International Journal of Applied Engineering Research ISSN 0973-4562 Volume 13, Number 6 (2018) pp. 3988-3998


# Decision Making for Hotel Selection using Rough Set Theory: A case study of Indian Hotels

**Haresh Kumar Sharma, Samarjit kar**

*Department of Mathematics, National Institute of Technology Durgapur, West Bengal 713 209, India.*

**Abstract**

The role of service quality to meet customer satisfaction is imperative for decisions and policy making by hotel authorities. This study presents a case study on developing strategies which will maximize the profit by enticing more tourists in hotels. In this study, we have applied rough set theory (RST) on data related to hotel industry and derived "$if\ \ldots then \ldots$." decision rules to classify the attributes of hotel. Proposed method can help decision makers to understand tourist behavior and improve the service quality of hotel industry. We have also performed some statistical analysis like predictive analysis and exploratory data analysis to analyze the effects of different factors which can influence the decision making of the concerned authorities.

**Keywords**: Rough set, hotel, Multi Criteria decision making, exploratory data analysis, *Sta*tistics

## INTRODUCTION

Over the past decades the tourism and hospitality industry is increasing quantity worldwide (Mohajerani and Miremadi, 2012). Indian tourism and hospitality industry has evolved into one of the major industries of development in the services sectors of India, which is one of the important industries. The tourism industry in India is a vital source of foreign exchange, earnings and employment opportunities. It contributes to 6.23% to the National Gross Domestic Product (GDP), annual growth rate of 14.12% of foreign exchange to the Indian economy and 8.78 % of its total jobs in the country. Sharma and Kalotra (2016) reported that India is one of the quickest growing Asian economies, which means that the Indian tourism industry may be predicted to grow unexpectedly within the coming years. In 2013, it was revealed that tourism and travel sectors contribute Rs 2.21 trillion (US$ 36.21 billion) or 2.3% of the Indian gross domestic product. Furthermore, The Tourism Competition Report (2013), published by the World Economic Forum, has confirmed that India is at the top 11th place in the Asia Pacific region and sixty fifth inside the world travel and tourism competitiveness index 2013.

Hotel management is important for the customer satisfaction, policy and decision makers. Hotel analysis has become not only an integral part of the travel decision-making and policy formulation but has also become essential to estimate its dynamic impact Ohlan (2012, 2016, 2017). Also, quality and relationships analysis can provide reliable and valuable information for the hotel industry in allocating resources and for future planning of the travel industry. Data mining and some appropriate statistical tools have an effective impact in the planning of all activities of the hotel industry and it also defines the relationship with other variables. Therefore, an empirical study of Indian hotel industry has been opted as a topic of this research to focus on the issue of hotel quality analysis.

In the recent times the hotel quality has been investigated in various literature because of their various implications in tourism and travel. A number of studies have evaluated the quality of hotels attribute by using exploratory data analysis such as principle component, factor analysis and descriptive statistics and multiple regression techniques by Lahap et al. (2016), Ren et al. (2016), Xu and Li (2016), Li et al (2017), Lai and Hitchcock (2017) and Patiar et al. (2017). Furthermore, Hua and Yang (2017) explores the systematic effects of crime on hotel operating performance based on data from a sample of 404 Houston hotels using econometric methods. The empirical evidence shows that crime incidents have a significantly negative impact on hotel operating performance. Lafuente et al. (2014) use the fuzzy Delphi approach and fuzzy analytic hierarchy process focuses on understanding the luxury resort hotel industry in Taiwan and Macao to create a system of evaluation criteria. Moreover, Wang et al. (2016) employed logistic regression to analyze data gathered from 140 hotels in Taiwan. The empirical analysis show that compatibility, company length, technology competence, and vital mass are appreciably positively related to mobile hotel reservation systems, while complexity is significantly negatively related to mobile hotel reservation systems. Also, Rajaguru and Rajesh (2016) explain the role of value for money and service quality in customer satisfaction using hierarchical regression analysis.

In addition, various studies have focused on the impact of service quality in the hotel industry (Mei et al, 1999, Tsang and Qu, 2000, Ekinci et al., 2003, Juwaheer, 2004, Mohsin, et al. 2005, Akbaba, 2006, Liou, 2009, Liou and Tzeng 2010, Akincilar and Dagdeviren 2014, Masiero et al. 2015, Li et al. (2016), Mardani et al. 2016, Rianthong 2016, Viglia et al. (2016) and Yu et al. 2016). Other researchers have also explored the relationship of hotel attributes. For example, Geetha et al. (2017) establish the relationship between customer ratings and customer sentiments for hotels. They found the consistency between actual customer feelings and customer ratings. In recent years, multi criteria decision-making approaches have been successfully applied in the service quality analysis of hotel industry. Benitez et al. (2007) evaluated the dynamical service quality of three hotels surveys in Gran Canaria island using a fuzzy multi-attribute decision-making method. Chen (2016) has explained the growth rate for inbound tourism market at Taiwan. In which on the basis of total tourist arrival the growth rate of sale and financial status of hotels has been calculated. However, the empirical literature which involves exploratory and predictive data analysis has some limitations related to statistical and distributional assumptions. Therefore, this study has adopted a new qualitative data mining rough set approach for discovering the relationships among dependent and independent variables. The main advantage of the rough set approach is due to the fact that the qualitative nature of the data being analyzed constructs it difficult for employing standard statistical methods Liou et al. (2016). Rough set theory has been consistently employed in a variety of research areas for the extraction of decision rules (Law and Au 1998, 2000, Au and Law 2000 and Goh, Law 2003 and Liou et al. 2016). Also, several variants of the rough set approach have also emerged in the literature by Goh et al. (2008), Lin (2010), Xiaoya and Zhiben (2011) and Celotto et al. (2012). Moreover, Li et al. (2011) analyzed and predicted tourism in Tangshan city of China through the rough set model. Golmohammadi and Ghareneh (2011) analyze the importance of travel attributes by rough set approach. Celotto et al. (2015) applied rough set theory to summarize tourist evaluations of a destination.

To the best our knowledge, hotel quality analysis not widely studied in the rough set literature. The current study fills these significant gaps in the research by exploring the quality analysis of hotel industry applying modern data mining techniques. The main objective of the study is to investigate the hotel quality using different variables using rough set approach. Also, regression analysis and cluster analysis are also conducted to examine the relationship between different variables and finally a reliable relationship analysis is presented. In this research, we have considered the indispensable factors on the quality of a hotel by using attribute reduction, regression and cluster analysis.

The remaining of the paper is structured as follows. The first section discusses the different empirical methodology used in the study. The next section describes the data for the analysis. The following presents the empirical findings. Next section describes the discussion and the policy implications of hotel management and the final section summarize the conclusions.

## EMPIRICAL METHODOLOGY

In this article, we applied three different models based on the relationship (rough set theory, regression model and cluster analysis) to model the hotel quality analysis for the hotel. In addition, we calculated the dynamic impact of hospitality management by the seven different variables, location, hospitality, facilities, cleanliness, value for money, food quality and price. A brief discussion of various methods is described in the following subsections.

### *Rough set theory:*

The rough set theory has been proposed by Pawlak (1982). The rough set principle is rooted in the hypothesis that every member of the X is associated with the assured amount of data and the attribute used for object description express the facts of data. Information can have indiscernibility while the object has an identical description. For the assessment of a vague description of the member RST is the amazing mathematical tool. The adjective vague signify the data quality that is ambiguity or uncertainty that chase from information granulation. The indiscernibility relation induces a separation of the universe into pieces of indiscernible (similar) objects, named elementary set. The RST can be expressed in two approximations set known as lower and upper approximation of a set.

Let $U = \{x_1, x_{2,}\, . . . , x_n\}$ be the non-empty set of objects known as universe and $A = \{a_1, a_{2,}\, . . . , a_n\}$ is a non-empty set of attributes, then $S = (U,\ A,\ C,\ D)$ is called an information system where $C$ and $D$ are condition and decision attribute, respectively. For any $\mathrm{P} \subseteq \mathrm{A}$ there exist an indiscernibility relation defined as $\mathrm{IND(P)} = \{(\mathrm{x,y}) \in \mathrm{U} \times \mathrm{U} | \forall \mathrm{b} \in \mathrm{P}, \mathrm{b(x)} = \mathrm{b(y)}\}$, where (x, y) is a couple of instances, $b(x)$ represents the value of attribute b for instance x and $IND(P)$ indicate indiscernibility relation (Pawlak 2002). The two principle idea of rough set theory are lower and upper approximation. In RST, any subset $R \subseteq U$ is symbolized by its $P$-lower and $P$-upper approximation of $R$. The lower approximation and upper approximation of set $R$ represents by $\underline{P}(R)$ and $\overline{P}(R)$ respectively; where

$$\underline{\mathrm{P}}(\mathrm{R}) = \{x | [x]_P \subseteq \mathrm{R}\} \qquad (1)$$

$$\overline{\mathrm{P}}(\mathrm{R}) = \{x | [x]_P \cap \mathrm{R} \neq \emptyset\} \qquad (2)$$

The objects in $\underline{\mathrm{P}}(\mathrm{R})$ is known as the set of all members of U which can be surely classified as an object of R in the knowledge P whereas objects in $\overline{\mathrm{P}}(\mathrm{R})$ is the set of all elements of U that can be probably classified as an object of R involving knowledge P.

### *Reduct and positive region*

The objective of the attribute reduction is to recognize the important attribute and to remove the unnecessary attribute of the information system. Attribute reduction is an essential part of the information system which could understand all objects discernible by way of data set and cannot be minimized anymore. Any information system may contain one or more reduct.

The positive region is an essential part of the RST (Pawlak, 1991). The C-Positive region of D contains the set of all cases of $U$ that are certainly classified into partition of U/D by using knowledge from C. The C-Positive region of D defined as:

$$POS_C(\mathrm{D}) = \cup_{R \in U/D}\, \underline{C}R$$

If attribute set $E \subseteq C$ is called a reduct of C with respect to D if following two conditions are satisfied

$$(i)\ POS_C(D) = POS_E(D),$$

$$(ii)\ POS_C(D) \neq POS_{E-\{e\}}(D), for\ any\ e \in E.$$

The core is known as intersections of all reducts of C. The core contains the set of all indispensable attributes.

### *Accuracy of approximation and dependency of attributes (Pawlak, 1991)*

The accuracy of approximation of any subset can be denoted in the following manner:

$$\alpha_P(R) = \frac{|\underline{P}(R)|}{|\overline{P}(R)|}, \qquad (3)$$

where $0 \leq \alpha_P(R) \leq 1$. If the accuracy of approximation of set is equal to 1 then set is called crisp otherwise set is rough.

The dependency of attributes is based on the total member in the lower approximation to the total member in-universe, and it is described as follow:

$$\gamma_C(D) = \frac{|POS_C(D)|}{|U|} \qquad (4)$$

If $\gamma_C(D) = 1$ we say that D depends completely on C, and if $0 \leq \gamma_C(D) \leq 1$, we say that D depends partially on C. furthermore, if $\gamma_C(D) = 0$ then D is entirely independent from C.

**Decision rules:**

Decision rules are used to preserve the core semantics of the feature set from the information of particular problem which is an additional significant characteristic of RST. The reduction of needless situations from the decision rules is termed as attribute reduction.

Following steps are used for information table exploration.

1. Collection of data.
2. Computation of lower and upper approximation of universal set.
3. Determine C-positive region of D.
4. Calculate reduct and core of attribute sets.

The decision rules can be obtained from information table. Rules can be considered as "if $p_j = r$ then $d = q$". Here, condition attribute is $p_j$ which consider the value *r* and decision attribute is d considered the value *q*.

***Predictive analysis***

The parametric model regression analysis is applied to identify whether variables are related or not. Although the regression modeling is frequently used in the literature (e.g. [38,4]). The regression analysis appears to be more powerful than the other modeling process. In addition, this process differs from the correlation analysis. The concept of regression can be defined as a systematic co-movement among the selected variables over the future.

***Exploratory data analysis***

To understand the complicated behavior of the multivariate relationship of grouping items, mostly preferred tools belong to exploratory data analysis. From which cluster analysis is the oldest one, in which no expectations have been considered regarding the number and structure of groups. During the cluster analysis main aim is to evaluate likely groups of objects. That can be done by the help of cluster analysis methodology of identifying the unsupervised configuration of data without any consideration of prior hypothesis, so that on resemblance bases grouping of items in the system can be formed (Singh et al. 2004). Hierarchical clustering is another popular methodology of cluster analysis. In this method, clustering followed a hierarchy from most resembled pair to forming higher groups gradually. Resembles between the groups calculated as Euclidean distance and denoted by the difference on the bases on quantitative value of groups (Johanson and Wichern 2002). Using Euclidean distance through Ward's method hierarchical cluster analysis can be performed on normally distributed data. In order to reduce the sum of square of considered two groups which formed gradually at each stage, so that Ward's method utilize analysis of variance to calculate the distances between groups.

***Data and Variables***

Since the Indian government opened its Hotel market to visitors from international in the month of January to March this article adopts only data before avoiding the potential bias arising from the structural change in the tourism market. Hence, the empirical study focuses on the hotel industry in India. From the tourist point of view, it is one of the important things to reserve the best hotel for their journey. For the travelers, the possibility of getting the best hotel is maximised by selecting certain related attributes of the hotel industry. This article investigates the influence of overall rating (OR) on location (L), hospitality (H), facilities (F), cleanliness (C), value for money (VM), food (F) and price (P) using India international tourist hotel data. All variables are used according to availability and the objective of the study. In our empirical analysis, all different variables are described and defined in Table 1 and 2 with their descriptive statistics of different variables. The data related to hotel industry are obtained from best tourism website (https://www.makemytrip.com). The tourism websites usually have the reviews based on customer's experiences with various destinations, hotels and various services they went through. When new customers want to go to any places as tourists or want to avail any of the services, they can browse these websites and have a look at these reviews and have decisions based on them. These decisions making becomes very difficult when there is large number of options available for hotels, tourist destinations or any of the services available. The proposed approach is helpful for a suitable hotel selection for the existing scenario of the city. When analyzing hotel reviews, location, service, facilities, value, cleanliness, and food are generally considered. The proposed method can be used to recognize the appropriate hotel on the basis of existing data.

**Table 1:** Variables Description

| Variables | Description |
|---|---|
| L (Location) | The geographical position of the hotel for transportation. |
| H (Hospitality) | The quality of greeting and treating tourist. |
| F (Facilities) | Physical characteristics associated with the hotel like travel desk, eating place, parking, and so on. |
| C (Cleanliness) | Cleaning condition of a hotel. |
| VM (value for money) | The extent of performance and effectiveness of the charge with the facility their offering. |
| F (Food) | Customer's health acceptable standard quality of food. |
| P (Price) | Room fare according to hotel facilities. |
| OR (overall quality) | Overall quality rankings. |

**Table 2:** Descriptive statistics

| Variables | Mean | Standard deviation |
|---|---|---|
| Location | 4.36462264 | 0.496455063 |
| Hospitality | 4.20566037 | 0.540550919 |
| Facilities | 4.238679245 | 0.438020267 |
| Cleanliness | 4.26509434 | 2.650334592 |
| Value for money | 4.35471698 | 2.78862388 |
| Food | 2.39292452 | 1.896416409 |
| Price | 9279.386792 | 4677.048409 |
| Overall rating | 4.29528301 | 0.361246414 |

*Empirical findings and discussion:*

This section provides the empirical results of the hotel quality analysis the impact of overall rankings on different attributes using rough set, regression and cluster analysis. The overall empirical analysis has been evaluated by using Minitab 16 for regression and cluster analysis. Also, Rough Set Data Explorer (ROSE2) software (Predki et al., 1998) has been applied for rough set analysis.

*Rough set analysis*

To generate an information system of rough set, estimated dependent and independent variables were used as a conditional and decision variables. In next step, variables are normalized for the rough set analysis. The information system of the normalized values (NV) is classified into three qualitative classes, excellent (class 1), very good (class 2) and good (class 3). The normalized decision table was then used for rough set analysis to study the dependency of attributes and accuracy of the approximation. The obtained results of quality of classification and accuracy of approximation are reported in Table 3. Attribute selection is one of most important step in this study, which can reveal the efficiency of indispensable attributes to data. We can find a smaller attribute set which can describe their important role in the decision table. The core element of decision Table is hospitality. The accuracy of approximation for the three decision classes (excellent, very good and good) is shown in Table 3. The accuracy of approximation for the decision class 'excellent' in demand is consistent than the decision class 'very good' and 'good' in the overall ranking and the accuracy of approximation for the decision class very good in overall ranking is consistent than the decision class very good and excellent in. The overall dependency between conditional and decision attributes is 91%. In our analysis, it is assumed that all the attributes are of equal importance for the hotel quality analysis. Some of the attributes are essential than the others during the data analysis.

International Journal of Applied Engineering Research ISSN 0973-4562 Volume 13, Number 6 (2018) pp. 3988-3998


**Table 3:** Accuracy and quality of approximation

| Class | No of objects | Lower Approximation | Upper approximation | Accuracy of Approximation | Quality of approximation |
|---|---|---|---|---|---|
| Excellent | 83 | 80 | 90 | 0.888 | |
| Very good | 88 | 78 | 96 | 0.812 | 0.915 |
| Good | 41 | 36 | 44 | 0.818 | |

Using Rough set theory training data were analyzed and decision rules were generated. These rules were applied to relationships among condition and decision attributes. Furthermore, 48 certain decision rules were obtained from the information system. The algorithm for the creation of decision rules was applied by Predki, Slowinski, Stefanowski, Susmaga, Wilk (1998). Total 16 decision rules are found to be more accurate since support is greater than 6. The reduced decision rules consists of 16 rules, where eight rules correspond to class excellent, five rules to class very good and three rules to class good. The estimated results of reduced rules are presented in Table 4. From Table 4, we can see that, rule 1 identified by two attributes, “facilities” and “food”, which means that. “If facilities of the hotel are excellent and quality of food is fair THEN overall rating of the hotel will be excellent.” The support and coverage factor of Rule 1 was 40, and 48.19(%), respectively. Rules 2 suggest that “If a hotel with excellent location, excellent cleanliness, and excellent value for money, then the overall rating of the hotel will be excellent.” Rule 2 identified three attributes, “location”, “cleanliness”, and “value for money”. The support and coverage factor of rule 2 was 26 and 31.33(%), respectively.

*Significance of the condition attributes*

The information system (IS) can generate only one reduct from here is only one reduct which necessarily means, every class is determined by all the conditional attributes of the IS. The importance of the conditional attributes can be understood by their presence in the decision rules. If a conditional attribute is frequently used in some decision rules and if the corresponding support values of these rules are higher, then the attribute will be very important as far as the decision of the tourist is concerned. The frequency of the conditional attributes and support values in the generated minimum cover rules are shown in Table 6. More the number of objects (hotels) matching these rules, high will be the importance of an attribute. Here, in Table 6, rules 1 and 6 imply facilities, cleanliness, and food are the most significant attributes as far as the tourist decisions are concerned with support values 40 and 34 respectively. Table 6 also specifies that for most of the rules, facilities and food are the most important condition rules. Therefore interpretation that can be drawn considering the data in Table 6 is that higher the customer satisfaction for facilities and food more will be the tourist attention to the hotel. The robustness of decision rules is evaluated using corrected classified accuracy. The results of corrected classified are given in Table 5. The overall accuracy of the model is 80.95.

**Table 4:** Estimated decision rules from an information system

| Rule no. | Rule explanation | support | coverage |
|---|---|---|---|
| 1 | (facilities = Excellent) & (food = Fair) => (overall rating = Excellent) | 40 | 48.19 |
| 2 | location = Excellent) & (Cleanliness = Excellent) & (value for money = Excellent) => (overall rating = Excellent) | 26 | 31.33 |
| 3 | facilities = Excellent & (food = Good) => (overall rating = Excellent) | 16 | 19.28 |
| 4 | facilities = Excellent) & (price = High) => (overall rating = Excellent) | 22 | 26.51 |
| 5 | (Cleanliness = Very Good) & (value for money = Excellent) & (food = Fair) => (overall rating = Excellent) | 16 | 19.28 |
| 6 | (facilities = Excellent) & (Cleanliness = Excellent) => (overall rating = Excellent) | 34 | 40.96 |
| 7 | (hospitality = Excellent) & (price = Very High) => (overall rating = Excellent) | 13 | 15.66 |
| 8 | (Cleanliness = Good) & (value for money = Excellent) => (overall rating = Excellent) | 9 | 10.84 |
| 9 | (hospitality = Very Good) & (Cleanliness = Good) & (food = Fair) & (price = Medium) => (overall rating = Very Good) | 14 | 15.91 |
| 10 | (location = Very Good) & (hospitality = Very Good) & (facilities = Very Good) => (overall rating = Very Good) | 16 | 18.18 |
| 11 | (hospitality = Very Good) & (facilities = Very Good) & (Cleanliness = Excellent) => (overall rating = Very Good) | 10 | 11.36 |
| 12 | (facilities = Very Good) & (value for money = Very Good) & (food = Good) => (overall rating = Very Good) | 8 | 10 |
| 13 | location = Excellent & (facilities = Very Good & (Cleanliness = Good) & (value for money = Very Good) => (overall rating = Very Good) | 11 | 12.50 |
| 14 | (location = Good) & (value for money = Good) & (price = Low) => (overall rating = Good) (facilities = Good) & (food = Very Poor) => (overall rating = Good) | 8 | 19.51 |
| 15 | (location = Very Good) & (facilities = Good) & (price = Low) => (overall rating = Good) | 7 | 17.07 |
| 16 | | 6 | 14.63 |

**Table 5:** Classification matrix for three decision classes

| | Class 1 | Class 2 | Class 3 |
|---|---|---|---|
| Class 1 | 78 | 5 | 0 |
| Class 2 | 6 | 62 | 18 |
| Class 3 | 0 | 11 | 30 |
| ND | 2 | | |

## REGRESSION ANALYSIS

On the basis of previous literatures, the model is obtained by employing the following framework:

Overall rating = α + $\alpha_1$L + $\alpha_2$H + $\alpha_3$ F + $\alpha_4$ C - $\alpha_5$ VM - $\alpha_6$ + $\alpha_7$ P + ∈

Where ∈ is the error term OR is the overall rating, L is the location of the hotel, H is the hospitality, F is the facilities, C is the Cleanliness, F is the food quality of the hotel and P is the price of the hotel. Estimated results of regression analysis are reported in Table 6. The obtained results indicate that the impact of location, hospitality, facilities, Cleanliness and value for money all are statistically significant. The positive facilities effect on overall hotel rankings. As expected, the location of the hotel also has a good impact on the overall ranking. On the other hand, the coefficients on food quality and price of the hotel are associated negatively with overall rankings. Hence, food and price variables do not seem to have been helpful to the overall rankings. Moreover, an F-statistic result confirmed that the overall regression model is significant for the hotel predictive analysis purpose since p-value, 0.000 is significant. Now, the next step of estimation is to perform robustness analysis for the model. $R^2$ and residual analysis are used for the robustness checking. Furthermore, $R^2$ value, 0.89 is higher for the model. It explained that all variables together explain approximately 89% of the variance in the overall ranking. The quantile-quantile (QQ) plot is used for the residual analysis. The Estimated Figure 1 shows that the statistical independence for residuals. For better understanding Figure 1 compares the actual and fitted values for the model. The straight line suggests the best fit curve of the model is good fitted with the actual curve for the regression model. Consequently, residuals are independent and follow normal distribution then the model is accepted in the robustness test. Therefore, a regression model is found to be more appropriate for the present study based on the above empirical findings.

The regression equation for the variables is

Overall rating = 0.703 + 0.0823 location + 0.0937 hospitality + 0.660 facilities
+ 0.0616 Cleanliness - 0.0566 value for Money - 0.0018 food + 0.000003 price

Table 6: Regression results

| Variables | Coefficients | SE | t-stat | Prob. |
|---|---|---|---|---|
| Constant | 0.7034 | 0.1078 | 6.53 | 0.000 |
| Location | 0.08234 | 0.02631 | 3.13 | 0.002 |
| Hospitality | 0.09375 | 0.03291 | 2.85 | 0.005 |
| Facilities | 0.65999 | 0.02839 | 23.25 | 0.000 |
| Cleanliness | 0.06158 | 0.01835 | 3.36 | 0.001 |
| Value for money | -0.05656 | 0.01947 | -2.91 | 0.004 |
| Food | -0.00175 | 0.01063 | -0.16 | 0.869 |
| Price | 0.000003 | 0.000001 | 1.57 | 0.118 |
| Adj. $R^2$ | 0.89 | | | |
| F-stats (Prob.) | 249.55(0.000) | | | |

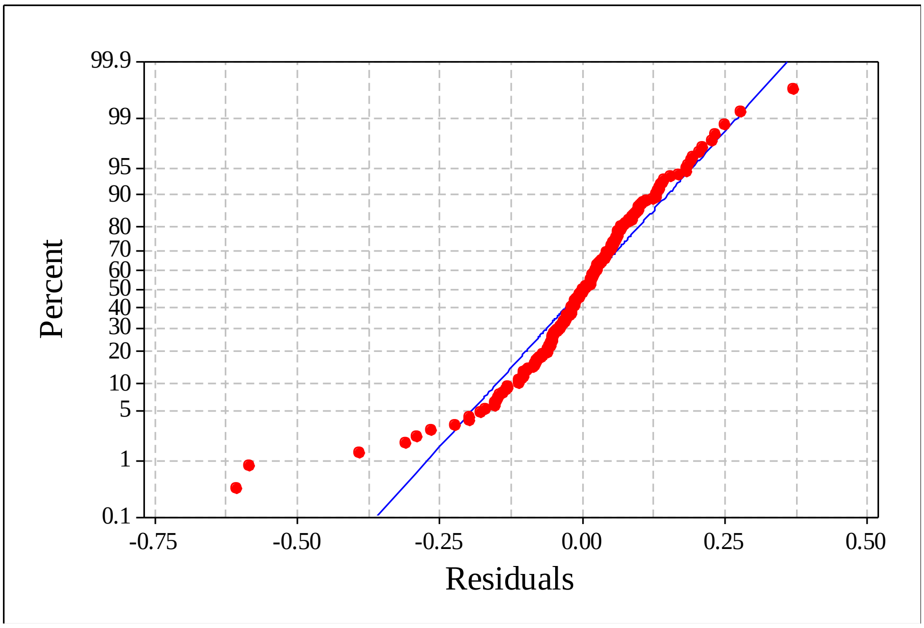


**Figure 1:** QQ plot for the regression model.

*Cluster analysis*

The multivariate statistical technique, cluster analysis was applied to explore the structure of the variables by underlying the group relationships. Figure 2 shows the cluster relation through a diagram, dendrogram, in the dendrogram, all the eight variables on the hotel quality were grouped into three statistically significant clusters. The clustering methodology generated three groups of hotel quality in a very simplest way, as the hotel variables in these groups have similar quality characteristics and ordinary environment source types. Cluster 1 (location, hospitality), cluster 2 (facilities, overall ranking and price) and cluster 3 (Cleanliness, value for money and food) correspond to a relatively moderate hotel quality, very high hotel quality and low hotel quality, respectively. This means that to quickly evaluate the quality of the hotel, three variables of each cluster can serve as a whole hotel quality as well. It is clear that cluster analysis techniques are useful in providing a reliable classification of hotel facilities throughout the system and will probably be able to better design the future sample strategy. Thus, the number of sampling variables in the monitoring network had decreased. There is other empirical evidence (Li et al. (2017)), where cluster analysis approach has successfully opted in the hotel industry.

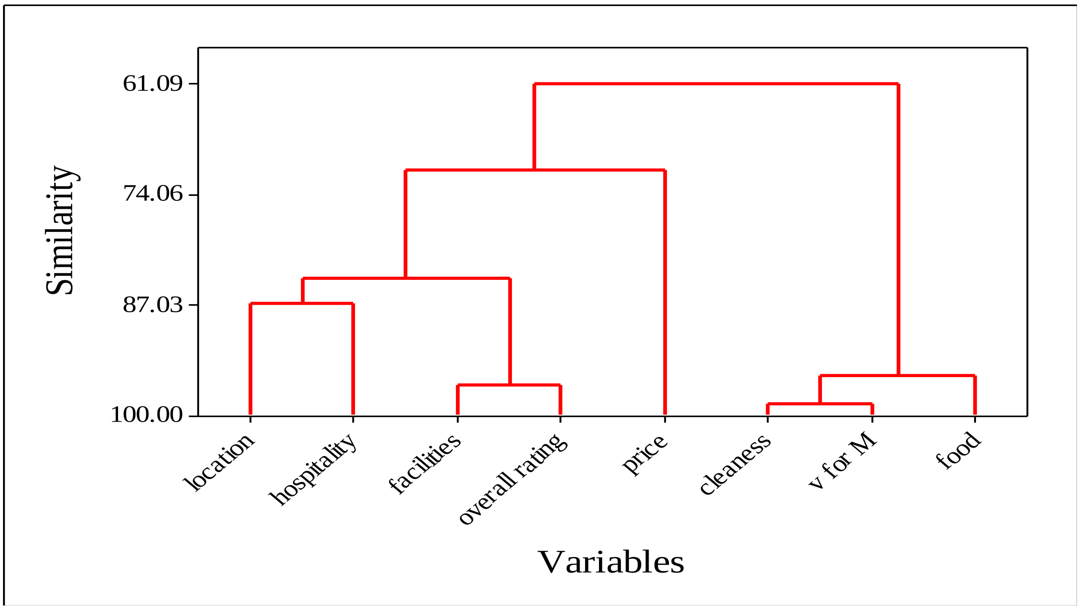


**Figure 2:** Cluster dendrogram for the variables.

## DISCUSSION AND POLICY IMPLICATIONS

In this paper Rough set theory (RST) was applied to data on the hotel industry. Rough set theory can easily handle the uncertainty of attribute-based data set without any statistical and distributional assumptions. The rough set theory is certainly applicable to a wide variety of data whereas, no such evidence has been found for tourist behavior of the hotel industry. We also consider inconsistencies within attributes.

In this article, a case study is performed on hotel industry by applying the concepts of the rough set on the related data. Here, the uncertainty involved with attribute-based data sets are processed with rough sets. Although rough set theory can be applied to a wide range of applications, however to the best of our knowledge, understanding tourist behavior for hotel booking using rough set is yet to be implemented. In this study, we have used rough set theory to analyze survey data related to hotels and to understand tourist behavior while selecting a hotel for booking. The related results of our study show that rough sets can be effectively applied for mining the data when a tourist chooses a hotel for booking. In case of RST methodology, analysis can avoid conditional attribute if there has been no change observed in the quality of approximation. But in this study, every attribute has been selected as a reduct attribute or core attribute. This means that all the attributes are important factors for the quality of classification as far as the hotel data is concerned. Here, Table 3 reports the accuracy of approximation and quality of classification of hotel data set. From Table 3, we observe that quality of classification for tourist behavior is 0.92, which implies that boundary region contains very few objects of the information system. In this case, quality of classification closer to 1 also suggests that a better dependency amongst all conditional attributes exist for every member of the class. In this study, a total of 16 minimum cover rules is used whose support and coverage value are respectively more than 6 and 10 (%). This implies that a majority of the data obtained from the case study can be exclusively classified by 16 rules.

In this paper, we have used Rough set theory in a hotel booking case study. The results associated with this study reveals that proposed approach provides a substantial evidence of rough set theory and is feasible to adopt. In this paper, a novel decision rule approach has been implemented to derive the results from a hotel data set. In contrast to the several classical statistical techniques such as logistic regression, discriminant analysis, etc., the importance of the rough set theory is that it does not require any statistical assumptions related to any distribution. Moreover, RST can handle inconsistence of both variables and attributes. Consequently, this study has implemented the surveyed data with qualitative and quantitative attributes to analyze the booking of hotel rooms from a customer's point of view.

Additionally, the study reveals that all variables together explain approximately 89% of the variance in the overall ranking. In this study, we have also considered the indispensable factors on the quality of a hotel by using attribute reduction, regression and cluster analysis. According to the attribute reduction, there is a close relationship between conditional and decision attributes. The obtained results indicate that the impact of location, hospitality, facilities, Cleanliness and Value for money all are statistically significant. Also, the positive facilities effect on overall hotel rankings. In addition, the regression results indicate that the location of the hotel also has a good impact on the overall ranking. The food quality and price of the hotel are associated negatively with overall rankings. Hence, food and price variables do not seem to have been helpful to the overall rankings. Although the presence of rough set analysis suggests that one variable can be useful in predicting relationships another variable. Moreover, the regression model is significant for the hotel predictive analysis purpose. It is evident that the cluster analysis technique is useful in offering a reliable classification of hotel features in the whole system and will make possible to design a future sampling strategy optimally. Thus, the number of sampling variables in the monitoring network was reduced. Cluster analysis evidence suggested that variables have effects on overall rankings. The important policy implications of cluster analysis are that the overall rankings are strongly dependent on the hotel quality analysis. According to (Mei et al., 1999, Tsang and Qu, 2000, Ekinci et al., 2003 and Juwaheer, 2004), overall rankings are important regarding hotel quality analysis. Moreover, hotel analysis has become a very popular in the hospitality sector.

## CONCLUSIONS

The purpose of the research was to model the rough set, regression and cluster analysis for the quality analysis of hotel industry. Seven key explanatory variables (e.g. location, hospitality, Cleanliness, facilities, value for money, food and price) for the hotel were incorporated into the modeling of the hotel industry. In robustness analysis, rough set and regression models appeared to establish the reliable and accurate results in terms of corrected classified accuracy and adjusted R2. The main contribution of the paper is that the more accurate and consistent results of the rough set approach with hotel variables without any particular statistical assumptions. Moreover, the study extends the hospitality literature by investigating overall rankings and their influence factors for a hotel data. Also, variables are relatively more dependent on overall rankings for hotel quality analysis. Furthermore, the major part of hotel industry has been initiated to promote hospitality sector and customer satisfaction.

The investigation of the firm relationship between overall rankings and their different factors has important policy implications. Hence, much empirical studied have been conducted for the hotel quality analysis. But these empirical studies have not yielded a consensus on the relationship between different variables. There is no literature that applied rough set theory for the relationships of hotel quality analysis. Therefore, this study demonstrates the relationships between hotel variables for overall rankings.

This work attempts the vital applications of the rough set theory in the form of decision rules. The decision rules have been conducted through the rough set information system. The results, as compared with statistical analyses are more advanced and dynamic. Our empirical analysis reveals that the derived decision rules are easier to understand compared with statistical time series methods without any distributional assumptions.

## REFERENCES

[1] Au, N. and Law, R. (2000): The application of rough sets to sightseeing expenditures, Journal of Travel Research, 39(1), 70-77.

[2] Akbaba, A (2006): Measuring service quality in the hotel industry: A study in a business hotel in Turkey Hospitality Management 25, 170–192.

[3] Akincilar, A and Dagdeviren, M (2014): A hybrid multi-criteria decision making model to evaluate hotel websites International Journal of Hospitality Management 36, 263– 271.

[4] B.Predki, R.Slowinski, J.Stefanowski, R.Susmaga, Sz.Wilk (1998) ROSE - Software Implementation of the Rough Set Theory. In: L.Polkowski, A.Skowron, eds. Rough Sets and Current Trends in Computing, Lecture Notes in Artificial Intelligence, vol. 1424. Springer-Verlag, Berlin, 605-608. http://idss.cs.put.poznan.pl/site/64.html.

[5] B.Predki, Sz.Wilk (1999) Rough Set Based Data Exploration Using ROSE System. In: Z.W.Ras, A.Skowron, eds. Foundations of Intelligent Systems, Lecture Notes in Artificial Intelligence, vol. 1609. Springer-Verlag, Berlin, 172-180.

[6] https://www.makemytrip.com.

[7] Benitez, JM, Martin, J C, Roman, C (2007): Using fuzzy number for measuring quality of service in the hotel industry Tourism Management 28, 544–555.

[8] Chou TY, Hsu, CL, Chen, MC (2008) A fuzzy multi-criteria decision model for international tourist hotels location selection (2008) International Journal of Hospitality Management 27, 293–301

[9] Chen C M (2011): Influence of uncertain demand on product variety: evidence from the international tourist hotel industry in Taiwan Tourism Economics 17(6), 1275-1285.

[10] Celotto, E., Ellero, A. and Ferretti, P. (2012): Short-medium term tourist services demand to forecast with rough ret theory, Procedia Economics, and Finance, 3, 62-67.

[11] Celotto, E, Ellero A and Ferretti P (2015): Conveying tourist ratings into an overall destination evaluation Procedia - Social and Behavioral Sciences 188, 35 – 41.

[12] Chen, C. M., Lin, Y. C. and Yang, H. W. (2017): The effects of advertising on market share instability in the hotel industry Tourism Economics 23(1), 214-222.

[13] Chen, M. H. (2016): A quantile regression analysis of tourism market growth effect on the hotel industry International Journal of Hospitality Management 52, 117–120

[14] Ekinci, Y., Prokopaki, P., Cobanoglu, C. (2003): Service quality in Cretan accommodations: marketing strategies for the UK holiday market. International Journal of Hospitality Management 22, 47–66.

[15] Goh, C. and Law, R. (2003) Incorporating the rough sets theory into travel demand analysis. Tourism Management. 24, 511-517.

[16] Golmohammadi, A and Ghareneh, N. S. (2011): Importance analysis of travel attributes using a rough set-based neural network: The case of Iranian tourism industry Journal of Hospitality and Tourism Technology 2(2), 155-171.

[17] Geetha, M, Singha, P and Sinha, S. (2017): Relationship between customer sentiment and online customer ratings for hotels - An empirical analysis Tourism Management 61, 43-54.

[18] Hua Nan and Yang Yang (2017): Systematic effects of crime on hotel operating performance Tourism Management 60, 257-269.

[19] Johanson, R. A. and Wichern, D. W. (2002): Applied multivariate statistical analysis. Fifth Edition.

[20] Juwaheer, T. D. (2004) Exploring international tourists' perceptions of hotel operations by using a modified SERVQUAL approach: a case study of Mauritius. Managing Service Quality 14 (5), 350–364.

[21] Law, R. and Au, N. (1998): A rough set approach to hotel expenditure decision rules induction, Journal of Hospitality and Tourism Research, 22(4), 359-375.

[22] Law, R. and Au, N. (2000): Relationship modeling in tourism shopping: a decision rules induction approach, Tourism Management, 21(3), 241-249.

[23] Liou J. J. H. (2009): A novel decision rules approach for customer relationship management of the airline market, Expert System with Applications, 36, 4374-4381. DOI: 10.1016/j.eswa.2008.05.002.

[24] Liou J. H and Tzeng G. H. (2010): A Dominance-based rough set approach to customer behavior in the airline market, Information Sciences, 180, 2230-2238. DOI: 10.1016/j.ins.2010.01.025.

[25] Lahap J, Ramli N S, Said, N M, Radzi S M, and Zain RA (2016): A Study of Brand Image towards Customer's Satisfaction in the Malaysian Hotel Industry Procedia - Social and Behavioral Sciences 224, 149 – 157.

[26] Liou, J. J. H., Chuang, Y. C., Hsu, C. C. (2016): Improving airline service quality based on rough set theory and flow graphs. Journal of Industrial and Production Engineering. 33(2), 123-133.

[27] Li-Fei Chen, C.T. Tsai, (2016): Data mining framework based on rough set theory to improve location selecting decisions: A case study of a

restaurant chain, Tourism Management, 53, 197- 206. DOI: 10.1016/j.tourman.2015.10.001.

[28] Li. L, Peng, M, Jiang, N and Law, R (2017): An empirical study on the influence of economy hotel website quality on online booking intentions International Journal of Hospitality Management 63 1-10.

[29] Lafuente A M G, Merigó J. M. and Vizuete Vizuete (2014): Analysis of luxury resort hotels by using the Fuzzy Analytic Hierarchy Process and the Fuzzy Delphi Method (2014) Economic Research-Ekonomska Istraživanja, 27(1), 244–266

[30] Lahap J, Ramli N S, Said, N M, Radzi S M, and Zain R. A. (2016): A Study of Brand Image towards Customer's Satisfaction in the Malaysian Hotel Industry Procedia - Social and Behavioral Sciences 224, 149 – 157.

[31] Lai I. K. W. and Hitchcock M. (2017): Sources of satisfaction with luxury hotels for new, repeat, and frequent travelers: A PLS impact-asymmetry analysis Tourism Management 60, 107-129.

[32] Mei, A.W.O., Dean, A.M., White, C.J. (1999): Analyzing service quality in the hospitality industry. Managing Service Quality 9 (2), 136–143.

[33] Mohsin, Asad, Ryan and Chris, (2005): Service Quality Assessment of 4-star Hotels in Darwin, Northern Territory, Australia". Journal of Hospitality and Tourism Management.

[34] Mohajerani, P., and Miremadi, A. (2012): Customer satisfaction modeling in the hotel industry: A case study of Kish Island in Iran. International Journal of Marketing Studies, 4(3), 134-152.

[35] Masiero L, Heo C Y and Pan, B. (2015): Determining guests' willingness to pay for hotel room attributes with a discrete choice model International Journal of Hospitality Management 49, 117–124.

[36] Mardani, A, Zavadskas, E. K, Streimikiene, D, Jusoh, A, Nor, K. M. D. and Khoshnoudi, M. (2016): Using fuzzy multiple criteria decision making approaches for evaluating energy-saving technologies and solutions in five-star hotels: A new hierarchical framework Energy 117, 131-148.

[37] Omurca, S. I. (2013): An intelligent supplier evaluation, selection and development system. Applied Soft Computing. 13, 690-697.

[38] Ohlan, R. (2012): ASEAN India free trade agreement in goods: An assessment African Journal of Social Sciences 2(3), 66–84.

[39] Ohlan, R. (2016): Rural transformation in India in the decade of miraculous economic growth. Journal of Land and Rural Studies 4(2), 188–205.

[40] Ohlan R. (2017): The relationship between tourism, financial development and economic growth in India Future Business Journal, 3(1), 9–22.

[41] Pawlak, Z. (1982): Rough sets. *International Journal of Computer and Information Science*. 11, 341- 356.

[42] Predki, B., Wong, S. K. M., Stefanowski, J., Susmaga, R., Wilk, Sz., (1998): ROSE-software implementation of the rough set theory. In L. Pollkowski, A. Skowron (Eds.). Rough Sets and Current Trends in Computing. *Lecture Notes in Artificial Intelligence. Berlin: Springer*. 605-608.

[43] Predki B., Sz.Wilk (1999): Rough Set Based Data Exploration Using ROSE System. In: Z.W.Ras, A.Skowron, eds. Foundations of Intelligent Systems, Lecture Notes in Artificial Intelligence, vol. 1609. Springer-Verlag, Berlin, 172-180.

[44] Pawlak, Z. (2002): Rough sets and intelligent data analysis, Information Sciences, 147, 1-12. DOI: 10.1016/S0020-0255(02)0019 7-4.

[45] Patiar A, Ma Emily, Kensbock Sandra, and Cox Russell (2017): Students' perceptions of quality and satisfaction with virtual field trips of hotels Journal of Hospitality and Tourism Management 31, 134-141

[46] Rianthong, N, Dumrongsiri, A and Kohda, Y. (2016): Optimizing customer searching experience of online hotel booking by sequencing hotel choices and selecting online reviews: A mathematical model approach, Optimizing customer searching experience of online hotel booking by Tourism Management Perspectives 20, 55–65.

[47] Rajaguru, R. and Rajesh, G. (2016): Value for money and service quality in customer satisfaction The social sciences 11(19), 4613-4616.

[48] Ren L, Qiuc H, Wangd P, Lin P M.C. (2016): Exploring customer experience with budget hotels: Dimensionality and satisfaction International Journal of Hospitality Management 52, 13–23.

[49] Singh, KP, Malik, A, Mohan, D, Sinha, S. (2004): Multivariate statistical techniques for the evaluation of spatial and temporal variations in water quality of Gomti River (India)—a case study Water Research 38, 3980–3992.

[50] Sharma N and Kalotra A. (2016): Hospitality Industry in India: A Big Contributor to Indias Growth, International Journal of Emerging Research in Management &Technology, 202-210.

[51] Tsang, N., Qu, H. (2000): Service quality in China's hotel industry: a perspective from tourists and hotel managers. International Journal of Contemporary Hospitality Management 12 (5), 316–326.

[52] Tsai W H, Hsu J L, Chen C. H, Lin, W. R. and Chen, S. P. (2010): An integrated approach for selecting corporate social responsibility programs and costs evaluation in the international tourist hotel International Journal of Hospitality Management 29, 385–396.

International Journal of Applied Engineering Research ISSN 0973-4562 Volume 13, Number 6 (2018) pp. 3988-3998
© Research India Publications. http://www.ripublication.com

[53] Viglia G, Mauri, A. and Carricano, M. (2016): The exploration of hotel reference prices under dynamic pricing scenarios and different forms of competition International Journal of Hospitality Management 52, 46–55.

[54] Wang, Y. S, Li H .T, Li, C. R and Zhang D. Z. (2016): Factors affecting hotels' adoption of mobile reservation systems: A technology-organization-environment framework Tourism Management 53, 163-172.

[55] World Travel and Tourism Council (2015): Travel and Tourism Economic Impact India (PDF).

[56] Yu S M, Wang J and Wang J Q (2016) An Interval Type-2 Fuzzy Likelihood-Based MABAC Approach and Its Application in Selecting Hotels on a Tourism Website Int. J. Fuzzy Syst. DOI 10.1007/s40815-016-0217-6.

[57] Xiaoya, H. and Zhiben, J. (2011): Research on an econometric model for domestic tourism income based on the rough set, *Springer*, 259-266.

[58] Xu Xun, Li Yibai (2016): The antecedents of customer satisfaction and dissatisfaction toward various types of hotels: A text mining approach International Journal of Hospitality Management 55, 57–69.

[59] Salamo, M., Lopez-Sanchez, M. (2011): Rough set based approaches to feature selection For Case-Based Reasoning classifiers. Pattern Recognition Letters, 32(2), 280-292.