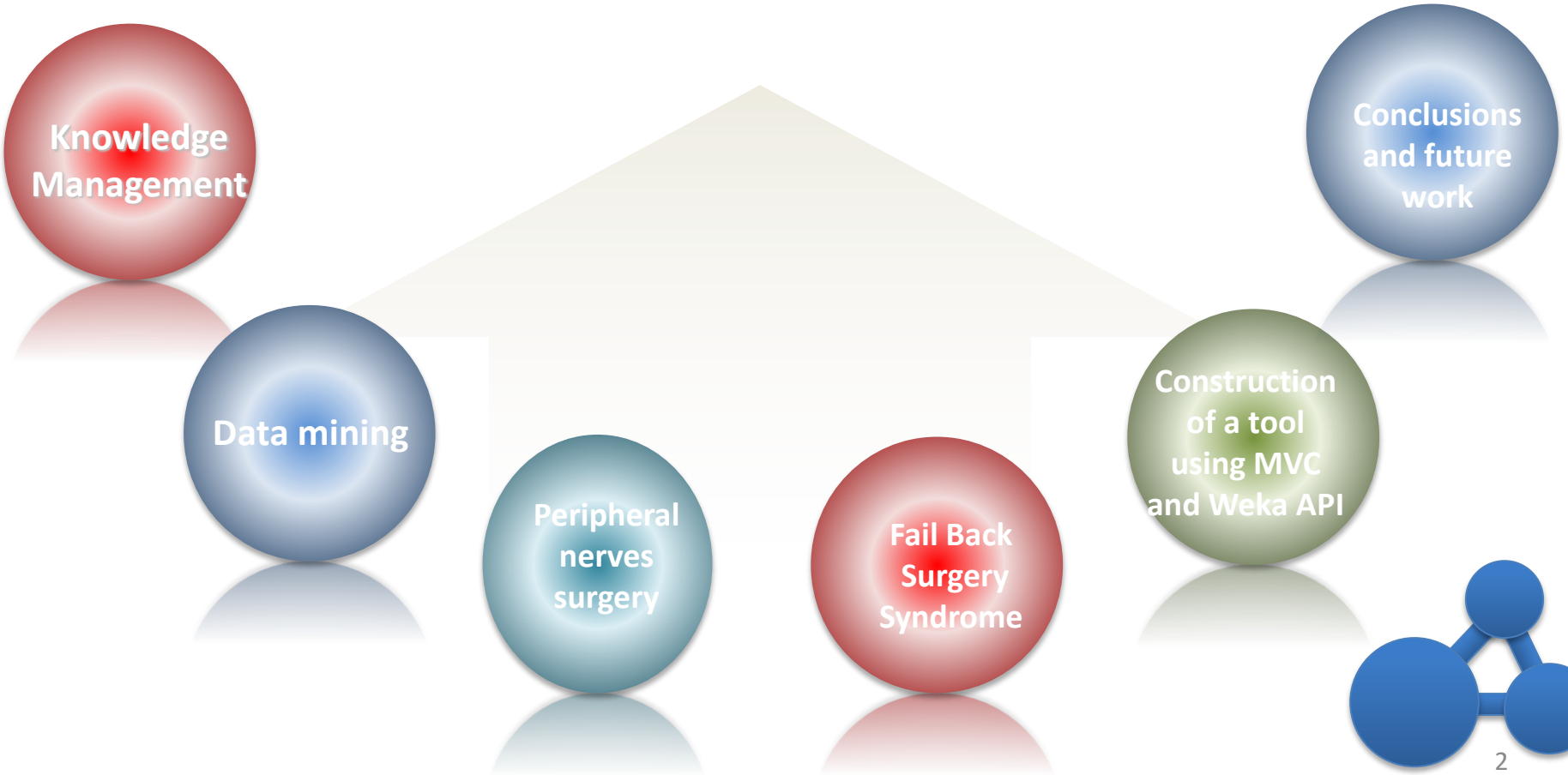
Prof.Asoc. Alda Kika
Department of Informatics
Faculty of Natural Sciences
University of Tirana

Knowledge Discovery in Medical Data

Content


Knowledge Management
Conclusions and future work
Data mining
Construction of a tool using MVC and Weka API
Peripheral nerves surgery
Fail Back Surgery Syndrome

The need to discover knowledge in medicine

The needs

- Medical Informatics and biomedical computing have grown in measure over the past decade.
- The characteristics of medical data: uncompleted, heterogeneous.
- To take clinical decision different attributes should be considered.

However, there is a lack of effective analysis tools.

Challenges of medical Informatics

- To acquire proficiency in understanding and interpreting clinical information.
- To apply this through making prudent treatment decisions.
- Predictability in surgery improves the overall results and the expectations of both surgeon and patient, influencing positively their relationship reflected also in preoperative informed consent.

Knowledge Discovery

Knowledge Management

"a processes cycle to identify, transfer, store and disseminate the knowledge in order to reuse, awareness, share and learn it across the organization"

(Efraim Turban et al., 2001; Davenport et al., 1998).

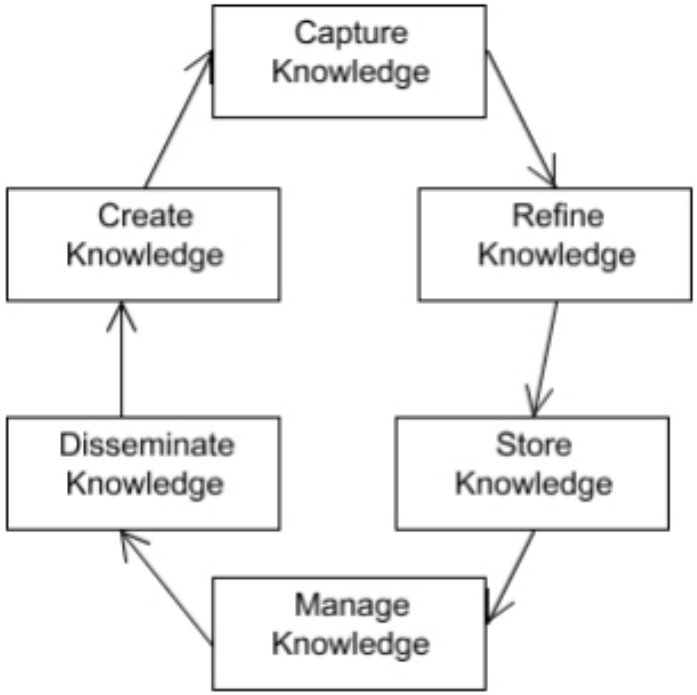
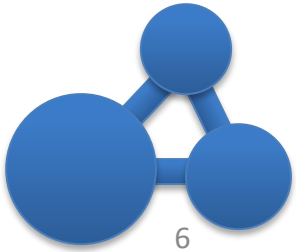


Figure 1: Processes of knowledge management

Knowledge Discovery

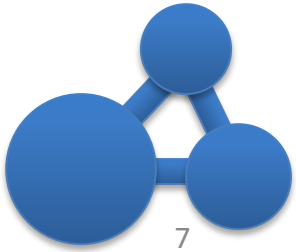
- *Knowledge Discovery* is an automatic, exploratory analysis and modeling of large data repositories. It is the organized process of identifying valid, novel, useful, and understandable patterns from large and complex data sets.

- *Data mining* (DM) is the core of the KDD process, involving the inferring of algorithms that explore the data, develop the model and discover previously unknown patterns.



Data Mining

- “Data mining is the analysis of (often large) observational data sets to find unsuspected relationships and to summarize the data in novel ways that are both understandable and useful to the data owner.” (Hall M. and Eibe F., 2009)
- Data mining classification techniques may be useful for the medical decision support.



Steps to discover knowledge

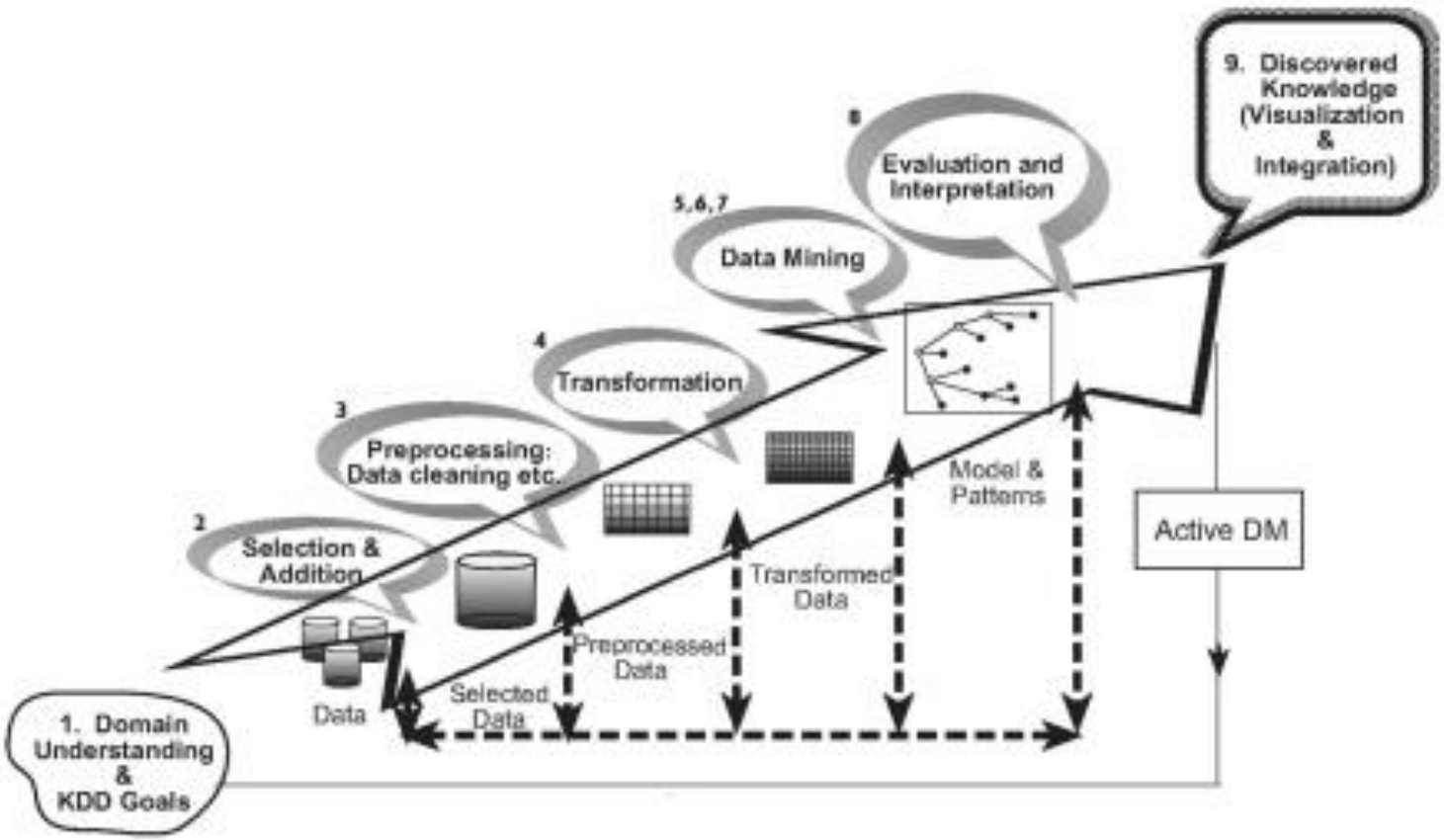


Figure 2: Steps to discover knowledge

Introduction to peripheral nerves surgery

Peripheral nerve needs surgical treatment in case of traumatic injury, entrapment or tumor.

Data of patients operated on for nerve injury at Service of Neurosurgery, Univ Hosp Centre MT, Tirana

Factors that affect nerve recovery process:

- **The time** passed from the onset of the disease to surgery
  - Minimum 0
  - Maximum 231 months
  - Mean 19.81 months
- **distance** of repaired nerve from the muscle

  The interval of the values is spread from 0 to 220 mm with an average value of 21.8 mm.

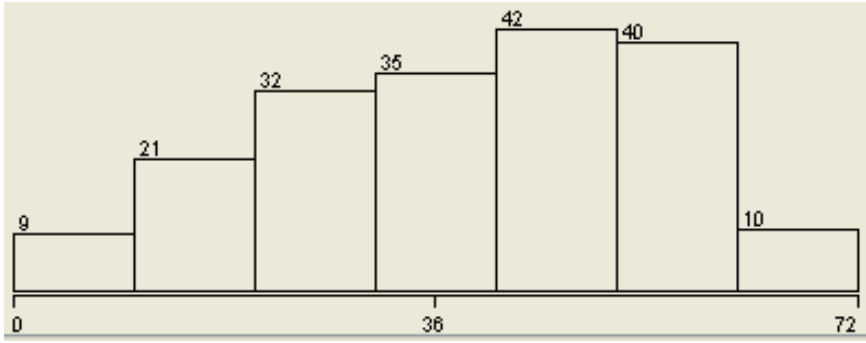
Clinical data

- **Type of nerve repair**

  Others: decompression, tumor, neuro-vascular decompression, tumor removal (100 cases)

  Neurolisis (45 cases), Graft (40 cases), Neuroraphia (4 cases).

- **Sex** - 110 males : 79 females.
- **Age**
  - Maximum 72 years old
  - Minimum <1 year old
  - Mean 38.878



Clinical data

- **Lesion type**

Lesion Type	Patients
entrapment	91
traumatic	90
tumor	8

- **Location**

Location	Patients
High, proximal	141
Low, distal	31
Brachial plexus	15

Clinical data

Graft length (numeric attribute):

The interval of the values is spread from 1 to 12 cm, with an average value of 3.2 cm.

The result of the operation (based on the valuation index) is a nominal attribute:

high (27 cases), medium (83 cases), low (79 cases).

Clinical data

Evaluation

To evaluate nerve recovery we use an integer value from **0-3:**

0 : patient doesn't feel any recovery at all

1 : the range of the recovery in the interval 0 -35%

2 : the range of the recovery is between 35% and 70 %

3 : the range of the recovery is greater than 70 %

The evaluation is done at the end of one month, three months, six months and one year.

The result is expressed as: low, middle , high.

Weka

- **WEKA(**Waikato **Environment** for Knowledge Analysis**)**
- **Weka** is a collection of machine learning algorithms for data mining tasks
- Weka is an open source Java software
- Package designed to perform the task of data mining and distributed under the terms of the GNU General Public License.

Algorithms

J48 (is the implementation of the C4.5 decision tree algorithm)

These models take a”divide-and-conquer” approach: a complex problem is decomposed in simpler sub-models and, recursively, this technique is applied to each sub-problem

Powerful classification algorithms and is often used in medicine for the prediction.

Naïve Bayes

Strong assumption that all the attributes are independent.

The algorithm tends to perform well in many class prediction scenarios.

PART

Constructs a rule set on the basis of information gain ratio.

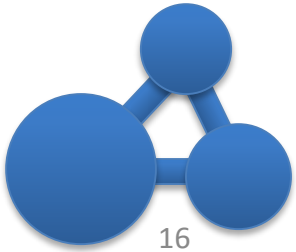
This implies that the obtained rule sets are biased with the correctness of each classification problem.

Evaluation Criteria

Let *p*i be the predicted values and *a*i the actual values

- Kappa statistics: The Kappa parameter measures pair wise agreement between two different observers, corrected for an expected chance agreement.
- Mean absolute error averages the magnitude of the individual errors without taking account of their sign.

$$\frac{|\,p1 - a1\,| + \ldots + |\,pn - an\,|}{n}$$

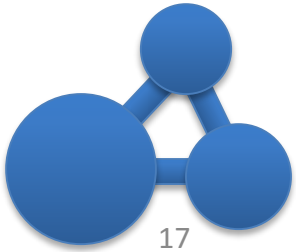


Evaluation Criteria

- The root means squared error is the square root of the average quadratic loss.

$$\sqrt{(((p1- a1)2+...+(pn- an)2)/n)}$$

- TP Rate shows the percentage of instances whose predicted values of the class attribute are identical with the actual values.
- FP Rate shows the percentage of instances whose predicted values of the class attribute are not identical with the actual values.



Evaluation Criteria

- Precision is defined as the number of positive instances retrieved over the total number of instances declared positive by the classifer.

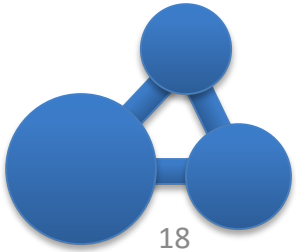
  $$Precision=TP/(TP+FP)$$

  where TP is True Positives, FP is False Positives

- Recall is defined as the number of true positive instances retrieved over the total number of instances that are positive in the set.

  $$Recall =TP/(TP+FN)$$

  where FN is False Negatives



Algorithms Comparision (10 cross validation)

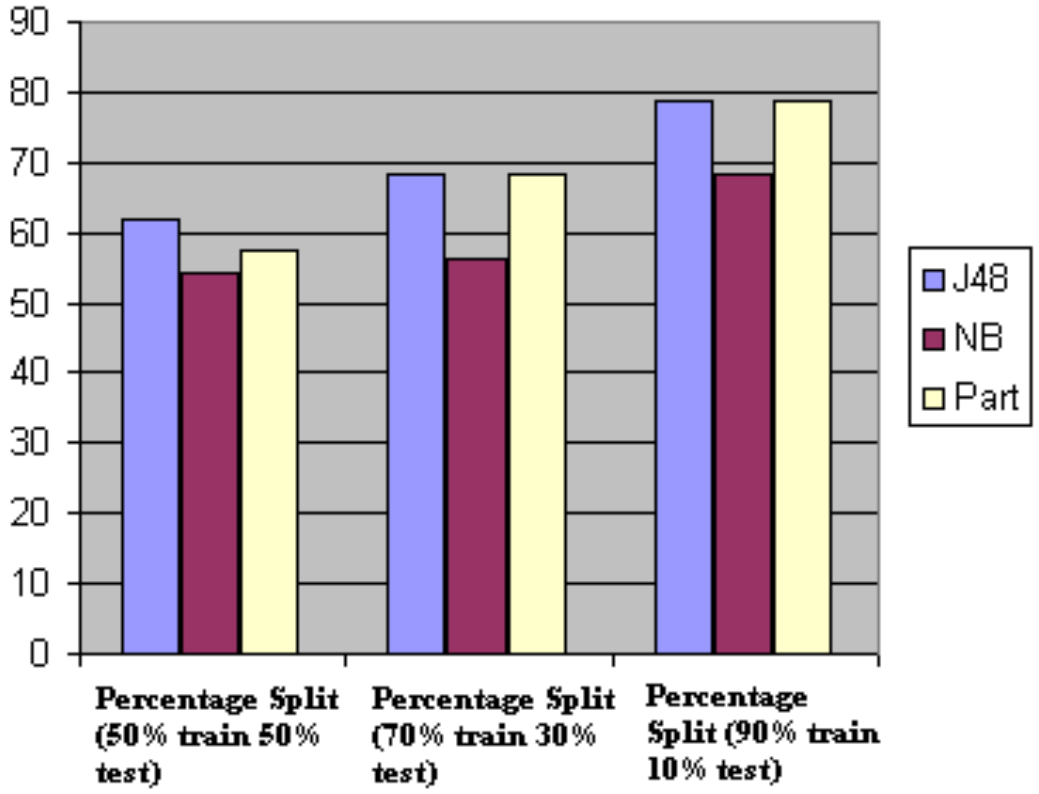
Evaluation Criteria	Classifiers		
	J48	NB	Part
Kappa statistic	0.3328	0.2374	0.2855
Mean absolute error	0.3368	0.3426	0.3435
Root Mean Squared Error	0.4389	0.4709	0.4431
Prediction accuracy	60.3175 %	52.9101 %	57.1429 %

Table 1: Comparison of Performance using 10 cross validation

Classifier	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area
J48	0.603	0.271	0.6	0.603	0.592	0.657
NB	0.529	0.28	0.579	0.529	0.528	0.669
Part	0.571	0.282	0.566	0.571	0.564	0.657

Table 2: Detailed accuracy by class (average values)

Comparison of Algorithms Accuracy using percentage split



Bayesian Network

A Bayesian network, belief network or directed acyclic graphical model is a probalistic graphical model that represents a set of random variables and their conditional dependencies via a directed acyclic graph.

The Bayesian network is defined as a couple (G,P) where G is directed acyclic graph. Each node represents an important random variable and P=P(X) represents probability distribution of these variables.

Construction of Bayesian Network

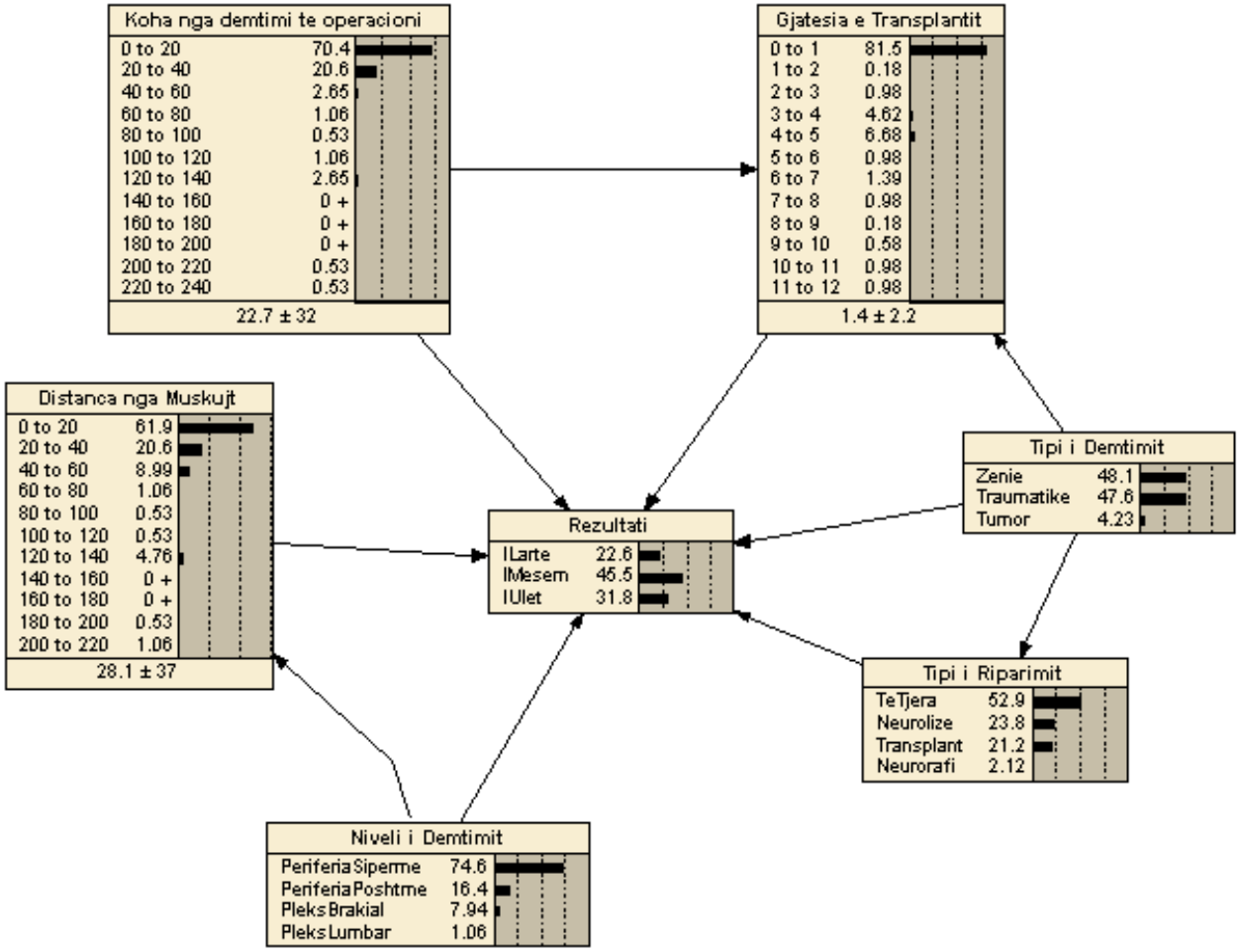
Netica program is used.

The core principle the **concept of causality**.

This concept is reflected in the model through the direction of the arches.

The result of the surgery is affected by the main factors as: type and the level of injury, time from injury to surgical repair, type of repair, distance from injury to effector muscle and length of transplanted nerve when applied.

Construction of Bayesian Network



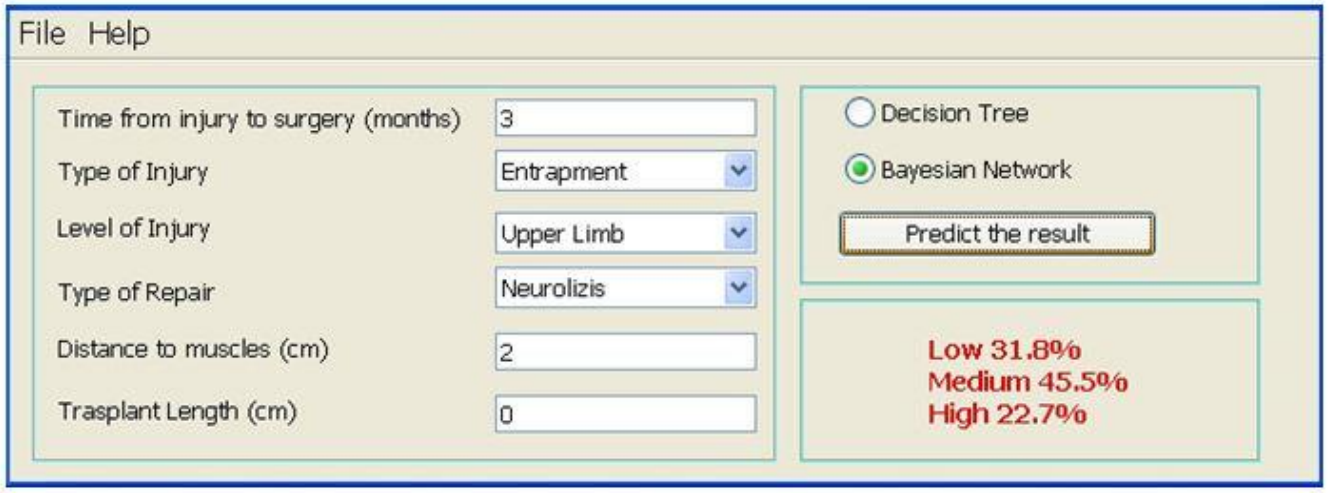
Sensitive Analysis

Sensivity is measured on the term of entropy reduction

Node	Reduction of Entropy (%)
Result	100
Distance to muscle	2.52
Time from nerve injury to surgery	1.44
Graft length	0.898
Type of repair	0.756
Location	0.668
Type of lesion	0.572

Clinical decision support systems

- A Clinical decision support systems is a computer program meant to help medical experts in taking better clinical decisions.



Conclusion and future work

- In our study, three machine learning methods were applied with the purpose to predict the results of peripheral nerve surgery. The most accurate learning method was evaluated.
- The study showed that J48 and Part algorithms have the highest accuracy, while Naïve Bayes has the most stable changing rate.

Conclusion and future work

- Bayesian Network can be used to model important components that are related to each other through conditional probability tables.
- The decision tree algorithm as well as Bayesian network is used to construct a system which will help in the clinical decision support. This will enable the surgeons to predict surgical results in preoperative period according to the specific characteristics of the single patient.

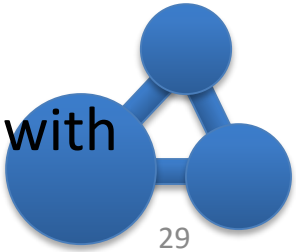
Conclusion and future work

In our future work we will try to train and test on larger and more diversified data sets (different surgeons and hospitals).

The homogeneity of the type of nerve lesion might influence the level of accuracy; therefore we are interested to try the algorithms for the separate categories.

Fail Back Surgery Syndrome

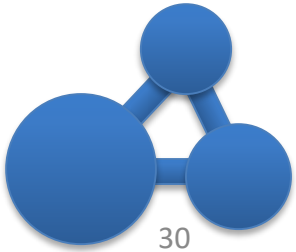
- Failed back syndrome (FBS), also called "failed back surgery syndrome" (FBSS), refers to chronic back and/or leg pain that occurs after back (spinal) surgery.
  - complex
  - multifactorial in origin
- the social and economic impact:
  - The degenerative disease of the lumbar spine has an important impact on work-day lost due to health reason
  - High treatment cost, specially if complicated with FBSS.



Aim of the study

- The aim of this study is to investigate different learning models and to select the most suited algorithm for predicting the attributes that influence this syndrome and to find new rules .

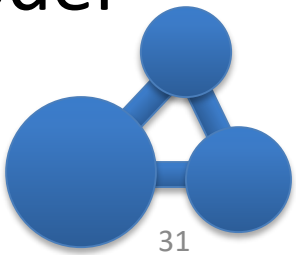
- The data are unbalanced and we will try not the usual approaches to achieve this aim.



Constructing the predictive model

To construct a predictive model we followed four step procedure:

- familiarization with the data
- data processing
- modelling
- evaluation of performance of the model



Clinical data(1/5)

- Data gathered from patients operated in two centers:
  - 258 pts at the ”Galeazzi” Institute, Neurosurgery Operative Unit, in Milan
  - 47 patients at the Neurosurgery Service of the Mother Theresa Hospital of Tirana.

- Inclusion criteria:
  1. Patients operated for lumbar disc herniation.
  2. Pain not due to recurrence, or persistence of certain hernia
- Exclusion criteria: they are excluded from the study (only the carbamazepine group / clorimipramina) patients with prostatic hypertrophy, narrow angle glaucoma, cardiac arrhythmias due toincompatibility with assunzioine tricyclics).

Clinical data(2/5)

Number	Attribute Name	Description	Values
1.	Level	Level of the intervertebral disc which has herniated.	L5-S1,L4-L5,L3-L4, L2-L3,L1-L2
2.	Side	Side of the radicle involved (right DX; left SX).	DX,SX
3.	Age	Age of the patient in years	Numeric Min. 17, Max. 86 Mean 46.8 Standard Deviation 12.8
4.	Sex	Sex of the patient (m – male; f - female)	m,f

Clinical data(3/5)

5.	Progressive Therapy	Steroid and non-steroid anti-inflamatory drugs administered before surgery	NSAID, Steroids, NSAID and steroids, NSAID and Physical therapy, None.
6.	Smoke	Patient's habit	no, yes, ex
7.	Time first pain to surgery	Time from onset of symptoms to surgery	Numeric Min 1, Max. 48 Mean 7.5 Standard Deviation 9.4
8.	Previous episodes of pain	Any similar episodes of pain in the same distribution before the final that made the patient undergo operation	no, yes

Clinical data(4/5)

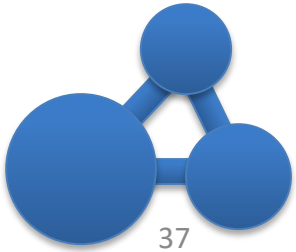
9.	Preoperative VAS	Intensity of pain before surgery	0,1,2,3,4,5,6,7,8,9,10
10.	Postoperative VAS at 1 month	Intensity of pain 1 month after surgery	0,1,2,3,4,5,6,7,8,9,10
11.	Postoperative VAS at 1 year	Intensity of pain 1 year after surgery	0,1,2,3,4,5,6,7,8,9,10
12.	Lasegue grading	Angle of possible flexion of the extended lower limb	no, 10.0, 15.0, 20.0, 30.0, 35.0, 40.0, 45.0, 50.0,55.0, 60.0, 70.0 , 75.0 ,80.0
13.	Hypo or absent OTR	Attenuated or absent osteo-tendinous reflex	no, yes
14.	Motor deficit	Lowered strength of the interested muscles	no, yes

Clinical data(5/5)

15.	Controls Sphincters	Capacity to control urinary and anal sphincters	no, yes
16.	Sensitive deficit	Alteration of the sensibility	no, yes
17.	Paravertebral contracture	Involuntary contracture of the scaffolding muscles of the vertebral column	no, yes
18.	Type of radiological examination	Radiological examinations done before surgery	MR,CT,CT and MR,CT dynamic X-ray.

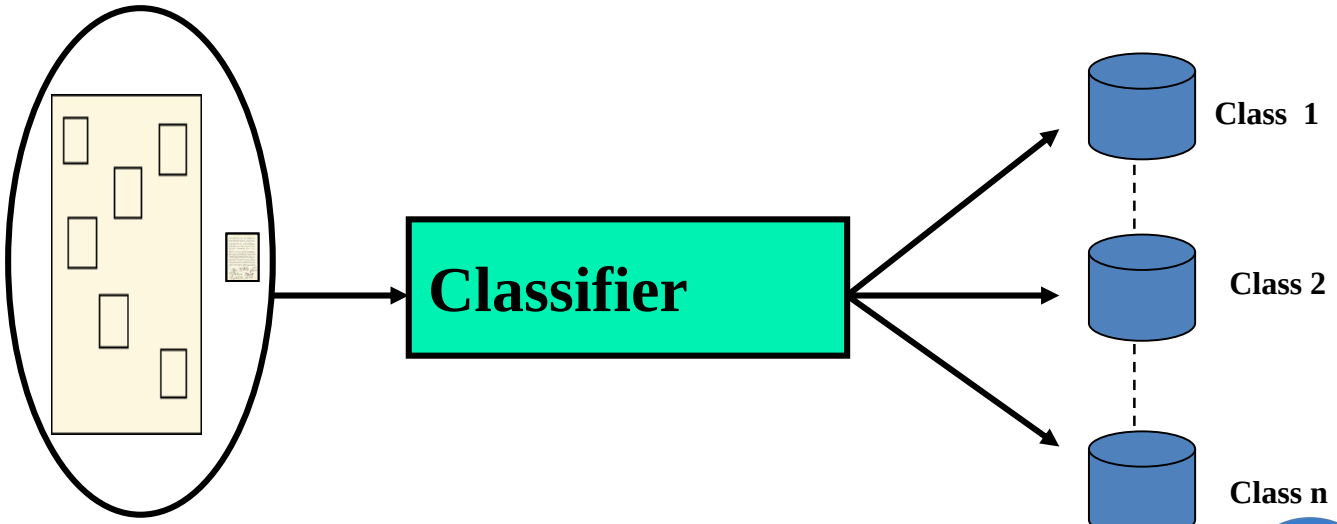
The classification

- The classification techniques of Data mining can be used in the clinical decision support.
  - Diagnostic and prevention
  - prognosis
  - Decision support



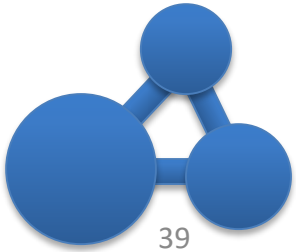
The classification

- Requires prior knowledge of the data (supervised learning).
- The set of records (training set) contains a special attribute which is the label of the class.



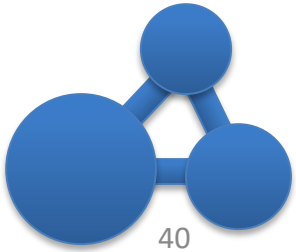
Used Classifiers

- **Classification algorithms of WEKA**
  - **Algorithms based on the trees**
    - ADTree, BFTree, DecisionStump, FT, J48, J48graft, LADTree, LMT, NBTree, RandomForest, RandomTree, REPTree, SimpleCart.
  - **NaiveBayes**
  - **SMO**
  - **ZeroR**
  - **MultilayerPerceptron.**



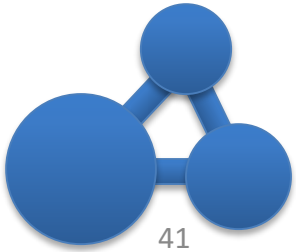
Esperiments (1/7)

- The validation method
  - 10 fold cross validation
  - hold out with percentage 50%, 66% e 85%.



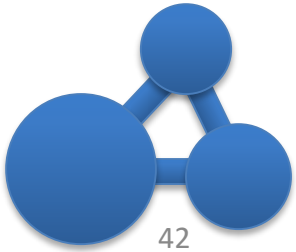
Evaluation methods and tools

- We will create new tools extending Weka which
  - repeate every algorithm 10 times
  - randomize the data in every repetition
  - register the media of every criteria for evaluation for each evaluation method.

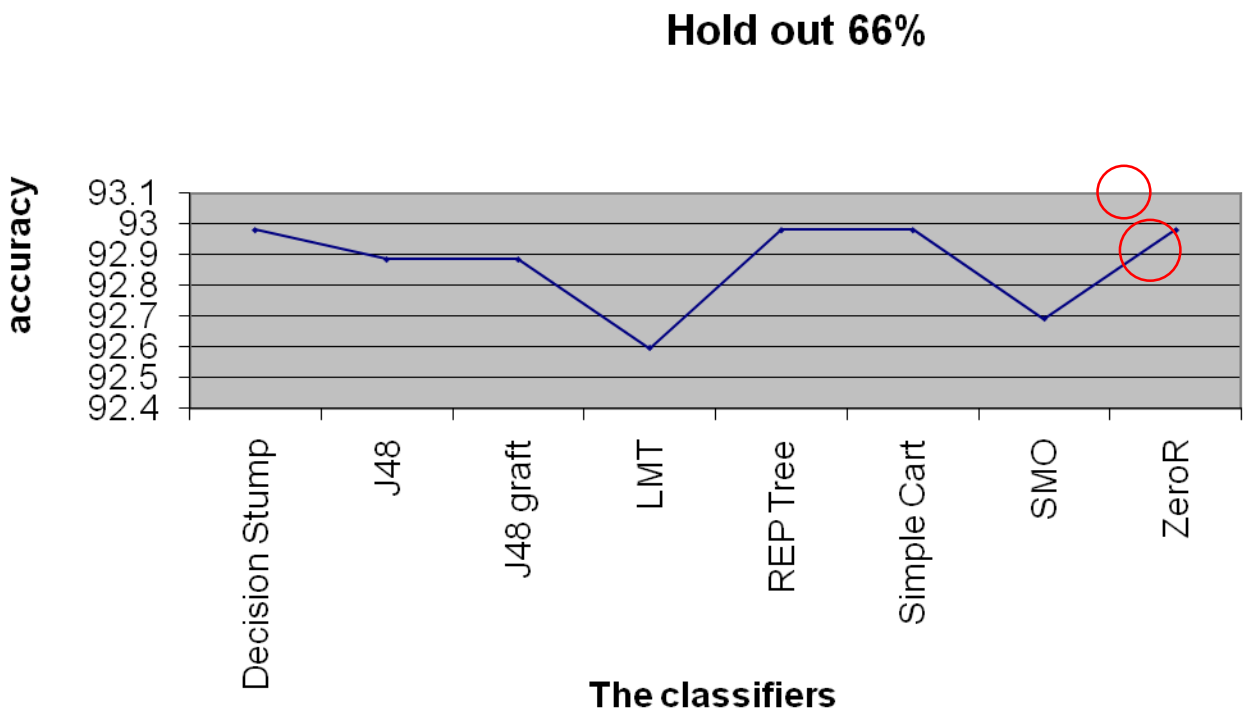


Esperiments (2/7)

- For every classifiers the media of the following criteria has been evaluated :
  - Accuracy
  - Kappa statistics
  - Mean absolute error
  - Root mean-square
  - TP and FP rates
  - Precision
  - F-measure.



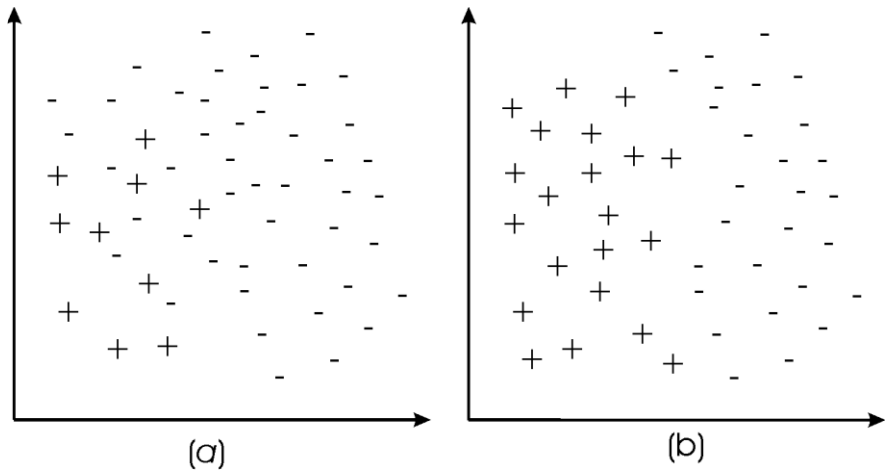
Esperiments (3/7)



- Zero R is the best classifier (the greatest accuracy)
- Accuracy is not significant because of the unbalanced data

The unbalanced data

- The problem of unbalanced data is presented when the data set is dominated by a class or classes that have significantly more data than the other.



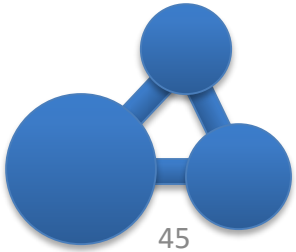
The approaches to solve the problem of unbalanced data

The approaches:

- resampling methods
- new algorithms
- feature selection methods

The evaluation criteria:

- F-Measure
- AUC



The first experiment

- Algorithms with greatest F-Measure:
  J48, J48Graft, RandomForest, REPTree e SMO.

- The selected algorithms have been applied for 10 times randomizing the data using the method 10 fold cross validation and the media of the AUC values has been evaluated for every attribute.

Esperiments (4/7)

Attributes	Maximal Value	Classifiers
Level	0,573	RandomForest
Side	0,500	SMO
Age	0,500	SMO
Sex	0,518	RandomForest
Progressive Therapy	0,526	RandomForest
Smoke	0,500	SMO
Time first pain to surgery	0,544	RandomForest
Previous episodes of pain	0,500	SMO
Preoperative VAS	0,682	RandomForest
Postoperative VAS at 1 month	0,632	RandomForest
Postoperative VAS at 1 year	0,536	RandomForest
Lasegue grading	0,500	SMO
Hypo or absent OTR	0,500	SMO
Motor deficit	0,500	SMO
Controls Sphincters	0,500	SMO
Sensitive deficit	0,566	RandomForest
Paravertebral contracture	0,569	RandomForest

Esperiments (5/7)

- The next experiment evaluate AUC of the groups composed of different attributes.
- In the table is presented the combination of groups composed of two, three, four and five attributes.

Attributes	**AUC**
All	0.646
Preoperative VAS	0.682
Preoperative VAS , Type of radiological examination	0.736
Preoperative VAS , Paravertebral contracture, Type of radiological examination	0.739
Level, Progressive Therapy, Time first pain to surgery, Hypo or absent OTR	0.756
Level, Progressive Therapy, Time first pain to surgery, Hypo or absent OTR, Paravertebral contracture	0.744

Esperiments (6/7)

- Maximizing F-Measure

Attributes	**FMeasure**
All	0.894
Time first pain to surgery	0.895
Time first pain to surgery, Paravertebral contracture	0.905
Postoperative VAS at 1 year, Motor deficit, Paravertebral contracture	0.909
Sex, Time first pain to surgery, Postoperative VAS at 1 year, Paravertebral contracture	0.919
Level, Sex, Previous episodes of pain, Sensitive deficit, Paravertebral contracture	0.921

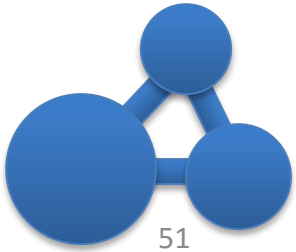
Esperiments (7/7)

Attributes	False Positive	False Negative
All	21.4	2.0
Time first pain to surgery	21.2	1.8
Time first pain to surgery, Paravertebral contracture	20.2	0.8
Postoperative VAS at 1 year, Motor deficit, Paravertebral contracture	19.1	2.5
Sex, Time first pain to surgery, Postoperative VAS at 1 year, Paravertebral contracture	20.4	5.1
Level, Sex, Previous episodes of pain, Sensitive deficit, Paravertebral contracture	18.0	0.0

Attributes	False Positive	False Negative
All	22.0	0.0
Preoperative VAS	21.9	0.9
Preoperative VAS , Type of radiological examination	21.9	0.9
Preoperative VAS , Paravertebral contracture, Type of radiological examination	18.9	9.4
Level, Progressive Therapy, Time first pain to surgery, Hypo or absent OTR	18.9	8.7

Conclusions

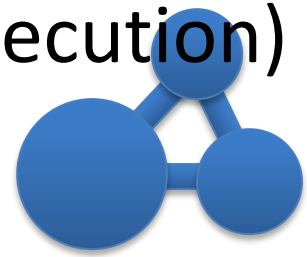
- When the data are unbalanced accuracy is not statistically significant.
- Two other evaluation criteria to find the best classifier are used: F-Measure and AUC.
- The most accurate algorithm is Random Forest.
- The maximal value of AUC is 0.756, which means that the classifier predicts well.



Future Work

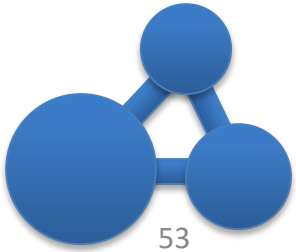
We are planning to expand our work in these directions:

- developing models based on more predictors
- application of other machine learning techniques and approaches
- training and testing on larger and more diversified data sets.
- Improving the performance of the application (moving into a cloud platform, paralel excecution)



Creting a user interface tool

- The problem with these programs is that they don't have a user interface and the user that will use them should know java programming in order to change the environment setting.
- We will changed them in order to have a more user friendly interface and to be more usable even by users who don't have programming background.



The subsystems of Weka

The simple command line interface allows a more experienced user to input commands directly into a command line.
The Explorer interface is a graphical interface that allows the user to perform the main data mining tasks and to visualise the results in a number of forms.
The Experimenter interface allows the user to compare and analyse the performance of different data mining algorithms.
The Knowledge Flow interface is offered as an alternative to the Explorer interface but does not contain all the functionality offered by the Explorer. It is based around a data flow style of layout.

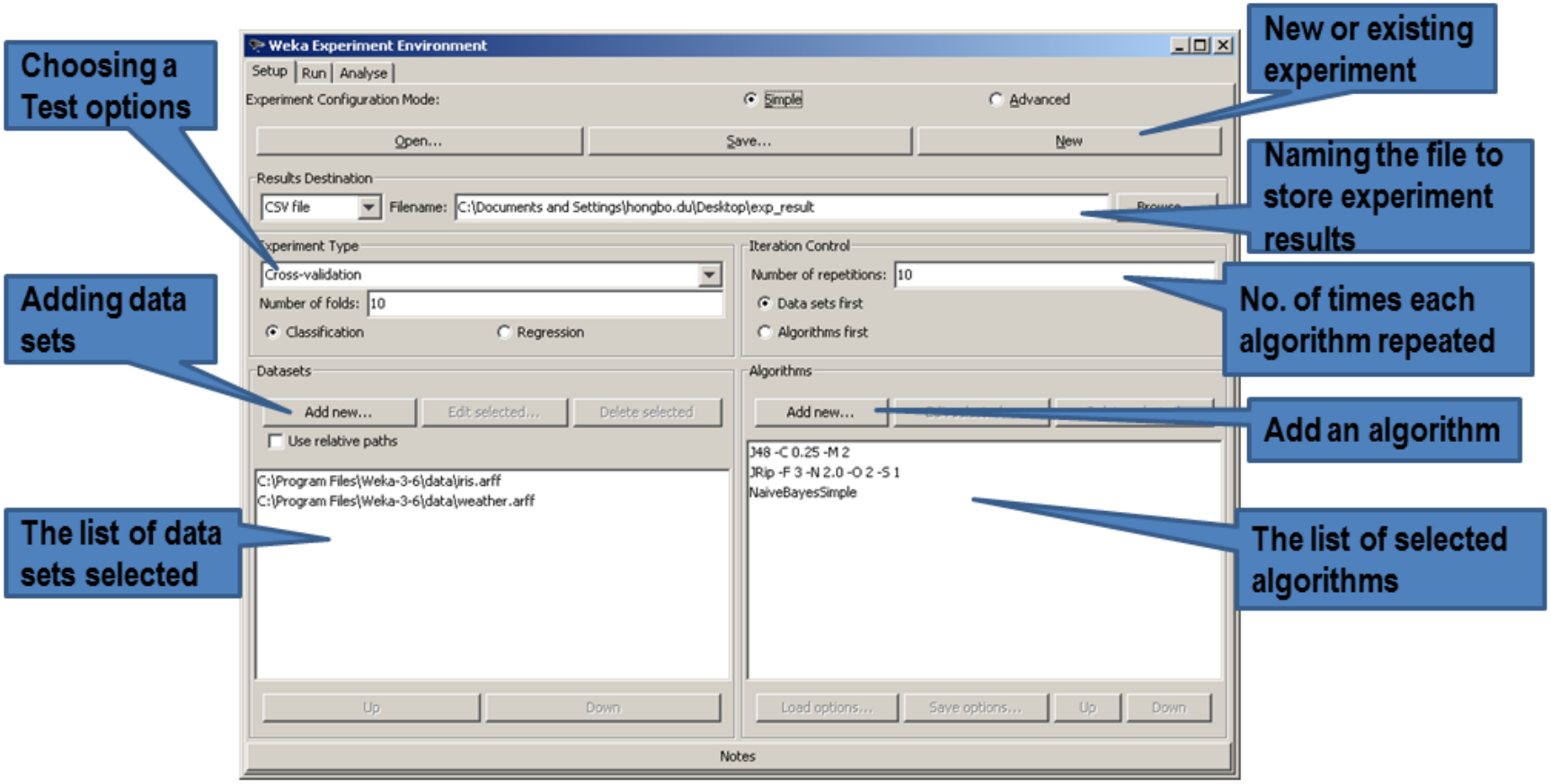
User interface of Weka

Weka is a very useful program for a wide range of users with a lot of functionalities.However there are some disavantages regarding its user interface which is not intuitive.

Some of its diasavantages are listed below:

- Both the appearance and interaction methods of each of the four interfaces are significantly different, and this adds to the complexity of the overall design.
- The system may seem far more intuitive if there was more consistency through the different functions offered by each of the sections.
- The interface should be more user-friendly while still offering all the functionality that is available through the program.
- The design process will need to include the application of user interface design standards to ensure that the interface is as effective as possible.
- It will be important to ensure that the interface makes the software far more accessible for new users, whist not making it frustrating for more experienced users to continue using.

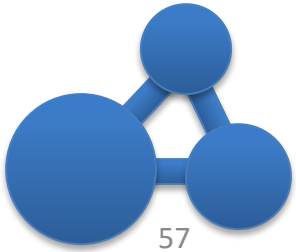
Experimenter of Weka



The user interface and the process of setting the environment is not intuitive!

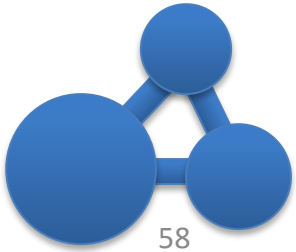
The results are displayed only in textual view.

- The aim is to build an application with user interface using the Java language and Weka API and designed built with MVC architecture, simple to use and extendable in the future.
- Results of the analysis of the data are displayed in two different views.
- The application allow to add other views or change the controller.



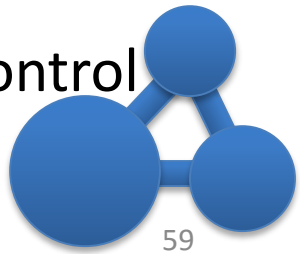
Designing the tool extending Weka

- The tool that we have created by using Java and Weka API is designed with the architectural design pattern Model-View-Controller.
  - highly modular architecture that allows for louse coupling of the proposed subsystems.
  - each individual subsystem is designed to be as much independent as possible, so we can easily extend or change the behavior of the actual system in the future.



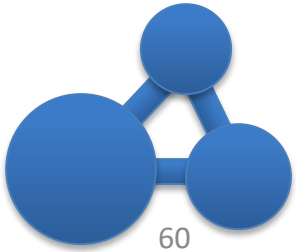
Model View Controller in designing the applications with user interface

- Model-View-Controller ("MVC") is the recommended architectural design pattern for interactive applications
- MVC organizes an interactive application into three separate modules:
  - one for the application model with its data representation and business logic,
  - the second for views that provide data presentation and user input, and
  - the third for a controller to dispatch requests and control flow.



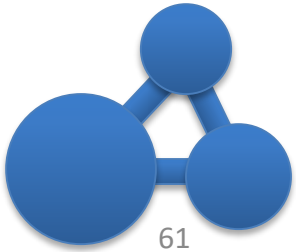
Model View Controller

- The intent of MVC is to keep neatly separate objects into one of tree categories
  - Model
    - The data, the business logic, rules, strategies, and so on
  - View
    - Displays the model and usually has components that allows user to edit change the model
  - Controller
    - Allows data to flow between the view and the model
    - The controller mediates between the view and model

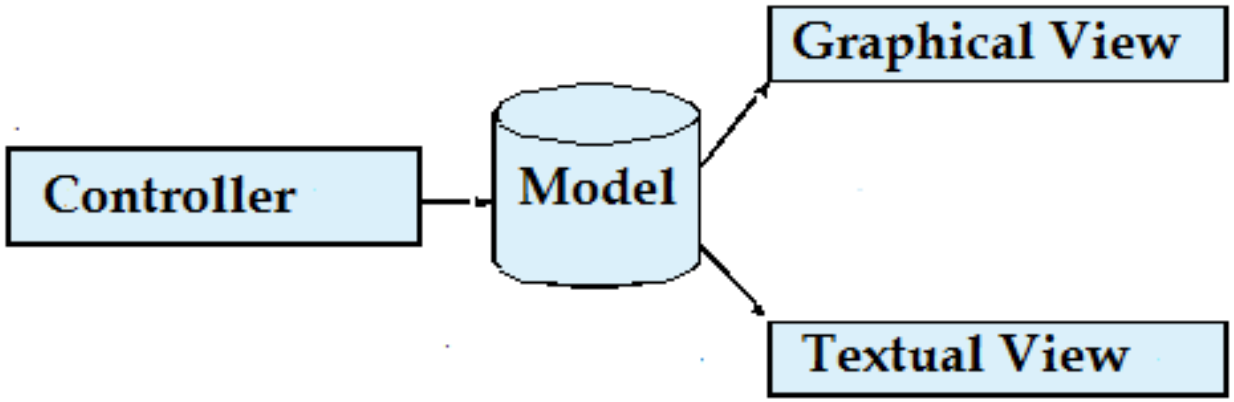


Advantages of MVC

- Clarity of design
  - easier to implement and maintain
- Modularity
  - changes to one don't affect the others
  - can develop in parallel once you have the interfaces
- Multiple views
  - games, spreadsheets, powerpoint, Eclipse, UML reverse engineering, ….

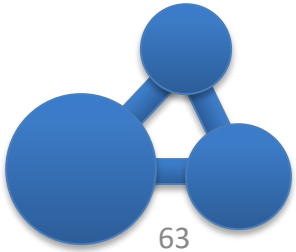


WekaTool Application MVC Architecture

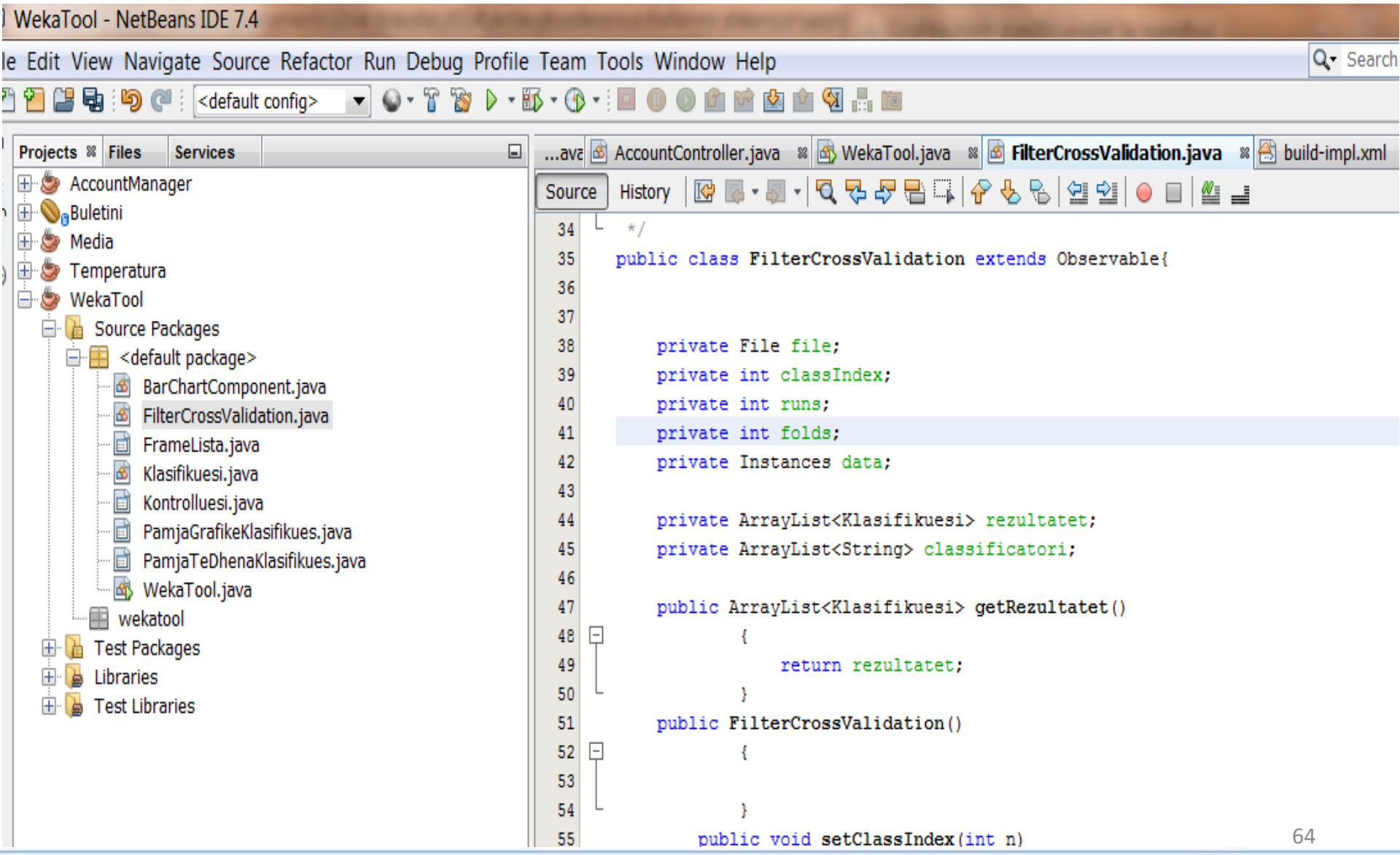


FilterCrossValidation Class

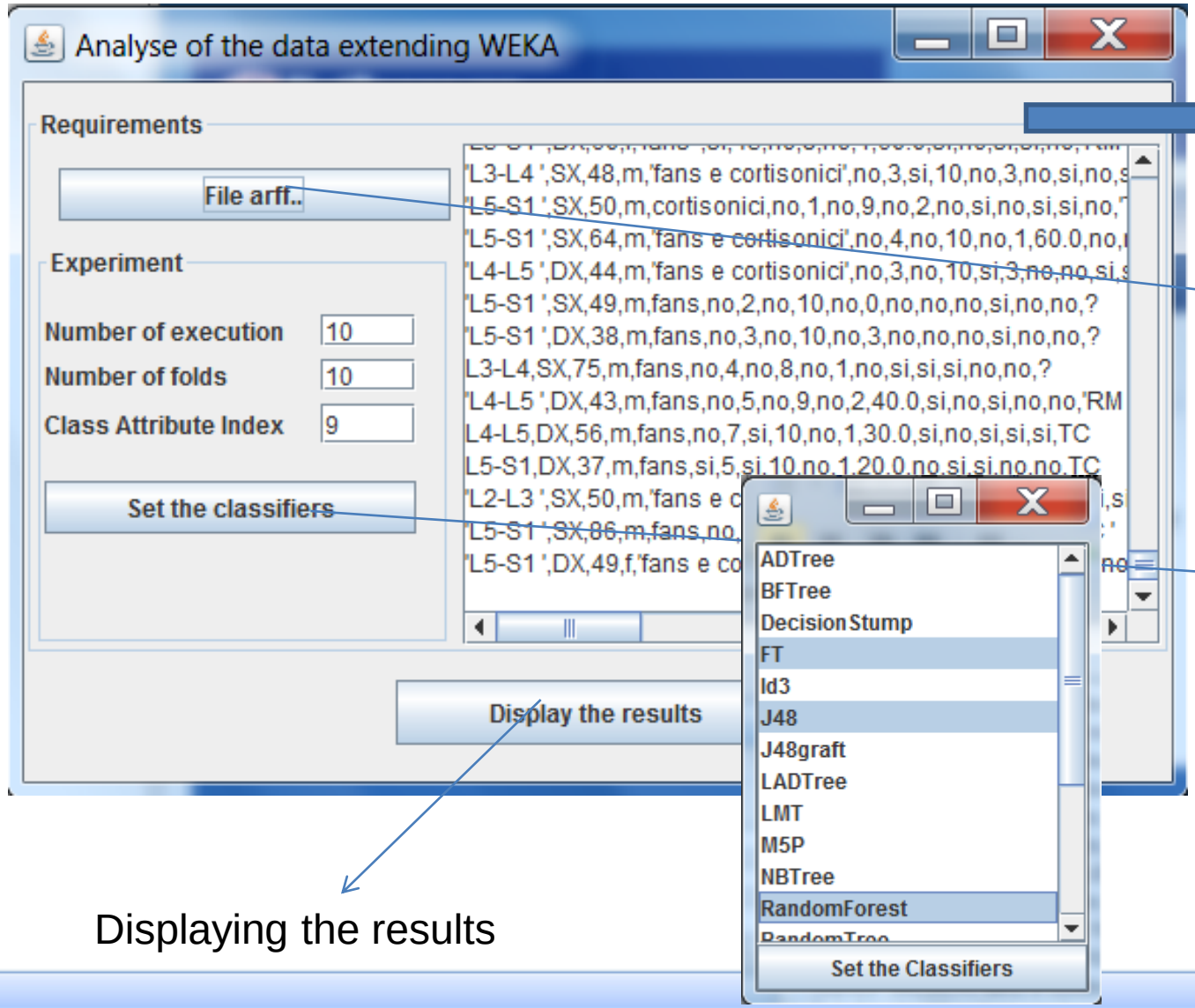
- Represents the model in the WekaTool application
- Class **FilterCrossValidation** extends class **Observable** and acts as a model in the application
  - Method ***setChanged***() of class **Observable** set model's ***changed*** flag
  - Method ***notifyObservers***() of class **Observable** notify all **Observers** (Views) of the change.
  - An **Observable** object must invoke method ***setChanged***() before invoking method ***notifyObservers***()
  - Method ***notifyObservers***() invokes method ***update***() of interface **Observer** for each registered **Observer**



Model


WekaTool - NetBeans IDE 7.4
le Edit View Navigate Source Refactor Run Debug Profile Team Tools Window Help
Search
<default config>
Projects
Files
Services
AccountManager
Buletini
Media
Temperatura
WekaTool
Source Packages
<default package>
BarChartComponent.java
FilterCrossValidation.java
FrameLista.java
Klasifikuesi.java
Kontrolluesi.java
PamjaGrafikeKlasifikues.java
PamjaTeDhenaKlasifikues.java
WekaTool.java
wekatool
Test Packages
Libraries
Test Libraries
...ava
AccountController.java
WekaTool.java
FilterCrossValidation.java
build-impl.xml
Source
History
34 */
35 public class FilterCrossValidation extends Observable{
36
37
38 private File file;
39 private int classIndex;
40 private int runs;
41 private int folds;
42 private Instances data;
43
44 private ArrayList<Klasifikuesi> rezultatet;
45 private ArrayList<String> classificatori;
46
47 public ArrayList<Klasifikuesi> getRezultatet()
48 {
49 return rezultatet;
50 }
51 public FilterCrossValidation()
52 {
53
54 }
55 public void setClassIndex(int n)

The tool with MVC archiecture



Controller

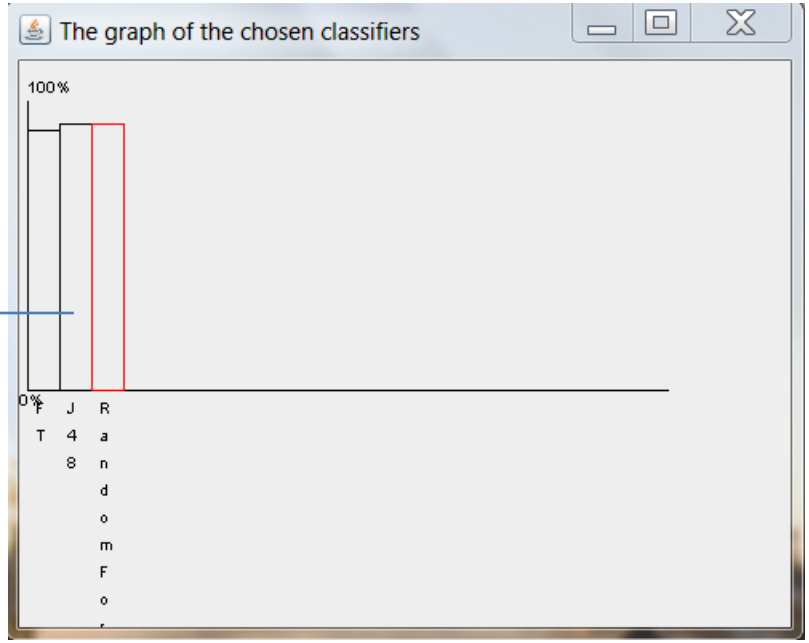
Define the arff file and display it

After determining the values of the experiment the user should choose the classifiers

Displaying the results

The results displayed wih two views

Graphical view where the classifier with the greatest accuracy is represented with a red bar



Textual view

The analysis of the chosen classifiers

weka.classifiers.trees.FT

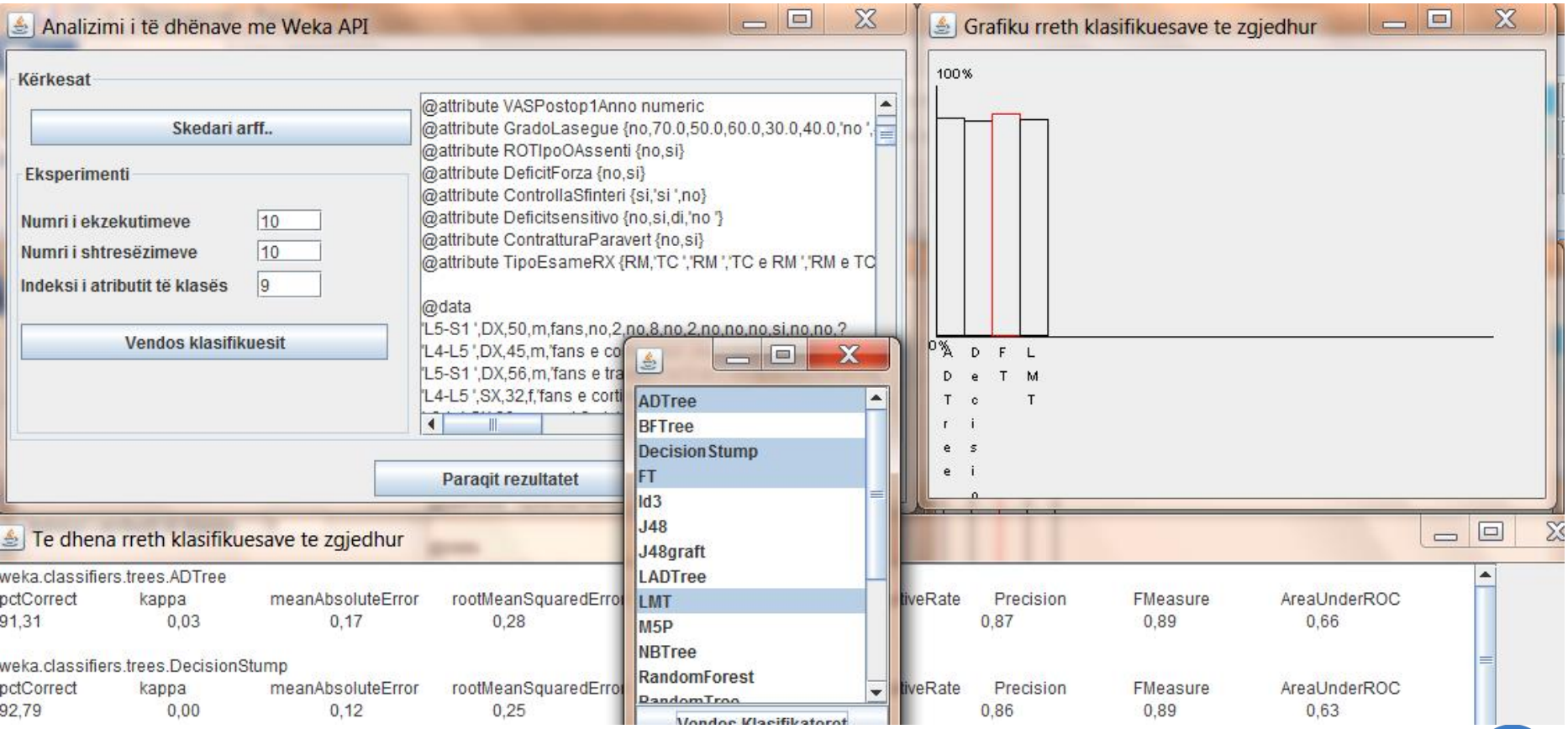
pctCorrect	kappa	meanAbsoluteError	rootMeanSquaredError	TruePositiveRate	FalseNegativeRate	Precision	FMeasure	AreaUnderROC
90,07	0,13	0,11	0,30	0,90	0,10	0,88	0,89	0,63

weka.classifiers.trees.J48

pctCorrect	kappa	meanAbsoluteError	rootMeanSquaredError	TruePositiveRate	FalseNegativeRate	Precision	FMeasure	AreaUnderROC
92,26	0,04	0,13	0,26	0,92	0,08	0,88	0,89	0,61

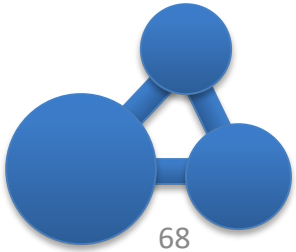
weka.classifiers.trees.RandomForest

pctCorrect	kappa	meanAbsoluteError	rootMeanSquaredError	TruePositiveRate	FalseNegativeRate	Precision	FMeasure	AreaUnderROC
92,33	0,02	0,13	0,26	0,92	0,08	0,87	0,89	0,66


Analizimi i të dhënave me Weka API
Kërkesat
Skedari arff..
Eksperimenti
Numri i ekzekutimeve
10
Numri i shtresëzimeve
10
Indeksi i atributit të klasës
9
Vendos klasifikuesit
Paraqit rezultatet
@attribute VASPostop1Anno numeric
@attribute GradoLasegue {no,70.0,50.0,60.0,30.0,40.0,'no '
@attribute ROTIpoOAssenti {no,si}
@attribute DeficitForza {no,si}
@attribute ControllaSfinteri {si,'si ',no}
@attribute Deficitsensitivo {no,si,di,'no '}
@attribute ContratturaParavert {no,si}
@attribute TipoEsameRX {RM,'TC ','RM ','TC e RM ','RM e TC
@data
'L5-S1 ',DX,50,m,fans,no,2,no,8,no,2,no,no,no,si,no,no,?
'L4-L5 ',DX,45,m,'fans e co
'L5-S1 ',DX,56,m,'fans e tra
'L4-L5 ',SX,32,f,'fans e corti
ADTree
BFTree
DecisionStump
FT
Id3
J48
J48graft
LADTree
LMT
M5P
NBTree
RandomForest
Grafiku rreth klasifikuesave te zgjedhur
100%
0%
A D T r e e
D e c i s i o
F T
L M T
Te dhena rreth klasifikuesave te zgjedhur
weka.classifiers.trees.ADTree
pctCorrect kappa meanAbsoluteError rootMeanSquaredError tiveRate Precision FMeasure AreaUnderROC
91,31 0,03 0,17 0,28 0,87 0,89 0,66
weka.classifiers.trees.DecisionStump
pctCorrect kappa meanAbsoluteError rootMeanSquaredError tiveRate Precision FMeasure AreaUnderROC
92,79 0,00 0,12 0,25 0,86 0,89 0,63

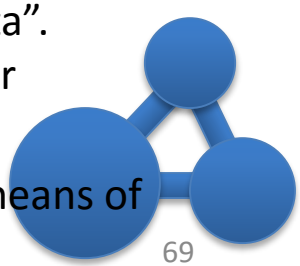
Conclusion

- Weka is a very useful to analyse the medical data with the techniques of data mining but we can extend it to improve its user interface and the way that the user interect with the system or develop new tools.
- In our study we have improved the user interface with the architecture MVC.



References

- **A. Kika**, R. Alimehmeti, February 2014, "Prognosis of failed back surgery syndrome based on feature extraction method", International Journal of Innovative Research in Computer and Communication Engineering(IF= 1.386). ISSN(Print):2320–9798; ISSN(Online):2320-9801,Vol. 2, Issue 2, Pages 2916-2921.
- **A. Kika**, R. Alimehmeti, E. Zekaj, N. Cesa-Bianchi, A. De Santis, M. Petrela (November 10–12, 2011), "Techniques of data mining to predict the factors that influence in the failed back surgery syndrome". 7th International Meeting of Euro-Mediterranean Medical Informatics and Telemedicine, Academy of Sciences, Tirane. Proceedings: Nuove Tecnologie in Medicina, Applicazioni Informatiche e Telematiche in Medicina, Anno 11, Nr 1-2,. Pg 3-8.
- **A. Kika** : Using Bayesian Networks for discovering knowledge from medical data. Proceedings:Scientific Conference : Role of Information and Communication Technology in the development of Albanian Society. pp: 287-294. Tiranë, 2010. ISBN: 978-99956-40-43-9.
- **A. Kika**, J. Lula (8-9 Qershor 2012), "Discovering knowledge in unbalanced data". International Conference "Information System an Technology Innovation: their application in Economy", Tiranë. Proceedings ISBN:978-99995-6377-8-1.
- **A. Kika**, E. Xhina, R. Alimehmeti, "Developing e-health data analysis tool by means of WEKA", eGeH'14 e-government & e-health, 10 – 12 July 2014, Desio,Italy.



Thank you!

Questions?