Shall We Dance?

Using Naive Bayes and Spotify Data to Predict If a Song is a Good Dance Song

Val Booth
vxb4825@rit.edu

1. OVERVIEW

In this project I analyzed audio features and contextual features of popular dance songs and non-dance songs available on Spotify to try and predict if any given song would be a dance hit. A dance hit is a song that encourages "the kind of dancing people do on their own, just getting into the beat, without any practiced knowledge of dancing", not a song you would say, slow dance to with a partner [10]. An example of a dance song from the data set used would be "Just Dance", a 2008 song by Lady Gaga, whereas "Bad Day", a 2005 song by Daniel Powter, would be a non-dance song.

Spotify is one of the leaders of the music streaming industry, with their "Discover Weekly" feature, based on data mining songs and user activity, being a huge success for the company [21]. Spotify's ability to data mine their songs is due in part to their acquisition of the EchoNest, an industry leading music intelligence company [2]. Spotify, in turn, allows data mining on their data through the exposure of their feature rich developer API, without which this project wouldn't be possible [3].

2. RELATED WORKS

2.1 Music Theory

2.1.1 *The Affective Value of Pitch and Tempo*

In 1936, Department of Psychology at the University of Minnesota Doctorate Kate Hevner performed a study about the relationship between emotion and pitch and tempo in music [11]. Dr. Hevner had come up with 8 emotional groups to describe music; each group containing a set of adjectives. I believe her Group 6 of "merry, bright, vivacious, cheerful, happy, gay, joyous, and carefree" and Group 7 of "soaring, triumphant, elated, exciting, impetuous, restless, stirring, spirited and dramatic" best described danceable music. She found a selection of music pieces that "would sound pleasing" at different pitch and tempo levels, and then had a pianist perform one variety for observers (so no observer saw both varieties of the same piece), who could vote for the words in any one of her predefined music groups. In her study, slow tempos were between 63 and 80 BPM, fast tempos were between 102-152 BPM, and pitch was changed by one octave in each direction.

Fast tempos strongly affected sorting into both Group 6 and Group 7. Pitch had a weaker effect- the higher the pitch, the more likely it was labeled with Group 6, but the less likely it was labeled Group 7. Of all of the music features she studied in regards to emotion or expressiveness (also including Mode, Rhythm, Harmony and Melody), "pitch and tempo show themselves to be of the greatest importance in carrying the expressiveness in music. Tempo plays the largest part of any of the factors. It yields majorities on all groups and its effects are clear cut and consistent." Relevant portions of her "Relative Weights For Musical Characteristics For Each Affective State in Terms of D./P.E.$_{\mathrm{D.}}$ From The Six Experiments" table are recreated below, although she "warn[s] again that our findings are in the nature of broad generalizations and averages which represent the trend rather than the specific function."

Musical Element	happy bright	exciting elated
Mode	major 24	—
Tempo	high 20	fast 21
Pitch	fast 6	low 9
Rhythm	flowing 10	firm 2
Harmony	simple 16	complex 14
Melody	—	descending 7

2.1.2 *Determining "Danceability" From Audio Signals*

Given its citation in a paper by EchoNest employee Paul Lamere, it would seem that the EchoNest (and later, Spotify) "danceability" metric is based on the dissertation about music complexity by Dr. Sebastian Streich [14, 20]. Streich was working off of a paper by Jennings, et. al. which analyzed the "Detrended Variance Fluctuation Exponent", the result of "Detrended Fluctuation Analysis" (DFA) on music audio files. DFA studies the periodic trends and rhythm complexity in songs, and was originally used to categorize music into genres [12]. Based on Jenning et.al.'s classifier, Streich saw correlations between their DFA output and "danceability", as the classifier was able to easily distinguish dance music styles like Techno or Brazilian Forro from more "high art" genres. He believed viewing DFA output as "danceability" made the results more "immediately accessible for human music perception". He found that higher DFA exponent values referred to a higher rythmic complexity (not danceable), whereas low exponent values referred to low rhythmic complexity (highly danceable). Using human evaluators and comparing to metadata about the song like genre, Striech believed he "[could] say that there is a strong evidence for the danceability descriptor being related with semantic properties of the music tracks it was computed on" [20].

2.2 Song Recommendation and Music Similarity

2.2.1 What Works in Music Reccomendation

The Co-Founder and CTO of EchoNest, Brian Whitman, wrote a blog post about how industry was doing music recommendation at the time, how the EchoNest was attempting the task, and what he thought did and didn't work [23]. He claimed there are 4 different sources of music knowledge: "(1) activity data, (2) critical or editorial review, (3) acoustic analysis, and (4) text analysis." The EchoNest focused on the latter two. For acoustic analysis, which is currently available as Spotify "Get Audio Features For a Track" API endpoint, "low level information can be combined through some useful applications of machine learning that Tristan and has [sic] team have built over the years to 'understand' the song at a higher level. We emit song attributes such as danceability, energy, key, liveness, and speechiness, which aim to represent the aboutness of the song in single floating point scalars. These attributes are either heuristically or statistically observed from large testbeds: we work with musicians to label large swaths of ground truth audio against which to test and evaluate our models." [19] However, he believed that acoustic analysis was not sufficient on its own to do music recommendation. The EchoNest addressed this issue by scraping the web for reviews and other text content that it could use to tags songs in its database in order to get "cultural analysis" [23].

2.2.2 A Model-Based Approach to Music Similarity

The idea of needing both cultural and content information about a song in order to achieve music knowledge was echoed by future EchoNest employee Paul Lamere's research article, "A Model-Based Approach to Music Constructing Similarity Functions" [22]. He explained how using purely audio content could erroneously mark two songs as similar: "A common example might be the confusion of a classical lute timbre, with that of an acoustic guitar string that might be found in folk, pop, or rock music. These two sounds are relatively close together in almost any acoustic feature space [...] but would likely be placed very far apart by any listener familiar with western music. This may lead to the unlikely confusion of rock music with classical music". However using cultural data of a song, for example genre or other human descriptions, is limited by the availability of such data; using cultural data alone could limit a data set to only popular music.

Lamere's goal for a music recommendation system would be one that was highly sensitive to the library of the listener- it should be more fine tuned to the nuances between classical music, if the user had a lot of classical pieces in their library, but could have coarser similarity between other genres. In his words, "if the user only listens to dance music, they will care about fine separation of rythmic or acoustic styles and will be less sensitive to the nuances of pitch clases, keys, or intonations used in classical music". He also noted that evaluating any music similarity metric is difficult as "we are trying to emulate a subjective perceptual judgement". An ideal test for the model he proposed would be large-scale listening tests, although due to the subjective nature of the task it is sometimes difficult to get human annotators to agree.

3. DATA

3.1 Song Set

The song set represented a total of 1614 songs available on Spotify from the 1980s to 2016, and was comprised of 1042 non-dance songs and 572 dance songs. Song selection came from quality curated lists from Billboard, a company known in the music industry for their popular song listing, and music enthusiast site DigitalDreamDoor.com, and represented both dance and non-dance songs popular in the United States.

Billboard has year end "Hot 100" and "Dance/Club" song lists for the years 2006-2016, although data from 2012 was unfortunately unusable, as will be discussed in section 4.2.1. Hot 100 songs are: "[Any given] year's most popular songs across all genres, ranked by radio airplay audience impressions [in the U.S] as measured by Nielsen Music, sales data [in the U.S.] as compiled by Nielsen Music and streaming activity data provided by online music sources." [9] Dance/Club songs are: "[Any given] year's most popular songs played in dance clubs, compiled from reports from a national sample of club DJs." [8]

In order to represent a broader time range of music, the 1980s, 1990s, and early 2000s "Greatest Songs" and "Greatest Dance Songs" lists were used from DigitalDreamDoor.com. DigitalDreamDoor.com is a best music list compilation site run as a hobby by one music enthusiast, who uses various websites and feedback from users to curate lists [7].

Dance songs mostly came from "Dance/Club" and "Greatest Dance Songs" lists, while non-dance songs were taken from "Hot 100" and "Greatest Songs" lists. Exceptions for song classification can be found in section 4.1.1.

3.2 Song Features

Algorithms were run on song content features (also known as "audio features") from Spotify's "Get Audio Features for a Track" endpoint, along with context features such as genres taken from each song's artist from Spotify's "Get an Artist" endpoint [19, 18]. A short description of each feature and some of its values for the data set are listed below. For all modes, bin start is used.

- **Genres**: A list of genres that are associated with the artist of the song. Each genre was its own feature, with blanks for songs/artists with less than the max number of genres. Genres are found for Spotify artists through a mixture of "machine listening" with manual classifying, and natural language processing on web scrapes [15, 23].

  Min Genre Count: 0

Max Genre Count: 26

Average Genre Count: 7-8

Unique Genre Count: 331

Total Genre Count on Spotify: 1,387 (As of January 2016 [15])

- **Acousticness**: A confidence measure from 0.0 to 1.0 of whether the track is acoustic. 1.0 represents high confidence the track is acoustic.

  Mean: 0.117153684

  Median: 0.0413

  Mode: 0 (Count: 439; Bin Size .02)

  Min: 0.00000984

  Max: 0.965

- **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. Measured from 0.0 - 1.0 where 1.0 is the most confident.

  Mean: 0.655596035

  Median: 0.668

  Mode: 0.69 (Count: 106; Bin Size .02)

  Min: 0.127

  Max: 0.973

- **Energy**: Energy is a measure from 0.0 to 1.0 and represents a perceptual measure of intensity and activity. Typically, energetic tracks feel fast, loud, and noisy. Perceptual features contributing to this attribute include dynamic range, perceived loudness, timbre, onset rate, and general entropy.

  Mean: 0.728242813

  Median: 0.744

  Mode: 0.69 (Count: 86; Bin Size .02)

  Min: 0.0565

  Max: 0.996

- **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". The closer the instrumentalness value is to 1.0, the greater likelihood the track contains no vocal content. Values above 0.5 are intended to represent instrumental tracks, but confidence is higher as the value approaches 1.0.

  Mean: 0.050712019

  Median: 0.00000181

  Mode: 0 (Count: 1367; Bin Size .02)

  Min: 0

  Max: 0.954

- **Key**: A nominal feature representing the key the track is in. Integers map to pitches using standard Pitch Class notation [1]. Due to the Circle of Fifths, amateur pianist colleague Adam McCarthy believed that although the feature is an integer, it should not be treated as an ordinal attribute.

  Mode: 1 (Key of C♯; Count: 200)

  Min: 0 (Key of C)

  Max: 11 (Key of B)

- **Liveness**: Detects the presence of an audience in the recording. Higher liveness values represent an increased probability that the track was performed live.

  Mean: 0.190192379

  Median: 0.13

  Mode: 0.09 (Count: 225; Bin Size .02)

  Min: 0.0176

  Max: 0.99

- **Mode**: A nominal feature, 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.

  Mathematical Mode: 1 (Count: 987)

- **Speechiness**: Speechiness 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.

  Mean: 0.089905328

  Median: 0.0532

  Mode: 0.035 (Count: 376; Bin Size .01)

  Min: 0.0239

  Max: 0.576

- **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.

  Mean: 121.9838147

  Median: 123.822

  Mode: 127.5 (Count: 252; Bin Size 5)

  Min: 51.316

  Max: 210.75

- **Valence**: A measure from 0.0 to 1.0 describing the musical positiveness conveyed by a track. Tracks with high valence sound more positive (e.g. happy, cheerful, euphoric), while tracks with low valence sound more negative (e.g. sad, depressed, angry).

  Mean: 0.566031537

  Median: 0.577

  Mode: .725 (Count: 131; Bin Size .05)

  Min: 0.0373

  Max: 0.977

3.2.1 Song Features Removed

Although loudness, time signature, and duration (ms) were also available, they were removed because they were either irrelevant or unuseful. Loudness, per the endpoint documentation, is "useful for comparing relative loudness of tracks", which could have more to do with recording quality than the nature of the song [19]. Additionally the vast majority of songs in the set had a time signature value of 4 with some songs dubiously having a time signature of 5 (amateur pianist colleague Adam McCarthy was very suspicious that Spotify believed "Carry On" by fun. had a 5 time signature as opposed to 4), and although EDM dance songs may have very long durations, such as 7 minutes, many pop dance songs will have average, around 3 minute length durations, so EDM enthusiast colleague Adam McCarthy recommended that the feature be removed.

4. DATA ISSUES

4.1 Issues

4.1.1 Dance Songs Included

In order to reduce data gathering and data cleaning, along with having reputable sources for whether or not a song was a dance song, the scope of the songs included in the set was very limited. Given the parameters of the songs included, the classifier can more accurately be said to be trying to distinguish American popular dance songs from other American popular hits. Many other cultures are well known for their dance music, and unfortunately the training data did not include, for example, popular salsa or rumba songs. An extension of this project could be to run these same methods on dance and non-dance hits from other countries with even larger dance cultures than the United States (e.g. Columbia, where salsa is popular).

Despite best efforts even with the limited song scope of the project, it was a struggle to prevent bias in the song set towards non-dance songs. This was partly because of the scarcity of lists of well labeled dance songs, and partly due to finding those dance songs on Spotify. DigitalDreamDoor only had 52 songs on its "Greatest Dance Songs of the 1980s" list, despite having 93 songs in their "Greatest Songs of the 1980s" list. Even worse, although Billboard had a top 100 end of the year chart for dance songs for 2006-2011 and 2013-2016, only 432 of those songs had Spotify ids, with many of those songs being duplicates of songs in the DigitalDreamDoor "Greatest Dance Songs 2000s" set. The lack of ids mostly seemed to be caused Billboard not containing the preview for songs on Spotify ("Bad Romance" by Lady Gaga, a year end hot Dance/Club song of 2010, is on Spotify but not available for preview on Billboard) but also because of the obscurity of some of the dance songs ("The One That Got Away (Wamdue/Valentin Mixes)" by Natasha Bedingfield is not available on Spotify).

In order to make the data set less biased and because Billboard only included the top 100 dance songs for any given year, reclassification was done for some of the dance songs, which is detailed in 4.2.2.

4.1.2 Genres

Genre is regarded as one of the most important contextual pieces of information about a song used in industry for music recommendation, as discussed in section 2.2.

That being said, it can be difficult to find insight from genres due to the large number of genres and their meaningfulness in general and related to their specific artist. For someone without deep domain knowledge of musical categorization, genre names such as "Contemporary Urban", "Deep Southern Trap", "Bouncy House", or "Escape Room" can be difficult to understand without further research. The number of genres also helps prevent understanding as well; this data set of 1614 songs resulted in 331 distinct genres alone, which is less than a third of all of the genres represented on Spotify [15].

Furthermore, Spotify only allows the user to retrieve genre information at the artist level, which for this project may be too coarse of information. One example is the artist Fergie, who Spotify identifies as being "dance pop", "hip pop", and "pop" among other genres. While this may be representative for two of the songs in the set, "Clumsy - Pajon Rock Mix" and "Glamorous - PoetNameLife Dance Remix", the set also includes her more acoustic, less "dance pop" ballad "Big Girls Don't Cry (Personal)". Admittedly, the audio features of "Big Girls" represent that it is more acoustic, as it has a value of .21 compared to the other songs' .09 and .003, respectively.

Although Fergie is largely a "dance pop" artist, for other artists there may have been genres generated from too small a selection of their song catalog. For example, there were 9 different types of "christmas" genres represented by the data set, with blink-182's "All the Small Things" being one example of "heavy christmas". Blink-182 has released 2 Christmas songs, "Won't Be Home For Christmas" and "Happy Holidays, You Bastard", but "All the Small Things", along with most of blink-182's catalog, is not particularly in the "heavy christmas" genre. An improvement for this project would be to have genre information on a song, as opposed to artist, level.

4.1.3 Tempo

Tempo was recognized as the most important audio feature in theoretical studies of songs' moods, which related to danceability, as discusssed in section 2.1.1, and in practical audio analysis of what made a song danceable, as explained in section 2.1.2.

While perusing the data set, however, there seemed to be some anomalies for tempo values for songs, where what was returned by Spotify didn't seem to match the actual rhythm of the song. Two examples from both ends of the spectrum are the song "Rehab" by Amy Winehouse, and "FourFiveSeconds" by Rhianna, Kanye West, and Paul McCartney. According to Spotify, "Rehab" has a tempo of 72 BPM while "FourFiveSeconds" has a tempo of 205 BPM. Amateur pianist colleague Adam McCarthy believed that these tempos seemed to be off by some factor of 2: based on listening alone we'd estimate "Rehab" to be around 144 BPM, while "FourFiveSeconds" should around 100 BPM. Developers for the Spotify Web API have admitted that audio features are extracted by machine learning models and are sometimes off [16]. With no way to verify the actual tempo, the Spotify values were used in the algorithm.

That being said, there were also many probably valid outliers at both ends of the tempo spectrum, such as the slower "If I Was Your Man" by Bruno Mars at 72 BPM and "End of the World" by R.E.M. at 205 BPM.

4.2 Data Gathering and Cleaning

4.2.1 Data Gathering

EchoNest, now a part of Spotify, employee Paul Lamere has an open source project that can load a user's Spotify playlist and display the audio features for each song, retrieved from Spotify's "Get Audio Features for a Track" endpoint [13, 19]. A local, cloned and edited version of that project was used to retrieve `JSON` objects with all of the information listed in section 3.2 along with the song and artist name for general data set understanding.

The local project was run directly on DigitalDreamDoor's Spotify account's playlists for "Greatest Dance Songs 2000s", "100 Greatest Dance Songs of the 1990s", "Greatest Dance Songs of the 1980s", "Greatest Songs of the 1980s", "Greatest Songs of the 1990s" and "Greatest Songs of the 2000s". For the Billboard songs a Python script was run that used BeautifulSoup to parse out the Spotify ids used in the preview feature on all of the Billboard web pages for the "Hot 100" and "Dance/Club" from 2006-2011 and 2013-2016 [4, 17]. The format of the 2012 top Dance/Club songs Billboard webpage was drastically different from all other years listed and only featured one Spotify id, so both "Hot 100" and "Dance/Club" songs for 2012 were not included in the data set. Once retrieved, those ids were then fed into a further edited version of the local "Sort Your Music" project to retrieve those `JSON` objects as well.

4.2.2 Data Cleaning

`JSON` objects needed to be converted to `csv` files in order to be used with Weka, which was done with a third party site covertcsv.com [5, 6]. The `csv` file was then loaded into Microsoft Excel for further processing. Songs with duplicate Spotify song ids were removed and songs without duplicate ids but containing duplicates of every other feature were also removed. If one duplicate was marked as danceable, but the other song was marked as non-danceable, the danceable copy was kept. Songs with duplicate song and artist titles but different audio features were kept as separate data points and not further edited, as they all had the same danceable class value, or removed.

In order to reduce the bias of the data set and because of the exclusivity of the Billboard dance songs in particular, over 50 songs were reclassified from non-danceable to danceable. These were nearly all songs originally from Billboard, and included numerous songs by The Black Eyed Peas and Pitbull. Only 4 songs needed to be reclassified as non-danceable from danceable, including "Unwritten" by Natasha Bedingfield in 2006 and "Read My Mind" by The Killers in 2007. Songs were reclassified based on recognition and best judgment with some feedback from colleague Adam McCarthy.

Finally, the artist and song names were removed as they were not going to be used as features and Weka cannot read `csv` files that contain data with apostrophes.

5. DATA MINING

5.1 Overview

In order to determine the best subsections of features to test on, I considered not only what related works explored in section 2 had claimed were important indicators of a dance song, but also the visualization of each individual audio feature compared to its danceable class. You can see this visualization of audio features composed by Weka in figure 1 [5].

Based off other related works, I had decided to try and focus in on genres, tempo, key and "danceability", and I wanted to focus on valence and energy as they seemed to have a positive correlation coefficient with danceable songs counts for higher ranges than it did non-dance songs. Based on these factors and the scattered nature of genres (i.e., not every song had the same number of genres, some having no genres) I decided to test and compare every algorithm with the following sets of features:

- Only genres
- All available genres, all audio features
- One Genre (the first listed), all audio features
- Only audio features
- Energy, key, tempo, valence, "danceability"
- Energy, key, tempo, valence
- Energy, key, tempo
- Key, tempo
- Energy, tempo
- Energy, key

I initially ran my data set through both Hierarchical Clustering and k-means clustering for $k = 2, 3, 4...8$ for the above feature sets, but did not find them to be very successful. The best misclassification rate for a clustering algorithm was 37%, although it ranged up to 63%. Seeing as the classifier algorithms were much more accurate, only the 1R Rule, Decision Tree (J.48), Naive Bayes and Random Forests will be investigated in this paper.

5.2 Comparison of Classifiers

All classifiers had their best values for the most amount of features. For larger feature sets Random Forest had surprisingly high accuracy rates and AUC sizes, although it began to suffer when more features were removed.

The following metrics were graphed for every feature subsets for the 1R, Decision Tree, Naive Bayes, and Random Forest algorithms: classification accuracy in figure 2, the TP rate for dance songs in figure 3, and the AUC in figure 4. For the 1R data, the following features were chosen for the 1 rule in this order: 2nd listed genre, 2nd listed genre, 1st listed genre, tempo, tempo, tempo, energy, tempo, tempo.

6. CONCLUSIONS

6.1 Validation

In order to help validate the data, I once again edited Paul Lamere's "Sort Your Music" instead to display whether or not I would predict any song in a user's playlist as a dance hit. I used Naive Bayes as my data model with the full set of audio features and genres, as it had the highest values for all three metrics out of the Non-Random Forest Classifiers. 255 songs were assessed, by myself and colleagues Adam McCarthy and George Herde. On the limited validation data set the classifier had a $66.\overline{6}\%$ accuracy. The full confusion matrix can be seen at figure 5.

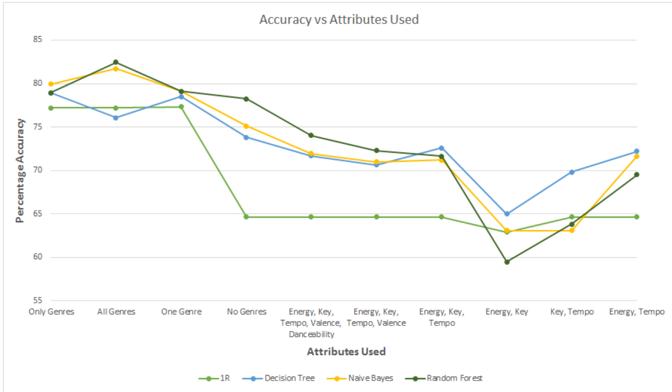

Figure 2: Classification Rate vs. Classifier and Feature Set

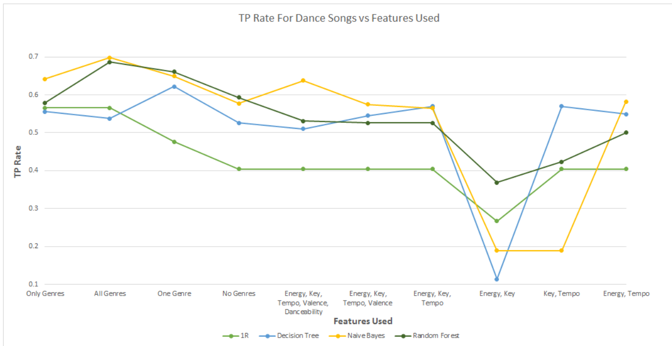

Figure 3: True Positive Rate For Dance Songs vs. Classifier and Feature Set

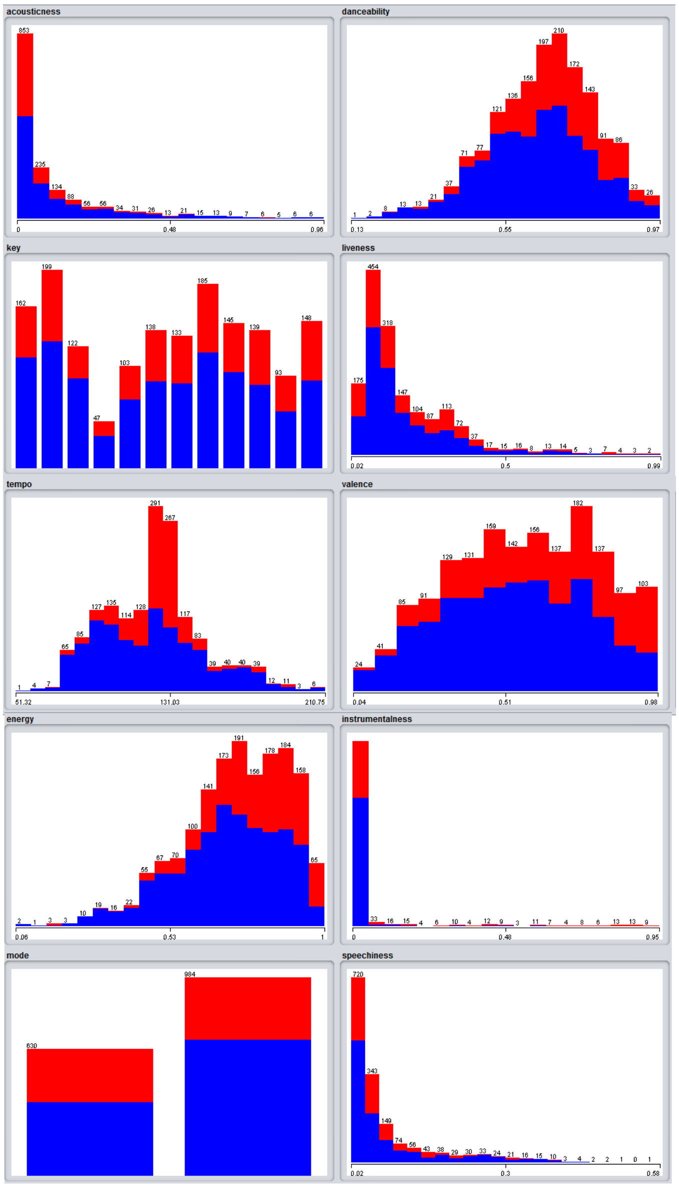

Figure 1: Visualization of data points compared to danceable class. Non-danceable is represented by blue and danceable is represented by red.

6.2 Comparison to Related Works

Despite some concerns about the large and varied genre set space discussed in section 4.1.2, all classifiers not only did better with some genre information than no genre information, they also did nearly as well with only genre information as they did with genre and audio features. The successfulness of the classifiers with only genre information strongly proves Brian Whitman and Paul Lamere's comments, used in section 2.2, about the necessity of genre information in music information retrieval works.

Comparing the results of this project with Dr. Hevner's work, explained in section 2.1.1, is more difficult, as she only had access to more basic music properties (mode, tempo, pitch, rhythm, harmony, melody) compared to EchoNest and Spotify's more involved ones ("danceability", energy, valence). Furthermore, not every algorithm reacted the same way to just key and tempo. The Decision Tree algorithm retained most of its accuracy, TP rate for dance songs, and AUC from using Energy, Tempo, and Key to just using Key and Tempo. Naive Bayes, on the other hand, performed worse across the three metrics from switching to Energy, Tempo, and Key to just Tempo and Key.

The usefulness of Streich and Spotify's "danceability" feature, detailed in section 2.1.2, in each classifier is likewise difficult to judge. Most classifiers did nearly just as well across all three metrics with energy, key, tempo, valence, and

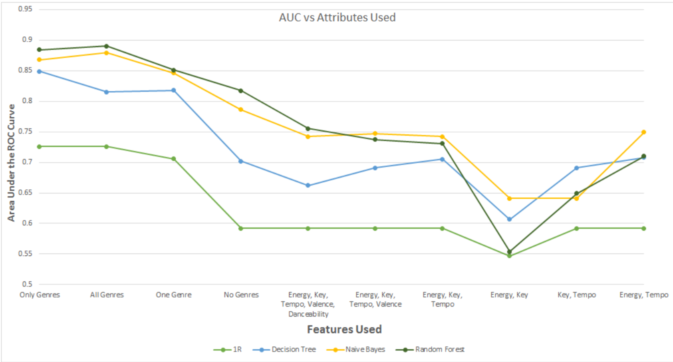

Figure 4: Area Under Curve vs. Classifier and Feature Set

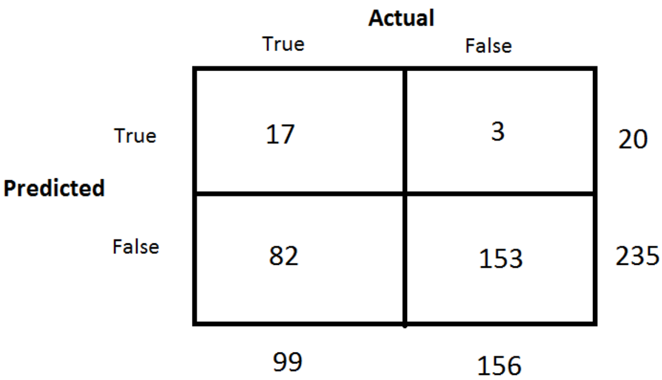

Figure 5: Confusion Matrix for Validating Classifier

danceability as they did with just energy, key, tempo, and valence (with the exception of Naive Bayes TP rate, which spiked between No Genres and just energy, key, tempo, and valence). Instead of "danceability" not being very useful in determining whether a song is a dance song or not, it may be that there's some overlap between energy, key, tempo, and valence and "danceability" measuring. Visualizing the feature set in Weka, seen in figure 6, seems to show positive correlation between energy, valence, and "danceabilty", and between higher "danceability" and number of dance songs [5].

6.3 Final Conclusions

Music understanding is a deeply complex data mining field in which a more holistic view of songs (often accompanied by more features) gives more refined knowledge of the music. This topics in this paper only cover the tip of the iceberg for music understanding- all of the features discussed in this paper have to do with the nature of the song. Another big area, especially in music reccomendation, is "collaborative filtering", or "friend-to-friend music recommendations enabled by social networks". EchoNest CTO Brian Whitman believes collaborative filtering techniques "are extremely valuable in music discovery", even if they are outside the scope of this paper.

Many of the ideas about music information retrieval and

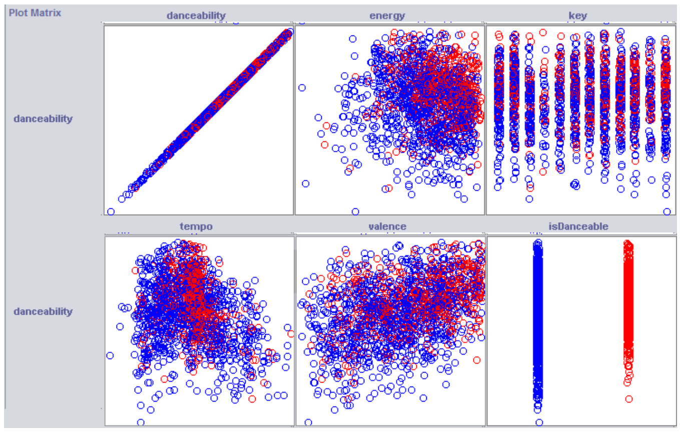

Figure 6: Visualization of the distribution of dance and non-dance songs comparing "danceability" and other audio features. Non-dance songs are represented by blue points, dance songs by red.

even "danceability" seen in previous works were corroborated by the results of my data mining classifiers. The more information we can read from audio signals or machine learn from music, however, the better. Dr. Kate Hevner's work, discussed in section 2.1.1, didn't have access to the idea of "energy" of a song, which is a relatively recent development by the team from EchoNest [23]. Although tempo and pitch, as she predicted, seem to be very important towards a song's mood, which can relate to "danceability", for all classifiers there were higher AUC and TP rate numbers for just energy and tempo as compared to energy and key.

Despite the discussion of genres and other contextual information in section 2.2, I was shocked by the effectiveness of genres alone in the classifiers. My skepticism mostly came from my concerns explored in section 4.1.2, although the usefulness of genres does make sense. Many cultural aspects of music as subjective, and genres are the easiest ways for humans to label the cultural and contextual aspects of songs that computers can't pick up on.

As more information about music understanding and music recommendation is found, the more companies (like Billboard and Spotify) and music enthusiasts (like myself and the people behind DigitalDreamDoor) can create and share enjoyable music and dance experiences. Spotify's "Discover Weekly" shows that a lot of progress has been made, but with such a large space of music listeners and music producers, there is always more work to be done.

7. REFERENCES

[1] Pitch (class) - open music theory. `http://openmusictheory.com/pitch(Class).html`.

[2] `http://the.echonest.com/pressreleases/spotify-acquires-echo-nest/`, 2014.

[3] `https://developer.spotify.com/`, 2017.

[4] `https://www.python.org/`, 2017.

[5] `http://www.cs.waikato.ac.nz/ml/weka/`, 2017.

[6] `http://convertcsv.com/`, 2017.

[7] About DigitalDreamDoor.com. `https://digitaldreamdoor.com/pages/about_us_ddd.html`, 2017.

[8] Billboard Music. Dance/club songs | year-end charts (2006). http://www.billboard.com/charts/year-end/2006/hot-dance-club-play-songs, 2017.

[9] Billboard Music. HOT 100 SONGS | YEAR-END CHARTS (2006). http://www.billboard.com/charts/year-end/2006/hot-100-songs, 2017.

[10] D. Book. Which 1960s songs are very danceable? - StraightDope Message Board. http://www.billboard.com/charts/year-end/2006/hot-dance-club-play-songs.

[11] K. Hevner. The affective value of pitch and tempo in music. *The American Journal of Psychology*, 49(4):621–630, 1937.

[12] H. D. Jennings, P. C. Ivanov, A. d. M. Martins, P. da Silva, and G. Viswanathan. Variance fluctuations in nonstationary time series: a comparative study of music genres. *Physica A: Statistical Mechanics and its Applications*, 336(3):585–594, 2004.

[13] P. Lamere. Sort your music. http://sortyourmusic.playlistmachinery.com/, 2016.

[14] F. Maillet, D. Eck, G. Desjardins, P. Lamere, et al. Steerable playlist generation by learning song similarity from radio station playlists. In *ISMIR*, pages 345–350, 2009.

[15] Patch, Nick. Meet the man classifying every genre of music on spotify - all 1,387 of them. https://www.thestar.com/entertainment/2016/01/14/meet-the-man-classifying-every-genre-of-music-on-spotify-all-1387-of-them.html, Jan 2016.

[16] H. Rawlinson. Time signature - issue #379 - spotify/web-api. https://github.com/spotify/web-api/issues/379#issuecomment-271863819, 2017.

[17] L. Richardson. Beautiful soup: We called him tortoise because he taught us. https://www.crummy.com/software/BeautifulSoup/, 2017.

[18] Spotify Developer. Get an artist. https://developer.spotify.com/web-api/get-artist/, 2017.

[19] Spotify Developer. Get audio features for several tracks. https://developer.spotify.com/web-api/get-several-audio-features/, 2017.

[20] S. Streich. *Music complexity: a multi-faceted description of audio content.* PhD thesis, Universitat Pompeu Fabra, 2006.

[21] M. Swant. Even spotify is surprised by the huge success of its discover weekly playlists. http://www.adweek.com/digital/even-spotify-surprised-huge-success-its-discover-weekly-playlists-173129/, 2016.

[22] K. West and P. Lamere. A model-based approach to constructing music similarity functions. *EURASIP J. Appl. Signal Process.*, 2007(1):149–149, Jan. 2007.

[23] B. Whitman. How music recommendation works - and doesn't work. https://notes.variogr.am/2012/12/11/how-music-recommendation-works-and-doesnt-work/, Dec 2012.