



UNIVERSITÀ DEGLI STUDI DI SALERNO

Dipartimento di Informatica

Corso di Laurea Triennale in Informatica

TESI DI LAUREA

**Dall'intuito al dato:
Analisi dei Dati e Machine Learning
nella Football Analytics**

RELATORE

Prof. Fabio Palomba

Università degli Studi di Salerno

CANDIDATO

Ilian Fasolino

Matricola: 0512108680

Questa tesi è stata realizzata nel



*A mia mamma,
saremo liberi in tutto.*

Abstract

Per decenni, l'intuito di grandi calciatori, allenatori e manager professionisti ha guidato il gioco del calcio. Tuttavia, l'intuito può condurre ad interpretazioni e scelte errate, generando così una cascata di risultati negativi.

In risposta a questa sfida, il settore calcistico sta vivendo una vera e propria rivoluzione grazie all'analisi dei Big Data, i quali evitano una considerevole percentuale di decisioni errate, con impatti positivi su tutti i fronti.

Questa tesi si propone di esplorare ed analizzare l'impatto del Machine Learning e dell'analisi dei dati nel contesto del calcio, ponendo particolare attenzione alle applicazioni pratiche che hanno rivoluzionato le dinamiche di questo gioco, la gestione delle squadre e le prestazioni dei singoli giocatori.

Inoltre, l'indagine si estende anche ad altri ambiti, come lo scouting per la scoperta di giovani talenti, la salute mentale dei giocatori tramite le neuroscienze, la prevenzione degli infortuni e la riabilitazione degli atleti.

Questo studio fornisce un quadro dettagliato non solo del contesto applicativo delle varie tecnologie attuali, ma anche delle prospettive a medio termine, contribuendo alla comprensione di come questi strumenti influenzeranno questa pratica sportiva, offrendo spunti per sviluppi futuri e ricerche nel campo.

Indice

Elenco delle Figure	iii
1 Introduzione	1
1.1 Contesto applicativo	1
1.2 Motivazioni e Obiettivi	2
1.3 Struttura della Tesi	2
2 Background e Stato dell'arte	4
2.1 Intelligenza Artificiale (AI)	4
2.2 Machine Learning (ML)	5
2.3 Deep Learning	6
2.4 Analisi dei dati	6
2.5 Natural Language Processing (NLP)	7
2.6 Regressione	7
2.7 Clustering	8
3 Metodologia di ricerca	9
3.1 Obiettivo e domande di ricerca	9
3.2 Formulazione delle query di ricerca	10
3.3 Estrazione e pulizia dei dati	10
3.4 Fase di Snowballing	13

3.5 Analisi dei dati	13
4 Applicazioni pratiche	15
4.1 Scouting	15
4.1.1 Scout Advisor	17
4.1.2 Data Driven Scouting	17
4.1.3 Il caso Philadelphia Union	19
4.2 Match Analysis	19
4.2.1 Dispositivi indossabili	20
4.2.2 Strumenti di visualizzazione dati	22
4.2.3 AIDA: Artificial Intelligence Data Analyst	23
4.3 Alimentazione, Prevenzione e Riabilitazione	24
4.4 Mental Training	25
4.5 Arbitraggio Virtuale	26
4.6 Fan Engagement	26
5 Sfide e limitazioni	29
5.1 Il caso di Kevin De Bruyne	31
6 Conclusioni e Sviluppi Futuri	34
Bibliografia	37

Elenco delle figure

3.1	Processo di estrazione.	11
4.1	Illustrazione del dispositivo IMU creato da PlayerMaker, con la rispettiva companion app dove è possibile visualizzare tutti i dati dell'atleta.	22
4.2	Detta "heatmap", è una rappresentazione grafica di dati dove i valori vengono indicati da variazioni di colori, molto utilizzata in quanto facilita la comprensione e l'interpretazione dei diversi pattern. . . .	23
4.3	Questa immagine mostra come l'emittente finlandese YLE utilizza un software AI-based chiamato "Tactic PRO" per tutta la sua copertura televisiva di Euro 2024.	27
4.4	Questa immagine mostra l'app "Immersive" per Apple Vision Pro, applicazione che grazie alla realtà aumentata arricchisce l'esperienza del tifoso rivoluzionando il broadcasting odierno.	27
5.1	Questa grafico delle sequenze d'attacco mostra il numero di coinvolgimenti di De Bruyne nelle sequenze di attacco che poi vengono concluse con un tiro in porta, nella stagione 22/23.	32

5.2 Questi dati dimostrano il ruolo cruciale di tale giocatore, sia nelle finalizzazioni offensive e soprattutto nella creazione e costruzione del gioco.	32
6.1 Questa immagine mostra la differenza tra dati evento e dati di tracking.	35

CAPITOLO 1

Introduzione

1.1 Contesto applicativo

Negli ultimi anni i Big Data sono entrati con una certa dirompenza nel calcio, catalizzando il processo di scelte tecniche degli allenatori e cambiando lo studio teorico dello sport in sé.

Quest'ultimo, man mano che le tecnologie migliorano e si sviluppano diventando davvero alla portata di tutti, muterà sempre di più, ridisegnando i confini anno dopo anno, aprendo nuove prospettive già nel futuro prossimo.

Grazie all'uso di tecniche di Machine Learning, le squadre ora possono identificare schemi e tendenze che sarebbero stati impossibili da individuare solo con l'occhio umano. Ad esempio, è possibile analizzare dettagliatamente le prestazioni di un giocatore durante un'intera stagione, individuando punti di forza e debolezze specifiche che possono essere migliorate con allenamenti mirati. Inoltre, l'analisi dei dati può essere utilizzata per prevedere infortuni, ottimizzare le strategie di gioco e migliorare la gestione delle risorse umane all'interno del team.

Tutto questo offre anche un vantaggio competitivo significativo. Le squadre che utilizzano i dati per guidare le loro decisioni sono in grado di adattarsi rapidamente alle nuove situazioni, rispondere in modo più efficace alle sfide e sfruttare al meglio le

opportunità. Questo non solo avviene sul campo, ma anche nella gestione economica e operativa del club, dove l'analisi dei dati può portare a una migliore pianificazione finanziaria (specialmente nelle varie spending review), a un marketing più efficace ed una maggiore soddisfazione dei tifosi.

1.2 Motivazioni e Obiettivi

La motivazione principale di questa ricerca è contribuire al processo di crescita dello sviluppo tecnologico tramite il ML e l'analisi dei dati nel contesto calcistico, fornendo una panoramica completa e portando alla luce un'ampia gamma di scenari e applicazioni che unite stanno portando alla rivoluzione di questo sport; Attraverso casi studio dettagliati di squadre e in generale società che hanno raggiunto grandi risultati in tempi più che dimezzati rispetto a squadre che non avevano mai adottato alcun tipo di approccio così avanzato ai dati.

Inoltre, questa tesi si propone di sottolineare l'importanza nell'avere un approccio data-driven: prendere decisioni basate su fatti concreti, tattiche e fattori chiave che sono stati ben compresi e analizzati, a discapito di approcci utilizzati nel passato come l'esperienza personale o il mero intuito.

1.3 Struttura della Tesi

Di seguito è riportata la struttura della tesi:

- **Background e Stato dell'arte.** In questo capitolo vengono approfondite le tematiche principali e i concetti chiave che ruotano attorno a questa ricerca, si analizzano le definizioni di Intelligenza Artificiale, Machine Learning, Deep Learning, analisi dei dati, regressione e clustering.
- **Metodologia di ricerca.** Qui viene fatta una panoramica completa del processo di ricerca, attraversando ogni singola fase.
- **Applicazioni pratiche.** Questo capitolo riporta applicazioni specifiche di Machine Learning nel calcio, illustrando casi di successo e impatti tangibili nei vari

aspetti calcistici più importanti come ad esempio lo scouting, l’analisi dei match, la riabilitazione fisica, il mental training, l’arbitraggio e il fan engagement.

- **Sfide e limitazioni.** In questo capitolo si discute delle varie sfide etiche, tecniche e pratiche associate all’uso del ML nel calcio, come la privacy dei giocatori e soprattutto la validità dei modelli, guardando più da vicino gli aspetti critici e le questioni che potrebbero sorgere in futuro nell’implementazione di queste tecnologie.
- **Conclusioni e Sviluppi futuri.** In questo capitolo vengono illustrati i possibili sviluppi futuri e tendenze di queste tecnologie unite ad approcci data-driven.

CAPITOLO 2

Background e Stato dell'arte

In questo capitolo viene fornita una base concettuale di quelli che sono i principali argomenti correlati a questa ricerca.

2.1 Intelligenza Artificiale (AI)

L'AI può essere vista come un determinato sistema che usa sia un modello basato sulle proprie azioni e sia un modello basato sul mondo esterno, con l'intento di trovare alternative e prendere delle decisioni. Queste decisioni si basano sull'esperienza delle azioni passate, permettendo al sistema di agire nel modo più appropriato ed affidabile possibile, nel presente e nel futuro.[1] Paragonata al cervello umano l'AI può comportarsi in diversi modi, ovvero può agire come un essere umano, pensare e risolvere problemi come un essere umano, in modo logico e razionale sulla base delle informazioni disponibili per ottenere il miglior risultato possibile.

L'AI si divide in due tipi: quella forte e quella debole.[2]

L'AI debole agisce e pensa come se avesse una vera e propria capacità di pensiero, ma senza averla realmente. Non cerca di emulare il nostro pensiero, ma di risolvere problemi complessi che richiedono grandi capacità di calcolo e memoria, come ad esempio la traduzione di testi. Essa è utile per affrontare problemi che già conosciamo,

ma che sono troppo complessi per essere risolti da persone senza l'aiuto di una macchina.

Invece l'AI forte ha capacità cognitive e consapevolezza dei propri limiti. Si basa su sistemi o algoritmi esperti che possono riprodurre le conoscenze e le prestazioni umane in specifici settori. Anche se non emula i processi di pensiero umani, l'AI forte cerca di sviluppare una forma di intelligenza più avanzata. Attualmente, gran parte dello sviluppo avviene nel campo del Machine Learning, dove le macchine imparano dai dati per migliorare le proprie prestazioni nel tempo.

2.2 Machine Learning (ML)

Il Machine Learning è un sottoinsieme dell'Intelligenza Artificiale che si occupa di creare algoritmi capaci di apprendere dai dati e migliorare le loro performance in base ai dati che utilizzano. La popolarità del ML è dovuta principalmente alla grande disponibilità dei dati e alla potenza degli strumenti che computano tutto questo.

Di seguito verranno mostrati i quattro tipi principali di apprendimento nel Machine Learning: supervisionato, non supervisionato, semi-supervisionato e per rinforzo.

- **Apprendimento Supervisionato.** L'algoritmo impara da dati etichettati, ovvero dati dove ogni etichetta indica il valore della variabile dipendente che l'algoritmo deve apprendere. Chi progetta l'algoritmo fornisce in input sia i dati e sia le etichette, da qui il concetto di apprendimento supervisionato (o guidato).
- **Apprendimento Semi-Supervisionato.** In questo caso solo alcuni dati sono etichettati, così facendo l'algoritmo deve imparare ad identificare le etichette mancanti.
- **Apprendimento Non Supervisionato.** L'algoritmo impara dai dati senza conoscere minimamente il valore delle variabili dipendenti, oltre che a trovare senza alcun aiuto da parte del progettista quelli che sono i pattern e le strutture all'interno dei dati.

- **Apprendimento per Rinforzo.** L'algoritmo non impara da un set di dati di esempio come negli altri casi, ma bensì prende decisioni in sequenza e alla fine di ogni sequenza riceve una ricompensa che lo incoraggia a migliorare, ed è per questo motivo che egli impara dagli errori e dalle decisioni precedenti rinforzandosi man mano.

2.3 Deep Learning

Il Deep Learning è una sotto-categoria del Machine Learning e si riferisce a quella branca dell'Intelligenza Artificiale che utilizza algoritmi ispirati alla struttura e alla funzione del cervello umano. Questo metodo specifico di ML impiega reti neurali artificiali disposte in strati successivi per apprendere dai dati in modo iterativo. Una rete neurale è un sistema che può modificare la sua struttura (nodi e interconnessioni) basandosi su dati esterni e informazioni interne che attraversano la rete durante la fase di apprendimento.

2.4 Analisi dei dati

Nel contesto della raccolta e dell'analisi dei dati correlati al contesto calcistico, possiamo distinguere quattro fonti di dati principali che includono:

- **Dati di Prestazione.** Informazioni dettagliate sulle azioni dei giocatori durante le partite e gli allenamenti, come passaggi, tiri, dribbling e movimenti senza palla. Tutti questi dati vengono quasi sempre raccolti utilizzando tecnologie di tracciamento video, sistemi GPS e dispositivi indossabili, tecnologie che verranno trattate in dettaglio nel prossimo capitolo.
- **Dati Biometrici.** Dati raccolti prettamente tramite wearables che monitorano parametri vitali e fisiologici come la frequenza cardiaca (e la sua variabilità), l'ossigeno nel sangue, la velocità e la distanza percorsa in una certo lasso di tempo.
- **Dati di Partita.** Statistiche generali delle partite come il possesso palla, le palle inattive, i tiri in porta, i corner e le sanzioni. Attualmente questi dati vengono

raccolti ancora per la grande maggioranza manualmente ed in alcuni casi da software specifici.

- **Dati di Allenamento.** Dati raccolti durante le varie sessioni di allenamento giornaliero dei giocatori, inclusi i carichi di lavoro, le prestazioni e gli esercizi svolti.

2.5 Natural Language Processing (NLP)

L'elaborazione del linguaggio naturale, nota come NLP (Natural Language Processing), unisce la linguistica computazionale e la modellazione del linguaggio umano basata su regole con modelli statistici e di apprendimento automatico per permettere a computer e dispositivi digitali di riconoscere, comprendere e generare testo e parlato.

In quanto sottocampo dell'Intelligenza Artificiale (AI), l'NLP è fondamentale per le applicazioni e i dispositivi che possono:

- Tradurre il testo da una lingua all'altra.
- Rispondere ai comandi digitati o vocali.
- Riconoscere o autenticare gli utenti in base alla voce.
- Riassumere grandi quantità di testo.
- Valutare l'intenzione o il sentimento del testo o del parlato.
- Generare testo, grafica o altri contenuti su richiesta, spesso in tempo reale.

La maggior parte delle persone ha interagito con l'NLP tramite sistemi GPS a comando vocale, assistenti digitali, software di dettatura Speech to Text, chatbot per il servizio clienti e altre comodità per i consumatori.

2.6 Regressione

Il problema di regressione riguarda la creazione di un modello che usa un algoritmo di apprendimento, chiamato regressore, per fare previsioni sui nuovi dati

basandosi su un set di dati di addestramento, chiamato training set. I regressori sono fondamentalmente delle funzioni matematiche che cercano di descrivere i dati e prevedere risultati futuri.

Esistono diversi tipi di regressori, ognuno con assunzioni e caratteristiche specifiche. La scelta del metodo di regressione dipende da vari fattori, come il numero di variabili indipendenti (predittori), la forma della relazione tra le variabili e il tipo di variabile dipendente che vogliamo prevedere.

2.7 Clustering

Il clustering è una tecnica di ML non supervisionato utilizzata per raggruppare oggetti simili in insiemi chiamati cluster. Il clustering può aiutare a identificare gruppi di giocatori con caratteristiche simili o segmentare tattiche di gioco.

I principali algoritmi di clustering includono:

- **k-means.** Algoritmo che divide i dati in k cluster predefiniti, minimizzando la distanza tra i cluster.
- **DBSCAN (Density-Based Spatial Clustering of Applications with Noise).** Algoritmo che raggruppa punti vicini basandosi sulla densità di punti.
- **Clustering Gerarchico.** Metodo che costruisce una gerarchia di cluster, creando una rappresentazione a forma di albero che mostra come i cluster sono composti.

CAPITOLO 3

Metodologia di ricerca

La metodologia di ricerca scelta è una revisione sistematica della letteratura (SLR), in quanto questo modus operandi permette di identificare, valutare ed analizzare la letteratura pubblicata finora con l'obiettivo di indagare su un campo di ricerca specifico.

La SRL è formata principalmente da una serie di tecniche utilizzate durante il processo di ricerca: definizione della query di ricerca, definizione dei criteri di inclusione ed esclusione, definizione dei criteri di qualità ed infine estrazione dei dati.

3.1 Obiettivo e domande di ricerca

L'obiettivo della SLR è individuare in che modo tecnologie come il Machine Learning e il Data Analysis, siano attualmente utilizzate a supporto della Football Analytics, in che modo la sta migliorando e come possono esser affrontate questioni limitanti come la privacy e la validità dei modelli.

La fase di ricerca ha avuto inizio con l'identificazione di due domande ben specifiche, che rappresentano i cardini su cui tutto il processo di ricerca si è basato:

- *Quali metodi di Machine Learning e Data Analysis vengono utilizzati a supporto della Football Analytics?*
- *Quali vantaggi e svantaggi comportano queste tecnologie nel mondo del calcio?*

Nella prossima sezione vedremo come, grazie a queste domande, si è arrivati a ravvisare le keywords indicate per formulare la query di ricerca.

3.2 Formulazione delle query di ricerca

Per effettuare le ricerche è stato selezionato Google Scholar, un database ampiamente riconosciuto ed utilizzato per la ricerca accademica; permette la ricerca di articoli scientifici, tesi, libri e altri tipi di pubblicazioni scientifiche.

Sono stati considerati i termini “Football Analytics”, “Machine Learning” e “Data Analysis” in lingua inglese, includendo sia tutte le rispettive abbreviazioni e i loro sinonimi al fine di ricavare un numero di risultati maggiore. Come sinonimi sono state considerate le parole: Soccer, Match Analyst, Artificial Intelligence, AI, ML, Deep Learning, DL, Natural Language Processing, NLP.

La query è stata formulata unendo attraverso un operatore logico OR, tutti i termini facenti parte dello stesso concetto, mentre tramite operatore logico AND sono state accorpate le unioni ottenute.

Per estrarre i dati dal database di Google Scholar si è arrivati a tale formulazione:

**(‘Football Analytics’ OR ‘Soccer’) AND (‘Machine Learning’ OR ‘ML’) AND
('Data Analysis' OR 'Match Analyst') AND ('Artificial Intelligence' OR 'AI') AND
('Deep Learning' OR 'DL') AND ('Natural Language Processing' OR NLP)**

3.3 Estrazione e pulizia dei dati

La query nel database di Google Scholar ha prodotto 1830 risultati.

Per rimuovere gli articoli non pertinenti all’obiettivo della ricerca, sono stati applicati criteri di esclusione e inclusione, permettendo di filtrare le risorse che rispondono alle domande di ricerca di una revisione sistematica della letteratura, evitando così risultati che non rispecchiano lo scopo della ricerca.

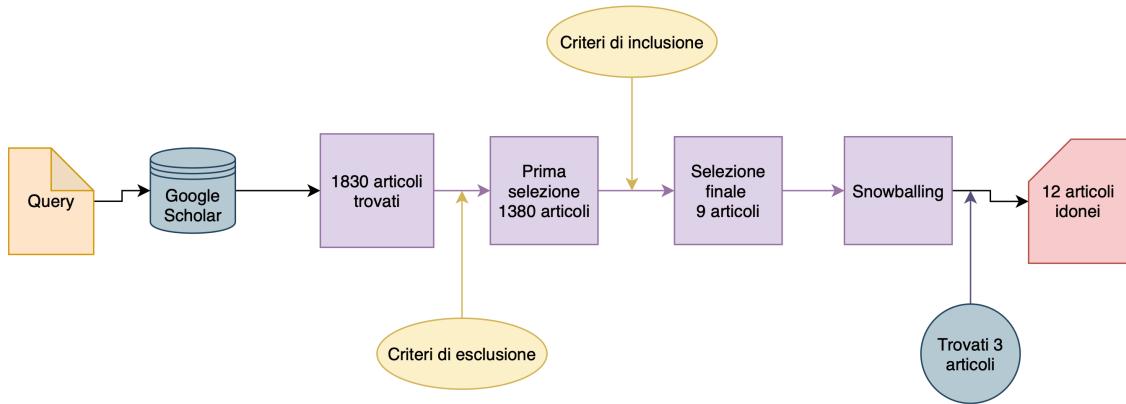


Figura 3.1: Processo di estrazione.

In questo studio, quelli che sono i vincoli facenti parte dei criteri di esclusione degli articoli sono rispettivamente:

- articoli non in inglese;
- articoli prodotti prima del 2015;
- elaborati formati da meno di 3 pagine;
- elaborati come tesi e sondaggi;
- articoli non accessibili completamente;

Grazie a questi criteri sono stati esclusi 450 articoli, dopodiché per i criteri di inclusione sono stati considerati gli articoli che applicavano metodi di ML/AI nel contesto calcistico e della Football Analysis.

Tutta questa fase di selezione è avvenuta manualmente, dove nella fase finale sono stati considerati oltre ai titoli anche i vari abstract, arrivando all'idoneità dei seguenti 9 articoli:

Tabella 3.1: Risultato della selezione

Game Plan: What AI can do for Football, and What Football can do for AI[3]
Performance evaluation of the English Premier Football League with data envelopment analysis[4]
Artificial intelligence and machine-learning approaches in sports: Concepts, applications, challenges, and future perspectives[5]
Artificial intelligence and machine learning in sport research: An introduction for non-data scientists[6]
In-game behaviour analysis of football players using machine learning techniques based on player statistics[7]
Sports analytics review: Artificial intelligence applications, emerging technologies, and algorithmic perspective[8]
Mixing data analytics with intuition: Liverpool Football Club scores with integrated intelligence[9]
Artificial intelligence (AI) in sports[10]
A Comparative Study for Football Analytics With Data Mining and Artificial Intelligence Techniques[11]

Tabella 3.2: Criteri di inclusione ed esclusione

Criteri di esclusione	Criteri di inclusione
Non è scritto in lingua inglese	È scritto in lingua inglese
Prodotto prima del 2015 (incluso)	Prodotto dopo il 2015
Non è formato da almeno 3 pagine	È formato da almeno 3 pagine Illustra metodi di Data Analysis e ML a supporto della Football Analytics

Tale processo ha portato di conseguenza alla fase di snowballing che vedremo nella prossima sezione.

3.4 Fase di Snowballing

Lo Snowballing è una tecnica di ricerca usata per individuare articoli rilevanti su uno specifico argomento; tale metodologia prevede l'uso delle liste di riferimenti bibliografici o citazioni di un articolo per identificare ulteriori articoli, escludendo quelli che non soddisfano i criteri di esclusione e inclusione precedentemente stabiliti come la lingua, il tipo, e l'anno di pubblicazione.

Con una sola iterazione, tale processo ha portato all'identificazione dei seguenti 3 articoli:

Tabella 3.3: Risultati della prima (e unica) iterazione della fase di snowballing

Applications of artificial intelligence in the game of football: The global perspective[12]
Conceptual structure and current trends in artificial intelligence, machine learning, and deep learning research in sports: a bibliometric review[13]
Deep Learning and Transfer Learning Architectures for English Premier League Player Performance Forecasting[14]

3.5 Analisi dei dati

Prima del processo di estrazione ed analisi dei dati, sono state valutate la qualità e la completezza dei documenti che sono stati selezionati nelle fasi precedenti, formulando le seguenti domande:

- *I metodi di Machine Learning e Data Analysis si evincono chiaramente?*
- *I vantaggi/svantaggi date da queste tecnologie implementate nel calcio e precisamente campo della Football Analytics vengono descritti in modo chiaro?*

Per ognuna di queste domande è stato associato un valore numerico che valuti la qualità e la completezza di ogni fonte. Ogni domanda ha tre possibili risposte esclusive: "Sì", "Parzialmente" e "No". Alla risposta "Sì" è stato attribuito un valore di 1, a "Parzialmente" un valore di 0.5 e a "No" un valore di 0. Il punteggio complessivo è stato calcolato sommando i punteggi delle risposte alle due domande, accettando gli articoli con un punteggio almeno pari a 1.

Dopo aver effettuato gli opportuni calcoli, si è arrivati ad individuare i seguenti articoli [3.4] che presentano un punteggio inferiore ad 1, e pertanto sono stati esclusi.

Tabella 3.4: Articoli che presentano un punteggio complessivo inferiore ad 1

A Comparative Study for Football Analytics With Data Mining and Artificial Intelligence Techniques[11]

Deep Learning and Transfer Learning Architectures for English Premier League Player Performance Forecasting[14]

CAPITOLO 4

Applicazioni pratiche

Questo capitolo riporta applicazioni specifiche di Machine Learning e analisi dei dati nel calcio, illustrando casi di successo e impatti tangibili nei vari aspetti calcistici più importanti.

4.1 Scouting

Nel calcio odierno, il ruolo di un giocatore non viene più visto come una posizione ma bensì una funzione. In passato ad esempio, esisteva il trequartista come posizione da ricoprire, oggigiorno invece esistono giocatori che durante la partita, in vari momenti, vanno a fare i trequartisti. E per questo motivo si è passati da giocatori che hanno delle posizioni fisse da rispettare a giocatori che svolgono determinati compiti e determinate funzioni. Questo cambiamento ha portato il concetto di calciatore ad essere sempre più universale e camaleontico in base alle situazioni, rendendolo capace di dimostrare di più le sue qualità a seconda delle varie mansioni da svolgere. Ora un club quando si trova davanti ad un nuovo giocatore che deve valutare per un acquisto o meno, non si focalizzerà nel comprendere tale calciatore che posizione ricopre, se ad esempio è un terzino o una ala, ma piuttosto si concentrerà nel carpire principalmente quali siano le sue qualità migliori: se è un “costruttore” o

un “invasore”, se in base ai loro moduli, in questo calcio diventato via via sempre più fluido, può spingere al massimo le sue capacità offensive e così via.

Un approccio di scouting moderno che si adatta al calcio contemporaneo non parte più dai tradizionali ruoli ma si concentra sulle funzioni e sulle skills dei giocatori. Tali valutazioni sono state rese decisamente più veloci ed efficaci grazie all’utilizzo dei dati statistici, allo sviluppo di tecniche moderne di Machine Learning e alla metodologia del clustering dei giocatori, con il quale vengono classificati proprio in base alle loro funzioni sul terreno di gioco. Per citare qualche esempio più pratico, se un club ha ceduto un “One to one explorer” (che eccelle nei duelli individuali uno contro uno) e ha bisogno di trovarne un altro, non è importante che sia stato impiegato come ala in uno schema 4-3-3 o come esterno del 4-4-2, ma deve possedere capacità molto simili al giocatore appena ceduto, ed è per questo motivo che i cluster giocano un ruolo fondamentale.

Tramite questi studi[15] da parte dei club si è scoperto che un giocatore come Joao Cancelo, definito da chiunque come terzino, possa trovarsi nel cluster come “Chance creator” in quanto sostanzialmente la sua funzione in campo è quella di playmaker, indipendentemente da quella che è la sua posizione di partenza.

Non sono le grandi squadre, ma l’AI può aiutare nello scouting a livello globale. Tecniche come il Natural Language Processing[16] possono essere usate per tracciare senza sosta notizie e articoli di ogni lingua da tutto il globo, permettendo ai club di scovare giovani promesse e battere la concorrenza sul tempo. Oltre all’aspetto prestazionale dei calciatori o del team, l’AI può essere di supporto soprattutto nella valutazione economica dei giocatori, permettendo tramite tecniche ed algoritmi di analisi predittiva di stimare il potenziale valore di mercato di un giocatore sia in base alle sue performance e sia in base ad altri parametri esterni, diventati oramai indispensabili come il seguito sui social media e quanto appeal possiede verso i brand ed operazioni commerciali simili.

In sostanza le squadre adesso hanno una visione decisamente più precisa e chiara sulla gestione dei loro budget e sulle offerte da formulare ai procuratori, e con queste tecnologie il calcio odierno sta riducendo tutti quelli che sono i flussi intermediati presenti durante ogni singolo processo di acquisto di un nuovo giocatore.

Di seguito verranno illustrati esempi riguardo le applicazioni tecnologiche nello

scouting calcistico mondiale odierno.

4.1.1 Scout Advisor

Il Siviglia ad inizio 2024 ha ufficializzato la partnership con IBM ed utilizzando la loro AI Generativa proprietaria hanno dato vita a “Scout Advisor”, uno strumento che migliora il processo di reclutamento dei giocatori. Questo tool è stato costruito su “watsonx”, la piattaforma AI e dati di IBM, ed è stato progettato per integrarsi con la suite esistente di strumenti di generazione dei dati del club. Il dipartimento dati del club ha collaborato per mesi con il team di ingegneria dei clienti di IBM per sviluppare questa applicazione, utilizzando l’elaborazione del linguaggio naturale (NLP) e i modelli di base per analizzare grandi quantità di informazioni nel database di scouting. Questo include sia dati quantitativi come altezza, peso, velocità, numero di gol e minuti giocati, sia dati qualitativi non strutturati come le analisi scritte in oltre 200.000 rapporti di scouting del club degli ultimi decenni.

Le capacità NLP di Scout Advisor permettono al club di usare più modelli di linguaggio di grandi dimensioni (LLM) per migliorare l’accuratezza e l’efficienza nell’identificazione dei giocatori. Basandosi sulle descrizioni degli scout, egli genera elenchi di candidati in base ai tratti ricercati e riassume i rapporti per ogni giocatore. Inoltre, collega ogni giocatore al database attuale per fornire ulteriori dettagli sui loro dati di prestazione quantitativi.

L’integrazione di "watsonx" nel sistema attuale ha portato un enorme aumento del ROI (Return On Investment) del Siviglia. Scout Advisor riesce a fornire supporto cruciale nelle decisioni chiave in tempi davvero brevi.

4.1.2 Data Driven Scouting

Per la realizzazione di questa ricerca ho avuto l’opportunità di entrare in contatto con Léon, uno studente dell’Università di Scienze Applicate di Pforzheim in Germania, che insieme a un collega dell’Università di Scienze Applicate di Regensburg (OTH), ha sviluppato un innovativo strumento di scouting per il calcio. Questo tool, concepito durante la pandemia, mira a migliorare l’efficienza e la precisione nello

scouting dei giocatori, riducendo gli errori di trasferimento e i flop dal punto di vista dei club.

Léon e il suo collega hanno iniziato con l’idea di facilitare lo scouting specifico delle squadre. Dopo ricerche approfondite e colloqui con esperti del settore, hanno cercato una formula per ridurre gli errori di trasferimento. Con il supporto del dipartimento di imprenditorialità dell’OTH, noto come “Start Up Lab”, hanno partecipato al concorso Danube Cup, raggiungendo la finale. Attualmente, stanno collaborando con uno scout croato per migliorare la loro rete di contatti e penetrare nel mercato dei Balcani.

L’obiettivo principale dello strumento è fornire un algoritmo e una base di dati modellati sulle esigenze specifiche dei club. Non si limitano alle statistiche avanzate dei giocatori, ma cercano talenti che si adattino a schemi tattici specifici. Se una squadra come il Benfica si focalizza su talenti del Sud America e del Portogallo, lo strumento raccoglie e analizza una vasta quantità di dati su questi giocatori. Un esempio pratico è l’analisi degli attaccanti: uno con un valore di xG (eXpected Goals) di 17 che ha segnato solo 5 gol, e un altro con un xG di 9 che ha segnato 12 gol. Il club potrebbe essere tentato di preferire il secondo attaccante per il suo rendimento. Tuttavia, lo strumento considera anche il contesto di squadra, come le occasioni create, il possesso palla e le tattiche utilizzate. Questo permette di valutare se un attaccante con meno gol potrebbe essere più efficace in una squadra diversa.

Lo stesso principio si applica ai difensori. Ad esempio, un difensore che ha vinto il 75% dei suoi tackle deve essere valutato nel contesto delle situazioni di gioco della sua squadra. Se il valore dei tackle è più alto quando la squadra è in vantaggio o in pareggio, questo può indicare una maggiore solidità difensiva. Mentre i dati avanzati sono essenziali, non sempre forniscono un quadro completo delle circostanze di gioco. Come sottolineato da Léon è fondamentale integrare questi dati con analisi qualitative ed analisi che contestualizzano, per comprendere appieno le performance di un giocatore. Avere dati sui compagni di squadra e sulle decisioni tattiche del mister può offrire una prospettiva più accurata delle capacità di un giocatore.

Lo strumento sviluppato da Léon e dal suo collega è progettato per inviare ai club dati dettagliati e analisi sui giocatori, rendendo le decisioni di scouting più informate. Questo approccio integrato e personalizzato migliora l’accuratezza nella selezione

dei giocatori, ottimizzando anche in questo caso il ROI e supportando le operazioni sportive dei club.

4.1.3 Il caso Philadelphia Union

Gli Union, squadra della Major League Soccer (MLS), ha adottato un approccio innovativo al reclutamento dei giocatori grazie all’uso della data analytics. Un elemento chiave è l’algoritmo “Philadelphia Union Value” (PUV), sviluppato internamente, che pesa le statistiche dei giocatori in base alle caratteristiche necessarie per il loro sistema di gioco, facilitando scelte di mercato più mirate. Questo approccio permette agli Union di ottenere un vantaggio competitivo nonostante il budget limitato, come dimostrato dall’acquisizione di Dániel Gazdag, un centrocampista ungherese proveniente da un campionato minore, individuato proprio grazie all’algoritmo PUV.

La filosofia del club si basa su tre pilastri: sviluppare talenti interni, promuovere la coesione del team e innovare costantemente, da qui l’impegno nell’investire nella data analytics con (come quanto detto precedentemente) l’obiettivo di acquisire un vantaggio competitivo non da poco rispetto ai grandi club che posseggono budget tre volte più grandi quello di Philadelphia.

4.2 Match Analysis

Lo sviluppo massiccio di Big Data porterà capacità strabilianti con l’AI nell’analisi dei dati storici di partite delle squadre che utilizzano queste tecnologie. Tali club possono confrontare i propri dati delle precedenti partite con quelle delle squadre che affronteranno, in modo da raccomandare allenamenti, formazioni e strategie ottimali per esaltare i propri punti di forza e sfruttare al meglio le vulnerabilità fisiche e strategiche degli avversari.

Tramite processi di analisi in real time, queste tecnologie possono fungere da supporto ad un reale assistente, dando raccomandazioni precise su come aggiustare gli schemi, proponendo innumerevoli scenari da poi discuterne col CT portando maggiore consapevolezza.

Unendo l'utilizzo di dati storici alle varie tecniche di Deep Learning, l'Intelligenza Artificiale può persino prevedere le strategie avversarie tramite l'identificazione di modelli ricorrenti nelle tattiche avversarie, importante per le squadre durante la preparazione di piani di gioco partita dopo partita. Tutti questi dati in principio venivano registrati tramite le varie in-venue cameras¹ presenti negli stadi. Adesso invece, oltre questi strumenti presenti in loco, grazie alla modellazione AI è ora possibile raccogliere tutti questi dati letteralmente da qualsiasi match registrato, da una semplice trasmissione televisiva, capace persino di prevedere dove si trovano i giocatori e tracciarli anche quando si trovano fuori inquadratura.

Ciò ha reso il calcio più “data-driven” e molto più intrattenente, aspetti che vedremo in seguito.

4.2.1 Dispositivi indossabili

I giocatori indossano più che delle semplici maglie. I dispositivi indossabili per il monitoraggio delle prestazioni, come i localizzatori GPS e i cardiofrequenzimetri, forniscono dati in tempo reale sulla forma fisica e lo sforzo del giocatore. Queste informazioni aiutano gli allenatori a gestire i carichi dei giocatori e a ridurre al minimo il rischio di infortuni. Grazie alla miniaturizzazione dei dispositivi indossabili, raccogliere i dati delle performance non è stato mai così semplice: basta una semplice canotta o un activity tracker. Attualmente però, le leggi della FIFA, ovvero l'ente che governa il calcio mondiale, non permette l'utilizzo di wearables durante le partite ufficiali e per questo motivo tutti i dati raccolti dalle società provengono esclusivamente dagli allenamenti quotidiani.

Un caso studio di successo in questo campo è sicuramente quello di PlayerMaker, azienda creatrice di un sistema di valutazione proprietario chiamato PRS (Proprietary Rating System), sviluppato con l'obiettivo di fornire valutazioni oggettiva delle abilità dei calciatori. Questo PRS si basa su molteplici parametri, tecnologie e funzioni come ad esempio:

¹Specifiche telecamere usate per catturare immagini e video per vari scopi operativi e/o promozionali.

- **Tecnologia di Tracciamento.** Playermaker ha sviluppato un dispositivo indossabile montato sul piede (teoricamente chiamato “IMU”, che verrà definito in seguito) per tracciare i dati di movimento degli atleti.²
- **Raccolta e Analisi dei Dati.** Il PRS raccoglie dati dalle attività di allenamento e dalle partite reali per fornire valutazioni oggettive, valutando sia le abilità fisiche che tecniche dei giocatori durante scenari reali di gioco, come ad esempio controllo a due piedi, primo tocco, dribbling, velocità, agilità e molti altri.
- **Benchmarking.** Tramite la companion app fornisce un contesto per comprendere le valutazioni rispetto ai benchmark di età e genere, oltre a suggerimenti per migliorare le abilità.
- **Metodologia del Sistema.** Il sistema si basa su misurazioni oggettive supportate da AI, registrate durante le situazioni di gioco senza interferenze, e sviluppate con benchmark globali.
- **Segmentazione dei Giocatori.** I giocatori sono segmentati per genere, età e categoria di posizione per una comparazione più accurata.
- **Tecnologia Cloud.** Utilizza la tecnologia cloud per fornire risultati comparativi immediati, rappresentati in grafici radar per una visione complessiva delle abilità critiche dei giocatori.

Inoltre l’azienda tramite il loro Whitepaper[17] approfondisce il prodotto da loro venduto sia come dispositivo IMU in sé e per sé, ma anche sul suo funzionamento riguardo gestione dati e il Machine Learning.

IMU sta per Inertial Measurement Unit, un dispositivo elettronico che misura e riporta la velocità specifica, l’orientamento e le forze gravitazionali utilizzando una combinazione di accelerometri, giroscopi e talvolta magnetometri. Nei sistemi di monitoraggio delle prestazioni sportive, come quelli sviluppati da Playermaker, l’IMU è utilizzato per tracciare i movimenti del giocatore in modo dettagliato e preciso. Questi dati, che vengono anche chiamati biomeccanici, vengono raccolti

²Oltre che ad essere lo strumento con cui l’azienda monetizza, aumentando il loro ROI basato dopo anni e anni di investimenti per la ricerca e lo sviluppo.



Figura 4.1: Illustrazione del dispositivo IMU creato da PlayerMaker, con la rispettiva companion app dove è possibile visualizzare tutti i dati dell'atleta.

Fonte: playermaker.com/blog/cityplay-review/

tramite questo IMU montato sul piede del giocatore. Nel giro di pochi anni l'azienda è riuscita a raccogliere ed analizzare (tramite algoritmi avanzati di ML per trasformare dati grezzi in valutazioni numeriche oggettive) dati pari a 6500 partite con circa 2000 giocatori, ovvero sono arrivati a coprire 7,1 miliardi di metri e 267 milioni di tocchi di palla. Quantità enormi di informazioni che vengono raccolti, analizzati, confrontati e forniti in tempo reale in cloud tramite grafici che aiutano ogni singolo giocatore a comprendere le proprie capacità e le aree di miglioramento.

4.2.2 Strumenti di visualizzazione dati

I numeri da soli possono essere travolgenti. Qui entrano in gioco gli strumenti di visualizzazione dei dati[18], trasformando statistiche complesse in grafici comprensibili. Queste immagini rendono più facile sia per i giocatori che per gli allenatori cogliere le intuizioni derivate dai dati.

Secondo il CEO di Real Analytics Ian McHale, azienda produttrice dell'omonima piattaforma basata sull'AI per l'industria dell'analytics calcistica sovvenzionata da Nvidia per coprire i costi di calcolo, di quanto oramai un club facente parte della Premier League[4] potrebbe guardare i campionati sudamericani e vedere come i

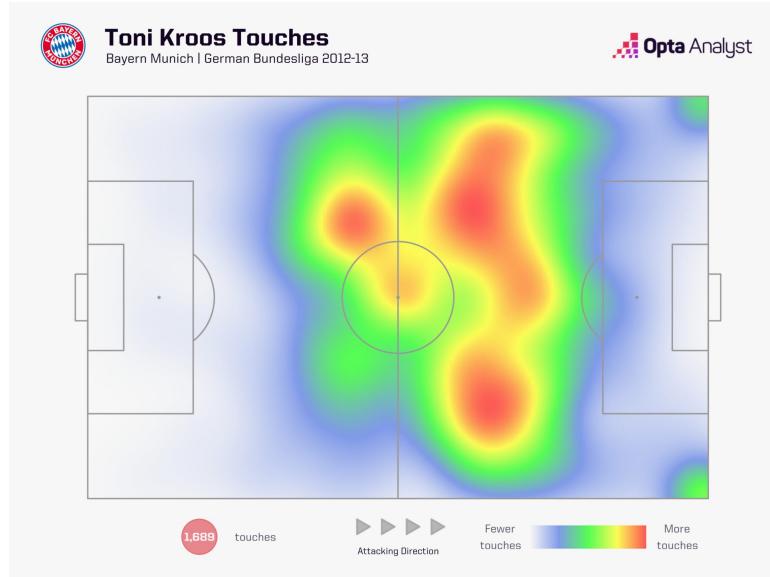


Figura 4.2: Detta "heatmap", è una rappresentazione grafica di dati dove i valori vengono indicati da variazioni di colori, molto utilizzata in quanto facilita la comprensione e l'interpretazione dei diversi pattern.

Fonte: theanalyst.com

giocatori si comportano con e senza la palla, come corrono, e come reagiscono quando sono sotto la pressione dei giocatori avversari.

Avere una visione reale di come il giocatore potrebbe adattarsi nel campionato più veloce del mondo diventa fondamentale.

4.2.3 AIDA: Artificial Intelligence Data Analyst

La nota azienda Soccerment ha creato una nuova applicazione chiamata AIDA, che fornisce analisi calcistiche complesse con l'ausilio dei Large Language Models (LLM). Essa si prefigge di ridurre il gap tra la complessità di tale sport con il tempo necessario per approfondire altrettante complesse dinamiche di gioco, uno strumento che può agevolare soprattutto i match analyst e vari professionisti del settore a prendere decisioni più informate basate sull'oggettività dei dati.

AIDA vuole porsi nell'esatto centro tra l'utente e gli LLMs, garantendo la produzione di risultati ben orientati, minimizzando i rischi di "allucinazione", ovvero fornire risposte errate e fuori contesto, problema tipico degli LLMs quando si richiede un eccessivo sforzo cognitivo.

La sua tecnologia è attualmente basata sui modelli “GPT-3.5 Turbo - 16k” e “GPT-4 Turbo 128k” accessibili tramite le API di OpenAI, permettendo una comprensione e generazione di testo molto sofisticata. Soccerment possiede un dataset enorme e in continua evoluzione, includendo dati provenienti da 29 campionati, ovvero circa 15 mila giocatori, in modo da avere ampio contesto e analisi sempre rilevanti.

4.3 Alimentazione, Prevenzione e Riabilitazione

Non solo in campo, ma ogni aspetto dei giocatori anche fuori dal campo viene misurato e analizzato con il fine di perfezionare le loro prestazioni ed ottimizzare il loro allenamento. Grazie all’analisi di training passati, delle caratteristiche e predisposizioni genetiche e delle ore di sonno³, si può indicare in modo minuzioso l’alimentazione che tale giocatore in quella settimana o quel periodo dovrebbe seguire.

Nel caso di prevenzione di infortuni, l’AI può svolgere un ruolo fondamentale.

Con la raccolta dei dati relativi al carico di allenamento, alle condizioni fisiche dei giocatori e allo storico degli infortuni, si possono identificare i fattori di rischio e personalizzare i programmi di allenamento per minimizzare il rischio di lesioni. Infatti l’AI può persino aiutare a determinare se un giocatore sta sforzando troppo una precisa parte del corpo, e quindi se ha bisogno di riposo, per prevenire ciò che nel ramo della fisioterapia e della medicina dello sport viene definito overtraining.

Ad esempio il Liverpool FC per prevenire infortuni utilizza Zone7, software di una società americana di AI con sede nella Silicon Valley. Tale software analizza dati di gioco, allenamento, livelli biometrici e stress, fornendo segnali di rischio e raccomandazioni pratiche [9]. Da quando il Liverpool ha adottato questa tecnologia, il numero di giornate perse per infortunio è diminuito di oltre un terzo.

Qualsiasi turnover⁴ del Liverpool non è mai casuale ma basato su dati raccolti da Zone7. Questo approccio sta permettendo ai così detti Reds di giocare tutte le partite della stagione riducendo al minimo il rischio di infortuni, dimostrando come l’uso

³Monitorato tramite appositi smartwatch o il noto anello OURA, popolare nel settore sportivo.

⁴Rotazione strategica dei titolari tra le varie partite al fine di dosare le forze e limitare gli affaticamenti dei giocatori.

intelligente della tecnologia possa dare un vantaggio competitivo e ottimizzare il benessere e le prestazioni dei giocatori. Riguardo la riabilitazione, l’AI può supportare la creazione di programmi di recupero personalizzati, tramite algoritmi di ML che analizzano i dati sul progresso del recupero di un giocatore e permettendo ai fisioterapisti di adattare i programmi in base alle esigenze specifiche del giocatore, monitorando l’efficacia del trattamento valutando il momento ottimale per il ritorno in campo. Riabilitazioni di alto livello, spesso integrano sistemi di realtà virtuale che creano tramite AI degli ambienti di simulazione sicuri in cui mentre il giocatore si allena durante il recupero, vengono costantemente analizzati i suoi movimenti per fornire feedback immediati allo staff medico, in modo da migliorare le cure in base alle problematiche.

4.4 Mental Training

In Europa, club come PSV Eindhoven e AZ Alkmaar stanno investendo in test cognitivi per migliorare i loro rispettivi settori giovanili.

Sviluppati da BrainsFirst, questi test si basano su giochi che utilizzano le neuroscienze per misurare le performance cognitive dei giovani giocatori al di fuori del campo e l’AI per analizzare dati complessi in un tempo irrisorio. Il main goal è quello di individuare le capacità cerebrali necessarie per eccellere ad alti livelli nel calcio, misurando circa 53 fattori cognitivi. I risultati rinvenuti da questi test aiutano le squadre nel carpire i talenti che presentano caratteristiche cognitive necessarie per diventare dei veri e propri campioni.

Secondo i neuroscienziati le capacità cognitive cruciali possono prevedere il potenziale futuro dei giocatori. Il ruolo delle neuroscienze è fondamentale in quanto fornisce un metodo oggettivo per valutare i giovani calciatori eliminando i pregiudizi nel processo di scouting. Questa tecnologia ha permesso a club come quelli sopracitati di ottenere successi significativi, come la vittoria nella UEFA Youth League e della Premier League International Cup.

4.5 Arbitraggio Virtuale

Con l'avvento oramai evidente del VAR e della goal-line technology questi strumenti hanno reso il calcio più oggettivo e preciso da un punto di vista delle decisioni arbitrali. Non è ancora chiaro se in un futuro non troppo lontano vedremo gli arbitri sostituiti completamente da sistemi di arbitraggio virtuale, ma ciò che ci appare più probabile è che col tempo il rapporto che sta avvenendo tra l'arbitro umano e gli assistenti virtuali basati sull'AI sta diventando via via sempre più consolidato. Assistenti che possono consigliare decisioni accurate da prendere, analizzando quantità ingenti di dati, dal tracking dei movimenti dei giocatori, alle possibili infrazioni di essi o ai contatti sospetti.

Evoluzioni del genere possono cambiare anche l'aspetto "sociale" del gioco: i giocatori avrebbero maggiore fiducia nelle decisioni arbitrali in quanto supportate da analisi di dati oggettivi e verrebbero inoltre scoraggiati dal simulare falli o situazioni di gioco, dato che, con una sorveglianza così sofisticata la probabilità di esser scoperti diventa esponenziale. Ambedue queste caratteristiche renderebbero sia il gioco più intrattenente, facendo diminuire perdite di tempo per via di lamentele e proteste, e soprattutto porterebbe ad un gioco più pulito, ponendo al centro le vere abilità dei giocatori a discapito di tattiche fasulle e comportamenti malevoli.

4.6 Fan Engagement

L'analisi del calcio non è solo per le squadre ma è anche per i tifosi. Con l'avvento delle app e dei siti web basati sui dati, i sostenitori possono accedere a statistiche e analisi tecnico-tattiche ben approfondite, migliorando la loro comprensione del gioco e favorendo una profonda immersività con le loro squadre preferite.

Che si tratti di video, podcast, testi, o altri tipi di contenuti oramai fruibili da chiunque, tutto viene adattato in termini di argomento, lunghezza, durata e stile per rispondere alle preferenze dell'utente. Come nel settore riabilitativo, anche nel mondo dell'engagement sportivo l'AI unito alla realtà virtuale (o aumentata) sta portando a sviluppi e scoperte interessanti.

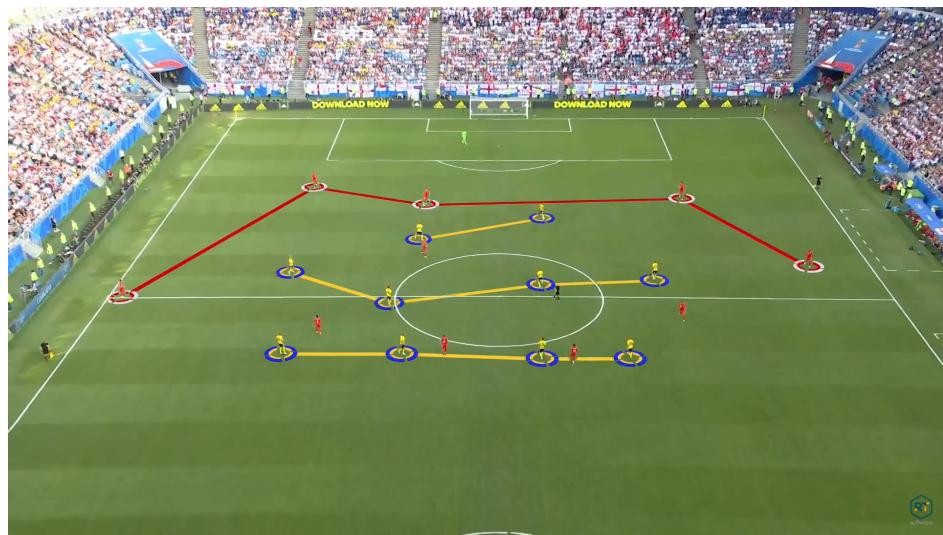


Figura 4.3: Questa immagine mostra come l'emittente finlandese YLE utilizza un software AI-based chiamato “Tactic PRO” per tutta la sua copertura televisiva di Euro 2024.

Fonte:

sportsvideo.org/2024/06/20/broadcaster-yle-adopts-tactic-pro-for-euro-2024



Figura 4.4: Questa immagine mostra l'app “Immersive” per Apple Vision Pro, applicazione che grazie alla realtà aumentata arricchisce l'esperienza del tifoso rivoluzionando il broadcasting odierno.

Fonte: [reddit.com/r/augmentedreality](https://www.reddit.com/r/augmentedreality)

Applicativi come “Immersiv” permettono ai fan di vivere le partite con un approccio decisamente rivoluzionario, non solo come se fossero effettivamente allo stadio, ma allo stesso tempo gli vengono fornite statistiche secondo per secondo su ogni singolo giocatore.

CAPITOLO 5

Sfide e limitazioni

In questo capitolo verranno stilati ed illustrati quelli che sono i 5 punti fondamentali dove l'uso delle tecnologie raccontate fino ad ora trovano ancora dei limiti o delle criticità.

1. **Privacy dei giocatori e sicurezza dei dati.** La raccolta di dati personali e sensibili dei giocatori, come informazioni sulla salute, abitudini di allenamento, e aspetti personali, può sollevare preoccupazioni sulla privacy. Concentrandoci principalmente nell'ambiente dell'Unione Europea e dell'EFTA¹, i club e le associazioni sportive devono saldamente rispettare il GDPR², assicurandosi che la raccolta e la conservazione dei dati personali dei giocatori, staff e tifosi siano conformi alle normative vigenti. Ovviamente tali norme devono esser rispettate previo consenso e presa di conoscenza da parte dei soggetti interessati sull'uso e la gestione di tali informazioni. Ogni volta che vengono introdotti ed implementati nuovi sistemi e tecnologie che potrebbero comportare gravi rischi per la sicurezza e la privacy dei dati personali (come informazioni biometriche

¹EFTA sta per European Free Trade Association, e include anche stati Extra-UE come Svizzera, Norvegia, Liechtenstein e Islanda.

²GDPR sta per General Data Protection Regulation ed è il regolamento generale sulla protezione dei dati dell'UE, enfatizzando il consenso e la trasparenza dell'uso dei dati nei confronti dell'utente.

tramite sistemi di tracking ad esempio), i club devono svolgere uno specifico processo chiamato DPIA³. Egli permette di quantificare e moderare i rischi per la protezione dei dati prima dell’implementazione.

2. **Equità e discriminazione.** L’uso di algoritmi di Machine Learning potrebbe portare a situazioni di discriminazione nella valutazione dei giocatori o nelle decisioni di selezione basate su caratteristiche demografiche. Per questo motivo è importante monitorare attentamente e regolare gli algoritmi per evitare discriminazioni ingiustificate.
3. **Validità dei modelli.** I modelli di Machine Learning devono essere accurati e validati per garantire che le previsioni e le raccomandazioni siano affidabili. La validità dei modelli può essere compromessa da dati di allenamento non rappresentativi o da cambiamenti nelle condizioni del gioco che non sono stati considerati durante la creazione del modello. Seppur i modelli linguistici sono agli inizi e stanno rapidamente migliorando, in questa fase stanno ancora generando inesattezze casuali, rendendo difficile fare affidamento completamente su di loro.
4. **Accettazione culturale.** Un punto da non sottovalutare minimamente, è che l’introduzione di tecnologie così avanzate nel calcio potrebbe incontrare resistenza da parte di giocatori, allenatori, tifosi o altre parti interessate ad avere un approccio conservativo; in quanto la cultura del calcio, da sempre basata sull’esperienza e sull’intuito, potrebbe richiedere tempo per accettare l’uso diffuso di tecnologie di analisi dei dati, creando dibattiti polarizzanti e schieramenti che possono rallentare o meno il progresso (come ad esempio avviene ancora oggi con la goal-line technology, introdotta nel lontano 2014 durante la Coppa del Mondo FIFA in Brasile).
5. **Interpretazione dei risultati.** I risultati ottenuti attraverso l’analisi dei dati e il ML richiedono un’interpretazione accurata da parte degli allenatori e il loro staff tecnico. Per questo motivo servono formazioni lunghe ed adeguate⁴ per

³DPIA sta per Data Protection Impact Assessment, senza di esso i diritti e la libertà delle persone fisiche coinvolte verrebbero compromessi.

⁴Oltre che continui aggiornamenti.

comprendere e applicare le informazioni generate dai modelli in modo efficace.

Nonostante queste enormi sfide e limiti dati dall'inserimento di tecnologie più o meno acerbe, in un contesto dove solo dopo i primi anni del 2000 ci sono stati piccoli passi in avanti, molteplici sono i punti a favore di tali progressi, sebbene portino ancora con sé dei coni d'ombra da non sottostimare.

Uno tra questi su cui c'è da porre particolare attenzione riguarda la democratizzazione del successo.

Se fino ad oggi il monitoraggio continuo degli atleti era presente solo nel così detto calcio che conta, queste tecnologie ora sono alla portata di tutti, anche tra i non professionisti. Ad esempio, come quelle utilizzate da PlayerMaker illustrate come caso studio nel capitolo 4.2.1, portando una vera e propria democratizzazione di queste tecnologie, soluzioni sempre più economiche, automatizzate e scalabili. Il data-driven approach quindi, permette a squadre con budget inferiori (come nel caso dei Philadelphia Union 4.1.3) di competere con squadre più ricche, grazie all'utilizzo di dati e algoritmi intelligenti.

5.1 Il caso di Kevin De Bruyne

Il centrocampista del Manchester City, Kevin De Bruyne, ha ottenuto il rinnovo del contratto fino al 2025 letteralmente senza avvalersi di un agente, ma utilizzando invece l'analisi dei dati per dimostrare la sua importanza per la squadra. De Bruyne ha commissionato, ad un team di data analyst, di creare un rapporto dettagliato che evidenziasse il suo contributo attuale e potenziale alla squadra, confrontando anche la situazione finanziaria e strategica del club con altre grandi squadre europee. Questa strategia innovativa ha portato De Bruyne a ottenere un significativo aumento di stipendio, arrivando a un totale di circa 24 milioni di euro all'anno, eliminando le elevate spese di commissioni di un agente.

Da questo episodio si evince che l'uso dei data analyst potrebbe segnare un cambiamento radicale nel calcio, ridimensionando il ruolo degli agenti e introducendo un modello salariale basato su statistiche e analisi tecniche. Questo modus operandi potrebbe essere seguito da altri calciatori, portando a negoziazioni contrattuali più

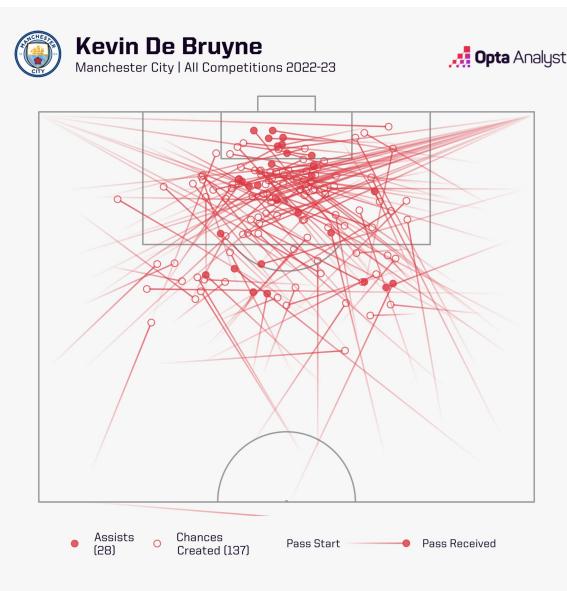


Figura 5.1: Questa grafico delle sequenze d'attacco mostra il numero di coinvolgimenti di De Bruyne nelle sequenze di attacco che poi vengono concluse con un tiro in porta, nella stagione 22/23.

Fonte: theanalyst.com

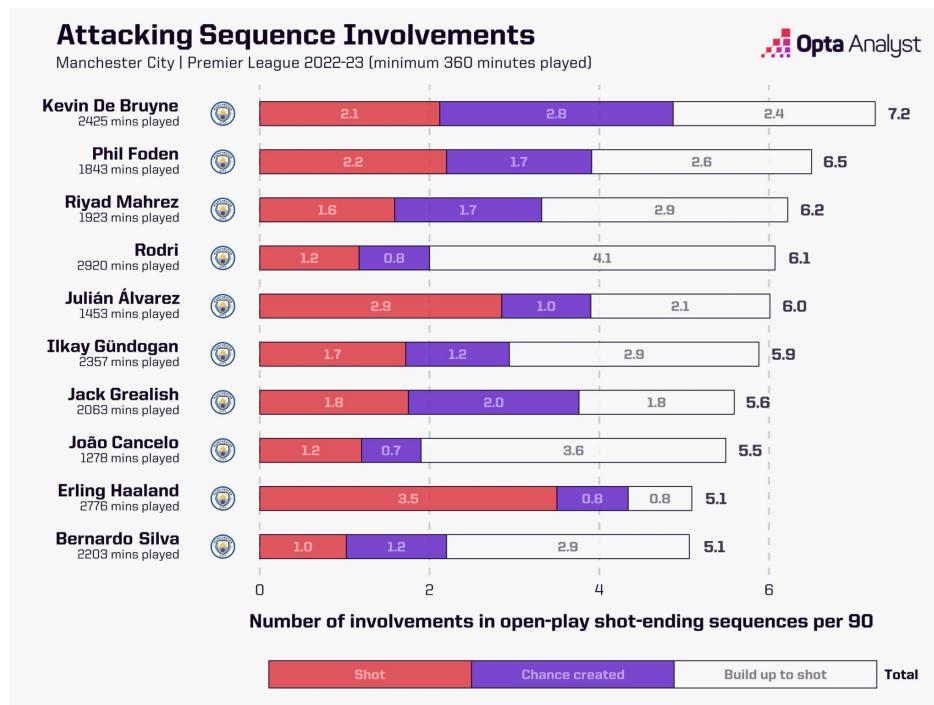


Figura 5.2: Questi dati dimostrano il ruolo cruciale di tale giocatore, sia nelle finalizzazioni offensive e soprattutto nella creazione e costruzione del gioco.

Fonte: theanalyst.com

trasparenti ed efficienti. Tale avvento ha dimostrato come queste tecnologie possano anche rivoluzionare le trattative contrattuali, riducendo i costi e migliorando la meritocrazia nel settore. Esistono alcuni esempi precedenti che includono l'uso di società di analisi dati per trasferimenti, come nel caso di Memphis Depay, ma il metodo di De Bruyne è pionieristico in quanto è riuscito a gestire un negoziato senza un agente. Nel futuro prossimo ciò porterà a un cambiamento significativo nel calciomercato, dove i giocatori potrebbero rappresentarsi da soli, basandosi su dati concreti per giustificare le loro richieste monetarie.

CAPITOLO 6

Conclusioni e Sviluppi Futuri

Una delle più grandi rivoluzioni che avremo nel futuro prossimo nell'ambito di questa ricerca, è la fusione dei dati evento coi dati di tracking 6.1. Ci sarà un avvicinamento sempre più forte alle reali funzioni di un giocatore, unendo informazioni tecniche (come tiri e dribbling), a quelle atletico-fisiche (come numero di scatti) e specialmente alle informazioni tattiche (come ad esempio il numero di ricezioni fra le linee e il numero di inserimenti senza palla).

Questi tramite lo sviluppo dei wearables e la Computer Vision, metteranno in luce aspetti spazio-temporali e strategici senza precedenti. Dato che via via potremo eseguire simulazioni in tempi ancora più brevi e calcoli in pochi istanti di massicce quantità di dati, l'AI aprirà le porte a nuove frontiere tattiche, modellando le dinamiche di gioco con una minuziosità impensabile per un essere umano.

Potrebbe dare in output schemi tattici mai visti prima, magari addirittura controintuitivi o apparentemente bizzarri, modelli di gioco dove è preferibile cedere la palla all'avversario o dove il portiere deve spostarsi in zone poco comuni per tale ruolo.

Già solo l'apprendimento da queste intuizioni potrebbe portare ad una nuova era del calcio, il calcio 4.0, tutto ciò rappresenta una vera e propria sfida per i calciatori, allenatori e staff tecnici che dovranno esser bravi e capaci ad individuare e applicare

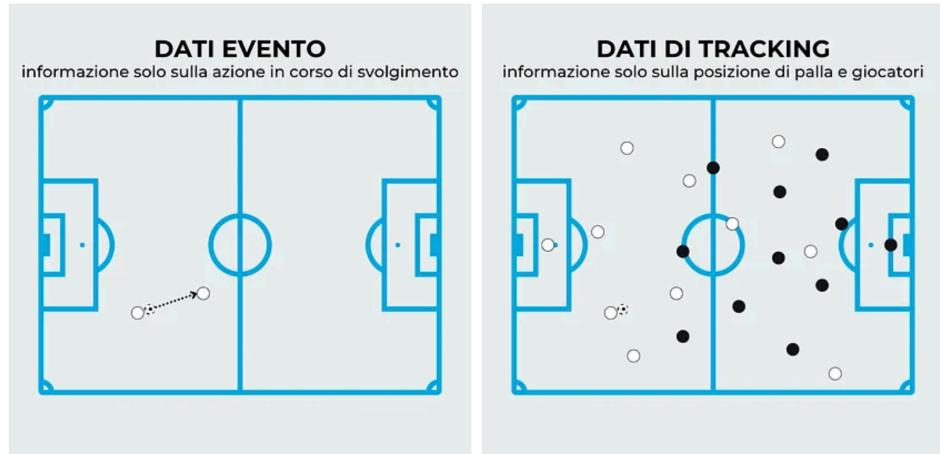


Figura 6.1: Questa immagine mostra la differenza tra dati evento e dati di tracking.

Fonte: calcioanalyst.substack.com

queste nuove strategie prima di tutti gli altri.

Il 3 novembre 2022 è una data che rimarrà per sempre nella storia dell'informatica. Il rilascio al pubblico di ChatGPT da parte di OpenAI ha portato alla consapevolezza che l'AGI (Artificial General Intelligence) è alle porte, ovvero ad un AI che sarà in grado di capire, imparare e applicare tali conoscenze con un approccio molto simile agli esseri umani.

Giorno dopo giorno si assiste a progressi tecnologici sempre più impensabili, fino a che in futuro non saranno più decifrabili. Tutto questo perchè la mente umana ha sempre avuto la tendenza di ragionare in termini lineari piuttosto che in termini esponenziali, abitudine che probabilmente ha fondamenta sin dalla nostra evoluzione. Basti pensare che per milioni e milioni di anni fino al secolo scorso, i mutamenti ambientali avvenivano sempre in modo prevedibile e graduale, mentre oramai non è più così. Questa sorta di differenza tra le nostre intuizioni lineari e la realtà che cresce esponenzialmente ci sorprenderà e ci richiederà un adattamento continuo, mentre la tecnologia evolve ad un ritmo sempre più veloce.

L'analisi del calcio non è più un campo di nicchia, ma un punto di svolta. I team che abbracciano strategie basate sui dati e investono in strumenti e analisi all'avanguardia hanno maggiori probabilità di trionfare sul campo. Nel mondo in continua evoluzione del calcio, l'arte dell'analisi è la chiave per trasformare le statistiche in strategie vincenti. Come visto in precedenza, con il modello De Bruyne 5.1 dei data analyst, un calciatore godrebbe oltre che di un notevole risparmio di

costi, anche di una maggiore efficienza nell’ambito della contrattazione con il club; due fattori importanti in quanto **i numeri, a differenza delle parole dei procuratori, difficilmente possono essere contestati.**

Bibliografia

- [1] S. Russell and P. Norvig, *Artificial Intelligence: A Modern Approach*, 3rd ed. Prentice Hall, 2010. (Citato a pagina 4)
- [2] J. C. Flowers, "Strong and weak ai: Deweyan considerations," in *AAAI Spring Symposium: Towards Conscious AI Systems*, vol. 22877, 2019. (Citato a pagina 4)
- [3] K. Tuyls, S. Omidshafiei, P. Muller, Z. Wang, J. Connor, D. Hennes, I. Graham, W. Spearman, T. Waskett, D. Steel *et al.*, "Game plan: What ai can do for football, and what football can do for ai," *Journal of Artificial Intelligence Research*, vol. 71, pp. 41–88, 2021. (Citato a pagina 12)
- [4] C. P. Barros and S. Leach, "Performance evaluation of the english premier football league with data envelopment analysis," *Applied Economics*, vol. 38, no. 12, pp. 1449–1458, 2006. (Citato alle pagine 12 e 22)
- [5] F. J. Reis, R. K. Alaiti, C. S. Vallio, and L. Hespanhol, "Artificial intelligence and machine-learning approaches in sports: Concepts, applications, challenges, and future perspectives," *Brazilian Journal of Physical Therapy*, p. 101083, 2024. (Citato a pagina 12)
- [6] N. Chmait and H. Westerbeek, "Artificial intelligence and machine learning in sport research: An introduction for non-data scientists," *Frontiers in sports and active living*, vol. 3, p. 682287, 2021. (Citato a pagina 12)

- [7] A. García-Aliaga, M. Marquina, J. Coteron, A. Rodríguez-González, and S. Luengo-Sánchez, "In-game behaviour analysis of football players using machine learning techniques based on player statistics," *International Journal of Sports Science & Coaching*, vol. 16, no. 1, pp. 148–157, 2021. (Citato a pagina 12)
- [8] I. Ghosh, S. Ramasamy Ramamurthy, A. Chakma, and N. Roy, "Sports analytics review: Artificial intelligence applications, emerging technologies, and algorithmic perspective," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 13, no. 5, p. e1496, 2023. (Citato a pagina 12)
- [9] U. Lichtenhaller, "Mixing data analytics with intuition: Liverpool football club scores with integrated intelligence," *Journal of Business Strategy*, vol. 43, no. 1, pp. 10–16, 2022. (Citato alle pagine 12 e 24)
- [10] O. G. Ratiu, D. Badau, C. G. Carstea, A. Badau, and F. Paraschiv, "Artificial intelligence (ai) in sports," in *Proceedings of the 9th WSEAS International Conference on Artificial Intelligence, Knowledge Engineering and Data Bases*, ser. AIKED'10. Stevens Point, Wisconsin, USA: World Scientific and Engineering Academy and Society (WSEAS), 2020, p. 93–97. (Citato a pagina 12)
- [11] M. A. M. Al-Asadi, "A comparative study for football analytics with data mining and artificial intelligence techniques," 2022. (Citato alle pagine 12 e 14)
- [12] K. Rathi, P. Soman, A. V. Koul, and K. Manu, "Applications of artificial intelligence in the game of football: The global perspective," *Researchers World*, vol. 11, no. 2, pp. 18–29, 2020. (Citato a pagina 13)
- [13] C. Dindorf, E. Bartaguz, F. Gassmann, and M. Fröhlich, "Conceptual structure and current trends in artificial intelligence, machine learning, and deep learning research in sports: a bibliometric review," *International Journal of Environmental Research and Public Health*, vol. 20, no. 1, p. 173, 2022. (Citato a pagina 13)
- [14] D. Frees, P. Ravella, and C. Zhang, "Deep learning and transfer learning architectures for english premier league player performance forecasting," *arXiv preprint arXiv:2405.02412*, 2024. (Citato alle pagine 13 e 14)

- [15] S. Research and A. Gagliardi, *The Clustering Project: Un nuovo approccio allo scouting moderno.* Soccerment Research, 2022. (Citato a pagina 16)
- [16] P. Ciancarini, D. D. Russo, D. V. Lomonaco, and G. Zhou, "Machine learning come supporto per la valutazione dei requisiti agili." (Citato a pagina 16)
- [17] E. Morag, E. Amit, M. Gad, H. Maayan, S. Barrett, Y. Wegrzyn, N. Ganz, and N. Levi, "Playermaker's proprietary rating system," Playermaker, UK, Tech. Rep., September 2022. [Online]. Available: <https://playermakerprd.wpengine.com/wp-content/uploads/2023/02/Playermaker-PRS-Whitepaper.pdf> (Citato a pagina 21)
- [18] D. Memmert and D. Raabe, *Data analytics in football: Positional data collection, modelling and analysis.* Routledge, 2018. (Citato a pagina 22)

Questa tesi ha contribuito a piantare un albero in Kenya tramite il progetto Treedom.

<https://www.treedom.net/it/user/sesalab/event/se-sa-random-forest>