



Application of Supervised Learning Algorithm for Music Listening Data Classification

Penerapan Algoritma Supervised Learning untuk Klasifikasi Data Music Listening

Eri Mardiani^{1*}, Nur Rahmansyah², Ira Kurniati³, Andy Setiawan⁴, Diah Widiastuti⁵,
Muhammad Ridwan⁶, Muhammad Zidan Rosyid⁷, Ari Febriansyah⁸

¹Program Studi Informatika, FTKI, Universitas Nasional, Indonesia,

²Program Studi Animasi Politeknik Negeri Media Kreatif, Indonesia,

³Program Studi Bisnis Digital, Fakultas Ekonomi dan Bisnis,
Institut Teknologi dan Bisnis Swadharma, Jakarta, Indonesia

^{4,6,7,8}Program Studi Akuntansi, Fakultas Ekonomi dan Bisnis,
Universitas Pembangunan Nasional Veteran Jakarta, Indonesia,
⁵Program Studi Teknik Mesin, Universitas Nasional, Indonesia

Email: ¹erimardiani1@gmail.com, ²nur_rahmansyah@polimedia.ac.id, ³irakurniati4@gmail.com,
⁴andysetiawan2285@upnvj.ac.id, ⁵diahwidiastuti712@gmail.com, ⁶muhammadridwann69@gmail.com,
⁷muhammadzidanrosyid@gmail.com, ⁸arfebrian123@gmail.com

Received Jun 30th 2023; Revised Aug 05th 2023; Accepted Sept 25th 2023
Corresponding Author: Eri Mardiani

Abstract

The development of information technology is increasing rapidly until now the latest technologies are being developed, the use of data as processing is now increasingly being applied to various fields, with the use of data technological developments can be carried out that can improve quality and efficiency, data processing can be applied in all fields and one one is in the field of music. Music has become a part of everyday human life. There are already many media that present various kinds of songs with various genres, starting from Pop, Jazz, Rock, R&B, and other genres. By using data mining to process data and analyze music data we can predict the diversity of music listening preferences and by using the Orange Data Mining application tools can help the public to find out what music is interested in by predicting it through several methods in orange tools, with the Naive Algorithm Bayes has a better level of accuracy.

Keyword: Big Data, Data Mining, Music, Orange Tools

Abstrak

Perkembangan teknologi informasi semakin pesat hingga saat ini terus dikembangkan teknologi-teknologi terbaru, penggunaan data sebagai pengolahan kini semakin banyak diterapkan pada berbagai bidang, dengan penggunaan data dapat dilakukan perkembangan teknologi yang dapat meningkatkan kualitas dan efisiensi, pengolahan data dapat diterapkan di semua bidang dan salah satunya adalah di bidang musik. Musik sudah menjadi bagian dari kehidupan manusia sehari-hari. Sudah banyak media-media yang menyajikan berbagai macam lagu dengan genre yang beragam pula mulai dari Pop, Jazz, Rock, R&B, dan genre-genre lainnya. Dengan menggunakan data mining untuk mengolah data maupun menganalisis data-data musik kita dapat memprediksi keanekaragaman preferensi mendengarkan musik dan dengan menggunakan tools aplikasi Orange Data Mining dapat membantu masyarakat untuk mengetahui musik apa saja yang diminat dengan memprediksi melalui beberapa metode dalam tools orange, dengan Algoritma Naive Bayes memiliki tingkat akurasi yang lebih baik.

Kata Kunci: Big Data, Data Mining, Music, Orange Tools

1. PENDAHULUAN

Musik sudah menjadi bagian dari kehidupan manusia sehari-hari. Sudah banyak media-media yang menyajikan berbagai macam lagu dengan genre yang beragam pula mulai dari Pop, Jazz, Rock, R&B, dan

genre-genre lainnya [3]. Dengan musik, emosi seseorang dapat terpengaruhi, sehingga dapat menjadi alat penenang, dan dapat membantu seseorang lebih rileks. Banyaknya penelitian yang meneliti tentang manfaat mendengarkan lagu pun turut menjadi pendorong manusia tidak lepas dari musik dalam melakukan kesehariannya.

Musik adalah seni yang terbuat dari kumpulan irama dan suara yang selaras serta harmonis. Dengan , musik seringkali menjadi cara seseorang untuk menenangkan diri dari kerasnya kehidupan. Oleh karena itu dibuat lah kumpulan data (dataset) tentang Music Listening Preferences Diversity [1]. Kumpulan data ini dikumpulkan sebagai bagian dari proyek penelitian untuk mengungkap beragam pengalaman dan preferensi mendengarkan musik yang dimiliki individu berdasarkan faktor-faktor seperti jenis kelamin, ras, atau kebangsaan.

Berdasarkan data Music Listening Preferences Diversity yang olah menggunakan aplikasi orange pada Participan ID (X) dan What is your age (Y) mempengaruhi orang yang mendengarkan musik dengan pengetahuan sebelumnya tentang trek dan artist musik dengan titik tertinggi yaitu pada titik participad Id 86 dan Age prefer not to say, dan terendah pada participian Id 31, Age 18-24 years old. Sedangkan untuk orang yang mendengarkan musik dengan trek dan artis terkini atau sekarang titik tertingginya adalah participian Id 34 dan Age 55-64 years old dan terendah pada participian Id 3 dan Age 18-24 years old.

Dan juga melihat penelitian mengenai genre musik seperti pada penelitian yang dilakukan menggunakan datamining mengenai dengan judul “Identifikasi Genre Musik dengan Menggunakan Metode Random Forest”. penelitian tersebut sedikitnya memiliki tujuan yang sama yaitu untuk memahami genre musik apa yang minati [3]. Metode yang tepat untuk digunakan untuk Identifikasi Genre Musik sehingga mengetahui metode apa yang pas untuk mengetahui orang yang mendengarkan musik dengan metode yang sesuai pada penelitian ini.

Penelitian yang dilakukan oleh G. A.V. Mastrika Giri [15]telah membangun sistem rekomendasi musik otomatis dengan metode Case-Based Reasoning(CBR)dengan menggunakan data listening history dan konteks pengguna. Metode CBR,memanfaatkan pengalaman sebelumnya untuk memecahkan masalah baru, adalah cara yang efektif untuk mendukung pemecahan masalah. Teori ini menyatakan bahwa ketika pembelajar dihadapkan pada masalah yang kompleks, mereka dapat menggunakan kasus-kasus sebelumnya untuk menginterpretasikan situasi baru dan memperoleh solusi [2]. Permasalahan yang ditemui dalam penelitian sebelumnya yang dilakukan oleh G. A. V. Mastrika Giri [15] adalah nilai precision berdasarkan preferensi pengguna tidak tinggi, hanya 66%. Berdasarkan analisis hasil rekomendasi, terdapat banyak hasil rekomendasi yang tidak sesuai dengan genre yang biasa didengarkan atau artis favorit dari pendengar.Maka dari itu diperlukan sebuah sistem rekomendasi yang bisa memberikan rekomendasi musik sesuai dengan artis yang disukai oleh pendengar.

2. METODOLOGI PENELITIAN

Ada banyak kemajuan teknologi yang tersedia dengan komputer. Komputer tidak hanya membantu kita dalam menyelesaikan pekerjaan, tetapi juga dapat menjadi alat yang menyenangkan dan berguna untuk digunakan[13]. Untuk lebih memaksimalkan pekerjaan teknologi berbasis internet sangat mendukung pekerjaan secara online. Selain itu, dengan teknologi ineternet pekerjaan menjadi jauh lebih cepat dan lebih mudah selesai daripada penggunaan komputer tradisional. Perkembangan teknologi informasi sudah sedemikian pesat. Teknologi Internet menghubungkan ribuan jaringan komputer individual dan organisasi di seluruh dunia [8].

Perkembangan zaman memberikan peningkatan data yang semakin banyak data yang dibutuhkan oleh setiap manusia [10]. Dengan banyaknya data yang ada, maka semakin sulit juga untuk menganalisis data tersebut dengan cara manual. Oleh karena itu, manusia membutuhkan bantuan data mining untuk mengumpulkan informasi menarik dari data yang tersedia. Dengan menggunakan data mining maka akan mendapatkan penemuan pengetahuan yang menarik dari sejumlah besar data yang disimpan dalam basis data, gudang data [7].

Big Data Analytics dapat menawarkan keuntungan besar bagi bisnis yang menggunakannya secara efektif, termasuk pengambilan keputusan yang lebih baik dan lebih cepat, lebih banyak kepuasan pelanggan, penemuan peluang baru, pengoptimalan proses, dan pembuatan barang dan layanan yang lebih baik [4].

Studi Literatur Tentang Naive Bayes Algorithm for Sentiment analysis Windows PhoneStore Application Reviews dapat disimpulkan bahwa, menganalisa pengguna aplikasi Windows Phone Store dengan mengklasifikasikan review ke dalam kategori opini positif atau negative [7].

Dataset digunakan untuk melihat dampak dari data categorical (Gender, Race/ethnicity, Parental level of education, Lunch, and Test preparation course) terhadap data numeric (nilai matematika, membaca, dan menulis) [5]. Adapun penelitiann ini menggunakan 5 algoritma Data Mining dengan spesifikasi Teknik Klasifikasi [5].

2.1 K-Nearest Neighbor

Algoritma K-Nearest Neighbor (k-NN) merupakan sebuah model algoritma yang digunakan untuk melakukan klasifikasi terhadap objek yang berdasarkan dari data pembelajaran yang jaraknya paling dekat dengan objek tersebut [14]. Ataupun dapat dipahami juga bahwa k-nearest neighbor adalah salah satu algoritma yang paling sederhana dan banyak digunakan. Titik data akan diklasifikasikan berdasarkan kesamaan kelompok tertentu dari titik data lain yang berdekatan. Sehingga, algoritma ini akan memberikan hasil yang kompetitif [4]. Teorema K-NN dapat dilihat pada persamaan 1.

$$d_i = \sqrt{\sum_{i=1}^n (x_{ij} - p_j)^2} \quad (1)$$

Keterangan

d_i	= Jarak sampel
x_{ij}	= Data sampel pengetahuan
p_j	= Data input var ke-j
n	= Jumlah sampel

2.2 Naïve Bayes

Salah satu metode data mining ialah klasifikasi Naive Bayes. Naive Bayes Classifier adalah metode klasifikasi yang berakar pada teorema Bayes. Naive Bayes merupakan metode pengklasifikasian berdasarkan probabilitas sederhana dan dirancang agar dapat dipergunakan dengan asumsi antar variabel penjas saling bebas (independen) [5]. Pada algoritma ini pembelajaran lebih ditekankan pada pengestimasian probabilitas. Tujuan dari metode Naive Bayes adalah untuk menemukan probabilitas ketika kita mengetahui probabilitas tertentu lainnya. Hasil dari perhitungan data mining menggunakan metode klasifikasi Naive Bayes akan makin berguna jika penyajiannya menarik dan dapat dipahami dengan baik oleh penerima data [8]. Teorema Naive Bayes dapat ditunjukkan pada persamaan 2 [5]:

$$P(H|X) = \frac{P(X|H).P(H)}{P(X)} \quad (2)$$

Keterangan

X	= Data dengan class yang belum diketahui
H	= Hipotesis data merupakan suatu class spesifik
$P(H X)$	= Probabilitas hiotesis H berdasar kondisi X (Posteriori Probabilitas)
$P(H)$	= Probabilitas Hipotesis H (Prior Probabilitas)
$P(X H)$	= Probabilitas X berdasarkan kondisi hipotesis H
$P(X)$	= Probabilitas X

2.3 Decision Tree

Decision tree adalah algoritma machine learning yang menggunakan seperangkat aturan untuk membuat keputusan dengan struktur seperti pohon yang memodelkan kemungkinan hasil, biaya sumber daya, utilitas dan kemungkinan konsekuensi atau resiko. Konsepnya adalah dengan cara menyajikan algoritma dengan pernyataan bersyarat, yang meliputi cabang untuk mewakili langkah-langkah pengambilan keputusan yang dapat mengarah pada hasil yang menguntungkan. Klasifikasi ini menggunakan observasi pada node untuk menemukan target pada leaves. Decision Tree merupakan salah satu metode klasifikasi yang paling populer karena mudah untuk diinterpretasi oleh manusia dengan kemampuannya untuk mem-break down proses pengambilan keputusan yang kompleks menjadi lebih simple [9]. Teorema Decision Tree ditunjukkan pada persamaan 3 [5].

$$Entropy(S) = \sum_{i=1}^n p_i * \log_2 p_i \quad (3)$$

Keterangan:

S	= Himpunan kasus
A	= Atribut
n	= Jumlah patisi S
P_i	= Proporsi dari S_i terhadap S

2.4 Ensemble Method

Ensemble Method adalah algoritma dalam pembelajaran mesin (machine learning) dimana algoritma ini sebagai pencarian solusi prediksi terbaik dibandingkan dengan algoritma yang lain karena metode ensemble ini menggunakan beberapa algoritma pembelajaran untuk pencapaian solusi prediksi yang lebih baik daripada

algoritma yang bisa diperoleh dari salah satu pembelajaran algoritma konstituen saja. Tidak seperti ensemble statistika dalam mekanika statistika biasanya selalu tak terbatas. Ensemble Pembelajaran hanya terdiri dari seperangkat model alternatif yang bersifat terbatas, namun biasanya memungkinkan untuk menjadi lebih banyak lagi struktur fleksibel yang ada diantara alternatif model itu sendiri. Evaluasi prediksi dari ensemble biasanya memerlukan banyak komputasi daripada evaluasi prediksi model tunggal (single model) [11].

$$F(x) = \sum_{t=1}^T \alpha_t h_t(x) \quad (4)$$

Keterangan:

- $h(x)$ = Pengklasifikasi dasar atau lemah
- α = Tingkat pembelajaran (learning rate)
- $F(x)$ = Hasil, berupa pengklasifikasi kuat atau akhir

2.5 Linear Regression

Linear regression merupakan salah satu algoritma yang memodelkan suatu persamaan untuk menghitung estimasi. Pada metode ini bertujuan untuk mencari pola pada nilai numerik, sehingga data yang dibutuhkan berupa data numerik agar dapat diolah dengan model algoritma ini. Linear regression memprediksi nilai data yang tidak diketahui dengan menggunakan nilai data lain yang terkait dan diketahui. Secara matematis memodelkan variabel yang tidak diketahui atau tergantung dan variabel yang dikenal atau independen sebagai persamaan linier. Teorema Linear Regression ditunjukkan pada persamaan 5 [7]:

$$Y = a + bX \quad (5)$$

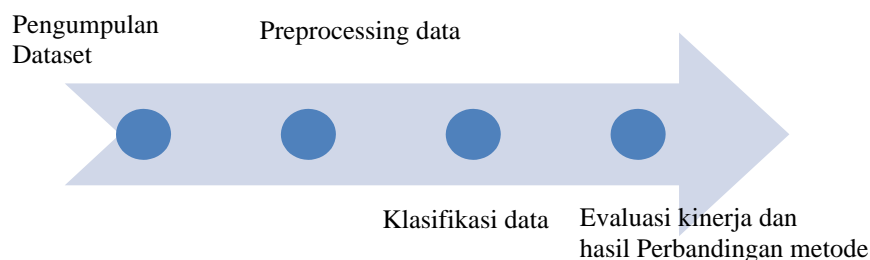
Mengidentifikasi Atribut dan Label

Keterangan:

- Y = Variabel terikat (dependen)
- a = konstanta
- b = koefisien regresi (kemiringan); besaran Respon yang ditimbulkan oleh variabel
- X = Variabel tidak terikat (Independen)

2.6 Tahapan Penelitian

Data yang dikumpulkan untuk penelitian ini yaitu primer dan sekunder. Teknik pengumpulan data primer berasal dari studi literatur seperti buku maupun jurnal. Adapun Teknik pengumpulan data sekunder pada penelitian ini diperoleh dari media online maupun sumber – sumber lainnya. Studi pustaka akan digunakan oleh peneliti guna menjabarkan dan menganalisis data yang sudah dikumpulkan yang memiliki kaitan dengan penelitian ini [6].



Gambar 1. Tahapan Penelitian

Tahapan pada penelitian ini terdiri dari 4 tahap yaitu pengumpulan dataset, pengujian data, proses prediksi, evaluasi kinerja dan hasil perbandingan metode [3].

3. HASIL DAN PEMBAHASAN

3.1 K-Nearest Neighbor

Pada Data Sampler, menggunakan proporsi data sebesar 10% dari total data seperti pada tabel 1. Pada bagian Predictions dengan menggunakan model kNN, menunjukkan bahwa AUC sebesar 51,9%, CA sebesar 86,6%, F1 sebesar 80,4%, Precision sebesar 75,0%, dan Recall sebesar 86,6%. Hasil prediksi dengan menggunakan model kNN ini sudah cukup baik dengan persentase keseluruhan yang berada di atas 50%, dan dengan total rata-rata persentase sebesar 76,1%.

Pada Confusion Matrix, dapat terlihat bahwa untuk data 'My listening of the tracks and seeing of the artist now' data yang benar sebanyak 84 data, dan tidak ada data yang salah, sehingga persentase keberhasilan

data ‘My listening of the tracks and seeing of the artist now’ sebesar 100,0% (84 data benar dari 84 data). Hasil Confusion Matrix, untuk data ‘My previous knowledge/familiarity with the tracks and the artist’ tidak ada data yang benar, sedangkan 13 data salah dan masuk ke data ‘My listening of the tracks and seeing of the artist now’, sehingga persentase keberhasilan data Enrolled sebesar 0,0% (0 data benar dari 13 data).

Tabel 1. Sampling Dataset

Participant ID	What is your age?	To which gender identity do you most identify? (If not listed, please specify in "Other")	According to the Fitzpatrick scale (https://en.wikipedia.org/wiki/Fitzpatrick_scale), How would you classify your skin type?	Where are you from?	What is the highest degree or Level of school you have completed?	Have you received any formal musical training (beyond the usual music lessons in school)?	Do you regularly engage in playing, Djing or producing any kind of music?	Would you describe Your musical taste as varied?	Do you often Listen to Electronic Music?	Within Electronic music, Would you describe your musical taste as varied?
1	25-34 years old	Male	5	South America	Master's degree (e.g. MA, MS, MEd)	Yes	4	5	4	3
2	35-44 years old	Female	2	Europe	Doctorate (e.g. PhD, EdD)	No	5	5	5	5
3	18-24 years old	Male	2	Europe	Master's degree (e.g. MA, MS, MEd)	Yes	5	4	5	4
4	45-54 years old	Male	3	Europe	Doctorate (e.g. PhD, EdD)	Yes	4	5	5	5
5	25-34 years old	Male	2		Bachelor's degree (e.g. BA, BS)	No	1	5	3	5
6	35-44 years old	Male	3	South America	Master's degree (e.g. MA, MS, MEd)	Yes	1	5	5	4
7	25-34 years old	Male	2		Bachelor's degree (e.g. BA, BS)	Yes	3	5	4	4
8	45-54 years old	Female	4	North America	Master's degree (e.g. MA, MS, MEd)	No	1	4	5	4
9	35-44 years old	Male	1	Europe	Doctorate (e.g. PhD, EdD)	No	4	5	4	5
10	25-34 years old	Male	4	South America	Bachelor's degree (e.g. BA, BS)	Yes	5	4	5	5
11	18-24 years old	Male	3	Europe	Bachelor's degree (e.g. BA, BS)	No	5	3	5	4
12	25-34 years old	Male	I prefer not to say	North America	High school degree or equivalent	No	4	4	3	2
13	18-24 years old	Male	1	Oceania	High school degree or equivalent	No	3	4	5	5
14	25-34 years old	Female	2	Europe	Master's degree (e.g. MA, MS, MEd)	No	1	4	4	4
15	25-34 years old	Female	2	Europe	Master's degree (e.g. MA, MS, MEd)	No	2	3	4	2
16	25-34 years old	Male	3	Europe	Master's degree (e.g. MA, MS, MEd)	Yes	5	5	5	5
17	35-44 years old	Female	I prefer not to say	Europe	Master's degree (e.g. MA, MS, MEd)	No	1	5	4	4
18	45-54 years old	Male	3		Master's degree (e.g. MA, MS, MEd)	No	1	5	3	4
19	35-44 years old	Male	1	Europe	Doctorate (e.g. PhD, EdD)	Yes	2	4	3	4
20	35-44 years old	Male	1	Europe	Doctorate (e.g. PhD, EdD)	Yes	4	3	2	3

Participant ID	What is your age?	To which gender identity do you most identify? (If not listed, please specify in "Other")	According to the Fitzpatrick scale (https://en.wikipedia.org/wiki/Fitzpatrick_scale), How would you classify your skin type?	Where are you from?	What is the highest degree or Level of school you have completed?	Have you received any formal musical training (beyond the usual music lessons in school)?	Do you regularly engage in playing, Djing or producing any kind of music?	Would you describe Your musical taste as varied?	Do you often Listen to Electronic Music?	Within Electronic music, Would you describe your musical taste as varied?
21	25-34 years old	Female	3	Europe	Master's degree (e.g. MA, MS, MEd)	Yes	1	4	4	4
22	18-24 years old	Male	1	Europe	Master's degree (e.g. MA, MS, MEd)	Yes	5	5	4	4

Pada bagian Data Table yang terhubung dengan Data Sampler, menggunakan Data Train sebanyak 97 dengan adanya missing data (data yang hilang) sebesar 2,5% seperti pada tabel 2.

Tabel 2. Predictions

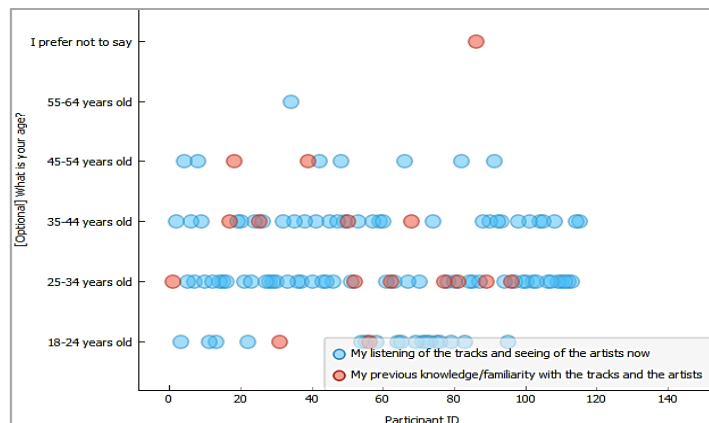
Investigating Influencers of Variation in Music Preference	According to the Fitzpatrick scale	Participant id	What is your age?	To which gender identity do you most identify?	Where are you from?	What is the highest degree or	Have you received any formal musical training	Do you regularly engage in playing
My listening of the tracks now	3	6	35-44 years old	Male	South America	Master's degree (e.g. MA, MS, MEd)	Yes	1
My listening of the tracks now	2	2	35-44 years old	Female	Europe	Doctorate (e.g. PhD, EdD)	No	5
My listening of the tracks now	I prefer not to say	58	18-24 years old	Male	North America	High school degree or equivalent	No	4
My listening of the tracks now	2	35	35-44 years old	Male	Europe	Master's degree (e.g. MA, MS, MEd)	Yes	3
My listening of the tracks now	3	113	25-34 years old	Male	Europe	Doctorate (e.g. PhD, EdD)	No	3
My listening of the tracks now	2	96	25-34 years old	Male	Europe	Master's degree (e.g. MA, MS, MEd)	Yes	5
My listening of the tracks now	3	67	25-34 years old	Male	South America	Doctorate (e.g. PhD, EdD)	Yes	3
My listening of the tracks now	2	14	25-34 years old	Female	Europe	Master's degree (e.g. MA, MS, MEd)	No	1
My listening of the tracks now	2	29	25-34 years old	Male	South America	Master's degree (e.g. MA, MS, MEd)	Yes	3
My listening of the tracks now	1	42	45-54 years old	Male	Europe	Master's degree (e.g. MA, MS, MEd)	Yes	3
My listening of the tracks now	1	70	25-34 years old	Male	Asia	Doctorate (e.g. PhD, EdD)	Yes	2
My listening of the tracks now	3	59	35-44 years old	Male	South America	Bachelor's degree (e.g. BA, BS)	No	5
My listening of the tracks now	I prefer not to say	55	18-24 years old	Male	Europe	Master's degree (e.g. MA, MS, MEd)	No	5
My listening of the tracks now	2	15	25-34 years old	Female	Europe	Master's degree (e.g. MA, MS, MEd)	No	2
My listening of the tracks now	I prefer not to say	102	25-34 years old	I prefer not to say	Europe	Master's degree (e.g. MA, MS, MEd)	Yes	5
My listening of the tracks now	3	50	35-44 years old	I prefer not to say	Europe	Bachelor's degree (e.g. BA, BS)	Yes	5
My listening of the tracks now	4	105	35-44 years old	I prefer not to say		Doctorate (e.g. PhD, EdD)	Yes	5
My listening of the tracks now	2	3	18-24 years old	Male	Europe	Master's degree (e.g. MA, MS, MEd)	Yes	5
My listening of the tracks now	2	31	18-24 years old	Male	Europe	Bachelor's degree (e.g. BA, BS)	Yes	5
My listening of the tracks now	2	87	25-34 years old	Male	Europe	Bachelor's degree (e.g. BA, BS)	Yes	5
My listening of the tracks now	3	115	35-44 years old	Male	Europe	Bachelor's degree (e.g. BA, BS)	Yes	2
My listening of the tracks now	2	76	18-24 years old	Male	Europe	High school degree or equivalent	Yes	5

Investigating Influencers of Variation in Music Preference	According to the Fitzpatrick scale	Participant id	What is your age?	To which gender identity do you most identify?	Where are you from?	What is the highest degree or	Have you received any formal musical training	Do you regularly engage in playing
My listening of the tracks now	1	92	35-44 years old	Male	Europe	High school degree or equivalent	No	1
My listening of the tracks now	1	111	25-34 years old	Female	Europe	Master's degree (e.g. MA, MS, MEd)	No	1
My listening of the tracks now	1	40	25-34 years old	Male	Oceania	Master's degree (e.g. MA, MS, MEd)	No	1

3.2 Naïve Bayes

Menggunakan data yang sama mencoba menggunakan Naive Bayes, untuk hasil Naive Bayes menunjukkan AUC 0.836, CA 0.496, F1 0.557, Precision 0.870, Recall 0.496. Artinya, tingkat presisi memprediksi listening of the tracks and seeing of the artist diatas 87 %.

Menurut prediksi dan data aktual 42 orang akan mendengarkan musik sesuai dengan listening of the tracks and seeing of the artist now atau trek dan artis masa kini (true positif) dan 15 orang akan mendengarkan dengan previous knowledge/familiarity with the tracks and the artis atau pengetahuan mengenai trek musik dan artisnya. Menurut prediksi sebanyak 57 orang akan mendengarkan sesuai dengan pengalaman, sedangkan menurut fakta akan mendengarkan sesuai trek terkini (false positif) dan menurut prediksi 1 orang akan mendengarkan musik sesuai dengan trek dan artis musik terkini, sedangkan menurut fakta akan mendengarkan musik dengan pengetahuan atau pengalaman sebelumnya mengenai trek (false negatif).



Gambar 2. Hasil Prediksi

Participan ID (X) dan What is your age (Y) mempengaruhi orang yang mendengarkan musik dengan pengetahuan sebelumnya tentang trek dan artis musik dengan titik tertinggi yaitu pada titik participad Id 86 dan Age prefer not to say, dan terendah pada participan Id 31, Age 18-24 years old. Sedangkan untuk orang yang mendengarkan musik dengan trek dan artis terkini atau sekarang titik tertingginya adalah participan Id 34 dan Age 55-64 years old dan terendah pada participan Id 3 dan Age 18-24 years old.

3.3 Decision Tree dan Ensemble Method

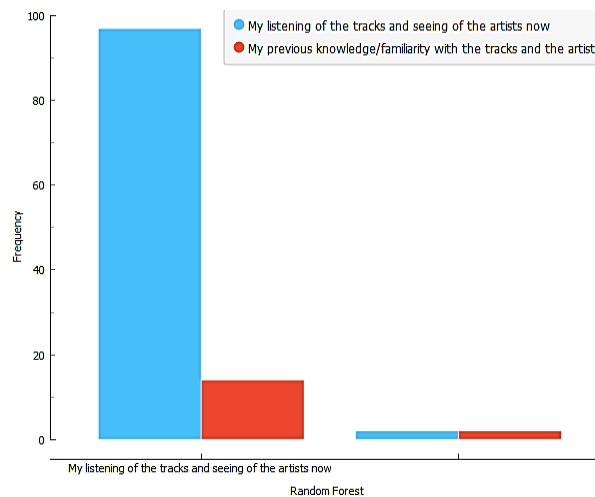
Dari 3 model algoritma yang ada Tree, Random Forest, dan AdaBoost maka model algoritma yang paling bagus dan baik adalah Random Forest karena dilihat dari AUC nya sebesar 0.601 atau 60%. Model algoritma AdaBoost memiliki hasil 0.597 atau 59% dan yang paling kecil adalah Model algoritma Tree yaitu 0,503 atau 50%.

Karena Random Forest memiliki model algoritma yang paling bagus, maka ditampilkan hasilnya untuk klasifikasi. Menurut prediksi dan data aktual 97 orang akan mendengarakan musik dengan trek dan artis terkini (true positif) dan 2 orang akan mendengarkan musik dengan pengetahuan sebelumnya mengenai trek dan artisnya (true negatif). Sedangkan 2 orang menurut data prediksi akan medengarkan musik dengan pengetahuan sebelumnya mengenai trek dan artis terkini nyatanya orang tersebut mendengarkan musik dengan trek dan artis terkini (false positif) dan 14 orang menurut prediksi mendengarkan musik dengan trek dan artis terkini sedangkan nyatanya orang tersebut mendengarkan musik melalui pengetahuan sebelumnya tentang trek dan artist tersebut.



Gambar 3. Hasil Scatter Plot

Participan ID (X) dan Familiaty Score (Y) mempengaruhi listening of the tracks and seeing of the artist now dengan titik tertinggi (90,0,76) dan terendah pada titik (20,0,06). Variabel Participan ID (X) dan Familiaty Score (Y) mempengaruhi previous knowledge/familiarity with the tracks and the artis dengan titik tertinggi (52,1.00) dan terendahnya pada (1,0,24).



Gambar 4. Hasil Variabel Random Forest

Variable Random Forest mempengaruhi cara mendengarkan musik dimana prediksi Random Forest untuk mendengarkan musik dengan listening of the tracks and seeing of the artist now sebanyak 97 orang dan melalui previous knowledge/familiarity with the tracks and the artis sebanyak 14 orang. Selain itu, prediksi Random Forest pada cara mendengarkan melalui previous knowledge/familiarity with the tracks and the artis sebanyak 2 orang dan 2 orang lagi melalui listening of the tracks and seeing of the artist now.

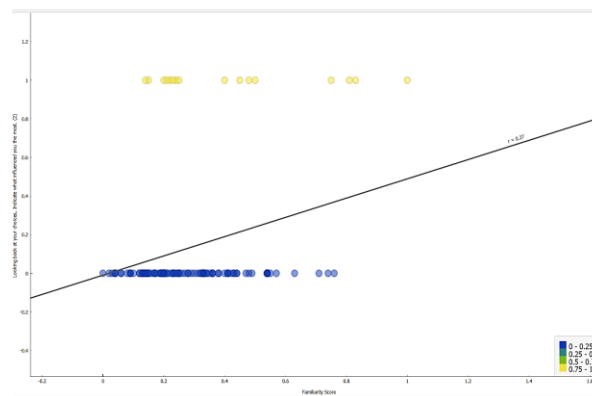
3.4 Linear Regression

Pada bagian select columns ini berisi untuk variabel apa yang bakal diuji dan juga targetnya apa. Pada bagian features dengan berbentuk numeric yaitu ada BOTH task 1, BOTH task 2, BOTH task 3, BOTH task 4, Familiarity Score, ARTIST Task 1, ARTIST Task 2, ARTIST Task 3, ARTIST Task 4, TRACK Task 1, TRACK Task 2, TRACK Task 3, TRACK Task 4, [Optional] Do you often Listen to Electronic Music?, [Optional] Within Electronic music, Would you describe your musical taste as varied?, Looking back at your choices, Which one did you weight the most: dengan targetnya yaitu Looking back at your choices, Indicate what influenced you the most (2).

Pada bagian rank ini untuk melakukan pengurutan pada variabel dimana yang nilainya paling tinggi berada diatas dan variabel tersebut adalah Familiarity Score dan TRACK Task 1 dengan targetnya yaitu Looking back at your choices, Indicate what influenced you the most (2). Pada bagian correlations juga untuk mengurutkan variabel mana yang tertinggi dan variabel tersebut sama seperti pada bagian rank yaitu Familiarity Score dan TRACK Task 1 dengan targetnya yaitu Looking back at your choices, Indicate what influenced you the most (2).

Pada bagian select columns yang baru, memilih 2 variabel saja yang dimana dari hasil paling tinggi saat pengecekan pada bagian rank dan correlations. mengambil variabel Familiarity Score dan TRACK Task 1 dengan targetnya yaitu Looking back at your choices, Indicate what influenced you the most (2).

Pada Bagian Test and Score, pada gambar diatas menunjukkan bahwa R2 yang dihasilkan memiliki nilai yang rendah. Ada model MSE, RMSE, MAE merupakan nilai-nilai untuk menghitung bagaimana tingkat kekuatan regression. Jadi kita menggunakan model R2 dimana korelasinya antara variabel Familiarity Score dan TRACK Task 1 dengan targetnya yaitu Looking back at your choices, Indicate what influenced you the most (2) sebesar 0,063 atau sebesar 6,3%. Nilai tersebut membuktikan bahwa sebagian besar variabilitas dalam data tidak dapat dijelaskan oleh model dan tidak menjadi efektif dalam memprediksi variabel target pada data testing.



Gambar 5. Scatter Plot

Dapat kita lihat pada gambar diatas yang dimana dibuka dari bagian Scatter plot dimana menunjukkan garis regresi dengan sumbu x Familiarity Score dan sumbu Y Looking back at your choices, Indicate what influenced you the most (2) sebagai target. Gambar diatas menunjukkan bahwa Familiarity score dapat mempengaruhi target karena hasilnya positif yaitu sebesar $r = 0,27$. Artinya semakin tinggi familiarity score maka akan mempengaruhi Looking back at your choices, Indicate what influenced you the most (2) juga semakin tinggi. Dengan begitu maka ada hubungan positif antara variabel independen (x) dan Variabel dependen (y) tapi tidak dapat menentukan sebab akibat atau hubungan kausal antara keduanya.

Berdasarkan penelitian pada dasarnya seseorang tidak hanya mendengarkan musik yang menjadi faktor utamanya, melainkan juga mengkonsumsi jenis-jenis musik lain, karena pendengar menyesuaikan musik yang di dengar sesuai dengan tren terkini, hanya beberapa yang berdasarkan pengalaman, sehingga musik yang mereka ketahui, berdasarkan dengan kondisi tren saat ini. Pemilihan musik untuk masyarakat global dipandang sebagai pengaruh media. Karena media massa yang saat ini dapat merambah ke berbagai penjuru dunia dan berbagai lapisan masyarakat dengan sangat mudah dan cepat merupakan wahana sosialisasi yang sangat efektif dan efisien untuk mengakrabkan musik dengan masyarakat.

4. KESIMPULAN

Berdasarkan dari hasil analisis dan pengujian data untuk memprediksi keanekaragaman preferensi mendengarkan musik yang menggunakan algoritma K-Nearest, Naive Bayes, Decision tree dan ensemble method serta linear regression maka dapat disimpulkan bahwa Algoritma Naive Bayes memiliki tingkat akurasi yang lebih baik dibandingkan algoritma lainnya dengan target Looking back at your choices, Indicate what influenced you the most (2). Rata rata nilai algoritma Naive Bayes dari lima algoritma lainnya yaitu sebesar 87%. kemudian K-Nearest memiliki nilai rata rata sebesar 76,1%. Diikuti dengan rata rata nilai decision tree dan ensemble method sebesar 56,7%, dan terakhir Linear regression yaitu 6,3% untuk nilai rata ratanya.

REFERENSI

- [1] <https://www.kaggle.com/datasets/thedevastator/music-listening-preferences-diversity>
- [2] A. W. Tawfik, H. Alhoori, C. W. Keene, C. Bailey and M. Hogan(2018). "Using a Recommendation System to Support Problem," Technology, Knowledge and Learning, vol. 23, no. 1, pp. 177-187, 2018
- [3] Abid As Sarofi, Muhammad, Irhamah, Mukarromah, Adatul. (2020). Identifikasi Genre Musik dengan Menggunakan Metode Random Forest, Jurnal Sains dan Seni ITS Vol. 9, No.1,
- [4] FH Pratama, A Triayudi, E Mardiani.(2022). Data Mining K-Medoids Dan K-Means Untuk Pengelompokan Potensi Produksi Kelapa Sawit di Indonesia. JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika) 7 (4), 1294-1310

-
- [5] Hozairi, Anwari, Syariful Alim (2021). IMPLEMENTASI ORANGE DATA MINING UNTUK KLASIFIKASI KELULUSAN MAHASISWA DENGAN MODEL K-NEAREST NEIGHBOR, DECISION TREE SERTA NAIVE BAYES. *Jurnal Ilmiah NERO* Vol. 6 No. 2
- [6] Indriyanti, Ichsan, Nurul., Fatah, Haerul., Wahyuni., Tri., Ermawati, Erni. Implementasi Orange Data Mining Untuk Prediksi Harga Bitcoin. *JURNAL RESPONSIF*, Vol. 4 No.2 Agustus 2022, pp. 118~125
- [7] Indriyawati, Henny, Khoirudin. 2019. PENERAPAN METODE REGRESI LINIER DALAM KOHERENSI PENGOLAHAN DATA BAHAN BAKU TIANDRA STORE GUNA MENINGKATKAN MUTU PRODUKSI. *Proceeding SINTAK 2019 Universitas Semarang* <https://www.unisbank.ac.id/ojs/index.php/sintak/article/view/7603>
- [8] Mardiani, Eri, Ferdan Akbar Ramadhan (2023). Design Information System Sales of Nuts and Bolts at PT. Catur Naga Steelindo. *Jurnal SITEKIN: Jurnal Sains, Teknologi dan Industri*, Vol. 20, No. 2, June 2023, pp.729 – 735
- [9] Mardiani, Eri, Nur Rahmansyah, Sari Ningsih, Dhieka Avrilia Lantana, Adisti Suryaningtyas Putri Wirawan, Sifonne Adi Wijaya, Dinda Nurkhaliza Putri. (2023) Komparasi Metode KNN, Naive Bayes. Decision Tree, Ensemble, Linear Regression Terhadap Analisis Performa Pelajar Sma. *Jurnal INNOVATIVE: Journal Of Social Science Research* Volume.3 Nomor.2 Tahun.2023 Page.13880-13892
- [10] Mardiani, Eri., Nur Rahmansyah, Ira Kurniati (2023). Website Design At SDN Cipete Utara 07. *Jurnal SITEKIN: Jurnal Sains, Teknologi dan Industri*, Vol. 20, No. 2, June 2023, pp.891 – 898
- [11] Matondang, Nurhafifah, Eri Mardiani, Wahyudi, Praptiningsih, Akhmad Saebani, 2019. *Aplikasi Komputer*, Jakarta: Mitra Wacana Media
- [12] MA Djamaludin, A Triayudi, E Mardiani. (2022) Analisis Sentimen Tweet KRI Nanggala 402 di Twitter menggunakan Metode Naïve Bayes Classifier. *Jurnal JTIK (Jurnal Teknologi Informasi dan Komunikasi)* 6 (2) 2022 pp 2580-1643
- [13] Rahmansyah, Nur., Mulyani, Deta., Mardiani, Eri., Rahman, Adityo. Perancangan Sistem Transaksi Berbasis Web Pada UKM Pangkas Rambut Tasik. *Jurnal JUNSIBI: Jurnal Sistem informasi Bisnis* April 2022, 22-31
- [14] G. A.V. Mastrika Giri. (2018). Klasifikasi Musik Berdasarkan Genre dengan Metode K-Nearest Neighbor. *Jurnal Universitas Udayana: Jurnal Ilmu Komputer* VOL. XI No. 2. <https://ojs.unud.ac.id/index.php/jik/article/view/42015/25852>
- [15] G. A.V. Mastrika Giri and I. G. A. G. Arya Kadyanan (2020), "Music Recommendation System using Case-Based Reasoning Method," *International Journal of Science and Research (IJSR)*, vol. 10, no. 1, pp. 1047 -1050, 2021.