

ANALISIS KLASIFIKASI *ARTIST MUSIC* MENGGUNAKAN MODEL REGRESI LOGISTIK BINER DAN ANALISIS DISKRIMINAN

Andrea Tri Rian Dani¹, Vita Ratnasari², Ludia Ni'matuzzahroh², Igar Calveria Aviantholib², Raditya Novidianto³, Narita Yuri Adrianingsih⁴

¹ Program Studi Statistika, Jurusan Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Mulawarman

² Departemen Statistika, Fakultas Sains dan Analitika Data, Institut Teknologi Sepuluh Nopember,

³ Badan Pusat Statistik Kabupaten Poso

⁴ Jurusan Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Tribuana Kalabahi.

e-mail: andreatriandani98@gmail.com

Abstrak

Penciri dari sebuah lagu merupakan sebuah aspek yang penting yang harus dijaga keasliannya oleh seorang penyanyi. Menggunakan fitur *Spotify* API, kita bisa mengekstrak karakteristik atau unsur dari suatu lagu yang dinyanyikan oleh seorang penyanyi. Terdapat delapan (8) unsur yang bisa kita peroleh dari hasil ekstraksi sebuah lagu, yaitu: *Danceability*, *Energy*, *Loudness*, *Speechiness*, *Acousticness*, *Liveness*, *Valence*, dan *Tempo*. Berdasarkan hasil ekstraksi tersebut, kita bisa memberikan label *artist music* menggunakan metode analisis klasifikasi. Pada penelitian ini, labelnya adalah *artist music*, yaitu Ariana Grande dan Taylor Swift. Penelitian ini bertujuan untuk mendapatkan klasifikasi label *artist music* dengan menggunakan metode regresi logistik biner dan analisis diskriminan. Variabel respon yang digunakan pada penelitian ini yaitu *Artist Music* (Y) yang dikategorikan menjadi dua kategori yaitu Ariana Grande (Y=0) dan Taylor Swift (Y=1). Data akan dibagi menjadi data *training* dan data *testing* dengan proporsi data 90:10 dan 80:20. Berdasarkan hasil analisis, model regresi biner yang dibangun, dengan proporsi data *training testing* yaitu 90:10 memiliki ketepatan klasifikasi untuk data *testing* sebesar 90,00%.

Kata Kunci: Analisis Diskriminan, Klasifikasi, Regresi Logistik Biner

Abstract

Characteristics of a song are an important aspect that must be kept authentic by a singer. Using the *Spotify* API feature, we can extract the characteristics or elements of a song sung by a singer. There are eight (8) elements that we can get from the extraction of a song, namely: *Danceability*, *Energy*, *Loudness*, *Speechiness*, *Acousticness*, *Liveness*, *Valence*, and *Tempo*. Based on the extraction results, we can label the music artist using the classification analysis method. In this study, the labels are music artists, namely Ariana Grande and Taylor Swift. This study aims to obtain the classification of music artist labels using binary logistic regression methods and discriminant analysis. The response variable used in this study is *Artist Music* (Y) which is categorized into two categories, namely Ariana Grande (Y=0) and Taylor Swift (Y=1). The data will be divided into training and testing data with the proportion of data 90:10 and 80:20. Based on the results of the analysis, the binary regression model that was built, with the proportion of training testing data that is 90:10 has a classification accuracy for data testing of 90.00%.

Keywords: Discriminant Analysis, Classification, Binary Logistic Regression

1. PENDAHULUAN

Karakteristik dari suatu lagu merupakan aspek penting yang harus dijaga orisinalitasnya oleh seorang penyanyi. Masalah yang sering kita jumpai adalah ketika kita mendengar sebuah lagu, terkadang kita tidak bisa menebak pasti siapa penyanyinya. Pendengar dapat dengan mudah mengatakan siapa penyanyinya namun kebanyakan yang terjadi adalah salah. Kita tahu bahwa seorang penyanyi tentunya selalu memberikan variasi terhadap musiknya namun tetap dengan mempertahankan identitas uniknya. Hal inilah yang menjadi permasalahan menarik, bagaimana kita bisa memberikan label *artist music* berdasarkan identitas uniknya tersebut (karakteristik) (Al-Beitawi et al., 2020).

Menggunakan fitur *Spotify* API kita bisa mengekstrak karakteristik dari suatu lagu yang dinyanyikan oleh seorang penyanyi. Karakteristik yang bisa kita peroleh diantaranya: *Acousticness, Danceability, Energy, Loudness, Liveness, Speechiness, Tempo*, dan *Valence* (Santana et al., 2020). Berdasarkan hasil ekstrak yang diperoleh, kita bisa membangun suatu model klasifikasi untuk memberikan label *artist music*. Selain itu kita bisa melakukan eksplorasi data secara umum untuk melihat ciri-ciri seorang penyanyi membawakan lagunya.

Metode klasifikasi merupakan salah satu metode statistika yang digunakan untuk membangun suatu model sebagai *prototype* yang kemudian disimpan dan digunakan untuk melakukan pengenalan atau klasifikasi pada suatu obyek baru ke dalam suatu kelompok yang telah memiliki label (Fitria & Azis, 2018; Mattjik & Sumertajaya, 2011; Rajagukguk et al., 2015). Metode klasifikasi termasuk dalam *Supervised Learning*, yang artinya label keluaran telah dikelompokkan, sehingga fungsi pemisah antara label yang satu dengan label lainnya dapat dicari dengan mempelajari pola dari data kelas-kelas/atribut yang telah ada untuk mengklasifikasikan data yang baru (Giovani et al., 2020; Suwardika, 2017). Pada kasus ini, labelnya adalah *artist music* yaitu Ariana Grande dan Taylor Swift. Klasifikasi disebut juga sebagai metode pengalokasian. Tujuan klasifikasi adalah untuk memilih atau memasukkan pengamatan (obyek baru) ke dalam kelompok yang telah mempunyai label kelompok (Hosmer & Lemeshow, 2000; Ketut et al., 2017). Beberapa metode klasifikasi yang terkenal dan sering digunakan oleh para peneliti dan akademisi diantaranya adalah model regresi logistik dan analisis diskriminan. Kedua metode ini memiliki masing-masing keunggulan dan kekurangan dalam praktek kasus riilnya.

Regresi logistik merupakan metode yang menghubungkan antara variabel respon yang bersifat kategorik dengan variabel prediktor yang dapat bersifat numerik dan kategorik (Hapsary, Subiyanto, & Firdaus, 2021). Regresi logistik merupakan salah satu bentuk spesial dari analisis regresi yang dapat digunakan untuk mengklasifikasikan. Regresi logistik akan menghasilkan nilai probabilitas untuk masing-masing label yang dapat digunakan sebagai dasar untuk klasifikasi. Beberapa penelitian terbaru yang menggunakan model Regresi logistik diantaranya: (Novianti & Purnami, 2012; Octisari et al., 2021; Rajagukguk et al., 2015; Suwardika, 2017). Pada penelitian ini, model regresi logistik yang digunakan adalah regresi logistik biner, dikarenakan variabel respon hanya memiliki 2 kategori.

Analisis diskriminan merupakan salah satu metode klasifikasi yang hampir mirip dengan regresi logistik (Mattjik & Sumertajaya, 2011). Perbedaannya adalah pada analisis diskriminan variabel prediktor hanya bisa bersifat numerik. Analisis diskriminan dapat digunakan untuk mengklasifikasikan suatu obyek kedalam kategori atau label kelas tertentu (Nabila et al., 2021; Nugroho, 2008). Beberapa penelitian terbaru yang menggunakan model analisis diskriminan diantaranya: (Annas & Irwan, 2015; Nabila et al., 2021).

Penelitian terkait yang dipublikasikan oleh (Jung, 2018) yang berjudul *The Data Science of Kpop: Understanding BTS through data dan Artificial Intelligence (AI)*, mencoba mengklasifikasikan *artist music* menggunakan pendekatan *Machine Learning*. Berdasarkan

hasil analisis, diperoleh hasil bahwa setiap *artist* memberikan ciri terhadap setiap lagunya dari hasil ekstraksi *Spotify* API, sedemikian sehingga model klasifikasi yang dibangun mampu mengklasifikasikan label *artist music* yang berbeda. Berangkat dari hal itu, kami mencoba menggunakan metode klasifikasi yang *classic*, apakah akan memberikan hasil klasifikasi yang baik dalam mengklasifikasikan label *artist music*, yaitu regresi logistik dan analisis diskriminan.

Berdasarkan uraian diatas, pada penelitian ini akan dibandingkan kedua metode tersebut dalam mengklasifikasikan *artist music* dengan label kelas Ariana Grande dan Taylor Swift. Data akan dibagi menjadi data *training* dan data *testing* dengan proporsi pembagian yang berbeda, yaitu 90:10 dan 80:20. Berdasarkan ketepatan klasifikasi nantinya akan diinterpretasikan dan dijelaskan model klasifikasi mana dengan proporsi berapa yang baik dalam mengklasifikasikan *artist music* berdasarkan unsur-unsur yang ada pada lagu-lagunya. Ariana Grande dan Taylor Swift dipilih karena keduanya merupakan salah satu penyanyi yang sangat terkenal dan lagu-lagunya sangat *familiar* di telinga pendengar. Ariana Grande adalah seorang penyanyi, penulis lagu dan aktris asal Amerika Serikat, sama halnya dengan Taylor Swift. Kedua penyanyi ini sering mendapatkan penghargaan dan masuk nominasi berkat kesuksesan lagu-lagu mereka.

2. METODE PENELITIAN

Pada bagian ini, akan dijelaskan terkait metode analisis yang digunakan, data penelitian, variabel penelitian, serta tahapan analisis data yang dilakukan:

2.1 Data dan Variabel Penelitian

Pada penelitian ini menggunakan *dataset* hasil ekstraksi dari *Spotify* API yang dapat didownload pada *link* berikut:

<https://www.kaggle.com/qiaozhenwu/kpopwesternsongs>.

Tabel 1 *Dataset* Olah

Danceability	Energy	Loudness	Speechiness	Acousticness	Liveness	Valence	Tempo	Artist	Code
0.737	0.802	-4.771	0.0878	0.468	0.0931	0.682	144.015	arianagrande	1
0.602	0.658	-5.934	0.0558	0.0233	0.237	0.268	145.031	arianagrande	1
0.824	0.624	-9.728	0.0827	0.781	0.1	0.248	102.012	taylor	0
0.532	0.623	-9.208	0.0331	0.538	0.0925	0.403	89.937	taylor	0
0.553	0.413	-10.037	0.0334	0.687	0.0594	0.379	110.107	taylor	0
0.568	0.656	-5.413	0.0433	0.0211	0.213	0.364	100.049	arianagrande	1
0.648	0.621	-5.831	0.179	0.207	0.918	0.371	76.002	arianagrande	1
0.358	0.639	-4.938	0.0382	0.257	0.964	0.314	103.289	arianagrande	1
0.688	0.573	-5.03	0.208	0.0686	0.0749	0.553	110.034	arianagrande	1
0.877	0.547	-5.705	0.18	0.0122	0.028	0.687	99.034	arianagrande	1
0.515	0.476	-9.843	0.0328	0.314	0.171	0.412	135.886	taylor	0
...
0.767	0.92	-3.445	0.0997	0.0673	0.375	0.598	104.9	arianagrande	1
0.662	0.747	-6.926	0.0736	0.028	0.138	0.487	150.088	taylor	0

Variabel yang digunakan pada penelitian ini terdiri dari variabel respon dan prediktor. Pada variabel ini digunakan satu variabel respon (Y) yang berskala kategorik dan delapan variabel prediktor. Variabel-variabel yang digunakan disajikan pada Tabel 2.

Tabel 2 Variabel Penelitian

No.	Variabel	Notasi	Jenis Data	Skala Data	
1	Respon	$Y=Artist\ Music$	Kategorik	Nominal	
		0: Ariana Grande 1: Taylor Swift			
		$X_1=Danceability$			Rasio
		$X_2=Energy$			Rasio
2	Prediktor	$X_3=Loudness$	Numerik	Interval	
		$X_4=Speechiness$		Rasio	
		$X_5=Acousticness$		Rasio	
		$X_6=Liveness$		Rasio	
		$X_7=Valence$		Rasio	
		$X_8=Tempo$		Rasio	

Definisi operasional dari variabel prediktor yang ditunjukkan pada tabel 2 dapat dijelaskan sebagai berikut:

1. *Danceability*

Variabel *Danceability* menjelaskan seberapa besar kesesuaian dalam sebuah lagu untuk bisa membuat menari berdasarkan kombinasi tempo, ritme, *beat*, dan lain sebagainya. Nilai dari variabel ini berkisar diantara 0,00 sampai dengan 1,00, dimana semakin mendekati 1,00, maka lagu tersebut semakin cocok didengarkan sambil menari/berjoget.

2. *Energy*

Variabel *Energy* menjelaskan seberapa besar sebuah lagu dapat meningkatkan emosi dari hati pendengarnya. Variabel *Energy* diperoleh dengan meningkatkan jumlah *instrument*, mengatur dinamika, ataupun irama pada lagu tersebut. Biasanya semakin energik suatu lagu, maka nilai *Energy* nya pun akan semakin membesar.

3. *Loudness*

Variabel *Loudness* menjelaskan kenyaringan atau kekerasan dari suatu lagu yang direpresentasikan dalam desibel (dB). Nilai dari variabel ini berada di rentang -60 dan 0 dB.

4. *Speechiness*

Variabel *Speechiness* menjelaskan kehadiran kata-kata yang dinyanyikan dalam sebuah lagu. Nilai *Speechiness* diatas 0,66 menjelaskan lagu yang dinyanyikan mungkin seluruhnya terbuat dari kata-kata yang diucapkan.

5. *Acousticness*

Variabel *Acousticness* menjelaskan apakah sebuah lagu tersebut termasuk lagu yang akustik atau tidak. Nilai dari variabel ini berada pada rentang 0,00 sampai dengan 1,00. Semakin mendekati 1,00 maka lagu tersebut cenderung bernuansa akustik, dan begitu pula sebaliknya.

6. *Liveness*

Variabel *Liveness* menjelaskan kehadiran suara penonton dalam rekaman sebuah lagu. Nilai *Liveness* yang lebih tinggi menunjukkan peningkatan kemungkinan bahwa lagu dinyanyikan secara langsung (*live*).

7. *Valence*

Variabel *Valence* menjelaskan indikasi emosi dari sebuah lagu (bahagia atau sedih). Nilai dari variabel ini berada pada rentang 0,00 sampai dengan 1,00. Nilai *Valence* yang mendekati 1,00 menjelaskan emosi positif, sedangkan semakin mendekati 0,00 mewakili emosi negatif.

8. *Tempo*

Variabel *Tempo* menjelaskan cepat atau lambatnya sebuah lagu. Variabel *Tempo* diperoleh dengan meningkatkan jumlah *instrument*, mengatur dinamika, ataupun *Beat*. *Beat* sendiri dapat dijelaskan sebagai ketukan dasar yang menunjukkan banyaknya ketukan dalam periode waktu tertentu.

2.2 Tahapan Penelitian

Tahapan analisis data yang dilakukan pada penelitian ini sebagai berikut:

- Melakukan identifikasi data awal dengan menampilkan statistika deskriptif untuk masing-masing variabel, baik variabel respon (Y) dan variabel prediktor (X).
- Melakukan pemeriksaan asumsi Non-Multikolinearitas menggunakan nilai *Variance Inflation Factor* (VIF).
- Melakukan randomisasi data. Hal ini bertujuan agar semua data memiliki kesempatan yang sama untuk bisa menjadi data *training* dan data *testing*. Proses pembagian data *training* dan data *testing* dengan 2 proporsi pembagian yaitu 90:10 dan 80:20.
- Melakukan pemodelan dengan menggunakan model regresi logistik biner dimana *link function* yang digunakan adalah Logit.
- Melakukan pemodelan dengan menggunakan model analisis diskriminan.
- Evaluasi performansi model klasifikasi pada penelitian ini menggunakan *Classification Accuracy* (Tabel 3) yaitu 1-APER.

Tabel 3 *Classification Accuracy*

<i>Actual</i>	<i>Predicted</i>	
	<i>Positive = Class 0</i>	<i>Negative = Class 1</i>
<i>Positive = Class 0</i>	<i>True Positive</i> (TP)	<i>False Negative</i> (FN)
<i>Negative = Class 1</i>	<i>False Positive</i> (FP)	<i>True Negative</i> (TN)

Apparent Error Rate (APER) menyatakan proporsi obyek yang salah diklasifikasikan oleh fungsi klasifikasi, sehingga untuk melihat proporsi obyek yang tepat diklasifikasikan dapat menggunakan:

$$APER = \frac{\text{jumlah objek yang salah klasifikasi}}{\text{jumlah klasifikasi yang dilakukan}} \times 100\%$$

$$Accuracy = 1 - APER$$

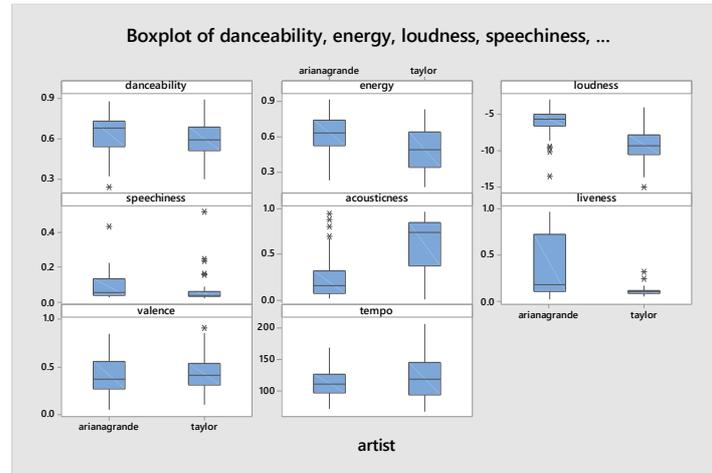
Model klasifikasi yang baik dapat dilihat dari nilai *Accuracy* yang mendekati 1, yang berarti proporsi obyek yang salah diklasifikasikan bisa sekecil mungkin (Ainurrochmah et al., 2019).

3. HASIL DAN PEMBAHASAN

Pada bagian ini akan dibahas mengenai hasil dari proses klasifikasi label *artist music* dengan menggunakan metode regresi logistik biner dan analisis diskriminan.

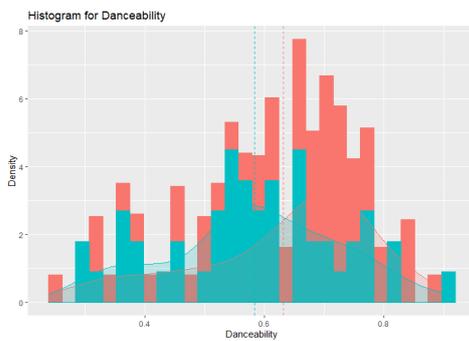
3.1 Statistika Deskriptif

Statistika deskriptif dilakukan guna memberikan gambaran awal dan proses eksplorasi awal dari masing-masing variabel yang digunakan.

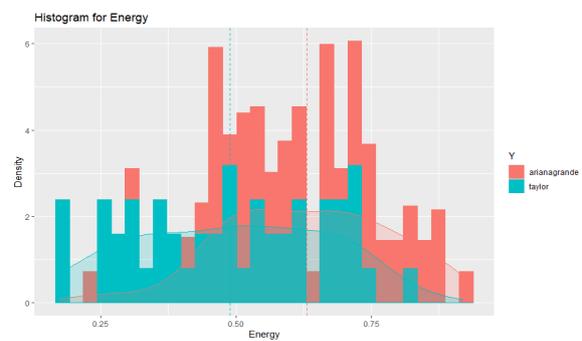


Gambar 1: *Boxplot* untuk Variabel Prediktor

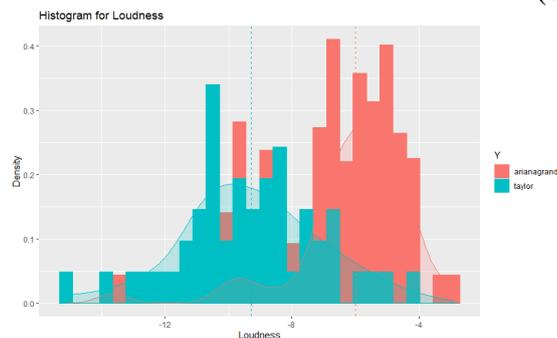
Berdasarkan *Boxplot* pada Gambar 1, dapat kita ketahui secara visual bahwa terdapat karakteristik yang berbeda dari lagu-lagu yang dinyanyikan oleh Ariana Grande dan Taylor Swift. Sebagai contoh untuk variabel *Danceability*, Ariana Grande memiliki rata-rata lagu dengan unsur *Danceability* yang lebih tinggi dibandingkan Taylor Swift. Hal yang sama pada beberapa variabel lainnya, yaitu *Energy* dan *Loudness*. Merujuk dari Gambar 2, untuk lebih jelasnya dapat dilihat pada Gambar 2. Gambar 2(a) menjelaskan jika untuk unsur *Danceability*, Ariana Grande cenderung memiliki lagu-lagu dengan unsur kesesuaian yang dapat membuat kita menari berdasarkan kombinasi tempo, ritme, *beat*, dan lain sebagainya.



(a)



(b)



(c)

Gambar 2: Histogram by Group

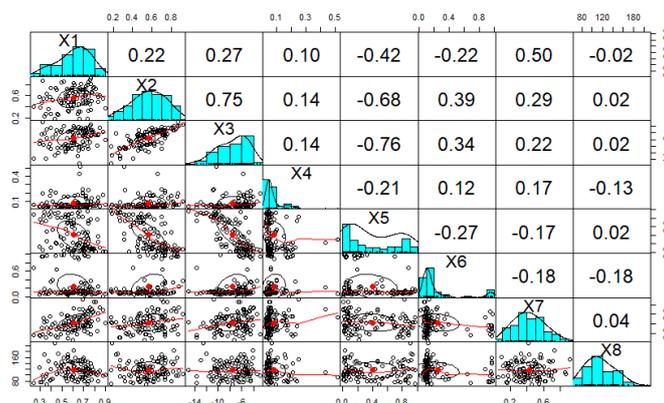
Gambar 2(b) menjelaskan jika untuk unsur *Energy*, Ariana Grande cenderung memiliki lagu-lagu dengan unsur kesesuaian yang dapat membuat kita meningkatkan emosi dari hati. Namun di beberapa variabel lainnya, Taylor Swift memiliki rata-rata lagu dengan unsur yang lebih tinggi, diantaranya pada variabel *Acousticness*, *Tempo*, dan *Valence*. Hal ini menunjukkan bahwa setiap lagu yang dibawakan oleh kedua penyanyi ini memiliki karakteristik dan ciri tersendiri. Tabel uraian statistika deskriptif seperti Mean, Minimum, dan Maksimum ditampilkan pada Tabel 4.

Tabel 4 Nilai VIF

Variabel	X1	X2	X3	X4	X5	X6	X7	X8
Mean	0,61	0,56	-7,58	0,08	0,40	0,25	0,42	116,16
Minimum	0,24	0,17	-15,06	0,03	0,008	0,03	0,05	68,53
Maksimum	0,89	0,92	-2,89	0,52	0,97	0,98	0,92	207,48

3.2 Pemeriksaan Asumsi Non-Multikolinearitas

Pada beberapa kasus, diharapkan bahwa tidak terdapat masalah multikolinearitas pada variabel prediktor. Secara visual dan dengan menghitung nilai korelasi Pearson, disajikan pada *Heatmap Correlation* pada Gambar 3.



Gambar 3: Heatmap Correlation

Berdasarkan Gambar 3 diindikasikan jika tidak terjadi masalah multikolinearitas. Untuk lebih meyakinkan, pemeriksaan Non-Multikolinearitas dilakukan menggunakan nilai *Variance Inflation Factor* (VIF) yang disajikan pada Tabel 5.

Tabel 5 Nilai VIF

Variabel	X1	X2	X3	X4	X5	X6	X7	X8
VIF	1,73	3,38	2,94	1,18	2,43	1,71	1,63	1,28

Berdasarkan Tabel 5, diketahui jika tidak terdapat nilai VIF untuk seluruh variabel prediktor yang lebih dari 10, sehingga dapat disimpulkan jika tidak terjadi masalah multikolinearitas.

3.3 Analisis Klasifikasi

Pada bagian ini, akan dilakukan analisis klasifikasi menggunakan dua metode yang diusulkan, yaitu regresi logistik biner dan analisis diskriminan. Data akan dibagi menjadi data *training* dan *testing* dengan proporsi pembagian yang berbeda, yaitu 90:10 dan 80:20.

Tabel 6 Proporsi 80:20

Metode Klasifikasi	<i>Training (80)</i>	<i>Testing (20)</i>
	1-APER	1-APER
Regresi Logistik Biner	0,87	0,85
Analisis Diskriminan	0,85	0,84

Tabel 7 Proporsi 90:10

Metode Klasifikasi	<i>Training (90)</i>	<i>Testing (10)</i>
	1-APER	1-APER
Regresi Logistik Biner	0,88	0,90
Analisis Diskriminan	0,84	0,89

Berdasarkan Tabel 6 dan Tabel 7, terlihat bahwa model klasifikasi yang terbaik untuk mengklasifikasikan label *artist music* adalah model regresi logistik biner, dimana dengan menggunakan akurasi kebaikan, yaitu 1-APER, model regresi logistik biner memberikan ketepatan klasifikasi yang lebih baik dibandingkan analisis diskriminan. Berikut ditampilkan hasil ketepatan klasifikasi dari model regresi logistik biner dengan proporsi data 90:10 pada data *testing*.

Tabel 8 Data *Testing*

Aktual	Prediksi	
	Ariana Grande	Taylor Swift
Ariana Grande	4	0
Taylor Swift	1	5

Berdasarkan hasil pemodelan, maka dapat dituliskan model regresi logistiknya sebagai berikut:

$$\pi(x) = \frac{\exp(11.0456 + 3.8787 X_1 - 4.4115 X_2 + \dots + -0.0178 X_8)}{1 + \exp(11.0456 + 3.8787 X_1 - 4.4115 X_2 + \dots + -0.0178 X_8)}$$

dan

$$1 - \pi(x) = \frac{1}{1 + \exp(11.0456 + 3.8787 X_1 - 4.4115 X_2 + \dots + -0.0178 X_8)}$$

4. KESIMPULAN

Berdasarkan hasil analisis, didapatkan kesimpulan jika model klasifikasi yang terbaik dalam mengklasifikasikan label *artist music* pada penelitian ini adalah model regresi logistik biner. Berdasarkan dua kombinasi proporsi pembagian data, model regresi logistik biner memberikan performa dan akurasi yang lebih baik dibandingkan analisis diskriminan. Model regresi logistik biner yang dibangun, menggunakan nilai 1-APER dengan proporsi data 90:10 memiliki ketepatan klasifikasi untuk data *training* adalah sebesar 88,00% dan untuk data *testing* adalah sebesar 90,00%.

DAFTAR PUSTAKA

- Al-Beitawi, Z., Salehan, M., & Zhang, S. (2020). What Makes a Song Trend? Cluster Analysis of Musical Attributes for *Spotify* Top Trending Songs. In *Journal of Marketing Development and Competitiveness* (Vol. 14, Issue 3, pp. 79–91). <https://doi.org/10.33423/jmdc.v14i3.3065>
- Ainurrochmah, A., Hayati, M. N., & Satriya, A. M. A. (2019). Perbandingan Klasifikasi Analisis Diskriminan Fisher dan Metode Naïve Bayes. *Jurnal Aplikasi Statistika & Komputasi Statistik*, 11(2), 37-48.
- Annas, S., & Irwan. (2015). Penerapan Analisis Diskriminan dalam Pengelompokan Desa Miskin di Kabupaten Wajo. *SCIENTIFIC PINISI*, 1(1), 34–43.
- Fitria, A., & Azis, H. (2018). Analisis Kinerja Sistem Klasifikasi Skripsi menggunakan Metode Naïve Bayes Classifier. *Prosiding Seminar Nasional Ilmu Komputer Dan Teknologi Informasi*, 3(2), 102–106.
- Giovani, A. P., Ardiansyah, A., Haryanti, T., Kurniawati, L., & Gata, W. (2020). Analisis Sentimen Aplikasi Ruang Guru Di Twitter Menggunakan Algoritma Klasifikasi. *Jurnal Teknoinfo*, 14(2), 115–124. <https://doi.org/10.33365/jti.v14i2.679>
- Hapsary, Maharani Shandra Ayu; Subiyanto, S., & Firdaus, H. S. (2021). Analisa Prediksi Perubahan Penggunaan Lahan Dengan Pendekatan Artificial Neural Network Dan Regresi Logistik Di Kota Balikpapan. *Geodesi Undip*, 10(2), 88–97.
- Hosmer, D. W., & Lemeshow, S. (2000). *Applied Logistic Regression* (2nd ed.). John Wiley & Sons, Inc.
- Jung, H., (2018). The Data Science of K-Pop: Understanding BTS through data and A.I. (<https://towardsdatascience.com/the-data-science-of-k-pop-understanding-bts-through-data-and-a-i-part-1-50783b198ac2>) diakses pada tanggal 01 April 2022.
- Ketut, I., Suniantara, P., & Rusli, M. (2017). Klasifikasi Waktu Kelulusan Mahasiswa Stikom Bali Menggunakan Chaid Regression – Trees dan Regresi Logistik Biner. *Statistika*, 5(1), 27–32.
- Mattjik, A. A., & Sumertajaya, I. M. (2011). Sidik Peubah Ganda dengan Menggunakan SAS. In *Sidik Peubah Ganda Dengan menggunakan SAS*. Institut Pertanian Bogor (IPB Press).
- Nabila, R., Himmati, R., & Erdkhadifa, R. (2021). Perbandingan Regresi Logistik Multinomial dan Analisis Diskriminan. *Journal of Islamic Tourism*, 1(2), 135–150.
- Novianti, F. A., & Purnami, S. W. (2012). Analisis Diagnosis Pasien Kanker Payudara Menggunakan Regresi Logistik dan Support Vector Machine (SVM) Berdasarkan Hasil Mamografi. *Jurnal SAINS Dan Seni ITS*, 1(1), D147–D152.
- Nugroho, S. (2008). *Statistika Multivariat Terapan* (1st ed.). UNIB Press.
- Octisari, S. K., Wijaya, M., & Baroroh, A. A. (2021). Pendekatan Regresi Logistik Pada Perataan Laba. *Eksis: Jurnal Ilmiah Ekonomi Dan Bisnis*, 12(1), 49–53. <https://doi.org/10.33087/eksis.v12i1.236>
- Rajagukguk, N., Ispriyanti, D., & Wilandari, Y. (2015). Perbandingan Metode Klasifikasi Regresi Logistik Biner Dan Naive Bayes Pada Status Pengguna Kb Di Kota Tegal Tahun 2014. *Gaussian*, 4(2), 365–374.
- Santana, I. A. P., Pinhelli, F., Donini, J., Catharin, L., Mangolin, R. B., Da Costa, Y. M. E. G., Delisandra Feltrim, V., & Domingues, M. A. (2020). Music4All: A New Music Database and Its Applications. *International Conference on Systems, Signals, and Image Processing*, 399–404. <https://doi.org/10.1109/IWSSIP48289.2020.9145170>

Suwardika, G. (2017). Pengelompokan Dan Klasifikasi Pada Data Hepatitis Dengan Menggunakan Support Vector Machine (SVM), Classification And Regression Tree (Cart) Dan Regresi Logistik Biner. *Journal of Education Research and Evaluation*, 1(3), 183–191. <https://doi.org/10.23887/jere.v1i3.12016>