

Penerapan Metode *Content Based Filtering* Dan *K-Nearest Neighbor* Dalam Sistem Rekomendasi Musik

I Made Teja Sarmandana^{a1}, I Made Widiartha^{a2}, Luh Arida Ayu Rahning Putri^{a3}, I Gede Santi Astawa^{a4}

^aInformatika, Universitas Udayana
Bali, Indonesia
¹tejamade01@gmail.com
²madewidiartha@unud.ac.id
³rahningputri@unud.ac.id
⁴santi.astawa@unud.ac.id

Abstract

Current technological developments are able to change the way the younger generation enjoys music, where music can now be packaged in digital form, which is a new innovation in the music industry in Indonesia. Given the large amount of music data available on the internet, a system that provides services for users to search for their favorite music is really needed. The recommendation system will provide relevant information based on the preferences that the user wants to search for. *Content Based Filtering* recommends that users utilize the information contained in the data to use as parameters. The *K-Nearest Neighbor (K-NN)* algorithm is a method of classifying objects based on the closest training data to the object under test. In this study, accuracy testing techniques were used to measure the performance of the classification that has been carried out. The classification process that was created succeeded in obtaining the highest accuracy value at 90,49% with a value of $k=9$ which shows that the classification and recommendation process can run quite well.

Keywords: Accuracy, Content Based Filtering, Classification, Recommendation System, K-Nearest Neighbor, Music

1. Pendahuluan

Masyarakat di penjuru dunia sangat dekat dengan yang namanya musik. Musik mampu menjadi peneman kita dalam melaksanakan aktivitas sehari-hari serta menjadi media yang mampu membawa kita ke perasaan senang, sedih, haru, dan beragam emosi lainnya. Teknologi saat ini berkembang hingga mampu mengubah cara generasi muda dalam menikmati musik dimana saat ini musik sudah mampu dikemas dalam bentuk digital yang menjadi inovasi baru di dalam industri musik di Indonesia [1]. Masyarakat saat ini dapat mendengarkan musik dimanapun kapanpun secara *online* karena telah tersedia banyak media layanan *streaming* musik secara *online*. Indonesia dipandang mampu menjadi pasar yang potensial dengan 220 juta lebih penduduknya. Hal tersebut dibuktikan dengan survey yang menunjukkan sekitar 46,9 juta atau 35,5% dari populasi pengguna internet di Indonesia menikmati musik secara *online* [2]. Ada pula survey yang dilakukan oleh *DailySocial* kepada para responden yang menunjukkan 85 persen responden menyatakan bahwa mereka sering menikmati musik secara *online* dalam waktu kurang lebih 6 bulan terakhir dimana sebagian besar pengguna layanan *streaming* musik tersebut berusia 16 sampai 24 tahun yang termasuk dalam kategori milenial atau "Generasi Z" [3]. Hal tersebut mendorong aplikasi-aplikasi penyedia layanan *streaming* musik seperti *Spotify*, *Joox*, *Youtube Music*, dan masih banyak lagi agar meningkatkan *user experience* untuk menarik *user-user* baru supaya ikut menggunakan aplikasi mereka serta menambah nilai kenyamanan dalam menggunakan aplikasi tersebut.

Mengingat banyaknya data musik yang tersedia di Internet, sistem yang menyediakan layanan bagi *user* untuk mencari musik favoritnya sangat dibutuhkan. Sistem rekomendasi menjadi jawaban untuk menghindari masalah kelebihan informasi tersebut [4]. Sistem rekomendasi merupakan sistem yang memperkirakan beberapa item atau data yang nantinya akan direkomendasikan dari yang teratas. Sistem rekomendasi akan menyajikan informasi yang sesuai dengan preferensi yang dicari oleh pengguna[5]. Sistem rekomendasi dirasa akan berguna menjadi pemecahan masalah dalam penelitian ini yang dimana menggunakan metode *Content-Based Filtering*. *Content-based filtering* merekomendasikan pengguna dengan memanfaatkan informasi yang terdapat pada data untuk

digunakan sebagai parameternya. Metode ini memilih dan memberi peringkat item berdasarkan kesamaan atribut itemnya. Keuntungan dari metode ini adalah bahwa informasi dalam setiap bagian data diketahui dalam representasi, sehingga pengguna bisa mendapatkan informasi tentang informasi yang dianggap relevan baginya [6]. Metode ini dijalankan dengan menggunakan algoritma *K-Nearest Neighbor* (K-NN).

Algoritma K-NN merupakan suatu teknik yang digunakan untuk mengklasifikasikan objek berdasarkan data pembelajaran yang memiliki jarak paling dekat dengan objek yang sedang diuji. Algoritma ini akan mengelompokkan hasil perhitungan dengan data pembelajaran yang memiliki kesamaan tertinggi dalam jarak yang telah ditentukan. Jarak antara data uji dan data latih diukur menggunakan persamaan *Euclidean* [7]. K-NN adalah salah satu algoritma yang mudah dan efektif untuk digunakan pada data besar [8]. Selain itu algoritma K-NN juga merupakan pendekatan yang paling mudah digunakan untuk menghitung jarak. Dalam menguji kinerja penerapan algoritma K-NN untuk sistem rekomendasi, kita dapat mengevaluasi sejauh mana keberhasilan algoritma ini dalam memberikan rekomendasi yang sesuai kepada pengguna [9].

Pada penelitian ini sistem rekomendasi musik dibuat menggunakan metode *Content-Based Filtering* untuk merekomendasikan musik berdasarkan dengan preferensi pengguna dengan menerapkan algoritma *K-Nearest Neighbor* untuk membandingkan dan mencari nilai kedekatan dari musik yang disukai pengguna dengan musik yang ada pada dataset. Dengan dibuatnya sistem rekomendasi ini diharapkan masyarakat di luar sana khususnya kelompok milenial dan “Generasi Z” mendapatkan musik yang sesuai dengan preferensi mereka dengan cepat.

2. Metode Penelitian

Adapun metode penelitian yang dilakukan pada penelitian ini yaitu:

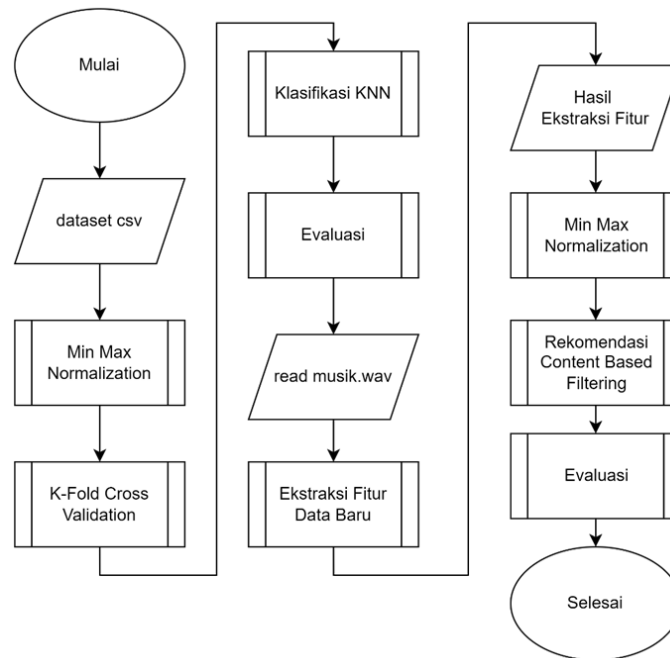
a. Pengumpulan Data

Data yang digunakan adalah data sekunder yang diambil dari *website kaggle*. Terdapat dua dataset yang akan digunakan yaitu dataset yang terdapat 4 folder yang menyimpan file musik yang telah dikelompokkan berdasarkan genrenya yaitu *blues, classical, metal, dan pop*, serta terdapat file CSV yang menyimpan hasil ekstraksi fitur dari seluruh file musik pada dataset ini. Total data musik pada dataset ini adalah sebanyak 400 data musik dengan format *.wav* yang di setiap genrenya terdiri dari 100 data musik dengan frekuensi 22050Hz yang berdurasi 30 detik pada setiap file audionya. Dataset ini akan digunakan untuk membangun model klasifikasi K-NN.

Satu dataset lagi merupakan folder yang menyimpan file musik preferensi user untuk digunakan dalam proses pencarian rekomendasi musik. Total data musik pada dataset ini adalah sebanyak 80 data musik dengan format *.wav* yang dengan genre yang sama seperti dataset sebelumnya namun pada dataset ini, tiap genrenya terdiri dari 20 data musik dengan frekuensi yang sama dan durasi yang sama dengan dataset sebelumnya. Data musik preferensi user ini yang nantinya akan diekstrak fiturnya untuk mendapatkan fitur *Zero Crossing Rate, Harmonic Percussive* dan MFCC yang kemudian akan dijadikan sebagai musik acuan yang akan dicari rekomendasinya.

b. Desain Penelitian

Metode *Content-Based Filtering* adalah sebuah teknik dalam sistem rekomendasi yang dapat menghasilkan rekomendasi dari preferensi *user* berdasarkan fitur yang terkait dengan item lain yang dibandingkan. Pada penelitian ini preferensi *user* adalah musik yang pernah didengarkan oleh user dan telah melewati proses ekstraksi fitur. Proses pencarian rekomendasinya dibantu dengan K-NN untuk menentukan genre dari data musik preferensi user. Langkah-langkah dalam penelitian yaitu diawali dengan proses pembuatan model K-NN sampai mendapatkan parameter k terbaik lalu dilanjutkan dengan proses pencarian rekomendasi musik menggunakan parameter k terbaik tersebut. *Flowchart* desain penelitian dapat dilihat pada Gambar 1.



Gambar 1. Desain Penelitian

c. *Min Max Normalization*

Tahap ini merupakan proses mengubah nilai-nilai fitur dataset ke rentang nilai antara 0 - 1. Tujuan utamanya adalah untuk memastikan bahwa semua fitur memiliki skala yang seragam, sehingga tidak ada fitur yang mendominasi yang lain dalam proses analisis atau pemodelan. Min Max Normalization dapat dilakukan dengan menerapkan rumus yang dapat dilihat pada Persamaan 1 [10].

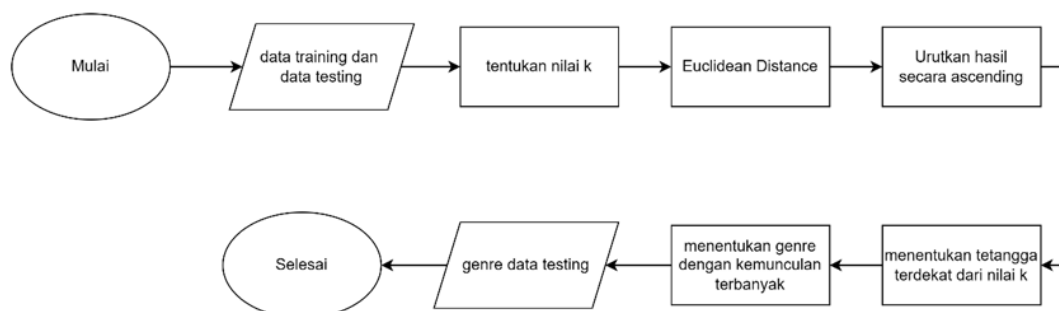
$$normalized(x) = \frac{minRange + (x - minValue)(maxRange - minRange)}{maxValue - minValue} \quad (1)$$

d. *K-Fold Cross Validation*

Proses dari teknik *K-Fold Cross Validation* dilakukan untuk memisahkan data menjadi sama rata tergantung nilai fold yang dimasukkan. Teknik *K-Fold Cross Validation* memiliki tujuan agar setiap data memiliki kesempatan untuk menjadi *data training* dan *data testing*.

e. *Klasifikasi K-NN*

Proses klasifikasi akan dilakukan terhadap dataset yang telah di split menggunakan *K-Fold Cross Validation*. Parameter k untuk menentukan jumlah tetangga terdekat dari *data testing* terhadap *data training* yang telah dihitung jaraknya dengan menggunakan metrik *Euclidean Distance*. K-NN akan menentukan label atau genre dari *data testing* dengan mengambil genre dengan kemunculan terbanyak pada jumlah tetangga terdekat yang telah ditetapkan. Proses klasifikasi memiliki tujuan untuk mendapatkan nilai k yang terbaik yang digunakan sebagai parameter untuk memprediksi genre dari data baru sekaligus menentukan rekomendasinya. Flowchart dari proses klasifikasi menggunakan K-NN dapat dilihat pada Gambar 2.



Gambar 2. Flowchart K-NN

f. Evaluasi

Tahap evaluasi merupakan tahap penting dimana kita bisa menilai kinerja dari algoritma dan metode yang digunakan untuk merekomendasikan musik ini. Tahap evaluasi ini dilakukan untuk menguji parameter k yang berbeda agar mendapatkan nilai k dengan *accuracy* tertinggi menggunakan data yang telah di-split menggunakan *K-Fold Cross Validation*. Pada penelitian ini menggunakan teknik evaluasi *accuracy* yang dapat dilihat pada Persamaan 2 [11].

$$Accuracy = \frac{Total\ Prediksi\ yang\ Benar}{Total\ Sampel} \quad (2)$$

Pada penelitian ini terdapat dua proses evaluasi yaitu evaluasi *accuracy* untuk mendapatkan nilai k terbaik dan evaluasi *accuracy* pada proses klasifikasi menggunakan data preferensi *user*.

1. Evaluasi *Accuracy* untuk Mendapatkan Nilai k Terbaik

Nilai k yang digunakan dapat ditentukan secara bebas oleh peneliti [12]. Dalam tahap pengujian ini, nilai k yang digunakan dimulai dari k=1 sampai dengan akar dari jumlah *data training* karena pembatasan ini diharapkan mampu memperoleh akurasi maksimal dengan waktu komputasi yang lebih singkat karena beberapa penelitian telah membuktikan bahwa nilai k yang terlalu tinggi juga cenderung menghasilkan *accuracy* yang semakin rendah [13].

Nilai k yang akan digunakan adalah bilangan ganjil untuk menghindari dua genre berbeda yang memiliki jumlah kemunculan yang sama pada tetangga terdekat. Oleh karena itu, skenario pengujian *accuracy* akan dilakukan dengan data yang telah dibagi menjadi 5 fold yang mana *data training* pada penelitian ini sebanyak 320 data training, sehingga nilai k yang dipilih adalah bilangan ganjil dalam rentang 1 hingga 17 [13].

Pada penelitian ini menggunakan teknik pengujian *accuracy* yang dapat dilihat pada persamaan (2). *Accuracy* digunakan untuk mengukur seberapa akurat model klasifikasi dalam mengklasifikasikan sampel untuk semua kelas [11].

2. Evaluasi *Accuracy* pada Data Preferensi User

Evaluasi ini dilakukan dengan tujuan untuk menilai seberapa efektif sistem tersebut dalam menentukan genre dari data preferensi *user*. Data preferensi *user* akan dinormalisasi menggunakan *Min Max Normalization* kemudian akan diklasifikasikan dengan K-NN. Nilai yang digunakan adalah nilai k yang telah ditetapkan sebagai nilai k terbaik berdasarkan evaluasi penentuan nilai k terbaik. Hasil dari K-NN tersebut akan dicari akurasi dan akurasi tersebutlah yang akan dijadikan tingkat keberhasilan sistem dalam menentukan genre dari data preferensi *user*.

g. Ekstraksi Fitur

Pada langkah ini dilakukan ekstraksi fitur pada musik-musik yang pernah didengarkan oleh *user* seperti yang dijelaskan pada data penelitian untuk menjalankan proses pencarian rekomendasi musik. Fitur-fitur yang akan diekstrak sesuai dengan fitur-fitur yang terdapat pada dataset musik yang digunakan yaitu *zero crossing rate*, *harmony percussive source separation* serta MFCC. Panjang vektor MFCC yang akan diekstrak adalah sepanjang 13 vektor fitur. Setelah seluruh fitur berhasil diekstrak, maka selanjutnya akan dibuat satu DataFrame yang menampung hasil dari ekstraksi fitur tersebut.

h. Rekomendasi *Content Based Filtering*

Setelah *data file* audio musik sebelumnya telah didapatkan fiturnya dan menjadi data vektor, data tersebut akan dicari kedekatannya dengan kelas genre terdekat diantara genre *Blues*, *Classical*, *Metal* dan *Pop* dengan menentukan nilai k. Nilai k dapat mempengaruhi tingkat akurasi perhitungan sehingga diperlukan perhitungan untuk menemukan nilai optimal k. *Data input* akan dihitung jaraknya dengan *data training* dan akan di sortir mulai dari yang terdekat sampai yang terjauh. Setelah diketahui jarak antara *data input* dan *data training* maka selanjutnya menentukan nilai k yang digunakan untuk menentukan kelas dari *data input*. Penentuan nilai k dilakukan dengan menggunakan beberapa nilai k yang nantinya akan dioptimasi dengan memperhatikan nilai k mana yang memiliki jarak terdekat. Setelah mengetahui kelas label genre maka rekomendasi musik diambil dari data-data yang termasuk di dalam kelas label genre tersebut dengan mengambil 5 data musik dengan jarak terdekat akan dijadikan output dengan menggunakan *Euclidean Distance* yang akan diurutkan mulai dari yang terdekat.

3. Hasil dan Diskusi

Implementasi sistem dilakukan dengan menggunakan bahasa pemrograman *python* dan beberapa *library* yang didukung oleh *python*. *Code* yang dibuat dijalankan di *google colab*.

3.1. Data Penelitian

Data yang digunakan untuk membangun model klasifikasi K-NN adalah file CSV yang menyimpan hasil ekstraksi fitur dari data-data musik pada dataset ini. Total data musik pada dataset ini adalah sebanyak 400 data musik yang terdiri dari 4 genre yaitu blues, classical, metal, dan pop dan di setiap genrenya terdiri dari 100 data musik.

Data musik preferensi user digunakan sebagai data input dalam proses pencarian rekomendasi musik. Total data musik pada dataset ini adalah sebanyak 80 data musik dengan format *.wav* yang dengan genre yang sama seperti dataset sebelumnya namun pada dataset ini, tiap genrenya terdiri dari 20 data musik. Data musik preferensi user ini yang nantinya akan diekstrak fiturnya untuk mendapatkan fitur Zero Crossing Rate, Harmonic Percussive dan MFCC yang kemudian akan dijadikan sebagai musik acuan yang akan dicari rekomendasinya.

Output yang akan dimunculkan adalah 5 data musik terdekat akan dijadikan sebagai rekomendasi musik. Tingkat keberhasilan rekomendasi dapat dilihat dari evaluasi model klasifikasi K-NN menggunakan nilai k terbaiknya dalam memprediksi genre dari data musik preferensi user.

3.2. Evaluasi

Tahap evaluasi ini dilakukan untuk menguji parameter k yang berbeda agar mendapatkan nilai k dengan *accuracy* tertinggi menggunakan data yang telah di-split menggunakan *K-Fold Cross Validation*.

a. Evaluasi *accuracy* untuk mendapatkan nilai k terbaik

Hasil pengujian *accuracy* dari kelima *fold* tersebut kemudian akan dirata-ratakan untuk mendapatkan *accuracy* nilai k secara keseluruhan. *Accuracy* tersebut disimpan ke dalam *DataFrame*. Hasil perhitungan *accuracy* dapat dilihat pada Tabel 1.

Tabel 1. *Accuracy* K-NN dengan beberapa Nilai k

Nilai k	<i>Accuracy</i>
1	0,9009
3	0,9018
5	0,9032
7	0,9043
9	0,9049
11	0,9035
13	0,9022
15	0,9005
17	0,8984

Berdasarkan hasil yang ditampilkan Tabel 1, didapatkan nilai *accuracy* dari yang terendah hingga yang tertinggi dengan rentang *accuracy* 0,8984 atau 89,84% sampai 0,9049 atau 90,49%. *Accuracy* tertinggi pada klasifikasi K-NN diperoleh pada nilai k=9 dengan *accuracy* berada di angka 0,9049 atau 90,49% sedangkan *accuracy* terendahnya diperoleh pada nilai k=17 dengan *accuracy* berada di angka 0,8984 atau 89,84%. Oleh karena itu, pada proses pencarian rekomendasi akan menggunakan nilai k=9 untuk menentukan genre dari data musik preferensi user.

b. Evaluasi pada Data Preferensi User

Evaluasi ini dilakukan untuk mengetahui keberhasilan sistem dalam menentukan genre dari data preferensi user. Data preferensi user akan dinormalisasi menggunakan *Min Max Normalization* kemudian akan diklasifikasikan dengan K-NN dan terakhir akan dilakukan perhitungan *accuracy*. Hasil evaluasi pada data preferensi user dapat dilihat pada Tabel 2.

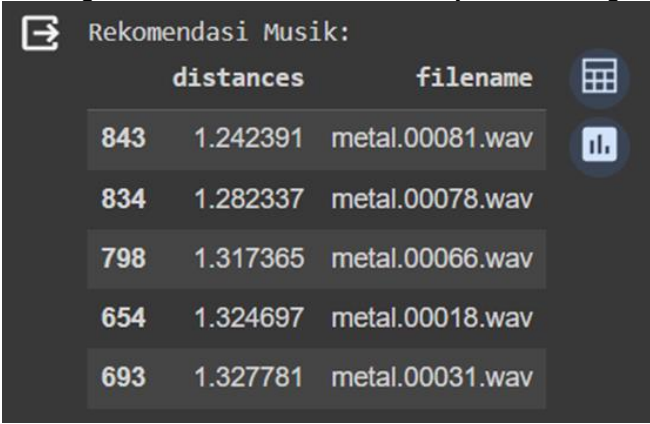
Tabel 2. Hasil *accuracy* K-NN dengan beberapa Nilai k

Nilai k	<i>Accuracy</i>
9	0,9018

Berdasarkan hasil yang ditampilkan Tabel 2, didapatkan tingkat keberhasilan sistem dalam menentukan genre pada data preferensi user dengan nilai *accuracy* sebesar 0,9018 atau 90,18%.

3.3. Hasil Rekomendasi Musik

Berdasarkan hasil klasifikasi sebelumnya, percobaan pada data baru telah mendapatkan hasil bahwa data tersebut termasuk ke dalam label *genre* 'metal' maka data yang memiliki label 'metal' akan dipertahankan dan 5 data dengan label 'metal' teratas akan dijadikan sebagai rekomendasi musik.



Rekomendasi Musik:

	distances	filename
843	1.242391	metal.00081.wav
834	1.282337	metal.00078.wav
798	1.317365	metal.00066.wav
654	1.324697	metal.00018.wav
693	1.327781	metal.00031.wav

Gambar 3. Hasil Rekomendasi Musik

4. Kesimpulan

Penelitian ini telah mampu melakukan rekomendasi dengan menerapkan metode *Content-Based Filtering* dan *K-Nearest Neighbor*. Kesimpulan pada penelitian ini yaitu:

1. Hasil evaluasi pada klasifikasi yang dijalankan dengan menggunakan K-NN telah berhasil mendapatkan nilai *accuracy* tertinggi berada pada nilai $k=9$ dengan *accuracy* sebesar 0,9049 atau 90,49% sedangkan *accuracy* terendahnya diperoleh pada nilai $k=17$ dengan *accuracy* berada di angka 0,8984 atau 89,84%.
2. Hasil evaluasi pada data preferensi *user* yang dijalankan menggunakan K-NN dengan menggunakan nilai k terbaik berdasarkan evaluasi sebelumnya yaitu $k=9$ berhasil mendapatkan *accuracy* sebesar 0,9018 atau 90,18%.

References

- [1] D. Noviani, R. Pratiwi, S. Silvianadewi, M. B. Alexandri, And M. A. Hakim, "Pengaruh Streaming Musik Terhadap Industri Musik Di Indonesia," 2020.
- [2] E. P. Utomo And N. Probosini, "Pengaruh Sikap Dan Norma Subjektif Terhadap Intensitas Penggunaan Aplikasi Streaming Pada Generasi 'Z,'" *Jurnal Ilmu Sosial Dan Humaniora*, Vol. 9, No. 2, P. 241, Sep. 2020, Doi: 10.23887/Jish-Undiksha.V9i2.18581.
- [3] A. , R. Abdurokhim & Hidayat, "Pengaruh Sikap Dan Norma Subjektif Pada Intensitas Generasi 'Z' Menggunakan Aplikasi Media Streaming," *Jurnal Ilmiah Indonesia*, Januari, Vol. 2020, No. 1, Pp. 9–22, 2020, [Online]. Available: [Http://Cerdika.Publikasiindonesia.Id/Index.Php/Cerdika/Indexhttp://Cerdika.Publikasiindonesia.Id/-9-](http://Cerdika.Publikasiindonesia.Id/Index.Php/Cerdika/Indexhttp://Cerdika.Publikasiindonesia.Id/-9-)
- [4] H.-C. Chen And A. L. P. Chen, "A Music Recommendation System Based On Music And User Grouping *," 2005.
- [5] N. Ula, C. Setianingsih, And R. A. Nugrahaeni, "Sistem Rekomendasi Lagu Dengan Metode Content-Based Filtering Berbasis Website Web-Based Song Recommendation System Using Content-Based Filtering," 2021.
- [6] A. I. Putra And R. R. Santika, "Edumatic: Jurnal Pendidikan Informatika Implementasi Machine Learning Dalam Penentuan Rekomendasi Musik Dengan Metode Content-Based Filtering," Vol. 4, No. 1, 2020, Doi: 10.29408/Edumatic.V4i1.2162.
- [7] M. M. Baharuddin, H. Azis, And T. Hasanuddin, "Analisis Performa Metode K-Nearest Neighbor Untuk Identifikasi Jenis Kaca," *Ilkom Jurnal Ilmiah*, Vol. 11, No. 3, Pp. 269–274, Dec. 2019, Doi: 10.33096/Ilkom.V11i3.489.269-274.

- [8] H. A. Sheeren, L. W. Santoso, And J. Andjarwirawan, "Sistem Rekomendasi Pembelian Laptop Dengan K-Nearest Neighbor (Knn)," 2019.
- [9] M. Teknologi Informasi, P. Pascasarjana, And U. Teknologi Yogyakarta, "This Work Is Licensed Under A Creative Commons Attribution-Sharealike 4.0 International License Sistem Rekomendasi Pada Tokopedia Menggunakan Algoritma K-Nearest Neighbor," 2022, Doi: 10.31294/Jtk.V4i2.
- [10] D. Azzahra Nasution, H. H. Khotimah, And N. Chamidah, "Perbandingan Normalisasi Data Untuk Klasifikasi Wine Menggunakan Algoritma K-Nn," 2019.
- [11] M. Rangga, A. Nasution, And M. Hayaty, "Perbandingan Akurasi Dan Waktu Proses Algoritma K-Nn Dan Svm Dalam Analisis Sentimen Twitter," *Jurnal Informatika*, Vol. 6, No. 2, Pp. 212–218, 2019, [Online]. Available: [Http://Ejournal.Bsi.Ac.Id/Ejournal/Index.Php/Ji](http://ejournal.bsi.ac.id/ejournal/index.php/ji)
- [12] R. L. Hasanah *Et Al.*, "Klasifikasi Penerima Dana Bantuan Desa Menggunakan Metode Knn (K-Nearest Neighbor)," *Jurnal Techno Nusa Mandiri*, Vol. 16, No. 1, P. 1, 2019, [Online]. Available: [Http://Nusamandiri.Ac.Id/](http://nusamandiri.ac.id/)
- [13] A. B. Hassanat, M. A. Abbadi, G. A. Altarawneh, And A. A. Alhasanat, "Solving The Problem Of The K Parameter In The Knn Classifier Using An Ensemble Learning Approach," 2014. [Online]. Available: [Http://Sites.Google.Com/Site/ljcsis/](http://sites.google.com/site/ljcsis/)

This page is intentionally left blank.