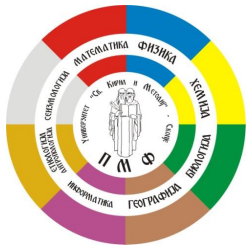


Природно-математички факултет - УКИМ
"Алгоритми и примена"
Семинар (10.12.2016)

АЛГОРИТАМ ЗА АНАЛИЗА НА ВЛИЈАНИЕ НА ПОДАТОЧНИ МНОЖЕСТВА ВО КЛАСИФИКАЦИСКИ ПРОБЛЕМ

дипл.инж. Билјана Заева
Asseco SEE ДООЕЛ Македонија - Скопје
biljana.zaeva@yahoo.com



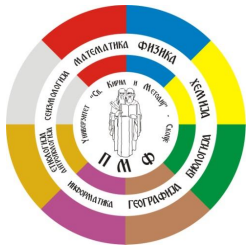
АЛГОРИТАМ ЗА АНАЛИЗА НА ВЛИЈАНИЕ НА ПОДАТОЧНИ МНОЖЕСТВА ВО КЛАСИФИКАЦИСКИ ПРОБЛЕМ

Вовед

❖ Што е Дијабетска Ретинопатија?



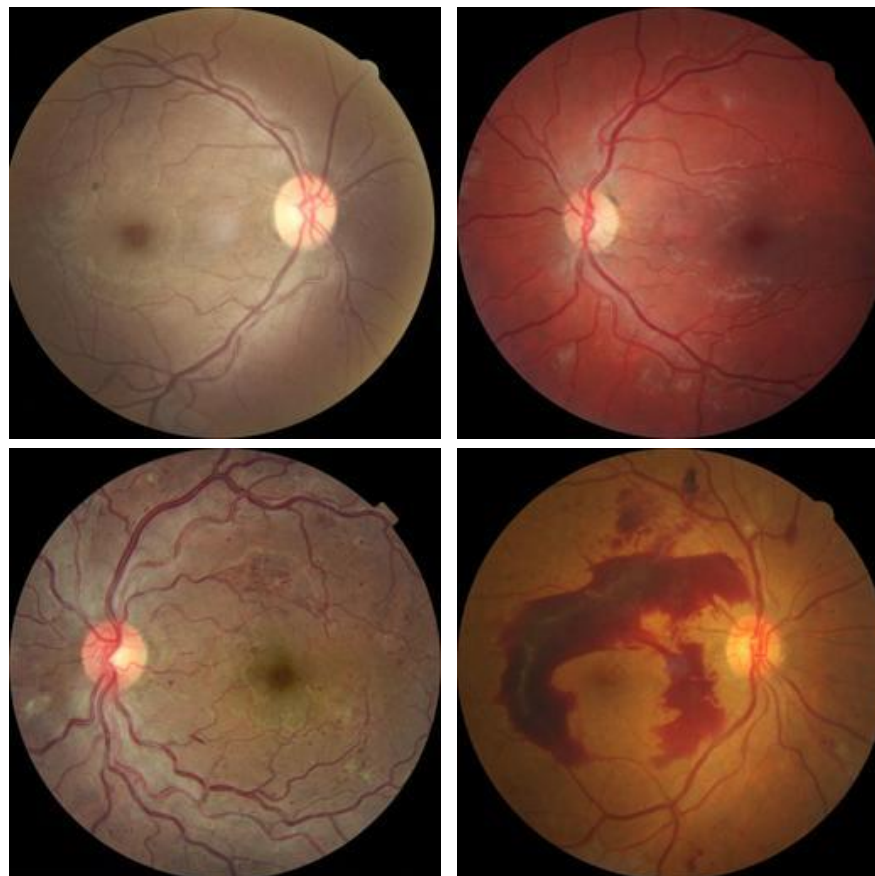
- ❖ Фудус слики
- ❖ Зошто?

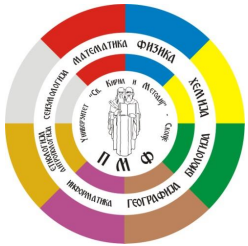


АЛГОРИТАМ ЗА АНАЛИЗА НА ВЛИЈАНИЕ НА ПОДАТОЧНИ МНОЖЕСТВА ВО КЛАСИФИКАЦИСКИ ПРОБЛЕМ

Вовед

❖ 4 фази на Дијабетска Ретинопатија

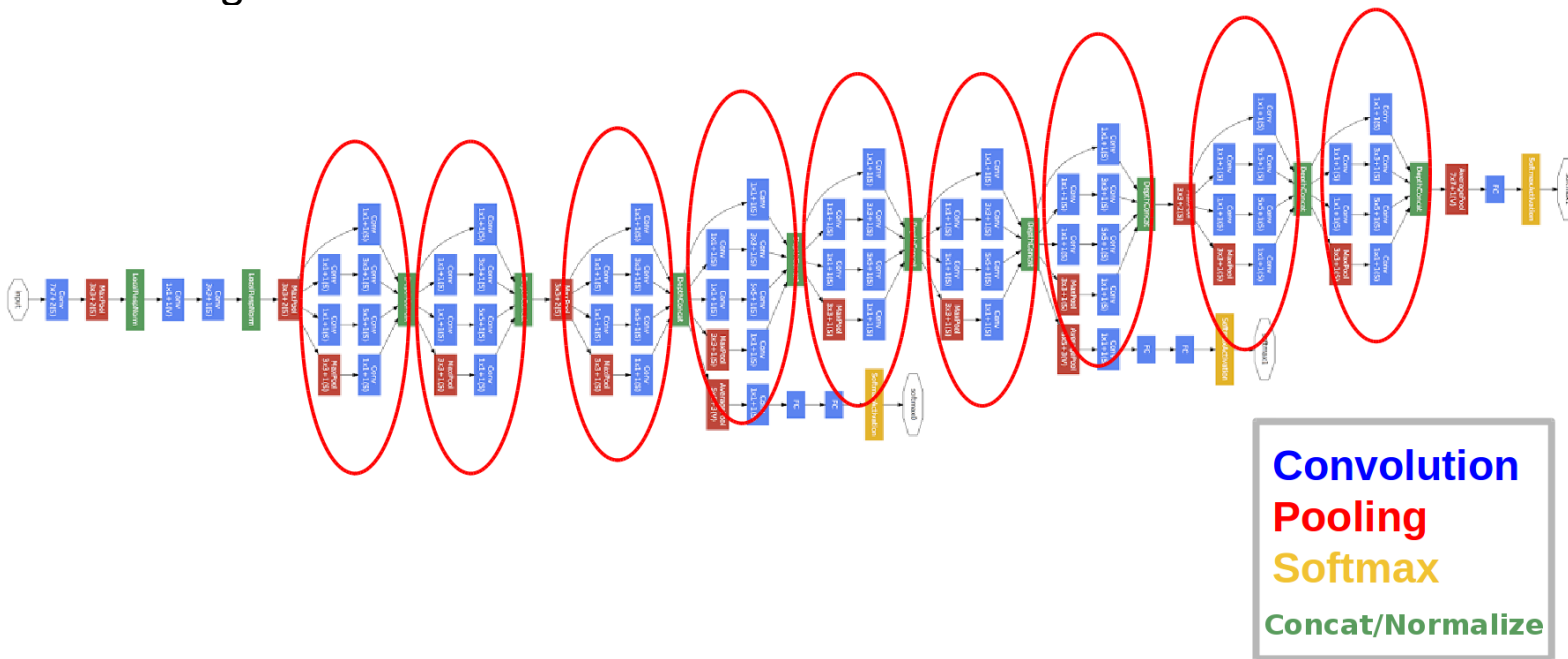




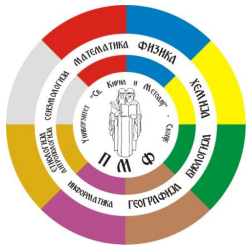
АЛГОРИТАМ ЗА АНАЛИЗА НА ВЛИЈАНИЕ НА ПОДАТОЧНИ МНОЖЕСТВА ВО КЛАСИФИКАЦИСКИ ПРОБЛЕМ

Вовед

❖ Зошто GoogLeNet?



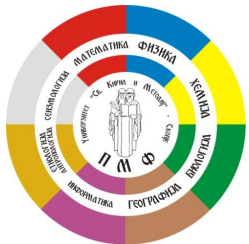
❖ Користена технологија - DIGITS



АЛГОРИТАМ ЗА АНАЛИЗА НА ВЛИЈАНИЕ НА ПОДАТОЧНИ МНОЖЕСТВА ВО КЛАСИФИКАЦИСКИ ПРОБЛЕМ

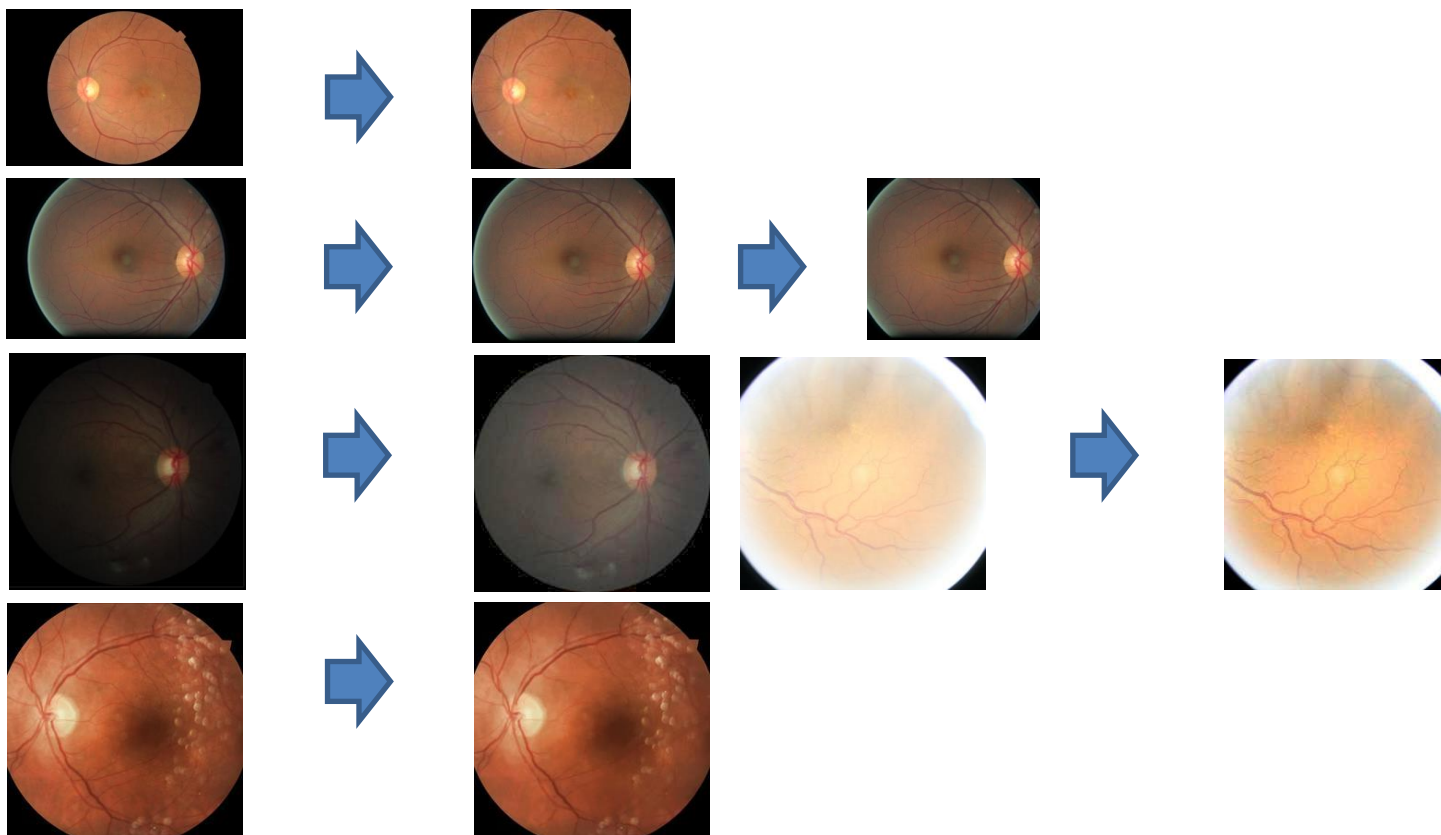
Методологија

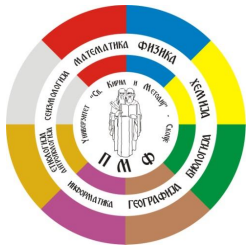
- ❖ Предпроцесирање
- ❖ Оригинално податочно множество
- ❖ Хипотези
- ❖ Нови податочни множества
- ❖ 9 модели на GoogLeNet
- ❖ Анализа на резултатите



АЛГОРИТАМ ЗА АНАЛИЗА НА ВЛИЈАНИЕ НА ПОДАТОЧНИ МНОЖЕСТВА ВО КЛАСИФИКАЦИСКИ ПРОБЛЕМ

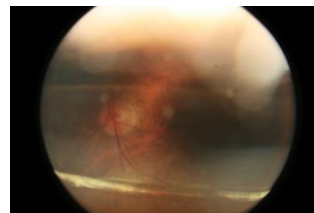
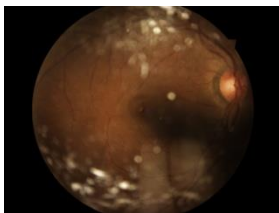
Предпроцесирање





АЛГОРИТАМ ЗА АНАЛИЗА НА ВЛИЈАНИЕ НА ПОДАТОЧНИ МНОЖЕСТВА ВО КЛАСИФИКАЦИСКИ ПРОБЛЕМ

Предпроцесирање



Слика 19 - оригинал



Слика 20 - flip



Слика 21 - mirror

Слика 22 - flip + mirror



Слика 23 - flip + mirror + 90



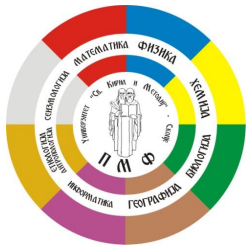
Слика 24 - mirror + 90



Слика 25 - flip + 90

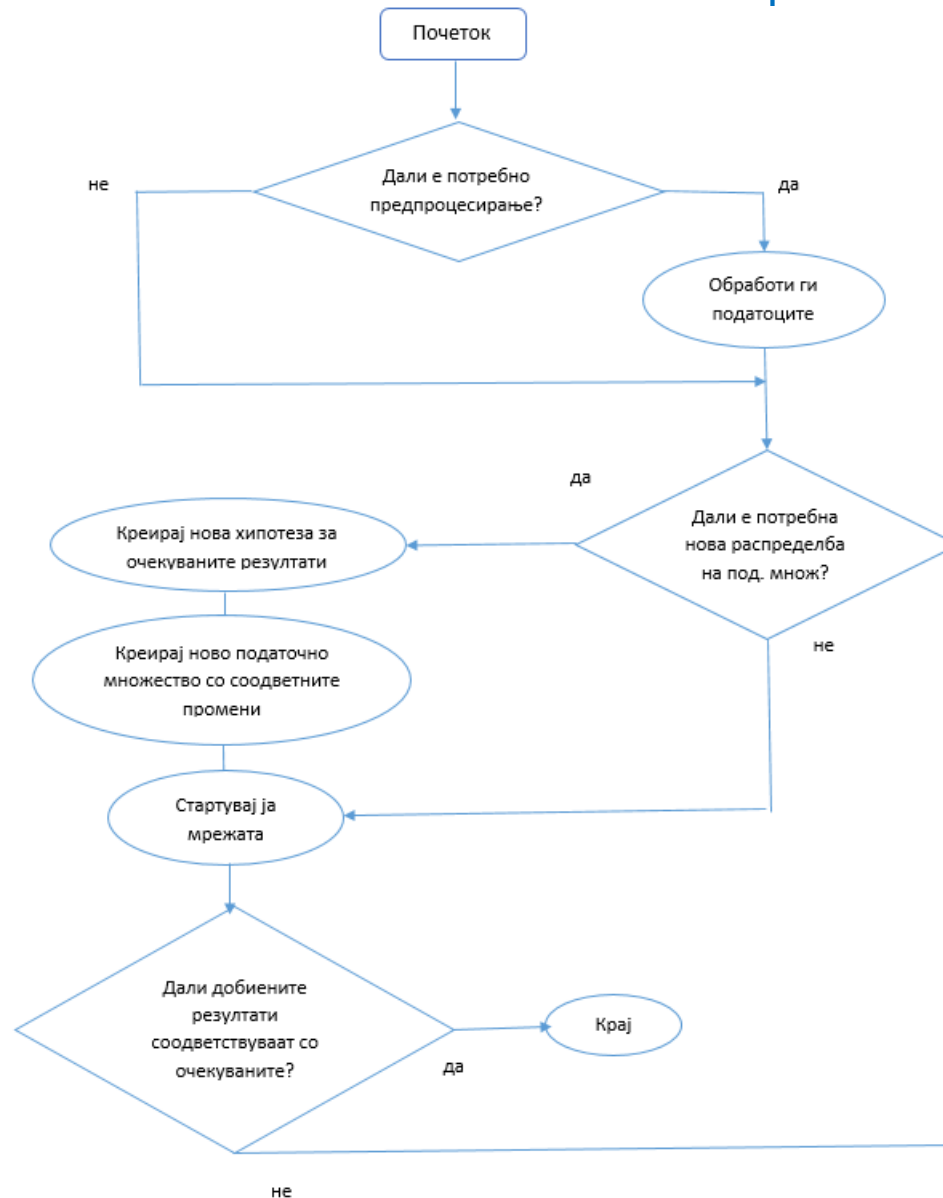


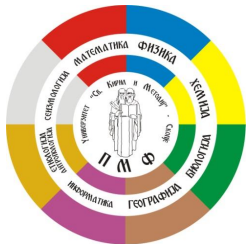
Слика 26 - оригинал + 90



АЛГОРИТАМ ЗА АНАЛИЗА НА ВЛИЈАНИЕ НА ПОДАТОЧНИ МНОЖЕСТВА ВО КЛАСИФИКАЦИСКИ ПРОБЛЕМ

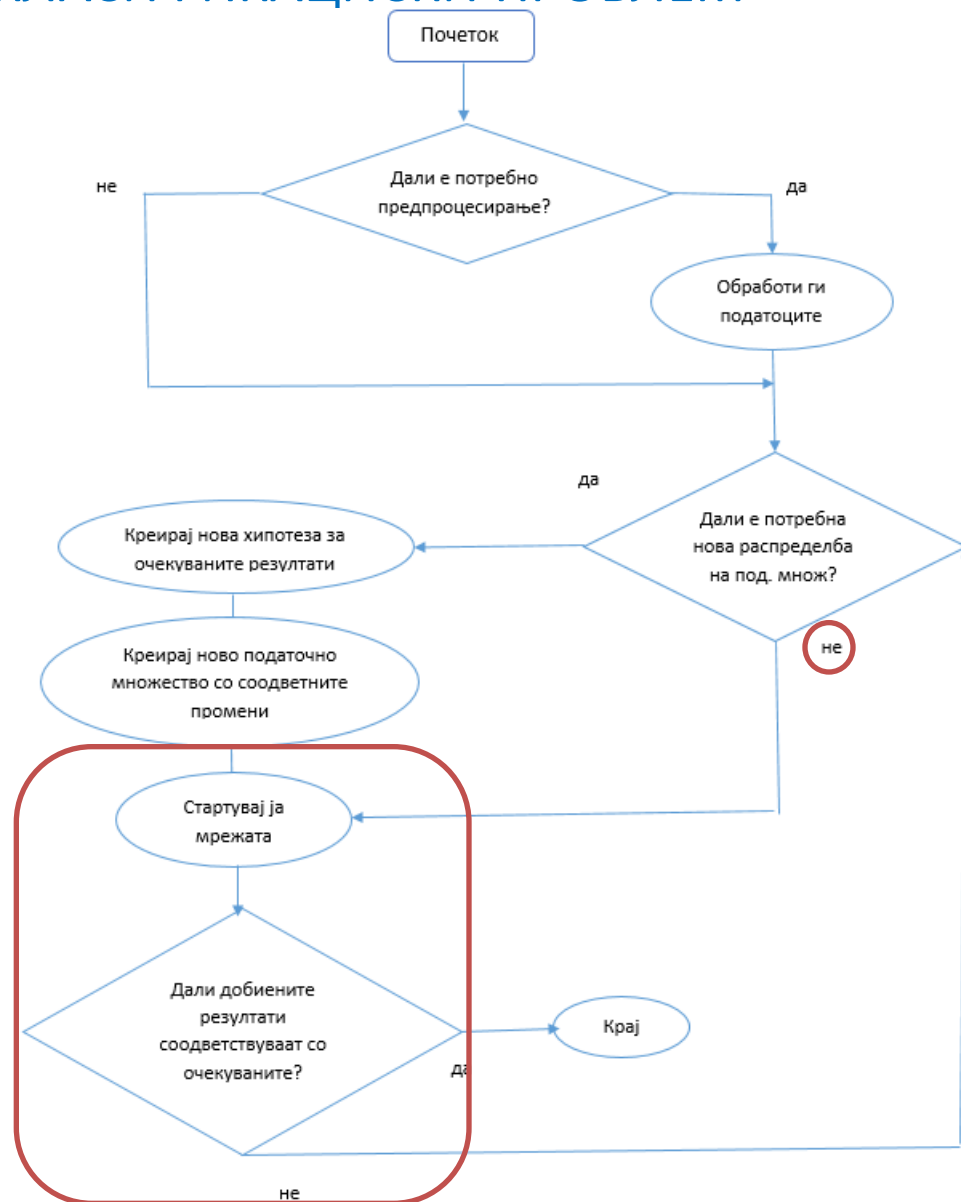
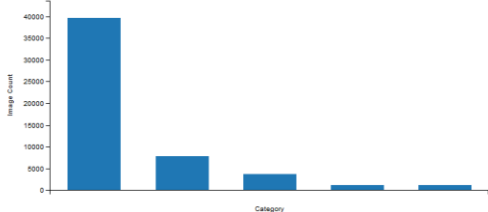
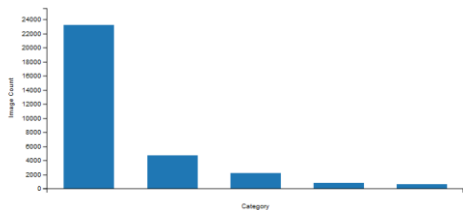
Алгоритам

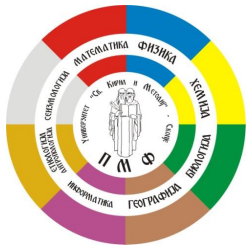




АЛГОРИТАМ ЗА АНАЛИЗА НА ВЛИЈАНИЕ НА ПОДАТОЧНИ МНОЖЕСТВА ВО КЛАСИФИКАЦИСКИ ПРОБЛЕМ

Оригинално податочно множество





АЛГОРИТАМ ЗА АНАЛИЗА НА ВЛИЈАНИЕ НА ПОДАТОЧНИ МНОЖЕСТВА ВО КЛАСИФИКАЦИСКИ ПРОБЛЕМ

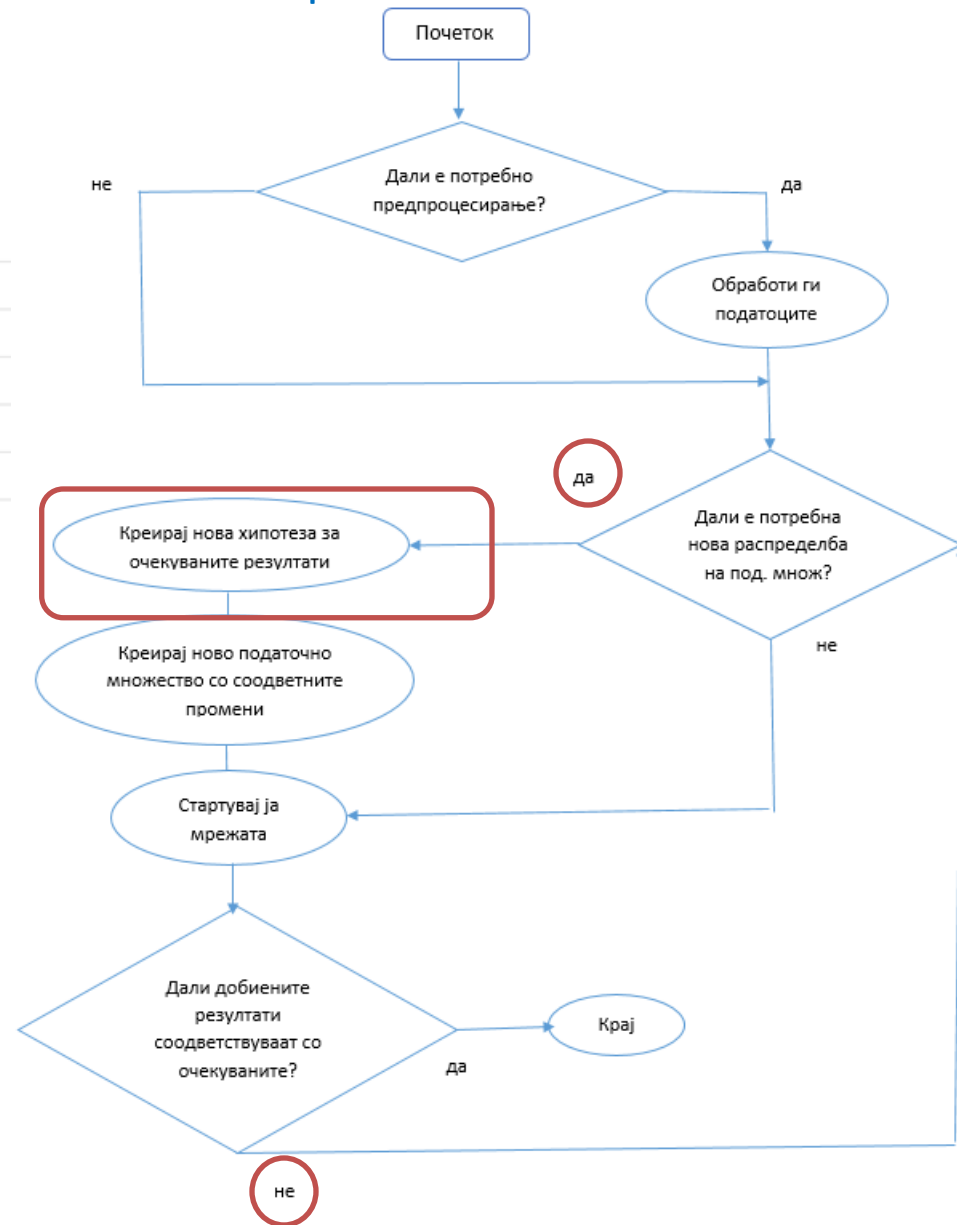
Резултати

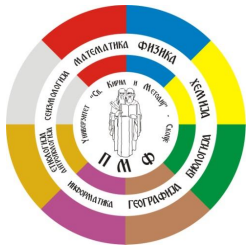
Confusion matrix

	faza1	faza2	faza3	faza4	zdravi	Per-class accuracy
faza1	0	4	0	0	3758	0.0%
faza2	0	28	0	0	7833	0.36%
faza3	0	29	0	0	1185	0.0%
faza4	0	39	0	0	1167	0.0%
zdravi	0	54	0	0	39479	99.86%

Хипотеза 1

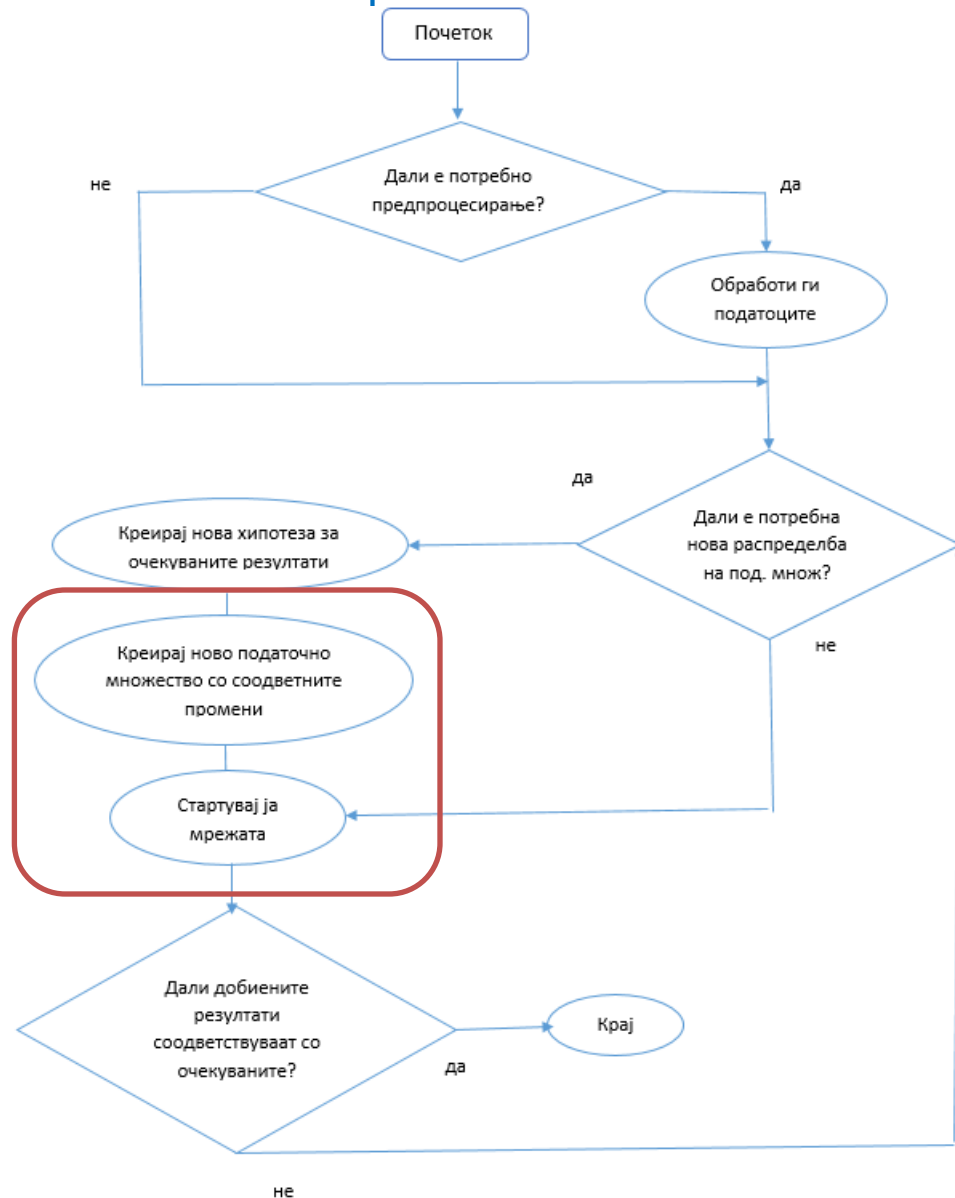
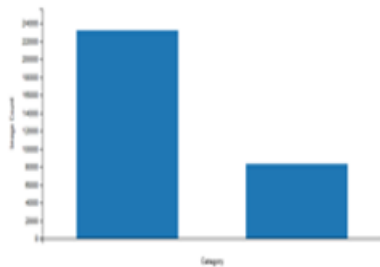
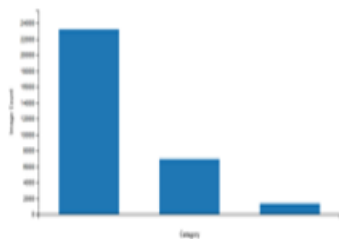
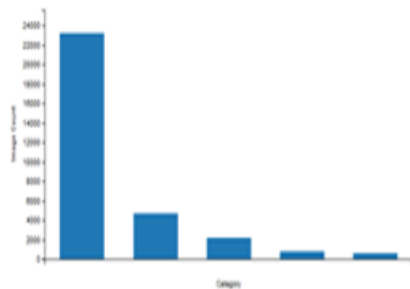
“Доколку се намали бројот на класите, ќе се зголеми успешноста на мрежата при класификација”

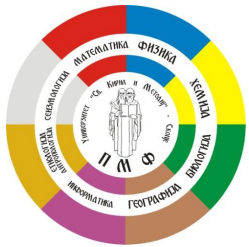




АЛГОРИТАМ ЗА АНАЛИЗА НА ВЛИЈАНИЕ НА ПОДАТОЧНИ МНОЖЕСТВА ВО КЛАСИФИКАЦИСКИ ПРОБЛЕМ

BASELINE податочно множество





АЛГОРИТАМ ЗА АНАЛИЗА НА ВЛИЈАНИЕ НА ПОДАТОЧНИ МНОЖЕСТВА ВО КЛАСИФИКАЦИСКИ ПРОБЛЕМ

Резултати

Confusion matrix

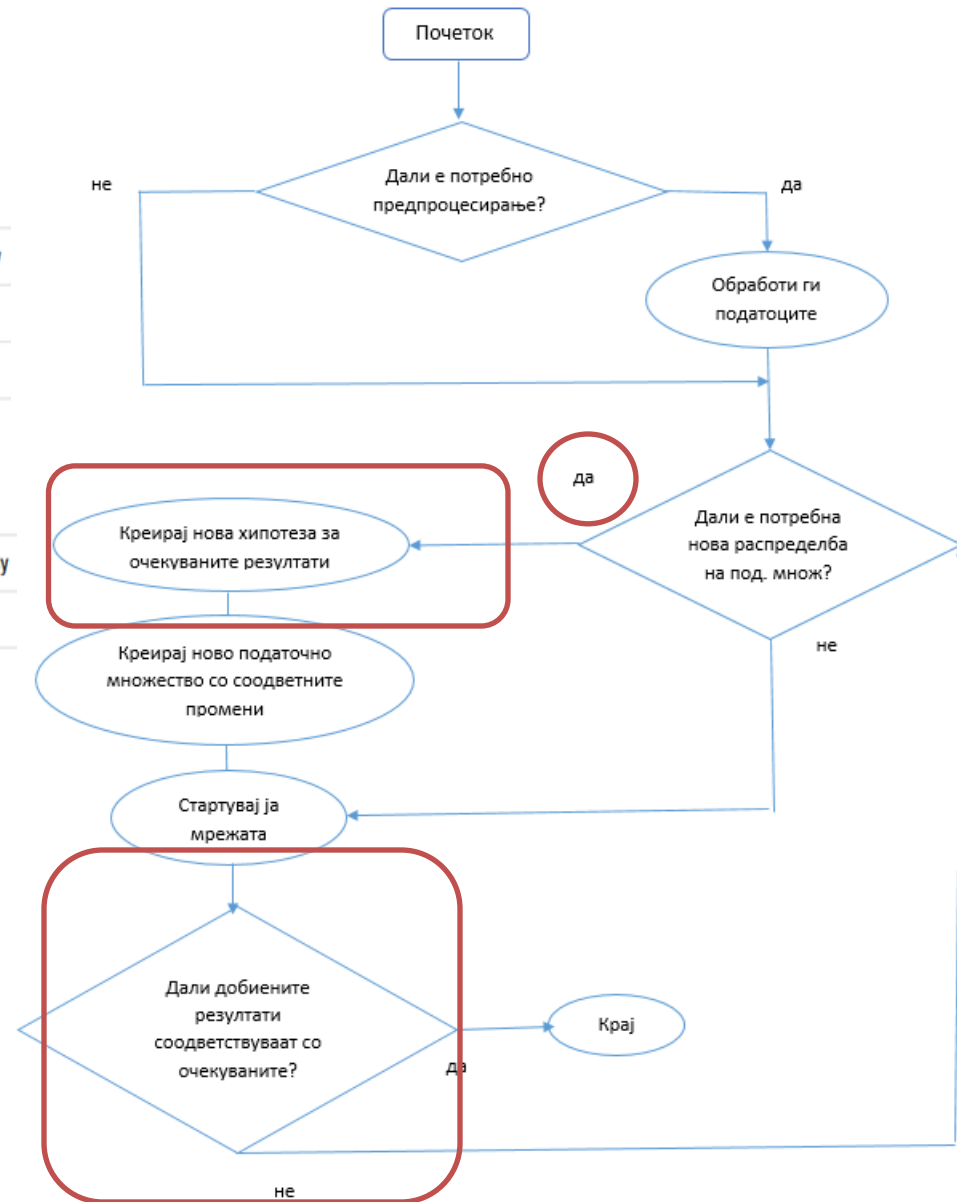
	faza1	faza3	zdravi	Per-class accuracy
faza1	53	0	11570	0.46%
faza3	94	0	2326	0.0%
zdravi	63	2	39468	99.84%

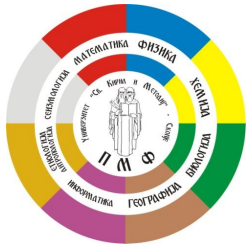
Confusion matrix

	bolni	zdravi	Per-class accuracy
bolni	1402	12641	9.98%
zdravi	1224	38309	96.9%

Хипотеза 2

“Доколку го агументираме (зголемиме) податочното множество, ќе се зголеми успешноста на мрежата при класификација”

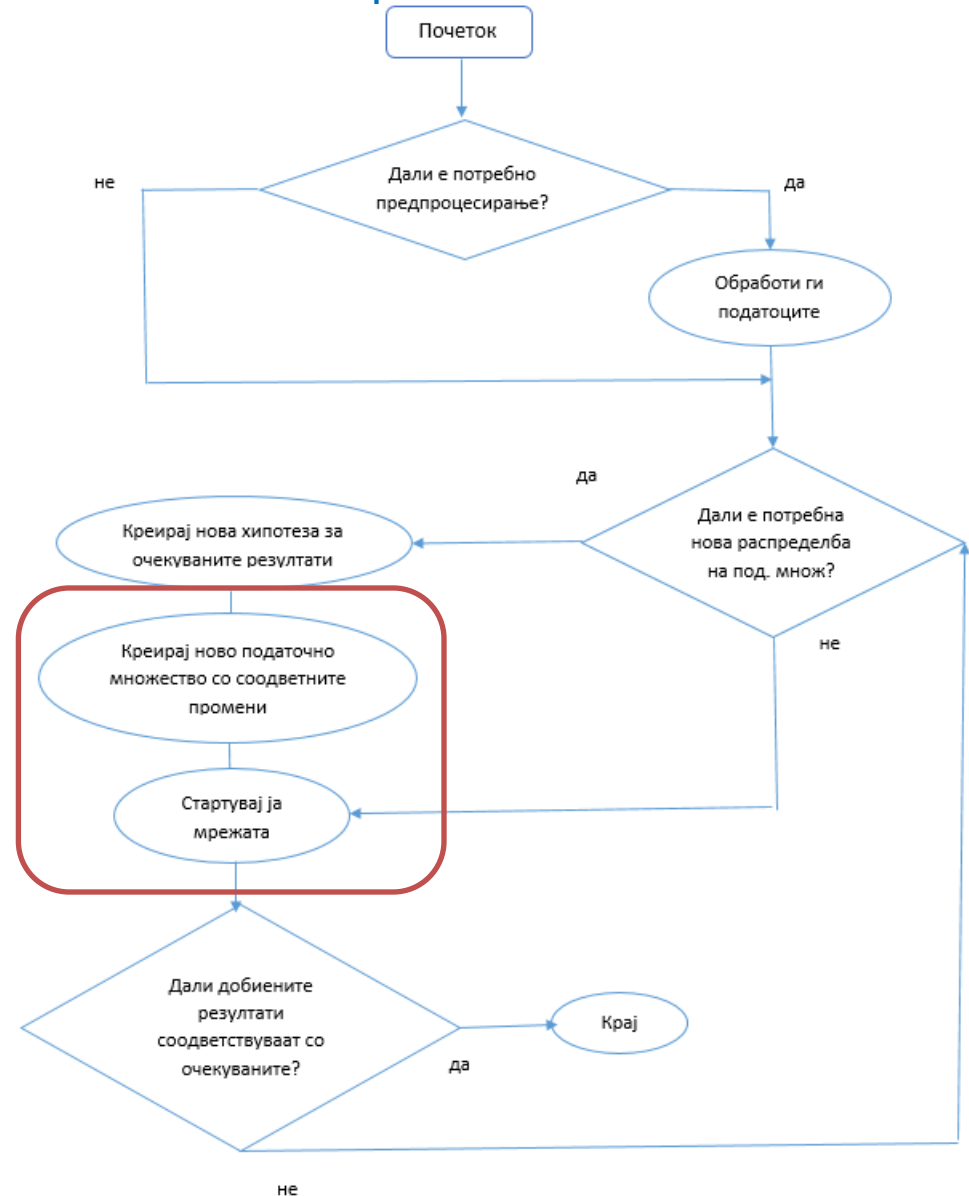
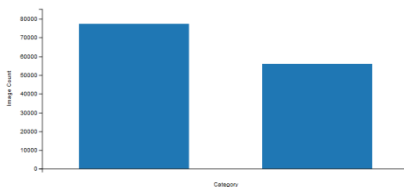
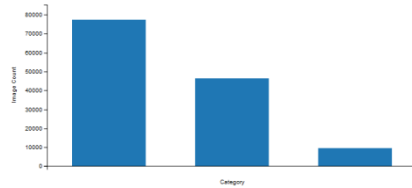
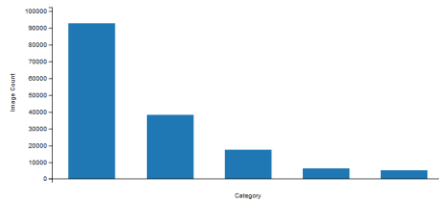


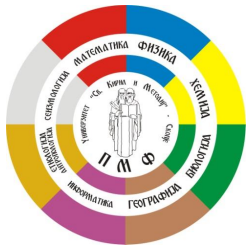


АЛГОРИТАМ ЗА АНАЛИЗА НА ВЛИЈАНИЕ НА ПОДАТОЧНИ МНОЖЕСТВА ВО КЛАСИФИКАЦИСКИ ПРОБЛЕМ

Аугументирано небалансирано

податочно множество





АЛГОРИТАМ ЗА АНАЛИЗА НА ВЛИЈАНИЕ НА ПОДАТОЧНИ МНОЖЕСТВА ВО КЛАСИФИКАЦИСКИ ПРОБЛЕМ

Резултати

Confusion matrix

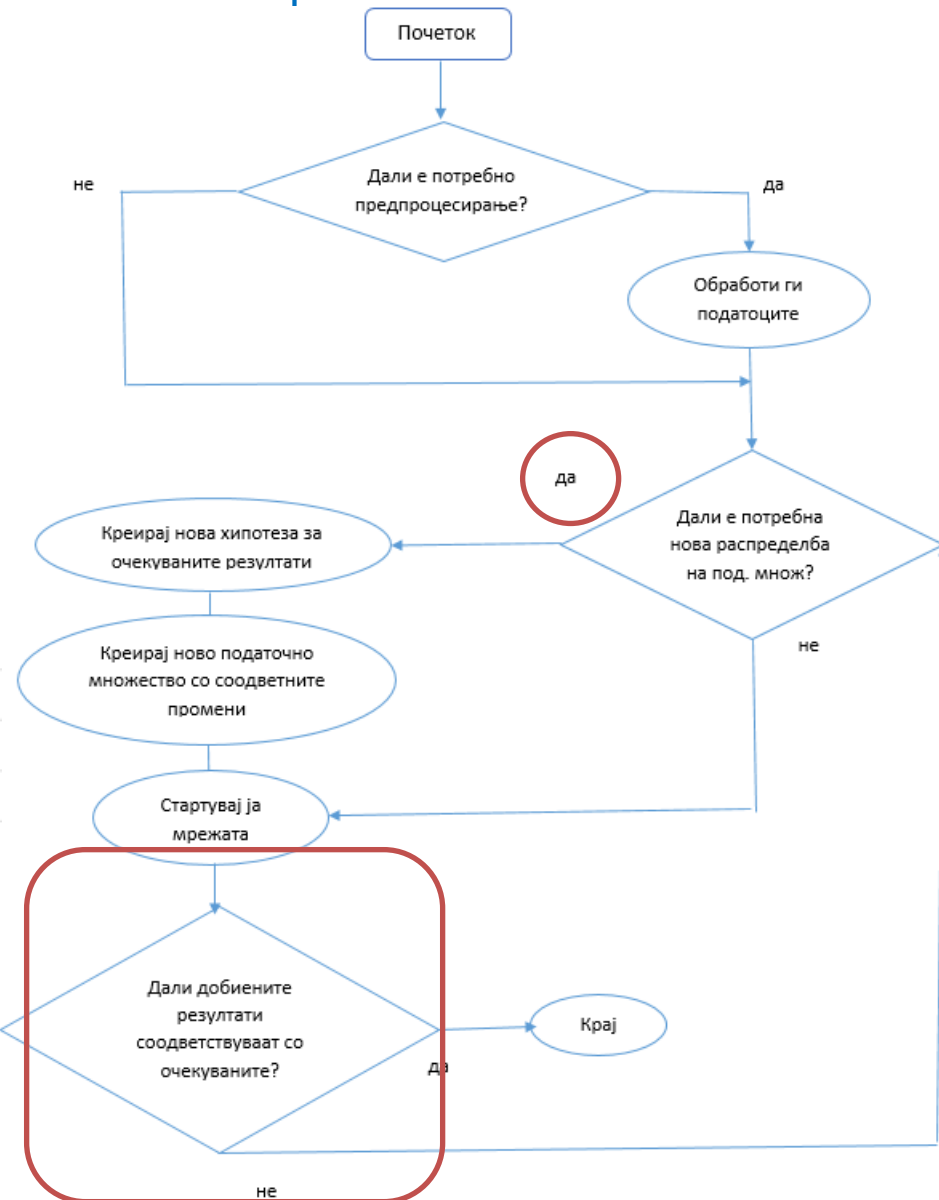
	faza1	faza2	faza3	faza4	zdravi	Per-class accuracy
faza1	1	213	0	8	3540	0.03%
faza2	7	2394	110	137	5213	30.45%
faza3	0	639	126	70	379	10.38%
faza4	0	345	58	395	408	32.75%
zdravi	20	1394	5	189	37925	95.93%

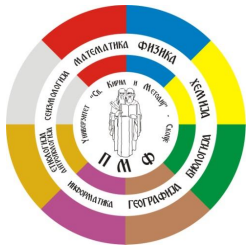
Confusion matrix

	faza1	faza3	zdravi	Per-class accuracy
faza1	2578	361	8684	22.18%
faza3	886	737	797	30.45%
zdravi	1641	217	37675	95.3%

Confusion matrix

	bolni	zdravi	Per-class accuracy
bolni	5170	8873	36.82%
zdravi	3076	36457	92.22%

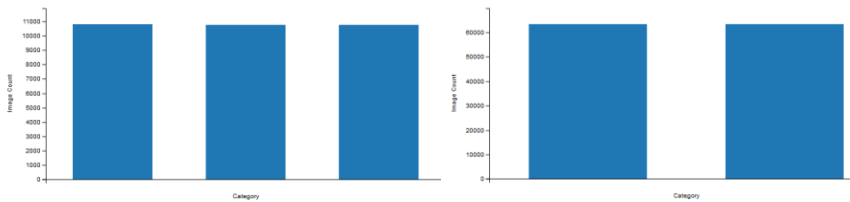
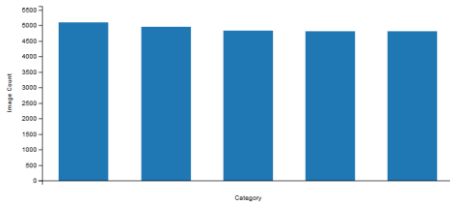




АЛГОРИТАМ ЗА АНАЛИЗА НА ВЛИЈАНИЕ НА ПОДАТОЧНИ МНОЖЕСТВА ВО КЛАСИФИКАЦИСКИ ПРОБЛЕМ

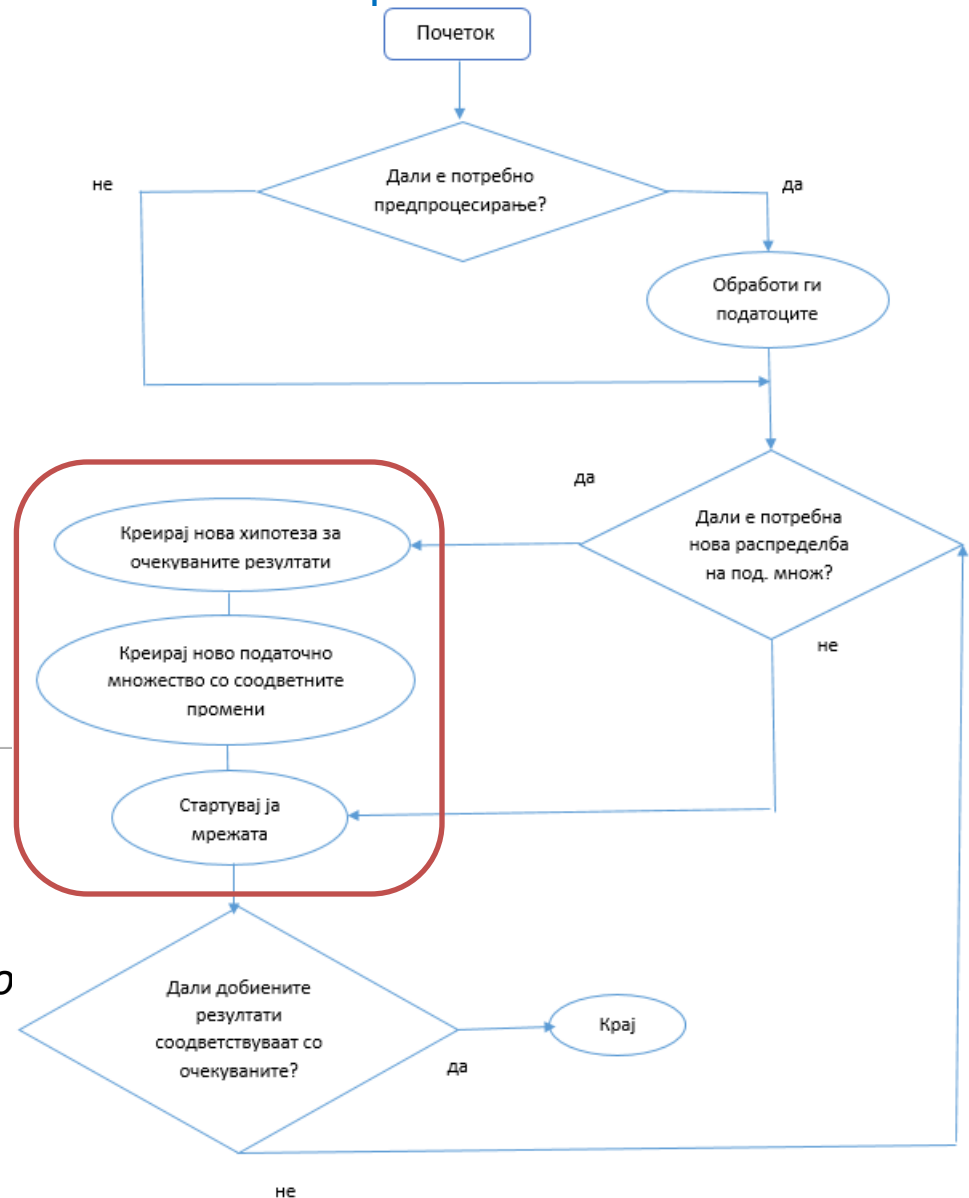
Аугументирано балансирано

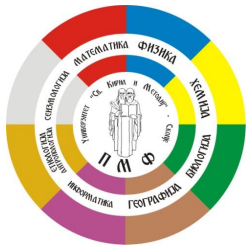
податочно множество



Хипотеза 3

“Доколку податочното множество балансирано, ќе се зголеми успешноста на мрежата при класификација”





АЛГОРИТАМ ЗА АНАЛИЗА НА ВЛИЈАНИЕ НА ПОДАТОЧНИ МНОЖЕСТВА ВО КЛАСИФИКАЦИСКИ ПРОБЛЕМ

Резултати

Confusion matrix

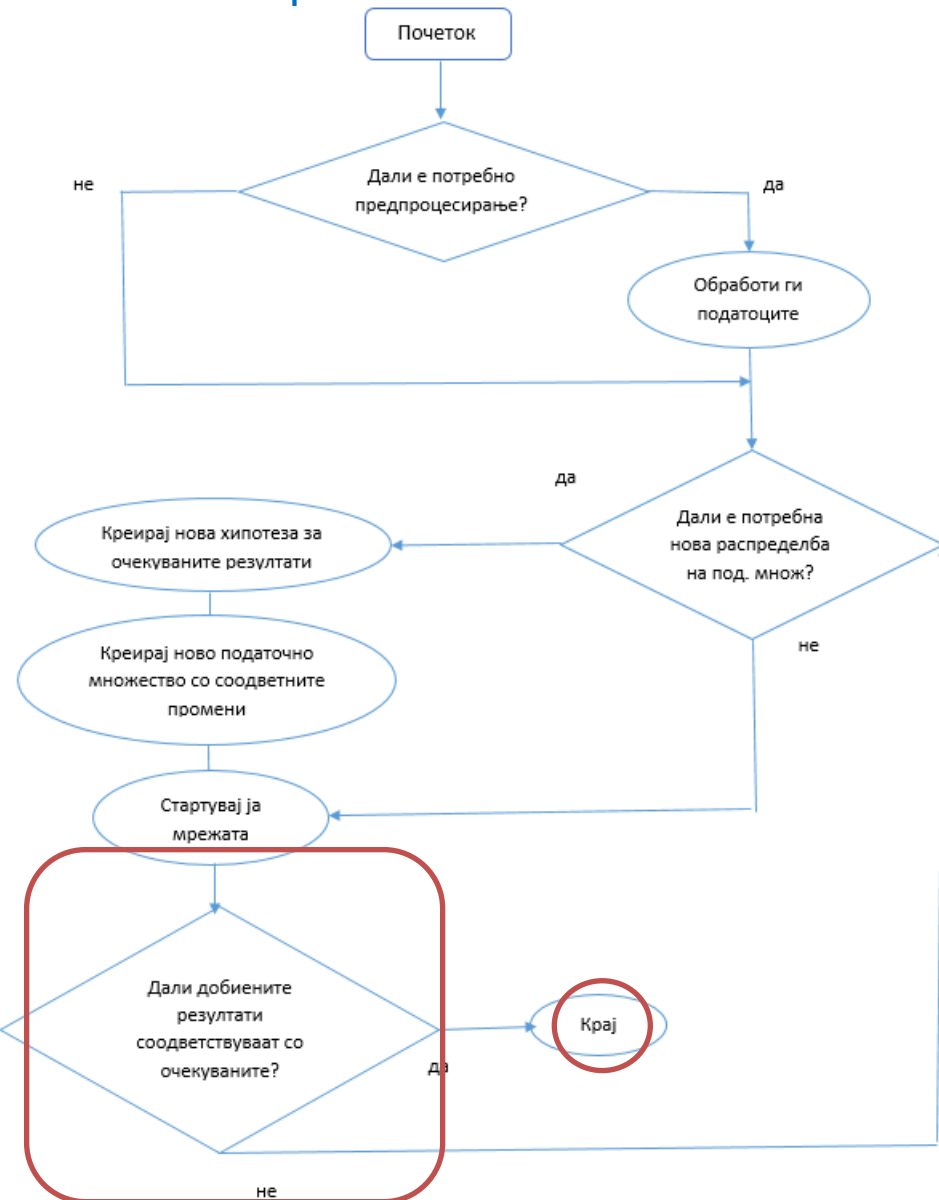
	faza1	faza2	faza3	faza4	zdravi	Per-class accuracy
faza1	54	934	51	103	2620	1.44%
faza2	82	3313	679	630	3157	42.14%
faza3	6	499	345	169	195	28.42%
faza4	3	304	156	590	153	48.92%
zdravi	604	9027	367	1388	28147	71.2%

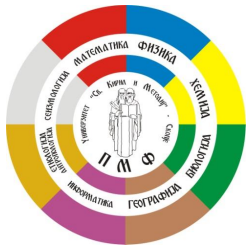
Confusion matrix

	faza1	faza3	zdravi	Per-class accuracy
faza1	2112	1418	8093	18.17%
faza3	499	1292	629	53.39%
zdravi	2819	1585	35129	88.86%

Confusion matrix

	bolni	zdravi	Per-class accuracy
bolni	6621	7422	47.15%
zdravi	6391	33142	83.83%





АЛГОРИТАМ ЗА АНАЛИЗА НА ВЛИЈАНИЕ НА ПОДАТОЧНИ МНОЖЕСТВА ВО КЛАСИФИКАЦИСКИ ПРОБЛЕМ

Анализа и Заклучок

- ❖ Во однос на **распределбата** по класи
- ❖ Во однос на **аугментирањето** на податочните множества
- ❖ Во однос на **балансираноста** на податочните множества

- ❖ Идна работа: *До кој степен треба да се аугментира податочното множество и во кој момент треба да се прекине?*