Fernando Espinosa

PHY 250 - Econophysics

Radio Win Prediction: A Musical Stock Analysis

Do big music hits significantly influence the value of various radio broadcasting company (MSN, RAIO, ROIAK) stocks? Can we predict whether a song will become a hit? Based on a list of songs by up and coming artists with the most radio play, we will give a prediction – based on their musical characteristics (Spotify API) – on whether these songs will, after already being on the “Top 100” list of the year given by Billboard.com, stay on the list for at least one more week. A statistical machine learning algorithm from the Python library Sklearn will be implemented to make this prediction. A song is defined as a ‘hit’ by the amount of radio play it gets, using the (unofficial) Billboard API to see the songs with the most radio play in the past year and using Google Finance data to see value of those radio stocks at the time.

First, we must identify whether these hits do have a significant effect on the value of 3 radio stocks - NYSEMKT:MSN (Emerson Radio Corporation), OTCMKTS:RAIO (RadioIO Inc.), and NASDAQ:ROIAK (Radio One, Inc.). To do this, historical data on these stocks was downloaded from Google Finance in two ranges for each stock – 1. From a year before to the date of the song charting, and 2. A week after the song has charted to the present. This way, the difference in date between the two stock price ranges is only a week, so any significant difference in prices is more likely to be due to a greater number of hit songs. A ‘hit’ is being defined as a song that stays on the Billboard Top 100 chart for more than a week. At the time of December 31, 2016, the Billboard Top 100 songs of the year that stayed on said chart for more than a week numbered 81, a large fraction of the top 100. This number was found by the code:



Before these 81 songs charted, the mean stock price over the course of a year for the MSN at closing was $0.85. From the week after they charted to the present, the mean stock price for the MSN at closing increased to $1.16. These values were found by the code:

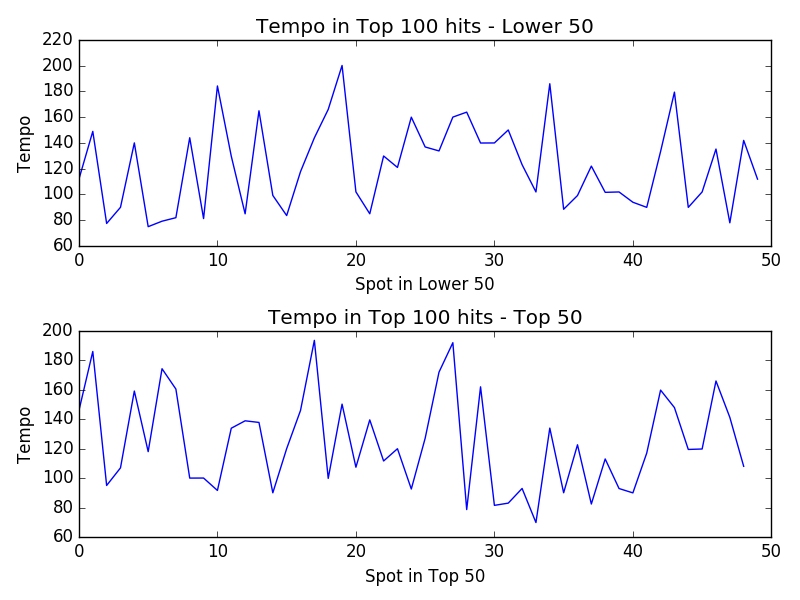
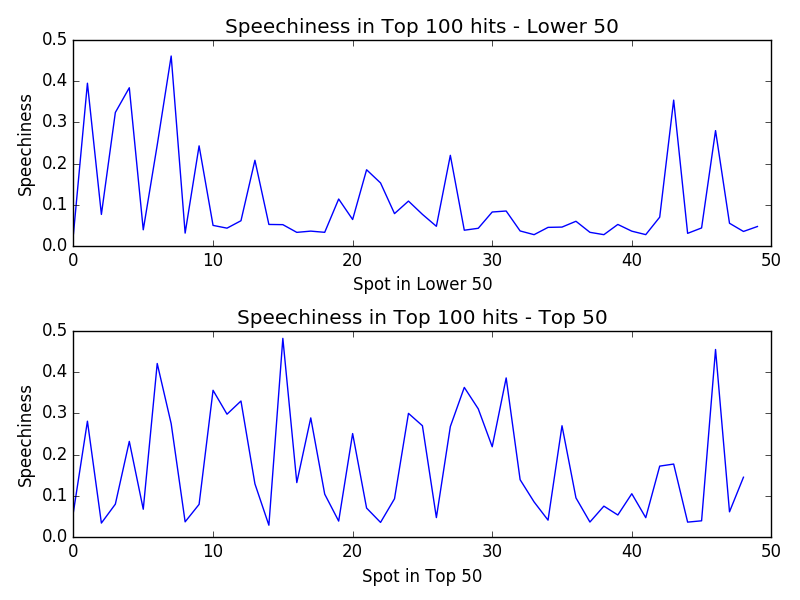
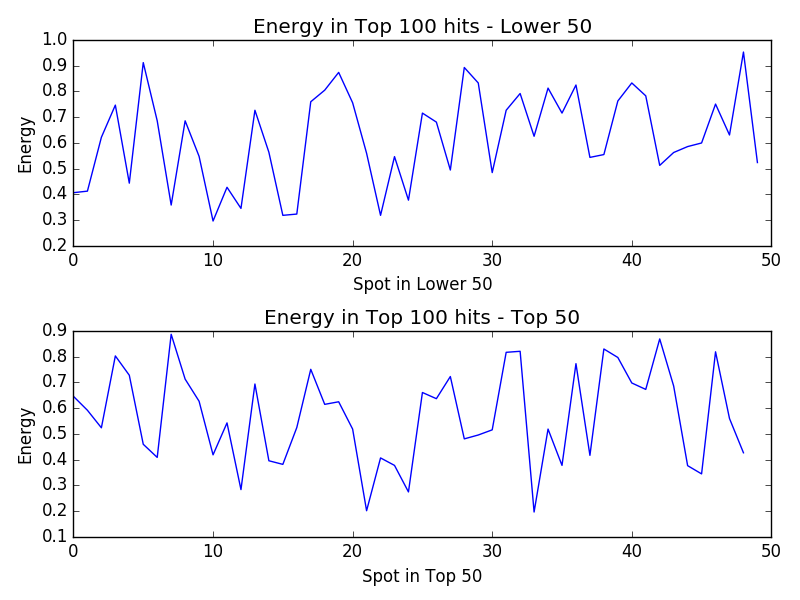
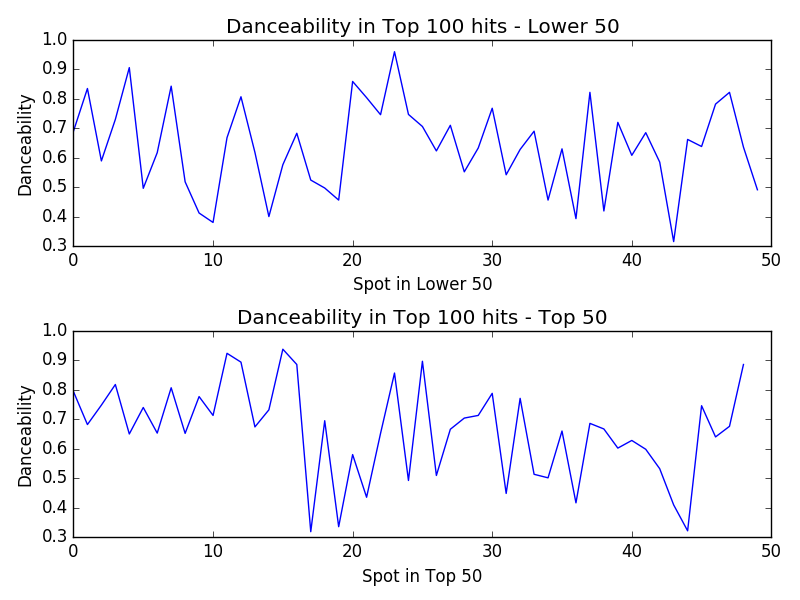


An increase of $0.30 cents (27%) is certainly statistically significant for this particular stock, but we will repeat this process for the other two. There may also be other factors at stake that are unknown, causing the radio stocks to go up, such as investments or the fact that the second range of values is in a new year. The calculated change in RAIO value was $-0.002: a 5% decrease. The calculated change in ROIAK value was $0.36: a 13% increase. Because the increases are 3-5x bigger than the one decrease that is bordering on statistical significance (RAIO), we can move forward with the hypothesis that there is a *correlation* between a large amount of ‘hits’ and a sharp increase in the price of the MSN and ROIAK stock prices.

Next, data related to the musical characteristics of all of these songs was explored to determine if there is a correlation between being a Top 100 song and their characteristics, of which we will define and use: ‘Danceability’, ‘Energy’, ‘Tempo’, and ‘Speechiness’. There are many other musical characteristics that could determine whether a song will go on to be a success, but these are the ones to be explored in this paper. Danceability is scored as a measure of tempo and time signature, energy as a function of the intensity/volume/ of vocals/instruments, tempo in terms of beats per minute, and ‘speechiness’ in terms of the portion of the song with dominant vocal tracks. All of the above seem intuitive, except for speechiness. There has been a great deal of research in the area of Neuroscience as to the effects of vocal vs instrumental-based music on memory recollection (Janata P, Tomic ST, Rakowski SK)(Janata, P.) and emotion (Jäncke, Lutz.), a great deal of the most influential research coming from Petr Janata’s work at the UC Davis Center for Mind and Brain. The author of this paper also happens to work as an undergraduate research assistant at Dr. Janata’s lab. The collective body of research on this topic seems to point to the hypothesis that melodies are much easier for the human brain to remember than vocals, and that both vocal and instrumental songs can develop very strong musico-emotional memories, leading to a personal attachment to the song. This could be a reason that songs with certain melody characteristics become ‘hits’ – they have strong ties to a listener’s autobiographical emotional memory. For this to be true in our data, low scores for ‘speechiness’ (due to the research pointing to melodies being easier to remember than vocals) and high scores for ‘energy’ – the intensity of the lead track (whether it be vocal or instrumental) should be obtained. All of these musical characteristics are included in a programmatic musical analysis tool created by Spotify/EchoNest - given a song name and artist, the API will return the characteristics in JSON format, in the form of a score from 0 -1. All of the songs in the top 100 will be plotted in terms of their given characteristics, in order to observe whether the top 50 differ significantly from the bottom 50. This data was plotted by the code on the following page.



And replacing ‘Danceability’ with the other 3 characteristics, in order to plot them as well. The results looked like this:



Danceability Mean (Lower 50): 0.63754

Danceability Mean (Top 50): 0.661693877551

Energy Mean (Lower 50): 0.62058

Energy Mean (Top 50): 0.569816326531

Tempo Mean (Lower 50): 122.16726

Tempo Mean (Top 50): 124.138979592

Speechiness Mean (Lower 50): 0.108064

Speechiness Mean (Top 50): 0.171426530612

These results appear to show no real difference in musical characteristics within the top 100 (between the bottom and top 50), as there was no difference in mean big enough to be significant. However, the results do show low scores for mean Speechiness, and slightly above average scores for mean Energy, as predicted by the research in emotional music recollection. Danceability also scored above average. With this in mind, a model was built to attempt to predict if a song will be a hit, based only on the three mentioned characteristics (removing tempo from the equation).

After obtaining the scores for all 3 characteristics, for all 100 songs, for every year since 2008 – to a .csv file, the features are set to be the scores for danceability, energy, and speechiness, and the output to be predicted is set to be whether the number of weeks on the top 100 chart (‘successes’) in the code to follow – is equal to 1 or greater than 1. We then build a machine learning classification model, using Sklearn and Modelrithm in Python code on the following page.







With a 92% accuracy and precision score, as well as a .95 F-beta score, both the SVC (Support Vector Classifier) and Random Forest Classifier (ensemble method) outputted a very accurate prediction of whether the test data set was going to stay in the top 100 charts for at least another week. These results could be used to hypothesize that since a majority of the predicted outputs are positive, there will be an overall increase in trend for the stock price of the 2 aforementioned radio broadcasting companies. This model could also be tweaked to predict ‘big hits’ as defined by songs which stay on the Top 100s charts for 1, 2, 3, etc. months.

References/Tools:

1. Billboard API: https://github.com/guoguo12/billboard-charts
2. Janata, P. “The Neural Architecture of Music-Evoked Autobiographical Memories.” Cereb Cortex 2009; 19 (11): 2579-2594. 15 Mar. 2017
3. Janata P, Tomic ST, Rakowski SK. “Characterization of music-evoked autobiographical memories.” *Memory*. 2007 Nov; 15(8):845-60. 15 Mar. 2017
4. Jäncke, Lutz. “Music, Memory and Emotion.” *Journal of Biology* 7.6 (2008): 21. *PMC*. Web. 15 Mar. 2017.
5. Matplotlib - https://matplotlib.org
6. Modelrithm - https://github.com/ferdavid1/Modelrithm (*This is a previous project of mine)*
7. NumPy – https://numpy.org
8. Pandas – https://pandas.pydata.org
9. Sklearn – https://scikit-learn.org
10. Spotify API wrappings for Python: https://github.com/plamere/spotipy