**DSC1 5330 Final Project Report:**

**Decoding What makes a Chartbuster:**

**Predictive Modeling of Song Popularity based on**

**Spotify Audio Features**

**By**

**Yusuf Saadu**

**Abstract**

In the dynamic and competitive realm of the music industry, akin to the intricacies of stock market investments, the selection of songs for promotion is pivotal. This research project explores the feasibility of a predictive model that employs Spotify's extensive audio features database to compute a song's popularity score. This innovative approach is designed to aid music labels and artists in making data-driven decisions about which tracks to promote, optimizing resource allocation, and enhancing the probability of commercial success. Grounded in empirical data, the model prioritizes quantitative audio features such as speechiness, acousticness, instrumentalness, liveness, valence, and tempo, capitalizing on Spotify's influential role in the music streaming domain.

Utilizing historical data sourced through Spotify's API, this research project looks to refine a predictive algorithmic model capable of generating a precise popularity score for any given song based on a given set of quantified audio features. Such scores are intended to signal which songs merit widespread release and marketing investment. While the project remains imperfect and not yet fully applicable in a real-world context, the preliminary results are promising and can be applied in a corporate setting, suggesting that audio features significantly influence song popularity but are not the sole determinants. It becomes clear that additional factors like genre, artist public profile, timing, and virality also play crucial roles and must be integrated into the predictive model to fully encapsulate a song's potential success on platforms like Spotify. The industry's ever-evolving nature demands that the model be continuously updated with new data to remain relevant. This research project paper offers a compelling synthesis of data science and music industry analytics, marking a step towards a more strategic and informed approach to predicting chart-topping hits.

**Introduction, Problem Statement and General Description**

In the cutthroat world of the music business, selecting the right song for promotion is akin to picking the ideal stock for investment. The stock market has seen numerous models aiming to forecast stock behaviors and reduce risks. Taking inspiration from this, the same logic can be applied to the music industry, a realm where the significance of a chart-topping hit reverberates beyond just song sales. From live performances, streaming royalties, and movie/tv show/advertising licensing to other revenue streams, the cascading benefits of a hit song are immense. The converse is also true; an incorrect song choice can adversely impact an artist's brand and drain a label's resources.

The aim of my project is to address the critical question: How can artists or labels identify the potential chartbusters from a variety of produced tracks? I proposed the solution of a predictive model harnessing Spotify's vast audio features to generate a song's popularity score. This score would aim to guide the music labels and artists in their investment and resource allocation decisions. The primary objective is crafting a predictive algorithm that gauges a song's potential popularity grounded on quantitative audio features. The emphasis is on attributes such as speechiness, acousticness, instrumentalness, liveness, valence, and tempo. Given Spotify's dominance in the music streaming landscape, leveraging their platform’s data offers a realistic popularity and commercial potential index for songs.

Historical Spotify data, amassed via its API and gotten from Kaggle, serves as the foundation. The aspiration is to fine-tune the model to a level where any artist's song can be examined, its features quantified, and fed into the model, yielding an accurate popularity score. Consequently, songs with peak scores are primed for mass release, bolstered by a promotional budget. The overarching aim is to amplify the odds of a song resonating with a vast audience in today's saturated music market. This endeavor offers an exciting intersection of data science and the music industry. While still purely theoretical and not really ready for real world or real time application, the results can provide invaluable insights, streamlining the song selection process and augmenting the chances of commercial success.

At the conclusion of the project, based on the performance metrics I’ve been able to elicit from the models I’ve trained, it can be ascertained that there is still much to be done when it comes to predicting the popularity of a song and it can not be solely based on its audio features. The results of the project show that the audio features do have a significant effect on the popularity of a song, however it does not paint the whole picture. Other variables need to be brought in such as genre popularity, artiste persona, virality, etc. to more fully capture how well we can predict the popularity of a song on the music streaming platform Spotify. Not to mention, the popularity of a song is constantly changing in the current fast paced or rather “microwave” nature of the industry, what is popular today may not be popular tomorrow, next week, month or year. The model, if perfected, will have to be constantly fed and trained on new data to keep it up to date, current and in vogue.

**Literature Review, Related Work & Background Info**

Navigating the intricate landscape of the music industry is a formidable endeavor, one made more tenable through the related research articles I was able to procure. The task of predicting song popularity has been akin to deciphering the enigmatic shifts of the stock market, it hasn’t been as easy as I thought it’d be. To comprehend the underpinnings of this industry, I relied heavily on empirical data as well as the substantive academic literature that offered historical and methodological perspectives. Historically, scholars endeavored to unravel the complexities behind a song's commercial success. David Huron's profound investigations into the psychology of sound shed light on how particular melodies or rhythms could potentially captivate emotions. Similarly, Philip Ball’s seminal work, "The Music Instinct," provided invaluable insights into the science of musical harmony and its consistent impact on global audiences.

In the realm of modern research, the study by Araujo, Cristo, and Giusti (2019) elucidated the intricacies of predicting music popularity on streaming platforms. Their findings, complemented by the methodological innovations proposed by Yutong et al. (2021) regarding factor extraction and model blending, enriched my understanding of contemporary predictive techniques. Beniwal and colleagues (2023) took a targeted approach by focusing on Hindi hit songs, revealing the cultural nuances in song popularity, while Kamal et al. (2021) offered a classification-based approach, emphasizing the potential of categorizing songs based on their likelihood of success. These academic pursuits, coupled with seminal investigations like that of Pham & Kyauk in 2015, forged a robust framework for my inquiry.

The rise of Spotify in the digital age into a position of dominance has been a phenomenon to witness. Various studies that dissect its API and audio features, such as in the work by Martin-Guiterrez et al. (2020), which presents a multimodal deep learning architecture for song popularity prediction, have been instrumental. By tapping into attributes such as speechiness and valence, these studies have provided a clear direction for my project. It is imperative to acknowledge that the realm of music does not live or die solely by its auditory components. External determinants wield significant influence on a song's trajectory. The erudite literature, epitomized by Essa et al. (2022), which harnessed machine learning techniques for predicting song popularity, underscored the necessity of an encompassing approach.

As my exploration into predicting song popularity progressed, I found myself consistently enlightened by the expansive academic canon. Each research contribution, with its methodological rigor and distinctive insights, informed my project approach. My unwavering objective was to synthesize empirical data with the profound revelations from these scholarly discourses, striving for acuity in the multifaceted domain of music.

**Data**

This project utilizes a dataset sourced from Kaggle that was compiled utilizing Spotify API. The dataset has more than a million rows and 20 columns with no null values. The dependent/predicted variable was **Popularity** which measures how many plays a track gets as well as how recently people listened to it. It ranges from 0 to 100 with the higher numerical value being the more popular and vice versa. Meanwhile the other variables in the dataset include:

* Genre: Serves as a categorical variable that delineates the musical style or category of the track.
* Danceability: A numeric variable indicating the suitability of a track for dancing based on various musical elements, ranging from 0.0 (least danceable) to 1.0 (most danceable).
* Energy: A numeric variable measuring the intensity and activity level of a track, ranging from 0.0 (least energetic) to 1.0 (most energetic). Energetic tracks are typically fast, loud, and noisy.
* Key: Categorizes the key to the track using integers that map to pitches in standard Pitch Class notation, e.g., 0 = C, 1 = C♯/D♭, 2 = D, etc.
* Loudness: Quantifies the overall loudness of a track in decibels (dB), averaged across the entire track, typically ranging between -60 and 0 dB. It is useful for comparing the relative loudness of tracks.
* Mode: A binary variable indicating the modality of a track, where 1 represents major, and 0 represents minor.
* Speechiness: Refers to the proportion of spoken words in a track. Values closer to 1.0 indicate a higher prevalence of speech-like recording.
* Acousticness: A confidence measure ranging from 0.0 to 1.0, indicating the likelihood that the track is acoustic. A value of 1.0 denotes high confidence in the track being acoustic.
* Instrumentalness: Denotes the likelihood that a track contains no vocals, with values closer to 1.0 indicating a higher probability of being instrumental.
* Liveness: Indicates the presence of an audience in the recording. A value above 0.8 suggests a high probability that the track is a live recording.
* Valence: Depicts the musical positiveness conveyed by a track. Higher values indicate more positive-sounding tracks, e.g., happy, cheerful, euphoric, while lower values suggest more negative-sounding tracks, e.g., sad, depressed, angry.
* Tempo: Represents the estimated overall tempo of a track in beats per minute (BPM), reflecting the speed or pace of a piece of music.
* Duration\_Ms: Specifies the duration of the track in milliseconds.
* Time\_Signature: Denotes an estimated overall time signature of a track, specifying how many beats are in each bar (or measure) following musical notation conventions.

**The independent variables deemed significant and actually used for this project include Danceability, Loudness, Mode, Speechiness, Acousticness, Instrumentalness, Liveness, Valence, Tempo, Duration\_ms, and Time Signature. After testing significance, multicollinearity and predictive power, these were the ones that stood out and showed that they could play a part in building the predictive model.**

**Project Methodology & Technical Approach**

The problem of tackling and predicting song popularity was honestly a much harder and arduous of an undertaking than initially anticipated. However, the project roadmap towards its final completion it has been broken down into the following steps:

1. **Defining the Problem and Scope:** The first step for this project was to define the problem I was trying to tackle and how it could generate value in data driven decision making. I not only had to decipher how the problem of song selection by labels and artistes was such a potentially risky decision, but I had to find the right metric and way to gauge popularity of the songs. I decided to use only one streaming platform Spotify since they were the biggest and most dominant.
2. **Data Sourcing:** The next step was finding the right data source to base this project on. I swept through the reliable Kaggle for a dataset before settling on three potential candidates. I did some research into the Spotify API and didn’t think I could do the scraping myself to get the data, so I found clean datasets on Kaggle that had what I was looking for in terms of audio features, variables and size. I initially chose the largest one with one million entries but issues with conducting exploratory data analysis on the dataset using Tableau led me to switch to a different but similar dataset. However, upon deciding to use Python through Google CoLab, I picked the hefty 1 million dataset back up as I saw that the platform could handle processing the large amount of data.
3. **Literature Research and Review:** Finding research literature for this particular project idea was daunting and straining however through the use of search optimization in Google Scholar I was able to find a bunch of papers that had to do with predicting song popularity. I then had to filter it by date picking only the most recent ones and the ones I could actually gain access to the full article whether it was through the school or payment. After that I was able to have a respectable list of research project papers that I could read through and look for guidance on what part to take for my project. They also provided valuable insight on possible datasets I could use and what kinds of models or algorithms I could employ.
4. **Exploratory Data Analysis:** This entailed putting the dataset through various tests and processing to find nulls, outliers, and other summary statistics. I tested for significance and multicollinearity so as to carry out feature selection in terms of the independent variables I’d use for the project. It also involved visualizing the chosen variables of interest in boxplot, histogram, and correlation matrices so as to get a better understanding of the data I’m dealing with. I also calculated mutual information and F scores for each variable ranking the variables from highest to lowest by feature importance to get a sense as to which variables would be the best in helping me predict popularity and which ones I’d need to drop.
5. **Data Cleaning and Preprocessing:** Upon carrying out the exploratory data analysis and settling on the list of variables I wanted to use to train the predictive model, I ran the Shapiro Wilk’s and D’Agostino’s tests for normality. It then came to my attention by way of the tests, histogram and box plot visualizations that some of my variables were not normally distributed and in need of transformations to better enable the performance of the model. This was of course done to normalize distribution and account for the effect of outliers. After making sure there were no null values, I carried out log transformation on speechiness, instrumentalness, and liveness. I also standardized loudness, tempo and duration using Standard scaler. Finally for the categorical variables like mode and time signature, I encoded them using One Hot Encoding. After that, I checked the distribution again and decided to carry out further Box Cox transformations on the duration, log-transformed speechiness and log-transformed liveness variables. Finally, the next step was to split the dataset into training and test for both the dependent and independent variables. The dataset split was 65% for the training set and 35% for the test set.
6. **Model Selection and Training:** After splitting the datasets into the training and test sets, I ran the training set through 9 different algorithms namely including:
   * *Linear Regression*: I used Linear Regression because it's ideal for a straightforward and interpretable model, especially if the relationship between Spotify's audio features and song popularity was linear. It's a fundamental approach for understanding direct correlations in data.
   * *Decision Tree:* I used Decision Tree because it can effectively map out complex decision rules that might’ve existed in the data. Its tree structure is great for understanding hierarchical relationships and making the model's decisions transparent and easy to interpret.
   * *Random Forest*: Random Forest was my choice due to its robustness in handling datasets with a multitude of features. It builds on decision trees but enhances accuracy and prevents overfitting through ensemble learning, making it more reliable for complex datasets.
   * *Ridge Regression*: I found Ridge Regression suitable for this dataset as it would help to address multicollinearity issues, which I thought would be common when dealing with numerous audio features such as loudness and energy. By including a regularization term, it ensured the model didn’t overfit and maintained generalizability.
   * *Lasso Regression*: I used Lasso Regression because it not only helped in preventing overfitting but also performed feature selection. This was particularly useful in identifying the most impactful features from Spotify's data that influence song popularity.
   * *K-Nearest Neighbors (KNN)*: KNN was a straightforward, intuitive choice for predicting song popularity based on the concept that similar songs might share similar popularity. This method was beneficial for capturing local patterns in the dataset.
   * *XGBoost Regressor:*  XGBoost stands out due to its exceptional efficiency and accuracy, especially in handling complex datasets like Spotify's audio features. Its advanced form of gradient boosting constructs trees sequentially to improve upon previous mistakes, effectively capturing the intricate non-linear relationships within the data. This makes it a highly powerful tool for predicting song popularity.
   * *Elastic Net*: Elastic Net was a balanced approach that combined the strengths of both ridge and lasso regression. It was particularly effective for the dataset where features were correlated, and I needed both regularization and feature selection capabilities.
   * *Neural Network*: I used Neural Network because of its ability to model extremely complex and non-linear relationships. Given the complexity and size of the Spotify dataset, a neural network should be capable of capturing deep patterns that simpler models might miss.
7. **Model Testing:** After training each model using different algorithms, I tested each trained model using the test dataset set to predict the popularity value of each song in the test set. After producing the predictions, I measured each model’s performance using each of the following evaluation metrics:

* *Root Mean Squared Error (RMSE)* - RMSE is a good choice because it provides a clear indication of the model's prediction accuracy, penalizing larger errors more severely, which is crucial for a reliable prediction of song popularity.
* *R-squared (R²)* - R² is valuable as it quantifies the proportion of variance in song popularity that is predictable from the audio features, offering a direct measure of the model's explanatory power.
* *Mean Absolute Error (MAE)* - MAE is beneficial for its simplicity and interpretability in representing the average magnitude of errors in predicting song popularity, without overly penalizing larger errors unlike RMSE.
* *Explained Variance* - Explained Variance is ideal as it measures how well our model captures the variability in song popularity, indicating the proportion of variation explained by the model.
* *Median Absolute Error* - Median Absolute Error is a robust metric for its resistance to outliers, offering a more reliable middle-ground measurement of the model's prediction errors in song popularity.
* *Adjusted R-squared* - Adjusted R-squared is a crucial metric for this model as it accounts for the number of predictors, providing a more accurate measure of the model's explanatory power for song popularity, especially when comparing models with different numbers of predictors.

1. **Result Interpretation and Evaluation-** After getting all the training and testing done, I finally settled on the Random Forest, XGBoost Regressor and Neural Network since they seemed to perform the best out of all the models based on the performance evaluation metrics namely RMSE and Adjusted R- Squared.
2. **Project Conclusion-** As stated earlier, the results of the project proved that while the audio features proved significant in determining a song’s popularity, they failed to paint the whole picture.

**Exploratory Data Analysis Insights**

As I mentioned earlier, I’ve done a bunch of data analysis on Spotify data. I checked the number of rows and columns (1159764, 20), looked over the first few rows of the dataset to see what it was like, checked and quantified the number of nulls for each variable. Surprisingly there were no null values in any of the columns. I then checked out the list of columns in the dataset to see which one I needed and which I didn’t. I ended up dropping these ones because they didn’t seem to have anything to do with the predictive model I was trying to construct: 'Unnamed: 0', 'artist\_name', 'track\_name','track\_id','genre', 'year'. So, all I was left with were some numerical variables. I then found the count, mean, standard deviation, minimum, maximum, 25th percentile, 50th percentile and 75th percentile for each variable remaining in the dataset. I made sure to double check for columns with missing values, there were none. I then constructed a correlation matrix in the form of a heat map to showcase each variable’s correlation with others in the dataset as a means of checking for possible multicollinearity.

A screenshot of a graph

Description automatically generated

After meticulously analyzing the data, several observations stood out. Notably, the data demonstrated a powerful positive correlation between energy and loudness (0.78). This suggested that tracks radiating with higher energy are intrinsically louder, emphasizing a close-knit relationship between a song's vibrancy and its volume. Similarly, the correlation between danceability and valence (0.52) unveiled an interesting insight: the more danceable a song is, the more likely it exudes positivity or cheerfulness. On the flip side, the data highlighted strong negative correlations as well. Tracks pulsating with higher energy levels seemingly had a lesser acoustic touch, a correlation pegged at -0.75. This trend provided an inkling that the more energetic tracks in this dataset might lean more towards electronic vibes or perhaps are layered with richer instrumentation. Furthermore, the loudness of a song seems inversely proportional to its acousticness, corroborated by a -0.75 correlation. This relationship strengthens the hypothesis of louder tracks being more electronically inclined.

The dataset also shed light on a subtler relationship between valence and mode, with a correlation of -0.32. This indicated that songs resonating in a major mode, coded as 1, are intrinsically associated with a higher degree of positivity. Conversely, tracks in a minor mode, coded as 0, evoked more somber or melancholic undertones. Venturing into moderate correlations, it was observed that danceability and energy shared a 0.14 correlation. This insinuated that while danceable tracks are somewhat energetic, the association isn't overpoweringly strong. The interplay between loudness and tempo (0.27) subtly hinted at louder tracks perhaps having a brisker tempo. Intriguingly, several features showcased weak or negligible correlations. A case in point is popularity, our dependent variable, which didn't demonstrate a strong correlation with any other audio features. This suggests the multifaceted nature of popularity, possibly influenced by a myriad of factors, both intrinsic to the song and external. The most robust associations revolve around loudness, energy, and acousticness, painting a vivid picture of a song's inherent nature. Elements like a track's popularity remain enigmatic, influenced by an eclectic blend of factors.

After that I tried to decipher the importance of each variable in predicting song popularity by generating both mutual information and F scores. The results were very interesting. Based on the resulting Feature importance ranking table showcasing mutual information scores for different musical features, 'duration\_ms' emerged as the feature with the highest mutual information value of 0.040805, suggesting it shared a significant relationship with the popularity of a song. Following this, 'tempo' and 'instrumentalness' held relatively higher scores, with values of 0.021862 and 0.018756, respectively. In contrast, the 'key' feature exhibited the lowest mutual information score, a mere 0.002905, indicating its minimal relevance or predictive power concerning song popularity. The table thus provided a hierarchy of features based on their relevance, from 'duration\_ms' at the top to 'key' at the bottom, which were very instrumental understanding the intervariable dynamics.

The resulting F-scores table showed the impact of the various audio features variables on song popularity. The F-score is a statistical metric used to measure the degree of differentiation a particular feature provides among various levels of the target variable, in this case, song popularity. A higher F-score indicates that the feature significantly distinguishes between popular and less popular songs, while a lower score suggests the feature might not be as influential. 'Instrumentalness' took the lead with an F-score of 27,759.797438, indicating it is a paramount feature in determining song popularity. This suggested that the level of instrumentality in a song plays a significant role in its popularity rating. Following this was 'duration\_ms' with an F-score of 16,513.091298, hinting at the possibility that the length of a song could be a strong predictor of its popularity. 'Loudness' also emerged as a crucial feature with an F-score of 12,704.955055, implying songs with particular loudness levels might resonate more with the audience.

Further down the list, attributes like 'danceability', 'acousticness', and 'liveness' also had notable F-scores, suggesting they have moderate influence over a song's popularity. On the other hand, 'tempo' and 'key' registered exceptionally low F-scores of 7.637441 and 0.176776, respectively. This implies that the tempo of a song and its key might not be pivotal determinants of its popularity among listeners.

Lastly, I generated histograms and boxplots to understand the distribution and outlier composition of each variable in the dataset.

A group of graphs showing different types of data

Description automatically generated

A group of blue squares

Description automatically generated

The dataset revealed several interesting distributions about the musical features. 'Danceability', 'Energy', and 'Valence' generally follow a normal distribution, with the majority of songs clustering around the median. The 'Key' feature has a uniform distribution due to its categorical nature representing various musical keys. Interestingly, 'Mode' is majorly skewed towards the major mode, suggesting that most songs in this collection are set in a major key. Features like 'Speechiness', 'Acousticness', 'Instrumentalness', and 'Liveness' lean positively, indicating songs are primarily musical, electronic, vocal-centric, and more of studio recordings rather than live. 'Tempo' displays a multimodal pattern, especially around 100-120 BPM, while 'Duration\_ms' is positively skewed, hinting at most songs falling within the typical song duration.

Outliers are evident in certain features. 'Loudness' and 'Duration\_ms' stand out, with some tracks being exceptionally quiet and others having atypical durations. An unusual observation is a song with nearly zero tempo. While songs with time signatures other than 4/4 are technically outliers, this is primarily because of the overwhelming popularity of the 4/4 rhythm in modern music.

Box plots from the dataset gave a closer look at the medians and interquartile range (IQR). 'Speechiness' and 'Liveness' show a lower median, suggesting that more than half of the tracks have minimal speech and are studio recordings. In contrast, 'Mode' and 'Acousticness' leaned towards a higher median, showing that many songs are in the major mode and possess significant acoustic elements. 'Key' and 'Loudness' exhibit medians more centrally located within their range.

In essence, this dataset encapsulates characteristics common to popular music: a predominant 4/4-time signature, average tempos of 100-120 BPM, and a bias towards higher energy levels. The data underscores prevailing trends in the musical style of the dataset, with a significant leaning towards major modes. Outliers in certain features emphasize the dataset's inclusivity, capturing songs that deviate from the norm, enriching its diversity.

**Test and Evaluation**

As I approached the final steps of the project, the culmination of my efforts provided me with 9 fully trained models each ready to be tested to predict a song’s popularity. Along with several other tests run on the dataset I had found the answers to the research questions I set out with at the start of this project, namely:

* *What audio features most significantly predict a song's popularity on platforms like Spotify?*
* *How can predictive models informed by stock market forecasting techniques be adapted for the music industry? Can it have high accuracy?*
* *What is the correlation between Spotify's audio feature data and the commercial success of a song across different revenue streams?*
* *How can music labels and artists use quantitative data to make informed decisions about which songs to promote?*
* *To what extent can historical Spotify data be used to create an accurate popularity scoring system for new music tracks?*
* *What challenges exist in creating a predictive algorithm for song popularity that is robust across various music genres and artist profiles?*
* *How does the predictive model account for the dynamic and evolving trends in music consumption and audience preferences?*

A screenshot of a computer code

Description automatically generated A screenshot of a computer program

Description automatically generated

When it came to the idea of determining how the spotify audio features could affect the popularity of a song, my approach was a supervised learn regression model utilizing a variety of algorithms to train and predict how popular a song can be. I chose this approach over a more general classification approach because I wanted to get into the minutae of the matter in terms of what variables stood out more in predictive and how much of an effect they had if any. That way by isolating specific audio features that proved essential in determining popularity, they could be manipulated for better results by the end user. My mutual info and Anova F score results above showcase the rank by which the variables help in predicting song popularity as well as how well they change along with it. Below you will find that all the variables used in this project to train the models proved to be significant in predicting song popularity though it also shows that they do not even come close to painting the whole picture. The prob F- statistic proves the model to be statistically significant as a whole so I believe the main takeaway is that there is so much more to be added when trying to predict a song’s popularity. I’m on the right track but I need more variables rather than just audio features.

A screenshot of a computer

Description automatically generated

As I mentioned earlier, each model was tested on the test set and their predictions were evaluated using the following performance metrics:

* *Root Mean Squared Error (RMSE)*
* *R-squared (R²)*
* *Mean Absolute Error (MAE)*
* *Explained Variance*
* *Median Absolute Error* *(MEDAE)*
* *Adjusted R-squared* *(Adj. R²)*

The results of these tests on their predictions showed how well the models were good at efficiently predicting the songs’ popularity as well as how much of the popularity’s variance they explained. Basically, they showed how much of the picture we’d mapped out when it came to song popularity. After testing each model for their performance metric, I tried tuning the hyperparameters for the most promising models trained using the Random Forest, XGBoost regressor and Neural Net algorithms. However apart from XGBoost they all took too much time, memory and crashed so I scratched them. I also played around with transformations, involving the dependent variable and not just to see how it affected performance. I played around with the dataset split ratio as well before finally settling on what I had. Ultimately, I ran one last round of model training and compiled the performance metrics of each algorithm trained model into one table color coding the lowest value as yellow and the highest as red.

**Results and Evaluation**

A screenshot of a computer

Description automatically generated

Given that I already explained each performance metric in a previous section I won’t reiterate them here but basically, I wanted the lowest possible RMSE, MAE and Median Absolute Error or MedAE and the highest possible R-Squared, Explained Variance, and Adjusted R-Squared, hence all the color coding to allow easier comparison. Based on the results seen above it’s obvious that the best performing model is the XGBoost Regressor that I used Random search on to find the best possible parameters ergo Best XGBoost Regressor. It boasts the highest R-Squared, Explained Variance, and Adjusted R-Squared indicating the best fit among all the models while also having the lowest RMSE and MAE indicating that it predicted the song popularity values for each set of audio features with relative accuracy. Also, MedAE was among the lowest so it’s still stellar. Other notable performances were Neural Net and Random Forest for their relatively high R-square and Adjusted R-square showcasing that they captured the change in song popularity fairly well. Alternatively, Decision Tree was a disaster showcasing a negative R-Squared, Explained Variance, and Adjusted R-Squared showcasing that it was a very poor fit for the dataset. It performed the total opposite of our best model.

Something to take away from this would be that complex models like the Best XGBoost Regressor and Neural Network seem to perform better, which could indicate that the relationship between audio features and song popularity is non-linear and possibly involves interactions between features. Also, the Adjusted R-squared for the Best XGBoost Regressor being higher than the R-squared seemed to suggest that the number of predictors used by the model were perfect and didn’t excessively inflate the performance metric.

Diving deeper into the XGBoost Regressor as it counts as my final model since it performed the best, analyzing each performance metric showed that when it came to predicting the popularity of a song on the platform using Spotify audio features, the metric which ranges from 0-100, the model’s prediction were about 14.16 units aways from the true popularity score. That’s not bad considering the constantly changing nature of a song’s popularity. And for Adjusted R-squared, it showed that the model was able to explain about 20.5% of the variance in song popularity. Though that’s not much, it is significant showing that how a song sounds does have an effect on its popularity but they aren’t the only things affecting it namely non audio factors such as marketing efforts, artist popularity/public profile, societal trends/virality/taste, timing, platform recommendation algorithms, playlist placement, radio airplay, Live performances, social media like TikTok, commercials, use in viral challenges/memes/skits, etc. Honestly, I’d classify this project as a moderate success.

**Final Thoughts & Conclusions: Challenges and Obstacles Encountered**

A major challenge of this project was handling the size of the dataset. When I was initially trying to analyze the data on tableau it crashed and never worked so I decided to switch to a backup dataset that was smaller but had similar variables, basically a subset. I was going to go forward with this new smaller dataset however after running it through the wringer on my python notebook I decided to give my original enormous dataset another try on this new platform. I will admit that the processing times have been long and tedious with some singular cells of code for a model running for about 20 to 30 minutes. I even had to drop the support vector model because it was taking too long and crashing the notebook.

Another challenge I faced was the fact that my evaluation metrics ended up overall being poor. Nothing seemed to be fitting well according to the Adjusted R-square and the errors weren’t as minute as I would have wanted in a perfect world. With the completion of the project and after analyzing the evaluation metric table I set up to comparing the results of all the different models, I still feel like there’s so much that can be done. Even after I did all the transformations I felt were necessary to the different variables and playing around with them, in the end the performance metrics I was looking for didn’t seem to materialize which just led me to come to the conclusion and accept the fact that just the Spotify audio features can’t be used to fully predict the popularity of a song on the platform. I even tried Hyperparameter tuning on the best performing XGBoost regressor only to find that there was barely a noticeable improvement in model performance. That just sealed it for me and forced me to agree with the research literature I found that much more goes into predicting a song’s popularity than just how it sounds.

With that said, the model is not useless and has a bunch of significance in a business setting for an entity such as a music label or artiste management. It would still be a very useful tool for any artiste looking to decide on what song to release next to the public. As I mentioned earlier in the paper, an ideal scenario would be to quantify the audio features of each song up for consideration, feed them to the model and use those numbers to gauge what song would work best in the current music scene. Of course, with the constant change in people’s music taste the model will have to be constantly trained on new data to stay current. Also, I thought of transforming the model into a binary classification one rather than prediction just to give more certainty. I already started work on that and I found a success rate of about 70-75% in terms of accuracy. When it comes down to it, anyone trying to replicate this project would have to have extremely fast processors with enormous memory to push it further. I wish I’d been able to perfectly tune the other models to give them the best possible chance but the engine and accelerator I used couldn’t handle that. I really enjoyed this project and hope to revisit it someday and improve upon it.

**Research Literature References**

* Araujo, C. S., Cristo, M., & Giusti, R. (2019). Predicting Music Popularity on Streaming Platforms. *Anais Do Simpósio Brasileiro de Computação Musical (SBCM 2019)*. <https://doi.org/10.5753/sbcm.2019.10436>
* Yutong, G., Yutong, S., & Jiaqian, W. (2021, May 26). Popularity Prediction of Music Based on Factor Extraction and Model Blending. Ieeexplore.ieee.org. <https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=9434816>
* Beniwal, R., Gupta, S., Shekhar, S., & Batra, S. (2023, August 7). Hindi Hit Songs Prediction Using Machine Learning Algorithms. Ieeexplore.ieee.org. <https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=10205626>
* Kamal, J., Priya, P., Anala, M. R., & Smitha, G. R. (2021). A Classification Based Approach to the Prediction of Song Popularity. 2021 International Conference on Innovative Computing, Intelligent Communication and Smart Electrical Systems (ICSES), Innovative Computing, Intelligent Communication and Smart Electrical Systems (ICSES), 2021 International Conference On, 1–5. <https://doi.org/10.1109/ICSES52305.2021.9633884>
* Martin-Guiterrez, D., Hernández-Peñaloza, G., Belmonte-Hernandez, A., & Álvarez García, F. (2020, February 24). A Multimodal End-to-End Deep Learning Architecture for Music Popularity Prediction. Ieeexplore.ieee.org. <https://ieeexplore.ieee.org/abstract/document/9007339/citations#citations>
* Pham, J., & Kyauk, E. (2015). Predicting Song Popularity. <https://cs229.stanford.edu/proj2015/140_report.pdf>
* Essa, Y., Usman, A., Garg, T., & Singh, K. (2022). Predicting Song Popularity Using Machine Learning Algorithm. <https://ijsret.com/wp-content/uploads/2022/05/IJSRET_V8_issue2_281.pdf>