

Sports Analytics - une promenade dans le pays des statistiques de sport

Christophe Ley

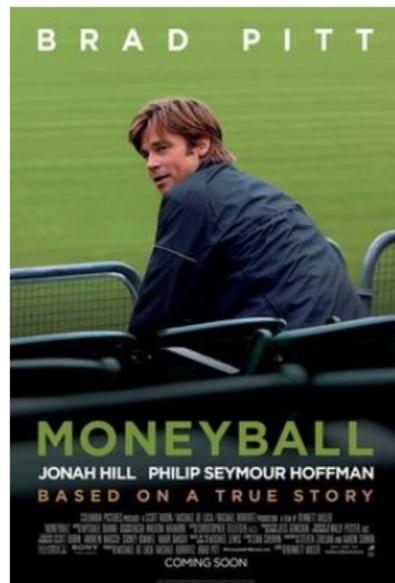
Université du Luxembourg

Académie Lorraine des Sciences
12 Mars 2025



- 1 La science des données à la rencontre du sport
- 2 Médaille d'or : analyses et tactiques
- 3 Médaille d'argent : prédictions et paris sportifs
- 4 Médaille de bronze : prévention de blessures

L'avènement des sports analytics



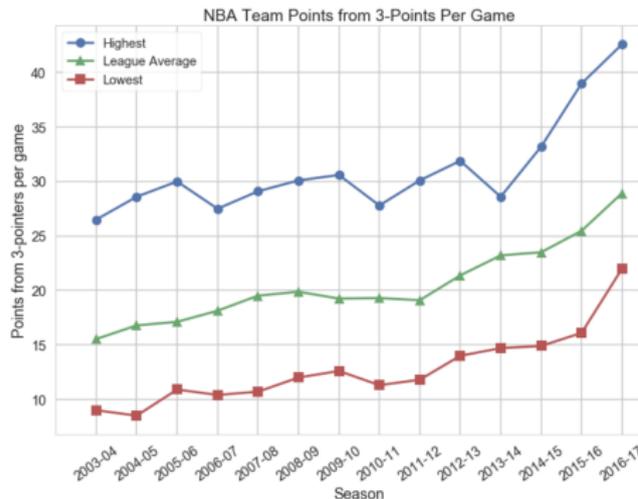
Après-saison 2001 : Billy Beane est choqué par la mauvaise performance de son équipe Oakland Athletics \Rightarrow engage l'économiste Peter Brand qui utilise des "sabermetrics"

Grand succès \Rightarrow a révolutionné le sport professionnel

Exemple flagrant : le basketball

Le fameux head coach des San Antonio Spurs, Gregg Popovich, a entamé la révolution des tirs à 3 points dans la NBA.

Year	NBA Avg. Pct. of 3PA from Corners	Spurs Pct. of 3PA from Corners	Spurs Rank
2000-01	.239	.345	1
2001-02	.235	.364	1
2002-03	.257	.400	1
2003-04	.261	.363	2
2004-05	.272	.431	1
2005-06	.270	.408	1
2006-07	.271	.384	1
2007-08	.264	.414	1



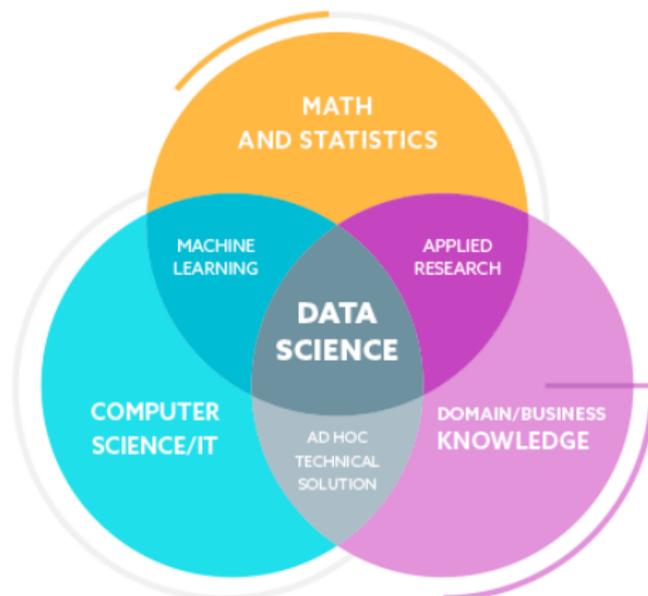
Big data - big motivation



Droits d'image : STATS, VectorStock

La science des données

La science des données est l'extraction de connaissances à partir de données, combinant des idées issues du Machine Learning, des statistiques, de l'informatique, de l'ingénierie et des mathématiques...

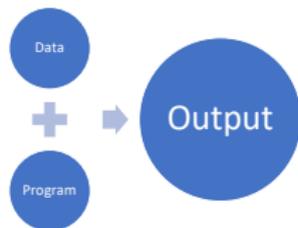


Crédits d'image : data-flair.training et denologix.com

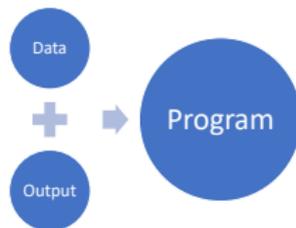
Qu'est-ce que le Machine Learning ?

Définition donnée par Arthur Samuel en 1959 : **Field of study that gives computers the ability to learn without being explicitly programmed.**

Traditional Programming



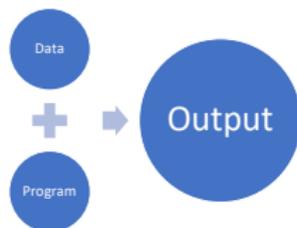
Machine Learning



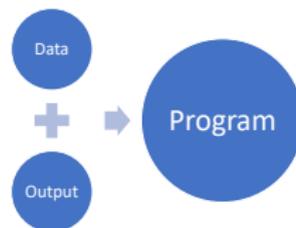
Qu'est-ce que le Machine Learning ?

Définition donnée par Arthur Samuel en 1959 : **Field of study that gives computers the ability to learn without being explicitly programmed.**

Traditional Programming



Machine Learning

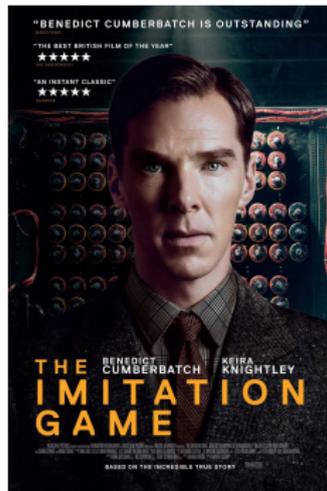


C'est une forme d'Intelligence Artificielle !

Mais c'est quoi l'Intelligence Artificielle ?

Capacité d'un ordinateur numérique ou d'un robot contrôlé par ordinateur à effectuer des tâches généralement associées à des êtres intelligents
(source : <https://www.britannica.com/technology/artificial-intelligence>).

Origines : Alan Turing fin des années 1940.



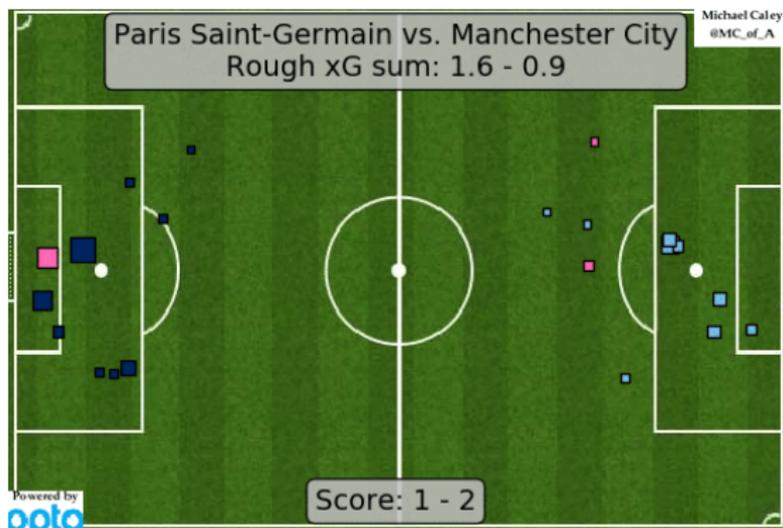
Capacités inouïes



Source : "A Neural Algorithm of Artistic Style" de Gatys, Ecker and Bethge 

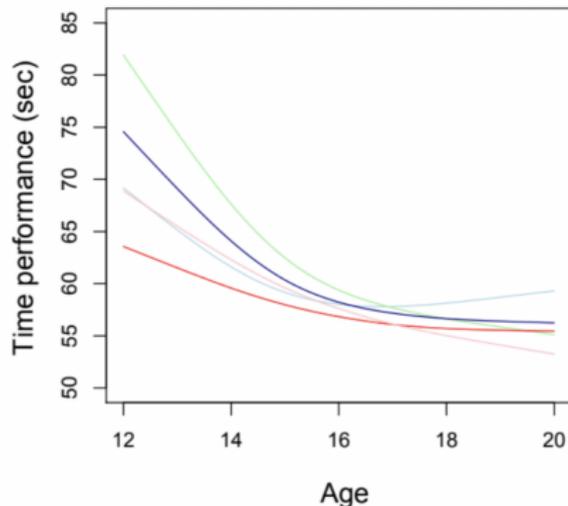
- 1 La science des données à la rencontre du sport
- 2 Médaille d'or : analyses et tactiques**
- 3 Médaille d'argent : prédictions et paris sportifs
- 4 Médaille de bronze : prévention de blessures

Expected goals (xG)



- Mesure pour chaque tir la probabilité d'un but
- Raconte davantage sur un match que le score final, p.ex. 2-1 peut être en xG 2.3-0.5 ou bien 1.8-4.5
- Sur plusieurs matchs : informatif pour les entraîneurs
- Rankings alternatifs

Identification de talents en natation



- Extrait de "Morphology, Talent Identification, Training Methods... Among the Many Ways of looking at Performance in Swimming" par Leroy et Pla dans "Science Meets Sports - When Figures are More Than Numbers"
- Collaboration avec la Fédération Française de Natation
- Analyse par séries chronologiques et clustering

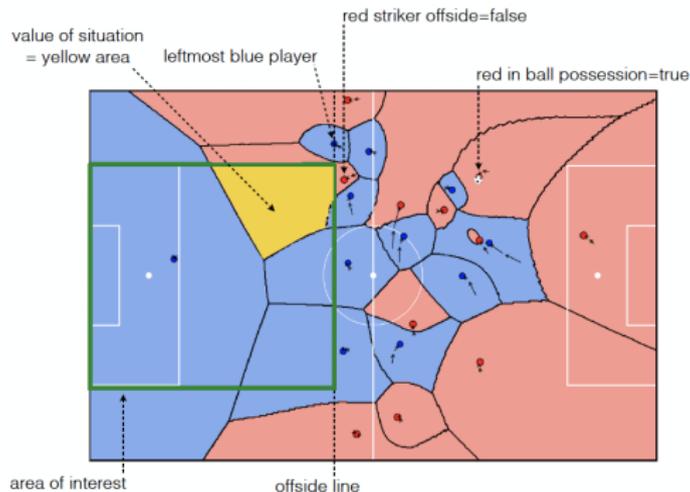
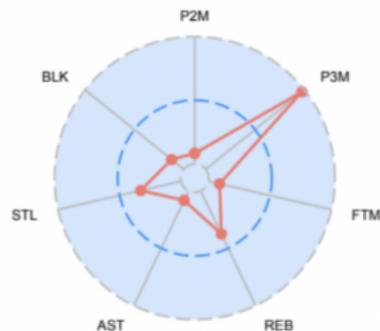


Figure 8: Retrieval of interesting situations.

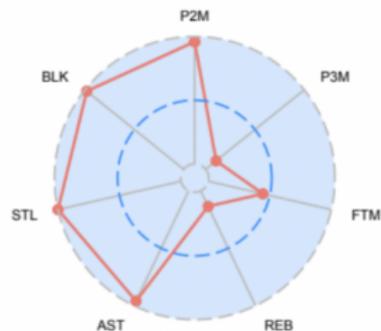
Extrait de "Analysing Positional Data" de Brefeld, Lasek et Mair dans "Science Meets Sports - When Figures are More Than Numbers"

Visualisations au basketball

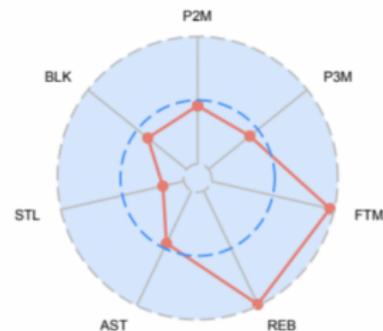
Karl-Anthony Towns



Anthony Davis



Joel Embiid



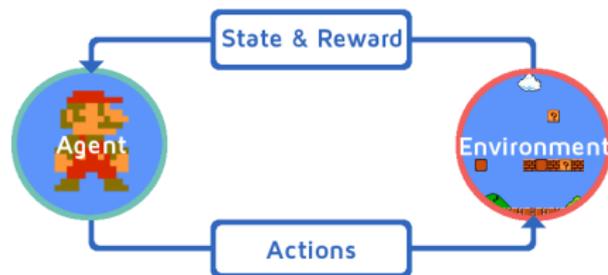
- Extrait de "Advances in Basketball Statistics" de Manisera, Sandri et Zuccolotto dans "Science Meets Sports - When Figures are More Than Numbers"
- Plot radial

Suprématie des données et de l'IA : les échecs



Comment une IA peut-elle devenir si imbattable ?

Reinforcement Learning



Un agent est supposé trouver le meilleur chemin, celui qui maximise une certaine récompense.

La machine apprend par sa propre expérience et s'améliore.

Exemple encore plus frappant : le jeu de go

Le superordinateur de Google a une nouvelle fois eu raison de l'être humain en battant de peu le champion chinois Ke Jie. Deux autres matchs sont prévus dans la semaine.



Le champion du jeu de Go Ke Jie défait contre l'IS Deepmind le 23 mai 2017

Crédit image : Sciences et Avenir

- 1 La science des données à la rencontre du sport
- 2 Médaille d'or : analyses et tactiques
- 3 Médaille d'argent : prédictions et paris sportifs**
- 4 Médaille de bronze : prévention de blessures

Sur base de facteurs déterminants nous cherchons à prédire une certaine quantité appelée réponse !

Sur base de facteurs déterminants nous cherchons à prédire une certaine quantité appelée réponse !

En termes mathématiques : à chaque ensemble de prédicteurs $(x_1, x_2, \dots, x_p)_i$ correspond une réponse associée $y_i, i = 1, \dots, n$.

Sur base de facteurs déterminants nous cherchons à prédire une certaine quantité appelée réponse !

En termes mathématiques : à chaque ensemble de prédicteurs $(x_1, x_2, \dots, x_p)_i$ correspond une réponse associée $y_i, i = 1, \dots, n$.

La machine apprend grâce à des *training data* la relation entre facteurs déterminants et la réponse.

Nota bene : la réponse peut être quantitative (régression) et qualitative (classification).

Comment construire un modèle ML pour prédire un match de foot ?

Comment construire un modèle ML pour prédire un match de foot ?



1 Définir le but :

Comment construire un modèle ML pour prédire un match de foot ?



- 1 Définir le but : prédire le nombre de buts de chaque équipe

Comment construire un modèle ML pour prédire un match de foot ?



- 1 Définir le but : prédire le nombre de buts de chaque équipe
- 2 Trouver des facteurs déterminants :

Comment construire un modèle ML pour prédire un match de foot ?



- 1 Définir le but : prédire le nombre de buts de chaque équipe
- 2 Trouver des facteurs déterminants : âge moyen des joueurs, résultats précédents, niveau de joueurs, ...

Comment construire un modèle ML pour prédire un match de foot ?



- 1 Définir le but : prédire le nombre de buts de chaque équipe
- 2 Trouver des facteurs déterminants : âge moyen des joueurs, résultats précédents, niveau de joueurs, ...
- 3 Collecter des données :

Comment construire un modèle ML pour prédire un match de foot ?



- 1 Définir le but : prédire le nombre de buts de chaque équipe
- 2 Trouver des facteurs déterminants : âge moyen des joueurs, résultats précédents, niveau de joueurs, ...
- 3 Collecter des données : résultats de matchs précédents

Comment construire un modèle ML pour prédire un match de foot ?



- 1 Définir le but : prédire le nombre de buts de chaque équipe
- 2 Trouver des facteurs déterminants : âge moyen des joueurs, résultats précédents, niveau de joueurs, ...
- 3 Collecter des données : résultats de matchs précédents
- 4 Entraîner l'IA :

Comment construire un modèle ML pour prédire un match de foot ?



- 1 Définir le but : prédire le nombre de buts de chaque équipe
- 2 Trouver des facteurs déterminants : âge moyen des joueurs, résultats précédents, niveau de joueurs, ...
- 3 Collecter des données : résultats de matchs précédents
- 4 Entraîner l'IA : algorithme qui relie les facteurs déterminants au résultat de chaque match

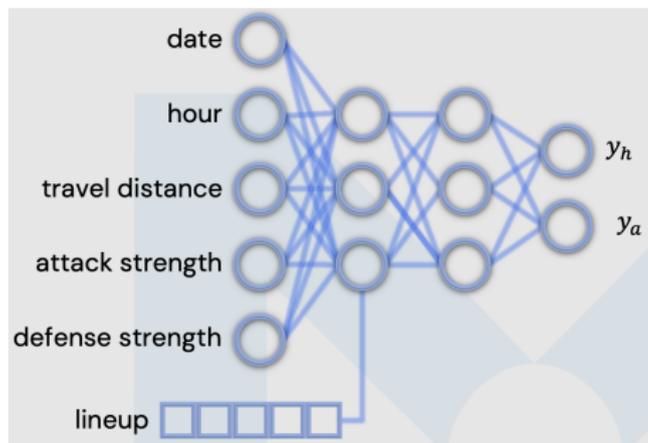
Thèse de Florian Felice

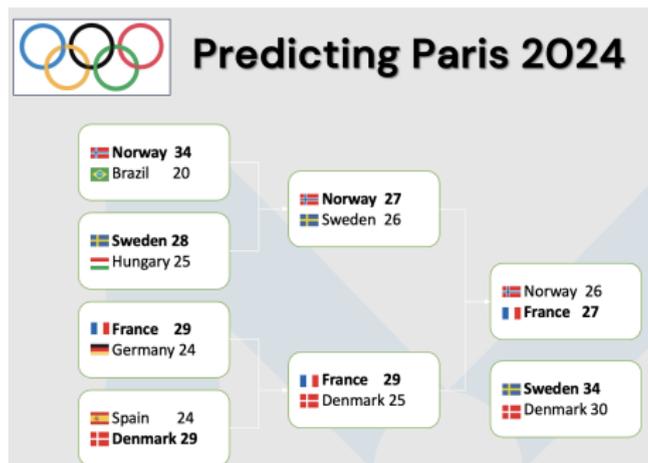
Entraîner une IA sur des matchs historiques :

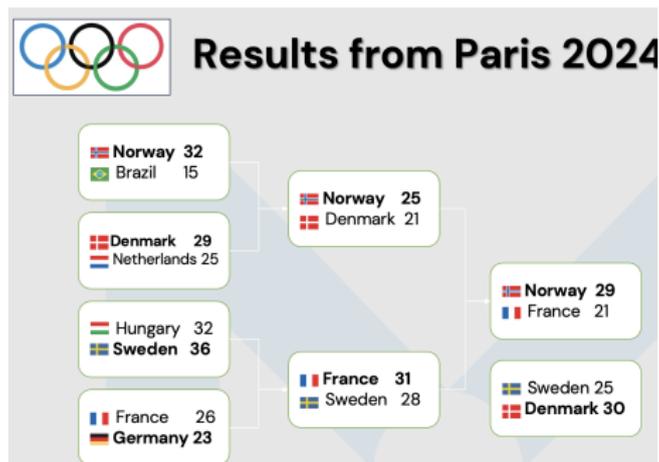
- Info sur les matchs (date, heure, ...)
- Info sur l'équipe (distance)
- Forces de jeu (attaque, défense)
- Composition de l'équipe

⇒ Scores des deux équipes (y_h, y_a)

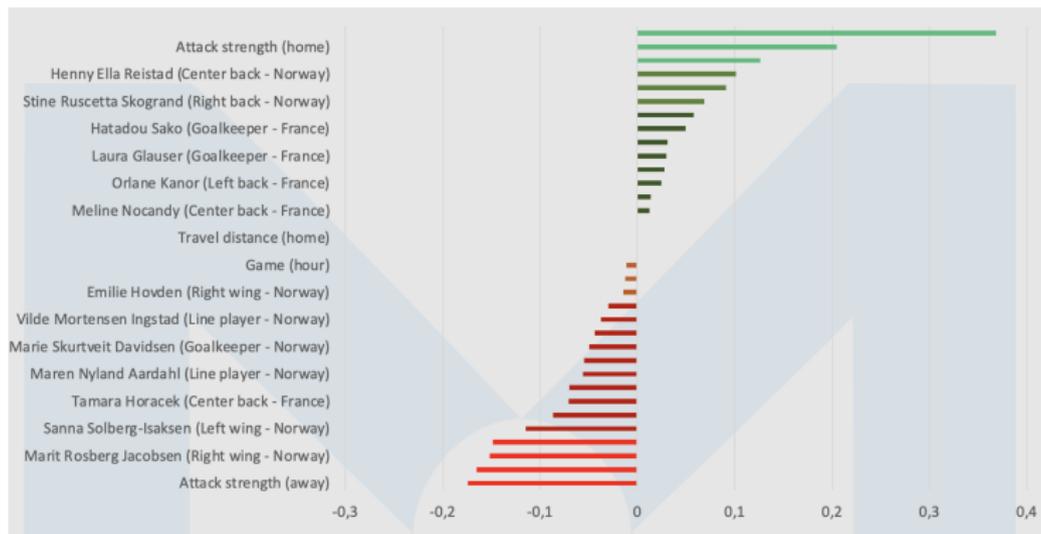
Prédiction de matchs de handball



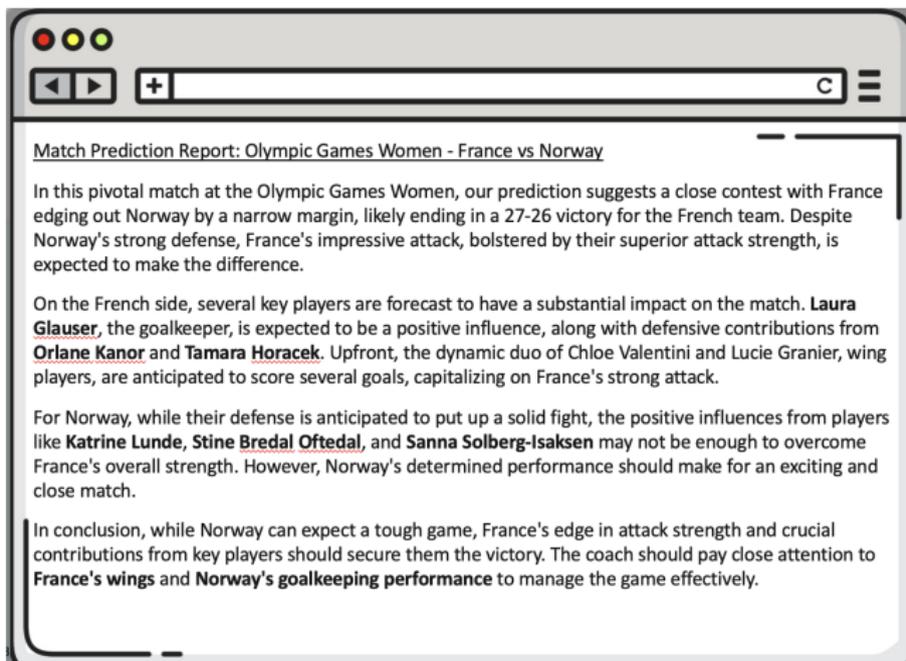




Expliquer les prédictions



Expliquer les prédictions



Match Prediction Report: Olympic Games Women - France vs Norway

In this pivotal match at the Olympic Games Women, our prediction suggests a close contest with France edging out Norway by a narrow margin, likely ending in a 27-26 victory for the French team. Despite Norway's strong defense, France's impressive attack, bolstered by their superior attack strength, is expected to make the difference.

On the French side, several key players are forecast to have a substantial impact on the match. **Laura Glauser**, the goalkeeper, is expected to be a positive influence, along with defensive contributions from **Orlane Kanor** and **Tamara Horacek**. Upfront, the dynamic duo of Chloe Valentini and Lucie Granier, wing players, are anticipated to score several goals, capitalizing on France's strong attack.

For Norway, while their defense is anticipated to put up a solid fight, the positive influences from players like **Katrine Lunde**, **Stine Bredal Oftedal**, and **Sanna Solberg-Isaksen** may not be enough to overcome France's overall strength. However, Norway's determined performance should make for an exciting and close match.

In conclusion, while Norway can expect a tough game, France's edge in attack strength and crucial contributions from key players should secure them the victory. The coach should pay close attention to **France's wings** and **Norway's goalkeeping performance** to manage the game effectively.

- 1 La science des données à la rencontre du sport
- 2 Médaille d'or : analyses et tactiques
- 3 Médaille d'argent : prédictions et paris sportifs
- 4 Médaille de bronze : prévention de blessures

PRESS RELEASE

Joint research project: Shoe cushioning as a possible risk factor for running injuries?

28 October 2020

5minutes

METABOLIC DISEASES

PREVENTIVE MEDICINE



- 848 coureurs amateurs
- analyse de survie, modèle à risques compétitifs
- étude menée par le Luxembourg Institute of Health avec Decathlon
- nouvelle analyse dans la thèse de Katarzyna Szczerba

- Problème multifactoriel

Prévenir les blessures - tout sauf évident !

- Problème multifactoriel
- Données de sources différentes et de types différents : facteurs orthopédiques, images, technologie portable, questionnaires, entraînements, matches, ...

Prévenir les blessures - tout sauf évident !

- Problème multifactoriel
- Données de sources différentes et de types différents : facteurs orthopédiques, images, technologie portable, questionnaires, entraînements, matches, ...
- Blessures = événement rare (heureusement...)

Prévenir les blessures - tout sauf évident !

- Problème multifactoriel
- Données de sources différentes et de types différents : facteurs orthopédiques, images, technologie portable, questionnaires, entraînements, matches, ...
- Blessures = événement rare (heureusement...)
- Nouveau trend vers le machine learning (Van Eetvelde et al. 2021, J. Exper. Ortho.)

Prévenir les blessures - tout sauf évident !

- Problème multifactoriel
- Données de sources différentes et de types différents : facteurs orthopédiques, images, technologie portable, questionnaires, entraînements, matchs, ...
- Blessures = événement rare (heureusement...)
- Nouveau trend vers le machine learning (Van Eetvelde et al. 2021, J. Exper. Ortho.)
- Essentiel : algorithme doit être interprétable !

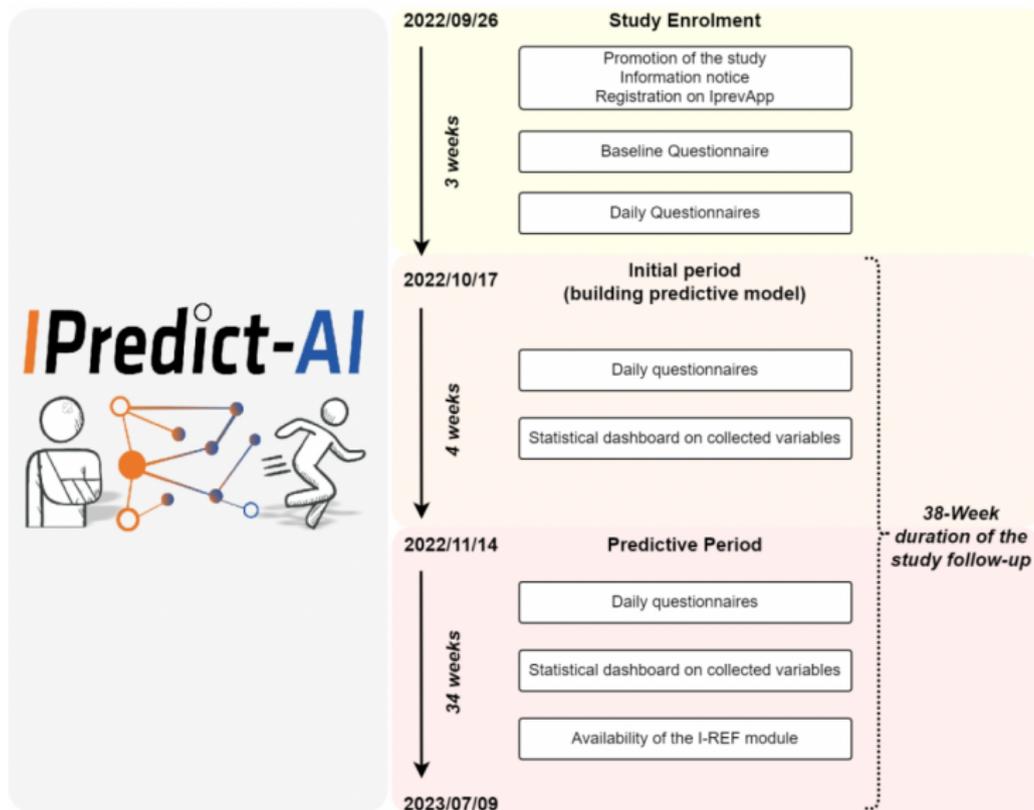
Prévenir les blessures - tout sauf évident !

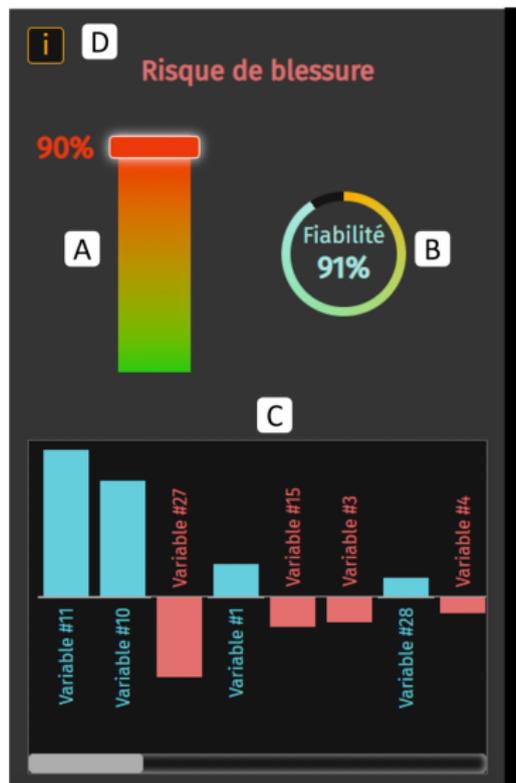
- Problème multifactoriel
- Données de sources différentes et de types différents : facteurs orthopédiques, images, technologie portable, questionnaires, entraînements, matchs, ...
- Blessures = événement rare (heureusement...)
- Nouveau trend vers le machine learning (Van Eetvelde et al. 2021, J. Exper. Ortho.)
- Essentiel : algorithme doit être interprétable !
- Bonne communication avec le staff sportif et volonté des athlètes

Avec Pierre-Eddy Dandrieux, Laurent Navarro, David Blanco, Alexis Ruffault, Antoine Bruneau, Joris Chapon, Karsten Hollander et Pascal Edouard, nous avons fait une étude prospective appelée **IPredict-AI - Injury Prediction with Artificial Intelligence**.

- Athlètes de la Fédération Française d'Athlétisme
- Suivi sur 38 semaines
- Questionnaires quotidiens sur leur activité sportive, état psychologique, sommeil, blessure, et degré d'utilisation de notre I-REF
- I-REF (Injury Risk Estimation Feedback) donne chaque jour une estimation du risque de blessure
- On mesure le nombre de jours perdus par 1000h d'activité sportive

Design d'étude





Extrait de Dandrieux et al., BMJ Open



Sports – Training and Research in Data Science Methods for ANalytics and INjury Prevention Group

S-TRAINING brings together an inter-disciplinary group of experts in sports science, sports medicine and data science that works together to address various sports-data-related challenges. The general research approach in this network can be summarized as follows: we propose to develop innovative Data Science methods to address injury prevention and coordination/tactics/performance analyses from an inter-disciplinary perspective fuelled by expert knowledge in sports science and sports medicine. A particular strength of our network is the diversity of the datasets we will analyse, diversity in terms of sports, gender, athletes (professionals/amateurs) and heterogeneous data sources (obtained via modern data collection techniques such as wearables or Virtual Reality).

More concretely, our group pursues the following main goals:

1. Develop data-driven predictions for negative events in sports such as acute and chronic injuries or health complaints, and close knowledge gaps in injury aetiology by looking at potentially influencing factors such as technique, training regimen, psychological, biomechanical, performance-related and gender-specific aspects of the athletes.
2. Develop a computational understanding of optimised coordination for performance at two distinct levels: (i) coordination of limbs and joint motions of athletes to allow for high performances, and (ii) coordination of players of a team to allow for winning games.

By providing open access to the research solutions developed throughout the project, a large range of sport clubs and athletes (both professional and recreational), particularly in commercially weaker sports, will benefit from this research.

La Science des Données a révolutionné le sport par le domaine sports analytics.



Sports – Training and Research in Data Science Methods for ANalytics and INjury Prevention Group

S-TRAINING brings together an inter-disciplinary group of experts in sports science, sports medicine and data science that works together to address various sports-data-related challenges. The general research approach in this network can be summarized as follows: we propose to develop innovative Data Science methods to address injury prevention and coordination/tactics/performance analyses from an inter-disciplinary perspective fuelled by expert knowledge in sports science and sports medicine. A particular strength of our network is the diversity of the datasets we will analyse, diversity in terms of sports, gender, athletes (professionals/amateurs) and heterogeneous data sources (obtained via modern data collection techniques such as wearables or Virtual Reality).

More concretely, our group pursues the following main goals:

1. Develop data-driven predictions for negative events in sports such as acute and chronic injuries or health complaints, and close knowledge gaps in injury aetiology by looking at potentially influencing factors such as technique, training regimen, psychological, biomechanical, performance-related and gender-specific aspects of the athletes.
2. Develop a computational understanding of optimised coordination for performance at two distinct levels: (i) coordination of limbs and joint motions of athletes to allow for high performances, and (ii) coordination of players of a team to allow for winning games.

By providing open access to the research solutions developed throughout the project, a large range of sport clubs and athletes (both professional and recreational), particularly in commercially weaker sports, will benefit from this research.

La Science des Données a révolutionné le sport par le domaine sports analytics.

Jamais 100% de contrôle par les données car les athlètes sont humain(e)s !