Can data from Twitter produce useful feature suggestions in the audio streaming service field? - An example of Spotify

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

ABSTRACT

This research paper examined whether relevant feature suggestions could be found in short messages (tweets) retrieved from Twitter and whether a natural language processing algorithm (BERT) could identify these at the example of Spotify. The data was collected via an API using only messages mentioning Spotify's help account. Then it was filtered further. All 10,000 tweets were labelled whether they contained a need and clustered according to the size of the proposed feature and which idea they contained. Results showed that 5.22% of all tweets contained a need and the algorithm had a f1-score of 53%. The findings suggest that Twitter can be a source of ideas for Spotify. Furthermore, the analysis of the most suggested ideas showed that many of them can be considered valuable. There were ideas of different sizes present, with smaller ones forming the majority. The data was found to be of considerable velocity as users reacted quickly to events and problems. Variability was displayed in the form of trending suggestions. Results suggest that this need mining approach may serve as a tool to more efficiently collect user suggestions compared to standardly used methods.

Graduation Committee members:
First supervisor: Dr. Dorian E. Proksch
Second supervisor: Dr. Tim G. Schweisfurth

Keywords
Data-driven innovation, user-based innovation, machine learning, BERT, Twitter, customer needs, need mining

1. INTRODUCTION

Spotify is a company providing audio streaming software. It has up to 360 million users per year and is available on most smart devices. Its users stream and download songs from a vast collection of more than 82 million tracks. There are many more features such as the automatic creation of playlists, shared session listening and podcasts. Based on a freemium subscription model, users can pay for extra benefits such as no advertisements and the ability to download songs. Spotify is the largest audio streaming service in the world, having 32% of the total market share (Götting, 2021).

A challenge for Spotify is to generate new features that customers need. It can be hard to find lead users and think of innovative functions in an abstract manner. The occurrence of the need and subsequent communication of it is usually timely dispersed, which leads to needs not being reported because the channel to report is not available immediately. Surveys often fail to produce significant insights because respondents cannot remember the need they had. This problem can be solved by using Twitter, which enables immediate communication of the need. Twitter is a social media platform where over 500 million tweets are sent every day (Mathioudakis & Koudas, 2010). There, users can post short text messages or other media. Users can reply to each other's tweets, repost them and like them. With a mainly English-speaking user base and a tendency to tweet anything that its users find notable, it is a suitable environment to search for how Spotify can use customers to gain new feature ideas. Many Spotify users are also Twitter users, and Spotify has a support account dedicated to customers there.

The tweets may contain dissatisfaction about the application not fitting their needs and explicit requests for features which can be used to improve Spotify's research and development. A complaint voiced by a customer is essentially a user need. This data can be gathered by an AI algorithm according to some criteria and may help identify needs for new features (Kuehl et al., 2016). Current research suggests that there is a positive relationship between a firm's presence on social media and the probability of introducing a product innovation (Bertschek & Kesler, 2018). User needs are extremely valuable for product improvement as they provide input which innovators are not aware of and help assess the desirability of new features.

The promise of big data is that it can save vast amounts of time and money if traditional ways were used by identifying the need-containing tweets among hundreds of non-relevant ones. Traditionally, employees searched for users expressing needs, generating labor costs which an AI makes redundant. The possible drawback of AI concerns the ability to identify relevant data with high reliability. This filtering problem becomes less relevant with the evolution of AI techniques but is still essential. Thus follows the main research question with three sub-questions:

Can a company like Spotify use data mined by an AI on Twitter to generate useful feature suggestions?

i. Can Twitter be a source of input?
ii. Are the selected tweets promising to aid innovation at Spotify?
iii. Can the AI reliably find relevant input?

2. LITERATURE REVIEW

2.1 Data-driven innovation management

Data-driven innovation management can be defined as an innovation approach that uses data collection to improve innovation efforts. Over the last two decades, the increasing amount of collected data has led to changes in the way businesses operate, by changing the landscape of information systems (Lies, 2019). The total amount of data in the world grew from 2 Zettabytes in 2010 to 79 Zettabytes in 2021 (IDC, 2021). More companies adopt data centric innovation approaches as infrastructure turns digital, giving rise to new technologies. These technologies open up a new world of possibilities for innovation (Fichman et al., 2014). Especially social media, Industry 4.0, Internet of Things, and Big Data in general change the ways that innovation emerges. Data-driven innovation management is enabled and driven by IT, but also constrained by it. A company needs to fit its IT with their current routines and organization to successfully make use of it (Pentland & Feldman, 2008). As a complement, a user-centered innovation approach can be used more extensively, as social media provides lots of data for analysis of customer needs. User-centered innovation is an approach which assumes that a company can gain unique insights for innovation by asking and analyzing user's needs, analyzing the way they use current products and by monitoring their behavior during that process (Trabucchi et al., 2018). Users are seen as two-fold actors: They provide information that improves the innovation, and they are also the group that will use the final product. Data collected from users can be used in all phases of the innovation process and tailor the product better to the customer (Karat, 1996).

2.2 Social media

Social media is a broad term referring to some form of internet-based communication on a platform. Users can use this to engage in conversations, share information and create content. There are many kinds of social media platforms, such as sites for photo-sharing, social networking, blogs and wikis (Kaplan & Haenlein, 2010). They usually are available at no cost. The value for personal users lies in the possibility to satisfy their social needs and creativity. For professionals, the billions of users make for a large possible reach with many different types of users. Also, an audience can be built and interactions with customers are possible.

From an innovation perspective, social media can be extremely relevant (Brandtzaeg & Følstad, 2016). Users are vital for innovation, as they act as informants buts also as consumers and critics (Von Hippel, 2006). Since social media provides a space for users to communicate with companies, it is a promising opportunity to gain new insights. User input can provide lower cost of knowledge and is generally faster than other forms of input. There are different ways companies incorporate user knowledge: Many companies use idea generation contests, questionnaires, co-development with online users. More casual but not less valuable approaches include reading comments made by users and engaging in discussions (Roberts & Piller, 2016). Social media is a source of huge amounts of data which companies should not neglect as it can improve innovation efforts (Bertschek & Kesler, 2018). Moreover, traditional methods of evaluating user needs do not produce effective and efficient results. Methods such as questionnaires or interviews are time inefficient and cost intensive. Datasets from Twitter come at a low price for companies compared to what they usually spend on innovation activities. Machine learning as a method of analysis is cheap compared to other methods and can analyze larger datasets which may increase accuracy (Kuehl et al., 2016). Research has been done on this topic, by Misopoulus (2014) and

Kühl (2016). Misopoulus used sentiment analysis to analyze tweets and users' needs, but Kühl was the first to use machine learning for detecting customer need elicitation. In his research, Kühl (2016) identified need-containing micro blog instances, quantified and evaluated them. This research paper will orient itself along the lines of his research. He focused on e-mobility in Germany, also gathering data from Twitter.

2.3 Big data

"Big data is the information asset characterized by such a high volume, velocity and variety to require specific technology and analytical methods for its transformation into value."(De Mauro et al., 2016). Big data is of such complexity and size that traditional methods for analysis do not suffice (Ghani et al., 2019). Pertaining to the definition of big data, three more attributes have come into consideration in the last years, making it 6 V's that define big data: Volume is the size of the data, velocity is the speed with which it is received and variety signals that there are different types of data. Variability is about variance in the flow and amount of generated data, value is the gain that can be derived from it and veracity is the data's quality (Emmanuel & Stanier, 2016). There is no consensus on a definition, due to the novelty of big data. Big data has come into existence as a side product of consumption (Trabucchi et al., 2018) and is therefore available at nearly no cost. The challenge of efficiently using big data lies in eliminating low quality data from the dataset and simultaneously discovering high quality data (Erevelles et al., 2016). Due to its complexity and size, analysis can identify trends and problems which can be used by companies as indicators for innovation possibilities as well as which trajectory development should follow. Especially user-generated big data (UGBD) can provide more comprehensive insights than companies would have using traditional approaches (Trabucchi et al., 2018). UGBD are all types of data which are created by users when using a product or service (Saura et al., 2021). These data are for example about purchases made or scrolling behavior. Analysis could e. g. uncover trends among customers or show the optimal layout of a website according to customers behavior on it. Another way UGBD is used is sentiment analysis: Sentiment analysis has been viewed as Natural Language Processing task at different levels, beginning on a document level, then a sentence level and newly also on the phrase level (Agarwal et al., 2011). The approach tries to identify the opinions of authors on specific subjects, often using advanced analysis methods such as algorithms (Lies, 2019). Most sentiment analysis is done on data from social media. The short messages of users expressing opinions are a gold mine for companies, as they give indication to what customers think of them. This feedback can be used to take subsequent actions such as reviewing the product (Feldman, 2013). Organizations are affected by beliefs and perceptions, which gives almost every company an incentive to perform sentiment analysis (Liu, 2012).

3. METHODOLOGY

This section outlines the research methods used to analyze the tweets. These will be used as data source to examine whether they contain useful feature suggestions for the audio-streaming company Spotify. To do this, the data will firstly be prepared for the machine learning algorithm. After preparation, the needs are turned into feature suggestions and clustered according to their size. Then, the BERT model will be used to try to automatically find needs. The dataset for training of the algorithm will consist of 5000 tweets. The algorithm will then analyze 5000 tweets and classify them whether they are needs. The programs Microsoft Excel and Python will be used to perform the research. The following figure gives an overview of the steps taken and is an adaptation of Kühl's (2016) methodology:

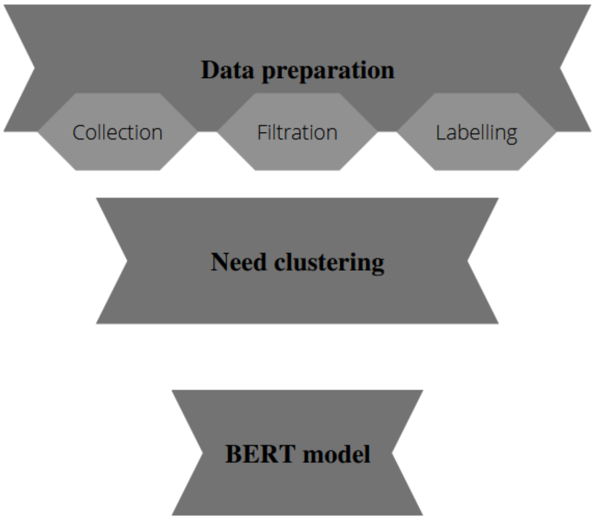

Figure 1. Methodology adapted from Kühl (2016)

3.1 Preparation

The data will be prepared similarly to the way Kühl (2016) did. The preparation process will consist of three steps to make the data useable by the machine learning algorithm. The data will firstly be collected from a Twitter API, then it will be filtered and labelled. The needs will then be clustered, and a machine learning algorithm will be used.

3.1.1 Data collection

The data will be retrieved from Twitter using an API available through platform's website. It contains a rich data source, and the data is available without legal restrictions. This streaming API can retrieve data according to specifiable criteria. The latest tweets matching these criteria at the time will be downloaded and used as the dataset. The number of tweets will be 10,000 after filtration. In this research, only English tweets made mentioning Spotify's customer support account (@SpotifyCares) were collected. The tweets used in the final dataset were made in a timeframe from 26th October 2021 to 30th March 2022.

3.1.2 Data filtration

Retweets and duplicates will not be considered relevant. The dataset will be filtered further to make the algorithm more precise by increasing the richness of information. Not all tweets matching the criteria mentioned above will be used as there are tweets that are not relevant due to their short length. To increase the chance of a tweet mentioning a need, they will be filtered according to the following criteria: Tweets may not be shorter than fifty characters, as these tweets are very unlikely to contain a need due to them being insufficiently long to elicit a need. Since each tweet in this dataset mentions "@SpotifyCares", a tweet will always have 13 characters already used for the mention. Thus, the actual minimum character length threshold is 63 characters. Tweets with more than 3 "@" signs were removed. These tweets mentioned more than three accounts and have a low probability to contain a need, as feature suggestions usually mention only Spotify's help account. Including these tweets would worsen the ratio of needs and ideas to irrelevant tweets.

Tweets that mention an URL will be excluded due to them often being advertisements and because they give more context to the tweet which cannot be used by the learning algorithm, which could lead to false or incomplete interpretation. Tweets containing specific keywords will be eliminated to reduce the chance of spam and low-quality data. Filtering can accidentally eliminate tweets containing needs but not fitting the criteria, so caution is taken to not alter the dataset too drastically. There were changes made to the dataset of which an overview follows:

There are numerous keywords which were used to filter the dataset to limit spam and to remove tweets about topics such as the Joe Rogan controversy which do not provide feature suggestions and reduce the dataset's quality. Numerous keywords related to this controversy, which ensued after Joe Rogan uploaded an episode of his podcast with Robert Malone in which both disseminated false information about the coronavirus. The reasons for excluding the following keywords are in brackets:

1. "Rogan" (Joe Rogan, a celebrity who had controversy arise from a podcast)
2. "JRE" (The Joe Rogan Experience, the podcast which received criticism on Twitter)
3. "censor" (Also related to JRE, some users advocated for Spotify not to censor the controversial episode)
4. "Neil" (Neil Young, a musician who left Spotify after the JRE controversy in protest to it and thus his music disappeared, leading to demands of bringing him back as well as insults towards him)
5. "misinformation" (Also used in relation to the JRE controversy, some users demanded that Spotify should not tolerate the public spread of misinformation about the coronavirus)
6. "disinformation" (Also used in relation to the JRE controversy, some users demanded that Spotify should not tolerate the public spread of disinformation about the coronavirus)
7. "SUGA" (A member of the K-Pop group BTS, users spammed about him)
8. "Namjoon" (A member of the K-Pop group BTS, users spammed about him)
9. "DARARI" (A K-Pop artist, who users spammed about)
10. "BTS" (A K-Pop group, which users spammed about)
11. "Jungkook" (A K-Pop artist, who users spammed about)
12. "Butter" (A song by the K-Pop group BTS)
13. "GeffenRecords" (A record label of a K-Pop artist, who users spammed about)
14. "Jaehyun" (A K-Pop artist, who users spammed about)
15. "JIN" (A K-pop artist, who users spammed about)
16. "crash" (Tweets containing "crash" usually reference the crash of the Spotify application or their servers, and thus exhibit a low probability to also contain a relevant feature suggestion)
17. "Taylor" (Taylor Swift, who users spammed about after she released an album)
18. "all too well" (The name of a song in the new album of Taylor Swift, which users spammed about)
19. "ALLTOOWELL" (Another variation which references the song Taylor Swift released)

Furthermore, all tweets starting from 19:00 on 8.3.2021 to 0:00 on 10.3.2022 were removed due to a service error where many users were logged out, which resulted in more than two thousand irrelevant tweets. In addition, tweets sent from 18:30 until 22:00 on 16.11.2021 were removed due to a service outage, which left users unable to connect to Spotify's servers and caused hundreds of users to tweet about it, which was too complex to remove using filters. Some tweets containing needs will be lost due to the removals, but the significantly higher number of irrelevant tweets makes this necessary.

For the creation of the dataset, 32062 tweets in total were used. The data preparation process thus removed 22062 irrelevant tweets. The number of tweets removed is large due to the amount of spam which is present in the dataset. 31.19% of the total tweets collected are still in the dataset.

3.1.3 Data labelling

The dataset of 10,000 tweets is labeled whether it contains a need or innovative idea by using "1" for tweets that contain a need or idea and "0" for tweets which do not. The decision of whether a tweet contains a need or suggestion is made as objectively as possible. Tweets are labeled containing a need or idea if the tweet elicits a need or idea to some degree. Implicit needs are also considered needs as well as e. g. tweets where the user switches to a competitor due to being dissatisfied while giving a reason. Not considered a need are tweets that report crashes or need help from support, the reason being the tweets' irrelevance to the research question.

3.2 Clustering

The tweets which contain a need or idea are individually assessed. Each tweet is transformed into a feature suggestion that the tweet exhibits. Similar tweets are given the same name to show which innovations are the most popular. General categories such as type of innovation are not used due to the small scope that is applied. Moreover, this research only considers relevant needs and ideas, and a broad categorization is too shallow to allow insights. Due to the variance of feature suggestions, narrow categories would be too specific. Instead, feature suggestions are categorized by size of the innovation. There are three categories for size, Small, Medium and Large. A suggestion is thought to be a small innovation if it presents an innovation which is largely implemented already by Spotify on other devices or by other companies, or if it is a small feature which does not have a large impact on the current state of Spotify and their applications. Medium innovations are those which are larger size than small innovations but lack potential or size to be a large innovation. Large innovations are innovations which are complex, of large size or new.

3.3 Machine learning algorithm

The algorithm used is a BERT model. It is a state-of-the-art language model developed by researchers at Google. It bidirectionally encodes representations from transformers. The model learns to recognize the context of the word based on the words next to it. The algorithm will learn from the first dataset and will be applied to the second dataset to select tweets according to the algorithm's learnings. The order of the 10,000 tweets is randomized by the model to prevent the machine from putting too much learning emphasis on tweets about topics which disproportionately occurred during the first dataset compared to the second. Moreover, the machine would become less accurate in correctly identifying needs in the second dataset.

4. RESULTS

This section focuses on the analysis and key findings of the tweets containing a need. The dataset of 10,000 tweets contained 511 with a need. These tweets were then transformed into the specific need for a feature they elicited, and afterwards the size of the feature as an innovation was determined. Following this, the research question whether these suggestions may be relevant to Spotify is discussed. Lastly, the analysis of the performance measures of the BERT model follows.

4.1 User needs and ideas

In the final dataset, there were 511 tweets containing a need or idea. This means that 5.11% of all tweets contained one. While labelling, similar ideas were given the same feature suggestion name so that the frequency an idea is mentioned can be analyzed.

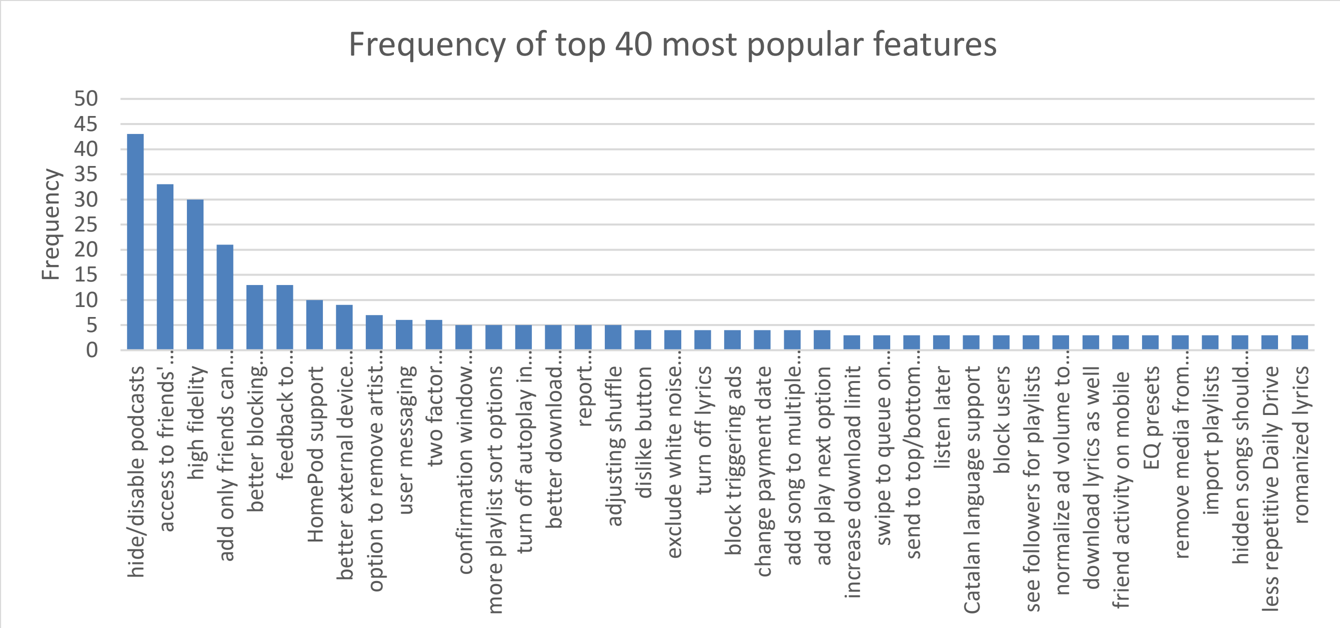

Figure 2. Graph showing the top 40 most frequent feature suggestions

The graph in Figure 2 shows that there are a few popular suggestions, after which follows a long chain of many less frequent suggestions. In total there are 216 unique feature suggestions. There are 7 ideas with more than 10 mentions. Most suggestions have a low frequency of suggestion, but this resulted in more variety, which means that more unique ideas are present in the dataset. The ten most popular suggestions are listed and explained to gain insight into what is needed by users.

4.1.1 Hide/disable podcasts

The most suggested feature with 43 mentions is the ability to hide or disable podcasts. Spotify shows podcasts in the main screen, and users can neither move where the podcasts are shown, nor can they choose not to have them displayed. Spotify started implementing podcasts into their app in 2018 and has since experienced strong growth in their podcast offering and revenue (Steele, 2019).

Example tweet: "@SpotifyCares @spotifypodcasts PLEASE LET ME TURN OFF PODCASTS. I do not want to see this in my app. I'm paying the subscription. Don't force me to switch to another service. Let me use it how I like."

4.1.2 Access to friends' individual playlists

The second most asked for feature is the ability to access individual playlists, which are playlists based upon a user's activity. Users can only access their own individual playlists, not ones of friends. This need appeared in the dataset shortly before the annual release of Spotify Wrapped in December, a summary which shows yearly usage statistics, such as top songs and favorite artists. Users wanted to see their friends' top songs playlists among others, which resulted in 33 mentions of this feature suggestion.

Example tweet: "@SpotifyCares is there a way I can listen to someone else's top songs from 2021? We've tried sharing link and manually going to their profile to listen, but each time it just brings up my own top songs playlist. Many thanks"

4.1.3 High fidelity

High fidelity audio quality has been recognized by Spotify as a need and it was announced by Spotify that the feature would be available in select markets by the end of 2021. This has not happened and thus users are continuing to suggest lossless audio. The feature has been suggested 30 times.

Example tweet: "@SpotifyCares what happened to Spotify hifi/studio or whatever it was gonna be called. I need my hq audio please ?"

4.1.4 Add only friends can add playlist option

Spotify has two options on who can add two playlists. Either only the owner can add, or the playlist is made collaborative, and everyone can add to the playlist. Many users complained about unauthorized users adding to their playlists and suggested the option to choose who can add to their playlists. The feature request was mentioned 21 times. Spotify included this feature in a later update (Steele, 2022).

Example tweet: "@SpotifyCares is there a way to make a Playlist collaborative but only with some people and not with the millions of users you have? Thank you ?"

4.1.5 Feedback to recommendations

Spotify recommends many different things in various areas of the application. It suggests podcasts, songs, artist, playlists and more. Users expressed their dissatisfaction of the missing option to give feedback. They often reported that even if they hid the recommendation, the algorithm would continue suggesting similar content. Thus, a feature to give feedback in some way to recommendations was desired. The idea received 13 mentions.

Example tweets: "@SpotifyCares How on Earth do I stop Satanic podcasts being suggested as something I'd be in2? I don't even want 2 risk clicking (2 block) it in case I have even more foisted on me by some stupid algorithm that thought it would be a good match. Listening 2 metal != satanist ?"

“@SpotifyCares Why does your Release Radar keep giving me live recording Pixies songs for eight weeks in a row(no exaggeration), despite my marking them with "I don't like Pixies" every time? "Spotify cares"... you don't even care enough to write a working like/dislike button.”

4.1.6 Better blocking synchronization

Users can click a “do not play” button on artist profiles and can click “hide” on songs in playlists which they do not want to listen to. Many users voiced their dissatisfaction with this feature as it does not constitute a complete block of artists and songs, which means that the songs and artists are still visible but greyed out, and songs with a feature from a blocked artist will still play. Hidden songs will also still be recommended f. e. in Discover Weekly, a playlist meant to find new music. 13 users voiced their dissatisfaction with the current way “blocking” works.

Example tweets: “@SpotifyCares is there a way to permanently mute artists? I'm trying to avoid music from abusers (e.g. Ryan Adams) but y'all keep putting them in my Spotify Radio playlists!”

“@hadyngreen @MattAutomata @SpotifyCares "okay okay we get it, no Jordan Peterson, BUT what about Jordan Peterson ON SOMEONE ELSE'S CHANNEL???"”

4.1.7 HomePod support

Users have been asking for HomePod support which would enable using it as a music playback device. A HomePod is an Apple product which is similar in function to Amazon’s Alexa. 10 tweets mentioned HomePod support.

Example tweet: “@SpotifyCares After waiting for 2 years, there's still no support for homepod integration, I might change to other platforms which do support that, sadly.”

4.1.8 Better external device integration

There are many similar but slightly different feature suggestions relating to the integration of external devices. They were grouped under this feature name due to the larger number of missing features for external integration, which constitute a general lack of features and seamless use when using external devices. The suggestions were mentioned 9 times. More example tweets are provided to show that there seem to different problems which are all related to external device integration.

Example tweets: “So @Spotify just outright ignores the "Autoplay" off setting now when playing on a remote device? Both Cast and Connect. Very annoying. @SpotifyCares”

“@SpotifyCares Appreciate it but it's a software UX issue, not an account one.To get my watch to play offline, I need to first connect to my phone via Bluetooth, set the watch as the playback device, then switch to offline.Surely just clicking play should make it the active playback device?”

“@SpotifyCares @Spotify why doesn't spotify kids have the ability to broadcast to a Google home speaker? I'd like to listen to my music at the same time my kids are listening to theirs but kids won't broadcast only my main account.”

4.1.9 Option to remove artist videos in Wrapped

Each year in December, Spotify creates a “(Year) Wrapped”, which is a visualized summary of statistics also including most listened artists. Artist had the option to record a video of themselves which would play when the artist is shown in the ranking. 7 users voiced their dissatisfaction and requested an option to turn artist videos off.

Example tweet: “@SpotifyCares hi, is there a way to turn off the artist videos in my wrapped playlist? They're incredibly annoying, i just wanna listen to music, not to people going on about how thankful they are to the listeners ?”

4.1.10 Two factor authentication

Two factor authentication is a common security feature where users must use another device as authenticator, to confirm they are logging into the device at hand. The need for this security feature was expressed 6 times.

Example tweet: “Hey @Spotify / @SpotifyCares given how many attempts other people make to reset my password, it would be really nice if you could add 2FA. You're one of the few services I have left that doesn't do it! Seems like a glaring security oversight on your part.”

The research question whether Twitter can serve as a source of ideas can be answered with yes. The dataset contained 511 needs which shows that there is a considerable amount of data to be found on Twitter. The graph displaying the top 40 feature suggestions exhibits different levels of popularity among individual suggestions. Thus, it seems possible to measure to an extent what features users need the most. With 216 unique ideas, there is variety in the dataset which has been found to increase firms’ innovation performance (Ghasemaghaei & Calic, 2020). The top ten suggestions show that users also directly ask for features as seen in the example tweets. Another notable occurrence was that some feature suggestions only appeared in a timeframe around a specific event. The feature “Access to friends' individual playlists” only appeared around the time Spotify 2021 Wrapped became available, when friends could not see each other’s personalized playlists. Users were also found to report recently arisen needs, such as in the case of the “Add only friends can add playlist option” feature, where bots added to every collaborative playlist as there were no protection measures. The need grew quickly in mentions shortly after the bots started. This shows that there is considerable data velocity and a possibility to monitor for recently emerged needs.

4.2 Size of innovation

All specific feature suggestions were labelled according to the size of the innovation. The three categories were Small, Medium, and Large. The attributes are meant to show the distribution of innovation size of unique ideas in the dataset. This can give insights into whether Twitter can also be used to generate ideas which surpass the operational level and have implications for strategic decisions. The evaluation of the variables relevance to Spotify will be performed in section 4.3. The distribution of innovation size is analyzed, medium and large feature suggestions are further discussed. Due to there being 23 medium-size features, only select ones are outlined.

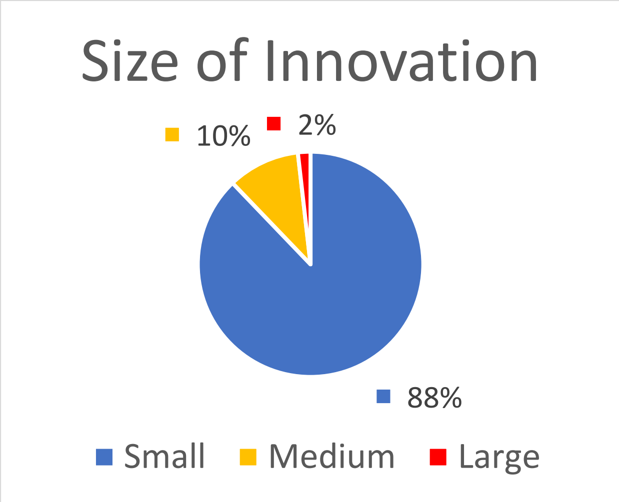

Figure 3. Graph showing the distribution of innovation size among unique ideas in the dataset.

The distribution of the size of the feature suggestions can be seen in the pie chart in Figure 3. Small suggestions are the largest category with 88%. Medium-size feature suggestions account for 10% of the total. Large innovations make up 2%. Small feature suggestions are often quick to implement and can be an improvement in the app's usability. 195 such innovative ideas were found, with some of them already having been implemented, such as the "clear queue button" and the "block users" feature (Steele, 2021).

The number of medium-size suggestions is considerable with 23 suggestions. Thus, it may be possible to scan Twitter in search of medium-size innovations. The 4th most mentioned feature suggestion in all categories is "add only friends can add playlist option", which was implemented in March 2022 (Steele, 2022). It is the most mentioned medium-size feature. Users can now allow other users access individually, meaning that only select people will be able to edit the playlist. "feedback to recommendations" is the second-most mentioned medium-size feature and it is not clear how the innovation would work. The idea is broader and leaves room for Spotify to address the need without being overly specific.

Large innovations are less common, with 4 ideas being classified as such. All except one tweet were very similar in their suggestions, the topic being user messaging and the sharing of media in the app. The topically different tweet suggests a complete redesign of Spotify's app in a simpler layout. One of the main features of social media platforms is user messaging and other direct user engagement. The notion suggests that Spotify should attempt to do so. This would be a change which strongly impacts Spotify's strategic direction as it opens another dimension to the app and Spotify's value proposition. Since there is no other music platform, as this tweet suggests: "totally wasted opportunity for @SpotifyCares not to be a social media platform. What other social media platform is there out there that's based on music? What if you could recommend songs and podcast episodes within the app to your friends list?" This may be a niche in the social media industry. Large innovations are rare, which suggests that larger amounts of data are needed for Twitter to be a source of large-size innovations.

The analysis shows that there are clear differences in distribution among the different sizes of the innovations. Twitter seems to be a good source of many small innovative ideas. Medium-size ideas can also be found, but less frequently. Large ideas are rare, but there are such ideas present in the dataset.

4.3 Relevance of feature suggestions

The previous two sections showed that Twitter is a source of ideas, this section discusses whether these feature suggestions are promising to help understand customer's needs and relevant for innovation.

Innovative ideas and customer needs-understanding are important for companies to build and sustain competitive advantage in the audio streaming industry. For Spotify, using Twitter to gain additional input could be beneficial, as the algorithm is faster and resource-conserving than traditional idea generation methods. An algorithm like the BERT model used can find needs on its own after a training dataset, at extremely low cost. Thus, it is important to determine whether tweets might be relevant to Spotify, which would make this approach commercially and innovation-wise relevant.

In the dataset, tweets were found containing suggestions which were implemented by Spotify afterwards. Moreover, the ideas were valuable to Spotify, which proves that the matching tweets would have been valuable to Spotify as well. Three ideas were found to have been implemented. There seem to be relevant feature suggestions in the dataset which have not been implemented yet as well, such as two factor authentication. Two factor authentication is a basic security feature, yet Spotify does not have it. There were 93 mentions of the word "hacked" in the dataset and the feature was mentioned 6 times. This shows that there is a clear need by users for more security since most tweets containing "hacked" implicitly show the need. Complaint analysis suggests that voiced complaints are only the tip of the iceberg, which shows that support for this need is likely higher (Stauss & Seidel, 2008). Since the innovation is already used by many other companies, it is easy-to-implement and likely to produce the desired effect of less others being hacked, thus avoiding dissatisfaction according to Kano theory (Sauerwein et al., 1996). This is due to customers being used to two-factor authentication as standard, whose non-existence will lead to dissatisfaction as it is considered a must-be requirement by them.

High fidelity streaming has been suggested frequently and Spotify is the only large audio streaming service not to have it. They made announcements of a select release of the feature until the end of 2021 but did not release it. Spotify can be considered the laggard adopter, when applying technology adoption life cycle theory (Meade & Rabelo, 2004). This feature suggestion does not provide an advantage over competitors, it a feature addition which makes it their streaming quality equal to competitors. Also, Apple is offering high fidelity as part of their standard plan (Inc., 2021), which is indicative that customers will be less willing to pay a premium for high fidelity now. TIDAL has been offering plans with high fidelity since 2017, yet their price equivalent tier offers up to 1411kbps, which is lossless audio quality (Welch, 2021). Spotify can stream at up to 320kbps, which shows that competitors are ahead with the implementation of this feature as well. Thus, it is also relevant.

Following the spamming of collaborative playlists, the feature to let only friends add was suggested 21 times. Spotify implemented such a feature, which proves that this suggestion was relevant. Furthermore, shortly after the bots started adding to random playlists, users started tweeting about it and suggested features to stop it. This shows that there is velocity in the data and that there is also a considerable amount of response.

Data velocity can be further illustrated with the triggering of a need by the release of Spotify 2021 Wrapped. Users complained about wanting to remove artist videos and not being able to share their playlists shortly after its release. The opportunity to scan explicitly for feedback to specific events seems to be present in the dataset. This is supported by two of the top ten feature suggestions only being suggested in December and January. Furthermore, in the filtration process of this research, two timeframes were omitted in the dataset as malfunctions triggered thousands of users to mention this in their tweets, which occurred immediately after the occurrence of the malfunction.

The research question of relevance is supported by the evidence found in the dataset. There is considerable velocity of the data, the is variability in the ideas and there are mentioning trends, the data seems to be valuable due to Spotify's implementation of some ideas, and veracity is also given, as there are many good feature suggestions and 5.11% of all data are needs.

4.4 BERT model performance measures

The machine learning algorithm BERT (Bidirectional Encoder Representations from Transformers) was used to try to predict tweets containing needs. There were 5000 tweets used to train the algorithm and then the algorithm tried to predict the needs in the remaining 5000 tweets.

	precision	recall	f1-score	support
0	0.98	0.97	0.97	4744
1	0.50	0.55	0.53	256
accuracy			0.95	5000
macro avg	0.74	0.76	0.75	5000
weighted avg	0.95	0.95	0.95	5000

Figure 4. BERT statistical performance measures

The algorithm's performance was measured with 4 indicators: Precision, recall, f-score and accuracy. Figure 4 contains a table of the results. In total, 256 needs were detected out of 511. Precision measures how many of the predicted needs are needs and how many tweets classified not needs are not needs. The precision scores for 'no need' and 'need' are 0.98 and 0.50 respectively. This means that 98% of tweets predicted to be 'no need' were correctly predicted. 50% of the predicted needs were needs. Thus, scanning Twitter for needs would result in 50% of tweets classified 'need' would contain usable ideas. Moreover, it is a considerable number of ideas which can be generated nearly automatically after a first training.

Recall shows how many actual needs were detected. The recall was 0.97 for 'no need' and 0.55 for 'need'. Of all tweets in the dataset containing a need, 55% were discovered by the BERT algorithm. For tweets without a need, 97% of them were classified as such.

F-score is a measure based on precision and recall, which makes it more comprehensive and significant. The formula can be seen below in Figure 5. The f1-score is 0.97 for 'no need' and 0.53 for 'need'. This means that the 53% of needs were predicted correctly according to the formula. The dataset containing needs classified as such by the algorithm is qualitatively high as observed through the similar precision and recall values, which are in optimal constellation to correctly identify needs.

$$F_\beta = \frac{(1+\beta^2) \cdot \text{true positive}}{(1+\beta^2) \cdot \text{true positive} + \beta^2 \cdot \text{false negative} + \text{false positive}}$$

Figure 5. F-score Formula

Accuracy determines how many needs and not needs were predicted correctly. According to the f1-score, 95% of all tweets were predicted accurately as containing a need or not.

The most insightful indicators are precision, recall and the f-score of "1", the classifier that a tweet contains a need. The objective of the algorithm used, is to explore how well needs are identified. Thus, the scores for "0" (no need) are less relevant to answering the research questions.

5. DISCUSSION

The research objective was to answer the question whether data from Twitter can produce relevant feature suggestions using the example of Spotify. The data was prepared using filters and then the BERT model was applied. Results showed that there was a significant percentage of tweets containing needs with 5.11%. Furthermore, the analysis determined that at least some of the ideas from tweets are relevant to Spotify. The machine learning algorithm was able to identify over half of all needs and displayed good statistical results, thus it is feasible to use an AI in this case. The dataset was found to be variable, with the presence of trends and 216 unique ideas. The tweets were found to be of high velocity as they were very responsive, such as in the case where a server outage led to thousands of tweets being made. The quality of the dataset is high, as there were 511 needs. It can be assumed that Spotify could gain insights from Twitter and would be likely to benefit.

Related research such as Kühl's paper on e-mobility with a similar approach found tweets to be a viable source of customer needs as well. Though it may be noted that his research used keywords for data collection, rather than using all data available. The research also focused on a topic, instead of a company. In addition, the algorithms used were less sophisticated. Nevertheless, the research is still useful. The algorithm used in this paper performed slightly better with an f1-score of .53 compared to .466. Moreover, the different methods of data collection mentioned lead to our dataset being broader, thus likely making it harder to identify needs. Following this, the score of .53 is even more impressive. As algorithms become more complex, it is likely that detection rates increase further, making need mining more effective. In practice, this means that need mining approaches could become more popular. Natural language processing models make it a promising source of needs and feature suggestions which can be used without in-depth technical knowledge. While 68.81% of tweets were removed from the dataset, which suggests that there needs to be some amount of work done by a human operator to make the dataset qualitative, this is not the case. The algorithm only needs a learning dataset to function, which naturally benefits from larger size, but seemed to work well with n=5000 given the result. Nevertheless, the remaining 5000 tweets were only filtered and labelled to track the algorithm performance and analyze the dataset regarding qualities. Any amount of data can be processed by the algorithm, thus making it infinitely scalable. This approach is very efficient both financially and timewise as the largest costs are the creation of the training dataset timewise and the costs for the download of data from Twitter. It could be used for many purposes, such as monitoring trends on regarding user's tweets or used as source of new ideas. In comparison to traditional methods of need collection, such as surveys and questionnaires, the BERT model is much more efficient as it uses less resources.

It seems possible to apply this model in the same way to other companies in the same industry. The most important prerequisite to successfully applying this model is a customer support account on Twitter. The main Spotify Twitter account is likely mentioned too often in unrelated tweets, which would significantly decrease the quality of the dataset. Thus, it may be possible to use this approach even in other industries, given a support account. There is no evidence which suggests that other industries would be unable to apply this model similarly. High-tech companies are assumed to benefit more from online need mining as their customers will likely be more adept at using Twitter to elicit needs. Also, success may be dependent on the presence of a company on Twitter, as users need to be aware of the company's presence.

6. LIMITATIONS AND FUTURE DIRECTION

While the research delivered promising results, there are some limitations that should be kept in mind. Firstly, the question whether the suggested features are relevant to Spotify cannot be answered with certainty, as Spotify was not requested to evaluate them. Thus, only evidence such as recently implemented features exists, which confirms that some tweets are valuable, but not how many. Also related to relevance, it may be that Spotify is only looking for totally new ideas which have not been mentioned before. It was not analyzed whether suggestions were new, as a list of all previously mentioned ideas would be needed, which was not available and too large to create. Spotify has set up a website where users can share ideas and vote for their favorites. This could mean that there is a better channel established already, which could make Twitter data less useful. Yet it does not necessarily mean that this lowers the number of users voicing their needs on Twitter. Twitter is a channel which allows needs to be elicited with low effort by users, by just typing a tweet and mentioning Spotify. Spotify's idea website though requires more effort: The user needs to become aware of the channel's existence and needs to log in with their Spotify credentials to submit the idea. Whether the existence of the idea website has an impact on where users voice their needs is unknown. While it is likely that the conclusions from this case with Spotify can be applied to all audio streaming service companies, the study was not done on other companies, thus further research is needed to ascertain. Another assumption may be that the data is generated by a random group of people, which is wrong. Twitter users are predominantly male and young compared to the general population (Mellon & Prosser, 2017; Mislove et al., 2011). The demographics of users in this dataset is unknown. Another limitation concerns the labelling of tweets as need or not needs manually. While the coder tries to be as objective as possible in their coding, the decision is still subjectively made. It may be that tweets were falsely classified by the researcher, which may skew the indicators used. Furthermore, the dataset only includes the timeframe from 26th October 2021 to 30th March 2022, which is a short time span. To validate the findings of this study further empirically in further research, a larger sample size with needs being n>=2000 over a timeframe of more than a year should be used. In addition, only English tweets were included in the dataset, thus no inference about non-English users can be made. In the future, different language groups could be compared to uncover whether there are differences in need elicitation among users. A new research approach could be taken, to apply the algorithm to a large amount of data and analyze the tweets labelled needs if they are needs and whether they are relevant. This could give insights especially for managers in practice, as a company would likely use the algorithm this way to gain the most ideas with least effort. Instead of deepening the literature specific to tech companies, the methodology may also be tested in the future in more industries, such as Kuehl did with e-mobility. Especially interesting could be low growth industries as needs might be rare, and if the viability of this approach varies.

7. CONCLUSION

The main point to draw from this research paper is that relevant customer needs were found from Twitter data using the BERT model. There were many unique ideas which also varied in their size. The algorithm performed well and more than half of all tweets identified as needs actually were a need. Human involvement in the process is still required to filter falsely classified need tweets, but only to a small degree. Twitter seems like a good source of needs, which can be extracted easily. Innovation managers may consider automatic need gathering a new tool for quickly and efficiently creating an innovation funnel but also to analyze recent trends. Traditional methods of idea gathering do not compare financially to machine learning methods, which are the future. As models get better and understand human language more naturally, gathering needs on the internet becomes even more valuable.

8. ACKNOWLEDGMENTS

I would like to thank my first supervisor Dr. Dorian E. Proksch for his communication and guidance while writing the thesis. His friendliness and feedback made this thesis a pleasure. I also thank my bachelor circle members Yucheng Chen, Selina Noorlander and Theodor Manole for the insights I could gain by reviewing our progress together, and especially Luisa Rensing for her invaluable support.

9. REFERENCES

1. Agarwal, A., Xie, B., Vovsha, I., Rambow, O., & Passonneau, R. J. (2011). *Sentiment analysis of twitter data.* Paper presented at the Proceedings of the workshop on language in social media (LSM 2011).
2. Bertschek, I., & Kesler, R. (2018). *Let the user speak: Is feedback on Facebook a source of firms' innovation?*
3. Brandtzaeg, P. B., & Følstad, A. (2016). Social media use and innovations: introduction to the special issue. *The Journal of Media Innovations, 3*(1), 1-3.
4. De Mauro, A., Greco, M., & Grimaldi, M. (2016). A formal definition of Big Data based on its essential features. *Library Review, 65*, 122-135. doi:10.1108/LR-06-2015-0061
5. Emmanuel, I., & Stanier, C. (2016). *Defining Big Data*. Paper presented at the Proceedings of the International Conference on Big Data and Advanced Wireless Technologies, Blagoevgrad, Bulgaria. https://doi.org/10.1145/3010089.3010090.
6. Erevelles, S., Fukawa, N., & Swayne, L. (2016). Big Data consumer analytics and the transformation of marketing. *Journal of Business Research, 69*(2), 897-904. doi:https://doi.org/10.1016/j.jbusres.2015.07.001.
7. Feldman, R. (2013). Techniques and applications for sentiment analysis. *Commun. ACM, 56*(4), 82–89. doi:10.1145/2436256.2436274
8. Fichman, R. G., Dos Santos, B. L., & Zheng, Z. (2014). Digital Innovation as a Fundamental and Powerful Concept in the Information Systems Curriculum. *MIS Quarterly, 38*(2), 329-A315. Retrieved from https://www.jstor.org/stable/26634929.
9. Ghani, N. A., Hamid, S., Targio Hashem, I. A., & Ahmed, E. (2019). Social media big data analytics: A survey. *Computers in Human Behavior, 101*, 417-428. doi:https://doi.org/10.1016/j.chb.2018.08.039
10. Ghasemaghaei, M., & Calic, G. (2020). Assessing the impact of big data on firm innovation performance: Big data is not always better data. *Journal of Business Research, 108*, 147-162.
11. Götting, M. C. (2021). Share of music streaming subscribers worldwide in the 1st quarter of 2021, by company. Retrieved from https://www-statista-com.ezproxy2.utwente.nl/statistics/653926/music-streaming-service-subscriber-share/.
12. IDC, Statista (2021). Volume of data/information created, captured, copied, and consumed worldwide from 2010 to 2025 (in zettabytes). In (pp. Graph): Statista.
13. Apple Inc. (2021). Apple introduces the Apple Music Voice Plan [Press release]. Retrieved from https://www.apple.com/newsroom/2021/10/apple-introduces-the-apple-music-voice-plan/.
14. Kaplan, A. M., & Haenlein, M. (2010). Users of the world, unite! The challenges and opportunities of Social Media. *Business Horizons, 53*(1), 59-68. doi:https://doi.org/10.1016/j.bushor.2009.09.003.
15. Karat, J. (1996). User centered design: quality or quackery? *interactions, 3*(4), 18-20.
16. Kuehl, N., Scheurenbrand, J., & Satzger, G. (2016). *Needmining: Identifying micro blog data containing customer needs*.
17. Lies, J. (2019). Marketing intelligence and big data: Digital marketing techniques on their way to becoming social engineering techniques in marketing. *International Journal of Interactive Multimedia & Artificial Intelligence, 5*(5).
18. Liu, B. (2012) Sentiment analysis and opinion mining. In*: Vol. 5. Synthesis Lectures on Human Language Technologies* (pp. 1-184).
19. Mathioudakis, M., & Koudas, N. (2010). *Twittermonitor: trend detection over the twitter stream.* Paper presented at the Proceedings of the 2010 ACM SIGMOD International Conference on Management of data.
20. Meade, P. T., & Rabelo, L. (2004). The technology adoption life cycle attractor: Understanding the dynamics of high-tech markets. *Technological Forecasting and Social Change, 71*(7), 667-684.
21. Mellon, J., & Prosser, C. (2017). Twitter and Facebook are not representative of the general population: Political attitudes and demographics of British social media users. *Research & Politics, 4*(3).
22. Mislove, A., Lehmann, S., Ahn, Y.-Y., Onnela, J.-P., & Rosenquist, J. (2011). *Understanding the demographics of Twitter users.* Paper presented at the Proceedings of the International AAAI Conference on Web and Social Media.
23. Pentland, B. T., & Feldman, M. S. (2008). Designing routines: On the folly of designing artifacts, while hoping for patterns of action. *Information and Organization, 18*(4), 235-250. doi:https://doi.org/10.1016/j.infoandorg.2008.08.001.
24. Roberts, D. L., & Piller, F. T. (2016). Finding the right role for social media in innovation. *MIT Sloan Management Review, 57*(3), 41-47.
25. Sauerwein, E., Bailom, F., Matzler, K., & Hinterhuber, H. H. (1996). *The Kano model: How to delight your customers.* Paper presented at the International working seminar on production economics.

26. Saura, J. R., Ribeiro-Soriano, D., & Palacios-Marqués, D. (2021). From user-generated data to data-driven innovation: A research agenda to understand user privacy in digital markets. *International Journal of Information Management, 60*, 102331. doi:https://doi.org/10.1016/j.ijinfomgt.2021.102331.

27. Stauss, B., & Seidel, W. (2008). Discovering the "customer annoyance iceberg" through evidence controlling. *Service Business, 2*(1), 33-45.

28. Steele, B. (2019, 23.12.2019). Spotify's podcast power play. *engadget*. Retrieved from https://www.engadget.com/2019-12-23-spotify-year-in-podcasts-2019.html?guccounter=1&guce_referrer=aHR0cHM6Ly93d3cuZ29vZ2xlLmNvbS8&guce_referrer_sig=AQAAACFkvfGq7UieAXsPtfWKMqQNdbEqNO9rnFPWEfPgpm-KRtBVzrWUFdB6Ty7eBYL7hxNzvXQKriIwsCB6KbT48q5y78HWBzCx1QAy4HKNTcBis2ncUTC8RBm663EF56_IxKPVOAIvF90wOi7N0sx1T6nB2psZbBbly4wYW0jNzx6u.

29. Steele, B. (2021). Spotify finally lets you block other users. Retrieved from https://www.engadget.com/spotify-block-user-140005922.html.

30. Steele, B. (2022). Spotify adds access controls to collaborative playlists. Retrieved from https://www.engadget.com/spotify-collaborative-playlists-access-controls-private-setting-130052125.html?guccounter=1&guce_referrer=aHR0cHM6Ly93d3cuZ29vZ2xlLmNvbS8&guce_referrer_sig=AQAAAN4t4fDBsSXzBHRKLFxpEfZetH_WvSnosZsw7ksYnErBHMoyXN5cNN0Izwfmymm3ElQPChcN2GKzIlrSodkJ0x05fMCi5G2rSzopuNusNAZnOq4jiAQvQSuhZWRLZfBZkM7mT6N74FVN72751hzrmvcl41jRvMLILhIN-YcDyi9e.

31. Trabucchi, D., Buganza, T., Dell'Era, C., & Pellizzoni, E. (2018). Exploring the inbound and outbound strategies enabled by user generated big data: Evidence from leading smartphone applications. *Creativity and Innovation Management, 27*(1), 42-55.

32. Von Hippel, E. (2006). *Democratizing innovation*: the MIT Press.

33. Welch, C. (2021). Tidal launches free streaming and splits HiFi into two plans. *The Verge*. Retrieved from https://www.theverge.com/2021/11/17/22787051/tidal-free-hifi-plus-announced-features-artist-payments.