

# Modely rozhodovania pri určovaní diagnóz v medicíne

Ing. Tomasz Kanik

Školiteľ: doc. RNDr. Štefan Peško, CSc.

Pracovisko: KMMOA, FRI, ŽU

Študijný program: 9.2.9 Aplikovaná informatika

# Ciele práce

- ▶ identifikácia problémovej skupiny pacientov,
- ▶ zlepšenie kvality rozhodovacích modelov,
- ▶ zvýšenie celkovej úspešností diagnóz.

# Praktický problém

Pacient	Príznak A	Príznak B	Príznak C	Stav
P <sub>1</sub>	5	7	9	dobrý
P <sub>2</sub>	6	7	8	zlý
P <sub>3</sub>	4	7	4	dobrý
P <sub>4</sub>	7	3	4	priemerný
P <sub>5</sub>	9	2	4	priemerný
P <sub>6</sub>	8	9	3	priemerný
P <sub>7</sub>	1	5	8	dobrý
P <sub>8</sub>	1	4	9	zlý
P <sub>9</sub>	1	6	7	zlý
P <sub>10</sub>	6	7	8	priemerný

# Praktický problém

Pacient	Príznak A	Príznak B	Príznak C	Stav
P <sub>1</sub>	5	7	9	dobrý
P <sub>2</sub>	6	7	8	zlý
P <sub>3</sub>	4	7	4	dobrý
P <sub>4</sub>	7	3	4	priemerný
P <sub>5</sub>	9	2	4	priemerný
P <sub>6</sub>	8	9	3	priemerný
P <sub>7</sub>	1	5	8	dobrý
P <sub>8</sub>	1	4	9	zlý
P <sub>9</sub>	1	6	7	zlý
P <sub>10</sub>	6	7	8	priemerný

# Praktický problém

Pacient	Príznak A	Príznak B	Príznak C	Stav
P <sub>1</sub>	5	7	9	dobrý
P <sub>2</sub>	6	7	8	zlý
P <sub>3</sub>	4	7	4	dobrý
P <sub>4</sub>	7	3	4	priemerný
P <sub>5</sub>	9	2	4	priemerný
P <sub>6</sub>	5	5	3	priemerný
P <sub>7</sub>	1	5	8	dobrý
P <sub>8</sub>	1	4	9	zlý
P <sub>9</sub>	1	6	7	zlý
P <sub>10</sub>	6	7	8	priemerný

**Nekonzistencia údajov**

# Praktický problém

Pacient	Príznak A	Príznak B	Príznak C	Stav
P <sub>1</sub>	5	7	9	dobry
P <sub>2</sub>	6	7	8	zly
P <sub>3</sub>	4	7	4	dobry
P <sub>4</sub>	4	3	4	priemerny
P <sub>5</sub>	9	2	4	priemerny
P <sub>6</sub>	8	9	3	priemerny
		5	8	dobry
		4		
		6		
		7		

Zmena hodnoty podmienených atribútov môže spôsobiť zmenu stavu

Aj poradie hodnôt rozhodovacieho atribútu je významné pre lekára

# Zvolený prístup

- ▶ VC-DRSA:
  - používa rozhodovacie pravidla (teória RSA);
  - zohľadňuje poradie preferencie v rozhodovacích pravidlách (teória DRSA);
  - umožňuje relaxáciu striktných pravidiel DRSA pomocou miery konzistencie;
- ▶ Identifikácia nekonzistentných inštancií vedie k určení problémovej skupiny pacientov
- ▶ Modifikácie teórie Rough Set sú často aplikované v expertných systémoch (napr. WMSS)

# Medicína

Problémová skupina  
pacientov

Kvalita rozhodovacích  
modelov

Celková úspešnosť  
diagnóz



# Model

Nekonzistentné  
inštancie

Kvalita rozhodovacích  
pravidiel

Celková úspešnosť  
klasifikácie



# VDRSA – miera konzistencie

- Miera konzistencie objektu vyjadruje informáciu o jeho príslušnosti do zvolenej množiny rozhodnutí.
- Miery rozdeľujeme na maximalizačné a minimalizačné.
  - Pre maximalizačnú mieru konzistencie platí, že čím je hodnota vyššia tým viac konzistentný je skúmaný objekt.
  - Pre minimalizačnú mieru konzistencie platí opak.

# Základné pojmy VC-DRSA

Majme  $n$  tried rozkladu univerza  $U$ :  $X = \{X_1, X_2, \dots, X_n\}$ , potom

dolná aproximácia:

$$X_i^{\geq} = X_i \cup X_{i+1} \cup \dots \cup X_n$$

horná aproximácia:

$$X_i^{\leq} = X_1 \cup X_2 \cup \dots \cup X_i$$

negácia aproximácie:

$$\neg X_i^{\leq} = X - X_i^{\geq}, \quad \neg X_i^{\geq} = X - X_i^{\leq}$$

pozitívny kužeľ dominácie:

$$D_P^+(x) = \{y \in U : y D_P x\}$$

negatívny kužeľ dominácie:

$$D_P^-(x) = \{y \in U : x D_P y\}$$

kde

$D_P$  je relácia dominácie na množine podmienených atribútov  $P$ ,  
 $x$  je objekt rozhodovania ( $x \in U$ )

# Známe miery konzistencie

- ▶ Inuiguchi et al., 2009

- $$\beta_{X \geq}^P(y) = \frac{|D_P^-(y) \cap X \geq|}{|D_P^-(y) \cap X \geq| + |D_P^+(y) \cap \neg X \geq|},$$

$$\beta_{X \leq}^P(y) = \frac{|D_P^+(y) \cap X \leq|}{|D_P^+(y) \cap X \leq| + |D_P^-(y) \cap \neg X \leq|}.$$

- $$\mu_{X \geq}^P(y) = \frac{|D_P^+(y) \cap X \geq|}{|D_P^+(y)|}, \quad \mu_{X \leq}^P(y) = \frac{|D_P^-(y) \cap X \leq|}{|D_P^-(y)|}.$$

# Známe miery konzistencie

- ▶ Błaszczński et al., 2007 & 2009

- $\epsilon_{X \geq}^P(y) = \frac{|D_P^+(y) \cap \neg X \geq|}{|\neg X \geq|}, \quad \epsilon_{X \leq}^P(y) = \frac{|D_P^-(y) \cap \neg X \leq|}{|\neg X \leq|}$

- $\epsilon'_{X \geq}{}^P(y) = \frac{|D_P^+(y) \cap \neg X \geq|}{|X \geq|}, \quad \epsilon'_{X \leq}{}^P(y) = \frac{|D_P^-(y) \cap \neg X \leq|}{|X \leq|}$

# Príklad výpočtu konzistencie

$q_1, q_2, q_3$  – podmienené atribúty

$d$  – rozhodovací atribút

$X_1, X_2, X_3$  – triedy rozkladu

objekt	$q_1$	$q_2$	$q_3$	$d$	trieda
$y_1$	5	7	9	3	$X_3$
$y_2$	6	7	8	1	$X_1$
$y_3$	4	7	4	3	$X_3$
$y_4$	7	3	4	2	$X_2$
$y_5$	9	2	4	2	$X_2$
$y_6$	8	9	3	2	$X_2$
$y_7$	1	5	8	3	$X_3$
$y_8$	1	4	9	1	$X_1$
$y_9$	1	6	7	1	$X_1$

# Príklad výpočtu konzistencie

objekt	$q_1$	$q_2$	$q_3$	d	trieda
$y_1$	5	7	9	3	$X_3$
$y_2$	6	7	8	1	$X_1$
$y_3$	4	7	4	3	$X_3$
$y_4$	7	3	4	2	$X_2$
$y_5$	9	2	4	2	$X_2$
$y_6$	8	9	3	2	$X_2$
$y_7$	1	5	8	3	$X_3$
$y_8$	1	4	9	1	$X_1$
$y_9$	1	6	7	1	$X_1$

$$X_1^{\leq} = \{y_2, y_8, y_9\}$$

$$X_2^{\geq} = \neg X_1^{\leq} = \{y_1, y_3, y_4, y_5, y_6, y_7\}$$

$$D_P^+(y_2) = \{y_2\}$$

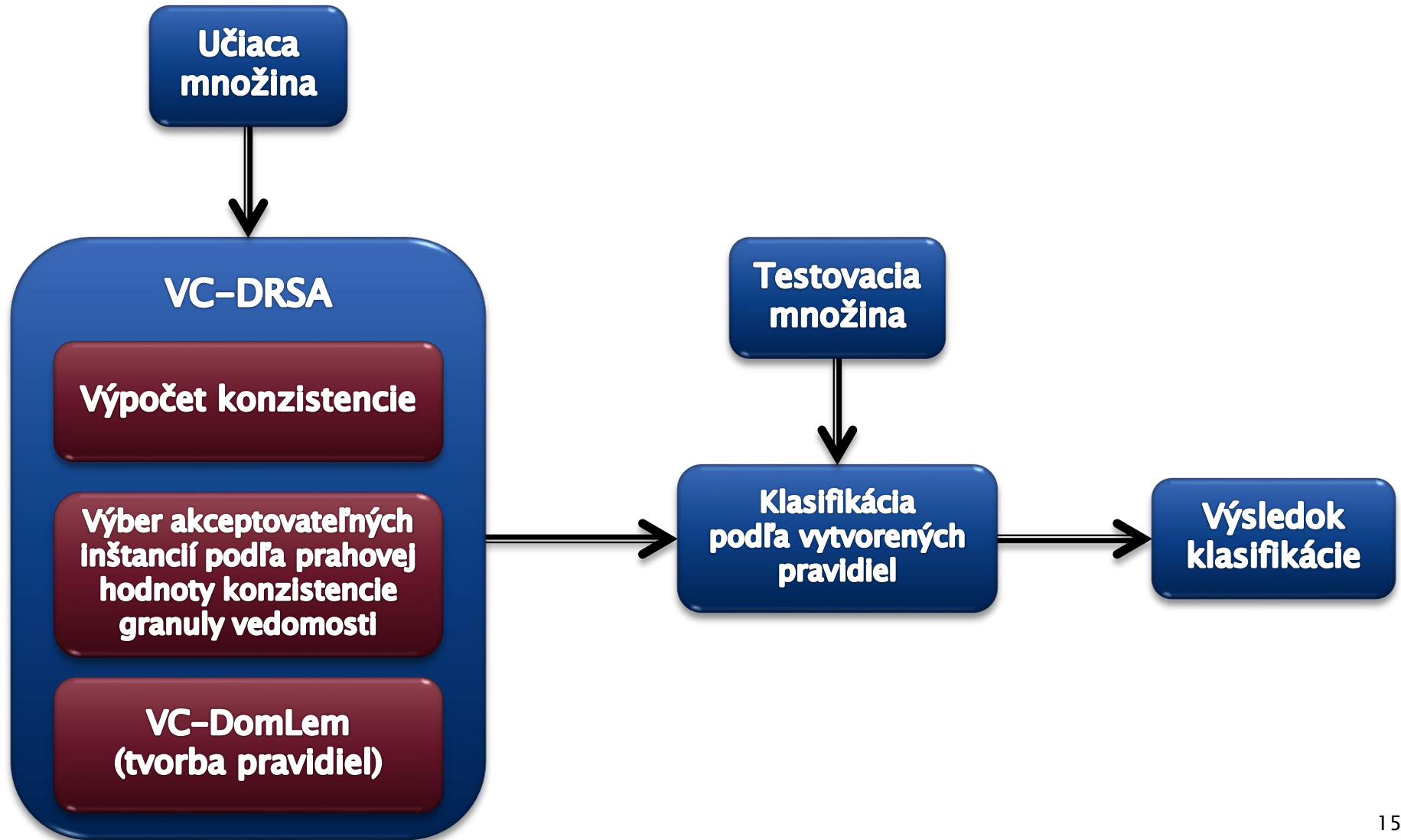
$$D_P^-(y_2) = \{y_2, y_3, y_7, y_9\}$$

---


$$\beta_{X_1^{\leq}}^P(y_2) = \frac{1}{1+2} = \frac{1}{3} \quad \mu_{X_1^{\leq}}^P(y_2) = \frac{1}{4}$$

$$\epsilon_{X_1^{\leq}}^P(y_2) = \frac{2}{6} = \frac{1}{3} \quad \epsilon'_{X_1^{\leq}}^P(y_2) = \frac{2}{3}$$

# Schéma aplikácie VC-DRSA



# Miera neistoty

- ▶ Autori Deng et al., 2011
- ▶ Definovali kvantitatívne vyjadrenie globálnej nekonzistencie objektu
- ▶ Miera neistoty objektu vyjadruje do akej miery objekt patri vyššej alebo nižšej triedy rozkladu podľa poradia preferencie tried



# Známé miery neistoty

- ▶ Deng et al., 2011 & 2013

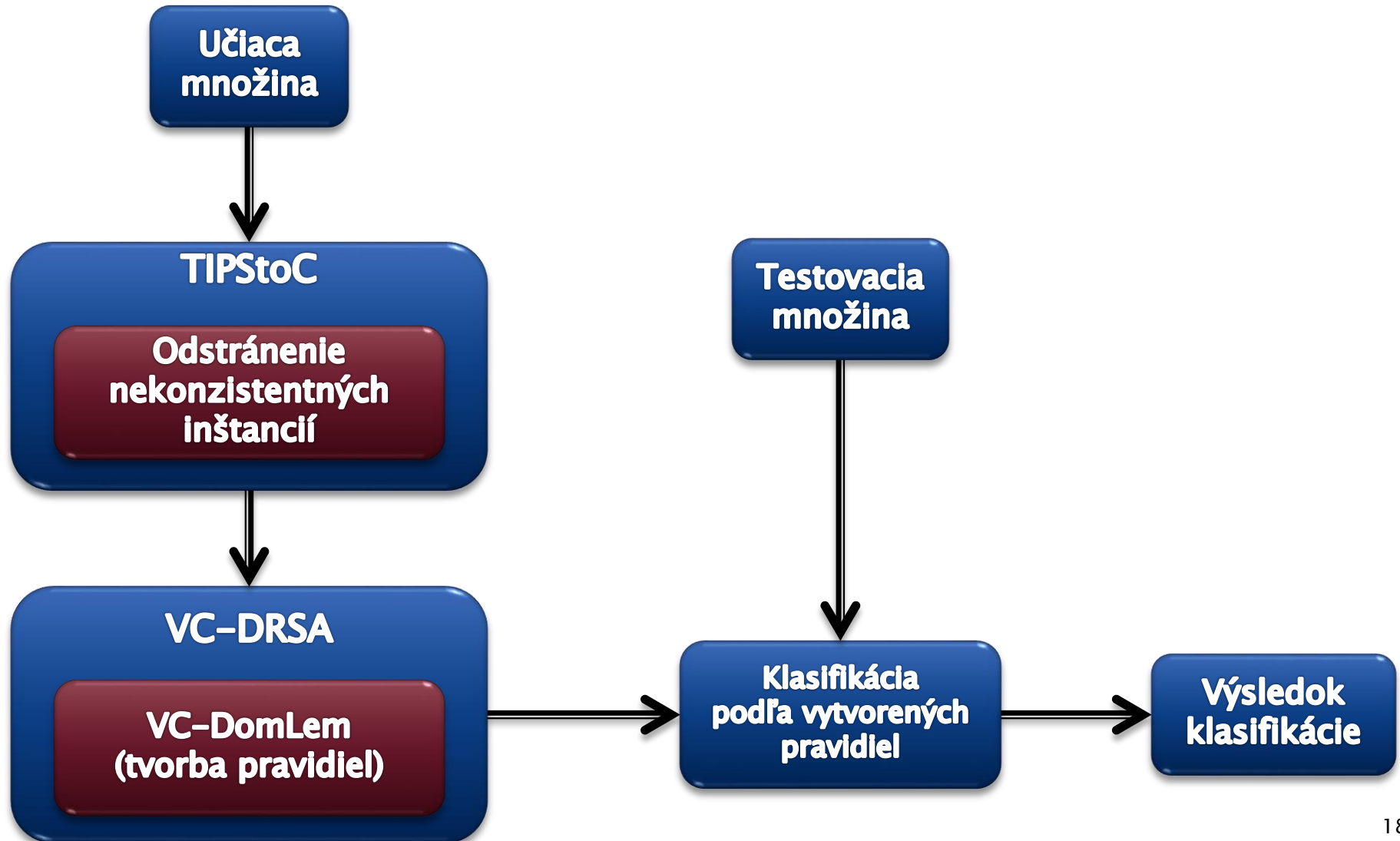
$$\alpha(y) = \max \left\{ \alpha^-(y) = \frac{|D_P^-(y) \cap X_{i+1}^{\geq}|}{|D_P^+(y) \cap X_i^{\leq}|}, \quad \alpha^+(y) = \frac{|D_P^+(y) \cap X_{i-1}^{\leq}|}{|D_P^-(y) \cap X_i^{\geq}|} \right\},$$

$$\epsilon(y) = \max \left\{ \epsilon^-(y) = \frac{|D_P^-(y) \cap X_{i+1}^{\geq}|}{|X_{i+1}^{\geq}|}, \quad \epsilon^+(y) = \frac{|D_P^+(y) \cap X_{i-1}^{\leq}|}{|X_{i-1}^{\leq}|} \right\},$$

$$\mu(y) = \max \left\{ \mu^-(y) = \frac{|D_P^-(y) \cap X_{i+1}^{\geq}|}{|D_P^-(y)|}, \quad \mu^+(y) = \frac{|D_P^+(y) \cap X_{i-1}^{\leq}|}{|D_P^+(y)|} \right\},$$

kde  $y \in X_i$

# Schéma aplikácie TIPStoC



# Výpočet miery neistoty pre pacientov trpiacich infekciou obličiek

	Q <sub>1</sub>	Q <sub>2</sub>	Q <sub>3</sub>	Q <sub>4</sub>	Q <sub>5</sub>	d	$\alpha$
V <sub>1</sub>	8	I	28	M	7,7	OKD	
V <sub>2</sub>	30	I	10	M	8,5	OKD	
V <sub>3</sub>	12	I	10	M	8,5	OKD	
V <sub>4</sub>	24	I	16	F	8,9	OKD	
V <sub>5</sub>	245	I	17	F	8,9	OKD	
V <sub>6</sub>	7	I	51	M	7,0	OKD	
V <sub>7</sub>	-	I	51	M	7,0	GN	
V <sub>8</sub>	511	I	55	F	9,5	GN	
V <sub>9</sub>	30	I	56	F	9,5	GN	
V <sub>10</sub>	53	I	69	F	9,3	AN	
V <sub>11</sub>	39	I	42	F	9,0	AN	

	Q <sub>1</sub>	Q <sub>2</sub>	Q <sub>3</sub>	Q <sub>4</sub>	Q <sub>5</sub>	d	$\alpha$
V <sub>12</sub>	46	C	43	F	9,0	AN	
V <sub>13</sub>	40	I	43	M	9,3	AN	
V <sub>14</sub>	113	C	57	F	9,5	AN	
V <sub>15</sub>	201	I	58	F	9,5	AN	
V <sub>16</sub>	132	I	10	F	8,9	GN	
V <sub>17</sub>	156	I	10	F	8,9	GN	
V <sub>18</sub>	34	I	52	F	8,2	AN	
V <sub>19</sub>	30	I	52	F	8,2	AN	
V <sub>20</sub>	2	I	53	M	8,5	GN	
V <sub>21</sub>	16	C	42	F	9,3	OKD	
V <sub>22</sub>	8	C	60	M	8,8	PKD	

# Výpočet miery neistoty pre pacientov trpiacich infekciou obličiek

$Q_1 = \langle 2, 562 \rangle$  - deň výskytu infekcie

$Q_2 = \{C, I\}$  - status liečby

$Q_3 = \langle 10, 69 \rangle$  - vek pacienta

$Q_4 = \{M, F\}$  - pohlavie

$Q_5 = \langle 7.0, 9.8 \rangle$  - interval náhodnosti vplyvu

$d = \{OKD, GN, AN, PKD\}$  - druh patológie obličiek

$X_1 = OKD, X_2 = GN, X_3 = AN, X_4 = PKD$

# Výpočet miery neistoty

Pre  $V_{18} \in X_3$

$$X_3^{\leq} = \{V_{17}, V_5, V_8, V_9, V_{16}\}$$

$$X_{3-1}^{\leq} = \{V_{17}, V_5, V_8, V_9, V_{16}\}$$

$$X_3^{\geq} = \{V_{87}, V_{19}, V_{11}\}$$

$$X_{3+1}^{\geq} = \emptyset$$

$$D_P^-(V_{18}) = \{V_{18}, V_{19}\}$$

$$D_P^+(V_{18}) = \{V_{18}, V_8\}$$

$$\alpha^-(V_{18}) = \frac{|D_P^-(V_{18}) \cap \emptyset|}{|D_P^+(V_{18}) \cap X_3^{\leq}|} = 0$$

$$\alpha^+(V_{18}) = \frac{|D_P^+(V_{18}) \cap X_{3-1}^{\leq}|}{|D_P^-(V_{18}) \cap X_3^{\geq}|} = \frac{1}{2}$$

$$\alpha(V_{18}) = \max \{\alpha^+(V_{18}), \alpha^-(V_{18})\} = 0.5$$

# 1. iterácia s mierou $\alpha$

	Q <sub>1</sub>	Q <sub>2</sub>	Q <sub>3</sub>	Q <sub>4</sub>	Q <sub>5</sub>	d	$\alpha$
V <sub>1</sub>	8	I	28	M	7,7	OKD	
V <sub>2</sub>	30	I	10	M	8,5	OKD	
V <sub>3</sub>	12	I	10	M	8,5	OKD	
V <sub>4</sub>	24	I	16	F	8,9	OKD	
V <sub>5</sub>	245	I	17	F	8,9	OKD	2,0
V <sub>6</sub>	7	I	51	M	7,0	OKD	
V <sub>7</sub>	-	I	51	M	7,0	GN	
V <sub>8</sub>	511	I	55	F	9,5	GN	3,0
V <sub>9</sub>	30	I	56	F	9,5	GN	1,0
V <sub>10</sub>	53	I	69	F	9,3	AN	
V <sub>11</sub>	39	I	42	F	9,0	AN	1,0

	Q <sub>1</sub>	Q <sub>2</sub>	Q <sub>3</sub>	Q <sub>4</sub>	Q <sub>5</sub>	d	$\alpha$
V <sub>12</sub>	46	C	43	F	9,0	AN	
V <sub>13</sub>	40	I	43	M	9,3	AN	
V <sub>14</sub>	113	C	57	F	9,5	AN	
V <sub>15</sub>	201	I	58	F	9,5	AN	
V <sub>16</sub>	132	I	10	F	8,9	GN	1,0
V <sub>17</sub>	156	I	10	F	8,9	GN	0,5
V <sub>18</sub>	34	I	52	F	8,2	AN	0,5
V <sub>19</sub>	30	I	52	F	8,2	AN	2,0
V <sub>20</sub>	2	I	53	M	8,5	GN	
V <sub>21</sub>	16	C	42	F	9,3	OKD	
V <sub>22</sub>	8	C	60	M	8,8	PKD	

# 1. iterácia s mierou $\alpha$

	Q <sub>1</sub>	Q <sub>2</sub>	Q <sub>3</sub>	Q <sub>4</sub>	Q <sub>5</sub>	d	$\alpha$
V <sub>1</sub>	8	I	28	M	7,7	OKD	
V <sub>2</sub>	30	I	10	M	8,5	OKD	
V <sub>3</sub>	12	I	10	M	8,5	OKD	
V <sub>4</sub>	24	I	16	F	8,9	OKD	
V <sub>5</sub>	245	I	17	F	8,9	OKD	2,0
V <sub>6</sub>	7	I	51	M	7,0	OKD	
V <sub>7</sub>	-	I	51	M	7,0	GN	
<del>V<sub>8</sub></del>	<del>511</del>	<del>I</del>	<del>55</del>	<del>F</del>	<del>9,5</del>	<del>GN</del>	<del>3,0</del>
V <sub>9</sub>	30	I	56	F	9,5	GN	1,0
V <sub>10</sub>	53	I	69	F	9,3	AN	
V <sub>11</sub>	39	I	42	F	9,0	AN	1,0

	Q <sub>1</sub>	Q <sub>2</sub>	Q <sub>3</sub>	Q <sub>4</sub>	Q <sub>5</sub>	d	$\alpha$
V <sub>12</sub>	46	C	43	F	9,0	AN	
V <sub>13</sub>	40	I	43	M	9,3	AN	
V <sub>14</sub>	113	C	57	F	9,5	AN	
V <sub>15</sub>	201	I	58	F	9,5	AN	
V <sub>16</sub>	132	I	10	F	8,9	GN	1,0
V <sub>17</sub>	156	I	10	F	8,9	GN	0,5
V <sub>18</sub>	34	I	52	F	8,2	AN	0,5
V <sub>19</sub>	30	I	52	F	8,2	AN	2,0
V <sub>20</sub>	2	I	53	M	8,5	GN	
V <sub>21</sub>	16	C	42	F	9,3	OKD	
V <sub>22</sub>	8	C	60	M	8,8	PKD	

## 2. iterácia s mierou $\alpha$

	Q <sub>1</sub>	Q <sub>2</sub>	Q <sub>3</sub>	Q <sub>4</sub>	Q <sub>5</sub>	d	$\alpha$
V <sub>1</sub>	8	I	28	M	7,7	OKD	
V <sub>2</sub>	30	I	10	M	8,5	OKD	
V <sub>3</sub>	12	I	10	M	8,5	OKD	
V <sub>4</sub>	24	I	16	F	8,9	OKD	
V <sub>5</sub>	245	I	17	F	8,9	OKD	2,0
V <sub>6</sub>	7	I	51	M	7,0	OKD	
V <sub>7</sub>	-	I	51	M	7,0	GN	
<del>V<sub>8</sub></del>	<del>511</del>	<del>I</del>	<del>55</del>	<del>F</del>	<del>9,5</del>	<del>GN</del>	
V <sub>9</sub>	30	I	56	F	9,5	GN	1,0
V <sub>10</sub>	53	I	69	F	9,3	AN	
V <sub>11</sub>	39	I	42	F	9,0	AN	

	Q <sub>1</sub>	Q <sub>2</sub>	Q <sub>3</sub>	Q <sub>4</sub>	Q <sub>5</sub>	d	$\alpha$
V <sub>12</sub>	46	C	43	F	9,0	AN	
V <sub>13</sub>	40	I	43	M	9,3	AN	
V <sub>14</sub>	113	C	57	F	9,5	AN	
V <sub>15</sub>	201	I	58	F	9,5	AN	
V <sub>16</sub>	132	I	10	F	8,9	GN	
V <sub>17</sub>	156	I	10	F	8,9	GN	0,5
V <sub>18</sub>	34	I	52	F	8,2	AN	
V <sub>19</sub>	30	I	52	F	8,2	AN	1,0
V <sub>20</sub>	2	I	53	M	8,5	GN	
V <sub>21</sub>	16	C	42	F	9,3	OKD	
V <sub>22</sub>	8	C	60	M	8,8	PKD	



## 2. iterácia s mierou $\alpha$

	Q <sub>1</sub>	Q <sub>2</sub>	Q <sub>3</sub>	Q <sub>4</sub>	Q <sub>5</sub>	d	$\alpha$
V <sub>1</sub>	8	I	28	M	7,7	OKD	
V <sub>2</sub>	30	I	10	M	8,5	OKD	
V <sub>3</sub>	12	I	10	M	8,5	OKD	
V <sub>4</sub>	24	I	16	F	8,9	OKD	
<del>V<sub>5</sub></del>	<del>245</del>	<del>I</del>	<del>17</del>	<del>F</del>	<del>8,9</del>	<del>OKD</del>	<b>2,0</b>
V <sub>6</sub>	7	I	51	M	7,0	OKD	
V <sub>7</sub>	-	I	51	M	7,0	GN	
<del>V<sub>8</sub></del>	<del>511</del>	<del>I</del>	<del>55</del>	<del>F</del>	<del>9,5</del>	<del>GN</del>	
V <sub>9</sub>	30	I	56	F	9,5	GN	<b>1,0</b>
V <sub>10</sub>	53	I	69	F	9,3	AN	
V <sub>11</sub>	39	I	42	F	9,0	AN	

	Q <sub>1</sub>	Q <sub>2</sub>	Q <sub>3</sub>	Q <sub>4</sub>	Q <sub>5</sub>	d	$\alpha$
V <sub>12</sub>	46	C	43	F	9,0	AN	
V <sub>13</sub>	40	I	43	M	9,3	AN	
V <sub>14</sub>	113	C	57	F	9,5	AN	
V <sub>15</sub>	201	I	58	F	9,5	AN	
V <sub>16</sub>	132	I	10	F	8,9	GN	
V <sub>17</sub>	156	I	10	F	8,9	GN	<b>0,5</b>
V <sub>18</sub>	34	I	52	F	8,2	AN	
V <sub>19</sub>	30	I	52	F	8,2	AN	<b>1,0</b>
V <sub>20</sub>	2	I	53	M	8,5	GN	
V <sub>21</sub>	16	C	42	F	9,3	OKD	
V <sub>22</sub>	8	C	60	M	8,8	PKD	

## 2. iterácia s mierou $\alpha$

	Q <sub>1</sub>	Q <sub>2</sub>	Q <sub>3</sub>	Q <sub>4</sub>	Q <sub>5</sub>	d	$\alpha$
V <sub>1</sub>	8	I	28	M	7,7	OKD	
V <sub>2</sub>	30	I	10	M	8,5	OKD	
V <sub>3</sub>	12	I	10	M	8,5	OKD	
V <sub>4</sub>	24	I	16	F	8,9	OKD	
<del>V<sub>5</sub></del>	<del>245</del>	<del>I</del>	<del>17</del>	<del>F</del>	<del>8,9</del>	<del>OKD</del>	<b>2,0</b>
V <sub>6</sub>	7	I	51	M	7,0	OKD	
V <sub>7</sub>	-	I	51	M	7,0	GN	
<del>V<sub>8</sub></del>	<del>511</del>	<del>I</del>	<del>55</del>	<del>F</del>	<del>9,5</del>	<del>GN</del>	
V <sub>9</sub>	30	I	56	F	9,5	GN	<b>1,0</b>
V <sub>10</sub>	53	I	69	F	9,3	AN	
V <sub>11</sub>	39	I	42	F	9,0	AN	

	Q <sub>1</sub>	Q <sub>2</sub>	Q <sub>3</sub>	Q <sub>4</sub>	Q <sub>5</sub>	d	$\alpha$
V <sub>12</sub>	46	C	43	F	9,0	AN	
V <sub>13</sub>	40	I	43	M	9,3	AN	
V <sub>14</sub>	113	C	57	F	9,5	AN	
V <sub>15</sub>	201	I	58	F	9,5	AN	
V <sub>16</sub>	132	I	10	F	8,9	GN	
V <sub>17</sub>	156	I	10	F	8,9	GN	<b>0,5</b>
V <sub>18</sub>	34	I	52	F	8,2	AN	
V <sub>19</sub>	30	I	52	F	8,2	AN	<b>1,0</b>
V <sub>20</sub>	2	I	53	M	8,5	GN	
V <sub>21</sub>	16	C	42	F	9,3	OKD	
V <sub>22</sub>	8	C	60	M	8,8	PKD	

### 3. iterácia s mierou $\alpha$

	Q <sub>1</sub>	Q <sub>2</sub>	Q <sub>3</sub>	Q <sub>4</sub>	Q <sub>5</sub>	d	$\alpha$
V <sub>1</sub>	8	I	28	M	7,7	OKD	
V <sub>2</sub>	30	I	10	M	8,5	OKD	
V <sub>3</sub>	12	I	10	M	8,5	OKD	
V <sub>4</sub>	24	I	16	F	8,9	OKD	
<del>V<sub>5</sub></del>	<del>245</del>	<del>I</del>	<del>17</del>	<del>F</del>	<del>8,9</del>	<del>OKD</del>	
V <sub>6</sub>	7	I	51	M	7,0	OKD	
V <sub>7</sub>	-	I	51	M	7,0	GN	
<del>V<sub>8</sub></del>	<del>511</del>	<del>I</del>	<del>55</del>	<del>F</del>	<del>9,5</del>	<del>GN</del>	
V <sub>9</sub>	30	I	56	F	9,5	GN	1,0
V <sub>10</sub>	53	I	69	F	9,3	AN	
V <sub>11</sub>	39	I	42	F	9,0	AN	

	Q <sub>1</sub>	Q <sub>2</sub>	Q <sub>3</sub>	Q <sub>4</sub>	Q <sub>5</sub>	d	$\alpha$
V <sub>12</sub>	46	C	43	F	9,0	AN	
V <sub>13</sub>	40	I	43	M	9,3	AN	
V <sub>14</sub>	113	C	57	F	9,5	AN	
V <sub>15</sub>	201	I	58	F	9,5	AN	
V <sub>16</sub>	132	I	10	F	8,9	GN	
V <sub>17</sub>	156	I	10	F	8,9	GN	
V <sub>18</sub>	34	I	52	F	8,2	AN	
V <sub>19</sub>	30	I	52	F	8,2	AN	1,0
V <sub>20</sub>	2	I	53	M	8,5	GN	
V <sub>21</sub>	16	C	42	F	9,3	OKD	
V <sub>22</sub>	8	C	60	M	8,8	PKD	

# 3. iterácia s mierou $\alpha$

	Q <sub>1</sub>	Q <sub>2</sub>	Q <sub>3</sub>	Q <sub>4</sub>	Q <sub>5</sub>	d	$\alpha$
V <sub>1</sub>	8	I	28	M	7,7	OKD	
V <sub>2</sub>	30	I	10	M	8,5	OKD	
V <sub>3</sub>	12	I	10	M	8,5	OKD	
V <sub>4</sub>	24	I	16	F	8,9	OKD	
<del>V<sub>5</sub></del>	<del>245</del>	<del>I</del>	<del>17</del>	<del>F</del>	<del>8,9</del>	<del>OKD</del>	
V <sub>6</sub>	7	I	51	M	7,0	OKD	
V <sub>7</sub>	-	I	51	M	7,0	GN	
<del>V<sub>8</sub></del>	<del>511</del>	<del>I</del>	<del>55</del>	<del>F</del>	<del>9,5</del>	<del>GN</del>	
<del>V<sub>9</sub></del>	<del>30</del>	<del>I</del>	<del>56</del>	<del>F</del>	<del>9,5</del>	<del>GN</del>	
V <sub>10</sub>	53	I	69	F	9,3	AN	
V <sub>11</sub>	39	I	42	F	9,0	AN	

	Q <sub>1</sub>	Q <sub>2</sub>	Q <sub>3</sub>	Q <sub>4</sub>	Q <sub>5</sub>	d	$\alpha$
V <sub>12</sub>	46	C	43	F	9,0	AN	
V <sub>13</sub>	40	I	43	M	9,3	AN	
V <sub>14</sub>	113	C	57	F	9,5	AN	
V <sub>15</sub>	201	I	58	F	9,5	AN	
V <sub>16</sub>	132	I	10	F	8,9	GN	
V <sub>17</sub>	156	I	10	F	8,9	GN	
V <sub>18</sub>	34	I	52	F	8,2	AN	
V <sub>19</sub>	30	I	52	F	8,2	AN	
V <sub>20</sub>	2	I	53	M	8,5	GN	
V <sub>21</sub>	16	C	42	F	9,3	OKD	
V <sub>22</sub>	8	C	60	M	8,8	PKD	

# Konzistentná tabuľka

	Q <sub>1</sub>	Q <sub>2</sub>	Q <sub>3</sub>	Q <sub>4</sub>	Q <sub>5</sub>	d	$\alpha$
V <sub>1</sub>	8	I	28	M	7,7	OKD	
V <sub>2</sub>	30	I	10	M	8,5	OKD	
V <sub>3</sub>	12	I	10	M	8,5	OKD	
V <sub>4</sub>	24	I	16	F	8,9	OKD	
V <sub>6</sub>	7	I	51	M	7,0	OKD	
V <sub>7</sub>	-	I	51	M	7,0	GN	
V <sub>10</sub>	53	I	69	F	9,3	AN	
V <sub>11</sub>	39	I	42	F	9,0	AN	

	Q <sub>1</sub>	Q <sub>2</sub>	Q <sub>3</sub>	Q <sub>4</sub>	Q <sub>5</sub>	d	$\alpha$
V <sub>12</sub>	46	C	43	F	9,0	AN	
V <sub>13</sub>	40	I	43	M	9,3	AN	
V <sub>14</sub>	113	C	57	F	9,5	AN	
V <sub>15</sub>	201	I	58	F	9,5	AN	
V <sub>16</sub>	132	I	10	F	8,9	GN	
V <sub>17</sub>	156	I	10	F	8,9	GN	
V <sub>18</sub>	34	I	52	F	8,2	AN	
V <sub>19</sub>	30	I	52	F	8,2	AN	
V <sub>20</sub>	2	I	53	M	8,5	GN	
V <sub>21</sub>	16	C	42	F	9,3	OKD	
V <sub>22</sub>	8	C	60	M	8,8	PKD	

# Navrhnuté miery neistoty

$$\epsilon^\dagger(y) = \max \left\{ \epsilon^{\dagger(-)}(y) = \frac{|D_P^-(y) \cap X_{i+1}^{\geq}|}{|X_i^{\leq}|}, \quad \epsilon^{\dagger(+)}(y) = \frac{|D_P^+(y) \cap X_{i-1}^{\leq}|}{|X_i^{\geq}|} \right\}.$$

$$\epsilon^\ddagger(y) = \max \left\{ \epsilon^{\ddagger(-)}(y) = \frac{|D_P^-(y) \cap X_{i-1}^{\leq}|}{|X_i^{\leq}|}, \quad \epsilon^{\ddagger(+)}(y) = \frac{|D_P^+(y) \cap X_{i+1}^{\geq}|}{|X_i^{\geq}|} \right\}.$$

$$\epsilon^\star(y) = \max \left\{ \epsilon^{\star(-)}(y) = \frac{|D_P^-(y) \cap X_{i-1}^{\leq}|}{|X_{i-1}^{\leq}|}, \quad \epsilon^{\star(+)}(y) = \frac{|D_P^+(y) \cap X_{i+1}^{\geq}|}{|X_{i+1}^{\geq}|} \right\}.$$

$$\mu^\star(y) = \max \left\{ \mu^{\star(-)}(y) = \frac{|D_P^-(y) \cap X_i^{\leq}|}{|D_P^-(y)|}, \quad \mu^{\star(+)}(y) = \frac{|D_P^+(y) \cap X_i^{\geq}|}{|D_P^+(y)|} \right\}.$$

kde  $y \in X_i$

# Navrhnuté miery neistoty

$$\beta(y) = \max \left\{ \beta^{(-)}(y), \beta^{(+)}(y) \right\}.$$

$$\beta^{(-)}(y) = \frac{|D_P^+(y) \cap X_i^{\leq}|}{|D_P^+(y) \cap X_i^{\leq}| + |D_P^-(y) \cap X_{i-1}^{\leq}|},$$

$$\beta^{(+)}(y) = \frac{|D_P^-(y) \cap X_i^{\geq}|}{|D_P^-(y) \cap X_i^{\geq}| + |D_P^+(y) \cap X_{i+1}^{\geq}|}.$$

# Porovnanie počtov odstránených objektov pre miery neistoty

Miera neistoty	Odstránené objekty	Počet
$\alpha$	$V_8, V_5, V_{19}$	3
$\epsilon$	$V_{17}, V_8, V_{16}, V_{19}$	4
$\mu$	$V_5, V_{18}, V_{19}, V_{11},$	4
$\epsilon^\dagger$	$V_5, V_8, V_{19}$	3
$\epsilon^\ddagger$	$V_8, V_{17}, V_{19}, V_5$	4
$\epsilon^*$	$V_8, V_{17}, V_{19}, V_5$	4
$\mu^*$	$V_{17}, V_8, V_9, V_{16}$	4
$\beta$	$V_{18}, V_{17}, V_{19}, V_5, V_8$	5



# Porovnanie počtov odstránených objektov pre miery neistoty

Miera neistoty	Odstránené objekty	Počet
$\alpha$	$V_8, V_5, V_{19}$	3
$\epsilon$	$V_{17}, V_8, V_{16}, V_{19}$	4
$\mu$	$V_5, V_{18}, V_{19}, V_{11},$	4
$\epsilon^\dagger$	$V_5, V_8, V_{19}$	3
$\epsilon^\ddagger$	$V_8, V_{17}, V_{19}, V_5$	4
$\epsilon^*$	$V_8, V_{17}, V_{19}, V_5$	4
$\mu^*$	$V_{17}, V_8, V_9, V_{16}$	4
$\beta$	$V_{18}, V_{17}, V_{19}, V_5, V_8$	5

# Navrhnutá modifikácia algoritmu TIPStoC (Deng et al. 2011)

- ▶ Zrýchlenie vyhľadávania nekonzistentných objektov znížením počtu testovaných podmienok zo 7 na 3.
- ▶ Efektívna implementácia vyhľadávania nekonzistentných objektov.

# Použité databázy pre overenie mier a algoritmov

Databáza	Počet objektov	Počet atribútov	Počet tried
balance	625	4	3
cpu	209	6	4
housing	506	13	4
car	1728	6	4
australian	690	14	2
breast-w	699	9	2
denbosch	119	8	2
ESL	488	4	9
windsor	546	10	4
SWD	1000	10	4
LEV	1000	4	5
breast-c	286	7	2
ERA	1000	4	9

# Použitý server

## AMAZON EC2 On-Demand Instances

<b>Instance Family</b>	Compute optimized
<b>Instance Type</b>	c3.8xlarge
<b>Processor Arch</b>	64-bit
<b>vCPU</b>	32
<b>ECU</b>	108
<b>Memory (GiB)</b>	60
<b>Instance Storage (GB)</b>	2 x 320 SSD
<b>EBS-optimized Available</b>	-
<b>Network Performance</b>	10 Gigabit*4
<b>Physical Processor</b>	Intel Xeon E5-2680 v2
<b>Intel® AES-NI</b>	Yes
<b>Intel® AVX†</b>	Yes
<b>Intel® Turbo</b>	Yes

# Výsledky porovnania mier

Databáza	VC-DRSA	$\alpha$	$\beta$	$\epsilon$	$\epsilon^\dagger$	$\epsilon^\ddagger$	$\epsilon^*$	$\mu$	$\mu^*$
balance	76.37 [8] ± 0.9794	81.67 [5] ± 0.6461	82.07 [4] ± 0.8377	<b>86.05 [1]</b> ± 0.9126	78.92 [7] ± 0.5374	83.70 [2.5] ± 0.9105	83.70 [2.5] ± 0.9464	79.45 [6] ± 0.8146	75.82 [9] ± 0.9134
cpu	94.11 [6] ± 0.7944	95.18 [4] ± 0.5470	<b>95.89 [1]</b> ± 0.2826	92.07 [8] ± 0.4832	95.55 [2] ± 0.2144	94.35 [5] ± 0.2281	93.98 [7] ± 0.2429	95.26 [3] ± 0.5634	91.18 [9] ± 0.4423
housing	70.50 [4] ± 0.3332	70.70 [7] ± 0.8893	69.72 [6] ± 0.6372	68.64 [9] ± 0.6450	<b>71.53 [1]</b> ± 0.4102	70.41 [5] ± 0.7517	70.52 [3] ± 0.6532	70.59 [2] ± 0.5879	68.68 [8] ± 0.9810
car	97.32 [3] ± 0.2115	96.93 [5] ± 0.4472	96.52 [6] ± 0.5131	94.82 [7.5] ± 0.3141	<b>98.92 [1]</b> ± 0.3352	94.82 [7.5] ± 0.5564	94.27 [9] ± 0.3260	97.33 [2] ± 0.4424	97.10 [4] ± 0.3156
australian	75.41 [4] ± 0.5403	75.83 [3] ± 0.2345	74.89 [5] ± 0.4874	73.22 [6] ± 0.3611	72.95 [7] ± 0.2694	72.75 [8.5] ± 0.2900	72.75 [8.5] ± 0.3001	<b>76.39 [1]</b> ± 0.4558	76.09 [2] ± 0.1544
breast-w	96.25 [4] ± 0.2598	94.67 [6] ± 0.0875	93.51 [8] ± 0.1131	95.35 [5] ± 0.2718	<b>98.61 [1]</b> ± 0.1653	97.50 [2] ± 0.2253	96.50 [3] ± 0.2260	94.66 [7] ± 0.0811	91.63 [9] ± 0.1815
denbosch	85.21 [8] ± 0.3633	86.36 [5] ± 0.3623	87.20 [3] ± 0.4792	85.36 [7] ± 0.2821	84.83 [9] ± 0.5151	87.20 [3] ± 0.4679	87.20 [3] ± 0.4741	<b>87.85 [1]</b> ± 0.1704	85.70 [6] ± 0.4618

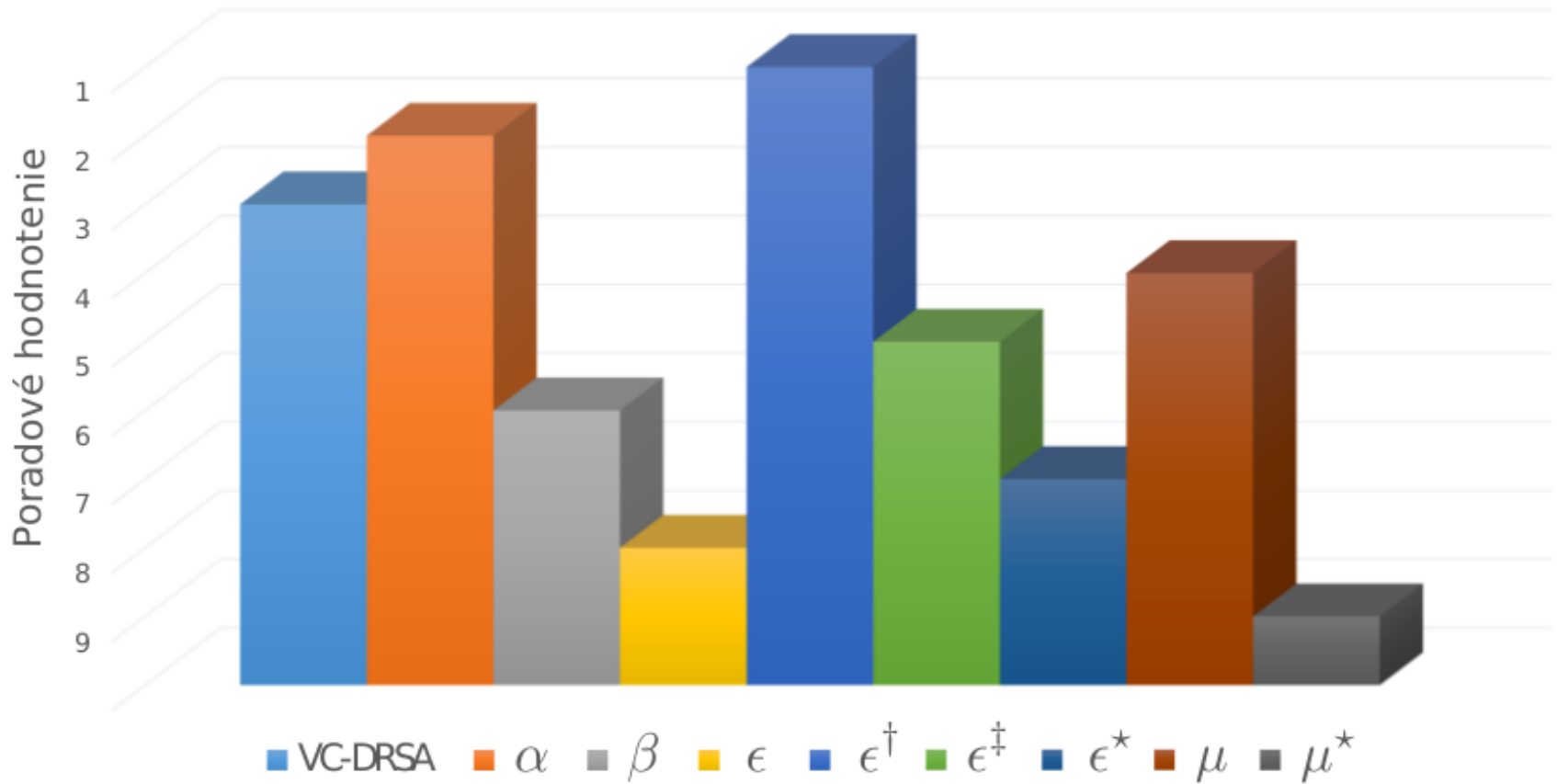
# Výsledky porovnania mier

Databáza	VC-DRSA	$\alpha$	$\beta$	$\epsilon$	$\epsilon^\dagger$	$\epsilon^\ddagger$	$\epsilon^*$	$\mu$	$\mu^*$
ESL	63.86 [2] ± 0.2665	<b>69.52</b> [1] ± 0.6876	57.71 [8] ± 0.7091	60.23 [7] ± 0.4070	63.66 [3] ± 0.2763	61.40 [6] ± 0.4292	56.20 [9] ± 0.6052	63.01 [5] ± 0.3701	63.51 [4] ± 0.9079
windsor	<b>53.97</b> [1] ± 0.2320	52.51 [3] ± 0.4260	44.24 [8] ± 0.8216	50.99 [4] ± 0.3161	53.23 [2] ± 0.2834	47.46 [6.5] ± 0.8265	47.46 [6.5] ± 0.8233	48.87 [5] ± 0.1525	42.26 [9] ± 0.6525
SWD	57.19 [5] ± 0.4883	57.62 [4] ± 0.7517	48.78 [7] ± 0.5363	58.28 [2] ± 0.6002	<b>58.30</b> [1] ± 0.5847	58.08 [3] ± 0.4973	42.40 [8] ± 0.6084	49.29 [6] ± 0.5918	41.00 [9] ± 0.7467
LEV	55.48 [3] ± 0.4562	<b>63.57</b> [1] ± 0.3228	62.60 [2] ± 0.5894	54.70 [5.5] ± 0.6041	54.70 [5.5] ± 0.6015	54.70 [5.5] ± 0.6065	54.70 [5.5] ± 0.6042	52.40 [8] ± 0.5438	51.60 [9] ± 0.4634
breast-c	76.45 [2] ± 1.0187	75.53 [6] ± 0.6553	73.27 [8] ± 0.4222	74.44 [7] ± 1.0493	<b>76.80</b> [1] ± 0.5812	75.56 [4.5] ± 0.7988	75.56 [4.5] ± 0.8268	75.78 [3] ± 0.6771	70.64 [9] ± 0.6625
ERA	26.13 [2.5] ± 0.5493	<b>28.38</b> [1] ± 0.8456	21.08 [7] ± 0.6034	19.09 [8] ± 0.6752	26.13 [2.5] ± 0.6578	25.78 [5] ± 0.5594	26.04 [4] ± 0.6844	22.19 [6] ± 0.7783	16.58 [9] ± 0.8400
Priem.Por.	4.04	3.92	5.62	5.92	<b>3.31</b>	4.92	5.65	4.23	7.38

# Výsledky porovnania mier

Databáza	VC-DRSA	$\alpha$	$\beta$	$\epsilon$	$\epsilon^\dagger$	$\epsilon^\ddagger$	$\epsilon^*$	$\mu$	$\mu^*$
ESL	63.86 [2] ± 0.2665	<b>69.52</b> [1] ± 0.6876	57.71 [8] ± 0.7091	60.23 [7] ± 0.4070	63.66 [3] ± 0.2763	61.40 [6] ± 0.4292	56.20 [9] ± 0.6052	63.01 [5] ± 0.3701	63.51 [4] ± 0.9079
windsor	<b>53.97</b> [1] ± 0.2320	52.51 [3] ± 0.4260	44.24 [8] ± 0.8216	50.99 [4] ± 0.3161	53.23 [2] ± 0.2834	47.46 [6.5] ± 0.8265	47.46 [6.5] ± 0.8233	48.87 [5] ± 0.1525	42.26 [9] ± 0.6525
SWD	57.19 [5] ± 0.4883	57.62 [4] ± 0.7517	48.78 [7] ± 0.5363	58.28 [2] ± 0.6002	<b>58.30</b> [1] ± 0.5847	58.08 [3] ± 0.4973	42.40 [8] ± 0.6084	49.29 [6] ± 0.5918	41.00 [9] ± 0.7467
LEV	55.48 [3] ± 0.4562	<b>63.57</b> [1] ± 0.3228	62.60 [2] ± 0.5894	54.70 [5.5] ± 0.6041	54.70 [5.5] ± 0.6015	54.70 [5.5] ± 0.6065	54.70 [5.5] ± 0.6042	52.40 [8] ± 0.5438	51.60 [9] ± 0.4634
breast-c	76.45 [2] ± 1.0187	75.53 [6] ± 0.6553	73.27 [8] ± 0.4222	74.44 [7] ± 1.0493	<b>76.80</b> [1] ± 0.5812	75.56 [4.5] ± 0.7988	75.56 [4.5] ± 0.8268	75.78 [3] ± 0.6771	70.64 [9] ± 0.6625
ERA	26.13 [2.5] ± 0.5493	<b>28.38</b> [1] ± 0.8456	21.08 [7] ± 0.6034	19.09 [8] ± 0.6752	26.13 [2.5] ± 0.6578	25.78 [5] ± 0.5594	26.04 [4] ± 0.6844	22.19 [6] ± 0.7783	16.58 [9] ± 0.8400
Priem.Por.	4.04	3.92	5.62	5.92	<b>3.31</b>	4.92	5.65	4.23	7.38

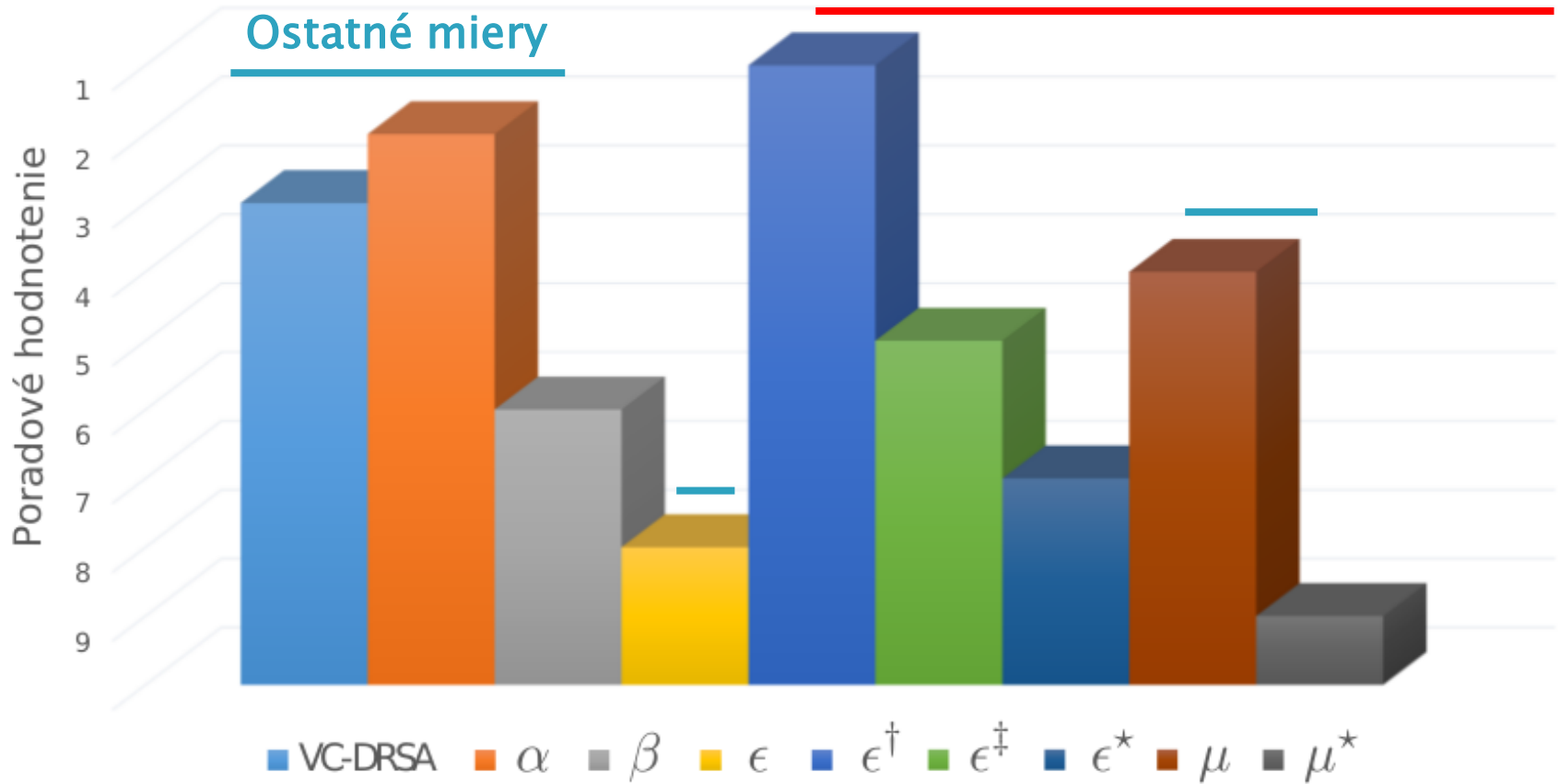
# Výsledky porovnania mier





# Výsledky porovnania mier

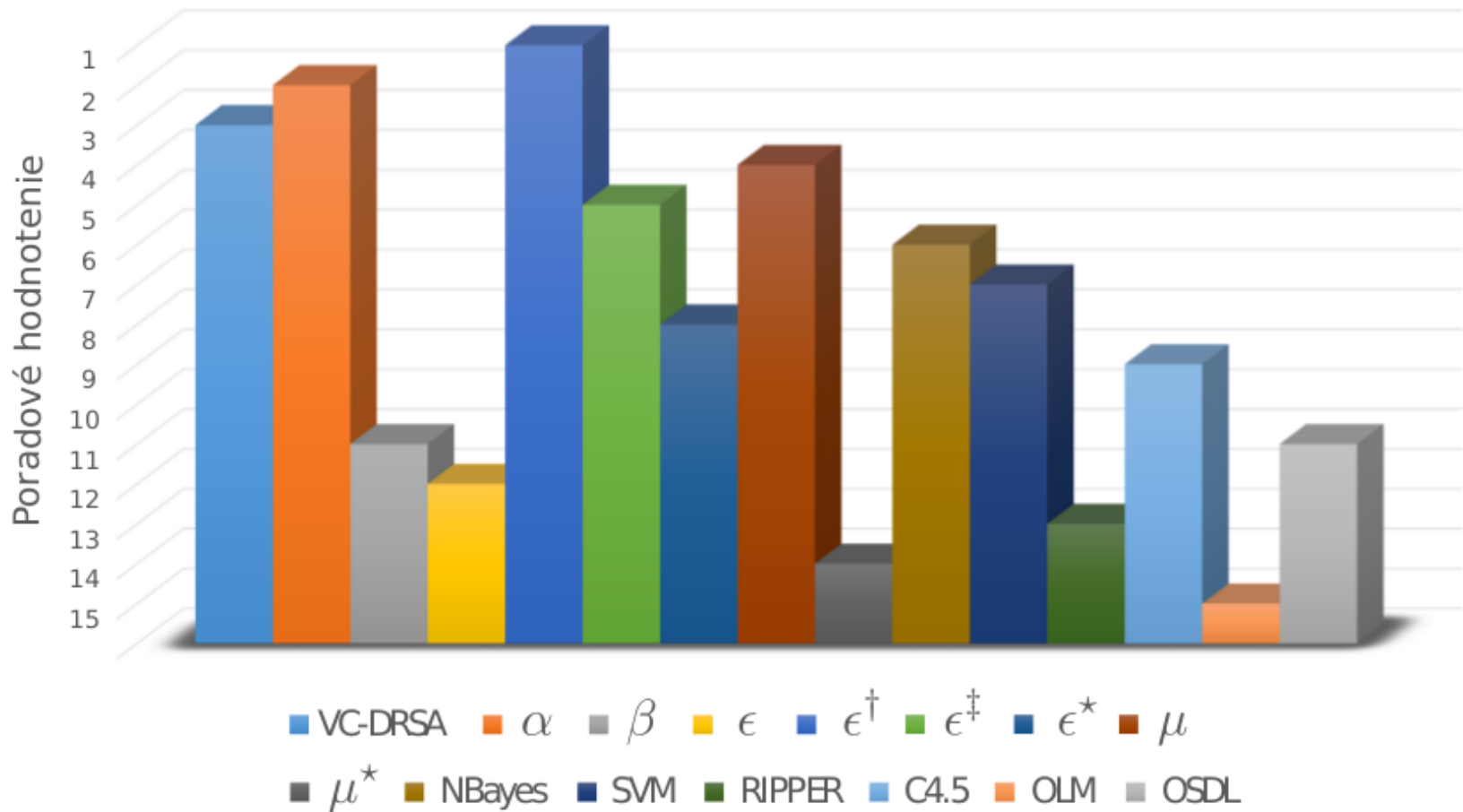
Navrhované riešenie



# Ďalšie porovnávané algoritmy

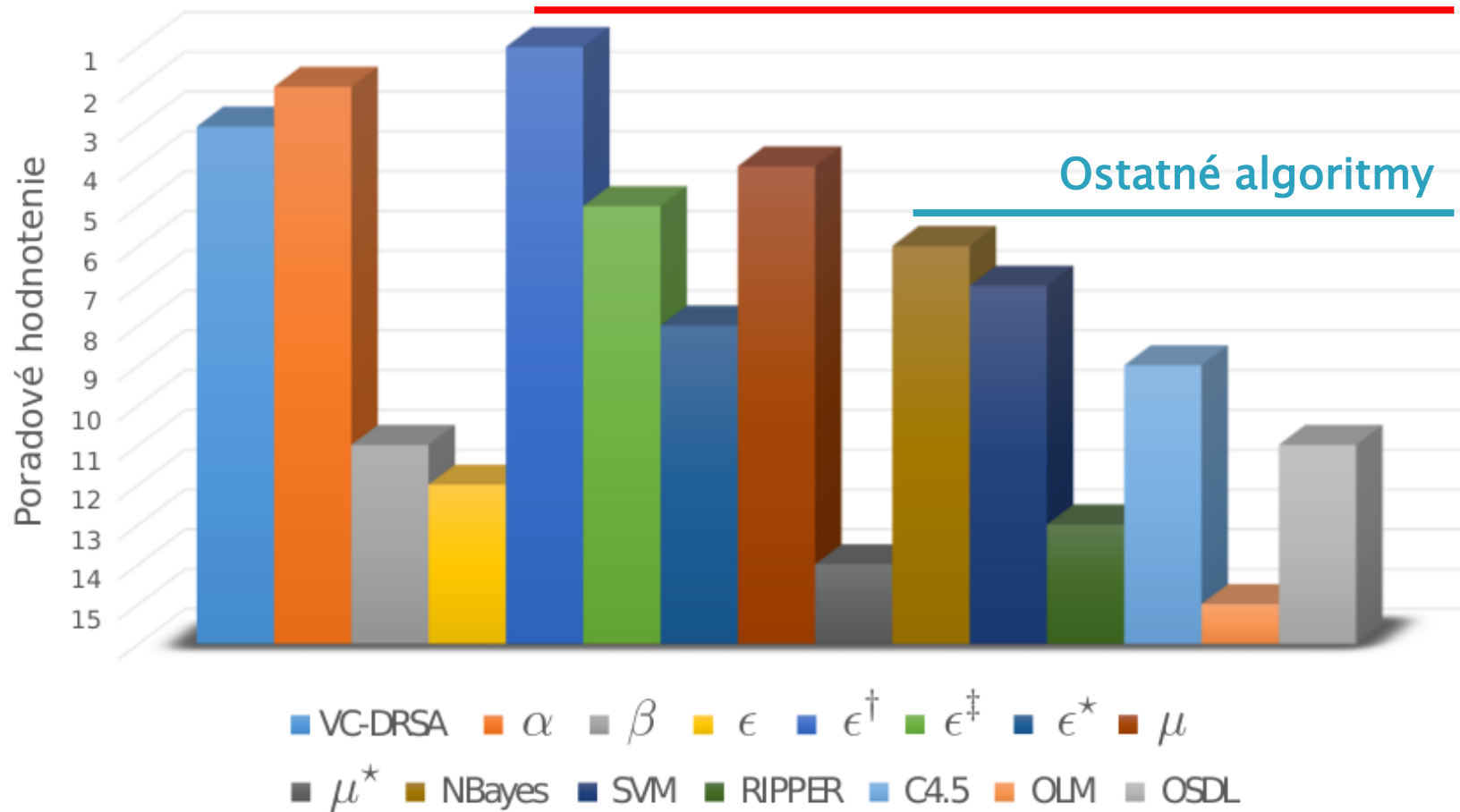
- **Naïve Bayes** - implementovaný podľa (John and Langley, 1995).
- **SVM** - Support Vector Machine z lineárnym jadrom, implementovaný podľa (Platt, 1998).
- **RIPPER** - klasifikátor rozhodovacích pravidiel, implementovaný podľa (Cohen, 1995).
- **C4.5** - klasifikátor rozhodovacích stromov, implementovaný podľa (Quinlan, 1993).
- **OLM** - Ordinal Learning Model, implementovaný podľa (Ben-David, 1992).
- **OSDL** - Ordinal Stochastic Dominance Learner, implementovaný podľa (Lievens et al., 2008). Algoritmus je vo verzii vyváženej, ktorá dosahuje lepšie výsledky na nekonzistentných údajoch.

# Celkové vyhodnotenie



# Celkové vyhodnotenie

Navrhované riešenie



# Prínos práce

- ✓ Jednoduchý spôsob identifikácie skupiny problémových pacientov
- ✓ Štatisticky významné zlepšenie kvality rozhodovacích pravidiel (modelov)
- ✓ Zvýšenie celkovej úspešnosti klasifikácie
- ✓ Návrh metodiky porovnávania mier neistoty pre budúce výskumy

Ďakujem za  
pozornosť

Ďakujem za  
pozornosť

Ďakujem za  
pozornosť



# TIPStoC

**Input:** množina objektov univerzum  $U$

miera neistoty  $\Delta$

**Output:** konzistentná množina objektov univerzum

**Define:**  $\Phi$  - množina nekonzistentných objektov na odstránenie s  $U$

$\Theta_I$  - množina nekonzistentných objektov

$\Theta_M$  - množina hodnôt uvažovanej miery neistoty  $\forall \varphi : \varphi \in \Theta_I$

$\sigma$  - objekt s maximálnou neistotou  $\max_{\sigma \in \Theta_I} \{\Theta_M(\sigma)\}$

# TIPStoC

```
1: procedure TIPSTOC( $U, \Delta$ )
2:    $\Phi := \emptyset$ 
3:    $[\Theta_I, \Theta_M] := \text{findInconsistentObject}(U, \Delta)$ 
4:   while  $\Theta_I \neq \emptyset$  do
5:      $\sigma := \text{maxInconsistentObject}(\Theta_I, \Theta_M)$ 
6:      $\Phi := \Phi \cup \{\sigma\}$ 
7:      $\Theta_I := \Theta_I - \{\sigma\}$ 
8:      $[\Theta_I, \Theta_M] := \text{findInconsistentObject}(\Theta_I, \Delta)$ 
9:   end while
10:  return  $U - \Phi$ 
11: end procedure
```

▷ vracia konzistentnú podmnožinu univerzum

# TIPStoC

1: **procedure** TIPSTOC( $U, \Delta$ )

2:      $\Phi := \emptyset$

3:      $[\Theta_I, \Theta_M] := \text{findInconsistentObject}(U, \Delta)$

Pôvodne vynechaný krok

4:     **while**  $\Theta_I \neq \emptyset$  **do**

5:          $\sigma := \text{maxInconsistentObject}(\Theta_I, \Theta_M)$

6:          $\Phi := \Phi \cup \{\sigma\}$

7:          $\Theta_I := \Theta_I - \{\sigma\}$

8:          $[\Theta_I, \Theta_M] := \text{findInconsistentObject}(\Theta_I, \Delta)$

9:     **end while**

10:     **return**  $U - \Phi$

▷ vracia konzistentnú podmnožinu univerzum

11: **end procedure**

# TIPStoC – findInconsistentObject()

**Input:**  $\Theta$  - množina potenciálne nekonzistentných objektov,  
 $\Delta$  - zvolená miera neistoty,

**Output:**  $\Theta_I$  - množina nekonzistentných objektov,  
 $\Theta_M$  - množina hodnôt mier neistoty  $\forall \varphi : \varphi \in \Theta_I$

**Define:**  $\sigma$  - objekt z množiny  $\Theta$

$\varphi$  - pomocná premenná

$\Delta$  - uvažovaná miera neistoty  $\Delta(\varphi) = \max_{\varphi \in \Theta} \{\Delta^-(\varphi), \Delta^+(\varphi)\}$

```

1: function FINDINCONSISTENTOBJECT( $\Theta, \Delta$ )
2:    $\Theta_I := \emptyset, \Theta_M := \emptyset$ 
3:   for all  $\sigma \in \Theta$  do
4:     if  $\sigma \in X_1$  then                                     ▷ 1 je najnižšia rozhodovacia trieda  $X$ 
5:        $\varphi := \Delta^-(\sigma)$ 
6:     else
7:       if  $\sigma \in X_n$  then                                     ▷  $n$  je najvyššia rozhodovacia trieda  $X$ 
8:          $\varphi := \Delta^+(\sigma)$ 
9:       else                                                       ▷ pre  $\sigma \in X_i, i = 2, \dots, n - 1$ 
10:         $\varphi := \Delta(\sigma)$ 
11:      end if
12:    end if
13:    if  $\varphi > 0$  then                                           ▷  $\sigma$  je nekonzistentné
14:       $\Theta_I := \Theta_I \cup \{\sigma\}$ 
15:       $\Theta_M := \Theta_M \cup \{\varphi\}$ 
16:    end if
17:  end for
18:  return  $[\Theta_I, \Theta_M]$ 
19: end function

```

```

1: function FINDINCONSISTENTOBJECT( $\Theta, \Delta$ )
2:    $\Theta_I := \emptyset, \Theta_M := \emptyset$ 
3:   for all  $\sigma \in \Theta$  do
4:     if  $\sigma \in X_1$  then
5:        $\varphi := \Delta^-(\sigma)$ 
6:     else
7:       if  $\sigma \in X_n$  then
8:          $\varphi := \Delta^+(\sigma)$ 
9:       else
10:         $\varphi := \Delta(\sigma)$ 
11:      end if
12:    end if
13:    if  $\varphi > 0$  then
14:       $\Theta_I := \Theta_I \cup \{\sigma\}$ 
15:       $\Theta_M := \Theta_M \cup \{\varphi\}$ 
16:    end if
17:  end for
18:  return  $[\Theta_I, \Theta_M]$ 
19: end function

```

▷ 1 je najnižšia rozhodovacia trieda  $X$

▷  $n$  je najvyššia rozhodovacia trieda  $X$

▷ pre  $\sigma \in X_i, i = 2, \dots, n - 1$

▷  $\sigma$  je nekonzistentné

**Pôvodne 7 podmienok IF**