

REKOMENDASI PEKERJAAN BIDANG EKONOMI : SISTEM REKOMENDASI MENGGUNAKAN CONTENT BASED

Abdur Ro'uf ^{1*}; Hasyim Asy'ari ²; Maysas Yafi Urrohman ³; Febriane Devi Rahmawati ⁴

Informatika^{1,2,3,4}
Institut Teknologi Dan Bisnis Widya Gama Lumajang, Indonesia^{1,2,3,4}
info@itbwigalumajang.ac.id^{1,2,3,4}
abdurrouf.ar34@gmail.com^{1*}



Ciptaan disebarluaskan di bawah Lisensi Creative Commons Atribusi-Non Komersial 4.0 Internasional.

Abstract— The recommendation system was developed to assist students of the Institut Teknologi dan Bisnis Widya Gama Lumajang, particularly those from the Faculty of Economics and Business, in determining their preferred career options. This system helps students by providing various job references that match their individual criteria. The data was collected from a tracer study, which includes information such as academic grades, non-academic achievements, job positions, company names, salaries received. From the total dataset, 1,120 records were deemed valid and used in the research process. The aim of this research is to assist students by providing job recommendations based on similar criteria between current students and alumni. The method applied in this study is quantitative experimental research based on data mining, with the main approach being Content-Based filtering and the MLP (Multi-Layer Perceptron) Classifier algorithm. The data was split into two parts: 65% for training and 35% for testing. This division aims to allow the model to learn from most of the data while also being tested for accuracy using unfamiliar data. The recommendation model was developed using the MLP Classifier algorithm with a hidden_layer_size configuration of 100 neurons and a max_iter of 200 iterations. For the initial test, 10 sample data points were used to evaluate the model's performance. During training, the loss value was monitored to assess how well the model understood the data and adjusted its internal weights. With this configuration, the system is expected to provide accurate job recommendations based on the user's profile and academic history.

Keywords: Content-Based, Economics, MLP Classifier, Recommendation System.

Abstrak— Sistem rekomendasi dibangun untuk memudahkan mahasiswa Institut Teknologi Dan Bisnis Widya Gama Lumajang khususnya bagi mahasiswa fakultas ekonomi dan bisnis dalam menentukan pekerjaan pilihan mereka sehingga mahasiswa terbantu oleh sistem ini dengan berbagai referensi pekerjaan yang sesuai dengan kriteria yang dimiliki oleh mahasiswa. Dengan mengumpulkan data dari data tracer study yang mencakup beberapa informasi seperti nilai akademik, prestasi non-akademik, jabatan pekerjaan, nama perusahaan tempat bekerja, serta besaran gaji yang diterima. Dari total data, sebanyak 1.120 data dinyatakan valid dan digunakan dalam proses penelitian. Tujuan dari penelitian ini adalah membantu mahasiswa dalam bentuk rekomendasi pekerjaan yang didasari pada kriteria yang sama pada mahasiswa dan alumni, metode yang digunakan dalam penelitian ini adalah kuantitatif eksperimental berbasis data mining dengan pendekatan utama yang digunakan ialah Content-Based dan algoritma MLP (Multi Layer Perceptron) Classifier. Data ini dibagi menjadi dua bagian, 65% untuk pelatihan dan 35% untuk pengujian. Pembagian ini bertujuan agar model dapat belajar dari sebagian besar data, dan teruji keakuratannya dengan data yang belum dikenalnya. Model rekomendasi dikembangkan menggunakan algoritma MLP Classifier dengan konfigurasi hidden_layer_size sebanyak 100 neuron dan max_iter sebanyak 200 iterasi. Untuk uji awal, digunakan 10 sampel data untuk melihat kinerja model. Selama pelatihan, nilai loss dipantau untuk menilai sejauh mana model memahami data dan menyesuaikan bobot internalnya. Dengan konfigurasi ini, diharapkan sistem dapat memberikan rekomendasi pekerjaan yang tepat berdasarkan profil dan riwayat akademik pengguna.

Kata kunci: Content Based, Ekonomi, MLP Classifier, Sistem Rekomendasi.

PENDAHULUAN

Pada tahun 2024 menegaskan adanya mismatch yang begitu besar baik secara vertikal maupun horizontal antara sistem pendidikan dan dunia kerja, mismatch vertikal adalah pendidikan yang ditempuh terlalu tinggi atau terlalu rendah dengan pekerjaan yang tersedia (Graha Nusantara, 2025). Survei Kita Lulus dan Populix tahun 2024 menambahkan laporan bahwa 46% perusahaan menyatakan kesulitannya dalam menemukan tenaga kerja yang sesuai bidang keilmuannya, sebagian besar hasil survei menyebutkan rendahnya keterampilan teknis dan lemahnya soft skill sebagai hambatan utama. Pekerjaan di bidang ekonomi menjadi profesi yang masih populer di kalangan mahasiswa saat ini, terdapat 739 (tujuh ratus tiga puluh sembilan) (Jobstreet, 2025) lowongan pekerjaan di bidang ekonomi yang ada di Indonesia menjadikan informasi tersebut banyak dikunjungi oleh mahasiswa terlebih lagi oleh mahasiswa yang memiliki bidang ilmu ekonomi seperti rumpun ilmu ekonomi, manajemen bisnis, akuntansi bisnis dan lain-lain yang masih berhubungan dengan rumpun ilmu terapan. Informasi lowongan pekerjaan tersedia pada situs web yang menyediakan 739 (tujuh ratus tiga puluh sembilan) dan tentunya mereka harus bersaing satu sama lain untuk dapat diterima, jika lamaran pekerjaan pada suatu perusahaan di tolak maka harus mencari kembali atau harus menaruh kembali lowongan pekerjaan pada perusahaan tentu ini akan menghabiskan banyak waktu untuk menemukan perusahaan yang sesuai dengan keinginannya.

Dalam hal lulusan dengan pendidikan sarjana strata 1 (satu) khususnya bidang ekonomi membutuhkan sebuah sistem yang dapat membantu mereka dalam hal mencari informasi pekerjaan yang sesuai dengan kriteria dari lulusan pendidikan sarjana strata 1 (satu) dalam mencari pekerjaan. Berdasarkan data yang disajikan oleh halaman blog post Databoks kata data (Databoks Kata Data, 2021) menjelaskan bahwa terdapat sarjana yang menganggur hampir satu juta di tahun 2021 hal ini menjadi peringatan penting bagi pemerintah untuk mencegah agar tidak semakin bertambah. Dalam membangun profesi jalan yang paling tepat ialah dengan membangun profesi sesuai dengan bidang ilmunya dikarenakan kemampuan dan skil sudah terbentuk semenjak di bangku perkuliahan. Penelitian ini dilakukan dengan mengambil data mahasiswa Institut Teknologi Dan Bisnis Widya Gama Lumajang yang sudah tersedia dan dalam memonitoring lulusan dalam laporan tracer study untuk mengetahui

sebaran lulusan baik dalam pekerjaan, bidang pekerjaan, posisi dan gaji.

Penelitian yang dilakukan oleh (Nurfalah et al., 2022) mengembangkan sistem rekomendasi yang dapat membantu pengguna dalam menemukan event online yang relevan sesuai dengan preferensi mereka, menggunakan algoritma cosine similarity, data dibagi 70% training, 30% testing mendapatkan hasil precision 83% recall 71% dan F1 score 77%. sistem rekomendasi lowongan pekerjaan yang personal dan relevan (Fitria et al., 2024) untuk membantu pencari kerja menemukan pekerjaan yang sesuai dengan kualifikasi mereka sekaligus mengurangi kesenjangan antara pendidikan dan kebutuhan pasar kerja, algoritma yang digunakan yaitu cosine similarity dengan hasil sistem dapat menampilkan 10 rekomendasi teratas dengan rata-rata presisi 53%. Dalam penelitian ini (Raharjo et al., 2022) sistem rekomendasi yang dapat menemukan pekerja dengan perusahaan secara efisien dengan mencocokkan preferensi pekerja terhadap lowongan kerja dengan memanfaatkan algoritma *cosine similarity*, di penelitian lain yang dilakukan oleh (Levid et al., 2025) mempermudah pencari kerja dengan preferensi dan kompetensi di era digital. Sistem rekomendasi ini dibangun dengan menggunakan sistem berbasis konten (Yusuf & Cherid, 2021). Sistem rekomendasi mempunyai beberapa macam diantaranya sistem rekomendasi berbasis collaborative, sistem rekomendasi berbasis content, sistem rekomendasi berbasis demographic dan sistem rekomendasi berbasis knowledge (Muhammad Alkaff et al., 2020) pada penelitian ini sistem rekomendasi yang digunakan ialah sistem rekomendasi berbasis konten (Humairo et al., 2023), pada penelitian yang berfokus pada sistem rekomendasi berbasis konten sangat sesuai (Fajriansyah et al., 2021) dikarenakan model sistemnya mencari kemiripan dengan item (Permana et al., 2024) yang sudah ada, data lulusan sebelumnya yang telah dikumpulkan melalui data tracer study perguruan tinggi Institut Teknologi Dan Bisnis Widya Gama Lumajang.

Tujuan peneliti membangun sistem rekomendasi ini ialah untuk membantu lulusan Fakultas Ekonomi Dan Bisnis menemukan pekerjaan yang sesuai dengan profil akademik dan non akademik mereka, dalam penelitian ini juga untuk menerapkan algoritma Multi-Layer Perceptron (MLP) Classifier untuk mengklasifikasikan dan merekomendasikan jenis pekerjaan berdasarkan kemiripan data dengan lulusan sebelumnya dan mengurangi kesenjangan antara lulusan dan dunia kerja bagi lulusan Fakultas Ekonomi Dan Bisnis dengan menyediakan sistem

pendukung keputusan berbasis data lulusan yang valid dan relevan.

BAHAN DAN METODE

Pada penelitian ini jenis dan pendekatan dilakukan dengan metode kuantitatif eksperimental berbasis data mining (Yanisa Putri et al., 2024) yang didapatkan dari data mulai dari lulusan tahun 2019 sampai dengan lulusan tahun 2023 dari fakultas ekonomi dan bisnis Institut Teknologi Dan Bisnis Widya Gama Lumajang tersebar di berbagai wilayah melalui tracer study yang ada pada bidang alumni dan kerjasama, pengumpulan data menggunakan teknik *cluster sampling* yaitu mengelompokkan data dengan kriteria alumni dengan data perusahaan lengkap dan data perusahaan tidak lengkap, dari total data sebanyak 1.120 data dan data yang digunakan yaitu sebanyak 1000 data.

Pendekatan pada pendekatan sistem rekomendasi menggunakan algoritma Content-Based Similarity (Huda et al., 2022) dan dipadukan dengan menambahkan teknik *machine learning* (MLP Classifier). Dalam membangun serta mengevaluasi sistem yang diharapkan mampu merekomendasikan pekerjaan kepada lulusan selanjutnya dengan mencari kesamaan pada data profil dan riwayat pendidikan maupun non formal tentu dengan melewati beberapa tahapan. Yang pertama pengumpulan data yang digunakan dalam penelitian ini yaitu 1.100 data lulusan yang mencakup nilai akademik, nilai non akademik seperti yang ada pada surat keterangan pendamping ijazah atau SKPI, posisi atau jabatan dan perusahaan selanjutnya pra-pemrosesan data dengan membersihkan data dan menambahkan atribut level jabatan, skill dan level tingkat kecocokan. Tahapan selanjutnya ialah percobaan model yang dibagi menjadi training 65% dan testing 35% model tersebut dilatih dengan menggunakan MLP Classifier (Musto et al., 2018).

A. Pengumpulan Data Set

Pengumpulan data set sebagai kebutuhan awal untuk mendukung terbentuknya sistem rekomendasi dan sebagai pendukung dalam melatih algoritma yang digunakan. Data yang dapat dikumpulkan dalam proses ini terdiri dari lulusan seperti nilai akademik, nilai non akademik, nama perusahaan bekerja, jabatan serta daftar gaji, semua data ini diperoleh dari data yang diambil dari data tracer study dan juga mengirimkan google form untuk mendapatkan data dari lulusan. Data yang dikumpulkan mulai dari lulusan 2019 sampai

dengan lulusan 2023 dengan menggunakan teknik *cluster sampling* data dengan populasi 1.120 dan dapat digunakan sebanyak 1000 data seperti pada tabel 1 dibawah ini.

Tabel 1. Data Set

No. ID	Jabatan	Nilai Non Akademik	Nilai Akademik	Level kecocokan
1.	Admin Kredit	4	84	4
2.	Admin Keuangan	4.8	90	5
3.	Teller	4.7	87	4
4.	Admin Perpajakan	4.6	86	4
5.	Admin Pembayaran	3.8	83	3
6.	Karyawan Produksi	3.3	81	3
7.	Kasir	2.8	78	2.7
8.	HRD	3.7	82	3.4

Source : (Hasil Penelitian, 2025)

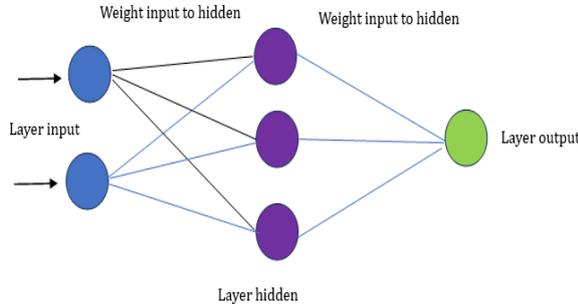
Ada beberapa perlakuan khusus dalam pengumpulan data set sehingga pada saat digunakan memudahkan model sistem untuk mempelajarinya, dalam penentuan nilai non akademik nilai ditentukan dengan bidang ilmu yang ditekuninya. Nilai non akademik dikumpulkan dari kegiatan di luar pembelajaran selama di bangku perkuliahan dengan memberikan level tingkat kecocokan untuk memudahkan score sampai pada level berapa seperti level 1 (satu) sampai dengan 5 (lima), nilai akademik di ambil dari data nilai beberapa mata kuliah yang mendukung pekerja tersebut dengan jabatan yang saat ini mereka tempati jika terdapat nilai akademik yang nilainya bagus maka pada tingkat level kecocokan juga akan menghasilkan score yang tinggi pula, penentuan level kecocokan pada jabatan yang saat ini di tempatnya mempengaruhi beberapa aspek seperti pada nilai non akademik dan juga nilai akademik

Setelah semua data terkumpul pada tahap ini dataset diubah dalam bentuk csv sebagai bentuk file yang akan di input.

B. Metodologi

Content-based similarity digunakan dalam penelitian ini karena kemampuan dalam find of the similar data (Hong et al., 2013) yang terdapat dalam MLP Classifier dengan menggunakan fitur library sklearn dan algoritma ini sudah terhubung dengan Neural Network. Tugas utama dari MLP Classifier yaitu untuk klasifikasi data (He et al., 2017). Sistem rekomendasi tepat sekali digunakan dalam membangun sistem pendukung keputusan, sistem

ini bekerja dengan memanfaatkan kecocokan data set dan juga memanfaatkan peringkat pada hasil training (Purkar et al., 2021). Berikut gambar ilustrasi MLP Classifier.



Source : (Hasil Penelitian, 2025)
Gambar 1. MLP Classifier

Content-based similarity memiliki sistem yang membangun hubungan dua arah berdasarkan kesamaan data (Crismastiana Koloman et al., 2023) sehingga dapat bekerja secara optimal pada sistem rekomendasi.

$$sim(i, j) = \frac{\sum_{u=1}^m R_{ui} - \underline{R}_i \quad R_{uj} - \underline{R}_j}{\sum_{u=1}^m R_{ui} - \underline{R}_i \quad \sum_{u=1}^m R_{uj} - \underline{R}_j} \quad (1)$$

- Sim (i,j) = nilai similarity
- m = total user pemberi rating item
- \underline{R}_i dan \underline{R}_j = rata-rata item i dan j
- R_{u_i} dan R_{u_j} = rating yang diberikan

Sistem rekomendasi adalah sebuah sistem yang memberikan suatu informasi sebagai pendukung pengambilan keputusan dengan menampilkan data informasi kepada pengguna berdasarkan kemiripan data, minat pengguna dan juga relevansi (Purkar et al., 2021), sudah banyak contoh penggunaan sistem rekomendasi yang diaplikasikan yang bertujuan memberikan informasi seperti pada artikel, film, kuliner dan juga rekomendasi pekerjaan (Jepriana & Hanief, 2020).

Precision terdiri dari item yang digunakan dalam daftar rekomendasi, sebagaimana yang dinyatakan dalam persamaan yang mengukur tingkat item, sedangkan recall dihitung berdasarkan jumlah item yang dikonsumsi dalam daftar rekomendasi dari total jumlah data yang digunakan (Silveira et al., 2019) , persamaan dari precision, recall dan f1-score sebagai berikut.

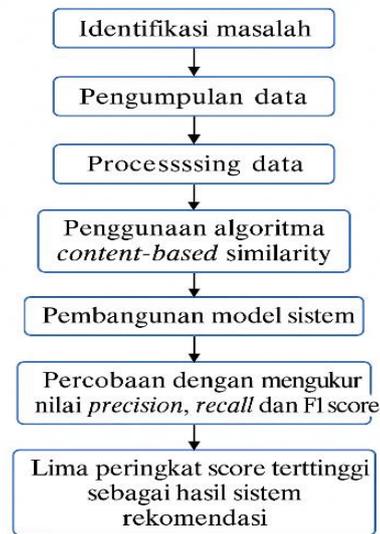
$$precision = \frac{|C_u \cap R_u|}{|R_u|} \quad (2)$$

$$recall = \frac{|C_u \cap R_u|}{|C_u|} \quad (3)$$

$$F1-score = 2x \frac{precision \times recall}{precision + recall} \quad (4)$$

- C_u = item yang disukai
- R_u = item yang direkomendasikan

Dalam percobaan suatu sistem rekomendasi sistem akan bekerja memberikan rekomendasi berdasarkan hasil tingkat kemiripan data atau peringkat dari percobaan (Yanti & Yahfizham, 2024) yang dilakukan. Berikut adalah alur penelitian :



Source : (Hasil Penelitian, 2025)
Gambar 2. Alur Penelitian

Dalam penelitian ini data yang digunakan dengan memiliki atribut yang dilengkapi dengan nilai akademik dan non akademik dimana nilai akademik di seleksi hanya pada nilai mata kuliah – mata kuliah tertentu sebagai poin besar kemudian untuk nilai non akademik di kumpulkan dengan menyeleksi nilai – nilai yang berhubungan dengan skill di bidang ekonomi dan masih berhubungan dengan nilai akademik. Tujuan dari nilai non akademik adalah semata-mata untuk poin tambahan bagi pengguna dalam mengukur implementasi ilmu yang sudah diperoleh selama menempuh pendidikan di perguruan tinggi.

Dalam penelitian ini, peneliti berhasil mengumpulkan sebanyak 2.000 (dua ribu) data lulusan yang diperoleh melalui tracer study dari Fakultas Ekonomi dan Bisnis di Institut Teknologi dan Bisnis Widya Gama Lumajang. Data tersebut mencakup informasi penting seperti nilai akademik,

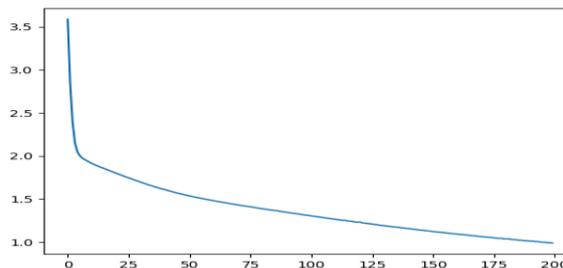
nilai non-akademik, posisi atau jabatan yang ditempati, nama perusahaan tempat bekerja, serta kisaran gaji yang diterima oleh para lulusan. Setelah proses pengumpulan data selesai, dilakukan tahapan pra-pemrosesan atau preprocessing untuk membersihkan data dari nilai yang tidak relevan atau duplikat, serta menyesuaikan format agar sesuai dengan kebutuhan sistem. Dari total 2.000 data yang terkumpul, sebanyak 1.120 (seribu seratus dua puluh) data dinyatakan valid dan layak digunakan untuk keperluan pelatihan dan pengujian model.

Dalam pembagian data, peneliti menggunakan metode pembagian proporsional, yaitu sebesar 35% data digunakan sebagai data testing atau pengujian dan 65% sisanya digunakan sebagai data training atau pelatihan. Pembagian ini bertujuan agar sistem dapat dilatih secara optimal dan juga diuji secara menyeluruh untuk menilai performanya. Selain itu, untuk proses evaluasi awal, peneliti mengambil 10 (sepuluh) data sebagai sampel untuk dilakukan percobaan model. Pada proses pelatihan model, parameter `max_iter` atau jumlah maksimum iterasi ditentukan sebanyak 200 kali, artinya proses pelatihan akan dilakukan hingga 200 siklus atau sampai model mencapai konvergensi.

Sementara itu, arsitektur jaringan yang digunakan dalam MLP (Multi-Layer Perceptron) Classifier terdiri dari `hidden_layer_size` (Ginting & Pratama, 2023) sebanyak 100 neuron pada satu lapisan tersembunyi. Penggunaan neuron dalam Multi-Layer Perceptron ditentukan sendiri sesuai dengan kondisi model yang telah dibangun, penentuan neuron dalam sistem rekomendasi ialah menandakan bahwa model tersebut berapa kali diuji coba dengan tujuan model mempelajari sistem dengan ketentuan – ketentuan. Konfigurasi ini dipilih dengan tujuan agar model memiliki kapasitas yang cukup dalam mempelajari pola data dan menghasilkan rekomendasi yang akurat. Selama pelatihan, peneliti juga memantau nilai loss atau kesalahan pada tiap iterasi untuk mengevaluasi sejauh mana model belajar dari data dan menyesuaikan bobotnya.

Dengan metode ini, diharapkan sistem rekomendasi dapat memberikan hasil yang optimal dan bermanfaat bagi lulusan dalam menemukan pekerjaan yang sesuai dengan profil mereka, memanfaatkan sistem rekomendasi dengan berbasis konten adalah salah satu metode yang sesuai dan dapat digunakan dikarenakan model mempelajari kesamaan data yang pengguna dengan data yang ada dalam data training. Dalam training

model machine learning menunjukkan pada gambar bahwa hasil iterasi yang didapatkan sumbu x menunjukkan jumlah epoch atau iterasi ke berapa kalinya dalam training model sistem ini kemudian sumbu y menunjukkan nilai loss, nilai loss dalam training ini menunjukkan menunjukkan ketika mendekati iterasi ke 200 (dua ratus) yaitu loss sebesar 0.69971150 artinya bahwa model *machine learning* bekerja dengan baik dan dapat digunakan. Hasil dari training dapat dilihat pada gambar dibawah ini.



Source : (Hasil Penelitian, 2025)

Gambar 3. Loss Iterasi

HASIL DAN PEMBAHASAN

Pada penelitian ini, training dilakukan untuk bertujuan melatih sistem rekomendasi dalam memberikan hasil yang terbaik dan dapat digunakan oleh pengguna menjadi dasar atas pengambilan keputusan. Dalam training ini dilakukan uji coba beberapa kali untuk mendapatkan hasil yang terbaik.

Tujuan dalam uji coba ini ialah untuk mengajari sistem dalam mengambil keputusan dari data dan dapat membuat keputusan atau prediksi tanpa diprogram secara eksplisit, mengukur akurasi rekomendasi untuk mengetahui seberapa tepat rekomendasi yang diberikan oleh sistem dibandingkan dengan preferensi nyata pengguna. Kualitas model akan dinilai menggunakan matrik evaluasi seperti precision, recall, f1-score atau mean average precision (Pratama & Hasrullah, 2025).

Dalam percobaan pertama model neural network pertama menggunakan arsitektur yang memiliki lapisan tersembunyi atau hidden layers dengan jumlah neuron layer 1 300, layer 2 200, layer 3 100, lapisan di antara input dan output bertugas mempelajari representasi atau pola dari data semakin banyak dan kompleks hidden layersnya maka semakin baik sistem untuk mengenali polanya, dan iterasi pada percobaan pertama menggunakan 200 iterasi artinya model dilatih selama 200 kali melewati proses update

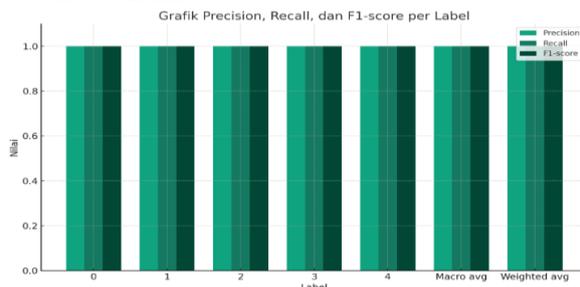
arsitektur yang memiliki lapisan 300, 200, 100 hasil yang didapatkan pada percobaan pertama yaitu precision 1.00, recall 1.00 dan f1-score 1.00 seperti pada tabel 1.

Tabel 2. Percobaan Pertama

	Precisio n	Recal l	F1- score	suppor t
0	1.00	1.00	1.00	34
1	1.00	1.00	1.00	5
2	1.00	1.00	1.00	45
3	1.00	1.00	1.00	156
4	1.00	1.00	1.00	40
Accuracy			1.00	280
Macro avg	1.00	1.00	1.00	280
Weighted avg	1.00	1.00	1.00	280

Source : (Hasil Penelitian, 2025)

Dari hasil ini tentu sempurna namun dalam praktik machine learning hasil yang sempurna patut dicurigai dan perlu adanya analisis lebih lanjut. Ketidakwaajaran hasil percobaan dapat dilihat pada tabel dibawah ini.



Source : (Hasil Penelitian, 2025)

Gambar 4. Grafik Hasil Model Pertama

Grafik bar menunjukkan nilai precision, recall dan F1-score untuk masing-masing label (termasuk *macro avg* dan *weighted avg*). Karena semua nilainya sempurna (1.00), tinggi batang pada grafik semuanya mencapai nilai maksimum.

Dalam percobaan pertama menghasilkan nilai evaluasi yang sempurna dan itu tidak wajar dalam hal ini terdapat beberapa kesalahan yaitu data testing terlalu mirip atau sama dengan data training menyebabkan data latih dan data uji bercampur sehingga model bisa menghafal daripada belajar, kemudian data *test_size* memiliki ukuran kecil artinya harus diperbesar, maka dari itu pada percobaan kedua data testing akan dipisah dan tidak dibuat begitu mirip dengan data training, kemudian dalam *test_size* akan ditambah yaitu menjadi 35% artinya dalam percobaan sebelumnya naik sebesar 10% pada percobaan kedua dengan arsitektur layers 1 100, layers 2 100, layer 3 100 dan dengan iterasi sebanyak 200 kali. Hasil analisis anomali dari model pertama yaitu kemungkinan

besar terjadi data leakage, dimana data testing terlalu mirip atau bahkan sama dengan data training. Hal ini membuat model “menghafal” data alih-alih mempelajarinya. Selain itu, ukuran *test_size* terlalu kecil, sehingga tidak cukup merepresentasikan variasi data. Oleh karena itu, dilakukan percobaan kedua dengan pemisahan data yang lebih baik dan penambahan *test_size* menjadi 35%.

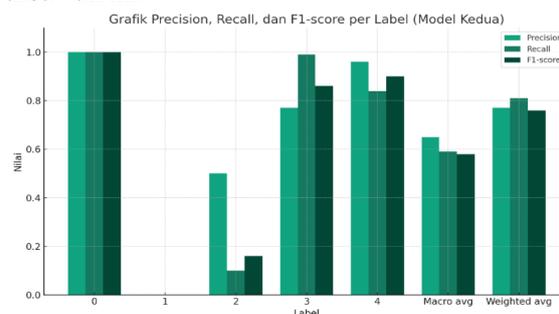
Secara teknis, nilai ini mengindikasikan bahwa model dapat memprediksi semua kelas dengan tepat 100%, tanpa ada kesalahan klasifikasi sama sekali. Namun, hasil ini mencurigakan dan tidak realistis dalam konteks *machine learning* yang umum karena model hampir pasti mengalami data *leakage* atau *overfitting*. Kemungkinan penyebabnya ialah data *training* dan *testing* terlalu mirip, atau bahkan terjadi percampuran data (*leakage*), sehingga model tidak belajar secara generalisasi, tapi justru menghafal data. Proporsi data uji (*test_size*) terlalu kecil, sehingga tidak representatif terhadap variabilitas data keseluruhan.

Tabel 3. Percobaan Kedua

	Precisio n	Recal l	F1- score	suppor t
0	1.00	1.00	1.00	39
1	0.00	0.00	0.00	6
2	0.50	0.10	0.16	63
3	0.77	0.99	0.86	221
4	0.96	0.84	0.90	63
Accuracy			0.81	392
Macro avg	0.65	0.59	0.58	392
Weighted avg	0.77	0.81	0.76	392

Source : (Hasil Penelitian, 2025)

Dari hasil percobaan kedua matrik evaluasi menunjukkan hasil yang berbeda dengan percobaan pertama, percobaan kedua menunjukkan precision 0.77, recall 0.81 dan f1-score 0.76 dengan tingkat akurasi yang dihasilkan 81%, penyajian dalam bentuk tabel dapat dilihat dibawah ini.



Source : (Hasil Penelitian, 2025)

Gambar 5. Grafik Hasil Model Kedua

Grafik diatas menunjukkan nilai *precision*, *recall* dan *f1-score* dimana terlihat perbedaan dengan percobaan model pertama yaitu adalah tinggi dari batang grafik beresonansi memperlihatkan kinerja model yang saling berkesinambungan.

Artinya dari hasil yang didapatkan *precision* 0.77 menandakan bahwa sistem masih relevan untuk pengguna dan sistem berpeluang sedikit memberikan rekomendasi yang salah, *recall* 0.81 sistem berhasil merekomendasikan semua item yang mungkin disukai pengguna *f1-score* 0.76 menerangkan bahwa hasil gabungan dari *precision* dan *recall* dengan hasil ini menunjukkan sistem bekerja dengan cukup baik dan cukup seimbang antara tidak salah dalam memberikan rekomendasi dan tidak melewatkan item yang bagus. Dengan tingkat akurasi setinggi 81% sistem ini sudah dapat memprediksi dengan baik dapat dikatakan model yang digunakan sudah menggambarkan prediksi yang tepat.

Terdapat anomali signifikan pada kelas 1 dan kelas 2 Kelas 1 (*support*: 6) memiliki *precision*, *recall*, dan *f1-score* semuanya 0.00. Ini mengindikasikan bahwa model tidak mampu mengenali atau memprediksi kelas ini sama sekali. Kelas 2 memiliki *recall* sangat rendah (0.10) dan *f1-score* 0.16, yang menunjukkan bahwa model hanya mampu mengenali sebagian kecil dari data kelas ini. Hal ini dapat disebabkan oleh jumlah data yang tidak seimbang (kelas minoritas seperti kelas 1 jauh lebih sedikit dari kelas mayoritas seperti kelas 3). Fitur input untuk kelas tersebut kurang representatif, sehingga sulit dikenali oleh model. Model terlalu bias terhadap kelas mayoritas, menyebabkan kelas minoritas terabaikan.

Secara umum, sistem mampu memberikan rekomendasi dengan akurasi cukup tinggi (81%), menandakan bahwa model memiliki kemampuan prediksi yang cukup baik dan layak digunakan. *Precision* 0.77 menunjukkan bahwa sebagian besar rekomendasi yang diberikan oleh sistem adalah relevan, meskipun masih ada peluang terjadinya kesalahan rekomendasi. *Recall* 0.81 menunjukkan bahwa sistem cukup berhasil dalam menemukan item yang benar-benar relevan dan sesuai dengan kebutuhan pengguna. *F1-score* 0.76 menggambarkan keseimbangan antara kemampuan model dalam memberikan hasil yang benar (*precision*) dan kelengkapan (*recall*). Nilai evaluasi per kelas juga menunjukkan variasi performa kelas 1 adalah anomali karena memiliki nilai *precision*, *recall*, dan *f1-score* = 0.00, yang berarti model gagal mengenali kelas ini, kelas 3 memiliki performa paling stabil dan tinggi (*f1-score*

0.86), kemungkinan karena memiliki jumlah data (*support*) paling besar, yaitu 221 data. Kelas 1 adalah anomali yang perlu diperhatikan secara khusus karena model sama sekali tidak mampu memprediksi kelas tersebut (semua nilai metrik = 0). Hal ini bisa disebabkan oleh jumlah data terlalu sedikit (hanya 6 data/sampel). Distribusi tidak seimbang (*imbalanced class*), sehingga model tidak cukup "belajar" dari data tersebut. Kemungkinan perlu dilakukan teknik *oversampling*, *undersampling*, atau penyesuaian bobot kelas (*class weight*) untuk memperbaiki performa di kelas minoritas.

Dengan adanya pembagian data yang lebih baik pada percobaan kedua, performa model menjadi lebih realistis. Namun, perlu perhatian khusus terhadap anomali prediksi pada kelas dengan jumlah data kecil yang gagal diklasifikasi dengan baik. Hal ini penting agar sistem rekomendasi tidak bias dan tetap adil bagi semua jenis data pengguna.

KESIMPULAN

Penelitian ini mengenai sistem rekomendasi dengan model yang sudah ditentukan dan diuji sebanyak dua kali, dalam percobaan pertama dengan *test_size* yaitu sebanyak 25% dan iterasi sebanyak 200 (dua ratus) kali memberikan hasil dari matrik evaluasi *precision* 1.00, *recall* 1.00 dan *f1-score* 100 hasil sangat sempurna namun dalam *machine learning* hasil yang begitu sempurna harus analisis kembali terbukti terdapat beberapa kekurangan dalam hal ini yaitu terdapat data *testing* yang begitu mirip dengan data training dan jumlah *test_size* yang perlu ditambahkan. Percobaan kedua, menganalisis dari hasil percobaan pertama peneliti mencoba kembali sistem ini dengan *test_size* 35% dan arsitektur layer 100, 100, 100 dan iterasi sebanyak 200 kali dari hasil percobaan kedua sistem memberikan hasil *precision* 0.77, *recall* 0.81 dan *f1-score* 0.76 dengan tingkat akurasi 81%. Dari hasil ini kesimpulannya adalah model sistem dapat dilatih dengan neuron - neuron yang tepat dengan mencoba arsitektur layer yang berbeda dan jumlah *test_size* yang tidak sedikit minimal 20% sampai dengan 35% maka model tersebut terlihat sampai berapa tingkat akurasi yang dimiliki model sistem tersebut. Secara garis besar model ini sudah dapat menggambarkan sistem dengan tingkat akurasi atau banyaknya rekomendasi yang tepat namun dalam penelitian ini dapat lebih optimal dengan memastikan fitur input benar-benar mencerminkan preferensi dan karakteristik item, menambahkan

fitur interaksi pengguna seperti *click*, *rating* dan *history* menggunakan model *hybrid* dengan menggabungkan *content based* dengan *collaborative filtering* dan juga *matrix factorization*.

REFERENCE

- Crismastiana Koloman, Raihan Maulana, Raisya Dwi Zahra Putri, & Wahyu Abadi Harahap. (2023). Sistem Rekomendasi Pekerjaan di bidang IT Menggunakan Algoritma Content-Based Filtering. *Journal of Creative Student Research*, 1(6), 78–88. <https://doi.org/10.55606/jcsrpolitama.v1i6.2992>
- Databoks Kata Data. (2021, Mei). *BPS: Sarjana yang Mengganggu Hampir 1 Juta Orang pada Februari 2021*. <https://databoks.katadata.co.id/ketenagakerjaan/statistik/b52bbe8b99077f1/bps-sarjana-yang-mengganggu-hampir-1-juta-orang-pada-februari-2021>
- Fajriansyah, M., Adikara, P. P., & Widodo, A. W. (2021). *Sistem Rekomendasi Film Menggunakan Content Based Filtering*. 5(6), 2188–2199. <https://doi.org/10.37729/intek.v8i1.6286>
- Fitria, A., Zaman, S., & Yaqin, M. A. (2024). *Sistem Rekomendasi Lowongan Pekerjaan Menggunakan Content-based filtering*. 10(3), 421–427. <https://doi.org/10.26418/jp.v10i3.83801>
- Ginting, E. T. B., & Pratama, I. (2023). *Sistem Rekomendasi Jurusan SMK Menggunakan Metode Content-Based Filtering Di Kabupaten Sleman*. 3(2), 291–300. <https://doi.org/10.47233/jsit.v3i2.954>
- Graha Nusantara. (2025, July 29). *Krisis Kecocokan Pendidikan dan Dunia Kerja di Indonesia: Mengapa Banyak Lulusan Tak Terserap?* https://grahanusantara.id/krisis-kecocokan-pendidikan-dan-dunia-kerja-di-indonesia-mengapa-banyak-lulusan-tak-terserap?utm_source=chatgpt.com
- He, X., Liao, L., Zhang, H., Nie, L., Hu, X., & Chua, T.-S. (2017). *Neural Collaborative Filtering* (No. arXiv:1708.05031). arXiv. <https://doi.org/10.48550/arXiv.1708.05031>
- Hong, W., Zheng, S., Wang, H., & Shi, J. (2013). A Job Recommender System Based on User Clustering. *Journal of Computers*, 8(8), 1960–1967. <https://doi.org/10.4304/jcp.8.8.1960-1967>
- Huda, A. A., Fajarudin, R., & Hadinegoro, A. (2022). *Sistem Rekomendasi Content-based Filtering Menggunakan TF-IDF Vector Similarity Untuk Rekomendasi Artikel Berita*. *Building of Informatics, Technology and Science (BITS)*, 4(3), 1679–1686. <https://doi.org/10.47065/bits.v4i3.2511>
- Humairo, A., Herdiani, A., & Puspitasari, S. Y. (2023). *Pembangunan Recommender System Menggunakan Content Based Filtering pada Aplikasi Service Desk*. *LOGIC: Jurnal Penelitian Informatika*, 1(1), 20. <https://doi.org/10.25124/logic.v1i1.6427>
- Jepriana, I. W., & Hanief, S. (2020). *ANALISIS DAN IMPLEMENTASI METODE ITEM-BASED COLLABORATIVE FILTERING UNTUK SISTEM REKOMENDASI KONSENTRASI DI STMIK STIKOM BALI*. 9.
- Jobstreet. (2025, Mei). *Terdapat 739 lowongan pekerjaan di bidang ekonomi*. *Jobstreet*. https://id.jobstreet.com/id/ekonomi-jobs?utm_source=chatgpt.com
- Levid, J. F., Wijaya, D., Irsyad, H., & Rahman, A. (2025). *Penerapan Smart, Edas, Dan Cosine Similarity Dalam Rekomendasi Lowongan Pekerjaan Di Era Digital*. 3(3), 85–92. <https://doi.org/10.58369/biit.v2i3.128>
- Muhammad Alkaff, Husnul Khatimi, & Andi Eriady. (2020). *Sistem Rekomendasi Buku Menggunakan Weighted Tree Similarity dan Content Based Filtering*. 20(1), 193–202. <https://doi.org/DOI:10.30812/matrik.v20i1.617>
- Musto, C., Franza, T., Semeraro, G., De Gemmis, M., & Lops, P. (2018). *Deep Content-based Recommender Systems Exploiting Recurrent Neural Networks and Linked Open Data*. *Adjunct Publication of the 26th Conference on User Modeling, Adaptation and Personalization*, 239–244. <https://doi.org/10.1145/3213586.3225230>
- Nurfalah, F., Asriyanik, & Pambudi, A. (2022). *Sistem Rekomendasi Event Online Menggunakan Metode Content Based Filtering*. *Elkom: Jurnal Elektronika dan Komputer*, 15(2), 271–279. <https://doi.org/10.51903/elkom.v15i2.736>
- Permana, R. M., Hadiana, A. I., & Sabrina, P. N. (2024). *Rekomendasi Pemilihan Sepeda Motor Menggunakan Metode Content Based Filtering Dan Item Based Collaborative Filtering*. 12(2), 207–217. <https://doi.org/0.33592/jutis.v12i2.5149>
- Pratama, R. V., & Hasrullah, H. (2025). *Pengembangan Sistem Rekomendasi Buku untuk Meningkatkan Minat Baca dengan Pendekatan Hybrid Filtering*. *Jurnal Inovasi*

- Global*, 3(1), 2182–2191.
<https://doi.org/10.58344/jig.v3i1.255>
- Purkar, M., Joshi, O., Salape, A., Patil, A., Kulkarni, V., & Futane, P. (2021). Recommendation System for Workers & Customers for Informal Jobs. *SSRN Electronic Journal*.
<https://doi.org/10.2139/ssrn.3833762>
- Raharjo, P. N., Handojo, A., & Juwiantho, H. (2022). *Sistem Rekomendasi Content Based Filtering Pekerjaan dan Tenaga Kerja Potensial menggunakan Cosine Similarity*. 10(2), 1–6.
<https://doi.org/10.11591/eei.v10i5.3157>
- Silveira, T., Zhang, M., Lin, X., Liu, Y., & Ma, S. (2019). How good your recommender system is? A survey on evaluations in recommendation. *International Journal of Machine Learning and Cybernetics*, 10(5), 813–831.
<https://doi.org/10.1007/s13042-017-0762-9>
- Yanisa Putri, K. S., I Made Agus Dwi Suarjaya, & Wayan Oger Vihikan. (2024). Sistem Rekomendasi Skincare Menggunakan Metode Content Based Filtering dan Collaborative Filtering. *Decode: Jurnal Pendidikan Teknologi Informasi*, 4(3), 764–774.
<https://doi.org/10.51454/decode.v4i3.601>
- Yanti, F. R., & Yahfizham, Y. (2024). Implementasi Sistem Informasi Pengolahan Data Alumni Membantu Akreditasi dan Bursa Kerja Metode Content Based Filtering. *Journal of Information System Research (JOSH)*, 5(4), 1102–1114.
<https://doi.org/10.47065/josh.v5i4.5546>
- Yusuf, M., & Cherid, A. (2021). *Implementasi Algoritma Cosine Similarity Dan Metode TF-IDF Berbasis PHP Untuk Menghasilkan Rekomendasi Seminar*. 9(1), 8–16.