A GIS-based decision-support system for hotel room rate estimation and temporal price prediction: The hotel brokers context

Slava Kisilevich[a,*], Daniel Keim[a], Lior Rokach[b,c],

[a] *Data Mining, Databases and Visualization, Department of Computer and Information Science, University of Konstanz, 78457 Konstanz, Germany*
[b] *Deutsche Telekom Laboratories at Ben-Gurion University, Beer-Sheva, Israel*
[c] *Department of Information Systems Engineering, Ben-Gurion University of the Negev, P.O.B. 653, Beer-Sheva, Israel 84105*

Abstract

The vastly increasing number of online hotel room bookings are not only intensifying the competition in the travel industry as a whole, but also prompt travel intermediates (i.e. e-companies that aggregate information about different travel products from different travel suppliers) into a fierce competition for the best prices of travel products, i.e. hotel rooms. An important factor that affects revenues is the ability to conclude profitable deals with different travel suppliers. However, the profitability of a contract not only depends on the communication skills of a contract manager. It significantly depends on the objective information obtained about a specific travel supplier and his/her products. While the contract manager usually has a broad knowledge of the travel business in general, collecting and processing specific information about travel suppliers is usually a time and cost expensive task. Our goal is to develop a tool that assists the travel intermediate to acquire the missing strategic information about individual hotels in order to leverage profitable deals. We present a GIS-based decision-support system that can both, estimate objective hotel room rates using essential hotel and locational characteristics and predict temporal room rate prices. Information about objective hotel room rates allow for an objective comparison and provide the basis for a realistic computation of the contract's profitability. The temporal prediction of room rates can be used for monitoring past hotel room rates and for adjusting the price of the future contract. This paper makes three major contributions. First, we present a GIS-based decision support system, the first of its kinds, for hotel brokers. Second, the DSS can be applied to virtually any part of the world, which makes it a very attractive business tool in real-life situations. Third, it integrates a widely used data mining framework that provides access to dozens of ready to run algorithms to be used by a domain expert and it offers the possibility of adding new algorithms once they are developed. The system has been designed and evaluated in close cooperation with a company that develops travel technology solutions, in particular inventory management and pricing solutions for many well-known websites and travel agencies around the world. This company has also provided us with real, large datasets to evaluate the system. We demonstrate the functionality of the DSS using the hotel data in the area of Barcelona, Spain. The results indicate the potential usefulness of the proposed system.

Keywords: Hedonic methods, Hotels, Price prediction, Geographic Information Systems, Regression Analysis, Data Mining

1. Introduction

With the ongoing penetration of the Internet and mobile technologies into all aspects of our lives, the number of online users is growing rapidly. As a result, consumer behavior is changing towards online shopping, which provides such benefits as product and price comparisons, ease of use, speed of purchase transaction, and trust [1, 2]. This trend is especially noticeable in the travel domain. More and more online travel websites have been emerging, including hotel advertisements and websites that aggregate information about hotel room rates around the world [3]. The advantage of travel aggregates, which are also referred to as travel intermediates or brokers is that they allow customers to simultaneously gather information about many hotels at their travel destination. Thus, the user can compare prices easily rather than having to search for single information about individual hotels and having to visit each hotel's website.

The competition between travel intermediates is very intense and there are many risk factors that can degrade revenues such as the quality of the website (ease of use, visual attractiveness) [4], the speed of execution, the level of user satisfaction [5], the lack of innovative tools and services, and the level of professionalism of their employees. How-

*Corresponding Author: Tel.: +49 7531 88 3536
Email addresses: slaks@dbvis.inf.uni-konstanz.de (Slava Kisilevich), keim@dbvis.inf.uni-konstanz.de (Daniel Keim), liorrk@bgu.ac.il (Lior Rokach)

ever, the most important factor is the ability to contract with different travel suppliers. This factor is characterized by two underlying issues: contracting with as many travel suppliers as possible and concluding profitable contracts. While the first issue is mostly organizational, the second issue is related to the personal ability of contract managers to conclude contracts and their comprehensive knowledge of the travel business. This knowledge, however, is dependant on the strategic information available about a specific travel supplier and his or her products. In reality, strategic decisions are reached using a limited amount of information due to the inability to acquire and process sufficient information in a sufficiently short time. With this in mind, we aim in this paper to improve the decision-making capability of hotel brokers by introducing a GIS-based decision support system. Our decision support system enables the broker to objectively estimate hotel room rates based on the intrinsic and locational characteristics as well as historic room rates of a given hotel or hotels with similar characteristics.

The analysis of product prices and factors that influence the price has been widely used in finance, economics and real estate property assessments since Rosen [6] formulated the property of price as the weighted sum of the different characteristics composing the product. In the hedonic pricing model (usually analyzed by linear regression) that he proposed, independent variables are the product characteristics relevant for the analysis, while the price serves as a dependent variable. Therefore, by finding the hotels with the same characteristics that affect hotel prices, it will be possible to compare room rates between similar hotels.

For understanding the factors that affect property prices and hotel room rates in particular, the use of the hedonic pricing theory has received much attention (e.g. [7, 8, 9, 10, 11, 12]). However, the results show that there is no universal solution as to what characteristics should be included and what analytical methods should be applied [13]. Sometimes the results are even contradictory [14]. Among the various reasons for differences in results, we can name such factors as: empirical methods selected for the analysis (linear and non-linear regression estimators, parametric and non-parametric algorithms); data quality and completeness; region of application; and characteristics included in a model.

The various studies in property valuation, including the hotel domain, showed the importance of considering such locational characteristics as the relative distance of a property to a city center or the distance to business centers in the models. Moreover, hotels have their distinguishing properties, such as the proximity to the waterfront. However, including locational characteristic into the model is very difficult for several reasons. First, the definition of locational properties is usually an ill-structured problem since it is difficult to agree on the definite spatial resolution (distance, areas, spatial density), which may or may not influence the results. It is easier to answer the question about non-spatial characteristics like *Is there a hairdryer in the room?* than answering the question *How many points of interest are there around the hotel?* since *around* is not precisely defined in terms of distance. Second, the precision and availability of the spatial data limit its use in the analysis.

For these reasons, a completely automated solution process [9] is not feasible since the guidance of an expert is paramount in the case of ill-structured problems and the task at hand. Clearly, there is a need for an interactive decision-support system (DSS) [15, 16, 17] that would help the analyst in testing different hypotheses regarding price factors for selected hotels. In this system, the analyst will be able to select the region of investigation by accessing all the necessary data from his/her corporate database. It would allow him/her to add additional data that he/she thinks is important in the analysis. Such data, for example, could be points of interest around hotels, transportation availability, historical places or information about the proximity of a hotel to the waterfront, etc. Enabling the analyst to build different models and apply various algorithms, the system will help the analyst decide about the desirability of a hotel and the objective room rate.

Geographic Information System (GIS) technology has proven to be useful for businesses. Its addition to a business decision-making environment improves the performance of the decision-maker [18]. Moreover, the importance of GIS in property valuation has been discussed in numerous works (e.g. [19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30]). However, the usage of GIS in these instances was mostly limited, either to utilizing spatial queries and distance measures or it was limited to one location only. A few works discuss the implementation requirements of a GIS. In any case, these were not robust due to the lack of appropriate technology, incompatibility in programming interfaces, or were implemented by integrating different components by data import and export facilities, which Denzer [31] called *null integration*. At the same time, no work, to the best of our knowledge, presented a robust and easy-to-use GIS-based solution that can be used in real life scenarios.

In contrast to past attempts, we provide a flexible and highly interactive GIS-based decision support system with rich functionality. Our system is integrated with a real GIS that provides support in order to input and layer spatial data; to represent complex spatial relations; to analyze spatial data; and to output spatial data in the form of maps [32]. The problem of spatial data acquisition, a crucial factor in past research, has been solved by utilizing OpenStreetMap crowdsourcing data [33], which comprises contributions from thousands of individuals around the world. Although, some features like the proximity of a hotel to the seafront are not available in OpenStreetMap, the analyst can use simple user interface to decide about this feature and to note that the hotel is being examined further with regard to this information. The integrated data mining package provides the domain expert with ac-

cess to dozens of available, ready to run algorithms.

The contribution of the paper can be summarized as follows:

1. The decision support system that we developed meets the real world business requirements of hotel brokerage companies for estimating a hotel's objective room rates.
2. Due to the integration of OpenStreetMap public data, the proposed DSS can be applied on different parts of the world and is not constrained to analyzing a specific region, as is usually the case in instances cited in the academic literature.
3. The provision of attributes that are usually difficult if not impossible to acquire or determine, e.g. proximity to the waterfront or attributes that are based on spatial characteristics (radius, distance), is facilitated by an interface that allows the domain expert to provide the desired information.
4. The framework is not limited to a predefined hotel characteristic and a single regression estimator but can use a variety of linear and non-linear estimators available in the data mining package that is embedded into the framework along with the capability of selecting the characteristics the domain expert believes to be important in a particular situation. This feature is particularly important since economic theory does not provide guidance about selecting characteristics and determining how these characteristics relate functionally to their product price and what are the best algorithms to apply [13, 34].

2. Related Work

In this section we review works related to the field of property valuation and the characteristics that influence the hotel room prices. The methods, which are used in both fields, are very common and are based on hedonic pricing models.

There are two main approaches inherent in the hedonic model. The first seeks to estimate how individual characteristics influence the overall price of a property or a room (Section 2.1). The second approach deals with generating and evaluating a model that can be used in the price prediction (Section 2.2) that is close to the goal of our paper. Finally, we show related works in which hedonic pricing models are integrated with a GIS (Section 2.3).

2.1. Determinants of room rates

The influence of a hotel location on room rates and the price contribution of a specific attribute were investigated in [35]. Initially, Bull included five independent variables in the hedonic model (hotel star rating, *age* of a building, availability of a restaurant, distance from the city center, and a binary variable *side* that represents whether a hotel faces a river side). However, *age* and *side* were excluded from the final model because their influence on the variance was insignificant. The results showed that the distance from the center is the strongest spatial determinant of hotel room rates (the room rate fell per kilometer from the center). In addition, the availability of a restaurant and hotel rating in stars increased the room rate.

Israeli [36] studied the influence of the number of rooms, star rating and corporate affiliation on room prices using 215 hotels and 30,000 rooms in three regions in Israel (Tel-Aviv, Jerusalem, and Eilat). The star rating was found to be the most consistent determinant in hotel price differences. However, the brand affiliation showed contradictory results. While for hotels in the Tel-Aviv area, brand affiliation had no influence on the price, brand affiliation increased the room rates in the Jerusalem area. In Eilat the chief factor in hotel price differences was the price discounts. The number of rooms and consequently the size of a hotel was another significant factor - the larger the hotel, the higher the room prices.

Examining 15 bed and breakfasts with a total of 36 rooms in Walworth County, Wisconsin, Monty and Skidmore [7] used a hedonic pricing model and regression analysis to study the influence of hotel characteristics on price and willingness to pay. The results showed that location is the strongest determinant for willingness to pay. The price of a bed and breakfast accommodation increases if it is located less than a mile from a city center. Other significant price determinants are room sizes, availability of hot tub, and private bathroom. Swimming pool, theme rooms, air conditioning, fireplace, kitchen appliances, the overall number of rooms in the accommodation, and gift certificates were found to be insignificant.

The relationship between availability of certain hotel attributes and room rates for single and double rooms was investigated in [8] using the data from about 74 hotels in and around Oslo, Norway. Among the attributes included in the model were the availability of mini-bars and hairdryers which proved to be the strongest determinant of room rates. In the case of single rooms, the rate was significantly higher in chain affiliated hotels but lower in hotels that offered room service. In the case of double rooms, chain affiliation had no influence on the room rate, while the distance from the center of Oslo was a significant factor for a decrease in prices. Such attributes as swimming pool or availability of restaurant had no influence on the prices.

By applying quantile regression analysis, it was shown in a study about hotels in Taiwan [10] that the age of hotels is negatively related to the hotel price, while hotel size has a positive influence on the price. It was also shown that chain affiliation and distance from the center of the city had no influence on the room rates. In another study that included 73 hotels in Taipei [11], it was found that room rates were significantly influenced by hotel location, TV, Internet access, and availability of the fitness center, while breakfast, business centers or swimming pools did not influence the room price.

A recent study on the effect of the location on the room prices at airport hotels in the US, [12] showed that hotel prices are affected by the proximity of a hotel to an airport or to central business districts. The room rates were higher in hotels that were affiliated with a chain and in hotels that provided free parking. However, the room rates were lower in hotels that provided free breakfast.

2.2. Property valuation

To improve estimations, research in property valuation has lately begun to concentrate on comparisons between the performance of various algorithms in addressing such issues as data normality, multicollinearity, heteroscedasticity, non-linearity, spatial dependency and spatial heterogeneity.

Methods based on neural networks were compared to the multiple regression in [37, 38, 26, 39] and showed that non-parametric methods do generally not outperform traditional multiple regression methods. In other studies the results proved the opposite [40].

To address the issue of a possible price variability over a large area (spatial non-stationarity), [41] examined the effect of geographically weighted regression (GWR) [42] on predictive accuracy. This method generates a separate regression equation for each data point, and gives more weights to points located near the given data point. Both, the R-squared goodness of fit and the predicted accuracy of GWR, were higher than the traditional linear regression model. In a recent study, [43] tested several models including GWR. The authors report that according to the goodness of fit measure, the performance of GWR was better, while its coefficients were correlated. This, according to the author, reduced confidence in the method.

2.3. GIS Integration

Sarip [27] utilized MapInfo Professional GIS software to facilitate integration of spatial data into an artificial neural network model of property valuation. The GIS was used for measuring distances between properties, spatial queries and thematic mapping. However, due to the lack of a common programming interface, the spatial modeling and price valuation tasks were divided into several heterogeneous components.

Kaboudan and Sarkar [29] proposed modeling the prediction of individual property prices using average neighborhood home prices instead of individual home prices. Three different neighborhood specifications were defined: census tract, assessor's parcel number, and zip code. However, the authors encountered a problem of proper resolution and extension of neighborhoods that was not possible to define without using a GIS. The authors used ArcGIS and its usage was limited to only two preparational tasks: geocoding of house addresses and boundary definition of neighborhoods according to each of the three specifications.

García et al. [30] integrate a GIS into an automated process for property valuation. Although, the GIS was specifically developed for the problem of property valuation, its usage was limited in several aspects. First, the area of applicability was limited to properties in Albacete, Spain. Second, only after the artificial neural network model had been trained automatically, the domain expert was able to use the GIS to select a property for valuation. The GIS was developed in a SciViews graphics environment of R software, which is designed to enable software development of generic GUI-based solutions. Such solutions do not allow development of fully functional geographical information systems.

3. Problem Domain

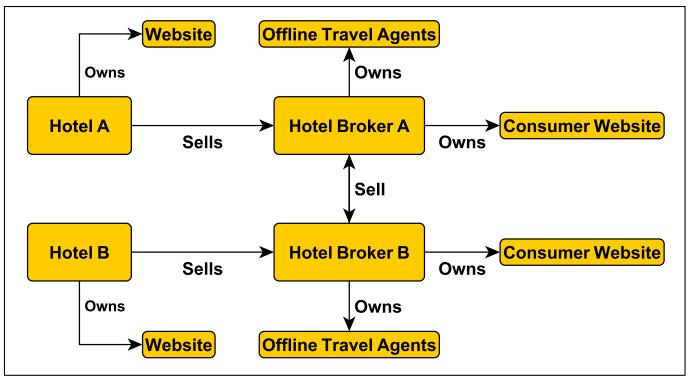

Figure 1: Interaction between hotels and hotel intermediates

Profitability of any travel intermediate is directly related to the discount rate contracts that are acquired and to the intermediate's ability of selling the product to customers. Travel intermediates are dependent on their staff of professional and highly paid hotel contract managers to negotiate the best contract. Since the number of hotels in the world is rather large and the negotiation process is long, any particular travel intermediate has a relatively small amount of contractors it can assign to any of the available destinations. Consequently, a contractor is faced with two challenges: (1) to identify hotels that fit the profile of their end customers, and (2) to identify hotels to which managers would be inclined to give better rates during negotiations.

The interaction between hotels and hotel intermediates is schematically depicted in Figure 1. A hotel usually has its own website where it directly promotes its rooms. The website is the most profitable selling channel because no intermediates are involved. However, the exposure of a hotel web page to a vast audience is limited because customers prefer using one or two travel web sites to compare hotel price lists rather than to search for individual hotels. Therefore, hotels are interested in having other channels

advertise them in order to increase exposure to end customers.

As depicted in Figure 1, *Hotel A* is exposed through the *Hotel Broker A* channel, while *Hotel B* is exposed through the *Hotel Broker B* channel. Similarly, hotel brokers promote their products through consumer websites and offline travel agents. The hotel broker may also sell hotel nights to other broker company if that brokerage company does not have already a contract with the hotel. The hotel intermediate can obtain the best price by working directly with the hotel. The hotels sell room nights to the hotel brokers in the form of discount rate contracts. Hotel brokers are committed (as part of the contract) to retaining the prices that appear on their online channels similar to the prices provided by hotels through their own websites. Therefore, the revenue of the travel intermediates constitutes the difference between the final hotel price and the contract cost. Consequently, the travel intermediates are very much interested in concluding the contract at the lowest possible price and to deal with the hotels directly rather than buying rooms from other hotel brokers. This means that, the analysis of existing contracts and knowledge about similar hotels will facilitate the decision-making about the profitability of a future contract. If a *Hotel Broker B* knows that *Hotel A* is identical to *Hotel B* (whose contract they have already acquired) in terms of characteristics that determine the hotel prices, then this knowledge will provide the leverage to negotiate a profitable deal with *Hotel A*. The proposed decision support system is designed to help the hotel brokerage company acquire the knowledge it needs about *Hotel A*. In addition, the same approach can also help in analyzing the profitability of existing deals by finding hotels similar in terms of their characteristics but different in regard to the prices they advertise.

4. Data

The hotel data below was provided by Travel Global Systems (TGS)[1], a travel service provider and hotel brokerage company. The data is divided into a static and a dynamic components. The static data includes the names of hotels, their internal IDs, their location coordinates in World Geodetic System (WGS84), hotel facilities, room amenities, and hotel categories. The dynamic component includes the room prices for one night that customers received during their search for accommodation, the date of search, and the date of order. The type of room desired was not specified in the data. This is why we assume that the average price of a hotel is related to a standard room type, most common in most of the hotels. Consequently, we selected only those room amenities that corresponded to a standard room. Table 1 presents the complete list of attributes available for analysis.

[1] http://www.travelholdings.com/

Each amenity and facility type has an internal identification number. However, preprocessing was required since some of the amenities and facilities that referred to the same entity were represented by different IDs and names. For example, what was referred to as *Wireless Internet* in one hotel, was referred to as *High-speed Internet* at another. We manually processed all the amenities and facilities and merged those that referred to the same entity providing a mapping between corporate IDs and those used in our system.

5. Models

As was discussed in Section 1, hedonic pricing models are usually used for property valuation and for determining the influence of individual characteristics on room rates. The room rate is averaged during a time period selected for analysis and expressed through the linear or non-linear combination of property (hotel) characteristics. However, hotel room prices are very volatile and prices may drastically change from season to season. Room rates depend also on the gap between the day a customer searches for available rooms and the day he/she wants to check-in. In general, it can be expected that the larger the time interval between the search date and the check in date, the lower the price of a hotel room. Consequently, the price estimation that uses only the hedonic model is insufficient for accurately estimating hotel room prices. Moreover, the hedonic pricing model does not cope with daily price variations. Therefore, we propose two models. The first model, referred to as *static*, is based on hedonic pricing where the prices are expressed through static characteristics. The second model, referred to as *dynamic*, is based on historical hotel room rates only. The following sections describe the structure of these models in detail.

5.1. Static model

During the discussion with the TGS representatives, including contract managers, about their requirements, we were asked to enhance support for analyzing locational characteristics. The commonly used *distance to a city center* is too general a measure to capture price differences between hotels. It is also imprecise because it is difficult to precisely determine whether the city center is a geographical location or simply a virtual concept. And it is very possibly that there may be more than one city center. It was therefore decided to introduce more geographical relations such as density, area, and distance. According to the requirements, the static model should capture the difference between regions with few and with many hotels; regions with few and with many points of interest; whether a hotel is located in one of these regions; the area of such regions; and the distance from points of interest to hotels.

To address the areal relationship between hotels, we first had to transform the point-based geographical space into a region-based representation. For this, we used Voronoi

tessellation [44]. The Voronoi tessellation decomposes the metric space into regions of nearest neighbors using a set of generating points. Every point in a region is closest to the generating point that generated the given region. In our study, this set of points can be any external data important for determining hotel prices (e.g. museums, historical places, transportation locations). The example of a model generated by Voronoi tessellation using museum data is presented in Figure 2. Red polygons are the generated regions and museum sites are the generating points used to generate the regions. Figure 3 shows the location of the hotels with respect to the generated regions. The size of the region indicates the relative density of the generating points. Therefore, the larger area indicates the region of low density of a specific point of interest, while the smaller area indicates a point of interest's higher density. Consequently, we can judge the relative popularity of a hotel with respect to the region in which it is located. The advantage of Voronoi tessellation over other possible clustering approaches is that it does not have any controlling parameters and produces only one solution if the number and location of generating points does not change. The following generating points were included into the overall model: museums, historical places, places of worship, transportation, restaurants and pubs.

In addition to the intrinsic hotel attributes presented in Table 1, the following locational attributes were introduced:

- **Nearest Object Count** - Counts the number of objects that are nearest to hotels for each of the generating points and for each hotel. In other words, each hotel that happens to be located the nearest to a point of interest gets score of 1 for a specific point of interest and incrementing its score for each point of interest for which it is the nearest.
- **Hotels in Neighborhood** - The number of hotels in the neighborhood of a given hotel. The neighborhood was defined as a radius of a user-specified size. Three default sizes were defined: 100m, 200m, and 500m. We would like to stress the difference between regions generated by Voronoi Tesselation and the neighborhoods. Regions provide some useful information about popularity of hotels relative to a point of interest. Neighborhoods provide useful information about the popularity of a specific hotel relative to the points of interest around the hotel within the specified radius.
- **Objects in Neighborhood** - The number of objects (museums, restaurants, etc) in the neighborhood of a given hotel. The same radius size as in **Hotels in Neighborhood** was used.
- **Hotel-Object Distance** - The distance from a hotel to an object in km.
- **Region Coverage** - A region that covers a hotel.
- **Hotels-Area** - The density of hotels in each region as a number of hotels in the region divided by the area of the region in square km.

The price variable was specified as an average market price by dividing the average hotel room price by the average hotel room price of all hotels selected for analysis. The non-normalized real and estimated prices were then recovered by multiplying the average market price and the predicted price on the denominator (the average hotel room price of all hotels). The average hotel room price was calculated as follows. First, the price of a hotel room for a given day was calculated as an average price at a given search date and all the combinations of check-in dates. Then, the total average price of a hotel was calculated as an average of hotel room prices over all dates for which users performed the search.

The above procedure can be formalized as following: Let $p_{i,j,k.l}$ denotes the unit daily price reported by the system for hotel i and check-in date k as a response to lth search performed on search date j. Thus the average room price for a given hotel on a certain search and check-in dates is computed as:

$$p_{i,j,k.*} = \sum_{\forall l} p_{i,j,k.l} \tag{1}$$

Then the average room price for a given hotel on a certain search date is calculated as:

$$p_{i,j,*.*} = \sum_{\forall k} p_{i,j,k.*} \tag{2}$$

Note that the last equation goes over all possible check-in dates stored in the database (i.e. all check-in dates in which at least one user has asked for a quotation).

Based on Equation (2) it is possible now to calculate the grand average of hotel i, namely:

$$p_{i,*,*.*} = \sum_{\forall j} p_{i,j,*.*} \tag{3}$$

Finally we calculate the normalized price value of hotel i as following:

$$y_i = \frac{p_{i,*,*.*}}{\sum_{\forall i} p_{i,*,*.*}} \tag{4}$$

Note that for Equation (4) we assume that all hotels in the database are available during the entire period. If the target hotel i was available for only a portion of the time, then it should be normalized accordingly. Namely the values that are summed up in the denominator of Equation (4), should refer to the availability dates of the target hotel.

We combine the intrinsic hotel attributes and the locational attributes into a scoring function that predicts the normalized price value of hotel i:

$$\hat{y}_i = f(a_1(i), \ldots, a_n(i))$$

where $a_k(i)$ specify the value of attribute k for hotel i. For learning the function f we can use any supervised learning method that is capable to learn numeric target attributes. Various methods might use different loss functions for training the model. For example a linear regression can be used to model the function as:

$$\hat{y}_i = \sum_{k=1}^{n} w_k \times a_k(i)$$

where the w_k coefficients are found by minimizing the mean least square errors loss function, namely:

$$err = \frac{1}{m} \sum_{i=1}^{m} (y_i - \hat{y}_i)^2$$

where m indicates the number of hotels that are included in the database. Linear regression is probably the simplest model function which can be used for our purposes. In this paper we examine other much more complicated methods such as neural networks.

5.2. Dynamic model

The dynamic model captures the temporal variability of hotel room prices. For each day on which at least one customer performed a search for available rooms, the target price (dependent variable) was expressed as the average room price of a target hotel within a check-in window of 7 to 21 days divided by grand average (see Equation 3) of a target hotel. We chose to analyse a check-in window of 7 to 21 days because, according to our analysis, most of the rooms are usually ordered in this window. In addition, we calculated the average prices for a specific search date of a target hotel with a check-in time interval of 7 to 21 days of the other hotels. The results were included in the model as independent variables. In cases where no search for available rooms was performed for a specific day, the entry was marked as a missing value. Therefore, the model consisted of N rows (the number of available search dates from the database times the number of hotels). Each row composed of M columns, each representing an average hotel price at a specific search date, and one column as a dependent attribute with the target hotel price at a specific search date.

6. The Decision Support System

Our GIS-based DSS follows the design guidelines of a general purpose GIS-based DSS and integrates the following characteristics: analytical and spatial modeling capabilities; spatial and non-spatial data management; domain knowledge; spatial display and reporting capabilities [45]. Moreover, like most modern DSSs, our DSS supports different stakeholders. This necessitates special consideration in regard to usability and ease of use during the system design stage. Therefore, we implemented our system by following general DSS guidelines and spatial DSS planning as suggested in [32]:

1. The user interface is powerful and easy to use.
2. The system combines analytical models and data in a flexible manner.
3. The system explores the solution space by using the models and generating feasible solutions.
4. The system inputs, represents, and outputs spatial data.
5. The system output appears in different forms (maps, non-spatial statistics).

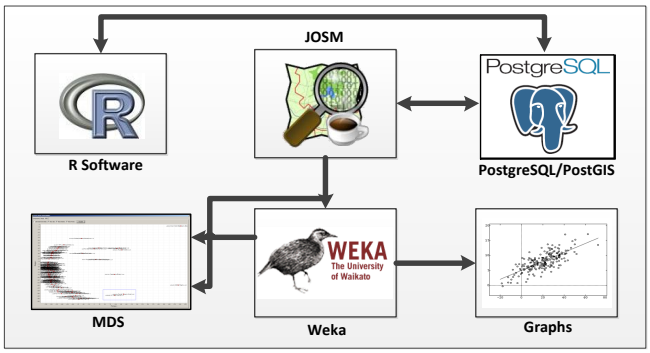

Figure 4: System components

Figure 4 presents the six main components that constitute our system (the detailed explanation of each component is provided in the following sections):

1. Java OpenStreetMap Editor (JOSM) - a GIS-based framework.
2. R Software - a statistical package.
3. PosgreSQL/PostGIS - a DBMS with spatial support.
4. Weka - a data mining framework.
5. MDS - a Multidimensional Scaling component for exploratory data analysis.
6. Graphs - a number of components that visualize the price estimation results.

Since the decision making process depends on many intermediate tasks (e.g., data integration, evaluation, visualization), component integration is an important issue during the development process and has a great impact on the system's performance, usage and acceptance by stakeholders when it is deployed [46]. The heterogeneous components presented in Figure 4 were integrated into a single software solution since this was the most effective way to achieve maximum flexibility and ease of use. Sugumaran and Degroote [45] argue that this type of integration is "less common because it is not very likely for a single piece of software to have tools out of the box to meet

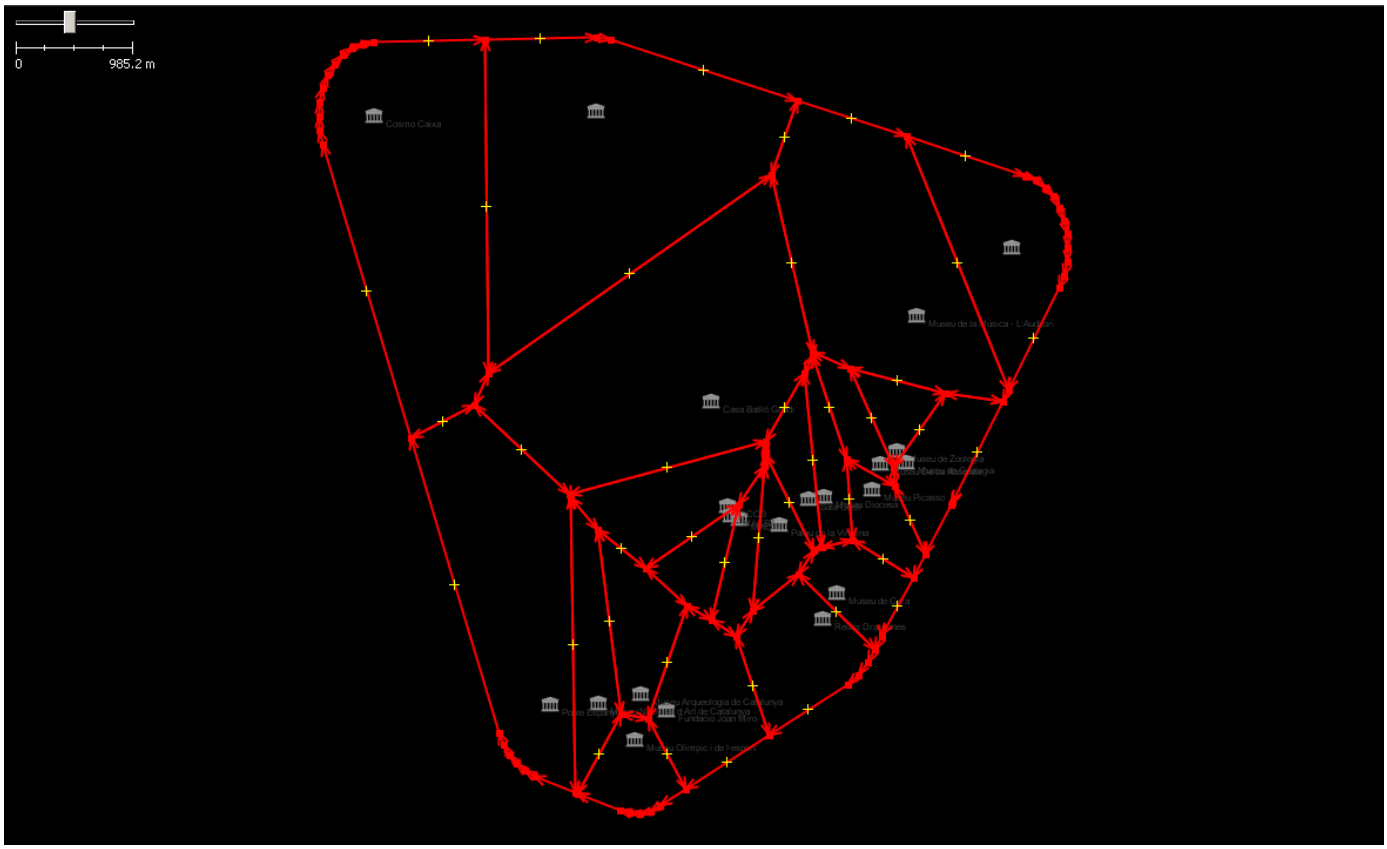

Figure 2: Spatial Model of Museums using Voronoi Tesselation: red polygons are the generated regions, museums are the generating points

all the necessary functionality requirements, and it is often expensive in the short term to develop all functionality within a single piece of software". While there is no need to address the first part of the statement because the facts speak for themselves and it is possible to meet all the necessary functionality requirements, the second half of the authors' statement is questionable since the authors forgot to take into account the power of free and open source software [47]. Since all the components presented in Figure 4 are free and open source, we were able, within a relatively short period of time, to almost seamlessly integrate them without making too many changes to the original source code[2].

6.1. Java OpenStreetMap Editor as a GIS platform

Many GIS frameworks that handle spatial data (e.g., OpenJump[3], UDig[4] or MapWindow GIS[5]) are freely available. Albeit the JOSM's functionality is comparable to general purpose GIS frameworks (e.g. it can present the spatial data in different layers and it features extendibility through its plug-in interface), what makes JOSM superior to all other GIS candidates is its inherent support of OpenStreetMap (OSM) data. This application is the primary source of external data in our DSS and its ability to discern different kinds of (OpenStreetMap) data is one of the prerequisites for effective decision support according to [48].

The main view of JOSM is presented in Figure 5, where our interface to the decision support system is outlined by the black rectangle at the bottom right corner.

6.1.1. Data Integration

The data collection process is an integral part of JOSM. JOSM reads the data from the OpenStreetMap database by selecting the boundary of the area. The data can then be saved and loaded locally into the proprietary OSM XML format. In order to obtain data for a desired region, the data manager uses the functionality provided by JOSM.

The OpenStreetMap data exists as two different types: (1) point data (*nodes*), which has coordinates expressed in longitude and latitude, and (2) *ways*, which express areal features that themselves are referenced through *nodes*. The geographical features have a list of attributes that come in a *key=value* form and determine different charac-

[2]The complete system was developed in a two-month period by two undergraduate students at the Department of Information Systems Engineering of the Ben-Gurion University of the Negev

[3]http://www.openjump.org/

[4]http://udig.refractions.net/

[5]http://www.mapwindow.org/

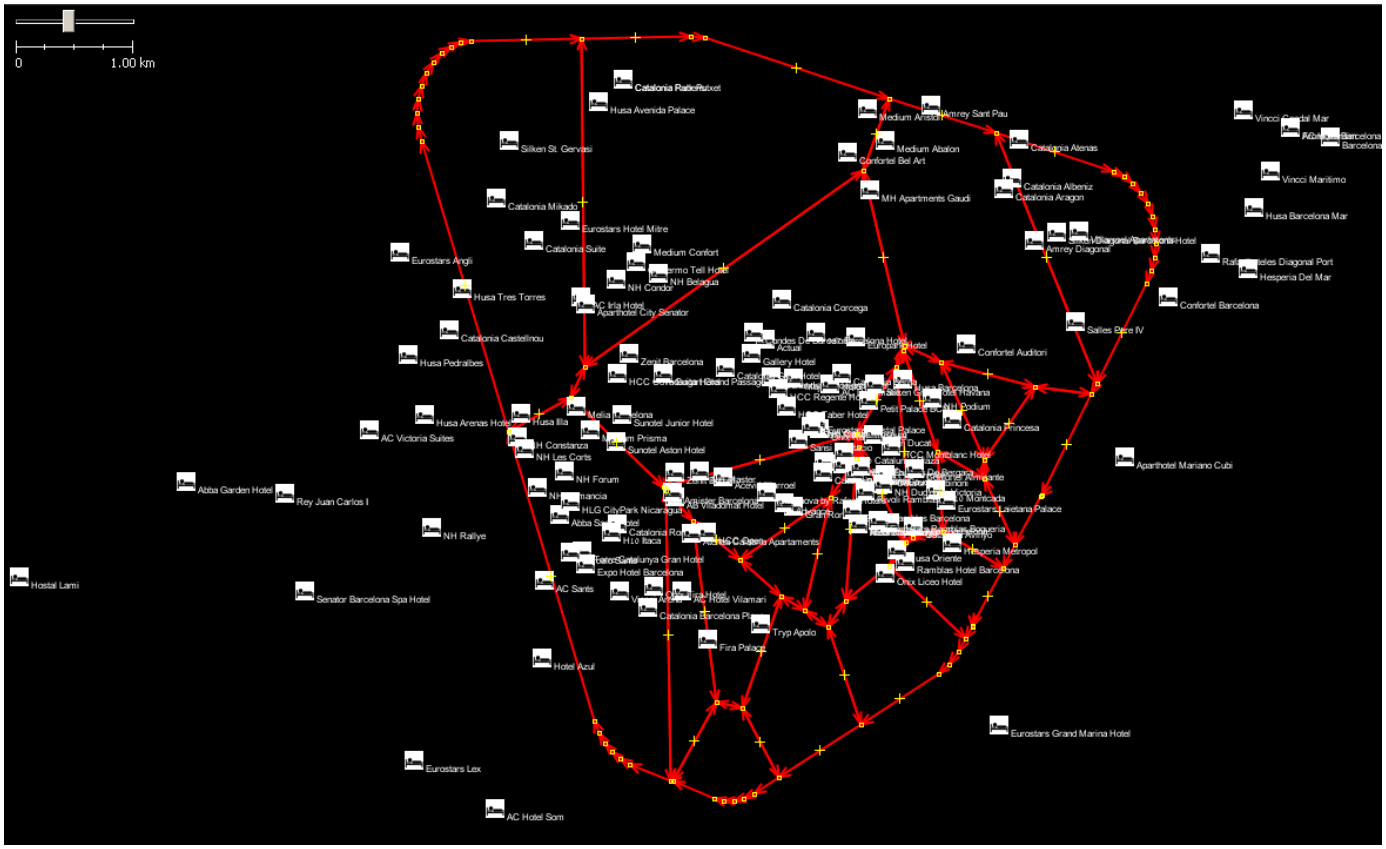

Figure 3: Spatial Model of Museums using Voronoi Tesselation: popularity of a hotel's location can be measured by the size of the (museum) region it is located

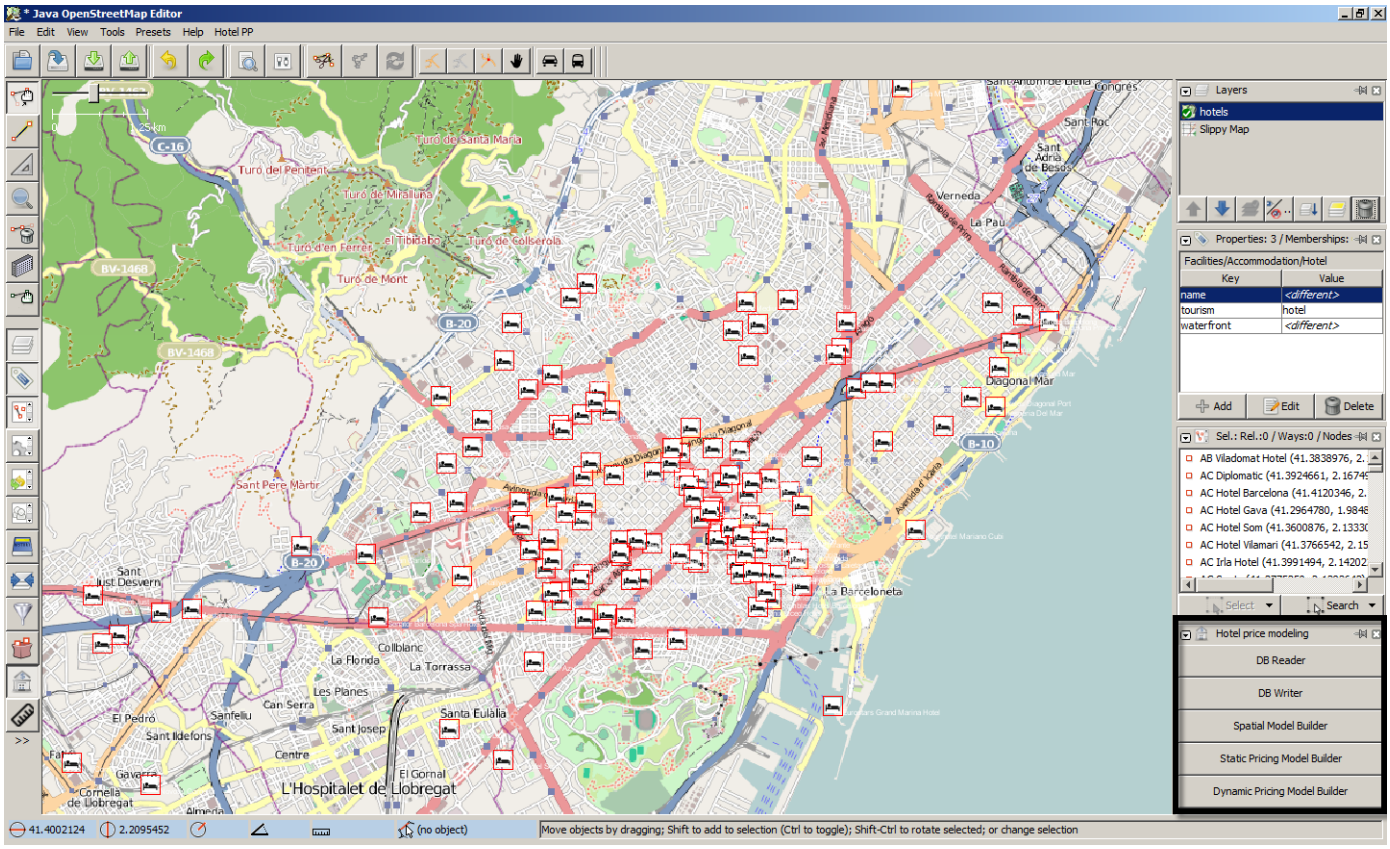

Figure 5: JOSM main view

teristics of the feature. The majority of widely used attributes are officially accepted, while some attributes can be used internally by an application. JOSM differentiates between types of features and attaches a specific icon to a feature that was recognized. This is extremely helpful when the user prepares the data for modeling since various types of data will be depicted by different icons. This facilitates data management. For example, hotels are tagged by a key named *tourism* with the value *hotel*, while restaurants are tagged by a key named *amenity* and a value *restaurant*[6]. An example of how hotels are represented in JOSM can be seen in Figure 5. We have introduced our own attribute *waterfront*, which is assigned to a hotel by the domain expert when a hotel is near a waterfront.

We have designed two components for easy data integration: the data reader and data writer. The data reader component consists of three parts: (1) database connection, (2) layer selection, and (3) data type selection. The database connection component allows the user to connect to the database and to select and read from the corresponding database table. The layer selection enables the user to select the existing layer or to create a new layer where the data will be read such that different types of data can be positioned in different layers. With the data type selection part, the user selects one of three types of data supported by the system. (1) general points, consisting of any data that has longitude and latitude coordinates, are used for importing any external spatial data into the system. (2) OSM points is similar to general points but this data contains an additional field for attributes in a *key=value* form that can be represented by JOSM. (3) Spatial model data is the areal data that consists of polygons and is created by a spatial model builder component (Section 6.1.2). After the general spatial data is read and presented in one of the layers, the user can annotate it with the official or custom attributes thus turning the general data into the form recognizable by JOSM. It is also at this stage that a specific attribute, recognizable by our system, can be attached to the data (e.g., *waterfront* property).

With the data writer component, the user can write the data back to the table. The data is read from the currently active layer. First, the user selects the database. The data can be written to an already existing table or to a new table by providing a name of a table. The user can also provide the description of the table that will be stored along with the data facilitating the search for a specific table in the database. Additional functionality is available for the table management in order to delete an existing table or the contents in an existing table.

[6]For a complete list of official attributes please see `http://wiki.openstreetmap.org/wiki/Map_Features`

6.1.2. Spatial Model Builder

In order to generate a spatial model using a spatial model builder component, the following steps are taken: (1) the user selects the database and the source table where the point-based data is located (points of interests, museums, historical places, etc.). (2) The user provides the name of the model table where the spatial model will be stored. We decided to simplify the process of spatial model creation by combining a model generation and table writer in one step. To achieve this, we call up the database stored procedure that invokes the spatial model creation algorithm in R framework using PL/R procedural language for PostgreSQL[7]. When the model is generated, it is written directly to a table provided in the spatial model builder component. The generated model is stored as a collection of polygons in a spatially-enabled PostgreSQL database.

6.1.3. Price Modeling

The price modeling components shown in Figure 6 (static model) and 7 (dynamic model) are the most important components available for the analyst. They allow the analyst to select the hotel features that will build up the pricing model for the static model or to select hotels and search dates in the dynamic model.

The static price modeling component consists of eight parts. First, the analyst connects to the database (this part is labeled as 1) that holds all the required information about hotels, prices, amenities, facilities, and spatial models. Second, the analyst retrieves the list of hotels he/she is interested in (labeled as 2) and selects the hotels that would be part of a model and hotels that will be used for price estimation (and which will not be part of a model). Parts 3 and 6 are responsible for retrieving the amenities and facilities of the selected hotels. The analyst has complete control over the final list of amenities and facilities that will be included in the model. If the hotel category (stars) is important for the model, the analyst uses part 4 to control this. Part 5 is called *Point and Spatial Model* and is the most versatile in the whole price modeling component. The analyst selects the spatial characteristics using two types of data: the point data that was used for generating the spatial model as explained in Section 6.1.2 and the spatial models stored in the corresponding tables. Next, the analyst selects the desired radius size(s). The definition of the radius size allows the analyst to answer such questions as: *How many points of interest/museums/bus stops are in a radius of 200 meters around the hotel.* The hotel density in the specified radius can also be calculated. In part 7, the analyst retrieves the hotel prices and specifies the period for which the pricing model has to be built.

The dynamic price modeling component consists of three parts (Figure 7). The first part (labeled as 1), the database control, is similar in functionality to that of the component of the static price model. The information about hotels and available search dates are represented in the part labeled as 2. In order to include a hotel into a test set, the domain expert should select at least one search date from the list. Hotels for which no search date was selected are automatically assigned to a training set. Hotels for which the search date was partially selected will also be included in the training set with the search dates that were not included in the test set.

In both components, the domain expert saves the training and test sets (if provided) in files (labeled as 8 in Figure 6 and 3 in Figure 7) with the format recognized by the data mining package embedded into the system (see below).

6.2. Weka

As the analytical component in our DSS, we integrated Weka [49], a free and open source data mining and machine learning framework. Weka supports different data mining goals (e.g. classification, clustering, regression) [50] and includes a vast collection of machine learning algorithms. Of these the most important for our task are those that perform regression estimation. In addition, some of the data mining algorithms handle missing values in the data. Since the dynamic model presented in Section 5.2 contains missing values for some of the searching dates, it is extremely important that we have algorithms that support missing values at our disposal. In cases in which the desired algorithm does not support missing values, Weka provides filters that replace missing values with means and modes.

Weka also supports several file formats for loading data as well as data loading from the database. However, the most common approach to loading data is through the ARFF column-base file format[8], that includes the data and describes the attributes and their data types in plain text. As was mentioned in Section 6.1.3, the characteristics selected in the static model or the hotel prices selected in the dynamic model are saved into file(s) in ARFF format. We found this way of data interchange more flexible than other options such as in-memory or database storage. There were several reasons for this choice. First, the domain expert has more control over the models that he/she produces. Second, since several evaluations with different models or parameters will usually be required, in-memory data interchange is not effective because the data will be lost when a new experiment is started. Third, it is easier to manage a file system than a database, and consequently, less integration and coding effort are required for embedding Weka.

[7] http://www.joeconway.com/plr/

[8] http://www.cs.waikato.ac.nz/ ml/weka/arff.html

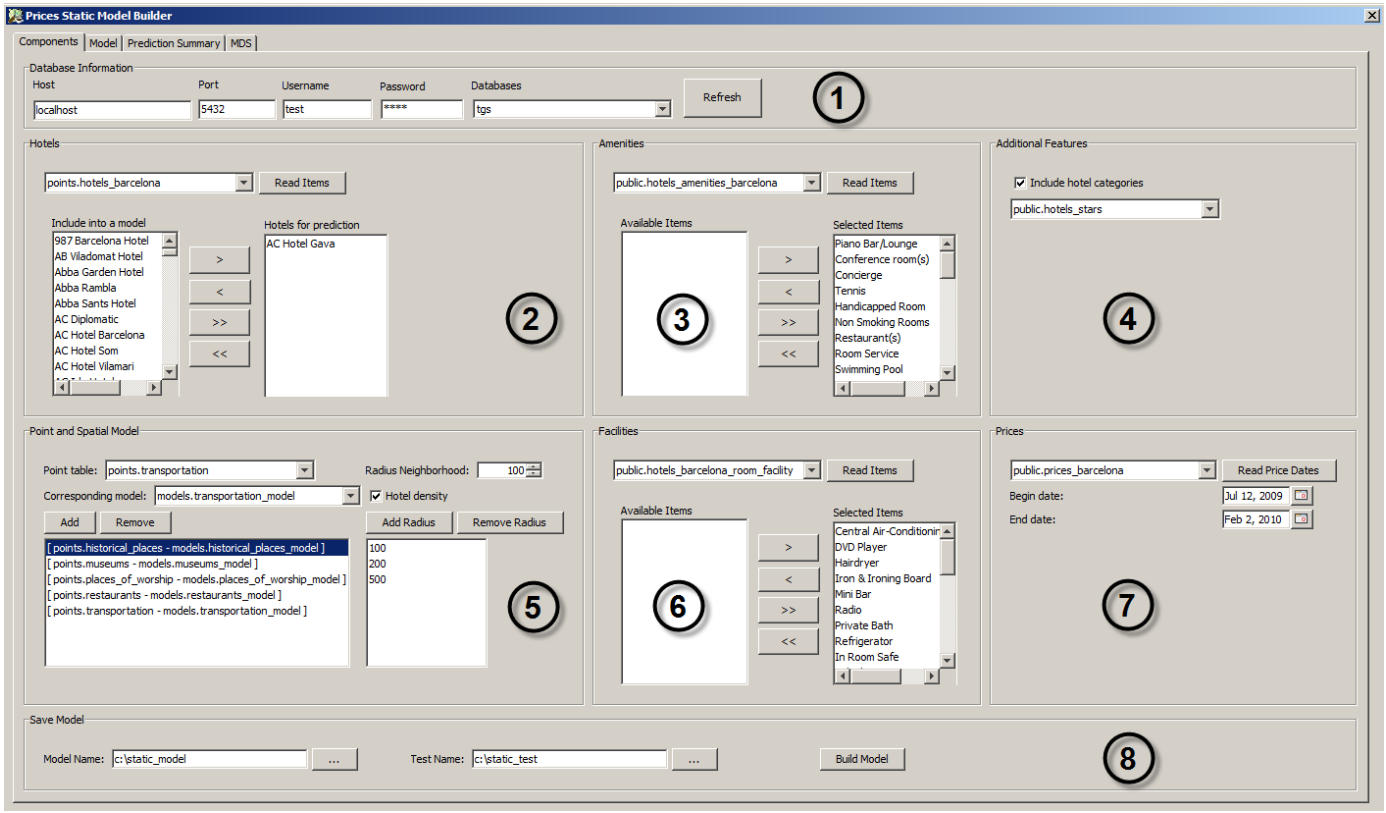

Figure 6: Components of the static model: hotels, amenities, facilities, spatial components and prices

6.3. Multidimensional Scaling

In order to facilitate the exploratory analysis of the hotel data, we implemented multidimensional scaling (MDS) [51], a powerful technique for investigating multivariate data by transforming the multidimensional data into two dimensions and then preserving the relative distance between objects (hotels in our case). Using graphical representation, MDS enables the analyst to observe the similarities of objects. Consequently, with MDS, the analyst can determine which hotels are more similar to each other in terms of their characteristics and also compare their average relative market price. The component inputs data in two modes: it can read the static model that was previously generated and stored in the file system or read the data currently loaded in Weka. An example presented in Figure 8 shows the relative similarities between hotels using the characteristics of the static model.

6.4. Graphs

Weka outputs the results of evaluation into the result window. While the information in the result window is comprehensive and includes test results, error measurements and various statistics, it is not intuitive for a non-expert. Moreover, there is a lack in flexibility in aggregating the results of repeated evaluations in a way that allows comparison between previous evaluations. In addition, there is no support for visualization of time series data as in the case of dynamic price estimation. As a result of these drawbacks, we enriched the visualization step with two components. The first component displays the results of evaluations as presented in Figure 9. This component stores the name of the classifier (i.e., algorithm) that was used along with the selected parameters: the hotel ID and name that was evaluated; its original price (actual price); the price predicted by the algorithm; correlation coefficient and error estimates (mean absolute error and root mean squared error). The second component visualizes the results of the time series price estimation using the dynamic model as presented in Figure 11 (a detailed explanation is given in Section 7). The visualization was implemented using JFreeChart chart library[9] and supports interactive zooming.

7. System Evaluation

The goal of this section is to demonstrate a possible scenario in which the system is applied. In general, the

[9] http://www.jfree.org/

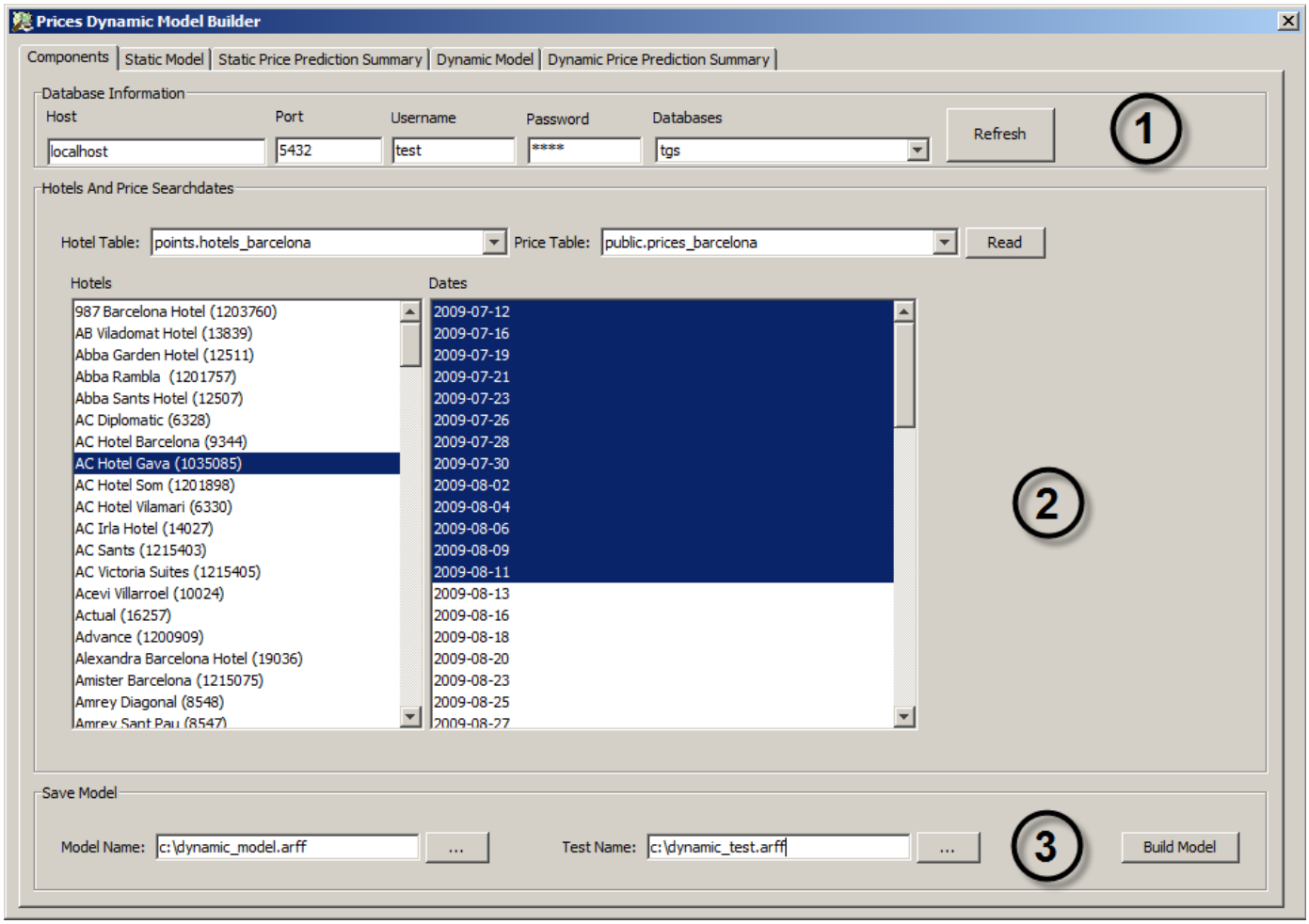

Figure 7: Components of the dynamic model: hotels and prices

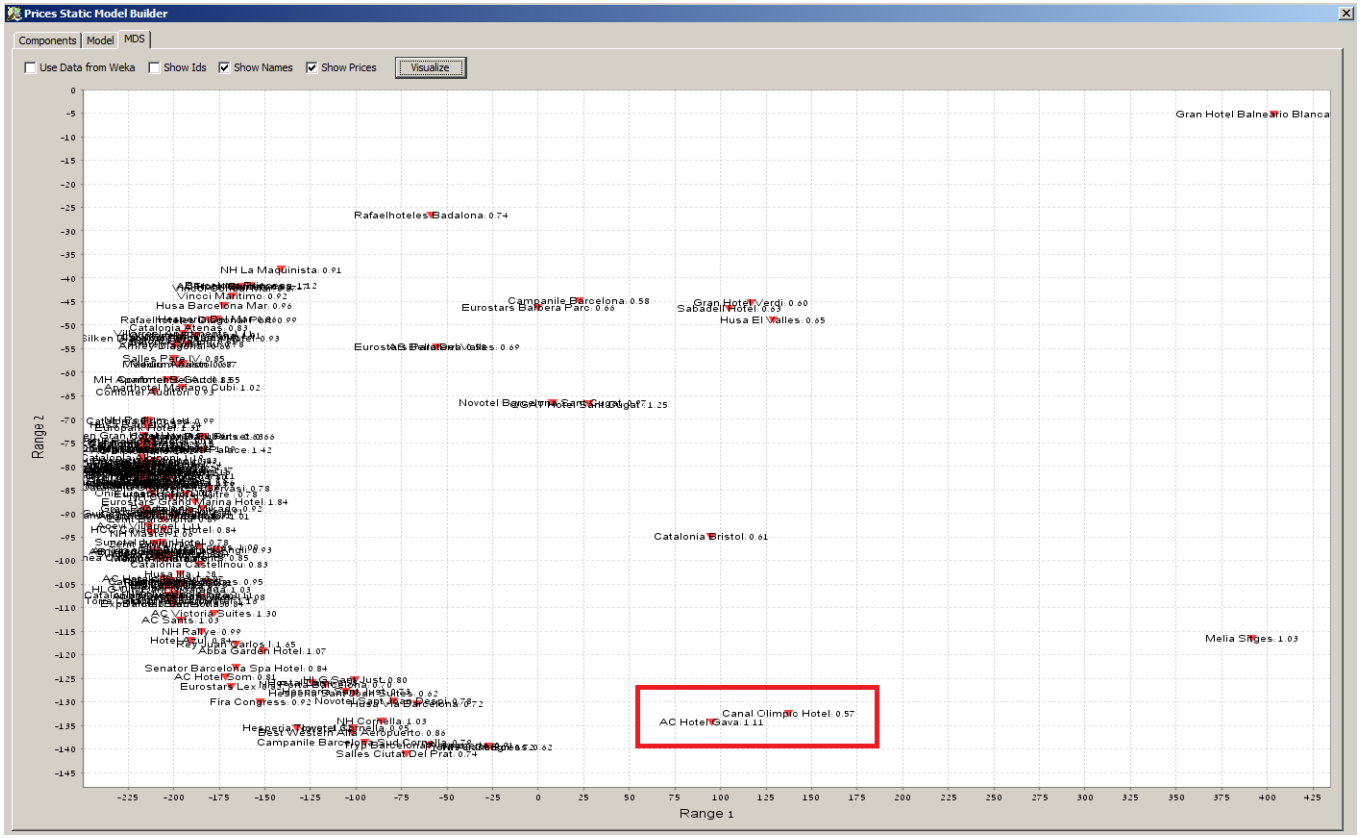

Figure 8: Multidimensional scaling using hotel characteristics

Classifier	Params	Hotel Id	Hotel Name	Actual Price	Predicted Price	Correlation Coefficient	Mean Absolute Error	Root Mean Squared Error
weka.classifiers.meta.AdditiveRegression	S 1.0 -I 10 -W weka.classifiers.functions.IsotonicRegression	1035085	AC Hotel Gava	144.309244	89.99791299…	1.0	0.2227	0.2956
weka.classifiers.meta.AdditiveRegression	S 1.0 -I 10 -W weka.classifiers.functions.IsotonicRegression	1214697	Canal Olimpic Hotel	75.0199679…	78.666772	1.0	0.2227	0.2956
weka.classifiers.functions.LibSVM	S 3 -K 2 -D 3 -G 0.0 -R 0.0 -N 0.5 -M 40.0 -C 1.0 -E 0.0010 -P 0.1	1035085	AC Hotel Gava	144.309244	121.12599	-1.0	0.2723	0.2883
weka.classifiers.functions.LibSVM	S 3 -K 2 -D 3 -G 0.0 -R 0.0 -N 0.5 -M 40.0 -C 1.0 -E 0.0010 -P 0.1	1214697	Canal Olimpic Hotel	75.0199679…	122.8191489…	-1.0	0.2723	0.2883
weka.classifiers.lazy.LWL	U 0 -K -1 -A \"weka.core.neighboursearch.LinearNNSearch -A \"weka.core.EuclideanDistance -…	1035085	AC Hotel Gava	144.309244	110.3158209…	1.0	0.2648	0.2648
weka.classifiers.lazy.LWL	U 0 -K -1 -A \"weka.core.neighboursearch.LinearNNSearch -A \"weka.core.EuclideanDistance -…	1214697	Canal Olimpic Hotel	75.0199679…	110.0553349…	1.0	0.2648	0.2648
weka.classifiers.functions.MultilayerPerceptron	L 0.3 -M 0.2 -N 500 -V 0 -S 0 -E 20 -H a	1035085	AC Hotel Gava	144.309244	142.3555989…	1.0	0.0415	0.0493
weka.classifiers.functions.MultilayerPerceptron	L 0.3 -M 0.2 -N 500 -V 0 -S 0 -E 20 -H a	1214697	Canal Olimpic Hotel	75.0199679…	83.876492	1.0	0.0415	0.0493
weka.classifiers.functions.IsotonicRegression		1035085	AC Hotel Gava	144.309244	102.631484	0.0	0.266	0.2715
weka.classifiers.functions.IsotonicRegression		1214697	Canal Olimpic Hotel	75.0199679…	102.631484	0.0	0.266	0.2715
weka.classifiers.functions.LinearRegression	S 0 -R 1.0E-8	1035085	AC Hotel Gava	144.309244	105.8875589…	1.0	0.2577	0.2604
weka.classifiers.functions.LinearRegression	S 0 -R 1.0E-8	1214697	Canal Olimpic Hotel	75.0199679…	103.673428	1.0	0.2577	0.2604

Figure 9: Aggregated evaluation results: classifier, parameters, hotel ID, hotel name, actual price, predicted price, correlation coefficient, mean absolute error, root mean squared error.

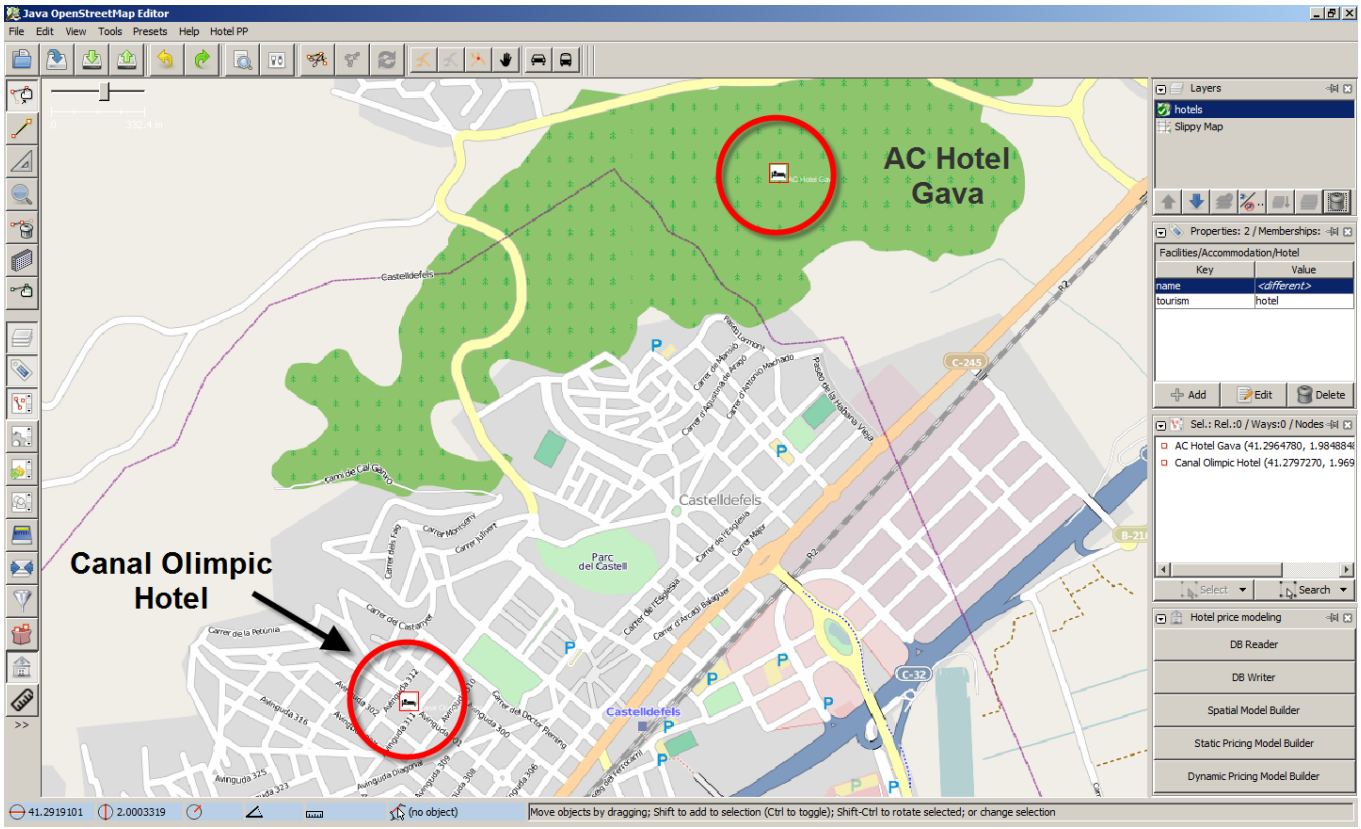

Figure 10: Geographical location of hotels under investigation

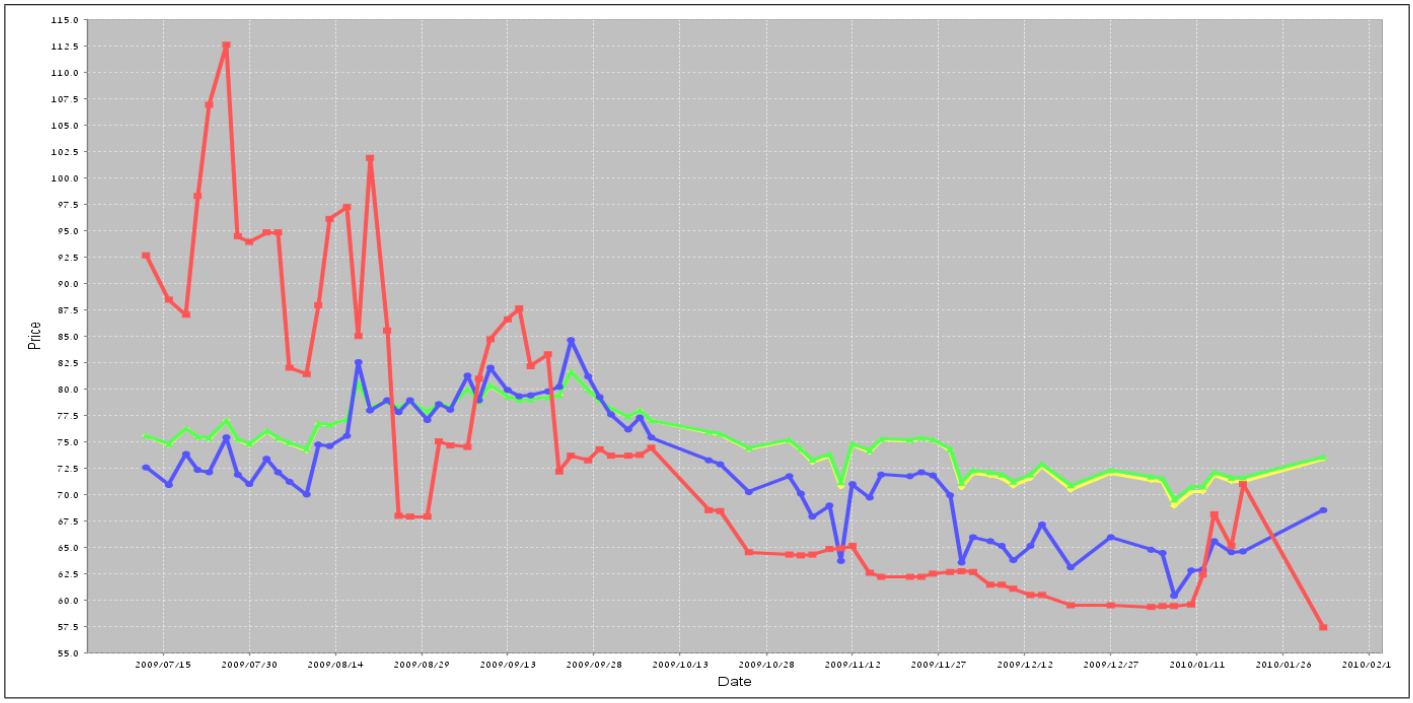

Figure 11: Price prediction for Canal Olimpic Hotel. Red line - real price; blue line - predicted price using dynamic model (LibSVM nu-SVR); green line - arithmetic average of the dynamic and static price components; yellow line - geometric mean of the dynamic and static price components

task of the domain expert is to estimate the price of a hotel under investigation utilizing any of the available algorithms provided in the data mining framework. However, since the predictive power of algorithms is data dependent, it is not possible to suggest a priori one of the algorithms. Therefore, we apply common steps from data mining methodology to analyze the hotel prices of selected hotels: exploratory analysis; a model evaluation and verification using cross-validation; and testing [52, 53, 54].

The system evaluation was performed with the cooperation of a company that develops travel technology solutions, in particular inventory management and pricing solutions for many well-known websites and travel agencies around the globe. The company provided us with a real database of hotels and their pricing for a period of three years. An expert who works for the company and is the primary user of the proposed system, participated in the evaluation process. For the demonstration, we used data from 168 hotels in Barcelona, Spain.

Let us assume that the domain expert wishes to explore hotels in Barcelona using the corporate database of hotels that are already contracted with the company. First, the hotels are visualized in the GIS as depicted in Figure 5. This allows the domain expert to see the spatial distribution of hotels in the area of Barcelona. In the next step, he/she acquires the visual representation of similarities between hotels in terms of their characteristics (amenities, facilities, locational attributes) by applying multidimensional scaling on the selected hotels (Figure 8).

Two hotels (Canal Olimpic Hotel, 3-stars, and AC Hotel Gava, 4 stars) outlined in red rectangle especially attract his/her attention because they are situated relatively close to each other, but far enough from the majority of the other hotels. This observation suggest that these two hotels share many characteristics. The two hotels are also located close to each other geographically (the straight line distance is 2.2 km) as shown in Figure 10. Next, the domain expert decides to analyze the prices of the two hotels. In order to do so, a model with information about these hotels must be created. The domain expert consequently selects 166 hotels in the region of Barcelona and creates two training sets: one for the static model using the control panel as shown in Figure 6 and the other for the dynamic model utilizing the control panel as shown in Figure 7. The dynamic model contains 10,250 records that correspond to different dates for which the customers searched for a room rate in any of 166 hotels selected in the training set. The two hotels under investigation are included in the test set and will be used when the models are being evaluated.

In order to estimate the predictive power of the model and to select the best algorithm that minimizes the error between actual and predicted prices, we used a 10-fold cross validation of the training set. Seven regression algorithms were used including linear, non-linear regression and non-parametric ones. Table 2 shows the results of the evaluation for static and dynamic models using Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) criteria. The smaller values of MAE and RMSE indicate a smaller error between the actual and the predicted prices. The predictive performance of the *additive regression with isotonic regression* is the best among other algorithms when applied on the static model and improves on 32.7% (MAE) and 29% (RMSE) compared to the worst case of Locally Weighted Learning with Linear Regression. In the case of the dynamic model, LibSVM nu-SVR and Locally Weighted Learning with Linear Regression, both yield the least MAE error of 0.1229, which improves the worst case with Multilayer Perceptron by more than 25.4%. Additive regression with isotonic regression yields the best results in terms of RMSE and improves the worse case with Multilayer Perceptron by more than 17.9%. Judging by the results of the prediction error, the domain expert is likely to select additive regression with isotonic regression for predicting the prices of the hotels in the static model and Locally Weighted Learning with Linear Regression for predicting the prices of the hotels in the dynamic model. However, important here are also runtime considerations. The difference between the running time of LibSVM nu-SVR (5.7 sec) and the additive regression with isotonic regression (1087.9 sec) on the static model using 10-fold cross-validation is 18 min, while the performance increase of additive regression with isotonic regression is only 12.4%. Likewise, the running time of the Multilayer Perceptron using 10-fold cross validation on the dynamic model is 9.1 hours, while the running time of LibSVM nu-SVR is 71.8 minutes.

In order to estimate the prices of two hotels, the domain expert applies the static model built using additive regression with isotonic regression on the test set that includes the two hotels. The average price of a room in AC Hotel Gava between July 2009 and February 2010 is €144.31. Using the hedonic model this room price is estimated at €90. The opposite price trend is shown for the Canal Olimpic Hotel where, instead of an average room price of €75, additive regression with isotonic regression estimates its price as €78.67, which decreases the price difference between 3-star and 4-star hotels from €69.31 ($144.31 - 75$) to only €11.33 ($90 - 78.67$). Next, the domain expert is interested in estimating hotel prices for each day for which customers performed a room search. In this case, the algorithm with the best predictive performance on the dynamic model, the LibSVM nu-SVR, was selected and applied on the test set.

After evaluating the dynamic model on the test set, estimated prices are presented for each hotel separately. Let us focus on the Canal Olimpic Hotel, whose real price was significantly lower than the price estimated by the static model. Figure 11 presents four price trends for the Canal Olimpic Hotel. The original average price for each search date is shown by the red line; the prices estimated using the dynamic model only are represented by the blue line. The effect on the price estimation when combining price estimation using static and dynamic models, are rep-

resented by the green and yellow lines. We used two approaches to combine the price estimation of the two models - arithmetic average (green line) and geometric mean (yellow line). In the case of the Canal Olimpic Hotel, the graph shows that while its price was relatively high in July 2009, the price decreased considerably afterwards. Both the estimated prices using the dynamic model alone or in combination with the static model show that the price of this hotel is somewhat underpriced.

8. Discussion

The proposed decision support system has three essential features. It uses: (1) JOSM, a GIS-based framework that was initially designed to support the very narrow task of creating and editing OpenStreetMap data; (2) OpenStreetMap data as an external data source in the process of determining hotel prices, and (3) a data mining framework instead of pure statistical approaches for price analysis. The advantage of using JOSM over other general purpose GIS tools was discussed in Section 6.1. However, the other two features require further discussion.

Since OpenStreetMap data retrieval is naturally supported by JOSM, it simplifies the process of data acquisition. In comparison, [9] applied a complex process of data collection. The authors used Virtual Earth Interactive SDK to measure the number of restaurants and shopping destinations that were in proximity to the hotels. To answer the question whether the hotel was located near the beach, [9] used image classification of satellite data and manually validated the results by using on-demand human annotators through the Amazon Mechanical Turk[10] paid service. Apart from the considerable degree of effort involved in implementing the task, the solution is by no means scalable and hardly replicable. While such a solution maybe considered as creative and able to fulfill research needs, it is clearly not applicable in real world situations, which is the primary goal of our research. The advantages of our approach are obvious. First, OpenStreetMap data is readily available and has a great deal of content. It contains information about transportation such as buses and trains, points of interest, restaurants and pubs, places of worship and historical sites. These elements are very useful since they are determining factors in the modeling of hotel prices. Second, the spatial data can be displayed in the system such that the analyst can decide what parts are relevant for the analysis and what data should be included into the model. Third, the absence of some functionality such as determining whether the hotel is located near a waterfront, is substituted by the domain expert himself without the need for applying costly image classification methods and expensive human annotators. However, the completeness and correctness of the OpenStreetMap data must still be closely examined because the project was only recently established and data is contributed by volunteers. There is much concern in the research over the credibility and completeness of volunteered geographic information [55].

A study [56] conducted on German data showed that there is a difference in terms of data completeness between cities and rural areas. However, the difference has decreased substantially in recent years due to the increase in new members willing to participate in the project (the number of participants doubled within one year and stands at over 200,000 members in January 2010). Moreover, the data in large cities is rich enough. In a recent study on OpenStreetMap coverage in England [57], it was shown that OpenStreetMap covers 65% of the area of England. As in the German study [56], the coverage is better in urban areas. It was shown in [58] that OpenStreetMap is quite accurate and comparable to geographical information produced by commercial companies. Moreover, OpenStreetMap data has been already used in place of proprietary and commercial datasets [56].

The advantage of using data mining over pure statistical analysis is explained by the type of problem we deal with. Statistical analysis usually deals with well-structured problems, small data sets, data integrity and a confirmatory type of analysis [59]. Moreover, statistical analysis depends on many assumptions, like normality, independence, homogeneity, that should be met prior to applying statistical methods. Confirmatory analysis implies "clean room" experiments with careful testing of each of the underlying model parameters using different statistical criteria (e.g. significance level, R-square fitting), which is hardly achievable in real life scenarios. Recall from Section 1 and 5, the problem of hotel price estimation is an ill-structured problem with different types of data (spatial and non-spatial), amount of input parameters, and missing values. Here, the use of heterogeneous data and exploratory analysis using different algorithms for price estimation are more appropriate. This is also due to the fact that data mining approaches can handle high-dimensional data with a high degree of sparseness, multicollinearity, outliers, and missing values, which statistical approaches cannot easily handle [60].

9. Conclusion

This paper described the problem brokerage companies face in the hotel business. The competition and revenue issues are pushing these companies towards developing non-standard solutions. We presented a practical approach for implementing a GIS-based decision support system to analyze hotel value and estimate objective room rates. We proposed two types of models. The first static model is based on hedonic pricing theory and composed of intrinsic hotel characteristics (e.g. amenities, facilities) and various locational characteristics (e.g. museums, restaurants around a hotel, etc). The second dynamic model contains

[10] http://www.mturk.com/

historical hotel room rates. We discussed in detail the requirements and components of a decision support system that is designed to be used in real business scenarios. We showed that the solution can be considerably simplified by using free and open source tools such as the Java OpenStreetMap Editor (JOSM), R statistical package and the Weka data mining framework. We also simplified the process of external spatial data acquisition by using OpenStreetMap data.

The effectiveness of the tool can only be assessed if it is really used by domain experts to improve their decision making and if it attains real (monetary) results. We developed the system by closely following the guidelines and suggestions from the top management at TGS. We consulted domain experts working at TGS and acquired a thorough understanding of their needs. We therefore hope that the system meets their expectations.

In our future work, we plan to enhance the system with different analytical components. We also intend to closely work with the hotel domain experts to identify problems that have not been yet covered by the current prototype.

Acknowledgements

This work was partially funded by the German Research Society (DFG) under grant GK-1042 (Research Training Group "Explorative Analysis and Visualization of Large Information Spaces"), and by the Priority Program (SPP) 1335 ("Visual Spatio-temporal Pattern Analysis of Movement and Event Data"). The authors wish to thank Dana Hendelsman and Maya Elman for their help in system implementation.

References

[1] C. Park, Y. Kim, Identifying key factors affecting consumer purchase behavior in an online shopping context, International Journal of Retail & Distribution Management 31 (1) (2003) 16–29.

[2] D. Gefen, E. Karahanna, D. Straub, Trust and TAM in online shopping: An integrated model, Mis Quarterly (2003) 51–90.

[3] W. Kim, D. Kim, Factors affecting online hotel reservation intention between online and non-online customers, International Journal of Hospitality Management 23 (4) (2004) 381–395.

[4] H. Van der Heijden, Factors influencing the usage of websites: the case of a generic portal in The Netherlands, Information & Management 40 (6) (2003) 541–549.

[5] C. Flavián, M. Guinalíu, R. Gurrea, The role played by perceived usability, satisfaction and consumer trust on website loyalty, Information & Management 43 (1) (2006) 1–14.

[6] S. Rosen, Hedonic prices and implicit markets: product differentiation in pure competition, The Journal of Political Economy 82 (1) (1974) 34–55.

[7] B. Monty, M. Skidmore, Hedonic pricing and willingness to pay for bed and breakfast amenities in Southeast Wisconsin, Journal of Travel Research 42 (2) (2003) 195.

[8] C. Thrane, Examining the determinants of room rates for hotels in capital cities: The Oslo experience, Revenue & Pricing Management 5 (4) (2007) 315–323.

[9] B. Li, A. Ghose, P. G. Ipeirotis, Stay elsewhere? improving local search for hotels using econometric modeling and image classification, in: 11th International Workshop on Web and Databases (WebDB), 2008.

[10] W. Hung, J. Shang, F. Wang, Pricing determinants in the hotel industry: Quantile regression analysis, Hospitality Management 29 (3) (2010) 378–384.

[11] C. Chen, R. Rothschild, An application of hedonic pricing analysis to the case of hotel rooms in Taipei, Tourism Economics 16 (3) (2010) 685–694.

[12] S. Lee, S. Jang, Room Rates of US Airport Hotels: Examining the Dual Effects of Proximities, Journal of Travel Research.

[13] R. Butler, The specification of hedonic indexes for urban housing, Land Economics 58 (1) (1982) 96–108.

[14] S. Sirmans, D. Macpherson, E. Zietz, The composition of hedonic pricing models, Journal of Real Estate Literature 13 (1) (2005) 1–44.

[15] J. Shim, M. Warkentin, J. Courtney, D. Power, R. Sharda, C. Carlsson, Past, present, and future of decision support technology, Decision support systems 33 (2) (2002) 111–126.

[16] D. Arnott, G. Pervan, A critical analysis of decision support systems research, Journal of Information Technology 20 (2) (2005) 67–87.

[17] N. Karacapilidis, An overview of future challenges of decision support technologies, Intelligent Decision-making Support Systems (2006) 385–399.

[18] M. Crossland, B. Wynne, W. Perkins, Spatial decision support systems: An overview of technology and a test of efficacy, Decision Support Systems 14 (3) (1995) 219–235.

[19] P. Longley, G. Higgs, D. Martin, A GIS-based appraisal of council tax valuations, Journal of Property Valuation and Investment 11 (4) (1993) 375–383.

[20] D. Fung, H. Kung, M. Barber, The application of GIS to mapping real estate values, Appraisal Journal 63 (1995) 445–445.

[21] M. Rodriguez, C. Sirmans, A. Marks, Using geographic information systems to improve real estate analysis, Journal of Real Estate Research 10 (2) (1995) 163–173.

[22] P. Wyatt, The development of a GIS-based property information system for real estate valuation, International Journal of Geographical Information Science 11 (5) (1997) 435–450.

[23] W. McCluskey, W. Deddis, A. Mannis, D. McBurney, R. Borst, Interactive application of computer assisted mass appraisal and geographic information systems, Journal of Property Valuation and Investment 15 (5) (1997) 448–465.

[24] G. Thrall, GIS applications in real estate and related industries, Journal of Housing Research 9 (1) (1998) 33–59.

[25] G. Castle, R. Joseph, GIS in real estate: Integrating, analyzing, and presenting locational information, Appraisal Institute, 1998.

[26] A. Din, M. Hoesli, A. Bender, Environmental variables and real estate prices, Urban Studies 38 (11) (2001) 1989.

[27] A. Sarip, Integrating Artificial Neural Networks and GIS for single-property valuation, in: Eleventh-PRRES Conference. Pacific Rim Real Estate Society, Melbourne, Citeseer, 2005.

[28] E. Natividade-Jesus, J. Coutinho-Rodrigues, C. Antunes, A multicriteria decision support system for housing evaluation, Decision Support Systems 43 (3) (2007) 779–790.

[29] M. Kaboudan, A. Sarkar, Forecasting prices of single family homes using GIS-defined neighborhoods, Journal of Geographical Systems 10 (1) (2008) 23–45.

[30] N. García, M. Gámez, E. Alfaro, ANN+ GIS: An automated system for property valuation, Neurocomputing 71 (4-6) (2008) 733–742.

[31] R. Denzer, Generic integration of environmental decision support systems-state-of-the-art, Environmental Modelling & Software 20 (10) (2005) 1217–1223.

[32] P. Densham, Spatial decision support systems, Geographical information systems: Principles and applications 1 (1991) 403–412.

[33] M. Haklay, P. Weber, OpenStreetMap: user-generated street maps, IEEE Pervasive Computing (2008) 12–18.

[34] C. Martins-Filho, O. Bin, Estimation of hedonic price functions via additive nonparametric regression, Empirical Economics 30 (1) (2005) 93–114.

[35] A. Bull, Pricing a motels location, International Journal of Con-

temporary Hospitality Management 6 (6) (1994) 10–15.
[36] A. Israeli, Star rating and corporate affiliation: their influence on room price and performance of hotels in Israel, International Journal of Hospitality Management 21 (4) (2002) 405–424.
[37] E. Worzala, M. Lenk, A. Silva, An exploration of neural networks and its application to real estate valuation, Journal of Real Estate Research 10 (2) (1995) 185–201.
[38] S. McGreal, A. Adair, D. McBurney, D. Patterson, Neural networks: the prediction of residential values, Journal of Property Valuation and Investment 16 (1) (1998) 57–70.
[39] J. Zurada, A. Levitan, J. Guan, Non-conventional approaches to property value assessment, Journal of Applied Business Research 22 (3).
[40] V. Limsombunchai, C. Gan, M. Lee, House price prediction: Hedonic price model vs. artificial neural network, American Journal of Applied Sciences 1 (3) (2004) 193–201.
[41] C. Bitter, G. Mulligan, S. Dallerba, Incorporating spatial variation in housing attribute prices: a comparison of geographically weighted regression and the spatial expansion method, Journal of Geographical Systems 9 (1) (2007) 7–27.
[42] A. Fotheringham, C. Brunsdon, M. Charlton, Geographically weighted regression: the analysis of spatially varying relationships, John Wiley & Sons Inc, 2002.
[43] M. Löchl, K. Axhausen, Modelling hedonic residential rents for land use and transport simulation while considering spatial effects, Journal of Transport and Land Use 3 (2) (2010) 39–63.
[44] A. Okabe, B. Boots, K. Sugihara, S. Chiu, Spatial tessellations: Concepts and applications of Voronoi diagrams, New York: John Wiley & Sons, 2000.
[45] R. Sugumaran, J. Degroote, Spatial Decision Support Systems: Principles and Practices, CRC Press, 2010.
[46] S. Liu, A. Duffy, R. Whitfield, I. Boyle, Integration of decision support systems to improve decision support performance, Knowledge and Information Systems 22 (3) (2010) 261–286.
[47] T. O'Reilly, Lessons from open-source software development, Communications of the ACM 42 (1999) 32–37.
[48] G. Andrienko, N. Andrienko, P. Jankowski, D. Keim, M. Kraak, A. MacEachren, S. Wrobel, Geovisual analytics for spatial decision support: Setting the research agenda, International Journal of Geographical Information Science 21 (8) (2007) 839–858.
[49] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, I. Witten, The WEKA data mining software: An update, ACM SIGKDD Explorations Newsletter 11 (1) (2009) 10–18.
[50] O. Maimon, L. Rokach, Data Mining and Knowledge Discovery Handbook, Springer, 2010.
[51] J. Kruskal, M. Wish, Multidimensional scaling, Sage Publications, Inc, 1978.
[52] J. Shao, Linear model selection by cross-validation, Journal of the American Statistical Association (1993) 486–494.
[53] R. Kohavi, A study of cross-validation and bootstrap for accuracy estimation and model selection, in: International joint Conference on artificial intelligence, Vol. 14, Citeseer, 1995, pp. 1137–1145.
[54] X. Gao, Y. Asami, C. Chung, An empirical evaluation of spatial regression models, Computers & Geosciences 32 (8) (2006) 1040–1051.
[55] A. Flanagin, M. Metzger, The credibility of volunteered geographic information, GeoJournal 72 (3) (2008) 137–148.
[56] D. Zielstra, A. Zipf, A Comparative Study of Proprietary Geodata and Volunteered Geographic Information for Germany, in: 13th AGILE International Conference on Geographic Information Science., 2010.
[57] M. Haklay, C. Ellul, Completeness in volunteered geographical information–the evolution of OpenStreetMap coverage in England (2008-2009), Journal of Spatial Information Science (0) (2011) In–revision.
[58] M. Haklay, How good is volunteered geographical information? A comparative study of OpenStreetMap and Ordnance Survey datasets, Environment and Planning B: Planning and Design 37 (4) (2010) 682–703.
[59] D. Hand, Data mining: statistics and more?, The American Statistician 52 (2).
[60] D. Brusilovsky, E. Brusilovskiy, White paper: Data mining: The means to competitive advantage, `http://www.connectis.ca/download/bis/bis_data_mining_whitepaper.pdf` (2008).

Table 1: Complete list of hotel-related attributes available for analysis

Facilities	**Amenities**	**Other**
Air Condition	24-Hour Front Desk	Number of Rooms
Satellite TV	Babysitter Services	Hotel Category
Hairdryer	Baggage Hold	Hotel Name
Iron & Ironing Board	Barber/Beauty Salon	Standard Room Rate
Mini Bar	Breakfast Room	Waterfront (derived attribute)
Clock-Radio	Cafe	
Private Bath	Car Rental Desk	
Refrigerator	Children care/activities	
In Room Safe	Coffee Shop	
Telephone	Concierge	
Fully Equipped Kitchen	Conference room(s)	
Microwave	Currency Exchange	
Wake-Up Service	Dry Cleaning Service	
Internet Access	Elevator(s)	
CD- Stereo system	Free Newspaper	
Shower only	Game Room	
Trouser Press	Gift/Sundry Shop	
In-Room Pay Movies	Handicapped Room	
Shared Bath	Horse Back Riding	
Coffee/Tea Making Facilities	Interior Corridors	
Individual Climate Control	Laundry/Valet	
Work Desk	Limited Medical Services	
1 Bed and 1 Sofa Bed	Massage Treatments	
Wheelchair Accessible	Multilingual Staff	
Balcony	Non Smoking Rooms	
Hydromassage Bathtubs	Parking	
Living Room	Parking (Fee)	
Crib on Request-Fee May Apply	Piano Bar/Lounge	
Soundproof Room/Windows	Playground/Play Area	
	Pool Bar	
	Restaurant(s)	
	Room Service	
	Safe Deposit Box	
	Shuttle to Airport	
	Swimming Pool	
	Tour Desk	
	Wedding services	
	Wireless High Speed Internet	

Table 2: Static & Dynamic Model Evaluation using 10-fold cross validation of the training set

Algorithm	**Static Model**		**Dynamic Model**	
	MAE	**RMSE**	**MAE**	**RMSE**
Isotonic Regression	0.1567	0.2088	0.1393	0.2232
LibSVM epsilon-SVR	0.1534	0.2056	0.124	0.2198
LibSVM nu-SVR	0.1509	0.203	0.1229	0.2204
Linear Regression	0.1881	0.24	0.1301	0.2182
Locally Weighted Learning with Linear Regression	0.1964	0.2468	0.1229	0.2201
Additive Regression with Isotonic Regression	0.1322	0.1738	0.1256	0.2091
Multilayer Perceptron	0.1913	0.2449	0.1647	0.2548