

ATP 2000-2017

Seminarski rad u okviru kursa
Istraživanje podataka
Matematički fakultet

Marija Mijailović
mi14199@alas.matf.bg.ac.rs
Miroslav Mišljenović
mr12260@alas.matf.bg.ac.rs

jun 2018.

Sažetak

U ovom radu analizirali smo skup podataka "ATP - rezultati turnira od 2000-2017". Obradili smo pravila pridruživanja, klasterovanje, klasifikaciju i predstavili sve navedene metode odgovarajućom vizualizacijom. Skup podataka je preuzet sa <https://www.kaggle.com/gmadevs/atp-matches-dataset>.

Sadržaj

| | | |
|----------|------------------------------|----------|
| 1 | Uvod | 1 |
| 2 | Analiza podataka | 1 |
| 3 | Pravila pridruživanja | 3 |
| 4 | Klasterovanje | 5 |
| 5 | Klasifikacija | 5 |

1 Uvod

Skup podataka ATP mečeva podeljen je u 17 zasebnih .csv fajlova i svaki od njih prikazuje individualne statistike za svaki turnir u toku te godine.

2 Analiza podataka

U ovom poglavlju sledi kratak pregled najistaknutijih atributa ovog skupa podataka. Svaki red u skupu, označava jedan meč i sve informacije o tom meču.

U tabeli [1](#) prikazani su podaci o turniru.

U tabeli [2](#) prikazani su podaci o pobedniku meča.

U tabeli [3](#) prikazani su podaci o gubitniku meča.

| Ime kolone | Objašnjenje |
|---------------|--|
| tourney_id | id turnira |
| tourney_name | ime turnira |
| surface | podloga(Grass, Clay, Hard) |
| tourney_level | nivo turnira(Grand Slam, Finals, Masters, Tour Series, Challenger) |
| round | runda(Round of 16, Quarterfinal...) |
| minutes | trajanje meča u minutima |

Tabela 1: Podaci o turnirima

| Ime kolone | Objašnjenje |
|--------------------|---|
| winner_seed | nosilac na turniru |
| winner_entry | ulaznica(WildCard, Qualified, LuckyLoser, ProtectedRanking) |
| winner_name | ime pobjednika |
| winner_ht | visina pobjednika |
| winner_ioc | zemlja porijekla pobjednika |
| winner_age | godine pobjednika |
| winner_rank | ATP rang pobjednika |
| winner_rank_points | ATP poeni pobjednika |
| w_ace | broj asova pobjednika |
| w_df | broj duplih grešaka pobjednika |
| w_svpt | broj poena dobijenih na servis pobjednika |
| w_1stIn | broj ubačenih prvih servisa pobjednika |
| w_1stWon | broj poena dobijenih nakon ubačenog prvog servisa pobjednika |
| w_2ndWon | broj poena dobijenih nakon ubačenog drugog servisa pobjednika |
| w_SvGms | broj gemova u kojima je servirao pobjednik |
| w_bpSaved | broj spašenih brejk lopti pobjednika |
| w_bpFaced | broj izgubljenih gemova posle brejka pobjednika |

Tabela 2: Podaci o pobjednicima

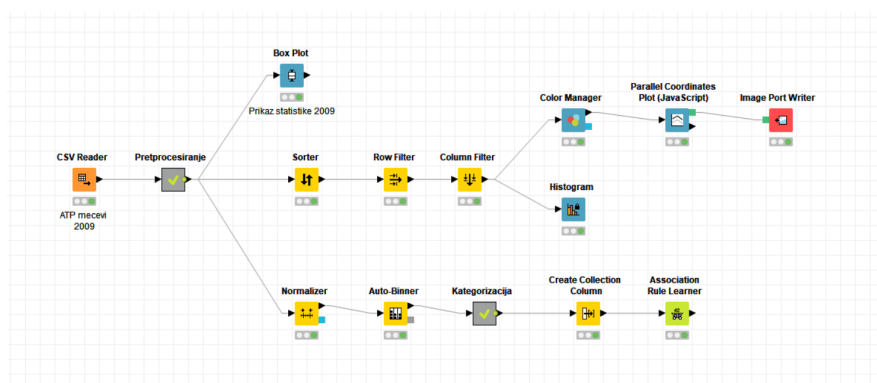
| Ime kolone | Objašnjenje |
|-------------------|--|
| loser_seed | nosilac na turniru |
| loser_entry | ulaznica(WildCard, Qualified, LuckyLoser, ProtectedRanking) |
| loser_name | ime gubitnika |
| loser_ht | visina gubitnika |
| loser_ioc | zemlja porijekla gubitnika |
| loser_age | godine gubitnika |
| loser_rank | ATP rang gubitnika |
| loser_rank_points | ATP poeni gubitnika |
| l_ace | broj asova gubitnika |
| l_df | broj duplih grešaka gubitnika |
| l_svpt | broj poena dobijenih na servis gubitnika |
| l_1stIn | broj ubačenih prvih servisa gubitnika |
| l_1stWon | broj poena dobijenih nakon ubačenog prvog servisa gubitnika |
| l_2ndWon | broj poena dobijenih nakon ubačenog drugog servisa gubitnika |
| l_SvGms | broj gemova u kojima je servirao gubitnik |
| l_bpSaved | broj spašenih brejk lopti gubitnika |
| l_bpFaced | broj izgubljenih gemova posle brejka gubitnika |

Tabela 3: Podaci o gubitnicima

S obzirom na veliki broj raspoloživih godina, prvo smo se detaljno upoznali sa podacima i šta nam koja godina pruža i koji su najzanimljiviji atributi za svaku godinu. U zavisnosti od toga smo, po potrebama metoda, koristili različite godine, ali svuda smo se ograničili na četiri maksimalno.

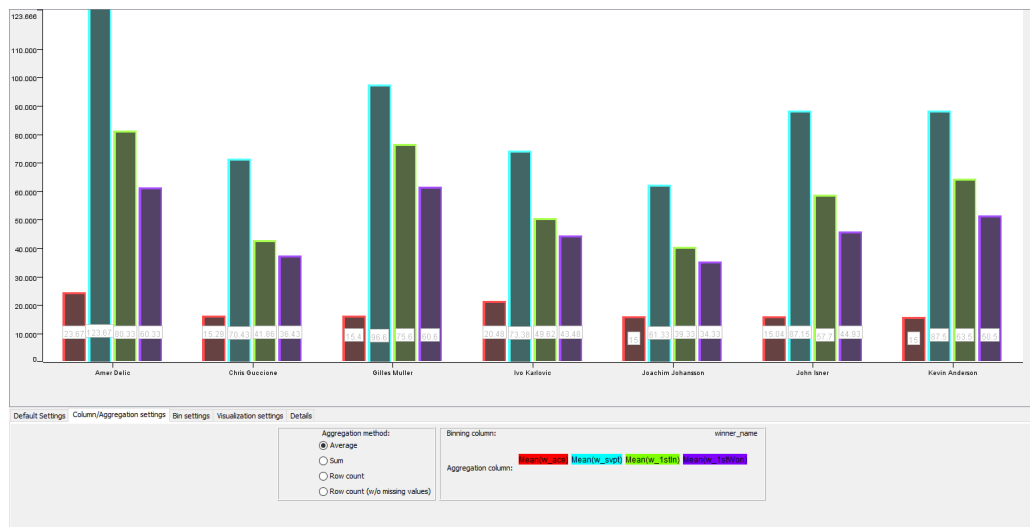
3 Pravila pridruživanja

Pravila pridruživanja smo obradili u programskom alatu KNIME (slika 1). Odlučili smo se za 2009. godinu, jer su rezultati reprezentativniji u odnosu na ostale godine.

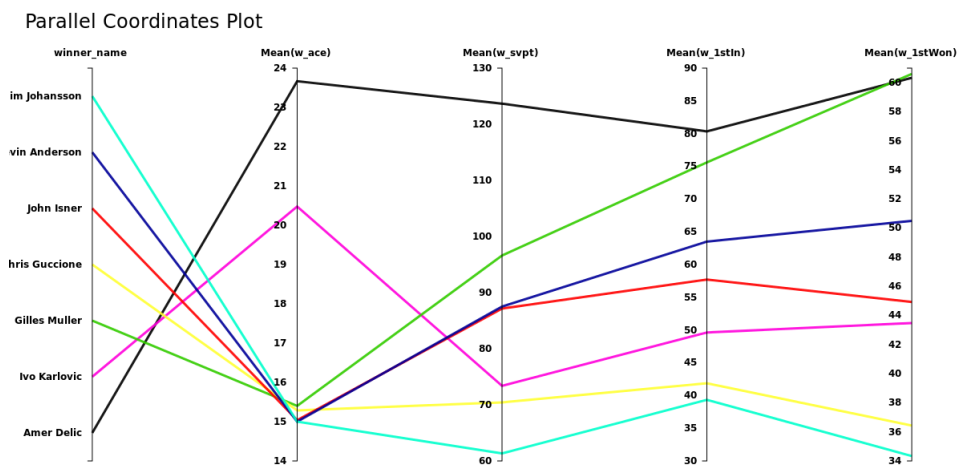


Slika 1: KNIME implementacija

Na slikama 2 i 3 grafički su prikazani rezultati za sedam tenisera koji su imali prosečno najviše asova po meču na kome su pobedili. Izabrali smo četiri parametra za svakog igrača: broj asova porednika, broj dobijenih poena na servis porednika, procenat ubačenog prvog servisa porednika i broj osvojenih poena nakon ubačenog prvog servisa porednika. Na histogramu i grafiku paralelnih koordinata mogu se videti i uporediti rezultati.



Slika 2: Histogram



Slika 3: Paralelne koordinate

Iznenadenje je pojavljivanje Amera Delića u prvih sedam, jer je to autorima nepoznat igrač. Uvidom u podatke, utvrđeno je da je on te godine odigrao samo osam mečeva, a pobedio je samo tri puta (što je kriterijum po kome je birano najboljih sedam).

U tri kategorije smo podelili sledeća četiri atributa: broj asova pobednika, broj duplih servis grešaka pobednika, broj osvojenih poena nakon ubačenog prvog servisa pobednika, broj spašenih brejk lopti pobednika. Na slici 4 se mogu videti pravila pridruživanja dobijena na osnovu te kategorizacije, sortirani po Lift meri. Za pouzdanost smo uzeli vrednost 0.4, a za minimalnu podršku vrednost 0.15. Analizirali smo podatke za

sve godine i rezultati su prilično uniformni. Za 2009. godinu je dobijena druga najveća Lift mera (1.36) i odnosi se na pravilo [ACE 2, WON 2, DF 1] -> [BPS 1]. U 2004. godini smo dobili najveću vrednost Lift mere (1.441) za pravilo [BPS 1, ACE 1, DF 1] -> [WON 1].

▲ Frequent itemsets/Association rules - 2:3 - Association Rule Learner

File Hilite Navigation View

Table "default" - Rows: 26 Spec - Columns: 6 Properties Flow Variables

| Row ID | D Support | D Confide... | D ▼ Lift | S Conseq... | S implies | (...) Items |
|--------|-----------|--------------|----------|-------------|-----------|---------------------|
| rule2 | 0.177 | 0.837 | 1.36 | BPS 1 | <--- | [ACE 2,WON 2,DF 1] |
| rule4 | 0.197 | 0.816 | 1.326 | BPS 1 | <--- | [ACE 2,DF 1] |
| rule6 | 0.207 | 0.857 | 1.279 | ACE 1 | <--- | [BPS 2,WON 2,DF 1] |
| rule10 | 0.217 | 0.83 | 1.239 | ACE 1 | <--- | [BPS 2,DF 1] |
| rule7 | 0.207 | 0.955 | 1.174 | WON 2 | <--- | [BPS 2,ACE 1,DF 1] |
| rule13 | 0.241 | 0.925 | 1.137 | WON 2 | <--- | [BPS 2,DF 1] |
| rule14 | 0.246 | 0.909 | 1.118 | WON 2 | <--- | [BPS 2,ACE 1] |
| rule0 | 0.177 | 0.9 | 1.107 | WON 2 | <--- | [BPS 1,ACE 2,DF 1] |
| rule1 | 0.177 | 0.878 | 1.1 | DF 1 | <--- | [BPS 1,ACE 2,WON 2] |
| rule17 | 0.325 | 0.868 | 1.088 | DF 1 | <--- | [BPS 1,ACE 1] |
| rule18 | 0.419 | 0.867 | 1.087 | DF 1 | <--- | [BPS 1,WON 2] |
| rule21 | 0.532 | 0.864 | 1.083 | DF 1 | <--- | [BPS 1] |
| rule9 | 0.212 | 0.878 | 1.08 | WON 2 | <--- | [ACE 2,DF 1] |
| rule16 | 0.31 | 0.875 | 1.077 | WON 2 | <--- | [BPS 2] |
| rule12 | 0.236 | 0.857 | 1.074 | DF 1 | <--- | [BPS 1,ACE 1,WON 2] |
| rule5 | 0.202 | 0.872 | 1.073 | WON 2 | <--- | [BPS 1,ACE 2] |
| rule3 | 0.197 | 0.851 | 1.066 | DF 1 | <--- | [BPS 1,ACE 2] |
| rule8 | 0.207 | 0.84 | 1.053 | DF 1 | <--- | [BPS 2,ACE 1,WON 2] |
| rule20 | 0.448 | 0.827 | 1.037 | DF 1 | <--- | [ACE 1,WON 2] |
| rule24 | 0.665 | 0.833 | 1.025 | WON 2 | <--- | [DF 1] |
| rule25 | 0.665 | 0.818 | 1.025 | DF 1 | <--- | [WON 2] |
| rule23 | 0.547 | 0.816 | 1.023 | DF 1 | <--- | [ACE 1] |
| rule15 | 0.266 | 0.831 | 1.022 | WON 2 | <--- | [ACE 2] |
| rule19 | 0.448 | 0.82 | 1.009 | WON 2 | <--- | [ACE 1,DF 1] |
| rule11 | 0.217 | 0.8 | 1.002 | DF 1 | <--- | [BPS 2,ACE 1] |
| rule22 | 0.542 | 0.809 | 0.995 | WON 2 | <--- | [ACE 1] |

Slika 4: Pravila pridruživanja

4 Klasterovanje

5 Klasifikacija