Spring 4-17-2022

Predicting and Composing a Top Ten Billboard Hot 100 Single with Descriptive Analytics and Classification

Matthew DeAmon
mdeamon@bgsu.edu

Predicting and Composing a Top Ten Billboard Hot 100 Single with Descriptive Analytics and Classification

Matt DeAmon

HONORS PROJECT

Submitted to the Honors College
at Bowling Green State University in partial fulfillment of the requirements
for graduation with

UNIVERSITY HONORS 17 APRIL 2022

Kyle Moninger, Advisor

Department of Applied Statistics and Operations Research

Dr. Kenneth Thompson, Advisor

College of Musical Arts

Abstract

In late 20th and early 21st century Western popular music, there are cyclical structures, sounds, and themes that come and go with historical trends. Not only do the production techniques utilized reflect technological advancements (the Yamaha DX7, the Roland 808, etc.), the art form reflects contemporary cultural attitudes through lyrics and stylistic choice. Through this lens, pop songs can serve as historical artifacts for their unique ability to captivate listeners based on their generally acceptable and familiar elements, both upon release and with future audiences. It raises the questions: "Can a chronological analysis of artistic choices reveal trends in songwriting and popular music composition?"; "Based on collected analysis, could forecast data suggest criteria that a future hit song may fit?"; and "How could the next 'hit song' sound, based on the calculated criteria from trend analysis and forecasting techniques?" By manually listening to and analyzing Billboard songs for each of the last 50 years and employing an assortment of feature selection and classification techniques, a random forest model predicts some of the significant characteristics of a potential future hit song. This prediction provided the framework for an original composition.

Introduction

It is a feat for any form of artistic expression to achieve mainstream success for three reasons. First, popular art requires a delicate balance between familiarity (so as not to alienate new audiences) and novelty (so as to keep them interested). "Popularity" is not necessarily synonymous with "groundbreaking," "innovative," or "avant-garde." Critics and musicologists may assess a song's greatness by its originality, musicality, or lyrical depth. However, these are the opinions of a select few; casual listeners judge music with wide, varying levels of scrutiny.

Second, the distribution of art is a business, with buy-in not only from the original artists, but also the producers, managers, promoters, and other collaborating artists. To pass through so many listeners and maintain enough support to be completed is a testament to the original artists and the industry as a whole. Ultimately, audiences have a say in the success or failure of a work, and as such, a unique quality of popular music lies in its ability to share a relationship with the collective audience and individuals.

Third, if recognizability and notoriety are any reflection of the zeitgeist, popular music captures it. Throughout the decades, historical limitations and technological progress in music are an unfolding story that, through close listening, can reveal more about the given period. For example, many songs of the 1970s have a runtime less than the average Top 25 track runtime. This may be because AM radio required a song to be edited to fit for radio airplay. Another example is the prominence of electronic drums in late 1980s tracks. This "gated reverb" effect is ubiquitous with the 80s sound, emulating the energy and vibrance present in contemporary popular culture.

Similarly, songs can be popular through their association with other events. "At This Moment" by Billy Vera & The Beaters has a noticeably different sound from other songs in 1987. Initially released in 1981, its success is linked to usage in an episode of "Family Ties," a popular 1980s sitcom. Interestingly, this also exemplifies how a song's influences extend beyond the initial point of release. 1970s soul artists like The Manhattans and Bootsy Collins have been compared to Bruno Mars and Anderson .Paak on their 2021 collaboration, "Silk Sonic," which features characteristics of 70s soul such as heavy orchestration, Latin auxiliary percussion, and disco grooves.

All of this serves to suggest that popularity is a nebulous term. Therefore, the measurability of popular music can be quite challenging. Of the countless metrics used, the Billboard Magazine charts are among the most well-known. To calculate which songs appear in their chart, Billboard Magazine analyzes the number of streams, radio rotation, album/track purchases, social media activity, and other measures of mainstream success in a digital distribution age. With the introduction of new recording technologies, Billboard has adjusted its metrics: for example, in 2013, online views were included in the measure of a song's success in response to the "Harlem Shake" popular dance trend.

The broad definition of popularity, joined with historical context, observed cyclical trends, and the prominence of the Billboard Magazine Year-End Top 100, converge to prompt further analysis into what characteristics of a song make it popular.

Methodology

Project Overview

This project was comprised of three parts: data collection, forecasting, and song composition. Popular songs selected for analysis were derived from the Billboard Year-End Hot 100 charts.

Only the top 25 songs of the Billboard Year-End Top 100 released between 1971 and 2021 (inclusive) were used in this study. Given the project duration and resource constraints, any amount beyond this would have been infeasible. Each of the 1275 tracks was analyzed manually for approximately 32 data points. Loosely, these data points can be grouped into three categories: Composition, Instrumentation, and Production.

Composition encompasses the layout of the track. These are the most objective and quantifiable observations for each song since they are generally accepted from listener to listener more so than the other categories:

- **Runtime** – the given length of a track. When possible, the official full version was analyzed (i.e., not an EP recording, demo, or radio edit). For composite tracks which combine two or more songs often played together, only the first song was recorded. For example, Queen's "We Will Rock You/We Are the Champions" are separate compositions but run together on the album. Only the length of "We Will Rock You" was used (2:02).
- **Genre** – the stylistic category for the track and/or artist. Predetermined genres were drawn from the Recording Industry Association of America (RIAA). As of 2015, there are 22 official genres to categorize this music. For the simplicity of this project, 18 were selected, but in practice, only 13 were used.
- **Beats Per Minute (BPM)** – the set tempo for a song. Tempo was extracted from tunebat.com, an online music metatag database. Because tempo is strict and rhythmic in nature, one may assume that tempo is the most specific and unalterable measure of a song. However, the pulse or "feel" of a song is changeable based on artistic choice. For example, "I Knew You Were Trouble" by Taylor Swift clocks in at 77 BPM, likely because the chorus drops to a half-time feel. Half-time is a technique where the emphasized downbeat is extended over two measures. Commonly used in 4/4 time signatures, this appears when the 2 and 4 downbeat falls on 3.
- **Groove** – to combat confusion with half-time and capture more about a song's style beyond its BPM, a complementary measure was introduced. 15 different grooves were established based on the syncopation and percussion techniques used in each song. This facilitated the composition portion of this project. To review each groove and examine reference tracks, please consult the Appendix.
- **Time Signature** – the meter in a composition. While originally intended to capture 6/8, 3/4, 2/4, and 4/4, this was ruled redundant and unnecessary through the data collection stage. An overwhelming majority of American popular music is in 4/4.
- **Tonality** – a variable to indicate whether a song is written in a major or minor key.
- **Starting Instrument** – a categorical variable that records which instrument is heard at the beginning of a song. "Ambience" refers to synthesized sounds played only at the introduction of a song. This varied by decade: the 1980s primarily featured billowing synthesizer sounds,

the 1990s used talking soundbites, and 2000s and beyond would use a mix of both. "Other" was typically reserved for strings (orchestra, fiddles, etc.)

- **Extended Intro** – a binary variable indicating whether the beginning of a song has a different tempo or groove from the rest of the song. These introductions fall outside the established meter. For example, "Maggie May" by Rod Stewart features an acoustic guitar solo in 3/4, noticeably slower than the rest of the song.
- **Verses** – the number of unique phrases following the same lyrical melody in a song.
- **Pre-Chorus** – the number of repeated phrases or lines falling between a verse and the chorus.
- **Refrains** – the number of choruses in a song.
- **Bridges** – the number of contrasting sections in a song using a different progression from the rest of a song. In the 1980s and 1990s, rap was popularly used in the bridge of a song.
- **Solo** – the number of sections highlighting a singular voice (guitar, synthesizer, vocals, drums, etc.) in a song.
- **Outro** – a categorical variable designed to capture how a song is finished. Eight different outros were recorded.
- **Key Change** – a binary variable indicating if the song changes key, either major to minor, minor to major, or within the tonality.
- **Breakdown** – a binary variable indicating whether the song includes a portion where the backing instruments are temporarily stripped away to focus attention on one or two voices.

Instrumentation encompasses the presence or absence of an instrument in a song. Moreover, it also covers any identifiable and predictable trends in the presence or absence of each track, typically around the vocals. A majority of these measures are binary variables. The following were used to measure Instrumentation:

- **Lead Singer** – a categorical variable indicating the gender of the singer (Male, Female, Both, and None)
- **Harmonies** – the use of overlaid vocal tracks to lift the main vocalist with supporting melody lines.
- **Backups** – the presence of additional voices (either the main vocalist's or another singer) appearing outside the vocalist's main melodic line.
- **Vocalization** – singing a melody without using words. Popular in the 2010s, these "oohs," "woahs," and "ahs" typically appear post-chorus.
- **Falsetto** – any instance where the main vocalist switches from chest singing to throat singing. While this is normally observed in male singers, I have used the definition of falsetto liberally to also measure artists who incorporate whistle tones or exceptionally high singing parts.
- **Ad Libs** – an impromptu similar to vocalization without the intent of the vocal line to be repeated by the audience. A single vocalist shouting "Yeah," "Woo," or "Hey" is a prime example of a vocal ad lib. In the 1970s, these were used commonly near the end of a track during the fade out.

- **Drums?** – whether or not the song includes drums.
- **Auxiliary Percussion** – whether or not the song includes any auxiliary percussion, including (but not limited to) orchestral percussion, clapping, snapping, tambourine, and claves.
- **Bass?** – whether or not the song includes a bass instrument.
- **Rhythm Guitar?** – whether or not the song includes a rhythm guitar.
- **Lead Guitar?** – whether or not the song includes a lead guitar.
- **# Keyboards** – the number of keyboard voices used in the song. For example, "Do You Know What I Mean" by Lee Michaels repeats the main riff on a chapel-style organ doubled by a concert piano. This song receives a # Keyboards score of 2. A 0 indicates no keyboards were used.
- **Orchestra/Strings?** – whether or not any stringed instrument voices beyond the bass and guitars were used. Songs with a fiddle or orchestral accompaniment fit this category. Synthesized orchestra sounds, such as those used in "Crank That (Soulja Boy)" by Soulja Boy would qualify for this category.
- **Horns?** – whether or not any brass instruments are used in a piece. This can include a single trumpet or French horn, or more commonly, this could be a 3- or 5-piece brass section.
- **Winds?** – whether or not any woodwind instruments are used in a piece. In American popular music, the tenor saxophone is commonplace. Flutes, soprano saxophones, and other woodwinds are used on occasion. "Horns?" and "Winds?" demonstrate very high variable interaction. In other words, songs that exhibit one of these features will very likely include the other.
- **Sample/Cover?** – whether or not the song borrows from a pre-recorded track. Samples deliberately take portions of an earlier song and re-use it for artistic purposes in an entirely new piece. For example, "Regulate" by Warren G and Nate Dogg samples Michael McDonald's "I Keep Forgettin'." Covers, by contrast, are instances whether artists take the original song and reproduce it in a different style. The "Baby I Love Your Way/Freebird Medley" by Will to Power combines Peter Frampton's "Baby I Love Your Way" with Lynyrd Skynyrd's "Freebird" to create a new ballad.
- **Spoken/Soundbites** – whether or not the song includes talking or spoken words. Rap verses are not simply spoken word. "The Streak" by Ray Stevens humorously uses this to recount the witness testimonies of a streaker in his 1974 #8 hit. Other artists will use it to acknowledge producers during the song, such as the announcement of "We the best music" at the beginning of "I'm the One" by DJ Khaled.

Production encompasses the stylistic choices that enhance the chosen instrumentation in a song. These are the most subjective measures in the dataset. A 0-10 scoring scale was used to interpret the acoustic vs. synthesized quality of an instrument. A score of 0 was reserved for tracks where an instrument was absent; after all, the exclusion of a track is an artistic choice that must be reflected in a dataset. 1 is for instruments with no added electronic sounds (acoustic guitar, upright bass, no drum machines), while 10 captured songs with heavily synthesized sounds (EQ bass, drum machines, etc.) For an example, please see the Appendix. This manufactured sound scale (M-Scale) is solely based on my listening, and as such is only useful for me when preparing for the composition portion of this project.

- **Bass M-Scale**: the natural (acoustic) or manufactured (synthetic) production modifications of the bass in a song.
- **Lead Guitar M-Scale**: the natural (acoustic) or manufactured (synthetic) production modifications of the lead guitar in a song.
- **Rhythm Guitar M-Scale**: the natural (acoustic) or manufactured (synthetic) production modifications of the rhythm guitar in a song.
- **Drums M-Scale**: the natural (acoustic) or manufactured (synthetic) production modifications of the drums in a song.
- **Keyboards M-Scale**: the natural (acoustic) or manufactured (synthetic) production modifications of the keyboard in a song.

Interdisciplinarity

Per the definition provided by the BGSU Honors College, interdisciplinary research involves "the combining of methods and insights of two or more academic disciplines into the pursuit of a common task." This project draws upon theory and applications from the realms of music composition and statistics. At its core, the project creatively combines the challenge of writing popular music (an arbitrary art) with the precision and predictive capabilities of analytical programs. Combination of disciplines for this project is a must, and as such, it is interwoven at every stage.

Data Collection

Music Sample Selection

The Central Limit Theorem states that a sufficiently large enough sample from a random independent population tends toward a normal distribution. The equation follows:

$\sum x=(n)(\mu) + (z)(\sqrt{n})(\sigma)$, where
$(n)(\mu)$ = population mean of the random variable, x
σ = population standard deviation

Using the Central Limit Theorem, a desired confidence interval of 99%, and 3% margin of error, a population of 5100 songs would require a minimum sample of 1354 songs, or about 27 songs per year. Given the time constraints for this project, I settled for 25 observations for each year.

The Billboard Year-End Top 100 charts use an aggregation of streaming, broadcasting, and sales to determine the top songs of the year. This can lead to some error in properly measuring popularity strictly within a given calendar year. Additionally, songs released after late November are not considered in the standings because the holiday season impacts streaming success, and a charting single cannot compare in significance to songs that have been released for a longer period of time.

Several Spotify users with anonymous aliases have created public-facing playlists with the top 100 Billboard Year-End singles of each year. This expedited the song searching process and allowed for quicker listening. The Length, Tonality, and BPM of each song was extracted from tunebat.com, a free database for harmonic music data. In as many instances as possible, the entire song was played and analyzed for its Structure, Production, and Instrumentation. However, in some cases, only portions of repetitive songs were played, which may give way to a small factor of error.

Song Characteristics and Trends

There are several observations that can be drawn just by dissecting the collected data. In no particular order, some are shared below.

Ballads. For a majority of the dataset, "Ballad" is consistently present in the Top 10. Songs categorized as using the Ballad groove may not capture every song colloquially called the "Ballad." "Careless Whisper" by Wham! is recorded as using a "16th pattern," though its prominent saxophone solo, soaring vocals, and dramatic chorus suggest it could also be identified as a "Ballad." The largest gap of ballads is between 2008 and 2011.

Auxiliary percussion over the years. Many different production techniques change between genres, years, and artists. Auxiliary percussion is notable for its constantly evolving role in the songs. Heavy orchestration in the 1970s resulted in timpani, chimes, and triangles. Synthesizer effects of the 1980s introduced claves and inverted cymbals. These continued into the 1990s, which switched periodically with hand claps and synthesized triangles (discerned from regular triangles by their uniformity and short decay). The 2010s and early 2020s are observed to use a combination of handclaps and finger snaps on the downbeats.

Dominant grooves. The dominant groove associated with a decade of music does not emerge until the later years. Disco is a prime example. In the mid-1970s, "Four on the Floor" grooves had mild success. From 1977-1979, a significant portion of the Top 25 songs followed this style. Additionally, the first two years of the following decade have carryover influences as the dominant style is shifting. Another example is the tresillo (triplet subdivided over two), which was popular in the 2010s and famously used in "Shape of You" by Ed Sheeran.

Data Analysis

Initial Obstacles

The major obstacle confronted in the analysis phase pertained to selecting a model that could produce sufficient predictions and could be built within the allotted timeframe for project completion. In an ideal environment, there would be ample time to tune models of different varieties using a validation dataset, then compare the best with each other to identify the "best of the best" models with a test dataset. Given the semester schedule and the immense length undertaken to collect data, the decision tree model was selected due to its simplicity and low requirements to process.

Despite collecting over 1200 records, several challenges had to be overcome before beginning the song analysis. First, the problem of fitting data to the model that could provide a full prediction was exacerbated by the "Curse of Dimensionality." This occurs when the number of dimensions (columns) comes at odds against the number of records (rows). For each additional dimension, more records are required in the sample to accurately depict the population.

The complete dataset includes 32 observations, already pushing the limits of a 1275 song dataset. After preprocessing, this leaves approximately 78 variables a computer model must juggle to form its predictions. Using all 78 dimensions overloads the processor and offers too much noise to sift through. Therefore, feature selection was required.

Additionally, the proposed plan to prescribe the features of a song without user input was abandoned due to complexity and time/resource constraints. Multivariate calculus and multiclass classification fall outside the foundations of my undergraduate curriculum. Per the advice of my advisors, we opted for a model that was feasible and reflective of my coursework. To that end, we recoded the 25 songs from each year based on their Top 10 status (Top 10 Hit Song = 1, Not a Top 10 Hit = 0), creating a binary classification problem. Not only is this kind of model more workable, but it also matches my curriculum.

Aggregated by year, the quantitative variables were averaged to provide insight into production trends between the selected time periods. Graphs and commentary are referenced throughout this section and provided in the Appendix.

Feature Selection

Redundant and unnecessary variables were removed to avoid confusion. For example, a Drums M-Scale score of 0 means there are no drums, which is also recorded as N in the "Drums?" dimension. The "Drums?," "Bass?," and "Harmonies?" dimensions were removed because an overwhelming majority of Hit and non-Hit songs includes these 3 variables. This was deduced by simple observation.

Binary variables can be removed based on their significance when measured against our binary target classification variable, "HIT_SONG." Logistic regression with least absolute shrinkage and selection operators (LASSO) regularization was used to determine the strength of each variable's impact on the Hit/No Hit standing for a track. Graph 2A illustrates the coefficients of each binary variable, ordered by their absolute value. The relative strength of each coefficient, ranging from -1 to 1, reveals which variables play a more significant role based on a song's Hit status due to their absence or presence. For example, the variable "Backups?" measures where there are backup vocals (Yes/No). The coefficient 0.27325233 indicates that, when backup vocals are used in a song, its likelihood of being a Top 10 Hit increases by 0.27. Conversely, if a song includes a "Sample/Cover," its likelihood of being a Top 10 Hit decreases by -0.25722527. All told, I kept the 7 most significant binary variables in the dataset.

Numerical variables can be reduced with Principal Component Analysis (PCA), a technique that measures how much of the variation can be explained by each additional variable. It does not dictate which variables should be removed; rather, it indicates how much a variable contributes to a song's status as Hit/No Hit. From this, I decided to drop the number of Keyboards and the M-Scales for Keyboards, Rhythm Guitar, and Lead Guitar.

While working through this phase, I recognized that the variation between counts of Verse, Pre-Chorus, Chorus, Bridge and Solos was negligible. In fact, by concatenating these measures into one, I learned there were only 253 unique song structures. And, with some manipulation, these 253 could be reduced to 13. This conversion of five distinct numeric variables into one categorical provided greater insight.

Categorical variables are not stored as numbers; however, counting the expected number of instances for each category and comparing it to its actual recorded amount sidesteps this issue. The chi-squared goodness-of-fit test tabulates these actual and expected appearances and determines whether the difference in amounts is significant. This was not the case for many variables. For example, the "Lead Singer" categorical variable broke into four possibilities: Male, Female, Both, or None. Songs sung by a Male were just as frequently recorded as Hit songs as they were *not* Hit songs. The most significant variable was Outro, with 8 possible categories, followed by Tonality (Major/Minor/Both).

In the end, the following variables were carried forward, based on their significance in a song's Hit/No Hit status: Runtime (in seconds), BPM, Drums M-Scale, Bass M-Scale, Tonality, Auxiliary Percussion, Backups, Vocalizing, Falsetto, Sample/Cover, and Outro.

Decision Tree, Random Forest, and Extreme Gradient Boosting (XGBoost)

The aforementioned Decision Tree model was used to establish a baseline and provide a glimpse into the purity potential for each branch split. With no constraints, the decision tree takes in a new characteristic at each internal node and, based on the samples being measured, separates the data records based on the criteria. The "purest" decision tree—that is, the tree which splits the records into as many leaves as possible with little overlap between records on different leaves—has 12 branches and uses log2(25) to determine how many features to consider at each tree split. The single decision tree has low sensitivity and specificity rates, (58.78% and 52.94%, respectively), and with a 50.9% accuracy rate, its predictive power is not much better than a coin flip. Sensitivity measures the model's ability to accurately capture relevant items—that is, Hit songs that are in fact Hits. On the other hand, specificity is the model's ability to avoid irrelevant items—that is, correctly categorizing songs that aren't Hit songs as "No Hit."

Fortunately, there is an advanced analysis method that takes the power behind decision trees and compares it with the wisdom of the masses. Random forests are a collection of smaller, simpler decision trees, each built with their own subset of the training dataset and measuring their Hit/No

Hit status using different dimensions. For example, one tree may have 45 random songs and split the data based on their BPM, Outro, and Backup vocals. Simultaneously, another tree could have 135 songs and split the data based on their BPM, Falsetto vocals, Tonality, and Auxiliary Percussion. This process is called bagging, a portmanteau of bootstrapping and aggregating; each tree works with its smaller observations and dimensions and the predictive results are aggregated to classify a song as Hit/No Hit.

Heavy overfitting can occur with a random forest. With overfitting, a model is trained to perform too well according to the training dataset—once unseen data is introduced, such as that of the new test dataset, the model's accuracy critically underperforms. The Python programming language offers a function called "product," which allows the user to input a list of combinations and the device can try each combination until it returns the highest value for the set inputs. In this example, I provided a set of estimator values, the maximum number of features, and the maximum depth for each tree. The program tested each combination until it could tune the hyperparameters of the tree to provide an accuracy of 58.3%, higher than our single decision tree.

Even so, hyperparameters can only be refined before the classification splitting occurs. A non-linear tuning technique, called Extreme Gradient Boosting (XGBoosting), builds on the mistakes of prior predictions to continuously update weights when classifying songs as Hit/No Hit. With each successive tree build, the errors of the previous are used to guide the model and parse off boundaries between predicted Hits and No Hits until a sufficient border between them is produced. For this reason, Extreme Gradient Boosting is very susceptible to overfitting. Fortunately, the Python XGBoost library can call a function that will try a set number of possibilities to determine the best fit for the given dataset. It acts very similar to the previously mentioned "product" function in the base Python library, only with less manual entry by the user. With the best parameters selected, the boosting yields a 90%+ accuracy rate for the training dataset.

As expected, the model performed much lower with the test dataset. I adjusted some of the "best" parameters to mimic early stopping, which keeps the model from continued fitting past the point of effectiveness. Other adjusted parameters for the XGBoost included raising the required number of samples for each internal node. Even though trees are designed to split until reaching the lowest impurity level, the perfect splitting means that each record from the sample has its own leaf. This is perfectly fit for the training dataset but performs poorly on test data. Additionally, I reduced the number of estimators (trees) from 300 to 250 and applied a minimum child weight of 1.5 (up from 1) to force the trees to only build extra branches that could sustain more samples.

From a songwriting standpoint, I found a song's key to be a critical guiding component to songwriting. Therefore, I input two suggested songs with a variety of characteristics until the model returned a predicted score of Hit (1).
The song has to include or resemble the following features:

Tonality: Minor
Runtime: 3:35-3:47
BPM: 110
Structure: 2-2-3-1-0 (Verse I; Pre-Chorus; Verse II; Pre-Chorus; Bridge; Chorus)
Outro: Isolated Track or Ritardando
Drums M-Scale: 6
Bass M-Scale: 3
Vocals: Backups, **Harmonies**

Tonality: Major
Runtime: 2:55-3:08
BPM: 120-30
Structure: 2-0-3-1-0 (Verse I; Chorus; Verse II; Chorus; Bridge; Chorus)
Outro: Ad Lib Fade
Drums M-Scale: 7
Auxiliary Percussion - Yes
Bass M-Scale: 6
# of Keyboards: 1
Keyboard M-Scale: 4
Vocals: Backups, Harmonies

Music Composition

Writing

With two "blueprints" in hand—one for a minor key song, one for a major—music composition took less than a week to complete. What could not be explicitly predicted by the random forest model, we relied on contemporary music trend observations to inspire the songwriting process. We opted for the Major key blueprint.

Our chosen key was A major, and BPM was near 120. The chord progression in the verses follows I-VI-V-I. I, IV, and V were used in a monophonic rhythm (every instrument emphasized the same beats) to build excitement. A 16th pattern circulating the tom drums, with snare on 2 and 4 provide the backbeat. A simple electronic keyboard effect and treble-heavy guitars produced an ethereal tone.

Verse I: A | F#m | E | A

A | F#m | E | A/C#m

Chorus: D E | A E D | D E | A E D

D E | A G#m F#m | D E

Post-Chorus: A | A

Bridge: E | C#m | B | A

Recording

To record the written piece, a team of volunteer student musicians were assembled from the BGSU Honors College and from other universities in Ohio. These undergraduates deserve appropriate recognition for their time and talent—without their assistance, the final phase of this project would not be possible.

Gabriella Spatz – Vocals; Hayden B Mesnick – Keyboards; Daniel Yang – Bass;
Elijah Stewart – Guitar; Evan Harnak – Guitar (Case Western Reserve University); Christian Harsa – Photography/Video (the Ohio State University)

The group gathered on April 9, 2022 and recorded the song between 11AM-1PM in the Stanton Recording Studio of the Kuhlin Center at Bowling Green State University. To minimize studio reservation time, the best quality audio was recorded in as few takes as possible. Editing, balancing, and mixing took place outside the reservation space.

The recorded song matched all the predetermined characteristics of a “Hit Song” by the model. Thanks to artistic liberty, additional features were included to emulate the stylistic choices made by popular artists of the 2020s. For example, heavy reverb was applied to the lead guitar riff, and Vocalizing is included in the choruses, and there is ample space for ad lib solos on the outro. After edits and final compilation, the track clocked in around 3:06 long.

Closing and Further Research

Conclusions

Whether the final song is Top 10 Hit material or not may forever remain a mystery. Other variables—celebrity/notoriety status, cultural trends, business connections—place an intangible role in the final placement of a song on the Billboard charts. Still, this was a creative exercise combining art and science. Descriptive statistics reveal that, given an equal amount of Top 10 Hits and non-Hits, some characteristics are more frequently counted, thus suggesting a minor influence on the success of a popular song. To a statistical model, these differences are negligible; various tests for significance remove features that play a small role in shaping a song's chart success.

Project Strengths and Limitations

The project strikes an interesting and novel balance between statistical analysis and music composition. While neither of these are perfect substitutes for one another, they work in tandem to aid the decision-making process for a songwriter. Additionally, what cannot be calculated in the model due to computing capabilities and complexity is still captured, and an individual can use their best judgment to fill in the missing gaps of the prediction.

That said, there are significant setbacks faced in creating the model and collecting data. For example, the time constraint and the limitations of my education withheld me from exploring different modeling techniques that could have provided more accurate predictions (e.g. Time Series Analysis). To fit within the scope of an Honors Project, the techniques used had to reflect the culmination of my studies up to this point. In the future, I would love to explore this data using time series analysis methods like Auto-Regressive Integrated Moving Average (ARIMA).

Additionally, the curse of dimensionality played a heavy hand in the early stages of analysis. 1275 records may seem adequate, but when combined with over 30 features, the model is left confused by noise. More records would be the best solution to the overfitting, but this was not feasible given the manual task of data collection. As such, intense feature selection/ transformation was used; outside of an educational context, this could bring about serious consequences.

Applications and Further Research

A lofty extension of this project would incorporate machine learning to mimic the process of data collection. After piping in the manually recorded songs, perhaps a model could be trained to "listen" to songs 26-100 on the Billboard 100 for each year. Combined with other statistical techniques, this could provide greater accuracy over a wider range of time and require less manpower to complete.

Appendix

Groove Chart

Below is the predetermined guide that can help illustrate how the Groove variable was designed and used. It is important to reiterate that the Groove variable is very subjective to my own listening bias and is not intended to fully categorize a song. Also, a note on data reduction – swing, shuffle, and half-time shuffle songs have unique feels, but they each share the fundamental driving pulse characterized by a weaker leading beat. To suggest they are the same is an oversimplification but yields better predictive results.

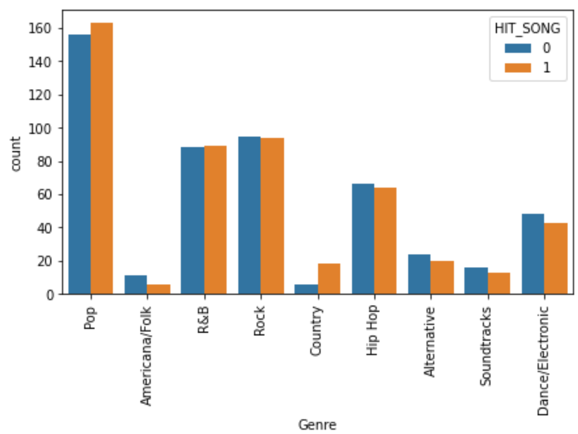
Name	Example Track	Further Description
Straight	Living After Midnight – Judas Priest Take Me Home Tonight – Eddie Money	Emphasis on 1,2,3 and 4, no swing, locked on tempo
Loose	Why Can't This Be Love – Van Halen Thinking Out Loud – Ed Sheeran	Emphasis on 2 and 4, moderate swing feel, but locked on tempo
Disco (Four on the Floor)	Last Dance – Donna Summer I Only Wanna Be with You – Bay City Rollers	Disco beat = four bass hits. Heavy orchestration, syncopated bassline, guitar pedal effects
Shuffle	Pride and Joy – Stevie Ray Vaughn (standard) Rosanna – Toto (half-time) Holiday Road – Lindsey Buckingham (swing)	Includes standard shuffle, half-time shuffle, and swing; accentuated downbeats lag
Syncopated	Go Your Own Way – Fleetwood Mac Pompeii – Bastille	Drum parts mimic or oppose other instruments; voices
16^{th} Pattern	I Keep Forgettin' – Michael McDonald Eternal Light – Free Nationals	16^{th} note drum pattern on hi hat, syncopated bass and keyboards
Dance Pulse	That's All – Genesis Hot N Cold – Katy Perry	Disconnected hi-hat, snare, and drumbeats; highly syncopated vocal melodies; repetitive and choppy basslines
Fusion	Another Rainy Day in New York City – Chicago Shape of You – Ed Sheeran	Bossa nova and tresillo rhythmic patterns; horns and winds; auxiliary percussion like bongos and timbales
Trap	This is America – Childish Gambino Hotline Bling – Drake	Broken "16^{th} Pattern"; high-pitched clap and drum hits, tightly packed in bursts; synthetic effects like orchestral hits
Ballad	Piano Man – Billy Joel I Remember You – Skid Row I Write the Songs – Barry Manilow	Slower than typical songs of the same genre; most likely to include a key change or heavy orchestration
Funk	Slide – Goo Goo Dolls Damn I Wish I Was Your Lover – Sophie B. Hawkins	A combination of "16^{th} pattern" and "Loose"
Offbeat	Take Me Home, Country Roads – John Denver Single Ladies (Put a Ring on It) – Beyonce	Perceived rhythm falls on off beats (the "ands" of 1, 2, 3, and 4)

Sample M-Scale: Lead Guitars

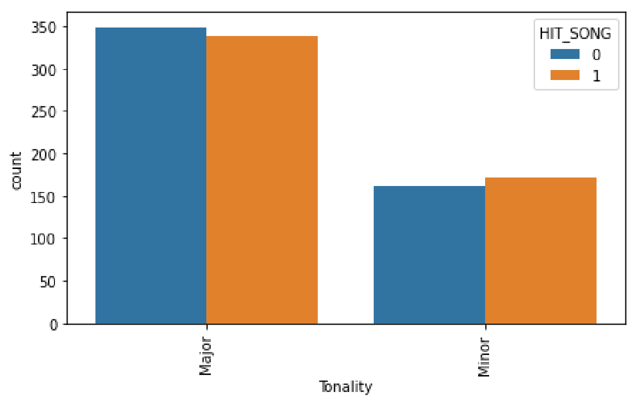
Lead Guitar M-Scale	Example Track	Rationale
0	"Poison" – Bell Biv DeVoe	No Guitar
1	"You Were Meant For Me" – Jewel	Acoustic guitar
2	"Lucid Dreams" – Juice Wrld	Acoustic with minor effects
3	"The Streak" – Ray Stevens	Acoustic electric
4	"Ride Like the Wind" – Christopher Cross	Electric, mild effects
5	"Don't Go Breaking My Heart" – Elton John	Electric, pedal effects
6	"Coming Up" – Paul McCartney	Electric, post-production effects
7	"Jessie's Girl" – Rick Springfield	Electric, minor distortion
8	"The Reflex" – Duran Duran	Electric, distortion
9	"How You Remind Me" – Nickelback	Electric, heavy distortion
10	"I Hope" – Gabby Barrett	Heavy distortion or synthesized

Exhibit 1: Categorical Variable Frequency Analysis.

1A



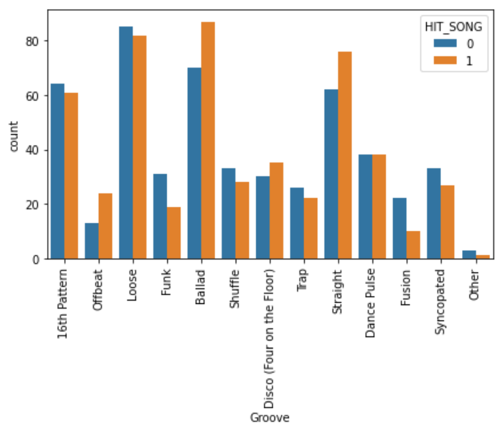
1B



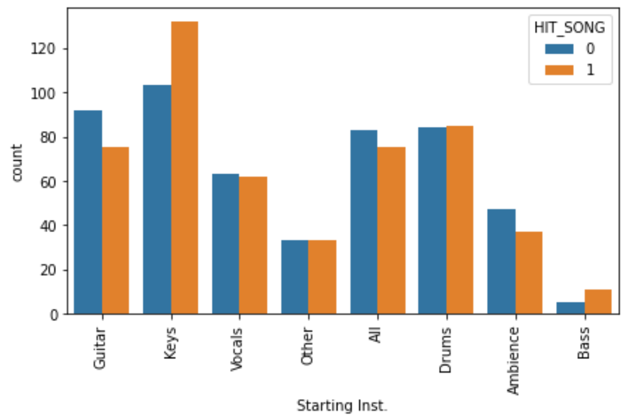
1A: Genre. American popular music is a genre in itself, varying with time and implementing different stylistic choices that reflect largely acceptable tastes of the given period. However, using "pop music" in effect created a miscellaneous category that limited certain genres from making a measurable impact as a different song genre. In other words, perhaps a portion of the "Pop" songs are technically soft rock, country rock, or other strands of music styles with temporary acclaim. If these were all re-categorized as "Rock," would the difference between Hit Rock songs and Not Hit Rock songs change? Regardless, some categories demonstrate a greater difference in frequency: more Country songs are measured as hits than Country songs that aren't in the Top 10. This suggests that, while country music may not be fully integrated into the mainstream, songs that do break through often achieve high recognition.

1B: Tonality. The most common key for popular music is Major—roughly twice as many songs as there are in the Minor counterpart. If a song includes a key change, it may cross from Major to Minor or vice versa, but typically a song will modulate within the same key. For example, Bon Jovi's "Livin' On a Prayer" goes up three half steps from E minor to G minor. By a slim margin depicted in the graph, more songs that had a Minor key were Top 10 Hits (out of total Minor key songs) vs. songs in a Major key.

1C



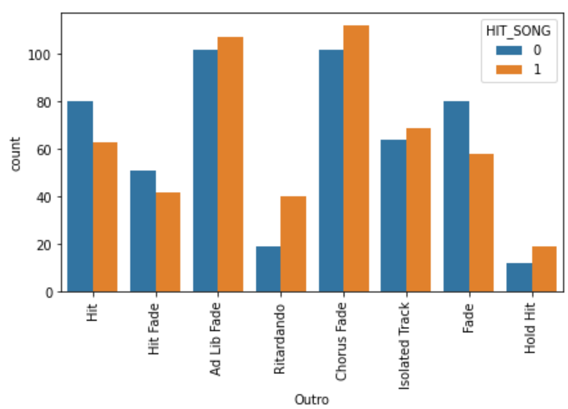
1D



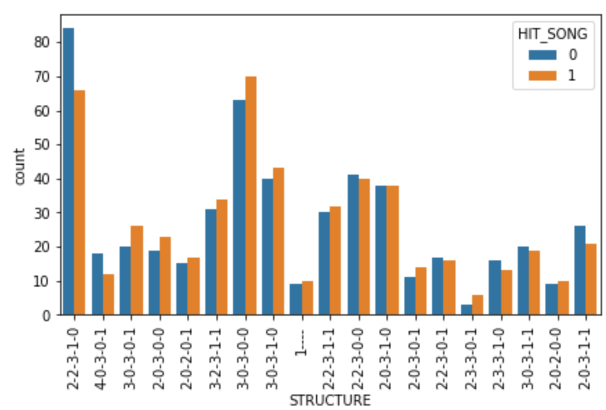
1C: Groove. As described in the Introduction and elaborated in the Groove Chart of the Appendix, this variable is meant to further explain the genre categorization but is subject to oversimplifying the musical feel of a song. Nevertheless, there are clear grooves that are prominent among Top 10 Hits; Ballads, Disco, Straight, and Offbeat rhythms are the most common Grooves observed in this dataset.

1D: Starting Instrument. To my surprise, many of the starting instruments of a song do not reveal any significant trends. A keyboard or bass voice at the beginning of a track may be more frequent in songs measure as Top 10 Hits.

1E



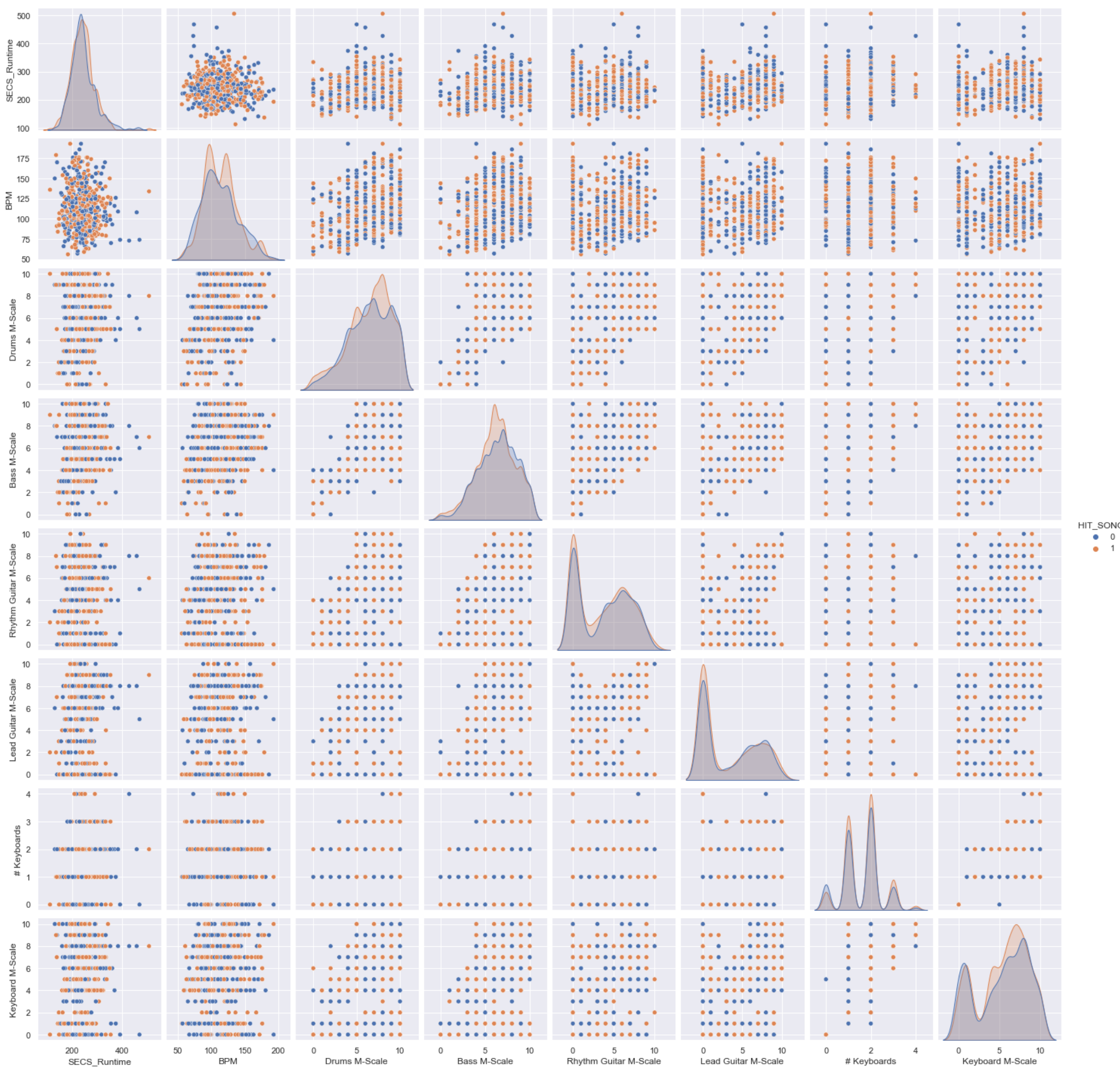
1F



1E: Outro. The variation between each of the different outros and their effect on a song's Hit/No Hit status are easily revealed here. Whether a simple Fade is less common in Top 10 Hits, or a song with a Ritardando is twice as often recorded as a Hit, this categorical variable is significant.

1F: Structure. The creation of the Structure variable is outlined in the Feature Selection portion of this paper. Here, the count of each Structure is illustrated in a "Verse–Pre-Chorus–Chorus–Bridge–Solo" count. For example, the most common Song Structure is 2 Verses, 2 Pre-Choruses, 3 Chorus, 1 Bridge, and no Solos. (2-2-3-1-0).

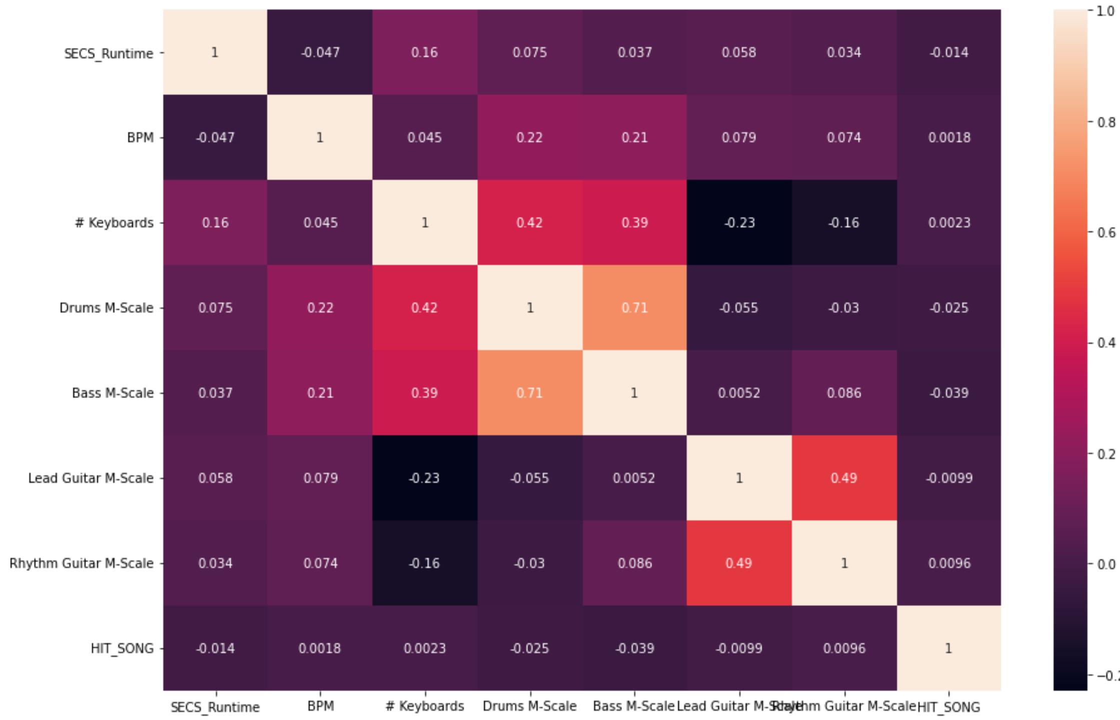
Exhibit 2A: Numeric Pairs Plot.



As a standalone graph, the pairs plot simply depicts the bivariate relationship between each variable and is only mildly effective. The main diagonal shows how songs rated Hit (orange) or No Hit (blue) compare with respect to the exhibited criteria. For example, the bottom right-most graph shows how there are more songs with a Keyboard M-Scale hovering around 4 and 6 that are Hits vs. those with the same characteristics that are not Hits.

To build this, I only entered the quantitative variables the dataset: Runtime (in seconds), BPM, Drums/Bass/Rhythm Guitar/Lead Guitar/Keyboard M-Scale, and Number of Keyboards.

Exhibit 2B: Heat Map (Correlation Matrix).

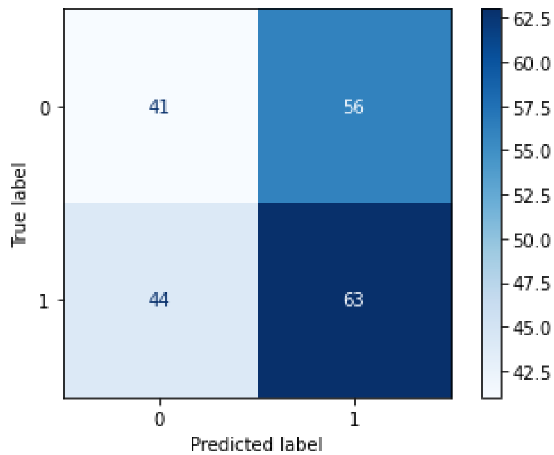


In conjunction with the pairs plot, the correlation matrix provides insight into any potential interaction effect between numerical variables. Relatively speaking, the stronger correlations appear between Bass M-Scale/Drums M-Scale and Rhythm Guitar M-Scale/Lead Guitar M-Scale. Drums and Bass were present in virtually every song of the Top 25 dataset, so these were included in the model.

Exhibit 3: Confusion Matrices.

A confusion matrix is a quick and simple measure of a model’s performance. Using a combination of mathematical equations with the numbers stored in each box, one can easily know a model’s sensitivity, specificity, misclassification rate, and accuracy.

3A: One Decision Tree



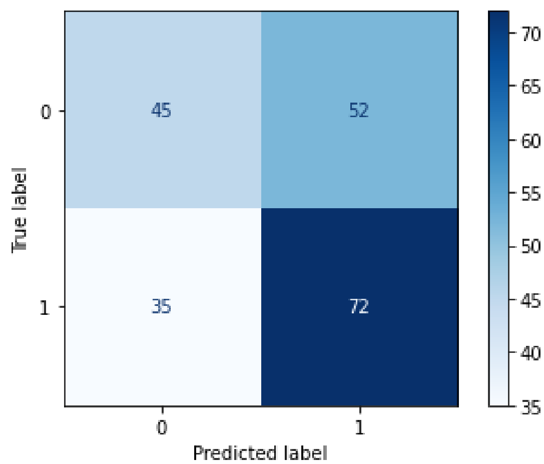
True Negative (top left): the number of Non-Top 10 Hit songs accurately labeled as "No Hit."

False Negative (top right): the number of songs falsely labeled as "Hit" when they were, in fact, not a Top 10 Hit.

False Positive (bottom left): the number of songs falsely labeled as "No Hit" when they were, in fact, a Top 10 Hit.

True Positive (top left): the number of Top 10 Hit songs accurately labeled as "Hit."

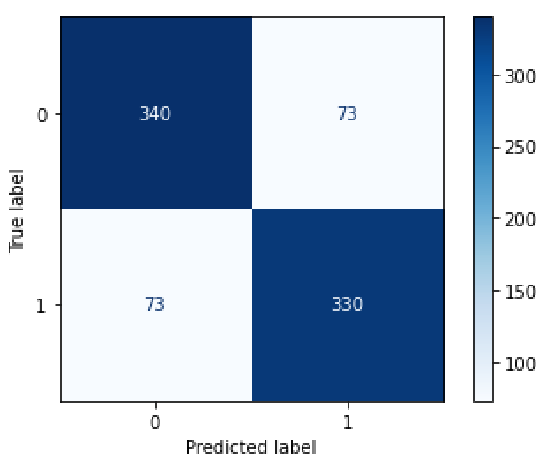
3B: Random Forest Technique – 150 estimators




```
Classification accuracy on test set
with max features = sqrt and max_depth =
8: 0.583
```

The tuned random forest model could accurately label Hits and Non-Hits (True Negatives and True Positives) at a 58.3% accuracy rate, a substantial increase from the 50.9% accuracy of a single tree.

3C: Random Forest with Extreme Gradient Boosting



The confusion matrix to the left illustrated the pred ictive accuracy of the xgboosted model with the **training** dataset. When new data was introduced, the model's accuracy improved to 62.1%, a minor improvement from the random forest classification model.

4: Copyrights and Considerations

The creative works referenced throughout this paper are the property of their original owners and creators. They are appropriately credited and referenced in accordance with the Copyright Act of 1976, which provides allowance for “fair use” such as criticism, comment, scholarship, and research.