A DATA MINING DECISION SUPPORT TOOL TO STUDY ELECTRICITY RETAIL CONTRACTS

Vera Figueiredo veraf@dee.isep.ipp.pt
Fátima Rodrigues fr@dei.isep.ipp.pt
Ricardo Pinto rjp@dee.isep.ipp.pt
Zita Vale zav@dee.isep.ipp.pt

GECAD – Knowledge Engineering and Decicion Support Group
ISEP – Polythecnic Institute of Porto
Rua Dr. António Bernardino de Almeida
4200-072 Porto, Portugal

Abstract – This paper presents a Decision Support System (DSS) to help both retail companies and electricity consumers on the definition of the best retail contracts and tariffs. This DSS is supported by a Consumer Characterization Framework. This one is based on data mining techniques, applied to obtain useful knowledge about electricity consumers, from large amounts of consumption data stored in utilities. This knowledge is acquired following an innovative and systematic approach able to identify different consumers' classes, represented by a load profile, and its characterization using decision trees. This framework generates inputs to use in the knowledge base and in the database of the DSS. The rule sets derived from the decision trees form the knowledge base of the DSS. The load profiles together with the information about contracts and electricity prices form the database of the DSS. This DSS will be able to perform the classification of different consumers, present their load profile and test different electricity tariffs and contracts. The final outputs of the decision support system are a comparative economic analysis between different contracts, and advice about the most economic contract to each consumer class. The presentation of the DSS is completed with some application examples using a real data base of consumers from the Portuguese distribution company.

Keywords: Electricity retail markets, electricity contracts, data mining, load profiles, rule base, decision support system.

1 INTRODUCTION

The full liberalization of most of the electricity markets in Europe and around the world creates a new environment were several retail companies compete for the electricity supply of end users. According to [1] retail markets have been much less successful than wholesale markets and they rarely give consumers the opportunity to buy electrical energy at spot prices. A more active participation of demand side would make electricity markets more efficient and more competitive and promote an optimal allocation of resources. For this purpose, in [1], the author points out the creation of tools to support retailers and consumers of electrical energy very as an important issue. Another issue is the definition of new rules and structures, concerning data collection and description, and the definition of communication protocols between the different participants in the market. This new structures will increase significantly the amounts of data collected by the participants in the market. This data grows up in a dynamic form and can play an important role in the decision support and in the definition of market strategic behavior. The importance of this data increases with the progressive replacement of traditional electricity meters by real-time-meters. The development of frameworks and tools able to extract useful knowledge from this data and the use of this knowledge in decision support, can be a competitive advantage for the participants in the market and an important pass to get a more active demand side participation. The freedom on definition of a larger portfolio of contracts by retail companies, and the freedom consumers have on the choice between different contracts and different companies increases the need of decision support tools to help both sides. It is possible to find in the literature some previous works dedicated to this problem. In [2] data mining techniques are applied to the problem of load profiling. In [3] a load research project followed by load profiling is presented and the results of this work are used to support tariff design. In [4] consumers classes and their load profiles are defined by clustering techniques and the results are used to study different contracts for producers. In [5] we present an automatic classification and characterization of electricity consumers' framework, able to deal with large amounts of data, and perform the classification of different consumers and present their load profiles. In the present paper the framework is used to provide inputs to a Decision Support System (DSS) able to test different tariffs and contractual structures and provide economical analysis results. This DSS is able to use simple knowledge about consumer characteristics to perform classification, load profiling and cost calculation important to help retailers and consumers on the selection of the Best Contract. This tool is important to increase demand participation in electricity markets. The paper is organized as follows: in section two a brief description of the consumer characterization framework developed and developed and carefully detailed in [5] is made. Section three presents the description of the DSS and in section four some practical results complete this description. Finally, in section five we present our conclusions.

2 CONSUMER CHARACTERIZATION

2.1 Consumer Characterization Methodology

The knowledge about how and when consumers use electricity is essential to develop an efficient DSS. This knowledge must be obtained from historical data and must be updated to follow the changes on consumer's behavior. To generate this kind of knowledge and keep it regularly updated a comprehensive methodology was developed. Due to the large amount of data predicted to be available in the future and the necessity of easy actualization, our methodology provides a clear separation of different steps based on the application of Data Mining techniques. The proposed methodology is based on the previous load profiling projects [6, 7] and on the structure of the Knowledge Discovery in Databases process. The major steps identified are:

1. <u>Load Research Project</u>: in this phase a representative sample of the population of electricity consumers is identified, the most relevant attributes to be measured, the cadence for data collection is defined and the measurement equipment is selected and installed. The duration of the measurement campaigns is also defined in this step. Finally the collected data is gathered in a large database.

2. <u>Data Cleaning and Pre-processing</u>: In real problems, like this, involving a large number of measurements, spread over a large geographic area, collecting data during a considerable period of time (ideally not less than a year) different kind of problems will affect the quality of the database. The most relevant and frequent are communication problems, outages, failure of equipment and irregular atypical behavior of some consumers. The result will be a very large database with problems like noise, missing values and outliers. At this point it is important to eliminate these problems so they can't affect the final results. For this purpose the first step is to estimate all the missing values. Statistical regression techniques and artificial neural networks are used on the data completion phase to estimate the missing data. Other important point was the detection of outliers that would interfere with the determination of the representative load diagrams. These outliers are detected using several interactive graphics replaced using adequate statistical regression techniques. The whole process in detailed in [5].

3. <u>Development of the Consumer Characterization Model</u>: The consumer characterization model is developed according to the attributes available, the techniques selected and the expected precision of the obtained results. In this work our major goal is to obtain relevant knowledge about consumer's behavior to use in the DSS. This knowledge is based on consumer classification, load profiling and consumer characterization. In this step the data mining techniques are selected and the general model is designed. This process is detailed in [6, 7].

4. <u>Model Evaluation</u>: this step is of major importance because it will provide the confidence on the knowledge obtained before using it in the DSS. We performed this evaluation using real data to evaluate the precision of the algorithms, using cluster validity indexes for clustering and cross validation for the decision trees. These results are presented in [5].

2.2 Structure of the Consumer Characterization Framework (CCF)

Figure 1 presents the general structure of the framework. Each step of this framework is based on the major steps of the previous methodology [5]. The major innovation of this framework is the combination of supervised and unsupervised learning to obtain useful knowledge from large amounts of mean less data. This data (after being cleaned, pre-processed and reduced) is used by the data mining model to obtain the division of the initial data set in classes. Each of these classes are represented by their load profile and described by a rule set. The Data Mining model also performs the identification of the most relevant attributes for consumer characterization which will be used as inputs in the DSS.

3 DECISION SUPPORT SYSTEM

3.1 Introduction

According to [1] the development of decision support tools is of major importance to show consumers the potential savings they can get by assuming a more active participation in the electricity markets. Also a company with the opportunity to trade energy to consumers, in a competitive scenario, must face a series of decisions and evaluations to attain the best tariff structure and to study the portfolio of contracts to offer to the consumer. Taking these decisions and correctly evaluating the options is a complex and time consuming process. So, tools providing assistance in evaluation and decision making are of considerable value to retail companies and consumers. Under this scenario comes the development of the decision tool for the most adequate contract (Best Contract) selection presented in this paper.

The DSS must easily permit the introduction of contractual consumer characteristics and the most relevant attributes on the description of consumer's behavior. These input parameters will start a decision process that ends with the presentation of the Best Contract. This decision is completed by a comparative analysis between all the options available to make clearer the choice of the best and the potential gain obtained with the change.

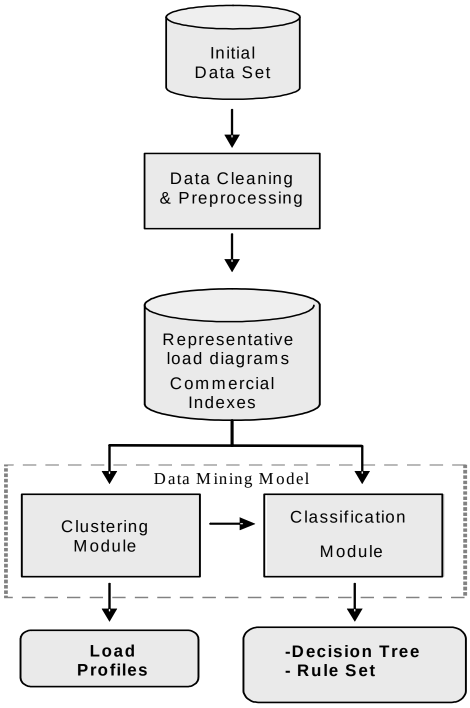

Figure 1: Structure of the Consumer Characterization Framework

3.2 *DSS Architecture*

The internal structure of the DSS is composed by:

- a knowledge base where the rules sets obtained by the CCF are stored;
- a data base where all the information about load profiles and electricity tariff structures, available or being tested in the company, are gathered;
- a working memory that contains all the information about the contract that is either supplied by the user or inferred by the system during the session;
- a reasoning base that matches the facts contained in the working memory with the knowledge contained in the knowledge base, to draw conclusions;
- a user interface to easily collect consumer characteristics as inputs, start the decision process and clearly present the output results.

In Figure 2 the structure of the DSS is presented. Besides the main modules previously presented, the DSS has also an interface to the CCF that provides means of storing and updating the knowledge base and the data base, to assure their actualization. This function will gain more importance as real-time-meters are being introduced in smaller consumers and more actual data is collected and also as new tariffs and contracts need to be tested and presented to consumers.

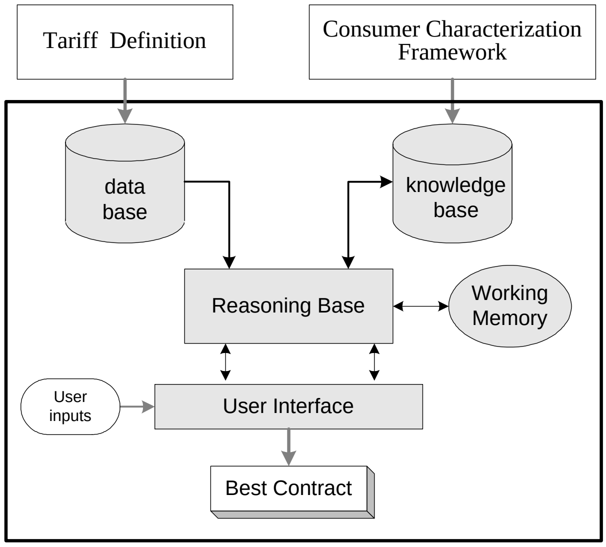

Figure 2: The structure of the DSS

3.2.1 *Knowledge Base*

The rules that define and characterize the client's load profiles are stored in the knowledge base. These rules are created by the CCF and consider the load shape indexes: load factor index (d1), night impact index (d3), presented in Table 1, and the contractual input: contracted power (represents the maximum active power the consumer can use). These parameters are the data inputs needed by the DSS Reasoning Base to select the Best Contract.

To get an easier and more practical data manipulation by the DSS, we have simplified the input parameters into a set of intervals. The user does not need to define a specific value for each parameter; he needs only to select a categorical value like: low, medium, high, very high and ultra high.

TABLE I
NORMALIZED LOAD SHAPE INDEXES

Parameter	Definition	Period of definition
Load Factor	$d_1 = \frac{P_{av,day}}{P_{\max,day}}$	1 day
Night Impact	$d_3 = \frac{1}{3}\frac{P_{av,night}}{P_{av,day}}$	1 day (8 hours night, from 11 p.m. to 7 a.m.)

This kind of simplification was necessary to allow the practical application of the DSS to small consumers without real-time-meters. While the electricity meters are not replaced by real-time-meters a LV consumer does not know the exact value of his load factor or night impact. On the other hand, it is possible, based on simple information about consumption habits, to decide to which of the categories presented below, these indexes might belong. This simplification will lose some precision but will permit the application of the DSS system on LV consumers.

The load factor and the night impact were classified into the following discretized intervals.

$$\begin{cases} d1 \in [0{,}2;0{,}3] \rightarrow low \\ d1 \in [0{,}3;0{,}4] \rightarrow medium \\ d1 \in [0{,}4;0{,}5] \rightarrow high \\ d1 \in [0{,}5;0{,}6] \rightarrow veryhigh \\ d1 \geq 0{,}6 \rightarrow ultra.high \end{cases} \quad \begin{cases} d3 \in [0;0{,}2] \rightarrow low \\ d3 \in [0{,}2;0{,}3] \rightarrow medium \\ d3 \in [0{,}3;0{,}4] \rightarrow high \\ d3 \in [0{,}4;0{,}6] \rightarrow veryhigh \\ d3 \geq 0{,}6 \rightarrow ultrahigh \end{cases}$$

The resulting knowledge base is much simpler, easy to manipulate and permits to perform a classification on low voltage consumers. In Figure 3 we present as an example the rule set obtained for winter-working days classes using a data base of Portuguese consumers. To each one of these classes we have a different load profile that will be used as reference to the calculations of costs and observe potential savings.

If	d1	High	and	d3	Very High	then	Class 1
If	d1	High	and	d3	Ultra High	then	Class 1
If	d1	Very High	and	d3	Very High	then	Class 1
If	d1	Very High	and	d3	Ultra High	then	Class 1
If	d1	Very High	and	d3	Medium	then	Class 3
If	d1	Ultra High				then	Class 3
If	d1	Low				then	Class 4
If	d1	Medium				then	Class 4
If	d1	High	and	d3	Low	then	Class 7
If	d1	Very High	and	d3	Medium	then	Class 7
If	d1	Very High	and	d3	Low	then	Class 7
If	d1	High	and	d3	Medium	then	Class 9
If	d1	High	and	d3	High	then	Class 9

Figure 3: Discretized Rule Set for the Winter Working Days classification model.

The knowledge base is composed by four different rule sets corresponding to different load conditions as winter weekends and working days and summer weekends and working days. Different calculations are performed for winter or summer.

3.2.2 Data Base

The implementation of the DSS started with the definition of a clear separation between data storage and organization and the reasoning base that performs the calculations to obtain the Best Contract. This separation is very important to have a flexible DSS able to deal with large amounts of data [8]. The information gathered in this database is composed by the load profiles obtained by the CCF and the tariff structures corresponding to different contracts to be used in the economical study.

In the example presented in this paper the data base comprehends different types of contracts:

<u>Fixed Rate (**FR**)</u>: electricity hourly prices based on the Portuguese regulated tariffs for 2004 [9] considering a fixed rate contract.

<u>Time-of-Use (**TOU**) Contracts</u>: electricity hourly prices and peak and off peak intervals based on the Portuguese regulated tariffs for 2004 [9]. These contracts have different prices for peak and off peak hours and consider the possibility of a weekly based plan: **TOU-WC** (different cycle for working days and weekends); or a daily plan: **TOU-WDC** (same plan for all the week days).

<u>Tailored Contracts (**TC**)</u>: these contracts were developed according to each load profile and follow the TOU tariff principle. They are designed to become more attractive to the consumers and at the same time consider the retailer interests. TC contracts are developed not to a single consumer but to a group of consumers represented by a load profile. This is a strategy of consumer satisfaction with a smaller portfolio of contracts and lower management expenses. The DSS can be easily adjusted to perform different test simulation to different tailored contracts with different profit levels. Besides the retailer profit these contracts can also include an insurance factor considering the level of risk shared by both parts involved in the contract. Figure 4 presents an example of a tailored contract for class 7 developed to test the DSS.

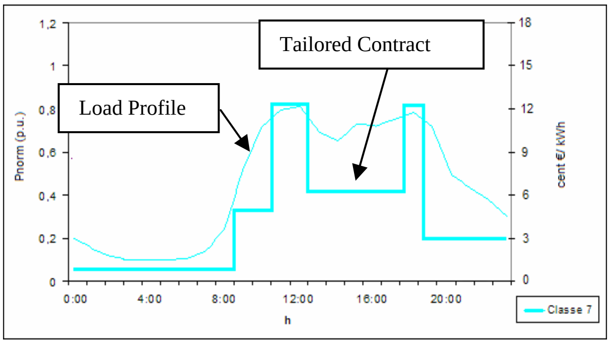

Figure 4: Tailored Contract (TC) developed to Class 7

Observing the figure we can see that the contract follows the load profile and we have more then just two different prices in a day. These prices change in order to satisfy the consumer and to create incentives to change consumption habits like reducing demand at peak hours using higher prices or creating more attractive prices to night consumers. This is a hypothetical test situation but the DSS is flexible enough to easily allow changes in the test contracts to study different structures and profit levels.

The apllication of this TC can be an important contribution to increase demand elasticity wihout significant investments in metering and communication systems. This can also be an important mean for sharing part of the risk of market price spikes with consumers. The possibility of developing a TC for a large group of consumers (represented by a consumer class), will contribute to the increase of adequacy of electricity prices. This leads to consumer satisfaction without significant investments on consumer managment resourses. This concept is already being apllied in the United Kingdom with the designation of Standard Contracts.

Real-time-pricing (RTP): the example presented in this paper is based on the Spanish market spot prices for the year 2003 [10]. The real time pricing is introduced as a test example considering only a different price profile for a typical winter and a typical summer day. In the future these prices must be actualized in a real time basis for the 24 hours of the next day, as market clearing prices are computed. This, of course, will depend on a significant evolution of communication systems between market and retail companies. Nevertheless, to increase the relevance of the results, we used real data from 2003 Spanish spot market. To create the typical price profiles the hierarchical clustering algorithm two-step [11] was used to perform the search for the typical winter and summer prices. These were computed to be used by the DSS as test values. The winter and summer price curves computed are presented in Figures 5 and 6.

In the future these values can be updated in the data base in order to have more accurate information on the most adequate tariff.

To obtain the cost of electricity at the LV consumer level to the spot prices must be added other costs. We followed the technique of additive tariffs presented in [12]. After computing the prices of different activities (transmission, distribution, commercialization and metering) using the additive methods, based the Portuguese regulated prices, we computed the electricity prices at the LV end users.

To this method must be added the retailer profit that can be variable according to the retailer commercial strategy and an insurance factor related to the risk of participation in the spot market. The final electricity price per day, to be paid by the LV end user, is computed using expression (1).

$$T = \sum_{h=1}^{24} W_h . TW_h + TPc \quad \text{Euros/Day} \qquad (1)$$

where T is the electricity price per day

W_h is the active energy consumed at hour k

TW_h is the active energy price at hour h

TP_C is the cost of contracted power or other retailer profit and insurance factor

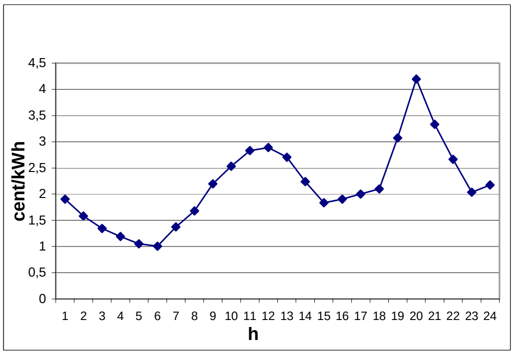

Figure 5: Typical winter electricity spot prices profile

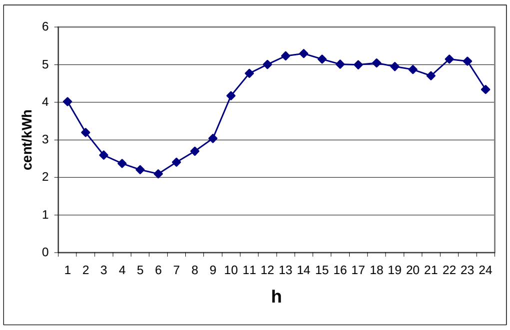

Figure 6: Typical summer electricity spot prices profile

3.2.3 Reasoning Base

In this module is made the computation of the electricity prices per week, for each test contract. We have the load profiles with hourly active energy consumption for each class in the database.

For each test contract we have the corresponding hourly electricity prices. In the knowledge base the inputs introduce by the user will give the consumer class and its load profile. With this information the reasoning base will compute the final electricity price per week to be paid by the consumer considering each different test contract. Further more, these final values are presented and compared to provide the best contract advice. In another interface the tailored contract is introduced in an hourly price base and its weekly cost is also computed by the reasoning base. With these results a final conclusion about the best contract can be made. This tailored contract function is separated to increase the flexibility of the DSS allowing easier changes on the tailored contract. Finally this reasoning base present graphical daily price profiles for the Best Contract as the ones presented in Figures 10 and 11.

3.2.4 Interface

An interactive and easy to use interface was developed to allow the user to interact with the DSS. In this interface the inputs are introduced, the calculations are selected and the results are presented. Figure 7 presents the interface with a simulation case as example.

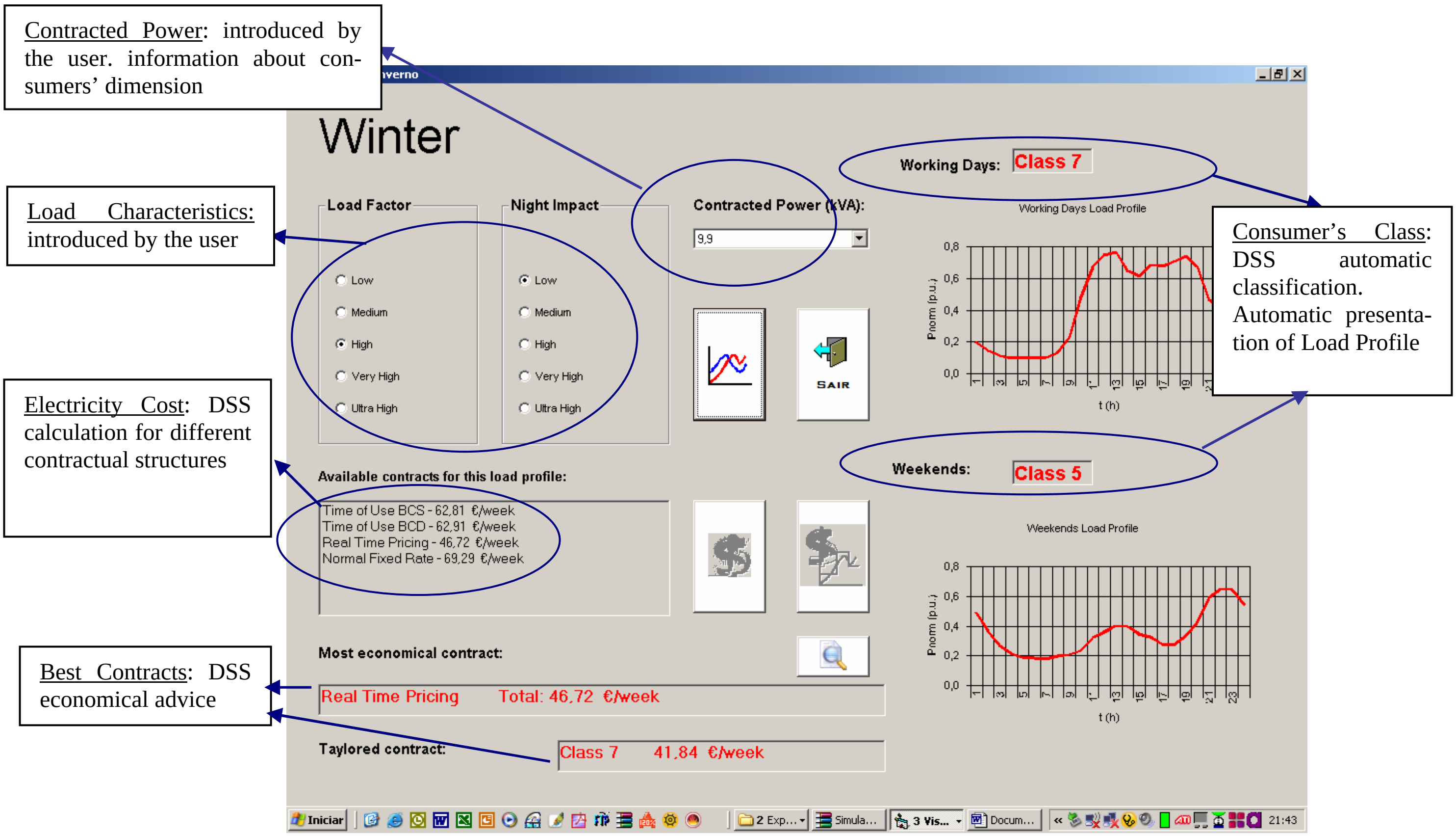

Figure 7: The DSS User Interface

4 SIMULATION RESULTS

The DSS was tested and validated running a large number of simulations for different possible clients and the results obtained were very satisfactory. The DSS is flexible enough to study different types of contracts and adjust them in a very simple way. The system is able to deal with large amounts of data and expandable to work with real time actualizations of the database.

As an example lets consider the results obtained by the DSS on the simulation presented in Figure 7:

Inputs- Consumer Characteristics: d1-High; d3-Low; CP-9.9 KVA
Outputs: Classification: Working days: Class 7
Weekends: Class 5
Best Contract: Tailored Contract (TC)

The graphical comparative Economical Analysis of different contracts is presented in Figure 8.

We can conclude that for consumers with High load factor (d1) and Low night impact (d3) the most economical contract is the proposed tailored contract (TC), presented in Figures 9 and 10, followed by the real-time-pricing (RTP).

The other existing contracts are more expensive, TOU-WC and TOU-WDC present the same value for this type of client, and FR is the most expensive.

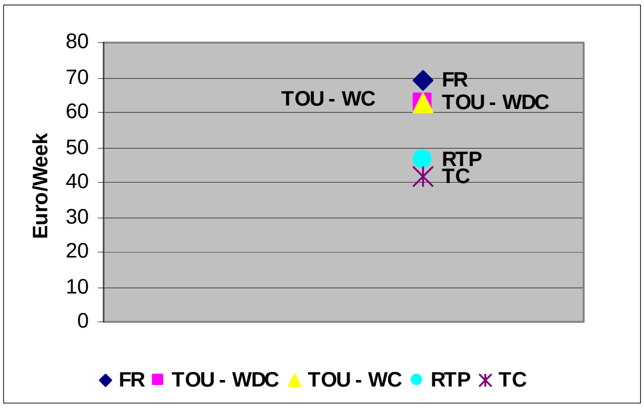

Figure 8: Comparative analysis of different contracts weekly cost

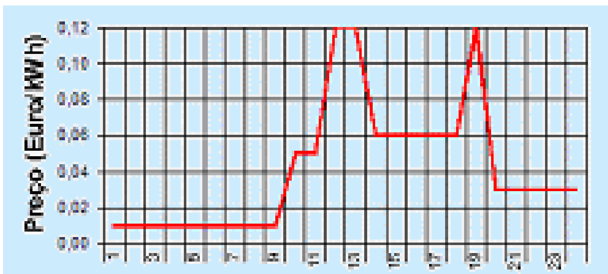

Figure 9: TC for working days

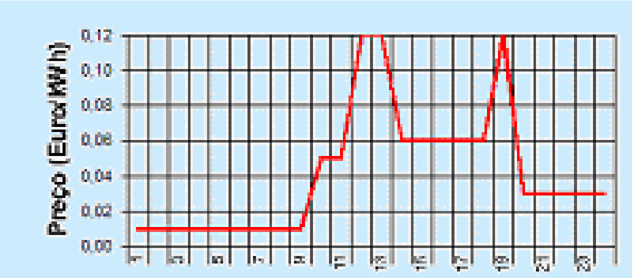

Figure 10: TC for weekends

From this example we can conclude that the electricity prices existing at the moment can be improved. There is a possibility to have lower prices, as presented in Figure 8. We can see that, as expected, FR is the most expensive and gives no signals to consumers about market price changes, so there is no incentive to increase demand elasticity. Significant advantages can be obtained with RTP and TC. This difference can be reduced increasing the retailer profit, so the final prices will be defined by the level of competition in the market.

This simulation has been repeated for a large number of different situations with different input factors, and RTP and TC are always the best option.

The results obtained are clear and as expected show the large potential of the use of tailored contracts and real time pricing. In the future, these will improve the economical efficiency of the market. The practical application of these new types of contracts depends on the replacement of metering systems, communications between consumers and utilities, and the development of data exchange and management tools.

5 CONCLUSIONS

A robust and flexible Decision Support Tool (DSS) for the study, test and selection of the most adequate retail electricity contract is presented. This DSS uses as inputs the most relevant load shape indexes and the contracted power of a consumer and provides as outputs the consumer load profile, an economical comparative analysis of different contracts and finally the decision about the most adequate contract. The knowledge base of the DSS is the result of a Consumer Characterization Framework based on Data Mining techniques developed to extract useful knowledge from large amounts of consumers' data. Its database is composed by the different classes' load profiles and the contracts that are being tested.

The DSS has been tested and validated with real data from the Portuguese distribution company. The results obtained after running a large number of simulations are very satisfactory. The DSS is able to perform an automatic classification of consumers from only three different inputs and perform an economical comparative analysis. It provides correct advice about the Best Contract option. From the simulations performed we are able to conclude that the tailored contracts and real time pricing are the most adequate contracts, so an investment on metering, communication and data management systems should be considered to make this kind of contracts possible in practical applications.

The diversification of pricing, specially the use of real time pricing will increase consumers' response due to price changes, and improve the overall efficiency of electricity markets and resource allocation. This DSS is useful both for retail companies and for electricity consumers. In the future we intend to provide the DSS as a Web based application to be consulted anywhere and by anyone using a simple Web browser. This will permit LV clients to simulate different scenarios.

REFERENCES

[1] Kirschen D., "Demand-Side View of Electricity Markets", IEEE Transactions on Power Systems, Vol. 18, NO2, pp 520-526, May 2003.

[2]Pitt, B., D. Kirschen., "Applications of Data Mining Techniques to Load Profiling", in Proc. IEEE PICA, Santa Clara, CA, May 1999.

[3]C.S., Chen, J.C., Hwang, C.W., Huang, "Application of Load Survey to Proper Tariff Design", IEEE Transactions on Power Systems, Vol. 12, No. 4, November 1997, pp 1746-1751.

[4]Chicco G., Napoli R., Postulache P., Scutariu M. & Toader C., "Customer Characterization Options for Improving the Tariff Offer", in IEEE Transactions on Power Systems, Vol, 18, nº 1, February, pp. 381-387, 2003.

[5]Figueiredo, V., Rodrigues, F., Vale, Z., Gouveia, J.B., "An Electric Energy Characterization Framework based on Data Mining Techniques", accepted for publication in the IEEE Transactions on Power Systems.

[6]Rodrigues F., Figueiredo V., Duarte F.J. & Vale Z., "A Comparative Analysis of Clustering Algorithms Applied to Load Profiling", in Lecture Notes in Artificial Intelligence (LNAI 2734), Springer-Verlag, pp.73-85, 2003.

[7]Figueiredo V., Duarte F.J., Rodrigues F., Vale Z., Gouveia, J. et al., "Electric Energy Customer Characterization by Clustering", in Proc of ISAP 2003 , Lemnos, Greece, September, 2003.

[8]Turban Efraim, Aranson Jay, "Decision Support Systems and Intelligent Systems", Prentice Hall, 1998.

[9]Portuguese Regulatory Authority (ERSE), "Tariffs and Prices for Electric Energy and other Services in 2004", Lisbon, December of 2003. available in : www.erse.pt

[10]Spanish Market Operator (OMEL), 2004, available in : www.omel.es

[11]Chiu T., Fang D. et al, "A Robust and Scalable Clustering Algorithm for Mixed Type Attributes in Large Database Environment" in Proc. of the 7th ACM SIGKDD Inter. Conf. on KDD and Data Mining, pp 263-268, USA, 2001.

[12]Apolinário et al, "Application of Additive tariffs to the electrical sector", 8th Luso-Espanhol Congress of Electrical Engineering, Vilamoura, Julho de 2003, Portugal.