Sentiment Prediction Based on Analysis of Customers Assessments in Food Serving Businesses

Zoltan Geler[a], Miloš Savić[b], Brankica Bratić[b], Vladimir Kurbalija[b], Mirjana Ivanović[b], Weihui Dai[c]

[a]*Faculty of Philosophy, University of Novi Sad, Novi Sad, Serbia*

[b]*Faculty of Sciences, University of Novi Sad, Novi Sad, Serbia*

[c]*School of management, Fudan University, Shanghai, China*
Corresponding author: Zoltan Geler, zoltang@ff.uns.ac.rs

Human activities and behaviour in different domains are usually influenced by other people's actions and opinion. Nowadays, it is evident that there is a growing research interest in *sentiment analysis, evaluation and prediction*. Content from web sources and social media is frequently used when people want to see others' opinion about different things. Our research is focused on ML-based sentiment analysis of food services reviews data. The comparison of several regression models with regards to prediction of customer satisfaction of restaurant and food services is presented. The experimental data collected from food serving businesses located in Shanghai Lujiazui Commercial Zone includes keywords extracted from the customers' written reviews. Additionally, the data are spatially labelled enabling to conduct separate analyses for different geographical regions. As a conclusion, the keywords extracted from the customer's reviews were suitable for the prediction of three observed satisfaction criteria: food taste, service, and environment.

Keywords: sentiment analysis; opinion mining; regression; machine learning

Introduction

Most human activities, interactions and behaviour are influenced by opinions of other people. In numerous domains our choices are based on others' perceptions and evaluations of a particular phenomenon in the domain. People often seek the opinion of others when they need to decide. In the last two decades research in sentiment analysis, evaluations, and prediction, as well as opinion mining, generally is growing constantly. This research usually intertwines social, computer and management sciences in order to increase positive effects in business and society as a whole (B. Liu, 2012). Distilling textual information contained in diverse web sites is not an easy task. Consequently, the rapid development of innovative techniques and methods for better processing huge volume of customers' opinion data available in the social networks and media is evident (Song et al., 2020; Zhang et al., 2020).

Organizations as well as individuals frequently use the content from social media when they want to see someone's opinion on a topic, such as buying a consumer product, choosing a good restaurant, booking a holiday apartment, etc. In this way they are not limited just to asking friends and family for opinions, but they can also consider many opinions and reviews available on the Web (Oh et al., 2020).

There are two groups of prevalent approaches for sentiment analysis (Saad & Saberi, 2017). The first group implies multiple techniques based on machine learning (ML). These techniques support different activities with available data that give more accurate information about the polarity of sentiments. Our research results presented in this paper are based on application of ML techniques. The second group is concentrated on the lexicon-based approach which is a linguistically inclined method (Badica & Vladutu, 2018; Colhon et al., 2014).

Besides the already mentioned application to review-related data, ML-based sentiment analysis and opinion-mining are applied in many other domains as well (Bakshi et al., 2016). These methods can be applied to detect heated or antagonistic language in written communication (Spertus, 1997). They can also be used as filters for ads content, banning inappropriate ads and bringing up the ones with positive sentiments (X. Jin et al., 2007). Some analyses explored the possibility of using sentiment analysis in information extraction for discarding information in subjective sentences (Riloff et al., 2005). Sentiment analysis found its place in scientometrics as well, for example in citation analysis where one can determine if the citation is related to the supporting evidence or to the research that the authors dismiss (Piao et al., 2007). Various methods of sentiment analysis can be applied in the analysis of public mood and stock returns (Chang, 2020).

The focus of this paper is on ML-based sentiment analysis of food services and restaurants reviews data. The necessity for such analyses increases with the growth of social media popularity (Grosse et al., 2015). Consequently, many researchers have already devoted themselves to this task. Topics in this field include prediction of the sentiment based on plain text restaurant reviews (Kang et al., 2012), examination of the influence of review sentiments on restaurant star ratings (Gan et al., 2017), examination of the influence of a person’s ethnic culture on the restaurant reviews (Nakayama & Wan, 2019), identification of restaurant features based on its reviews (Yu et al., 2017), etc. Many of these topics emerged very recently, making the whole field challenging.

In this paper we present a novel method for predicting the customers’ satisfaction with restaurants and food services. The method is based on regression models trained from datasets containing spatially labelled instances. One instance describes an individual restaurant (or some other food business object) by a set of keywords extracted from customers’ written reviews. The main innovative characteristic of our method is that various competitive regression models are not formed considering only the whole training dataset, but also for spatial clusters of instances identified by an unsupervised clustering technique. More concretely, our sentiment prediction approach relies on the expectation-maximization clustering algorithm to identify spatial clusters in the training dataset, additionally employing the 10-fold cross validation procedure to estimate the number of spatial clusters in the training data (i.e., the number of spatial clusters is not specified in advance). Then, regression models are trained and evaluated for each of identified spatial clusters and compared to the global models trained on the whole training dataset. In this way, our sentiment prediction approach enables spatially-personalized predictions taking specificities of different geographical regions into account.

The rest of the paper is organized as follows. Section 2 gives an overview of the related work in the field. The methods used for analyses are summarized in Section 3. Section 4 describes the data used in the experiments and presents the obtained results. Finally, the conclusions and remarks are given in Section 5.

Related Work

The public opinion shared through social networks and media turns out to be a crucial reason for choosing a particular restaurant i.e. a food serving place. The analysis of meal experience gains much attention for several reasons:

- it refers to a series of events a guest experiences when eating out (Kotschevar & Withrow, 2007),
- it represents an event in everyday life of humans (Mäkelä, 2000),
- all individuals have their own experience of meals (Warde & Martens, 2000),
- all guests' feelings (from arrival to leaving of the restaurant) should be considered as a part of meal experience (Heung & Gu, 2012; Kotschevar & Withrow, 2007),
- all events before and after dinning can also influence the total experience (N.-H. Jin et al., 2011).

The overall costumers' satisfaction depends on numerous factors. Some authors highlight food quality, physical environment and service as the major components (Dulen, 1999; Susskind & Chan, 2000). A similar study emphasizes: service quality, product quality, and price as well as situational/environmental factors and personal factors (Choi & Chu, 2001). As expected, the food quality is the most important attribute of restaurant experience (Sulek & Hensley, 2004), but it is extremely important to satisfy all the other customers' expectations (Peri, 2006).

As already mentioned, the quality of service is an important part of meal experience. However, it is hard to evaluate because the quality depends not only on service outcome, but also on the process of service delivery (Markovic et al., 2010). However, three important quality elements can be identified and measured: environmental elements (e.g. music, interior design, distance between tables), employees (e.g. professional skills, kindness, reliability) and customers (e.g. interaction with other customers) (Wu & Liang, 2009).

Understanding and handling all the important characteristics of restaurant service quality properly requires an adequate measurement methodology. There are several instruments developed for this purpose. One of the first and well-known methodologies is SERVQUAL proposed by Parasuraman et al. (1985). The instrument consists of two sections, where each contains 22 items. The first section measures the customers' expectations, while the second measures the customers' perceptions of the service. Another well-known instrument called DINESERV was presented in (Stevens et al., 1995). This instrument is considered as a simple and reliable tool for assessing customers' view of a service quality. DINESERV instrument consists of 29 items which are measured on a 7-point scale. Both instruments give insight into five main dimensions of service quality: tangibles, reliability, responsiveness, assurance and empathy.

All recent studies based on these instruments gave generally expected results. The research presented in (Saad Andaleeb & Conway, 2006) showed that responsiveness of the employees, price and food quality are the most significant factors for customers' satisfaction. A similar study by Kim et al. (2009) points out five dimensions which significantly influence the customers' satisfaction: food quality, service quality, price and value, atmosphere and convenience. The research presented in (Wu & Liang, 2009) reported that customer satisfaction was positively influenced by restaurant employees. Moreover, the results of the study conducted by Liu & Jang (2009) underline the following aspects for customer satisfaction: food quality, service reliability, environmental cleanliness, interior design, and neat and well-dressed employees.

Many researchers and studies point out the importance of word-of-mouth communication to the overall restaurant reputation (Gersch et al., 2011; Kimes, 2008; Ladhari et al., 2008; Pantelidis, 2010). However, with the rise of social media and different online forums the traditional word-of-mouth transformed to the more popular and accessible electronic-word-of-mouth (Riedl & Konstan, 2002) also commonly known as word-of-mouse (Gersch et al., 2011). The necessity of maintaining a good reputation on the Internet was noticed at the beginning of the century (Gelb & Sundaram, 2002). Some authors argue that the failure to keep up with technological changes is the main reason for the failure of restaurants (Camillo et al., 2008), while others just highlight the importance of analysis of electronic guest comments (Mhlanga, 2015).

Recently, text and data mining was used in various service industries to analyse he customers' opinions: airline review data (Hong & Park, 2019), hotel reviews (Chanwisitkul et al., 2018), camps rating (Senožetnik et al., 2019), opinions in app store user reviews (Genc-Nayebi & Abran, 2017) after-sale service analysis (Y. Dai et al., 2020), enrichment of user profiles from user's posts (Benkhelifa et al., 2020) etc. The same trend is also identified in restaurants' opinion mining, where different techniques of data mining were applied to the freely available pool of customers' comments (Gan et al., 2017; Jia, 2018; Yu et al., 2017).

This study will also focus on the extraction of costumers' opinions from online comments, while paying special attention to the geo-locations of the food serving businesses.

Methods

The main idea of this research is to explore regression models trained on spatially-labelled data to predict customer satisfaction from textual comments made on social media. The study is based on a dataset containing the most frequent keywords in online comments of different food serving points. Food serving points in our case are restaurants and other different food serving businesses located in Shanghai Lujiazui Commercial Zone (Lin et al., 2019). Each instance in the dataset includes the following three numeric attributes reflecting customer satisfaction with the corresponding food serving points: the average satisfaction with meal quality (TASTE attribute), service quality (SERVICE attribute) and interior design (ENVIRONMENT attribute). These attributes were individually expressed on a scale from 1 (lowest satisfaction) to 10 (highest satisfaction). Additionally, the dataset contains the GPS coordinates of the considered food serving points that enable their division into spatially-bounded clusters.

A regression model predicts the value of a numeric attribute from the values of other (usually numeric) attributes. To perform regression based on keywords in textual comments we have transformed them into appropriate real-valued vector representations (detailed explanations are given in the next section). Various regression models implemented in the WEKA machine learning package (Frank et al., 2009) were trained to predict TASTE, SERVICE and ENVIRONMENT attributes from the experimental dataset. We considered the following methods:

(1) SMOreg (SMO) – regression based on support vector machines. The main idea of SMOreg is to find a function minimizing absolute prediction errors, where errors below a predefined threshold are discarded (i.e., the loss function is epsilon-insensitive) and, simultaneously, the flatness of the function is maximized.

(2) Random forest (RandF) – regression based on classical random forests. A random forest is an ensemble of regression trees learned from bootstrapped samples of the training data. Additionally, the random forest method employs so-called feature bagging meaning that a random subset of features is selected for learning individual regression trees. The final regression result is the average prediction of the regression trees in the ensemble.
(3) Random tree (RandT) – regression based on a regression tree that is constructed considering a fixed number of randomly selected attributes at each node. The information gain metric is used to grow unpruned regression trees. The implementation provided by WEKA supports predictions based on a hold-out set that is back fitted into the learned random tree.
(4) REPTree (REPT) – a fast regression tree learner. This algorithm uses either information gain or information variance as the training dataset splitting criterion when growing the regression tree. The WEKA implementation of the algorithms additionally supports reduced-error pruning and backfitting.
(5) M5P – regression based on model trees constructed by the M5 algorithm. A model tree is a decision tree whose leaves are linear regression models. This means that the regression result is the outcome of the linear model selected by the decision tree.
(6) MP – regression based on classic neural networks with fully connected layers (multi-layer perceptrons). Neural networks in WEKA are learned using the standard backpropagation algorithm with the sigmoid function as the activation function. Early stopping to prevent overfitting is also supported.

Our selection of regression models includes the three most common regression approaches taken by both researchers and industrial practitioners: support vector machines, random forests and neural networks. Our selection additionally contains both linear and non-linear models, as well as singleton and ensemble models. The selection of discrete tree-based models is motivated by the fact that in the context of our work they enable explainable recommendations.

The predictive power of the trained regression models was evaluated using mean absolute error (MAE) and root mean square error (RMSE) measures. Although those measures are interchangeable when comparing two different regression models, we have used both since MAE is easily interpretable, while RMSE is more sensitive to a small number of relatively large errors when MAE values of two compared methods are approximately equal. Let M denote a regression model, X a sequence of N real-valued vectors where X_i denotes the i-th vector in the sequence, and let Y be a sequence of N real values where Y_i represents ground truth prediction for X_i. MAE and RMSE for M are then defined as follows:

$$MAE(M) = \frac{1}{N}\sum_{i=1}^{N}|Y_i - M(X_i)| \quad (1)$$

$$RMSE(M) = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(Y_i - M(X_i))^2} \quad (2)$$

Lower values of MAE and RMSE indicate better (more precise) regression models. MAE and RMSE for the selected regression models were estimated using the stratified 10-fold cross-validation procedure that was repeated 10 times. This means that each of the regression models was trained 100 times for 100 different partitions of the dataset into the training part encompassing 9 stratified folds (90% of the dataset) and the test part consisting of one stratified fold (10% of the dataset). Then, MAE and RMSE for a regression model were obtained by averaging the MAE/RMSE scores obtained on 100 test folds. The stratified cross-validation was employed in order to ensure approximately equal distributions of the target variables in formed folds and to obtain more robust estimates of the error measures.

The stratified 10-fold cross validation was also used to compare different regression models. We say that a regression model *A* wins against a regression model *B* if (1) the average value of MAE/RMSE for *A* is lower than the average value of MAE/RMSE for *B* and (2) the paired t-test detects statistically significant differences in the average MAE/RMSE scores obtained from 100 test folds (scores are paired per fold). For each regression model we determined its Win-Lose (W-L) score (the number of statistically significant wins minus the number of statistically significant losses) considering 3 predicted attributes (TASTE, SERVICE and ENVIRONMENT) and two different vector-based representations of keywords (binary and Inverse Document Frequency (IDF), see the next section of the paper). This means that 6 different instances of the same model (one for each particular combination of an attribute that is being predicted and a vector-based representation of keywords that is used) were compared against 30 instances of other considered models (5 other models each with 6 instances).

The spatial clusters of food serving objects were determined using the Expectation-Maximization (EM) clustering algorithm implemented in WEKA. The WEKA implementation of EM employs the 10-fold cross validation procedure to estimate the number of clusters in data. For our dataset WEKA EM detects 4 clusters (labelled from C0 to C3), where one cluster (C1) contains less than 2% of food serving points. The other three clusters contain 57% (C0), 31% (C2) and 10% (C3) of the instances in the dataset. Consequently, we trained, evaluated, and compared the selected regression models considering both the whole dataset and spatial clusters of non-negligible size (C3, C2, and C0).

Dataset and Results

Preprocessing of the Original Dataset

The original dataset consists of 1600 instances, each representing a customer service object i.e. a food serving point such as a restaurant, store, hotel, etc. The dataset has 19 attributes: object type, name, address, average price, taste score, environment score, service score, the number of comments, the number of high praises, and the 10 most frequent keywords that appeared in the object's comments, each keyword being a separate attribute. In the following text, the instance's 10 most frequent keywords will be referred to as *top-10 keywords list*. Figure 1 shows the distribution of instances over different object types. As can be seen, most of the instances are restaurants and daily food services.

[Figure 1 near here]

In order to make the dataset suitable for our research goal, we had to reorganize the data in an appropriate way. Since the *name* and the *address* attributes do not hold research relevant information, we started the data pre-processing with the elimination of these two attributes and reorganized the keyword related attributes. Namely, the 10 last attributes, that held most frequent keywords, were not very suitable for many machine learning algorithms, since their values were unstructured strings. To make them interpretable, we transformed the 10 textual attributes to a bag-of-words representation. In this representation, each keyword that appeared as a value of the 10 attributes became an attribute itself, whose values represent the frequencies of the given keyword among different instances. So, the 10 attributes were transformed to a set of new attributes, whose cardinality is equal to the number of distinct keywords. After this transformation we ended up with 517 new attributes in total.

After the transformation of the attributes set, we faced the question of the new attributes' values. As already said, the values should represent keyword frequencies, but the frequencies could be represented in different ways. Our first method was based on using Inverse Document Frequency (IDF) value of a given keyword for all instances that had the particular keyword in their top-10 keywords list. To all the other instances the value 0 was assigned. The IDF value of a given keyword is calculated based on the formula shown in Equation (3), where k represents a keyword for which the IDF is calculated, N is the total number of instances and n_k is the number of instances in which k is listed among the 10 most frequent keywords. In the rest of the paper, the dataset that was created by using this method will be referred to as the *IDF dataset*.

$$IDF(k) = log\frac{N}{n_k} \tag{3}$$

For the purpose of making adequate transformations of the original dataset to be used by the planned ML algorithms we have applied one more method for calculating frequency values. This method is even simpler than the first one. The value 0 was assigned to instances which did not have the given keyword in their top-10 keywords list and the value 1 was assigned to the instances which did have the keyword in their top-10 keywords list. We named this new dataset the *01 dataset*.

After these pre-processing steps, we ended up with two distinct datasets: the IDF dataset and the 01 dataset. In the rest of this section we will give a detailed presentation of the results of the analyses conducted upon both. It is worth mentioning that both datasets are very sparse, since we use bag-of-words representation of keywords.

Experimental Results

In this section, we will present the results of the experiments on the whole dataset and the spatial clusters of non-negligible size (C3, C2, C0), both for the 01 and the IDF representation of the attributes. In all four cases, in order to identify the most suitable regression model, we will compare the average values and standard deviations of the mean absolute error (MAE) and the root mean square error (RMSE) for each combination of the considered models (SMO, RandF, RandT, REPT, M5P, MP) and the predicted attributes (TASTE, SERVICE, and ENVIRONMENT - denoted as ENVIRO).

Tables 1, 4, 7, and 10 present the average values and standard deviations of MAE obtained for the whole dataset, and the clusters C3, C2, and C0, respectively. The average values and standard deviations of RMSE are shown in Tables 2, 5, 8, 11. In these tables, the best (lowest) results are marked in bold and the worst (highest) are underlined.

The performance of the examined algorithms was also compared relying on the counts of statistically significant wins and losses obtained by applying the corrected resampled t-test with a significance level of 0.001 (Bouckaert & Frank, 2004). These results are presented in Tables 3, 6, 9, and 12.

A brief summary of the results is given at the end of this section.

Experimental Results - The Whole Dataset

From Table 1, it can be seen that the lowest average MAE value (0.541) is found for the combination of the RandF algorithm and the TASTE attribute, for both the 01 and the IDF dataset. The average error was the highest for the SERVICE attribute when applying the MP regression model (0.738 for the 01, and 0.798 for the IDF dataset). The differences between the results of the RandF and MP models are between 0.12 and 0.23. The results of the other algorithms do not differ by more than 0.09 from the average MAE values of the RandF model.

In general, the RandF model gave the best average MAE values and the MP model gave the worst results for all three attributes. In addition, the MP algorithm produced noticeably higher standard deviations than the other considered models (Table 1): the difference ranges from about twice as large (TASTE attribute, 01 dataset) to nearly 9.5 times larger (SERVICE attribute, IDF dataset).

Based on the graphical representation of the obtained average values and standard deviations of MEA given in Figure 2, it can be seen that there is no significant difference between the 01 and the IDF datasets.

[Table 1 near here]

[Figure 2 near here]

The results obtained for RMSE (Table 2) are consistent with those acquired for MAE: while the lowest average error was produced by the RandF model for the TASTE attribute (0.679), the MP model generated the worst results for the SERVICE attribute (0.895 and 0.955 for the 01 and the IDF datasets, respectively). The differences between the results of these algorithms are between about 0.13 and 0.24. The average RMSE values of the other models do not differ from the RandF model's by more than 0.09.

Figure 3 presents a graphical representation of the average values and standard deviations of RMSA given in Table 2. Similarly, as in the case of the MAE, there is no significant difference between the 01 and the IDF datasets.

Similar to MAE, the obtained RMSE results indicate that RandF is the best and MP is the worst model in predicting the values of the attributes. Again, the MP algorithm produced noticeably higher standard deviations than the other considered models (Table 2): the difference ranges from about 2.5 times as large (TASTE attribute, 01 dataset) to nearly 8.5 times larger (SERVICE attribute, IDF dataset).

[Table 2 near here]

[Figure 3 near here]

The counts of statistically significant wins and losses (Table 3) confirm the advantage of the RandF over the other considered models in predicting the values of the attributes, and highlight the general inferiority of the MP model.

[Table 3 near here]

Experimental Results - Cluster C3

As Table 4 shows, in the case of cluster C3, the lowest average MAE results were obtained with the RandF model when predicting the values of the TASTE attribute

(0.568 for the 01 dataset, and 0.569 for the IDF dataset), similarly as in the case of the whole dataset. Nonetheless, in this cluster, for the SERVICE and the ENVIRONMENT attributes, the SMO model produced slightly better average results than the RandF algorithm (for both the 01 and the IDF dataset).

The worst results were again generated by the MP algorithm. In this cluster, however, at the dataset level, instead of the SERVICE attribute (as in the case of the entire datasets), the highest average errors were achieved with the ENVIRONMENT attribute (0.739 for the 01 dataset, and 0.738 for the IDF dataset).

There are some other notable discrepancies in the results obtained over the entire datasets. First, the differences between the models' average MAE values are a bit smaller (less than 0.14). Secondly, the standard deviations of the MP model are not significantly different from the standard deviations of the other models - their values are not more than about 1.25 times higher (Table 4).

[Table 4 near here]

The average values and standard deviations of RMSE in cluster C3 are presented in Table 5. Clearly, the lowest average value (0.694) was achieved by the RandF model for both the 01 and the IDF dataset. However, when looking at the results of the individual attributes, we notice that in the case of the ENVIRONMENT attribute the SMO model slightly outperformed the RandF model.

For each attribute, the highest average RMSE value was generated by the MP model. In accordance with the MAE results, in cluster C3, among all combinations of models and attributes, the worst results were obtained with the MP model when predicting the ENVIRONMENT attribute (0.872 for the 01 dataset, and 0.871 for the IDF dataset) - instead of the SERVICE attribute as in the case of the entire datasets.

In this cluster, similarly as in the case of the MAE results, the differences between the models' average RMSE values are a bit smaller (less than 0.12) than on the entire datasets. Moreover, the standard deviations of the MP model are much smaller (compared to the entire datasets) and they don’t differ significantly from the deviations of the other algorithms - they are not more than 1.35 times larger (Table 5).

[Table 5 near here]

According to Table 6, the advantage of the RandF model over the other examined algorithms and the inadequacy of the MP model for the task of predicting the values of attributes considered in cluster 3 is also supported by the numbers of statistically significant wins and losses.

[Table 6 near here]

Experimental Results - Cluster C2

When it comes to cluster C2, the best MAE results still come from the RandF model (Table 7) with the TASTE attribute: 0.570 for the 01 dataset, and 0.569 for the IDF dataset. Moreover, this model produced the lowest average MAE values for the other two attributes, too, just like in the case of the entire datasets.

As before, the worst results were produced by the MP model and the ENVIRONMENT attribute (0.724 for both the 01 and the IDF dataset), in a similar manner as in cluster C3. The differences between the MP model’s and the other algorithms’ average MAE values are smaller than in the case of the entire datasets (the upper bound of the differences in this cluster is about 0.16, but when it comes to the whole dataset it is 0.23).

The MP model also stands out in terms of standard deviation, namely, it produced values that are higher than the deviations of the other models by up to almost 3 times (see the results for the TASTE attribute in Table 7).

[Table 7 near here]

When it comes to the average RMSE values (presented in Table 8), the smallest errors in cluster C2 were obtained by applying the RandF algorithm as before. However, it was not in combination with the TASTE attribute (like in cluster C3 and the entire datasets), but rather with the ENVIRONMENT attribute (0.692 for the 01 dataset, and 0.691 for the IDF dataset). In the same way as in the case of MAE results, this model dominates in the case of the TASTE and the SERVICE attributes, too.

The worst RMSE results show similarity to the values obtained over the whole datasets: they were produced by the MP model in combination with the SERVICE attribute (0.877 for both the 01 and the IDF dataset). The differences between the results of the RandF and MP models are between 0.13 and 0.18.

The standard deviation of RMSE is the largest in the case of the MP model (Table 8). The values obtained are slightly higher than those in cluster C3, but are significantly lower than the results generated over the entire datasets.

[Table 8 near here]

Considering the numbers of statistically significant wins and losses shown in Table 9, we can conclude that the superiority of the RandF model and the inferiority of the MP approach is confirmed in this cluster as well.

[Table 9 near here]

Experimental Results - Cluster C0

In terms of the best average MAE values, there is no deviation from the pattern observed within the other clusters or the entire datasets: they were achieved in the process of predicting the values of the TASTE attribute utilizing the RandF model (0.540 for both the 01 and the IDF dataset - see Table 10). The RandF algorithm gave the lowest average errors in the case of the other two attributes, too.

As before, the worst MAE results were produced by the MP model for all three attributes. In both the 01 and the IDF dataset, the highest average error was achieved with the SERVICE attribute (0.921). The differences with the results of the RandF model are the largest in this cluster: in the case of the SERVICE attribute they reach almost 0.35.

Looking at the standard deviation values presented in Table 10, it is easy to notice the significant difference between the MP model and the rest of the algorithms in the case of the SERVICE attribute. For both the 01 and the IDF dataset the MP model produced about 21-28 times higher deviations than the other models. These values are significantly larger than those observed in the other clusters and even than the results obtained for the entire datasets.

[Table 10 near here]

Just like for the other clusters and for the entire datasets, the RandF model proved to be the best predictor of all three attributes when it comes to RMSE (Table 11). The only exception was observed in cluster C3 where, when predicting the ENVIRONMENT attribute, the SMO model gave slightly better results. In cluster C0, the best results were obtained for the TASTE attribute (0.674 for the 01, and 0.675 for the IDF dataset), in the same manner as in the case of cluster C3 and the entire datasets.

The highest average RMSE values were observed when using the MP model to predict the values of the SERVICE attribute (1.071 for both the 01 and the IDF dataset).

The MP approach proved to be the worst solution in the case of the other two attributes as well. The differences between the results of the RandF and MP models are between 0.13 and 0.36.

The detected standard deviations of the MP model (Table 11) are much higher than the results of the other models in the case of RMSE, too. When it comes to the SERVICE attribute, the MP model generated about 18-24 times higher RMSE deviations than the other algorithms. Again, these values are significantly larger than those observed in the other clusters and even than the results obtained for the entire datasets.

[Table 11 near here]

As in the previous cases, the statistical comparison of the obtained MAE and RMSE values confirmed the advantage of the RandF model and the inferiority of the MP approach in the C2 cluster as well. Table 12 shows that the number of statistically significant wins is the highest in the case of the RandF model and that the MP algorithm is among the worst ones in this respect, too.

[Table 12 near here]

Experimental Results - Summary

Based on the presented results, as far as the performance of the examined regression models is concerned, there is no significant difference between the whole dataset and the three spatial clusters. The similarity of the result obtained for the whole dataset and the clusters is illustrated in Figures 4 and 5 which show the average values and standard deviations of MAE and RMSE for the 01 representation. The results obtained with the IDF representation do not differ significantly from the 01 results. A noticeable divergence can be noticed only in the case of the SERVICE attribute and the MP method.

[Figure 4 near here]

[Figure 5 near here]

The best average results were generated by the random forest algorithm (RandF) predicting the values of the TASTE attribute. The only exception was observed in the case of cluster C2: when it comes to RMSE, the RandF algorithm was slightly more successful in predicting the values of the ENVIRO attribute than the values of the TASTE attribute.

The worst results were always achieved by the multi-layer perceptron (MP) model in combination with the SERVICE or the ENVIRO attribute.

The advantage of the RandF model and the inferiority of the MP algorithm were also confirmed by the results of the corrected resampled t-test.

Conclusion

In this paper we analysed how the satisfaction of restaurant and food service customers can be predicted by using six different regression algorithms: SMOreg, random forest (RandF), random tree (RandT), REPTree, M5P and MP. The regression models were trained upon data that contained keywords extracted from the customer's written reviews. We were predicting three different aspects of customer satisfaction: satisfaction with food taste, with service and with environment. Additionally, the data contained geographical locations of each restaurant and food service. We used this information to create spatial clusters upon which we conducted separate analyses in order to determine the extent to which the results differ among different geographical areas.

The experiments showed that the best results are mostly achieved by the random forest algorithm, while the MP approach reported the worst results. Satisfaction with food taste turned out to be the easiest to predict, as that MAE and RMSE values were the lowest for the TASTE attribute. However, the results are not bad for the other two satisfaction criteria (service and environment) either. We can conclude that the keywords extracted from the customer reviews were suitable for the prediction of all three satisfaction criteria.

The results for the different spatial clusters were similar to the overall results, meaning that there was no significant difference between the results for different geographical locations. This result can be exploited for producing smaller and more effective regressors based on a particular cluster. Such regressors (trained on data from one cluster) will be significantly faster than regressors trained on the whole dataset, while the accuracy will be similar.

The results obtained in this paper could represent a good starting point for other researchers to consider the selection of adequate methods and perform experiments on other similar datasets.

References

Badica, C., & Vladutu, G. (2018). Application of Meaningful Text Analytics to Online Product Reviews. *2018 20th International Symposium on Symbolic and Numeric Algorithms for Scientific Computing (SYNASC)*, 327–334. https://doi.org/10.1109/SYNASC.2018.00057

Bakshi, R. K., Kaur, N., Kaur, R., & Kaur, G. (2016). Opinion mining and sentiment analysis. *2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom)*, 452–455.

Benkhelifa, R., Biskri, I., Laallam, F. Z., & Aïmeur, E. (2020). User content categorisation model, a generic model that combines text mining and semantic models. *International Journal of Computational Science and Engineering*, *21*(4), 536–555. https://doi.org/10.1504/IJCSE.2020.106867

Bouckaert, R. R., & Frank, E. (2004). Evaluating the Replicability of Significance Tests for Comparing Learning Algorithms. In H. Dai, R. Srikant, & C. Zhang (Eds.), *Advances in Knowledge Discovery and Data Mining* (Vol. 3056, pp. 3–12). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-24775-3_3

Camillo, A. A., Connolly, D. J., & Woo Gon Kim. (2008). Success and Failure in Northern California. *Cornell Hospitality Quarterly*, *49*(4), 364–380. https://doi.org/10.1177/1938965508317712

Chang, Q. (2020). The sentiments of open financial information, public mood and stock returns: an empirical study on Chinese growth enterprise market. *International Journal of Computational Science and Engineering*, *23*(2), 103–114. https://doi.org/10.1504/IJCSE.2020.110550

Chanwisitkul, P., Shahgholian, A., & Mehandjiev, N. (2018). The Reason Behind the Rating: Text Mining of Online Hotel Reviews. *2018 IEEE 20th Conference on Business Informatics (CBI)*, 149–157. https://doi.org/10.1109/CBI.2018.00025

Choi, T. Y., & Chu, R. (2001). Determinants of hotel guests' satisfaction and repeat patronage in the Hong Kong hotel industry. *International Journal of Hospitality Management*, *20*(3), 277–297. https://doi.org/10.1016/S0278-4319(01)00006-8

Colhon, M., Bădică, C., & Şendre, A. (2014). Relating the Opinion Holder and the Review Accuracy in Sentiment Analysis of Tourist Reviews. In R. Buchmann, C. V. Kifor, & J. Yu (Eds.), *Knowledge Science, Engineering and Management.*

KSEM 2014. Lecture Notes in Computer Science, vol 8793 (pp. 246–257). Springer, Cham. https://doi.org/10.1007/978-3-319-12096-6_22

Dai, Y., Wang, Y., Xu, B., Wu, Y., & Xian, J. (2020). Research on image of enterprise after-sales service based on text sentiment analysis. *International Journal of Computational Science and Engineering*, *22*(2/3), 346–354. https://doi.org/10.1504/IJCSE.2020.107367

Dulen, J. (1999). Quality control. *Restaurants & Institutions*, *109*(5), 38–52.

Frank, E., Hall, M., Holmes, G., Kirkby, R., Pfahringer, B., Witten, I. H., & Trigg, L. (2009). Weka-A Machine Learning Workbench for Data Mining. In *Data Mining and Knowledge Discovery Handbook* (pp. 1269–1277). Springer US. https://doi.org/10.1007/978-0-387-09823-4_66

Gan, Q., Ferns, B. H., Yu, Y., & Jin, L. (2017). A Text Mining and Multidimensional Sentiment Analysis of Online Restaurant Reviews. *Journal of Quality Assurance in Hospitality & Tourism*, *18*(4), 465–492. https://doi.org/10.1080/1528008X.2016.1250243

Gelb, B. D., & Sundaram, S. (2002). Adapting to "word of mouse." *Business Horizons*, *45*(4), 21–25. https://doi.org/10.1016/S0007-6813(02)00222-7

Genc-Nayebi, N., & Abran, A. (2017). A systematic literature review: Opinion mining studies from mobile app store user reviews. *Journal of Systems and Software*, *125*, 207–219. https://doi.org/10.1016/j.jss.2016.11.027

Gersch, M., Hewing, M., & Schöler, B. (2011). Business Process Blueprinting – an enhanced view on process performance. *Business Process Management Journal*, *17*(5), 732–747. https://doi.org/10.1108/14637151111166169

Grosse, K., González, M. P., Chesñevar, C. I., & Maguitman, A. G. (2015). Integrating argumentation and sentiment analysis for mining opinions from Twitter. *AI Communications*, *28*(3), 387–401. https://doi.org/10.3233/AIC-140627

Heung, V. C. S., & Gu, T. (2012). Influence of restaurant atmospherics on patron satisfaction and behavioral intentions. *International Journal of Hospitality Management*, *31*(4), 1167–1177. https://doi.org/10.1016/j.ijhm.2012.02.004

Hong, J.-W., & Park, S.-B. (2019). The Identification of Marketing Performance Using Text Mining of Airline Review Data. *Mobile Information Systems*, *2019*, 1–8. https://doi.org/10.1155/2019/1790429

Jia, S. (Sixue). (2018). Behind the ratings: Text mining of restaurant customers' online reviews. *International Journal of Market Research*, *60*(6), 561–572. https://doi.org/10.1177/1470785317752048

Jin, N.-H., Lee, S.-M., & Huffman, L. (2011). *What Matter Experiential Value in Casual-dining Restaurants?* https://scholarworks.umass.edu/gradconf_hospitality/2011/Poster/121/

Jin, X., Li, Y., Mah, T., & Tong, J. (2007). Sensitive webpage classification for content advertising. *Proceedings of the 1st International Workshop on Data Mining and Audience Intelligence for Advertising - ADKDD '07*, 28–33. https://doi.org/10.1145/1348599.1348604

Kang, H., Yoo, S. J., & Han, D. (2012). Senti-lexicon and improved Naïve Bayes algorithms for sentiment analysis of restaurant reviews. *Expert Systems with Applications*, *39*(5), 6000–6010. https://doi.org/10.1016/j.eswa.2011.11.107

Kim, W. G., Ng, C. Y. N., & Kim, Y. (2009). Influence of institutional DINESERV on customer satisfaction, return intention, and word-of-mouth. *International Journal of Hospitality Management*, *28*(1), 10–17. https://doi.org/10.1016/j.ijhm.2008.03.005

Kimes, S. E. (2008). The Role of Technology in Restaurant Revenue Management. *Cornell Hospitality Quarterly*, *49*(3), 297–309. https://doi.org/10.1177/1938965508322768

Kotschevar, L. H., & Withrow, D. (2007). *Management by Menu* (4th ed.). John Wiley & Sons, Inc.

Ladhari, R., Brun, I., & Morales, M. (2008). Determinants of dining satisfaction and post-dining behavioral intentions. *International Journal of Hospitality Management*, *27*(4), 563–573. https://doi.org/10.1016/j.ijhm.2007.07.025

Lin, W., Jiang, Y., Fu, J., Hao, H., Hu, L., Zou, Y., & Dai, W. (2019). Study on Chinese "Feng Shui" characteristics of commercial community based on big data analytics. *21st International Conference on IT Applications and Management*, 114–122.

Liu, B. (2012). Sentiment Analysis and Opinion Mining. In *Synthesis Lectures on Human Language Technologies* (Vol. 5, Issue 1). https://doi.org/10.2200/S00416ED1V01Y201204HLT016

Liu, Y., & Jang, S. (Shawn). (2009). Perceptions of Chinese restaurants in the U.S.: What affects customer satisfaction and behavioral intentions? *International Journal of Hospitality Management*, *28*(3), 338–348. https://doi.org/10.1016/j.ijhm.2008.10.008

Mäkelä, J. (2000). Cultural definitions of the meal. In H. L. Meiselman (Ed.), *Dimensions of the meal : the science, culture, business, and art of eating* (pp. 7–18). Aspen Publishers, Inc.

Markovic, S., Jankovic, S. R., & Segaric, K. (2010). Does Restaurant Performance Meet Customers' Expectations? An Assessment of Restaurant Service Quality Using a Modified Dineserv Approach. *Tourism & Hospitality Management*, *16*(2), 181–195. https://ssrn.com/abstract=2063559

Mhlanga, O. (2015). Electronic meal experience: a gap analysis of online Cape Town restaurant comments. *African Journal of Hospitality*, *4*(1). http://www.ajhtl.com/uploads/7/1/6/3/7163688/article_15_vol_4_1_2015.pdf

Nakayama, M., & Wan, Y. (2019). The cultural impact on social commerce: A sentiment analysis on Yelp ethnic restaurant reviews. *Information & Management*, *56*(2), 271–279. https://doi.org/10.1016/j.im.2018.09.004

Oh, H.-K., Jung, J., Park, S., & Kim, S.-W. (2020). A robust reputation system using online reviews? *Computer Science and Information Systems*, *17*(2), 487–507. https://doi.org/10.2298/CSIS191122007O

Pantelidis, I. S. (2010). Electronic Meal Experience: A Content Analysis of Online Restaurant Comments. *Cornell Hospitality Quarterly*, *51*(4), 483–491. https://doi.org/10.1177/1938965510378574

Parasuraman, A., Zeithaml, V. A., & Berry, L. L. (1985). A Conceptual Model of Service Quality and Its Implications for Future Research. *Journal of Marketing*, *49*(4), 41–50. https://doi.org/10.1177/002224298504900403

Peri, C. (2006). The universe of food quality. *Food Quality and Preference*, *17*(1–2), 3–8. https://doi.org/10.1016/j.foodqual.2005.03.002

Piao, S., Ananiadou, S., Tsuruoka, Y., Sasaki, Y., & McNaught, J. (2007). Mining opinion polarity relations of citations. *International Workshop on Computational Semantics (IWCS)*, 366–371.

Riedl, J., & Konstan, J. (2002). *Word of mouse : the marketing power of collaborative filtering*. Warner Books.

Riloff, E., Wiebe, J., & Phillips, W. (2005). Exploiting Subjectivity Classification to Improve Information Extraction. *Proceedings of AAAI-05 , the 20th National*

Conference on Artificial Intelligence, 1106–1111. https://www.microsoft.com/en-us/research/publication/exploiting-subjectivity-classification-to-improve-information-extraction/

Saad Andaleeb, S., & Conway, C. (2006). Customer satisfaction in the restaurant industry: an examination of the transaction-specific model. *Journal of Services Marketing*, *20*(1), 3–11. https://doi.org/10.1108/08876040610646536

Saad, S., & Saberi, B. (2017). Sentiment Analysis or Opinion Mining: A Review. *International Journal on Advanced Science, Engineering and Information Technology*, *7*(5), 1660–1666. https://doi.org/10.18517/ijaseit.7.4.2137

Senožetnik, M., Bradeško, L., Šubic, T., Herga, Z., Urbančič, J., Škraba, P., & Mladenić, D. (2019). Estimating point-of-interest rating based on visitors geospatial behaviour. *Computer Science and Information Systems*, *16*(1), 131–154. https://doi.org/10.2298/CSIS171212011S

Song, M., Qiao, L., & Law, R. (2020). Formation path of customer engagement in virtual brand community based on back propagation neural network algorithm. *International Journal of Computational Science and Engineering*, *22*(4), 454–465. https://doi.org/10.1504/IJCSE.2020.109405

Spertus, E. (1997). Smokey: Automatic Recognition of Hostile Messages. *Proceedings of IAAI-97, the 9th Conference on Innovative Application of Artificial Intelligence,* 1058–1065. https://www.microsoft.com/en-us/research/publication/smokey-automatic-recognition-of-hostile-messages/

Stevens, P., Knutson, B., & Patton, M. (1995). Dineserv: A Tool for Measuring Service Quality in Restaurants. *Cornell Hotel and Restaurant Administration Quarterly*, *36*(2), 56–60. https://doi.org/10.1177/001088049503600226

Sulek, J. M., & Hensley, R. L. (2004). The Relative Importance of Food, Atmosphere, and Fairness of Wait. *Cornell Hotel and Restaurant Administration Quarterly*, *45*(3), 235–247. https://doi.org/10.1177/0010880404265345

Susskind, A. M., & Chan, E. K. (2000). How Restaurant Features Affect Check Averages: A Study of the Toronto Restaurant Market. *Cornell Hotel and Restaurant Administration Quarterly*, *41*(6), 56–83. http://scholarship.sha.cornell.edu/articles/380/

Warde, A., & Martens, L. (2000). *Eating Out: Social Differentiation, Consumption and Pleasure*. Cambridge University Press. https://doi.org/10.1017/CBO9780511488894

Wu, C. H.-J., & Liang, R.-D. (2009). Effect of experiential value on customer satisfaction with service encounters in luxury-hotel restaurants. *International Journal of Hospitality Management*, *28*(4), 586–593. https://doi.org/10.1016/j.ijhm.2009.03.008

Yu, B., Zhou, J., Zhang, Y., & Cao, Y. (2017). *Identifying Restaurant Features via Sentiment Analysis on Yelp Reviews*.

Zhang, Y., Liu, F., Koura, Y. H., & Wang, H. (2020). Analysing rumours spreading considering self-purification mechanism. *Connection Science*, 1–14. https://doi.org/10.1080/09540091.2020.1783640

Table 1. Average values and standard deviations of MAE on the whole dataset

	01			IDF		
	TASTE	**SERVICE**	**ENVIRO**	**TASTE**	**SERVICE**	**ENVIRO**
SMO	0.614 ±0.050	0.636 ±0.056	0.605 ±0.053	0.614 ±0.050	0.636 ±0.056	0.605 ±0.053
RandF	**0.541** ±0.045	0.572 ±0.046	0.551 ±0.045	**0.541** ±0.044	0.572 ±0.046	0.551 ±**0.044**
RandT	0.609 ±0.053	0.657 ±0.048	0.636 ±0.055	0.609 ±0.047	0.655 ±0.050	0.635 ±0.055
REPT	0.558 ±**0.044**	0.602 ±0.049	0.594 ±0.047	0.558 ±0.044	0.602 ±0.049	0.594 ±0.047
M5P	0.593 ±0.046	0.628 ±0.049	0.631 ±0.051	0.593 ±0.046	0.628 ±0.049	0.631 ±0.051
MP	0.662 ±0.098	<u>0.738</u> ±<u>0.164</u>	0.734 ±0.153	0.680 ±0.210	<u>0.798</u> ±<u>0.429</u>	0.734 ±0.153

Table 2. Average values and standard deviations of RMSE on the whole dataset

	01			**IDF**		
	TASTE	**SERVICE**	**ENVIRO**	**TASTE**	**SERVICE**	**ENVIRO**
SMO	0.768 ±0.058	0.800 ±0.065	0.761 ±0.064	0.768 ±0.058	0.800 ±0.065	0.761 ±0.064
RandF	**0.679** ±0.049	0.717 ±0.051	0.688 ±0.048	**0.679** ±0.049	0.716 ±0.051	0.688 ±0.048
RandT	0.758 ±0.059	0.817 ±0.056	0.791 ±0.061	0.757 ±0.053	0.814 ±0.057	0.791 ±0.061
REPT	0.698 ±**0.048**	0.751 ±0.054	0.736 ±0.051	0.698 ±**0.048**	0.751 ±0.054	0.736 ±0.051
M5P	0.721 ±0.050	0.768 ±0.052	0.763 ±0.049	0.721 ±0.050	0.768 ±0.052	0.763 ±0.049
MP	0.810 ±0.116	<u>0.895</u> ±<u>0.179</u>	0.887 ±0.169	0.827 ±0.215	<u>0.955</u> ±<u>0.427</u>	0.887 ±0.169

Table 3. Statistically significant wins and losses of MAE and RMSE on whole dataset

	MAE				RMSE		
	W	**L**	**W-L**		**W**	**L**	**W-L**
RandF	30	0	30	**RandF**	30	0	30
REPT	22	6	16	**REPT**	24	6	18
M5P	12	14	-2	**M5P**	16	12	4
SMO	11	12	-1	**SMO**	7	16	-9
RandT	5	20	-15	**RandT**	4	20	-16
MP	0	28	-28	**MP**	0	27	-27

Table 4. Average values and standard deviations of MAE in cluster C3

	01			**IDF**		
	TASTE	**SERVICE**	**ENVIRO**	**TASTE**	**SERVICE**	**ENVIRO**
SMO	0.587 ±0.164	0.610 ±0.177	0.604 ±0.173	0.587 ±0.164	0.610 ±0.177	0.604 ±0.173
RandF	**0.568** ±0.173	0.613 ±0.180	0.631 ±0.178	**0.569** ±0.175	0.612 ±0.178	0.632 ±0.177
RandT	0.626 ±0.166	0.687 ±0.179	0.691 ±**0.158**	0.629 ±**0.161**	0.689 ±0.172	0.691 ±0.175
REPT	0.595 ±0.180	0.649 ±0.172	0.678 ±0.186	0.595 ±0.180	0.649 ±0.172	0.678 ±0.186
M5P	0.595 ±0.173	0.656 ±0.169	0.689 ±0.180	0.595 ±0.173	0.656 ±0.169	0.689 ±0.180
MP	0.646 ±0.196	0.711 ±<u>0.199</u>	<u>0.739</u> ±0.197	0.646 ±0.196	0.711 ±<u>0.199</u>	<u>0.738</u> ±0.197

Table 5. Average values and standard deviations of RMSE in cluster C3

	01			**IDF**		
	TASTE	**SERVICE**	**ENVIRO**	**TASTE**	**SERVICE**	**ENVIRO**
SMO	0.707 ±0.181	0.751 ±0.203	0.744 ±0.188	0.707 ±0.181	0.751 ±0.203	0.744 ±0.188
RandF	**0.694** ±0.193	0.750 ±0.200	0.753 ±0.191	**0.694** ±0.195	0.750 ±0.198	0.753 ±0.191
RandT	0.756 ±0.179	0.834 ±0.203	0.823 ±**0.158**	0.762 ±**0.171**	0.835 ±0.194	0.820 ±0.176
REPT	0.721 ±0.188	0.783 ±0.184	0.806 ±0.189	0.721 ±0.188	0.783 ±0.184	0.806 ±0.189
M5P	0.722 ±0.181	0.784 ±0.183	0.810 ±0.184	0.722 ±0.181	0.784 ±0.183	0.810 ±0.184
MP	0.779 ±0.212	0.855 ±<u>0.219</u>	<u>0.872</u> ±0.213	0.778 ±0.212	0.854 ±<u>0.219</u>	<u>0.871</u> ±0.213

Table 6. Statistically significant wins and losses of MAE and RMSE in cluster C3

	MAE				RMSE		
	W	**L**	**W-L**		**W**	**L**	**W-L**
RandF	26	2	24	**RandF**	24	0	24
SMO	22	2	20	**SMO**	20	0	20
REPT	8	10	-2	**M5P**	9	10	-1
M5P	6	10	-4	**REPT**	8	10	-2
RandT	0	14	-14	**RandT**	0	17	-17
MP	0	24	-24	**MP**	0	24	-24

Table 7. Average values and standard deviations of MAE in cluster C2

	01			**IDF**		
	TASTE	**SERVICE**	**ENVIRO**	**TASTE**	**SERVICE**	**ENVIRO**
SMO	0.647 ±0.100	0.648 ±0.105	0.588 ±0.095	0.647 ±0.100	0.648 ±0.105	0.588 ±0.095
RandF	**0.570** ±0.093	0.589 ±0.108	0.572 ±**0.086**	**0.569** ±0.093	0.588 ±0.107	0.571 ±**0.087**
RandT	0.639 ±0.099	0.681 ±0.109	0.639 ±0.105	0.653 ±0.096	0.679 ±0.116	0.641 ±0.098
REPT	0.603 ±0.092	0.631 ±0.105	0.631 ±0.096	0.603 ±0.092	0.631 ±0.105	0.631 ±0.096
M5P	0.606 ±0.089	0.632 ±0.107	0.641 ±0.092	0.606 ±0.089	0.632 ±0.107	0.641 ±0.092
MP	0.694 ±<u>0.260</u>	0.715 ±0.149	<u>0.724</u> ±0.138	0.694 ±<u>0.260</u>	0.715 ±0.149	<u>0.724</u> ±0.138

Table 8. Average values and standard deviations of RMSE in cluster C2

	01			IDF		
	TASTE	**SERVICE**	**ENVIRO**	**TASTE**	**SERVICE**	**ENVIRO**
SMO	0.774 ±0.100	0.809 ±0.116	0.726 ±0.105	0.774 ±0.100	0.809 ±0.116	0.726 ±0.105
RandF	0.693 ±0.099	0.739 ±0.121	**0.692** ±0.094	0.692 ±0.100	0.738 ±0.120	**0.691** ±0.095
RandT	0.779 ±0.098	0.858 ±0.115	0.785 ±0.111	0.796 ±0.094	0.856 ±0.123	0.786 ±0.102
REPT	0.722 ±0.095	0.781 ±0.110	0.758 ±0.093	0.722 ±0.095	0.781 ±0.110	0.758 ±0.093
M5P	0.722 ±0.093	0.779 ±0.114	0.758 ±**0.085**	0.722 ±0.093	0.779 ±0.114	0.758 ±**0.085**
MP	0.832 ±<u>0.268</u>	<u>0.877</u> ±0.169	0.865 ±0.158	0.832 ±<u>0.268</u>	<u>0.877</u> ±0.169	0.865 ±0.158

Table 9. Statistically significant wins and losses of MAE and RMSE in cluster C2

	MAE				RMSE		
	W	**L**	**W-L**		**W**	**L**	**W-L**
RandF	28	0	28	**RandF**	30	0	30
M5P	12	8	4	**M5P**	14	8	6
REPT	12	8	4	**REPT**	14	8	6
SMO	12	8	4	**SMO**	12	14	-2
RandT	2	18	-16	**RandT**	2	18	-16
MP	0	24	-24	**MP**	0	24	-24

Table 10. Average values and standard deviations of MAE in cluster C0

	01			**IDF**		
	TASTE	**SERVICE**	**ENVIRO**	**TASTE**	**SERVICE**	**ENVIRO**
SMO	0.648 ±0.060	0.694 ±0.067	0.659 ±0.064	0.648 ±0.060	0.694 ±0.067	0.659 ±0.064
RandF	**0.540** ±0.048	0.576 ±0.055	0.550 ±0.054	**0.540** ±0.048	0.576 ±0.055	0.551 ±0.054
RandT	0.618 ±0.057	0.662 ±0.066	0.638 ±0.060	0.610 ±0.055	0.665 ±0.073	0.636 ±0.067
REPT	0.574 ±0.051	0.622 ±0.063	0.615 ±0.056	0.574 ±0.051	0.622 ±0.063	0.615 ±0.056
M5P	0.584 ±**0.045**	0.624 ±0.053	0.618 ±0.053	0.584 ±**0.045**	0.624 ±0.053	0.618 ±0.053
MP	0.661 ±0.104	<u>0.921</u> ±<u>1.500</u>	0.731 ±0.193	0.661 ±0.105	<u>0.921</u> ±<u>1.500</u>	0.731 ±0.193

Table 11. Average values and standard deviations of RMSE in cluster C0

	01			**IDF**		
	TASTE	**SERVICE**	**ENVIRO**	**TASTE**	**SERVICE**	**ENVIRO**
SMO	0.799 ±0.072	0.854 ±0.081	0.828 ±0.080	0.799 ±0.072	0.854 ±0.081	0.828 ±0.080
RandF	**0.674** ±0.059	0.714 ±0.065	0.687 ±0.064	**0.675** ±0.059	0.713 ±0.065	0.688 ±0.064
RandT	0.766 ±0.065	0.818 ±0.072	0.797 ±0.073	0.755 ±0.062	0.823 ±0.082	0.797 ±0.078
REPT	0.709 ±0.061	0.761 ±0.068	0.755 ±0.061	0.709 ±0.061	0.761 ±0.068	0.755 ±0.061
M5P	0.712 ±**0.056**	0.759 ±0.063	0.751 ±0.059	0.712 ±**0.056**	0.759 ±0.063	0.751 ±0.059
MP	0.812 ±0.121	<u>1.071</u> ±<u>1.487</u>	0.881 ±0.206	0.812 ±0.122	<u>1.071</u> ±<u>1.487</u>	0.881 ±0.206

Table 12. Statistically significant wins and losses of MAE and RMSE in cluster C0

MAE				RMSE			
	W	**L**	**W-L**		**W**	**L**	**W-L**
RandF	28	0	28	**RandF**	28	0	28
M5P	16	6	10	**M5P**	16	6	10
REPT	16	6	10	**REPT**	16	6	10
RandT	10	18	-8	**RandT**	10	18	-8
SMO	2	24	-22	**MP**	0	16	-16
MP	0	18	-18	**SMO**	0	24	-24

Figure captions

Figure 1. Instances distribution over object types

Figure 2. Graphical representation of the average values and standard deviations of MAE on the whole dataset

Figure 3. Graphical representation of the average values and standard deviations of RMSE on the whole dataset

Figure 4. Comparison of the average MAE values and standard deviations obtained for the 01 representation: the whole dataset (ALL) and the clusters (C0, C2, C3)

Figure 5. Comparison of the average RMSE values and standard deviations obtained for the 01 representation: the whole dataset (ALL) and the clusters (C0, C2, C3)