

An Empirical Research Approach on Confusion Matrix Using Existing Musical Industry Dataset

Dr R.Naveenkumar, Associate Professor, Department of Computer Science and Engineering Brainware University
Kolkata WestBengal

E-Mail: rnaveenkumarooty@gmail.com

Shubhadeep Pandit, Department of Computer Science and Engineering Brainware University Kolkata WestBengal

Abstract :

This paper analyzes the sales of existing music albums from 1973 to 2019. The data shows that the music industry has experienced many changes over the years, with the most significant change being the introduction of digital platforms. Digital platforms have disrupted the traditional music industry model, leading to a decline in sales of physical albums. The paper also analyzes the sales of each medium through which music was distributed worldwide with different machine learning algorithm classifiers consistently exhibiting inadequate predictive power. the lack of a good prognosis highlights a fundamental failure toward well-infected persons in a class. the presence of non-zero false positive rates indicates negative data are distributed among positive errors. analysis of the data reveals worrying inefficiencies in classifier performance

Keywords: Precession, Recall, Accuracy, Track, Digital industry, Cost efficiency and confusion matrix

Introduction :

The music industry has experienced so many turns regarding the medium through which music albums are sold in the market. At first, the global music industry experienced 8 eight-track medium of sharing music but as time changed different types of I have taken a dataset which contains information about the revenue of the music industry from from 1973 to 2019. The changes done in the music industry were very crucial as its revenue was getting to higher highs and digital platforms were released which disrupted the high-selling albums which cost very high at that time[1].

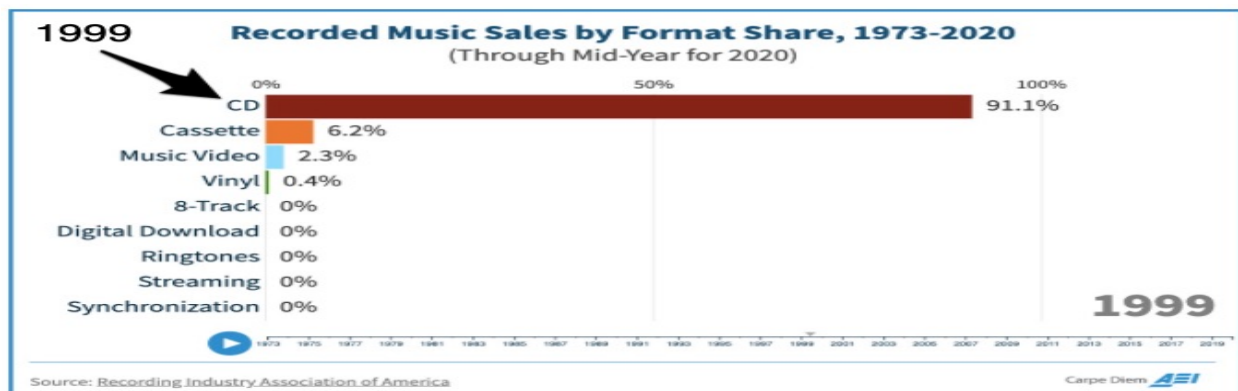
It represents the data analysis of the start of the music industry from cassettes to the continuing era of digital platforms. It also represents the sales of each medium through which the music was distributed throughout the world[2]. This dataset covers the whole 40 years duration of sales in past which is a great period for analyzing the path of sales of music albums. We would be able to derive anything related to the topic regarding any analysis.Music Distribution between 1973 to 2000 taken from existing survey Between 1973 and 2000, the landscape of music distribution experienced a seismic shift, catalyzed by technological innovation and changing consumer behavior[3]. At the beginning of this period, vinyl records reigned supreme as the primary medium for music distribution. Record stores served as the epicenter of music exchange, offering a plethora of LPs and singles[4]. However, as the 1970s progressed, cassette tapes

began to gain ground, offering listeners newfound portability and the ability to create their own mixtapes. But the most significant transformation came with the advent of compact discs (BLACK DISKS) in the 1980s, which marked a quantum leap in audio fidelity and durability. The BLACK DISK's rise was meteoric, eventually displacing vinyl and cassettes as the format of choice in the late 1990s[5].

At the same time, the role of radio and television in promoting music was revolutionized, with the launch of MTV in the 1980s bringing music videos to the forefront of cultural importance. Artists now had a visual medium to complement their audio creations, further increasing sales and influencing popular culture[6]. The retail landscape was also evolving, with traditional record stores facing competition from large retailers such as Walmart and the emergence of online platforms such as Amazon, heralding the coming digital disruption. Meanwhile, the late 1990s saw the rise of digital technology, especially MP3, which made it easier to compress and distribute music files over the Internet[7]. Napster, the pioneering peer-to-peer file sharing service launched in 1999, disrupted the industry by allowing users to share music freely, albeit illegally.

The seismic event sparked legal battles and raised questions about intellectual property rights in the digital age. However, it also laid the groundwork for legitimate digital distribution platforms such as iTunes, which debuted in 2001, and streaming services such as Spotify, which emerged in the late 2000s. In conclusion, the period from 1973 to 2000 witnessed a paradigm shift in distribution of music, from the analogue era dominated by vinyl and cassette tapes to the digital age powered by BLACK DISKS, MP3s and online streaming[8].

This transformative journey not only revolutionized the way music was consumed and distributed, but also reshaped the very fabric of the music industry, setting the stage for further disruption and innovation for decades to come[9].



MUSIC INDUSTRY ALBUM DISTRIBUTION EXPLANATION

Music Distribution between 2000 to 2019.

Between 2000 and 2019, the music distribution landscape underwent a revolution, changing the way music was created, consumed and distributed. The turn of the millennium marked the beginning of the digital age, characterized by the widespread adoption of the Internet and digital technologies. This era saw the proliferation of online music stores and streaming services, fundamentally changing the way people accessed and enjoyed music[10].

With the advent of platforms such as iTunes, which launched in 2001, consumers gained the ability to purchase and download individual tracks or albums digitally, heralding the decline of physical formats such as BLACK DISKS. At the same time, illegal file-sharing platforms that rose to prominence in the late 1990s continued to pose challenges to the music industry's traditional revenue models, fueling legal battles and calls for stronger copyright enforcement. However, amid these challenges, legitimate streaming services have begun to emerge, offering users access to vast libraries of music for a monthly subscription fee. Spotify's launch in 2008 marked a watershed moment for the industry, popularizing the subscription-based streaming model and paving the way for competitors like Apple Music, Amazon Music and Tidal.

These platforms not only provided convenient access to music, but also introduced features such as personalized playlists and algorithm-driven recommendations, enhancing the overall listening experience. As streaming services grew in popularity, they emerged as a dominant force in music distribution, surpassing both physical sales and digital downloads by the end of the decade. This shift towards streaming had profound implications for artists and record labels, who had to adapt to new revenue models based on streaming royalties and concert tours rather than album sales. In addition, social media platforms such as MySpace, Facebook, and later Instagram and Twitter played an increasingly important role in the promotion and discovery of music, allowing artists to connect directly with fans and cultivate a devoted following.

The rise of user-generated content platforms such as YouTube has also provided a new avenue for artists to showcase their work and reach a global audience. Additionally, advances in digital recording technology have democratized the music production process, allowing artists to create professional-quality recordings from the comfort of their homes or small studios. This democratization of music production, along with the ease of distribution offered by online platforms, has led to the proliferation of independent artists and the diversification of musical styles and genres. Additionally, the globalization of music distribution facilitated by the Internet and streaming services has fostered cross-cultural collaboration and the exchange of musical influences on a global scale. However, amid the digital revolution, concerns have arisen about the impact of streaming on artists' earnings and the sustainability of the music industry ecosystem.

Critics argued that streaming platforms paid artists disproportionately low royalties, particularly for independent and niche musicians, leading to calls for reform and greater transparency in royalty calculations. However, the rise of independent distribution platforms like Bandcamp and Patreon have provided artists with alternative sources of income, allowing them to maintain more control over their music and directly monetize their fan base. Looking ahead, the future

of music distribution is likely to be shaped by further advances in technology, changes in consumer behavior and continued efforts to address the challenges and opportunities presented by the digital environment. As the industry continues to evolve, one thing remains certain: music will continue to be a powerful force that crosses borders and connects people, regardless of distribution medium or platform.

DATASET DESCRIPTION

Set of attributes: 5

Sum of weights: 3008

Instances: 3008

Relation : Music data

Selecting Format attribute-

Type:Nominal

Distinct : 24

This attribute helps in recognizing the name of different ways in which audio was being distributed between the year 1973 to 2019.

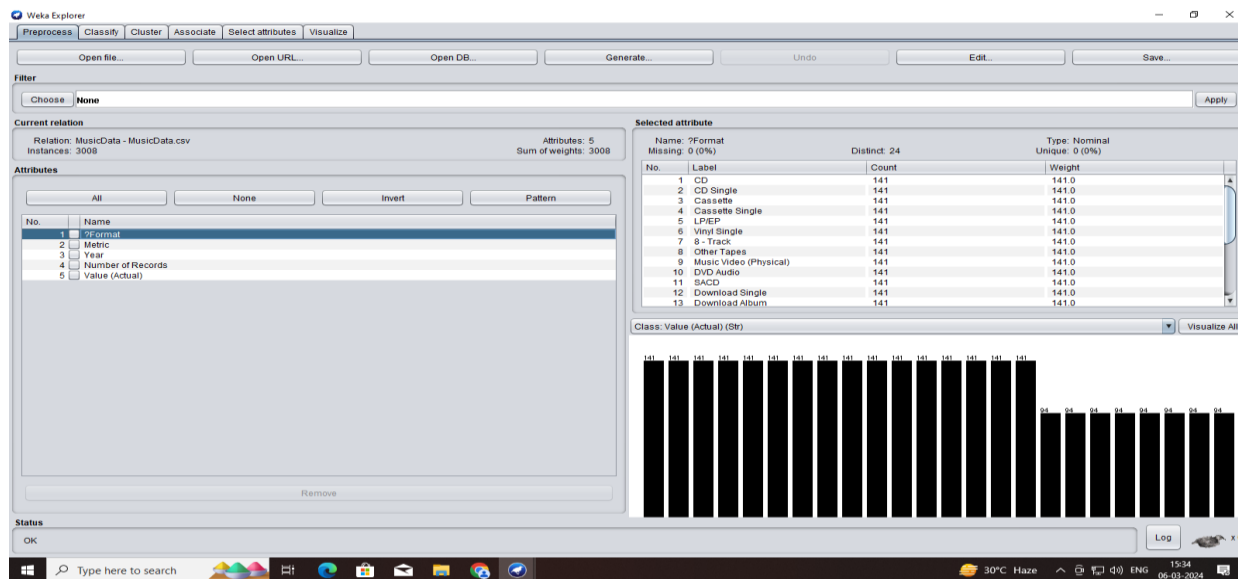


FIGURE 1: INTERFACE PICTURE

No.	Label	Count	Weight
1	CD	141	141
2	CD Single	141	141
3	Cassette	141	141
4	Cassette Single	141	141
5	LP/EP	141	141
6	Vinyl Single	141	141
7	8 - Track	141	141
8	Other Tapes	141	141
9	Music Video (Physical)	141	141
10	DVD Audio	141	141
11	SACD	141	141
12	Download Single	141	141
13	Download Album	141	141
14	Kiosk	141	141
15	Download Music Video	141	141
16	Ringtones & Ringback	141	141
17	Paid Subscriptions	94	94
18	Limited Tier Paid Subs	94	94
19	On-Demand Streaming	94	94
20	Other Ad-Supported S	94	94
21	Other Digital	94	94
22	Paid Subscription	94	94
23	SoundExchange Distri	94	94
24	Synchronization	94	94

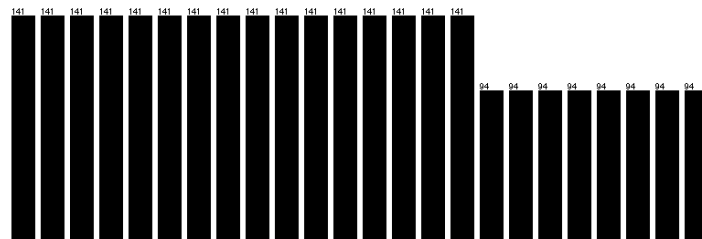


FIGURE 2 : FORMAT ATTRIBUTE PICTURE

Selecting Metric attribute-

Type:Nominal

Distinct : 3

It provides the total measure of units of music albums sold during the time period of 1973 to 2019.

No.	Label	Count	Weight
1	Units	846	846.0
2	Value	1081	1081.0
3	Value (Adjusted)	1081	1081.0



FIGURE 3 : METRIC ATTRIBUTE PICTURE

Selecting Year attribute-

Type:Numeric

Distinct : 47

It provides the minimum and maximum number of the albums sold in the year and the mean and the standard deviation provided by the dataset.

Statistic	Value
Minimum	1973
Maximum	2019
Mean	1996
StdDev	13.567

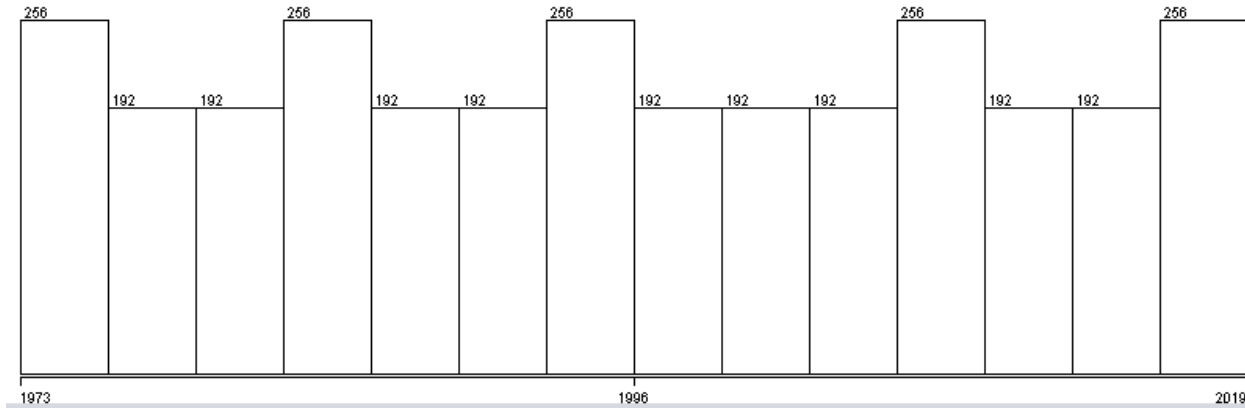


FIGURE 4: YEAR ATTRIBUTE PICTURE

Selecting Number of Records attribute-

Type:Numeric

Distinct : 1

Statistic	Value
Minimum	1
Maximum	1
Mean	1
StdDev	0

Selecting Value(Actual)attribute-

Type:String

Distinct : 1138

No Picture For The Attribute Value Is Found .

No Analysis Was Found For The Classifiers: Part And One R

Therefore The Analysis Of The Confusion Matrix Is Taken Into Processing And Analysis.

Confusion Matrix For Zero R

Scheme: weka.classifiers.rules.ZeroR

Relation: MusiBlack Diskata - MusiBlack Diskata.csv

Instances: 3008

Attributes: 5

?Format

Metric

Year

Number of Records

Value (Actual)

Test mode: 5-fold cross-validation

==== Classifier model (full training set) ====

ZeroR predicts class value: BLACK DISK

Time taken to build model: 0 seconds

==== Stratified cross-validation ====

==== Summary ====

Correctly Classified Instances	0	0 %
Incorrectly Classified Instances	24	100 %
Kappa statistic		-0.0435
Mean absolute error	0.0814	
Root mean squared error	0.2039	
Relative absolute error	100 %	
Root relative squared error	100 %	
Total Number of Instances	24	
Ignored Class Unknown Instances	2984	

==== Detailed Accuracy By Class ====

TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0.000	0.826	0.000	0.000	0.000	-0.406	0.100	0.000	BLACK DISK
0.000	0.217	0.000	0.000	0.000	-0.107	0.100	0.000	BLACK DISK Single
0.000	0.000	0.000	0.000	0.000	0.000	0.100	0.000	Cassette
0.000	0.000	0.000	0.000	0.000	0.000	0.100	0.000	Cassette Single

0.000	0.000	0.000	0.000	0.000	0.000	0.100	0.000	LP/EP
0.000	0.000	0.000	0.000	0.000	0.000	0.100	0.000	Vinyl Single
0.000	0.000	0.000	0.000	0.000	0.000	0.100	0.000	8 - Track
0.000	0.000	0.000	0.000	0.000	0.000	0.100	0.000	Other Tapes
0.000	0.000	0.000	0.000	0.000	0.000	0.100	0.000	Music Video (Physical)
0.000	0.000	0.000	0.000	0.000	0.000	0.100	0.000	DVD Audio
0.000	0.000	0.000	0.000	0.000	0.000	0.100	0.000	SABBLACK DISK
0.000	0.000	0.000	0.000	0.000	0.000	0.100	0.000	Download Single
0.000	0.000	0.000	0.000	0.000	0.000	0.100	0.000	Download Album
0.000	0.000	0.000	0.000	0.000	0.000	0.100	0.000	Kiosk
0.000	0.000	0.000	0.000	0.000	0.000	0.100	0.000	Download Music Video
0.000	0.000	0.000	0.000	0.000	0.000	0.100	0.000	Ringtones & Ringbacks
0.000	0.000	0.000	0.000	0.000	0.000	0.100	0.000	Paid Subscriptions
0.000	0.000	0.000	0.000	0.000	0.000	0.100	0.000	Limited Tier Paid Subscription
0.000	0.000	0.000	0.000	0.000	0.000	0.100	0.000	On-Demand Streaming
0.000	0.000	0.000	0.000	0.000	0.000	0.100	0.000	Other Ad-Supported Streaming
0.000	0.000	0.000	0.000	0.000	0.000	0.100	0.000	Other Digital
0.000	0.000	0.000	0.000	0.000	0.000	0.100	0.000	Paid Subscription
0.000	0.000	0.000	0.000	0.000	0.000	0.100	0.000	SoundExchange Distributions
0.000	0.000	0.000	0.000	0.000	0.000	0.100	0.000	Synchronization

Weighted Avg. 0.000 0.043 0.000 0.000 0.000 -0.021 0.100 0.000

=== Confusion Matrix ===

a b c d e f g h i j k l m n o p q r s t u v w x <-- classified as
0 1 0 | a = BLACK DISK
1 0 | b = BLACK DISK Single
1 0 | c = Cassette
1 0 | d = Cassette Single
0 1 0 | e = LP/EP
1 0 | f = Vinyl Single
1 0 | g = 8 - Track
1 0 | h = Other Tapes

01000000000000000000000000 | i = Music Video (Physical)
10000000000000000000000000 | j = DVD Audio
10000000000000000000000000 | k = SABLACK DISK
10000000000000000000000000 | l = Download Single
10000000000000000000000000 | m = Download Album
10000000000000000000000000 | n = Kiosk
10000000000000000000000000 | o = Download Music Video
10000000000000000000000000 | p = Ringtones & Ringbacks
10000000000000000000000000 | q = Paid Subscriptions
10000000000000000000000000 | r = Limited Tier Paid Subscription
10000000000000000000000000 | s = On-Demand Streaming (Ad-Supported)
01000000000000000000000000 | t = Other Ad-Supported Streaming
10000000000000000000000000 | u = Other Digital
10000000000000000000000000 | v = Paid Subscription
10000000000000000000000000 | w = SoundExchange Distributions
01000000000000000000000000 | x = Synchronization

Conculsion For ZERO R classifier

Analyzing the metrics for the given confusion matrix for the "zero r" classifier:

1. TP Rate (True Positive Rate):

All classes have a TP Rate of 0, indicating that there are no true positive predictions for any class.

2. FP Rate (False Positive Rate):

- The number of FPs varied between studies, with values ranging from 0.000 to 0.826.

- FP Rate represents the proportion of negative cases that are incorrectly classified as positive.

3. Precision:

- Accuracy values for all classes are 0, indicating no true positive prediction. Precision refers to the accuracy of a good forecast.

4. Recall:

- Recall values for all classes are 0, indicating poor true prediction. Recall measures the ability of a classifier to recognize all positive samples.

5. F-Measure:

- F-measure values for all classes are 0, because precision and recall are 0. F-measure is the harmonic mean of precision and recall and is used to measure the accuracy of the classifier.

6. MCC (Matthews Correlation Coefficient):

-MCC values for all classes are negative or 0, indicating no relationship between predicted and actual distributions. A value of 0 indicates random predictions, and a negative value indicates anticorrelation.

7. ROC Area:

- The ROC area for all classes was 0.100, indicating that the model performed poorly in terms of classification ability. ROC Area Represents the area under the Receiver Operating Characteristic curve and is used to evaluate the performance of the binary classifiers.

8. PRC Area (Precision-Recall Curve Area):

- The PRC area of all subjects is 0.000, indicating poor performance in terms of accuracy and recall. PRC Area represents the area under the Precision-Recall curve and is an additional measure of classifier performance.

9. Weighted Avg:

- The weighted values of TP Rate, Precision, Recall, F-Measure, and MCC are all 0, indicating that they are all underperforming across all groups.

- a weighed in. The FP Rate was 0.043, which shows some false positives across all classes, and contributes to the overall FP Rate.

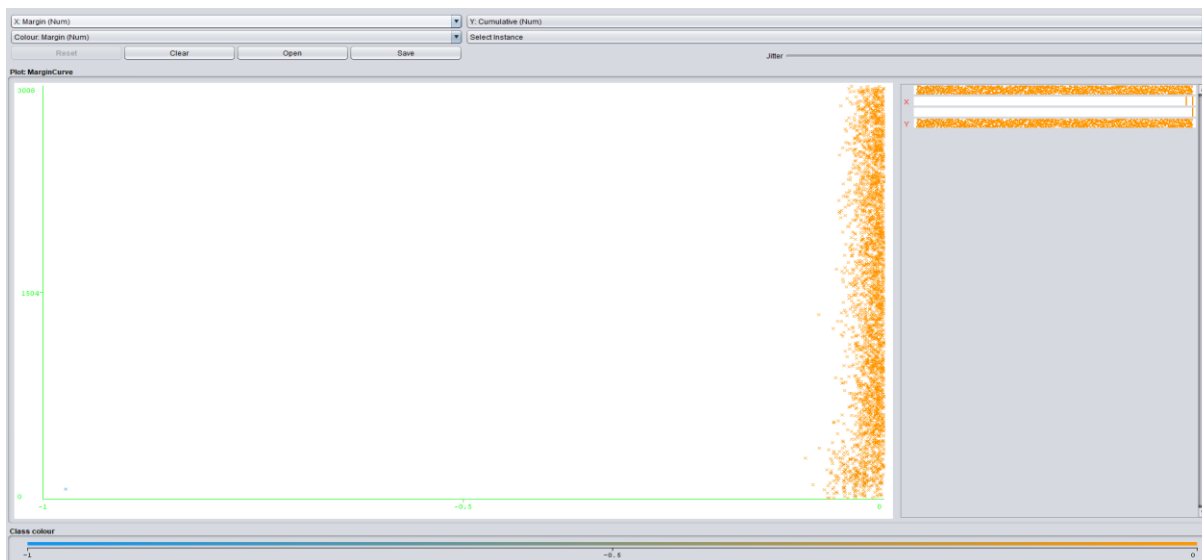


FIGURE 1 FOR CURVE LINE

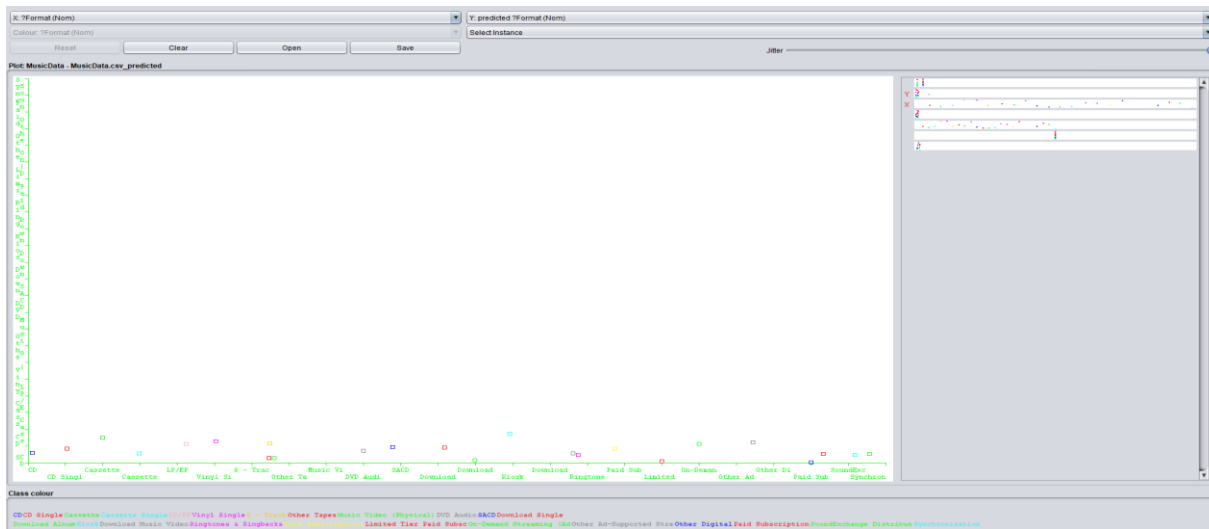


FIGURE 2 FOR CURVE LINE

Conclusion:

In summary, the analysis of the data reveals worrying inefficiencies in classifier performance. Among all tested metrics one class 'a - Black Disk' is performing higher in comparison with another class matrix with the example of a different confusion matrix it is clear that the class performing better than other classes where TP Rate, FP Rate, Precision, Recall, F-Measure, MCC, ROC Area, and PRC Area, the classifier consistently exhibits inadequate predictive power. Notably, accuracy, the lack of a good prognosis highlights a fundamental failure toward well-infected persons in a class. Furthermore, the presence of non-zero false positive rates indicates that negative data are distributed among positive errors, further increasing sampling errors. Similarly, low or negative values in metrics such as accuracy, recall, F-measures, MCC, etc. confirm that classification cannot meaningfully distinguish between groups, reflecting minimum values, indicating a lack of discriminative power. In particular, the overall poor performance of the classifier across various evaluation metrics indicates that its ability to classify samples correctly is significantly reduced, warranting further investigation and it can reconsider its underlying process or characteristics.

Reference

1. J. M. Abbazio, A. Boddie, and E. Ogihara, "Music Libraries and an Expanding Repertory: Suggested Strategies for Building Diverse Music Library Collections," *Notes* 78, no. 3 (2022):1–28. doi:10.1353/not.2022.0005
2. W. E. Anderson, P. Bristah, E. A. Davis, J. Gottlieb, K. D. Underwood, and Music Library Association, *A Basic Music Library: Essential Scores and Sound Recordings*, Third. (Chicago: American Library Association, 1997).
3. A. Asher, T. Shelton, J. Heckathorn, and H. Huet, "Patron-Driven Acquisition of Print Monographs and Music Scores: A Pilot Program Case Study," *Music Reference Services Quarterly* 21, no. 4 (2018):185–208. doi:10.1080/10588167.2018.1539605
4. R. V. Brittin and J. M. Standley, "Researchers in Music Education/Therapy: Analysis of Publications, Citations, and Retrievability of Work," *Journal of Research in Music Education* 45, no. 1 (1997):145–60. doi:10.2307/3345471
5. M. J. Citron, "Women and the Western Art Canon: Where are We Now?," *Notes* 64, no. 2 (2007):209–15. doi:10.1353/not.2007.0167
6. J. C. Clark, "Format Preferences of Performing Arts Students," *The Journal of Academic Librarianship* 39, no. 3 (2013):297–307. doi:10.1016/j.acalib.2013.02.005
7. E. J. Cox, R. Cross, and J. Ballestro, "Music Score Approval Plans in Research Libraries: A Survey of Librarian Satisfaction with and Without Approval Plans," *Library Collections, Acquisitions, & Technical Services* 33, no. 2–3 (2009):80–94. doi:10.1080/14649055.2009.10766236
8. P. Daub, "The RLG Music Conspectus: Its History and Applications, in *Collection Assessment in Music Libraries*, ed. J. Gottlieb, (Canton, MA: Music Library Association, 1994), 7–24.
9. C. J. Duncan and G. M. O’Gara, "Building Holistic and Agile Collection Development and Assessment." *Performance Measurement and Metrics* 16, no. 1 (2015): 62–85. doi:10.1108/PMM-12-2014-0041
10. K. C. Fountain and L. Frederiksen, "Just Passing Through: Patron-Initiated Collection Development in Northwest Academic Libraries." *Collection Management* 35, no. 3–4 (2010): 185–95. doi:10.1080/01462679.2010.486745
11. V. Graziano, "LGBTQ Collection Assessment: Library Ownership of Resources Cited by Master’s Students," *College & Research Libraries* 77, no. 1 (2016):114–27. doi:10.5860/crl.77.1.114
12. R. Griscom, "Periodical Use in a University Music Library: A Citation Study of Theses and Dissertations Submitted to the Indiana University School of Music from 1975-1980," *The Serials Librarian* 7, no. 3 (1983):35–52. doi:10.1300/J123v07n03_05
13. P. L. K. Gross and E. M. Gross, "College Libraries and Chemical Education," *Science* 66, no. 1713 (1927):385–89. doi:10.1126/science.66.1713.385
14. C. B. Hancock and H. E. Price, "Sources Cited in the Journal of Research in Music Education: 1953 to 2015," *Journal of Research in Music Education* 68, no. 2 (2020):216–40. doi:10.1177/0022429420920579

15. K. Heath and D. Novak, "Implementing a Music Scores Approval Plan at Carnegie Mellon University: A Retrospective Analysis 2010-2014." *Music Reference Services Quarterly* 20, no. 1 (2017): 8–41. doi:10.1080/10588167.2017.1274606
16. E. Henry, R. Longstaff, and D. Van Kampen, "Collection Analysis Outcomes in an Academic Library," *Collection Building* 27, no. 3 (2008):113–17. doi:10.1108/01604950810886022
17. L. Hooper, "The Art of Crafting Music Score Approval Plans: An Ongoing Process," *Collection Management* 41, no. 4 (2016):228–35. doi:10.1080/01462679.2016.1227007
18. J. E. Knievel and C. Kellsey, "Citation Analysis for Collection Development: A Comparative Study of Eight Humanities Fields," *The Library Quarterly* 75, no. 2 (2005):142–68. doi:10.1086/431331
19. L. Kuypers-Rushing, "Identifying Uniform Core Journal Titles for Music Libraries: A Dissertation Citation Study," *College & Research Libraries* 60, no. 2 (1999):153–63. doi:10.5860/crl.60.2.153
20. K. Lai, "A Revival of the Music Conspectus: A Multi-Dimensional Assessment for the Score Collection," *Notes (Music Library Association)* 66, no. 3 (2010):503–18. doi:10.1353/not.0.0310
21. J. Mayer, "Serving the Needs of Performing Arts Students: A Case Study," *Portal: Libraries and the Academy* 15, no. 3 (2015):409–31. doi:10.1353/pla.2015.0036
22. C. E. McGuire, "Of Programs and Prima Donnas: Investigating British Music with the Musical Festivals Database," *Notes* 73, no. 3 (2017):432–72. doi:10.1353/not.2017.0001
23. B. Nettl, *Heartland Excursions: Ethnomusicological Reflections on Schools of Music* (Urbana: University of Illinois Press, 1995).
24. T. van Oort and J. Noordegraaf, "Structured Data for Performing Arts History: An Introduction to a Special Issue of Data Papers: Arts and Media," *Research Data Journal for the Humanities and Social Sciences* 5, no. 2 (2020):1–12. doi:10.1163/24523666-bja10008
25. M. Ruppel, "Tying Collection Development's Loose Ends with Interlibrary Loan." *Collection Building* 25, no. 3 (2006): 72–77. doi:10.1108/01604950610677530
26. C. Shirkey, "Taking the Guesswork Out of Collection Development: Using Syllabi for a User-Centered Collection Development Method," *Collection Management* 36, no. 3 (2011):154–64. doi:10.1080/01462679.2011.580046
27. W. Weber, "The History of Musical Canon," in *Rethinking Music*, eds. Nicholas Cook and Mark Everist, (Oxford: Oxford University Press, 1999), 336–55.
28. D. Zager, "Essential Partners in Collection Development: Vendors and Music Librarians," *Notes* 63, no. 3 (2007):565–75. doi:10.1353/not.2007.0049