ORIGINAL RESEARCH

Can machine learning techniques predict customer dissatisfaction? A feasibility study for the automotive industry

Stefan Meinzer*[1], Ulf Jensen[1], Alexander Thamm[2], Joachim Hornegger[3], Björn M. Eskofier[1]

[1]*Digital Sports Group, Pattern Recognition Lab, Department of Computer Science, Friedrich-Alexander-University Erlangen-Nürnberg, Erlangen, Germany*
[2]*Department of Sales Management and Marketing, Faculty of Business Administration, University of Eichstaett-Ingolstadt, Ingolstadt, Germany*
[3]*Pattern Recognition Lab, Department of Computer Science, Friedrich-Alexander-University Erlangen-Nürnberg, Erlangen, Germany*

ABSTRACT

The automotive industry is in the strongest competition ever, as this sector gets disrupted by new arising competitors. Providing services to maximum customer satisfaction will be one of the most crucial competitive advantages in the future. Around 1 Terabyte of objective data is created every hour today. This volume will significantly grow in the future by the increasing number of connected services within the automotive industry. However, customer satisfaction determination is solely based on subjective questionnaires today without taking the vast amount of objective sensor and service process data into account. This work presents an industrial application that fills this lack of research and thus provides a solution with a high practical impact to survive in the tough competition of the automotive industry. Therefore, the work addresses these fundamental business questions: 1) Can dissatisfied customers be classified based on data that is produced during every service visit? 2) Can the dissatisfaction indicators be derived from service process data? A machine learning problem is set up that compared 5 classifiers and analyzed data from 19,008 real service visits from an automotive company. The 105 extracted features were drawn from the most significant available sources: warranty, diagnostic, dealer system and general vehicle data. The best result for customer dissatisfaction classification was 88.8% achieved with the SVM classifier (RBF kernel). Furthermore, the 46 most potential indicators for dissatisfaction were identified by the evolutionary feature selection. Our system was capable of classifying customer dissatisfaction solely based on the objective data that is generated by almost every service visit. As the amount of these data is continuously growing, we expect that the presented data-driven approach can achieve even better results in the future with a higher amount of data.

Key Words: Customer satisfaction prediction, Classification models, Data consolidation logic, Automotive industry, Monte Carlo sampling, Evolutionary feature selection

1. INTRODUCTION

Every hour, around 1 TB of data is produced by cars and dealers within the automotive industry today, growing to 1 GB in the future created by self-driving cars in every second.[1] Thus, the automotive industry is looking intensively for methods to handle this vast amount of data and analyze it

---

to secure maximum customer benefit. Increasing perceived customer satisfaction or, respectively, decreasing dissatisfaction was identified as the number one management target within the automotive industry.[2, 3] Knowing about dissatisfied customers is vitally important to create competitive leads, especially in the after-sales sector.[4] For the US service industries, an estimated revenue of 6-8 billion USD is annually created by automotive after-sales services.[5] Consequently, the identification of potentially dissatisfied customers, before the service interaction ends, is of great value to enable proactive actions before the customer actually gets dissatisfied. A satisfied customer will retain and thus is of great interest from a business perspective. This is underlined by literature that shows that the costs for new customer acquisition are at least double[6] to 5-6 times higher than maintaining the customer.[7] Studies indicate that profits can be increased by almost 100% only by retaining 5% of the customers.[8]

Traditionally, customer satisfaction with a dealer visit is assessed by customer surveys using questionnaires after the customer left the dealer.[2, 9–11] The use of questionnaires to measure dissatisfaction has several drawbacks. First, the customer is already dissatisfied at this time as the customer-service interaction is already completed. Second, not every customer can be assessed as only a little sample of customers receives a survey. However, for every visit, service and repair data are produced and can be used for further analysis.[12, 13]

Therefore, the essential question for the automotive industries is: Can dissatisfied customers be classified based on data that is produced during every service visit?

Beside this customer value, the optimization of processes based on the knowledge of customer dissatisfaction would generate huge company benefits. Repair routines could be improved and potential service failures could be eliminated. In order to realize this company value, a second question arises: Can the dissatisfaction indicators be derived from service process data?

The following literature review aims to identify existing approaches to increase customer satisfaction in the service industry in general. The use of machine learning techniques to improve the customer relationship was prevalent especially in the telecommunication business. The prediction of customer churn, as a consequence of dissatisfaction, using machine learning algorithms was thereby most prominent. Datta et al.[14] developed a prediction model for churning customers in mobile phone services using a cascade neural network model. Customers that will discontinue using the cellular phone services were predicted. To predict the time of churn was the goal by further research.[15] They compared ordinal regression models with survival analysis for tenure prediction of mobile phone customers whereby the regression models showed most significant results.

Existing resaerch[16] aimed to develop a propensity-to-churn score for mobile phone subscribers in Taiwan. They compared various data mining techniques and identified decision trees as most promising to model churning customers. Predicting latent churn customers was also scope of research.[6] The authors investigated different techniques to predict customer churn and concluded that support vector machines (SVM) showed the highest accuracy.

In the health care sector, Yarnold et al.[17] were among the first who used machine learning techniques to predict overall patient dissatisfaction. Nonlinear decision trees were used to classify patient satisfaction based on survey response data of emergency departments. Sun et al.[18] had the objective to identify process measures that significantly influence patient satisfaction in emergency departments. Logistic regression models were applied based on questionnaire data, patient characteristics and process measures. Critical features of satisfaction, such as explanation of diagnostic results, could be identified. Boudreaux et al.[19] investigated the main predictors of patient satisfaction in municipal emergency departments. Their results were based on telephone interview data and sociodemographics of hospital patients that were analyzed with logistic regression. The findings revealed the main indicators for overall patient satisfaction like patients perceptions of care and further features. An extensive overview on the determination of perceived patient satisfaction in the health care sector can be found in literature.[20]

Within the automotive industry, similar approaches were used for customer recommendations, targeted marketing leads and customer satisfaction. Mavridou et al.[20] aimed to improve the car configurator. They used association rules based on customer choices to create an individualized recommendation system. Chan et al.[21] focused on the product development process within the automotive industry. They mapped survey response data from customer studies to design attributes according to customer preferences. Further research investigated the possibilities to create marketing leads from customer satisfaction, sociodemographic and accounting data.[22] They used the techniques of self-organizing maps and decision trees for an individualized customer segmentation. In addition, the goal to make customer preferences more transparent is a research area.[23] The authors applied Bayesian Networks solely on survey response data. For instance individualized pricing or region specific marketing concepts were derived. Various researches focused on modeling customer satisfaction using vehicle failure data.[12, 24, 25] They limited their focus on quality and reliability related

satisfaction and extracted features from warranty data.[12] Bandaru *et al.*[24, 25] used extracted and calculated features from service and sales data to model customer satisfaction.

Models that focus on the automotive industry are scarce.[10, 12] Machine learning techniques were not applied in order to classify overall dissatisfied customers or to identify the most critical indicators based on continuously created, objective data. Further existing research within the area of artificial intelligence and machine learning is focusing on sensor data in order to improve diagnostic procedures[26, 27] or enhancing pedestrians, road and traffic sign detection.[28, 29] Due to the limitations of determining subjective customer satisfaction using questionnaires and the overall customer satisfaction as a fundamental management target in the automotive industry, there is an unmet need for a system to classify dissatisfied customers from objective, technical data.[10] Such an approach has not been presented so far as the literature review expressed.

Customer satisfaction is one of the most important managerial goals in the automotive industry and helps to increase competitive advantages. To address this high revenue po-tential, dissatisfied customers need to be predicted and the indicators for dissatisfaction need to be identified. The large amount of data that is produced every day by dealers, customers and cars can be used to achieve this goal.[2, 3]

This work investigates five machine learning approaches for their performance in predicting dissatisfied customers and identifying general indicators for customer dissatisfaction based on objective data. The presented concept can be transferred to other service industries, such as the health care sector where the customer satisfaction is an important management target and sensor data are continuously produced.

2. METHOD

This section describes the materials and methods to classify customer dissatisfaction and to identify indicators for customer dissatisfaction. First, the necessary data sources are described. Second, the consolidation of the data is presented. Third, the feature extraction is illustrated. Fourth, the data balancing procedure is shown. Finally, the experimental design to prepare and model the data is explained. The complete data analysis process is shown as an overview in Figure 1 and will be described in more detail.
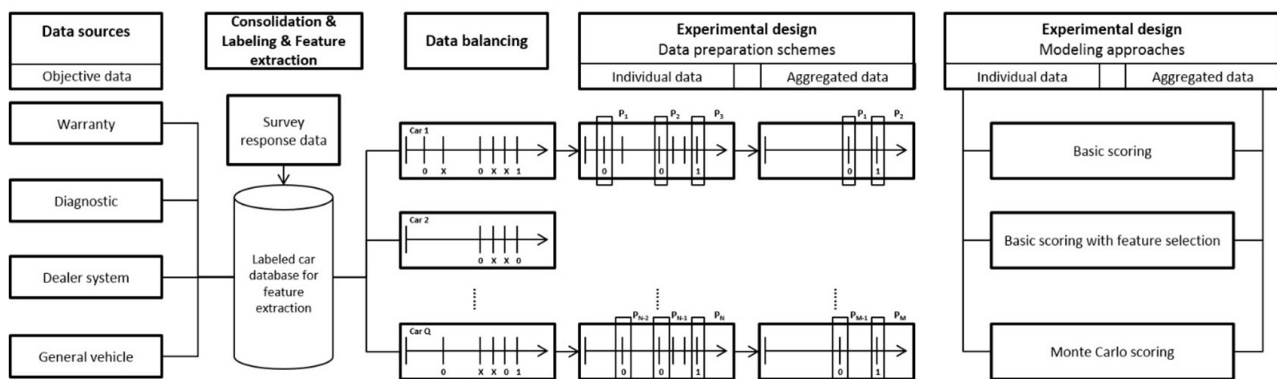

Figure 1. Overview about the data sources for the consolidated database, the data balancing procedure, the resulting data preparation and modeling approaches. The label 0 was used for a satisfied and label 1 for a dissatisfied customer; Service visit without survey response are represented by X, and the patterns of data deduced from service visits by P.

2.1 Data sources

Since unique identifiers to aggregate all information for each car are not present, an automated way to consolidate and validate the data on a service visit level was developed. Generally, warranty data, diagnostic data, dealer system data and general vehicle data have been consolidated. A detailed description of the transfer of information to the automotive manufacturer's can be found in literature.[13, 30] The data sources, used for labeling and feature extraction are illustrated in Figure 1.

2.1.1 *Warranty data*

Warranty claims identify all repairs that are performed and paid by the automotive company within the warranty period or as goodwill.

Dealers have to claim for all work they do by using different codes. Those codes consider all parts that are replaced, mechanical work and the time needed for the several repair steps. All this information and the amount of money the dealer requests is transferred via the claims. An overview of the features is given in Table 1.

Table 1. Overview of the features used for classification, corresponding feature categories and data sources

Feature	Feature category	Feature description	Data sources	Category information	Exemplary feature
1	SV_CNT	Amount of service visits per car	Warranty Diagnostic Dealer system	Calculated	Visits = 5
2-57	CHARACS	General characteristics of the car	General vehicle	Extracted	Milage = 12,000 miles
58	LENGTH	Length of stay in workshop per service	Warranty Diagnostic	Calculated	Length of stay = two days
59-62	SOURCE	Identifier of data sources	All	Calculated	Data source = Diagnostic
63-65	RR	Identification of repeat repairs	Warranty Diagnostic	Calculated	Warranty repeat repair = true
66-69	WGS	Description of warranty level of repair	Warranty	Extracted	Warranty stage = Goodwill
70-95	DIAG	Repair details from diagnostic data	Diagnostic	Extracted	Result = Software update
96-104	WTY	Repair details from warranty data	Warranty	Extracted	Repair = Engine part replaced
105	DIFF	Time gap to previous service visit	Warranty Diagnostic Dealer system	Calculated	Previous visit = 197 days

2.1.2 *Diagnostic data*

The data transfer of diagnostic protocols is automatically triggered on each service visit caused by an electronic related defect. The system captures details of the errors (stored as fault codes) set in the car. Repair plans are derived from these fault codes and provided to the mechanics. Diagnostic data comprises the results and errors that are read out from the car as well as the repair procedure steps. A detailed description can be found in literature.[13] The features extracted from this sources are given in Table 1.

2.1.3 *Dealer system data*

Additional information about dealers' work is retrieved from the dealer management systems. Spare parts that were paid by the customer or regular maintenance information is provided. An overview about different exemplary features from this source is given in Table 1.

2.1.4 *General vehicle data*

This data source covers all general car and dealer information such as car type, extras, power, *etc.* It is collected after a car is produced and sold to the customer. Examples for these car characteristics are presented in Table 1.

2.1.5 *Survey response data*

After each service visit the customer may be contacted for a customer satisfaction survey if he gives the dealer his approval. The survey is conducted by an external agency that collects data for the main automotive companies. A random sample of customers is asked to fill out a questionnaire. Thus, a representative set of service visits from all dealers is drawn assuming that this sample covers the different perceived quality levels of the visits. The sample size is based on various factors like the size of the dealer or the type of customer (business, private, *etc.*).

There are regular maintenance interval visits included as well as repair incidents. The perceived overall satisfaction with a service visit is determined. A customer can receive a survey multiple times for different visits. This data source is needed in order to label the service visits as satisfactory or dissatisfactory.

2.2 Consolidation and labeling

2.2.1 *Data consolidation*

In order to analyze the data, the five different data sources needed to be combined as illustrated in Figure 1. All systems store the information of a service visit in combination with the vehicle identification number (VIN) and a timestamp. However, the timestamp is not unique in all data sources. There can be multiple timestamps identifying different service visits (*e.g.* diagnostic readouts differ from warranty claiming). The data in the warranty and dealer management system is created manually, the others are automatically produced. This situation yields a biased timestamp. This bias needed to be integrated into the consolidation logic. A window of 5 days and 50 kilometers was tolerated to consider the data as one visit. Otherwise a new visit is assumed.

The aggregation procedure was split up in three steps and is explained in the following.

Step 1: Data sources consolidation

   (1) Data sets for all vehicles were extracted from the five data sources introduced in Section 2.1.

   (2) Check all data sources for matching cars (VIN).

   (3) Vehicles with survey responses were selected.

Step 2: Specified matching

   (1) Data sets with identical timestamps and VIN were matched.

   (2) Vehicles with different timestamps but identical VIN were transferred to Step 3.

Step 3: Tolerance matching

   (1) Biased timestamps with identical VIN were filtered and the difference between the timestamps was calculated.

  (2) Vehicles with a difference less than 5 days or 50 kilometers were consolidated into one service visit.

  (3) Vehicles that did not fulfill these criteria were considered as separate service visits.

2.2.2 *Labeling*

Each service visit can be identified with the unique timestamp of the survey and the VIN. The survey design requires to send out the questionnaire not later than 30 days after a service visit. Thus, survey data and service visits using a tolerance window of 30 days were mapped. The overall satisfaction of the particular customer was considered to be the class label. The overall satisfaction rate was measured using a likert scale from 1-5 (where 5 is best). In order to guarantee a two-class classification problem, the results of this question have been transferred into a binary coding. Service visits with satisfaction rates of 1 to 3 were identified as dissatisfactory and labeled with 1 whereas 4 and 5 represent the service visits of satisfied customers that were labeled with 0. This is a common way of interpretation and presentation in a management manner.[19]

2.3 Feature extraction

Features were derived from the consolidated and labeled data sources. In order to provide a holistic overview, the features were categorized as presented in Table 1. Generally, the features provide an encompassing view about the service visits of the car. Technical information such as the repair cause, repair results, affected parts or general vehicle characteristics were included. Furthermore, features that may potentially be perceived by the customer were calculated, such as repeat repairs, the length of stay or the time between actual and previous visit. In total 105 features were extracted for the analysis and presented in Table 1.

2.4 Data balancing

Dissatisfied customers were underrepresented in the dataset. For the analysis, too many patterns about satisfied customers were included during the learning process. The class imbalance challenge has gained attention previously in research on machine learning.[31] In the dataset, 5,048 events with label 1 corresponded to 13,960 service visits with label 0. As the minor group of dissatisfied customers should be classified as good as possible, strategies to cope with this class imbalance were needed. Thereby, the probability of False Negatives, meaning dissatisfied customers were classified as satisfied, should be reduced to a minimum. Various machine learning algorithms that do not take class imbalance into account tend to be biased.[31, 32] In order to increase the performance of the classification of dissatisfied customers, a random undersampling of the satisfied customers was used.[33] Compared to alternative solutions, such as cost functions, that was shown to be an easy and effective procedure.

In order to achieve equal populations of the two classes, a selection probability of roughly 36.1% within the satisfied customer group was needed to achieve a subsample of 5,048 observations. Thereby, an optimal population of car owners whose satisfaction level changed from satisfied to dissatisfied within a service history has been generated.
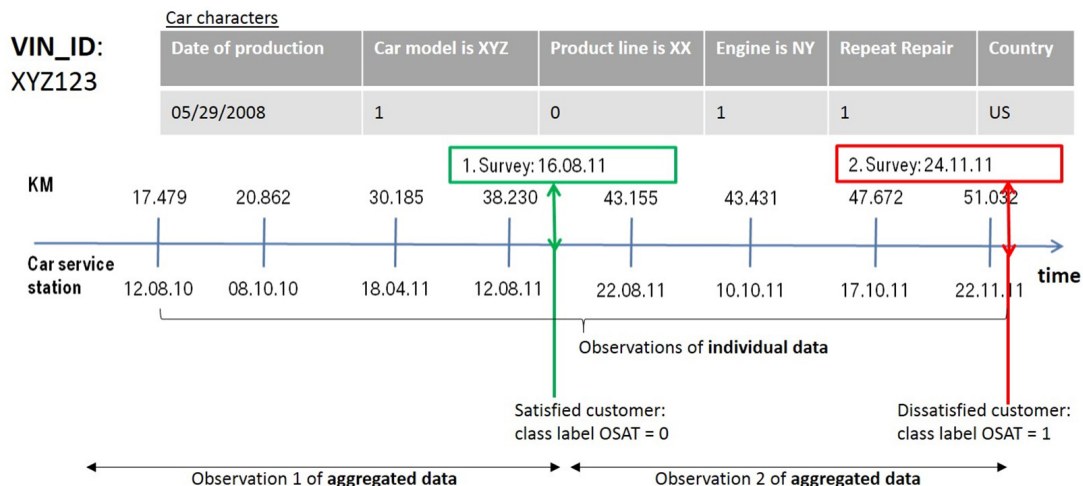

Figure 2. Overview about the two data preparation schemes illustrated by one exemplary car

2.5 Experimental design

The experimental design of this study contained two data preparation schemes: Individual and aggregated data. Three modeling approaches were used: Basic scoring, basic scoring with evolutionary feature selection and a Monte Carlo scoring. The two data preparation schemes aimed to provide insight into whether the performance of the classification can be increased by considering the individual service visits or

combining them as a history of events. The three modeling approaches addressed the question about the derivation of potential indicators of dissatisfaction. In the experimental design five classifiers were compared using different parameter settings: AdaBoost, k nearest neighbor (kNN), SVM with linear and RBF-kernel and Random Forest. These five classifiers were chosen as they were evaluated to be suited as state-of-the-art classifiers for industrial applications.[34]

2.5.1 *Data preparation schemes*

Two different data preparation schemes consisting of individual and aggregated data were applied. Individual data means that training and scoring was based on the original longitudinal data structure of cars and service visits. Consequently, the dataset consisted of multiple dealer visits per car. This approach is typically used today to analyze survey results.

Aggregated data means, that all service visits of one customer (represented by the VIN) that remain the customer as satisfied were aggregated to one observation. Means were used for numeric features and modes for categorical features. The following, second record identified the visit that is labeled as dissatisfactory if available.

An exemplary overview on one car, its general characteristics and the corresponding service visits is given in Figure 2. The figure illustrates the two data preparation schemes.

2.5.2 *Modeling approaches*

The three benchmarked modeling approaches were: Basic scoring, basic scoring with evolutionary feature selection and a Monte Carlo scoring. In general, constant and quasi constant items, meaning features with very low entropy (for example high amount of null values), were eliminated. Furthermore, features with a correlation coefficient of more than 0.95 were removed.

Basic scoring compared different classifiers for their performance on customer dissatisfaction classification. The complete feature set was used.

The basic scoring with evolutionary feature selection[35] experiment had the main goal to identify the discriminant features and thus the main indicators for customer dissatisfaction. The feature selection must be fast in calculation to secure a practical implementation. Therefore, it was performed using a simple decision stump that uses only one split, therefore being fast and avoiding overfitting.[36] The initial feature set is randomly generated, whereby each feature set had a feature activation probability of $p_i = 0.50$. During the next maximum 50 iterations a population of 5 individual feature sets competed for the best classification accuracy using mutation. Mutation means the random activation and deactivation of features with $p_m = 1$/random of

features. Furthermore, features can interchange within the individual feature sets with the probability of $p_c = 0.50$. One of the five feature sets will result in the highest accuracy and remain for the next iteration. If the accuracy of one of the new feature sets is better than before, the algorithm proceeds to the next iteration. Once there was no improvement over 10 iterations, an early stopping rule was applied and the final subset of features was reported.

A Monte Carlo resampling approach was conducted in order to evaluate the classification variances of the applied predictions. Thereby, 500 random subsamples, each consisting of 1,000 balanced observations (500 satisfied and 500 dissatisfied customers), were drawn and evaluated using the grid search from the basic scoring approach. In each iteration the respective classification results and parameter sets were stored.

While the two basic scoring approaches generate one observation per parameter set, the Monte Carlo Scoring is generating 500 observations as a consequence of the iterations for each parameter setting Thereby, statistical tests could be applied to these iterations. In order to investigate if the accuracy distribution of training, test and validation sample are significantly different, Friedmans' ANOVA has been applied for the 500 observations within the best parameter set for each classifier of the Monte Carlo Scoring with significance level of $\alpha = .05$.[37] In order to investigate if there are significant differences between the accuracy distributions of the classifiers which is necessary to compare the 5 evaluated classifiers, the non-parametric Kruskal-Wallis-Test with significance level of $\alpha = .05$ has been applied for the best parameter set.[38, 39] Both statistical tests were performed for each data preparation schema. All experiments were conducted using RapidMiner toolbox version 5.3.013 (RapidMiner, Cambridge, MA, USA). A PC with 32 cores and 512 GB RAM was used for the analysis.

2.5.3 *Experimental setup and evaluation*

In order to achieve the intended goal of differentiating satisfied from dissatisfied customers during a service visit, a specific selection of machine learning approaches was chosen. An overview of the most common machine learning techniques can be found in literature.[40, 41] In this work, the focus is on the following five classifiers based on the above described design. First, the performance of the AdaBoost classifier was investigated with Nit ε [10, 20, $\cdots$, 100] which corresponds to literature findings.[42, 43] Second, the performance of the kNN classifier was investigated with number of nearest neighbors [k = 1, 2, $\cdots$, 10]. Third, the performance of the SVM was investigated using the linear kernel and fourth, using the RBF kernel. Both SVM classi-

fiers were investigated with kernel parameter C in the range [2-5, 2-4, $\cdots$, 25].[44] For the RBF kernel the parameter space [2-5, 2-4, $\cdots$, 25] was used for γ. Fifth, the performance of the Random Forest[45] was investigated using 10 to 1,000 trees increasing with 10 steps on a quadratic scale [10, 20, 50, $\cdots$, 1,000]. Each set of free parameters was evaluated.

From the current dataset, a training (0.70), test (0.20) and validation (0.10) sample was conducted and the accuracy for test and validation was reported. For each parameter set, the classification rate of the five classifiers was investigated. For the feature selection approach, the partitioning was performed after the evolutionary feature selection has been done on the full feature set. Within the Monte Carlo Scoring, the partitioning has been done within each iteration before the classification models were applied to the feature set resulting from the Monte Carlo iteration.

Table 2. Results of basic scoring on individual and aggregated data

Classifier	Individual data		Aggregated data	
	Test	Validation	Test	Validation
AdaBoost	50.0	68.8	82.9	82.9
kNN	57.0	58.0	55.1	54.9
SVM (Linear)	72.3	70.8	80.4	81.7
SVM (RBF)	**77.2**	**76.4**	**88.8**	**88.8**
Random Forest	75.0	74.7	81.5	78.3

3. RESULTS

The results of the experimental design are presented in Tables 2-4. The classifiers in the tables are sorted in alphabetic order where the classifiers with the highest accuracy in test

and validation partition are marked. For the results of the Monte Carlo scoring in Table 4 the standard deviation σ is given in brackets over all iterations.

Table 2 presents the results for the basic scoring. The SVM using the RBF kernel was the best performing classifier for the individual as well as for the aggregated data with a test and validation accuracy of 77.2% (76.4%) and 88.8% (88.8%) respectively.

Table 3. Results of basic scoring with feature selection on individual and aggregated data

Classifier	Individual data		Aggregated data	
	Test	Validation	Test	Validation
AdaBoost	70.0	70.7	83.3	79.3
kNN	55.8	55.9	84.2	80.9
SVM (Linear)	73.2	72.1	78.9	80.7
SVM (RBF)	72.6	74.9	**86.1**	**88.8**
Random Forest	**75.0**	**74.6**	87.2	82.5

Table 3 summarizes the results of the basic scoring with evolutionary feature selection. Based on the individual data the Random Forrest showed best results with an accuracy of 75.0% for the testing and 74.6% for the validation data. The SVM with RBF kernel performed best for aggregated data with an accuracy of 86.1% (88.3%). In order to secure the practical application of the feature selection approach, a fast computing time is required. Therefore, decision stump has been applied within the feature selection procedure using one split. This approach secures the necessary computing performance but has the disadvantage that interaction effects of the features are limited to the one split criteria and thereby may decrease the accuracy.

Table 4. Results of Monte Carlo scoring on individual and aggregated data

Classifier	Individual data*		Aggregated data*	
	Test	Validation	Test	Validation
AdaBoost	66.0 (0.06)	66.6 (0.06)	79.5 (0.06)	79.7 (0.04)
kNN	54.6 (0.05)	54.5 (0.04)	55.0 (0.05)	55.6 (0.03)
SVM (Linear)	69.6 (0.04)	69.9 (0.03)	80.3 (0.04)	80.2 (0.03)
SVM (RBF)	**75.0 (0.04)**	**72.2 0.03)**	**84.1 (0.04)**	**83.8 (0.03)**
Random Forest	71.2 (0.05)	71.8 (0.03)	79.9 (0.04)	79.8 (0.03)

Note. * Training, test and validation accuracy distributions are significantly different with $p < .05$ by Friedmans' ANOVA.

Results for the Monte Carlo scoring are given in Table 4. The SVM with RBF kernel showed the best classification results for individual data with an accuracy of 75.0% and 72.2% for test and the validation partition respectively. For the aggregated dataset, it was also the SVM with RBF kernel that achieved highest accuracy with 84.1% and 83.3% for test and validation partition. Both show a standard deviation

of the accuracy over all iterations of $\sigma = 0.04$ and $\sigma = 0.03$. Results of Friedmans' ANOVA allow the significant interpretation of the accuracy measures as the null hypothesis could be rejected that the accuracy distributions over training, test and validation samples are equal. This is valid for both data preparation scheme with a p-value of $< .05$. The results of the Kruskal-Wallis-Test are as follows: For the individual data,

the accuracy distribution over the 5 classifiers is significantly different for the test sample. For the validation sample, there is no significant difference in the accuracy distribution of the SVM (RBF) and Random Forest. For the aggregated data, there is no significant difference in the accuracy distribution for the test and validation sample of the Random Forest and AdaBoost, the SVM (Linear) and AdaBoost and the Random Forest and SVM (Linear). The other accuracy distributions are significantly different with a p-value of $< .05$.

Overall, best classification rates with 88.8% for test and validation were achieved with the basic scoring approach on the aggregated dataset performing the SVM with a RBF kernel with a parameter set of C = 16, $\gamma = 0.03125$ (see Table 2). The second best results with a classification rate of 88.3% were returned by the SVM using a RBF kernel based on the basic scoring with evolutionary feature selection using the aggregated data with an optimal parameter set of C = 4, $\gamma = 0.03125$ (see Table 3). The feature selection selected 46 features. Thereby, 23 features were selected from the category CHARACS, 15 from DIAG, 5 from WGS, 2 from RR and 1 feature was selected from the category SOURCE.

4. DISCUSSION

There is a significant cross-industrial need to identify dissatisfied customers, since companies are no longer competing only on product quality but even more on a service quality level.[46–48] This work showed a method to classify dissatisfied customers from the automotive industry based on data produced within a service process with a maximum classification rate of 88.8%. Generally, the approach is designed to secure the practical implication by considering high accuracy and fast computing power as necessary success factors.

The paper contributes to the existing research as it predicts customers dissatisfaction solely based on objective data. This allows a proactive customer treatment in the future in order to achieve the management target to increase the overall customer satisfaction. Furthermore, sustainability to keep satisfaction high can be achieved once the indicators for dissatisfaction could be identified. The complexity of the presented data analysis was caused by isolated data sources and the resulting need for consolidation was shown. A stepwise data consolidation logic to overcome this challenge was derived and proven for the automotive industry. This methodology is not limited to the car sector but can be transferred to service industries, such as the health care or the telecommunication sectors.

The presented work answers two research questions. First, can dissatisfied customers be classified based on data that is produced during a service visit? Overall, the schema that aggregates the service events achieved best results throughout all experiments. Thereby, the SVM classifier (RBF kernel) outperformed the other classifiers and showed best results of 88.8% accuracy for test and validation data. Consequently, dissatisfied customers could be classified based on a service visit even before the service interaction is ending. Furthermore, the results showed that it is not a single event that turns a customer from satisfied into dissatisfied but rather the whole history of experiences the customer had. This is based on the fact that the aggregation of the service history yielded higher classification rates in all modeling approaches.

Second research question, can the dissatisfaction indicators be derived from service process data? This question is answered by the results from the experimental design that achieved results of 86.1% for the test and 88.3% for the validation partition based on basic scoring with evolutionary feature selection on aggregated data. Specific features that are highly important for the perceived dissatisfaction of a customer could be derived. The selected features were investigated in more detail. Most of the features were related to technical information from general car characteristics, warranty and diagnostic data. Furthermore, two features were related to the category of repeat repair. Based on these findings the potential indicators of dissatisfaction could be deduced from the technical data that were created when the car is still at the dealer. Additionally, repeat repairs in general are critical events that could be important to identify customers that may need specific care. This matches to the literature findings that showed repeat repairs to be highly influential for the perceived satisfaction.[9,49]

Generally, the presented method of linking technical data to customer perceptions is a novel approach that could be used in individual business sectors and thus create value for the service industry in the future. For instance, the need in the health care sector has already been identified.[10,17,50] With the consolidation of diagnostic results with the subjective patient expectations, a personalized patient treatment could be developed. The knowledge about potential dissatisfaction indicators generates huge business value as it prevents customer churn and thus reduces reacquisition costs.[6] A proactive approach which takes special care of potentially dissatisfied customers has a very low effect in case of false positives. If satisfied customers are classified as dissatisfied, the business result will be a more specific treatment. Thus, only the dissatisfied customers that are not identified have a real negative impact.

5. SUMMARY AND OUTLOOK

This work provided an approach to answer two important questions that help the automotive industry to increase the customer satisfaction and thus securing competitive leads. Can dissatisfied customers be classified based on data that is produced during a service visit? Can the dissatisfaction indicators be derived from service process data? In the presented data analysis procedure, the challenge of combining technical data from various sources with subjective survey responses was solved with a specific data consolidation logic. An experimental design consisting of three modeling approaches combined with two data preparation schemes was performed to answer the above research questions. The methodology that classified customer dissatisfaction best was the SVM classifier with a RBF kernel, which resulted in an accuracy of 88.8%. The indicators for dissatisfaction were identified within the feature selection experiment and the SVM with a RBF kernel setup achieved the best results. Knowing these indicators allows a proactive customer treatment as dissatisfied customers can directly be identified before the customer-service interaction is completed. Thus, actions to increase the satisfaction can be implemented immediately and secure a competitive advantage.

Furthermore, managerial implications can be derived based on the dissatisfaction indicators that have been identified by the evolutionary feature selection. In total, 46 indicators that have a high influence on dissatisfaction were identified. Most of them are related to car characteristics, such as the car model. Thereby, individual customer treatment campaigns can be designed according to this information. Additionally, 21 indicators were related to technical processes (diagnostics, warranty and data source information). Management can use this information to optimize the internal technical processes as these have a significant influence on the perceived service quality. Furthermore, these technical indicators need to be analyzed in detail in order to avoid service failures. Repeat repairs can be one of these service failures. They were identified as a main dissatisfaction indicator. Management should develop campaigns to reduce the rate of repeat repairs, especially the customer perceived ones. Therefore, a transparent and detailed explanation of the conducted work by the service advisors may be one opportunity. In general, the management of the automotive industry can benefit significantly

from the derived indicators for dissatisfaction.

The presented approach is intended to be implemented in business areas of the automotive industry. Therefore, the runtime of the application is of significant importance regarding the practical use. In order to allow a short runtime, cross validation has not been performed for the evaluation of the classifiers as this would increase the runtime significantly with the available machine. Therefore, Monte Carlo Scoring and the split of all observations into three subsamples (training, test and validation) instead of two (only training and test) have been applied. Statistical tests were applied to the Monte Carlo Scoring in order to investigate the significant comparison of the accuracy distributions. Further research should investigate the improvement of the classification rates by applying n-fold cross validation on a more powerful hardware environment.

The developed approach allows to identify potentially dissatisfied customers based on service process data received from dealers. Thereby, it is possible to classify dissatisfied customers before the service interaction ends and the customer is leaving the dealer. However, in future service visits at a dealer may decrease due to the implementation of remote features in connected cars. Further research should make use of online data that will identifiy the status of the car and is transferred online from the cars the the automotive companies. Thereby, real time customer services can be established that react on potential failures in the car and solve them online based on this received data.

The presented method was designed for the automotive industry and will be of high importance for other service industries. A closely related business where the approach might be suitable for is the health care sector. Research showed that satisfying patient-physician interaction is highly important to survive in the tough competition.[46] Diagnostic data that were collected during the treatment could be consolidated with patient satisfaction survey responses. Thus, dissatisfied patients may be classified and indicators for dissatisfaction may be derived. The adaptation of the method for the health care sector and further industries should be investigated. Therefore, survey responses from other service industries are needed and should be combined with service process data using the presented consolidation logic and the experimental design.

REFERENCES

[1] Mearian L. Self-driving cars could create 1GB of data a second. Computerworld 2013 July Accessed on 09/08/2016. Available from: http://www.forbes.com/sites/bernardmarr/2016/05/24/ big-data-a-game-changer-in-healthcare/3/#633a02d97 c0a

[2] Brito EPZ, Aguilar RLB, Brito LAL. Customer choice of a car maintenance service provider: A model to identify the service

attributes that determine choice. International Journal of Operations & Production Management. 2007; 27(5): 464-81. https://doi.org/10.1108/01443570710742366

[3] Hünecke P, Gunkel M. The influence of After-Sales service determinants on brand loyalty within the premium automotive industry: An empirical comparison of three countries. Service Science. 2012 December; 4(4): 365-81. https://doi.org/10.1287/serv.1120.0029

[4] Gebauer H, Krempl R, Fleisch E. Service development in traditional product manufacturing companies. European Journal of Innovation Management. 2008 May; 11(2): 219-40. https://doi.org/10.1108/14601060810869875

[5] Gaiardelli P, Saccani N, Songini L. Performance measurement of the after-sales service network - evidence from the automotive industry. Computers in Industry. 2007 September; 58(7): 698-708. https://doi.org/10.1016/j.compind.2007.05.008

[6] Xia GE, Jin WD. Model of customer churn prediction on support vector machine. Systems Engineering-Theory & Practice. 2008 January; 28(1): 71-7. https://doi.org/10.1016/S1874-8651(09)60003-X

[7] Bhattacharya CB. When customers are members: Customer retention in paid membership contexts. Journal of the Academy of Marketing Science. 1998 December; 26(1): 31-44. https://doi.org/10.1177/0092070398261004

[8] Reichheld F, Sasser WJ. Zero defections: Quality comes to services. Harvard business review. 1990; 68(5): 105-11. PMid:10107082.

[9] Meinzer S, Prenninger J, Eberl M, et al. Increase customer satisfaction and quality management by applying predictive analytics on diagnostic data. In: Text- and Data Mining for quality analytics in the Automotive Industry, Heyer G, Luy JF, and Jahn A, eds., Leipzig, Germany; 2010 September. p. 17-27.

[10] Meinzer S, Prenninger J, Vesel P, et al. Translating satisfaction determination from health care to the automotive industry. Service Business. 2016 December; 10(4): 651-85. https://doi.org/10.1007/s11628-015-0284-z

[11] Müller W. Gaining competitive advantage through customer satisfaction. European Management Journal. 1991 June; 9(2): 201-11. https://doi.org/10.1016/0263-2373(91)90085-5

[12] Chougule R, Khare VR, Pattada K. A fuzzy logic based approach for modeling quality and reliability related customer satisfaction in the automotive domain. Expert Systems with Applications. 2013 February; 40(2): 800-10. https://doi.org/10.1016/j.eswa.2012.08.032

[13] Kohl J, Kotucz A, Prenninger J, et al. Using multivariate split analysis for an improved maintenance of automotive diagnosis functions. In: Proceedings of the 15th European Conference on Software Maintenance and Reengineering (CSMR). Mens T, Kanellopoulos Y, Winter A, eds., Oldenburg, Germany. 2011 March. p. 305-8. https://doi.org/10.1109/csmr.2011.42

[14] Datta P, Masand B, Mani D, et al. Automated cellular modeling and prediction on a large scale. Artificial Intelligence Review. 2000 December; 14(6): 485-502. https://doi.org/10.1023/A:1006643109702

[15] Gopal RK, Meher SK. Customer churn time prediction in mobile telecommunication industry using ordinal regression. In: Pacific-Asia Conference on Knowledge Discovery and Data Mining, Springer: Lecture Notes in Computer Science. 2008; 5012: 884-9. https://doi.org/10.1007/978-3-540-68125-0_88

[16] Hung SY, Yen DC, Wang HY. Applying data mining to telecom churn management: Expert Systems with Applications. 2006 October; 31(3): 515-24. https://doi.org/10.1016/j.eswa.2005.09.080

[17] Yarnold PR, Michelson EA, Thompson DA, et al. Predicting patient satisfaction: a study of two emergency departments. Journal of Behavioral Medicine. 1998 December; 21(6): 545-63. PMid:9891254. https://doi.org/10.1023/A:1018796628917

[18] Sun BC, Adams J, Orav EJ, et al. Determinants of patient satisfaction and willingness to return with emergency care. Annals of Emergency Medicine. 2000 May; 35(5): 426-34. https://doi.org/10.1016/S0196-0644(00)70003-5

[19] Boudreaux ED, Ary RD, Mandry CV, et al. Determinants of patient satisfaction in a large, municipal ED: The role of demographic variables, visit characteristics, and patient perceptions. American Journal of Emergency Medicine. 2000 July; 18(4): 394-400. PMid:10919526. https://doi.org/10.1053/ajem.2000.7316

[20] Mavridou E, Kehagias DD, Tzovaras D, et al. Mining affective needs of automotive industry customers for building a mass-customization recommender system. Journal of Intelligent Manufacturing. 2013 April; 24(2): 251-65. https://doi.org/10.1007/s10845-011-0579-4

[21] Chan KY, Kwong C, Wong T. Modelling customer satisfaction for product development using genetic programming: Journal of Engineering Design. 2011 June; 22(1): 55-68. https://doi.org/10.1080/09544820902911374

[22] Lee JH, Park SC. Intelligent profitable customers segmentation system based on business intelligence tools. Expert systems with applications. 2005 July; 29(1): 145-52. https://doi.org/10.1016/j.eswa.2005.01.013

[23] Salini S, Kenett RS. Bayesian networks of customer satisfaction survey data. Journal of Applied Statistics. 2009 October; 36(11): 1177-89. https://doi.org/10.1080/02664760802587982

[24] Bandaru S, Deb K, Khare V, et al. Quantitative modeling of customer perception from service data using evolutionary optimization. In: Proceedings of the 13th annual conference on Genetic and evolutionary computation, Krasnogor N, eds., Dublin, Ireland; 2011 July. p. 1763-70. https://doi.org/10.1145/2001576.2001813

[25] Bandaru S, Gaur A, Deb K, et al. Development, analysis and applications of a quantitative methodology for assessing customer satisfaction using evolutionary optimization. Applied Soft Computing. 2015 May; 30: 265-78. https://doi.org/10.1016/j.asoc.2015.01.014

[26] Console L, Picardi C, Theseider-Dupré D. Temporal decision trees: model-based diagnosis of dynamic systems on-board. Journal of Artificial Intelligence Research. 2003 July; 19: 469-512.

[27] Struss P, Price C. Model-based systems in the automotive industry. AI magazine. 2003 Winter; 24(4): 17-34.

[28] Broggi A, Berte S. Vision-based road detection in automotive systems: A real-time expectation-driven approach. Journal of Artificial Intelligence Research. 1995 December; 3: 325-48.

[29] Kehl M, Enzweiler M, Froehlich B, et al. Vision-Based Road Sign Detection. In: Proceedings of the 18th International Conference on Intelligent Transportation Systems, Medina JJS, eds., Gran Canaria, Spain; 2015 Sep 15-18. p. 505-10. https://doi.org/10.1109/itsc.2015.89

[30] Prenninger J. Advanced Diagnostics and Predictive Analytics of Vehicle Data. In: Proceedings of the IECON 2013 - The 39th Annual Conference of the IEEE Industrial Electronics Society, Taiber J and Huang V, eds., Vienna, Austria; 2013 November 10-13. p. 1-21.

[31] Li Q, Mao Y. A review of boosting methods for imbalanced data classification: Pattern Analysis and Applications. 2014 November; 17(4): 679-93. https://doi.org/10.1007/s10044-014-0392-8

[32] Kotsiantis S, Kanellopoulos D, Pintelas P. Handling imbalanced datasets-A review: GESTS International Transactions on Computer Science and Engineering. 2006; 30(1): 25-36.

[33] Garcia V, Sanchez J, Mollineda R, et al. The class imbalance problem in pattern classification and learning. In: Proceedings of the Pattern Analysis and Learning Group, Ferrer-Troyano FJ, eds., Castellón de la Plana, Spain; 2007 Summer. p. 283-91.

[34] Reif M, Shafait F, Goldstein M, et al. Automatic classifier selection for non-experts. Pattern Analysis and Applications. 2014 February; 17(1): 83-96. https://doi.org/10.1007/s10044-012-0 280-z

[35] Wang CM, Huang YF. Evolutionary-based feature selection approaches with new criteria for data mining: A case study of credit approval data. Expert Systems with Applications. 2009 April; 36(3): 5900-8. https://doi.org/10.1016/j.eswa.2008.07.026

[36] Shah M, Marchand M, Corbeil J. Feature selection with conjunctions of decision stumps and learning from microarray data. IEEE transactions on Pattern Analysis and Machine Intelligence. 2012 January; 34(1): 174-86. PMid:21576745. https://doi.org/10.1109/TP AMI.2011.82

[37] Rodríguez-Fdez I, Canosa A, Mucientes M, et al. STAC: A web platform for the comparison of algorithms using statistical tests. In: Proceedings of the 2015 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE). Yazici A, et al., eds., Istanbul, Turkey; 2015 Aug 2-5. p. 1-8. https://doi.org/10.1109/fuzz-ieee.2015 .7337889

[38] Kruskal WH, Wallis WA. Use of ranks in one-criterion variance analysis. Journal of the American statistical Association. 1952 December; 47(260): 583-621. https://doi.org/10.1080/01621459.195 2.10483441

[39] McKight PE, Najab J. Kruskal-Wallis Test. In: The Corsini Encyclopedia of Psychology, Weiner IB and Craighead WE, eds. 2010, John Wiley & Sons. https://doi.org/10.1002/978047047921 6.corpsy0491

[40] Duda RO, Hart PE, Stork DG. Pattern Classification. Second ed. 2000, New York, NY, USA: Willey-Interscience.

[41] Hastie T, Tibshirani R, Friedman J. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Second ed. 2011, New York, NY, USA: Springer Series in Statistics.

[42] Polikar R. Bootstrap-inspired techniques in computational intelligence. IEEE Signal Processing Magazine. 2007 July; 24(4): 59-72. https://doi.org/10.1109/MSP.2007.4286565

[43] Schapire RE, Freund Y, Bartlett P, et al. Boosting the margin: A new explanation for the effectiveness of voting methods. Annals of Statistics. 1998 October; 26(5): 1651-86. https://doi.org/10.1 214/aos/1024691352

[44] Wu KP, Wang SD. Choosing the kernel parameters for support vector machines by the inter-cluster distance in the feature space. Pattern Recognition. 2009 May; 42(5): 710-7. https://doi.org/10.101 6/j.patcog.2008.08.030

[45] Oshiro TM, Perez PS, Baranauskas JA. How many trees in a random forest?, In: Machine Learning and Data Mining in Pattern Recognition. Lecture Notes in Computer Science, Vol. 7376, Perner P, eds. 2012, Springer. p. 154-68. https://doi.org/10.1007/978-3-6 42-31537-4_13

[46] Corbin CL, Kelley SW, Schwartz RW. Concepts in service marketing for healthcare professionals. The American Journal of Surgery. 2001 January; 181(1): 1-7. https://doi.org/10.1016/S0002-961 0(00)00535-3

[47] Sivakumar K, Li M, Dong B. Service quality: The impact of frequency, timing, proximity, and sequence of failures and delights. Journal of Marketing. 2014 January; 78(1): 41-58. https://doi. org/10.1509/jm.12.0527

[48] Yarris LM, Fu R, Frakes B, et al. How accurately can emergency department providers estimate patient satisfaction? Western Journal of Emergency Medicine. 2012 September; 13(4): 351-7. PMid:22942936. https://doi.org/10.5811/westjem.2011.1 0.6644

[49] Biehal GJ. Consumers' prior experiences and perceptions in auto repair choice. Journal of Marketing. 1983 Summer; 47(3): 82-91. https://doi.org/10.2307/1251199

[50] Holgers KM, Zöger S, Svedlund K. Predictive factors for development of severe tinnitus suffering-further characterisation. International Journal of Audiology. 2005 October; 44(10): 584-92. PMid:16315449. https://doi.org/10.1080/14992020500190 235