

Analisis Kemiripan Portofolio Mahasiswa Rekognisi Pembelajaran Lampau Dengan Capaian Pembelajaran Menggunakan Cosine Similarity

I Putu Gede Abdi Sudiatmika¹, Made Yessi Puspitha², Ida Ayu Gde Suwiprabayanti Putra³, Made Suci Ariantini⁴

^{1,2}Politeknik Negeri Bali, ³Universitas Udayana, ⁴INSTIKI Indonesia
e-mail: ¹sudiatmika.abdi@pnb.ac.id, ²yessipuspitha@pnb.ac.id, ³iagsuwiprabayantiputra@unud.ac.id, ⁴suci.ariantini@instiki.ac.id

Diajukan: 06 Mei 2026; Direvisi: 13 Mei 2026; Diterima: 20 Mei 2026

Abstrak

Proses pencocokan dokumen portofolio calon mahasiswa Rekognisi Pembelajaran Lampau dengan capaian pembelajaran program studi umumnya masih dilakukan secara manual oleh asesor. Kondisi ini berpotensi menimbulkan subjektivitas, perbedaan interpretasi, dan ketidakkonsistenan dalam penilaian awal. Penelitian ini bertujuan untuk menganalisis kemiripan dokumen portofolio mahasiswa Rekognisi Pembelajaran Lampau dengan capaian pembelajaran pada bidang Akuntansi Perpajakan dan Akuntansi Manajerial menggunakan metode TF-IDF dan Cosine Similarity. Data penelitian terdiri dari 40 dokumen portofolio calon mahasiswa dan 8 dokumen capaian pembelajaran program studi. Tahapan penelitian meliputi preprocessing teks, pembobotan kata menggunakan Term Frequency-Inverse Document Frequency, perhitungan kemiripan menggunakan Cosine Similarity, dan pengelompokan hasil ke dalam kategori tinggi, sedang, dan rendah. Hasil penelitian menunjukkan 17 dokumen memiliki kemiripan tinggi, 16 dokumen sedang, dan 7 dokumen rendah. Metode ini dapat membantu analisis awal portofolio secara lebih objektif, terstruktur, dan efisien.

Kata kunci: Cosine Similarity, TF-IDF, Portofolio, Rekognisi Pembelajaran Lampau, Capaian Pembelajaran.

Abstract

The process of matching portfolio documents of Recognition of Rekognisi Pembelajaran Lampau (RPL with study program learning outcomes is generally still conducted manually by assessors. This condition may lead to subjectivity, different interpretations, and inconsistency in the initial assessment. This study aims to analyze the similarity between portfolio documents of Recognition of Prior Learning students and learning outcomes in the fields of Tax Accounting and Managerial Accounting using TF-IDF and Cosine Similarity. The research data consisted of 40 applicant portfolio documents and 8 study program learning outcome documents. The research stages included text preprocessing, word weighting using Term Frequency-Inverse Document Frequency, similarity calculation using Cosine Similarity, and classification of the results into high, medium, and low categories. The results showed that 17 documents had high similarity, 16 documents had medium similarity, and 7 documents had low similarity. This method can support a more objective, structured, and efficient initial portfolio analysis

Keywords: cosine Similarity, TF-IDF, Portfolio, Recognition of Prior Learning, Learning Outcomes.

1. Pendahuluan

Rekognisi Pembelajaran Lampau merupakan salah satu mekanisme dalam pendidikan tinggi yang memberikan pengakuan terhadap capaian pembelajaran seseorang yang diperoleh melalui pendidikan formal, nonformal, informal, maupun pengalaman kerja[1]. Program ini memberikan kesempatan bagi calon mahasiswa yang telah memiliki pengalaman kerja atau kompetensi tertentu untuk melanjutkan pendidikan formal melalui proses pengakuan capaian pembelajaran yang relevan dengan program studi tujuan. Dalam regulasi pendidikan tinggi, Rekognisi Pembelajaran Lampau telah diatur melalui Permendikbudristek Nomor 41 Tahun 2021[2], sedangkan petunjuk teknis penyelenggaraannya pada perguruan tinggi akademik diperbarui melalui Keputusan Direktur Jenderal Pendidikan Tinggi, Riset, dan Teknologi Nomor 91/E/KPT/2024[3].

Dalam proses penerimaan mahasiswa Rekognisi Pembelajaran Lampau, dokumen portofolio menjadi salah satu komponen penting untuk menilai kesesuaian pengalaman calon mahasiswa dengan capaian pembelajaran program studi[4]. Portofolio dapat berisi riwayat pekerjaan, deskripsi tugas, sertifikat pelatihan, pengalaman profesional, serta dokumen pendukung lain yang menunjukkan kompetensi calon mahasiswa[5]. Pada Program Studi Akuntansi Perpajakan dan Akuntansi Manajerial, dokumen portofolio perlu dianalisis untuk mengetahui keterkaitannya dengan bidang kompetensi seperti perpajakan, pelaporan pajak, pengelolaan dokumen pajak, penyusunan laporan keuangan, penganggaran, analisis biaya, audit, dan pengendalian internal. Analisis ini penting karena hasilnya dapat membantu asesor dalam menentukan apakah pengalaman calon mahasiswa memiliki kesesuaian dengan capaian pembelajaran yang ditetapkan oleh program studi.

Permasalahan yang muncul adalah proses pencocokan antara dokumen portofolio dan capaian pembelajaran umumnya masih dilakukan secara manual. Proses manual membutuhkan ketelitian tinggi karena asesor harus membaca, memahami, dan membandingkan banyak dokumen dengan standar capaian pembelajaran. Selain memerlukan waktu, penilaian manual juga berpotensi menimbulkan perbedaan interpretasi antarasesor, terutama jika dokumen portofolio ditulis dengan gaya bahasa, istilah, dan tingkat kelengkapan yang berbeda. Kondisi tersebut menunjukkan perlunya pendekatan berbasis teknologi yang dapat membantu proses analisis awal secara lebih terstruktur dan objektif, tanpa menggantikan peran asesor sebagai pembuat keputusan akhir.

Salah satu pendekatan yang dapat digunakan untuk membantu analisis dokumen teks adalah text mining[6]. Text mining memungkinkan dokumen teks diolah menjadi data terstruktur melalui tahapan seperti case folding, tokenizing, stopword removal, stemming, pembobotan kata, dan perhitungan kemiripan dokumen[7]. Penelitian sebelumnya menunjukkan bahwa kombinasi Term Frequency-Inverse Document Frequency dan Cosine Similarity banyak digunakan dalam pengolahan dokumen teks karena dapat mengukur tingkat kesamaan antar dokumen berdasarkan bobot kata yang muncul di dalamnya[8]. Pada penelitian terkait temu kembali informasi, tahapan praproses teks, pembobotan Term Frequency-Inverse Document Frequency, dan perhitungan Cosine Similarity digunakan untuk memperoleh dokumen yang relevan berdasarkan nilai kemiripan tertinggi[9].

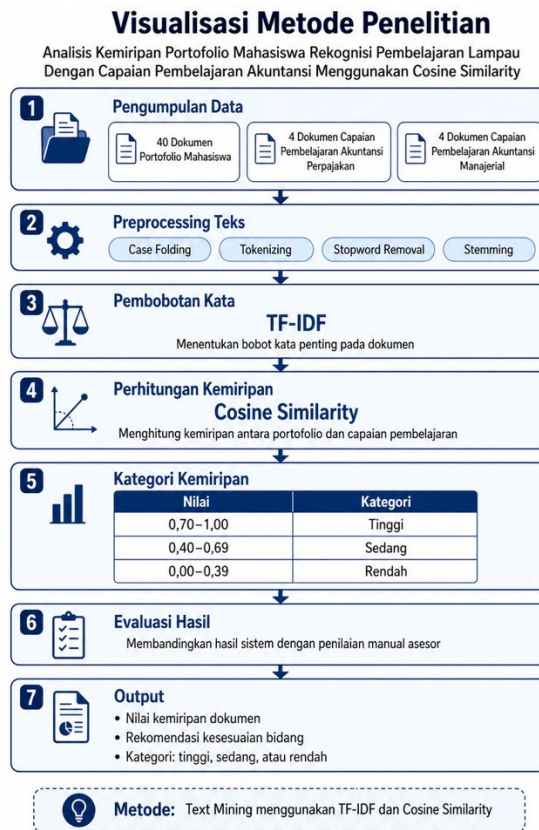
Metode Cosine Similarity dipilih dalam penelitian ini karena mampu menghitung tingkat kemiripan antara dua dokumen berdasarkan representasi vektor[10]. Sebelum dilakukan perhitungan kemiripan, dokumen portofolio dan dokumen capaian pembelajaran diberi bobot menggunakan Term Frequency-Inverse Document Frequency agar kata-kata penting memiliki pengaruh lebih besar dalam proses perbandingan[11]. Dalam konteks penelitian ini, dokumen portofolio calon mahasiswa dibandingkan dengan capaian pembelajaran Program Studi Akuntansi Perpajakan dan Akuntansi Manajerial[12]. Nilai kemiripan yang dihasilkan kemudian dikelompokkan ke dalam kategori tinggi, sedang, dan rendah untuk memberikan gambaran awal mengenai kesesuaian portofolio dengan capaian pembelajaran program studi[13].

Kontribusi dari penelitian ini adalah penerapan metode Cosine Similarity untuk menganalisis kemiripan dokumen portofolio mahasiswa Rekognisi Pembelajaran Lampau dengan capaian pembelajaran pada bidang akuntansi, khususnya Akuntansi Perpajakan dan Akuntansi Manajerial. Berbeda dari penelitian kemiripan dokumen yang umumnya digunakan untuk deteksi plagiasi, pencarian dokumen, atau temu kembali informasi, penelitian ini menerapkan pendekatan tersebut pada konteks seleksi awal mahasiswa Rekognisi Pembelajaran Lampau. Hasil penelitian diharapkan dapat menjadi dasar pengembangan alat bantu analisis portofolio yang mampu memberikan informasi awal kepada asesor secara lebih cepat, objektif, dan efisien sebelum dilakukan proses verifikasi dan penilaian akademik secara menyeluruh.

2. Metode Penelitian

Penelitian ini menggunakan pendekatan text mining untuk menganalisis kemiripan antara dokumen portofolio calon mahasiswa Rekognisi Pembelajaran Lampau dan dokumen capaian pembelajaran program studi. Metode yang digunakan adalah Term Frequency-Inverse Document Frequency (TF-IDF) untuk pembobotan kata dan Cosine Similarity untuk menghitung nilai kemiripan antar dokumen[14]–[16]. Objek penelitian difokuskan pada calon mahasiswa Rekognisi Pembelajaran Lampau yang mendaftar pada bidang Akuntansi Perpajakan dan Akuntansi Manajerial.

Tahapan penelitian ini dilakukan secara sistematis mulai dari pengumpulan data dokumen portofolio mahasiswa Rekognisi Pembelajaran Lampau, preprocessing teks, pembobotan kata menggunakan TF-IDF, perhitungan kemiripan menggunakan Cosine Similarity, pengelompokan kategori kemiripan, evaluasi hasil, hingga penentuan output penelitian. Alur tahapan penelitian ditunjukkan pada Gambar 1.



Gambar 1. Metode Penelitian

Berdasarkan Gambar 1, penelitian diawali dengan pengumpulan data berupa 40 dokumen portofolio mahasiswa, 4 dokumen capaian pembelajaran Akuntansi Perpajakan, dan 4 dokumen capaian pembelajaran Akuntansi Manajerial. Selanjutnya, seluruh dokumen melalui tahap preprocessing teks yang meliputi case folding, tokenizing, stopword removal, dan stemming[17]. Dokumen yang telah diproses kemudian diberi bobot menggunakan TF-IDF untuk menentukan bobot kata penting. Setelah itu, nilai kemiripan antara dokumen portofolio dan dokumen capaian pembelajaran dihitung menggunakan Cosine Similarity[18]. Hasil perhitungan dikelompokkan ke dalam kategori tinggi, sedang, dan rendah, kemudian dibandingkan dengan penilaian manual asesor sebagai bentuk evaluasi hasil[19].

Data yang digunakan dalam penelitian ini berupa dokumen teks yang terdiri dari dokumen portofolio calon mahasiswa dan dokumen capaian pembelajaran program studi. Dokumen portofolio berisi deskripsi pengalaman kerja, kompetensi, pelatihan, sertifikat, dan aktivitas profesional calon mahasiswa. Sementara itu, dokumen capaian pembelajaran berisi kompetensi yang harus dicapai mahasiswa pada bidang Akuntansi Perpajakan dan Akuntansi Manajerial.

Tabel 2. Komposisi Data Penelitian

Jenis Data	Jumlah Dokumen	Keterangan
Portofolio calon mahasiswa Rekognisi Pembelajaran Lampau	40 dokumen	Berisi pengalaman kerja, keahlian, pelatihan, dan sertifikat
Capaian pembelajaran Akuntansi Perpajakan	4 dokumen	Berisi kompetensi bidang perpajakan
Capaian pembelajaran Akuntansi Manajerial	4 dokumen	Berisi kompetensi bidang akuntansi manajemen
Total	48 dokumen	Digunakan untuk proses analisis kemiripan

Tahap preprocessing dilakukan untuk membersihkan dan menyiapkan data teks agar dapat diproses secara komputasional. Preprocessing terdiri dari case folding, tokenizing, stopword removal, dan stemming. Case folding dilakukan untuk mengubah seluruh huruf menjadi huruf kecil. Tokenizing digunakan untuk memecah kalimat menjadi kata-kata. Stopword removal digunakan untuk menghapus kata umum yang tidak memiliki makna penting, seperti “yang”, “dan”, “di”, “ke”, dan “dari”[20]. Stemming digunakan untuk mengubah kata berimbuhan menjadi kata dasar.

Proses stemming pada penelitian ini dilakukan terhadap dokumen berbahasa Indonesia menggunakan pustaka Sastrawi. Stemming digunakan untuk mengubah kata berimbuhan menjadi kata dasar, seperti “menghitung” menjadi “hitung”, “menyusun” menjadi “susun”, dan “pelaporan” menjadi “lapor”. Pemilihan Sastrawi dilakukan karena pustaka tersebut mendukung proses stemming untuk bahasa Indonesia, sehingga sesuai dengan karakteristik dokumen portofolio dan capaian pembelajaran yang digunakan dalam penelitian ini.

Tabel 1. Tahapan Preprocessing Penelitian

Tahapan	Teknik/Library	Bahasa	Keterangan
Case folding	Python string processing	Indonesia	Mengubah seluruh huruf menjadi huruf kecil
Tokenizing	Python text processing	Indonesia	Memecah kalimat menjadi token kata
Stopword removal	Sastrawi StopWordRemoverFactory	Indonesia	Menghapus kata umum bahasa Indonesia
Stemming	Sastrawi StemmerFactory	Indonesia	Mengubah kata berimbuhan menjadi kata dasar
Pembobotan kata	TF-IDF	Indonesia	Memberikan bobot pada kata dalam dokumen

Setelah tahap preprocessing, dokumen yang telah bersih diberi bobot menggunakan metode TF-IDF[11]. TF-IDF digunakan untuk menentukan tingkat kepentingan suatu kata dalam dokumen berdasarkan frekuensi kemunculan kata pada satu dokumen dan penyebarannya pada seluruh dokumen. Kata yang sering muncul dalam satu dokumen tetapi jarang muncul pada dokumen lain akan memiliki bobot yang lebih tinggi. Rumus TF-IDF yang digunakan adalah sebagai berikut:

$$TFIDF(t, d) = TF(t, d) \times IDF(t) \tag{1}$$

Keterangan:

- TF(t,d) adalah frekuensi kemunculan kata t dalam dokumen d.
- IDF(t) adalah nilai inverse document frequency dari kata t.
- TFIDF(t,d) adalah bobot kata t pada dokumen d.

Setelah setiap dokumen direpresentasikan dalam bentuk vektor TF-IDF, tahap berikutnya adalah menghitung nilai kemiripan antara dokumen portofolio dan dokumen capaian pembelajaran menggunakan Cosine Similarity. Metode ini menghitung kemiripan berdasarkan sudut antara dua vektor dokumen. Semakin besar nilai Cosine Similarity, semakin tinggi tingkat kemiripan antar dokumen. Nilai Cosine Similarity berada pada rentang 0 sampai 1. Rumus Cosine Similarity yang digunakan adalah:

$$\text{Cosine Similarity}(A, B) = \frac{A \cdot B}{\|A\| \times \|B\|} \tag{2}$$

- A adalah vektor dokumen portofolio calon mahasiswa.
- B adalah vektor dokumen capaian pembelajaran.
- A·B adalah hasil perkalian dot product antara vektor A dan B.
- ||A|| dan ||B|| adalah panjang masing-masing vektor dokumen.

Setiap dokumen portofolio calon mahasiswa dibandingkan dengan dokumen capaian pembelajaran Akuntansi Perpajakan dan Akuntansi Manajerial. Apabila nilai kemiripan terhadap dokumen Akuntansi Perpajakan lebih tinggi, maka portofolio calon mahasiswa cenderung lebih sesuai dengan bidang Akuntansi Perpajakan. Sebaliknya, apabila nilai kemiripan terhadap dokumen Akuntansi Manajerial lebih tinggi, maka portofolio calon mahasiswa cenderung lebih sesuai dengan bidang Akuntansi Manajerial.

Hasil perhitungan Cosine Similarity kemudian dikelompokkan ke dalam tiga kategori, yaitu tinggi, sedang, dan rendah. Kategori ini digunakan untuk memudahkan interpretasi hasil kemiripan dokumen.

Tabel 2. Kategori Nilai Cosine Similarity

Nilai Cosine Similarity	Kategori Kemiripan	Interpretasi
0,70 – 1,00	Tinggi	Portofolio sangat sesuai dengan capaian pembelajaran
0,40 – 0,69	Sedang	Portofolio cukup sesuai dengan capaian pembelajaran
0,00 – 0,39	Rendah	Portofolio kurang sesuai dengan capaian pembelajaran

Hasil akhir dari penelitian ini berupa nilai kemiripan setiap dokumen portofolio terhadap capaian pembelajaran Akuntansi Perpajakan dan Akuntansi Manajerial. Nilai tersebut digunakan untuk menentukan kategori kemiripan dan kecenderungan kesesuaian bidang calon mahasiswa.

Evaluasi hasil dilakukan dengan cara membandingkan hasil kategori kemiripan dari sistem dengan penilaian manual asesor. Penilaian manual asesor digunakan sebagai pembanding untuk mengetahui apakah hasil analisis menggunakan Cosine Similarity sudah sesuai dengan pertimbangan manusia. Tingkat kesesuaian dihitung berdasarkan jumlah dokumen yang memiliki hasil kategori sama antara sistem dan asesor. Rumus tingkat kesesuaian adalah:

$$\text{Tingkat Kesesuaian} = \frac{\text{Jumlah Data Sesuai}}{\text{Jumlah Seluruh DataB}} \times 100\% \tag{2}$$

Dengan metode ini, hasil penelitian tidak dimaksudkan untuk menggantikan keputusan asesor, tetapi sebagai alat bantu analisis awal. Asesor tetap berperan dalam melakukan verifikasi akhir terhadap dokumen portofolio, sertifikat, pengalaman kerja, dan bukti kompetensi calon mahasiswa Rekognisi Pembelajaran Lampau. Pendekatan ini diharapkan dapat membantu proses analisis portofolio menjadi lebih cepat, terstruktur, dan objektif

3. Hasil dan Pembahasan

Bagian ini menyajikan hasil analisis kemiripan dokumen portofolio mahasiswa Rekognisi Pembelajaran Lampau dengan capaian pembelajaran pada bidang Akuntansi Perpajakan dan Akuntansi Manajerial. Analisis dilakukan melalui tahapan preprocessing teks, pembobotan kata menggunakan TF-IDF, perhitungan kemiripan menggunakan Cosine Similarity, serta pengelompokan hasil ke dalam kategori kemiripan tinggi, sedang, dan rendah. Data yang digunakan terdiri dari 40 dokumen portofolio calon mahasiswa Rekognisi Pembelajaran Lampau, 4 dokumen capaian pembelajaran Akuntansi Perpajakan, dan 4 dokumen capaian pembelajaran Akuntansi Manajerial.

3.1. Hasil Preprocessing Teks

Tahap preprocessing dilakukan untuk membersihkan data teks sebelum proses pembobotan kata dan perhitungan kemiripan. Proses ini meliputi case folding, tokenizing, stopword removal, dan stemming. Tujuan preprocessing adalah mengubah dokumen teks menjadi bentuk yang lebih terstruktur sehingga dapat diproses menggunakan metode TF-IDF dan Cosine Similarity.

Tabel 3. Tahapan Preprocessing Teks

Tahapan	Hasil
Teks awal	Saya memiliki pengalaman menyusun laporan pajak, menghitung PPh, dan membuat faktur pajak perusahaan.
Case folding	saya memiliki pengalaman menyusun laporan pajak, menghitung pph, dan membuat faktur pajak perusahaan
Tokenizing	saya, memiliki, pengalaman, menyusun, laporan, pajak, menghitung, pph, membuat, faktur, pajak, perusahaan
Stopword removal	pengalaman, menyusun, laporan, pajak, menghitung, pph, membuat, faktur, pajak, perusahaan
Stemming	alam, susun, lapor, pajak, hitung, pph, buat, faktur, pajak, usaha

Berdasarkan Tabel 3, proses preprocessing mampu mengurangi kata-kata yang tidak relevan dan mempertahankan kata-kata penting yang berkaitan dengan kompetensi calon mahasiswa. Kata seperti **pajak**, **lapor**, **hitung**, **PPh**, dan **faktur** menjadi kata penting yang dapat merepresentasikan pengalaman kerja pada bidang Akuntansi Perpajakan. Hasil preprocessing ini kemudian digunakan sebagai dasar dalam proses pembobotan kata menggunakan TF-IDF.

3.2. Hasil Perhitungan Cosine Similarity

Setelah preprocessing dilakukan, setiap dokumen portofolio dan dokumen capaian pembelajaran diubah menjadi vektor menggunakan TF-IDF. Selanjutnya, nilai kemiripan dihitung menggunakan Cosine Similarity. Setiap dokumen portofolio dibandingkan dengan dua kelompok capaian pembelajaran, yaitu Akuntansi Perpajakan dan Akuntansi Manajerial. Nilai tertinggi digunakan sebagai dasar untuk menentukan kecenderungan kesesuaian bidang.

Tabel 4. Hasil Perhitungan Cosine Similarity

Kode Dokumen	Akuntansi Perpajakan	Akuntansi Manajerial	Nilai Tertinggi	Rekomendasi Kesesuaian	Kategori
P01	0,82	0,54	0,82	Akuntansi Perpajakan	Tinggi
P02	0,47	0,79	0,79	Akuntansi Manajerial	Tinggi
P03	0,68	0,61	0,68	Akuntansi Perpajakan	Sedang
...
P40	0,41	0,73	0,73	Akuntansi Manajerial	Tinggi

Tabel 4 menampilkan sebagian hasil perhitungan Cosine Similarity sebagai representasi dari 40 dokumen portofolio yang dianalisis. Rekapitulasi keseluruhan hasil disajikan pada Tabel 7 dalam bentuk distribusi kategori kemiripan.

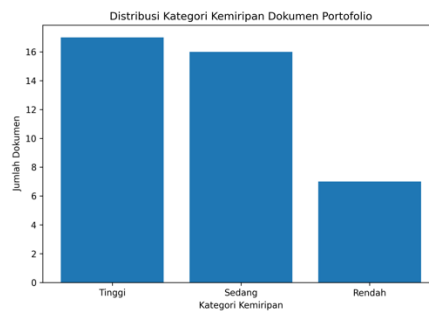
3.2.1. Distribusi Kategori Kemiripan

Hasil perhitungan terhadap 40 dokumen portofolio menunjukkan bahwa dokumen dapat dikelompokkan ke dalam tiga kategori, yaitu tinggi, sedang, dan rendah. Kategori tinggi menunjukkan bahwa dokumen portofolio memiliki tingkat kemiripan yang kuat dengan capaian pembelajaran. Kategori sedang menunjukkan adanya sebagian kesesuaian, sedangkan kategori rendah menunjukkan bahwa portofolio belum cukup merepresentasikan capaian pembelajaran yang dituju.

Tabel 5. Distribusi Kategori Kemiripan Dokumen Portofolio

Kategori Kemiripan	Jumlah Dokumen	Persentase
Tinggi	17	42,5%
Sedang	16	40,0%
Rendah	7	17,5%
Total	40	100%

Berdasarkan Tabel 5, sebanyak 17 dokumen atau 42,5% berada pada kategori kemiripan tinggi. Sebanyak 16 dokumen atau 40,0% berada pada kategori sedang, sedangkan 7 dokumen atau 17,5% berada pada kategori rendah. Jika kategori tinggi dan sedang digabungkan, maka terdapat 33 dokumen atau 82,5% yang memiliki tingkat kemiripan cukup baik terhadap capaian pembelajaran program studi.



Gambar 2. Distribusi Kategori Kemiripan Dokumen Portofolio

Gambar 2 memperlihatkan bahwa kategori kemiripan tinggi dan sedang mendominasi hasil analisis. Kondisi ini menunjukkan bahwa sebagian besar dokumen portofolio calon mahasiswa Rekognisi Pembelajaran Lampau memiliki keterkaitan dengan capaian pembelajaran pada bidang akuntansi. Namun, dokumen dengan kategori rendah tetap perlu menjadi perhatian karena dapat menunjukkan kurangnya relevansi pengalaman kerja, kurang lengkapnya deskripsi portofolio, atau penggunaan istilah yang tidak sesuai dengan capaian pembelajaran program studi.

3.2.2. Rekomendasi Kesesuaian Bidang

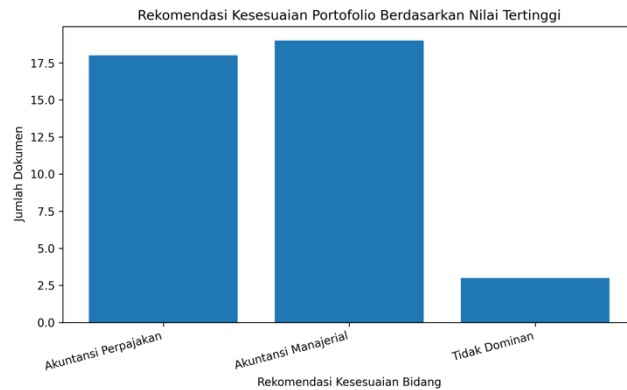
Selain mengelompokkan dokumen berdasarkan kategori kemiripan, penelitian ini juga menganalisis kecenderungan kesesuaian dokumen portofolio terhadap dua bidang, yaitu Akuntansi Perpajakan dan Akuntansi Manajerial. Rekomendasi kesesuaian ditentukan berdasarkan nilai Cosine Similarity tertinggi dari masing-masing dokumen.

Tabel 6. Distribusi Rekomendasi Kesesuaian Bidang

Rekomendasi Kesesuaian	Jumlah Dokumen	Persentase
Akuntansi Perpajakan	18	45,0%
Akuntansi Manajerial	19	47,5%
Tidak Dominan	3	7,5%
Total	40	100%

Berdasarkan Tabel 6, sebanyak 18 dokumen atau 45,0% lebih sesuai dengan bidang Akuntansi Perpajakan, sedangkan 19 dokumen atau 47,5% lebih sesuai dengan bidang Akuntansi Manajerial. Sementara itu, terdapat 3 dokumen atau 7,5% yang tidak memiliki kecenderungan kuat terhadap salah satu

bidang. Dokumen yang tidak dominan umumnya memiliki nilai kemiripan rendah pada kedua bidang, sehingga memerlukan pemeriksaan lebih lanjut oleh asesor.



Gambar 3. Rekomendasi Kesesuaian Portofolio Berdasarkan Nilai Tertinggi

Gambar 3 menunjukkan bahwa distribusi rekomendasi antara Akuntansi Perpajakan dan Akuntansi Manajerial relatif seimbang. Dokumen yang lebih sesuai dengan Akuntansi Perpajakan umumnya memuat kata-kata seperti pajak, SPT, PPh, PPN, faktur pajak, e-Filing, dan administrasi perpajakan. Sebaliknya, dokumen yang lebih sesuai dengan Akuntansi Manajerial cenderung memuat kata-kata seperti laporan keuangan, anggaran, biaya, jurnal, neraca, audit, dan analisis keuangan.

3.2.3. Evaluasi Hasil dengan Penilaian Asesor

Evaluasi dilakukan dengan membandingkan hasil kategori kemiripan yang diperoleh dari sistem dengan penilaian manual asesor. Penilaian asesor digunakan sebagai acuan pembandingan untuk mengetahui sejauh mana hasil analisis menggunakan TF-IDF dan Cosine Similarity sesuai dengan pertimbangan manusia. Evaluasi ini penting karena hasil sistem hanya berfungsi sebagai alat bantu analisis awal, sedangkan keputusan akhir tetap berada pada asesor.

Tabel 7. Hasil Perbandingan Sistem dengan Penilaian Asesor

Keterangan	Jumlah
Jumlah dokumen portofolio	40
Hasil sesuai dengan penilaian asesor	35
Hasil tidak sesuai dengan penilaian asesor	5
Tingkat kesesuaian	87,5%

Berdasarkan Tabel 7, sebanyak 35 dari 40 dokumen portofolio memiliki kategori yang sama antara hasil sistem dan penilaian manual asesor. Dengan demikian, tingkat kesesuaian hasil sistem terhadap penilaian asesor adalah sebesar 87,5%. Hasil ini menunjukkan bahwa metode TF-IDF dan Cosine Similarity dapat digunakan sebagai alat bantu dalam proses analisis awal dokumen portofolio mahasiswa Rekognisi Pembelajaran Lampau.

Meskipun tingkat kesesuaian yang diperoleh cukup tinggi, terdapat 5 dokumen yang memiliki hasil berbeda antara sistem dan asesor. Perbedaan tersebut dapat terjadi karena sistem menghitung kemiripan berdasarkan kesamaan kata, sedangkan asesor dapat mempertimbangkan konteks, makna pengalaman kerja, bukti pendukung, serta relevansi kompetensi secara lebih mendalam. Oleh karena itu, hasil sistem tidak digunakan sebagai keputusan akhir, tetapi sebagai rekomendasi awal yang perlu diverifikasi kembali oleh asesor.

3.3. Pembahasan

Hasil penelitian menunjukkan bahwa metode Cosine Similarity dapat digunakan untuk membantu menganalisis kemiripan dokumen portofolio mahasiswa Rekognisi Pembelajaran Lampau dengan capaian pembelajaran program studi. Melalui pembobotan TF-IDF, kata-kata penting dalam dokumen dapat diberi bobot sesuai tingkat kemunculan dan keunikannya. Kata-kata yang sering muncul pada dokumen tertentu tetapi tidak terlalu umum pada dokumen lain akan memiliki bobot lebih tinggi, sehingga lebih berpengaruh dalam proses perhitungan kemiripan.

Pada bidang Akuntansi Perpajakan, dokumen portofolio dengan nilai kemiripan tinggi umumnya berisi pengalaman yang berkaitan dengan pelaporan pajak, penghitungan PPh, PPN, faktur pajak, SPT, e-

Filing, dan administrasi perpajakan. Hal ini menunjukkan bahwa dokumen tersebut memiliki kesesuaian dengan capaian pembelajaran yang menekankan kemampuan dalam bidang perpajakan. Sementara itu, pada bidang Akuntansi Manajerial, dokumen dengan nilai kemiripan tinggi umumnya berisi pengalaman terkait penyusunan laporan keuangan, pengelolaan anggaran, pencatatan transaksi, analisis biaya, audit internal, dan evaluasi kinerja keuangan.

Dokumen dengan kategori sedang menunjukkan bahwa sebagian pengalaman calon mahasiswa memiliki keterkaitan dengan capaian pembelajaran, tetapi belum sepenuhnya kuat. Kondisi ini dapat terjadi karena deskripsi pengalaman kerja belum ditulis secara rinci atau hanya memuat sebagian kompetensi yang relevan. Dokumen dengan kategori rendah menunjukkan bahwa isi portofolio memiliki sedikit kesamaan kata dengan capaian pembelajaran. Hal ini dapat disebabkan oleh pengalaman kerja yang kurang relevan, dokumen portofolio yang terlalu singkat, atau penggunaan istilah yang berbeda dari dokumen capaian pembelajaran.

Secara keseluruhan, penelitian ini menunjukkan bahwa pendekatan text mining menggunakan TF-IDF dan Cosine Similarity dapat membantu proses analisis awal dokumen portofolio mahasiswa Rekognisi Pembelajaran Lampau. Pendekatan ini dapat mempercepat proses pencocokan dokumen, mengurangi subjektivitas pada tahap awal, dan memberikan gambaran awal mengenai kecenderungan kesesuaian portofolio terhadap bidang Akuntansi Perpajakan atau Akuntansi Manajerial.

4. Kesimpulan

Berdasarkan hasil penelitian, metode TF-IDF dan Cosine Similarity dapat digunakan untuk menganalisis kemiripan dokumen portofolio mahasiswa Rekognisi Pembelajaran Lampau dengan capaian pembelajaran pada bidang Akuntansi Perpajakan dan Akuntansi Manajerial. Proses preprocessing yang meliputi case folding, tokenizing, stopword removal, dan stemming membantu mengubah dokumen teks menjadi lebih terstruktur sebelum dilakukan pembobotan kata dan perhitungan kemiripan.

Hasil perhitungan terhadap 40 dokumen portofolio menunjukkan bahwa sebanyak 17 dokumen berada pada kategori kemiripan tinggi, 16 dokumen berada pada kategori sedang, dan 7 dokumen berada pada kategori rendah. Selain itu, hasil rekomendasi kesesuaian bidang menunjukkan bahwa 18 dokumen lebih sesuai dengan Akuntansi Perpajakan, 19 dokumen lebih sesuai dengan Akuntansi Manajerial, dan 3 dokumen tidak memiliki kecenderungan dominan terhadap salah satu bidang..

Meskipun demikian, hasil analisis menggunakan Cosine Similarity tetap perlu diverifikasi oleh asesor karena metode ini hanya mengukur kemiripan berdasarkan kesamaan kata, bukan penilaian kompetensi secara menyeluruh. Penelitian selanjutnya dapat mengembangkan pendekatan ini menggunakan metode semantic similarity atau model berbasis transformer agar sistem mampu memahami kesamaan makna antar dokumen dengan lebih baik.

Daftar Pustaka

- [1] K. R. Penerapan Permendikbudristek Nomor, R. Nasir, C. Author, S. Muhammadiyah Kalabahi, N. Tenggara Timur, Dan I. Abstrak, "Rekognisi Pembelajaran Lampau Di Perguruan Tinggi: Kajian Penerapan Permendikbudristek Ri Nomor 41 Tahun 2021," *Bul. Edukasi Indones.*, Vol. 1, No. 01, 2022.
- [2] Rahmad Nasir, "Rekognisi Pembelajaran Lampau Di Perguruan Tinggi," *Bul. Edukasi Indones.*, Vol. 1, No. 01, 2022, Doi: 10.56741/Bei.V1i01.20.
- [3] D. Jatmoko, W. Widiyatmoko, Y. Widiyono, Dan M. F. Rais, "Analisis Pendukung Dan Penghambat Kelulusan Tepat Waktu Mahasiswa Rekognisi Pembelajaran Lampau," *J. Penelit. Inov.*, Vol. 4, No. 3, 2024, Doi: 10.54082/Jupin.513.
- [4] Lukman Rais *Et Al.*, "Sosialisasi Peningkatan Kapabilitas Aparatur Pemerintah Melaluipendidikan Rekognisi Pembelajaran Lampau Dikabupaten Sorong," *Abdimas Papua J. Community Serv.*, Vol. 6, No. 1, 2024.
- [5] M. Ali *Et Al.*, "Sosialisasi Sistem Pendidikan Rekognisi Pembelajaran Lampau (Rpl) Di Pemerintah Kota Sorong," *Abdimas Papua J. Community Serv.*, Vol. 6, No. 1, 2024, Doi: 10.33506/Pjcs.V6i1.3128.
- [6] H. Hassani, C. Beneki, S. Unger, M. T. Mazinani, Dan M. R. Yeganegi, "Text Mining In Big Data Analytics," *Big Data Cogn. Comput.*, Vol. 4, No. 1, 2020, Doi: 10.3390/Bdcc4010001.
- [7] T. Yuniarti, J. Astuti, F. Faujiyah, Dan M. Zaiyar, "Pendekatan Text Mining Dalam Menilai Sentimen Publik Pada Baterai Kendaraan Listrik," *J. Serambi Eng.*, Vol. 9, No. 4, 2024.
- [8] A. Karami, M. Lundy, F. Webb, Dan Y. K. Dwivedi, "Twitter And Research: A Systematic Literature Review Through Text Mining," *Ieee Access*, Vol. 8, 2020, Doi: 10.1109/Access.2020.2983656.

- [9] B. S. Tanjung, "Pendekatan Text Mining Sebagai Sistem Pendeteksi Pemberitaan Palsu Yang Tersebar Dalam Twitter," *Semin. Nas. Ilmu Terap. 2018 – Univ. Widya Kartika*, 2018.
- [10] S. Lumbansiantar, S. Dwiasnati, Dan N. S. Fatonah, "Penerapan Metode Cosine Similarity Dalam Mendeteksi Plagiarisme Pada Jurnal," *Format J. Ilm. Tek. Inform.*, Vol. 12, No. 2, 2023, Doi: 10.22441/Format.2023.V12.I2.007.
- [11] N. Setiawan Dan F. Amin, "Sistem Temu Kembali Informasi Jurnal Ilmiah Unisbank Dengan Metode Cosine Similarity," *J. Teknol. Dan Sist. Inf. Bisnis*, Vol. 6, No. 3, 2024, Doi: 10.47233/Jteksis.V6i3.1391.
- [12] R. P. Pratama, M. Faisal, Dan A. Hanani, "Deteksi Plagiarisme Pada Dokumen Jurnal Menggunakan Metode Cosine Similarity," *Smartics J.*, Vol. 5, No. 1, 2019, Doi: 10.21067/Smartics.V5i1.2848.
- [13] L. Meilina, I. N. S. Kumara, Dan I. N. Setiawan, "Literature Review Klasifikasi Data Menggunakan Metode Cosine Similarity Dan Artificial Neural Network," *Maj. Ilm. Teknol. Elektro*, Vol. 20, No. 2, 2021, Doi: 10.24843/Mite.2021.V20i02.P15.
- [14] A. Hanani, "Algoritma Term Frequency – Inverse Document Frequency (Tf-Idf)," *Researchgate*, No. December, 2023.
- [15] L. Havrlant Dan V. Kreinovich, "A Simple Probabilistic Explanation Of Term Frequency-Inverse Document Frequency (Tf-Idf) Heuristic (And Variations Motivated By This Explanation)," *Int. J. Gen. Syst.*, Vol. 46, No. 1, 2017, Doi: 10.1080/03081079.2017.1291635.
- [16] D. Septiani Dan I. Isabela, "Sintesia: Jurnal Sistem Dan Teknologi Informasi Indonesia Analisis Term Frequency Inverse Document Frequency (Tf-Idf) Dalam Temu Kembali Informasi Pada Dokumen Teks," *Sintesia J. Sist. Dan Teknol. Inf. Indones.*, Vol. 1, No. 2, 2022.
- [17] I. P. G. A. Sudiatmika, P. S. Saputra, R. L. Rahardian, Dan K. H. S. Dewi, "Sentiment Analysis Of Tourist Reviews On Google Maps For Pura Besakih Using Machine Learning Algorithms," *J. Mandiri It*, Vol. 14, No. 1, 2025.
- [18] B. Herwijayanti, D. E. Ratnawati, Dan L. Muflikhah, "Klasifikasi Berita Online Dengan Menggunakan Pembobotan Tf-Idf Dan Cosine Similarity," *J. Pengemb. Teknol. Inf. Dan Ilmu Komput.*, Vol. 2, No. 1, 2018.
- [19] P. Xia, L. Zhang, Dan F. Li, "Learning Similarity With Cosine Similarity Ensemble," *Inf. Sci. (Ny)*, Vol. 307, 2015, Doi: 10.1016/J.Ins.2015.02.024.
- [20] A. Riyani, M. Zidny Naf'an #2, Dan A. Burhanuddin, "Penerapan Cosine Similarity Dan Pembobotan Tf-Idf Untuk Mendeteksi Kemiripan Dokumen," *Jlk*, Vol. 2, No. 1, 2019.