

Pembuatan Model Rekomendasi Gerakan Olahraga Berbasis Content-Based Filtering Dengan Integrasi Clustering dan Klasifikasi

Andi Wibowo¹, Wati Erawati²

^{1,2}Universitas Bina Sarana Informatika
e-mail: ¹wibowoandi001@gmail.com, ²wati.wti@bsi.ac.id

Diterima	Direvisi	Disetujui
06-10-2025	04-11-2025	22-12-2025

Abstrak - Sistem rekomendasi gerakan olahraga menjadi esensial dalam mempromosikan gaya hidup sehat dengan memberikan saran aktivitas yang personal. Latar belakang ini mendorong pengembangan sistem rekomendasi yang mampu memahami karakteristik pengguna dan menyediakan rekomendasi aktivitas yang relevan. Tujuan penelitian ini adalah merancang dan mengimplementasikan model rekomendasi terintegrasi melalui tiga metode utama: *clustering*, klasifikasi, dan *content-based filtering*. Tahap pertama, *clustering*, menggunakan algoritma K-Means untuk mengelompokkan 3.000 partisipan ke dalam kluster 'Aktif' dan 'Pasif' berdasarkan profil kebugaran, yang kemudian menjadi fitur penting dalam tahap klasifikasi. Selanjutnya, berbagai model machine learning dievaluasi untuk memprediksi jenis aktivitas olahraga utama yang sesuai dengan profil pengguna. Hasil evaluasi menunjukkan bahwa model K-Nearest Neighbors (KNN) adalah yang terbaik dengan akurasi 0.4875, mengungguli Decision Tree, Random Forest, dan Deep Learning. Model KNN yang terlatih disimpan untuk *deployment*. Tahap terakhir adalah *content-based filtering* yang merekomendasikan tiga aktivitas serupa dengan aktivitas utama hasil prediksi, berdasarkan fitur konten seperti intensitas, tingkat kebugaran, kalori, dan langkah harian menggunakan Cosine Similarity. Sistem berhasil memprediksi aktivitas utama (misalnya *cycling*) dan memberikan rekomendasi relevan (seperti *running*, *basketball*, dan *tennis*) berdasarkan input profil pengguna.

Kata Kunci: Sistem Rekomendasi, Content-Based Filtering, Flask, Personalisasi Kebugaran.

Abstract - Sport movement recommendation systems are essential for promoting healthy lifestyles by providing personalized activity suggestions. This background motivated the development of a recommendation system capable of understanding user characteristics and delivering relevant activity recommendations. The objective of this research is to design and implement an integrated recommendation system using three main methods: clustering, classification, and content-based filtering. The first stage, clustering, employed the K-Means algorithm to group 3,000 participants into 'Active' and 'Passive' clusters based on their fitness profiles, which then became crucial features in the subsequent classification stage. Next, various machine learning models were evaluated to predict the primary sport activity type suitable for a user's profile. Evaluation results indicated that the K-Nearest Neighbors (KNN) model performed best with an accuracy of 0.4875, outperforming Decision Tree, Random Forest, and Deep Learning models. The trained KNN model was saved for deployment. The final stage involved content-based filtering, which recommends three additional activities similar to the main predicted activity, based on content features like intensity, fitness level, calories burned, and daily steps, quantified using Cosine Similarity. The system successfully predicted the main activity (e.g., cycling) and provided relevant recommendations (e.g., running, basketball, tennis) based on user profile input.

Keywords: Recommendation System, Content-Based Filtering, Fitness Personalization.

PENDAHULUAN

Tingkat kebugaran jasmani masyarakat Indonesia yang masih rendah menjadi isu krusial yang perlu perhatian serius, mengingat dampaknya terhadap kualitas hidup dan risiko kesehatan jangka panjang. Laporan Nasional *Sport Development Index (SDI) 2022* dari Kementerian Pemuda dan Olahraga (Kemenpora) mengindikasikan bahwa mayoritas anak muda Indonesia berada dalam kategori kebugaran "kurang" hingga "kurang sekali".

Kondisi serupa juga teramati pada kelompok usia dewasa, di mana pola hidup aktif dan sehat yang konsisten masih belum menjadi kebiasaan umum. Hal ini selaras dengan penelitian (Geralda Adhianto et al., 2023) yang menyoroti hubungan signifikan antara minimnya aktivitas fisik dengan penurunan kebugaran jasmani pada peserta didik, menunjukkan bahwa kurangnya pemahaman tentang pentingnya aktivitas fisik turut berkontribusi pada rendahnya tingkat kebugaran mereka.

Rendahnya kebugaran jasmani ini berimplikasi langsung pada peningkatan risiko berbagai Penyakit Tidak Menular (PTM) yang mengancam. Penyakit seperti jantung koroner, stroke, hipertensi, dan diabetes tipe 2 telah menjadi momok serius bagi masyarakat Indonesia. Data Organisasi Kesehatan Dunia (WHO) mengungkapkan bahwa sekitar 71% kematian di Indonesia disebabkan oleh PTM, termasuk stroke, obesitas, dan komplikasi kesehatan lainnya. Fakta ini menegaskan betapa mendesaknya upaya peningkatan kesadaran dan penyediaan fasilitas untuk menjaga kebugaran tubuh secara terstruktur dan berkelanjutan. Dari perspektif kesehatan, menjaga kebugaran fisik terbukti mampu mengurangi risiko penyakit kronis seperti diabetes, penyakit jantung, dan tekanan darah tinggi (Dwi Putranto & Yasin Efendi, 2024). Dengan demikian, peningkatan kebugaran jasmani melalui aktivitas fisik rutin adalah langkah strategis untuk menekan beban PTM di Indonesia, sekaligus meningkatkan kualitas hidup masyarakat secara menyeluruh.

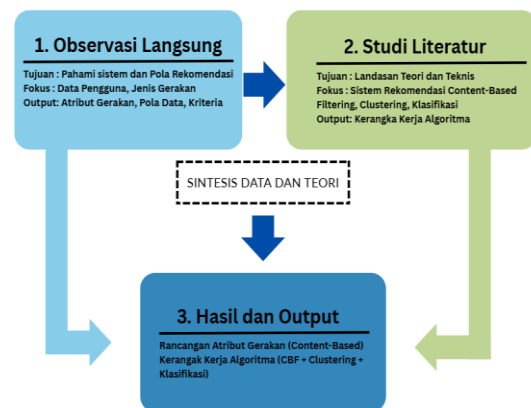
Dalam konteks ini, perkembangan teknologi menawarkan potensi besar sebagai solusi. Inovasi di bidang kecerdasan buatan (*Artificial Intelligence/AI*) dan *machine learning* telah membuka jalan baru untuk mendukung pola hidup sehat. Di tengah minimnya kesadaran masyarakat akan pentingnya kebugaran, kehadiran sistem digital yang mampu memberikan rekomendasi olahraga secara personal menjadi sangat relevan dan dibutuhkan. Inovasi kesehatan digital, khususnya dengan integrasi AI, berperan sebagai pilar penting dalam diseminasi informasi kesehatan. AI memungkinkan interaksi yang lebih mendalam antara manusia dan mesin, dengan kemampuan menganalisis data, memberikan rekomendasi personal, serta mempercepat proses diagnosis dan edukasi kesehatan secara efektif (Subiyantoro, 2024). Hal ini didukung oleh penelitian (Bron et al., 2024) yang menyimpulkan bahwa metode *machine learning* efektif digunakan untuk mempersonalisasi strategi persuasif dalam intervensi *mHealth*, yang secara signifikan dapat mendorong dan meningkatkan aktivitas fisik pengguna. Menurut artikel MedisInfo, integrasi AI dalam inovasi kesehatan digital tidak hanya meningkatkan efisiensi penyebaran informasi kesehatan, tetapi juga berkontribusi pada peningkatan kesadaran dan pengetahuan masyarakat mengenai pentingnya pola hidup sehat (Arif & Baharudin, 2025).

Memanfaatkan teknologi sebagai alat bantu yang efektif dalam memberikan panduan berolahraga yang sesuai dengan kondisi fisik dan tujuan individu, serta dapat diakses kapan saja dan di mana saja, menjadi krusial. Oleh karena itu, "Pembuatan Model Rekomendasi Gerakan Olahraga berbasis *Content-Based Filtering* dengan Integrasi *Clustering* dan Klasifikasi" merupakan solusi potensial untuk mengatasi permasalahan kebugaran jasmani yang rendah di Indonesia. Integrasi ketiga metode ini memungkinkan sistem tidak hanya merekomendasikan gerakan

berdasarkan kesamaan karakteristik konten gerakan dengan profil pengguna, tetapi juga mampu mengelompokkan pengguna atau gerakan olahraga berdasarkan pola tertentu (*clustering*) serta mengklasifikasikan preferensi atau kondisi khusus pengguna (klasifikasi) untuk rekomendasi yang lebih presisi dan adaptif. Pendekatan hibrida ini diharapkan dapat memberikan rekomendasi yang sangat personal, relevan, dan dinamis, sehingga mendorong masyarakat untuk lebih aktif dan mencapai gaya hidup sehat yang berkelanjutan.

METODE PENELITIAN

Penelitian ini menggunakan pendekatan kualitatif dan kuantitatif melalui dua metode utama: observasi dan studi literatur.



Gambar.1 Tahapan Metode Penelitian

Metode observasi dilakukan dengan mengamati secara langsung beberapa aplikasi atau platform kebugaran digital yang sudah ada. Observasi ini bertujuan untuk memahami bagaimana sistem rekomendasi yang ada beroperasi, terutama dalam hal pola rekomendasi konten, bagaimana aplikasi mengumpulkan data pengguna (seperti usia, berat badan, tinggi badan, dan keaktifan olahraga), serta jenis-jenis gerakan olahraga yang biasanya direkomendasikan berdasarkan input tersebut. Selain itu, aspek presentasi informasi kepada pengguna juga diamati untuk merancang alur interaksi yang user-friendly dan informatif pada sistem yang dikembangkan. Hasil observasi menjadi dasar dalam merancang atribut-atribut gerakan olahraga yang relevan untuk proses *Content-Based Filtering* dan mengidentifikasi potensi pola untuk *Clustering* serta kriteria untuk Klasifikasi.

Metode dokumentasi dilakukan melalui studi pustaka mendalam terhadap berbagai literatur ilmiah, jurnal, dan artikel penelitian yang membahas penerapan sistem rekomendasi dengan metode *Content-Based Filtering* serta integrasi *Clustering* dan Klasifikasi, khususnya yang berkaitan dengan bidang kebugaran dan kesehatan digital. Fokus utama dokumentasi adalah pada:

1. Konsep dan Algoritma *Content-Based Filtering*

Memahami bagaimana sistem memproses deskripsi atau fitur dari item (gerakan olahraga) dan membandingkannya dengan profil preferensi pengguna, termasuk teknik pengukuran kesamaan.

2. Teknik *Clustering*

Mempelajari algoritma unsupervised learning yang digunakan untuk mengelompokkan data (baik pengguna maupun gerakan olahraga) berdasarkan kesamaan karakteristik, seperti K-Means atau hirarki *clustering*.

3. Teknik Klasifikasi

Memahami algoritma supervised learning yang digunakan untuk memprediksi kategori atau label berdasarkan fitur input, yang dapat diterapkan untuk mengidentifikasi jenis preferensi atau kondisi pengguna.

4. Strategi Integrasi

Mempelajari berbagai pendekatan untuk mengintegrasikan *Content-Based Filtering* dengan *Clustering* dan Klasifikasi guna meningkatkan kualitas rekomendasi.

5. Pra-pemrosesan Data

Mempelajari metode untuk membersihkan, menormalisasi, dan mengubah data gerakan olahraga serta data pengguna menjadi format yang siap diproses oleh ketiga algoritma.

6. Metode Evaluasi Model Rekomendasi Hibrida

Mempelajari metrik yang umum digunakan untuk mengevaluasi performa sistem rekomendasi yang kompleks. Informasi yang diperoleh dari dokumentasi ini menjadi dasar konseptual dan teknis dalam merancang, membangun, dan mengevaluasi model rekomendasi gerakan olahraga yang personal, akurat, dan adaptif.

HASIL DAN PEMBAHASAN

Bagian ini menyajikan hasil dari proses pembuatan sistem rekomendasi gerakan olahraga, dimulai dari alur sistem secara keseluruhan hingga detail implementasi setiap tahapan, serta contoh studi kasus dan keterkaitannya dengan kompetensi program studi.

1. Alur Pembuatan Sistem Rekomendasi

Pembuatan sistem rekomendasi gerakan olahraga ini dirancang dengan pendekatan tiga tahapan utama yang terintegrasi secara sekuensial untuk menghasilkan rekomendasi yang personal dan relevan. Diagram alur sistem secara keseluruhan.

2. Memahami Dataset

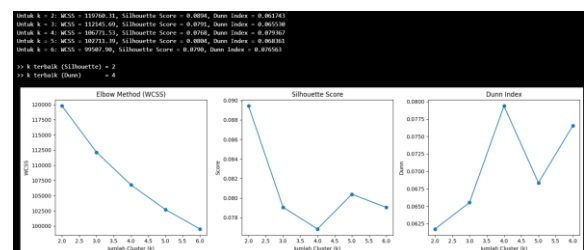
Dalam penelitian ini menggunakan dataset yang berisikan data pelacakan kesehatan dan kebugaran dari 3.000 partisipan, mencakup aktivitas harian, indikator kesehatan penting, dan faktor gaya hidup. Dataset tersebut didapatkan dari *website* Kaggle dengan nama "*Health and fitness dataset*" dan dapat di lihat pada tautan

<https://www.kaggle.com/datasets/evan65549/health-and-fitness-dataset>.

Analisis korelasi antar fitur dalam dataset menunjukkan pentingnya memahami hubungan linear antar variabel, yang dikategorikan dari korelasi positif kuat (mendekati 1.00) hingga korelasi negatif kuat (mendekati -1.00), serta korelasi sedang dan lemah. Namun, analisis ini memiliki keterbatasan dalam menjelaskan distribusi data yang lebih kompleks, seperti fenomena bimodalitas, yang mengindikasikan adanya sub-populasi tersembunyi. Oleh karena itu, analisis ini menjadi fondasi awal untuk tahapan pemilihan fitur dan pemodelan, tetapi perlu dilengkapi dengan metode yang lebih canggih seperti *clustering* untuk mengungkap pola non-linear dan mendapatkan pemahaman data yang lebih mendalam.

3. Tahap 1: Pembuatan Model *Clustering*

Tahapan *clustering* merupakan komponen fundamental dalam sistem rekomendasi ini, yang bertujuan untuk memahami dan melakukan segmentasi pengguna berdasarkan karakteristik kebugaran mereka. Proses ini esensial untuk mengidentifikasi kelompok-kelompok partisipan dengan profil serupa, yang nantinya akan digunakan untuk mempersonalisasi rekomendasi aktivitas olahraga. Algoritma yang digunakan dalam tahap ini adalah *K-Means Clustering*, dengan jumlah kluster optimal ditentukan menggunakan metode *elbow*. Hasil dari model *clustering* ini adalah label *cluster_aktivitas* untuk setiap pengguna, yang kemudian akan berperan sebagai fitur tambahan yang krusial dalam model klasifikasi pada tahap berikutnya, meningkatkan relevansi prediksi sistem.



Sumber: (Penelitian 2025)

Gambar.2 Hasil Visualisasi Clustering

Berdasarkan hasil proses *clustering*, partisipan dalam dataset berhasil dikelompokkan ke dalam dua kluster utama: 'Aktif' dan 'Pasif'. Setiap kluster menunjukkan karakteristik yang berbeda secara signifikan pada berbagai fitur kebugaran dan gaya hidup. Pemilihan

jumlah kluster $k=2$ didukung oleh evaluasi WCSS yang menunjukkan nilai 119760.31, menandakan tingkat kekompakan cluster yang cukup baik dengan variansi internal yang relatif kecil. Selain itu, *Silhouette Score* untuk $k=2$ mencapai 0.0894, nilai tertinggi dibandingkan jumlah cluster lainnya, sehingga mengindikasikan pemisahan cluster yang lebih optimal secara relatif. Meskipun nilai *Dunn Index* pada $k=2$ adalah 0.0617 (lebih rendah dibanding $k=4$), hasil ini tetap konsisten dengan temuan *Silhouette Score* yang menegaskan bahwa pembagian ke dalam dua kluster memberikan keseimbangan terbaik antara kekompakan intra-kluster dan pemisahan antar-kluster. Dengan demikian, $k=2$ dipilih sebagai jumlah *cluster* yang paling representatif untuk menggambarkan segmentasi partisipan dalam dataset ini.

4. Tahap 2: Klasifikasi

Setelah berhasil melakukan clustering dan menyimpan hasil label kluster kedalam dokumen CSV, tahap selanjutnya adalah Klasifikasi. Fokus utama adalah pengembangan dan evaluasi model prediktif yang mampu mengklasifikasikan atau memprediksi jenis aktivitas olahraga yang paling sesuai untuk seorang pengguna berdasarkan profil dan kluster kebugaran mereka. Data yang telah diperkaya dengan label kluster dari hasil Cluster sebelumnya digunakan untuk melatih model klasifikasi ini, sehingga dapat memberikan prediksi aktivitas yang personal dan relevan.

Untuk mengidentifikasi model klasifikasi terbaik dalam memprediksi aktivitas olahraga, membandingkan kinerja empat algoritma: K-Nearest Neighbors (KNN), Random Forest, Decision Tree, dan Neural Network (*Deep Learning*). Setiap algoritma memiliki karakteristik unik. KNN bekerja berdasarkan mayoritas kelas dari tetangga terdekat, unggul dalam kesederhanaan tetapi lambat pada data besar. Random Forest menggunakan ensemble learning dari banyak pohon keputusan, yang membuatnya sangat akurat dan tangguh terhadap outlier namun sulit diinterpretasi. Decision Tree membangun serangkaian aturan yang mudah dipahami, tetapi rentan terhadap *overfitting*. Terakhir, Neural Network melalui arsitektur Sequential dengan lapisan-lapisan neuron (64 neuron, 32 neuron, dan lapisan output dengan fungsi aktivasi *Softmax*) mampu menangani pola data yang sangat kompleks, meskipun memerlukan komputasi intensif dan interpretasinya sulit (*black box*).

Model Neural Network kami dibangun dengan arsitektur *Sequential*, menggunakan Dense layer dengan 64 *neuron* dan ReLU di lapisan pertama, lalu 32 neuron dengan ReLU di lapisan tersembunyi kedua, dan diakhiri dengan lapisan output yang menggunakan *Softmax* untuk klasifikasi multi-kelas. Sementara itu, Decision Tree Classifier, Random Forest Classifier, dan K-Nearest Neighbors Classifier diimplementasikan menggunakan fungsi standar dari Scikit-learn. Untuk memastikan konsistensi, parameter *random_state* diatur

pada angka 42 pada model Decision Tree dan Random Forest.

Untuk mengukur dan membandingkan kinerja setiap model, digunakan metrik evaluasi standar yang memberikan gambaran komprehensif, terutama karena adanya potensi ketidakseimbangan kelas dalam dataset. Metrik yang digunakan meliputi:

a. Metrik Akurasi (*Accuracy*)

Mengukur persentase prediksi yang benar terhadap seluruh data. Akurasi merupakan metrik paling dasar yang memberikan gambaran umum mengenai seberapa sering model membuat prediksi yang tepat.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

b. Metrik Presisi (*Precision*)

Mengukur proporsi prediksi positif yang benar-benar merupakan data positif. Presisi relevan digunakan ketika penting untuk meminimalkan *false positive*, yaitu kasus di mana model secara keliru memprediksi kelas positif.

$$Precision = \frac{TP}{TP + FP}$$

c. Metrik Sensitivitas (*Recall*)

Mengukur proporsi data positif yang berhasil dikenali oleh model. *Recall* sangat penting ketika tujuan utamanya adalah meminimalkan *false negative*, yaitu kasus di mana model gagal mengenali kelas positif yang sebenarnya.

$$Recall = \frac{TP}{TP + FN}$$

d. Metrik F1-Score

Merupakan rata-rata harmonik dari precision dan recall. Metrik ini sangat berguna ketika dibutuhkan keseimbangan antara precision dan recall, terutama pada kasus klasifikasi dengan imbalanced dataset di mana hanya melihat akurasi saja bisa menyesatkan.

5. Tahap 3: Content-Based Filtering

Setelah model klasifikasi berhasil memprediksi jenis aktivitas olahraga utama yang sesuai untuk seorang pengguna, tahap selanjutnya adalah membangun sistem rekomendasi yang memberikan opsi tambahan. Untuk tujuan ini, sistem mengimplementasikan metode Content-Based Filtering (CBF). CBF adalah

pendekatan sistem rekomendasi yang bekerja dengan menyarankan item (dalam hal ini, jenis aktivitas olahraga) berdasarkan kemiripan atribut atau karakteristik konten antara item tersebut. Ini berarti sistem akan merekomendasikan aktivitas yang memiliki profil fitur serupa dengan aktivitas hasil prediksi utama. Berikut merupakan langkah-langkah yang dilakukan:

a. Pembuatan Profil Aktivitas

Pada langkah ini, seluruh aktivitas olahraga dikelompokkan berdasarkan *activity_type*. Kemudian, rata-rata dari fitur-fitur konten yang relevan (*intensity_encoded*, *fitness_level*, *calories_burned*, *daily_steps*) dihitung untuk setiap *activity_type*. Hasil perhitungan rata-rata ini membentuk 'profil aktivitas' untuk setiap jenis olahraga.

activity_type	intensity_encoded	fitness_level	calories_burned	daily_steps
Running	2	9.2	105	500
Swimming	2	85	95	0

Sumber: (Penelitian 2025)

Gambar.3 Contoh Profil Aktivitas

b. Perhitungan Kemiripan (*Similarity Calculation*)

Setiap profil aktivitas diperlakukan sebagai vektor numerik dalam ruang multidimensional yang ditentukan oleh fitur-fitur konten. Untuk mengukur kemiripan antar aktivitas, digunakan metrik *Cosine Similarity*. *Cosine similarity* menghitung kosinus sudut antara dua vektor, di mana nilai yang mendekati 1 menunjukkan kemiripan tinggi dan 0 menunjukkan tidak ada kemiripan. Proses ini menghasilkan sebuah matriks kemiripan yang mencatat tingkat kemiripan antara setiap pasangan *activity_type*.

c. Pemberian Rekomendasi

Tahap akhir dalam proses Content-Based Filtering adalah pemberian rekomendasi aktivitas olahraga kepada pengguna. Berdasarkan *activity_type* yang telah diprediksi oleh model klasifikasi K-Nearest Neighbors (KNN), sistem akan menjalankan serangkaian langkah. Pertama, sistem akan mengambil baris yang sesuai dari matriks kemiripan yang telah dihasilkan pada langkah perhitungan kemiripan sebelumnya. Baris ini secara spesifik merepresentasikan tingkat kemiripan antara aktivitas prediksi utama dengan seluruh aktivitas lainnya yang ada dalam dataset. Selanjutnya, aktivitas-aktivitas tersebut akan diurutkan dari tingkat kemiripan tertinggi hingga terendah. Dari daftar yang telah terurut ini, sistem akan memberikan tiga rekomendasi aktivitas yang paling mirip, dengan catatan bahwa aktivitas yang merupakan hasil prediksi utama itu sendiri akan dicekualikan untuk menghindari rekomendasi yang bersifat duplikat dan tidak informatif.

Logika *Content-Based Filtering* diintegrasikan ke dalam aplikasi web menggunakan framework Flask, yang terletak pada *app.py* di direktori *activity-recommendation-app/*. Aplikasi *Flask* ini memiliki struktur folder yang jelas untuk *backend* (*app.py*), *template* HTML (*templates/dashboard.html*), dan aset statis (*static/*). Khususnya, model klasifikasi (*knn_model_new.pkl*) dan label *encoder* (*le_activity_encoder_new.pkl*) juga disimpan dalam direktori utama aplikasi, siap untuk dimuat oleh *app.py*.

```
cbf_df = df[["activity_type", "intensity_encoded", "fitness_level", "calories_burned", "daily_steps"]]
activity_profiles = cbf_df.groupby("activity_type").mean(numeric_only=True)
similarity_matrix = cosine_similarity(activity_profiles)
similarity_df = pd.DataFrame(similarity_matrix, index=activity_profiles.index, columns=activity_profile)

# Rekomendasi
similar = similarity_df[predicted_activity].sort_values(ascending=False)
top_similar = similar.drop(predicted_activity).head(3).index.tolist()
```

Sumber: (Penelitian 2025)

Gambar.4 Implementasi CBF di *app.py*

Untuk mengilustrasikan fungsionalitas dan kinerja sistem rekomendasi secara langsung, dilakukan sebuah studi kasus dengan menggunakan antarmuka pengguna yang telah dikembangkan. Pengguna diminta untuk memasukkan profil pribadi mereka, yang kemudian akan diproses oleh model klasifikasi dan sistem rekomendasi. Pada studi kasus ini, *input* profil pengguna yang dimasukkan adalah sebagai berikut:

Sumber: (Penelitian 2025)

Gambar.5 Dashboard Rekomendasi

Hasil studi kasus ini secara jelas menunjukkan bahwa sistem mampu memprediksi "*Cycling*" sebagai aktivitas utama yang paling sesuai dengan profil pengguna yang relatif pasif. Selanjutnya, sistem memberikan tiga rekomendasi aktivitas lain yang memiliki kemiripan konten tertinggi: "*Running*" dengan skor kemiripan 40.07%, "*Basketball*" dengan skor 39.66%, dan "*Tennis*" dengan skor 26.74%. Hal ini mengilustrasikan bahwa sistem rekomendasi telah berfungsi sebagaimana mestinya. Dengan memanfaatkan Content-Based Filtering berdasarkan karakteristik pengguna dan atribut konten aktivitas, sistem berhasil menyediakan personalisasi aktivitas olahraga yang terukur dan terurut berdasarkan tingkat relevansi.

KESIMPULAN

Proyek sistem rekomendasi ini berhasil mengintegrasikan tiga tahapan utama: *clustering*, klasifikasi, dan *content-based filtering*. Pada tahap *clustering*, data pengguna berhasil dikelompokkan ke dalam kluster 'Aktif' dan 'Pasif' berdasarkan karakteristik kebugaran, yang kemudian digunakan sebagai fitur penting dalam tahap klasifikasi. Evaluasi kluster menunjukkan perbedaan profil yang jelas antar kelompok, memvalidasi proses segmentasi.

Dalam tahap klasifikasi, berbagai model machine learning dievaluasi untuk memprediksi *activity_type* yang sesuai. Hasil evaluasi menunjukkan bahwa *K-Nearest Neighbors (KNN) Classifier* menjadi model terbaik dengan akurasi 0.4875, mengungguli *Decision Tree*, *Random Forest*, dan *Deep Learning Model* yang memiliki akurasi lebih rendah (0.43, 0.4495, dan 0.202 secara berurutan). Kinerja KNN yang superior mengindikasikan kemampuannya yang lebih baik dalam menangkap pola dalam data yang diberikan.

Selanjutnya, sistem rekomendasi berbasis *content-based filtering* berhasil diimplementasikan untuk memberikan tiga rekomendasi aktivitas lain yang mirip dengan aktivitas utama hasil prediksi klasifikasi. Fitur seperti *intensity_encoded*, *fitness_level*, *calories_burned*, dan *daily_steps* digunakan untuk menghitung kemiripan antar aktivitas menggunakan *Cosine Similarity*. Studi kasus yang dilakukan menunjukkan bahwa sistem dapat memprediksi aktivitas utama (misalnya, *Cycling*) dan memberikan rekomendasi relevan (misalnya, *Running*, *Basketball*, *Tennis*) berdasarkan *input* profil pengguna.

Secara keseluruhan, proyek ini berhasil menunjukkan implementasi alur *Data Science* yang komprehensif, mulai dari pemahaman data, pra-pemrosesan, *clustering*, klasifikasi, hingga pengembangan sistem rekomendasi berbasis konten yang terintegrasi penuh dalam aplikasi web *Flask*. Proyek ini membuktikan kemampuan penerapan konsep-konsep *machine learning* yang diperoleh dari *coding camp* untuk memecahkan masalah nyata dalam personalisasi rekomendasi kebugaran.

REFERENSI

- Aisyiah, J., & Cahyani, L. (2024). Sistem Rekomendasi Program Studi Menggunakan Metode Hybrid Recommendation (Studi Kasus: MAN Sumenep). *Jurnal Eksplora Informatika*, 12(1), 59–72. <https://doi.org/10.30864/eksplora.v12i1.992>
- Ardilia Putri, K., & Hapsari Handayani, H. (2024). Perbandingan Hasil Klasifikasi Tutupan Lahan Menggunakan Metode Artificial Neural Network (ANN), Support Vector Machine (SVM), dan Random Forest (RF) dengan Bahasa Pemrograman R Comparison of Land Cover Classification Results Using Artificial Neural Network (ANN), Support Vector Machine (SVM), And Random Forest (RF) Methods with R Programming Language. 19(2), 349–360.
- Arif, T., & Baharudin. (2025). Jurnal Keperawatan Muhammadiyah sebagai Keamanan Inovasi Kesehatan Digital sebagai Perwujudan Artificial Intelligence (AI) dalam Penyebaran Informasi Kesehatan. In *Jurnal Keperawatan Muhammadiyah* (Vol. 10, Issue 1). <https://journal.um-surabaya.ac.id/index.php/JKM>
- Brons, A., Wang, S., Visser, B., Kröse, B., Bakkes, S., & Veltkamp, R. (2024). Machine Learning Methods to Personalize Persuasive Strategies in mHealth Interventions That Promote Physical Activity: Scoping Review and Categorization Overview. In *Journal of Medical Internet Research* (Vol. 26). JMIR Publications Inc. <https://doi.org/10.2196/47774>
- Cholil, S. R., Handayani, T., Prathivi, R., & Ardianita, T. (2021). IJCIT (Indonesian Journal on Computer and Information Technology) Implementasi Algoritma Klasifikasi K-Nearest Neighbor (KNN) Untuk Klasifikasi Seleksi Penerima Beasiswa. In *IJCIT (Indonesian Journal on Computer and Information Technology)* (Vol. 6, Issue 2).
- DBS Foundation. (2025). About us. Retrieved June 17, 2025. <https://www.dbs.com/foundation/>.
- Dwi Putranto, A., & Yasin Efendi, M. (2024). Pentingnya Kebugaran Jasmani untuk Meningkatkan Kekuatan Tubuh dalam Mengikuti Kegiatan Belajar dan Mengajar Di SMP Dharma Karya UT. <https://jurnal.umj.ac.id/index.php/SEMNASFIP/article/viewFile/23593/11083>
- Fajriansyah, M., Adikara, P. P., & Widodo, A. W. (2021). Sistem Rekomendasi Film Menggunakan Content Based Filtering (Vol. 5, Issue 6). <http://j-ptiik.ub.ac.id>
- Fauziyah, S., & Sugiarti, Y. (2022). Literature Review: Analisis Metode Perancangan Sistem Informasi Akademik Berbasis Web. *Jurnal Ilmiah Ilmu Komputer Fakultas Ilmu Komputer*, 8(2). <http://ejournal.fikom-unasman.ac.id>
- Geralda Adhianto, K., Ahmad Arief, N., & Studi Pendidikan Jasmani Kesehatan dan Rekreasi, P. (2023). HUBUNGAN AKTIVITAS FISIK TERHADAP KEBUGARAN JASMANI PESERTA DIDIK SEKOLAH MENENGAH PERTAMA THE RELATIONSHIP OF PHYSICAL ACTIVITY TO PHYSICAL FITNESS OF JUNIOR HIGH SCHOOL STUDENTS. *Jambura Journal of Sports Coaching*, 5(2).
- Jaya Purnama, J., Mahmud Nawawi, H., & Rosyida, S. (2020). KLASIFIKASI MAHASISWA HER BERBASIS ALGORITMA SVM DAN DECISION

- TREE*. 7(6).
<https://doi.org/10.25126/jtiik.202073080>
- Mulyatno, Pujitresnani, A., Kristioko Legowo, D., Firman, A., & Ryan Mahendra, A. (2024). *Pemberdayaan Siswa Sekolah Menengah Kejuruan Melalui Pelatihan Pengenalan Machine Learning*.
<https://jurnalpengabdianmasyarakatbangsa.com/index.php/jpmba/index>
- Putriana, P., Suarna, N., Prihartono, W., Informatika, T., Ikmi, S., Jln, C., No, P., Karyamulya, B., Kota, K., Jawa, C., & Indonesia, B. (2023). ANALISIS CLUSTERING PRESTASI ATLET PADA BERBAGAI CABANG OLAHRAGA MENGGUNAKAN ALGORITMA K-MEANS (STUDI KASUS: DINAS PEMUDA DAN OLAHRAGA KABUPATEN CIREBON). In *Jurnal Mahasiswa Teknik Informatika* (Vol. 7, Issue 6).
- Rilastiyo Budi, D., Deri Listiandi, A., Widanita, N., & Anggraeni, D. (2020). *Edukasi Kesehatan Mengenai Aktivitas Olahraga dan Pola Istirahat Bagi Wanita Usia Dewasa dan Lansia*.
- Rohmah, L., & Nur Muhammad, H. (2021). *Tingkat Kebugaran Jasmani dan Aktivitas Fisik Siswa Sekolah*.
<https://ejournal.unesa.ac.id/index.php/jurnal-pendidikan-jasmani>
- Soebroto, A. A. (2019). *Buku Ajar AI, Machine Learning & Deep Learning*.
<https://www.researchgate.net/publication/348003841>
- Subiyantoro, S. (2024). *2024-Buku-Ajar-Artificial-Intelligence*.
- Sukma Pradana, D., Prajoko, & Hartawan, G. P. (2022). *Perbandingan Algoritma Content-Based Filtering dan Collaborative Filtering dalam Rekomendasi Kegiatan Ekstrakurikuler Siswa*.