



Data Mining techniques application in Power Distribution Utilities

Authors:

Sérgio Ramos
Zita A. Vale

GECAD – Knowledge Engineering and Decision Support Research Group
Engineering Institute of Porto – Polytechnic Institute of Porto

INTRODUCTION

Content of Presentation:

- **Electricity Market Liberalization Environment**
- **MV Customers Characterization – An overview**
- **Data Mining Techniques Application**
- **Clustering and Consumers Characterization**
 - **Case Study**
- **Future Work**

ELECTRICITY MARKET LIBERALIZATION

- ✓ Total freedom in choosing the electricity supplier
- ✓ Consumers and suppliers are exposed to price risk
- ✓ Distribution and retail companies are looking for better tariff rates
- ✓ Competitive environment among retail companies to sell the electricity



ELECTRICITY MARKET LIBERALIZATION

- ✓ Increase of demand elasticity due to the electricity price volatility
- ✓ Electricity customers more concerned about their consumption behaviour
- ✓ Knowledge about customers' daily load profile is essential for leadership in this new context
- ✓ Deeper relationship between Customer and Electricity Supplier



ELECTRICITY MARKET LIBERALIZATION

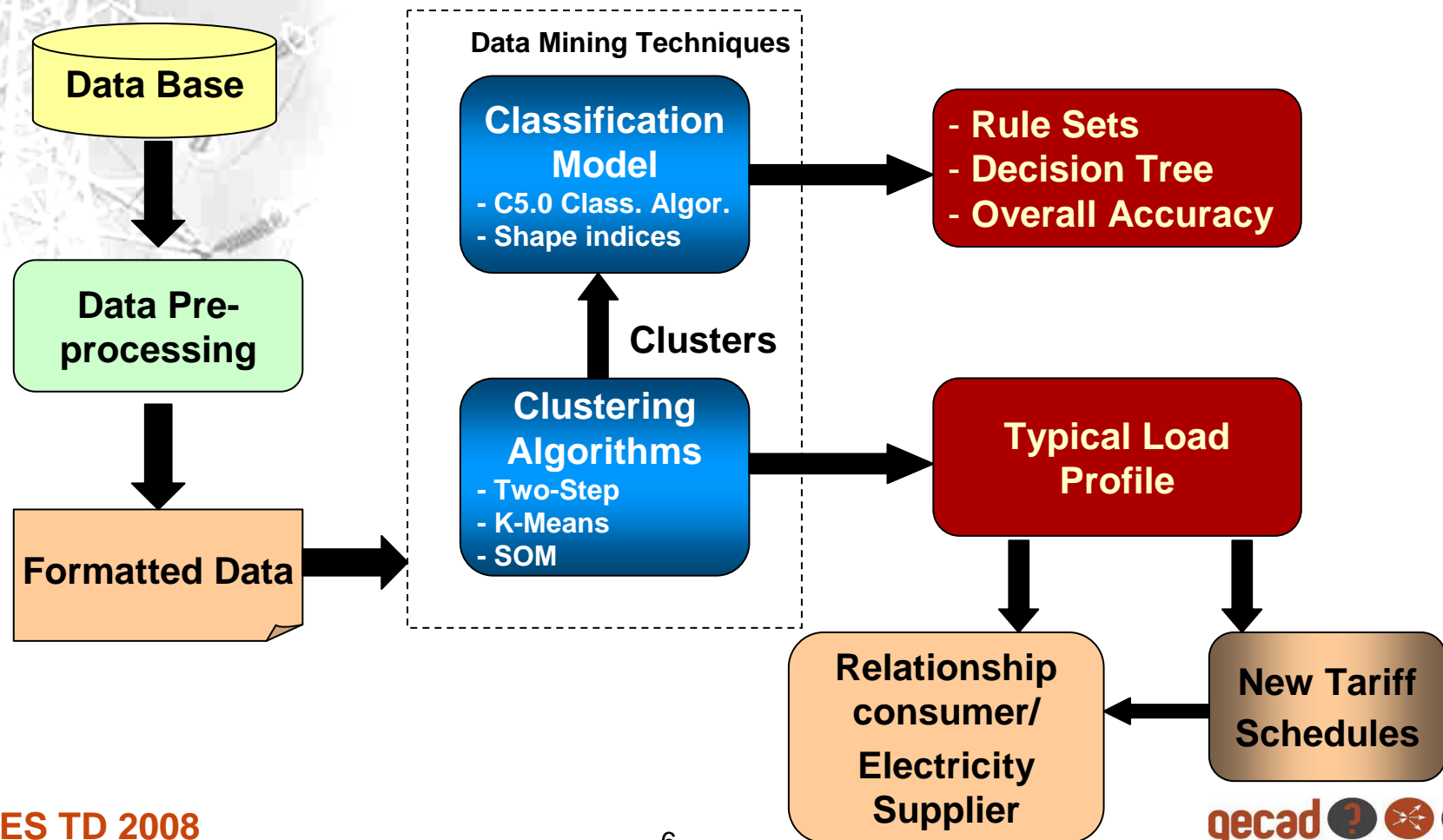
CONSUMERS' CHARACTERIZATION

▪ Advantages:

- ✓ Design of new tariffs, contracts, products and services
- ✓ Creation of incentive actions to the energy efficiency

MV COSTUMERS CHARACTERIZATION

DETERMINATION AND CHARACTERIZATION OF MV CONSUMERS LOAD PROFILE USING DATA MINING TECHNIQUES



LOAD STUDY

Data description:

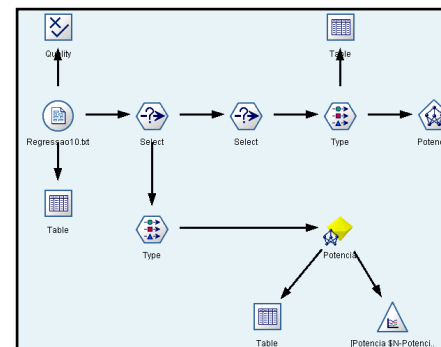
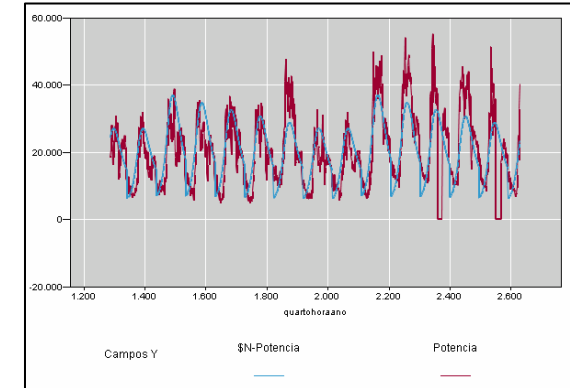
- ✓ Sample of 229 MV Consumers
- ✓ Collect period of data
(3 months in the Summer / Winter – from the Portuguese Distribution Company)
- ✓ Consumed power recorded with a cadence of 15 minutes
- ✓ 96 values obtained per day

$I^{(m)} = \{I_1^{(m)}, \dots, I_{96}^{(m)}\}$
with $m = \text{number of customers}$

DATA PREPARATION

Data-Cleaning

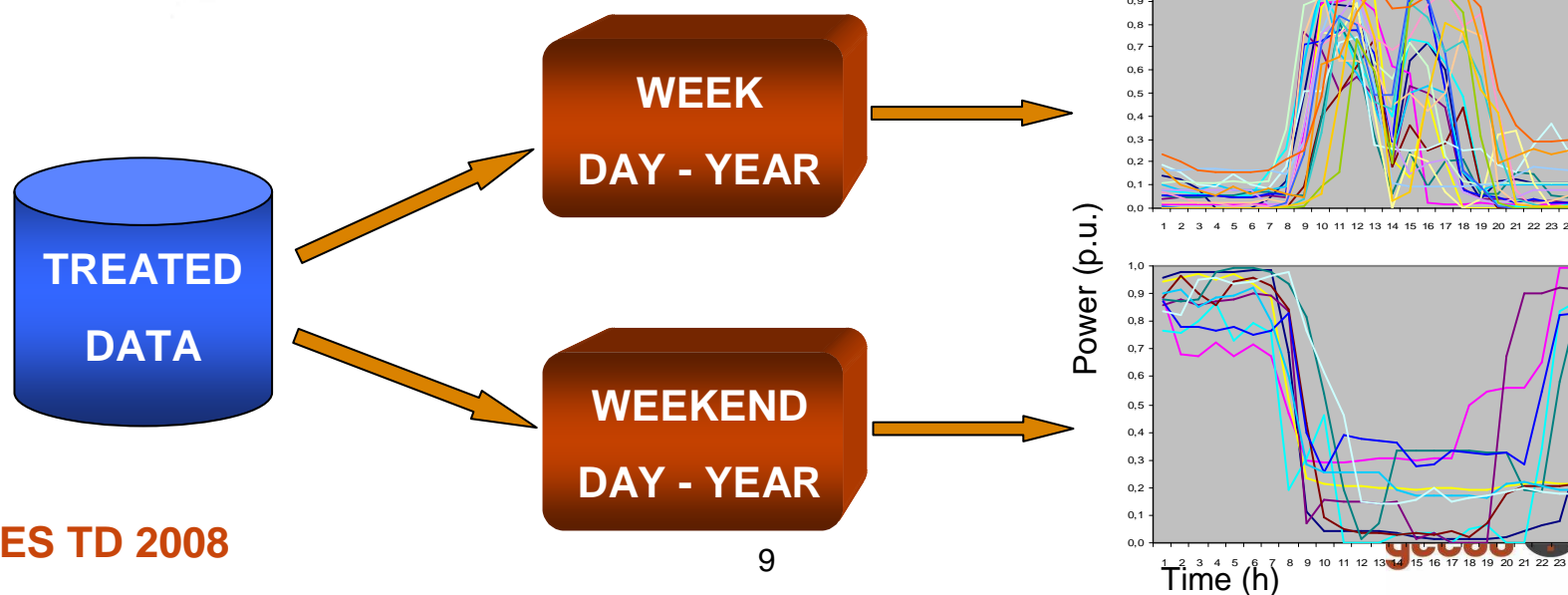
- ✓ **21 customer's files were discarded**
 - Some damaged files were detected
 - Customers without registered values
 - 208 customers remained to be analyzed
- ✓ **To estimate missing values of measures a multi layer perceptron – MLP – artificial neural net was used**
- ✓ **The errors of the metered load curves are attenuated without making significant alterations in the real measures**



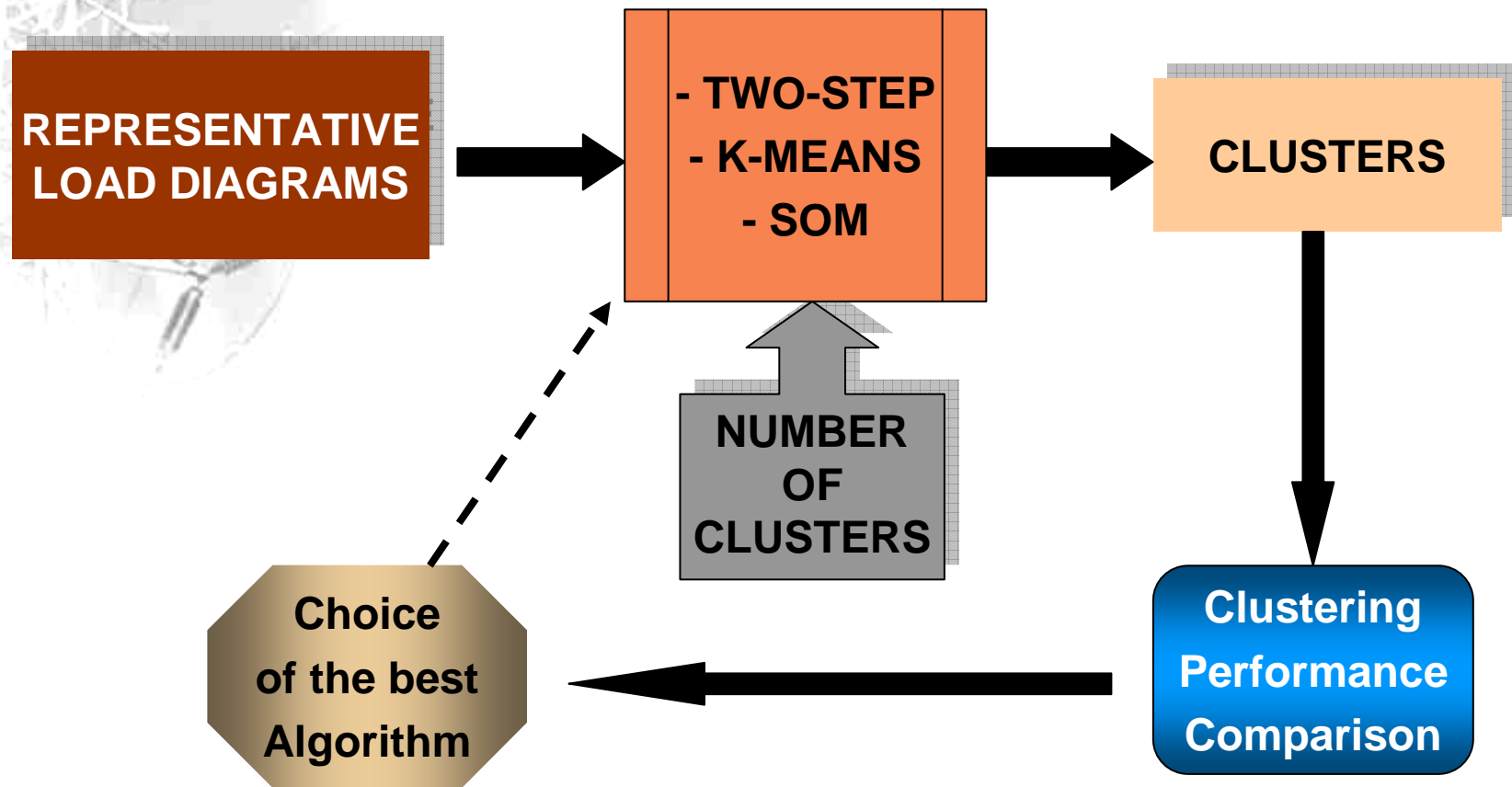
DATA PREPARATION

Pre – processing data:

- ✓ The Power consumption was normalized to the [0-1] range
- ✓ A representative load diagram has been built for each customer by averaging the related load diagrams



CLUSTERING PROCESS



CLUSTERING PROCESS

Choice of the Clustering Algorithm:

Mean Index Adequacy (MIA):

$$MIA = \sqrt{\frac{1}{K} \times \sum_{k=1}^K d^2(r^{(k)}, L^{(k)})}$$

Cluster Dispersion Index (CDI):

$$CDI = \frac{\sqrt{\frac{1}{K} \sum_{k=1}^K \left[\frac{1}{2 \cdot n^{(k)}} \sum_{n=1}^{n^{(k)}} d^2(l^{(m)}, L^{(k)}) \right]}}{\sqrt{\frac{1}{2K} \sum_{k=1}^K d^2(r^{(k)}, R)}}$$

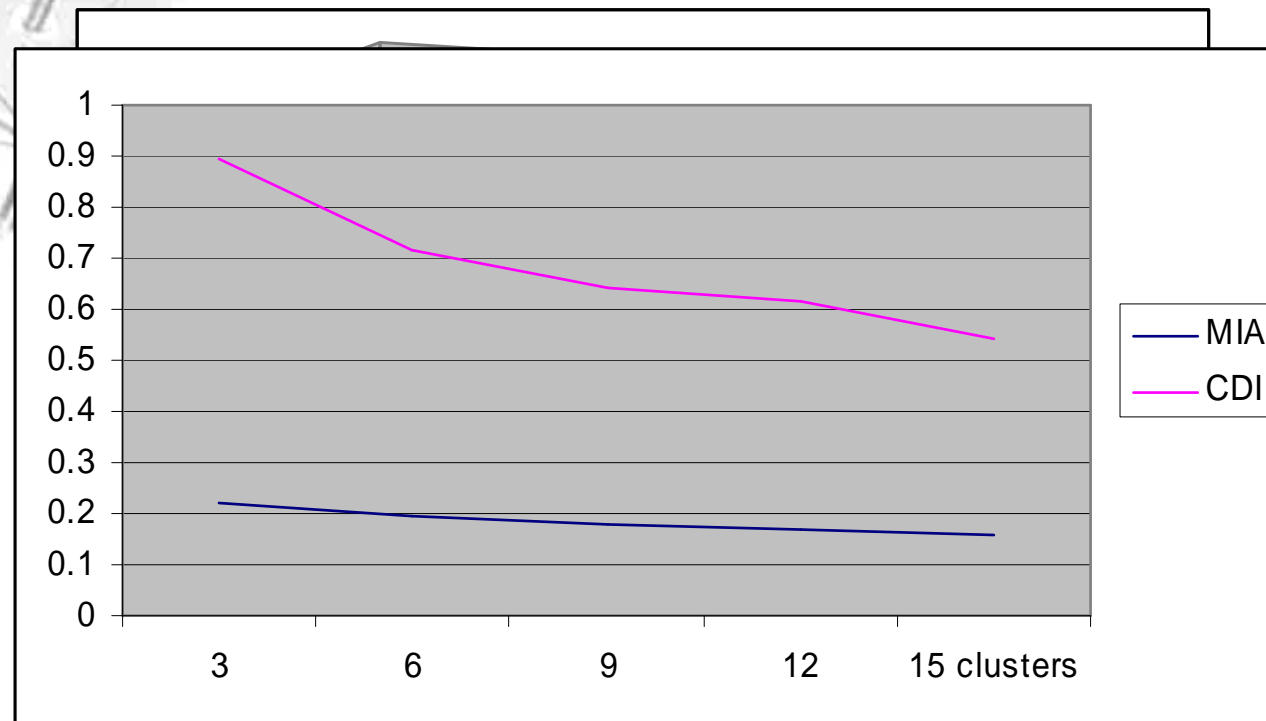
CLUSTERING PROCESS

Comparison of the clustering performance:

Two-Step Cluster Analysis

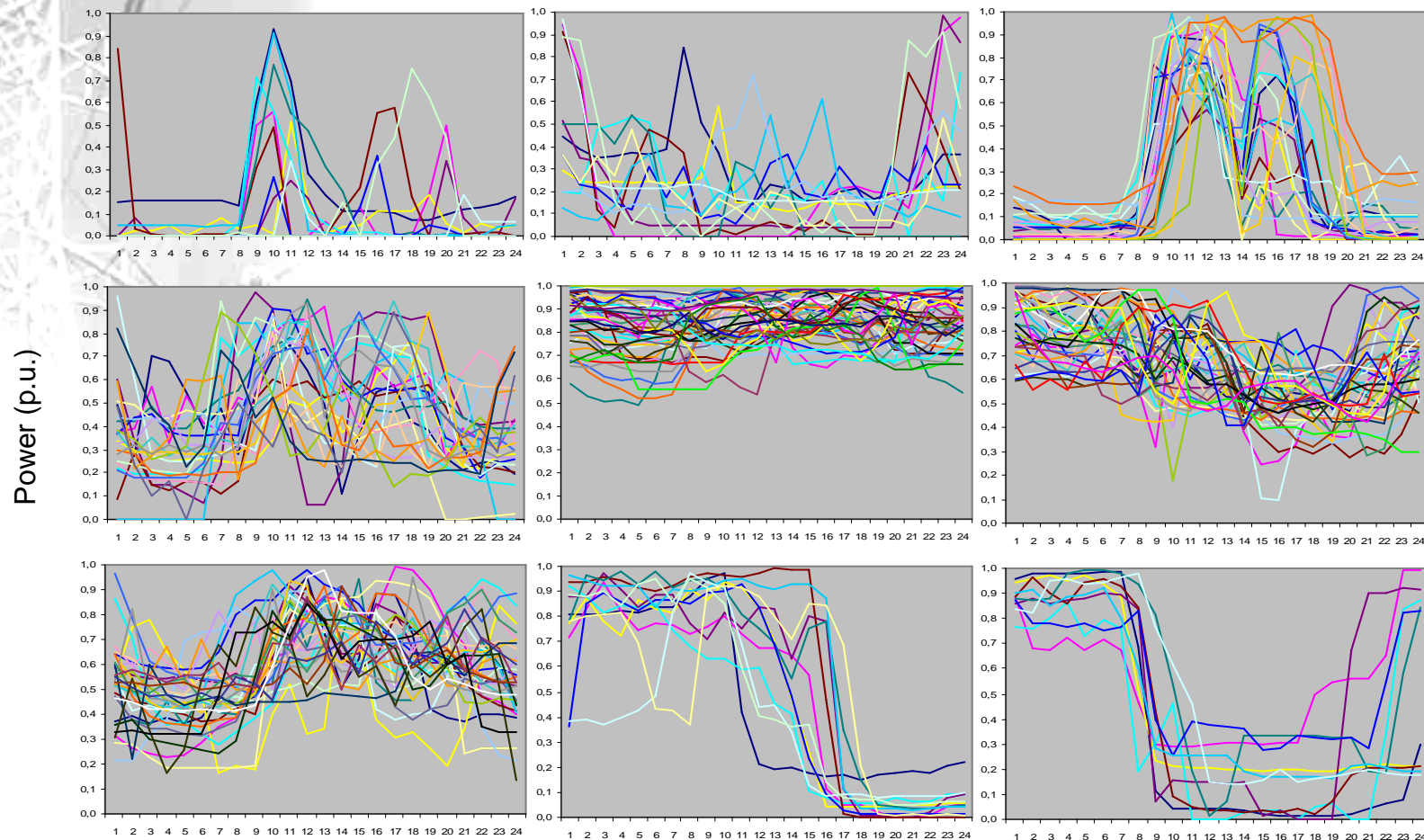
K-means

Kohonen Net – Self Organizing Features Maps



TWO-STEP CLUSTERING APPLICATION

- ✓ Using the **Two-step** cluster algorithm the clusters were obtained using the representative load diagrams

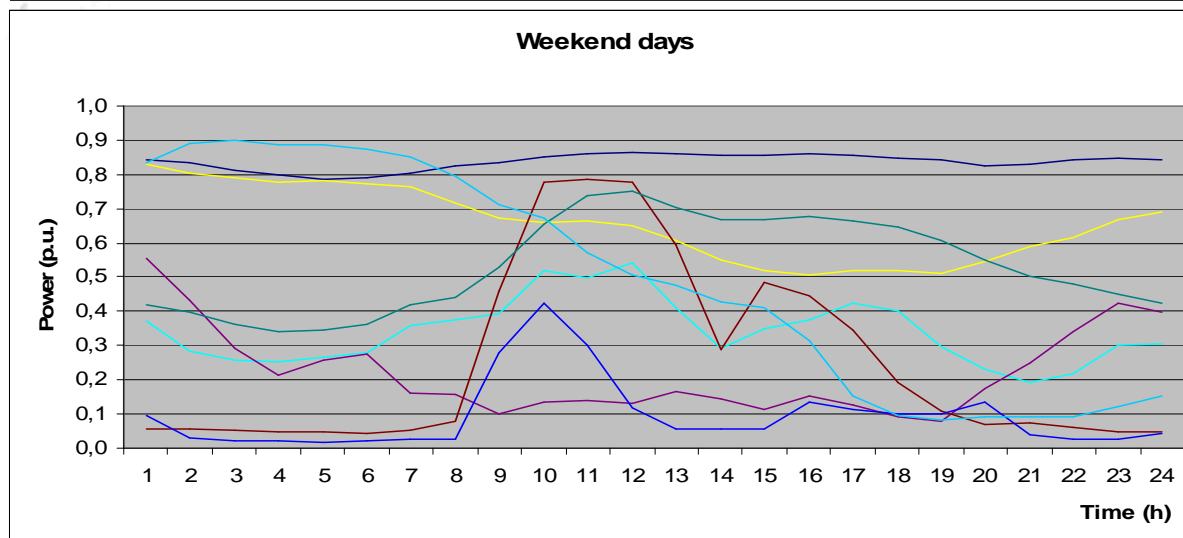
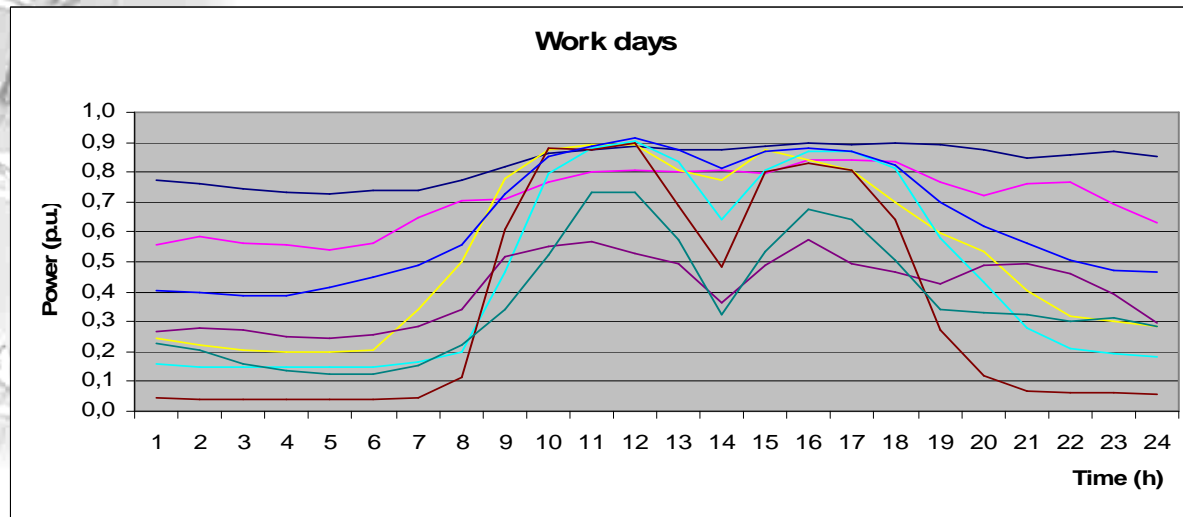


Working days

Time (h)

REPRESENTATIVE LOAD DIAGRAMS

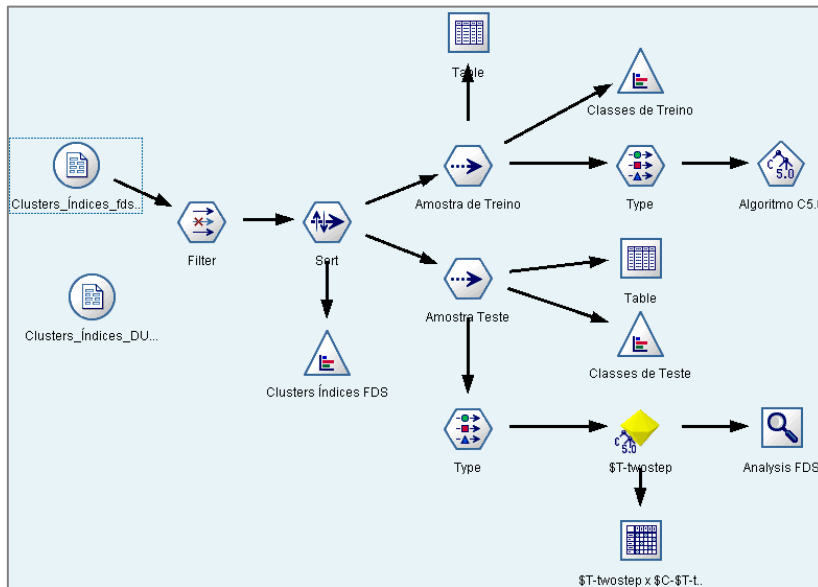
✓ Representative diagram for each cluster



CLASSIFICATION MODEL

Objective:

- ✓ To build a classification model, that applied to new unclassified records, will allow to foresee the class to which it belongs
- ✓ In the future it will allow to attribute to each new consumer the consumption profile that best represents it



CLASSIFICATION MODEL

C5.0 Algorithm

- ✓ **Decision Tree:**
 - **Application simplicity**
 - **Result in tree form**
 - **Generation of rules**

$$f1 = \frac{P_{av,day}}{P_{max,day}}$$

$$f2 = \frac{P_{min,day}}{P_{max,day}}$$

$$f3 = \frac{P_{min,day}}{P_{av,day}}$$

$$f4 = \frac{1}{3} \frac{P_{av,night}}{P_{av,day}}$$

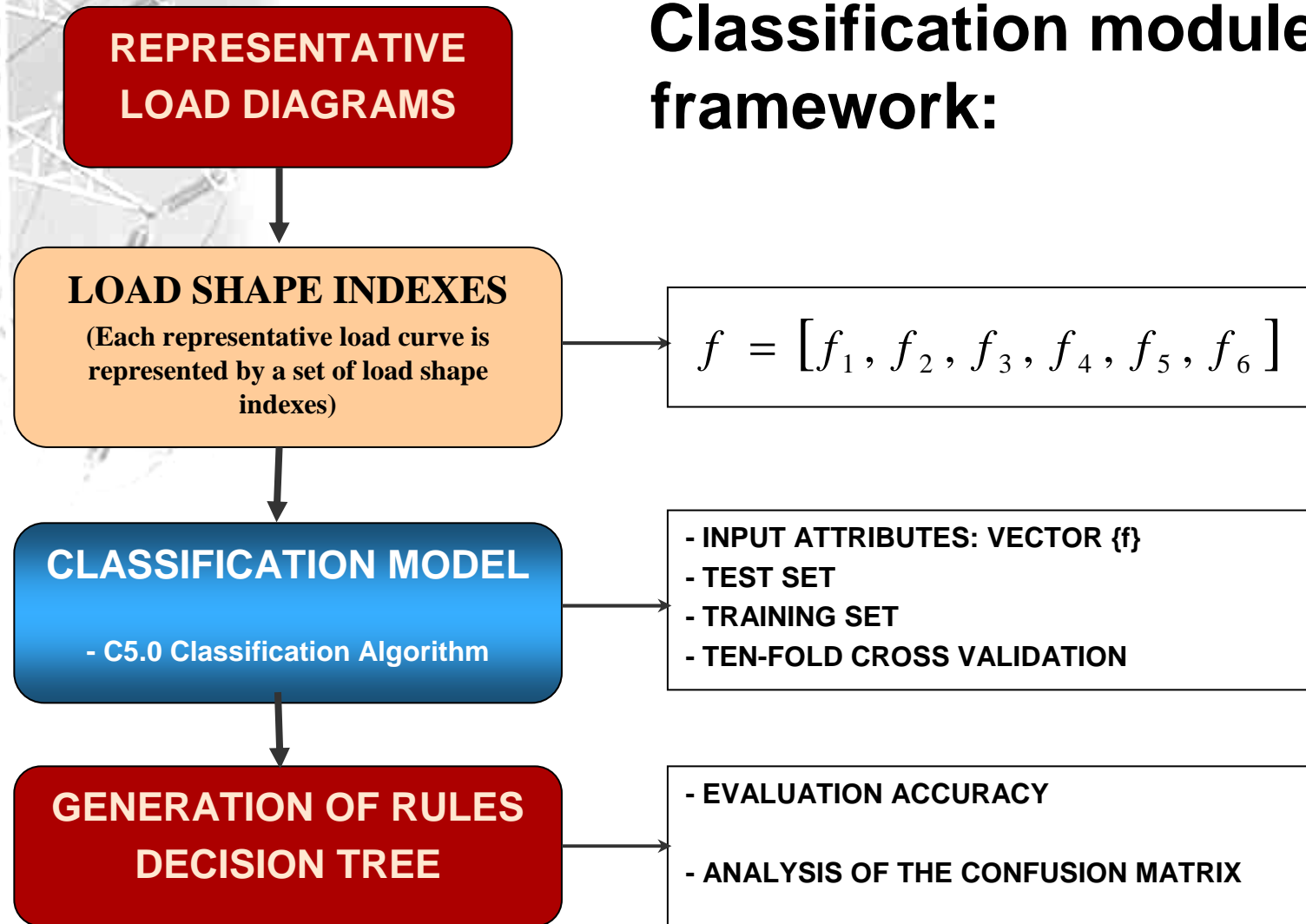
$$f5 = \frac{1}{8} \frac{P_{av,lunch}}{P_{av,day}}$$

Derive from the daily load diagrams

- ✓ **Give information about:**
 - **The daily load curve shape**
 - **The consumption pattern of each consumer**

CLASSIFICATION MODEL

Classification module framework:



CLASSIFICATION

✓ Rule set for the working days classification model:

if $f_3 \leq 0,48$ and $f_2 \leq 0,13$ and $f_5 \leq 0,55$ and $f_1 \leq 0,35$ and $f_4 \leq 0,31$	then cluster 8
if $f_3 \leq 0,48$ and $f_2 \leq 0,13$ and $f_5 \leq 0,55$ and $f_1 \leq 0,35$ and $f_4 > 0,31$	then cluster 9
if $f_3 \leq 0,48$ and $f_2 \leq 0,13$ and $f_5 \leq 0,55$ and $f_1 > 0,35$	then cluster 5
if $f_3 \leq 0,48$ and $f_2 \leq 0,13$ and $f_5 > 0,55$ and $f_2 \leq 0,06$	then cluster 7
if $f_3 \leq 0,48$ and $f_2 \leq 0,13$ and $f_5 > 0,55$ and $f_2 > 0,06$	then cluster 6
if $f_3 \leq 0,48$ and $f_2 > 0,13$ and $f_4 \leq 0,24$	then cluster 4
if $f_3 \leq 0,48$ and $f_2 \leq 0,13$ and $f_4 > 0,24$	then cluster 5
if $f_3 > 0,48$ and $f_3 \leq 0,78$ and $f_2 \leq 0,44$	then cluster 3
if $f_3 > 0,48$ and $f_3 \leq 0,78$ and $f_2 > 0,44$	then cluster 2
if $f_3 > 0,48$ and $f_3 > 0,78$	then cluster 1

✓ Overall accuracy:

94,83%

CONSUMER-SUPPLIER RELATIONSHIP

- ✓ **The Knowledge can be used by the Retail Companies**
 - **Identify diagrams' peaks**
 - **Develop specific consumer's contracts**
 - **Optimization of the offers of electric power purchase**

CONSUMER-SUPPLIER RELATIONSHIP

- ✓ **The Knowledge can be used by the electric power consumers**
 - **Choice of the electricity supplier with the best tariff schedule proposal**
 - **Modulation of their electric consumption habits**
 - **In the execution of electric energy interruption contracts**

FURTHER WORK

- ✓ **Compare the efficiency of the C5.0 algorithm with different classification algorithms**
- ✓ **The design of new prices categories in order to adequately adapt the tariff schedules to the cluster consumption pattern**
- ✓ **Formulation of new tariffs schedules in articulation with electricity markets**



Data Mining techniques application in Power Distribution Utilities

Authors:

Sérgio Ramos
Zita A. Vale

GECAD – Knowledge Engineering and Decision Support Research Group
Engineering Institute of Porto – Polytechnic Institute of Porto