



STATSBOMB

COMMENT LES DONNÉES STATSBOMB PEUVENT
AMÉLIORER LES MODÈLES DE PRÉDICTION

Quel que soit leur niveau de complexité, l'efficacité des modèles de prédiction dépend notamment des données utilisées. Si vous introduisez des données de mauvaise qualité dans un modèle, vos résultats sont inévitablement compromis. Il n'existe aucune technique mathématique magique permettant d'utiliser efficacement des informations incomplètes et incohérentes.

Chez StatsBomb, nos fondateurs et plusieurs de nos cadres supérieurs connaissent l'importance de pouvoir accéder aux meilleures données possible de par leur expérience dans l'industrie du pari. Ainsi, nous savons que la qualité et la profondeur des données sont essentielles. Dans les secteurs ultracompetitifs, tout le monde cherche à obtenir le moindre avantage, et le fait de disposer de données de meilleure qualité pour votre modèle de prédiction vous donnera sans aucun doute une longueur d'avance.

Lorsque nous avons commencé à recueillir nos propres données en 2017, nous étions sûrs que les ensembles de données et modèles disponibles pouvaient être grandement améliorés. Tant du point de vue du football professionnel ou l'industrie du pari, il manquait de nombreux éléments que nous considérons comme importants : positionnement des gardiens de but et des défenseurs lors des tirs, événements de pressing ou encore le pied avec lequel chaque passe est jouée. Nous avons donc ajouté ces points de données uniques à notre spécification de données initiale.

Au total, les données StatsBomb enregistrent plus de 3 400 événements par match en moyenne, soit plus du double que de nombreux autres fournisseurs de données, sur l'ensemble des plus de 80 compétitions dans le monde. Les informations contextuelles fournies par les points de données supplémentaires que nous recueillons font de notre modèle Expected Goals (xG) le plus riche et le plus fiable du marché.

L'importance des Expected Goals (xG) pour les modèles de prédiction

Les modèles sur les Expected Goals (xG) sont devenus la norme du marché pour évaluer la qualité intrinsèque d'une équipe et prédire ses performances futures. Il a été démontré qu'ils s'avéraient être de meilleurs indicateurs des performances futures que les points actuels, la différence de but actuelle, les différences de tirs et d'autres éléments statistiques de base.

Les modèles standards sur les Expected Goals comprennent des éléments tels que :

- La distance par rapport au but
- L'angle par rapport au but
- La partie du corps avec laquelle le tir est effectué
- Le type de passe et d'action précédant le tir
(passe en profondeur, centre, coup de pied arrêté, dribble, etc.)

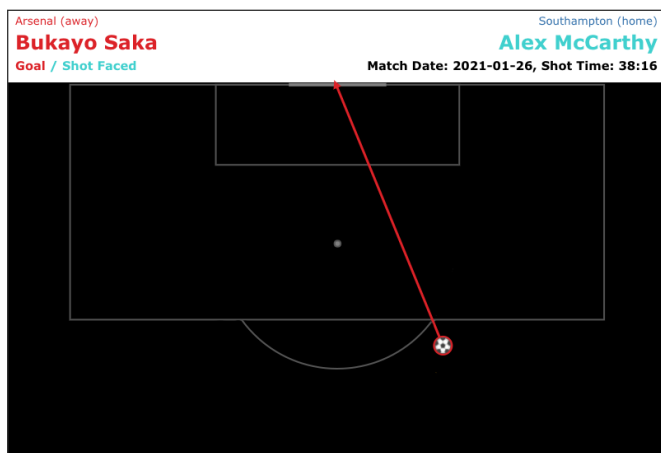
Ces données couvrent de nombreux éléments clés qui sont nécessaires pour déterminer les chances de réussite d'un tir, et sur un échantillon important, les modèles n'incluant que ces éléments donnent une assez bonne indication de la qualité intrinsèque d'une équipe et des performances futures probables. Ces modèles oublient cependant certains éléments que la plupart considéreraient comme essentiels pour déterminer les chances qu'un tir se transforme en but.

L'une des principales innovations dans notre spécification des données est la technologie Freeze Frames, qui permet d'enregistrer la position de tous les joueurs à l'image au moment d'un tir. Ceci donne des informations supplémentaires essentielles sur les tirs qui permettent à notre modèle sur les buts attendus de prendre en compte : ►

- la position du gardien de but et la situation dans laquelle il se trouve (sur ses appuis, en mouvement, au sol, etc.)
- la position des défenseurs

Il est évident qu'un tir dans un but ouvert, quelle que soit la distance, a plus de chance de se transformer en but qu'un tir avec le gardien de but sur ses appuis et dans son but. Ou qu'un tir effectué avec quatre défenseurs directement entre le tireur et le but a moins de chance de réussite qu'un tir sans défenseur, mais il s'agit de choses que les précédents modèles ne prenaient pas en compte.

Pour vous donner une idée, l'image sur la gauche représente les informations utilisées dans les modèles standards sur les buts attendus des autres fournisseurs, tandis que l'image sur la droite affiche toutes les informations supplémentaires utilisées dans le modèle Expected Goals de StatsBomb.



Nous cherchons continuellement à nous améliorer et à innover, et depuis le lancement initial de nos données, nous avons ajouté un autre point de données à notre modèle sur les buts attendus : la hauteur du ballon lors du tir. Encore une fois, cela peut sembler évident, mais il y a une différence entre un tir au pied pris à hauteur de cheville et un tir au pied pris en demi-volée à hauteur d'épaule. Cela est maintenant pris en compte dans notre modèle.

C'est grâce à ces innovations que le modèle Expected Goals de StatsBomb génère constamment des valeurs qui semblent plus conformes à la réalité du football, et qui permettent d'obtenir un meilleur aperçu de la qualité intrinsèque des équipes et des joueurs.

Innovation continue

Mais il ne s'agit pas du seul modèle avancé que nous offrons. Entre notre modèle sur la réussite attendue des passes et notre nouveau modèle sur la progression holistique du ballon, nous continuons à offrir des informations qui comptent. Notre équipe de sciences des données hautement qualifiée est dédiée à la création de nouveaux modèles améliorés et la recherche.

Notre engagement pour l'innovation a cette année permis le lancement de StatsBomb 360, notre nouveau produit révolutionnaire. StatsBomb 360 offre encore plus d'informations contextuelles grâce à l'utilisation de notre technologie Freeze Frame pour tous les événements liés au ballon, indiquant ainsi la position de tous les joueurs à l'image lors de chaque événement. Ceci nous permet d'étudier des choses que nous étions jusqu'alors incapables de directement analyser sur la base des données d'événements standard : ►

- Passes éliminant des lignes
- Réceptions de ballons dans l'espace
- Distance par rapport à tous les défenseurs
- Lignes de passe
- Défenseurs éliminés par chaque passe
- Formation défensive lors de chaque événement

Et nous ne sommes qu'à la surface de ce qu'il est possible de réaliser. Les premiers retours de clients indiquent que StatsBomb 360 fournit 80 à 90 % des informations généralement obtenues à partir des données de tracking en étant bien plus facilement utilisables. À la différence des données de tracking, ces informations peuvent facilement être obtenues sur un grand nombre de championnats. Lors de la saison 2021-22, nous collecterons des données 360 dans 38 championnats du monde entier.

Depuis son lancement en 2017, les données StatsBomb ont établi une nouvelle norme de données événementielles dans le football, offrant des informations aux détails sans précédent à plus de 100 clubs professionnels, fédérations, agences, médias et organisations de pari avec lesquels nous collaborons. Il s'agit du seul ensemble de données conçu par et pour les professionnels du football, et tout ce qui en fait le bon choix pour les clubs et les fédérations en fait également le choix logique pour les modèles de prédiction avancés. Aucun autre fournisseur de données n'offre un tel niveau de détails et de qualité. ►

Ce qui rend les données StatsBomb uniques

		STATSBOMB	AUTRES FOURNISSEURS DE DONNÉES SUR LES ÉVÈNEMENTS
Données sur les événements	Évènements par match	3 400 en moyenne	Moins de 2 000 (selon les fournisseurs)
	Évènements de pressing et durée des pressings	✓	✗
	Position des joueurs lors des tirs (gardien de but compris)	✓	✗
	Le pied avec lequel chaque passe est jouée	✓	✗
	Conduites de balle	✓	✗
	Hauteur du ballon au moment du tir	✓	✗
	Plus haut niveau de détail sur tous les championnats couverts	✓	✗
	Données en temps réel	3e trimestre 2021	✗
	Position des joueurs lors de chaque événement	✓ (avec StatsBomb 360)	✗

Ce qui rend les données StatsBomb uniques (cont.)

Modèles dérivés et analyse	Plate-forme d'analyse avancée	✓	✗
	xG avec gardiens de but et défenseurs	✓	✗
	Séquences	✓	✓ (Pas toutes)
	Modèle sur la réussite des passes	✓	✗
	Modèle sur les performances des gardiens de but	✓	✗
	Modèle holistique de progression du ballon	✓	✗
	Outil de scouting avancé	✓	✓ (Pas toutes)
	Nombre d'utilisateurs illimité	✓	✗

Nous sommes confiants dans le fait que les données StatsBomb sont le meilleur ensemble de données sur le football disponible, et nous les avons utilisées pour créer certains des modèles les plus fiables et les plus utiles du marché, mais nous pensons également que nous offrons les données les plus précises du marché.

Nos collecteurs de données suivent une formation approfondie avant de commencer à collecter des données réelles, et nous avons créé de puissants outils qui combinent facteur humain et technologie de vision par ordinateur afin d'accélérer la collecte de données et ajouter un processus de validation supplémentaire. De plus, nous avons mis en place un processus d'assurance qualité rigoureux afin de nous assurer que nos clients reçoivent les données les plus précises et exactes possibles. ►

Processus de qualité suivi par les données StatsBomb (cont.)

Formation des collecteurs de données	Les collecteurs de données suivent 4 semaines de formation intensive puis analysent des matchs en mode formation pendant un minimum de 120 heures avant de pouvoir collecter des données « réelles »
	Les nouveaux collecteurs sont suivis de près par les responsables d'équipe et membres chargés de l'assurance qualité
	Retours constants entre l'équipe qualité et les analystes
	Les collecteurs de données sont rétribués financièrement sur la base des indicateurs clés de qualité

StatsBomb Data Quality Process (cont.)

Collecte des données Spécification	La spécification des données dicte les règles et la validation
	Toutes les données doivent exister / être complètes conformément à la spécification
	La validation de pair permet de garantir que tous les événements sont correctement enregistrés du point de vue des deux équipes, p. ex., chaque dribble doit avoir une élimination par un dribble correspondant
	Validation automatisée intégrée au logiciel - emplacement, équipe, faits (validés à partir de sources externes)
	Les erreurs sont automatiquement signalées pour vérification par l'équipe d'assurance qualité
Vérification de la qualité des données (35 % de mise à jour post vérification pour assurance qualité)	Vérification passive - les vérificateurs évaluent des échantillons de données pour chaque collecteur de données et vérifient chaque point de donnée, corrigent les erreurs et transmettent leurs retours aux collecteurs. Les données doivent à nouveau être collectées en dessous de certains seuils
	Vérification active - le vérificateur analyse un match sans accès aux données correspondantes puis compare ensuite les résultats à la collecte originale
	Données mises à jour lors de cette étape et les lacunes du collecteur sont signalées pour reformation
Importation de hub de données StatsBomb (7 % de mise à jour à la suite de vérifications post-importation)	La validation automatisée vérifie que les données sont complètes - les valeurs correspondant au match sont complètes (selon la spécification des données), la position des joueurs lors des événements est collectée, vérification qu'aucun joueur n'a été supprimé, vérification que chaque champ de données apparaît au format demandé
	Les erreurs sont bloquées et signalées à l'équipe d'assurance qualité
Analyse approfondie	Analyse des tendances des collecteurs au fil du temps
	Comparaison des données sur les événements et modèles d'analyse par rapport aux autres fournisseurs de données
	Analyse de cohérence - analyse d'un match par plusieurs collecteurs

Nos données sont accessibles par différents moyens et sous différents formats : API, JSON, XML, CSV. Nous avons également créé des paquets pour les utilisateurs de R et de Python afin de faciliter l'intégration des données. Les informations contenues dans tous nos graphiques et tableaux sur notre plate-forme d'analyse StatsBomb IQ peuvent également être téléchargées au format CSV. Quelle que soit la façon dont vous souhaitez accéder aux données, nous avons une solution.

Nous savons que le fait d'avoir des données précises et détaillées est vital pour créer des modèles de prédiction robustes et notre ensemble de données est conçu pour répondre à ce besoin, avec plus d'événements, plus de points de données, plus de détails et plus de précision que les autres fournisseurs de données. Nous pensons que cela fait des données StatsBomb le choix évident pour les organismes de pari.

Vous voulez en savoir plus sur les avantages des données StatsBomb ?

Rendez-vous sur www.statsbomb.com/fr pour demander une démonstration. SB