NBA MVP Analysis - Prediction

Urh Peček

2021/22

Kazalo

1. Uvod

1.1 Predstavitev naloge

Po vseh letih nagrade najkoristnejšega igralca rednega dela sezone (MVP) v košarki se še vedno pojavlja veliko vprašanj in razprav o definiciji igralca, ki prejme tovrstno nagrado. Je to igralec, ki ima najboljšo individualno statistiko, najboljši igralec najboljše ekipe rednega dela ali pa gre za druge indikatorje, ki jih tvori več spremenljivk skupaj ali pa jih celo številke ne morejo opisati? Glede definicije najkoristnejšega igralca verjetno ni dokončnega odgovora, lahko pa nam pomagajo orodja kot so modeli strojnega učenja, ki se lahko s primernimi statističnimi podatki uporabijo za iskanje vzorcev in logiko izbire MVP igralca.

Pri vsem skupaj je ključno razumevanje kako se v ligi NBA odloča o izbiri najkoristnejšega igralca in s tem kako uporabiti njegovo individualno in ekipno statistiko. Skozi čas se je spreminjalo pravilo, kdo so odločevalci, ki podeljujejo nagrado MVP. Danes v ligi NBA odločevalce sestavlja 100 neodvisnih medijskih osebnosti, ki niso povezane z ekipami ali igralci. Vsak odločevalec izbira MVP-ja po sistemu točkovnega glasovanja in sicer izbere 5 igralcev in jih razvrsti na lestvici od 1 do 5: 1. - 10 točk, 2. - 7 točk, 3. - 5 točk, 4. - 3 točke in 5. - 1 točka. Igralec, ki prejme največje število točk (seštevek vseh odločevalcev), je izbran za najkoristnejšega igralca rednega dela sezone in tako prejme nagrado MVP.

V namen vpogleda v atribute in pregled ter razumevanje vpliva atributov, ki (ne) definirajo najboljšega igralca rednega dela sezone lige NBA in pomoč pri napovedi igralca, ki bo to nagrado prejel v naslednjih letih, bodo v tej nalogi preučeni zgodovinski podatki kandidatov za MVP prejšnjih sezon in na podlagi njih uporabljeni različni modeli strojnega učenja, ki bodo pomagali pri razumevanju atributov in njihovega vpliva na izbiro MVP ter pomoč pri napovedi za prihajajoče sezone.

Nalogo je mogoče obravnavati kot regresijski problem, kjer napovedujemo delež glasov, ki ga igralec prejme, torej številsko ciljno spremenljivko. Igralec, ki ima v posamezni sezoni napovedan najvišji delež glasov, bi napovedano prejel nagrado MVP. Za ocenjevanje napovedi lahko primerjamo tako napovedan delež glasov igralca z njegovim dejanskim deležem glasov kot tudi binarno spremenljivko MVP, ki označuje ali je igralec nagrado prejel ali ne.

Napoved kot tudi vpliv spremenljivk bomo pridobili z uporabo različnih modelov strojnega učenja. Za oceno kakovosti modela oziroma napovedne moči spremenljivk, lahko uporabimo nekoliko prilagojeno metodo navzkrižnega preverjanja, kjer na vsakem koraku učne podatke predstavljajo statistični podatki, z obema ciljnima atributoma, igralcev določenega števila sezon, napovedujemo pa delež glasov in nagrado MVP za naslednjo, še ne vključeno, sezono. Celotne učne podatke tako lahko razdelimo na število delov ekvivalentno številu obravnavanih sezon in na vsakem koraku model ocenimo na vseh preteklih sezonah glede na tisto, na kateri preverimo napovedi modela. To storimo za "vsako" sezono posebej in pridobimo napovedi za "vsako" od sezon na podlagi katerih lahko ocenimo oziroma izračunamo kakovost modela oziroma napovedno moč vključenih spremenljivk.

Za oceno kakovosti modela na podlagi deleža glasov lahko uporabimo regresijsko metriko MAE (povprečna absolutna napaka), ki zavzame vrednosti $[0, \infty]$ in manjša kot je vrednost, boljši je model.

$$MAE = \frac{1}{n} \sum_{i=1}^{n} |y_i - \hat{y}_i|$$

V kolikor pa primerjamo zgolj napovedane končne dobitnike nagrade z dejanskimi dobitniki pa lahko kot metriko uporabimo natančnost oziroma zaradi t.i. redkih dogodkov občutljivost.

$$\text{Občutljivost} = \frac{\text{Pravilno napovedani zmagovalci}}{\text{Pravilno napovedani zmagovalci} + \text{Napačno napovedani ne zmagovalci}}$$

Najboljši napovedni model doseže čim višjo stopnjo občutljivosti in ima obenem čim nižjo vrednost MAE.

1.2 Cilji naloge

Atribute oziroma statistike, ki jih bomo uporabili za napoved deleža prejetih glasov in tako prejemnika nagrade MVP in katerih vpliv bomo preučevali lahko na grobo razdelimo v 3 sklope. Prvi sklop predstavljajo ekipne statistike posameznega igralca, drugi sklop predstavljajo osnovne igralčeve statistike, ki jih vsi najbolje poznamo, v tretjem sklopu pa so nekatere napredne statistike, ki združujejo več osnovnih atributov in si jih ni tako enostavno predstavljati.

V skladu z opisanimi sklopi bomo predstavili atribute, analizirali njihov vpliv na MVP oziroma delež prejetih glasov (Share) in prek opisanega prečnega preverjanja na učnih podatkih preučili kako dobro napovejo prejemnika nagrade MVP. Cilj okoli katerega se bo naša raziskava vrtela bo torej napovedati delež glasov, ki ga igralec prejme tj. MVP glasovi določenega igralca / skupni MVP glasovi, in na podlagi tega napovedati igralca, ki prejme nagrado MVP tj. igralca z največjim deležem glasov v sezoni.

Da bi najbolje razumeli in predstavili vpliv posameznih atributov tako znotraj posameznega sklopa kot med sklopi bomo primerjali kakovost napovedi posameznih podmnožic spremenljivk in jih primerjali med seboj. S pomočjo tega bomo preverili katere spremenljivke sploh značilno vplivajo na odločitev o igralcu MVP in poskusili odgovorili na nekatera vprašanja, ki si jih ljubitelji NBA košarke, tudi v skladu z nagrado MVP, pogosto postavljajo. Ali so v očeh ameriške strokovne javnosti pomembnejše absolutne vrednosti kot so skupne točke, podaje skoki ali so pomembnejše relativne vrednosti teh atributov v smislu realizacije kot so odstotek meta, odstotek izkoriščenih priložnosti za podajo in skok. Podobno bomo preverili tudi ali so pomembnejše absolutne vrednosti atributov ali relativne vrednosti glede na odigrane minute. Poskusili bomo preveriti tudi kolikšen delež h deležu prejetih glasov prispevajo ekipne statistike, koliko osnovne statistike, koliko nekoliko prilagojene osnovne statistike in koliko napredne igralčeve statistike.

V zadnjem delu naloge pa bomo na podlagi celotnih testnih podatkov, ki jih predstavljajo sezone 1980/81 - 2020/21, s pridobljenim znanjem o vplivu spremenljivk ter primernem modelu za napoved deleža glasov oziroma igralca MVP, iz določenih kandidatov napovedali MVP igralca lige NBA sezone 2021/22.

1.3 Viri

Podatki za nalogo so bili pridobljeni s spletne strani basketball-reference.com. Za vsako od sezon 1980/81, kjer je bil že uveden met za tri točke in prvo leto, ko so MVP igralca izbirali novinarji (in ne igralci), do najnovejše sezone 2021/22 so bili pridobljeni igralci, ki so bili v ožjem izboru za nagrado MVP (od 9 do 31 igralcev) in pripisane so jim bile različne tako individualne kot ekipne statistike, ki bodo predstavljene tekom naloge. Glede na v predstavitvi naloge predstavljeno metodo napovedovanja in ocenjevanja modela učno množico vseskozi večamo in začnemo z ničelno, zato bomo za prvo sezono za katero preverjamo napovedi modela uporabili sezono 1989-1990 in tako ocenjevali napovedi in vpliv spremenljivk na podlagi 33 sezon.

Bolj kot ne zgolj za idejo in morda kakšen namig h celotnem pristopu smo si pomagali z nekaj članki strani towardsdatascience.com [Duarte Feire 25.3.2021 in David Yoo 11.1.2022] ter tudi člankom strani static1.squarespace.com [Tongan Wu] ter diplomsko nalogo [Jernej Luci 2018]. Večinski del so naše ideje, hipoteze, njihove utemeljitve in izvedbe.

2. Opisna analiza atributov in vpliv na nagrado Share in MVP

Kot rečeno smo atribute igralcev na grobo razdelili v tri sklope, znotraj katerih pa lahko ponovno ustvarimo več sklopov na podlagi opisanih domnev oziroma zastavljenih vprašanj. Eden od načinov preverjanja in moči vpliva posameznega atributa oziroma množice atributov, ki je enostavni model linearne regresije. Ta ima sicer nekatere predpostavke, ki omejujejo njegovo kakovost napovedi in uporabo, vendar na podlagi njega najlažje razberemo moč atributov pri napovedovanju Share-a. Prek regresijskih koeficientov modela lahko enostavno razberemo način in velikost linearnega vpliva posameznega atributa na Share (kar pri majhni napovedni moči ne gre pretirano upoštevati), prek popravljenega determinacijskega koeficienta R^2 pa razberemo delež pojasnjene variabilnosti Share-a s strani izbrane množice atributov. Znotraj posameznih sklopov bomo tudi s pomočjo elastic net modela preverili katere spremenljivke sploh značilno vplivajo na igralca MVP oziroma vrednost Sgare.

Moč napovedi atributov bomo preverjali tudi na podlagi opisanega prečnega preverjanja in izbranih metrik kakovosti modela. Kot izbrane modele strojnega učenja smo vključili že omenjen elastic net, naključne gozdove, light gradient boost in extreme gradient boost. Ti modeli v primerjavi z enostavnim linearnim niso odvisni od linearne povezanosti atributov z odvisno spremenljivko, multikoreliranosti in nekaterih ostalih predpostavk in tako niso tako močno odvisni od izbire atributov. Tako bomo lahko določen nabor spremenljivk lahko primerjali z večjim, in tako primerjali njihovo moč napovedi oziroma (ne)moč napovedi spremenljivk, ki so izpuščene.

Za oceno, oziroma pomoč pri njej, posamičnega vpliva spremenljivk na Share in posledično MVP smo v začetku, pred podrobnejšo analizo atributov, za vsakega izmed atributov izračunali več meritev, to so: korelacija s Share, Mutual Information s Share, pomembnost atributa, ki jo pridobimo z enostavnega naključnega gozda z vsemi atributi, p-vrednost atributa in popravljena vrednost R^2 iz samostojnega linearnega modela, p-vrednost atributa in psevdo R^2 (McFadden) iz samostojnega regresijskega modela za MVP ter MAE, občutljivost in število pravilnih napovedi na podlagi prečnega preverjanja z samostojnim regresijskim modelom.

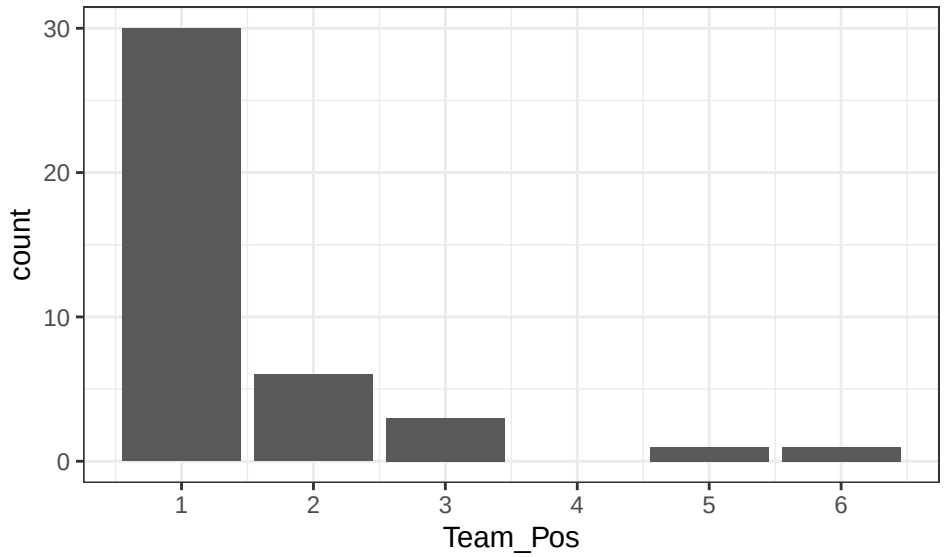
Poleg vseh teh meritev, ki nam služijo zgolj kot nekakšna dodatna informacija se bomo v nadaljevanju osredotočali na grafične prikaze, vključeni bodo le najbolj zanimivi, ter povprečne absolutne vrendosti atributov ločene glede na indikator MVP ter njihove absolutne ter relativne razlike, ki podajajo večjo oziroma boljšo informacijo.

2.1 Ekipna statistika

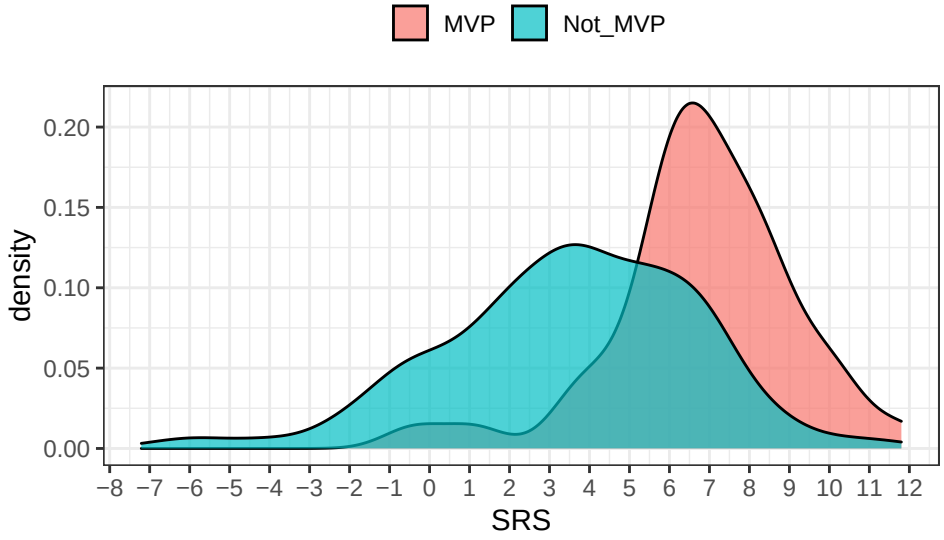
Spremenljivke, ki smo jih vključili pod sklop ekipne statistike:

- W (Team Wins) - Zmage ekipe v celotnem rednem delu sezone (82 tekem)
- L (Team Losses) - Porazi ekipe v celotnem rednem delu sezone (82 tekem)
- W/L% (Team Win-Loss Percentage) - Odstotni delež zmag ekipe v rednem delu sezone
- GB (Team games behind) - Število zmag manj od prvo uvrščene ekipe v konferenci
- PS/G (Team points per game) - Povprečno število danih točk ekipe na tekmo
- PA/G (Team opponent points per game) - Povprečno število prejetih točk ekipe na tekmo
- SRS (Simple Rating System) - Ekipna ocena, ki upošteva povprečno razliko točk in moč razporeda izražena v točkah nad/pod povprečjem
- Team_Pos (Team position in Conference) - Mesto na lestvici ekipe znotraj konference

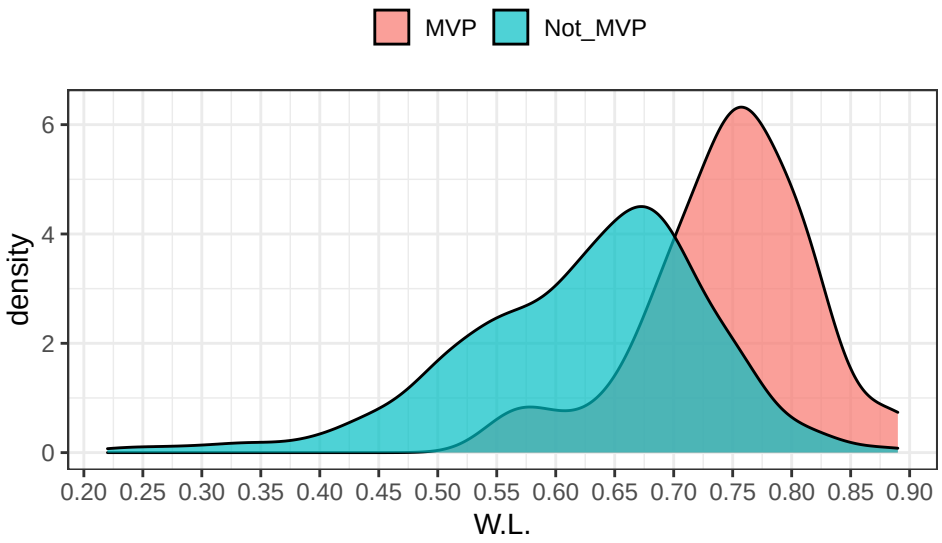
Hitro vidimo, da so nekatere spremenljivke med seboj močno korelirane. Mesto na lestvici je denimo močno povezano z zmagami in porazi ekipe, odstotnim deležem zmag, številom zmag zaostanka in tudi SRS. Ekipne zmage in porazi ter deleži zmag so skoraj popolnoma korelirani (včasih se igrajo podaljški, ki se štejejo drugače).



V naših podatkih kar 73% končnih MVP igralcev prihaja iz prvo uvrščene ekipe v konferenci in 95% jih prihaja iz ene izmed prvih treh ekip znotraj konference. Povprečni MVP igralec je v ekipi, ki je na 1'5. mestu na konferenčni lestvici in povprečni igralec, ki ne prejme nagrade MVP je v ekipi, ki je na 4. mestu na konferenčni lestvici. Povprečna odstotna razlika med MVP in ne MVP igralci je kar 172% in v samostojnem modelu spremenljivka pojasni 11% variabilnosti Share. Mesto na lestvici ekipe znotraj konference je torej gotovo eno izmed bolj pomembnih spremenljivk. Morda pa velja opomniti, kot bomo še nekajkrat, da povezava s Share-om oziroma MVP-jem popolnoma enosmerna, saj je pričakovati, da najboljši igralec svojo ekipo dela dobro, oziroma najboljšo.



Močno povezana spremenljivka z mestom ekipe na lestvici je tudi SRS, ki spremenljivki Team_Pos doda še nekaj informacije. 75% MVP-jev prihaja iz ekipe z vrednostjo SRS višjo od 5.8 in 75% ne MVP-jev prihaja iz ekipe z vrednostjo SRS nižjo od 75%. Razlika v povprečni vrednosti SRS med MVP-ji in ne MVP-ji je 48%. Torej tudi spremenljivka SRS igra pomembno vlogo pri napovedovanju Share oziroma MVP, vendar bi glede na vse izmerjene metrike lahko rekli da nekoliko manjšo, čeprav nosi več informacije. Zaključili bi lahko, da odločevalci bolj kot na točkovno razliko tekem in težavnost razporeda bolj gledajo zgolj na končne zmage in tako razvrstitev ekipe na lestvici.



Kot rečeno so spremenljivke ekipnih zmag, porazov in deleža zmag skoraj popolnoma korelirane s spremenljivko Team_Pos oziroma SRS in velja, da povprečni MVP prihaja iz ekipe, ki v sezoni doseže 10 (12.1% delež zmag) zmag več kot ekipa povprečnega ne MVP igralca. Povprečno število danih in prejetih točk ekipe ne vpliva značilno na vrednost Share ali MVP.

Na podlagi izračunanih meritev med MVP-ji in ne MVP-ji najbolje ločita in pojasnjujeta vrednost Share spremenljivki Team_Pos in W/L%, ki je tudi v najbolj linearnem razmerju s Share. Da preverimo, ali je izmed ekipne statistike dovolj upoštevati zgolj delež zmag in ne tudi ostalih atributov, lahko preverimo ocene koeficientov oziroma značilnost atributov v elastic net modelu. Izkaže se, da na vrednost Share in MVP (močno) značilno vplivata zgolj ugotovljena atributa deleža zmag in mesta na lestvici, ki pa sta močno korelirana. Medtem ko število porazov in zmag, tekem zaostanka in SRS vplivajo zelo pogojno.

Da zgolj za informacijo preverimo delež pojasnjene variabilnosti Share s strani ekipnih statistik, čeprav predpostavimo linearni vpliv, je torej dovolj, da v linearni vključimo zgolj eno od spremenljivk W/L% ali Team_Pos. Odločimo se za spremenljivko W/L% in izkaže se, da je delež pojasnjene variabilnosti Share enak 12.2%.

Pri prečnem preverjanju z modeli uporabljenimi na vseh atributih dobimo naslednje rezultate:

Metrika	Elastic.Net	Random.Forest	Light.GB	Extreme.GB
MAE	0.189	0.196	0.203	0.205
Občutljivost	0.531	0.375	0.344	0.188
Pravilne napovedi	17/32	12/32	11/32	6/32

2.2 Osnovna statistika

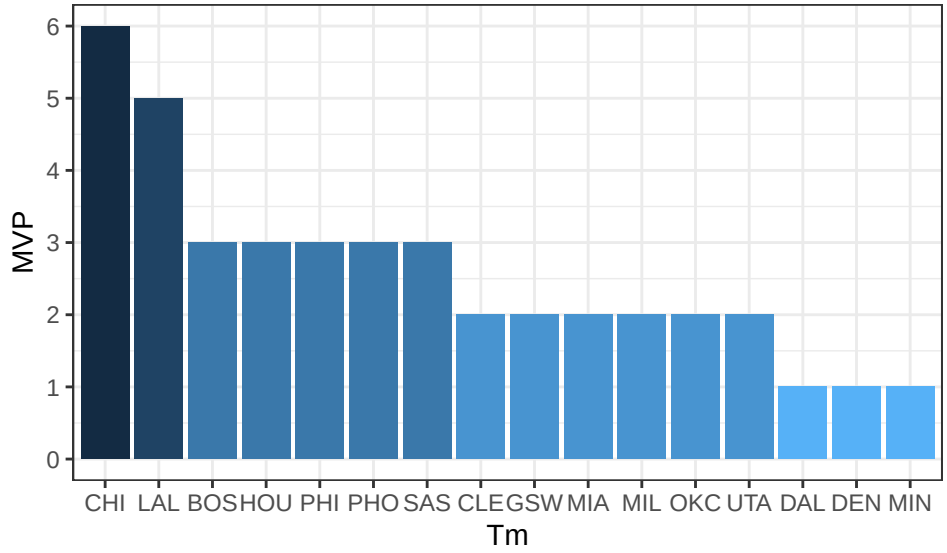
Osnovno igralčevo statistiko bomo dodatno razdelili in sicer na 4 sklope, prek česar bomo lahko tudi bolje primerjali na kateri sklop izmed osnovnih statistik se osredotočajo odločevalci pri nagradi MVP. Osnovno statistiko bomo razdelili na osnovni opis igralca, absolutne vrednosti metov na koš, relativne vrednosti metov na koš in ostale statistike. Dodatno bomo preverili tudi ali se odločevalci odločajo bolj na podlagi absolutnih vrednosti ali vrednosti na odigrane minute.

2.2.1 Osnovni opis

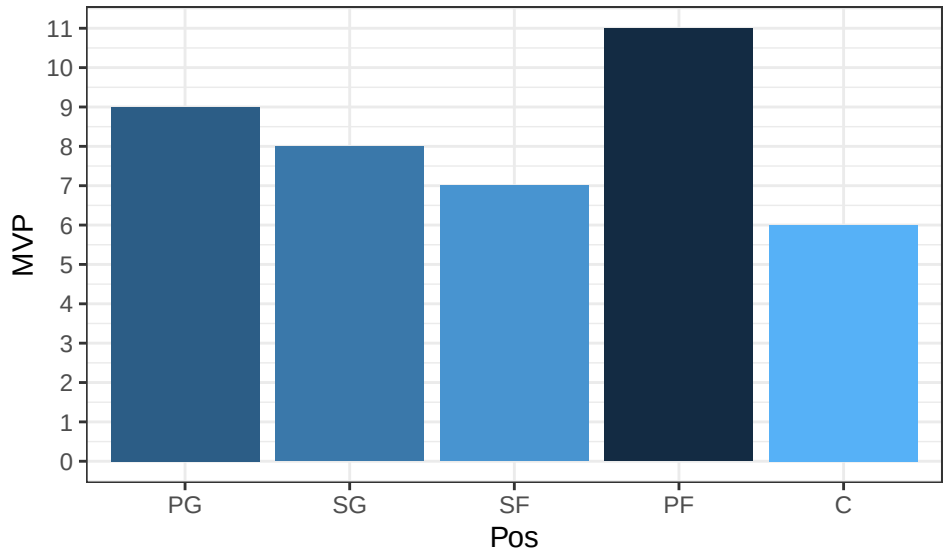
- Age - Starost

- Tm (Team) - Ekipa
- G (Games) - Število odigranih tekem (82 tekem)
- GS (Games Started) - Število tekem začetih v prvi peterki (82 tekem)
- MP (Minutes Played) - Povprečno število odigranih minut na tekmo (tekma ima 48 minut)
- Pos (Position) - Igralčeva pozicija (PG, SG, SF, PF, C)

Ker odločevalci niso povezani z ekipami, ekipa igralca v tem smislu ne bi smela odločati o izbiri MVP igralca. Lahko pa ekipa odloča v smislu njene uspešnosti oziroma kakovosti, kar je povezano z zgornjo spremenljivko Team_Pos oziroma W/L%, ki se skozi daljše obdobje ne bi smela preveč nihati. Od sezone 1980/81 do 2020/21 največ MVP-jev prihaja iz ekipe Chicago Bulls, in sicer 6, pri čemer je bil kar 5-krat to Michael Jordan in 1-krat Derrick Rose. Bullsom sledi ekipa Los Angeles Lakers s trikratnim dobitnikom nagrade Magicom Johnsonom in 1-kratnima dobitnika Shaquilleom O'Nealom in Kobejem Bryantom. Vse ostale ekipe so dokaj izenačene in tudi na podlagi novejše zgodovine lahko rečemo, da ekipe ne vpliva na izbiro MVP.

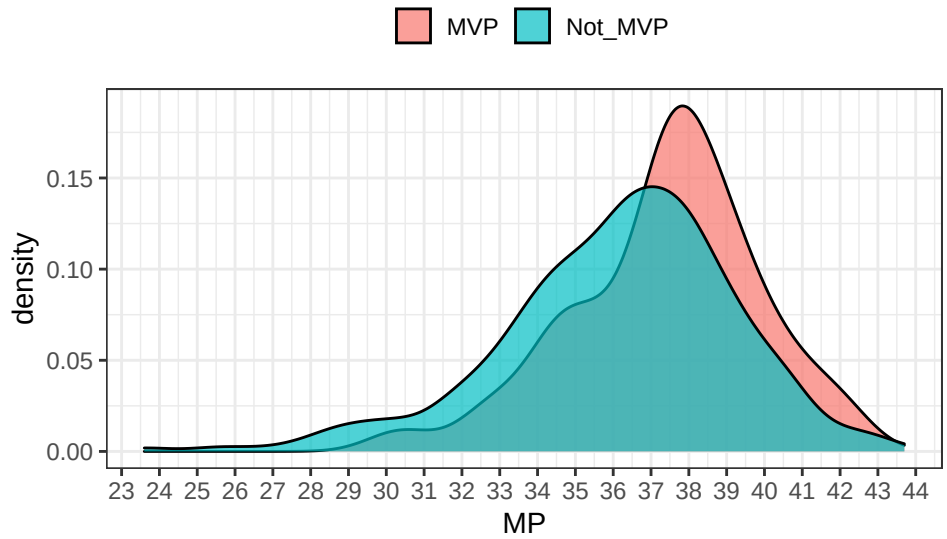


Pozicije igralcev so med prejemniki nagrad MVP zastopane precej enakovredno (PG = 9, SG = 8, SF = 7, PF = 11, C = 6) in tako lahko sklepamo, da tudi pozicija igralca (in s tem tudi njegova velikost) ne vpliva značilno na prejemnika nagrade MVP.



Starost igralca, število odigranih in začetih tekem prav tako nimajo značilnega vpliva na MVP oziroma Share. Večina igralcev, ki pridejo v ožji izbor za MVP so tako ali tako zvezdniki svoje ekipe in začenjajo tekme ter odigrajo večinski del sezone. Slednji dve spremenljivki bi pa definitivno (pozitivno) vplivali v primeru celotnega nabora igralcev in ne samo ožjega izbora na podlagi katerega se mi odločamo.

Glede števila odigranih minut v povprečju na tekmo bi moralo veljati podobno kot pri številu (začetih) tekem. Ožji izbor igralcev za MVP na igrišču preživi večino časa, morda centri nekoliko manj, vendar to ne bi smelo poglavitno vplivati na izbiro MVP oziroma delež glasov. Na spodnjem grafu pa vendarle vidimo, da je razlika med skupinama obstaja, vendar je majhna. Po drugi strani pa število odigranih minut pa lahko vpliva posredno oziroma v kombinaciji z nekaterimi drugimi spremenljivkami, kar pa bomo preverili v nadaljevanju.



Na podlagi zgornjega in tudi značilnosti koeficientov elastic net modela, nobeden izmed osnovnih "opisnih" atributov ne vpliva (močno) značilno na MVP ali Share. Tudi izračun ostalih statistik ne kaže, da bi te atributi značilno vplivali. Za informacijo, enostavni linearni model z vsemi navedenimi spremenljivkami pojasni 15% variabilnosti Share, vendar najverjetneje prihaja do velikega preprileganja.

Pri prečnem preverjanju z modeli uporabljenimi na vseh atributih dobimo naslednje rezultate:

Metrika	Elastic.Net	Random.Forest	Light.GB	Extreme.GB
MAE	0.191	0.176	0.2	0.224
Občutljivost	0	0.25	0.156	0.125
Pravilne napovedi	0/32	8/32	5/32	4/32

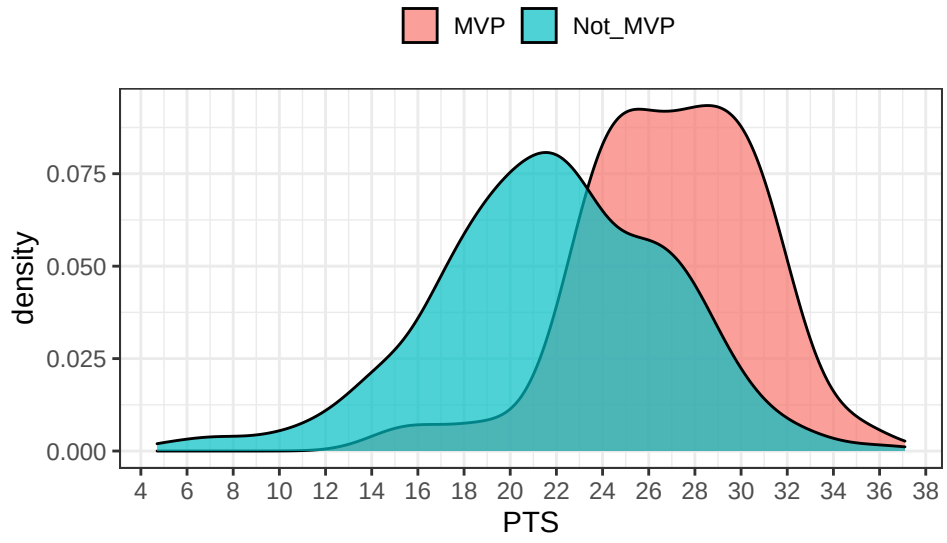
Vidimo, da so rezultati oziroma kakovosti napovedi zelo slabe, veliko slabše kot pri ekipnih atributih. To pomeni, da opisne statistike igralcev res razložijo minimalen del Share oziroma MVP.

2.2.2 Absolutne vrednosti metov

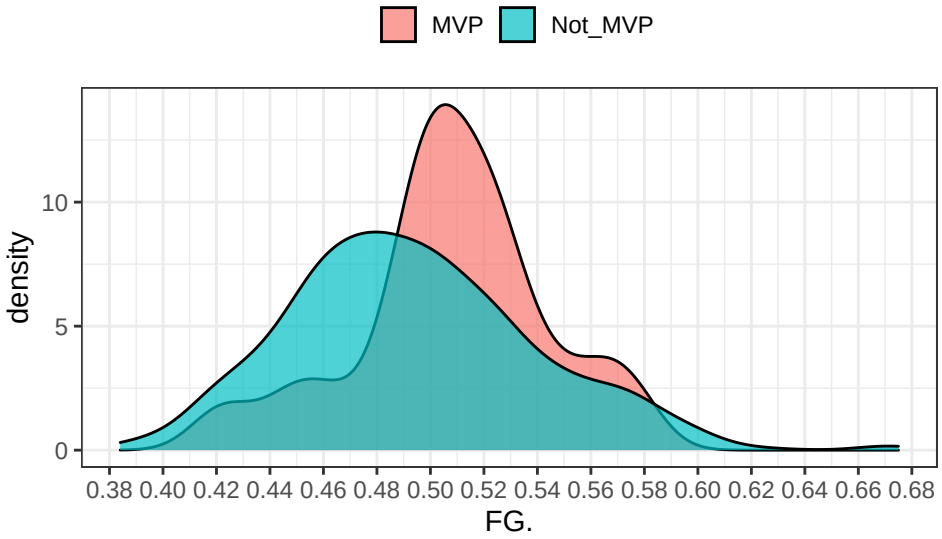
- PTS (Points Per Game) - Povprečno število točk na tekmo
- FG (Field Goals Per Game) - Povprečno število zadetih metov na tekmo
- FGA (Field Goal Attempts Per Game) - Povprečno število metov na tekmo
- 3P (3-Point Field Goals Per Game) - Povprečno število zadetih metov za 3 točke na tekmo
- 3PA (3-Point Field Goal Attempts Per Game) - Povprečno število metov za 3 točke na tekmo
- 3PAr (3-Point Attempt Rate) - Povprečni delež metov za 3 točke na tekmo
- 2P (2-Point Field Goals Per Game) - Povprečno število zadetih metov za 2 točki na tekmo
- 2PA (2-Point Field Goal Attempts Per Game) - Povprečno število metov na tekmo
- FT (Free Throws Per Game) - Povprečno število zadetih prostih metov na tekmo
- FTA (Free Throw Attempts Per Game) - Povprečno število prostih metov na tekmo
- FTr (Free Throw Attempt Rate) - Povprečni delež prostih metov na tekmo

V vseh primerih velja, da sta spremenljivki števila metov in števila zadetih metov močno korelirani. Obenem bi rekli, da bi zadeti meti morali razložiti večji del variabilnosti Share oziroma MVP, saj le ti štejejo, pri zgolj metih pa je poleg informacije odstotne uspešnosti izpuščena pomembna informacija absolutne uspešnosti.

Če najprej pogledamo točke, je razlika med MVP-ji in ne MVP-ji 18% in velja, da 75% MVP-jev v povprečju doseže več kot 24.5 točk na tekmo in 75% ne MVP-jev v povprečju doseže manj kot 25.5 točk na tekmo. Razlika med skupinama je torej precejšnja in tudi po osebnem mnenju oziroma znanju bi rekli, da so točke eden izmed ključnih dejavnikov pri napovedovanju MVP.



Pri številu vseh metov, torej prosti meti, 2 točki in 3 točke, v skladu z zgornjim premislekom velja, da zadeti meti pojasnijo nekoliko (čeprav ne veliko) večji delež variabilnosti Share oziroma MVP. Razlika med MVP igralci in ne MVP igralci je 15% in sicer MVP-ji v povprečju izmed 19 metov zadenejo 9.3 metov ter ne MVP-ji izmed 17.3 metov zadenejo 8.8 metov.

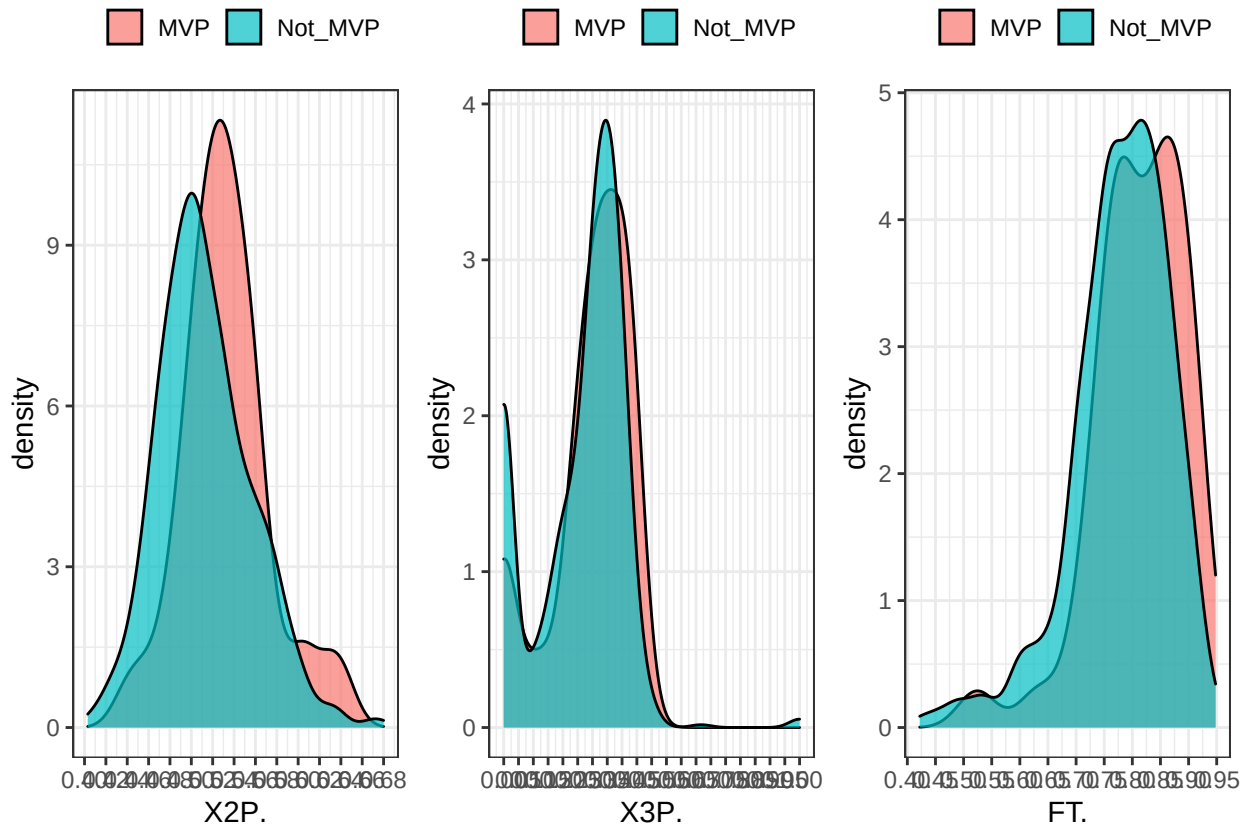


Vse tri omenjene spremenljivke so močno korelirane in v kolikor se je potrebno odločiti za samo eno, bi se odločili za povprečne točke, ki nosi največ in najbolj pomembno informacijo.

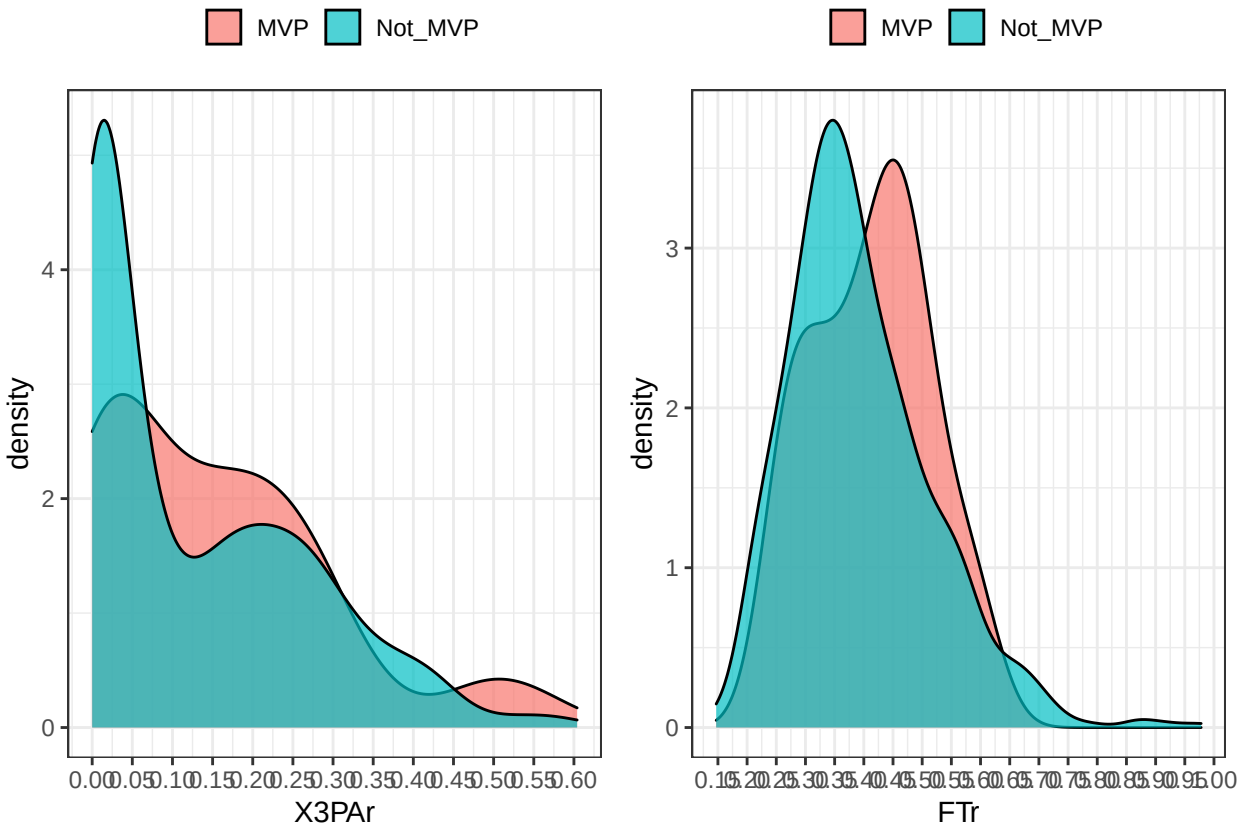
Pri metu za 3 točke velja, da MVP-ji v povprečju vržejo 0.7 oziroma 23% trojke na tekmo več kot ne MVP-ji, zadenejo pa 0.3 trojke oziroma 27% več kot ne MVP-ji. Upoštevajoč grafični prikaz, razlika med skupinama ni močno značilna. Če bi pogledali glede na igralske pozicije, so opazne razlike le pri igralcih na pozicijah C in FG, kjer več mečejo (in so uspešnejši) MVP-ji, vendar se v podrobnosti zaradi relativno majhnega nabora igralcev ne bomo spuščali.

Podobno kot pri metu za 3 točke velja tudi pri metu za dve točki, le da je razlika med MVP-ji in ne MVP-ji še nekoliko manjša in sicer 14% oziroma 17%. Met za dve točki je močno koreliran tudi s številom točk in ostalimi zgornjimi spremenljivkami.

Pri prostih metih velja, da MVP-ji v povprečju mečejo večje število prostih metov, in jih več tudi zadenejo (7.9 proti 6.4 oziroma 6.4 proti 4.9). Razlika je 20% oziroma 22%, kar je precej in glede na to dobimo občutek, da je povezanost z MVP-jem močna.



Če pa pogledamo še delež metov igralcev, ki pride iz za črte za 3 točke oziroma proste mete velja, da MVP-ji v povprečju večji delež metov opravijo za 3 točke (15% proti 13%) in razlika je 15%, medtem ko razlika pri prostih metih ni značilna.



Opozoriti velja, da je povezanost vseh zgornjih spremenljivk z MVP-jem oziroma Share-om v bistvu dvostranska ali morda celo obratna. Pričakovati gre, da bodo najboljši igralci največ metali in tako dosegli največje število točk/metov. Vendar pa je to posledica njihove kakovosti in zaupanja ekipe oziroma trenerja in si tako na podlagi svoje kakovosti verjetno zaslužijo tolikšno število metov v primerjavi z ostalimi. Tako smo razrešili nekolikšen dvom o smiselnosti upoštevanja teh spremenljivk. Na to se velja spomniti tudi v prihodnje, pri podobnih atributih.

S pomočjo značilnosti koeficientov elastic net modela, povzamemo, da so vsi atributi značilni. Zanimivo, je razlika med značilnostjo atributov pri napovedovanju zgolj MVP in Share precejšnja. Za najpomembnejši atribut se izkaže delež metov za 3 točke in pa tudi število zadetih metov, tako za 2 kot za 3 točke, ki sta seveda močno korelirana s številom točk. Za informacijo, enostavni linearni model z vsemi navedenimi spremenljivkami pojasni 23% variabilnosti Share.

Pri prečnem preverjanju z modeli uporabljenimi na vseh atributih dobimo naslednje rezultate:

Metrika	Elastic.Net	Random.Forest	Light.GB	Extreme.GB
MAE	0.175	0.167	0.173	0.175
Občutljivost	0.281	0.125	0.125	0.125
Pravilne napovedi	9/32	4/32	4/32	4/32

V primerjavi z že preverjenima sklopoma spremenljivk so rezultati precej boljši, vsaj v smislu Share. V smislu pravilne napovedi MVP, ki je odvisna od Share so rezultati še vedno precej slabi. Glede na vrednost MAE bi lahko torej rekli, da osnovne igralske statistike pojasnijo večji delež Share kot ekipne staitstike. Še posebej pa nas zanima kakovost napovedi v primerjavi z relativnimi in minutnimi statistikami meta, kar bomo preverili naslednje.

2.2.3 Relativne vrednosti metov

- FG% (Field Goal Percentage) - Odstotni delež zadetih metov iz igre
- 3P% (3-Point Field Goal Percentage) - Odstotni delež zadetih metov za 3 točke
- FT% (Free Throw Percentage) - Odstotni delež zadetih prostih metov
- 2P% (2-Point Field Goal Percentage) - Odstotni delež zadetih metov za 2 točki
- eFG% (Effective Field Goal Percentage) - Odstotni delež zadetih metov, kjer prilagodimo, da so meti za 3 točke več vrednosti kot meti za 2 točki
- TS% (True Shooting Percentage) - Merilo učinkovitosti metanja, ki upošteva mete za 2 točki, 3 točke in proste mete

Vidimo, da so vse spremenljivke močno povezane. Načeloma največ informacije nosi TS%, kasneje eFG% in potem FG%. Po hitrem premisleku bi morale te relativne spremenljivke močneje odločati o MVP-ju kot zgornje absolutne spremenljivke, še posebej, ker že imamo ožji izbor kandidatov in se s tem "znebimo" nekaterih nepotrebnih osamelcev. Preverili bomo ali zgornja predpostavka velja.

Upoštevajoč spodnje grafe so razlike v porazdelitvah podobne kot pri absolutnih vrednostih pripadajočih spremenljivk. Vendar, velja upoštevati skalo na x osi, ki kaže, da sta v resnici porazdelitvi precej prekrivajoči. V kolikor pogledamo relativne razlike povprečnih vrendosti skup posameznih spremenljivk je razlika povsod veliko manjša kot pri absolutnih atributih. Razlika je povsod 3% - 4% (ne v smislu absolutne razlike vrednosti spremenljivke).

Tako velja, da je v primerjavi z absolutnimi vrednostmi spremenljivk, relativna razlika med skupinama v vseh primerih močno na strani absolutnih spremenljivk. Tudi v samostojnih linearnih modelih absolutne spremenljivke pojasnijo značilno večji delež variabilnosti Share in so bolj korelirane z njim.

Če pogledamo koeficientie elastic net modela vidimo, da atributa eFG% in FG% ne pojasnita značilno Share oziroma se ne razlikuje med skupinama. Zanimivo pa je atribut TS% močno značilen. Močno značilna sta tudi atributa deleža zadetih prostih metov metov za dve točki, medtem ko je atribut deleža zadetih metov za 3 točke mejno značilen.

Pri prečnem preverjanju z modeli uporabljenimi na vseh atributih, katerim smo dodali še spremenljivki 3PAr in FTr iz zgornjega sklopa, dobimo naslednje rezultate:

Metrika	Elastic.Net	Random.Forest	Light.GB	Extreme.GB
MAE	0.191	0.193	0.199	0.207
Občutljivost	0.062	0.188	0.188	0.125
Pravilne napovedi	2/32	6/32	6/32	4/32

Na podlagi zgornjih rezultatov, predvsem vrednosti MAE, ko tudi informacije, da enostavni linearni model z vsemi navedenimi spremenljivkami pojasni zgolj 7% variabilnosti Share lahko zaključimo, da odločevalce bolj kot odstotek zadetih metov in tako izkoristek napadov oziroma metov zanimajo absolutne vrednosti metov, torej same točke oziroma zadeti meti. To pojasni tudi pogled na košarko v Ameriki, kje se vrti vse okoli velikih številk, v primerjavi z Evropo, kjer je vsak med skrbno premišljen. Še vedno, pa so deleži pravilno napovedanih igralcev MVP nizki, kljub boljšim napovedim vrendosti Share.

2.2.4 Vrednosti metov na odigrane minute

- PTS/Min (Points Per Game) - Povprečno število točk na odigrano minut
- FG/Min (Field Goals Per Game) - Povprečno število zadetih metov na odigrano minut
- FGA/Min (Field Goal Attempts Per Game) - Povprečno število metov na odigrano minut
- 3P/Min (3-Point Field Goals Per Game) - Povprečno število zadetih metov za 3 točke na odigrano minut
- 3PA/Min (3-Point Field Goal Attempts Per Game) - Povprečno število metov za 3 točke na odigrano minut

- 2P/Min (2-Point Field Goals Per Game) - Povprečno število zadetih metov za 2 točki na odigrano minut
- 2PA/Min (2-Point Field Goal Attempts Per Game) - Povprečno število metov na odigrano minut
- FT/Min (Free Throws Per Game) - Povprečno število zadetih prostih metov na odigrano minut
- FTA/Min (Free Throw Attempts Per Game) - Povprečno število prostih metov na odigrano minut

Podobno kot pri primerjavi absolutnih in relativnih vrednostih metov bomo preverili tudi vpliv vrednosti metov na povprečno odigrano minuto in ali le ti odločevalce zanima bolj kot absolutne vrednosti. Velikih razlik v porazdelitvah atributov, pogojno na skupini igralcev, ni, vendar pa bi bile lahko razlike v prekrivanju porazdelitev skupin, kar pa zaradi različnih merskih enot ne moremo primerjati.

Na podlagi koeficientov elastic net modela večjih razlik v značilnosti spremenljivk v primerjavi z absolutnimi atributi ni. Še vedno so vsi atributi značilni. Pri prečnem preverjanju z modeli uporabljenimi na vseh atributih, katerim smo dodali še spremenljivki 3PAr in FTr, dobimo naslednje rezultate:

Metrika	Elastic.Net	Random.Forest	Light.GB	Extreme.GB
MAE	0.175	0.178	0.182	0.18
Občutljivost	0.219	0.125	0.188	0.156
Pravilne napovedi	7/32	4/32	6/32	5/32

Na podlagi zgornjih rezultatov, predvsem vrednosti MAE, ko tudi informacije, da enostavni linearni model z vsemi navedenimi spremenljivkami pojasni 22% variabilnosti Share lahko zaključimo, da je razlika v pojasnjevalni moči metov na minuto in absolutnih vrendostih metov minimalna ali pa celo ponvno nekoliko na strani absolutnih vrednosti. Vidimo torej, da spremenljivka odigranih minut ne vpliva na Share ali MVP posredno, prek ostalih spremenljivk. Torej, načeloma je vseeno ali primerjamo igralce na odigrane minute ali pa na odigrano tekmo saj so razlike minimalne, morda celo bolje da jih primerjamo absolutno, na odigrano tekmo.

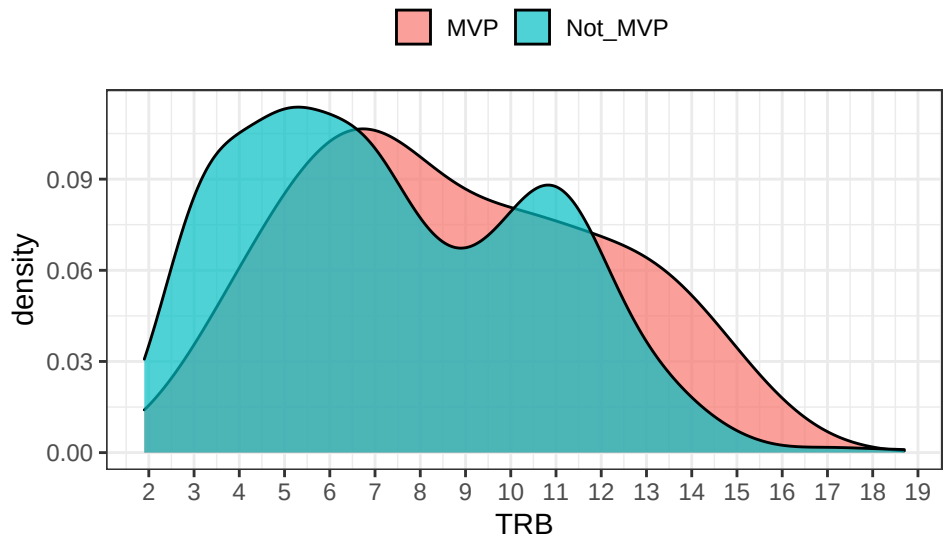
2.2.4 Ostale statistike

- TRB (Total Rebounds Per Game) - Povprečno število skokov na tekmo
- ORB (Offensive Rebounds Per Game) - Povprečno število napadalnih skokov na tekmo
- DRB (Defensive Rebounds Per Game) - Povprečno število obrambnih skokov na tekmo
- AST (Assists Per Game) - Povprečno število podaj na tekmo
- STL (Steals Per Game) - Povprečno število ukradenih žog na tekmo
- BLK (Blocks Per Game) - Povprečno število blokad na tekmo
- TOV (Turnovers Per Game) - Povprečno število izgubljenih žog na tekmo
- PF (Personal Fouls Per Game) - Povprečno število prekrškov na tekmo

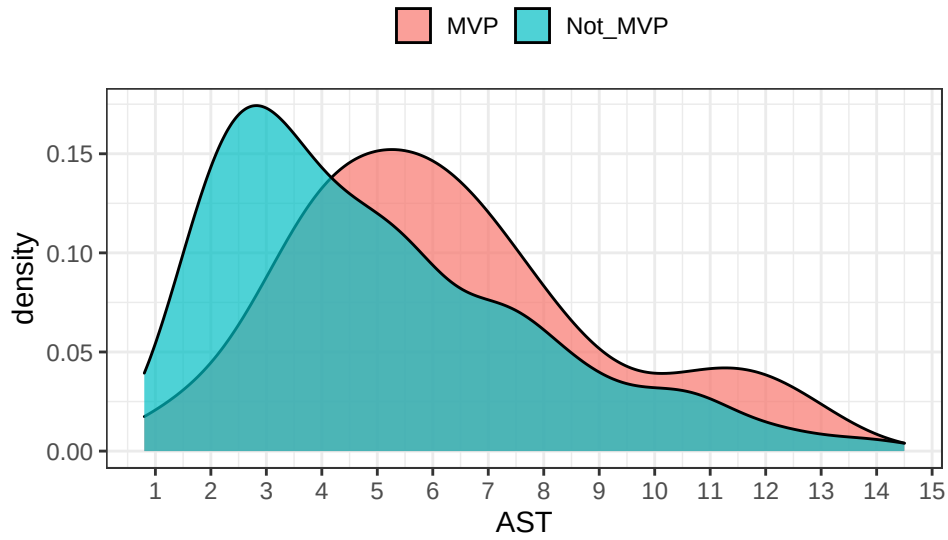
Na prvi pogled na MVP nagrado najbolj vplivata spremenljivki števila skokov in števila podaj. Ti dve spremenljivki sta poleg števila točk v splošni javnosti največkrat opazovani. Druge spremenljivke najverjetneje zgolj posredno "pridejo" z igralcem in jih odločevalci ne upoštevajo v večji meri.

Povprečno skupno število skokov, število napadalnih skokov in število obrambnih skokov so med seboj močno korelirani, oziroma napadalni in obrambni skoki skupaj tvorijo skupne skoke. Največji delež skokov seveda predstavljajo obrambni skoki in sicer 75%. V povprečju MVP-ji dosegajo 1.3 oziroma 19% obrambnih skokov več kot ne MVP-ji, pri napadalnih skokih pa je to razmerje 0.2 oziroma 5% skokov. Skupaj MVP-ji dosegajo 1.3 oziroma 15% več skokov (8.7 proti 7.4).

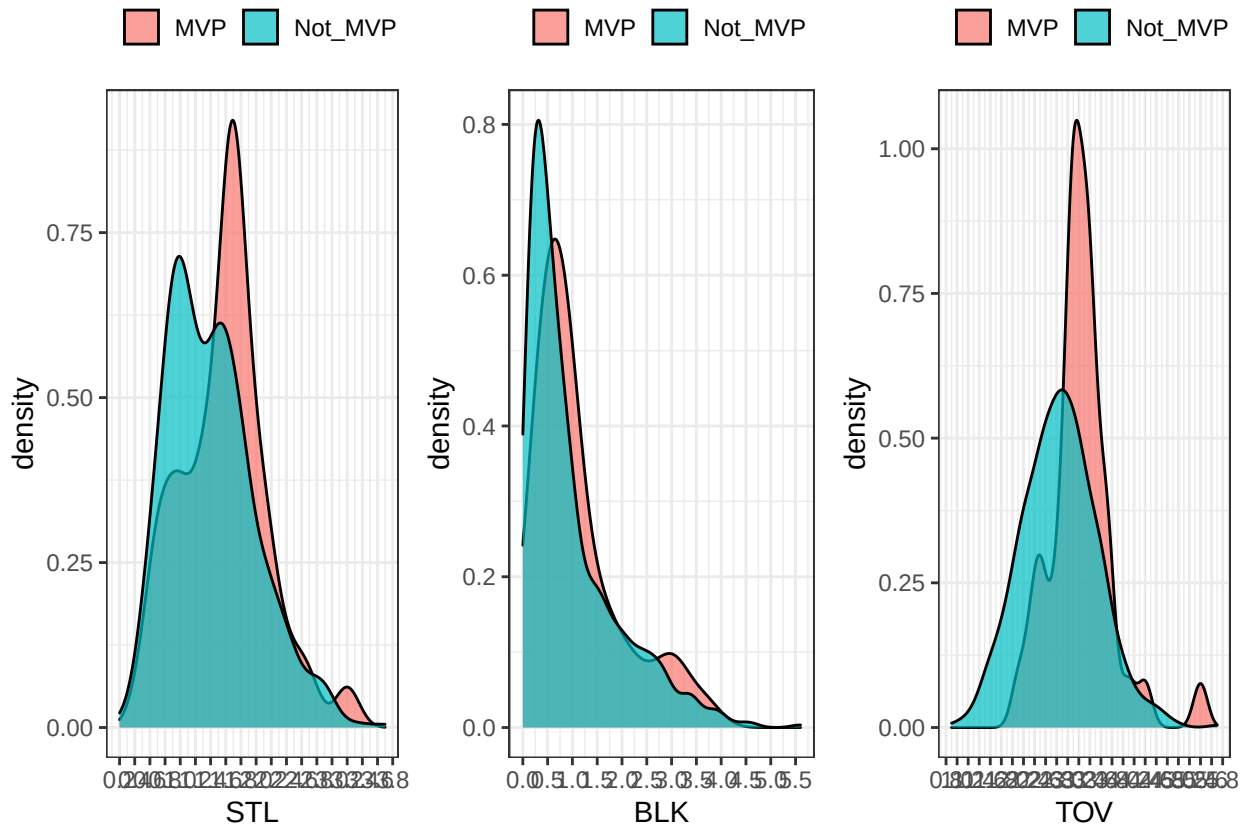
Predpostavljamo torej, da skoki igrajo vlogo pri odločanju o nagradi MVP. Po našem prepričanju je vseeno ali skoke obravnavamo skupno ali ločeno, saj čeprav je odstotna razlika v napadalnih skokih majhna, ti enako ali pa še bolj prispevajo h "cenjenju" igralca kot obrambni skoki. Tudi razlika v porazdelitvi vseh treh vrst skokov med skupinama igralcev je podobna.



Posebej velika razlika med skupinama je tudi pri podajah, kjer MVP-ji v povprečju dosežejo 6.3 podaje na tekmo, ne MVP-ji pa 4.9 podaje na tekmo. Razlika je 21%. Tudi podaje tako predstavljajo pomemben kriterij vlogo pri odločevalcih.



Razlika med skupinama je pri atributih ukradenih žog, blokad, izgubljenih žog in prekrškov po absolutni vrendosti majhna (zaradi majhnih vrednosti atributov), relativna razlika pa je okoli 11%. Sami menimo, da ti atributi ne vplivajo značilno na vrednost Share ali MVP ter bi lahko zgolj zmedli model. Opozoriti velja morda na zanimivost, da MVP-ji izgubijo večje število žog kot ne MVP-ji, kar pa je posledica večje posesti žoge in tako večje možnosti izgube žoge vendar morda enakega deleža glede na število posesti (v nadaljevanju).



Pri koeficientih elastic net modela se izkaže, da so vse spremenljivke značilne, vendar ukradene žoge in blokade mejno, kot smo ugotovili že sami. Pri prečnem preverjanju z modeli uporabljenimi na vseh atributih dobimo naslednje rezultate:

Metrika	Elastic.Net	Random.Forest	Light.GB	Extreme.GB
MAE	0.19	0.181	0.191	0.194
Občutljivost	0.281	0.219	0.219	0.219
Pravilne napovedi	9/32	7/32	7/32	7/32

Z vsemi atributi pojasnimo 19% variabilnosti Share, kar je v primerjavi z atributi absolutnega meta (23%) relativno veliko. Vidimo pa, da so rezultati oziroma kakovosti napovedi v primerjavi z ostalimi relativno slabi ali pa povprečni. To pomeni, da skoki, podaje in ostali atributi razložijo nekolikšen del Share in MVP, vendar ne pripomorejo bistveno. Še vedno, ostajajo napovedi MVP-ja slabe.

2.3 Napredna statistika

Tudi napredne igralske statistike bomo razdelili v dva sklopa. Prvi sklop bo predstavljal zgoraj opisane ostale osnovne statistike, vendar izražene odstotkovno, drugi sklop atributov pa združene spremenljivke, izračunane prek več različnih faktorjev.

2.3.1 Relativne ostale osnovne statistike

- TRB% (Total Rebounds Percentage) - Odstotni delež skokov med igranjem

- ORB% (Offensive Rebounds Percentage) - Odstotni delež napadalnih skokov med igranjem
- DRB% (Defensive Rebounds Percentage) - Odstotni delež obrambnih skokov med igranjem
- AST% (Assists Percentage) - Odstotni delež košov kot posledica podaje med igranjem
- STL% (Steals Percentage) - Odstotni delež nasprotnikove posesti, ki so se končale z ukradeno žogo med igranjem
- BLK% (Blocks Percentage) - Odstotni delež nasprotnikovih metov, ki so se končali z blokado med igranjem
- TOV% (Turnovers Percentage) - Odstotni delež izgubljenih žog med igranjem

Vidimo, da imamo opravka s podobnimi statistikami kot pri relativnih vrednostih metov, le da so te izražene v odstotnem deležu glede na ekipo. Ponovno bomo relativne statistike primerjali z zgornjimi absolutnimi in videli čemu dajejo odločevalci večji poudarek, čeprav smo predspotavili, da ne vplivajo značilno.

Ponovno pri vseh atributih velja podobno kot pri absolutnih vrendostih, le da so odstotne razlike razlike med skupinama (ne v smislu absolutne razlike vrednosti spremenljivke) za odtenek manjše. Porazdelitve atributov so torej podobne, le nekoliko bolj se porazdelitvi skupin prekrivata. V kolikor pogledamo pojasnjen delež variabilnosti Share v samostojnem linearnem modelu velja, da je pri skokih ponovno podobno kot pri metu, da absolutne vrednosti spremenljivk pojasnijo večji delež variabilnosti.

Tudi rezultati oziroma koeficienti elastic net modela kažejo na enako značilnost atributov kot pri njihovih absolutnih vrednostih. Razlika je samo v izgubljenih žogah, kjer atribut postane neznačilen, kar dodatno kaže na to, da imajo MVP-ji več žogo v svojih rokah in tako enak delež izgubljenih žog kot ostali igralci.

Pri prečnem preverjanju z modeli uporabljenimi na vseh spremenljivkah dobimo naslednje rezultate:

Metrika	Elastic.Net	Random.Forest	Light.GB	Extreme.GB
MAE	0.202	0.197	0.199	0.203
Občutljivost	0.344	0.188	0.188	0.156
Pravilne napovedi	11/32	6/32	6/32	5/32

Kljub informaciji, da atributi v osnovnem linearnem modelu pojasnijo 23% variabilnosti Share, na zgornjih rezultatih vidimo, da imajo odstotne vrednosti atributov, torej nekoliko naprednejše, v primerjavi s prej predstavljenimi osnovnimi statistikami veliko manjšo napovedno moč, vsaj pri Share. Tudi to potrdi domnevo, da američani oziroma odločevalci bolj kot na izkoriščanje igralnih minut na parketu gledajo na absolutne vrednosti atributov oziroma številke.

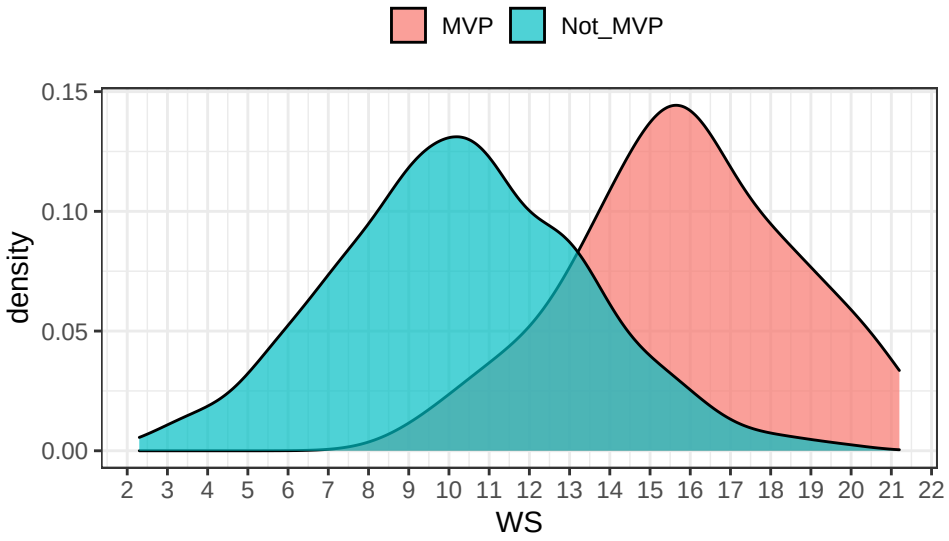
2.3.2 Združene statistike

- WS (Win Shares) - Ocena števila zmag, ki jih je prispeval igralec
- WS.48 (Win Shares Per 48 Minutes) - Ocena števila zmag, ki jih prispeva igralec na 48 odigranih minut povprečje lige je približno 0.100
- OWS (Offensive Win Shares) - Ocena števila zmag, ki jih je prispeval igralec v napadu
- DWS (Defensive Win Shares) - Ocena števila zmag, ki jih je prispeval igralec v obrambi
- BPM (Box Plus/Minus) - Ocena števila točk na 100 posesti, ki jih je igralec prispeval nad povprečnim igralcem lige, prevedeno v povprečno ekipo
- OBPM (Offensive Box Plus/Minus) - Ocena števila točk v napadu na 100 posesti, ki jih je igralec prispeval nad povprečnim igralcem lige, prevedeno v povprečno ekipo
- DBPM (Defensive Box Plus/Minus) - Ocena števila točk v obrambi na 100 posesti, ki jih je igralec prispeval nad povprečnim igralcem lige, prevedeno v povprečno ekipo
- VORP (Value over Replacement Player) - Ocena števila točk na 100 posesti ekipe, ki jih je igralec prispeval nad igralcem na nadomestni ravni (-2,0), prevedena v povprečno ekipo in sorazmerno razporejena v sezoni 82 tekem
- PER (Player Efficiency Rating) - Merilo produkcije na minuto, standardizirano tako, da je povprečje lige 15, sešteje vse igralčeve pozitivne dosežke, odšteje negativne dosežke in vrne minutno oceno igralčeve uspešnosti

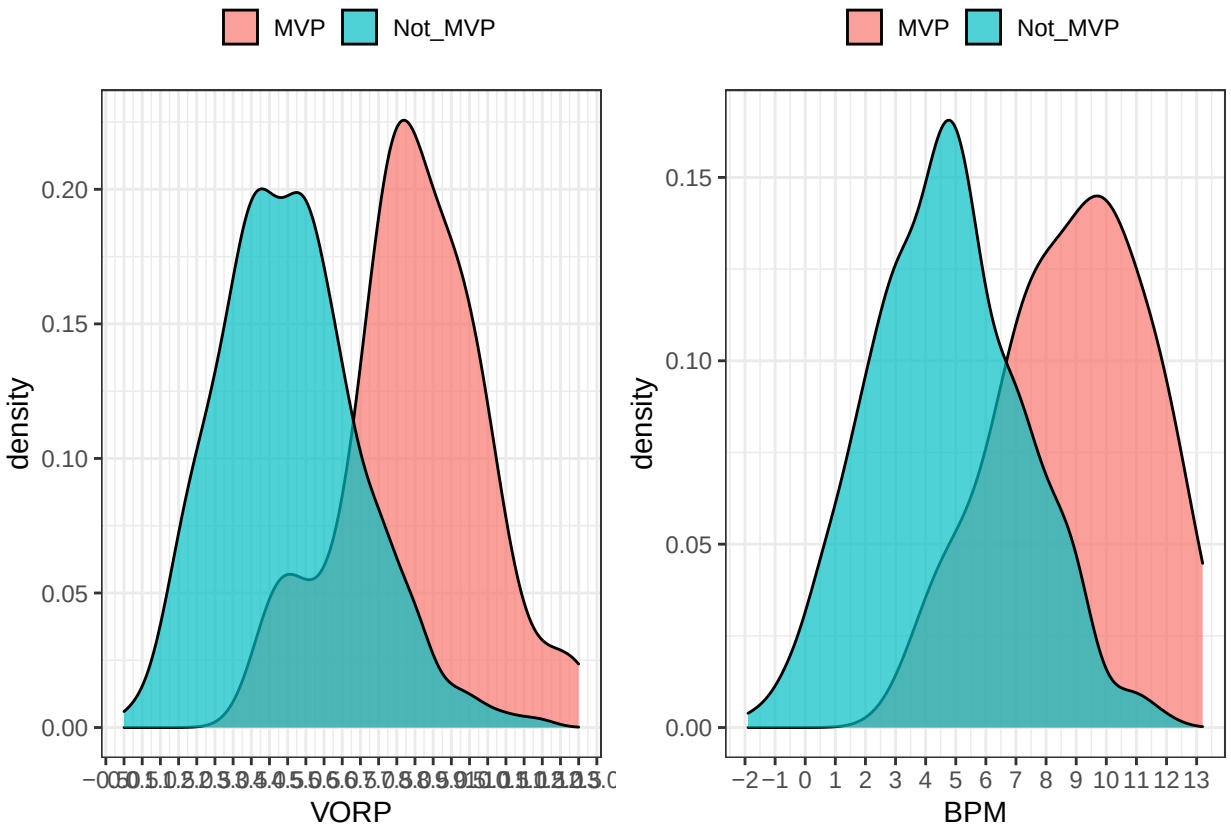
- USG% (Usage Percentage) - Ocena deleža ekipnih akcij, ki jih vključujejo igralca, ko je bil na parketu

Vidimo, da opisane spremenljivke združujejo več faktorjev oziroma spremenljivk in tako najverjetneje najbolj opisujejo Share oziroma MVP.

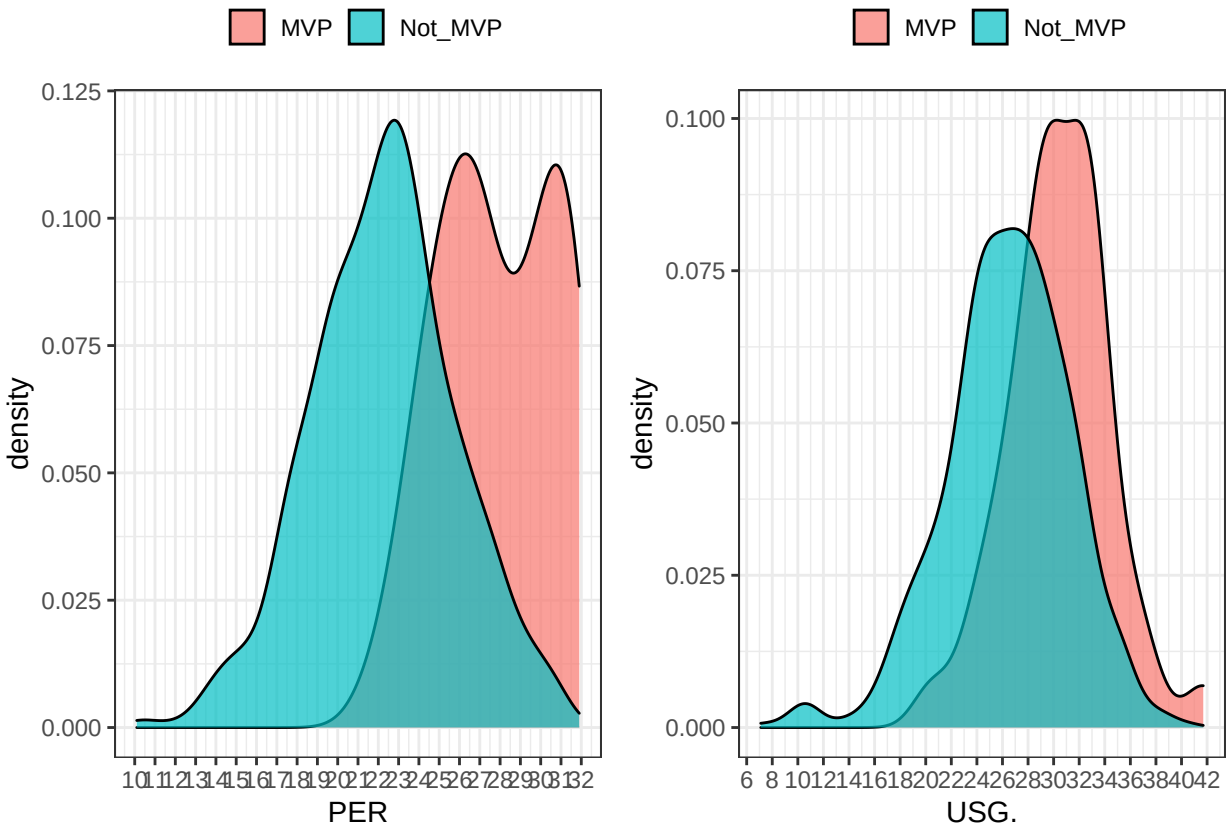
Močno korelirani so atributi WS, WS.48, OWS in DWS. Največji delež variabilnosti Share v samostojnem linearnem modelu pojasni WS. Ponovno vidimo, da je pomembnejša absolutna vrednost v primerjavi z relativno vrednostjo (WS.48). Podobno kot pri skokih velja, da je je večja razlika med skupinama in tako pojasnjena variabilnost v napadu v primerjavi z obrambo. MVP-ji v povprečju prispevajo 5.5 zmag več kot ne MVP-ji (15.8 proti 10.3). Razlika je 35% medtem ko je razlika na 48 odigranih minut manjša in sicer 30%. V napadu MVP-ji v povprečju prispevajo 4.2 zmage več kot ne MVP-ji (39%) medtem ko je razlika v obrambi 1.3 zmage (26%). Tudi na podlagi visoke vrendosti $R^2 = 0.38$ linearnega modela s samostojnim atributom WS, vidimo, da atribut močno vpliva na Share in tako MVP.



Tudi BPM atribut je podobno kot že velikokrat, razdeljen na skupno vrendost, napad in obrambno. Podobno informacijo nosi tudi atribut VORP, tako da so te atributi med seboj močno korelirani. Velja, da MVP-ji v povprečju na 100% prispevajo 4.5 točk več kot ne MVP-ji (8.9 proti 6.4). Razlika je kar 48%. Ponovno, število točk v napadu je večje od števila točk v obrambi, le da je tokrat razlika med skupinama na strani obrambe. Podobno velja tudi za vrednosti VORP, kjer MVP-ji dosežejo 3.4 točke več na 100 posesti in razlika je 43%.



Na koncu pa poglejmo še atributa PER in USG%. Razlika med skupinama je v povprečju pri teh dveh atributih nekoliko manjša in sicer 20% in 13%. MVP-ji so na minuto produktivni za 5.5 bolj kot ne MVP-ji (oboji so visoko nad povprečjem - nabor kandidatov). Pri USG% pa ponovno velja, kot že velikokrat, možnost napačne interpretacije, vendar pa so MVP-ji v povprečju udeleženi v 30.5% akcij ekip, ne MVP-ji pa v 36.4% akcij svojih ekip, razlika je 13%.



Na podlagi koeficientov elastic net modela razberemo, da sta zanimivo BPM in VORP atributa mejno statistično značilna, medtem ko so ostali atributi močno značilni. V osnovnem linearnem modelu vsi atributi skupaj pojasnjujejo kar 47% variabilnosti Share.

Pri prečnem preverjanju z modeli uporabljenimi na vseh spremenljivkah dobimo naslednje rezultate:

Metrika	Elastic.Net	Random.Forest	Light.GB	Extreme.GB
MAE	0.15	0.131	0.138	0.141
Občutljivost	0.625	0.531	0.406	0.5
Pravilne napovedi	20/32	17/32	13/32	16/32

OB 47% pojasnjeni variabilnosti Share v osnovnem linearnem modelu tudi pri zgornjih rezultatih vidimo znatno izboljšanje napovedi. Vidimo torej, da ti napredni atributi prispevajo večinski delež h pojasnjevanju Share in tako odločevalčevih izbirah MVP. Velja seveda, da so ti atributi pridobljeni iz združenih statistik vendar torej na nek poseben način, ki se v posameznih statistikah, ali pa vsaj odločevalčevem mišljenju ne pokaže enostavno.

3. Napoved sezone 2021/22

V zadnjem delu naloge pa na podlagi celotnih testnih podatkov (1980/81 - 2020/21) napovejmo MVP igralca lige NBA sezone 2021/22. Glavni kandidati za MVP igralca so Nikola Jokić (Denver Nuggets), Jel Emiid (Philadelphia 76ers), Giannis Antetokounmpo (Milwaukee Bucks), Devin Booker (Phoenix Suns), Luka Dončić (Dallas Mavericks), Jayson Tatum (Boston Celtics), Ja Morant (Memphis Grizlies), Stephen Curry (Golden State Warriors), Chris Paul (Phoenix Suns), DeMar DeRozan (Toronto Raptors), Kevin Durant (Brooklyn Nets) in LeBron James (Los Angeles Lakers).

Za izbiro modela oziroma vključenih atributov tokrat preizkusimo še s prečnim preverjanjem na vseh razpoložljivih spremenljivkah in poglejmo rezultate.

Metrika	Elastic.Net	Random.Forest	Light.GB	Extreme.GB
MAE	0.139	0.122	0.113	0.12
Občutljivost	0.688	0.625	0.656	0.625
Pravilne napovedi	22/32	20/32	21/32	20/32

Vidimo, da ob uporabi popolnoma vseh atributov v primerjavi zgolj z združenimi atributi ponovno pride do vidnega izboljšanja napovedi in tako združeni atributi vseeno ne nosijo vse informacije (ki jo imamo na voljo) glede odločitve o podanih glasovih in tako vrednosti Share, vplivajo tudi ostali, bolj osnovni atributi. Tudi delež pojasnjene variabilnosti Share v osnovne linearnem modelu z vsemi atributi se precej poveča in je enak 60%. Za najboljšega izmed modelov bi lahko izbrali light gradient boost model, ki ima najboljše uteženo povprečje med MAE in pravilnimi napovedmi.

Sedaj pa z uporabo light gradient boost modela na vseh atributih poglejmo napovedi za sezono 2021/22 in jih primerjajmo z dejanskimi.

Igralec	Ekipa	Dejanski MVP	Napovedan MVP	Dejanski Share	Napovedan Share
Nikola Jokić	DEN	MVP	MVP	0.875	0.748
Joel Embiid	PHI	Not MVP	Not MVP	0.706	0.397
Giannis Antetokounmpo	MIL	Not MVP	Not MVP	0.595	0.590
Devin Booker	PHO	Not MVP	Not MVP	0.216	0.199
Luka Dončić	DAL	Not MVP	Not MVP	0.146	0.144
Jayson Tatum	BOS	Not MVP	Not MVP	0.043	0.135
Ja Morant	MEM	Not MVP	Not MVP	0.010	0.204
Stephen Curry	GSW	Not MVP	Not MVP	0.004	0.009
Chris Paul	PHO	Not MVP	Not MVP	0.002	0.133
DeMar DeRozan	CHI	Not MVP	Not MVP	0.001	0.047
Kevin Durant	BRK	Not MVP	Not MVP	0.001	0.103
LeBron James	LAL	Not MVP	Not MVP	0.001	0.180

Vidimo, da smo pravilno napovedali MVP igralca, to je Nikola Jokić, vendar pa smo delež glasov (Share) zanj nekoliko podcenili. Pri deležu glasov smo se najbolj zmotili pri Joelu Embiidu, ki smo ga z dejanskega drugega mesta uvrstili na tretje mesto za Giannisa Antetokounmpa. Visok (previsok) delež glasov smo napovedali predvsem tistim z najmanjšim deležem glasov. Povprečna absolutna razlika med napovedano vrednostjo Share in dejansko vrednostjo Share tj. MAE je bila 0.101. Naš model torej nekoliko preveč optimistično napoveduje nizke vrednosti Share. To je popolnoma razumljivo in je posledica majhnih razlik v statističnih podatkih končnega nabora kandidatov za nagrado MVP, katerih razlike se ne odražajo v deležu glasov v realnem svetu, kjer imajo odločevalci praviloma le dva ali tri favorita med katerimi se večina odloča oziroma jih uvršča na najvišja mesta, čeprav nimajo nujno tako boljše statistike od ostalih.

4. Zaključek

V namen razumevanja izbire najkoristnejšega igralca rednega dela lige NBA, tako imenovanega MVP igralca smo z uporabo nekaterih modelov strojnega učenja preučili vpliv in moč vpliva izbranih atributov oziroma statistik, ki jih lahko pripišemo posameznem igralcu. Raziskavo smo zastavili kot napovedni problem deleža prejeta glasov za nagrado MVP in tako napoved MVP igralca posamezne sezone.

Na podlagi sezon 1980/81 - 2020/21 smo opisno predstavili različne atribute, preverili njihov vpliv na delež prejetih glasov in nagrado MVP in preverili moč vpliva na delež prejetih MVP glasov iz katerega smo izračunali tudi moč vpliva na nagrado MVP.

Ugotovili smo, da največji delež variabilnosti Share pojasnijo iz več atributov sestavljene statistike oziroma atributi kot so ocena števila zmag, ki jih je prispeval igralec, ocena števila točk na 100 posesti, ki jih je prispeval igralec, ocena števila točk na 100 posesti, produkcija igralca na minuto in ocena deleža ekipnih akcij, ki so vključevale igralca.

Na podlagi statistik igralcev v povezavi z metom na koš ter nekaterih ostalih statistik kot so skoki, podaje, ukradene žoge in blokade smo odgovorili tudi na vprašanje kaj pri odločevalcih nagrade MVP bolj velja: absolutne vrednosti statistik, njihove relativne vrednosti tj. različni odstotki, ki predstavljajo kakovost izvedbe, ali pa vrednosti statistik glede na igralni čas. Pokazalo se je, da največ štejejo absolutne vrendosti statistik kot so same točke, skoke in podaje in ne odstotek zadetih metov, odstotek uspešnih podaj ali pobranih žog pod obročem. S tem smo morda tudi razložili zakaj je ameriška košarka taka kot je, igralci in ekipe potemtakem stremijo h čim večjim številkam in ne kakovostni realizaciji.

Na koncu smo kot na učnih podatkih, torej vseh obravnavanih sezonah zgradili tudi napovedni model in napovedali deleže prejetih glasov ter na podlagi njih nagrado MVP za redni del sezone 2021/22. Tega smo pravilno napovedali da je Nikola Jokić, bolj pa velja povedati, da naš model nekoliko preveč optimistično napoveduje nizke vrednosti deležev glasov. To je posledica predvsem majhnih razlik v statističnih podatkih končnega nabora kandidatov za nagrado MVP, katerih razlike se ne odražajo v deležu glasov v realnem svetu.

V zakup velja vzeti tudi, da je "popolne" napovedi v našem primeru zelo težko (nemogoče) doseči, saj je izbira odločevalcev subjektivna in temelji tudi na nekaterih spremenljivkah, ki niso merljive in jih v naše modele ne moremo vključiti in včasih ne sovpada popolnoma z statistikami, ki so nam na voljo in tako napovedmi, ki bi bile nekako bolj logične in jih algoritem lahko napove.