

UNIVERZA V LJUBLJANI  
FAKULTETA ZA RAČUNALNIŠTVO IN INFORMATIKO

Darko Pevec

# **Ocenjevanje zanesljivosti posameznih klasifikacij z lokalnimi metodami**

DIPLOMSKO DELO  
NA INTERDISCIPLINARNEM UNIVERZITETNEM ŠTUDIJU  
RAČUNALNIŠTVO IN MATEMATIKA

Mentor: prof. dr. Igor Kononenko

Ljubljana, 2009



Št. naloge: 00012/2009

Datum: 01.09.2009

Univerza v Ljubljani, Fakulteta za računalništvo in informatiko ter Fakulteta za matematiko in fiziko izdaja naslednjo nalogo:

Kandidat: **DARKO PEVEC**

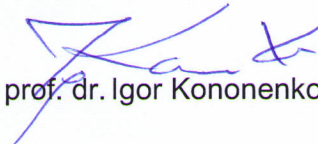
Naslov: **OCENJEVANJE ZANESLJIVOSTI POSAMEZNIH KLASIFIKACIJ Z  
LOKALNIMI METODAMI**  
**EVALUATING RELIABILITY OF INDIVIDUAL CLASSIFICATIONS WITH  
LOCAL METHODS**

Vrsta naloge: Diplomsko delo univerzitetnega študija

Tematika naloge:

Naloga je razviti metode za ocenjevanje zanesljivosti posameznih klasifikacij z lokalnimi metodami ter varianco bagginga. Osnova za razvoj lokalnih metod so metode za ocenjevanje zanesljivosti regresijskih napovedi, ki jih je razvil Zoran Bosnić v okviru svojega doktorata. Kandidat naj preveri ocenjevanje zanesljivosti za različne klasifikatorje na umetnih in na realnih bazah podatkov ter naj primerja ocenjevanje zanesljivosti klasifikacij z obrnjeno transdukcijo, ki jo je razvil v okviru svojega doktorata Matjaž Kukar, z razvitimi lokalnimi metodami ter varianco bagginga. Kandidat bo implementiral metodo v okviru izbranega sistema in ovrednotil možnosti uporabe na več problemskih domenah.

Mentor:

  
prof. dr. Igor Kononenko



Dekan Fakultete za računalništvo in informatiko:

  
prof. dr. Franc Solina

Dekan Fakultete za matematiko in fiziko:

akad. prof. dr. Franc Forstnerič







Rezultati diplomskega dela so intelektualna lastnina Fakultete za računalništvo in informatiko Univerze v Ljubljani. Za objavljanje ali izkoriščanje rezultatov diplomskega dela je potrebno pisno soglasje Fakultete za računalništvo in informatiko ter mentorja.



**Namesto te strani vstavite** original izdane teme diplomskega dela s podpisom mentorja in dekana ter žigom fakultete, pred vezavo.

[ *Prazna stran.* ]



# IZJAVA O AVTORSTVU

diplomskega dela

Spodaj podpisani Darko Pevec,

z vpisno številko 63050205,

sem avtor diplomskega dela z naslovom:

Ocenjevanje zanesljivosti posameznih klasifikacij z lokalnimi metodami

S svojim podpisom zagotavljam, da:

- sem diplomsko delo izdelal samostojno pod mentorstvom prof. dr. Igorja Kononenka
- so elektronska oblika diplomskega dela, naslov (slov., angl.), povzetek (slov., angl.) ter ključne besede (slov., angl.) identični s tiskano obliko diplomskega dela
- soglašam z javno objavo elektronske oblike diplomskega dela v zbirki „Dela FRI”.

V Ljubljani, dne 01.09.2009

Podpis avtorja:



# Zahvala

Prvi na seznamu zahval je seveda Igor Kononenko, ki me je kot mentor sprejel pod svoje okrilje. Pri tem delu mi je podaril mnogo dobrih nasvetov, zanimivih idej, potrpežljivih razlag in obilico novega znanja, pa tudi nekaj modrosti. Kolega Zoran Bosnić je zaslužen, da mi je uspelo obvladati programsko okolje za statistično programiranje R. Ravno tako se mu zahvaljujem za napotke in navodila pri statistični obdelavi rezultatov, ponavadi ob skodelici kave. Za konec pa še zahvala mojim najdražjim, za vso pomoč in podporo ter vsem, ki so me prenašali, me vzpodbujali in mi venomer stali ob strani.



*Prikrajšani popotnici,  
moji mali sopotnici.*



# Kazalo

<b>Povzetek</b>	<b>1</b>
<b>Abstract</b>	<b>2</b>
<b>1 Uvod</b>	<b>3</b>
1.1 Motivacija in cilji . . . . .	3
1.2 Pregled področja . . . . .	4
1.3 Pregled vsebine . . . . .	5
<b>2 Strojno učenje in klasifikacija</b>	<b>7</b>
2.1 Odločitvena drevesa . . . . .	8
2.2 Naivni Bayesov klasifikator . . . . .	9
2.3 Metoda najbližjih sosedov . . . . .	9
2.4 Umetne nevronske mreže . . . . .	10
2.5 Naključni gozdovi odločitvenih dreves . . . . .	11
2.6 Bagging z odločitvenimi drevesi . . . . .	11
2.7 Metoda podpornih vektorjev . . . . .	12
<b>3 Ocenjevanje zanesljivosti</b>	<b>15</b>
3.1 Ocenjevanje zanesljivosti kot sprememba porazdelitve verjetnosti . . . . .	15
3.2 Ocena lokalne napake . . . . .	17
3.3 Metoda lokalnega prečnega preverjana . . . . .	18
3.4 Zanesljivost kot gostota učnih primerov . . . . .	18
3.5 Varianca modela bagging . . . . .	19
3.6 Obratna transdukcija . . . . .	20
<b>4 Metodologija testiranja</b>	<b>21</b>
4.1 Testne podatkovne množice . . . . .	21
4.2 Statistične metode . . . . .	22

<b>5</b>	<b>Rezultati testiranj ocen zanesljivosti</b>	<b>25</b>
5.1	Q-Q diagrami . . . . .	25
5.2	Rezultati statističnih testov . . . . .	34
<b>6</b>	<b>Diskusija in zaključki</b>	<b>41</b>
6.1	Pregled rezultatov . . . . .	41
6.2	Praktični napotki . . . . .	43
6.3	Nadaljnje delo . . . . .	44
	<b>Literatura</b>	<b>45</b>
<b>A</b>	<b>Podrobni rezultati testiranj</b>	<b>47</b>
A.1	Korelacijski koeficienti ocen <i>CNK</i> . . . . .	48
A.2	Korelacijski koeficienti ocen <i>LCV</i> . . . . .	54
A.3	Korelacijski koeficienti ocen <i>LCV<sub>d</sub></i> . . . . .	60
A.4	Korelacijski koeficienti ocen <i>DENS</i> . . . . .	66
A.5	Korelacijski koeficienti ocen <i>BAGV</i> . . . . .	67
A.6	Korelacijski koeficienti ocen <i>TRANS<sub>1</sub></i> . . . . .	68
A.7	Korelacijski koeficienti ocen <i>TRANS<sub>2</sub></i> . . . . .	74
A.8	Korelacijski koeficienti ocen <i>TRANS<sub>n</sub></i> . . . . .	80
A.9	Korelacijski koeficienti referenčnih ocen zanesljivosti . . . . .	86

# Seznam uporabljenih kratic in simbolov

<b>BAG</b>	bagging odločitvenih dreves
<b>BAGV</b>	varianca modela bagging
<b>CNK</b>	ocena lokalne napake (CNK)
<b>DENS</b>	zanesljivost kot gostota učnih primerov
<b>DT</b>	odločitveno drevo (decision tree)
<b>kNN</b>	metoda najbližjih sosedov (k-nearest neighbors)
<b>LCV</b>	metoda lokalnega prečnega preverjanja (local cross validation)
<b>NB</b>	naivni Bayes-ov klasifikator
<b>NN</b>	nevronska mreža kot klasifikator (neural network)
<b>RF</b>	random forest tj. naključni gozd klasifikatorjev
<b>RT</b>	obratna transdukcija (reverse transduction)
<b>SVM</b>	metoda podpornih vektorjev (support vector machine)



# Povzetek

Diplomsko delo obravnava področje ocenjevanja zanesljivosti posameznih napovedi klasifikacijskih modelov, zgrajenih z uporabo nadzorovanega učenja.

V splošnem zanesljivost pomeni sposobnost predvidljivega opravljanja določenih funkcij, v kontekstu strojnega učenja pa pod tem terminom mislimo na pojem točnosti, tj. sposobnost podajanja točne napovedi. Ker se v praksi soočamo s problemom, da mere zanesljivosti niso kvantitativno definirane, zanesljivost le ocenjujemo. Ocene zanesljivosti posameznih napovedi nam lahko nudijo dodatno koristno informacijo, ki je lahko v pomoč pri presoji o uporabnosti posameznih napovedi.

Za potrebe naloge smo razvili metode za ocenjevanje zanesljivosti posameznih klasifikacij z lokalnimi metodami ter varianco bagginga. Ocenjevanje zanesljivosti preverimo za različne klasifikatorje na umetnih in na realnih bazah podatkov in naredimo primerjavo razvitih metod z ocenjevanjem zanesljivosti klasifikacij z obrnjeno transdukcijo.

Metode smo preizkusili na dvajset umetnih in realnih problemskih domelah, uporabili smo sedem klasifikacijskih metod, ocene pa smo tvorili z enajstimi različnimi razdaljami oz. merami. Dobljene ocene smo analizirali s statističnimi metodami in ugotovili, da opravljeni testi ne dajejo jasnih rezultatov, saj opisna statistična metoda Q-Q diagramov v mnogih primerih ne podpira izmerjene korelacije. Korelacijski koeficienti kažejo, da sta najboljši razviti metodi *LCV* in *BAGV*. Dobre rezultate sta pokazali še oceni  $TRANS_1$  in *CNK*, ostale metode pa so dosegle bistveno slabše rezultate.

## Ključne besede:

Strojno učenje, klasifikacija, ocenjevanje zanesljivosti, klasifikacijska točnost.



# Abstract

In this thesis we take upon different approaches for estimating reliability of individual classification predictions made by classifiers based on supervised learning.

The general definition of the term *reliability* is the ability to perform required functions under stated conditions. In machine learning, we refer to *accuracy*, as in the ability to provide accurate predictions. We face the problem that measures of reliability are not quantitatively defined. We can therefore only conceive estimates. Reliability estimates of individual predictions provide valuable information that could be beneficial in individual predictions assessment of use.

For the needs of our thesis we develop several methods for reliability estimation based on existing approaches of local methods and the variance of a bagged model. We test our methods on various available real-life and artificial datasets and compare our methods with those based on inverse transduction.

Methods were tested on 20 different datasets on 7 classification models and the estimates were calculated using 11 measures of similarity. We applied three statistical methods to our results. We came to a conclusion that these tests do not give clear results, as Q-Q plots only vaguely support calculated correlation. Correlation tests show potential of estimates *LCV* and *BAGV* as they demonstrated best on average performance. Second-comers with good result were estimates *TRANS<sub>1</sub>* and *CNK*, while other estimates failed to excel.

## Key words:

Machine learning, classification, reliability estimates, classification accuracy.



# Poglavje 1

## Uvod

### 1.1 Motivacija in cilji

Pojem *zanesljivost* je zelo širok, v splošnem pa ponavadi pomeni sposobnost predvidljivega opravljanja določenih funkcij. Pri strojnem učenju se ponavadi osredotočamo na pojem *točnosti*, tj. na sposobnost podajanja točne napovedi.

V praksi se soočamo s problemom, da mere zanesljivosti niso kvantitativno definirane, zato zanesljivost le ocenjujemo. Ocene zanesljivosti posameznih napovedi nam lahko nudijo dodatno koristno informacijo, ki je lahko v pomoč pri presoji o uporabnosti posameznih napovedi. V primerih, ko imajo napovedi pomembne posledice, preproste metode ki dajo povprečene rezultate niso dovolj zadovoljive. Takrat bi želeli dobljene napovedi podkrepiti vsaj še z kakšno bolj kredibilno razlago. Poskusimo demonstrirati: Ko ljudje napovedujemo, si večina pravi „*Mislím, da je odgovor . . . , vendar nisem čisto prepričan*” ali „*Odgovor je . . . , zagotovo!*” Običajno ne razmišljamo o povprečni točnosti napovedi dotične osebe; običajno podzavestno ocenjujemo zanesljivost posamezne izjave, na osnovi preteklih podobnih primerov. Pri kritičnih napovedih (medicinska diagnostika in prognostika, borza, oz. finance) kjer so v igri človeška življenja in premoženje bi informacije o zanesljivosti posameznih napovedi bile zelo dobrodošle. Zato je precej intuitivno iskati metode ocenjevanja zanesljivosti posameznih napovedi na bolj lokalni način.

Naloga tega diplomskega dela je razviti metode za ocenjevanje zanesljivosti posameznih klasifikacij z lokalnimi metodami ter varianco bagginga. Ocenjevanje zanesljivosti smo preverili za različne klasifikatorje na umetnih in na realnih bazah podatkov. Razvite metode smo primerjali z metodo obrnjene transdukcije in na podlagi referenčne ocene zanesljivosti ovrednotili možnosti uporabe razvitih metod na več problemskih domenah.

## 1.2 Pregled področja

Najbolj relevantne raziskave so s področja ocenjevanja zanesljivosti posameznih napovedi. Primeren kriterij delitve različnih pristopov je glede na to, ali je pristop vezan na določen model napovedovanja. Pristopi, ki so vezani na določen model, so ponavadi manj splošni, so pa podprti z natančnimi matematičnimi ali verjetnostnimi lastnostmi.

Pristopi, neodvisni od uporabljenih modelov strojnega učenja, ne znajo izkoristiti dodatnih parametrov in znanja, ki so vezani na posamezen model, so pa zato bolj splošni in je njihova uporabnost širša. Ti pristopi se osredotočajo na opazovanje sprememb napovedi ob spremembah zunanjih parametrov nadzorovanega učenja. Najpogosteje se opazujejo spremembe ob spremembah učne množice ali atributov. Osnova za naše delo so že razvite lokalne metode ocenjevanja zanesljivosti regresijskih napovedi, ki so predstavljene v [2]. Pri regresiji so ti pristopi definirani kot metrike nad opazovanimi učnimi parametri. Zaradi hevristične zasnove te metrike niso omejene na zaprt interval vrednosti in posledično ne morejo imeti verjetnostne interpretacije.

Osnovna ideja večine obstoječih lokalnih metod ocenjevanja zanesljivosti posameznih napovedi je transdukcija. Transduktivno sklepanje je sklepanje od konkretnega h konkretnemu. Navadno iz učnih primerov sklepamo na vrednost odvisne spremenljivke novega primera, transdukcijo pa lahko uporabimo tudi v obratni smeri, tako da v učno množico dodamo nove primere, za katere ne poznamo prave vrednosti odvisne spremenljivke [12]. Na tej osnovi je bila v [13] razvita metoda ocenjevanja zanesljivosti posameznih klasifikacijskih napovedi na osnovi obrnjene transdukcije, s katero je narejena primerjava prej omenjenih lokalnih metod in variance bagging-a.

Za namene primerjave teh metod pri klasifikaciji je bilo potrebno obstoječe metode nekoliko prilagoditi. Spremeniti jih je bilo potrebno tako, da delujejo nad prostorom verjetnostnih porazdelitev. Pri klasifikaciji so napovedi omejene na zaprtem intervalu. Če spretno izbiramo način merjenja razdalj med napovedmi, ohranimo zaprtost intervala zaloge vrednosti, in potem imajo lahko takšne ocene zanesljivosti verjetnostno interpretacijo.

Nove metode smo implementirali v okviru okolja za statistično programiranje R [16]. Ker je to okolje v široki uporabi tudi na področju strojnega učenja, obstajajo razširitveni paketi, v katerih so klasifikacijski modeli že implementirani. Pakete razvijajo različne skupine raziskovalcev, zato obstajajo malenkostne razlike v uporabi in obnašanju klasifikatorjev. Za sistematično testiranje je bilo zaradi tega potrebno dodatno izgraditi še ogrodje za klasifikatorje, tako da se navzven vsi klasifikatorji obnašajo enako.

## 1.3 Pregled vsebine

V 2. poglavju si najprej na kratko ogledamo klasifikacijske modele, katerih napovedi ocenjujemo. Navedemo tudi, znotraj katerih paketov se nahajajo in s kakšnimi parametri jih je potrebno klicati, da pridobimo napovedi željene oblike. V poglavju 3 namenimo nekoliko pozornosti merjenju razdalj v prostoru verjetnosti, saj je to osnova večine razvitih metod. Razvijemo lokalne metode ocenjevanja zanesljivosti in povzemamo metodo obrnjene transdukcije. Nove metode želimo s slednjo primerjati, za kar potrebujemo določena orodja. Metodologijo testiranja obravnava 4. poglavje. Za vrednotenje ocen zanesljivosti posameznih napovedi uporabimo statistične teste. Namen naših testov je kvalitativno in kvantitativno ovrednotiti razvite metode ocenjevanja zanesljivosti. Z njimi v 5. poglavju ovrednotimo stopnjo korelacije med ocenami zanesljivosti in klasifikacijsko napako. V zadnjem poglavju sledijo še diskusija, pregled rezultatov in zaključki.



## Poglavje 2

# Strojno učenje in klasifikacija

Metod strojnega učenja je veliko. V osnovi se delijo glede na način uporabe naučenega znanja: klasifikacija, regresija, učenje asociacij in logičnih relacij, učenje sistemov enačb in razvrščanje [12]. Najpogostejši uporabi strojnega učenja sta klasifikacija in regresija.

Učni algoritmi na vhodu dobijo predznanje in učne primere. Zatem preiščejo prostor možnih hipotez in kot rezultat vrnejo zaključno hipotezo. Tako pri klasifikaciji kot pri regresiji je naloga *klasifikatorjev* in *prediktorjev* določanje neznanne razredne spremenljivke novih vhodnih objektov. Najpogosteje se uporablja atributna predstavitev teh objektov, oziroma učnih primerov. Atributi so lahko zvezni ali diskretni in njihovo število je dano v naprej. Atributna predstavitev je v [12] definirana z množico atributov  $A$ :

$$A = A_i, i = 0 \dots a;$$

za vsak diskretni atribut  $A_i$  imamo množico možnih vrednosti:

$$\forall A_i \text{ diskreten} : \mathcal{V}_i = V_1, \dots, V_{n_i};$$

za vsak zvezni atribut  $A_i$  imamo interval možnih vrednosti:

$$\forall A_i \text{ zvezen} : \mathcal{V}_i = [Min_i, Max_i];$$

razred je podan z atributom  $A_0$ : če rešujemo klasifikacijski problem, potem je  $A_0$  diskretni atribut, če pa rešujemo regresijski problem, je  $A_0$  zvezni atribut; en učni primer je vektor vrednosti atributov, kjer je razred označen z  $r_{(j)} = v_{(0,j)}$ :

$$u_j = \langle r_{(j)}, v_{(1,j)}, \dots, v_{(a,j)} \rangle;$$

množica učnih primerov je podana kot množica vektorjev:

$$\mathcal{U} = \{u_j, j = 1 \dots n\}.$$

Pri klasifikaciji je razredna spremenljivka  $r_j$  diskretna, na drugi strani, pri regresiji pa je ta spremenljivka zvezna. Zaradi te osnovne razlike moramo obstoječe metode ocenjevanja zanesljivosti posameznih napovedi, ki so razvite za uporabo nad regresijskimi modeli, „prevesti“ za uporabo nad klasifikacijskimi modeli. To naredimo v naslednjem poglavju, v tem poglavju pa si na kratko ogledamo klasifikatorje, katere uporabljamo pri testiranju. Večino definicij povzamemo po [12]. V podrobnosti in izboljšave se ne spuščamo, saj naš cilj ni doseči čim večjo klasifikacijsko točnost ampak nasprotno. Statistično želimo analizirati korelacijo ocen zanesljivosti s klasifikacijsko napako, za to pa potrebujemo dovolj slabih napovedi.

Omenili smo že, da uporabljamo že implementirane klasifikatorje, ki so znotraj dopolnilnih paketov okolja za statistično programiranje R. Navedemo znotraj katerih paketov se nahajajo in s kakšnimi parametri jih je potrebno klicati, da pridobimo napovedi v obliki *verjetnostne distribucije po razredih* (Več o tem v razdelku 3.1).

## 2.1 Odločitvena drevesa

Odločitveno drevo je sestavljeno iz notranjih vozlišč, ki ustrezajo atributom, vej, ki ustrezajo podmnožicam vrednosti atributov, in listov, ki ustrezajo razredom. Ena pot v drevesu od korena do lista ustreza enemu odločitvenemu pravilu. Pri tem so pogoji, tj. pari atribut–podmnožica vrednosti, ki jih obiščemo na poti, konjuktivno povezani. Algoritmi odločitvenih dreves za gradnjo uporabljajo ocene informativnosti za izbiro posameznih atributov in ustreznih podmnožic njihovih atributov. Pri odločitvenih drevesih imamo za dani diskretni atribut  $A$  in njegove vrednosti  $V, V_1, \dots, V_n$  konjuktivne pogoje oblike  $A = V$  ali  $A \in V_1, \dots, V_n$ . Zvezne attribute je potrebno bodisi vnaprej diskretizirati, ali pa tvoriti pogoje za dani zvezni atribut  $A$  in vrednosti  $V, V_1$  in  $V_2$  oblike  $A > V$  ali  $A \leq V$  oziroma  $V_1 \geq A \geq V_2$ . V okolju za statistično programiranje R najdemo implementacijo odločitvenih dreves v paketu `rpart`. Verjetnostno distribucijo dobimo tako, da učnemu algoritmu podamo parameter `method="class"` [3].

## 2.2 Naivni Bayesov klasifikator

Naloga Bayesovega klasifikatorja je izračunati pogojne verjetnosti za vsak razred pri danih vrednostih atributov, za dani novi primer, ki ga želimo klasificirati. Naivni Bayesov klasifikator predpostavi pogojno neodvisnost pri danem razredu. To omogoči, da učna množica ponavadi zadošča za zanesljivo oceno vseh potrebnih verjetnosti za izračun končne pogojne verjetnosti vsakega razreda. Osnovna formula je izpeljana s pomočjo Bayesovega pravila:

$$P(r_k|V) = P(r_k) \prod_{i=1}^a \frac{P(r_k|v_i)}{P(r_k)}$$

Naloga učnega algoritma je potem s pomočjo učne množice aproksimirati apriorne verjetnosti razredov  $P(r_k)$ ,  $k = 1 \dots n_0$  in pogojne verjetnosti  $P(r_k|v_i)$  razredov  $r_k$ ,  $k = 1 \dots n_0$ , pri danih vrednostih  $v_i$  atributa  $A_i$ ,  $i = 1 \dots a$ .

V okolju za statistično programiranje R najdemo implementacijo naivnega Bayesovega klasifikatorja `naiveBayes` v paketu `e1071`, za pravilno delovanje pa potrebujemo še paket `class`. Verjetnostno distribucijo dobimo tako, da klasifikacijskemu algoritmu `predict` podamo parameter `type="raw"`. Ker naivni Bayesov klasifikator računa pogojne verjetnosti diskretnih spremenljivk, je potrebna diskretizacija zveznih atributov. V uporabljeni implementaciji je diskretizacija že vgrajena.

## 2.3 Metoda najbližjih sosedov

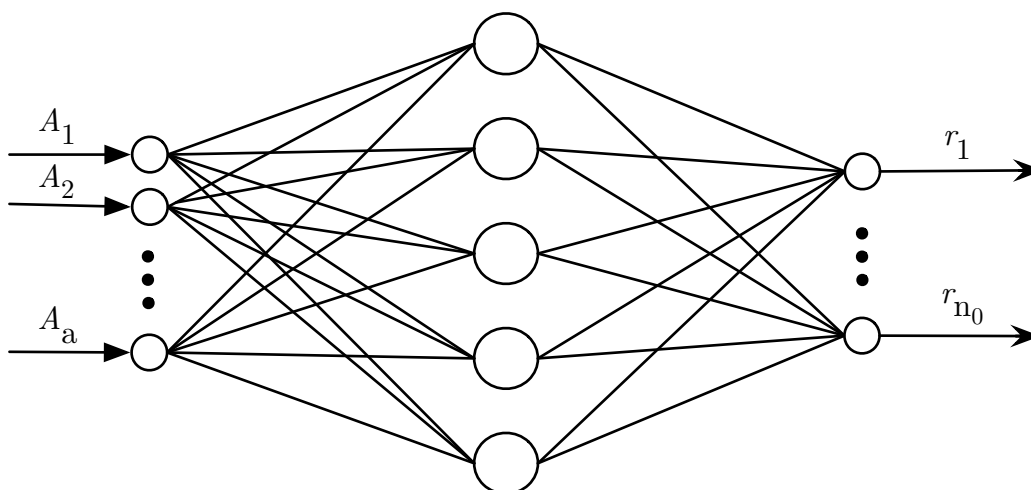
Najpreprostejša varianta algoritma najbližjih sosedov (kNN) kot znanje uporablja kar množico vseh učnih primerov, učni algoritem si samo zapomni vse primere. Pri klasifikaciji novega primera se iz učne množice poišče določeno število najbolj podobnih, tj. najbližjih primerov - sosedov. Nov primer klasificiramo v razred, ki mu pripada največ bližnjih sosedov. Zaradi ustrezne metrike v prostoru atributov je pri tem potrebno normalizirati vrednosti zveznih atributov in definirati razdaljo (metriko) med vrednostmi vsakega diskretnega atributa.

Implementacija metode najbližjih sosedov se v okolju R nahaja v paketu `kkn`. Pri testiranju smo okolico nastavili na pet sosedov s parametrom `k=5`, rezultat v obliki verjetnostne distribucije smo pa črpali iz spremenljivke (dobljenega odgovora) `prob`.

## 2.4 Umetne nevronske mreže

Algoritmi umetnih nevronske mreže skušajo oponašati biološke nevronske mreže tako, da nevron abstrahirajo na preprost element, ki zna nad vhodnimi signali izvršiti preprosto funkcijo, na primer uteženo seštevanje, in rezultat normalizirati s pragovno funkcijo. Take preproste umetne nevrone potem lahko povežemo v poljubno kompleksne umetne nevronske mreže. Za klasifikacijo se najpogosteje uporablja usmerjene večnivojske umetne nevronske mreže. Te kaskadno povezujejo več nivojev nevronov: vhodni nevroni (ki ustrezajo atributom), eden ali več nivojev skritih nevronov in izhodni nevroni (ki ustrezajo zalogi vrednosti razrednega atributa). Naloga učnega algoritma je nastaviti uteži na povezavah med nevroni tako, da je klasifikacijska napaka čim manjša. Med klasifikacijo novega primera nevronska mreža na vhodu dobi vrednosti atributov, s katerimi nastavi vrednosti vhodnih nevronov, za dani novi primer. Zatem vsak nevron na naslednjem nivoju izračuna svoj izhod kot uteženo vsoto svojih vhodnih signalov. Neuron z največjim odzivom na zadnjem nivoju določi razred novemu primeru.

V okolju R so umetne nevronske mreže z enim skritim nivojem implementirane v paketu `nnet` [15]. Pri testiranju so bile uporabljene privzete nastavitve; nastaviti je bilo potrebno le število nevronov skrite plasti, kar dosežemo s parametrom `size=5`. Shematski prikaz uporabljene nevronske mreže prikazuje slika 2.1. Da dobimo verjetnostno distribucijo, klasifikacijskemu algoritmu podamo parameter `type="raw"`.



Slika 2.1: Umetna nevronska mreža s petimi nevroni v skriti plasti.

## 2.5 Naključni gozdovi odločitvenih dreves

Metoda naključnih gozdov je namenjena izboljševanju napovedne točnosti drevesnih algoritmov. Originalno je bila razvita za odločitvena drevesa [5]. Ideja je generirati zaporedje odločitvenih dreves, tako da se pri izbiri najboljšega atributa v vsakem vozlišču naključno izbere relativno majhno število atributov, ki vstopajo v izbor za najboljši atribut. Avtor predlaga naključno izbiro toliko atributov v vsakem vozlišču, kolikor znaša logaritem od števila atributov plus 1. Lahko pa je to število kar 1, kar pomeni popolnoma naključno izbiro atributa v vsakem vozlišču vsakega drevesa. Število tako zgrajenih dreves je ponavadi 100, lahko tudi več.

Vsako drevo, ki je zgrajeno na učni množici, se zatem uporabi za klasifikacijo novega primera po metodi glasovanja - vsako drevo ima en glas, ki ga nameni razredu, v katerega bi to drevo klasificiralo nov primer. Iz vseh glasov zatem dobimo verjetnostno distribucijo po vseh razredih.

V okolju R je metoda naključnih gozdov implementirana v paketu `randomForest` [6]. Za rezultat v obliki verjetnostne distribucije je potrebno klasifikacijskemu algoritmu podati parameter `type='prob'`.

## 2.6 Bagging z odločitvenimi drevesi

Izraz bagging pride iz izraza „bootstrap aggregating“ [4]. Bootstrap je bolj splošen princip razmnoževanja učnih primerov, kadar jih nimamo dovolj za učenje, pri baggingu pa generiramo serijo različnih učnih množic. Če ima učna množica  $n$  primerov, potem vsakič  $n$  krat naključno izberemo primer iz učne množice z vračanjem. To pomeni, da se isti učni primer v tako generirani učni množici lahko večkrat ponovi, nekaterih primerov iz originalne učne množice pa tako generirana množica sploh ne vsebuje. Nad vsako tako generirano učno množico zatem požemo učni algoritem (v našem primeru za gradnjo odločitvenega drevesa) in tako dobimo veliko število potencialno različnih hipotez. Serijo zgeneriranih hipotez potem uporabimo za klasifikacijo novega primera tako, da kombiniramo napovedi vseh hipotez.

Implementacija metode `bagging` je znotraj okolja R v paketu `ipred`. Za rezultat v obliki verjetnostne distribucije je potrebno klasifikacijskemu algoritmu podati parameter `type='prob'`, enako kot pri metodi naključnih dreves.

## 2.7 Metoda podpornih vektorjev

Naloga učnega algoritma je izračunati vrednosti koeficientov vnaprej podane diskriminantne funkcije, ki predpostavlja hiperploskev med dvema razredoma v prostoru atributov. Hiperploskev lahko definiramo samo v zveznem prostoru, torej morajo biti vsi atributi zvezni. Če imamo več kot dva razreda, potrebujemo za vsak par razredov eno hiperploskev. Diskriminantne funkcije so lahko linearne, kvadratične, polinomske itd. Na primer, če je diskriminantna funkcija linearna, potem učni algoritem skuša primere iz dveh razredov med seboj ločiti s hiperravnino. Koeficiente hiperploskev določimo tako, da je klasifikacijska napaka čim manjša.

Običajna, tako imenovana Fisherjeva linearna diskriminantna funkcija predpostavlja normalno porazdelitev primerov znotraj vsakega razreda ter maksimizira razdaljo med povprečnim učnim primerom iz prvega razreda in povprečnim učnim primerom iz drugega razreda. Pri tem upošteva različni varianci dveh razredov in poišče optimalno klasifikacijsko mejo med razredoma.

Metoda podpornih vektorjev (SVM) je ena najbolj uspešnih metod za klasifikacijo, je pa potrebno optimizirati več kriterijev:

- Potrebno je maksimizirati razdaljo podpornih vektorjev od ločitvene hiperravnine. Podporni vektorji so tisti učni primeri, ki so najbližji hiperravnini in ti določajo rob okoli hiperravnine. Torej po tej metodi želimo maksimizirati širino roba.
- Ker mnogi klasifikacijski problemi niso eksaktno rešljivi z linearno funkcijo, metoda SVM uporablja implicitno transformacijo atributov v (potencialno veliko, lahko tudi neskončno) množico novih atributov s t.i. jedrnimi funkcijami. Od izbire ustreznega jedra je odvisna uspešnost metode SVM.
- Ker lahko vsak klasifikacijski problem rešimo eksaktno z dovolj zapleteno funkcijo, je treba upoštevati tudi kompleksnost rešitve, kar pomeni minimizacijo vsote uteži atributov.
- Kljub nelinearni transformaciji atributov se moramo zadovoljiti z napačno klasifikacijo nekaterih primerov. Seveda pa želimo število napačno klasificiranih primerov minimizirati.

Potrebno je torej poiskati kompromis med vsemi kriteriji. Metoda SVM uporablja različne optimizacijske algoritme, ki optimizirajo vse kriterije hkrati.

Še vedno pa obstajajo parametri, ki utežujejo posamezne kriterije, in je treba njihovo najboljšo vrednost določiti empirično.

Implementacija metode SVM se v okolju R nahaja v paketu LIBSVM [7] [8]. Ker se ne ukvarjamo z optimizacijo, smo vse nastavitve prepustili avtorjem paketa. Glavna prednastavljena parametra sta uporaba radialnih jedrnih funkcij (oblike  $e^{-\gamma*|u-v|^2}$ ;  $\gamma = \frac{1}{a}$ ), uteži posameznih kriterijev pa so enake. Algoritem vrne verjetnostno distribucijo napovedi v atributu 'probabilities', če nastavimo parameter `probability=T`.



## Poglavje 3

# Ocenjevanje zanesljivosti

V tem poglavju opredelimo način ocenjevanja zanesljivosti. Takoj zatem predstavimo razvoj lokalnih metod ocenjevanja zanesljivosti *CNK*, *LCV*, *DENS*, *BAGV* ter nekoliko starejše metode ocenjevanja zanesljivosti z obratno transdukcijo — *TRANS*.

### 3.1 Ocenjevanje zanesljivosti kot sprememba porazdelitve verjetnosti

Zanesljivost ocenjujemo kot spremembe porazdelitve verjetnosti, kakor je opisano v [13]. Vsi klasifikatorji znajo vračati svoje napovedi v obliki verjetnostnih distribucij, zato se osredotočamo na kvalitativno merjenje sprememb takšnih napovedi.

Ukvarjamo se s podprostori vektorskih prostorov  $\mathbb{R}^n$ . Prostor porazdelitev verjetnosti  $\mathbb{P}^n \subset \mathbb{R}^n$  je  $(n - 1)$ -dimenzionalni hipertetraeder v  $\mathbb{R}^n$ , za katerega velja

$$P \in \mathbb{P}^n \iff P = (p_1, p_2, \dots, p_n) \text{ in } \sum_{i=1}^n p_i = 1; \forall i : p_i \in [0, 1]$$

Za oceno zanesljivosti želimo, da je omejena na interval  $[0, 1]$ , kjer ocena 0 pomeni popolnoma nezanesljivo klasifikacijo, ocena 1 pa popolnoma zanesljivo klasifikacijo. S tem pridobimo podobnost z verjetnostjo, kar je pa ravno komplementarno definiciji razdalje. Zaradi tega so naše mere oblike  $1 - \text{razdalja}$ .

Iz [13] in [14] smo izbrali različne metrike in njim podobne razdalje, s pomočjo katerih so nastale v nadaljevanju navedene *mere bližine*  $f$ , za katere velja:

1. funkcija  $f(P, Q)$  je realna funkcija:  $\mathbb{P}^n \times \mathbb{P}^n \longrightarrow \mathbb{R}$ ,
2.  $\forall P, Q \in \mathbb{P}^n : 0 \leq f(P, Q) \leq 1$ ,
3.  $\forall P, Q \in \mathbb{P}^n : f(P, Q) = 1 \Leftrightarrow P = Q$ .

- mera na osnovi manhattanske razdalje

$$f_{\text{man}}(P, Q) = 1 - \sum_i \frac{|p_i - q_i|}{2} \quad (3.1)$$

- mera na osnovi evklidske razdalje

$$f_{\text{evk}}(P, Q) = 1 - \frac{\sqrt{\sum_i (p_i - q_i)^2}}{\sqrt{2}} \quad (3.2)$$

- mera na osnovi maksimalne razdalje

$$f_{\text{max}}(P, Q) = 1 - \max_i |p_i - q_i| \quad (3.3)$$

- mera na osnovi kota med porazdelitvama

$$f_{\text{kot}}(P, Q) = 1 - \max_i |p_i - q_i| \quad (3.4)$$

- mera na osnovi kosinusa kota med porazdelitvama

$$f_{\text{cos}}(P, Q) = 1 - 2 \arccos \left( \frac{\sum_i p_i q_i}{\sqrt{\sum_i p_i^2} \sqrt{\sum_i q_i^2}} \right) \quad (3.5)$$

- mera na osnovi Bhattacharyy razdalje

$$f_{\text{bha}}(P, Q) = \sum_i \sqrt{p_i q_i} \quad (3.6)$$

- mera na osnovi skalarnega produkta

$$f_{\text{ska}}(P, Q) = \sum_i p_i q_i \quad (3.7)$$

- mera na osnovi harmonične srednje vrednosti

$$f_{\text{har}}(P, Q) = \sum_i \frac{2p_i q_i}{p_i + q_i} \quad (3.8)$$

- mera na osnovi Hellingerjeve razdalje

$$f_{\text{hel}}(P, Q) = 1 - \sum_i \frac{(\sqrt{p_i} - \sqrt{q_i})^2}{2} \quad (3.9)$$

- mera na osnovi separacijske razdalje

$$f_{\text{sep}}(P, Q) = 1 - \max_i \left(1 - \frac{p_i}{q_i}\right) \quad (3.10)$$

- mera na osnovi Jensen–Shannon divergence

$$f_{\text{jsd}}(P, Q) = 1 - \sum_i \frac{p_i \log\left(\frac{2p_i}{p_i + q_i}\right) + q_i \log\left(\frac{2q_i}{p_i + q_i}\right)}{2} \quad (3.11)$$

## 3.2 Ocena lokalne napake

Naj bo  $K$  klasifikatorjeva napoved v obliki verjetnostne distribucije možnih vrednosti razredne spremenljivke za dani neoznačeni primer  $(x, -)$  in naj bo  $N$  množica najbližjih sosedov testnega primera  $(x, -)$ ,  $N = \{(x_1, C_1), \dots, (x_k, C_k)\}$ , kjer je  $C_i$  pravilna označba  $i$ -tega najbližjega sosedu. Ta pristop lokalnega ocenjevanja je zasnovan na označitvah najbližjih sosedov. Ocena lokalne napake za neoznačeni testni primer je definirana kot povprečna mera bližine med  $k$  najbližjimi sosedi in klasifikatorjevo napovedjo  $K$ :

$$CNK = \frac{\sum_{i=1}^k f(C_i, K)}{k} \quad (3.12)$$

Pri testiranju smo za izračun ocene  $CNK$  uporabljali pet najbližjih sosedov, ravno toliko sosedov pa jemlje pri svojih napovedih tudi klasifikator KNN (razdelek 2.3). Ker oba ustvarjata napovedi na enak način, sta posledično popolnoma korelirana. Tako par klasifikator KNN in ocena  $CNK$  tvorita kontrolni test odziva uporabljenih mer bližine.

### 3.3 Metoda lokalnega prečnega preverjanja

Ocena zanesljivosti  $LCV$  temelji na ideji postopka prečnega preverjanja [2]. Za neoznačeni primer  $(x, -)$  poteka izračun ocene zanesljivosti po sledečem postopku:

- določimo množico  $k$  najbližjih sosedov  $N = \{(x_1, C_1), \dots, (x_k, C_k)\}$
- za vsakega soseda  $(x_i, C_i) \in N$ 
  - zgradimo klasifikacijski model  $M$  na množici  $N \setminus (x_i, C_i)$
  - za  $x_i$  z modelom  $M$  izračunamo napoved  $K_i$
  - za  $x_i$  izračunamo mero bližine  $f(C_i, K_i)$
- izračunamo oceno zanesljivosti:

$$LCV = \sum_{i=1}^k \frac{f(C_i, K_i)}{k} \quad (3.13)$$

Pri testiranju smo uporabili oceno  $LCV$ , ki glede na velikost domene dinamično prilagaja število najbližjih sosedov  $k$  tako, da je njihovo število enako  $\frac{1}{10}$  moči učne množice. Namesto povprečenja lokalnih napak najbližjih sosedov smo uporabili tudi izboljšavo, da se prispevki posameznih napak uteženo povprečijo glede na razdaljo med testnim primerom in njegovimi sosedi. Tako izboljšano metodo bomo označevali z  $LCV_d$ .

### 3.4 Zanesljivost kot gostota učnih primerov

Metoda, ki skuša napovedati klasifikacijsko napako na osnovi gostote učnih primerov, sloni na osnovni predpostavki, ki povezuje klasifikacijsko napako in gostoto učne množice. Predpostavlja, da je napaka napovedi manjša za primere, ki se nahajajo v bolj gostih predelih učne množice, in višja za primere, ki ležijo v manj gostih predelih učne množice. Predpostavka temelji na tem, da ima učni algoritem v gostem predelu učne množice na razpolago več lokalnih informacij, zato je zaupanje v posamezne napovedi sorazmerno količini informacij, ki so na voljo za njihov izračun. Metoda ima to slabost, da upošteva samo razdaljo do učnih primerov v atributnem prostoru, ne pa tudi njihovih označb. Zaradi tega je zelo občutljiva na šum v podatkih in na situacije, kjer so bistveno različni primeri težko ločljivi med seboj.

Gostoto prostora danega testnega primera ocenimo kot aproksimacijo funkcije gostote verjetnosti pojavitve primera. Za aproksimiranje verjetnostne porazdelitve se uporablja metoda Parzenovih oken z Gaussovimi jedrnimi funkcijami. Ob tem se metodo računanja Gaussovih jeder v večdimenzionalnem prostoru poenostavi tako, da se število dimenzij zmanjša s kombinacijo uporabe funkcije razdalje in dvorazsežnih Gaussovih jeder [9].

Če imamo podano učno množico  $L = \{(x_1, C_1), \dots, (x_l, C_l)\}$ , potem gostoto verjetnosti za neoznačeni testni primer  $(x, -)$  izračunamo kot

$$DENS(x) = \frac{\sum_{e \in L} \kappa(D(x, e))}{l} \quad (3.14)$$

pri čemer je  $D$  izbrana funkcija razdalje,  $\kappa$  pa je izbrana jedrna funkcija (v našem primeru je to funkcija normalne porazdelitve).

### 3.5 Varianca modela bagging

Ta metoda opazovanja variance napovedi znotraj agregiranih modelov strojnega učenja je bila prvič uporabljena za ocenjevanje zanesljivosti agregiranih nevronske mreže. Glede na to, da lahko to metodo posplošimo na poljuben model v okviru postopka bagging, je bila metoda v delu [2] razvita za delo z drugimi regresijskimi modeli. Mi jo bomo spremenili za delo s klasifikacijskimi modeli.

Naj bo podan agregiran klasifikator, ki je sestavljen iz  $m$  klasifikacijskih modelov, kjer  $i$ -ti klasifikator prispeva svojo napoved  $K_i$  v obliki vektorja verjetnostne distribucije možnih razredov. Označbo neoznačenega primera  $(x, -)$  tedaj tvorimo kot povprečno vrednost posameznih napovedanih vektorjev:

$$K = \sum_{i=1}^m \frac{K_i}{m} \quad (3.15)$$

Ocena zanesljivosti  $BAGV$  pa je definirana kot varianca verjetnostnih distribucij po razredih, katere vrnejo klasifikacijski modeli.

$$BAGV = \frac{1}{m} \sum_{k=1}^m (K_k - K)^2 \quad (3.16)$$

### 3.6 Obratna transdukcija

Transduktivno sklepanje je sklepanje od konkretnega h konkretnemu — iz učnih primerov sklepamo na vrednost odvisne spremenljivke novega primera. Transdukcijo lahko uporabimo tudi v obratni smeri, tako da v učno množico dodamo nove primere, za katere ne poznamo prave vrednosti odvisne spremenljivke, in opazujemo spremembe.

Metoda ocenjevanja zanesljivosti z obratno transdukcijo temelji na ideji, da lahko zanesljivost klasifikatorja ocenimo kot funkcijo spremembe porazdelitve verjetnosti po razredih. Spremembo povzročimo na tri načine in sicer tako, da v učno množico dodamo testni primer s primerno izbranim razredom. V osnovi imamo izbiro, ali testni primer uvrstimo v *isti* razred, v katerega ga je uvrstil klasifikator sam, ali pa ga uvrstimo v *nekvalificiranega*. Pri slednjem imamo več različnih možnosti izbire [13].

Po predhodnem klasificiranju testnega primera  $(x, \_)$  in pridobitvi verjetnostne distribucije  $K$  se omejimo na tri skrajnosti:

- učnemu algoritmu se doda v učno množico primer, označen s tistim razredom, v katerega je klasifikator dani primer predhodno uvrstil (označimo z  $r_1$ ) S tem merimo, ali klasifikator bolj zaupa v svojo napoved, če primer že pozna;
- učnemu algoritmu se doda v učno množico primer, označen z najbolj verjetnim nekvalificiranim razredom, torej v drugi najbolj verjeten razred,  $r_2$ . S tem opazujemo, ali se klasifikator „zmede“ pri novem klasificiranju;
- učnemu algoritmu se doda v učno množico primer, označen z najmanj verjetnim nekvalificiranim razredom  $r_n$ . Namen je še bolj „zmesti“ klasifikator pri svojih odločitvah.

Ko dodamo označeni testni primer v učno množico, ponovno poženemo učni algoritem. Z novo hipotezo ponovno klasificiramo testni primer, tako pridobimo napovedi  $K_{r_1}$ ,  $K_{r_2}$  in  $K_{r_n}$  ter izmerimo spremembe teh napovedi. Mera bližine  $f$  med prvotno napovedjo  $K$  in transducirano napovedjo tvori oceno zanesljivosti:

$$TRANS_1 = f(K, K_{r_1}) \quad (3.17)$$

$$TRANS_2 = f(K, K_{r_2}) \quad (3.18)$$

$$TRANS_n = f(K, K_{r_n}) \quad (3.19)$$

## Poglavje 4

# Metodologija testiranja

Ocene zanesljivosti smo testirali na 20 standardnih testnih domenah in s sedmimi klasifikacijskimi modeli, opisanimi v 2. poglavju.

Testiranje ocen zanesljivosti je potekalo v obliki prečnega preverjanja, na način *izloči-enega*. Vsakemu učnemu primeru, ki je bil izključen iz učne množice, je bila izračunana napoved na danem modelu, nato pa še ocene zanesljivosti z razvitimi lokalnimi metodami ( $CNK$ ,  $LCV$ ,  $LCV_d$ ,  $DENS$ ), varianco bagginga ( $BAGV$ ) in vsemi tremi inačicami metode ocenjevanja z obrnjeno transdukcijo ( $TRANS_1$ ,  $TRANS_2$  in  $TRANS_n$ ). Vse metode, izvemši  $BAGV$  in  $DENS$ , so bile izračunane z uporabo 11 različnih funkcij bližine, opisanih v razdelku 3.1. Da ugotovimo, ali so razvite metode sploh kakorkoli boljše, potrebujemo referenčno oceno zanesljivosti. Vzeli smo kar klasifikatorjevo lastno napoved, kar je pravzaprav najbolj preprosta verjetnostna mera zanesljivosti klasifikacije. Večja kot je verjetnost, ki jo napove klasifikator za nek razred, bolj je prepričan, da je njegova napoved točna.

Ob zaključku prečnega preverjanja, so bile na ta način izračunane ocene zanesljivosti za vse primere iz podatkovne množice. Na vsakem koraku prečnega preverjanja je bila učna množica povrnjena v prvotno stanje, tako da se z vsakim korakom postopka ni spreminjala, oz. se iterativno širila z vsakim dodatnim učnim primerom.

### 4.1 Testne podatkovne množice

Testne podatkovne množice smo izbrali iz obsežnega repozitorija UCI Machine Learning Repository [1]. Uporabljene so bile domene s čim bolj raznolikih področij, kot so zdravstvo, ekologija, tehnične aplikacije in celo družboslovje. Tabela 4.1 prikazuje ključne lastnosti testnih domen. Po vrsti si sledijo število

učnih primerov, število diskretnih in zveznih atributov ter število možnih razredov odvisne spremenljivke.

	#primerov	#diskretnih atr.	#zveznih atr.	#razredov
housevotes	435	16	0	2
wine	178	0	13	3
parkinsons	195	0	22	2
zoo	101	16	0	7
tictactoe	958	9	0	2
postoperative	90	7	1	3
monks	432	6	0	2
iriset	150	0	4	3
glass	214	0	9	6
hungarian	294	7	6	2
ecoli	336	0	7	8
heart	270	12	1	2
haberman	306	3	0	2
flag	194	28	0	10
wdbc	569	0	30	2
bcw	699	9	0	2
sonar	208	0	60	2
soybean	47	21	0	4
hepatitis	155	14	5	2
lungcancer	32	56	0	3

Tabela 4.1: Lastnosti uporabljenih testnih domen.

## 4.2 Statistične metode

Dobljene ocene zanesljivosti na testnih podatkovnih množicah kvalitativno in kvantitativno ovrednotimo z uporabo statističnih metod. Izračunane ocene zanesljivosti primerjamo s klasifikacijsko točnostjo, tj. verjetnostjo, ki jo klasifikator vrne za pravilni razred. Kvalitativne podatke pridobimo s *Q-Q diagrami*, kvantitativne vrednosti pa v obliki Pearsonovih korelacijskih koeficientov, katere podkrepimo z Wilcoxonovim parnim testom.

### 4.2.1 Q-Q diagrami

Pri opisni statistiki so Q-Q diagrami grafična metoda za grafično analizo verjetnostnih distribucij populacij. V osnovi se uporabljajo za grafično oceno prileganja poljubne populacije normalni porazdelitvi. Metoda nam ravno tako omogoča primerjavo dveh poljubnih distribucij, zato z njo ocenjujejo podobnost med ocenami zanesljivosti in točnostjo klasifikacij.

Pri teh diagramih so eden ob drugega postavljeni kvantili dotičnih verjetnostnih distribucij. Če se populaciji ujemata v svojih porazdelitvah, se to na diagramu vidi kot premica  $y = x$ . V paketu R se diagram prikliče z ukazom `qqplot(x, y)` [9].

### 4.2.2 Pearsonovi korelacijski koeficienti

Korelacijski koeficient (korelacija) je številska mera, ki predstavlja moč linearne povezanosti dveh spremenljivk in tako označuje odvisnost dveh spremenljivk v statističnih populacijah. Pearsonove korelacijske koeficiente računamo na podlagi kovariance in standardnih odklonov zavzetih vrednosti opazovanih spremenljivk.

Korelacijski koeficienti so primerni za kvantitativno primerjavo in vrednotenje povezanosti izračunanih ocen zanesljivosti in izmerjene klasifikacijske točnosti. Za nas jih, v paketu R, izračuna standardna funkcija `cor.test`; potrebno ji je le podati parameter `method='pearson'` [9].

### 4.2.3 Wilcoxon–Mann–Whitney test

S tem testom preverimo tisto, kar nas v osnovi najbolj zanima — ali se ocene zanesljivosti pravilno klasificiranih testnih primerov ločijo od ocen zanesljivosti napačno klasificiranih primerov.

Wilcoxon–Mann–Whitney test je neparametričen statistični test, uporablja pa se za testiranje ničelne hipoteze o enakosti aritmetičnih sredin dveh populacij (nasprotna hipoteza je, da sta različni). Nadomešča klasični parni t-test, saj ga uporabljamo takrat, ko ni izpolnjen pogoj normalnosti primerjanih populacij (kar ugotovimo v 5.1). Wilcoxon–Mann–Whitney test predpostavlja, da sta populaciji zvezni ter da so vzorci naključni in neodvisni [11]. V paketu R ta test izvedemo s pomočjo funkcije `wilcox.test(pos, neg)` [11].



## Poglavje 5

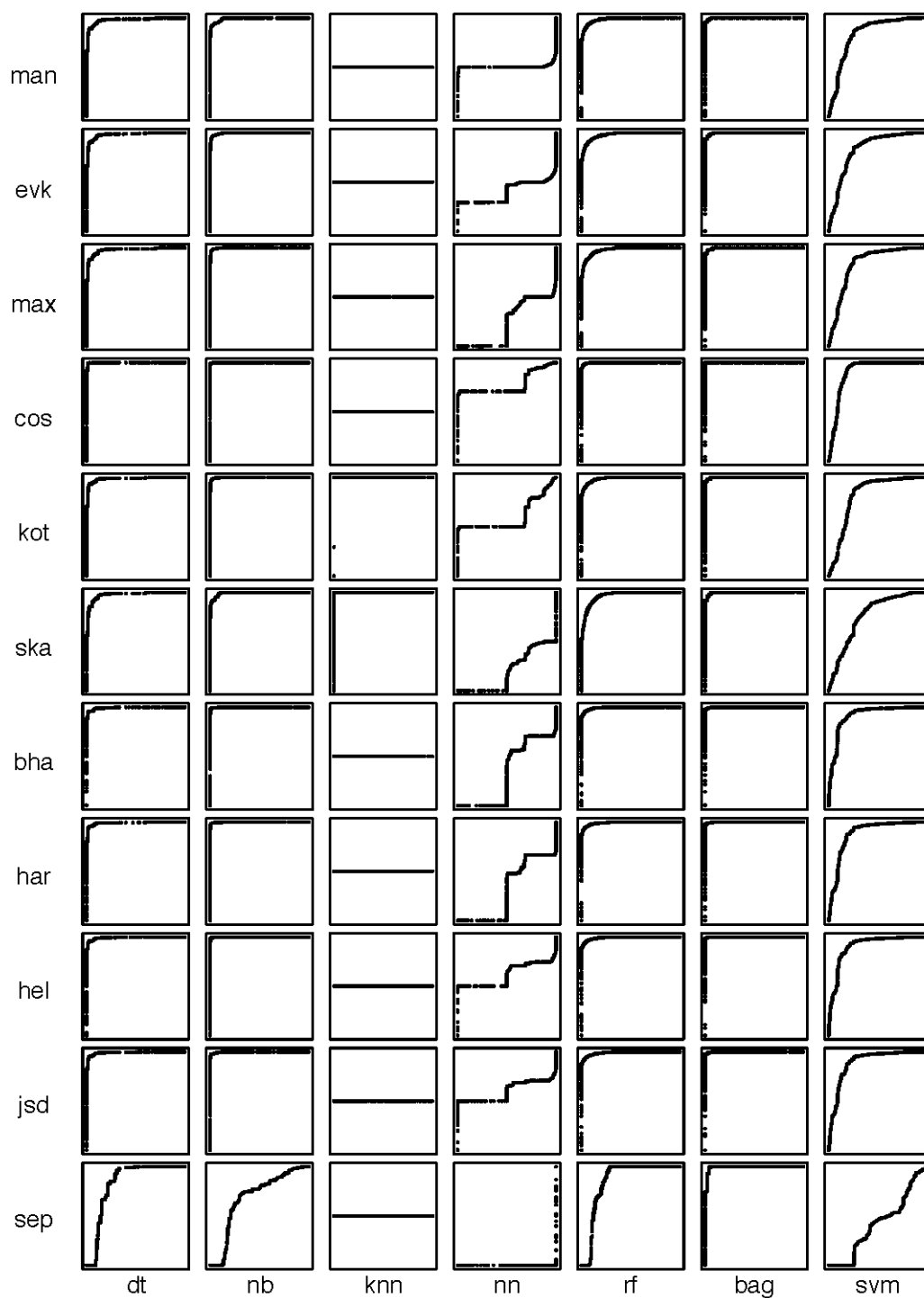
# Rezultati testiranj ocen zanesljivosti

Namen naših testov je kvalitativno in kvantitativno ovrednotiti razvite metode ocenjevanja zanesljivosti. Na dvajsetih testnih učnih množicah izračunamo ocene zanesljivosti vseh napovedi klasifikatorjev DT, NB, KNN, NN, RF, BAG in SVM ter hkrati zabeležimo klasifikatorjevo lastno zanesljivost (verjetnost najbolj verjetnega razreda). Dobljene vrednosti podvržemo statističnim testom in tako pridobimo osnovne kvalitativne in kvantitativne podatke za primerjavo in vrednotenje.

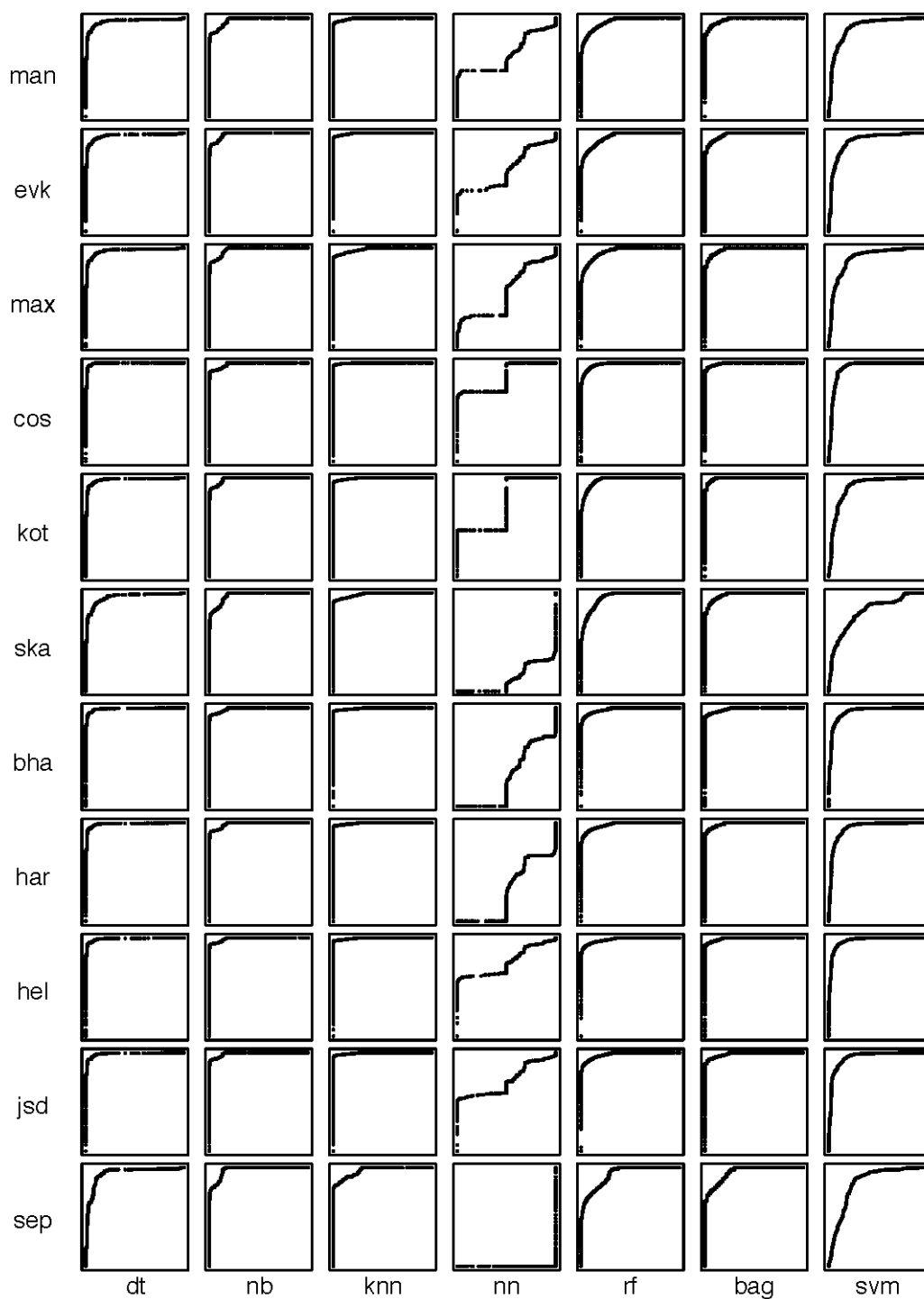
Najprej si ogledamo Q-Q diagrame vseh implementiranih ocen zanesljivosti in ocenimo podobnost distribucij točnosti klasifikacij ter ocen zanesljivosti. Sledi povzetek in interpretacija Pearsonovih korelacijskih koeficientov in rezultatov Wilcoxonovega parnega testa. Razvite metode primerjamo z metodo obrnjene transdukcije in z referenčno oceno ovrednotimo možnosti uporabe na uporabljenih problemskih domenah.

### 5.1 Q-Q diagrami

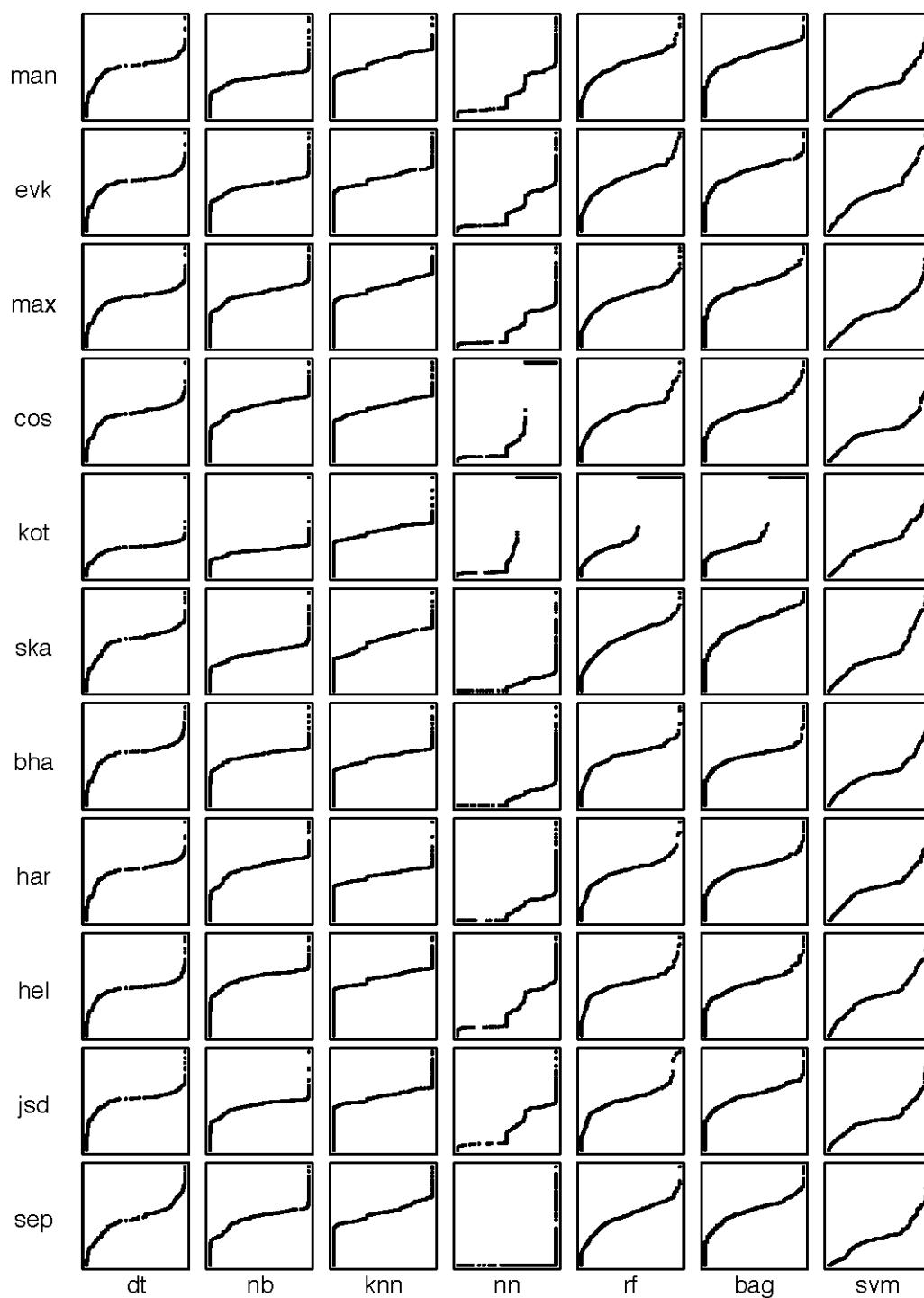
Na sledečih Q-Q diagramih leži klasifikacijska točnost na abscisi, izračunane ocene zanesljivosti pa na ordinati. Podatki domen so združeni za vsak posa-



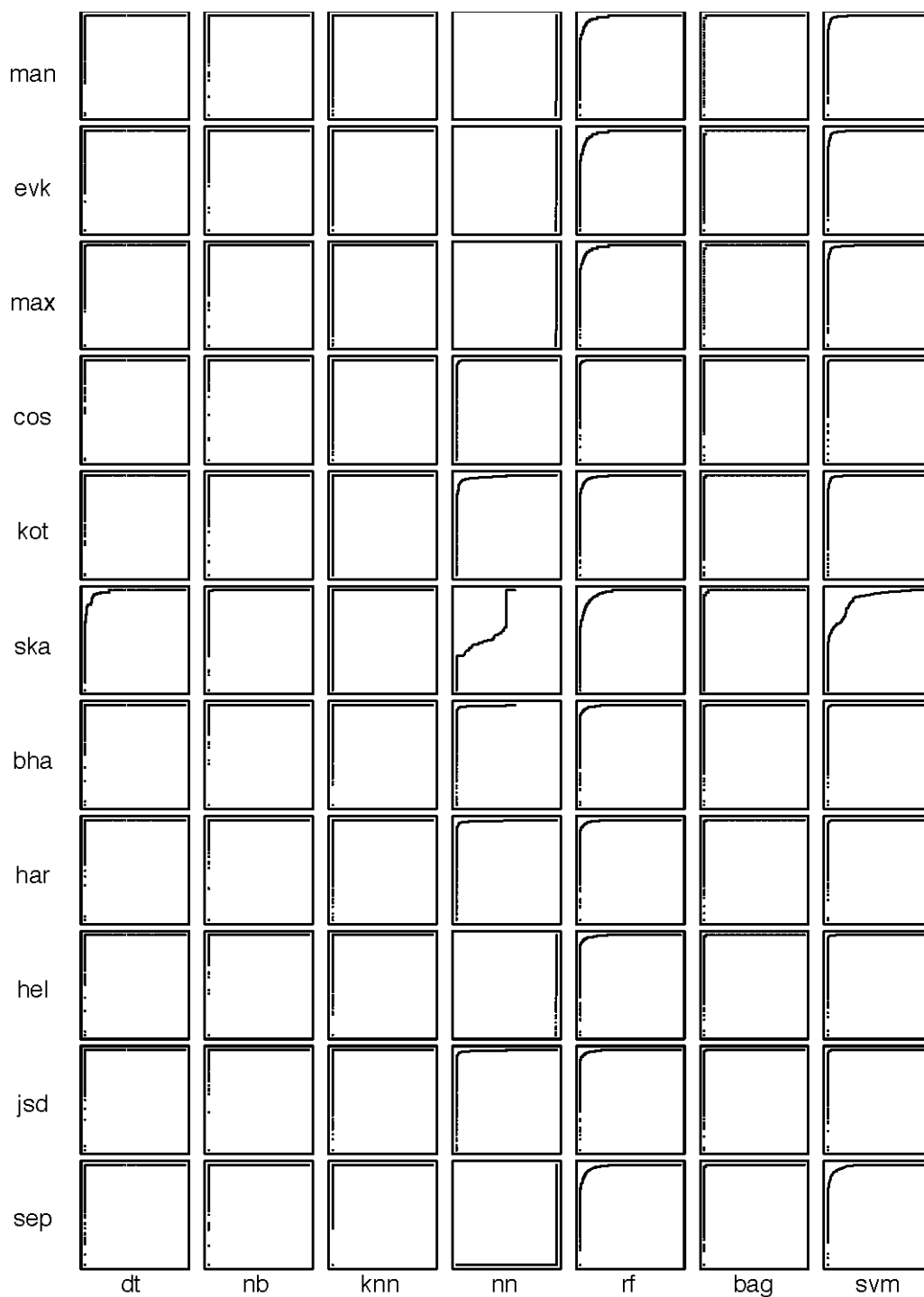
Slika 5.1: Q-Q diagram ocene zanesljivosti  $CNK$  in klasifikacijske napake — z različnimi merami, po klasifikatorjih. Podatki domen so združeni po klasifikatorjih.



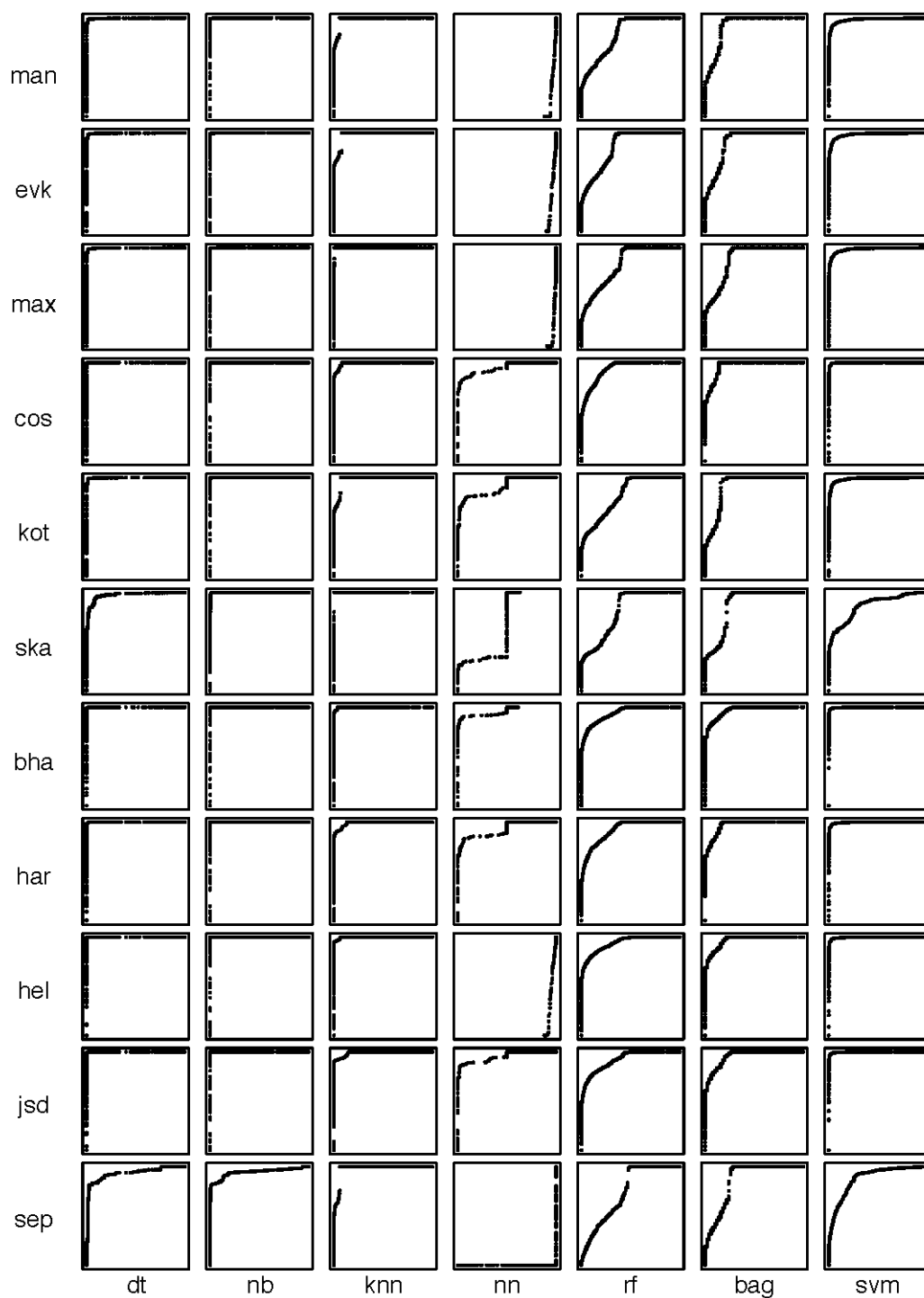
Slika 5.2: Q-Q diagram ocene zanesljivosti *LCV* in klasifikacijske napake — z različnimi merami, po klasifikatorjih. Podatki domen so združeni po klasifikatorjih.



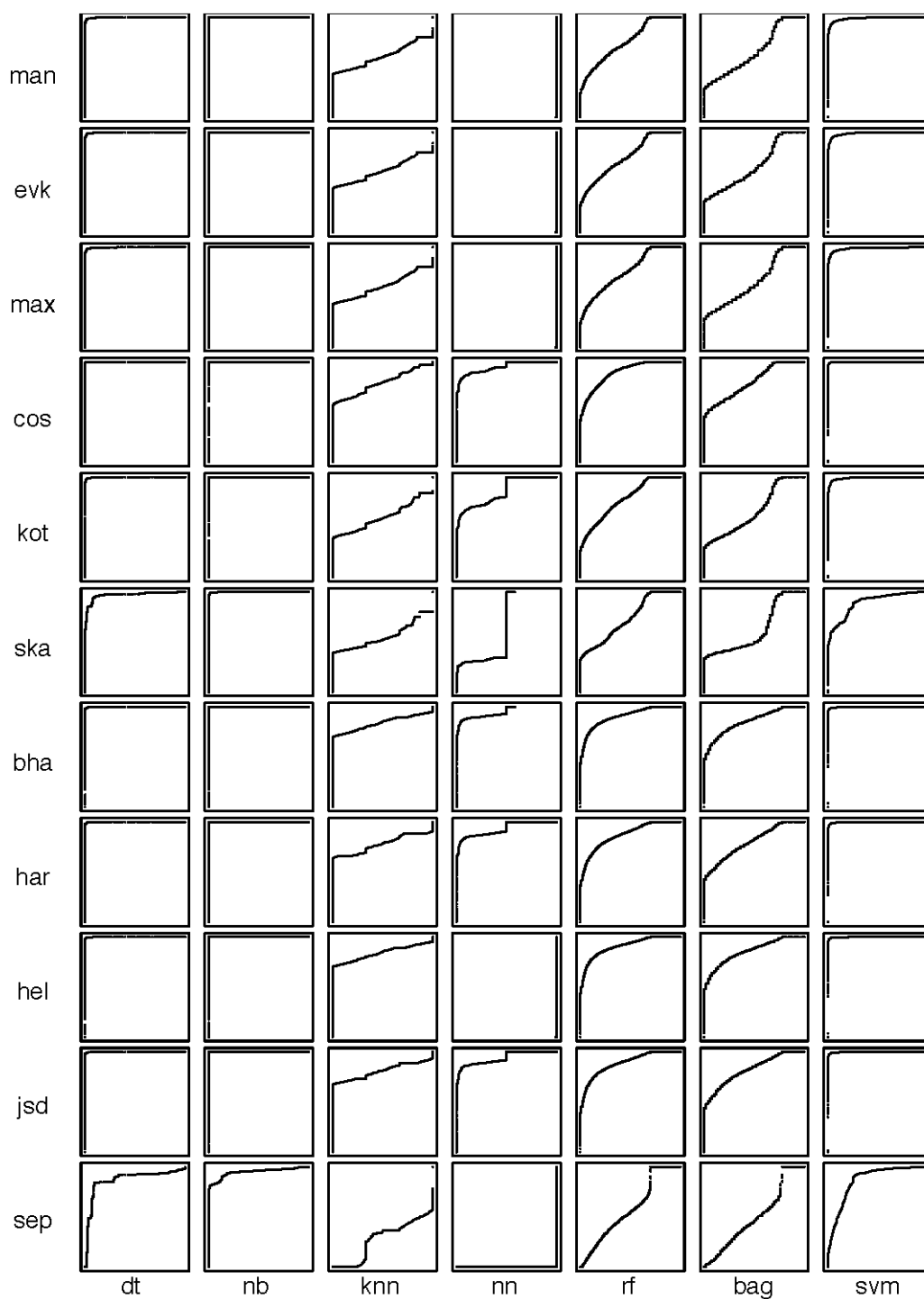
Slika 5.3: Q-Q diagram ocene zanesljivosti  $LCV_d$  in klasifikacijske napake — z različnimi merami, po klasifikatorjih. Podatki domen so združeni po klasifikatorjih.



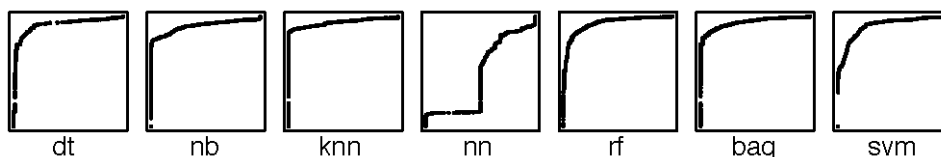
Slika 5.4: Q-Q diagram ocene zanesljivosti  $TRANS_1$  in klasifikacijske napake — z različnimi merami, po klasifikatorjih. Podatki domen so združeni po klasifikatorjih.



Slika 5.5: Q-Q diagram ocene zanesljivosti  $TRANS_2$  in klasifikacijske napake — z različnimi merami, po klasifikatorjih. Podatki domen so združeni po klasifikatorjih.

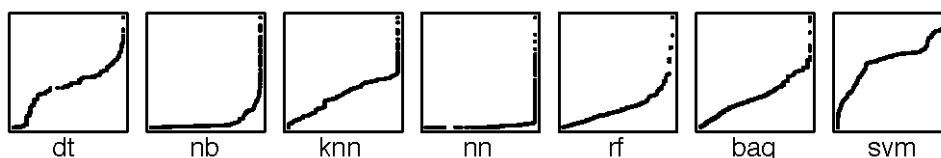


Slika 5.6: Q-Q diagram ocene zanesljivosti  $TRANS_n$  in klasifikacijske napake — z različnimi merami, po klasifikatorjih. Podatki domen so združeni po klasifikatorjih.



Slika 5.7: Q-Q diagram ocene zanesljivosti *DENS*

mezni klasifikator in prikazani za vsako mero posebej.



Slika 5.8: Q-Q diagram ocene zanesljivosti *BAGV*

Na prvi pogled opazimo da pretežno malo diagramov izkazuje podobnost distribucije ocen zanesljivosti in točnosti klasifikacij. Zaradi tega pri testiranju ločljivosti množic ocen pravilno in nepravilno klasificiranih primerov ne moremo uporabljati klasičnega parnega t-testa, marveč uporabimo Wilcoxon–Mann–Whitney test.

Na Q-Q diagramih ocen metode *CNK* (slika 5.1) opazimo določeno stopnjo prileganja opazovanih distribucij z uporabo separacijske razdalje (enačba 3.10) pri klasifikatorjih DT, NB, RF in SVM. Določeno manjšo stopnjo prileganja zaznamo v celotnem stolpcu klasifikatorja SVM, v stolpcu nevronske mreže pa vidimo zanimiv vzorec klasifikatorjevega obnašanja.

Diagrami ocen metode *LCV* (slika 5.2) v povprečju ne izkazujejo očitnih podobnosti. Poleg obnašanja nevronske mreže vidimo v stolpcih modelov RF in SVM nekaj znakov ujemanja, največ z uporabo skalarnega produkta (enačba 3.7) in separacijske razdalje.

Pri ocenah  $LCV_d$  izkazujejo Q-Q diagrami največji odstotek prileganj vseh ocen na vseh klasifikatorjih. Dobro prileganje vidimo v stolpcu SVM, najboljše izmed kombinacij pa pri kombinaciji modela SVM z mero na osnovi evklidske razdalje (enačba 3.2).

Q-Q diagrami ocen metode  $TRANS_1$  (slika 5.4) so skoraj identični diagramom metode *LCV*: nekaj podobnosti zaznamo v stolpcih RF in SVM ter v vrsticah skalarnega produkta in separacijske razdalje.

Pri diagramih metode  $TRANS_2$  (slika 5.5) so omembe vredni diagrami

pri modelih RF in BAG, ponovno pa vidimo odziv modela SVM z merama skalarnega produkta in separacijske razdalje.

Na Q-Q diagramu ocen  $TRANS_n$  (slika 5.6) vidimo odziv pri polovici diagramov. Diagram je podoben diagramu metode  $TRANS_2$ , vendar s to razliko, da vidimo odziv v celotnem stolpcu metode KNN.

Vsi diagrami ocen metode  $DENS$  (slika 5.7) le opazno podpirajo osnovno predpostavko te metode (razdelek 3.4), vendar ne signifikantno.

Diagrami ocen metode  $BAGV$  (slika 5.8) izkazujejo zanimanja vredne podobnosti. Diagrami modelov KNN in BAG so precej podobni premicam, za ogled so zanimivi tudi modeli DT, RT in SVM.

## 5.2 Rezultati statističnih testov

[%]	DT	NB	KNN	NN	RF	BAG	SVM	povprečje
<i>LCV</i>	80/3	76/0	<b>81/1</b>	43/11	65/1	69/1	50/3	<b>66/3</b>
<i>BAGV</i>	<b>85/5</b>	70/0	<b>80/0</b>	40/10	75/0	<b>80/0</b>	35/40	66/8
<i>TRANS<sub>1</sub></i>	39/0	64/0	80/5	19/5	<b>78/0</b>	<b>80/0</b>	47/30	58/6
<i>CNK</i>	72/5	67/6	10/0	55/14	60/2	67/0	<b>70/7</b>	57/5
<i>TRANS<sub>2</sub></i>	59/4	43/6	28/10	17/10	14/43	9/32	36/22	29/18
<i>LCV<sub>d</sub></i>	30/19	38/10	21/20	35/20	10/30	12/24	22/13	24/19
<i>TRANS<sub>n</sub></i>	54/2	37/5	30/33	5/7	16/43	11/40	29/22	26/22
<i>DENS</i>	15/10	10/20	10/10	30/15	10/20	15/25	15/0	15/14
referenca	60/5	<b>80/5</b>	60/0	<b>80/0</b>	60/0	<b>80/0</b>	45/40	66/7

Tabela 5.1: Povprečni odstotki značilnih pozitivnih in negativnih korelacij med *ocenami zanesljivosti* in *klasifikacijsko točnostjo*.

V našem primeru je značilna korelacija tista, kjer je  $p$ -vrednost manjša od 0.05 tako za Pearsonov koeficient korelacije, kot za Wilcoxonov test. Podrobni rezultati za vsak par ocena – domena se nahajajo v tabelah dodatka A. Povzetki rezultatov vsake ocene so prikazani v tabelah 5.2 in 5.3, povzetki ocen parov ocena/klasifikator pa v tabeli 5.4. Zgornja sumarna tabela 5.1 prikazuje ključne povzetke vseh rezultatov, na podlagi katerih lahko enostavno rangiramo testirane metode ocenjevanja zanesljivosti.

Vse tabele prikazujejo uspešnost ocen glede na odstotek domen, pri katerih je določena ocena dosegla statistično značilno pozitivno ali negativno korelacijo. Rezultati potrjujejo naša pričakovanja o smeri korelacije med razvitimi ocenami zanesljivosti in klasifikacijsko točnostjo. Pri veliki večini testov prevladuje število značilno pozitivnih korelacij. Najboljše povprečne rezultate izkazuje razvita metoda *LCV*, saj je dosegla 66% pozitivnih in zgolj 3% negativnih primerov korelacije med oceno zanesljivosti in točnostjo klasifikacij. Zelo blizu ji sledi ocena *BAGV* z enakim odstotkom pozitivnih testov (66%) vendar nekoliko višjim odstotkom negativnih (8%).

Najslabše rezultate beleži metoda *DENS*, kjer rezultati izkazujejo zgolj 15% odstotkov pozitivnih korelacij in skoraj enak odstotek negativnih (14%).

Izmed metod ocenjevanja zanesljivosti na osnovi obratne transdukcije, je najboljša metoda *TRANS<sub>1</sub>* na tretjem mestu. Z rezultatom 58 % pozitivnih in 6% negativnih korelacij je primerljiva z oceno *CNK* (57/5). Metodi *TRANS<sub>2</sub>*

in  $TRANS_n$  sta se obnesli precej slabše (29/18 in 26/22, ustrezno) in izkazuje podobne rezultate kot metoda  $LCV_d$  (24/19).

Odebeljene vrednosti v tabeli 5.1 označujejo kombinacije ocena/model, katere v povprečju izkazuje boljše rezultate kot referenčne ocene zanesljivost posameznih klasifikatorjev. Rezultati prikazujejo, da je najboljša ocena zanesljivosti za odločitvena drevesa  $BAGV$  (85/5 proti referenčni oceni DT 60/5), za metodo najbližjih sosedov imata oceni  $LCV$  in  $BAGV$  ekvivalenten rezultat (81/1 in 80/0 proti referenci 60/0), za naključna drevesa ocena  $TRANS_1$  (78/0 proti 60/0) in za metodo podpornih vektorjev ocena  $CNK$  (70/7 proti 45/40). Za ostale modele, tj. za naivnega Bayesa, umetne nevronske mreže in model bagging, razvite ocene niso presegle rezultatov klasifikatorjevih lastnih ocen zanesljivosti.

Tabela 5.4 prikazuje odstotke povprečnih pozitivnih in negativnih korelacij, združenih po testnih podatkovnih množicah. V tej tabeli lahko iz vrstic razberemo, katera ocena je bila najbolj uspešna pri določeni testni podatkovni množici, zato smo te vrednosti odebelili. Če pogledamo kje ležijo te vrednosti po stolpcih, vidimo najboljše rezultate metod  $CNK$ ,  $LCV$ ,  $BAGV$  in  $TRANS_1$ . V tabeli ravnotako vidimo, da pri različnih podatkovnih množicah različne ocene zanesljivosti dosegajo boljše rezultate kakor druge. Izstopajoča vrednost je visok rezultat metode  $DENS$  na množici **haberman**, kjer je ta dosegla 86% značilnih pozitivnih korelacij in 0 % negativnih. Če bolj podrobno pogledamo vrstico testne množice, vidimo, da so ostale metode dosegale precej nižje rezultate kot povprečno.

V tabelah podrobnih rezultatov v dodatku A lahko najdemo ostale podatke. Na primer iz tabele A.34 lahko za prej omenjeno izstopajočo vrednost ocene  $DENS$  naknadno izračunamo dosežen povprečni koeficient korelacije, ki ima v tem primeru vrednost 0.21.

Podrobni rezultati kontrolnega para ocene  $CNK$  in klasifikatorja KNN (tabela A.1) prikazujejo da v mnogih primerih statistične metode odpovedo zaradi enakosti opazovanih meritev (izmerjena standardna deviacija je 0) in potrdijo naš razmislek o neprimernosti ocenjevanja zanesljivosti dotičnega para ocena/klasifikator. (razdelek 3.2).

Ob pregledu rezultatov tabele 5.2 smo najprej opazili izstopajoče vrednosti pri modelu KNN, v vrstici ocene  $CNK$ ska; v bolj podrobni tabeli A.6 pa nas zmoti stolpec modela KNN. Statistična metoda za izračun korelacijskih koeficientov v teh primerih pri merah kosinus in skalarni produkt ni odpovedala (kot je pričakovano), zato je primer vreden dodatne pozornosti. Doda-

Poglavje 5: Rezultati testiranj ocen zanesljivosti

[%]	DT	NB	KNN	NN	RF	BAG	SVM	povprečje
<i>CNK</i> man	80/5	75/5	0/0	55/5	70/0	75/0	75/0	<b>61/2</b>
<i>CNK</i> evc	80/5	75/5	0/0	60/0	70/0	75/0	75/0	<b>62/1</b>
<i>CNK</i> max	80/5	75/5	0/0	65/5	70/0	70/0	75/0	<b>62/2</b>
<i>CNK</i> cos	80/5	75/0	0/0	30/35	75/0	75/0	75/5	<b>59/6</b>
<i>CNK</i> kot	80/5	75/0	35/0	30/35	75/0	75/0	65/10	<b>62/7</b>
<i>CNK</i> ska	85/5	80/5	80/5	75/0	80/0	80/0	85/5	<b>81/3</b>
<i>CNK</i> bha	70/0	65/5	0/0	65/5	45/0	55/0	65/15	<b>52/4</b>
<i>CNK</i> har	75/0	70/5	0/0	70/5	50/0	70/0	75/5	<b>59/2</b>
<i>CNK</i> hel	70/0	65/5	0/0	65/5	45/0	55/0	65/15	<b>52/4</b>
<i>CNK</i> jsd	75/0	70/5	0/0	65/5	50/0	70/0	70/10	<b>57/3</b>
<i>CNK</i> sep	15/25	15/30	0/0	20/50	30/20	40/5	45/15	<b>24/21</b>
<i>LCV</i> man	80/0	75/0	85/0	30/20	70/0	70/0	55/5	<b>66/4</b>
<i>LCV</i> evc	80/0	75/0	85/0	30/15	65/0	70/0	55/0	<b>66/2</b>
<i>LCV</i> max	75/0	75/0	85/0	20/15	65/0	70/5	55/5	<b>64/4</b>
<i>LCV</i> cos	85/0	80/0	75/0	25/20	70/0	70/0	55/0	<b>66/3</b>
<i>LCV</i> kot	80/0	75/0	85/0	30/20	70/0	80/0	55/10	<b>68/4</b>
<i>LCV</i> ska	85/5	85/0	85/0	75/0	80/0	80/0	50/5	<b>77/1</b>
<i>LCV</i> bha	80/5	75/0	80/5	70/0	60/0	65/0	45/0	<b>68/1</b>
<i>LCV</i> har	80/5	75/0	80/0	70/0	70/0	70/0	45/0	<b>70/1</b>
<i>LCV</i> hel	80/5	75/0	80/5	55/15	60/0	65/0	45/0	<b>66/4</b>
<i>LCV</i> jsd	80/5	75/0	80/5	45/15	65/0	70/0	45/0	<b>66/4</b>
<i>LCV</i> sep	75/5	75/0	70/0	25/5	35/10	50/5	45/5	<b>54/4</b>
<i>LCV<sub>d</sub></i> man	35/10	40/10	35/10	30/35	10/25	15/15	25/10	<b>27/16</b>
<i>LCV<sub>d</sub></i> evc	30/10	40/10	25/15	20/30	10/25	15/15	20/15	<b>23/17</b>
<i>LCV<sub>d</sub></i> max	30/10	40/10	30/10	20/20	10/25	15/15	20/15	<b>24/15</b>
<i>LCV<sub>d</sub></i> cos	30/25	35/10	15/30	25/45	5/35	5/25	25/10	<b>20/26</b>
<i>LCV<sub>d</sub></i> kot	40/5	45/10	40/5	25/15	10/10	10/5	30/10	<b>29/9</b>
<i>LCV<sub>d</sub></i> ska	45/5	50/10	40/5	70/5	30/5	30/5	30/15	<b>42/7</b>
<i>LCV<sub>d</sub></i> bha	20/35	35/10	5/35	50/5	5/40	5/40	20/15	<b>20/26</b>
<i>LCV<sub>d</sub></i> har	30/30	35/10	5/30	70/0	5/40	5/35	25/10	<b>25/22</b>
<i>LCV<sub>d</sub></i> hel	20/35	35/10	5/35	20/30	5/40	5/40	20/15	<b>16/29</b>
<i>LCV<sub>d</sub></i> jsd	15/35	15/15	5/35	25/35	5/50	5/40	15/15	<b>12/32</b>
<i>LCV<sub>d</sub></i> sep	40/5	45/5	25/10	25/5	10/35	20/25	15/10	<b>26/14</b>
<i>DENS</i>	15/10	10/20	10/10	30/15	10/20	15/25	15/0	<b>15/14</b>
<i>BAGV</i>	85/5	70/0	80/0	40/10	75/0	80/0	35/40	<b>66/8</b>

Tabela 5.2: Odstotek značilnih pozitivnih in negativnih korelacij med *ocenami zanesljivosti* in *klasifikacijsko točnostjo*.

[%]	DT	NB	KNN	NN	RF	BAG	SVM	povprečje
<i>TRANS<sub>1</sub></i> man	45/0	70/0	80/5	25/5	80/0	80/0	50/30	<b>61/6</b>
<i>TRANS<sub>1</sub></i> evc	45/0	70/0	80/5	20/5	80/0	80/0	50/30	<b>61/6</b>
<i>TRANS<sub>1</sub></i> max	45/0	70/0	80/5	20/5	80/0	80/0	50/30	<b>61/6</b>
<i>TRANS<sub>1</sub></i> cos	35/0	55/0	80/5	15/5	80/0	80/0	50/30	<b>56/6</b>
<i>TRANS<sub>1</sub></i> kot	45/0	75/0	80/5	20/5	80/0	80/0	50/30	<b>61/6</b>
<i>TRANS<sub>1</sub></i> ska	60/5	60/0	80/5	25/0	85/0	80/0	40/35	<b>61/6</b>
<i>TRANS<sub>1</sub></i> bha	25/0	60/0	80/5	20/5	75/0	80/0	50/30	<b>56/6</b>
<i>TRANS<sub>1</sub></i> har	30/0	55/0	80/5	20/5	80/0	80/0	50/30	<b>56/6</b>
<i>TRANS<sub>1</sub></i> hel	25/0	60/0	80/5	20/5	75/0	80/0	50/30	<b>56/6</b>
<i>TRANS<sub>1</sub></i> jsd	30/0	55/0	80/5	20/5	75/0	80/0	50/30	<b>56/6</b>
<i>TRANS<sub>1</sub></i> sep	40/0	70/0	80/5	0/10	65/0	80/0	25/20	<b>51/5</b>
<i>TRANS<sub>2</sub></i> man	70/0	50/0	25/5	20/10	10/40	10/35	40/20	<b>32/16</b>
<i>TRANS<sub>2</sub></i> evc	70/0	50/0	25/5	20/10	10/40	10/35	40/20	<b>32/16</b>
<i>TRANS<sub>2</sub></i> max	70/0	50/0	25/5	20/10	10/40	10/35	40/20	<b>32/16</b>
<i>TRANS<sub>2</sub></i> cos	55/10	45/0	35/5	15/10	15/30	10/35	45/20	<b>31/16</b>
<i>TRANS<sub>2</sub></i> kot	60/10	55/0	35/0	15/10	15/30	10/30	45/20	<b>34/14</b>
<i>TRANS<sub>2</sub></i> ska	60/10	60/0	50/5	20/5	45/15	25/10	40/35	<b>43/11</b>
<i>TRANS<sub>2</sub></i> bha	55/0	40/5	20/15	15/10	10/55	0/35	30/20	<b>24/20</b>
<i>TRANS<sub>2</sub></i> har	60/0	40/5	25/15	15/10	10/55	10/35	30/20	<b>27/20</b>
<i>TRANS<sub>2</sub></i> hel	55/0	40/5	20/15	15/10	10/55	0/35	30/20	<b>24/20</b>
<i>TRANS<sub>2</sub></i> jsd	60/5	40/5	20/15	15/10	10/55	10/35	30/20	<b>26/21</b>
<i>TRANS<sub>2</sub></i> sep	35/5	5/50	25/20	15/20	5/55	5/35	30/25	<b>17/30</b>
<i>TRANS<sub>n</sub></i> man	60/0	40/0	30/35	5/5	10/45	5/40	35/20	<b>26/21</b>
<i>TRANS<sub>n</sub></i> evc	60/5	40/0	20/35	5/10	10/60	0/50	30/20	<b>24/26</b>
<i>TRANS<sub>n</sub></i> max	60/0	40/0	30/35	5/5	10/45	5/40	25/20	<b>25/21</b>
<i>TRANS<sub>n</sub></i> cos	45/5	40/0	35/20	5/10	35/35	25/40	40/25	<b>32/19</b>
<i>TRANS<sub>n</sub></i> kot	50/5	45/0	35/20	10/10	35/35	25/40	40/25	<b>34/19</b>
<i>TRANS<sub>n</sub></i> ska	55/5	60/0	55/15	10/0	55/25	35/30	40/35	<b>44/16</b>
<i>TRANS<sub>n</sub></i> bha	55/0	35/5	25/40	5/5	5/45	5/40	20/20	<b>21/22</b>
<i>TRANS<sub>n</sub></i> har	55/0	35/5	25/40	5/5	5/45	5/40	20/20	<b>21/22</b>
<i>TRANS<sub>n</sub></i> hel	55/0	35/5	25/40	5/5	5/45	5/40	20/20	<b>21/22</b>
<i>TRANS<sub>n</sub></i> jsd	55/0	35/5	25/40	5/5	5/45	10/40	20/20	<b>22/22</b>
<i>TRANS<sub>n</sub></i> sep	40/5	5/30	20/45	0/15	5/50	0/40	25/20	<b>14/29</b>

Tabela 5.3: Odstotek značilnih pozitivnih in negativnih korelacij med *ocenami zanesljivosti* in *klasifikacijsko točnostjo*.

tna analiza posameznih testnih primerov pri kombinaciji *CNK*/KNN pokaže da dotični meri preveč ohlapno sledita definiciji razdalje oz. njenega komplementa. Izračun ocene *CNK* z mero na osnovi skalarnega produkta je v tem primeru  $f_{ska}(P, P) = \sum_i p_i^2$ . Na preprostem primeru demonstriramo vzrok odstopanja rezultatov: naj bo  $(0.9, 0.1)$  rezultat metode KNN za nek testni primer, ki pripada prvemu razredu (napoved je pravilna). Točnost klasifikacije je v tem primeru 0.9, ocena *CNK*ska pa  $0.9^2 + 0.1^2 = 0.82$ . Zaradi takšnih odstopanj statistična metoda izračuna korelacijskih koeficientov ni javila napake, temveč nas je le na svojevrsten način opozorila na nepričakovano obnašanje uporabljenih mer.

	[%]	CNK	LCV	LCV <sub>a</sub>	DENS	BAGV	TRANS <sub>1</sub>	TRANS <sub>2</sub>	TRANS <sub>n</sub>	Ref.
housevotes	82/6	78/4	26/44	14/29	<b>86/0</b>	57/13	60/16	34/16	71/14	71/14
wine	78/0	96/0	60/0	0/0	57/14	77/13	51/10	42/5	<b>100/0</b>	<b>100/0</b>
parkinsons	<b>75/1</b>	73/0	66/3	0/57	86/14	53/13	8/35	8/39	71/14	71/14
zoo	84/0	97/0	9/27	14/0	71/14	96/0	52/1	69/0	<b>100/0</b>	<b>100/0</b>
tictactoe	78/6	29/22	21/26	0/43	<b>86/0</b>	58/14	22/55	21/57	86/14	86/14
postoperative	0/17	<b>17/5</b>	0/0	0/0	14/14	<b>13/1</b>	0/6	1/5	0/14	0/14
monks3	19/17	<b>38/6</b>	23/9	0/0	43/0	21/16	27/17	26/19	29/29	29/29
iriset	83/0	96/0	4/6	0/29	86/14	99/0	44/4	23/3	<b>100/0</b>	<b>100/0</b>
glass	75/0	<b>82/0</b>	13/0	0/0	57/14	68/1	3/26	26/12	71/0	71/0
hungarian	39/12	<b>84/3</b>	4/18	14/0	71/14	57/0	26/8	10/36	71/0	71/0
ecoli	82/1	91/0	36/0	29/0	71/14	86/0	48/3	44/4	<b>100/0</b>	<b>100/0</b>
heart	43/4	75/0	14/21	29/0	<b>86/0</b>	57/0	26/25	26/43	57/14	57/14
haberman	23/19	66/0	0/73	<b>86/0</b>	71/14	57/0	32/1	19/21	57/0	57/0
flag	<b>78/0</b>	68/3	51/1	0/29	71/0	73/0	27/38	39/16	71/0	71/0
wdbc	74/1	91/0	62/6	14/43	<b>100/0</b>	65/14	30/18	30/56	71/14	71/14
bcw	79/5	90/10	4/81	14/14	86/14	61/0	75/3	45/3	<b>86/0</b>	<b>86/0</b>
sonar	70/6	55/4	42/34	0/14	<b>86/0</b>	51/14	18/42	17/53	71/14	71/14
soybean	16/0	14/0	29/0	14/0	14/0	<b>39/0</b>	13/14	14/1	29/0	29/0
hepatitis	68/1	<b>82/0</b>	14/38	71/14	86/14	71/14	25/38	25/47	86/14	86/14
lungcancer	0/1	<b>5/0</b>	0/0	0/14	0/0	0/0	1/4	0/0	0/0	0/0

Tabela 5.4: Odstotek značilnih pozitivnih in negativnih korelacij med ocenami zanesljivosti in klasifikacijsko točnostjo predstavljeni po domenah.



# Poglavje 6

## Diskusija in zaključki

Pojem zanesljivost pri strojnem učenju pomeni točnost, tj. sposobnost podajanja točne napovedi. Ocene zanesljivosti posameznih napovedi nam lahko nudijo dodatno koristno informacijo, ki je lahko v pomoč pri presoji o uporabnosti posameznih napovedi. Razvili smo ocene zanesljivosti posameznih klasifikacijskih napovedi, ki temeljijo na lokalnih metodah in varianci bagginga.

Iz lokalnih metod za ocenjevanje zanesljivosti regresijskih napovedi smo razvili metode *CNK*, *LCV*, *DENS* in *BAGV* za ocenjevanje zanesljivosti posameznih napovedi pri klasifikaciji. Naredili smo primerjavo novih metod z nekoliko starejšo metodo ocenjevanja zanesljivosti z obratno transdukcijo – *TRANS*.

Metode smo preizkusili na dvajset umetnih in realnih problemskih domenah, uporabili smo sedem klasifikacijskih metod, ocene pa smo tvorili z enajstimi različnimi merami. Vse metode so bile implementirane v okviru programskega okolja R.

Največja ovira pri razvoju in izvajanju testov je bila časovna zahtevnost. Izračun vseh ocen zanesljivosti je trajal približno pet dni na 2.4 GHz procesorju Intel Core Duo. Vsaka pozno odkrita napaka programske kode je bila zato časovno zelo draga.

### 6.1 Pregled rezultatov

Dobljene ocene smo analizirali s statističnimi metodami in ugotovili, da opravljeni testi ne dajo popolnoma jasnih rezultatov. Opisna statistična metoda Q-Q diagramov v mnogih primerih ne podpira izmerjene korelacije.

Kvalitativne meritve smo pridobili tako, da smo izračunali Pearsonove korelacijske koeficiente med ocenami zanesljivosti in točnostjo napovedi ter da smo

z Wilcoxonovim parnim testom preizkusili ločljivost ocen zanesljivosti pravilno in napačno klasificiranih testnih primerov (razdeleka 4.2.2 in 4.2.3).

Kvantitativni rezultati kažejo, da je v povprečju najboljša razvita metoda  $LCV$ , saj je dosegla 66% pozitivnih in zgolj 3% negativnih primerov korelacije med oceno zanesljivosti in točnostjo klasifikacij. Po vrsti ji sledijo  $BAGV$  (66/8),  $TRANS_1$  (58/6) in  $CNK$  (57/5). Ostale metode so dosegle bistveno slabše rezultate:  $TRANS_2$  (29/18),  $TRANS_n$  (26/22),  $LCV_d$  (24/19) in  $DENS$  (15/14).

Tudi ko uporabljamo različne ocene sprememb porazdelitev verjetnosti, je pričakovano, da se v različnih domenah različno obnašajo. Enkrat se obnesejo boljše metode, drugič slabše — kakor je zapisano v [13]. Zaradi tega nas zmotijo nadpovprečno visoke stopnje korelacije natančnosti klasifikacije z mero na osnovi skalarnega produkta, zapazimo pa jih tudi pri ostalih pristopih ocenjevanja zanesljivosti (tabeli 5.2 in 5.3). Pri ocenah  $CNK$  in  $LCV_d$  dotična mera izkazuje približno 20% boljše rezultate, pri  $LCV$ ,  $TRANS_2$  in  $TRANS_n$  pa približno 10%.

Par klasifikator  $KNN$  in ocena  $CNK$  tvorita kontrolni test odziva uporabljenih mer bližine, saj oba ustvarjata napovedi na enak način. Za sumljivo mero skalarnega produkta smo pri analizi podatkov kontrolnega para ugotovili, da mera preveč ohlapno sledi definiciji komplementa razdalje.

Zaradi zgornjih argumentov sumimo v ustreznost uporabe skalarnega produkta v namene ocenjevanja zanesljivosti. Če iz meritev odstranimo dvomljivo mero, dobimo tabelo 6.1. Rangiranje se spremeni le modelu  $LCV_d$ , opazimo pa tudi, da se rezultati med prej podobnimi rezultati celo nekoliko strnejo in tako še podkrepijo nastali sum.

Izmed ocen na osnovi obratne transdukcije, je ocena  $TRANS_1$  v obeh obravnavanih primerih uvrščena na tretje mesto in po rezultatih testov tesno primerljiva z oceno  $CNK$ . Rezultati ocen  $TRANS_2$  in  $TRANS_n$  so slabši in primerljivi z rezultati ocen  $LCV_d$  in  $DENS$ .

Q-Q diagrami prikazujejo nekoliko drugačno sliko, saj prikazujejo podobnost verjetnostnih porazdelitev ocen zanesljivosti in klasifikacijske točnosti (razdelek 5.1). V povprečju ne izkazujejo močne podobnosti opazovanih distribucij. Največjo podobnost najdemo v kombinaciji ocene  $LCV_d$  in klasifikatorja  $SVM$  (slika 5.3), kar pa pri kvalitativnih rezultatih ni zaznavno.

[%]	DT	NB	KNN	NN	RF	BAG	SVM	povprečje
<i>LCV</i>	80/3	76/0	81/2	40/13	63/1	68/1	50/3	<b>65/3</b>
<i>BAGV</i>	<b>85/5</b>	70/0	<b>80/0</b>	40/10	75/0	<b>80/0</b>	35/40	66/8
<i>TRANS<sub>1</sub></i>	37/0	64/0	80/5	18/6	<b>77/0</b>	<b>80/0</b>	48/29	58/6
<i>CNK</i>	71/5	66/7	4/0	53/15	58/2	66/1	<b>69/8</b>	55/5
<i>TRANS<sub>2</sub></i>	59/3	42/7	26/10	17/11	11/46	8/35	36/21	28/19
<i>TRANS<sub>n</sub></i>	54/2	35/5	27/35	5/8	13/45	9/41	28/21	24/22
<i>DENS</i>	15/10	10/20	10/10	0/15	10/20	15/25	15/0	15/14
<i>LCV<sub>d</sub></i>	29/20	37/10	19/22	31/22	8/33	10/26	22/13	22/21
referenca	60/5	<b>80/5</b>	60/0	<b>80/0</b>	60/0	<b>80/0</b>	45/40	66/7

Tabela 6.1: Povprečeni odstotki značilnih pozitivnih in negativnih korelacij med *ocenami zanesljivosti* in *klasifikacijsko točnostjo*, brez ocen zanesljivosti z mero skalarnega produkta.

## 6.2 Praktični napotki

Tabela 5.4 prikazuje, da nobena metoda ocenjevanja zanesljivosti ni bila uspešna na vseh domenah. V povprečju je za uporabo najbolj zanimiva ocena *LCV*, saj se pri naših testih izkaže za najboljšo metodo pri polovici testnih domen. Ravno tako je obetavna drugouvrščena metoda *BAGV*, ki je bila najboljša pri 25% testnih množicah. Pri nekaterih domenah sta se obnesli tudi metodi *CNK* in *TRANS<sub>1</sub>*, zato je pred uporabo na specifični domeni priporočljivo narediti primerjavo omenjenih ocen.

Kljub temu da je ocena *LCV* najboljša v povprečju, je informacija, katera ocena je boljša za kateri uporabljeni model, tudi koristna. Iz tabele 6.1 razberemo da je metoda *BAGV* boljša od osnovne klasifikatorjeve ocene zanesljivosti pri odločitvenih drevesih (85/5 proti 60/6), metodi najbližjih sosedov (80/0) in metodi bagging (80/0), ocena *TRANS<sub>1</sub>* pri naključnih gozdovih (77/0 proti 60/0) ter ocena *CNK* pri metodi podpornih vektorjev (69/8 proti 45/40). Pri naivnemu Bayesu, nevronske mreže in modelu bagging nismo uspeli doseči višje korelacije, kot jo je dosegla njihova osnovna ocena.

Pri izbiri mere bližine velja upoštevati princip večkratne razlage, saj se vse mere, če izključimo mero na osnovi skalarnega produkta, obnašajo podobno različno. Zato predlagamo uporabo utežene kombinacije ali pa preprostega povprečja večih različnih mer.

## 6.3 Nadaljnje delo

Glavne ideje za nadaljnje delo so:

- ugotoviti ali lahko z vrednostmi pridobljenih ocen zanesljivosti dosežemo popravke prvotnih napovedi in tako izboljšamo klasifikacijsko točnost;
- preveriti učinek kalibriranja napovednih verjetnosti na različnih modelih. NN so sicer znane po tem, da ne potrebujejo kalibracije, medtem ko SVM in NB potrebujeta vsak drugačno. Zato je možno, da bo vpliv kalibracije na referenčno metodo, mogoče pa tudi na ostale metode, precejšen;
- raziskati potencial dinamične izbire najbolj primerne ocene. V [2] dose-gajo boljše rezultate z metaučenjem v prostoru domen in modelov ter z notranjim prečnim preverjanjem;
- organizirati razvite algoritme v okolju R v modularno knjižnico za okolje R in prispevati paket javno dostopni zbirki.

# Literatura

- [1] A. Asuncion, D.J. Newman, *UCI Machine Learning Repository*, dostopno na naslovu <http://archive.ics.uci.edu/ml/> (2007).
- [2] Z. Bosnić, *Ocenjevanje zanesljivosti posameznih napovedi z analizo občutljivosti regresijskih modelov*, doktorska disertacija, Fakulteta za računalništvo in informatiko, Ljubljana (2007).
- [3] Breiman, Friedman, Olshen, and Stone, *Classification and Regression Trees*, Wadsworth (1984).
- [4] L. Breiman, *Bagging predictors*, *Machine Learning* 24 (2), 1230–140, (1996).
- [5] L. Breiman, *Random Forests*, *Machine Learning* 45(1), 5–32, (2001).
- [6] L. Breiman, *Manual On Setting Up, Using, And Understanding Random Forests V3.1*, dostopno na naslovu [http://oz.berkeley.edu/users/breiman/Using\\_random\\_forests\\_V3.1.pdf](http://oz.berkeley.edu/users/breiman/Using_random_forests_V3.1.pdf) (2002).
- [7] N. Christiannini, J. Shawe-Taylor, *Support Vector Machines and Other Kernel-Based Learning Methods*, Cambridge University Press (2000).
- [8] C. Chang, C. Lin, *LIBSVM: a library for support vector machines*, dostopno na naslovu <http://www.csie.ntu.edu.tw/~cjlin/libsvm> (2001).
- [9] B.S. Everitt, T. Hothorn, *A Handbook of Statistical Analyses Using R*, Chapman & Hall/CRC (2006).
- [10] Hechenbichler K. and Schliep K.P., *Weighted k-Nearest-Neighbor Techniques and Ordinal Classification*, Discussion Paper 399, SFB 386, Ludwig-Maximilians University Munich (2004).

## LITERATURA

---

- [11] G. K. Kanji, *100 statistical tests*, SAGE Publications, (2006).
- [12] I. Kononenko, *Strojno učenje*, Založba fakultete za elektrotehniko in fakultete za računalništvo in informatiko, Ljubljana (1997).
- [13] M. Kukar, *Ocenjevanje zanesljivosti klasifikacij in cenovno občutljivo kombiniranje metod strojnega učenja*, doktorska disertacija, Fakulteta za računalništvo in informatiko, Ljubljana (2001).
- [14] J. Lin, *Divergence measures based on the shannon entropy*, IEEE Transactions on Information Theory 37 (1), 145–151, (1991).
- [15] Ripley, B. D., *Pattern Recognition and Neural Networks*, Cambridge (1996).
- [16] R Development Core Team, *A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria (2006).

# Dodatek A

## Podrobni rezultati testiranja

V prihodnjih tabelah so prikazani rezultati vseh opravljenih testov. Tabelirani rezultati so združeni po metodah ocenjevanja zanesljivosti in razčlenjeni po merah bližine.

Senčene celice označujejo statistično značilne ( $\alpha \leq 0.05$ ) korelacijske koeficiente, podkrepljene s pozitivnim testom ločljivosti ocen zanesljivosti pozitivnih in negativnih klasifikacij. Oznaka „/“ zaznamuje teste kjer izračun korelacijskih koeficientov ni bil možen zaradi neustreznih podatkov (standardna deviacija enaka 0).

Zadnji dve vrstici (+ in -) prikazujeta odstotek značilnih pozitivnih in negativnih testov posameznega klasifikatorja pri vseh podatkovnih množicah.

## A.1 Korelacijski koeficienti ocen *CNK*

	<i>CNK</i> <sub>man</sub>						
	DT	NB	KNN	NN	RF	BAG	SVM
housevotes	0.346	0.614	/	0.166	0.557	0.382	-0.442
wine	0.871	0.488	/	0.933	0.770	0.808	0.624
parkinsons	0.794	0.896	/	0.192	0.764	0.765	-0.823
zoo	0.863	0.593	/	0.705	0.825	0.797	0.470
tictactoe	0.735	0.738	/	0.469	0.577	0.493	-0.396
postoperative	-0.236	0.165	/	0.098	-0.047	0.105	-0.462
monks3	/	-0.393	/	0.317	0.237	/	-0.153
irisset	0.654	0.464	/	0.577	0.551	0.562	0.772
glass	0.503	0.637	/	0.470	0.505	0.504	0.412
hungarian	0.336	0.198	/	0.091	0.374	0.343	0.117
ecoli	0.609	0.523	/	0.563	0.561	0.516	0.509
heart	0.328	0.197	/	0.249	0.316	0.305	-0.395
haberman	0.049	0.019	/	-0.165	0.244	0.299	-0.025
flag	0.430	0.526	/	0.385	0.466	0.310	-0.179
wdbc	0.768	0.739	/	0.119	0.732	0.729	-0.647
bcw	0.711	0.529	/	0.013	0.579	0.591	0.463
sonar	0.771	0.787	/	0.382	0.569	0.572	-0.743
soybean	1.000	-0.021	/	-0.009	0.997	1.000	0.387
hepatitis	0.607	0.315	/	0.671	0.450	0.486	-0.529
lungcancer	0.050	0.100	/	-0.020	0.220	0.148	-0.024
+	<b>80 %</b>	<b>75 %</b>	<b>0 %</b>	<b>55 %</b>	<b>70 %</b>	<b>75 %</b>	<b>30 %</b>
-	<b>5 %</b>	<b>5 %</b>	<b>0 %</b>	<b>5 %</b>	<b>0 %</b>	<b>0 %</b>	<b>45 %</b>

Tabela A.1: Korelacijski koeficienti med oceno *CNK*<sub>man</sub> in točnostjo klasifikacij.

## A.1 Korelacijski koeficienti ocen *CNK*

	<i>CNK<sub>evk</sub></i>						
	DT	NB	KNN	NN	RF	BAG	SVM
housevotes	0.346	0.614	/	0.778	0.557	0.382	-0.442
wine	0.872	0.488	/	0.931	0.757	0.806	0.619
parkinsons	0.794	0.896	/	0.409	0.764	0.765	-0.823
zoo	0.860	0.534	/	0.677	0.794	0.761	0.340
tictactoe	0.735	0.738	/	0.777	0.577	0.493	-0.396
postoperative	-0.235	0.150	/	0.075	-0.054	0.093	-0.457
monks3	/	-0.393	/	0.457	0.237	/	-0.153
irisset	0.654	0.464	/	0.577	0.551	0.562	0.772
glass	0.491	0.632	/	0.414	0.479	0.491	0.389
hungarian	0.336	0.198	/	0.329	0.374	0.343	0.117
ecoli	0.596	0.530	/	0.556	0.542	0.491	0.474
heart	0.328	0.197	/	0.469	0.316	0.305	-0.395
haberman	0.049	0.019	/	-0.033	0.244	0.299	-0.025
flag	0.344	0.491	/	0.157	0.321	0.185	-0.349
wdbc	0.768	0.739	/	0.543	0.732	0.729	-0.647
bcw	0.711	0.529	/	0.814	0.579	0.591	0.463
sonar	0.771	0.787	/	0.686	0.569	0.572	-0.743
soybean	1.000	-0.021	/	-0.009	0.989	0.998	0.377
hepatitis	0.607	0.315	/	0.751	0.450	0.486	-0.529
lungcancer	0.037	0.063	/	-0.060	0.222	0.120	-0.008
+	<b>80 %</b>	<b>75 %</b>	<b>0 %</b>	<b>60 %</b>	<b>70 %</b>	<b>75 %</b>	<b>30 %</b>
-	<b>5 %</b>	<b>5 %</b>	<b>0 %</b>	<b>0 %</b>	<b>0 %</b>	<b>0 %</b>	<b>45 %</b>

Tabela A.2: Korelacijski koeficienti med oceno *CNK<sub>evk</sub>* in točnostjo klasifikacij.

	<i>CNK<sub>max</sub></i>						
	DT	NB	KNN	NN	RF	BAG	SVM
housevotes	0.346	0.614	/	0.922	0.557	0.382	-0.442
wine	0.871	0.488	/	0.933	0.770	0.808	0.624
parkinsons	0.794	0.896	/	0.485	0.764	0.765	-0.823
zoo	0.859	0.550	/	0.705	0.793	0.746	0.299
tictactoe	0.735	0.738	/	0.883	0.577	0.493	-0.396
postoperative	-0.236	0.165	/	0.098	-0.047	0.105	-0.462
monks3	/	-0.393	/	0.761	0.237	/	-0.153
irisset	0.654	0.464	/	0.577	0.551	0.562	0.772
glass	0.485	0.637	/	0.403	0.479	0.488	0.383
hungarian	0.336	0.198	/	0.516	0.374	0.343	0.117
ecoli	0.596	0.529	/	0.560	0.545	0.508	0.483
heart	0.328	0.197	/	0.609	0.316	0.305	-0.395
haberman	0.049	0.019	/	0.135	0.244	0.299	-0.025
flag	0.301	0.508	/	0.153	0.306	0.169	-0.342
wdbc	0.768	0.739	/	0.774	0.732	0.729	-0.647
bcw	0.711	0.529	/	0.935	0.579	0.591	0.463
sonar	0.771	0.787	/	0.845	0.569	0.572	-0.743
soybean	1.000	-0.021	/	-0.009	0.997	0.999	0.397
hepatitis	0.607	0.315	/	0.736	0.450	0.486	-0.529
lungcancer	0.050	0.100	/	-0.020	0.220	0.148	-0.024
+	<b>80 %</b>	<b>75 %</b>	<b>0 %</b>	<b>70 %</b>	<b>70 %</b>	<b>70 %</b>	<b>30 %</b>
-	<b>5 %</b>	<b>5 %</b>	<b>0 %</b>	<b>0 %</b>	<b>0 %</b>	<b>0 %</b>	<b>45 %</b>

Tabela A.3: Korelacijski koeficienti med oceno *CNK<sub>max</sub>* in točnostjo klasifikacij.

## Poglavje A: Podrobni rezultati testiranj

	<i>CNK<sub>cos</sub></i>						
	DT	NB	KNN	NN	RF	BAG	SVM
housevotes	0.286	0.549	-0.054	-0.836	0.381	0.228	-0.454
wine	0.867	0.467	0.118	0.845	0.426	0.610	0.469
parkinsons	0.726	0.868	0.183	-0.037	0.659	0.662	-0.836
zoo	0.805	0.523	0.104	0.708	0.685	0.797	0.435
tictactoe	0.700	0.778	0.037	-0.716	0.426	0.381	-0.405
postoperative	-0.227	0.143	0.062	0.128	-0.071	0.054	-0.485
monks3	/	-0.437	0.041	-0.858	0.235	/	-0.158
irisset	0.559	0.398	-0.017	0.480	0.454	0.508	0.730
glass	0.499	0.637	-0.074	0.484	0.417	0.456	0.409
hungarian	0.322	0.137	-0.042	-0.168	0.300	0.270	0.196
ecoli	0.593	0.474	0.123	0.510	0.428	0.418	0.376
heart	0.320	0.124	0.083	-0.038	0.309	0.280	-0.408
haberman	0.074	0.008	0.043	-0.264	0.180	0.250	0.044
flag	0.417	0.496	-0.047	0.323	0.313	0.265	-0.146
wdbc	0.759	0.735	-0.052	-0.071	0.610	0.628	-0.646
bcw	0.654	0.527	0.007	-0.403	0.389	0.434	0.406
sonar	0.752	0.776	-0.076	-0.663	0.598	0.530	-0.750
soybean	0.967	-0.022	0.989	-0.019	0.884	0.973	0.395
hepatitis	0.560	0.217	0.039	0.233	0.408	0.380	-0.531
lungcancer	0.004	-0.053	-0.080	-0.226	0.295	0.082	0.059
+	<b>80 %</b>	<b>75 %</b>	<b>5 %</b>	<b>30 %</b>	<b>75 %</b>	<b>75 %</b>	<b>30 %</b>
-	<b>5 %</b>	<b>0 %</b>	<b>0 %</b>	<b>35 %</b>	<b>0 %</b>	<b>0 %</b>	<b>50 %</b>

Tabela A.4: Korelacijski koeficienti med oceno *CNK<sub>cos</sub>* in točnostjo klasifikacij.

	<i>CNK<sub>kot</sub></i>						
	DT	NB	KNN	NN	RF	BAG	SVM
housevotes	0.343	0.609	0.215	-0.852	0.539	0.373	-0.420
wine	0.872	0.486	0.390	0.944	0.691	0.770	0.605
parkinsons	0.789	0.895	0.247	-0.023	0.790	0.769	-0.828
zoo	0.874	0.556	0.353	0.681	0.838	0.812	0.487
tictactoe	0.742	0.792	0.270	-0.782	0.584	0.498	-0.384
postoperative	-0.263	0.166	0.014	0.100	-0.047	0.102	-0.496
monks3	/	-0.408	-0.261	-0.897	0.251	/	-0.133
irisset	0.655	0.470	0.144	0.569	0.560	0.569	0.786
glass	0.532	0.659	0.106	0.533	0.524	0.530	0.484
hungarian	0.347	0.199	0.089	-0.174	0.402	0.365	0.204
ecoli	0.603	0.537	0.293	0.564	0.546	0.508	0.478
heart	0.336	0.204	0.222	-0.035	0.374	0.346	-0.391
haberman	0.085	0.042	0.104	-0.229	0.258	0.318	0.042
flag	0.452	0.526	0.100	0.388	0.435	0.340	-0.089
wdbc	0.764	0.739	0.160	-0.099	0.732	0.725	-0.634
bcw	0.703	0.525	0.221	-0.395	0.568	0.581	0.466
sonar	0.775	0.782	0.142	-0.742	0.660	0.623	-0.735
soybean	1.000	-0.021	1.000	-0.009	0.972	0.996	0.464
hepatitis	0.620	0.315	0.166	0.205	0.484	0.495	-0.513
lungcancer	0.021	0.002	-0.094	-0.114	0.320	0.143	0.012
+	<b>80 %</b>	<b>75 %</b>	<b>35 %</b>	<b>30 %</b>	<b>75 %</b>	<b>75 %</b>	<b>30 %</b>
-	<b>5 %</b>	<b>0 %</b>	<b>0 %</b>	<b>35 %</b>	<b>0 %</b>	<b>0 %</b>	<b>45 %</b>

Tabela A.5: Korelacijski koeficienti med oceno *CNK<sub>kot</sub>* in točnostjo klasifikacij.

## A.1 Korelacijski koeficienti ocen *CNK*

	<i>CNK</i> <sub>ska</sub>						
	DT	NB	KNN	NN	RF	BAG	SVM
housevotes	0.377	0.624	0.575	1.000	0.659	0.511	-0.428
wine	0.871	0.545	0.580	0.975	0.821	0.827	0.696
parkinsons	0.809	0.901	0.862	1.000	0.894	0.836	-0.866
zoo	0.980	0.639	0.970	0.706	0.959	0.875	0.551
tictactoe	0.809	0.919	0.848	1.000	0.824	0.670	-0.410
postoperative	-0.221	0.220	0.224	0.141	0.019	0.229	-0.447
monks3	/	-0.362	-0.200	1.000	0.258	/	-0.169
irisset	0.690	0.591	0.614	0.597	0.636	0.634	0.850
glass	0.587	0.680	0.647	0.720	0.707	0.658	0.689
hungarian	0.445	0.332	0.575	1.000	0.601	0.575	0.461
ecoli	0.647	0.617	0.629	0.648	0.702	0.653	0.630
heart	0.389	0.319	0.586	1.000	0.610	0.595	-0.412
haberman	0.213	0.112	0.409	1.000	0.424	0.498	0.236
flag	0.514	0.543	0.485	0.571	0.659	0.497	0.056
wdbc	0.773	0.746	0.736	1.000	0.797	0.808	-0.635
bcw	0.715	0.536	0.639	1.000	0.697	0.693	0.548
sonar	0.791	0.788	0.744	1.000	0.911	0.826	-0.752
soybean	1.000	-0.021	1.000	-0.009	1.000	0.999	0.705
hepatitis	0.690	0.408	0.583	1.000	0.679	0.663	-0.544
lungcancer	0.131	0.069	0.410	-0.067	0.559	0.327	0.063
+	<b>85 %</b>	<b>80 %</b>	<b>80 %</b>	<b>75 %</b>	<b>80 %</b>	<b>80 %</b>	<b>45 %</b>
-	<b>5 %</b>	<b>5 %</b>	<b>5 %</b>	<b>0 %</b>	<b>0 %</b>	<b>0 %</b>	<b>45 %</b>

Tabela A.6: Korelacijski koeficienti med oceno *CNK*<sub>ska</sub> in točnostjo klasifikacij.

	<i>CNK</i> <sub>bha</sub>						
	DT	NB	KNN	NN	RF	BAG	SVM
housevotes	0.257	0.545	0.077	0.979	0.418	0.234	-0.495
wine	0.850	0.502	/	0.861	0.726	0.733	0.485
parkinsons	0.749	0.851	-0.040	0.843	0.540	0.642	-0.782
zoo	0.782	0.663	-0.189	0.746	0.809	0.803	0.401
tictactoe	0.629	0.590	/	0.980	0.386	0.317	-0.417
postoperative	-0.021	0.264	-0.142	0.159	-0.030	0.108	-0.418
monks3	/	-0.404	/	0.981	0.250	/	-0.179
irisset	0.590	0.387	/	0.562	0.373	0.427	0.679
glass	0.496	0.592	-0.044	0.434	0.431	0.418	0.339
hungarian	0.256	0.107	0.050	0.792	0.180	0.194	-0.124
ecoli	0.600	0.436	0.023	0.512	0.494	0.476	0.464
heart	0.271	0.069	0.026	0.859	0.105	0.165	-0.391
haberman	-0.050	-0.079	0.150	0.355	0.109	0.152	-0.180
flag	0.376	0.479	-0.174	0.313	0.421	0.268	-0.223
wdbc	0.763	0.724	/	0.889	0.634	0.667	-0.630
bcw	0.656	0.528	0.083	0.979	0.466	0.504	0.386
sonar	0.683	0.777	-0.040	0.962	0.288	0.361	-0.734
soybean	0.957	-0.021	/	-0.009	0.985	0.999	0.193
hepatitis	0.494	0.216	-0.049	0.938	0.290	0.360	-0.545
lungcancer	0.131	0.067	0.003	-0.105	0.045	0.071	0.100
+	<b>70 %</b>	<b>65 %</b>	<b>10 %</b>	<b>70 %</b>	<b>45 %</b>	<b>55 %</b>	<b>25 %</b>
-	<b>0 %</b>	<b>5 %</b>	<b>5 %</b>	<b>0 %</b>	<b>0 %</b>	<b>0 %</b>	<b>55 %</b>

Tabela A.7: Korelacijski koeficienti med oceno *CNK*<sub>bha</sub> in točnostjo klasifikacij.

Poglavje A: Podrobni rezultati testiranj

		<i>CNKhar</i>					
	DT	NB	KNN	NN	RF	BAG	SVM
housevotes	0.291	0.572	-0.029	0.979	0.445	0.267	-0.481
wine	0.864	0.480	0.050	0.878	0.692	0.735	0.498
parkinsons	0.761	0.872	-0.048	0.852	0.634	0.690	-0.815
zoo	0.802	0.634	0.206	0.746	0.797	0.805	0.426
tictactoe	0.681	0.665	/	0.974	0.435	0.367	-0.416
postoperative	-0.105	0.207	0.025	0.141	-0.049	0.093	-0.464
monks3	/	-0.412	/	0.975	0.222	/	-0.184
irisset	0.607	0.408	-0.021	0.532	0.436	0.506	0.716
glass	0.495	0.623	0.115	0.461	0.449	0.441	0.388
hungarian	0.287	0.129	-0.069	0.794	0.239	0.233	-0.029
ecoli	0.609	0.470	0.207	0.520	0.506	0.472	0.463
heart	0.298	0.106	0.021	0.856	0.179	0.204	-0.404
haberman	0.002	-0.037	0.044	0.347	0.143	0.197	-0.103
flag	0.403	0.501	0.314	0.353	0.403	0.270	-0.208
wdbc	0.767	0.741	-0.083	0.891	0.658	0.683	-0.652
bcw	0.678	0.536	0.096	0.987	0.478	0.510	0.400
sonar	0.735	0.785	0.085	0.956	0.416	0.438	-0.751
soybean	0.977	-0.021	/	-0.009	0.991	1.000	0.249
hepatitis	0.529	0.238	0.047	0.952	0.335	0.379	-0.546
lungcancer	0.067	0.059	0.364	-0.112	0.131	0.083	0.050
+	<b>75 %</b>	<b>70 %</b>	<b>10 %</b>	<b>75 %</b>	<b>50 %</b>	<b>70 %</b>	<b>30 %</b>
-	<b>0 %</b>	<b>5 %</b>	<b>0 %</b>	<b>0 %</b>	<b>0 %</b>	<b>0 %</b>	<b>50 %</b>

Tabela A.8: Korelacijski koeficienti med oceno *CNKhar* in točnostjo klasifikacij.

		<i>CNKhel</i>					
	DT	NB	KNN	NN	RF	BAG	SVM
housevotes	0.257	0.545	/	0.831	0.418	0.234	-0.495
wine	0.850	0.502	/	0.861	0.726	0.733	0.485
parkinsons	0.749	0.851	/	0.241	0.540	0.642	-0.782
zoo	0.782	0.663	/	0.746	0.809	0.803	0.401
tictactoe	0.629	0.590	/	0.877	0.386	0.317	-0.417
postoperative	-0.021	0.264	/	0.159	-0.030	0.108	-0.418
monks3	/	-0.404	/	0.865	0.250	/	-0.179
irisset	0.590	0.387	/	0.562	0.373	0.427	0.679
glass	0.496	0.592	/	0.434	0.431	0.418	0.339
hungarian	0.256	0.107	/	0.171	0.180	0.194	-0.124
ecoli	0.600	0.436	/	0.512	0.494	0.476	0.464
heart	0.271	0.069	/	0.366	0.105	0.165	-0.391
haberman	-0.050	-0.079	/	-0.225	0.109	0.152	-0.180
flag	0.376	0.479	/	0.313	0.421	0.268	-0.223
wdbc	0.763	0.724	/	0.418	0.634	0.667	-0.630
bcw	0.656	0.528	/	0.711	0.466	0.504	0.386
sonar	0.683	0.777	/	0.800	0.288	0.361	-0.734
soybean	0.957	-0.021	/	-0.009	0.985	0.999	0.193
hepatitis	0.494	0.216	/	0.787	0.290	0.360	-0.545
lungcancer	0.131	0.067	/	-0.105	0.045	0.071	0.100
+	<b>70 %</b>	<b>65 %</b>	<b>0 %</b>	<b>65 %</b>	<b>45 %</b>	<b>55 %</b>	<b>25 %</b>
-	<b>0 %</b>	<b>5 %</b>	<b>0 %</b>	<b>5 %</b>	<b>0 %</b>	<b>0 %</b>	<b>55 %</b>

Tabela A.9: Korelacijski koeficienti med oceno *CNKhel* in točnostjo klasifikacij.

## A.1 Korelacijski koeficienti ocen $CNK$

	$CNK_{jsd}$						
	DT	NB	KNN	NN	RF	BAG	SVM
housevotes	0.276	0.561	/	0.823	0.434	0.252	-0.491
wine	0.859	0.488	/	0.869	0.709	0.735	0.488
parkinsons	0.758	0.863	/	0.344	0.596	0.673	-0.802
zoo	0.792	0.650	/	0.747	0.803	0.805	0.410
tictactoe	0.660	0.632	/	0.857	0.415	0.343	-0.418
postoperative	-0.067	0.232	/	0.150	-0.043	0.100	-0.447
monks3	/	-0.409	/	0.839	0.231	/	-0.185
irisset	0.603	0.399	/	0.540	0.407	0.476	0.700
glass	0.496	0.612	/	0.450	0.442	0.431	0.370
hungarian	0.274	0.117	/	0.270	0.212	0.216	-0.076
ecoli	0.607	0.454	/	0.515	0.501	0.473	0.463
heart	0.289	0.087	/	0.455	0.145	0.186	-0.400
haberman	-0.019	-0.055	/	-0.184	0.127	0.176	-0.140
flag	0.391	0.492	/	0.334	0.405	0.264	-0.216
wdbc	0.766	0.736	/	0.489	0.650	0.679	-0.647
bcw	0.670	0.534	/	0.810	0.475	0.509	0.393
sonar	0.716	0.782	/	0.784	0.361	0.406	-0.746
soybean	0.967	-0.021	/	-0.009	0.988	1.000	0.223
hepatitis	0.513	0.227	/	0.830	0.314	0.370	-0.548
lungcancer	0.093	0.063	/	-0.111	0.091	0.076	0.072
+	<b>75 %</b>	<b>70 %</b>	<b>0 %</b>	<b>65 %</b>	<b>50 %</b>	<b>70 %</b>	<b>25 %</b>
-	<b>0 %</b>	<b>5 %</b>	<b>0 %</b>	<b>5 %</b>	<b>0 %</b>	<b>0 %</b>	<b>55 %</b>

Tabela A.10: Korelacijski koeficienti med oceno  $CNK_{jsd}$  in točnostjo klasifikacij.

	$CNK_{sep}$						
	DT	NB	KNN	NN	RF	BAG	SVM
housevotes	-0.263	-0.309	/	-0.168	0.321	0.386	-0.375
wine	-0.154	0.075	/	0.178	0.315	0.553	0.311
parkinsons	-0.165	0.392	/	-0.172	0.043	0.235	-0.472
zoo	0.840	0.023	/	0.007	0.456	0.582	0.564
tictactoe	0.635	-0.349	/	-0.243	-0.384	0.257	-0.330
postoperative	0.147	-0.121	/	0.163	-0.147	-0.073	-0.083
monks3	/	-0.107	/	-0.856	0.350	/	-0.055
irisset	0.130	-0.078	/	0.271	0.250	0.415	0.212
glass	0.027	0.117	/	-0.024	0.117	0.200	-0.021
hungarian	-0.266	-0.284	/	-0.446	-0.192	0.047	-0.335
ecoli	0.100	-0.256	/	0.069	0.372	0.438	0.232
heart	-0.198	-0.195	/	-0.456	-0.275	-0.097	-0.289
haberman	-0.215	-0.186	/	-0.467	-0.181	0.040	-0.221
flag	-0.055	-0.039	/	0.300	0.287	0.339	-0.165
wdbc	-0.145	0.177	/	-0.229	0.260	0.444	-0.418
bcw	-0.236	-0.074	/	-0.110	0.316	0.390	-0.403
sonar	-0.078	0.176	/	-0.432	-0.484	-0.236	-0.582
soybean	0.968	0.129	/	0.053	0.277	0.504	0.005
hepatitis	-0.206	-0.305	/	-0.095	-0.117	0.261	-0.442
lungcancer	0.010	-0.074	/	-0.560	-0.355	-0.167	-0.021
+	<b>15 %</b>	<b>15 %</b>	<b>0 %</b>	<b>15 %</b>	<b>30 %</b>	<b>40 %</b>	<b>5 %</b>
-	<b>25 %</b>	<b>30 %</b>	<b>0 %</b>	<b>55 %</b>	<b>20 %</b>	<b>5 %</b>	<b>55 %</b>

Tabela A.11: Korelacijski koeficienti med oceno  $CNK_{sep}$  in točnostjo klasifikacij.

## A.2 Korelacijski koeficienti ocen *LCV*

	<i>LCV</i> man						
	DT	NB	KNN	NN	RF	BAG	SVM
housevotes	0.212	0.612	0.494	-0.173	0.550	0.430	-0.158
wine	0.920	0.723	0.619	0.452	0.847	0.826	0.652
parkinsons	0.495	0.759	0.437	0.067	0.511	0.583	-0.240
zoo	0.833	0.396	0.782	0.604	0.854	0.736	0.373
tictactoe	0.008	0.389	0.116	-0.951	0.426	0.050	0.104
postoperative	0.015	-0.032	0.281	-0.021	0.107	0.302	-0.281
monks3	/	0.043	0.624	-0.631	0.399	/	-0.409
irisset	0.728	0.805	0.817	0.468	0.804	0.790	0.788
glass	0.493	0.512	0.405	0.234	0.475	0.367	0.251
hungarian	0.482	0.371	0.510	-0.019	0.537	0.593	0.165
ecoli	0.581	0.607	0.401	0.099	0.615	0.498	0.476
heart	0.444	0.397	0.578	0.074	0.567	0.532	-0.272
haberman	0.296	0.079	0.442	0.645	0.399	0.511	0.029
flag	0.343	0.625	0.400	0.301	0.272	0.159	-0.246
wdbc	0.672	0.587	0.534	0.045	0.753	0.742	-0.262
bcw	0.635	0.536	0.567	-0.823	0.709	0.728	0.197
sonar	0.501	0.504	0.161	0.227	0.191	0.232	-0.379
soybean	0.994	1.000	0.880	0.670	0.982	0.973	0.135
hepatitis	0.602	0.420	0.487	0.202	0.567	0.549	-0.051
lungcancer	-0.127	-0.238	-0.220	0.089	0.226	0.500	0.117
+	<b>80 %</b>	<b>75 %</b>	<b>85 %</b>	<b>30 %</b>	<b>70 %</b>	<b>70 %</b>	<b>35 %</b>
-	<b>0 %</b>	<b>0 %</b>	<b>0 %</b>	<b>20 %</b>	<b>0 %</b>	<b>0 %</b>	<b>25 %</b>

Tabela A.12: Korelacijski koeficienti med oceno *LCV* man in točnostjo klasifikacij.

## A.2 Korelacijski koeficienti ocen *LCV*

	<i>LCV</i> <sub>evk</sub>						
	DT	NB	KNN	NN	RF	BAG	SVM
housevotes	0.212	0.612	0.494	0.017	0.550	0.430	-0.158
wine	0.923	0.723	0.620	0.438	0.843	0.824	0.646
parkinsons	0.495	0.759	0.437	0.103	0.511	0.583	-0.240
zoo	0.828	0.379	0.759	0.608	0.823	0.695	0.245
tictactoe	0.008	0.389	0.116	-0.951	0.426	0.050	0.104
postoperative	0.021	-0.063	0.271	-0.005	0.088	0.287	-0.262
monks3	/	0.043	0.624	-0.631	0.399	/	-0.409
iriset	0.728	0.805	0.817	0.431	0.804	0.790	0.792
glass	0.494	0.488	0.396	0.141	0.455	0.341	0.247
hungarian	0.482	0.371	0.510	-0.012	0.537	0.593	0.165
ecoli	0.555	0.602	0.385	0.184	0.593	0.452	0.453
heart	0.444	0.397	0.578	0.109	0.567	0.532	-0.272
haberman	0.296	0.079	0.442	0.645	0.399	0.511	0.029
flag	0.229	0.614	0.316	0.044	0.004	-0.096	-0.330
wdbc	0.672	0.587	0.534	0.155	0.753	0.742	-0.262
bcw	0.635	0.536	0.567	-0.745	0.709	0.728	0.197
sonar	0.501	0.504	0.161	0.238	0.191	0.232	-0.379
soybean	0.994	1.000	0.880	0.656	0.972	0.966	0.115
hepatitis	0.602	0.420	0.487	0.274	0.567	0.549	-0.051
lungcancer	-0.145	-0.236	-0.297	0.068	0.198	0.491	0.106
+	<b>80 %</b>	<b>75 %</b>	<b>85 %</b>	<b>30 %</b>	<b>65 %</b>	<b>70 %</b>	<b>35 %</b>
-	<b>0 %</b>	<b>0 %</b>	<b>0 %</b>	<b>15 %</b>	<b>0 %</b>	<b>0 %</b>	<b>20 %</b>

Tabela A.13: Korelacijski koeficienti med oceno *LCV*<sub>evk</sub> in točnostjo klasifikacij.

	<i>LCV</i> <sub>max</sub>						
	DT	NB	KNN	NN	RF	BAG	SVM
housevotes	0.212	0.612	0.494	0.191	0.550	0.430	-0.158
wine	0.920	0.723	0.619	0.421	0.847	0.826	0.652
parkinsons	0.495	0.759	0.437	0.127	0.511	0.583	-0.240
zoo	0.820	0.396	0.778	0.529	0.842	0.698	0.217
tictactoe	0.008	0.389	0.116	-0.951	0.426	0.050	0.104
postoperative	0.015	-0.032	0.281	-0.030	0.107	0.302	-0.281
monks3	/	0.043	0.624	-0.631	0.399	/	-0.409
iriset	0.728	0.805	0.817	0.361	0.804	0.790	0.788
glass	0.501	0.513	0.404	0.066	0.464	0.357	0.236
hungarian	0.482	0.371	0.510	-0.005	0.537	0.593	0.165
ecoli	0.567	0.607	0.394	0.084	0.607	0.476	0.444
heart	0.444	0.397	0.578	0.133	0.567	0.532	-0.272
haberman	0.296	0.079	0.442	0.645	0.399	0.511	0.029
flag	0.123	0.624	0.295	0.066	-0.064	-0.175	-0.329
wdbc	0.672	0.587	0.534	0.228	0.753	0.742	-0.262
bcw	0.635	0.536	0.567	-0.614	0.709	0.728	0.197
sonar	0.501	0.504	0.161	0.240	0.191	0.232	-0.379
soybean	0.994	1.000	0.880	0.538	0.982	0.971	0.108
hepatitis	0.602	0.420	0.487	0.308	0.567	0.549	-0.051
lungcancer	-0.127	-0.238	-0.220	-0.040	0.226	0.500	0.117
+	<b>75 %</b>	<b>75 %</b>	<b>85 %</b>	<b>20 %</b>	<b>65 %</b>	<b>70 %</b>	<b>35 %</b>
-	<b>0 %</b>	<b>0 %</b>	<b>0 %</b>	<b>15 %</b>	<b>0 %</b>	<b>5 %</b>	<b>25 %</b>

Tabela A.14: Korelacijski koeficienti med oceno *LCV*<sub>max</sub> in točnostjo klasifikacij.

Poglavje A: Podrobni rezultati testiranj

	<i>LCV<sub>cos</sub></i>						
	DT	NB	KNN	NN	RF	BAG	SVM
housevotes	0.189	0.644	0.406	-0.544	0.393	0.290	-0.178
wine	0.894	0.686	0.597	0.334	0.629	0.707	0.524
parkinsons	0.457	0.759	0.315	0.135	0.335	0.446	-0.284
zoo	0.737	0.353	0.703	0.747	0.760	0.720	0.302
tictactoe	-0.070	0.215	-0.026	/	0.286	-0.063	0.057
postoperative	0.004	-0.115	0.248	-0.039	0.069	0.225	-0.224
monks3	/	-0.115	0.598	-0.073	0.270	/	-0.418
irisset	0.672	0.819	0.826	0.248	0.834	0.817	0.829
glass	0.469	0.494	0.313	0.215	0.354	0.290	0.170
hungarian	0.461	0.313	0.400	-0.127	0.432	0.498	0.124
ecoli	0.521	0.524	0.280	0.277	0.471	0.396	0.366
heart	0.444	0.328	0.486	0.008	0.486	0.451	-0.290
haberman	0.289	0.071	0.379	-0.028	0.323	0.412	0.028
flag	0.406	0.610	0.397	0.157	0.273	0.254	-0.211
wdbc	0.662	0.576	0.400	-0.028	0.678	0.670	-0.307
bcw	0.581	0.528	0.462	-0.881	0.572	0.620	0.155
sonar	0.478	0.477	-0.028	-0.151	0.051	0.080	-0.419
soybean	0.966	1.000	0.465	0.801	0.942	0.830	0.131
hepatitis	0.567	0.390	0.407	0.264	0.510	0.457	-0.088
lungcancer	-0.207	-0.230	-0.191	0.262	0.258	0.482	0.116
+	<b>80 %</b>	<b>75 %</b>	<b>75 %</b>	<b>25 %</b>	<b>70 %</b>	<b>70 %</b>	<b>35 %</b>
-	<b>5 %</b>	<b>5 %</b>	<b>0 %</b>	<b>20 %</b>	<b>0 %</b>	<b>0 %</b>	<b>20 %</b>

Tabela A.15: Korelacijski koeficienti med oceno *LCV<sub>cos</sub>* in točnostjo klasifikacij.

	<i>LCV<sub>kot</sub></i>						
	DT	NB	KNN	NN	RF	BAG	SVM
housevotes	0.214	0.614	0.491	-0.544	0.515	0.421	-0.152
wine	0.925	0.721	0.627	0.339	0.825	0.824	0.644
parkinsons	0.509	0.744	0.464	0.134	0.541	0.568	-0.211
zoo	0.850	0.386	0.785	0.751	0.836	0.743	0.391
tictactoe	0.036	0.370	0.139	/	0.427	0.150	0.122
postoperative	-0.022	-0.061	0.289	-0.056	0.090	0.305	-0.219
monks3	/	0.084	0.599	-0.089	0.433	/	-0.412
irisset	0.754	0.806	0.831	0.246	0.739	0.803	0.838
glass	0.507	0.520	0.429	0.206	0.496	0.412	0.306
hungarian	0.484	0.368	0.516	-0.116	0.508	0.591	0.181
ecoli	0.574	0.603	0.415	0.302	0.594	0.467	0.476
heart	0.451	0.397	0.585	0.008	0.558	0.550	-0.255
haberman	0.301	0.100	0.446	-0.028	0.404	0.457	0.048
flag	0.446	0.621	0.447	0.357	0.416	0.326	-0.168
wdbc	0.653	0.589	0.531	-0.027	0.759	0.749	-0.242
bcw	0.626	0.534	0.564	-0.881	0.676	0.724	0.201
sonar	0.498	0.505	0.178	-0.150	0.288	0.394	-0.327
soybean	0.997	1.000	0.859	0.705	0.971	0.951	0.228
hepatitis	0.625	0.417	0.501	0.264	0.592	0.562	-0.026
lungcancer	-0.158	-0.238	-0.165	0.233	0.286	0.478	0.123
+	<b>80 %</b>	<b>75 %</b>	<b>85 %</b>	<b>30 %</b>	<b>70 %</b>	<b>80 %</b>	<b>40 %</b>
-	<b>0 %</b>	<b>0 %</b>	<b>0 %</b>	<b>20 %</b>	<b>0 %</b>	<b>0 %</b>	<b>25 %</b>

Tabela A.16: Korelacijski koeficienti med oceno *LCV<sub>kot</sub>* in točnostjo klasifikacij.

## A.2 Korelacijski koeficienti ocen *LCV*

	<i>LCV</i> ska						
	DT	NB	KNN	NN	RF	BAG	SVM
housevotes	0.269	0.608	0.516	0.941	0.583	0.452	-0.143
wine	0.915	0.724	0.628	0.595	0.855	0.832	0.718
parkinsons	0.596	0.759	0.605	0.661	0.720	0.702	-0.153
zoo	0.935	0.401	0.844	0.530	0.878	0.755	0.492
tictactoe	0.163	0.439	0.276	0.925	0.624	0.203	0.277
postoperative	-0.210	-0.040	0.293	-0.017	0.019	0.314	-0.148
monks3	/	0.337	0.756	0.971	0.663	/	-0.430
irisset	0.756	0.782	0.802	0.489	0.798	0.795	0.842
glass	0.537	0.528	0.529	0.225	0.641	0.538	0.514
hungarian	0.519	0.420	0.592	0.771	0.658	0.659	0.258
ecoli	0.620	0.615	0.512	0.162	0.718	0.614	0.653
heart	0.464	0.416	0.631	0.793	0.678	0.653	-0.220
haberman	0.316	0.175	0.506	0.862	0.479	0.572	0.192
flag	0.553	0.625	0.518	0.693	0.711	0.589	0.031
wdbc	0.668	0.588	0.574	0.773	0.782	0.776	-0.218
bcw	0.644	0.537	0.594	0.992	0.741	0.747	0.232
sonar	0.510	0.506	0.342	0.935	0.729	0.641	-0.110
soybean	0.993	1.000	0.880	0.538	0.981	0.974	0.552
hepatitis	0.704	0.430	0.565	0.777	0.700	0.666	0.076
lungcancer	-0.140	-0.237	0.271	-0.096	0.459	0.357	-0.136
+	<b>85 %</b>	<b>85 %</b>	<b>85 %</b>	<b>75 %</b>	<b>80 %</b>	<b>80 %</b>	<b>45 %</b>
-	<b>5 %</b>	<b>0 %</b>	<b>0 %</b>	<b>0 %</b>	<b>0 %</b>	<b>0 %</b>	<b>10 %</b>

Tabela A.17: Korelacijski koeficienti med oceno *LCV*ska in točnostjo klasifikacij.

	<i>LCV</i> bha						
	DT	NB	KNN	NN	RF	BAG	SVM
housevotes	0.170	0.647	0.410	0.931	0.467	0.326	-0.202
wine	0.841	0.696	0.560	0.394	0.739	0.779	0.532
parkinsons	0.444	0.752	0.216	0.690	0.230	0.330	-0.356
zoo	0.667	0.368	0.696	0.548	0.809	0.706	0.250
tictactoe	-0.208	0.265	-0.103	0.930	0.232	-0.160	0.003
postoperative	0.101	-0.020	0.238	0.006	0.145	0.243	-0.089
monks3	/	-0.306	0.654	0.970	0.081	/	-0.407
irisset	0.636	0.832	0.778	0.423	0.776	0.771	0.612
glass	0.443	0.437	0.263	0.133	0.327	0.236	0.114
hungarian	0.428	0.301	0.344	0.766	0.340	0.456	0.071
ecoli	0.485	0.515	0.246	0.105	0.526	0.335	0.349
heart	0.409	0.312	0.441	0.799	0.341	0.262	-0.303
haberman	0.278	0.029	0.305	0.870	0.250	0.351	0.007
flag	0.321	0.598	0.372	0.060	0.165	0.142	-0.236
wdbc	0.675	0.564	0.413	0.759	0.690	0.678	-0.364
bcw	0.574	0.530	0.441	0.986	0.635	0.665	0.134
sonar	0.459	0.452	-0.102	0.938	-0.149	-0.110	-0.461
soybean	0.951	1.000	0.874	0.584	0.982	0.971	0.049
hepatitis	0.456	0.384	0.300	0.807	0.243	0.354	-0.131
lungcancer	0.026	-0.211	-0.193	-0.016	0.135	0.507	0.119
+	<b>80 %</b>	<b>75 %</b>	<b>80 %</b>	<b>70 %</b>	<b>60 %</b>	<b>65 %</b>	<b>25 %</b>
-	<b>5 %</b>	<b>0 %</b>	<b>5 %</b>	<b>0 %</b>	<b>0 %</b>	<b>0 %</b>	<b>20 %</b>

Tabela A.18: Korelacijski koeficienti med oceno *LCV*bha in točnostjo klasifikacij.

Poglavje A: Podrobni rezultati testiranj

	<i>LCVhar</i>						
	DT	NB	KNN	NN	RF	BAG	SVM
housevotes	0.182	0.639	0.428	0.922	0.464	0.338	-0.187
wine	0.884	0.699	0.589	0.397	0.772	0.782	0.544
parkinsons	0.441	0.758	0.274	0.751	0.283	0.403	-0.324
zoo	0.720	0.372	0.718	0.535	0.810	0.718	0.289
tictactoe	-0.135	0.279	-0.051	0.927	0.258	-0.099	0.030
postoperative	0.124	-0.057	0.249	-0.006	0.144	0.252	-0.159
monks3	/	-0.211	0.635	0.970	0.171	/	-0.412
irisset	0.667	0.831	0.807	0.426	0.804	0.797	0.695
glass	0.457	0.489	0.300	0.146	0.356	0.269	0.124
hungarian	0.456	0.320	0.396	0.739	0.403	0.503	0.088
ecoli	0.527	0.546	0.286	0.074	0.526	0.396	0.367
heart	0.432	0.337	0.482	0.787	0.419	0.361	-0.307
haberman	0.290	0.039	0.360	0.874	0.294	0.401	0.008
flag	0.330	0.615	0.376	0.150	0.170	0.138	-0.252
wdbc	0.670	0.577	0.433	0.734	0.702	0.693	-0.333
bcw	0.597	0.532	0.483	0.987	0.645	0.675	0.148
sonar	0.489	0.477	-0.042	0.938	-0.100	-0.045	-0.470
soybean	0.973	1.000	0.869	0.566	0.985	0.968	0.050
hepatitis	0.516	0.397	0.375	0.862	0.371	0.419	-0.113
lungcancer	-0.087	-0.227	-0.213	-0.022	0.188	0.531	0.137
+	<b>80 %</b>	<b>75 %</b>	<b>80 %</b>	<b>70 %</b>	<b>70 %</b>	<b>70 %</b>	<b>25 %</b>
-	<b>5 %</b>	<b>0 %</b>	<b>0 %</b>	<b>0 %</b>	<b>0 %</b>	<b>0 %</b>	<b>20 %</b>

Tabela A.19: Korelacijski koeficienti med oceno *LCVhar* in točnostjo klasifikacij.

	<i>LCVhel</i>						
	DT	NB	KNN	NN	RF	BAG	SVM
housevotes	0.170	0.647	0.410	0.110	0.467	0.326	-0.202
wine	0.841	0.696	0.560	0.412	0.739	0.779	0.532
parkinsons	0.444	0.752	0.216	0.162	0.230	0.330	-0.356
zoo	0.667	0.368	0.696	0.629	0.809	0.706	0.250
tictactoe	-0.208	0.265	-0.103	-0.950	0.232	-0.160	0.003
postoperative	0.101	-0.020	0.238	0.028	0.145	0.243	-0.089
monks3	/	-0.306	0.654	-0.615	0.081	/	-0.407
irisset	0.636	0.832	0.778	0.468	0.776	0.771	0.612
glass	0.443	0.437	0.263	0.227	0.327	0.236	0.114
hungarian	0.428	0.301	0.344	-0.011	0.340	0.456	0.071
ecoli	0.485	0.515	0.246	0.092	0.526	0.335	0.349
heart	0.409	0.312	0.441	0.163	0.341	0.262	-0.303
haberman	0.278	0.029	0.305	0.488	0.250	0.351	0.007
flag	0.321	0.598	0.372	0.065	0.165	0.142	-0.236
wdbc	0.675	0.564	0.413	0.189	0.690	0.678	-0.364
bcw	0.574	0.530	0.441	-0.677	0.635	0.665	0.134
sonar	0.459	0.452	-0.102	0.328	-0.149	-0.110	-0.461
soybean	0.951	1.000	0.874	0.760	0.982	0.971	0.049
hepatitis	0.456	0.384	0.300	0.343	0.243	0.354	-0.131
lungcancer	0.026	-0.211	-0.193	0.130	0.135	0.507	0.119
+	<b>80 %</b>	<b>75 %</b>	<b>80 %</b>	<b>55 %</b>	<b>60 %</b>	<b>65 %</b>	<b>25 %</b>
-	<b>5 %</b>	<b>0 %</b>	<b>5 %</b>	<b>15 %</b>	<b>0 %</b>	<b>0 %</b>	<b>20 %</b>

Tabela A.20: Korelacijski koeficienti med oceno *LCVhel* in točnostjo klasifikacij.

## A.2 Korelacijski koeficienti ocen $LCV$

	$LCV_{jsd}$						
	DT	NB	KNN	NN	RF	BAG	SVM
housevotes	0.175	0.643	0.418	0.081	0.462	0.330	-0.195
wine	0.868	0.696	0.576	0.432	0.761	0.779	0.533
parkinsons	0.441	0.756	0.243	0.149	0.254	0.366	-0.342
zoo	0.695	0.369	0.707	0.620	0.806	0.714	0.265
tictactoe	-0.171	0.275	-0.079	-0.950	0.242	-0.130	0.017
postoperative	0.115	-0.047	0.244	0.020	0.146	0.248	-0.129
monks3	/	-0.264	0.646	-0.622	0.123	/	-0.409
iriset	0.653	0.834	0.794	0.465	0.793	0.790	0.660
glass	0.449	0.472	0.282	0.228	0.338	0.251	0.114
hungarian	0.446	0.311	0.372	0.016	0.377	0.484	0.077
ecoli	0.510	0.531	0.266	0.072	0.523	0.367	0.355
heart	0.424	0.323	0.462	0.190	0.387	0.316	-0.309
haberman	0.286	0.033	0.337	0.536	0.273	0.379	0.007
flag	0.326	0.609	0.374	0.113	0.162	0.138	-0.244
wdbc	0.672	0.573	0.420	0.212	0.696	0.686	-0.351
bcw	0.589	0.531	0.463	-0.667	0.640	0.670	0.139
sonar	0.481	0.467	-0.073	0.318	-0.127	-0.082	-0.473
soybean	0.962	1.000	0.872	0.742	0.984	0.969	0.043
hepatitis	0.493	0.391	0.344	0.338	0.315	0.390	-0.122
lungcancer	-0.040	-0.221	-0.207	0.132	0.167	0.525	0.131
+	<b>80 %</b>	<b>75 %</b>	<b>80 %</b>	<b>45 %</b>	<b>65 %</b>	<b>70 %</b>	<b>25 %</b>
-	<b>5 %</b>	<b>0 %</b>	<b>5 %</b>	<b>15 %</b>	<b>0 %</b>	<b>0 %</b>	<b>20 %</b>

Tabela A.21: Korelacijski koeficienti med oceno  $LCV_{jsd}$  in točnostjo klasifikacij.

	$LCV_{sep}$						
	DT	NB	KNN	NN	RF	BAG	SVM
housevotes	0.128	0.378	0.415	0.203	0.423	0.381	-0.167
wine	0.721	0.674	0.486	0.010	0.495	0.614	0.228
parkinsons	0.250	0.700	0.181	0.335	0.229	0.380	-0.264
zoo	0.784	0.296	0.626	0.039	0.671	0.590	0.497
tictactoe	-0.513	0.313	-0.084	0.983	0.024	-0.248	0.122
postoperative	0.110	0.109	0.002	-0.022	0.107	0.235	-0.142
monks3	/	-0.093	0.438	0.851	-0.434	/	-0.412
iriset	0.447	0.694	0.636	-0.209	0.389	0.453	0.194
glass	0.317	0.544	0.341	-0.050	0.204	0.162	0.155
hungarian	0.392	0.320	0.366	-0.061	0.229	0.390	0.235
ecoli	0.340	0.545	0.350	0.020	0.350	0.250	0.177
heart	0.450	0.247	0.303	0.031	0.093	0.019	-0.218
haberman	0.264	0.116	0.096	0.002	-0.286	-0.051	0.069
flag	0.115	0.626	0.314	0.120	-0.212	0.028	-0.148
wdbc	0.466	0.438	0.395	0.064	0.357	0.456	-0.541
bcw	0.600	0.267	0.469	-0.133	0.373	0.482	0.227
sonar	0.381	0.490	-0.119	0.532	-0.295	-0.193	-0.598
soybean	0.785	-0.031	0.181	0.026	0.437	0.631	0.053
hepatitis	0.244	0.348	0.292	0.602	0.194	0.248	-0.009
lungcancer	0.155	-0.202	-0.110	-0.154	0.113	0.516	0.332
+	<b>75 %</b>	<b>75 %</b>	<b>70 %</b>	<b>25 %</b>	<b>35 %</b>	<b>50 %</b>	<b>30 %</b>
-	<b>5 %</b>	<b>0 %</b>	<b>0 %</b>	<b>5 %</b>	<b>10 %</b>	<b>5 %</b>	<b>20 %</b>

Tabela A.22: Korelacijski koeficienti med oceno  $LCV_{sep}$  in točnostjo klasifikacij.

### A.3 Korelacijski koeficienti ocen $LCV_d$

	$LCV_{dman}$						
	DT	NB	KNN	NN	RF	BAG	SVM
housevotes	-0.065	-0.161	-0.049	0.104	-0.180	-0.099	-0.178
wine	0.757	0.409	0.252	0.383	0.131	0.396	0.155
parkinsons	0.343	0.536	0.276	-0.162	0.396	0.381	-0.085
zoo	-0.311	-0.166	-0.209	0.204	-0.346	0.071	-0.347
tictactoe	-0.031	0.270	0.085	-0.916	0.414	0.023	-0.016
postoperative	-0.005	0.078	0.111	0.060	0.001	0.086	-0.025
monks3	/	0.020	0.673	-0.654	0.385	/	/
iriset	-0.104	0.069	0.009	0.149	-0.057	-0.004	-0.097
glass	0.088	0.242	0.050	-0.138	-0.097	-0.093	0.008
hungarian	-0.055	0.024	-0.008	-0.035	-0.223	-0.135	-0.126
ecoli	0.295	0.317	0.159	-0.032	-0.139	-0.061	-0.223
heart	0.107	0.136	0.093	-0.172	-0.241	-0.168	-0.138
haberman	-0.121	-0.239	-0.143	-0.473	-0.260	-0.108	-0.292
flag	0.239	0.591	0.167	0.307	0.069	0.077	-0.262
wdbc	0.201	0.195	0.205	-0.222	0.140	0.174	-0.246
bcw	-0.184	-0.021	-0.161	-0.893	-0.272	-0.226	-0.358
sonar	0.198	0.411	-0.097	0.208	-0.255	-0.219	-0.435
soybean	0.904	0.046	-0.419	0.594	0.174	0.223	0.266
hepatitis	-0.194	-0.031	-0.230	0.349	-0.490	-0.401	-0.060
lungcancer	-0.123	-0.215	0.111	0.268	0.404	0.490	-0.003
+	<b>35 %</b>	<b>40 %</b>	<b>35 %</b>	<b>30 %</b>	<b>10 %</b>	<b>15 %</b>	<b>0 %</b>
-	<b>10 %</b>	<b>10 %</b>	<b>10 %</b>	<b>35 %</b>	<b>25 %</b>	<b>15 %</b>	<b>35 %</b>

Tabela A.23: Korelacijski koeficienti med oceno  $LCV_{dman}$  in točnostjo klasifikacij.

### A.3 Korelacijski koeficienti ocen $LCV_d$

	$LCV_{d}evk$						
	DT	NB	KNN	NN	RF	BAG	SVM
housevotes	-0.065	-0.161	-0.049	0.139	-0.180	-0.099	-0.178
wine	0.746	0.409	0.252	0.361	0.089	0.365	0.136
parkinsons	0.343	0.536	0.276	-0.093	0.396	0.381	-0.085
zoo	-0.409	-0.198	-0.298	0.182	-0.478	-0.080	-0.551
tictactoe	-0.031	0.270	0.085	-0.916	0.414	0.023	-0.016
postoperative	-0.004	0.064	0.103	0.071	-0.005	0.076	-0.014
monks3	/	0.020	0.673	-0.654	0.385	/	/
irisset	-0.104	0.069	0.009	0.137	-0.057	-0.004	-0.105
glass	0.028	0.219	0.020	-0.190	-0.153	-0.142	-0.019
hungarian	-0.055	0.024	-0.008	-0.033	-0.223	-0.135	-0.126
ecoli	0.169	0.297	0.093	-0.020	-0.245	-0.190	-0.296
heart	0.107	0.136	0.093	-0.150	-0.241	-0.168	-0.138
haberman	-0.121	-0.239	-0.143	-0.473	-0.260	-0.108	-0.292
flag	0.072	0.577	0.013	0.098	-0.090	-0.109	-0.291
wdbc	0.201	0.195	0.205	-0.113	0.140	0.174	-0.246
bcw	-0.184	-0.021	-0.161	-0.873	-0.272	-0.226	-0.358
sonar	0.198	0.411	-0.097	0.219	-0.255	-0.219	-0.435
soybean	0.894	0.046	-0.419	0.587	0.096	0.165	0.219
hepatitis	-0.194	-0.031	-0.230	0.366	-0.490	-0.401	-0.060
lungcancer	-0.140	-0.212	0.076	0.240	0.386	0.464	-0.013
+	<b>30 %</b>	<b>40 %</b>	<b>25 %</b>	<b>20 %</b>	<b>10 %</b>	<b>15 %</b>	<b>0 %</b>
-	<b>10 %</b>	<b>10 %</b>	<b>15 %</b>	<b>30 %</b>	<b>25 %</b>	<b>15 %</b>	<b>35 %</b>

Tabela A.24: Korelacijski koeficienti med oceno  $LCV_{d}evk$  in točnostjo klasifikacij.

	$LCV_{d}max$						
	DT	NB	KNN	NN	RF	BAG	SVM
housevotes	-0.065	-0.161	-0.049	0.183	-0.180	-0.099	-0.178
wine	0.757	0.409	0.252	0.357	0.131	0.396	0.155
parkinsons	0.343	0.536	0.276	0.035	0.396	0.381	-0.085
zoo	-0.392	-0.167	-0.224	0.192	-0.391	-0.042	-0.546
tictactoe	-0.031	0.270	0.085	-0.916	0.414	0.023	-0.016
postoperative	-0.005	0.078	0.111	0.031	0.001	0.086	-0.025
monks3	/	0.020	0.673	-0.654	0.385	/	/
irisset	-0.104	0.069	0.009	0.148	-0.057	-0.004	-0.097
glass	0.068	0.242	0.044	-0.198	-0.130	-0.116	-0.014
hungarian	-0.055	0.024	-0.008	-0.030	-0.223	-0.135	-0.126
ecoli	0.240	0.317	0.125	-0.044	-0.191	-0.125	-0.278
heart	0.107	0.136	0.093	-0.117	-0.241	-0.168	-0.138
haberman	-0.121	-0.239	-0.143	-0.473	-0.260	-0.108	-0.292
flag	0.025	0.591	0.021	0.108	-0.119	-0.163	-0.300
wdbc	0.201	0.195	0.205	0.035	0.140	0.174	-0.246
bcw	-0.184	-0.021	-0.161	-0.836	-0.272	-0.226	-0.358
sonar	0.198	0.411	-0.097	0.230	-0.255	-0.219	-0.435
soybean	0.904	0.046	-0.419	0.508	0.173	0.199	0.218
hepatitis	-0.194	-0.031	-0.230	0.370	-0.490	-0.401	-0.060
lungcancer	-0.123	-0.215	0.111	0.080	0.404	0.490	-0.003
+	<b>30 %</b>	<b>40 %</b>	<b>30 %</b>	<b>20 %</b>	<b>10 %</b>	<b>15 %</b>	<b>0 %</b>
-	<b>10 %</b>	<b>10 %</b>	<b>10 %</b>	<b>20 %</b>	<b>25 %</b>	<b>15 %</b>	<b>35 %</b>

Tabela A.25: Korelacijski koeficienti med oceno  $LCV_{d}max$  in točnostjo klasifikacij.

Poglavje A: Podrobni rezultati testiranj

	$LCV_d\cos$						
	DT	NB	KNN	NN	RF	BAG	SVM
housevotes	-0.077	-0.194	-0.122	-0.729	-0.267	-0.186	-0.191
wine	0.680	0.320	0.120	0.211	-0.222	0.066	-0.046
parkinsons	0.321	0.534	0.244	-0.010	0.367	0.344	-0.119
zoo	-0.423	-0.240	-0.349	0.073	-0.513	-0.004	-0.403
tictactoe	-0.100	0.085	-0.051	/	0.279	-0.082	-0.057
postoperative	-0.009	0.047	0.079	-0.035	-0.017	0.042	-0.014
monks3	/	-0.124	0.638	-0.109	0.264	/	/
irisset	-0.231	-0.018	-0.113	-0.365	-0.158	-0.106	-0.205
glass	0.064	0.241	-0.013	0.015	-0.165	-0.147	-0.031
hungarian	-0.089	-0.029	-0.092	-0.330	-0.277	-0.208	-0.139
ecoli	0.206	0.171	-0.002	0.161	-0.324	-0.254	-0.320
heart	0.055	0.018	-0.070	-0.272	-0.321	-0.270	-0.137
haberman	-0.148	-0.238	-0.200	-0.155	-0.290	-0.158	-0.301
flag	0.283	0.577	0.154	0.223	0.080	0.147	-0.249
wdbc	0.168	0.173	0.130	-0.290	0.038	0.065	-0.285
bcw	-0.204	-0.028	-0.209	-0.909	-0.317	-0.274	-0.388
sonar	0.118	0.386	-0.226	-0.343	-0.308	-0.291	-0.456
soybean	0.775	0.046	-0.473	0.324	-0.303	-0.231	0.221
hepatitis	-0.286	-0.080	-0.313	0.173	-0.528	-0.470	-0.082
lungcancer	-0.194	-0.212	0.179	0.324	0.392	0.427	-0.017
+	<b>30 %</b>	<b>35 %</b>	<b>15 %</b>	<b>25 %</b>	<b>10 %</b>	<b>5 %</b>	<b>0 %</b>
-	<b>25 %</b>	<b>10 %</b>	<b>30 %</b>	<b>45 %</b>	<b>30 %</b>	<b>25 %</b>	<b>35 %</b>

Tabela A.26: Korelacijski koeficienti med oceno  $LCV_d\cos$  in točnostjo klasifikacij.

	$LCV_d\cot$						
	DT	NB	KNN	NN	RF	BAG	SVM
housevotes	-0.058	-0.118	-0.045	-0.694	-0.076	-0.121	-0.174
wine	0.772	0.412	0.268	0.219	0.101	0.215	0.168
parkinsons	0.351	0.368	0.284	-0.032	0.266	0.166	-0.071
zoo	-0.243	-0.175	-0.186	0.079	-0.341	0.096	-0.273
tictactoe	-0.036	0.258	0.105	/	0.345	0.131	0.006
postoperative	-0.008	0.065	0.125	-0.040	0.003	0.091	-0.023
monks3	/	0.060	0.650	-0.089	0.277	/	/
irisset	-0.089	0.069	0.015	-0.358	-0.153	0.014	-0.073
glass	0.139	0.253	0.074	0.027	0.079	0.144	0.053
hungarian	-0.027	0.044	0.003	-0.076	0.070	0.115	-0.116
ecoli	0.317	0.305	0.179	0.166	0.073	-0.057	-0.139
heart	0.133	0.145	0.134	-0.061	0.102	0.117	-0.149
haberman	-0.095	-0.233	-0.115	-0.028	0.126	0.102	-0.276
flag	0.349	0.588	0.236	0.333	0.216	0.223	-0.221
wdbc	0.118	0.171	0.209	-0.060	0.095	0.079	-0.229
bcw	-0.178	-0.021	-0.157	-0.903	-0.153	-0.122	-0.336
sonar	0.235	0.412	-0.076	-0.316	-0.039	0.174	-0.399
soybean	0.911	0.046	-0.438	0.315	0.035	0.079	0.338
hepatitis	-0.139	-0.031	-0.206	0.234	-0.182	-0.107	-0.040
lungcancer	-0.161	-0.217	0.110	0.295	0.428	0.496	0.001
+	<b>40 %</b>	<b>45 %</b>	<b>40 %</b>	<b>25 %</b>	<b>10 %</b>	<b>10 %</b>	<b>5 %</b>
-	<b>5 %</b>	<b>10 %</b>	<b>5 %</b>	<b>15 %</b>	<b>10 %</b>	<b>5 %</b>	<b>35 %</b>

Tabela A.27: Korelacijski koeficienti med oceno  $LCV_d\cot$  in točnostjo klasifikacij.

### A.3 Korelacijski koeficienti ocen $LCV_d$

	$LCV_d$ ska						
	DT	NB	KNN	NN	RF	BAG	SVM
housevotes	-0.039	-0.153	-0.017	0.745	-0.121	-0.071	-0.166
wine	0.784	0.414	0.312	0.606	0.447	0.506	0.371
parkinsons	0.374	0.536	0.327	0.280	0.461	0.425	-0.011
zoo	0.081	-0.153	0.007	0.197	-0.043	0.173	-0.075
tictactoe	0.109	0.342	0.233	0.913	0.593	0.160	0.104
postoperative	-0.037	0.075	0.163	0.029	0.019	0.181	0.004
monks3	/	0.286	0.813	0.966	0.647	/	/
irisset	-0.002	0.084	0.070	0.337	0.046	0.069	0.128
glass	0.193	0.251	0.165	-0.055	0.165	0.105	0.283
hungarian	0.024	0.065	0.095	0.555	-0.042	0.022	-0.088
ecoli	0.402	0.334	0.325	0.032	0.351	0.257	0.347
heart	0.163	0.181	0.304	0.757	0.152	0.181	-0.094
haberman	-0.092	-0.199	0.000	0.288	-0.108	0.063	-0.252
flag	0.486	0.592	0.363	0.682	0.620	0.514	-0.059
wdbc	0.219	0.197	0.239	0.551	0.256	0.268	-0.210
bcw	-0.175	-0.021	-0.134	0.476	-0.222	-0.181	-0.321
sonar	0.214	0.413	0.080	0.862	0.186	0.160	-0.240
soybean	0.913	0.046	-0.419	0.508	0.295	0.240	0.605
hepatitis	0.016	-0.015	-0.050	0.807	-0.206	-0.140	0.010
lungcancer	-0.106	-0.214	0.393	0.027	0.547	0.423	-0.167
+	45 %	50 %	40 %	75 %	30 %	30 %	20 %
-	5 %	10 %	5 %	0 %	5 %	5 %	25 %

Tabela A.28: Korelacijski koeficienti med oceno  $LCV_d$ ska in točnostjo klasifikacij.

	$LCV_d$ bha						
	DT	NB	KNN	NN	RF	BAG	SVM
housevotes	-0.124	-0.222	-0.145	0.682	-0.301	-0.199	-0.188
wine	0.485	0.273	-0.009	0.364	-0.264	-0.016	-0.205
parkinsons	0.298	0.526	0.221	0.211	0.353	0.325	-0.094
zoo	-0.520	-0.264	-0.421	0.151	-0.586	-0.087	-0.545
tictactoe	-0.230	0.071	-0.119	0.919	0.227	-0.174	-0.116
postoperative	-0.001	0.099	0.037	0.072	-0.018	0.018	0.022
monks3	/	-0.310	0.691	0.965	0.073	/	/
irisset	-0.296	-0.078	-0.172	0.188	-0.239	-0.180	-0.273
glass	-0.028	0.186	-0.054	-0.163	-0.208	-0.202	-0.049
hungarian	-0.166	-0.079	-0.131	0.435	-0.312	-0.259	-0.151
ecoli	0.004	0.131	-0.050	-0.090	-0.362	-0.320	-0.396
heart	-0.118	-0.060	-0.187	0.671	-0.382	-0.362	-0.047
haberman	-0.190	-0.240	-0.263	-0.126	-0.316	-0.213	-0.323
flag	0.153	0.560	0.076	0.103	-0.048	0.036	-0.225
wdbc	0.078	0.152	0.106	0.493	-0.017	0.017	-0.335
bcw	-0.249	-0.034	-0.230	0.248	-0.334	-0.294	-0.420
sonar	-0.076	0.361	-0.278	0.863	-0.352	-0.364	-0.461
soybean	0.758	0.046	-0.447	0.539	-0.177	-0.060	0.174
hepatitis	-0.399	-0.122	-0.363	0.811	-0.557	-0.507	-0.111
lungcancer	0.048	-0.191	0.252	0.115	0.370	0.328	-0.020
+	20 %	35 %	5 %	50 %	10 %	5 %	0 %
-	35 %	10 %	35 %	5 %	35 %	40 %	35 %

Tabela A.29: Korelacijski koeficienti med oceno  $LCV_d$ bha in točnostjo klasifikacij.

Poglavje A: Podrobni rezultati testiranj

	$LCV_d$ har						
	DT	NB	KNN	NN	RF	BAG	SVM
housevotes	-0.093	-0.195	-0.119	0.672	-0.271	-0.176	-0.195
wine	0.637	0.323	0.090	0.369	-0.187	0.108	-0.091
parkinsons	0.312	0.532	0.234	0.402	0.360	0.338	-0.127
zoo	-0.469	-0.226	-0.359	0.171	-0.513	-0.026	-0.467
tictactoe	-0.161	0.124	-0.071	0.916	0.257	-0.115	-0.087
postoperative	0.003	0.076	0.064	0.059	-0.012	0.033	0.004
monks3	/	-0.219	0.675	0.965	0.164	/	/
irisset	-0.242	-0.014	-0.118	0.209	-0.177	-0.116	-0.232
glass	0.015	0.225	-0.025	-0.136	-0.183	-0.175	-0.046
hungarian	-0.129	-0.039	-0.102	0.443	-0.294	-0.231	-0.147
ecoli	0.161	0.197	0.005	-0.079	-0.325	-0.255	-0.371
heart	-0.024	0.010	-0.117	0.678	-0.357	-0.325	-0.106
haberman	-0.171	-0.242	-0.227	-0.102	-0.305	-0.185	-0.316
flag	0.190	0.580	0.108	0.175	-0.019	0.046	-0.250
wdbc	0.147	0.174	0.132	0.540	0.024	0.059	-0.307
bcw	-0.217	-0.027	-0.211	0.309	-0.319	-0.277	-0.416
sonar	0.051	0.385	-0.242	0.864	-0.345	-0.340	-0.483
soybean	0.804	0.046	-0.447	0.530	-0.117	-0.027	0.181
hepatitis	-0.341	-0.083	-0.333	0.828	-0.548	-0.491	-0.100
lungcancer	-0.062	-0.207	0.200	0.107	0.382	0.399	-0.008
+	<b>30 %</b>	<b>35 %</b>	<b>5 %</b>	<b>70 %</b>	<b>10 %</b>	<b>5 %</b>	<b>0 %</b>
-	<b>30 %</b>	<b>10 %</b>	<b>30 %</b>	<b>0 %</b>	<b>35 %</b>	<b>35 %</b>	<b>35 %</b>

Tabela A.30: Korelacijski koeficienti med oceno  $LCV_d$ har in točnostjo klasifikacij.

	$LCV_d$ hel						
	DT	NB	KNN	NN	RF	BAG	SVM
housevotes	-0.124	-0.222	-0.145	0.181	-0.301	-0.199	-0.188
wine	0.485	0.273	-0.009	0.317	-0.264	-0.016	-0.205
parkinsons	0.298	0.526	0.221	-0.135	0.353	0.325	-0.094
zoo	-0.520	-0.264	-0.421	0.152	-0.586	-0.087	-0.545
tictactoe	-0.230	0.071	-0.119	-0.913	0.227	-0.174	-0.116
postoperative	-0.001	0.099	0.037	0.103	-0.018	0.018	0.022
monks3	/	-0.310	0.691	-0.638	0.073	/	/
irisset	-0.296	-0.078	-0.172	0.042	-0.239	-0.180	-0.273
glass	-0.028	0.186	-0.054	-0.183	-0.208	-0.202	-0.049
hungarian	-0.166	-0.079	-0.131	-0.036	-0.312	-0.259	-0.151
ecoli	0.004	0.131	-0.050	-0.107	-0.362	-0.320	-0.396
heart	-0.118	-0.060	-0.187	-0.166	-0.382	-0.362	-0.047
haberman	-0.190	-0.240	-0.263	-0.500	-0.316	-0.213	-0.323
flag	0.153	0.560	0.076	0.107	-0.048	0.036	-0.225
wdbc	0.078	0.152	0.106	-0.197	-0.017	0.017	-0.335
bcw	-0.249	-0.034	-0.230	-0.891	-0.334	-0.294	-0.420
sonar	-0.076	0.361	-0.278	0.245	-0.352	-0.364	-0.461
soybean	0.758	0.046	-0.447	0.642	-0.177	-0.060	0.174
hepatitis	-0.399	-0.122	-0.363	0.400	-0.557	-0.507	-0.111
lungcancer	0.048	-0.191	0.252	0.288	0.370	0.328	-0.020
+	<b>20 %</b>	<b>35 %</b>	<b>5 %</b>	<b>20 %</b>	<b>10 %</b>	<b>5 %</b>	<b>0 %</b>
-	<b>35 %</b>	<b>10 %</b>	<b>35 %</b>	<b>30 %</b>	<b>35 %</b>	<b>40 %</b>	<b>35 %</b>

Tabela A.31: Korelacijski koeficienti med oceno  $LCV_d$ hel in točnostjo klasifikacij.

### A.3 Korelacijski koeficienti ocen $LCV_d$

	$LCV_{djsd}$						
	DT	NB	KNN	NN	RF	BAG	SVM
housevotes	-0.132	-0.251	-0.176	0.180	-0.317	-0.224	-0.176
wine	0.408	0.156	-0.089	0.304	-0.293	-0.095	-0.222
parkinsons	0.295	0.485	0.221	-0.147	0.353	0.323	0.005
zoo	-0.575	-0.355	-0.497	-0.156	-0.633	-0.245	-0.598
tictactoe	-0.199	0.046	-0.100	-0.867	0.229	-0.153	-0.116
postoperative	-0.002	0.093	0.028	0.107	-0.019	0.014	0.020
monks3	/	-0.270	0.685	-0.645	0.117	/	/
irisset	-0.305	-0.128	-0.205	-0.079	-0.252	-0.203	-0.279
glass	-0.080	0.091	-0.074	-0.190	-0.215	-0.214	-0.053
hungarian	-0.172	-0.093	-0.139	-0.034	-0.313	-0.266	-0.151
ecoli	-0.134	-0.006	-0.115	-0.210	-0.375	-0.340	-0.399
heart	-0.128	-0.104	-0.211	-0.171	-0.382	-0.364	-0.035
haberman	-0.191	-0.240	-0.277	-0.509	-0.317	-0.219	-0.324
flag	0.090	0.497	-0.016	0.149	-0.059	0.006	-0.217
wdbc	0.045	0.075	0.084	-0.224	-0.025	0.002	-0.303
bcw	-0.260	-0.071	-0.246	-0.899	-0.342	-0.305	-0.405
sonar	-0.110	0.345	-0.271	0.206	-0.350	-0.361	-0.445
soybean	0.626	0.046	-0.456	0.538	-0.269	-0.162	0.150
hepatitis	-0.410	-0.174	-0.366	0.399	-0.557	-0.511	-0.111
lungcancer	0.005	-0.158	0.269	0.362	0.366	0.311	-0.026
+	15 %	15 %	5 %	25 %	10 %	5 %	0 %
-	35 %	15 %	35 %	35 %	45 %	40 %	30 %

Tabela A.32: Korelacijski koeficienti med oceno  $LCV_{djsd}$  in točnostjo klasifikacij.

	$LCV_{dsep}$						
	DT	NB	KNN	NN	RF	BAG	SVM
housevotes	-0.084	-0.082	-0.043	0.212	-0.201	-0.101	-0.233
wine	0.689	0.403	0.280	0.008	0.146	0.346	0.067
parkinsons	0.288	0.538	0.229	0.335	0.347	0.344	-0.129
zoo	0.272	-0.100	0.044	0.034	-0.134	0.224	-0.085
tictactoe	-0.528	0.198	-0.104	0.978	0.054	-0.250	0.005
postoperative	0.133	0.173	0.050	-0.004	0.071	0.171	-0.097
monks3	/	-0.106	0.472	0.853	-0.433	/	/
irisset	0.176	0.084	0.158	-0.209	-0.003	0.044	-0.113
glass	0.097	0.288	0.132	-0.110	-0.019	0.053	0.033
hungarian	0.032	0.045	-0.007	-0.058	-0.233	-0.144	-0.083
ecoli	0.270	0.310	0.238	0.020	0.021	0.056	-0.046
heart	0.250	0.094	0.025	0.037	-0.296	-0.277	-0.146
haberman	-0.045	-0.131	-0.176	-0.037	-0.373	-0.241	-0.247
flag	0.111	0.591	0.275	0.120	-0.229	0.023	-0.186
wdbc	0.216	0.218	0.202	0.065	0.123	0.161	-0.467
bcw	-0.060	0.039	-0.162	-0.126	-0.343	-0.264	-0.289
sonar	0.287	0.434	-0.255	0.504	-0.394	-0.362	-0.589
soybean	0.762	0.042	-0.428	0.027	0.111	0.399	0.143
hepatitis	-0.178	0.016	-0.168	0.622	-0.544	-0.468	-0.049
lungcancer	0.163	-0.196	-0.064	-0.154	0.236	0.520	0.145
+	40 %	40 %	25 %	25 %	10 %	20 %	0 %
-	5 %	10 %	10 %	5 %	35 %	25 %	25 %

Tabela A.33: Korelacijski koeficienti med oceno  $LCV_{dsep}$  in točnostjo klasifikacij.

## A.4 Korelacijski koeficienti ocen *DENS*

	<i>DENS</i>						
	DT	NB	KNN	NN	RF	BAG	SVM
housevotes	-0.016	-0.142	-0.080	0.306	-0.181	-0.134	-0.020
wine	0.029	-0.067	0.111	0.061	0.008	0.007	0.070
parkinsons	-0.158	-0.206	-0.138	0.011	-0.199	-0.164	-0.038
zoo	0.139	-0.174	-0.289	-0.129	-0.185	-0.208	-0.218
tictactoe	-0.220	-0.026	-0.170	-0.025	-0.355	-0.337	0.061
postoperative	-0.046	-0.118	-0.003	-0.168	-0.025	-0.041	-0.094
monks3	/	/	/	/	/	/	/
iriset	-0.303	-0.239	-0.187	-0.242	-0.223	-0.216	-0.199
glass	0.048	0.098	-0.053	0.149	0.043	0.086	-0.033
hungarian	0.017	-0.052	-0.044	0.127	0.040	0.029	0.005
ecoli	0.239	0.161	0.180	0.250	0.287	0.302	0.340
heart	-0.104	-0.134	-0.100	0.205	-0.096	-0.058	-0.122
haberman	0.113	0.133	0.196	0.399	0.184	0.164	0.184
flag	-0.086	-0.229	-0.134	-0.265	-0.223	-0.177	-0.133
wdbc	-0.044	-0.040	-0.139	0.215	-0.124	-0.115	0.071
bcw	-0.049	-0.073	-0.063	0.688	-0.075	-0.077	0.002
sonar	0.119	-0.233	-0.043	-0.123	0.060	0.136	0.018
soybean	0.487	-0.079	0.029	-0.079	-0.062	0.033	-0.181
hepatitis	0.431	0.220	0.260	-0.266	0.420	0.407	-0.067
lungcancer	-0.287	-0.188	-0.555	-0.301	-0.543	-0.429	0.111
+	15 %	10 %	10 %	30 %	10 %	15 %	5 %
-	10 %	20 %	10 %	15 %	20 %	25 %	10 %

Tabela A.34: Korelacijski koeficienti med oceno *DENS* in točnostjo klasifikacij.

A.5 Korelacijski koeficienti ocen *BAGV*

	<i>BAGV</i>						
	DT	NB	KNN	NN	RF	BAG	SVM
housevotes	0.536	0.255	0.574	0.044	0.550	0.622	-0.305
wine	0.563	0.507	0.477	-0.043	0.582	0.651	-0.765
parkinsons	0.433	0.199	0.709	-0.267	0.591	0.481	-0.377
zoo	0.812	0.538	0.723	0.474	0.610	0.775	-0.796
tictactoe	0.484	0.391	0.756	-0.013	0.667	0.668	-0.357
postoperative	-0.211	0.257	0.143	-0.004	-0.010	0.062	-0.169
monks3	/	0.656	0.320	0.011	0.641	/	-0.405
irisset	0.679	0.446	0.458	0.366	0.706	0.701	-0.788
glass	0.321	0.248	0.514	0.163	0.387	0.467	-0.592
hungarian	0.332	/	0.419	0.500	0.526	0.450	-0.230
ecoli	0.526	0.039	0.539	0.437	0.557	0.600	-0.518
heart	0.367	0.292	0.534	0.159	0.424	0.513	-0.244
haberman	0.243	0.046	0.305	0.543	0.351	0.393	-0.239
flag	0.522	0.423	0.413	-0.200	0.451	0.507	-0.540
wdbc	0.539	0.330	0.694	0.432	0.646	0.664	-0.537
bcw	0.530	0.226	0.587	0.250	0.646	0.638	-0.311
sonar	0.415	0.183	0.582	0.110	0.445	0.453	-0.719
soybean	0.951	-0.022	0.541	0.467	0.245	0.621	-0.470
hepatitis	0.614	0.449	0.599	-0.502	0.555	0.544	-0.188
lungcancer	-0.079	-0.111	0.291	-0.039	0.213	-0.073	-0.062
+	<b>85 %</b>	<b>70 %</b>	<b>80 %</b>	<b>40 %</b>	<b>75 %</b>	<b>80 %</b>	<b>0 %</b>
-	<b>5 %</b>	<b>0 %</b>	<b>0 %</b>	<b>10 %</b>	<b>0 %</b>	<b>0 %</b>	<b>75 %</b>

Tabela A.35: Korelacijski koeficienti med oceno *BAGV* in točnostjo klasifikacij.

## A.6 Korelacijski koeficienti ocen $TRANS_1$

	$TRANS_1$ man						
	DT	NB	KNN	NN	RF	BAG	SVM
housevotes	0.304	0.326	0.527	/	0.656	0.590	0.446
wine	0.208	0.521	0.589	-0.507	0.886	0.787	0.906
parkinsons	0.208	0.140	0.823	/	0.790	0.685	0.405
zoo	0.824	0.634	0.944	0.786	0.972	0.823	0.777
tictactoe	0.057	0.434	0.846	/	0.896	0.861	0.359
postoperative	-0.153	0.249	0.161	0.204	0.092	0.267	-0.404
monks3	/	0.810	-0.200	/	0.659	/	0.339
irisset	0.337	0.505	0.651	0.446	0.660	0.596	0.708
glass	0.045	0.173	0.613	-0.115	0.706	0.574	0.365
hungarian	0.319	0.188	0.557	/	0.571	0.573	0.223
ecoli	0.313	0.229	0.612	0.511	0.718	0.663	0.518
heart	-0.007	0.295	0.562	/	0.591	0.557	0.228
haberman	0.192	0.115	0.373	/	0.433	0.414	0.215
flag	0.253	0.105	0.452	0.216	0.788	0.691	0.498
wdbc	0.125	0.194	0.728	/	0.800	0.755	0.639
bcw	0.119	0.163	0.623	/	0.729	0.685	0.477
sonar	-0.027	0.183	0.723	/	0.835	0.669	0.703
soybean	0.850	1.000	1.000	1.000	0.965	0.986	0.807
hepatitis	0.337	0.352	0.548	/	0.722	0.678	0.419
lungcancer	0.262	-0.074	0.229	-0.090	0.228	0.059	0.334
+	45 %	70 %	80 %	25 %	80 %	80 %	80 %
-	0 %	0 %	5 %	5 %	0 %	0 %	0 %

Tabela A.36: Korelacijski koeficienti med oceno  $TRANS_1$ man in točnostjo klasifikacij.

## A.6 Korelacijski koeficienti ocen $TRANS_1$

	$TRANS_{1evk}$						
	DT	NB	KNN	NN	RF	BAG	SVM
housevotes	0.304	0.326	0.527	/	0.656	0.590	0.446
wine	0.211	0.521	0.589	-0.508	0.887	0.784	0.908
parkinsons	0.208	0.140	0.823	/	0.790	0.685	0.405
zoo	0.824	0.642	0.957	0.786	0.974	0.798	0.772
tictactoe	0.057	0.434	0.846	/	0.896	0.861	0.359
postoperative	-0.153	0.248	0.162	0.201	0.085	0.267	-0.405
monks3	/	0.810	-0.200	/	0.659	/	0.339
iriset	0.337	0.505	0.651	0.442	0.661	0.596	0.722
glass	0.044	0.171	0.616	-0.119	0.695	0.570	0.367
hungarian	0.319	0.188	0.557	/	0.571	0.573	0.223
ecoli	0.321	0.222	0.605	0.505	0.712	0.653	0.537
heart	-0.007	0.295	0.562	/	0.591	0.557	0.228
haberman	0.192	0.115	0.373	/	0.433	0.414	0.215
flag	0.239	0.102	0.448	0.117	0.768	0.680	0.487
wdbc	0.125	0.194	0.728	/	0.800	0.755	0.639
bcw	0.119	0.163	0.623	/	0.729	0.685	0.477
sonar	-0.027	0.183	0.723	/	0.835	0.669	0.703
soybean	0.850	1.000	1.000	1.000	0.972	0.973	0.810
hepatitis	0.337	0.352	0.548	/	0.722	0.678	0.419
lungcancer	0.230	-0.074	0.187	-0.088	0.191	0.032	0.328
+	45 %	70 %	80 %	20 %	80 %	80 %	80 %
-	0 %	0 %	5 %	5 %	0 %	0 %	0 %

Tabela A.37: Korelacijski koeficienti med oceno  $TRANS_{1evk}$  in točnostjo klasifikacij.

	$TRANS_{1max}$						
	DT	NB	KNN	NN	RF	BAG	SVM
housevotes	0.304	0.326	0.527	/	0.656	0.590	0.446
wine	0.208	0.521	0.589	-0.507	0.886	0.787	0.906
parkinsons	0.208	0.140	0.823	/	0.790	0.685	0.405
zoo	0.824	0.634	0.944	0.786	0.967	0.816	0.757
tictactoe	0.057	0.434	0.846	/	0.896	0.861	0.359
postoperative	-0.153	0.249	0.161	0.204	0.092	0.267	-0.404
monks3	/	0.810	-0.200	/	0.659	/	0.339
iriset	0.337	0.505	0.651	0.446	0.660	0.596	0.708
glass	0.049	0.173	0.613	-0.123	0.702	0.572	0.365
hungarian	0.319	0.188	0.557	/	0.571	0.573	0.223
ecoli	0.305	0.229	0.612	0.507	0.717	0.661	0.511
heart	-0.007	0.295	0.562	/	0.591	0.557	0.228
haberman	0.192	0.115	0.373	/	0.433	0.414	0.215
flag	0.242	0.105	0.452	0.105	0.776	0.677	0.472
wdbc	0.125	0.194	0.728	/	0.800	0.755	0.639
bcw	0.119	0.163	0.623	/	0.729	0.685	0.477
sonar	-0.027	0.183	0.723	/	0.835	0.669	0.703
soybean	0.850	1.000	1.000	1.000	0.965	0.986	0.807
hepatitis	0.337	0.352	0.548	/	0.722	0.678	0.419
lungcancer	0.262	-0.074	0.229	-0.090	0.228	0.059	0.334
+	45 %	70 %	80 %	20 %	80 %	80 %	80 %
-	0 %	0 %	5 %	5 %	0 %	0 %	0 %

Tabela A.38: Korelacijski koeficienti med oceno  $TRANS_{1max}$  in točnostjo klasifikacij.

Poglavje A: Podrobni rezultati testiranj

	$TRANS_1\cos$						
	DT	NB	KNN	NN	RF	BAG	SVM
housevotes	0.285	0.188	0.438	/	0.531	0.486	0.342
wine	0.179	0.418	0.455	-0.522	0.762	0.564	0.739
parkinsons	0.136	0.083	0.660	/	0.630	0.451	0.333
zoo	0.608	0.555	0.817	0.789	0.837	0.619	0.681
tictactoe	-0.008	0.352	0.770	/	0.814	0.718	0.281
postoperative	-0.149	0.226	0.115	0.210	0.075	0.231	-0.358
monks3	/	0.834	-0.200	/	0.639	/	0.363
irisset	0.332	0.329	0.568	0.409	0.511	0.439	0.498
glass	-0.021	0.120	0.492	-0.042	0.577	0.432	0.313
hungarian	0.242	0.080	0.432	/	0.447	0.443	0.186
ecoli	0.280	0.025	0.519	0.432	0.575	0.523	0.310
heart	-0.023	0.148	0.441	/	0.459	0.429	0.174
haberman	0.149	0.110	0.274	/	0.332	0.287	0.151
flag	0.206	0.084	0.336	0.130	0.654	0.527	0.428
wdbc	0.060	0.158	0.634	/	0.661	0.639	0.512
bcw	0.062	0.110	0.520	/	0.598	0.542	0.347
sonar	-0.012	0.131	0.616	/	0.727	0.564	0.603
soybean	0.850	1.000	1.000	1.000	0.931	0.830	0.795
hepatitis	0.268	0.262	0.439	/	0.605	0.516	0.299
lungcancer	0.336	-0.022	0.058	-0.103	0.061	0.021	0.391
+	35 %	55 %	80 %	15 %	80 %	80 %	80 %
-	0 %	0 %	5 %	5 %	0 %	0 %	0 %

Tabela A.39: Korelacijski koeficienti med oceno  $TRANS_1\cos$  in točnostjo klasifikacij.

	$TRANS_1kot$						
	DT	NB	KNN	NN	RF	BAG	SVM
housevotes	0.299	0.287	0.517	/	0.653	0.580	0.424
wine	0.213	0.518	0.564	-0.485	0.890	0.752	0.901
parkinsons	0.225	0.127	0.806	/	0.772	0.652	0.424
zoo	0.804	0.631	0.950	0.789	0.957	0.767	0.806
tictactoe	0.059	0.447	0.840	/	0.923	0.861	0.353
postoperative	-0.154	0.260	0.173	0.210	0.079	0.268	-0.401
monks3	/	0.832	-0.200	/	0.761	/	0.372
irisset	0.336	0.484	0.637	0.444	0.627	0.561	0.721
glass	0.036	0.167	0.589	-0.107	0.681	0.554	0.402
hungarian	0.308	0.192	0.534	/	0.564	0.555	0.222
ecoli	0.339	0.205	0.603	0.517	0.686	0.634	0.540
heart	0.002	0.302	0.542	/	0.588	0.547	0.229
haberman	0.197	0.131	0.360	/	0.416	0.398	0.216
flag	0.285	0.095	0.418	0.278	0.757	0.648	0.513
wdbc	0.126	0.191	0.719	/	0.792	0.749	0.660
bcw	0.119	0.148	0.606	/	0.710	0.673	0.469
sonar	-0.023	0.180	0.711	/	0.846	0.682	0.691
soybean	0.850	1.000	1.000	1.000	0.987	0.956	0.833
hepatitis	0.347	0.341	0.533	/	0.710	0.654	0.413
lungcancer	0.232	-0.058	0.145	-0.097	0.119	-0.002	0.305
+	45 %	75 %	80 %	20 %	80 %	80 %	80 %
-	0 %	0 %	5 %	5 %	0 %	0 %	0 %

Tabela A.40: Korelacijski koeficienti med oceno  $TRANS_1kot$  in točnostjo klasifikacij.

## A.6 Korelacijski koeficienti ocen $TRANS_1$

	$TRANS_{1ska}$						
	DT	NB	KNN	NN	RF	BAG	SVM
housevotes	0.338	0.303	0.560	/	0.727	0.651	0.440
wine	0.329	0.612	0.557	0.066	0.933	0.842	0.928
parkinsons	0.491	0.122	0.852	/	0.809	0.732	0.578
zoo	0.911	0.614	0.962	0.786	0.983	0.858	0.889
tictactoe	0.672	0.602	0.847	/	0.968	0.897	0.453
postoperative	-0.932	0.184	0.227	0.203	0.041	0.271	0.174
monks3	/	0.736	-0.200	/	0.956	/	0.408
iriset	0.422	0.597	0.592	0.534	0.671	0.644	0.903
glass	0.292	0.144	0.621	0.107	0.764	0.650	0.592
hungarian	0.397	0.432	0.553	/	0.646	0.614	0.323
ecoli	0.433	0.516	0.627	0.635	0.749	0.708	0.701
heart	0.213	0.414	0.567	/	0.675	0.651	0.320
haberman	0.266	0.068	0.393	/	0.468	0.497	0.260
flag	0.542	0.076	0.455	0.676	0.857	0.728	0.508
wdbc	0.323	0.267	0.732	/	0.822	0.803	0.766
bcw	0.331	0.222	0.624	/	0.752	0.749	0.584
sonar	-0.047	0.217	0.736	/	0.909	0.805	0.719
soybean	0.850	1.000	1.000	1.000	0.997	0.996	0.803
hepatitis	0.569	0.368	0.574	/	0.735	0.708	0.385
lungcancer	-0.033	-0.074	0.322	-0.090	0.246	0.050	-0.104
+	<b>60 %</b>	<b>60 %</b>	<b>80 %</b>	<b>25 %</b>	<b>85 %</b>	<b>80 %</b>	<b>75 %</b>
-	<b>5 %</b>	<b>0 %</b>	<b>5 %</b>	<b>0 %</b>	<b>0 %</b>	<b>0 %</b>	<b>0 %</b>

Tabela A.41: Korelacijski koeficienti med oceno  $TRANS_{1ska}$  in točnostjo klasifikacij.

	$TRANS_{1bha}$						
	DT	NB	KNN	NN	RF	BAG	SVM
housevotes	0.286	0.326	0.439	/	0.536	0.502	0.401
wine	0.144	0.410	0.575	-0.537	0.725	0.628	0.840
parkinsons	0.126	0.142	0.701	/	0.676	0.538	0.248
zoo	0.652	0.624	0.816	0.768	0.869	0.746	0.688
tictactoe	-0.027	0.313	0.820	/	0.669	0.678	0.300
postoperative	-0.149	0.212	0.064	0.203	0.048	0.215	-0.298
monks3	/	0.801	-0.200	/	0.190	/	-0.226
iriset	0.333	0.352	0.607	0.414	0.624	0.549	0.488
glass	0.046	0.139	0.489	-0.049	0.612	0.465	0.290
hungarian	0.281	0.058	0.488	/	0.441	0.476	0.210
ecoli	0.264	0.039	0.529	0.414	0.636	0.603	0.333
heart	-0.029	0.153	0.492	/	0.439	0.415	0.177
haberman	0.147	0.089	0.244	/	0.338	0.279	0.148
flag	0.222	0.117	0.370	0.058	0.661	0.583	0.428
wdbc	0.047	0.154	0.672	/	0.682	0.622	0.479
bcw	0.065	0.129	0.571	/	0.665	0.550	0.335
sonar	-0.006	0.086	0.660	/	0.668	0.412	0.643
soybean	0.796	1.000	1.000	1.000	0.719	0.873	0.736
hepatitis	0.219	0.263	0.474	/	0.624	0.575	0.292
lungcancer	0.378	-0.065	0.087	-0.082	0.154	0.243	0.409
+	<b>25 %</b>	<b>60 %</b>	<b>80 %</b>	<b>20 %</b>	<b>75 %</b>	<b>80 %</b>	<b>75 %</b>
-	<b>0 %</b>	<b>0 %</b>	<b>5 %</b>	<b>5 %</b>	<b>0 %</b>	<b>0 %</b>	<b>5 %</b>

Tabela A.42: Korelacijski koeficienti med oceno  $TRANS_{1bha}$  in točnostjo klasifikacij.

Poglavje A: Podrobni rezultati testiranj

	<i>TRANS</i> <sub>1</sub> har						
	DT	NB	KNN	NN	RF	BAG	SVM
housevotes	0.286	0.326	0.472	/	0.565	0.533	0.404
wine	0.157	0.420	0.599	-0.524	0.789	0.677	0.851
parkinsons	0.128	0.138	0.732	/	0.715	0.582	0.259
zoo	0.652	0.617	0.844	0.780	0.902	0.753	0.695
tictactoe	-0.015	0.313	0.839	/	0.732	0.765	0.300
postoperative	-0.149	0.219	0.080	0.203	0.071	0.234	-0.318
monks3	/	0.801	-0.200	/	0.314	/	-0.226
irisset	0.333	0.352	0.628	0.423	0.632	0.556	0.497
glass	0.028	0.138	0.527	-0.060	0.636	0.494	0.298
hungarian	0.281	0.061	0.504	/	0.471	0.498	0.210
ecoli	0.272	0.062	0.551	0.434	0.658	0.612	0.339
heart	-0.029	0.155	0.513	/	0.471	0.447	0.178
haberman	0.150	0.089	0.285	/	0.362	0.321	0.149
flag	0.214	0.112	0.390	0.100	0.707	0.610	0.432
wdbc	0.051	0.157	0.691	/	0.724	0.675	0.497
bcw	0.067	0.129	0.583	/	0.682	0.595	0.350
sonar	-0.016	0.113	0.678	/	0.700	0.512	0.657
soybean	0.842	1.000	1.000	1.000	0.863	0.949	0.750
hepatitis	0.249	0.268	0.491	/	0.653	0.599	0.301
lungcancer	0.373	-0.063	0.120	-0.090	0.149	0.126	0.405
+	<b>30 %</b>	<b>55 %</b>	<b>80 %</b>	<b>20 %</b>	<b>80 %</b>	<b>80 %</b>	<b>75 %</b>
-	<b>0 %</b>	<b>0 %</b>	<b>5 %</b>	<b>5 %</b>	<b>0 %</b>	<b>0 %</b>	<b>5 %</b>

Tabela A.43: Korelacijski koeficienti med oceno *TRANS*<sub>1</sub>har in točnostjo klasifikacij.

	<i>TRANS</i> <sub>1</sub> hel						
	DT	NB	KNN	NN	RF	BAG	SVM
housevotes	0.286	0.326	0.439	/	0.536	0.502	0.401
wine	0.144	0.410	0.575	-0.537	0.725	0.628	0.840
parkinsons	0.126	0.142	0.701	/	0.676	0.538	0.248
zoo	0.652	0.624	0.816	0.768	0.869	0.746	0.688
tictactoe	-0.027	0.313	0.820	/	0.669	0.678	0.300
postoperative	-0.149	0.212	0.064	0.203	0.048	0.215	-0.298
monks3	/	0.801	-0.200	/	0.190	/	-0.226
irisset	0.333	0.352	0.607	0.414	0.624	0.549	0.488
glass	0.046	0.139	0.489	-0.049	0.612	0.465	0.290
hungarian	0.281	0.058	0.488	/	0.441	0.476	0.210
ecoli	0.264	0.039	0.529	0.414	0.636	0.603	0.333
heart	-0.029	0.153	0.492	/	0.439	0.415	0.177
haberman	0.147	0.089	0.244	/	0.338	0.279	0.148
flag	0.222	0.117	0.370	0.058	0.661	0.583	0.428
wdbc	0.047	0.154	0.672	/	0.682	0.622	0.479
bcw	0.065	0.129	0.571	/	0.665	0.550	0.335
sonar	-0.006	0.086	0.660	/	0.668	0.412	0.643
soybean	0.796	1.000	1.000	1.000	0.719	0.873	0.736
hepatitis	0.219	0.263	0.474	/	0.624	0.575	0.292
lungcancer	0.378	-0.065	0.087	-0.082	0.154	0.243	0.409
+	<b>25 %</b>	<b>60 %</b>	<b>80 %</b>	<b>20 %</b>	<b>75 %</b>	<b>80 %</b>	<b>75 %</b>
-	<b>0 %</b>	<b>0 %</b>	<b>5 %</b>	<b>5 %</b>	<b>0 %</b>	<b>0 %</b>	<b>5 %</b>

Tabela A.44: Korelacijski koeficienti med oceno *TRANS*<sub>1</sub>hel in točnostjo klasifikacij.

## A.6 Korelacijski koeficienti ocen $TRANS_1$

	$TRANS_{1jsd}$						
	DT	NB	KNN	NN	RF	BAG	SVM
housevotes	0.286	0.326	0.459	/	0.553	0.522	0.402
wine	0.150	0.414	0.595	-0.528	0.764	0.658	0.844
parkinsons	0.127	0.140	0.717	/	0.699	0.566	0.252
zoo	0.652	0.619	0.833	0.774	0.891	0.748	0.691
tictactoe	-0.021	0.313	0.836	/	0.703	0.731	0.300
postoperative	-0.149	0.215	0.069	0.203	0.061	0.227	-0.305
monks3	/	0.801	-0.200	/	0.257	/	-0.226
iriset	0.333	0.352	0.620	0.418	0.631	0.555	0.491
glass	0.037	0.137	0.509	-0.056	0.626	0.484	0.293
hungarian	0.281	0.059	0.498	/	0.458	0.489	0.210
ecoli	0.268	0.050	0.542	0.425	0.650	0.609	0.335
heart	-0.029	0.154	0.504	/	0.456	0.433	0.177
haberman	0.148	0.089	0.266	/	0.352	0.305	0.149
flag	0.218	0.114	0.380	0.079	0.688	0.598	0.430
wdbc	0.049	0.155	0.681	/	0.706	0.654	0.486
bcw	0.066	0.129	0.576	/	0.675	0.578	0.341
sonar	-0.012	0.099	0.668	/	0.682	0.472	0.649
soybean	0.830	1.000	1.000	1.000	0.811	0.926	0.741
hepatitis	0.235	0.265	0.482	/	0.641	0.591	0.295
lungcancer	0.376	-0.065	0.100	-0.087	0.147	0.173	0.408
+	<b>30 %</b>	<b>55 %</b>	<b>80 %</b>	<b>20 %</b>	<b>75 %</b>	<b>80 %</b>	<b>75 %</b>
-	<b>0 %</b>	<b>0 %</b>	<b>5 %</b>	<b>5 %</b>	<b>0 %</b>	<b>0 %</b>	<b>5 %</b>

Tabela A.45: Korelacijski koeficienti med oceno  $TRANS_{1jsd}$  in točnostjo klasifikacij.

	$TRANS_{1sep}$						
	DT	NB	KNN	NN	RF	BAG	SVM
housevotes	0.301	0.300	0.568	/	0.153	0.384	0.135
wine	0.112	0.529	0.579	-0.531	0.208	0.265	0.450
parkinsons	0.242	0.137	0.826	/	0.301	0.158	0.122
zoo	0.796	0.633	0.949	0.031	0.381	0.591	0.630
tictactoe	0.036	0.436	0.558	/	0.494	0.217	0.245
postoperative	-0.149	0.266	0.176	0.134	0.009	0.145	-0.328
monks3	/	0.826	-0.200	-0.046	-0.092	/	0.364
iriset	0.337	0.498	0.645	-0.012	0.206	0.403	0.327
glass	0.006	0.177	0.494	-0.277	0.329	0.219	0.127
hungarian	0.307	0.209	0.552	/	0.116	0.346	0.204
ecoli	0.339	0.255	0.621	0.103	0.260	0.405	-0.071
heart	0.002	0.322	0.560	/	0.551	0.287	0.212
haberman	0.187	0.133	0.408	/	0.097	0.142	0.172
flag	0.180	0.102	0.443	-0.121	0.285	0.325	0.420
wdbc	0.187	0.200	0.727	/	0.257	0.341	0.106
bcw	0.102	0.153	0.621	/	0.322	0.291	0.129
sonar	-0.056	0.186	0.726	/	0.777	0.577	0.223
soybean	0.850	1.000	1.000	0.066	-0.051	0.256	0.697
hepatitis	0.378	0.350	0.549	/	0.331	0.596	0.357
lungcancer	0.376	-0.074	0.244	-0.179	-0.096	-0.018	0.314
+	<b>40 %</b>	<b>70 %</b>	<b>80 %</b>	<b>0 %</b>	<b>65 %</b>	<b>80 %</b>	<b>45 %</b>
-	<b>0 %</b>	<b>0 %</b>	<b>5 %</b>	<b>10 %</b>	<b>0 %</b>	<b>0 %</b>	<b>0 %</b>

Tabela A.46: Korelacijski koeficienti med oceno  $TRANS_{1sep}$  in točnostjo klasifikacij.

## A.7 Korelacijski koeficienti ocen $TRANS_2$

	$TRANS_{2man}$						
	DT	NB	KNN	NN	RF	BAG	SVM
housevotes	0.288	0.282	0.288	/	0.377	0.239	0.499
wine	0.524	0.439	0.090	-0.375	-0.094	0.037	0.619
parkinsons	0.144	0.018	-0.078	/	-0.724	-0.405	0.152
zoo	0.436	0.505	0.164	0.392	0.597	0.211	0.732
tictactoe	0.503	0.379	-0.083	/	-0.729	-0.176	0.293
postoperative	-0.437	-0.227	0.002	0.016	-0.176	-0.046	-0.141
monks3	/	0.252	0.648	/	-0.342	/	-0.421
irisset	0.272	0.599	0.206	0.258	0.110	0.093	0.522
glass	0.070	0.111	0.144	-0.029	-0.570	-0.292	0.236
hungarian	0.070	0.091	0.304	/	-0.053	0.120	0.200
ecoli	0.425	0.193	0.237	0.264	0.185	0.163	0.464
heart	0.189	0.280	0.186	/	-0.433	-0.135	0.093
haberman	0.064	0.082	0.425	/	0.154	0.146	0.300
flag	0.196	-0.111	0.175	-0.433	-0.208	-0.395	0.443
wdbc	0.308	0.216	0.128	/	0.024	0.116	0.452
bcw	0.587	0.231	0.234	/	0.318	0.280	0.358
sonar	0.339	0.128	0.247	/	-0.808	-0.395	0.562
soybean	1.000	0.261	0.045	-0.512	-0.031	-0.169	0.345
hepatitis	0.550	0.213	-0.125	/	-0.606	-0.632	0.180
lungcancer	-0.451	0.118	-0.172	0.028	0.032	-0.002	-0.271
+	<b>70 %</b>	<b>50 %</b>	<b>25 %</b>	<b>15 %</b>	<b>10 %</b>	<b>10 %</b>	<b>60 %</b>
-	<b>0 %</b>	<b>0 %</b>	<b>5 %</b>	<b>15 %</b>	<b>40 %</b>	<b>35 %</b>	<b>0 %</b>

Tabela A.47: Korelacijski koeficienti med oceno  $TRANS_{2man}$  in točnostjo klasifikacij.

## A.7 Korelacijski koeficienti ocen $TRANS_2$

	$TRANS_{2evk}$						
	DT	NB	KNN	NN	RF	BAG	SVM
housevotes	0.288	0.282	0.288	/	0.377	0.239	0.499
wine	0.527	0.439	0.090	-0.373	-0.119	0.022	0.619
parkinsons	0.144	0.018	-0.078	/	-0.724	-0.405	0.152
zoo	0.436	0.493	0.134	0.367	0.537	0.194	0.701
tictactoe	0.503	0.379	-0.083	/	-0.729	-0.176	0.293
postoperative	-0.436	-0.241	-0.002	0.006	-0.197	-0.060	-0.133
monks3	/	0.252	0.648	/	-0.342	/	-0.421
iriset	0.272	0.599	0.206	0.251	0.110	0.093	0.530
glass	0.067	0.110	0.125	-0.043	-0.642	-0.351	0.222
hungarian	0.070	0.091	0.304	/	-0.053	0.120	0.200
ecoli	0.396	0.191	0.223	0.252	0.110	0.110	0.461
heart	0.189	0.280	0.186	/	-0.433	-0.135	0.093
haberman	0.064	0.082	0.425	/	0.154	0.146	0.300
flag	0.193	-0.116	0.079	-0.526	-0.519	-0.560	0.435
wdbc	0.308	0.216	0.128	/	0.024	0.116	0.452
bcw	0.587	0.231	0.234	/	0.318	0.280	0.358
sonar	0.339	0.128	0.247	/	-0.808	-0.395	0.562
soybean	1.000	0.261	0.045	-0.510	-0.094	-0.191	0.308
hepatitis	0.550	0.213	-0.125	/	-0.606	-0.632	0.180
lungcancer	-0.453	0.118	-0.238	0.027	-0.082	-0.010	-0.309
+	<b>70 %</b>	<b>50 %</b>	<b>25 %</b>	<b>15 %</b>	<b>10 %</b>	<b>10 %</b>	<b>60 %</b>
-	<b>0 %</b>	<b>0 %</b>	<b>5 %</b>	<b>15 %</b>	<b>40 %</b>	<b>35 %</b>	<b>0 %</b>

Tabela A.48: Korelacijski koeficienti med oceno  $TRANS_{2evk}$  in točnostjo klasifikacij.

	$TRANS_{2max}$						
	DT	NB	KNN	NN	RF	BAG	SVM
housevotes	0.288	0.282	0.288	/	0.377	0.239	0.499
wine	0.524	0.439	0.090	-0.375	-0.094	0.037	0.619
parkinsons	0.144	0.018	-0.078	/	-0.724	-0.405	0.152
zoo	0.436	0.505	0.164	0.392	0.597	0.211	0.716
tictactoe	0.503	0.379	-0.083	/	-0.729	-0.176	0.293
postoperative	-0.437	-0.227	0.002	0.016	-0.176	-0.046	-0.141
monks3	/	0.252	0.648	/	-0.342	/	-0.421
iriset	0.272	0.599	0.206	0.258	0.110	0.093	0.522
glass	0.067	0.111	0.144	-0.062	-0.570	-0.294	0.238
hungarian	0.070	0.091	0.304	/	-0.053	0.120	0.200
ecoli	0.396	0.193	0.237	0.262	0.175	0.159	0.465
heart	0.189	0.280	0.186	/	-0.433	-0.135	0.093
haberman	0.064	0.082	0.425	/	0.154	0.146	0.300
flag	0.197	-0.111	0.174	-0.501	-0.210	-0.425	0.434
wdbc	0.308	0.216	0.128	/	0.024	0.116	0.452
bcw	0.587	0.231	0.234	/	0.318	0.280	0.358
sonar	0.339	0.128	0.247	/	-0.808	-0.395	0.562
soybean	1.000	0.261	0.045	-0.512	-0.030	-0.169	0.350
hepatitis	0.550	0.213	-0.125	/	-0.606	-0.632	0.180
lungcancer	-0.451	0.118	-0.172	0.028	0.032	-0.002	-0.271
+	<b>70 %</b>	<b>50 %</b>	<b>25 %</b>	<b>15 %</b>	<b>10 %</b>	<b>10 %</b>	<b>60 %</b>
-	<b>0 %</b>	<b>0 %</b>	<b>5 %</b>	<b>15 %</b>	<b>40 %</b>	<b>35 %</b>	<b>0 %</b>

Tabela A.49: Korelacijski koeficienti med oceno  $TRANS_{2max}$  in točnostjo klasifikacij.

Poglavje A: Podrobni rezultati testiranj

	$TRANS_2\cos$						
	DT	NB	KNN	NN	RF	BAG	SVM
housevotes	0.285	0.175	0.257	/	0.381	0.183	0.465
wine	0.509	0.321	0.099	-0.337	-0.124	0.002	0.574
parkinsons	0.114	-0.005	-0.050	/	-0.648	-0.359	0.110
zoo	0.416	0.501	0.140	0.408	0.663	0.197	0.651
tictactoe	0.439	0.392	-0.188	/	-0.501	-0.206	0.284
postoperative	-0.396	-0.139	-0.019	0.022	-0.129	-0.114	-0.066
monks3	/	0.606	0.674	/	-0.151	/	-0.429
irisset	0.240	0.466	0.262	0.214	0.098	0.061	0.456
glass	0.048	0.083	0.161	0.059	-0.542	-0.280	0.239
hungarian	0.050	0.036	0.279	/	-0.013	0.059	0.218
ecoli	0.371	0.101	0.196	0.246	0.194	0.101	0.377
heart	0.172	0.147	0.116	/	-0.419	-0.195	0.136
haberman	0.030	0.123	0.385	/	0.190	0.066	0.215
flag	0.157	-0.116	0.203	-0.456	-0.164	-0.345	0.441
wdbc	0.294	0.165	0.091	/	-0.033	0.066	0.385
bcw	0.571	0.118	0.211	/	0.294	0.239	0.291
sonar	0.327	0.073	0.212	/	-0.802	-0.446	0.519
soybean	0.966	0.042	0.038	-0.448	0.093	-0.154	0.326
hepatitis	0.506	0.156	-0.126	/	-0.513	-0.570	0.192
lungcancer	-0.435	0.126	-0.145	0.043	0.072	0.041	-0.370
+	<b>55 %</b>	<b>45 %</b>	<b>35 %</b>	<b>15 %</b>	<b>15 %</b>	<b>10 %</b>	<b>60 %</b>
-	<b>10 %</b>	<b>0 %</b>	<b>5 %</b>	<b>10 %</b>	<b>30 %</b>	<b>35 %</b>	<b>5 %</b>

Tabela A.50: Korelacijski koeficienti med oceno  $TRANS_2\cos$  in točnostjo klasifikacij.

	$TRANS_2kot$						
	DT	NB	KNN	NN	RF	BAG	SVM
housevotes	0.288	0.254	0.309	/	0.426	0.261	0.504
wine	0.533	0.457	0.115	-0.343	0.001	0.089	0.657
parkinsons	0.146	0.015	0.019	/	-0.652	-0.290	0.190
zoo	0.551	0.502	0.236	0.391	0.676	0.269	0.735
tictactoe	0.505	0.460	-0.049	/	-0.515	-0.086	0.321
postoperative	-0.434	-0.180	0.029	0.024	-0.176	-0.039	-0.118
monks3	/	0.539	0.631	/	-0.142	/	-0.428
irisset	0.280	0.588	0.253	0.267	0.145	0.123	0.611
glass	0.087	0.114	0.227	-0.006	-0.523	-0.222	0.261
hungarian	0.074	0.096	0.347	/	0.062	0.160	0.210
ecoli	0.439	0.210	0.280	0.289	0.259	0.192	0.478
heart	0.195	0.284	0.222	/	-0.360	-0.077	0.133
haberman	0.070	0.107	0.442	/	0.220	0.186	0.296
flag	0.210	-0.112	0.205	-0.391	-0.169	-0.355	0.478
wdbc	0.308	0.209	0.155	/	0.073	0.158	0.471
bcw	0.576	0.199	0.259	/	0.351	0.310	0.376
sonar	0.326	0.134	0.283	/	-0.785	-0.332	0.581
soybean	1.000	0.206	0.057	-0.496	0.126	-0.079	0.338
hepatitis	0.552	0.209	-0.061	/	-0.515	-0.570	0.217
lungcancer	-0.471	0.107	-0.146	0.036	0.077	0.027	-0.395
+	<b>60 %</b>	<b>55 %</b>	<b>35 %</b>	<b>15 %</b>	<b>15 %</b>	<b>10 %</b>	<b>60 %</b>
-	<b>10 %</b>	<b>0 %</b>	<b>0 %</b>	<b>10 %</b>	<b>30 %</b>	<b>30 %</b>	<b>5 %</b>

Tabela A.51: Korelacijski koeficienti med oceno  $TRANS_2kot$  in točnostjo klasifikacij.

## A.7 Korelacijski koeficienti ocen $TRANS_2$

	$TRANS_2$ ska						
	DT	NB	KNN	NN	RF	BAG	SVM
housevotes	0.374	0.313	0.340	/	0.531	0.351	0.511
wine	0.561	0.577	0.138	0.426	0.190	0.214	0.708
parkinsons	0.346	0.115	0.160	/	-0.471	-0.056	0.465
zoo	0.960	0.507	0.434	0.392	0.827	0.409	0.850
tictactoe	0.743	0.603	-0.076	/	0.255	0.079	0.468
postoperative	-0.484	0.012	0.089	0.035	-0.148	0.088	0.195
monks3	/	0.741	0.619	/	0.586	/	0.408
iriset	0.374	0.608	0.317	0.328	0.210	0.185	0.876
glass	0.206	0.197	0.386	0.329	0.011	0.164	0.549
hungarian	0.310	0.405	0.427	/	0.337	0.278	0.349
ecoli	0.536	0.532	0.407	0.431	0.473	0.353	0.709
heart	0.192	0.431	0.288	/	-0.112	0.111	0.288
haberman	0.115	0.071	0.481	/	0.362	0.324	0.322
flag	0.523	-0.105	0.399	0.118	0.558	0.289	0.536
wdbc	0.342	0.296	0.196	/	0.187	0.256	0.637
bcw	0.632	0.229	0.298	/	0.414	0.381	0.520
sonar	0.321	0.198	0.315	/	-0.744	-0.080	0.629
soybean	1.000	0.261	0.066	-0.512	0.417	0.087	0.553
hepatitis	0.692	0.347	0.053	/	-0.273	-0.435	0.351
lungcancer	-0.557	0.114	0.021	0.028	0.394	0.198	-0.558
+	<b>60 %</b>	<b>60 %</b>	<b>50 %</b>	<b>20 %</b>	<b>45 %</b>	<b>30 %</b>	<b>75 %</b>
-	<b>10 %</b>	<b>0 %</b>	<b>5 %</b>	<b>5 %</b>	<b>15 %</b>	<b>5 %</b>	<b>0 %</b>

Tabela A.52: Korelacijski koeficienti med oceno  $TRANS_2$ ska in točnostjo klasifikacij.

	$TRANS_2$ bha						
	DT	NB	KNN	NN	RF	BAG	SVM
housevotes	0.285	0.236	0.173	/	0.154	0.065	0.472
wine	0.458	0.256	-0.056	-0.397	-0.454	-0.271	0.449
parkinsons	0.109	-0.006	-0.398	/	-0.719	-0.607	0.056
zoo	0.055	0.509	-0.113	0.412	0.340	-0.025	0.630
tictactoe	0.409	0.176	-0.319	/	-0.802	-0.503	0.245
postoperative	-0.424	-0.240	-0.093	-0.007	-0.127	-0.081	-0.044
monks3	/	-0.404	0.662	/	-0.558	/	-0.407
iriset	0.206	0.472	0.059	0.183	-0.098	-0.105	0.256
glass	0.034	0.095	-0.200	0.015	-0.632	-0.442	0.200
hungarian	0.030	0.019	0.048	/	-0.316	-0.129	0.206
ecoli	0.332	0.054	0.008	0.204	-0.110	-0.089	0.333
heart	0.138	0.137	-0.089	/	-0.521	-0.370	0.067
haberman	0.007	0.114	0.293	/	-0.075	-0.116	0.220
flag	0.140	-0.115	-0.045	-0.516	-0.366	-0.421	0.391
wdbc	0.263	0.173	-0.049	/	-0.229	-0.138	0.356
bcw	0.586	0.281	0.080	/	0.109	0.074	0.230
sonar	0.369	0.024	-0.023	/	-0.810	-0.570	0.461
soybean	0.943	0.183	-0.070	-0.472	-0.415	-0.442	0.258
hepatitis	0.514	0.175	-0.333	/	-0.651	-0.642	0.097
lungcancer	-0.369	0.200	-0.288	0.006	-0.040	0.053	-0.204
+	<b>55 %</b>	<b>40 %</b>	<b>20 %</b>	<b>10 %</b>	<b>10 %</b>	<b>0 %</b>	<b>50 %</b>
-	<b>0 %</b>	<b>5 %</b>	<b>15 %</b>	<b>15 %</b>	<b>55 %</b>	<b>35 %</b>	<b>0 %</b>

Tabela A.53: Korelacijski koeficienti med oceno  $TRANS_2$ bha in točnostjo klasifikacij.

Poglavje A: Podrobni rezultati testiranj

	<i>TRANS</i> <sub>2</sub> har						
	DT	NB	KNN	NN	RF	BAG	SVM
housevotes	0.285	0.236	0.220	/	0.241	0.133	0.476
wine	0.490	0.287	0.029	-0.379	-0.343	-0.141	0.492
parkinsons	0.115	-0.004	-0.268	/	-0.750	-0.559	0.060
zoo	0.208	0.505	-0.014	0.409	0.452	0.070	0.650
tictactoe	0.428	0.176	-0.232	/	-0.811	-0.396	0.246
postoperative	-0.413	-0.218	-0.069	0.010	-0.132	-0.103	-0.090
monks3	/	-0.399	0.678	/	-0.538	/	-0.407
irisset	0.224	0.482	0.141	0.212	0.003	-0.012	0.276
glass	0.028	0.084	-0.045	0.015	-0.623	-0.390	0.206
hungarian	0.042	0.022	0.161	/	-0.260	-0.045	0.206
ecoli	0.374	0.077	0.090	0.217	-0.013	0.014	0.355
heart	0.159	0.145	0.022	/	-0.525	-0.309	0.068
haberman	0.013	0.114	0.333	/	-0.004	-0.028	0.221
flag	0.155	-0.117	0.090	-0.484	-0.279	-0.409	0.395
wdbc	0.284	0.173	0.027	/	-0.147	-0.042	0.363
bcw	0.584	0.286	0.148	/	0.201	0.160	0.244
sonar	0.353	0.053	0.112	/	-0.819	-0.542	0.481
soybean	0.969	0.164	0.004	-0.476	-0.284	-0.337	0.278
hepatitis	0.520	0.163	-0.247	/	-0.652	-0.663	0.102
lungcancer	-0.395	0.151	-0.208	0.028	-0.011	0.039	-0.216
+	<b>60 %</b>	<b>40 %</b>	<b>25 %</b>	<b>10 %</b>	<b>10 %</b>	<b>10 %</b>	<b>50 %</b>
-	<b>0 %</b>	<b>5 %</b>	<b>15 %</b>	<b>15 %</b>	<b>55 %</b>	<b>35 %</b>	<b>0 %</b>

Tabela A.54: Korelacijski koeficienti med oceno *TRANS*<sub>2</sub>har in točnostjo klasifikacij.

	<i>TRANS</i> <sub>2</sub> hel						
	DT	NB	KNN	NN	RF	BAG	SVM
housevotes	0.285	0.236	0.173	/	0.154	0.065	0.472
wine	0.458	0.256	-0.056	-0.397	-0.454	-0.271	0.449
parkinsons	0.109	-0.006	-0.398	/	-0.719	-0.607	0.056
zoo	0.055	0.509	-0.113	0.412	0.340	-0.025	0.630
tictactoe	0.409	0.176	-0.319	/	-0.802	-0.503	0.245
postoperative	-0.424	-0.240	-0.093	-0.007	-0.127	-0.081	-0.044
monks3	/	-0.404	0.662	/	-0.558	/	-0.407
irisset	0.206	0.472	0.059	0.183	-0.098	-0.105	0.256
glass	0.034	0.095	-0.200	0.015	-0.632	-0.442	0.200
hungarian	0.030	0.019	0.048	/	-0.316	-0.129	0.206
ecoli	0.332	0.054	0.008	0.204	-0.110	-0.089	0.333
heart	0.138	0.137	-0.089	/	-0.521	-0.370	0.067
haberman	0.007	0.114	0.293	/	-0.075	-0.116	0.220
flag	0.140	-0.115	-0.045	-0.516	-0.366	-0.421	0.391
wdbc	0.263	0.173	-0.049	/	-0.229	-0.138	0.356
bcw	0.586	0.281	0.080	/	0.109	0.074	0.230
sonar	0.369	0.024	-0.023	/	-0.810	-0.570	0.461
soybean	0.943	0.183	-0.070	-0.472	-0.415	-0.442	0.258
hepatitis	0.514	0.175	-0.333	/	-0.651	-0.642	0.097
lungcancer	-0.369	0.200	-0.288	0.006	-0.040	0.053	-0.204
+	<b>55 %</b>	<b>40 %</b>	<b>20 %</b>	<b>10 %</b>	<b>10 %</b>	<b>0 %</b>	<b>50 %</b>
-	<b>0 %</b>	<b>5 %</b>	<b>15 %</b>	<b>15 %</b>	<b>55 %</b>	<b>35 %</b>	<b>0 %</b>

Tabela A.55: Korelacijski koeficienti med oceno *TRANS*<sub>2</sub>hel in točnostjo klasifikacij.

## A.7 Korelacijski koeficienti ocen $TRANS_2$

	$TRANS_{2jsd}$						
	DT	NB	KNN	NN	RF	BAG	SVM
housevotes	0.285	0.236	0.199	/	0.200	0.103	0.474
wine	0.477	0.269	-0.008	-0.385	-0.409	-0.205	0.466
parkinsons	0.113	-0.005	-0.334	/	-0.747	-0.595	0.057
zoo	0.146	0.508	-0.063	0.410	0.400	0.025	0.638
tictactoe	0.419	0.176	-0.273	/	-0.816	-0.455	0.245
postoperative	-0.417	-0.230	-0.083	0.003	-0.129	-0.095	-0.068
monks3	/	-0.402	0.674	/	-0.554	/	-0.407
irisset	0.213	0.476	0.106	0.200	-0.042	-0.052	0.263
glass	0.028	0.088	-0.118	0.014	-0.635	-0.418	0.202
hungarian	0.037	0.020	0.111	/	-0.295	-0.087	0.206
ecoli	0.356	0.065	0.052	0.209	-0.067	-0.034	0.343
heart	0.150	0.140	-0.030	/	-0.531	-0.343	0.068
haberman	0.009	0.114	0.315	/	-0.043	-0.070	0.220
flag	0.149	-0.117	0.031	-0.500	-0.326	-0.422	0.392
wdbc	0.274	0.171	-0.008	/	-0.190	-0.089	0.359
bcw	0.586	0.283	0.117	/	0.159	0.120	0.235
sonar	0.362	0.039	0.052	/	-0.816	-0.562	0.469
soybean	0.957	0.172	-0.026	-0.474	-0.363	-0.391	0.266
hepatitis	0.517	0.167	-0.288	/	-0.661	-0.663	0.099
lungcancer	-0.383	0.172	-0.242	0.019	-0.025	0.047	-0.208
+	<b>60 %</b>	<b>40 %</b>	<b>20 %</b>	<b>10 %</b>	<b>10 %</b>	<b>10 %</b>	<b>50 %</b>
-	<b>5 %</b>	<b>5 %</b>	<b>15 %</b>	<b>15 %</b>	<b>55 %</b>	<b>35 %</b>	<b>0 %</b>

Tabela A.56: Korelacijski koeficienti med oceno  $TRANS_{2jsd}$  in točnostjo klasifikacij.

	$TRANS_{2sep}$						
	DT	NB	KNN	NN	RF	BAG	SVM
housevotes	0.042	-0.332	0.191	0.009	0.187	0.143	0.176
wine	0.386	-0.349	0.028	-0.484	-0.211	-0.055	-0.604
parkinsons	-0.081	-0.297	-0.280	-0.004	-0.809	-0.552	-0.376
zoo	-0.250	-0.356	-0.040	0.031	0.085	0.035	0.328
tictactoe	-0.209	-0.433	-0.100	0.236	-0.960	-0.279	-0.002
postoperative	-0.439	-0.337	-0.072	-0.001	-0.124	-0.117	-0.172
monks3	/	-0.570	0.659	0.165	-0.700	/	-0.405
irisset	0.210	-0.504	0.073	-0.191	0.051	0.060	-0.633
glass	0.005	0.007	-0.100	-0.240	-0.473	-0.368	-0.144
hungarian	0.032	-0.228	0.210	/	-0.301	-0.019	0.123
ecoli	0.170	-0.493	0.139	0.007	0.089	0.106	-0.202
heart	0.140	-0.125	0.137	/	-0.562	-0.262	-0.128
haberman	0.047	0.118	0.403	/	-0.195	0.042	0.329
flag	0.099	0.100	-0.328	-0.412	-0.361	-0.365	0.261
wdbc	0.361	0.182	0.093	-0.007	-0.047	0.047	-0.453
bcw	0.569	0.218	0.196	-0.239	0.242	0.232	-0.159
sonar	0.350	-0.228	0.090	0.194	-0.892	-0.545	-0.245
soybean	-0.850	-0.997	0.047	0.023	-0.490	-0.272	-0.486
hepatitis	0.306	-0.222	-0.283	-0.103	-0.731	-0.704	0.036
lungcancer	-0.329	0.079	-0.202	-0.058	0.079	0.045	-0.112
+	<b>35 %</b>	<b>5 %</b>	<b>25 %</b>	<b>15 %</b>	<b>5 %</b>	<b>5 %</b>	<b>20 %</b>
-	<b>5 %</b>	<b>50 %</b>	<b>20 %</b>	<b>20 %</b>	<b>55 %</b>	<b>35 %</b>	<b>35 %</b>

Tabela A.57: Korelacijski koeficienti med oceno  $TRANS_{2sep}$  in točnostjo klasifikacij.

## A.8 Korelacijski koeficienti ocen $TRANS_n$

	$TRANS_{n,man}$						
	DT	NB	KNN	NN	RF	BAG	SVM
housevotes	0.288	0.282	0.140	/	0.289	0.124	0.499
wine	0.448	-0.021	0.194	-0.578	-0.258	-0.091	0.424
parkinsons	0.144	0.018	-0.311	/	-0.694	-0.551	0.145
zoo	0.460	0.569	0.198	0.192	0.640	0.291	0.739
tictactoe	0.405	0.379	-0.847	/	-0.756	-0.662	0.287
postoperative	-0.246	-0.248	-0.042	0.096	-0.076	0.061	-0.213
monks3	/	0.252	0.200	/	-0.439	/	-0.422
irisset	0.016	-0.029	0.318	0.166	-0.096	0.030	0.280
glass	0.095	0.062	0.300	-0.094	-0.160	0.078	0.269
hungarian	0.075	0.091	-0.217	/	-0.286	-0.470	0.198
ecoli	0.028	0.276	0.078	0.165	0.327	0.019	0.302
heart	0.189	0.280	-0.412	/	-0.565	-0.503	0.119
haberman	0.071	0.082	-0.079	/	0.079	-0.323	0.307
flag	0.349	-0.081	0.431	-0.477	0.067	0.012	0.347
wdbc	0.308	0.216	-0.433	/	-0.703	-0.591	0.451
bcw	0.537	0.231	0.065	/	0.232	0.134	0.356
sonar	0.302	0.128	-0.208	/	-0.795	-0.624	0.553
soybean	0.965	-0.022	-0.036	0.031	0.520	-0.284	0.143
hepatitis	0.550	0.213	-0.328	/	-0.629	-0.593	0.218
lungcancer	-0.108	-0.140	-0.330	-0.032	-0.372	0.034	-0.154
+	60 %	40 %	30 %	5 %	10 %	5 %	55 %
-	0 %	0 %	35 %	5 %	45 %	40 %	0 %

Tabela A.58: Korelacijski koeficienti med oceno  $TRANS_{n,man}$  in točnostjo klasifikacij.

A.8 Korelacijski koeficienti ocen  $TRANS_n$

	$TRANS_{n, evk}$						
	DT	NB	KNN	NN	RF	BAG	SVM
housevotes	0.288	0.282	0.140	/	0.289	0.124	0.499
wine	0.436	-0.021	0.079	-0.571	-0.541	-0.340	0.406
parkinsons	0.144	0.018	-0.311	/	-0.694	-0.551	0.145
zoo	0.470	0.562	0.091	0.179	0.546	0.213	0.673
tictactoe	0.405	0.379	-0.847	/	-0.756	-0.662	0.287
postoperative	-0.245	-0.273	-0.069	0.095	-0.095	0.039	-0.217
monks3	/	0.252	0.200	/	-0.439	/	-0.422
irisset	0.010	-0.029	0.200	0.149	-0.312	-0.155	0.280
glass	0.075	0.067	0.121	-0.100	-0.531	-0.283	0.258
hungarian	0.075	0.091	-0.217	/	-0.286	-0.470	0.198
ecoli	0.029	0.264	-0.118	0.141	-0.021	-0.336	0.277
heart	0.189	0.280	-0.412	/	-0.565	-0.503	0.119
haberman	0.071	0.082	-0.079	/	0.079	-0.323	0.307
flag	0.327	-0.091	0.320	-0.599	-0.482	-0.406	0.324
wdbc	0.308	0.216	-0.433	/	-0.703	-0.591	0.451
bcw	0.537	0.231	0.065	/	0.232	0.134	0.356
sonar	0.302	0.128	-0.208	/	-0.795	-0.624	0.553
soybean	0.972	-0.022	-0.064	0.031	0.272	-0.447	0.094
hepatitis	0.550	0.213	-0.328	/	-0.629	-0.593	0.218
lungcancer	-0.087	-0.136	-0.367	-0.023	-0.353	-0.012	-0.192
+	<b>60 %</b>	<b>40 %</b>	<b>20 %</b>	<b>5 %</b>	<b>10 %</b>	<b>0 %</b>	<b>50 %</b>
-	<b>5 %</b>	<b>0 %</b>	<b>35 %</b>	<b>10 %</b>	<b>60 %</b>	<b>50 %</b>	<b>0 %</b>

Tabela A.59: Korelacijski koeficienti med oceno  $TRANS_{n, evk}$  in točnostjo klasifikacij.

	$TRANS_{n, max}$						
	DT	NB	KNN	NN	RF	BAG	SVM
housevotes	0.288	0.282	0.140	/	0.289	0.124	0.499
wine	0.448	-0.021	0.194	-0.578	-0.258	-0.091	0.424
parkinsons	0.144	0.018	-0.311	/	-0.694	-0.551	0.145
zoo	0.549	0.569	0.198	0.192	0.635	0.278	0.704
tictactoe	0.405	0.379	-0.847	/	-0.756	-0.662	0.287
postoperative	-0.246	-0.248	-0.042	0.096	-0.076	0.061	-0.213
monks3	/	0.252	0.200	/	-0.439	/	-0.422
irisset	0.016	-0.029	0.318	0.166	-0.096	0.030	0.280
glass	0.079	0.062	0.300	-0.109	-0.165	0.067	0.269
hungarian	0.075	0.091	-0.217	/	-0.286	-0.470	0.198
ecoli	0.018	0.276	0.078	0.152	0.320	0.004	0.278
heart	0.189	0.280	-0.412	/	-0.565	-0.503	0.119
haberman	0.071	0.082	-0.079	/	0.079	-0.323	0.307
flag	0.324	-0.081	0.431	-0.563	0.069	-0.022	0.319
wdbc	0.308	0.216	-0.433	/	-0.703	-0.591	0.451
bcw	0.537	0.231	0.065	/	0.232	0.134	0.356
sonar	0.302	0.128	-0.208	/	-0.795	-0.624	0.553
soybean	0.965	-0.022	-0.036	0.031	0.510	-0.284	0.149
hepatitis	0.550	0.213	-0.328	/	-0.629	-0.593	0.218
lungcancer	-0.108	-0.140	-0.330	-0.032	-0.372	0.034	-0.154
+	<b>60 %</b>	<b>40 %</b>	<b>30 %</b>	<b>5 %</b>	<b>10 %</b>	<b>5 %</b>	<b>45 %</b>
-	<b>0 %</b>	<b>0 %</b>	<b>35 %</b>	<b>5 %</b>	<b>45 %</b>	<b>40 %</b>	<b>0 %</b>

Tabela A.60: Korelacijski koeficienti med oceno  $TRANS_{n, max}$  in točnostjo klasifikacij.

Poglavje A: Podrobni rezultati testiranj

	$TRANS_n\cos$						
	DT	NB	KNN	NN	RF	BAG	SVM
housevotes	0.285	0.175	0.133	/	0.319	0.107	0.464
wine	0.417	-0.019	0.243	-0.557	0.144	0.208	0.317
parkinsons	0.114	-0.005	-0.146	/	-0.612	-0.427	0.109
zoo	0.341	0.521	0.294	0.236	0.819	0.380	0.620
tictactoe	0.329	0.392	-0.846	/	-0.554	-0.495	0.281
postoperative	-0.238	-0.249	-0.031	0.074	-0.102	0.044	-0.181
monks3	/	0.606	0.200	/	-0.157	/	-0.429
irisset	-0.017	0.040	0.393	0.155	0.160	0.161	0.249
glass	0.060	0.047	0.443	0.017	0.331	0.376	0.284
hungarian	0.050	0.036	-0.082	/	-0.084	-0.395	0.220
ecoli	0.112	0.261	0.263	0.143	0.562	0.362	0.187
heart	0.172	0.147	-0.339	/	-0.504	-0.442	0.139
haberman	0.031	0.123	0.034	/	0.215	-0.248	0.218
flag	0.319	-0.086	0.514	-0.490	0.593	0.441	0.304
wdbc	0.294	0.165	-0.348	/	-0.597	-0.463	0.384
bcw	0.527	0.118	0.067	/	0.246	0.121	0.290
sonar	0.290	0.073	-0.114	/	-0.789	-0.592	0.510
soybean	0.885	-0.022	-0.011	0.031	0.697	-0.036	0.258
hepatitis	0.506	0.156	-0.239	/	-0.542	-0.516	0.238
lungcancer	-0.015	-0.148	-0.276	-0.031	-0.329	-0.039	-0.324
+	45 %	40 %	35 %	5 %	35 %	25 %	60 %
-	5 %	0 %	20 %	10 %	35 %	40 %	5 %

Tabela A.61: Korelacijski koeficienti med oceno  $TRANS_n\cos$  in točnostjo klasifikacij.

	$TRANS_n\cot$						
	DT	NB	KNN	NN	RF	BAG	SVM
housevotes	0.288	0.254	0.164	/	0.359	0.156	0.503
wine	0.453	-0.021	0.238	-0.525	0.143	0.200	0.451
parkinsons	0.146	0.015	-0.138	/	-0.618	-0.438	0.183
zoo	0.493	0.561	0.308	0.197	0.776	0.358	0.765
tictactoe	0.416	0.460	-0.841	/	-0.567	-0.515	0.315
postoperative	-0.249	-0.262	0.001	0.100	-0.072	0.092	-0.198
monks3	/	0.539	0.200	/	-0.189	/	-0.428
irisset	0.021	-0.006	0.365	0.190	0.159	0.157	0.319
glass	0.097	0.055	0.449	-0.066	0.318	0.372	0.312
hungarian	0.077	0.096	-0.088	/	-0.095	-0.397	0.209
ecoli	0.092	0.277	0.265	0.183	0.555	0.355	0.314
heart	0.195	0.284	-0.334	/	-0.498	-0.443	0.153
haberman	0.075	0.107	0.042	/	0.193	-0.228	0.303
flag	0.366	-0.079	0.524	-0.443	0.591	0.442	0.366
wdbc	0.308	0.209	-0.368	/	-0.618	-0.486	0.471
bcw	0.530	0.199	0.109	/	0.283	0.177	0.374
sonar	0.290	0.134	-0.098	/	-0.762	-0.575	0.575
soybean	0.969	-0.022	-0.005	0.031	0.653	-0.038	0.237
hepatitis	0.552	0.209	-0.243	/	-0.548	-0.518	0.250
lungcancer	-0.084	-0.148	-0.275	-0.034	-0.319	-0.039	-0.299
+	50 %	45 %	35 %	10 %	35 %	25 %	60 %
-	5 %	0 %	20 %	10 %	35 %	40 %	5 %

Tabela A.62: Korelacijski koeficienti med oceno  $TRANS_n\cot$  in točnostjo klasifikacij.

A.8 Korelacijski koeficienti ocen  $TRANS_n$

		$TRANS_{n\text{ska}}$					
	DT	NB	KNN	NN	RF	BAG	SVM
housevotes	0.374	0.313	0.199	/	0.484	0.243	0.514
wine	0.519	0.608	0.317	-0.066	0.631	0.548	0.743
parkinsons	0.346	0.115	0.072	/	-0.439	-0.221	0.463
zoo	0.934	0.635	0.468	0.192	0.857	0.450	0.858
tictactoe	0.677	0.603	-0.846	/	0.198	-0.227	0.463
postoperative	-0.635	-0.239	0.074	0.134	-0.101	0.165	0.237
monks3	/	0.741	0.200	/	0.684	/	0.408
irisset	0.272	0.599	0.520	0.250	0.458	0.324	0.877
glass	0.244	0.196	0.608	0.213	0.640	0.567	0.584
hungarian	0.316	0.405	0.143	/	0.332	-0.250	0.348
ecoli	0.317	0.481	0.495	0.309	0.682	0.609	0.695
heart	0.192	0.431	-0.239	/	-0.329	-0.319	0.299
haberman	0.125	0.071	0.272	/	0.349	-0.057	0.328
flag	0.564	-0.080	0.573	0.044	0.769	0.668	0.542
wdbc	0.342	0.296	-0.243	/	-0.380	-0.268	0.637
bcw	0.589	0.229	0.144	/	0.348	0.252	0.519
sonar	0.280	0.198	-0.020	/	-0.713	-0.497	0.627
soybean	0.974	0.043	0.041	0.031	0.801	0.384	0.562
hepatitis	0.692	0.347	-0.112	/	-0.298	-0.355	0.357
lungcancer	-0.174	-0.153	-0.156	-0.032	-0.307	-0.092	-0.588
+	55 %	60 %	55 %	10 %	55 %	35 %	75 %
-	5 %	0 %	15 %	0 %	25 %	30 %	0 %

Tabela A.63: Korelacijski koeficienti med oceno  $TRANS_{n\text{ska}}$  in točnostjo klasifikacij.

		$TRANS_{n\text{bha}}$					
	DT	NB	KNN	NN	RF	BAG	SVM
housevotes	0.285	0.236	0.044	/	0.004	-0.055	0.469
wine	0.376	-0.020	0.202	-0.621	-0.423	-0.287	0.363
parkinsons	0.109	-0.006	-0.578	/	-0.712	-0.652	0.055
zoo	0.394	0.563	0.152	0.230	0.621	0.310	0.494
tictactoe	0.283	0.176	-0.829	/	-0.807	-0.752	0.235
postoperative	-0.162	-0.269	-0.087	0.059	-0.085	-0.054	-0.178
monks3	/	-0.404	0.197	/	-0.617	/	-0.407
irisset	-0.003	-0.029	0.386	0.127	-0.090	0.039	0.245
glass	0.086	0.065	0.298	-0.018	-0.260	0.130	0.238
hungarian	0.035	0.019	-0.408	/	-0.458	-0.548	0.206
ecoli	0.088	0.258	0.106	0.130	0.224	0.101	0.193
heart	0.138	0.137	-0.530	/	-0.598	-0.567	0.080
haberman	0.019	0.114	-0.251	/	-0.198	-0.475	0.227
flag	0.318	-0.091	0.396	-0.544	-0.005	0.021	0.287
wdbc	0.263	0.173	-0.540	/	-0.687	-0.674	0.355
bcw	0.508	0.281	-0.090	/	-0.015	-0.089	0.231
sonar	0.325	0.024	-0.433	/	-0.822	-0.673	0.446
soybean	0.879	-0.022	-0.042	0.033	0.241	-0.221	0.091
hepatitis	0.514	0.175	-0.446	/	-0.665	-0.630	0.167
lungcancer	0.021	-0.099	-0.456	-0.014	-0.475	-0.136	-0.277
+	55 %	35 %	25 %	5 %	5 %	5 %	40 %
-	0 %	5 %	40 %	5 %	45 %	40 %	0 %

Tabela A.64: Korelacijski koeficienti med oceno  $TRANS_{n\text{bha}}$  in točnostjo klasifikacij.

Poglavje A: Podrobni rezultati testiranj

	$TRANS_n$ har						
	DT	NB	KNN	NN	RF	BAG	SVM
housevotes	0.285	0.236	0.091	/	0.126	0.026	0.474
wine	0.404	-0.021	0.199	-0.605	-0.334	-0.134	0.375
parkinsons	0.115	-0.004	-0.449	/	-0.726	-0.615	0.058
zoo	0.396	0.558	0.165	0.224	0.674	0.309	0.570
tictactoe	0.305	0.176	-0.848	/	-0.826	-0.733	0.236
postoperative	-0.174	-0.256	-0.075	0.070	-0.098	-0.014	-0.179
monks3	/	-0.399	0.200	/	-0.609	/	-0.407
iriset	-0.002	-0.029	0.372	0.143	-0.086	0.033	0.247
glass	0.079	0.061	0.298	-0.025	-0.164	0.132	0.247
hungarian	0.045	0.022	-0.304	/	-0.419	-0.518	0.206
ecoli	0.082	0.263	0.095	0.132	0.342	0.084	0.206
heart	0.159	0.145	-0.467	/	-0.606	-0.544	0.081
haberman	0.020	0.114	-0.169	/	-0.078	-0.408	0.229
flag	0.332	-0.088	0.403	-0.517	0.120	0.070	0.295
wdbc	0.284	0.173	-0.461	/	-0.725	-0.644	0.362
bcw	0.530	0.286	-0.021	/	0.108	0.014	0.244
sonar	0.315	0.053	-0.304	/	-0.827	-0.668	0.466
soybean	0.904	-0.022	-0.043	0.031	0.518	-0.252	0.124
hepatitis	0.520	0.163	-0.378	/	-0.666	-0.636	0.170
lungcancer	-0.031	-0.130	-0.351	-0.027	-0.411	-0.048	-0.267
+	55 %	35 %	25 %	5 %	5 %	5 %	40 %
-	0 %	5 %	40 %	5 %	45 %	40 %	0 %

Tabela A.65: Korelacijski koeficienti med oceno  $TRANS_n$ har in točnostjo klasifikacij.

	$TRANS_n$ hel						
	DT	NB	KNN	NN	RF	BAG	SVM
housevotes	0.285	0.236	0.044	/	0.004	-0.055	0.469
wine	0.376	-0.020	0.202	-0.621	-0.423	-0.287	0.363
parkinsons	0.109	-0.006	-0.578	/	-0.712	-0.652	0.055
zoo	0.394	0.563	0.152	0.230	0.621	0.310	0.494
tictactoe	0.283	0.176	-0.829	/	-0.807	-0.752	0.235
postoperative	-0.162	-0.269	-0.087	0.059	-0.085	-0.054	-0.178
monks3	/	-0.404	0.197	/	-0.617	/	-0.407
iriset	-0.003	-0.029	0.386	0.127	-0.090	0.039	0.245
glass	0.086	0.065	0.298	-0.018	-0.260	0.130	0.238
hungarian	0.035	0.019	-0.408	/	-0.458	-0.548	0.206
ecoli	0.088	0.258	0.106	0.130	0.224	0.101	0.193
heart	0.138	0.137	-0.530	/	-0.598	-0.567	0.080
haberman	0.019	0.114	-0.251	/	-0.198	-0.475	0.227
flag	0.318	-0.091	0.396	-0.544	-0.005	0.021	0.287
wdbc	0.263	0.173	-0.540	/	-0.687	-0.674	0.355
bcw	0.508	0.281	-0.090	/	-0.015	-0.089	0.231
sonar	0.325	0.024	-0.433	/	-0.822	-0.673	0.446
soybean	0.879	-0.022	-0.042	0.033	0.241	-0.221	0.091
hepatitis	0.514	0.175	-0.446	/	-0.665	-0.630	0.167
lungcancer	0.021	-0.099	-0.456	-0.014	-0.475	-0.136	-0.277
+	55 %	35 %	25 %	5 %	5 %	5 %	40 %
-	0 %	5 %	40 %	5 %	45 %	40 %	0 %

Tabela A.66: Korelacijski koeficienti med oceno  $TRANS_n$ hel in točnostjo klasifikacij.

A.8 Korelacijski koeficienti ocen  $TRANS_n$

	$TRANS_{n,jsd}$						
	DT	NB	KNN	NN	RF	BAG	SVM
housevotes	0.285	0.236	0.072	/	0.069	-0.010	0.471
wine	0.389	-0.020	0.201	-0.611	-0.397	-0.208	0.369
parkinsons	0.113	-0.005	-0.518	/	-0.731	-0.644	0.056
zoo	0.395	0.560	0.157	0.227	0.657	0.310	0.527
tictactoe	0.292	0.176	-0.844	/	-0.827	-0.758	0.236
postoperative	-0.167	-0.264	-0.083	0.064	-0.102	-0.034	-0.179
monks3	/	-0.402	0.200	/	-0.620	/	-0.407
iriset	-0.003	-0.029	0.384	0.136	-0.089	0.036	0.246
glass	0.082	0.062	0.298	-0.021	-0.208	0.138	0.242
hungarian	0.041	0.020	-0.355	/	-0.448	-0.540	0.206
ecoli	0.086	0.261	0.101	0.131	0.309	0.097	0.198
heart	0.150	0.140	-0.501	/	-0.611	-0.561	0.080
haberman	0.018	0.114	-0.209	/	-0.138	-0.444	0.228
flag	0.326	-0.089	0.397	-0.530	0.079	0.059	0.290
wdbc	0.274	0.171	-0.498	/	-0.721	-0.669	0.358
bcw	0.524	0.283	-0.055	/	0.052	-0.034	0.236
sonar	0.322	0.039	-0.368	/	-0.826	-0.677	0.454
soybean	0.891	-0.022	-0.043	0.031	0.442	-0.242	0.104
hepatitis	0.517	0.167	-0.411	/	-0.674	-0.644	0.168
lungcancer	-0.006	-0.118	-0.395	-0.022	-0.446	-0.092	-0.274
+	55 %	35 %	25 %	5 %	5 %	10 %	40 %
-	0 %	5 %	40 %	5 %	45 %	40 %	0 %

Tabela A.67: Korelacijski koeficienti med oceno  $TRANS_{n,jsd}$  in točnostjo klasifikacij.

	$TRANS_{n,sep}$						
	DT	NB	KNN	NN	RF	BAG	SVM
housevotes	0.042	-0.332	-0.033	/	-0.052	-0.060	0.214
wine	0.058	0.132	0.113	-0.561	-0.496	-0.426	-0.024
parkinsons	-0.081	-0.297	-0.842	/	-0.802	-0.739	-0.395
zoo	0.569	0.037	0.151	-0.349	0.209	0.128	0.503
tictactoe	-0.612	-0.433	-0.847	/	-0.959	-0.898	-0.007
postoperative	-0.119	-0.250	-0.093	0.038	-0.164	-0.187	-0.033
monks3	/	-0.570	0.200	/	-0.887	/	-0.405
iriset	0.120	-0.263	0.223	-0.383	-0.034	0.086	-0.167
glass	0.127	0.110	0.335	-0.294	-0.242	0.035	-0.035
hungarian	0.082	-0.228	-0.549	/	-0.635	-0.602	0.118
ecoli	0.177	-0.282	0.069	0.033	-0.143	-0.063	-0.217
heart	0.140	-0.125	-0.559	/	-0.658	-0.628	-0.082
haberman	0.079	0.118	-0.385	/	-0.494	-0.489	0.329
flag	0.369	-0.139	0.428	-0.391	-0.104	-0.118	-0.002
wdbc	0.361	0.182	-0.729	/	-0.819	-0.796	-0.445
bcw	0.435	0.218	-0.096	/	0.021	-0.010	-0.154
sonar	0.307	-0.228	-0.731	/	-0.888	-0.754	-0.300
soybean	0.940	0.035	0.111	0.028	-0.497	0.019	-0.355
hepatitis	0.306	-0.222	-0.576	/	-0.739	-0.697	0.059
lungcancer	-0.238	0.099	-0.322	-0.142	-0.423	-0.104	-0.122
+	40 %	5 %	20 %	0 %	5 %	0 %	15 %
-	5 %	30 %	45 %	15 %	50 %	40 %	30 %

Tabela A.68: Korelacijski koeficienti med oceno  $TRANS_{n,sep}$  in točnostjo klasifikacij.

## A.9 Korelacijski koeficienti referenčnih ocen zanesljivosti

	klasifikatorjeva zanesljivost						
	DT	NB	KNN	NN	RF	BAG	SVM
housevotes	0.354	0.278	0.551	1.000	0.711	0.611	0.529
wine	0.315	0.598	0.560	0.669	0.945	0.807	0.938
parkinsons	0.484	0.123	0.840	1.000	0.827	0.750	0.553
zoo	0.882	0.626	0.962	0.329	0.994	0.865	0.984
tictactoe	0.776	0.559	0.847	1.000	0.976	0.902	0.425
postoperative	-0.627	0.223	0.212	0.014	0.072	0.289	0.392
monks3	/	0.767	-0.200	1.000	1.000	/	0.406
irisset	0.431	0.581	0.610	0.678	0.712	0.617	0.901
glass	0.325	0.153	0.609	0.555	0.732	0.669	0.585
hungarian	0.392	0.399	0.549	1.000	0.606	0.597	0.457
ecoli	0.430	0.543	0.627	0.554	0.739	0.721	0.703
heart	0.211	0.407	0.562	1.000	0.655	0.641	0.504
haberman	0.254	0.120	0.383	1.000	0.448	0.466	0.286
flag	0.517	0.068	0.454	0.584	0.850	0.699	0.595
wdbc	0.321	0.247	0.730	1.000	0.811	0.817	0.762
bcw	0.327	0.200	0.621	1.000	0.736	0.741	0.567
sonar	0.002	0.206	0.733	1.000	0.905	0.725	0.709
soybean	0.850	1.000	1.000	1.000	1.000	1.000	0.890
hepatitis	0.523	0.365	0.564	1.000	0.701	0.678	0.586
lungcancer	-0.052	-0.074	0.290	0.027	0.211	0.102	-0.545
+	<b>60 %</b>	<b>60 %</b>	<b>80 %</b>	<b>60 %</b>	<b>80 %</b>	<b>80 %</b>	<b>85 %</b>
-	<b>5 %</b>	<b>0 %</b>	<b>5 %</b>	<b>0 %</b>	<b>0 %</b>	<b>0 %</b>	<b>0 %</b>

Tabela A.69: Korelacijski koeficienti med referenčnimi klasifikatorjevimi ocenami zanesljivosti in točnostjo klasifikacij.