Music Genre Prediction

Pranav Manjunath

11/12/2020

Summary

The goal of this study is to understand the various Spotify audio features that strongly impact and determine the genre of a song, as well as to build a parametric and non parametric model that can accurately predict the genre of a song based on these audio features. Multinomial Logistic Regression (MLR) and Gradient Boosting Method (GBM) algorithms have been used and the GBM model performs better, having an overall accuracy of 54.02%. Audio features such as Danceability, Acousticness, and Speechness have the highest relative influence when predicting the genre of a song.

Introduction

A music genre is a category that identifies music as belonging to a shared tradition or set of conventions. As genres lacks clear boundaries, it becomes difficult to accurately classify songs into a particular musical genre. This classification however becomes extremely crucial for music platforms and musicians. Spotify, a Swedish-based audio streaming and media services provider, created an API that extracts multiple audio features of a song, calculated by their own algorithm. Upon inputting a song's ID, the API would output the audio feature values for given song. A detailed description of the various audio features is mentioned in the appendix.

The questions of interest for this project are:

1. What are the important features that impact the genre of a song?
2. Is it possible to build a model that can accurately predict the genre of a song based on the Spotify Music Features? Does the data follow the parametric assumptions? Would a non parametric model perform better than a parametric model?

Data

The original dataset was taken from Kaggle [1], consisted of 122382 songs and over 2800 unique genres. There are 20 columns in the data, namely, Name, Danceability, Energy, Key, Loudness, Mode, Speechness, Acousticness, Valence, Tempo, Duration_ms, Time Signature, Liveness, Instrumentalness, Type, ID, Uri, Ref_Track, and URL_features. For this project, the columns Type, ID, Uri, Ref_Track and URL_Features have been removed from the dataset. A detailed description of the various audio features is present in the Appendix. This dataset contained no missing values however had duplicate entries which were removed.

[1] Data Link: https://www.kaggle.com/grasslover/spotify-music-genre-list?select=songDb.tsv

Data Preprocessing

a. Combining the Genres. The first step of data preprocessing was to combine the genres into its parent genre. A new variable (Overall_Genre) was created to store the parent genre and used as the response variable in this study. Genres that did not have one particular parent genre (genres with multiple parent genres also excluded) and had less than 2000 songs were removed from the data. Throughout this paper the term Overall_Genre will be represented by genre. Detailed description of which genres were grouped is present in the appendix.

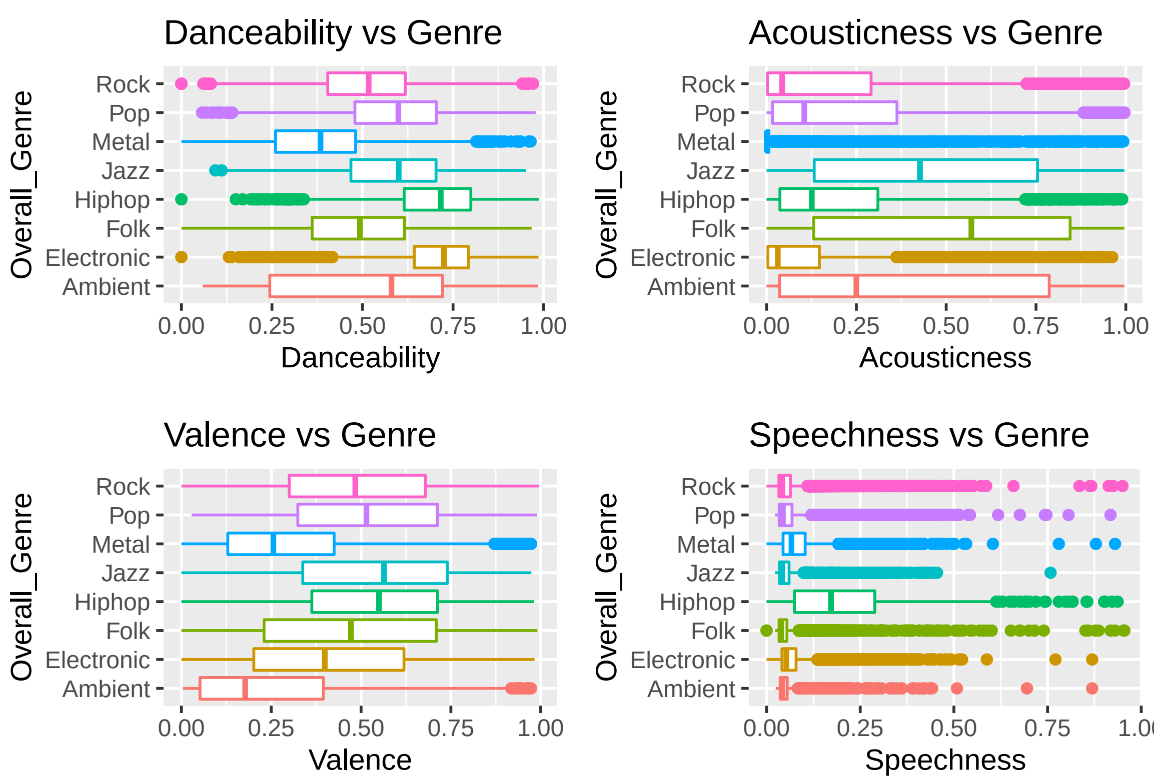
b. Filtering. Genres with less than 2000 songs were removed from the dataset. After filtering the data, there were 8 genres present.

c. Data Types. The variable "Overall_Genre" (Response Variable) was converted to a factor variable. "Key" and "Time Signature" was converted to discrete variables while "Mode" was converted into a binary value.

d. Final Data. The final dataset used for the study consists of 63302 songs and 8 genres. The genres used are Rock, Hip-hop, Metal, Ambient, Folk, Electronic, Pop, and Jazz. The table below indicates the number of songs present in each Genre.

Ambient	Electronic	Folk	Hiphop	Jazz	Metal	Pop	Rock
2326	7739	4883	7131	2844	8693	10014	19672

Exploratory Data Analysis



The above box-plots illustrates the various features across the genres. The findings from the EDA also has justification in real life. Key Findings:

1. Hiphop and Electronic music have the highest median Danceability values when compared to the other variables. This is also supported as Electronic (such as EDM, House) and Hiphop music are known for dancing.
2. As Folk music uses real instruments and very less electronic sounds, it can be justified that Folk has the highest median acousticness value when compared to other genres. The genre with the second highest median acousticness value is Jazz.
3. It can be noticed that Jazz has the highest median valence value, then comes Hip-hop. Valence measures the positivity in the music. Interestingly, Ambient Music has the lowest median valence value, which is in fact surprising as ambient music is known to calm the mind and spread happiness.
4. As Hiphop is a combination of DJing, Rapping, Breakdancing, and Graffiti writing, it justifies the above boxplot as it has the highest median Speechness value when compared to the other genres.

Model

The dataset was split into a 80:20 stratified training testing set and was used on two models. One major assumption made while conducting the study is that each song belongs to only one genre. Transformations and interactions were tested but did not improve the model's performance. Hence, I focused on only main effects as predictors.

Multinomial Logistic Regression (MLR)

The final model for multinomial logistic regression is

$\log(\pi_{i,j}/\pi_{i,1}) = \beta_{0,j} + \beta_{1,j}$ * Danceability + $\beta_{2,j}$ * Energy + $\beta_{3,j}$ * Loudness + $\beta_{4,j}$ * Acousticness + $\beta_{5,j}$ * Instrumentalness + $\beta_{6,j}$ * Valence + $\beta_{7,j}$ * Duration + $\beta_{8,j}$ * Tempo + $\beta_{9,j}$ * Mode + $\beta_{10,j}$ * Speechness + $\beta_{11,j}$ * Key + $\beta_{12,j}$ * Time_Signature

The summary of the model is present in the Appendix. To identify significant variables, the 95% confident intervals was calculated. For majority of the genres, Tempo, time_signature, and Key contain 0 in the interval, indicating that it is not significant. For all genres, Danceability is the only feature that does not contain 0 in its 95% confident interval, indicating that it is a significant variable. The ROC curves (on the training dataset) are present in the Appendix .The Multi ROC curve has an AUC value of 0.8234.

Below is the Classification Report of the model on the testing dataset.

```
Overall Statistics

               Accuracy : 0.4853
                 95% CI : (0.4781, 0.4924)
    No Information Rate : 0.3108
    P-Value [Acc > NIR] : < 2.2e-16

                  Kappa : 0.3422

 Mcnemar's Test P-Value : < 2.2e-16

Statistics by Class:
```

	Class: Ambient	Class: Electronic	Class: Folk	Class: Hiphop	Class: Jazz	Class: Metal	Class: Pop	Class: Rock
Sensitivity	0.220947	0.55536	0.24317	0.58953	0.117233	0.56655	0.050599	0.7497
Specificity	0.987807	0.93615	0.96918	0.94581	0.989246	0.94249	0.974346	0.5807
Pos Pred Value	0.408488	0.54781	0.39732	0.58004	0.338983	0.61058	0.270463	0.4464
Neg Pred Value	0.970821	0.93795	0.93875	0.94778	0.959713	0.93179	0.845202	0.8373
Prevalence	0.036711	0.12225	0.07711	0.11266	0.044928	0.13731	0.158222	0.3108
Detection Rate	0.008111	0.06789	0.01875	0.06642	0.005267	0.07779	0.008006	0.2330
Detection Prevalence	0.019857	0.12393	0.04719	0.11451	0.015538	0.12741	0.029601	0.5220
Balanced Accuracy	0.604377	0.74576	0.60618	0.76767	0.553240	0.75452	0.512473	0.6652

Figure 1: MLR Classification Report

The model has an overall accuracy of 48.53%. Interestingly, except for the genre Rock, the other genres have a higher specificity value than sensitivity value. Similarly, the Negative Predicted Value for all the genres are above 80% and greater than its respective Positive Predicted Value, indicating that the model is able to accurately predict when a song does not belong to a specific genre. It can also be seen that Hip-hop has the highest balanced accuracy of 76.76%. While observing the binned residual plots, few plots (shown in the appendix) while having 95% of its points inside the confidence interval, showed significant patterns such as a parabolic curve. Upon adding transformations and interactions, there were no difference, inferring that this data might not follow the assumptions of the parametric model.

Gradient Boosting Methods (GBM)

GBM is a non parametric, ensemble method that is used for classification and regression analysis. An ensemble model is a strong learner created by combining multiple weak learners. One of the benefits of implementing GBM is that the model also computes the relative importance/influence scores for each predictor after constructing the boosting trees. Relative influence indicates how valuable each feature is in the construction of the boosted decision trees within the model. A variable with a high importance value indicates that the attribute is used to make key decisions. This value is calculated for a single tree and then averaged across all of the decision trees within the model.

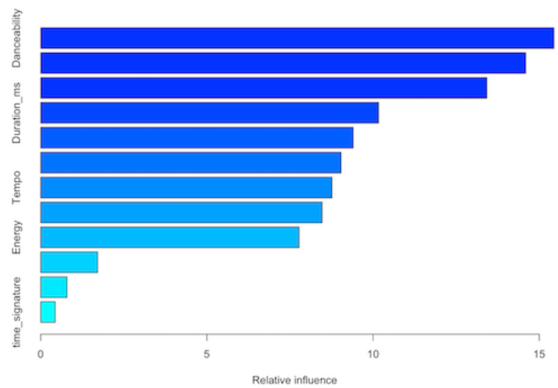
Using the same formula present in MLR, I built a GBM model with the following additional hyperparameters: n_trees = 1000, shrinkage = 0.15, interaction.depth = 2. Below is the Classification Report of the GBM model on the testing dataset.

```
Overall Statistics

               Accuracy : 0.5402
                 95% CI : (0.5331, 0.5473)
    No Information Rate : 0.3108
    P-Value [Acc > NIR] : < 2.2e-16

                  Kappa : 0.4225

 Mcnemar's Test P-Value : < 2.2e-16

Statistics by Class:

                     Class: Ambient Class: Electronic Class: Folk Class: Hiphop Class: Jazz Class: Metal Class: Pop Class: Rock
Sensitivity                 0.32999           0.59199     0.29577       0.68864     0.32825      0.62255    0.16178      0.7385
Specificity                 0.98879           0.94779     0.96924       0.94444     0.98362      0.95513    0.93943      0.6806
Pos Pred Value              0.52874           0.61230     0.44547       0.61146     0.48527      0.68830    0.33425      0.5105
Neg Pred Value              0.97483           0.94344     0.94277       0.95982     0.96887      0.94082    0.85638      0.8523
Prevalence                  0.03671           0.12225     0.07711       0.11266     0.04493      0.13731    0.15822      0.3108
Detection Rate              0.01211           0.07237     0.02281       0.07758     0.01475      0.08548    0.02560      0.2295
Detection Prevalence        0.02291           0.11819     0.05120       0.12688     0.03039      0.12420    0.07658      0.4496
Balanced Accuracy           0.65939           0.76989     0.63250       0.81654     0.65594      0.78884    0.55061      0.7096
```

Figure 2: GBM Classification Report

The GBM model has an overall accuracy of 54.02%, performing better than the MLR model. Similar to the MLR model, expect for the genre Rock, the other genres have a higher Specificity value than Sensitivity value. Each genre has a greater balanced accuracy in the GBM model when compared to the MLR model-Hip-hop having the highest of 81.65%. The plot below shows the Relative Variable Importance constructed using the GBM model.

	var	rel.inf
Danceability	Danceability	15.44
Acousticness	Acousticness	14.59
Speechness	Speechness	13.42
Duration_ms	Duration_ms	10.17
Instrumentalness	Instrumentalness	9.401
Valence	Valence	9.035
Tempo	Tempo	8.763
Loudness	Loudness	8.467
Energy	Energy	7.774
Key	Key	1.712
Mode	Mode	0.7892
time_signature	time_signature	0.4349



As shown in the figure above, Danceability, Acousticness, and Speechness have the top three relative importance value when compared to the other variables, indicating that these are important attribute is used to make key decisions in predicting the Genre of a song. On the contrast, features such as Time Signature, Key and Mode have the lowest relative importance, indicating that these variables do not play a major role in predicting the genre of a song. These findings are also consistent with the MLR model as well.

The variables initially had high multicollinearity. After scaling up and centering the continuous predictors, majority of the variables did not have any problems with multicollinearity (VIF Score < 10).

Discussions

1. From this study, the most important variables that impact the genre of the song are Danceability, Acousticness, and Speechness. Variables such as the Time Signature, Key and Mode do not play a significant role in predicting the genre of a song.
2. When comparing both GBM and MLR, the non parametric model GBM performs better in predicting the genre of a song. However, the overall accuracy of GBM is only 54.02%, still requires more fine tuning to improve the performance. From these results, it cannot be said that GBM model can accurately predict the genre of a song.

Application

To apply this project in the real world, I built a R Shiny App that takens in the inputs of the musician's song features and returns the predicted genre as well as the top 10 songs similar to the one inputted. The Genre prediction is done by the GBM model and song similarities are calculated using Cosine Similarity. There is also a link to access Spotify from the application. Application Screenshot attached in the appendix.

Conclusion

This study focuses on the application of MLR and GBM in predicting the musical genre song based on the audio features extracted by the Spotify Algorithm. The GBM Model produced a better overall accuracy of 54.02%. To understand the influence of various features on the response variable, I looked at the confidence intervals for the MLR model and also the variable importance from the GBM model. The study concluded that Danceability, Acousticness, and Speechness have the top 3 highest relative influence values when compared with the other variables. There are a few limitations with this study. 1) Classifying the songs into one parent genre can be difficult as songs can belong to multiple genres. 2) There are frequent overlaps between various genres. 3) A genre can consist of multiple genres. For example, the genre Indie Rock is a combination of the genres Indie and Rock. For future scope, application of Deep Learning algorithms might help in accurately predict the genre of a song. Another approach that removes the assumption that one song belongs to only one genre is Fuzzy Logic. As songs can also belong to one or more genres, Fuzzy Logic aims to identify the percentage fit that a song belongs to each genre.

Appendix

Spotify Audio Features

Variable	Description
Name	Song Name
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.
Energy	Energy is a measure from 0.0 to 1.0 and represents a perceptual measure of intensity and activity.
Key	The estimated overall key of the track.
Loudness	The overall loudness of a track in decibels (dB).
Mode	Mode indicates the modality (major or minor) of a track.
Speechness	Speechiness detects the presence of spoken words in a track.
Acousticness	A confidence measure from 0.0 to 1.0 of whether the track is acoustic.
Instrumentalness	Predicts whether a track contains no vocals.
Liveness	Detects the presence of an audience in the recording.
Valence	A measure from 0.0 to 1.0 describing the musical positiveness conveyed by a track.
Tempo	The overall estimated tempo of a track in beats per minute (BPM).
Duration_ms	The duration of the track in milliseconds.
Time_Signature	An estimated overall time signature of a track.
Genre	Genre of the Song

Sub Genres that belong to the main Genre

Ambient - “deepchill-out”,“organicambient” “futureambient” “darkambient” “ambientworship” “spaceambient”

Electronic -“tropicalhouse” “darktechno” “deepdeephouse” “nuelectro” “minimaltechno” “electrolatino” “experimentalhouse” “microhouse” “electronicore” “canadianelectropop” “deepdubtechno” “greekhouse” “turkishelectronic” “brazilianelectronica” “bouncyhouse” “tijuanaelectronic” “chicagohouse” “hiphouse” “deepgroovehouse” “darkprogressivehouse” “hardminimaltechno” “czechelectronic” “electronica” “russianelectronic” “electronic” “witchhouse” “germantechno” “nordichouse”

Folk - “portuguesefolk” “runefolk” “dronefolk” “romanianfolk” “spanishfolk” “psychedelicfolk” “nlfolk” “frenchfolk” “turkishfolk” “balticfolk” “traditionalbritishfolk” “irishfolk” “bretonfolk” “traditionalenglishfolk” “norwegianfolk” “flemishfolk” “corsicanfolk” “polishfolk” “catalanfolk” “galicianfolk” “czechfolk” “americanfolkrevival” “okinawanfolk” “traditionalfolk” “finnishfolk”

Hip-Hop - “ghanaianhiphop” “albanianhiphop” “arabichiphop” “undergroundlatinhiphop” “italianundergroundhiphop” “finnishhiphop” “hiphop” “norwegianhiphop” “frenchhiphop” “chiphop” “experimentalhiphop” “austrianhiphop” “venezuelanhiphop” “estonianhiphop” “outerhiphop” “indonesianhiphop” “nativeamericanhiphop” “scottishhiphop” “peruvianhiphop” “lgbtq+hiphop” “turkishhiphop” “australianundergroundhiphop” “italianhiphop” “portlandhiphop” “czechhiphop” “hungarianhiphop” “balkantrap” “dmvrap” “trapfrancais” “rapmetalcore” “emorap” “darktrap” “rapdominicano” “countryrap” “poprap” “romaniantrap” “rapmetal” “j-rap” “undergroundrap”

Rock - “slovakrock” “ostrock” “mexicanrock-and-roll” “canadianrock” “italianpost-rock” “colombianrock” “geekrock” “basquerock” “rockkapak” “swedishrockabilly” “wrock” “icelandicrock” “ecuadorianalternativerock” “deepsoftrock” “uknoiserock” “portugueserock” “gothicrock” “rocktico” “panamanianrock” “czechrock” “germanpost-rock” “garagerock” “psychedelicrock” “rocknacional” “rock” “post-rocklatinoamericano” “germanhardrock” “yugoslavrock” “atmosphericpost-rock” “australianrock” “classicgaragerock” “medievalrock” “alternativerootsrock” “deepchristianrock” “rockcatala” “swedishhardrock” “southernrock” “kindierock” “russianpunk” “indonesianpoppunk” “noisepunk” “deeppoppunk” “hardcorepunk” “chicagopunk” “ukpost-punk” “deeppower-poppunk” “darkpost-punk” “cyberpunk” “steampunk” “ukdiypunk” “mexicanpoppunk” “norwegianpunk” “frenchpunk” “hungarianpunk” “germanpunk” “crustpunk” “swedishpoppunk” “protopunk” “acousticpunk” “minneapolispunk” “christianpunk” “polishpunk” “slovakindie” “birminghamindie” “liverpoolindie” “torontoindie” “elpasoindie” “belfastindie” “hongkongindie” “ecuadorianindie” “alaskaindie” “essexindie” “portugueseindie” “minneapolisindie” “detroitindie” “peruvianindie” “buffalonyindie” “monterreyindie” “dominicanindie” “indietico” “albuquerqueindie” “israeliindie” “tassieindie” “northeastenglandindie” “columbusohioindie” “sanantonioindie” “colombianindie” “swissindie” “brightonindie” “icelandicindie” “okindie” “argentineindie” “coventryindie” “leedsindie” “czechindie” “malmoindie” “chicagoindie” “indienapoletano” “norwegianindie” “vegasindie” “rvaindie” “indianindie” “milanindie” “dallasindie” “michiganindie” “cornwallindie” “seattleindie” “phoenixindie” “idahoindie” “dunedinindie” “dublinindie” “halifaxindie” “westernmassindie” “victoriabcindie” “romanianindie” “guatemalanindie” “melbourneindie” “sydneyindie” “stlindie” “aucklandindie” “welshindie”

Jazz - “deeplatinjazz” “jazzfusion” “russianjazz” “vocaljazz” “neosoul-jazz” “finnishjazz” “britishjazz” “estonianjazz” “smoothjazz” “germanjazz” “jazzelectricbass” “deepsmoothjazz” “spiritualjazz” “contemporaryjazz” “bossanovajazz” “southafricanjazz”

Metal - “celticmetal” “groovemetal” “dronemetal” “belgianmetal” “japanesedeathmetal” “gothicmetal” “norwegianmetal” “bostonmetal” “metalguitar” “neo-tradmetal” “melodicdeathmetal” “melodicmetalcore” “sludgemetal” “technicalblackmetal” “austrianmetal” “metallichardcore” “latinmetal” “norwegiandeathmetal” “finnishblackmetal” “polishblackmetal” “chaoticblackmetal” “turkishmetal” “gothicsymphonicmetal” “welshmetal” “vikingmetal” “greekblackmetal” “koreanmetal” “peruvianmetal” “slavicmetal” “atmosphericpost-metal” “swedishdeathmetal” “instrumentalprogressivemetal” “post-metal” “darkblackmetal” “post-blackmetal” “symphonicpowermetal” “depressiveblackmetal” “thrashmetal”

Pop - “canadianpop” “belgianpop” “lapop” “dreampop” “undergroundpowerpop” “mathpop” “mandepop” “deepbrazilianpop” “thaipop” “v-pop” “vaporpop” “channelpop” “ukpop” “christianpop” “icelandicpop” “gauzepop” “swedishsynthpop” “grungepop” “skyroom” “cumbiapop” “portuguesepop” “etherpop” “indonesianpop” “bedroompop” “germanpop” “danishpop” “polynesianpop” “noisepop” “ukalternativepop” “azeripop” “classicdanishpop” “classicswedishpop” “greenlandicpop” “popping” “newwavepop” “europop” “hip-pop” “polishpop” “koreanpop” “regionalmexicanpop” “czechpop” “australianpop”

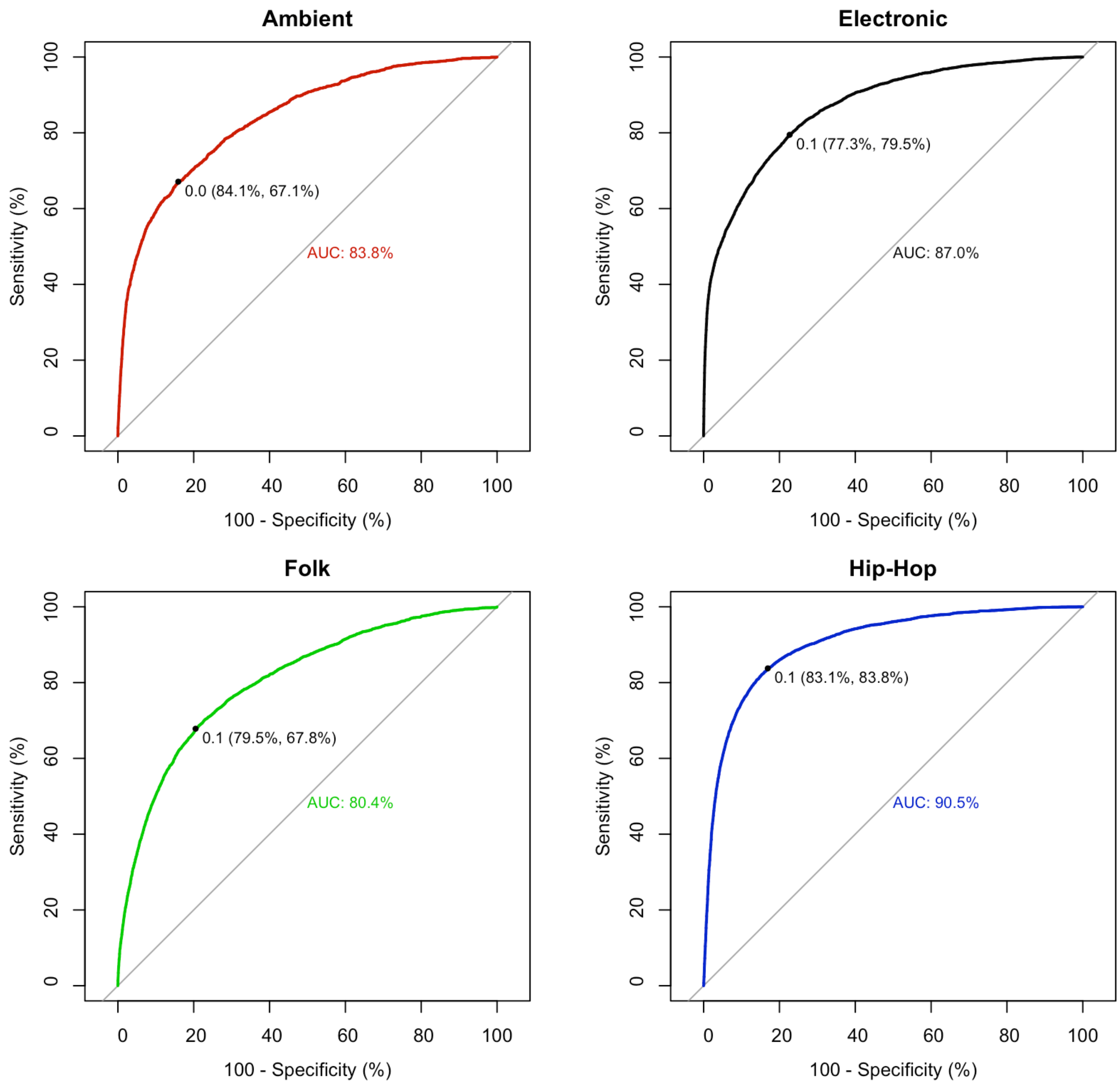


Figure 3: ROC Curves MLR

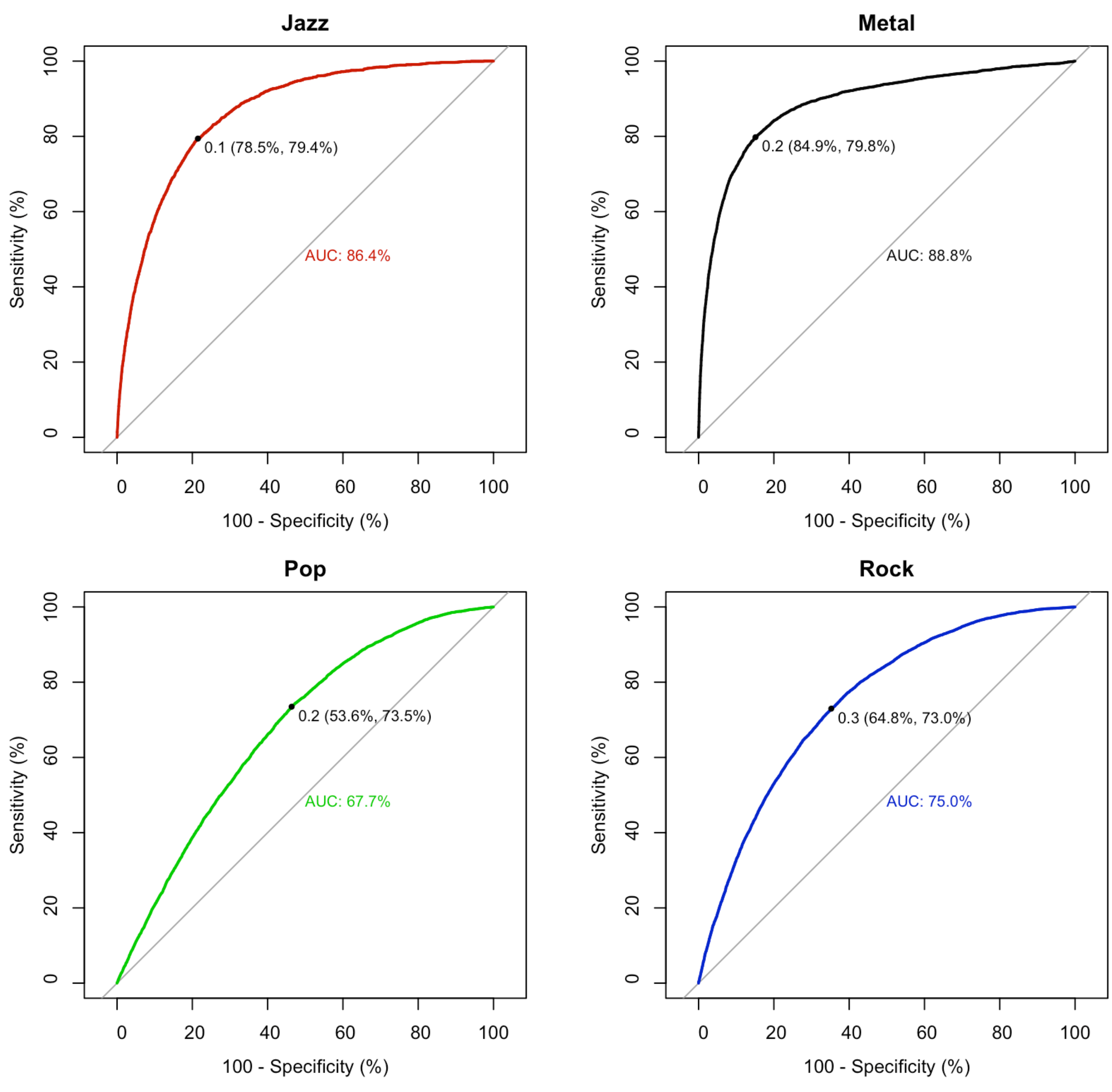


Figure 4: ROC Curves MLRs

Coefficients:

	(Intercept)	Key	Mode	Danceability	Energy	Loudness	Speechness	Acousticness	Instrumentalness	Valence	time_signature	Tempo
Electronic	-5.0151577	0.0107486632	0.2710437	6.933428	1.679622	0.15831497	1.0519155	-0.98405220	-0.02354597	0.6162628	0.26992305	0.0074675387
Folk	5.0291788	0.0103372142	0.5184283	-6.076086	-3.633039	0.20033129	-0.1761172	1.74985589	-1.74025071	5.5913866	0.04170568	0.0019060761
Hiphop	-0.8221329	-0.0028705034	0.2726860	3.947724	-2.021926	0.29055707	11.0281823	-0.30405027	-3.08076266	1.7730497	0.50955755	0.0006824062
Jazz	0.3185808	-0.0002777096	0.2330363	-1.303563	-2.202049	0.09107464	0.7416136	1.33205117	-0.21812887	5.4052476	0.08791880	-0.0020435361
Metal	4.0650940	0.0045451307	0.2925997	-6.523404	3.155629	0.13186877	0.4292019	-1.79009699	-1.32000179	0.4685665	-0.05593053	-0.0003894877
Pop	3.0833796	0.0129209482	0.7197714	-1.426690	-2.158275	0.21418785	-0.1503453	-0.03730723	-2.53136177	3.2773621	0.25399721	0.0048909212
Rock	4.5568091	0.0140904823	0.7906081	-4.552421	-1.466763	0.18340584	-2.6671776	-0.35736391	-2.46474836	3.5905596	0.30176948	0.0051825049

Std. Errors:

	(Intercept)	Key	Mode	Danceability	Energy	Loudness	Speechness	Acousticness	Instrumentalness	Valence	time_signature	Tempo
Electronic	0.3253064	0.008799571	0.06271533	0.2059148	0.1972346	0.010524860	0.2537030	0.1363394	0.09183428	0.1583254	0.08122423	0.001297621
Folk	0.2205174	0.009242453	0.06620280	0.2071300	0.2106360	0.008970677	0.3028631	0.1274981	0.09819034	0.1755102	0.05784281	0.001173986
Hiphop	0.3322577	0.009290854	0.06645666	0.2187196	0.2156281	0.011758431	0.1659503	0.1377541	0.12647487	0.1700971	0.08758597	0.001256132
Jazz	0.2928437	0.009967551	0.07092533	0.2233221	0.2342479	0.009671890	0.3490485	0.1365821	0.10098071	0.1843659	0.06897033	0.001324499
Metal	0.2301426	0.008919314	0.06356845	0.1943899	0.2032703	0.010044759	0.2804929	0.1529622	0.09209990	0.1685362	0.06016577	0.001165485
Pop	0.2115707	0.008520711	0.06107698	0.1843019	0.1860089	0.009103167	0.2028648	0.1189598	0.09328565	0.1588468	0.06201077	0.001139934
Rock	0.1739427	0.008198848	0.05856512	0.1736154	0.1730570	0.008105870	0.1965570	0.1129008	0.08556930	0.1543300	0.05537359	0.001089035

Residual Deviance: 123411.8
AIC: 123579.8

Figure 5: Summary of MLR Model

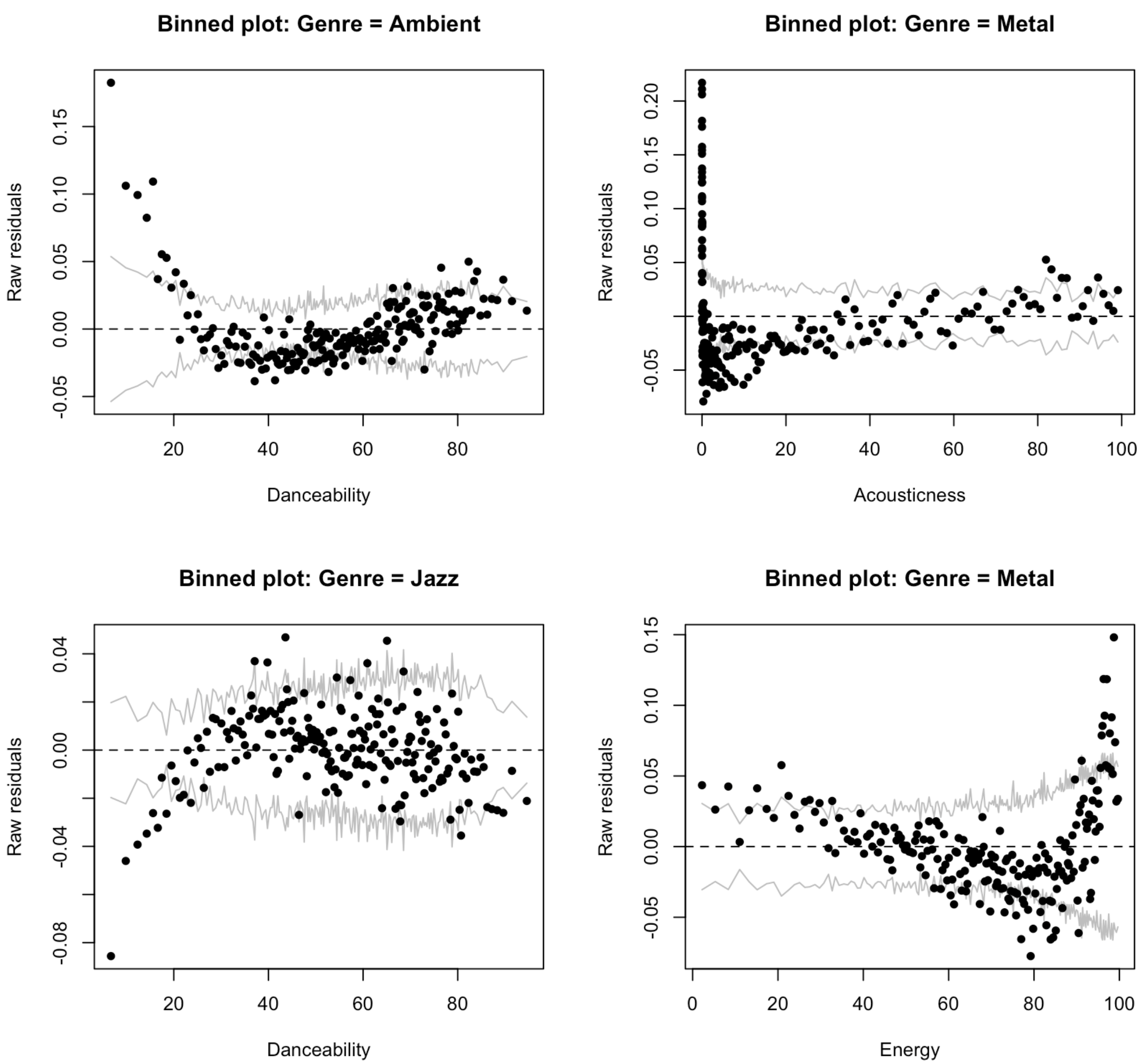
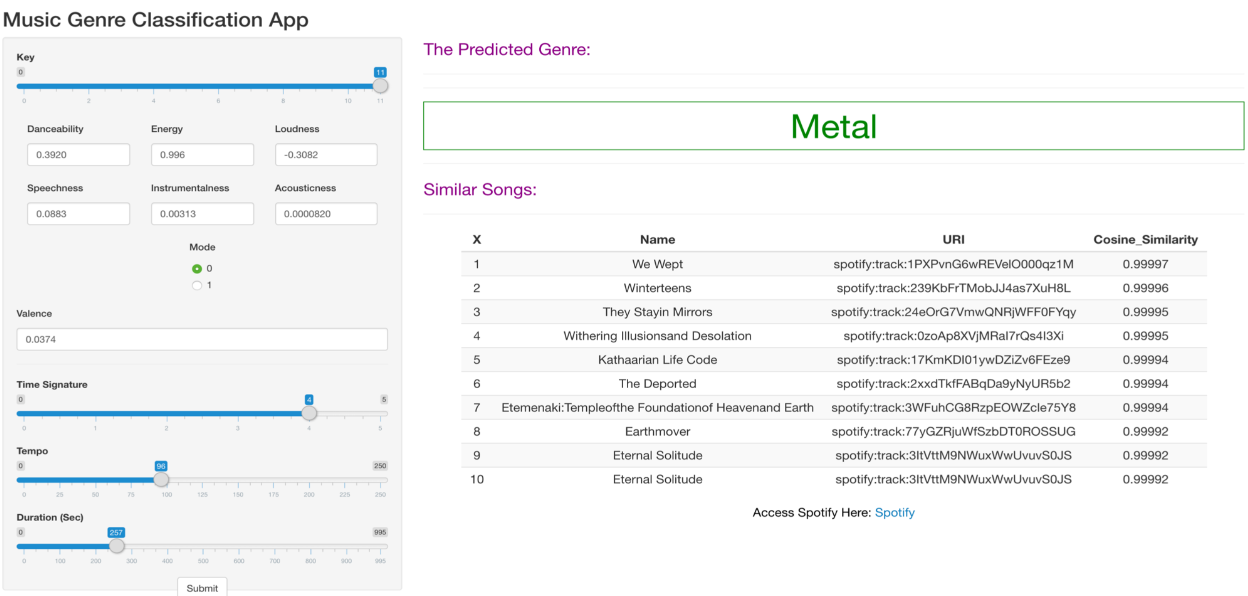


Figure 6: Few MLR Binned Residual Plots



X	Name	URI	Cosine_Similarity
1	We Wept	spotify:track:1PXPvnG6wREVelO000qz1M	0.99997
2	Winterteens	spotify:track:239KbFrTMobJJ4as7XuH8L	0.99996
3	They Stayin Mirrors	spotify:track:24eOrG7VmwQNRjWFF0FYqy	0.99995
4	Withering Illusionsand Desolation	spotify:track:0zoAp8XVjMRaI7rQs4I3Xi	0.99995
5	Kathaarian Life Code	spotify:track:17KmKDI01ywDZiZv6FEze9	0.99994
6	The Deported	spotify:track:2xxdTkfFABqDa9yNyUR5b2	0.99994
7	Etemenaki:Templeofthe Foundationof Heavenand Earth	spotify:track:3WFuhCG8RzpEOWZcle75Y8	0.99994
8	Earthmover	spotify:track:77yGZRjuWfSzbDT0ROSSUG	0.99992
9	Eternal Solitude	spotify:track:3ItVttM9NWuxWwUvuvS0JS	0.99992
10	Eternal Solitude	spotify:track:3ItVttM9NWuxWwUvuvS0JS	0.99992

Figure 7: Application Screenshot