FUZZY DECISION TREES (FDT)

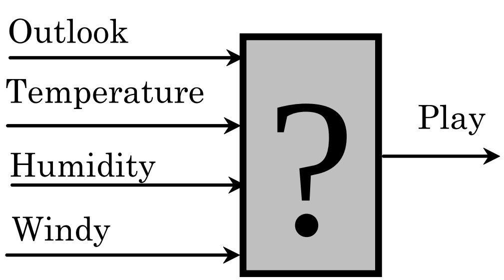
DECISION MAKING SUPPORT SYSTEM BASED ON FDT

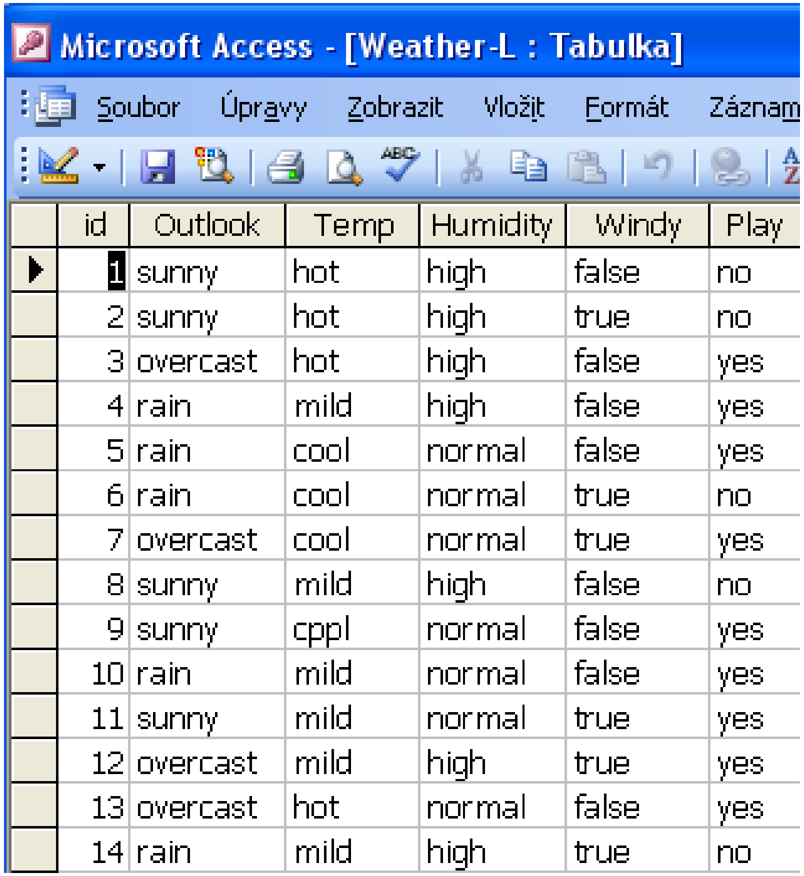
Decision Support Systems

Decision Support Systems are a specific class of computer-based information systems that support your decision-making activities. A decision support system analyzes data and provide interactive information support to professionals during the decision-making process.

Decision making implies selection of the best decision from a set of possible options. In some cases, this selection is based on past experience. Past experience is used to analyze the situations and the choice made in these situations.

DECISION MAKING BY ANALYSIS OF PREVIOUS SITUATIONS




Microsoft Access - [Weather-L : Tabulka]

Soubor Úpravy Zobrazit Vložit Formát Záznam

id	Outlook	Temp	Humidity	Windy	Play
1	sunny	hot	high	false	no
2	sunny	hot	high	true	no
3	overcast	hot	high	false	yes
4	rain	mild	high	false	yes
5	rain	cool	normal	false	yes
6	rain	cool	normal	true	no
7	overcast	cool	normal	true	yes
8	sunny	mild	high	false	no
9	sunny	cppl	normal	false	yes
10	rain	mild	normal	false	yes
11	sunny	mild	normal	true	yes
12	overcast	mild	high	true	yes
13	overcast	hot	normal	false	yes
14	rain	mild	high	true	no

Our goal is **building a model** for the **recognition** of the new situation:

Outlook	Temperature	Humidity	Windy	Play
sunny	cool	high	true	?

Several Models for Solving this Task

- k-nearest neighbors (k-NN)
- Regression Models a Support Vector Machine
- Naïve-Bayes Classifications Models
- Neural Networks
- **Decision Trees**

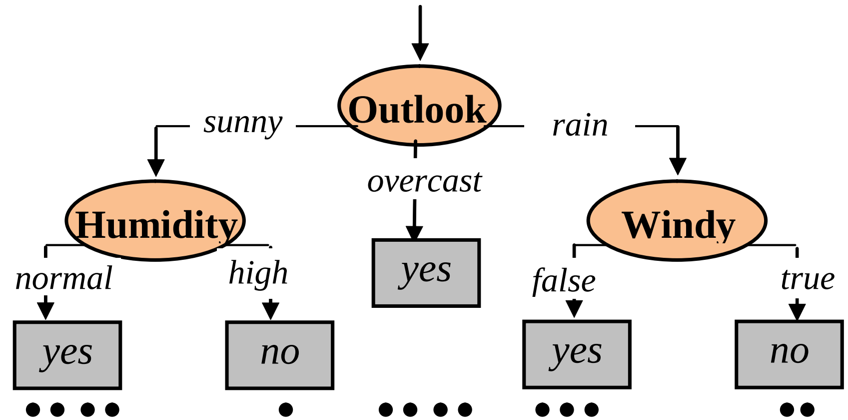
Decision Trees (1). Introduction

Decision Tree is a flow-chart like structure in which internal node represents test on an attribute, each branch represents outcome of test and each leaf node represents class label.

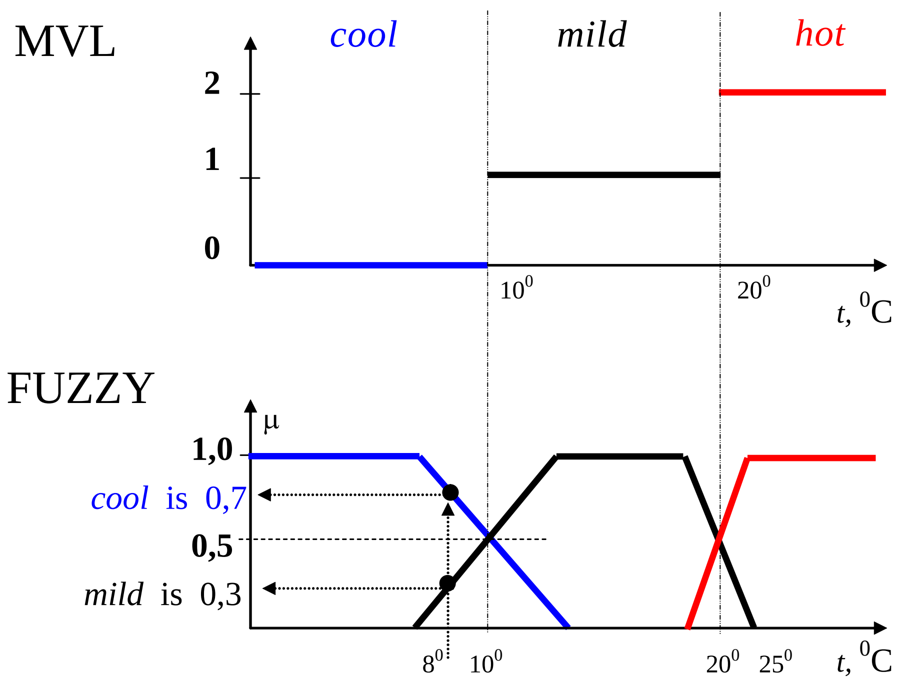
Microsoft Access - [Weather-L : Tabulka]

Soubor Úpravy Zobrazit Vložit Formát Záznam

id	Outlook	Temp	Humidity	Windy	Play
1	sunny	hot	high	false	no
2	sunny	hot	high	true	no
3	overcast	hot	high	false	yes
4	rain	mild	high	false	yes
5	rain	cool	normal	false	yes
6	rain	cool	normal	true	no
7	overcast	cool	normal	true	yes
8	sunny	mild	high	false	no
9	sunny	cppl	normal	false	yes
10	rain	mild	normal	false	yes
11	sunny	mild	normal	true	yes
12	overcast	mild	high	true	yes
13	overcast	hot	normal	false	yes
14	rain	mild	high	true	no



MVL VS. FUZZY. CLUSTERING

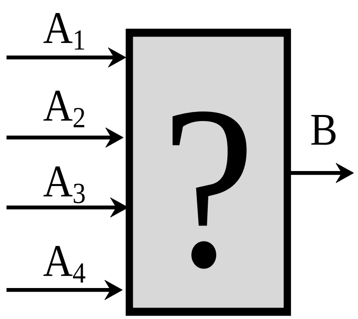


Numeric attribute A_2	*Linguistic attribute* A_2 (*Temperature*)		
	$A_{2,1}$ (*cool*)	$A_{2,2}$ (*mild*)	$A_{2,3}$ (*hot*)
$x_1 = 8$	0.7	0.3	0.0
$x_2 = 10$	0.5	0.5	0.0
$x_3 = 15$	0.0	1.0	0.0
$x_4 = 20$	0.0	0.5	0.5
$x_5 = 23$	0.0	0.8	0.2
$x_6 = 25$	0.0	0.0	1.0

An object *x* can belong simultaneously to more than one class and do so to varying degrees called memberships

H.-M. Lee, C.M. Chen, etc, An Efficient Fuzzy Classifier with Feature Selection Based on Fuzzy Entropy, *Journal of IEEE Trans. on Systems, Man and Cybernetics*, Part B, vol.31(3), 2001, pp. 426-432

FUZZY DATA *



N	Attribute A_1			Attribute A_2			Attribute A_3		Attribute A_4		Output B		
	A_{11}	A_{12}	A_{13}	A_{21}	A_{22}	A_{23}	A_{31}	A_{32}	A_{41}	A_{42}	B_1	B_2	B_3
Cost	*Cost* (A_1)=2.5			*Cost* (A_2)=2.0			*Cost*(A_3)=1.7		*Cost*(A_4)=1.8				
1.	0.9	0.1	0.0	**1.0**	**0.0**	**0.0**	0.8	0.2	0.4	0.6	0.0	0.8	0.2
2.	0.8	0.2	0.0	0.6	0.4	0.0	0.0	1.0	0.0	1.0	0.6	0.4	0.0
3.	0.0	0.7	0.3	0.8	0.2	0.0	0.1	0.9	0.2	0.8	0.3	0.6	0.1
4.	0.2	0.7	0.1	0.3	0.7	0.0	0.2	0.8	0.3	0.7	0.9	0.1	0.0
5.	0.0	0.1	0.9	**0.7**	**0.3**	**0.0**	0.5	0.5	0.5	0.5	0.0	0.0	1.0
6.	0.0	0.7	0.3	0.0	0.3	0.7	0.7	0.3	0.4	0.6	0.2	0.0	0.8
7.	0.0	0.3	0.7	0. 0	0.0	1.0	0.0	1.0	0.1	0.9	0.0	0.0	1.0
8.	0.0	1.0	0.0	0.0	0.2	0.8	0.2	0.8	0.0	1.0	0.7	0.0	0.3
9.	1.0	0.0	0.0	1.0	0.0	0.0	0.6	0.4	0.7	0.3	0.2	0.8	0.0
10.	0.9	0.1	0.0	0.0	0.3	0.7	0.0	1.0	0.9	0.1	0.0	0.3	0.7
11.	0.7	0.3	0.0	1.0	0.0	0.0	1.0	0.0	0.2	0.8	0.3	0.7	0.0
12.	0.2	0.6	0.2	0.0	1.0	0.0	0.3	0.7	0.3	0.7	0.7	0.2	0.1
13.	0.9	0.1	0.0	0.2	0.8	0.0	0.1	0.9	1.0	0.0	0.0	0.0	1.0
14.	0.0	0.9	0.1	0.0	0.9	0.1	0.1	0.9	0.7	0.3	0.0	0.0	1.0
15.	0.0	0.0	1.0	0.0	0.0	1.0	1.0	0.0	0.8	0.2	0.0	0.0	1.0
16.	1.0	0.0	0.0	0.5	0.5	0.0	0.0	1.0	0.0	1.0	0.5	0.5	0.0
17.											?	?	?

*Y. Yuan, M.J. Shaw, Induction of Fuzzy Decision Trees, *Fuzzy Sets and Systems*, **69**, 1995, pp.125-139

Measuring the value of each input attribute requires resource costs (money or time):

Cost (A_1), *Cost* (A_2), *Cost* (A_3), *Cost* (A_4) .

Our goal is find a method for transform values of input attributes into the value of output attribute with *minimal resources:*

$$\textit{sum Cost}\ (A_i) \to \textit{minimum}$$

Algorithms ID3 and C4.5 (by Prof. Ross Quinlan)

We compared the information gain and classical concepts of information theory (information and entropy).

These mathematical expression are similar and common.

Algorithm	Prof. Ross Quinlan	Information theory
ID3	$Gain(\mathrm{A}) = \sum_{l=1}^{m_b} -\frac{N_{b_l}}{N}\log_2\frac{N_{b_l}}{N} - \sum_{j=1}^{m_a}\frac{N_{a_j}}{N}\times\sum_{l=1}^{m_b} -\frac{N_{b_l/a_j}}{N_{a_j}}\log_2\frac{N_{b_l/a_j}}{N_{a_j}}$	**I(B;A) = H(B) – H(B|A)** (Abs.)
C4.5	$GainRatio(\mathrm{A}) = Gain(\mathrm{A}) / SplitInfo(\mathrm{A})$, $SplitInfo(\mathrm{A}) = \sum_{j=1}^{m_a} -\frac{N_{a_j}}{N}\times\log_2\frac{N_{a_j}}{N}$.	**s(A_i|B) = I(B;A) / H(A)** (Rel.)

Review of information estimation

Entropy Calculation

$$H = -\sum_{i=1}^{m_i} p_i \times \log_2 p_i$$

Another Entropies:
Hybrid, Yager, Koufmann, Kosko, ...

$$-\sum_{i=1}^{m_i} \frac{\sum_{j=1}^{N} \mu_{i,j}}{N} \times \log_2 \frac{\sum_{j=1}^{N} \mu_{i,j}}{N}$$

- H.Ichihashi, 1996
- H-M.Lee, 2001

$$-\frac{1}{N}\sum_{i=1}^{m_i}\sum_{j=1}^{N}\left(\mu_{i,j} \times \log_2 \mu_{i,j} + \left(1-\mu_{i,j}\right) \times \log_2\left(1-\mu_{i,j}\right)\right)$$

- A.de Luca and S.Termini, 1972
- Y.Yuan and M.Shaw, 1995
- X.Wang etc, 2000

New Cumulative Information Estimations

We have proposed new cumulative information estimations

	Personal	*Joint*	*Conditional*	*Mutual*
Information	$\mathbf{I}(A_{i1,j1})$	$\mathbf{I}(A_{i2,j2}, A_{i1,j1})$	$\mathbf{I}(A_{i2,j2} \mid A_{i1,j1})$	$\mathbf{I}(A_{i2,j2}; A_{i1,j1})$
Entropy	$\mathbf{H}(A_{i1})$	$\mathbf{H}(A_{i2}, A_{i1})$	$\mathbf{H}(A_{i2} \mid A_{i1})$	$\mathbf{I}(A_{i2}; A_{i1})$

Levashenko V., Zaitseva E. Usage of new information estimations for induction of fuzzy decision trees. *Intelligent Data Engineering and Automated Learning, Lecture Notes in Computer Science*, **2412**, 2002, 493-499

New Criteria of Choice Expanded Attributes

Unordered FDT

$$\frac{\mathbf{I}(B;\, A_{i1,j1},\ldots,A_{iq\text{-}1,\, j\, q\text{-}1},\, A_{iq})}{Cost\ (A_{iq})} \rightarrow max$$

Ordered FDT

$$\frac{\mathbf{I}(B;\, A_{i1},\ldots,A_{iq\text{-}1},\, A_{iq})}{Cost\ (A_{iq})} \rightarrow max$$

Stable FDT

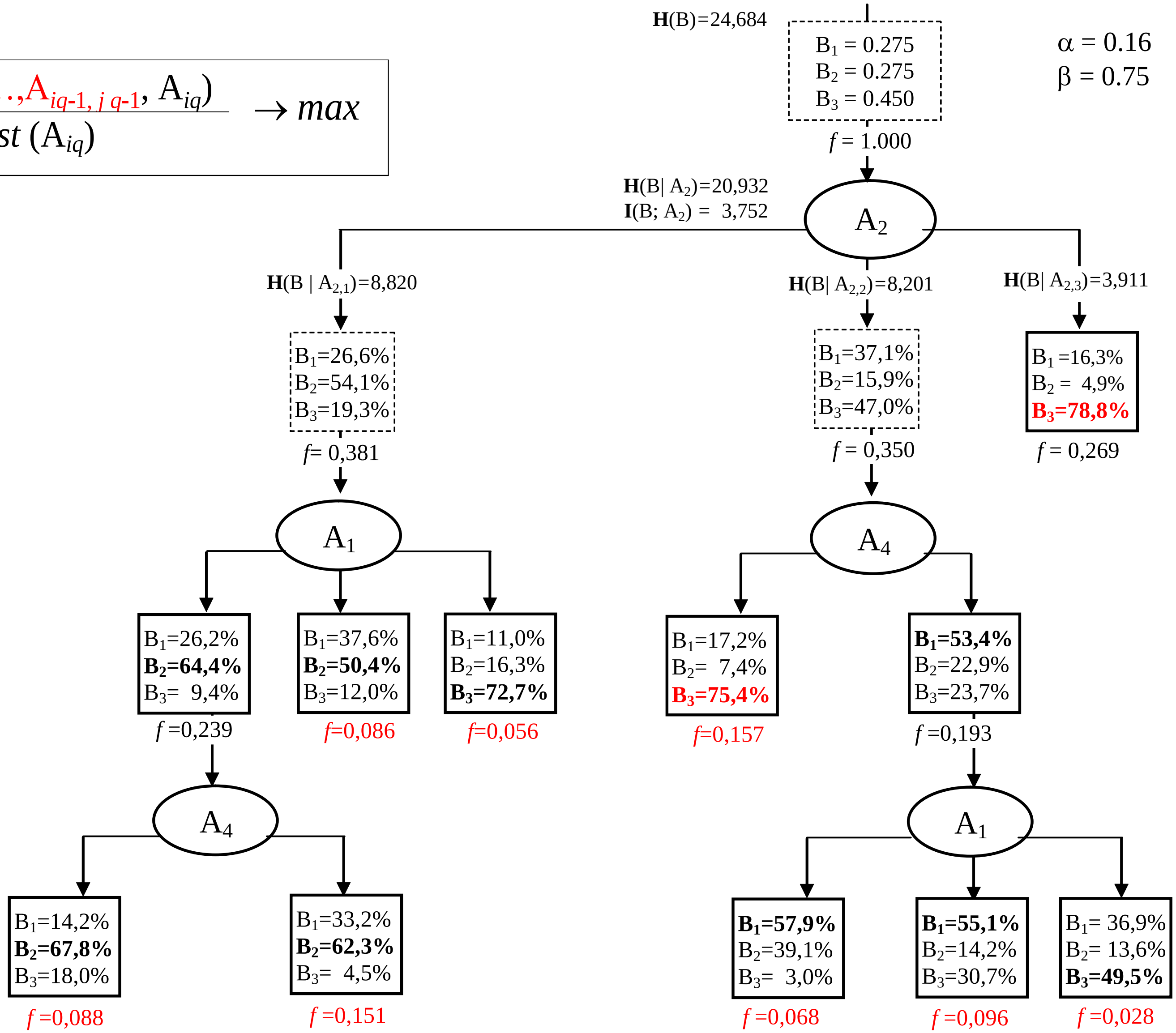
$$\frac{\mathbf{I}(A_{iq};\, B,\, A_{i1},\ldots,A_{iq\text{-}1})}{Cost\ (A_{iq})} \rightarrow max$$

etc.

Levashenko V., Zaitseva E., Fuzzy Decision Trees in Medical Decision Making Support System, *Proc. of the IEEE Fed. Conf. on Computer Science and Information Systems*, 2012, 213-219

UNORDERED FUZZY DECISION TREES

$$\frac{\mathbf{I}(B;\ A_{i1,j1},\ldots,A_{iq-1,\,j\,q-1},\ A_{iq})}{Cost\ (A_{iq})} \rightarrow max$$



Fuzzy Decision Rules (1). A priori

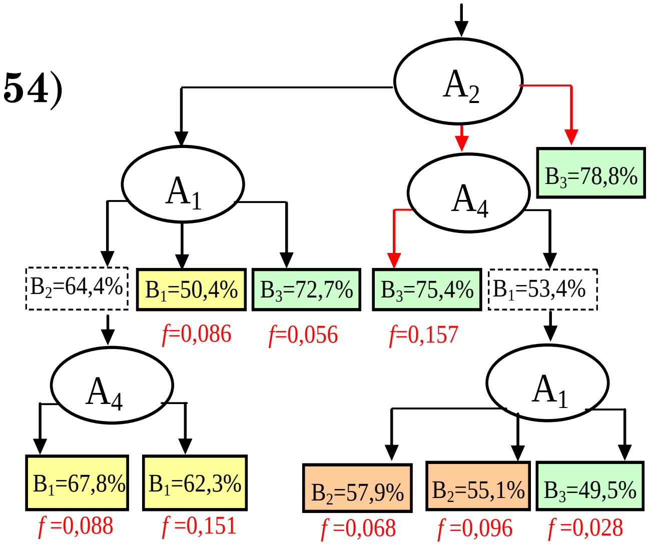
- **Fuzzy Decision Rule is path from root to leaf:**

If (A_2 is $A_{2,3}$) **then** B (with degree of truth [0.169 0.049 **0.788**])

If (A_2 is $A_{2,2}$) and (A_4 is $A_{4,1}$) **then** B is B_3 (with degree of truth **0.754**)

…..

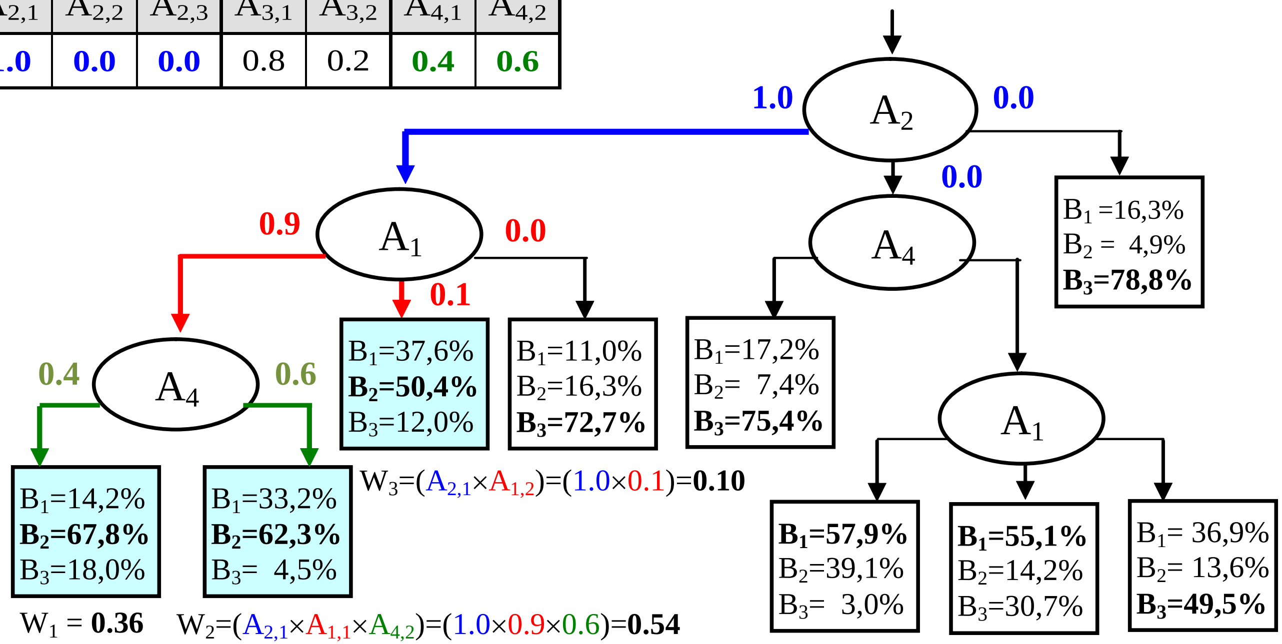
Input attribute A_3 have not influence to attribute B (for given thresholds $\alpha = 0{,}16$ и $\beta = 0{,}75$).



FUZZY DECISION RULES (2). A POSTERIORI

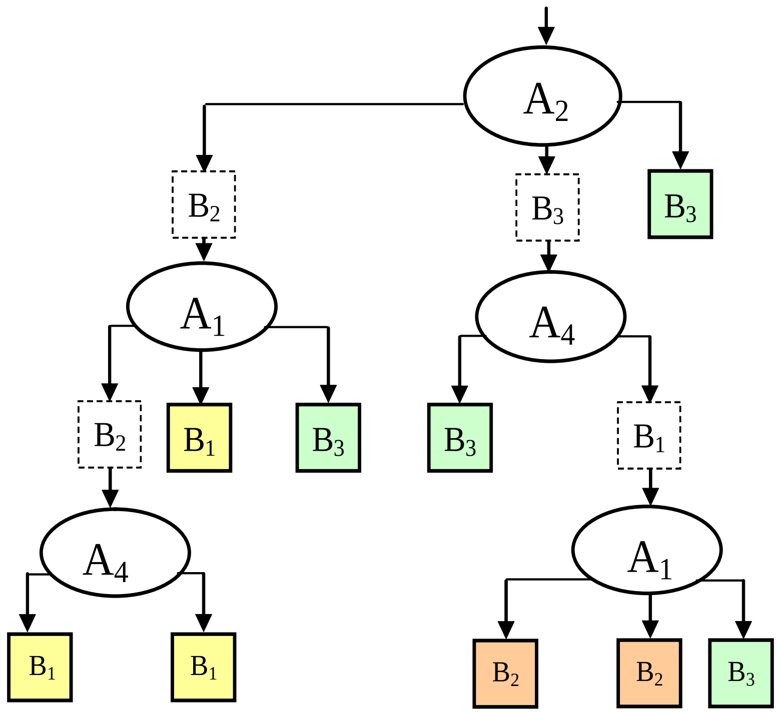
- Fuzzy Decision Rule is path from root to leaf
- One example describes by several Fuzzy Decision Rules

A_1			A_2			A_3		A_4	
$A_{1,1}$	$A_{1,2}$	$A_{1,3}$	$A_{2,1}$	$A_{2,2}$	$A_{2,3}$	$A_{3,1}$	$A_{3,2}$	$A_{4,1}$	$A_{4,2}$
0.9	**0.1**	**0.0**	**1.0**	**0.0**	**0.0**	0.8	0.2	**0.4**	**0.6**



$$\begin{bmatrix} B_1=14{,}2\% \\ B_2=67{,}8\% \\ B_3=18{,}0\% \end{bmatrix} \times \mathbf{0.36} + \begin{bmatrix} B_1=33{,}2\% \\ B_2=62{,}3\% \\ B_3=\ 4{,}5\% \end{bmatrix} \times \mathbf{0.54} + \begin{bmatrix} B_1=37{,}6\% \\ B_2=50{,}4\% \\ B_3=12{,}0\% \end{bmatrix} \times \mathbf{0.10} = \begin{bmatrix} B_1=26{,}8\% \\ \mathbf{B_2=63{,}1\%} \\ B_3=10{,}1\% \end{bmatrix}$$

Decision Tables

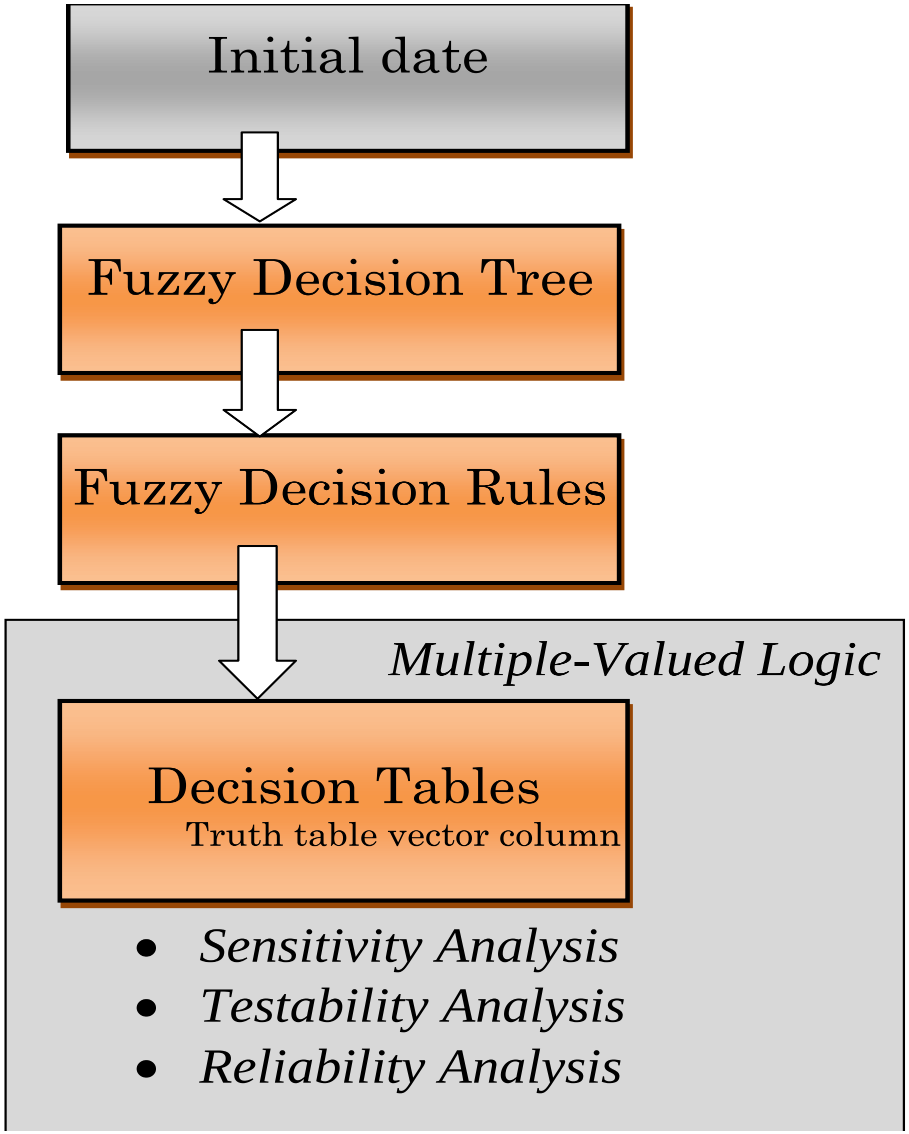


A_1	A_2	A_3	A_4	**X**
0	0	0	0	**1**
0	0	0	1	**1**
0	0	1	0	**1**
0	0	1	1	**1**
0	1	0	0	**2**
0	1	0	1	**0**
0	1	1	0	**2**
0	1	1	1	**0**
0	2	0	0	**2**
0	2	0	1	**2**
0	2	1	0	**2**
0	2	1	1	**2**
1	0	0	0	**1**
1	0	0	1	**1**
1	0	1	0	**1**
1	0	1	1	**1**
1	1	0	0	**2**
1	1	0	1	**0**
1	1	1	0	**2**
1	1	1	1	**0**
1	2	0	0	**2**
1	2	0	1	**2**
1	2	1	0	**2**
1	2	1	1	**2**
2	0	0	0	**2**
2	0	0	1	**2**
2	0	1	0	**2**
2	0	1	1	**2**
2	1	0	0	**2**
2	1	0	1	**2**
2	1	1	0	**2**
2	1	1	1	**2**
2	2	0	0	**2**
2	2	0	1	**2**
2	2	1	0	**2**
2	2	1	1	**2**

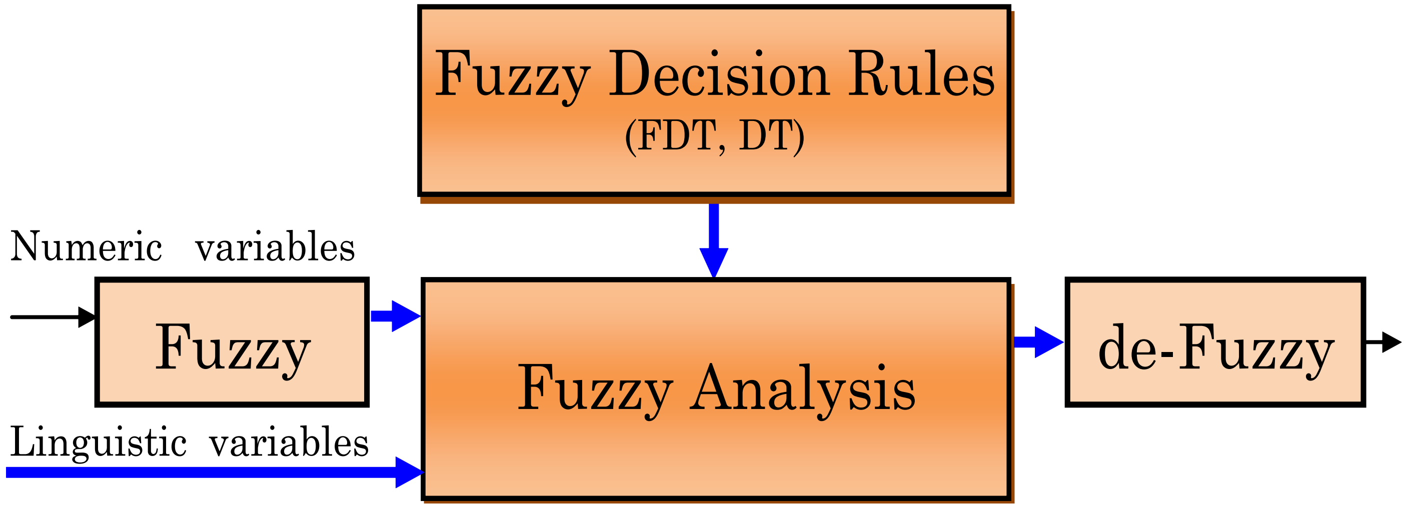
Truth table vector column:

$$\mathbf{X} = [1111\ 2020\ 2222 \quad 1111\ 2020\ 2222 \quad 2222\ 2222\ 2222]^T$$

Basic of Knowledge Representation

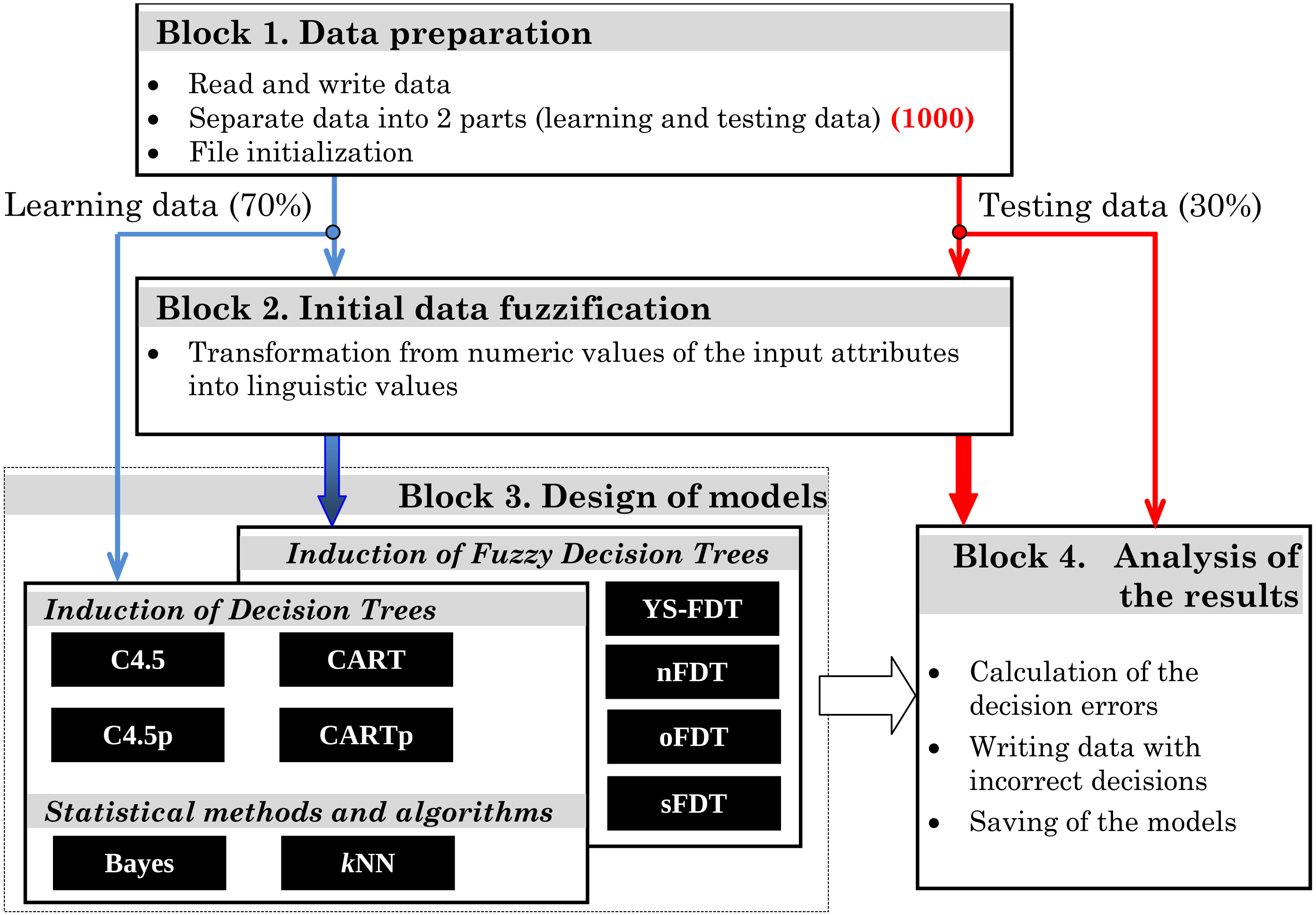


FUZZY DECISION MAKING SUPPORT SYSTEM



SOFTWARE FOR EXPERIMENTAL INVESTIGATIONS

We create software application Multiprognos by C++ ver. 5.02



Results of Experiment

		Initial parameters			15-0	5-4	18-1	8-1	7-13	6-20	0,1884	0,4793	0,2135	0,3065	0,5106	0,6465
	DB Name	TS	NoIA	NoOI	nFDT	YS	C45p	CART	Bayes	kNN	nFDT	YS	C45p	CART	Bayes	kNN
1	abalone	4177	8	28	0,7625	0,7663	0,7299	0,7452	0,7567	0,842	0,2908	0,3247	0	0,1365	0,2391	1
2	balance	625	4	3	0,122	0,3282	0,2213	0,2364	0,1029	0,3769	0,0697	0,8223	0,4321	0,4872	0	1
3	blood	748	4	2	0,2363	0,2389	0,2263	0,2281	0,2493	0,3953	0,0592	0,0746	0	0,0107	0,1361	1
4	breast	106	9	6	0,4017	0,4454	0,359	0,3567	0,361	0,3586	0,5177	1	0,0464	0,021	0,0684	0,0419
5	bupa	345	6	2	0,4105	0,4224	0,3517	0,3527	0,4448	0,4339	0,6316	0,7594	0	0,0107	1	0,8829
6	car	1728	6	4	0,1147	0,2195	0,0831	0,1731	0,1466	0,1338	0,2317	1	0	0,6598	0,4655	0,3717
7	cmc	1473	9	3	0,5209	0,5292	0,47	0,4959	0,5063	0,6186	0,3425	0,3984	0	0,1743	0,2443	1
8	crx	653	15	2	0,1323	0,1333	0,1348	0,1357	0,2312	0,2799	0,0617	0,068	0,0776	0,0833	0,6904	1
9	diagnosis	120	6	2	0,0002	0,1578	0,0072	0,0074	0,0476	0	0,0013	1	0,0456	0,0469	0,3016	0
10	ecoli	336	7	8	0,1816	0,2408	0,2918	0,2036	0,152	0,279	0,1008	0,3023	0,476	0,1757	0	0,4324
11	faults	1941	27	7	0,3104	0,3016	0,2598	0,2859	0,4005	0,3593	0,3596	0,2971	0	0,1855	1	0,7072
12	forestfires	517	12	3	0,5217	0,5214	0,5218	0,5179	0,5617	0,6355	0,0323	0,0298	0,0332	0	0,3724	1
13	german	1000	20	2	0,2619	0,2977	0,2756	0,2828	0,2534	0,3503	0,0877	0,4572	0,2291	0,3034	0	1
14	german24	1000	24	2	0,257	0,2908	0,2698	0,2683	0,2796	0,3928	0	0,2489	0,0943	0,0832	0,1664	1
15	glass	214	9	7	0,3731	0,4491	0,3307	0,3335	0,5361	0,354	0,2103	0,5785	0,0048	0,0184	1	0,1177
16	haberman	306	3	2	0,2652	0,2462	0,2749	0,2625	0,2537	0,4077	0,1176	0	0,1777	0,1009	0,0464	1
17	heart	270	13	2	0,1852	0,2575	0,2443	0,221	0,1635	0,3126	0,1455	0,6304	0,5419	0,3856	0	1
18	ilpd	579	10	2	0,285	0,2854	0,2837	0,2828	0,4419	0,3991	0,0138	0,0163	0,0057	0	1	0,731
19	image	2100	18	7	0,1385	0,1805	0,0443	0,0827	0,2043	0,071	0,5961	0,8539	0,0178	0,2535	1	0,1817
20	ionosphere	351	33	2	0,1217	0,1146	0,1081	0,0972	0,183	0,1375	0,2855	0,2028	0,127	0	1	0,4697
21	iris	150	4	3	0,036	0,0356	0,0525	0,0493	0,046	0,0699	0,0117	0	0,4927	0,3994	0,3032	1
22	magic04	19020	10	2	0,1638	0,1837	0,1344	0,1558	0,2262	0,2805	0,2012	0,3374	0	0,1465	0,6283	1
23	monks1	556	6	2	0,1227	0,1111	0,1574	0,1667	0,287	0,3889	0,3155	0,2857	0,4047	0,4286	0,738	1
24	monks2	601	6	2	0,2708	0,3449	0,3218	0,3194	0,3843	0,2199	0,3096	0,7603	0,6198	0,6052	1	0
25	monks3	554	6	2	0	0,1111	0,0278	0,2222	0,0278	0,2685	0	0,3608	0,0903	0,7217	0,0903	0,872
26	nursery	12960	8	5	0,0853	0,0833	0,0289	0,4271	0,0926	0,0982	0,132	0,1273	0	0,9321	0,1491	0,1622
27	parkinsons	195	22	2	0,0959	0,1469	0,1498	0,1289	0,2993	0,0839	0,0557	0,2925	0,3059	0,2089	1	0
28	pendigits	10992	16	10	0,0834	0,0926	0,0795	0,0909	0,1593	0,0247	0,4361	0,5045	0,4071	0,4918	1	0
29	pima	768	8	2	0,2417	0,2527	0,2547	0,2544	0,2462	0,3585	0,0135	0,1064	0,1233	0,1208	0,0515	1
30	seeds	210	7	3	0,0909	0,0968	0,0776	0,0939	0,0973	0,0933	0,4043	0,5836	0	0,4954	0,5988	0,4772
31	sonar	208	60	2	0,1593	0,2879	0,2679	0,2657	0,313	0,173	0	0,8367	0,7066	0,6923	1	0,0891
32	tae	151	5	3	0,3583	0,5151	0,447		0,542	0,4977	0,2213	0,886	0,5973		1	0,8122
33	thyroid	215	5	3	0,0739	0,1059	0,078	0,0955	0,0339	0,0568	0,5556	1	0,6125	0,8556	0	0,3181
34	tic	958	9	2	0,1346	0,2813	0,1441	0,1373	0,2958	0,3472	0	0,69	0,0447	0,0127	0,7582	1
35	venicle	846	18	4	0,3303	0,4108	0,2896	0,3048	0,5458	0,366	0,1746	0,483	0,0188	0,077	1	0,3114
36	vertebral2	310	6	2	0,1714	0,1818	0,1916	0,1892	0,2202	0,2846	0	0,0919	0,1784	0,1572	0,4311	1
37	vertebral3	310	6	3	0,1993	0,3027	0,1915	0,1903	0,1788	0,3412	0,1262	0,7629	0,0782	0,0708	0	1
38	vowel	990	10	11	0,3447	0,4892	0,2424	0,3959	0,3394	0,0461	0,6739	1	0,443	0,7894	0,6619	0
39	wdbc	569	30	2	0,0368	0,0705	0,0641	0,0704	0,0674	0,0719	0	0,816	0,661	0,8136	0,7409	0,8499
40	wine	178	13	3	0,0325	0,0432	0,0923	0,0973	0,0288	0,0678	0,054	0,2102	0,927	1	0	0,5693
41	wine-red	1599	11	6	0,4106	0,4467	0,4072	0,4291	0,4567	0,4771	0,0486	0,5651	0	0,3133	0,7082	1
42	wine-white	4898	11	7	0,4749	0,497	0,4333	0,4784	0,5568	0,4773	0,3819	0,5487	0,0679	0,4083	1	0,4
43	wpbc	194	33	2	0,221	0,249	0,24	0,2334	0,3378	0,4269	0	0,136	0,0923	0,0602	0,5673	1
44	yeast	1484	8	10	0,4468	0,5077	0,4262	0,4383	0,4325	0,5944	0,1225	0,4845	0	0,0719	0,0375	1

Machine Learning Repository

http://archive.ics.uci.edu/ml/

Normalization into [0; 1]

$$x_{new} = \frac{x_{old} - x_{min}}{x_{max} - x_{min}}$$

Results of Experiment (2)

- We have selected 44 databases from Machine Learning Repository http://archive.ics.uci.edu/ml/

  blood, breast, bupa, cmc, diagnosis, haberman, heart, ilpd, nursery, parkinsons, pima, thyroid, vertebral2, vertebral3, wdbc, wpbc, etc.

- We have calculated normalized error of misclassification:

Methods' Name	Rating
Fuzzy Decision Trees	0,1884
Method C4.5	0,2135
Method CART	0,3065
Fuzzy Decision Trees (Y. Yuan & M. Show)	0,4793
Naïve-Bayes Classifications Models	0,5106
k-nearest neighbors	0,6465

Implementation into Projects

1. **Project FP7-ICT-2013-10**. Regional Anesthesia Simulator and Assistant (RASimAs), Reg. No. 610425, **Nov. 2013-2016**.
2. **Project APVV SK-PL**. Support Systems for Medical Decision Making, Reg. No. SK-PL-0023-12, **2013-2014**.
3. **Project TEMPUS**. Green Computing & Communications (GreenCo), Reg. No. 530270-TEMPUS-1-2012-1-UK-TEMPUS-JPCR, **2012-2015**.
4. **Project NATO**. Intelligent Assistance Systems: Multisensor Processing and Reliability Analysis, NATO Collaborative Linkage Grant, Canada-Slovensko-Czech-Belarus, Reg. No. CBP.EAP.CLG 984, **2011-2012**.