How to discover the new Frenkie de Jong using data?

Valuing actions in football

Friends of Tracking
May 2020

About us

Lotte Bransen

Lead Data Scientist at SciSports

@LotteBransen

l.bransen@scisports.com

Jan Van Haaren

Chief Product & Technology Officer at SciSports

Research Fellow at KU Leuven

@JanVanHaaren

j.vanhaaren@scisports.com

Scouts are faced with a multitude of questions

What is the player's current and potential ability?

What is the player's expected transfer fee?

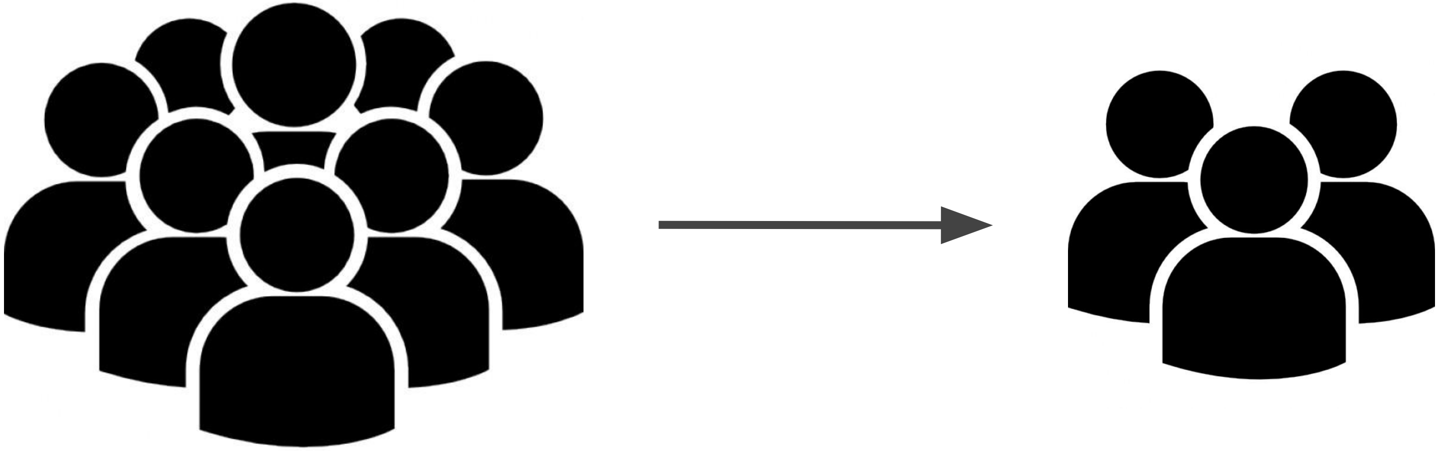
What is the player's playing style?

How valuable are the player's actions?

Insights from data can help football clubs to recruit suitable replacements

More than 100,000 professional football players around the world

Shortlist of 5 to 10 suitable players



Outline

1. Why go beyond traditional statistics to assess football players?
2. How to assess the performances of football players?
3. What does the VAEP framework have to offer?
4. What will you learn in our tutorials?

Outline

1. **Why go beyond traditional statistics to assess football players?**
2. How to assess the performances of football players?
3. What does the VAEP framework has to offer?
4. What will you learn in our tutorials?

What is the player's current and potential ability?

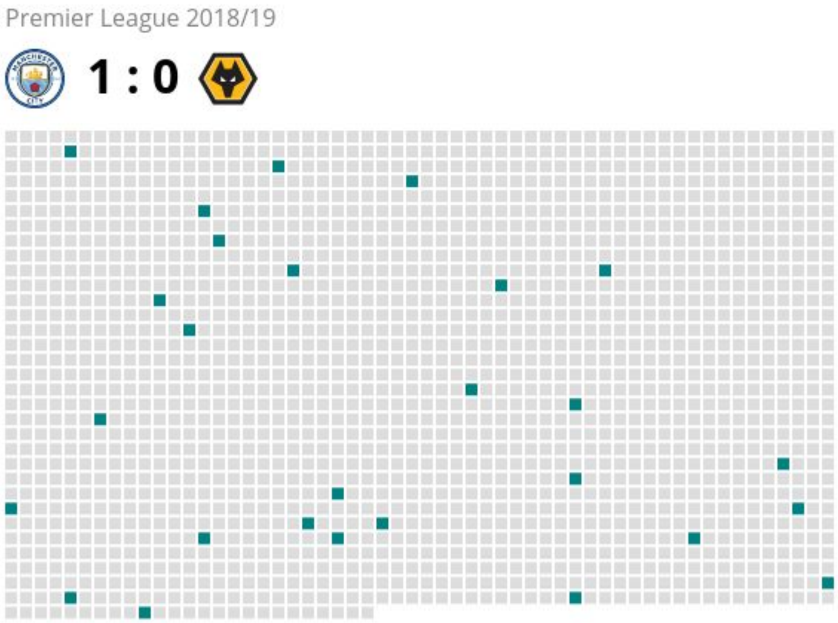
What is the player's expected transfer fee?

What is the player's playing style?

How valuable are the player's actions?

Expected goals models quantify the quality of a goal-scoring opportunity

Fewer than 1% of the on-the-ball actions in a football match are shots



Visualization by Pieter Robberechts

Traditional statistics fail to consider the context

% ACCURATE EXPAND

1	J. Stones, 25 Manchester City	94.82
2	A. Laporte, 25 Manchester City	94.2
3	N. Otamendi, 32 Manchester City	94.19
4	Eric García, 19 Manchester City	94.07
5	Rodri Hernández, 23 Manchester City	93.64

Passing accuracy

79.7%


MCI
0-0
CHE
25:26
SPORT1
TECNO
mobile


44:38 BAR 1 1 VIL
FOX HD
QATAR AIRWAYS
beko.com
KONAMI

Traditional statistics fall short as they fail to account for the context of the actions

Key Pass? **No**
Assist? **No**

Long ball from Fàbregas to Hazard

Key Pass? **Yes**
Assist? **Yes**

Short sideward pass from Busquets to Messi

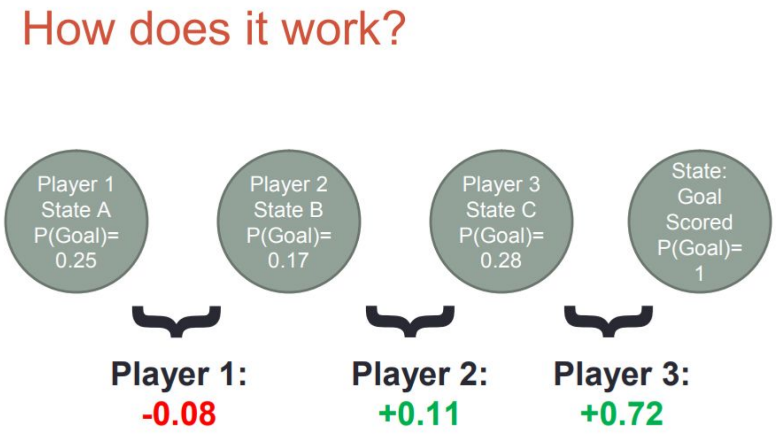
Outline

1. Why go beyond traditional statistics to assess football players?
2. **How to assess the performances of football players?**
3. What does the VAEP framework have to offer?
4. What will you learn in our tutorials?

In 2011, Sarah Rudd first proposed a framework to value each player action


A Framework for Tactical Analysis and Individual Offensive Production Assessment in Soccer Using Markov Chains

Sarah Rudd
On Football Research and Consulting



Several action value frameworks have appeared

Actions Speak Louder Than Goals: Valuing Player Actions in Soccer

Tom Decroos[1], Lotte Bransen[2], Jan Van Haaren[2], and Jesse Davis[1]

[1]KU Leuven, {tom.decroos, jesse.davis}@cs.kuleuven.be
[2]SciSports, {l.bransen, j.vanhaaren}@scisports.com

February 21, 2018

karun.in/blog

Introducing Expected Threat (xT)

Modelling team behaviour in possession to gain a deeper understanding of buildup play.

- Karun Singh (@karun1710)

Decomposing the Immeasurable Sport: A deep learning expected possession value framework for soccer

Javier Fernández
F.C. Barcelona
javier.fernandezr@fcbarcelona.cat

Luke Bornn
Simon Fraser University, Sacramento Kings
lbornn@sfu.ca

Dan Cervone
Los Angeles Dodgers
dcervone@gmail.com

Attacking Contributions: Markov Models for Football

By Derrick Yam | February 21, 2019 | StatsBomb Labs

03.10.19

BLOG: Introducing a Possession Value framework

Article by Nils Mackay

Share

AMERICAN SOCCER ANALYSIS

GOALS ADDED: INTRODUCING A NEW WAY TO MEASURE SOCCER

May 04, 2020

These frameworks differ in several aspects

- Hand-crafted vs data-driven
- Offensive contribution vs offensive + defensive contribution
- Pitch location vs more expressive game states
- Event data vs tracking data

Maaike Van Roy, Pieter Robberechts, Tom Decroos, Jesse Davis. **Valuing On-the-Ball Actions in Soccer: A Critical Comparison of xT and VAEP.** *AAAI 2020 Workshop on Artificial Intelligence in Team Sports, 2020.*

These frameworks differ in several aspects

- Hand-crafted vs **data-driven**
- Offensive contribution vs **offensive + defensive contribution**
- Pitch location vs **more expressive game states**
- **Event data** vs tracking data

Maaike Van Roy, Pieter Robberechts, Tom Decroos, Jesse Davis. **Valuing On-the-Ball Actions in Soccer: A Critical Comparison of xT and VAEP.** *AAAI 2020 Workshop on Artificial Intelligence in Team Sports, 2020.*

Outline

1. Why go beyond traditional statistics to assess football players?
2. How to assess the performances of football players?
3. **What does the VAEP framework have to offer?**
4. What will you learn in our tutorials?

Football data comes in several different flavors

Matchsheet data

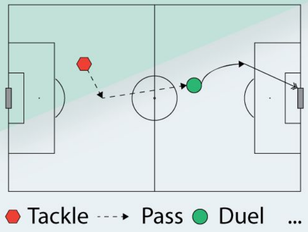
Line-ups, substitutions, goals, cards,...

Brazil 1
Belgium 2

Goals: 13’ Fernandinho (OG) 0-1, 31’ De Bruyne 0-2, 76’ Renato Augusto 1-2
Brazil: Alisson, Fagner, Silva, Miranda, Marcelo, Fernandinho, Paulinho (73’ Renato Augusto), Coutinho, Willian (46‘ Firminho), Neymar, Jesus (58’ Costa)
Belgium: Courtois, Meunier, Alderweireld, Kompany, Verthongen, Witsel, Fellaini, Chadli (83’ Vermaelen), De Bruyne, Hazard, Lukaku (87’ Tielemans)
Yellow cards: 47’ Alderweireld, 71’ Meunier, 85‘ Fernandinho, 90’ Fagner
Red cards: None

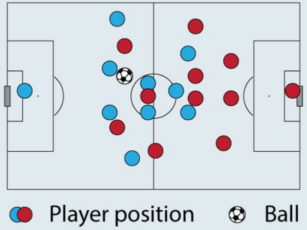
Ball event data

Event type, timestamp, spatial location and meta information of on the ball actions



Tracking data

Cameras capture positions of all players and the ball at all times



Visualization by Pieter Robberechts

Availability decreases as granularity increases

High availability

Freely available for all professional matches

Commercially available for professional matches

Proprietary, available for a single team or teams within the same league

Low availability

Limited availability

Matchsheet data

Line-ups, substitutions, goals, cards,...

Brazil	**1**
Belgium	**2**

Goals: 13' Fernandinho (OG) 0-1, 31' De Bruyne 0-2, 76' Renato Augusto 1-2

Brazil: Alisson, Fagner, Silva, Miranda, Marcelo, Fernandinho, Paulinho (73' Renato Augusto), Coutinho, Willian (46' Firminho), Neymar, Jesus (58' Costa)

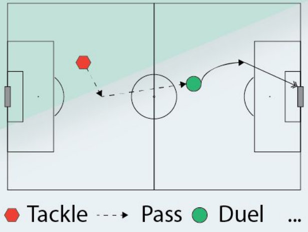
Belgium: Courtois, Meunier, Alderweireld, Kompany, Verthongen, Witsel, Fellaini, Chadli (83' Vermaelen), De Bruyne, Hazard, Lukaku (87' Tielemans)

Yellow cards: 47' Alderweireld, 71' Meunier, 85' Fernandinho, 90' Fagner

Red cards: None

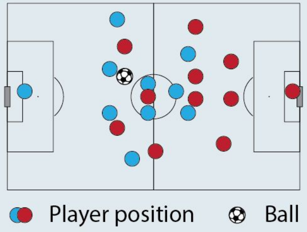
Ball event data

Event type, timestamp, spatial location and meta information of on the ball actions



Tracking data

Cameras capture positions of all players and the ball at all times



Low granularity

Limited granularlity

High-level summary

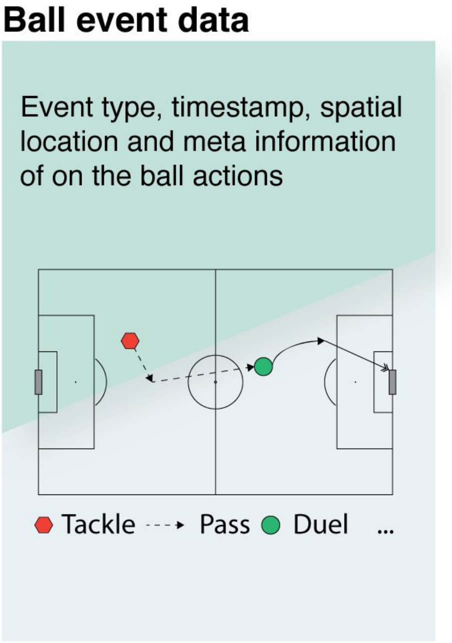
Spatio-temporal description of all on the ball events

Exact spatial movements of all players and the ball

High granularity

Visualization by Pieter Robberechts

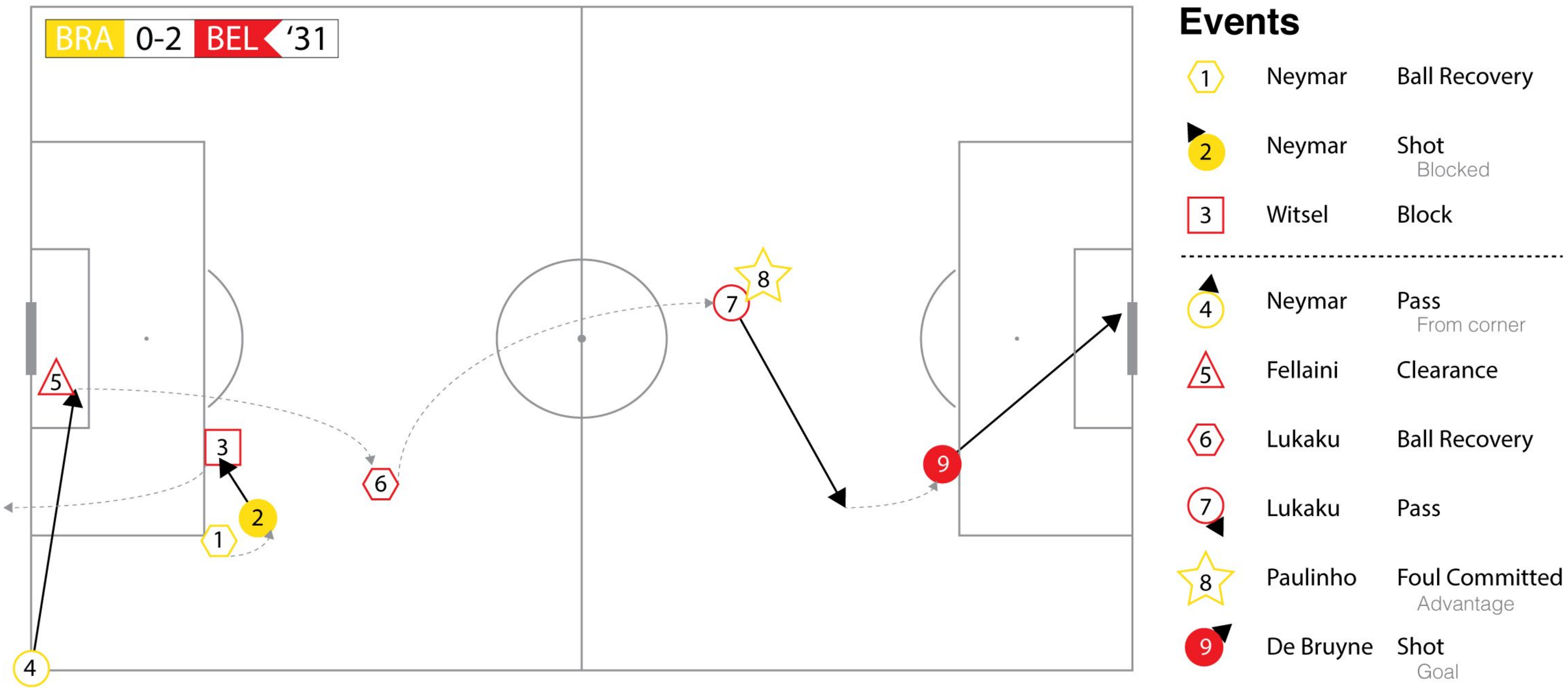
Event data is useful in the recruitment process



- ✔ Widely available for hundreds of competitions
- ✔ Increasingly information rich
- ✔ Easier to process and analyze than player tracking data

Visualization by Pieter Robberechts

VAEP uses event data to value actions



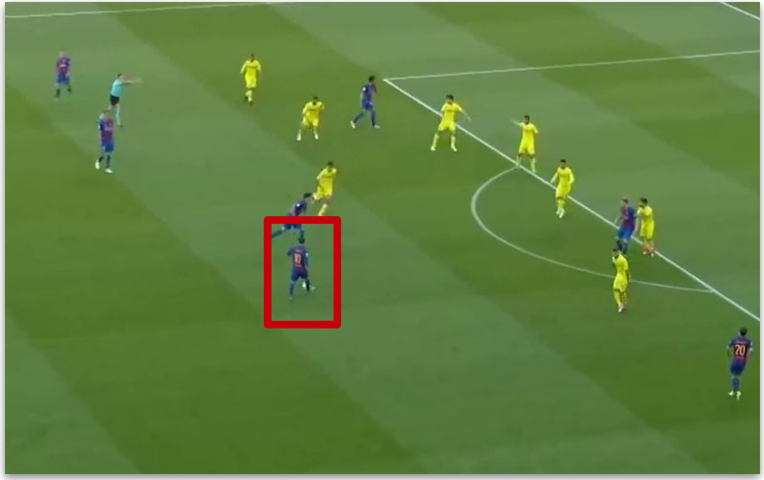
Visualization by Pieter Robberechts

Compute the difference between the values of the pre-action and post-action game states

Pre-action game state

How valuable is this game state?

Post-action game state



How valuable is this game state?

How valuable is this game state?

Pre-action game state

- How likely is it that the team in possession will **score** a goal from this game state?
- How likely is it that the team in possession will **concede** a goal from this game state?

Determine the likeliness of **scoring** a goal

How likely is it to **score** a goal from this game state?

Retrieve a large number of highly similar game states in earlier matches

Determine the likeliness of **scoring** a goal

How likely is it to **score** a goal from this game state?

Retrieve a large number of highly similar game states in earlier matches **using event data**

Determine the likeliness of **scoring** a goal

How likely is it to **score** a goal from this game state?

0.03

"**3** out of each 100 similar game states in the past led to a goal"

Determine the likeliness of **conceding** a goal

How likely is it to **concede** a goal from this game state?

0.02

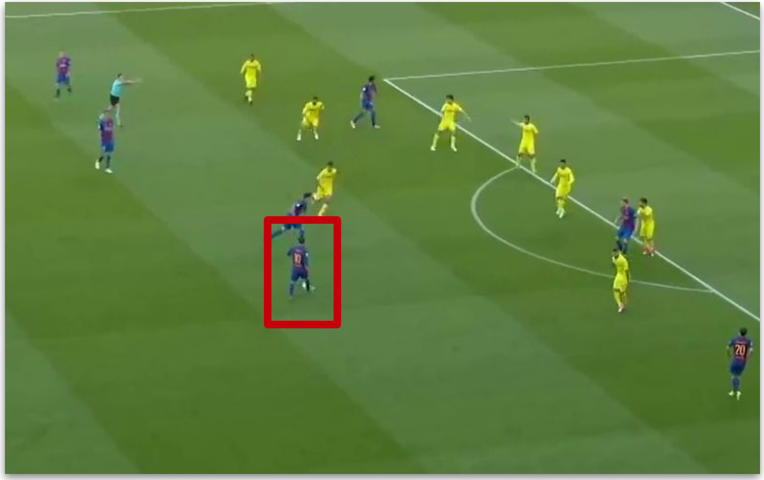
"**2** out of each 100 similar game states in the past led to a goal"

Compute the difference between the values of the pre-action and post-action game states

Pre-action game state

How valuable is this game state?

Post-action game state



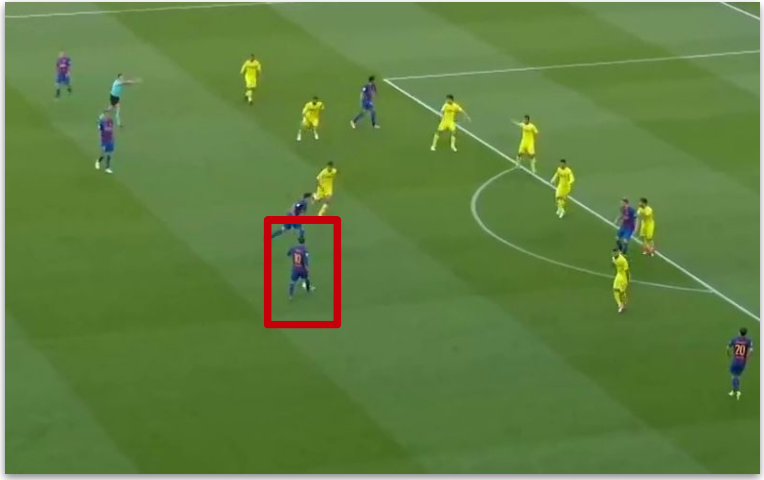
How valuable is this game state?

Compute the difference between the values of the pre-action and post-action game states

Pre-action game state

Likeliness to **score** a goal?
Likeliness to **concede** a goal?

Post-action game state



Likeliness to **score** a goal?
Likeliness to **concede** a goal?

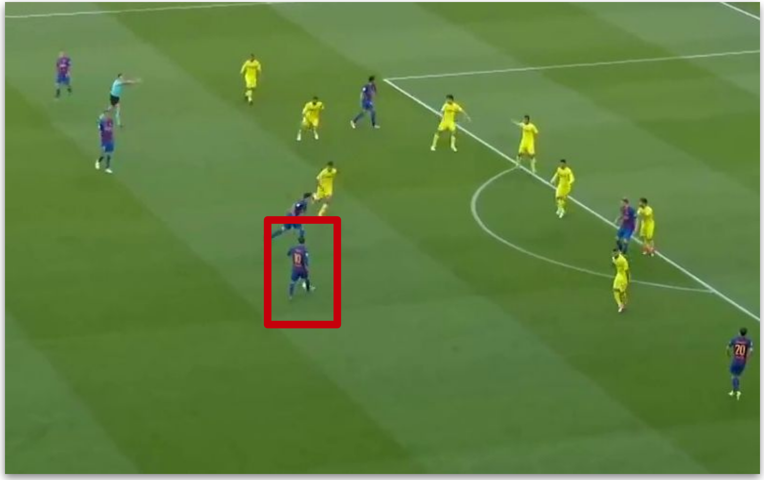
Compute the difference between the values of the pre-action and post-action game states

Pre-action game state

Likeliness to **score** a goal = 0.03

Likeliness to **concede** a goal = 0.02

Post-action game state



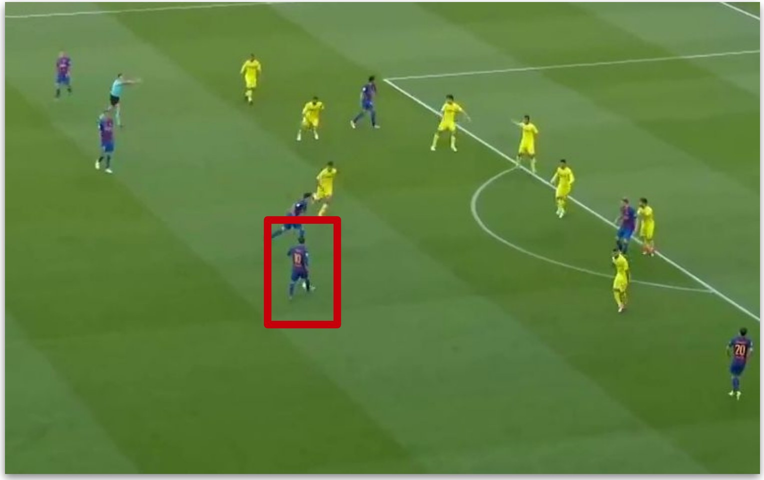
Likeliness to **score** a goal = 0.05

Likeliness to **concede** a goal = 0.01

Compute the difference between the values of the pre-action and post-action game states

Pre-action game state

Post-action game state



Likeliness to **score** a goal = 0.03
Likeliness to **concede** a goal = 0.02

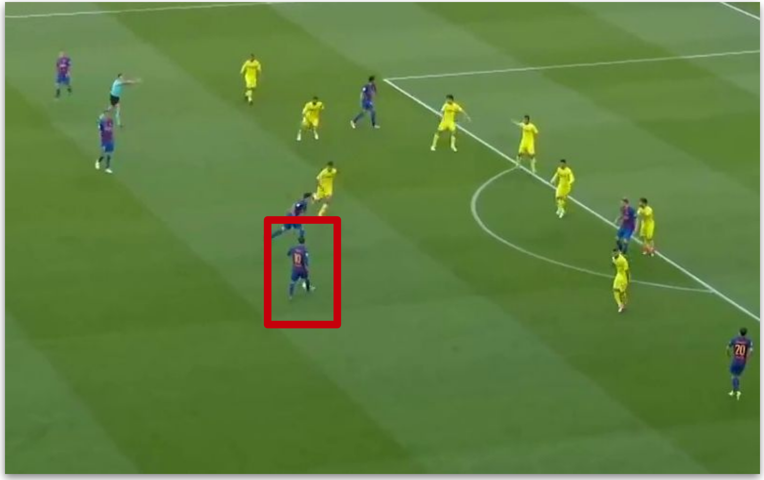
Score increase = 0.02
Concede decrease = 0.01

Likeliness to **score** a goal = 0.05
Likeliness to **concede** a goal = 0.01

Compute the difference between the values of the pre-action and post-action game states

Pre-action game state

Post-action game state



Likeliness to **score** a goal = 0.03
Likeliness to **concede** a goal = 0.02

Score increase = 0.02
Concede decrease = 0.01
Contribution rating = **0.03**

Likeliness to **score** a goal = 0.05
Likeliness to **concede** a goal = 0.01

Describe a game state as a sequence of 3 events

Game as sequence of ~1600 events

…	**Dribble Neymar**	**Duel Neymar**	**Pass Neymar**	**Pass Busquets**	**Take-on Messi**	**Shot Messi**	…

Describe a game state as a sequence of 3 events

Game as sequence of ~1600 events

…	Dribble Neymar	Duel Neymar	Pass Neymar	**Pass Busquets**	Take-on Messi	Shot Messi	…

Pre-action game state as subsequence of 3 events

	Dribble Neymar	**Duel Neymar**	**Pass Neymar**				

Describe a game state as a sequence of 3 events

Game as sequence of ~1600 events

...	Dribble Neymar	Duel Neymar	Pass Neymar	Pass Busquets	Take-on Messi	Shot Messi	...

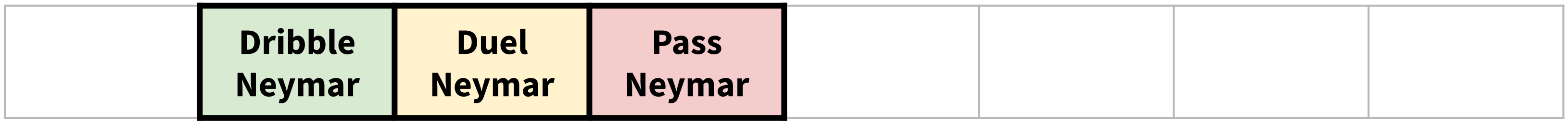
Pre-action game state as subsequence of 3 events

	Dribble Neymar	Duel Neymar	Pass Neymar				

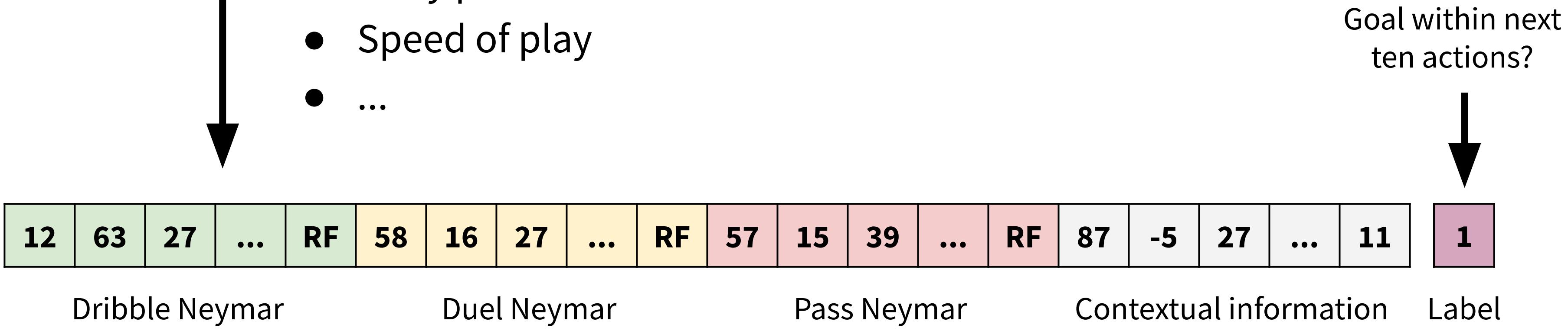
Post-action game state as subsequence of 3 events

		Duel Neymar	Pass Neymar	Pass Busquets			

Describe a game state as a feature vector



- Start and end location
- Type of event
- Body part used
- Speed of play
- ...



Predict whether a game state will yield a goal

Action 1	12	• • •	97	• • •	55
	...		...		...
	RF		H		RF
Action 2	58		63		38
	...		...		...
	RF		RF		RF
Action 3	57	• • •	82	• • •	49
	...		...		...
	RF		LF		LF
Context	87		96		68
	...		...		...
	11		37		24
Label	1	• • •	0	• • •	0

Game states in earlier matches and their labels

Predict whether a game state will yield a goal

Action 1	12	• • •	97	• • •	55
	...		...		...
	RF		H		RF
Action 2	58		63		38
	...		...		...
	RF		RF		RF
Action 3	57	• • •	82	• • •	49
	...		...		...
	RF		LF		LF
Context	87		96		68
	...		...		...
	11		37		24
Label	1	• • •	0	• • •	0

Game states in earlier matches and their labels

Probabilistic classifier

Predict whether a game state will yield a goal

	12	• • •	97	• • •	55
Action 1	...		...		...
	RF		H		RF
	58		63		38
Action 2	...		...		...
	RF		RF		RF
	57	• • •	82	• • •	49
Action 3	...		...		...
	RF		LF		LF
	87		96		68
Context	...		...		...
	11		37		24
Label	1	• • •	0	• • •	0

Game states in earlier matches and their labels

Probabilistic classifier

58	
...	Action 1
RF	
57	
...	Action 2
RF	
67	
...	Action 3
RF	
78	
...	Context
29	
?	Prediction

Game state to be valued

Predict whether a game state will yield a goal

	12	• • •	97	• • •	55
Action 1	...		...		...
	RF		H		RF
	58		63		38
Action 2	...		...		...
	RF		RF		RF
	57	• • •	82	• • •	49
Action 3	...		...		...
	RF		LF		LF
	87		96		68
Context	...		...		...
	11		37		24
Label	1	• • •	0	• • •	0

Game states in earlier matches and their labels

Probabilistic classifier

58	
...	Action 1
RF	
57	
...	Action 2
RF	
67	
...	Action 3
RF	
78	
...	Context
29	
0.03	Prediction

Game state to be valued

Traditional statistics fall short as they fail to account for the context of the actions

Key Pass? **No**
Assist? **No**

Long ball from Fàbregas to Hazard

Key Pass? **Yes**
Assist? **Yes**

Short sideward pass from Busquets to Messi

In contrast, our VAEP Ratings do account for the context of actions

"The Fàbregas pass is about 7 times more valuable than the Busquets pass"

Long ball from Fàbregas to Hazard

Short sideward pass from Busquets to Messi

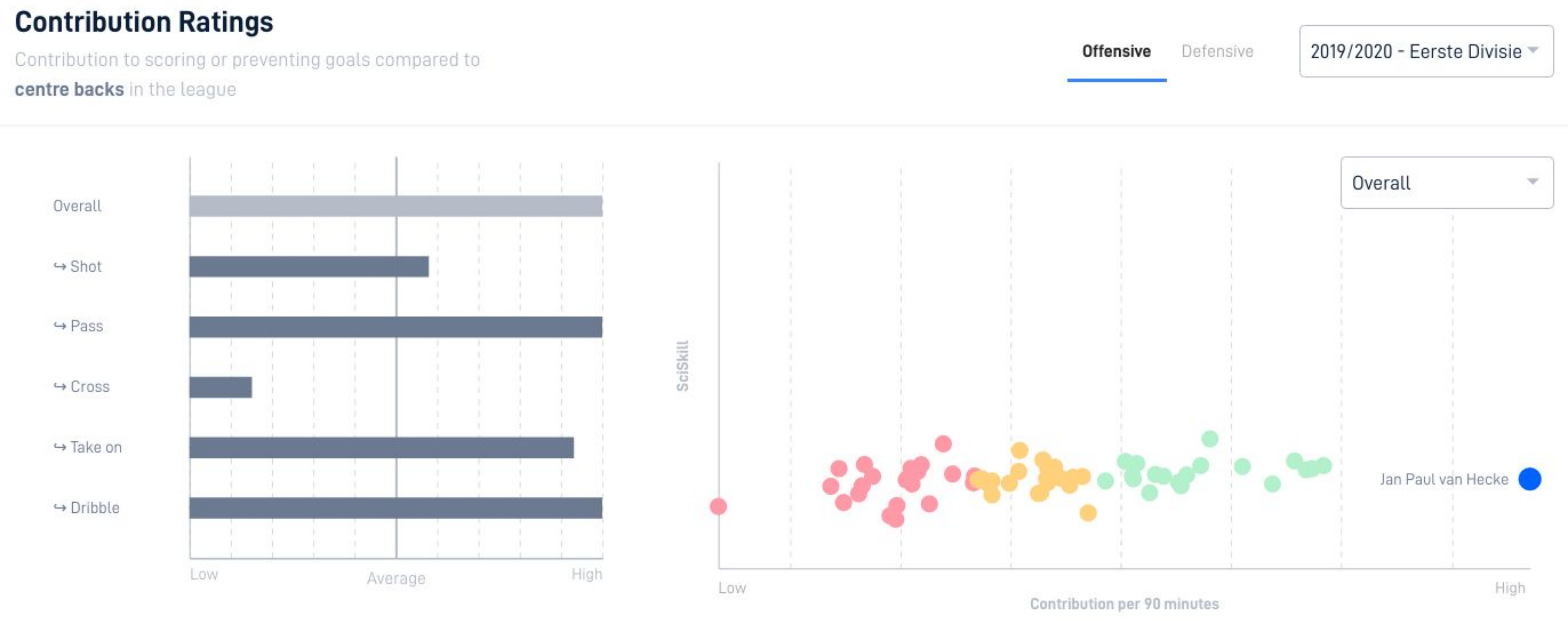
VAEP identifies the most valuable players

Rank	Player	Team	Goals/90	Assists/90	VAEP/90
1	R. Mahrez	Manchester City FC	0.42	0.42	0.89
2	K. De Bruyne	Manchester City FC	0.31	0.59	0.79
3	S. Mané	Liverpool FC	0.57	0.24	0.69
4	S. Agüero	Manchester City FC	0.97	0.12	0.68
5	P. Aubameyang	Arsenal FC	0.62	0	0.62
6	D. Ings	Southampton FC	0.63	0.04	0.62
7	Mohamed Salah	Liverpool FC	0.61	0.08	0.61
8	J. Vardy	Leicester City FC	0.72	0.19	0.59
9	R. Sterling	Manchester City FC	0.46	0.04	0.59
10	Richarlison	Everton FC	0.36	0.07	0.56

Table 1: The top-10 players who played at least 900 minutes in the 2019/2020 English Premier League season up until April 2020 in terms of our VAEP player ratings.

VAEP identifies hidden gems

Who is the new Dutch hidden gem?



Action ratings help answer more questions

- What are the most interesting young players in a given region?
- What are our next opponent's weaknesses?
- How does the player perform in important matches?
- How does the player perform in the last minutes of a match?
- How does the player perform in bad weather circumstances?
- How does the player perform in different tactical systems?

For more information on the VAEP ratings

- Tom Decroos, Lotte Bransen, Jan Van Haaren, Jesse Davis. **Actions Speak Louder than Goals: Valuing Player Actions in Soccer.** *International Conference on Knowledge Discovery and Data Mining 2019 (KDD 2019).* **Best Paper Award!**

- Interactive tool by Pieter Robberechts: https://dtai.cs.kuleuven.be/sports/vaep?toggle=explore.

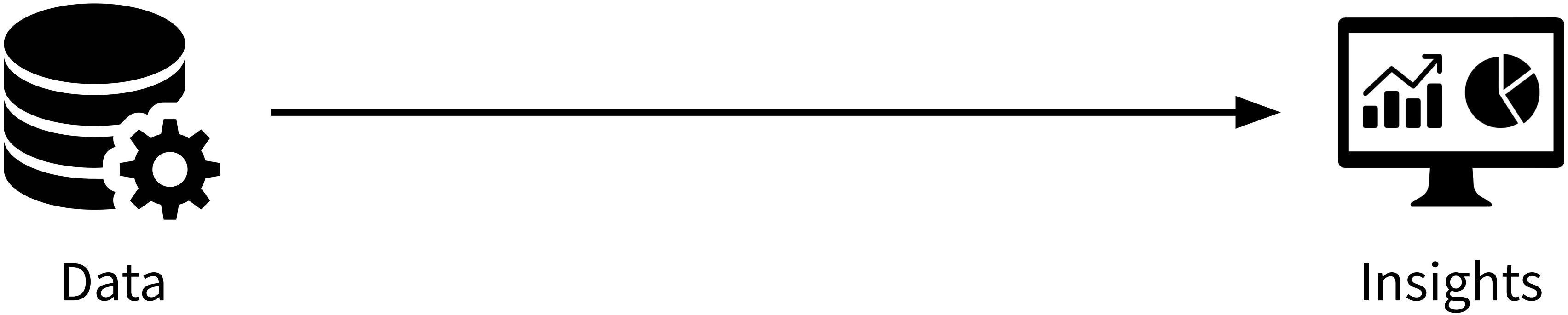
Outline

1. Why go beyond traditional statistics to assess football players?
2. How to assess the performances of football players?
3. What does the VAEP framework have to offer?
4. **What will you learn in our tutorials?**

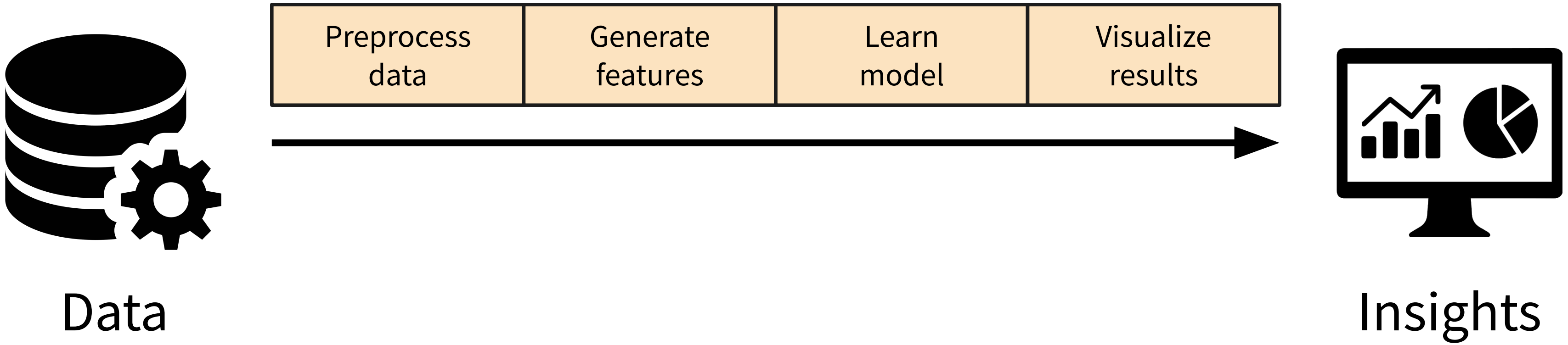
We always build football metrics in an agile way

1. Consult practitioners to genuinely understand their needs
2. Develop a prototype metric
3. Collect feedback from the practitioners
4. Improve the prototype metric

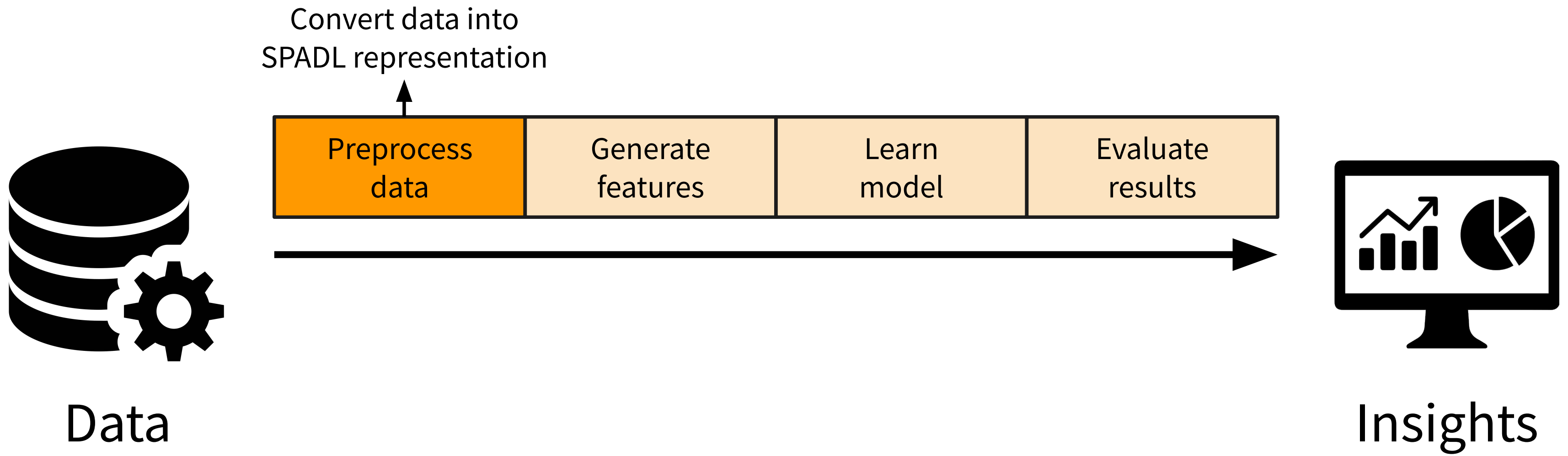
Start with building an end-to-end solution



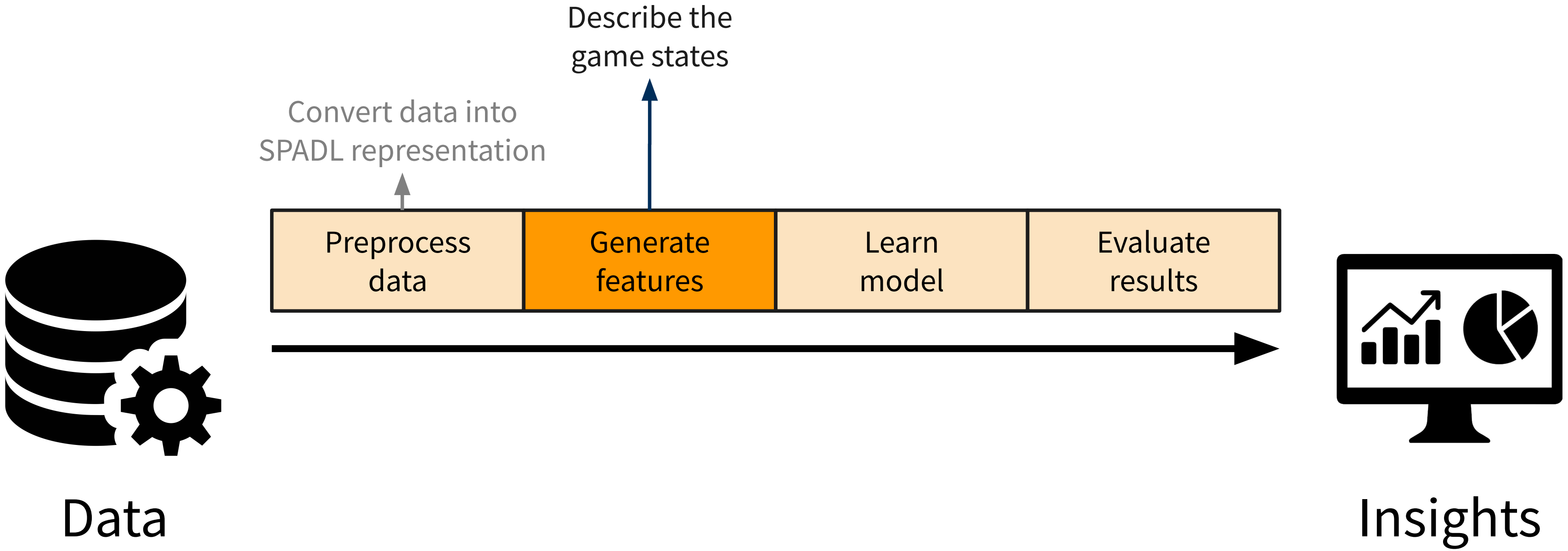
Tutorial 1 builds the end-to-end pipeline



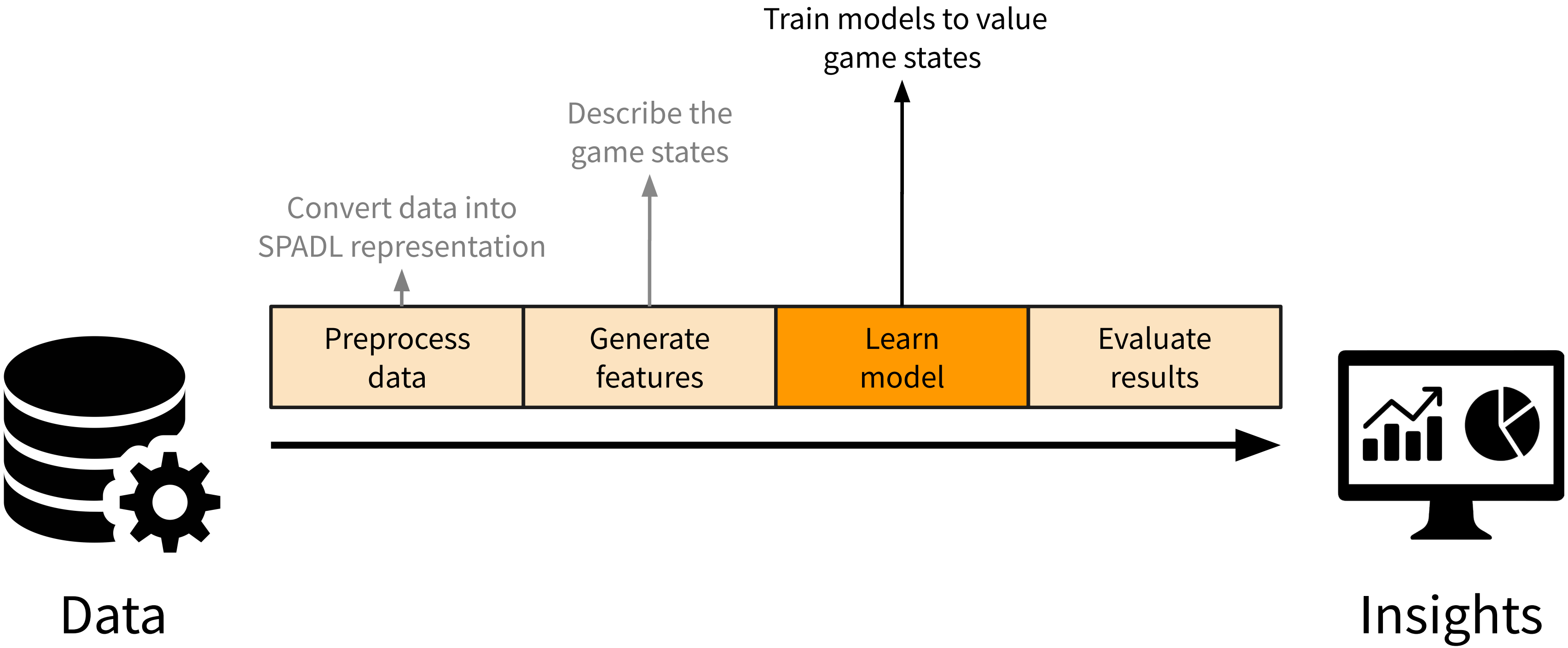
Tutorial 1 also touches upon data preprocessing



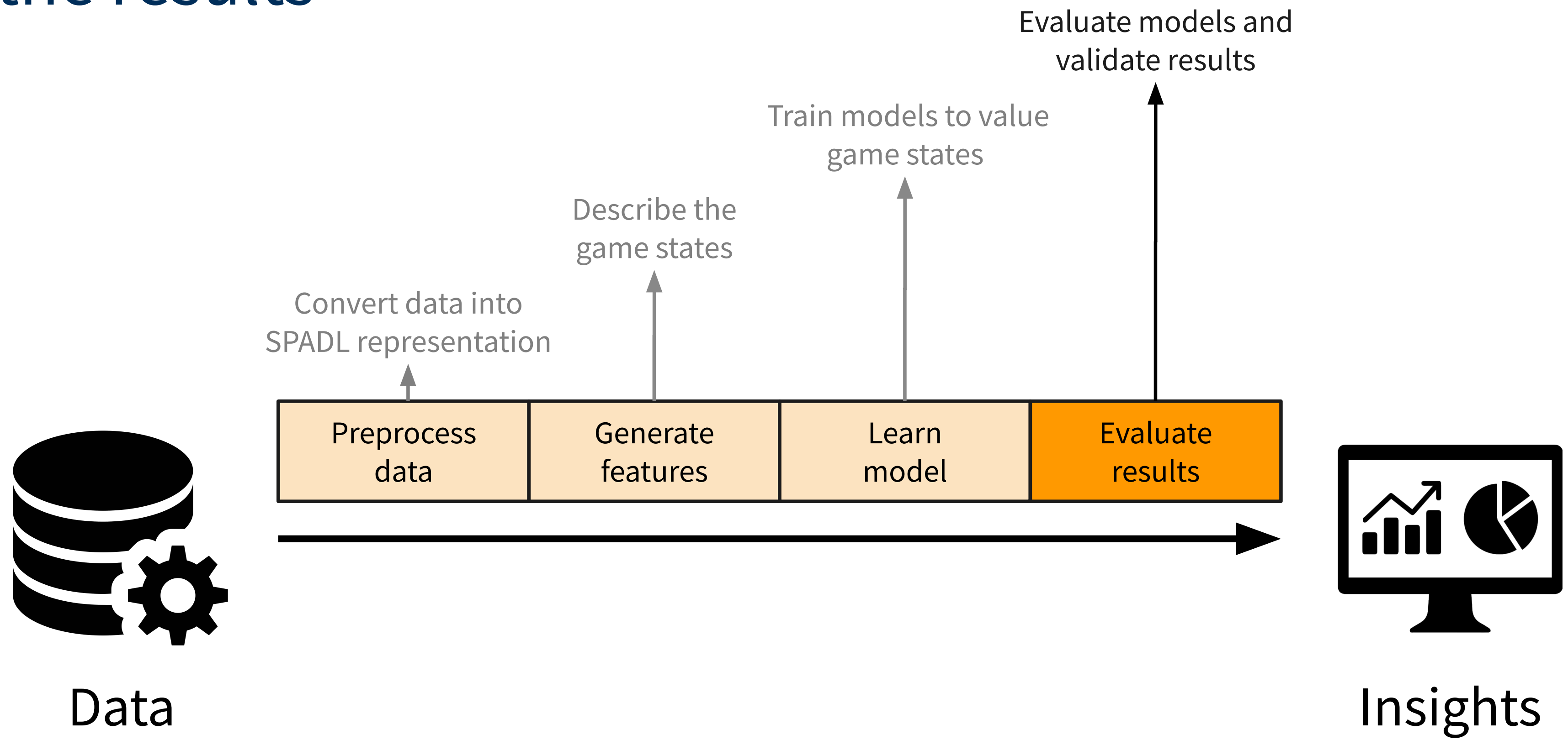
Tutorial 2 shows how to represent the game states and generate features for your model



Tutorial 3 trains machine learning models to value game states and actions



Tutorial 4 explains how to evaluate and visualize the results



We will share tutorials that follow these steps

1. From raw event data to action values - the complete pipeline
2. Feature representation - how to describe game states and actions
3. Valuing game states and actions - how to train your machine learning model
4. Model evaluation and validation - how to evaluate and validate your results

GitHub repository: https://github.com/SciSports-Labs/fot-valuing-actions

Papers and blog posts to check out

- Sarah Rudd. **A Framework for Tactical Analysis and Individual Offensive Production Assessment in Soccer Using Markov Chains.** *http://nessis.org/nessis11/rudd.pdf*
- Karun Singh. **Introducing expected threat.** *https://karun.in/blog/expected-threat.html.*
- Javier Fernández, Luke Bornn, Dan Cervone. **Decomposing the Immeasurable Sport: A deep learning expected possession value framework for soccer.** *MIT Sloan Sports Analytics, 2019.*
- Maaike Van Roy, Pieter Robberechts, Tom Decroos, Jesse Davis. **Valuing On-the-Ball Actions in Soccer: A Critical Comparison of xT and VAEP.** *AAAI 2020 Workshop on Artificial Intelligence in Team Sports, 2020.*
- Derrick Yam. **Attacking Contributions: Markov Models for Football.** *https://statsbomb.com/2019/02/attacking-contributions-markov-models-for-football/*
- Nils Mackay. **BLOG: Introducing a Possession Value framework.** *https://www.optasportspro.com/news-analysis/blog-introducing-a-possession-value-framework/*
- Aditya Kothari. **xPo.** *https://thecomeonman.github.io/xPo/*
- John Muller. **Goals added: introducing a new way to measure soccer.** *https://www.americansocceranalysis.com/home/2020/4/22/37ucr0d5urxxtryn2cfhzormdziphq*

Questions or comments? Please get in touch!

Lotte Bransen

Lead Data Scientist at SciSports

@LotteBransen

l.bransen@scisports.com

Jan Van Haaren

Chief Product & Technology Officer at SciSports

Research Fellow at KU Leuven

@JanVanHaaren

j.vanhaaren@scisports.com