

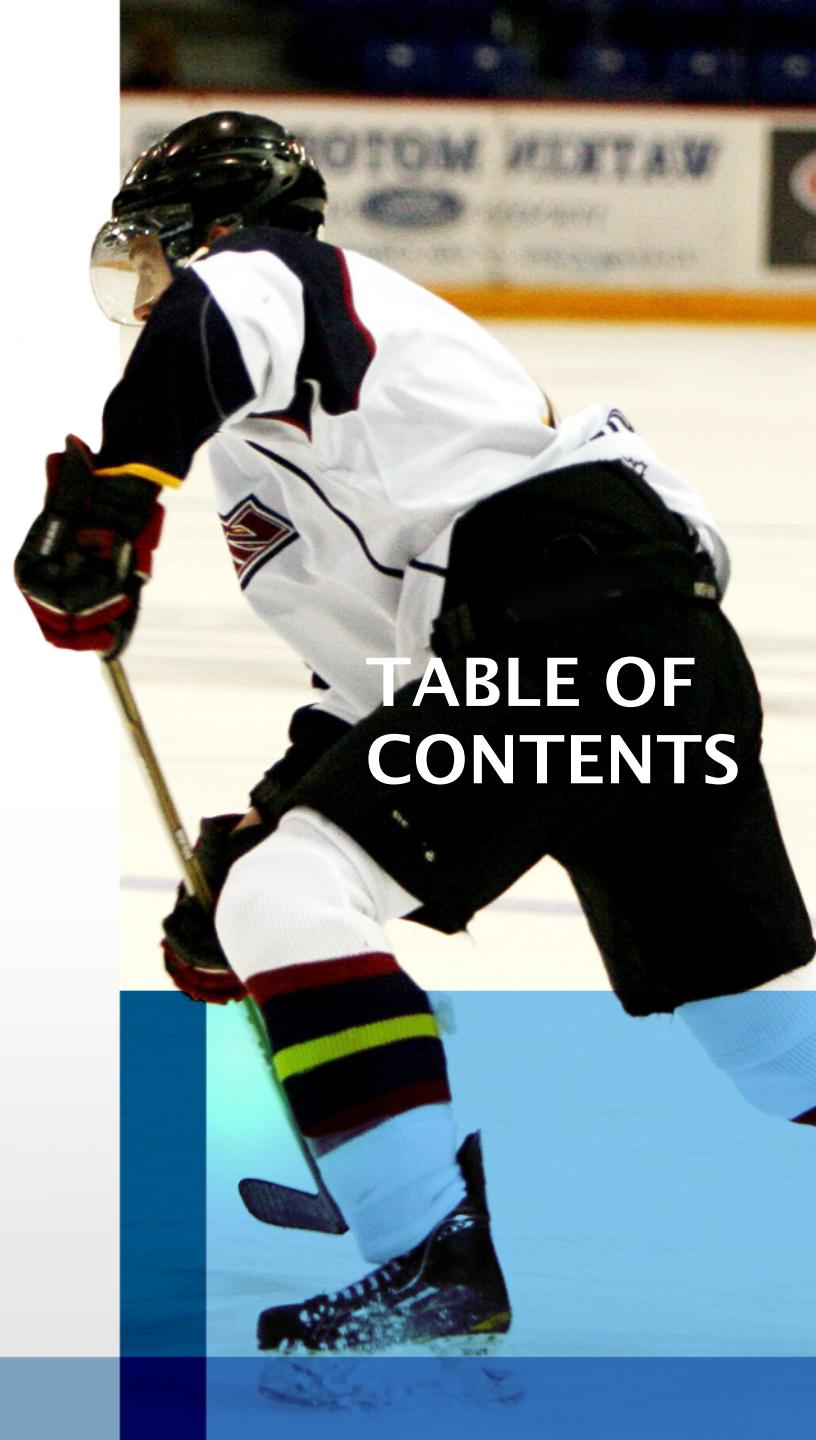


# HOW TO BECOME A NHL ANALYST FROM SCRATCH

IRONHACK FINAL PROJECT - MACHINE LEARNING



- 
- 01 | SCENARIO
  - 02 | WORKFLOW
  - 03 | MODEL COMPARISON
  - 04 | STRUCTURE OF THE LEAGUE
  - 05 | PLAYOFF PREDICTION
  - 06 | DREAM TEAM
  - 07 | PLAYOFF PREDICTION



# SCENARIO



*Could I beat any average NHL fan in a NHL Pool game without having watched a game?*

*Be able to predict which teams are going to make the playoffs*

*Be able to predict the players performance*



## STRUCTURE OF THE LEAGUE

32 TEAMS

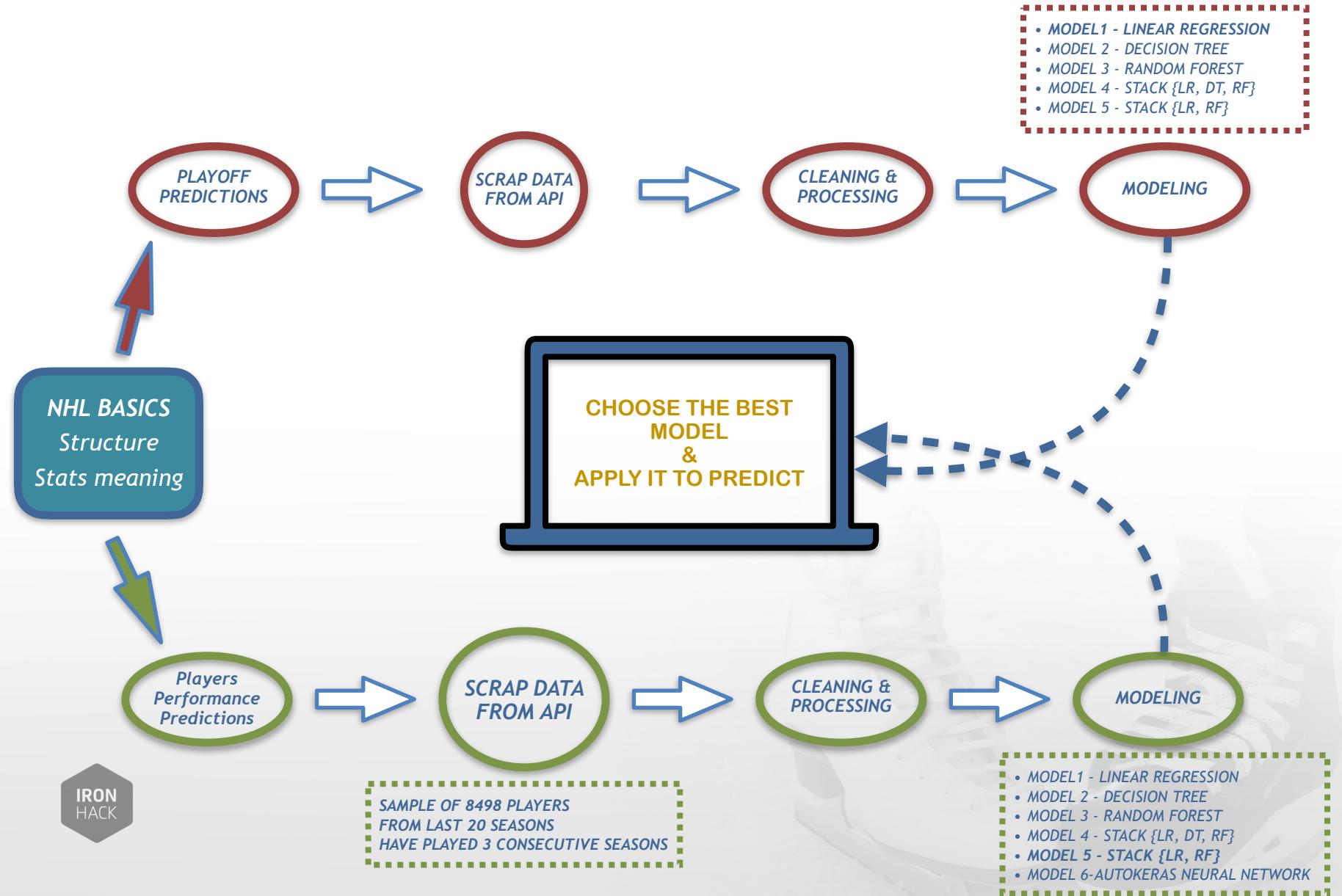
- 20 PLAYERS IN THE ROASTER - 3 MORE TO COVER INJURIES
- 6 PLAYERS IN THE LINE UP
  - 1 CENTER —————→ *Points = Goals + Assists*
  - 2 FORWARDS - LEFT & RIGHT —————→ *Points = Goals + Assists*
  - 2 DEFENSES —————→ *Points = Goals + Assists*
  - 1 GOALIE —————→ *Points = N° Wins*

82 GAMES IN REGULAR SEASON



PLAYOFF WITH THE BEST 16 TEAMS  
IN REGULAR SEASON

# WORKFLOW



# MODEL COMPARISONS



## PLAYOFFS

### MINIMUM SQUARE ERROR

|                                |         |
|--------------------------------|---------|
| • MODEL 1 - LINEAR REGRESSION  | 0.00379 |
| • MODEL 2 - DECISION TREE      | 0.00835 |
| • MODEL 3 - RANDOM FOREST      | 0.00484 |
| • MODEL 4 - STACK {LR, DT, RF} | 0.00381 |
| • MODEL 5 - STACK {LR, RF}     | 0.00381 |

### ROOT MINIMUM SQUARE ERROR

|                                |        |
|--------------------------------|--------|
| • MODEL 1 - LINEAR REGRESSION  | 0.0615 |
| • MODEL 2 - DECISION TREE      | 0.0914 |
| • MODEL 3 - RANDOM FOREST      | 0.0695 |
| • MODEL 4 - STACK {LR, DT, RF} | 0.0617 |
| • MODEL 5 - STACK {LR, RF}     | 0.0617 |

MY MODEL IS WRONG BY JUST AROUND 0.3 POINTS PER SEASON

## PLAYERS PERFORMANCES

### MINIMUM SQUARE ERROR

|                                |        |
|--------------------------------|--------|
| • MODEL 1 - LINEAR REGRESSION  | 0.0235 |
| • MODEL 2 - DECISION TREE      | 0.0490 |
| • MODEL 3 - RANDOM FOREST      | 0.0248 |
| • MODEL 4 - STACK {LR, DT, RF} | 0.0232 |
| • MODEL 5 - STACK {LR, RF}     | 0.0232 |
| • MODEL 6 - NEURAL NETWORK     | 0.0243 |

### ROOT MINIMUM SQUARE ERROR

|                                |        |
|--------------------------------|--------|
| • MODEL 1 - LINEAR REGRESSION  | 0.1533 |
| • MODEL 2 - DECISION TREE      | 0.2215 |
| • MODEL 3 - RANDOM FOREST      | 0.1574 |
| • MODEL 4 - STACK {LR, DT, RF} | 0.1524 |
| • MODEL 5 - STACK {LR, RF}     | 0.1524 |
| • MODEL 6 - NEURAL NETWORK     | 0.1560 |

MY MODEL IS WRONG BY JUST AROUND 1.3 POINTS PER SEASON

## Which teams will make the 2022 PlayOffs?



**TEAM TO BEAT**



COLORADO  
**AVALANCHE**

# Dream Team



# BUSINESS INSIGHTS



AS A TEAM MANAGER

***IN WHICH PLAYERS SHOULD THE TEAM INVEST***

***WHICH TEAMS ARE GOING TO BE OUR DIRECT RIVALS***

FOR FUTURE DATA ANALYSIS JOBS

***WORKING WITH UNKNOWN DATA***

***SCRAPPING DATA FROM API - LIVE DATA***

***WORKING WITH NEURAL NETWORKS***

***ANALYZE AND UNDERSTAND DIFFERENT  
REGRESSION MODELS***



# THANK YOU

# SCENARIO



*I've never watched a single NHL game in my life and I know absolutely nothing about the league. I wanted to develop some models with which I could beat any average NHL fan in a NHL Pool game*

*The main objective of the project is to be able to predict which teams are going to make the playoffs at the end of the current season.*

*Once I achieve that, I will predict the players performance during this season and create my own team capable of winning the league at the end of the season.*

# STRUCTURE OF THE LEAGUE



## 32 TEAMS

- **20 PLAYERS IN THE ROASTER - 3 MORE TO COVER INJURIES**
- **6 PLAYERS IN THE LINE UP**
  - **1 CENTER**
  - **2 FORWARDS - LEFT & RIGHT**
  - **2 DEFENSES**
  - **1 GOALIE**



**82 GAMES IN REGULAR SEASON**



**PLAYOFF WITH THE BEST 16 TEAMS IN REGULAR SEASON**