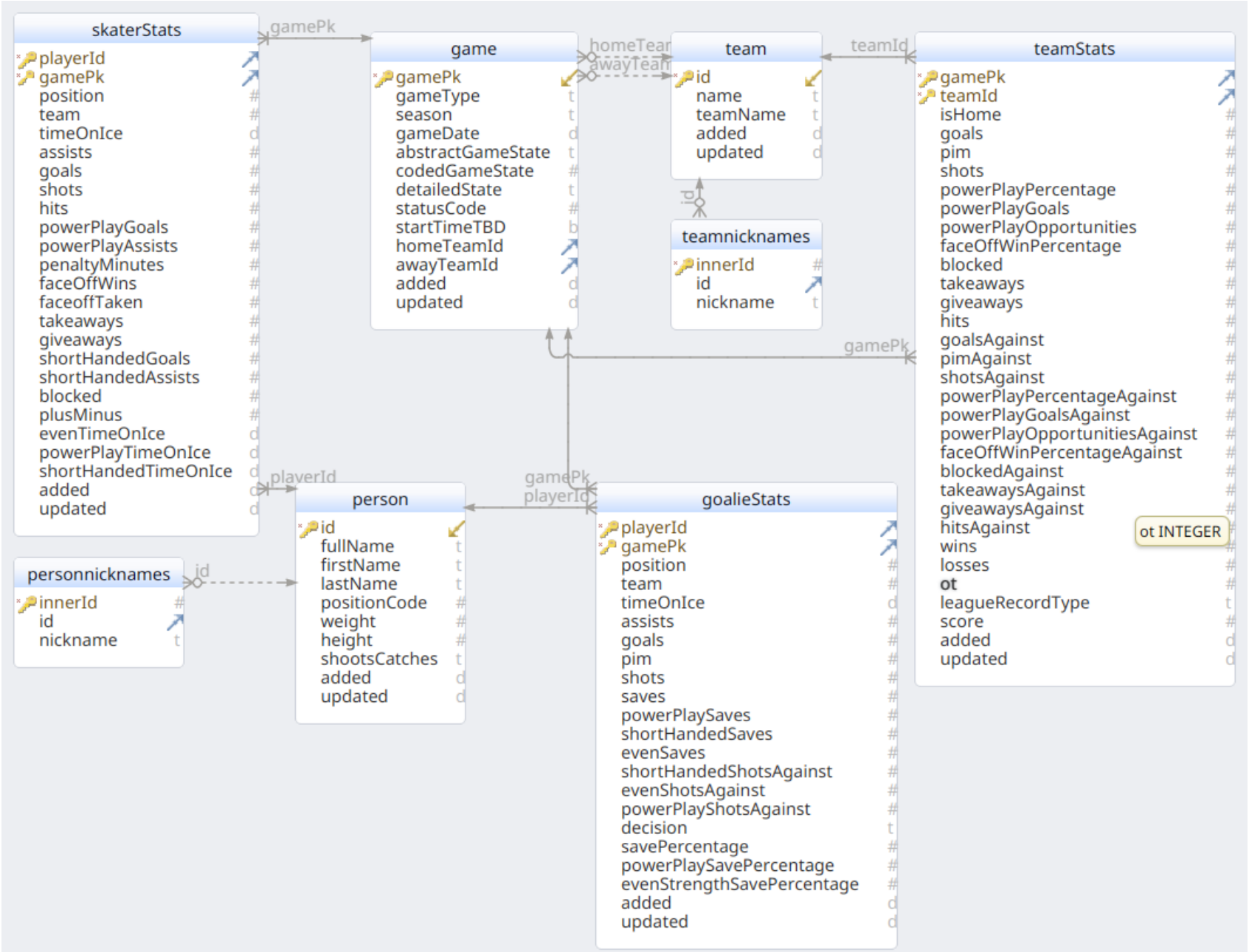
Since this is no trivial dataset with relations between game, teams, players and player stats storing it in CSV or other plain file format would be overkill. That's why SQL format was chosen, which allows:

- no need to write complex python code for parsing and choosing columns to data frame
- easy filtering of data by years, seasons or other categories
- updating data easy as a pie

Storing data set

To be type agnostic another python framework was chosen to design and model database in easy to migrate manner. SQLAlchemy allows for shaping SQL database using python classes. Following schema was proposed:

Materials and Methods



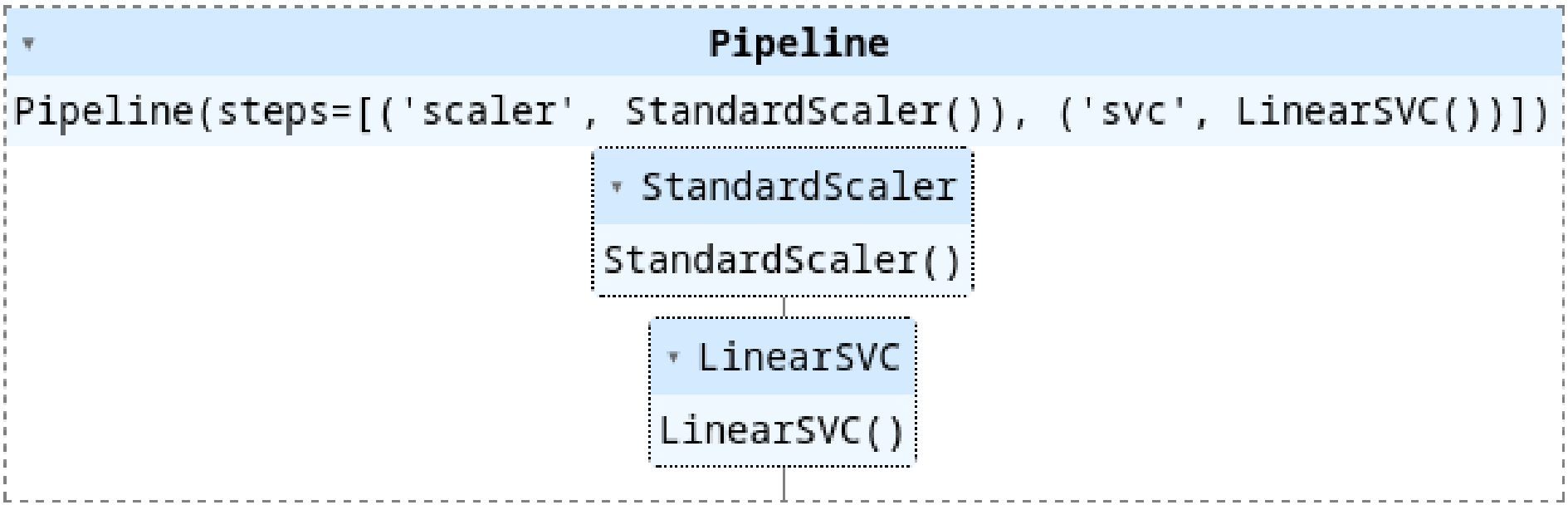
Shaping data frame

From person and skaterStats table following columns were chosen as candidates for data frame:

- person.id
- person.positionCode
- person.weight,
- person.height,
- person.shootsCatches,
- skaterStats.timeOnIce,
- skaterStats.assists,
- skaterStats.goals,
- skaterStats.shots,
- skaterStats.hits,
- skaterStats.powerPlayGoals
- skaterStats.powerPlayAssists,
- skaterStats.penaltyMinutes,
- skaterStats.faceOffWins,
- skaterStats.faceoffTaken,
- skaterStats.takeaways,
- skaterStats.giveaways,
- skaterStats.shortHandedGoals,
- skaterStats.shortHandedAssists,
- skaterStats.blocked,
- skaterStats.plusMinus,
- skaterStats.evenTimeOnIce,
- skaterStats.powerPlayTimeOnIce,
- skaterStats.shortHandedTimeOnIce

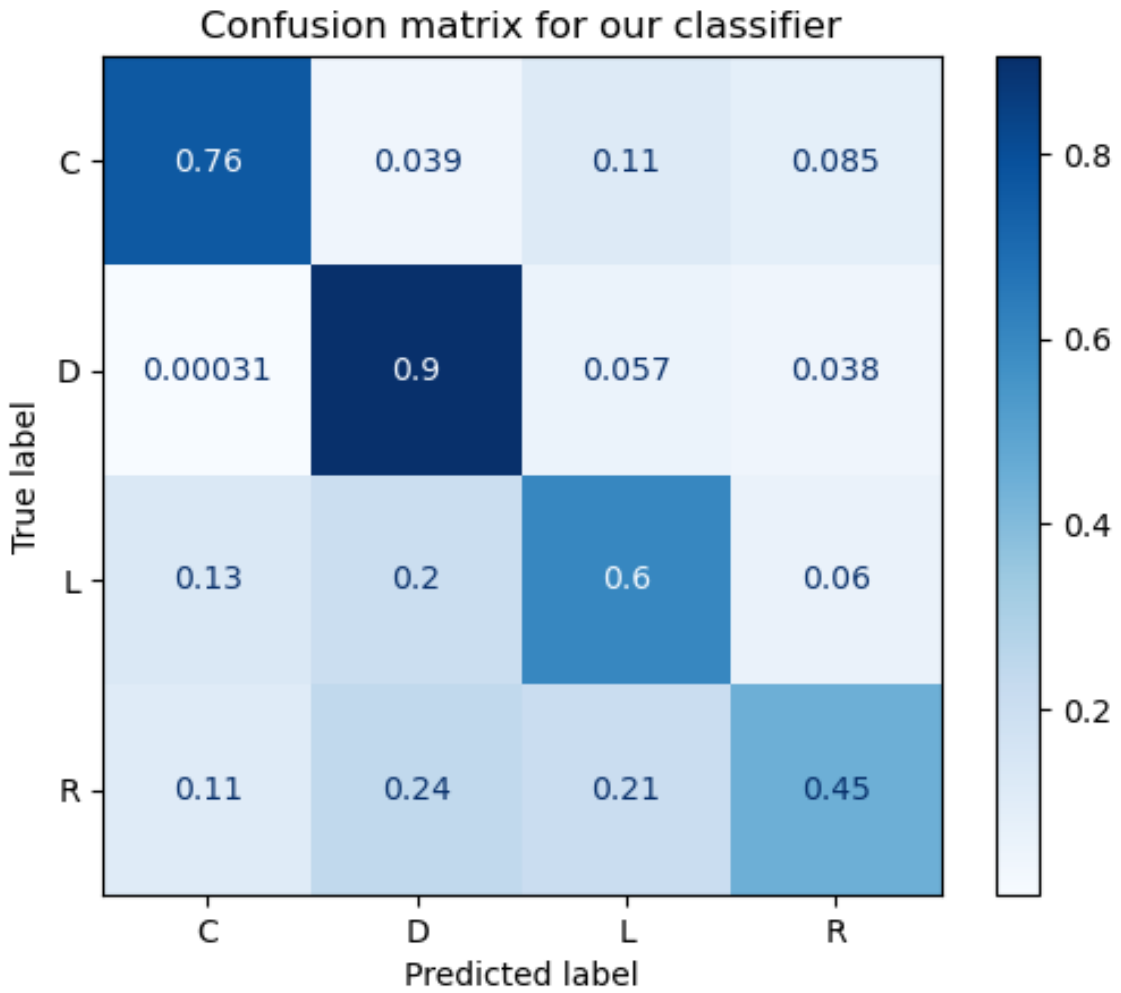
Method of modeling

Scalable Vector Machine was used to o estimate the relationship between a dependent variable and independent variables with standard scaler as standardization of a dataset is a common requirement for many machine learning estimators: they might behave badly if the individual features do not more or less look like standard normally distributed data.



Results

Achieved model accuracy at first attempt without optimizations is 73.52%.



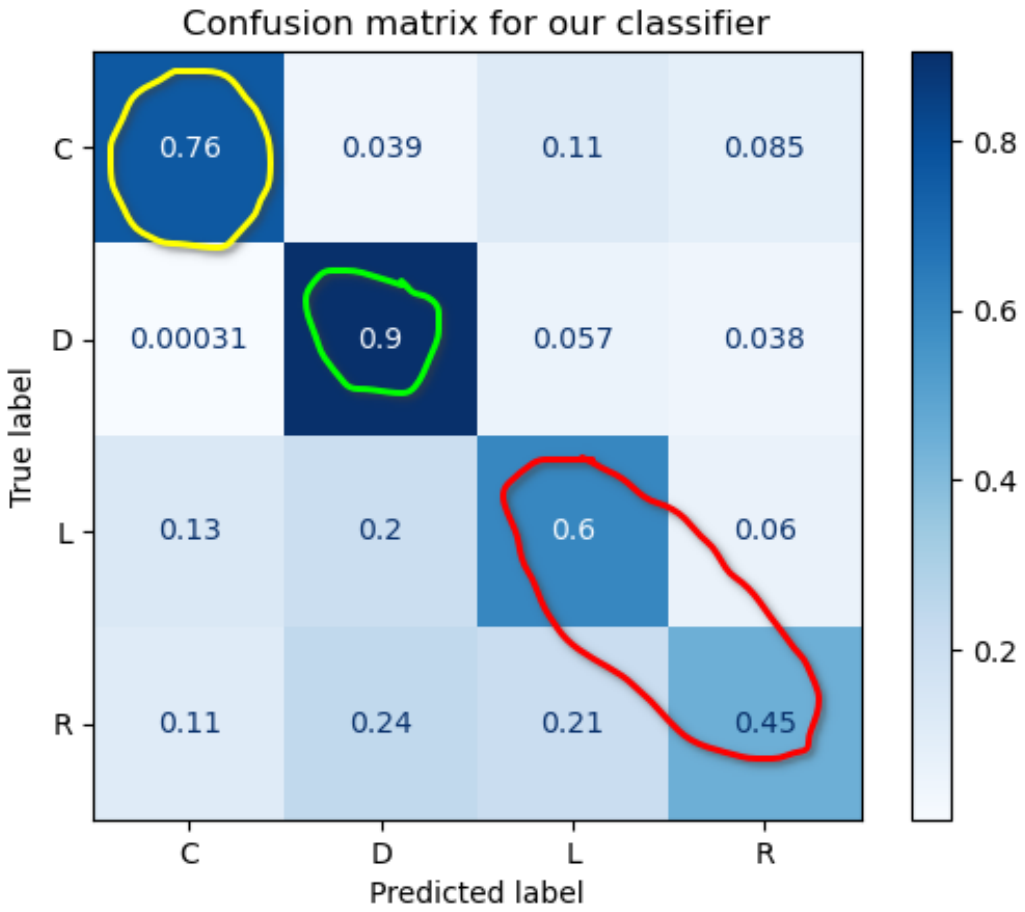
Results and Discussion

Accuracy of prediction per position

- defense position (D) has 90% accuracy
- center position (C) has 76% accuracy
- left wing position (L) has 60% accuracy
- right wing position (R) has 45% accuracy

Optimize

Model optimization should focus on areas marked with yellow and red. The worst case here is predicting right wing player position as the accuracy of it is only 45% so it's probably more accurate to toss the coin.



Conclusion

- Sophisticated model would help NHL teams to figure out which players they should acquire and sell to keep necessary backup for line positions,
- NHL teams could use this model to figure out which players should play on which position to unleash their true potential based on statistics,
- According to Variance Treshold all of the columns are important for position classification,
- Defence position classification is almost certainly true,

References

- Matsuzawa, Takehiro. 2017. Using Machine Learning to Predict Future Points in the NHL
- Matt Eland, Predicting Hockey Penalties with Azure Machine Learning