

SVEUČILIŠTE U ZAGREBU
FAKULTET ELEKTROTEHNIKE I RAČUNARSTVA

DIPLOMSKI RAD br. 2101

**IMPLEMENTACIJA UČINKOVITOG NEIZRAZITOG
KLASIFIKACIJSKOG SUSTAVA TEMELJENOGA NA
ALGORITMU IVTURS**

Romano Barilar

Zagreb, lipanj 2020.

SVEUČILIŠTE U ZAGREBU
FAKULTET ELEKTROTEHNIKE I RAČUNARSTVA

DIPLOMSKI RAD br. 2101

**IMPLEMENTACIJA UČINKOVITOG NEIZRAZITOG
KLASIFIKACIJSKOG SUSTAVA TEMELJENOGA NA
ALGORITMU IVTURS**

Romano Barilar

Zagreb, lipanj 2020.

DIPLOMSKI ZADATAK br. 2101

Pristupnik: **Romano Barilar (0036496368)**

Studij: Računarstvo

Profil: Računarska znanost

Mentor: izv. prof. dr. sc. Alan Jović

Zadatak: **Implementacija učinkovitog neizrazitog klasifikacijskog sustava temeljenoga na algoritmu IVTURS**

Opis zadatka:

Neizraziti klasifikacijski sustavi koriste neizrazita pravila kako bi precizno klasificirali podatke u ciljne klase. Oni se sastoje od baze podataka u koju se spremaju sva neizrazita pravila te od metode za rasuđivanje. Algoritam IVTURS (od engl. linguistic fuzzy rule-based classification system based on Interval-Valued fuzzy reasoning method with TUning and Rule Selection) jedan je od najuspješnijih primjera neizrazitih klasifikacijskih sustava koji koristi intervalnu neizrazitu logiku, gdje intervali dodaju još jednu dimenziju neodređenosti nad neizrazitom logikom. U diplomskom radu potrebno je dati osvrt na metode koje se koriste u ovom području te ostvariti vlastitu implementaciju algoritma IVTURS u programskom jeziku po izboru. Nakon implementacije, potrebno je poraditi na optimizaciji algoritma IVTURS kako bi se postigla što brža izgradnja modela sa što točnijim rezultatima. Pritom se mogu ponuditi i izmjene dijelova postojećeg algoritma koje treba zasnovati na postojećoj srodnoj literaturi iz područja koje koristi nesavršeno znanje. Učinkovitost implementacije algoritma potrebno je ispitati na većem broju skupova podataka koje su koristili i izvorni autori algoritma. Vrednovanje rezultata treba provesti primjerenim statističkim testovima. U implementacijskom smislu, osim samog algoritma, potrebno je izraditi i programsko rješenje koje omogućuje učitavanje skupova podataka, izbor parametara algoritma te prikaz rezultata vrednovanja.

Rok za predaju rada: 30. lipnja 2020.

Sadržaj

Uvod.....	1
1 Neizraziti klasifikacijski sustavi.....	2
1.1 Neizrazitost i intervali	2
1.2 Intervalno neizraziti skupovi i dodaci	2
1.3 Neizraziti i intervalno neizraziti klasifikacijski sustavi bazirani na pravilima....	4
1.4 Metoda rasuđivanja neizrazitih klasifikacijskih sustava	5
2 IVTURS	7
2.1 IVTURS-ova metoda rasuđivanja.....	7
2.2 Inicijalizacija sustava	11
2.3 Odabir pravila	14
2.4 Intervalno neizraziti klasifikacijski sustav.....	16
2.5 Optimizacija sustava	18
2.6 Klasifikacija primjera.....	22
3 Promjene uvedene u algoritam IVTURS	24
3.1 Choquetovi integrali.....	24
4 Ispitivanje rada klasifikatora	28
4.1 Friedmanov test	30
4.2 Holmov test	31
4.3 Test Wilcoxonovih rangova	32
4.4 Postavke klasifikatora <i>IVTURS</i>	33
4.5 Usporedba s <i>FURIA</i> -om, <i>C4.5</i> , radom o <i>IVTURS</i> -u i <i>KEEL</i> -om	34
4.6 Sustav u ovisnosti o hiperparametru k	36
4.7 Choquetovi integrali.....	38
4.8 Usporedba s klasifikatorom <i>FOATFRB</i>	39

4.9	Usporedba s klasifikatorom <i>BPMOGA</i>	40
4.10	Usporedba s klasifikatorom <i>BTSKFC</i>	41
	Zaključak	43
	Literatura	44
	Naslov, sažetak i ključne riječi.....	47
	Title, abstract and keywords.....	48
	Privitak	49
	Format ulaznih podataka.....	49
	GUI.....	50

Uvod

Klasifikacija podataka jedan je od mnogih problema strojnog učenja, pa sukladno s tim postoje različiti pristupi rješavanju spomenutog problema. Za potrebe ovog rada korišten je klasifikacijski sustav utemeljen na pravilima. Takvi sustavi sastoje se od baze podataka i metode rasuđivanja. Metoda rasuđivanja način je na koji sustav koristi bazu podataka za klasificiranje primjera.

IVTURS je intervalno neizraziti klasifikacijski sustav definiran u radu pod nazivom *IVTURS: A Linguistic Fuzzy Rule-Based Classification System Based On a New Interval-Valued Fuzzy Reasoning Method With Tuning and Rule Selection* [1].

Koristeći neizrazita klasifikacijska pravila i potpuno novu metodu rasuđivanja, *IVTURS* ostvaruje *state-of-the-art* rezultate na raznim skupovima podataka. Pravila se stvaraju u tri faze, pri čemu je posljednja, ona koja koristi evolucijski algoritam *CHC*, najvažnija. Prve dvije faze pronalaze inicijalna pravila dok treća bira i optimira najvažnija pravila koja će se i koristiti. Algoritam je objavljen u radu [1], a nastao je na temelju rada [2].

Ovaj diplomski rad analizira i proučava segmente rada algoritma *IVTURS*. Za rad je napisana implementacija *IVTURS*-a u programskim jezicima python/cython, koja se u daljnjem tekstu naziva reimplementacijom *IVTURS*-a. Osim reimplementacije, uvedene su i promjene, koristeći Choquetove integrale s kojima su definirane dvije nove verzije algoritma. Prva verzija koristi funkcije *IVREF* i Choquetove integrale, dok druga koristi samo Choquetove integrale. Također se u radu analizira duljina i specifičnost pravila ovisno o parametru dubine stabla pretraživanja.

Rezultati reimplementacije usporedit će se s rezultatima iz originalnog rada [1], implementacije *IVTURS*-a iz *KEEL*-a [5], implementacije *FURIA*-e [7] iz *KEEL*-a, implementacije C4.5 [18] iz *KEEL*-a te novijih algoritama *BPMOGA* [19], *FOATFRB* [20] i *BTSKFC* [21]. Testovi koji će se koristiti za usporedbu rezultata su Friedmanov [14], Holmov [16] i Wilcoxonov test [17].

1 Neizraziti klasifikacijski sustavi

1.1 Neizrazitost i intervali

Neizrazita logika (eng. *fuzzy logic*) je prvi put opisana i definirana u radu *Fuzzy sets* koji je 1965. objavljen u časopisu *Information and control*. Njegov autor i otac neizrazite logike je matematičar L. A. Zadeh [9][10]. U klasičnoj se logici pripadnost elemenata opisuje diskretnim vrijednostima 1 ili 0, ovisno o tome pripada li element određenom skupu ili ne. Zadeh je pripadnost neizrazitim skupovima definirao kao vrijednost iz zatvorenog intervala $[0, 1]$. Takva definicija dopušta neodređenost u definiranju skupova i funkcija pripadnosti. Koristeći kontinuirane vrijednosti lakše je opisati stvarni svijet i ljudsko razmišljanje. Jedan od problema klasičnih skupova je potreba za definicijom tvrdih granica između skupova [9]. Problem tvrdih granica vrlo je jednostavno riješiti neizrazitom logikom u kojoj se neodređenost granica lako definira pomoću raznih oblika funkcija pripadnosti. Intervalno neizrazita logika je proširenje neizrazite logike. Njena pripadnost skupovima nije više vrijednost već podinterval iz zatvorenog intervala $[0, 1]$. Korištenjem podintervala se umjesto jedne vrijednosti u neizrazitu logiku uvodi još jedna razina nesigurnosti. Činjenica da je teško odrediti pripada li element nekom skupu može podrazumijevati da je vrlo teško i odrediti točnu vrijednost kojom element pripada skupu. Korištenjem intervala taj problem se olakšava jer izbor više nije toliko precizan.

1.2 Intervalno neizraziti skupovi i dodaci

Neizraziti skup (neizraziti skup prvog tipa) A je skup uređenih parova definiran nad univerzalnim skupom X prema izrazu (1) [9][10]:

$$A = \{(x, \mu_A(x)) | x \in X\} \quad (1)$$

gdje je $\mu_A(x)$ funkcija pripadnosti (eng. *membership function*) određena kao $\mu_A(x): X \rightarrow [0, 1]$.

Neka je $L([0,1])$ skup svih zatvorenih podintervala iz $[0,1]$. Intervalno neizraziti skup (neizraziti skup drugog tipa) A je mapiranje definirano nad univerzalnim skupom X (2) [1][10]:

$$A = \{(x, A(x)) | x \in X\} \quad (2)$$

gdje je $A(x) = [\underline{A}(x), \overline{A}(x)] \in L([0,1])$ a $\underline{A}(x): X \rightarrow [0, 1]$ i $\overline{A}(x): X \rightarrow [0, 1]$ su preslikavanja koja se nazivaju donja odnosno gornja granica pouzdanosti.

Intervalna aritmetika koju *IVTURS* koristi definirana je na sljedeći način [1]:

1. zbrajanje – $[\underline{x}, \overline{x}] + [\underline{y}, \overline{y}] = [\underline{x} + \underline{y}, \overline{x} + \overline{y}]$
2. oduzimanje – $[\underline{x}, \overline{x}] - [\underline{y}, \overline{y}] = [\underline{y} - \overline{x}, \overline{y} - \underline{x}]$
3. množenje – $[\underline{x}, \overline{x}] \cdot [\underline{y}, \overline{y}] = [\underline{x} \cdot \underline{y}, \overline{x} \cdot \overline{y}]$
4. dijeljenje – $\frac{[\underline{x}, \overline{x}]}{[\underline{y}, \overline{y}]} = \left[\min\left(\min\left(\frac{\underline{x}}{\underline{y}}, \frac{\overline{x}}{\overline{y}}\right), 1\right), \min\left(\max\left(\frac{\underline{x}}{\underline{y}}, \frac{\overline{x}}{\overline{y}}\right), 1\right) \right]$
5. množenje s jediničnim intervalom - $[\underline{x}, \overline{x}] \cdot [y] = [\underline{x} \cdot y, \overline{x} \cdot y]$

Funkcija $T: (L([0,1]))^2 \rightarrow L([0,1])$ je intervalna t-norma ako zadovoljava sljedeća svojstva [1][11]:

1. komutativnost – $T(x, y) = T(y, x)$
2. asocijativnost – $T(x, T(y, z)) = T(y, T(x, z))$
3. monotonost – ako je $\underline{x} \leq \underline{q}, \overline{x} \leq \overline{q}, \underline{y} \leq \underline{r}, \overline{y} \leq \overline{r}$ onda je $T(x, y) \leq T(q, r)$
4. neutralni element – $T(x, [1, 1]) = x$

gdje su $x = [\underline{x}, \overline{x}], y = [\underline{y}, \overline{y}], q = [\underline{q}, \overline{q}], r = [\underline{r}, \overline{r}] \in L([0,1])$ a $[1, 1]$ je neutralni element.

Funkcija $S: (L([0,1]))^2 \rightarrow L([0,1])$ je intervalna s-norma ako zadovoljava sljedeća svojstva [11]:

1. komutativnost – $S(x, y) = S(y, x)$
2. asocijativnost – $S(x, S(y, z)) = S(y, S(x, z))$
3. monotonost – ako je $\underline{x} \leq \underline{q}, \overline{x} \leq \overline{q}, \underline{y} \leq \underline{r}, \overline{y} \leq \overline{r}$ onda je $S(x, y) \leq S(q, r)$
4. neutralni element – $S(x, [0, 0]) = x$

gdje su $x = [\underline{x}, \overline{x}], y = [\underline{y}, \overline{y}], q = [\underline{q}, \overline{q}], r = [\underline{r}, \overline{r}] \in L([0,1])$ a $[0, 0]$ je neutralni element.

Jezične varijable su varijable koje uzimaju riječi iz prirodnog jezika kao svoje vrijednosti [9][10]. Kod neizrazite se logike jezične varijable vezuju uz pojedine skupove definirane nad istom domenom. Sastoje se od uređene petorke (X, T, U, G, M) [9][10]:

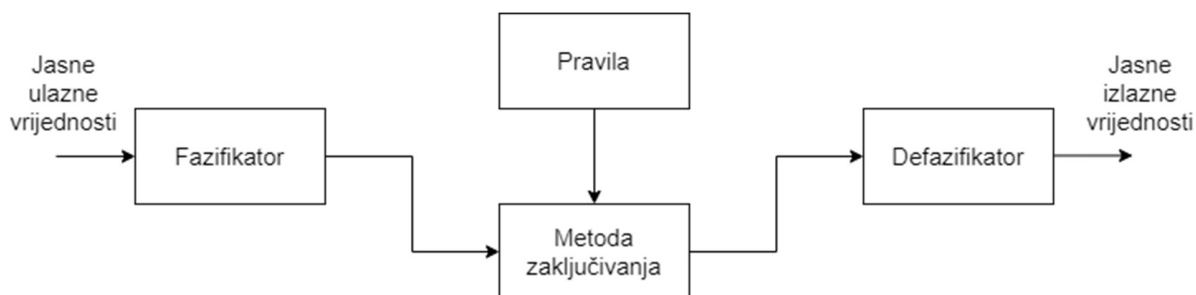
1. X – naziv jezične varijable
2. T – skup vrijednosti (riječi) koje varijabla može poprimiti (eng. *linguistic terms*)
3. U – domena nad kojom su definirani skupovi koji se povezuju s vrijednostima T (eng. *universal set*)
4. G – sintaksna pravila za generiranje jezičnih vrijednosti (eng. *syntactic rule*)
5. M – semantička funkcija koja povezuje jezične vrijednosti sa skupovima definiranim nad U (eng. *semantic rule*)

1.3 Neizraziti i intervalno neizraziti klasifikacijski sustavi bazirani na pravilima

U ovom radu obrađuje se intervalno neizraziti klasifikacijski sustavi. Neizraziti klasifikacijski sustavi i intervalno neizraziti klasifikacijski sustavi funkcioniraju na jednak način i sastoje se od načelno jednakih dijelova. Najveća je razlika u matematičkim izračunima pri čemu jedni koriste neizrazite, a drugi intervalno neizrazite skupove. Stoga su pojedini dijelovi različiti i osposobljeni za bolji rad s neizrazitim vrijednostima tj. neizrazitim intervalima. Neizraziti klasifikacijski sustavi nazivaju se sustavima prvog tipa, a intervalno neizraziti sustavi sustavima drugog tipa.

Općeniti neizraziti klasifikacijski sustav utemeljen na pravilima prikazan je na slici 1.1. Pravila su glavni dio ovakvih sustava [10]. Ulazne vrijednosti su jasne (eng. *crisp*) tj. realne vrijednosti. Kako sustav zaključuje s neizrazitim pravilima, antecedenti pravila koriste neizrazite vrijednosti. Ulazne je vrijednosti stoga najprije potrebno fazificirati (eng. *fuzzify*). Ulazni podaci opisuju se jezičnim varijablama. Svaka jezična varijabla odgovara jednom atributu iz ulaznog prostora. U varijabli se nalaze interval vrijednosti atributa i podjela tog intervala po skupovima. Fazifikator (eng. *fuzzifier*) koristi jezične varijable za fazificiranje ulaznih podataka. Neizraziti ulazni podaci kombiniraju se s

pravilima u aktivacije antecedent dijelova pravila. Potom sustav pomoću metode rasuđivanja uz aktivacije donosi zaključak. Zaključak ne mora biti jasna vrijednost, stoga se po potrebi može koristiti defazifikator (eng. *defuzzifier*). Sama metoda rasuđivanja može biti definirana na mnoge načine. U sljedećem poglavlju prikazana je ona koju autori *IVTURS*-a koriste kao temelj za svoju.



Slika 1.1 Prikaz neizrazitog klasifikacijskog sustava baziranog na pravilima[10]

Neizraziti klasifikacijski sustavi bazirani na pravilima već su se pokazali kao dobro rješenje za klasifikacijske probleme [1][2][3]. Neizrazita i intervalno neizrazita logika područje je koje budi veliki interes svjetskih znanstvenika [10].

1.4 Metoda rasuđivanja neizrazitih klasifikacijskih sustava

Općenita metoda rasuđivanja koju autori *IVTURS*-a nasljeđuju može se podijeliti u četiri koraka [1][3]. $x = (x_1, \dots, x_n)$ primjer je koji se klasificira, M je broj klasa, L broj pravila, RW_j težina pojedinog pravila, j iterira po pravilima, a k po klasama :

1. Stupanj podudaranja (3), $\mu_{A_j}(x)$ – računa se jačina aktivacije antecedent dijela pravila. Vrijednosti pripadnosti svih funkcija povezuju se koristeći t-normu (eng. *conjunction operator*).

$$\mu_{A_j}(x) = T\left(\mu_{A_{j1}}(x_1), \dots, \mu_{A_{jn}}(x_n)\right) \quad (3)$$

2. Stupanj asocijacije (4), b_j^k – za svako se pravilo računaju vrijednosti pouzdanosti kojom dano pravilo klasificira primjer u svaku od klasa. Koristeći kombinacijski operator (eng. *combination operator*) spajaju se stupnjevi podudaranja s težinama pravila.

$$b_j^k = h(\mu_{A_j}(x), RW_j) \quad (4)$$

3. Stupanj klasifikacijske sigurnosti uzorka za sve klase (5), Y_k – koristi se agregacijska funkcija (eng. *aggregation function*) kako bi se kombinirale asocijacijske vrijednosti izračunate u prošlim koracima.

$$Y_k = f(b_1^k, \dots, b_L^k) \quad (5)$$

4. Klasifikacija (6) – koristi se funkcija odlučivanja (eng. *decision function*) na svim stupnjevima klasifikacije kako bi se odredilo kojoj klasi podatak najvjerojatnije pripada.

$$F(Y_1, \dots, Y_M) = \underset{k}{\operatorname{arg\,max}}(Y_k) \quad (6)$$

2 IVTURS

IVTURS-ov sustav za klasifikaciju stvara se kroz tri faze. Prva faza, koja se temelji na algoritmu *FARC-HD* [2], inicijalizira početni sustav koji se u sljedećim fazama unapređuje. Druga faza je transformiranje inicijalnog sustava u intervalni sustav. Treća faza je odabir pravila i ugađanje parametara s ciljem poboljšanja klasifikacije sustava. Vrlo bitan dio sustava je metoda rasuđivanja, koju su autori u ovom obliku osmislili i prvi put iskoristili za *IVTURS* [1]. Na slici 2.1 je prikazan *IVTURS*-ov tijek.



Slika 2.1 Prikaz *IVTURS*-ovog tijeka [1]

U ovom poglavlju prikazat će se stvaranje i učenje sustava *IVTURS*. Svi primjeri u kojima su potrebni podaci koristit će skup podataka Iris. Iris je odabran zbog svoje jednostavnosti, zbog koje će se lako iskoristiti za prikaz najjednostavnijih izračuna unutar sustava.

Skup podataka Iris sastoji se od 150 primjera mjerenih pomoću četiri numerička atributa i podijeljenih u tri klase. Atributi i klase su prikazani u tablicama 2.1.

Tablice 2.1 Prikaz atributa i klase skupa podataka Iris

Atribut	Interval
<i>SepalLength</i> – <i>SP</i>	[4.3, 7.9]
<i>SepalWidth</i> – <i>SW</i>	[2.0, 4.4]
<i>PetalLength</i> – <i>PL</i>	[1.0, 6.9]
<i>PetalWidth</i> – <i>PW</i>	[0.1, 2.5]

Klasa
<i>Iris-setosa</i>
<i>Iris-versicolor</i>
<i>Iris-virginica</i>

2.1 *IVTURS*-ova metoda rasuđivanja

Metoda rasuđivanja koju koristi *IVTURS* temelji se na općenitoj metodi, ali je proširena za rad s intervalnim vrijednostima [1]. Bitan alat *IVTURS*-ove metode rasuđivanja su funkcije *IV-REF* (eng. *interval valued restricted equivalence functions*).

IV-REF-ovi su intervalno proširenje *REF-ova* (eng. *restricted equivalence function*).

Koriste se kako bi se omogućilo ugađanje stupnja jednakosti između dvaju vrijednosti

[1]. *REF-ovi* su funkcije $[0, 1]^2 \rightarrow [0, 1]$, takve da [1][23]:

1. $REF(x, y) = REF(y, x), \forall x, y \in [0, 1]$
2. $REF(x, y) = 1$, akko $x = y$
3. $REF(x, y) = 0$, akko $x = 1$ i $y = 0$ ili $x = 0$ i $y = 1$
4. $REF(x, y) = REF(c(x), c(y)), \forall x, y \in [0, 1]$
5. $\forall x, y, z \in [0, 1]$, ako $x \leq y \leq z$ onda $REF(x, y) \geq REF(x, z)$ i $REF(y, z) \geq REF(x, z)$

$c(x)$ je jaka negacija (eng. *strong negation*), definirana kao $c(x) = 1 - x$. Autori za izgradnju *REF-ova* koriste automorfizme (eng. *automorphism*) [1]. Automorfizmi su kontinuirane i striktno rastuće funkcije $\phi: [0, 1] \rightarrow [0, 1]$ tako da $\phi(0) = 0$ i $\phi(1) = 1$. Točnije, autori koriste automorfizam $\phi(x) = x^a$, gdje se promjenom parametara a može izgraditi mnogo različitih funkcija [1].

Ako su ϕ_1, ϕ_2 automorfizmi, *REF* se gradi po izrazu (7),

$$REF(x, y) = \phi_1^{-1}(1 - |\phi_2(x) - \phi_2(y)|) \quad (7)$$

s jakom negacijom kao (8):

$$c(x) = \phi_2^{-1}(1 - \phi_2(x)) \quad (8)$$

IV-REF-ovi su funkcije $L([0, 1])^2 \rightarrow L([0, 1])$ takve da [1]:

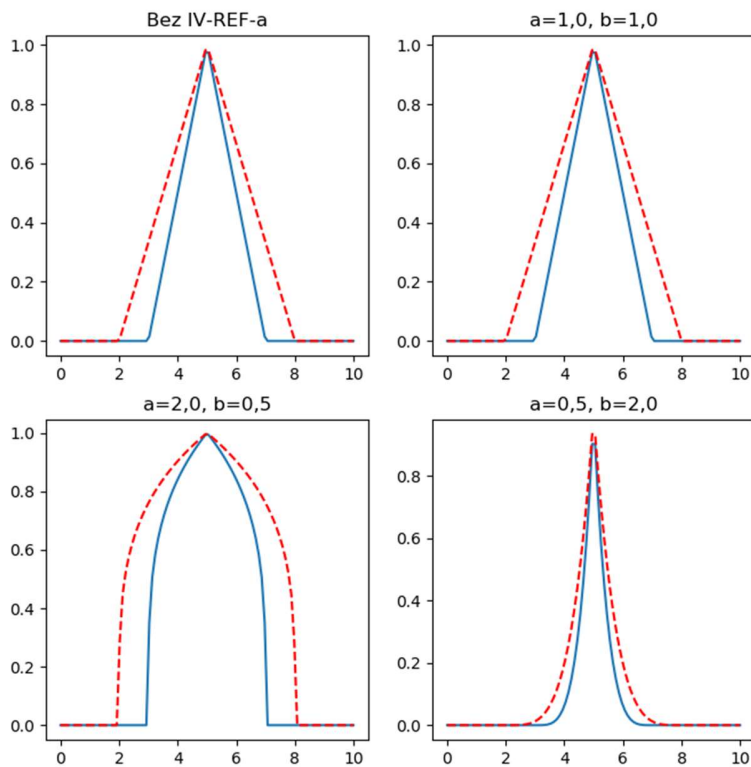
1. $IV-REF(x, y) = IV-REF(y, x), \forall x, y \in L([0, 1])$
2. $IV-REF(x, y) = 1_L$, akko $x = y$
3. $IV-REF(x, y) = 0_L$, akko $x = 1_L$ i $y = 0_L$ ili $x = 0_L$ i $y = 1_L$
4. $IV-REF(x, y) = IV-REF(N(x), N(y)), \forall x, y \in L([0, 1])$
5. $\forall x, y, z \in L([0, 1])$, ako $x \leq y \leq z$ onda $IV-REF(x, y) \geq IV-REF(x, z)$ i $IV-REF(y, z) \geq IV-REF(x, z)$

$N(x), N(y)$ su intervalne negacije. Prikaz generalnog izgleda *IV-REF-a* je dan izrazom (9), a prikaz *IV-REF-a* s gore navedenim *REF-om* izrazom (10), gdje su T i S , t-norma te s-norma.

$$IV-REF(x, y) = \left[T \left(REF(\underline{x}, \underline{y}), REF(\bar{x}, \bar{y}) \right), S \left(REF(\underline{x}, \underline{y}), (REF(\bar{x}, \bar{y})) \right) \right] \quad (9)$$

$$IV-REF(x, y) = \left[T(\phi_1^{-1}(1 - |\phi_2(\underline{x}) - \phi_2(\underline{y})|)), \phi_1^{-1}(1 - |\phi_2(\bar{x}) - \phi_2(\bar{y})|)), S(\phi_1^{-1}(1 - |\phi_2(\underline{x}) - \phi_2(\underline{y})|)), \phi_1^{-1}(1 - |\phi_2(\bar{x}) - \phi_2(\bar{y})|)) \right] \quad (10)$$

Ovakva definicija *IV-REF*-a omogućuje transformaciju oblika ulaznog intervala ovisno o parametrima automorfizma. Na slici 2.2 prikazana je intervalna trokutasta funkcija pripadnosti i njen izgled nakon izlaza iz *IV-REF*-a. Pripadnost za danu funkciju uspoređuje se sa savršenim intervalom kao kod *IVTURS*-ove metode rasuđivanja. Parametar *a* pripada automorfizmu ϕ_1 , a parametar *b* automorfizmu ϕ_2 .



Slika 2.2 Prikaz *IV-REF*-a s raznim parametrima

Metoda rasuđivanja koju koristi *IV-REF* prikazana je ispod [1], $x = (x_1, \dots, x_n)$ je primjer za klasifikaciju, M broj klasa, L broj pravila, $[\underline{RW}_j, \overline{RW}_j]$ intervalna težina pojedinog pravila, j iterira po pravilima a k po klasama:

1. Intervalni stupanj podudaranja (11), $[\underline{A}_j(x_p), \overline{A}_j(x_p)]$ – *IV-REF*-ovi se koriste kako bi izračunali sličnost funkcija podudaranja sa savršenim intervalom $[1, 1]$. Sličnosti za sve atribute kombiniraju se t-normom u stupanj podudaranja. Autori koriste intervalno množenje kao t-normu. Promjenom parametara *IV-REF*-a može se mijenjati izgled ulaznih funkcija pripadnosti.

$$[\underline{A}_j(x_p), \overline{A}_j(x_p)] = T \left(IV-REF \left([\underline{A}_{j1}(x_1), \overline{A}_{j1}(x_1)], [1, 1] \right), \dots, IV-REF \left([\underline{A}_{jn}(x_n), \overline{A}_{jn}(x_n)], [1, 1] \right) \right) \quad (11)$$

2. Intervalni stupanj asocijacije (12), $[\underline{b}_j^k, \overline{b}_j^k]$ – kombinira se težina pravila i stupanj podudaranja. Autori predlažu intervalno množenje kao kombinacijski operator:

$$[\underline{b}_j^k, \overline{b}_j^k] = h \left([\underline{A}_j(x), \overline{A}_j(x)], [\underline{RW}_j, \overline{RW}_j] \right) \quad (12)$$

gdje se težina računa kao:

$$[\underline{RW}_j, \overline{RW}_j] = \frac{\sum_{x \in C_j} [\underline{A}_j(x), \overline{A}_j(x)]}{\sum_{p=1}^P [\underline{A}_j(x_p), \overline{A}_j(x_p)]} \quad (13)$$

3. Intervalni stupanj klasifikacijske sigurnosti uzorka za sve klase (14), $[\underline{Y}_k, \overline{Y}_k]$ – agregacijskom funkcijom spajaju se stupnjevi asocijacije. Kao agregacijska funkcija koristi se intervalno zbrajanje:

$$[\underline{Y}_k, \overline{Y}_k] = f \left([\underline{b}_1^k, \overline{b}_1^k], \dots, [\underline{b}_L^k, \overline{b}_L^k] \right) \quad (14)$$

4. Klasifikacija (15) – funkcijom odlučivanja se odabire klasa kojoj primjer pripada. Za usporedbu vrijednosti klasa i odabir najbolje autori koriste totalni poredak (eng. *total order relationship*), definiran u [1][4].

$$F\left(\left[\underline{Y}_1, \overline{Y}_1\right], \dots, \left[\underline{Y}_M, \overline{Y}_M\right]\right) = \underset{k}{\operatorname{arg\,max}}\left(\left[\underline{Y}_k, \overline{Y}_k\right]\right) \quad (15)$$

Ako su $[\underline{x}, \overline{x}], [\underline{y}, \overline{y}] \in L([0,1])$ i $s([\underline{x}, \overline{x}]) = \underline{x} + \overline{x} - 1$ skor (eng. *score*) te $h([\underline{x}, \overline{x}]) = 1 - (\overline{x} - \underline{x})$ stupanj točnosti (eng. *accuracy degree*) $[\underline{x}, \overline{x}]$, totalni se poredak određuje na sljedeći način:

1. ako je $s([\underline{x}, \overline{x}]) > s([\underline{y}, \overline{y}])$ onda $[\underline{x}, \overline{x}] > [\underline{y}, \overline{y}]$
2. ako je $s([\underline{x}, \overline{x}]) = s([\underline{y}, \overline{y}])$ onda
 - a. ako je $h([\underline{x}, \overline{x}]) = h([\underline{y}, \overline{y}])$ onda $[\underline{x}, \overline{x}] = [\underline{y}, \overline{y}]$
 - b. ako je $h([\underline{x}, \overline{x}]) > h([\underline{y}, \overline{y}])$ onda $[\underline{x}, \overline{x}] > [\underline{y}, \overline{y}]$

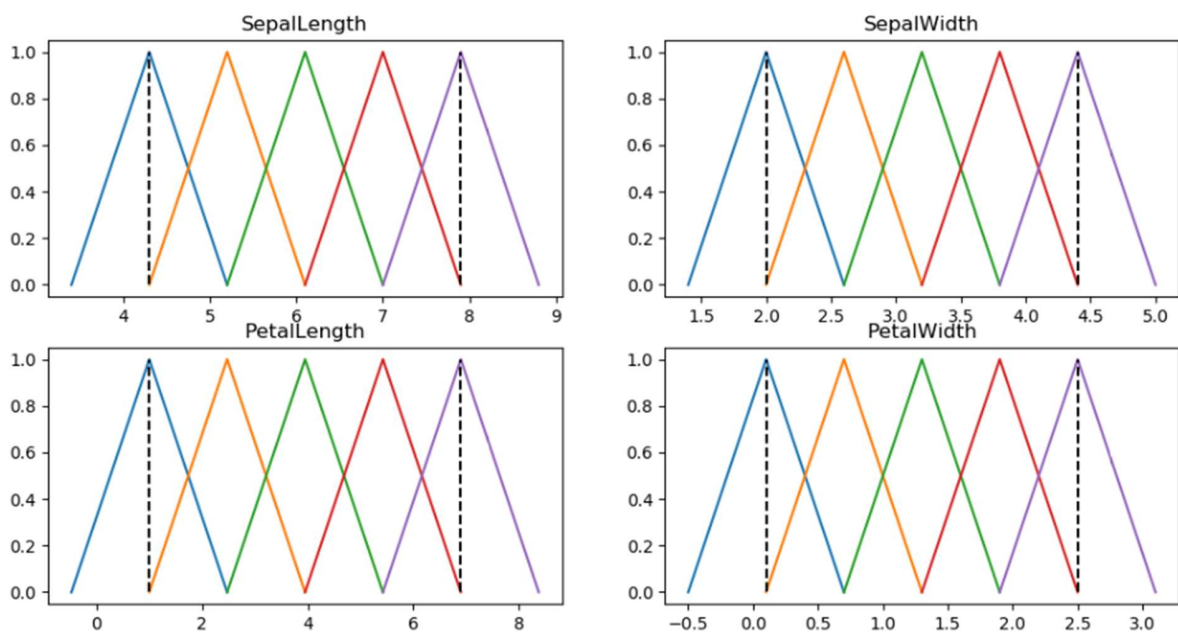
2.2 Inicijalizacija sustava

Općenito, sustavi drugog tipa mogu se stvoriti na dva načina. Prvi način je stvaranje sustava prvog tipa koji služi kao temelj za naknadnu nadogradnju u sustav drugog tipa. Drugi način je izgradnja sustava drugog tipa „od nule“. *IVTURS* koristi prvi način i sustav prvog tipa stvara se na temelju algoritma *FARC-HD* [1]. *FARC-HD* se sastoji od tri faze, pri čemu se prve dvije faze koriste kod inicijalizacije *IVTURS*-a, a treća faza kod optimizacije *IVTURS*-a. *IVTURS* je zapravo sustav stvoren na isti način kao *FARC-HD*, ali je transformiran i optimiran za rad s intervalima.

Ulazni podaci sustava opisuju se jezičnim varijablama, pri čemu se za svaku atributnu jezičnu varijablu definira nekolicina funkcija pripadnosti. Autori *IVTURS*-a za svaki atribut stvaraju pet pripadnih funkcija i sve su funkcije trokutastog oblika. Na slici 2.3 prikazane su funkcije pripadnosti definirane za skup podataka Iris ovisno o atributnim intervalima prikazanim u tablici 2.1. Pet funkcija pripadnosti prikazano je u različitim bojama, dok su crnom isprekidanom crtom prikazane donja i gornja granica intervala u kojem se vrijednosti atributa nalaze. Uz svaku funkciju veže se lingvistička

vrijednost. Lingvističke vrijednosti jednake su za sve varijable i vežu se uz funkcije istim redoslijedom kojim su nabrojane:

1. malen (eng. *low*), L
2. srednje malen (eng. *medium low*), ML
3. osrednji (eng. *medium*), M
4. srednje velik, (eng. *medium high*), MH
5. velik, (eng. *high*), H



Slika 2.3 Prikaz funkcija pripadnosti za attribute skupa podataka Iris

Nakon definiranja pripadnih funkcija, uporabom stabla pretraživanja pronalaze se česte kombinacije funkcija pripadnosti kod ulaznih podataka. Česte kombinacije su skupovi funkcija koje prekrivaju što više ulaznih podataka. Cilj je pronaći kombinacije koje prekrivaju podatke većinom iz jedne klase kako bi se one mogle iskoristiti u pravilima za tu klasu. Stablo u svakoj razini sadrži kombinacije dobivene spajanjem kombinacija iz prošle razine. U prvoj razini nalaze se svi jednočlani skupovi istovjetni pojedinačnim funkcijama pripadnosti. Već se u drugoj razini skupovi kombiniraju i dolazi do raznovrsnijih skupova. Unutar istog skupa ne mogu se nalaziti funkcije koje se odnose na isti atribut. Kako bi se smanjio broj mogućih kombinacija, ograničava se dubina stabla i uvode se dvije metrike za ocjenu skupova. Potpora (eng. *support*) i

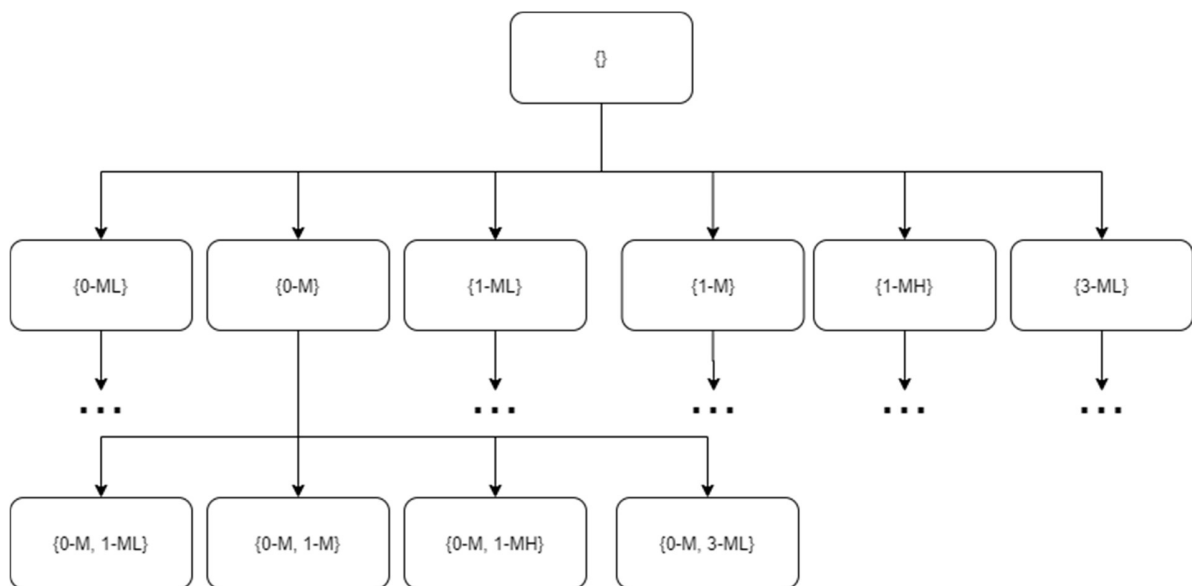
pouzdanost (eng. *confidence*) računaju se za svaku kombinaciju kako bi se, ovisno o njima, zaustavilo proširivanje nedovoljno ili dovoljno čestih skupova [2]. Potpora se računa prema izrazu (16), ona se koristi kako bi odredili prekriva li skup dovoljno podataka kako bi ga smatrali čestim. U brojniku se nalazi suma pripadnosti skupova, prema podacima koje skup prekriva, dok je u nazivniku broj svih ulaznih podataka. Vrijednost pripadnosti cijelog skupa za pojedini podatak računa se t-normom, kojom se kombiniraju sve pripadnosti funkcija koje se nalaze u skupu. Autori koriste produktnu t-normu.

$$Support(A) = \frac{\sum_{x_p} \mu_A(x_p)}{|N|} \quad (16)$$

Pouzdanost se računa prema izrazu (17) i koristi se kako bi se odredile kombinacije funkcija koje su dovoljno pouzdane i koje se ne moraju dalje proširivati [2]. U nazivniku je ista suma kao i u brojniku potpore, a u brojniku je također jednaka suma koja je ograničena na podatke jedne klase.

$$Confidence(A) = \frac{\sum_{x_p \in Klasa\ i} \mu_A(x_p)}{\sum_{x_p} \mu_A(x_p)} \quad (17)$$

Na slici 2.4 nalazi se prikaz stabla pretraživanja za skup podataka Iris. Korijen stabla je prazan skup iz kojeg u prvoj razini proizlaze jednočlani skupovi funkcija pripadnosti. Prikazano je kako se proširivao jednočlani skup $\{0-M\}$, pri čemu je svaki atribut označen brojem na sljedeći način *SepalLength* – 0, *SepalWidth* – 1, *PetalLength* – 2 i *PetalWidth* – 3. Skupovi koji su nastali iz skupa $\{0-M\}$ dalje se proširuju ili ne proširuju, ovisno o potpori i pouzdanosti skupova.



Slika 2.4 Prikaz dijela stabla pretraživanja za skup podataka Iris

Iz početne inicijalizacije proizlaze tri hiperparametra sustava - dubina stabla, maksimalna pouzdanost i minimalna potpora. Autori određuju njihove vrijednosti kao 3, 0,8 i 0,05. Važno je naglasiti da se za minimalnu potporu ne uzima izravno hiperparametar, već se minimum za svaku klasu množi s udjelom pojedine klase u sveukupnim podacima. Time se izbjegavaju problemi stvaranja pravila kod neuravnoteženosti klasa [1]. Također, maksimalna dubina ne predstavlja maksimalni broj atributa u pravilima, već broj funkcija unutar kombinacije nakon kojeg se kombinacija neće pokušavati proširiti. Ako se prethodnim proširenjima stvorila kombinacija veća od maksimalne dubine, ona će se uvažiti, ali se neće proširiti.

2.3 Odabir pravila

Nakon pronalaska čestih kombinacija one postaju antecedent dijelovi pravila dok se za konzekvent postavljaju klase za koje su kombinacije pronađene. Pronađene kombinacije su često redundantne i mnogo ih je više nego što je potrebno za rad klasifikatora. Stoga sljedeća faza iz početnih pravila odabire onu kombinaciju koja najbolje opisuje podatke iz svake klase. Odabir se provodi pronalaskom podgrupa (eng. *subgroup discovery*).

Iterirajući po klasama, računaju se težine prekrivenih i neprekrivenih podataka. Početno svakom podatku težina je jednaka jedan, a računa se prema izrazu (18)[2]:

$$w(x, i) = \frac{1}{i+1} \quad (18)$$

Dalje se računa težina za svako pravilo proizašlo iz prošle faze. Težina pravila se računa prema izrazu (19)[2]:

$$wWRAcc''(R) = \frac{n''(A \cdot C_j)}{n'(C_j)} \cdot \left(\frac{n''(A \cdot C_j)}{n''(A)} - \frac{n(C_j)}{N} \right) \quad (19)$$

$n''(A)$ i $n''(A \cdot C_j)$ su sume umnoška težina i stupnjeva podudaranja podataka s pravilom. Gdje je $n''(A)$ suma svih podataka, dok je $n''(A \cdot C_j)$ suma svih točno prekrivenih podataka (podataka koji pripadaju istoj klasi kao i pravilo). $n'(C_j)$ je suma svih težina točno prekrivenih podataka. $n(C_j)$ je broj podataka iz iste klase kao i pravilo, a N je broj svih podataka. Nakon računanja težina svih pravila najbolje se pravilo odabire i izbacuje iz budućih računanja. Također se povećava brojač prekrivanja za svaki podatak koje to pravilo pokriva. Nakon povećanja brojača ponovno se računaju težine podataka. U slučaju da je prekriven više od k puta, podatak se izbacuje i ne ulazi u buduća računanja. Autori su vrijednost hiperparametra k postavili na 2. Iteriranje i odabir pravila se zaustavljaju nakon što su svi podaci odbačeni ili sva pravila odabrana.

Neka su podaci koji će se koristiti za izračun težina prikazani u tablici 2.2.

Tablica 2.2 Podaci iz skupa podataka Iris

SepalLength	SepalWidth	PetalLength	PetalWidth	w_i	Klasa
5,8	4,0	1,2	0,2	1,0	Iris-setosa
5,4	3,9	1,7	0,4	1,0	Iris-setosa
5,5	3,5	1,3	0,2	0,5	Iris-setosa
6,0	2,2	4,0	1,0	0,5	Iris-versicolor

Pravilo za koje će se računati težina proizlazi iz skupa $\{0-M, 3-ML\}$:

Ako je SepalLength osrednji i PetalWidth srednje malen onda je klasa Iris-setosa.

U tablici 2.3 prikazane su vrijednosti funkcija pripadnosti, njihov umnožak tj. podudaranje antecedent dijela pravila s podacima i umnožak podudaranja s težinama

primjera. Funkcija izračuna težine pravila prikazana je pod (20). Težina svakog pravila može poprimiti vrijednost između -1 i 1. Pravilo sa težinom -0,0184 ne predstavlja dobro pravilo i vjerojatno ne bi bilo odabrano za upotrebu u klasifikatoru.

Tablica 2.3 Vrijednosti pripadnosti i umnošci

$\mu_0(x)$	$\mu_3(x)$	$\mu_0(x) \cdot \mu_3(x)$	$w_i \cdot \mu_0(x) \cdot \mu_3(x)$
0,667	0,167	0,113	$1,0 \cdot 0,113$
0,222	0,5	0,111	$1,0 \cdot 0,111$
0,333	0,167	0,056	$0,5 \cdot 0,056$
0,889	0,5	0,445	$0,5 \cdot 0,445$

$$wWRAcc''(R) = \frac{0,113+0,111+0,028}{1,0+1,0+1,0} \cdot \left(\frac{0,113+0,111+0,028}{0,113+0,111+0,028+0,223} - \frac{3}{4} \right) = -0,0184 \quad (20)$$

U ovom primjeru svi su podaci prekriveni pravilom i njihova se težina smanjuje ovisno o odabranim hiperparametrima.

Primjer pravila koja su pronađena za skup podataka *Iris* prilikom jednog pokretanja algoritma prikazana su u tablici 2.4.

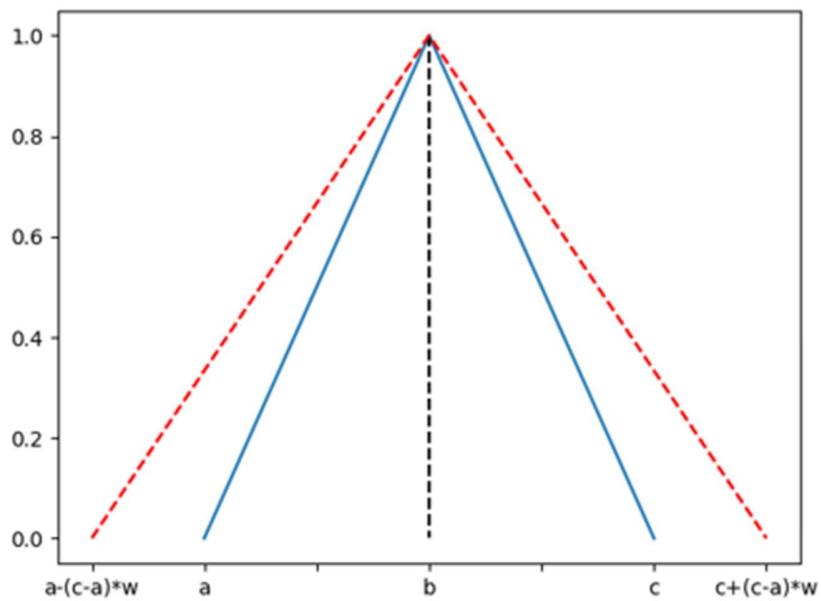
Tablica 2.4 Pravila za skup podataka *Iris*

<i>id</i>	<i>Antecedent</i>	<i>Konzekvent</i>
1	<i>PW malen (3-1)</i>	<i>Iris-setosa (0)</i>
2	<i>PL malen (2-1)</i>	<i>Iris-setosa (0)</i>
3	<i>PW osrednji (3-3)</i>	<i>Iris-versicolor (1)</i>
4	<i>PL osrednji (2-3)</i>	<i>Iris-versicolor (1)</i>
5	<i>PW srednje velik (3-4)</i>	<i>Iris-virginica (2)</i>
6	<i>PW velik (3-5)</i>	<i>Iris-virginica (2)</i>
7	<i>SW srednje malen (1-2), PW srednje velik (3-4)</i>	<i>Iris-virginica (2)</i>
8	<i>PL velik (2-5)</i>	<i>Iris-virginica (2)</i>

2.4 Intervalno neizraziti klasifikacijski sustav

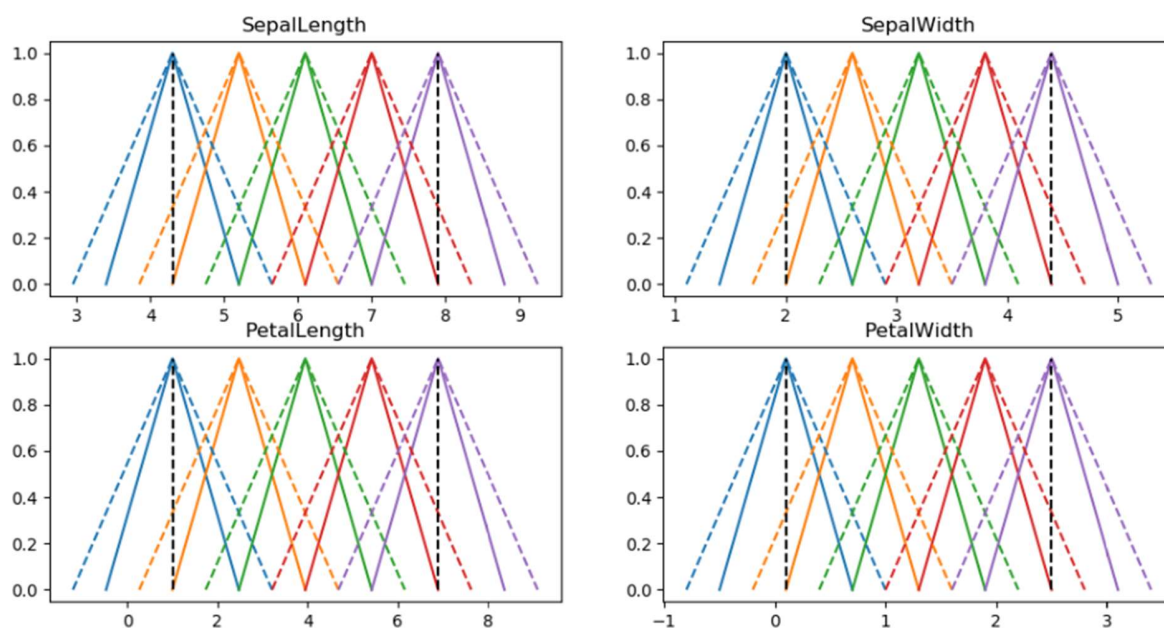
Nakon odabira pravila i inicijalizacije sustava prvog tipa potrebno je formirati sustav drugog tipa. Za njega se transformiraju funkcije pripadnosti kako bi pripadnost umjesto dosadašnje jedne vrijednosti definirali kao interval. Kada su funkcije pripadnosti trokutastog oblika njih opisuju tri vrijednosti (točke), dvije granične i jedna vršna za sto postotnu pripadnost. Kako bi se proširile, dvije nove granične točke

pomiču se od sredine funkcije. Za pomicanje se koristi faktor w , koji se množi s područjem koje prekrivaju početne granične točke. Autori za faktor w predlažu vrijednost 0,25, čime transformirana funkcija prekriva upola veće područje. Originalna funkcija postaje gornja granica proširene intervalne funkcije dok nova funkcija s povećanim područjima postaje donja granica. Slika 2.5 prikazuje originalnu funkciju (plavu) i proširenu funkciju (crvenu) s točkama koje ih određuju.



Slika 2.5 Prikaz originalne i proširene funkcije

Slika 2.6 prikazuje transformirane funkcije sa slike 2.3.



Slika 2.6 Transformirane funkcije

2.5 Optimizacija sustava

Optimiranjem sustava želi se postići što bolja klasifikacija i izbaci ona pravila koja nisu potrebna za optimalnu klasifikaciju. Korištenjem funkcija *IV-REF* moguće je definirati puno različitih funkcija mijenjanjem parametara a i b [1]. Autori su stoga ograničili parametre na vrijednosti u intervalu $[0,01, 100]$. Cilj je pronaći parametre za *IV-REF*-e svakog atributa kako bi točnost klasifikatora bila maksimalna. Nadalje, kako bi se izbacila redundantna i suvišna pravila, način optimiranja mora moći smanjiti broj pravila. Autori rada za optimiranje sustava predlažu korištenje evolucijskog modela *CHC* zbog njegovih dobrih rezultata kod rješavanja sličnih i kompleksnih problema [1].

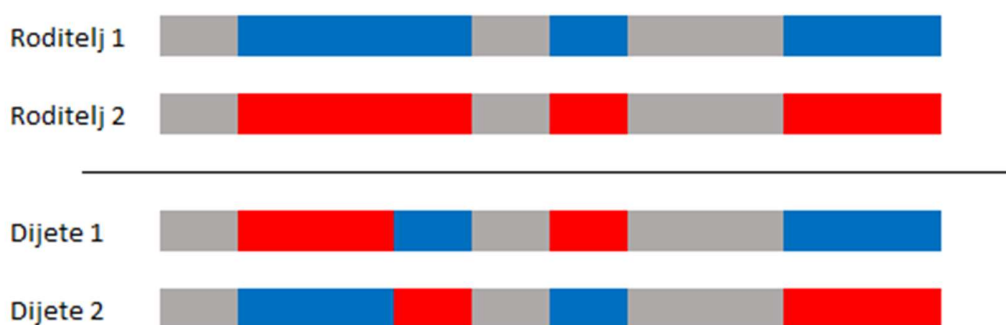
Pet je najvažnijih segmenata rada evolucijskog modela *CHC* [1][2]:

- 1) Kodiranje – budući da je cilj pronaći parametre i smanjiti broj pravila, kromosomi evolucijskog modela moraju biti definirani za oboje. Autori odabiru podjelu kromosoma u dva dijela. Prvi će dio imati gene koji

predstavljaju korištenje pojedinog pravila. Jednostavnom notacijom istina/laž (eng. *true/false*) naznačit će se koristi li dani kromosom to pravilo ili ne. U drugom dijelu svaki parametar *IV-REF*-a će se binarno kodirati s n bitova. Autori predlažu korištenje 30 bitova. Kako bi se smanjio interval mogućih kodiranih vrijednosti parametara *IV-REF*, parametri će biti vrijednosti iz intervala $[0,01, 1,99]$, a pravi parametri računat će se prema izrazu (21):

$$a = \begin{cases} a, & \text{ako je } 0 < a \leq 1 \\ \frac{1}{2-a}, & \text{ako je } 1 < a < 2 \end{cases} \quad (21)$$

- 2) Inicijalna populacija – kako bi se sva pravila našla unutar inicijalne populacije evolucijskog modela, u populaciju se postavlja kromosom sa svim genima postavljenim na 1. Time taj kromosom koristi sva pravila i svi *IV-REF*-ovi su jednaki početnim funkcijama pripadnosti. Autori predlažu broj kromosoma u inicijalnoj populaciji jednak 50.
- 3) Evaluacija kromosoma – postoje razne evaluacijske metrike koje se mogu koristiti za ocjenu kromosoma, a autori su se odlučili za točnost klasifikatora (eng. *accuracy rate*).
- 4) Operator križanja – budući da se kromosom sastoji od dva dijela, potrebna su i dva operatora križanja:
 - a) Za dio kromosoma koji se odnosi na pravila koristi se polovično uniformno križanje (eng. *half uniform crossover, HUX*). Kod *HUX*-a križanjem nastaje dvoje djece i svako dijete nasljeđuje jednu polovicu gena koji su različiti kod roditelja. Geni koje će dijete naslijediti od kojeg roditelja biraju se nasumično. Na slici 2.7 je prikazan izgled roditelja i



Slika 2.7 Prikaz roditelja i djece prilikom *HUX* križanja

djece, pri čemu siva boja označava jednake gene dok plava i crvena označavaju različite gene kod jednog odnosno drugog roditelja.

- b) Dio kromosoma zadužen za parametre funkcija *IV-REF* koristi *PBX- α* [22] (eng. *parent centric BLX- α*). Ako su odabrani roditelji $X=(x_1, \dots, x_N)$ i $Y=(y_1, \dots, y_N)$, djeca nastala *PBX- α* su $Z_1=(z_1^1, \dots, z_n^1)$ i $Z_2=(z_1^2, \dots, z_n^2)$ [22], gdje z_i^1 nastaje iz intervala $[l_i^1, u_i^1]$ (22):

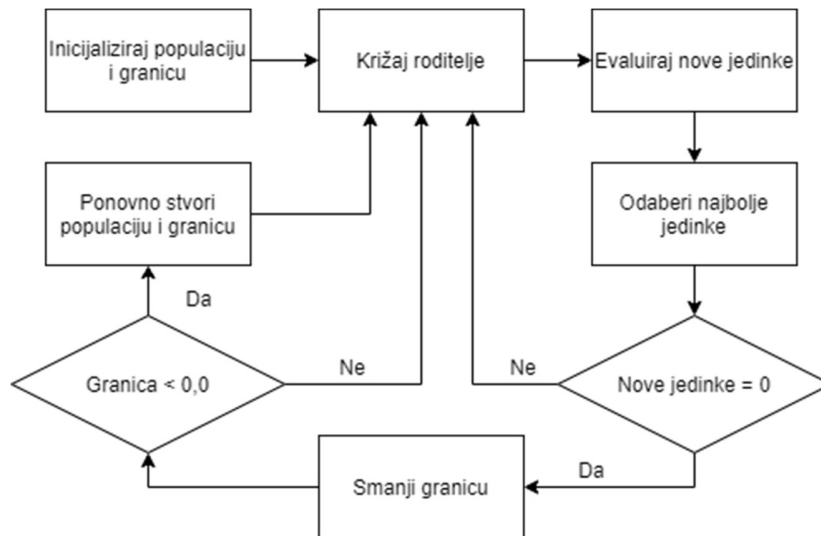
$$l_i^1 = \max\{a_i, x_i - I \cdot \alpha\}, u_i^1 = \max\{b_i, x_i + I \cdot \alpha\} \quad (22)$$

a z_i^2 se izabire iz intervala $[l_i^2, u_i^2]$ (23):

$$l_i^2 = \max\{a_i, y_i - I \cdot \alpha\}, u_i^2 = \max\{b_i, y_i + I \cdot \alpha\} \quad (23)$$

gdje je za oba intervala $I = |x_i - y_i|$. Ovakvom definicijom maksimizira se vjerojatnost da će vrijednost djeteta biti što sličnija roditeljima ili malo pomaknuta iz njihovih intervala. Ovisno o α mijenja se veličina područja u kojem se može nalaziti vrijednost djeteta.

- 5) Ponovno pokretanje – da bi se izbjegao lokalni optimum, koristi se ponovno pokretanje evolucijskog modela [2]. Kada kromosomi postanu previše slični, najbolji kromosom se sačuva, a ostali se ponovno generiraju. Time se ostvaruje elitizam preživljavanja najboljeg i omogućuje pretraživanje (potencijalno) nepretraženog dijela prostora rješenja. Kako bi se utvrdilo da su kromosomi suviše slični, prilikom odabira roditelja računa se Hammingova udaljenost između njih (eng. *hamming distance*). Ako je njihova udaljenost manja od vrijednosti L , odabir se odbacuje i ne stvaraju se novi kromosomi. Vrijednost L inicijalno je postavljena na najveću moguću Hammingovu udaljenost između dva roditelja. Svaki put kada u jednoj iteraciji ne nastane nijedno novo dijete L se smanjuje ili za 1 ili za postotak inicijalne vrijednosti, ovisno o implementaciji. Ako L postane 0 ili manji, model se ponovno pokreće.

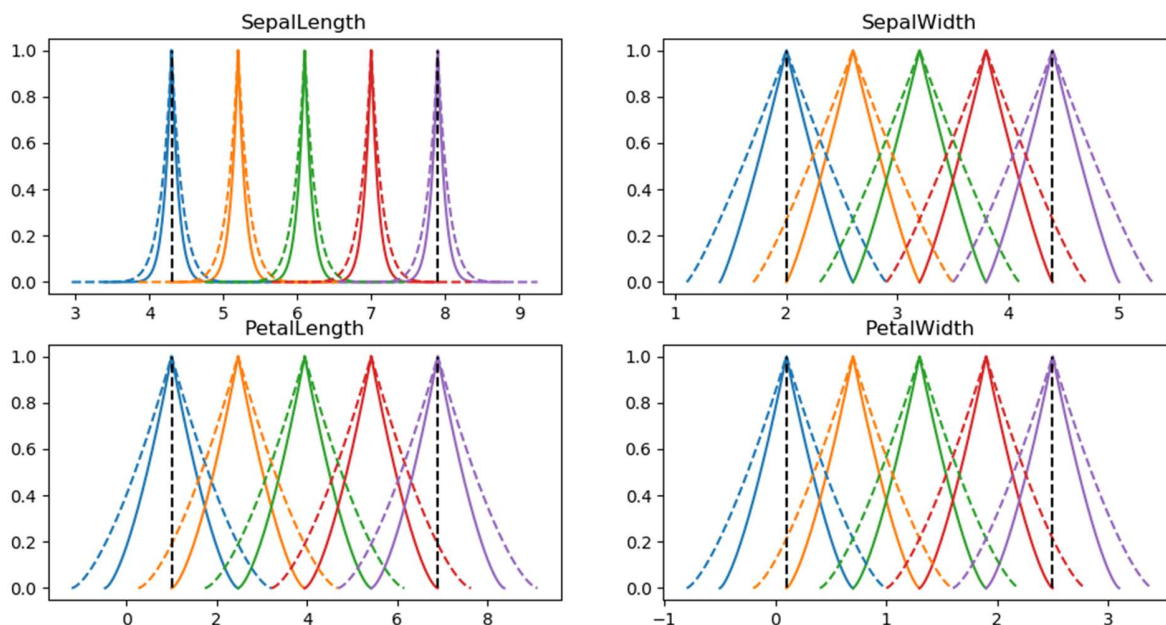


Slika 2.8 Prikaz toka rada CHC-a

Ograničenje rada evolucijskog modela koji koristi sustav *IVTURS*, za razliku od mnogih drugih sustava, nije broj iteracija, nego broj maksimalnih evaluacija kromosoma. Autori ga postavljaju na 20000. Slika 2.8 prikazuje tijek evolucijskog modela *CHC* koji se koristi u *IVTURS*-u. Nakon optimizacije za klasifikaciju sustavom koji sadrži pravila navedena u tablici 2.4 koriste se prvo (1), četvrto (4), šesto (6), sedmo (7) i osmo (8) pravilo. Pronađeni parametri *IVREF*-a prikazani su u tablici 2.5, a izgled funkcija pripadnosti uspoređenih sa savršenim intervalom prikazan je na slici 2.9.

Tablica 2.5 Prikaz parametara *IVREF*-a

	<i>SepalLength</i>	<i>SepalWidth</i>	<i>PetalLength</i>	<i>PetalWidth</i>
a	0,386	0,963	0,722	0,816
b	1,707	1,129	1,085	1,128



Slika 2.9 Prikaz funkcija pripadnosti uspoređenih sa savršenim intervalom

2.6 Klasifikacija primjera

Pravila koja se koriste za klasifikaciju podataka prikazana su u tablici 2.4. Prikaz parametara *IVREF*-a nalazi se u tablici 2.5. Klasifikacija sustava provest će se na prvom primjeru iz tablice 2.2. Vrijednosti pripadnosti za dani primjer prikazane su u tablici 2.6. Pripadnosti uspoređene funkcijama *IVREF* nalaze se u tablici 2.7.

Tablica 2.6 Vrijednosti pripadnosti

	<i>malen</i>		<i>srednje malen</i>		<i>osrednji</i>		<i>srednje velik</i>		<i>velik</i>	
	D	G	D	G	D	G	D	G	D	G
<i>SepalLength</i>	0	0	0,33	0,55	0,66	0,77	0	0,11	0	0
<i>SepalWidth</i>	0	0	0	0	0	0,11	0,66	0,77	0,33	0,55
<i>PetalLength</i>	0,86	0,91	0,13	0,42	0	0	0	0	0	0
<i>PetalWidth</i>	0,83	0,88	0,16	0,44	0	0	0	0	0	0

Tablica 2.7 Usporedba pripadnosti *IVREF* funkcijama

	<i>malen</i>		<i>srednje malen</i>		<i>osrednji</i>		<i>srednje velik</i>		<i>velik</i>	
	D	G	D	G	D	G	D	G	D	G
<i>SepalLength</i>	0	0	0	0	0,03	0,11	0	0	0	0
<i>SepalWidth</i>	0	0	0	0	0	0,07	0,62	0,74	0,27	0,50

<i>PetalLength</i>	0,80	0,86	0,04	0,27	0	0	0	0	0	0
<i>PetalWidth</i>	0,77	0,84	0,08	0,32	0	0	0	0	0	0

Stupnjevi podudaranja, težina pravila i stupnjevi asocijacije za pravila prikazani su u tablici 2.8.

Tablica 2.8 Prikaz stupnjeva podudaranja, težina pravila i stupnjeva asocijacije

Pravilo/klasa	$[\underline{A}_j(x_p), \overline{A}_j(x_p)]$	$[\underline{RW}_j, \overline{RW}_j]$	$[\underline{b}_j^k, \overline{b}_j^k]$
1/0	[0,0,0,0]	[1,0,1,0]	[0,0,0,0]
2/0	∅	[0,982,1,0]	∅
3/1	∅	[1,0,1,0]	∅
4/1	[0,0,0,0]	[0,810,0,915]	[0,0,0,0]
5/2	∅	[0,758,0,873]	∅
6/2	[0,079,0,326]	[0,989,1,0]	[0,078,0,326]
7/2	[0,0,0,0]	[0,652,0,858]	[0,0,0,0]
8/2	[0,0,0,0]	[0,991,1,0]	[0,0,0,0]

Ovakav klasifikator netočno klasificira prvi primjer iz tablice 2.2. Kao što je prikazano u tablici 2.9, primjer je klasificiran kao *Iris-virginica* dok zapravo pripada u klasu *Iris-setosa*.

Tablica 2.9 Prikaz intervalnih klasifikacijskih sigurnosti

$\sum_{j \in k=0} [\underline{b}_j^0, \overline{b}_j^0]$	$\sum_{j \in k=1} [\underline{b}_j^1, \overline{b}_j^1]$	$\sum_{j \in k=2} [\underline{b}_j^2, \overline{b}_j^2]$	Klasa
[0,0,0,0]	[0,0,0,0]	[0,078,0,326]	<i>Iris-virginica</i>

3 Promjene uvedene u algoritam IVTURS

U ovom su poglavlju navedene promjene koje su uvedene u rad sustava *IVTURS* u svrhu postizanja boljih rezultata.

3.1 Choquetovi integrali

Odabir agregacijske funkcije prikazane pod (5) i (14) iznimno je važan za rad neizrazitih sustava. Njena uloga u metodi rasuđivanja je agregacija svih zaključaka pravila za pojedinu klasu kako bi se dobila sigurnost klasifikacije u pojedinu klasu. U [12] je osmišljena nova metoda rasuđivanja u kojoj se Choquetov integral koristi kao agregacijska funkcija. Choquetov integral već se u ranijim istraživanjima pokazao kao dobar odabir za klasifikaciju [12][13].

Agregacijska funkcija $A: [0, 1]^n \rightarrow [0, 1]$ mora imati sljedeća svojstva [13]:

1. A je rastuća funkcija za svaki argument - $\forall i \in \{1, \dots, n\}$ ako je $x_i \leq y$ onda $A(x_1, \dots, x_n) \leq A(x_1, \dots, x_{i-1}, y, x_{i+1}, \dots, x_n)$. Ako se sigurnost u zaključak od jednog pravila zamjeni sa sigurnošću koja je jednaka ili veća, očekuje se da i sveukupni zaključak ostane isti ili da se poveća.
2. A zadovoljava granične uvjete - $\forall i \in \{1, \dots, n\}$ ako je $x_i = 1, y_i = 0$ onda $A(x_1, \dots, x_n) = 1, A(y_1, \dots, y_n) = 0$. U slučaju potpunih informacija o zaključcima očekuje se kako će i sveukupni zaključak biti potpun, suprotno u slučaju nikakvih informacija zaključak također mora biti nikakav.

IVTURS koristi intervalne neizrazite skupove, pa je stoga potrebna intervalna agregacijska funkcija $A: L(0, 1)^n \rightarrow L(0, 1)$ koja ima jednaka, ali intervalna svojstva:

1. $A(x_1, \dots, x_n) = [\underline{A}(\underline{x}_1, \dots, \underline{x}_n), \bar{A}(\bar{x}_1, \dots, \bar{x}_n)]$ je rastuća funkcija za svaki argument - $\forall i \in \{1, \dots, n\}$ ako je $\underline{x}_i \leq \underline{y}$ i $\bar{x}_i \leq \bar{y}$ onda $\underline{A}(\underline{x}_1, \dots, \underline{x}_n) \leq \underline{A}(\underline{x}_1, \dots, \underline{x}_{i-1}, \underline{y}, \underline{x}_{i+1}, \dots, \underline{x}_n)$ i $\bar{A}(\bar{x}_1, \dots, \bar{x}_n) \leq \bar{A}(\bar{x}_1, \dots, \bar{x}_{i-1}, \bar{y}, \bar{x}_{i+1}, \dots, \bar{x}_n)$

2. $A(x_1, \dots, x_n) = [A(\underline{x}_1, \dots, \underline{x}_n), \bar{A}(\bar{x}_1, \dots, \bar{x}_n)]$ zadovoljava granične uvjete - $\forall i \in \{1, \dots, n\}$ ako je $x_i = [1, 1], y_i = [0, 0]$ onda $A(x_1, \dots, x_n) = [1, 1], A(y_1, \dots, y_n) = [0, 0]$.

Agregacijske funkcije su u početku bile monotono rastuće, ali naknadnim generalizacijama pronađene su razne agregacijske funkcije sa slabijim tipovima rasta [13]. Neke od njih su t-norme, s-norme, funkcije preklapanja (eng. *overlap functions*) i grupiranja (eng. *grouping functions*), kopule i semikopule (eng. *copulas and semicopulas*), težinske (kvazi-) aritmetičke funkcije srednje vrijednosti (eng. *weighted (quasi-) arithmetic means*) te operatori OWA (eng. *OWA operators*) [13].

Neizrazita mjera (eng. *fuzzy measure*) je vrlo važan dio Choquetova integrala [13]. Njom se definiraju veze između elemenata koji se agregiraju. Neizrazita mjera $m: 2^N \rightarrow [0, 1]$ zadovoljava sljedeća svojstva, gdje je $X = \{x_1, \dots, x_N\}$ konačni skup:

1. m je rastuća funkcija - ako je $A \subseteq B \subseteq X$ onda je $m(A) \leq m(B) \leq 1$
2. m zadovoljava granične uvjete - $m(\emptyset) = 0, m(X) = 1$

Dvije mjere iz [12] prikazane su ispod. Neizrazita mjera m_{PM} je mjera snage (eng. *power measure*) prikazana pod (24)[12][13]:

$$m_{PM}(X) = \left(\frac{|X|}{n}\right)^q, q > 0 \quad (24)$$

Neizrazita mjera m_{WM} je težinska mjera srednje vrijednosti (eng. *weighted mean*) (25), gdje je $m(\{1\}) = w_1, \dots, m(\{n\}) = w_n$, a vektor težina $w = (w_1, \dots, w_n) \in [0, 1]^n$:

$$m_{WM}(X) = \sum_{i \in X} m(\{i\}) \quad (25)$$

Diskretni Choquetov integral s obzirom na neizrazitu mjeru m prikazan je pod (26)[13]:

$$C_m(x) = \sum_{i=1}^n (x_i \cdot m(A_i) - x_{i-1} \cdot m(A_i)) \quad (26)$$

gdje su:

1. $x = (x_1, \dots, x_n) \in [0, 1]^n$, x je sortirani vektor tako da $x_1 \leq \dots \leq x_n$
2. $x_0 = 0$

3. $A_{(i)} = \{(i), \dots, (n)\}$, je podskup svih indeks od $n - i + 1$ najvećih elemenata

Kao i agregacijske funkcije, razni tipovi Choquetovih integrala su se pojavili u literaturi [13]. U [13] su navedene razne generalizacije integrala, ali je samo jedna ostvarila statistički bolje rezultate od sustava *IVTURS*. Stoga će se za promjenu *IVTURS*-ove metode rasuđivanja koristiti generalizacija koja je pokazala bolje rezultate.

Generalizirani integral (27) koji je pokazao bolje rezultate od *IVTURS*-a je integral

$C_{F_1 F_2}^m$:

$$C_{F_1 F_2}^m(x) = \min \left\{ 1, x_{(1)} + \sum_{i=2}^n F_1 \left(x_{(i)}, m(A_{(i)}) \right) - F_2 \left(x_{(i-1)}, m(A_{(i)}) \right) \right\} \quad (27)$$

Gdje su F_1 i F_2 fuzijske funkcije, a m je neizrazita mjera. Sljedeća svojstva moraju biti ispunjena:

1. neizrazita mjera m mora biti simetrična neizrazita mjera
2. funkcija F_1 dominira nad funkcijom F_2
3. funkcija F_1 je (1,0)-rastuća funkcija

U [13] se kao F_1 koristila agregacijska funkcija F_{GL} (28) a kao F_2 t-norma T_M (29).

Neizrazita mjera koju su koristili je mjera snage m_{PM} (24). Takvom definicijom nastaje

$C_{F_{GL} T_M}^{m_{PM}}$ integral (30):

$$F_{GL}(x, y) = \sqrt{\frac{x \cdot (y+1)}{2}} \quad (28)$$

$$T_M(x, y) = \min\{x, y\} \quad (29)$$

$$C_{F_{GL} T_M}^{m_{PM}}(x) = \min \left\{ 1, x_{(1)} + \sum_{i=2}^n \sqrt{\frac{x_{(i)} \cdot (m(A_{(i)})+1)}{2}} - \min\{x_{(i-1)}, m(A_{(i)})\} \right\} \quad (30)$$

Kako bi se poviše definirani Choquetov integral (30) mogao koristiti u *IVTURS*-u, potrebno ga je proširiti za rad s intervalima. Intervalni ulazni vektor x je skup elemenata nastalih iz $L([0,1])^n$ koji su sortirani koristeći totalni poredak naveden u *IVTURS*-ovoj metodi rasuđivanja. Izračun neizrazite mjere se ne mijenja, ali njen izlaz postaju dvije jednake vrijednosti, jedna za donju, a druga za gornju granicu intervala.

Dvije nove metode rasuđivanja koristeći diskretni intervalni Choquetov integral $C_{FGLTM}^{m_{PM}}$ kao agregacijsku funkciju. Kod prve metode, u potpunosti će se preuzeti *IVTURS*-ova metoda rasuđivanja, dok će se kod druge metode odbaciti korištenje *IV-REF*-ova. Time će se intervalni stupanj podudaranja (11) kod druge metode računati izravno koristeći t-normu s funkcijama podudaranja. Cilj ovih promjena je doznati ostvaruje li Choquetov integral bolje rezultate od *IVTURS*-a sam ili s *IV-REF*-ovima.

Za najbolje rezultate kod korištenja neizrazite mjere m_{PM} potrebno je parametar q podijeliti tako da svaka klasa ima svoju zasebnu vrijednost [12]. Odabir vrijednosti parametara q predstavlja vrlo važan segment ovakvog pristupa. Kako bi se riješio taj problem, u [12] se predlaže korištenje evolucijskog modela *CHC* kako bi se pronašle optimalne vrijednosti. Parametar q se kodira te križa na jednak način kao i parametri funkcija *IV-REF*.

4 Ispitivanje rada klasifikatora

Za ispitivanje će se koristiti 45 skupova podataka preuzetih s [5][6]. Skupovi podataka razlikuju se po broju primjeraka, broju klasa, broju atributa i tipovima atributa. Njihov prikaz nalazi se u tablici 4.1. Neće se svi skupovi koristiti za ispitivanje svih klasifikatora. Usporedba s pojedinim klasifikatorima izvodit će se na skupovima podataka kao i u njihovim originalnim radovima.

Tablica 4.1 Prikaz skupova podataka

Ime skupa	Primjeri	Broj atributa	Numerički atributi	Kategorički atributi	Broj klasa	Kratice
Australian	690	14	8	6	2	AUS
Balance	625	4	4	0	3	BAL
Balance_2_vs_3	576	4	4	0	2	BAL2
Breast	683	10	10	0	2	BRE
Cleveland	297	13	13	0	5	CLE
Contraceptive	1473	9	9	0	3	CON
Crx	653	15	6	9	2	CRX
Dermatology	358	34	34	0	6	DER
Ecoli	336	7	7	0	8	ECO
Flare	1066	11	0	11	6	FLA
Fourclass	862	2	2	0	2	FOU
German	1000	20	7	13	2	GER
Glass	214	9	9	0	6	GLA
Glass_0_1_2_3_vs_4_5_6	214	9	9	0	2	GLA2
Haberman	306	3	3	0	2	HAB
Hayes-Roth	160	4	4	0	3	H-R
Heart	270	13	13	0	2	HEA
Ionosphere	351	33	33	0	10	ION
Iris	150	4	4	0	3	IRI
Led7digit	500	7	7	0	10	LED
Magic	1902	10	10	0	2	MAG
Mammographic	830	5	5	0	2	MAM
Monk	432	6	6	0	2	MON
New-Thyroid	215	5	5	0	3	N-T
Pageblocks	548	10	10	0	5	PAG
Penbased	1992	16	16	0	10	PEN
Pima	768	8	8	0	2	PIM
Ringnorm	7400	20	20	0	2	RIN
Saheart	462	9	8	1	2	SAH
Sonar	208	60	60	0	2	SON
Specfthart	267	44	44	0	2	SPE
Tae	151	5	5	0	3	TAE

Tae_1_vs_2	99	5	5	0	2	TAE2
Titanic	2201	3	3	0	2	TIT
Twonorm	740	20	20	0	2	TWO
Vehicle	846	18	18	0	4	VEH
Vehicle_1_vs_rest	846	18	18	0	2	VEH2
Vowel	990	13	10	3	11	VOW
Waveform_1_2_vs_3	5000	21	21	0	2	WAV
Wine red	1599	11	11	0	11	WIR
Wine	178	13	13	0	3	WIN
Wisconsin	683	9	9	0	2	WIS
Yeast	1484	8	8	0	10	YEA
Yeast_1_vs_2	892	8	8	0	2	YEA2
Zoo	101	16	0	16	7	ZOO

U ovom poglavlju reimplementacije *IVTURS-a* će se usporediti sa sličnim algoritmima i sustavima baziranim na pravilima. Također će se ispitati utjecaj dubine kod stabla pretraživanja i ispitati točnost dvije nove verzije bazirane na Choquetovim integralima. Po uzoru na [15], statistički testovi koji će se koristiti su Friedmanov test [14] (eng. *Friedman test*), Holmov test [16] (eng. *Holm's test*) i test Wilcoxonovih rangova [17] (eng. *Wilcoxon Signed-Ranks Test*). Nivo značajnosti α bit će postavljen na 0,05.

Klasifikatori koji će se koristiti za usporedbu su:

1. *FURIA – FURIA* [7] je nastala na temelju *RIPPER-a* [8]. Indukcija pravila koje *FURIA* koristi provodi se kroz tri faze. U prvoj se stvaraju pravila dodavanjem uvjeta u antecedent dio. Druga faza je fazifikacija inicijalno stvorenih pravila. Kod treće se faze za svako pravilo stvaraju dva dodatna i bira se ono najbolje. Implementacija ovog klasifikatora je preuzeta iz *KEEL-a*.
2. *C4.5 – C4.5* [18] je jedan od najpopularnijih klasifikatora. Za klasifikaciju gradi se stablo odluke tako da u svakom čvoru odabire atribut iz ulaznih podataka po kojem se želi podijeliti podatke. Nakon dovoljno podjela dolazi se do posljednjih razina u kojima svaki list određuje klasu kojoj podatak pripada. U ovom radu koristi se verzija *C4.5* koja stablo odluke pretvara u niz pravila za klasifikaciju. Implementacija ovog klasifikatora je preuzeta iz *KEEL-a*.
3. *FOATFRB – FOATFRB* [19] koristi optimizaciju voćne mušice (eng. *fruit fly optimization*) kako bi optimirao skup pravila za klasifikaciju. Kao čestice oko

kojih se skupljaju mušice odabiru se kombinacije nasumično generiranih pravila. Kombinacije i sama pravila iterativno se približavaju drugim boljim česticama/mušicama. Postoje dvije verzije klasifikatora: *FOATFRB* za definiciju pravila koristi trokutaste funkcije, dok druga verzija *FOAOFRB* koristi ortogonalne. Implementacija ovog klasifikatora nije pronađena već su rezultati preuzeti iz izvornog rada.

4. *BPMOGA* – objavljen u [20], *BPMOGA* (eng. *bi-phased multi-objective genetic algorithm based classifier*) koristi dva evolucijska modela za generiranje i optimiranje neizrazitih pravila. Prvi evolucijski model stvara pojedinačna pravila koja se u drugom evolucijskom modelu kombiniraju i optimiraju s ciljem poboljšanja klasifikatora kao cjeline. Implementacija ovog klasifikatora nije pronađena već su rezultati preuzeti iz izvornog rada.
5. *BTSKFC* – objavljen u [21], *BTSKFC* je kratica za Bayesov Takagi-Sugeno-Kang neizraziti klasifikator (eng. *Bayesian Takagi-Sugeno-Kang Fuzzy Classifier*). *BTSKFC* koristi neizrazita pravila koja svoje zaključke donose na temelju vektora parametara u konzekvent dijelu i pripadnosti primjera antecedent dijelu. Svi zaključci pravila se normaliziraju u jednu sumu iz intervala $[-1, 1]$. Zbog toga, za razliku od *IVTURS*-a, *BTSKFC* može klasificirati jedino podatke podijeljene u 2 klase. Implementacija ovog klasifikatora nije pronađena već su rezultati preuzeti iz izvornog rada.

4.1 Friedmanov test

Friedmanov test umjesto realnih vrijednosti podataka koje se žele analizirati koristi rangove između pojedinih skupova vrijednosti [14][15]. Takav pristup olakšava analizu podataka kojima je razdioba nepoznata, a da se razdioba mora pretpostaviti.

Test se provodi u četiri koraka [14]:

1. Podaci se organiziraju u tablicu u kojoj svaki red predstavlja blok podataka, a svaki stupac eksperimentalni tretman koji se proučava.
2. Podaci se rangiraju u svakom bloku s k rangova, pri čemu je k jednak broju stupaca.

3. Za svaki stupac računaju se prosječni rangovi. Srednja vrijednost prosječnih rangova je $\frac{1}{2}(k + 1)$, a varijanca $\frac{(k^2-1)}{(12 \cdot n)}$, gdje je n broj redaka po kojima se računa prosjek.
4. Hipoteza kako su svi eksperimentalni tretmani iz jednake razdiobe, tj. da su statistički jednaki provjerava se računajući χ_r^2 (31)[14][15], gdje je \bar{r}_j prosječni rang stupca j :

$$\chi_r^2 = \frac{12 \cdot n}{k \cdot (k+1)} \left(\sum_{j=1}^k \bar{r}_j^2 - \frac{k \cdot (k+1)^2}{4} \right) \quad (31)$$

Očekivana razdioba prosječnih rangova χ_r^2 je hi-kvadratna razdioba χ_{k-1}^2 (eng. *chi-squared distribution*) s $k - 1$ stupnjeva slobode. U slučaju da je vjerojatnost za opažanjem vrijednosti jednake ili veće od χ_r^2 , uzimajući u obzir χ_{k-1}^2 razdioba, manja od nekog nivoa značajnosti α , hipoteza se može odbaciti.

Kod uspoređivanja klasifikatora stupci su različiti klasifikatori, a redovi skupovi podataka. Kada neki klasifikatori imaju jednaki rezultat u redu, njihov rang se postavlja kao prosjek rangova koje bi trebali pokriti. z-vrijednost usporedbe dvaju klasifikatora i i j za Friedmanov test je (32)[15]:

$$z = \frac{(\bar{r}_i - \bar{r}_j)}{\sqrt{\frac{k \cdot (k+1)}{6 \cdot n}}} \quad (32)$$

4.2 Holmov test

Friedmanov test jedino ukazuje postoje li razlike između eksperimentalnih pristupa, ali ne otkriva koji pristup je statistički bolji od ostalih. Stoga, kada Friedmanov test ukaže na postojanje razlika, Holmov test će se koristiti kako bi utvrdili koji pristupi su različiti. Test se provodi na sljedeći način [15]:

1. Odabire se kontrolni eksperimentalni pristup.
2. Računaju se z vrijednosti i p -vrijednosti između kontrolnog te ostalih pristupa. p -vrijednosti se uzimaju iz normalne razdiobe.
3. p -vrijednosti se postavljaju u redosljed od najmanje do najveće $p_1 < p_2 < \dots < p_{k-1}$ i uz njih se vežu hipoteze $H_1 < H_2 < \dots < H_{k-1}$. Svaka hipoteza

određuje kako su i -ti pristup i svi iza njega statistički jednaki kontrolnom pristupu.

4. Krenuvši od najmanje, uspoređuju se p -vrijednosti s korigiranom α_i vrijednosti (33):

$$\alpha_i = \frac{\alpha}{(k-i)} \quad (33)$$

U slučaju kada je p_i -vrijednost manja od α_i vrijednosti, H_i hipoteza se odbacuje i prelazi se na sljedeće vrijednosti. Kada se jedna od hipoteza ne može odbaciti postupak se zaustavlja i prihvaća se hipoteza koja se ne može odbaciti te sve hipoteze koje slijede iza nje.

4.3 Test Wilcoxonovih rangova

Wilcoxonov test koristi se za usporedbu eksperimentalnih pristupa u parovima. Test se sastoji od rangiranja razlika između dva pristupa, neovisno o predznaku. Provodi se na sljedeći način [15]:

1. Računaju se razlike između dvaju pristupa.
2. Razlike se rangiraju s obzirom na njihovu apsolutnu vrijednost.
3. Sumiraju se rangovi u dvije sume. R^+ suma sadrži sve rangove pozitivnih razlika dok R^- sadrži sve rangove negativnih razlika.
4. Računa se statistika T (34):

$$T = \min(R^-, R^+) \quad (34)$$

5. Hipoteza kako je razlika između pristupa uzrok slučajnosti može se odbaciti na sljedeća dva načina [15]:
 - a. Za manje n -ove potrebno je pronaći kritične vrijednosti u statističkim tablicama i usporediti ih s T . Ako je T manji ili jednak kritičnoj vrijednosti hipoteza se odbacuje.
 - b. Za veće n -ove čije se kritične vrijednosti ne mogu pronaći u statističkim tablicama T prati normalnu razdioba i hipoteza se može odbaciti računajući z -vrijednost (35) ili p -vrijednost:

$$Z = \frac{T - \frac{1}{4} \cdot n \cdot (n+1)}{\sqrt{\frac{n(n+1)(2n+1)}{24}}} \quad (35)$$

4.4 Postavke klasifikatora *IVTURS*

Hiperparametri koji se koriste kod testiranja *IVTURS*-a jednaki su onima u originalnom radu [1]:

1. *Conjunction operator* – kod obje implementacije se koristi intervalno množenje
2. *Combination operator* – kod obje implementacije se koristi intervalno množenje
3. *Aggregation operator* – kod obje implementacije se koristi intervalno zbrajanje
4. *Decision function* – kod obje implementacije se koristi totalni poredak
5. Broj funkcija pripadnosti – kod obje implementacije se koristi pet funkcija pripadnosti kako bi opisali svaki atribut
6. Oblici funkcija pripadnosti – kod obje implementacije oblik je trokutasti
7. Minimalna potpora – minimalna potpora kod obje implementacije je 0,05
8. Maksimalna pouzdanost – maksimalna pouzdanost kod obje implementacije je 0,8
9. Maksimalna dubina stabla – maksimalna dubina kod obje implementacije je 3
10. Faktor w – kod obje implementacije faktor w je postavljen na 0,25
11. Broj prekrivanja primjera podataka, k – k je kod obje implementacije 2
12. Broj bitova za kodiranje parametara – broj bitova za kodiranje kod obje implementacije je 30
13. Veličina populacije modela *CHC* – kod obje implementacije veličina populacije je 50
14. Maksimalni broj evaluacija modela *CHC* – kod obje implementacije maksimalni broj je 20000
15. Parametar α – parametar α za *PBX- α* križanje kod obje implementacije je 1,0
16. Vrijednost za promjenu parametra L – reimplementacija koristi 10% inicijalne vrijednost parametra L , dok *KEEL*-ova implementacija koristi konstantu 1,0

Reimplementacija i implementacija u *KEEL*-u imaju vrlo slične postavke hiperparametara te vrlo sličan, ali ne i isti način izračuna pojedinih dijelova klasifikatora.

4.5 Usporedba s *FURIA*-om, *C4.5*, radom o *IVTURS*-u i *KEEL*-om

Ova usporedba sastoji se od originalnih rezultata navedenih u radu o *IVTURS*-u, reimplementacije napisane za ovaj rad, implementacije *IVTURS*-a u *KEEL*-u, implementacije *FURIA*-je u *KEEL*-u i implementacije pravila *C4.5* iz *KEEL*-a.

Usporedba se provodi na petnaest skupova podataka i prikazana je u tablici 4.2. Način ispitivanja preuzet je iz originalnog rada o *IVTURS*-u i provodi se jedna peterostruka unakrsna validacija. Reimplementacija *IVTURS*-a označena je s *RIV*, *FURIA* je označena s *FUR*, *KEEL*-ova implementacija je označena s *KIV*, pravila *C4.5* su označena s *C4.5*, a podaci iz rada su označeni s *RAD*.

Tablica 4.2 Prikaz točnosti prve usporedbe

Skup	<i>RAD</i>		<i>KIV</i>		<i>FUR</i>		<i>C4.5</i>		<i>RIV</i>	
	Test	Train	Test	Train	Test	Train	Test	Train	Test	Train
AUS	0,8580	0,9004	0,8507	0,9036	0,8521	0,8894	0,8579	0,8804	0,8348	0,8928
BAL	0,8576	0,9184	0,8512	0,9196	0,8191	0,8836	0,8064	0,8868	0,8608	0,9080
CLE	0,5960	0,8544	0,5896	0,8459	0,5757	0,6195	0,5489	0,6557	0,5967	0,7806
CRX	0,8714	0,9142	0,8621	0,9123	0,862	0,892	0,8345	0,879	0,8519	0,9092
ECO	0,7858	0,8906	0,7977	0,8913	0,8036	0,904	0,7648	0,8512	0,8294	0,8813
GER	0,7310	0,8535	0,709	0,8534	0,721	0,7672	0,693	0,7472	0,7090	0,8238
HAB	0,7285	0,8072	0,7548	0,8071	0,735	0,772	0,7054	0,7679	0,7903	0,8041
H-R	0,8023	0,9128	0,7937	0,9218	0,8125	0,8953	0,8	0,9	0,8000	0,9203
HEA	0,8815	0,9361	0,8518	0,9388	0,7925	0,8898	0,8074	0,8703	0,8185	0,9306
IRI	0,9600	0,9817	0,96	0,985	0,9533	0,9833	0,96	0,9766	0,9600	0,9800
N-T	0,9535	0,9884	0,9534	0,9813	0,9441	0,9918	0,9162	0,9593	0,9535	0,9826
SAH	0,7099	0,7955	0,7143	0,8003	0,7229	0,7559	0,7055	0,7656	0,6925	0,7973
TAE	0,5034	0,7318	0,5501	0,7251	0,5027	0,5082	0,4498	0,5894	0,5032	0,7433
WIN	0,9719	0,9930	0,9604	0,9971	0,9158	0,9929	0,9326	0,9887	0,9500	0,9986
WIS	0,9649	0,9850	0,9663	0,9846	0,9619	0,9893	0,956	0,9784	0,9650	0,9828
Prosjek	0,8117	0,8975	0,8110	0,8978	0,7982	0,8489	0,7826	0,8464	0,8077	0,8890

Friedmanovi rangovi za ovu usporedbu prikazani su u tablici 4.3.

Tablica 4.3 Friedmanovi rangovi za prvu usporedbu

	<i>RAD</i>	<i>KIV</i>	<i>FUR</i>	<i>C4.5</i>	<i>RIV</i>
Skup	Test	Test	Test	Test	Test
AUS	1	4	3	2	5
BAL	2	3	4	5	1
CLE	2	3	4	5	1
CRX	1	2	3	5	4
ECO	4	3	2	5	1
GER	1	3,5	2	5	3,5
HAB	4	2	3	5	1
H-R	2	5	1	3,5	3,5
HEA	1	2	5	4	3
IRI	2,5	2,5	5	2,5	2,5
N-T	1,5	3	4	5	1,5
SAH	3	2	1	4	5
TAE	2	1	4	5	3
WIN	1	2	5	4	3
WIS	3	1	4	5	2
Prosjek	2,06	2,6	3,33	4,33	2,66

Za dane rangove, vrijednost χ_r^2 je jednaka 18,1866. p-vrijednost za takav rezultat hi-razdiobe χ_4^2 s četiri stupnja slobode je 0,001135. Može se zaključiti kako je $p < \alpha$, čime se dokazuje kako ovih pet pristupa nisu statistički jednaki. Za daljnju usporedbu klasifikatora koristit će se Holmov test prikazan u tablici 4.4. Kontrolni klasifikator je *RAD*.

Tablica 4.4 Prikaz Holmovog testa

	i	z_i	p_i	α_i	H_i
<i>C4.5</i>	1	3,9259	0,000086	0,0125	Odbačena
<i>FUR</i>	2	2,193	0,028307	0,01667	Nije odbačena
<i>RIV</i>	3	1,0392	0,274306	0,025	Nije odbačena
<i>KIV</i>	4	0,9237	0,35564	0,05	Nije odbačena

S ovakvim rezultatima Holmov test pronalazi razlike između *C4.5* i ostalih, ali rezultati *FURIA*-e, reimplementacije, *KEEL*-ove implementacije i izvornog rada su statistički

jednaki. Ovakav zaključak je suprotan onom iz izvornog *IVTURS*-ovog rada gdje je zaključeno kako je *IVTURS* bolji od *FURIA*-e.

4.6 Sustav u ovisnosti o hiperparametru k

U ovom potpoglavlju ispitat će se ovisnost sustava o hiperparametru k - dubini pretraživačkog stabla kod pretraživanja čestih kombinacija. Isprobane su četiri vrijednosti hiperparametra k - 3, 4, 5, 6. Povećanjem hiperparametra k omogućuje se generiranje pravila koja imaju više kombinacija funkcija pripadnosti u antecedent dijelu pravila. Očekuje se kako će se povećanjem generirati specifičnija pravila što za sobom vuče i više pravila u sustavu. Takav pristup može dovesti do prenaučenosti, gdje pravila previše oponašaju podatke za učenje. Korištenjem potpore (16) i pouzdanosti (17) ograničava se prenaučenosť te se preferiraju općenita pravila. U tablici 4.5 prikazan je prosječan broj pravila i prosječna veličina svakog pravila za gore navedene vrijednosti hiperparametra k . Iako se s povećanjem hiperparametra k povećava i broj pravila i veličina pravila, pojedini skupovi koji se mogu objasniti manjim brojem pravila ne uče nepotrebna pravila te uče konstantan broj pravila.

Tablica 4.5 Prikaz prosječnog broja pravila i prosječne veličine pravila

k	3		4		5		6	
	Pravila	Veličina	Pravila	Veličina	Pravila	Veličina	Pravila	Veličina
AUS	64,60	2,70	78,20	3,54	72,60	3,96	78,80	4,56
BAL	112,80	1,84	107,60	1,86	110,20	1,84	111,00	1,85
CLE	79,40	2,87	81,40	3,66	87,00	4,36	93,60	4,73
CRX	76,00	2,66	78,40	3,42	81,20	3,96	82,20	4,38
ECO	52,80	2,48	50,60	2,88	55,60	3,01	54,60	3,06
GER	155,20	2,74	156,20	3,43	160,00	4,00	170,60	4,43
HAB	32,80	1,74	32,20	1,80	30,40	1,77	29,40	1,74
H-R	46,80	2,12	45,20	2,01	45,20	1,98	45,60	2,04
HEA	44,80	2,66	47,40	3,09	45,60	3,44	46,80	3,60
IRI	7,00	1,00	7,40	1,07	7,20	1,03	7,40	1,05
N-T	16,40	1,76	17,00	1,79	16,20	1,77	17,80	1,96
SAH	55,20	2,59	53,80	2,86	56,60	2,89	53,60	2,82
TAE	44,80	2,57	45,20	2,67	46,20	2,80	42,80	2,66
WIN	16,40	1,81	15,80	1,83	15,20	1,76	16,40	1,72

WIS	32,60	1,36	32,00	1,35	32,80	1,40	31,40	1,36
Prosjek	55,84	2,19	56,56	2,48	57,47	2,66	58,80	2,80

U tablici 4.6 prikazane su točnosti za sustave s gore navedenim vrijednostima hiperparametra k korištenjem predložene reimplementacije *IVTURS-a*.

Tablica 4.6 Prikaz točnosti u ovisnosti o hiperparametru k

k	3		4		5		6	
Skup	Test	Train	Test	Train	Test	Train	Test	Train
AUS	0,8609	0,8873	0,8536	0,9033	0,86	0,90	0,82	0,91
BAL	0,8720	0,9124	0,8656	0,9108	0,86	0,91	0,87	0,91
CLE	0,5667	0,8169	0,5267	0,8565	0,59	0,85	0,52	0,85
CRX	0,8595	0,8969	0,8687	0,8985	0,85	0,91	0,84	0,90
ECO	0,8088	0,8896	0,7412	0,8888	0,79	0,89	0,75	0,88
GER	0,7460	0,8265	0,7260	0,8420	0,71	0,84	0,71	0,84
HAB	0,7613	0,8180	0,7355	0,8238	0,74	0,82	0,72	0,82
H-R	0,8000	0,9156	0,7875	0,9141	0,79	0,92	0,75	0,92
HEA	0,8000	0,9278	0,8519	0,9296	0,83	0,93	0,81	0,93
IRI	0,9400	0,9850	0,9733	0,9783	0,95	0,99	0,96	0,98
N-T	0,9442	0,9837	0,9442	0,9884	0,95	0,98	0,99	0,98
SAH	0,7140	0,7973	0,7011	0,8033	0,73	0,80	0,69	0,80
TAE	0,5097	0,7433	0,5355	0,7117	0,57	0,72	0,55	0,72
WIN	0,9056	0,9986	0,9556	0,9986	0,97	1,00	0,95	1,00
WIS	0,9620	0,9806	0,9547	0,9795	0,96	0,99	0,97	0,98
Prosjek	0,8034	0,892	0,8014	0,8951	0,8102	0,8956	0,7924	0,895

Za dane točnosti Friedmanovi rangovi mogu se vidjeti u tablici 4.7.

Tablica 4.7 Prikaz rangova za točnosti u ovisnosti o hiperparametru k

k	3	4	5	6
Skup	Test	Test	Test	Test
AUS	1	4	3	2
BAL	2	4	3	1
CLE	4	3	2	1
CRX	3	1	4	2
ECO	4	1	2	3
GER	1	2	4	3
HAB	1	4	3	2
H-R	3	4	2	1

HEA	4	2	1	3
IRI	4	3	2	1
N-T	3	2	1	4
SAH	4	2	3	1
TAE	3	1	2	4
WIN	1	2	4	3
WIS	3	3	1	3
Prosjek	2,73	2,53	2,47	2,27

Za rangove u ovisnosti o hiperparametru k vrijednost χ_r^2 je jednaka 3,98. p -vrijednost za takav rezultat hi-razdiobe χ_3^2 s tri stupnja slobode je 0,2636. Takvim rezultatom dolazi se do zaključka kako su svi sustavi statistički jednaki i ne isplati se povećati hiperparametar k s vrijednosti 3 na vrijednost 6.

4.7 Choquetovi integrali

Usporedit će se dvije verzije koje koriste Choquetove integrale s reimplementacijom *IVTURS*-a. Njihove točnosti prikazane su u tablici 4.8. Verzija koja koristi integrale i *IVREF* funkcije označena je s *QIV*, dok je verzija koja koristi samo integrale označena s *QFRM*.

Tablica 4.8 Točnosti za RIV, QIV i QFRM

	RIV		QIV		QFRM	
AUS	0,8348	0,8928	0,8507	0,8924	0,8246	0,8587
BAL	0,8608	0,9080	0,8288	0,9164	0,8256	0,8956
CLE	0,5967	0,7806	0,5500	0,8262	0,5667	0,7527
CRX	0,8519	0,9092	0,8076	0,8989	0,8366	0,8678
ECO	0,8294	0,8813	0,8059	0,9030	0,7647	0,8590
GER	0,7090	0,8238	0,7300	0,8405	0,7110	0,7925
HAB	0,7903	0,8041	0,7452	0,8213	0,7452	0,7934
H-R	0,8000	0,9203	0,8625	0,9141	0,7438	0,9250
HEA	0,8185	0,9306	0,8481	0,9343	0,8259	0,9176
IRI	0,9600	0,9800	0,9800	0,9767	0,9667	0,9817
N-T	0,9535	0,9826	0,9256	0,9849	0,8930	0,9709
SAH	0,6925	0,7973	0,7161	0,8136	0,7097	0,7724
TAE	0,5032	0,7433	0,5484	0,7600	0,5032	0,6883
WIN	0,9500	0,9986	0,9278	1,0000	0,9444	0,9958
WIS	0,9650	0,9828	0,9562	0,9828	0,9474	0,9718
Prosjek	0,8077	0,8890	0,8055	0,8977	0,7872	0,8695

U tablici 4.9 prikazani su Friedmanovi rangovi za točnosti u tablici 4.9.

Tablica 4.9 Friedmanovi rangovi za RIV, QIV i QFRM

	<i>RIV</i>	<i>QIV</i>	<i>QFRM</i>
AUS	2	1	3
BAL	1	2	3
CLE	1	3	2
CRX	1	3	2
ECO	1	2	3
GER	3	1	2
HAB	1	2,5	2,5
H-R	2	1	3
HEA	3	1	2
IRI	3	1	2
N-T	1	2	3
SAH	3	1	2
TAE	2,5	1	2,5
WIN	1	3	2
WIS	1	2	3
Prosjek	1,7666	1,7666	2,466

Za navedene klasifikatore vrijednost χ_r^2 je jednaka 4,9. p -vrijednost za takav rezultat hi-razdiobe χ_2^2 s dva stupnja slobode je 0,08629. p -vrijednost nije manja od α stoga se može zaključiti kako su klasifikatori statistički jednaki.

4.8 Usporedba s klasifikatorom FOATFRB

S obzirom na to da nije pronađen kod implementacije algoritma FOATFRB, usporedba s njim provesti će se na podacima koji su objavljeni u njegovom izvornom radu [19]. U [19], FOATFRB je ispitan na 27 skupova podataka. U ovom radu uspoređuje se s 26 skupova jer nije pronađen jedan od skupova. Skupovi podataka, točnosti, razlike (rezultati drugog klasifikatora oduzeti od prvog) i rangovi za RIV te FOATFRB prikazani su u tablici 4.10.

Tablica 4.10 Wilcoxonov test za FOATFRB i RIV

Skupovi	FOATFRB	RIV	Razlika	Rangovi
AUS	0,8824	0,8652	-0,0172	-12,00
BAL	0,8736	0,8508	-0,0228	-15,00
CLE	0,6955	0,5500	-0,1455	-26,00
CON	0,5390	0,5277	-0,0113	-9,00
CRX	0,9170	0,8348	-0,0822	-23,00
DER	0,9489	0,9556	0,0067	5,00
ECO	0,8120	0,7912	-0,0208	-14,00
GER	0,7380	0,7510	0,0130	10,00
HAB	0,7621	0,7548	-0,0073	-6,00
H-R	0,8342	0,7625	-0,0717	-22,00
HEA	0,9050	0,8519	-0,0531	-21,00
ION	0,9071	0,9083	0,0012	1,00
IRI	0,9667	0,9800	0,0133	11,00
MAG	0,8033	0,8073	0,0040	3,00
N-T	0,9588	0,9545	-0,0043	-4,00
PEN	0,9009	0,9201	0,0192	13,00
PIM	0,7914	0,7442	-0,0472	-20,00
SAH	0,7145	0,7255	0,0110	8,00
SPE	0,8920	0,7926	-0,0994	-24,00
TAE	0,5237	0,5563	0,0326	18,00
TIT	0,8006	0,7765	-0,0241	-16,00
TWO	0,9585	0,9499	-0,0086	-7,00
VEH	0,7320	0,6318	-0,1002	-25,00
WIN	0,9719	0,9444	-0,0275	-17,00
WIR	0,6173	0,5806	-0,0367	-19,00
WIS	0,9649	0,9623	-0,0026	-2,00
Prosjek	0,8235	0,7973		

Suma R^+ je 69, a suma R^- je 282 stoga je T 69. n je jednak 26, nivo značajnosti α je 0,05. Kritična vrijednost $T_{\alpha,n}$ je 98. Budući da je T manji od kritične vrijednosti, nulta hipoteza se može odbaciti. RIV i $FOATFRB$ nisu statistički jednaki, već je $FOATFRB$ statistički bolji. Ovakav zaključak jednak je onom u [19]. Autori u svom radu ne navode izvor rezultata za $IVTURS$ ali su oni jednaki onim iz izvornog rada.

4.9 Usporedba s klasifikatorom *BPMOGA*

S obzirom na to da nije pronađen kod implementacije algoritma *BPMOGA*, usporedba s njim provesti će se na podacima koji su objavljeni u njegovom izvornom radu [20]. U [20], *BPMOGA* je ispitan na 21 skupu podataka. U ovom radu uspoređuje se sa 16

skupova jer BPMOGA može raditi sa skupovima kod kojih nedostaju pojedine vrijednosti. Kako *IVTURS* to ne može, skupovi s nedostajućim vrijednostima su izbačeni. Skupovi podataka, točnosti, razlike (rezultati drugog klasifikatora oduzeti od prvog) i rangovi za *RIV* te *BPMOGA* prikazani su u tablici 4.11.

Tablica 4.11 Wilcoxonov test za *BPMOGA*-u i *RIV*

Skupovi	<i>BPMOGA</i>	<i>RIV</i>	Razlika	Rangovi
CRX	0,8564	0,85741	0,0010	1
PIM	0,7414	0,75959	0,0182	8
ECO	0,7850	0,7947	0,0097	4
FLA	0,7489	0,73101	-0,0179	-7
GLA	0,7004	0,6759	-0,0245	-10
HAB	0,7448	0,7616	0,0168	6
IRI	0,9573	0,9633	0,0060	3
LED	0,7012	0,7274	0,0262	11
MON	0,8929	0,9886	0,0957	16
N-T	0,9372	0,9428	0,0055	2
SON	0,7232	0,7991	0,0759	15
VEH	0,6812	0,6329	-0,0483	-14
VOW	0,6133	0,5893	-0,0240	-9
WIN	0,9423	0,95446	0,0122	5
YEA	0,5851	0,5426	-0,0425	-13
ZOO	0,9532	0,9227	-0,0305	-12
Prosjek	0,7852	0,7902		

Suma R^+ je 71, a suma R^- je 65, pa je T 65. n je jednak 16, nivo značajnosti α je 0,05. Kritična vrijednost $T_{\alpha,n}$ je 29. Budući da T nije manji od kritične vrijednosti, nulta hipoteza se ne može odbaciti. *RIV* i *BPMOGA* su statistički jednaki te je njihova razlika rezultat slučajnosti..

4.10 Usporedba s klasifikatorom *BTSKFC*

S obzirom na to da nije pronađen kod implementacije algoritma *BTSKFC*, usporedba s njim provesti će se na podacima koji su objavljeni u njegovom izvornom radu [21]. U [21], *BTSKFC* je ispitan na 20 skupova podataka. U ovom radu uspoređuje se sa 17 skupova jer nisu pronađena tri skupa. Skupovi podataka, točnosti, razlike (rezultati

drugog klasifikatora oduzeti od prvog) i rangovi za *RIV* te *BTSKFC* prikazani su u tablici 4.12.

Tablica 4.12 Wilcoxonov test za *BTSKFC* i *RIV*

Skupovi	<i>BTSKFC</i>	<i>RIV</i>	Razlika	Rangovi
AUS	0,8688	0,85507	-0,0137	-8
BAL2	0,9505	0,9395	-0,0110	-7
BRE	0,9600	0,9623	0,0023	3
FOU	0,9770	0,8775	-0,0995	-17
GLA2	0,9545	0,9364	-0,0181	-9
HAB	0,7587	0,7397	-0,0190	-10
HEA	0,8562	0,8463	-0,0099	-6
MAM	0,8313	0,8011	-0,0302	-13
N-T	0,9545	0,9496	-0,0050	-4
PIM	0,7722	0,7529	-0,0193	-11
RIN	0,9744	0,9067	-0,0677	-16
SAH	0,7021	0,7022	0,0001	1
TAE2	0,7005	0,6600	-0,0405	-14
VEH2	0,7823	0,7803	-0,0021	-2
WAV	0,9121	0,8689	-0,0432	-15
WIS	0,9721	0,9641	-0,0080	-5
YEA2	0,7452	0,7655	0,0203	12
Prosjeak	0,8631	0,8416		

Suma R^+ je 16, a suma R^- je 137 stoga je T 16. n je jednak 17, nivo značajnosti α je 0,05. Kritična vrijednost $T_{\alpha,n}$ je 34. T je manji od kritične vrijednosti, stoga se nulta hipoteza može odbaciti. *RIV* i *BTSKFC* nisu statistički jednaki te njihova razlika nije rezultat slučajnosti. *BTSKFC* je statistički bolji klasifikator.

Zaključak

Reimplementacija *IVTURS*-a pokazala se statistički jednakom kao izvorni *IVTURS*. S malo lošijim prosječnim rezultatima, ali i statistički jednako važnim rezultatima parira algoritmu u izvorno objavljenom radu. Nadalje, reimplementacija ima statistički jednake rezultate i brže radi od jedine dostupne verzija *IVTURS*-a iz *KEEL*-ovog repozitorija algoritama. Promjene uvedene s Choquetovima integralima nisu se pokazale kao dobra zamjena za originalnu agregacijsku funkciju. Sustav pomoću potpore i pouzdanosti ograničava prenaučenosť te omogućuje bolju generalizaciju. Utvrđeno je kako reimplementacija daje statistički lošije rezultate od *FOATFRB*-a i *BTSKFC*-a što je jednako zaključcima iz originalnih radova *FOATFRB*-a i *BTSKFC*-a koji zaključuju kako su njihovi rezultati statistički bolji od *IVTURS*-a. Također, pokazano je da *IVTURS* daje statistički jednake rezultate kao algoritam *BPMOGA*.

Literatura

- [1] Sanz J.A., Fernández A., Bustince H., Herrera F., *IVTURS: A Linguistic Fuzzy Rule-Based Classification System Based On a New Interval-Valued Fuzzy Reasoning Method With Tuning and Rule Selection*, IEEE Transactions on fuzzy systems, 21, 3 (2013), str. 399-411
- [2] Alcalá-Fdez J., Alcalá R., Herrera F., *A Fuzzy Association Rule-Based Classification Model for High-Dimensional Problems with Genetic Rule Selection and Lateral Tuning*, IEEE Transactions on fuzzy systems, 19, 5 (2011), str. 857-872
- [3] Cordon O., Jesus M.J.d., Herrera F., *A proposal on reasoning methods in fuzzy rule-based classification systems*, International Journal of Approximate Reasoning, 20 (1999), str. 21-45
- [4] Xu Z.S., Yager R.R., *Some geometric aggregation operators and their relationship to t-norms in interval-valued fuzzy sets*, International Journal of General Systems, 35 (2006), 4, str. 417-433
- [5] Alcalá-Fdez J., Fernandez A., Luengo J., Derrac J., García S., Sánchez L., Herrera F., *KEEL Data-Mining Software Tool: Data Set Repository, Integration of Algorithms and Experimental Analysis Framework*, Journal of Multiple-Valued Logic and Soft Computing, 17, 2-3 (2011), str. 417-433
- [6] KEEL-dataset repository, (2004), poveznica:
<https://sci2s.ugr.es/keel/datasets.php>, pristupljeno: 15. ožujka 2020.
- [7] Hühn J., Hüllermeier E., *FURIA: An algorithm for unordered fuzzy rule induction*, Data Mining and Knowledge Discovery, 19, 3 (2009), str. 293–319
- [8] Cohen W. W., *Fast effective rule induction*, Machine Learning Proceedings of the Twelfth International Conference, Tahoe City, (1995), str. 115 - 123
- [9] Čupić M., Dalbello Bašić B., Golub M., *Neizrazito, evolucijsko i neuroračunarstvo*, Zagreb 2013.
- [10] Mendel J.M., *Uncertain Rule-based Fuzzy Systems*, 2. izdanje, Los Angeles: Pearson 2017.

- [11] Deschrijver G., Cornelis C., Kerre E. E., *On the Representation of Intuitionistic Fuzzy t -Norms and t -Conorms*, IEEE Transactions on fuzzy systems, 12, 1 (2004), str. 45-61
- [12] Barrenechea E., Bustince H., Fernandez J., Pararnian D., Sanz J. A., *Using the Choquet Integral in the Fuzzy Reasoning Method of Fuzzy Rule-Based Classification Systems*, Axioms, 2, 2 (2013), str. 208-223
- [13] Dimuro G. P., Fernández J., Bedregal B., Mesiar R., Sanza J. A., Lucca G., Bustince H., *The state-of-art of the generalizations of the Choquet integral: From aggregation and pre-aggregation to ordered directionally monotone functions*, Information Fusion, 57 (2020), str. 27-43
- [14] Friedman M., *The Use of Ranks to Avoid the Assumption of Normality Implicit in the Analysis of Variance*, Journal of the American Statistical Association, 32 (1937), str. 675-701
- [15] Demšar J., *Statistical Comparisons of Classifiers over Multiple Data Sets*, Journal of Machine Learning Research, 7 (2006), str. 1-30
- [16] Holm S., *A Simple Sequentially Rejective Multiple Test Procedure*, Scandinavian Journal of Statistics, 6, 2 (1979), str. 65-70
- [17] Wilcoxon F., *Individual comparisons by ranking methods*, Biometrics, 1 (1945), str. 80-83
- [18] Quinlan J. R., *C4.5: programs for machine learning*, 2. izdanje, Elsevier 2014.
- [19] Mousavi S. M., Tavana M., Alikar N., Zandieh M., *A tuned hybrid intelligent fruit fly optimization algorithm for fuzzy rule generation and classification*, Neural Computing and Applications, 31 (2019), str. 873-885
- [20] Dutta D, Sil J., Dutta P., *A bi-phased multi-objective genetic algorithm based classifier*, Expert Systems With Applications, 146 (2020)
- [21] Gu X., Chung F.-L., Wang S., *Bayesian Takagi-Sugeno-Kang Fuzzy Classifier*, IEEE Transactions on fuzzy systems, 25, 6 (2017), str. 1655-1671
- [22] Lozano M., Herrera F., Krasnogor N., Molina D., *Real-Coded Memetic Algorithms with Crossover Hill-Climbing*, Evolutionary computation, 12, 3 (2004), str. 273-302

[23] Bustince H., Barrenechea E., Pagola M., *Restricted equivalence functions*,
Fuzzy Sets and Systems, 157, 17 (2006), str. 2333-2346

Naslov, sažetak i ključne riječi

Naslov: Implementacija učinkovitog neizrazitog klasifikacijskog sustava temeljenoga na algoritmu *IVTURS*

Ključne riječi: *IVTURS*, neizraziti klasifikacijski sustavi, klasifikacija, neizrazita logika, Choquetovi integrali, Friedmanov test, Holmov test, Test Wilcoxonovih rangova

Za rad je napisana implementacija neizrazitog klasifikacijskog sustava *IVTURS* u programskim jezicima python/cython. Prvo poglavlje rada obrađuje općenite teorijske pojmove vezane uz *IVTURS*. Drugo poglavlje rada obrađuje sam *IVTURS*, tj. njegovu metodu rasuđivanja, generiranje i optimiranje pravila. U drugom se poglavlju također nalazi prikaz kako sustav klasificira ulazne primjere. Treće poglavlje uvodi promjene u radu sustava u vidu korištenja Choquetovih integrala kao agregacijske funkcije. Za kraj se u četvrtom poglavlju ispituje rad sustava i uspoređuje s drugim klasifikatorima. Reimplementacija se uspoređuje s *FURIA*-om, *C4.5*, implementacijom *IVTURS*-a iz *KEEL*-a, rezultatima izvornog rada o *IVTURS*-u, *FOATFRB*-om, *BPMOGA*-om i *BTSKFC*-om. Također se uspoređuju dvije nove verzije koje su nastale korištenjem Choquetovih integrala i analizira se utjecaj hiperparametra dubine stabla petraživanja na prosječnu dužinu, specifičnost pravila te prenaučenosť klasifikatora. Rezultati ukazuju na reimplementaciju statistički jednaku izvornoj, uz veću brzinu implementacije u odnosu na *KEEL*-ovu, a predložene i uvedene promjene nisu poboljšale klasifikaciju.

Title, abstract and keywords

Title: Implementation of an Efficient Fuzzy Rule-Based Classification System Based on *IVTURS* Algorithm

Keywords: *IVTURS*, fuzzy rule-based classification systems, classification, fuzzy logic, Choquet integrals, Friedman test, Holm test, Wilcoxon signed ranks test

A fuzzy rule-based classification system implementation of *IVTURS* algorithm written in python / cython programming languages was achieved in this thesis. The first chapter deals with general theoretical concepts related to *IVTURS*. The second chapter deals with *IVTURS* itself, i.e., its reasoning method, generation and optimization of rules. The second chapter also provides an overview of how the system classifies input examples. The third chapter introduces changes in the system using Choquet integrals as an aggregation function. Finally, the fourth chapter examines the system and compares it with other classifiers. Reimplementation is compared with *FURIA*, *C4.5*, *IVTURS* implementation from *KEEL*, results presented in the original *IVTURS* paper, *FOATFRB*, *BPMOGA* and *BTSKFC* algorithms. We also compare two new versions created using Choquet integrals and analyze the influence of the search tree depth parameter on the average length, rule specificity, and classifier overfitting. The results indicate comparative performance with original implementation and changes that do not improve the classification process.

Privitak

Format ulaznih podataka

Format ulaznih podataka mora biti *KEEL*-ov format podataka, definiran u [5][6].

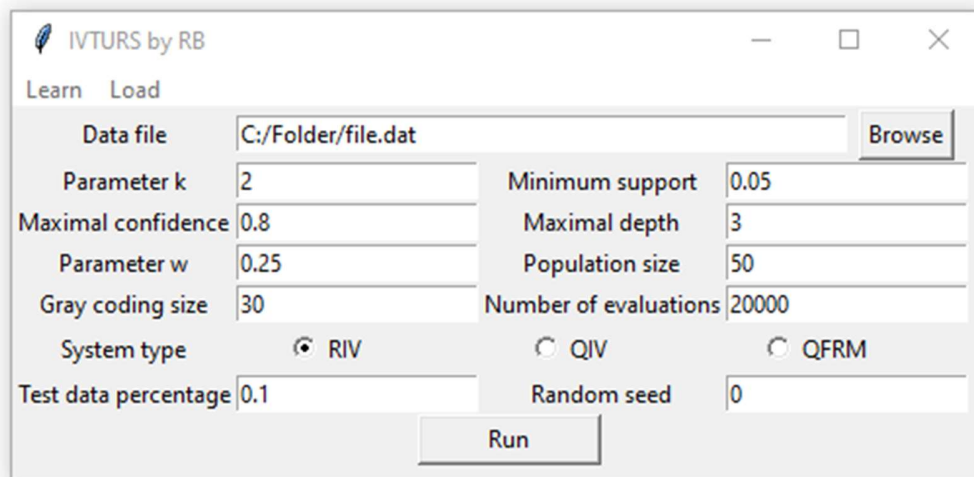
Format sadrži pet ključnih pojmova. Pojmovi se zapisuju na početak retka i određuju što redak predstavlja. Pojmovi su:

1. *@relation* – predstavlja redak u kojem se nalazi ime skupa podataka
2. *@attribute* – predstavlja redak u kojem se definira jedan od atributa. Klasa je također definirana kao atribut i treba biti dodana kao kategorički tip. Atribut je opisan imenom i tipom, gdje postoje tri tipa:
 - 1) Numerički atribut kontinuiranih vrijednosti – označen pojmom *real*, sadrži i interval vrijednosti koje atribut može poprimiti zapisan između uglatih zagrada.
 - 2) Numerički atribut diskretnih vrijednosti – označen pojmom *integer*, sadrži i interval vrijednosti koje atribut može poprimiti zapisan između uglatih zagrada.
 - 3) Kategorički atribut – nije označen pojmom, predstavljaju ga vitičaste zagrade u kojima se nalaze sve moguće vrijednosti atributa.
3. *@inputs* – predstavlja redak u kojem su nabrojani svi atributi, tj. ulazni atributi po kojima se određuju izlazni atribut - klasa.
4. *@outputs* – predstavlja red u kojem su izlazni atributi, u ovom slučaju atribut koji predstavlja klasu.
5. *@data* – predstavlja redak nakon kojeg su nabrojane instance podataka gdje se svaki podatak nalazi u zasebnom retku. Vrijednosti atributa za svaki primjerak moraju biti odijeljene zarezom i zapisane redoslijedom kojim su definirani ulazni atributi (*@inputs*).

U slučaju kada je bilo koje ime ili vrijednost, skup od više riječi ne treba koristiti razmak, već riječi treba odijeliti nekim drugim znakom, npr - ili _.

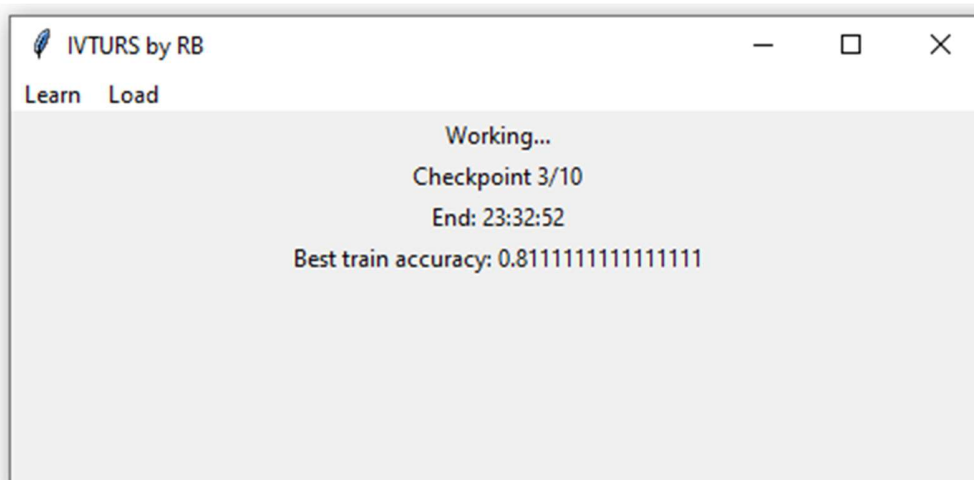
Grafičko sučelje

Grafičko sučelje sastoji se od dvije glavne funkcije. Prva, funkcija učenja („Learn“), ovisno o odabranim parametrima i ulaznim podacima uči novi klasifikator kojikorisnik ima priliku spremiti nakon što se učenje završi. Na slici P.1 prikazan je prozor odabira parametara i datoteka za učenje klasifikatora.

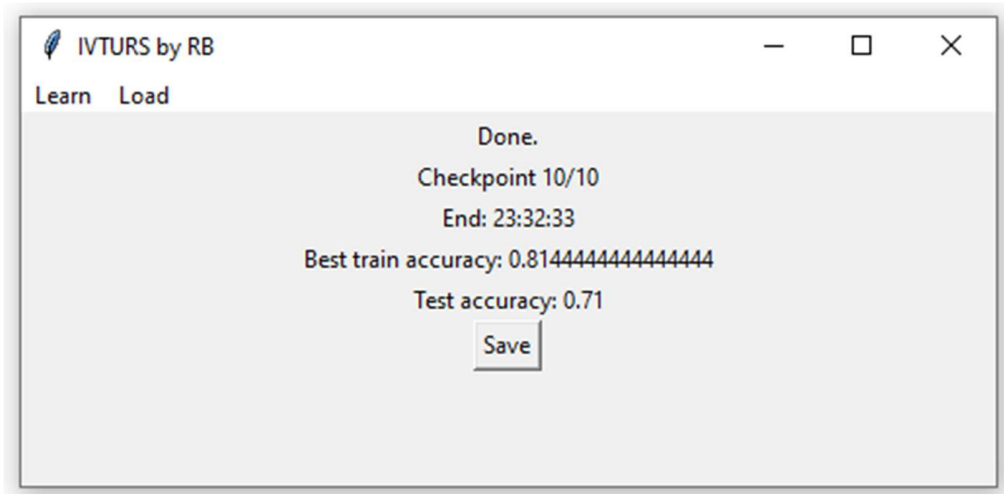


Slika P.1 Prikaz prozora za funkciju učenja

Na slikama P.2 i P.3 prikazane su informacije koje su prikazivane tijekom te nakon učenja klasifikatora.

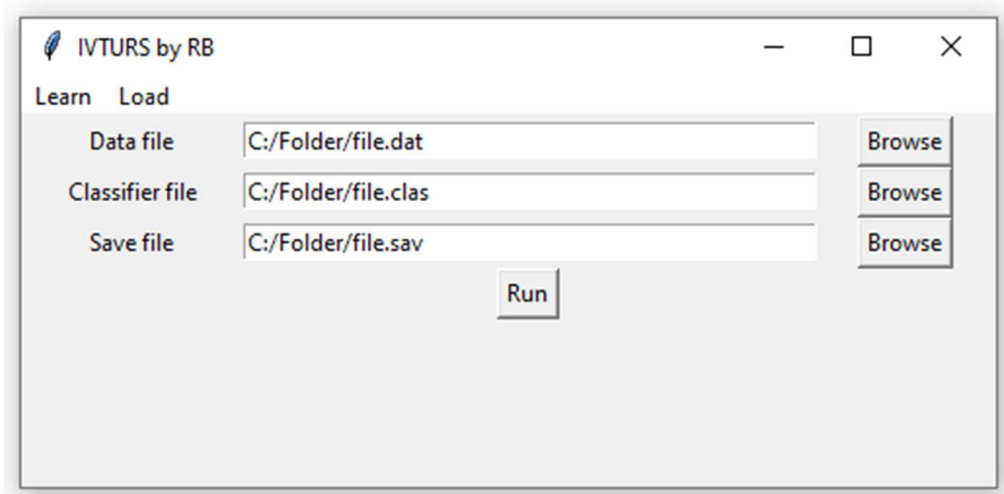


Slika P.2 Prikaz prozora tijekom učenja



Slika P.3 Prikaz prozora nakon učenja

Druga funkcija u grafičkom sučelju je učitavanje klasifikatora („Load“). Funkcija učitavanja prima podatke, klasifikator i datoteku u koje će se spremiti izlaznu datoteku klasa kojima pripadaju podaci. Na slici P.4 se nalazi prikaz prozora funkcije učitavanja.



Slika P.4 Prikaz prozora učitavanja