

Informatics Tools, AI Models and Methods Used for Automatic Analysis of Customer Satisfaction

George KOVÁCS¹, Diana BOGDANOVA², Nafissa YUSSUPOVA², Maxim BOYKO²

¹ Computer and Automation Research Institute,
Kende u. 13-17, Budapest, 1111, Hungary,
kovacs.gyorgy@sztaki.mta.hu

² Ufa State Aviation Technical University,
K. Marx 12, Ufa, 450000, Russia,
dianochka7bog@mail.ru, yussupova@ugatu.ac.ru, maxim.boyko87@gmail.com

Abstract: Customer satisfaction is getting more and more importance world-wide. Informatics tools and methods are used to research customer satisfaction based on a detailed analysis of consumer reviews. The examined reviews are written in natural languages and some Artificial Intelligence (AI) techniques such as Text Mining, Aspect Sentiment Analysis, Data Mining and Machine Learning are used for the study. As input for running the investigations, we use different internet resources in which the accumulated customer reviews are available. These are for example yelp.com, tripadvisor.com and tophotels.ru, etc. To see and show the efficacy of the proposed approach, we have carried out experiments on hotel client satisfaction. The results have proven the effectiveness of the proposed approach to decision support in product quality management and support applying them instead of traditional methods of qualitative and quantitative research of customer satisfaction.

Keywords: quality management; customer satisfaction research; decision support system; sentiment analysis.

1. Introduction

Quality assurance is currently realized by means of a process approach based on the model of a quality management system [1]. It describes the interaction of the company and the customer during the process of product production and consumption. To correct the parameters of product quality in order to improve it for the customer, the models include feedback. For companies, one aspect of feedback during the process of quality management is information about the level of customer satisfaction, expressed in the form of customer reviews of the product quality. That is why customer satisfaction is the key information in quality management that influences decision-making.

To collect data and to evaluate customer satisfaction, the International Quality Standard ISO 10004 recommends using the following methods: personal and phone interviews, discussion groups, mail surveys (postal questionnaires), online research and survey (questionnaire survey) [2]. However, these methods of collecting and analyzing customer opinions show a number of drawbacks.

A general drawback of the recommended methods is the need for a large amount of manual work: preparing questions, creating a respondent database, mailing questionnaires and

collecting results, conducting personal interviews, preparing a report based on the results. All this increases the research costs. Due to their discreteness these methods do not allow for the continuous monitoring of customer satisfaction. For this reason, the data analysis is limited to one time period and does not give an insight into the trends and dynamics of customer satisfaction. This also has a negative influence on the speed of managerial decision making, which depends on the arrival rate of up-to-date information about customer opinions.

Existing scales of customer satisfaction and their subjectivity perception raise questions. Values of customer satisfaction expressed in the form of abstract satisfaction indices make it difficult to understand, compare and interpret the results. Methods of analysis of data collected through the recommended ISO 10004 procedures permit only the detection of linear dependencies.

To increase the effectiveness of product quality management, we suggest approaching the research of customer satisfaction through the use of Informatics, as AI technologies. Applying Text Mining tools for analyzing customers' reviews posted on the Internet is not novel. There are many studies concerning models and methods for data collection, sentiment analysis and information extraction. Recent studies show acceptable accuracy of methods for sentiment classification. Gräbner

et al. [3] proposed a system that performs the sentiment classification of customer reviews on hotels. The precision values are 84% for positive and 92% for negative reviews. Lexicon-based method [4] allowed the correct classification of reviews with a probability of about 90%. These achievements make sentiment analysis applicable for an application on quality management and customer satisfaction research.

Jo and Oh [5] and Lu et al. [6] considered the problems of automatically discovering products' aspects and sentiments estimation for these aspects, which are evaluated in reviews. For solving these problems, they suggested methods based on Latent Dirichlet Allocation [7] and its modifications.

The main drawback of most social monitoring systems and frameworks for automatic analysis of reviews is that they can provide entirely only a quantitative survey of customer reviews, i.e., they can provide measurement of the degree of customer satisfaction with a product and its aspects, sometimes with a model [9]. Qualitative survey were usually only conducting the extraction of products' aspects. However, estimation of the significance of each products' aspects for the customer is missed. The information about products' aspects that influence customers' satisfaction and relative importance of products' aspects for the customers is missing, as well as an insight into customer expectations and perceptions.

The most related work to this problem is [8]. It is dedicated to the topic of aspect ranking, which aims to automatically identify important aspects of product from online consumer reviews. Most proposals used a probabilistic model with a large number of parameters that lead to low robustness of the model. Total weighting values of aspects are calculated as the average of the weighting values by each review. Finally, significance values of aspects are estimated independently of sentiments of opinions. In real life we can speak about bad "signal connection", in a review, but we usually omit comments in the case of good "signal connection", as it should be caused by the phone. In our paper, we estimate significance values of aspects in accordance with their positive and negative sentiments.

In this paper, for qualitative survey is used a novel approach based on transformation results of sentiment analysis and aspect-based

sentiment analysis, such as sentiment labels of reviews and mentions about product's aspects in reviews, into boolean data. After that, boolean data is processed with a data mining tool – decision tree. Qualitative survey aims to identify how the sentiment of reviews depends on the sentiment of different products' aspects. In other words, how overall customer satisfaction with product depends on the customer satisfaction with a product's aspects. Decision tree performs this aim and identifies latent relations between the sentiment of reviews and sentiment of a product's aspects. Also using the decision tree allows to estimate the significance of product's aspects for the customers. The output of the qualitative survey contains significant values of the product's aspects for customers, and identifies latent relations between satisfaction with the product and satisfaction with each product's aspect. These were produced as rules extracted by the decision tree. The availability of both quantitative and qualitative surveys allows realizing Intelligent Decision Support System for Quality Management in accordance with quality standard ISO 10004.

2. Quality Management – Customer Satisfaction

Figure 1 represents the algorithm of the suggested approach to quality management based on research into customer satisfaction using Artificial Intelligence applications. It consists of four main stages. The first stage includes the collection of reviews from Internet resources, data cleansing and loading data into the database. The second stage comprises the processing and analysis of the collected reviews. It includes marking reviews by their emotional response, i.e., sentiment (for example, negative and positive), identifying product aspects, and evaluating the sentiment of the comments on the separate aspects. Following the stage of data processing utilizing visualization tools, quantitative research is carried out. A qualitative research of customer satisfaction is undertaken by means of building models based on decision trees, where the review's sentiment serves as a dependent variable, and sentiment comments on separate product aspects as independent variables. Managerial decision development and making is carried out on the basis of these four stages.

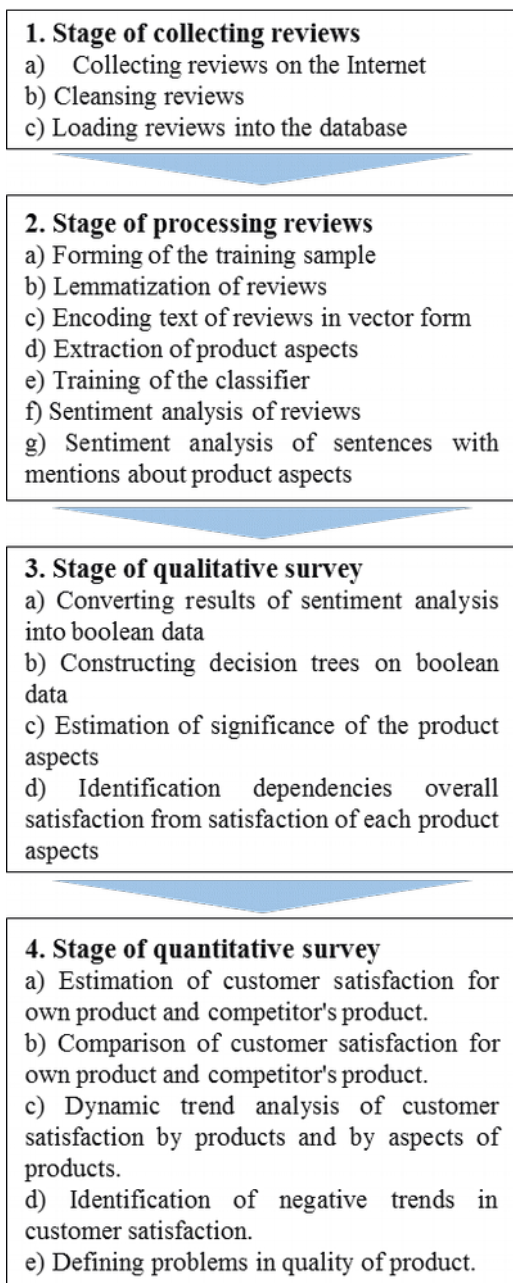


Figure 1. The algorithm of AI based quality management of customer satisfaction.

3. Applicable AI Means

3.1 Collecting important data

Nowadays there are a large number of Internet resources where users can leave their opinions about goods and services. The most popular examples are tophotels.ru (635,000 reviews), yelp.com (53 million reviews), tripadvisor.com (travels, 130 million reviews). Similar resources continue to gain popularity. Their advantage as a source of information for satisfaction evaluation lies in their purpose – the accumulation of customer reviews. As opposed to social network services, the web

pages of review sources use XML that determines the structure typical for a review. Such a structure includes separate blocks with the name of a product or company and a review, and other blocks with additional information. In the case of several input sources, information/data should be managed by up-to-date tools, as e.g. cloud computing. Therefore, all reviews are clearly identified in relation to the review object. It significantly simplifies the process of data collection and excludes the problem of key word ambiguity. One further advantage is that many of such resources monitor the reviews and check the objectivity of the authors.

There are two main types of collecting Internet data on customer reviews: 1) by using API (application programming interface) and 2) by web parsing. API is a set of ready-to-use tools – classes, procedures, functions – provided by the application (Internet resource) for use in an external software product. Unfortunately, only few resources that accumulate reviews have their own API. In this case, to collect reviews we can use the second method for data collection – web parsing. Web parsing is a process of automated analysis and content collection from xml-pages of any Internet resource using special programs or script. In this paper is used the second method for data collection – web data extraction. It is a process of automated content collection from HTML-pages of any Internet resource using special programs or script. Related work is presented in [11].

3.2 Analysis of sentiments

After the data has been collected and cleaned, we can start their processing with the help of Text Mining tools. Sentiment Analysis is used to evaluate the author's product satisfaction. Sentiment stands for the emotional evaluation of an author's opinion in respect to the object that is referred to in the text. We can distinguish three main approaches to Sentiment Analysis: 1) linguistic, 2) statistical, and 3) combined. The linguistic approach is based on using rules and sentiment vocabulary. It is quite time-consuming due to the need of compiling sentiment vocabularies, patterns and making rules for identifying sentiments. But the main drawback of the approach is the impossibility of obtaining a quantitative evaluation of the sentiment. The statistical approach is based on the methods of supervised and non-supervised machine

learning. The combined approach refers to a combined use of the first two approaches.

We use the methods of supervised machine learning: Bayesian classification and Support Vector Machines. Software implementation is simple, and does not require generating linguistic analyzers or sentiment vocabularies. Text sentiment evaluation can be expressed quantitatively. To apply these methods, a training sample was created. To describe an attribute space, vector representation of review texts was used with the help of the bag-of-words model. Bit vectors - presence or absence of the word in the review text, and frequency vectors – the number of times that a given word appears in the text of the review, served as attributes. Lemmatization, a procedure of reducing all the words of the review to their basic forms, was also used. More details can be found in [12].

3.3 Aspects of sentiment analysis

Sentiment Analysis of reviews allows to evaluate general customer product or company satisfaction. However, it does not make clear what exactly the author of the review likes and what not. To answer this question, it is necessary to perform an Aspect-based Sentiment Analysis. An aspect means characteristics, attributes, qualities, properties that characterize the product, for example, a phone battery or delivery period, etc. However, one sentiment object can have a great number of aspects. Furthermore, aspects in the text can be expressed by synonyms (battery and accumulator). In such cases it is useful to combine aspects into aspect groups.

An Aspect-based Sentiment Analysis of a review is a more difficult task and consists of two stages – identifying aspects and determining the sentiment of the comment on them. To complete the task of the Aspect-based Sentiment Analysis, a simple and effective algorithm has been developed (see Figure 2).

A frequency vocabulary (based on the corpus) that helps to compare the obtained frequencies with word frequencies is used to identify aspects. The nouns with maximum frequency deviations are candidates for inclusion into aspect groups. Division of the noun set into aspect groups was carried out manually. We should note that if a sentence includes nouns from several aspect groups, then it will appear in each of them. The results of Sentiment

Aspect extraction

Input: set of reviews D

1. Extract all nouns S from the set of reviews D .
2. Count the frequency of nouns
 $\forall t = 1, |S|: f_t = N_t / N$ in the whole set of reviews D , where N – number of appearances of all words, N_t – number of appearances of the t noun.
3. Count the difference $\forall t: \Delta_t = f_t - f_t^v$ between the counted frequencies f_t and vocabulary frequencies f_t^v .
4. Sort the set of nouns S in descending order Δ_t .
5. Divide the set of nouns S from $\Delta_t > 0$ into aspect groups.

Output: set of aspect groups and aspect words

Aspect-based sentiment classification

Input: sentiment classifier, set of aspect groups and aspect words

1. Divide a set of reviews into set of sentences.
2. Perform sentiment classification for each sentence.
3. Check each sentence for the condition: if a sentence has a sentiment score (negative or positive) greater than a threshold h and contains at least one noun from any aspect group, then this sentence is labeled as an opinion (negative or positive) about the given product's aspect.

Output: labeled sentences with mentions about product's aspects

$\{Neg_{i_1, \dots, i_m}, Pos_{i_1, \dots, i_m}\}$

Figure 2. Aspect-based Sentiment Analysis.

Analysis and Aspect Sentiment Analysis can be represented in the form of text variables $Obj = (Rev_i, Sent_i, Neg_i^1, \dots, Neg_i^j, Pos_i^1, \dots, Pos_i^j)$, where Obj is a sentiment object or a product, Rev_i the text of the i review, $Sent_i$ the sentiment of i review, Neg_i^j the negative sentences about the j aspect in the i review, Pos_i^j the positive sentences about the j aspect in the i review, i the review number, j the aspect group number.

3.4 A well established way: decision tree

The following section focuses on an algorithm of the processing of data obtained with help of Sentiment Analysis and Aspect-based Sentiment Analysis. The task of the developed algorithm is the mining of data that can be used for decision support in product quality management. To realize this algorithm, we use the Data Mining method, i.e. the decision tree, since this tool can be easily understood and results can be clearly interpreted; it also can explain situations by means of Boolean logic.

The algorithm of processing of data obtained by means of Sentiment Analysis is presented on Figure 3.

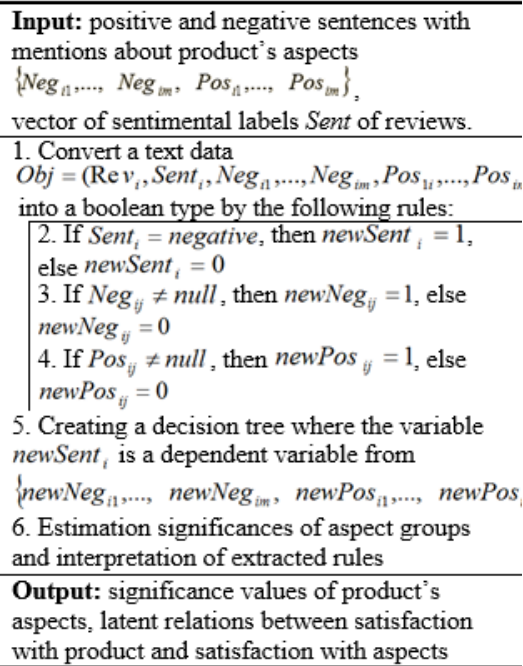


Figure 3. Algorithm of data mining.

The algorithm we have described allows us to understand which sentiment comments on product aspects influence the whole text sentiment or, in other words, what product aspects influence customer satisfaction and in what way. Our decision tree model allows us to consider the influence not only of separate sentiment comments on aspects but also of their mutual presence (or absence) in the text on customer satisfaction. The decision tree model also enables us to detect the most significant product aspects that are essential for the customer. The logical constructions (called rules) that we have obtained can be expressed both in the form of Boolean functions in a disjunctive normal form and in natural language.

A decision tree model can help to predict sentiment in dependence on various inputting aspect comments of different sentiments. In fact, it makes it possible to evaluate experimentally customer satisfaction in dependence on satisfaction with different product attributes. As the final result, prediction and analysis of the influence of different inputting variants on customer satisfaction allows us to distribute the company's budget effectively to maintain a high product quality.

As a measure of the customer satisfaction with product is used a ratio of positive reviews to the sum of positive and negative reviews. The score

of customer satisfaction *CS* by product is calculated by (1):

$$CS = \frac{Z^{pos}}{Z^{pos} + Z^{neg}} \cdot 100\% \quad (1)$$

where Z^{pos} – number of positive reviews, Z^{neg} – number of negative reviews.

As a measure of the customer satisfaction with product's aspect groups is used a ratio of positive sentences with mentions of a product's aspect to the sum of positive and negative sentences with mentions of a product's aspect. The score of customer satisfaction cs_j with j product's aspect group is calculated by (2):

$$cs_j = \frac{z_j^{pos}}{z_j^{pos} + z_j^{neg}} \cdot 100\% \quad (2)$$

where z_j^{pos} – number of positive sentences containing mention about the j product's aspect group, z_j^{neg} – number of negative comments containing mention about the j product's aspect group. Unlike indices in [2] (which often represent the average subjective values obtained with using rating scales) proposed measures show the ratio of positive / negative reviews to total number of reviews. It gives more clearer for understanding of monitoring results.

Significance of aspects group shows how much the sentiment of a review depends on the aspect group in positive and negative sentences, i.e., significance of product's aspects for customers. Let the number of aspect groups is $g/2$, then the number of independent sentimental variables g . According to the methodology described in [13] the equation (3) for calculating the significance of variable m is:

$$Sign_m = \frac{\sum_{j=1}^{k_m} \left(E_{m,j} - \sum_{i=1}^{q_{m,j}} E_{m,j,i} \cdot \frac{Q_{m,j,i}}{Q_{m,j}} \right)}{\sum_{l=1}^g \sum_{j=1}^{k_l} \left(E_{l,j} - \sum_{i=1}^{q_{l,j}} E_{l,j,i} \cdot \frac{Q_{l,j,i}}{Q_{l,j}} \right)} \cdot 100\% \quad (3)$$

where k_l – number of nodes that were split by attribute l , $E_{l,j}$ – entropy of the parent node, split by attribute l , $E_{l,j,i}$ – subsite node for j , which was split by attribute l , $Q_{l,j}$, $Q_{l,j,i}$ – number of examples in the corresponding nodes, $q_{l,j}$ – number of child nodes for j parent node.

4. Real Data Experiments

Effectiveness evaluation of the developed approach was performed on the data obtained from 635,824 reviews of hotels and resorts in

Russian. The reviews were collected from a popular Internet resource for the period of 2003-2013. The initial structure of the collected data consisted of the following fields: hotel name; country name; resort name; date of visit; opinion of the hotel; author evaluation of food; author evaluation of service; review number. The data was preliminarily processed and loaded into the database SQL Server 2012.

To classify segments, we used a binary scale (negative and positive) on the hypothesis that the absence of negative is positive. A training sample of positive and negative opinions was created using the collected data on the author's evaluation of accommodation, food and service. The Internet resource *tophotels.ru* uses a five-point grading scale. A review can have a maximum total of 15 points, a minimum of 3 points. The training sample included 15,790 negative reviews that had awarded 3 and 4 points, and 15,790 positive reviews that had awarded 15 points. We did not use author evaluation for further data processing. The marking of the remaining 604,244 reviews was carried out using a trained classifier.

For the purpose of effectively creating a sentiment classifier, we evaluated the accuracy of the classification of machine learning algorithms and some peculiarities of their structure (Table 1). The criterion *Accuracy* (ratio of the number of correctly classified examples to their total number) was used to assess classification accuracy. Accuracy evaluation was performed on two sets of data. The first set (Test No. 1) represented a training sample consisting of strong positive and strong negative opinions. It was tested by cross validation by dividing the data into 10 parts. The second set (Test No. 2) included reviews covering different points and was marked manually (497 positive and 126 negative

reviews). It was used only for the accuracy control of the classifier that had been trained on the first data set.

To assess the influence of the negative particles “not” and “no”, we used tagging; for example, the phrase “not good” was marked as “not_good”, and was regarded by the classifier as one word. This technique allowed the increase of sentiment classification accuracy.

Table 1. Results of methods for sentiment classification

Machine learning methods	Vector	Test No. 1	Test No. 2
SVM (linear kernel)	Frequency	94.2%	83.1%
SVM (linear kernel)	Binary	95.7%	84.1%
NB	Binary	96.1%	83.7%
NB	Frequency	97.6%	92.6%
NB (exceptional words)	Frequency	97.7%	92.7%
Bagging NB	Frequency	97.6%	92.8%
NB (tagging “not” and “no”)	Frequency	98.1%	93.6%

For the marking of reviews and the Sentiment Analysis, we created a classifier on the basis of the NB method, with frequency vectors as attribute space, and with the use of lemmatization and tagging of the negative particles “not” and “no”.

Using the algorithm we had developed we extracted from all reviews the key words that were divided into seven basic aspect groups (see Figure 4): *beach/swimming pool, food, entertainment, place, room, service, transport*. The following step was extracting and marking sentences with words from aspect groups by sentiment. However, not all sentences with aspects have a clearly expressed sentiment; therefore, the sentences which do not show a clearly expressed sentiment were filtered out.

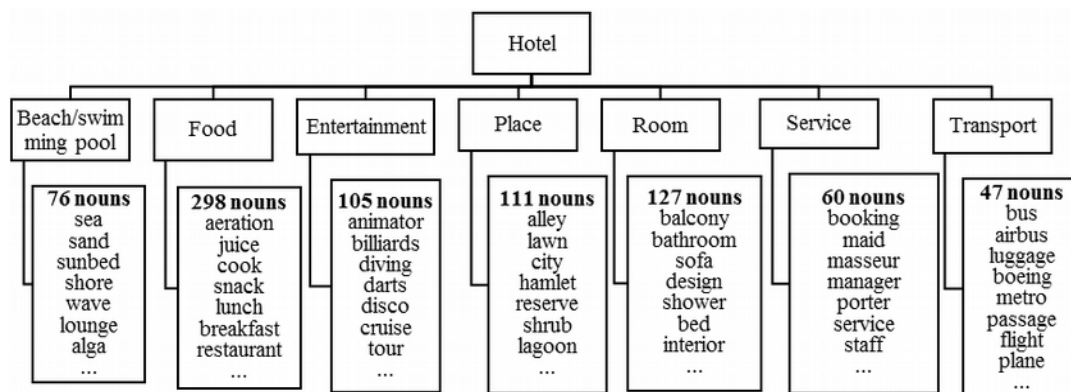


Figure 4. Aspect groups of the object “hotel”

We will give an example of our qualitative and quantitative research for two 5 star hotels “A” (1,692 reviews) and “B” (1,300 reviews) located in the resort Sharm el-Sheikh (63,472 reviews) in Egypt. First, we will describe our quantitative research of consumer satisfaction dynamics, then we will compare this with the average satisfaction in the whole resort, detect negative trends in the different hotel aspects and identify problems in the quality of hotel services.

The dynamics of customer satisfaction is represented in Figure 5. Concerning Hotel “A”, there is a positive upward satisfaction trend beginning in 2009; it reaches the average resort level in 2013. Concerning Hotel “B”, in 2012 there was a sharp satisfaction decline and a similarly sharp increase in 2013. We can also notice this trend in a monthly schedule (Figure 6). Satisfaction decrease for Hotel “B” started in June 2012 and stopped in October 2012. Then, customer satisfaction with Hotel “B” grew to a level that was higher than the average resort level, being ahead of its competitor – Hotel “A”.

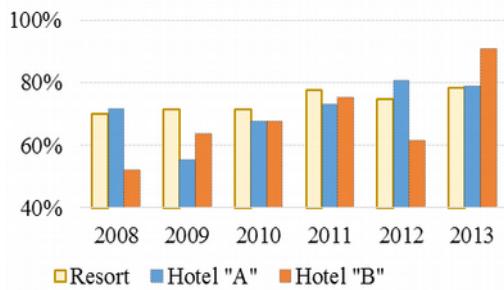


Figure 5. Yearly dynamics of customer satisfaction.

To find reasons for the Hotel “B” satisfaction decrease, we will examine the diagrams in Figure 7. We can see that in 2012, Hotel “B” on average was second to Hotel “A” in such aspects as “Room” ($\Delta 12\%$), “Place” ($\Delta 8\%$), “Service” ($\Delta 5\%$), “Beach/swimming pool” ($\Delta 3\%$) and “Entertainment” ($\Delta 3\%$). Besides, in 2012, Hotel “B” had more registered cases of food poisoning as well as cases of theft in August 2012. We should also note that one of the reasons of client dissatisfaction with Hotel “B” as a place was the beginning of the renovation of the hotel building and the rooms. These measures, however, were rewarded in 2013, when customer satisfaction with Hotel “A” aspects equaled the average resort level.

In 2013, customer satisfaction with Hotel “B” exceeded the average level in all aspects (Figure 8). Customer satisfaction with Hotel “A” dropped to lower than average values in such aspects as “Service” ($\Delta 3\%$), “Food”

($\Delta 3\%$), “Beach/swimming pool” ($\Delta 3\%$) and “Transport” ($\Delta 4\%$). The manager of Hotel “A” could be advised to direct efforts to increase the quality of all aspects, but would this be the most effective solution? Which aspects are the most significant for the customer and should consequently be improved in the first place? Is it possible to offset the dissatisfaction with the service, for example, by healthier food or an animated evening performance and achieve client satisfaction? A qualitative research of the Sentiment Analysis results can give answers to these questions.

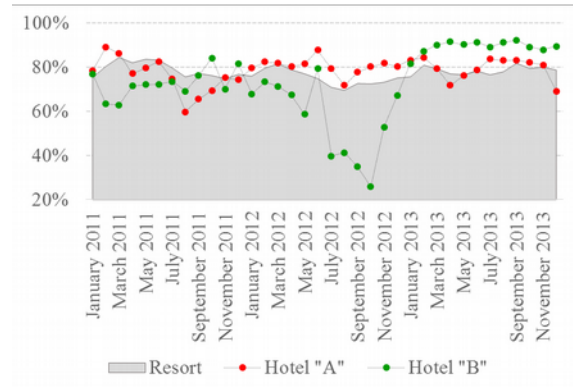


Figure 6. Monthly dynamics of customer satisfaction.

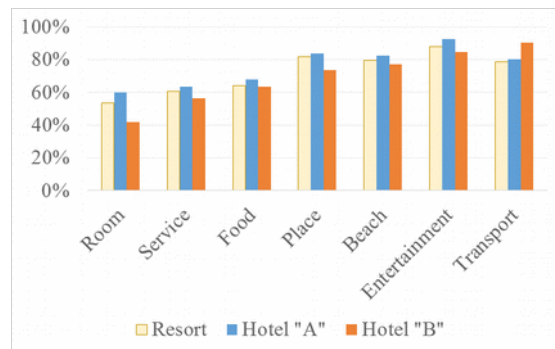


Figure 7. Comparison of customer satisfaction by aspects in 2012.

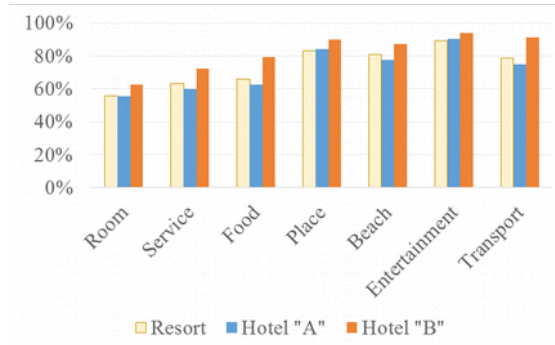


Figure 8. Comparison of customer satisfaction by aspects in 2013

Decision trees were created using algorithm C4.5 and Data Mining tool – Deductor [13].

The first step was creating a tree for the total sample of the reviews on the given resort to detect general principles. Constructed decision tree is presented on Figure 9. Extracted rules that have a confidence >80% are represented in Table 3. The second step is developing decision trees for the sample of Hotel "A" and Hotel "B" reviews to identify principles on the hotel level. Aspect significance is represented in Table 2.

Table 2. Significance of product aspect groups

Aspect group	Sentiment of mention	Significance values		
		Resort	Hotel "A"	Hotel "B"
Service	Negative	34.8%	60.2%	-
	Positive	0.7%	-	-
Food	Negative	30.3%	27.2%	30.3%
	Positive	16%	-	-
Entertainment	Negative	-	-	-
	Positive	8.5%	12.7%	12.4%
Room	Negative	4%	-	57.3%
	Positive	2.1%	-	-
Beach/swimming pool	Negative	0.2%	-	-
	Positive	2.5%	-	-
Place	Negative	-	-	-
	Positive	1%	-	-
Transport	Negative	-	-	-
	Positive	-	-	-

Analyzing values of aspect significance (Table 2), we can say that the main factors of consumer dissatisfaction are a low service level, problems with food, and complaints about the hotel rooms. The most critical aspect for Hotel "B" is "Room". Without negative opinions on the aspect "Room", the reviews would be positive with a probability of 95.5% (Rule No. 10, Table 3). That is why the

performed repair work facilitated a significant increase of consumer satisfaction. The most critical aspect for Hotel "A" is "Service", which corresponds with the findings for the resort as a whole.

The aspects which are significant both for the resort and for the two hotels and contributing to customer satisfaction are good food and amusing entertainment activities. The combination of these aspects can counterbalance negative emotions from the service or complaints concerning hotel rooms and leave the client with a favorable impression of the time spent in the hotel (Rules No. 5, 7, 11, Table 3). We should note that positive opinions about service, beach/swimming pool or place do not have a powerful influence on sentiment. That means the consumer a priori awaits a high-level service, well-kept place and beach/swimming pool as a matter of course.

Our qualitative research enabled us to detect the main ways for Hotel "A" to increase customer satisfaction (see Table 3). The problematic aspects identified in the course of our quantitative research correspond to the most significant aspects detected during the qualitative research stage. A search for alternative aspects that can lead to customer satisfaction in the presence of negative opinions about the significant aspects "Service" and "Food" was carried out. To accomplish this, the rules (see Table 3) containing negative sentiment in problem aspects, but which eventually lead to a positive review, were filtered out by the decision tree.

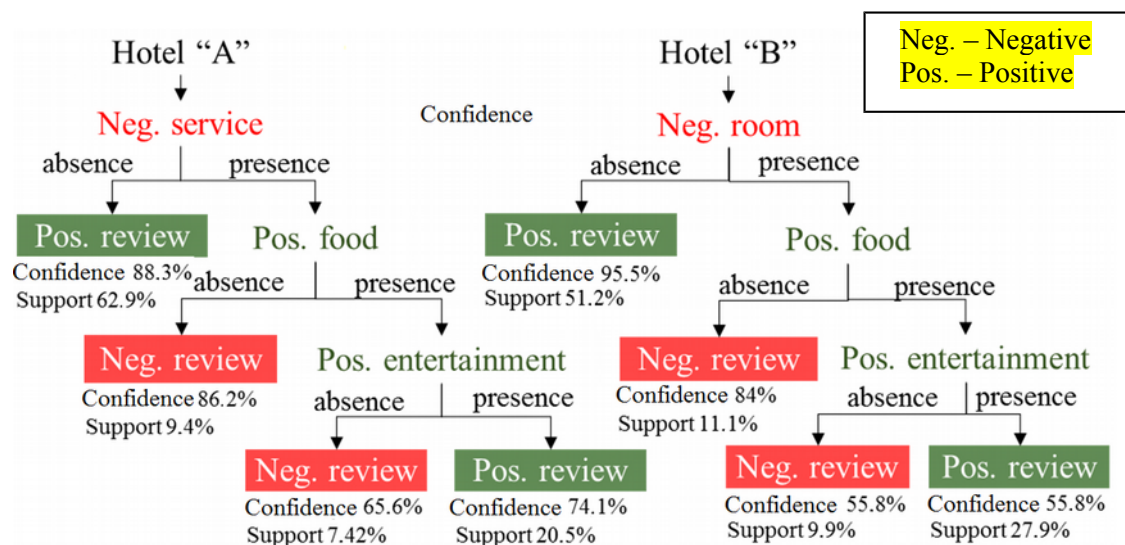


Figure 9. Decision trees for hotels

Table 3. Rules extracted by using decision trees.

#	Rules	Support	Confidence
<i>Extracted rules on resort reviews</i>			
1	$\text{Food}^+ \cap \overline{\text{Service}}^- \cap \overline{\text{Food}}^- = \text{Positive review}$	37.2%	97.4%
2	$\text{Food}^+ \cap \overline{\text{Service}}^- \cap \overline{\text{Food}}^- \cap \text{Beach}^+ = \text{Positive review}$	11%	86.2%
3	$\text{Food}^+ \cap \overline{\text{Service}}^- \cap \overline{\text{Food}}^- \cap \overline{\text{Room}}^- = \text{Positive review}$	10.6%	83.9%
4	$\overline{\text{Food}}^+ \cap \overline{\text{Service}}^- \cap \overline{\text{Entertainment}}^+ = \text{Negative review}$	6.9%	92.3%
5	$\text{Food}^+ \cap \overline{\text{Service}}^- \cap \overline{\text{Food}}^- \cap \overline{\text{Entertainment}}^+ = \text{Positive review}$	5.8%	88.4%
<i>Extracted rules on Hotel "A" reviews</i>			
6	$\overline{\text{Service}}^- = \text{Positive review}$	62.9%	88.3%
7	$\text{Food}^+ \cap \overline{\text{Service}}^- \cap \overline{\text{Entertainment}}^+ = \text{Positive review}$	20.5%	74.1%
8	$\overline{\text{Food}}^+ \cap \overline{\text{Service}}^- = \text{Negative review}$	9.4%	86.2%
9	$\text{Food}^+ \cap \overline{\text{Service}}^- \cap \overline{\text{Entertainment}}^+ = \text{Negative review}$	7.2%	65.6%
<i>Extracted rules on Hotel "B" reviews</i>			
10	$\overline{\text{Room}}^- = \text{Positive review}$	51.2%	95.5%
11	$\text{Food}^+ \cap \overline{\text{Room}}^- \cap \overline{\text{Entertainment}}^+ = \text{Positive review}$	27.9%	81%
12	$\overline{\text{Food}}^+ \cap \overline{\text{Room}}^- = \text{Negative review}$	11.1%	84%
13	$\text{Food}^+ \cap \overline{\text{Room}}^- \cap \overline{\text{Entertainment}}^+ = \text{Negative review}$	9.9%	55.8%

In order of preference, the manager of Hotel "A" should first of all make decisions on increasing the service quality, and then on increasing the quality of food and beach/swimming pool maintenance. Transport problems – concerning flights, early check-in, and baggage storage – are not significant and can be solved within the frames of service improvement. The process of service quality increase can take much time; organizing entertainment and animated programs together with solving problems in connection with restaurant service and beach/swimming pool maintenance can serve as immediate measures to increase client satisfaction. Specification of managerial decisions can be performed on the basis of the information on existing problems contained in negative reviews. The extracted opinions on aspects can be used by hotel managers for improve specific service areas.

5. Conclusion

Poor quality of products and services contributes to a decrease of customer satisfaction. On the other hand, under the conditions of stiff competition, there are no barriers for the consumer to change the supplier of goods and services. All these things can cause loss of clients and a decrease of a company's efficiency indexes.

Therefore, maintaining high-quality standards should be provided by effective managerial decisions and based on opinion mining as a feedback.

The suggested conception of decision support based on the developed approach of text data processing and analysis allows performing quantitative and qualitative surveys of customer satisfaction using information technology in the form of computer-aided procedures, and making effective managerial decisions on product quality management. The present conception allows effective reduction of labor intensity of customer satisfaction research that makes it available for use by a wide range of companies.

A prototype of IDSS was developed on the basis of the suggested conception. The performed experiment has proved its efficacy for solving real problems of quality management and consistency of the results obtained. IDSS enables companies to make decisions on quality control based on analytical processing of text data containing implicit information on client satisfaction.

Future research on the given topic can be devoted to automatic annotating of text data, representing text amount of review in the form of a summary, and extracting useful and unique information.

REFERENCES

1. ISO 9000:2008. **The quality management system. Fundamentals and vocabulary.**
2. ISO10004:2010. **Quality management. Customer satisfaction. Guidelines for monitoring and measuring.**

3. GRÄBNER, D., M. ZANKER, G. FLIEDL, M. FUCHS, **Classification of Customer Reviews based on Sentiment Analysis**, Proceedings of the International Conference in Helsingborg, Springer Vienna, 2012, pp. 460-470.
4. TABOADA, M., J. BROOKE, M. TOFILOSKI, K. D. VOLL, M. STEDE, **Lexicon-Based Methods for Sentiment Analysis**, Computational Linguistics, June 2011, vol. 37(2), pp. 267-307.
5. JO, Y., A. OH, **Aspect and Sentiment Unification Model for Online Review Analysis**, Proceedings of the fourth ACM international conference on Web search and data mining (WSDM '11), ACM New York, Feb. 2011, pp. 815-824.
6. LU, B., M. OTT, C. CARDIE, B. TSOU, **Multi-aspect Analysis with Topic Models**, Proceedings of the 2011 IEEE 11th International Conference on Data Mining Workshops (ICDMW '11), Dec. 2011, pp. 81-88.
7. BLEI, D. M., A. Y. NG, M. I. JORDAN, **Latent Dirichlet allocation**, Journal of Machine Learning Research, Jan. 2003, vol. 3 (4-5), pp. 993-1022.
8. YU, J., Z.-J. ZHA, M. WANG, T.-S. CHUA, **Aspect Ranking: Identifying Important Product's aspects from Online Consumer Reviews**, Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies (HLT '11), June 2011, pp. 1496-1505.
9. HORVÁTH, L., I. RUDAS, **New Method for Intellectual Content Driven Generic Product Model Generation** 2014 IEEE International Conference on Systems, Man and Cybernetics (SMC). IEEE, 2014. pp. 1660-1665.
10. RUDAS, I., **Cloud Technology-Based Education with Special Emphasis on Using Virtual Environment** Proceedings of the 14th WSEAS Conf. 2014.01.29-2014.01.31. Cambridge. USA, p. 23.
11. THOMSEN, J., E. ERNST, C. BRABRAND, M. SCHWARTZBACH, **WebSelf: A Web Scraping Framework**, Proceedings of the 12th international conference on Web Engineering (ICWE 2012), July 2012, pp. 347-361.
12. YUSSUPOVA, N., D. BOGDANOVA, M. BOYKO, **Applying of Sentiment Analysis for Texts in Russian Based on Machine Learning Approach**, IMMM 2012, Venice, Italy, pp. 8-14.
13. EBERT, S., N. T. VU, H. SCHÜTZE, **CIS-positive: Combining Convolutional Neural Networks and SVMs for Sentiment Analysis in Twitter** In: Proceedings of the 9th International Workshop on Semantic Evaluation. SemEval 2015.
14. TURNEY, P. D., Y. NEUMAN, D. ASSAF, Y. COHEN, **Literal and Metaphorical Sense Identification through Concrete and Abstract Context**. In: EMNLP., 2011, pp. 680-690.
15. DURRANI, N., H. SCHMID, A. FRASER, P. KOEHN, H. SCHÜTZE, **The Operation Sequence Model – Combining N-Gram-based and Phrase-based Statistical Machine Translation**. Computational Linguistics, vol. 41(2), 2015, pp. 157-186.
16. YUSSUPOVA, N., M. BOYKO, D. BOGDANOVA, A. HILBERT, **A Decision Support Approach based on Sentiment Analysis Combined with Data Mining for Customer Satisfaction Research**, The International Journal on Advances in Intelligent Systems is Published by IARIA vol 8, no 1&2, 2015.