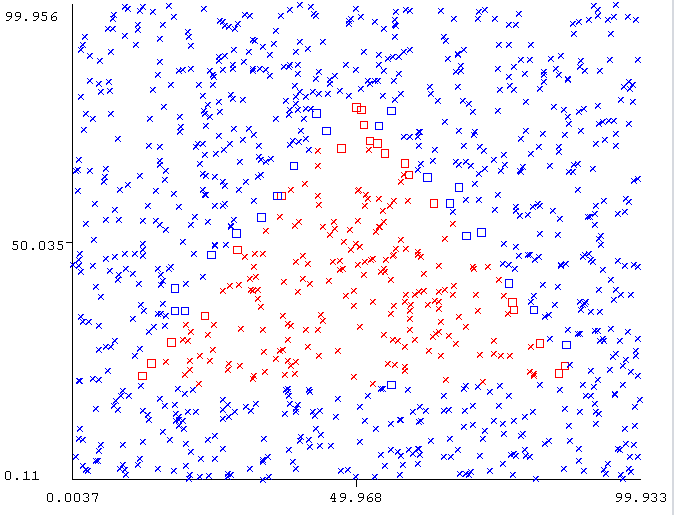
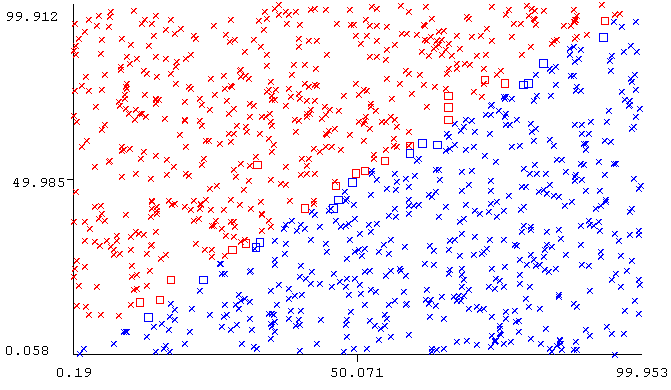
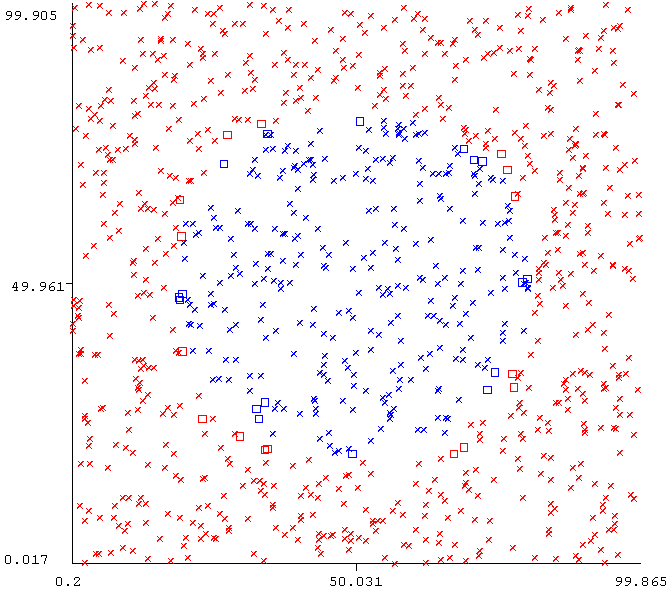
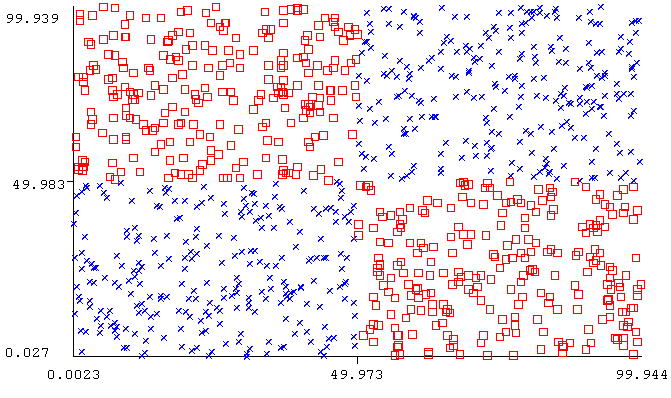
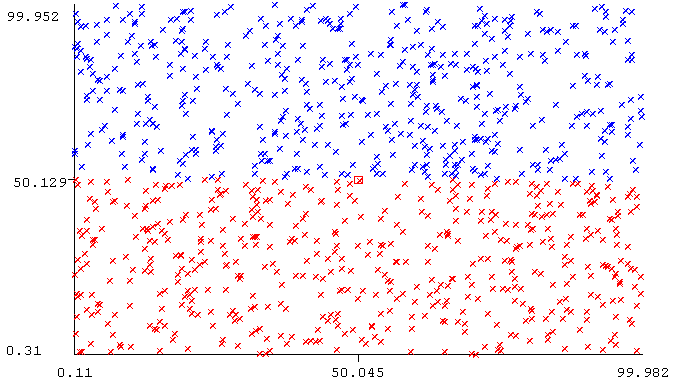
**Projekt WEKA**

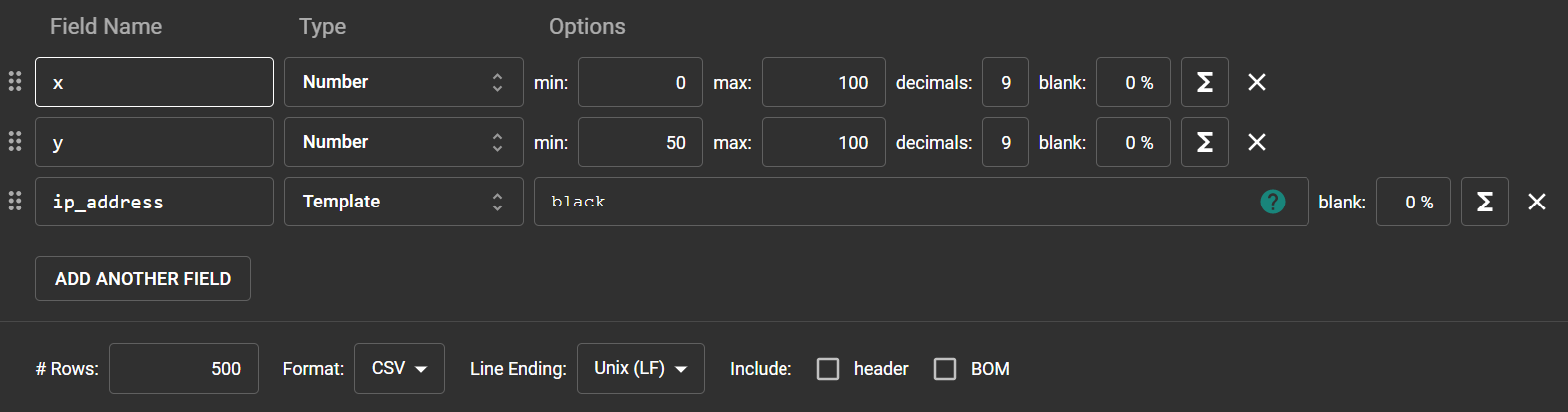
**Jméno a příjmení:** Kateřina Černá

Součástí projektu bylo dolování znalostí z dat pomocí systému WEKA. Během projektu jsem si musela vytvořit pět textových souborů umělých dat podle přede definovaných příkladů.

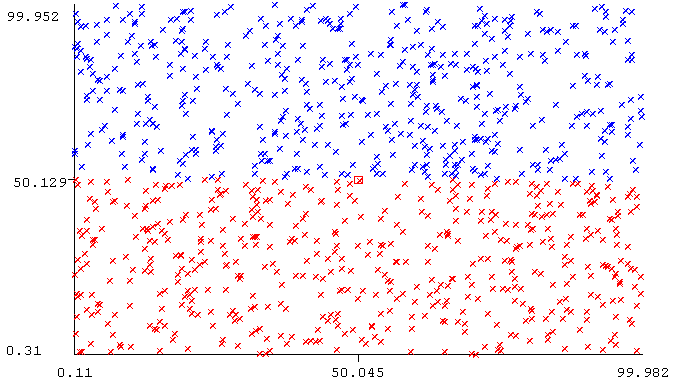


Data jsem si tvořila dvěma způsoby: generování dat pomocí webové stránky a generování dat pomocí programu Matlab.

První obrázek je plocha v polovině podle osy Y rozdělená na dvě barvy. Tyto náhodná data jsem získala pomocí webové stránky <https://www.mockaroo.com/> , kde jsem si zadala parametry podle toho, jaká data bych chtěla vygenerovat. Nejdříve jsem si dala data „black“ jako název souboru data\_black.csv a poté podobným způsobem, ale s jinými parametry „white“ jako název souboru data\_white.csv. a následně jsem tyto data spojila.



**Plochy (percentage split 10 %)**



Úroveň stromu je 2 a velikost stromu je 3.

**J48**

Correctly Classified Instances 999 99.9 %

Incorrectly Classified Instances 1 0.1 %

Kappa statistic 0.998

Mean absolute error 0.001

Root mean squared error 0.0316

Relative absolute error 0.2 %

Root relative squared error 6.3246 %

Total Number of Instances 1000

**Nativní Bayes**

Correctly Classified Instances 994 99.4 %

Incorrectly Classified Instances 6 0.6 %

Kappa statistic 0.988

Mean absolute error 0.0695

Root mean squared error 0.1432

Relative absolute error 13.9076 %

Root relative squared error 28.6356 %

Total Number of Instances 1000

**NB Tree – Multinomial**

Correctly Classified Instances 757 75.7 %

Incorrectly Classified Instances 243 24.3 %

Kappa statistic 0.514

Mean absolute error 0.2411

Root mean squared error 0.4728

Relative absolute error 48.2294 %

Root relative squared error 94.5608 %

Total Number of Instances 1000

**IB1 (1-NN)**

Correctly Classified Instances 943 97.2165 %

Incorrectly Classified Instances 27 2.7835 %

Kappa statistic 0.9443

Mean absolute error 0.0573

Root mean squared error 0.1645

Relative absolute error 11.4691 %

Root relative squared error 32.9071 %

Total Number of Instances 970

**IBk (k-NN)**

**K = 2**

Correctly Classified Instances 911 93.9175 %

Incorrectly Classified Instances 59 6.0825 %

Kappa statistic 0.8784

Mean absolute error 0.1019

Root mean squared error 0.2145

Relative absolute error 20.3858 %

Root relative squared error 42.8916 %

Total Number of Instances 970

**K = 3**

Correctly Classified Instances 928 95.6701 %

Incorrectly Classified Instances 42 4.3299 %

Kappa statistic 0.9134

Mean absolute error 0.1131

Root mean squared error 0.2129

Relative absolute error 22.6132 %

Root relative squared error 42.5729 %

Total Number of Instances 970

**K = 4**

Correctly Classified Instances 940 96.9072 %

Incorrectly Classified Instances 30 3.0928 %

Kappa statistic 0.9381

Mean absolute error 0.1157

Root mean squared error 0.2079

Relative absolute error 23.1367 %

Root relative squared error 41.5796 %

Total Number of Instances 970

**K = 5**

Correctly Classified Instances 907 93.5052 %

Incorrectly Classified Instances 63 6.4948 %

Kappa statistic 0.8701

Mean absolute error 0.1329

Root mean squared error 0.2267

Relative absolute error 26.5871 %

Root relative squared error 45.3355 %

Total Number of Instances 970

**Multilayer Perception**

Correctly Classified Instances 937 96.5979 %

Incorrectly Classified Instances 33 3.4021 %

Kappa statistic 0.932

Mean absolute error 0.0576

Root mean squared error 0.1478

Relative absolute error 11.5282 %

Root relative squared error 29.5652 %

Total Number of Instances 970

**PART**

Correctly Classified Instances 954 98.3505 %

Incorrectly Classified Instances 16 1.6495 %

Kappa statistic 0.967

Mean absolute error 0.0165

Root mean squared error 0.1284

Relative absolute error 3.299 %

Root relative squared error 25.6865 %

Total Number of Instances 970

**KStar**

Correctly Classified Instances 943 97.2165 %

Incorrectly Classified Instances 27 2.7835 %

Kappa statistic 0.9443

Mean absolute error 0.069

Root mean squared error 0.1532

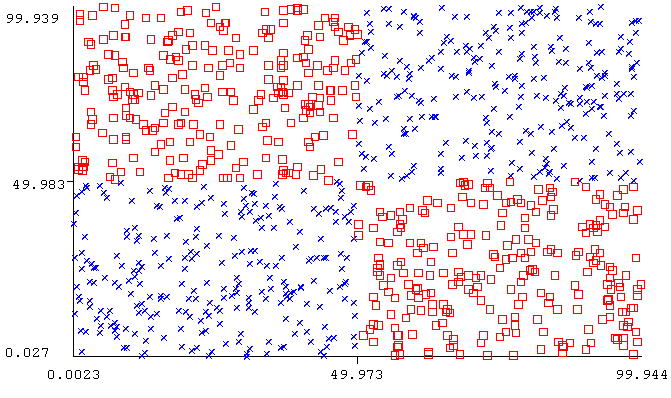
Relative absolute error 13.7957 %

Root relative squared error 30.6465 %

Total Number of Instances 970

Tímto obdobným způsobem jsem vygenerovala data pro obrázek tile.

**Dlaždice – tile (percentage split 66 %)**



Úroveň stromu je 1 a velikost stromu je 1.

**J48**

Correctly Classified Instances 157 46.1765 %

Incorrectly Classified Instances 183 53.8235 %

Kappa statistic 0

Mean absolute error 0.5015

Root mean squared error 0.5019

Relative absolute error 100.0009 %

Root relative squared error 100.0014 %

Total Number of Instances 340

**Nativní Bayes**

Correctly Classified Instances 154 45.2941 %

Incorrectly Classified Instances 186 54.7059 %

Kappa statistic -0.0177

Mean absolute error 0.5022

Root mean squared error 0.5027

Relative absolute error 100.1304 %

Root relative squared error 100.1721 %

Total Number of Instances 340

**NB Tree – Multinomial**

Correctly Classified Instances 191 56.1765 %

Incorrectly Classified Instances 149 43.8235 %

Kappa statistic 0.1375

Mean absolute error 0.5043

Root mean squared error 0.5192

Relative absolute error 100.5653 %

Root relative squared error 103.4406 %

Total Number of Instances 340

**IB1 (1-NN)**

Correctly Classified Instances 335 98.5294 %

Incorrectly Classified Instances 5 1.4706 %

Kappa statistic 0.9704

Mean absolute error 0.0162

Root mean squared error 0.1211

Relative absolute error 3.2247 %

Root relative squared error 24.1279 %

Total Number of Instances 340

**IBk (k-NN)**

**K = 2**

Correctly Classified Instances 334 98.2353 %

Incorrectly Classified Instances 6 1.7647 %

Kappa statistic 0.9646

Mean absolute error 0.0228

Root mean squared error 0.1242

Relative absolute error 4.5427 %

Root relative squared error 24.749 %

Total Number of Instances 340

**K = 3**

Correctly Classified Instances 336 98.8235 %

Incorrectly Classified Instances 4 1.1765 %

Kappa statistic 0.9764

Mean absolute error 0.0318

Root mean squared error 0.1199

Relative absolute error 6.35 %

Root relative squared error 23.8949 %

Total Number of Instances 340

**K = 4**

Correctly Classified Instances 332 97.6471 %

Incorrectly Classified Instances 8 2.3529 %

Kappa statistic 0.9528

Mean absolute error 0.0356

Root mean squared error 0.1272

Relative absolute error 7.1078 %

Root relative squared error 25.3437 %

Total Number of Instances 340

**K = 5**

Correctly Classified Instances 335 98.5294 %

Incorrectly Classified Instances 5 1.4706 %

Kappa statistic 0.9705

Mean absolute error 0.0415

Root mean squared error 0.132

Relative absolute error 8.266 %

Root relative squared error 26.2946 %

Total Number of Instances 340

**Multilayer Perception**

Correctly Classified Instances 285 83.8235 %

Incorrectly Classified Instances 55 16.1765 %

Kappa statistic 0.6735

Mean absolute error 0.2371

Root mean squared error 0.3331

Relative absolute error 47.2818 %

Root relative squared error 66.3646 %

Total Number of Instances 340

**PART**

Correctly Classified Instances 157 46.1765 %

Incorrectly Classified Instances 183 53.8235 %

Kappa statistic 0

Mean absolute error 0.5015

Root mean squared error 0.5019

Relative absolute error 100.0009 %

Root relative squared error 100.0014 %

Total Number of Instances 340

**KStar**

Correctly Classified Instances 908 93.6082 %

Incorrectly Classified Instances 62 6.3918 %

Kappa statistic 0.8722

Mean absolute error 0.1331

Root mean squared error 0.2295

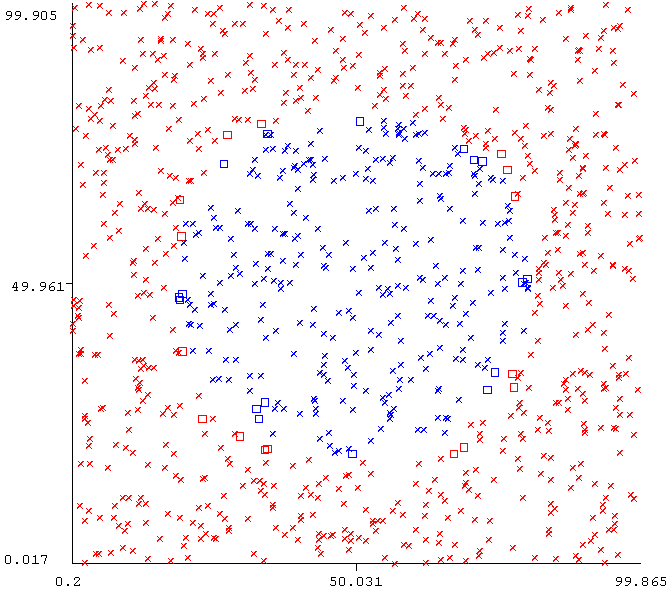
Relative absolute error 26.624 %

Root relative squared error 45.9 %

Total Number of Instances 970

Zbytek dat jsem generovala pomocí specifických příkazů v programu Matlab, Tuto tvorbu dat / konkrétního kódu jsme provedli ve skupině s Janem Mrákavou, Ondřejem Šabrňákem, Martinem Patočkou, Adamem Přikrylem a Michaelou Mikuláškovou. Veškeré funkce pro tvorbu jsou přiloženy k tomuto protokolu. Zde jsou výstupy a příloze jsou kódy, pomocí kterých jsem tyto data získala.

**Kruh (percentage split 66 %)**



**J48**

Correctly Classified Instances 331 97.3529 %

Incorrectly Classified Instances 9 2.6471 %

Kappa statistic 0.9372

Mean absolute error 0.0286

Root mean squared error 0.1597

Relative absolute error 6.7645 %

Root relative squared error 35.011 %

Total Number of Instances 340

**Nativní Bayes**

Correctly Classified Instances 277 81.4706 %

Incorrectly Classified Instances 63 18.5294 %

Kappa statistic 0.4533

Mean absolute error 0.2303

Root mean squared error 0.3144

Relative absolute error 54.3883 %

Root relative squared error 68.9394 %

Total Number of Instances 340

**NB Tree – Multinomial**

Correctly Classified Instances 170 50 %

Incorrectly Classified Instances 170 50 %

Kappa statistic -0.2403

Mean absolute error 0.4641

Root mean squared error 0.5427

Relative absolute error 109.5857 %

Root relative squared error 118.9974 %

Total Number of Instances 340

**IB1 (1-NN)**

Correctly Classified Instances 338 99.4118 %

Incorrectly Classified Instances 2 0.5882 %

Kappa statistic 0.9859

Mean absolute error 0.0074

Root mean squared error 0.0766

Relative absolute error 1.7415 %

Root relative squared error 16.794 %

Total Number of Instances 340

**IBk (k-NN)**

**K = 2**

Correctly Classified Instances 333 97.9412 %

Incorrectly Classified Instances 7 2.0588 %

Kappa statistic 0.9514

Mean absolute error 0.0228

Root mean squared error 0.1182

Relative absolute error 5.3795 %

Root relative squared error 25.9077 %

Total Number of Instances 340

**K = 3**

Correctly Classified Instances 335 98.5294 %

Incorrectly Classified Instances 5 1.4706 %

Kappa statistic 0.9647

Mean absolute error 0.024

Root mean squared error 0.1054

Relative absolute error 5.6696 %

Root relative squared error 23.1133 %

Total Number of Instances 340

**K = 4**

Correctly Classified Instances 331 97.3529 %

Incorrectly Classified Instances 9 2.6471 %

Kappa statistic 0.9372

Mean absolute error 0.029

Root mean squared error 0.1158

Relative absolute error 6.8557 %

Root relative squared error 25.4003 %

Total Number of Instances 340

**K = 5**

Correctly Classified Instances 332 97.6471 %

Incorrectly Classified Instances 8 2.3529 %

Kappa statistic 0.9437

Mean absolute error 0.0356

Root mean squared error 0.1302

Relative absolute error 8.4005 %

Root relative squared error 28.5388 %

Total Number of Instances 340

**Multilayer Perception**

Correctly Classified Instances 236 69.4118 %

Incorrectly Classified Instances 104 30.5882 %

Kappa statistic 0.3999

Mean absolute error 0.3187

Root mean squared error 0.4085

Relative absolute error 75.2573 %

Root relative squared error 89.5591 %

Total Number of Instances 340

**PART**

Correctly Classified Instances 331 97.3529 %

Incorrectly Classified Instances 9 2.6471 %

Kappa statistic 0.9372

Mean absolute error 0.0312

Root mean squared error 0.1597

Relative absolute error 7.3666 %

Root relative squared error 35.0053 %

Total Number of Instances 340

**KStar**

Correctly Classified Instances 335 98.5294 %

Incorrectly Classified Instances 5 1.4706 %

Kappa statistic 0.9645

Mean absolute error 0.1252

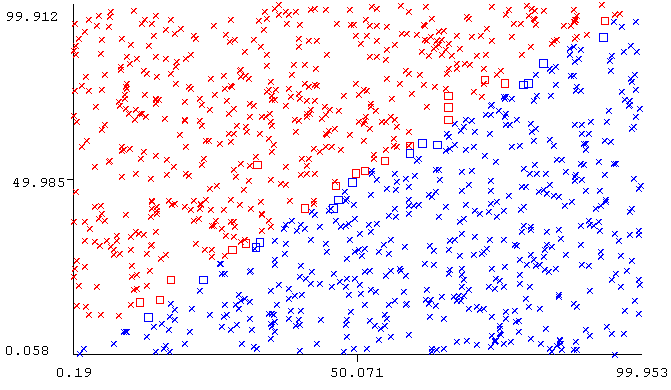
Root mean squared error 0.1811

Relative absolute error 29.5661 %

Root relative squared error 39.7092 %

Total Number of Instances 340

**Šikmé plochy (percentage split 66 %)**



**J48**

Correctly Classified Instances 326 95.8824 %

Incorrectly Classified Instances 14 4.1176 %

Kappa statistic 0.9175

Mean absolute error 0.0427

Root mean squared error 0.1942

Relative absolute error 8.5082 %

Root relative squared error 38.6953 %

Total Number of Instances 340

**Nativní Bayes**

Correctly Classified Instances 331 97.3529 %

Incorrectly Classified Instances 9 2.6471 %

Kappa statistic 0.947

Mean absolute error 0.1875

Root mean squared error 0.2432

Relative absolute error 37.4035 %

Root relative squared error 48.4561 %

Total Number of Instances 340

**NB Tree – Multinomial**

Correctly Classified Instances 338 99.4118 %

Incorrectly Classified Instances 2 0.5882 %

Kappa statistic 0.9882

Mean absolute error 0.023

Root mean squared error 0.0844

Relative absolute error 4.5849 %

Root relative squared error 16.8109 %

Total Number of Instances 340

**IB1 (1-NN)**

Correctly Classified Instances 332 97.6471 %

Incorrectly Classified Instances 8 2.3529 %

Kappa statistic 0.9528

Mean absolute error 0.025

Root mean squared error 0.1532

Relative absolute error 4.9805 %

Root relative squared error 30.5212 %

Total Number of Instances 340

**IBk (k-NN)**

**K = 2**

Correctly Classified Instances 328 96.4706 %

Incorrectly Classified Instances 12 3.5294 %

Kappa statistic 0.9294

Mean absolute error 0.0331

Root mean squared error 0.1485

Relative absolute error 6.5945 %

Root relative squared error 29.5838 %

Total Number of Instances 340

**K = 3**

Correctly Classified Instances 329 96.7647 %

Incorrectly Classified Instances 11 3.2353 %

Kappa statistic 0.9352

Mean absolute error 0.0367

Root mean squared error 0.1479

Relative absolute error 7.329 %

Root relative squared error 29.4806 %

Total Number of Instances 340

**K = 4**

Correctly Classified Instances 330 97.0588 %

Incorrectly Classified Instances 10 2.9412 %

Kappa statistic 0.9411

Mean absolute error 0.0393

Root mean squared error 0.1479

Relative absolute error 7.843 %

Root relative squared error 29.4694 %

Total Number of Instances 340

**K = 5**

Correctly Classified Instances 331 97.3529 %

Incorrectly Classified Instances 9 2.6471 %

Kappa statistic 0.947

Mean absolute error 0.042

Root mean squared error 0.1538

Relative absolute error 8.3861 %

Root relative squared error 30.6401 %

Total Number of Instances 340

**Multilayer Perception**

Correctly Classified Instances 340 100 %

Incorrectly Classified Instances 0 0 %

Kappa statistic 1

Mean absolute error 0.0089

Root mean squared error 0.0523

Relative absolute error 1.7697 %

Root relative squared error 10.4202 %

Total Number of Instances 340

**PART**

Correctly Classified Instances 324 95.2941 %

Incorrectly Classified Instances 16 4.7059 %

Kappa statistic 0.9055

Mean absolute error 0.0544

Root mean squared error 0.2128

Relative absolute error 10.8482 %

Root relative squared error 42.409 %

Total Number of Instances 340

**KStar**

Correctly Classified Instances 330 97.0588 %

Incorrectly Classified Instances 10 2.9412 %

Kappa statistic 0.9411

Mean absolute error 0.0907

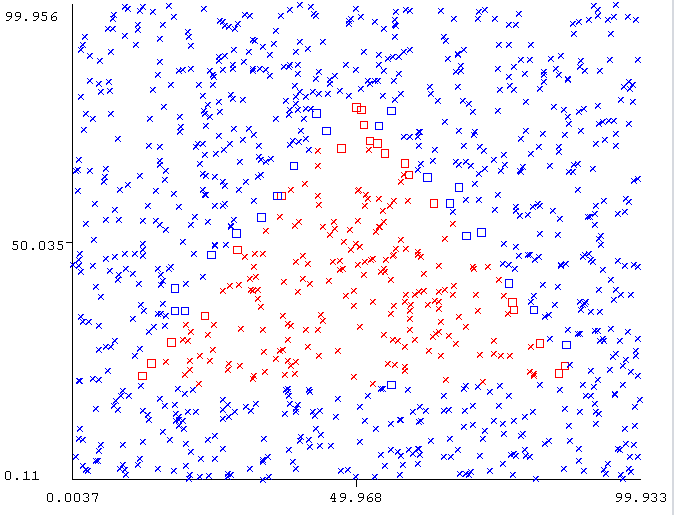
Root mean squared error 0.172

Relative absolute error 18.0853 %

Root relative squared error 34.273 %

Total Number of Instances 340

**Trojúhelník (percentage split 66 %)**



**J48**

Correctly Classified Instances 324 95.2941 %

Incorrectly Classified Instances 16 4.7059 %

Kappa statistic 0.8704

Mean absolute error 0.0505

Root mean squared error 0.2124

Relative absolute error 14.7089 %

Root relative squared error 49.3934 %

Total Number of Instances 340

**Nativní Bayes**

Correctly Classified Instances 265 77.9412 %

Incorrectly Classified Instances 75 22.0588 %

Kappa statistic 0.1241

Mean absolute error 0.2014

Root mean squared error 0.3299

Relative absolute error 58.6152 %

Root relative squared error 76.714 %

Total Number of Instances 340

**NB Tree – Multinomial**

Correctly Classified Instances 193 56.7647 %

Incorrectly Classified Instances 147 43.2353 %

Kappa statistic -0.0387

Mean absolute error 0.4247

Root mean squared error 0.5944

Relative absolute error 123.5918 %

Root relative squared error 138.2327 %

Total Number of Instances 340

**IB1 (1-NN)**

Correctly Classified Instances 329 96.7647 %

Incorrectly Classified Instances 11 3.2353 %

Kappa statistic 0.9112

Mean absolute error 0.0338

Root mean squared error 0.1796

Relative absolute error 9.8267 %

Root relative squared error 41.77 %

Total Number of Instances 340

**IBk (k-NN)**

**K = 2**

Correctly Classified Instances 323 95 %

Incorrectly Classified Instances 17 5 %

Kappa statistic 0.8568

Mean absolute error 0.0448

Root mean squared error 0.1714

Relative absolute error 13.0401 %

Root relative squared error 39.8703 %

Total Number of Instances 340

**K = 3**

Correctly Classified Instances 326 95.8824 %

Incorrectly Classified Instances 14 4.1176 %

Kappa statistic 0.8836

Mean absolute error 0.0495

Root mean squared error 0.1676

Relative absolute error 14.3985 %

Root relative squared error 38.9838 %

Total Number of Instances 340

**K = 4**

Correctly Classified Instances 324 95.2941 %

Incorrectly Classified Instances 16 4.7059 %

Kappa statistic 0.8635

Mean absolute error 0.0503

Root mean squared error 0.1569

Relative absolute error 14.6505 %

Root relative squared error 36.5013 %

Total Number of Instances 340

**K = 5**

Correctly Classified Instances 325 95.5882 %

Incorrectly Classified Instances 15 4.4118 %

Kappa statistic 0.8748

Mean absolute error 0.0591

Root mean squared error 0.1701

Relative absolute error 17.197 %

Root relative squared error 39.5651 %

Total Number of Instances 340

**Multilayer Perception**

Correctly Classified Instances 268 78.8235 %

Incorrectly Classified Instances 72 21.1765 %

Kappa statistic 0.391

Mean absolute error 0.2388

Root mean squared error 0.3551

Relative absolute error 69.5012 %

Root relative squared error 82.5825 %

Total Number of Instances 340

**PART**

Correctly Classified Instances 320 94.1176 %

Incorrectly Classified Instances 20 5.8824 %

Kappa statistic 0.8294

Mean absolute error 0.0588

Root mean squared error 0.1936

Relative absolute error 17.1192 %

Root relative squared error 45.0218 %

Total Number of Instances 340

**KStar**

Correctly Classified Instances 325 95.5882 %

Incorrectly Classified Instances 15 4.4118 %

Kappa statistic 0.8714

Mean absolute error 0.1164

Root mean squared error 0.1947

Relative absolute error 33.8805 %

Root relative squared error 45.2696 %

Total Number of Instances 340

**Výsledky**

Cílem projektu bylo otestovat účinnost jednotlivých náhodných dat podle jednotlivých algoritmů.

Jako specifické algoritmy byly použity:

Následující tabulka ukazuje průměr přesností všech použitých algoritmů. Mezi nejlepší a nejvýhodnější algoritmus patří IBk.

|  |  |
| --- | --- |
| **Data** | **Průměrná přesnost (%)** |
| J48 | 86,918 |
| Nativní Bayes | 80,284 |
| NB Tree – Multinomial | 68,208 |
| IB1 (1-NN) Ibk (k-NN) | **97,908** |
| Multilayer Perception | 85,728 |
| PART | 86,248 |
| KStar | 96,392 |