



## THE SELECTION OF OPTIMAL DATA MINING METHOD FOR SMALL-SIZED HOTELS

### IZBOR OPTIMALNE METODE DUBINSKE ANALIZE PODATAKA ZA MALE HOTELE

Angelina Njeguš<sup>1</sup>, Vanja Nikolić<sup>2</sup>, Verka Jovanović<sup>1</sup>

<sup>1</sup>Singidunum University, 32 Danijelova St., Belgrade, Serbia

<sup>2</sup>Hotel Crowne Plaza Belgrade, Serbia

#### Abstract:

Small-sized hotels that prevail in the tourist destination of Serbia rarely use any kind of property management or intelligence systems. The issue that pervades throughout this paper is related to the ways in which they can benefit from data mining. This paper discusses data mining practical application of making predictions of future monthly in-house nights for small hotels. The selected data mining algorithms have been analyzed and compared in order to choose the optimal method for application of this case study. An empirical application of methods demonstrates that it can generate reasonably accurate forecasts and can be useful to managers in their evaluation of the future occupancy rate. Furthermore, it considers which of the four algorithms is best suited for other applications in the hospitality industry.

#### Key words:

Data mining, hospitality, occupancy rate.

#### Apstrakt:

Mali hoteli, koji zauzimaju značajno mesto na tursitičkom tržištu Srbije, retko koriste neki vid savremenih sistema za upravljanje imovinom hotela ili inteligentne sisteme. Pitanje koje se prožima kroz ovaj rad tiče se načina na koji mali hoteli mogu imati koristi od tehnika dubinske analize podataka (*data mining*). U ovom radu se razmatra praktična primena algoritama za dubinsku analizu podataka u cilju predviđanja mesečne popunjenosti kapaciteta hotela. Izabrani algoritmi se analiziraju i upoređuju u cilju izbora optimalnih metoda za primenu ove studije slučaja. Empirijska primena metoda pokazuje da se generišu prilično tačne prognoze, što može biti od velike pomoći rukovodstvu u proceni budućih stopa popunjenosti hotelskih kapaciteta. Takođe se razmatraju i druge mogućnosti u okviru kojih se četiri izabrana algoritma za dubinsku analizu podataka mogu primeniti u hotelskoj industriji.

#### Ključne reči:

dubinska analiza podataka (*data mining*), hotelijerstvo, stopa popunjenosti kapaciteta.

## 1. INTRODUCTION

Significant contribution of the tourism industry to the overall economy imposes the need for accurate tourism forecasting. The planning process relies heavily on accurate forecasts in order to reduce risk in the decision-making process, and is therefore important to hotel managers (Njeguš, 2010; Nikolić, 2011). Namely, managers use demand forecasting to determine operational requirements, such as staffing and capacity, and financial viability to build a new hotel *etc.* Small and medium-sized hotels are aware of the fact that they cannot compete with the large hotel chains that constantly invest in advanced information technologies. However, the usage of those systems can reduce costs, increase revenue, and generate great business opportunities.

Forecasting estimates the future value based on customer record's patterns and is essential in the hotel industry. The decision-makers offer valuable inputs in the decision making process to a hotelier. The hotels mostly use the occupancy forecasting model. Arrival and departure patterns are needed for budgeting appropriate resources, employees, and work shifts to make the most of their resources (Aghazadeh, 2007). Seasonality is also a major part of forecasting process (Zakhary *et al.*,

2009). Apart from having additional staff and supplies during the busiest periods, hotels should ensure that guests are fully satisfied and that their stay is optimal. Another thing to forecast is the room type preference. If more guests prefer double occupancy instead of single occupancy, it would be the most beneficial to have more of that type of room available.

Demand and revenue forecasting are the most important for managing hotel business. Demand forecasting establishes selling procedures and criteria for the revenue forecast. Such information should be gathered on a daily basis to track the stay patterns for future forecasting efforts. In this article, we use demand forecasting since it provides information to managers about in-house nights, guests' arrivals, country of origin, occupancy levels, *etc.*

Fueled by the proliferation of centralized reservation and property management systems, hotel corporations can accumulate and analyze large amounts of customer data. However, small and medium-sized privately owned hotels, which prevail in any tourist destination, usually do not use advanced intelligence systems. The issue that pervades throughout this paper is related to seeking the ways to improve small hotel's robustness, so that hotel managers can predict with some certainty the monthly occupancy rate.



Even though numerous studies have applied various forecasting models to tourism demand analysis, the application of data mining methods in tourism sector is still in its infancy. Within seventy world tourism journals, during the period from 1980-2007, only 174 articles dealt with tourism demand forecasting. Among those articles, only 14 used data mining techniques, and the first was published in 1999. Furthermore, 9 of 14 articles were published during the period 2003-2007 (Law *et al.*, 2007).

This paper discusses the data mining practical application of making predictions of future monthly in-house nights for small hotels. Also, four data mining algorithms were analyzed and compared in order to choose the optimal method for application of this case study. Furthermore, it considers which of the four algorithms is best suited for other applications in the hospitality industry.

## 2. DATA MINING ALGORITHMS FOR DEMAND FORECASTING

Data mining can be defined as the usage of statistical and mathematical models for business purposes in order to discover the hidden patterns and their interpretation into valuable and useful knowledge. Some of general data mining models are association, classification, clustering, forecasting, regression *etc.* For each type of data mining models, there are numerous machine learning techniques, such as association rule, decision tree, linear/logistic regression, neural networks *etc.* These methods are used for creation of predictions and data analysis using mathematical algorithm to create robust prediction models. Data mining discovers typical patterns and influential factors in data, that can't be easily discovered by managers.

Machine learning studies the development of techniques and algorithms that allow computers to learn how to perform tasks based on empirical data. There are hundreds of prediction methods, which differ in their capabilities, results, computational costs, and comprehensibility. One of the most commonly studied tasks in machine learning is classification, which refers to mapping an instance represented by a set of attributes into a class. Classifiers usually take a dataset as input, and generate a mathematical expression as output, rule set or a decision tree that embodies the knowledge contained within the data. The knowledge extracted can be used either directly by the system or can be applied to new instances.

Some algorithms can be used in the hotel industry. Since small-sized hotels usually do not use some kind of intelligence systems, the issue that pervades throughout this work is how small-sized hotels can exploit data mining potential. Small and medium-sized privately owned hotels mostly prevail in the tourist destination of Belgrade, the capital of Serbia.

In this paper, several data mining algorithms, which are available from open source tool called "Weka", are analyzed and compared. Weka provides numerous classification algorithms. However, only tree and rule algorithms are used here as they have easily understandable behaviour. Out of the available data mining algorithms, we used M5P Tree, M5 Rules, JRip Rules, and J48 Tree, to forecast the number of guest overnight stays for an actual small-sized hotel.

### 2.1 M5P ALGORITHM OVERVIEW

M5P is modified Quinlan's M5 algorithm (Quinlan, 1992) for inducing trees of regression models (Machado *et al.*, 2011). It can be used to predict a numeric target (class) attribute. M5P

combines a conventional decision tree with the possibility of linear regression functions at the nodes. This algorithm is chosen because available data are numeric, and the aim is to find numeric prediction for the next year. The basic idea of this algorithm is to express the class as a linear combination of attributes, with predetermined weights (Witten *et al.*, 2005):

$$X = w_0 + w_1 a_1 + w_2 a_2 + \dots + w_k \quad (1)$$

$w_0, w_1 \dots w_k$   $a_1, a_2 \dots a_k$  where X is a class, are attribute values (in this case study: guest arrivals, and in-house nights per months), and are weights. Weights are calculated from based on the training data. When predicting a class for one instance (in this example for one country), the formula is as follows:

$$w_0 a_0^{(1)} + w_1 a_1^{(1)} + w_2 a_2^{(1)} + \dots + w_k a_k^{(1)} = \sum_{j=0}^k w_j a_j^{(1)} \quad (2)$$

### 2.2 M5 RULE ALGORITHM

The main problem in linear regression is that the attributes must be numeric so that the model obtained will also be numeric (simple equations in a dimensions). As a solution to this problem, decision trees have been used in data mining for a long time as a learning technique, *i.e.* models are learned from data. A decision tree divides the attribute space into cluster (leaf of the tree). Each branch of the tree has a condition which reads as follows attribute  $\leq$  value or attribute  $>$  value, that serve to make selections until a leaf is reached. Each leaf represents the average value of the instances covered by the leaf and each leaf is then turned into the rule.

One of the common methods for generating rules from the trees is M5 algorithm. The M5 builds regression trees whose leaves are composed of multivariate linear models, and the nodes of the tree are chosen over the attribute that maximizes the expected error reduction as a function of the standard deviation of output parameter (Dolado *et al.*, 2007). There are two main rule algorithm types: the association rule, and the classification rule. Rule algorithms are commonly used in data mining application for several reasons (Nagabhushana, 2006):

- ♦ It produces clear and understandable results.
- ♦ It supports undirected data mining (data mining of extremely large data sets).
- ♦ It works on variable length data (no need for summarization).
- ♦ The computation it uses is the simple one.

The M5 construct the tree in three steps:

1. Generates a regression tree using the training data, and calculates a linear model for each node of the tree generated.
2. Tries to simplify the regression tree deleting the nodes of the linear models whose attributes do not increase the error.
3. Reduces the size of the tree without reducing the accuracy.

There are some similarities between M5P tree algorithm and M5 rule algorithm (the tree in booth algorithm is made from Quinlan's M5 algorithm). After the tree is built instead of a single rule, as is usually done, a full model tree at each stage is built and the best leaf is turned into a rule. It uses the basic separate method where it looks for a rule that explains part of the training instances, removes the instances covered by that rule, and continues until no instances remained. Unlike most other rule



algorithms, M5 can handle numeric values, generate output that can be easily read and used, and give several rules.

### 2.3 JRIP RULE ALGORITHM

JRip implements a propositional rule learner, “Repeated Incremental Pruning to Produce Error Reduction (RIPPER)”, as proposed by (Cohen, 1995). JRip is a rule learner. It is similar to the commercial rule learner RIPPER. It constructs a rule set in which all positive examples are covered, and its algorithm performs efficiently on large data sets. Before building a rule, the current set of training examples is partitioned into two subsets, a growing set (usually 2/3), and a pruning set (usually 1/3). The rule is constructed from the examples in the growing set. The rule set begins with an empty rule set and rules are added incrementally to the rule set until no negative examples are covered. After growing a rule from the growing set, condition is deleted from the rule in order to improve the performance of the rule set on the pruning examples. To prune a rule, RIPPER considers only a final sequence of conditions from the rule, and selects the deletion that maximizes the function (Witten, 2005).

Unlike M5P and M5 that are producing numeric results, depending on numeric values of the attribute in training set, the RIPPER and its heir JRIP have if – then – else, structure, but can’t handle numeric classes. Thus, we used modified training set, on which the numeric predictions are transferred to nominal prediction with conversion rate from Table 1.

### 2.4 J48 TREE ALGORITHM

J48 is an open source Java implementation of the C4.5 algorithm in the WEKA data mining tool.

It is a fast algorithm that can quickly solve classification problems on large scale data sets. J48 builds tree from training data using information entropy (the disorder of data), *i.e.* information gain (an information-based measure that takes into account different numbers, and different probabilities of test outcomes).

At each node, a criterion for attribute split is the attribute with maximal difference in entropy, or maximal variation of numeric or nominal value of test outcome. The learning algorithm decides on which attribute to put in the node of the tree based on its information gain. The idea of this approach is that if the attribute that carries the most information is selected, this will minimize the size of the tree. In information theory, the amount of information is measured by means of entropy. The concept of entropy in information theory describes how much information there is in an event. The idea of entropy was introduced by Shannon in 1948. An intuitive understanding of information entropy relates to the amount of uncertainty about an event associated with the given probability distribution. Shannon defined the entropy in terms of discrete random variable  $X$ , with possible states  $x_1, x_2, \dots, x_n$  as:

$$H(X) = \sum_{i=1}^n p(x_i) \log_2 \left( \frac{1}{p(x_i)} \right) = - \sum_{i=1}^n p(x_i) \log_2 p(x_i) \quad (5)$$

That is, the entropy of the variable  $X$  is the sum, over all possible outcomes  $x_i$  of  $X$  of the product of the probability of outcome  $x_i$  times the log of the inverse of the probability of  $x_i$ . Using the entropy as a measure for information gain the approach can be described as (Gašperin, 2007):

- ♦ Chose an attribute that has the highest entropy value.
- ♦ Create a separate tree branch for each value of the chosen attribute.

- ♦ Divide the instances into subgroups so as to reflect the attribute values of the chosen node.
- ♦ For each group, terminate the attribute selection process if all members of a subgroup belong to the same class or a subgroup contains a single node.
- ♦ For each subgroup that has not been labeled as terminal repeat the above process.

The main issue with the decision trees is, that the algorithm is making new branches deep enough to perfectly fit all the training data. These results in the decision tree perfectly fit the training data, but decrease its performance over the test examples. To avoid this over fitting, a post-pruning approach will be used. With this approach, we let the algorithm to grow a full tree and then remove the branches that will cause the over fitting.

The reasons why this algorithm has been chosen for test possibilities of this case study are:

- ♦ C4.5 algorithm has an extremely wide usage possibility.
- ♦ Any versions of it are used in almost every company that has focus on consumers.
- ♦ It can easily handle not only numerical but also nominal values.

## 3. THE PROPOSED DATA MINING METHODOLOGY FOR SMALL HOTELS

The selected hotel for the case study is relatively new (from January 2008) with 15 rooms and suites. The Property Management System the hotel uses is purchased from domestic vendor, and does not have the option to record and store variety of guest information. Some deep analysis of guests or possibility to apply data mining methods to segmentation or pattern behaviour recognition is unavailable. Since the small hotels are obliged to send monthly statistical reports of guest arrivals and overnight stays to the Statistical Office of the Republic of Serbia, these reports can be the only data source for tourist demand. Therefore, the main data input is the statistical number of arrivals and in house overnight stays per year, month and per country, which can be used to predict the expected fluctuation of arrivals and overnight stays in order to fully exploit the capacity of the small hotel.

Since data are not organized in a way that facilitates the extraction, and analysis of information, the suggested data mining methodology includes the following steps: Data selection and preprocessing; Application of data mining algorithms; Interpretation of the results; and Assimilation and evaluation of the results.

### 3.1 DATA SELECTION AND PREPROCESSING

Since open source tool “Weka” uses table approach of instances and attributes (rows and columns), the original data were slightly modified, in order to be properly used. In this case study, the instances are country origin of guests, and attributes are:

- ♦ Region - every country is placed in a region according to its geographical location (The criterion for grouping countries is taken from United Nation’s geographical region division of Europe).
- ♦ Year of business - starting from the first year.
- ♦ Arrivals and in-house nights for every month of the year.
- ♦ Expected nights in January (the month is taken by random chance).

The data from the statistical report are formatted in a way so that Weka tool can manipulate them. The ways in which source data were modified are:





1. During data analysis some instances (countries) were not taken into consideration, such as the group of other European countries and the group of other world countries. The main reason is that these two groups are too heterogeneous. On the other hand, the countries outside Europe that had their classification became the new class called "The Rest of the World".
2. Depending on which machine learning algorithm is used, the class, that needs to be predicted (in-house nights), is expressed in two ways. In the first case, the class has numeric values, while in the second case, the class has nominal values, or categories, which are derived from numeric values (Table 1).

Number of in-house nights	Nominal value of the class
0-5	Very low
6-10	Low
11-20	Low Medium
21-35	Medium
36-50	High Medium
51-90	High

Table 1. Conversion of numeric data into nominal values.

3. Statistical data from the first three business years are used to generate two training sets (one for nominal and the other for numeric class). The newest data, from 2012, are used for two test sets (also one for nominal and the other for numeric class). Based on these data, the effectiveness of the analyzed machine learning algorithm is evaluated. Note that the two training sets, as two test sets, are almost identical except for the last class attribute.
4. In all sets, the number of nights in January next year was taken as a predicted class. For example, predicted class for 2008 are actual values of nights in January 2009, and for 2009 predicted class are guest nights in January 2010 and so on.
5. Class "segment importance" is introduced according to the management philosophy of the country from which their top customer comes from.
6. In order to eliminate long stays from data (because in this case it corrupts normal fluctuation of arrivals and nights), the filter was applied. First, the twelve new attributes, that represent the ratio between attribute in-house nights and attribute arrivals for every month, are introduced. The coefficient 8.00 was chosen for limited value. Finally, all the instances whose values cross the limit of 8.00, in any of twelve new attributes, were removed. These twelve new attributes were then removed, and the remaining instances taken as a training data set.

After data preprocessing, the effectiveness of four classifiers was analyzed and measured in order to choose the best one.

### 3.2 INTERPRETATION OF THE RESULTS

In Weka, every algorithm has an option of changing its mining parameters in order to increase its mining effectiveness. For each algorithm, some adjustments of parameters have been made in order to increase its predictive power.

The M5P algorithm is configured to generate regression tree. The main reason why regression tree is chosen instead of model tree is that it produces better results for small-sized hotel business. Since predictions are made for January, which is period

a low- season period for hotels, it is decided to choose regression tree. Regression tree works better with a fewer number of nights, whereas a model tree works better with more number of nights.

Figure 1 represents the graphical output for M5P algorithm of this case study. The algorithm created 5 clusters, from LM1 to LM5. Each leaf of the tree provides further information within brackets. For instance, there are 37 instances for LM1, and an approximate error in that leaf is 8.57%. The error for each node is the average of the difference between the predicted value and the actual value of each instance of the training set that reaches the node. This error is weighted according to the number of instances that reach that node.

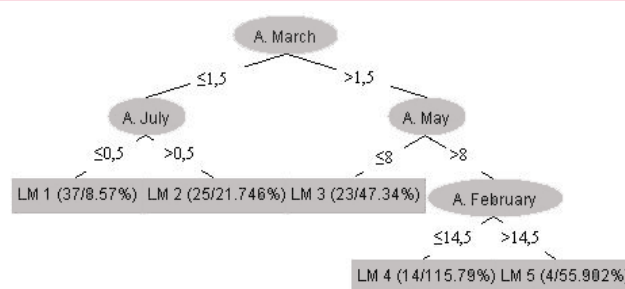


Fig 1. M5P tree output

The following results represent the predicted number of nights for January. For example, every instance that has attribute value of A. March > 1.5, and A. May > 8, and A. February > 14.5, will produce the case LM 5 of 16.0932. The M5 rule algorithm is also set to build the regression tree, while leaving other parameters default. The results are as follows:

Rule 1: IF A. March ≤ 1.5 and A. July ≤ 0.5 THEN January of next year = + 1.5103 [37/8.57%]

Rule 2: IF A. May ≤ 8 and N. November ≤ 4.5 THEN January of next year = + 4.1271 [27/17.652%]

Rule 3: IF A. March ≤ 12 N. May ≤ 18 THEN January of next year = + 8.5912 [20/33.254%]

Rule 4: IF A. February ≤ 14.5 THEN January of next year = + 18.9947 [15/74.123%]

Rule 5: January of next year = + 42.75 [4/100%]

Unlike M5P algorithm that produces a decision tree, M5 rule has if-then-else structure. Every instance is tested, and if it satisfies criteria of some of the first 4 rules, that rule will be implemented; otherwise rule 5 will be implemented.

In order to optimize JRip classifier, we set it to check the error rate, with stopping criteria at 50%. In order to maximize accuracy, we performed a 10-fold cross-validation experiment. In each of the 10 repetitions, the 9 folds were used for training and 1 fold for pruning. This leads to three classification rules, such as:

If N. May > 25 and N. August ≤ 3 then January of next year = Medium (2.0/0.0) Else

If N. March > 26 and A. October ≤ 16 then January of next year = High Medium (5.0/0.0) Else

If N. May > 19 then January of next year = Low Medium (11.0/5.0) Else

January of next year = Very Low (85.0/9.0)



Confusion matrix is probably the best indicator of predictive performance of nominal classifiers. It shows distribution of true positive and false positive predicted values. Figure 2 represents confusion matrix for JRip classifier. True positive values, or values that have been correctly predicted, are diagonally stretched, while other values, outside diagonal, are misclassified errors. In this case, there are 28 instances with class Very Low, 25 have been correctly classified, while 3 have been classified as Low, and 1 as Medium.

<i>a</i>	<i>b</i>	<i>c</i>	<i>d</i>	<i>e</i>	←	<i>Classified as</i>
25	0	3	1	0	<i>a</i> =	<i>Very Low</i>
1	0	1	0	0	<i>b</i> =	<i>Low</i>
0	0	2	0	1	<i>c</i> =	<i>Low Medium</i>
0	0	2	0	0	<i>d</i> =	<i>Medium</i>
1	0	0	0	1	<i>e</i> =	<i>High Medium</i>

Fig 2. Confusion matrix of JRip algorithm

Figure 2 shows that the two biggest mistakes are at classes Very Low and High Medium. The difference between classes of Very Low and Low is much smaller than difference between Low Medium and Medium. According to Table 1, the reason for this lies in the fact that not every class covers equal numeric values.

For J48 algorithm parameters are adjusted: confidence factor set to 0.2 (stands for amount of pruning, with smallest values for greater pruning), minimum number of instance per leaf set to 3, number of folds set to 10, and other values set to default. As a result, we have a 4 level tree, as shown in Figure 3.

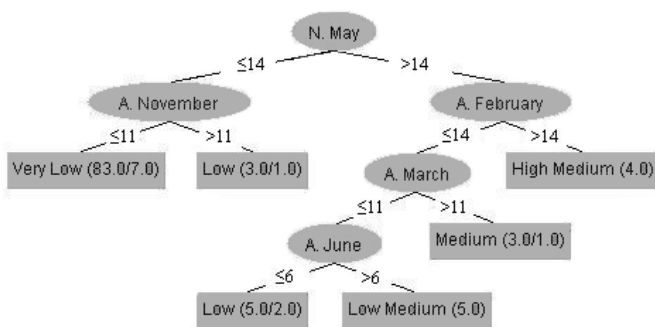


Fig 3. J48 algorithm tree output

With parameters, whose values are set, the tree complexity is greatly reduced (the other results of parameter vary from 5 to 6 level tree with much more leaves).

<i>a</i>	<i>b</i>	<i>c</i>	<i>d</i>	<i>e</i>	←	<i>Classified as</i>
25	3	0	0	1	<i>a</i>	<i>Very Low</i>
1	1	0	0	0	<i>b</i>	<i>Low</i>
0	0	0	2	1	<i>c</i>	<i>Low Medium</i>
0	1	0	1	0	<i>d</i>	<i>Medium</i>
1	0	0	0	1	<i>e</i>	<i>High Medium</i>

Fig 4. J48 confusion matrix

## 4. ASSIMILATION AND EVALUATION OF RESULTS

Results collected from all used algorithms are shown in Table 2.

Algorithm: Description	M5P Tree	M5 rule	JRip rule	J48 Tree
Mean absolute error	6.2442	7.3372	0.1343	0.1326
Root mean squared error	12.3693	13.0638	0.2845	0.2973
Relative absolute error	76.5392%	89.9376%	77.7593%	76.7554%
Root relative squared error	89.4866%	94.5116%	99.7423%	104.2223%

Table 2. Comparison of algorithm results

As presented in Table 2, algorithms JRip and J48 have significantly smaller mean absolute error and root mean squared error rate. This is a result of deriving nominal from numeric values. Nominal values of the class attribute are not separated by the same range. If class values were all of the same range (1-5,6-10,11-15..., or 1-10,11-20,21-30...) there would be:

- ♦ **Too small range of values** - there is a great chance that some values that occur in the test don't occur in the training set, and therefore, data mining process could not be tested properly. Also, there would be too many values that would produce big decision trees with many leaves covered by small number of instances per leaf. This is quite opposite from Occam's razor concept that is frequently applied in data mining. Occam's razor concept is governed by the principle that simple solutions are always better than complicated ones.
- ♦ **Too big range of values** - could produce very good predictive results, but would be practically useless. For instance, every instance with actual value of 0 would be predicted as a class that has range of 0-25, and it would have the same weight as the class with the actual value of for example 24.

In this case, the performance of JRip and J48 is not optimal for in-house night forecasting. However, it shows that these algorithms can be used for other purposes in hospitality, in particular for guest behaviour analyses. Otherwise, M5P tree and M5 rule algorithms produce outputs that can be graphically presented and easily used. Transformation from numeric to nominal classes in these two algorithms is avoided. The question that remains is whether the hotel needs one universal way of calculating the predicted values for the next year, or it should use separate set of rules or tree for every predicted month. In the first case, the advantage is on the side of M5 rule algorithm because its output predicts a greater variety of values. In the second case, the advantage is on the side of M5P algorithm because it has better predictive performance. For instance, in January, the month with the lowest season period of year, MP5 pays closer attention to smaller values which correspond to hotels in-house nights in January.

## 5. SUMMARY

This paper analyses different machine learning models for demand forecasting in the hotel industry. Since small-sized hotels in Serbia rarely use any kind of intelligence systems, the most frequent issues hotel managers are confronted with include lack of data sources, tools, and dilemmas related to choos-



ing the optimal data mining model for forecasting. This paper also analyses four data mining models available from open source tool Weka, in order to give optimal solutions in demand forecasting. Also, data mining methodology is suggested to managers of small-sized hotels. The predictive power of data mining algorithms is around 70-75%. Although this may seem low, we should note that this research was limited to only 120 instances. Out of these 120 instances, 17 aren't used because they stray too much from the normal flow of tourism demand (long stays). The performance of data mining process depends directly on the number of available cases (instances) that can be used. As presented in the paper, even small-sized hotels, without additional or with small investments, can benefit from data mining. Its usage cannot guarantee optimal business results, but can substantially reduce risks of making bad decisions. The results show that there is no single best algorithm that could beat other algorithms in all situations. To conclude, M5P and M5 are suitable for tourism demand forecasting, while JRip and J48 give better results in guest behaviour analyses.

## REFERENCES

- Aghazadeh, S.M. (2007). Revenue Forecasting Models for Hotel Management. *Journal of Business Forecasting Methods and Systems*, 26(3), 33-37.
- Cohen, W. (1995). Fast effective rule induction. Proceedings of the 12th International Conference on Machine Learning, 115-123.
- Dolado, J.J., Rodriguez, D., Riquelme, J., Ferrer-Troyano, F., & Cuadrado, J.J. (2007). A Two Stage Zone Regression Method for Global Characterization of a Project Database. Available at: <http://www.cc.uah.es/drg/b/DolRodetAlIdea07.pdf> (accessed 5.7.12).
- Gašperin, M. (2007). Case Study on the use of Data Mining Techniques in Food Science using Honey Samples. Available at: <http://kt.ijs.si/PetraKralj/DM-2007/seminar-Gasperin.pdf> (accessed 28.06.2012).
- Law, R., Mok, H., & Goh, C. (2007). Data Mining in Tourism Demand Analysis: A Retrospective Analysis. Proceedings of the 3rd International Conference on Advanced Data Mining and Applications, 508-515.
- Machado, A., Carvalho, D., Clua, E., Duarte, C.G., Montanari, M.V., & Reis, W.M.P. (2011). An Architecture Based on M5P Algorithm for Multiagent Systems. Available at: <http://www.lbd.dcc.ufmg.br/colecoes/enia/2011/0024.pdf> (accessed 5.07.12).
- Nagabhushana, S. (2006). *Data Warehousing OLAP and Data Mining*. New Delhi, India: New Age International Publishers.
- Nikolić, V. (2011). Researching New Technologies in Hospitality for Customer Satisfaction Purpose. Proceedings of International Scientific and Technical Symposium: Strategic Quality Management in Hospitality, 181-195.
- Njeguš, A. (2010). Importance of Business Intelligence in Tourism. *Singidunum Review* 5 (2), 129-137.
- Quinlan, J.R. (1992). Learning with continuous classes. Proceedings of the Australian Joint Conference on Artificial Intelligence, 343-348.
- Quinlan, J.R. (1996). Improved Use of Continuous Attributes in C4.5. *Journal of Artificial Research* 4, 77-90.
- Witten, I.H., & Frank, E. (2005). *Data Mining Practical Machine Learning Tools and Techniques*. CA: Morgan Kaufmann, Elsevier.
- Zakhary, A., Atiya, A.F., El-Shishiny, H., & Gayar, N.E. (2011). Forecasting hotel arrivals and occupancy using Monte Carlo Simulation. *Journal of Revenue and Pricing Management*, 10 (4), 344-366.