

FIFA 18

Податочно Множество

Проект по Интелигентни Системи

Давид Симеоновски 141208

Божидар Стеваноски 151502

Андреј Јанчевски 151003

Опис на податочното множество

- Вкупно 185 атрибути за 17 994 фудбалери во FIFA 18.
 - Податоци за секој од играчите, како возраст, клуб, лига, националност, плата и физички атрибути.
 - Сите фудбалски атрибути, како шутирање, одбрана и дриблање.
 - Специјални атрибути како специјални движења и интернационална репутација.
 - Човечки особини и однесување на фудбалерите.
 - Генерални и потенцијални вештини за секоја позиција.

Остварени задачи

- Претпоцесирање на податоци
- Визуелизација на податоци
- Статистичка обработка
- Кластерирање по вештини и позиции
- Предвидување на ранкинг на клубови
- Анализа на асоцијации
- Reverse-engineering на атрибути
- Предвидување на позиција

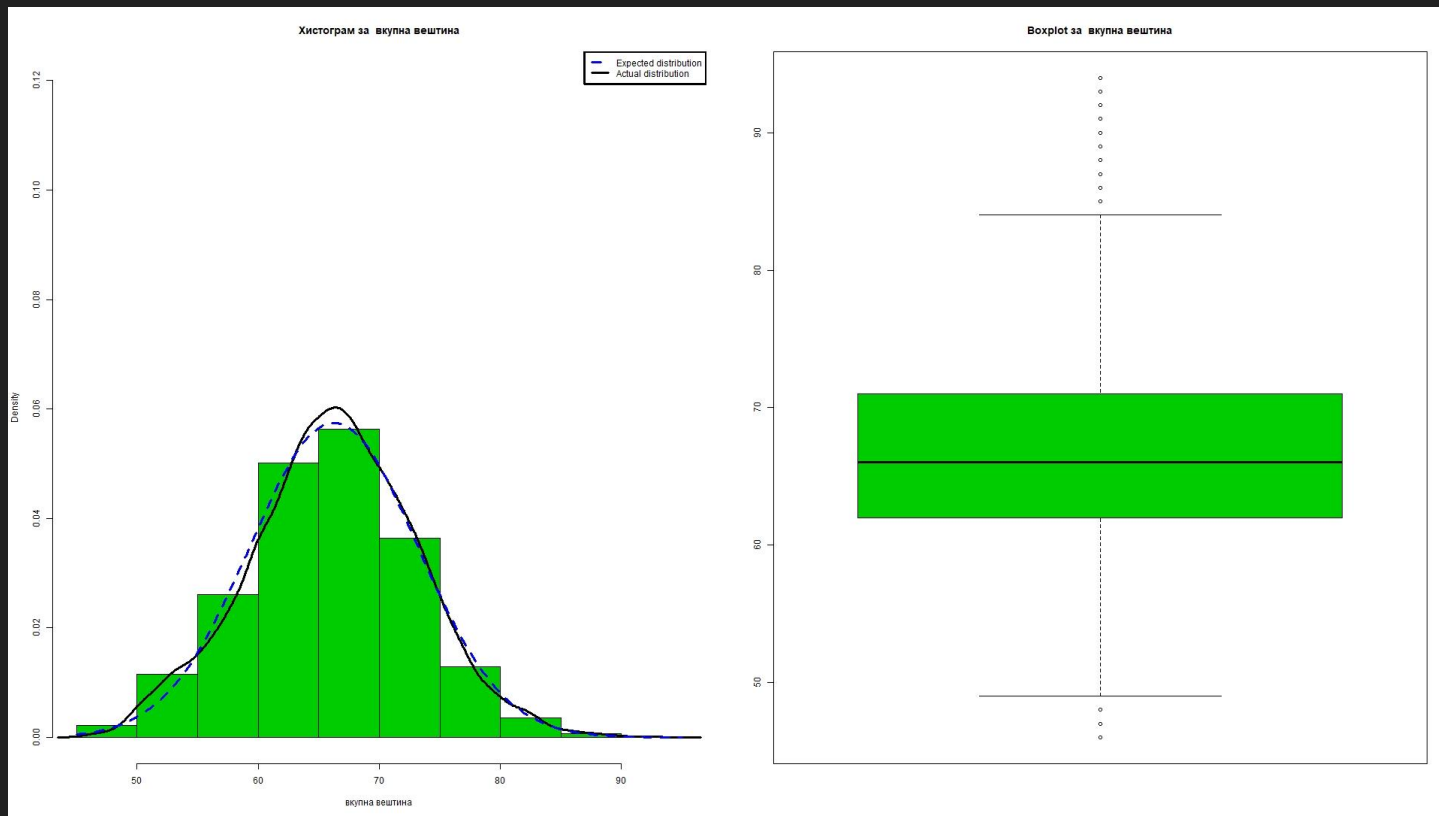
Претпроцесирање на податоците

- Справување со вредности што недостасуваат
 - Откупна клаузула
 - Голмански вештини кај играчите и фудбалски вештини кај голманите
- Детекција и справување со outliers - фудбалери со барем 4 екстремни вредности кај основните атрибути
 - Роналдо, Меси, Суарез, Касемиро, Ферфелис
- Нормализација на нумеричките атрибути
 - Min-max нормализација во интервал [0, 1]

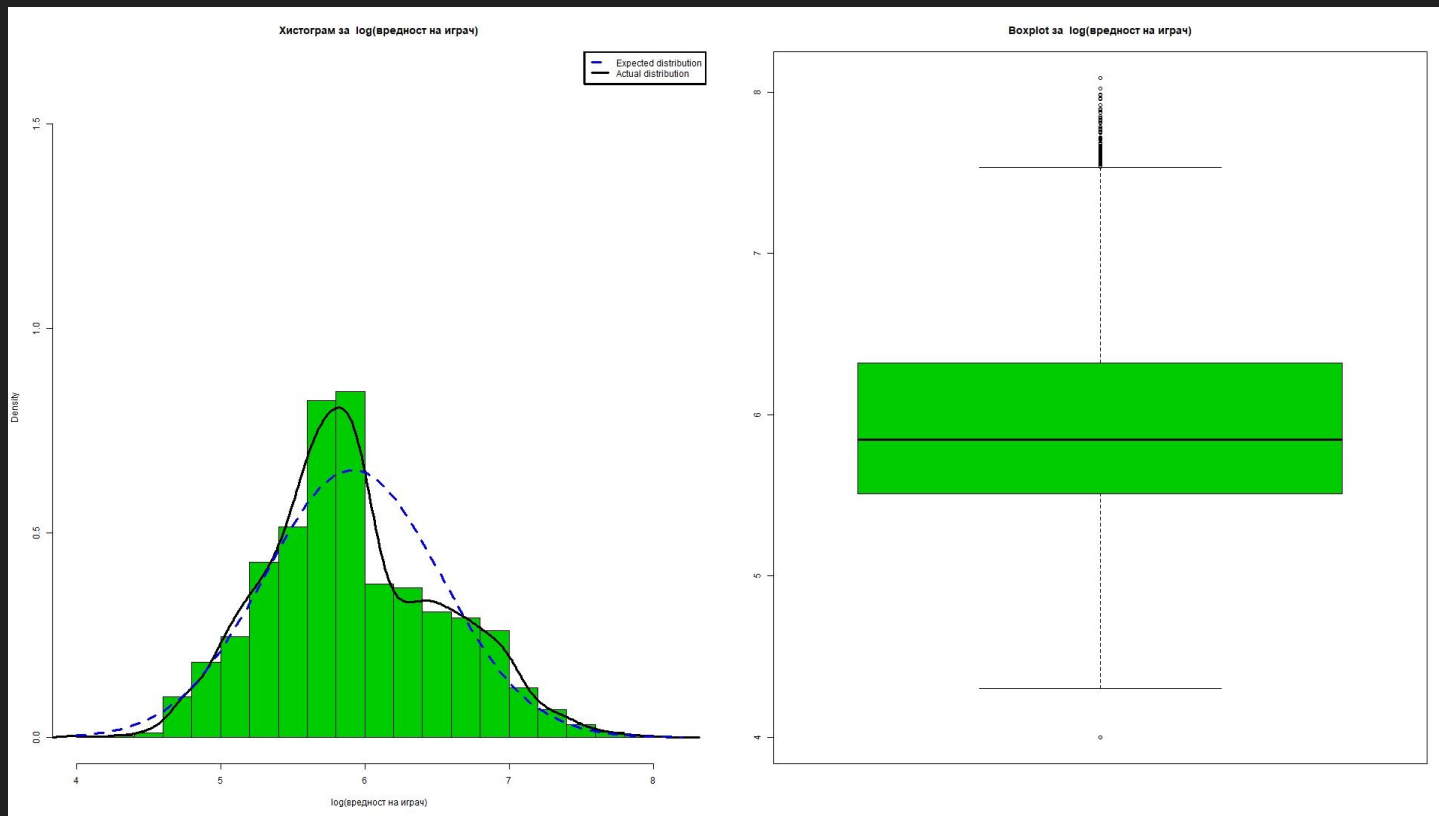
Претпроцесирање на податоците

- **Анализа на распределбите на основните атрибути**
 - За номинални атрибути беше анализирана фреквенцијата на појавување и споредена со рамномерна распределба користејќи хи-квадрат тест
 - За непрекинати атрибути беше анализирана густината на распределба и споредена со рамномерна распределба користејќи Kolmogorov-Smirnov тест
- **Примена на различни техники на визуелизација**
 - хистограм
 - boxplot
 - barplot
 - scatterplot
 - pie chart
 - radar chart

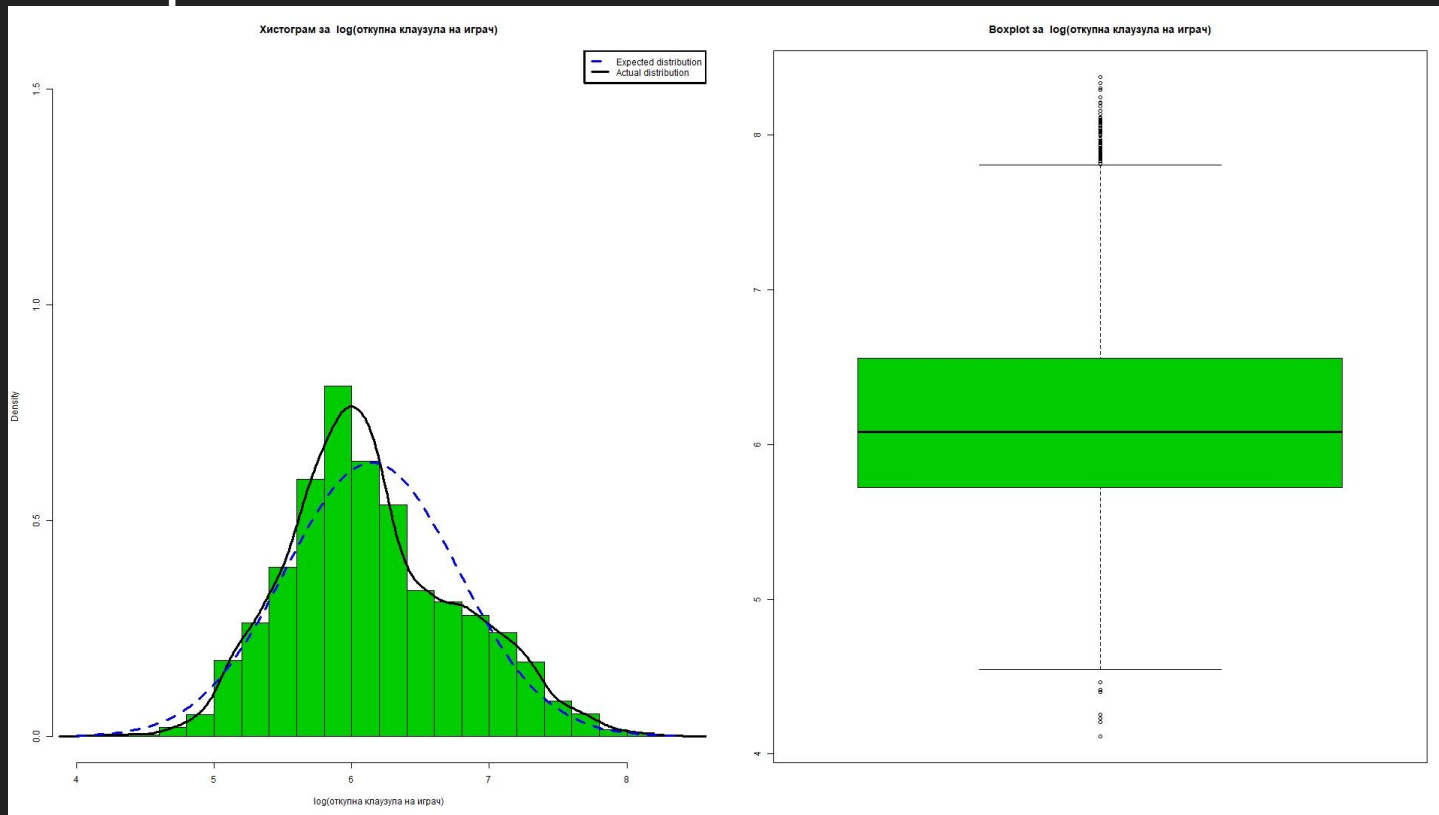
Генерална вештина на фудбалерите



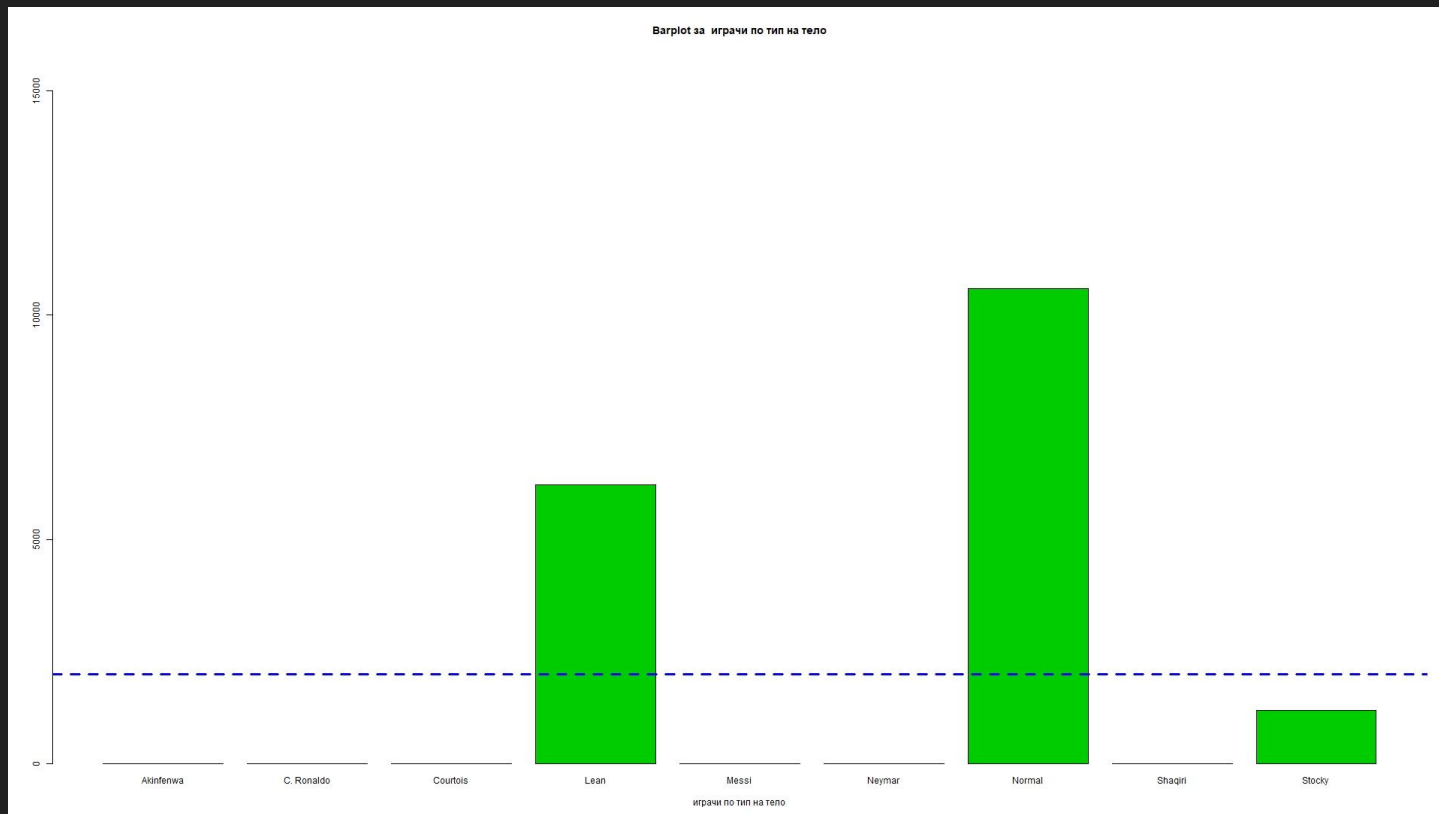
Логаритам од вредноста на фудбалерите



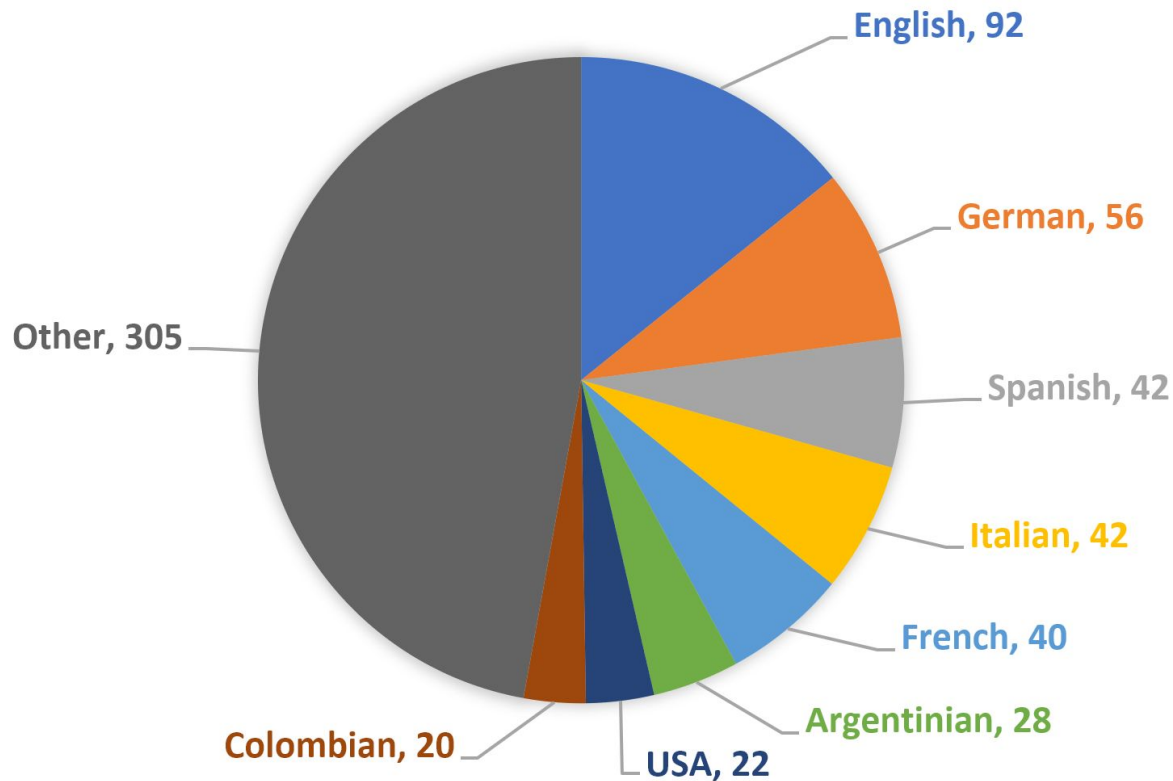
Логаритам од откупната клаузула на фудбалерите



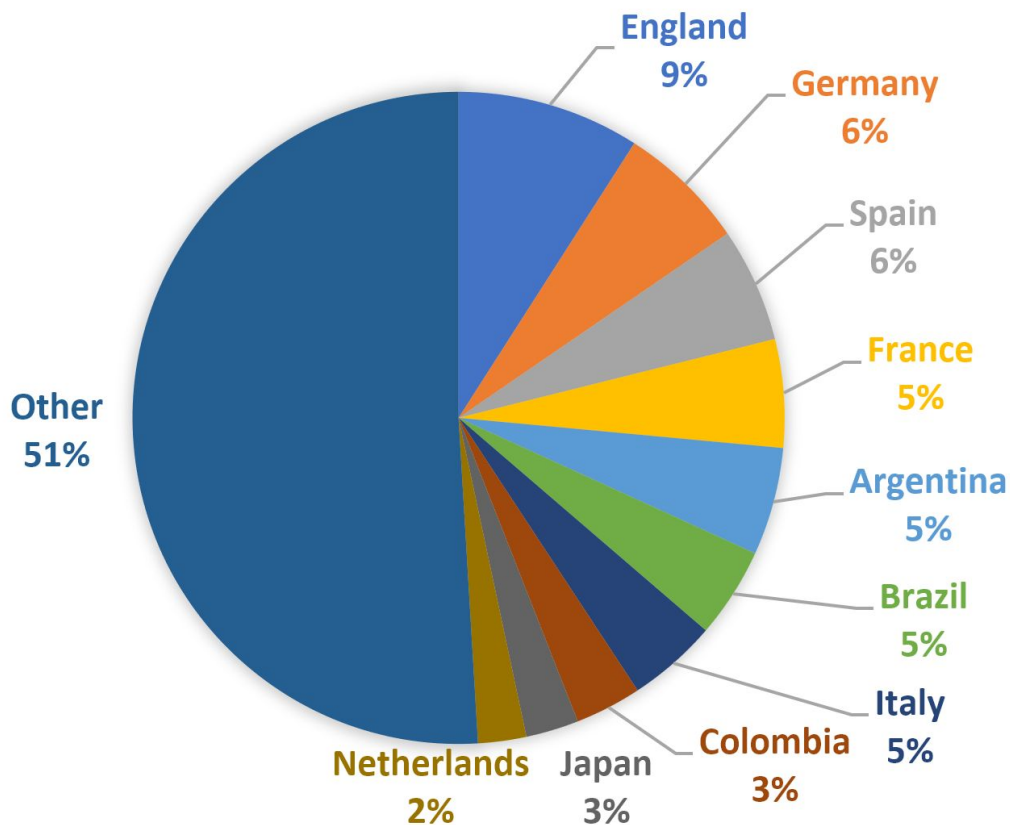
Градба на тело на фудбалерите



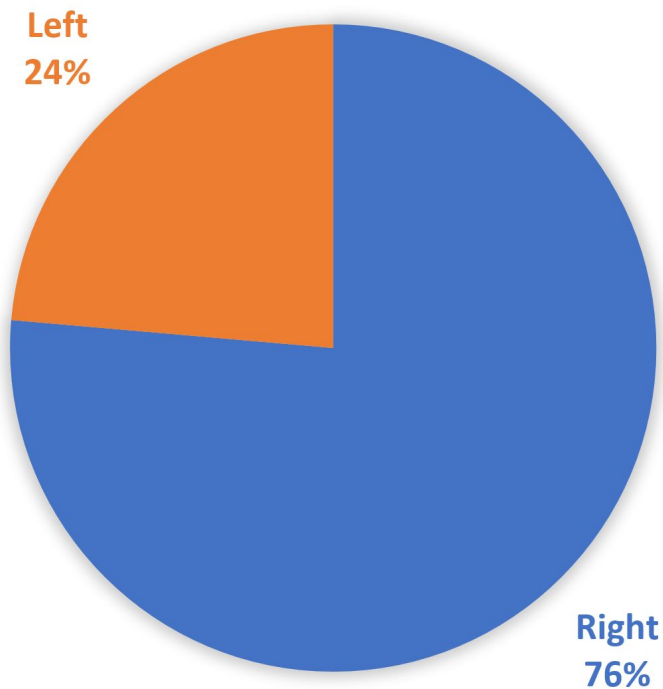
Број на клубови по држава



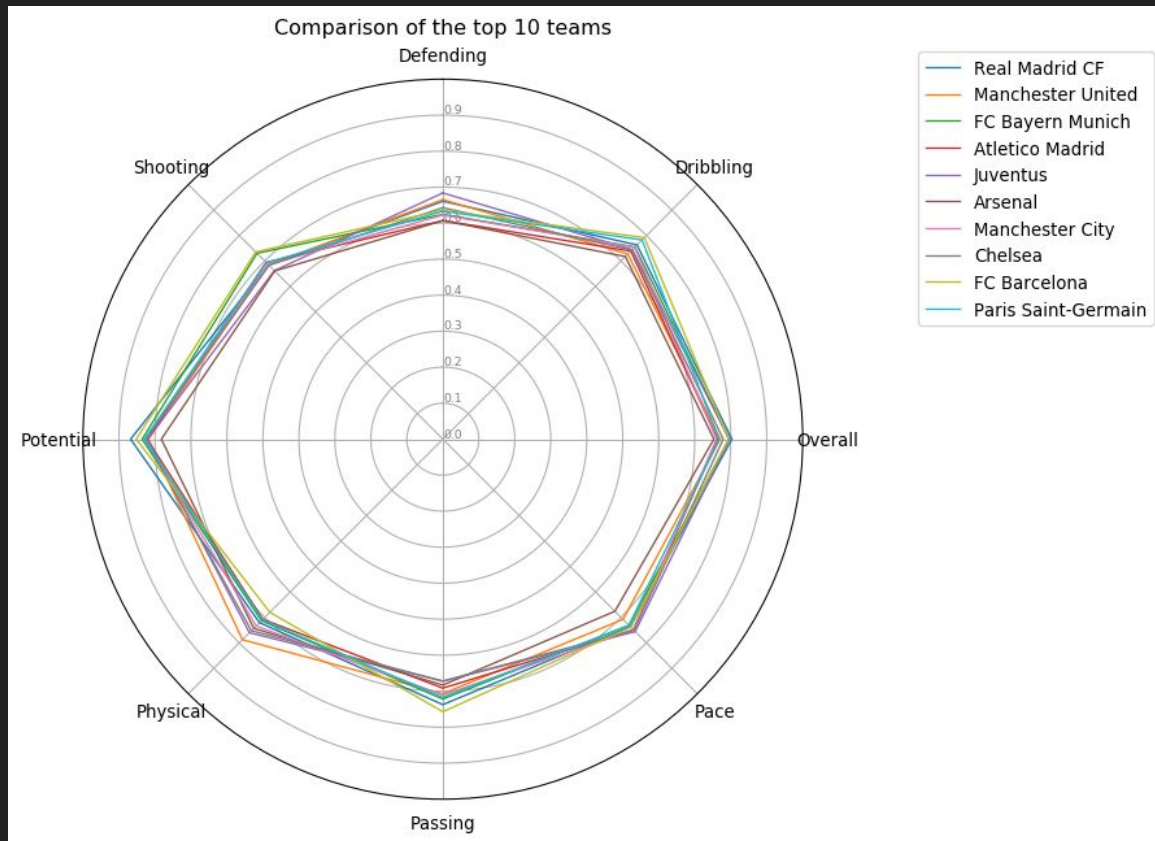
Број на играчи по националност



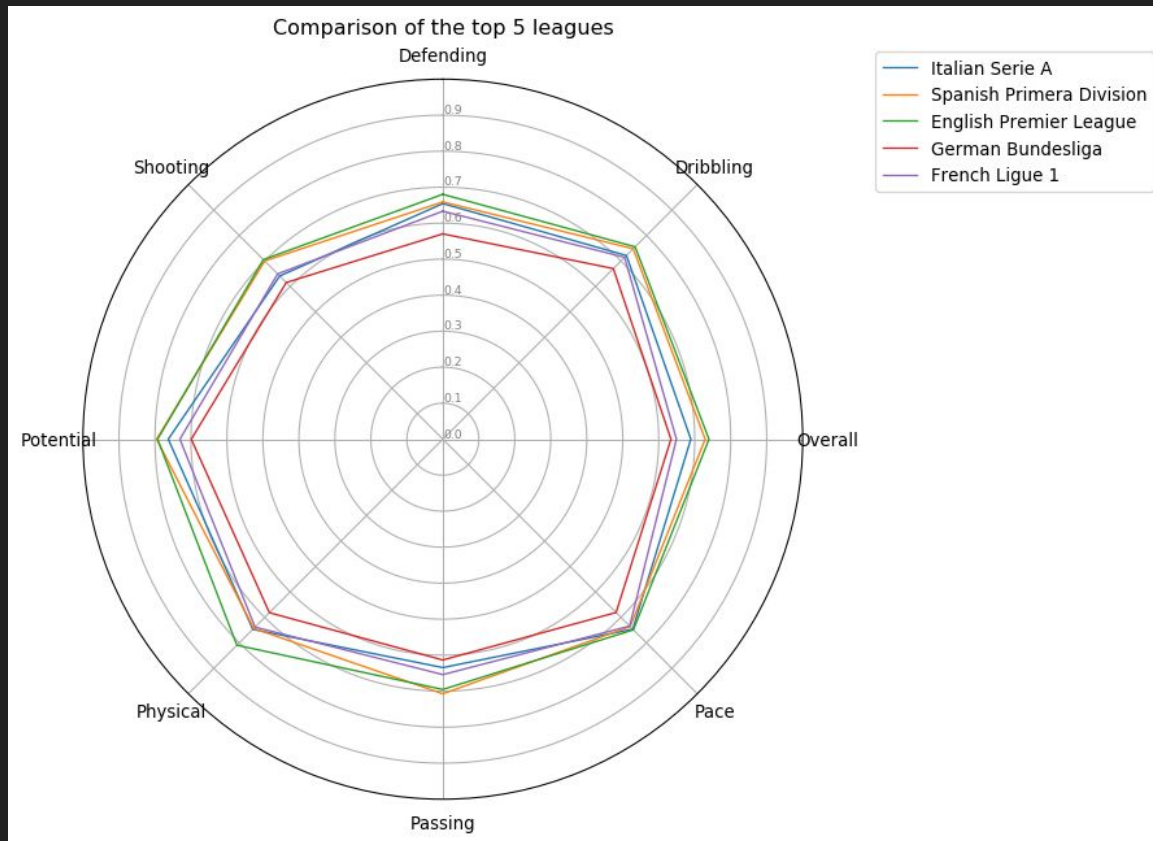
Сооднос на играчи по доминантна нога



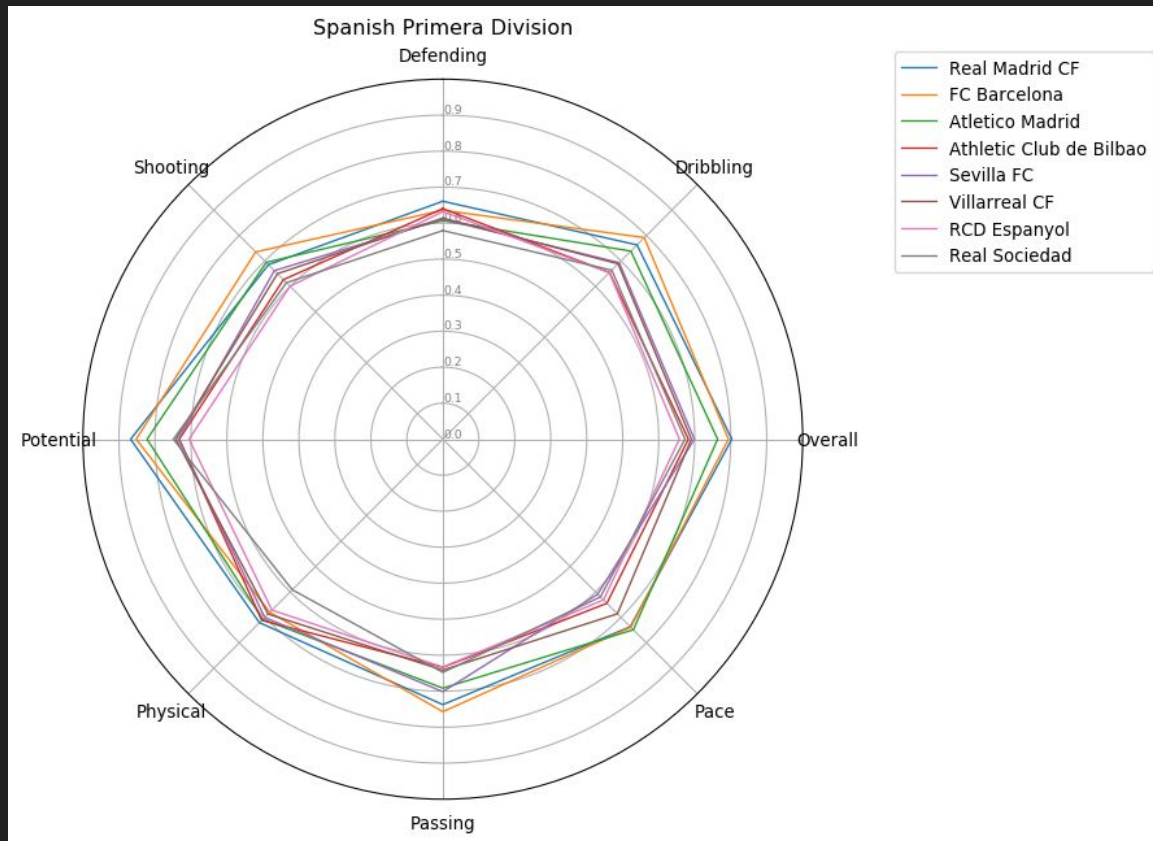
Споредба на вештини на најдобрите 10 клубови



Споредба на вештини на најдобрите 5 лиги



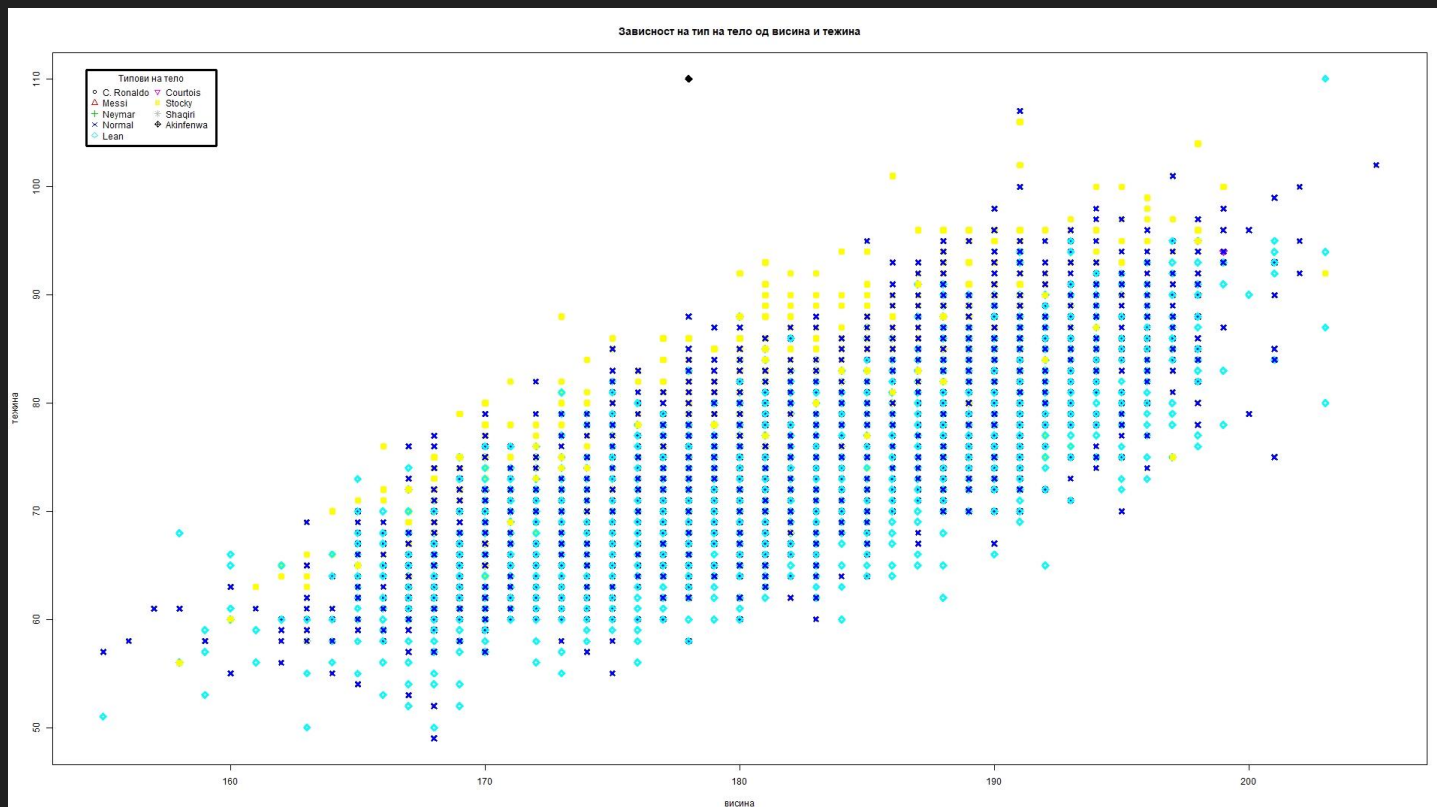
Споредба на тимовите во Шпанска лига



Корелација меѓу парови од атрибути

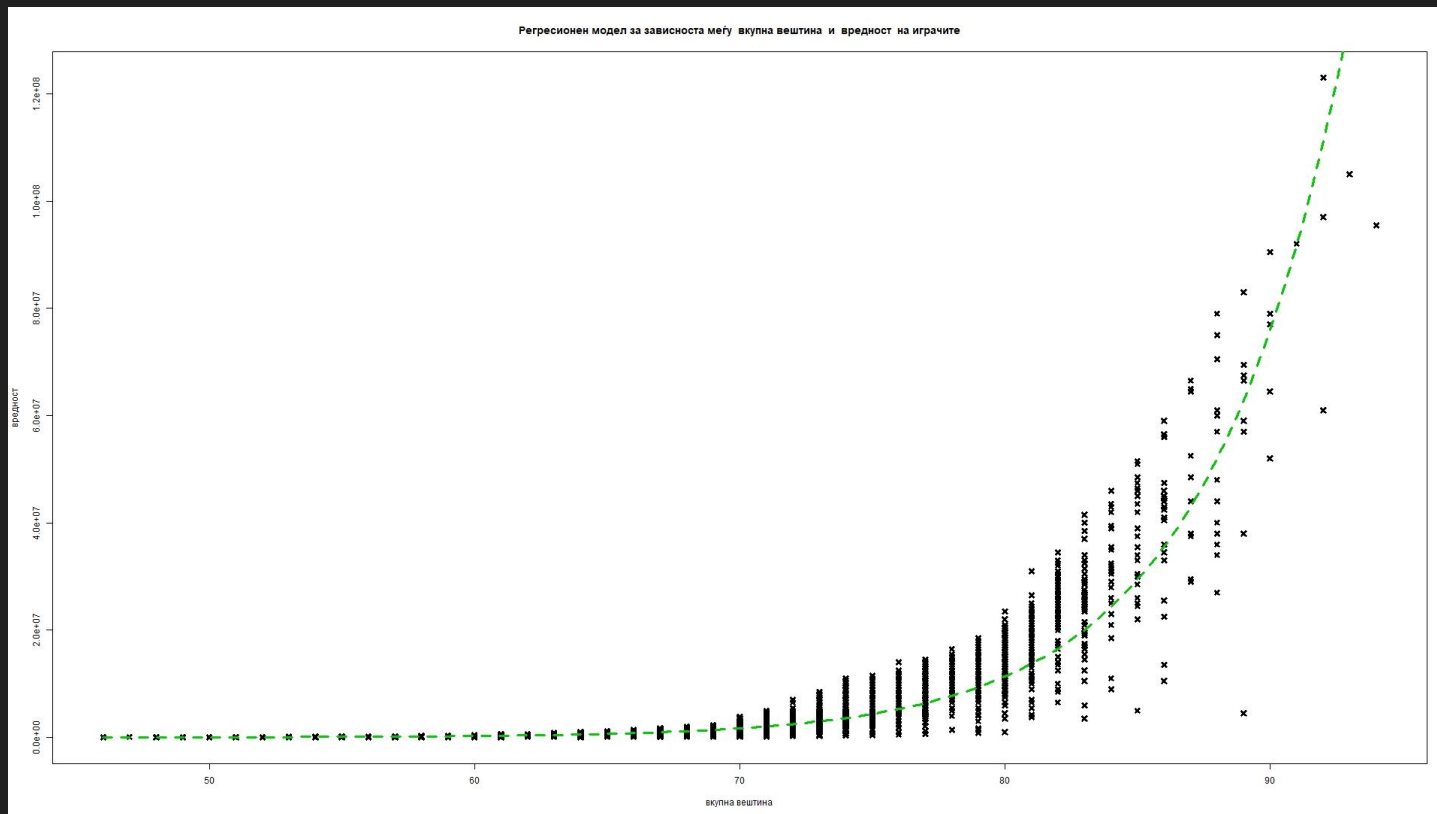
- Линеарна регресија
- Пресметување на коефициентот на корелација
- Потврда на зависност означува редундантност во податочното множество

Зависност на градба на тело од тежина и висина



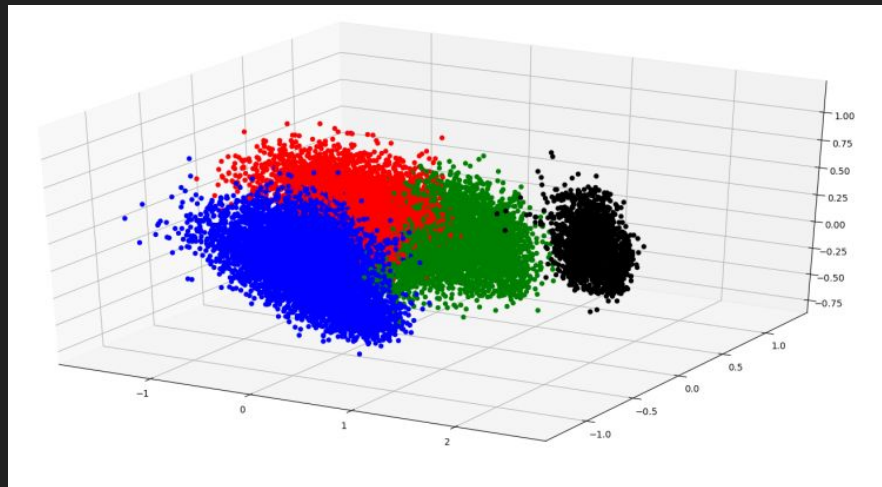
Зависност на вредност од генерална вештина

$R = 0.94$



Кластерирање на играчите според вештините

- Репрезентација на секој играч во повеќе-димензионален простор
 - Кластерирање на 34-D вектори со општи вештини
 - Хиерархиско - 0.747305
 - K - means - 0.728354
 - Кластерирање на 27-D вектори со вештини за позиции
 - Хиерархиско - 0.559631
 - K - means - 0.725075
- Визуелизација по трите најзначајни компоненти после Principal Component Analysis (PCA)



Рангирање на клубовите според просечна вештина

- За секој клуб, се пресметува просек на генералната вештина од најдобрите 20 играчи.
- Клубовите од лигата, се подредуваат според така добиениот просек, и овие рангирања се споредуваат со пласманите на тимовите од вистинскиот свет.

Најдобрите 15 тимови според просечна вештина

Rank	Club	Average overall
1	Real Madrid CF	0.803125
2	Juventus	0.795833
3	FC Bayern Munich	0.795833
4	FC Barcelona	0.79375
5	Chelsea	0.779167
6	Manchester United	0.770833
7	Paris Saint-Germain	0.767708
8	Atletico Madrid	0.763542
9	Manchester City	0.760417
10	Arsenal	0.753125
11	Tottenham Hotspur	0.74375
12	Borussia Dortmund	0.742708
13	Liverpool	0.723958
14	Napoli	0.720833
15	Roma	0.714583

Рангирање на клубовите од Шпанската лига

Rank	Club	Average overall	2017 Standings	Difference
1	Real Madrid CF	0.803125	Real Madrid CF	0
2	FC Barcelona	0.79375	FC Barcelona	0
3	Atletico Madrid	0.763542	Atletico Madrid	0
4	Sevilla FC	0.698958	Sevilla FC	0
5	Villarreal CF	0.690625	Villarreal CF	0
6	Athletic Club de Bilbao	0.682292	Real Sociedad	+1
7	Real Sociedad	0.672917	Athletic Club de Bilbao	-1
8	RCD Espanyol	0.657292	RCD Espanyol	0
9	Valencia CF	0.654167	Deportivo Alaves	+8
10	SD Eibar	0.635417	SD Eibar	0
11	RC Deportivo de La Coruna	0.632292	Malaga CF	+3
12	UD Las Palmas	0.630208	Valencia CF	-3
13	RC Celta de Vigo	0.626042	RC Celta de Vigo	0
14	Malaga CF	0.620833	UD Las Palmas	-2
15	Real Betis Balompie	0.61875	Real Betis Balompie	0
16	CD Leganes	0.601042	RC Deportivo de La Coruna	-5
17	Deportivo Alaves	0.598958	CD Leganes	-1
18	CA Osasuna	0.555208	Real Sporting de Gijon	+1
19	Real Sporting de Gijon	0.552083	CA Osasuna	-1
20	Granada CF	0.551042	Granada CF	0
			Absolute error:	26

Рангирање на клубовите од Германската лига

Rank	Club	Average overall	2017 Standings	Difference
1	FC Bayern Munich	0.795833	FC Bayern Munich	0
2	Borussia Dortmund	0.742708	RB Leipzig	+6
3	Bayer 04 Leverkusen	0.665625	Borussia Dortmund	-1
4	Borussia Monchengladbach	0.661458	TSG 1899 Hoffenheim	+3
5	FC Schalke 04	0.652083	1. FC Koln	+6
6	VfL Wolfsburg	0.65	Hertha BSC Berlin	+3
7	TSG 1899 Hoffenheim	0.64375	SC Freiburg	+9
8	RB Leipzig	0.634375	SV Werder Bremen	+5
9	Hertha BSC Berlin	0.626042	Borussia Monchengladbach	-5
10	Eintracht Frankfurt	0.622917	FC Schalke 04	-5
11	1. FC Koln	0.621875	Eintracht Frankfurt	-1
12	1. FSV Mainz 05	0.608333	Bayer 04 Leverkusen	-9
13	SV Werder Bremen	0.60625	FC Augsburg	+1
14	FC Augsburg	0.590625	Hamburger SV	+1
15	Hamburger SV	0.588542	1. FSV Mainz 05	-3
16	SC Freiburg	0.552083	VfL Wolfsburg	-10
17	FC Ingolstadt 04	0.533333	FC Ingolstadt 04	0
18	SV Darmstadt 98	0.50625	SV Darmstadt 98	0
			Absolute error:	68

Анализа на асоцијации

- Креирање на множество од трансакции и извлекување на асоцијативни правила од човечките особини и однесувањето на фудбалерите
- Apriori алгоритам
- Минимална поддршка поставена на 0.005
- Минимална доверба поставена на 0.45
- Пример добиени правила:
 - (Technical dribbler trait, Selfish trait, Acrobat speciality) ==> (Flair trait) ; confidence=0.857 ; support=0.005
 - (Speedster speciality) ==> (Skilled dribbling trait) ; confidence=0.759 ; support=0.036
 - (Aerial threat speciality) ==> (Strength speciality) ; confidence=0.667 ; support=0.014

Изведување на општите вештини

- Цел - да се откријат формулите со кои во FIFA 18 се пресметуваат општите вештини на фудбалерите користејќи ги специфичните
- Повеќедимензионална линеарна регресија
- Резултати:
 - $\text{Shooting} = \text{round}(0.45 * \text{Finishing} + 0.2 * \text{Long shots} + 0.05 * \text{Penalties} + 0.05 * \text{Positioning} + 0.2 * \text{Shooting} + 0.05 * \text{Volleys})$
 - $\text{Passing} = \text{round}(0.2 * \text{Crossing} + 0.05 * \text{Curve} + 0.05 * \text{Free kick accuracy} + 0.15 * \text{Long passing} + 0.35 * \text{Short passing} + 0.2 * \text{Vision})$
 - $\text{Dribbling} = \text{round}(0.1 * \text{Agility} + 0.05 * \text{Balance} + 0.3 * \text{Ball control} + 0 * \text{Composure} + 0.5 * \text{Dribbling} + 0.05 * \text{Reactions})$
 - $\text{Defending} = \text{round}(0.1 * \text{Heading accuracy} + 0.2 * \text{Interceptions} + 0.3 * \text{Marking} + 0.1 * \text{Sliding tackle} + 0.3 * \text{Standing tackle})$
 - $\text{Physical} = \text{round}(0.2 * \text{Aggression} + 0.05 * \text{Jumping} + 0.25 * \text{Stamina} + 0.5 * \text{Strength})$

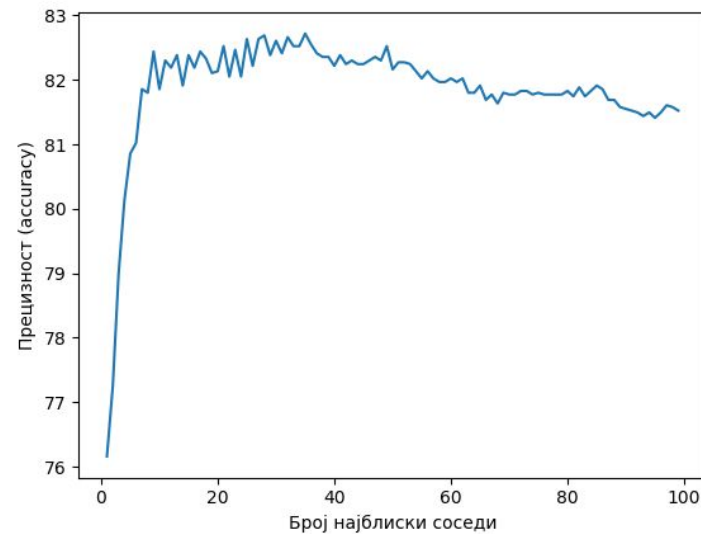
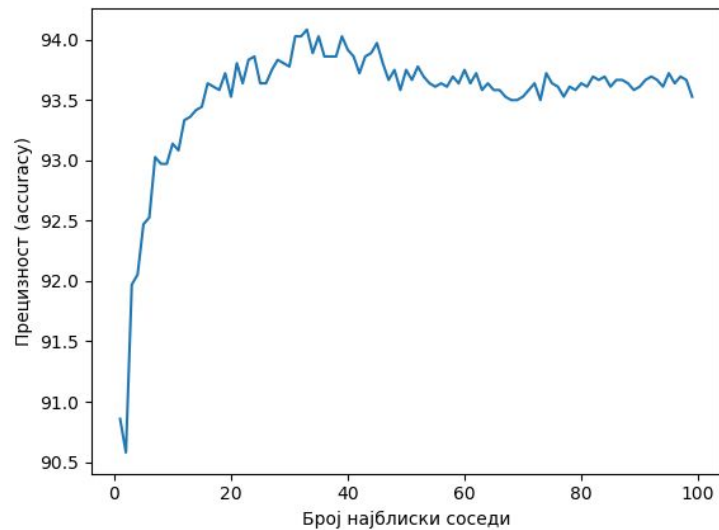
Предвидување на позиција според општи вештини

- Предвидување на генерална позиција
 - Вкупно 4 позиции - напад, среден ред, одбрана и голман
 - Методи на класификација искористени:
 - Невронска мрежа - 96.22%
 - SVM - 94.11%
 - Дрво на одлука - 90.61%
 - К најблиски соседи - 94.06%
- Предвидување на специфична позиција
 - Вкупно 27 можни позиции
 - Методи на класификација искористени:
 - Невронска мрежа - 91.00%
 - SVM - 90.25%
 - Дрво на одлука - 75.10%
 - К најблиски соседи - 82.72%

Невронска мрежа			
4 класи		27 класи	
Архитектура	Точност	Архитектура	Точност
34-4	93.53%	34-27	88.55%
34-32-4	95.72%	34-32-27	91.00%
34-32-32-4	96.22%	34-32-32-27	90.50%
34-32-32-32-4	95.69%	34-32-32-32-27	89.36%

СВМ							
4 класи				27 класи			
РБФ (Radial Basis Function) јадро	Линеарно јадро	Полиномно јадро		РБФ (Radial Basis Function) јадро	Линеарно јадро	Полиномно јадро	
		Квадратно јадро	Кубно јадро			Квадратно јадро	Кубно јадро
93.33%	94.11%	92.14%	87.50%	82.97%	90.25%	79.86%	74.49%

Дрво на одлука		K најблиски соседи	
4 класи	27 класи	4 класи	27 класи
90.61%	75.10%	94.08%	82.72%



ПРАШАЊА?

