High Enough? Explaining and Predicting Traveler Satisfaction Using Airline Reviews

Emanuel Lacic
KTI
Graz University of Technology
Graz, Austria
elacic@know-center.at

Dominik Kowald
Know-Center
Graz University of Technology
Graz, Austria
dkowald@know-center.at

Elisabeth Lex
KTI
Graz University of Technology
Graz, Austria
elisabeth.lex@tugraz.at

ABSTRACT

Air travel is one of the most frequently used means of transportation in our every-day life. Thus, it is not surprising that an increasing number of travelers share their experiences with airlines and airports in form of online reviews on the Web. In this work, we thrive to explain and uncover the features of airline reviews that contribute most to traveler satisfaction. To that end, we examine reviews crawled from the Skytrax air travel review portal. Skytrax provides four review categories to review airports, lounges, airlines and seats. Each review category consists of several five-star ratings as well as free-text review content. In this paper, we conduct a comprehensive feature study and we find that not only five-star rating information such as airport queuing time and lounge comfort highly correlate with traveler satisfaction but also inferred features in the form of the review text sentiment. Based on our findings, we create classifiers to predict traveler satisfaction using the best performing rating features. Our results reveal that given our methodology, traveler satisfaction can be predicted with high accuracy. Additionally, we find that training a model on the sentiment of the review text provides a competitive alternative when no five-star rating information is available. We believe that our work is of interest for researchers in the area of modeling and predicting user satisfaction based on available review data on the Web.

Keywords

traveler satisfaction; airline reviews; skytrax; user satisfaction prediction; feature analysis; sentiment analysis; clustering analysis

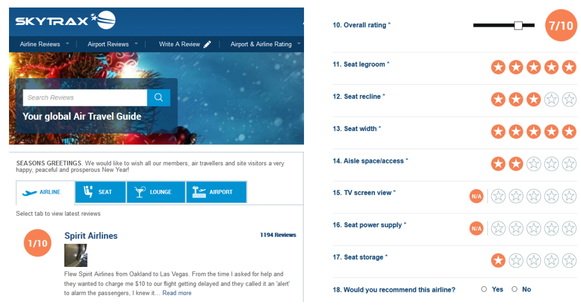
1. INTRODUCTION

In the last decades, air travel has become one of the most frequently used means of transportation. The International Air Transport Association (IATA) expects traveler numbers to reach 7.3 billion by 2034, representing a 4.1% average annual growth in demand for air connectivity [1]. At the same time, an increasing number of airlines is competing for market shares, which raises the need to attract customers while balancing costs and services.

[1]http://www.iata.org/pressroom/pr/Pages/2014-10-16-01.aspx

HT '16, July 10-13, 2016, Halifax, NS, Canada

DOI: http://dx.doi.org/10.1145/2914586.2914629



(a) Review text (b) Ratings

Figure 1: The Skytrax airline (a) review and (b) rating portal. Within each of the four review categories, users state their traveler satisfaction via several rating features, a review text, an overall rating and a binary signal indicated by the *Would you recommend this airline/airport?* checkbox.

A growing number of customers (i.e., travelers) share their experiences and viewpoints on airlines and airports in form of online reviews in order to help others to better judge airline and airport quality. Such reviews may consist of free-text reviews combined with ratings (e.g., by means of 5-star ratings). As a consequence, a vast amount of airline review data is available on the Web, which is not only of interest for the airline industry but also for researchers working on analyzing the impact of the factors/features contributing to user satisfaction [2, 3, 4].

In this paper, we present work-in-progress on a recently started project that aims at explaining and predicting traveler satisfaction using airline review data. Specifically, it is our goal to identify critical features that contribute to air travel satisfaction based on rating and textual reviews. Our idea is to exploit these features in order to predict whether a traveler is satisfied with her airline/airport choice based on the given ratings and/or textual review. This is summed up in the following two research questions that guide our work:

RQ1: Which airline review features are most indicative for traveler satisfaction?

RQ2: To what extent can we predict traveler satisfaction using the available rating and inferred sentiment of airline reviews?

Explaining traveler satisfaction. To better explain how the features contribute to traveler satisfaction, and thus, to address *RQ1*, we exploit real-world airline review data, which was crawled from the Skytrax portal. As shown in Figure 1, in Skytrax users can (a) enter review text and (b) rate various services. Moreover, the user can state her final traveler satisfaction not only using an overall rating between 1 and 10 but also using a checkbox to indicate if she would recommend this airline or airport to other travelers. In terms of rating features, we explore features derived from four different review categories, namely airport, lounge, airline and seat reviews. In terms of inferred features, we extract and make use of the review text sentiment (see Section 3).

To identify which review features are most indicative for traveler satisfaction, we conduct a feature analysis in which we correlate rating and the inferred sentiment with the overall rating given by the user. We find that airport *queuing* time, lounge *comfort*, airline *cabin staff* quality and seat *legroom* space are factors that highly impact the overall traveler satisfaction. We also find that the sentiment of the review content is a good indicator to determine whether a traveler was satisfied with the travel. Additionally, we perform clustering and cluster labelling of the textual content in order to identify topics, which are discussed in the reviews. In the long run, this may help to extend the rating schema. For example, if many users discuss the topic "immigration" in their textual review, the rating portal could introduce a novel rating feature, which enables users to rate the quality of the immigration service.

Predicting traveler satisfaction. We utilize the available rating information as well as the sentiment of the textual reviews as features for our prediction study (*RQ2*). We formulate the prediction task as a binary classification problem of the final traveler satisfaction signal indicated by the *Would you recommend this airline/airport?* checkbox (see Figure 1).

We find strong performance in predicting the traveler satisfaction using the individual rating features. By using a combination of the best performing rating features, we demonstrate that the prediction accuracy can even be increased. Additionally, we show that a classifier, which solely uses the sentiment of the review text, provides a competitive performance in terms of prediction accuracy. This could especially be beneficial in cases where rating features are missing. In terms of metrics, we report the prediction accuracy by means of the F1-score and AUC (i.e., area under ROC curve).

Significance of this work. With this study, we aim at explaining which airline review features have the most impact on predicting traveler satisfaction. Our findings can provide guidance for stakeholders in the airline industry, as well as for researchers, who study online review data to better understand what is important to travellers and what impacts user satisfaction.

2. RELATED WORK

Since Heskett et al. [8] established a relationship between traveler satisfaction and profitability, research on the airline service quality has become an important issue for the airline industry. As a consequence, the authors of [14] claim that it is crucial to continuously collect and evaluate data about traveler satisfaction and how it relates to the provided service quality in order to be competitive in the airline industry. However, most work that conduct research in airline service quality rely on gathered offline data coming from on-site questionnaires [17, 11], airline submissions [18] or in-depth interviews [19].

Nowadays, online reviews are getting more popular and as a consequence, there is the opportunity to leverage them as a rich and powerful source of information. In fact, there is a lot of valuable hidden information available in online reviews [12]. As such, Web sites like the already mentioned Skytrax portal, AirlineRatings [2] and TripAdvisor [3] are important for the airline industry to study how service quality is perceived by the travelers. Furthermore, this data may be a valuable source for researchers that aim at better understanding the factors that contribute to user satisfaction.

Review categories	Airports	Lounges	Airlines	Seats
# Users	11,834	1,598	29,645	1,147
# Reviews	17,721	2,264	41,396	1,258
Traveler Satisfaction	22.12%	36.04%	53.38%	36.41%

Table 1: Statistics of the Skytrax dataset showing how many reviews were given by the users in the four categories. Additionally, we report the traveler satisfaction in the categories as the relative number of reviews that were indicated as airlines or airports that would be recommended to other travelers.

One recent work going into that direction is the one described in [20], in which the authors mined review data about airlines' in-flight services from the Skytrax portal. By grouping travelers via feature-based and clustering-based modelling, the authors showed that inferences can be captured to explain how travelers evaluate in-flight services. Another recent work of Yao et al. [21] presented a research framework to extract and explore information on a user's opinion about airline service features from a large static corpus of online review texts.

In our work, we perform a comprehensive feature analysis using rating features and inferred sentiment from airport, lounge, airline and seat reviews in order to explain which features actually contribute to traveler satisfaction. Moreover, we show how these features can be utilized to predict traveler satisfaction. Our methods and results provide practical insights on how to build upon work like [21] in order to predict traveler satisfaction using online airline reviews.

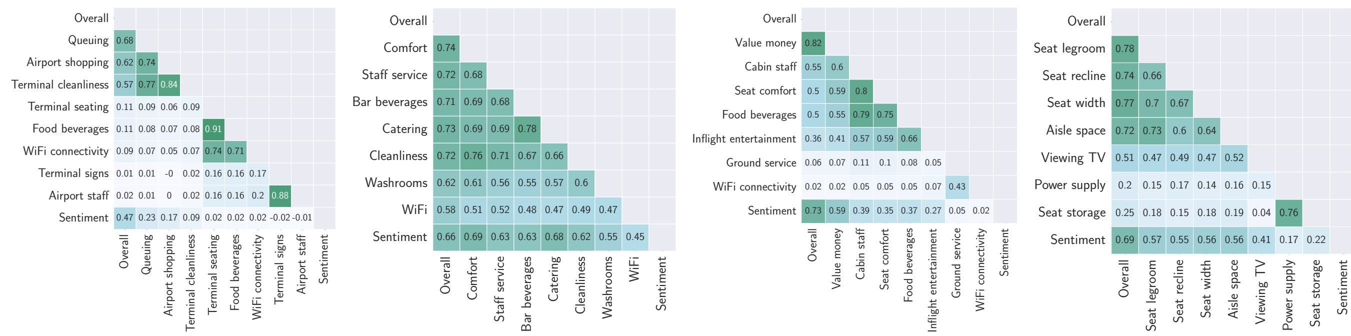
3. AIRLINE REVIEW DATA

Within the air travel industry, the London-based company Skytrax has established itself as a leader in conducting air travel research. Skytrax provides international audits and airport rankings and gives traveler-based satisfaction awards in its yearly *World Airport Awards* and *World Airline Awards*. Their airport and airline review Web portal has positioned itself as one of the most popular independent review sites within the air travel industry. In this work, we incorporate a recent publicly made available airline review dataset [4] scraped from Skytrax's Web portal. This dataset contains not only rating features and textual content of airline reviews but also features that indicate the final traveler satisfaction (see Figure 1).

Rating features. The rating data gathered from Skytrax is divided into four different review categories: (1) airport, (2) lounge, (3) airline, and (4) seat reviews. Each review category has 7 - 8 individual rating features that map the perceived quality of a specific service. The individual rating features are based on a 5-star scale and are accompanied by an additional overall rating on a 1 - 10 scale. Table 1 shows the statistics of the dataset and reveals that most reviews are targeted at airports and airlines, and less at specific seats or lounges.

Inferred features. The posted review text can also contain valuable information about the perceived service quality and satisfaction of a traveler [12]. To that end, we manually enriched the available dataset by inferring the sentiment from the review text. Based on recent research, which compared several sentiment analysis tools [16], we used the AlchemyAPI [5] for this task. For each review, the API was called using the textual content and the re-

[2]http://http://www.airlineratings.com/
[3]http://www.tripadvisor.com
[4]https://github.com/quankiquanki/skytrax-reviews-dataset
[5]http://www.alchemyapi.com/



(a) Airport reviews (b) Lounge reviews (c) Airline reviews (d) Seat reviews

Figure 2: Pearson correlation of the airline review features with the overall rating given by the users. Aside from rating and review sentiment, each review category features an overall rating, which indicates traveler satisfaction (*RQ1*). *Note:* all correlations values higher than .02 have a p-value < .001.

turned value, denoting if the sentiment is positive or negative, was added to the review data. As we will show in this paper, the sentiment of the review text further helps in explaining and predicting traveler satisfaction and is especially useful when rating features are missing.

Traveler satisfaction. We use the overall rating to evaluate how the different review features influence the traveler's satisfaction. Furthermore, in order to make a final decision on how a traveler was satisfied with an airline or airport, we utilize the binary signal represented as the *Would you recommend this airline/airport?* checkbox of Skytrax. As such, Table 1 also shows how travelers are satisfied based on the four review categories. For example, airport reviews mostly resulted in a negative traveler satisfaction, whereas airline reviews almost contain the same amount of satisfactory and unsatisfactory experiences.

4. EXPLAINING TRAVELER SATISFACTION

In this section, we aim to answer the first research question of our work (*RQ1*) and determine the review features that contribute the most to traveler satisfaction.

4.1 Methodology

As already outlined in Section 3, each review category reveals an overall rating, which states how a user perceived an airport, lounge, airline or seat during the travel. For example, the *Dalaman* airport, located in south-west Turkey, received the worst overall rating with a mean of 2.17. On the contrary, the best rated airport is the *Singapore's Changi* airport with an average overall rating of 7.09. With respect to airlines, *Bangkok Airways* was the best rated one with a mean overall rating of 7.99, whereas *Air Canada Rouge* is the worst rated airline with a mean of 2.54.

In order to determine which features actually influence these overall scores, we conduct a feature analysis in which we correlate the rating and inferred features (i.e., the sentiment) with the overall rating given by the user. To explore the influences of these features, we use the Pearson's product-moment correlation coefficient [10]. In this respect, we further correlate the ratings of the features among each other because we believe that knowing how features influence not only the overall rating but also the rating of other features, helps us in even better understanding the factors that contribute to traveler satisfaction.

In addition to the correlation analysis of rating and inferred features, we further incorporate the textual content of online airline reviews. Our aim is to uncover additional features that could be introduced to the rating schema. To that end, we perform clustering and cluster labeling of the review content in order to identify topics, which are discussed in reviews. In contrast to [20], we do not cluster the content with the commonly used k-means approach but rather using Suffix Tree Clustering (STC) [22], an approach that focuses on the problem of cluster labeling. We justify our choice since this clustering technique merges base clusters with high textual overlaps and was shown to outperform group average agglomerative hierarchical clustering, k-means, buckshot, fractionation and single-pass algorithms [22, 15].

4.2 Results

Figure 2 shows the results of our feature correlation analysis on rating and inferred (i.e., sentiment) features based on the four categories.

Airport reviews. In airport reviews, the *overall rating* mostly correlates with ratings assigned to *queuing*, *airport shopping* and *terminal cleanliness*. Looking at the review text, we hypothesize that this is usually caused by short or long queuing times, the availability or quality of airport shopping and how dirty or worn out the airport facilities (e.g., toilets, restaurant, passages, etc.) are. A mild correlation with the sentiment of the review text can also be found. One interesting observation is that there is a relationship between the traveler satisfaction of *terminal seats* and the offered *foods and beverages* as well as available *WIFI connectivity*. It can also be observed that the *airport staff* and *terminal signs* ratings correlate. By looking again in the review text, we think that staff politeness and professionalism, in combination with experiencing issues with signs (e.g., unclear, contradictory, etc.), plays a mayor role in that relationship.

Lounge reviews. Compared to airport reviews, the overall satisfaction within lounge reviews highly correlates with most rating features. The top four features are the perceived *lounge comfort*, available *catering* services, *staff service* quality and the area *cleanliness*. A probably expected observation is that the perceived *catering* quality is mostly in relation with the availability of *beverages*. An interesting finding in lounge reviews is that the sentiment not only correlates with the overall traveler satisfaction but also with the various rating features that denote specific services provided in lounges.

Airline reviews. The top associated rating feature in airline reviews is *value-for-money*. We also find that satisfaction in respect to *seat comfort*, as well as the availability of *food and beverages* is related to how a traveler perceives the *cabin staff*. The extracted sentiment from the review text mostly correlates with the *overall rating*, being here the second best correlating feature and as such a strong signal for traveler satisfaction.

Seat reviews. With respect to the overall satisfaction of a traveler's seat, the best correlating features are *legroom*, *width*, *recline* and *aisle space*. The correlation of these distinctive features suggests that for a traveler, the available personal space is key. Another interesting observation is that how a traveler is satisfied with the available *seat storage* is highly associated with the availability and satisfaction of a *power supply*. The review sentiment, similar as in the case of lounge and airline reviews, is again a strong indicator for the traveler satisfaction denoted by the overall rating.

Extracting review topics. With respect to clustering and cluster labelling, in Figure 3, we report a snapshot of our preliminary results using the Suffix Tree Clustering (STC) approach. By utilizing STC, additional textual features (i.e., cluster labels) can be extracted from the review content. For example, we see in Figure 3 that travelers write about *boarding time* when experiencing negative traveler satisfaction, which in turn results into a negative review about the specific airline. On the contrary, travelers seem to be satisfied with airports when, for example, a smooth *immigration* is ensured and when *gates* are labeled well and easy to reach. Consequently, existing rating schemes could be extended with such cluster labels if they reflect recurring points of discussion in textual reviews.

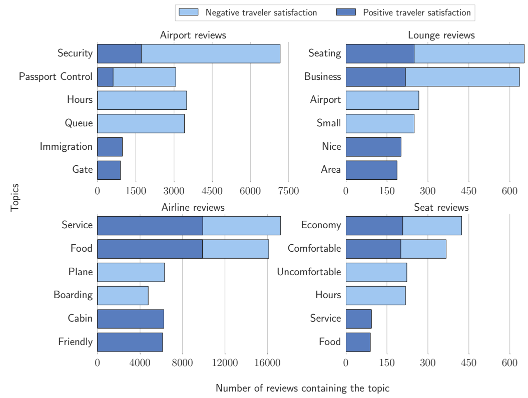


Figure 3: Snapshot of our preliminary clustering and cluster labeling analysis of the review content using the Suffix Tree Clustering (STC) approach. The extracted topics are grouped by having a positive or negative traveler satisfaction denoted by the *Would you recommend this airline/airport?* checkbox of Skytrax.

5. PREDICTING TRAVELER SATISFACTION

In this section, we aim to address our second research question (*RQ2*) in order to determine the features that can be exploited to predict the final traveler satisfaction. Therefore, we formulate the prediction task as a binary classification problem. Given that reviews are marked as either *positive* or *negative* traveler satisfaction by means of the *Would you recommend this airline/airport?* checkbox of Skytrax, we aim to predict this outcome using the available rating and inferred features.

5.1 Methodology

We performed our experiments using several standard classification algorithms (e.g., NaiveBayes, C4.5, Random Forest, CART, etc. [1, 23]) provided by the popular machine learning tool WEKA [6]. In this work, however, we report the results of the Hoeffding Tree.

Introduced by Domigos and Hulten [5], the Hoeffding Tree algorithm is an incremental decision tree learner for large data streams. The tree itself tracks only attribute statistics in its leafs and uses it to grow and make classification decisions for incoming data. When sufficient statistics have accumulated at each leaf, a node-splitting approach determines whether a node-split should happen and the leaf be replaced with a new decision node. We chose this algorithm due to its practical advantage for real-time data mining [9].

In order to evaluate the classification performance, we sorted the reviews of the four categories in chronological order and used the 20% most recent reviews for testing and the rest for training. Next, using each of the four training sets, we examined whether the final user satisfaction of a target review from the corresponding test set could be predicted. With this procedure, we aim to simulate a real-world environment in which future reviewing behavior should be predicted based on past reviews. To determine the best performing features for traveler satisfaction prediction, we trained and evaluated the classification model in the following three settings.

Firstly, for each single rating feature, we created a separate classifier and evaluated its performance. Secondly, we used a combination of features that highly correlate with the traveler satisfaction while having a low inter-correlation [7] (e.g., *overall* rating, *queuing* rating and *airport shopping* rating in case of airport reviews). Thirdly, we trained a model solely based on the inferred review text sentiment. In order to finally quantify the prediction performance, we used a set of well-known information retrieval metrics. In particular, we report the prediction accuracy by means of the F1-score (F1) and Area Under the ROC curve (AUC) [13].

5.2 Results

In this section, we present a discussion on runtime considerations, as well as our prediction results of the Hoeffding Tree algorithm for the individual rating features, the combination of rating features and the review text sentiment.

Runtime considerations. When training and testing the different classification approaches, we achieved the best accuracy performance using the Hoeffding Tree algorithm. Moreover, we found a maximum model training runtime of 0.06 seconds for this classifier in case of the rating feature combination for airline reviews. This clearly underpins our choice for the Hoeffding Tree classifier since runtime is crucial when freshly mined online review data should be instantly incorporated in the classification model. In other words, the Hoeffding Tree can build a competitive model in reasonable time and it enables incremental data updates without re-training the model. This is crucial for real-time data mining applications [9].

Individual rating features. Our prediction results based on the review categories are shown in Table 2. In general, we find strong accuracy performance in predicting the traveler satisfaction using the overall rating feature (e.g., F1 = 0.939 for seat reviews). Furthermore, the performance of rating features that have shown a high correlation with the overall rating (see *RQ1*) also perform reasonably well in terms of satisfaction prediction. For example, using the *value-for-money* feature (F1 = 0.863) in airline reviews provides higher prediction accuracy than using the overall rating (F1 = 0.838).

Airport reviews		
Feature	F1	AUC
Overall	0.963	0.948
Queuing	0.869	0.875
Airport shopping	0.859	0.876
Terminal cleanliness	0.828	0.814
Terminal seating	0.791	0.534
Food beverages	0.792	0.530
WiFi connectivity	0.774	0.519
Terminal signs	0.800	0.502
Airport staff	0.678	0.499
Combination	**0.967**	**0.976**
Airport Sentiment	0.719	0.715
Lounge reviews		
Feature	F1	AUC
Overall	0.834	0.878
Comfort	0.762	0.839
Staff service	0.768	0.819
Bar beverages	0.783	0.838
Catering	0.783	0.829
Cleanliness	0.773	0.817
Washrooms	0.750	0.826
WiFi	0.743	0.795
Combination	**0.837**	**0.884**
Lounge Sentiment	0.773	0.822
Airline reviews		
Feature	F1	AUC
Overall	0.838	0.971
Value money	**0.863**	0.940
Cabin staff	0.794	0.884
Seat comfort	0.750	0.843
Food beverages	0.741	0.827
Inflight entertainment	0.693	0.754
Ground service	0.622	0.533
WiFi connectivity	0.615	0.509
Combination	0.842	**0.975**
Airlne Sentiment	0.839	0.896
Seat reviews		
Feature	F1	AUC
Overall	**0.939**	**0.985**
Seat legroom	0.872	0.919
Seat width	0.847	0.890
Aisle space	0.840	0.895
Seat recline	0.802	0.855
Viewing TV	0.730	0.759
Seat storage	0.711	0.576
Power supply	0.647	0.529
Combination	0.925	0.984
Seat Sentiment	0.812	0.849

Table 2: Classification results using the Hoeffding Tree algorithm for each of the four review categories. The accuracy performance of each single rating feature is reported, as well as the performance when the rating features are combined. Additionally, we report the accuracy, which is achieved by only using review text sentiment as the sole feature. All results are reported by means of the F1-score and AUC (*RQ2*).

This finding indicates that travelers perceive the received value for the spent money as the strongest influence on their final satisfaction with a flight. In contrast, we observe that features with a weak correlation to the overall rating also reach low AUC estimates below 0.6, which is only slightly above random guessing.

Combination of rating features. Overall, the combination of rating features results in strong prediction results with respect to F1-score and AUC. The best performance is achieved with airport and lounge reviews. While being the second best performing feature in airline and seat reviews, the prediction accuracy is still high and does not differ that much from the best performing feature. In case of airport, lounge and airline reviews the feature combination even shows the best AUC performance.

Review text sentiment. Compared to other rating features, review text sentiment is a competitive feature when predicting traveler satisfaction. For example, we can observe that for airline reviews, the sentiment is the third best performing feature (F1 = 0.839), outperforming even the overall rating. Especially in cases where rating features are missing, such a performance is highly beneficial.

6. CONCLUSION AND FUTURE WORK

In this paper, we discuss how online reviews can be an important source of information to explain (*RQ1*) and predict (*RQ2*) traveler satisfaction. We utilized data crawled from the Skytrax portal in order to show that rating features such as airport *queuing* time, lounge *comfort*, airline *cabin staff* quality and seat *legroom* size highly contribute to the overall traveler satisfaction. Moreover, we found a strong correlation between review text sentiment and the final traveler satisfaction (*RQ1*). Based on these findings, we trained several classifiers and we report the results of the Hoeffding Tree algorithm, which not only provides strong accuracy performance but also a competitive runtime. The algorithm is especially suited for real-world settings, where the goal is to continuously mine and predict traveler satisfaction using online reviews. As such, our proposed methods and findings of this work should be of interest for researchers in the area of modeling and predicting user satisfaction based on review data on the Web. To sum up, we found not only that traveler satisfaction can indeed be predicted with high accuracy but also that inferred features such as the extracted sentiment bear great potential in explaining and predicting traveler satisfaction (*RQ2*).

Limitations and future work. In our opinion, a limitation of this work is the lack of a direct comparison with other incremental classifiers such as Incremental Tree Induction (i.e., ITI, the successor of ID5R) or FlexDT (Flexible Decision Tree based on fuzzy logic). As such, we plan to conduct an extensive comparison between different incremental classifiers when mining and predicting user satisfaction using online reviews. Moreover, we aim to continue our preliminary investigations of extracting review topics presented in Figure 3 by further analyzing the textual content of online airline reviews. In this respect, we plan to extend the topic extraction process conducted on the review text with additional approaches like TextRank (i.e., one of the most well-known graph-based approaches for keyphrase extraction) and Topical PageRank (runs TextRank multiple times for topics induced by a Latent Dirichlet Allocation from the text). Therefore, it is not only our intend to uncover additional features that help in explaining traveler satisfaction but also to integrate them in the process of predicting traveler satisfaction. With respect to our prediction study, we plan to incorporate further approaches known from research on recommender systems such as Collaborative Filtering or Matrix Factorization.

Acknowledgments. The authors would like to thank Dieter Theiler and Simone Kopeinik for valuable comments on this work. This work is supported by the Know-Center and the EU funded project Learning Layers (FP7, Grant Agreement 318209). The Know-Center is funded within the Austrian COMET Program - Competence Centers for Excellent Technologies - under the auspices of the Austrian Ministry of Transport, Innovation and Technology, the Austrian Ministry of Economics and Labor and by the State of Styria. COMET is managed by the Austrian Research Promotion Agency (FFG).

7. REFERENCES

[1] L. Breiman, J. Friedman, C. J. Stone, and R. A. Olshen. *Classification and regression trees*. CRC press, 1984.

[2] P. Chatterjee. Online reviews: do consumers use them? *Advances in Consumer Research*, 2001.

[3] P.-Y. Chen, S.-y. Wu, and J. Yoon. The impact of online recommendations and consumer feedback on sales. *ICIS 2004 Proceedings*, page 58, 2004.

[4] Y. Chen, S. Fay, and Q. Wang. Marketing implications of online consumer product reviews. *Business Week*, 7150:1–36, 2003.

[5] P. Domingos and G. Hulten. Mining high-speed data streams. In *Proceedings of the sixth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 71–80. ACM, 2000.

[6] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten. The weka data mining software: an update. *ACM SIGKDD explorations newsletter*, 11(1):10–18, 2009.

[7] M. A. Hall. *Correlation-based feature selection for machine learning*. PhD thesis, The University of Waikato, 1999.

[8] J. L. Heskett, L. Schlesinger, et al. Putting the service-profit chain to work. *Harvard business review*, 72(2):164–174, 1994.

[9] G. Hulten, L. Spencer, and P. Domingos. Mining time-changing data streams. In *Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 97–106. ACM, 2001.

[10] J. Lee Rodgers and W. A. Nicewander. Thirteen ways to look at the correlation coefficient. *The American Statistician*, 42(1):59–66, 1988.

[11] J. J. Liou and G.-H. Tzeng. A dominance-based rough set approach to customer behavior in the airline market. *Information Sciences*, 180(11):2230–2238, 2010.

[12] B. Pang and L. Lee. Opinion mining and sentiment analysis. *Foundations and trends in information retrieval*, 2(1-2):1–135, 2008.

[13] D. M. W. Powers. Evaluation: from precision, recall and f-measure to roc, informedness, markedness and correlation. *International Journal of Machine Learning Technology*, 2011.

[14] G. C. Saha and Theingi. Service quality, satisfaction, and behavioural intentions: A study of low-cost airline carriers in thailand. *Managing Service Quality: An International Journal*, 19(3):350–372, 2009.

[15] S. Sambasivam and N. Theodosopoulos. Advanced data clustering methods of mining web documents. *Issues in Informing Science and Information Technology*, 3:563–579, 2006.

[16] J. Serrano-Guerrero, J. A. Olivas, F. P. Romero, and E. Herrera-Viedma. Sentiment analysis: A review and comparative analysis of web services. *Information Sciences*, 311:18–38, 2015.

[17] N. M. Suki. Passenger satisfaction with airline service quality in malaysia: A structural equation modeling approach. *Research in Transportation Business & Management*, 10:26–32, 2014.

[18] S. Tiernan, D. L. Rhoades, and B. Waguespack Jr. Airline service quality: Exploratory analysis of consumer perceptions and operational performance in the usa and eu. *Managing Service Quality: An International Journal*, 18(3):212–224, 2008.

[19] I. Vlachos and Z. Lin. Drivers of airline loyalty: Evidence from the business travelers in china. *Transportation Research Part E: Logistics and Transportation Review*, 71:1–17, 2014.

[20] I. Yakut, T. Turkoglu, and F. Yakut. Understanding customers' evaluations through mining airline reviews. *International Journal of Data Mining & Knowledge Management Process*, 2015.

[21] B. Yao, H. Yuan, Y. Qian, and L. Li. On exploring airline service features from massive online review. In *Proceedings of the 12th International Conference on Service Systems and Service Management (ICSSSM)*, pages 1–6. IEEE, 2015.

[22] O. Zamir and O. Etzioni. Web document clustering: A feasibility demonstration. In *Proceedings of the 21st annual international ACM SIGIR conference on Research and development in information retrieval*, pages 46–54, 1998.

[23] Y. Zhao and Y. Zhang. Comparison of decision tree methods for finding active objects. *Advances in Space Research*, 41(12):1955–1959, 2008.