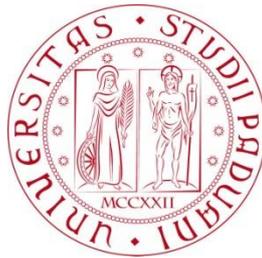


Università degli Studi di Padova
Dipartimento di Scienze Statistiche
Corso di Laurea Triennale in
Statistica per l'Economia e l'Impresa



RELAZIONE FINALE
**MODELLO DI PREVISIONE DEGLI ESITI CALCISTICI
BASATO SU METRICHE RELATIVE AI SINGOLI
GIOCATORI**

Relatore Prof. Matteo Grigoletto
Dipartimento di Scienze Statistiche

Laureando: Tommaso Lazzari
Matricola N2035271

Anno Accademico 2023/2024

Indice

1.Introduzione del problema e analisi della letteratura	1
1.1. Modelli di conteggio proposti.....	1
1.1.1. Modello Poisson	1
1.1.2. Modello Weibull.....	2
1.2. Dinamicità delle valutazioni	2
1.2.1. Ponderamento esponenziale decrescente.....	2
1.2.2. Approccio stocastico	2
1.2.3. Approccio deterministico	3
1.3. Sistemi di valutazione proposti.....	3
1.3.1. ELO-rating.....	4
1.3.2. Pi-rating.....	4
1.3.3. GAP-rating	5
1.4. Introduzione ad un modello player-based.....	6
2. Descrizione dei dati utilizzati, delle correzioni apportate dagli autori e della stima delle interazioni tra giocatori	9
2.1. Dati	9
2.1.1. Valutazioni WhoScored individuali	9
2.1.2. Dati relativi agli eventi dell'incontro.....	10
2.1.3. Quote dei bookmakers.....	11
2.2. Correzione delle valutazioni WhoScored	11
2.2.1. Raw WhoScored rating (RWS)	11
2.2.2. Adjusted WhoScored rating (AWS).....	12
2.3. Includere le valutazioni dei giocatori in un modello di predizione dei risultati di squadra	15
2.3.1. Stima del livello di interazione tra giocatori avversari, escluso il portiere.....	15
2.3.2. Stima del livello di interazione tra giocatori e portiere avversario.....	17
2.3.3. Esempi	18
3. Descrizione dei modelli di previsione	21
3.1. Gestione dei dati	21
3.2. Modello di Skellam	22
4. Valutazione dei modelli sulla base dei risultati ottenuti	25
4.1. Scoring rules.....	25
4.1.1. Score di Brier.....	27
4.1.2. Accuratezza	27
4.2. Scommesse	29
5.Conclusioni	33
6.Riferimenti bibliografici	35

Capitolo 1

Introduzione del problema e analisi della letteratura

I modelli per la previsione di risultati calcistici utilizzano tipicamente lo storico delle informazioni riguardanti le squadre in competizione per ottenerne delle valutazioni. Queste valutazioni vengono poi a loro volta usate per generare le stime delle probabilità dell'esito finale della partita (vittoria, pareggio, sconfitta) o del relativo risultato (0-0, 1-0, 1-1, 0-1, etc.).

Questa relazione prenderà come riferimento l'approccio e le novità proposte da [Holmes e McHale \(2023\)](#).

1.1. Modelli di conteggio proposti

Trattandosi i punteggi di calcio di dati di tipo discreto, i modelli implementati sono di tipo conteggio.

1.1.1. Modello Poisson

Per esempio, quello che può essere considerato il più noto modello per risultati di partite di calcio, il modello [Dixon e Coles \(1997\)](#) (che a sua volta è basato sul modello di [Maher \(1982\)](#)), usa lo storico dei dati riguardanti i goal segnati e subiti per fornire ad ogni squadra una valutazione della capacità offensiva e una valutazione di quella difensiva. Il modello, poi, per stimare la probabilità di un dato risultato del match, stima il numero di goal segnati dalle due formazioni avversarie con l'ausilio di due distribuzioni di Poisson indipendenti, le cui medie sono funzione dell'interazione delle valutazioni delle capacità offensiva assegnata a una squadra, di quella difensiva assegnata alla squadra avversaria e di un parametro, che attribuisce alla formazione che gioca in casa il vantaggio del "fattore campo".

1.1.2. Modello Weibull

Un'altra distribuzione proposta dalla letteratura è la distribuzione bivariata di Weibull di conteggio. La distribuzione di Weibull è una distribuzione discreta che [McShane et al. \(2008\)](#) hanno concluso adattarsi alle variabili casuali di conteggio di eventi che occorrono in un determinato lasso di tempo, quando gli intervalli tra un evento e l'altro possono essere assunti identicamente distribuiti e indipendenti. [Boshnakov et al. \(2016\)](#) hanno valutato significativamente migliori i risultati ottenuti con un modello basato su questa distribuzione, il quale, tra le altre cose, a differenza del modello Poisson, permette di adattarsi anche a dati che presentano sovra-dispersione o sotto-dispersione. Essi hanno poi elaborato *ad hoc* una modifica alla funzione di probabilità, come fatto da [Koopman e Lit \(2015\)](#) per le distribuzioni di Poisson proposte da [Maher \(1982\)](#), per ottenere una variabile bivariata di Weibull in grado di descrivere anche la componente di dipendenza tra i goal segnati dalle due squadre.

Gli studi concernenti modelli di previsione calcistica si sono successivamente focalizzati su come meglio attribuire le valutazioni delle componenti offensiva e difensiva alle squadre (paragrafo 1.3) e di come permettere a queste di variare nel tempo (paragrafo 1.2).

1.2. Dinamicità delle valutazioni

Il modello di [Maher \(1982\)](#) assumeva costanti i livelli di gioco delle squadre: è evidente che non sia realistico assumere che le forze in gioco non siano dinamiche nel tempo.

1.2.1. Ponderamento esponenziale decrescente

[Dixon e Coles \(1997\)](#) applicano un sistema di ponderamento decrescente nella funzione di massima-verosimiglianza in modo tale che minore è il tempo trascorso dalle partite giocate, maggiore è la considerazione che verrà fornita alle informazioni da questa apportate e viceversa.

1.2.2. Approccio stocastico

[Crowder et al. \(2002\)](#), [Owen \(2011\)](#) e [Koopman e Lit \(2015\)](#) adottano modelli che consentono alle valutazioni delle squadre di variare stocasticamente da partita a partita.

I primi e i terzi (i cui modelli si basano sulla doppia distribuzione di Poisson proposta da [Maher \(1982\)](#)) per modellare i valori delle valutazioni del gioco nelle due metà campo delle squadre

propongono un processo autoregressivo di primo ordine, le cui innovazioni, incorrelate serialmente per definizione, permettono che il processo evolva indipendentemente per ogni squadra. Il secondo, invece, applica alle valutazioni un processo Random Walk, per cui le valutazioni del livello offensivo e quello difensivo di una squadra in un tempo $t+1$ si assumono distribuite come una normale con media pari al valore degli stessi parametri stimati al tempo t e varianza pari alla varianza dell'evoluzione di questi parametri, che, per semplicità, è considerata costante nel tempo e comune alle diverse squadre.

1.2.3. Approccio deterministico

[Baker e McHale \(2015\)](#) assumono, diversamente, che queste componenti varino deterministicamente nell'arco di un lungo periodo temporale. Sostengono che l'adozione di un approccio stocastico nella modellazione di queste variazioni nel tempo comporti delle complicazioni. Prima di tutto, in un modello per la stima della forza di una squadra bisogna considerare che potrebbero verificarsi degli importanti cambiamenti repentini, per esempio il cambio di proprietà o di gestione della squadra, la quale potrebbe a sua volta comportare una grossa iniezione di liquidità e conseguenti operazioni sul mercato. Un semplice modello stocastico farebbe difficoltà a gestire degli shock così estremi in intervalli temporali così limitati, essi infatti consisterebbero in una deviazione randomica dalla distribuzione normale, sulla quale sono solitamente basati i modelli stocastici. A questa problematica è possibile, tuttavia, ovviare permettendo alle innovazioni di seguire una distribuzione le cui code risultano essere più pesanti, come una *t di student*. In secondo luogo, per evitare che le stime divergano con un processo *random walk*, la componente stocastica deve essere *mean-reverting* (deve, cioè, presentare la caratteristica di non discostarsi mai eccessivamente da una data media). Ma intorno a quale media dovrebbe "ruotare"? L'utilizzo di una media globale non sarebbe ideale, poiché verrebbe attribuito a tutte le squadre non solo un livello complessivo di gioco, ma anche la proporzione tra capacità offensive e difensive mediamente uguali. Diversamente, si potrebbe attribuire una media specifica ad ogni squadra, ma il livello medio delle singole squadre varia anch'esso nel lungo periodo.

1.3. Sistemi di valutazione proposti

Per quanto riguarda l'attribuzione delle valutazioni alle squadre sono stati proposti diversi metodi, tra i quali ve ne sono alcuni che prevedono un continuo aggiornamento dei *rating* dei giocatori o delle squadre in base alle performance ottenute sul campo.

1.3.1. ELO-rating

Il sistema di rating Elo utilizzato per modelli calcistici, che fu inizialmente sviluppato per fornire una valutazione quantitativa ai giocatori di scacchi ([Elo, 1978](#)), e che, ad oggi, è stato adottato da numerosi altri sport, valuta la solidità delle squadre basandosi sui risultati ottenuti e prevede un metodo per aggiornare le valutazioni delle squadre man mano che si hanno a disposizione nuovi risultati. Per esempio, il sistema, proposto da [Hvattum e Arntzen \(2010\)](#) per l'aggiornamento delle valutazioni, prevede che il valore del rating ad un tempo t sia funzione del rating stesso al tempo $t-1$ e di un fattore, il quale si basa sulla differenza tra un punteggio -pari a 1 (in caso di vittoria), 0.5 (pareggio) o 0 (sconfitta)- e la proporzione di goal della partita segnati dalla squadra di cui si sta aggiornando il *rating* ponderata per un parametro che tiene conto della differenza reti dell'incontro.

I *pi-rating* di [Constantinou et al. \(2012\)](#) e il *GAP rating* di [Wheatcroft \(2020\)](#) seguono in maniera simile il metodo in cui le valutazioni delle squadre vengono aggiornate.

1.3.2. Pi-rating

Questo sistema anziché lo storico delle due squadre, di cui si vuole prevedere l'esito dell'incontro, prende in considerazione tutte le partite – anche di diversi campionati e paesi – in cui la differenza dei *rating* delle due formazioni è corrispondente.

Ciò implica che è possibile che la previsione di un incontro non prenda nemmeno in considerazione informazioni apportate dalle due squadre coinvolte.

In questa maniera si ovvia alle problematiche legate all'assenza di dati per squadre appena promosse o retrocesse ed è possibile utilizzare l'intera osservazione a disposizione perché, mentre non è possibile confrontare squadre con *rating* uguali ma militanti in campionati diversi, è possibile paragonare partite in cui la differenza di valutazione delle squadre è uguale anche se disputate in campionati differenti.

Con la versione originale dei *pi-rating* ([Constantinou et al., 2012](#)) le valutazioni delle squadre sono determinate da: due parametri -uno, che determina quanto una nuova informazione disponibile influenzi la valutazione e un altro, che determina quanto le partite giocate in casa influenzano la valutazione del gioco in trasferta della squadra stessa- e una funzione decrescente -che mira a limitare esponenzialmente l'effetto della differenza goal, valorizzando maggiormente la vittoria indipendentemente dallo scarto nei punteggi.

L'interessante estensione proposta ([Constantinou 2019](#)) prevede l'implementazione del “*form factor*”, che ha l'obiettivo di attribuire alle squadre una valutazione provvisoria in quei brevi periodi in cui sono sotto/sovra-performanti. Per queste valutazioni ci si è avvalsi di un parametro che rappresenta il numero massimo di partite anomale consecutive che non allerta il “*form factor*”, la correzione della valutazione e una funzione che determina la velocità con cui la correzione della valutazione diminuisce per ogni partita consecutiva sotto/sopra le aspettative, basata sull'assunzione che, se la nuova condizione di squadra persistesse nel tempo, la valutazione reale man mano si allineerebbe con quella provvisoria.

1.3.3. GAP rating

I *GAP rating* considerano un campionato in cui militano N squadre, che si scontrano tutte un determinato numero di volte nel corso della stagione sportiva. Per una data partita, S_h e S_a sono una metrica delle performance offensive della squadra di casa e della squadra in trasferta, rispettivamente, dove la definizione di performance offensiva è data dall'utente e derivata da dati statistici, come possono essere, ad esempio, il numero di goal segnati, il numero di tiri in porta, la percentuale di passaggi completati o una combinazione di essi. Ad ogni squadra vengono assegnati quattro *GAP rating* differenti: uno per le capacità offensive in casa (H^a), uno per le capacità offensive in trasferta (A^a), uno per le capacità difensive in casa (H^d) e uno per le capacità difensive in trasferta (A^d).

I *GAP rating* offensivi corrispondono alla performance offensiva attesa che la data squadra dovrebbe avere contro la squadra media del dato campionato; quelli difensivi alla performance offensiva attesa che l'avversario medio dovrebbe avere contro la squadra valutata.

Le valutazioni vengono poi aggiornate nella seguente maniera: per la i -esima squadra in una partita casalinga contro la j -esima.

$$H_i^a = \max\left(H_i^a + \lambda \Phi_1\left(S_h - \frac{H_i^a + A_j^d}{2}\right), 0\right)$$

$$A_i^a = \max\left(A_i^a + \lambda (1 - \Phi_1)\left(S_h - \frac{H_i^a + A_j^d}{2}\right), 0\right)$$

$$H_i^d = \max\left(H_i^d + \lambda \Phi_1\left(S_a - \frac{A_j^d + H_i^d}{2}\right), 0\right)$$

$$A_i^d = \max\left(A_i^d + \lambda (1 - \Phi_1) \left(S_a - \frac{A_j^a + H_i^d}{2}\right), 0\right)$$

E per aggiornare la j-esima squadra

$$A_j^a = \max\left(A_j^a + \lambda \Phi_2 \left(S_a - \frac{A_j^a + H_i^d}{2}\right), 0\right)$$

$$H_j^a = \max\left(H_j^a + \lambda (1 - \Phi_2) \left(S_a - \frac{A_j^a + H_i^d}{2}\right), 0\right)$$

$$A_j^d = \max\left(A_j^d + \lambda \Phi_2 \left(S_h - \frac{H_i^a + A_j^d}{2}\right), 0\right)$$

$$H_j^d = \max\left(H_j^d + \lambda (1 - \Phi_2) \left(S_h - \frac{H_i^a + A_j^d}{2}\right), 0\right)$$

dove $\lambda > 0$, $0 < \Phi_1 < 1$ e $0 < \Phi_2 < 1$ sono parametri da selezionare.

Per una determinata partita si può dire che una squadra ha ecceduto / non mantenuto le aspettative se la sua performance risulta essere migliore / peggiore del proprio *GAP rating* offensivo e del *GAP rating* difensivo della squadra e, in questo caso, vengono entrambi aggiornati positivamente / negativamente.

Il parametro λ determina il peso da attribuire all'aggiornamento apportato dall'ultima partita di ogni squadra, mentre Φ_1 e Φ_2 determinano l'influenza di una partita giocata in casa sulla valutazione della trasferta e viceversa, rispettivamente. Se $\Phi_1 = 0$, la valutazione delle abilità di gioco in casa non viene influenzata dai risultati ottenuti in trasferta. L'operatore max è incluso al fine di evitare che i *rating* possano risultare negativi.

1.4. Introduzione ad un modello player-based

In seguito alla “*Soccer Prediction Challenge*” ([Dubitzky et al. 2019](#)) sono stati pubblicati una moltitudine di articoli che adottano tecniche di machine learning. Nonostante gli sforzi della comunità di ricercatori che opera nell'ambito del machine learning, i guadagni marginali in termini di

accuratezza predittiva sono estremamente limitati. Per esempio, il modello più performante della “*Soccer Prediction Challenge*” ha ottenuto una precisione del 53.88%, mentre il peggiore del 50.49%. Come hanno constatato [Berrar et al. \(2019\)](#), vincitori della “*Soccer Prediction Challenge*” la conoscenza del settore è un elemento fondamentale per il successo in questo ambito e gli algoritmi di machine learning hanno la non–attraente proprietà di non rappresentare la realtà. Come quelli di [Maher \(1982\)](#) e di [Dixon e Coles \(1997\)](#), il modello qui proposto ha, al contrario, l’attraente peculiarità di rappresentare il modo in cui effettivamente il gioco del calcio è giocato.

Nonostante i benefici siano chiari, ci sono stati pochi tentativi di utilizzare un modello basato sulle caratteristiche e le statistiche dei singoli giocatori per predire risultati delle partite di calcio. Un’importante novità del modello che [Holmes e McHale \(2023\)](#) propongono è che non utilizzano una semplice media dei *rating* dei giocatori nelle due squadre come base per la generazione delle previsioni. Costruiscono, invece, un modello per simulare il livello a cui ogni giocatore di una squadra delle due in competizione interagisce con ciascuno di quelli della squadra avversaria.

Pur essendo stati fatti progressi nel permettere alle valutazioni delle squadre di variare, è stata prestata poca attenzione al motivo per cui queste varino e al meccanismo fisico che conduce a queste variazioni di partita in partita e di stagione in stagione. Sembra logico supporre che la principale causa di questa natura dinamica sia il fatto che cambiano gli interpreti del gioco all’interno delle squadre.

L’articolo considerato in questa relazione non fonda sulle valutazioni delle squadre la base per predirne i risultati futuri, bensì sulle valutazioni dei singoli giocatori. Così facendo, si riprodurrà direttamente il meccanismo che causa il variare degli equilibri delle squadre. Il progetto si concretizza in un modello performante sia dal punto di vista di accuratezza predittiva sia quando testato sul mercato delle scommesse.

Capitolo 2

Descrizione dei dati utilizzati, delle correzioni apportate dagli autori e della stima delle interazioni tra giocatori

Per il modello proposto da [Holmes e McHale \(2023\)](#) sono necessarie tre tipologie di dati: i rating dei giocatori, gli eventi delle partite e le quote dei *bookmaker*. I *rating* dei giocatori, dopo opportune modifiche, verranno utilizzati, assieme ai dati riguardo gli eventi delle partite, per ottenere delle valutazioni complessive delle squadre avversarie; le quote dei *bookmaker* per testare il modello finale. Ogni *dataset* è stato ottenuto per tutte le stagioni sportive dal 2013/14 al 2020/21.

2.1. Dati

2.1.1. Valutazioni WhoScored individuali

Il progetto prevede delle valutazioni individuali dei giocatori in campo (gli schieramenti e i singoli giocatori sono annunciati con un minimo di 30 minuti di anticipo rispetto alla partita e, solitamente, sono conosciuti con largo anticipo), che fungano da *input* per la previsione dei risultati delle partite. Gli autori hanno raccolto i *rating* delle performance delle partite pubblicate dal sito web *WhoScored.com*. In totale sono state usate 1.505.177 valutazioni individuali relative a 24.167 giocatori diversi. Queste valutazioni si riferiscono all'intervallo temporale compreso tra il 14 Luglio 2013 e il 31 Maggio 2021. Nonostante il metodo di calcolo di queste valutazioni non sia completamente di dominio pubblico, i concetti generali sono descritti nel sito web di WhoScored¹.

¹ Vedere <https://www.whoscored.com/Explanations>

Per riassumere, ogni giocatore comincia la partita con una valutazione pari a 6. Durante la partita la valutazione del giocatore viene aggiornata positivamente o negativamente in base alle sue azioni che impattano il match in maniera positiva o negativa per la propria squadra. I giocatori possono ottenere una valutazione massima pari a 10. In tutto ci sono stati 1084 casi in cui un giocatore ha ottenuto la valutazione massima. Il noto trio “MSN”, Lionel Messi, Luis Suarez e Neymar, detiene il record con, rispettivamente, 52, 20 e 22 valutazioni massime. Lo sfortunato record della valutazione minima registrata è detenuto da Oier Olazàbal, che era il portiere di Granada nella sconfitta per 9-1 con il Real Madrid nel 2015, che gli è valsa una valutazione di appena 1.89. Queste valutazioni sono molto diffuse tra appassionati e media specializzati. La **Figura 1** rappresenta l’istogramma delle valutazioni ottenute dai giocatori, divisi tra titolari e non.

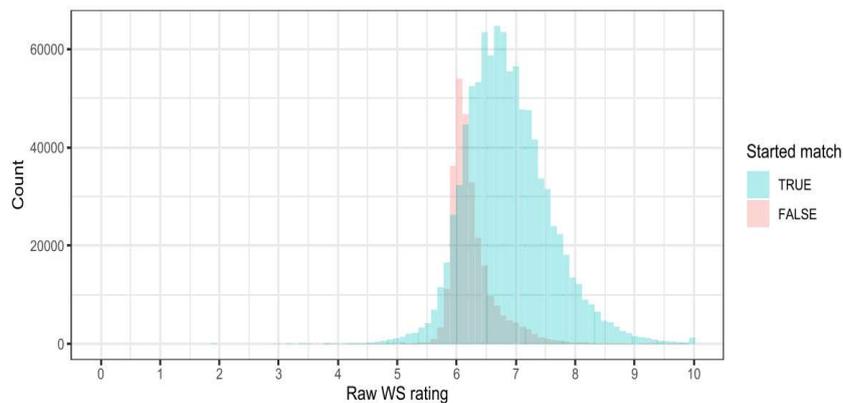


Figura 1. Istogramma delle valutazioni WhoScored ottenute da tutti i giocatori. I giocatori sono divisi in titolari e sostituti.

2.1.2. Dati relativi agli eventi dell’incontro

I dati relativi agli eventi dell’incontro descrivono tutte le azioni (tiri, passaggi, falli, palle rubate, etc.) occorse durante una partita e stanno diventando sempre più diffuse nella letteratura calcistica. Questa tipologia di dato è stata usata per adattare una serie di modelli multinomiali utili alla stima del livello di interazione tra due giocatori avversari. I dati sono stati raccolti da *InStat* .

Per i modelli multinomiali sono necessari i dati riguardanti tutti i tiri e tutte le azioni difensive: duelli aerei, parate, scontri sul terreno di gioco, intercettazioni, eccetera. Inoltre, sono necessari i dati riguardanti i ruoli coperti da ogni giocatore e le formazioni delle due squadre, che tengano conto anche delle eventuali modifiche avvenute durante il periodo di gioco.

Il modello finale di previsione usa i dati raccolti dai migliori cinque campionati europei dalla stagione sportiva 2015/16 a quella 2020/21. I modelli multinomiali sono addestrati sui dati degli stessi campionati relativi alle stagioni sportive 2013/14 e 2014/15 per essere sicuri che le previsioni siano completamente *out-of-sample*. Il *dataset* finale comprende 764.712 azioni difensive e 108.286 tiri.

Le informazioni riguardanti i risultati, i punteggi, le identità dei giocatori di entrambe le squadre e le formazioni delle squadre sono state anch'esse ottenute da *InStat*.

2.1.3. Quote dei bookmakers

Infine, lo storico delle quote di *Bet365* (un *bookmaker*) relative alle scommesse che sono state usate per testare la capacità predittive del modello di [Holmes e McHale \(2023\)](#) e le quote per le vittorie casalinghe, i pareggi e le vittorie in trasferta sono state ottenute dal sito web [football-data.co.uk](#).

2.2. Correzione delle valutazioni *WhoScored*

Prima di descrivere il disegno di modellazione, è stato riportato il sistema con cui sono state attribuite le valutazioni ai giocatori e che sono state successivamente usate come input per il nostro modello di previsione.

2.2.1. Raw *WhoScored* rating (*RWS*)

Considerando la media delle valutazioni di *WhoScored* ottenute, ci si può fare un'idea della qualità di un giocatore e di come potrebbe performare in partite future.

Sia il rating Raw *WhoScored* (*RWS*) di un giocatore la media delle sue valutazioni *WhoScored* ottenute in due anni. La **Tabella 1** mostra i migliori giocatori per valutazioni *RWS* nei *dataset*.

Tabella 1. Migliori dieci medie di valutazioni WhoScored ottenute dai giocatori in una finestra mobile di due anni. I giocatori devono aver disputato almeno dieci partite e possono comparire solo una volta nella tabella.

Player	Date	Team	League	RWS
Neymar	09/04/2019	PSG	France Ligue 1	8.821
Lionel Messi	19/12/2018	Barcelona	Spain LaLiga	8.698
Hakim Ziyech	12/02/2020	Ajax	Netherlands Eredivisie	8.429
Carlos Vela	20/08/2020	Los Angeles FC	USA Major League Soccer	8.394
Cristiano Ronaldo	02/09/2016	Real Madrid	Spain LaLiga	8.240
James Tavernier	07/12/2020	Rangers	Scotland Premiership	8.174
Kylian Mbappé	05/12/2020	PSG	France Ligue 1	8.149
Robert Lewandowski	23/05/2021	Bayern	Germany Bundesliga	8.098
Luuk de Jong	19/09/2019	Sevilla	Spain LaLiga	8.073
Zlatan Ibrahimovic	21/02/2017	Man Utd	England Premier League	8.026

Nonostante alcuni profili di alto livello di giocatori riconosciuti come i migliori in attività nelle stagioni calcistiche considerate assicuri la significatività di questo indice, ci sono alcuni nomi inaspettati nella **Tabella 1** (in ordine alfabetico: Carlos Vela, James Tavernier e Luuk de Jong). È evidente che considerare la semplice media aritmetica delle valutazioni di *WhoScored* delle partite disputate dal giocatore può essere fuorviante. Innanzitutto, appare evidente che questa procedura non considera la qualità del campionato in cui militano i diversi giocatori, le valutazioni ottenute da giocatori in campionati diversi non sono, infatti, direttamente confrontabili. Constatato ciò, è dunque necessaria una correzione di queste valutazioni per poter tenere conto dei diversi campionati e dei relativi livelli di competizione. In secondo luogo, alcuni giocatori hanno disputato un numero limitato di partite. Utilizzare la semplice media aritmetica delle valutazioni ottenute in queste poche partite per stimare come potrebbero performare in futuro, è probabile che porti ad ottenere risultati distorti, valutazioni irrealisticamente elevate o basse. Infine, è probabile che le prestazioni più recenti dei giocatori siano più significative per valutarne le condizioni e predirne l’impatto nelle partite future.

2.2.2. Adjusted WhoScored ratings (AWS)

Queste problematiche possono essere risolte da un modello di regressione, che [Holmes e McHale \(2023\)](#) usano per generare dei rating corretti, che chiamiamo Adjusted WhoScored Ratings (*AWS*). La variabile dipendente equivale alla *RWS* della partita. Le covariate includono delle variabili dummy che indicano il giocatore, il campionato e il fatto che la partita si disputi in casa o in trasferta. Nello specifico, si supponga di osservare y_1, \dots, y_N valutazioni WhoScored. Siano $p(i)$ il giocatore che ottiene il rating y_i , $l(i)$ il campionato in cui viene disputata la partita e $h(i)$ una funzione indicatrice che indica se il giocatore stava giocando in casa.

Allora il modello può essere scritto come

$$y_i = \alpha_0 + h(i)\alpha_1 + \beta_{p(i)} + \gamma_{l(i)} + e_i, e_i \sim N(0, \sigma^2) \quad (1)$$

dove i coefficienti β e γ sono le stime delle valutazioni per ogni giocatore e campionato, rispettivamente, mentre α_0 è l'intercetta e α_1 rappresenta un parametro che indica il vantaggio nel giocare nel campo di casa.

Per tenere conto di giocatori con un campione di partite giocate limitato bisogna far sì che le valutazioni si contraggano in direzione della valutazione media aggiungendo partite fittizie. In queste partite simulate assumiamo che il giocatore giochi nel suo attuale campionato e che il parametro che indica il vantaggio del campo sia pari a 0.5, che equivale a considerare un campo neutrale. Incrementando il peso di queste pseudo-osservazioni, è possibile aggiustare il livello di contrazione. Sia ω il peso, quando $\omega = 0$ il modello non ha contrazione. Quando ω aumenta il livello di contrazione aumenta. Questo è un iper-parametro che deve essere calibrato.

Per tener conto dell'evoluzione delle capacità di un giocatore nel corso della sua carriera, le osservazioni vanno pesate in modo tale che quelle relative a partite disputate più nel passato abbiano meno impatto sulla stima dei coefficienti di regressione rispetto alle osservazioni più recenti. Gli autori dell'articolo applicano uno schema di ponderamento esponenziale alle osservazioni, come fecero [Dixon e Coles \(1997\)](#) e altri dopo di loro. E includono solo partite disputate nei ψ anni prima del giorno della valutazione, dove il peso è pari a $e^{(-\phi t/3.5)}$ e t è il numero di giorni compresi tra l'osservazione e il giorno della valutazione.

Per calibrare gli iper-parametri ω , ψ e ϕ , [Holmes e McHale \(2023\)](#) hanno puntato a minimizzare lo scarto quadratico medio della predizione delle performance future di un giocatore attraverso i dati di validazione. Hanno constatato che gli iper-parametri ottimali sono $\psi = 2.00$, $\phi = 0.0062$ e $\omega = 7.00$. Il valore stimato di ϕ è tale per cui una partita disputata un anno fa abbia un peso sulla stima del *AWS* di un giocatore approssimativamente pari alla metà di quello della partita più recente.

Tabella 2. Migliori 10 AWS rating ottenuti dai giocatori. Non c'è un minimo di partite giocate necessarie per apparire nella tabella.

Player	Date	Team	League	AWS
Lionel Messi	19/05/2021	Barcelona	Spain LaLiga	1.799
Neymar	20/02/2018	PSG	France Ligue 1	1.603
Cristiano Ronaldo	07/08/2015	Real Madrid	Spain LaLiga	1.397
Robert Lewandowski	23/05/2021	Bayern	Germany Bundesliga	1.225
Kylian Mbappé	01/11/2020	PSG	France Ligue 1	1.166
Kevin De Bruyne	20/01/2021	Man City	England Premier League	1.106
Eden Hazard	10/05/2019	Chelsea	England Premier League	1.100
Zlatan Ibrahimovic	20/08/2016	Man Utd	England Premier League	1.091
Hakim Ziyech	11/03/2019	Ajax	Netherlands Eredivisie	1.075
Harry Kane	21/01/2018	Tottenham	England Premier League	1.053

La **Tabella 2** elenca i migliori 10 giocatori secondo l'indice *AWS*. Grazie alla simulazione di partite fittizie, non è necessario filtrare i giocatori in base al numero minimo di partite giocate per apparire nella tabella. La lista è composta dall'élite calcistica, assicurandoci sulla qualità dell'indice *AWS*. Le inaspettate presenze nella **Tabella 1** sono ora sparite dalla top 10. Vela e Tavernier giocavano nella MLS e Scottish Premiership nel momento in cui hanno ottenuto le loro valutazioni migliori. De Jong è stato trasferito dall'Eredivisie, massima serie professionistica del campionato olandese di calcio, il 1 Luglio 2019, poco prima di ottenere la propria valutazione *RWS* massima. Di conseguenza, le valutazioni *RWS* ottenute da questi tre giocatori erano indici di prestazioni in campionati relativamente poco competitivi. Le loro valutazioni *AWS*, tenendo conto della più bassa competitività dei rispettivi campionati, sono meno elevate.

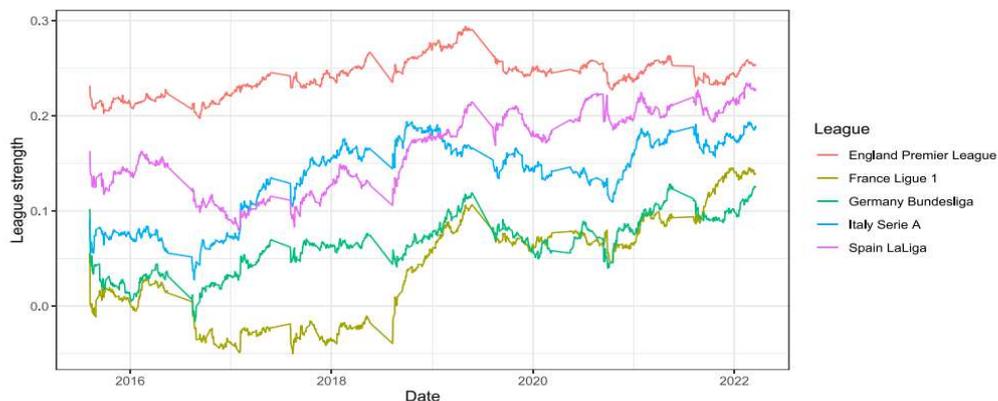


Figura 2. Andamento del livello dei diversi campionati nel tempo. Notare che viene riportato il valore negativo della stima effettiva, dato che un valore minore implica un campionato più competitivo. Il campionato di riferimento è l'English Championship.

Una potenziale problematica delle valutazioni *AWS* messa in luce dalla **Tabella 2** è che gli attaccanti dominano queste classifiche. Infatti, la maggior parte dei migliori 50 giocatori per *AWS* è attaccante. Ovviamente, potrebbe essere che i migliori giocatori del mondo siano attaccanti; dopotutto, sono i giocatori che ottengono le retribuzioni più elevate e che comportano i costi di trasferimento maggiori. Ma è anche possibile che le valutazioni *AWS* presentino un bias a favore degli attaccanti. Si potrebbe utilizzare un diverso sistema di valutazione, ma, come dimostrato dalle performance predittive del nostro modello, la qualità delle valutazioni *AWS* è elevata.

Un aspetto marginale interessante sono le stime del livello di competitività dei diversi campionati. Anche questa è un'importante area di studio, infatti, quando i club vogliono valutare per il proprio mercato giocatori che militano in campionati differenti dal loro, è importante che siano certi che i giocatori di interesse possano performare allo stesso livello anche nel nuovo campionato. La **Figura 2** mostra la stima del coefficiente di correzione relativo al campionato in cui gioca il giocatore che si

sta valutando (dove il secondo campionato inglese, English Championship, è il campionato di riferimento). Non è una sorpresa che la Premier League è il campionato che risulta essere più competitivo tra quelli considerati. Ogni valutazione individuale ottenuta in Premier League viene considerata all'incirca 0.25 volte di più rispetto alla stessa valutazione grezza ottenuta nell' English Championship. Un altro aspetto interessante è la forte crescita della Ligue 1 francese. Durante le stagioni sportive 2016/17 e 2017/18, una data valutazione grezza ottenuta nella English Championship valeva più della stessa ottenuta in Ligue 1 (il coefficiente stimato era negativo). Mentre, dalla stagione 2021/22 una valutazione ottenuta in Ligue 1 ha più valore di una uguale ottenuta non solo in English Championship, ma anche in Bundesliga.

2.3. Includere le valutazioni dei giocatori in un modello di predizione dei risultati della squadra

In questo paragrafo vengono presentate le modalità con cui viene preso in considerazione l'impatto dei singoli giocatori nei modelli di previsione dei risultati delle partite. La specificazione in questione ha due benefici. Il primo è che il modello replica il modo in cui i giocatori avversari interagiscono all'interno di una partita nella realtà. Il secondo è che questa specificazione è facilmente applicabile a diversi modelli dal momento che produce una singola covariata.

Solitamente, in ogni partita, in base alle formazioni delle due squadre, un singolo giocatore spenderà la maggior parte del tempo in competizione con un singolo giocatore avversario, mentre con gli altri interagirà in maniera più marginale.

2.3.1. Stima del livello di interazione tra giocatori avversari, escluso il portiere

Il primo modello multinomiale stima il livello di interazione tra giocatori avversari usando dati di eventi in cui sono coinvolti due giocatori: contrasti (vinti o persi), intercettazioni, duelli aerei e parate. Supponiamo ci siano M eventi indicizzati da $j=1, \dots, M$. Siano pos_j^A e pos_j^D le posizioni dell'attaccante e del difensore che sono coinvolti nell'evento j -esimo, rispettivamente. Siano, inoltre, $form_j^A$ e $form_j^D$ le formazioni che la squadra in attacco e quella in difesa hanno adottato rispettivamente nel momento in cui è occorso l'evento j -esimo. La variabile dipendente è la posizione

del giocatore in difesa. Le variabili indipendenti comprendono la posizione del giocatore attaccante e le formazioni delle due squadre.

Per comodità espositiva indichiamo con $pos_j^D = i = 1, \dots, 25$ le 25 possibili modalità che può assumere la variabile pos_j^D .

Se riorganizziamo i dati in M vettori $pos_j^D = (pos_{j1}^D, \dots, pos_{j25}^D)$, con $j = 1, \dots, M$, con $pos_{ji}^D = 1$ se nel j -esimo evento è stato coinvolto il giocatore difensore che gioca nella i -esima posizione e $pos_{ji}^D = 0$ altrimenti, con $i = 1, \dots, 25$ si può assumere per pos_j^D un modello multinomiale con funzione di probabilità

$$\begin{aligned}
 p_{POS_j^D}(pos_j^D; \pi_j) &= Pr(POS_j^D = pos_j^D) \\
 &= Pr(POS_{j1}^D = pos_{j1}^D, \dots, POS_{j25}^D = pos_{j25}^D) \\
 &= \pi_{j1}^{pos_{j1}^D} \cdot \dots \cdot \pi_{j25}^{pos_{j25}^D}
 \end{aligned} \tag{2}$$

che può essere anche scritta come

$$\begin{aligned}
 p_{POS_j^D}(pos_j^D; \pi_j) &= \exp\left(pos_{j1}^D \cdot \log(\pi_{j1}) + \dots + pos_{j25}^D \cdot \log(\pi_{j25})\right) \\
 &= \exp\left(pos_{j1}^D \cdot \log(\pi_{j1}) + \dots + pos_{j24}^D \cdot \log(\pi_{j24}) + \left(1 - \sum_{i=1}^{24} pos_{ji}^D\right) \cdot \log(\pi_{j25})\right) \\
 &= \exp\left(pos_{j1}^D \cdot \log\left(\frac{\pi_{j1}}{\pi_{j25}}\right) + \dots + pos_{j24}^D \cdot \log\left(\frac{\pi_{j24}}{\pi_{j25}}\right) + \log(\pi_{j25})\right)
 \end{aligned} \tag{3}$$

Ci sono 25 posizioni possibili (escluso il portiere), dunque vengono stimati 24 modelli logit. Il modello di regressione logistica con modalità di riferimento, infatti, assume che, per ogni i , l' i -esimo logit rispetto alla modalità di riferimento, che, nel nostro caso, sarà sempre il trequartista centrale (CAM), sia pari al predittore lineare specifico per ogni modalità $i \neq CAM$.

Per esempio, il modello logit, che stima la probabilità che il giocatore che copre il ruolo di terzino destro (RB) sia coinvolto difensivamente nel j -esimo evento, date le due formazioni e la posizione dell'attaccante coinvolto, è

$$\log\left(\frac{P(pos_j^D=RB)}{P(pos_j^D=CAM)}\right) = \text{int}^{RB} + \beta_{pos_j^A}^{RB} + \beta_{form_j^D}^{RB} + \beta_{form_j^A}^{RB} \tag{4}$$

Da questa specificazione risulta

$$\pi_{ji} = \frac{\exp(x_j \beta_i)}{1 + \sum_{h=1}^{24} \exp(x_j \beta_h)} \quad (5)$$

dove $x_j = (1, pos_j^A, form_j^D, form_j^A)$, $\beta_i = (int^i, \beta_{pos_j^A}^i, \beta_{form_j^D}^i, \beta_{form_j^A}^i)^T$, $i = 1, \dots, 24$

Quando si generano previsioni con questo modello, vengono calcolate 25 probabilità diverse da zero. Tuttavia, è possibile che siano schierati solo dieci giocatori contemporaneamente sul campo². Di conseguenza, è necessario correggere questi dieci valori in modo che sommino a uno.

Il valore π_{ji} ($i \neq portiere$) rappresenta la probabilità che, date le formazioni delle due squadre, il giocatore difensivo che ricopre il ruolo $pos_j^D = i = 1, \dots, 25$ interagisca con il giocatore avversario che ricopre il ruolo pos_j^A , la quantità misura, quindi, il livello di interazione generale tra i giocatori nelle due posizioni, date le due formazioni. In altre parole, il valore π_{ji} ($i \neq portiere$) risponde alla domanda: date le due formazioni e la posizione dell'attaccante coinvolto nel j -esimo evento, qual è la probabilità che sia stato coinvolto l' i -esimo difensore?

2.3.2. Stima del livello di interazione tra giocatori e portiere avversario

In maniera simile vengono utilizzati i dati relativi ai tiri per determinare il livello di interazione tra un giocatore sul campo e il portiere avversario. Questo modello è necessario perché i portieri tendono a non interagire con duelli sul campo, che sono stati usati come base per il nostro modello del sottoparagrafo precedente. Questa volta la variabile dipendente è il ruolo del giocatore che tira. Il ruolo del giocatore difensore è sempre il portiere, quindi solo le formazioni delle due squadre sono le variabili indipendenti considerate. Come nel modello precedente correggiamo i valori in base ai 10 giocatori effettivamente in campo. Il valore π_{iGK} rappresenta dunque la probabilità che un determinato tiro verrà effettuato dal giocatore i della squadra attaccante e sarà una misura del livello di interazione tra il giocatore i e il portiere avversario $i = 1, \dots, 25$.

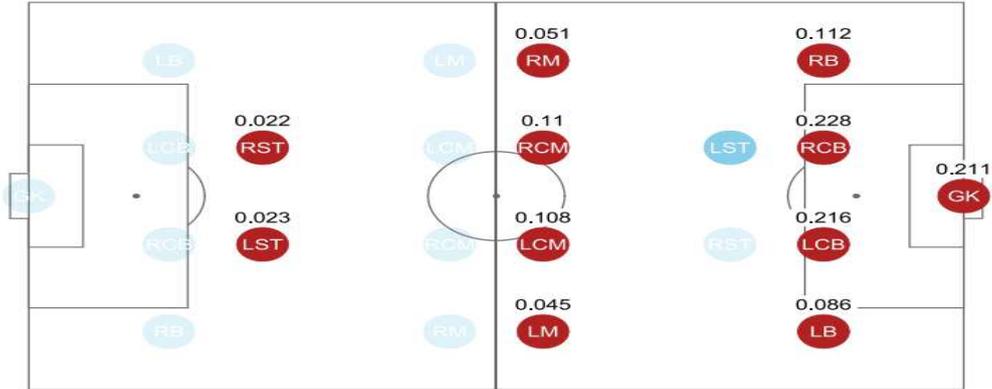
2 Si noti che le probabilità stimate dei giocatori non presenti sul terreno di gioco sono estremamente basse.

2.3.3. Esempi

La **Figura 3** mostra i risultati di questi modelli multinomiali per due casi esempio. Il primo grafico mostra come un attaccante sinistro (LST), in una squadra che gioca in formazione 4-4-2, interagisca con ognuno dei giocatori avversari, che a loro volta giocano con un modulo 4-4-2. Il 22.8% delle sue interazioni sono con il difensore centrale destro (RCB), mentre solo il 2.3% sono con un attaccante sinistro (LST) avversario. Vediamo un 21.1% di possibilità che l'attaccante sinistro tenti una conclusione in porta, che indica il suo livello di interazione con il portiere.

Il secondo grafico mostra che un'ala sinistra (LW) in una formazione 4-3-3 ha il 27.4% delle sue interazioni con il terzino destro (RB) avversario (che sta giocando in formazione 4-4-2) e il 12.2% contro il centrocampista destro (RM) avversario. Notiamo che le interazioni di un attaccante sinistro (LST) in una formazione 4-4-2 con un attaccante destro (RST) avversario rappresentano il 2.2% delle sue azioni di gioco. Per un'ala sinistra (LW) in una 4-3-3 questo dato aumenta al 3.1%, suggerendo che un'ala giocherà, mediamente, più difensivamente di un attaccante, che sembra coerente con la realtà. Inoltre, c'è un 20.8% di possibilità che un'ala sinistra (LW) tenti una conclusione in porta.

Team formation: 442
 Opponent formation: 442
 Player position: LST



Team formation: 433
 Opponent formation: 442
 Player position: LW

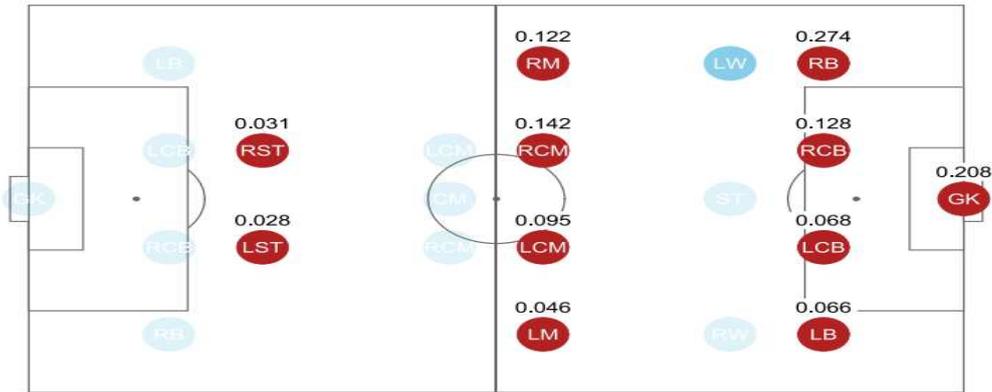


Figura 3. Esempi dei livelli di interazione in due casi differenti. La squadra difensiva è colorata in rosso.

Per misurare la differenza tra due squadre in una partita viene proposta la seguente statistica:

$$\Delta = \sum_j \sum_i \pi_{ji} (AWS_j - AWS_i) \quad (6)$$

dove AWS_i è il rating AWS dell' i -esimo giocatore della squadra di casa e AWS_j è il rating AWS del j -esimo giocatore della squadra in trasferta. π_{ji} è il livello di interazione stimato dai modelli multinomiali sopra descritti tra il giocatore i -esimo e il giocatore j -esimo. La prima sommatoria fornisce una differenza pesata tra il rating AWS di un giocatore della squadra di casa e ciascuno dei rating AWS dei giocatori avversari. La seconda sommatoria calcola la prima somma per ciascuno dei giocatori della squadra che gioca in casa.

Capitolo 3

Descrizione dei modelli di previsione

3.1. Gestione dei dati

Essendo i modelli multinomiali stati addestrati sui dati delle stagioni sportive 2013/14 e 2014/15, vengono utilizzati i rimanenti dati (2015/16-2020/21) per modellare i risultati. Ciò ci assicura che le probabilità generate dai modelli multinomiali siano anch'esse *out-of-sample*.

Come è usuale fare, viene usato il primo 80% dei dati delle stagioni 15/16-20/21 per l'addestramento del modello finale e il rimanente 20% per testarlo.

Ci sono diversi parametri da calibrare: gli iper-parametri dei modelli di valutazione dei singoli giocatori, i pesi che utilizziamo nella funzione di verosimiglianza per dare più considerazione alle osservazioni più recenti e le soglie ottimali per le strategie di investimento nei siti di scommesse. Dividiamo nuovamente il set che useremo per l'addestramento per ottimizzare questi parametri e assicurarci che siano effettivamente *out-of-sample*, tenendoci il rimanente 20% per il test di validazione. La **Figura 4** illustra queste separazioni graficamente. Di conseguenza, è possibile riportare i nostri risultati come completamente *out-of-sample* in tutti gli esperimenti seguenti.

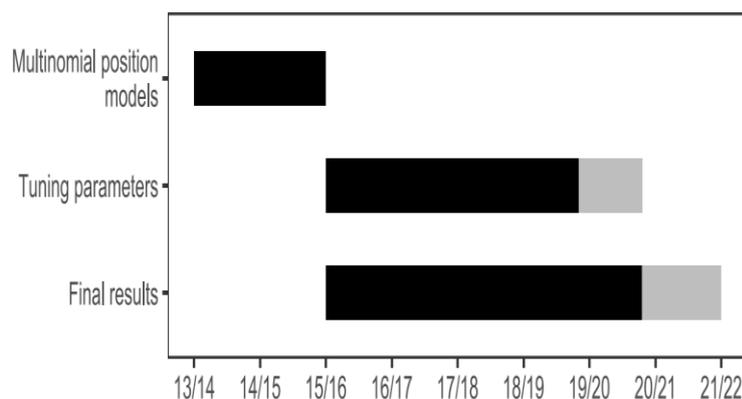


Figura 4. Illustrazione della suddivisione dei dati nelle tre principali fasi del progetto. Il nero indica la porzione di dati utilizzata per il *train* e il grigio la porzione utilizzata per il *test*.

Durante la pandemia di Covid-19 le partite di calcio furono giocate a porte chiuse, con gli stadi privi di spettatori e tifoserie. Queste partite sono state rimosse dai campioni per evitare che i parametri che indicano il vantaggio del campo di casa risultassero distorti. Rimaniamo così con un campione di 6824 partite tra il 12 Agosto 2016 e il 23 Maggio 2021. Di questo campione, l'ultimo 20% è stato utilizzato per il test di validazione finale.

3.2. Modello di Skellam

Nella letteratura concernente i modelli di previsione nel calcio c'è molta attenzione allo stimare il risultato finale delle partite. L'idea originaria di [Maher \(1982\)](#) fu quella di valutare separatamente il potenziale offensivo delle due squadre per la singola partita. Il potenziale offensivo della squadra di casa è stimato come funzione di una valutazione delle capacità offensive della squadra di casa e di una valutazione delle capacità difensive di quella in trasferta e viceversa per il potenziale offensivo della squadra avversaria. Il modello originale usava regressioni di Poisson indipendenti e, da quel momento, i ricercatori hanno sviluppato piani di modellazione sempre più complessi come, per esempio, regressioni bivariate di Weibull ([Boshnakov et al., 2016](#)) (di cui al sottoparagrafo 1.1.2.)

Le variabili Δ , di cui al capitolo precedente, che rappresentano la differenza tra le valutazioni complessive delle squadre, non riescono ad inserirsi perfettamente in un algoritmo di questo genere. Invece di prevedere i goal segnati da ogni squadra, è possibile stimare la differenza reti usando una regressione di Skellam.

La distribuzione di Skellam ([Skellam, 1946](#)), infatti, permette di descrivere la variabile casuale discreta risultante dalla differenza di due variabili casuali indipendenti di Poisson: date due variabili casuali di Poisson con valori attesi, rispettivamente, pari a λ_1 e λ_2 , la funzione di probabilità della differenza GD (differenza goal) dei due conteggi indipendenti risulta essere

$$\begin{aligned}
 p(GD; \lambda_1, \lambda_2) &= \sum_{n=-Inf}^{Inf} p(GD + n; \lambda_1) p(n; \lambda_2) \\
 &= e^{-(\lambda_1 + \lambda_2)} \sum_{n=\max(0, -GD)}^{Inf} \frac{\lambda_1^{GD+n} \lambda_2^n}{n!(GD+n)!}
 \end{aligned} \tag{7}$$

in cui la seconda sommatoria viene presa in considerazione solamente quando $n \geq 0$ e $n + GD \geq 0$, dal momento che la funzione di probabilità di una variabile casuale di Poisson vale 0 per valori negativi del conteggio.

È poi possibile dimostrare che l'Eq. (7) implica che

$$\frac{p(GD; \lambda_1, \lambda_2)}{p(-GD; \lambda_1, \lambda_2)} = \left(\frac{\lambda_1}{\lambda_2}\right)^{GD} \quad (8)$$

da cui

$$p(GD; \lambda_1, \lambda_2) = e^{-(\lambda_1 + \lambda_2)} \left(\frac{\lambda_1}{\lambda_2}\right)^{\frac{GD}{2}} I_{|GD|}(2\sqrt{\lambda_1 \lambda_2}) \quad (9)$$

dove $I_k(z)$ è la modified Bessel function di primo tipo.

Le probabilità di ognuno degli esiti può essere facilmente ottenuta sommando le probabilità delle differenza reti: pertanto la probabilità di vittoria della squadra di casa sarà data dalla sommatoria $\sum_{gd=1}^{Inf} p(GD = gd; \lambda_1, \lambda_2)$, un pareggio da $p(GD = 0; \lambda_1, \lambda_2)$ e la vittoria della squadra in trasferta dalla sommatoria $\sum_{gd=-1}^{-Inf} p(GD = gd; \lambda_1, \lambda_2)$.

Per la partita i , siano GD_i la differenza goal ipotizzata, Δ_i^{Start} la differenza ponderata dei ratings AWS dei giocatori titolari definita nell'Eq. (6) e Δ_i^{Sub} la differenza non ponderata nei ratings AWS dei giocatori in panchina³. Dal momento che la distribuzione di Skellam rappresenta la differenza tra due variabili casuali di Poisson indipendenti, bisogna stimare un coefficiente per ognuna delle due variabili. Questi coefficienti naturalmente corrispondono allo scoring rate della squadra di casa e di quella in trasferta, che è stato incluso nella modello con la seguente notazione.

Di conseguenza, viene stimato

$$GD_i \sim Skellam(\lambda_{ih}, \lambda_{ia}) \quad (10)$$

$$\lambda_{ih} = \exp(\beta_{0h} + \beta_{1h}\Delta_i^{Start} + \beta_{2h}\Delta_i^{Sub}) \quad (11)$$

$$\lambda_{ia} = \exp(\beta_{0a} + \beta_{1a}\Delta_i^{Start} + \beta_{2a}\Delta_i^{Sub}) \quad (12)$$

3 Viene impiegata la differenza non ponderata per i giocatori non titolari perché non è noto prima della partita se i sostituti verranno impiegati, per quanti minuti o che ruolo copriranno

Le stime dei coefficienti di questo modello sono illustrate nella **Tabella 4**. Notare che le variabili sono state standardizzate in modo tale da avere media 0 e varianza 1 nel momento di inserirle nel modello.

Tabella 4. Risultati del modello di regressione di Skellam. Le variabili sono standardizzate per ottenere una media pari a 0 e una varianza unitaria. I relativi *p-value* sono forniti nella colonna di destra.

<i>Dependent variable:</i>		
Goal-difference	Coefficient	<i>p-value</i>
β_{0h}	0.3670	0.0000
β_{1h}	0.2493	0.0000
β_{2h}	0.0583	0.0088
β_{0a}	0.0500	0.0686
β_{1a}	-0.3403	0.0000
β_{2a}	-0.0635	0.0269
Observations	5459	
Log-likelihood	- 10374.17	
AIC	20760.34	

Si nota un significativo vantaggio per la squadra che gioca in casa e che l'effetto si allinea quantitativamente con la letteratura passata, la stima del coefficiente di intercetta del predittore lineare della squadra che gioca in casa ($\beta_{0h} = 0.3670$) è maggiore del corrispettivo della squadra in trasferta ($\beta_{0a} = 0.0500$). La somma ponderata delle differenze nelle valutazioni dei singoli giocatori è altamente significativa. La differenza nelle valutazioni dei giocatori non titolari è a sua volta altamente significativa, con un coefficiente stimato leggermente minore. Questo è coerente con il fatto che il risultato della partita verrà impattato maggiormente dalle capacità dei titolari delle due squadre. Osserviamo inoltre un significativo vantaggio per i padroni di casa, risultando le variabili Δ positive se la squadra che gioca in casa è stata valutata meglio di quella in trasferta e viceversa, i coefficienti stimati risultano, coerentemente, positivi per la squadra di casa ($\beta_{1h} = 0.2493$, $\beta_{0h} = 0.0583$) e negativi per quella in trasferta ($\beta_{1a} = -0.3403$, $\beta_{2a} = -0.0635$).

Capitolo 4

Valutazione dei modelli sulla base dei risultati che si sono realizzati

Per confrontare modelli per la previsione degli esiti di incontri calcistici gli autori propongono l'utilizzo delle *scoring rules* e una valutazione basata sui risultati che sono stati effettivamente ottenuti con una strategia di investimento su siti di scommesse.

4.1. Scoring rules

Quando si utilizzano le *scoring rules* per valutare dei modelli di previsione, bisogna prestare particolare attenzione; questi strumenti, infatti, si basano sui punteggi che i modelli di previsione riescono ad ottenere massimizzando delle funzioni di utilità o minimizzando delle funzioni di perdita date. Queste funzioni, tuttavia, possono essere fortemente personalizzate in base alle esigenze e alle disponibilità dell'utente finale che dovrebbe usufruirne; nel caso in esame, ad esempio, la forma di queste funzioni potrebbe dipendere dal grado di avversione al rischio dell'investitore o dalla sua disponibilità economica o, ancora, dalla sua percezione dell'esposizione finanziaria. La forte personalizzazione di queste funzioni comporta delle problematiche. Innanzitutto, la capacità predittiva di due modelli che stimano la probabilità di uno stesso evento verrà classificata in ordini diversi in base alla modalità di valutazione. In secondo luogo, adattare un modello massimizzando una data funzione di punteggio piuttosto che un'altra, comporta la stima di parametri, e quindi di probabilità, differenti ([Johnstone et al. 2013](#)).

Dal momento che sia il rating *AWS* sia i pesi basati sui livelli di interazione sono nuovi innesti al nostro modello, proviamo diverse varianti per verificare se, effettivamente, queste due componenti apportano miglorie alle capacità predittive del modello stesso.

Sia Δ_{full}^{Start} la somma delle differenze degli *AWS rating* dei titolari ponderata per il livello di interazione tra di essi, la variabile definita nell'Eq. (6).

Sia Δ_{adj}^{Start} la somma delle differenze degli *AWS rating* dei titolari non ponderata per il livello di interazione.

In maniera analoga, sia Δ_{raw}^{Start} la somma non ponderata delle differenze dei *RWS rating* dei titolari.

Infine, siano Δ_{adj}^{Sub} e Δ_{raw}^{Sub} le somme delle differenze dei *rating AWS* e *RWS*, rispettivamente, dei titolari e dei potenziali sostituti disponibili per la data partita.

Si considerino ora quattro modelli così definiti:

-Il modello $skellam_{full}$ è il modello illustrato nella **Tabella 4**. Questo usa Δ_{full}^{Start} e Δ_{adj}^{Sub} come covariate; esso include dunque sia l'effetto della ponderazione per il livello di interazione, sia i ratings *AWS*.

-Il modello $skellam_{adj}$ rimuove il ponderamento per il livello di interazione ed utilizza i ratings *AWS*; quindi usa Δ_{adj}^{Start} e Δ_{adj}^{Sub} come covariate.

-Il modello $skellam_{raw}$ rimuove anche gli aggiustamenti effettuati ai ratings WhoScored; quindi usa Δ_{raw}^{Start} e Δ_{raw}^{Sub} come covariate.

-Il modello $skellam_{team}$ è un modello basato sulla valutazione delle squadre, che prevede variabili *dummy* per ogni squadra, le quali possono assumere valore 1 o -1 se l'incontro avviene in casa o in trasferta, rispettivamente. Il valore 0 indica che la squadra non è coinvolta nella partita. Come per i ratings dei giocatori, le abilità di una squadra sono aggiornate utilizzando i dati disponibili prima della partita. In questo modello includiamo anche la ponderazione basata sul tempo trascorso dalla partita che fornisce le informazioni prese in considerazione.

Il confronto tra modelli così definiti permette di valutare se le novità introdotte nel modello apportano migliorie alla capacità predittive dello stesso e rispetto ai modelli basati su valutazioni complessive di squadre proposte dalla letteratura analizzata nel Capitolo 1.

Nella **Tabella 5** vengono riportate l'accuratezza e lo *score di Brier* ottenuti.

4.1.1. Score di Brier

Lo *score di Brier* ([Brier, 1950](#)) è la regola più utilizzata nella letteratura dei modelli di previsione e fu inizialmente messo a punto per valutare dei modelli di previsione meteorologica.

Supponiamo che in ognuna delle n occorrenze considerate un evento possa verificarsi solamente in una delle r classi o categorie, che devono essere scelte in modo tale da essere mutualmente esclusive; nel nostro caso, queste categorie potrebbero essere vittoria della squadra di casa, pareggio e sconfitta della squadra di casa ($r = 3$).

Siano $f_{i1}, f_{i2}, \dots, f_{ir}$ le probabilità stimate che l'evento i avvenga nella categoria $1, 2, \dots, r$, tali che

$$\sum_{j=1}^r f_{ij} = 1, i = 1, 2, 3, \dots, n \quad (13)$$

Allora lo *score di Brier* è definito come

$$P = \frac{1}{n} \sum_{j=1}^r \sum_{i=1}^n (f_{ij} - E_{ij})^2 \quad (14)$$

in cui E_{ij} assume valore 1 se l'evento i si è verificato nella categoria j -esima, 0 altrimenti.

Dalla formula risulta evidente che è possibile ottenere un punteggio P minimo di 0, quando tutte le previsioni risultano unitarie e perfettamente allineate con gli eventi osservati e un punteggio massimo di 2, che corrisponde alla peggior previsione possibile.

4.1.2. Accuratezza

L'accuratezza, invece, benché sia molto più intuitiva per il lettore, non è un punteggio vero e proprio. È un valore che non tiene conto della dimensione dell'errore; esso corrisponde infatti alla semplice proporzione di eventi correttamente categorizzati sui totali previsti. Nel caso in esame, vengono considerati gli eventi con la probabilità marginale maggiore; pertanto, se, per esempio, vengono stimate delle probabilità di vittoria della squadra di casa pari a 0.6, di pareggio pari a 0.3 e di sconfitta della squadra di casa 0.1, la previsione verrà considerata corretta nel caso in cui dovesse vincere la squadra di casa.

La **Tabella 5** mostra lo *score di Brier* per i diversi modelli e le probabilità dei *bookmaker*. [Holmes e McHale \(2023\)](#) usano le quote di *Bet365* ricavate da www-football-data.co.uk e rimuoviamo il margine di guadagno dei *bookmaker*, facendo in modo che le probabilità sommino a uno.

In termini di accuratezza, il *bookmaker* è quello che funziona meglio. Tuttavia, notiamo delle differenze minime tra le cifre. Un modello base che prevede per ogni partita la vittoria della squadra di casa risulta avere un'accuratezza del 45.11%. Il guadagno marginale che si ottiene andando dal modello più semplice tra quelli disponibili a quello più performante (i *bookmaker*) è sorprendentemente minimo. Perfino il modello di machine learning più performante del “*Soccer Prediction Challenge*” (Dubitzky et al., 2019) ha ottenuto un'accuratezza leggermente superiore al 53.88% (nonostante, ovviamente, sia stato ottenuto su un dataset differente e quindi non può essere direttamente comparato con l'accuratezza ottenuta dai modelli qui trattati).

Il *bookmaker* ottiene anche il miglior *score di Brier*. Tra i modelli implementati dagli autori dell'articolo su cui si basa questa relazione $skellam_{full}$ è quello più performante; il piccolo gap che lo pone sotto ai *bookmaker* in classifica è incoraggiante. Vediamo, inoltre, che il modello $skellam_{full}$ è migliore dello $skellam_{adj}$, viene verificato, quindi, che la ponderazione per il livello di interazione con i giocatori avversari migliora le performance del modello (nonostante lo faccia in maniera esigua). C'è inoltre un evidente miglioramento quando si sostituiscono i ratings *AWS* a quelli *RWS*, il che giustifica lo schema di aggiornamento e correzione delle valutazioni individuali dei giocatori. Il risultato più interessante è notare che tutti i modelli che utilizzano i ratings *AWS* ottengono risultati migliori rispetto al modello basato sulle valutazioni delle squadre.

Tabella 5. Scoring rules ottenuti dai diversi modelli di previsione. Sono inoltre illustrati i corrispondenti risultati ottenuti dalle probabilità implicite del *bookmaker* Bet365.

Model	<i>n</i>	Accuracy	Brier
Bet365	1350	52.74%	0.5877
$skellam_{full}$	1350	52.00%	0.5955
$skellam_{adj}$	1350	51.85%	0.5957
$skellam_{team}$	1350	52.89%	0.5962
$skellam_{raw}$	1350	51.33%	0.6029

Nota. L'accuratezza riportata per il modello $skellam_{team}$ non è coerente con la trattazione. Si tratta di un refuso.

4.2. Scommesse

Nel sottocapitolo precedente è stato dimostrato che il modello $skellam_{full}$ ottiene risultati simili ai *bookmaker*, se ci si basa sullo *score di Brier* e l'accuratezza. Tuttavia, diversamente dai *bookmaker*, uno scommettitore non deve necessariamente scommettere su tutte le partite.

Gli autori hanno utilizzato i modelli per scommettere e valutare che ritorno ci si può attendere dalle puntate su scommesse "1X2" (vittoria della squadra di casa, pareggio, vittoria della squadra in trasferta). Mentre i *bookmaker* hanno il vantaggio costruito con i valori delle quote, calcolate appositamente per avere un margine di guadagno, lo scommettitore ha il vantaggio che può decidere su cosa scommettere.

La strategia utilizzata per le puntate è basata sul *Kelly Criterion* ([Kelly, 1956](#)), che è lo stesso utilizzato da [Boshnakov et al. \(2016\)](#).

Il *Kelly Criterion* è nato dal desiderio di massimizzare i log-utili sul lungo periodo e si traduce in una strategia in cui lo scommettitore investe una frazione $f = \frac{(b+1)p-1}{b}$ del suo patrimonio totale,

nel quale p è la stima dello scommettitore della probabilità che un evento si verifichi (ad esempio la vittoria della squadra di casa) e b è la quota offerta dal *bookmaker* per lo stesso evento (da cui $1/(b+1)$ può essere interpretato sommariamente come la probabilità implicita stimata dal *bookmaker* che l'evento si verifichi).

Mentre la strategia di Kelly usuale prevede un rifinanziamento ricorsivo in seguito a ciascuna scommessa, gli autori propongono di allocare l'equivalente di una unità moltiplicata per f per ogni scommessa. Effettivamente l'investimento dopo ogni scommessa viene re-impostato ad uno.

In aggiunta è stata implementata un'ulteriore "protezione": la strategia si limita alle scommesse "di qualità", ossia quelle in cui il valore atteso di ogni scommessa è superiore ad una determinata soglia. Per ogni partita ci sono tre possibili eventi su cui puntare: vittoria della squadra di casa, pareggio, vittoria della squadra in trasferta⁴.

4 Per essere espliciti, ciò significa che possiamo scommettere su un massimo di tre possibili esiti (nella poco probabile eventualità in cui tutti e tre hanno un valore atteso positivo). Questo può anche significare che potremmo scommettere anche su eventi che non necessariamente sono quelli che i nostri modelli stimano come i più probabili.

La scommessa viene effettuata sull'evento A solo se

$$E[A] = P(A) \times Odds(A) - 1 > t,$$

in cui t rappresenta un parametro soglia e protegge la strategia di investimento nel momento in cui ci sia un'asimmetria informativa tra *bookmaker* e modello. Come menzionato nel capitolo 3, abbiamo suddiviso i dati tra quelli per addestrare i modelli, quelli per validarli e quelli per testarli. Un primo modello è stato adattato sui dati per l'addestramento (dati per la validazione esclusi) per determinare la soglia ottimale t . La soglia ottimale è poi determinata trovando il ritorno massimo dagli investimenti (ROI) sui dati di validazione (che, a questo punto, risultano essere *out-of-sample* rispetto ai dati di addestramento). Infine, l'intero modello è adattato utilizzando sia i dati di validazione sia quelli di addestramento e i risultati delle scommesse *out-of-sample* sono calcolati utilizzando la soglia predeterminata sui dati rimanenti.

Oltre a guardare al guadagno dell'investimento, è utile considerare anche l'indice di Sharpe. L'indice di Sharpe è uno strumento per misurare i guadagni attesi aggiustati per il rischio ed è definito come il tasso di guadagno per unità di volatilità. Come in finanza, l'indice di Sharpe è calcolato come il ROI su ogni scommessa diviso per la deviazione standard del ROI per ogni singola scommessa. Il risultato è poi annualizzato moltiplicandolo per \sqrt{n} , dove n è il numero totale di scommesse. Una regola generale è quella di valutare un investimento come buono se l'indice di Sharpe ha valore pari o superiore a 1 (e maggiore è, meglio è, dal momento che l'investimento ottiene rendimenti maggiori con esposizione al rischio minore).

Gli autori propongono un confronto tra i risultati ottenuti con questa strategia di Kelly rivisitata e quelle che prevedono un investimento fisso. Queste ultime strategie prevedono una puntata unitaria sull'evento che i modelli valutano come il più probabile. Esattamente come per la strategia di Kelly viene stimata una soglia ottimale e si riportano i corrispettivi risultati ottenuti.

Deve essere fatto notare che l'insieme delle scommesse con i due metodi potrebbe differire. Quando si applica il secondo metodo, infatti, l'utente scommette sull'evento che reputa più probabile; seguendo la strategia di Kelly, invece, l'utente scommette in base al valore atteso dell'esito della scommessa. Di conseguenza, questa scelta potrebbe non essere quella più probabile dal punto di vista dello scommettitore.

I risultati delle scommesse ottenuti basandosi sul modello *skellam_{full}* sono riportate nella **Tabella 6**. Data la letteratura sui modelli di previsione calcistica, questi risultati sono molto promettenti, specialmente se si considera l'alto numero di scommesse effettuate. Per esempio, [Koopman e Lit \(2015\)](#) hanno effettuato solo 50 scommesse in due stagioni calcistiche.

Si può constatare che solo la strategia più semplice – scommesse di importo costante senza soglia di sicurezza – restituisce una perdita. Entrambe le strategie che adottano il criterio di Kelly hanno ottenuto una buona performance, ottenendo ROI e indici di Sharpe molto promettenti su un numero di scommesse molto elevato. Inoltre, nonostante le strategie di scommesse sempre dello stesso importo con soglie $t=0$ e $t=0.1760$ abbiano entrambe ottenuto guadagni positivi, l'indice di Sharpe è minore di 1, che indica più rischio che guadagno atteso.

La **Figura 5** mostra la relazione tra la soglia scelta per il valore atteso minimo e il ROI ottenuto dal modello $skellam_{full}$ utilizzato con la strategia di Kelly rivisitata. È inoltre mostrato il numero di scommesse effettuate nella parte superiore del grafico. Il numero di puntate decresce all'aumentare della soglia di guadagno atteso minimo fissata, ma il ROI cresce a livelli veramente elevati.

Tabella 6. Risultati ottenuti con diverse strategie di scommesse usando il modello $skellam_{full}$

Strategy	t	N	Accuracy (%)	Stakes	Profits	ROI (%)	Sharpe
Kelly	0.1866	556	24.10	65.36	7.81	11.96	1.07
Kelly	0.0000	1457	29.44	105.42	6.03	5.72	1.02
Flat	0.1760	199	37.19	199.00	9.04	4.54	0.45
Flat	0.0000	568	45.25	568.00	16.93	2.98	0.59
Flat		1350	52.00	1350.00	-32.56	-2.41	-0.87

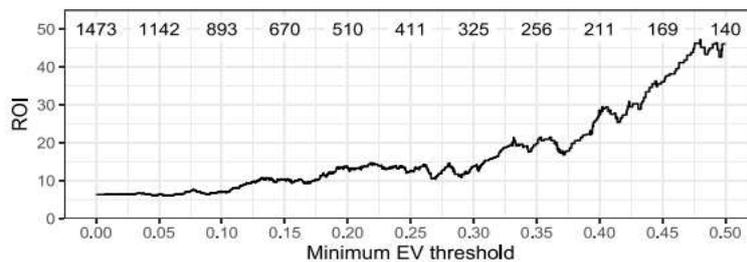


Figura 5. Grafico che illustra il ROI che si otterrebbe utilizzando il modello $skellam_{full}$ per scommettere con il criterio di Kelly modificato per diverse soglie di valore atteso minimo.

Capitolo 5

Conclusioni

In questa relazione è stato presentato un nuovo modello per la previsione di esiti di incontri calcistici. Il modello si basa sulle valutazioni e le interazioni dei singoli giocatori coinvolti nella partita a differenza dei modelli precedentemente proposti da [Maher \(1982\)](#) e [Dixon e Coles \(1997\)](#), i quali si fondavano su delle stime di valutazioni complessive di squadra.

Il progetto di [Holmes e McHale \(2023\)](#) prevede uno schema di attribuzione di *rating* individuali, il quale corregge delle valutazioni pubbliche per assicurarne la comparabilità nei diversi campionati. Successivamente, vengono introdotti dei modelli multinomiali utili alla stima del livello di interazione di ogni coppia di giocatori avversari in base ai propri ruoli e alle due formazioni.

La possibilità di implementare e utilizzare dei modelli così definiti è fortemente legata alla disponibilità e alla reperibilità di dati e informazioni, ma risolve il più grande problema dei modelli basati sulle valutazioni di squadra. Non è più necessario preoccuparsi di catturare il carattere dinamico del livello di gioco delle squadre: viene presa in considerazione direttamente la causa di questa variabilità, ossia i cambiamenti nei *roster*, nelle formazioni e nelle condizioni fisiche e sportive degli atleti.

Tuttavia, pur necessitando il modello della disponibilità di una grande mole di dati, database, come quello richiesto per l'implementazione del modello proposto nella relazione, oggi esistono e in futuro, con ogni probabilità, diventeranno sempre più ricchi di dati e accessibili.

Il modello è stato, successivamente, testato sia con delle *scoring rule*, che hanno suggerito un'ottima performance, anche quando confrontata con quella dei *bookmaker*, sia con la valutazione dei ricavi ottenuti con gli investimenti nel mercato delle scommesse sportive. Questi ultimi si sono rivelati positivi sulle scommesse 1X2.

Questi risultati hanno implicazioni sull'efficienza e sulla pratica di *trading* nel mondo calcistico. Per esempio, il modello qui proposto, potrebbe limitare la necessità dei *bookmaker* di affidarsi a *trader* esperti professionisti per correggere le proprie quote alla luce di nuove informazioni, come per esempio un cambiamento negli schieramenti iniziali delle due formazioni. Al momento ai *trader*

professionisti è richiesta una valutazione e una correzione soggettiva, il modello di cui si è discusso in questa relazione, invece, potrebbe farlo automaticamente.

Eventuali futuri studi, utilizzi e implementazioni di questo modello potrebbero rivelarsi interessanti. Si potrebbe, per esempio, stimare anche il livello di interazione tra i giocatori della stessa squadra: gli appassionati spesso notano che il livello di gioco di due o più giocatori quando sono contemporaneamente in campo è maggiore o inferiore alla somma dei livelli individuali in base al rapporto e all'intesa calcistica che vi è tra di loro. Un modello che include dei parametri dettati dall'interazione tra compagni di squadra riuscirebbe a catturare queste caratteristiche del gioco, altrimenti trascurate.

Un altro aspetto potenzialmente migliorabile del modello proposto è la scelta dei *rating* individuali. Qui sono state utilizzate le valutazioni proposte dal sito web WhoScored, ma queste potrebbero non essere perfette. Per esempio, è possibile che ci sia una distorsione delle valutazioni in favore degli attaccanti. Si potrebbe addirittura utilizzare il modello stesso per ottenere i *rating*: per esempio, un giudizio quantitativo al singolo giocatore potrebbe essere fornito dalla percentuale di passaggi riusciti nel corso della partita. Questa metrica potrebbe poi essere utilizzata dal modello e la performance finale risulterebbe indicatore dell'utilità della suddetta statistica per la previsione dei risultati finali degli incontri. Molte statistiche individuali potrebbero essere testate, confrontate e classificate in base alla loro utilità per la previsione.

Infine, il modello può essere usato anche per sviluppare dei programmi di *scouting* e reclutamento per le società sportive calcistiche, per ottenere una stima dell'impatto che il giocatore fonte di interesse avrebbe in una determinata squadra in un determinato campionato.

Riferimenti bibliografici

- Baker, R. D., & McHale, I. G. (2015). Time varying ratings in association football: the all-time greatest team is.. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*.
- Berrar, D., Lopes, P., & Dubitzky, W. (2019). Incorporating domain knowledge in machine learning for soccer outcome prediction. *Machine Learning*.
- Boshnakov, G., Kharrat, T., & McHale, I. (2016). A bivariate weibull count model for association football scores. *Journal of International Forecasting*.
- Brier, G. W. (1950). Verification of forecasts expressed in terms of probability. *Monthly Weather Review*.
- Constantinou, A. C. (2019). Dolores: a model that predicts football match outcomes from all over the world. *Machine Learning*.
- Crowder, M., Dixon, M., Ledford, A., & Robinson, M. (2002). Dynamic modelling and prediction of english football league matches for betting. *Journal of the Royal Statistical Society: Series D (the Statistician)*.
- Dixon, M. J., & Coles, S. G. (1997). Modelling association football scores and inefficiencies in the football betting market. *Journal of the Royal Statistical Society. Series C. Applied Statistics*.
- Dubitzky, W., Lopes, P., Davis, J., & Berrar, D. (2019). The Open International Soccer Database for machine learning. *Machine Learning*.
- Elo, A. E. (1978). *The rating of chessplayers, past and present*. New York: Arco Publishing.
- Holmes, B., McHale, I. G. (2023). Forecasting football match results using a player rating based model. *International Journal of Forecasting*.
- Hvattum, L. M., & Arntzen, H. (2010). Using elo ratings for match result prediction in association football. *International Journal of Forecasting*.
- Johnstone, D. J., Jones, S., Jose, V. R. R., & Peat, M. (2013). Measures of the economic value of probabilities of bankruptcy. *Journal of the Royal Statistical Society*

- Koopman, S. J., & Lit, R. (2015). A dynamic bivariate Poisson model for analysing and forecasting match results in the english premier league. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*.
- Maher, M. J. (1982). Modelling association football scores. *Statistica Neerlandica*.
- McShane, B., Adrian, M., Bradlow, E. T., & Fader, P. S. (2008). Count models based on Weibull interarrival times. *Journal of Business and Economic Statistics*.
- Owen, A. (2011). Dynamic bayesian forecasting models of football match outcomes with estimation of the evolution variance parameter. *IMA Journal of Management Mathematics*.
- Salvan, A., Sartori, N., Pace, L. (2020). *Modelli Lineari Generalizzati*.
- Skellam, J.G. (1946). The Frequency Distribution of the Difference Between Two Poisson Variates Belonging to Different Populations. *Journal of the Royal Statistical Society*.
- Wheatcroft, E. (2020). A profitable model for predicting the over/under market in football. *International Journal of Forecasting*.