



**19. konferenca
Dnevi slovenske informatike**

**ANALIZA ŠAHOVSKIH PARTIJ
S PROGRAMOMA ZA
RUDARJENJE PO PODATKIH
ORANGE CANVAS IN
SPECTOR**

*Vid Ogris,
Vladislav Rajkovič,
Davorin Kofjač*

16. 04. 2012



ANALIZA ŠAHOVSKIH PARTIJ S PROGRAMOMA ZA RUDARJENJE PO PODATKIH ORANGE CANVAS IN SPECTOR

Raziskava:

- Analiza potez šahistov
- Uporaba programov Orange Canvas in Spector



ANALIZA ŠAHOVSKIH PARTIJ S PROGRAMOMA ZA RUDARJENJE PO PODATKIH ORANGE CANVAS IN SPECTOR

Namen:

- Povezave med potezami Belega in Črnega
- Poiskati najprimernejšo statistično metodo za napoved zmagovalca



Podatki

- Baza podatkov z 1,74 milijona zapisi šahovskih partij (<http://www.top-5000.nl/pgn.htm>)
- PGN vrsta zapisa – algebraična notacija
- Naključno izbranih 5140 partij za obdelavo s programom Orange
- Celotna baza podatkov za obdelavo s programom Spector



Podatki

- 1. poteza Belega (1. white),
- 1. poteza Črnega (1. black),
- 2. poteza Belega (2. white),
- Ali je bil po tej potezi Belega šah? (DA/NE) (CHECK2),
- 2. poteza Črnega (2. black),
- Ali je bil po tej potezi Črnega šah? (DA/NE) (CHECK3),
- Končni zmagovalec partije (Winner).

1.white	1.black	2.white	CHECK2	2.black	CHECK3	Winner
d4	Nf6	c4	NO	g6	NO	Black
d4	Nf6	c4	NO	e6	NO	Draw
c4	Nf6	Nc3	NO	c5	NO	White
e4	g6	d4	NO	Bg7	NO	Black
e4	Nf6	e5	NO	Nd5	NO	White



Orange Canvas



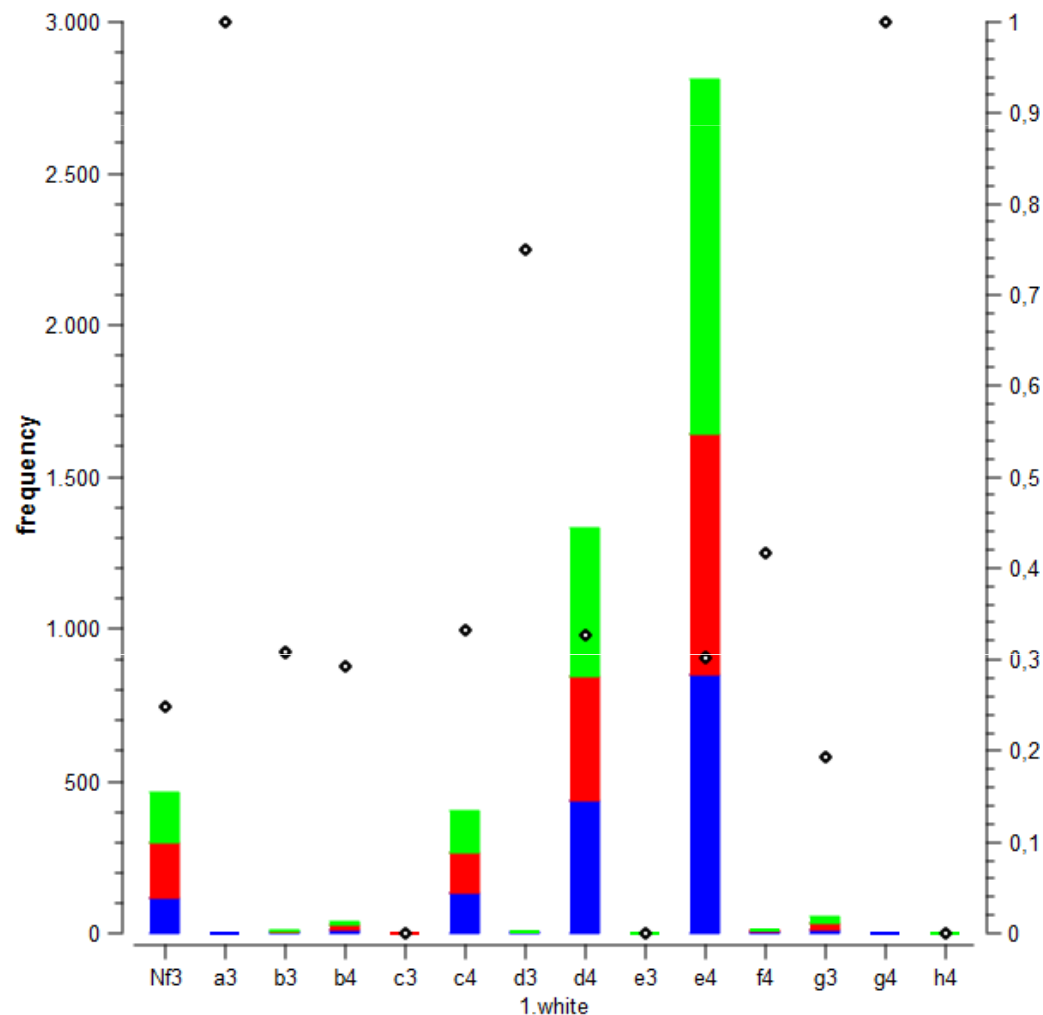
Analize

- 2010 - krat zmagovalec Beli
- 1571 - krat zmagovalec Črni
- 1559 - krat neodločen rezultat



Korelacije

- Beli na e4 - zmagovalec Beli v 41%, v 30% zmagovalec Črni
- Beli na d4 – zmagovalec Beli v 36%, v 32% zmagovalec Črni
- Beli na a3 oz. g4 – zmagovalec Črni

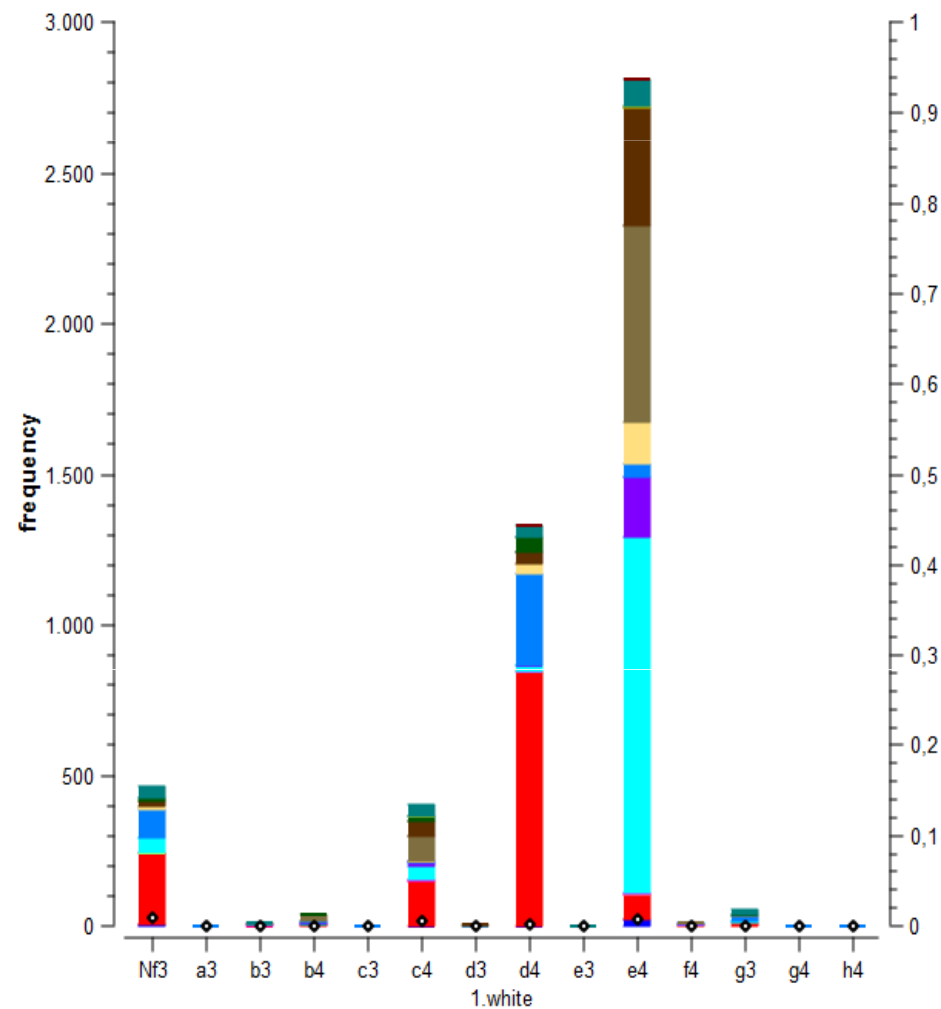




Korelacije

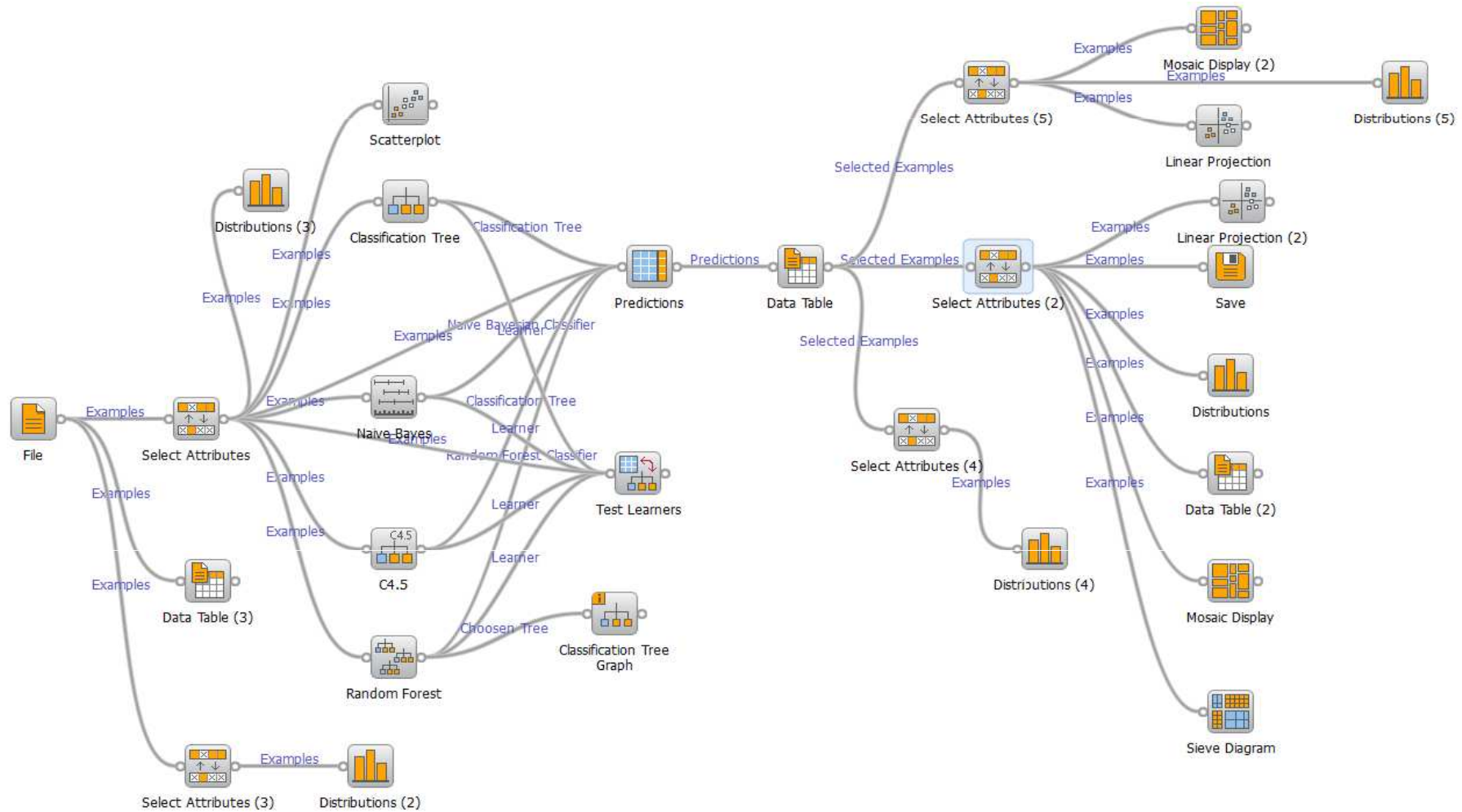
Odločitev Črnega glede na prvo potezo Belega

- Beli na e4, Črni na c5 v 42%
- Beli na e4, Črni na e5 v 23%
- Beli na e4, Črni na e6 v 13%
- Beli na d4, Črni na Nf6 v 63%





Proces analize v programu Orange Canvas





Statistične metode

- Klasifikacijsko drevo
- Naivni Bayes
- C4.5
- Naključni gozd
- Parameter Test Learners (k = 10)

	<i>Točnost klasifikacije</i>	<i>Občutljivost</i>	<i>Specifičnost</i>	<i>Površina pod ROC krivuljo</i>
Klasifikacijsko drevo	0.3951	0.0433	0.9574	0.5340
Naivni Bayes	0.3794	0.1795	0.8403	0.5439
C4.5	0.3963	0.0789	0.9375	0.5330
Naključni gozd	0.3963	0.1050	0.9137	0.5430



Predvidevanja

1. <i>white</i>	2. <i>white</i>	1. <i>black</i>	CHECK2	2. <i>black</i>	CHECK3	Winner	Naivni Bayes	C4.5	Klasifikacijsko drevo	Naključni gozd
<i>d4</i>	<i>c4</i>	<i>Nf6</i>	<i>NO</i>	<i>g6</i>	<i>NO</i>	<i>Black</i>	<i>Black</i>	<i>White</i>	<i>White</i>	<i>White</i>
<i>d4</i>	<i>c4</i>	<i>Nf6</i>	<i>NO</i>	<i>e6</i>	<i>NO</i>	<i>Draw</i>	<i>Draw</i>	<i>Draw</i>	<i>Draw</i>	<i>White</i>
<i>c4</i>	<i>Nc3</i>	<i>Nf6</i>	<i>NO</i>	<i>c5</i>	<i>NO</i>	<i>White</i>	<i>Black</i>	<i>White</i>	<i>White</i>	<i>White</i>
<i>e4</i>	<i>d4</i>	<i>g6</i>	<i>NO</i>	<i>Bg7</i>	<i>NO</i>	<i>Black</i>	<i>White</i>	<i>White</i>	<i>White</i>	<i>White</i>
<i>e4</i>	<i>e5</i>	<i>Nf6</i>	<i>NO</i>	<i>Nd5</i>	<i>NO</i>	<i>White</i>	<i>White</i>	<i>White</i>	<i>White</i>	<i>White</i>
<i>e4</i>	<i>d4</i>	<i>g6</i>	<i>NO</i>	<i>Bg7</i>	<i>NO</i>	<i>Draw</i>	<i>White</i>	<i>White</i>	<i>White</i>	<i>White</i>
<i>e4</i>	<i>d4</i>	<i>c6</i>	<i>NO</i>	<i>d5</i>	<i>NO</i>	<i>Draw</i>	<i>White</i>	<i>White</i>	<i>White</i>	<i>White</i>
<i>e4</i>	<i>d4</i>	<i>c6</i>	<i>NO</i>	<i>d5</i>	<i>NO</i>	<i>Black</i>	<i>White</i>	<i>White</i>	<i>White</i>	<i>White</i>
<i>e4</i>	<i>d4</i>	<i>c6</i>	<i>NO</i>	<i>d5</i>	<i>NO</i>	<i>Draw</i>	<i>White</i>	<i>White</i>	<i>White</i>	<i>White</i>
<i>e4</i>	<i>d3</i>	<i>e6</i>	<i>NO</i>	<i>Nf6</i>	<i>NO</i>	<i>Black</i>	<i>Black</i>	<i>Draw</i>	<i>White</i>	<i>Black</i>
<i>d4</i>	<i>Nf3</i>	<i>e6</i>	<i>NO</i>	<i>b6</i>	<i>NO</i>	<i>Draw</i>	<i>White</i>	<i>White</i>	<i>White</i>	<i>Draw</i>
<i>e4</i>	<i>d4</i>	<i>e6</i>	<i>NO</i>	<i>d5</i>	<i>NO</i>	<i>Black</i>	<i>White</i>	<i>White</i>	<i>White</i>	<i>White</i>



Rezultati glede na metode

- Favorizira se poteza Beli na e4 in d4 za zmago Belega
- Zmagovalec je v veliki večini Beli (več kot 70%)
- Velika odstopanja glede na dejanske rezultate



Spector



Spector

- Podatkovne datoteke v logične implementacije
- Poišče regularnosti in jih kvantitativno ovrednoti
- Analiza na vseh 1,7 milijona partijah
- "Če to in to, potem toliko bolj/manj verjetno tisto in tisto"
- Našel več 1000 Simpsonovih paradoksov



Spector rezultati

- Prva poteza Črnega ni nikoli ni enaka drugi potezi Črnega
- Prva poteza Črnega ni nikoli enaka prvi potezi Belega
- Po potezi Belega na e4 se poveča možnost zmage Črnega
- Beli na a4 in Črni na e5 → zmagovalec Črni v 7-ih primerih, 3-krat pa to ne drži → s 70% gotovostjo trdimo da bo v primeru da je Beli naredil potezo na a4 in Črni na e5 zmagal Črni, da pa je to zgolj naključje pa je verjetnost 0,0009
- Prva poteza Belega Nf3 → v 37% izid neodločen (naključje: $1,25 * 10^{-401}$)



Specter

Specter_demo - Sah

Specterize

Filter Auto link (off) Rows: 10000/*

	Rank	_1white	_1black	_2white	CHECK2	_2black	CHECK3	Winner
1		d4	Nf6	c4	NO	g6	NO	Black
2		d4	Nf6	c4	NO	e6	NO	Draw

Filter Auto link (off) Rows: 83

	Rank	Implicator	Implication	Implicant	Confirmations	Exceptions	Probability	Prior	*Improbability*	Regularity
1				1black != _2black	1736158	0	1.0		E-9999*0.000000	regularity
2				_1black != _2white	1736158	0	1.0		E-9999*0.000000	regularity
3				_1white != _2white	1736158	0	1.0		E-9999*0.000000	regularity
4				_2white != _2black	1736158	0	1.0		E-9999*0.000000	regularity
5				_1white != _2black	1736158	0	1.0		E-0877*1.213391	regularity
6		2 _1white == 'Nf3'	-->	Winner == 'Draw'	54873	92737	0.37174	0.31905	E-0401*1.258677	regularity
7		_1white == 'Nf3' & _1black == 'Nf6'	-->	Winner == 'Draw'	27428	43288	0.38786	0.31905	E-0326*2.816625	regularity
8		1white == 'e4'	-->	Winner == 'Black'	259341	574527	0.31101	0.29398	E-0253*8.274191	regularity

Table Parent

Fa1_algit	
Fa1_algit_spec...	Fa1_algit
Sah	
Sah_spector	Sah
Sonce	

48	_1white == 'a3' & _1black == 'd5'	-->	Winner == 'Black'	72	97	0.42604	0.29398	E-0004*1.781824	regularity
49	1white == 'a4' & _1black == 'e5'	-->	Winner == 'Black'	7	3	0.7	0.29398	E-0003*9.386708	regularity
50	_1white == 'b4' & _1black == 'f5'	-->	Winner == 'White'	45	41	0.52326	0.38696	E-0003*7.050836	regularity



Zaključek

- Za predikcijsko najboljšo metodo pri uporabi programa Orange se izkaže Naključni gozd (Random forest)
- Premalo število vhodnih podatkov (Orange)
- Kompleksnost igre
- Dober vpogled v delovanje programov za rudarjenje po podatkih
- Poučevanje šaha, predvidevanje zmagovalca med samo igro



Igrajte Beli na ...



Hvala za vašo pozornost !

Vprašanja?

Pripombe?

Predlogi?