MBDS Project – Spotify Song Recommendation System

Zilin Yin

Abstract

This study investigates the relationship between the audio features of a song and whether that song is liked or not. As a result of the study, some relationships do exist. Take me as an example, the danceability and acousticness are the two most influential factors that decide whether or not I would like a song, and I tend to like songs with less danceability and more acousticness. This is probably because I'm not a fan of dancing, and so high danceability does a negative effect on me. For different people, the relationships can be different because different people have different taste toward music. Overall, knowing that the relationships exist, people may use them for song recommendation systems, which is to recommend songs to people based on the collected data of the audio features of the songs that they used to listen to.

Hence the next part of the study investigates the performances of audio feature-based song recommendation systems based on different models. The first one is the 1st order logistic regression model, the second one is the 2nd order logistic regression model and the last one is the ensembled random forest model. Turns out that the ensembled random forest model is the one that shows the relationships between the audio features of a song and whether that song liked or not the best. Besides, for the accuracy of the predictions, the ensembled random forest model also has the highest score, with around 75% accuracy. Therefore, we may conclude that the ensembled random forest should be the best model for the song recommendation system among the three evaluated.

1. Introduction

This goal of this study is to investigate the relationship between the audio features of a certain song and whether or not the song would be liked by the target person, and if there are some relationships, how much can people depend on them for the song prediction model. The importance of this study, on a large scale, if such relationships exist, they can be extremely helpful for music companies like Spotify, ITUNES, etc. to recommend music to their customers that are more attached to their taste. In this way, they may attract more customers than before and gain huge amount of profit. For myself in person, I can use the fitted model to predict the likelihood of whether or not I would like a newly published song, such that I could save a long time from listening to all those newly published songs one by one.

2. Related Works

The oldest way that music companies use to recommend music to their customers is by using collaborating filtering models, which is to recommend music based on a set of customers' behaviors (Abhinav Ajitsaria, 2019). For example, after collecting a set of behaviors, the system would recommend to a person the songs that are often listened to by other people who are considered by the system having similar behaviors to that person. The problem with this method is that the prediction system would have to have a large amount of collected data on people's behaviors, which is relatively complicated.

Another way that music companies use is to recommend songs based on the contexts of the songs, which is strongly related to the idea of natural language processing. And the corresponding process flow is as follow:

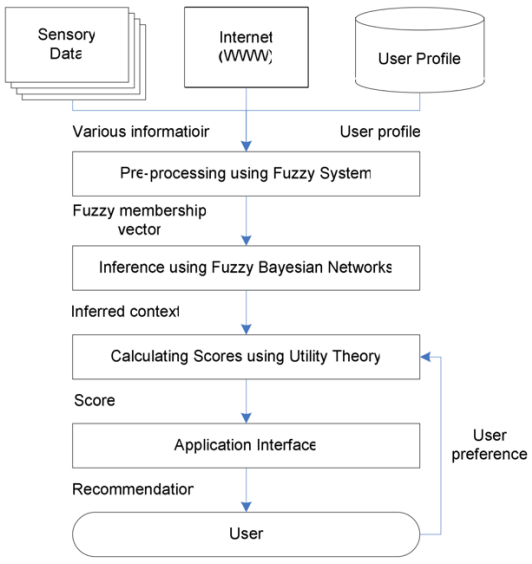

Figure 2.1. The process in context-aware music recommendation system (Han-Saem Park, Ji-Oh Yoo, Sung-Bae Cho, 2006)

Compared to the collaborating filtering models mentioned before, the NLP model requires fewer conditional data, just the text contents of the songs. But the problem is that when dealing with songs with various languages, or songs in minority language, NLP model may have issues classifying them and make predictions.

This leads to the last way that music companies tend to use to recommend songs, which is to recommend the same types to the songs that their customers based on the audio features of the songs.

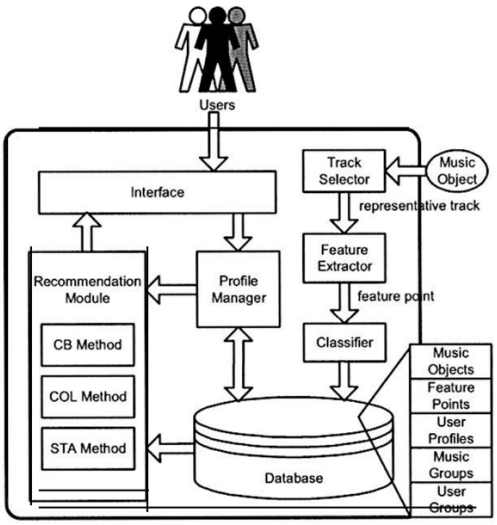

Figure 2.2. The system architecture of the music recommendation system (Hung-Chen Chen, Arbee L.P. Chen, 2001)

This is actually similar to the idea of this study. However, for this study, more audio features would be considered and thus there would be more variables in the model.

Hence, the study would first investigate the relationships between several different song audio features including danceability, energy, key, loudness, mode, speechiness, acousticness, instrumentalness, liveness, valence, tempo, duration (ms) and time signature. Then implement different methods for the prediction of liked songs based on the audio features. And finally, make the corresponding analysis based on the predictions made from different methods and see which one yields better results.

3. Data

There are two sets of data for this study. The first set of data includes the Spotify top 100 songs of 2018 and the second set of data includes the Spotify top 50 songs of 2019, as well as sets of their audio features. The data are collected from Kaggle. And the original data sources are Spotify Web API and Spotify Python library. Both sets of data are in csv format. Followings are overviews of the two datasets:

retop2018

id	name	artists	danceability	energy	key	loudness	mode	speechiness	acousticness	instrumentalness	liveness	valence	tempo	duration_ms	time_signature	Like
6DCZcSspjsKoFjzjrWoCd	God's Plan	Drake	0.754	0.449	7	-9.211	1	0.109	0.0332	8.29E-05	0.552	0.357	77.169	198973	4	1
3ee8Jmje8o58CHK66QrVC	SAD!	XXXTENTACION	0.74	0.613	8	-4.88	1	0.145	0.258	0.00372	0.123	0.473	75.023	166606	4	1
0e7ipj03S05BNilyu5bRz	rockstar (feat. 21 Savage)	Post Malone	0.587	0.535	5	-6.09	0	0.0898	0.117	6.56E-05	0.131	0.14	159.847	218147	4	1
3swc6WTsr7rl9DqQKQA55	Psycho (feat. Ty Dolla $ign)	Post Malone	0.739	0.559	8	-8.011	1	0.117	0.58	0	0.112	0.439	140.124	221440	4	1
2G7V7zsVDxg1yRsu7Ew9R	In My Feelings	Drake	0.835	0.626	1	-5.833	1	0.125	0.0589	6E-05	0.396	0.35	91.03	217925	4	1
7dt6x5M1jzdTEt8oCbisTT	Better Now	Post Malone	0.68	0.563	10	-5.843	1	0.0454	0.354	0	0.136	0.374	145.028	231267	4	1
58q2HKrzhC3ozto2nDdN4	I Like It	Cardi B	0.816	0.726	5	-3.998	0	0.129	0.099	0	0.372	0.65	136.048	253390	4	0
7ef4DlsgrMEH11cDZd32M	One Kiss (with Dua Lipa)	Calvin Harris	0.791	0.862	9	-3.24	0	0.11	0.037	2.19E-05	0.0814	0.592	123.994	214847	4	0
76cy1WJvNGJTj78UqeA5z	IDGAF	Dua Lipa	0.836	0.544	7	-5.975	1	0.0943	0.0403	0	0.0824	0.51	97.028	217947	4	1

Table 3.1. Spotify top100 2018 sample

There are 100 observations and 17 columns for this dataset.

top2019

	Track.Name	Artist.Name	Genre	Beats.Per.Minute	Energy	Danceability	Loudness..dB..	Liveness	Valence.	Length.	Acousticness..	Speechiness.	Popularity	Like
1	Senorita	Shawn Mendes	canadian pop	117	55	76	-6	8	75	191	4	3	79	1
2	China	Anuel AA	reggaeton flow	105	81	79	-4	8	61	302	8	9	92	0
3	boyfriend (with Social House)	Ariana Grande	dance pop	190	80	40	-4	16	70	186	12	46	85	0
4	Beautiful People (feat. Khalid)	Ed Sheeran	pop	93	65	64	-8	8	55	198	12	19	86	1
5	Goodbyes (Feat. Young Thug)	Post Malone	dfw rap	150	65	58	-4	11	18	175	45	7	94	1
6	I Don't Care (with Justin Bieber)	Ed Sheeran	pop	102	68	80	-5	9	84	220	9	4	84	1
7	Ransom	Lil Tecca	trap music	180	64	75	-6	7	23	131	2	29	92	0
8	How Do You Sleep?	Sam Smith	pop	111	68	48	-5	8	35	202	15	9	90	0
9	Old Town Road - Remix	Lil Nas X	country rap	136	62	88	-6	11	64	157	5	10	87	0
10	bad guy	Billie Eilish	electropop	135	43	70	-11	10	56	194	33	38	95	1

Table 3.2. Spotify top50 2019 sample

There are 50 observations and 14 columns for this dataset.

Note that the last column 'Like' in the dataset doesn't exist in the original dataset collected from Kaggle, and the 'Like' column is consisted of binary data, specifically, 1 means the song is liked, 0 means the song is not liked. The data in table1 is considered as the training data for this study and the data in table 2 is considered as the hold-out data or test data for the study. For the 2018 dataset, the 'Like' column is filled out by me in person, which means I'm taking myself as the sample of this study, whereas the 'Like' column for the 2019 data is filled out by a friend of mine. In such a way we can use the fitted model based on my personal preferences of songs on hold-out data. And we could see how it performs when it is implemented on a set of unseen data.

As shown in the tables, the values for different elements can be very different, and we need to normalize the data. The formula used for the normalization is as follow:

$$x^* = \frac{x-\mu_x}{\sigma_x} \qquad (3.1)$$

Where x* is normalized data and x is the original data.

Also, not all parameters in the dataset are valid to be put into the model. To be more specific, the id, name, artists columns in the 2018 dataset and the track name, artist and genre columns in the 2019 dataset are taken out when fitting the model as they are not considered as audio features. Lastly, when implementing logistic model and random forest on hold out data, the number of parameters used is less than before as the parameters of the 2019 data is partially different from that of the 2018 data, therefore, only the parameters that are in both dataset were taken into account when doing the hold-out data analysis. To be more specific, the parameters are energy, danceability, loudness, liveness, valence, acousticness and speechiness.

4. Methodology

The two methods for the study are logistic model for prediction, and machine learning algorithm, ensemble random forest specifically. The logistic model is used because whether a song is liked or not is represented by 0 and 1, which is binary data, and so logistic model may potentially be a good fit for this kind of binary data. For the logistic model, a graphical model diagram of it is as follow:

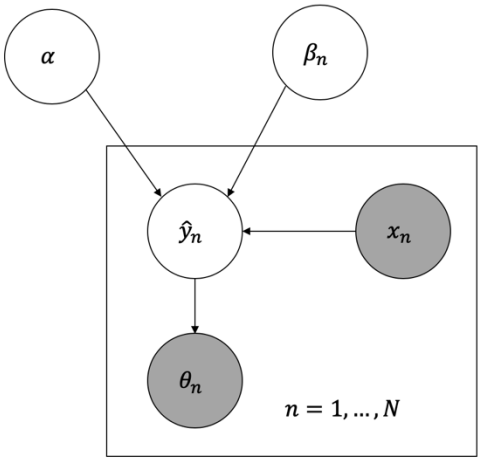

Figure 4.1. Graphical diagram of the logistic regression model

The figure 3 above shows how the logistic regression model for this study works. Where. the following priors are assumed:

$$\alpha \sim N(0, 1) \qquad (4.1)$$
$$\beta \sim N(0, 1) \qquad (4.2)$$

And the corresponding linear model:

$$\hat{y} = \beta x + \alpha \qquad (4.3)$$

And the sigmoid function for likelihood:

$$y = \frac{1}{1+e^{-(\beta x+ \alpha)}} \qquad (4.4)$$

Besides the simple logistic linear regression model, a nonlinear regression (2nd order) logistic model would be evaluated and see if the performance of the model would get better as compared to the linear one. For the nonlinear model, other than the variates as in the linear model, input data would include the multiple of the variates such that more interactions between variates are added to the model. The output of the logistic regression model should be n (denoted as the number of observations) values within the range from 0 to 1, representing the likelihood of the song is liked for each observation. One limitation of the logistic regression is that it tends to overfit, which means it sometimes overstates the accuracy of its predictions (Nick Robinson, 2018). For this study, the input of the model would be the Spotify top100 2018 dataset. And the outcome parameters would be implemented on the Spotify top50 2019 dataset.

For the ensemble random forest algorithm, the Spotify top100 2018 dataset is used as the training dataset, and the Spotify top50 2019 dataset is used as the test data set. The optimal number of decision trees is found by calculating the out of bag scores starting from 1 decision tree to 100 decision trees, and the number of decision trees with the highest score is used when fitting the ensemble random forest model. Once the model is fitted, it's implemented on the test dataset.

Once we have the outcomes of the above two different methods, the corresponding analysis could be made based on the outcomes, including the comparison of the accuracy, comparison of the relationships between the variables and the outcomes, etc.

5. Results

5.1. 1st order logistic regression model:

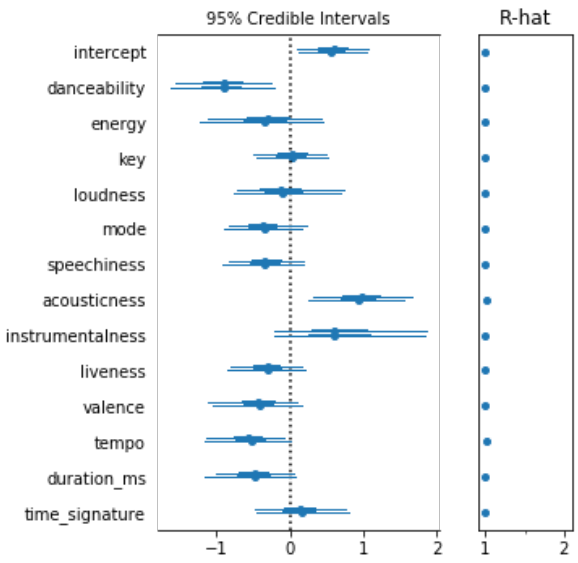

Figure 5.1. 1st order logistic regression model forest plot

Accuracy score	0.73
F1 score	0.7692307692307692
WAIC	130.07316394978048

Table 5.1. 1st order logistic regression mode scores

Parameters	Value	$\hat{R}$
α- intercept	0.5805420565161633	1.000747
β_1- danceability	-0.9161305309526839	0.999495
β_2- energy	-0.32777996833742473	1.001399
β_3- key	0.03825366789091607	0.999215
β_4- loudness	-0.10358210504978348	0.999763
β_5- mode	-0.3621347452405544	0.999065
β_6- speechiness	-0.33757658641703703	0.999005
β_7- acousticness	0.9668000775115154	1.003974
β_8- instrucmentalness	0.7211801819360918	0.999120
β_9- liveness	-0.3092883901582177	0.999169
β_{10}- valence	-0.4315530916747307	0.999130
β_{11}- tempo	-0.553741612036561	1.002627
β_{12}- duration (ms)	-0.492603113923586	1.000524
β_{13}- time signature	0.1284227038860472	0.999065

Table 5.2. 1st order logistic model parameters

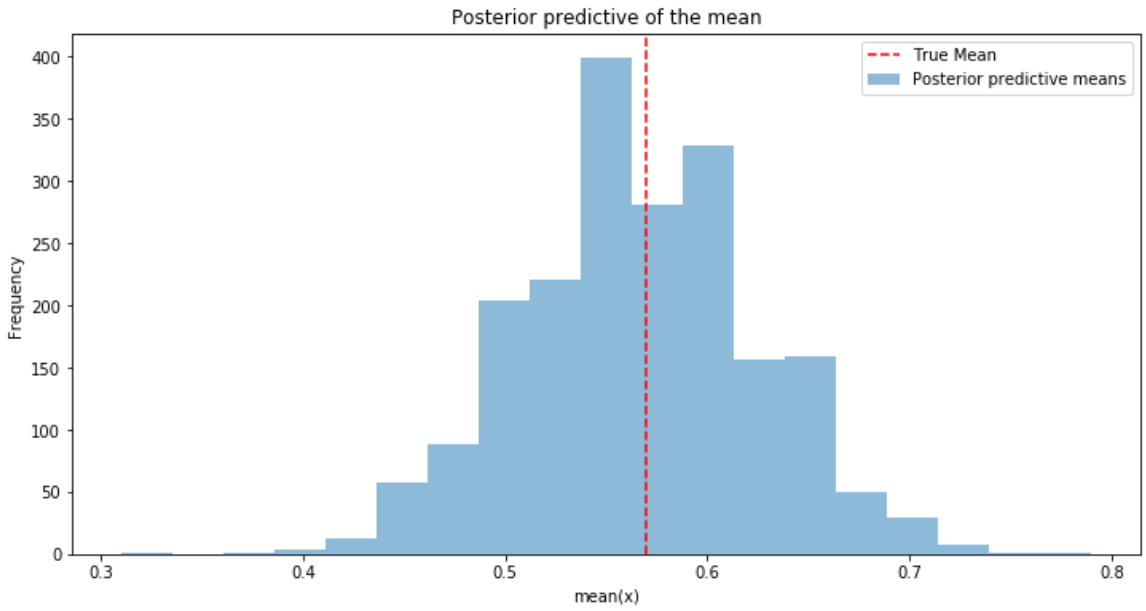

Figure 5.2. 1st order logistic regression mode posterior predictive of the mean

5.2. 2nd order logistic regression model:

Accuracy score	0.76
F1 score	0.8
WAIC	124.7434056186990

Table 5.3. 2nd order logistic regression. model scores

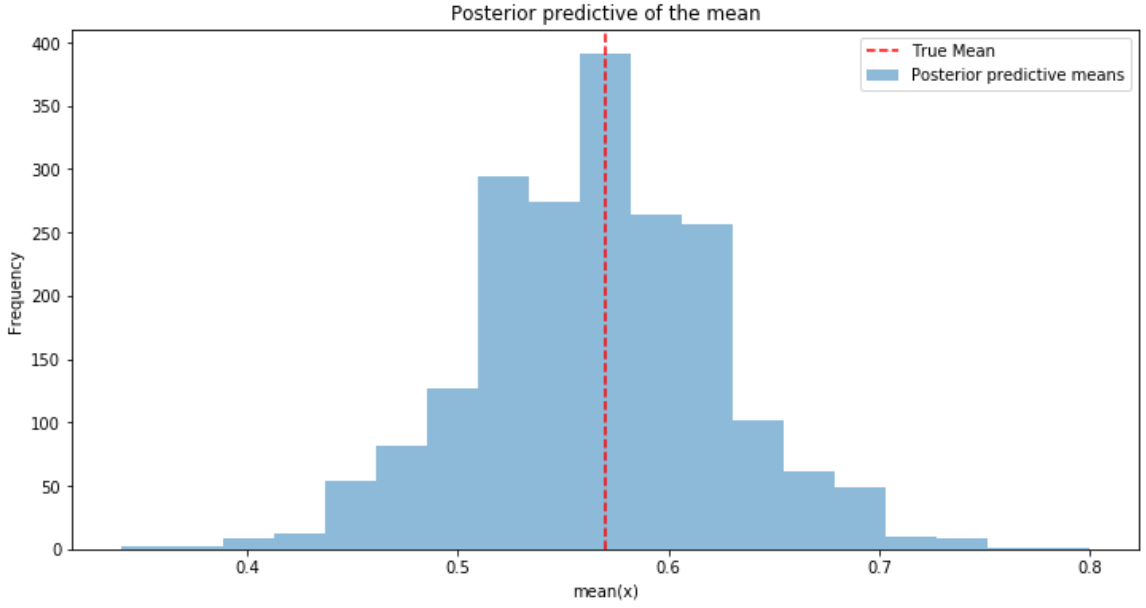

Figure 5.3. 2nd logistic regression model posterior predictive of the mean

Refer to the Appendix for the forest plot and the table of parameters.

5.3. Ensembled random forest:

Accuracy score	0.7666666666666667

Table 5.4. Ensembled random forest score

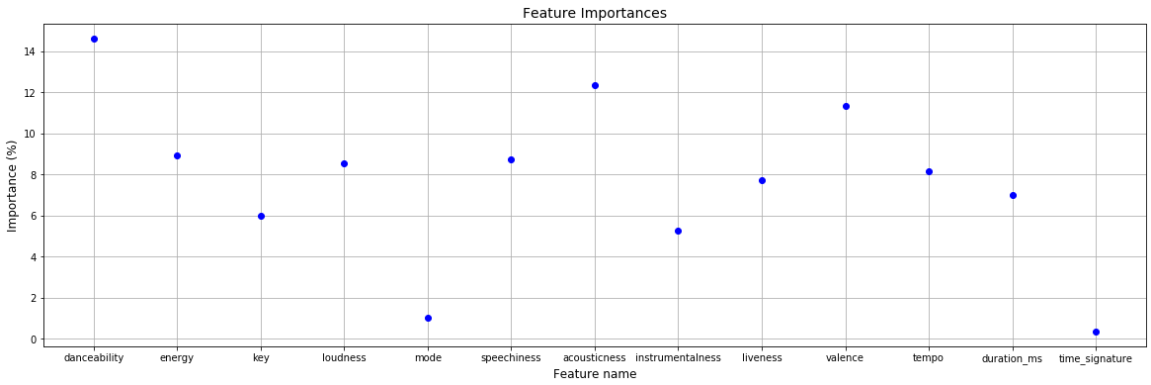

Figure 5.4. Ensembled random forest feature importance

5.4. Hold-out data:

Accuracy score	0.58

Table 5.5. 1st order logistic regression score

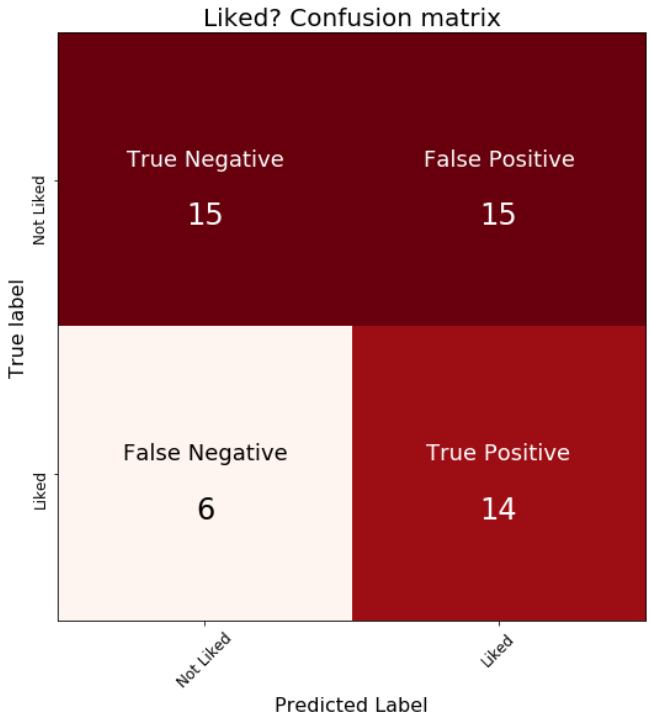

Figure 5.5. 1st order logistic regression model predictions check

Accuracy score	0.62

Table 5.6. 2nd order logistic regression model score

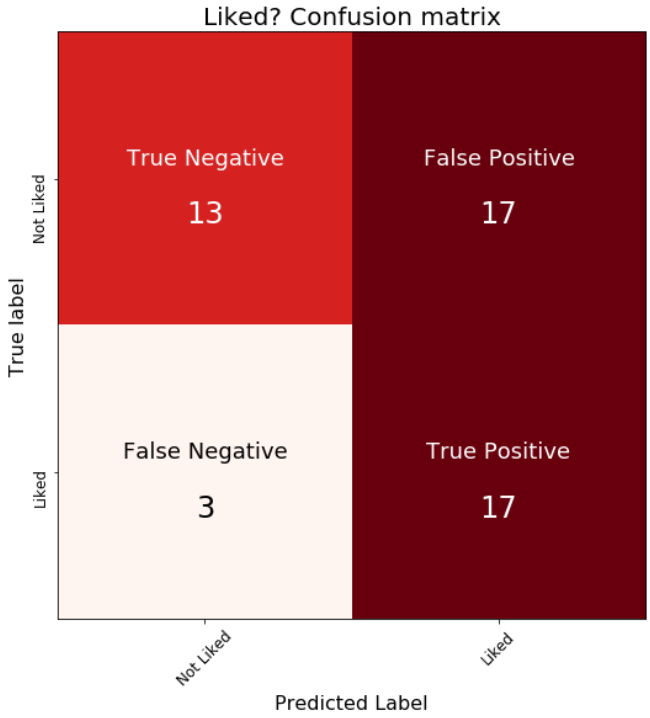

Figure 5.6. 2nd order logistic regression model predictions check

Accuracy score	0.74

Table 5.7. Ensembled random forest score

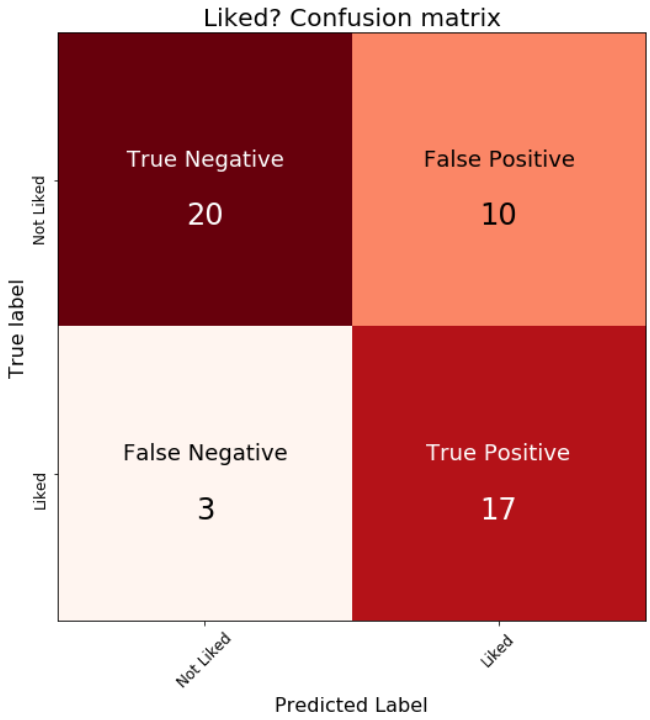

Figure 5.7. Ensembled random forest predictions check

6. Discussion

Based on figure 5.1, figure 5.4 and figure 7.1, we find out that danceability and acousticness are two most influential factors, which can be indicated by both figures, as the value of danceability increases, the likelihood of the song been liked decreases and as the. Value of acousticness increases, the likelihood of the song been liked also increases. How other parameters influence the likelihood can also be seen from the forest plot (figure 5.1 and figure 7.1) and the parameters table (table 5.2 and table 7.1). By looking at the forest plot, if the value is at the right part, then it's positively related, if it is at the left part of the plot, then it is negatively related. By looking at the parameters table, if the value of β can be regarded as how the corresponding parameter influence the likelihood of whether the song is liked or not, the larger the value of β, the more influential the corresponding parameter is, and a positive value means a positive relationship and a negative value means a negative relationship.

However, for some specific parameters, the importance seems to be inconsistent between the three. For example, the 'key' parameter, in logistic model it's the less important factor, whereas in random forest it is not. And the 'mode' and 'time signature' parameters are the two least important factors in random forest, whereas in logistic model, they are not. This indirectly shows that random forest may be better than logistic regression model for the study because the key of a song, thinking practically, shouldn't be the least important factor influencing whether that song is liked or not.

Then based on table 5.1 and table 5.3, we can see obviously that the scores of the 2nd order logistic model are higher than the 1st order logistic model, but not by a huge amount. Figure 5.2 and figure 5.3 show the posterior predictive of the mean visually for each model, we can see that figure 5.3, the plot of the 2nd order logistic model seems more fitted to the true mean compared to the 1st order logistic model. We can also compare the accuracy score of the random forest model with the other two. And as a result, the accuracy score of the random forest is the highest among the three.

As for the hold-out data, based on tables 5.5, 5.6 and 5.7, which are the accuracy score of the 1st order logistic model, 2nd order logistic model and ensembled random forest model on hold-out data respectively. And we can see that compared to the case before, the accuracy score of the random forest model is still the highest, but this time, the score is higher than that of the other two by a large amount. Figure. 5.5, 5.6 and 5.7 show the predictions of the three models visually.

These show that the relationship between audio features of a song and whether that song is liked or not does exist, and it may vary from person to person. Music companies could recommend new songs to their customers, based on the collected data of the audio features of the songs they used to listen to.

7. Conclusion

Based on the results of the study, we conclude that danceability and acousticness are the two most important factors that influence whether or not I would like a song, where danceability is negatively related and acousticness is positively related. For the accuracy of the predictions, the ensembled random forest model has the best performance, as well as for the hold out data, with a prediction accuracy of around 75%. Whereas the logistic models seem to have problems with the relationship between the parameters and the likelihood of whether a song is liked or not. And when dealing with hold-out data, the scores of 1st and 2nd order logistic regression models are low, as shown in table 5.5 and table 5.6. Therefore, the ensembled random forest model is considered the best model for the song recommendation system among the three evaluated.

8. Appendix

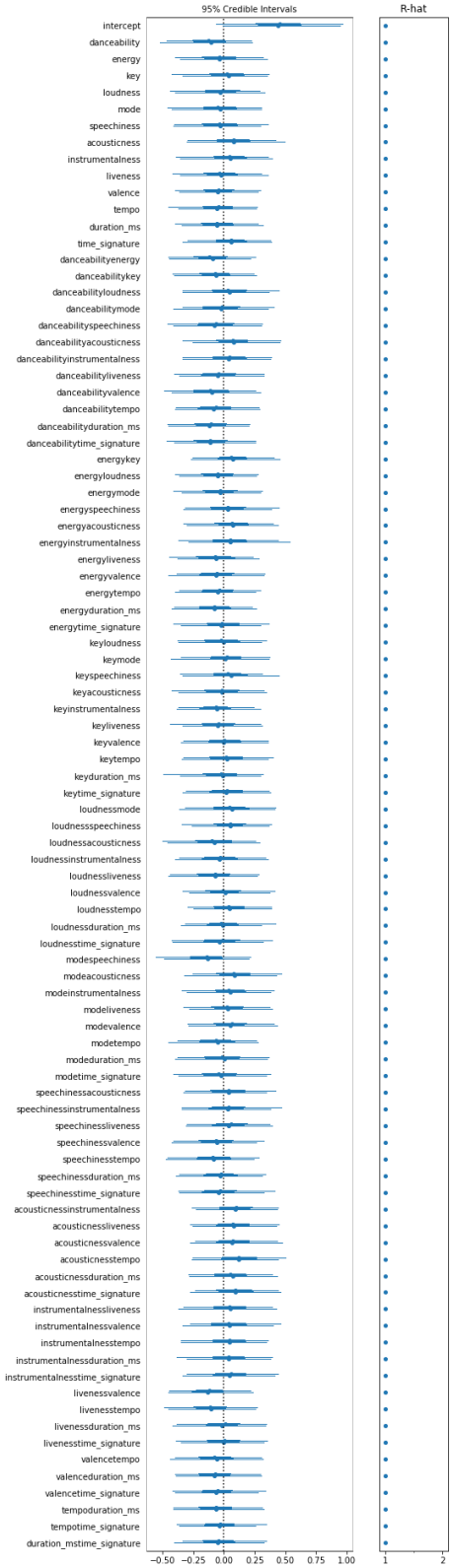
95% Credible Intervals
R-hat
intercept
danceability
energy
key
loudness
mode
speechiness
acousticness
instrumentalness
liveness
valence
tempo
duration_ms
time_signature
danceabilityenergy
danceabilitykey
danceabilityloudness
danceabilitymode
danceabilityspeechiness
danceabilityacousticness
danceabilityinstrumentalness
danceabilityliveness
danceabilityvalence
danceabilitytempo
danceabilityduration_ms
danceabilitytime_signature
energykey
energyloudness
energymode
energyspeechiness
energyacousticness
energyinstrumentalness
energyliveness
energyvalence
energytempo
energyduration_ms
energytime_signature
keyloudness
keymode
keyspeechiness
keyacousticness
keyinstrumentalness
keyliveness
keyvalence
keytempo
keyduration_ms
keytime_signature
loudnessmode
loudnessspeechiness
loudnessacousticness
loudnessinstrumentalness
loudnessliveness
loudnessvalence
loudnesstempo
loudnessduration_ms
loudnesstime_signature
modespeechiness
modeacousticness
modeinstrumentalness
modeliveness
modevalence
modetempo
modeduration_ms
modetime_signature
speechinessacousticness
speechinessinstrumentalness
speechinessliveness
speechinessvalence
speechinesstempo
speechinessduration_ms
speechinesstime_signature
acousticnessinstrumentalness
acousticnessliveness
acousticnessvalence
acousticnesstempo
acousticnessduration_ms
acousticnesstime_signature
instrumentalnessliveness
instrumentalnessvalence
instrumentalnesstempo
instrumentalnessduration_ms
instrumentalnesstime_signature
livenessvalence
livenesstempo
livenessduration_ms
livenesstime_signature
valencetempo
valenceduration_ms
valencetime_signature
tempoduration_ms
tempotime_signature
duration_mstime_signature
−0.50
−0.25
0.00
0.25
0.50
0.75
1.00
1
2

Figure 7.1. 2nd order logistic regression model forest plot

Parameters	Value	$\hat{R}$
α- intercept	0.44440612485366254	0.9991490884766311
β_1- danceability	-0.12411918635371279	0.9990650271139498
β_2- energy	-0.03208069092728778	0.9990082519484029
β_3- key	0.01610060544858965	0.9991528727426543
β_4- loudness	-0.018268545231546498	0.9990894848485539
β_5- mode	-0.04126420808812514	1.0008068509876475
β_6- speechiness	-0.03169912046572077	0.9990164041233117
β_7- acousticness	0.08284750582145534	0.9993223859403115
β_8- instrucmentalness	0.03727498946057921	0.9990117631906972
β_9- liveness	-0.028242465863717283	0.9990140652855237
β_{10}- valence	-0.038843237357282234	0.9990033301389321
β_{11}- tempo	-0.050791903157612256	0.9991210991043329
β_{12}- duration (ms)	-0.0593908997838544	0.9991979974945729
β_{13}- time signature	0.046934814047239704	0.9992324375424594
β_{14}	-0.11263107455801753	0.9997334909916001
β_{15}	-0.0794658366104305	0.9994350914146437
β_{16}	0.04979166515502661	0.9998917608154246
β_{17}	-0.0312715225445348	0.9992705422061168
β_{18}	-0.07071442155438327	0.9991726418819407
β_{19}	0.06763292083921599	0.9990055611352879
β_{20}	0.035096137329501516	0.9999356351377053
β_{21}	-0.036403846265075396	0.9991372981422622
β_{22}	-0.11195415121891128	0.9990116025440342
β_{23}	-0.07944527612375714	0.999016651447694
β_{24}	-0.11619814223834879	0.9991472741832688
β_{25}	-0.11261255052401833	0.9992226260823559
β_{26}	0.06165677896338652	0.9989995365920135
β_{27}	-0.04823445192833902	0.9999008901808004
β_{28}	-0.032166269198715215	0.9990455322241273
β_{29}	0.022885658840658026	0.9989995591646175
β_{30}	0.05840726207553523	0.9993340237446313
β_{31}	0.054122450496107546	1.000023671132261
β_{32}	-0.07900770307578499	0.9993604256557946
β_{33}	-0.06196591974445651	0.999029072176639
β_{34}	-0.04904388875868855	0.9992014977764578
β_{35}	-0.07610659548486398	1.0001074957584055
β_{36}	-0.022053632608011024	0.9990116324580466

β_{37}	-0.01634404100458539	0.9992615899240704
β_{38}	0.011734597356530607	1.00284558110954
β_{39}	0.03843302539553689	0.9990956086639696
β_{40}	-0.01351146924843893	0.9990226843846846
β_{41}	-0.0722876765950719	0.9992061480203817
β_{42}	-0.04266906428457604	0.9991225922574792
β_{43}	-0.002117215188949759	1.0003325781915857
β_{44}	0.02017047257067585	0.9995630154048726
β_{45}	-0.01422511262792928	0.9995660740049012
β_{46}	0.018549708131006773	1.0019227003939262
β_{47}	0.047242950112689926	0.9990110915705279
β_{48}	0.04939758024338588	0.9990234287318339
β_{49}	-0.08316587344091589	0.9993030304870518
β_{50}	-0.02831501022694543	0.9994664432666293
β_{51}	-0.07595829108721285	0.9992510556794697
β_{52}	-0.006021003990473002	1.0010598969358457
β_{53}	0.042165415221483364	0.9992980971391552
β_{54}	-0.00284873174111149	0.9991116335340747
β_{55}	-0.026981926405194156	0.9989995122281486
β_{56}	-0.15227424912208118	0.9990009927009406
β_{57}	0.09202625159365244	0.999635483982514
β_{58}	0.034847758876961225	0.9990015197467409
β_{59}	0.032996793576050835	0.9992418314997183
β_{60}	0.0657717838344529	0.9990294366962893
β_{61}	-0.05450638988615982	0.9999384641090096
β_{62}	-0.01041862601171034	0.9994945945492003
β_{63}	-0.03252346371727399	0.9994072702656873
β_{64}	0.04145724897306382	0.9995656638890831
β_{65}	0.03763075483140155	0.9999596081368308
β_{66}	0.058424045785817426	1.0003727161491536
β_{67}	-0.06103476321117744	0.9990475523912515
β_{68}	-0.08239697999791554	0.9990529627042344
β_{69}	-0.030580693381575593	0.9990214683409523
β_{70}	-0.035210036698878955	0.999295062774806
β_{71}	0.09960939029000172	0.9992893857982985
β_{72}	0.08064558946416199	0.9996189389661264
β_{73}	0.08173616895714493	0.999496902941724
β_{74}	0.11845743308074524	0.999040366335331
β_{75}	0.057791036693432864	0.9994785617765432
β_{76}	0.09394929514214556	0.9990988014838713

β_{77}	0.04370809367541987	0.9990077384816991
β_{78}	0.04386398679608692	1.0020415776989033
β_{79}	0.04369771328187332	0.999087506233844
β_{80}	0.03648608360425073	0.9990032333531083
β_{81}	0.04462780487561325	0.9990503679943747
β_{82}	-0.1289508321481651	1.000501689535455
β_{83}	-0.09819426043296069	0.9996678436310692
β_{84}	-0.017953162560852713	0.9999952771792686
β_{85}	-0.010860532866817901	0.9995701031820963
β_{86}	-0.058637646675288195	1.0008493179599007
β_{87}	-0.08058449829110435	0.998999742139081
β_{88}	-0.046283025872467354	0.9995099627077099
β_{89}	-0.060935105181576237	0.9990825360473171
β_{90}	-0.031597119368228566	0.9991424034405588
β_{91}	-0.049596036876311485	1.000273839706787

Table 7.1. 2nd order logistic regression model parameters

9. References

Han-Saem Park, Ji-Oh Yoo, Sung-Bae Cho. (2006). *A Context-Aware Music Recommendation System Using Fuzzy Bayesian Networks with Utility Theory*. Springer Link.

Hung-Chen Chen, Arbee L.P. Chen. (2001). *A music recommendation system based on music data grouping and user interests*. dl.acm.org.

Abhinav Ajitsaria. (July 10, 2019). *Build a Recommendation Engine With Collaborating Filtering*. Retrieved from https://realpython.com/build-recommendation-engine-collaborative-filtering/

Nadin Tamer. (2019). *Top Spotify Tracks of 2018*. Retrieved from https://www.kaggle.com/nadintamer/top-spotify-tracks-of-2018

Leonardo Henrique. (2019). *Top 50 Spotify Songs* – 2019. Retrieved from https://www.kaggle.com/leonardopena/top50spotify2019

Kaggle.com. http://organizeyourmusic.playlistmachinery.com/

Nick Robinson. (2018). *The Disadvantages of Logistic Regression*. Retrieved from https://www.theclassroom.com/disadvantages-logistic-regression-8574447.html