

DOCUMENTAȚIE SISTEME INTELIGENTE

ANALIZĂ STATISTICI NBA ȘI PREDICȚIE MVP

STUDENT : TORZ IULIAN MARIUS
SPECIALIZARE : CALCULATOARE 3

Contents

1. INTRODUCERE	3
1.1. Descriere generala	3
2. CONTEXT	4
2.1. Organizare	4
3. STATE OF THE ART	5
3.1. Alegere MVP curenta	5
3.2. Bias-uri	5
3.3. Abordarea problemei	6
4. IMPLEMENTARE	7
4.1. Analiza, curatare	7
4.2. Corelatii	8
4.3. Entropie	9
4.4. Modele	10
5. TESTARE	11
6. REZULTATE	12
6.1. Cat de mult afecteaza punctele un MVP?	12
6.2. Poate un jucator egoist sa castige MVP-ul?	13
6.3. Ce importanta are defensiva (rebounds) in alegerea MVP-ului?	13
6.4. Poate sa castige un jucator inefficient MVP-ul?	13
6.5. Care sunt statisticile minime?	14
7. CONCLUZII	14
BIBLIOGRAPHY	15

1. Introducere

1.1. Descriere generala

Scopul acestui proiect este de a dezvolta un model de machine learning care să poată prezice dacă un jucător din NBA a avut o performanță care merită premiul MVP (Most Valuable Player). Modelul va fi antrenat pe un set de date cuprinzând statistici ale jucătorilor din istoria NBA, luând în considerare factori precum punctele marcate, recuperările, pasele decisive, procentajul la aruncări, procentajul la libere, blocările, furturile și alte statistici relevante

Dataset-ul ales de mine a fost preluat de pe [kaggle](https://www.kaggle.com), datele sunt în format CSV , și conține toate datele de care am putea avea nevoie pentru a antrena modelul.

Din datele existente nu ne sunt folositoare chiar toate, multe dintre ele sunt redundante pentru ca exprima mai mult sau mai puțin aceleași statistici, spre exemplu aruncările de 2 puncte : 2P(aruncări de 2 puncte reușite), 2PA(aruncări de 2 puncte încercate), 2P%(procentul de aruncări reușite din totalul de aruncări), aceelasi lucru se întâmplă și cu aruncările de 3 puncte și aruncările libere (free throws). Toate aceste statistici referitoare la aruncari se pot concentra într-o singura statistica numita eFG% (Effective Field Goal Percentage) este o statistică care ajustează procentul aruncărilor pentru a ține cont de faptul că aruncările de trei puncte contează pentru trei puncte, în timp ce toate celelalte goluri de teren contează doar pentru două puncte.Scopul sau este da a masura eficienta arucatorilor relativ la aruncările pe care le încearcă. Dacă un jucător arunca mai des de la 3 puncte acesta poate să aibă o eficienta mai buna decat un jucător care arunca mai des de la 2 puncte chiar dacă numărul de aruncări reușite este similar.

$$eFG\% = \frac{FG + 0.5 * 3P}{FGA}$$

Figure 1. Formula eFG%

Luand toate aceste aspecte în considerare rezultatul final al proiectului este un model de machine learning care poate sa prezica cu o acuratețe destul de ridicată cum o sa prezint în capitolele următoare dacă un jucător a avut o performanță demnă de MVP pentru un anumit sezon sau nu.

Mandić Radivoj et al. au realizat o analiza a statisticilor colectate din NBA si Euroleague de-a lungul a 17 ani și aceștia au descoperit anumite statisticile interesante precum numărul aruncarilor de pedeapsa (free throws) este mai mare în Euroleague dar acest număr este în declin, deci se așteaptă ca în următorii ani sa se ajungă la o anumită egalitate, un alt fapt interesant este numărul de posesii, in NBA se joacă un mediu mult mai multe posesii pe meci față de Euroleague. [1]

Yves de Saá Guerra et al. au ajuns la concluzia ca competitivitatea în anii curenți este la un nivel mai ridicat fata de anii din trecut , după cercetarea lor, sezoanele din anii 90' și începutul anilor 2000 au fost cele mai puțin competitive din toata istoria NBA. [2]

2. Context

2.1.Organizare

Cum am enuntat și în introducere scopul proiectului este analiza statisticilor unui jucător de-a lungul sezonului, un sezon normal din NBA are 82 de meciuri (în pandemie sezonul a fost redus la 72 meciuri) excluzand meciurile din play offs deoarece premiul de MVP se acorda doar pentru performanța din sezonul regular.

Determinarea MVP-ului in viata reala nu este bazată strict pe statisticile jucatorilor dar și pe forma echipei în general, ultimul jucător care a câștigat titlul de MVP fără ca echipa sa să se califice în play-off a fost Kareem Abdul Jabbar în sezonul 1975-76, echipă având un W/L de 40-42.Există o varietate de statistici care sunt utilizate pentru a evalua performanța echipelor, cum ar fi numărul de puncte marcate pe meci, numărul de recuperări pe meci, numărul de pase decisive pe meci, procentajul la aruncări, procentajul la libere, blocările și furturile.

Kuan-Chieh Wang și Richard Zemel au folosit un model de machine learning pentru a clasifica tipurile de strategii pe care echipele le trateaza in ofensiva lor, aceștia au luat în calcul schimbarea sezoanelor și chiar și eventualele transferuri ale jucătorilor de la o echipa la alta. Datele folosite au fost preluate din sezonul 2013-2014 al echipei Toronto Raptors și s-a decis asupra a 11 clase de ofensiva, folosind un neural network aceștia au atins o precizie de 65.6% în funcție jucătorii care sunt pe teren, timpul din repriza la care se afla, daca mingea porneste dintr-o aruncare de la margine, dintr-un rebound sau daca echipa cealalta a marcat. [3]

2.2.Premii

NBA este formată din două conferințe: Conferința de Est și Conferința de Vest. Fiecare conferință este formată din trei divizii: Atlantic, Centrală și Sud-Estică și Pacifică, Mountain și Southwest.Fiecare divizie are cinci echipe, pentru un total de 30 de echipe în ligă. Alte premii care se acorda la sfarsitul sezonului sunt Rookie of the Year este acordat celui mai bun jucător nou din liga în acel sezon, Defensive Player of the Year este acordat celui mai bun jucător defensiv din liga în acel sezon,Sixth Man of the Year este acordat celui mai bun jucător care nu este titular din liga în acel sezon.

Un alt aspect care afecteaza decizia finală este importantă jucătorului în echipa, cand jucatorul este cu un nivel superior peste colegii sai și performanța acestuia în momentele cruciale ale meciurilor, reziliență și motivatia de a castiga, uneori și story line-ul din spatele unui jucător poate afecta decizia, spre exemplu revenirea din accidentare a unui jucător placut

de public (Kawhi Leonard) sau un potențial sezon de retragere pentru un veteran (Lebron James)

3. State of the art

3.1. Alegere MVP curentă

În prezent pentru alegerea MVP-ului ligii se iau în considerare următoarele lucruri: Jucătorii trebuie să fi jucat cel puțin 65 de meciuri în sezonul regulat. (modificare de la 58 la 65 din sezonul 2023/2024)

Jucătorii nu pot fi suspendați pentru o perioadă semnificativă de timp din cauza încălcării regulilor ligii. După ce se ating aceste criterii jucătorii cu cele mai bune statistici din echipele care au un record pozitiv în liga ($\text{winrate} > 50\%$), de obicei primii 10. Apoi are loc procesul de votare, un grup de 100 de jurnaliști sportivi votează pentru MVP.

Fiecare jurnaliști își selectează primii zece jucători, cu 10 puncte acordate pentru primul loc, 7 puncte pentru locul al doilea, 5 puncte pentru locul al treilea și așa mai departe.

Jucătorul cu cele mai multe puncte câștigă premiul MVP de aceea mult considera ca este un lucru imparțial și pot apărea bias-uri ca și recency bias, voters fatigue sau chiar playstyle bias.

Alegea MVP-ului este influentata în mod direct de către formarea echipelor All-Stars, aceste echipe sunt formate doar din voturile publicului dar în general doar cei mai bun jucatori din liga ajung în aceste echipe, influența selecției All-Star a fost studiata de catre Albert Alberto Arteta et al. testand modelul pe 115 de inputuri acestea au primit o acuratețe de 87.3%, pentru a ajunge la aceasta acuratețe impresionantă aceștia au folosit hibrid, folosindu-se de MLP, Adaboost și Random forest. [4]

3.2. Biasuri

Recency bias - Voturile pentru MVP sunt exprimate la finalul sezonului, imediat după ultimele meciuri. Performanțele puternice ale unui jucător în ultima etapă pot fi proaspete în mintea alegătorilor, depășind potențial performanțele sale din restul sezonului. Amintirea performanței unui jucător pe parcursul unui sezon de 82 de meciuri poate fi dificilă. Meciurile recente sunt pur și simplu mai ușor de reținut, ceea ce poate duce la subestimarea unui jucător care ar fi putut fi constant bun pe tot parcursul anului, dar nu a avut un final deosebit de puternic.

Voters fatigue bias - Votarea pentru același jucător ca MVP mai mulți ani la rând poate duce la un impuls inconștient de a căuta un nou candidat "demn". Alegătorii pot fi tentați să recompenseze jucători noi care au izbucnit recent, chiar dacă performanța lor generală nu este la fel de puternică ca a MVP-ului în exercițiu. Poate duce la o rotație mai

mare a câștigătorilor MVP, chiar dacă unii jucători își mențin un nivel dominant de performanță. Aceasta poate duce la o apreciere mai puțin consecventă a excelenței din ligă. Chiar dacă este incert dacă acesta este un bias real, eu personal consider ca a avut loc chiar în sezonul trecut 2022/2023, Nikola Jokic a câștigat MVP-ul doi ani la rand 2020/2021, 2021/2022, el este considerat ca fiind cel mai dominant și talentat jucător din ligă la ora actuală, chiar a câștigat al treilea sau MVP în 4 ani, dar să ne concentrăm pe sezonul 2022/2023, când a pierdut acest titlu în favoarea lui Joel Embiid. În acel sezon o statistică numită Box Plus Minus care concentrează toate statisticile unui jucător într-un singur număr arată ca performanța lui Jokic a fost mai bună decât cea a lui Embiid (640 vs 424 potrivit [statmuse](#)).

Playstyle bias - Alegătorii pot fi atrași de anumite stiluri de joc sau de jucători cu anumite caracteristici, favorizându-i pe aceștia în detrimentul altora cu performanțe similare, dar care nu sunt la fel de “flashy”. Un jucător care joacă baschet fundamental corect și eficient poate fi mai plictisitor față de unul care încearcă aruncări spectaculoase de la distanță cu șansa mică de reușită sau care performează multe dunk-uri, alley-oops.

3.3. Abordarea problemei

Să presupunem că avem o problemă tipică de clasificare, în care avem un set de puncte în spațiu și fiecărui punct i se atribuie o etichetă de clasă (MVP sau nu). Dacă o linie dreaptă poate împărți cele două clase, atunci avem o problemă separabilă liniară. Pe de altă parte, dacă o linie dreaptă nu este suficientă pentru a împărți cele două clase, atunci avem o problemă separabilă neliniară. Figura de mai jos prezintă date în spațiul bidimensional. Fiecărui punct i se atribuie o etichetă de clasă roșie sau albastră. Figura din stânga arată o problemă separabilă liniară care necesită o limită liniară pentru a distinge cele două clase. Figura din dreapta arată o problemă separabilă neliniară, unde este necesară o limită de decizie neliniară. [5]

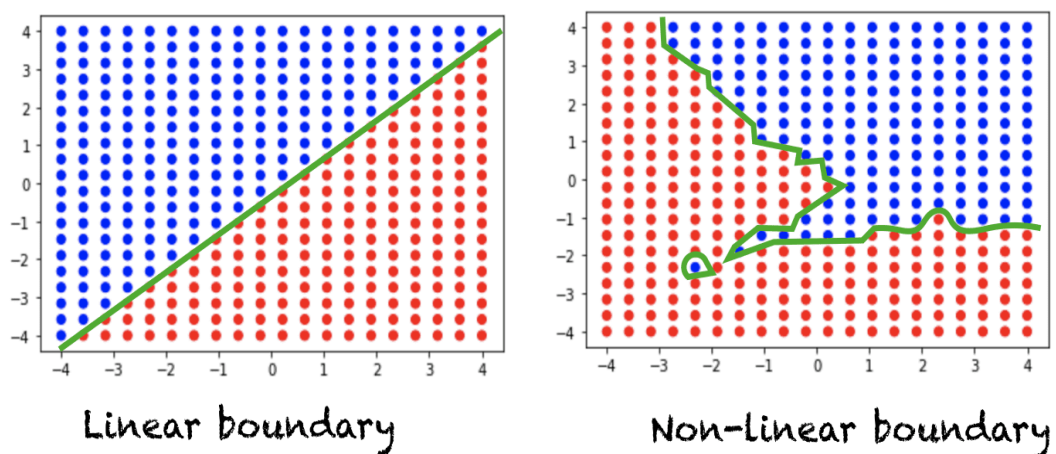


Figure 2. Comparare Liniar vs Non-liniar

O alta problema care a fost abordată în contextul de machine learning a fost predicția rezultatelor meciurilor, Fadi Thabtah et al. au investigat aceasta categorie în lucrarea lor, iar rezultatele cele mai bune le-au avut folosind ANN (Artificial Neural Network), rezultatele lor au depășit precizia de 73.4% până atunci considerată cea mai bună posibilă. [6]

Tot problema predicției rezultatelor meciurilor a fost studiată și de către Ge Cheng et al. de aceasta data ei au pornit tratat modelul ca o problemă de clasificare, aceasta fiind antrenat cu principiul entropiei maxime, numit de ei NBAME. Folosind această strategie ei au atins o acuratețe de 74.4%, puțin mai ridicată decât alte modele din vremea aceea. [7]

4. Implementare

4.1. Analiza, curatare

Pentru implementarea acestui proiect am folosit Python 3.9, librăriile cu ajutorul cărora am dezvoltat proiectul sunt pandas, numpy, pentru prelucrarea dataframe-ului, sklearn pentru accesul ușor la modele și antrenarea lor, matplotlib și seaborn pentru grafice și heatmap. corelații, descriere curatare

Am început prin curatarea dataset-ului, începând cu verificarea valorilor lipsa, am observat că valorile lipsa apar la procentajele aruncărilor jucătorilor care nu au foarte multe meciuri și/sau minute jucate. Am eliminat jucătorii cu mai puțin de 2 meciuri jucate sau mai puțin de 2 minute jucate în average. Dacă au mai rămas jucători cu valori lipsa am înlocuit cu 0-uri Toate campurile lipsa sunt procente, sunt missing values deoarece jucătorul nu are nici o aruncare din categoria respectivă).

Unii jucători aveau caracterul "*" după numele lor, documentația nu spunea ce înseamnă acest lucru dar prin diferite asocieri mi-am dat seama că se referă la faptul că acei jucători fac parte din [Hall of Fame](#) (premiu acordat jucătorilor legendari după retragere), am eliminat acest aspect din numele jucătorilor.

Iar ultimul lucru pe care l-am făcut pentru a modifica baza de date este transformarea/binarizarea coloanei de MVP, aceasta conținea valori true și false iar eu le-am transformat în 1 și 0.

4.2. Corelații

Analizând matricea de corelații putem observa o mulțime de legături între diferite statistici, unele care le-am putea prezice, altele nu așa mult. De exemplu ne-am așteptat să fie o corelație foarte mare dintre aruncări încercate, aruncări reușite și procentajul de succes al aruncărilor cum se poate observa și în imagine (0.9), între aruncări și puncte (0.98) sau între

rebounds totale, ofensive și defensive (0.88) dar și unele corelații la care nu ne-am fi gândit precum corelataia dintre punctele marcate și turnovers (atunci cand un jucător pierde mingea inofensiva) avem o corelație de 0.83. O explicație pe care eu as sugera-o este ca deoarece jucătorii care înscriu cel mai mult sunt predispuși sa se piardă mingea de mai multe ori, ei fiind cei care coordonează atacul și au posesia mingea în majoritatea timpului, în mod teoretic. În trecut poziția responsabilă pentru coordonarea atacului era Point Guard (PG) dar odată cu schimbarea timpului și trecerea generațiilor, rolurile jucatorilor s-au diluat, in era noastra jucătorul cu cele mai bune statistice ajunge sa conduca mingea și implicit partea ofensiva a echipei. Field goals attempted are o corelație mare cu toate statisticile legate de aruncări, fie ele de 2 puncte, 3 puncte sau aruncari libere.

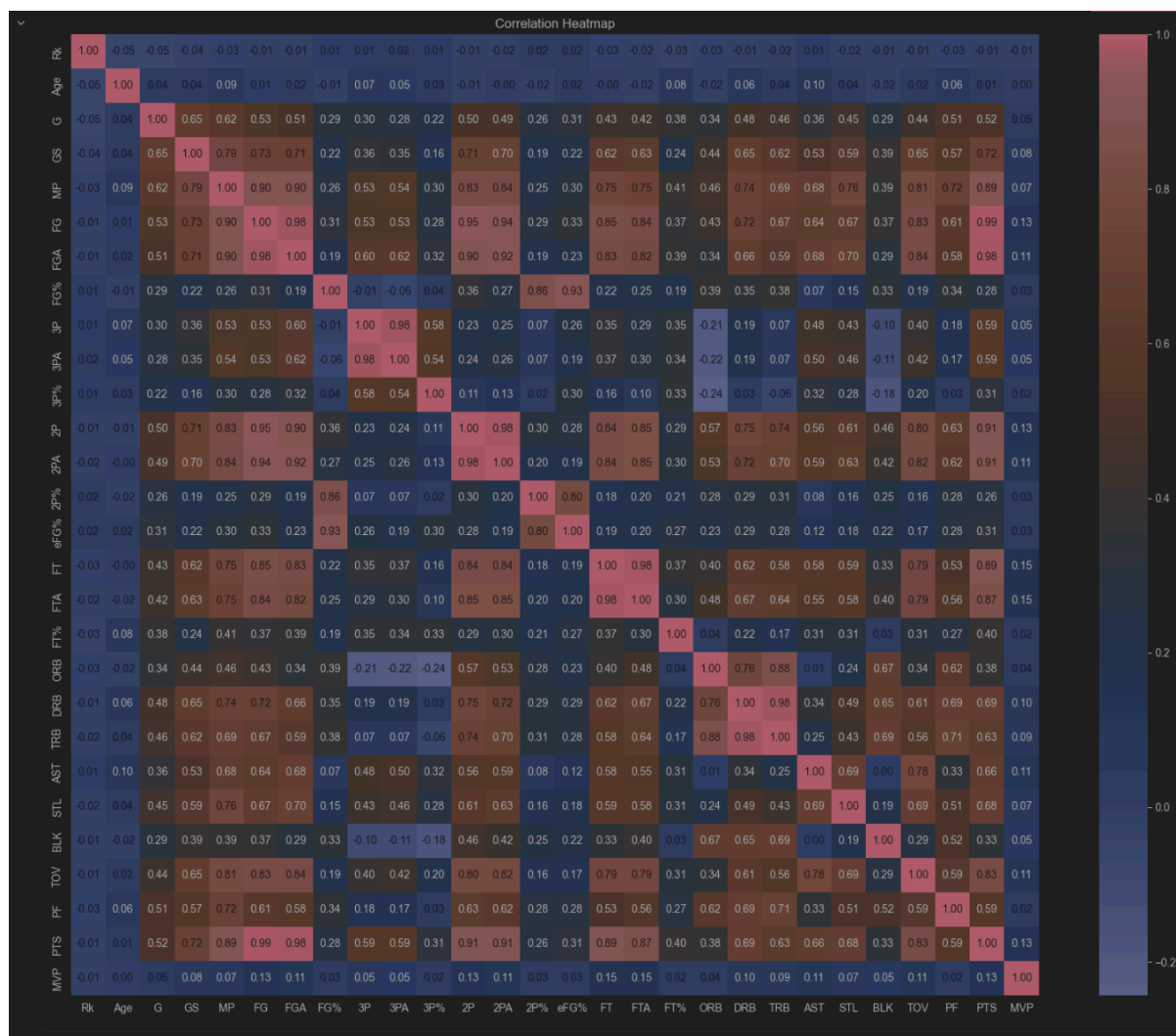


Figure 3. Matricea de corelații

4.3.Entropie

Entropia a fost măsurată pentru toate coloanele numerice, folosind formula:

$$H = \sum P(x) * \log \log \frac{1}{P(x)}$$

Figure 4. Formula entropie

Aceasta informatie de aduce cunoștințe cu privire la informația pe care o putem obține din o anumită coloana, cu cat entalpia este mai mare cu atat informația din aceea coloana variază. Entropia cea mai mare este prezentă în coloanele următoare:

MP (minutes played): 8.51
FT% (free throw %): 8.23
eFG%(effective field goal %): 8.09

4.4.Modele

Modelele folosite au fost importate din biblioteca sklearn, am folosit unele dintre cele mai populare modele și am comparat rezultatele. Cel mai mare acuratețe am obtinut-o folosind Logistic Regression (99.86%). De aceea am decis sa folosești acest model că modelul meu final. Regresia logistică este o metodă statistică utilizată pentru problemele de clasificare binară, ceea ce înseamnă că prezice probabilitatea ca o intrare dată să aparțină uneia dintre cele două clase. Spre deosebire de regresia liniară, care prezice o valoare continuă, regresia logistică prezice o probabilitate că o anumită mostră aparține unei anumite clase. Rezultatul ei este o functie sigmoida, ieșirea funcției sigmoide este o probabilitate care poate fi programată pentru a lua o decizie binară. De exemplu, dacă ieșirea este mai mare de 0.5(threshold), modelul prezice clasa pozitivă; altfel, prezice clasa negativă.(în cazul nostru dacă statisticile caracterizează un jucător care ar putea sau nu sa fie MVP). [8]

Regresia logistică este un proces de modelare a probabilității unui rezultat discret având în vedere o variabilă de intrare. Cele mai frecvente regresii logistice modelează un rezultat binar; ceva care poate lua două valori, cum ar fi adevărat/fals, da/nu și așa mai departe, în cazul nostru MVP sau nu.

Todd G. Nick și Kathleen M Campbell au folosit regresia logica în medicina și spun ca ca atunci când există o singură variabilă predictivă într-un model de regresie logistică, modelul este denumit regresie logistică simplă(dacă o boala exista sau nu în corp). Atunci când există

mai mulți predictorii (de exemplu, factorii de risc și tratamente), incluzând atât variabile categoricale, cât și variabile continue ca predictorii, modelul este denumit regresie logistică multiplă sau multivariabilă. [9]

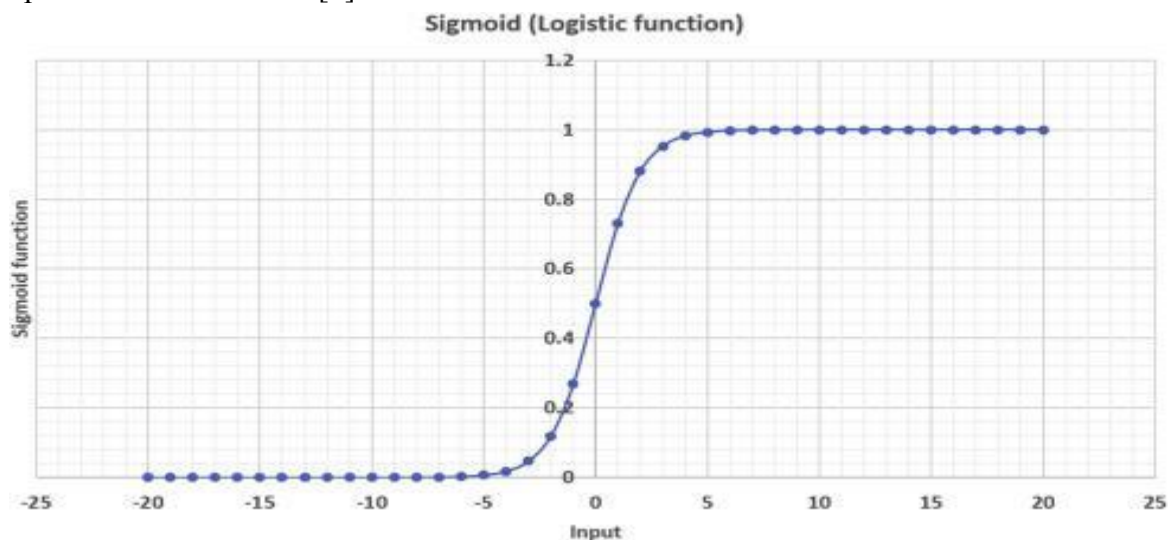


Figure 5. Logistic Regression

4.5. Ce se întâmplă dacă modificăm valoarea threshold-ului?

Atunci modelul devine mai conservativ cu predicția de MVP, consecința acestui aspect este că mai puțini jucători vor fi prezis ca MVP, deci va apărea o scădere în cazurile true positives și o posibilă creștere a cazurilor false negative, de aceea acuratețea poate să scadă dacă dataset-ul este balansat din punct de vedere al numărului de MVPs, un număr semnificativ mai mare de jucători nu au premium de MVP, doar o selecție mică sunt MVPs. (25 MVPs din 14573).

Pentru antrenarea modelelor dataset-ul a fost împărțit în 2 părți, 80% pentru antrenare și 20% pentru testare. Principalul motiv pentru care împărțim setul de date este să evaluăm modelul pe date nevăzute. Acest lucru ajută la simularea modului în care modelul va performa în scenarii reale unde întâlnește date noi. Un alt motiv pentru care se face acest “split” are legătura cu suprainvatarea, dacă antrenăm modelul pe un subset și îl testăm pe altul putem evita suprainvatarea datelor.

Chiar dacă am menționat folosirea unui threshold, acesta nu a adus îmbunătățiri pentru acuratețea modelului, a fost mai mult un aspect experimental în cadrul proiectului, de aceea în testarea cu datele propuse de mine pentru judecarea dacă un jucător a avut o performanță demnă pentru MVP am decis să nu folosesc acel threshold.

5. Testare

Am testat mai multe modele, cu diferite coloane, mai întâi am antrenat modelul cu toate coloanele care țin de statisticile unui jucător, excluzând doar, numele, poziția, echipa și sezonul, acuratețea acestor modele a fost ușor mai mică și mai variată (diferența mai mare de acuratețe între cel mai bun model și cel mai slab model 6%~) decât atunci când am antrenat modelul cu statisticile pe care eu, și comunitatea de baschet le consideră cele mai importante pentru a judeca performanța unui jucător, și anume: punctele, assist-urile, și rebound-urile + eficiența aruncărilor (eFG%) cu aceste 4 coloane, variația acurateței a scăzut puțin, maxim 1%~ între modele.

În general pentru a evalua modelele folosite în proiect am folosit eroare pătrată medie (Mean squared error), eroarea medie absolută (mean absolute error) pentru cele 4 statistici la care ne-am referit atunci când am discutat despre impactul asupra acurateții. În mare parte am avut o eroare destul de mică relativ la range-ul pe care se întind valorile (sub 0.2 la un PRA + eficiența de 53~).

Am experimentat și cu scăderea sau ridicarea pragului de threshold pentru care modelul prezicea dacă un jucător este MVP sau nu, am descoperit că sweet spot-ul se afla la 0.5, lucru reafirmat și la laborator, dar și în lucrări științifice. Un mod popular de antrenare a unui clasificator sensibil la costuri fără o matrice de costuri cunoscută este să se pună accent pe modificarea ieșirilor de clasificare atunci când sunt făcute predicții pe date noi. Acest lucru se face de obicei prin stabilirea unui prag pentru clasa pozitivă, sub care este prezisă clasa negativă. Valoarea acestui prag este optimizată folosind un set de validare, iar astfel matricea de costuri poate fi învățată din datele de antrenament.

6. Rezultate

Înainte de a enumera rezultatele testelor mele cu date introduse manual vreau să descriu puțin „gold standard”-ul când vine vorba de statistici în NBA, toți jucătorii își doresc să atingă renumitul triple double, asta înseamnă că în toate cele 3 statistici care măsoară direct performanța lor, points-rebounds-assists aceștia au un număr peste 10, de aici se trage numele de triple double(digits). Un singur jucător din era noastră a reușit să aibă ca average de-a lungul unui sezon triple double Russel Westbrook în sezonul 2016-2017, având 31.6 puncte, 10.7 rebound-uri și 10.4 assist-uri, în acel sezon el a câștigat MVP-ul chiar dacă echipa sa Oklahoma City Thunder a terminat pe locul 6, ultimul loc de play-off direct fără calificare.

Acestea fiind spuse am testat următoarele lucruri:

- Cat de mult afectează punctele un MVP?
- Poate un jucător egoist să câștige MVP-ul?
- Ce importanță are defensivă (rebounds) în alegerea MVP-ului?
- Poate să câștige un jucător ineficient MVP-ul?
- Care sunt statisticile minime?

6.1. Cat de mult afectează punctele un MVP?

Chiar în acest dataset avem un pasator extraordinar care a câștigat de 2x MVP, Steve Nash, noi vom analiza sezonul 2004-2005 în care a câștigat premiul cu doar 15.5 puncte în medie.

```
'PTS': [15.5],  
'AST': [11.5],  
'TRB': [3.3],  
'eFG%': [.58]
```

Modelul consideră că aceste statistici nu sunt suficiente, așa că am îmbunătățit puțin statisticile până când am obținut un rezultat pozitiv. Am decis să modific doar asist-urile pentru a ajunge la breakpoint, și am ajuns la 24.5 (2.13x) ! de asist-uri, deci mai mult de dublu. Dacă inversăm statisticile de puncte cu cele de asist-uri

```
'PTS': [24.5],  
'AST': [15.5],  
'TRB': [3.3],  
'eFG%': [.58]
```

atunci modelul consideră că nu este demn de MVP, din acest test putem să spunem că un asist are o pondere mai mare în alegere, ceea ce are sens, având în vedere că un asist duce la o marcă de 2 sau chiar 3 puncte.

6.2. Poate un jucător egoist să câștige MVP-ul?

Pentru această curiozitate am decis să observăm sezonul 1997-1998 în care Michael Jordan a câștigat MVP-ul cu următoarele statistici

```
'PTS': [28.7],  
'AST': [3.5],  
'TRB': [5.8],  
'eFG%': [.493]
```

Aceste statistici nu sunt foarte spectaculoase, predicția modelului este că nu merită titlul, iar pentru a obține o predicție pozitivă, trebuie să creștem eficiența eFG% și punctele până la numere ridicole de mari, ne mai întâlnim în viața reală, deci modelul defavorizează jucătorii egoiști.

6.3.Ce importanta are defensiva (rebounds) in alegerea MVP-ului?

Am sa folosesc statisticile din întrebarea trecută pentru a vedea dacă un jucător egoist dar care se descurca bine în apărare are șanse mai mari de reusita. Pentru a primi aprobarea am ridicat numărul rebound urilor la 18~ (foarte mare) ceea ce ne arată ca nu este foarte important ca jucatorul sa fie bun in defensiva ci mai mult in ofensiva. Acest lucru poate sa fie influențat și de votarea pentru MVP, de obicei defensiva este considerata partea plictisitoare a baschetului, de aceea și premium separat Defensive Player of the Year.

6.4.Poate sa castige un jucător ineficient MVP-ul?

Pentru a ne răspunde la aceasta intrebare vom prelua statisticile lui Derrick Rose, castigatorul MVP cu unul dintre cele mai mici eFG% din dataset-ul nostru, sezonul 2010-2011.

'PTS': [25.0],

'AST': [7.7],

'TRB': [4.1],

'eFG%': [.485]

Chiar dacă eficienta sa este una mica pentru standardele la care ne raportăm noi, daca crestem numărul de puncte suficient de mult modelul va ajunge la un compromis și va avea un ouput pozitiv, undeva la 34 de puncte a fost pragul la care am ajuns. Putem sa spunem ca eficiență nu are un impact major asupra alegerii, într-un meci real din NBA dacă un jucător nu are o eficienta buna, acesta cel mai probabil nu o sa ajungă la praguri de puncte ridicate deci este justificata aceasta neglijență asupra eFG%.

6.5.Care sunt statisticile minime?

Am discutat la început despre triple doubles, dar acum este timpul sa testam care este minimum necesar pe care modelul îl considera semn pentru premiul de MVP. După nenumărate teste am ajuns la concluzia că statisticile ideale sunt balansate, un jucător care este un marcator bun, un pasator decent care își face treaba în apărare în timp ce este și mediocru în ceea ce privește eficiența are șanse mult mai mari să câștige premium decât un jucător care este un marcator pur dar îi lipsesc celelalte aspecte ale jocului.

'PTS': [26.9],

'AST': [10.1],

'TRB': [8.4],

'eFG%': [.65]

7. Concluzii

Modelul cantareste statisticile jucatorilor cu ponderi diferite în calculul pentru premium de MVP, de obicei el considera ca mulți jucători din trecut nu au meritat premiul din cauza statisticilor lor. Acest lucru ne poate duce la două concluzii : în NBA a devenit mai ușor sa se înscrie puncte deoarece arbitri dictează mai ușor free throw-uri, jucătorii nu mai încearcă așa de mult sa opreasca aruncările adversarilor deci defensiva a trecut pe un loc inferior sau o alta teorie viabilă este aceea ca pur și simplu jucătorii sunt mai talentati în zilele noastre, au antrenamente mai stricte, diete, facilități care îi ajuta sa isi dezvolte jocul într-o direcție pe care atleții din trecut nu au avut oportunitatea sa le trăiască.

Personal eu consider ca este o mixtură între cele două cauze prezentate anterior: e de o parte, modificările în regulament și în arbitraj au dus la creșterea numărului de puncte marcate și au favorizat un stil de joc mai ofensiv, ceea ce stimulează scorurile mai mari. Pe de altă parte, nu putem ignora faptul că jucătorii din zilele noastre beneficiază de avansuri tehnologice și științifice care le permit să-și maximizeze potențialul. Antrenamentele sunt mai bine structurate, dietele sunt personalizate și monitorizate constant, iar facilitățile și echipamentele de recuperare sunt mult mai avansate

Evaluarea meritelor pentru premiul MVP trebuie să țină cont de contextul specific al fiecărei ere în NBA. Fiecare generație de jucători a avut provocările și oportunitățile sale unice. În acest sens, recunoașterea talentului și a contribuțiilor jucătorilor trebuie să fie realizată cu o înțelegere nuanțată a acestor diferențe.

În lucrarea lor, Yilmaz Mustafa și Chatterjee Sangit au descoperit ca de obicei campioanele sezonului NBA au avut mai multe puncte pe meci dar și mai multe puncte marcate împotriva lor decât media ligii, a mai observat ca numărul de assist-uri pe meci si rebound-uri pe meci nu afectează neapărat campioana. [10]

Bibliography

- [1] R. Mandić, S. Jakovljević; F. Erčulj and E. Štrumbelj "Trends in NBA and Euroleague basketball: Analysis and comparison of statistical data from 2000 to 2017", *Public Library of Science San Francisco*, 2019.
- [2] Y. de Saá Guerra, J. M. M. González, N. A. López, S. S. Montesdeoca, D. R. Ruiz and J. M. G. Manso, "Analysis of competitiveness in the NBA regular seasons," *Baltic Journal of Sport and Health Sciences*, vol. 1, no. 80, 2011.
- [3] K.-C. Wang and R. Zemel, Classifying NBA offensive plays using neural networks, 42Analytics, 2016.
- [4] A. A. Albert, L. F. de Mingo López, K. Allbright and N. Gómez Blas, "A hybrid machine learning model for predicting USA NBA all-stars," *Electronics*, vol. 11, no. 1, p. 97, 2021.
- [5] C. M. Bishop, Pattern Recognition and Machine Learning (Information Science and Statistics), Springer, 2006.
- [6] T. Fadi, Z. Li and A. Neda "NBA game result prediction using feature analysis and machine learning" *Annals of Data Science*, vol. 6, no. 1, pp. 103-116, 2019.
- [7] G. Cheng, Z. Zhang, M. N. Kyebambe and N. Kimbugwe, "Predicting the outcome of NBA playoffs based on the maximum entropy principle," *Entropy*, vol. 18, no. 12, p. 450, 2016.
- [8] A. M. Saeed, "A Gentle Introduction To Sigmoid Function," *Start Machine Learning*, 2021.
- [9] T. G. Nick and K. M. Campbell, "Logistic regression," *Topics in biostatistics*, pp. 273-301, 2007.
- [10] M. R. Yilmaz and S. Chatterjee, "Patterns of NBA team performance from 1950 to 1998," *Journal of Applied Statistics*, vol. 27, no. 5, pp. 555-566, 2000.