

# FUTURA

## Les algorithmes peuvent prédire le vainqueur de la Coupe du monde, vrai ou faux ?

### Podcast écrit et lu par Melissa Lepoureau

*N.B. La podcastrice s'est efforcée, dans la mesure du possible, d'indiquer par quel personnage ou personnalité sont prononcées les citations. Néanmoins, certaines de ces dernières échappent à sa connaissance et devront rester anonymes.*

*[Une musique d'introduction détendue et jazzy. Une série de voix issues de films se succèdent, s'exclamant alternativement « C'est vrai », ou « C'est faux ». L'intro se termine sur la voix du personnage de Karadoc issu de Kaamelott, s'exclamant d'un air paresseux « Ouais, c'est pas faux. »]*

Est-ce qu'on pourrait vraiment prédire le vainqueur de la Coupe du monde grâce aux algorithmes ?

Pendant des décennies, lorsqu'une Coupe du monde approchait, tout le monde se posait la même question : qui va gagner ? Les supporters débattaient pendant des heures, les journalistes multipliaient les pronostics, les anciens joueurs partageaient leur intuition et certains allaient même jusqu'à consulter des voyants ou à s'en remettre à des mascottes improbables. On se souvient par exemple de Paul le poulpe, devenu une célébrité mondiale en 2010 après avoir prédit plusieurs résultats du Mondial. Mais aujourd'hui, une nouvelle génération de pronostiqueurs est entrée dans le jeu. Ils ne portent ni maillot, ni costume de consultant télé, et ils ne possèdent aucun don de voyance. Leur terrain de jeu, ce sont les bases de données. Leur langage, ce sont les mathématiques. Et leur arme secrète, ce sont les algorithmes. Alors la question mérite d'être posée : les algorithmes peuvent-ils vraiment prédire le vainqueur de la Coupe du monde ? À première vue, cela semble presque absurde. Après tout, le football est probablement l'un des sports les plus imprévisibles au monde. Un ballon qui touche le poteau, un carton rouge inattendu, une blessure à la dernière minute ou une séance de tirs au but peuvent bouleverser le destin d'une compétition entière. Pourtant, depuis une quinzaine d'années, les scientifiques des données s'intéressent de plus en plus au football. Leur objectif n'est pas de lire l'avenir, mais de mesurer les probabilités. Car en réalité, prédire n'est pas forcément annoncer avec certitude ce qui va se produire. C'est plutôt estimer ce qui a le plus de chances d'arriver. Pour comprendre comment ces modèles fonctionnent, imaginons qu'un algorithme doive évaluer les chances de victoire de chaque nation engagée dans une Coupe du monde. La première étape consiste à collecter une quantité gigantesque d'informations. Les résultats des matchs internationaux des dernières années, les performances récentes des équipes, les statistiques offensives et défensives, les classements mondiaux, la valeur marchande des joueurs, leur expérience dans les grandes compétitions, leur âge, leur forme du moment, les blessures éventuelles, et même parfois des données économiques ou démographiques concernant les pays représentés. Une fois toutes ces données réunies, les chercheurs

construisent des modèles statistiques capables d'évaluer la force réelle de chaque sélection. Mais ce n'est pas tout. Certains modèles intègrent également les cotes des bookmakers. Cela peut sembler étonnant, mais les bookmakers sont eux-mêmes de véritables machines à statistiques. Leurs cotes reflètent des milliers d'informations analysées en permanence. Les intégrer dans un modèle permet donc parfois d'améliorer considérablement la qualité des prévisions. Ensuite intervient l'intelligence artificielle. Des algorithmes d'apprentissage automatique sont entraînés sur les résultats des précédentes Coupes du monde et des grands tournois internationaux. Ils apprennent progressivement quelles variables semblent avoir le plus d'influence sur les résultats réels. Est-ce que l'expérience compte davantage que la jeunesse ? Les équipes les plus chères gagnent-elles plus souvent ? Le classement FIFA est-il réellement pertinent ? Quels sont les facteurs qui permettent de distinguer un futur champion d'un simple outsider ? Les modèles cherchent des corrélations et des tendances invisibles à l'œil nu. Une fois cette phase terminée, la machine peut commencer à simuler le tournoi. Et c'est là que les chiffres deviennent impressionnants. Certains chercheurs réalisent non pas cent simulations, ni mille, mais cent mille simulations complètes de la compétition. Dans chaque scénario, tous les matchs sont joués virtuellement. Les probabilités évoluent à chaque tour. Une équipe peut être éliminée prématurément dans certaines simulations et soulever le trophée dans d'autres. Au bout de cent mille tournois virtuels, on obtient une estimation statistique extrêmement précise des chances de chaque nation. Prenons un exemple simple. Si une équipe remporte quinze mille simulations sur cent mille, cela signifie qu'elle possède environ quinze pour cent de chances de devenir championne du monde. Cela paraît peu, mais dans un tournoi aussi relevé, c'est souvent suffisant pour être considérée comme favorite. Et c'est là qu'apparaît un premier élément surprenant. Même les plus grands favoris dépassent rarement les vingt pour cent de chances de victoire. Autrement dit, avant le début d'une Coupe du monde, aucune équipe n'est proche de la certitude. Même les meilleures nations restent très loin du cent pour cent. Cela montre à quel point ce tournoi demeure imprévisible. Les chercheurs qui développent ces modèles ont d'ailleurs connu des succès remarquables. Certains avaient correctement identifié les États-Unis comme favorites de la Coupe du monde féminine 2019. D'autres ont régulièrement placé les futurs vainqueurs parmi leurs principaux prétendants. Mais ils ont également connu des échecs. L'Argentine en 2022 n'était pas systématiquement annoncée comme favorite absolue par tous les modèles. L'Espagne féminine en 2023 non plus. Pourtant, ces équipes ont fini par triompher. Pourquoi ? Parce qu'un algorithme ne prévoit pas un scénario unique. Il estime une distribution de probabilités. Une équipe ayant dix pour cent de chances de gagner peut parfaitement devenir championne. Dix pour cent, ce n'est pas zéro. C'est simplement moins probable. C'est un peu comme lancer un dé. Si le chiffre six a une chance sur six d'apparaître, cela ne signifie pas qu'il n'arrivera pas au prochain lancer. Cela signifie seulement qu'il n'est pas l'issue la plus fréquente sur le long terme. Le football fonctionne de manière similaire, avec énormément de hasard à court terme. D'ailleurs, plus un sport comporte peu de points marqués, plus l'incertitude est élevée. Au basket-ball, où une équipe marque souvent plus de cent points, la meilleure équipe gagne généralement. Au football, un seul but peut suffire à faire basculer tout un match. Cette part de hasard est précisément ce qui rend ce sport si fascinant. Alors, après tout ce parcours, revenons à notre question de départ. Les algorithmes peuvent-ils prédire le vainqueur de la Coupe du monde ? Si par prédire on entend annoncer avec certitude l'équipe qui soulèvera le trophée avant même le premier match, alors la réponse est non. Aucun algorithme, aussi sophistiqué soit-il, ne peut éliminer totalement l'incertitude, les surprises et les événements imprévus qui font la beauté du

football. Mais si par prédire on entend identifier les équipes qui ont statistiquement les meilleures chances de gagner et évaluer leurs probabilités avec une précision supérieure à l'intuition humaine, alors la réponse devient oui. Les algorithmes ne voient pas l'avenir. Ils calculent les futurs les plus plausibles. Et c'est précisément pour cette raison qu'ils sont souvent meilleurs que les experts, les supporters... et même que les célèbres animaux devins.

Et vous, vous avez d'autres idées reçues à debunker ? Envoyez-les nous sur les apps audio ou en vocal sur Instagram, et nous les inclurons dans de futurs épisodes. Pensez à vous abonner à Science ou Fiction et à nos autres podcasts pour ne plus manquer un seul épisode, et n'hésitez pas à nous laisser un commentaire et une note pour nous dire ce que vous en pensez et soutenir notre travail. À bientôt !