

FAKTOR-FAKTOR YANG MEMENGARUHI PENGGUNAAN APLIKASI MUSIK SPOTIFY DENGAN METODE REGRESI LOGISTIK BAYES

Anak Agung Sagung Bulan Indraswari¹, Ni Luh Putu Ari Sekartini²
sagungbulan5@gmail.com¹, arisekartini0207@gmail.com²
Universitas Udayana

ABSTRAK

Penelitian ini bertujuan untuk mengidentifikasi dan menganalisis faktor-faktor yang memengaruhi penggunaan aplikasi spotify di era digital yang semakin berkembang pesat. Fokus utama studi ini adalah menerapkan metode regresi logistik Bayes untuk mendapatkan estimasi parameter yang lebih optimal dibandingkan metode klasik. Menggunakan 100 data sekunder yang bersumber dari datasets Kaggle. Melibatkan satu variabel respons (status penggunaan spotify) dan enam variabel prediktor yaitu usia, waktu mendengarkan musik, bermain musik, menciptakan musik, mendengarkan musik saat bekerja, dan genre. Parameter diestimasi menggunakan pendekatan Bayes dengan simulasi Markov Chain Monte Carlo. Analisis menunjukkan bahwa dari enam variabel yang diuji terdapat tiga variabel yang berpengaruh signifikan yaitu usia (X1), waktu mendengarkan musik (X2) dan kemampuan menciptakan musik (X4). Model regresi yang terbentuk memiliki keakuratan total klasifikasi sebesar 76%, dengan akurasi spesifik untuk kategori pengguna spotify mencapai 89,39%.

Kata Kunci: Spotify, Regresi Logistik Multinomial Bayes, Hamiltonian Monte Carlo (Hmc), Usia, Menciptakan Musik, Mendengarkan Musik.

PENDAHULUAN

Musik adalah salah satu hiburan yang relatif mudah dan sering diakses di era digital. Banyak pengembang aplikasi musik berlomba-lomba menghadirkan layanan terbaik untuk memuaskan para penikmat musik, termasuk di Indonesia. Perkembangan industri musik digital di Indonesia semakin pesat dengan kehadiran berbagai platform layanan streaming. Era digital tidak hanya mengubah industri musik tetapi juga memengaruhi gaya hidup masyarakat serta proses kreatif pembuatan musik itu sendiri.

Kini, hampir semua orang dapat mendengarkan musik secara streaming melalui perangkat digital seperti ponsel pintar. Hal ini memberikan kebebasan kepada para pendengar untuk memilih lagu favorit mereka kapan saja dan di mana saja, tanpa terbatas ruang dan waktu. Secara global, diperkirakan lebih dari 4,8 miliar orang secara teratur mendengarkan musik melalui berbagai platform, baik secara online maupun offline dengan mayoritas pendengar kini mengandalkan layanan streaming (IFPI, 2023). Layanan streaming seperti Langit Musik, Spotify, Joox, Apple Music, atau YouTube Musik menjadi pilihan utama bagi para penikmat musik. Di antara platform tersebut, Spotify telah menjadi layanan yang paling diminati. Berdasarkan laporan kuartal, jumlah pengguna aktif bulanan Spotify meningkat dari 68 juta pada kuartal pertama tahun 2015 menjadi 626 juta pengguna pada kuartal kedua tahun 2024 (Spotify, 2024).

Musik juga memiliki peran penting dalam mendukung produktivitas. Banyak individu mendengarkan musik saat belajar atau bekerja karena musik diyakini dapat meningkatkan suasana hati dan fokus. Pilihan genre yang beragam, seperti pop, rock, metal, atau genre lainnya, memberikan keleluasaan bagi pendengar untuk memilih musik yang sesuai dengan preferensi dan suasana hati mereka. Dengan segala kemudahan dan fitur yang ditawarkan, platform seperti Spotify telah menjadi bagian integral dalam kehidupan modern.

Metode statistika yang dapat digunakan untuk menyelidiki hubungan antara penggunaan spotify dan berbagai faktor yang mungkin memengaruhinya adalah dengan analisis regresi. Karena variabel respons pada kasus ini berupa data kategori maka analisis yang sesuai yaitu analisis regresi logistik (Hosmer et al., 2013). Dalam model regresi, estimasi parameter sangat penting dilakukan untuk memperoleh hasil yang sesuai dan akurat. Terdapat dua metode estimasi parameter yang dapat digunakan yaitu metode klasik dan metode Bayes (Lukman et al., 2021). Metode klasik hanya menggunakan informasi berdasarkan data sampel tanpa memperhitungkan informasi awal, sehingga hasil estimasi sangat bergantung pada kualitas data. Sedangkan metode Bayes menggabungkan fungsi likelihood dan distribusi prior yang memberikan hasil yang lebih optimal (Gelman et al., 2013). Oleh karena itu, penelitian ini bertujuan untuk menerapkan metode regresi logistik menggunakan penduga estimasi Bayes untuk mengetahui faktor-faktor yang memengaruhi penggunaan aplikasi spotify.

METODE

Data yang digunakan pada penelitian ini sebanyak 100 data yaitu konsumen pengguna aplikasi musik spotify. Dalam penelitian ini, data yang digunakan merupakan data sekunder yang diperoleh dari datasets kaggle. Variabel penelitian ini meliputi satu variabel respons yaitu status penggunaan spotify dan enam variabel prediktor dengan rincian sebagai berikut.

Tabel 1. Variabel Penelitian

Variabel	Keterangan	Kategori	Skala
Y	Status Penggunaan Spotify	0: Tidak Menggunakan 1: Menggunakan	Nominal
X1	Usia		Rasio
X2	Waktu Mendengarkan Musik		Rasio
X3	Bermain Alat Musik	0: Tidak Bermain 1: Bermain	Nominal
X4	Menciptakan Musik	0: Tidak Menciptakan 1: Menciptakan	Nominal
X5	Mendengarkan Musik Saat Bekerja	0: Tidak 1: Mendengarkan	Nominal
X6	Genre	0: Pop 1: Rock 2: Metal 3: Lainnya	Nominal

Dalam penelitian ini, adapun langkah-langkah analisis data yang dilakukan sebagai berikut:

1. Menentukan fungsi likelihood yang mengikuti bentuk Bernoulli dengan persamaan (Kurniawati & Tri Sutanto MSi, 2019):

$$f(y|\beta) = \prod_{i=1}^{100} \left(\frac{e^{\beta_0 + \beta_1 x_{i1} + \dots + \beta_k x_{ik}}}{1 + e^{\beta_0 + \beta_1 x_{i1} + \dots + \beta_k x_{ik}}} \right)^{y_i} \left(1 - \frac{e^{\beta_0 + \beta_1 x_{i1} + \dots + \beta_k x_{ik}}}{1 + e^{\beta_0 + \beta_1 x_{i1} + \dots + \beta_k x_{ik}}} \right)^{1-y_i}$$

2. Menentukan distribusi prior dari parameter yaitu dengan menggunakan distribusi prior normal karena parameter β_j dapat berada pada rentang $(-\infty, \infty)$ dan tidak ada informasi sebelumnya tentang β_j (Lukman et al., 2021).

$$f(\beta_j) = \frac{1}{\sqrt{2\pi\sigma_j^2}} \exp \left\{ -\frac{1}{2\sigma_j^2} (\beta_j - \mu_0)^2 \right\}$$

- Menentukan distribusi posterior yaitu dengan mengalikan informasi fungsi likelihood dengan distribusi prior sesuai persamaan (1).

$$f(\beta|data) \propto \prod_{i=1}^{100} \left(\frac{e^{\beta_0 + \beta_1 x_{i1} + \dots + \beta_k x_{ik}}}{1 + e^{\beta_0 + \beta_1 x_{i1} + \dots + \beta_k x_{ik}}} \right)^{y_i} \left(1 - \frac{e^{\beta_0 + \beta_1 x_{i1} + \dots + \beta_k x_{ik}}}{1 + e^{\beta_0 + \beta_1 x_{i1} + \dots + \beta_k x_{ik}}} \right)^{1-y_i} \times \prod_{j=0}^8 \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left\{ -\frac{1}{2\sigma^2} (\beta_j - \mu_j)^2 \right\}. \quad (1)$$

- Melakukan simulasi Markov Chain Monte Carlo menggunakan algoritma Hamiltonian Monte Carlo (HMC) dengan bantuan software Rstudio dan menguji kekonvergenan hasil simulasi menggunakan trace plot dan plot autokorelasi. Jika pada plot autokorelasi nilai lag sudah mendekati nol dan trace plot tidak membentuk pola periodik maka data dapat dikatakan konvergen (Susila, 2016).
- Melakukan pengujian parameter menggunakan credible interval. Parameter yang diperoleh memiliki pengaruh yang signifikan terhadap prediktor jika credible interval tidak mengandung nilai nol.
- Membentuk model regresi logistik Bayes dari parameter yang signifikan.
- Interpretasi dan membuat kesimpulan mengenai faktor-faktor yang berpengaruh signifikan.

HASIL DAN PEMBAHASAN

Berdasarkan data yang diperoleh, gambaran karakteristik masing-masing variabel dapat dilihat pada Tabel 2.

Tabel 2. Tabulasi Silang Variabel Kategori

Variabel(X)		Status Penggunaan Spotify (Y)		Total
		Tidak Menggunakan	Menggunakan	
Bermain Alat Musik (X3)	Tidak	9	13	
	Ya	25	53	
Total		34	66	100
Menciptakan Musik (X4)	Tidak	19	50	
	Ya	15	16	
Total		34	66	100
Mendengarkan Musik saat Bekerja (X5)	Tidak	26	54	
	Ya	8	12	
Total		34	66	100
Genre (X6)	Pop	3	8	
	Rock	10	16	
	Metal	2	9	
	Lainnya	19	33	
Total		34	66	100

Sumber: data diolah

Tabel 2 menyajikan frekuensi dari masing-masing variabel kategorik terhadap variabel respon. Dapat dilihat bahwa mayoritas responden menggunakan spotify sebagai aplikasi streaming musik sebanyak 66 orang. Responden yang bisa bermain alat musik lebih dominan dibanding tidak bermain alat musik dalam menggunakan spotify. Responden pengguna spotify lebih dominan tidak bekerja di bidang musik atau tidak bisa menciptakan musik. Frekuensi pengguna aplikasi spotify cenderung tidak mendengarkan musik pada saat bekerja. Genre yang lebih diminati oleh responden merupakan lainnya.

Tabel 3. Statistik Deskriptif Variabel Kontinu

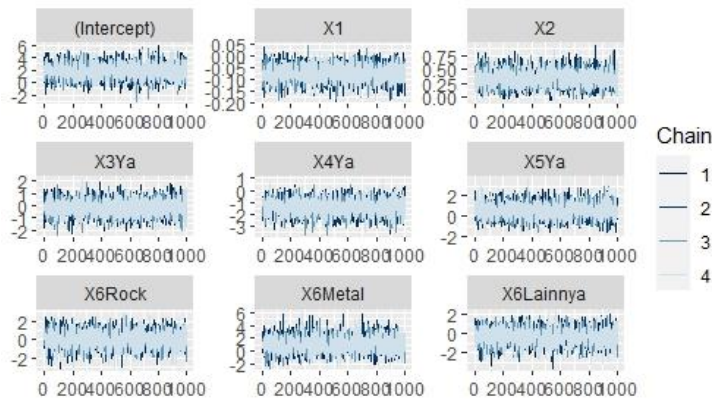
Variabel	Minimum	Q1	Median	Mean	Q3	Maksimum
Usia (X1)	14	17	19	22.3	23.25	63
Waktu Mendengarkan Musik (X2)	0	2	3	3.9	5	24

Sumber: data diolah

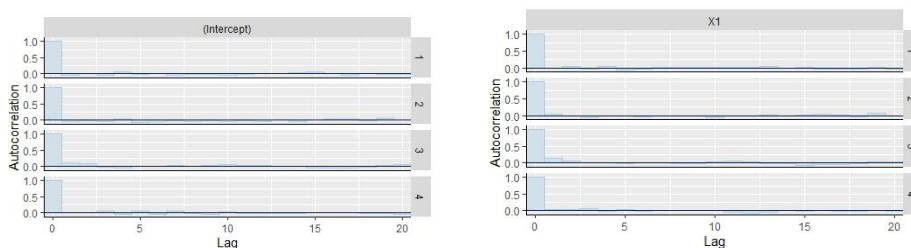
Pada Tabel 3 disajikan statistik deskriptif variabel kontinu. Untuk rata-rata usia responden yaitu 22 tahun dengan usia minimum 14 tahun dan usia maksimum adalah 63 tahun. Q1 untuk usia adalah 17 tahun yang berarti kurang lebih 25% dari keseluruhan sampel berusia kurang dari atau paling tinggi berusia 17 tahun sedangkan Q3 sebesar 23 tahun yang berarti kurang lebih 75% dari seluruh sampel berusia kurang dari 23 tahun. Untuk rata-rata waktu mendengarkan musik sebesar 3,9 jam dengan minimum mendengarkan musik 0 jam dan maksimum mendengarkan musik 24 jam. Q1 untuk waktu mendengarkan musik adalah 2 jam yang berarti kurang lebih 25% dari seluruh sampel mendengarkan lagu kurang dari atau paling lama 2 jam, sedangkan Q3 sebesar 5 jam yang berarti kurang lebih 75% dari seluruh sampel mendengarkan lagu kurang dari 5 jam.

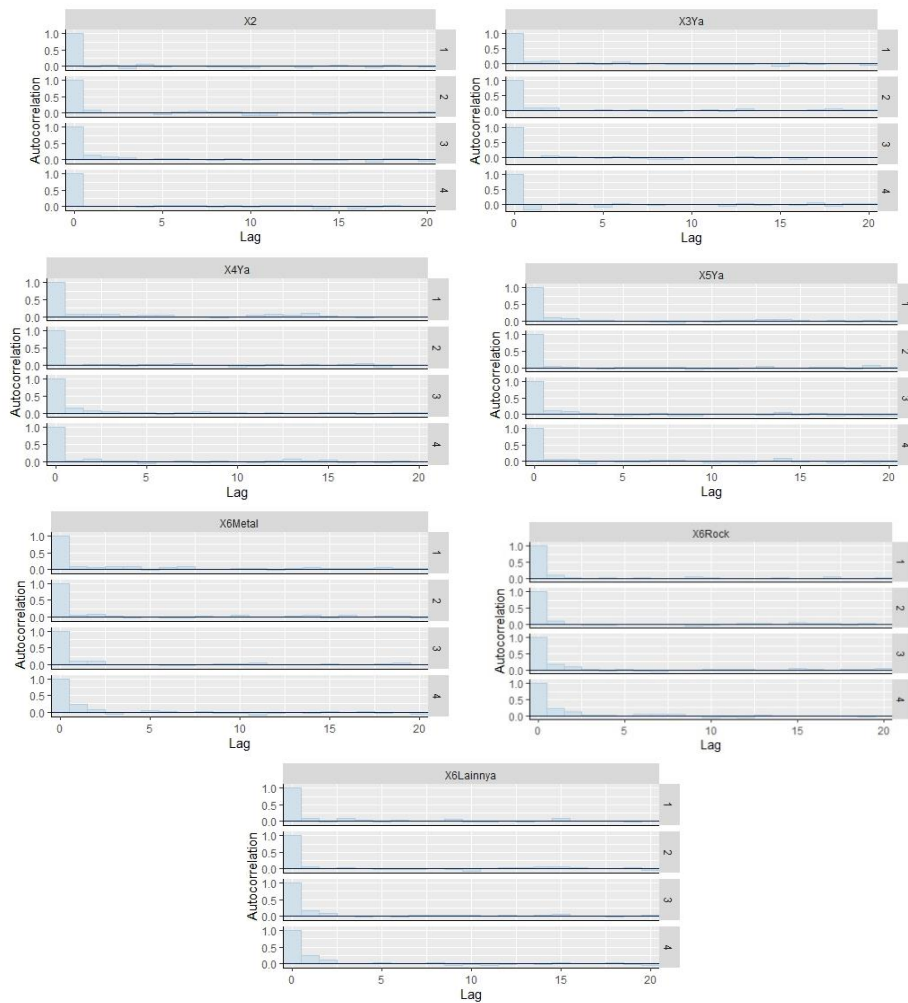
Simulasi MCMC

Distribusi posterior pada persamaan (1) memiliki bentuk analitik yang sulit, maka untuk pengambilan sampel dari distribusi posterior dapat dilakukan dengan menerapkan simulasi Markov Chain Monte Carlo (MCMC). Proses simulasi dilakukan dengan bantuan software RStudio sebanyak 4.000 sampel posterior menggunakan algoritma Hamiltonian Monte Carlo (HMC) pada library rstanarm. Selanjutnya dilakukan pengujian kekonvergenan dengan melihat trace plot dan plot autokorelasi sebagai berikut.



Gambar 1. Trace Plot





Gambar 2. Plot Autokorelasi

Berdasarkan Gambar 1 pada trace plot dapat dilihat bahwa sampel untuk setiap parameter berkumpul di sekitar nilai tertentu tanpa ada tren naik atau turun yang jelas menunjukkan bahwa rantai sudah mengeksplorasi distribusi posterior dengan baik serta pada Gambar 2 autokorelasi menurun dengan tajam di lag awal dan mendekati nol yang menandakan sampel relatif independen dan berkualitas baik maka sampel dapat dikatakan konvergen.

Estimasi Parameter dan Pengujian Hipotesis

Estimasi parameter yang digunakan pada penelitian ini adalah nilai mean atau rata-rata sampel. Nilai estimasi parameter disajikan dalam bentuk Tabel 4 berikut.

Tabel 4. Estimasi Parameter

Variabel	Mean	Kuantil 2.5%	Kuantil 97.5%	Keterangan
<i>Intercept</i>	1.8	-0.52	4.22	—
X_1 (Usia)	-0.1	-0.14	-0.01	Signifikan
X_2 (Waktu Mendengarkan Musik (jam))	0.3	0.08	0.64	Signifikan
X_3 (Ya)	-0.1	-1.34	1.06	Tidak Signifikan
X_4 (Ya)	-1.4	-2.75	-0.23	Signifikan
X_5 (Ya)	0.6	-0.91	2.20	Tidak Signifikan
X_6 (Rock)	-0.1	-1.94	1.67	Tidak Signifikan
X_6 (Metal)	1.0	-1.38	3.73	Tidak Signifikan
X_6 (Lainnya)	-0.3	-2.03	1.22	Tidak Signifikan

Pengujian hipotesis akan dilakukan dengan pendekatan credible interval 95%. Akan dihitung dengan batas bawah yaitu kuantil ke 2.5% dengan batas atas yaitu kuantil 97.5%. Parameter yang dinyatakan signifikan jika credible interval tidak mengandung nol.

Hipotesis yang akan digunakan dalam penelitian ini adalah

$$H_0 : \beta_j = 0$$

$$H_1 : \beta_j \neq 0, \text{ dengan } j = 1, 2, 3, \dots 6.$$

Berdasarkan tabel estimasi tersebut menyatakan bahwa terdapat variabel yang berpengaruh signifikan, di antaranya variabel Usia (X_1), waktu mendengarkan musik dengan satuan jam (X_2), dan Bisa menciptakan musik dengan kategori Ya (X_4).

Pembentukan Model Regresi Logistik

Tujuan dari pembentukan model regresi adalah untuk mengetahui hubungan antara variabel bebas dengan variabel respons. Langkah dalam memodelkan regresi logistik Bayes yaitu mendapatkan parameter-parameter dari model yang diperoleh dari proses iterasi MCMC yang telah dilakukan sebelumnya. Sehingga diperoleh hasil persamaan regresi sebagai berikut

$$\pi(x) = \frac{\exp(-0,1x_1 + 0,3x_2 - 1,4x_4)}{1 + \exp(-0,1x_1 + 0,3x_2 - 1,4x_4)}$$

dan bentuk logit dari persamaan tersebut adalah

$$g(x) = \ln\left(\frac{\pi(x)}{1-\pi(x)}\right) = -0,1x_1 + 0,3x_2 - 1,4x_4$$

Pada regresi logistik, koefisien-koefisien yang didapatkan sulit untuk diinterpretasikan secara langsung sehingga perlu dilakukan interpretasi dengan menggunakan nilai odds ratio namun tetap memperhatikan nilai dari $\exp(\beta)$.

1. Pada $\exp(\beta_1)=0,904$

Variabel usia dalam memengaruhi penggunaan aplikasi musik spotify 0,904 kali lebih signifikan sebagai faktor yang memengaruhi penggunaan aplikasi spotify.

2. Pada $\exp(\beta_2)=1,349$

Variabel waktu mendengarkan musik dalam memengaruhi penggunaan aplikasi musik spotify 1,349 kali lebih signifikan sebagai faktor yang memengaruhi penggunaan aplikasi spotify.

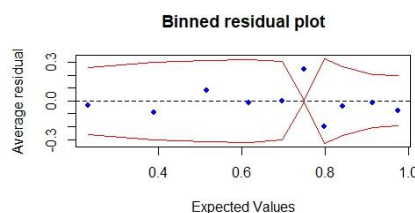
3. Pada $\exp(\beta_4)=0,246$

Variabel menciptakan musik dengan kategori bisa menciptakan musik memengaruhi penggunaan aplikasi musik spotify 0,246 kali lebih lebih besar dibandingkan tidak menciptakan musik sebagai faktor yang memengaruhi penggunaan aplikasi spotify.

Berdasarkan interpretasi dapat dikatakan bahwa variabel yang berpengaruh secara signifikan terhadap penggunaan aplikasi musik spotify adalah usia, waktu mendengarkan musik dan menciptakan musik. Sedangkan variabel bermain alat musik, mendengarkan musik saat bekerja, dan genre tidak memiliki pengaruh secara signifikan terhadap penggunaan aplikasi musik spotify.

Diagnostik Model

Diagnostik model digunakan untuk mengetahui kecocokan model dengan menggunakan binned residual plot diperoleh hasil sebagai berikut.



Gambar 3. Binned Residual Plot

Pada Gambar 3 garis merah memperlihatkan rentang nilai residu yang diharapkan dalam model yang baik dengan batas ± 3 standar error bands sedangkan titik-titik biru pada plot mewakili rata-rata residu. Suatu model dikatakan sesuai harapan apabila 95% titik berada di dalam batas standar error bands. Pada plot diatas, menunjukkan bahwa sebagian besar titik berada di dalam batas standar error bands dan hanya ada satu titik berada di luar batas standar error bands.

Klasifikasi dan Keakuratan Model

Tabel 5. Klasifikasi Model

Pengamatan	Prediksi		Total
	Tidak menggunakan (0)	Menggunakan (1)	
Tidak Menggunakan (0)	17	17	34
Menggunakan (1)	7	59	66
Total	24	76	100

Sumber: data diolah

Tabel 5 menunjukkan hasil klasifikasi dari model regresi logistik biner dengan cut point 0,5. Angka 17 menunjukkan jumlah data yang benar-benar "Tidak menggunakan" dan diklasifikasikan sebagai "Tidak menggunakan" oleh model. Sedangkan angka 59 menunjukkan jumlah data yang benar-benar "Menggunakan" dan diklasifikasikan sebagai "Menggunakan" oleh model. Sedangkan jumlah data yang sebenarnya "Tidak menggunakan" tetapi diklasifikasikan sebagai "Menggunakan" oleh model adalah sebesar 17 dan jumlah data yang sebenarnya "Menggunakan" tetapi diklasifikasikan sebagai "Tidak menggunakan" oleh model adalah sebesar 7.

Tingkat keakuratan total klasifikasi sebesar 76% yang diperoleh dengan cara membagi jumlah prediksi benar dengan jumlah keseluruhan data, yaitu $((17+59))/100 \times 100\%$. Sedangkan tingkat keakuratan klasifikasi masing-masing grup sebesar $59/66 \times 100\% = 89,39\%$ untuk klasifikasi status menggunakan spotify (1) dan sebesar $17/34 \times 100\% = 50\%$ untuk klasifikasi status tidak menggunakan spotify (0).

KESIMPULAN

Berdasarkan hasil penelitian yang telah diperoleh, maka dapat ditarik beberapa kesimpulan sebagai berikut:

1. Model regresi logistik yang diperoleh untuk penggunaan spotify adalah sebagai berikut.

$$\pi(x) = \frac{e^{\beta_1 x_1 + \beta_2 x_2 + \beta_4 x_4}}{1 + e^{\beta_1 x_1 + \beta_2 x_2 + \beta_4 x_4}} = \frac{e^{-0,1x_1 + 0,3x_2 - 1,4x_4}}{1 + e^{-0,1x_1 + 0,3x_2 - 1,4x_4}}$$

Bentuk logit dari persamaan tersebut adalah

$$g(x) = \ln\left(\frac{\pi(x)}{1 - \pi(x)}\right) = -0,1x_1(\text{usia}) + 0,3x_2(\text{jam}) - 1,4x_4(\text{ya})$$

2. Dari enam variabel bebas yang digunakan, diperoleh tiga variabel yang signifikan berpengaruh terhadap penggunaan aplikasi musik spotify, di antaranya variabel usia (X_1), variabel waktu mendengarkan musik (X_2), dan variabel menciptakan musik (X_4) dengan kategori ya. Variabel waktu mendengarkan musik memiliki hubungan positif dengan penggunaan aplikasi musik spotify.

DAFTAR PUSTAKA

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2013). Bayesian Data Analysis (Third). Chapman and Hall/CRC. <https://doi.org/10.1201/b16018>

- Hosmer, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied Logistic Regression: Vol. Third*. Wiley. <https://doi.org/10.1002/9781118548387>
- IFPI. (2023). *Global Music Report 2023*.
- Kurniawati, D., & Tri Sutanto MSi, H. (2019). Faktor-Faktor Yang Mempengaruhi Anemia Remaja Putri Dengan Menggunakan Bayesian Regresi Logistik dan Algoritma Metropolis-Hasting. *Jurnal Ilmiah Matematika*, 7(1).
- Lukman, P. A., Abdullah, S., & Rachman, A. (2021). Bayesian Logistic Regression and Its Application for Hypothyroid Prediction in Post-radiation Nasopharyngeal Cancer Patients. *Journal of Physics: Conference Series*, 1725(1). <https://doi.org/10.1088/1742-6596/1725/1/012010>
- Spotify. (2024). *Spotify Q2 2024 Earnings Report*. <https://investors.spotify.com/financials/default.aspx>
- Susila, M. R. (2016). *Pemodelan Regresi Logistik Biner Bivariat Bayesian Untuk Respon Yang Unbalance Studi Kasus: Konsumen Produk Low Price Software Antivirus Perusahaan 'X'*. Institut Teknologi Sepuluh Nopember.