Twitter as a potential goldmine: A data-driven product innovation approach in the wearable device domain based on machine learning

Author: Eric Wan
University of Twente
P.O. Box 217, 7500AE Enschede
The Netherlands

ABSTRACT

The recognition, assessment, and integration of customer feedback is an essential component for businesses to design their products and services appropriately. Although commonly applied methods provide the necessary information, they often require extensive manual labor, lack automation as well as scalability, and hence, are not suitable for ongoing use. As an alternative for identifying customer needs, this paper investigates the reliability of the BERT model, a machine learning algorithm that uses a trained neural network to classify a given dataset. Prior research from Kuehl et al. (2016) already adopted a similar approach within the e-mobility domain. This paper expands on the efforts of Kuehl et al. by applying the approach for the wearable device domain. In total, 10,000 tweets – short messages retrieved from the social media platform Twitter – were manually assessed, from which 130 'need tweets' were identified. The needs were categorized into eight different groups and provided relevant product development ideas and inspirations. Furthermore, an analysis of the BERT model showed that the prediction and classification of the tweets are inconsistent and unreliable. Further research in this domain with adjustments in the dataset and expanding into other domains are suggested.

Graduation Committee members:
First supervisor: Dr. Dorian E. Proksch
Second supervisor: Dr. Kjartan Sigurdsson

Keywords
Data mining, social media analysis, customer need elicitation, data-driven product development, needmining, wearable device domain

1. INTRODUCTION

Today's world is characterized by fast-changing trends and constantly improving technologies that are influenced and driven by customers' preferences and desires. As a consequence, identifying and understanding the various needs of customers has become increasingly important for businesses to generate higher profits (Singh, 2006), to become more competitive (Cai, 2009), and thus, to succeed within a market (Limehouse, 1999). The significance of changing customer perspectives can be illustrated by the highly competitive smartphone industry (Cecere, 2015). With the aim of becoming the leader within the smartphone industry and thereby realizing an advantageous competitive position as well as financial benefits, the companies that operate in that market are constantly contesting against each other with their products and services. Comparing older generation smartphones with newer ones, changes in hardware components and software features are easily noticeable. For instance, the size of the display has increased over the last decades. Or another feature that was once popular and then gradually disappeared in current smartphone designs is the headphone jack, an electrical connector used for analog audio input. One influencing factor that has directed this path of change is alternating customer preferences throughout the years (Ha, 2016). And to cope with the ever-changing nature of customer needs, companies engage in extensive market research. They continuously adapt and improve as a response to newly identified customer needs, sometimes even leading to the invention of a new technology, product, or service. Mainly, conventional approaches are applied in this regard, e.g., interviews, surveys, or focus groups analysis. And even though these methods provide relevant insights and new perspectives from the customers' point of view, the main disadvantages are its cost (Fisher et al., 2014) and time-consumption (Hauser and Griffin, 1993) in applying them. They lack automation and require extensive manual labor in order to function, resulting in low scalability and making it inconvenient to use them on a continuous basis.

Hence, in search of more convenient and viable alternatives, researchers have focused intensely on data analysis of social media content from platforms such as Twitter to elicit customer needs more accurately and efficiently (Kuehl et al., 2020). Especially, as the number of social media users has risen to roughly 3.78 billion users (Tankovska, 2021), a sufficient amount of data is available to be analyzed. Evidently, a large proportion of the content that is posted online has no managerial or strategical implications for a company. But considering that a substantial share of individuals publishes their opinions about a specific product or service (Misopoulos et al., 2014), both positively and negatively, openly on social media platforms, can turn the seemingly meaningless data into a potential goldmine filled with countless ideas and inspirations for product development and innovation (Edvardsson et al., 2011).

Due to yet unresolved obstacles – namely the correct filtering process of the social media content, which is followed by a meaningful deduction – only market-specific attempts were made, for instance, in the e-mobility domain (Kuehl et al., 2016), raising the question of whether this data analysis approach would work in another sector outside the e-mobility domain. Therefore, in this paper, the aim is to partly fill this domain-gap and address the mentioned obstacles by using data from the Twitter platform and related to the Fitbit watch, which operates within the wearable device domain. The filtering process will be done through a supervised machine learning algorithm. By analyzing the data, conclusions will be drawn to evaluate how practical and useful this approach is.

This paper intends to answer the following two research questions:

(i) What are the most prominent customer needs, and to what extent can they be generalized for other domains?

(ii) How reliable is the detection of customer needs by using the proposed machine learning algorithm?

The remainder of this paper is structured as follows: At first, a summary of the used literature will be given, containing the most important work related to this topic. After that, a section explaining the methodology of this paper will be provided. In the next section, the results of the machine learning algorithm will be presented, followed by a separate conclusion and a discussion part.

2. LITERATURE REVIEW

To provide a basic foundation, this section will clarify the concept of data-driven product innovation and how recent approaches are used to identify customer needs. Furthermore, the relevance of customer needs in the context of successful business development will be discussed. Social media data analysis will be elaborated – especially a short introduction to the work of Kuehl et al. (2016), related to the concept of *Needmining*, will be provided. Lastly, a brief definition of customer needs will be given, and the selection of Fitbit as representer of the wearable device sector will be explained.

2.1 Data-driven product innovation

Technological advancements in the information and telecommunication sector, especially social media networks, led to increased data generation from end-users and, thus, partly co-created the era of big data (Elmer et al., 2015). Big data is a "collection of massive and complex data sets and data volume that include huge quantities of data […]" (Anuradha, 2015). This boost in data availability, from which a remarkable amount is related to product and service reviews from customers, created an incentive for companies and scholars to find ways to exploit it (Trabucchi et al., 2019; Del Vecchio et al., 2018; Saarijarvi et al., 2014), converting the unstructured data into critical input (Bharadwaj et al., 2015), thereby identifying meaningful trends and needs. As a consequence of these newly introduced data-driven approaches, traditional innovation approaches such as comprehensive customer surveys, interviews with lead customers, and focus group analysis, are being challenged (Geum et al., 2016). However, both the data-driven approach and the conventional methods emphasize the importance of customer needs identification and view customer satisfaction as one of the essential pillars of a successful company (Hoyer et al., 2001). Consequently, the end-users, or rather the customers of a company, have become a driving factor of product innovation (Troilo et al., 2017). The term *innovation* has various forms of definitions and usages. Among those, the most suitable for this research paper is referring to the one of Rogers (1998), stating that "innovation is the process of introducing new ideas to the firm which result in increased firm performance." In sum and defined for the purpose of this research, the concept of data-driven product innovation describes the process of analyzing the selected social media data with the use of machine learning algorithms in a way that hidden and obscure customer needs that can potentially lead to an increased firm performance are identified.

2.2 Relevance of customer needs analysis for the overall performance of companies.

Even prior to the development of more sophisticated machine learning algorithms and the proliferation of social media along with the occurrence of big data, companies already put efforts into customer satisfaction (Hoyer et al., 2001), showing that customer-centered perspectives were and still are worthwhile using. Evidently, throughout the years various, different models were designed to actively include the customer into the business processes, in particular product and service development, in order to meet the expectations of customers and satisfy them for the purpose of loyalty and retention (Singh, 2006). To name some examples of customer-centered approaches that have become notably popular over the last decades in all sorts of industries and geographical locations, there is the *Quality Function Development (QFD)*(Chan et al., 2002) – a method for defining customer requirements and translating them into measurable design targets in order to adjust operational specifications and processes – or the Kano model (Kano et al., 1984) – a method used for grouping and prioritizing product features based on the extent to which they are likely to satisfy customers.

The researchers Urban and Hauser (2004) emphasize the importance of recognizing newly developed customer needs early on to benefit from being first. They recommend keeping the customers involved as an ongoing process, even when it is a rather huge investment, including interviews with lead customers, maintaining and monitoring user groups, etc. These disadvantages are outweighed by the benefits it provides. According to Hoyer et al. (2001), achieving a satisfied customer base is crucial for any successful business as it leads to repeat purchases, brand loyalty, and positive word of mouth. Especially, the last aspect is further strengthened by the ease of spreading information via social media networks, resulting in electronic word-of-mouth (eWOM), recognized as one of the most influential resources of information transmission (Jalilvand et al., 2011). Standard measures of a company's performance, such as financial stability (Singh, 2006) and competitiveness (Cai, 2009), are also enhanced by implementing a customer-centered approach. A quote that illustrates the relevance of customers in the context of business-related success is provided by Zairi (2000), stating that "Customers are the purpose of what we do and rather than them depending on us, we very much depend on them […]". In accordance with this positive tendency towards customer-centered approaches, this paper attempts to identify customer needs that may have an impact on the business of the selected company.

2.3 Social media data analysis

As mentioned in previous sub-sections, the rapid increase in data availability that was generated from social media platforms provided companies with the opportunity to learn about individual customers as well as broader networks of customers and a new possibility to retrieve important insights to support their business decisions (Moe et al., 2017; Urbinati et al., 2019). Traditional methods are still in use, but one striking element of big data analysis is its speed and scalability. In case of a successfully implemented big data analysis, customer needs and sentiment towards a product or service are recognized earlier (Davenport et al., 2012), giving the company more time to adapt to these findings and exploit the benefit of being first. Apart from the duration of the process, big data also provides a more cost-efficient alternative for analysis compared to the traditional methods as the data is oftentimes publicly available or at least cheaper to acquire, enabling companies to implement big data analysis continuously.

On the contrary, due to the characteristics of big data, including massive unstructured sample size and very high dimensionality, it is inherently coupled with challenges. The main challenge in using big data analyses is the lack of understanding of the given data, making traditional small datasets easier to comprehend and analyze (Birch-Jensen, 2020). Highly computational and statistical models are required to structure and make sense of the data under investigation (Fan et al., 2014). Besides, since the data is oftentimes collected from multiple sources at different time points, human interaction is necessary to develop more adaptive algorithms and procedures.

Since the field of big data analysis, in particular social media data analysis, is a rather recent trend, no general approach has been discovered yet. Therefore, among various different approaches, this research paper is in conformity with the work of Kuehl et al. (2016) and their contribution to the broader concept of *Needmining*. Kuehl et al. (2016) were the first to introduce a method for analyzing user-generated content from Twitter by utilizing machine learning techniques, providing a wide application range. "Machine learning is an evolving branch of computational algorithms that are designed to emulate human intelligence by learning from the surrounding environment" (El et al., 2015); an environment which is created by big data collection. The *needmining* approach allows for a simple customer need separation, meaning that this algorithm is trained to distinguish if a tweet – a short message transmitted via the microblogging platform Twitter – contains a customer need or not. In a later publication, Kuehl et al. (2019) extended this simple separation by adding a classification feature, making it possible to classify between more detailed and specific needs. Both versions of the needmining approach are supervised machine learning methods, in which the training of the data set is handled by a human. Up until now, only market-specific data analysis attempts were made, e.g., one within the e-mobility domain by Kuehl et al. or another one within the airline industry conducted by Misopoulus (2014). As mentioned before, this paper adds to the current research by analyzing the feasibility of this approach within the wearable device sector.

2.4 Definition of customer needs

Due to psychological phenomena and different situational circumstances, every human being has very unique needs and expectations as a customer of a product (Scharnbacher et al., 2010). Preferences for usage, design, price, or specific (technical) features of a product or service, can thus differ remarkably from customer to customer. Thus, under the premise that customer needs are very individualistic, and the machine learning algorithm needs a well-trained training set in order to properly function, a clear distinction has to be drawn in terms of what is considered to be a customer need and what is not. According to Kotler and Armstrong (2001), customer requirements can be separated into three different categories – namely, *needs*, *wants*, and *demands*. *Needs* describe basic human requirements of survival, such as food and shelter. *Wants* are specific desires for things or products outside of the realm of survival, and *demands* are *wants* that are backed up with the ability to acquire them. For reasons of simplicity, customer needs will be regarded as all three categories combined as the differentiation between the categories does not provide a significant benefit. In addition, and in conformity with Kuehl et al. (2016), "any information, regardless of the level of granularity, is valuable information […]". In section 3, the methodology part, a more elaborated and practical definition will be provided as to what is regarded as a need. The distinction between a need and no need is essential because it lays the foundation for the machine learning algorithm.

2.5 Wearable devices as evaluation domain

Among newly technological trends during the recent decades, wearable devices have become popular, and their usage is increasingly extended (Jeong et al., 2017). According to Vailshery (2021), about 722 million wearables devices were connected worldwide, with an expected forecast to achieve the one billion milestone in 2022. In particular, fitness trackers and smartwatches are seen as commonly used accessories, providing more functionalities and practicalities than traditional watches (Chuah et al., 2016). In addition to telling the time and setting alarms and timers, these wearable devices usually include a heart rate monitor with other health-related measures to detect abnormalities and diseases, different programs to effectively track and record your training sessions, various interface designs to choose from, other customization settings to change the layout, an interface to the customer's preference and allow the usage of apps from a linked smartphone device. Next to the technological specifications and features, an important part of successful wearable devices is their aesthetics (Hsiao et al., 2018). Because the devices are physically worn and therefore observable by others, the appearance, style, size, and other design factors are important to consider. Therefore, the market of wearable devices, in this case, the fitness-tracker and smartwatch sector, offers a multitude of targets for customer feedback acquisition as both aesthetic and technical features can be rated and criticized. Another relevant aspect of the wearable device industry is its user-base's tendency to share their opinion about certain products publicly on social media platforms, providing a sufficient data set for customer needs identification.

For the purpose of this research, Fitbit, a company founded in 2007 and acquired by Google in November 2019 with its main focus on the creation of fitness-trackers and smartwatches, is chosen to be examined. It fulfills the aforementioned conditions of being in the wearable device industry, and because of its user-base's active social media behavior, a sufficient amount of data can be retrieved from the Twitter database. In addition, Fitbit offers a variety of different smartwatches and fitness trackers that all have overlapping functionalities, making it possible to generalize product-specific feedback and apply it to different products. Furthermore, even though the products of Fitbit are physical, its main features are software-based, meaning that improvements based on customer feedback can easily be implemented by software updates. Lastly, as Fitbit is in direct and high competition with various companies, including large firms such as Apple, Samsung, Garmin, etc., the results of this paper may have relevance in terms of business-related decisions and implications.

3. METHODOLOGY

This section elaborates the research steps related to the collection and preparation of the Twitter data for the subsequent analysis and application of the machine learning algorithm. Twitter is selected for the aim of this research due to several reasons. Firstly, it is a social media platform, enabling users of this platform to share their thoughts, feelings, and moments with the public through so-called tweets – short messages containing photos, GIFs, videos, links, hashtags, and up to 280 text characters. The Twitter database has been checked to ensure that a sufficient amount of data related to the Fitbit company and its products is available. In addition to that, the data from Twitter is publicly available to everyone, and thus, no legal restrictions limit the extraction process. Lastly, as Twitter content has been in frequent use, statistical software programs such as Python provide readily available routines to facilitate the transfer of Twitter data. The preparation process can be divided into four steps: Data retrieval, filtering, labeling, and pre-processing.

3.1 Data retrieval

Twitter provides developers and researchers with a streaming API, allowing them to fetch real-time data and also past tweets. API is an acronym for 'application programming interface,' and the underlying process is ubiquitous nowadays. In essence, the streaming API is a tool that facilitates the extraction of information from the Twitter database and makes it possible to undertake initial filtering. Through a selection process based on pre-defined keywords, only relevant data will be fetched via the streaming API. An appropriate time period is determined. For instance, one could choose the period right after the release of a new software update for the Fitbit watch to analyze the satisfaction level of customers and to identify possible weak points of the new update.

The keywords that are being used for this research are as follows: "Fitbit", "fitbit" and "FitBit". The time period under investigation is from 01/04/2020 until 01/04/2021, so a one-year period. Figure 1 displays how many tweets were involved in each data preparation step, including data retrieval, filtering, and labeling.

3.2 Data filtering

Through the use of the streaming API tool, the data from the Twitter database has already been filtered in a way that only English tweets were shown, and retweets – republication or forwarding of an existent tweet – are excluded, reducing the number of duplicates. This filtering step happens before the actual retrieval of the data. Subsequently, the retrieved data will be further filtered in order to enhance the structure and reduce the amount of unrelated and unnecessary tweets.

As a first step, the occurrence of duplicates is checked again. The filtering step before the data retrieval partly removed duplicates, but in this step, the removal will be repeated to ensure that all remaining duplicates will be removed because they do not provide further need information. One tweet with the same information is sufficient. Secondly, a list with stopwords – words that, if contained in a tweet, have a high likelihood of not containing a need – was created. In this case, the list includes words that have a notion of being either spam or promotion, e.g., "deal", "sale, or "win". In addition to stopwords, tweets with promoting URLs are also excluded because they are unlikely to provide need information (Kuehl et al., 2016). Lastly, and in accordance with empirical research, tweets containing less than 25 text characters will be excluded and ignored as they, on average, do not provide insights in terms of customer needs. Eventually, the data filtering rules will be adjusted whilst working with the actual data. Unexpected results may make it difficult, e.g., a too-small dataset or, on the other extreme, too much information. Nonetheless, it is obvious that the more restrictions are being used, the higher is the risk of losing valuable customer need information will be.

3.3 Data labeling

The remaining tweets will represent the complete dataset, which will be used for the machine learning algorithm. From this resulting dataset, 10,000 tweets will be randomly selected and analyzed under the criterium of whether they contain a customer need. In case they do contain a customer need, a '1' will be used to mark these tweets; in case it does not contain a customer need, a '0' will be assigned to the tweet. It is intended that only one individual, the researcher, will carry out the data labeling step. Due to the limited number of coders, tweets that could be either containing a need or not, depending on how the coder argues, may be difficult to label.

time	tweet	isneed
18.02.21 00:20	@fitbit A lot of the mindfulness items that are available through the app are audio. Do you have stuff for hard of hearing or deaf people? #accessiblemindfulness	1
16.08.20 07:33	@fitbit @FitbitSupport Hi, When are you planning to support Arabic on fitbit devices and wearables?	1
03.04.20 00:16	Whoever invented the FitBit charger needs an ass whoopin from my grandmother when she was in her prime. This short cord is ridiculous	1
23.03.21 07:21	I just achieved my daily @Fitbit step goals of 10749	0
10.07.20 23:19	Omg the CEO and co-founder of Fitbit just superswiped me on Bumble I knew I was meant to be a sugar baby? If you need me I'll be on a yacht sipping a martini (don't need me)	0
21.03.21 16:56	March 21, 2021 #Fitbit activity: 0 steps taken, 0 kilometers walked/ran, and 1394 calories burned.	0

Figure 1. Example of data labeling

Therefore, to overcome this grey area, only the tweets containing very clear customer needs will be labeled as '1'. The ones that are vague and prone to individual interpretation will be categorized as '0'. The above example clarifies how the tweets are labelled in Excel and partly illustrates what is regarded as a need and what is regarded as no need.

Figure 2 summarizes the remaining tweets after each data preparation step. At the beginning, a total of 416,840 tweets was recorded, which in the end was reduced by 255,211 tweets by the different data filtering techniques. The randomized training set – a data set that is trained to then be used as basis for the machine learning algorithm – contained 10,000 tweets, from which only 130 tweets (1.3%) were labeled as 'containing a need'.

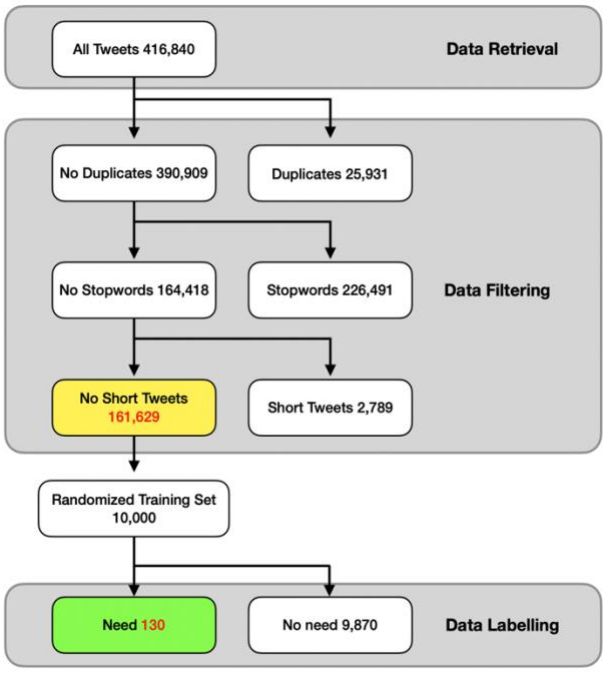

Figure 2. Number of used Tweets in the individual steps

3.4 Data pre-processing

Within this step, the data is pre-processed, and the text is broken down to its basic structure as preparation for the algorithm. Conventional and simple pre-processing includes adjusting the words in a way that only the stem of the word is shown without any capital letters or stops words. But for this research, an extended approach – namely the BERT model – is applied, in which stop-words are kept. The words are not simply tokenized, stemmed down, or put in lower case.

The acronym stands for Bidirectional Encoder Representations from Transformers, which was developed by Devlin et al. (2018). This algorithm can run on Python, which will also be applied in this research. BERT offers the opportunity to train the data set in a way that it recognizes the presence of customer needs. It is able to analyze correlations, and thus stop-words, such as "a", "or", "to" are not ignored as they are also an indicator for the presence of a customer need.

3.5 Manual categorization

After completion of the initial labeling and the assignment of the two labels, the training data set will be further categorized by the coder himself. First of all, each tweet will be checked again, whether a clear customer need is contained or not. Again, the ones that do not contain a clear customer need will be excluded and filtered out. Then, the contents of the tweets labeled with a '1' will be analyzed, and common software, as well as hardware features will be identified. Each tweet also receives a number that corresponds to the type of feature. For instance, Fitbit tweets that address needs concerning the in-built heartrate-monitor, a common feature in smartwatches as well as in fitness-trackers, will be put in one category and be assigned a suitable label number. This categorization allows for a more fine-grained identification of customer needs. Based on the number of needs within one category, certain features of a product are emphasized by customers, which may indicate what aspects require further analysis.

An initial distinction between software- and hardware-features is made, and also a category for other needs, which cannot be easily put into another category because of their uniqueness, is created. It is expected that during the categorization process of the training set, more categories will become more present, either because of the expression used by the customer or the frequency at which they occur. Therefore, in case it is necessary, more detailed categories will be added that allows an appropriate distinction between the customer needs.

3.6 Supervised machine learning

In this last step, the BERT algorithm will be run in Python with the purpose of assigning the corresponding probabilities of containing a customer need for every single tweet. As previously mentioned, only the tweets that show a clear presence of a customer need will be kept. Thus, the model will be modified in a way that only tweets with at least a 50% chance probability of containing a customer need will be labeled with a '1'. The remaining tweets will be labeled '0'.

The BERT algorithm is capable of providing a variety of different measures and scores. For the purpose of this paper, four relevant performance metrics are chosen: accuracy, precision, recall, and the different f-scores. In the following results section, the calculated metrics will be shown and put into context, revealing how effective the algorithm worked for the dataset.

4. RESULTS

This section will give an overview of all relevant findings from the conducted qualitative research and provide a summary of the calculated performance metrics from the applied machine learning algorithm. As shown before in Figure 2, 130 tweets were labeled as '1', meaning that of all 10,000 tweets, only 1.30% of them were considered as containing a need.

4.1 Identified categories

During this step, the needs expressed in the 130 tweets were categorized into one of the seven categories. Figure 3 displays a chart of the created categories and how often a need has occurred. It is possible that one tweet contained multiple needs; thus, the number of occurrences is greater than the number of actual need tweets.

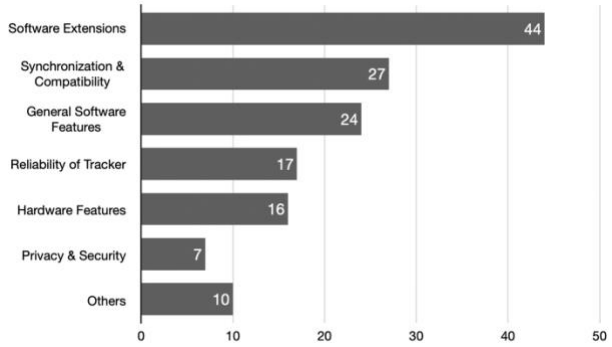

Figure 3. Occurrences of need-tweets for each category

As mentioned in the methodology section, the categorization step was initiated by creating three distinct categories: 'Software Features', 'Hardware Features' and 'Others'. Throughout the process, frequent reoccurrences of the same need became more visible and were grouped in suitable categories, which are elaborated in the following.

General Software Features

This category encompasses all basic and customizable software features that are available to the user. Among those, the most frequently mentioned complaints were the inseparability of language from location and the inability to change to the preferred decimal separator (comma or dot). Next to that, users complained that the watch face designs are 'outdated' and newer, more aesthetic ones should be made. Other software adjustments or bug fixes that do not require the construction of a completely new program or function are also grouped in this category.

Software Extensions

This category is a branch of the previous one and covers all need tweets that were related to software extensions – programs or functions that users desire but which have not been implemented. Examples include a 'drink water notification', which notifies the user in certain time intervals to hydrate him or herself. Furthermore, some users wished for a 'walking detected' function that sends the wearer of the watch a notification to record the walk. Thus, the watch automatically recognizes the movement made by the wearer, detects that it is walking, and then sends the request to record the walk. These software upgrades are not changing the existing general settings of the device but add more features through the construction of new programs or functions.

Synchronization & Compatibility

The last software-related category consists of two similar features. The first one is regarding the ability to synchronize the Fitbit device with a smartphone via a Bluetooth connection and the corresponding speed of data transfer between both devices. About 20 users complained that they had to restart their Fitbit device multiple times a day to get a stable Bluetooth connection. Besides, a few users could not transfer the data to their smartphones, even under a stable Bluetooth connection. The other feature is related to the compatibility of Fitbit devices with other services and apps. Users have wished for more collaborations with fitness-related services or apps such as the current collaboration with MyFitnessPal, an application that enables the tracking of a user's diet with in-depth information about nutrients, comprehensible graphs, and other dietary features.

It is shown on figure 3 that the software features have the highest occurrences of needs, totaling 95 occurrences – 44 in 'Software Extensions', 27 in 'Synchronization & Compatibility' and 24 in 'General Software Features' – which is a majority of 65.52% of all recorded needs.

Hardware Features

Needs that are directed at the physical product are grouped under this category. Among the needs of this category, the most frequently mentioned ones are related to the quality of the display. Some users have complained that the display lacks durability and breaks after short-term use. Another need is the weather sensitivity of Fitbit devices. Apparently, the interface is not usable under colder conditions. Also, some users experienced skin irritations presumably caused by the material of the band.

Reliability of Tracker

This category covers all needs related to the reliability of the Fitbit devices and is thus a sub-group of 'Hardware Features'. Due to its high occurrence, it seemed more adequate to create a separate category for this need. An essential component in all Fitbit devices is the in-built sensors. They provide the necessary data to enable the device to execute the basic functions such as monitoring the heart rate, detecting movement, or counting the steps – in short, the functions that are inherently associated with Fitbit devices. And 17 tweets point out that the data of the in-built sensors are 'far off' and overall 'imprecise'.

Both categories related to the hardware features have a combined occurrence of 33, resulting in 22.76% of all recorded needs.

Privacy & Security

While using a Fitbit device, the user creates data with the in-built sensors and other software functions. The personal information is stored on the device itself and later transferred to the connected smartphone and then to the database of Fitbit. Some users are concerned about how the data will be stored and used by the company, especially because the data is highly sensitive and private. Not only the heart rate and workouts are tracked, but also your sleeping pattern, your diet (in case you track it manually), and most alarming your location might be tracked as well if the user activates the GPS function. In addition, the users have become even more skeptical after the acquisition of Google.

Others

The last category includes all other needs that were too specific and unrelated to be put in one of the previous categories. The majority of the needs are ideas or inspirations for new products or the desire for older generation devices. For instance, one user requests to re-launch the outdated 'pocket-clip' – a small Fitbit device that is not worn as a watch but instead can be clipped to the user's pocket. Another idea raised by mothers is a specifically designed Fitbit that has features to control and check the vital signs of their babies.

These last two categories combined have an occurrence of 17, which is 11.72% of all needs, making it the lowest rate of occurrences.

4.2 Other labeling observations

In addition to the identification of expressed needs and categories, several other observations and abnormalities were found that may have practical implications for future work in this field or at least are helpful for understanding the results of the performance metrics.

4.2.1 Daily reports

First of all, a pattern that stands out and occurs relatively often, considering that the coded dataset only consists of 10,000 tweets, are the daily reports or updates from the Fitbit users. The daily report is a feature offered by Fitbit that shortly summarizes the activities that a user has performed throughout the day. The Fitbit app allows you to easily share your 'achievements' with friends and followers on social media platforms. Several examples of this can be seen below in Figure 4.

I got 24660 seconds of sleep, tracked with my @fitbit. @achievement
I just achieved my daily @Fitbit step goals of 7771 August 10, 2020
Step goal for April 7, 2020 achieved! 11263 total steps today via @Fitbit.
Calorie burn goal achieved! 2818 calories burned on May 6, 2020 via Fitbit

Figure 4. Examples of daily report

The preferred goal can be chosen by the user, and thus, the format of the shared message will slightly differ. In total, 721 daily reports were identified, which is a relatively high occurrence. The daily report itself is intended to motivate users to reach their goals and share them with their friends, but for the identification of new customer needs, they do not contain any relevant information and should be excluded from the data set.

4.2.2 Information overload

Twitter users have the opportunity to write tweets containing up to 280 characters, allowing them to express various thoughts and opinions in one message. Evidently, the majority of identified need tweets contain a notable amount of unnecessary information, which is oftentimes filled with an emotional feeling – indicating an encountered inconvenience or discomfort from the use of the product or service. Going through the need tweets, it can be shown that, on average, those tweets hold more information, which is not helpful for the identification of needs. On the contrary, they do provide a sentimental indication, so whether something is seen as positive or negative.

The need tweet with the shortest message has a character count of 40, which is approximately 8-10 words. On the other extreme, some need tweets contained up to the maximum of 280 characters. The average character count of all need tweets is 179, and need tweets that contained more than 200 characters are 54.

4.2.3 Missing product-model specification

Fitbit has a variety of smartwatches and fitness trackers, but these are rarely mentioned explicitly by the customers. The need is not addressed to a specific model but instead to Fitbit as a whole. For some categories such as 'Privacy & Security', it is reasonable because the customers are concerned about company-wide issues, which are not directed at a specific product model. Unexpectedly, need tweets within the 'Hardware Features' category are also addressed at Fitbit as a company. The underlying reason for this may be the overlapping of similar technological features across different models, and thus, a high interconnectedness is present.

4.2.4 Need-indicating patterns

The tweets purposely directed at either Fitbit, Fitbit-health or Fitbit-support, so the official Twitter pages of Fitbit, which are tagged with a '@'-symbol, have a high likelihood of containing relatively neutral statements, which in some instances can be converted into customer needs. Out of the 130 need tweets, 69 of them were either directed at Fitbit or Fitbit-support – so, slightly more than half of all need tweets.

Another observation that occurred frequently is the gray area described in the methodology section – tweets that are formulated in a manner that could be either containing a need or not, depending on the argumentation used. In particular, tweets directed at Fitbit-support are prone to interpretation. Oftentimes, the customers contact the support to complain about malfunctions of their device. Technical difficulties are the cause, but if an actual need can be deducted or the customer just criticizes a faulty production or natural wear and tear is unclear.

4.3 Performance metrics

After the completion of the data labeling step, in which all 10,000 tweets were checked on the presence of a customer need, the BERT machine learning algorithm was applied. Afterward, 3,300 tweets were randomly selected from the data set and were used to verify the trained neural network for classification. The BERT algorithm made predictions on the 3,300 randomly selected tweets, whether a need is present or not, and these predictions were cross-checked with the actual dataset. In figure 5 below, a confusion matrix is displayed that shows the results of the algorithm. The green-colored quadrants represent the number of correctly made predictions, while the red-colored quadrants represent the number of incorrectly made predictions.

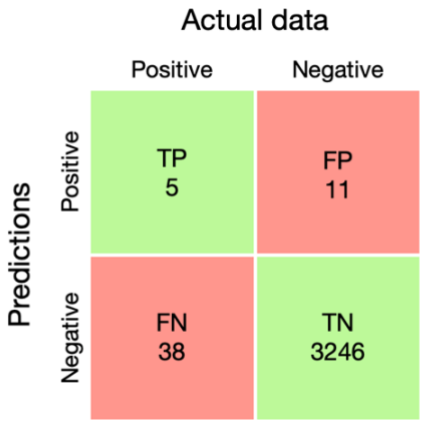

Figure 5. Confusion-matrix: Predictions vs Actual data

According to the confusion matrix, the BERT algorithm correctly predicted an actual need tweet 5 times (True Positive) and also correctly predicted 3,246 tweets that do not contain a need (True Negative). Furthermore, the algorithm falsely classified 11 tweets and labeled them as a need tweet (False Positive). Lastly, 38 need tweets were not identified by the model (False Negative). Given these numbers, the four performance metrics were calculated to evaluate how accurately this algorithm has worked with the provided dataset. The results are listed in figure 6.

Accuracy	Precision	Recall	F-score (β = 0.5)	F-score (β = 1)	F-score (β = 2)
0.9861	0.3125	≈ 0.1163	≈ 0.2804	≈ 0.1695	≈ 0.133

Figure 6. Values of the performance metrics

The accuracy describes the fraction of all instances, which were correctly classified by the BERT algorithm. With a score of 0.9861, the accuracy of the algorithm appears to be significantly high but note that this result may be deceiving and not meaningful.

Because due to the small number of need tweets (130) and the large presence of tweets not containing a need (9,870), the likelihood of having true negatives is increased, as well. Thus, this metric on its own is not sufficient to explain the actual accuracy of the algorithm.
The precision and recall metrics are both constituents of the f-score, as can be seen in figure 7. Precision describes the fraction of correctly predicted needs among all need tweets, while the recall metric, also known as sensitivity, describes the fraction of correctly predicted needs among all identified needs.

$$F_\beta = (1 + \beta^2) * \frac{precision * recall}{(\beta^2 * precision) + recall}$$

Figure 7. F-score formula

The f-score gives an overall indication for how well the model worked, while a value of '1' is considered to be perfect and a value of '0' is considered as a misaligned model. The formula can be slightly alternated by taking different values for 'β':

- β < 1, emphasis is put on precision
- β = 1, harmonic f-score, emphasis on both precision and recall
- β > 1, emphasis is put on recall

For the evaluation of the accuracy of the machine learning algorithm, we take the harmonic f-score (β = 1), giving both precision and recall an equal amount of emphasis. With a score of approximately 0.1695, the accuracy is relatively low. Also, compared to the work of Kuehl et al. (2016) that recorded a harmonic f-score of 0.466, the score is significantly lower.

5. DISCUSSION

5.1 Significance of the identified needs and the level of generalization

The variety and scope of the identified needs are noticeably large. The needs range from minor inconveniences such as a short charging cable (hardware feature) to more customer-infuriating needs like a missing training program or the incorrectness of the heart rate monitor. Especially the category related to software extensions appears to have valuable insights from customers and their perspective on which features should be implemented in the future. Software extensions, as well as general software features, also have the benefit of being fixed by software updates; thus, they do not necessarily require new hardware upgrades, making them achievable in the short term. Hardware features, including the reliability of the tracker, are bound to the physical product and should be considered for the development of newer product models. An actual assessment of the feasibility of these customer needs cannot be provided by this paper, but they may give useful hints at malfunctions and weaknesses of current Fitbit products.

In regard to the first research question, all seven identified categories were already described in prior sections. In addition, it was shown that software-related features are the most prominent customer needs, followed by hardware features and lastly, privacy and security concerns combined with other needs. Despite the fact that the tweets are rarely product-specific, and the formulations are on a surface-level, meaning that no in-depth technical reviews are given, generalizations for other domains are most probably only possible to a limited extent.

The BERT algorithm, which has been used for this dataset, is strongly wired to the provided needs of Fitbit users; thus, a direct transfer to a different data set is not feasible. But note that categories identified in this domain may be taken over to other products or services. For instance, data privacy and security are a target for customer feedback in many other technology-related industries. Also, the identified needs can be projected onto other companies, which also operate in the wearable device industry.

5.2 Elaboration on the performance metrics

With a percentage of 1.30, the occurrence of customer needs in the training dataset for Fitbit products is relatively low compared to the work of Kuehl et al. (2016) for the e-mobility domain – which reached a percentage of about 13.86 (332 need tweets in a dataset of 2.396 tweets). Due to the low occurrence of need tweets, the number of true negatives (correctly predicted tweets containing no need) is high. Because of the high number of true negatives, the accuracy score of the analysis is very high (= 0.9861). The reason for this might be the relatively small training dataset, resulting in an insufficient number of identified need tweets. Hence, because of the small number of needs within the training dataset, the chance that the algorithm predicts a tweet as a need tweet is low, and therefore, the majority of tweets will be labeled as '0' (containing no need). The accuracy metric on its own is therefore not significant and, for this paper's research, not meaningful.

On the other hand, the calculated f-score, which makes use of both recall and precision, is a somewhat better indicator of the overall accuracy of the algorithm. With a value of approximately 0.1695, the f-score indicates a badly trained neural network for classification. Both scores for recall and precision are low; thus, the resulting f-score is low, even if different β-distributions are used. Concrete values of how often the algorithm has made wrong predictions can be found in section 4.3.

6. LIMITATIONS AND FUTURE DIRECTION

The results of the algorithm were significantly worse than initially expected. The model predicted most of the need tweets wrong, and thus, a lot of needs were not discovered by the neural network. The use of the BERT algorithm, at least for this paper and conducted research, has only little to no practical relevance due to the high number of errors. But they still indicate what the possible errors in the research were and may give advice and direction for future projects.

To investigate the usefulness and potential of the BERT algorithm, future research in different domains are suggested. By attempting to use this approach for different sectors, where the success of the product or service is dependent on other factors, more positive results may result. Note that the customer community and the manner in which they formulate their reviews do have an impact on the algorithm as well. Some products or services tend to have customer reviews that are more descriptive, and some are more colloquially written. As an example, it became clear that even among the different approaches and investigated industries within the bachelor circle of this thesis topic, the BERT algorithm resulted in different prediction scores – two researchers were more successful than the other two.

Measures to prevent inconsistency and to provide a better training set with the intention of achieving a higher prediction rate should be implemented. Especially in the data labeling step, in which the tweets are coded as either '1' (containing a need) and '0' (not containing a need), some improvements can be made. To counter inconsistency in the way the tweets are labeled, more coders should review the presence of a need instead of one researcher. By cross-checking the information, possible needs are not overlooked and the tweets, which are located somewhere in the gray area, can be better discussed.

Also, sentence structures that are most likely unrelated to need tweets, such as the daily reports of Fitbit users, should be removed in a separate filtering step.

7. CONCLUSION

Identifying customer needs for Fitbit products by using the BERT machine learning algorithm has not shown any reliable results and thus, has no practical implications. On the contrary, the needs that were identified and put into the respected categories are examples of real-life desires and wishes of customers that most certainly are useful to some extent. They could be either directly implemented, though the technical feasibility remains unknown and not within the scope of this thesis, or they can be used as inspiration for future directions for the development of the products. Especially, software-based needs provide a variety of missing functionalities in current product models.

This paper partly filled the domain-gap and showed that the machine learning algorithm is not adapted well enough to be applied to the social media data from Fitbit users. Possibly, with suitable adjustments to the training set, the algorithm might have performed better. It may have reached a sufficient number of correct predictions, and thus, an integration of this algorithm as a continuous process might have worked well.

Lastly, this algorithm still relies on human interaction and requires a data set from which it can be trained. Further research may find ways to extend the thus made *Needmining* efforts and transform the binary classification model into a model that can distinguish and categorize needs on its own.

8. ACKNOWLEDGEMENTS

Herewith I would like to express my gratitude towards my first supervisor Dr. Dorian E. Proksch, for providing me with the research data and his guidance as well as his expertise throughout this whole thesis process. Besides, I want to thank my bachelor circle group colleagues Joel Tyhaar, Justus Jürgens, and Marc van Huizing for reviewing my work and providing me with a lot of useful insights and ideas.

9. REFERENCES

Anuradha, J. (2015). A brief introduction on Big Data 5Vs characteristics and Hadoop technology. Procedia computer science, 48, 319-324. doi: https://doi.org/10.1016/j.procs.2015.04.188

Bharadwaj, N., and C. H. Noble. 2015. Innovation in data-rich environments. Journal of Product Innovation Management 32: 476–78. doi: 10.1111/jpim.12266

Birch-Jensen, A. (2020). The evolving role of customer focus in quality management.

Cai, S. (2009). The importance of customer focus for organizational performance: a study of Chinese companies. International Journal of Quality & Reliability Management. doi: https://doi.org/10.1108/02656710910950351

Cecere, G., Corrocher, N., & Battaglia, R. D. (2015). Innovation and competition in the smartphone industry: Is there a dominant design?. Telecommunications Policy, 39(3-4), 162-175. doi: https://doi.org/10.1016/j.telpol.2014.07.002

Chan, L. K., & Wu, M. L. (2002). Quality function deployment: A literature review. European journal of operational research, 143(3), 463-497. doi: https://doi.org/10.1016/S0377-2217(02)00178-9

Chuah, S. H. W., Rauschnabel, P. A., Krey, N., Nguyen, B., Ramayah, T., & Lade, S. (2016). Wearable technologies: The role of usefulness and visibility in smartwatch adoption. Computers in Human Behavior, 65, 276-284. doi: https://doi.org/10.1016/j.chb.2016.07.047

Davenport, T. H., Barth, P., & Bean, R. (2012). How 'big data' is different.

Del Vecchio, P., Di Minin, A., Petruzzelli, A.M., Panniello, U. and Pirri, S. (2018), "Big Data for open innovation in SMEs and large corporations: trends, opportunities, and challenges", Creativity and Innovation Management, Vol. 27 No. 1, pp. 6-22. doi: https://doi.org/10.1111/caim.12224

Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805.

Edvardsson, B., Kristensson, P., Magnusson, P., & Sundström, E. (2012). Customer integration within service development—A review of methods and an analysis of insitu and exsitu contributions. Technovation, 32(7-8), 419-429. doi: https://doi.org/10.1016/j.technovation.2011.04.006

El Naqa, I., & Murphy, M. J. (2015). What is machine learning?. In machine learning in radiation oncology (pp. 3-11). Springer, Cham. doi: https://doi.org/10.1007/978-3-319-18305-3_1

Elmer, G., Langlois, G., & Redden, J. (Eds.). (2015). Compromised data: From social media to big data. Bloomsbury Publishing USA.

Fan, J., Han, F., & Liu, H. (2014). Challenges of big data analysis. National science review, 1(2), 293-314. doi: https://doi.org/10.1093/nsr/nwt032

Fisher, M., Houghton, M. and Jain, V. (2014), Cambridge IGCSE® Business Studies Coursebook, Cambridge University Press, available at: https://books.google.de/books?id=PdtkAwAAQBAJ

Geum, Y., Jeon, H. and Lee, H. (2016), "Developing new smart services using integrated morphological analysis: integration of the market-pull and technology-push approach", Service Business, Vol. 10 No. 3, pp. 531-555.

Ha, N. (2016). Smartphone industry: The new era of competition and strategy.

Hauser, J.R. and Griffin, A. (1993), "The Voice of the Customer", Marketing Science, Vol. 12, pp. 1–27.

Hoyer, W. D. & MacInnis, D. J., 2001, Consumer Behaviour. 2nd ed., Boston, Houghton Mifflin Company.

Hsiao, K. L., & Chen, C. C. (2018). What drives smartwatch purchase intention? Perspectives from hardware, software, design, and value. Telematics and Informatics, 35(1), 103-113. doi: https://doi.org/10.1016/j.tele.2017.10.002

Jalilvand, M. R., Esfahani, S. S., & Samiei, N. (2011). Electronic word-of-mouth: Challenges and opportunities. Procedia Computer Science, 3, 42-46. doi: https://doi.org/10.1016/j.procs.2010.12.008

Jeong, S. C., Kim, S. H., Park, J. Y., & Choi, B. (2017). Domain-specific innovativeness and new product adoption: A case of wearable devices. Telematics and Informatics, 34(5), 399-412. doi: https://doi.org/10.1016/j.tele.2016.09.001

Kano, N. (1984). Attractive quality and must-be quality. Hinshitsu (Quality, The Journal of Japanese Society for Quality Control), 14, 39-48.

Kotler, P. and Armstrong, G. (2001), Principles of Marketing, World Wide Web Internet And Web Information Systems, Vol. 42, available at:http://doi.org/10.2307/1250103.

Kuehl, N., Scheurenbrand, J., & Satzger, G. (2016). Needmining: Identifying micro blog data containing customer needs. arXiv preprint arXiv:2003.05917.

Labrinidis, A., & Jagadish, H. V. (2012). Challenges and opportunities with big data. Proceedings of the VLDB Endowment, 5(12), 2032-2033. doi: https://doi.org/10.14778/2367502.2367572

Limehouse, D. (1999). Know your customer. Work study. doi: https://doi.org/10.1108/00438029910262518

McAfee, A., Brynjolfsson, E., Davenport, T. H., Patil, D. J., & Barton, D. (2012). Big data: the management revolution. Harvard business review, 90(10), 60-68.

Moe, W. W., & Schweidel, D. A. (2017). Opportunities for innovation in social media analytics. Journal of Product Innovation Management, 34(5), 697-702. doi: https://doi.org/10.1111/jpim.12405

Rogers, M., & Rogers, M. (1998). The definition and measurement of innovation.

Saarijarvi, H., C. Gronroos, and H. Kuusela. 2014. Reverse use of consumer data: Implications for service-based business models. Journal of Services Marketing 28: 529–37.

Scharnbacher, K., & Kiefer, G. (2010). Kundenzufriedenheit: Analyse, Messbarkeit und Zertifizierung. Walter de Gruyter.

Singh, H. (2006). The importance of customer satisfaction in relation to customer loyalty and retention. Academy of Marketing Science, 60(193-225), 46.

Tankovska, H. (2021, January 28). Number of global social network users 2017-2025. Statista. https://www.statista.com/statistics/278414/number-of-worldwide-social-network-users/

Trabucchi, D., & Buganza, T. (2019). Data-driven innovation: switching the perspective on Big Data. European Journal of Innovation Management.

Troilo, G., De Luca, L.M. and Guenzi, P. (2017), "Linking data-rich environments with service innovation in incumbent firms: a conceptual framework and research propositions", Journal of Product Innovation Management, Vol. 34 No. 5, pp. 617-639.

Urban, G. L., & Hauser, J. R. (2004). “Listening in” to find and explore new combinations of customer needs. Journal of Marketing, 68(2), 72-87.

Urbinati, A., Bogers, M., Chiesa, V., & Frattini, F. (2019). Creating and capturing value from Big Data: A multiple-case study analysis of provider companies. Technovation, 84, 21-36.

Vailshery, L. S. (2021, January 22). Global connected wearable devices 2016-2022. Statista. https://www.statista.com/statistics/487291/global-connected-wearable-devices/#:~:text=The%20number%20of%20connected%20wearable

Van Rijsbergen, C. (1979). Information retrieval: theory and practice. In Proceedings of the Joint IBM/University of Newcastle upon Tyne Seminar on Data Base Systems (pp. 1-14).

Wu, J., Li, H., Lin, Z., & Zheng, H. (2017). Competition in wearable device market: the effect of network externality and product compatibility. Electronic Commerce Research, 17(3), 335-359. doi: https://doi.org/10.1007/s10660-016-9227-6

Zairi, M., 2000, Managing Customer Dissatisfaction Through Effective Complaint Management Systems, The TQM Magazine, 12 (5), pp. 331-335.