

Машинско учење

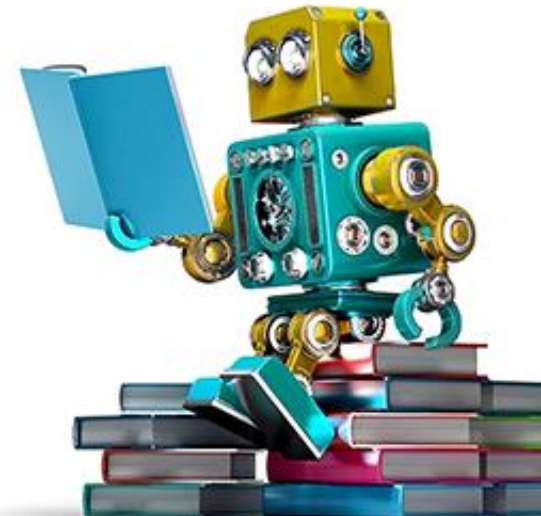
Преваре са кредитним картицама

студент: Милан Симовић 3169/2022
ментор: доц. др Предраг Тадић

Увод

- Бинарна класификација
- На располагању 1 000 000 примера
- Седам обележја – три континуална и четири бинарна

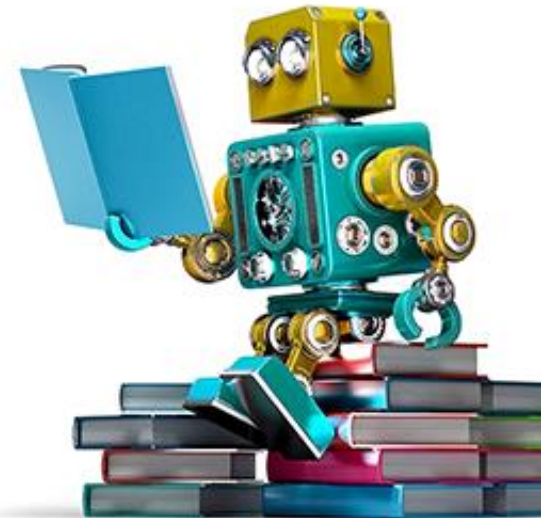
Обележја	
Ред. бр.	Опис
1.	Растојање места трансакције од куће власника картице
2.	Растојање места трансакције од места претходне трансакције
3.	Однос тренутне цене и медијане плаћања
4.	Куповина у истој продавници
5.	Коришћен чип
6.	Унет пин
7.	Онлајн наруџбина



Експлоративна анализа података

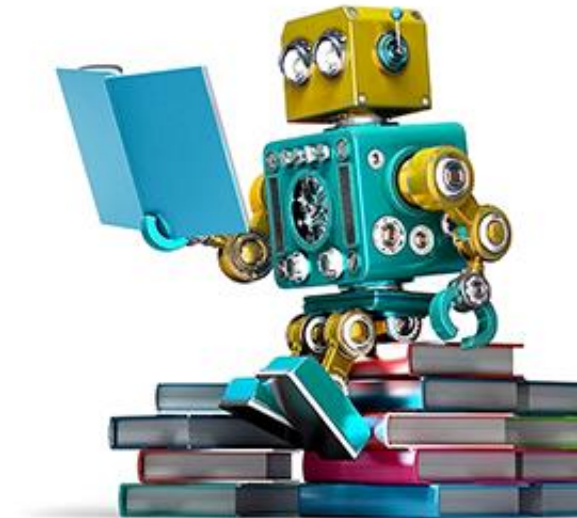
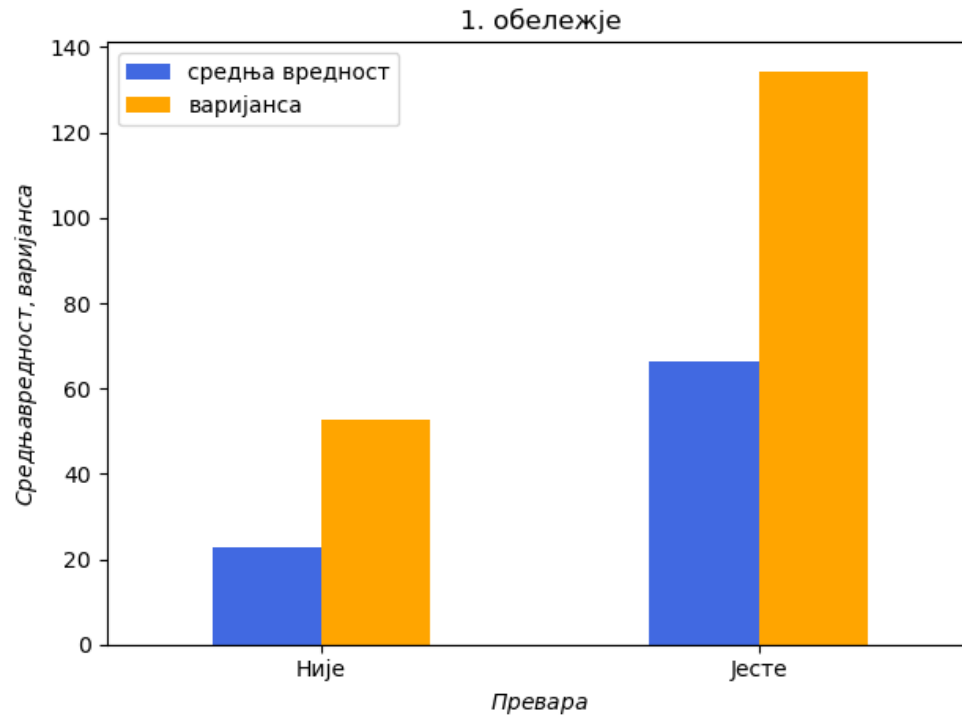
- Скуп података је неуравнотежен

Уравнотеженост података



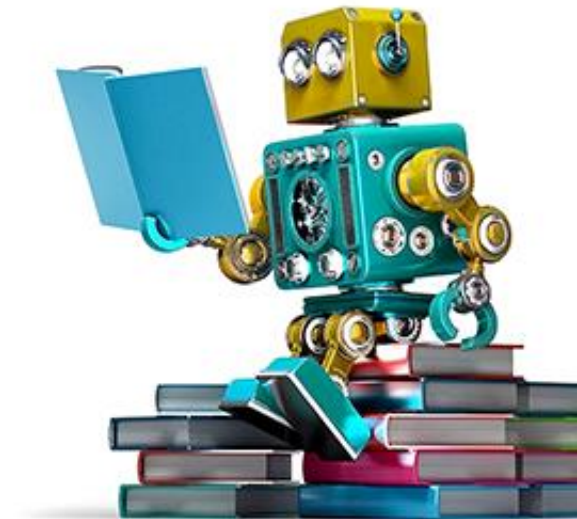
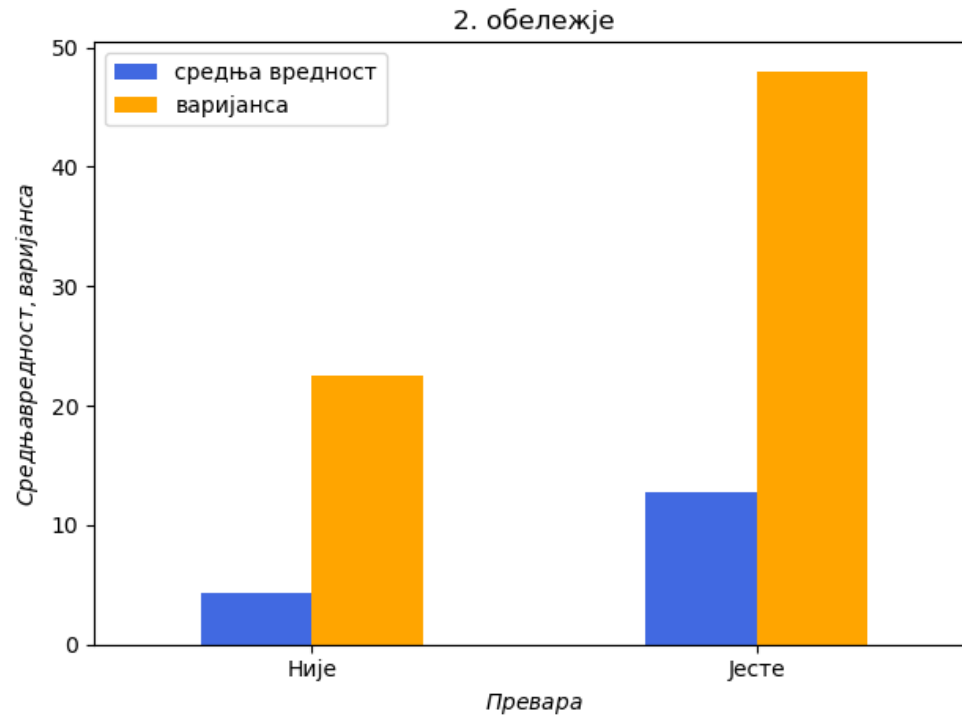
Експлоративна анализа података

Процена средње вредности и варијансе



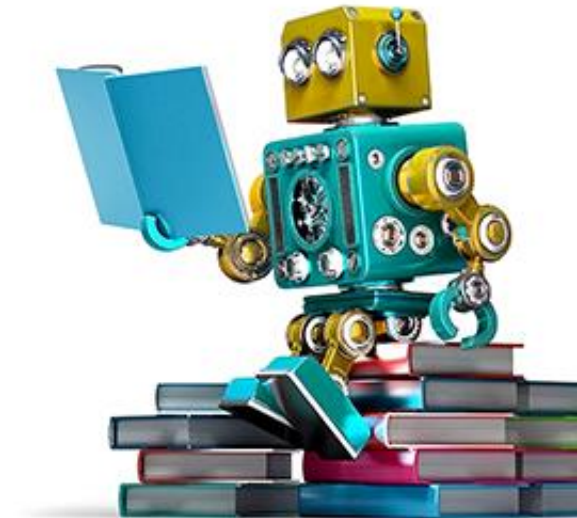
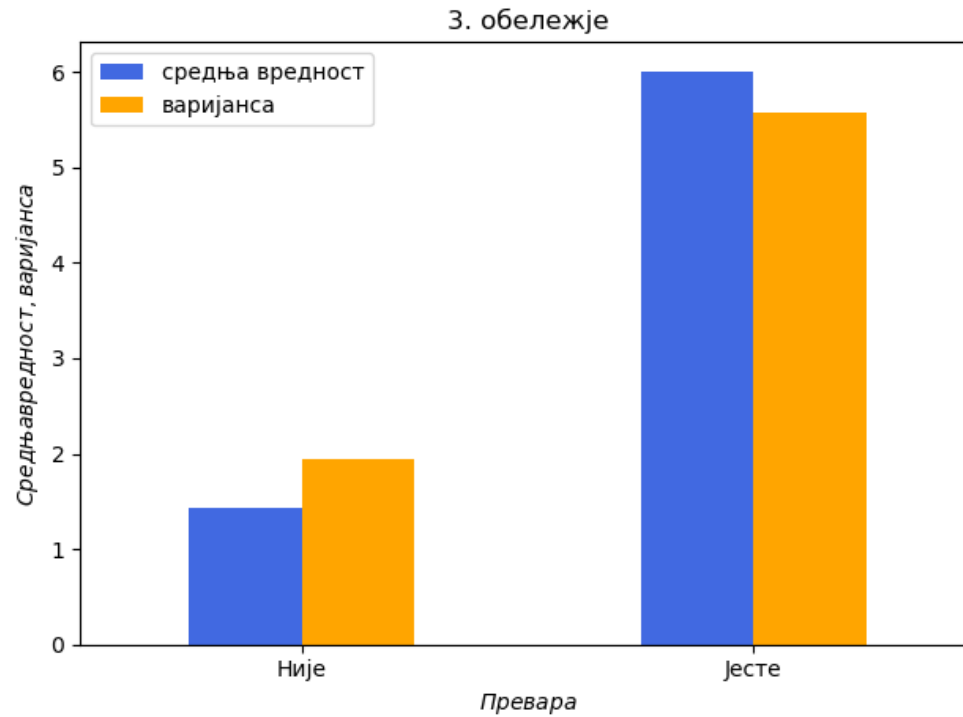
Експлоративна анализа података

Процена средње вредности и варијансе

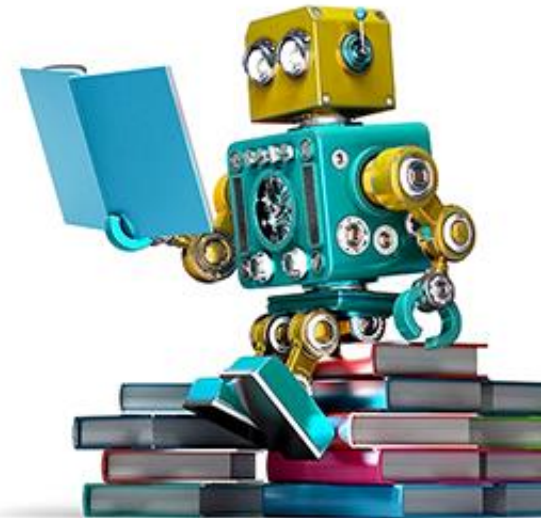
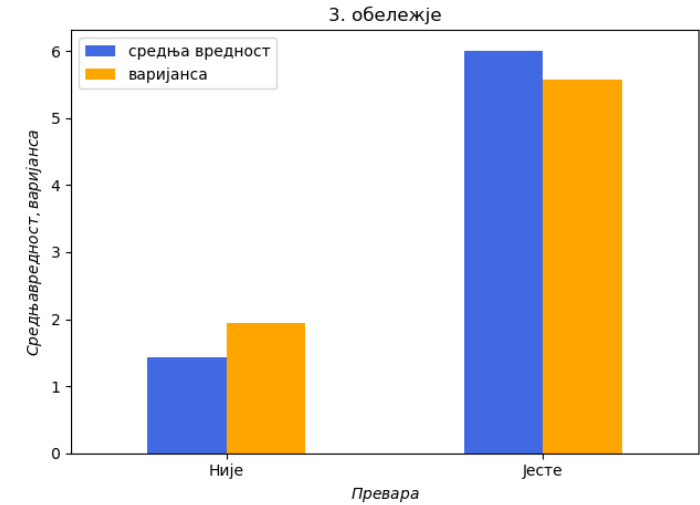
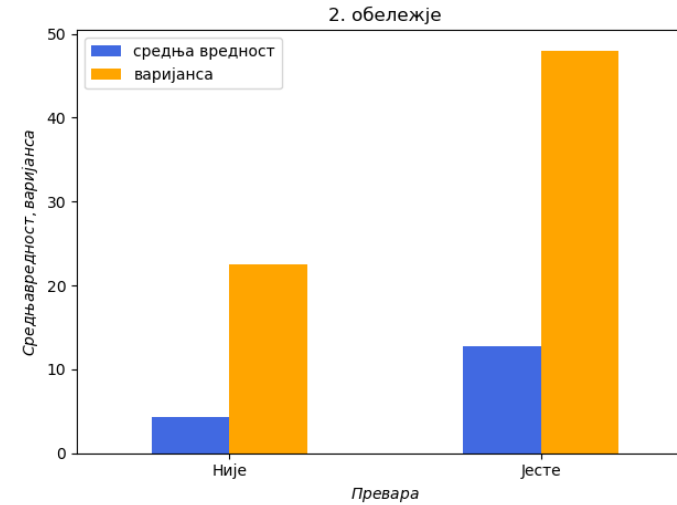
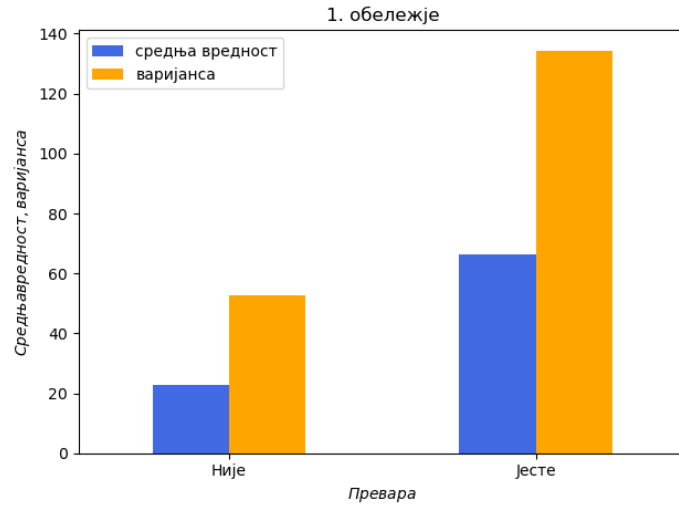


Експлоративна анализа података

Процена средње вредности и варијансе

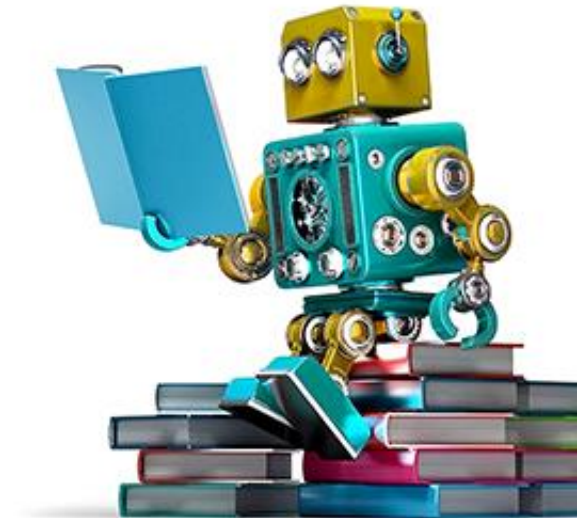
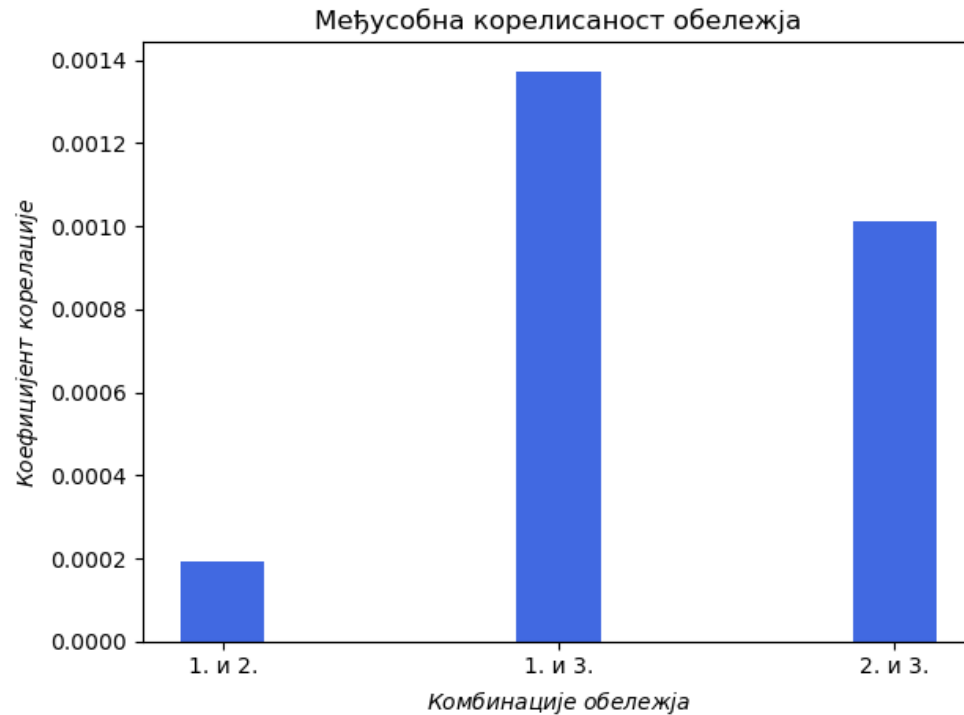


Експлоративна анализа података



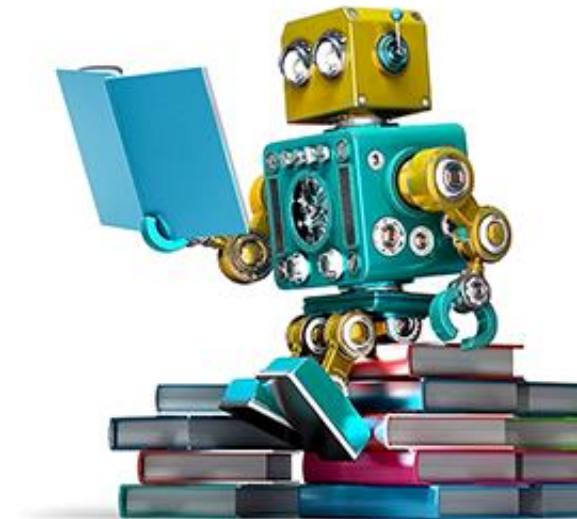
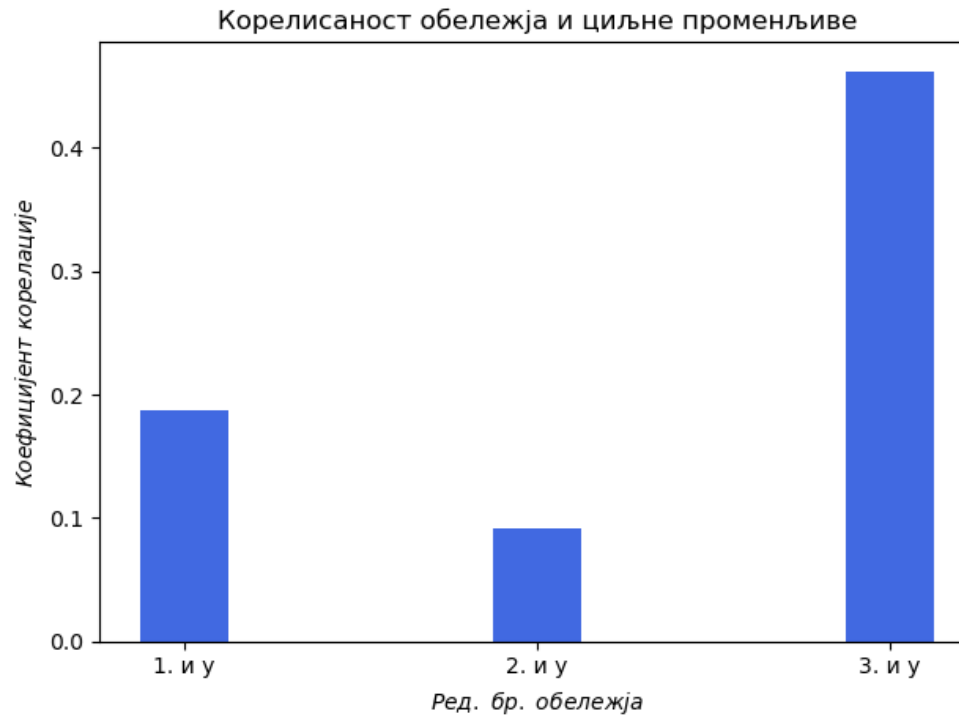
Експлоративна анализа података

Процена међусобне корелисаности обележја



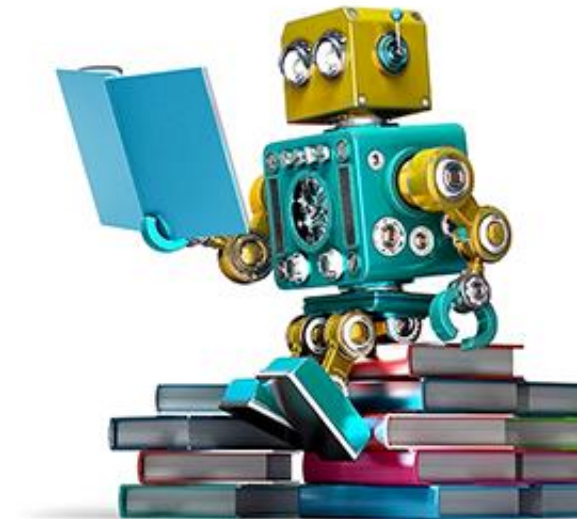
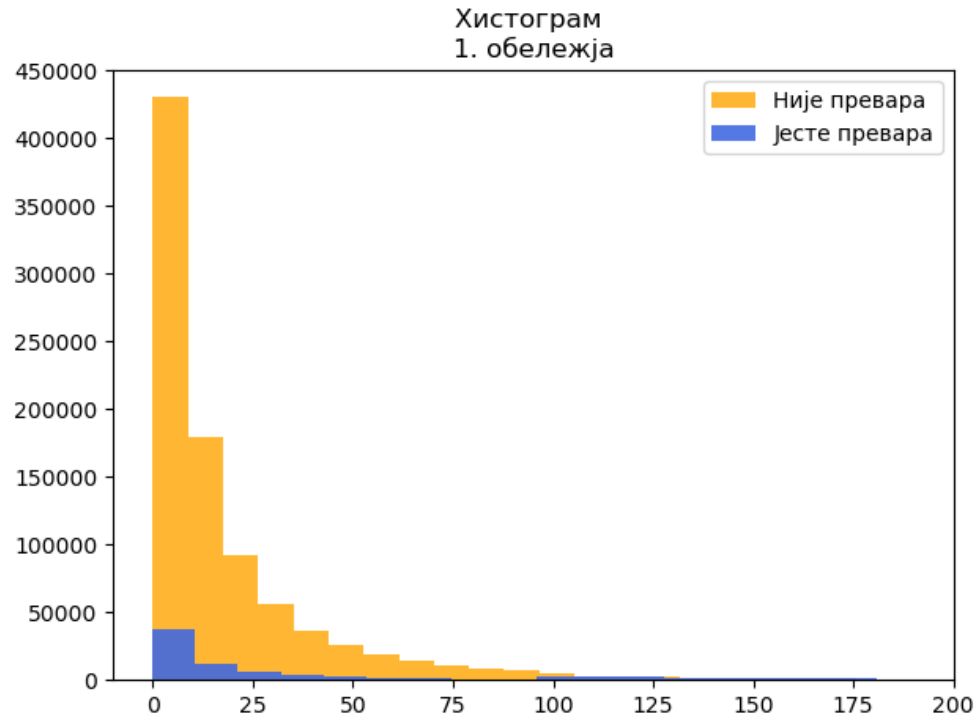
Експлоративна анализа података

Процена корелисаности обележја и циљне променљиве



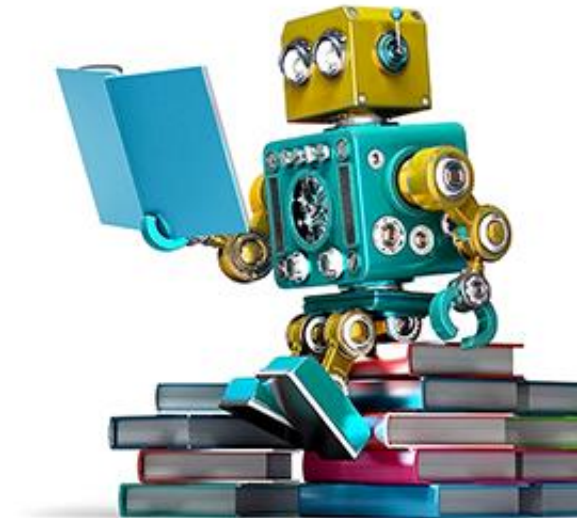
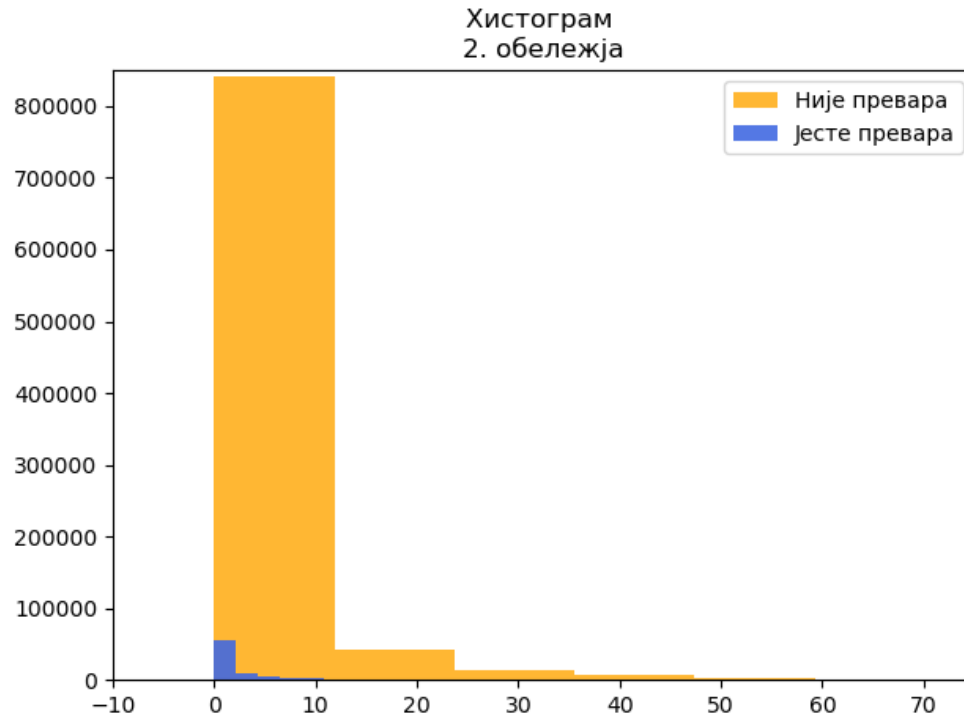
Експлоративна анализа података

Хистограми обележја за позитивне и негативне примере



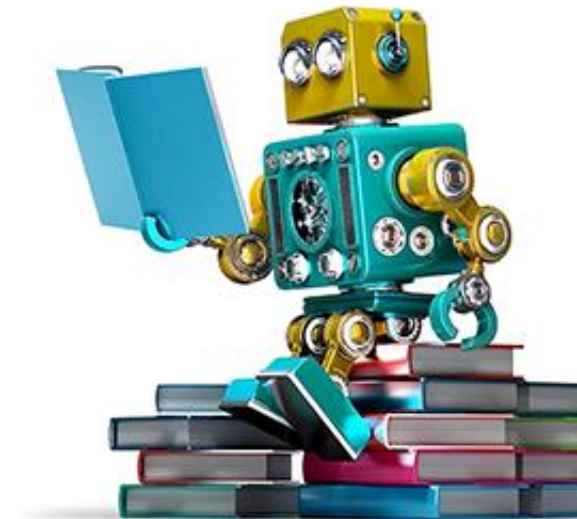
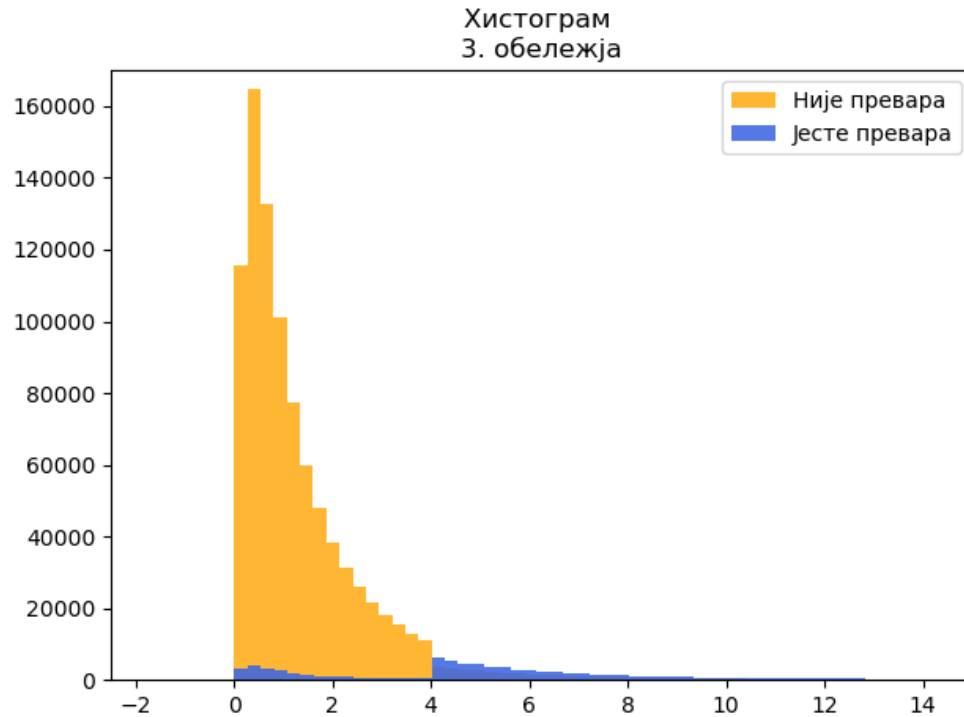
Експлоративна анализа података

Хистограми обележја за позитивне и негативне примере



Експлоративна анализа података

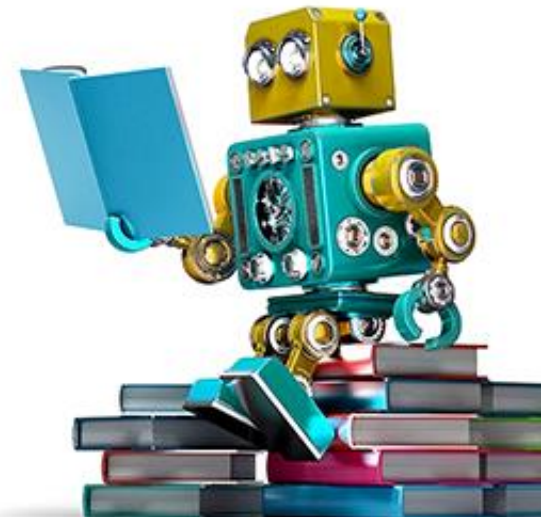
Хистограми обележја за позитивне и негативне примере



Класификација према бројности

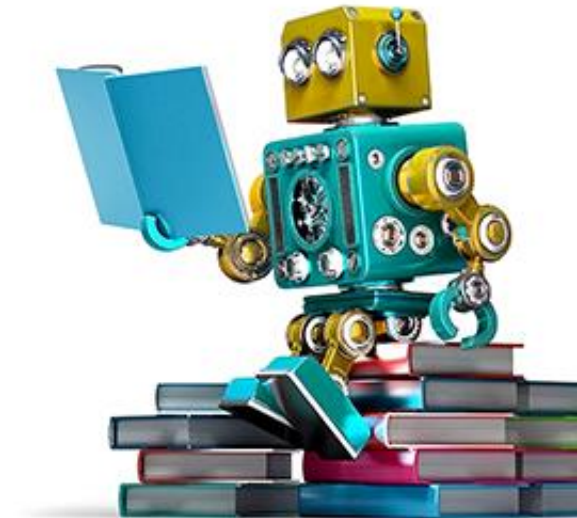
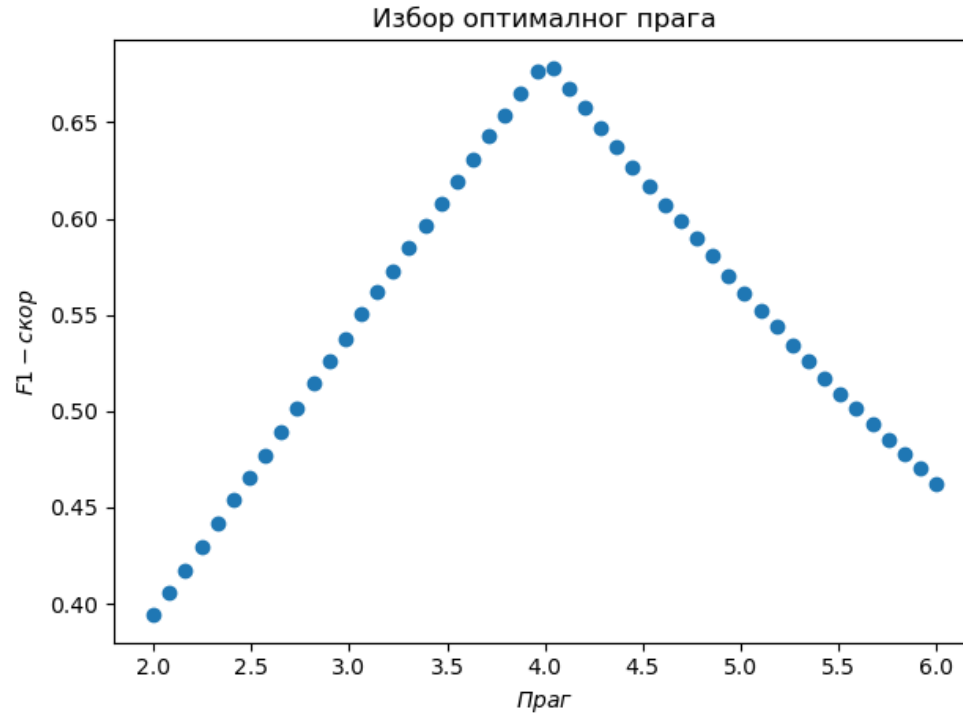
	\hat{N}	\hat{P}
N	166605	15914
P	15934	1547

Тачност	0,84
Прецизност	0,09
Осетљивост	0,09
F1-скор	0,09



Класификација на основу информативног обележја

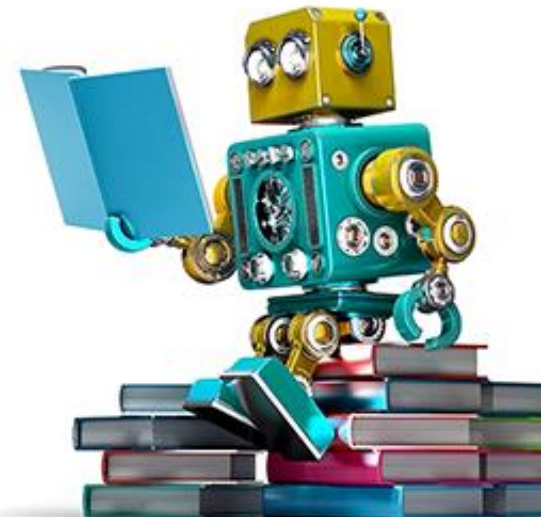
Процена оптималног прага за 3. обележје



Класификација на основу информативног обележја

	\hat{N}	\hat{P}
N	175008	7511
P	4520	12961

Тачност	0,94
Прецизност	0,63
Осетљивост	0,74
F1-скор	0,68

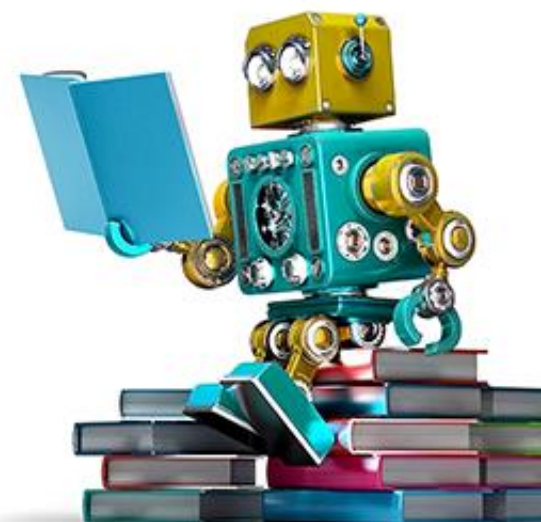


Логистичка регресија

Без 4. и 5. обележја

	\hat{N}	\hat{P}
N	181219	1300
P	7370	10111

Тачност	0,96
Прецизност	0,89
Осетљивост	0,58
F1-скор	0,70

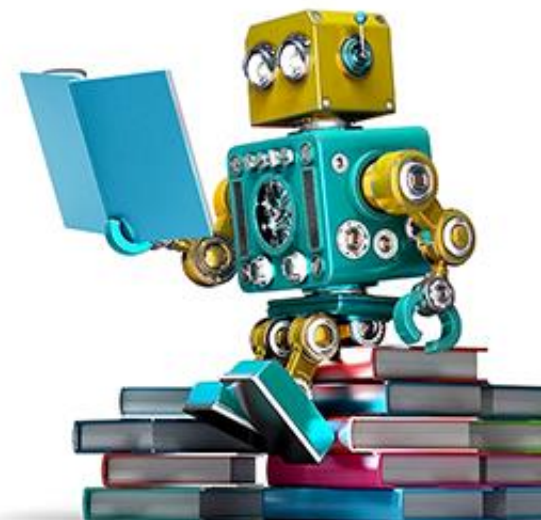


Логистичка регресија

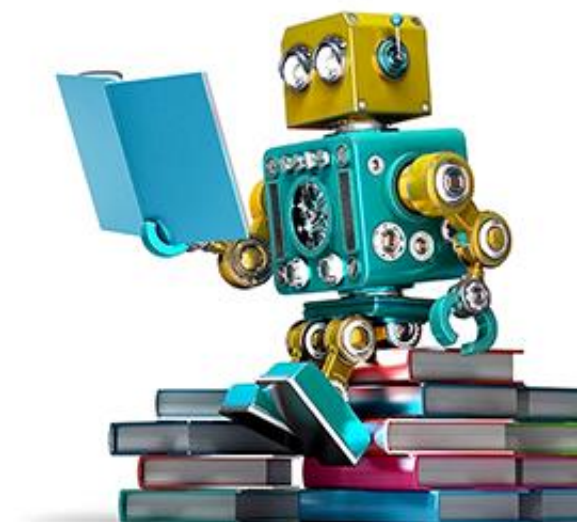
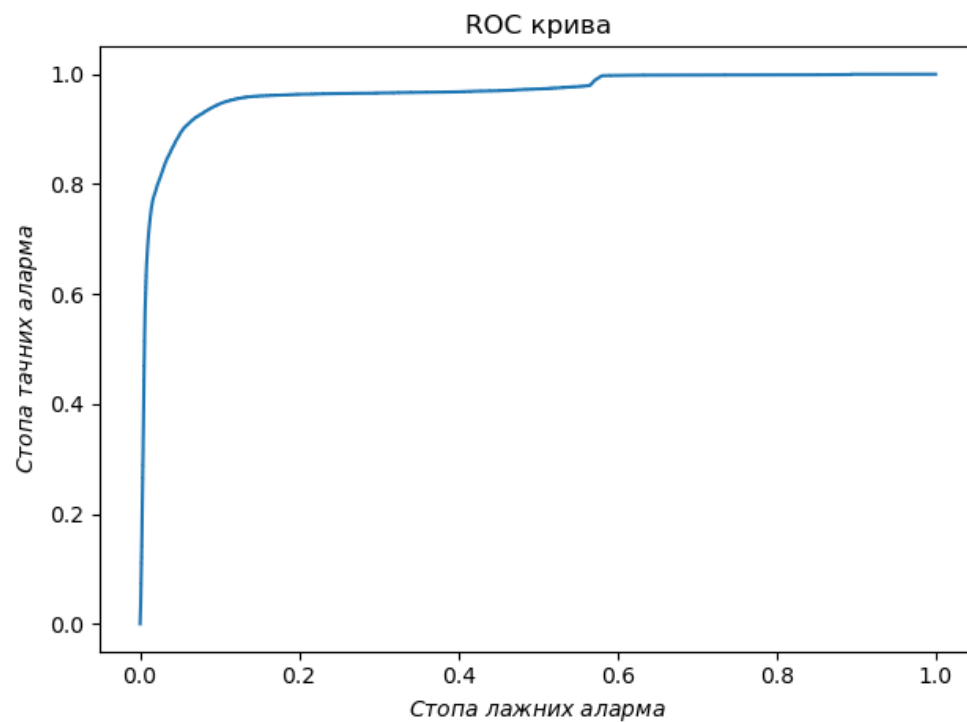
Без 4. обележја

	\hat{N}	\hat{P}
N	181268	1251
P	7039	10442

Тачност	0,96
Прецизност	0,89
Осетљивост	0,60
F1-скор	0,72



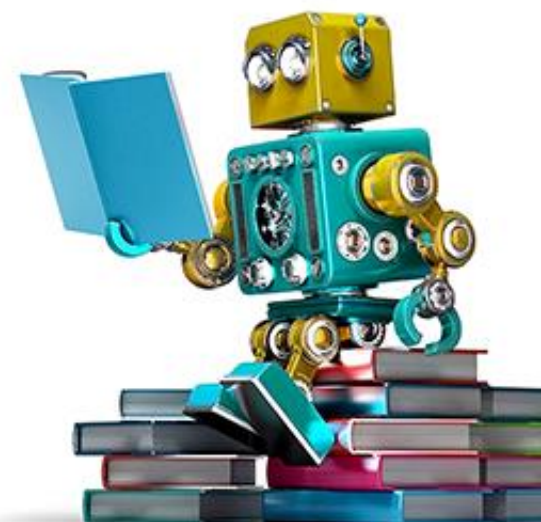
Логистичка регресија



Гаусовски наивни Бејз

	\hat{N}	\hat{P}
N	179752	2767
P	7179	10284

Тачност	0,95
Прецизност	0,79
Осетљивост	0,59
F1-скор	0,67



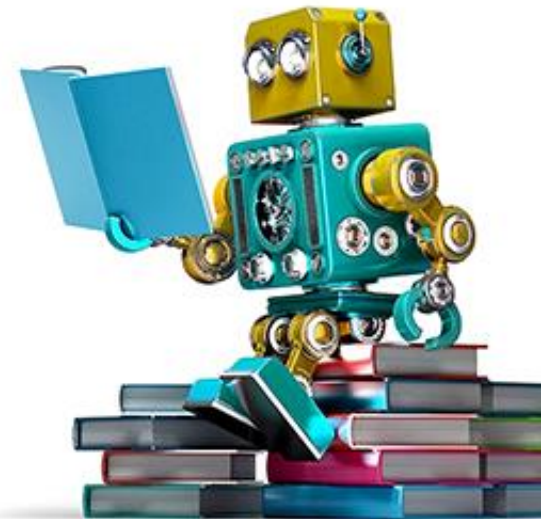
Метода носећих вектора

- Примењен ради пробе

	\hat{N}	\hat{P}
N	182405	114
P	264	17217

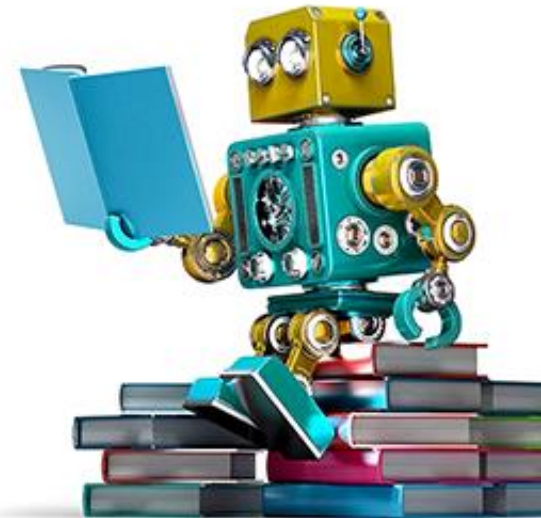
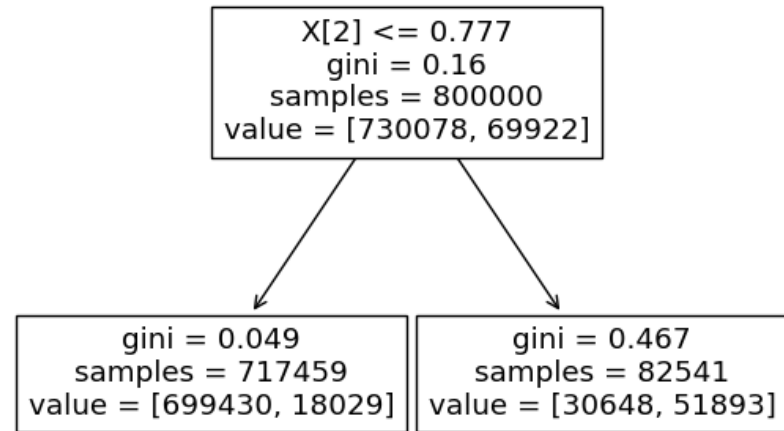
Тачност	0,99
Прецизност	0,99
Осетљивост	0,98
F1-скор	0,99

- За 800 000 података потребно ≈ 1398 s



Стабло

Стабло дубине један

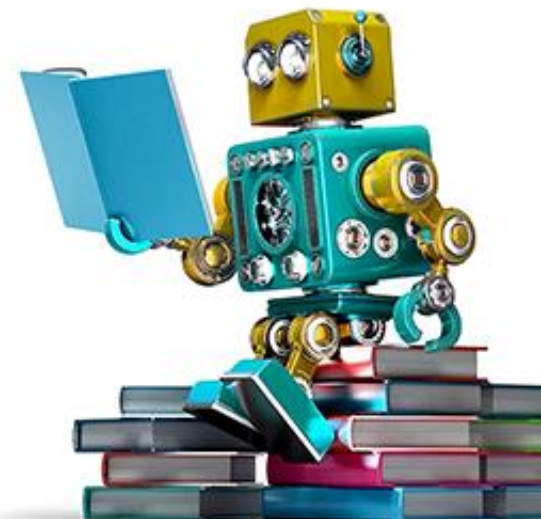


Стабло

Стабло дубине један

	\hat{N}	\hat{P}
N	174960	7559
P	4423	13058

Тачност	0,94
Прецизност	0,63
Осетљивост	0,75
F1-скор	0,69

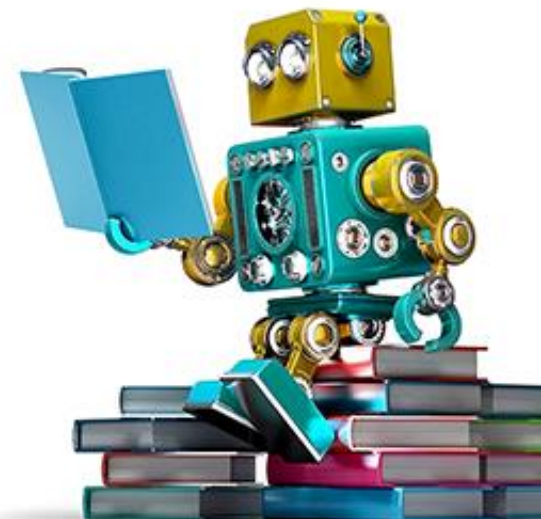


Стабло

Стабло дубине два

	\hat{N}	\hat{P}
N	181200	1319
P	5315	12166

Тачност	0,97
Прецизност	0,90
Осетљивост	0,70
F1-скор	0,79

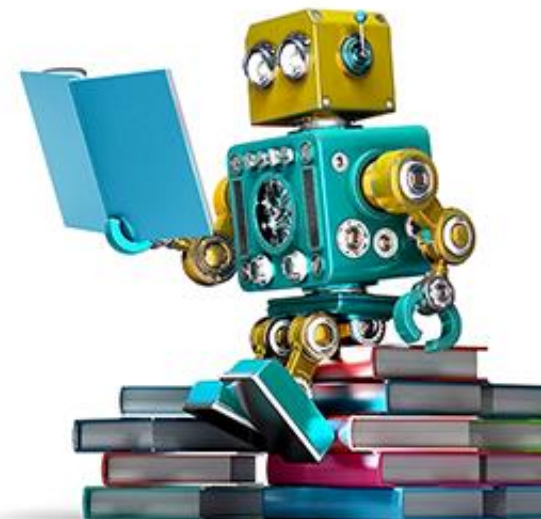


Стабло

Стабло дубине три

	\hat{N}	\hat{P}
N	180060	2459
P	1717	15764

Тачност	0,98
Прецизност	0,87
Осетљивост	0,90
F1-скор	0,88

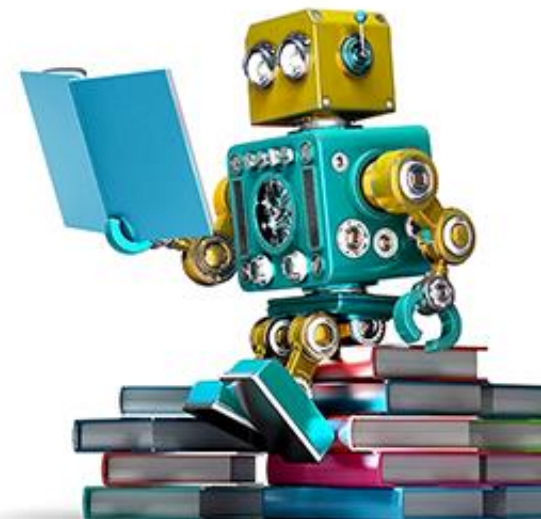


Стабло

Стабло дубине пет

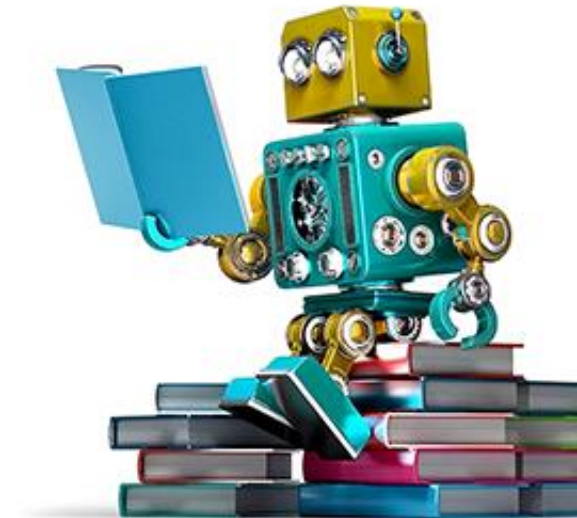
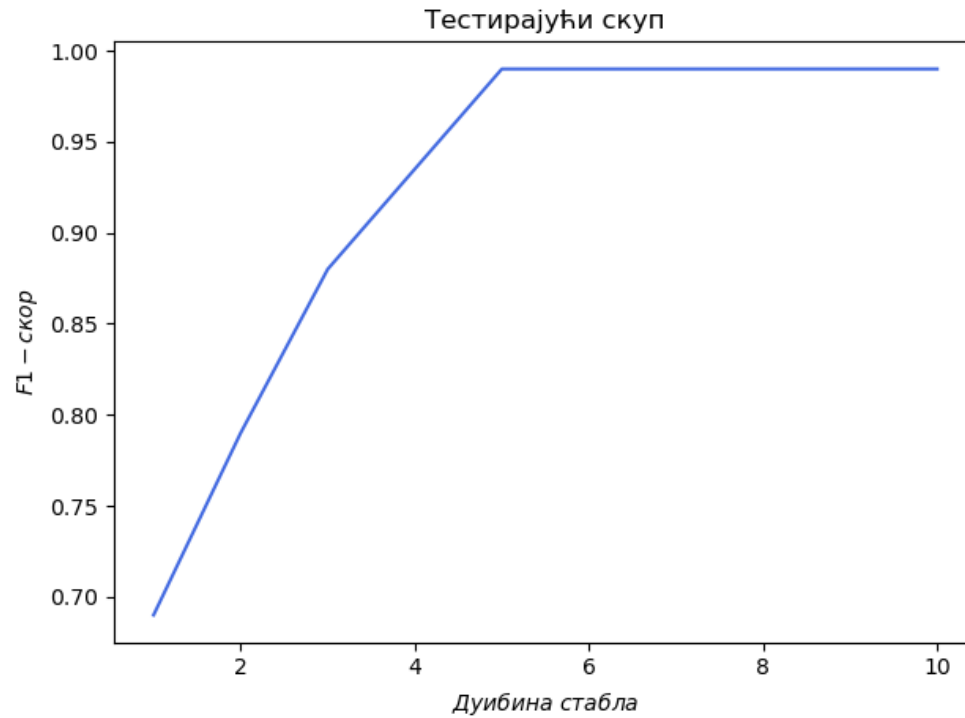
	\hat{N}	\hat{P}
N	182346	173
P	77	17404

Тачност	0,99
Прецизност	0,99
Осетљивост	0,99
F1-скор	0,99



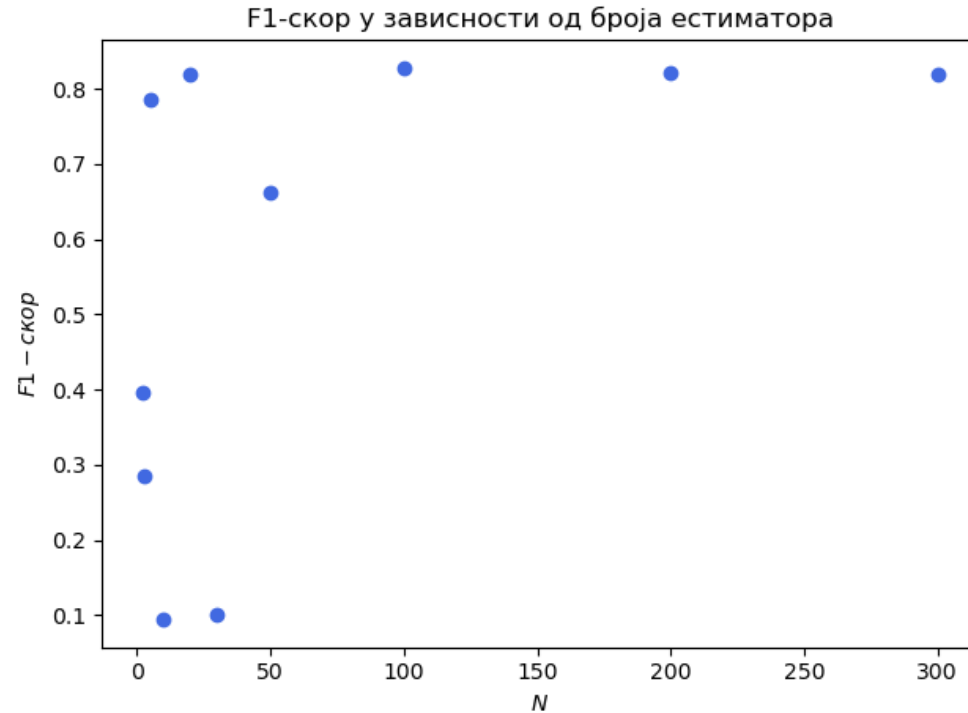
Стабло

$F1$ -скор у зависности од дубине стабла (тестирајући скуп)

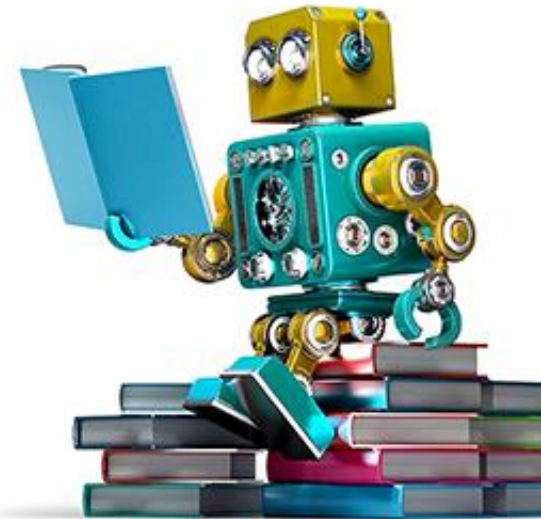


Случајне шуме

$$n_{\max}=2 \quad d_{\max}=3$$



≈ 80 s за 100 чланова ансамбла



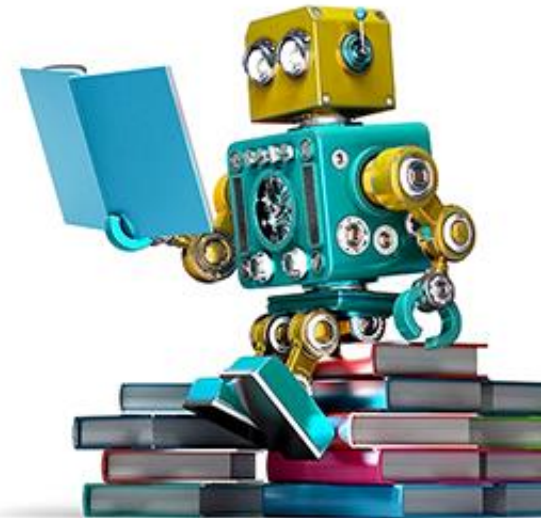
Случајне шуме

$N=10$ $d_{\max}=4$ $n_{\max}=3$

	\hat{N}	\hat{P}
N	182519	0
P	689	16792

Тачност	0,99
Прецизност	1,00
Осетљивост	0,96
F1-скор	0,98

≈ 15 s



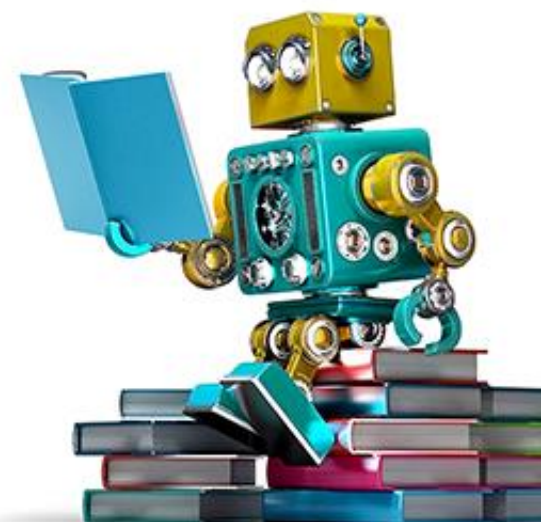
Случајне шуме

$N=10$ $d_{\max}=5$ $n_{\max}=3$

	\hat{N}	\hat{P}
N	182364	155
P	153	17328

Тачност	0,99
Прецизност	0,99
Осетљивост	0,99
F1-скор	0,99

≈ 15 s



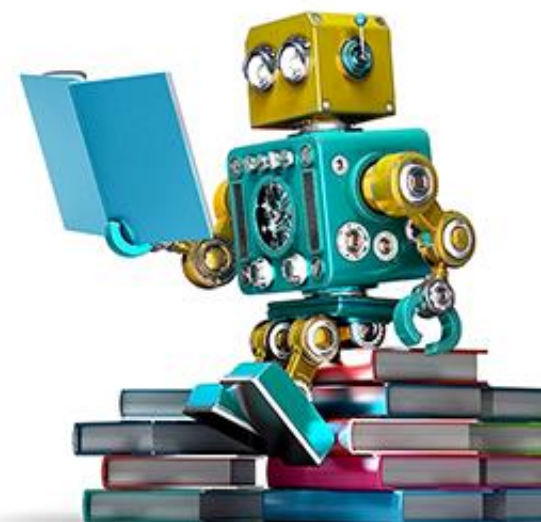
XGBoost метода

$N=10$ $d_{\max}=4$ $n_{\max}=3$

	\hat{N}	\hat{P}
N	182352	167
P	79	17402

Тачност	0,99
Прецизност	0,99
Осетљивост	0,99
F1-скор	0,99

$\approx 8 \text{ s}$



Хвала на пажњи!

