



Università degli Studi di Salerno



Dipartimento di Informatica

Fondamenti di Data Science e Machine Learning 2023/2024

Documentazione

EA Sports FC 24

Realizzato da: Arcangeli Giovanni, Di Maio Marco

Sommario

Introduzione.....	4
L'importanza delle valutazioni dei giocatori	4
Motivazione del Progetto	4
Obiettivi del Progetto	5
Struttura della Relazione	5
Background.....	6
La Serie FIFA	6
Evoluzione e Innovazioni	6
Importanza delle Valutazioni dei Giocatori	6
Machine Learning e Previsione delle Valutazioni	6
Significato e Impatto del Progetto	7
Stato dell'Arte	7
Introduzione all'Intelligenza Artificiale e al Machine Learning	7
Applicazioni del Machine Learning nello Sport.....	7
Previsione delle Performance degli Atleti	7
La Serie FIFA e l'Analisi dei Giocatori	7
Metodologie di Valutazione Tradizionali.....	8
Introduzione del Machine Learning nelle Valutazioni.....	8
Modelli di Machine Learning Utilizzati	8
Regressione Lineare	8
Random Forest.....	8
Gradient Boosting.....	8
XGBoost.....	8
Voting Regressor.....	8
Applicazioni Pratiche e Sfide	9
Applicazioni Pratiche	9
Sfide.....	9
Conclusione	9
Metodologie.....	9
Introduzione.....	9
Raccolta e Preprocessing dei Dati	9
Raccolta dei Dati	9
Pulizia dei Dati	10
Gestione delle Colonne Categoriali.....	10
Imputazione e Standardizzazione dei Dati	10

Creazione di Nuove Feature	10
Riduzione della Dimensionalità con PCA	10
Addestramento dei Modelli	10
Divisione del Dataset.....	10
Modelli Utilizzati.....	11
Valutazione dei Modelli.....	11
Metriche di Valutazione.....	11
Confronto dei Risultati	12
Produzione di Previsioni.....	12
Implementazione delle Previsioni	12
Conclusione	12
Risultati	12
Introduzione.....	12
Valutazione delle Performance dei Modelli.....	12
Random Forest Regressor.....	12
Gradient Boosting Regressor.....	13
Voting Regressor.....	13
Randomized Search CV	14
Confronto tra i Modelli.....	14
Interpretazione delle Previsioni.....	14
Valutazioni Future	14
Accuratezza delle Previsioni	15
Conclusione	15
Conclusioni e Lavori Futuri.....	15
Conclusioni.....	15
Lavori Futuri.....	16
1. Integrazione di Nuove Variabili.....	16
2. Utilizzo di Modelli di Deep Learning.....	16
3. Analisi dei Trend Temporali	16
4. Ottimizzazione dei Modelli Ensemble	16
5. Interpretabilità dei Modelli.....	16
6. Valutazione Continua e Aggiornamento dei Modelli.....	16
7. Espansione ad Altri Ambiti Sportivi.....	16
Conclusione Finale	16

Introduzione

Il mondo dei videogiochi è in continua evoluzione, e tra i titoli più popolari e influenti nel panorama dei giochi sportivi vi è senza dubbio la serie FIFA. Questo gioco, sviluppato e pubblicato da Electronic Arts sotto l'etichetta EA Sports, rappresenta una simulazione realistica del calcio, combinando gameplay avvincente e dettagli tecnici accurati. La serie FIFA è celebre per il suo realismo, dovuto in gran parte alle valutazioni dettagliate dei giocatori, che riflettono le loro abilità reali. Queste valutazioni sono aggiornate annualmente, rispecchiando le performance e i progressi dei giocatori nel mondo reale.

Nel contesto di questo progetto, l'obiettivo principale è sviluppare un modello predittivo che possa stimare con accuratezza le valutazioni complessive dei giocatori di FIFA per i prossimi tre anni. Questa capacità predittiva ha molteplici applicazioni pratiche, sia per i giocatori che per gli sviluppatori del gioco. Per i giocatori, conoscere in anticipo le potenziali valutazioni future dei calciatori può influenzare le loro decisioni strategiche nel gioco. Per gli sviluppatori, invece, un modello predittivo accurato può aiutare a mantenere il gioco realistico e bilanciato, migliorando ulteriormente l'esperienza dell'utente.

L'importanza delle valutazioni dei giocatori

Le valutazioni dei giocatori di FIFA sono determinate da una combinazione di fattori che includono la velocità, il dribbling, il tiro, la difesa, il passaggio e la fisicità. Questi attributi sono poi combinati per generare una valutazione complessiva, che rappresenta la qualità generale del giocatore nel gioco. Le valutazioni sono fondamentali non solo per la rappresentazione accurata dei calciatori nel gioco, ma anche per la strategia dei giocatori di FIFA. Ad esempio, un giocatore con un'elevata valutazione complessiva sarà considerato un elemento prezioso in qualsiasi squadra.

La capacità di predire le valutazioni future dei giocatori rappresenta un vantaggio significativo. Nel contesto dei giochi di simulazione sportiva come FIFA, le previsioni accurate possono migliorare la giocabilità e la realistica rappresentazione del gioco. Inoltre, queste previsioni possono offrire insights preziosi agli sviluppatori del gioco per migliorare ulteriormente la precisione delle valutazioni dei giocatori.

Motivazione del Progetto

La motivazione principale di questo progetto risiede nella possibilità di sfruttare tecniche avanzate di machine learning per creare modelli predittivi accurati e affidabili. Con l'aumento della disponibilità di dati e delle capacità computazionali, le tecniche di machine learning sono diventate strumenti potenti per l'analisi e la previsione di dati complessi. Applicare queste tecniche per predire le valutazioni dei giocatori di FIFA non solo è un campo di studio interessante, ma offre anche numerosi vantaggi pratici.

Una delle principali sfide del progetto è rappresentata dalla complessità dei dati e dalle numerose variabili che influenzano le valutazioni dei giocatori. Ogni giocatore è unico e le sue performance possono essere influenzate da un'ampia gamma di fattori, tra cui la forma fisica, le condizioni psicologiche, il livello di competizione e altri aspetti esterni. Nonostante queste sfide, l'uso di tecniche di machine learning avanzate offre la possibilità di modellare queste variabili complesse e di fornire previsioni accurate.

Origine del Dataset

Il dataset utilizzato in questo progetto è stato ottenuto da Kaggle, una piattaforma online nota per la condivisione di dataset e competizioni di data science. Il dataset specifico, denominato "EA Sports

FC 24 Complete Player Dataset", è stato creato da Stefano Leone e contiene informazioni dettagliate sui giocatori presenti nel gioco EA Sports FC 24. Questo dataset è disponibile al seguente link: [Kaggle - EA Sports FC 24 Complete Player Dataset](#).

Il dataset comprende diverse informazioni sui giocatori, tra cui le valutazioni degli attributi, le statistiche delle performance, i dati anagrafici e altre informazioni rilevanti. Per questo progetto, abbiamo focalizzato l'attenzione sul file `male_players.csv`, che contiene i dati dei giocatori di sesso maschile.

Obiettivi del Progetto

Gli obiettivi principali di questo progetto sono i seguenti:

1. **Preprocessing dei Dati:** Pulire e preparare i dati dei giocatori di FIFA per l'analisi. Questo include la gestione dei valori mancanti, la normalizzazione dei dati e la creazione di nuove caratteristiche utili per i modelli predittivi.
2. **Ingegneria delle Caratteristiche:** Identificare e creare nuove caratteristiche che possano migliorare la capacità predittiva dei modelli. Questo include l'uso di tecniche di riduzione della dimensionalità come l'analisi delle componenti principali (PCA).
3. **Addestramento dei Modelli:** Utilizzare vari algoritmi di machine learning per addestrare modelli predittivi sulle valutazioni future dei giocatori. I modelli considerati includono RandomForestRegressor, GradientBoostingRegressor e VotingRegressor.
4. **Valutazione dei Modelli:** Valutare le performance dei modelli utilizzando metriche standard come l'errore quadratico medio (RMSE) e il coefficiente di determinazione (R^2). Confrontare le performance dei diversi modelli per identificare il più accurato.
5. **Previsioni:** Applicare il modello migliore per fare previsioni sulle valutazioni future dei giocatori.
6. **Documentazione e Presentazione dei Risultati:** Documentare il processo, i metodi utilizzati e i risultati ottenuti in un rapporto dettagliato. Presentare i risultati in modo chiaro e comprensibile, utilizzando grafici e tabelle per supportare le conclusioni.

Struttura della Relazione

Questa relazione è strutturata in diverse sezioni, ognuna delle quali affronta un aspetto specifico del progetto:

1. **Background:** Una panoramica della serie FIFA e dell'importanza delle valutazioni dei giocatori.
2. **Stato dell'Arte:** Un'analisi delle tecniche esistenti e delle ricerche correlate nel campo della previsione delle valutazioni dei giocatori.
3. **Metodologia:** Dettagli sul preprocessing dei dati, l'ingegneria delle caratteristiche e le tecniche di machine learning utilizzate.
4. **Risultati:** Presentazione e analisi dei risultati ottenuti dai vari modelli predittivi.
5. **Conclusioni e Lavori Futuri:** Considerazioni finali e possibili direzioni future per migliorare ulteriormente il modello predittivo.

In conclusione, questo progetto rappresenta un'analisi approfondita dell'uso delle tecniche di machine learning per predire le valutazioni future dei giocatori di FIFA, offrendo insights preziosi sia per i giocatori che per gli sviluppatori del gioco.

Background

La Serie FIFA

La serie FIFA, sviluppata e pubblicata da Electronic Arts sotto l'etichetta EA Sports, rappresenta uno dei franchise di videogiochi più venduti e longevi nella storia del gaming. Il primo titolo della serie è stato rilasciato nel 1993, e da allora, FIFA è diventato un punto di riferimento per gli appassionati di calcio e di videogiochi. La serie è conosciuta per la sua simulazione realistica del calcio, offrendo ai giocatori la possibilità di controllare squadre e giocatori reali in una vasta gamma di modalità di gioco.

Evoluzione e Innovazioni

Nel corso degli anni, la serie FIFA ha subito numerose evoluzioni e innovazioni. Ogni nuova edizione del gioco ha introdotto miglioramenti significativi in termini di grafica, gameplay, intelligenza artificiale e modalità di gioco. Le innovazioni tecnologiche hanno permesso di creare una rappresentazione sempre più realistica del calcio, con dettagli accurati sui movimenti dei giocatori, le dinamiche del pallone e l'atmosfera degli stadi.

Un aspetto cruciale che ha contribuito al successo della serie FIFA è l'accuratezza delle valutazioni dei giocatori. Queste valutazioni, che riflettono le abilità e le performance dei giocatori nel mondo reale, sono aggiornate annualmente per garantire una rappresentazione fedele dei giocatori all'interno del gioco. Le valutazioni sono basate su una combinazione di fattori, tra cui velocità, dribbling, tiro, difesa, passaggio e fisicità.

Importanza delle Valutazioni dei Giocatori

Le valutazioni dei giocatori in FIFA non sono solo un elemento di gioco, ma rappresentano anche una componente essenziale per la strategia dei giocatori. Ad esempio, un giocatore con una valutazione complessiva elevata sarà considerato un elemento prezioso per qualsiasi squadra, influenzando le decisioni dei giocatori in modalità come Ultimate Team, Carriera e Pro Club. Le valutazioni determinano anche il comportamento dei giocatori controllati dall'intelligenza artificiale, influenzando il realismo e la difficoltà del gioco.

Le valutazioni dei giocatori sono determinate da un team di esperti di EA Sports, che analizzano le performance dei giocatori nel mondo reale e aggiornano le loro valutazioni di conseguenza. Questo processo assicura che le valutazioni riflettano accuratamente le abilità attuali dei giocatori, ma allo stesso tempo rappresenta una sfida significativa, data la natura dinamica e imprevedibile delle performance sportive.

Machine Learning e Previsione delle Valutazioni

Con l'aumento della disponibilità di dati e delle capacità computazionali, le tecniche di machine learning sono diventate strumenti potenti per l'analisi e la previsione di dati complessi. Nel contesto della serie FIFA, l'uso del machine learning per predire le valutazioni future dei giocatori rappresenta un'opportunità unica per migliorare ulteriormente la precisione e l'affidabilità delle valutazioni.

Il machine learning consente di analizzare grandi quantità di dati storici sulle performance dei giocatori e di identificare pattern e tendenze che possono essere utilizzati per fare previsioni accurate. Queste previsioni possono aiutare gli sviluppatori del gioco a mantenere il realismo e l'equilibrio del

gioco, offrendo allo stesso tempo ai giocatori insights preziosi sulle potenziali performance future dei calciatori.

Significato e Impatto del Progetto

Questo progetto ha un significato e un impatto significativo sia per la comunità dei giocatori di FIFA che per gli sviluppatori del gioco. Per i giocatori, la capacità di predire le valutazioni future dei calciatori può influenzare le loro decisioni strategiche nel gioco, migliorando la loro esperienza di gioco. Per gli sviluppatori, un modello predittivo accurato può aiutare a mantenere il gioco realistico e bilanciato, migliorando ulteriormente la soddisfazione degli utenti.

Inoltre, il progetto rappresenta un'opportunità per esplorare l'applicazione delle tecniche di machine learning in un contesto pratico e rilevante, contribuendo al progresso della ricerca nel campo dell'analisi dei dati sportivi.

In conclusione, questo progetto rappresenta un'analisi approfondita dell'uso delle tecniche di machine learning per predire le valutazioni future dei giocatori di FIFA, offrendo insights preziosi sia per i giocatori che per gli sviluppatori del gioco. La combinazione di dati accurati, tecniche avanzate di machine learning e un'analisi dettagliata dei risultati rende questo progetto un contributo significativo al campo dei giochi di simulazione sportiva.

Stato dell'Arte

Introduzione all'Intelligenza Artificiale e al Machine Learning

L'Intelligenza Artificiale (IA) e il Machine Learning (ML) rappresentano due dei campi più dinamici e innovativi della tecnologia moderna. L'IA si riferisce alla capacità di un sistema di eseguire compiti che normalmente richiederebbero intelligenza umana, come la percezione visiva, il riconoscimento vocale, il processo decisionale e la traduzione tra lingue. Il Machine Learning, un sottocampo dell'IA, si concentra sulla progettazione e lo sviluppo di algoritmi che permettono ai computer di imparare dai dati e di fare previsioni o decisioni senza essere esplicitamente programmati per ogni compito.

Applicazioni del Machine Learning nello Sport

L'applicazione del Machine Learning nello sport è cresciuta esponenzialmente negli ultimi anni. Le tecniche di ML sono utilizzate per analizzare le performance degli atleti, prevedere risultati di partite, ottimizzare le strategie di gioco e persino migliorare l'esperienza dei fan attraverso la personalizzazione dei contenuti. In particolare, nel calcio, l'analisi dei dati ha rivoluzionato il modo in cui gli allenatori, i giocatori e i club prendono decisioni, basandosi su informazioni dettagliate e approfondite ottenute dai dati delle partite.

Previsione delle Performance degli Atleti

Uno degli aspetti più affascinanti del Machine Learning applicato allo sport è la previsione delle performance degli atleti. Modelli predittivi possono essere addestrati su dati storici delle performance per prevedere future prestazioni, infortuni e sviluppi della carriera. Questo tipo di analisi è particolarmente utile per gli scout, gli allenatori e i manager dei club, che possono utilizzare queste informazioni per prendere decisioni strategiche su trasferimenti, formazione e gestione dei giocatori.

La Serie FIFA e l'Analisi dei Giocatori

La serie di videogiochi FIFA di EA Sports ha sempre puntato a una simulazione realistica del calcio, e parte di questo realismo deriva dall'accuratezza delle valutazioni dei giocatori. Ogni anno, EA Sports aggiorna le valutazioni dei giocatori basandosi sulle loro performance reali. Queste valutazioni

includono attributi come velocità, dribbling, tiro, difesa, passaggio e fisicità, che insieme determinano la valutazione complessiva di un giocatore.

Metodologie di Valutazione Tradizionali

Tradizionalmente, le valutazioni dei giocatori di FIFA sono determinate da un team di esperti che analizzano le performance dei giocatori nelle partite reali. Questo processo è in parte soggettivo e può essere influenzato da vari fattori, come il contesto della partita, la qualità dell'avversario e le condizioni fisiche del giocatore. Sebbene questo approccio abbia garantito una certa accuratezza, presenta limitazioni in termini di scalabilità e oggettività.

Introduzione del Machine Learning nelle Valutazioni

Con l'aumento della disponibilità di dati dettagliati sulle performance dei giocatori e dei progressi nelle tecniche di Machine Learning, EA Sports e altri sviluppatori di giochi hanno iniziato a esplorare l'uso del ML per migliorare le valutazioni dei giocatori. I modelli di Machine Learning possono analizzare grandi quantità di dati e identificare pattern che potrebbero non essere evidenti agli esperti umani. Questo approccio può aumentare l'accuratezza e l'oggettività delle valutazioni, garantendo al contempo la scalabilità necessaria per aggiornare le valutazioni di migliaia di giocatori.

Modelli di Machine Learning Utilizzati

Diversi modelli di Machine Learning possono essere utilizzati per prevedere le valutazioni dei giocatori. Tra i più comuni ci sono i seguenti:

Regressione Lineare

La regressione lineare è uno dei modelli più semplici e interpretabili. Tuttavia, la sua capacità di catturare la complessità dei dati delle performance sportive è limitata. Questo modello assume una relazione lineare tra le variabili indipendenti (attributi del giocatore) e la variabile dipendente (valutazione complessiva), che potrebbe non essere sempre il caso nel contesto sportivo.

Random Forest

Il Random Forest è un algoritmo di apprendimento supervisionato che utilizza un ensemble di alberi decisionali. È robusto agli outlier e alle variabili non lineari, il che lo rende particolarmente adatto per i dati delle performance sportive, che possono essere altamente variabili e complessi. Il Random Forest combina i risultati di molti alberi decisionali per migliorare la precisione delle previsioni e ridurre il rischio di overfitting.

Gradient Boosting

Il Gradient Boosting è un'altra tecnica di ensemble che crea un modello predittivo forte combinando molti modelli deboli, solitamente alberi decisionali. Ogni nuovo albero corregge gli errori commessi dai precedenti, migliorando progressivamente la precisione delle previsioni. Questo modello è molto potente e spesso fornisce risultati eccellenti nei concorsi di data science.

XGBoost

XGBoost (Extreme Gradient Boosting) è una versione ottimizzata del Gradient Boosting che è stata progettata per essere altamente efficiente, flessibile e portatile. È particolarmente noto per le sue prestazioni e velocità, ed è ampiamente utilizzato nelle competizioni di machine learning.

Voting Regressor

Il Voting Regressor è un modello di ensemble che combina le previsioni di diversi modelli base (ad esempio, Random Forest, Gradient Boosting e XGBoost) prendendo una media ponderata delle

previsioni. Questo approccio può migliorare la robustezza e l'accuratezza delle previsioni, sfruttando i punti di forza di ciascun modello base.

Applicazioni Pratiche e Sfide

L'uso del Machine Learning per la previsione delle valutazioni dei giocatori in FIFA presenta diverse applicazioni pratiche, ma anche alcune sfide:

Applicazioni Pratiche

1. **Aggiornamenti delle Valutazioni:** Automatizzare il processo di aggiornamento delle valutazioni dei giocatori basandosi su dati delle performance in tempo reale.
2. **Scouting e Reclutamento:** Fornire ai club strumenti avanzati per identificare talenti emergenti e prendere decisioni informate sui trasferimenti.
3. **Personalizzazione del Gioco:** Migliorare l'esperienza del giocatore personalizzando le valutazioni e le dinamiche di gioco in base alle prestazioni recenti dei calciatori.

Sfide

1. **Qualità dei Dati:** La qualità e la completezza dei dati sono cruciali per l'accuratezza delle previsioni. Dati mancanti o errati possono compromettere i risultati.
2. **Interpretabilità:** I modelli di Machine Learning possono essere complessi e difficili da interpretare, il che può limitare la trasparenza delle valutazioni.
3. **Adattamento al Cambiamento:** Le performance dei giocatori possono variare notevolmente da una stagione all'altra, e i modelli devono essere in grado di adattarsi rapidamente a questi cambiamenti.

Conclusione

Il Machine Learning offre opportunità significative per migliorare la precisione e l'oggettività delle valutazioni dei giocatori in FIFA, superando le limitazioni dei metodi tradizionali basati su giudizi umani. Tuttavia, l'implementazione efficace di queste tecniche richiede attenzione alla qualità dei dati, alla scelta dei modelli e alla gestione delle complessità operative. Il nostro progetto mira a esplorare queste opportunità e sfide, contribuendo alla ricerca e allo sviluppo di soluzioni avanzate per l'analisi delle performance sportive.

Metodologie

Introduzione

Le metodologie rappresentano il cuore di qualsiasi progetto di machine learning, determinando il successo delle previsioni e la validità dei risultati. In questo progetto, abbiamo seguito un processo sistematico che includeva la raccolta e il preprocessing dei dati, l'addestramento dei modelli, la valutazione delle performance e la produzione di previsioni. Di seguito, dettagliamo ciascuna di queste fasi, illustrando le tecniche e gli strumenti utilizzati per garantire l'accuratezza e la robustezza delle nostre previsioni sulle valutazioni dei giocatori di FIFA.

Raccolta e Preprocessing dei Dati

Raccolta dei Dati

Il dataset utilizzato, `male_players.csv`, è stato scaricato da Kaggle. Questo file contiene una vasta gamma di informazioni sui giocatori maschili, tra cui le valutazioni degli attributi, le statistiche delle

performance, i dati anagrafici e altre informazioni rilevanti. La qualità e la completezza di questo dataset erano cruciali per il successo del progetto.

Pulizia dei Dati

Il processo di pulizia dei dati ha incluso diverse fasi chiave:

- **Rimozione dei portieri:** I portieri sono stati rimossi dal dataset poiché le loro valutazioni e attributi differiscono significativamente dai giocatori di campo.
- **Gestione delle etichette dei giocatori (player_tags) e delle caratteristiche (player_traits):** Queste colonne sono state trasformate in variabili dummy per essere incluse nei modelli di machine learning.
- **Ordinamento dei dati:** I dati sono stati ordinati per player_id e fifa_version per mantenere la coerenza temporale delle valutazioni.
- **Creazione delle variabili di overall per i 3 anni successivi:** Sono state create 3 ulteriori colonne, una per ogni anno successivo, relative agli overall.
- **Rimozione di colonne irrilevanti:** Colonne non necessarie per l'analisi sono state rimosse per semplificare il dataset.

Gestione delle Colonne Categoriali

Le colonne categoriali come player_positions, preferred_foot, club_name, league_name, nationality_name, work_rate, e body_type sono state codificate utilizzando il LabelEncoder per convertire le categorie in valori numerici. Questo passaggio era essenziale per permettere ai modelli di machine learning di elaborare queste informazioni.

Imputazione e Standardizzazione dei Dati

Per affrontare i valori mancanti nelle colonne numeriche, è stata utilizzata l'imputazione con la media (SimpleImputer). Successivamente, le colonne numeriche sono state standardizzate utilizzando lo StandardScaler per garantire che tutte le variabili avessero una media di 0 e una deviazione standard di 1, facilitando il processo di apprendimento dei modelli.

Creazione di Nuove Feature

Abbiamo creato nuove feature per migliorare la capacità predittiva dei modelli:

- **Trend dell'overall dei giocatori (overall_trend):** Calcolato come la differenza media dell'overall tra le diverse versioni di FIFA.
- **Valutazione media degli attributi principali (average_rating):** Calcolata come la media delle valutazioni di pace, shooting, passing, dribbling, defending, e physic.

Riduzione della Dimensionalità con PCA

Per gestire l'alta dimensionalità del dataset e ridurre il rischio di overfitting, è stata utilizzata l'Analisi delle Componenti Principali (PCA) per ridurre la dimensionalità mantenendo il 95% della varianza. Le componenti principali ottenute sono state aggiunte al dataset originale.

Addestramento dei Modelli

Divisione del Dataset

Il dataset preprocessato è stato diviso in set di addestramento, validazione e test. La suddivisione è stata effettuata con una proporzione di 80% per l'addestramento, 10% per la validazione e 10% per il

test. Questa suddivisione ha permesso di valutare le performance dei modelli in modo robusto e di prevenire l'overfitting.

Modelli Utilizzati

Abbiamo addestrato diversi modelli di machine learning per prevedere le valutazioni dei giocatori nei tre anni successivi. I modelli utilizzati sono stati:

1. **Random Forest Regressor:**

- Tre modelli distinti (rf1, rf2, rf3) per prevedere le valutazioni dei giocatori per ciascuno dei tre anni successivi.
- Addestrati utilizzando l'algoritmo RandomForestRegressor con 100 alberi.

2. **Gradient Boosting Regressor:**

- Tre modelli distinti (gb1, gb2, gb3) per le previsioni annuali.
- Addestrati utilizzando l'algoritmo GradientBoostingRegressor con 100 iterazioni.

3. **Voting Regressor:**

- Tre modelli distinti (vr1, vr2, vr3) che combinano le previsioni di Random Forest e Gradient Boosting.
- Implementati utilizzando l'algoritmo VotingRegressor per combinare i modelli base.

4. **Randomized Search CV:**

- Tre modelli distinti (best_rs1, best_rs2, best_rs3) ottimizzati con RandomizedSearchCV.
- Parametri ottimizzati includono il numero di stimatori, la profondità massima, il numero minimo di campioni per divisione e il numero minimo di campioni per foglia.

5. **Grid Search CV:**

- Tre modelli distinti ottimizzati con GridSearchCV (successivamente rimosso a causa dei tempi di addestramento elevati).
- Parametri ottimizzati simili a quelli di Randomized Search CV.

Valutazione dei Modelli

Metriche di Valutazione

Per valutare le performance dei modelli, abbiamo utilizzato due metriche principali:

1. **Root Mean Squared Error (RMSE):**

- Misura la deviazione quadratica media tra i valori predetti e i valori reali.
- Fornisce un'indicazione della precisione delle previsioni dei modelli.

2. **R² Score (Coefficient of Determination):**

- Indica la proporzione della varianza nei dati dipendenti che è spiegata dal modello.
- Un R² alto indica che il modello spiega bene la varianza dei dati.

Confronto dei Risultati

I risultati dei modelli sono stati confrontati per identificare il migliore in termini di RMSE e R^2 . Questo confronto ha evidenziato che il Voting Regressor ha fornito le migliori performance complessive per ciascuno dei tre anni successivi.

Produzione di Previsioni

Implementazione delle Previsioni

Utilizzando il Voting Regressor come modello finale, abbiamo generato previsioni per le valutazioni dei giocatori nei tre anni successivi. Queste previsioni sono state integrate nel dataset originale per creare un dataset finale che include sia le valutazioni storiche che le previsioni future.

Conclusione

Le metodologie adottate in questo progetto hanno permesso di sfruttare appieno le tecniche di machine learning per prevedere le future valutazioni dei giocatori di FIFA. Il processo sistematico di raccolta, pulizia, preprocessing, addestramento e valutazione dei modelli ha garantito risultati accurati e robusti. L'uso del Voting Regressor come modello finale ha dimostrato di essere la scelta più efficace, fornendo previsioni affidabili che possono essere utilizzate per vari scopi pratici, come il miglioramento del gameplay, il supporto alle decisioni di scouting e la personalizzazione dell'esperienza utente.

Risultati

Introduzione

In questa sezione, presenteremo i risultati ottenuti dai vari modelli di machine learning utilizzati per prevedere le valutazioni dei giocatori di FIFA nei tre anni successivi. L'analisi dettagliata delle performance dei modelli, delle metriche di valutazione e delle previsioni finali consentirà di comprendere l'efficacia delle tecniche adottate e l'affidabilità delle previsioni. Questa sezione si suddivide in tre sottosezioni principali: valutazione delle performance dei modelli, confronto tra i modelli e interpretazione delle previsioni.

Valutazione delle Performance dei Modelli

Random Forest Regressor

Abbiamo addestrato tre modelli Random Forest Regressor (rf1, rf2, rf3) per prevedere le valutazioni dei giocatori nei tre anni successivi. Di seguito sono riportati i risultati:

- **Modello 1 (rf1):**
 - RMSE: 0.324761
 - R^2 : 0.894738
- **Modello 2 (rf2):**
 - RMSE: 0.372375
 - R^2 : 0.860832
- **Modello 3 (rf3):**
 - RMSE: 0.428664
 - R^2 : 0.816064

I modelli Random Forest hanno dimostrato una buona capacità predittiva, con R^2 superiori a 0.8 per tutti e tre gli anni. Tuttavia, la precisione delle previsioni tende a diminuire leggermente con l'aumentare del periodo di previsione.

Gradient Boosting Regressor

Anche per il Gradient Boosting Regressor abbiamo addestrato tre modelli distinti (gb1, gb2, gb3). I risultati sono stati i seguenti:

- **Modello 1 (gb1):**
 - RMSE: 0.313148
 - R^2 : 0.902131
- **Modello 2 (gb2):**
 - RMSE: 0.359492
 - R^2 : 0.870295
- **Modello 3 (gb3):**
 - RMSE: 0.419146
 - R^2 : 0.824141

I modelli Gradient Boosting hanno ottenuto performance leggermente migliori rispetto ai Random Forest, soprattutto per il primo anno di previsione. Questo suggerisce che il Gradient Boosting potrebbe essere più adatto per catturare le complessità nei dati dei giocatori.

Voting Regressor

Il Voting Regressor, che combina le previsioni di Random Forest e Gradient Boosting, ha prodotto i seguenti risultati:

- **Modello 1 (vr1):**
 - RMSE: 0.314708
 - R^2 : 0.901153
- **Modello 2 (vr2):**
 - RMSE: 0.358150
 - R^2 : 0.871262
- **Modello 3 (vr3):**
 - RMSE: 0.414145
 - R^2 : 0.828313

Il Voting Regressor ha dimostrato di essere il modello più robusto, fornendo previsioni accurate e stabili per tutti e tre gli anni. Le sue performance sono comparabili a quelle del Gradient Boosting, ma con una maggiore consistenza tra i diversi periodi di previsione.

Randomized Search CV

Infine, abbiamo utilizzato Randomized Search CV per ottimizzare i parametri dei modelli Random Forest. I risultati sono stati i seguenti:

- **Modello 1 (best_rs1):**
 - RMSE: 0.337865
 - R^2 : 0.886071
- **Modello 2 (best_rs2):**
 - RMSE: 0.403687
 - R^2 : 0.836443
- **Modello 3 (best_rs3):**
 - RMSE: 0.462813
 - R^2 : 0.785591

Sebbene Randomized Search CV abbia migliorato leggermente le performance rispetto ai modelli base di Random Forest, i risultati sono stati inferiori rispetto ai modelli di Gradient Boosting e Voting Regressor.

Confronto tra i Modelli

Il confronto tra i vari modelli di machine learning ha evidenziato che il Voting Regressor è il modello più efficace per prevedere le valutazioni dei giocatori di FIFA nei tre anni successivi. Questo modello ha ottenuto i migliori risultati sia in termini di RMSE che di R^2 per tutti e tre i periodi di previsione.

In particolare:

- **Per il primo anno (y1)**, il Voting Regressor ha ottenuto un RMSE di 0.314708 e un R^2 di 0.901153, leggermente inferiore al Gradient Boosting ma con una maggiore consistenza.
- **Per il secondo anno (y2)**, il Voting Regressor ha raggiunto un RMSE di 0.358150 e un R^2 di 0.871262, superando sia il Random Forest che il Gradient Boosting.
- **Per il terzo anno (y3)**, il Voting Regressor ha ottenuto un RMSE di 0.414145 e un R^2 di 0.828313, dimostrando la sua robustezza anche per le previsioni a lungo termine.

Interpretazione delle Previsioni

Le previsioni ottenute dai modelli Voting Regressor sono state integrate nel dataset originale, fornendo una visione completa delle valutazioni storiche e future dei giocatori. Questo consente di analizzare l'evoluzione delle performance dei giocatori e di identificare potenziali trend o cambiamenti nel tempo.

Valutazioni Future

Le previsioni delle valutazioni future dei giocatori sono cruciali per diverse applicazioni pratiche. Ad esempio, possono essere utilizzate per:

- **Supporto alle Decisioni di Scouting:** I club di calcio possono utilizzare le previsioni per identificare i giocatori promettenti e prendere decisioni informate sugli acquisti e le cessioni.

- **Personalizzazione del Gameplay:** I dati sulle previsioni possono essere integrati nei videogiochi di simulazione di calcio per migliorare l'esperienza utente, rendendo il gioco più realistico e dinamico.
- **Analisi delle Performance:** Le previsioni possono aiutare gli allenatori e gli analisti a monitorare lo sviluppo dei giocatori e a pianificare le sessioni di allenamento in modo più efficace.

Accuratezza delle Previsioni

L'accuratezza delle previsioni è stata validata utilizzando set di dati di test, garantendo che i modelli siano in grado di generalizzare bene a dati non visti durante l'addestramento. I risultati elevati di R^2 indicano che i modelli sono stati in grado di spiegare gran parte della varianza nelle valutazioni dei giocatori, mentre i bassi valori di RMSE suggeriscono che le previsioni sono vicine ai valori reali.

Conclusione

I risultati ottenuti dai modelli di machine learning sviluppati in questo progetto dimostrano la loro efficacia nel prevedere le valutazioni future dei giocatori di FIFA. Il Voting Regressor, in particolare, si è rivelato il modello più robusto e accurato, fornendo previsioni affidabili per tutti e tre gli anni successivi. Queste previsioni possono essere utilizzate per una varietà di applicazioni pratiche, contribuendo a migliorare le decisioni nel mondo del calcio e a personalizzare l'esperienza dei videogiochi di simulazione sportiva.

I risultati di questo progetto evidenziano l'importanza di un processo metodologico rigoroso, dalla raccolta e pulizia dei dati all'addestramento e valutazione dei modelli. La combinazione di tecniche di preprocessing, ottimizzazione dei modelli e valutazione rigorosa ha permesso di ottenere previsioni accurate e utili, dimostrando il potenziale del machine learning nell'analisi e previsione delle performance sportive.

Conclusioni e Lavori Futuri

Conclusioni

Il presente progetto ha esplorato l'utilizzo di tecniche avanzate di machine learning per prevedere le valutazioni future dei giocatori di FIFA, concentrandosi su tre orizzonti temporali: uno, due e tre anni. Attraverso un'analisi approfondita e un'approccio metodologico rigoroso, sono stati sviluppati e valutati diversi modelli di machine learning, tra cui Random Forest, Gradient Boosting, Voting Regressor e Randomized Search CV.

I risultati ottenuti hanno dimostrato che il Voting Regressor, una combinazione di Random Forest e Gradient Boosting, è il modello più efficace per prevedere le valutazioni dei giocatori nei tre anni successivi. Questo modello ha ottenuto performance superiori in termini di RMSE e R^2 , indicando una maggiore accuratezza e robustezza delle previsioni rispetto agli altri modelli testati.

Il processo di previsione è stato validato utilizzando set di dati di test, garantendo la capacità dei modelli di generalizzare a dati non visti. Le previsioni delle valutazioni future dei giocatori forniscono informazioni preziose per una varietà di applicazioni pratiche, tra cui il supporto alle decisioni di scouting, la personalizzazione del gameplay nei videogiochi di simulazione sportiva e l'analisi delle performance dei giocatori.

Lavori Futuri

Sebbene i risultati ottenuti siano promettenti, vi sono diverse direzioni future che possono essere esplorate per migliorare ulteriormente la precisione e l'utilità delle previsioni delle valutazioni dei giocatori di FIFA. Di seguito sono riportate alcune delle principali aree di ricerca e sviluppo future:

1. Integrazione di Nuove Variabili

L'integrazione di nuove variabili e fonti di dati potrebbe migliorare ulteriormente le previsioni. Ad esempio, dati su infortuni, trasferimenti, performance nelle partite internazionali e altre metriche di performance avanzate potrebbero fornire una visione più completa delle capacità dei giocatori e del loro potenziale futuro.

2. Utilizzo di Modelli di Deep Learning

I modelli di deep learning, come le reti neurali convoluzionali (CNN) e le reti neurali ricorrenti (RNN), potrebbero essere esplorati per catturare pattern più complessi nei dati. Questi modelli hanno dimostrato un grande potenziale in vari campi e potrebbero offrire miglioramenti significativi nelle previsioni delle valutazioni dei giocatori.

3. Analisi dei Trend Temporali

Un'analisi più dettagliata dei trend temporali delle performance dei giocatori potrebbe fornire informazioni utili per migliorare le previsioni. Ad esempio, l'utilizzo di modelli di serie temporali, come ARIMA o LSTM, potrebbe catturare meglio le dinamiche temporali delle valutazioni dei giocatori.

4. Ottimizzazione dei Modelli Ensemble

L'ottimizzazione dei modelli ensemble potrebbe essere ulteriormente esplorata. Ad esempio, l'uso di tecniche come il boosting adattivo o l'ottimizzazione bayesiana potrebbe migliorare le performance del Voting Regressor o di altri modelli ensemble.

5. Interpretabilità dei Modelli

L'interpretabilità dei modelli di machine learning è un aspetto cruciale, soprattutto quando si utilizzano modelli complessi come il Gradient Boosting o il Voting Regressor. Tecniche di interpretazione dei modelli, come SHAP (SHapley Additive exPlanations) o LIME (Local Interpretable Model-agnostic Explanations), potrebbero essere utilizzate per comprendere meglio l'importanza delle variabili e il funzionamento dei modelli.

6. Valutazione Continua e Aggiornamento dei Modelli

I modelli di machine learning devono essere continuamente valutati e aggiornati per mantenere la loro accuratezza nel tempo. La costruzione di un sistema di aggiornamento automatico dei modelli, basato su nuovi dati che diventano disponibili, potrebbe garantire che le previsioni rimangano accurate e rilevanti.

7. Espansione ad Altri Ambiti Sportivi

Sebbene questo progetto si sia concentrato sui giocatori di FIFA, le tecniche sviluppate potrebbero essere applicate ad altri ambiti sportivi. Ad esempio, la previsione delle performance future di atleti in altri sport come il basket, il baseball o il tennis potrebbe beneficiare delle metodologie e dei modelli presentati in questo progetto.

Conclusione Finale

In conclusione, il progetto ha dimostrato l'efficacia delle tecniche di machine learning nel prevedere le valutazioni future dei giocatori di FIFA. I risultati ottenuti evidenziano il potenziale di queste

tecniche per supportare le decisioni nel mondo del calcio e per migliorare l'esperienza utente nei videogiochi di simulazione sportiva. Tuttavia, vi sono ancora molte opportunità di miglioramento e di espansione della ricerca, che potrebbero portare a risultati ancora più significativi e applicazioni pratiche più ampie.