
1 Príloha B – Zoznam odborných a vedeckých prác zaoberajúcich sa selekciou atribútov

Tab. 1 Odborné a vedecké články na tematiku klasifikácie malvéru.

článok	dáta	interpretácia	metóda	úspešnosť
[32]	sekvencie bajtov	obraz	CNN (konvolučná neurónová sieť)	97.5%
[8]	frekvencia bajtov, 4-gramy bajtov, mená funkcií, assembly atribúty, frekvencia a počet inštrukcií, denzita pixelov, počet segmentov	n-gramy, textové reťazce, obraz	Ensemble DT (rozhodovacích stromov)	0.0028 (log loss), 99.83% (presnosť)
[33]	4-gramy bajtov	hash mapa - obraz	CNN	99.0% (presnosť)
[104]	n-gramy bajtov	obraz	CNN	69.1%
[2]	n-gramy inštrukcií, metadáta, spustiteľný súbor	obraz, text	CNN, LSTM (Long short-term memory), LR (logistická regresia)	99.36% (presnosť)
[102]	sekvencie bajtov	text	DRN (Dilated residual network)	98.6%
[101]	spustiteľný súbor	obraz	CNN	94.0%
[34]	sekvencie bajtov	vektor entropie	CNN, MRCNN (Mask regions with CNN features)	98.96%
[111]	grafy volaní funkcií	Vektor atribútov grafu	RF (náhodný les)	99.33% (presnosť)
[110]	grafy volaní funkcií	graf	Graf uniqueness matching	77.44%

[109]	spustiteľný súbor	obraz	CNN	98.0%
[108]	spustiteľný súbor	obraz	CNN	98.63%
[107]	kombinácia mnohých statických a dynamických dát	n-gramy	DT, KNN (k-najbližších susedov), LR, NB (naivný Bayes), RF, SVM (support vektor machine)	RF – 97% DT – 97%, KNN – 73%, LR – 71%, SVM – 70%, NB – 65%
[106]	sekvencie bajtov	?	CNN	94.91%
[105]	metadáta, inštrukcie	n-gramy	CNN + FFNN (dopredná neurónová sieť)	93.0%
[116]	spustiteľný súbor	obraz	RF	95.62%
[47]	sekvencie bajtov	4-gramy	Ensemble DT, LR	98.83%
[115]	byte, sekvencie .asm súboru	textové reťazce	Zarovňavanie génových sekvencií	96-98%
[61]	spustiteľný súbor	obraz	KNN, NN (neurónová sieť), SVM	96.6%, 95.6%, 94.6%
[27]	kombinácia statických dát	obrázky, n-gramy, textové reťazce	Ensemble DT	99.77% (presnosť), 0.0096% (log loss)
[22]	inštrukcie	n-gramy	SVM, NB, DT, RF	98% pre SVM
[39]	systémové volania	vektor	CNN + RNN (rekurentná neurónová sieť), FFNN, CNN	89.4%, 79.8%, 89.2%
[37]	kombinácia statických a dynamických dát	?	Latentá Dirichletova alokácia, Hierarchické Dirichletove procesy	93.4%, 90.5%

[118]	inštrukcie	control flow graf, vektor pre pageRank	pageRank + RF	72.0%
[117]	spustiteľný súbor	obraz, graf entropie	Meranie podobností histogramov	97.9%
[123]	API volania	graf	NB + algoritmus pre podobnosť grafov	93.0%
[122]	sekvencie API volaní	textové reťazce	Zarovnávanie génových sekvencií	?
[121]	sekvencie inštrukcií	obraz	vlastný algoritmus pre podobnosť okolia farebných obrázkov	98.96% a 97.32%
[120]	pakety	stavová matica	Evolučné algoritmy - XCS, UCS, GAS, SLAVE a neevolučné (DT, SVM, KNN, NB	94.53%, 99.7%, 99.19%, 84.91%, a 99.42%, 99.01%, 99.55%, 99.70% (presnosť)
[119]	API volania	graf správania sa, vektor modality	RF	92.07%
[41]	frekvencia API volaní	?	RF	66.8%
[138]	spustiteľný súbor	sekvencie n- gramov	DT, SVM, FFNN	91.25%, 96.64%, 88.31%
[49]	API volania, spustiteľný súbor	n-gramy, textové reťazce	LR, NN	89.3%, 91.5%
[6]	spustiteľný súbor, inštrukcie , PE hlavičky, stopy po vykonávaní	n-gramy	NB, KNN, SVM, DT	mnoho rôznych výsledkov
[129]	grafy volaní funkcií	graf	KNN, SVM	95.31%, 90.86%

[128]	spustiteľný súbor	n-gramy	DT, FFNN, SVM	91.25%, 88.31%, 96.64%
[3]	kombinácia statických a dynamických dát	rôzne	SVM, KNN, DT, RF	92.65%, 94.67%, 95.67%, 97.05%
[126]	inštrukcie, systémové volania	vektor (MIST formát)	PHMM (profilový skrytý Markov model)	96.4%
[125]	pakety	graf správania sa	DT	94.57%
[124]	API volania	vektor	FFNN	Najlepší priemer 95.36%
[131]	statické aj dynamické dáta	rôzne	SVM, DT, RF, KNN	mnoho rôznych výsledkov
[130]	rôzne dynamické dáta	vektor	NB, RI (rule induction), DT, KNN	79.88%, 91.30%, 92.83%, 95.30%
[134]	Sekvencie inštrukcií	n-gramy	používajú tzv. subfamily centroid vektor pre rátanie podobnosti	99.05%
[139]	spustiteľný súbor	obraz-> GIST atribúty	KNN	99.2% (presnosť)
[133]	API volania, rôzne frekvencie	textové reťazce	SVM, KNN, DT, RF	94.3%, 90%, 94.5%, 94.5%
[132]	inštrukcie, systémové volania	vektor (MIST formát), lokálne senzitívne hashe	Nearest prototype classification (alternatíva ku KNN)	98.1%
[42]	statické a dynamické dáta	obraz-> GIST atribúty a JSON-> vektor	KNN	98%, 95%
[43]	dynamické dáta	vektor	SVM	83.3%
[136]	systémové volania	graf	Hľadanie najväčšieho spoločného podgrafu	veľa rôznych výsledkov

[135]	sekvencie API volaní	textové reťazce, frekvencie	SVM, KNN, DT, RF	94.5%, 93.1%, 92.5%, 95.7%
[55]	disasemblovaný súbor, spustiteľný súbor	dĺžka funkcií, textové reťazce (PSI)	NB, SVM, KNN, DT, RF	94.31%, 97.77%, 97.63%, 98.15%, 96.69%
[59]	textové reťazce	binárny vektor	NB, SVM, KNN, DT, RF	90.4%, 95.5%, 95.4%, 95.9%, 94.8%
[137]	grafy volaní funkcií	graf	KNN, grafové algoritmy	91.3%
[97]	rôzne dynamické dáta	vektor	SVM	88.0%

2 Príloha C – nástroje na analýzu programov

Tab. 2 nástroje na statickú analýzu programov.

Názov	Údaje
pype32 [113]	PE metadáta, informácie o sekciách, zdrojoch a overlay dátach, importoch a exportoch. Umožňuje zapisovať (vytvárať sekcie, rozširovať ich, pridať priečinky, zdroje...)
Pefile [83]	PE metadáta (aj importy a exporty, informácie o sekciách a zdrojoch (resources)), textové reťazce
Pescanner [80]	PE metadáta, informácie o sekciách, importoch a zdrojoch, podozrivé importy
LIEF [84]	PE metadáta (aj importy a exporty, informácie o overlay časti a symboly)
Pev [81]	PE metadáta (aj importy a exporty, zdroje a sekcie), textové reťazce
Dumpbin [75]	PE metadáta (aj importy a exporty, priečinky a sekcie), disasemblovaný a hexadecimálny súbor
strings – linux [114]	textové reťazce
Strings – sysinternals [103]	textové reťazce
IDA, ida python [77]	Disasemblovaný súbor
Sigcheck [112]	Entropia
Radare, Rabin [79]	Disasemblovaný súbor, metadáta (aj importy a exporty, sekcie, priečinky), textové reťazce
Mastiff [76]	Metadáta (aj importy, exporty, sekcie, zdroje), textové reťazce, výstupy z Virustotal API, fuzzy hashe, Yara pravidlá, získanie súborov z resources sekcie a hexdump súboru.
multiScanner [127]	Výstupy z Virustotal API, metadáta (používa Exif, PEFile, TrID, Libmagic, Pdftinfo), importy, exporty, sekcie, zdroje, Yara pravidlá
Pyew [165]	Metadáta pre PE a ELF typy spustiteľných súborov, textové reťazce, funkcie

Generic parser [166]	Metadáta (aj informácie o sekciách aj s entropiou, importy, exporty, priečinky)
Peframe [82]	Podozrivé atribúty súboru a výstup z Virustotal API
Objdump [78]	Metadáta (aj sekcie, priečinky, importy, exporty), disasemblovanie, demangling, dump sekcií
Exiftool [70]	Metadáta pre viacero formátov spustiteľných súborov (len z hlavičky)
Hexdump [74]	Hexadecimálny súbor
Virustotal API (private) [90]	Ohodnotenia programu 70 antivírusovými programami, počet pozitívnych hodnotení, hash súboru
Virustotal API (premium) [90]	Okrem výstupu z private verzie aj výstup z exiftool (metadáta), importy, exporty, informácie o sekciách aj s entropiou, zdrojoch a overlay, výstup z IDS a sandboxu (otvorené porty, navštívené adresy, vytvorené procesy, manipulácia s registrami a súbormi, načítané knižnice).
Procmon [91]	Dynamické údaje: manipulácia so súbormi, registrami, procesmi, načítané knižnice, komunikácia cez porty.

3 Príloha D – Podrobné výsledky z analýz

V nasledujúcej tabuľke po počte atribútov nasledujú stĺpce obsahujúce najhorší a najlepší výsledok pre stromové klasifikátory, a za nimi sú stĺpce pre najhorší a najlepší výsledok pre SVM klasifikátory.

Tab. 3 Klasifikácia troch skupín atribútov pre konsenzus.

Počet atribútov	Stromy min.	Stromy max.	SVM min.	SVM max.
486136	96.29%	96.29%	95.88%	96.9%
5567	95.98%	98.96%	95.78%	98.14%
766	94.43%	98.46%	88.55%	95.17%

Tab. 4 Časy metód selekcie pri výbere 1000 atribútov z 486136.

Metóda	Čas
Chi-square	0:00:02.468936
Vzájomná informácia (MI)	0:04:38.693122
ANOVA F-skóre	0:00:04.859748
Gini skóre	0:30:12.824917
Fisher skóre	0:00:25.470692
Laplacian skóre	0:04:32.685737
SPEC	0:37:08.983870
Relieff	0:12:00.287992
XGBoost zisk (gain)	0:18:30.276268
XGBoost rozdelenie (split)	0:18:41.889669
LGBM rozdelenie	0:18:43.820059
LGBM zisk	0:19:07.769344
RFC	0:00:09.110058
SVC L1	0:03:42.932795
SGD L1	0:04:34.672729
SGD L2	0:00:46.568237

SGD elasticnet	0:02:42.066376
SVM linear	0:05:45.145463
FCBF (vybral 79)	5:31:39.259762

Tab. 5 časy metód selekcie pri výbere 1000 atribútov z 5567.

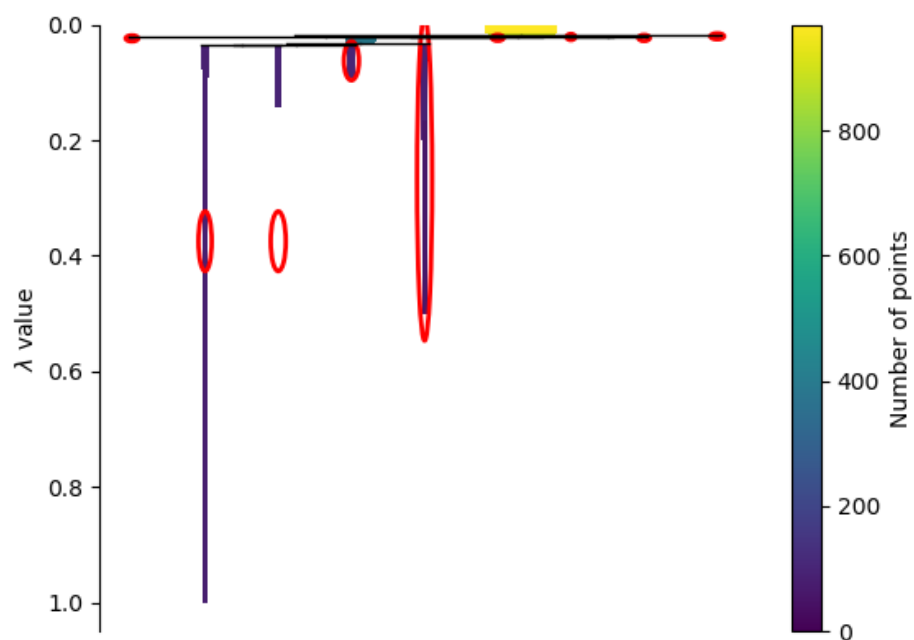
Metóda	Čas
MIFS	4:09:29.529556
mRMR	4:32:43.128295
CIFE	4:11:28.817448
JMI	5:18:14.093954
CMIM	4:30:27.683071
DISR	14:34:12.942039
Trace ratio	0:00:01.734508
CatBoost	0:07:18.986581
RGF	0:01:32.694567
Chi-square	0:00:00.015627
Vzájomná informácia (MI)	0:00:03.234622
ANOVA F-skóre	0:00:00.046879
Gini skóre	0:01:53.289886
Fisher skóre	0:00:00.328141
Laplacian skóre	0:00:03.000219
SPEC	0:00:25.970721
Relieff	0:00:10.657062
XGBoost zisk	0:00:14.079189
XGBoost rozdelenie	0:00:14.704237
LGBM rozdelenie	0:00:15.766818
LGBM zisk	0:00:15.766818
RFC	0:00:01.375096

SVC L1	0:00:07.344310
SGD L1	0:00:00.671925
SGD L2	0:00:00.125001
SGD elasticnet	0:00:00.453150
SVM linear	0:00:01.765759
SVC L2	0:01:46.148716

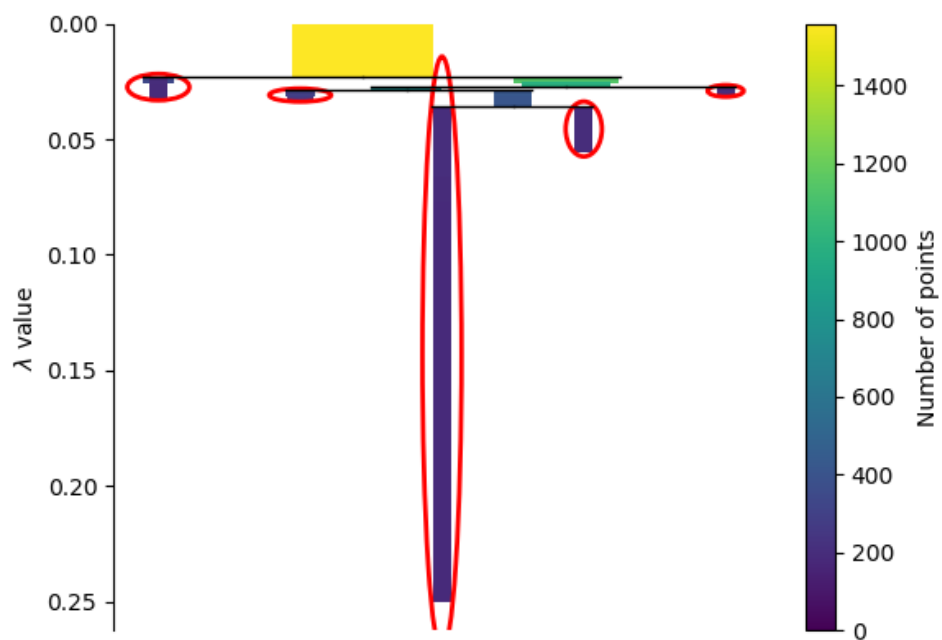
Tab. 6 Porovnanie presnosti pre zhukovanie (klastrovanie) a konsenzus na dátach bez outlierov.

Počet atribútov	Stromy min.	Stromy max.	SVM min.	SVM max.
Klastrovanie				
486136	94.95%	94.95%	96.71%	98.36%
5567	95.05%	99.45%	96.08%	98.46%
766	94.3%	98.9%	90.12%	96.5%
Konsenzus				
486136	95.64%	95.64%	95.62%	96.63%
5567	95.53%	99.01%	95.75%	97.93%
766	94.1%	98.56%	88.81%	95.52%

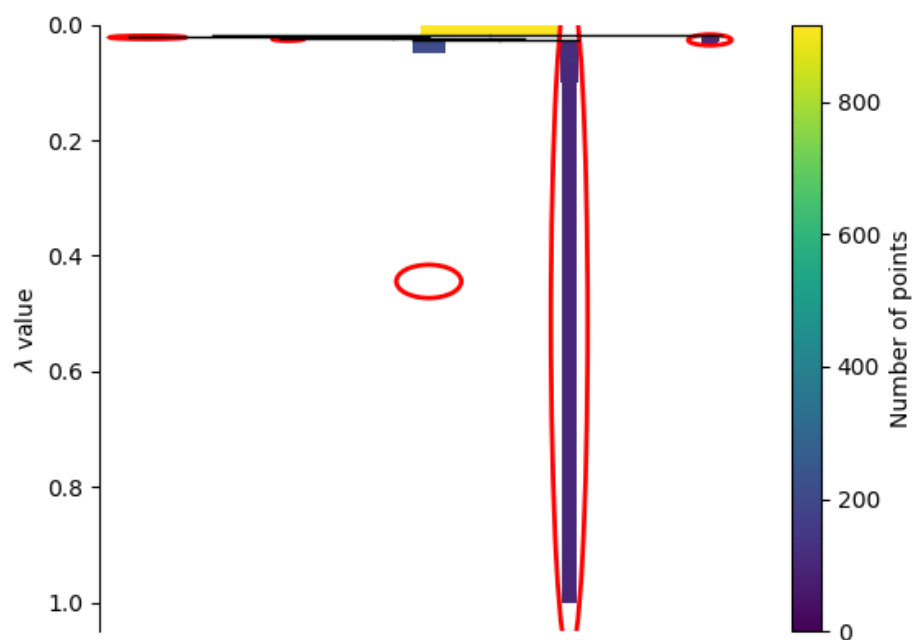
Obr. 1 Dendogram klastrov s vyznačenými vybranými triedami – prvý dataset.



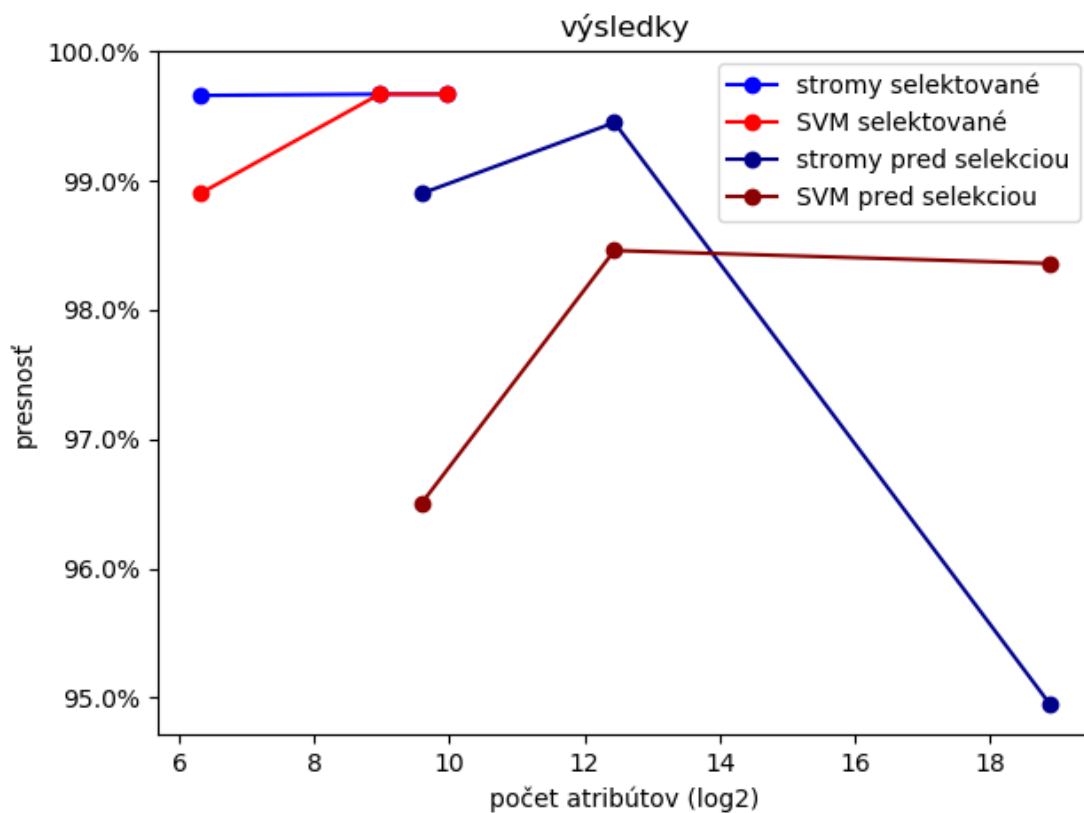
Obr. 2 Dendrogram klastrov s vyznačenými vybranými triedami – druhý dataset.



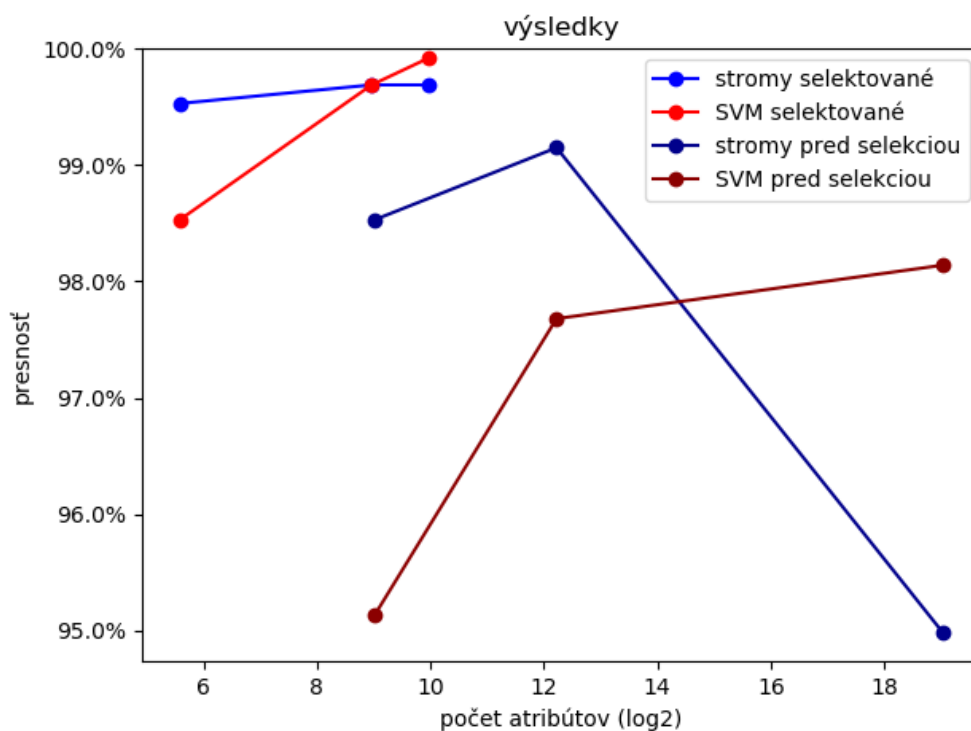
Obr. 3 Dendrogram klastrov s vyznačenými vybranými triedami – tretí dataset.



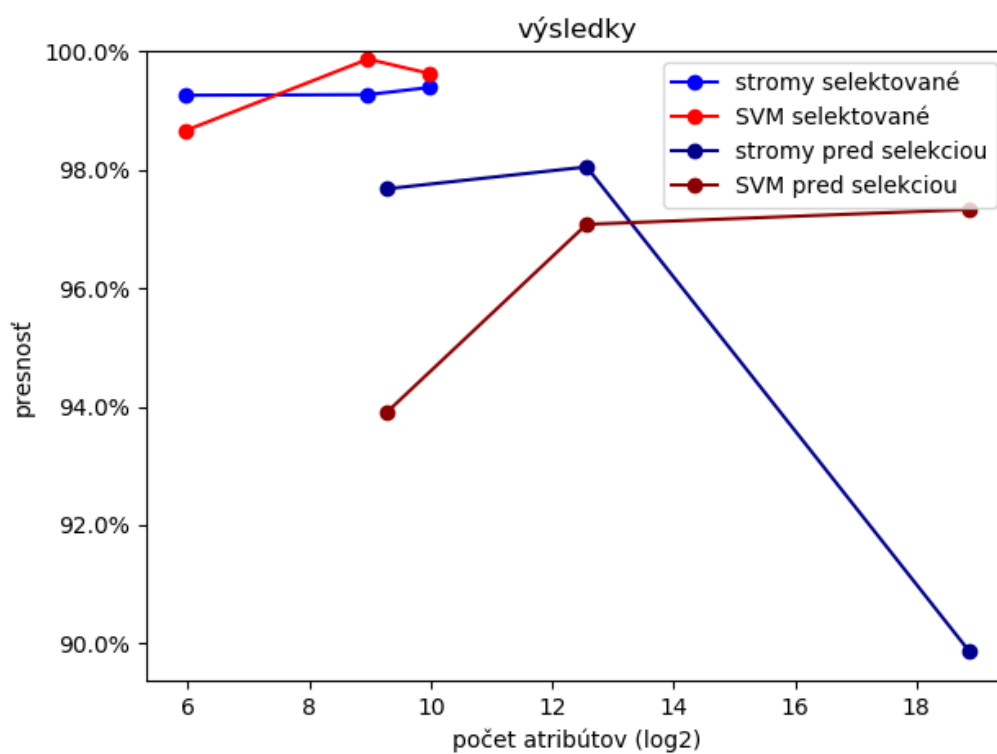
Obr. 4 Výsledky selekcie a presnosť pred selekciou – prvý dataset.



Obr. 5 Výsledky selekcie a presnosť pred selekciou – druhý dataset.



Obr. 6 Výsledky selekcie a presnosť pred selekciou – tretí dataset.



Nasledujúce tabuľky predstavujú selekcie pri označovaní konsenzom. Prvý stĺpec predstavuje použitú metódu selekcie. Ostatné stĺpce popisujú výsledky pri použití selektovaných atribútov: najhorší a najlepší výsledok pre stromové klasifikátory, a potom aj pre SVM klasifikátory. V dvojiciach stĺpcov teda sú rozsahy výsledkov ktoré sa pre jednotlivé typy klasifikátorov dosiahli.

Tab. 7 Selekcia 79 atribútov z 486136.

Metóda	Stromy min.	Stromy max.	SVM min.	SVM max.
ANOVA_F-skóre	32.89%	40.36%	39.53%	40.25%
Chi-square	94.54%	97.82%	84.31%	92.56%
FCBF	96.6%	98.66%	94.22%	98.26%
Fisher skóre	32.89%	40.25%	39.53%	40.25%
Gini	20.64%	30.97%	30.97%	30.97%
Laplacian skóre	80.51%	92.16%	36.63%	60.51%
LGBM zisk	97.32%	99.27%	95.03%	98.55%
LGBM rozdelenie	97.74%	98.95%	92.46%	96.79%
Vzájomná informácia (MI)	91.11%	95.76%	58.02%	86.98%
ReliefF	84.52%	93.27%	18.88%	67.27%
RFC	90.2%	92.78%	83.47%	88.43%
SPEC	76.8%	90.29%	48.74%	72.66%
XGBoost zisk	96.8%	99.06%	95.66%	97.22%
XGBoost rozdelenie	97.11%	99.07%	94.84%	98.46%
SGD elasticnet	77.5%	90.41%	69.61%	80.6%
SGD L1	86.79%	94.62%	84.94%	94.6%
SGD L2	50.02%	53.36%	36.12%	52.19%
SVC L1	95.02%	98.34%	91.96%	95.48%

SVM	86.77%	94%	28.24%	67.06%
pôvodne	96.29%	96.29%	95.88%	96.9%

Tab. 8 Selekcia 500 atribútov z 486136.

Metóda	Stromy min.	Stromy max.	SVM min.	SVM max.
ANOVA_F-skóre	51.33	90.15	41.38	53.27
Chi-square	96.5	98.34	85.24	97.4
Fisher skóre	50.11	89.96	41.38	53.27
Gini	76.08	94.82	59.7	78.95
Laplacian skóre	56.02	92.35	36.22	61.86
LGBM zisk	97.73	99.28	92.57	99.07
LGBM rozdelenie	97.84	99.37	92.02	98.65
Vzájomná informácia (MI)	92.26	96.99	60.23	94.43
ReliefF	90.41	96.79	37.04	86.06
RFC	97.11	98.76	90.9	96.91
SPEC	87.85	94.1	55.77	88.95
XGBoost zisk	97.73	99.48	93.7	99.28
XGBoost rozdelenie	97.73	99.48	94.01	99.28
SGD elasticnet	94.55	98.65	95.46	98.25
SGD L1	96.18	98.86	96.18	98.97
SGD L2	95.38	98.75	95.38	98.65
SVC L1	97.21	98.64	96.91	99.48
SVM	92.78	97.21	56.17	91.11
pôvodne	96.29%	96.29%	95.88%	96.9%

Tab. 9 Selekcia 1000 atribútov z 486136.

Metóda	Stromy min.	Stromy max.	SVM min.	SVM max.
ANOVA_F-skóre	75.47	94.61	42.52	80.74
Chi-square	96.70	98.44	83.48	97.73
Fisher skóre	76.09	94.51	42.52	80.74
Gini	94.46	98.44	68.47	96.29
Laplacian skóre	59.06	92.65	26.53	65.9
LGBM zisk	97.63	99.26	93.08	99.06
LGBM rozdelenie	97.42	99.16	94.02	98.66
Vzájomná informácia (MI)	92.16	97.31	57.3	94.45
ReliefF	92.68	97.32	58.72	94.94
RFC	96.6	98.85	91.62	98.05
SPEC	90.94	95.86	56.68	90.49
XGBoost zisk	97.63	99.48	95.05	99.28
XGBoost rozdelenie	97.63	99.48	94.84	99.28
SGD elasticnet	96.81	98.65	97.12	98.96
SGD L1	96.49	98.66	97.52	98.86
SGD L2	96.7	98.65	97.12	98.96
SVC L1	96.6	99.27	96.91	99.58
SVM	93.41	98.13	78.86	96.9
pôvodne	96.29%	96.29%	95.88%	96.9%

Tab. 10 Selekcia 79 atribútov z 5567.

Metóda	Stromy min.	Stromy max.	SVM min.	SVM max.
ANOVA_F-skóre	58.39	90.37	42.73	48.64
CatBoost	96.59	98.85	89.6	97.21
Chi-square	93.7	97.4	73.49	88.33

CIFE	90.41	96.17	78.86	93.31
CMIM	86.47	94.63	28.25	70.87
DISR	95.04	98.44	80.82	94.54
Fisher skóre	58.07	90.36	42.73	48.64
Gini	77.93	95.13	62.38	81.36
HSIC lasso (vybral 19)	84.21	95.13	77.47	86.78
JMI	89.48	95.55	60.58	89.78
Laplacian skóre	82.59	93.07	55.44	75.15
LGBM zisk	97.42	99.27	92.16	97.41
LGBM rozdelenie	96.69	99.17	90.08	96.18
MIFS	81.19	91.75	75.77	82.54
MRMR	94.21	98.24	85.64	92.38
Vzájomná informácia (MI)	91.21	96.39	58.41	87.39
ReliefF	85.24	92.66	17.65	67.56
RFC	96.59	98.74	82.95	95.36
RGF	97.21	99.07	94.32	98.24
SPEC	86.35	95.16	32.81	77.14
Trace ratio	57.23	90.37	42.73	48.64
XGBoost zisk	96.9	98.85	92.46	96.29
XGBoost rozdelenie	97.62	98.85	93.07	97.73
SGD elasticnet	92.69	98.24	92.77	97.51
SGD L1	92.78	98.35	92.77	97.09
SGD L2	94.12	98.13	94.21	98.15
SVC L1	92.66	98.35	94.94	98.35
SVC L2	94.83	98.55	93.07	98.34
SVM	95.86	98.35	92.58	98.13
pôvodne	95.98%	98.96%	95.78%	98.14%

Tab. 11 Selekcia 500 atribútov z 5567.

Metóda	Stromy min.	Stromy max.	SVM min.	SVM max.
ANOVA_F-skóre	94.21	98.54	76.28	92.79
CatBoost	97.11	99.06	94.11	98.45
Chi-square	93.4	97.73	58.22	93.1
CIFE	94.74	98.55	84.7	98.14
CMIM	94.95	98.34	84.05	97.62
DISR	95.24	98.75	70.21	96.49
Fisher skóre	94.01	98.54	76.28	92.79
Gini	96.18	98.74	81.84	96.49
JMI	91.65	96.9	59.19	94.13
Laplacian skóre	89.71	96.57	80.8	92.88
LGBM zisk	97.32	99.06	94.84	98.54
LGBM rozdelenie	97.21	99.17	94.95	98.35
MIFS	86.68	97.83	92.7	95.87
MRMR	95.97	98.85	92.97	98.25
Vzájomná informácia (MI)	91.86	97	60.12	93.82
ReliefF	89.28	95.87	33.64	85.77
RFC	96.38	98.96	94.01	98.04
RGF	97.94	99.17	95.67	98.65
SPEC	90.4	95.97	52.04	92.16
Trace ratio	94.01	98.43	76.28	92.79
XGBoost zisk	97.73	99.17	94.03	98.35
XGBoost rozdelenie	97.73	99.17	93.72	98.35
SGD elasticnet	95.56	98.55	96.2	98.75
SGD L1	95.66	98.54	95.34	98.34
SGD L2	95.56	98.34	96.58	98.96
SVC L1	96.08	98.96	97.32	99.38
SVC L2	95.88	98.86	96.99	99.38

SVM	95.77	98.85	95.76	99.17
pôvodne	95.98%	98.96%	95.78%	98.14%

Tab. 12 Selekcia 1000 atribútov z 5567.

Metóda	Stromy min.	Stromy max.	SVM min.	SVM max.
ANOVA_F-skóre	95.66	98.75	84.12	97.22
CatBoost	97.01	98.95	95.05	98.76
Chi-square	93.4	97.62	54.25	93.58
CIFE	95.06	98.64	89.45	98.55
CMIM	94.42	98.55	96.28	98.35
DISR	95.35	98.86	64.65	97
Fisher skóre	95.66	98.85	84.12	97.22
Gini	95.97	98.95	76.07	97.62
JMI	91.44	97.11	55.16	94.33
Laplacian skóre	90.11	97	78.57	94.83
LGBM zisk	97.11	99.06	95.16	98.45
LGBM rozdelenie	97.21	99.17	94.75	98.35
MIFS	92.45	98.54	94.85	98.25
MRMR	95.87	98.65	92.85	98.55
Vzájomná informácia (MI)	91.34	96.79	55.79	94.24
ReliefF	91.76	97.31	59.11	94.85
RFC	96.07	98.96	92.47	98.14
RGF	97.01	98.96	95.57	98.55
SPEC	92.17	97.94	63.69	93.59
Trace ratio	95.66	98.75	84.12	97.22
XGBoost zisk	97.62	99.27	93.82	98.35
XGBoost rozdelenie	97.62	99.27	93.82	98.35
SGD elasticnet	95.67	98.76	96.09	98.75

SGD L1	95.46	98.76	95.68	98.55
SGD L2	95.76	98.65	95.89	98.45
SVC L1	96.08	98.86	96.6	99.17
SVC L2	95.56	98.75	97.42	99.38
SVM	95.46	98.85	97.01	99.17
pôvodne	95.98%	98.96%	95.78%	98.14%

Tab. 13 Selekcia 79 atribútov z 766.

Metóda	Stromy min.	Stromy max.	SVM min.	SVM max.
ANOVA_F-skóre	91.84	97.62	73.12	89.19
CatBoost	94.84	97.93	86.69	93.4
Chi-square	92.56	97	71.64	89.15
CIFE	83.97	92.77	68.04	83.93
CMIM	88.45	95.56	72.06	83.99
DISR	93.09	97.21	78.42	92.36
Fisher skóre	92.15	97.62	73.12	89.19
Gini	93.6	97.62	76.61	91.63
HSIC lasso	93.07	97.81	90.92	95.35
JMI	91.75	97.19	63.19	89.67
Laplacian skóre	77.43	92.57	49.95	66.13
LGBM zisk	94.73	98.54	85.96	95.57
LGBM rozdelenie	94.84	98.46	83.92	93.39
MIFS	56.67	83.75	46.27	56.09
MRMR	93.61	97.94	82.9	93.92
Vzájomná informácia (MI)	91.54	96.78	62.77	89.88
ReliefF	92.78	97.11	71.13	88.95
RFC	94.53	98.35	85.33	94.95
RGF	95.14	98.43	88.33	95.47

SPEC	88.53	94.84	57.96	84.1
Trace ratio	92.25	97.82	73.12	89.19
XGBoost zisk	95.25	98.46	87.3	95.27
XGBoost rozdelenie	94.64	98.25	87.28	94.54
SGD elasticnet	90.32	97.41	85.78	93.61
SGD L1	86.59	95.76	67.51	81.62
SGD L2	91.22	97.41	87.73	94.42
SVC L1	91.13	97.31	92.03	95.78
SVC L2	92.67	97.43	90.8	94.63
SVM	93.18	97.39	89.26	96.3
pôvodne	94.43%	98.46%	88.55%	95.17%

Tab. 14 Selekcia 500 atribútov z 766.

Metóda	Stromy min.	Stromy max.	SVM min.	SVM max.
ANOVA_F-skóre	94.74	98.55	87.52	95.16
CatBoost	95.05	98.46	88.43	95.46
Chi-square	94.53	98.35	88.86	95.17
CIFE	91.25	97.21	87.31	94.23
CMIM	94.54	98.35	90.53	95.16
DISR	94.84	98.46	90.92	95.16
Fisher skóre	94.63	98.55	87.01	95.16
Gini	94.84	98.46	88.87	95.48
JMI	94.95	98.55	90.6	95.78
Laplacian skóre	94.95	98.25	90.93	95.38
LGBM zisk	95.05	98.44	88.63	95.79
LGBM rozdelenie	95.05	98.44	88.63	95.79
MIFS	90.32	96.7	87.93	94.32
MRMR	94.94	98.56	90.61	95.67

Vzájomná informácia (MI)	95.05	98.46	90.53	95.69
ReliefF	94.85	98.35	90.61	94.85
RFC	95.05	98.35	90.21	95.36
RGF	95.26	98.46	88.11	94.43
SPEC	94.22	98.66	87.41	95.05
Trace ratio	94.74	98.55	87.73	95.16
XGBoost zisk	95.16	98.35	86.71	94.75
XGBoost rozdelenie	95.16	98.35	86.6	94.75
SGD elasticnet	94.12	98.23	91.55	95.98
SGD L1	93.81	98.33	91.55	95.36
SGD L2	95.25	98.35	92.27	96.4
SVC L1	94.33	98.25	91.34	95.6
SVC L2	94.44	98.34	91.95	95.99
SVM	94.12	98.35	91.25	95.46
pôvodne	94.43%	98.46%	88.55%	95.17%

Nasledujúce tabuľky predstavujú selekcie pre označovanie zhľukovaním.

Tab. 15 Selekcia 79 atribútov z 486136.

Metóda	Stromy min.	Stromy max.	SVM min.	SVM max.
ANOVA_F-skóre	45.29	45.94	44.52	45.73
Chi-square	94.43	98.48	86.32	96.06
Fisher skóre	45.29	45.94	44.52	45.73
Gini	55.6	56.59	35.97	43
Laplacian skóre	43.14	86.53	29.29	33.11
LGBM zisk	97.49	99.45	95.26	98.02
LGBM rozdelenie	97.82	99.56	94.45	97.16

Vzájomná informácia (MI)	92.23	97.02	60.11	88.16
ReliefF	86.21	94.32	16.95	66.72
RFC	89.67	98.25	86.52	94.86
SPEC	77.78	91.83	50.13	74.8
XGBoost zisk	97.71	99.23	94.59	98.21
XGBoost rozdelenie	98.36	99.56	96.94	98.67
SGD elasticnet	43.95	46.63	37.95	46.59
SGD L1	47.28	54.67	43.59	53.53
SGD L2	50.09	59.71	53.22	58.99
SVC L1	96.57	98.57	96.91	98.35
SVM	86.84	95.76	42.66	83.59
pôvodne	94.95%	94.95%	96.71%	98.36%

Tab. 16 Selekcia 500 atribútov z 486136.

Metóda	Stromy min.	Stromy max.	SVM min.	SVM max.
ANOVA_F-skóre	49.01	92.67	45.4	54.87
Chi-square	96.05	98.81	88.48	97.8
Fisher skóre	49.45	92.45	45.4	54.87
Gini	74.92	96.59	53.9	82.23
Laplacian skóre	38.27	87.73	28.85	34.98
LGBM zisk	97.81	99.45	92.38	99.11
LGBM rozdelenie	97.27	99.67	94.32	98.47
Vzájomná informácia (MI)	91.8	97.58	58.59	93.98
ReliefF	89.61	96.95	39.55	87.05
RFC	95.93	99.13	93.2	98.59
SPEC	87.51	96.3	47.95	89.65

XGBoost zisk	97.92	99.56	94.39	99
XGBoost rozdelenie	97.92	99.56	94.07	99
SGD elasticnet	91.4	99.23	92.4	98.14
SGD L1	89.75	99.01	93.59	98.25
SGD L2	92.17	98.9	93.6	98.58
SVC L1	95.94	99.11	97.93	99.67
SVM	90.02	96.71	62.99	91.34
pôvodne	94.95%	94.95%	96.71%	98.36%

Tab. 17 Selekcia 1000 atribútov z 486136.

Metóda	Stromy min.	Stromy max.	SVM min.	SVM max.
ANOVA_F-skóre	79.97	95.52	45.83	83.9
Chi-square	96.38	99.01	86.17	97.48
Fisher skóre	79.76	95.4	45.83	83.9
Gini	90.28	99.01	74.96	94.62
Laplacian skóre	39.5	88.06	18.97	41.84
LGBM zisk	97.59	99.45	95.42	99.11
LGBM rozdelenie	97.27	99.67	94.64	98.78
Vzájomná informácia (MI)	91.24	97.46	70.74	94.74
ReliefF	92.53	97.59	58.76	94.61
RFC	95.72	99.12	91.59	99.12
SPEC	89.28	96.71	60.22	93.07
XGBoost zisk	97.92	99.56	94.72	98.79
XGBoost rozdelenie	97.92	99.56	94.72	98.79
SGD elasticnet	93.17	99.23	97.27	99.12
SGD L1	93.16	99.35	97.61	99.01

SGD L2	92.28	99.12	97.93	99.67
SVC L1	96.27	99.12	98.04	99.56
SVM	92.89	98.56	83.21	96.92
pôvodne	94.95%	94.95%	96.71%	98.36%

Tab. 18 Selekcia 79 atribútov z 5567.

Metóda	Stromy min.	Stromy max.	SVM min.	SVM max.
ANOVA_F-skóre	51.37	93.32	45.84	53.11
Chi-square	93.21	97.82	73.04	91.57
Fisher skóre	50.36	93.32	45.84	53.11
Gini	79.51	96.24	58.12	81.3
Laplacian skóre	69.64	87.61	48.07	65.24
LGBM zisk	97.59	99.66	94.83	98.11
LGBM rozdelenie	96.92	99.66	89.78	98.03
Vzájomná informácia (MI)	91.89	97.13	62.13	88.7
ReliefF	86.98	93.26	15.76	67.17
RFC	93.07	98.8	86.77	91.84
SPEC	86.29	95.21	33.46	74.26
XGBoost zisk	96.27	99.55	92.71	97.91
XGBoost rozdelenie	97.69	99.66	94.74	97.82
SGD elasticnet	91.04	99.22	89.78	94.59
SGD L1	91.03	98.78	87.2	95.14
SGD L2	90.49	98.78	87.11	95.14
SVC L1	93.5	98.8	96.71	98.9
SVM	95.74	98.89	92.45	98.67
pôvodne	95.05%	99.45%	96.08%	98.46%

Tab. 19 Selekcia 500 atribútov z 5567.

Metóda	Stromy min.	Stromy max.	SVM min.	SVM max.
ANOVA_F-skóre	88.6	97.71	63.42	90.7
Chi-square	93.12	98.24	57.08	94.09
Fisher skóre	89.05	97.83	63.42	90.7
Gini	93.65	99.01	84.12	96.61
Laplacian skóre	88.94	95.96	78.19	94.86
LGBM zisk	96.93	99.66	95.39	98.67
LGBM rozdelenie	96.82	99.66	94.19	98.23
Vzájomná informácia (MI)	91.36	97.81	59.57	94.42
ReliefF	89.69	96.07	41.89	87.61
RFC	95.71	99.55	94.28	98.58
SPEC	89.93	96.39	57.54	92.45
XGBoost zisk	97.16	99.55	94.52	98.25
XGBoost rozdelenie	97.16	99.55	94.08	98.25
SGD elasticnet	94.64	99.45	96.83	99.01
SGD L1	94.75	99.23	96.28	98.79
SGD L2	94.85	99.12	97.26	98.67
SVC L1	95.15	99.44	97.49	99.55
SVM	95.95	99.33	97.71	99.34
pôvodne	95.05%	99.45%	96.08%	98.46%

Tab. 20 Selekcia 1000 atribútov z 5567.

Metóda	Stromy min.	Stromy max.	SVM min.	SVM max.
ANOVA_F-skóre	93.2	99.13	82.3	96.5
Chi-square	92.67	98.13	58.87	94.18
Fisher skóre	93.42	99.02	82.3	96.5

Gini	94.18	99.45	85.06	97.39
Laplacian skóre	89.17	97.6	87.39	95.85
LGBM zisk	96.49	99.56	96.29	98.79
LGBM rozdelenie	96.28	99.66	94.44	98.47
Vzájomná informácia (MI)	91.04	97.59	66.29	94.08
ReliefF	91.47	98.24	58.98	94.29
RFC	95.27	99.55	93.42	98.46
SPEC	91.69	98.47	63.43	93.42
XGBoost zisk	97.37	99.66	93.95	98.47
XGBoost rozdelenie	97.37	99.66	94.18	98.47
SGD elasticnet	95.18	99.34	96.5	99.12
SGD L1	94.86	99.55	96.95	99.11
SGD L2	94.53	99.34	97.26	99.01
SVC L1	95.05	99.55	97.49	99.23
SVM	95.17	99.34	96.95	99.44
pôvodne	95.05%	99.45%	96.08%	98.46%

Tab. 21 Selekcia 79 atribútov z 766.

Metóda	Stromy min.	Stromy max.	SVM min.	SVM max.
ANOVA_F-skóre	91.13	97.94	72.86	88.95
Chi-square	92.1	97.7	73.59	91.23
Fisher skóre	91.04	98.26	72.86	88.95
Gini	93.64	98.46	70.74	89.16
Laplacian skóre	54.31	86.98	37.8	43.87
LGBM zisk	95.07	98.89	87.07	96.39
LGBM rozdelenie	95.07	98.9	86.05	93.89

Vzájomná informácia (MI)	92.44	97.48	70.23	91.24
ReliefF	92	97.7	78.42	89.51
RFC	94.62	99.01	84.21	95.41
SPEC	89.05	95.18	66.66	85.98
XGBoost zisk	95.29	99.01	86.6	96.6
XGBoost rozdelenie	95.28	99.11	88.29	95.83
SGD elasticnet	92.23	98.67	84.15	93.97
SGD L1	83.57	94.39	48.44	71.89
SGD L2	91.9	98.57	89.4	93.65
SVC L1	94.74	98.8	95.06	97.47
SVM	94.73	99.12	92	96.18
pôvodne	94.3%	98.9%	90.12%	96.5%

Tab. 22 Selekcia 500 atribútov z 766.

Metóda	Stromy min.	Stromy max.	SVM min.	SVM max.
ANOVA_F-skóre	94.08	99	91	96.51
Chi-square	94.41	99.01	91.23	96.27
Fisher skóre	94.31	98.9	91.32	96.51
Gini	94.74	99	88.71	96.29
Laplacian skóre	94.08	98.8	91.45	96.16
LGBM zisk	94.85	98.79	89.91	95.62
LGBM rozdelenie	94.85	98.79	89.91	95.62
Vzájomná informácia (MI)	94.31	99.22	89.47	96.72
ReliefF	94.52	98.79	91.32	96.51
RFC	94.41	99.12	91.87	96.39
SPEC	93.21	99	90.36	95.62

XGBoost zisk	94.51	98.9	88.23	95.5
XGBoost rozdelenie	94.51	98.9	88.45	95.5
SGD elasticnet	94.31	99	93.1	97.16
SGD L1	95.4	99.34	92.33	96.3
SGD L2	94.52	98.9	92.99	97.05
SVC L1	94.97	99.11	92.98	96.61
SVM	93.43	98.78	91.68	96.84
pôvodne	94.3%	98.9%	90.12%	96.5%

Nasledujúce tabuľky sú pre druhý dataset.

Tab. 23 Selekcia 48 atribútov z 536109.

Metóda	Stromy min.	Stromy max.	SVM min.	SVM max.
ANOVA_F-skóre	39.39	42.22	31.70	41.91
Chi-square	95.68	97.99	67.43	84.32
FCBF	95.84	98.29	94.20	97.60
Fisher skóre	39.39	42.22	31.70	41.91
Gini	64.51	86.22	54.37	70.77
Laplacian skóre	72.79	82.44	31.24	44.53
LGBM zisk	97.31	99.15	94.36	97.14
LGBM rozdelenie	97.84	99.30	92.58	97.83
Vzájomná informácia (MI)	90.56	95.82	60.89	81.75
ReliefF	87.85	94.04	19.57	51.81
RFC	94.44	97.76	89.19	93.59
SPEC	75.55	84.91	37.44	62.58
XGBoost zisk	97.23	98.84	92.51	95.75
XGBoost rozdelenie	97.84	99.53	94.21	97.91

SGD elasticnet	90.41	95.81	52.68	76.90
SGD L1	89.17	95.90	58.44	80.73
SGD L2	89.79	96.58	61.82	79.75
SVC L1	97.84	98.84	97.14	98.53
SVM	90.33	94.65	69.02	83.56
pôvodne	94.98%	94.98%	95.75%	98.14%

Tab. 24 Selekcia 500 atribútov z 536109.

Metóda	Stromy min.	Stromy max.	SVM min.	SVM max.
ANOVA_F-skóre	41.06	43.00	31.78	42.84
Chi-square	95.75	98.61	83.45	96.91
Fisher skóre	41.06	43.08	31.78	42.84
Gini	64.51	85.91	54.29	70.38
Laplacian skóre	72.95	82.37	31.24	44.37
LGBM zisk	98.07	99.61	97.75	99.22
LGBM rozdelenie	97.91	99.46	96.52	98.53
Vzájomná informácia (MI)	93.74	97.68	61.13	92.27
ReliefF	91.18	96.59	77.18	88.46
RFC	96.22	98.84	95.91	98.45
SPEC	80.35	93.35	60.50	79.26
XGBoost zisk	97.68	99.69	97.37	98.91
XGBoost rozdelenie	97.68	99.69	97.37	98.91
SGD elasticnet	95.82	99.53	87.77	98.37
SGD L1	95.82	99.15	92.26	98.53
SGD L2	95.66	99.22	87.23	98.76
SVC L1	97.76	99.30	97.36	99.69
SVM	96.76	98.99	97.84	99.07

pôvodne	94.98%	94.98%	95.75%	98.14%
---------	--------	--------	--------	--------

Tab. 25 Selekcia 1000 atribútov z 536109.

Metóda	Stromy min.	Stromy max.	SVM min.	SVM max.
ANOVA_F-skóre	41.06	43.00	31.78	42.84
Chi-square	94.98	98.53	80.29	96.29
Fisher skóre	41.06	43.08	31.70	42.84
Gini	64.43	85.75	53.98	70.38
Laplacian skóre	72.87	82.44	31.32	43.97
LGBM zisk	97.53	99.69	96.67	98.84
LGBM rozdelenie	97.61	99.69	96.36	98.99
Vzájomná informácia (MI)	93.51	97.91	58.60	94.13
ReliefF	92.35	97.21	76.96	92.11
RFC	96.06	99.15	96.29	98.45
SPEC	81.90	94.04	61.05	80.74
XGBoost zisk	97.92	99.61	97.45	98.99
XGBoost rozdelenie	97.92	99.61	97.45	98.99
SGD elasticnet	96.44	98.99	93.26	98.91
SGD L1	96.83	99.07	93.34	98.84
SGD L2	96.52	99.15	92.10	98.91
SVC L1	98.07	99.22	98.06	99.92
SVM	96.76	99.30	95.75	99.53
pôvodne	94.98%	94.98%	95.75%	98.14%

Tab. 26 Selekcia 48 atribútov z 4781.

Metóda	Stromy min.	Stromy max.	SVM min.	SVM max.
--------	-------------	-------------	----------	----------

ANOVA_F-skóre	73.09	88.00	43.84	60.72
Chi-square	95.83	97.83	68.32	84.24
Fisher skóre	73.55	87.69	43.84	60.72
Gini	90.11	95.51	48.16	66.14
Laplacian skóre	64.21	84.75	42.14	56.85
LGBM zisk	97.84	98.92	88.70	94.82
LGBM rozdelenie	97.69	99.22	86.01	95.37
Vzájomná informácia (MI)	90.41	95.90	56.96	81.59
ReliefF	86.70	92.80	30.59	63.74
RFC	96.06	98.76	78.32	90.64
SPEC	87.40	93.81	49.38	68.79
XGBoost zisk	97.76	98.76	84.81	93.51
XGBoost rozdelenie	98.07	99.22	88.71	95.67
SGD elasticnet	91.50	96.04	69.54	83.69
SGD L1	91.96	96.05	74.03	85.45
SGD L2	92.04	96.67	76.28	88.40
SVC L1	94.90	98.61	91.26	97.14
SVM	92.51	98.07	89.74	96.06
pôvodne	95.52%	99.15%	91.96%	97.68%

Tab. 27 Selekcia 500 atribútov z 4781.

Metóda	Stromy min.	Stromy max.	SVM min.	SVM max.
ANOVA_F-skóre	93.35	96.90	66.07	83.98
Chi-square	94.59	98.37	65.41	95.83
Fisher skóre	93.19	97.06	66.07	83.98
Gini	93.29	97.36	59.48	87.85
Laplacian skóre	75.44	90.41	70.22	79.28

LGBM zisk	96.91	99.38	90.65	96.51
LGBM rozdelenie	97.53	99.38	89.96	96.98
Vzájomná informácia (MI)	93.82	98.07	60.37	91.95
ReliefF	89.40	95.98	49.29	86.47
RFC	96.14	99.22	85.00	96.59
SPEC	89.63	97.43	64.51	90.34
XGBoost zisk	97.15	99.30	91.73	94.97
XGBoost rozdelenie	97.15	99.30	91.73	94.89
SGD elasticnet	96.21	98.92	94.74	97.83
SGD L1	94.66	98.53	94.12	97.61
SGD L2	95.67	99.07	95.52	98.53
SVC L1	96.68	98.61	97.37	98.99
SVM	93.66	98.61	95.74	98.76
pôvodne	95.52%	99.15%	91.96%	97.68%

Tab. 28 Selekcia 1000 atribútov z 4781.

Metóda	Stromy min.	Stromy max.	SVM min.	SVM max.
ANOVA_F-skóre	94.12	97.37	70.88	89.16
Chi-square	94.36	98.60	65.43	96.29
Fisher skóre	94.12	97.67	70.88	89.16
Gini	94.52	98.30	75.72	90.95
Laplacian skóre	89.64	96.05	86.01	89.79
LGBM zisk	96.14	99.22	89.80	96.67
LGBM rozdelenie	96.14	99.22	88.11	96.98
Vzájomná informácia (MI)	93.20	98.07	64.06	93.89
ReliefF	94.67	98.84	73.72	95.28

RFC	96.14	99.15	88.33	97.14
SPEC	93.20	97.99	68.70	93.59
XGBoost zisk	95.90	99.22	83.70	95.43
XGBoost rozdelenie	95.90	99.22	83.70	95.20
SGD elasticnet	96.37	99.07	96.21	98.06
SGD L1	96.06	98.99	96.21	98.06
SGD L2	96.37	99.15	96.44	98.45
SVC L1	97.06	98.92	96.83	98.83
SVM	96.29	99.14	96.13	98.53
pôvodne	95.52%	99.15%	91.96%	97.68%

Tab. 29 Selekcia 48 atribútov z 518.

Metóda	Stromy min.	Stromy max.	SVM min.	SVM max.
ANOVA_F-skóre	94.12	97.37	70.08	81.14
Chi-square	94.60	98.14	74.37	86.34
Fisher skóre	91.34	96.44	70.08	81.14
Gini	92.66	97.44	75.92	87.48
Laplacian skóre	87.80	96.04	75.07	85.62
LGBM zisk	96.29	98.68	87.78	94.66
LGBM rozdelenie	95.91	98.76	87.40	94.89
Vzájomná informácia (MI)	94.13	98.07	76.34	89.96
ReliefF	94.98	98.07	80.84	93.12
RFC	93.82	97.76	81.05	90.56
SPEC	72.24	83.83	46.53	64.29
XGBoost zisk	96.22	98.68	86.30	94.12
XGBoost rozdelenie	95.98	98.61	85.77	94.74

SGD elasticnet	93.67	97.36	81.66	88.56
SGD L1	47.74	67.09	37.13	57.91
SGD L2	95.29	97.98	88.55	93.21
SVC L1	94.83	98.45	89.26	94.82
SVM	94.83	98.22	90.73	97.06
pôvodne	95.36%	98.53%	91.97%	95.13%

Tab. 30 Selekcia 500 atribútov z 518.

Metóda	Stromy min.	Stromy max.	SVM min.	SVM max.
ANOVA_F-skóre	95.21	98.53	91.97	95.05
Chi-square	95.37	98.53	91.97	95.05
Fisher skóre	95.45	98.60	91.97	94.97
Gini	95.44	98.60	91.97	95.05
Laplacian skóre	95.06	98.45	92.27	95.13
LGBM zisk	95.06	98.45	91.81	94.66
LGBM rozdelenie	95.06	98.45	91.81	94.59
Vzájomná informácia (MI)	95.21	98.53	91.97	94.97
ReliefF	95.21	98.45	91.81	94.97
RFC	95.52	98.61	91.51	94.43
SPEC	95.21	98.60	91.73	95.12
XGBoost zisk	95.37	98.84	92.04	94.74
XGBoost rozdelenie	95.37	98.84	92.04	94.89
SGD elasticnet	95.37	98.37	91.58	95.28
SGD L1	95.21	98.53	91.89	94.89
SGD L2	95.44	98.53	91.50	95.28
SVC L1	95.06	98.53	91.81	95.12
SVM	95.21	98.53	91.81	95.13

pôvodne	95.36%	98.53%	91.97%	95.13%
---------	--------	--------	--------	--------

Nasledujúce tabuľky sú pre tretí dataset.

Tab. 31 Selekcia 62 atribútov z 478831.

Metóda	Stromy min.	Stromy max.	SVM min.	SVM max.
ANOVA_F-skóre	79.92	86.98	58.29	74.81
Chi-square	90.25	96.46	75.78	88.20
FCBF	91.83	97.08	91.72	94.40
Fisher skóre	79.57	87.45	58.29	74.81
Gini	89.39	94.27	82.51	88.92
Laplacian skóre	77.59	87.82	49.02	81.36
LGBM zisk	95.60	98.90	92.33	97.68
LGBM rozdelenie	95.22	99.26	90.26	96.70
Vzájomná informácia (MI)	92.08	95.87	56.31	86.75
ReliefF	83.68	92.54	52.53	81.78
RFC	91.59	96.34	88.08	94.63
SPEC	77.99	87.33	37.66	67.76
XGBoost zisk	94.75	98.29	90.50	96.70
XGBoost rozdelenie	95.10	99.26	91.35	97.18
SGD elasticnet	91.21	95.73	86.01	93.30
SGD L1	90.61	95.38	85.31	91.61
SGD L2	90.73	95.85	87.47	92.80
SVC L1	93.40	98.05	96.94	98.29
SVM	88.33	91.84	79.87	90.54
pôvodne	89.87%	89.87%	90.78%	97.33%

Tab. 32 Selekcia 500 atribútov z 478831.

Metóda	Stromy min.	Stromy max.	SVM min.	SVM max.
ANOVA_F-skóre	92.55	96.46	88.34	95.13
Chi-square	89.63	96.09	82.73	94.15
Fisher skóre	92.42	96.82	88.34	95.13
Gini	92.45	98.17	90.38	95.26
Laplacian skóre	86.95	93.17	58.77	90.26
LGBM zisk	93.65	99.02	87.83	98.17
LGBM rozdelenie	93.04	99.27	88.93	98.30
Vzájomná informácia (MI)	90.35	95.61	49.89	92.67
ReliefF	87.20	94.98	65.62	91.48
RFC	92.44	97.44	89.77	95.86
SPEC	85.52	92.60	42.75	81.90
XGBoost zisk	93.28	99.14	90.86	97.44
XGBoost rozdelenie	93.28	99.14	90.86	97.56
SGD elasticnet	93.41	97.68	95.62	98.79
SGD L1	92.81	98.18	96.10	98.54
SGD L2	93.05	98.18	95.73	98.42
SVC L1	93.52	98.66	96.46	99.87
SVM	93.42	98.05	95.26	97.81
pôvodne	89.87%	89.87%	90.78%	97.33%

Tab. 33 Selekcia 1000 atribútov z 478831.

Metóda	Stromy min.	Stromy max.	SVM min.	SVM max.
ANOVA_F-skóre	92.07	97.31	90.63	96.58
Chi-square	89.63	96.09	81.53	94.02
Fisher skóre	91.83	97.31	90.63	96.58
Gini	91.60	98.29	90.13	96.10

Laplacian skóre	88.53	93.66	56.71	91.35
LGBM zisk	92.91	99.39	85.28	98.17
LGBM rozdelenie	92.30	99.15	84.17	97.81
Vzájomná informácia (MI)	90.24	96.33	57.17	93.55
ReliefF	88.55	95.11	66.70	92.07
RFC	91.83	98.65	89.52	96.46
SPEC	87.22	94.02	49.94	86.87
XGBoost zisk	93.42	99.14	89.89	97.68
XGBoost rozdelenie	93.42	99.14	89.89	97.56
SGD elasticnet	93.54	98.17	95.26	98.42
SGD L1	93.05	98.17	96.47	98.18
SGD L2	93.17	98.17	95.13	98.42
SVC L1	93.53	98.53	94.88	99.63
SVM	92.68	97.93	95.51	98.06
pôvodne	89.87%	89.87%	90.78%	97.33%

Tab. 34 Selekcia 62 atribútov z 6070.

Metóda	Stromy min.	Stromy max.	SVM min.	SVM max.
ANOVA_F-skóre	41.36	42.45	36.84	42.21
Chi-square	91.33	97.07	67.67	82.19
Fisher skóre	41.36	42.57	36.84	42.21
Gini	91.84	97.31	89.55	95.01
Laplacian skóre	38.05	40.50	29.95	39.74
LGBM zisk	94.86	98.05	93.65	97.92
LGBM rozdelenie	95.10	98.41	88.93	95.85
Vzájomná informácia (MI)	92.20	96.09	58.03	86.14

ReliefF	86.34	92.66	34.89	83.10
RFC	94.40	97.33	91.12	95.38
SPEC	72.20	82.46	48.83	69.65
XGBoost zisk	95.35	97.93	88.69	96.70
XGBoost rozdelenie	95.35	98.90	90.14	96.35
SGD elasticnet	94.26	97.92	94.14	96.82
SGD L1	90.87	97.32	91.59	96.10
SGD L2	94.38	97.93	92.09	97.32
SVC L1	93.88	98.41	97.55	98.66
SVM	93.89	97.57	94.52	97.19
pôvodne	92.44%	98.05%	85.26%	97.08%

Tab. 35 Selekcia 500 atribútov z 6070.

Metóda	Stromy min.	Stromy max.	SVM min.	SVM max.
ANOVA_F-skóre	93.05	97.45	92.81	96.71
Chi-square	92.66	96.95	76.61	92.09
Fisher skóre	93.29	97.45	92.81	96.71
Gini	93.06	97.44	90.38	96.46
Laplacian skóre	86.59	95.86	83.81	94.39
LGBM zisk	94.63	98.05	86.36	97.32
LGBM rozdelenie	93.78	98.66	85.02	97.68
Vzájomná informácia (MI)	90.48	95.72	51.75	92.43
ReliefF	88.30	95.72	42.88	91.20
RFC	93.42	97.56	89.77	96.34
SPEC	85.26	93.90	48.24	78.81
XGBoost zisk	94.63	98.29	89.41	96.59

XGBoost rozdelenie	94.63	98.29	89.41	96.59
SGD elasticnet	94.38	98.05	95.49	97.93
SGD L1	94.51	98.05	95.61	97.57
SGD L2	94.14	97.81	95.86	98.05
SVC L1	93.90	98.16	96.09	98.66
SVM	93.53	98.06	95.99	98.06
pôvodne	92.44%	98.05%	85.26%	97.08%

Tab. 36 Selekcia 1000 atribútov z 6070.

Metóda	Stromy min.	Stromy max.	SVM min.	SVM max.
ANOVA_F-skóre	92.81	97.44	90.26	97.44
Chi-square	92.66	97.19	69.08	92.81
Fisher skóre	92.81	97.44	90.26	97.44
Gini	92.81	97.93	87.69	95.97
Laplacian skóre	87.80	96.84	86.48	96.35
LGBM zisk	93.90	98.42	84.88	97.92
LGBM rozdelenie	93.65	98.17	84.29	97.56
Vzájomná informácia (MI)	90.36	96.11	57.79	92.70
ReliefF	90.84	96.45	51.02	94.26
RFC	92.81	97.81	86.59	97.57
SPEC	87.32	94.40	53.82	90.48
XGBoost zisk	93.17	98.05	88.79	96.83
XGBoost rozdelenie	93.17	98.05	88.79	96.83
SGD elasticnet	93.16	97.93	95.73	98.41
SGD L1	93.90	98.05	95.25	98.06
SGD L2	93.53	97.81	96.09	98.42

SVC L1	93.29	98.05	95.36	97.93
SVM	93.54	97.93	95.01	97.82
pôvodne	92.44%	98.05%	85.26%	97.08%

Tab. 37 Selekcia 62 atribútov z 622.

Metóda	Stromy min.	Stromy max.	SVM min.	SVM max.
ANOVA_F-skóre	93.05	97.68	84.44	92.45
Chi-square	92.92	97.45	83.45	93.91
Fisher skóre	93.17	97.68	84.44	92.45
Gini	94.75	97.68	81.68	94.91
Laplacian skóre	85.41	94.64	58.83	77.83
LGBM zisk	94.02	97.92	83.93	94.40
LGBM rozdelenie	94.26	97.44	82.24	94.76
Vzájomná informácia (MI)	93.78	97.68	79.06	93.42
ReliefF	93.54	97.20	77.33	93.06
RFC	94.75	97.32	80.70	93.43
SPEC	63.76	71.52	48.58	63.26
XGBoost zisk	94.75	97.81	83.71	94.79
XGBoost rozdelenie	94.38	97.56	85.18	93.81
SGD elasticnet	93.54	97.56	90.02	92.44
SGD L1	93.66	97.56	90.11	93.29
SGD L2	94.14	97.31	87.49	92.31
SVC L1	94.26	97.68	87.36	94.66
SVM	93.78	97.44	89.63	93.91
pôvodne	93.17%	97.68%	88.3%	93.91%

Tab. 38 Selekcia 500 atribútov z 622.

Metóda	Stromy min.	Stromy max.	SVM min.	SVM max.
ANOVA_F-skóre	93.65	97.80	87.19	93.54
Chi-square	94.26	97.69	85.38	93.68
Fisher skóre	93.53	97.44	87.19	93.54
Gini	93.89	97.32	87.56	93.79
Laplacian skóre	93.65	97.68	89.40	92.57
LGBM zisk	93.41	97.80	89.40	92.57
LGBM rozdelenie	93.41	97.80	86.11	92.22
Vzájomná informácia (MI)	93.90	97.69	86.84	93.31
ReliefF	93.78	97.68	85.24	93.17
RFC	93.78	97.68	85.25	92.45
SPEC	93.78	97.44	86.10	93.91
XGBoost zisk	93.05	97.56	86.72	91.72
XGBoost rozdelenie	93.05	97.56	86.72	91.72
SGD elasticnet	94.14	97.81	89.52	94.52
SGD L1	93.89	97.68	89.63	93.91
SGD L2	93.89	97.56	89.04	94.41
SVC L1	93.30	97.68	88.42	93.78
SVM	93.89	97.68	85.86	93.43
pôvodne	93.17%	97.68%	88.3%	93.91%