

# Pravila pridruživanja o FIFA 18 igračima

## Uvod

Skup o FIFA 18 igračima, preuzet sa adrese <https://www.kaggle.com/thec03u5/fifa-18-demo-player-dataset>, sastoji se od 3 tabele. Tabela Player Personal Data na prvi pogled izgleda ovako:

field1	Unnamed: 0	ID	Name	Age	Photo	Nationality	Flag
0	0	20801	Cristiano Ronaldo	32	<a href="https://cdn.sofifa.org/48/18/players/20801.png">https://cdn.sofifa.org/48/18/players/20801.png</a>	Portugal	<a href="https://cdn.sofifa.org/flags/38.png">https://cdn.sofifa.org/flags/38.png</a>
1	1	158023	L. Messi	30	<a href="https://cdn.sofifa.org/48/18/players/158023.png">https://cdn.sofifa.org/48/18/players/158023.png</a>	Argentina	<a href="https://cdn.sofifa.org/flags/52.png">https://cdn.sofifa.org/flags/52.png</a>
2	2	190871	Neymar	25	<a href="https://cdn.sofifa.org/48/18/players/190871.png">https://cdn.sofifa.org/48/18/players/190871.png</a>	Brazil	<a href="https://cdn.sofifa.org/flags/54.png">https://cdn.sofifa.org/flags/54.png</a>
3	3	176580	L. Suárez	30	<a href="https://cdn.sofifa.org/48/18/players/176580.png">https://cdn.sofifa.org/48/18/players/176580.png</a>	Uruguay	<a href="https://cdn.sofifa.org/flags/60.png">https://cdn.sofifa.org/flags/60.png</a>
4	4	167495	M. Neuer	31	<a href="https://cdn.sofifa.org/48/18/players/167495.png">https://cdn.sofifa.org/48/18/players/167495.png</a>	Germany	<a href="https://cdn.sofifa.org/flags/21.png">https://cdn.sofifa.org/flags/21.png</a>
5	5	188545	R. Lewandowski	28	<a href="https://cdn.sofifa.org/48/18/players/188545.png">https://cdn.sofifa.org/48/18/players/188545.png</a>	Poland	<a href="https://cdn.sofifa.org/flags/37.png">https://cdn.sofifa.org/flags/37.png</a>
6	6	193080	De Gea	26	<a href="https://cdn.sofifa.org/48/18/players/193080.png">https://cdn.sofifa.org/48/18/players/193080.png</a>	Spain	<a href="https://cdn.sofifa.org/flags/45.png">https://cdn.sofifa.org/flags/45.png</a>
7	7	183277	E. Hazard	26	<a href="https://cdn.sofifa.org/48/18/players/183277.png">https://cdn.sofifa.org/48/18/players/183277.png</a>	Belgium	<a href="https://cdn.sofifa.org/flags/7.png">https://cdn.sofifa.org/flags/7.png</a>
8	8	182521	T. Kroos	27	<a href="https://cdn.sofifa.org/48/18/players/182521.png">https://cdn.sofifa.org/48/18/players/182521.png</a>	Germany	<a href="https://cdn.sofifa.org/flags/21.png">https://cdn.sofifa.org/flags/21.png</a>
9	9	167664	G. Higuaín	29	<a href="https://cdn.sofifa.org/48/18/players/167664.png">https://cdn.sofifa.org/48/18/players/167664.png</a>	Argentina	<a href="https://cdn.sofifa.org/flags/52.png">https://cdn.sofifa.org/flags/52.png</a>

Overall	Potential	Club	Club Logo	Value	Wage	Special
94	94	Real Madrid CF	<a href="https://cdn.sofifa.org/24/18/teams/243.png">https://cdn.sofifa.org/24/18/teams/243.png</a>	€95.5M	€565K	2228
93	93	FC Barcelona	<a href="https://cdn.sofifa.org/24/18/teams/241.png">https://cdn.sofifa.org/24/18/teams/241.png</a>	€105M	€565K	2154
92	94	Paris Saint-Germain	<a href="https://cdn.sofifa.org/24/18/teams/73.png">https://cdn.sofifa.org/24/18/teams/73.png</a>	€123M	€280K	2100
92	92	FC Barcelona	<a href="https://cdn.sofifa.org/24/18/teams/241.png">https://cdn.sofifa.org/24/18/teams/241.png</a>	€97M	€510K	2291
92	92	FC Bayern Munich	<a href="https://cdn.sofifa.org/24/18/teams/21.png">https://cdn.sofifa.org/24/18/teams/21.png</a>	€61M	€230K	1493
91	91	FC Bayern Munich	<a href="https://cdn.sofifa.org/24/18/teams/21.png">https://cdn.sofifa.org/24/18/teams/21.png</a>	€92M	€355K	2143
90	92	Manchester United	<a href="https://cdn.sofifa.org/24/18/teams/11.png">https://cdn.sofifa.org/24/18/teams/11.png</a>	€64.5M	€215K	1458
90	91	Chelsea	<a href="https://cdn.sofifa.org/24/18/teams/5.png">https://cdn.sofifa.org/24/18/teams/5.png</a>	€90.5M	€295K	2096
90	90	Real Madrid CF	<a href="https://cdn.sofifa.org/24/18/teams/243.png">https://cdn.sofifa.org/24/18/teams/243.png</a>	€79M	€340K	2165
90	90	Juventus	<a href="https://cdn.sofifa.org/24/18/teams/45.png">https://cdn.sofifa.org/24/18/teams/45.png</a>	€77M	€275K	1961

SLIKA 1: TABELA PLAYER PERSONAL DATA

Field1 i Unnamed: 0 imaju identične vrednosti i označavaju redni broj igrača, ali pošto imamo ID igrača u tabeli, njega ćemo koristiti za identifikaciju igrača, a ova dva atributa ćemo ukloniti. Uklonićemo i attribute Photo, Flag i Club Logo, jer su to adrese slika igrača i neće nam biti korisni u daljoj analizi, kao i atribut Special, koji nema nikakvu posebnu namenu. To ćemo uraditi pomoću Filter taba iz Var čvora.

Potential je atribut koji takođe nećemo koristiti u skupu, ali on označava ocenu koju igrač može dostići ako se odgovarajuće trenira u Carrer mode.

Kako bi nam atributi Value i Wage bili korisniji u analizi, uklonićemo oznake za evro, hiljadu i milion i pretvoriti ih u neprekidne attribute i nazvaćemo ih Value in Euro i Wage in Euro. Atribut Value označava vrednost igrača, a Wage njegovu platu.

U tabeli se nalaze i atributi:

- Name, koji označava ime igrača
- Age, godine igrača
- Nationality, nacionalnost igrača
- Overall, njegovu ukupnu ocenu u FIFA
- Club, klub za koji igrač nastupa

Tabela Player Attribute Data izgleda ovako:

field1	Acceleration	Aggression	Agility	Balance	Ball control	Composure	Crossing	Curve	Dribbling	Finishing
0	89	63	89	63	93	95	85	81	91	94
1	92	48	90	95	95	96	77	89	97	95
2	94	56	96	82	95	92	75	81	96	89
3	88	78	86	60	91	83	77	86	86	94
4	58	29	52	35	48	70	15	14	30	13
5	79	80	78	80	89	87	62	77	85	91
6	57	38	60	43	42	64	17	21	18	13
7	93	54	93	91	92	87	80	82	93	83
8	60	60	71	69	89	85	85	85	79	76
9	78	50	75	69	85	86	68	74	84	91
Free kick accuracy	GK diving	GK handling	GK kicking	GK positioning	GK reflexes	Heading accuracy	ID			
76	7	11	15	14	11	88	20801			
90	6	11	15	14	8	71	158023			
84	9	9	15	15	11	62	190871			
84	27	25	31	33	37	77	176580			
11	91	90	95	91	89	25	167495			
84	15	6	12	8	10	85	188545			
19	90	85	87	86	90	21	193080			
79	11	12	6	8	8	57	183277			
84	10	11	13	7	10	54	182521			
62	5	12	7	5	10	86	167664			
Interceptions	Jumping	Long passing	Long shots	Marking	Penalties	Positioning	Reactions	Short passing		
29	95	77	92	22	85	95	96	83		
22	68	87	88	13	74	93	95	88		
36	61	75	77	21	81	90	88	81		
41	69	64	86	30	85	92	93	83		
30	78	59	16	10	47	12	85	55		
39	84	65	83	25	81	91	91	83		
30	67	51	12	13	40	12	88	50		
41	59	81	82	25	86	85	85	86		
85	32	93	90	63	73	79	86	90		
20	79	59	82	12	70	92	88	75		
Short passing	Shot power	Sliding tackle	Sprint speed	Stamina	Standing tackle	Strength	Vision	Volleys		
83	94	23	91	92	31	80	85	88		
88	85	26	87	73	28	59	90	85		
81	80	33	90	78	24	53	80	83		
83	87	38	77	89	45	80	84	88		
55	25	11	61	44	10	83	70	11		
83	88	19	83	79	42	84	78	87		
50	31	13	58	40	21	64	68	13		
86	79	22	87	79	27	65	86	79		
90	87	69	52	77	82	74	88	82		
75	88	18	80	72	22	85	70	88		

SLIKA 2: TABELA PLAYER ATTRIBUTE DATA

Ova tabela sadrži fizičke, tehničke, mentalne i golmanske sposobnosti igrača.

Tabela Player Playing Position Data izgleda ovako:

field1	CAM	CB	C...	CF	CM	ID	LAM	LB	LCB	LCM	LDM	LF	LM	LS	LW
0	89...	53...	62...	91...	82...	20801	89...	61...	53...	82...	62...	91...	89...	92...	91...
1	92...	45...	59...	92...	84...	158023	92...	57...	45...	84...	59...	92...	90...	88...	91...
2	88...	46...	59...	88...	79...	190871	88...	59...	46...	79...	59...	88...	87...	84...	89...
3	87...	58...	65...	88...	80...	176580	87...	64...	58...	80...	65...	88...	85...	88...	87...
4	\$n...	\$n...	\$n...	\$n...	\$n...	167495	\$n...	\$n...	\$n...	\$n...	\$n...	\$n...	\$n...	\$n...	\$n...
5	84...	57...	62...	87...	78...	188545	84...	58...	57...	78...	62...	87...	82...	88...	84...
6	\$n...	\$n...	\$n...	\$n...	\$n...	193080	\$n...	\$n...	\$n...	\$n...	\$n...	\$n...	\$n...	\$n...	\$n...
7	88...	47...	61...	87...	81...	183277	88...	59...	47...	81...	61...	87...	87...	82...	88...
8	83...	72...	82...	81...	87...	182521	83...	76...	72...	87...	82...	81...	81...	77...	80...
9	81...	46...	52...	84...	71...	167664	81...	51...	46...	71...	52...	84...	79...	87...	82...

LWB	Preferred Positions	RAM	RB	RCB	RCM	RDM	RF	RM	RS	RW	RWB	ST
66...	ST LW	89...	61...	53...	82...	62...	91...	89...	92...	91...	66...	92...
62...	RW	92...	57...	45...	84...	59...	92...	90...	88...	91...	62...	88...
64...	LW	88...	59...	46...	79...	59...	88...	87...	84...	89...	64...	84...
68...	ST	87...	64...	58...	80...	65...	88...	85...	88...	87...	68...	88...
\$n...	GK	\$n...	\$n...	\$n...	\$nu...	\$nu...	\$n...	\$n...	\$n...	\$n...	\$nu...	\$n...
61...	ST	84...	58...	57...	78...	62...	87...	82...	88...	84...	61...	88...
\$n...	GK	\$n...	\$n...	\$n...	\$nu...	\$nu...	\$n...	\$n...	\$n...	\$n...	\$nu...	\$n...
64...	LW	88...	59...	47...	81...	61...	87...	87...	82...	88...	64...	82...
78...	CDM CM	83...	76...	72...	87...	82...	81...	81...	77...	80...	78...	77...
55...	ST	81...	51...	46...	71...	52...	84...	79...	87...	82...	55...	87...

ILUSTRACIJA 3: PLAYER PLAYING POSITION DATA

Ova tabela sadrži ocene igrača na svim pozicijama, kao i njegove omiljene pozicije. Iz Preferred Positions ćemo izvući samo prvu i preimenovaćemo atribut u Position.

Sada ćemo skup spojiti u jednu tabelu i izvršićemo preprocesiranje podataka.

## Preprocesiranje podataka

### Čišćenje duplikata

Faza čišćenja je bila zamišljena da se uradi malo kasnije u radu, ali posmatranjem skupa, primećeno je da postoje duplikati u skupu, pa je prvo urađeno čišćenje duplikata. Njih bismo svakako otkrili kasnije i da nismo slučajno приметили duplikate. To je urađeno pomoću čvora Distinct. Duplikati su eliminisani kako nam ne bi uticali na statistike, pogotovo kasnije kad radimo pravila pridruživanja.

	ID	Name	Record_Count				
				26	207862	M. Ginter	8
1	227796	C. Pulisic	8	27	207494	J. Lingard	8
2	226753	A. Onana	8	28	206591	M. Weiser	8
3	216547	Rafa	8	29	206225	D. Cheryshev	8
4	223874	V. Rongier	8	30	206222	Bigas	8
5	216778	Rúben Semedo	8	31	206198	D. Heintz	8
6	222943	Jony	8	32	204923	M. Sabitzer	8
7	223627	O. Al Soma	8	33	204240	B. Reynet	8
8	221087	Pau López	8	34	203980	K. Fortounis	8
9	225199	V. Koziello	8	35	203895	Pozuelo	8
10	219576	R. Battaglia	8	36	202648	Sergi Darder	8
11	216475	Morales	8	37	201878	V. Eysseric	8
12	227234	L. Tousart	8	38	201144	M. Badelj	8
13	216189	Capa	8	39	201143	A. Mandi	8
14	213400	D. Laxalt	8	40	201119	Ismaily	8
15	213135	D. Origi	8	41	201095	A. Marchesín	8
16	212977	N. Stark	8	42	199626	André André	8
17	212911	Hernâni	8	43	199602	J. Guidetti	8
18	212198	Bruno Fernand...	8	44	199304	Daniilo	8
19	212153	R. Gagliardini	8	45	199069	V. Aboubakar	8
20	211119	Pedro Santos	8	46	199042	C. Aránguiz	8
21	210930	Carles Gil	8	47	198614	Raúl Navas	8
22	210679	Paulo Oliveira	8	48	198584	K. Mbodj	8
23	210333	A. Szymanowski	8	49	198329	Rodrigo	8
24	209307	Tiago Volpi	8	50	198140	R. Funes Mori	8
25	208574	F. Kostić	8	51	198113	M. Höger	8
26	207862	M. Ginter	8	52	197083	D. Caligiuri	8

SLIKA 4: DUPLIKATI

### Izdvajanje karatkeristika

Pomoću Data Audit čvora možemo videti neke osnovne statistike.

Za neprekidne attribute imamo histograme, minimalnu, maksimalnu, srednju vrednost, standardnu devijaciju, pomenost raspodele, medijanu i modu.

Field	Sample Graph	Measurement	Min	Max	Mean	Std. Dev	Skewness ▾	Median	Mode	Unique	Valid
Shot power		Continuous	3	94	55.585	17.397	-0.688	59	68	—	17857
Composure		Continuous	5	96	57.841	12.927	-0.704	60	60	—	17835
Acceleration		Continuous	11	96	64.552	14.917	-0.789	67	68	—	17845
Sprint speed		Continuous	11	96	64.771	14.648	-0.822	67	68	—	17815
Heading accuracy		Continuous	4	94	52.287	17.416	-0.876	56	58	—	17854
Stamina		Continuous	12	95	63.207	15.940	-0.902	66	72	—	17821
Dribbling		Continuous	2	97	54.969	19.005	-1.031	60	65	—	17798
Short passing		Continuous	10	92	58.225	14.936	-1.043	62	64	—	17780
Ball control		Continuous	8	95	58.032	16.820	-1.213	63	64	—	17788
ID		Continuous	16	241219	207654.488	32334.820	-2.297	214106	16*	—	17929

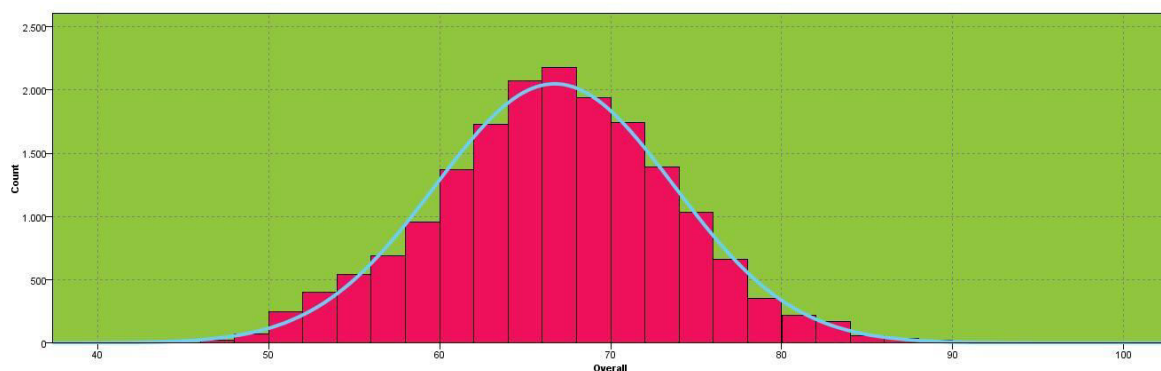
SLIKA 5: DATA AUDIT ČVOR

Vidimo da je najveći broj atributa pomeran ulevo, tj. levi rep raspodele je duži od desnog u odnosu na srednju vrednost.

Overall		Continuous	46	94	66.214	6.969	0.007	66	66	—	17929
Wage in Euro		Continuous	1000	565000	11647.305	23178.616	7.054	4000	1000	—	17681

SLIKA 6: NAJVEĆA I NAJMANJA POMERENOST RASPODELE

Najmanje je pomeran rep raspodele za atribut Overall, a najviše Wage in Euro i oba su pomerena udesno.



SLIKA 7: HISTOGRAM OVERALL SA KRIVOM NORMALNE RASPODELE

Ukupna ocena igrača ima neku približno normalnu raspodelu. Slično izgledaju i raspodele većine pozicija.

A Nationality	Nominal	--	--	--	--	--	--	England	165	17981
---------------	---------	----	----	----	----	----	----	---------	-----	-------

SLIKA 8: NOMINALNI ATRIBUT

Za nominalne atribute nemamo nešto puno informacija, ali vidimo da imamo igrače iz 165 država.

Value	Proportion	%	Count
England		9.09	1629
Germany		6.33	1135
Spain		5.63	1009
France		5.43	973
Argentina		5.36	961
Brazil		4.51	809
Italy		4.45	797
Colombia		3.3	592
Japan		2.62	469
Netherlands		2.39	429
Republic of Ireland		2.33	417
United States		2.12	380
Chile		2.09	374
Sweden		2.05	367
Portugal		2.01	360
Mexico		2.01	360
Denmark		1.93	346
Poland		1.88	337
Norway		1.86	333
Korea Republic		1.84	330
Saudi Arabia		1.84	329
Russia		1.7	305
Scotland		1.67	300
Turkey		1.62	291
Belgium		1.51	271
Austria		1.48	265
Switzerland		1.3	233
Australia		1.27	227
Uruguay		0.85	152
Serbia		0.74	132

SLIKA 9: RASPOREĐENOST IGRAČA PO NACIONALNOSTIMA

Najviše igrača ima iz Engleske, Nemačke i Španije sa od po 1000 igrača, a iz Srbije ima 132 igrača u skupu.

Pomoću Quality taba možemo videti i ekstreme i autlajere.

Field	Measurement	Outliers $\nabla$	Extremes
⬠ GK positioning	Continuous	617	0
⬠ GK reflexes	Continuous	610	0
⬠ GK diving	Continuous	592	0
⬠ GK kicking	Continuous	588	0
⬠ GK handling	Continuous	538	0
⬠ Wage in Euro	Continuous	175	127
⬠ ID	Continuous	147	110
⬠ Sprint speed	Continuous	125	0
⬠ Acceleration	Continuous	99	0
⬠ Composure	Continuous	57	0
⬠ Reactions	Continuous	47	0
⬠ Short passing	Continuous	46	0
⬠ Balance	Continuous	39	0
⬠ Overall	Continuous	33	0
⬠ LM	Continuous	29	0
⬠ RM	Continuous	29	0
⬠ Jumping	Continuous	26	0
⬠ Strength	Continuous	24	0
⬠ Stamina	Continuous	18	0
⬠ Age	Continuous	16	0
⬠ CAM	Continuous	14	0
⬠ LAM	Continuous	14	0
⬠ LW	Continuous	14	0
⬠ RAM	Continuous	14	0
⬠ RW	Continuous	14	0
⬠ Agility	Continuous	9	0
⬠ CF	Continuous	9	0
⬠ LF	Continuous	9	0
⬠ RF	Continuous	9	0
⬠ CM	Continuous	8	0
⬠ LCM	Continuous	8	0

SLIKA 10: AUTLAJERI I EKSTREMNE VREDNOSTI

Ekstremne vrednosti ima samo polje plata, što je i očekivano, s obzirom na oblik raspodele. Izdvojićemo prvih nekoliko ekstrema.



	Name	Wage in Euro
1	L. Messi	565000
2	Cristiano Ronaldo	565000
3	L. Suárez	510000
4	G. Bale	370000
5	R. Lewandowski	355000
6	L. Modrić	340000
7	T. Kroos	340000
8	S. Agüero	325000
9	Sergio Ramos	310000
10	K. Benzema	295000
11	E. Hazard	295000
12	K. De Bruyne	285000
13	Neymar	280000
14	G. Higuaín	275000
15	I. Rakitić	275000
16	A. Sánchez	265000
17	M. Özil	265000
18	Iniesta	260000
19	Marcelo	250000
20	Sergio Busquets	250000

SLIKA 11: EKSTREMI PLATE

Izdvojícemo još i autlejere za Overall:

	Name	Overall
1	Cristiano Ronaldo	94
2	L. Messi	93
3	Neymar	92
4	L. Suárez	92
5	M. Neuer	92
6	R. Lewandowski	91
7	De Gea	90
8	E. Hazard	90
9	T. Kroos	90
10	G. Higuaín	90
11	Sergio Ramos	90
12	G. Buffon	89
13	A. Sánchez	89
14	K. De Bruyne	89
15	T. Courtois	89
16	L. Modrić	89
17	G. Bale	89
18	S. Agüero	89
19	G. Chiellini	89
20	A. Griezmann	88

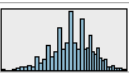
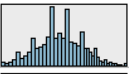
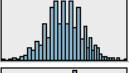
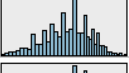
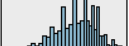
SLIKA 12: AUTLAJERI ZA UKUPNU OCENU IGRAČA

Rad sa nedostajućim vrednostima

Field	Measurement	Null Value	Empty String	White Space /	Blank Value
Strength	Continuous	124	0	0	0
Vision	Continuous	104	0	0	0
Volleys	Continuous	41	0	0	0
CAM	Continuous	2059	0	0	0
CB	Continuous	2059	0	0	0
CDM	Continuous	2059	0	0	0
CF	Continuous	2059	0	0	0
CM	Continuous	2059	0	0	0
LAM	Continuous	2059	0	0	0
LB	Continuous	2059	0	0	0
LCB	Continuous	2059	0	0	0
LCM	Continuous	2059	0	0	0
LDM	Continuous	2059	0	0	0
LF	Continuous	2059	0	0	0
LM	Continuous	2059	0	0	0
LS	Continuous	2059	0	0	0
LW	Continuous	2059	0	0	0
LWB	Continuous	2059	0	0	0
Position	Nominal	0	0	0	0
RAM	Continuous	2059	0	0	0
RB	Continuous	2059	0	0	0
RCB	Continuous	2059	0	0	0
RCM	Continuous	2059	0	0	0
RDM	Continuous	2059	0	0	0
RF	Continuous	2059	0	0	0
RM	Continuous	2059	0	0	0
RS	Continuous	2059	0	0	0
RW	Continuous	2059	0	0	0
RWB	Continuous	2059	0	0	0
ST	Continuous	2059	0	0	0
Club	Nominal	0	248	248	248

SLIKA 13: NEDOSTAJUĆE VREDNOSTI I PRAZNI STRINGOVI

Isto u Quality tabu Data Audit čvora možemo videti nedostajuće vrednosti. Prvo smo tamo gde je klub nepoznat stavili Unknown. Dalje, vidimo da ocene pozicija imaju po 2059 nedostajućih vrednosti. Međutim, razgledanjem skupa smo videli da Manuel Neuer nema ocene za druge pozicije i zato ćemo prvo izvući golmane i skupa i pogledati neke njihove osobine.

⚡ GK diving		Continuous	39	91	65.310	7.890	0.068	--	1998
⚡ GK handling		Continuous	43	91	62.879	7.908	0.225	--	1997
⚡ GK kicking		Continuous	35	95	61.536	7.889	0.252	--	2005
⚡ GK positioning		Continuous	38	91	63.051	8.824	-0.057	--	1998
⚡ GK reflexes		Continuous	37	90	66.183	8.358	0.016	--	1995

SLIKA 14: RASPODELE GOLMANSKIH ATRIBUTA KOD GOLMANA

Prvo što primećujemo je da golmanske osobine sad imaju mnogo lepše i približnije normalnoj raspodele. Međutim, sada ako pogledamo nedostajuće vrednosti:

Field	Measurement	Null Value	Valid Records	% Complete
⚡ CB	Continuous	2059	0	0
⚡ CDM	Continuous	2059	0	0
⚡ CF	Continuous	2059	0	0
⚡ CM	Continuous	2059	0	0
⚡ LAM	Continuous	2059	0	0
⚡ LB	Continuous	2059	0	0
⚡ LCB	Continuous	2059	0	0
⚡ LCM	Continuous	2059	0	0
⚡ LDM	Continuous	2059	0	0
⚡ LF	Continuous	2059	0	0
⚡ LM	Continuous	2059	0	0
⚡ LS	Continuous	2059	0	0
⚡ LW	Continuous	2059	0	0
⚡ LWB	Continuous	2059	0	0
A Position	Nominal	0	2059	100
⚡ RAM	Continuous	2059	0	0
⚡ RB	Continuous	2059	0	0
⚡ RCB	Continuous	2059	0	0
⚡ RCM	Continuous	2059	0	0
⚡ RDM	Continuous	2059	0	0
⚡ RF	Continuous	2059	0	0
⚡ RM	Continuous	2059	0	0
⚡ RS	Continuous	2059	0	0
⚡ RW	Continuous	2059	0	0
⚡ RWB	Continuous	2059	0	0
⚡ ST	Continuous	2059	0	0

SLIKA 15: NEDOSTAJUĆE VREDNOSTI KOD GOLMANA

Vidimo da svih onih 2059 nedostajućih vrednosti za pozicije imaju golmani! Nećemo naravno isključiti sve golmane iz skupa, samo ćemo isfiltrirati kolone sa ocenama pozicija.

Ostale nedostajuće vrednosti ćemo zameniti sa nasumičnim brojevima iz normalne raspodele za taj atribut, za sve kolone, osim za Value i Wage in Euro. Pošto se te dve raspodele drastično razlikuju od normalne, a

nisu približne ni uniformnoj raspodeli, promenili bismo oblik raspodele ako bismo nedostajuće vrednosti menjali nasumičnim vrednostima iz te dve raspodele. Takođe, ako bismo nekom konstantom menjali nedostajuće vrednosti, nagomilavale bi nam se te vrednosti u skupu. Zato ćemo, u slučaju kada u analizi koristimo platu i vrednost igrača, eliminisati one redove koji ne poseduju jedan od te dve vrednosti.

## **Uzorkovanje**

Za pravila pridruživanja u kojima se koristi ocena pozicije, korišćeni su svi igrači koji nisu golmani. Tamo gde se ne koriste ocene pozicija, koriste se i svi golmani. Ako se koristi kao telo ili glava plata ili cena fudbalera, onda se koristi uzorak koji ne sadrži igrače (i golmane i ostale) koji nemaju vrednost ili Value ili Wage in Euro.

Na kraju, uzeto je pristrasno uzorkovanje, tj. urađena su pravila pridruživanja za Lige 5, 5 najbolje rangiranih liga. Izdvojeni su igrači iz klubova koji su sezoni 2017/2018 igrali u Premijer ligi, La ligi, Bundesligi, Ligi 1 i Seriji A.

## **Transformacija podataka**

Svi podaci koji su korišćeni u pravilima pridruživanja su bili u tabelarnom formatu. Izvršena je diskretizacija svih korišćenih neprekidnih podataka. Svi neprekidni podaci su podeljeni u 3 intervala, tako da svaki interval ima jednak broj elemenata. Korišćen je Binning čvor za to. Dalje su svi kategorički podaci korišćeni u pravilima pridruživanja binarizovani korišćenjem čvora SetToFlag.

## Pravila pridruživanja

### 1.

Hoćemo da vidimo kako utiču ubrzanje (Acceleration) i brzina trčanja (Sprint speed) na poziciju koju igrač igra. Prvo je iskorišćen čvor Apriori, sa podrazumevanim vrednostima, podrška tela je 10%, a pouzdanost je 80%. Dobijeno je 8 pravila:

Consequent	Antecedent	Support...	Confid...	Rule Su...	Lift
Sprint speed Category_Fine	Position_GK Acceleration Category_Fine	11,189	99,252	11,105	3,06
Sprint speed Category_Fine	Position_GK	11,289	98,814	11,155	3,046
Acceleration Category_Fine	Position_GK Sprint speed Category_Fine	11,155	99,55	11,105	3,002
Acceleration Category_Fine	Position_GK	11,289	99,111	11,189	2,989
Acceleration Category_Fine	Sprint speed Category_Fine	32,439	82,875	26,884	2,499
Sprint speed Category_Fine	Acceleration Category_Fine	33,159	81,077	26,884	2,499
Acceleration Category_Great	Sprint speed Category_Great	36,237	82,715	29,974	2,299
Sprint speed Category_Great	Acceleration Category_Great	35,986	83,292	29,974	2,299
















SLIKA 16: PODRAZUMEVANI APRIORI

Najbolje pravilo je da igrač ako je golman i ima loše ubrzanje, onda on ima i lošu brzinu. U ovom skupu, 11% su golmani koji imaju loše ubrzanje, u 99% slučajeva oni imaju i lošu brzinu. Oko 11% skupa su zapravo golmani sa lošim ubrzanjem i brzinom. Lift mera od 3,06 nam govori da se 3 puta češće javlja loša brzina kod igrača koji su golmani sa lošim ubrzanjem, nego kod drugih igrača.

Takođe, oni igrači koji imaju lošu brzinu, imaju 2,5 veće šanse da imaju i loše ubrzanje. Loša brzina se javlja kod 32,5% igrača, od toga njih skoro 83% imaju i loše ubrzanje. Igrači koji imaju i lošu brzinu i ubrzanje javljaju se u oko 27% skupa.

Uočimo još da ako igrač ima odlično ubrzanje, ima oko 2 puta veće šanse da ima i odličnu brzinu. Trećinu skupa čine igrači koji imaju i odličnu brzinu i odlično ubrzanje.

Zašto smo dobili ovakve rezultate? Sve pozicije i ubrzanje i brzinu smo postavili da budu i glava i telo, a u skupu nalazimo samo golmane.

Value	Proportion	%	Count
CB		15.04	2697
ST		12.67	2271
GK		11.29	2024
CM		11.0	1973
CDM		7.73	1386
RM		7.51	1347
LM		7.36	1319
LB		7.3	1308
RB		6.66	1194
CAM		6.27	1124
RW		2.38	426
LW		2.36	424
CF		1.14	205
LWB		0.67	120
RWB		0.62	111

SLIKA 17: RASPOREĐENOST IGRAČA PO POZICIJAMA

Odgovor nalazimo kad pogledamo kako su raspoređene vrednosti pozicija. Golmana ima među najzastupljenijima u skupu, dok levi i desni krilni bek ne čine čak ni jedan posto skupa. Zato smo te pozicije i izbacili iz skupa, a podršku tela smo stavili da bude 1% da bi levo i desno krilo imali bar kakve takve šanse da se pojave u skupu. Pouzdanost smo ostavili da bude 80%. Sada smo dobili sledeće rezultate:

Consequent	Antecedent	Suppo...	Confide...	Rule ...	Lift
Sprint speed Category_Fine	Position_GK Acceleration Category_Fine	11,189	99,252	11,105	3,06
Sprint speed Category_Fine	Position_GK	11,289	98,814	11,155	3,046
Acceleration Category_Fine	Position_GK Sprint speed Category_Fine	11,155	99,55	11,105	3,002
Acceleration Category_Fine	Position_GK	11,289	99,111	11,189	2,989
Acceleration Category_Great	Position_LW Sprint speed Category_Great	1,768	94,322	1,668	2,621
Acceleration Category_Fine	Position_CB Sprint speed Category_Fine	7,948	86,807	6,899	2,618
Acceleration Category_Great	Position_RW Sprint speed Category_Great	1,673	93,333	1,562	2,594
Acceleration Category_Great	Position_RM Sprint speed Category_Great	4,981	90,594	4,512	2,517
Sprint speed Category_Great	Position_LW Acceleration Category_Great	1,835	90,881	1,668	2,508
Acceleration Category_Great	Position_LM Sprint speed Category_Great	4,78	90,198	4,311	2,506
Acceleration Category_Fine	Sprint speed Category_Fine	32,439	82,875	26,884	2,499
Sprint speed Category_Fine	Acceleration Category_Fine	33,159	81,077	26,884	2,499
Sprint speed Category_Great	Position_RW Acceleration Category_Great	1,729	90,323	1,562	2,493
Sprint speed Category_Great	Position_RM Acceleration Category_Great	5,159	87,459	4,512	2,414
Sprint speed Category_Great	Position_ST Acceleration Category_Great	5,684	87,242	4,958	2,408
Sprint speed Category_Great	Position_LM Acceleration Category_Great	4,964	86,854	4,311	2,397
Acceleration Category_Great	Position_CAM Sprint speed Category_Great	2,794	85,828	2,398	2,385
Sprint speed Category_Great	Position_RB Acceleration Category_Great	3,135	85,409	2,677	2,357
Acceleration Category_Great	Sprint speed Category_Great	36,237	82,715	29,974	2,299
Sprint speed Category_Great	Acceleration Category_Great	35,986	83,292	29,974	2,299
Sprint speed Category_Great	Position_LB Acceleration Category_Great	3,547	82,862	2,939	2,287
Acceleration Category_Great	Position_ST Sprint speed Category_Great	6,085	81,485	4,958	2,264
Acceleration Category_Great	Position_RB Sprint speed Category_Great	3,302	81,081	2,677	2,253
Acceleration Category_Great	Position_LB Sprint speed Category_Great	3,642	80,704	2,939	2,243

SLIKA 18: APRIORI 1% PODRŠKA TELA

Sad vidimo i neke druge pozicije u skupu. U oko 1,7% su igrači koji igraju na poziciji levog krila i imaju odličnu brzinu. Od toga oko 93% ima i odlično ubrzanje. Levi krilni igrači sa odličnom brzinom i

ubrzanjem čine oko 1,6% skupa. Zaključujemo i da ako su igrači levo krilni i imaju odličnu brzinu, 2,5 puta je verovatnije i da imaju odlično ubrzanje. Slično je i sa igračima koji igraju desno krilo.

Centralni bek koji ima lošu brzinu ima 2,5 puta veću šansu da ima i loše ubrzanje nego ostali igrači.

Oko 90 posto igrača koji igraju desnog veznog i imaju odličnu brzinu imaju i odlično ubrzanje. Takvi igrači čine oko 4 posto skupa. Slično je i sa levim veznim.

Napadači sa odličnim ubrzanjem čine oko 5,5 posto skupa. Od njih, 87 posto ima i odličnu brzinu. Napadači sa obe karakteristike čine oko 5 posto skupa.

Levi bek sa odličnim ubrzanjem ima oko 2 puta veće šanse da ima i odličnu brzinu. Isto je i sa desnim bekom.

Primenili smo i Carma čvor sa podrazumevanom podrškom i pouzdanošću pravila od 20%. Sledeći rezultati su dobijeni:

Consequent	Antecedent	Supp...	Confid...	Rule S...	Lift
Sprint speed Category_Fine	Acceleration Category_Fine	33,159	81,077	26,884	2,499
Acceleration Category_Fine	Sprint speed Category_Fine	32,439	82,875	26,884	2,499
Sprint speed Category_Great	Acceleration Category_Great	35,986	83,292	29,974	2,299
Acceleration Category_Great	Sprint speed Category_Great	36,237	82,715	29,974	2,299

SLIKA 19: PODRAZUMEVANA CARMA

Neka pravila su, naravno ista kao što smo i dobili prvi put. Ali što smo sad dobili samo 4 pravila i ni jedno od njih ne uključuje ni jednu poziciju? Pošto je podrška pravila 20%, to bi značilo da bar na jednu četvrtinu skupa bude primenljivo pravilo. Ali ni jedna pozicija sama ne čini 20% skupa (najzastupljenija je pozicija CB sa oko 15%). Radi poređenja, stavili smo minimalnu podršku pravila za jedan manju nego što je najmanja dobijena sa Apriori algoritmom, kako bi bilo uključeno to pravilo sa najmanjom podrškom, i istu pouzdanost. Sada smo dobili 25 pravila, umesto 24. Zašto? Zato što sada imamo i pravila koja uključuju više od jedne stavke u glavi, gde kod Apriori algoritma to nismo dobili. Analiziraćemo to jedno novo pravilo:

Consequent	Antecedent	Supp...	Confid...	Rule S...	Lift
Acceleration Category_Fine	Position_GK	11,289	98,37	11,105	3,659
Sprint speed Category_Fine					

SLIKA 20: CARMA SA VREDNOSTIMA DOBIJEM APRIORI ČVOROM

To je zapravo i najbolje pravilo koje smo dobili do sad. Golmani čine oko 11, 3 posto skupa. Od toga, čak 98 posto golmana ima i loše ubrzanje i lošu brzinu. Čak 11% skupa čine golmani sa lošom i brzinom i ubrzanjem, dakle, ako je igrač golman, on ima 3,7 puta veću mogućnost da ima lošu brzinu i ubrzanje.



## 2.

Sada smo hteli da proverimo uticaj brzine i ubrzanja na krila. Uzeta je ocena bolje pozicije (levog ili desnog krila) za svakog igrača (isključujući golmane, jer golmani nemaju ocene za druge pozicije). Podrazumevani Apriori algoritam daje sledeće rezultate:

Consequent	Antecedent	Suppor...	Confid...	Rule S...	Lift
AccelerationCategory_Fine	Winger ScoreCategory_Fine Sprint speedCategory_Fine	19,132	82,682	15,819	2,727
Sprint speedCategory_Fine	AccelerationCategory_Fine Winger ScoreCategory_Fine	18,441	85,782	15,819	2,631
AccelerationCategory_Great	Sprint speedCategory_Great Winger ScoreCategory_Great	20,993	86,763	18,214	2,562
Sprint speedCategory_Great	AccelerationCategory_Great Winger ScoreCategory_Great	21,823	83,463	18,214	2,459
Sprint speedCategory_Great	AccelerationCategory_Great	33,864	80,059	27,111	2,358

SLIKA 21: PODRAZUMEVANI APRIORI

Dakle, ako je igrač ocenjen kao loš krilni i ima lošu brzinu, on ima 2,7 puta veću šansu da ima i loše ubrzanje. Nešto manje od 16% skupa čine igrači koji čine loša krila sa lošom brzinom i ubrzanjem.

S druge strane, ako je igrač odlično ocenjen kao krilo i ima odličnu brzinu, onda on ima 2,5 veću šansu da ima i odlično ubrzanje.

Primenom Carma algoritma sa podrazumevanim vrednostima, dobija se 27 pravila. Ali do sada nismo dobijali neka baš mnogo zanimljiva pravila, jer skoro trećinu skupa, i prvog i drugog na koji smo primenljivali pravila pridruživanja, čine zajedno odlična brzina i ubrzanje. Malo manje čine loša brzina i loše ubrzanje. Zato smo sad filtrirali model, tako da glava bude samo ocena krila. Dobili smo 14 pravila:

Consequent	Antecedent	Suppo...	Confide...	Rule S...	Lift
Sprint speedCategory_Fine Winger ScoreCategory_Fine	AccelerationCategory_Fine	30,318	52,178	15,819	2,727
Winger ScoreCategory_Fine AccelerationCategory_Fine	Sprint speedCategory_Fine	32,606	48,515	15,819	2,631
Sprint speedCategory_Great Winger ScoreCategory_Great	AccelerationCategory_Great	33,864	53,788	18,214	2,562
Winger ScoreCategory_Great AccelerationCategory_Great	Sprint speedCategory_Great	33,945	53,658	18,214	2,459
Winger ScoreCategory_Fine	Sprint speedCategory_Fine AccelerationCategory_Fine	24,219	65,317	15,819	2,072
Winger ScoreCategory_Fine	AccelerationCategory_Fine	30,318	60,825	18,441	1,929
Winger ScoreCategory_Fine	Sprint speedCategory_Fine	32,606	58,677	19,132	1,861
Winger ScoreCategory_Great	Sprint speedCategory_Great AccelerationCategory_Great	27,111	67,185	18,214	1,815
Winger ScoreCategory_Great	AccelerationCategory_Great	33,864	64,445	21,823	1,741
Winger ScoreCategory_Great	Sprint speedCategory_Great	33,945	61,845	20,993	1,671
Winger ScoreCategory_Good	AccelerationCategory_Good	35,819	38,845	13,914	1,235
Winger ScoreCategory_Good	Sprint speedCategory_Good	33,449	38,797	12,977	1,233
Winger ScoreCategory_Great	Sprint speedCategory_Good	33,449	33,308	11,141	0,9
Winger ScoreCategory_Great	AccelerationCategory_Good	35,819	31,666	11,342	0,856

SLIKA 22: CARMA, FILTRIRANI PODRAZUMEVANI MODEL

Nešto manje od 16 posto skupa čine loša krila sa lošom brzinom i ubrzanjem. Dakle, oni sa lošom brzinom i ubrzanjem su 2 puta verovanije loši krilni igrači.

Skoro 2 puta je verovatnije da su odlični krilni igrači ako imaju odličnu brzinu i ubrzanje.

Skoro 40% igrača koji imaju dobru brzinu ili ubrzanje su dobra krila.

### 3.

Tražimo odličnog zadnjeg veznog, ocene karakteristika su postavljene za telo, a za glavu je postavljena ocena koju igrač ima kao zadnji vezni (najveća od 3 pozicije koje pripadaju zadnjem veznom). Podrazumevanim Apriorijem dobijemo 1232 pravila. Filtriravši samo ona koja daju odličnog zadnjeg veznog, dobijamo 778 pravila. Zato smo ponovo pokrenuli Apriori sa podrškom tela od 20%, a pouzdanost pravila 95%. Dobili smo jedno pravilo:

Consequent	Antecedent	Suppo...	Confid...	Rule S...	Lift
DM Score Category_Great	Interceptions Category_Great Long passing Category_Great	20,138	97,752	19,686	2,686

SLIKA 23: APRIORI, ZADNJI VEZNI

Dakle, ako igrač odlično čita igru i presreće pasove i ima odlične duge pasove, onda je on 98% odličan zadnji vezni.

Na ovaj skup nismo primenili Carma algoritam, jer on radi na drugačiji način i ne prihvata ulogu koju treba da ima neko polje.

#### 4.

Proverili smo kako utiču godine, cena igrača i njegova ukupna ocena jedno na drugo. Koristili smo podrazumevani Apriori:

Consequent	Antecedent	Suppo...	Confid...	Rule S...	Lift
Overall Category_Fine	Age Category_Very young Value in Euro Category_Normal	16,098	93,919	15,119	3,274
Overall Category_Great	Value in Euro Category_Cheaper Age Category_In his prime	15,063	99,437	14,978	2,703
Overall Category_Good	Age Category_Young Value in Euro Category_Expensive	11,022	93,121	10,264	2,697
Value in Euro Category_Normal	Overall Category_Fine	28,688	83,097	23,839	2,49
Value in Euro Category_Expensive	Age Category_Young Overall Category_Good	12,375	82,945	10,264	2,486
Value in Euro Category_Cheaper	Age Category_Young Overall Category_Great	12,137	81,026	9,834	2,437
Overall Category_Great	Value in Euro Category_Cheaper	33,254	86,098	28,631	2,341
Overall Category_Great	Age Category_Young Value in Euro Category_Cheaper	11,769	83,558	9,834	2,272

SLIKA 24: PODRAZUMEVANI APRIORI

Vidimo da veoma mladi igrači koji imaju srednju cenu, čak 94% njih je lošije ocenjeno. Međutim, skoro svi koji su jeftini igrači, ali su malo stariji, su odlični igrači. Mladi igrači koji su najskuplji, u čak 93 posto su i dobri igrači. A oko 28% celog skupa su igrači koji su najjeftiniji, a odlični su igrači.

Sa podrazumevanim Carma algoritmom, ali sa menjanjem podrške na početku traženja čestih skupova:

Consequent	Antecedent	Suppor...	Confid...	Rule S...	Lift
Value in Euro Category_Normal	Overall Category_Fine	28,688	83,097	23,839	2,49
Overall Category_Fine	Value in Euro Category_Normal	33,379	71,419	23,839	2,49
Value in Euro Category_Cheaper	Overall Category_Great	36,785	77,834	28,631	2,341
Overall Category_Great	Value in Euro Category_Cheaper	33,254	86,098	28,631	2,341
Overall Category_Good	Value in Euro Category_Expensive	33,367	77,938	26,006	2,257
Value in Euro Category_Expensive	Overall Category_Good	34,527	75,32	26,006	2,257

SLIKA 25: CARMA SA MENJANJEM PODRŠKE

Vidimo da je nešto manje od trećine skupa dobri igrači koji su skupi, a nešto manje su loši igrači koji imaju srednju cenu.

## 5.

Na kraju smo pokušali da vidimo da li ukupna ocena i pozicija utiču na Ligu. Za podršku tela je postavljeno 5%, a 10 za pouzdanost pravila. Dobijeno je 70 pravila, ali nisu toliko zanimljiva, osim ovih na dnu:

Consequent	Antecedent	Support...	Confide...	Rule S...	Lift
League = Ligue 1	Position = CB	15,042	20,188	3,037	0,958
League = Premier League	Position = RB	6,497	21,739	1,412	0,95
League = Ligue 1	Position = RM	6,956	19,797	1,377	0,939
League = La Liga	Position = ST	11,723	17,47	2,048	0,935
League = Bundesliga	Overall Cateogry = Good	33,51	17,387	5,826	0,933
League = La Liga	Position = CDM	7,98	17,257	1,377	0,924
League = Serie A	Overall Cateogry = Fine	30,367	17,209	5,226	0,92
League = La Liga	Position = CAM	5,968	17,16	1,024	0,919
League = Bundesliga	Position = ST	11,723	16,867	1,977	0,905
League = Bundesliga	Position = LB	7,768	16,818	1,306	0,902
League = Ligue 1	Position = LM	7,698	18,807	1,448	0,892
League = La Liga	Position = CB	15,042	15,728	2,366	0,842
League = Premier League	Overall Cateogry = Good	33,51	18,967	6,356	0,829
League = Serie A	Position = CAM	5,968	15,385	0,918	0,822
League = Premier League	Position = RM	6,956	18,782	1,306	0,821
League = Serie A	Position = LB	7,768	14,545	1,13	0,777
League = Bundesliga	Position = CM	10,664	14,238	1,518	0,764
League = Serie A	Position = LM	7,698	14,22	1,095	0,76
League = Premier League	Position = CDM	7,98	17,257	1,377	0,754
League = La Liga	Overall Cateogry = Fine	30,367	13,953	4,237	0,747
League = Ligue 1	Position = CB Overall Cateogry = Great	6,462	14,754	0,953	0,7
League = Serie A	Position = RM	6,956	12,183	0,847	0,651
League = Ligue 1	Overall Cateogry = Great	36,123	12,219	4,414	0,58

SLIKA 26: APRIORI 5, 10

Poslednje pravilo nam kaže da ako je igrač odličan, skoro 2 puta je manje verovatno da je iz Lige 1. Slično, malo manje od toga je ako je igrač desni vezni, da igra u Seriji A.

Carma algoritmom, sa 1% podrška pravila i 10% pouzdanost pravila, dobijemo dva relativno zanimljivija pravila:

Consequent	Antecedent	Suppor...	Confide...	Rule S...	Lift
Overall Cateogry_Fine	League_Ligue 1 Position_GK	2,507	53,521	1,342	1,762
Overall Cateogry_Good	League_Ligue 1 Position_CM	2,154	54,098	1,165	1,614

SLIKA 27: CARMA 1, 10

Pola golmana iz Lige 1 su loši, a isto pola centralnih veznih iz Lige 1 su dobri.

## **Zaključak**

Ovaj skup daje mnoge mogućnosti za igranje, ali to premašuje moje trenutne mogućnosti. Deo vremena posvećen ovom radu bio je zamenjen i igranjem FIFE, 19, doduše.

U ovom radu, malo korisnije rezultate nam je dao Carma algoritam, iako bih se ja pristrasno opredelila za Apriori, koji je nekako intuitivniji algoritam. Nekoliko urađenih pravila pridruživanja je izbačeno iz rada, pa bih Vam za kraj preporučila da ako šutirate slobodnjake, to nikako ne bude sa golmanima.