Specialised Tools for Automating Data Mining for Hospital Management

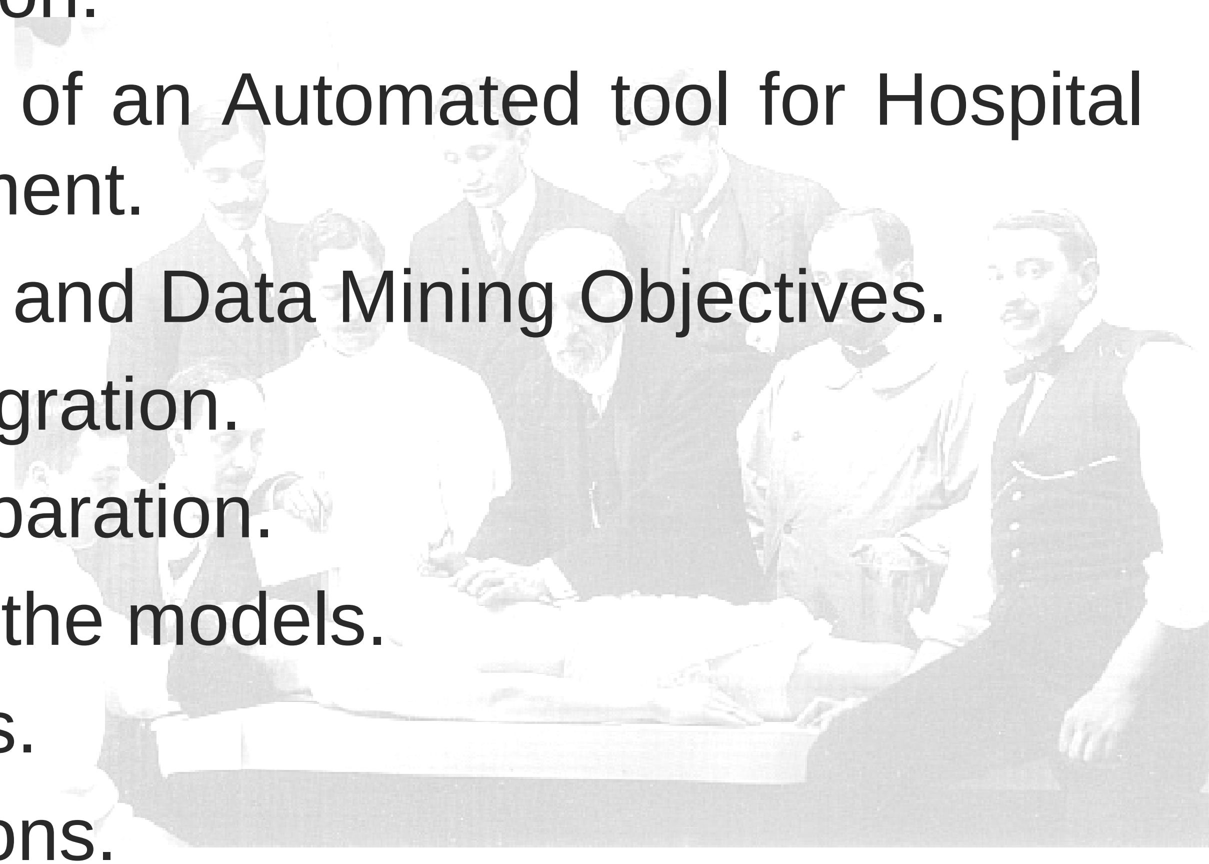
HCMC-2005 31st August – 2nd September
Craiova-Romania

José Daniel Llopis Llopis
*A **M**ulti-paradigm **I**nductive **P**rogramming Group (MIP)*
Universidad Politécnica de Valencia (UPV), Valencia (Spain)

DEPARTAMENTO DE SISTEMAS INFORMÁTICOS Y COMPUTACIÓN

UNIVERSIDAD POLITECNICA DE VALENCIA

Index



- Introduction.
- Structure of an Automated tool for Hospital Management.
- Business and Data Mining Objectives.
- Data Integration.
- Data Preparation.
- Learning the models.
- Examples.
- Conclusions.

Index

- Introduction.
- Structure of an Automated tool for Hospital Management.
- Business and Data Mining Objectives.
- Data Integration.
- Data Preparation.
- Learning the models.
- Examples.
- Conclusions.

Participants in the Project

Research Group

MIP-DSIC-UPV

"A Multi-paradigm Inductive Programming Group"

Dimensión Informática Company

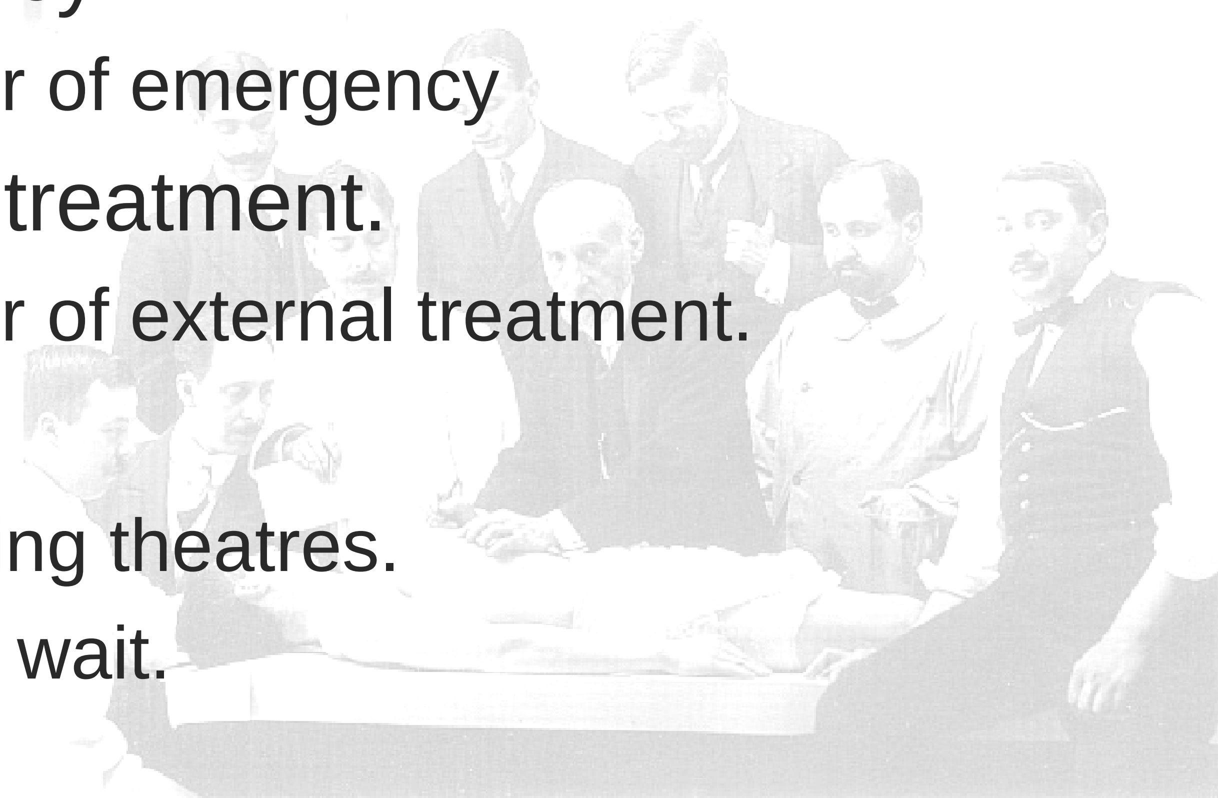
Clinic Hospital of Valencia

Motivations

- There are big amount of stored data.
- To improve the planning of resources:
  - Bed occupation.
  - Operating theatres.
  - Human and material resources.
  - The influence of certain diseases in the hospital's services.
- In order to minimise costs and improve the care given to the patients.
- Minimise the time of list wait of the patients.

Initial Hospital Areas

- Emergency.
  - Number of emergency
- External treatment.
  - Number of external treatment.
- Surgery
  - Operating theatres.
  - Lists of wait.

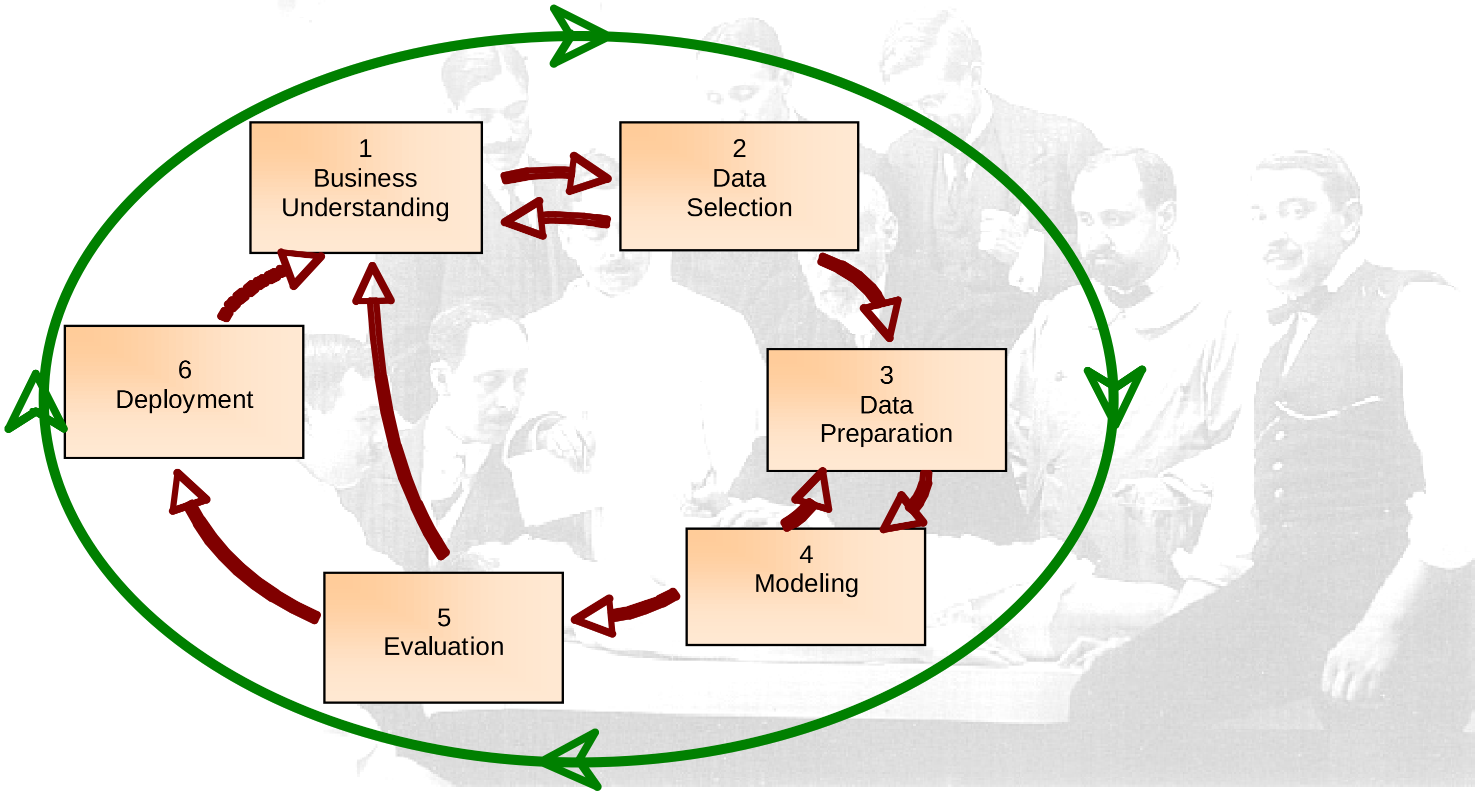


Index

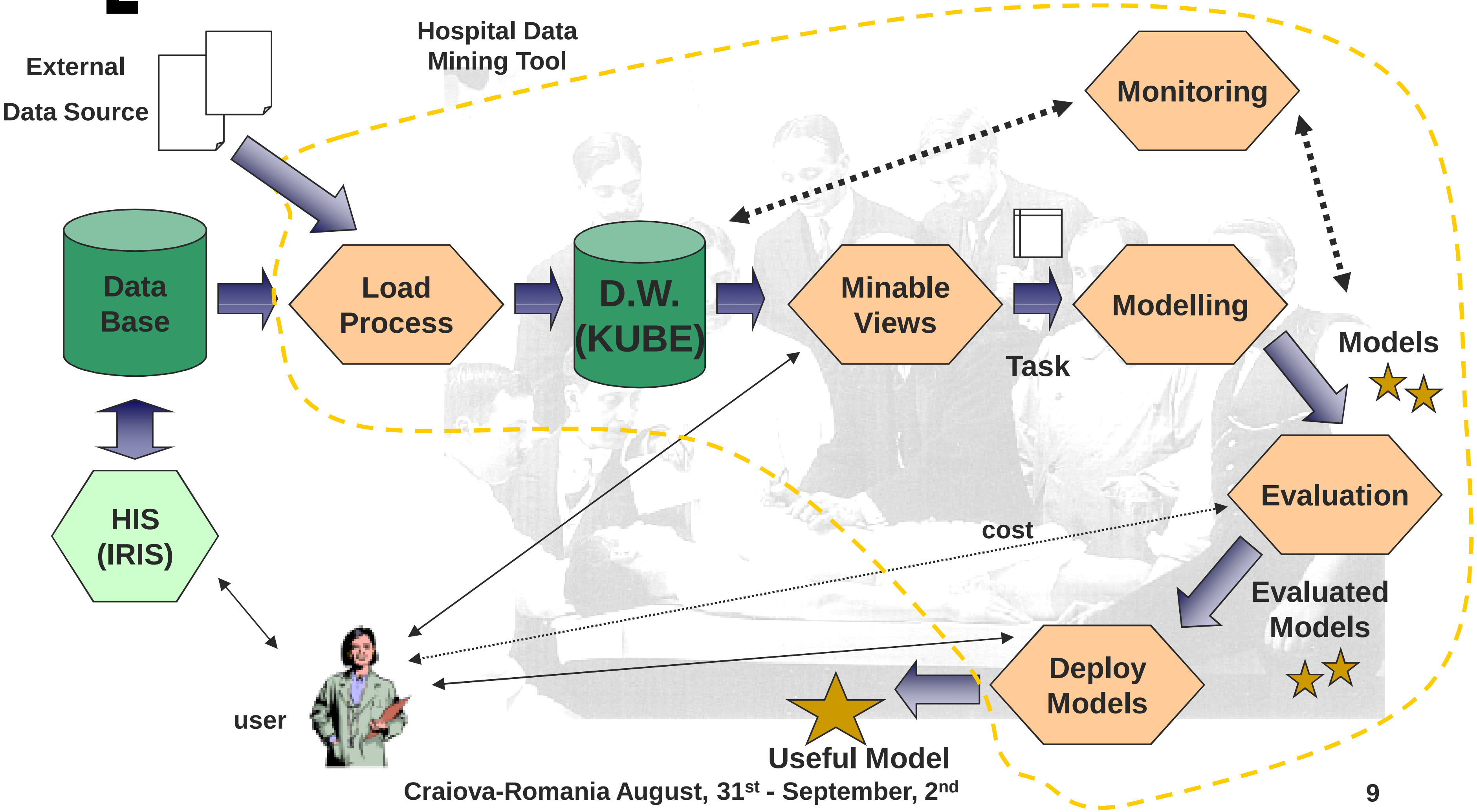
- Introduction.
- Structure of an Automated tool for Hospital Management.
- Business and Data Mining Objectives.
- Data Integration.
- Data Preparation.
- Learning the models.
- Examples.
- Conclusions.

CRISP-DM

CRISP-DM (CRoss-Industry Standard Process for Data Mining)



Automated Data Mining Tool



Index

- Introduction.
- Structure of an Automated tool for Hospital Management.
- **Business and Data Mining Objectives.**
- Data Integration.
- Data Preparation.
- Learning the models.
- Examples.
- Conclusions.

Business Objectives

- To optimise bed occupation.
- To improve the use of operating theatres, avoiding the cancellation of operations.
- To know how emergencies affect to the administration of the hospital departments or services (cancellation of operations, etc).
- To optimise the allocation of human and material resources to wards and shifts.
- To detect the influence of certain diseases in the hospital's services.
- To find clusters of patients.

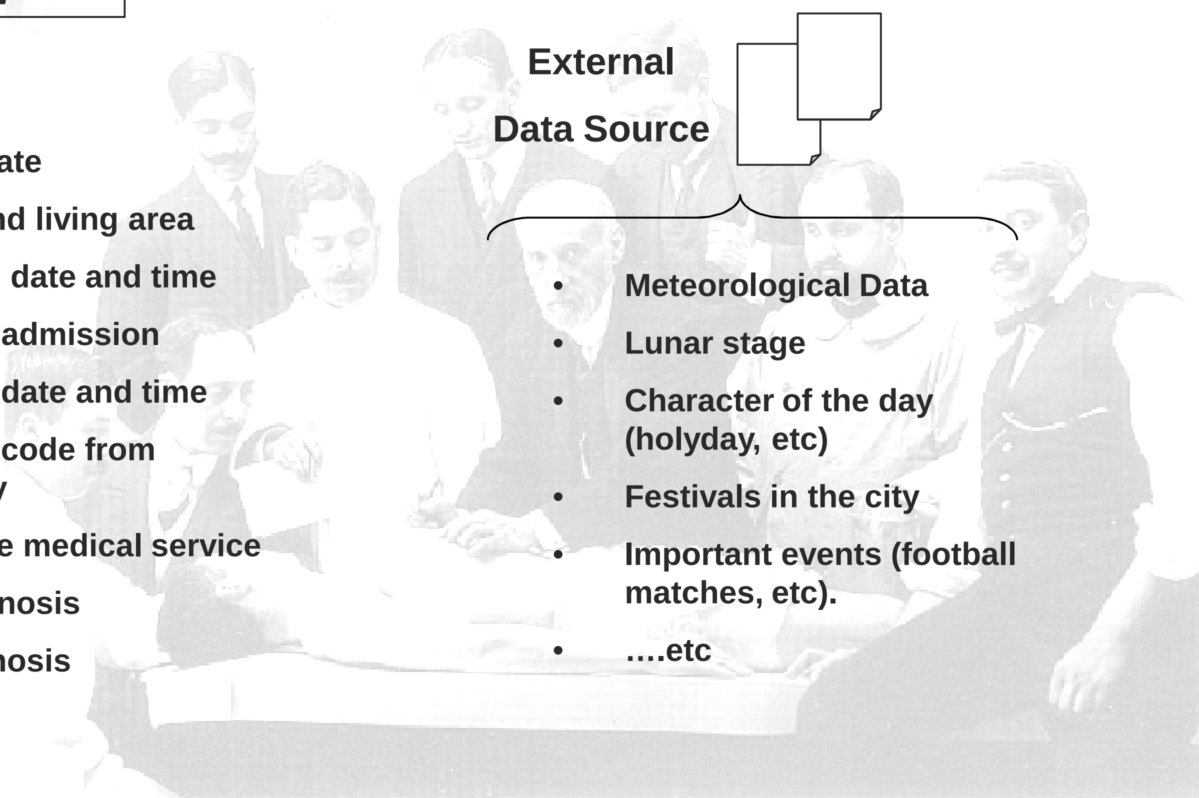
Data Mining Objectives

- To carry out global models about pressure emergencies by different time periods (daily, by shifts of work, by day of the week, etc).
- To generate a model for predicting the number of daily hospitalisations coming from emergencies.
- To obtain predictive models of global and partial use of beds by hospital service.
- To construct models for estimating how the resources of a hospital are affected by a certain disease (for instance, influenza).
- To carry out models to cluster patients (by age, by area, by pathology class, etc).

Index

- Introduction.
- Structure of an Automated tool for Hospital Management.
- Business and Data Mining Objectives.
- Data Integration.
- Data Preparation.
- Learning the models.
- Examples.
- Conclusions.

Internal and External Data



Internal Data of Hospital

Data Base

HIS (IRIS)

- Sex
- Birthday date
- Country and living area
- Admission date and time
- Reason of admission
- Discharge date and time
- Discharge code from emergency
- Code of the medical service
- Initial diagnosis
- Final diagnosis
- ….etc

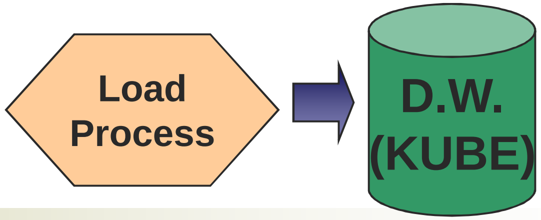
External Data of Hospital

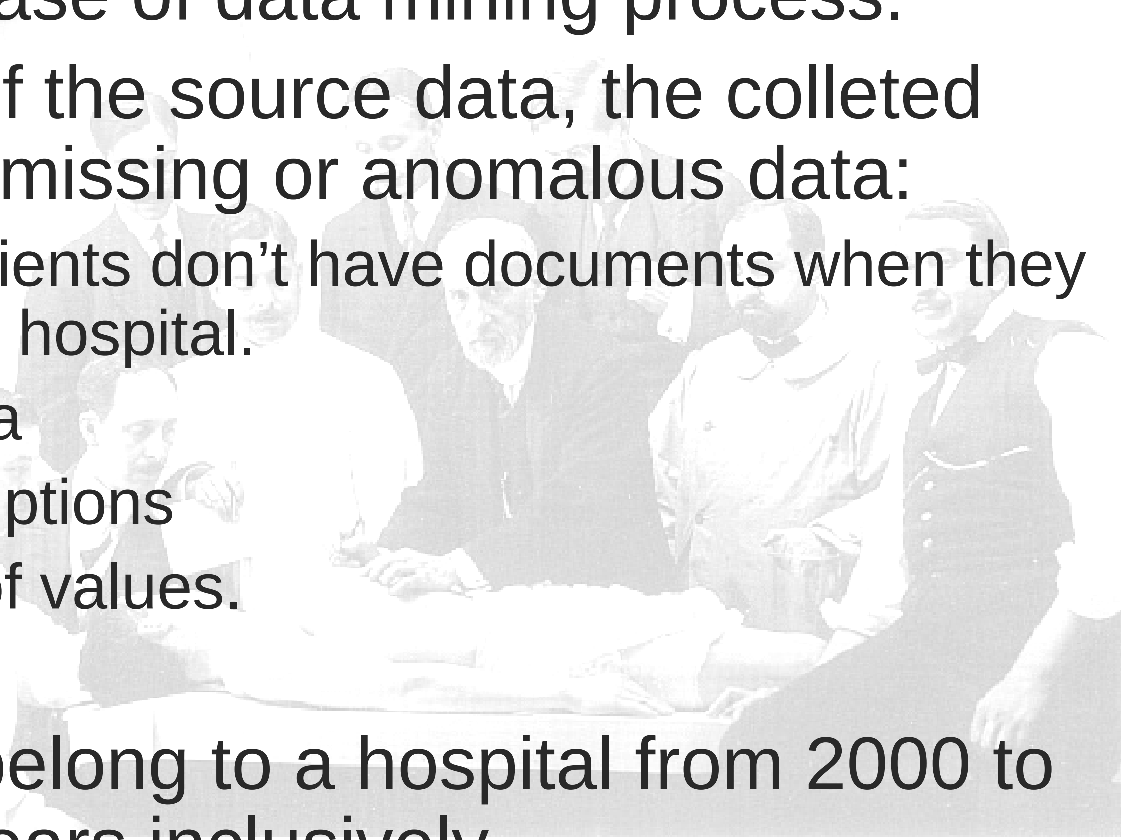
External Data Source

- Meteorological Data
- Lunar stage
- Character of the day (holyday, etc)
- Festivals in the city
- Important events (football matches, etc).
- ….etc

Index

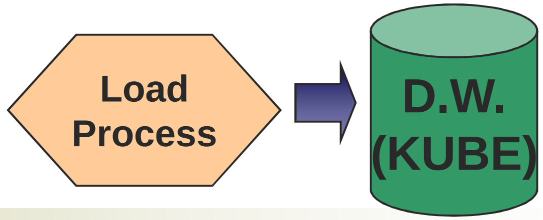
Data Preparation





- Important phase of data mining process.
- Bad quality of the source data, the colleted data contain missing or anomalous data:
  - Many of patients don't have documents when they arrive at the hospital.
  - Illegible data
  - Bad transcriptions
  - Repetition of values.
  - ..etc.
- These data belong to a hospital from 2000 to 2004, both years inclusively.

Data Preparation



- This preparation stage is reused from hospital to hospital, through the automation of all these processes in a data preparation module.
- We implemented SQL scripts for extracting data the different hospitals into the DW.
- These queries are 100% portable from one hospital DW to another, and all this effort is reused.
- We found attributes containing text, for example, an initial description of the pathology of the patient, since this kind of attributes cannot be directly dealt with classical learning methods, we employ to transform the text attributes in one or more discrete attribute.
- Finally, the data is converted into a standard format (the arff format of WEKA) by means scripts of python

Index

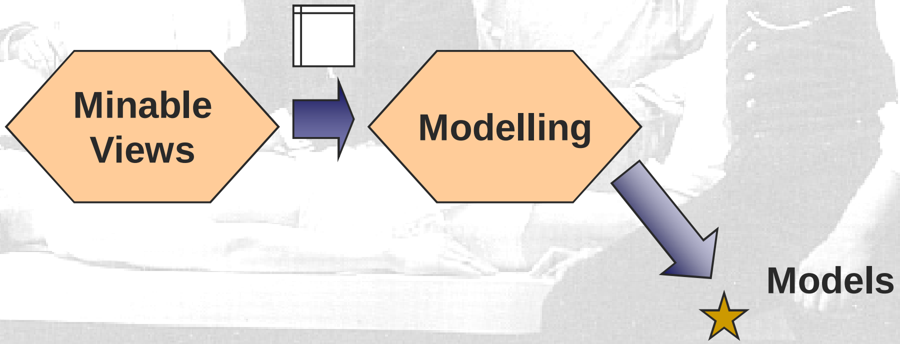
- Introduction.
- Structure of an Automated tool for Hospital Management.
- Business and Data Mining Objectives.
- Data Integration.
- Data Preparation.
- Learning the models.
- Examples.
- Conclusions.

WEKA

- WEKA is a collection of machine learning algorithms for data mining tasks.
- The algorithms can either be applied directly to a dataset or called from your own Java code.
- WEKA contains tools for data pre-processing, classification, regression, clustering, association rules, and visualization.
- WEKA is open source software issued under the GNU General Public License.
- WEKA is developed for the university of Waikato (New Zealand) [***http://www.cs.waikato.ac.nz/ml/weka/***]

Learning Methods

- We generated different minable views for some areas of hospital.
- The methods that we used of Weka:
  - Linear Regression
  - LeastMedSq
  - SMOreg
  - MultiLayerPercepton
  - Kstart
  - LWL
  - IBK
  - DecissionStump
  - Tree M5P



Predictive Models

- We obtained different predictive models:
  - Model of number of emergency admission per day.
  - Model of number of emergency admission per after day (from d+1 to d+7).
  - Model of number of emergency admission per week.
  - Model of number of emergency admission by different time periods.
  - Model of number of emergency admission per services of emergency.
  - Model of re-admission in emergency department.
  - ……etc.

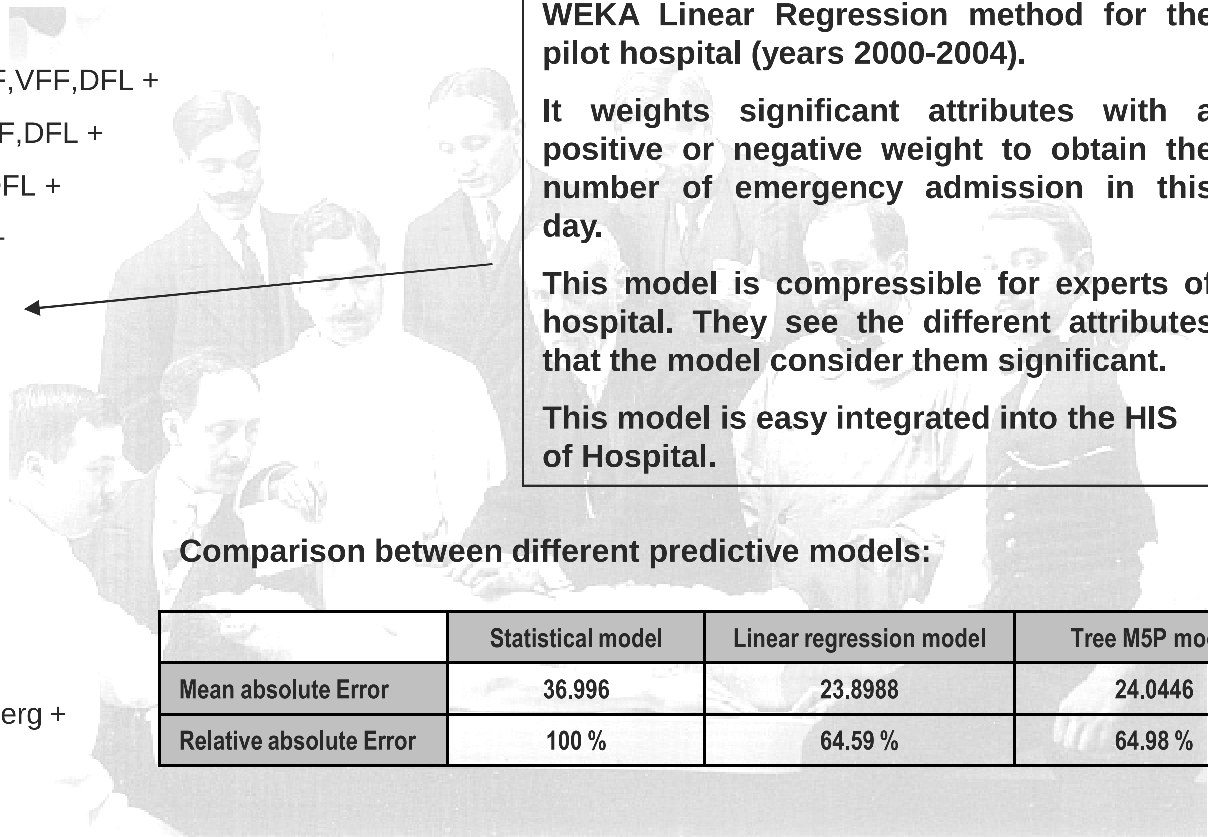
Index

- Introduction.
- Structure of an Automated tool for Hospital Management.
- Business and Data Mining Objectives.
- Data Integration.
- Data Preparation.
- Learning the models.
- Examples.
- Conclusions.

Initial Minable View

Attribute	SQL Type	Description/Values
day_week	nvarchar	Day of the week
type_day	nvarchar	Type of the day: F=Holiday, VFF=Before holiday, etc.
sport_events	nvarchar	TRUE/FALSE
average_temperature	float	Average temperature
rain	float	Amount of rain in mm^3.
numEmerg-1	integer	Number of admissions from the previous day.
numEmerg-2	integer	Number of admissions from the two previous days.
numEmerg-3	integer	Number of admissions from the three previous days.
numEmerg-4	integer	Number of admissions from the four previous days.
numEmerg-5	integer	Number of admissions from the five previous days.
numEmerg-6	integer	Number of admissions from the six previous days.
numEmerg-7	integer	Number of admissions from the seven previous days.
monthNumEmerg	integer	Average number of admissions in the same month of the year before.
dayWeekYearNumEmerg	integer	Average number of admissions in the same day of the week of all the year before.
daysBefHoli	integer	Number of holidays before the day.
daysAftHoli	integer	Number of holidays after the day.
numEmerg (class)	integer	Number of admissions in this day.

Predictive Model (LR)



numEmerg =

58.7937 * day_week=Monday +

14.806 * type_day=DFF,LL,NLF,VFF,DFL +

-11.7541 * type_day=LL,NLF,VFF,DFL +

32.3177 * type_day=NLF,VFF,DFL +

0.387 * average_temperature +

-1.6026 * rain +

0.2572 * numEmerg-1 +

0.076 * numEmerg-2 +

0.0678 * numEmerg-3 +

0.0748 * numEmerg-4 +

0.0473 * numEmerg-5 +

0.0702 * numEmerg-6 +

0.1193 * numEmerg-7 +

0.1496 * dayWeekYearNumEmerg +

-3.6549 * daysBefHoli +

50.438

We show the model obtained with the WEKA Linear Regression method for the pilot hospital (years 2000-2004).

It weights significant attributes with a positive or negative weight to obtain the number of emergency admission in this day.

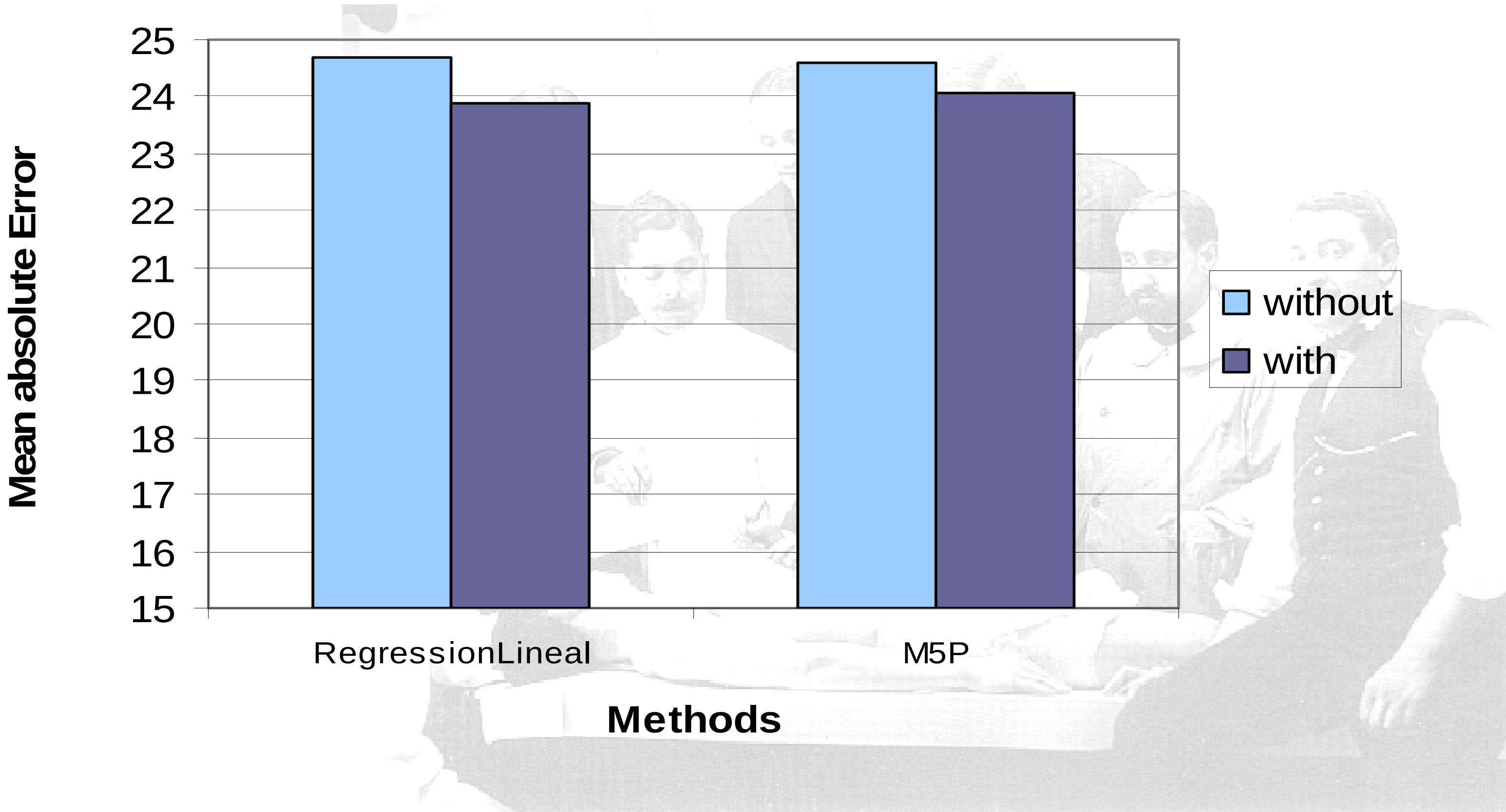
This model is compressible for experts of hospital. They see the different attributes that the model consider them significant.

This model is easy integrated into the HIS of Hospital.

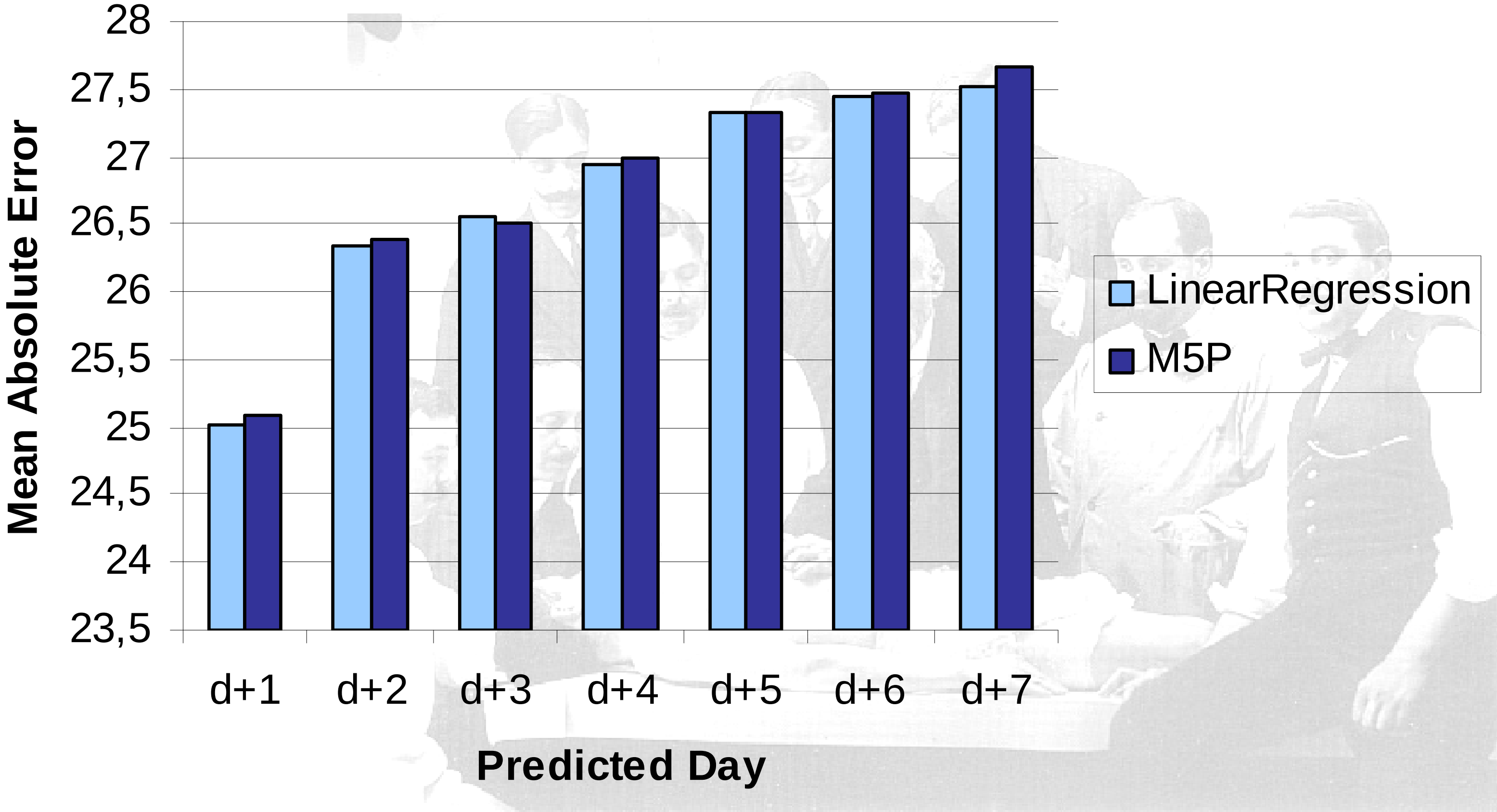
Comparison between different predictive models:

	Statistical model	Linear regression model	Tree M5P model
Mean absolute Error	36.996	23.8988	24.0446
Relative absolute Error	100 %	64.59 %	64.98 %

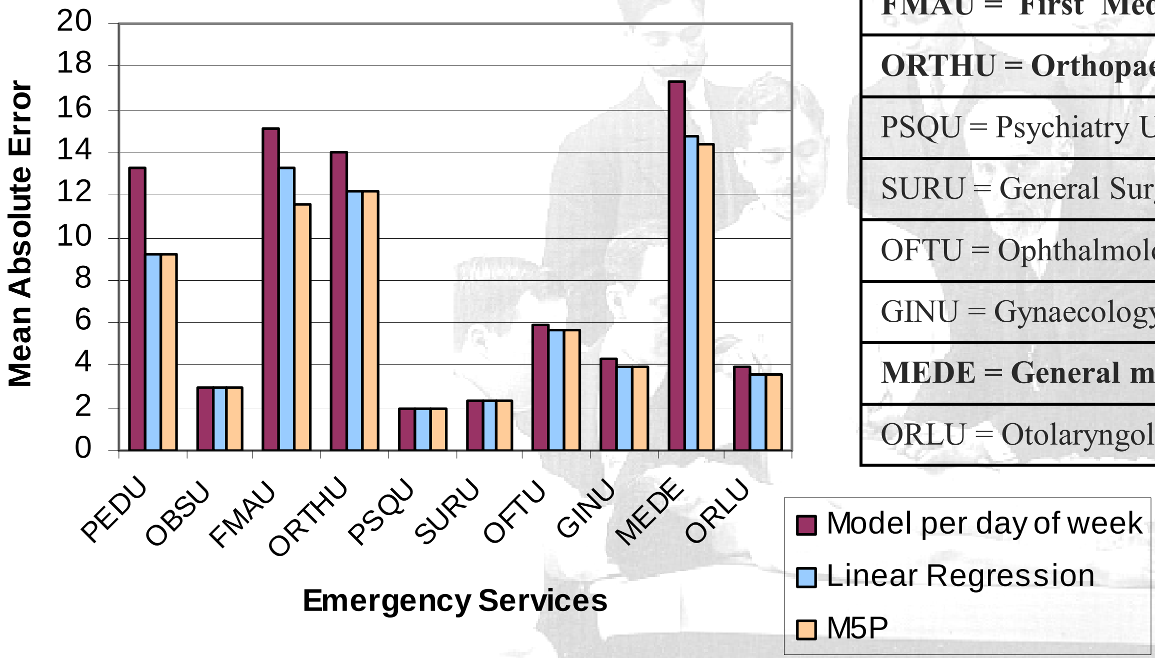
Without and With Meteorological Data



Prediction of number emergency admission from d+1 to d+7

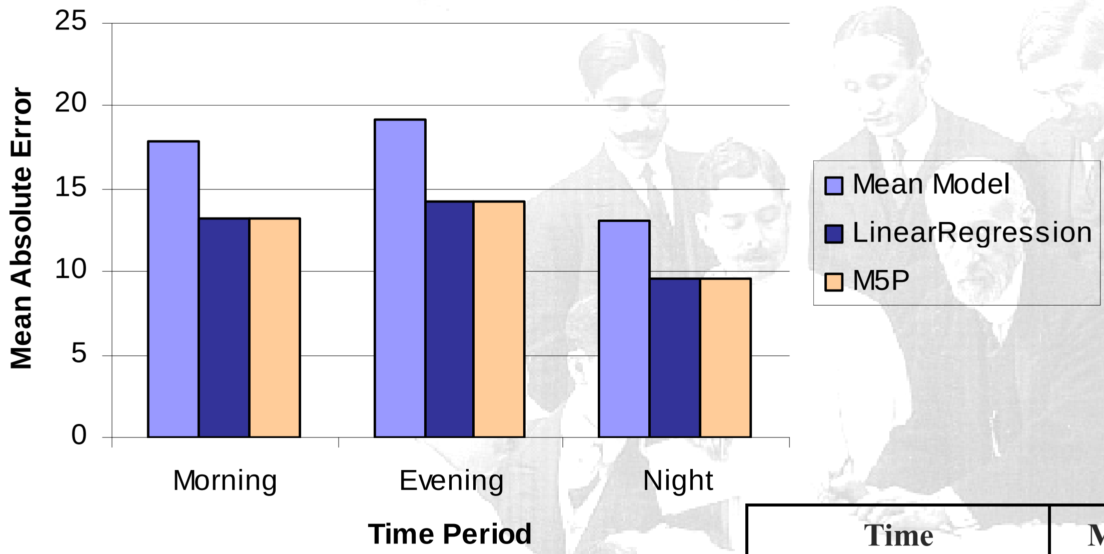


Prediction per Emergency Services



Service/Unit	Mean per day
PEDU = Pediatrics unit	68
OBSU = Obstetrics unit	11
FMAU = First Medical assistance Unit	**83**
ORTHU = Orthopaedics Unit	**148**
PSQU = Psychiatry Unit	5
SURU = General Surgery Unit	6
OFTU = Ophthalmology Unit	38
GINU = Gynaecology Unit	19
MEDE = General medicine Emergency	**104**
ORLU = Otolaryngology Unit	17

Prediction per Time Period



Time Period	Mean of number emergency admission
Morning	185
Evening	211
Night	96

Index

- Introduction.
- Structure of an Automated tool for Hospital Management.
- Business and Data Mining Objectives.
- Data Integration.
- Data Preparation.
- Learning the models.
- Examples.
- Conclusions.

Conclusions

- DM is still below its full potential in many areas. Healthcare, it's one of these areas.
- We identified which are the stages in the KDD process which could be reused and automated across different hospitals.
- In future, we like to extend the modules to modify or define new data mining objectives, not only the predefined data mining objectives identified in general.

Are there any questions ?

Thanks…..