



SISTEM PREDIKSI KECOCOKAN KARIR PASCA STUDI BERBASIS RESUME SCREENING MENGGUNAKAN METODE RANDOM FOREST DAN SVM

Yutika Amelia Effendi*, Afnan Nadhir, Hensa Hendy Anugerah Ebenezer Situngkir, Rafka
Kumara Aradea dan Stephanus Chandra Setiono

Teknik Robotika dan Kecerdasan Buatan, Fakultas Teknologi Maju dan Multidisiplin,
Universitas Airlangga, Surabaya, 60115, Indonesia

yutika.effendi@ftmm.unair.ac.id, afnan.nadhir-2023@ftmm.unair.ac.id, hensa.hendy.anugrah-
2023@ftmm.unair.ac.id, rafka.kumara.aradea-2023@ftmm.unair.ac.id, stephanus.chan.setiono-
2023@ftmm.unair.ac.id

*Penulis Koresponden

ABSTRAK

Fenomena *education-job mismatch* masih menjadi permasalahan serius bagi lulusan perguruan tinggi, di mana banyak individu bekerja pada bidang yang tidak selaras dengan latar belakang pendidikan dan kompetensinya. Proses penentuan karier juga umumnya masih bersifat subjektif dan belum memanfaatkan analisis data secara optimal. Penelitian ini bertujuan untuk merancang dan mengimplementasikan sistem prediksi kecocokan karier pasca-studi berbasis *resume screening* menggunakan pendekatan *machine learning*. Metode yang digunakan melibatkan dua algoritma klasifikasi terawasi, yaitu Random Forest dan Support Vector Machine (SVM), dengan representasi fitur menggunakan TF-IDF berbasis n-gram pada dataset sebanyak 13.389 resume. Hasil penelitian menunjukkan bahwa kedua model mampu memberikan performa yang baik, namun SVM memberikan hasil yang lebih unggul dengan akurasi sebesar 84,39% dan F1-score sebesar 83,36%, dibandingkan Random Forest dengan akurasi 81,96%. Analisis *feature importance* menunjukkan bahwa kompetensi teknis, pengalaman kerja, dan bidang studi merupakan faktor utama yang mempengaruhi kecocokan karier. Penelitian ini memberikan kontribusi berupa pendekatan prediktif berbasis data untuk mendukung pengambilan keputusan karier yang lebih objektif bagi mahasiswa dan lulusan.

Kata kunci: pembelajaran mesin, random forest, resume screening, support vector machine.

ABSTRACT

The *education-job mismatch* phenomenon remains a significant challenge for university graduates, where many individuals work in fields that are not aligned with their educational background and competencies. Career decision-making processes are also generally subjective and have not fully leveraged data-driven analysis. This study aims to design and implement a post-graduation career-fit prediction system based on resume screening using a machine learning approach. The proposed method employs two supervised classification algorithms, namely Random Forest and Support Vector Machine (SVM), with feature representation using TF-IDF based on n-grams on a dataset of 13,389 resumes. The results indicate that both models achieve strong performance; however, SVM outperforms Random Forest, achieving an accuracy of 84.39% and an F1-score of 83.36%, compared to 81.96% accuracy for Random Forest. Feature importance analysis reveals that technical skills, work experience, and field of study are the most influential factors in determining career fit. This study contributes a data-driven predictive approach to support more objective career decision-making for students and graduates.

Keywords: machine learning, random forest, resume screening, support vector machine.

Histori Artikel:

Diserahkan: 21 Januari 2026

Diterima setelah Revisi: 22 Mei 2026

Diterbitkan: 31 Mei 2026

1. PENDAHULUAN

Periode transisi dari dunia pendidikan ke dunia kerja sering kali dihadapkan pada permasalahan ketidaksesuaian antara latar belakang pendidikan dengan pekerjaan yang dijalani, yang dikenal sebagai fenomena *education-job mismatch*. Permasalahan ini masih menjadi isu yang signifikan di Indonesia. Data terbaru pada November 2025 menunjukkan bahwa sekitar 35,36% pekerja muda bekerja pada bidang yang tidak sesuai dengan tingkat pendidikan yang dimilikinya [1], sementara sekitar 1,01 juta lulusan sarjana berada dalam kondisi menganggur [2]. Kondisi ini menunjukkan bahwa pendidikan tinggi belum sepenuhnya mampu menjamin kesesuaian antara kompetensi lulusan dengan kebutuhan pasar kerja. Dampaknya tidak hanya berupa inefisiensi investasi pendidikan, tetapi juga memengaruhi produktivitas, tingkat upah, stabilitas kerja, serta kepuasan kerja.

Fenomena tersebut menegaskan perlunya pendekatan yang lebih objektif dalam membantu individu menentukan jalur karier yang sesuai. Selama ini, proses pengambilan keputusan karier masih cenderung subjektif dan belum memanfaatkan analisis data secara optimal. Oleh karena itu, diperlukan suatu sistem yang mampu memprediksi kecocokan karier berdasarkan profil individu, seperti latar belakang pendidikan, keterampilan, dan pengalaman kerja, sehingga keputusan karier dapat dilakukan secara lebih terukur dan berbasis data.

Perkembangan teknik *machine learning* membuka peluang untuk membangun model prediktif kecocokan karier. Algoritma seperti Random Forest dan Support Vector Machine (SVM) telah banyak digunakan dalam berbagai tugas klasifikasi karena kemampuannya dalam menangani data berdimensi tinggi dan hubungan nonlinier. Random Forest memiliki keunggulan dalam stabilitas model serta kemampuan mengidentifikasi *feature importance*, sedangkan SVM unggul dalam membentuk batas keputusan optimal pada ruang fitur berdimensi tinggi. Karakteristik ini menjadikan kedua algoritma relevan untuk digunakan dalam analisis data resume yang bersifat tidak terstruktur.

Penelitian terdahulu telah mengkaji fenomena *education-job mismatch* dan pengembangan sistem rekomendasi karier dari berbagai perspektif. Jiang dan Guo [3] menunjukkan bahwa hampir 25% lulusan perguruan tinggi di Tiongkok bekerja pada bidang yang tidak sesuai dengan jurusan mereka, yang berdampak pada penurunan kualitas luaran tenaga kerja. Penelitian lain oleh Assylzhan dkk. [4] mengembangkan sistem cerdas untuk menilai kesiapan karier mahasiswa berdasarkan aspek kepribadian dan pengembangan diri. Selain itu, Siswipraptini dkk. [5] mengembangkan model rekomendasi jalur karier personalisasi bagi mahasiswa teknologi informasi di Indonesia menggunakan pendekatan *machine learning*.

Dalam konteks prediksi karier berbasis data teks, Faruque dkk. [6] mengembangkan sistem prediksi karier berbasis Natural Language Processing (NLP) untuk mahasiswa ilmu komputer dan rekayasa perangkat lunak. Sementara itu, Senger dkk. [7] menyoroti pentingnya pengembangan model prediksi jalur karier yang lebih realistis dengan mempertimbangkan dinamika perjalanan karier individu. Pada sisi lain, penelitian mengenai resume screening berbasis *machine learning* juga telah dilakukan, seperti oleh T. K. dkk. [8], Saatci dkk. [9], dan Sheikh dkk. [10], yang menunjukkan bahwa integrasi NLP dan klasifikasi *machine learning* mampu meningkatkan efisiensi serta objektivitas proses seleksi kandidat.

Meskipun demikian, sebagian besar penelitian sebelumnya masih memiliki beberapa keterbatasan. Pertama, banyak penelitian hanya berfokus pada rekomendasi karier pada domain tertentu, seperti mahasiswa teknologi informasi atau ilmu komputer [5][6], sehingga generalisasi model ke berbagai bidang pekerjaan masih terbatas. Kedua, sebagian penelitian lebih menitikberatkan pada proses *resume ranking* atau *screening* untuk rekrutmen [8][9][10], bukan pada prediksi kecocokan karier pasca-studi secara komprehensif. Ketiga, penelitian terdahulu umumnya belum membandingkan performa beberapa algoritma klasifikasi secara mendalam pada dataset resume multi-kategori berskala besar.

Berdasarkan kondisi tersebut, terdapat celah penelitian berupa belum tersedianya sistem prediksi kecocokan karier berbasis *machine learning* yang memanfaatkan data resume lintas

bidang secara komprehensif dengan perbandingan algoritma klasifikasi untuk menghasilkan prediksi yang objektif dan akurat. Oleh karena itu, penelitian ini bertujuan untuk mengembangkan sistem prediksi kecocokan karier pasca-studi berbasis *resume screening* menggunakan algoritma Random Forest dan SVM.

Secara khusus, penelitian ini memiliki tiga kontribusi utama, yaitu: (1) mengembangkan model prediksi kecocokan karier berbasis data resume lintas bidang; (2) membandingkan kinerja algoritma Random Forest dan SVM menggunakan metrik evaluasi seperti *accuracy*, *precision*, *recall*, dan *F1-score*; serta (3) mengidentifikasi fitur-fitur utama dalam resume yang memengaruhi kecocokan karier melalui analisis *feature importance*. Dengan pendekatan tersebut, penelitian ini diharapkan dapat memberikan kontribusi dalam pengembangan sistem rekomendasi karier berbasis data yang lebih objektif, serta mendukung mahasiswa dan lulusan dalam mengambil keputusan karier yang lebih tepat dan sesuai dengan kompetensinya.

2. STUDI LITERATUR

2.1 Kecocokan Karier dan *Education-Job Mismatch*

Kecocokan karier (*career fit*) merupakan konsep penting dalam psikologi karier yang menekankan kesesuaian antara karakteristik individu, seperti minat, kemampuan, dan nilai, dengan tuntutan pekerjaan. Model seperti Holland's RIASEC mengelompokkan individu dan pekerjaan ke dalam enam tipe utama, yaitu *Realistic*, *Investigative*, *Artistic*, *Social*, *Enterprising*, dan *Conventional*, yang menjadi dasar konsep *person-job alignment*. Penelitian menunjukkan bahwa tingkat kecocokan karier yang tinggi berkorelasi positif dengan kepuasan kerja, retensi, dan produktivitas [11].

Namun demikian, fenomena *education-job mismatch* masih menjadi permasalahan global, termasuk di Indonesia. *Mismatch* dapat diklasifikasikan menjadi *horizontal mismatch* (ketidaksesuaian bidang studi dan pekerjaan) dan *vertical mismatch* (ketidaksesuaian tingkat pendidikan dengan tuntutan pekerjaan). OECD memperkirakan sekitar 35-40% lulusan muda di negara berkembang bekerja di bidang yang tidak relevan dengan pendidikannya [12]. Kondisi ini berdampak pada penurunan kepuasan kerja, peningkatan *turnover intention*, serta penurunan produktivitas hingga 10-15% [13]. Dengan demikian, konsep kecocokan karier tidak hanya mencakup kesesuaian bidang studi, tetapi juga integrasi kompetensi, pengalaman, dan aspirasi individu dengan kebutuhan pasar kerja.

2.2 *Machine Learning* dalam Prediksi Kecocokan Karier

Machine learning merupakan pendekatan dalam *artificial intelligence* yang memungkinkan sistem mempelajari pola dari data untuk melakukan prediksi atau klasifikasi [14]. Dalam konteks kecocokan karier, *machine learning* digunakan untuk memodelkan hubungan antara profil individu, seperti pendidikan, keterampilan, dan pengalaman, dengan jalur karier yang sesuai.

Pendekatan ini umumnya menggunakan *supervised learning*, di mana model dilatih menggunakan data berlabel untuk memprediksi kategori pekerjaan berdasarkan pola historis [15]. Algoritma seperti Random Forest dan SVM banyak digunakan karena kemampuannya dalam menangani data berdimensi tinggi dan kompleks [16]. Penelitian menunjukkan bahwa metode ensemble learning seperti Random Forest mampu mengidentifikasi faktor penting dalam kecocokan karier, termasuk keterampilan teknis dan pengalaman kerja [17]. Dengan demikian, *machine learning* memberikan pendekatan yang lebih objektif dan berbasis data dibandingkan metode konvensional dalam menentukan kecocokan karier.

2.3 *Related Work* Penelitian Terdahulu

Berbagai penelitian telah dilakukan terkait prediksi karier, rekomendasi pekerjaan, dan *resume screening* berbasis *machine learning*. Assylzhan dkk. [4] mengembangkan sistem cerdas untuk mengevaluasi kesiapan karier mahasiswa menggunakan pendekatan berbasis data psikometrik dan perkembangan personal. Penelitian tersebut menunjukkan bahwa *machine learning* mampu membantu proses pengambilan keputusan karier secara lebih objektif.

Pada domain rekomendasi jalur karier, Siswipraptini dkk. [5] mengembangkan model rekomendasi personalisasi untuk mahasiswa teknologi informasi di Indonesia menggunakan

pendekatan *machine learning*. Penelitian tersebut berfokus pada pemetaan kompetensi mahasiswa terhadap jalur karier tertentu. Sementara itu, Faruque dkk. [6] memanfaatkan NLP untuk memprediksi karier mahasiswa ilmu komputer berdasarkan data akademik dan keterampilan teknis.

Penelitian terkait *career path prediction* juga dilakukan oleh Senger dkk. [7], yang menekankan pentingnya model prediksi yang mampu merepresentasikan perjalanan karier secara realistis. Selain itu, Nawaz dan Amin [15] mengembangkan kerangka *machine learning* untuk prediksi karier dan pengembangan kewirausahaan, sedangkan Sai dkk. [16] mengembangkan sistem rekomendasi jalur karier berbasis *machine learning* dengan pendekatan klasifikasi.

Pada konteks *resume screening*, T. K. dkk. [8] mengembangkan sistem pemeringkatan resume berbasis *machine learning* untuk meningkatkan efisiensi proses rekrutmen. Saatci dkk. [9] menerapkan NLP untuk melakukan ekstraksi informasi resume secara otomatis, sedangkan Sheikh dkk. [10] mengembangkan sistem *AI-powered resume ranking* untuk meningkatkan objektivitas seleksi kandidat.

Meskipun berbagai penelitian tersebut menunjukkan bahwa *machine learning* efektif dalam prediksi karier dan *resume screening*, sebagian besar penelitian masih berfokus pada domain spesifik atau hanya menitikberatkan pada proses rekrutmen. Penelitian ini memiliki perbedaan dengan penelitian sebelumnya karena mengembangkan sistem prediksi kecocokan karier pasca-studi berbasis resume lintas bidang dengan membandingkan performa Random Forest dan SVM pada dataset multi-kategori berskala besar. Selain itu, penelitian ini juga melakukan analisis *feature importance* untuk mengidentifikasi faktor utama yang memengaruhi kecocokan karier.

2.4 Resume Screening Berbasis Machine Learning

Resume screening merupakan tahap awal dalam proses pencocokan kandidat dengan pekerjaan, namun sering kali tidak efisien karena melibatkan volume data yang besar dan tidak terstruktur. Pendekatan tradisional berbasis pencocokan kata kunci cenderung menghasilkan bias dan kurang mampu memahami konteks informasi [8].

Penggunaan *machine learning* dan NLP memungkinkan otomatisasi proses ini dengan mengekstraksi informasi dari resume dan merepresentasikannya dalam bentuk numerik. Model kemudian dapat melakukan klasifikasi atau pemeringkatan kandidat berdasarkan tingkat kecocokan dengan pekerjaan tertentu [9].

Beberapa penelitian menunjukkan bahwa integrasi teknik ekstraksi fitur, vektorisasi teks, dan algoritma klasifikasi seperti Random Forest dan SVM mampu meningkatkan akurasi serta objektivitas dalam proses seleksi [10]. Oleh karena itu, pendekatan ini menjadi dasar dalam pengembangan sistem prediksi kecocokan karier berbasis resume.

2.5 Pemrosesan Teks dan Representasi Fitur

Pemrosesan teks merupakan tahap penting dalam mengubah data tidak terstruktur menjadi representasi numerik yang dapat diproses oleh algoritma *machine learning*. Tahapan ini meliputi pembersihan teks, tokenisasi, penghapusan *stopword*, serta normalisasi kata.

Metode *Term Frequency-Inverse Document Frequency* (TF-IDF) banyak digunakan karena kemampuannya dalam memberikan bobot pada kata yang memiliki nilai informasi tinggi. Penggunaan n-gram (misalnya unigram dan bigram) memungkinkan model menangkap konteks kata secara lebih baik.

Dalam klasifikasi teks, SVM dikenal efektif dalam menangani fitur berdimensi tinggi, sementara Random Forest unggul dalam stabilitas dan interpretabilitas model. Kombinasi teknik representasi teks yang tepat dengan algoritma klasifikasi yang sesuai menjadi faktor kunci dalam meningkatkan performa prediksi kecocokan karier.

2.6 Algoritma Klasifikasi: Random Forest dan SVM

Random Forest dan SVM merupakan dua algoritma yang umum digunakan dalam klasifikasi berbasis data teks. Random Forest adalah metode ensemble learning berbasis decision tree yang bekerja dengan menggabungkan banyak pohon keputusan untuk meningkatkan akurasi dan mengurangi *overfitting*. Keunggulan utama Random Forest adalah kemampuannya dalam

menangani data kompleks serta menyediakan informasi *feature importance* untuk interpretasi model [18].

Sementara itu, SVM bekerja dengan mencari *hyperplane* optimal yang memisahkan kelas data dengan margin maksimum. Algoritma ini sangat efektif untuk data berdimensi tinggi dan banyak digunakan dalam klasifikasi teks [19][20]. Namun, SVM memerlukan penyesuaian parameter yang lebih sensitif dibandingkan Random Forest.

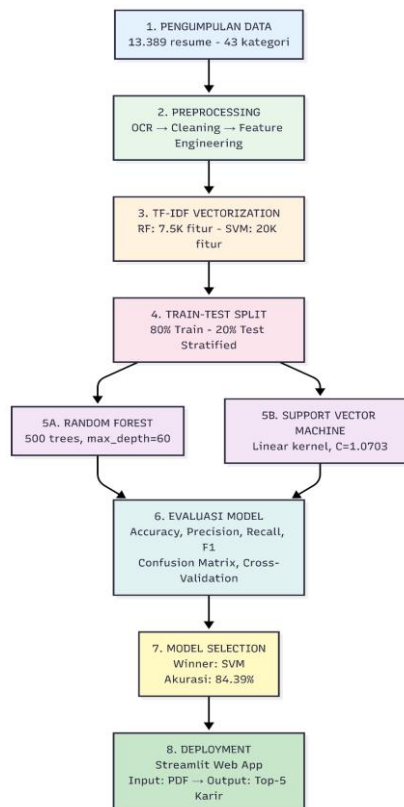
Secara umum, Random Forest unggul dalam interpretabilitas dan efisiensi pada dataset besar, sedangkan SVM lebih unggul dalam akurasi pada data berdimensi tinggi. Oleh karena itu, perbandingan kedua algoritma menjadi relevan untuk menentukan pendekatan terbaik dalam prediksi kecocokan karier.

3. METODE

3.1 Desain penelitian

Penelitian ini menggunakan pendekatan kuantitatif eksperimental dengan menerapkan metode *supervised machine learning* untuk memprediksi kecocokan karier pasca-studi berdasarkan data resume. Sistem yang dikembangkan memodelkan hubungan antara profil individu, meliputi latar belakang pendidikan, keterampilan, dan pengalaman kerja, dengan kategori karier yang sesuai.

Secara umum, alur penelitian terdiri dari tiga tahap utama, yaitu *input*, proses, dan *output*. Tahap *input* mencakup data resume sebagai variabel independen. Tahap proses melibatkan *preprocessing*, ekstraksi fitur, serta pelatihan model menggunakan algoritma Random Forest dan SVM. Tahap output berupa prediksi kecocokan karier dalam bentuk klasifikasi kategori pekerjaan. Diagram alur penelitian ditunjukkan pada Gambar 1.



Gambar 1. Diagram Alur Penelitian secara Umum

3.2 Dataset dan Pra-pemrosesan

Dataset yang digunakan merupakan *Resume Classification Dataset* yang bersumber dari repositori publik GitHub, terdiri dari 13.389 dokumen resume dengan 43 kategori karier. Setiap data memiliki dua atribut utama, yaitu Category sebagai label dan Text sebagai konten resume.

Data resume yang berasal dari format PDF dan gambar terlebih dahulu diekstraksi menggunakan *Optical Character Recognition (OCR)*, kemudian dilakukan *preprocessing* yang meliputi pembersihan teks, tokenisasi, penghapusan *stopword*, dan *lemmatization*.

Selain representasi teks utama, dilakukan pula rekayasa fitur tambahan, seperti jumlah keterampilan teknis (*technical skills count*), tingkat pendidikan (*education level*), kepadatan keterampilan (*skills density*), serta panjang teks resume. Tahapan ini bertujuan untuk meningkatkan kualitas representasi data sebelum digunakan dalam proses pembelajaran model. Gambar 2 menunjukkan contoh dataset *raw* yang digunakan.

```
Name: Howard Gerrard
Position: Accountant

Summary:
Experienced accountant with expertise in financial management, auditing, and reporting.
Skilled in handling financial transactions, budgeting, and ensuring data accuracy.

Work Experience:
- Accountant (2016-Present)
  Responsible for managing company finances, verifying financial records, and handling transactions.

- Accountant (2014-2016)
  Conducted financial audits, prepared reports, and managed budgeting processes.

Education:
- Bachelor Degree in Accounting

Skills:
- Accounting
- Financial Reporting
- Tax Management
- MS Word
```

Gambar 2. Contoh Dataset *Raw* yang digunakan

3.3 Representasi Fitur dan Model

Representasi teks dilakukan menggunakan metode TF-IDF dengan rentang n-gram (1,2). Konfigurasi TF-IDF disesuaikan untuk masing-masing algoritma, yaitu 7.500 fitur untuk RF dan 20.000 fitur untuk SVM, guna mengakomodasi karakteristik model yang berbeda.

Penelitian ini menggunakan dua algoritma klasifikasi, yaitu:

1. Random Forest. Random Forest merupakan metode ensemble learning berbasis decision tree yang menggabungkan beberapa pohon keputusan untuk meningkatkan akurasi dan mengurangi overfitting. Proses klasifikasi dilakukan melalui mekanisme *majority voting* dari seluruh pohon keputusan.
2. SVM. SVM merupakan algoritma yang bekerja dengan mencari *hyperplane* optimal untuk memisahkan kelas data dengan margin maksimum. Pada penelitian ini digunakan kernel linear dengan parameter penalti $C = 1.0703$, yang sesuai untuk data teks berdimensi tinggi.

Perbedaan konfigurasi TF-IDF antara Random Forest dan SVM dilakukan untuk mengoptimalkan performa masing-masing algoritma, di mana SVM cenderung optimal pada fitur berdimensi tinggi, sedangkan Random Forest lebih sensitif terhadap jumlah fitur yang berlebihan, seperti yang ditunjukkan pada Tabel 1 dan Tabel 2.

Tabel 1. Konfigurasi Hyperparameter TF-IDF dan RF

| Komponen | Parameter | Nilai |
|----------|-----------|-------|
|----------|-----------|-------|

| | | |
|---------------------------------|-----------------------|--------------------|
| TF-IDF Vectorizer | tfidf_analyzer | word |
| | tfidf_ngram_range | (1, 2) |
| | tfidf_min_df | 3 |
| | tfidf_max_df | 0.8 |
| | tfidf_max_features | 7500 |
| | tfidf_sublinear_tf | True |
| Random Forest Classifier | clf_n_estimators | 500 |
| | clf_max_depth | 60 |
| | clf_max_features | sqrt |
| | clf_min_samples_split | 5 |
| | clf_min_samples_leaf | 1 |
| | clf_max_samples | 0.8 |
| | clf_criterion | gini |
| | clf_class_weight | balanced_subsample |

Tabel 2. Konfigurasi Hyperparameter TF-IDF dan SVM

| Komponen | Parameter | Nilai |
|--------------------------|--------------------|---------------|
| TF-IDF Vectorizer | tfidf_analyzer | word |
| | tfidf_ngram_range | (1, 2) |
| | tfidf_min_df | 2 |
| | tfidf_max_df | 0.75 |
| | tfidf_max_features | 20000 |
| | tfidf_sublinear_tf | True |
| SVM Classifier | tfidf_norm | l2 |
| | clf_C | 1.0703 |
| | clf_loss | squared_hinge |
| | clf_penalty | l2 |
| | clf_dual | True |
| | clf_class_weight | None |

3.4 Proses Pelatihan dan Evaluasi

Dataset dibagi menjadi data latih dan data uji dengan rasio 80:20 menggunakan *stratified split* untuk menjaga keseimbangan distribusi kelas. Model dilatih menggunakan kedua algoritma secara paralel, kemudian dievaluasi menggunakan metrik: *accuracy*, *precision*, *recall*, *F1-score*. Selain itu, digunakan *confusion matrix* untuk menganalisis pola kesalahan klasifikasi serta *cross-validation* (validasi silang) untuk menguji konsistensi performa model. Model terbaik dipilih berdasarkan nilai evaluasi tertinggi.

3.5 Implementasi Sistem

Model yang telah dilatih diimplementasikan dalam bentuk aplikasi berbasis web menggunakan framework Streamlit. Sistem menerima input berupa dokumen resume dalam format PDF, kemudian menghasilkan rekomendasi lima kategori karier teratas beserta nilai *confidence score*. Implementasi ini bertujuan untuk memastikan bahwa model tidak hanya memiliki performa yang baik secara teoritis, tetapi juga dapat digunakan secara praktis sebagai alat bantu pengambilan keputusan karier.

4. HASIL DAN PEMBAHASAN

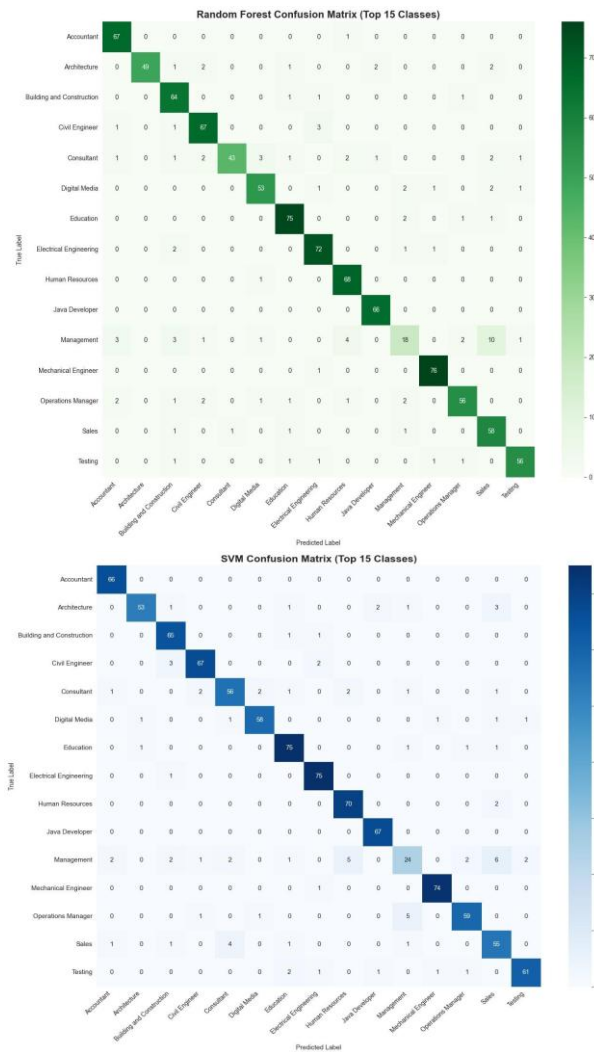
4.1 Hasil

Hasil pengujian yang ditunjukkan oleh Gambar 3(a) dan Gambar 3(b) menunjukkan bahwa kedua algoritma, SVM dan Random Forest, mampu melakukan klasifikasi resume dengan performa yang baik pada dataset multi-kategori. Evaluasi dilakukan menggunakan metrik *accuracy*, *precision*, *recall*, dan *F1-score*, serta dianalisis menggunakan *confusion matrix*.

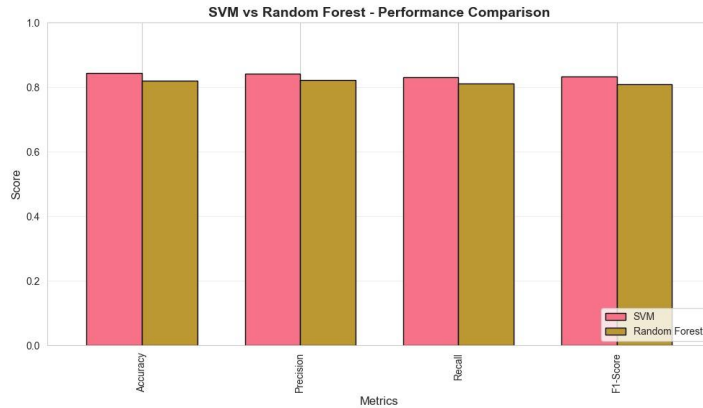
Model SVM menunjukkan performa terbaik dengan akurasi sebesar 84,39%, *precision* 84,29%, *recall* 83,16%, dan *F1-score* 83,36%. Sementara itu, Random Forest mencapai akurasi

81,96%, dengan *precision* 82,25%, *recall* 81,25%, dan *F1-score* 81,02%. Hasil ini menunjukkan bahwa SVM secara konsisten unggul pada seluruh metrik evaluasi seperti pada Gambar 4.

Analisis *confusion matrix* menunjukkan bahwa sebagian besar prediksi berada pada diagonal utama, yang mengindikasikan tingkat klasifikasi yang tinggi. Namun, kesalahan klasifikasi masih terjadi pada kategori yang memiliki kemiripan konteks, seperti *Management*, *Operations Manager*, dan *Sales*. Pola ini terlihat pada kedua model, meskipun lebih dominan pada Random Forest. Hasil validasi silang juga memperkuat temuan tersebut, di mana SVM mencapai nilai F1-macro sebesar 0,8174, lebih tinggi dibandingkan Random Forest sebesar 0,8044, yang menunjukkan bahwa SVM memiliki performa yang lebih konsisten pada seluruh kelas.

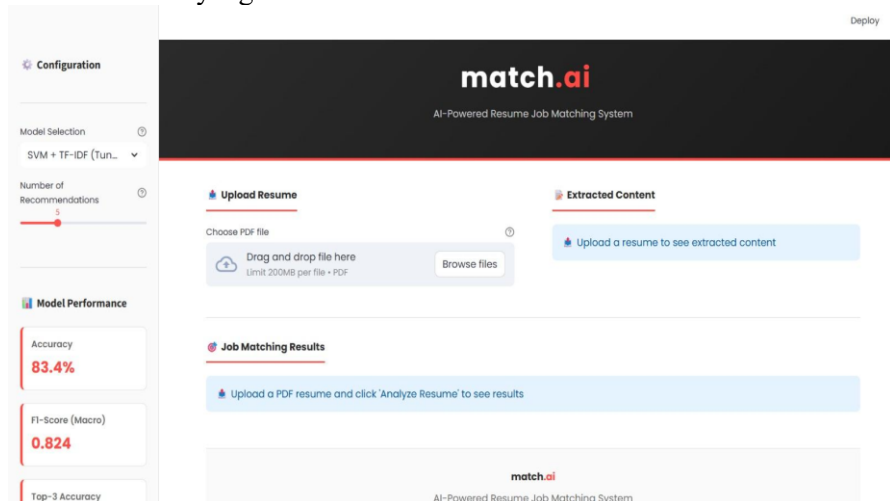


Gambar 3. Visualisasi *confusion matrix* : (a) Random Forest dan (b) SVM



Gambar 4. Komparasi Metrik Evaluasi RF dan SVM

Selanjutnya, desain UI/UX difokuskan pada kesederhanaan alur penggunaan, keterbacaan informasi, serta efisiensi interaksi, sehingga sistem dapat berfungsi sebagai alat bantu rekomendasi karir yang mudah digunakan dan intuitif sehingga memudahkan pengguna dalam memahami keluaran sistem seperti yang ditunjukkan oleh Gambar 5. Sementara itu, Tabel 3 menjelaskan fitur-fitur UI yang tersedia.



Gambar 5. User Interface

Tabel 3. Fitur - Fitur User Interface

| Komponen | Teknologi yang dipakai | Nilai |
|--------------------|----------------------------|---|
| Upload resume | Streamlit File Uploader | Menerima input file PDF resume (max 10MB) app.py |
| Ekstraksi teks | PyMuPDF | Mengonversi dokumen PDF menjadi teks app.py |
| Text preprocessing | Regular Expression | Pembersihan teks dan normalisasi app.py |
| Parsing resume | Pattern Matching | Mengidentifikasi bagian pendidikan, pengalaman, dan skill app.py |
| Pemilihan model | Streamlit Sidebar | Memilih algoritma RF atau SVM untuk prediksi app.py |
| Prediksi karir | Pre-trained Model (Pickle) | Menghasilkan top-5 rekomendasi karir dengan confidence score app.py |
| Visualisasi hasil | Streamlit Markdown | Menampilkan hasil prediksi secara interaktif app.py |

4.2 Pembahasan

Hasil penelitian menunjukkan bahwa SVM lebih unggul dalam klasifikasi resume berbasis teks dibandingkan Random Forest. Hal ini disebabkan oleh kemampuan SVM dalam membentuk *hyperplane* optimal pada ruang fitur berdimensi tinggi, sehingga mampu memisahkan kelas dengan margin yang lebih jelas. Representasi fitur berbasis TF-IDF dengan n-gram juga berperan penting dalam meningkatkan kemampuan model dalam menangkap pola semantik dari teks resume.

Sebaliknya, Random Forest menunjukkan performa yang stabil namun sedikit lebih rendah. Meskipun memiliki keunggulan dalam interpretabilitas melalui *feature importance*, pendekatan berbasis pohon keputusan cenderung kurang optimal pada data teks dengan dimensi fitur yang sangat besar. Hal ini menyebabkan peningkatan kesalahan klasifikasi pada kategori yang memiliki distribusi fitur yang timpang tindih.

Perbedaan performa antara kedua model relatif kecil, dengan selisih sekitar 2-3%, yang menunjukkan bahwa kedua algoritma sama-sama mampu menangkap pola dalam data dengan baik. Hal ini dipengaruhi oleh kualitas representasi fitur TF-IDF serta ukuran dataset yang cukup besar (13.389 data dengan 43 kategori), sehingga kedua model memiliki kapasitas pembelajaran yang memadai.

Namun demikian, keunggulan SVM dalam hal generalisasi menjadikannya lebih sesuai untuk implementasi sistem prediksi kecocokan karir berbasis teks. Di sisi lain, Random Forest tetap memiliki peran penting dalam memberikan interpretasi terhadap fitur-fitur yang berpengaruh, sehingga dapat digunakan sebagai model pendukung untuk meningkatkan transparansi sistem.

Berdasarkan temuan ini, pendekatan kombinasi (*hybrid approach*) berpotensi menjadi solusi yang lebih optimal, dengan memanfaatkan SVM sebagai model utama untuk prediksi dan Random Forest untuk analisis fitur. Pendekatan ini tidak hanya meningkatkan akurasi, tetapi juga memberikan nilai tambah dalam bentuk interpretabilitas, yang penting dalam sistem rekomendasi karir berbasis data.

5. KESIMPULAN

Penelitian ini berhasil mengembangkan sistem prediksi kecocokan karir pasca-studi berbasis *resume screening* menggunakan pendekatan *machine learning*. Model yang dibangun mampu memetakan hubungan antara profil lulusan, yang meliputi latar belakang pendidikan, keterampilan, dan pengalaman kerja, dengan kategori karir yang sesuai melalui representasi fitur berbasis TF-IDF.

Hasil evaluasi menunjukkan bahwa kedua algoritma, Random Forest dan SVM, memiliki performa yang baik dalam klasifikasi resume. Namun, SVM menunjukkan kinerja yang lebih unggul dan konsisten dengan akurasi sebesar 84,39% dan F1-score sebesar 83,36%, sehingga lebih sesuai untuk menangani data teks berdimensi tinggi. Analisis lebih lanjut menunjukkan bahwa kesalahan prediksi umumnya terjadi pada kategori dengan kemiripan konteks, sementara fitur seperti keterampilan teknis, pengalaman kerja, dan bidang studi menjadi faktor utama yang mempengaruhi kecocokan karir.

Secara keseluruhan, penelitian ini menunjukkan bahwa pendekatan berbasis machine learning dapat memberikan prediksi kecocokan karir yang lebih objektif dan terukur dibandingkan pendekatan konvensional. Sistem yang dikembangkan berpotensi digunakan sebagai alat bantu pengambilan keputusan karir bagi mahasiswa dan lulusan.

Penelitian selanjutnya dapat difokuskan pada perluasan dataset untuk meningkatkan generalisasi model, integrasi fitur non-tekstual seperti nilai akademik atau sertifikasi, serta eksplorasi metode *deep learning* berbasis transformer untuk meningkatkan pemahaman konteks semantik.

6. UCAPAN TERIMA KASIH

Terima kasih kepada riset grup *Robotics, Mechatronics, and Intelligent Systems*, Fakultas Teknologi Maju dan Multidisiplin, Universitas Airlangga atas bantuan dan masukannya selama penelitian ini dilakukan.

7. DAFTAR PUSTAKA

- [1] Z. Yonatan, "35% Pemuda RI Bekerja Tidak Sesuai Tingkat Pendidikan," *GoodStats*, 2025. [Online]. Available: <https://data.goodstats.id/statistic/35-pemuda-ri-bekerja-tidak-sesuai-tingkat-pendidikan-YyNTc>. [Accessed: Dec. 10, 2025].
- [2] Badan Pusat Statistik, "Keadaan Ketenagakerjaan Indonesia Februari 2025," Feb. 2025. [Online]. Available: <https://www.bps.go.id/id/infographic?id=1112>. [Accessed: Dec. 10, 2025].

- [3] S. Jiang and Y. Guo, "Reasons for college major-job mismatch and labor market outcomes: Evidence from China," *China Economic Review*, vol. 74, p. 101822, 2022, doi: 10.1016/j.chieco.2022.101822.
- [4] I. Assylzhan, M. Muratbekova, D. Amangeldi, N. Oryngozha, A. Ogorodova, and P. Shamoi, "Intelligent System for Assessing University Student Personality Development and Career Readiness," *arXiv preprint*, 2023, doi: 10.48550/arXiv.2308.15620.
- [5] P. C. Siswipraptini, L. Hendric, A. Ramadhan, and W. Budiharto, "Personalized Career-Path Recommendation Model for Information
- [6] S. H. Faruque, K. S. Akter, and S. Akter, "Unlocking Futures: A Natural Language Driven Career Prediction System for Computer Science and Software Engineering Students," *arXiv preprint*, 2024, doi: 10.48550/arXiv.2405.18139.
- [7] E. Senger, Y. Campbell, R. van der Goot, and B. Plank, "Toward more realistic career path prediction: evaluation and methods," *Frontiers in Big Data*, vol. 8, 2025, doi: 10.3389/fdata.2025.1564521.
- [8] T. K., U. V., S. M. Kadiwal, and S. Revanna, "Design and Development of Machine Learning Based Resume Ranking System," *Global Transitions Proceedings*, vol. 3, no. 2, 2021, doi: 10.1016/j.gltip.2021.10.002.
- [9] M. Saatci, R. Kaya, and R. Ünlü, "Resume Screening With Natural Language Processing (NLP)," *Alphanumeric Journal*, vol. 12, no. 2, 2024, doi: 10.17093/alphanumeric.1536577.
- [10] S. Sheikh, P. Adep, N. Aidasani, S. Chavan, and V. Darade, "AI-Powered Resume Ranking System: Enhancing Recruitment Efficiency through Natural Language Processing," *International Journal for Research Trends and Innovation*, vol. 10, no. 5, 2025.
- [11] H. A. Abdulkarem and A. M. Naemah, "Between Specialization and Employment: The Dilemma of Placing Highly Educated Individuals in Irrelevant Fields," *Al-Ghary Journal of Economic and Administrative Sciences*, vol. 20, no. 4, pp. 402–426, 2024, doi: 10.36325/ghjec.v20i4.17420.
- [12] M. Zahran, "Pendidikan Tinggi, Upah Rendah: 35% Pekerja Muda Salah Jurusan," *CNBC Indonesia*, Nov. 2025. [Online]. Available: <https://www.cnbcindonesia.com/research/20251104154304-128-682220/pendidikan-tinggi-upah-rendah-35-pekerja-muda-salah-jurusan>. [Accessed: Dec. 11, 2025].
- [13] A. de la Fuente, Y. Muñoz, P. Castro, and J. L. Arroyo, "Gamificación, estrategia compartida entre universidad, empresa y millennials," *REDU: Revista de Docencia Universitaria*, vol. 17, no. 2, pp. 73–73, 2019, doi: 10.4995/redu.2019.11479.
- [14] A. S. Chauhan and H. M. Henrietta, "Machine Learning Basics: A Comprehensive Guide: A Review," *Babylonian Journal of Machine Learning*, pp. 31–34, 2023, doi: 10.58496/bjml/2023/006.
- [15] M. Nawaz and N. Amin, "Machine Learning Framework for Career Prediction and Entrepreneurial Development," *Preprints*, 2025, doi: 10.20944/preprints202511.1625.v1.
- [16] G. Sai, M. L. Pandala, D. T. Veeranki, and P. Kumbha, "Career Compass: A Career Path Recommender Using Machine Learning," in *Proc. 5th Int. Conf. Electronics and Sustainable Communication Systems (ICESC)*, 2024, pp. 1937–1941, doi: 10.1109/ICESC60852.2024.10689897.
- [17] B. Thomas and A. K. John, "Machine Learning Techniques for Recommender Systems – A Comparative Case Analysis," *IOP Conference Series: Materials Science and Engineering*, vol. 1085, no. 1, p. 012011, 2021, doi: 10.1088/1757-899X/1085/1/012011.
- [18] A. Wakade, A. Wakde, R. Maywade, A. Pandey, J. Kumar, and A. K. Singh, "An Ensemble Learning Based Career Prediction Model," *Lecture Notes in Networks and Systems*, pp. 503–512, 2024, doi: 10.1007/978-3-031-60935-0_45.
- [19] M. Sheykhmousa, M. Mahdianpari, H. Ghanbari, F. Mohammadimanes, P. Ghamisi, and S. Homayouni, "Support Vector Machine Versus Random Forest for Remote Sensing Image Classification: A Meta-Analysis and Systematic Review," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 13, pp. 6308–6325, 2020, doi: 10.1109/JSTARS.2020.3026724.
- [20] J. T. Iorzua, T. Moses, C. I. Eke, O. J. Agushaka, D. K. Kwaghtyo, and T. Godswill, "A Machine Learning Based Approach to Course and Career Recommendation System: A Systematic Literature Review," *Journal of Computing Theories and Applications*, vol. 3, no. 1, pp. 1–16, 2025, doi: 10.62411/jcta.12603.