

Komparasi Algoritma Klasifikasi *Genre* Musik pada Spotify Menggunakan CRISP-DM

Salma Navisa
Sistem Informasi, Fakultas Sains
dan Teknologi
UIN Sunan Ampel
Surabaya, Indonesia
salmanavisa17@gmail.com

Luqman Hakim
Sistem Informasi, Fakultas Sains
dan Teknologi
UIN Sunan Ampel
Surabaya, Indonesia
manluq@protonmail.com

Aulia Nabilah
Sistem Informasi, Fakultas Sains
dan Teknologi
UIN Sunan Ampel
Surabaya, Indonesia
aulianabila13@gmail.com

Abstract— Before the modernization era in Indonesia, if someone wanted to listen to music they had to be in a place or be able to listen and see from a device that could not be carried anywhere, for example, the radio. Makes a person unable to choose the music genre as desired. With the advent of the internet in Indonesia and the development of the times, people can listen to music online with their taste or preferred music genre or listen to music with music genres that are played randomly. The data mining process has a variety of classification algorithms that are always evolving. Classification techniques in data mining functionality are data that can be categorized based on previously known class labels. Classification techniques have been used in various fields including research in the music industry. This study is aimed at comparing which classification algorithm produces the best performance based on the data mining process using CRISP-DM. The classification algorithms used for testing are Naive Bayes, K-NN and Random Forest. From the research conducted, the best accuracy results were obtained by the Naive Bayes algorithm with a value of 58.91%. The best performing algorithms are K-NN and Random Forest with a value of 0.528.

Keywords—Data Mining, Genre Classification, K-NN, Music Genre, Naive Bayes, Random Forest.

I. INTRODUCTION

Aktivitas mendengarkan musik dapat didengarkan di berbagai tempat atau media, misal di radio, televisi, toko buku, tempat makan, toko minimarket, dan lain-lain. Jika ditilik kembali pada zaman sebelum terjadinya modernisasi, seseorang yang ingin mendengarkan musik harus berada di suatu lokasi atau mendengarkan dan melihat dari peranti yang tidak bisa dibawa ke mana-mana, Namun, saat ini manusia tidak perlu harus repot-repot pergi ke suatu tempat atau mendengarkan dari radio. Sejak hadirnya internet di Dunia, khususnya di Indonesia [1], membuat masyarakat bisa mencari segala informasi yang dibutuhkan di dunia maya menggunakan koneksi internet hingga mendengarkan musik secara daring melalui penyedia layanan mendengarkan musik menggunakan koneksi internet secara digital yang bisa dilakukan di mana saja, kapan saja, dan pastinya menggunakan peranti yang mudah dibawa.

Penyedia layanan untuk mendengarkan musik secara digital menggunakan koneksi internet saat ini terdiri dari banyak pilihan. Banyaknya penyedia layanan tersebut memberikan kemudahan dan pilihan kepada masyarakat untuk digunakan. Kegiatan mendengarkan musik seolah-olah sudah menjadi rutinitas yang tidak boleh terlewatkan atau lupa dilakukan.

Melakukan aktivitas mendengarkan musik saat ini sudah sangat fleksibel yang dapat dilakukan di banyak peranti, misal seperti: ponsel cerdas; laptop; komputer tablet portabel; mau pun komputer pribadi. Kumpulan musik dikumpulkan menjadi satu yang selanjutnya akan disimpan dan dikelola pada pangkalan data (*database*) milik penyedia layanan untuk dilakukan proses pemilahan lebih lanjut seperti pengurutan berdasarkan nama artis/penyanyi, jenis album/mini album (ep)/single, tahun rilis, pemberian label *genre* musik [2], hingga daftar putar (*playlist*) musik yang dibuat oleh penyedia layanan. Contoh

salah satu ciri-ciri untuk menentukan suatu musik cocok dalam suatu *genre* yaitu musik yang memiliki tempo yang pelan sampai yang cepat.

Pengertian *genre* musik secara sederhana yaitu pengkategorian, tipe, atau jenis dari suatu musik. Pelabelan dalam suatu musik yang dilakukan oleh penyedia layanan digunakan untuk memudahkan pengelompokan atau pengklasifikasian suatu musik yang akan disajikan dalam layanan penyedia mendengarkan musik digital berbasis website atau aplikasi di ponsel cerdas. *Genre* musik ada karena terjadinya interaksi yang kompleks antara masyarakat, pihak pemasaran, sejarah, dan faktor budaya lokal [3]. Mendengarkan musik, khususnya berjenis klasik, penggunaannya tidak hanya diketahui sebagai terapi untuk mengurangi kadar stres [4], tetapi bisa juga digunakan untuk membuat progres pembelajaran lebih mudah fokus, lebih paham, dan antusias, menghilangkan kebosanan [5], menjadi lebih bersemangat, mengurangi rasa kesedihan, mengurangi rasa marah, serta mengurangi rasa takut dan cemas [6], [7].

Genre yang diterima oleh setiap pendengar dari peranti yang dimiliki akan mendapatkan tampilan rekomendasi yang berbeda, hal itu dikarenakan terjadinya proses pengklasifikasian berdasarkan perilaku pendengar kepada musik yang sering, jarang, bahkan tidak pernah didengarkan atau diputar sama sekali oleh pendengar. Pada penelitian terdahulu [8] menggunakan algoritma genetika untuk meningkatkan kinerja dari klasifikasi *genre* musik berbasis algoritma *back-propagation* mendapatkan peningkatan hasil akurasi klasifikasi *genre* menjadi 85.55% dengan parameter generasi 100, individu 50, crossover 0.6, dan peluang permutasi 0.01.

Pada penelitian ini, untuk melakukan klasifikasi *genre* musik pada Spotify peneliti akan menggunakan perbandingan hasil kinerja berupa akurasi dari tiga algoritma yaitu K-Nearest Neighbour, Naive Bayes, dan Random Forest. Penelitian menunjukkan hasil dari proses data mining mana yang memiliki kinerja yang paling baik diantara ketiga algoritma yang telah disebutkan sebelumnya menggunakan pola penelitian kerangka kerja CRISP-DM untuk data mining.

II. METHOD/MATERIAL

A. CRISP-DM

Penelitian klasifikasi *genre* musik pada Spotify disesuaikan dengan tahapan metode CRISP-DM menggunakan dataset “Songs in Spotify” [9] dari situs Kaggle yang dikumpulkan melalui Spotify API. Dilihat dari sejarah terkumpulnya, pengumpulan data dilakukan pada tanggal 6 Desember 2020 dengan total 42.305 data. Dataset dibagi menjadi *data training* sebesar 90% dan *data testing* sebesar 10% untuk mengetahui keakuratan antara model dengan data.

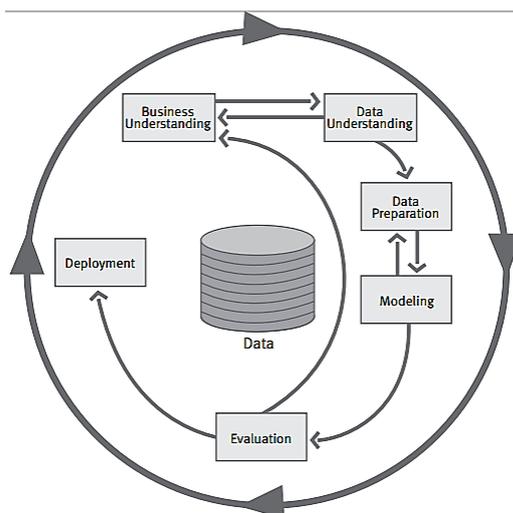


Figure 1. Tahapan Metode CRISP-DM

Sederhananya, metode CRISP-DM (Cross-Industry Standard Process for Data Mining) bertujuan untuk memberikan *blueprint* berupa tahapan dalam proses pengumpulan data. Tahapan tersebut dibagi menjadi enam yaitu *business understanding*, *data understanding*, *data preparation*, *modeling*, *evaluation*, dan *deployment* [10]. Figure 1 memaparkan tahapan proses data mining bekerja. Setiap arah panah menunjukkan siklus dan menggambarkan bahwa data mining dapat menghasilkan solusi, pengetahuan, dan pemahaman bisnis. Tahap awal, *business understanding* berfokus pada pemahaman tujuan penelitian berdasarkan perspektif bisnis yaitu mengklasifikasikan musik berdasarkan *genre*.

Pada tahap *data understanding* dimulai dengan pengumpulan data awal kemudian melakukan identifikasi dan eksplorasi data. Tahap ketiga, *data preparation* mencakup aktivitas untuk menyusun data mentah agar dapat digunakan dalam pemodelan dengan melakukan pemilahan, pembersihan, penyusunan, dan pengintegrasian data. Tahap *modelling*, berbagai teknik atau algoritma diterapkan berdasarkan jenis data yang dipilih. Pemodelan meliputi pemilihan teknik, pembuatan, dan penilaian model [11].

Peneliti menggunakan alat bantu perangkat lunak RapidMiner Studio versi 9.8 Educational Edition untuk preparation dan modelling data. Proses pemodelan data mining menggunakan 3 macam algoritma klasifikasi supervised learning yaitu K-Nearest Neighbor, Naive Bayes, dan Random Forest. Tahap Evaluasi dilakukan untuk mengukur kinerja model menggunakan operator *performance*. Proses evaluasi menghasilkan nilai *Accuracy* dan *Kappa Statistic* karena dataset yang digunakan bersifat klasifikasi polinomial yaitu variabel yang memiliki atribut lebih dari tiga [12]. Tahap terakhir, *deployment* yaitu menghasilkan laporan penerapan proses data mining dan langkah apa yang harus diambil selanjutnya [11].

B. K-Nearest Neighbour

K-Nearest Neighbour (K-NN) adalah metode pembelajaran mesin standar berbasis instans yang berisi kumpulan data skala besar [13]. Meskipun sederhana, K-NN sering digunakan untuk klasifikasi, dan juga dapat digunakan untuk estimasi dan prediksi. Dalam proses klasifikasi, K-NN memilih label yang paling

sering muncul dengan aturan menetapkan nilai x ke populasi yang memiliki titik data paling banyak diamati dari k -tetangga terdekat [14]. K-NN menetapkan nilai prediksi berdasarkan nilai k tetangga terdekat (yaitu, paling “mirip” dengannya) dari setiap titik *query* dengan menghubungkan karakteristik pengukuran [15]. Pengklasifikasian data dilakukan dengan menentukan jumlah tetangga, dari nilai k yang dipilih dengan *cross-validation* atau menguji kualitas klasifikasi pada data testing. *Cross-validation* dipilih karena proses klasifikasi beroperasi dengan perangkat lunak Rapidminer. Serta menggunakan matrik *euclidean distance* untuk mengukur besar jarak tiap *query* [14].

$$d_{Euclidean}(x, y) = \sqrt{\sum_i (x_i - y_i)^2}$$

Euclidean distance ditunjukkan dengan persamaan (1) dengan $x = x_1, x_2, \dots, x_i$, dan $y = y_1, y_2, \dots, y_i$ mewakili nilai atribut i dari dataset training dan testing. Sementara d didefinisikan sebagai jarak [13].

C. Naive Bayes

Algoritma Naive Bayes dapat digunakan sebagai pengklasifikasian karena perhitungan probabilistik yang sederhana dan memiliki performa yang baik. Setiap dokumen berisi istilah probabilitas berdasarkan jumlah kemunculan data. Naive bayes dapat mempelajari pola sekumpulan data kemudian dibandingkan dengan daftar kata baru. Kata tersebut dapat digunakan untuk klasifikasi data baru ke dalam kategori yang tepat sesuai probabilitas [16].

$$P(x_i | y) = \frac{1}{\sqrt{2\pi\sigma_y^2}} \exp\left(-\frac{(x_i - \mu_y)^2}{2\sigma_y^2}\right) \quad (2)$$

Naive bayes pada persamaan (2) menggunakan dua asumsi bahwa dengan label kelas, atribut tidak bergantung berdasarkan kondisi, dan tidak ada atribut asli yang mempengaruhi prediksi label. Gaussian NB mendukung penelitian dengan asumsi x_i, \dots, x_n mewakili n atribut dari instance x , label kelas. Mean serta varian data diwakili oleh μ_y dan σ_y^2 masing-masing diberi label kelas y . Oleh sebab itu probabilitas pengamatan dihitung dengan persamaan distribusi normal Bayesian [17].

D. Random Forest

Algoritma Random Forest dapat digunakan untuk klasifikasi dan regresi. Random Forest merupakan pengklasifikasian yang terdiri dari kumpulan atau kombinasi pengklasifikasi pohon terstruktur. Random forest dapat menghasilkan suatu klasifikasi yang dimulai dari input vector yang kemudian menuruni banyak pohon-pohon. Pohon-pohon tersebut merupakan hasil klasifikasi dari suara terbanyak yang menandakan class tersebut. Forest menghasilkan hasil yang kompetitif dengan peningkatan dan pengemasan yang adaptif, tetapi tidak secara progresif mengubah *training set*. Akurasi yang ditunjukkan adalah untuk mengurangi bias [18].

(3)

$$Gini = N_L \sum_k PkL (1 - PkL) + N_R \sum_k PkR (1 - PkR)$$

Pengklasifikasian random forest menggunakan indeks Gini sebagai pengukuran atribut yang dipilih sehubungan dengan kelas acak [19]. Kriteria indeks gini berfokus pada meminimalkan kemungkinan kesalahan klasifikasi. Indeks gini dapat ditulis sesuai rumus pada persamaan 3 dimana $k \in (1, \dots, K)$ mewakili kategori peristiwa, PkL menunjukkan proporsi kelas k di simpul kiri dan PkR sama dengan proporsi kelas k di simpul kanan [20].

E. Accuracy & Kappa Statistic

Accuracy dan *kappa statistic* merupakan hasil dari klasifikasi yang memiliki label dengan tipe atribut *polynomial*. Penjelasan dari *accuracy* adalah angka yang menunjukkan seberapa dekat nilai yang terukur (*measured*) dengan nilai sebenarnya (*reality*). Fungsi dari *accuracy* ialah untuk mengungkapkan kebenaran suatu pengukuran dan ditentukan secara absolut dan komparatif [21]. Perhitungan *accuracy* dihitung dengan mengambil persentase prediksi yang benar yang di atas jumlah dengan rincian sebagai berikut:

$$Accuracy = \frac{Sum\ of\ true\ positives + Sum\ of\ true\ negatives}{Total\ population} \quad (4)$$

Kappa Statistic adalah metrik yang tujuannya untuk membandingkan antara akurasi penelitian dengan akurasi yang diharapkan (*random accuracy*). *Kappa Statistic* ditujukan untuk klasifikasi [21]. Pada umumnya *kappa* dianggap sebagai ukuran yang lebih kuat daripada hitungan prediksi benar, karena perhitungannya yaitu dengan memperhitungkan prediksi benar yang terjadi secara keseluruhan. Perhitungan mencari nilai *Kappa* dapat ditulis sebagai berikut:

$$Kappa = \frac{Accuracy - Random\ Accuracy}{1 - Random\ Accuracy} \quad (5)$$

F. Cross Validation

Cross validation digunakan untuk validasi atau evaluasi kinerja model, dan berfungsi sebagai pemilihan rasio untuk memisahkan data pengujian dan pelatihan [22]. *Cross validation* dapat menghasilkan estimasi akurasi dari proses *dataset*. Penelitian ini menggunakan *10-fold cross validation*, supaya menghasilkan estimasi akurasi yang baik atau kurang bias. Cara kerja dari *k-fold cross validation* adalah dengan model dilatih oleh subset data *training*, dan divalidasi oleh subset data *testing*. Dengan *10-fold cross validation* data dibagi menjadi 10 lipatan (*fold*) dengan ukuran yang sama, sehingga memiliki 10 subset data untuk dilakukan evaluasi kinerja model. *10-fold cross validation* melakukan perulangan pengujian sebanyak 10 kali [23].

Validation	Dataset									
1	■									
2		■								
3			■							
4				■						
5					■					
6						■				
7							■			
8								■		
9									■	
10										■

Figure 2. 10-Fold Cross Validation

III. RESULTS AND DISCUSSION

Pada penelitian klasifikasi genre musik pada Spotify, menggunakan proses *Cross Industry Standard Process for Data Mining* (CRISP-DM). CRISP-DM merupakan salah satu standar untuk proses data mining yang terdapat beberapa fase dari pemahaman bisnis hingga evaluasi. Dengan adanya beberapa fase tersebut diharapkan dapat memperoleh pemodelan yang sesuai saat proses data mining melalui pemahaman data yang telah disiapkan sehingga menghasilkan informasi yang dituju. Fase-fase CRISP-DM pada penelitian ini dijelaskan sebagai berikut:

A. Fase Pemahaman Bisnis (*Business Understanding*)

Fase ini menyebutkan tujuan serta kebutuhan dalam lingkup penelitian yang dilakukan secara detail. Kemudian menerjemahkan tujuan dan batasan menjadi rumus dari permasalahan data mining. Pemahaman bisnis pada kasus klasifikasi genre musik pada Spotify adalah:

1. Kumpulan ribuan musik di Spotify dengan berbagai atribut skala angka karakteristik musik.
2. Terdapat beberapa faktor penentu yang berupa atribut pada dataset untuk proses implementasi data mining.
3. Terdapat sejumlah 42.306 musik pada dataset untuk diklasifikasikan genrenya.
4. Melakukan klasifikasi dari atribut analisis audio, yang akan memudahkan pendengar untuk memilih lagu yang ingin diputar.

B. Fase Pemahaman Data (*Data Understanding*)

Pada fase ini, diperlukan pengambilan atau pengumpulan data terlebih dahulu. Pengumpulan data pada penelitian ini berupa mengambil data pada situs dan *platform* yang terdiri dari kumpulan dataset, kemudian dilakukan analisis data untuk mengenali variabel sebelum melakukan pengolahan data secara menyeluruh. Data yang dikumpulkan diambil dari penggunaan musik pada Spotify dengan atribut sebagai berikut Tabel 1.

Tabel 1.
Tabel Atribut Genre Musik

Atribut	Deskripsi
id	id musik
type	atribut tipe
url	atribut <i>link</i>
song_name	judul musik
genre	jenis genre musik
time_signature	atribut beat dalam satu bar
duration_ms	durasi musik
danceability	atribut karakteristik lagu
energy	atribut karakteristik lagu
key	atribut kunci lagu
loudness	atribut karakteristik lagu
mode	atribut mode lagu
speechiness	atribut karakteristik lagu
acousticness	atribut karakteristik lagu
instrumentalness	atribut karakteristik lagu
liveness	atribut karakteristik lagu
valence	atribut karakteristik lagu
tempo	atribut karakteristik lagu

Dataset Song of Spotify degan total 42.306 data, memiliki 10 jenis genre diantaranya underground, dark trap, hiphop, trance, trap, tech house, dnb, psytrance, techno, dan hardstyle. Underground masuk dalam genre dengan lagu terbanyak sebesar 5,873 lagu. Sementara genre yang memperoleh lagu paling sedikit sebesar 2.936 data diduduki genre hardstyle.

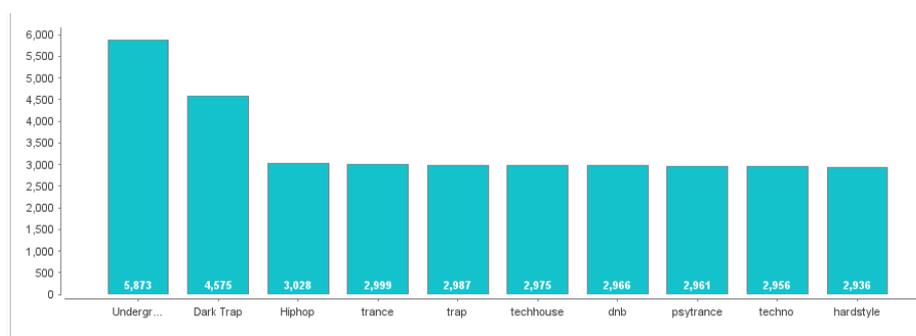


Figure 3. Total lagu berdasarkan genre

C. Fase Persiapan Data (*Data Preparation*)

Fase menyiapkan data awal, mengumpulkan data yang akan digunakan untuk keseluruhan fase-fase berikutnya (*data selection*). Memilih variabel untuk kasus yang ingin dianalisis. Melakukan perubahan pada variabel jika dibutuhkan.

Menyiapkan data awal untuk masuk pada fase *modeling*. Pada penelitian ini *dataset* yang digunakan adalah “*dataset of song in Spotify*”.

Pada fase *modeling*, data yang kurang lengkap atau *incomplete* dilakukan penambahan operator *declare missing value* dan *replace missing value*. Fungsi dari kedua operator tersebut adalah untuk menangani data yang terlewat saat penelitian dilakukan. Kesalahan tersebut dapat terjadi saat *entry data*, kesalahan instrumen pengumpulan data, penamaan yang tidak konsisten dan beberapa kesalahan lainnya. *Dataset* yang digunakan untuk klasifikasi genre musik terdapat beberapa atribut yang mengalami *incomplete data* sehingga menghasilkan *missing value*. Data tersebut terdiri dari atribut *song_name* dan *mode*.



Figure 4. Fase Persiapan Data

D. Fase Pemodelan (*Modeling*)

Memilih teknik pemodelan yang sesuai dengan kasus pada penelitian. Kemudian melakukan kalibrasi aturan model untuk hasil yang optimal. Teknik dapat digunakan pada permasalahan data mining yang sama. Selanjutnya dapat kembali ke fase *data preparation* jika perlu untuk menjadikan data sesuai kebutuhan. Fase pemodelan pada penelitian ini menggunakan salah satu teknik data mining yaitu klasifikasi dengan menggunakan 3 algoritma (Naive Bayes, K-NN, dan Random Forest).

Pada fase ini, dilakukan pemrosesan data *training* sesuai tujuan dan menggunakan teknik pemodelan naive bayes, k-nearest neighbour, dan random forest. Untuk menghasilkan keputusan dari klasifikasi *genre* musik pada Spotify. Proses pemodelan klasifikasi *genre* musik seperti Figure 8.

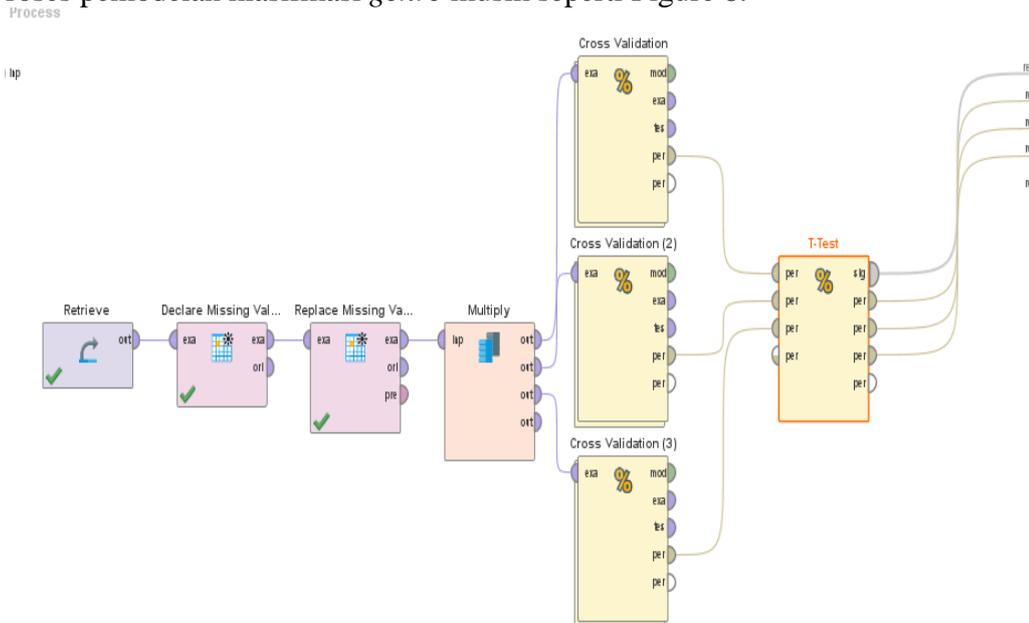


Figure 5. Proses *Modeling* Klasifikasi Genre Musik

Tabel 2.

Tabel Hasil Pengukuran 3 Algoritma Klasifikasi

Performance	Classification		
	Naive Bayes	K-NN	Random Forest
Accuracy	58.91%	33.30%	52.81%
Kappa	0.554	0.274	0.475

E. Fase Evaluasi (*Evaluation*)

Melakukan evaluasi pada satu model atau lebih yang digunakan pada fase pemodelan, agar mendapat efektivitas sebelum menerapkannya di lapangan. Mengoreksi sekaligus menetapkan apakah model yang digunakan telah sesuai dengan kebutuhan kasus [24]. Kemudian menentukan apakah ada permasalahan yang belum ditangan, selanjutnya mengambil keputusan yang terkait dengan hasil dari proses data mining. Setelah melakukan proses olah data pada dataset dengan beberapa teknik pemodelan, dilakukan proses evaluasi yaitu pengujian dari hasil proses data mining.

Pengujian dilakukan untuk menghasilkan akurasi terhadap proses klasifikasi menggunakan algoritma naive bayes, k-nearest neighbour, dan random forest. Nilai yang lebih kecil dari $\alpha = 0,050$ menunjukkan kemungkinan perbedaan yang signifikan antara nilai mean.

Tabel 3.

Tabel Komparasi 3 Algoritma Klasifikasi dengan T-Test

A	B	C	D
	0.333 +/- 0.007	0.528 +/- 0.009	0.528 +/- 0.009
0.333 +/- 0.007		0.000	0.000
0.589 +/- 0.007			0.000
0.528 +/- 0.009			

Keterangan:

B: Naive Bayes

C: K-NN

D: Random Forest

F. Fase Penyebaran (*Deployment*)

Fase terakhir dari crisp-dm yang mempresentasikan hasil dari model yang telah digunakan pada proses data mining. Penyajian dari hasil proses, mulai dari pengetahuan yang didapat selama proses sehingga dapat dipahami oleh pengguna. Fase pembuatan laporan hasil setelah melakukan evaluasi dari penelitian terhadap model klasifikasi genre musik.

IV. CONCLUSION

Berdasarkan hasil dari penelitian dan pembahasan terkait komparasi klasifikasi genre musik menggunakan *Cross Industry Standard Process for Data Mining* (CRISP-DM) dapat disimpulkan bahwa:

1. Klasifikasi genre musik pada *Spotify* menggunakan algoritma klasifikasi yaitu Naive Bayes, K-NN dan Random Forest. Validasi menggunakan *10-fold cross validation* dan uji menggunakan *t-test*. Performa dari ketiga algoritma kemudian dibandingkan, mana algoritma yang memiliki hasil performa yang baik.
2. Pengujian dilakukan dengan dataset “*dataset of song in Spotify*”, yang terdiri dari 42.305 data.
3. Hasil tertinggi nilai akurasi didapatkan oleh algoritma klasifikasi Naive Bayes yaitu sebesar 58.91%

Melakukan evaluasi dengan uji *t-test* algoritma dengan performa terbaik diperoleh dari algoritma K-NN dan Random Forest dengan nilai 0.528, sedangkan untuk algoritma Naive Bayes algoritma dengan nilai performa yang kurang baik.

V. ACKNOWLEDGEMENTS

Penulis berterima kasih kepada Andriy Samoshyn selaku pembuat datasets yang telah membuat kumpulan lagu dari penyedia layanan mendengarkan musik digital secara daring, Spotify dengan menggunakan Spotify API. Selain itu, terima kasih kepada tim peneliti yang telah bekerjasama dan memberikan kontribusi dalam beberapa pekerjaan yang dipaparkan dalam penelitian ini.

VI. REFERENCES

- [1] Admin, “Sejak Kapan Masyarakat Indonesia Nikmati Internet?,” *STEI*, Jun. 19, 2017. <https://stei.itb.ac.id/id/blog/2017/06/19/sejak-kapan-masyarakat-indonesia-nikmati-internet/> (accessed Jan. 26, 2021).
- [2] G. A. V. M. Giri, “Klasifikasi Musik Berdasarkan Genre dengan Metode K-Nearest Neighbor,” *Jurnal Ilmu Komputer*, vol. 11, no. 2, Art. no. 2, Oct. 2018, doi: <https://doi.org/10.24843/jik.2018.v11.i02.p05>.
- [3] P.-N. Tan, M. Steinbach, and V. Kumar, *Introduction to Data Mining*, 1st ed. Boston: Pearson Addison Wesley, 2006.
- [4] D. Kurnianingsih, J. Suroso, and A. Muhajirin, “EFEKTIFITAS TERAPI MUSIK KLASIK TERHADAP PENURUNAN STRES KERJA PERAWAT IGD DI RSUD DR. R. GOETHENG TAROENADIBRATA PURBALINGGA TAHUN 2013,” in *Seminar Nasional & Internasional*, Semarang, Indonesia, May 2013, pp. 166–172, Accessed: Jan. 26, 2021. [Online]. Available: <https://jurnal.unimus.ac.id/index.php/psn12012010/article/view/870>.
- [5] A. Roffiq, I. Qiram, and G. Rubiono, “MEDIA MUSIK DAN LAGU PADA PROSES PEMBELAJARAN,” *JPDI (Jurnal Pendidikan Dasar Indonesia)*, vol. 2, no. 2, pp. 35–40, Dec. 2017, doi: 10.26737/jpdi.v2i2.330.
- [6] E. Labbé, N. Schmidt, J. Babin, and M. Pharr, “Coping with Stress: The Effectiveness of Different Types of Music,” *Appl Psychophysiol Biofeedback*, vol. 32, no. 3, pp. 163–168, Dec. 2007, doi: 10.1007/s10484-007-9043-9.
- [7] S. Saarikallio and J. Erkkilä, “The role of music in adolescents’ mood regulation,” *Psychology of Music*, vol. 35, no. 1, p. 88–109, 2007, doi: <https://doi.org/10.1177/0305735607068889>.

- [8] A. Ramadhiat, I. I. Tritoasmoro, and I. Wijayanto, "Analisis Penggunaan Algoritma Genetika Untuk Meningkatkan Performansi Dari Klasifikasi Genre Musik Berbasis Jaringan Syaraf Tiruan Back-propagation," *eProceedings of Engineering*, vol. 3, no. 2, Art. no. 2, Aug. 2016, Accessed: Jan. 26, 2021. [Online]. Available: <https://openlibrarypublications.telkomuniversity.ac.id/index.php/engineering/article/view/302>.
- [9] *Dataset of songs in Spotify*, Kaggle, Dec. 06, 2020, [Online]. Available: <https://kaggle.com/mrmorj/dataset-of-songs-in-spotify>.
- [10] P. Chapman et al., "CRISP-DM 1.0: Step-by-step data mining guide." The CRISP-DM Consortium, Aug. 2000, [Online]. Available: <https://the-modeling-agency.com/crisp-dm.pdf>.
- [11] C. Shearer, "The CRISP-DM Model: The New Blueprint for Data Mining," *Journal of Data Warehousing*, vol. 5, no. 4, pp. 13–23, 2000.
- [12] RapidMiner Studio, "Performance - RapidMiner Documentation," Dec. 18, 2014. <https://docs.rapidminer.com/latest/studio/operators/validation/performance/performance.html> (accessed Jan. 27, 2021).
- [13] D. T. Larose, *Discovering knowledge in data: an introduction to data mining*. Hoboken, N.J: Wiley-Interscience, 2005.
- [14] R. C. Neath and M. S. Johnson, "Discrimination and Classification," in *International Encyclopedia of Education (Third Edition)*, P. Peterson, E. Baker, and B. McGaw, Eds. Oxford: Elsevier, 2010, pp. 135–141.
- [15] J. S. Richman, "Multivariate Neighborhood Sample Entropy: A Method for Data Reduction and Prediction of Complex Data," in *Methods in Enzymology*, vol. 487, Elsevier, 2011, pp. 397–408.
- [16] S. L. Ting, W. H. Ip, and A. H. C. Tsang, "Is Naïve bayes a good classifier for document classification?," *International Journal of Software Engineering and its Applications*, vol. 5, no. 3, pp. 37–46, 2011.
- [17] A. H. Jahromi and M. Taheri, "A non-parametric mixture of Gaussian naive Bayes classifiers based on local independent features," in *International Symposium on Artificial Intelligence and Signal Processing (AISP)*, Shiraz, Iran, Oct. 2017, pp. 209–212, doi: <https://doi.org/10.1109/AISP.2017.8324083>.
- [18] L. Breiman, "Random Forest," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001, doi: <https://doi.org/10.1023/A:1010933404324>.
- [19] M. Pal, "Random forest classifier for remote sensing classification," *International Journal of Remote Sensing*, vol. 26, no. 1, pp. 217–222, Jan. 2005, doi: <https://doi.org/10.1080/01431160412331269698>.
- [20] X. Xia, R. Togneri, F. Sohel, and D. Huang, "Random forest classification based acoustic event detection," in *IEEE International Conference on Multimedia and Expo (ICME)*, Hong Kong, Jul. 2017, pp. 163–168, doi: <https://doi.org/10.1109/ICME.2017.8019452>.
- [21] Dr. J. Arunadevi, S. Ramya, and M. R. Raja, "A study of classification algorithms using Rapidminer," *International Journal of Pure and Applied Mathematics*, vol. 119, no. 12, pp. 15977–15988, 2018.
- [22] S. Yadav and S. Shukla, "Analysis of k-Fold Cross-Validation over Hold-Out Validation on Colossal Datasets for Quality Classification," in *IEEE*

International Advance Computing Conference, IACC, Bhimavaram, India, Feb. 2016, pp. 78–83, doi: <https://doi.org/10.1109/IACC.2016.25>.

- [23] A. Ardiyansyah, P. A. Rahayuningsih, and R. Maulana, “Analisis Perbandingan Algoritma Klasifikasi Data Mining Untuk Dataset Blogger Dengan Rapid Miner,” *Jurnal Khatulistiwa Informatika*, vol. 6, no. 1, Art. no. 1, 2018, doi: <https://doi.org/10.31294/jki.v6i1.3799>.
- [24] S. E. Damayanti and S. Kuswayati, “ANALISIS DAN IMPLEMENTASI FRAMEWORK CRISP-DM (CROSS INDUSTRY STANDARD PROCESS FOR DATA MINING) UNTUK CLUSTERING PERGURUAN TINGGI SWASTA,” *E-Journal STT Bandung*, vol. 6, pp. 1–6, Dec. 2018.