JOHANNES KEPLER
UNIVERSITY LINZ

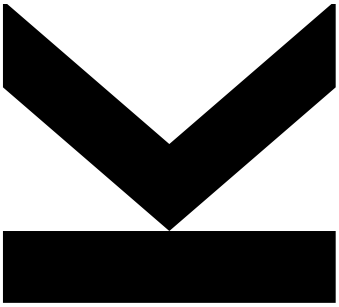
Submitted by
Jürgen Ratzenböck

Submitted at
Institut für Telekooperation

Supervisor
Univ. Prof. Mag. Dr. Gabriele Anderst-Kotsis

January 2018

Using Behavioral Analytics to reason about Customer Satisfaction in data-intensive Software systems



Master Thesis
to obtain the academic degree of
Diplom-Ingenieur
in the Master's Program
Computer Science

JOHANNES KEPLER
UNIVERSITY LINZ
Altenbergerstraße 69
4040 Linz, Österreich
www.jku.at
DVR 0093696

Affidavit

I hereby declare that the following dissertation "Using Behavioral Analytics to reason about Customer Satisfaction in data-intensive Software systems" has been written only by the undersigned and without any assistance from third parties.

Furthermore, I confirm that no sources have been used in the preparation of this thesis other than those indicated in the thesis itself. This printed thesis is identical with the electronic version submitted.

Linz, on October 29, 2017

Jürgen Ratzenböck

Acknowledgment

Hereby I would like to thank my supervisor Univ.-Prof. Mag. Dr. Gabriele Anderst-Kotsis for her support throughout the whole work on this thesis. She always provided me useful feedback and used her professional skills and knowledge about this topic to give me hints and advices when I needed it.

Abstract

Contents

Abbreviations

CRM Customer Relationship Management

GPS Global Positioning System

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Customer Satisfaction as Business Value

While companies were only focused on the quality of their products and services until the 1980s, increasing supply and resulting competition on the market caused a shift towards a more customer centric approach starting in the early 1990s [17]. The trend is visualized in figure 1.1.

Companies realized that the customer is an essential part of the company and has major influence on the success. Furthermore the competitive pressure has become higher over the last years due to the oversupply of products and services. Especially online services and products have become extremely popular and people usually have a variety of similar ones to choose from, which means that customers rarely stay at the same service provider due to convenience only since switching to the competitor has never been easier [21]. For businesses it gets harder to attract and recruit new customers. As a result, customer retention turns out to be a major business issue [15]. As of [4] the Boston Consulting Groups estimates that it costs 6.80$ to care about existing customers on the web and market them appropriately whereas acquiring a new web customer costs about 34$. The described shift impacts the strategy how products are presented and marketed to the customer. To achieve the goal of successfully binding customers to the provided products and services, it is not sufficient anymore to pursue a traditional mass marketing strategy like sending the same advertising prospect to all the customers in a given area. This can even lead to a negative attitude of customers towards a company and to a potential break up of the relationship [21, 17]. Instead companies have to treat customers individually as they differ in their personality, behavior and needs. In a company which is centered around the customer, products have to be built to meet the customer needs [4]. Apparently, no company can ever know each customer and his/her needs in depth at any point in time without collecting useful data about them. With the beginning of the 21st century a new hyped business term

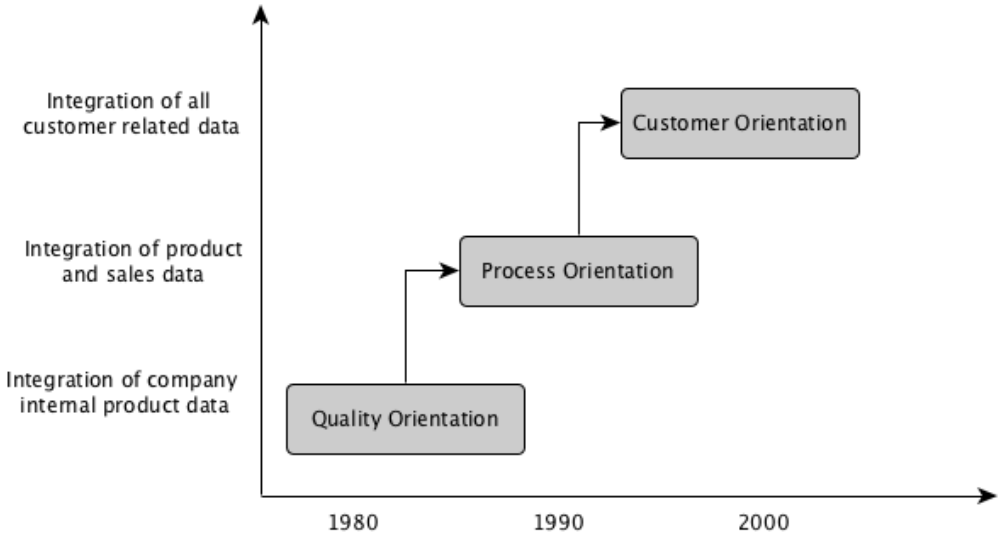


Figure 1.1: Development towards a customer centric company [17]

has been established, namely CRM (Customer Relationship Management). It can be described as the comprehensive process of knowing customers needs and wants, presenting the correct products and services via the preferred channels to them, allowing a convenient purchase of the selected product or service and providing good care after the purchase [21]. CRM can be seen as the core process to establish profitable long-term relationships, develop customer loyalty and thus increase the value of the company. Although it has become widely recognized, different researchers look at it from various viewpoints and there is no unique valid definition. Some of them put their focus more on the processes and tools whereas others put more emphasis on the marketing aspect. All these different definitions available in literature have a common property of understanding the customer in detail [16]. However, the way of creating profitable customer relationships is hard and takes place on a fine line. It can be imagined as a pyramid where each level is based on the fundament below and removing one level can lead to destruction of the whole construct. The pyramid is visualized in figure 1.2 and will be discussed in more detail.

1.1.1 Customer Use

With a seemingly unlimited number of marketing instruments and possibilities to influence people and drive them into a specific direction, it is often overseen that the basis for any customer relationship is the actual use of the product to solve an existing problem. If people do not see any sense of a product to solve a particular problem or make something easier in their lives, they will not buy it and any of higher levels on

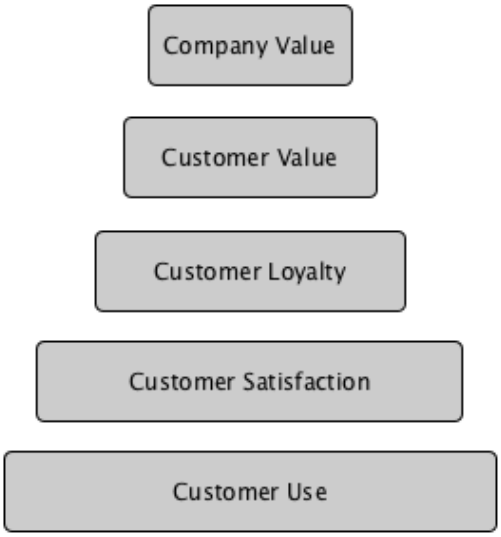


Figure 1.2: Illustration of the way towards increasing company value [17]

the pyramid will not work. Any successful product has its basis use case which solves a customers problem. However, in addition to this core use, products have to provide some extra value in order to communicate some unique value which makes the product standing out from the competitors. [17]. In the service sector this can be summarized as Service quality [9]. A good example for its customer use is definitely Apple Inc with their iPhone. In a simplified manner they manage to provide the base value of a current Smartphone and since many years they are able to stand out of the crowd with their unique and elegant design which gives customers a positive image and increases their reputation.

1.1.2 Customer Satisfaction

When people consider a product as useful for their purposes and decided to buy it, they experience some feeling after its use which can be either positive or negative. The general applicable definition of customer satisfaction according to [17] is the comparison between individual expectations of the customer (created by his/her ideal imagination, previous experience with the company, status of the company or publicly available ratings and recommendations of other customers) and the perceived experience after usage of the product. If his expectations were higher or equal, the customer can be considered as satisfied whereas vice versa the customer is dissatisfied since his expectations were not fulfilled. As obvious from this definition, the customer use or Service quality is an important metric which is also reflected by the pyramid figure. [9] proposed a detailed customer satisfaction model in the airline service business which is visualized in figure

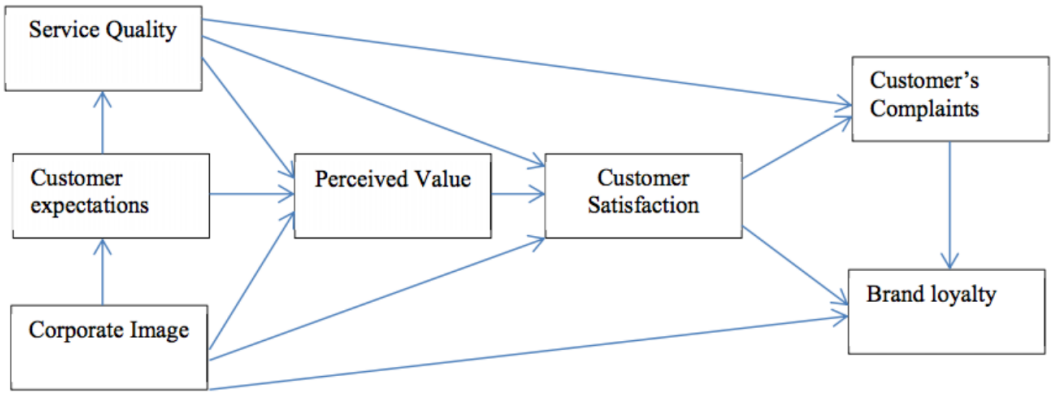


Figure 1.3: Customer satisfaction model [9]

1.3. The following few paragraphs will especially take an eye on the relationships and influences among customer satisfaction- and loyalty.

This model conforms in most parts with the opinions of [17] as it indicates a strong relationship between Service quality and customer satisfaction. The perceived value describes the subjectively perceived ratio between quality and price. If this number appears to be high for a customer, it indicates correct marketing activities. As a result the perceived value directly affects customer satisfaction as well. In addition [9] also claims that the reputation (indicated as Corporate Image in the figure) of the company influences satisfaction. Although this variable plays a role in this model it will not be considered as a core point in the practical work which puts its focus on service usage- and quality influences. What [17] does not outline is the reaction of a customer itself. According to [10] this can be called the response of a customer which has a specific focus. This focus is hard to define in a unique way since it can have various facades. A positive experience with a product leads to an emotional effect which can for instance create happiness as temporary condition but it can also have a behavioral impact which can yield changes in the usage or attitude towards the product. Moreover according to [9] there is the nasty statistic that a dissatisfied customer tells on average nine other people about his/her negative experiences with the product or company. With the heavy use of the Internet and Social Media this can get spread even much faster [17]. [10] mentions as third dimension that a customer response does not remain stable over time and varies among product usages. These changes over time are caused by several types of actions like an interaction with the product or after a purchase. This fact emphasizes that measuring customer satisfaction is an ongoing process which requires much more than an occasional estimation once in time to do it right. Although influencing factors like emotions or current mood are not accessible to the service provider at any time there is still some chance to analyze reflected information of users to gather insights into what causes a user to be more (dis)satisfied and to reason about his/her satisfaction level.

1.1.3 Customer Loyalty

A customer is referred as loyal if he/she shows a positive attitude towards the service provider and thus recommends the company and its products to other people and also is willing to stay with the company in the future. Only if these two conditions are both met a customer can be clearly considered as loyal. There is also the possibility that he/she stays with the company over a longer period and does several purchases but only because there are too few other possibilities on the market ore switching to a competitor requires too much effort [17]. The customer satisfaction model of [9] indicates the loyalty parameter with "brand loyalty" and agrees with the opinion of [17] with regard to a direct relationship between customer satisfaction and loyalty. Even though there is no clear evidence for causal relationship between customer satisfaction and customer loyalty so far, it has been found out that these two variables correlate positively. Customers who are satisfied with the overall product package including the provided services are more likely to buy again and stay with the company. In contrast, unsatisfied customers can cause big financial damage since they are not only vulnerable themselves to end the relationship but also influence other interested people and potential customers negatively [17]. Loyal customers positively impact profitability by increasing sales due to repetitive purchases, reducing marketing and operational costs because people know certain products and do not demand for much more specific information anymore [2]. The difficulty is to retain a rather high number of loyal customers since small mistakes can lead to an abrupt end of the customer relationship [4].

1.1.4 Customer- and Company Value

1.1.5 Customer data

Due to the ongoing innovation in computer and software technology, businesses can collect huge amounts of data of their customers and integrate it in a way to be able to gain deep knowledge on the behavior, situations, desires and problems of them. Advanced data mining techniques allow to analyze behavior patterns of customers and predict future events and actions taken by them. The computational power of todays hardware enables businesses to store every single touch point of each customer in database systems and disregarded of the communication channel integrate it into a data warehouse to provide a comprehensive view onto the profile of a customer [4]. Although the recent technology progress opens a wide range of new opportunities to record and monitor communication and interaction of customers with a business, it turns out that

measuring satisfaction of customers and pro-actively reacting to his/her situation is a complex task [17].

1.2 Thesis Statement

1.2.1 Problem

Since the strength of customer satisfaction is connected to the attitude and loyalty towards the providing company, it turns out to be an important indicator for revenue and company value. Research showed that an 5% increase of customer retention has a positive impact of 25 to 125% on the profit [2]. Many modern Internet businesses try to live the "Customer first" principle and therefore invest resources to make their customers as satisfied as possible. They usually employ a first-class customer support service to quickly tackle problems and complaints which undoubtedly is important but since on average only 5% of the customers actually contact customer support service when experiencing problems it demands for more sophisticated solutions to understand customers in detail [17]. The big issue in reasoning about satisfaction is that it cannot be directly measured and determined whether a customer is satisfied or not. Online surveys have become quite popular to collect data about attitude and satisfactory level of customers but they lack of covering every parameter which can have some influence on how people feel about a product or service. Moreover they only reflect the current situation and therefore do not include any past events and touch points. Exactly these experiences in the past could have changed customers minds [17]. To be able to explore a customer in more detail, new methods and techniques in the field of data mining have been proposed [17].

1.2.2 Research objectives

The research question this thesis deals with is to find out how to leverage useful data related to the behavior of customers, interaction with a service respectively product and reported data from product usage, to derive patterns which allow to make a statement about how (dis)satisfied a customer is. Furthermore the results should show which data have major influence on customers satisfactory level and thus outline the essential product characteristics a service provider has to pay attention to. If the results can be considered as reliable and valid, they can be used for early detection of unsatisfied customers and help to pro-actively solve their problems. The route towards the goal of the research looks the following:

- Define what customer satisfaction means and how to differentiate between satisfactory levels.

- Elaborate which data has potential to influence customer satisfaction regarding the illustration example used during this research.

- Implement statistical tests to which metrics in the data cause major changes in the situation and decision behavior of a customer.

- Implement a software solution to analyze behavior of customers and based on similar customer data do a predictive analysis of the satisfactory level and impact on loyalty.

1.3 Case Study: Tractive

As representative example the research relies on the data of a pet tracking company named Tractive GmbH. The company which was founded as a startup in 2012 in Pasching, Upper Austria produces hardware devices and software applications for different kinds of pets. Their most popular product is a GPS (Global Positioning System) device which can be put on a pet and allows the customer to track its position live on a mobile phone or via a website. This way Tractive helps pet owners worldwide not to lose their pets again since they can always have a eye on them via a smartphone or desktop computer. Since the company is quite successful with about 50k paying customers using a Tractive GPS device, there is also a lot of data available related to each customer. This starts with data related to the usage of the hardware device ranging to subscription data of a customer and his/her interaction with customer support service. Since the company is built up on a subscription model which requires customers to pay on a regular basis for the usage of the device, there is special interest in binding customers for a long time to the company. In essence, this means that satisfied customers are a key asset for the company. Therefore its model should fit well to the research objectives and promising results could support solving customer issues earlier and as a result affect drop out rate positively.

Figure 1.4: Tractive - Official company logo

1.4 Structure of the Thesis

The thesis will contribute to a more efficient way of measuring customer satisfaction using large sets of collected data related to usage behavior. Chapter 1 first gave an introduction about the general problem context and outlined the research objectives as well as the illustration example this thesis is based on. Chapter 2 will take a closer look at related research work. It will give an overview on existing approaches to determine customer satisfaction and reasons about their efficiency, problems and improvement potential. Moreover this chapter outlines necessary background research done to gain a deeper understanding on what it means to reason about satisfaction and sheds some light onto different methods suitable for the implementation part. After the fundament has been built, chapter 3 illustrates the practical implementation in detail. Chapter 4 evaluates the outcome from the implementation and analyzes why some approaches work better than others. Chapter 5 concludes the work by summarizing the important findings and gives a short look into future work.

Chapter 2

Existing Approaches in Behavior Analysis and Detection of Customer satisfaction

2.1 Interpretation of Customer satisfaction

The beginning of research work in terms of measuring and analyzing how satisfied users are with a computer system already dates back to the beginning of the 80s. At these times first researchers supposed that there is a relationship between user satisfaction and the success of a computer system. From this point on, a series of researches investigated which factors could influence user satisfaction and how to ask users to get a representative opinion [20]. Bailey & Pearson were first to propose a questionnaire consisting of 39 satisfaction dimensions which should help information system providers to evaluate whether a user is satisfied with the system or not. Ives, Olsen & Baroudi did some further research based on the existing work from Bailey & Pearson to reduce the effort for users to answer 39 dimensions. However technology changed dramatically as the shift towards the use of personal computers started a few years later and therefore new approaches were needed. According to Doll and Torkzadeh the previous developed methods did not focus on the satisfaction with a specific end-user application since they do not cover the human-machine interaction [20]. Thus, they developed a model specifying the end-user satisfaction as a second factor driven by five first-order factors namely content, format, accuracy, timeliness and ease of use. After testing four different models regarding validity and reliability using a confirmatory factor analysis approach against some sample data, they came up with the following model visualized in figure 2.1.

In the results for this model they claim that 74% of the variation in the five first-order factors is explained by the end-user satisfaction. [5]. This second-factor approach

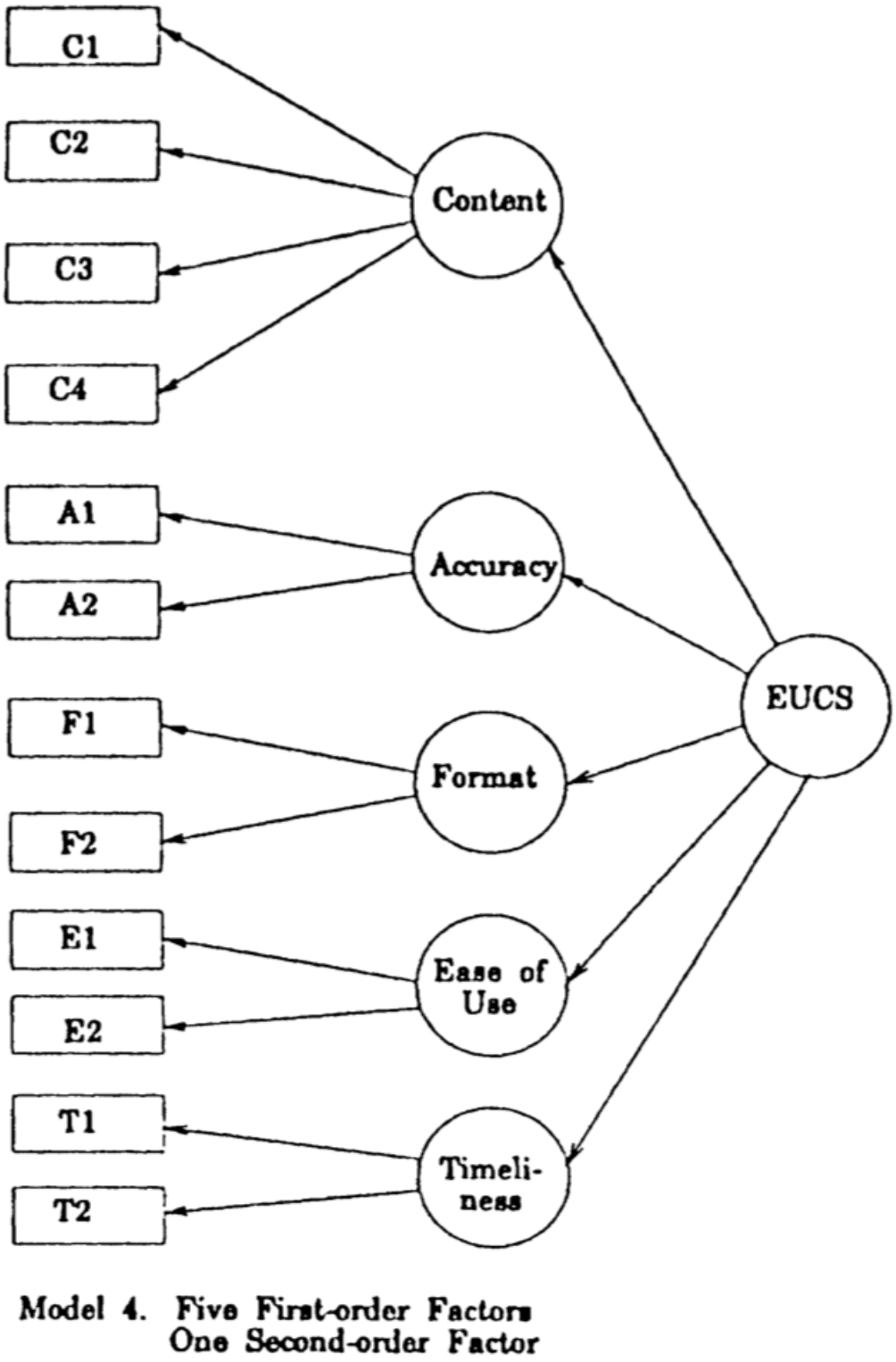


Figure 2.1: Customer Satisfaction - Five first-order factors [5]

turned out to be quite successful over the following years and has been widely used to reason about user satisfaction for specific applications. The web has been evolving rapidly and due to this reason [24] analyzed in their research in 2002 whether the existing method of Doll and Torkzadeh is still appropriate and applicable to web-based information systems. They recognized that some of the previously identified factors may not be relevant anymore for web-based information systems. Based on the work of Doll and Torkzadeh, a questionnaire was created to ask people how satisfied they are with a web based information system whereby they decided to choose Internet Portals as representative example. The results of this research showed that the existing five factors are still valid under these new circumstances. However, the research clearly has some limitations as it only considered Internet Portals and did not look on additional factors like privacy or security [24]. In 1999 [20] took a closer look on several existing measurement approaches and discussed their limitations in practice. In contrast to [24] the results of this research tend to be more critical regarding practical usefulness of approaches like the one from Doll and Torkzadeh. They criticize that surveyed users are considered with equal personality and behavior whereas in practice each individual user is different and this would therefore require an independent consideration in a survey. Furthermore the research claims that existing survey approaches are often too inflexible since a service provider has specific objectives and the survey has to put more focus on certain aspects than others which cannot be achieved without modifications. Meanwhile researchers recognized that besides extracting knowledge about customer satisfaction explicitly, there is another promising opportunity to implicitly understand behavior of customers based on how they present themselves, act and behave. Supported by the rapid growth of software technology and tools the era of statistical analysis and data mining in CRM started up in the beginning of the 21st century [16] [17]. The research paper of [16] wrapped up CRM quite comprehensively by taking a closer look onto 87 selected published journals dealing with CRM and Data Mining techniques. They tried to find out how the distribution of publications among the different areas of CRM looks like and as a result came to the conclusion that statistical and data mining techniques are especially in customer retention of great interest. A promising approach reasoning about dissatisfaction of customers and its connection to churn was proposed by [14]. This research project first of all analyzed influencing factors driving customers in the wireless telecommunications industry to stay or leave the service. After identifying major factors differentiating between satisfied and dissatisfied customers, statistical machine learning techniques as logistic regression, decision trees and neural networks were employed to predict churn rate for a selected time period. Later on [14] outlines a calculation model to determine under which circumstances it makes sense to offer an incentive to a potential churner and take the opportunity to pursue him/her to stay. Based on the variables of lost revenue due to a churner, acquisition costs of a new customer, the probability of staying after offered an incentive and the cost of offering an incentive to a customer, an cost saving per customer could be calculated. The results clearly showed that in most cases taking the effort of offering incentives and thus

preventing churn results in higher profitability. Due to the similarity, considering the recurring payment service type as well as properties and nature of competition on the market, with the illustration example outlined in 1.3, the research of [14] supports the practical work conveyed by this thesis. Another related research work published by [12] dealt with analysis of customer behavior and its impact on retention and profitability for a European Financial Institute. Using a data set of about 100k customers and advanced decision tree algorithm, namely Random Forest, was employed to perform a binary classification based on the properties of whether a customer

- buys a further banking product in the future,
- cancels a non-ending relationship with a purchased product,
- or causes a profit drop within a considered time period.

In a further stage the Random Forest algorithm was also used to analyze which of the identified independent variables affect the outcome factors related to customer retention remarkably. As surprising side effect it was found out that some of the variables show an influencing effect in all three binary classifiers although they seem quite contrary as the new-buy and cancellation of a product for instance.

The previous work from [14] and [12] already shows the power of statistical and machine learning techniques to predict future customer behavior and its effect on retention and profitability. It was clearly indicated that past customer behavior data turns out to be worthy in recognizing patterns, relationships and predicting the future. Since customer satisfaction is usually a rather complex construct consisting of several dimensions besides purchase, cancellation and profit evolution of a product, there is enough potential to reveal hidden patterns and answer business critical questions regarding customers behavior. Moreover statistical tools and data mining techniques have evolved further and demand for a reappraisal. According to [8] only half a percent of the data available in the digital universe is analyzed but in 2020 about 33% of all data may be valuable which supports the assumption that there is still a lot of work ahead.

2.2 Concepts

Analyzing a customer's attitude towards a product, recognizing whether his/her expectations are met and providing the desired value at the needed time, is usually a difficult challenge and requires a company to setup a well coordinated working process. [16] illustrated CRM as a framework comprising four core dimensions whereby on each

of them, integration of knowledge about customers can be used to increase company profitability [17]. Before looking in greater detail on the technical concepts of how to organize collected data about customers and finding useful information hidden in those large datasets, the thesis will take a look on the CRM dimensions proposed by [22]. The aim is to outline the various tasks of CRM but also clearly mark out at which particular dimension(s) the focus of this thesis lies.

2.2.1 Customer Identification

No successful company launched their first product and immediately interested customers recognized the potential of the product and bought it. In contrast potential customers first of all have to be found by a careful segmentation following by an analysis of these identified segments which should lead to concrete customer segments one wants to target [22] [16].

2.2.2 Customer Attraction

The concept of attracting customers as a major prerequisite to develop trust and commitment between a seller and a buyer has already been outlined by [6] in 1987. The goal of this part in the CRM Framework is to raise attention by potential customers and present outstanding features and differentiators of the product which lead to fulfilling needs of the customer [7].

Chapter 3

Approaches for analyzing and predicting Customer Satisfaction

This chapter of the thesis focuses on the design and implementation of promising approaches which were introduced and described in more detail within the previous chapter.

3.1 Identification of relevant data sources

On the one hand, statistical analysis and data-driven approaches as they were suggested in the previous chapter are rich tools to reveal gain new insights into the data and find hidden associations and relationships but on the other hand these approaches will not deliver any satisfying results if they work with wrong or invalid datasets as input. Therefore it is essential to invest enough resources on selecting relevant data sources and ensuring a high quality within this data before starting with any analytical approach. Considering right and valid data was not only mentioned once in literature as a key factor for successful data analysis [17].

3.1.1 Selecting the right data based on its representativeness regarding Customer Satisfaction

The approach this thesis followed for selecting relevant data sources is based on the definition of Customer Satisfaction as it was modeled in section 1.1.2. Furthermore, as already indicated in section 1.1.2, the focus is on the provided service quality and customer usage, since this can be represented in data and as a result is measurable. First of all, it started with a summarization of which properties and features of the

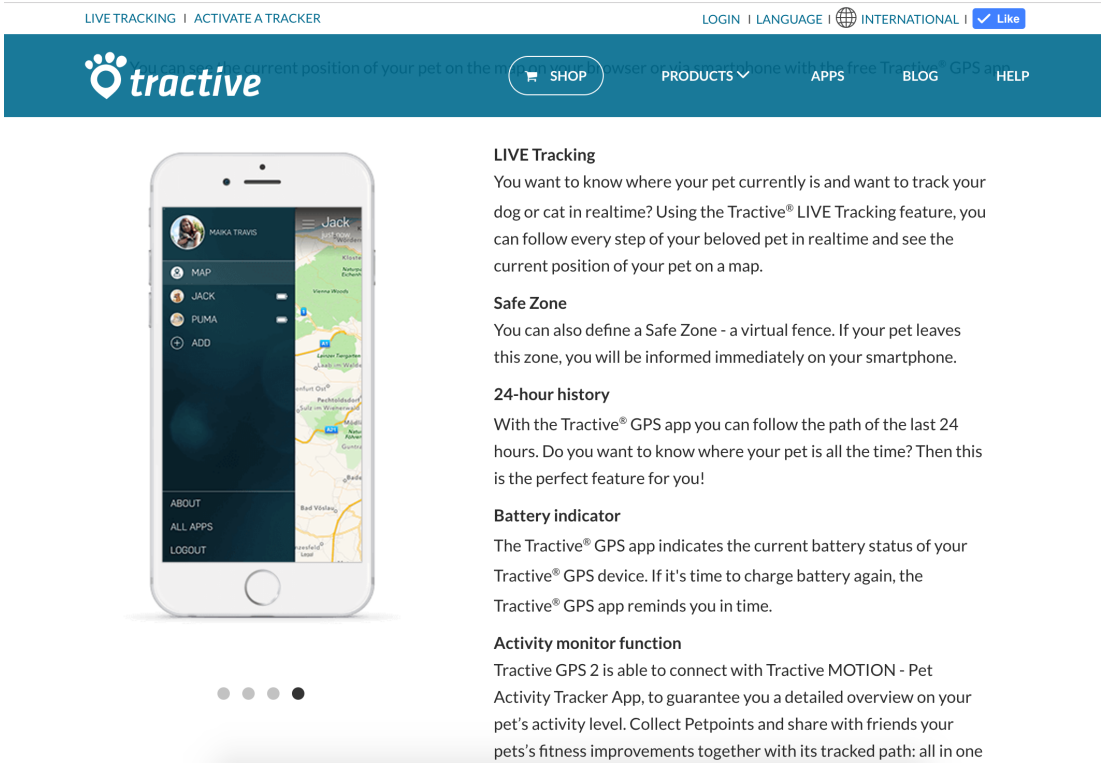


Figure 3.1: Extract of landing page https://tractive.com

main product considered in the use case scenario, namely the Tractive GPS device, are advertised actively to potential customers. As representative resource the landing web page of the company was used to extract this information. This website is the major source of customer conversion and lists all relevant characteristics and features of the product. Figure 3.1 shows an extract of this web page.

Other sources used in marketing like social media channels provide brief summarizations of this content and link to the main website. Therefore it can be implied that product descriptions on the landing web page are decisive for customers and drive their expectations with regard to the product. Table 3.1 gives an overview on important properties and features from a customer's perspective and assigns it to the data which will be useful for reasoning about Customer Satisfaction. It indicates what a customer can expect when he purchases the product.

The table excludes properties which are advertised but where the nature of the feature is static and stays always the same among all customers. Such properties are usually product characteristics which do not change due to customer use and as a result are not measurable in collected data. An example is the handy charger or the fact that the GPS device is one of the smallest and lightest devices for pet tracking available on the market. Since the analytical part of the thesis should provide highly objective results mainly based on user actions and the resulting usage experience, any customer specific

Feature / Characteristic	**Description**	**Nature of data suitable for predicting Customer Satisfaction**
Locate pet anytime, anywhere	See location of pet accurately on a map on the Smartphone	Position data (GPS, Mobile Cell)
Live Tracking	Follow the trace of a pet in realtime on a smartphone or on a web page	Live Tracking commands statistics
Safezones	Creating a virtual fence and get notified if pet leaves selected safe area	Number of times pet leaves and enters safezone, reliability data regarding notifications
Battery indication	If battery of GPS device is full or is nearly empty, it will be indicated in the Smartphone app	Reliability of data regarding battery notifications
Integrated light	Customers can turn on the integrated LED (Light-Emitting Diode) on the GPS device via the Smartphone	LED command statistics
100% waterproof	A product characteristic customers trust in, since many pets often get wet	Hardware defects due to water damage
Premium Customer Service	With a premium service plan, it is promised that customers get feedback within 24 hours on weekdays	Customer service data related to ticket resolving times

Table 3.1: Features / characteristics of product among with their nature of data suitable for predicting Customer Satisfaction.

data like personal or demographical data will not be included. Since customers with totally different cultural backgrounds can differ a lot in usage behaviors and attitude towards Customer Satisfaction, this kind of data will bias correlation or prediction results and is therefore not considered as relevant.

3.1.2 Elaboration on data collection regarding identified features

This section will have a detailed look into availability, representation and content of stored data for the remaining features from table 3.1. The procedure this thesis followed for this task started with an investigation to find out by which of the data collections each feature is represented best and how much value in the content is included. Following paragraphs assign identified data sources to categories to introduce a structure which makes it more understandable for the reader. It is not the aim of the following paragraphs to discuss every data attribute which could contribute to predicting Customer Satisfaction in depth. Instead the following paragraphs will rather explain briefly which type of data is stored in the collections and how the data is generated through customer behavior to get an idea why they can be important when it comes to Customer Satisfaction prediction.

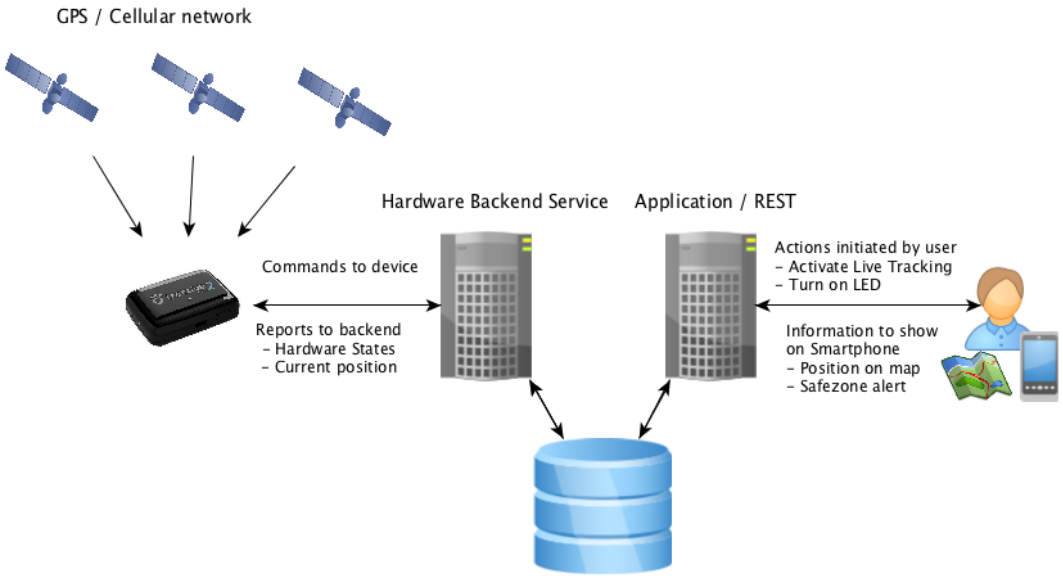


Figure 3.2: Overview on communication and data flow between client and device

3.1.2.1 Device related data

This group contains a bunch of potentially valuable data resulting from device usage. This type of data is created by sending of data from the device to a backend service and storing it in the database or sending commands from a client to the backend and finally to the device which should initiate some action. Before looking on concrete schemata of the data, figure 3.2 illustrates in a simplified way the communication and thus the data exchange between client and device.

The following list outlines the essential device related data collections leveraged in the analytical part of the thesis.

- **Device id reports** are sent automatically from time to time when the tracker is switched on and has network coverage. The Id reports contain information about the firmware- and hardware version of the device. Since there are bug fixes and improvements on the firmware quite often and updating it on a device is possible over the air, this can lead to differences in service quality among devices. Furthermore the device id reports allow to imply the activity level (eg.: the number of days in use) of a customer's tracker.

- **Device position reports**: Every time a tracker receives a position it is stored in the database. Along with the latitude and longitude of the position there is data indicating whether GPS or mobile cell network was used for localization, how strong the received signal was, as how accurate the position is classified, the

timestamp when the device received the position, the time where it was stored in the database and some further technical GPS data as the number of reachable satellites used for localization.

- **Device network reports** contain data about current network coverage. Since the device can be used in over 100 countries worldwide each network report provides the MCC (Mobile Country Code) which indicates where the device is currently used. Prior experiences showed that network coverage sometimes varies a lot among countries which can influence satisfaction of a customer. The mobile network generation which is limited to GSM and UMTS in the use data of this thesis can have an impact as well.

- **Device hardware reports**: These reports are usually sent along with the id reports described before and contain different kind of hardware states. Interesting attributes are for instance the current battery level- and voltage, temperature states or collected number of position, error and network logs for the particular device. Moreover hardware reports store specific hardware events for a point in time where they happened. Amongst others this includes valuable information as the time when the device has been switched on/off or battery is charging, full or low.

- **Geofence reports**: If a customer's pet with a mounted device leaves or enters a virtual fence defined in the app, a geofence report with trigger (called In-break in case of entering the safezone or Out-break otherwise) and source (GPS or mobile cell) is sent to the server which stores it into the database and initiates sending a push or mail notification to the customer. Both accuracy based on the trigger source and the number of In- and Out-breaks can influence Customer Satisfaction.

- **Server commands**: These are commands which are sent due to an action initiated by the user. The most prominent server command according to the statistics so far is the Live Tracking function where the device starts sending a new position every few seconds for a maximum of ten minutes which allows the user to follow his/her pet in real time. Further commands enable the user to turn on/off the integrated light or trigger a sound on the device. The server commands are very essential for customers of the Tractive GPS device and therefore several attributes indicating the quality are collected. For each server command the timestamps when the command was sent by the server to the device, when it was commanded by the device and the time of confirmation respectively cancellation by the device is recorded.

3.1.2.2 Notification related data

Notifications play an essential role for Tractive customers, since the information whether a pet has left a defined safezone or the battery state of the device is critical can be crucial for users. The assumption made in this thesis relies on the fact that reliability of notification delivery either via mail but even more important via push on the Smartphone has an influence on Customer Satisfaction. As mentioned in the device related data geofence reports are sent in case the device is not within the safezone area and stored in the database. Any mail- and push notification sending attempts no matter whether they were successful or not are stored in push logs. Each push log entry contains a reason for sending it, the message, a reference to the tracker in case of a safezone- or battery alert and the return status indicating the success state.

3.1.2.3 Customer service related data

Tractive advertises a first-class customer service which a lot of customers actually make use of. Tractive for instance promises feedback within 24 hours. Instead of responding to formless emails, Tractive employs the popular customer support service tool Zendesk® which is the central place for desires, complaints and support requests. Next to better data integration there is the advantage that all collected information regarding a particular support ticket is stored transparently and can be retrieved if needed. Thus, this enabled the opportunity to analyze the time passed until a ticket is handled which is also assumed to have an impact on the overall satisfaction of customers.

3.1.3 Quality of identified data and preprocessing

The relevant data sources are identified but before diving into the implementation of analytical processing the thesis took a look into quality of the data and modified or enriched it if considered as necessary. The aim was to provide correct and analytical oriented data as input for the actual analysis and data-driven approach.

3.1.3.1 Representation in database system

As [17] outline in their research, one of the major prerequisites for a successful data analysis is that the available data provides a unique view on the real world and prevents ambiguous data collections. Tractive uses the semi-structured NoSql database MongoDB as central storage system for any transactional customer and device related

data. This has the big advantage that identified data sources can be taken from the same database system and thus provide a unique interface to query sample data for analysis and prediction. Since this thesis considers data generated by customers using a Tractive GPS device, inconsistencies among transactional-driven data are not a problem as it often is the case for companies with a diverse product palette [17].

3.1.3.2 Transactional- vs. analytical oriented data

When taking a look on the nature of the selected data, it can be noticed that it results from operative application. Although this huge amount of transactional data provides the opportunity for intensive analyses, it is not always on the desired level of detail. This was realized during the first few analysis steps and made it necessary to bring some data collections in another format. Section 3.1.2.1 outlined the server commands as critical contributor to Customer Satisfaction. However, the collected data attributes belong to one particular server command of a device. This would require extra calculations for each device during analysis to aggregate command success rates or delays. Moreover other device related information like the model edition, batch number, country of use or SIM provider is stored in another collection and therefore not directly available.

Due to the importance of these server commands for its business, Tractive decided to store beginning with mid of May the server commands in addition in a suitable form to make analysis, statistical evaluation and trend analysis more convenient. Therefore a new collection, namely server command metrics, was created. From this point in time on, every new server command causes a new entry in the metrics collection with a reference to the user who initiate the action, to the device along with its properties to filter and the server command specific information delivered from the hardware device itself. An extract of the schema of this collection is shown in table 3.2.

After one month of storing those metric entries for each device when a new server command is executed, it was decided to make use of this data within the thesis instead of relying on the transactional entries. The main advantage using this preprocessed analytical oriented data will be revealed when diving deeper into the querying tasks during data analysis.

Alongside with prominent advertised features, the work in this data gathering and preprocessing task also took opinions, complaints or desires under consideration to get a feeling of what additional properties seem to be useful for customers and make them (un)happy. Due to the lack of automatism, feedback from Tractives customer support service employees was considered as source of trust. It turned out that the battery life of a device, which is not advertised explicitly on the landing web page of Tractive, was

Attribute names	**Description**
hw_edition	One of the three available editions. (Normal, Pink or Hunter)
batch_no	Batch number
sim_type	SIM provider
iso2	Country code of usage
cmd_success_rate	Percentage of successfully issued commands
cmd_cancelled_rate	Percentage of canceled commands
cmd_terminated_rate	Percentage of terminated commands
cmd_delay_to_commanded	Delay until command is received at device
cmd_delay_to_confirmed	Delay until command is successfully confirmed by the device
cmd_delay_to_pos_any	Delay until any position is available
cmd_delay_to_pos_new	Delay until a new position is available
cmd_duration	Whole duration of a command

Table 3.2: Important attributes of aggregated server command metrics to be used for data analysis

often mentioned in contact with customers. As a result of these findings, the author of this thesis decided to include battery life time of devices in the Customer Satisfaction prediction approaches.

Due to the reason that battery levels are only reported as single events within hardware reports it was necessary to compute battery life time with following approach.

3.2 Hypotheses-driven approach to gain knowledge about interrelationships among data

Based on the prepared data analysis work started with a top-down approach as it was described in section 2. The aim was to get an understanding which types of data influence the output of other data and how strong these correlations are. In essence, the overall goal was to reveal which of the data related to customer usage is expressive regarding Customer Satisfaction. For this type of analysis, hypotheses were explicitly defined and either verified or falsified by statistical measurements. It was planned to use results if they are promising for a manual feature selection in an automatic framework for Customer Satisfaction prediction.

3.2.1 Choosing target data approximating Customer Satisfaction

While section 3.1 explored the available data sources to identify those which are likely to be predictors for Customer Satisfaction, the work explained in this section tried to find data reasoning about an effect on Customer Satisfaction. This target data should be as expressive to distinguish satisfied and dissatisfied customers. Due to the lack of any explicit satisfaction responses, implicit data instead had to be found. The author of this thesis proceeded by finding behavioral patterns customers would follow if they feel pleased or disappointed and came up with following collected data:

- Recurring service active or canceled: The assumption hereby is that on the one hand satisfied customers will keep their service where they pay monthly, yearly or biennially active and are therefore able to use the device anytime. On the other hand dissatisfied customers are vulnerable to cancel the service and leave the relationship with the company. These thoughts are based on the theory of relationship between Customer Satisfaction and loyalty outlined in sections 1.1.2 and 1.1.3.

- Increased or diminished app usage: The assumed behavioral property resulting from (dis)satisfaction is that customers increase respectively diminish the usage of the smartphone app. According to the analytics data the Tractive GPS app for iOS and Android is most important for customers of a Tractive GPS device. As a result of good user experience it is expected that customers use the app more often and the same vice versa. The data indicating events for opening the app on a smartphone as well as bringing the app into foreground were taken as representative indicators.

- How many days GPS device is in use: The device id reports also mentioned in section 3.1.2.1 can be seen as an activity indicator when aggregated over a period of time. The thesis considered the number of days the device sent id reports as a meaningful number depending on increased of decreased satisfaction.

3.2.2 Formulation of Hypotheses and solving analysis problem

Before setting up the specific hypotheses and choosing the statistical tools to prove them, a so called CRM problem was defined for this top-down approach. From a business point of view it is important to find out key factors driving satisfaction of customers to put more effort in improving them and in the end increase company value by binding customers permanently. For an automatic customer satisfaction prediction those key factors should help for selecting the right features more easily. The procedure followed has been described in section 2. The upcoming paragraphs give a more detailed insight into the different hypotheses tested, how they were checked and which statistical tools were applied. The analysis task can be split into two types of data, namely categorical- and continuous data.

3.2.2.1 Analysis of categorical data

This first analysis done in the implementation part of this thesis is based on the relationship between customer- satisfaction and loyalty. Thus it took a look on the long-term relationship of a customer with Tractive. Only the service status "Active" which indicates that a customer is paying on a recurring basis and the service status "Terminated" which indicates a manual service cancellation by a customer were considered. This implies the fact, that this task relies on binary categorical data. The analysis goal was then stated as: Does Live Tracking success influence service termination behavior?

1. Hypotheses

H0 There is no significant difference in service termination behavior between customers who belong to the group of bad live-tracking users and customers who belong to the group of good live-tracking users.

H1 There is a significant difference between these two groups.

2. Analysis objects

   - Live Tracking Server commands related to a specific tracker
   - Service status of customers

3. Analysis problem

   - Select representative sample and split it up into two groups to compare, namely a treatment and control group. It is important to exclude any other influencing factors as best as possible and therefore define common base data shared among both groups.
   - Choose a suitable sample size

4. Analysis solution: Statistical test which should check whether the Null-hypothesis can be rejected and thus statistical significant difference can be shown. The analysis was done on 17.04.2017.

   - Group selection: The first group contains random customers suffering from bad Live Tracking which was defined by an overall success rate below 70%. In contrast, the control group consists of customers with a success rate greater or equal than 80%. These thresholds were set according to experience in customer support where complaints from users regarding Live Tracking usually show success rates around this threshold. As common base data only users from Germany who own at least one Tractive GPS device and do not have a premium service plan to be payed on a monthly basis which was created before 01.10.2016 and is at least valid until 01.10.2016 were considered.
   - Choose a suitable sample size: The sample size for the two groups was estimated based on findings of [3]. Regarding statistical significance a widely used α-error of 5% was used while a β-error of 10% should clearly reduce the probability of false-negatives meaning non-rejection of H0 although it could have been rejected. Since a statistical significant result does not necessarily say something about expressiveness of the computed result, a minimal rel-

	Service status = ACTIVE	**Service status = TERMINATED**
Good Live Tracking	54	8
Bad Live Tracking	43	19

Table 3.3: 2x2 table showing influence of Live Tracking on service status

evance level had to be set for choosing an appropriate sample size. A 20% decrease of active subscriptions due to bad Live Tracking was considered as relevant. Picking the right number from the proposed sample size table yielded a number of 79 per group which in total is 158.

- Querying data: At the time where this particular analysis was conducted no analytical view on the server command metrics was available and therefore the procedure was to calculate those metrics for the devices under consideration on the fly. Therefore a Javascript program was implemented to fetch randomly the necessary number of subscriptions which either had status "active" or "terminated" with the defined common data. Based on the device reference, server commands of type "Live Tracking" within the given time period were filtered and the success rate averaged. Based on the output data a post-processing step provided a simple 2x2 table which is illustrated in 3.3.

- Use statistical test to verify or falsify H0: There are different statistical tests for binary data whereby Fisher's exact test is proposed as most suitable for a rather small sample size as it was in this case [18]. The test works based on a 2x2 table as input and calculates a 95% confidence interval indicating where the true odd ratio lies. The odds ratio is a often used value when analyzing binary data and states the relative probability than an event occurs against that it does not occur [1]. The Null-hypothesis in Fisher's exact test can be rejected if the confidence interval does not include the odds-ratio 1. Then it is safe to claim that there is a difference between the two sample groups under consideration. The open source statistic program package R was used to execute the Fisher exact test for the 2x2 table and yielded a confidence interval of $[1.105933, 8.604534]$ which indeed allows rejecting H0. Based on this statistical experiment it is thus valid to say that there is a statistical significant and based on the minimum of 20% difference in terminations between treatment and control group also a relevant result.

- Interpreting the result: This data analysis confirms the expected importance of the Live Tracking function for Tractive's customers since terminating an active service is equal to ending the relationship with the company and can therefore be considered as a last resort in case of dissatisfaction. As a

result it can be stated that Live Tracking success rate qualifies as potentially promising feature for Customer Satisfaction.

3.2.2.2 Analysis of continuous numeric data

The following analysis tasks are all based on continuous data and therefore summarized in this section. Although the procedure remained the same as explained in the previous section, all sub-tasks share some common ground and therefore it was decided to make the analysis more modular. As a result of this decision, a NodeJS application with a MongoDB connection driver was created. Querying tasks were extracted into reusable Javascript functions. Before elaborating on the particular sub-tasks, the common parts are described following:

- Choose a suitable sample size: For an appropriate choice the width of the confidence interval, where the true correlation coefficient of the population lies, was considered. The author decided on 0.1 as its width for the experiments to ensure a high precision. Since the app usage was not considered as a direct measurement of a customer satisfaction value, a sample correlation coefficient of 0.4 was determined as relevant. Based on these parameters and the research of [13] a sample size of 1086 could be derived for the correlation analyses.

- Select sample users: The selection was done by randomly skipping users and ended when the maximum sample size of 1086 was reached. For every sub-task only users with exactly one active service were considered to make comparison with device related data easier. The output of this selection stage contained user- and device ids.

- Query data to compare: Data from the previous stage was used as input parameter for the according function to fetch the desired data rows for correlation analysis within a selected time period. The implemented queries heavily use the aggregation framework of MongoDB to provide the desired data in an aggregated form grouped by user id.

- Match data based on user id: The data collected was stored in JSON (Javascript Object Notation) arrays in-memory. A matching function finally generates the input vectors for the subsequent statistical calculations, by matching the nested JSON objects via the user id and writing data to one CSV file each.

- Use correlation analysis to reason about linear relationship: After finishing the querying part, the data was ready for correlation analysis. Therefore the sta-

Date of analysis	Commands	Attribute	Correlation coefficient
04.06.2017	LT	Command success rate	r = 0.04222696
04.06.2017	LT, Buzzer, LED, Position request	Command success rate	r = 0.1384539
10.06.2017	LT	Command delay	r = -0.01711778
10.06.2017	LT	Command cancellation rate	r = -0.0420426

Table 3.4: Results of correlation analysis between server command metrics and app usage

tistical open source program package R was used to calculate a Bravai-Pearson correlation coefficient and thus get a first impression on the relationship between the considered data rows.

Following correlation analyses were done with the aim to find influencing factors for Customer Satisfaction. The ordered list below shows the analysis goal implicitly formulating the hypotheses along with the data collections used for each analysis and the results retrieved.

1. Do server command metrics cause an increased respectively diminished app usage?

   a) Analysis objects

      - Server command metrics collection
      - App Events collection containing app startup and app foreground events

   b) Results: See table 3.4

2. Does bad signal strength of the device cause less activity?

   a) Analysis objects

      - Device hardware reports collection containing the reported RSSI (Received Signal Strength Indication) value at a particular point in time. Since RSSI values are not expressive enough when the device is indoor, only those reports with GPS used as sensor were considered during analysis to reduce bias.
      - Device Id reports: The number of days the device was in use which was retrieved through grouping by day.

   b) Results:

3. Does battery lifetime of the device influence user's activity level?

   a) Analysis objects

      - Each device hardware report contains information about the current battery level. Moreover there are some defined events carrying information about whether the battery is currently charging, full, low or critical. Unfortunately there is no direct information about an average battery lifetime available in the data. Therefore some pe-calculation had to be done to get this average value. Due to the reason that the devices do not always deliver the battery events reliably only an approximation is possible.

      - As for the signal strength, again the number of days the device was in use was used as target vector.

   b) Results:

3.2.3 Evaluation of approach and derived decisions

The very first analysis task for a randomly selected sample of subscriptions yielded a statistical significant difference between users tending to keep behave loyal in terms of keeping the subscription active and users tending to cancel their subscription with regard to the live tracking success rate. The asymptotic Chi-Square test as well as Fisher's exact test supposed to reject the Null hypothesis which confirmed the initial assumption that the live tracking success rate is a critical business factor for Tractive. However, due to complexity regarding the aggregation queries and hypothesis test possbility, this statistical test only considered two groups of subscription states, namely ACTIVE and TERMINATED. Therefore payment failures and as a result expired- as well as paused services were not considered. In the following hypotheses driven tests several correlation coefficients between a given behavior metric and an assumed Customer Satisfaction driver, like the app foreground resp. app startup events or the aggregated number of usage days of the GPS device, were calculated. For none of them the author of the thesis could derive any further influential factor for Customer Satisfaction. One of the main factors identified is the single metric used to explain Customer Satisfaction. Due to the results, the author realized that Customer Satisfaction is a much more complex and involved construct and it turned out to that the app usage or usage days of the GPS device are not sufficient to make a promising statement with regard to Customer Satisfaction. Although [11] came to the conclusion, after their survey results analysis, that mobile usage of customers in nowadays smartphone domi-

nated world has an influence on satisfaction, their proven hypotheses contained mobile engagement motiviation as primary factor. This engagement can be characterized as a three dimensional model consisting of the functionality of a mobile app driving a users efficiency to accomplish tasks, the ease of use and entertainment factor and a social component indicating whether it is possible to connect with friends via the app [23]. The results of [11] furthermore show that mobile engagement and perceived value can create a basic satisfaction level for a user which in turn leads to more engagement motivation and as a result increases satisfaction. This mobile engagement- and satisfaction model is not well represented by the single dimension of app opening events. As a result the data analysis tasks explained in the previous part of this thesis underperformed. The difficulty in finding a representative metric, which provides a better chance to get more accurate results, from the collected data at Tractive led to the following decision. In conjunction with the company the author decided to design a customer survey sent to users aksing them a few questions to find out how satisfied they are with the product. The goal was to incorporate the gathered knowledge back into the software system, extract potentially satisfaction driving features as mentioned in section 3.1 for those customers and learn from this data. Identified relevant data sources should be reused and further extended in this part of the thesis but the approach should be a different one. The upcoming section 3.3 will shed more light onto the details of the attempted approach.

3.3 Data-driven approach leveraging explicit feedback as target variable on Customer Satisfaction

Instead of using identified relevant features from the first approach to make predictions about a satisfaction level of customers, this approach should be fed with more features and find its way to the promising ones automatically. The goal was to work on a software framework which analyses data and selects the influential factors for Customer Satisfaction automatically while throwing away garbage data and redundant features. Furthermore, the lack of metrics in the data representing Customer Satisfaction identified as major issue during the first part of the thesis should be tackled by getting explicit feedback from customers on how satisfied they are. The expectation from this customer survey is that it provides a much more reliable quantitative measurement of the satisfaction level of a customer than any other kind of data Tractive collects in its databse system. As described in more detail in the satisfaction theory part of chapter 2 understanding why a customer behaves in a certain way can be quite subjective, involve psychological factors and vary among different types of customers. Therefore it has to be admitted that this can hardly be extracted from some objectively collected data and the actual source of truth is to ask people themselves. With enough survey

answers the framework should first analyse all observations retrieved by leveraging descriptive statistics and evaluate chances to predict satisfaction for arbitrary customers, who have not filled in the survey. With specific features selected machine learning algorithms can then be trained and evaluated on the data to find a model which is suited to do automated predictions.

3.3.1 Implementation of a survey to gather explicit feedback from customers

Before any further data analysis task was started the initial task was to create a customer survey to collect usable information which gives a reliable insight into how a customer rates his experience with the product so far. A requirement was to keep the effort for the customer to fill in the survey low which in first place means to require little time to fill in the duty part of the survey. Little time consumption correlates positively with response rate. According to [19] customer surveys are often too complex and as a result overwhelming for the user even though the company cannot derive better decision from it. With regard to customer satisfaction and loyalty [19] claim that the most essential number is the willingness of a customer to recommend the product to a friend or colleague. This metric should be on a scale of 0-10 to allow a computation of the NPS (Net-Promoter-Score), a widely used score to constitute growth of a company. In collaboration with responsible people at Tractive it was decided to limit the number of questions to the following two which were used later on to establish the target attribute in the prediction framework.

1. How satisfied are you with Tractive GPS? (Scale 1-5)

2. How likely are you to recommend Tractive to a friend or colleague? (Scale 0-10)

Next to the specifications of questions, the survey should be appealing, personalized and easy to use. The final landing page of the customer survey is shown in figure 3.3.

The goal of the company was to get customer feedback early on to be able to respond to unsatisfied customers quickly and in best case before they complain about a problem they have at customer support. As a result, the author of the thesis could arrange with the responsible people of Tractive to following procedure. When a new user converts successfully to a paying customer, he or she has to activate the GPS device before first usage. This enables the device to be actively used, sending positions and showing up on the map in the apps. Directly after activating the device, the backend service schedules an email in two weeks asking the customer to fill in the survey. To reach as many users as possible the survey was translated into the five major languages supported at Tractive,

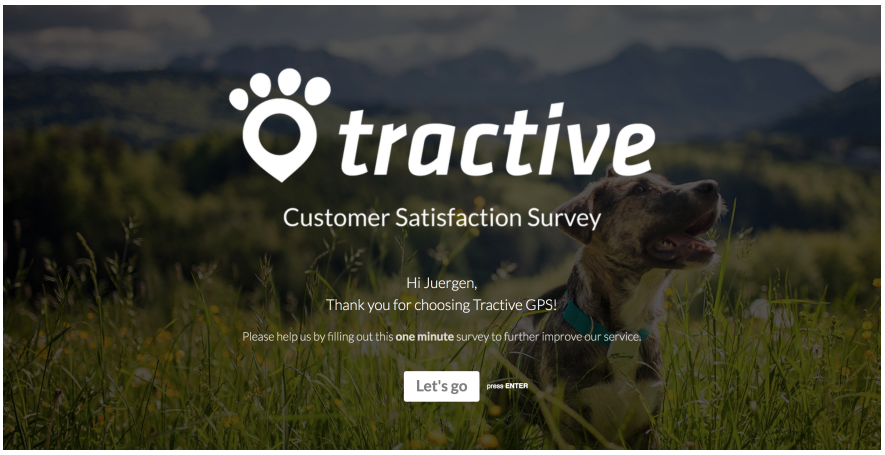


Figure 3.3: Customer Survey - Landing page

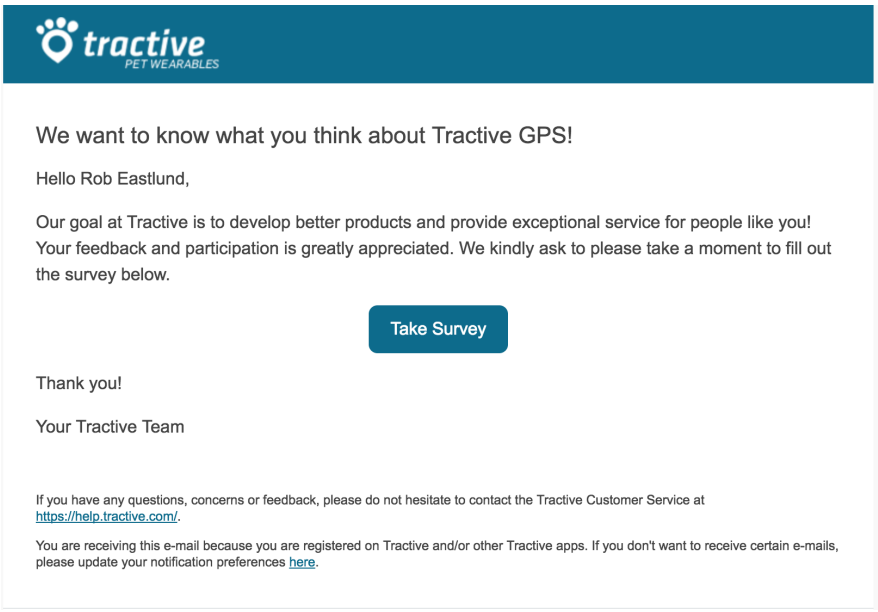


Figure 3.4: Customer Survey - Email

namely German, English, French, Italian and Spanish. Depending on the users stored demographical settings the correct language version is triggered. An example of how the email looks like is shown in figure 3.4.

When the user clicks on the ”Take Survey” button in the email, a browser window will open with the customer survey landing page. Behind the scenes In the URL (Uniform Resource Locator), the language version of the survey as well as the name of the customer is passed. For better personalization of the voluntary additional questions the name of the pet associated with the GPS device is sent as URL parameter. A necessary information is the unique id of the device since it allows to query all of the identified hardware- and position related data identified in section 3.1. For designing the customer survey the external tool Typeform was used. The following section describes briefly how the data gets extracted and integrated back into the internal database to

Attribute	Datatype	Description
_id	ObjectId	Unique identifier of document
survey_id	String	Typeform identifier for the particular language version
submit_date	DateTime	Date and time when customer submitted survey
user_id	ObjectId	Reference to the user
tracker_id	ObjectId	Reference to the tracker
rating	Integer	Overall satisfaction (scale: 1-5)
recommendation_score	Integer	Recommendation potential (scale: 0-10)

Table 3.5: Structure of a survey response represented in the company database

be processable. Furthermore a few statistical details will be pointed out to get an impression with regard to the pending satisfaction analysis and prediction task.

3.3.2 Results and interpretation of survey results

To fetch the survey results from Typeform, a nightly job was implemented which uses the provided API of Typeform and gets the results from the previous day. The essential metrics from the two duty questions are extracted and along with the user- , pet- and tracker data stored in a new collection in the main database of Tractive. A typical structure of a document in this collection is illustrated in table 3.5.

The rating and recommendation score are the two interesting numbers when it comes to predicting satisfaction for an arbitrary user afterwards. The customer survey was deployed to the productive environment on 03.07.2017 which yielded the first survey result submitted on 17.07.2017. As of 28.10.2017 17:42 UTC+2 following statistics regarding the survey could be extracted.

- Number of customer survey emails sent: 15695
- Number of customers who filled in the survey: 2182
- Percentage of users who filled in the survey: 13.9%

As these numbers show, the response rate of the customer survey fortunately is quite good. The collected amount of customer survey data so far should be sufficient for finding potential patterns in the data. In order to get a better understanding some descriptive statistic values for the satisfaction rating and the value indicating willigness of a customer to recommend Tractive were calculated. The results are shown in table 3.6.

Survey metric	Min.	1. Quartile	Median	Mean	3. Quartile	Max.	Variance	Standard dev.
Satisfaction	1	3	4	3.688	5	5	1.327	1.152
Recommendation	0	6	8	7.137	9	10	7.574	2.752

Table 3.6: Statistical summary - Overall satisfaction and recommendation score

Based on the results it can be followed that the average customer rates his or her satisfaction with the Tractive GPS product as mostly satisfied represented by a value close to 4. Similar is the recommendation likeliness value where the average lies between 7 and 8. Since the first quartile with a value of 6 is rather high, it can be followed that 75% of the customers are more likely to recommend Tractive to a friend or colleague. To close this statistical analysis of survey responses it can be stated that the majority of customers tend to be satisfied which had to be considered accordingly in the prediction framework outlined in more detail in the following section 3.3.3.

3.3.3 Software architecture of prediction framework

Chapter 4

Evaluation

Chapter 5

Conclusion

Bibliography

[1] J Martin Bland and Douglas G Altman. The odds ratio. *Bmj*, 320(7247):1468, 2000.

[2] John T Bowen and Shiang-Lih Chen. The relationship between customer loyalty and customer satisfaction. *International journal of contemporary hospitality management*, 13(5):213–217, 2001.

[3] Michael J Campbell, Steven A Julious, and Douglas G Altman. Estimating sample sizes for binary, ordered categorical, and continuous outcomes in two group comparisons. *BMJ: British Medical Journal*, 311(7013):1145, 1995.

[4] Injazz J Chen and Karen Popovich. Understanding customer relationship management (crm) people, process and technology. *Business process management journal*, 9(5):672–688, 2003.

[5] William J Doll, Weidong Xia, and Gholamreza Torkzadeh. A confirmatory factor analysis of the end-user computing satisfaction instrument. *MIS quarterly*, pages 453–461, 1994.

[6] F Robert Dwyer, Paul H Schurr, and Sejo Oh. Developing buyer-seller relationships. *The Journal of marketing*, pages 11–27, 1987.

[7] Chris Ellegaard and Thomas Ritter. Customer attraction and its purchasing potential. In *22nd IMP Conference, Milan*, 2006.

[8] John Gantz and David Reinsel. The digital universe in 2020: Big data, bigger digital shadows, and biggest growth in the far east. *IDC iView: IDC Analyze the future*, 2007(2012):1–16, 2012.

[9] Rahim Hussain, Amjad Al Nasser, and Yomna K Hussain. Service quality and customer satisfaction of a uae-based airline: An empirical investigation. *Journal of Air Transport Management*, 42:167–175, 2015.

[10] Asghar Afshar Jahanshani, Gashti Mohammad Ali Hajizadeh, Seyed Abbas Mirdhamadi, Khaled Nawaser, and Sayed Mohammad Sadeq Khaksar. Study the effects of customer service and product quality on customer satisfaction and loyalty. 2014.

[11] Young Hoon Kim, Dan J Kim, and Kathy Wachter. A study of mobile user engagement (moen): Engagement motivations, perceived value, satisfaction, and continued engagement intention. *Decision Support Systems*, 56:361–370, 2013.

[12] Bart Larivière and Dirk Van den Poel. Predicting customer retention and profitability by using random forests and regression forests techniques. *Expert Systems with Applications*, 29(2):472–484, 2005.

[13] Murray Moinester and Ruth Gottfried. Sample size estimation for correlations with pre-specified confidence interval. *The Quantitative Methods for Psychology*, 10(2):124–130, 2014.

[14] Michael C Mozer, Richard Wolniewicz, David B Grimes, Eric Johnson, and Howard Kaushansky. Predicting subscriber dissatisfaction and improving retention in the wireless telecommunications industry. *IEEE Transactions on neural networks*, 11(3):690–696, 2000.

[15] Friedemann W Nerdinger, Christina Neumann, and Susanne Curth. Kundenzufriedenheit und kundenbindung. In *Wirtschaftspsychologie*, pages 119–137. Springer, 2015.

[16] Eric WT Ngai, Li Xiu, and Dorothy CK Chau. Application of data mining techniques in customer relationship management: A literature review and classification. *Expert systems with applications*, 36(2):2592–2602, 2009.

[17] Bernd Knobloch Peter Neckel. *Customer Relationship Analytics*, volume 2. dpunkt.verlag, 2015.

[18] Michel Raymond and Francois Rousset. An exact test for population differentiation. *Evolution*, 49(6):1280–1283, 1995.

[19] Frederick F Reichheld. The one number you need to grow. *Harvard business review*, 81(12):46–55, 2003.

[20] Marie-Christine Roy, Lyne Bouchard, et al. *Developing and Evaluating Methods for User Satisfaction Measurement*. Faculté des Sciences de l'administration, Université Laval, 1998.

[21] Chris Rygielski, Jyun-Cheng Wang, and David C Yen. Data mining techniques for customer relationship management. *Technology in society*, 24(4):483–502, 2002.

[22] Ronald S Swift. *Accelerating customer relationships: Using CRM and relationship technologies*. Prentice Hall Professional, 2001.

[23] Kaan Varnali and Ayşegül Toker. Mobile marketing research: The-state-of-the-art. *International Journal of Information Management*, 30(2):144 – 151, 2010.

[24] Li Xiao and Subhasish Dasgupta. Measurement of user satisfaction with web-based information systems: an empirical study. *AMCIS 2002 Proceedings*, page 159, 2002.