Mikuš Filip

I-SUNS – Zadanie č. 2

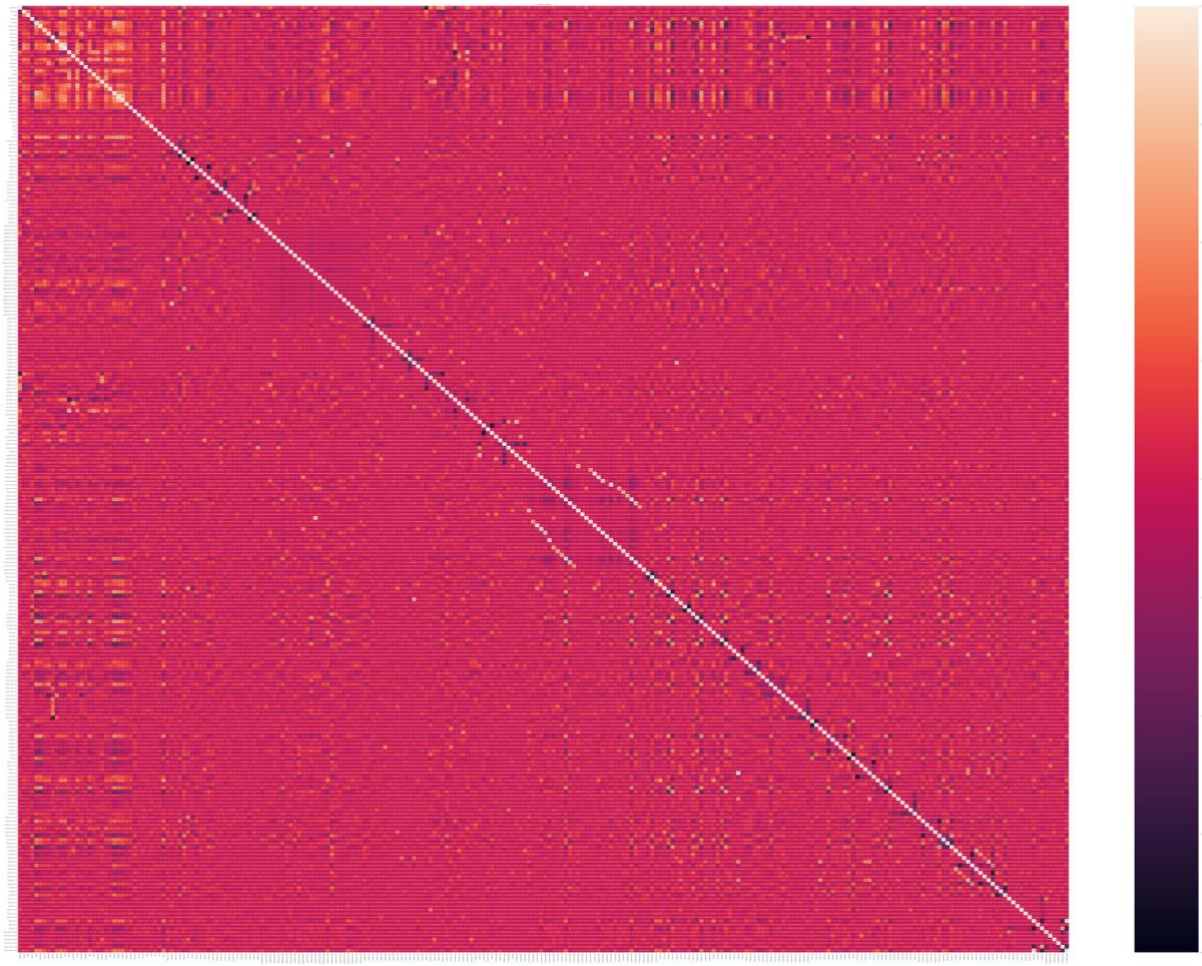
Výber datasetu

Pri trénovaní modelov boli použité ***dummy*** datasety s one hot encoding. Na normalizáciu bola použitá štandardná metóda ***MinMax.***

EDA

Korelačná matica

Akú majú závislosť jednotlivé dátové stĺpce?

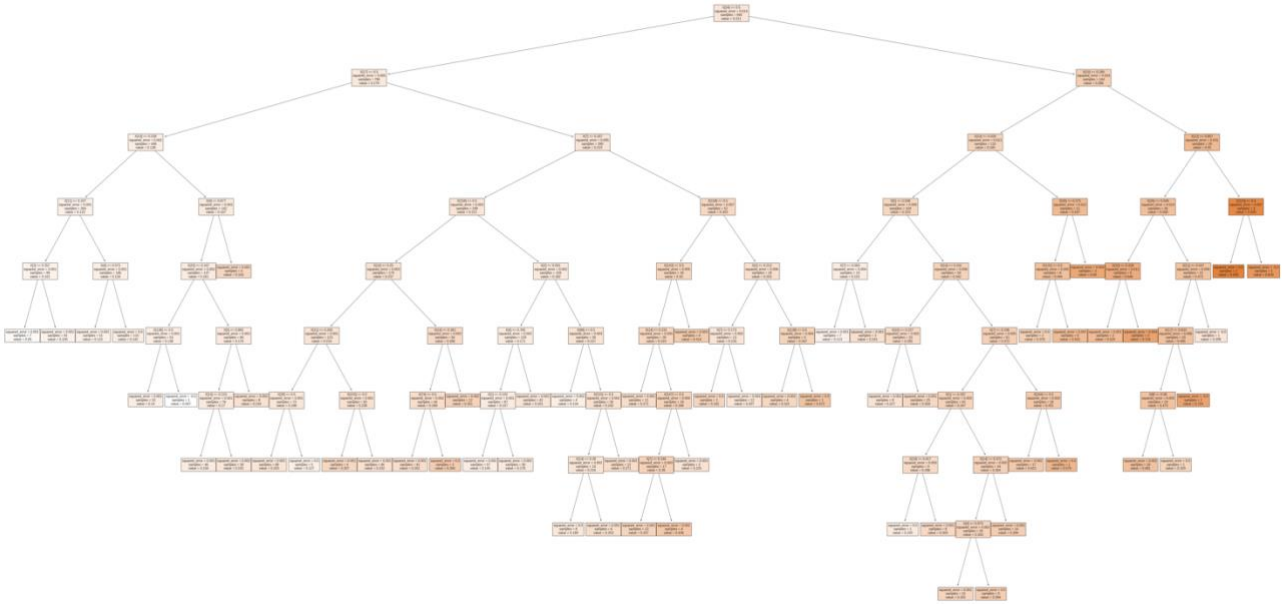


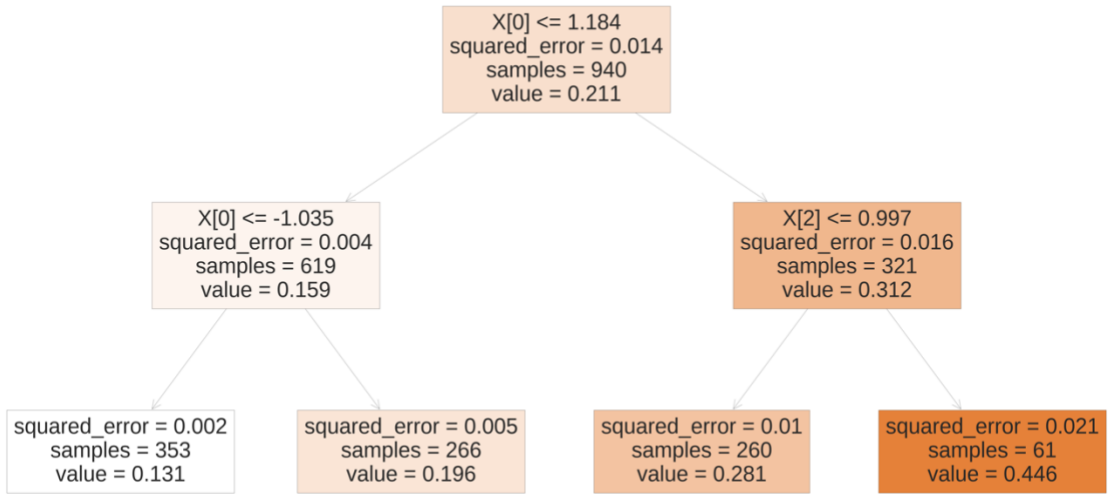
Mikuš Filip

Súbor experimentov - Rozhodovacie stromy (DT)

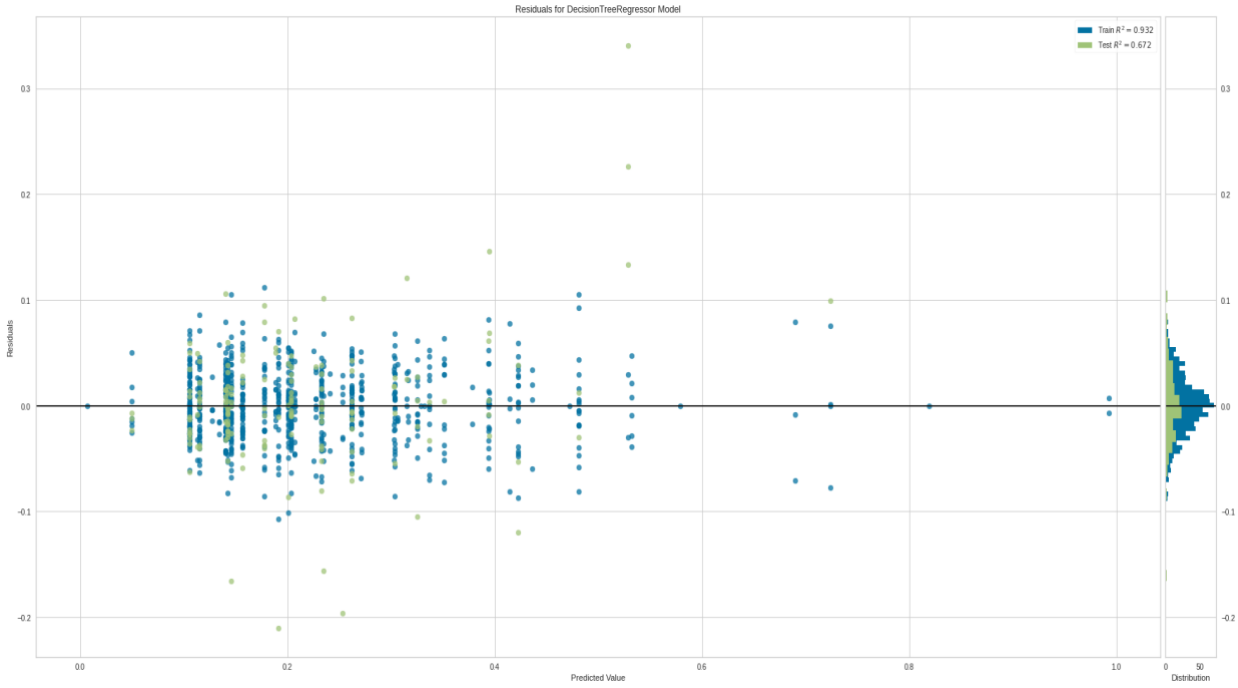
Max Depth	Max Leaf Nodes	Cross Validation	PCA Dimension Reduction	Train Time (ms)	Train MSE	Train R2	Test MSE	Test R2
default	default	-	-	-	0.000	1.000	0.005	0.606
41	55	5	-	-	0.001	0.932	0.004	0.672
default	default	-	3	-	0.000	1.000	0.007	0.378
2	None	5	3	-	0.006	0.541	0.005	0.560
41	55	5	10	15186	0.002	0.887	0.005	0.542
41	55	5	50	47340	0.001	0.910	0.007	0.366
41	55	5	100	78923	0.001	0.911	0.006	0.476
41	55	5	150	94350	0.001	0.913	0.007	0.392
41	55	5	200	132592	0.001	0.915	0.011	0.066

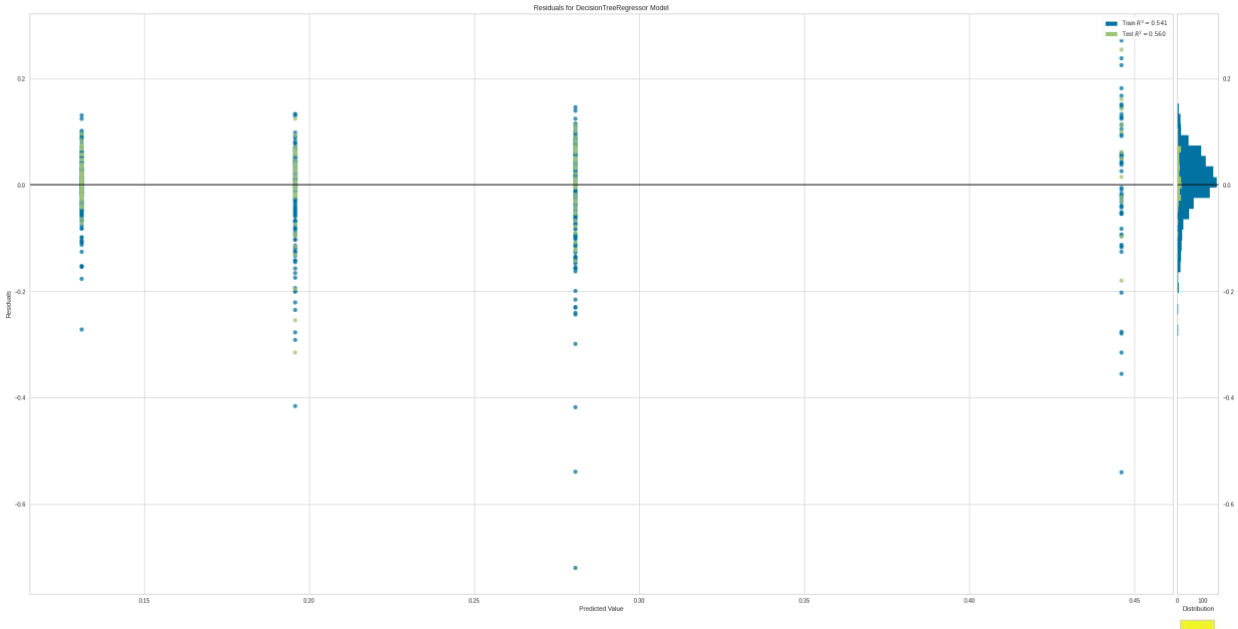
Súbor vizualizácii - Rozhodovacie stromy (DT)

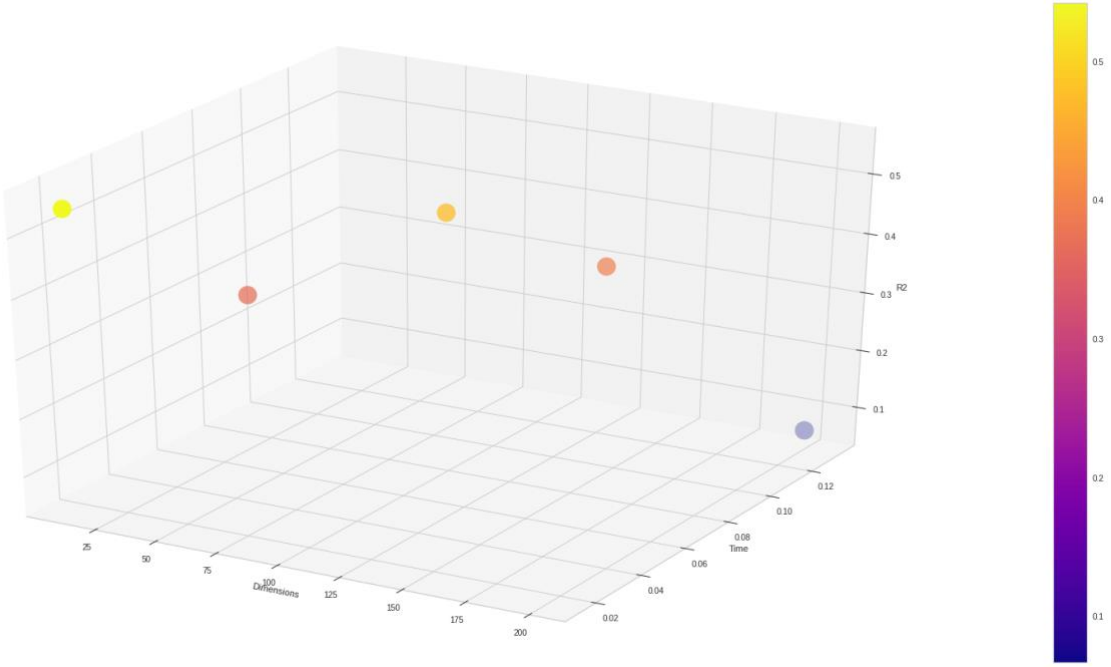




Mikuš Filip


Residuals for DecisionTreeRegressor Model
Train R^2 = 0.932
Test R^2 = 0.672
Residuals
Predicted Value
Distribution


Residuals for DecisionTreeRegressor Model
Train R^2 = 0.541
Test R^2 = 0.560
Residuals
Predicted Value
Distribution

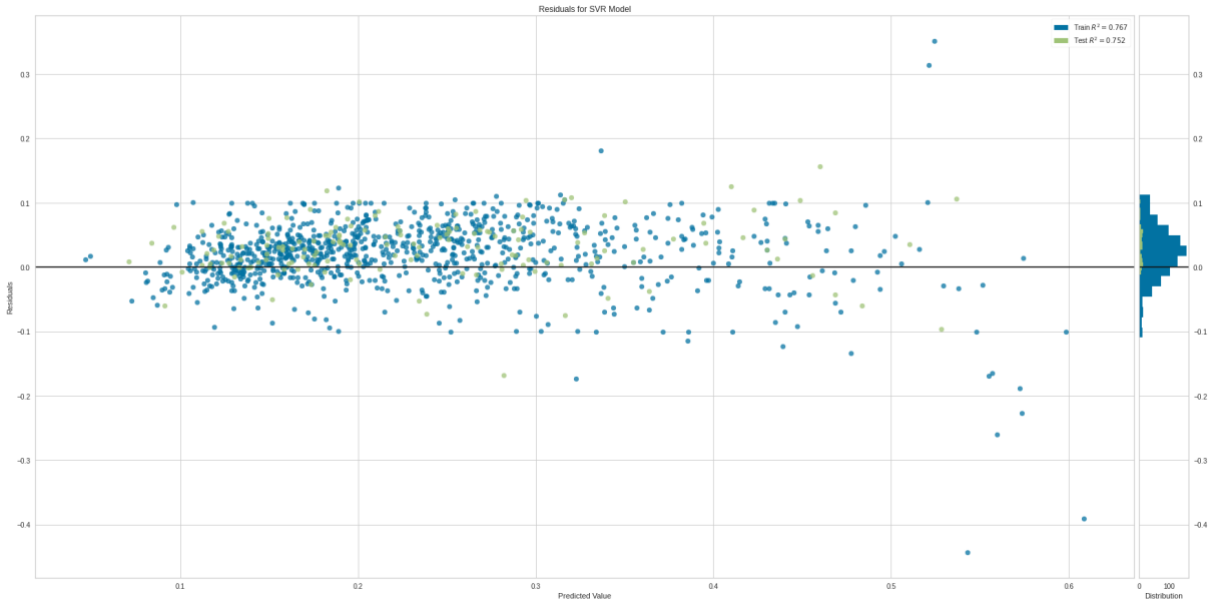

R2
Time
Dimensions

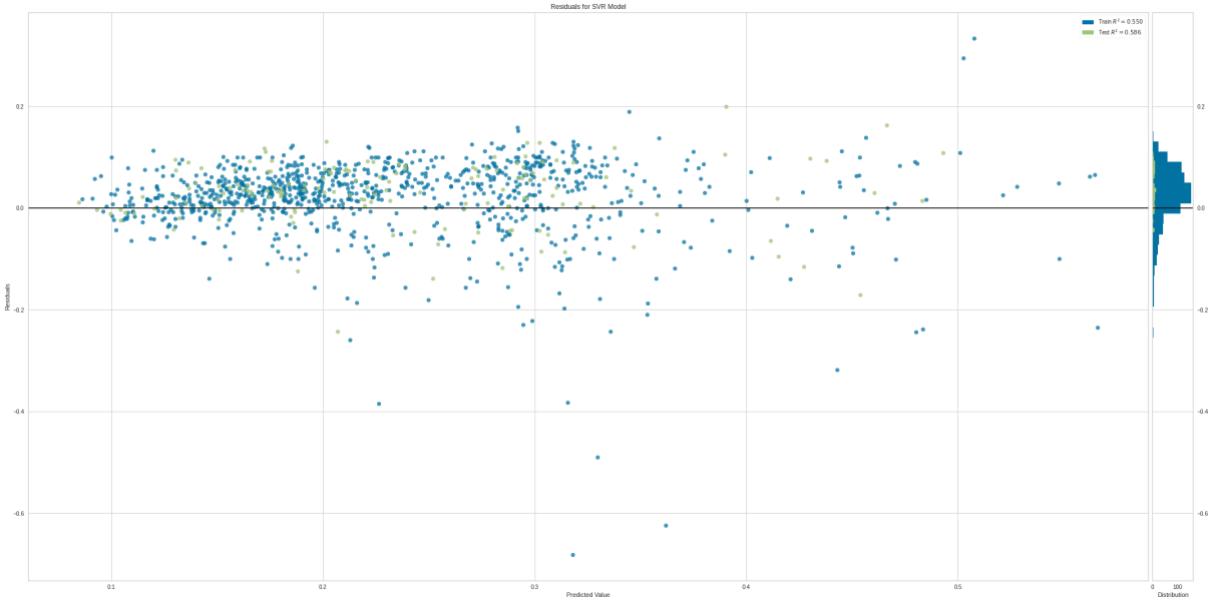
Mikuš Filip

Súbor experimentov - Stroje s podpornými vektormi (SVM)

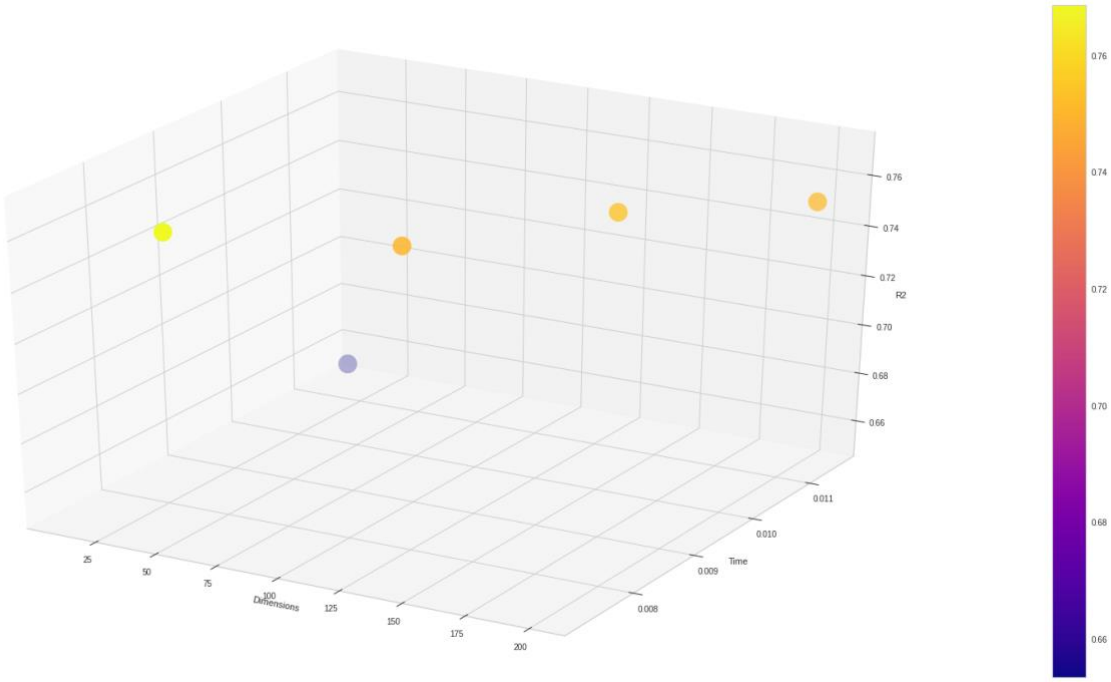
Kernel	Gamma	C	Epsilon	PCA Dimension Reduction	Train Time (ms)	Train MSE	Train R2	Test MSE	Test R2
rbf	default	default	default	-	-	0.003	0.811	0.003	0.702
linear	default	default	default	-	-	0.003	0.764	0.005	0.605
rbf	auto	default	default	-	-	0.003	0.751	0.003	0.729
rbf	auto	0.1	default	-	-	0.006	0.561	0.005	0.605
linear	scale	0.01	0.1	-	-	0.003	0.767	0.003	0.752
rbf	default	default	default	3	-	0.006	0.555	0.005	0.596
rbf	scale	0.5	0.1	3	-	0.006	0.550	0.005	0.586
linear	scale	0.01	0.1	10	11362	0.005	0.657	0.004	0.653
linear	scale	0.01	0.1	50	7219	0.003	0.758	0.002	0.769
linear	scale	0.01	0.1	100	8852	0.003	0.765	0.003	0.751
linear	scale	0.01	0.1	150	10265	0.003	0.768	0.003	0.754
linear	scale	0.01	0.1	200	11481	0.003	0.767	0.003	0.751

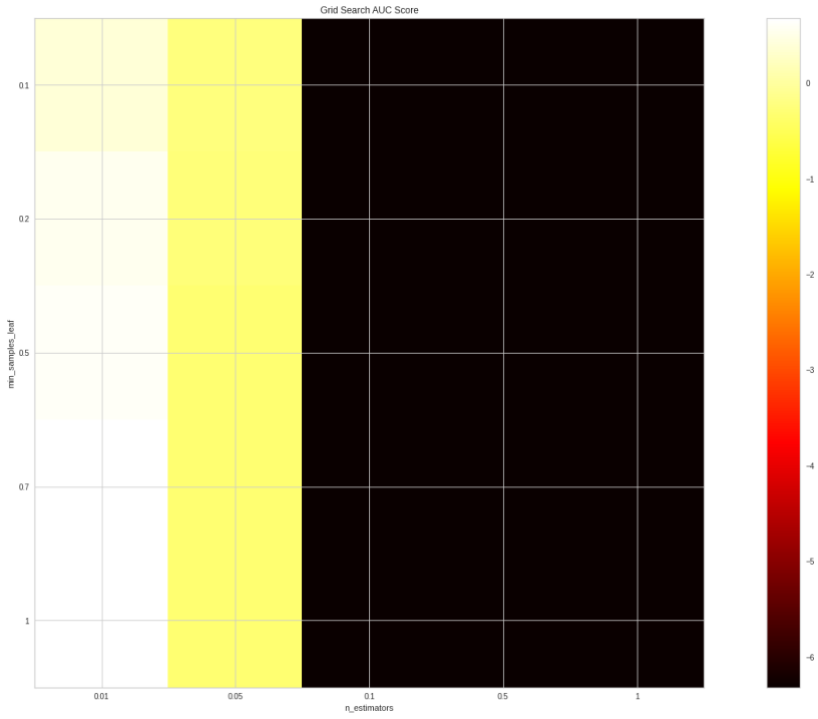
Súbor vizualizácii - Stroje s podpornými vektormi (SVM)





Mikuš Filip


0.76
0.74
0.72
R2
0.70
0.68
0.66
0.011
0.010
Time
0.009
0.008
25
50
75
100
Dimensions
125
150
175
200
0.76
0.74
0.72
0.70
0.68
0.66

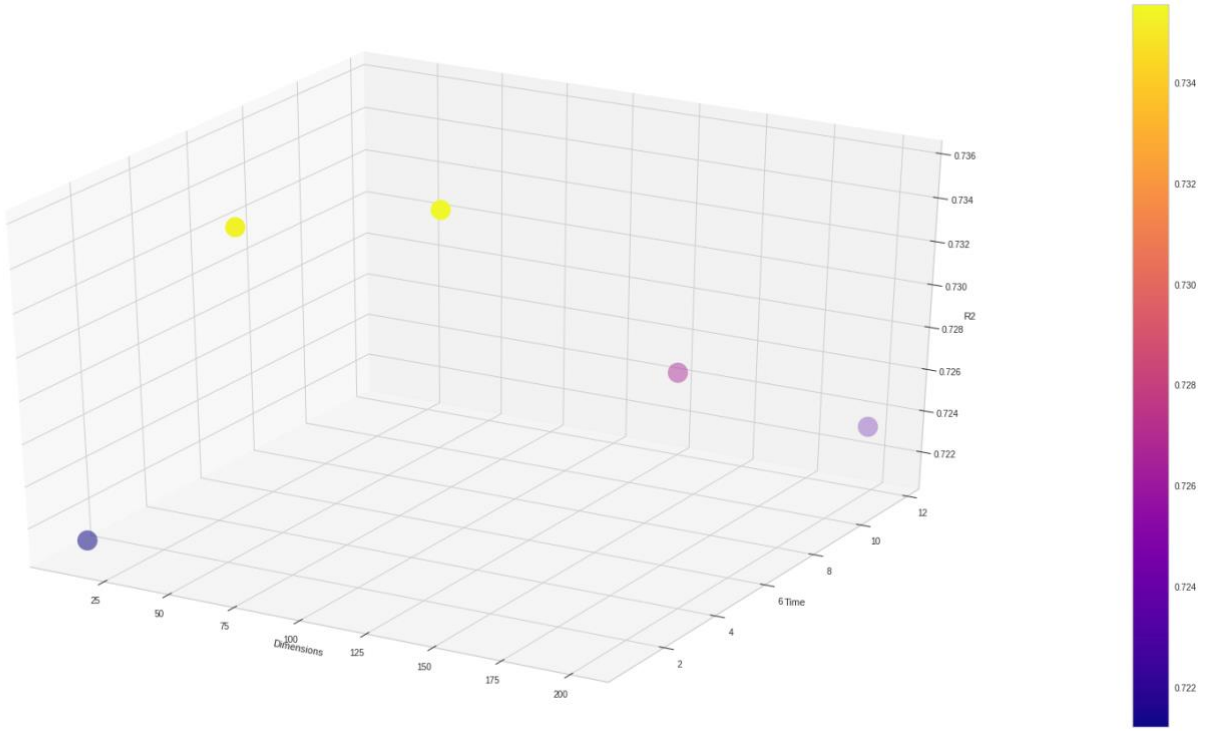

Grid Search AUC Score
min_samples_leaf
0.1
0.2
0.5
0.7
1
0.01
0.05
0.1
0.5
1
n_estimators
0
-1
-2
-3
-4
-5
-6

Mikuš Filip

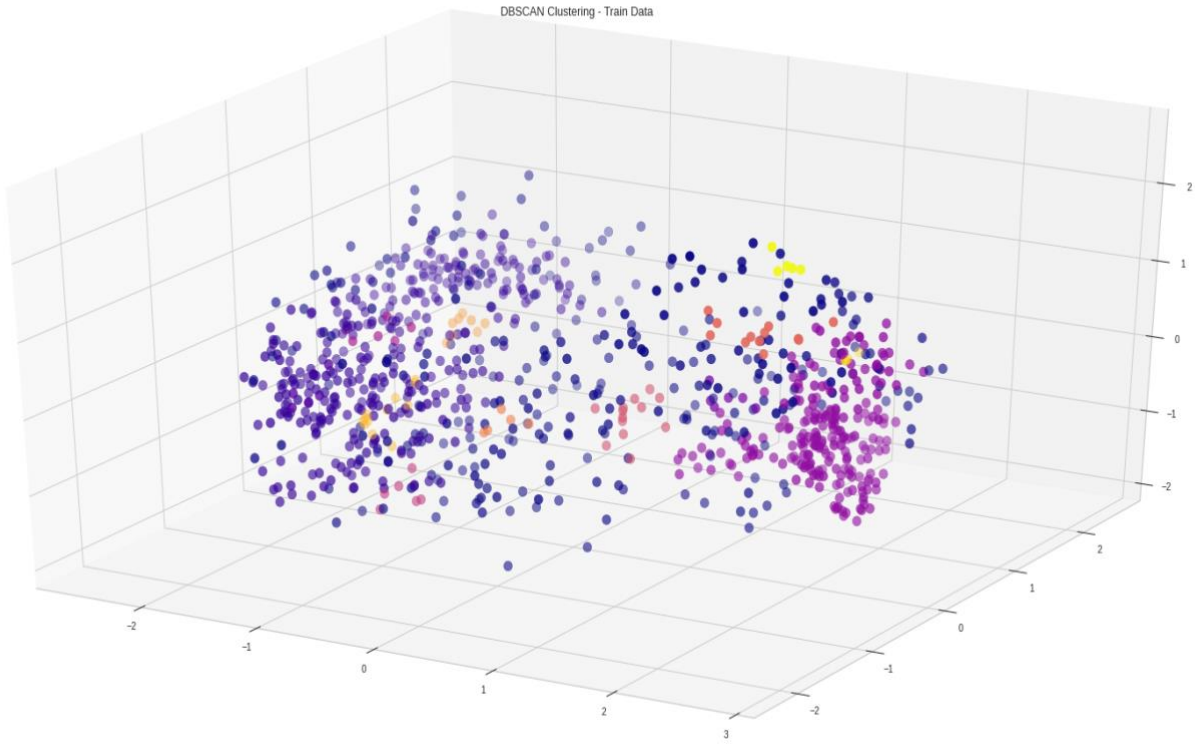
Súbor experimentov - Náhodný les (RF)

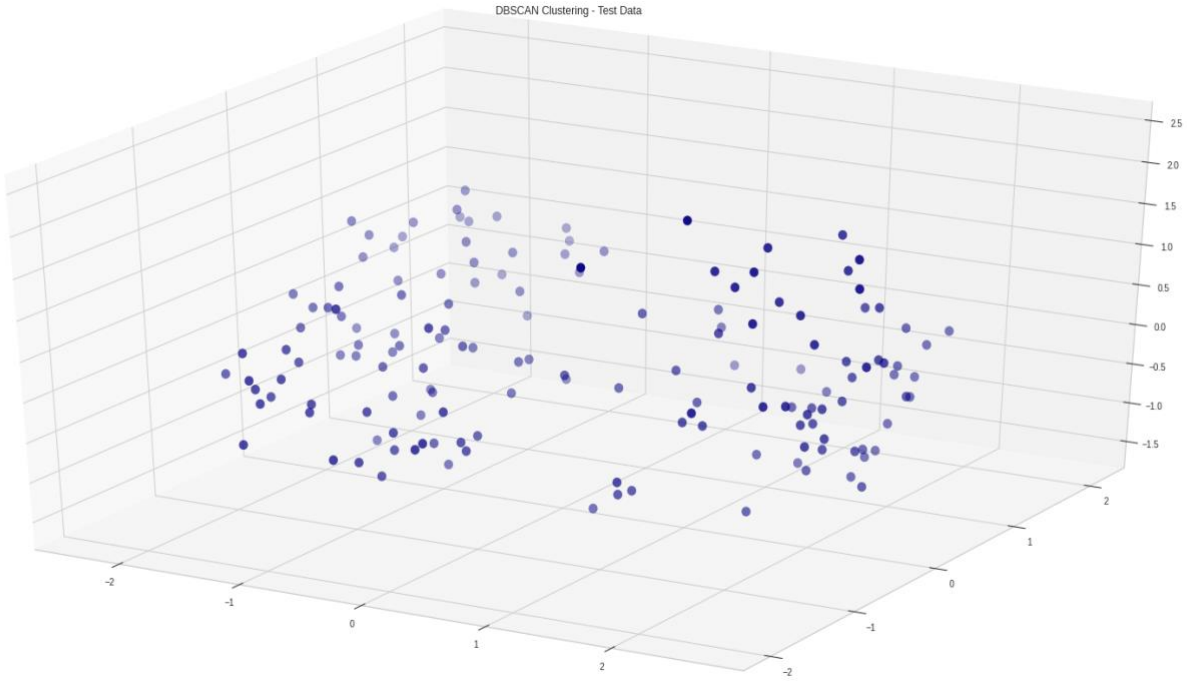
Bootstrap	Max Depth	Max Features	Min Samples Leaf	Min Samples Split	N Estimators	PCA Dimension Reduction	Train Time (ms)	Train MSE	Train R2	Test MSE	Test R2
default	default	default	default	default	default	-	-	0.003	0.980	0.002	0.863
True	110	3	3	8	100	-	-	0.004	0.741	0.003	0.758
default	default	default	default	default	default	3	-	0.001	0.933	0.004	0.627
True	100	2	3	12	100	3	-	0.003	0.763	0.004	0.630
default	default	default	default	default	default	10	684327	0.001	0.956	0.003	0.721
default	default	default	default	default	default	50	2684474	0.001	0.959	0.003	0.735
default	default	default	default	default	default	100	5185010	0.001	0.958	0.003	0.736
default	default	default	default	default	default	150	9205382	0.001	0.957	0.003	0.727
default	default	default	default	default	default	200	11703996	0.001	0.958	0.003	0.723

Súbor vizualizácii - Náhodný les (RF)

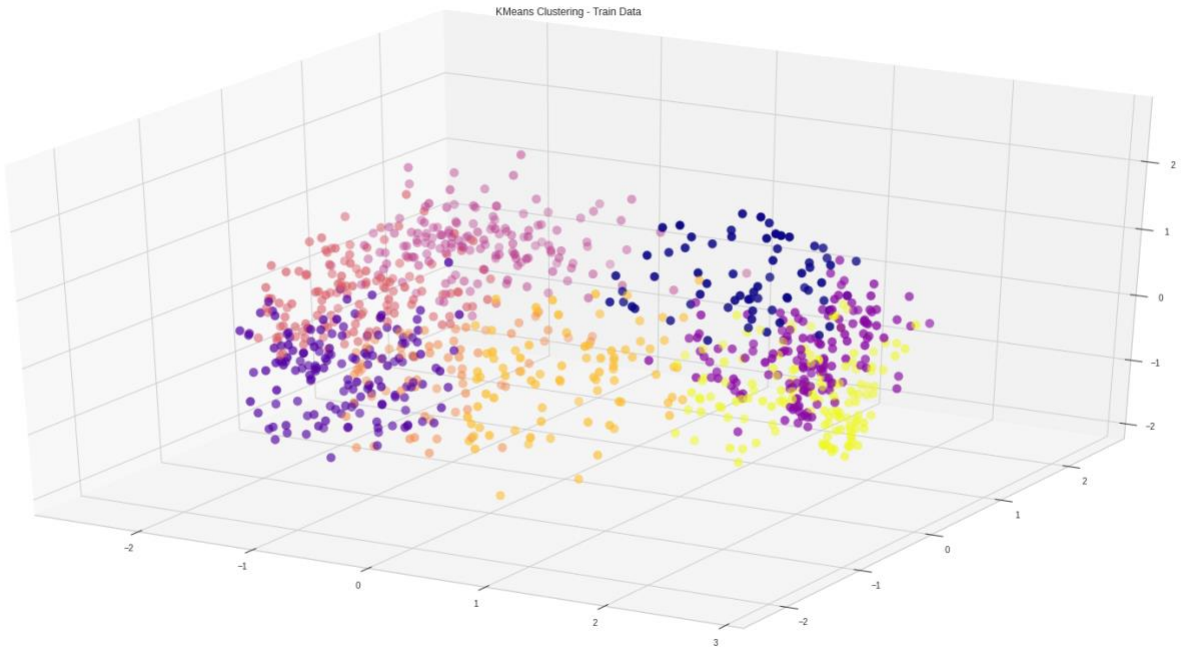


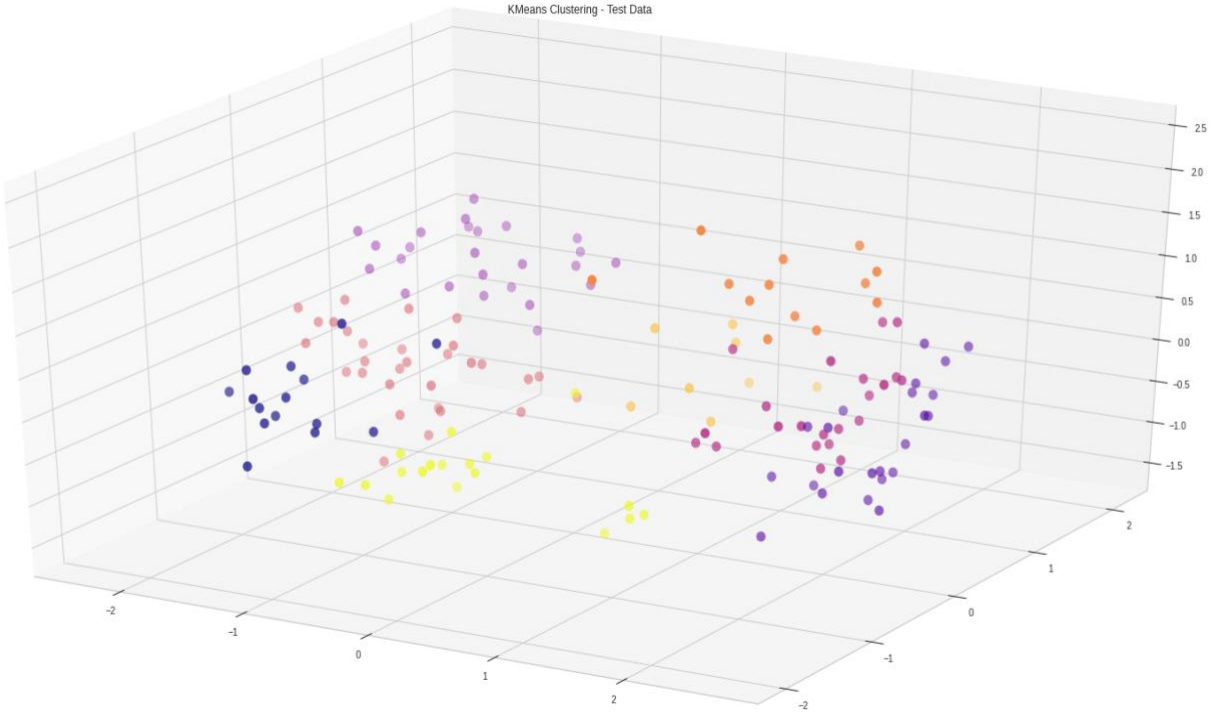
BONUS: Súbor experimentov - Zhlukovanie (DBSCAN)





BONUS: Súbor experimentov - Zhlukovanie (KMEANS)





BONUS: Súbor experimentov – Neurónová sieť (MLP-NN)

Neurons	Epochs	Activation Function	Loss Function	Optimizer	PCA Dimension Reduction	Train Time (s)	Train MSE	Train R2	Test MSE	Test R2
64-64	500	ReLu	MSE	ADAM	10	122	0.000	0.990	0.004	0.638
64-64	500	ReLu	MSE	ADAM	50	145	0.000	0.997	0.002	0.828
64-64	500	ReLu	MSE	ADAM	100	202	0.000	0.998	0.002	0.802
64-64	500	ReLu	MSE	ADAM	150	202	0.000	0.996	0.001	0.860
64-64	500	ReLu	MSE	ADAM	200	168	0.000	0.999	0.002	0.841