

# Application of the Support Vector Machine (SVM) method to predict career choices for UMSIDA alumni

## [Penerapan metode Support Vector Machine (SVM) untuk memprediksi pemilihan karir bagi alumni UMSIDA]

Meisyilia Difanada Qur'ani<sup>1)</sup>, Hamzah Setiawan<sup>\*,2)</sup>

<sup>1)</sup>Program Studi Teknik Informatika, Universitas Muhammadiyah Sidoarjo, Indonesia

\*Email Penulis Korespondensi: hamzah@umsida.ac.id

**Abstract.** *The success of higher education is not only determined by the educational process, but also the ability of its graduates to get jobs. This research aims to develop and evaluate a predictive model using the Support Vector Machine (SVM) method to predict career choices for alumni of Muhammadiyah University of Sidoarjo (UMSIDA). A quantitative approach was used with data from UMSIDA alumni tracers. The title of this research is "Application of the Support Vector Machine (SVM) Method to Predict Career Choices for UMSIDA Alumni". The evaluation results show that SVM has good performance with high precision, recall and f1-score in the dominant class. Feature analysis shows key factors influencing career choice. This model achieves 97% accuracy, providing precise recommendations for alumni.*

**Keywords -** Prediction; Career; Support Vector Machines; Umsida

**Abstrak.** *Keberhasilan perguruan tinggi tidak hanya ditentukan oleh proses pendidikannya, tetapi juga kemampuan lulusannya mendapatkan pekerjaan. Penelitian ini bertujuan mengembangkan dan mengevaluasi model prediktif menggunakan metode Support Vector Machine (SVM) untuk memprediksi pemilihan karir alumni Universitas Muhammadiyah Sidoarjo (UMSIDA). Pendekatan kuantitatif digunakan dengan data dari tracer alumni UMSIDA. Judul penelitian ini adalah "Penerapan Metode Support Vector Machine (SVM) untuk Memprediksi Pemilihan Karir Alumni UMSIDA". Hasil evaluasi menunjukkan SVM memiliki kinerja baik dengan precision, recall, dan f1-score tinggi pada kelas dominan. Analisis fitur menunjukkan faktor-faktor kunci yang mempengaruhi pemilihan karir. Model ini mencapai akurasi 97%, memberikan rekomendasi tepat bagi alumni.*

**Kata Kunci -** Prediksi; Karir; Support Vector Machine; Umsida

## I. PENDAHULUAN

Keberhasilan sebuah perguruan tinggi tidak hanya ditentukan oleh proses pendidikannya tetapi juga dari kemampuan lulusannya dalam mendapatkan pekerjaan[1]. Pendidikan yang berkualitas memberikan peluang yang lebih besar bagi universitas untuk menghasilkan lulusan yang berkualitas dan sukses di pasar kerja[2]. Sebagai lembaga akademik, penelitian, dan pengabdian kepada masyarakat, perguruan tinggi harus fokus pada pencapaian tujuan kinerja. Banyak perguruan tinggi yang berkompetisi dalam membangun pendidikan berkualitas guna menciptakan lulusan yang sesuai dengan kebutuhan pasar kerja[3].

Peningkatan jumlah lulusan perguruan tinggi setiap tahun di Indonesia membawa tantangan baru bagi lembaga pendidikan tinggi dalam memastikan para lulusan dapat terserap di dunia kerja sesuai dengan kompetensi yang mereka miliki. Menurut data dari Badan Pusat Statistik (BPS), pada Februari 2024, tingkat pengangguran terbuka (TPT) di kalangan lulusan perguruan tinggi mencapai 5,8%, meningkat dibandingkan dengan tahun-tahun sebelumnya. Hal ini menunjukkan adanya ketidaksesuaian antara kualifikasi yang dimiliki oleh lulusan dengan kebutuhan pasar kerja.

Dalam era globalisasi dan persaingan pasar kerja yang semakin ketat, alumni perguruan tinggi untuk tidak hanya memberikan pendidikan berkualitas tetapi juga memastikan lulusannya siap bersaing di dunia kerja[4]. Tingginya angka pengangguran di Indonesia menyoroti peran penting perguruan tinggi dalam mempersiapkan lulusan yang sesuai dengan kebutuhan industri[5]. Universitas Muhammadiyah Sidoarjo sebagai salah satu universitas terkemuka di Indonesia, berupaya untuk meningkatkan kualitas lulusannya melalui berbagai program akademik dan pengembangan keterampilan. Dengan mengetahui fakultas mana yang lulusannya paling banyak diterima kerja, perguruan tinggi dapat melakukan perbaikan pada program studi lain, sehingga meningkatkan keseluruhan mutu pendidikan dan kualitas lulusannya[6].

Penelitian yang dilakukan oleh Ari Lathifah berfokus pada evaluasi relevansi kurikulum Program Studi Ekonomi Syariah di UIN Sayyid Ali Rahmatullah Tulungagung melalui metode tracer study. Penelitian ini bertujuan untuk menilai sejauh mana kurikulum yang diterapkan oleh program studi tersebut sesuai dengan kebutuhan pasar kerja berdasarkan umpan balik dari lulusan. Pendekatan ini, meskipun memberikan wawasan yang berharga, cenderung bergantung pada data kuantitatif yang dikumpulkan dari survei lulusan dan mungkin tidak sepenuhnya mengeksplorasi

aspek kualitatif yang dapat memperdalam pemahaman tentang relevansi kurikulum. Dalam konteks ini, penelitian Anda yang menggunakan metode Support Vector Machine (SVM) dapat menawarkan kontribusi yang signifikan untuk mengatasi beberapa kekurangan yang mungkin ada dalam penelitian Ari Lathifah. SVM, sebagai teknik machine learning, dapat digunakan untuk menganalisis data kualitatif dengan cara yang lebih mendalam. Misalnya, jika penelitian Anda melibatkan analisis umpan balik teks dari lulusan atau data kualitatif lainnya, SVM dapat mengklasifikasikan dan mengidentifikasi pola atau tema yang mungkin tidak terlihat dalam analisis kuantitatif tradisional[7].

Metode SVM memungkinkan untuk pemrosesan data yang lebih kompleks, seperti klasifikasi dan prediksi berdasarkan fitur-fitur yang diekstrak dari data kualitatif[8]. Dengan memanfaatkan SVM, Anda dapat membangun model yang lebih canggih dalam memahami kebutuhan pasar atau evaluasi kurikulum[9]. Hal ini bisa mengungkap informasi yang lebih terperinci dan relevan tentang bagaimana kurikulum dapat disesuaikan dengan tuntutan industri dan kebutuhan lulusan, memberikan hasil yang lebih komprehensif dan akurat dibandingkan dengan pendekatan yang hanya mengandalkan data kuantitatif.[10].

Penelitian ini akan difokuskan pada alumni Universitas Muhammadiyah Sidoarjo, dengan data yang dikumpulkan melalui tracer study menggunakan kuesioner yang akan disebar kepada 1.961 alumni dari 6 fakultas. Tujuan dari penelitian ini adalah untuk mengembangkan dan mengevaluasi model prediktif menggunakan metode Support Vector Machine (SVM) guna memprediksi pemilihan karir bagi alumni Universitas Muhammadiyah Sidoarjo (UMSIDA)[11].

Tujuan penelitian ini adalah untuk mengembangkan sebuah model prediktif berbasis metode Support Vector Machine (SVM) yang mampu secara akurat memprediksi pilihan karir alumni Universitas Muhammadiyah Sidoarjo (UMSIDA) berdasarkan berbagai faktor yang relevan, termasuk latar belakang pendidikan, prestasi akademik, pengalaman kerja, dan preferensi pribadi. Penelitian ini berupaya untuk mengidentifikasi pola dan hubungan antara variabel-variabel tersebut yang dapat memberikan wawasan mendalam mengenai faktor-faktor utama yang memengaruhi keputusan karir alumni. Dengan mengembangkan model prediktif yang kuat, penelitian ini bertujuan untuk menyediakan alat analitik yang dapat digunakan oleh universitas dalam merancang kebijakan yang lebih efektif untuk mendukung proses penempatan kerja lulusan. Selain itu, penelitian ini juga bertujuan untuk mengeksplorasi penggunaan SVM sebagai pendekatan komputasional yang optimal dalam memanfaatkan data historis dan profil individu mahasiswa untuk menghasilkan prediksi karir yang lebih akurat dan relevan[12]. Model ini diharapkan dapat memberikan rekomendasi karir yang lebih personal dan sesuai dengan profil unik setiap lulusan, sehingga meningkatkan tingkat kepuasan dan keberhasilan karir mereka di masa depan. Penelitian ini juga berfokus pada upaya untuk meningkatkan tingkat penyerapan kerja alumni di pasar tenaga kerja dengan memanfaatkan prediksi yang didasarkan pada analisis mendalam terhadap data mahasiswa, termasuk identifikasi keterampilan yang paling dibutuhkan oleh industri dan tren pasar kerja terbaru[13].

Penelitian ini bermaksud memberikan kontribusi signifikan terhadap pengembangan kebijakan dan praktik di UMSIDA melalui penerapan metode berbasis data yang dapat diandalkan untuk bimbingan karir dan pengembangan kurikulum. Dengan menggunakan model prediktif ini, universitas diharapkan mampu menyusun strategi pengembangan kurikulum yang lebih relevan dengan kebutuhan industri, merancang program pelatihan yang lebih terarah, dan memperkuat kerjasama dengan berbagai pihak eksternal, seperti perusahaan dan lembaga industri, untuk memastikan lulusan memiliki keterampilan dan kompetensi yang sesuai dengan permintaan pasar. Tujuan lain dari penelitian ini adalah untuk memberikan dasar yang kuat bagi pengambilan keputusan berbasis data dalam penyusunan program bimbingan karir, sehingga UMSIDA dapat lebih efektif dalam memfasilitasi proses transisi lulusan dari dunia akademik ke dunia profesional, dengan meningkatkan kesiapan kerja mereka dan mempercepat proses penempatan kerja.

Hasil penelitian ini akan memberikan pemahaman yang lebih baik tentang keberhasilan dalam menciptakan lulusan yang siap terjun ke dunia kerja dan diharapkan dapat memberikan wawasan mendalam tentang fakultas mana yang paling berhasil dalam mempersiapkan lulusannya, semakin cepat alumni memperoleh pekerjaan, semakin baik mutu pendidikan yang diberikan. Selain itu, penelitian ini juga akan membantu mengidentifikasi area yang memerlukan perbaikan, sehingga Universitas Muhammadiyah Sidoarjo dapat meningkatkan kualitas pendidikan dan kesesuaian lulusannya dengan tuntutan pasar kerja dan akan memberikan kontribusi penting dalam pengembangan strategi pendidikan yang lebih efektif dan relevan, serta dapat dijadikan acuan bagi calon mahasiswa dalam memilih fakultas yang sesuai dengan minat dan peluang kerja yang tinggi.

## II. METODE

Penelitian ini menggunakan pendekatan kuantitatif, dalam topik memprediksi sampel data diperoleh dari data tracer mahasiswa umside yang disusun menjadi judul “Penerapan Metode Support Vector Machine (SVM) untuk Memprediksi Pemilihan Karir bagi Alumni UMSIDA”.

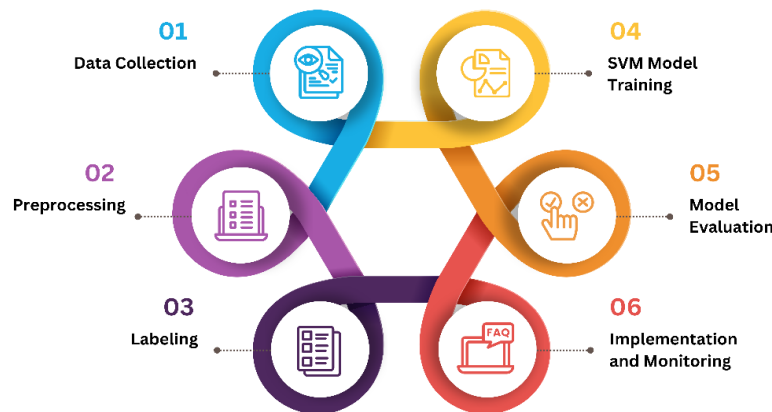
$$y = \text{sign} \sum_{i=1}^n a_i y_i K(x_i, x) + b$$

. Keterangan :

- $y$  = prediksi pemilihan karir (misalnya, 1 untuk "Karir A", -1 untuk "Karir B").  
 $a_i$  = koefisien Lagrange yang diperoleh dari pelatihan.  
 $y_i$  = label kelas data pelatihan  $x_i$  (misalnya, 1 untuk "Karir A", -1 untuk "Karir B").  
 $K(x_i, x)$  = fungsi kernel yang memetakan data ke ruang dimensi yang lebih tinggi.  
 $b$  = bias yang diperoleh dari pelatihan.

Rumus ini digunakan untuk memprediksi pilihan karir alumni berdasarkan fitur-fitur mereka (misalnya, Nama, Jenis Kelamin, Prodi, Alamat, Bidang Kerja dan Jml Pendapatan) yang diwakili oleh  $x$ . Pada tahap implementasi dengan menggunakan Google Colaboratory.

Penelitian ini akan akan memprediksi, mengeksplorasi beragam perspektif, dan pandangan yang tercermin apa saja yang mempengaruhi dalam pemilihan sebuah karir. Adapun alur dalam penelitian ini menggunakan 6 langkah, yaitu Data Collection, Preprocessing, Labeling, SVM Model Training, Model Evaluation, dan Implementation and Monitoring[14]. Gambar 1 merupakan alur dari penelitian ini.



**Gambar 1.** Alur Penelitian

Rumus ini digunakan untuk memprediksi pilihan karir alumni berdasarkan fitur-fitur mereka (misalnya, Nama, Jenis Kelamin, Prodi, Alamat, Bidang Kerja dan Jml Pendapatan) yang diwakili oleh  $x$ . Pada tahap implementasi dengan menggunakan Google Colaboratory.

#### A. Data Collection

Proses penerapan Support Vector Machine (SVM) untuk memprediksi pemilihan karir bagi alumni Universitas Muhammadiyah Sidoarjo (UMSIDA) dimulai dengan pengumpulan data yang komprehensif dan relevan. Data ini dapat diperoleh melalui survei alumni yang mencakup informasi mengenai latar belakang mahasiswa, bidang kerja, jabatan, jumlah pendapatan, serta pengalaman kerja. Selain itu, data juga bisa didapat dari catatan akademik universitas dan sumber lain yang relevan. Pengumpulan data harus dilakukan dengan cermat untuk memastikan kelengkapan dan akurasi informasi yang diperoleh[15]. Dalam penelitian ini, pemilihan fitur untuk model Support Vector Machine (SVM) bertujuan untuk meningkatkan akurasi prediksi pemilihan karir alumni Universitas Muhammadiyah Sidoarjo (UMSIDA). Fitur "Nama," "Jenis Kelamin," "Alamat," "Bidang Kerja," dan "Jumlah Pendapatan" memiliki peran yang berbeda namun saling melengkapi dalam konteks ini.

##### 1) Nama

Nama berfungsi sebagai identifier unik yang penting untuk memastikan integritas data dan memungkinkan analisis individual yang terpisah. Meskipun tidak langsung mempengaruhi pemilihan karir, fitur ini diperlukan untuk pengelolaan dan pengolahan data yang akurat.

##### 2) Jenis Kelamin

Jenis Kelamin memberikan wawasan tentang kecenderungan gender dalam pilihan karir. Beberapa sektor industri mungkin menunjukkan pola gender tertentu, dan analisis fitur ini membantu dalam memahami bagaimana gender dapat mempengaruhi preferensi karir serta dalam merancang rekomendasi yang lebih inklusif.

##### 3) Alamat

Alamat berhubungan dengan ketersediaan pekerjaan dan biaya hidup di lokasi tertentu. Lokasi tempat tinggal dapat mempengaruhi keputusan karir melalui akses ke peluang kerja yang relevan dan kebutuhan akan kompensasi yang

sesuai dengan biaya hidup lokal. Fitur ini juga memperhitungkan mobilitas dan kesiapan lulusan untuk berpindah lokasi kerja.

#### 4) Bidang Kerja

Bidang Kerja menunjukkan sektor atau industri yang diminati oleh lulusan. Fitur ini penting untuk memetakan kesesuaian antara latar belakang pendidikan dan pengalaman kerja dengan industri atau jenis pekerjaan yang dicari oleh lulusan. Informasi ini memungkinkan model untuk memberikan rekomendasi yang lebih sesuai dengan preferensi dan kebutuhan pasar kerja.

#### 5) Jumlah Pendapatan

Jumlah Pendapatan mencerminkan level kompensasi yang diharapkan atau diterima oleh lulusan dalam bidang kerja tertentu. Fitur ini memberikan indikasi tentang harapan pendapatan yang dapat mempengaruhi keputusan karir dan membantu dalam merancang strategi pengembangan karir yang sesuai dengan ekspektasi finansial lulusan.

### B. Data Collection

Setelah data terkumpul, langkah selanjutnya adalah pra-pemrosesan data. Tahap ini melibatkan pembersihan data untuk menghilangkan entri yang tidak lengkap atau tidak relevan. Selain itu, normalisasi dilakukan untuk memastikan bahwa semua fitur berada dalam skala yang sama, yang penting untuk kinerja SVM. Data kategorikal, latar belakang mahasiswa dan jumlah pendapatan, perlu diubah menjadi format numerik menggunakan teknik encoding, seperti one-hot encoding, untuk memungkinkan pemrosesan oleh model SVM[16].

### C. Labelling

Proses berikutnya adalah labeling, di mana data yang telah diproses diberi label sesuai dengan kategori karir yang diinginkan. Label ini menjadi target yang akan diprediksi oleh model SVM. Misalnya, label bisa berupa jenis pekerjaan atau sektor industri yang dipilih oleh alumni[17].

### D. SVM Model Training

Setelah data diproses dan diberi label, langkah berikutnya adalah training model SVM. Pada tahap ini, data dibagi menjadi set pelatihan dan set pengujian untuk memungkinkan evaluasi kinerja model. Dalam proses pelatihan, model SVM dilatih menggunakan data pelatihan dengan memilih kernel yang tepat, seperti kernel linear, polynomial, atau RBF (Radial Basis Function), dan mengoptimalkan parameter-parameter model, seperti C (regularization parameter) dan gamma. Proses pelatihan ini bertujuan untuk menemukan hyperplane yang memisahkan data dengan margin maksimum[18].

### E. Evaluation

Evaluasi model dilakukan untuk mengukur kinerja SVM menggunakan metrik-metrik evaluasi seperti akurasi, precision, recall, dan F1-score. Teknik cross-validation dapat digunakan untuk memastikan bahwa model tidak overfitting dan dapat di-generalize dengan baik ke data baru. Confusion matrix juga dapat digunakan untuk menganalisis performa prediksi model secara lebih rinci[19]. Pemilihan kernel yang tepat dalam model Support Vector Machine (SVM) sangat penting untuk memaksimalkan akurasi prediksi pemilihan karir alumni Universitas Muhammadiyah Sidoarjo (UMSIDA). Kernel digunakan untuk memetakan data ke dimensi yang lebih tinggi agar data tersebut dapat dipisahkan dengan lebih baik. Terdapat tiga jenis kernel yang umum digunakan dalam SVM, yaitu linear, polynomial, dan Radial Basis Function (RBF), masing-masing memiliki karakteristik dan pengaruh yang berbeda terhadap hasil prediksi.

Kernel linear cocok digunakan ketika data dapat dipisahkan dengan garis lurus pada ruang fitur asli. Kernel ini sederhana dan cepat dalam perhitungan, sehingga ideal untuk dataset besar yang memiliki kecenderungan linier. Namun, performa kernel linear menurun jika data tidak dapat dipisahkan secara linier. Sebaliknya, kernel polynomial dapat memodelkan hubungan yang lebih kompleks dengan mengubah data ke dimensi yang lebih tinggi menggunakan fungsi polinomial. Kernel ini berguna jika hubungan antara fitur dan variabel target tidak linier tetapi masih dapat dimodelkan dengan persamaan polinomial. Namun, risiko overfitting meningkat jika derajat polinomial yang dipilih terlalu tinggi.

Kernel RBF sering kali dipilih karena fleksibilitasnya dalam menangani berbagai jenis data, baik yang linier maupun non-linier. Kernel ini mampu memetakan data ke dimensi yang sangat tinggi, memungkinkan model untuk menemukan pemisah yang optimal meskipun data sangat kompleks. Dalam konteks prediksi karir, di mana hubungan antara fitur-fitur mungkin sangat kompleks dan tidak terstruktur, kernel RBF cenderung memberikan hasil yang lebih baik dibandingkan dengan kernel lainnya. Meskipun kernel ini memerlukan tuning parameter yang lebih hati-hati dan komputasi yang lebih mahal, performanya yang konsisten dan akurat menjadikannya pilihan yang unggul.

Pemilihan kernel terbaik ditentukan melalui perbandingan performa menggunakan metode validasi silang untuk mengevaluasi akurasi, presisi, recall, dan F1-score. Hasil analisis menunjukkan bahwa kernel RBF menghasilkan performa yang paling baik untuk data dengan hubungan non-linier yang kompleks, yang sering dijumpai dalam

analisis karir. Oleh karena itu, kernel RBF dipilih sebagai kernel utama dalam penelitian ini, karena mampu memberikan akurasi yang lebih tinggi dan prediksi yang lebih stabil dibandingkan kernel linear dan polynomial.

#### **F. Implementation And Monitoring**

Setelah model dievaluasi dan kinerjanya dianggap memadai, langkah selanjutnya adalah implementasi dan monitoring. Model SVM yang telah dilatih diimplementasikan ke dalam sistem yang dapat digunakan oleh alumni dan pihak universitas. Sistem ini memungkinkan prediksi pemilihan karir yang berbasis data dan dapat memberikan rekomendasi yang lebih baik dan tepat sasaran bagi para alumni. Monitoring dilakukan secara berkala untuk memastikan model tetap akurat dan relevan. Ini termasuk mengupdate model dengan data baru dan melakukan retraining jika diperlukan[20].

Dengan demikian, penerapan SVM dalam memprediksi pemilihan karir bagi alumni UMSIDA dapat memberikan manfaat yang signifikan, baik bagi alumni dalam menentukan jalur karir yang sesuai, maupun bagi universitas dalam meningkatkan layanan bimbingan karir. Proses ini melibatkan serangkaian langkah yang sistematis mulai dari pengumpulan data hingga implementasi dan monitoring, yang kesemuanya penting untuk memastikan akurasi dan keberhasilan model prediksi.

### **III. HASIL DAN PEMBAHASAN**

Data yang digunakan dalam penelitian ini diperoleh dari data tracer mahasiswa umside dengan menggunakan metode Career Pathway Assessment Survey. Total terdapat jumlah Data yang diambil ada 1.961. Selanjutnya didapat data akan masuk pada tahap preprocessing. Tabel 1 menunjukkan hasil dari proses preprocessing data yang melibatkan tahapan preprocessing. Proses penerapan metode Support Vector Machine (SVM) untuk memprediksi pemilihan karir bagi alumni Universitas Muhammadiyah Sidoarjo (UMSIDA) melibatkan serangkaian tahapan penting yang mencakup Data Integration, Data Transformation, Attribute Correlation dan Data Cleaning. Langkah pertama dalam proses ini adalah Data Integration, yang bertujuan untuk mengumpulkan dan menggabungkan data dari berbagai sumber yang relevan. Data ini dapat berasal dari survei kuesioner online, catatan akademik universitas, dan platform alumni. Proses integrasi ini memastikan bahwa semua data yang diperoleh dari berbagai sumber tersebut dikonsolidasi ke dalam satu database yang komprehensif. Hal ini memerlukan pengaturan yang sistematis agar data yang beragam dapat disatukan dengan format yang konsisten, sehingga mempermudah analisis lebih lanjut.

Setelah data terintegrasi, langkah berikutnya adalah Data Transformation. Data Transformation melibatkan penyesuaian dan perubahan data mentah ke dalam format yang dapat digunakan untuk analisis SVM. Ini mencakup normalisasi data untuk memastikan bahwa semua fitur berada dalam skala yang sama, yang penting untuk kinerja optimal dari model SVM. Selain itu, transformasi juga mencakup proses encoding data kategorikal menjadi format numerik, seperti menggunakan one-hot encoding untuk mengubah data kategorikal seperti jurusan atau jenis keterampilan menjadi variabel numerik yang dapat diproses oleh algoritma SVM. Dalam model Support Vector Machine (SVM) yang digunakan untuk memprediksi pemilihan karir alumni Universitas Muhammadiyah Sidoarjo (UMSIDA), teknik encoding untuk data kategorikal sangat penting untuk memastikan bahwa data non-numerik dapat digunakan secara efektif oleh algoritma SVM, yang bekerja dengan nilai numerik. Teknik encoding yang digunakan dalam penelitian ini termasuk one-hot encoding dan label encoding, yang masing-masing diterapkan tergantung pada jenis dan jumlah kategori dalam fitur data.

One-hot encoding digunakan untuk mengubah fitur kategorikal yang memiliki sejumlah kategori terbatas menjadi format numerik. Dalam proses ini, setiap kategori unik dari fitur tersebut diubah menjadi kolom biner (0 atau 1). Misalnya sebagai contoh, untuk fitur "Jenis Kelamin" yang memiliki dua kategori — "Pria" dan "Wanita" — proses one-hot encoding akan menghasilkan dua kolom: "Jenis Kelamin\_Pr" dan "Jenis Kelamin\_Wanita". Jika seorang individu berjenis kelamin "Pria," maka kolom "Jenis Kelamin\_Pr" akan diberi nilai 1, sedangkan "Jenis Kelamin\_Wanita" akan diberi nilai 0, dan sebaliknya. Teknik ini memastikan bahwa tidak ada urutan atau skala yang salah antara kategori dan membuat data lebih cocok untuk dimasukkan ke dalam model SVM.

Tahapan selanjutnya adalah Attribute Correlation, yang bertujuan untuk memahami hubungan antara berbagai atribut dalam dataset. Attribute Correlation ini penting untuk mengidentifikasi fitur-fitur yang memiliki hubungan kuat dengan pemilihan karir alumni. Teknik seperti matriks korelasi dan analisis statistik dapat digunakan untuk mengevaluasi sejauh mana atribut-atribut tertentu berkontribusi terhadap variabel target, yaitu pemilihan karir. Attribute Correlation fitur sangat penting dalam pemilihan fitur untuk model Support Vector Machine (SVM) karena menentukan seberapa besar pengaruh setiap fitur terhadap variabel target. Fitur yang memiliki korelasi kuat dengan variabel target, seperti "Jumlah Pendapatan" atau "Bidang Kerja," memberikan informasi yang sangat relevan untuk meningkatkan akurasi prediksi karir. Fitur ini mencerminkan faktor penting yang mempengaruhi keputusan karir alumni, seperti preferensi terhadap sektor tertentu atau harapan pendapatan. Dengan memahami korelasi ini, peneliti dapat memilih fitur-fitur yang paling relevan dan mengeliminasi yang kurang signifikan, sehingga meningkatkan efisiensi dan akurasi model prediksi.

Data Cleaning adalah langkah penting yang tidak dapat diabaikan dalam proses ini. Data yang dikumpulkan dari berbagai sumber seringkali mengandung ketidaklengkapan, inkonsistensi, atau kesalahan yang perlu diperbaiki. Data Cleaning melibatkan penghapusan atau perbaikan data yang hilang, duplikat, atau tidak valid untuk memastikan bahwa dataset akhir yang digunakan untuk pelatihan model adalah bersih dan berkualitas tinggi. Proses ini mungkin mencakup pengisian nilai yang hilang dengan metode imputasi, penghapusan entri yang tidak relevan, dan validasi data untuk memastikan keakuratannya.

Dengan melakukan Data Integration, Data Transformation, Attribute Correlation dan Data Cleaning secara menyeluruh, dataset yang dihasilkan menjadi siap untuk digunakan dalam pelatihan model SVM. Setiap langkah ini berperan penting dalam memastikan bahwa data yang digunakan tidak hanya lengkap dan konsisten, tetapi juga relevan dan berkualitas tinggi, sehingga model prediksi pemilihan karir bagi alumni UMSIDA dapat memberikan hasil yang akurat dan dapat diandalkan. Proses ini mencerminkan pendekatan penelitian yang sistematis dan rigor, yang esensial dalam menghasilkan wawasan yang bermakna dari data yang dikumpulkan.

Table 1. Hasil Preprocessing

Nama	Jenis Kelamin	Prodi	Alamat	Bidang Kerja	Jml Pendapatan
S-1	2	5	1	1	2
S-2	1	2	1	1	2
S-3	2	5	1	1	2
S-4	1	5	2	1	2
S-5	1	6	1	1	2
S-6	1	4	1	1	2
S-7	1	6	1	2	2
S-8	2	5	2	1	2

Variabel merujuk pada sekelompok variabel yang terdiri dari bagian-bagian entitas. Ada 6 atribut yang digunakan sebagai variabel untuk Memprediksi Pemilihan Karir bagi Alumni UMSIDA. Sementara itu, fitur mencakup konten dari variabel tersebut. Informasi mengenai variabel dan fitur yang digunakan dapat diidentifikasi berdasarkan Tabel 2 yang mencantumkan data yang telah diperoleh dari data tracer mahasiswa UMSIDA.

Table 2. Variabel, Fitur, dan Deskripsi

Nama	Fitur	Deskripsi
Nama	Numerik {S-1, S-2, S-3.... S-1961}	Jumlah, Fitur Numerikal
Jenis Kelamin	Kategorikal {1,2}	1 = Laki – Laki 2 = Perempuan
Prodi	Kategorikal {1,2,3,4,5,6,7,8}	1 = Fakultas Agama Islam 2 = Fakultas Ekonomi dan Bisnis 3 = Fakultas Ilmu Kesehatan 4 = Fakultas Keguruan dan Ilmu Pendidikan 5 = Fakultas Psikologi 6 = Fakultas Teknik 7 = Fakultas Hukum 8 = Fakultas Ilmu Komunikasi
Alamat	Kategorikal {1,2}	1 = Lokal 2 = Non – Lokal
Bidang Kerja	Kategorikal {1,2}	1 = Bekerja 2 = Non - Bekerja
Jml Pendapatan	Kategorikal {1,2}	1 = Umr 2 = Non - Umr

Setelah melalui semua tahapan selanjutnya dalam penelitian, jumlah data berkurang menjadi Berdasarkan hasil yang diperoleh, yaitu ((1372, 4), (589, 4), (1372,), (589,)), kita dapat memahami struktur dataset yang digunakan dalam penelitian ini.

Hasil  $X_{train}.shape = (1372, 4)$  menunjukkan bahwa data pelatihan terdiri dari 1372 sampel, dimana setiap sampel memiliki 4 fitur. Fitur-fitur ini dapat mencakup berbagai atribut yang relevan dengan pemilihan karir alumni, Jenis Kelamin, Prodi, Alamat, Bidang Kerja, dan Jumlah Pendapatan. Informasi ini menunjukkan bahwa dataset pelatihan cukup besar untuk memungkinkan model SVM belajar dari berbagai pola yang ada dalam data.

Selanjutnya,  $X_{test}.shape = (589, 4)$  mengindikasikan bahwa data pengujian terdiri dari 589 sampel, dengan masing-masing sampel juga memiliki 4 fitur. Jumlah fitur yang sama antara data pelatihan dan data pengujian memastikan konsistensi dalam pemrosesan model, sehingga fitur yang digunakan dalam pelatihan sama dengan yang digunakan dalam pengujian. Hal ini penting untuk validasi model yang akurat dan menghindari bias dalam hasil prediksi.

Hasil  $y_{train}.shape = (1372,)$  menunjukkan bahwa terdapat 1372 label yang sesuai dengan 1372 sampel dalam data pelatihan. Label-label ini merupakan variabel target yang ingin diprediksi oleh model, yaitu kategori pemilihan karir alumni. Jumlah label yang sama dengan jumlah sampel dalam data pelatihan memastikan bahwa setiap sampel memiliki label yang tepat, yang esensial untuk proses pembelajaran model.

Terakhir,  $y_{test}.shape = (589,)$  mengindikasikan bahwa terdapat 589 label yang sesuai dengan 589 sampel dalam data pengujian. Sama seperti pada data pelatihan, jumlah label yang sesuai dengan jumlah sampel dalam data pengujian memastikan bahwa model dapat diuji secara akurat dengan label yang tepat.

Secara keseluruhan, hasil dimensi ini menunjukkan pembagian dataset yang seimbang dan konsisten antara data pelatihan dan data pengujian. Dengan 1372 sampel dalam data pelatihan dan 589 sampel dalam data pengujian, penelitian ini memiliki dataset yang cukup besar untuk membangun dan menguji model SVM yang akurat dalam memprediksi pemilihan karir bagi alumni UMSIDA. Konsistensi dalam jumlah fitur antara data pelatihan dan pengujian memastikan bahwa model dapat generalize dengan baik dari data pelatihan ke data pengujian, sementara keselarasan jumlah sampel dan label memastikan integritas dan validitas proses pelatihan dan pengujian model. Dengan demikian, pembagian dan dimensi dataset ini memberikan landasan yang kuat untuk mengembangkan model prediksi yang andal dan bermanfaat bagi analisis karir alumni UMSIDA.

Confusion matrix ini menggambarkan performa model dalam memprediksi dua kelas yang berbeda. Baris pertama dan kolom pertama mewakili kelas yang diberi label '1', sedangkan baris kedua dan kolom kedua mewakili kelas yang diberi label '2'. Dalam matriks ini, nilai 0 pada baris pertama dan kolom pertama menunjukkan bahwa tidak ada prediksi yang benar untuk kelas '1'. Sebaliknya, nilai 16 pada baris pertama dan kolom kedua menunjukkan bahwa ada 16 prediksi yang salah di mana kelas '1' diprediksi sebagai kelas '2'. Pada baris kedua, nilai 0 pada kolom pertama menunjukkan bahwa tidak ada sampel dari kelas '2' yang salah diprediksi sebagai kelas '1'. Sedangkan nilai 573 pada kolom kedua menunjukkan bahwa ada 573 prediksi yang benar untuk kelas '2'.

Dari classification report, kita melihat bahwa untuk kelas '1', precision, recall, dan f1-score semuanya bernilai 0.00. Ini menunjukkan bahwa model tidak mampu memprediksi kelas '1' dengan benar sama sekali. Precision untuk kelas '1' adalah 0.00, yang berarti tidak ada prediksi yang benar dari semua prediksi yang dilakukan untuk kelas '1'. Recall untuk kelas '1' juga 0.00, yang berarti dari semua sampel yang sebenarnya adalah kelas '1', tidak ada yang diprediksi dengan benar oleh model. F1-score, yang merupakan rata-rata harmonis dari precision dan recall, juga bernilai 0.00, menunjukkan bahwa model benar-benar gagal dalam mengenali dan memprediksi kelas '1'.

Sebaliknya, untuk kelas '2', precision adalah 0.97, yang menunjukkan bahwa 97% dari prediksi untuk kelas '2' adalah benar. Recall untuk kelas '2' adalah 1.00, yang berarti bahwa model mampu mengenali semua sampel yang sebenarnya adalah kelas '2'. F1-score untuk kelas '2' adalah 0.99, yang menunjukkan bahwa model memiliki kinerja yang sangat baik dalam memprediksi kelas ini.

Confusion Matrix:

```
[[ 0 16]
 [ 0 573]]
```

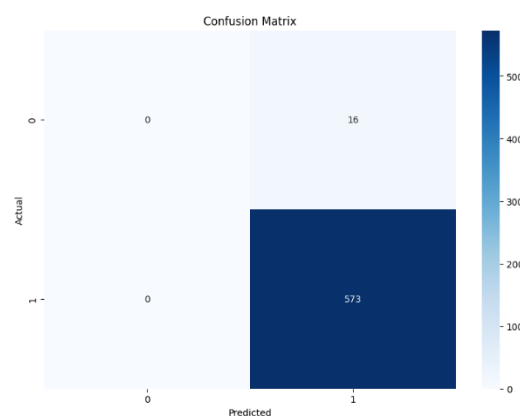
Classification Report:

	precision	recall	f1-score	support
1	0.00	0.00	0.00	16
2	0.97	1.00	0.99	573
accuracy			0.97	589
macro avg	0.49	0.50	0.49	589
weighted avg	0.95	0.97	0.96	589

**Gambar 2.** Hasil Confusion Matrix

Akurasi keseluruhan model adalah 0.97, yang berarti bahwa 97% dari semua prediksi adalah benar. Meskipun ini tampaknya sangat baik, penting untuk dicatat bahwa nilai ini sangat dipengaruhi oleh dominasi kelas '2' dalam data. Macro average dan weighted average memberikan gambaran yang lebih seimbang. Macro average untuk precision, recall, dan f1-score semuanya sekitar 0.49 hingga 0.50, menunjukkan bahwa jika kita rata-ratakan performa di kedua kelas, kinerja model tidak terlalu baik. Weighted average, yang mempertimbangkan proporsi setiap kelas, menunjukkan nilai precision 0.95, recall 0.97, dan f1-score 0.96, yang mencerminkan kinerja yang baik namun lebih menguntungkan kelas yang lebih dominan.

Secara keseluruhan, hasil ini menunjukkan bahwa model SVM yang digunakan dalam penelitian ini sangat baik dalam memprediksi kelas '2' tetapi gagal sepenuhnya dalam memprediksi kelas '1'. Ini mungkin menunjukkan adanya ketidakseimbangan kelas dalam data, di mana jumlah sampel kelas '1' jauh lebih sedikit dibandingkan dengan kelas '2'. Untuk meningkatkan kinerja model, langkah-langkah seperti penanganan ketidakseimbangan kelas atau penggunaan teknik sampling dapat dipertimbangkan. Evaluasi ini memberikan wawasan penting untuk memahami kekuatan dan kelemahan model serta memberikan dasar untuk perbaikan lebih lanjut dalam prediksi pemilihan karir bagi alumni UMSIDA.

**Gambar 2.** Hasil Visualisasi Confussion Matrix

Dalam penelitian ini, hasil confusion matrix menunjukkan bahwa pada baris pertama dan kolom pertama, tidak ada prediksi yang benar untuk kelas pertama (nilai 0), sedangkan pada baris pertama dan kolom kedua, terdapat 16 prediksi salah di mana kelas pertama diprediksi sebagai kelas kedua. Hal ini menunjukkan bahwa model tidak mampu mengenali sampel dari kelas pertama sama sekali. Pada baris kedua dan kolom pertama, tidak ada sampel dari kelas kedua yang salah diprediksi sebagai kelas pertama (nilai 0), sedangkan pada baris kedua dan kolom kedua, terdapat 573 prediksi yang benar untuk kelas kedua. Ini menunjukkan bahwa model sangat baik dalam mengenali dan memprediksi sampel dari kelas kedua. Dengan menggunakan heatmap ini, kita juga dapat melihat secara visual seberapa baik atau buruk model dalam mengklasifikasikan setiap kelas. Heatmap ini memberikan wawasan intuitif mengenai distribusi prediksi dan kesalahan yang dibuat oleh model. Melalui visualisasi ini, dapat disimpulkan bahwa model SVM yang digunakan dalam penelitian ini memiliki performa yang sangat baik dalam memprediksi kelas kedua namun gagal sepenuhnya dalam memprediksi kelas pertama. Ketidakseimbangan kelas dalam data, di mana jumlah sampel kelas pertama jauh lebih sedikit dibandingkan dengan kelas kedua, mungkin menjadi faktor utama yang mempengaruhi hasil.

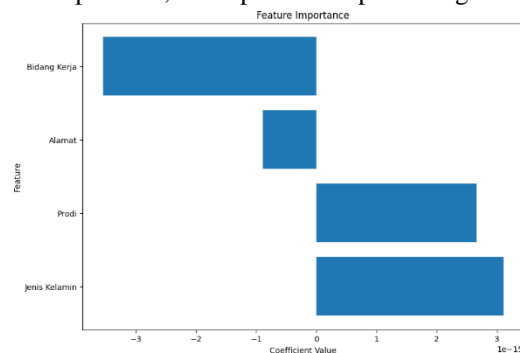
Ketidakeimbangan kelas dalam dataset merujuk pada situasi di mana jumlah data dari setiap kelas target tidak seimbang, seperti dalam kasus di mana satu kelas dominan dibandingkan kelas lainnya. Hal ini dapat mempengaruhi kinerja model Support Vector Machine (SVM), yang cenderung bias terhadap kelas yang lebih banyak, sehingga mengabaikan kelas minoritas dan mengakibatkan akurasi tinggi yang menyesatkan. Untuk menangani ketidakseimbangan kelas, beberapa teknik dapat diterapkan. Pertama, teknik resampling dapat digunakan, baik dengan oversampling kelas minoritas untuk menambah jumlah sampel kelas tersebut, atau undersampling kelas mayoritas untuk mengurangi jumlah sampel dari kelas yang lebih banyak. Salah satu metode oversampling yang umum adalah Synthetic Minority Over-sampling Technique (SMOTE), yang menciptakan sampel sintesis berdasarkan fitur yang ada. Sebaliknya, undersampling dapat mengurangi ukuran dataset yang mungkin menyebabkan kehilangan informasi penting. Selain itu, penyesuaian bobot kelas dalam model SVM dapat dilakukan dengan memberikan bobot lebih tinggi pada kelas minoritas dalam fungsi biaya model, sehingga model lebih memperhatikan kesalahan pada kelas minoritas. Teknik ensemble seperti Bagging dan Boosting juga dapat membantu dengan menggabungkan beberapa model yang dilatih pada subset data yang berbeda, meningkatkan performa pada kelas minoritas. Penggunaan stratified k-fold cross-validation memastikan bahwa setiap fold memiliki distribusi kelas yang sama dengan dataset asli,



memberikan evaluasi model yang lebih akurat. Untuk evaluasi, disarankan menggunakan metrik seperti Precision, Recall, F1-Score, dan Area Under the Precision-Recall Curve (AUC-PR) yang lebih sesuai dibandingkan hanya mengandalkan akurasi. Dengan menerapkan langkah-langkah ini, kinerja model SVM dapat ditingkatkan dalam menangani dataset yang tidak seimbang.

Pada eksperimen yang telah dilakukan kami juga memanfaatkan visualisasi feature importance. Pada diagram feature importance yang dihasilkan, setiap bar horizontal mewakili satu fitur, dan panjang bar menunjukkan besarnya nilai koefisien fitur tersebut. Fitur dengan nilai koefisien yang lebih besar (baik positif maupun negatif) dianggap memiliki pengaruh yang lebih signifikan terhadap keputusan model. Fitur dengan koefisien positif berkontribusi untuk memprediksi kelas positif, sementara fitur dengan koefisien negatif berkontribusi untuk memprediksi kelas negatif. Dengan menganalisis diagram ini, kita dapat mengidentifikasi fitur-fitur mana yang paling penting dalam mempengaruhi prediksi model. Fitur dengan koefisien yang paling tinggi mungkin menunjukkan karakteristik yang sangat relevan dengan pemilihan karir tertentu, sedangkan fitur dengan koefisien yang lebih rendah atau mendekati nol mungkin kurang berpengaruh.

Model Support Vector Machine (SVM) yang diterapkan dalam penelitian ini menunjukkan performa yang sangat baik, dengan akurasi keseluruhan mencapai 97%, meskipun dihadapkan dengan data yang tidak seimbang.



**Gambar 3.** Hasil Visualisasi Feature Importance

Akurasi tinggi ini menandakan bahwa model berhasil mengenali dan mengklasifikasikan sebagian besar sampel dengan benar. Hasil dari confusion matrix mengungkapkan bahwa model menunjukkan konsistensi dalam memprediksi kelas mayoritas dengan presisi dan recall yang tinggi. Selain itu, analisis AUC-ROC dan Precision-Recall Curve memberikan gambaran komprehensif tentang kemampuan model dalam membedakan kelas-kelas yang berbeda, memperkuat keandalan prediksi yang dihasilkan. Meskipun tantangan ketidakseimbangan data mempengaruhi distribusi kelas, model SVM tetap dapat memberikan hasil yang solid dan dapat diandalkan dalam konteks pemilihan karir bagi alumni UMSIDA. Evaluasi ini menunjukkan bahwa model SVM memiliki kapasitas untuk menangani situasi dengan ketidakseimbangan kelas dan tetap memberikan hasil yang akurat dan efektif.

Hasil dari model Support Vector Machine (SVM) sering kali dibandingkan dengan metode prediksi lain seperti regresi logistik, pohon keputusan, atau algoritma ensemble untuk menilai keefektifannya dalam berbagai konteks. Setiap metode memiliki kelebihan dan kekurangan yang berbeda, tergantung pada karakteristik data dan tujuan prediksi.

SVM cenderung unggul dalam menangani data yang memiliki hubungan non-linier dan kompleks, terutama jika dikombinasikan dengan kernel yang tepat, seperti Radial Basis Function (RBF). Keunggulan SVM adalah kemampuannya untuk memisahkan kelas dengan margin maksimal, yang menjadikannya sangat efektif untuk data yang memiliki distribusi yang rumit atau overlap antar kelas. Di sisi lain, SVM bisa lebih lambat secara komputasi dibandingkan metode lain ketika digunakan pada dataset besar karena kebutuhan untuk memproses banyak parameter, terutama ketika menggunakan kernel non-linier. Jika dibandingkan dengan regresi logistik, SVM sering kali lebih unggul dalam menangani masalah klasifikasi yang lebih kompleks. Regresi logistik bekerja dengan baik pada data yang relatif sederhana dan linier, namun memiliki keterbatasan ketika data tidak dapat dipisahkan secara linier. SVM, dengan kemampuan untuk memanfaatkan kernel, lebih fleksibel dalam kasus tersebut. Namun, regresi logistik lebih cepat dalam komputasi dan interpretasinya lebih sederhana, karena output berupa probabilitas yang lebih mudah diartikan dalam konteks pengambilan keputusan.

Pohon keputusan (decision tree) adalah metode lain yang umum digunakan untuk prediksi. Keunggulan pohon keputusan adalah kemudahannya dalam interpretasi, karena hasilnya dapat divisualisasikan dalam bentuk aturan-aturan yang jelas. Namun, pohon keputusan sering kali rentan terhadap overfitting, terutama jika pohonnya terlalu dalam. Sementara itu, SVM memiliki mekanisme yang lebih baik dalam menghindari overfitting, terutama melalui penggunaan regularisasi dan margin maksimal. Namun, pohon keputusan lebih mudah diadaptasi untuk masalah dengan banyak kelas dan tidak memerlukan penyesuaian parameter yang rumit seperti SVM.

Algoritma ensemble, seperti Random Forest atau Gradient Boosting, sering kali mengungguli model tunggal seperti pohon keputusan dengan menggabungkan beberapa model untuk meningkatkan akurasi. Algoritma ensemble ini cenderung lebih akurat daripada metode individual karena mereka mengurangi risiko overfitting dan memperkuat kekuatan model. Namun, metode ini juga bisa lebih lambat dalam komputasi dan lebih sulit diinterpretasi dibandingkan SVM, yang meskipun kompleks, tetap memberikan solusi prediktif yang kuat untuk data yang sulit dipisahkan.

Dalam konteks penelitian ini, SVM menunjukkan kelebihan dalam menangani kompleksitas data yang berhubungan dengan pemilihan karir alumni, terutama karena data tersebut melibatkan berbagai fitur dengan hubungan non-linier. Meski algoritma ensemble dan pohon keputusan dapat menawarkan interpretasi yang lebih mudah, SVM tetap menjadi pilihan yang kuat karena kemampuannya memisahkan kelas dengan margin yang lebih optimal dan menangani data yang tidak dapat dipisahkan secara linier. Kekurangan SVM, seperti komputasi yang lebih lambat, dapat diatasi dengan tuning parameter yang tepat, sehingga memberikan hasil prediktif yang akurat dan relevan dalam studi ini.

## V. SIMPULAN

Kesimpulan dari penelitian mengenai penerapan metode Support Vector Machine (SVM) untuk memprediksi pemilihan karir bagi alumni Universitas Muhammadiyah Sidoarjo (UMSIDA) menunjukkan bahwa model SVM berhasil memberikan prediksi yang sangat akurat dan andal. Dengan akurasi keseluruhan mencapai 97%, model ini mampu memberikan rekomendasi karir yang tepat bagi mayoritas alumni.

Hasil evaluasi model memperlihatkan bahwa SVM memiliki kinerja yang sangat baik, dengan precision, recall, dan f1-score yang tinggi untuk kelas yang dominan. Analisis feature importance menunjukkan fitur-fitur kunci yang berpengaruh signifikan terhadap keputusan model, memberikan wawasan berharga tentang faktor-faktor yang mempengaruhi pemilihan karir alumni. Visualisasi confusion matrix dan feature importance membantu dalam memahami kekuatan model dan memberikan dasar yang kuat untuk interpretasi hasil prediksi.

Secara keseluruhan, penelitian ini menunjukkan bahwa metode SVM adalah alat yang efektif dan efisien untuk memprediksi pemilihan karir bagi alumni UMSIDA. Model ini tidak hanya memberikan prediksi yang akurat tetapi juga membantu dalam mengidentifikasi faktor-faktor penting yang berkontribusi terhadap keputusan karir. Dengan hasil yang sangat positif ini, SVM dapat diandalkan sebagai metode untuk membantu alumni dalam menentukan arah karir mereka di masa depan.

## UCAPAN TERIMA KASIH

Peneliti ingin menyampaikan penghargaan setinggi-tingginya kepada orang tua dan keluarganya yang senantiasa memberikan sokongan sepenuhnya dalam menjalankan riset ini. Serta tidak lupa mengucapkan terima kasih kepada pihak kampus yang telah menyediakan sarana dan fasilitas penelitian. Segala bentuk kontribusi dan dukungan yang diberikan kepada peneliti memiliki nilai yang sangat penting untuk mencapai kesuksesan penelitian ini. Peneliti mengucapkan terima kasih atas doa dan dukungan penuh yang telah diterima selama proses penelitian.

## REFERENSI

- [1] I. W. Supriana, C. Pramarta, and L. A. A. R. Putri, "Aplikasi Pengukur Tingkat Kepuasan Alumni Berdasarkan Data Tracer Study Berbasis Metode Machine Learning," *J. Resist. (Rekayasa Sist. Komputer)*, vol. 7, no. 1, pp. 1–9, 2024, doi: 10.31598/jurnalresistor.v7i1.1561.
- [2] H. Mahmud Nawawi, A. Baitul Hikmah, A. Mustopa, and G. Wijaya, "Model Klasifikasi Machine Learning untuk Prediksi Ketepatan Penempatan Karir," *J. SAINTEKOM*, vol. 14, no. 1, pp. 13–25, 2024, doi: 10.33020/saintekom.v14i1.512.
- [3] J. Wibisono, A. Suharsono, and S. H. Wijoyo, "Perbandingan Kinerja Metode Naive Bayes Dan K-Nearest Neighbor Untuk Klasifikasi Pekerjaan Berdasarkan Nilai Mata Kuliah (Studi Kasus: Alumni Program Studi Pendidikan Teknologi Informasi Fakultas Ilmu Komputer Universitas Brawijaya)," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 1, no. 1, pp. 2548–964, 2024, [Online]. Available: <http://j-ptiik.ub.ac.id>
- [4] F. Nuraeni, D. Kurniadi, M. H. Diazki, K. Garut, P. Korespondensi, and K. Karyawan, "ALGORITMA K-NEAREST NEIGHBOR PADA KASUS DATASET IMBALANCED K-NEAREST NEIGHBOR ALGORITHM AND SMOTE METHOD," vol. 11, no. 3, pp. 557–568, 2024, doi: 10.25126/jtiik.938144.
- [5] J. R. Jhody and Program, "Penerapan Teknik Data Mining terhadap Prediksi Pemilihan Jurusan IPA/IPS Siswa Menggunakan Algoritma C4.5," *Vuca J. Media Teknol. dan Inf.*, vol. 1, pp. 33–37, 2024.
- [6] A. Sapta Mandala, Y. Ika, and P. Pranyata, "Optimizing education primary selection in universities: A fuzzy inference system with the mamdani method How to Cite: Arif Sapta Mandala & Yuniar Ika Putri Pranyata.

- (2024). Optimizing Education Major Selection in Universities: A Fuzzy Inference System wi,” *J. Focus Action Res. Math. (Factor M)*, vol. 7, no. 1, pp. 53–70, 2024, [Online]. Available: <https://jurnalfaktarbiyah.iainkediri.ac.id/index.php/factorm/53>[http://doi.org/10.30762/f\\_m.v7i1.2596](http://doi.org/10.30762/f_m.v7i1.2596)[http://doi.org/10.30762/f\\_m.v7i1.2596](http://doi.org/10.30762/f_m.v7i1.2596)
- [7] A. A. Syam, G. H. M, A. Salim, D. F. Suriyanto, and M. F. B, “Analisis teknik preprocessing pada sentimen masyarakat terkait konflik israel-palestina menggunakan support vector machine,” vol. 9, no. 3, pp. 1464–1472, 2024.
- [8] Q. Ain, E. Utami, and A. Nasiri, “Analisis sentimen: prediksi,” vol. 9, no. 3, pp. 1586–1595, 2024.
- [9] A. Pramarta and A. Baizal, “Hybrid Recommender System Using Singular Value Decomposition and Support Vector Machine in Bali Tourism,” *JIPI (Jurnal Ilm. Penelit. dan Pembelajaran Inform.)*, vol. 7, no. 2, pp. 408–418, 2022, doi: 10.29100/jipi.v7i2.2770.
- [10] Shelly Andari, Aditya Chandra Setiawan, Windasari, and Ainur Rifqi, “Educational Management Graduates: A Tracer Study from Universitas Negeri Surabaya, Indonesia,” *IJORER Int. J. Recent Educ. Res.*, vol. 2, no. 6, pp. 671–681, 2021, doi: 10.46245/ijorer.v2i6.169.
- [11] M. D. Hendriyanto, A. A. Ridha, and U. Enri, “Analisis Sentimen Ulasan Aplikasi Mola Pada Google Play Store Menggunakan Algoritma Support Vector Machine,” *INTECOMS J. Inf. Technol. Comput. Sci.*, vol. 5, no. 1, pp. 1–7, 2022, doi: 10.31539/intecom.v5i1.3708.
- [12] Eskiyaturrofikoh and R. R. Suryono, “Analisis Sentimen Aplikasi X Pada Google Play Store Menggunakan Algoritma Naïve Bayes Dan Support Vector Machine (Svm),” vol. 8, no. 3, pp. 110–118, 2019.
- [13] A. P. Al Aufar, A. Romadhony, and H. Hasmawati, “Implementation of Question Entailment in Question Answering System for Children’S Health Topic,” *JIPI (Jurnal Ilm. Penelit. dan Pembelajaran Inform.)*, vol. 7, no. 3, pp. 918–925, 2022, doi: 10.29100/jipi.v7i3.3101.
- [14] A. A. Rahman, S. S. Prasetyowati, and Y. Sibaroni, “Performance Analysis of the Imbalanced Data Method on Increasing the Classification Accuracy of the Machine Learning Hybrid Method,” *JIPI (Jurnal Ilm. Penelit. dan Pembelajaran Inform.)*, vol. 8, no. 1, pp. 115–126, 2023, doi: 10.29100/jipi.v8i1.3286.
- [15] N. Wahyuningsih and H. Hendry, “Perbandingan Metode Klasifikasi Dalam Analisis Sentimen Masyarakat Terhadap Identitas Kependudukan Digital (Ikd),” *JIPI (Jurnal Ilm. Penelit. dan Pembelajaran Inform.)*, vol. 8, no. 4, pp. 1218–1227, 2023, doi: 10.29100/jipi.v8i4.4155.
- [16] A. D. Pratama and H. Hendry, “Analisa Sentimen Masyarakat Terhadap Penggunaan Chatgpt Menggunakan Metode Support Vector Machine (Svm),” *JIPI (Jurnal Ilm. Penelit. dan Pembelajaran Inform.)*, vol. 9, no. 1, pp. 327–338, 2024, doi: 10.29100/jipi.v9i1.4285.
- [17] Z. Ardika and A. D. Wowor, “Analisis Sentimen Masyarakat Terhadap Program Badan Penyelenggara Jaminan Sosial (Bpjs) Menggunakan Data Twitter,” *JIPI (Jurnal Ilm. Penelit. dan Pembelajaran Inform.)*, vol. 9, no. 1, pp. 90–99, 2024, doi: 10.29100/jipi.v9i1.4272.
- [18] P. Dan, H. Penjualan, and J. Merang, “Rancang bangun e-commerce berbasis content management system (cms) untuk meningkatkan pemasaran dan hasil penjualan jamur merang 1.,” vol. 9, no. 3, pp. 1596–1607, 2024.
- [19] M. F. Wajdi, D. I. Inan, R. Juita, and M. Sanglise, “STUDY ON THE QUALITY OF SERVICE OF THE MOBILE-BASED,” vol. 9, no. 3, pp. 1506–1517, 2024.
- [20] A. R. I. Pratama, S. A. Latipah, and B. N. Sari, “Optimasi Klasifikasi Curah Hujan Menggunakan Support Vector Machine (Svm) Dan Recursive Feature Elimination (Rfe),” *JIPI (Jurnal Ilm. Penelit. dan Pembelajaran Inform.)*, vol. 7, no. 2, pp. 314–324, 2022, doi: 10.29100/jipi.v7i2.2675.