# AN APPLICATION OF SOFT COMPUTING TECHNIQUES FOR A HOTEL DECISION SUPPORT SYSTEM

Francisco Javier Martínez-de-Pisón Ascacíbar

Rubén Escribano García

Julio Fernández Ceniceros

Roberto Fernández Martínez

*Grupo EDMANS (http://www.mineriadatos.com)*

*Universidad de La Rioja. España*

José Ignacio Pérez Moneo

*Angestur Consultores e Innovación S.L (HotelOptimizer) (http://hoteloptimizer.com/)*

## Abstract

In last years, due to the spreading of websites specialized on online hotel booking, many clients wait until the last moment to make the reservations. Thus, at the last moment, they can compare the room prices in different closely hotels, with almost the same services, in order to save money. As a consequence, for hotel managers, it is a challenge to maximize profits by continuously changing room prices to present them more attractive than the competition hotels. This paper shows the results of a research project which principal aim is the application of soft computing techniques to obtain models useful to predict future bookings for each calendar day, by means of the use of the hotel historical reservation data and the characteristics of each day (month, day of the week, season, festivities in closely regions or cities, the weather, etc.). Built models will be part of a hotel decision support software which is been developed for *HotelOptimizer* company with EDMANS group support.

***Keywords***: *hotel decision support systems; soft-computing; hotel booking systems; data mining*

## Resumen

En los últimos años, debido a la proliferación de portales Web especializados en reservas *On-Line* de habitaciones de hotel, muchos clientes esperan hasta el último momento para realizar sus reservas. Así, pueden ahorrarse dinero comparando los precios de última hora de hoteles cercanos que ofrecen prácticamente los mismos servicios. Como consecuencia de esto, para los gerentes de hotel es un reto poder maximizar los beneficios mediante el ajuste continuo de los precios de las habitaciones de forma que éstos sean más atractivos para los clientes frente a los de hoteles de la competencia. En este artículo, se muestran los resultados de un proyecto de investigación donde se han aplicado técnicas de *soft-computing* para desarrollar modelos capaces de predecir las futuras reservas de cada día del año para un hotel cualquiera a partir de la base de datos de históricos de reservas del mismo y las características de cada día (mes, día de la semana, temporada del año, fiestas en comunidades o ciudades cercanas, climatología, etc.). Los modelos desarrollados formarán parte de un software de apoyo a la toma de decisiones para la gestión de hoteles que está desarrollando la empresa *HotelOptimizer* en colaboración con el Grupo EDMANS.

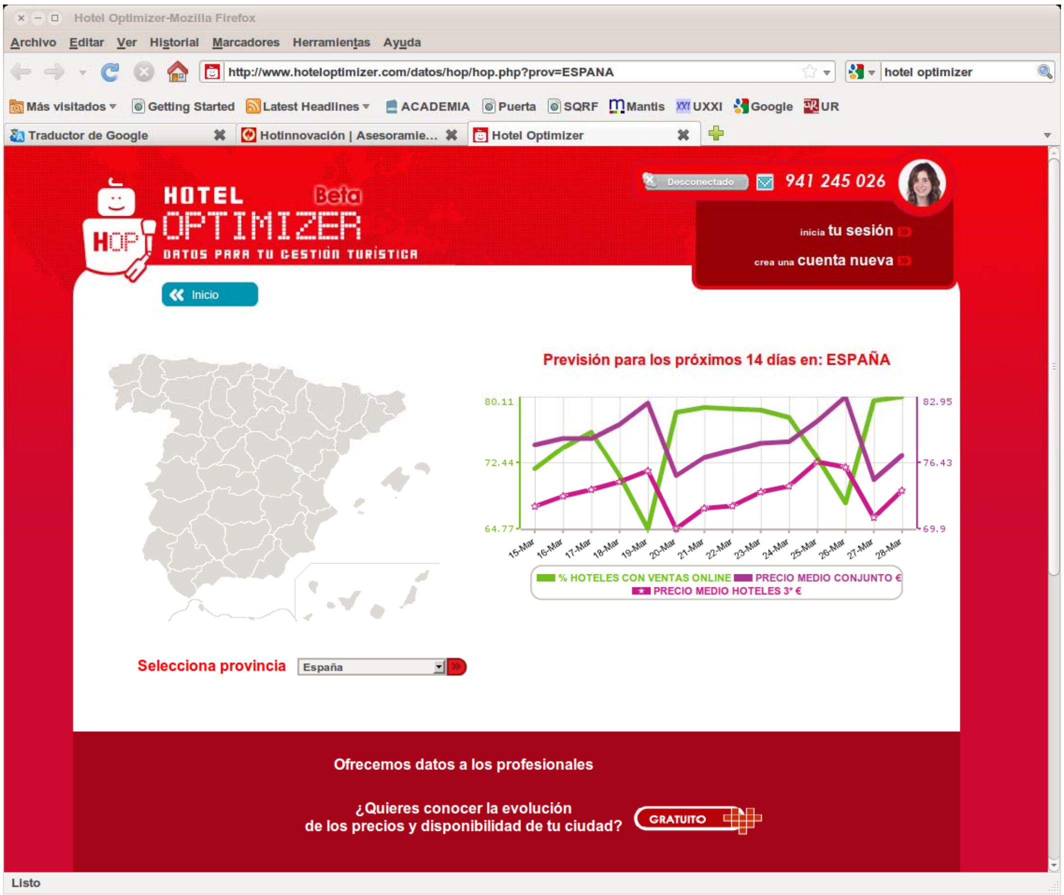
***Palabras clave:*** *sistemas de apoyo a la toma de decisiones en hoteles; computación flexible; reservas de hotel; minería de datos*

## 1. Introduction

Nowadays, due to the increase of Websites to online hotel booking, hotel managers need to change continuously the room prices in order to be competitive with other competition hotels. In this case, the aim consists to predict the precise room price to obtain maximum profit in each moment.

*HotelOptimizer* (http://www.hoteloptimizer.com) is a new company that is developing new decision support tools to help to the hotel managers. For example, they present a free Web-tool which permits to analyze a 14 days prediction of hotel prices in different Spanish regions (figure 1.).

**Figure 1: Web-tool where is shown a prediction of hotel prices for the next 14 days in Spain**


Hotel Optimizer-Mozilla Firefox
Archivo Editar Ver Historial Marcadores Herramientas Ayuda
http://www.hoteloptimizer.com/datos/hop/hop.php?prov=ESPANA
hotel optimizer
Más visitados Getting Started Latest Headlines ACADEMIA Puerta SQRF Mantis UXXI Google UR
Traductor de Google
Hotinnovación | Asesoramie...
Hotel Optimizer
Desconectado
941 245 026
HOTEL
Beta
OPTIMIZER
HOP
DATOS PARA TU GESTIÓN TURÍSTICA
inicia tu sesión
crea una cuenta nueva
Inicio
Previsión para los próximos 14 días en: ESPAÑA
80.11
72.44
64.77
82.95
76.43
69.9
15-Mar
16-Mar
17-Mar
18-Mar
19-Mar
20-Mar
21-Mar
22-Mar
23-Mar
24-Mar
25-Mar
26-Mar
27-Mar
28-Mar
% HOTELES CON VENTAS ONLINE
PRECIO MEDIO CONJUNTO €
PRECIO MEDIO HOTELES 3* €
Selecciona provincia
España
GRATUITO
Listo

Ofrecemos datos a los profesionales

¿Quieres conocer la evolución de los precios y disponibilidad de tu ciudad?

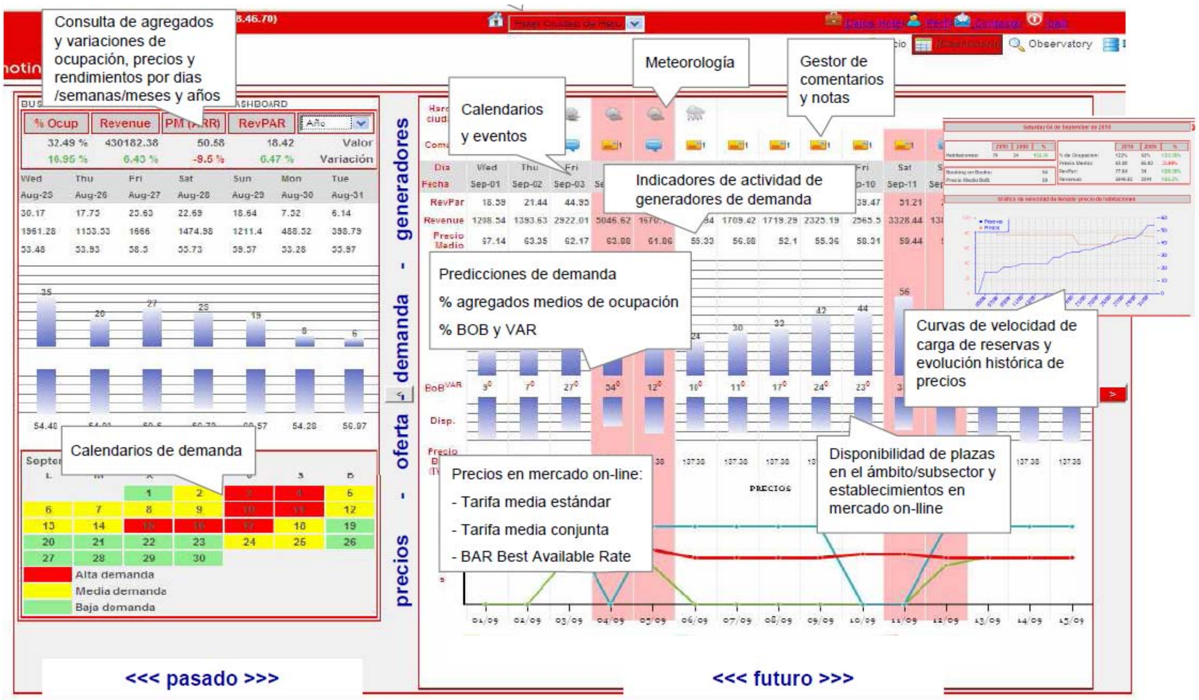
Also, they have a new commercial product, named "Hop Yield & Revenue Management Assistant (Hop YRMA)" (figure 2), which permits to:

1. Obtain business positioning in the market and management analysis occupation money: developments, trends, schedules and forecasts of demand for employment, reservations and price elasticity, price analysis and strategies of competitors, anticipating to the fluctuations in the availability of places in the area, planning of prices, availability, conditions and restrictions and updating of distribution channels.
2. Demand forecast: Hop YRMA has a demand forecasting system through complex formulas and identifying patterns using data mining techniques.

3. See details of the competition: Learn the strategies of competitors knowing their prices for the coming days and weeks with a single click (figure 3).
4. See demand schedules: Hop YRMA has a comprehensive database of events and issues that could to affect to the hotel.
5. Can be integrated with any Property Management Systems (PMS).
6. Have a Channel Manager connected to the hotel PMS and distribution channels allowing automatically link availability in all distributors and, even able to process reservations through the system connected with the hotel's own website.

One of the most important parts of this product is the capacity to forecast the future hotel demand. The basic idea was to use soft computing techniques to develop models useful to predict future room bookings for each calendar day, by means of the use of the hotel historical reservation data and the characteristics of each day (month, day of the week, season, festivities in closely regions or cities, weather, etc.).

**Figure 2: Hop Yield & Revenue Management Assistant (Hop YRMA)**


Consulta de agregados y variaciones de ocupación, precios y rendimientos por dias /semanas/meses y años
Meteorología
Gestor de comentarios y notas
Calendarios y eventos
Indicadores de actividad de generadores de demanda
Predicciones de demanda
% agregados medios de ocupación
% BOB y VAR
Curvas de velocidad de carga de reservas y evolución histórica de precios
Calendarios de demanda
Precios en mercado on-line:
- Tarifa media estándar
- Tarifa media conjunta
- BAR Best Available Rate
Disponibilidad de plazas en el ámbito/subsector y establecimientos en mercado on-lline
% Ocup
Revenue
PM (ARR)
RevPAR
Año
Valor
Variación
Alta demanda
Media demanda
Baja demanda
precios - oferta - demanda - generadores
Observatory
<<< pasado >>>
<<< futuro >>>

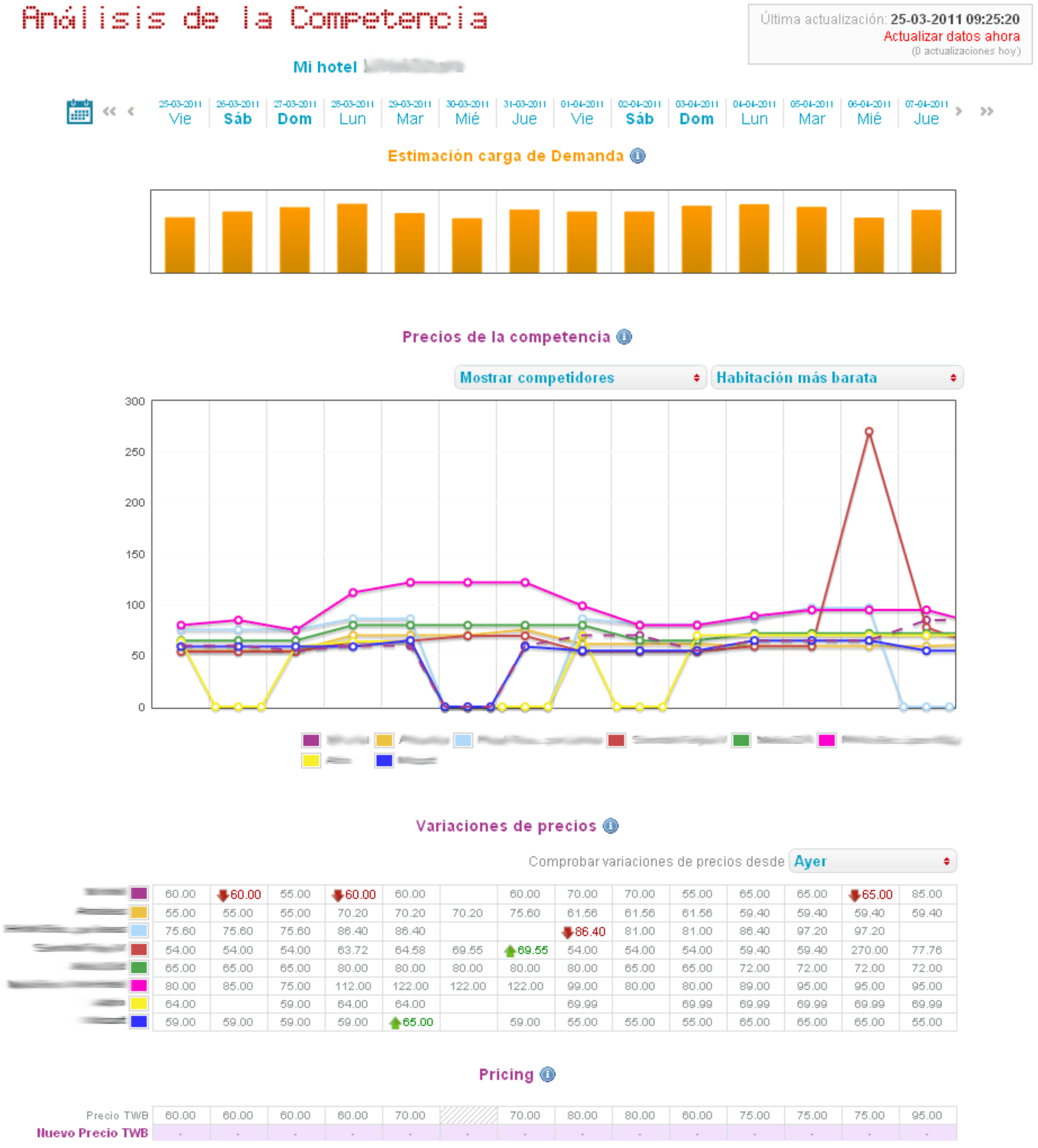
In this paper the results of the research project, named "CIERZO", are presented in which its principal objective was to build useful models to create demand calendars.

## 2. Developing Hotel Demand Calendars Using Soft Computing Techniques

### 2.1 Attributes Selection

The first step was to analyze visually many variables extracted from different sources: Macroeconomic and Social variables from the "*Instituto Nacional de Estadística (INE)*" databases, Meteorological databases from Websites, local and global festivities and other type of data included manually from different sites (figure 4).

**Figure 3: Example of analysis of competitors**



Análisis de la Competencia

Última actualización: **25-03-2011 09:25:20**
Actualizar datos ahora
(0 actualizaciones hoy)

Mi hotel

Variaciones de precios

Comprobar variaciones de precios desde Ayer

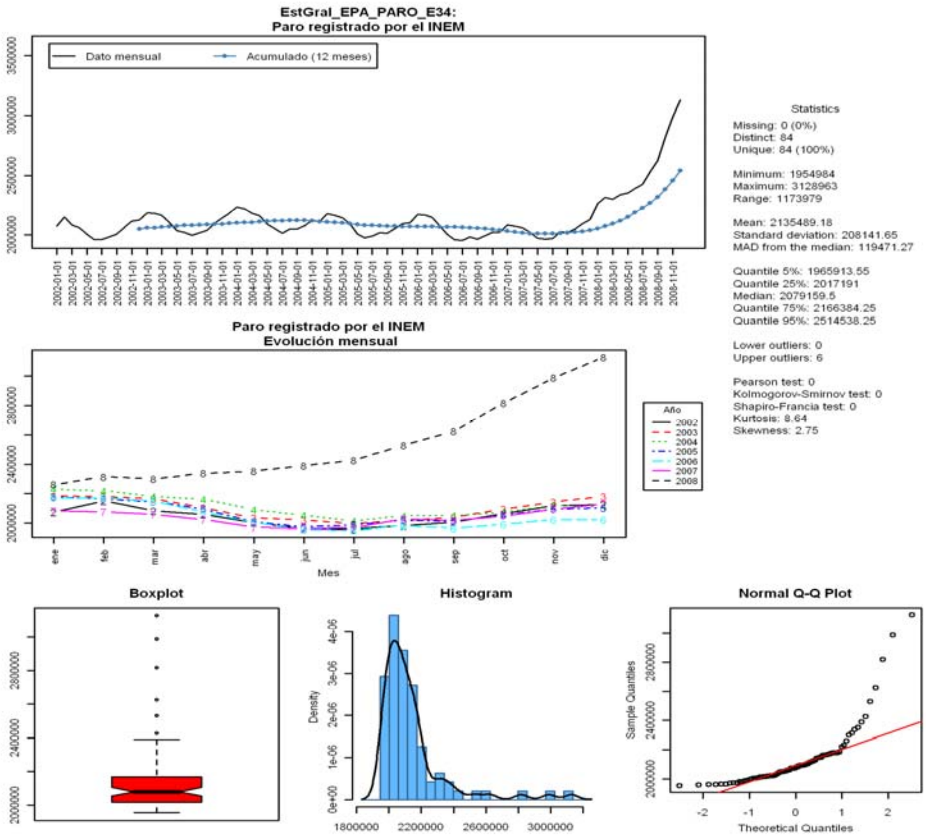
| | 25-03-2011 Vie | 26-03-2011 Sáb | 27-03-2011 Dom | 28-03-2011 Lun | 29-03-2011 Mar | 30-03-2011 Mié | 31-03-2011 Jue | 01-04-2011 Vie | 02-04-2011 Sáb | 03-04-2011 Dom | 04-04-2011 Lun | 05-04-2011 Mar | 06-04-2011 Mié | 07-04-2011 Jue |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 🟪 | 60.00 | ↓60.00 | 55.00 | ↓60.00 | 60.00 | | 60.00 | 70.00 | 70.00 | 55.00 | 65.00 | 65.00 | ↓65.00 | 85.00 |
| 🟧 | 55.00 | 55.00 | 55.00 | 70.20 | 70.20 | 70.20 | 75.60 | 61.56 | 61.56 | 61.56 | 59.40 | 59.40 | 59.40 | 59.40 |
| 🟦 | 75.60 | 75.60 | 75.60 | 86.40 | 86.40 | | | ↓86.40 | 81.00 | 81.00 | 86.40 | 97.20 | 97.20 | |
| 🟥 | 54.00 | 54.00 | 54.00 | 63.72 | 64.58 | 69.55 | ↑69.55 | 54.00 | 54.00 | 54.00 | 59.40 | 59.40 | 270.00 | 77.76 |
| 🟩 | 65.00 | 65.00 | 65.00 | 80.00 | 80.00 | 80.00 | 80.00 | 80.00 | 65.00 | 65.00 | 72.00 | 72.00 | 72.00 | 72.00 |
| 🟪 | 80.00 | 85.00 | 75.00 | 112.00 | 122.00 | 122.00 | 122.00 | 99.00 | 80.00 | 80.00 | 89.00 | 95.00 | 95.00 | 95.00 |
| 🟨 | 64.00 | | 59.00 | 64.00 | 64.00 | | | 69.99 | | 69.99 | 69.99 | 69.99 | 69.99 | 69.99 |
| 🟦 | 59.00 | 59.00 | 59.00 | 59.00 | ↑65.00 | | 59.00 | 55.00 | 55.00 | 55.00 | 65.00 | 65.00 | 65.00 | 55.00 |

Pricing

| | 25-03-2011 Vie | 26-03-2011 Sáb | 27-03-2011 Dom | 28-03-2011 Lun | 29-03-2011 Mar | 30-03-2011 Mié | 31-03-2011 Jue | 01-04-2011 Vie | 02-04-2011 Sáb | 03-04-2011 Dom | 04-04-2011 Lun | 05-04-2011 Mar | 06-04-2011 Mié | 07-04-2011 Jue |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Precio TWB | 60.00 | 60.00 | 60.00 | 60.00 | 70.00 | ▨ | 70.00 | 80.00 | 80.00 | 60.00 | 75.00 | 75.00 | 75.00 | 95.00 |
| **Nuevo Precio TWB** | - | - | - | - | - | - | - | - | - | - | - | - | - | - |

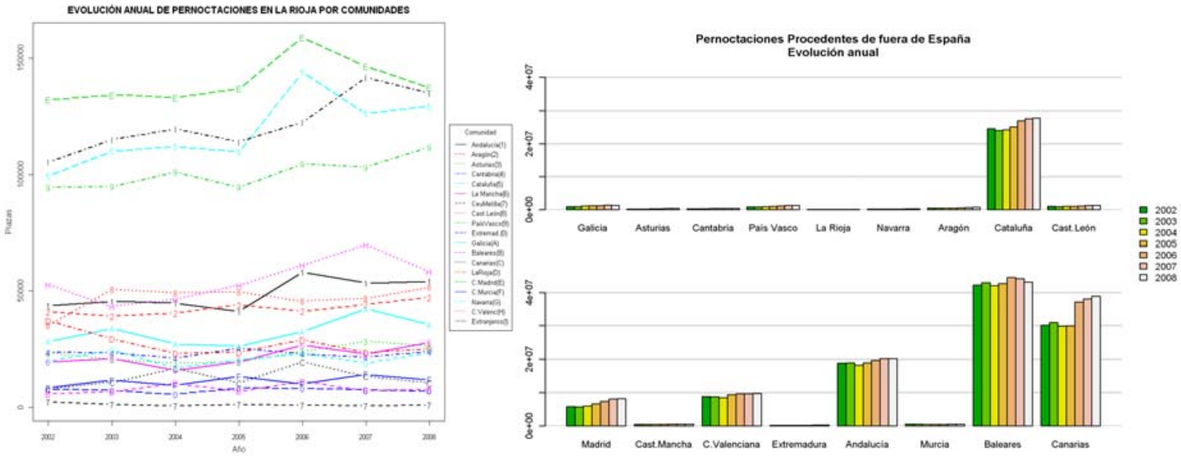
Also, evolution of hotel room prices and others hotels indicators were studied in different Spanish regions (figure 5). Finally, 119 of the most interesting attributes were selected.

**Figure 4: Example of unemployment evolution in Spain**



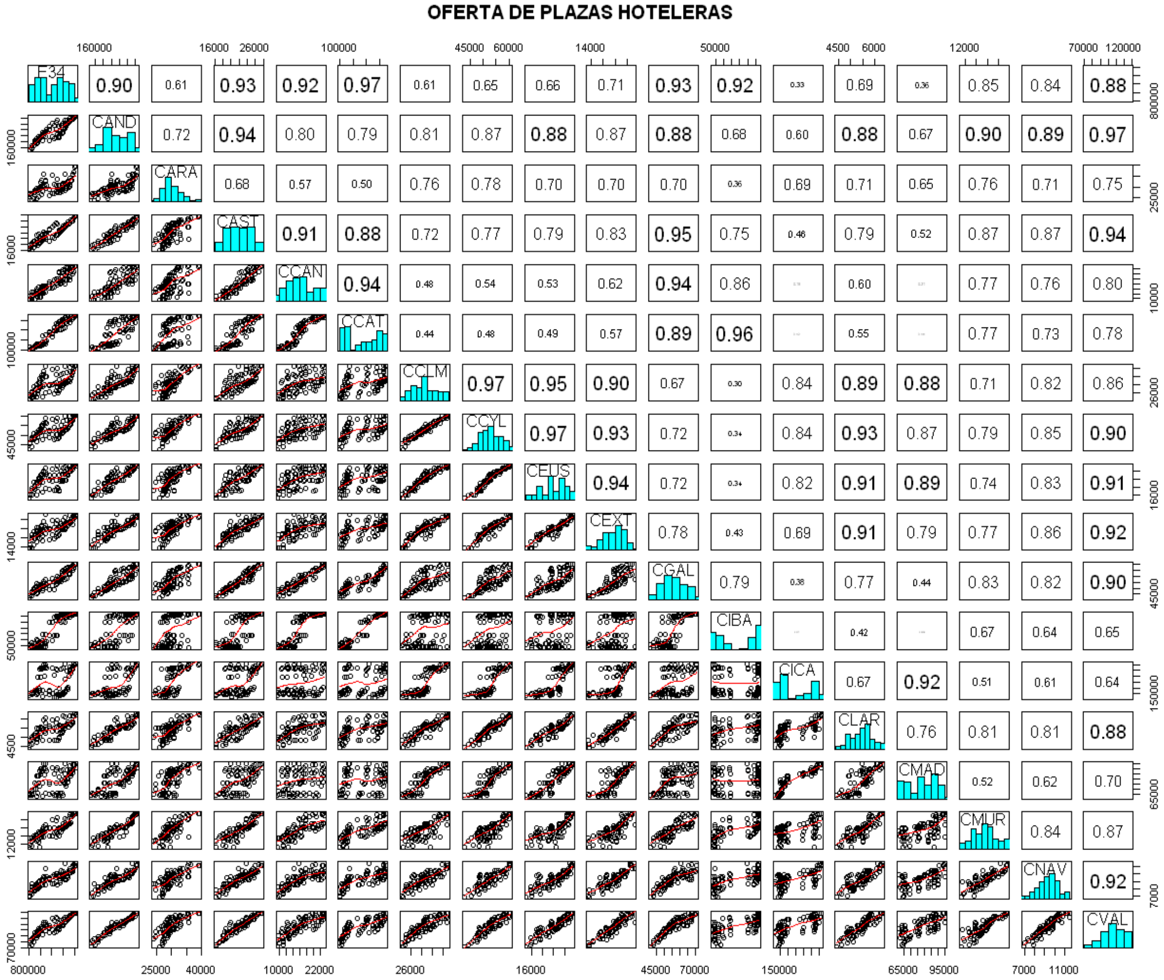
**Note: It is interesting to observe the strong increment from the end of 2007 when government was saying that it had not been crisis.**

**Figure 5: Evolution of overnights in La Rioja of people from different Spanish Regions (left) and Evolution of overnights in each Spanish region of people from different countries**



After this, advanced scatter plots (figure 6) were used to identify correlation between Macroeconomic indexes, meteorological data, local and global festivities, etc. in front of hotel overnights attributes and from different Spanish regions. At the end of this step, the list of the most important variables was reduced significantly.

**Figure 6: Example of scatter plot of room vacancy in different Spanish Regions**



## 2.2 Calendar Forecast for one specific Hotel with Data Mining Techniques

After the selection of the most important variables (Economical, Social, Festive, Meteorological and Hotel Indicators), the next aim was to search a methodology for develop models useful in order to create calendar forecast for a specific hotel.

The basic idea was to use the historical reservation data from each hotel to create specific models to predict the future bookings for each future calendar day. The design of the models was structured in 22 input attributes that define the characteristics of each day (month, day of the week, season, festivities in closely regions or cities, festivities in Spain, the weather, etc.) and one output variable with the number of bookings for the day. Due to confidentiality is not possible to show the name and type of each selected attribute.

First, models were created using data of the years 2007 to 2009. Data from before 2007 were not considered because of the different Spanish economic situation.

In order to find models that generate a low prediction error, a battery of Data Mining algorithms were used:

- M5P algorithm (M5P): Implements base routines for generating M5Model trees. A decision list for regression problems is generated using separate-and-conquer. In each iteration, it builds a model tree using M5 and makes the "best" leaf into a rule.

Quinlan's M5P can learn such piece-wise linear models. M5P also generates a decision tree that indicates when to use which linear model (Quinlan, 1992).

- Multilayer Perceptron (MLP): A classifier and predictor that uses backpropagation to classify instances. All nodes in this network are sigmoid, except when the class is numeric. In the latter case, the output nodes become unthresholded linear units (Haykin, 1999). Training is performed with networks that have between 1 and 40 neurons in the hidden layer.
- Linear Regression (LINREG): A class for using linear regression for prediction. It uses the Akaike criterion for variable selection and is able to deal with weighted instances (Wilkinson and Rogers, 1973).
- Simple Linear Regression (SIMPLR): Uses only the best attribute to obtain the model. It is useful for comparing with other algorithms.
- LeastMedSq (LMSQ): Implements a least median squared linear regression to make predictions. Least squared regression functions are generated from random sub-samples of the data. The least squared regression that has the lowest median squared error is chosen as the final model (Portnoy and Koenke, 1997).
- IBk (IBk): A version of the k-nearest neighbour algorithm. K is the number of neighbours to be used. It also permits the use of distance weighting. As it is a lazy algorithm, there is no training time (Aha and Kibler, 1991).

To obtain the best precision, ten models of each type of algorithm configuration were trained with 70% of the data from the training database and the remaining data (30%) were used to validate each model. By generating 10 models of each algorithm configuration, the influence of local minima was reduced and much more realistic errors were obtained.

The purpose of this work was to determine the algorithm configuration that provide the best rooms booking prediction or, in other words, the algorithm configuration that yields the lowest Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE) for other different days not used for model construction. These errors were:

$$RMSE = \sqrt{\frac{1}{n}\sum_{k=1}^{n}(y(k)-\hat{y}(k))^2} \qquad (1)$$

and

$$MAE = \frac{1}{n}\sum_{k=1}^{n}\left|y(k)-\hat{y}(k)\right| \qquad (2)$$

where $y$ and $\hat{y}$ were, respectively, the measured and predicted rooms booking and $n$ was the number of examples in the database used to validate the models.

The results of the training and validation process are shown in table 1. This table provides a summary of the validation errors arranged by the RMSE corresponding to ten trained models for the algorithm configurations. This table presents the mean (MEAN) and standard deviation (SD) of RMSE and MAE validation of ten models of each type of algorithm configuration.

As it is observed, the best algorithm was a Multilayer Perceptron with 20 neurons in the hidden layer but the errors were not acceptable (RMSEMEAN over 40%).

**Table 1: Results of the Training and Validation Process**

| Algorithm | RMSEMEAN | RMSESD | MAEMEAN | MAESD | TIME |
|---|---|---|---|---|---|
| MLP (Neurons=20) | 0.404 | 0.004 | 0.311 | 0.008 | 8628.589 |
| MLP (Neurons=15) | 0.405 | 0.003 | 0.313 | 0.006 | 1804.589 |
| MLP (Neurons=10) | 0.405 | 0.002 | 0.316 | 0.010 | 1951.41 |
| IBk (K=10) | 0.405 | 0.004 | 0.312 | 0.003 | 0.02 |
| LMSQ | 0.405 | 0.003 | 0.329 | 0.002 | 22.165 |
| MLP (Neurons=5) | 0.405 | 0.004 | 0.314 | 0.010 | 899.565 |
| MLP (Neurons=40) | 0.406 | 0.004 | 0.311 | 0.012 | 13739.393 |
| MLP (Neurons=30) | 0.406 | 0.003 | 0.312 | 0.011 | 8228.685 |
| MLP (Neurons=3) | 0.406 | 0.003 | 0.322 | 0.011 | 715.824 |
| MLP (Neurons=7) | 0.407 | 0.002 | 0.312 | 0.008 | 1009.002 |
| M5P | 0.407 | 0.005 | 0.316 | 0.004 | 4.122 |
| IBk (K=6) | 0.409 | 0.004 | 0.308 | 0.002 | 0.017 |
| IBk (K=5) | 0.412 | 0.003 | 0.308 | 0.002 | 0.018 |
| LINREG | 0.425 | 0.004 | 0.352 | 0.009 | 2709.669 |
| IBk (K=3) | 0.430 | 0.004 | 0.308 | 0.003 | 0.017 |
| SIMPLR | 0.443 | 0.004 | 0.307 | 0.003 | 0.02 |

## 2.3 Designing the Final Algorithm using Soft Computing Techniques

From the conclusions raised in the previous works, a new model was contemplated using soft computing techniques.

Soft computing consists in the use of computational techniques and intelligent systems in order to solve inexact and complex problems (Sedano et al., 2010). The involved computational techniques are stochastic and therefore suited to investigate real-world problems (Banerjee et al., 2010; Corchado and Herrero, 2011).

Evolutionary algorithms (EA), which comprise a fundamental component of soft computing, are said to be inspired by nature. One of the most representative techniques of soft computing is the genetic algorithms (GA): a large number of systematic methods used to solve optimization problems applying the principles of biological evolution, namely survival of the ‘fittest’, sexual reproduction and mutation (Mitchell, 1998). From this conception of GA, it has been possible to solve real-world problems which were impossible to tackle previously with the classic techniques.

The new algorithm was created with the following phases: selection of the most important variables, identification of similar days and average values obtained with robust estimation of reserves. In this case, we sought to optimize the parameters of the proposed model using genetic algorithms, with the idea of getting a model as accurate as possible.

The proposed model was based on an instance-based learning algorithm where the day’s Euclidean distance depended on: if there are parties or in nearby cities, If there are parties or not in nearby regions, the month of the year, the week of the month, the day of the week, the weather, etc.

The final algorithm was formed for these steps:

- Each attribute is normalized between 0 and 1 and multiplied with a weighting coefficient.
- Euclidean distance matrix is created with the computation of the Euclidean distance for each day with the others.
- K-nearest days are chosen for each date from the Euclidean distance matrix.

In order to obtain robust values, the mean room booking is calculated using the mean of the K-values that falls into the range limited by PERCENTIL and 1-PERCENTIL.

### 2.4 Searching the Best Algorithm's Parameters with Genetic Algorithms

Weighting coefficients, K, PERCENTIL and the most important attributes used to calculate the day's Euclidean distance were optimized with genetic algorithms.

First of all, 100 individual algorithm's parameters were initialised with random values.

For each individual the training data was randomly selected from the 70% of the database. It was used for calculates day's Euclidean distance matrix. Then, for each day, the most K-closed days were selected using this matrix. The other 30% of the data (test data) was used to calculate the correlation between the predictions of the room reservations and the actuals. In each case, correlation was computed ten times with different test and training data. The final objective function, to be minimised, was one minus the mean of the ten correlations.

Those of the many individuals in generation 0 which have the lowest objective function were selected and used as the basis for obtaining the next generation (generation 1) by means of crosses and mutations.

The new generation was made up as follows:

- 25% comprises the best individuals from the previous generation (parents of the new generation);
- 60% comprises individuals obtained by crossovers from selected parents. The crossover process involves changing various digits in the chromosomes of the variables to be modified. These chromosomes were made up of the digits for the variables with the decimal points removed, joined together in a single set.
- The remaining 15% was obtained by mutation, through the creation at random of chromosomes within the ranges established. The aim was to find new solutions in areas not previously explored.

This process was repeated over several generations until the objective function of the best individual was observed to remain the same or not to drop significantly from one generation to the next. If the error was suitable, the best individual was selected as the final solution.

## 3. Results

After 300 generations of individuals and 50 days of calculating, the maximum correlation achieved for the best model was 1-0.27142 = 0.72858. In figure 7 is possible to observe the evolution of the 100 individuals of the generations 1, 100, 200 and the final generation 300.

**Figure 7: Generations: 1, 100, 200 and 300 (Final)**

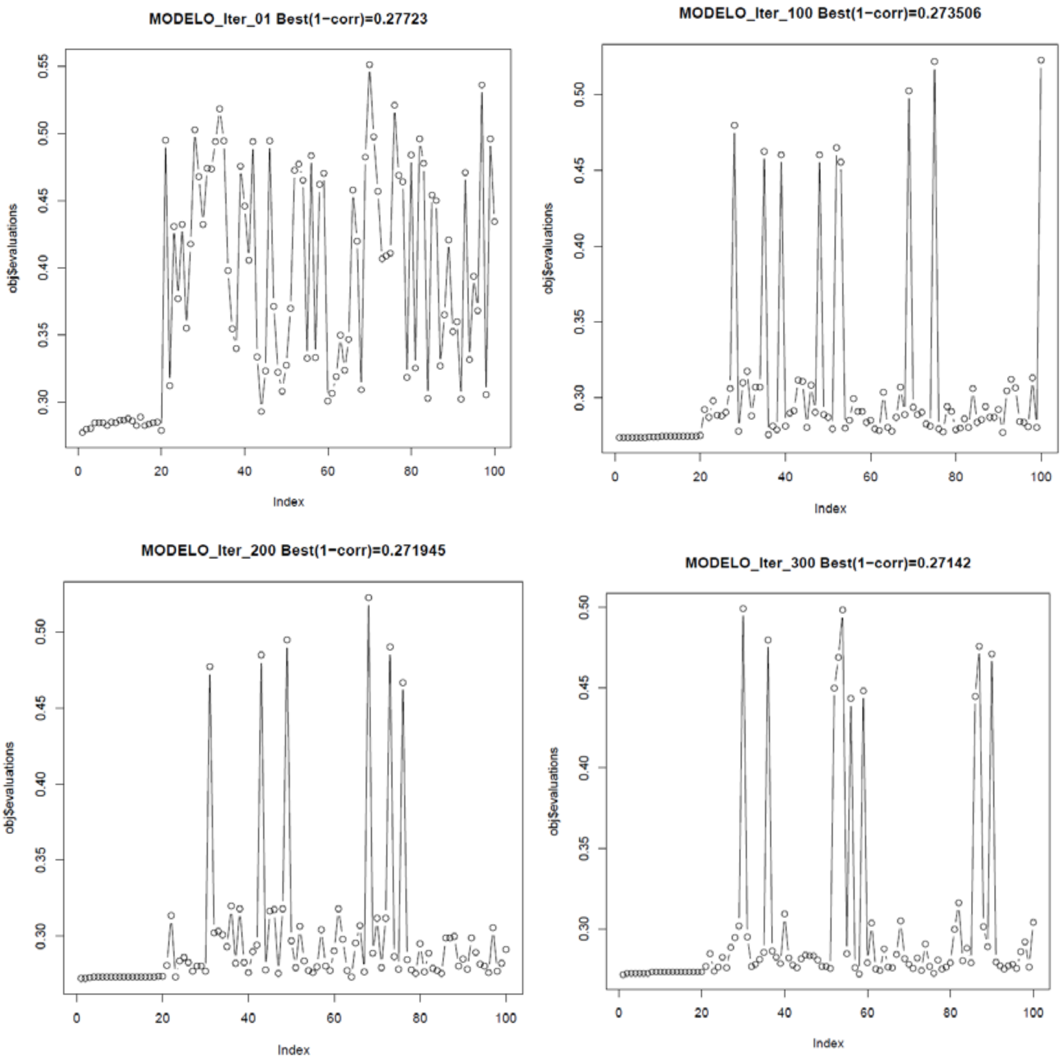
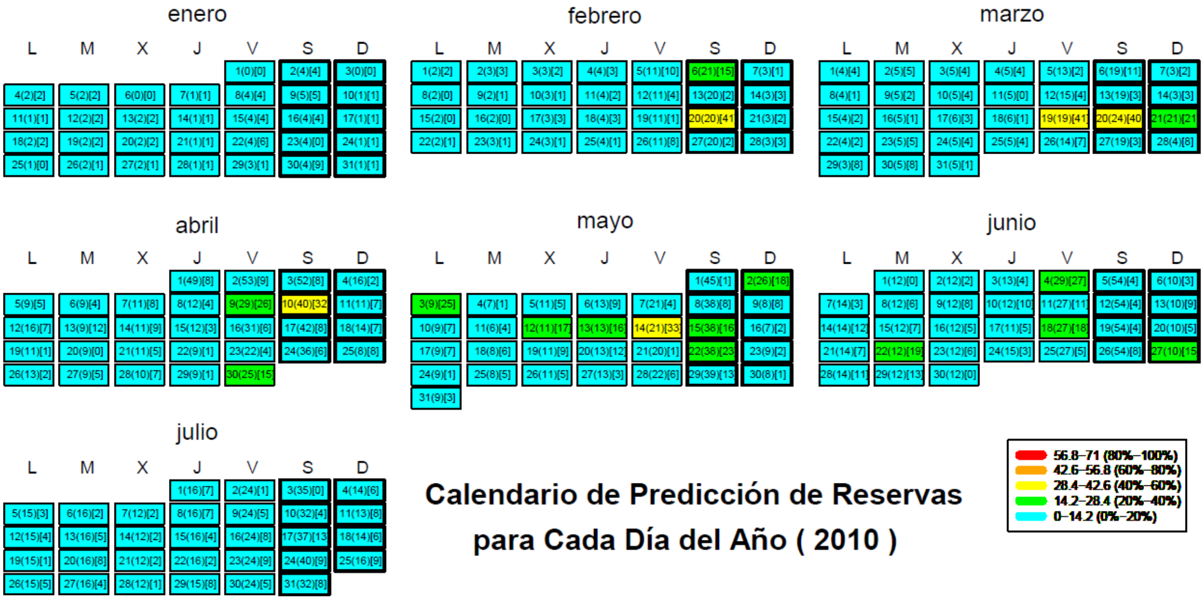


Figure 8 presents the predictions of the best model for the first seven months of the year 2010 contrasting to the real room bookings. Calendar shows the prediction for each day on this way:

- The number on the left is the month day.
- The number in brackets shows the prediction of reservations for that day as estimated by the model using the database for the years 2007 to 2009.
- The number between brackets indicates the absolute error or absolute difference of the value that predicts the model and the actual bookings that were made for that day.
- In colors is showed the error respect the maximum value. Box is blue if the error is below 20%, green if error is between 20% and 40%, and so on like is presented in the box on the right.

As can be seen, most of the days of the calendar had obtained less than 20% error. Fourteen (14) of them had an error between 20% and 40% (green), and five (5) yellow errors (range between 40% and 60%).

**Figure 8: Calendar with the room reservation predictions for the first seven months for 2010**



enero

| L | M | X | J | V | S | D |
|---|---|---|---|---|---|---|
| | | | | 1(0)[0] | 2(4)[4] | 3(0)[0] |
| 4(2)[2] | 5(2)[2] | 6(0)[0] | 7(1)[1] | 8(4)[4] | 9(5)[5] | 10(1)[1] |
| 11(1)[1] | 12(2)[2] | 13(2)[2] | 14(1)[1] | 15(4)[4] | 16(4)[4] | 17(1)[1] |
| 18(2)[2] | 19(2)[2] | 20(2)[2] | 21(1)[1] | 22(4)[6] | 23(4)[0] | 24(1)[1] |
| 25(1)[0] | 26(2)[1] | 27(2)[1] | 28(1)[1] | 29(3)[1] | 30(4)[9] | 31(1)[1] |

febrero

| L | M | X | J | V | S | D |
|---|---|---|---|---|---|---|
| 1(2)[2] | 2(3)[3] | 3(3)[2] | 4(4)[3] | 5(11)[10] | 6(21)[15] | 7(3)[1] |
| 8(2)[0] | 9(2)[1] | 10(3)[1] | 11(4)[2] | 12(11)[4] | 13(20)[2] | 14(3)[3] |
| 15(2)[0] | 16(2)[0] | 17(3)[3] | 18(4)[3] | 19(11)[1] | 20(20)[41 | 21(3)[2] |
| 22(2)[1] | 23(3)[1] | 24(3)[1] | 25(4)[1] | 26(11)[8] | 27(20)[2] | 28(3)[3] |

marzo

| L | M | X | J | V | S | D |
|---|---|---|---|---|---|---|
| 1(4)[4] | 2(5)[5] | 3(5)[4] | 4(5)[4] | 5(13)[2] | 6(19)[11] | 7(3)[2] |
| 8(4)[1] | 9(5)[2] | 10(5)[4] | 11(5)[0] | 12(15)[4] | 13(19)[3] | 14(3)[3] |
| 15(4)[2] | 16(5)[1] | 17(6)[3] | 18(6)[1] | 19(19)[41 | 20(24)[40 | 21(21)[21 |
| 22(4)[2] | 23(5)[5] | 24(5)[4] | 25(5)[4] | 26(14)[7] | 27(19)[3] | 28(4)[8] |
| 29(3)[8] | 30(5)[8] | 31(5)[1] | | | | |

abril

| L | M | X | J | V | S | D |
|---|---|---|---|---|---|---|
| | | | 1(49)[8] | 2(53)[9] | 3(52)[8] | 4(16)[2] |
| 5(9)[5] | 6(9)[4] | 7(11)[8] | 8(12)[4] | 9(29)[26] | 10(40)[32 | 11(11)[7] |
| 12(16)[7] | 13(9)[12] | 14(11)[9] | 15(12)[3] | 16(31)[6] | 17(42)[8] | 18(14)[7] |
| 19(11)[1] | 20(9)[0] | 21(11)[5] | 22(9)[1] | 23(22)[4] | 24(36)[6] | 25(8)[8] |
| 26(13)[2] | 27(9)[5] | 28(10)[7] | 29(9)[1] | 30(25)[15] | | |

mayo

| L | M | X | J | V | S | D |
|---|---|---|---|---|---|---|
| | | | | | 1(45)[1] | 2(26)[18] |
| 3(9)[25] | 4(7)[1] | 5(11)[5] | 6(13)[9] | 7(21)[4] | 8(38)[8] | 9(8)[8] |
| 10(9)[7] | 11(6)[4] | 12(11)[17] | 13(13)[16] | 14(21)[33] | 15(38)[16 | 16(7)[2] |
| 17(9)[7] | 18(8)[6] | 19(11)[9] | 20(13)[12] | 21(20)[1] | 22(38)[23 | 23(9)[2] |
| 24(9)[1] | 25(8)[5] | 26(11)[5] | 27(13)[3] | 28(22)[6] | 29(39)[13 | 30(8)[1] |
| 31(9)[3] | | | | | | |

junio

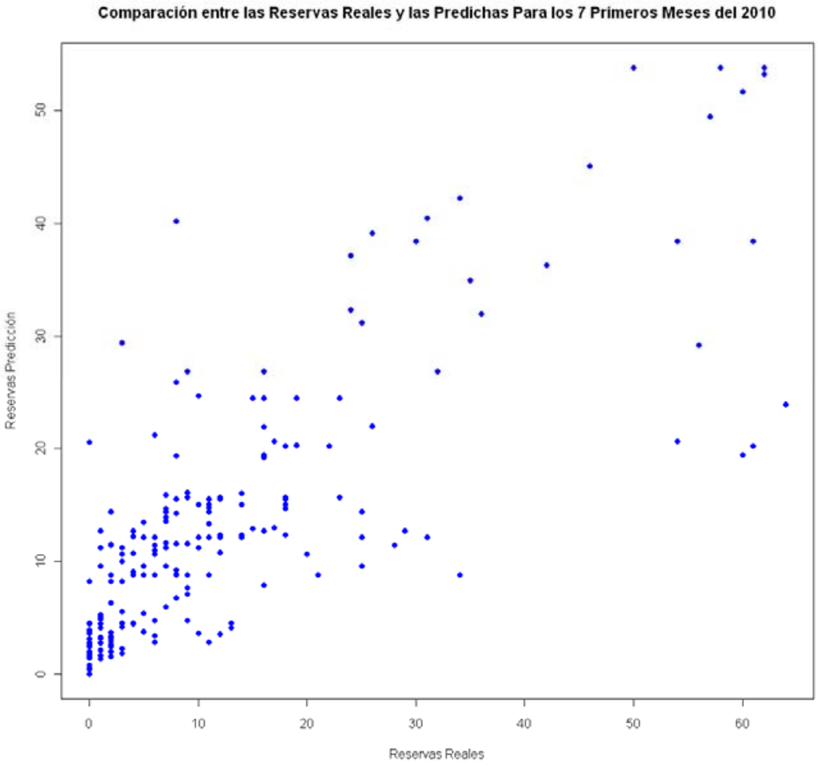
| L | M | X | J | V | S | D |
|---|---|---|---|---|---|---|
| | 1(12)[0] | 2(12)[2] | 3(13)[4] | 4(29)[27] | 5(54)[4] | 6(10)[3] |
| 7(14)[3] | 8(12)[6] | 9(12)[8] | 10(12)[10] | 11(27)[11] | 12(54)[4] | 13(10)[9] |
| 14(14)[12] | 15(12)[7] | 16(12)[5] | 17(11)[5] | 18(27)[18] | 19(54)[4] | 20(10)[5] |
| 21(14)[7] | 22(12)[19] | 23(12)[6] | 24(15)[3] | 25(27)[5] | 26(54)[8] | 27(10)[15 |
| 28(14)[11] | 29(12)[13] | 30(12)[0] | | | | |

julio

| L | M | X | J | V | S | D |
|---|---|---|---|---|---|---|
| | | | 1(16)[7] | 2(24)[1] | 3(35)[0] | 4(14)[6] |
| 5(15)[3] | 6(16)[2] | 7(12)[2] | 8(16)[7] | 9(24)[5] | 10(32)[4] | 11(13)[8] |
| 12(15)[4] | 13(16)[5] | 14(12)[2] | 15(16)[4] | 16(24)[8] | 17(37)[13 | 18(14)[6] |
| 19(15)[1] | 20(16)[8] | 21(12)[2] | 22(16)[2] | 23(24)[9] | 24(40)[9] | 25(16)[9] |
| 26(15)[5] | 27(16)[4] | 28(12)[1] | 29(15)[8] | 30(24)[5] | 31(32)[8] | |

Figure 9 shows an XY plot which represents the actual reserves of these months compared with that predicted values. In this case, the points tend to approach the main diagonal of a correlation graph obtained corroborating close to 0.80 (80%) that can be considered as excellent.

**Figure 9: Predictions vs. actual room bookings of the best model**



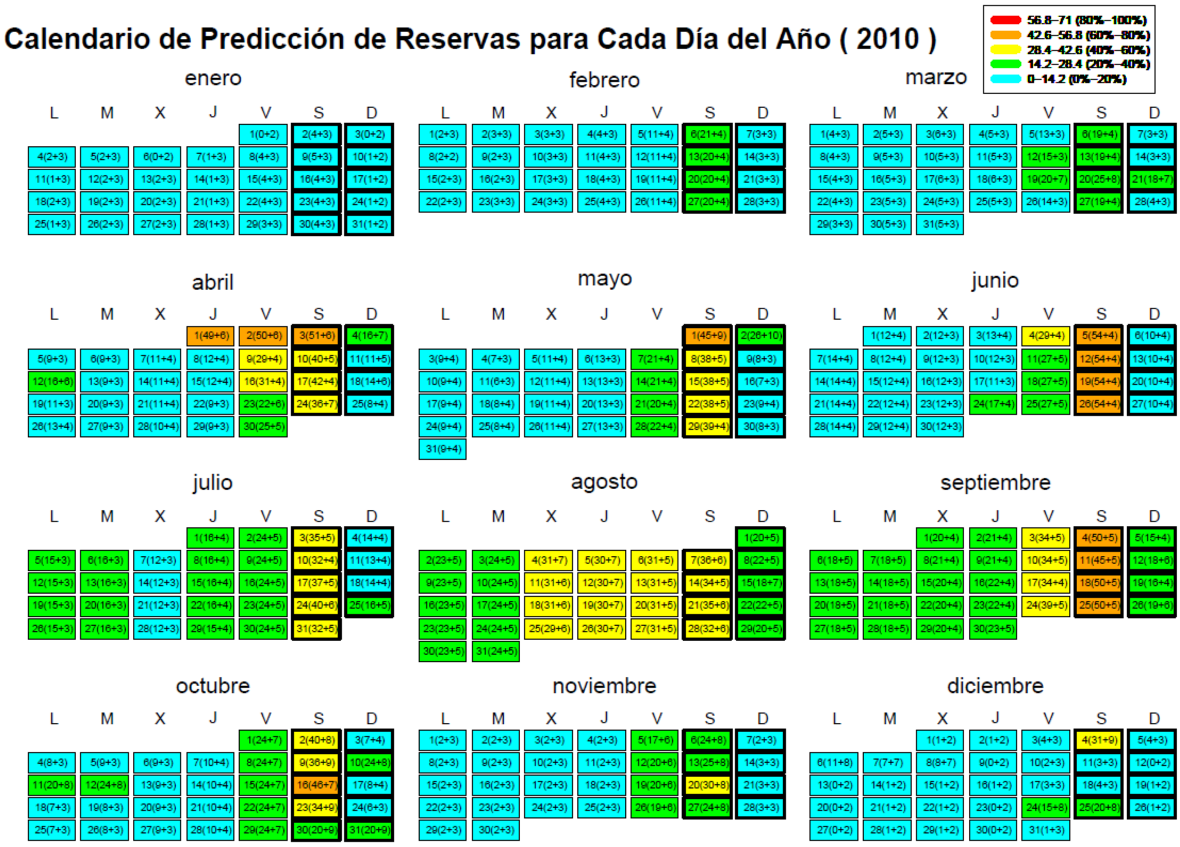
With the fitted model is possible to define a booking calendar like in the figure 10. This case is a Forecast Calendar for 2010 using data from years 2007 to 2009.

In brackets is shown the reservations prediction for each day (left) + the interval confidence at 95% of level (right). That is, a value 20+4 could be interpreted as: room bookings prediction for that day is 20±4, or whatever is the same, 95% of the days used to calculate the average value 20 have fallen into the interval [20-4, 20+4] = [16.24].

In colors, is shown the class of the day depending on the final reservations value predicted (cyan=very low, green=low, yellow=medium, orange=high, red=very high).

In this case, the maximum interval did not exceed the value of 10, but that does not mean they could be higher in the prediction errors due to abnormal days.

**Figure 10: Final forecast calendar to 2010**



Calendario de Predicción de Reservas para Cada Día del Año ( 2010 )

- 🟥 56.8–71 (80%–100%)
- 🟧 42.6–56.8 (60%–80%)
- 🟨 28.4–42.6 (40%–60%)
- 🟩 14.2–28.4 (20%–40%)
- 🟦 0–14.2 (0%–20%)

enero

| L | M | X | J | V | S | D |
|---|---|---|---|---|---|---|
| | | | | 1(0+2) | 2(4+3) | 3(0+2) |
| 4(2+3) | 5(2+3) | 6(0+2) | 7(1+3) | 8(4+3) | 9(5+3) | 10(1+2) |
| 11(1+3) | 12(2+3) | 13(2+3) | 14(1+3) | 15(4+3) | 16(4+3) | 17(1+2) |
| 18(2+3) | 19(2+3) | 20(2+3) | 21(1+3) | 22(4+3) | 23(4+3) | 24(1+2) |
| 25(1+3) | 26(2+3) | 27(2+3) | 28(1+3) | 29(3+3) | 30(4+3) | 31(1+2) |

febrero

| L | M | X | J | V | S | D |
|---|---|---|---|---|---|---|
| 1(2+3) | 2(3+3) | 3(3+3) | 4(4+3) | 5(11+4) | 6(21+4) | 7(3+3) |
| 8(2+2) | 9(2+3) | 10(3+3) | 11(4+3) | 12(11+4) | 13(20+4) | 14(3+3) |
| 15(2+3) | 16(2+3) | 17(3+3) | 18(4+3) | 19(11+4) | 20(20+4) | 21(3+3) |
| 22(2+3) | 23(3+3) | 24(3+3) | 25(4+3) | 26(11+4) | 27(20+4) | 28(3+3) |

marzo

| L | M | X | J | V | S | D |
|---|---|---|---|---|---|---|
| 1(4+3) | 2(5+3) | 3(6+3) | 4(5+3) | 5(13+3) | 6(19+4) | 7(3+3) |
| 8(4+3) | 9(5+3) | 10(5+3) | 11(5+3) | 12(15+3) | 13(19+4) | 14(3+3) |
| 15(4+3) | 16(5+3) | 17(6+3) | 18(6+3) | 19(20+7) | 20(25+8) | 21(18+7) |
| 22(4+3) | 23(5+3) | 24(5+3) | 25(5+3) | 26(14+3) | 27(19+4) | 28(4+3) |
| 29(3+3) | 30(5+3) | 31(5+3) | | | | |

abril

| L | M | X | J | V | S | D |
|---|---|---|---|---|---|---|
| | | | 1(49+6) | 2(50+6) | 3(51+6) | 4(16+7) |
| 5(9+3) | 6(9+3) | 7(11+4) | 8(12+4) | 9(29+4) | 10(40+5) | 11(11+5) |
| 12(16+6) | 13(9+3) | 14(11+4) | 15(12+4) | 16(31+4) | 17(42+4) | 18(14+6) |
| 19(11+3) | 20(9+3) | 21(11+4) | 22(9+3) | 23(22+6) | 24(36+7) | 25(8+4) |
| 26(13+4) | 27(9+3) | 28(10+4) | 29(9+3) | 30(25+5) | | |

mayo

| L | M | X | J | V | S | D |
|---|---|---|---|---|---|---|
| | | | | | 1(45+9) | 2(26+10) |
| 3(9+4) | 4(7+3) | 5(11+4) | 6(13+3) | 7(21+4) | 8(38+5) | 9(8+3) |
| 10(9+4) | 11(6+3) | 12(11+4) | 13(13+3) | 14(21+4) | 15(38+5) | 16(7+3) |
| 17(9+4) | 18(8+4) | 19(11+4) | 20(13+3) | 21(20+4) | 22(38+5) | 23(9+4) |
| 24(9+4) | 25(8+4) | 26(11+4) | 27(13+3) | 28(22+4) | 29(39+4) | 30(8+3) |
| 31(9+4) | | | | | | |

junio

| L | M | X | J | V | S | D |
|---|---|---|---|---|---|---|
| | 1(12+4) | 2(12+3) | 3(13+4) | 4(29+4) | 5(54+4) | 6(10+4) |
| 7(14+4) | 8(12+4) | 9(12+3) | 10(12+3) | 11(27+5) | 12(54+4) | 13(10+4) |
| 14(14+4) | 15(12+4) | 16(12+3) | 17(11+3) | 18(27+5) | 19(54+4) | 20(10+4) |
| 21(14+4) | 22(12+4) | 23(12+3) | 24(17+4) | 25(27+5) | 26(54+4) | 27(10+4) |
| 28(14+4) | 29(12+4) | 30(12+3) | | | | |

julio

| L | M | X | J | V | S | D |
|---|---|---|---|---|---|---|
| | | | 1(16+4) | 2(24+5) | 3(35+5) | 4(14+4) |
| 5(15+3) | 6(16+3) | 7(12+3) | 8(16+4) | 9(24+5) | 10(32+4) | 11(13+4) |
| 12(15+3) | 13(16+3) | 14(12+3) | 15(16+4) | 16(24+5) | 17(37+5) | 18(14+4) |
| 19(15+3) | 20(16+3) | 21(12+3) | 22(16+4) | 23(24+5) | 24(40+6) | 25(16+5) |
| 26(15+3) | 27(16+3) | 28(12+3) | 29(15+4) | 30(24+5) | 31(32+5) | |

agosto

| L | M | X | J | V | S | D |
|---|---|---|---|---|---|---|
| | | | | | | 1(20+5) |
| 2(23+5) | 3(24+5) | 4(31+7) | 5(30+7) | 6(31+5) | 7(36+6) | 8(22+5) |
| 9(23+5) | 10(24+5) | 11(31+6) | 12(30+7) | 13(31+5) | 14(34+5) | 15(18+7) |
| 16(23+5) | 17(24+5) | 18(31+6) | 19(30+7) | 20(31+5) | 21(35+6) | 22(22+5) |
| 23(23+5) | 24(24+5) | 25(29+6) | 26(30+7) | 27(31+5) | 28(32+6) | 29(20+5) |
| 30(23+5) | 31(24+5) | | | | | |

septiembre

| L | M | X | J | V | S | D |
|---|---|---|---|---|---|---|
| | | 1(20+4) | 2(21+4) | 3(34+5) | 4(50+5) | 5(15+4) |
| 6(18+5) | 7(18+5) | 8(21+4) | 9(21+4) | 10(34+5) | 11(45+5) | 12(18+6) |
| 13(18+5) | 14(18+5) | 15(20+4) | 16(22+4) | 17(34+4) | 18(50+5) | 19(16+4) |
| 20(18+5) | 21(18+5) | 22(20+4) | 23(22+4) | 24(39+5) | 25(50+5) | 26(19+6) |
| 27(18+5) | 28(18+5) | 29(20+4) | 30(23+5) | | | |

octubre

| L | M | X | J | V | S | D |
|---|---|---|---|---|---|---|
| | | | | 1(24+7) | 2(40+8) | 3(7+4) |
| 4(8+3) | 5(9+3) | 6(9+3) | 7(10+4) | 8(24+7) | 9(36+9) | 10(24+8) |
| 11(20+8) | 12(24+8) | 13(9+3) | 14(10+4) | 15(24+7) | 16(46+7) | 17(8+4) |
| 18(7+3) | 19(8+3) | 20(9+3) | 21(10+4) | 22(24+7) | 23(34+9) | 24(6+3) |
| 25(7+3) | 26(8+3) | 27(9+3) | 28(10+4) | 29(24+7) | 30(20+9) | 31(20+9) |

noviembre

| L | M | X | J | V | S | D |
|---|---|---|---|---|---|---|
| 1(2+3) | 2(2+3) | 3(2+3) | 4(2+3) | 5(17+6) | 6(24+8) | 7(2+3) |
| 8(2+3) | 9(2+3) | 10(2+3) | 11(2+3) | 12(20+6) | 13(25+8) | 14(3+3) |
| 15(2+3) | 16(2+3) | 17(2+3) | 18(2+3) | 19(20+6) | 20(30+8) | 21(3+3) |
| 22(2+3) | 23(2+3) | 24(2+3) | 25(2+3) | 26(19+6) | 27(24+8) | 28(3+3) |
| 29(2+3) | 30(2+3) | | | | | |

diciembre

| L | M | X | J | V | S | D |
|---|---|---|---|---|---|---|
| | | 1(1+2) | 2(1+2) | 3(4+3) | 4(31+9) | 5(4+3) |
| 6(11+8) | 7(7+7) | 8(8+7) | 9(0+2) | 10(2+3) | 11(3+3) | 12(0+2) |
| 13(0+2) | 14(1+2) | 15(1+2) | 16(1+2) | 17(3+3) | 18(4+3) | 19(1+2) |
| 20(0+2) | 21(1+2) | 22(1+2) | 23(0+2) | 24(15+8) | 25(20+8) | 26(1+2) |
| 27(0+2) | 28(1+2) | 29(1+2) | 30(0+2) | 31(1+3) | | |

## 4. Conclusions

Considering the high difficult to predict future room reservations, it could be considered than the results with the last model are quite good. Correlation with new data, not used to build the model, was 0.80. In contrast, the use of classical and advanced data mining algorithms, directly with the database, did not give satisfactory results.

The final model has been a heuristic algorithm based on the combination of three actions: selection of the best attributes, identification of similar days and average values obtained with robust estimation of reserves. Algorithm parameters have been optimized with genetic algorithms.

The optimization process was computed until the 500 generations but, in the last 200 generations, the best individual didn't improve (since the 300th generation). That comes to clearly indicate that the problem, for this type of hotel and these type of data, it had its own limitations because there was a high randomness in the rooms bookings. This came to mean that the models could be improved slightly but did not get very high precision due to the inherent properties of the data and the high uncertainty of the process.

Obviously, the time necessary for calculates 500 generations was very high (almost four months) but is possible to observe that in a few number of generations, the models' correlations were stable in the three principal decimal digits. In this manner, the methodology can be applied in order to obtain, in a few days, models that could be efficient enough to be used in hotel decision support systems.

## 5. Acknowledges

The authors thank for the financial support to the "Dirección General de Investigación" of the Spanish Ministry of Science and Innovation for the project DPI2007-61090, to the "Universidad de La Rioja" and "Banco Santander" for the project API11/13, and to the European Union for the project RFS-PR-06035.

Finally, the authors also thank the Autonomous Government of La Rioja for its support through the 3º Plan Riojano de I+D+I for the project FOMENTA 2010/13.

## 6. References

Aha, D. & Kibler, D. (1991). Instance based learning algorithms. *Machine Learning*, 6, 37-66.

Banerjee, T., Das, S., Roychoudhury, J. & Abraham, A. (2010). Implementation of a new hybrid methodology for fault signal classification using short -time fourier transform and support vector machines. In: Soft Computing Models in Industrial and Environmental Applications, 5th International Workshop (SOCO 2010), Advances in Intelligent and Soft Computing, vol 73, 219–225

Corchado, E. & Herrero. A., (2011). Neural visualization of network traffic data for intrusion detection. *Applied Soft Computing*, 11, (2), 2042-2056.

Haykin, S. (1999). *Neural networks, a comprehensive foundation (2nd ed.)*. EE.UU, New Jersey: Prentice Hall.

Koenker, R. & Portnoy, S. (1997). The Gaussian Hare and the Laplacian Tortoise: Computability of squared-error vs. absolute-error estimators, with discussion. *Statistical Science,* 12, 279-300.

Mitchell, M. (1998). *An introduction to genetic algorithms*. Cambridge: The MIT Press.

Quinlan, J. R. (1992). Learning with Continuous Classes. 5th Australian Joint Conference on Artificial Intelligence, Singapore, 343-348.

Sedano, J., Curiel, L., Corchado, E., De la Cal, E. & Villar, J. (2010). A soft computing method for detecting lifetime building thermal insulation failures. *Integrated Computer-Aided Engineering*, 17, (2), 103–115.

Wilkinson, G. N. & Rogers, C. E. (1973). Symbolic description of factorial models for analysis of variance. *Applied Statistics*, 22, 392-399.

## Correspondence:

Dr. Francisco Javier Martínez de Pisón Ascacíbar
Grupo EDMANS.  URL: http://www.mineriadatos.com
Área de Proyectos de Ingeniería. Departamento de Ingeniería Mecánica
Edificio Departamental. ETSII de Logroño. C/ Luís de Ulloa, 20, 26004 Logroño (España).
Phone: +34 941 299 232
Fax: + 34 941 299 794
E-mail: fjmartin@unririoja.es.