**Prezicerea rezultatului meciului în baza locului desfășurării**

**Babencu Daniel, IA-211**

*Universitatea Tehnică a Moldovei, Facultatea Calculatoare, Inginerie și Micro-Electronică, Chișinău, Moldova, Informatica aplicată.*

**Introducere**

Analiza efectuată se bazeaza pe setul de date în domeniul fotbalului, având ca scop evidențierea impactului diferitor factori asupra rezultatelor meciurilor. Obiectivul studiului este de a determina dacă există o relație vizibilă între locația unui meci de fotbal și rezultatul său final.

Prin examinarea atentă a acestui set de date, voi evidenția modele care ar putea contura influența locului de desfășurare a meciului asupra probabilității de victorie a echipei gazdă, a unui rezultat de egalitate sau a victoriei echipei oaspete.

Obiectivele principale ale acestei analize includ descifrarea dacă locul desfășurprii are o influență statistic semnificativă asupra rezultatelor meciurilor și dacă aceste efecte sunt consistente în diferite turnee. În plus, îmi propun să dezvălui tendințe temporale, explorând dacă a existat o evoluție în influența locurilor de desfășurare asupra rezultatelor de-a lungul timpului.

**Abstract**

În domeniul analizei sportive, analiza computațională bazată pe date joacă un rol crucial în dezvăluirea de modele și înțelegerea detaliilor. Acest studiu explorează datele de fotbal international din perioada 1871-2023. Scopul principal al analizei rezultatelor meciurilor de fotbal este să investigheze dacă locul desfășurării influențează rezultatul acestora și, în consecință, să ofere o bază solidă pentru realizarea de predicții în legătură cu rezultatele viitoarelor meciuri.

Metodologia utilizată în analiza setului de date de fotbal a constat în aplicarea regresiei logistice pentru a prezice rezultatele meciurilor. Analiza a inclus, de asemenea, utilizarea unui tabel de contingenta pentru a examina relațiile dintre variabile și interpretarea coeficienților pentru înțelegerea contribuției fiecărei variabile la predicții.

Pe baza rezultatelor analizei, se propune dezvoltarea unui model predictiv care să poată anticipa rezultatele meciurilor viitoare, având în vedere locul desfășurării. Acest model poate fi utilizat pentru a oferi predicții cu privire la câștigătorii meciurilor în funcție de locul în care acestea au loc.

Analiza rezultatelor meciurilor de fotbal poate oferi perspective valoroase asupra influenței locului desfășurării și poate contribui la îmbunătățirea strategiilor echipelor și a preciziei predicțiilor privind rezultatele meciurilor.

Analiza are scopul de a aduce o contribuție semnificativă la înțelegerea dinamicilor din fotbal și de a oferi informații utile persoanelor interesate în lumea fotbalului.

**Materiale și Metode**

**Descărcare și Pregătirea Setului de Date**

Am descărcat setul de date de fotbal de pe Kaggle[2], și l-am pregătit pentru analiză. Acest set cuprinde rezultatele meciurilor internaționale din perioada 1872-2023 și conține 11 coloane și 44934 de rînduri.

Am adăugat, de asemenea, coloane suplimentare pentru a facilita analiza ulterioară, precum coloana cu variabile de tip categorial pentru rezultatul meciurilor. De asemenea, am adăugat variabila home\_results, care indică dacă echipa de acasă a câștigat sau nu (1 pentru victorie, 0 altfel). Aceste adăugări au fost esențiale pentru a oferi un context mai detaliat în timpul analizei.

Pentru analiză am utilizat RStudio 2023.06.2 Build 56[1] și librariile tidyverse, dplyr, ggplot2, gridExtra.

**Analiza Datelor**

Cu setul de date pregătit, am investigat distribuțiile, relațiile și tendințele relevante. Am căutat să identific informații semnificative despre impactul locului de desfășurare asupra rezultatelor meciurilor de fotbal.

Am investigat dacă locul desfășurării meciului (pe teren propriu sau pe teren neutru) influențează rezultatul. Analiza mea s-a concentrat pe distribuția victoriilor, pierderilor și egalităților pentru echipele gazdă în aceste două contexte distincte.

Am examinat cum diferitele tipuri de turnee pot afecta rezultatele meciurilor. Am inclus în analiză diverse competiții și calificări, investigând dacă participarea la anumite evenimente poate fi asociată cu performanțe semnificative.

Am evaluat coeficienții de regresie ai variabilelor mele independente, cum ar fi terenul și tipul de turneu. Am identificat contribuțiile semnificative ale acestor variabile la predicția rezultatului meciurilor.

**Metode Statistice și Analize**

Pentru a anticipa rezultatul unui meci, am aplicat regresia logistică folosind variabila home\_results ca variabilă dependentă și alte caracteristici relevante (cum ar fi terenul neutru și tipul de turneu) drept variabile independente. Variabilele independente utilizate în regresia logistică includ informații precum dacă meciul s-a desfășurat pe teren neutru și tipul de turneu (exemplu: Calificari Campionatul Mondial, Cupa Națiunilor Africane).

Pentru a investiga influența terenului asupra rezultatelor meciurilor de fotbal, am utilizat un tabel de contingenta pentru a prezenta distribuția frecvențelor a trei categorii principale: victorii, pierderi și egalități ale echipei gazde. Am luat în considerare atât meciurile jucate pe teren propriu, cât și cele desfășurate pe teren neutru.

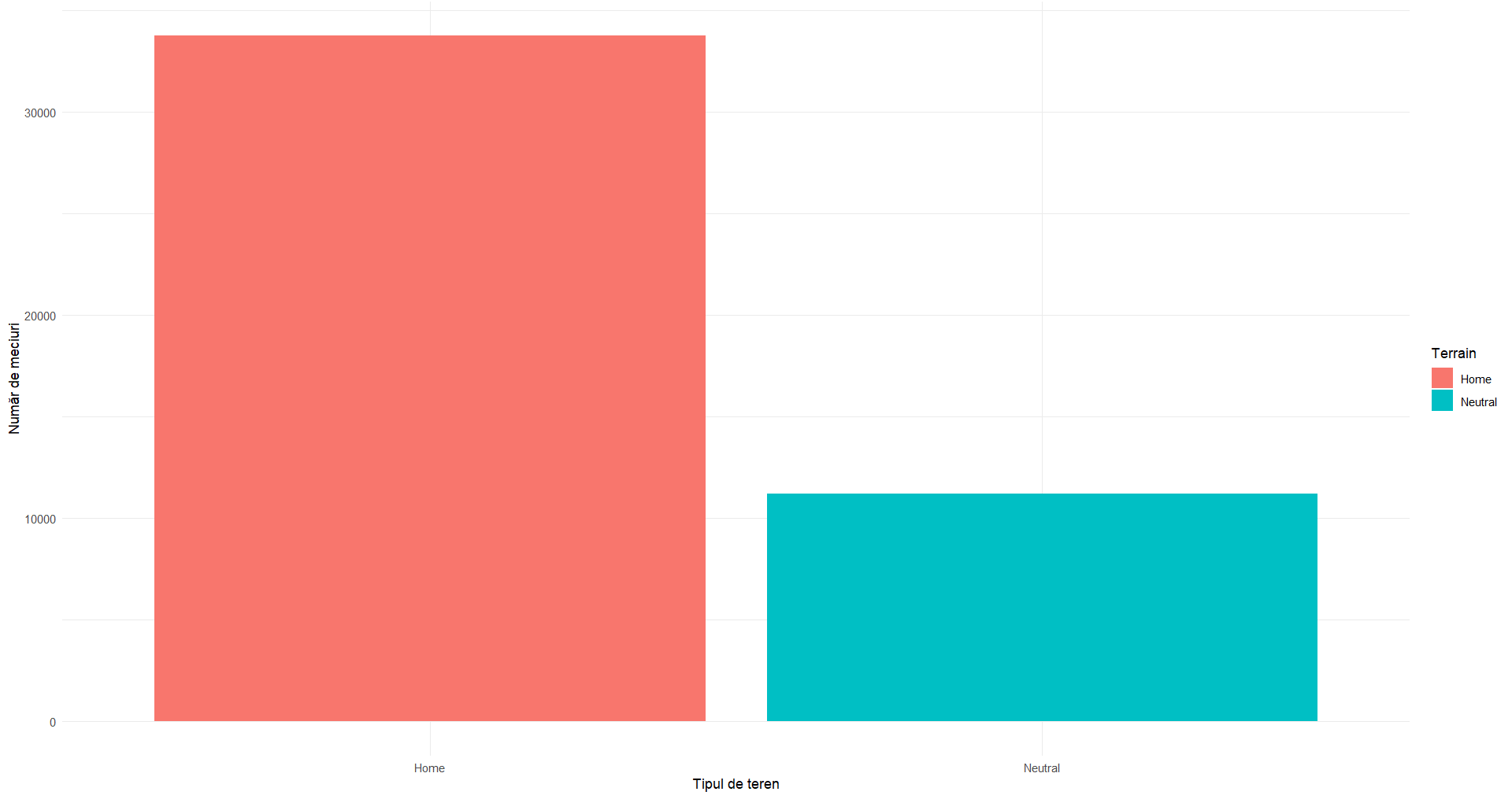
**Disponibilitatea Datelor și a Codului**

Codul este disponibil pe GitHub, pe linkul:

<https://github.com/Deo777/Analiza_Datelor>

**Rezultate**

Graficul din figura 1.1 prezintă distribuția numărului de meciuri jucate pe teren propriu și pe teren neutru în setul de date. Observăm că majoritatea meciurilor au fost jucate pe teren propriu, reprezentând 75.1% din totalul de meciuri, în timp ce meciurile pe teren neutru reprezintă 24.9%.



**Figura 1.1** Numărul de meciuri jucate în baza tipului de teren

În tabelul 1.1 este prezentată distribuția rezultatelor pentru meciurile jucate de echipele gazdă, desfășurate pe teren propriu și pe teren neutru. Datele din tabel reflectă numărul de victorii, pierderi și egalități înregistrate de echipele gazdă în cele două contexte.

|  |  |  |
| --- | --- | --- |
|  | Teren Propriu | Teren Neutru |
| Nr. de victorii echipa gazda | 17064 | 4818 |
| Nr. de pierderi echipa gazda | 8893 | 3823 |
| Nr. de egalitati echipa gazda | 7782 | 2554 |

**Tabelul 1.1** Rezultatele meciurilor echipei gazde

Teren Propriu:

Numărul de victorii ale echipei gazdă: 17,064 de victorii. Acest număr înalt sugerează că echipele gazdă are o performanță mai bună când joacă pe teren propriu.

Numărul de înfrângeri ale echipei gazdă: 8,893 de înfrângeri. Deși este un număr considerabil, este mai mic decât numărul de victorii, confirmând avantajul general al echipei gazdă.

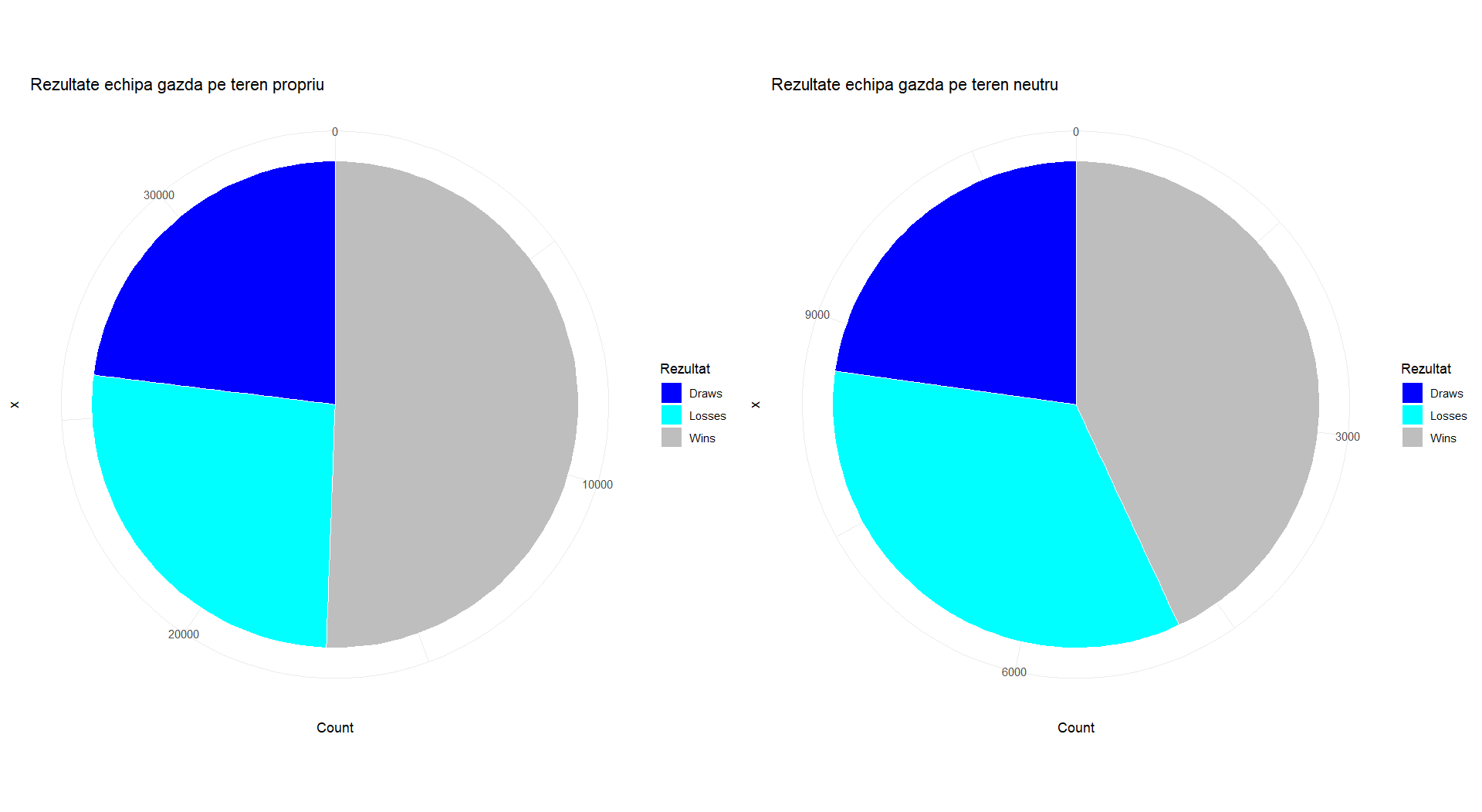
Numărul de egalități ale echipei gazdă: 7,782 de egalități. Echipele gazdă au avut și un număr semnificativ de rezultate de egalitate.

Teren Neutru:

Numărul de victorii ale echipei gazdă: 4,818 de victorii. Aici, numărul de victorii este semnificativ mai mic decât pe teren propriu, indicând că terenul neutru poate reduce avantajul echipei gazdă.

Numărul de înfrângeri ale echipei gazdă: 3,823 de înfrângeri. Similar cu terenul propriu, numărul de înfrângeri este mai mic decât numărul de victorii.

Numărul de egalități ale echipei gazdă: 2,554 de egalități. Echipele gazdă au înregistrat, de asemenea, un număr semnificativ de rezultate de egalitate pe teren neutru.



**Figura 1.2** Procentul Rezultatelor meciurilor echipei gazde

În figura 1.2 sunt prezentate rezultate meciurilor echipei gazde pe teren propriu și pe teren neutru, din tabelul 1.1, și noi putem observa că echipa gazdă are o performanță generală bună pe teren propriu comparativ cu pereformanța pe teren neutru.

|  |  |  |
| --- | --- | --- |
| Turneul | Rezultatul | Frecvența |
| AFC Asian Cup qualification | Away | 124 |
| African Cup of Nations qualification | Away | 373 |
| British Home Championship | Away | 174 |
| FIFA World Cup qualification | Away | 1925 |
| Friendly | Away | 4085 |
| UEFA Euro qualification | Away | 842 |
| UEFA Nations League | Away | 146 |

**Tabelul 1.2** Frecvența de câștig a echipei oaspete în cele mai mari turnee

|  |  |  |
| --- | --- | --- |
| Turneul | Rezultatul | Frecvența |
| African Cup of Nations qualification | Draw | 469 |
| British Home Championship | Draw | 108 |
| FIFA World Cup qualification | Draw | 1537 |
| Friendly | Draw | 3808 |
| UEFA Euro qualification | Draw | 524 |
| UEFA Nations League | Draw | 109 |

**Tabelul 1.3** Frecvența de egalitate în cele mai mari turnee

|  |  |  |
| --- | --- | --- |
| Turneul | Rezultatul | Frecvența |
| AFC Asian Cup qualification | Home | 291 |
| African Cup of Nations qualification | Home | 1083 |
| British Home Championship | Home | 233 |
| CECAFA Cup | Home | 102 |
| CFU Caribbean Cup qualification | Home | 213 |
| CONCACAF Nations League | Home | 103 |
| Copa América | Home | 177 |
| FIFA World Cup qualification | Home | 3771 |
| Friendly | Home | 7357 |
| Nordic Championship | Home | 135 |
| UEFA Euro qualification | Home | 1313 |
| UEFA Nations League | Home | 198 |

**Tabelul 1.4** Frecvența de câștig a echipei gazde în cele mai mari turnee

Rezultatele din tabelele de contingență prezentate mai sus indică distribuția frecvenței victoriilor, remizelor și înfrângerilor pentru echipele gazdă în diferite turnee.

În tabelul 1.2 sunt prezentate frecvețele câștigurilor a echipelor oaspete în cele mai mari turnee (Away):

Friendly: Cele mai multe meciuri au avut loc în condiții prietenoase, cu un număr semnificativ de victorii (4085), urmate de meciuri în care s-a înregistrat o remiză (3808). Aceasta indică că meciurile amicale sunt relativ echilibrate în ceea ce privește rezultatele.

FIFA World Cup qualification: Aici, vedem o frecvență semnificativă a meciurilor de calificare la Cupa Mondială, cu un număr impresionant de victorii în deplasare (1925).

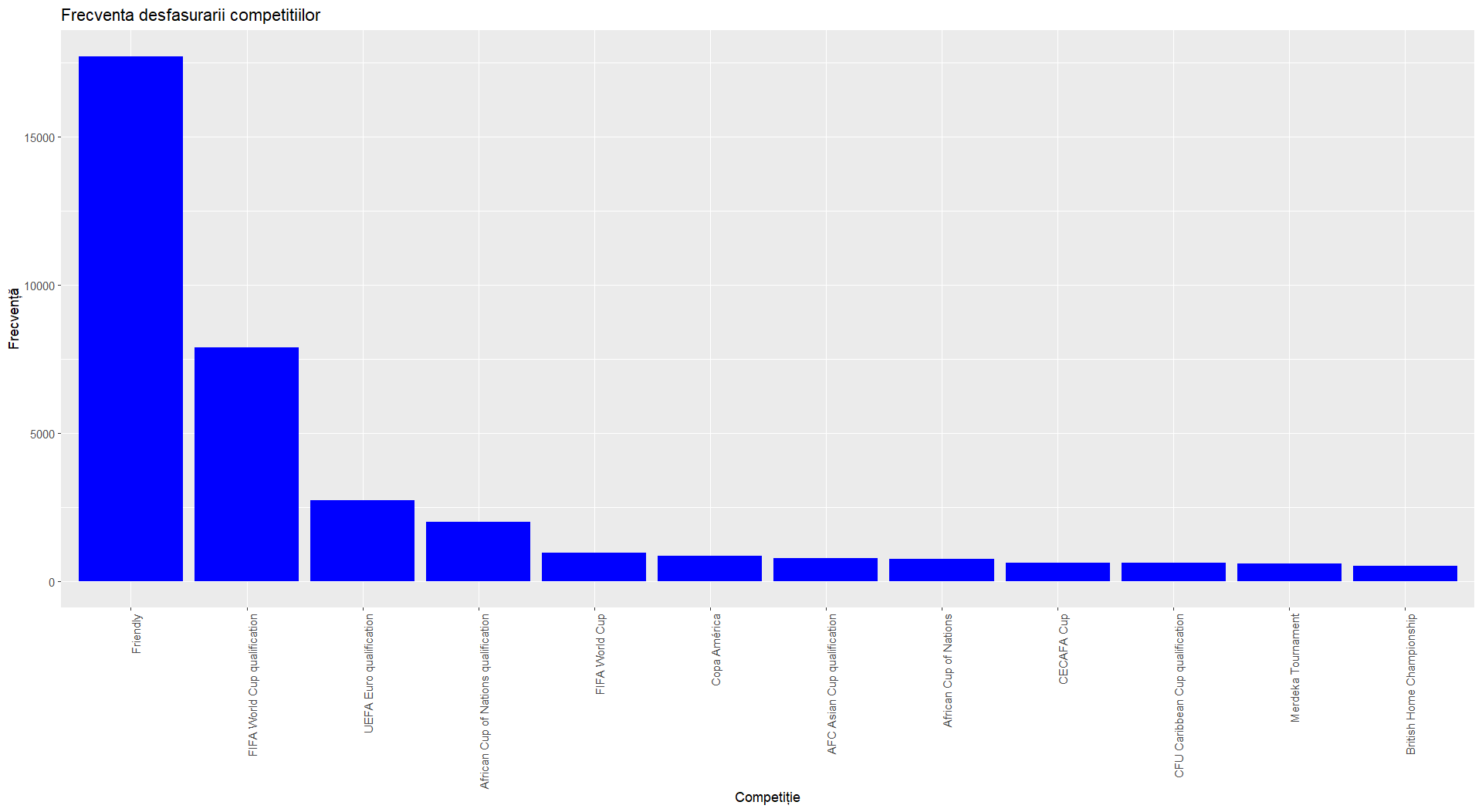
În tabelul 1.3 sunt prezentate frecvețele egalităților în cele mai mari turnee (Draw):

Friendly și FIFA World Cup qualification: Meciurile amicale și cele de calificare la Cupa Mondială au înregistrat cele mai mari frecvențe pentru remize, ceea ce sugerează că aceste competiții pot avea rezultate mai imprevizibile.

În tabelul 1.4 sunt prezentate frecvețele câștigurilor a echipelor gazde în cele mai mari turnee (Home):

Friendly și FIFA World Cup qualification: Meciurile amicale și cele de calificare la Cupa Mondială au cel mai mare număr de victorii pentru echipele gazdă (7357 și 3771, respectiv), indicând un avantaj semnificativ pentru echipele care joacă pe teren propriu în aceste competiții.

UEFA Nations League: Deși numărul total de meciuri este mai mic, această competiție arată, de asemenea, o frecvență semnificativă a victoriilor pentru echipele gazdă.

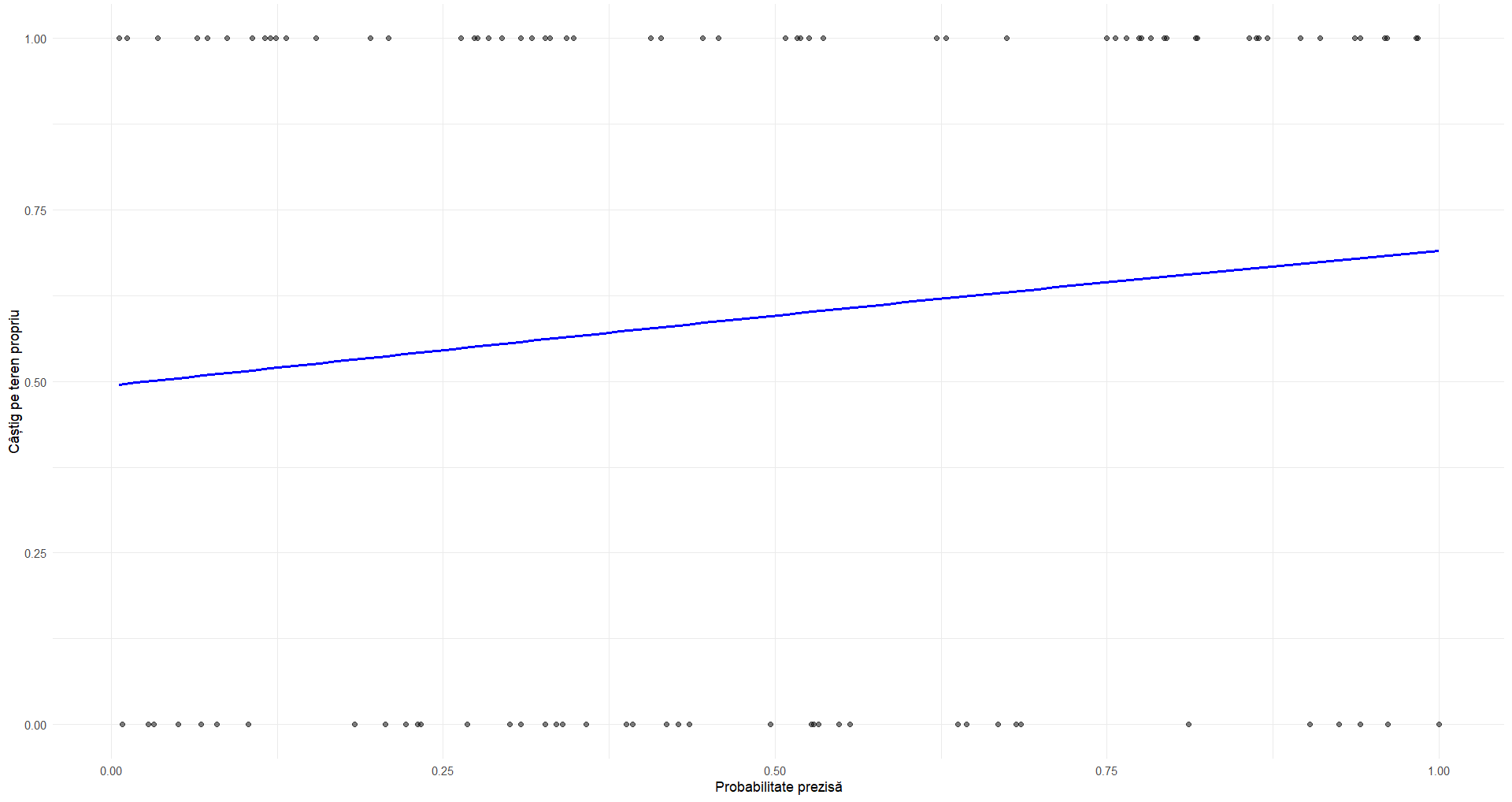


**Figura 1.3** Frecvența desfășurării competițiilor

Figura 1.3 prezintă frecvențele meciurilor pentru diferite turnee. Cea mai mare frecvență o au meciurile amicale, cu peste 15.000 de meciuri disputate. În al doilea rând se află meciurile de calificare pentru Cupa Mondială FIFA, cu puțin peste 7.500 de meciuri. Pe locul al treilea se află meciurile de calificare pentru Campionatul European UEFA, cu puțin peste 2.500 de meciuri.

În urma aplicării regresiei logistice pe setul tău de date privind fotbalul, am investigat modul în care diverse caracteristici influențează rezultatul unui meci, având ca variabilă dependentă rezultatul echipei gazdă (home\_results). Principalele caracteristici luate în considerare au fost prezența meciului pe teren neutru (neutral), tipul de turneu (tournament), și dacă echipa gazdă este favorizată prin faptul că meciul se desfășoară pe teren propriu (home\_team).

Rezultatele analizei au arătat că există o corelație semnificativă între variabilele independente menționate și rezultatul meciului. Coeficienții obținuți din modelul de regresie logistică oferă informații despre direcția și magnitudinea influenței fiecărei variabile asupra șanselor echipei gazdă de a obține un rezultat pozitiv.



**Figura 1.4**  Probabilitatea de câștig pe teren propriu

În figura 1.4 este o reprezentare grafică a relației dintre probabilitatea prezisă și variabila dependentă home\_results care indică dacă echipa gazdă a câștigat sau nu pe teren propriu.

|  |  |  |
| --- | --- | --- |
|  | Clasa Prezisă | |
| Clasa Actuală | 0 | 1 |
| 0 | 2835 | 1788 |
| 1 | 1784 | 2579 |

**Tabelul 1.5** Matricea de confuzie

În tabelul 1.5 este prezentată matricea de confuzie pentru modelul de predictie creat folosind regresia logistic.

* True Negative (TN): 2835
* False Positive (FP): 1788
* False Negative (FN): 1784
* True Positive (TP): 2579

Aceste valori pot fi interpretate astfel:

* Modelul a identificat corect 2835 de cazuri negative.
* Modelul a clasificat incorect 1788 de cazuri negative ca fiind pozitive.
* Modelul a clasificat incorect 1784 de cazuri pozitive ca fiind negative.
* Modelul a identificat corect 2579 de cazuri pozitive.

Acuratețea modelului:

Acuratețea modelului este o măsură a cât de bine modelul face predicții corecte pe datele de test. În acest caz, acuratețea modelului este aproximativ 60.25%, ceea ce înseamnă că modelul a prezis corect rezultatele în aproximativ 60% din cazuri.

**Bibliografie**

1. R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL [https://www.R-project.org/](https://www.r-project.org/). – accesat (12.09.2023)
2. Dataset: [International football results from 1872 to 2023](https://www.kaggle.com/datasets/martj42/international-football-results-from-1872-to-2017?select=results.csv) - sursa setului de date utilizat în analiză, disponibil pe Kaggle. – accesat (17.09.2023)
3. Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer. Disponibil la: [https://ggplot2.tidyverse.org](https://ggplot2.tidyverse.org/). – accesat(14.09.2023)
4. Wickham, H., și alții. (2019). *tidyverse: Easily Install and Load the 'Tidyverse'*. Disponibil la: [https://www.tidyverse.org](https://www.tidyverse.org/). – accesat(14.09.2023)
5. OpenIntro Statistics. (2019). [openintro: OpenIntro data sets and supplemental functions](https://cran.r-project.org/package=openintro). – accesat (04.11.2023)

# Materiale Suplimentare

Anexa 1. Variabilele setului de date

1. **date** - data meciului
2. **home\_team** - numele echipei gazdă
3. **away\_team** - numele echipei oaspete
4. **home\_score** - scorul echipei gazdă la sfârșitul timpului regulamentar, inclusiv timpul suplimentar, fără penalty-uri
5. **away\_score** - scorul echipei oaspete la sfârșitul timpului regulamentar, inclusiv timpul suplimentar, fără penalty-uri
6. **tournament** - numele turneului
7. **city** - numele orașului/localității/unității administrative unde s-a jucat meciul
8. **country** - numele țării unde s-a jucat meciul
9. **neutral** - coloană TRUE/FALSE care indică dacă meciul s-a jucat pe un teren neutru
10. **winner** - numele echipei câștigătoare
11. **winner\_categ** – categoria rezultatului (Draw, Home, Away)