

ALMA MATER STUDIORUM · UNIVERSITÀ DI BOLOGNA

SCUOLA DI SCIENZE
Corso di Laurea in Informatica per il Management

**ANALISI DI DATI CALCISTICI:
VALUTAZIONE E PREVISIONE
DEL VALORE ECONOMICO
DEI GIOCATORI DI RUOLO DIFENSIVO**

Relatore:
Chiar.mo Prof.
MARCO ROCCETTI

Presentata da:
LEONE DI LEONE

Sessione unica
Anno Accademico 2023-2024

Indice

Introduzione	1
1 I dati e la loro raccolta	3
1.1 Cosa sono i KPI	3
1.2 Fonte dei dati	5
1.3 Descrizione del database	6
1.3.1 Tabella "Giocatore"	7
1.3.2 Tabella "Statistiche_psicologiche"	12
1.3.3 Tabella "Statistiche_tattiche"	12
1.3.4 Tabella "Statistiche_tecniche"	13
2 Metodologia e modelli statistici	15
2.1 Introduzione alla regressione lineare	15
2.1.1 Significatività statistica e interpretazione del p-value	16
2.1.2 R-quadro e capacità predittiva del modello	17
2.1.3 Z-score e interpretazione dei coefficienti	18
2.2 Strumenti e linguaggi utilizzati	19
2.2.1 Caricamento e preparazione dei dati	19
2.2.2 Selezione dei KPI	21
2.2.3 Costruzione dei modelli	21
2.2.4 Analisi degli Outlier	22
3 Analisi dei risultati	23
3.1 Modello Generale per Tutti i Difensori	23
3.2 Modelli Specifici per Difensori Centrali e Terzini	25
3.2.1 Risultati modello Difensori centrali	26
3.2.2 Risultati modello Terzini	27
3.3 Selezione basata su p-value	29
3.3.1 Selezione Statistica	29
3.3.2 Selezione manuale	32
Conclusioni	39
Bibliografia	43

Introduzione

Negli ultimi anni, l'analisi dei dati ha rivoluzionato numerosi ambiti, compreso il mondo dello sport. L'uso di tecniche avanzate per lo studio delle prestazioni degli atleti ha permesso di migliorare il processo decisionale di club, scout e analisti. In particolare, il calcio ha visto un'evoluzione significativa nell'impiego delle statistiche per la valutazione oggettiva dei giocatori, rendendo sempre più centrale il ruolo della data analysis nelle strategie di mercato e nelle scelte tecnico-tattiche.

In questo contesto, la presente tesi è stata sviluppata nell'ambito di un progetto di ricerca finanziato dal Ministero dell'Università e della Ricerca (MUR) e co-diretto dal Prof. M. Rocchetti, volto all'analisi di dati sportivi.

L'uso dei dati e delle statistiche avanzate ha acquisito un ruolo sempre più centrale nell'analisi del calcio, consentendo una valutazione più oggettiva delle prestazioni dei giocatori. I Key Performance Indicators (KPI) si sono affermati come strumenti fondamentali per quantificare e confrontare il rendimento degli atleti, offrendo una base più solida rispetto alle tradizionali valutazioni soggettive. Tuttavia, mentre diversi studi hanno analizzato la relazione tra KPI e valore di mercato per alcune categorie di giocatori, come attaccanti e centrocampisti, la valutazione dei difensori, e in particolare dei difensori centrali, presenta sfide specifiche.

A differenza delle altre posizioni, il contributo di un difensore non si misura facilmente attraverso statistiche evidenti come gol, assist o dribbling. Elementi come posizionamento, lettura del gioco, organizzazione della linea difensiva e capacità di leadership hanno un peso rilevante nella determinazione del valore di mercato, ma risultano più difficili da quantificare attraverso i KPI tradizionali. Per questo motivo, risulta interessante indagare quali KPI influenzano effettivamente la valutazione economica dei difensori e se esistono differenze significative tra difensori centrali e terzini. Un'analisi di questo tipo può offrire spunti utili per club, scout e analisti sportivi, aiutandoli a interpretare meglio i dati nella valutazione dei giocatori.

Lo studio si propone di analizzare la relazione tra le prestazioni dei difensori e il loro valore di mercato, utilizzando i KPI per individuare quali metriche incidano maggiormente sulla valutazione economica di questi giocatori. In particolare, si vuole verificare se l'impatto dei KPI sul valore di mercato differisca tra difensori centrali e terzini, fornendo un quadro più dettagliato su come questi due ruoli vengano percepiti nel mercato

calcistico.

Per raggiungere questo obiettivo, è stato costruito un dataset basato sui difensori che militano in Serie A, analizzando le loro prestazioni nella stagione 2023/24. I dati relativi ai KPI sono stati raccolti da fonti specializzate nel settore dell'analisi calcistica, che forniscono informazioni dettagliate su diverse metriche di gioco, sia difensive che offensive. Il metodo analitico adottato è la regressione lineare multipla, che consente di stimare il peso di ciascun KPI nella determinazione del valore di mercato di un giocatore. Inoltre, l'analisi è stata condotta separatamente per difensori centrali e terzini, per verificare se l'influenza delle variabili dipenda dal ruolo specifico.

I risultati ottenuti evidenziano come il valore di mercato dei difensori sia effettivamente influenzato da diversi KPI, ma con differenze rilevanti tra le due categorie di giocatori analizzate. In particolare, il modello sviluppato ha mostrato una maggiore capacità predittiva per i terzini, il cui contributo alla squadra è più facilmente misurabile attraverso KPI come cross riusciti, dribbling e partecipazione alla fase offensiva. Al contrario, per i difensori centrali, la capacità predittiva del modello è risultata inferiore, suggerendo che il loro valore di mercato dipenda in modo più marcato da aspetti meno quantificabili, come leadership, esperienza e capacità di organizzare la linea difensiva.

Nonostante questi risultati, è importante considerare alcune limitazioni dello studio. Il modello utilizzato si basa su una regressione lineare multipla, che rappresenta una semplificazione delle dinamiche reali che determinano il valore di mercato di un giocatore. Relazioni più complesse tra KPI potrebbero essere meglio catturate da modelli più sofisticati, come algoritmi di machine learning. Inoltre, l'analisi è stata condotta su un campione limitato ai difensori di Serie A in una singola stagione, il che potrebbe ridurre la generalizzabilità dei risultati ad altri campionati o periodi temporali più estesi.

Nel corso della tesi verranno descritti i passaggi metodologici fondamentali che hanno portato allo sviluppo del modello di analisi. Si partirà dalla fase di raccolta e filtraggio dei dati, con particolare attenzione alla selezione dei giocatori e delle variabili indipendenti, per poi passare alla descrizione dei metodi analitici adottati, tra cui la regressione lineare multipla.

Infine, verranno discusse le principali limitazioni dello studio, legate sia alla disponibilità dei dati che alla metodologia utilizzata, e saranno proposti spunti per sviluppi futuri. Tra questi, l'estensione dell'analisi a più stagioni e campionati potrebbe consentire di ottenere risultati più generali e applicabili a un contesto calcistico più ampio. Inoltre, l'adozione di modelli più avanzati, come le reti neurali artificiali, potrebbe ampliare ulteriormente la portata dello studio, permettendo di cogliere relazioni più complesse tra le variabili e migliorando la capacità di prevedere il valore economico dei giocatori in contesti competitivi differenti.

Capitolo 1

I dati e la loro raccolta

1.1 Cosa sono i KPI

Negli ultimi anni, il concetto di Key Performance Indicators (KPI) ha assunto un ruolo centrale non solo in ambito aziendale, ma anche nel mondo dello sport e, in particolare, del calcio professionistico. L'uso di KPI ha permesso di trasformare le valutazioni delle prestazioni, un tempo soggettive, in misurazioni oggettive basate su dati numerici, rendendo l'analisi delle performance più precisa e strategica.

I KPI sono indicatori quantitativi che permettono di misurare il grado di successo rispetto a un determinato obiettivo. Come spiegato nell'articolo "*KPI: cosa sono, significato, esempi e come calcolarli (OFG, 2023)*", questi indicatori vengono utilizzati in ambito aziendale per monitorare il progresso di un'organizzazione rispetto ai suoi obiettivi strategici e operativi. Un KPI deve essere:

- Misurabile: espresso in numeri o percentuali.
- Rilevante: strettamente legato all'obiettivo che si intende analizzare.
- Comparabile: deve consentire confronti nel tempo o tra soggetti diversi.

Nel mondo del calcio, questa logica è stata applicata per valutare l'efficacia di un giocatore o di una squadra in modo oggettivo. Grazie ai KPI, gli allenatori, gli analisti e i dirigenti sportivi possono ottenere una visione chiara delle prestazioni e prendere decisioni strategiche basate su dati concreti.

Come evidenziato nell'articolo "*KPI: come scegliere quelli giusti (Auxiell, 2023)*", l'identificazione dei KPI più adatti dipende dal contesto di utilizzo e dagli obiettivi dell'analisi. Nel calcio, i KPI possono essere suddivisi in diverse categorie:

- KPI Tecnici: misurano la qualità delle giocate e delle azioni, come precisione nei passaggi, tiri in porta, dribbling riusciti.

- KPI Tattici: valutano l'adattabilità del giocatore al sistema di gioco, come il pressing medio, la posizione media in campo, il numero di recuperi palla.
- KPI Fisici: includono velocità media, distanza percorsa, intensità di corsa.
- KPI Psicologici: valutano aspetti più difficili da misurare, come la resilienza mentale o la capacità di mantenere alte prestazioni sotto pressione.

I KPI variano notevolmente in base al ruolo del giocatore e alla filosofia di gioco della squadra. Ad esempio, per un attaccante saranno fondamentali indicatori come gol segnati, expected goals e tiri in porta, mentre per un difensore centrale saranno più rilevanti intercetti, duelli aerei vinti e falli commessi.

Lo studio "*Moneyball and Soccer di Hughes (2021)*" evidenzia come, a differenza di altri sport, il calcio presenti una grande variabilità nelle prestazioni e sia influenzato da fattori di contesto difficili da quantificare. Questo significa che i KPI non devono mai essere interpretati in modo isolato, ma devono sempre essere contestualizzati.

Un esempio concreto riguarda il numero di contrasti vinti da un difensore centrale. A prima vista, si potrebbe pensare che un giocatore con un alto numero di contrasti sia un difensore più efficace. Tuttavia, questa metrica deve essere interpretata nel contesto del suo stile di gioco e del sistema difensivo adottato dalla squadra.

Un difensore che gioca in una squadra con una linea difensiva alta e aggressiva potrebbe essere coinvolto in un numero maggiore di duelli e, di conseguenza, registrare più contrasti vinti. D'altra parte, un difensore con eccellenti capacità di posizionamento e lettura del gioco potrebbe intercettare il pallone prima ancora che si renda necessario un contrasto fisico, risultando comunque molto efficace dal punto di vista difensivo, pur avendo un numero di contrasti inferiore. Pertanto, valutare un difensore solo sulla base del numero di contrasti vinti senza considerare il suo stile di gioco e quello della squadra potrebbe portare a una lettura distorta della sua reale efficacia difensiva.

Uno degli aspetti più interessanti dell'utilizzo dei KPI nel calcio è la loro applicazione nella valutazione economica dei giocatori. Come discusso nell'articolo "*KPI aziendali: definizione, a cosa servono e come stabilirli (Business Coaching Italia, 2023)*", i KPI sono strumenti essenziali per stimare il valore di un'azienda o di un reparto. Nel calcio, questa stessa logica è stata adottata per correlare il valore di mercato di un giocatore alle sue prestazioni sul campo.

Tuttavia, il livello di predittività dei KPI dipende molto dal ruolo del giocatore. Mentre per attaccanti e centrocampisti esistono indicatori facilmente correlabili al valore economico (gol, assist, precisione nei passaggi chiave), per i difensori centrali questa correlazione risulta più complessa. Il loro contributo al successo della squadra è meno immediato da quantificare e dipende da aspetti più difficili da misurare, come il posizionamento difensivo, la leadership e la capacità di organizzare la linea difensiva.

L'utilizzo dei KPI nel calcio ha trasformato radicalmente il modo in cui vengono analizzate le prestazioni dei giocatori. Grazie a queste metriche, è possibile adottare un approccio più oggettivo e basato su dati, riducendo la soggettività delle valutazioni tradizionali.

Tuttavia, affinché i KPI siano realmente efficaci, devono essere scelti con attenzione, sempre aggiornati e soprattutto contestualizzati in base alle caratteristiche del ruolo e della squadra. Nel caso specifico dei difensori, che sono il focus di questa ricerca, la sfida è identificare gli indicatori più significativi e verificare se sia possibile costruire un modello predittivo affidabile che colleghi le loro prestazioni al valore di mercato.

1.2 Fonte dei dati

Il database utilizzato per questa analisi è stato costruito attraverso un processo accurato di raccolta e selezione dei dati, con l'obiettivo di garantire una rappresentazione il più completa e affidabile possibile delle prestazioni dei difensori di Serie A. La principale fonte per la determinazione del valore economico dei giocatori è stata Transfermarkt, piattaforma di riferimento per le valutazioni di mercato, mentre per l'estrazione dei KPI statistici è stato utilizzato Sofascore, che fornisce dati dettagliati sulle performance individuali in campo. La selezione dei parametri chiave da analizzare è stata inoltre guidata dalla letteratura scientifica, in particolare dallo studio "*Moneyball and Soccer: An Analysis of the Key Performance Indicators of Elite Male Soccer Players by Position*", che ha offerto una base teorica solida per individuare le metriche più rilevanti per i diversi ruoli difensivi.

Per garantire una maggiore coerenza nell'analisi, il periodo di riferimento scelto è stato l'intero campionato di Serie A della stagione scorsa (2023/24), permettendo così di avere una visione completa delle prestazioni dei giocatori nell'arco di un'intera annata sportiva. Il database è stato strutturato in tabelle che raccolgono informazioni sia anagrafiche che prestazionali, includendo variabili legate agli aspetti offensivi, difensivi e di costruzione del gioco. Le statistiche selezionate sono state scelte con particolare attenzione per garantire un'elevata capacità predittiva, assicurandosi che rappresentassero al meglio le caratteristiche del ruolo difensivo e la loro influenza sul valore di mercato.

L'affidabilità delle fonti consultate ha consentito di costruire un dataset solido, utile per condurre analisi inferenziali e modelli predittivi robusti. La combinazione tra i dati di mercato di Transfermarkt e le metriche prestazionali di Sofascore ha reso possibile lo studio della relazione tra le performance in campo e la valutazione economica dei difensori, offrendo un approccio quantitativo alla determinazione del valore di mercato. Il risultato è un database che rappresenta in modo accurato il rendimento dei gioca-

tori, consentendo di esplorare con maggiore precisione il legame tra KPI e valutazione economica nel contesto calcistico.

1.3 Descrizione del database

Per la gestione e l'organizzazione dei dati raccolti, ho utilizzato un database relazionale MySQL, implementato tramite XAMPP. Questa soluzione mi ha permesso di strutturare le informazioni in tabelle relazionate tra loro, facilitando l'accesso, la manipolazione e l'analisi dei dati. Grazie all'utilizzo di SQL, è stato possibile eseguire query efficienti per filtrare, aggregare e unire i dati provenienti da diverse fonti, garantendo così una gestione ottimale delle informazioni relative ai KPI e al valore di mercato dei giocatori. L'uso di un database relazionale ha inoltre assicurato coerenza e integrità ai dati, riducendo al minimo il rischio di duplicazioni o errori durante le fasi di analisi.

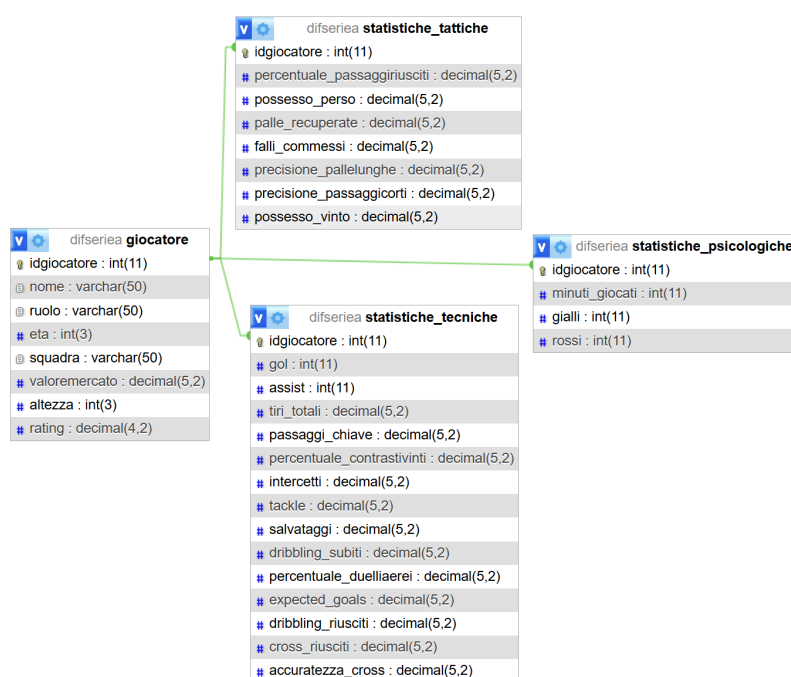


Figura 1.1: Schema relazionale del database dei Difensori di Serie A

L'immagine illustra il modello del database relazionale sviluppato per questa analisi, concepito come una raccolta strutturata di dati per lo studio dettagliato delle prestazioni dei difensori di Serie A. La costruzione di questa base di dati è stata il risultato di una pianificazione attenta e di un'accurata selezione delle variabili, finalizzata a rappresentare in modo completo e fedele le performance tattiche, tecniche e comportamentali

dei difensori. Questo approccio strutturato consente di collegare i vari aspetti delle prestazioni individuali a una valutazione economica, fornendo uno strumento integrato per analisi inferenziali e predittive.

Il database è composto da quattro tabelle principali:

- **Giocatore:** tabella principale che raccoglie i dati anagrafici e le informazioni di base.
- **Statistiche Psicologiche:** Analizzano il comportamento in campo e l'aspetto disciplinare.
- **Statistiche Tattiche:** valutano il contributo di un giocatore al sistema di gioco della squadra e la loro capacità di rispettare le indicazioni tattiche. Sono fondamentali per comprendere come un giocatore si integra nel collettivo.
- **Statistiche Tecniche:** si riferiscono alle abilità individuali di un giocatore con la palla, la loro precisione e la capacità di eseguire azioni di gioco in modo efficace. Sono utilizzati per valutare la competenza tecnica di un giocatore, sia in fase offensiva che difensiva.

1.3.1 Tabella "Giocatore"

La tabella "Giocatore" rappresenta la parte centrale del database, facendo da punto di connessione con tutte le altre tabelle. Contiene informazioni identificative e anagrafiche, nonché dati chiave per l'analisi. Contiene i seguenti campi:

- **idgiocatore:** identificatore univoco usato per garantire l'integrità delle relazioni tra le tabelle.
- **nome:** identifica il calciatore con i suoi dati anagrafici principali (nome e cognome).
- **ruolo:** distingue il difensore in "difensore centrale" o "terzino". È una distinzione importante per l'interpretazione delle statistiche.
- **eta:** indica l'età del giocatore, utile all'analisi per valutare l'esperienza maturata o un potenziale per il futuro.
- **squadra:** specifica il club di appartenenza del giocatore.
- **valoremercato:** stima economica del calciatore, espressa in milioni di euro.
- **altezza:** rappresenta l'altezza fisica del calciatore, espressa in centimetri.

- rating: valutazione sintetica del calciatore nella stagione 2023/24. Questo punteggio può derivare da analisi statistiche avanzate o da valutazioni di esperti.

Nella tabella di seguito sono rappresentati tutti i difensori presenti nel database su cui ho sviluppato l'analisi.

Tabella 1.1: Tabella completa dei difensori

Nome	Ruolo	Età	Valore di Mercato (mln €)
Alessandro Bastoni	Difensore Centrale	25	75,00
Bremer	Difensore Centrale	27	60,00
Giorgio Scalvini	Difensore Centrale	21	45,00
Alessandro Buongiorno	Difensore Centrale	25	45,00
Benjamin Pavard	Difensore Centrale	28	40,00
Yann Bissec	Difensore Centrale	24	30,00
Isak Hien	Difensore Centrale	26	30,00
Odilon Kossounou	Difensore Centrale	24	30,00
Mario Gila	Difensore Centrale	24	28,00
Pierre Kalulu	Difensore Centrale	24	26,00
Evan Ndicka	Difensore Centrale	25	25,00
Federico Gatti	Difensore Centrale	26	25,00
Fikayo Tomori	Difensore Centrale	27	22,00
Sam Beukema	Difensore Centrale	26	22,00
Matteo Gabbia	Difensore Centrale	25	20,00
Malick Thiaw	Difensore Centrale	23	20,00
Gianluca Mancini	Difensore Centrale	28	18,00
Jaka Bijol	Difensore Centrale	25	18,00
Jhon Lucumí	Difensore Centrale	26	18,00
Strahinja Pavlović	Difensore Centrale	23	17,00
Perr Schuurs	Difensore Centrale	25	15,00
Pietro Comuzzo	Difensore Centrale	19	15,00
Saúl Coco	Difensore Centrale	25	14,00
Luca Ranieri	Difensore Centrale	25	12,00
Alessio Romagnoli	Difensore Centrale	30	12,00
Amir Rrahmani	Difensore Centrale	30	12,00
Koni De Winter	Difensore Centrale	22	12,00
Sead Kolasinac	Difensore Centrale	31	10,00
Mario Hermoso	Difensore Centrale	29	10,00
Marin Pongračić	Difensore Centrale	27	10,00
Johan Vásquez	Difensore Centrale	26	10,00
Rafa Marín	Difensore Centrale	22	9,00

Nome	Ruolo	Età	Valore di Mercato (mln €)
Alessandro Circati	Difensore Centrale	21	8,50
Berat Djimsiti	Difensore Centrale	31	8,00
Alan Matturro	Difensore Centrale	20	8,00
Oumar Solet	Difensore Centrale	24	8,00
Isaak Touré	Difensore Centrale	21	8,00
Diego Coppola	Difensore Centrale	21	7,50
Stefan de Vrij	Difensore Centrale	32	7,00
Nicolò Casale	Difensore Centrale	26	7,00
Alberto Dossena	Difensore Centrale	26	7,00
Mattia Viti	Difensore Centrale	23	6,50
Tomás Palacios	Difensore Centrale	21	6,00
Ardian Ismajli	Difensore Centrale	28	6,00
Martin Erlić	Difensore Centrale	27	5,00
Sebastian Walukiewicz	Difensore Centrale	24	5,00
Federico Baschirotto	Difensore Centrale	28	5,00
Giovanni Leoni	Difensore Centrale	18	4,50
Thomas Kristensen	Difensore Centrale	23	4,50
Nicolás Valentini	Difensore Centrale	23	4,50
Flavius Daniliuc	Difensore Centrale	23	4,00
Patric	Difensore Centrale	31	4,00
Samuel Gigot	Difensore Centrale	31	4,00
Guillermo Maripán	Difensore Centrale	30	4,00
Danilo	Difensore Centrale	33	4,00
Andrea Carboni	Difensore Centrale	23	4,00
Mats Hummels	Difensore Centrale	36	3,50
Sebastiano Luperto	Difensore Centrale	28	3,50
Pablo Marí	Difensore Centrale	31	3,50
Francesco Acerbi	Difensore Centrale	36	3,00
Michael Svoboda	Difensore Centrale	26	3,00
Jay Idzes	Difensore Centrale	24	3,00
Yerry Mina	Difensore Centrale	30	3,00
Adam Obert	Difensore Centrale	22	3,00
Botond Balogh	Difensore Centrale	22	2,50
Paweł Dawidowicz	Difensore Centrale	29	2,50
Giangiaco­mo Magnani	Difensore Centrale	29	2,50
Lautaro Giannetti	Difensore Centrale	31	2,50
Saba Sazonov	Difensore Centrale	22	2,50
Kialonda Gaspar	Difensore Centrale	27	2,50
Marc Oliver Kempf	Difensore Centrale	29	2,50
Armando Izzo	Difensore Centrale	32	2,50

Nome	Ruolo	Età	Valore di Mercato (mln €)
Alessandro Vogliacco	Difensore Centrale	26	2,00
Alessandro Marcandalli	Difensore Centrale	22	2,00
Joël Schingtienne	Difensore Centrale	22	2,00
Giorgio Altare	Difensore Centrale	26	1,80
Yordan Osorio	Difensore Centrale	30	1,70
Daniele Ghilardi	Difensore Centrale	22	1,60
Rafael Tolói	Difensore Centrale	34	1,50
Adam Masina	Difensore Centrale	31	1,50
Juan Jesus	Difensore Centrale	33	1,50
Lautaro Valenti	Difensore Centrale	26	1,50
Fali Candé	Difensore Centrale	27	1,50
Edoardo Goldaniga	Difensore Centrale	31	1,20
Marin Šverko	Difensore Centrale	26	1,20
Luca Caldirola	Difensore Centrale	33	1,00
Kevin Bonifazi	Difensore Centrale	28	1,00
Gaby Jean	Difensore Centrale	24	1,00
Christian Kabasele	Difensore Centrale	33	1,00
Yllan Okou	Difensore Centrale	22	0,90
Daniilo D'Ambrosio	Difensore Centrale	36	0,80
José Luis Palomino	Difensore Centrale	35	0,70
Federico Dimarco	Terzino Sinistro	27	60,00
Theo Hernández	Terzino Sinistro	27	50,00
Andrea Cambiaso	Terzino Sinistro	24	40,00
Nuno Tavares	Terzino Sinistro	24	25,00
Carlos Augusto	Terzino Sinistro	26	22,00
Mathías Olivera	Terzino Sinistro	27	18,00
Juan Miranda	Terzino Sinistro	25	11,00
Juan Cabal	Terzino Sinistro	24	9,00
Angeliño	Terzino Sinistro	28	8,00
Robin Gosens	Terzino Sinistro	30	8,00
Antonino Gallo	Terzino Sinistro	25	5,50
Aarón Martín	Terzino Sinistro	27	5,00
Jordan Zemura	Terzino Sinistro	25	5,00
Fabiano Parisi	Terzino Sinistro	24	5,00
Domagoj Bradarić	Terzino Sinistro	25	4,00
Samuel Dahl	Terzino Sinistro	21	4,00
Luca Pellegrini	Terzino Sinistro	25	3,50
Hassane Kamara	Terzino Sinistro	30	3,50
Emanuele Valeri	Terzino Sinistro	26	3,00
Leonardo Spinazzola	Terzino Sinistro	31	2,50

Nome	Ruolo	Età	Valore di Mercato (mln €)
Giuseppe Pezzella	Terzino Sinistro	27	2,50
Martin Frese	Terzino Sinistro	27	2,00
Cristiano Biraghi	Terzino Sinistro	32	2,00
Liberato Cacace	Terzino Sinistro	24	2,00
Alberto Moreno	Terzino Sinistro	32	1,80
Tommaso Augello	Terzino Sinistro	30	1,80
Ridgeciano Haps	Terzino Sinistro	31	1,00
Charalampos Lykogiannis	Terzino Sinistro	31	1,00
Marco Sala	Terzino Sinistro	25	1,00
Franco Carboni	Terzino Sinistro	21	0,90
Gianluca Di Chiara	Terzino Sinistro	31	0,50
Dodô	Terzino Destro	26	22,00
Denzel Dumfries	Terzino Destro	28	20,00
Giovanni Di Lorenzo	Terzino Destro	31	15,00
Emerson Royal	Terzino Destro	26	14,00
Devyne Rensch	Terzino Destro	22	12,00
Stefan Posch	Terzino Destro	27	10,00
Kyle Walker	Terzino Destro	34	10,00
Jackson Tchatchoua	Terzino Destro	23	8,00
Davide Calabria	Terzino Destro	28	7,50
Emil Holm	Terzino Destro	24	7,00
Enrico Delprato	Terzino Destro	25	6,00
Alessandro Zanolì	Terzino Destro	24	6,00
Zeki Çelik	Terzino Destro	27	5,00
Brooke Norton-Cuffy	Terzino Destro	21	4,50
Filippo Terracciano	Terzino Destro	21	4,50
Matteo Darmian	Terzino Destro	35	4,00
Adam Marusic	Terzino Destro	32	4,00
Saud Abdulhamid	Terzino Destro	25	4,00
Samuele Birindelli	Terzino Destro	25	3,50
Gabriele Zappa	Terzino Destro	25	3,50
Kingsley Ehizibue	Terzino Destro	29	3,00
Ignace Van der Brempt	Terzino Destro	22	3,00
Pasquale Mazzocchi	Terzino Destro	29	3,00
Mërgim Vojvoda	Terzino Destro	29	3,00
Frédéric Guilbert	Terzino Destro	30	2,50
Antonio Candela	Terzino Destro	24	2,30
Richie Sagrado	Terzino Destro	20	2,00
Elseid Hysaj	Terzino Destro	30	2,00
Pedro Pereira	Terzino Destro	27	1,80

Nome	Ruolo	Età	Valore di Mercato (mln €)
Tyronne Ebuehi	Terzino Destro	29	1,50
Stefano Sabelli	Terzino Destro	32	1,50
Alessandro Florenzi	Terzino Destro	33	1,50
Junior Sambia	Terzino Destro	28	1,40
Francesco Zampano	Terzino Destro	31	1,10
Mattia De Sciglio	Terzino Destro	32	1,00
Davide Faraoni	Terzino Destro	33	1,00
Lorenzo De Silvestri	Terzino Destro	36	0,40

1.3.2 Tabella "Statistiche_psicologiche"

La tabella delle statistiche psicologiche esamina il comportamento del calciatore nel corso della stagione, evidenziando aspetti legati alla sua disciplina e resilienza mentale. Contiene i seguenti campi:

- `idgiocatore`: collegamento alla tabella "giocatore".
- `minuti_giocati`: numero totale di minuti giocati, ci aiuta a capire la continuità del giocatore.
- `gialli`: numero totale di cartellini gialli ricevuti, riflette il comportamento del giocatore in campo.
- `rossi`: rappresenta l'impatto disciplinare negativo del giocatore, riflettendo situazioni di espulsione diretta o doppia ammonizione.

1.3.3 Tabella "Statistiche_tattiche"

Le statistiche tattiche valutano quanto un calciatore incida sulle dinamiche di squadra. Contiene i seguenti campi:

- `idgiocatore`: collegamento alla tabella "giocatore".
- `percentuale_passaggiriusciti`: percentuale di passaggi completati con successo rispetto al totale dei passaggi tentati, indicativa dell'accuratezza nel palleggio.
- `possesso_perso`: numero di volte in cui il giocatore ha perso il controllo del pallone a favore degli avversari, sia per errore tecnico che per pressione avversaria.
- `palle_recuperate`: numero medio di volte a partita in cui il giocatore riconquista il possesso del pallone attraverso intercetti, contrasti vinti o errori avversari, senza commettere fallo.

- `falli_commessi`: numero medio di falli commessi dal giocatore a partita, indicativo dell'aggressività difensiva e della propensione al rischio.
- `precisione_pallelunghe`: percentuale di passaggi lunghi riusciti rispetto al totale tentato, utile per valutare la capacità di costruzione del gioco a distanza.
- `precisione_passaggicorti`: percentuale di passaggi corti riusciti rispetto al totale tentato, indicativa dell'abilità nel fraseggio ravvicinato.
- `possesso_vinto`: numero di volte in cui il giocatore recupera il possesso del pallone per la propria squadra in situazioni di duello diretto, sia a terra che aereo.

1.3.4 Tabella "Statistiche_tecniche"

Le statistiche tecniche analizzano le abilità individuali di un giocatore, valutano la competenza tecnica di un giocatore sia in fase offensiva che difensiva. Contiene i seguenti campi:

- `idgiocatore`: collegamento alla tabella "giocatore".
- `gol`: numero totale di gol segnati dal giocatore durante la stagione, indicatore chiave dell'efficacia offensiva.
- `assist`: numero totale di passaggi che hanno portato direttamente a un gol, rappresentando la capacità del giocatore di creare occasioni per i compagni
- `tiri_totali`: numero medio di tiri totali effettuati a partita, misura il volume di tentativi di conclusione verso la porta.
- `passaggi_chiave`: numero medio di passaggi per partita che creano un'occasione da gol, riflette la visione di gioco e la capacità di costruzione offensiva.
- `percentuale_contrastivinti`: percentuale di contrasti a terra vinti rispetto al totale dei contrasti tentati, rappresenta l'efficacia del giocatore nei duelli difensivi.
- `intercetti`: numero medio di palloni intercettati per partita, indica la capacità di interrompere le linee di passaggio avversarie
- `tackle`: numero medio di contrasti effettuati per partita, rappresenta la frequenza con cui il giocatore tenta di sottrarre il pallone agli avversari.
- `salvataggi`: numero medio di interventi difensivi decisivi per partita, inclusi respinte sulla linea di porta o blocchi di tiri avversari.

- `dribbling_subiti`: numero medio di volte per partita in cui il giocatore viene superato da un avversario in dribbling, evidenzia la sua vulnerabilità nei duelli individuali.
- `percentuale_duelliaerei`: percentuale di duelli aerei vinti rispetto al totale tentato, misura la capacità del giocatore di imporsi nei contrasti aerei.
- `expected_goals`: valore previsto dei gol che il giocatore avrebbe dovuto segnare sulla base della qualità delle occasioni avute, utilizzato per valutare l'efficacia realizzativa.
- `dribbling_riusciti`: numero medio di dribbling completati con successo per partita, misura l'abilità del giocatore nel superare l'avversario con la conduzione palla.
- `cross_riusciti`: numero medio di cross completati con successo per partita, rappresenta la precisione e l'efficacia dei traversoni effettuati dal giocatore.
- `accuratezza_cross`: percentuale di cross completati con successo rispetto al totale dei cross tentati, indica la precisione nei passaggi effettuati dalle fasce.

Capitolo 2

Metodologia e modelli statistici

2.1 Introduzione alla regressione lineare

L'analisi dei dati ha assunto un ruolo centrale nel calcio moderno, consentendo di misurare e valutare le prestazioni dei giocatori attraverso strumenti statistici avanzati. In questo contesto, la regressione lineare rappresenta una delle tecniche più utilizzate per modellare e comprendere le relazioni tra diverse variabili. Si tratta di un metodo statistico che consente di prevedere il valore di una variabile dipendente (ovvero quella che vogliamo stimare) in base a una o più variabili indipendenti, che agiscono come fattori esplicativi.

Nel caso specifico di questa analisi, l'obiettivo è determinare se esiste una relazione significativa tra il valore di mercato di un giocatore difensivo e una serie di Key Performance Indicators (KPI) che descrivono le sue prestazioni in campo. In altre parole, la regressione lineare permette di verificare se e in che misura le statistiche di gioco, come il numero di contrasti vinti, gli intercetti effettuati o la precisione nei passaggi, possano essere utilizzate per spiegare e prevedere la valutazione economica di un calciatore. Esistono due principali varianti della regressione lineare:

- Regressione lineare semplice, in cui il valore della variabile dipendente è spiegato da una sola variabile indipendente.
- Regressione lineare multipla, in cui la variabile dipendente è influenzata da più fattori contemporaneamente.

Poiché il valore di mercato di un calciatore non dipende da un singolo KPI, ma da una combinazione di diversi aspetti delle sue prestazioni, in questo studio ho adottato la regressione lineare multipla, che consente di includere più variabili esplicative nel modello, offrendo una rappresentazione più accurata della realtà.

Matematicamente, la regressione lineare multipla è espressa dalla seguente equazione:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n + \varepsilon$$

Dove:

- Y : la variabile dipendente, in questo caso il valore di mercato del giocatore.
- β_0 : intercetta, ovvero il valore di Y quando tutte le variabili indipendenti sono pari a zero.
- $\beta_1, \beta_2, \dots, \beta_n$: coefficienti di regressione, che indicano l'effetto di una variazione unitaria in una variabile indipendente su Y , mantenendo costanti le altre variabili.
- $X_1, X_2, \dots, X_n$: variabili indipendenti (i KPI selezionati).
- ε : errore residuo, che rappresenta la variazione in Y non spiegata dal modello.

L'adozione della regressione lineare multipla è giustificata dalla natura multidimensionale della valutazione economica di un calciatore. Il valore di mercato non è determinato esclusivamente da una singola statistica, ma è il risultato dell'interazione tra molteplici fattori. Ad esempio, un difensore potrebbe avere un'elevata percentuale di duelli aerei vinti, ma se al tempo stesso commette molti falli o ha una precisione nei passaggi relativamente bassa, il suo valore complessivo potrebbe risultare inferiore rispetto a un altro giocatore con caratteristiche più equilibrate.

Questo approccio consente di integrare l'intuizione calcistica tradizionale con un metodo analitico basato sui dati, rendendo informazioni complesse strumenti concreti per l'analisi. Per valutare l'efficacia del modello di regressione sviluppato, vengono calcolate metriche matematiche come R^2 , p-value e Z-score, che permettono di misurarne la precisione e l'affidabilità. Questi indicatori statistici forniscono informazioni essenziali per comprendere il modello e selezionare le variabili (KPI) più significative, contribuendo a una previsione accurata del valore di mercato dei giocatori. Nei prossimi paragrafi, approfondiremo il significato di queste metriche e il loro ruolo nell'analisi.

2.1.1 Significatività statistica e interpretazione del p-value

La significatività statistica è un concetto fondamentale nell'analisi dei dati e nella regressione lineare, poiché consente di determinare se l'effetto osservato di una variabile indipendente sulla variabile dipendente sia reale o semplicemente dovuto al caso. Per valutare questa significatività, si utilizza il p-value, un indicatore numerico che misura la probabilità che il risultato ottenuto sia frutto della casualità piuttosto che di una relazione effettiva tra le variabili.

Il **p-value** si basa sul valore $t_{\text{osservato}}$, che viene calcolato nel seguente modo:

$$t_{\text{osservato}} = Z = \frac{\hat{\beta}_i}{\text{SE}(\hat{\beta}_i)}$$

dove:

- $\hat{\beta}_i$ rappresenta il coefficiente stimato della variabile indipendente X_i ,
- $\text{SE}(\hat{\beta}_i)$ è l'errore standard associato a $\hat{\beta}_i$.

Una volta calcolato il valore di $t_{\text{osservato}}$, il **p-value** viene determinato come:

$$p\text{-value} = P(|t| > t_{\text{osservato}})$$

Questa espressione indica la probabilità che il valore assoluto della statistica del test t sia maggiore rispetto a quello osservato, assumendo che il coefficiente β_i sia effettivamente pari a zero (ossia, non abbia alcun impatto sulla variabile dipendente).

Il p-value viene poi confrontato con un livello di significatività prestabilito, tipicamente fissato a 0.05. Se il p-value di una variabile è inferiore a questa soglia, si considera che l'effetto di quella variabile sulla variabile dipendente sia statisticamente significativo, il che implica che è altamente improbabile che la relazione osservata sia dovuta al caso. Al contrario, se il p-value è superiore a 0.05, la variabile potrebbe non avere un'influenza rilevante e potrebbe essere esclusa dal modello senza comprometterne la qualità predittiva.

Tuttavia, è importante sottolineare che un p-value basso non implica necessariamente una forte correlazione tra le variabili, ma indica semplicemente che l'effetto della variabile indipendente sulla variabile dipendente è difficilmente attribuibile al caso. Per questo motivo, la valutazione del p-value deve essere sempre affiancata da altre metriche statistiche, come il coefficiente di determinazione R^2 e lo Z-score, per ottenere un quadro più completo sulla validità del modello di regressione.

2.1.2 R-quadro e capacità predittiva del modello

Il coefficiente di determinazione, noto come R^2 , è una delle metriche fondamentali per valutare la capacità di un modello di regressione di spiegare la variabilità della variabile dipendente. Esso fornisce una misura della bontà del modello, indicando in che percentuale le variazioni della variabile dipendente Y possono essere attribuite alle variabili indipendenti X_i incluse nel modello.

Matematicamente, R^2 è definito come:

$$R^2 = 1 - \frac{SS_{\text{residuo}}}{SS_{\text{totale}}}$$

dove:

- $SS_{\text{residuo}} = \sum (Y_i - \hat{Y}_i)^2$ rappresenta la somma dei quadrati dei residui, ovvero la differenza tra i valori osservati Y_i e quelli predetti dal modello $\hat{Y}_i$. Indica la parte della variabilità di Y che il modello non è stato in grado di spiegare.
- $SS_{\text{totale}} = \sum (Y_i - \bar{Y})^2$ è la somma dei quadrati totali, che rappresenta la variabilità complessiva dei dati rispetto alla loro media $\bar{Y}$.

Il valore di R^2 varia tra 0 e 1. Un valore vicino a 1 indica che il modello è in grado di spiegare la quasi totalità della variabilità della variabile dipendente, mentre un valore vicino a 0 suggerisce che il modello ha una scarsa capacità esplicativa.

Tuttavia, un R^2 elevato non garantisce necessariamente che il modello sia affidabile. Infatti, potrebbe adattarsi bene ai dati di addestramento ma avere una scarsa capacità predittiva su dati nuovi (overfitting). Per questo motivo, è fondamentale valutare R^2 insieme ad altre metriche statistiche, come il p-value dei coefficienti e lo Z-score, per avere un quadro più completo sulla bontà del modello.

2.1.3 Z-score e interpretazione dei coefficienti

Nell'analisi della regressione lineare multipla, oltre al p-value e al coefficiente di determinazione R^2 , un'altra misura fondamentale per valutare l'importanza delle variabili indipendenti è lo Z-score. Questo valore indica quanto un determinato coefficiente di regressione si discosta dalla media in termini di deviazioni standard, consentendo di confrontare l'impatto relativo delle variabili indipendenti all'interno del modello.

Lo Z-score viene calcolato secondo la seguente formula:

$$Z = \frac{\hat{\beta}_i}{SE(\hat{\beta}_i)}$$

dove:

- $\hat{\beta}_i$ è il coefficiente stimato della variabile indipendente X_i , ovvero la misura dell'impatto che tale variabile ha sulla variabile dipendente.
- $SE(\hat{\beta}_i)$ rappresenta l'errore standard associato alla stima del coefficiente, che indica la precisione della stima.

Uno Z-score elevato (in valore assoluto) suggerisce che la variabile indipendente ha un impatto significativo sulla variabile dipendente. In generale, valori di $|Z| > 2$ indicano che il coefficiente $\hat{\beta}_i$ è statisticamente significativo con un livello di confidenza superiore al 95%, mentre valori inferiori possono indicare che l'effetto della variabile potrebbe essere dovuto alla casualità.

Interpretare correttamente lo Z-score è essenziale per comprendere quali KPI abbiano un reale effetto sulla valutazione del valore di mercato dei difensori. Infatti, un valore alto

di Z-score suggerisce che il KPI ha una forte influenza nel modello, mentre un valore basso potrebbe indicare che la variabile non è particolarmente rilevante. Tuttavia, è sempre importante considerare lo Z-score insieme al p-value e al R^2 per ottenere un'analisi più completa della validità e dell'affidabilità del modello.

2.2 Strumenti e linguaggi utilizzati

Nell'ambito di questa analisi, è stato necessario l'utilizzo di diversi strumenti software e linguaggi di programmazione per gestire, elaborare e modellare i dati relativi ai KPI dei difensori e alla loro valutazione economica. La scelta degli strumenti si è basata su criteri di efficienza, affidabilità e capacità di supportare analisi statistiche avanzate.

L'ambiente di sviluppo principale utilizzato è stato Python, un linguaggio di programmazione ampiamente diffuso per l'analisi dei dati e il machine learning. In particolare, sono state impiegate librerie specializzate come pandas per la gestione dei dataset, numpy per le operazioni numeriche e statsmodels per la costruzione dei modelli di regressione lineare multipla. L'uso di matplotlib e seaborn ha inoltre permesso la visualizzazione dei dati attraverso grafici e rappresentazioni intuitive, fondamentali per l'interpretazione dei risultati.

Per la gestione e l'archiviazione dei dati, è stato adottato un database relazionale MySQL, implementato tramite XAMPP. Questo ha consentito di strutturare le informazioni in modo organizzato e di eseguire interrogazioni SQL efficienti per l'estrazione e l'aggregazione dei dati. Il database è stato progettato per garantire integrità e coerenza, minimizzando la ridondanza delle informazioni e facilitando l'integrazione tra le diverse fonti di dati.

Il processo di analisi dei difensori di Serie A si basa su una pipeline strutturata che integra diverse fasi: il caricamento dei dati, la loro pulizia e normalizzazione, la selezione delle variabili (KPI) e la creazione di modelli di regressione per identificare i fattori determinanti del valore di mercato dei giocatori. Il sistema è implementato attraverso una serie di script Python modulari, ciascuno con funzioni specifiche.

Il codice sorgente, la base di dati e i modelli sviluppati saranno visualizzabili all'interno di questa repository GitHub al seguente link:

<https://github.com/leonedi/tesiDifensori>

2.2.1 Caricamento e preparazione dei dati

L'analisi dei difensori di Serie A si basa su un dataset contenente un'ampia gamma di informazioni relative alle performance dei giocatori. Per garantire la coerenza e la completezza delle informazioni, i dati vengono caricati direttamente da un database relazionale. Questo database memorizza statistiche dettagliate su diversi aspetti del gioco, tra cui dati fisiologici, psicologici, tecnici, tattici e relativi alle performance difensive e

offensive.

Il caricamento dei dati avviene attraverso una connessione al database MySQL, che permette di eseguire interrogazioni mirate per estrarre solo le informazioni necessarie. L'integrazione delle diverse tabelle in un unico dataset consente di ottenere un quadro dettagliato delle caratteristiche e delle prestazioni di ciascun giocatore. Durante questa fase, è possibile applicare dei filtri per selezionare specifici gruppi di giocatori, come difensori centrali o terzini, in modo da facilitare analisi mirate per ogni ruolo.

Una volta caricati i dati dal database, il dataset viene sottoposto a un processo di pulizia per eliminare eventuali anomalie e garantire la qualità dell'analisi. Questo processo prevede la rimozione di valori mancanti, duplicati e incoerenze nei dati. La pulizia è una fase cruciale per evitare distorsioni nei risultati dell'analisi e garantire che ogni osservazione nel dataset sia accurata e affidabile.

Successivamente, i dati vengono organizzati in un formato strutturato, in cui ogni variabile è chiaramente definita e classificata. Questa strutturazione permette di facilitare le analisi successive, assicurando che tutte le informazioni siano disposte in modo coerente e facilmente accessibile. Inoltre, viene applicato un primo filtro sui dati per considerare solo i giocatori con un numero minimo di minuti giocati, in modo da escludere coloro che hanno avuto un impiego marginale nel campionato.

La decisione di applicare questo filtro è giustificata dalla necessità di basare l'analisi su un campione rappresentativo di giocatori che abbiano accumulato un minutaggio sufficiente da permettere una valutazione attendibile delle loro prestazioni. Giocatori con pochi minuti all'attivo potrebbero non fornire dati significativi e introdurre variabilità indesiderata nei modelli predittivi.

Per rendere il dataset adatto all'analisi statistica, i dati numerici vengono normalizzati e standardizzati. La normalizzazione è un passaggio essenziale per garantire che le variabili abbiano la stessa scala di riferimento, evitando che alcune caratteristiche abbiano un peso maggiore rispetto ad altre solo a causa delle loro unità di misura.

La standardizzazione dei dati prevede la trasformazione delle variabili numeriche in modo che abbiano media zero e deviazione standard pari a uno. Questo metodo assicura che tutte le variabili contribuiscano equamente all'analisi e che i modelli statistici non siano influenzati da scale di grandezza diverse.

Questa scelta metodologica è particolarmente importante nelle analisi di regressione, dove variabili con scale molto differenti potrebbero distorcere i risultati. Normalizzando e standardizzando i dati, si garantisce che ogni variabile venga valutata in modo equo, migliorando la stabilità e l'accuratezza dei modelli di previsione del valore di mercato dei giocatori.

Dopo la pulizia e la normalizzazione, il dataset finale è pronto per l'analisi. Ogni giocatore è rappresentato da un insieme di variabili che descrivono in dettaglio le sue caratteristiche e performance. Questo dataset strutturato consente di effettuare analisi statistiche approfondite, individuare correlazioni tra variabili e sviluppare modelli predittivi per stimare il valore di mercato dei giocatori.

La preparazione dei dati è quindi una fase fondamentale del processo, poiché garantisce la qualità delle informazioni e assicura che l'analisi venga condotta su un dataset accurato, coerente e privo di distorsioni. Questo approccio consente di ottenere risultati più affidabili e di migliorare la comprensione dei fattori che influenzano le prestazioni e il valore di mercato dei difensori di Serie A.

2.2.2 Selezione dei KPI

Una volta che i dati sono stati caricati, puliti e normalizzati, il passaggio successivo consiste nella selezione delle variabili indipendenti (KPI) che saranno utilizzate nei modelli di regressione per stimare il valore di mercato dei difensori. Questa fase è cruciale per garantire che il modello si concentri esclusivamente sulle metriche più significative, evitando ridondanze e riducendo il rischio di overfitting.

La selezione dei KPI avviene attraverso due approcci principali:

1. Selezione basata su criteri statistici: l'utilizzo del p-value consente di identificare i KPI con maggiore impatto statistico. Il calcolo dello Z-score per ciascun coefficiente permette di identificare le variabili più influenti, migliorando l'interpretabilità del modello.
2. Selezione personalizzata: è possibile scegliere manualmente i KPI sulla base dell'esperienza e della letteratura scientifica, includendo variabili che, pur non risultando statisticamente dominanti, sono note per avere un impatto rilevante sulle valutazioni dei giocatori.

2.2.3 Costruzione dei modelli

Per l'analisi, sono stati sviluppati diversi modelli di regressione lineare multipla, con l'obiettivo di comprendere la relazione tra il valore di mercato e i KPI selezionati. L'implementazione dei modelli segue un processo strutturato:

1. Suddivisione del dataset: i dati vengono separati in tre gruppi:
 - Tutti i difensori
 - Difensori Centrali

- Terzini

2. Applicazione di filtri: i giocatori con meno di 1700 minuti giocati vengono esclusi per garantire che l'analisi sia basata su campioni significativi.
3. Adattamento del modello: si utilizza la regressione lineare per stimare il valore di mercato come combinazione lineare dei KPI selezionati. La funzione obiettivo minimizza la somma dei quadrati dei residui per trovare i coefficienti ottimali.
4. Valutazione del modello: il coefficiente di determinazione R^2 viene calcolato per misurare la percentuale di variabilità del valore di mercato spiegata dai KPI. Il confronto tra diversi modelli aiuta a identificare la combinazione ottimale di variabili.

2.2.4 Analisi degli Outlier

Una parte fondamentale dell'analisi è l'identificazione degli outlier, ovvero quei giocatori il cui valore di mercato risulta significativamente diverso da quanto previsto dal modello. Gli outlier possono emergere per diversi motivi, tra cui fattori esterni non catturati dai KPI, come la popolarità mediatica, il contratto con il club o il marketing personale del giocatore.

Per individuare gli outlier viene adottato il criterio dei residui standardizzati: si calcolano le differenze tra il valore di mercato osservato e quello stimato dal modello, se il valore assoluto del residuo standardizzato supera una soglia (tipicamente 3), il giocatore viene considerato un outlier.

L'analisi degli outlier è utile per comprendere eventuali limitazioni del modello, individuare giocatori il cui valore di mercato è influenzato da dinamiche esterne, migliorare il modello, includendo eventuali variabili aggiuntive che potrebbero spiegare le discrepanze.

L'approccio adottato per la selezione dei KPI, la costruzione dei modelli di regressione e l'analisi degli outlier ha permesso di sviluppare un framework solido per l'analisi del valore di mercato dei difensori di Serie A.

Capitolo 3

Analisi dei risultati

Dopo aver descritto la metodologia utilizzata e gli strumenti impiegati per la costruzione del modello di regressione, in questo capitolo vengono presentati e analizzati i risultati ottenuti. L'obiettivo principale è valutare l'efficacia dei modelli sviluppati nel predire il valore di mercato dei difensori di Serie A, confrontando diverse strategie di selezione dei KPI e interpretando le metriche statistiche chiave.

3.1 Modello Generale per Tutti i Difensori

Per ottenere una prima visione d'insieme sull'impatto dei KPI sul valore di mercato dei difensori, è stato costruito un modello generale che include l'intero dataset senza distinguere tra difensori centrali e terzini. Questo approccio permette di individuare le variabili che, mediamente, influenzano maggiormente la valutazione economica di un difensore, prima di passare a modelli più specifici per ciascun ruolo.

Il modello sviluppato rappresenta un punto di partenza, costruito includendo tutti i KPI disponibili al fine di ottenere una visione complessiva delle variabili che potrebbero influenzare il valore di mercato dei difensori. Tuttavia, essendo un primo approccio esplorativo, il risultato ottenuto non è ancora del tutto solido, poiché la presenza di numerose variabili potrebbe introdurre rumore nei dati e ridurre la precisione delle stime.

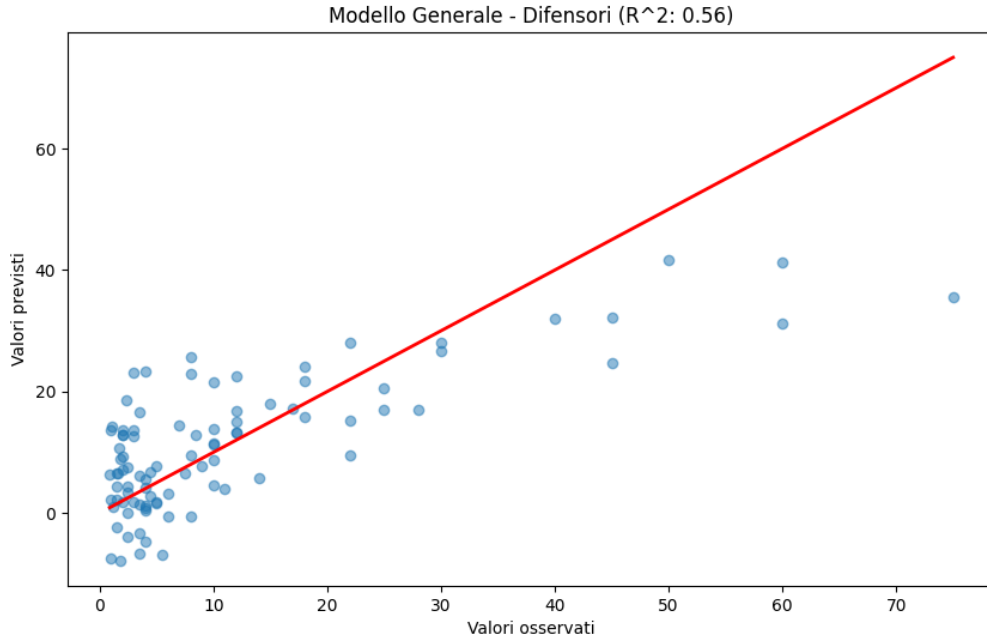


Figura 3.1: Prima analisi riguardante l'intero dataset dei difensori

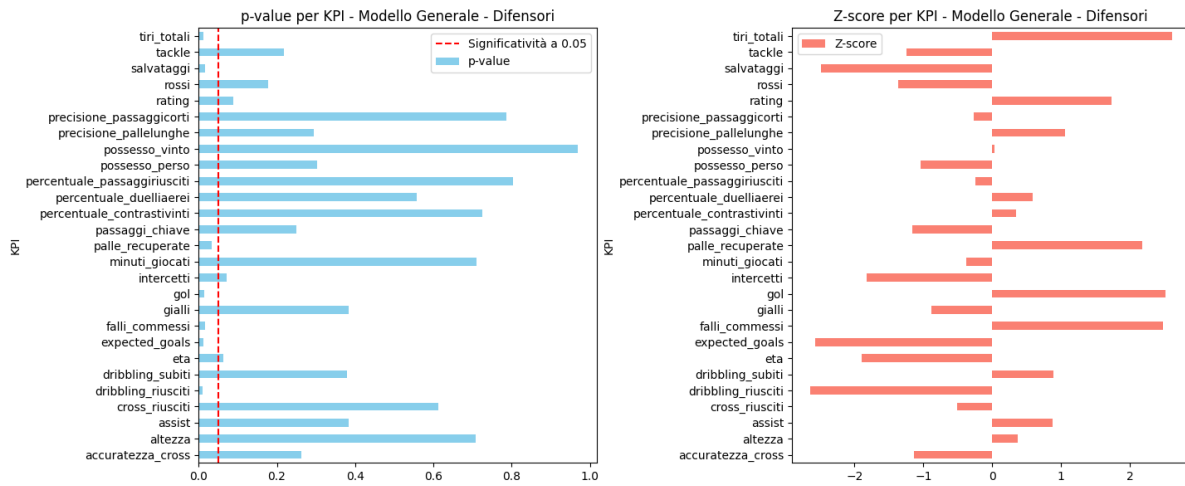


Figura 3.2: Valore dei p-value di tutti i KPI considerando tutti i difensori

Il modello iniziale sviluppato per l'analisi dei difensori ha prodotto un valore del coefficiente di determinazione R^2 pari a 0.56, indicando che il 56% della variabilità del valore di mercato dei giocatori può essere spiegata dai KPI selezionati. Sebbene questo risultato

possa suggerire una certa coerenza tra i valori previsti e quelli effettivi, un'analisi più approfondita ha evidenziato alcune criticità. In particolare, si osserva una dispersione significativa per i giocatori con un valore di mercato più elevato, segno che il modello fatica a catturare con precisione le dinamiche economiche di questi atleti.

Un ulteriore limite emerge dall'analisi dei p-value associati ai coefficienti di regressione. Molti KPI inclusi nel modello presentano un p-value superiore alla soglia di significatività statistica (0.05), suggerendo che il loro contributo alla spiegazione della variabilità del valore di mercato potrebbe essere dovuto più al caso che a un'effettiva relazione causale. Questo aspetto compromette l'affidabilità del modello, indicando che, pur fornendo un valore di R^2 relativamente accettabile, l'influenza di alcuni KPI potrebbe non essere realmente rilevante nella determinazione del valore economico dei difensori.

Nonostante queste limitazioni, il modello iniziale ha comunque rappresentato un primo passo importante per comprendere le dinamiche generali che legano le prestazioni dei difensori al loro valore di mercato. In particolare, ha permesso di individuare la necessità di affinare l'approccio, distinguendo tra le diverse tipologie di difensori. Per questo motivo, nelle analisi successive, il dataset e il modello di regressione sono stati suddivisi in due categorie distinte: difensori centrali e terzini. Questa scelta ha permesso di identificare KPI più specifici per ciascun ruolo, migliorando la capacità predittiva del modello e garantendo una maggiore coerenza nell'analisi dei risultati.

3.2 Modelli Specifici per Difensori Centrali e Terzini

Dopo aver analizzato il modello generale per tutti i difensori, è emersa la necessità di distinguere tra difensori centrali e terzini, due ruoli che, pur appartenendo alla stessa macro-categoria difensiva, presentano caratteristiche e responsabilità tattiche differenti. Questa suddivisione ha permesso di selezionare KPI più pertinenti per ciascun ruolo, con l'obiettivo di migliorare la capacità predittiva del modello e rendere più attendibile l'analisi della relazione tra prestazioni individuali e valore di mercato.

Entrambe le categorie sono state esaminate considerando tutti i KPI disponibili, con l'obiettivo di individuare quali metriche abbiano il maggiore impatto sul valore di mercato per ciascun ruolo. Questo approccio ha permesso di comprendere meglio le dinamiche economiche legate alle prestazioni dei giocatori e ha contribuito a migliorare la precisione del modello, affinando la selezione delle variabili più rilevanti per ogni tipologia di difensore.

3.2.1 Risultati modello Difensori centrali

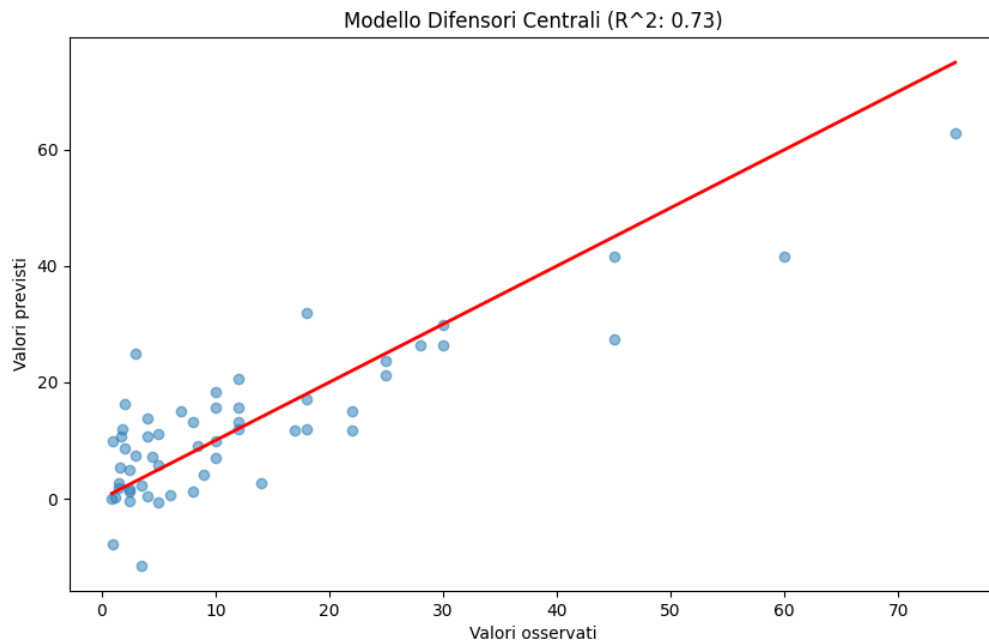


Figura 3.3: Analisi riguardante tutti i difensori centrali

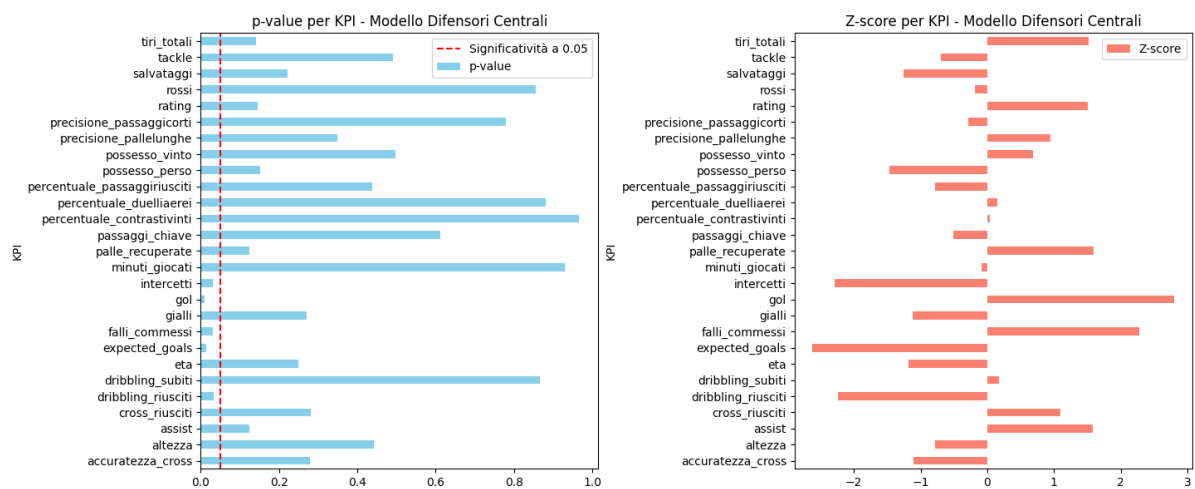


Figura 3.4: Valore dei p-value di tutti i KPI considerando i difensori centrali

L'analisi specifica sui difensori centrali ha prodotto un valore del coefficiente di determinazione R^2 pari a 0.73, un miglioramento significativo rispetto al modello generale, suggerendo che la segmentazione per ruolo abbia aumentato la capacità esplicativa del modello. Questo risultato indica che il 73% della variabilità del valore di mercato può essere spiegato dai KPI selezionati per i difensori centrali.

Tuttavia, nonostante il miglioramento dell' R^2 , l'analisi dei p-value ha evidenziato ancora alcune criticità. In particolare, alcuni KPI, pur mostrando una correlazione apparente con il valore di mercato, presentano un p-value superiore alla soglia di significatività, il che potrebbe indicare che la loro influenza non sia statisticamente rilevante. Ciò suggerisce che il modello, pur essendo più preciso rispetto alla versione generale, include alcune variabili non strettamente correlate al valore economico dei giocatori.

3.2.2 Risultati modello Terzini

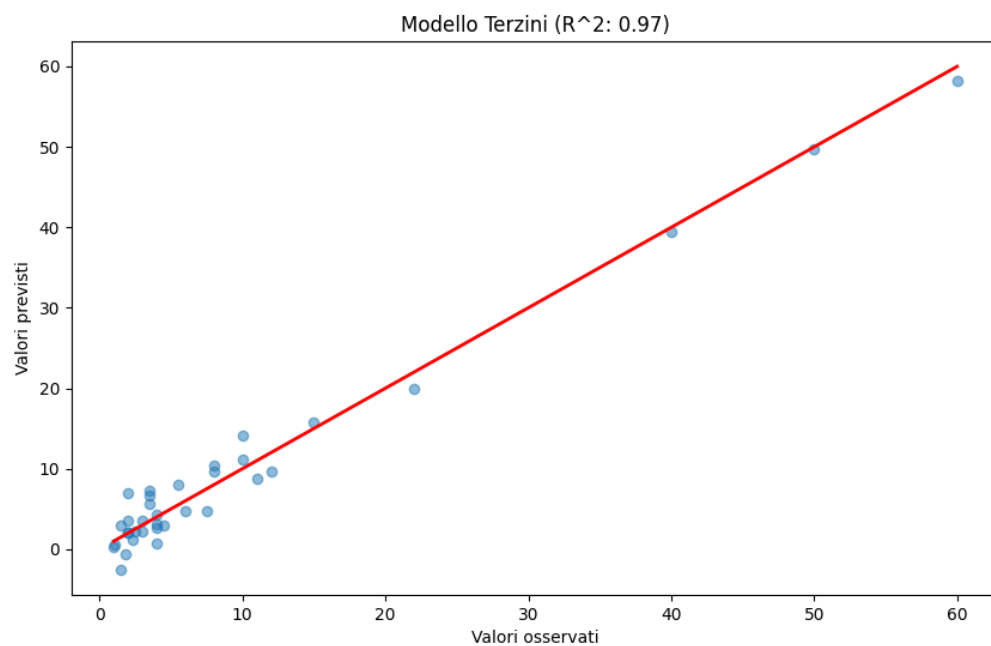


Figura 3.5: Analisi riguardante tutti i terzini

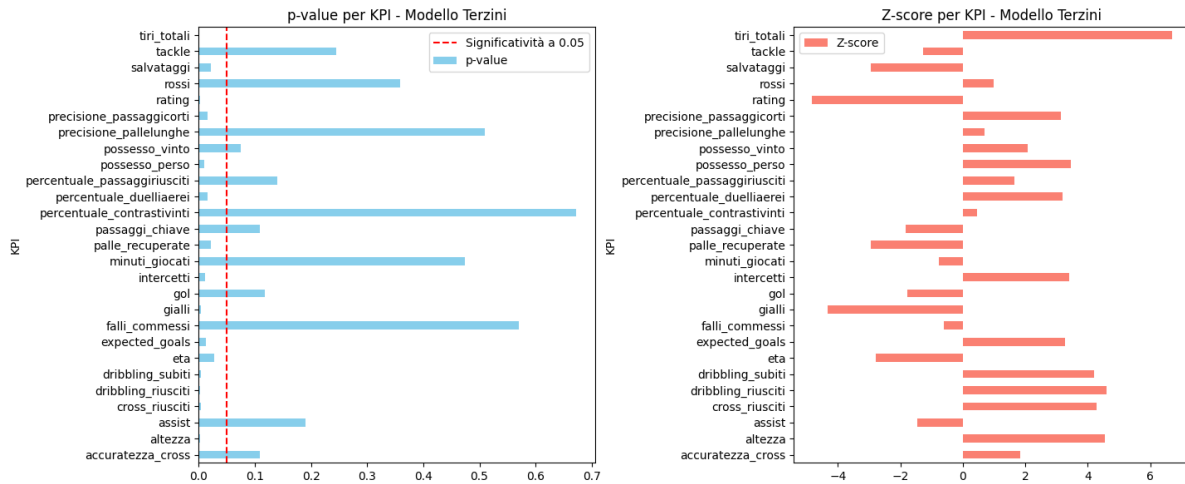


Figura 3.6: Valore dei p-value di tutti i KPI considerando i terzini

Il modello sviluppato per i terzini ha mostrato una capacità predittiva ancora superiore rispetto a quello dei difensori centrali, con un coefficiente di determinazione R^2 pari a 0.97. Questo risultato indica che quasi la totalità della variabilità del valore di mercato può essere spiegata dai KPI selezionati. Tuttavia, un R^2 così elevato deve essere interpretato con cautela: potrebbe derivare non solo dall'effettiva bontà del modello, ma anche dalla presenza di variabili collineari o da una sovradattabilità ai dati disponibili.

Nonostante l'elevato R^2 , l'analisi dei p-value mostra che non tutti i KPI selezionati risultano statisticamente significativi, suggerendo la possibilità di affinare ulteriormente il modello attraverso una selezione più rigorosa delle variabili. Inoltre, la differenza tra il modello dei difensori centrali e quello dei terzini evidenzia come il mercato valuti in maniera diversa i due ruoli, attribuendo un maggior peso ai KPI offensivi per i terzini rispetto ai difensori centrali.

Sebbene i risultati ottenuti non siano ancora ottimali, questa seconda analisi ha confermato l'importanza di distinguere i difensori centrali dai terzini per ottenere un modello predittivo più accurato. I valori di R^2 evidenziano un miglioramento nella capacità del modello di spiegare la relazione tra KPI e valore di mercato, suggerendo che il contesto tattico di ogni giocatore gioca un ruolo cruciale nella determinazione della sua valutazione economica. Il passo successivo consisterà quindi nel raffinare ulteriormente il modello, selezionando esclusivamente i KPI più significativi per ciascuna categoria, così da ottimizzare l'accuratezza della previsione.

3.3 Selezione basata su p-value

La terza analisi rappresenta un passo cruciale nel perfezionamento del modello, con l'obiettivo di affinare ulteriormente i risultati ottenuti nelle fasi precedenti. Le analisi iniziali hanno evidenziato come l'uso di tutti i KPI disponibili abbia portato a un'elevata capacità predittiva, ma con un numero significativo di variabili statisticamente insignificanti. Questo ha reso il modello meno interpretabile e più complesso, senza tuttavia migliorare in modo sostanziale la qualità delle previsioni.

Grazie alle osservazioni emerse nelle fasi precedenti, questa nuova analisi si concentra su una selezione più mirata delle variabili, eliminando quei KPI che non offrono un contributo significativo alla spiegazione del valore di mercato. L'obiettivo è costruire un modello più robusto e affidabile, riducendo il rumore statistico e migliorando la coerenza dei risultati, così da fornire una base più solida per la valutazione economica dei difensori.

3.3.1 Selezione Statistica

Per migliorare il modello precedente, abbiamo selezionato tutti quei KPI con un valore di p-value inferiore a una determinata soglia, includendo solo le variabili più significative nel determinare il valore di mercato. In questo modo, è stato possibile ottenere un risultato più affidabile, focalizzandosi sui KPI che mostrano una relazione più consistente con il valore economico dei difensori.

Risultati della Selezione Statistica dei Difensori Centrali

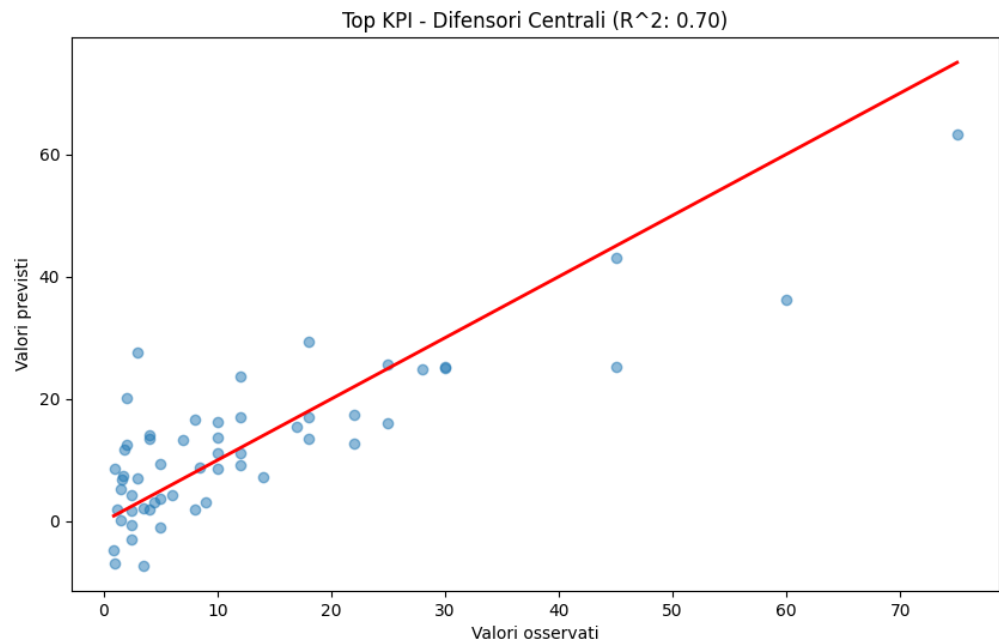


Figura 3.7: Valori osservati e previsti per i KPI selezionati - Difensori Centrali

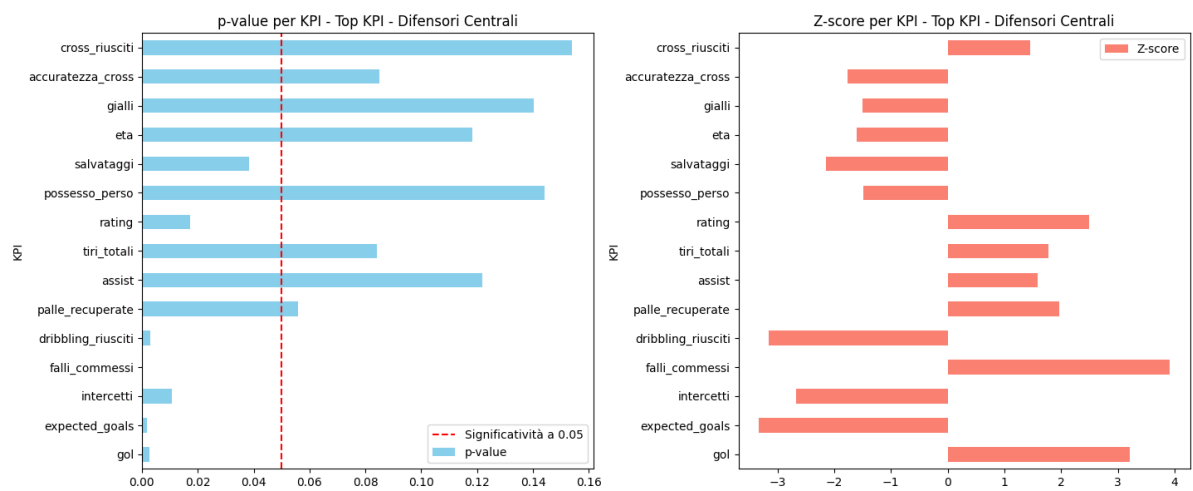


Figura 3.8: Distribuzione p-value e Z-score per i Top KPI selezionati - Difensori Centrali

L'applicazione della selezione statistica ha portato a una riduzione significativa del numero di KPI utilizzati nel modello per i difensori centrali, mantenendo solo quelli con un p-value inferiore alla soglia stabilita. Questo processo ha consentito di migliorare la qualità del modello, rimuovendo variabili che non apportavano un contributo rilevante alla spiegazione del valore di mercato.

L' R^2 del modello risultante si è ridotto rispetto alla versione precedente, a conferma del fatto che la selezione delle variabili porta a una minore capacità predittiva in termini assoluti, ma con un miglioramento nella significatività statistica delle variabili rimaste. Questo compromesso tra complessità del modello e robustezza statistica rappresenta un passo avanti verso un'analisi più affidabile e interpretabile.

Risultati della Selezione Statistica dei Terzini

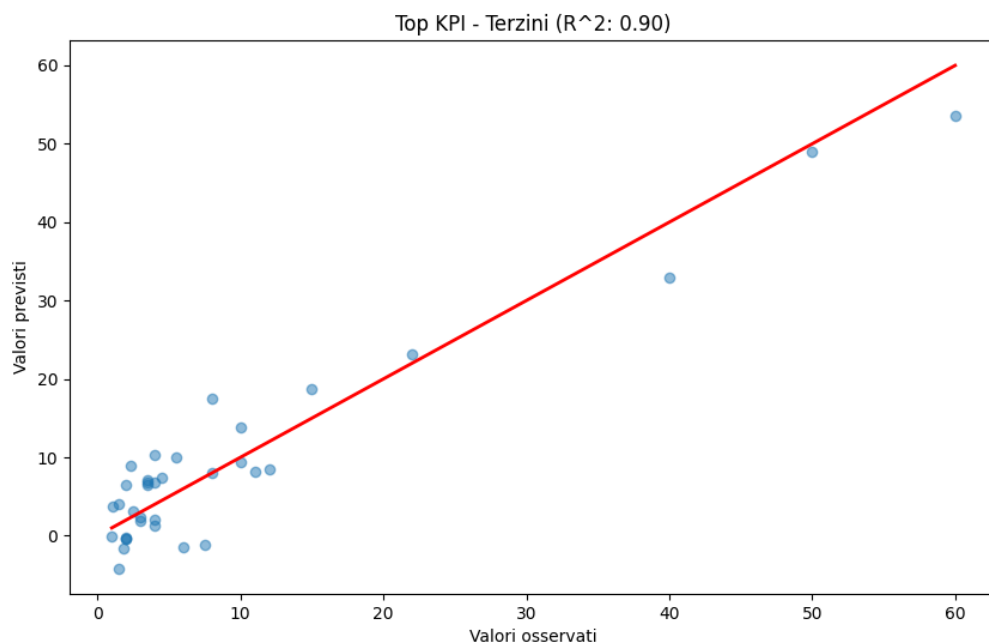


Figura 3.9: Valori osservati e previsti per i KPI selezionati - Terzini

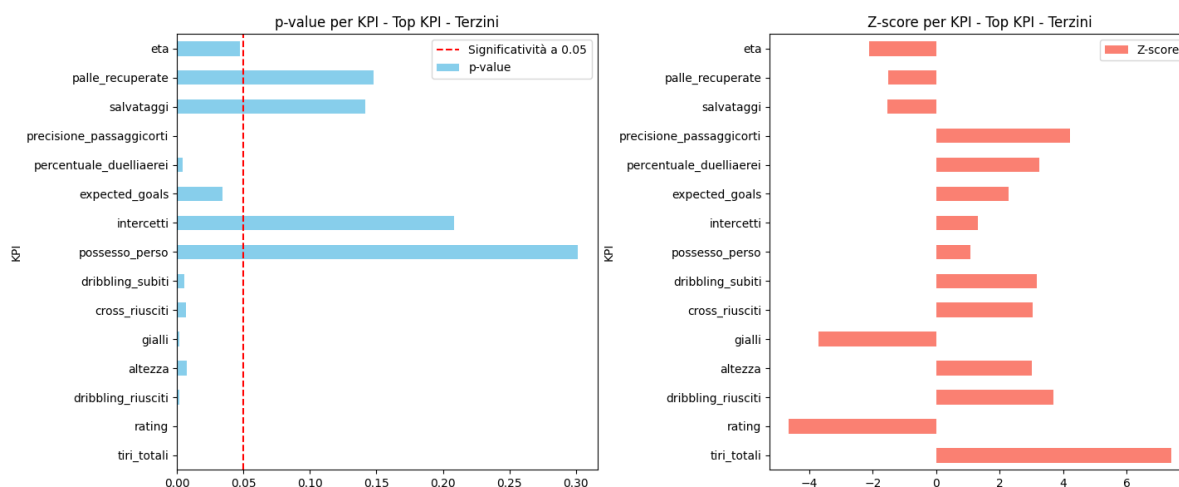


Figura 3.10: Distribuzione p-value e Z-score per i Top KPI selezionati - Terzini

Anche per i terzini, la selezione statistica ha permesso di individuare i KPI più rilevanti, rimuovendo quelli con un contributo trascurabile alla spiegazione del valore di mercato.

Un aspetto interessante è la conferma dell'importanza della capacità offensiva nel determinare il valore di mercato dei terzini.

Il modello per i terzini ha mostrato un R^2 superiore rispetto a quello dei difensori centrali, indicando che i KPI selezionati riescono a spiegare una quota maggiore della variabilità del valore di mercato. Questo risultato conferma quanto osservato nelle analisi precedenti, ovvero che il valore di mercato dei terzini è più facilmente modellizzabile rispetto a quello dei difensori centrali.

3.3.2 Selezione manuale

Dopo l'analisi appena conclusa, si è reso necessario un ulteriore affinamento del modello per ottenere risultati ancora più affidabili. Per migliorare la qualità della regressione, si è deciso di escludere ulteriormente quei KPI che presentavano un valore di p-value superiore alla soglia di significatività stabilita. Questo passaggio ha permesso di concentrare l'analisi solo sulle variabili con un impatto più rilevante sul valore di mercato, riducendo la complessità del modello e migliorandone la capacità predittiva.

Risultati della Selezione Manuale dei Difensori Centrali

Dopo l'analisi basata sulla selezione statistica, si è proceduto con un'ulteriore fase di affinamento del modello, adottando una selezione manuale dei KPI. In questa fase, sono stati mantenuti esclusivamente i KPI che nel modello precedente avevano mostrato un p-value inferiore a 0.05, eliminando quelli meno significativi. L'obiettivo era ottenere

un modello più interpretabile e robusto, focalizzandosi sulle variabili con un impatto effettivo sul valore di mercato.

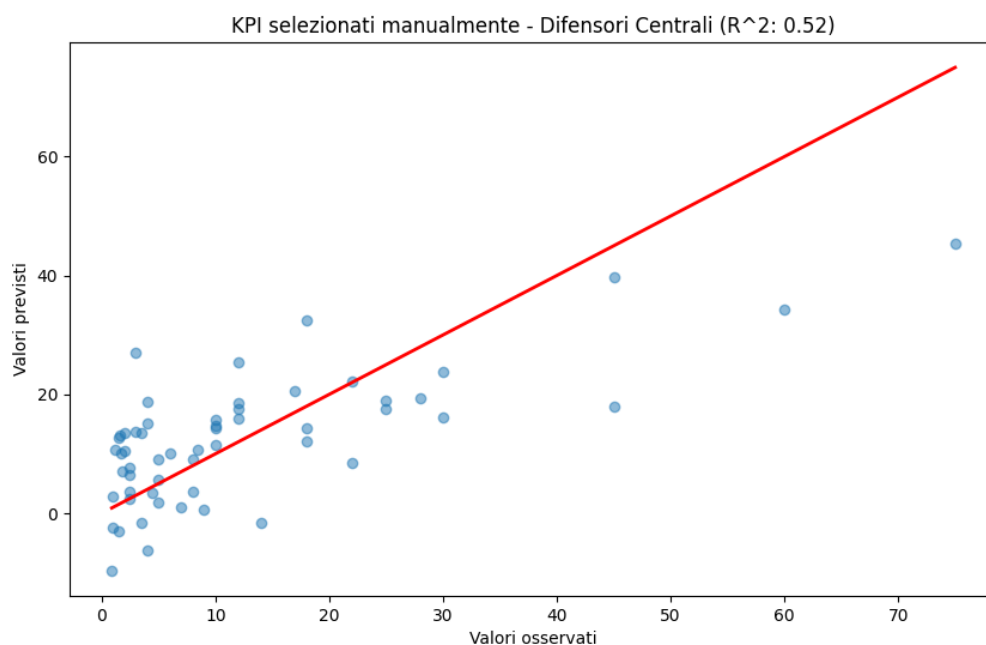


Figura 3.11: Valori osservati e previsti per i KPI selezionati - Difensori Centrali

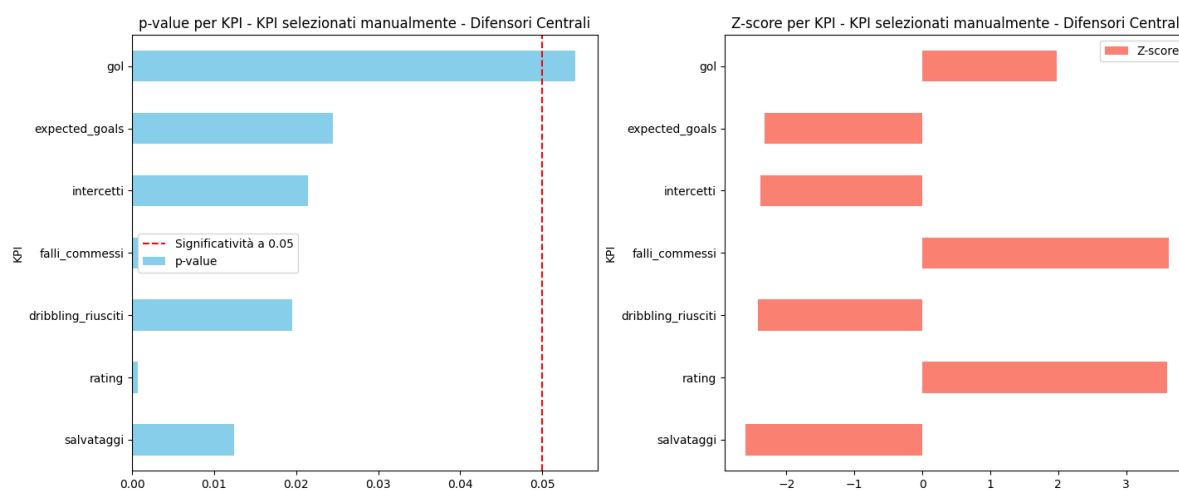


Figura 3.12: Distribuzione p-value e Z-score per i KPI selezionati manualmente - Difensori Centrali

Dopo la selezione manuale basata sui KPI con p-value inferiore a 0.05, il modello dei difensori centrali ha evidenziato una serie di variabili che risultano statisticamente significative nel determinare il valore di mercato. Tra queste, troviamo expected goals, gol, intercetti, salvataggi, rating e dribbling riusciti.

Tra le statistiche selezionate, expected goals e gol segnati suggeriscono che i difensori con una maggiore propensione offensiva possono avere una valutazione economica leggermente più alta, sebbene il loro peso nel modello sia limitato. Questo riflette il fatto che, pur essendo apprezzato un difensore capace di segnare su calci piazzati o inserimenti offensivi, il mercato non valuta queste caratteristiche come primarie rispetto alle capacità difensive.

Le statistiche difensive pure, come intercetti e salvataggi, giocano un ruolo importante ma con delle sfumature particolari. Gli intercetti indicano la capacità di leggere in anticipo le giocate avversarie, ma un numero elevato può anche suggerire che il difensore gioca in una squadra che subisce molte azioni offensive, piuttosto che essere necessariamente un segnale di qualità individuale. Lo stesso vale per i salvataggi, che sono spesso frutto di situazioni in cui la difesa è stata superata e richiede un intervento disperato: questo può essere un segnale positivo di reattività del giocatore, ma anche un sintomo di una struttura difensiva che concede troppe occasioni.

Il rating è spesso una combinazione di variabili già presenti nel modello, e il fatto che abbia uno Z-score negativo suggerisce che, a parità di altre caratteristiche, i giocatori con rating più alto non hanno necessariamente un valore di mercato maggiore. Questo potrebbe accadere se, ad esempio, la parte difensiva del rating ha più peso rispetto a quella offensiva, mentre il valore di mercato sembra essere più sensibile a KPI offensivi come tiri, expected goals, dribbling riusciti e cross.

Infine, il numero di dribbling riusciti da parte di un difensore centrale ha mostrato un effetto negativo, suggerendo che un'eccessiva propensione a portare palla e tentare l'uno contro uno non sia una caratteristica apprezzata in un centrale. Questo è coerente con l'idea che un difensore centrale debba essere sicuro e pragmatico nelle scelte, evitando rischi eccessivi in zone di campo delicate.

In sintesi, i KPI selezionati mostrano che i difensori centrali sono valutati in modo diverso rispetto ad altri ruoli: le statistiche offensive sono apprezzate, ma non determinanti, mentre le metriche difensive devono essere interpretate con attenzione, poiché il loro significato può variare a seconda del contesto di squadra. Questo spiega le difficoltà nel costruire un modello predittivo altamente preciso per i difensori centrali, in quanto molti aspetti fondamentali del loro gioco non sono facilmente catturabili attraverso le statistiche tradizionali.

Analisi degli Outliers per i Difensori Centrali

Per individuare la presenza di eventuali outliers, è stata condotta un'analisi basata sullo Z-score, adottando una soglia pari a 3 come criterio di identificazione. I risultati mostrano che nessun giocatore supera questo valore limite, confermando la coerenza tra i dati osservati e le previsioni del modello. Questo suggerisce che il dataset non presenta anomalie significative che potrebbero influenzare negativamente l'analisi, garantendo così una maggiore affidabilità nella valutazione della relazione tra KPI e valore di mercato.

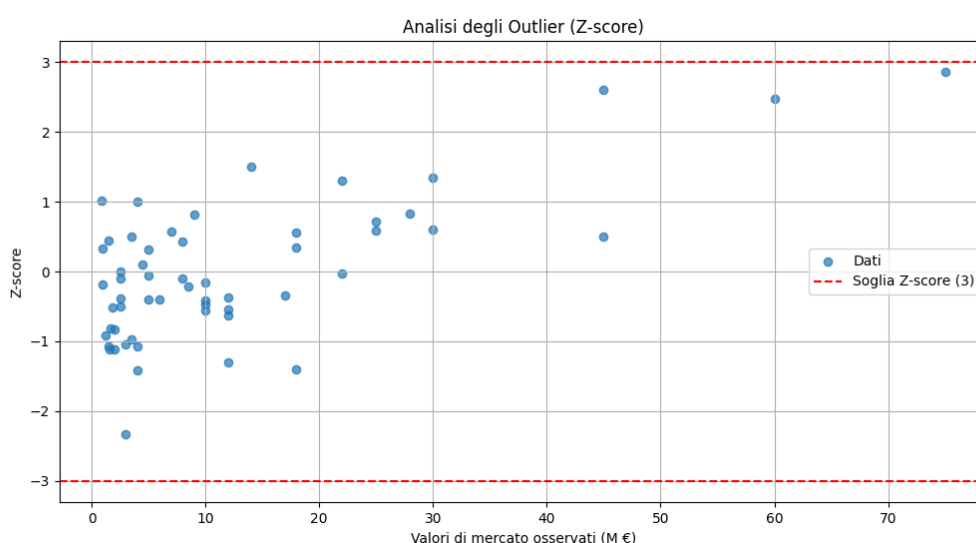


Figura 3.13: Analisi degli Outliers per i Difensori Centrali

L'assenza di outliers nel modello dei difensori centrali indica che i dati analizzati non presentano valori estremi che potrebbero distorcere i risultati della regressione. Questo aspetto contribuisce a garantire una maggiore affidabilità del modello, in quanto le stime ottenute non sono influenzate da osservazioni anomale che potrebbero falsare l'interpretazione dei coefficienti. Tuttavia, nonostante questa stabilità nei dati, il modello presenta comunque alcune limitazioni in termini di capacità predittiva, come evidenziato dal valore relativamente basso di R^2 .

Risultati della Selezione Manuale dei Terzini

Per quanto riguarda i terzini, la selezione manuale è stata effettuata con lo stesso criterio, mantenendo solo i KPI che avevano un p-value inferiore a 0.05 nel modello precedente.

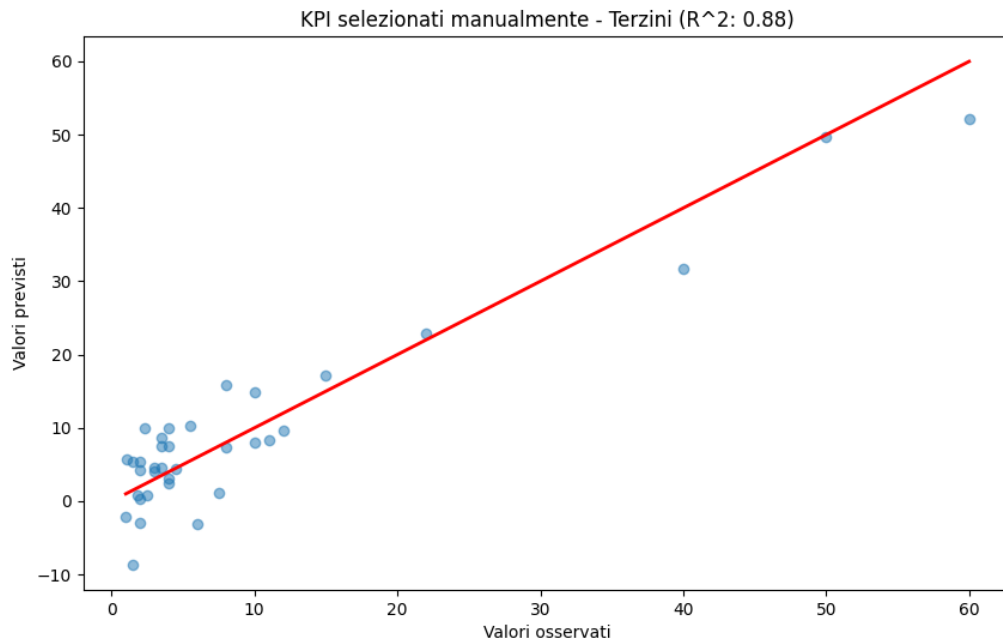


Figura 3.14: Valori osservati e previsti per i KPI selezionati - Terzini

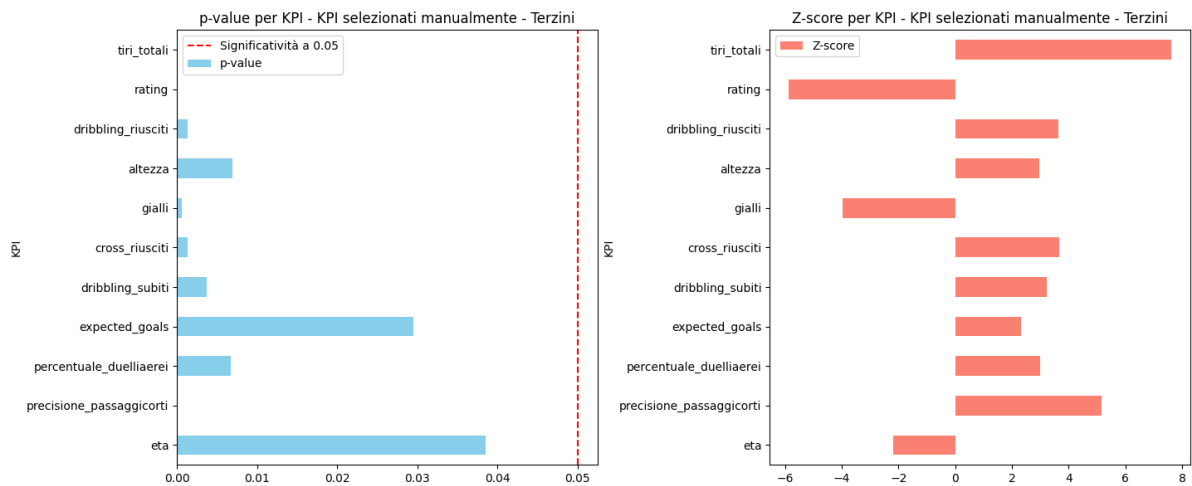


Figura 3.15: Distribuzione p-value e Z-score per i KPI selezionati manualmente - Terzini

In questo caso, il nuovo modello ha prodotto un R^2 di 0.88, confermando un'ottima capacità predittiva e una solida relazione tra KPI e valore di mercato.

L'analisi dei KPI dei terzini evidenzia un profilo statistico molto più variegato rispetto a quello dei difensori centrali, riflettendo il doppio ruolo che questi giocatori ricoprono in campo. I terzini, infatti, non solo devono garantire solidità difensiva, ma sono spesso chiamati a supportare attivamente la fase offensiva della squadra, rendendo la loro valutazione più legata a un mix di metriche offensive e difensive.

Uno degli aspetti più significativi emersi dall'analisi è l'impatto delle statistiche offensive sul valore di mercato di questi giocatori. KPI come tiri totali, expected goals e cross riusciti suggeriscono che i terzini più coinvolti nella costruzione del gioco e nella finalizzazione tendono ad avere una valutazione più alta. Questo riflette l'evoluzione moderna del ruolo, dove i giocatori che riescono a contribuire attivamente alla manovra d'attacco sono particolarmente apprezzati. I terzini con una buona capacità di cross e un'efficace partecipazione alle azioni offensive hanno più possibilità di essere considerati giocatori di alto valore.

Parallelamente, anche alcune statistiche difensive hanno un impatto rilevante nel modello. KPI come intercetti, dribbling subiti e percentuale di duelli aerei vinti forniscono un'indicazione della solidità difensiva del giocatore. Un numero elevato di intercetti suggerisce una buona capacità di lettura del gioco e di posizionamento, mentre una bassa percentuale di dribbling subiti indica un'efficace capacità di contrasto nei duelli individuali. Anche la percentuale di duelli aerei vinti ha mostrato un effetto positivo, suggerendo che i terzini con una maggiore prestanza fisica possono avere un vantaggio in termini di valutazione economica.

Un altro aspetto interessante è la presenza di KPI legati alla disciplina e all'esperienza, come il numero di cartellini gialli e l'età del giocatore. Un numero elevato di cartellini può essere un indicatore di un gioco troppo aggressivo o falloso, che potrebbe penalizzare il valore di mercato del giocatore. L'età, invece, ha un impatto negativo sul valore economico, con i giocatori più giovani che risultano generalmente più appetibili sul mercato rispetto a quelli più esperti.

Infine, il rating complessivo emerge come una delle variabili più influenti nel modello, suggerendo che una valutazione sintetica delle prestazioni, che incorpora diversi aspetti del gioco, possa essere un indicatore efficace per stimare il valore di mercato di un terzino.

Nel complesso, i KPI selezionati mostrano che i terzini vengono valutati con un approccio più completo rispetto ai difensori centrali. L'equilibrio tra contributo offensivo e affidabilità difensiva gioca un ruolo chiave nella determinazione del loro valore di mercato, e i giocatori in grado di eccellere in entrambi gli aspetti sono quelli che tendono ad avere le valutazioni più alte. Questo spiega anche perché il modello predittivo per i terzini ha ottenuto risultati significativamente migliori rispetto a quello dei difensori centrali, dimostrando come le statistiche disponibili siano più efficaci nel catturare il loro impatto sul campo.

In generale, l'uso del filtro per minutaggio si è rivelato cruciale per garantire una maggiore qualità dei dati e per ridurre il rischio di includere giocatori con statistiche non

rappresentative. Inoltre, la selezione dei KPI più significativi ha permesso di ottenere modelli più interpretabili e meno soggetti a overfitting.

Analisi degli Outliers per i Terzini

Per l'analisi degli outliers relativa ai terzini, è stato applicato lo stesso metodo basato sullo Z-score, utilizzando una soglia pari a 3 per identificare eventuali valori anomali. Anche in questo caso, nessun giocatore ha superato il limite stabilito, confermando la coerenza tra i dati osservati e le previsioni del modello.

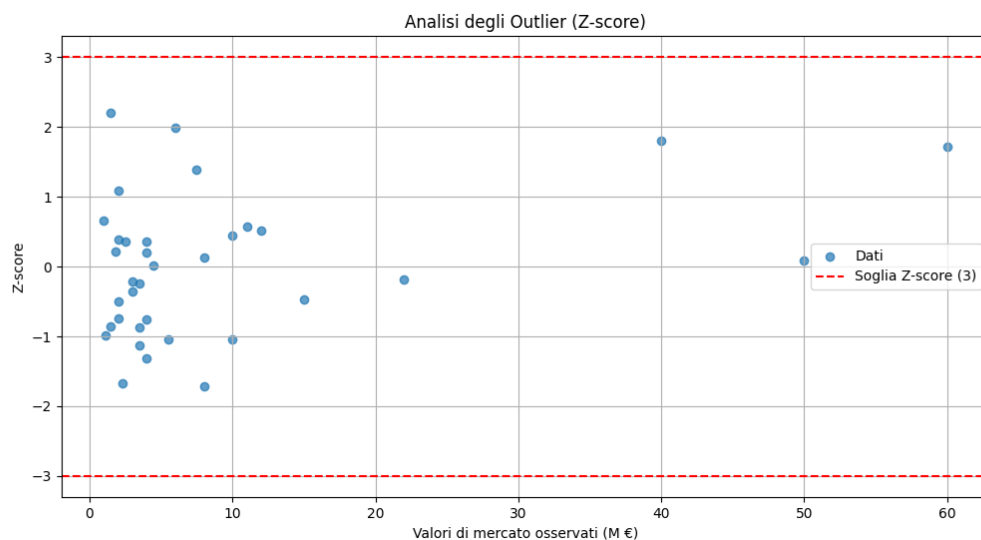


Figura 3.16: Analisi degli Outliers per i Terzini

Questo risultato evidenzia la solidità del modello, dimostrando che non è influenzato da anomalie nei dati. L'assenza di outliers conferma la stabilità della regressione, suggerendo che le previsioni ottenute sono coerenti con il comportamento generale dei dati e, di conseguenza, più affidabili nell'interpretazione delle relazioni tra KPI e valore di mercato.

Conclusioni

Lo studio condotto ha permesso di analizzare la relazione tra KPI e valore di mercato dei difensori di Serie A, sviluppando un modello di regressione lineare multipla per identificare i fattori più influenti nella valutazione economica dei giocatori. L'analisi ha evidenziato come alcune metriche, in particolare quelle relative alla fase difensiva e alla costruzione del gioco, abbiano un impatto significativo sul valore di mercato, sebbene con variazioni importanti tra difensori centrali e terzini.

L'analisi ha evidenziato una differenza significativa tra il modello sviluppato per i difensori centrali e quello per i terzini, con il primo che mostra una capacità predittiva inferiore rispetto al secondo. Questo risultato si discosta da quanto emerso negli studi precedenti su attaccanti e centrocampisti in termini di R^2 , che per i difensori centrali risulta notevolmente più basso.

Questa discrepanza può essere spiegata dal fatto che il valore di mercato dei difensori centrali sembra dipendere da fattori meno direttamente misurabili attraverso i KPI prestazionali tradizionali. A differenza di attaccanti, centrocampisti o anche terzini, i cui contributi si riflettono in azioni facilmente quantificabili come gol, assist o dribbling, il ruolo del difensore centrale è spesso legato a elementi meno tangibili come la capacità di guidare la linea difensiva, l'organizzazione tattica, il posizionamento e l'interpretazione del gioco. Questi aspetti non sempre emergono chiaramente dalle statistiche standard, il che potrebbe spiegare la minore capacità del modello di prevedere il valore economico di questi giocatori.

Al contrario, per i terzini il modello ha mostrato un R^2 decisamente più elevato. Questo può essere attribuito alla loro partecipazione più attiva alla fase offensiva, che li porta a incidere su KPI più direttamente correlabili al valore di mercato, come i cross riusciti, i dribbling e i tiri totali. La loro influenza sulla manovra di squadra è più misurabile rispetto a quella dei difensori centrali, rendendo il modello predittivo più efficace.

Questa differenza nei risultati suggerisce che il valore di mercato dei difensori centrali potrebbe essere influenzato in modo più marcato da fattori non inclusi nel modello, come esperienza, leadership, continuità ad alti livelli e reputazione, piuttosto che esclusivamente dalle prestazioni quantificabili attraverso i KPI. Nonostante questa limitazione, il modello sviluppato fornisce comunque indicazioni utili e rappresenta un primo pas-

so verso una comprensione più approfondita dei meccanismi che determinano il valore economico dei difensori.

L'approccio adottato ha fornito risultati interessanti, ma è necessario evidenziare alcune limitazioni che possono influenzare la portata delle conclusioni tratte.

Limitazioni dello studio

Sebbene lo studio offra un'analisi dettagliata della relazione tra KPI e valore di mercato, presenta alcune limitazioni che è opportuno considerare. Queste riguardano la metodologia statistica adottata, il campione analizzato e l'orizzonte temporale dello studio. Approfondiamo ciascun aspetto di seguito.

Scelta della metodologia statistica

In questo lavoro è stata utilizzata la regressione lineare multipla per modellare la relazione tra KPI e valore di mercato. Questo approccio presenta diversi vantaggi, tra cui la semplicità di implementazione, la trasparenza nell'interpretazione dei risultati e la possibilità di determinare il peso specifico di ogni KPI nella stima del valore economico di un giocatore. Tuttavia, questa metodologia ha anche delle limitazioni, poiché si basa sull'ipotesi che la relazione tra le variabili indipendenti (i KPI) e la variabile dipendente (il valore di mercato) sia strettamente lineare.

Nel contesto calcistico, però, questa ipotesi potrebbe non essere del tutto realistica. Il valore di mercato di un giocatore è influenzato da una combinazione di fattori complessi e interdipendenti, che non sempre seguono una relazione lineare. Si pensi, ad esempio, a un difensore con un numero elevato di intercetti: intuitivamente, si potrebbe concludere che il suo valore di mercato sia più alto rispetto a un collega con meno intercetti. Tuttavia, questa relazione non è sempre così diretta. Il numero di intercetti dipende anche dal sistema di gioco della squadra e dal tipo di pressione esercitata sugli avversari. Un difensore che milita in una squadra con una fase difensiva aggressiva potrebbe registrare un numero molto alto di intercetti non necessariamente perché ha una qualità superiore, ma perché il suo contesto di gioco gli offre più opportunità di intervenire.

Un'altra variabile che non segue una relazione lineare con il valore di mercato è l'età del giocatore. Un giovane talento può essere valutato molto più di un calciatore esperto, non solo per le sue capacità attuali, ma anche per il suo potenziale di crescita e la prospettiva di una lunga carriera. Tuttavia, questa tendenza non è sempre costante: un giocatore con molta esperienza e una reputazione consolidata potrebbe mantenere un alto valore di mercato nonostante l'età avanzata, mentre un giovane senza un rendimento stabile potrebbe vedere la sua valutazione oscillare in modo imprevedibile. Questi elementi introducono dinamiche che una semplice regressione lineare potrebbe non cogliere con precisione.

Per superare queste limitazioni, modelli più sofisticati come le reti neurali, le regressioni non lineari o gli algoritmi di machine learning (ad esempio, Random Forest o Gradient Boosting) potrebbero essere più adatti. Questi modelli permettono di catturare interazioni più complesse tra le variabili e di identificare schemi nascosti che la regressione lineare non è in grado di rilevare.

Campione analizzato

Un altro aspetto da considerare riguarda la composizione del dataset. I dati utilizzati in questo studio riguardano esclusivamente difensori della Serie A. Questo limita la generalizzabilità delle conclusioni, poiché il valore di mercato può variare in base a fattori specifici del campionato di riferimento, come il livello di competitività, lo stile di gioco delle squadre e le strategie di mercato adottate dai club. Un'analisi più ampia, che includa giocatori provenienti da diversi campionati europei, potrebbe fornire risultati più robusti e applicabili a un contesto calcistico più ampio.

Orizzonte temporale dello studio

Il dataset si riferisce esclusivamente alle prestazioni della stagione 2023/24, il che rappresenta un'ulteriore limitazione. Le prestazioni di un giocatore possono variare significativamente nel corso degli anni, e un solo campionato potrebbe non essere sufficiente per individuare tendenze stabili nel tempo. Fattori come infortuni, cambi di allenatore o modifiche tattiche all'interno della squadra possono influenzare il rendimento di un difensore in una singola stagione. Pertanto, un'estensione dell'analisi a più stagioni potrebbe migliorare l'affidabilità dei risultati e permettere di identificare pattern più consistenti.

Sviluppi futuri

Alla luce di queste considerazioni, futuri studi potrebbero migliorare l'analisi sotto diversi aspetti. In primo luogo, l'integrazione di tecniche di machine learning potrebbe aumentare la capacità predittiva del modello, permettendo di individuare relazioni più complesse tra KPI e valore di mercato. Inoltre, un ampliamento del dataset, includendo più stagioni e giocatori di campionati diversi, consentirebbe di ottenere risultati più generali e applicabili a contesti più ampi.

Un ulteriore sviluppo potrebbe riguardare l'inclusione di fattori extra-campo, come la popolarità mediatica dei giocatori, l'impatto del loro brand personale e il ruolo delle agenzie di procuratori nel determinare il valore di mercato. Questi elementi, difficilmente quantificabili con le metriche tradizionali, potrebbero rappresentare una variabile chiave per spiegare alcune discrepanze tra valore stimato e valore effettivo di mercato.

In conclusione, questo studio rappresenta un primo passo verso una comprensione più oggettiva del valore economico dei difensori, basata su dati e metriche quantificabili. Sebbene presenti alcune limitazioni, il modello sviluppato fornisce una base solida per analisi future, contribuendo al dibattito sull'utilizzo dei dati nel calcio moderno e offrendo spunti per migliorare le strategie di valutazione dei giocatori da parte di club e analisti sportivi.

Bibliografia

- [1] Adriano Gilardone, “Il significato del p-value”, disponibile su <https://adrianogilardone.com/p-value/>.
- [2] Ala, Christopher E.; Turner, Anthony N.; Vescovo, Chris J., *Importanza della forza e della potenza sugli indicatori chiave delle prestazioni nel calcio giovanile d'élite*, Journal of Strength and Conditioning Research, 34(7): 2006-2014, Luglio 2020. DOI: <https://doi.org/10.1519/JSC.0000000000002446>
- [3] Auxiell. (2024). KPI: come scegliere quelli giusti. Disponibile su: <https://www.auxiell.com/kpi-come-scegliere-quelli-giusti/>
- [4] Calcio in Pillole. (2024). Viaggio nell'universo Transfermarkt: il fenomeno delle valutazioni dei calciatori. Disponibile su: <https://www.calcioinpillole.com/viaggio-nelluniverso-transfermarkt-il-fenomeno-delle-valutazioni-dei-calciatori/>.
- [5] *Dalla Moneyball alla Football Analytics: la statistica scende in campo*. Ghigliottina, 12 Luglio 2022. Disponibile su: <https://www.ghigliottina.info/2022/07/12/dalla-moneyball-alla-football-analytics-la-statistica-scende-in-campo/>
- [6] Diario di un Analista, “Come identificare anomalie nei tuoi dati”, disponibile su <https://www.diariodiunanalista.it/posts/come-identificare-anomalie-nei-tuoi-dati/>.
- [7] Gobbi, N. (2024). *Valore di mercato e performance sportiva dei calciatori*. Tesi di Laurea, Università di Padova. Disponibile su: https://thesis.unipd.it/bitstream/20.500.12608/22117/1/Gobbi_Nicolo.pdf.
- [8] Hughes, M. D., et al. (2012). *Moneyball and soccer: An analysis of the key performance indicators of elite male soccer players by position*. In: *Notational Analysis of Sport*. Disponibile su: https://www.researchgate.net/publication/268186454_Moneyball_and_soccer__An_analysis_of_the_key_performance_indicators_of_elite_male_soccer_players_by_position
- [9] Investopedia, “Z-Score Definition”, disponibile su <https://www.investopedia.com/terms/z/zscore.asp>.

- [10] Kumawat, N. (2024). *Ricerca di parametri predittivi per la stima del valore economico degli attaccanti nel gioco del calcio: uno studio basato sulla regressione lineare*. Tesi di Laurea in Informatica per il Management, Sessione 2, Anno Accademico 2023/2024, Alma Mater Studiorum - Università di Bologna.
- [11] Milanato, D. (2008). *Regressione lineare*. In: *Demand Planning*. UNITEXT(). Springer, Milano. DOI: https://doi.org/10.1007/978-88-470-0822-9_15
- [12] Nacsport, “Key Performance Indicators in Soccer”, disponibile su <https://www.nacsport.com/blog/en-us/Tips/key-performance-indicators-soccer>.
- [13] OFG - Open for Growth. (2024). KPI: cosa sono? Significato, esempi e come calcolarli. Disponibile su: <https://blog.ofg.it/kpi-cosa-sono-significato-esempi-e-come-calcolarli-ofg>
- [14] Pascarella, L. (2024). *KPI e valore di mercato: Statistica inferenziale e modelli predittivi nel calcio professionistico applicati ai centrocampisti della Serie A (stagione 2023/24)*. Tesi di Laurea in Informatica per il Management, Sessione 2, Anno Accademico 2023/2024, Alma Mater Studiorum - Università di Bologna.
- [15] Plakias, S., Tsatalas, T., Armatas, V., Tsaopoulos, D., & Giakas, G. (2024). Tactical Situations and Playing Styles as Key Performance Indicators in Soccer. *Journal of Functional Morphology and Kinesiology*, 9(2), 88. DOI: <https://doi.org/10.3390/jfmk9020088>.
- [16] Rivista Undici. (2016). *Teoria del difensore*. Disponibile su: <https://www.rivistaundici.com/2016/12/20/teoria-difensore/>
- [17] Roccetti, M., Berveglieri, F., & Cappiello, G. (2024). *Football Data Analysis: The Predictive Power of Expected Goals*. In *Proceedings of the 25th Annual European GAME’ON Conference on Simulation, AI and in Computer Games*.
- [18] Roccetti, M., Tenace, M., Cappiello, G. (2025). *Prescient Perspectives on Football Tactics: A Case with Liverpool FC, Corners and AI*. In: Aiello, L.M., Chakraborty, T., Gaito, S. (eds) *Social Networks Analysis and Mining. ASONAM 2024*. Lecture Notes in Computer Science, vol 15214. Springer, Cham. Disponibile su: https://doi.org/10.1007/978-3-031-78554-2_15
- [19] SofaScore, “Statistiche Calcistiche”, disponibile su <https://www.sofascore.com/>.
- [20] Thiese, M. S., Ronna, B., & Ott, U. (2016). P value interpretations and considerations. *Journal of Thoracic Disease*, 8(9), E928-E931. DOI: <https://doi.org/10.21037/jtd.2016.08.16>. PMID: 27747028; PMCID: PMC5059270.

[21] Transfermarkt, “Valutazioni di Mercato”, disponibile su <https://www.transfermarkt.it/>.