

Analiza NBA Igrača Kroz Eksploraciju Podataka

Primena Naprednih Tehnika Istraživanja Podataka na NBA Draft Combine Statistike
Matematički fakultet

Andela Jovanović
mi251035@alas.matf.bg.ac.rs

Januar 2026.

Sažetak

Ovaj rad analizira NBA Draft Combine statistike primenom mašinskog učenja i statističkih metoda. Na osnovu spajanja tri izvora podataka formiran je skup sa 61 karakteristikom i 1.202 igrača. Identifikovana su dva dominantna fizička arhetipa uz tačnost klasifikacije veću od 98%, kao i 129 asocijativnih pravila između fizičkih karakteristika. Takođe, pokazano je da 10 glavnih komponenti objašnjava 90% varijanse podataka, što ukazuje na visoku redundanciju merenja.

Sadržaj

1 Uvod	3
1.1 Kontekst i Motivacija	3
1.2 Skup Podataka	3
2 Analiza skupa podataka	4
2.1 Priprema Podataka	4
2.2 Eksplorativna Analiza Podataka	4
2.3 Korelaciona Analiza	5
3 Detekcija Anomalija	6
4 Redukcija Dimenzionalnosti	7
4.1 Analiza Glavnih Komponenti (PCA)	7
4.2 Interpretacija Glavnih Komponenti	7
4.3 Vizualizacija u Redukovanom Prostoru	8
4.4 t-SNE: Alternativna Perspektiva	10
5 Klasterovanje	11
5.1 Određivanje Optimalnog Broja Klastera	11
5.2 K-Means Klasterovanje sa $k=2$	11
5.3 Alternativni pristupi klasterovanju	13
6 Klasifikacija	14
6.1 Predviđanje Stvarnih Košarkaških Pozicija	14
6.2 Predikcije Draft Statusa	16
7 Pravila Pridruživanja	17
7.1 Otkrića	17
7.2 Najinteresantnija Pravila	17
7.3 Praktična Primena Pravila	18
8 Vremenska Dimenzija	19
8.1 Šta Ova Stabilnost Znači?	20

9 Zaključak	21
9.1 Ograničenja	21
9.2 Budući pravci	21
9.3 Završna Refleksija	21
Literatura	22
Zahvalnice	22

1 Uvod

1.1 Kontekst i Motivacija

NBA Draft Combine je godišnji događaj gde se košarkaši podvrgavaju standardizovanim fizičkim i atletskim testovima. Ova merenja pružaju skautima i timovima objektivne podatke za evaluaciju kandidata, ali sa preko 50 različitih metrika, sama količina informacija može biti opterećujuća. Kako prepoznati koje karakteristike definišu uspešnog košarkaša? Koje kombinacije fizičkih atributa idu zajedno, a koje su retke i potencijalno vredne?

U periodu kada se odluke vredne milione dolara donose na osnovu ovih merenja, NBA timovi se susreću sa bitnim pitanjima. Nije dovoljno samo pogledati da li je igrač visok ili koliko brzo trči, potrebno je razumeti složene odnose između visine i agilnosti, snage i eksplozivnosti, veličine i brzine. Postoje li prirodne grupe igrača koje diktiraju različite stilove igre? Može li se iz fizičkih merenja predvideti na kojoj poziciji će igrač biti najkorisniji? I možda najvažnije - koji igrači predstavljaju statistički izuzetak koji može signalizirati potencijalni izuzetan talent?

Tradicionalni pristupi evaluaciji često se oslanjaju na iskustvo skauta i intuiciju, što ima svoju vrednost, ali nosi i rizik subjektivnosti. Ovaj projekat pokušava da dopuni taj proces sistematskim pristupom zasnovanim na mašinskom učenju i statističkoj analizi, kako bi iz podataka izvukli objektivne obrasce koji mogu pomoći u donošenju boljih odluka.

1.2 Skup Podataka

Za potrebe ove analize korišćen je skup podataka preuzet sa Kaggle platforme [1], koji obuhvata višegodišnje statistike NBA Draft Combine događaja. Finalni skup podataka je rezultat spajanja 1.202 igrača sa 61 karakteristikom, obuhvatajući period od 2000. do 2024. godine.

Atribut	Vrednost
Izvor	NBA baza podataka (Kaggle)
Broj zapisa	1.202 igrača (nakon spajanja)
Broj karakteristika	61 kolona
Vremenski period	Više sezona (2000-2024)
Tipovi podataka	Numerički, kategorički, vremenski

Tabela 1: Pregled skupa podataka

2 Analiza skupa podataka

2.1 Priprema Podataka

Polaznu tačku ove analize čine tri odvojena skupa podataka iz NBA baze, koji zajedno pružaju kompletan pregled fizičkih karakteristika, draft istorije i biografskih informacija igrača.

Prvi i najvažniji je skup statistika sa probnog kampa za draft¹ koji sadrži sve fizičke testove od osnovnih merenja poput visine i težine, preko atletskih testova kao što su vertikalni skok i trčanje na kratke staze, do specifičnijih merenja kao što su raspon ruku i veličina šake. Ovaj skup čini osnovu analize jer dokumentuje direktna fizička merenja 1.202 igrača kroz 47 različitih kolona.

Drugi skup, istorija izbora na draftu, donosi kontekst o tome kako su ovi igrači procenjeni od strane NBA timova to jest u kojoj rundi i na kojem mestu su izabrani, od strane kog tima. Ove informacije su ključne jer omogućavaju da se ispita da li fizičke karakteristike zaista predviđaju kako će skauti rangirati igrače.

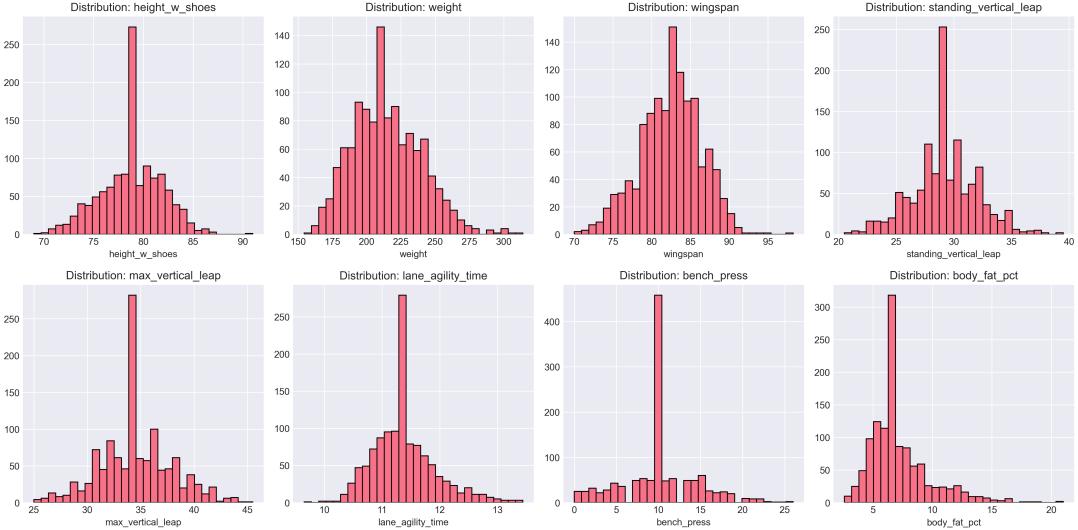
Konačno, treći skup sa opštim informacijama o igračima dodaje biografski kontekst kroz podatke o datumu rođenja, univerzitetu ili inostranoj ligi iz koje dolaze, i zemlji porekla.

Spajanjem ovih tabela preko jedinstvenog identifikatora igrača dobijen je skup podataka od 1.202 zapisa sa 61 kolonom. Međutim, kvalitet podataka je predstavljaо najveći izazov jer mnoge kolone su imale značajan broj nedostajućih vrednosti. Odlučeno je da se kolone sa više od 70% nedostajućih podataka uklone iz analize, jer bi njihova analiza bila statistički nepouzdana i mogla dovesti do pogrešnih zaključaka. Ovo je posebno uticalo na metrike šutiranja, koje su često bile nedostupne jer mnogi igrači ne učestvuju u tim testovima, bilo zbog povreda, saveta svojih agenata, ili zato što nisu bili pozvani na kompletno testiranje.

Za preostale nedostajuće vrednosti primenjena je strategija popunjavanja medijanom. Ova metoda je odabrana umesto proseka jer je medijana otpornija na ekstremne vrednosti koje su česte u sportskim podacima.

2.2 Eksplorativna Analiza Podataka

Pre primene složenih algoritama mašinskog učenja, bilo je neophodno razumeti strukturu i distribuciju samih podataka. Početna eksplorativna analiza je otkrila zanimljive obrasce koji potvrđuju da se radi o specijalnoj populaciji - elitnim sportistima, a ne opštoj populaciji.



Slika 1: Distribucije ključnih fizičkih karakteristika

Slika 1 prikazuje distribucije ključnih fizičkih merenja i odmah su vidljivi određeni obrasci. Većina

¹Draft Combine - godišnji dogadjaj gde se potencijalni NBA igrači podvrgavaju standardizovanim testovima

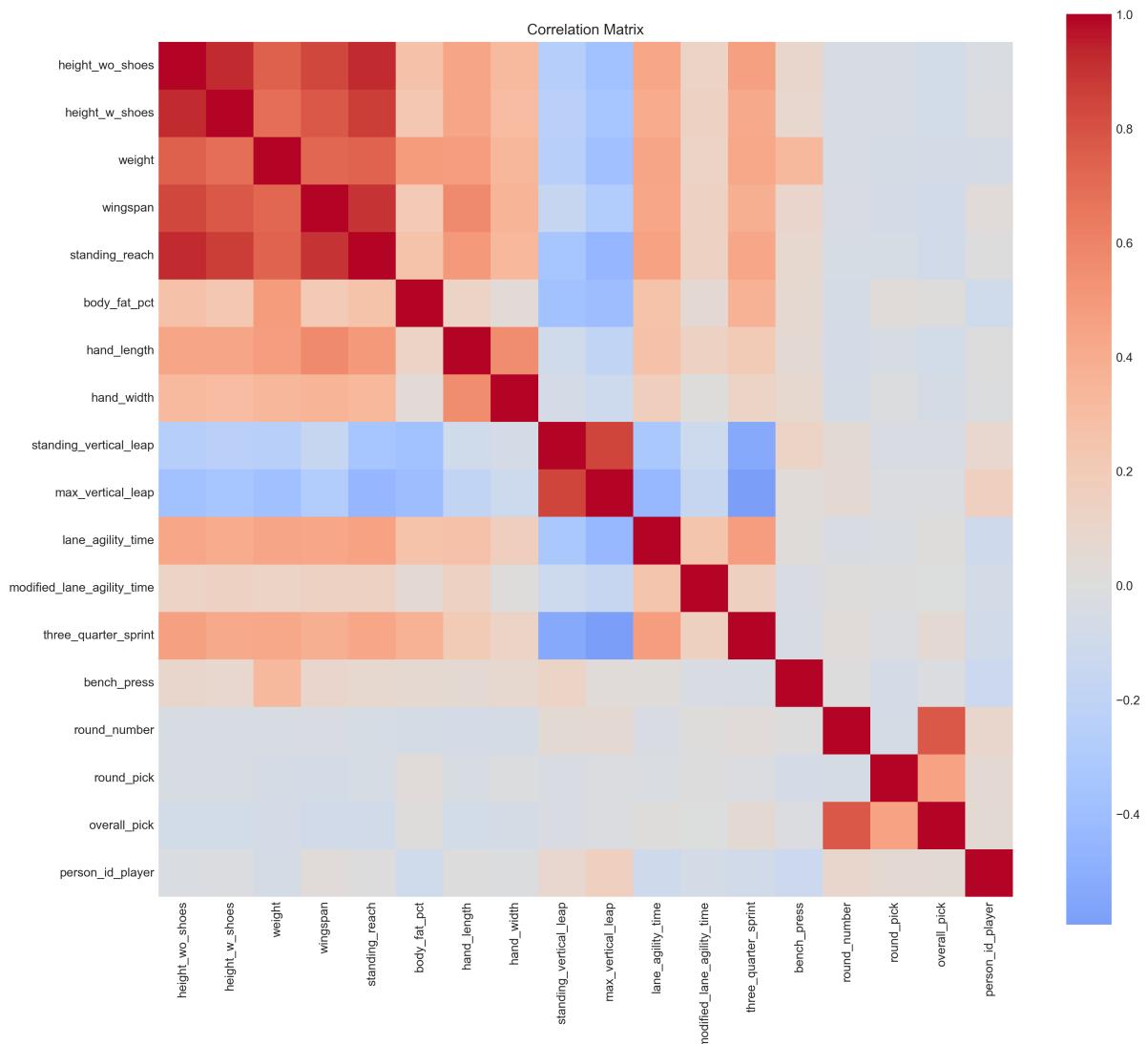
distribucija prati normalnu raspodelu, što je dobar znak za kvalitet podataka i kasnije statističke analize. Ovo nije iznenadjuće jer visina, težina i slične biološke karakteristike prirodno prate normalnu distribuciju u velikim populacijama, čak i kada se radi o vrhunskim sportistima.

Visina sa patikama se koncentriše oko 79-80 inča (≈ 201 cm) što je očekivano za NBA igrače. Međutim, interesantno je da su same granice prihvatljivosti relativno uske, budući da se većina igrača nalazi u opsegu od 75 do 85 inča (190–215 cm). Ovo govori o tome da NBA ima određene „ulazne standarde” što se tiče visine, izuzimajući retke izuzetke.

Težina pokazuje širu distribuciju jer se kreće od oko 175 do 275 lbs ($\approx 80 – 125$ kg), što pokazuje veliku raznolikost u tipovima igrača - od brzih, lakih bekova² do masivnih centara³. Raspon ruku takođe pokazuje normalnu distribuciju sa blagom asimetrijom ka većim vrednostima, što ima smisla jer se NBA timovi često fokusiraju na igrače sa dugim rukama koji mogu pokrивati veći prostor na terenu.

2.3 Korelaciona Analiza

Nakon razumevanja pojedinačnih distribucija, sledeći korak je bio ispitati kako karakteristike međusobno koreliraju.



Slika 2: Korelaciona matrica

²Bek (eng. guard) - spoljni igrač zadužen za organizaciju igre i šutiranje sa veće distance

³Centar (eng. center) - najviši igrač u timu, zadužen za igru pod košem, skokove i blokade

Najveće pozitivne korelacije postoje između različitih merenja veličine kao što su visina bez patika, visina sa patikama, raspon ruku i doseg u stojećem položaju. Ove korelacije su vrlo jake, često prelaze 0.7, što je razumljivo jer su sve determinisane osnovnom skeletnom strukturu i genetikom. Međutim, jačina ovih korelacija kod NBA igrača je zapravo veća nego u opštoj populaciji, što sugerira da NBA regrtuje morfološki homogeniju grupu igrača.

Ono što je još interesantnije su pozitivne korelacije između veličine i testova agilnosti. Veći, teži igrači su statistički sporiji u testovima trčanja na kratke staze i imaju duža vremena u testovima agilnosti. Ovo otkriva fundamentalni kompromis u košarci: dobijate na fizičkoj prisutnosti i dometu, ali gubite na brzini i pokretljivosti. Ovaj princip objašnjava zašto postoje različite pozicije u košarci jer sport zahteva i velike igrače koji dominiraju pod košem, i brze igrače koji kontrolisu tempo i spoljašnji deo terena.

Neočekivan nalaz je da rezultati u testu potiska sa klupe⁴ pokazuju vrlo slabe korelacije sa gotovo svim ostalim varijablama, uključujući visinu i težinu. Ovo sugerira da je snaga gornjeg dela tela relativno nezavisna karakteristika koja varira među igračima bez obzira na njihovu građu.

Možda najzanimljiviji uvid dolazi iz analize draft pozicije. Variable poput `overall_pick` (ukupna pozicija izbora) i `round_pick` (pozicija izbora u rundi) pokazuju slabe korelacije sa fizičkim merenjima, što ukazuje da NBA timovi pri odabiru igrača ne pridaju presudnu važnost fizičkim atributima. Veština, košarkaški IQ i procenjeni potencijal očigledno imaju veću ulogu u procesu drafa.

3 Detekcija Anomalija

Pre nego što krenemo sa grupisanjem igrača u prirodne kategorije, bilo je važno identifikovati statističke izuzetke, to jest igrače čije kombinacije fizičkih karakteristika ne odgovaraju tipičnim obrascima. Za ovo su korišćene dve metode: Isolation Forest [2]⁵ i Local Outlier Factor⁶, koje pristupaju problemu iz različitih perspektiva ali su dale veoma konzistentne rezultate.

Metoda	Parametri	Broj netipičnih vrednosti	Procenat (%)
Isolation Forest	Kontaminacija = 0.10	121	10.07
Local Outlier Factor (LOF)	20 suseda, kontaminacija = 0.10	121	10.07

Tabela 2: Rezultati detekcije anomalija

Obe metode su identifikovale 121 igrača, odnosno 10.07% ukupne populacije, kao netipične vrednosti. Ova konzistentnost između dva različita algoritma povećava poverenje u rezultate. Ali ko su zapravo ovi igrači i šta znači biti netipična vrednost u NBA kontekstu?

U većini analiza podataka, netipične vrednosti se smatraju greškama ili šumom koji treba ukloniti. Međutim, u sportu, a posebno u NBA-u, netipične vrednosti mogu biti izuzetno vredni. Ovo su igrači sa jedinstvenim kombinacijama fizičkih karakteristika - možda izuzetno visok igrač koji je neuobičajeno brz, ili relativno nizak igrač sa nenormalnim rasponom ruku, ili centar sa agilnošću beka.

Ovi profili mogu predstavljati dva scenarija. Prvi scenario je pozitivan i reč je o potencijalnim zvezdama, igračima koji prevazilaze tradicionalne pozicione okvire i imaju potencijal da transformišu način igre. Kevin Durant, na primer, kombinuje visinu centra sa veštinama beka, što ga čini statističkom anomalijom ali i jednim od najboljih igrača svoje generacije. Drugi scenario je rizičniji jer su ovo igrači koji ne odgovaraju nijednom utvrđenom profilu i mogu imati poteškoća da pronađu svoju ulogu u timu.

Za ostatak analize, ovih 121 igrača je isključeno kako bi se fokus stavio na „tipičnu“ populaciju od 1.081 igrača. Ovo omogućava jasnije definisane klastere i robusnije statističke zaključke o reprezentativnim profilima. Međutim, vrednost identifikacije ovih netipičnih vrednosti ne treba zanemariti jer za NBA tim koji traži „sledeću veliku stvar“, upravo ovi igrači mogu biti najveća prilika.

⁴Potisak sa klupe (eng. bench press) - test snage gornjeg dela tela gde igrač podiže tegove ležeći na klupi

⁵Isolation Forest - algoritam za detekciju anomalija zasnovan na izolaciji tačaka pomoću stabala odlučivanja

⁶Local Outlier Factor (LOF) - algoritam koji identificiše anomalije na osnovu lokalne gustine tačaka u prostoru

4 Redukcija Dimenzionalnosti

4.1 Analiza Glavnih Komponenti (PCA)

Sa 17 različitih numeričkih fizičkih karakteristika u skupu podataka, vizualizacija i razumevanje strukture podataka postaje izazovno. Tu nastupa analiza glavnih komponenti (eng. Principal Component Analysis - PCA), tehnika koja transformiše te karakteristika u novi skup nekoreliranih glavnih komponenti, rangirajući ih po tome koliko varijanse u podacima objašnjavaju.

Kada je PCA primenjena na naših 1.081 igrača, rezultati su bili zanimljivi i praktično veoma korisni. Prva glavna komponenta (PC1) sama po sebi objašnjava 31.69% ukupne varijanse u svim merenjima. Dodavanjem druge komponente (PC2) koja doprinosi dodatnih 12.48%, dolazimo do ukupno 44.17% varijanse objašnjene sa samo dve dimenzije. Ali još impresivnije je da svega 10 komponenti objašnjava 90% ukupne varijanse.

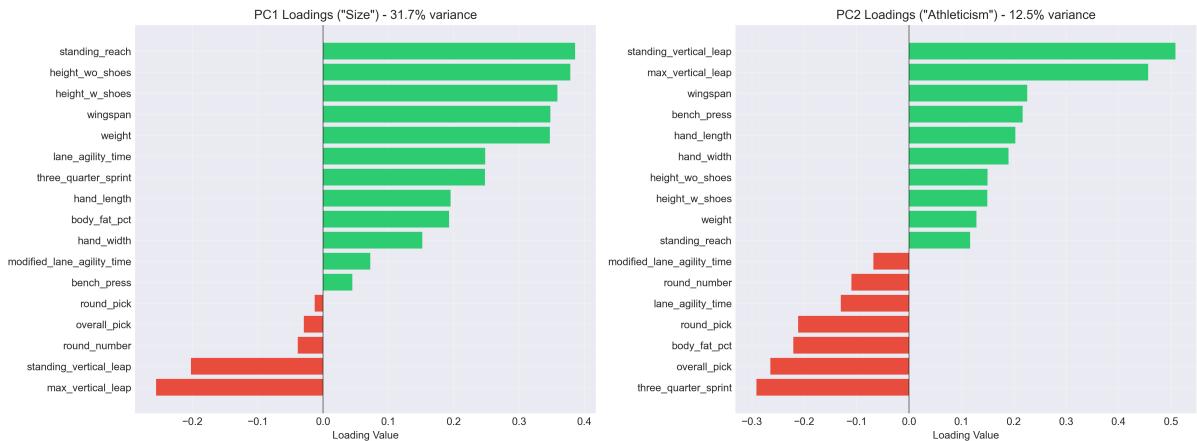
Metrika	Vrednost
Broj komponenti za 90% varijanse	10
Broj komponenti za 95% varijanse	11
Varijansa prve komponente (PC1)	31.69%
Varijansa druge komponente (PC2)	12.48%
Varijansa treće komponente (PC3)	10.78%
Varijansa četvrte komponente (PC4)	7.49%
Varijansa pete komponente (PC5)	6.94%

Tabela 3: PCA rezultati prvih pet komponennti

Možemo se fokusirati na 10 ključnih dimenzija sa gubitkom od samo 10% informacija. Ovo predstavlja redukciju od 67% uz minimalan gubitak, što direktno sugerise da postoji ogromna redundancija u trenutnim NBA Draft Combine merenjima.

4.2 Interpretacija Glavnih Komponenti

PCA takođe omogućava interpretaciju ovih komponenti analizom težinskih koeficijenata (eng. loadings) - doprinos originalnih karakteristika svakoj komponenti pa možemo razumeti šta svaka komponenta zapravo predstavlja.



Slika 3: Doprinosi originalnih karakteristika prvim dvema glavnim komponentama

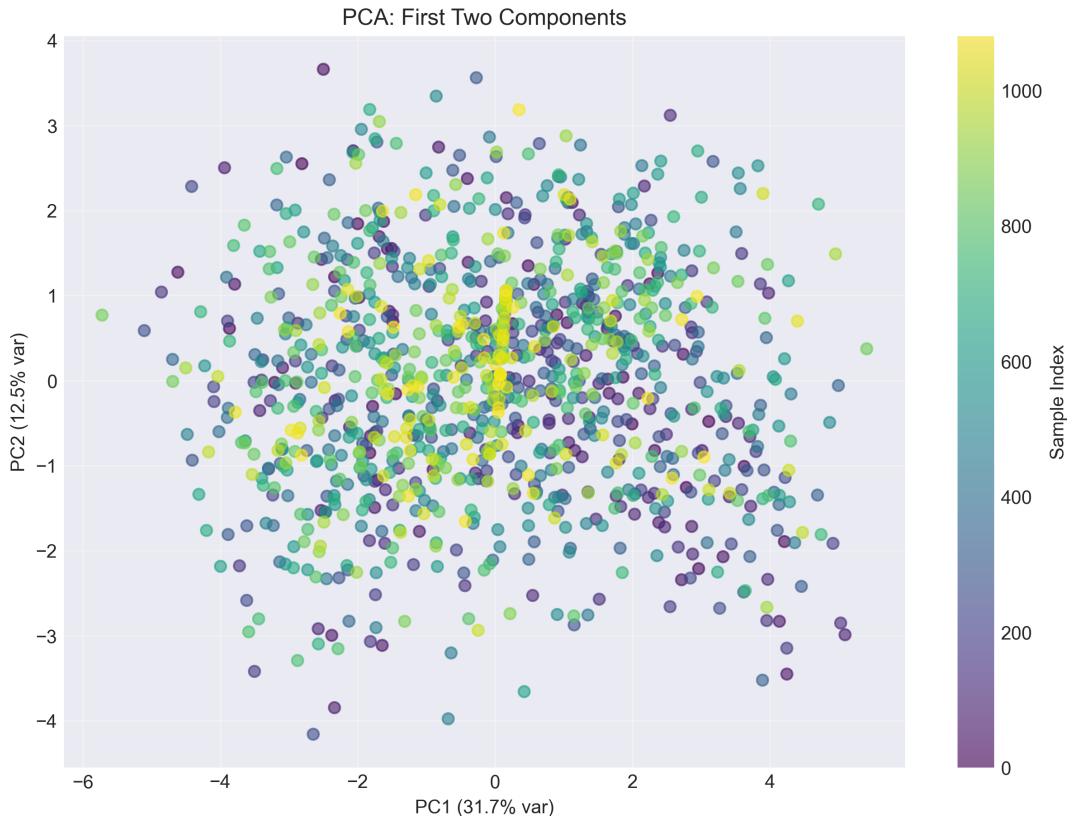
Na slici 3 možemo videti da je prva komponenta, PC1, gotovo u potpunosti definisana merama fizičke veličine gde najveće pozitivne doprinose daju doseg u stoju, visina sa i bez patika, raspon ruku i težina. Na suprotnom kraju, negativne doprinose imaju testovi agilnosti i vremena u trčanju na kratke staze. Ovo nam govori da PC1 zapravo meri spektar „veliki i spori” nasuprot „mali i brzi”.

Druga komponenta, PC2, priča drugaćiju priču. Ovde dominiraju merenja eksplozivnosti i snage kao što su vertikalni skok (iz mesta i maksimalni), potisak sa klupe, i širina šake. Ova komponenta meri atletizam nezavisno od veličine. Dva igrača mogu biti iste visine (neutralni na PC1), ali jedan može biti mnogo eksplozivniji (pozitivan na PC2).

Ove dve dimenzije zajedno objašnjavaju preko 44% ukupne varijanse i odgovaraju intuitivnim pitanjima koja se postavljaju o košarkašima: „Koliko si visoki?” (PC1) i „Koliko si atletičan?” (PC2).

4.3 Vizualizacija u Redukovanom Prostoru

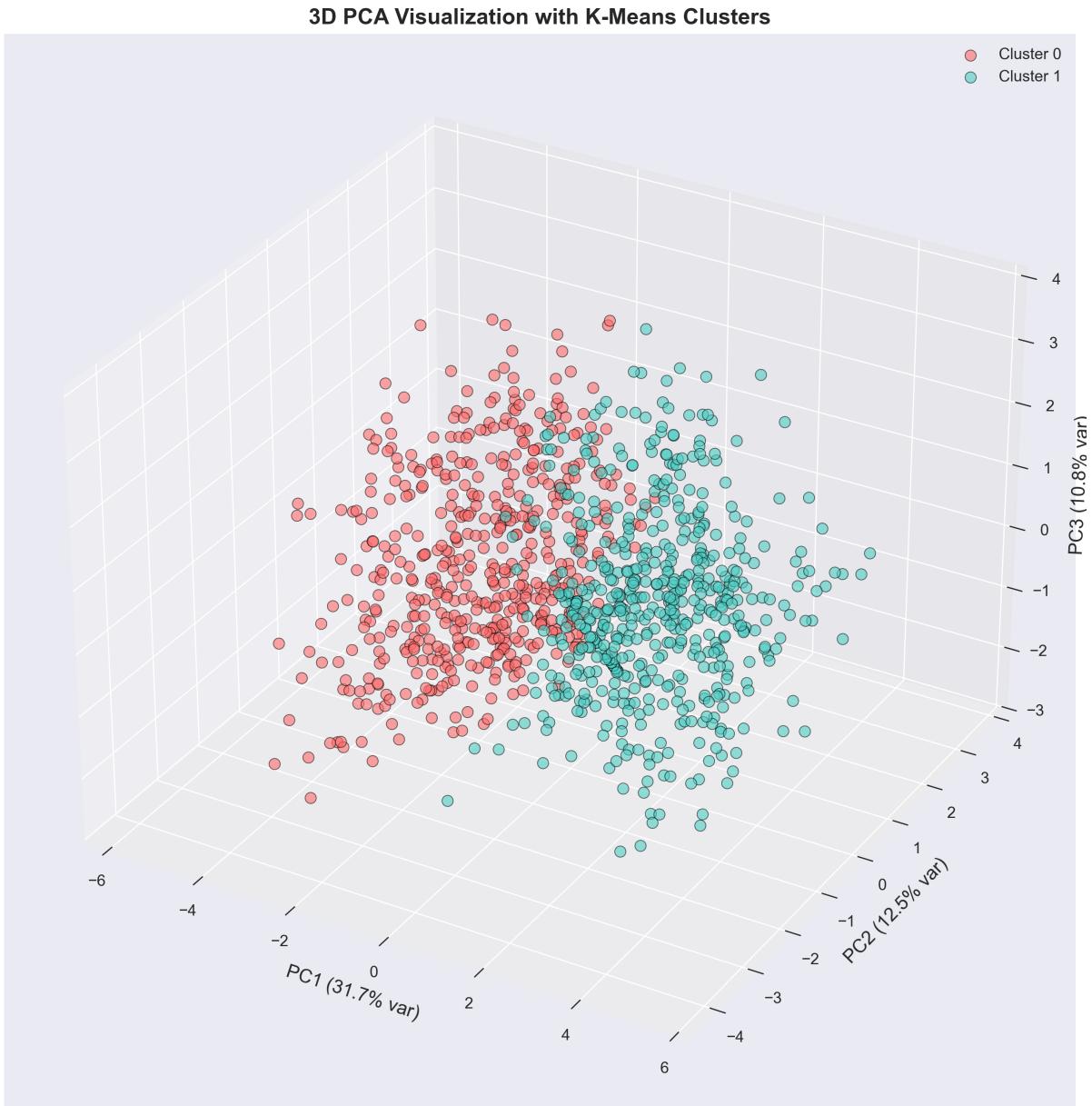
Sa redukcijom na 2 dimenzije, konačno možemo videti celokupan skup podataka i početi da vidimo obrasce koji bi inače bili sakriveni u višedimenzionalnom prostoru.



Slika 4: 2D PCA projekcija sa prve dve glavne komponente

Slika 4 prikazuje svih 1.081 igrača kao tačke u prostoru PC1 i PC2. Nešto što odmah upada u oči je da nema oštrih granica ili jasno razdvojenih klastera već vidimo kontinuiran oblak tačaka.

Ovaj kontinuitet je važan nalaz jer sugerije da podela pozicija u košarci nije binarni kategorija već fluidni spektar. Nema jasne linije gde prestaje bek a počinje krilo već postoje granični slučajevi, hibridi, igrači koji mogu igrati više pozicija. Ovo odražava modernu košarku gde su tradicionalne pozicije sve manje rigidne.



Slika 5: 3D PCA vizualizacija sa K-Means klasterima ($k=2$)

Dodavanjem treće dimenzije (PC3, koja doprinosi dodatnih 10.78% varijanze), dobijamo potpuniju sliku. Slika 5 prikazuje podatke u 3D prostoru gde su igrači obojeni prema klasterima koje će K-Means algoritam kasnije identifikovati. (Pogledati Sekciju 5.2) Već ovde možemo videti da se formiraju dve glavne grupe, jedna crvena (veći igrači) i jedna plava (perimetralni igrači), ali sa značajnim preklapanjem u srednjoj zoni. Ovo preklapanje su upravo oni igrači koji mogu funkcionisati u više uloga, poznatiji i kao igrači na više pozicija⁷.

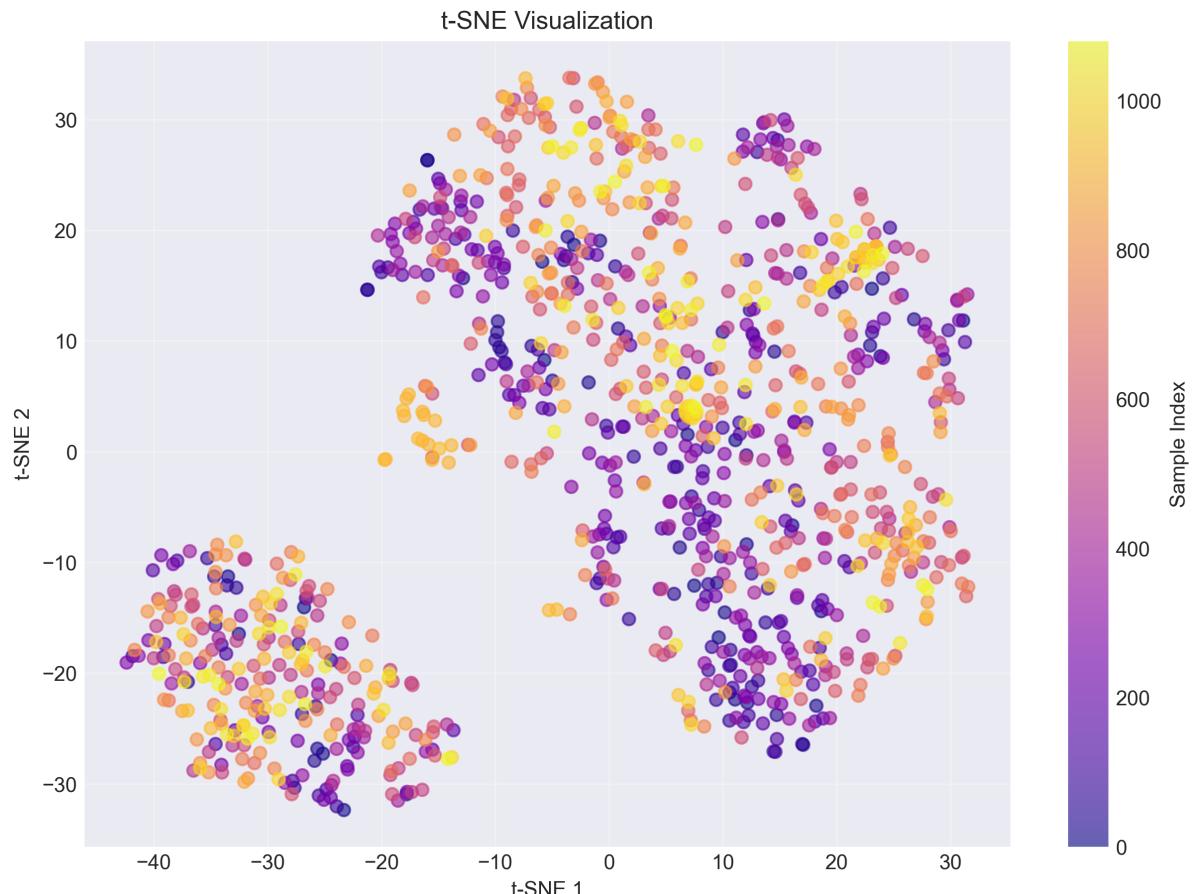
⁷Swingman - igrač koji može igrati na dve ili više pozicija, najčešće na pozicijama beka šutera i niskog krila

4.4 t-SNE: Alternativna Perspektiva

PCA je moćna tehnika, ali ona je linearna transformacija. Šta ako odnosi između igrača nisu linearni? Šta ako postoje kompleksnije strukture koje PCA ne može uhvatiti zbog svoje prirode?

Za istraživanje ove mogućnosti, primenjena je t-SNE [3]⁸ tehnika. Ona se fokusira na očuvanje lokalnih struktura gde pokušava da igrače koji su slični u višedimenzionalnom prostoru drži blizu i u 2D projekciji, čak i ako ta bliskost nije linearna.

t-SNE algoritam radi kroz iterativni proces koji započinje sa slučajnom projekcijom tačaka u 2D prostor, a zatim ih postepeno pomera tako da lokalne strukture budu očuvane. Ovo znači da dva igrača sa sličnim fizičkim profilima u originalnom prostoru treba da budu blizu i na t-SNE vizualizaciji, nezavisno od toga da li je ta sličnost vezana za linearu kombinaciju karakteristika.



Slika 6: t-SNE vizualizacija NBA igrača

Slika 6 prikazuje rezultat t-SNE projekcije primenjen na naših 1.081 igrača. Za razliku od PCA vizualizacije gde vidimo kontinualan oblak tačaka sa postepenim prelazom, t-SNE otkriva „ostrva“ igrača sa sličnim karakteristikama i jasnije definisane granice između različitih grupa. Algoritam je uspeo da otkrije lokalne gustine koje odgovaraju različitim podtipovima igrača unutar širih kategorija.

Ovo ne znači da je t-SNE bolja od PCA već jednostavno prikazuje druge aspekte strukture podataka.

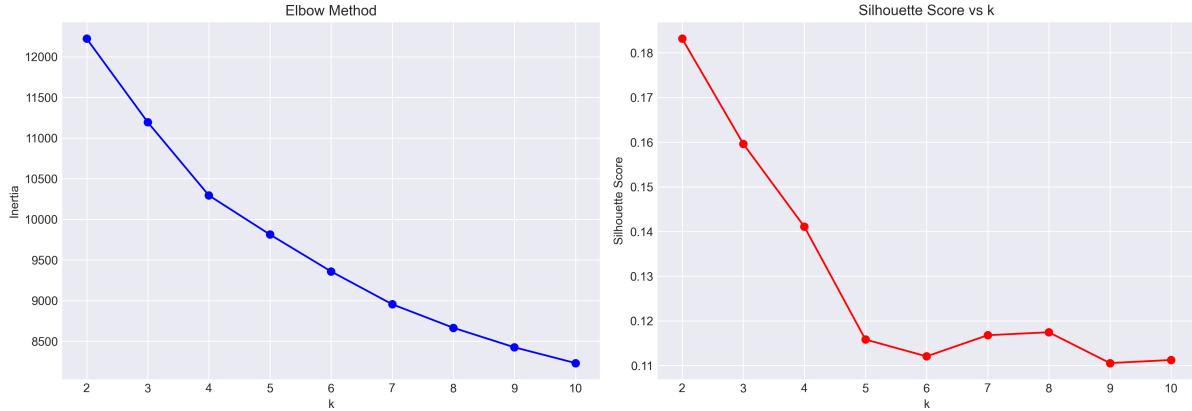
⁸t-distributed Stochastic Neighbor Embedding

5 Klasterovanje

5.1 Određivanje Optimalnog Broja Klastera

Sa jasnom slikom podataka u redukovanim prostoru, sledeći korak je bio formalizovati grupisanje igrača u prirodne arhetipove. Ali prvo pitanje je bilo koliko ima tih arhetipova? Da li košarkaši prirodno padaju u dve kategorije (veliki vs. mali), tri (odgovarajući pozicijama bek-kriko-centar), ili možda pet (po tradicionalnim pozicijama: plejmejker, bek šuter, nisko krilo, kriknji centar, centar⁹)?

Za odgovor na ovo pitanje korišćene su dve komplementarne tehnike: metoda lakta (eng. Elbow method) i analiza siluete [4] (eng. Silhouette analysis). Algoritam K-Means je pokretan za vrednosti k od 2 do 10, i za svaku vrednost su izračunate metrike kvaliteta klasterovanja.



Slika 7: Metoda lakta i skor siluete za različite vrednosti k

Slika 7 prikazuje rezultate. Na levom grafiku, **metoda lakta** pokazuje kako inercija (suma kvadratnih odstojanja tačaka od njihovih najbližih centroida) opada sa povećanjem broja klastera. Tačka pregiba gde dalje povećanje broja klastera ne donosi proporcionalno poboljšanje se javlja već na k=2, možda k=3. Ovo sugerira da dodatni klasteri nakon ovoga ne donose mnogo novih informacija.

Desni grafik, analiza siluete, potvrđuje ovaj nalaz još jasnije. **Skor siluete** meri koliko je svaki igrač sličniji igračima u svom klasteru nego igračima u drugim klasterima. Vrednosti blizu 1 znače odlično klasterovanje, blizu 0 znače preklapanje, a negativne vrednosti znače da je igrač verovatno pogrešno dodeljen. Najviši skor (0.184) je postignut tačno na k=2. Sa povećanjem k, skor opada, što znači da forsiranje više klastera zapravo pogoršava kvalitet grupisanja.

Vrednost od 0.1832 je zapravo realistična vrednost jer fizičke karakteristike variraju kontinualno, ne u diskretnim kategorijama. Nije kao da postoje dva potpuno odvojena tipa ljudi već umesto toga postoje spektar, ali sa dva jasna „vrha“ oko kojih se igrači grupišu.

5.2 K-Means Klasterovanje sa k=2

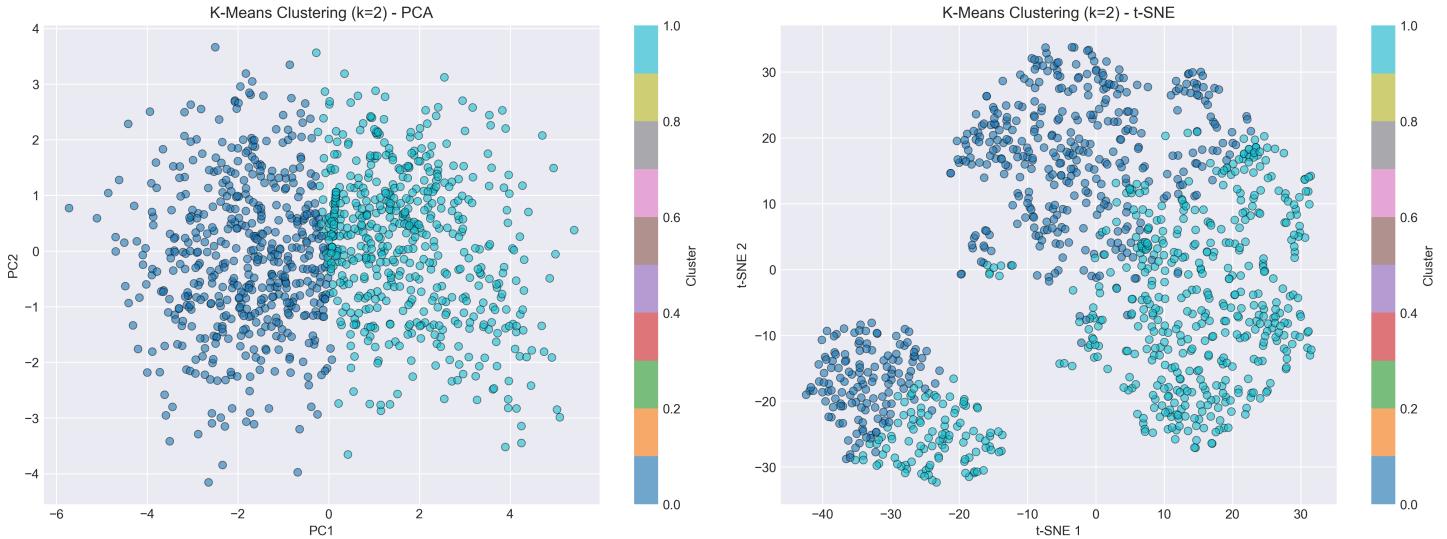
Na osnovu gornje analize, K-Means je pokrenut sa k=2, što je dalo jasnu podelu populacije:

Klaster	Broj igrača	Procenat
Klaster 0	512	47.4%
Klaster 1	569	52.6%

Tabela 4: Distribucija igrača po klasterima

Podela je gotovo ravnomerna gde 47.4% populacije pada u jednu grupu, a 52.6% u drugu.

⁹Tradicionalne košarkaške pozicije: plejmejker (eng. point guard - PG) - organizator igre; bek šuter (eng. shooting guard - SG) - spoljni šuter; nisko krilo (eng. small forward - SF) - svestrani igrač; kriknji centar (eng. power forward - PF) - snažniji kriknji igrač; centar (eng. center - C) - igrač pod košem



Slika 8: K-Means klasteri vizualizovani u PCA i t-SNE prostoru

Slika 8 pokazuje ove klastere iz dve perspektive. Na levoj strani, PCA projekcija jasno pokazuje razdvajanje klastera duž PC1 ose, što je očekivano jer smo već videli da PC1 meri „veličinu”. Klasteri se manje razlikuju duž PC2 ose (atletizam), što sugerije da oba arhetipa imaju podjednako velike raspone u atletskim sposobnostima.

Desna strana koristi t-SNE, drugu tehniku redukcije dimenzionalnosti koja je fokusirana na očuvanje lokalne strukture. Za razliku od PCA prikaza, t-SNE jasnije razdvaja klastere u kompaktne grupe, što potvrđuje da algoritam zaista prepoznaće dve distinktne populacije igrača.

Matematika je identifikovala dva klastera, ali da bismo razumeli šta oni predstavljaju potrebno je pogledati prosečne vrednosti fizičkih merenja za svaki klasster:

Karakteristika	Klaster 1	Klaster 0	Razlika
Visina (sa patikama, cm)	204.97	195.15	+9.82
Težina (kg)	103.95	89.67	+14.28
Raspon ruku (cm)	215.50	202.94	+12.56
Vert. skok (cm)	84.72	91.69	-6.97
Agilnost (sek)	11.56	11.14	+0.42

Tabela 5: Prosečne vrednosti fizičkih karakteristika po klasteringima

Vidimo da Klaster 0 obuhvata prosečno oko 10 cm više igrače, koji su takođe teži za oko 14 kg i imaju raspon ruku duži za 12.56 cm prosečno. Ovo su klasični „veliki igrači” koji dominiraju prostorom pod košem, pretežno u ulozi centara i krilnih centara¹⁰. Oni su sporiji u testovima agilnosti što je očekivani kompromis sa veličinom, ali ta razlika je minimalna (svega pola sekunde prosečno).

Međutim, interesantno je da su ovi visoki igrači u proseku slabiji u vertikalnom skoku. Ovaj nalaz je u skladu sa posmatranjima iz igre, budući da je sposobnost visokog skoka značajnija za igrače koji šutiraju sa veće distance, dok je njen značaj manji za igrače koji većinu vremena provode pod košem.

Klaster 0, sa prosečnom visinom od 195.15 cm i težinom od 89 kg, su spoljašnji igrači¹¹ u ulozi bekova i krila¹². Njihova prednost nije u veličini već u brzini, agilnosti i često u eksplozivnosti. Ovo su igrači koji diktiraju tempo igre, organizuju napad, i čuvaju spoljašnji deo terena.

¹⁰Krilni centar (eng. power forward) - pozicija između krila i centra, kombinuje snagu i pokretljivost

¹¹Perimetralni/spoljašnji igrači - igrači koji većinu vremena provode na spoljašnjem delu terena, dalje od koša

¹²Krilo (eng. forward) - pozicija između beka i centra, svestrani igrač koji igra i na perimetru i pod košem

5.3 Alternativni pristupi klasterovanju

Kako bi se proverilo da li su dobijeni rezultati specifični za K-Means algoritam, primenjeni su i alternativni pristupi klasterovanju.

Hijerarhijsko klasterovanje, koje konstruiše dendrogram postupnim spajanjem najbližih klastera, rezultovalo je identičnom podelom na dve dominantne grupe. Ovakav ishod ukazuje da dobijena podela nije artefakt izabranog algoritma, već predstavlja robusnu strukturu prisutnu u podacima. Dobijeni skor siluete iznosio je 0.1368.

DBSCAN algoritam [5], zasnovan na identifikaciji regionalne visoke gustine u prostoru podataka, dao je drugačiji rezultat, pri čemu je većina igrača svrstana u jedan veliki gasti klaster, uz manji broj izdvojenih netipičnih vrednosti. Ovakav nalaz sugerira da podaci nisu jasno razdvojeni u gustinskom smislu, odnosno da ne postoje izražene „praznine“ između grupa, već kontinuirani prelaz između fizičkih profila igrača, što je u skladu sa vizualizacijama prikazanim na slici 4. DBSCAN na taj način dodatno potvrđuje prethodne nalaze da se košarkaške pozicije mogu posmatrati kao kontinuum, a ne kao stroga diskretna kategorije. Ukupan broj tačaka označenih kao šum iznosio je 54.

Na osnovu ovih poređenja, K-Means algoritam, koji je ostvario najviši skor siluete i obezbedio najinterpretabilniju podelu, ostaje najpogodniji pristup za ovaj skup podataka i stoga je korišćen u daljim analizama.

6 Klasifikacija

K-Means algoritam je identifikovao dva fizička arhetipa među NBA igračima. Sledеće pitanje je koliko je taj obrazac jasno definisan? Ako treniramo na modele mašinskog učenja da prepozna razliku između ta dva klastera, koliko uspešno to mogu uraditi?

Ako klasifikatori mogu sa velikom tačnošću razlikovati klasterne, to potvrđuje da su klasteri dobro definisani i nisu proizvod slučajnosti. Ako tačnost bude niska, to bi sugerisalo da su klasteri veštački i da možda ne treba da im verujemo.

Za ovaj eksperiment, podaci su podeljeni na trening (70%) i test skupove (30%), i testirana su tri različita tipa algoritama: slučajna šuma [6]¹³ (eng. Random forest), mašina potpornih vektora (eng. Support Vector Machine (SVM))¹⁴, i neuronska mreža¹⁵ (eng. Neural network). Svi modeli su implementirani korišćenjem biblioteke scikit-learn [7].

Model	Skup karakteristika	Tačnost
Slučajna šuma	Pun skup (30 PC)	98.15%
Slučajna šuma	Redukovani (10 PC)	98.77%
SVM (RBF jezgro)	Pun skup (30 PC)	96.62%
SVM (RBF jezgro)	Redukovani (10 PC)	96.62%
Neuronska mreža	Pun skup (30 PC)	97.85%
Neuronska mreža	Redukovani (10 PC)	99.08%

Tabela 6: Performanse modela za predviđanje pripadnosti klasteru

Rezultati su impresivni jer svi modeli postižu tačnost preko 96%. Slučajna šuma greši u samo 1.85% slučajeva kada koristi svih 30 komponenti, dok Neuronska mreža na redukciji od 10 komponenti dostiže gotovo savršenih 99.08%. Ovo jasno potvrđuje da su klasteri koje je K-Means identifikovao ekstremno dobro definisani i jasno da modeli gotovo uvek mogu ispravno klasifikovati novog igrača.

Ali možda još interesantniji nalaz je to što se dešava kada redukujemo broj karakteristika sa 30 na 10 koristeći PCA. Umesto očekivanog pada u performansama, vidimo poboljšanje. Slučajna šuma se poboljšava za 0.62%, a Neuronska mreža čak za 1.23%.

Objašnjenje leži u konceptu „šuma u podacima“. Od naših 30 originalnih karakteristika, mnoge mere iste stvari ili su irrelevantne za razlikovanje između ova dva arhetipa. PCA, fokusirajući se na 10 komponenti koje objašnjavaju najveći deo varijanse, efektivno filtrira ovaj šum. Modeli dobijaju čistiji signal i mogu donositi bolje odluke.

6.1 Predviđanje Stvarnih Košarkaških Pozicija

Šta se dešava kada pokušamo da predvidimo stvarne košarkaške pozicije - bek, krilo i centar - samo na osnovu fizičkih merenja?

Ovaj zadatak je složeniji jer pozicije nisu striktno definisane fizičkim karakteristikama, dva igrača iste visine mogu igrati različite pozicije zavisno od veština koje poseduju. Ipak, očekivali bismo određenu prediktivnost jer visina i atletizam svakako utiču na to na kojoj poziciji će igrač biti najkorisniji.

Model	Tačnost
Slučajna šuma	85.56%
SVM (RBF)	80.28%
Neuronska mreža	67.22%

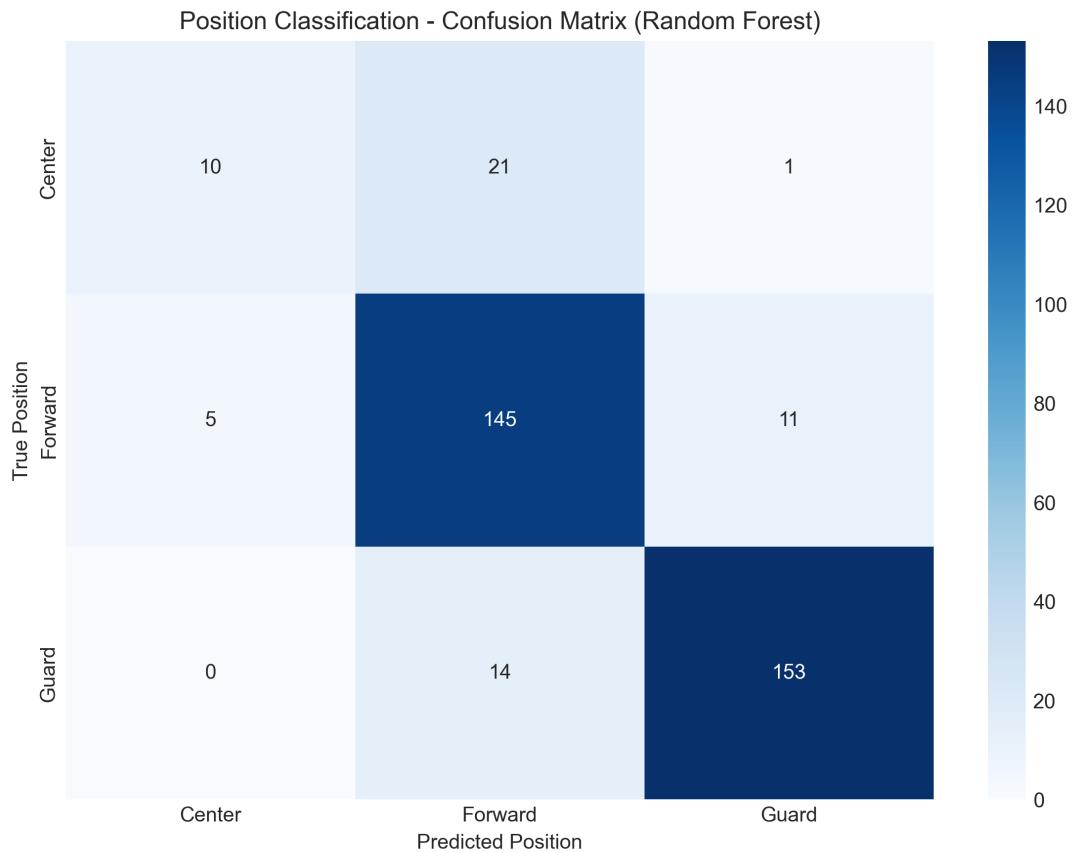
Tabela 7: Performanse modela za klasifikaciju pozicije

¹³Slučajna šuma - ansambl metoda zasnovana na kombinovanju više stabala odlučivanja

¹⁴SVM - algoritam koji pronađe optimalnu granicu razdvajanja između klasa

¹⁵Neuronska mreža - model inspirisan biološkim neuronima, višeslojna perceptronska mreža

Slučajna šuma postiže tačnost od 85.56%, što je impresivno ali veoma niže od 98% koje smo videli za klastere. Ova razlika nam govori da su fizičke karakteristike dobar ali ne i savršen prediktor pozicije.



Slika 9: Matrika konfuzije za klasifikaciju pozicije (Slučajna šuma)

Slika 9 otkriva gde model pravi greške. Pozicija beka se klasificuje sa najvišom preciznošću sa 153 od 167 bekova koji su ispravno klasifikovani, što daje 91.6% tačnost. Bekovi tipično imaju najjasnije fizičke karakteristike jer su manji, lakši, brži, pa rezultat ima smisla.

Pozicija centra takođe pokazuje dobru preciznost sa 10 od 32 ispravno klasifikovanih u ovoj relativno maloj klasi, mada se često brka sa krilo pozicijom. Pozicija krila ima najviše grešaka gde je 145 od 161 je ispravno klasifikovano (90.1%), ali 21 krilni igrač je pogrešno klasifikovan kao centar, a 11 kao bek.

Ovo odražava realnost moderne košarke gde je krilo najfluidnija pozicija. Igrači na poziciji niskog krila mogu biti dovoljno brzi da igraju kao bek u nekim situacijama, dok igrači na poziciji krilnog centra mogu igrati kao centar. Ova fluidnost je jedna od karakteristika evolucije NBA igre u poslednjih 20 godina gde pozicije postaju sve manje rigidne, a igrači sve više igraju bez pozicije¹⁶.

¹⁶Positionless basketball - moderni stil igre gde igrači nisu striktno vezani za jednu poziciju već mogu igrati više uloga

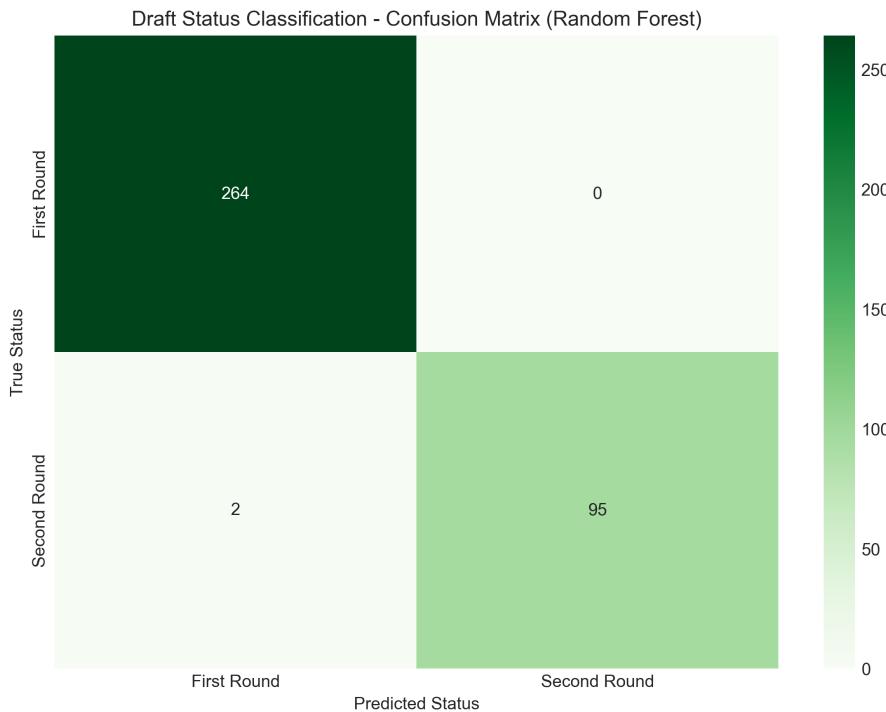
6.2 Predikcije Draft Statusa

Poslednji eksperiment testira možda najzanimljivije pitanje za NBA organizacije. Da li fizičke karakteristike mogu predvideti kako će igrač biti rangiran na draftu? Konkretno, može li model razlikovati igrače koji će biti izabrani u prvoj rundi (pozicije 1-30) od onih u drugoj rundi (pozicije 31-60)?

Model	Tačnost
Slučajna šuma	99.45%
SVM (RBF)	94.18%
Neuronska mreža	100.00%

Tabela 8: Performanse modela za klasifikaciju draft statusa

Neuronska mreža postiže savršenih 100%, a Slučajna šuma greši u samo 2 od 363 slučaja. Postignuta tačnost je izuzetno visoka i stoga zahteva oprez pri interpretaciji.



Slika 10: Matrica konfuzije za klasifikaciju draft statusa (Slučajna šuma)

Slika 10 pokazuje matricu konfuzije koja je gotovo dijagonalna sa samo 2 greške od 361 predviđanja. Od 266 igrača iz prve runde, 264 je ispravno klasifikovano (99.2%), dok je od 97 igrača iz druge runde, 95 ispravno klasifikovano (97.9%).

Postoji nekoliko mogućih interpretacija. Prva je da su fizičke karakteristike toliko važne da gotovo potpuno određuju kako će skauti rangirati igrače. Igrači draftovani u prvoj rundi su statistički superiorni po svim fizičkim metrikama, što sugerira da NBA organizacije rade dobar posao evaluacije fizičkih talenta.

Međutim, postoje i ograničenja ove interpretacije koja ne treba zanemariti. Skup podataka sadrži samo igrače koji su pozvani na probni kamp dok mnogi igrači iz druge runde ili nedraftovani nikada nisu ni testirani, što kreira veštačku separaciju. Takođe, može postojati neravnoteža u broju igrača između rundi (880 iz prve runde vs 322 iz druge u celom skupu podataka), što olakšava klasifikaciju.

Uprkos ovim ograničenjima, rezultat jasno potvrđuje hipotezu da su fizičke karakteristike ekstremno prediktivne za rani uspeh u draft procesu. Ovo ne znači da veštine, košarkaški IQ i mentalni aspekti nisu važni, ali sugerira da su fizički atributi primarni kriterijum selekcije koji NBA timovi koriste u ranoj fazi evaluacije.

7 Pravila Pridruživanja

Do sada smo koristili tehnike koje grupišu slične igrače ili predviđaju ishode. Ali postoji još jedan ugao analize i to je otkrivanje „pravila” koja povezuju različite fizičke karakteristike. Na primer, ako znamo da igrač ima visok vertikalni skok i dobru agilnost, šta još verovatno možemo očekivati od njega?

Za ovo koristimo Apriori algoritam [8], tehniku iz oblasti otkrivanja asocijativnih pravila¹⁷. Primjenjena na naše podatke, može otkriti koji fizički atributi se konzistentno pojavljuju zajedno.

Pre primene algoritma, kontinuirane numeričke karakteristike (poput visine ili telesne mase) diskretnizovane su u kategorijalne varijable. Svaka karakteristika podeljena je na tri nivoa i to „Low”, „Medium” i „High” na osnovu tercila raspodele, pri čemu donjih 33% vrednosti pripada kategoriji „Low”, srednjih 33% kategoriji „Medium”, a gornjih 33% kategoriji „High”. Ovakva transformacija omogućava algoritmu da identificuje asocijativne obrasce, kao što su veze između niske telesne mase, visokog vertikalnog skoka i povećane agilnosti.

7.1 Otkrića

Apriori¹⁸ je pokrenut sa pragovima od 10% za podršku (eng. support) (pravilo mora da se pojavi u bar 10% slučajeva da bi bilo razmatrano) i 70% za pouzdanost (eng. confidence) (ako je leva strana pravila istinita, desna strana mora biti istinita u bar 70% slučajeva).

Metrika	Vrednost
Ukupan broj transakcija	1.202
Ukupan broj diskretnih stavki	21
Identifikovanih čestih skupova predmeta	203
Generisanih pravila	129
Minimalna podrška	0.10
Minimalna pouzdanost	0.70
Prosečna pouzdanost	0.786
Prosečni podizaj	2.093
Maksimalni podizaj	3.273

Tabela 9: Sumarni rezultati Apriori analize

Algoritam je identifikovao ukupno 129 pravila koja zadovoljavaju oba kriterijuma. Prosečna vrednost pouzdanosti od 78.6% ukazuje na to da su dobijena pravila robusna, jer ne predstavljaju slučajne korelacije, već stabilne i konzistentne obrasce u podacima. Mera **podizaj** (eng. lift) kvantifikuje u kojoj meri je pojava desne strane pravila verovatnija kada je leva strana ispunjena, u poređenju sa njenom ukupnom učestalošću. Najveća zabeležena vrednost podizaja od 3.273 označava asocijaciju koja je više od tri puta jača od one koja bi se očekivala slučajnim putem, što ukazuje na izuzetno snažno pravilo.

7.2 Najinteresantnija Pravila

Uslov (Antecedent)	Posledica (Consequent)	Pouzd.	Podizaj
Med Standing Vert + Med Agility	Med Max Vert	0.931	3.273
Med Max Vert + Med Agility	Med Standing Vert	0.856	3.225
Med Max Vert + Low Bench Press	Med Standing Vert	0.739	2.784
Med Standing Vert + Med Max Vert	Med Agility	0.852	2.639
High Weight + Low Standing Vert	Low Max Vert	0.887	2.544
High Bench Press + High Max Vert	High Standing Vert	0.892	2.466

Tabela 10: Prvih 6 pravila po vrednosti podizaja

¹⁷Association rule mining - tehniku otkrivanja znanja koja pronalazi veze između stavki u velikim skupovima podataka

¹⁸Apriori - algoritam za pronalaženje čestih skupova stavki i generisanje asocijativnih pravila

Najjače identifikovano pravilo (podizaj = 3.273) pokazuje da igrači sa prosečnim vertikalnim skokom iz mesta i prosečnim vremenom u testu agilnosti sa velikom verovatnoćom (93.1% pouzdanost) imaju i prosečan maksimalni vertikalni skok. Vrednost lifta od 3.273 ukazuje da se ova kombinacija pojavljuje više od tri puta češće nego što bi se očekivalo pod pretpostavkom nezavisnosti karakteristika. Ovakav nalaz ukazuje na visok stepen konzistentnosti u atletskim performansama, budući da igrači koji ostvaruju prosečne rezultate u jednoj dimenziji eksplozivnosti teže da budu prosečni i u ostalim srodnim testovima.

Drugo pravilo predstavlja recipročan odnos u odnosu na prvo i dodatno potvrđuje uočenu konzistentnost atletskih sposobnosti. Treće pravilo otkriva interesantnu povezanost između eksplozivnosti donjih ekstremiteta i snage gornjeg dela tela: igrači sa prosečnim maksimalnim skokom i niskim rezultatima u testu potiska sa klupe i dalje ostvaruju prosečne vrednosti u vertikalnom skoku iz mesta. Ovakav obrazac sugerira da eksplozivnost skoka ne zavisi direktno od snage gornjeg dela tela.

Peto pravilo potvrđuje očekivani kompromis između telesne mase i eksplozivnosti. Igrači sa visokom telesnom masom i niskim vertikalnim skokom iz mesta sa velikom verovatnoćom (88.7%) ostvaruju i nizak maksimalni vertikalni skok. Ovaj nalaz je u skladu sa osnovnim biomehaničkim principima, prema kojima veća masa zahteva veću silu za postizanje iste visine skoka.

Šesto pravilo ukazuje da igrači koji kombinuju visoku eksplozivnost (maksimalni vertikalni skok) i visoku snagu (potisak sa klupe) najčešće ostvaruju dobre rezultate i u testu vertikalnog skoka iz mesta.

Ukupno posmatrano, dobijena asocijativna pravila otkrivaju sistemske obrasce u fizičkim karakteristikama NBA igrača. Uočavaju se prirodni „paketi“ osobina koje se pojavljuju zajedno, oblikovani kako fiziologijom ljudskog tela, tako i specifičnim zahtevima savremene košarkaške igre.

7.3 Praktična Primena Pravila

Otkrivena pravila imaju konkretnе praktične primene. Prvo, mogu služiti za validaciju merenja. Na primer, ako igrač ima visoku težinu ali pokazuje visok vertikalni skok, to je neuobičajena kombinacija jer pravilo sa 88.7% pouzdanošću predviđa suprotno. Ovakav slučaj može ukazivati na grešku u merenju ili na zaista izuzetnog kandidata koji zaslužuje dodatnu pažnju.

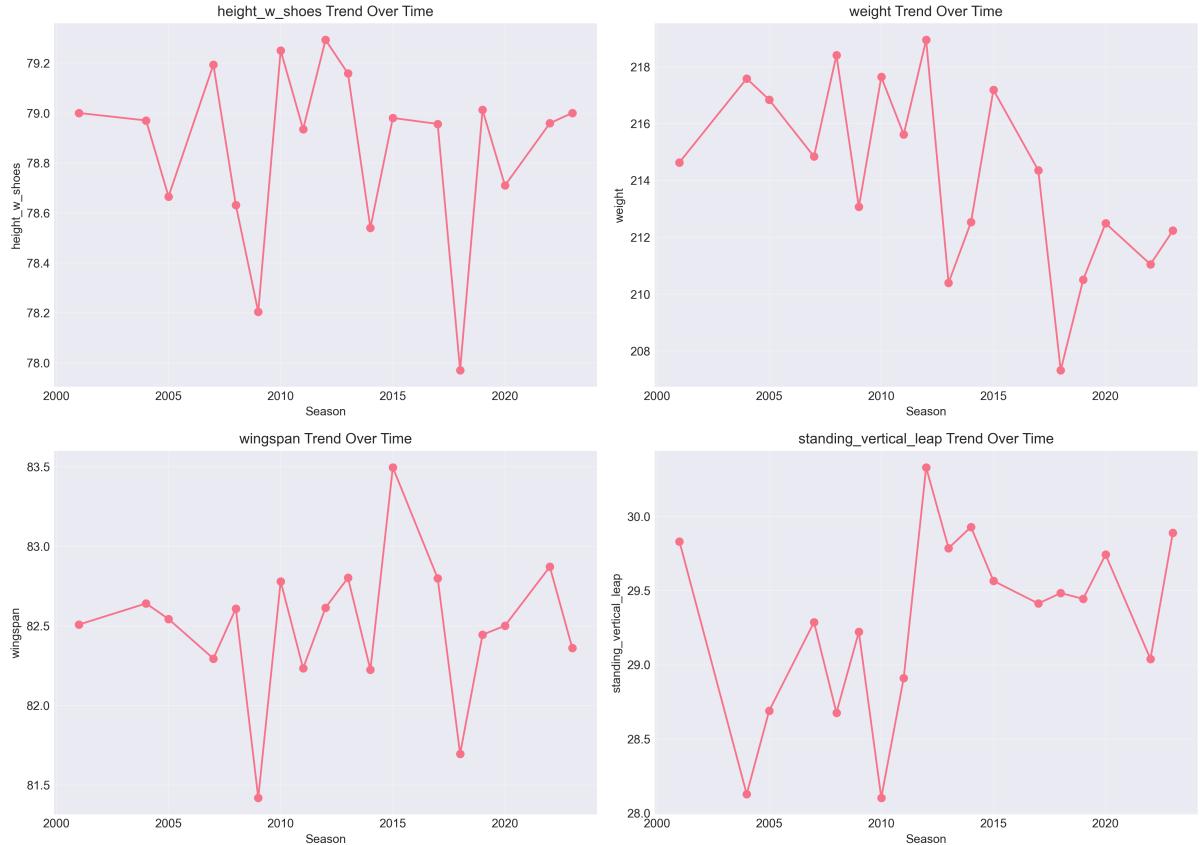
Drugo, pravila mogu pomoći u predviđanju nedostajućih vrednosti. Ako igrač nije učestvovao u testu maksimalnog vertikalnog skoka, ali ima prosečan vertikalni skok iz mesta i prosečnu agilnost, sa 93.1% pouzdanošću možemo pretpostaviti da bi i maksimalni vertikalni skok bio u prosečnom opsegu.

Treće, pravila formalizuju kompromise između fizičkih karakteristika. Pravilo o visokoj težini i niskom skoku (podizaj = 2.544) kvantificuje kompromis koji igrači na poziciji centra prave, žrtvujući eksplozivnost za fizičku dominaciju. Sa druge strane, pravilo o niskom potisku sa klupe i prosečnom skoku pokazuje da eksplozivnost nogu ne zahteva snagu gornjeg dela tela, što je relevantno za bekove koji mogu zanemariti taj aspekt treninga bez uticaja na njihovu sposobnost skakanja.

Konačno, pravila o konzistentnosti prosečnih vrednosti (prva četiri pravila sa vrednost podizajima 2.6-3.3) sugeriraju da većina igrača ima ujednačen atletski profil. Igrači retko imaju ekstremne vrednosti u jednoj dimenziji a prosečne u drugoj, što olakšava kategorizaciju i poređenje kandidata.

8 Vremenska Dimenzija

Jedan od jedinstvenih aspekata ovog skupa podataka jeste činjenica da obuhvata period duži od dve decenije (2000–2024). To omogućava ne samo analizu trenutnih fizičkih profila NBA draft kandidata, već i uvid u njihove promene kroz vreme. Postavlja se pitanje da li savremeni igrači postaju viši, brži i eksplozivniji, ili se osnovni fizički obrasci u velikoj meri zadržavaju nepromjenjenim.



Slika 11: Trendovi fizičkih karakteristika kroz sezone

Slika 11 prikazuje prosečne vrednosti četiri ključne metrike grupisane po sezonomama. Ono što uočavamo je relativna stabilnost većine merenja jer nema dramatičnih skokova ili padova. Ovo nam govori da osnovni fizički zahtevi NBA košarke ostaju konzistentni, uprkos promenama u stilu igre.

Visina sa patikama osciluje oko 201 cm kroz ceo period bez jasnog trenda. Visina je biološki ograničena karakteristika i nije nešto što se može značajno promeniti treningom ili evolucijom sporta u kratkom periodu od 25 godina. NBA nastavlja da regrutuje igrače iz istog spektra visina jer su ti zahtevi najbitniji za sport.

Telesna masa pokazuje dinamičniji obrazac, sa oscilacijama između približno 94 i 99 kg tokom posmatranog perioda, uz blagi pad u poslednjim sezonomama. Ovaj trend ka nižoj prosečnoj masi u skladu je sa evolucijom NBA igre ka bržem tempu, većem prostornom razmaku i snažnijem naglasku na perimetralnoj igri i šutiranju za tri poena. Savremeni igrači su u proseku pokretljiviji i vitkiji u poređenju sa igračima iz ranih 2000-ih, kada je igra bila sporija i više orientisana ka igri leđima prema košu.

Raspon ruku ostaje relativno stabilan, u intervalu od približno 207 do 212 cm, bez jasno izraženog dugoročnog trenda. Ovakav nalaz je očekivan, budući da je raspon ruku, slično kao i visina, u velikoj meri genetski determinisana karakteristika i stoga ne pokazuje sistematske promene tokom kraćih vremenskih intervala.

Najizraženiju varijabilnost pokazuje vertikalni skok iz mesta. Njegov maksimum zabeležen je oko 2012. godine, sa prosečnim vrednostima većim od 76 cm, nakon čega sledi postepeni pad koji traje do oko 2021.

godine, uz blag oporavak u kasnijim sezonomama. Ovakav obrazac može imati više objašnjenja. Jedna od mogućnosti jeste da savremeni trenažni pristupi stavljuju veći akcenat na funkcionalnu snagu, izdržljivost i višedimenzionalnu pokretljivost, a manji na čistu eksplozivnost u skoku. Alternativno, može biti prisutan i selekcioni efekat, budući da neki od najperspektivnijih igrača u poslednjim godinama preskaču NBA Draft Combine, opredeljujući se za individualne treninge i testiranja direktno sa timovima.

8.1 Šta Ova Stabilnost Znači?

Možda najznačajniji nalaz vremenske analize upravo je uočena stabilnost fizičkih profila. Uprkos značajnim promenama u stilu igre kao što su prelazak ka strategijama sa manjim igračima¹⁹, porast značaja šuta za tri poena i povećanje tempa igre, fundamentalne fizičke karakteristike NBA draft kandidata ostaju relativno konzistentne. Ovakav nalaz sugerije da, bez obzira na evoluciju taktika i strategija, osnovni fizički zahtevi košarke na najvišem nivou ostaju u velikoj meri nepromenjeni.

Blagi trend ka manjoj telesnoj masi u poslednjim godinama može se posmatrati kao odraz jednog aspekta savremene igre, u kojoj se veći prioritet daje pokretljivosti u odnosu na čistu snagu. Ipak, ova promena je postepena i umerena, a ne revolucionarna, što ukazuje da NBA nije prešla na potpuno drugačiji tip igrača, već na suptilnu adaptaciju unutar postojećeg spektra fizičkih profila.

Analiza takođe pokazuje da se godišnje draft klase prirodno razlikuju u pogledu fizičkih karakteristika i opštег kvaliteta. Pojedine klase sadrže veći udio visokih igrača, dok su druge izraženije po atletičnosti. Ove varijacije predstavljaju prirodnu posledicu stohastičke raspodele talenata i ne ukazuju na dugoročne strukturne trendove, već na godišnje oscilacije u populaciji draft kandidata.

¹⁹Small-ball - taktički pristup u kom tim koristi niže i brže igrače umesto tradicionalnih visokih centara

9 Zaključak

Fizički profili NBA draft kandidata mogu se svesti na dve fundamentalne dimenzije: veličinu i atletizam. Od preko 30 merenja, svega deset glavnih komponenti objašnjava oko 90% varijanse, dok je ostatak u velikoj meri redundantan. Uprkos popularnom narativu o „košarci bez pozicija”, klaster analiza otkriva dva jasna arhetipa koja i dalje dominiraju ligom.

Posebno značajan uvid ne leži u prosečnim vrednostima, već u odstupanjima. Autlajeri, igrači čije kombinacije atributa odstupaju od statističkih očekivanja su često upravo oni koji redefinišu igru i uvođe nove stilove igre.

9.1 Ograničenja

Ova analiza obuhvata isključivo fizičke karakteristike i ne uključuje tehničke veštine, košarkaški IQ niti mentalnu čvrstinu - faktore koji su podjednako presudni za uspeh u ligi. Skup podataka sadrži samo igrače pozvane na NBA Draft Combine, dok mnogi međunarodni igrači i utvrđeni talenti ovaj događaj preskaču. Nedostatak podataka o šutiranju (preko 70% nedostajućih vrednosti) ograničio je analizu jedne od ključnih veština moderne košarke. Konačno, korelacija između fizičkog profila i draft pozicije ne implicira kauzalnost; fizički atributi mogu biti posledica, a ne uzrok različitih razvojnih prilika.

9.2 Budući pravci

Prirodni nastavak ovog istraživanja bio bi povezivanje podataka sa probnog kampa sa karijernim statistikama, kako bi se procenilo u kojoj meri fizički atributi zaista predviđaju dugoročni uspeh u ligi. Analiza po pozicijama mogla bi otkriti specifične optimalne profile za svaku ulogu. Vremensko praćenje igrača omogućilo bi praćenje promena fizičkih karakteristika tokom karijere i identifikaciju tipova igrača koji duže ostaju zdravi i funkcionalno sposobni.

9.3 Završna Refleksija

Ovaj projekat je demonstrirao moć pristupa zasnovanog na podacima u domenu gde tradicionalno dominira subjektivna procena i iskustvo. Pokazali smo da sofisticirane tehnike mašinskog učenja mogu otkriti obrasci i pravila koja nisu očigledna golim okom, i da ti obrasci imaju stvarnu prediktivnu moć.

Međutim, cilj nije bio da se zameni ljudska ekspertiza statističkim modelima, već da se dopuni. Najbolji rezultati dolaze iz kombinacije gde skauti koji razumeju košarku i igrače na dubokom nivou, podržani objektivnim analizama koje pomažu da se izbegnu pristranosti i prepoznaju skriveni obrasci.

Ključna lekcija je da složenost nije uvek bolja od jednostavnosti. Pokazali smo da 10 pažljivo odabranih dimenzija može biti informativnije od 30 koreliranih merenja. U eri velikih podataka, ponekad je najvažnije znati šta zanemariti. „Manje je više” nije samo dizajnerski princip, već je to i statistička realnost kada je u pitanju kvalitet podataka i robusnost zaključaka.

Literatura

- [1] Wyatt Walsh. NBA Draft Combine Measurements Dataset. Kaggle, <https://www.kaggle.com/datasets/wyattwalsh/basketball>, 2024. Pristupljeno: Januar 2026.
- [2] Fei Tony Liu, Kai Ming Ting, and Zhi-Hua Zhou. Isolation forest. In *Proceedings of the 8th IEEE International Conference on Data Mining (ICDM)*, pages 413–422. IEEE, 2008. <https://ieeexplore.ieee.org/document/4781136>.
- [3] Laurens van der Maaten and Geoffrey Hinton. Visualizing data using t-SNE. *Journal of Machine Learning Research*, 9:2579–2605, 2008. <https://www.jmlr.org/papers/v9/vandermaaten08a.html>.
- [4] Peter J. Rousseeuw. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 20:53–65, 1987. <https://www.sciencedirect.com/science/article/pii/0377042787901257>.
- [5] Martin Ester, Hans-Peter Kriegel, Jörg Sander, and Xiaowei Xu. A density-based algorithm for discovering clusters in large spatial databases with noise. In *Proceedings of the 2nd International Conference on Knowledge Discovery and Data Mining (KDD)*, pages 226–231, 1996. <https://www.aaai.org/Papers/KDD/1996/KDD96-037.pdf>.
- [6] Leo Breiman. Random forests. *Machine Learning*, 45(1):5–32, 2001. <https://link.springer.com/article/10.1023/A:1010933404324>.
- [7] Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, et al. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011. <https://scikit-learn.org/stable/>.
- [8] Rakesh Agrawal and Ramakrishnan Srikant. Fast algorithms for mining association rules. In *Proceedings of the 20th International Conference on Very Large Data Bases (VLDB)*, pages 487–499, 1994. <https://www.vldb.org/conf/1994/P487.PDF>.

Zahvalnice

Ovaj projekat je urađen u okviru kursa Istraživanje Podataka (IP2) na Matematičkom fakultetu, Univerzitet u Beogradu. NBA baza podataka je preuzeta sa Kaggle platforme. Posebna zahvalnost za jednici otvorenog koda za pružanje biblioteka mašinskog učenja (scikit-learn, pandas, numpy, matplotlib, seaborn, mlxtend) koje su omogućile ovu analizu. Zahvaljujem se profesoru i asistentima kursa na smernicama i podršci tokom realizacije projekta.