

Tema: Predikcija Pobjednika na UEFA Euro 2024

1. Definicija Problema

Cilj projekta je razviti model mašinskog učenja koji može predvidjeti pobjednika UEFA Euro 2024 takmičenja na osnovu istorijskih podataka o prethodnim turnirima, performansama timova i drugih relevantnih faktora. Ovo predstavlja tipičan problem klasifikacije, gde će ulaz u model biti različite karakteristike timova (npr. broj golova postignutih/primljenih, posedom lopte, broj povreda ključnih igrača, ishod prethodnih mečeva), dok će izlaz biti predikcija o tome koji tim će osvojiti turnir (npr. pobjednički tim ili top nekoliko favorita).

Ključni aspekti ovog problema uključuju:

- **Podaci:** Potrebno je prikupiti podatke o prethodnim UEFA Euro takmičenjima, uključujući rezultate mečeva, statistike igrača, taktičke postavke timova, povrede igrača i druge relevantne faktore. Ovi podaci će poslužiti kao osnova za obučavanje i testiranje modela.
- **Predikcija pobjednika:** Centralni zadatak modela će biti predviđanje pobjednika UEFA Euro 2024 takmičenja. Ovo će zahtevati analizu istorijskih podataka kako bi se identifikovali ključni faktori koji su doprineli uspehu timova u prethodnim turnirima.
- **Klasifikacioni pristup:** Problematika se može posmatrati kao klasifikacioni problem, gde će model klasifikovati timove u različite kategorije (npr. pobjednički tim, polufinalisti, eliminisani u grupnoj fazi). Klase mogu biti definisane na osnovu rangiranja ili pozicija timova na turniru.
- **Relevantni faktori:** Važno je identifikovati i uključiti relevantne faktore koji mogu uticati na ishod turnira, kao što su trenutna forma timova, povrede ključnih igrača, taktičke strategije, kao i faktori van terena poput atmosfere u timu ili očekivanja javnosti.

2. Motivacija za problem koji se rešava

Ova predikcija može imati širok spektar koristi i primena:

- **Sportski analitičari:** Precizna predikcija pobjednika UEFA Euro takmičenja omogućava sportskim analitičarima da dublje razumeju faktore koji doprinose uspehu timova. Analiza performansi timova, taktičkih strategija i ključnih igrača može pružiti uvide u dinamiku fudbalskih turnira i pomoći u razvoju boljih strategija za buduće takmičenje.
- **Osobe koje se bave sportskim klađenjem:** Ovaj tip ljudi je uvek zainteresovan za tačne predikcije rezultata sportskih događaja kako bi ostvarili profit. Precizna predikcija pobjednika UEFA Euro 2024 takmičenja može im pružiti dragocene informacije za postavljanje pametnih opklada i smanjenje rizika od gubitka novca.
- **Sportski mediji:** Mediji su uvek u potrazi za intrigantnim pričama i unikatnim sadržajem koji će privući pažnju publike. Predikcija pobjednika UEFA Euro 2024 takmičenja može poslužiti kao tema za analitičke članke, intervju sa stručnjacima i predviđanja koja će angažovati navijače širom sveta.
- **Fudbalski navijači:** Navijači širom sveta su strastveni u praćenju UEFA Euro takmičenja i podržavanju svojih omiljenih timova. Precizna predikcija pobjednika može dodatno pobuditi uzbuđenje među navijačima i pružiti im dodatni razlog za praćenje i uživanje u turniru.

Ove koristi naglašavaju značaj razvijanja modela za predviđanje pobjednika UEFA Euro 2024 takmičenja ne samo za istraživače, već i za širu zajednicu koja je zainteresovana za fudbal i sportsku analitiku.

3. Skup podataka

- **Link do skupa podataka:** [UEFA Euro 2024 Dataset](#), [UEFA Euro 2024 Dataset 2](#)
- **Opis skupa podataka:**

Broj instanci: 500 - 1000

Broj atributa: 11

Opis atributa:

Tim domaćin: Ime domaćeg tima.

Tim gost: Ime gostujućeg tima.

Datum: Datum odigravanja meča.

Golovi domaćin: Broj golova postignutih od strane domaćeg tima.

Golovi gost: Broj golova postignutih od strane gostujućeg tima.

Posed lopte domaćin: Procenat vremena provedenog u posedu lopte od strane domaćeg tima.

Posed lopte gost: Procenat vremena provedenog u posedu lopte od strane gostujućeg tima.

Broj povreda domaćin: Ukupan broj povreda igrača domaćeg tima tokom meča.

Broj povreda gost: Ukupan broj povreda igrača gostujućeg tima tokom meča.

Ishod meča: Ciljno obeležje koje označava ishod meča (ciljno obeležje):

Klasifikacija: Pobjeda domaćeg tima (1), pobjeda gostujućeg tima (-1), nerešen rezultat (0).

Najznačajniji atributi za predikciju pobjednika UEFA Euro 2024 takmičenja su broj golova postignutih od strane domaćeg i gostujućeg tima, kao i procenat vremena provedenog u posedu lopte. Ciljno obeležje je ishod meča, gde su moguće tri klase: pobjeda domaćeg tima, pobjeda gostujućeg tima ili nerešen rezultat.

4. Način pretprocesiranja podataka

Za skup podataka UEFA Euro 2024, pretprocesiranje će uključivati nekoliko ključnih koraka. Prvo, biće neophodno obraditi eventualne nedostajuće vrednosti u podacima, poput praznih polja ili nedostajućih atributa za određene mečeve. Zatim će se numerički atributi, poput broja golova ili procenata poseda lopte, normalizovati kako bi se osigurala konzistentnost među različitim skalama vrednosti. Kategorički atributi, kao što su imena timova, moći će se kodirati u numeričke forme, na primer, korišćenjem tehnika kao što je one-hot encoding. Na kraju, podaci će se skalirati radi smanjenja varijabilnosti među atributima i unapređenja performansi modela tokom treniranja.

5. Metodologija – način rešavanja problema

- **Pregled podataka:** Prvi korak u metodologiji je detaljan pregled i analiza skupa podataka UEFA Euro 2024. Ovaj korak uključuje razumevanje strukture podataka, identifikaciju atributa koji su relevantni za predikciju pobjednika, kao i uočavanje eventualnih nedostataka ili anomalija u podacima.
- **Pretprocesiranje podataka:** Nakon pregleda podataka, sledi korak pretprocesiranja podataka u skladu sa prethodno opisanim metodama. Ovde će se obraditi nedostajuće vrednosti, normalizovati numerički atributi, kodirati kategorički atributi i skalirati podaci kako bi bili spremni za treniranje modela.
- **Izbor modela:** Zatim će se izabrati odgovarajući modeli mašinskog učenja za rešavanje problema predikcije pobjednika UEFA Euro 2024. Ovo može uključivati upotrebu različitih algoritama kao

što su Random Forest, Support Vector Machines, ili Neural Networks, uz mogućnost eksperimentisanja sa kombinacijom više modela radi poboljšanja performansi predikcije.

- **Treniranje modela:** Nakon izbora modela, sledi korak treniranja modela na podacima za obuku. Ovde će se koristiti skup podataka sa adekvatno izvršenim pretprocesiranjem kako bi se model naučio prepoznavati obrasce i relacije između ulaznih atributa i ciljnog obeležja (ishoda meča).
- **Evaluacija modela:** Konačno, model će biti evaluiran korišćenjem skupa podataka za testiranje. Različite metrike performansi, poput tačnosti, preciznosti i odziva biće korišćene za ocenu performansi modela. Ova evaluacija pomoći će u proceni koliko dobro model generalizuje na novim podacima i koliko pouzdano može predvideti pobednika UEFA Euro 2024.

Ulaz u model će biti pretprocesirani skup podataka koji sadrži informacije o atributima relevantnim za predikciju, dok će izlaz biti predikcija o tome koji tim će osvojiti turnir. Modeli mašinskog učenja će se koristiti za analizu podataka i kreiranje predikcija na osnovu naučenih relacija između ulaznih atributa i ciljnog obeležja.

6. Način evaluacije

- **Train/Validation/Test podela:** Pre evaluacije modela, skup podataka će biti podeljen na tri subseta: trening, validacija i test. Trening skup će se koristiti za treniranje modela, validacioni skup za podešavanje hiperparametara modela, dok će test skup biti zadržan kao nezavisni set podataka za konačnu evaluaciju performansi modela.
- **Metrika performansi:** Kao metriku performansi modela koristićemo tačnost, koja meri procenat tačno predviđenih rezultata u odnosu na ukupan broj instanci u test skupu. Ova metrika je pogodna za klasifikacione probleme poput predikcije pobednika UEFA Euro 2024 takmičenja, gde je cilj precizno predvideti ishod mečeva.
- **Evaluacija rezultata:** Nakon treniranja modela, biće evaluirani njegovi rezultati na test skupu korišćenjem definisane metrike tačnosti. Visoka tačnost modela na test skupu ukazaće na dobru generalizaciju modela na novim, neviđenim podacima, što je ključni cilj u razvoju modela za predikciju pobednika UEFA Euro 2024 takmičenja.

7. Tehnologije

Za realizaciju projekta biće korišćen Python kao glavni programski jezik, uz korišćenje biblioteka poput scikit-learn za implementaciju algoritama mašinskog učenja, Pandas za manipulaciju podacima, NumPy za matematičke operacije, i Matplotlib i Seaborn za vizualizaciju rezultata.

8. Relevantna literatura

Ovo su primeri radova sa sličnom temom predikcije pobednika na svetskim prvenstvima, koji mi mogu poslužiti kao inspiracija i referenca u mom istraživanju.

Linkovi ka tim radovima su: [Projekat 1](#), [Projekat 2](#) i [Projekat 3](#).