MARTIN-LUTHER
UNIVERSITÄT
HALLE-WITTENBERG

Juristische und Wirtschaftswissenschaftliche Fakultät

Fachgebiet Ökonometrie und empirische Wirtschaftsforschung

Bachelorthesis

Predicting music track skips

An empirical investigation using data from Spotify

Author:

Moritz Kai Phillip Deecke (217214661)

moritz.deecke@student.uni-halle.de

Supervisor:

Prof. Dr. Christoph Wunder

Berlin, June 3, 2022

Statutory Declaration

Hereby, I declare that I have developed and written this research completely by myself and that I have not used sources or means without declaration in the text. Any external thought, content, media, or literal quotation is explicitly marked and attributed to its respective owner or author.

As of the date of submission, this piece of document and its content have not been submitted anywhere else.

Berlin, June 3, 2022

MORITZ KAI PHILLIP DEECKE

ABSTRACT

An important objective for companies is to connect customers with the right products. Where previously this was achieved through marketing campaigns, in the era of big data statistical analysis techniques have taken on this role, by finding complex patterns in data that are difficult for humans to see. This thesis studies the recommendation of music on Spotify, a key player in the digital transformation of the music industry. Using data from the Music Streaming Sessions Dataset, released by Spotify in 2018, it is investigated whether predicting content that users will like is possible by using only content-based features, such as the acousticness or key of a track. Such content-based prediction is interesting, because it can still provide valuable insights when user data is missing, for example when entering new markets, or when such data is considered private, e.g. for some hypothetical legal reason. This thesis focuses on binary logistic regression to make predictions, and while this is a simplified modeling approach, it gives direct insights into which content features are the most relevant ones for predicting user engagement. Perhaps surprisingly it is found that most audio features play a secondary role in content-based prediction, while others, such as track duration, have an important part in it.

KEYWORDS

Music streaming; product feature; logistic regression; music information retrieval; session data; big data; machine learning; recommender systems.

Table of Contents

List of Tables

1 INTRODUCTION

The growth of digital platforms like Amazon, Uber, Google or Apple, shows the increasing weight and importance that these companies occupy in our society. Kenney and Zysman (2016) argue that these companies nowadays impact "[...] how we work, socialize and create value in the economy, and compete for the resulting profits." (p. 61). To illustrate the tremendous impact of today's major digital brands, Apple for example, valued by its market capitalization, exceeds the accumulated value of all 30 brands listed in the German stock index (Sommer, 2020).

Big digital companies are often associated with platform economies, a concept that radically gained popularity during the 21st century (Bardhi and Eckhardt, 2012; Gansky, 2010). As discussed in more detail in Section 2, a key characteristic of this industry is the amount of information generated that results from the use of their services: during each interaction with the platform, users and product characteristics are stored and processed, resulting in enormous amounts of data.

From that stored data, statements are derived with techniques of machine learning and statistical analysis in order to provide customers with personalized recommendations and offers. These personalized recommendations can be separated into two main modeling concepts: collaborative filters, which infer a user's preferences based on tastes of similar users, and content-based filters, investigated in this thesis, which use similarities in product features to obtain insights.

A digital platform in the context of media and music is Spotify, which provides a highly successful collaborative streaming service in the music industry. The company, founded in 2006, allows users to listen to music from artists around the globe. Spotify has the essential characteristics of a digital platform: it mainly does not create its own content and instead gives users access to songs uploaded by artists, with over 70 million tracks available. As of 2020, Spotify has obtained over 365 million monthly active users in 178 countries with this strategy (Spotify, 2021). Spotify is currently valued at $69 billion, exceeding the valuation of other traditional media brands like Thomson Reuters, Axel

Springer, or The New York Times (Turvill, 2020).

A major challenge for Spotify is to model how a given user will react to content. Ideally, Spotify wants to provide users with content that elicits a positive reaction to drive engagement in the user base. Keeping their customers engaged with the platform is crucial to its business model: as users keep using the platform, Spotify sells them monthly subscriptions and presents third-party advertisements. In turn, this revenue allows Spotify to maintain and extend its wide product range.

This thesis investigates data from the Spotify platform to discuss some critical questions around content-based recommendation. Namely, is it possible to make predictions from content-based audio features alone? And if yes, what features are essential to predict consumer behavior?

Content-based recommendation is a subfield of research in music information retrieval (MIR), a discipline that sits between engineering and musicology, and is concerned with analyzing music and associated metadata through computational methods that integrate ideas from signal processing and machine learning.

Content-based MIR research is particularly focused on the extraction of musical information from audio data. As Dieleman et al. (2013) describe, this is a demanding task because accurately described music is hard, given the various intersecting characteristics of genres to the influences on our preferences and the underlying consumption behavior; or in simple words, users often themselves don't know what new songs they would like to listen to and what are the significant elements of a track that relate to their taste.

Content-based recommendation is particularly interesting for several reasons. First, such methods do not suffer from the so-called "cold-start" problem, a situation where no historical data is available to make predictions from, such that insightful predictions from user-level data are impossible. Second, content-based predictions are privacy-preserving as they do not use any user-related data. Accounting for privacy is a growing concern in the digital platform industry; this can be seen from the fact

that countries exert greater pressure on companies, as with the EU's new data protection regulation. Third, collaborative filters are known to sometimes falsely pick up on atypical tracks, such as covers, remixes, or intro- and outro-tracks by artists (Dieleman et al., 2013).

Content-based MIR recommendations are typically solved using a two-stage approach: features are extracted from music audio signals and afterward used as input to a regressor or classifier (Dieleman, 2015). This thesis focuses on the second part and uses fixed music features released by Spotify as part of the Music Streaming Sessions Dataset (MSSD) (Brost et al., 2019), which are described in Section 3 in more detail. These fixed features are then used as input to a logistic regression, explained in Section 4. In doing so this thesis extends existing literature with new insights into the potential and limitations of predicting user engagement only from content-based determinants, and allowing insights into which of them play decisive roles in the actual performance of compositions.

While content-based recommendations have certain advantages, such as being privacy-preserving, they should not be understood as a Swiss army knife and instead more as an additional tool for consumer behavior predictions. However, as the experiments in Section 5 show, the evidence supports that it is possible to explain track skips using track features, although there are considerable differences in the feature's influence to predict certain customer behaviors. The results, for example, show that an increase in the instrumentalness negatively impacts track popularity, while some features like the acousticness or the speechiness have a positive influence. Section 6 concludes this thesis and discusses directions for future research.

2 PREDICTING DECISIONS FROM BIG DATA

The 21st-century commercial world is faster and generates more information from market interactions than the industry of the 20th century; around 1.7 MB of data is created every second for every person on earth (Domo, 2017).

In the "ordinary" business world, statistically evaluating the effect of campaigns from e.g. mobile

banners or other traditional forms of advertising onto single individuals is almost impossible. To gain such insights, companies had to rely on market research and, hence, individual consumers' self-assessment. In the so-called era of big data, companies have been able to automate their analysis using data generated by users. Such usage data tracks users' actual behavior and decisions at that very moment, allowing a much more personalized analysis. This, in turn, gives companies valuable information on their existing and potential customers, resulting in better marketing efforts and thereby improving engagement and conversion. This shift to predominantly digital processing of data can, for example, be observed for Amazon, the most prominent retailer worldwide on the Forbes List (Forbes, 2021): a study by MacKenzie et al. (2013) for McKinsey observed a shift in Amazons' business model from mass advertising toward an extensive, data-driven personalized advertising strategy.

Such changes are fueled by enormous growth in data over the last decade: humans jointly use the Google search engine around 3.5 billion times a day. Between 2010 and 2020, the combined data interactions went up by around 5000%, and now include more complex interactions such as streaming music data or video content (Domo, 2017). The collection of data has therefore turned into a big business: studies show that the data analytics market is set to reach a combined revenue of $103 billion by 2027 and will grow much faster than anticipated in 2020 because of the Coronavirus restrictions, which saw people spending more time on digital media services (Domo, 2020; Kobielus, 2018).

The emergence of such large amounts of data is often summarized under the term "big data". While such terminology exists, the understanding of what exactly big data constitutes is still open to debate, evidenced by a poll with 154 participants from 2012 conducted by SAP which reported widely differing interpretations of the term among C-suite executives (SAP, 2012). The most important definitions are highlighted in the following.

One interpretation of the "big data" term stems from the substantial growth of transaction data, including data from customers and supply chains, as well as the needed requirements to store and archive existing data. This definition comes from a more technological interpretation of big data that involves

the so-called three Vs: volume, variety, and velocity of big data (SAP, 2012). The three Vs stand for the following: volume refers to the amount of data, variety explains the diverse data structures or types of data, and velocity refers to the speed the data is produced and processed (Gartner, 2021).

A slightly different definition is found in Kwon et al. (2014) and Russom et al. (2011), who also characterize big data analytics by volume, variety, velocity, but add additional categories like the value, meaning that only some data is useful and can be used for analysis, and veracity, implying that data has to be of a certain quality, free from too many unusable or defective entries. Contrary to this, Kitchin and McArdle (2016) define the concept more widely, stating that "[...] the 3Vs meme is actually false and misleading and along with the term itself is partially to blame for the confusion over the definitional boundaries of big data." (p. 9).

An interesting perspective is to define the term big data by the challenges it creates, as done by Mills et al. (2012): "Addressing the challenge and capturing the opportunity [of big data] requires advanced techniques and technologies to enable the capture, storage, distribution, management, and analysis of the information.". From this, a crucial portion of the overall stream of big data would appear worthless if its potential is unused, and real value is only unlocked if data is organized in the right way and then applied to guide market decisions (Gandomi and Haider, 2015; Provost and Fawcett, 2013).

According to "Cost of a Data Breach Report 2021", an annual report by IBM (2020), the unused potential of big data is estimated to cost the US economy more than around $3 trillion yearly. Meanwhile, private companies that are understood to be pioneers of the big data age, such as the video streaming company Netflix, is estimated to save around $1 billion per year on customer retention due to them unlocking potential in their data.

To enable the right kind of evidence-based decision-making from big data, organizations need efficient processes to turn the vast volumes and fast and diverse data streams into meaningful insights. A company that arguably stands for this integration like no other is Amazon, which implements big data

analytics under their "everything-under-one-roof" model, which allows them to apply their services to a wide range of products (Provost and Fawcett, 2013). From toilet paper to screws, or even the newest music album — literally any existing good can be bought in Amazon's digital department store.

The more a potential retailer knows about the customer, the better they can predict what items have a high potential of being bought. This knowledge is used to compare individual buy decisions; for instance, if two customers consume the same set of items, their tastes are probably similar. This information can be exploited to make recommendations once additional items are bought by one person but not the other. When many such interactions occur, then recommendations are typically handled algorithmically by so-called recommender systems (Koren and Bell, 2015).

In detail, this means an increased focus on personalized marketing, including customized recommendations based on a user's personal history on a platform, and using this to highlight products a user may be interested in within banner ads, or the recommendation of items the consumer may like while shopping on the company's product page. One of the biggest market advantages of personalized marketing is delivering messages to potential customers by targeting the right customer subgroups for a particular strategy. For example, a company may want to advertise a certain clothing brand to one gender, different classes of products to pre-specified income brackets, or different pharmaceuticals to certain age groups. Doing this successfully is highly lucrative; research by MacKenzie et al. (2013) for McKinsey found that almost 35% of Amazon sales were due to algorithmic targeting of customers.

Big data analytics can also help companies find better demand pricing, for example, by automatically checking prices from competitors and adjusting theirs to be more competitive. Once again, a good example of this is Amazon, which has switched its pricing model to mirror prices from its competitors. Under this strategy, customers can rely on the best price being guarantee while buying from Amazon, simplifying the decision for the customer and building up consumer trust. This trust is mandatory because customers today are much more demanding, informed, and emancipated than in the past (Babka, 2016). In fact, Amazon is changing the prices of their products around 2.5 million times a

day, which results in a price change for an average product approximately every 20 minutes (Chen et al., 2016).

While big data analytics opens up the possibility of presenting an offer that becomes steadily more individualized and thus adapted to the customer, it can additionally be used to modify processes within companies; for example, internal pricing or the allocation of marketing resources can benefit from a data-driven strategy. Another potential lies in better forecasts of sales and therefore anticipating future returns on investments. Such evidence-based estimation of sales goals is possible based on former and current performances, and the resulting predictions can help create budgets and define the financial strategies of a company (McAfee et al., 2012).

This thesis focuses on analytics for music data, and several notable examples exist for applying big data strategies to the music industry. Shazam aims to match segments of recorded music with a database using MIR analysis methods to determine the correct title of a piece of music. With more than 15 million song identifications daily, Shazam can use this data to forecast which artists will receive more attention in future years (The Guardian, 2013). Apart from that, there is a whole research area, called "Hit Song Science", that deals only with predicting potentially successful songs (Casey et al., 2008). The purpose of such research is diverse, but major labels in the music industries are keen on finding solutions to this and allocating resources to produce the most promising hits. As Casey et al. (2008) describes, firms use "symbolic information with techniques from artificial intelligence to make consultative recommendations about new releases." (p. 669). Considering that labels spend almost 5.8 billion a year on music production, music analysis could help them invest in a more target-oriented way (IFPI, 2019).

A significant difficulty in predicting future success lies in the fact that audiences expect new tracks to be novel and different from their competitors. Research by Askin and Mauskapf (2017) shows that this makes it hard to produce a new hit track by just focusing on successful predecessors, which is the core idea in most statistical predictions. Their research focused on processing Billboards Top

100 from 1958 to 2013; Askin and Mauskopf found that songs with features too similar to others have difficulty succeeding in the charts. Different research by Interiano et al. (2018) focused on determining the dynamics of successful tracks by integrating musical features in their analysis, and analyzing more than 500.000 tracks that were released between 1985 and 2015 in the UK. Their results showed promising results that music features alone could predict the success of a song quite well, with an accuracy of around 75%. While they also used standard track features comparable to the features used in this thesis, this thesis investigates a much bigger and diverse set of tracks released by Spotify as part of the MSSD. This dataset is described in detail in the next section.

3 THE MUSIC STREAMING SESSIONS DATASET

3.1 Metadata Description

Section 2 highlighted the importance of data-driven decisions in today's economic world, establishing a relationship between big data analytics and valuable research for the use of big data in the music industry. This section desribes the dataset used in this research and defines the essential features for the analysis. This will help get better clarity about the impact of each feature and allow better interpretation of the results from this data.

In November 2018, Spotify and the International Web Search and Data Mining Conference hosted a challenge to encourage new MIR research, specifically to find out how users interact with the streamed content Spotify offers. Approximately 160 million consecutive user interactions called "listening sessions" are provided in the dataset, and each session contains between 10 to 20 tracks with information on whether users decided to listen to them or to skip them. In total, the uncompressed data makes up around 420 Gigabyte of memory. The dataset is of overall high quality, so it contains no missing fields. The streaming sessions are stored in logs, where each row contains session IDs, timestamps, contextual information about the stream, the track and context IDs, and the timing and type of user interactions within each session. Each session in the dataset is defined as a listening period with no

more than 60 seconds of inactivity between consecutive tracks.

The information about users' IDs and the track IDs is fully anonymized. Moreover, the Spotify research team excluded sessions that include tracks that did not meet a minimum popularity threshold. All sessions in the dataset are sampled from Spotify Radio, personalized recommendation mixes, the user's collections, and 100 of the most popular playlists inside the Spotify streaming universe. Hence, the dataset logs contain a mix of listening sessions based on a user's personally curated collections, expertly curated playlists, contextual, non-personalized recommendations, as well as personalized recommendations.

For each track in the dataset, Spotify provides audio features and metadata describing the individual piece of music in a quantitative way. In total, they provide around 3.7 million tracks, a much larger database compared to existing research by Interiano et al. (2018), which released around half a million tracks. The audio features include technical characteristics like the acousticness, a computational measure of confidence that a track was recorded acoustically; downbeats, which estimate timestamps for the first beat of a bar in a track; or the valence, a measure of how positive a track sounds. The full list of track features is shown in Table 2.

To better understand all existing track features, Table 1 contains explicit details for all of the features. The Spotify Developer API guidelines provide the schema for the track metadata; each row has the name of a given column and a short column description.

Figure 1 visualizes differences in the occurrence of the track features. For example, bounciness is approximately symmetric around 0.5 and only has values between zero and one, whereas the loudness assumes negative values with a long-tailed shape. Another interesting observation that can be made from the figure is that most of the songs have a duration of approximately 200 seconds, with some songs being as long as 30 minutes. It should be noted that all of the studied songs were published between 1950 and 2017; see the histogram plot for the release year.

3.2 User Data

The main research question of this thesis is to evaluate the impact of individual track features on consumer behavior. The dataset contains information about whether the user skips an individual track or whether they listened to it. In particular, it contains the following information about the consumption of a track: the Boolean field "skip 1" indicates if a given track was played very briefly or skipped immediately. Next, "skip 2" and "skip 3" tell whether a skip occurred after a relatively brief hearing session or in the middle of tracks, however Spotify gives no precise definition for these two fields.

While in principle more advanced predictions are possible that predict the different moments of skipping to determine which features are most useful at what time of each track, for simplicity reasons the research here focuses on whether tracks are listened to in full. Models are therefore learned to predict whether a track was skipped at any point during the session, and this information is also available in the official release by Spotify, denoted as "not skipped". The distribution of skipping to not-skipping is approximately 1:2, such that the dataset is relatively balanced with respect to this information. All other parameters in the user dataset, shown in Table 3, are not used in this research, which focuses on content-based prediction instead.

Once information about the skipping of tracks has been organized, prediction models, such as logistic regression, a popular model in the statistical literature (Cameron and Trivedi, 2005; Winkelmann and Boes, 2006), can be learned to predict the likelihood of a track being listened to. This model is explained in the next section.

4 METHODS

Here the methodology is presented that is needed to evaluate the main research question: whether the popularity of a track can be predicted from audio features alone. Section 4.1 introduces logistic regression, the model used in this thesis, followed by Section 4.2 which details how precisely model

parameters are found. Section 4.3 explains what is needed to carry out the evaluation of the model.

4.1 Logistic Regression

To identify the effect of all features on the skip rate, we apply a binary logistic regression approach to estimate the probability of positive engagements. To learn the model, the skip information from the previous section is used to model the binary outcomes as follows:

$$y = \begin{cases} 1 & \text{if track is not skipped,} \\ 0 & \text{if track is skipped.} \end{cases} \tag{1}$$

The dependent variable, whether a track is skipped or not, can therefore only take one of the two values. This can be modeled by a probability p, which estimates how likely either associated outcome is observed. As such:

$$\begin{aligned} P(y = 1) &= p, \\ P(y = 0) &= 1 - P(y = 1) = 1 - p. \end{aligned} \tag{2}$$

In a regression model, the probability to observe either outcome is computed from a set of mutually independent factors, which are parametrized as $\beta_1, \ldots, \beta_K$, one parameter for each feature value $x_1, \ldots, x_K$ in the data (in this thesis, these are the audio features described in Section 3). Moreover, a bias term β_0 is added to scale the data.

Once these are found via some optimization routine (how this is done is described in the next section), the parameters can be used to estimate a particular outcome. Given some features $(x_1, \ldots, x_K) \in \mathbb{R}^K$, this is computed as such:

$$x_1\beta_1 + \cdots + x_K\beta_K + \beta_0 = \sum_{k=1}^{K} \beta_k x_k + \beta_0. \tag{3}$$

The estimates made in equation (3) correspond to an ordinary least squares (OLS) model and can assume any value. In this thesis, we instead want to estimate probabilities, that take on values between zero and one. This is the major difference between OLS and binary logistic regression: while for an

OLS regression, the predicted values can be infinite, in this research, we aim to predict the probability of a binary outcome.

A simple solution to this problem is to use OLS to estimate logits instead, and subsequently use a sigmoid activation $F(t) = e^t/(1+e^t)$ to transform the predicted output. The final prediction then lies between zero and one:

$$p = F\Big(\sum_{k=1}^{K} \beta_k x_k + \beta_0\Big). \tag{4}$$

Sometimes it is necessary to make binary model predictions; in particular, this is needed for some quantitative metrics that are used to evaluate models in this thesis (see Section 4.3.2). In this case, one chooses a threshold τ, for example 0.5, and if the prediction p_n for an example exceeds this value, it is assumed that the prediction y_n^* from the model was positive, while if $p_n < \tau$, then $y_n^* = 0$.

Gaining insights into the individual relevance of the parameters is a key research question in this thesis. Crucially, it is possible to read off the influence of each regression parameter: when β_k has a large positive value, it has a positive influence on the likelihood of y_n^* being positive. When negative, this reduces p_n. And if $\beta_k \approx 0$, then the parameter does not significantly influence the outcome.

4.2 Finding the Parameters

Once features have been selected to make predictions from, we need to find out which individual values each of the parameters $\beta_0, \ldots, \beta_K$ of the model are well suited. This section describes how these can be found for large datasets like the one studied in this thesis. For this, it introduces the loss (Section 4.2.1) and stochastic gradient descent (Section 4.2.2).

4.2.1 Binary Cross Entropy

Whether a model can fit the data well can be determined by comparing true measurements of whether users interacted with songs, the so-called labels y_n, to the prediction p_n by the model. This comparison can be done via a loss function, which is small when predictions are close to the ground-truth, and large if not. Because the problem is binary, the labels can either take on $y_n = 0$ or $y_n = 1$, and the

binary cross entropy (BCE) is a suitable loss. For an example x_n this is defined as:

$$\text{BCE}(y_n, p_n) = -[y_n \log(p_n) + (1 - y_n) \log(1 - p_n)]. \tag{5}$$

Before the next section details how BCE is used to find the parameters $\beta_0, \ldots, \beta_K$, we look at some examples to better understand how BCE works. Ultimately, the loss function is supposed to return a higher loss when an inaccurate prediction is made, and smaller values closer to zero when the prediction is accurate. This can be seen for two extreme outcomes: first, assume that the actual value and the predicted value are the exact same, and $y_n = p_n$. Regardless of whether $y_n = 0$ or $y_n = 1$, when inserting them together with $p_n = 0$ or $p_n = 1$ into eq. (5), this results in a zero of the loss, indicating that there is no difference between the predicted and the actual value, and that the model parameters perfectly fit the actual observation:

$$\text{BCE}(1, 1) = \overbrace{1(-\log(1))}^{0} + \overbrace{(1-1)}^{0}\overbrace{(-\log(1-1),}^{-\log(0)=\infty} = 0, \tag{6}$$

$$\text{BCE}(0, 0) = \underbrace{0(-\log(0))}_{0} + \underbrace{(1-0)}_{1}\underbrace{(-\log(1-0)}_{0} = 0. \tag{7}$$

Next, we investigate the opposite case. When $y_n = 1 - p_n$, then the loss function is supposed to assume large values, since the prediction is completely wrong. In this case, we get $y_n = 1$ together with $p_n = 0$, as well as $y_n = 0$ and $p_n = 1$. In this scenario, the predicted parameters do not fit the observations well, and the loss function assumes infinite values:

$$\text{BCE}(1, 0) = \overbrace{1(-\log(0))}^{-\log(0)=\infty} + \overbrace{(1-1)}^{0}\overbrace{(-\log(1-0)}^{0} = \infty, \tag{8}$$

$$\text{BCE}(0, 1) = \underbrace{0(-\log(1))}_{0} + \underbrace{(1-0)}_{1}\underbrace{(-\log(1-1)}_{-\log(0)=\infty} = \infty. \tag{9}$$

Of course, looking at an individual example y_n and its prediction p_n is not particularly informative. Instead, a better estimate of the goodness of the fit is determined by evaluating BCE for multiple examples. To get the most accurate loss value, one would compute an average over all examples $x_1, \ldots, x_N$ available in the dataset:

$$\text{BCE}(y_1, \ldots, y_N, p_1, \ldots, p_N) = -\frac{1}{N} \sum_{n=1}^{N} \big[y_n \log(p_n) + (1 - y_n) \log(1 - p_n)\big]. \tag{10}$$

Minimizing the BCE loss is closely related to the maximization of the likelihood of the data under some Bernoulli model with parameters $\beta_0, \ldots, \beta_K$, for which:

$$L(\beta_0, \ldots, \beta_K) = \prod_{n=1}^{N} p(y_n = 1|\beta_0, \ldots, \beta_K)^{y_n} p(y_n = 0|\beta_0, \ldots, \beta_K)^{1-y_n} = \prod_{n=1}^{N} p_n^{y_n}(1-p_n)^{1-y_n}, \quad (11)$$

and taking the logarithm of this:

$$\log L = \sum_{n=1}^{N} \left[y_n \log p_n + (1-y_n) \log(1-p_n) \right] = -N\text{BCE}. \quad (12)$$

When N is very large, as is the case in this thesis, evaluating BCE for all examples requires a lot of time and becomes unfeasible. The following section describes an alternative implementation, which approximates the BCE loss with random examples from the dataset, and uses a gradient descent technique to optimize the model.

4.2.2 Stochastic Gradient Descent

For accurate predictions, suitable model parameters $\beta_0, \ldots, \beta_K$ are needed. As the previous section established, one way to test the suitability of parameters is to evaluate samples using the BCE loss. While computing the BCE loss from all examples in the dataset is not realistic for the large dataset studied in this thesis, one option is to use a few examples to approximate the BCE. For this, we fix a batch size M that is small compared to N, and compute predictions $p_1, \ldots, p_M$ for all samples $x_1, \ldots, x_M$ in the batch, and compare this to the $y_1, \ldots, y_M$ using the BCE loss.

This idea is the foundation of stochastic gradient descent (SGD): batches are sampled from the data over and over again, and each time the BCE is computed (Ruder, 2016). One key ingredient that is still missing is a connection that reduces the loss between the steps of this process: this is achieved via a technique called gradient descent, which computes gradients for the model parameters at every step, and takes small steps in the negative direction of gradients so that the loss becomes smaller and smaller.

For this, it helps to consider an example of a climber that wants to descent from some mountain into a valley. The valley here corresponds to the minimum of the loss function we are looking for in the

optimization of the $\beta_0, \ldots, \beta_K$. Even when the visibility is poor, the climber can look around and simply follow the downhill direction of the mountain he is on. If he continues to do this for a long time, he should be able to make it to the valley.

In the same way, SGD determines the loss at every step, then computes gradients, and moves along the negative direction, i.e. the downhill direction. SGD is also called a first-order optimization algorithm, which takes only the first derivative (of the loss function) into account when performing the updates on the parameters.

To return to the example of the climber, one problem is that the mountain can be surrounded by multiple valleys, with no downhill paths connecting them. In this case, there is no guarantee that the climber will arrive in any particular valley, only that he makes it some distance downhill. In the context of optimization, this means SGD can only find local optima.

There are some important steps to consider for SGD to work well to achieve a suitable optimization, such as choosing a small learning rate, which is used to reduce the gradient steps that are computed from the loss (Ruder, 2016). Similarly, it can help to run the optimization multiple times since the model can end up in local optimums, and better optimums may exist elsewhere.

Because of the size of the dataset, choosing a batch size is necessary. This is also where SGD gets the "stochastic" in its name from, due to the random sampling of examples from the dataset, which are then used to estimate BCE and compute the gradients. While there is no perfect choice for M, it should not be too small because then the loss value will be averaged poorly. On the other hand, it should not be too large because then computations become too expensive.

Besides these points, SGD provides some major benefits; for example, it requires very little compute power, and is easy to implement. At the same time, while it can be a problem that SGD terminates at an optimum that is not global, in the experiments of this thesis, there was no noticeable difference in model performance when repeating model training.

4.3 Model Evaluation

After explaining BCE (Section 4.2.1) and SGD (Section 4.2.2), here we focus on how data is organized to optimize models, and how they can be evaluated. Section 4.3.1 explains the concept of train and test splits, while Section 4.3.2 details how performance is measured.

4.3.1 Train and Test Splits

To enable evaluation, a standard strategy in statistical machine learning splits the dataset into two separate sets, the so-called train and test splits. The train set is used to optimize a model, and the test set is used to evaluate its predictions. The idea behind splitting training and testing data is that it gives insights into how well a model can handle new data: the training data represents data that the model has seen and, if trained on it long enough and with sensible optimization settings, knows how to predict. The real challenge however is to make accurate predictions for new data. For a real-world example, consider the Spotify application: here, the training data is accumulated as users use the platform. Once models are learned from this, it has to be seen how well this performs for new user visits, which is essentially nothing but a large test set. In the case of the data used in this research, datasets are already split into a train set and a test set, a manual split is therefore not necessary.

4.3.2 Measuring Performance

After the separation of the train and test set, the test portion of the data can be used to measure the model's performance. This process allows checking whether the fitted model is accurate and whether it makes good predictions on data it has previously not seen, which is after all the main purpose of a predictive model.

As explained in Section 4.1, when an observation is positive for an example and $y_n = 1$, then this indicates that a track was not skipped. When it is negative instead and $y_n = 0$, then it was skipped. Under a perfect fit of the data, the number of positive and negative predictions of the model would exactly correspond to the number of positive and negative ground-truth binary labels of the test set.

However, not only do the numbers need to match: it is also important to assign the right prediction to the right example.

Since a perfect model is unlikely to be produced under normal analysis conditions, it is necessary to analyze misstatements made by the model. Such incorrect predictions can occur in one of two ways. One type of error is called a false positive prediction, where the true value is negative y_n, but the assigned probability p_n exceeds the threshold for a positive prediction τ, in which case the predicted value is $y_n^* = 1$. The second type of error that can occur is a false negative, where $y_n = 1$, but the prediction is smaller than τ. When $y_n = 0$ and $y^* = 0$, or $y_n = 1$ and $y^* = 1$, a prediction was accurate.

These four cases motivate the concept of the confusion matrix, which can be used to easily quantify the accuracy of the model: this consists of a two by two matrix that contains four values associated with the accurate predictions and errors of a binary classifier. The four outcomes of the confusion matrix allow easily reading off whether a prediction was accurate (in this case, it sits on the diagonal of the matrix) or an error was made (these are found on the off-diagonals).

Table 4 shows all possible outcomes in a binary prediction: the true positive outcome (TP), understood as correctly predicted positive outcomes when $y_n = y_n^* = 1$; the true negative outcome (TN), being the number of correctly predicted negative outcomes $y_n = y_n^* = 0$, as well as the misstatement of the models, known under the terms false positive (FP) and false negative (FN), where $1 = y_n \neq y_n^* = 0$, or $0 = y_n \neq y_n^* = 1$, respectively.

Some important measures can be derived from the confusion matrix. First and foremost the accuracy is calculated as:

$$\text{ACC} = \frac{\text{TN} + \text{TP}}{\text{TN} + \text{TP} + \text{FN} + \text{FP}}, \tag{13}$$

which divides the number of all correct predictions TN and TP by the total number of observations in the dataset. The best obtainable ACC is one, with zero being the worst. A related metric is the error

rate (ERR):

$$\text{ERR} = \frac{\text{FN} + \text{FP}}{\text{TN} + \text{TP} + \text{FN} + \text{FP}}, \tag{14}$$

which can also be calculated directly from the accuracy by $\text{ERR} = 1 - \text{ACC}$.

While the confusion matrix is an interesting and intuitive way to analyze a binary classification model, it depends on how the threshold τ is chosen to produce the predictions y_n^*. If τ is small, more predictions become positive, and if it is large, many predictions are negative. Other metrics do not depend on the threshold, for example the area under the ROC-AUC curve, which is available in sklearn (Pedregosa et al., 2011). This metric evaluates whether the scores of the model p_n are correctly ordered: if all scores of positive examples are larger than the scores of negative ones, the ROC-AUC achieves an ideal value of 1. When the scores p_n are completely random, the ROC-AUC becomes 0.5.

5 EXPERIMENTS

Previous sections explained the main methodological ingredients used in this thesis: the logistic regression model (Section 4.1) and how the parameters of the model are found (Section 4.2) by combining the BCE loss with SGD. The next step is to combine these elements in the experiments, which let us investigate the main research question of this thesis: whether it is possible to predict the popularity of tracks on the Spotify platform from only the track features.

For the experiments, models are trained on the official train and test splits using SGD to optimize the BCE loss with a batch size of $M = 32$ and a learning rate of 0.0001. Because there is sufficient data in the training set, we simply sample randomly from this and stop training once the loss does no longer decrease, which was approximately the case after processing 1.5 million batches. Models and optimization are implemented via Pytorch (Paszke et al., 2019). Moreover, all features x_k are normalized to have zero mean and standard deviation:

$$x_{k,\text{normalized}} = \frac{x_k - \mu_k}{\sigma_k}, \tag{15}$$

where μ_k is the mean of each feature in the data, and σ_k is its standard deviation. This was found to be extremely important for robust optimization of the models and also enables comparisons between the features and their relative importance in Section 5.1.

Results for the logistic regression at a threshold of $\tau = 0.4$ are shown in Table 6. The model obtains an ACC of 0.641 and thus an ERR of 0.359. Accurate predictions appear more difficult for skipped examples, likely because there is fewer of these overall. The ACC is improved for larger τ, for example 0.5, but the predictions on negative examples become worse by this, pointing to the importance of the choice of threshold. However, as the ROC-AUC score of 0.549 shows (which is independent of the threshold), the model obtains a better than random prediction from the audio features.

5.1 Feature Importance

Here we investigate the relationship between the endogenous variables, the audio-level track features, and the dependent variable, whether a track was skipped. For this, Table 6 shows the values for every individual parameter that was used in the logistic regression. These have been averaged over five repetitions, and the standard deviation is reported alongside the parameter means.

Some important observations can be made here about the individual track features and their importance. Generally, if a parameter β_k has a value different from zero, then on average it has an impact on the prediction. One of the most influential features is the bounciness with $\beta_5 = 0.0637$, which means an increase in the bounciness of a track results in a track being less likely to be skipped. Similarly, the us_popularity_estimate with $\beta_1 = 0.0643$ has a large positive influence on user engagement. Other features that have a positive influence are, for example, acousticness, the key of a track, or the acoustic_vector_1, although interpreting this technical parameter is less straightforward.

In contrast to the above factors, which have a positive effect on the likelihood of not being skipped, other parameters make skipping more likely. For example dyn_range_mean with $\beta_7 = -0.0257$. Interestingly, the duration of a track has a negative influence on average with $\beta_1 = -0.1362$, indicating

that very long tracks are more likely to be skipped at some point, potentially due to some repetitive elements in them. Other examples of negative factors are instrumentalness or energy of a track.

These results are mostly in line with findings by Poor (2020). Although this study uses more advanced models for the analysis, it also highlighted a beneficial influence of features like acousticness or speechiness, and a negative impact of features such as the instrumentalness. This means that songs that include vocals and acoustic elements are generally considered to have a higher popularity. On the other hand, a song containing excessive instrumental elements, such as songs reissued as instrumental tracks, tend to be skipped more frequently by users. These results also correspond to classical findings around the potential of Pop music, which established higher market acceptance for moderately complex music structures (Hargreaves, 1984).

Lastly, it is also important to identify such features in a track that do not have any real influence: the regression parameters for flatness, liveness, loudness, mechanism or organism, as well as some acoustic vectors like the fourth or seventh, are all very close to zero. This lets us conclude that high or low levels of these features will neither positively nor negatively affect a song's performance.

6 CONCLUSION

This thesis carried out model analysis over tracks available on the Spotify platform. The main research question was to determine whether it is possible to predict if tracks are skipped or not using only the track features associated with each tune, and to identify which content-based features are the essential features in predicting consumer preferences.

To achieve this, an empirical study around engagement with music data was carried out using Spotify's Music Streaming Sessions Database, a large dataset with around 4 million unique tracks. To understand the dataset in the right context, this thesis discussed the importance of big data analytics in the commercial world and some beneficial aspects of data-driven decisions for companies and consumers. This lead to a discussion of the methods needed to analyze such data, such as loss functions

and gradient-based optimization.

The experiments used logistic regression to build a tool for binary predictions of track engagement, and the resulting models were carefully analyzed through the use of different metrics. Although the model did not score the highest obtainable accuracy, this research can nonetheless help contribute to a better understanding of the individual components that should be used in content-based music recommendation. In particular in some special market conditions content-based models can be helpful, for example when entering a new market. In such a situation, there would be no user data available to make predictions from, and having a reliable content-based model becomes very important. In future research that puts additional focus on this problem, it would be interesting to look at performance when splitting the data so that no track from the train set appears in the test set.

It is also important to mention that the track features in the study are sometimes relatively abstract, for example the acoustic vectors. These are of little help for music producers and songwriters, as such features are constructed from relatively complex engineering principles, making them less interpretable than more intuitive features like track duration or acousticness. Although acoustic vectors may work well for statistical models, future work could investigate excluding such relatively cryptic features and instead using only the quantities more known to musicians, such as instrumentalness, speechiness. Other options would be to develop content-based features that are more human-centric, for example via crowdsourcing.

From a pure machine learning perspective, further analysis using more complex model classes, for example deep learning or other non-linear approaches, are natural steps to extend this work. Further work could also predict consumer preferences by including more conditional information, and build recommendation systems guided by decisions made by users over time (Hansen et al., 2019). This allows better recommendations with suitable track features that fit potential time-driven patterns, for example a user might enjoy acoustic tracks, followed by something instrumental afterward.

Ultimately it is hard to formulate a one-fits-all recommendation model just from track features alone. Music is a social phenomenon, and an individual's taste is, therefore, a vital factor for the success of a song; for example, the preferences of a Metal fan will differ from that of German Schlager fans or people listening to Pop music. Therefore, the individual preferences or music tastes specific to each user will remain important factors, and by definition, cannot be determined from the track alone. For content-based model predictions, this makes investigating various genre types individually an interesting aspect to consider in future research.

REFERENCES

Askin, N., & Mauskapf, M. (2017). "What makes popular culture popular? Product features and optimal differentiation in music." *American Sociological Review*, *82*5, 910–944.

Babka, S. (2016). Der einfluss von social media auf den vertrieb. *In social media für führungskräfte* (pp. 133–144). Wiesbaden, DE: Springer Fachmedien Wiesbaden.

Bardhi, F., & Eckhardt, G. M. (2012). "Access-based consumption: The case of car sharing." *Journal of Consumer Research*, *39*4, 881–898.

Brost, B., Mehrotra, R., & Jehan, T. (2019). "The music streaming sessions dataset." *The World Wide Web Conference*, 2594–2600.

Cameron, A. C., & Trivedi, P. K. (2005). *Microeconometrics: Methods and applications*. Cambridge University Press.

Casey, M. A., Veltkamp, R., Goto, M., Leman, M., Rhodes, C., & Slaney, M. (2008). "Content-based music information retrieval: Current directions and future challenges." *Proceedings of the IEEE*, *96*4, 668–696.

Chen, L., Mislove, A., & Wilson, C. (2016). "An empirical analysis of algorithmic pricing on amazon marketplace." *Proceedings of the 25th International Conference on World Wide Web*, 1339–1349.

Dieleman, S. (2015). *Learning feature hierarchies for musical audio signals* (Doctoral dissertation). Ghent University.

Dieleman, S., van den Oord, A., & Schrauwen, B. (2013). "Deep content-based music recommendation." *Advances in Neural Information Processing Systems*, *26*.

Domo. (2017). 5.0 data never sleeps. Retrieved 2021 July 22, from https://www.domo.com/learn/infographic/data-never-sleeps-5

Domo. (2020). 8.0 data never sleeps. Retrieved 2021 July 12, from https://www.domo.com/learn/infographic/data-never-sleeps-8

Forbes. (2021, May 13). Global 2000: How the world's biggest public companies endured the pandemic. Retrieved 2021 Aug. 12, from https://www.forbes.com/lists/global2000/#22b4a7ff5ac0

Gandomi, A., & Haider, M. (2015). "Beyond the hype: Big data concepts, methods, and analytics." *International Journal of Information Management*, *35*2, 137–144.

Gansky, L. (2010). *The mesh: Why the future of business is sharing*. London, UK: Penguin.

Gartner. (2021, Aug. 8). Glossary: Big Data. Retrieved 2021 Aug. 11, from https://www.gartner.com/en/information-technology/glossary/big-data

Hansen, C., Hansen, C., Alstrup, S., Simonsen, J. G., & Lioma, C. (2019). "Towards data science: Modelling sequential music track skips using a multi-RNN approach." *arXiv preprint arXiv:1903.08408*.

Hargreaves, D. J. (1984). "The effects of repetition on liking for music." *Journal of Research in Music Education*, *32*1, 35–47.

IBM. (2020). Data breach report 2020. Retrieved 2021 July 21, from https://www.ibm.com/security/digital-assets/cost-data-breach-report/#/

IFPI. (2019). - international federation of the phonographic industry - global music report - the industry in 2019. Retrieved 2021 Aug. 7, from https://www.ifpi.org/wp-content/uploads/2020/07/Global_Music_Report-the_Industry_in_2019-en.pdf

Interiano, M., Kazemi, K., Wang, L., Yang, J., Yu, Z., & Komarova, N. L. (2018). "Musical trends and predictability of success in contemporary songs in and out of the top charts." *Royal Society Open science*, *5*5, 171274.

Kenney, M., & Zysman, J. (2016). "The rise of the platform economy." *Issues in Science and Technology*, *32*3, 61.

Kitchin, R., & McArdle, G. (2016). "What makes big data, big data? exploring the ontological characteristics of 26 datasets." *Big Data & Society*, *3*1, SAGE Publications UK: London, England.

Kobielus, J. (2018, Feb. 28). Wikibon's 2018 big data analytics trends and forecast. Retrieved 2021 Aug. 14, from https://wikibon.%20com/wikibons-2018-big-data-analytics-trends-forecast/

Koren, Y., & Bell, R. (2015). Advances in collaborative filtering. *Recommender systems handbook* (pp. 77–118). Wiesbaden, DE: Springer Fachmedien Wiesbaden.

Kwon, O., Lee, N., & Shin, B. (2014). "Data quality management, data usage experience and acquisition intention of big data analytics." *International Journal of Information Management*, *34*3, 387–394, Elsevier.

MacKenzie, I., Meyer, C., & Noble, S. (2013). "How retailers can keep up with consumers." *McKinsey & Company*, *18*, 1.

McAfee, A., Brynjolfsson, E., Davenport, T. H., Patil, D., & Barton, D. (2012). "Big data: The management revolution." *Harvard Business Review*, *90*10, 60–68.

Mills, S., Lucas, S., Irakliotis, L., Rappa, M., Carlson, T., & Perlowitz, B. (2012). "Demystifying big data: A practical guide to transforming the business of government." *TechAmerica Foundation, Washington*.

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., et al. (2019). "Pytorch: An imperative style, high-performance deep learning library." *Advances in Neural Information Processing Systems*, *32*, 8026–8037.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., et al. (2011). "Scikit-learn: Machine learning in python." *The Journal of Machine Learning Research*, *12*, 2825–2830.

Poor, A. (2020, Feb. 20). Retrieved 2021 Aug. 21, from https://towardsdatascience.com/predicting-spotify-track-skips-49cf4a48b2a5

Provost, F., & Fawcett, T. (2013). "Data science and its relationship to big data and data-driven decision making." *Big Data*, *1*1, 51–59.

Ruder, S. (2016). An overview of gradient descent optimization algorithms. Retrieved 2021 Aug. 12, from https://ruder.io/optimizing-gradient-descent/

Russom, P. et al. (2011). "Big data analytics." *TDWI Best Practices Report, Fourth Quarter*, *19*4, 1–34.

SAP. (2012, June 26). SAP News: Small and Midsize Companies Look to Make Big Gains With "Big Data," According to Recent Poll Conducted on Behalf of SAP. Retrieved 2021 July 28, from https://news.sap.com/2012/06/small-and-midsize-companies-look-to-make-big-gains-with-big-data-according-to-recent-poll-conducted-on-behalf-of-sap/

Sommer, U. (2020, Jan. 29). Handelsblatt: Apple ist jetzt wertvoller als alle Dax-Unternehmen zusammen. Retrieved 2021 Aug. 8, from https://www.handelsblatt.com/finanzen/anlagestrategie/trends/boersenwert-apple-ist-jetzt-wertvoller-als-alle-dax-unternehmen-zusammen/25484872.html?ticket=ST-3386104-5G0DCxCUlOYOdOoE3HKM-ap2

Spotify. (2021, June 30). Company Info. Retrieved 2021 Aug. 8, from https://newsroom.spotify.com/company-info/

The Guardian. (2013, Dec. 10). How shazam uses big data to predict music's next big artists. Retrieved 2021 Aug. 12, from https://www.theguardian.com/technology/datablog/2013/dec/10/shazam-big-data-prediction-breakthrough-music-artists

Turvill, W. (2020, Dec. 20). The news 50: Tech giants dwarf rupert murdoch to become the biggest news media companies in the english-speaking world. Retrieved 2021 Aug. 22, from https://www.pressgazette.co.uk/biggest-media-companies-world/

Winkelmann, R., & Boes, S. (2006). *Analysis of microdata*. Wiesbaden, DE: Springer Science & Business Media.

APPENDIX

Table 1: Track Features

Column Name	Column Description
Track ID	Unique identifier for the track played. Being linked with Track ID in the user logs
Release Year	Estimate of the year the track was released
US Popularity Estimate	Estimate of the US popularity percentile of the track as of October 12, 2018
Danceability	Danceability describes how suitable a track is for dancing based on a combination of musical elements including tempo, rhythm stability, beat strength, and overall regularity. A value of 0.0 is least danceable and 1.0 is most danceable
Energy	Represents a perceptual measure of intensity and activity. Typically, energetic tracks feel fast, loud, and noisy. For example, death metal has high energy, while a Bach prelude scores low on the scale
Tempo	The overall estimated tempo of a track in beats per minute (BPM). In musical terminology, tempo is the speed or pace of a given piece, and derives directly from the average beat duration

Continued on next page

Table 1 – continued from previous page

Column Name	Column Description
Loudness	The overall loudness of a track in decibels (dB). Loudness values are averaged across the entire track and are useful for comparing relative loudness of tracks
Speechiness	This detects the presence of spoken words in a track. The more exclusively speech-like the recording (e.g. talk show, audio book, poetry), the closer to 1.0 the attribute value
Instrumentalness	Predicts whether a track contains no vocals. "Ooh" and "aah" sounds are treated as instrumental in this context. Rap or spoken word tracks are clearly "vocal"
Liveness	Detects the presence of an audience in the recording. Higher liveness values represent an increased probability that the track was performed live
Acousticness	A confidence measure from 0.0 to 1.0 of whether the track is acoustic.
Key	The estimated overall key of the track. Integers map to pitches using standard Pitch Class notation. E.g. 0 = C, 1 = C/D, 2 = D, and so on

Continued on next page

Table 1 – continued from previous page

Column Name	Column Description
Mode	Indicates the modality (major or minor) of a track, the type of scale from which its melodic content is derived. Major is represented by 1 and minor is 0
Duration	The duration of the track in seconds
Time Signature	An estimated overall time signature of a track. The time signature (meter) is a notational convention to specify how many beats are in each bar (or measure)
Acoustic Vectors	Proposed by Spotify Research Team, see http://papers.nips.cc/paper/5004-deep-content-based-

Source: Based on Brost et al. (2019)

Table 2: Track Features from Metadata

track id	popularity	bounciness
energy	key	mechanism
speechiness	valence	duration
acousticness	danceability	flatness
liveness	mode	tempo
acoustic vector 1-8	release year	beat strength
dyn range mean	instrumentalness	loudness
organism	time signature	

Source: Based on Brost et al. (2019)

Table 3: User Features from Metadata

session id	session position	session length
track id clean	skip 1	skip 2
skip 3	not skipped	context switch
no pause before play	short pause before play	long pause before play
hist user behavior n seekfwd	hist user behavior n seekback	hist user behavior is shuffle
date	premium	context type
hour of day	hist user behaviour reason start	hist user behaviour reason end

Source: Based on Brost et al. (2019)

Table 4: Confusion Matrix

		Actual y_n		
		not skipped	skipped	Total
Predicted y^*	not skipped	TP	FP	TP+FP
	skipped	FN	TN	FN+TN
	Total	TP+FN	FP+TN	N

Source: Own Illustration

Table 5: Confusion Matrix Results

		Actual y_n		
		not skipped	skipped	Total
Predicted y_n^*	not skipped	149 523 367	74 885 089	224 408 456
	skipped	16 450 724	13 217 475	29 668 199
	Total	165 974 091	88 102 564	254 076 655

Source: Own Illustration

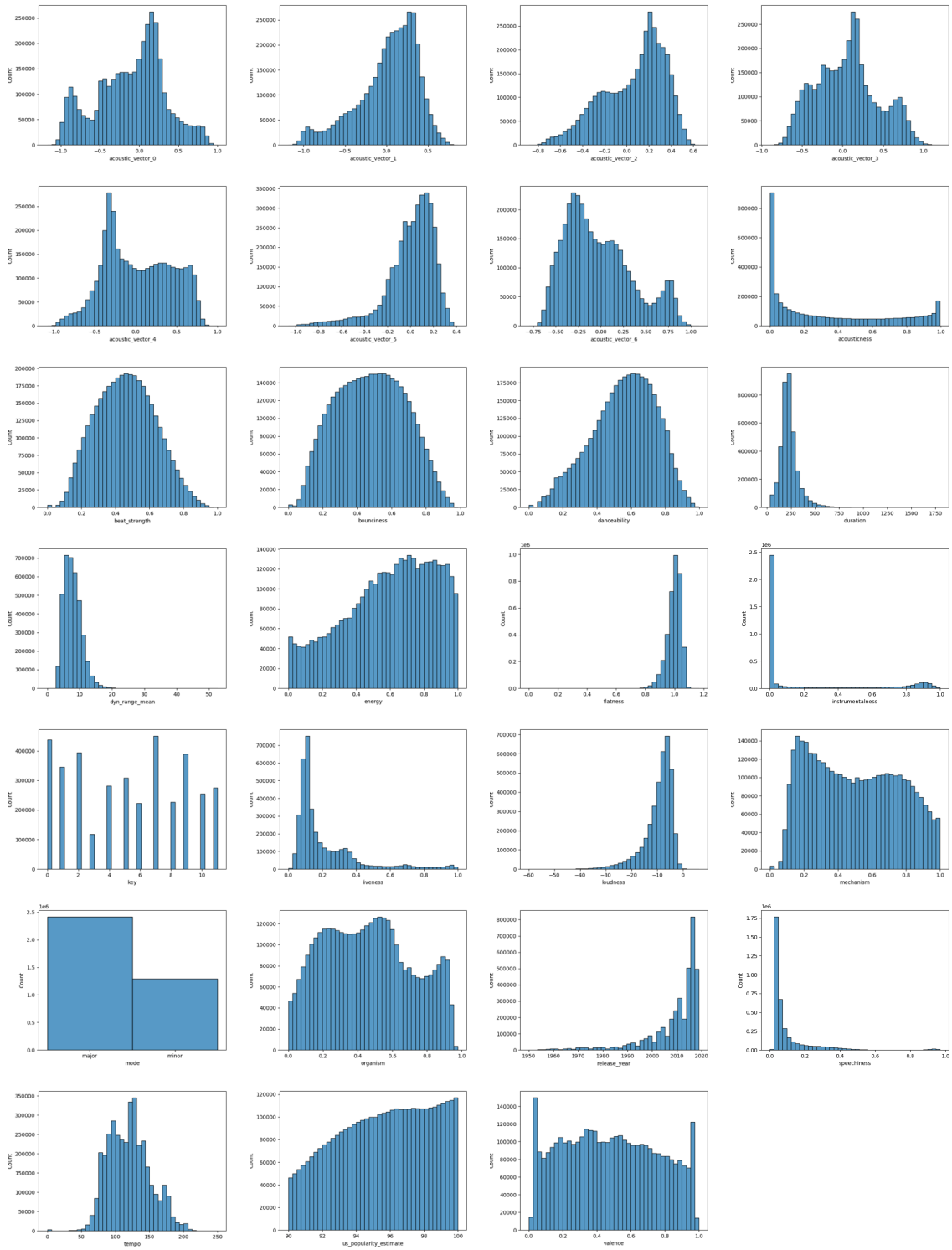
Table 6: Regression Table

	Dependent Variable		**Dependent Variable**
x_0 (Bias Term)	-0.6554	x_{15} (mode)	0.0264
	(0.00160)		(0.01833)
x_1 (duration)	-0.1362	x_{16} (organism)	0.0035
	(0.00176)		(0.00143)
x_2 (us_popularity_estimate)	0.0643	x_{17} (speechiness)	0.0281
	(0.00166)		(0.01115)
x_3 (acousticness)	0.0394	x_{18} (tempo)	0.0061
	(0.00146)		(0.00076)
x_4 (beat_strength)	0.0157	x_{19} (time_signature)	0.0016
	(0.01407)		(0.00208)
x_5 (bounciness)	0.0637	x_{20} (valence)	-0.0044
	(0.01306)		(0.00118)
x_6 (danceability)	-0.0083	x_{21} (acoustic_vector_0)	0.0044
	(0.00946)		(0.00116)
x_7 (dyn_range_mean)	-0.0257	x_{22} (acoustic_vector_1)	0.0879
	(0.00411)		(0.02685)
x_8 (energy)	-0.0221	x_{23} (acoustic_vector_2)	-0.0133
	(0.00691)		(0.01159)
x_9 (flatness)	-0.0084	x_{24} (acoustic_vector_3)	0.0243
	(0.00123)		(0.00334)
x_{10} (instrumentalness)	-0.0174	x_{25} (acoustic_vector_4)	-0.0095
	(0.00151)		(0.00029)

Table 6 continued from previous page

	Dependent Variable		**Dependent Variable**
x_{11} (key)	0.0334	x_{26} (acoustic_vector_5)	0.0320
	(0.00211)		(0.00480)
x_{12} (liveness)	-0.0004	x_{27} (acoustic_vector_6)	0.0142
	(0.00127)		(0.00779)
x_{13} (loudness)	-0.0083	x_{28} (acoustic_vector_7)	-0.0022
	(0.00119)		(0.01344)
x_{14} (mechanism)	-0.0027	x_{29} (acoustic_vector_8)	0.0246
	(0.00064)		(0.00377)
ROC-AUC Score	0,5491250		

Figure 1: Track Features Count



Source: Own Illustration