Issues in Information Systems
Volume 22, Issue 4, pp. 284-296, 2021

DOI: https://doi.org/10.48009/4_iis_2021_306-319

Music march madness: predicting the winner of locura de marzo

Kevin Mentzer, *Bryant University, kmentzer@bryant.edu*
Zachary Galante, *Bryant University, zachgalante62@gmail.com*
Brenna Rojek, *Bryant University, brojek@bryant.edu*
Rachel Cardarelli, *Bryant University, cardarellirachel@gmail.com*

Abstract

Each Spring, thousands of middle and high school students enrolled in Spanish classes vote for their favorite songs in the annual Locura De Marzo competition. This alternative March Madness competition gives us an opportunity to build and test models to predict which songs will win which furthers the Hit Song Science literature. Using decision trees and support vector machine (SVM) models we find similarities with the challenge of predicting the popular NCAA Basketball bracket including the importance of seed and the difficulty in predicting a "perfect" bracket.

Keywords: Hit Song Science, Bracketology, March Madness, Spotify, YouTube

Introduction

Predicting whether a song will turn into a hit or not has been of particular interest in the music industry. Careers and fortunes are made on the success of musicians generating content with mass appeal. However, studies on Hit Song Science (HSS) that have tried to build analytic models to accurately predict whether a song will become a hit have failed to produce the desired results (Pachet & Roy, 2008).

This work takes a new approach to predicting hits. Instead of looking holistically at the entire music market, we consider head-to-head matchups to determine whether a model can be developed that accurately predicts whether one song will be more successful than another. By merging studies that examine head-to-head matchups, i.e. Bracketology, with Hit Song Science, we build a model to predict these head-to-head matchups in a March Madness music competition. Using data from one of several popular music competitions organized by world language teachers, we train a model to predict the winner of this music competition.

Our findings suggest that the challenge of hit song prediction remains even down to a head-to-head matchup. Similar to studies trying to predict the popular NCAA basketball March Madness bracket, predicting a reliably accurate bracket for this music March Madness event remains challenging.

Literature Review

Hit Song Science

For years, producers and celebrities in the music industry have attempted to perfect the art of composing a hit song. However, there are many immeasurable factors included in the creation and production of a song that make it impossible to always produce hits. In recent years, Data Scientists have attempted to take a new approach to this problem using Hit Song Science. Hit Song Science (HSS) attempts to understand song characteristics (tempo, pitch, etc) and how they contribute to the making of a "Hit" or commercial success.

Current literature in Hit Song Science demonstrates the difficulty in predicting which songs will turn into hits. While there are concrete measurable factors such as loudness and tempo, there are many other factors to consider that are more difficult to measure. For example, exposure to a song, popularity amongst peers, and individuals' personal music preference are all factors that are highly correlated to the popularity of a song but cannot be easily measured or controlled during the creation and production of the song (Li et al., 2011). Similarly, a study by Mauch et al. which aimed to analyze the diversity amongst popular songs of each year since 1950 found that the diversity amongst charting songs predominantly shifts alongside cultural changes (Mauch et al., 2015). Thus, oftentimes the factors that determine a song's popularity are historical and unpredictable. Furthermore, Hit Song Science attempts to find connections between the concrete qualities of popular songs (Li, T., 2011) (Tough, 2013) in order to predict future hit songs. These features include measures such as pitch, gender of the musician, topic of the song, song form, and complexity of the song.

Hit Song Science was first introduced in 2003 by Polyphonic HMI (Human Media Interface). Polyphonic HMI collected data on popular songs from the 1950s on and created clusters to analyze similarities amongst the hits. Later studies, including one conducted by Ni et al. (Ni et al., 2011), attempted to study the trend of song popularity throughout the decades. The researcher was able to use time trends and the features of the song to create a semi-accurate model in predicting hit songs. This study drew attention to the importance of understanding the ever-changing style of music that is popular and recognizing that the popular preferences of the era in which one is predicting a hit in critical. In a related study, Interiano et al. found that only data from recent years contributed to the success of predicting a hit and showed that multi-decadal trends are not predictive (Interiano et al., 2018).

Pachet and Roy (Pachet & Roy, 2008) found that, despite any other claims, Hit Song Science fails to accurately predict hits with significance. The researchers used a dataset containing 32,000 songs and three categories of features: generic audio features, specific audio features, and human features. Their model attempted to predict a popularity level of low, medium, or high for each song but the result was inconclusive. The study suggests that the features that are widely used in Hit Song Science are not accurate predictors of song performance.

In contrast to Pachet and Roy's findings, Herremans, Martens, and Sörensen (Herremans et al., 2014) used song characteristics such as tempo and danceability along with a variety of modeling techniques including Naïve Bayes, decision trees, support vector machines, and logistic regression to predict hit songs. There results suggest that logistic regression worked best for detecting hit songs.

It is not just audio characteristics that have been studied. Singhi and Brown (Singhi & Brown, 2014) examined a set of rhyme and syllable characteristics of the lyrics to attempt to predict hit songs. Dhanaraj and Logan (Dhanaraj & Logan, 2005) found that lyric-based features were slightly more effective than audio-based features.

Kim, Suh, and Lee (Kim et al., 2014) take the novel approach of examining tweets to predict hit songs and find that song mentions using hashtags such as #nowplaying and #itunes could be used to predict hit songs while artist popularity is only weakly correlated.

Finally, genre of music may be an important factor to consider. Borg and Hokkanen (Borg & Hokkanen, 2011) suggest that hit song detection might need different approaches based on the genre of music being studied. On the other hand, Monterola, Abundo, Tugaff, and Venturina (Monterola et al., 2009) argue that neural networks work best regardless of genre.

Bracketology

Bracketology, or the science of predicting winners who participate in a head to head competition, is the foundation to produce predictive models for events such as the Division 1 NCAA's Basketball Tournament, commonly referred to as March Madness. This annual event is a popular office pool event (Kaplan & Garstka, 2001) with more than 55% of American workers participating in an office pool (Lowell, 2017) as well as a popular Kaggle challenge (*March Machine Learning Mania 2021 - NCAAM*, n.d.). In addition, it is estimated that as much as $12.5B is spent on wagers related to NCAA March Madness ("How Much Is Really Bet on March Madness?," n.d.). This all speaks to the high interest level in predicting an accurate bracket.

Given the context of March Madness there are 10^{20} possible bracket outcomes, making modeling a very challenging task (Jacobson et al., 2011). Decisions about which teams make the tournament, as well as how they are ranked (i.e. "Seeding") are determined by the NCAA Men's Basketball Selections Committee (*Men's Basketball Selections 101 - Committee | NCAA.Org - The Official Site of the NCAA*, n.d.).

Much has been studied in relation to the NCAA bracket prediction for feature selection. Magel and Unruh (Magel & Unruh, 2013) found that features including Free Throw Attempts, Defensive Rebounds, Assists, and Turnovers were best used to predict winners. In the March Madness tournament, upsets occur when a team with a higher seed loses to one with a significantly lower seed with the lower seed commonly referred to as a "Cinderella" story. Boulier and Stekler (Boulier & Stekler, 1999) found that seed alone can predict winners with a 73.5% accuracy, however this approach appears to work well only for the early rounds of the tournament (Stekler & Klein, 2012). Although seed of each team may be indicative of how the team will perform in the tournament, it is not a perfect feature (Jacobson et al., 2011)

There are many approaches to Bracketology and how to create the best model. Depending on the year, different models are more successful in predicting the outcome of the tournament. Amongst the best models are the Colley and Massey methods. The Colley Method (Colley, 2002) creates ranking for each team based on win percentage while the Massey method (Massey, 1997) ranks teams based on how each team performs against one another during the regular season.

Shi et al. (Shi et al., 2013), using random forests and support vector machine models, found that differences between teams was not significant and found very little difference between the various modeling techniques used. Fuqua (Fuqua, 2014) found similar results with model type being less important than variables used. As far as model accuracy, Shi et al (Shi et al., 2013) found that there appears to be a "glass ceiling" of 74-75% in accurately predicting a match regardless of modeling techniques used, confirming the overarching challenge of predicting the NCAA March Madness bracket.

Considering both Hit Song Science and Bracketology, this leads us to our research question:

RQ: Can a Bracketology approach to modeling lead to insights for Hit Song Science?

Research Methodology

Data Acquisition

Run by Señor Ashby, a Michigan middle school Spanish teacher, Locura de Marzo *(*Locura De Marzo 2021*, n.d.)* is an online music competition featuring Spanish songs where students from across the world are able to vote for their favorite songs. The competition, which began in 2017, runs each spring and songs compete in a head-to-head matchup with the winners moving on to the next round until a final winner is crowned the winner of the competition. The songs are found on popular platforms such as YouTube and Spotify, allowing for mass exposure.

Using the Locura de Marzo website we collected the song title and artist(s) for each of the 16 songs competing each year going back to 2016. In the fall of 2020, Señor Ashby ran a similar competition called Locotubre. However, we did not include the songs that competed in the 2020 Locotubre challenge due to the fact that this was a different style of competition since it was made up of top songs from previous years.

Using the data from the Locura De Marzo site, we were able to find the official song on Spotify and the official video on YouTube. Using the APIs from each platform, we obtained the following data for each song:

Table 1. Data Sources and Variables

Spotify Variables	YouTube Variables	Twitter Variable
Release Date	Favorite Count	Followers Count
Length	View Count	
Popularity	Like Count	
Acousticness	Dislike Count	
Danceability	Comment Count	
Energy	Tags	
Intrumentalness		**Locura De Marzo**
Liveness		Seed number
Loudness		Song Title
Speechiness		Artist Name
Tempo		
Key		
Mode		

Model Development

Predicting a bracket style tournament is a difficult task in itself, with many simulations resulting in low accuracy scores. In fact, the chances of predicting a perfect NCAA March Madness bracket are 1 in 120.2 billion (*The Absurd Odds of a Perfect NCAA Bracket | NCAA.Com*, n.d.).

In order to better understand the data available to us, we began by generating a correlation matrix (Figure 1) allowing us to quickly see which variables may be important to our models. Our first four variables indicate whether the song won in each of the rounds with 'Win R4' indicating whether the song was the overall winner or not. We see a moderate to high negative correlation between Seed, with a higher seed equal to a lower number causing the negative correlation, and likelihood to win rounds 1 and 2, with the correlation dropping down to a small to near negligible correlation for rounds 3 and 4. This indicates that overall seed is important early on in the competition while becoming less so as the competition moves on to subsequent rounds. We also see a moderate positive correlation between the YouTube video specific columns (Likes, Dislikes, and Views) and only a low correlation to number of Artist Followers suggesting that the song itself is much more important that the popularity of the artist.

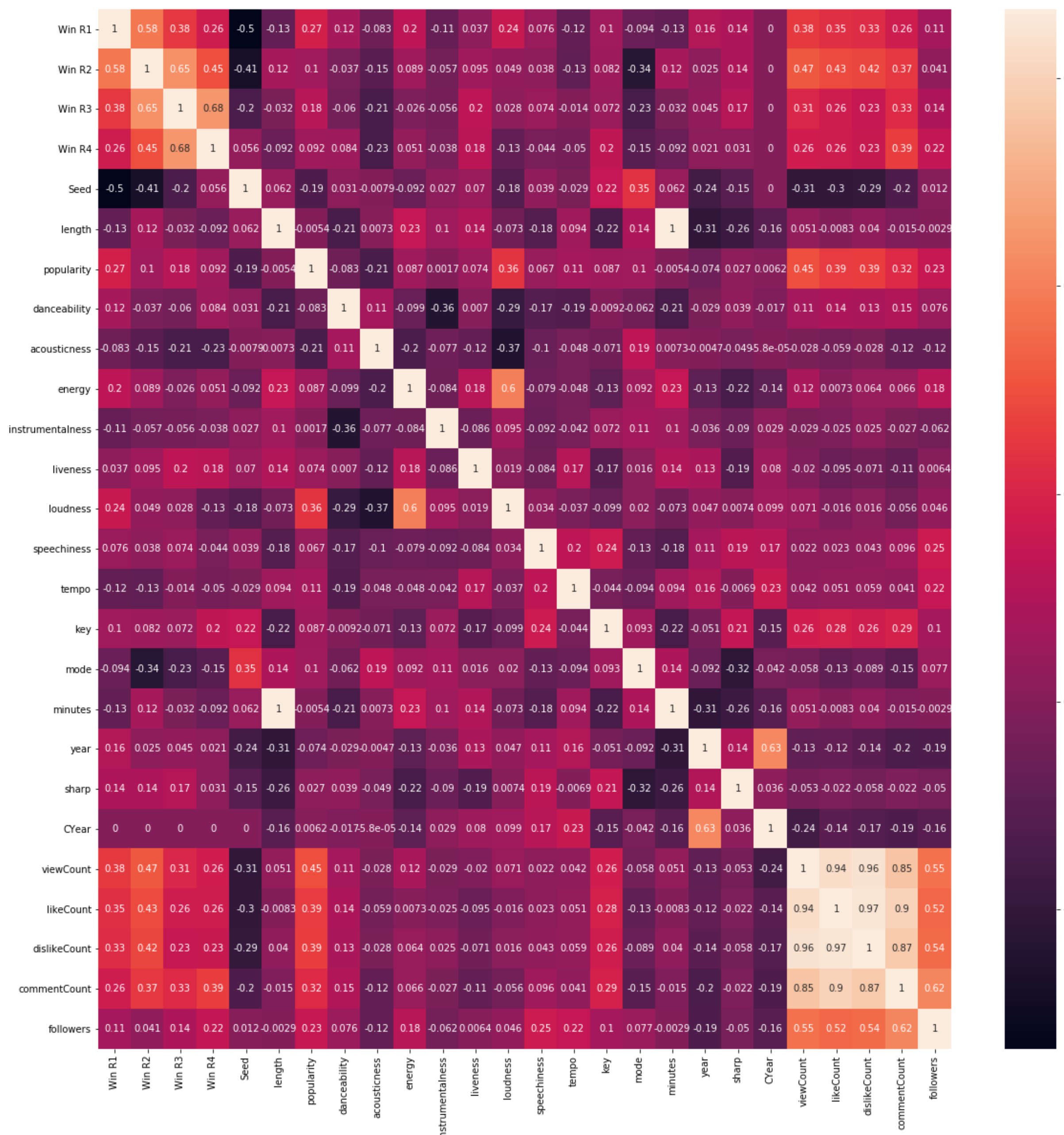

Figure 1: Feature Correlation Matrix

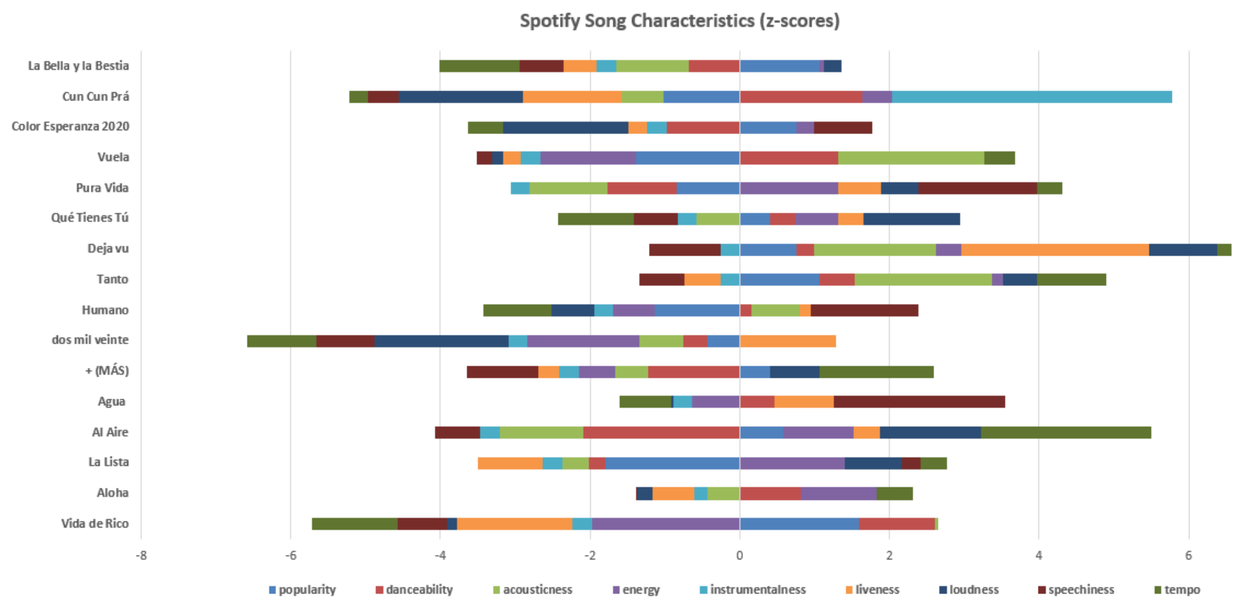

Figure 2: Differences in Song Characteristics

Following the correlation matrix, we turn to better understanding the song characteristics based on the related 9 Spotify attributes converting each Spotify measure to the z-score, we get a clearer understanding of how different each song's characteristic is from the other songs. The z-score represents the number of standard deviations away from the mean a song's attribute is. The stacked bar chart (Figure 2) shows how each characteristic contributes to the overall difference of each song. For example, Cun Cun Prá is 3.75 standard deviations from the mean in instrumentalness on the positive side indicating it is much more instrumental than the other songs, while Vida de Rico is two standard deviations from the mean in Energy on the negative side indicating it has much less energy than the other songs. When considering the stacked bar in entirety (so total distance from mean), songs such as Al Aire and Cun Cun Prá represent those that are the most different than the other songs while Aloha and Agua are the least different from the other songs.

Decision Tree

Next we modeled a decision tree (Figure 3). Our dependent, or target, variable was the chance of winning round 1. This allowed us to build a model with an even number of winners and losers, thereby emphasizing the traits needed to move past the first round.

The number of YouTube likes was the most important predictor with 16 winning songs, and only 2 losing songs, having more than 1.5M likes. Further, the combination of YouTube popularity with a seed number of 10 or higher proved unbeatable with all 15 songs winning the first round. However, we still had half of the winners, and 30 losing songs, with less than 1.5M likes represented in the left hand side of the decision tree. For those with less 1.5M likes, seed number was next important. Seed number greater than 11 (i.e. lowest ranked songs or potential Cinderella stories) were less likely to win with only 2 of the 18 in this category moving past the first round. While those ranked seed 1 through 11 still had a 50-50 shot at winning.

At this point in our model we are left with less popular songs with seeds 1 through 11 and our model finally detects song characteristics as important. Specifically, the song's key, the liveliness, and danceability were found to help predict their chance of winning.

While no final leaves in our decision tree have a mix of winners and losers, suggesting that the model is 100% accurate based on the training data, one should certainly assume that this model is over-fitted at this point due to the limited amount of data available.

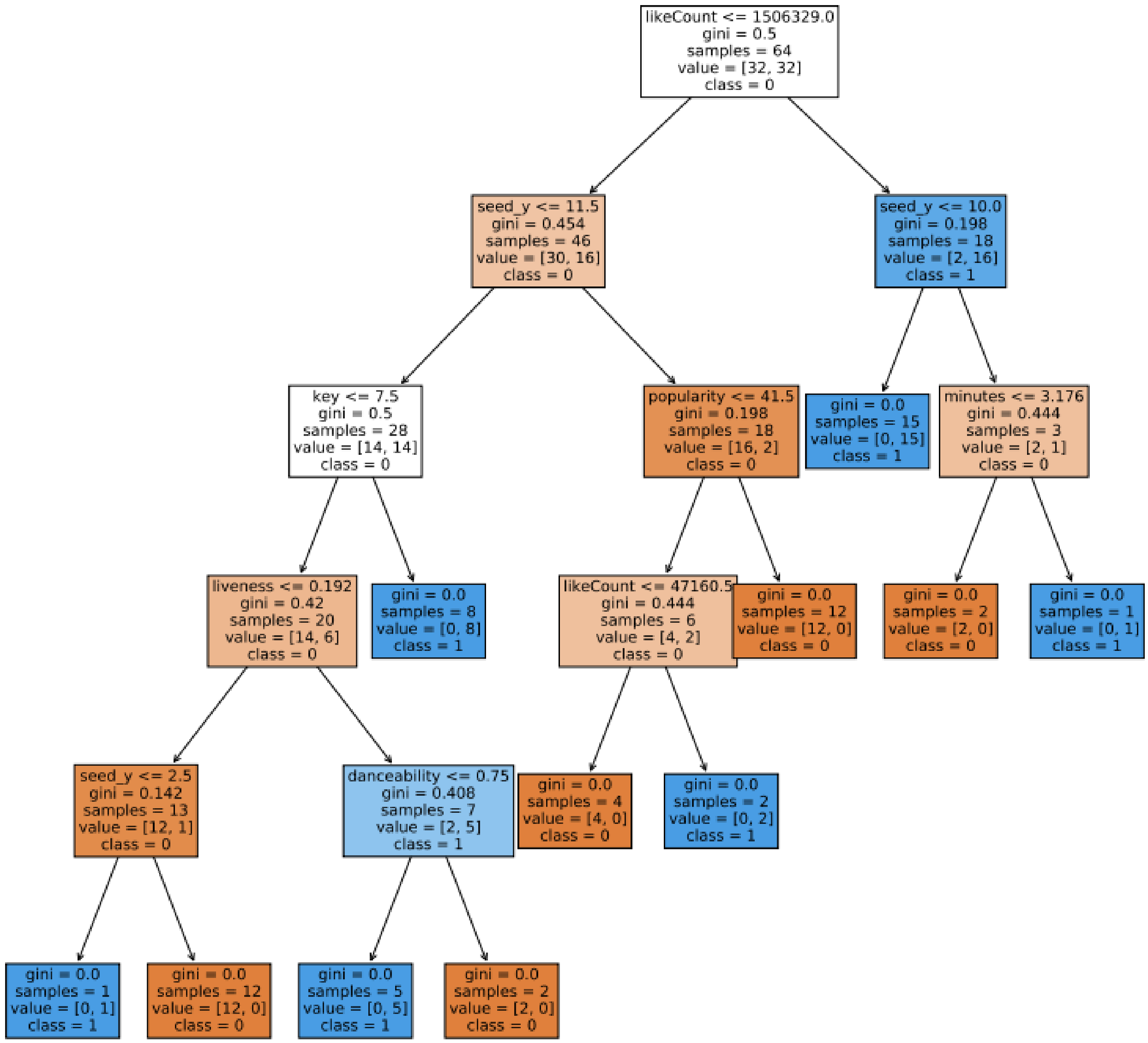

Figure 3: Decision Tree

Support Vector Machine

A Support Vector Machine (SVM) is a supervised machine learning algorithm that is frequently used for both classification and regression purposes. SVMs are based on the challenge of finding a hyperplane that best divides the dataset into two classes, which is well aligned with our challenge of finding winners and losers. Confidence in the prediction increases the further away from the hyperplane the data point is.

The decision to use a Support Vector Machine was made based on its ability to return the actual probabilities that a certain song would move onto the next round. Further, this is making use of the 'predict_proba' function within SKlearn to then select the higher of the two values. This feature, and its ability to return clear and concise results led to its selection for the overall model.

The modeling process is to train a Support Vector Machine on the results of each round from the previous Locura De Marzo tournaments. This includes every song and a binary result if they were victorious in a given round. This dataset can be thought of as the training set for the model, with the test set being the songs from the 2021 simulation.

Beginning with the 1st round results, the attributes for each song are used to train a Support Vector Machine on the 'R1' feature. Once this model was trained, the song pairings from the 2021 tournament were fit into the trained model based upon their assigned matchups for prediction. The result is a probability that each song would win the round when matched up against each other. Simulating the results for 100 iterations gives us our final prediction. Whichever probability was higher between the two songs in the given matchup would then move on to the next round. This process is repeated for each successive round.

Results

The actual results, based on student voting, from the 2021 Locura De Marzo tournament are shown in Figure 4 with winners both bold and having a check mark. Each line includes the seed number, song name, check mark if the song won the head-to-head matchup, and the % of votes the song received. As can be seen, the "Final 4" consisted of seed numbers 5, 6, 7, and 8 with none of the seed number1 through 4 making the final four round. Interestingly, this was the first year the number 1 seed did not make it into the final four. However, unlike the NCAA tournament, where Cinderella stories tend to not make it as far as the final four, in the Locura De Marzo tournament it is not uncommon to see high seeds in the final 4; in fact, the number14 seed has won the tournament in 2 of the past 5 years of the tournament.

As previously stated, predicting a perfect bracket, or just a reasonably accurate one, is an extremely difficult task based on the elimination style of the tournament. An incorrect prediction feeds into subsequent rounds compounding into subsequent predictions. For example, if the team or song that was picked to win the tournament loses in the first round, then all following games will be incorrect, this is referred to as forward propagation error. As a result, this makes the task of prediction much more difficult with models.

After running both the decision tree and the SVM, clear conclusions can be drawn from both models. When using the decision tree to predict the outcome of a first round matchup, features such as the YouTube Like Count and the original Locura De Marzo Seed number emerged as the most important attributes. For the Spotify attributes, the Key, Liveness, and Danceability proved to be important while the remaining attributes, including Popularity and Tempo, did not make an appearance on the decision tree. While the other YouTube features were highly correlated with Seed number, these did not emerge in the decision tree most likely due to the fact that they were also highly correlated with the YouTube Like Count. The number of Twitter followers the artist has also appeared to have no impact on the song success in the model.

When comparing the results from the model to the actual results of the 2021 tournament, the model performed well, given the difficulty of predicting a bracket style tournament. Looking at the results from the tournament as a whole may not give the best insight to the accuracy of the model so instead, it is more appropriate to look at the results round by round. For the first round of predictions, the model predicted 5/8 (62.5%) matchups correctly. This is a significant score, because it predicted the majority of matchups correctly, but it failed to predict the actual winner of the tournament to move onto the 2nd round. Ultimately, this resulted in the games further on in the tournament to be incorrectly predicted. However, the model was still able to correctly predict the outcome of one of the four 2nd round matchups. This was the final correctly predicted game in the bracket. After running the model 100 times, “Agua” was predicted to win the 2021 tournament. Although this is not the correct prediction, “Agua” did make it to the final four.

Figure 4. Actual Results

Figure 5: SVM Model Prediction for 2021 Locura De Marza

Conclusion

This is the first study we are aware of that applies Bracketology to the music domain. As prior studies have shown, bracket prediction is a challenging task, and we find that this extends into the music domain as well. Similar to the NCAA tournament models, seed appears to be an important feature, but just as others have found, its accuracy diminishes significantly with each round.

While predicting hit songs has been elusive, we offer a novel technique by comparing songs against each other. While Seed number and the number of YouTube likes emerged as important features, there is also evidence that the song-level attributes of Key, Liveliness, and Danceability were also significant. Overall song feature deviation from the mean didn't appear to be important with no observable differences between songs with similar levels of overall deviation.

This work highlights that prediction competitions with a Bracketology framework (i.e. head-to-head matchup) have merit outside of sports and have the potential to appeal to non-sports enthusiasts.

Limitations

One of the challenges of predicting a head-to-head matchup is the fact that if one team/song wins then the other team/song has to lose. This limits the number of modeling techniques available. For instance, while we used a decision tree to understand feature importance, it is challenging to apply it in a head to head matchup since both songs could theoretically end up being classified as the winner.

Accurately making predictions based on such a limited amount of data is challenging and these results are instead meant to be exploratory in nature to better understand Hit Song Science. We would expect the accuracy of this model to vary widely from year to year.

One of the factors that emerged as relevant was the seed number. This seed is chosen based on faculty voting. As such, this measure has bias built in. However, this same challenge exists in the NCAA tournament since seeds are also selected based on a biased pool of individuals.

The Locura De Marzo competition is a "celebration of Hispanic culture and language" (*HOW IT WORKS 2021*, n.d.). As such, this competition represents a subset of music and the features that are important for the success of popular Latin music may not transfer over to other genres.

Finally, teachers are encouraged, as part of the competition, to show the music videos to the students prior to voting. Some teachers do so while others choose to just play the songs for the students without the video. Therefore, components of the video are likely influencing the voting. However, without coding for the characteristics of the videos and understanding which votes were based on song only versus song and video, this influence could not be included in this study. In addition, through a shared drive, teachers share lesson plans to go along with certain songs, and voting can certainly be influenced by the lesson plans used.

Future Work

This work shows that Bracketology can be used beyond the scope of sports and may offer insight into the field of Hit Song Science. However, novel sources of data will be necessary to train models accurately and, in the meantime, we can expect wide variation in the accuracy of such models. Social media is making voting surveys in a bracket style manner popular, whether it be best ice cream (*WBZ Ice Cream Social Sweet 16*, n.d.) or best burger (Naik, n.d.). This work should be viewed as a starting point to build predictive models for a bracket style tournament and offers an exciting opportunity to bring challenges like this into the classroom as evidenced by Plant Madness (Miller, 2019) which brought a bracket style challenge into a horticultural class. We would expect new insights to emerge with each iteration of the Locura de Marzo competition furthering discovery of the traits that predict hit songs. These challenges offer students the opportunity to explore novel data sources for building predictive models and can appeal to those who might otherwise be turned off by a sports theme.

References

Borg, N., & Hokkanen, G. (2011). What makes for a hit pop song? What makes for a pop song. *Unpublished Thesis, Stanford University, California, USA*.

Boulier, B. L., & Stekler, H. O. (1999). Are sports seedings good predictors?: An evaluation. *International Journal of Forecasting*, *15*(1), 83–91.

Colley, W. (2002). *Colley's bias free college football ranking method*. Ph. D. Thesis, Princeton University, Princeton, NJ, USA.

Dhanaraj, R., & Logan, B. (2005). Automatic Prediction of Hit Songs. *ISMIR*, 488–491.

Fuqua, C. (2014). The Final Four Formula: A Binary Choice Logit Model to Predict the SemiFinalists of the NCAA Division I Men's Basketball Tournament. *Major Themes in Economics*, *16*(1), 31–49.

Herremans, D., Martens, D., & Sörensen, K. (2014). Dance hit song prediction. *Journal of New Music Research*, *43*(3), 291–302.

HOW IT WORKS 2021. (n.d.). Retrieved June 12, 2021, from http://www.senorashby.com/how-it-works-2021.html

How much is really bet on March Madness? (n.d.). *Bloomberg.Com*. Retrieved July 15, 2021, from http://www.bloomberg.com/graphics/2015-march-madness-gambling/

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

Jacobson, S. H., Nikolaev, A. G., King, D. M., & Lee, A. J. (2011). Seed distributions for the NCAA men's basketball tournament. *Omega*, *39*(6), 719–724.

Kaplan, E. H., & Garstka, S. J. (2001). March madness and the office pool. *Management Science*, *47*(3), 369–382.

Kim, Y., Suh, B., & Lee, K. (2014). # nowplaying the future Billboard: Mining music listening behaviors of Twitter users for hit song prediction. *Proceedings of the First International Workshop on Social Media Retrieval and Analysis*, 51–56.

Li, T., Ogihara, M., & Tzanetakis, G. (2011). *Music data mining*. CRC Press.

Lowell, A. (2017). Gamblization: The Rise of Sports Gambling and the Need to Repeal PASPA. *Pace Intell. Prop. Sports & Ent. LF*, *7*, 255.

Magel, R., & Unruh, S. (2013). *Determining factors influencing the outcome of college basketball games*.

March Machine Learning Mania 2021—NCAAM. (n.d.). Retrieved July 15, 2021, from https://kaggle.com/c/ncaam-march-mania-2021

Massey, K. (1997). Statistical models applied to the rating of sports teams. *Bluefield College*.

Mauch, M., MacCallum, R. M., Levy, M., & Leroi, A. M. (2015). The evolution of popular music: USA 1960–2010. *Royal Society Open Science*, *2*(5), 150081.

Men's Basketball Selections 101—Committee | NCAA.org—The Official Site of the NCAA. (n.d.). Retrieved June 12, 2021, from https://www.ncaa.org/about/resources/media-center/mens-basketball-selections-101-committee

Miller, C. T. (2019). Plant Madness: A Classroom Game Using Bracketology for Horticulture Plant Identification Courses. *HortTechnology*, *29*(2), 223–228.

Monterola, C., Abundo, C., Tugaff, J., & Venturina, L. E. (2009). Prediction of potential hit song and musical genre using artificial neural networks. *International Journal of Modern Physics C*, *20*(11), 1697–1718.

Naik, N. (n.d.). *Burger Bracket: Vote for the winning burger in the Midlands*. The State. Retrieved July 15, 2021, from https://www.thestate.com/living/food-drink/article252761858.html

Ni, Y., Santos-Rodriguez, R., Mcvicar, M., & De Bie, T. (2011). Hit song science once again a science. *4th International Workshop on Machine Learning and Music*.

Pachet, F., & Roy, P. (2008). Hit Song Science Is Not Yet a Science. *ISMIR*, 355–360.

Shi, Z., Moorthy, S., & Zimmermann, A. (2013). Predicting NCAAB match outcomes using ML techniques–some results and lessons learned. *ECML/PKDD 2013 Workshop on Machine Learning and Data Mining for Sports Analytics*.

Singhi, A., & Brown, D. G. (2014). Hit song detection using lyric features alone. *Proceedings of International Society for Music Information Retrieval*, 30.

Stekler, H. O., & Klein, A. (2012). Predicting the outcomes of NCAA basketball championship games. *Journal of Quantitative Analysis in Sports*, *8*(1).

The absurd odds of a perfect NCAA bracket | NCAA.com. (n.d.). Retrieved June 11, 2021, from https://www.ncaa.com/news/basketball-men/bracketiq/2021-03-14/perfect-ncaa-bracket-absurd-odds-march-madness-dream

Tough, D. (2013). Teaching modern production and songwriting techniques: What makes a hit song? *MEIEA Journal*, *13*(1).

WBZ Ice Cream Social Sweet 16. (n.d.). Retrieved July 15, 2021, from https://boston.cbslocal.com/wbz-ice-cream-social-sweet-16/