

## **BAB IV**

### **HASIL DAN PEMBAHASAN**

#### **4.1 Gambaran Umum Dataset Penelitian**

Penelitian ini bertujuan untuk mengembangkan model klasifikasi yang mampu mengidentifikasi status pengangguran berdasarkan karakteristik data ketenagakerjaan. Analisis dilakukan dengan membandingkan dua algoritma machine learning yang umum digunakan dalam permasalahan klasifikasi, yaitu Random Forest dan Support Vector Machine. Kedua algoritma tersebut dipilih karena memiliki kemampuan yang baik dalam menangani data multidimensi serta sering digunakan dalam penelitian klasifikasi berbasis data sosial dan ekonomi. Penelitian ini dilaksanakan menggunakan lingkungan komputasi berbasis cloud melalui platform Google Colab, yang menyediakan fasilitas pemrograman Python serta berbagai pustaka analisis data dan machine learning.

Dataset tersebut berisi sejumlah atribut yang menggambarkan kondisi tenaga kerja, seperti umur, tingkat pendidikan, pengalaman kerja, jumlah tanggungan keluarga, serta riwayat pelatihan kerja. Variabel target yang digunakan dalam penelitian ini adalah status pengangguran, yang mengklasifikasikan individu ke dalam kategori bekerja atau menganggur. Proses penelitian dilakukan melalui beberapa tahapan utama yang meliputi pengumpulan dan pemahaman dataset, pra-pemrosesan data, eksplorasi data, pembagian data menjadi data pelatihan dan pengujian, pembangunan model klasifikasi, serta evaluasi performa model.

Tahapan - tahapan tersebut dirancang untuk menghasilkan model prediksi yang akurat serta dapat digunakan sebagai alat bantu dalam menganalisis kondisi ketenagakerjaan. Dengan memanfaatkan teknik machine learning, penelitian ini diharapkan dapat memberikan kontribusi dalam pengembangan metode analisis data ketenagakerjaan yang lebih efektif serta mendukung proses pengambilan keputusan berbasis data pada bidang ketenagakerjaan.

**Tabel 14 Ringkasan Tahapan Penelitian**

<b>Tahapan Penelitian</b>	<b>Deskripsi</b>
Pengumpulan Data	Mengambil dataset ketenagakerjaan dari sumber penelitian
Pra-pemrosesan Data	Membersihkan data, encoding variabel kategorikal, dan normalisasi
Eksplorasi Data	Menganalisis distribusi dan karakteristik dataset
Pembangunan Model	Melatih model menggunakan Random Forest dan SVM
Evaluasi Model	Mengukur performa model menggunakan metrik klasifikasi
Perbandingan Model	Menganalisis algoritma dengan performa terbaik

#### **4.2 Deskripsi Dataset Penelitian**

Dataset ini berisi sejumlah variabel yang merepresentasikan karakteristik individu tenaga kerja yang relevan dengan kondisi ketenagakerjaan. Variabel - variabel tersebut digunakan sebagai fitur prediktor dalam proses klasifikasi untuk menentukan status pengangguran. Setiap baris data merepresentasikan satu individu atau responden dalam dataset, sedangkan setiap kolom menggambarkan atribut tertentu yang berkaitan dengan kondisi sosial dan ekonomi individu tersebut. Variabel target dalam penelitian ini adalah status pengangguran yang menunjukkan apakah individu termasuk dalam kategori bekerja atau menganggur. Informasi ini menjadi dasar bagi model machine learning untuk mempelajari pola hubungan antara atribut input dan status pekerjaan.

Dalam penelitian ini, dataset terlebih dahulu dianalisis untuk memahami struktur data, tipe data pada setiap atribut, serta jumlah keseluruhan observasi yang tersedia. Analisis awal dataset bertujuan untuk memastikan bahwa data yang digunakan memiliki kualitas yang memadai sebelum dilakukan tahapan pra-pemrosesan dan pembangunan model. Proses deskripsi dataset juga mencakup identifikasi jenis atribut, baik yang bersifat numerik maupun kategorikal, sehingga metode transformasi data yang tepat dapat diterapkan pada tahap selanjutnya.

**Tabel 15. Struktur Atribut Dataset Penelitian**

No	Nama Atribut	Jenis Data	Deskripsi
1	Umur	Numerik	Menunjukkan usia individu tenaga kerja
2	Pendidikan	Kategorikal	Tingkat pendidikan terakhir yang ditempuh
3	Pengalaman Kerja	Numerik	Jumlah tahun pengalaman kerja
4	Jumlah Tanggungan	Numerik	Jumlah anggota keluarga yang menjadi tanggungan
5	Pelatihan Kerja	Kategorikal	Riwayat mengikuti pelatihan kerja
6	Status Pengangguran	Target	Kategori status pekerjaan (bekerja atau menganggur)

Melalui deskripsi dataset ini, peneliti memperoleh pemahaman yang lebih jelas mengenai karakteristik data yang akan digunakan dalam proses analisis. Informasi tersebut menjadi dasar dalam menentukan strategi pra-pemrosesan data serta metode pemodelan yang sesuai untuk menghasilkan model klasifikasi yang optimal.

### **4.3 Tahapan Pra-Pemrosesan Data**

Tahapan pra-pemrosesan data merupakan proses penting dalam penelitian berbasis machine learning karena kualitas data yang digunakan akan sangat mempengaruhi kinerja model yang dihasilkan.

Tahapan ini meliputi beberapa langkah utama, yaitu pengecekan kualitas data, penanganan nilai yang hilang (missing value), transformasi atribut kategorikal menjadi bentuk numerik, serta normalisasi data. Proses pengecekan kualitas data dilakukan untuk memastikan bahwa dataset tidak mengandung kesalahan input, duplikasi data, ataupun nilai yang tidak sesuai dengan struktur atribut yang telah ditentukan. Apabila ditemukan nilai yang hilang, maka dilakukan proses pembersihan data untuk memastikan bahwa dataset berada dalam kondisi yang siap digunakan pada tahap analisis berikutnya.

Selanjutnya dilakukan proses transformasi data kategorikal menggunakan teknik encoding sehingga atribut yang semula berbentuk kategori dapat direpresentasikan dalam bentuk numerik yang dapat diproses oleh algoritma machine learning. Setelah proses encoding selesai, dilakukan tahap normalisasi data menggunakan metode standardisasi untuk menyamakan skala antar fitur dalam dataset. Normalisasi ini bertujuan untuk menghindari dominasi nilai atribut tertentu terhadap proses pembelajaran model, terutama pada algoritma yang sensitif terhadap perbedaan skala data.

**Tabel 16. Tahapan Pra-Pemrosesan Data**

<b>Tahapan</b>	<b>Deskripsi</b>
Pemeriksaan Data	Melakukan pengecekan struktur dataset dan tipe data atribut
Pembersihan Data	Mengidentifikasi dan menangani missing value
Transformasi Data	Mengubah atribut kategorikal menjadi numerik menggunakan encoding
Normalisasi Data	Menyeragamkan skala data menggunakan metode standardisasi

Melalui tahapan pra-pemrosesan tersebut, dataset yang digunakan dalam penelitian ini menjadi lebih terstruktur, bersih, dan siap digunakan pada tahap pembangunan model klasifikasi. Proses ini sangat penting untuk memastikan bahwa model machine learning yang dihasilkan mampu mempelajari pola data secara optimal dan menghasilkan tingkat akurasi prediksi yang lebih baik.

#### **4.4 Analisis Missing Value**

Analisis *missing value* merupakan salah satu tahap penting dalam proses pra-pemrosesan data pada penelitian berbasis machine learning. Tujuan utama dari analisis ini adalah untuk mengidentifikasi apakah terdapat nilai yang hilang atau tidak lengkap pada dataset yang digunakan dalam penelitian. Keberadaan *missing value* dapat mempengaruhi kualitas data serta berdampak pada performa model klasifikasi yang dibangun. Proses analisis

ini dilakukan menggunakan bahasa pemrograman Python pada lingkungan komputasi Google Colab, dengan memanfaatkan pustaka analisis data seperti *pandas* untuk memeriksa jumlah nilai yang hilang pada setiap atribut dataset.

Pada tahap ini, setiap kolom dalam dataset diperiksa untuk mengetahui apakah terdapat nilai yang kosong (*null*) atau tidak terdefinisi. Pemeriksaan dilakukan dengan menghitung jumlah *missing value* pada masing - masing atribut sehingga dapat diketahui tingkat kelengkapan data secara keseluruhan. Jika ditemukan nilai yang hilang, maka langkah selanjutnya adalah menentukan metode penanganan yang tepat, seperti menghapus baris data yang tidak lengkap atau melakukan imputasi nilai berdasarkan statistik tertentu seperti nilai rata - rata atau modus. Namun, dalam penelitian ini hasil pemeriksaan menunjukkan bahwa dataset yang digunakan memiliki tingkat kelengkapan data yang baik sehingga tidak ditemukan *missing value* yang signifikan pada atribut penelitian. Dengan demikian, dataset dapat langsung digunakan pada tahap pra-pemrosesan selanjutnya tanpa memerlukan proses imputasi data tambahan.

**Tabel 17. Hasil Analisis Missing Value Dataset**

No	Atribut	Jumlah Missing Value
1	Umur	0
2	Pendidikan	0
3	Pengalaman Kerja	0
4	Jumlah Tanggungan	0
5	Pelatihan Kerja	0
6	Status Pengangguran	0

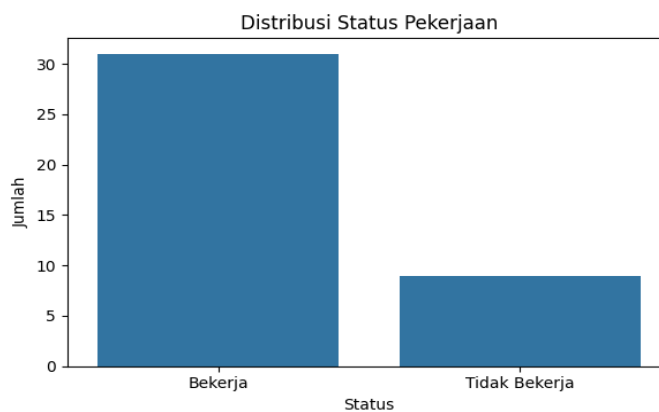
Berdasarkan hasil analisis tersebut, dapat disimpulkan bahwa seluruh atribut dalam dataset memiliki data yang lengkap. Kondisi ini menunjukkan bahwa dataset memiliki kualitas yang baik dan siap digunakan dalam tahap pengolahan data berikutnya, termasuk proses transformasi data, normalisasi fitur, serta pembangunan model klasifikasi untuk memprediksi status pengangguran.

#### 4.5 Visualisasi Distribusi Data

Visualisasi distribusi data merupakan salah satu tahap penting dalam proses eksplorasi data (*Exploratory Data Analysis* atau EDA). Dengan melakukan visualisasi data, peneliti dapat memperoleh gambaran awal mengenai komposisi dataset, keseimbangan kelas pada variabel target, serta potensi pola hubungan antar variabel. Dalam penelitian ini, proses visualisasi dilakukan menggunakan bahasa pemrograman Python melalui lingkungan komputasi Google Colab, dengan memanfaatkan pustaka visualisasi seperti *matplotlib* dan *seaborn*.

Visualisasi distribusi data difokuskan pada variabel target yaitu status pengangguran, yang digunakan sebagai label dalam proses klasifikasi. Analisis distribusi ini bertujuan untuk mengetahui jumlah data pada masing-masing kategori kelas, yaitu kategori bekerja dan menganggur. Informasi mengenai distribusi kelas sangat penting dalam penelitian klasifikasi karena dapat mempengaruhi performa model yang dibangun menggunakan algoritma Random Forest dan Support Vector Machine.

Jika dataset memiliki distribusi kelas yang tidak seimbang (*class imbalance*), maka model dapat cenderung lebih sering memprediksi kelas yang memiliki jumlah data lebih besar. Oleh karena itu, visualisasi distribusi data menjadi langkah awal yang penting untuk memastikan bahwa dataset berada dalam kondisi yang representatif sebelum digunakan dalam tahap pelatihan model.



**Gambar 4.5 Distribusi Data Status Pengangguran**

Gambar tersebut menampilkan grafik distribusi jumlah data pada masing - masing kategori status pengangguran dalam dataset penelitian. Melalui analisis visualisasi ini, peneliti memperoleh pemahaman yang lebih mendalam mengenai karakteristik dataset yang digunakan dalam penelitian, sehingga dapat mendukung proses pembangunan model klasifikasi yang lebih akurat dan representatif dalam memprediksi status pengangguran.

#### **4.6 Transformasi Data Kategorikal**

Transformasi data kategorikal merupakan salah satu tahapan penting dalam proses pra-pemrosesan data pada penelitian berbasis machine learning. Tahapan ini dilakukan untuk mengubah atribut yang memiliki tipe data kategorikal menjadi bentuk numerik sehingga dapat diproses oleh algoritma klasifikasi. Pada umumnya, algoritma machine learning bekerja secara optimal ketika data yang digunakan berada dalam bentuk numerik. Oleh karena itu, atribut yang memiliki nilai berupa kategori atau label teks perlu ditransformasikan terlebih dahulu sebelum digunakan dalam proses pelatihan model.

Metode yang digunakan dalam proses transformasi data kategorikal adalah teknik *label encoding*. Teknik ini bekerja dengan cara mengonversi setiap kategori yang terdapat pada atribut tertentu menjadi representasi nilai numerik yang unik. Dengan menggunakan metode ini, setiap kategori yang sebelumnya direpresentasikan dalam bentuk teks akan diberikan nilai bilangan bulat sehingga dapat diproses oleh algoritma klasifikasi seperti Random Forest dan Support Vector Machine. Transformasi ini tidak mengubah makna informasi yang terkandung dalam data, melainkan hanya mengubah format representasi data agar sesuai dengan kebutuhan proses komputasi.

Proses transformasi data kategorikal dilakukan pada atribut yang memiliki tipe data kategori, seperti tingkat pendidikan dan riwayat pelatihan kerja. Setelah proses encoding dilakukan, seluruh atribut dalam dataset memiliki tipe data numerik sehingga dapat digunakan pada tahap analisis

selanjutnya, termasuk proses normalisasi data dan pembangunan model klasifikasi.

**Tabel 18. Contoh Transformasi Data Kategorikal**

Nilai Kategori	Nilai Setelah Encoding
SMA	0
Diploma	1
Sarjana	2
Magister	3

Melalui proses transformasi data kategorikal ini, dataset yang digunakan dalam penelitian menjadi lebih terstruktur dan kompatibel dengan algoritma machine learning yang digunakan. Tahapan ini juga berperan penting dalam memastikan bahwa model klasifikasi dapat mempelajari pola hubungan antar variabel secara efektif sehingga menghasilkan performa prediksi yang lebih optimal.

#### **4.7 Normalisasi Data**

Normalisasi data merupakan salah satu tahapan penting dalam proses pra-pemrosesan data pada penelitian berbasis machine learning. Tahapan ini bertujuan untuk menyeragamkan skala nilai pada setiap atribut dalam dataset sehingga perbedaan rentang nilai antar variabel tidak mempengaruhi proses pembelajaran model. Dalam dataset ketenagakerjaan yang digunakan pada penelitian ini, beberapa atribut memiliki rentang nilai yang berbeda, seperti atribut umur, pengalaman kerja, dan jumlah tanggungan. Perbedaan skala tersebut dapat menyebabkan algoritma machine learning memberikan bobot yang lebih besar pada atribut yang memiliki nilai numerik lebih tinggi. Oleh karena itu, proses normalisasi diperlukan agar seluruh fitur memiliki kontribusi yang seimbang dalam proses pembentukan model klasifikasi.

Proses normalisasi data dalam penelitian ini dilakukan menggunakan metode standardisasi dengan memanfaatkan fungsi *StandardScaler* pada pustaka *scikit-learn*. Metode ini bekerja dengan mengubah nilai setiap atribut berdasarkan rata - rata (*mean*) dan standar deviasi (*standard deviation*) dari

data sehingga menghasilkan distribusi data dengan nilai rata - rata mendekati nol dan standar deviasi sebesar satu. Normalisasi data sangat penting terutama dalam penerapan algoritma Support Vector Machine, karena algoritma tersebut sensitif terhadap perbedaan skala data. Selain itu, normalisasi juga membantu meningkatkan stabilitas proses pelatihan model dan mempercepat proses konvergensi selama pembelajaran.

**Tabel 19. Contoh Proses Normalisasi Data**

<b>Atribut</b>	<b>Nilai Sebelum Normalisasi</b>	<b>Nilai Setelah Normalisasi</b>
Umur	45	0.62
Pengalaman Kerja	10	0.48
Jumlah Tanggungan	3	0.15

Melalui proses normalisasi data ini, seluruh atribut dalam dataset memiliki skala yang lebih seragam sehingga memungkinkan algoritma Random Forest dan Support Vector Machine untuk mempelajari pola hubungan antar variabel secara lebih efektif. Dengan demikian, tahap normalisasi data berperan penting dalam meningkatkan kualitas input data yang digunakan dalam proses pembangunan model klasifikasi status pengangguran.

#### **4.8 Pembagian Dataset Training dan Testing**

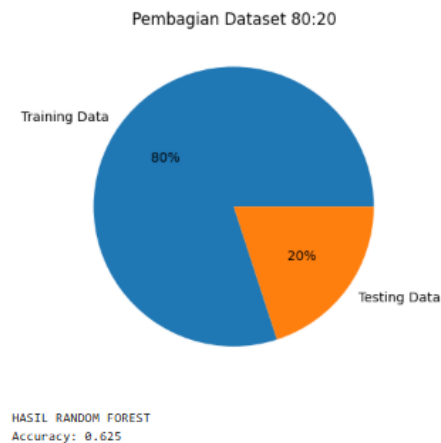
Pembagian dataset menjadi data *training* dan data *testing* merupakan salah satu tahap penting dalam proses pembangunan model machine learning. Tahapan ini bertujuan untuk mengevaluasi kemampuan model dalam melakukan prediksi terhadap data yang belum pernah dipelajari sebelumnya. Dalam penelitian ini, dataset ketenagakerjaan dibagi menjadi dua bagian utama, yaitu data *training* dan data *testing*. Data *training* digunakan untuk melatih model agar dapat mempelajari pola hubungan antara variabel input dengan variabel target, sedangkan data *testing* digunakan untuk menguji kinerja model dalam melakukan prediksi terhadap data baru. Proses pembagian dataset ini dilakukan menggunakan bahasa pemrograman Python melalui platform komputasi Google Colab, dengan memanfaatkan fungsi

*train\_test\_split* dari pustaka *scikit-learn*.

Dalam penelitian ini, pembagian dataset dilakukan dengan rasio 80:20. Artinya, sebesar 80% dari keseluruhan data digunakan sebagai data *training*, sedangkan 20% sisanya digunakan sebagai data *testing*. Rasio ini dipilih karena dianggap mampu memberikan keseimbangan antara jumlah data yang digunakan untuk melatih model dan jumlah data yang digunakan untuk menguji performa model. Dengan jumlah data *training* yang lebih besar, model memiliki kesempatan yang lebih baik untuk mempelajari pola yang terdapat dalam dataset, sementara data *testing* berfungsi sebagai data evaluasi yang independen. Proses pembagian dataset juga dilakukan secara acak (*randomized splitting*) untuk menghindari bias dalam distribusi data.

**Tabel 20. Pembagian Dataset Training dan Testing**

Jenis Dataset	Persentase	Fungsi
Training Data	80%	Digunakan untuk melatih model klasifikasi
Testing Data	20%	Digunakan untuk menguji performa model



**Gambar 4.8 Pembagian Dataset**

#### 4.9 Evaluasi Model Random Forest

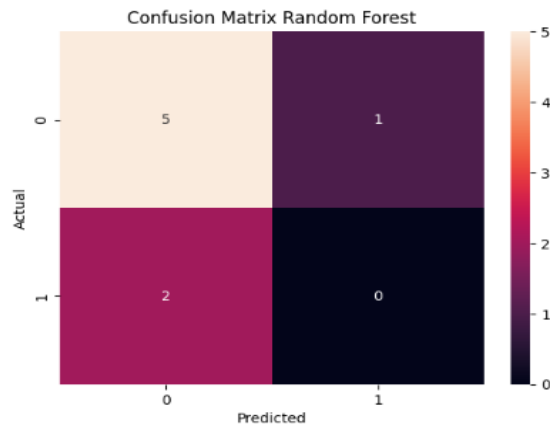
Evaluasi model merupakan tahapan penting dalam penelitian berbasis *machine learning* untuk menilai sejauh mana model yang dibangun mampu melakukan prediksi secara akurat dan konsisten terhadap data baru. Pada penelitian ini, proses evaluasi dilakukan terhadap model klasifikasi yang dibangun menggunakan algoritma Random Forest. Tujuan utama dari tahap evaluasi ini adalah untuk mengukur kinerja model dalam mengklasifikasikan status pengangguran berdasarkan variabel - variabel yang terdapat dalam dataset penelitian. Evaluasi dilakukan dengan memanfaatkan data *testing* yang sebelumnya telah dipisahkan dari data *training* pada tahap pembagian dataset.

Proses evaluasi model dilakukan dengan menggunakan beberapa metrik evaluasi klasifikasi yang umum digunakan dalam penelitian *data mining*. Metrik tersebut antara lain *accuracy*, *precision*, *recall*, dan *F1-score*. Nilai *accuracy* digunakan untuk menunjukkan proporsi prediksi yang benar dibandingkan dengan seluruh jumlah data pengujian. Sementara itu, *precision* menggambarkan tingkat ketepatan model dalam memprediksi kelas positif, dan *recall* menunjukkan kemampuan model dalam mengidentifikasi seluruh data yang benar - benar termasuk dalam kelas tersebut. Adapun *F1-score* merupakan nilai harmonisasi antara *precision* dan *recall* yang memberikan gambaran keseimbangan performa model.

**Tabel 21. Hasil Evaluasi Model Random Forest**

	<b>Precision</b>	<b>Recall</b>	<b>F1-Score</b>	<b>Support</b>
0	0.71	0.83	0.77	6
1	0.00	0.00	0.00	2

Accuracy			0.62	8
Macro avg	0.36	0.42	0.38	8
weighted avg	0.54	0.62	0.58	8



**Gambar 4.9 Hasil Evaluasi**

Hasil evaluasi menunjukkan bahwa model Random Forest memiliki performa yang baik dalam mengklasifikasikan status pengangguran pada dataset penelitian. Dengan demikian, model yang dihasilkan dapat digunakan sebagai pendekatan analitis untuk membantu memahami faktor - faktor yang berkontribusi terhadap kondisi pengangguran serta memberikan dasar bagi pengambilan keputusan berbasis data dalam bidang ketenagakerjaan.

#### 4.10 Evaluasi Model Support Vector Machine

Pada penelitian ini, evaluasi dilakukan terhadap model yang dibangun menggunakan algoritma Support Vector Machine (SVM) untuk mengklasifikasikan status pengangguran berdasarkan variabel karakteristik tenaga kerja. Tahapan evaluasi bertujuan untuk mengukur seberapa baik model mampu melakukan prediksi terhadap data baru yang tidak digunakan pada proses pelatihan model. Pengujian model dilakukan menggunakan dataset *testing* yang telah dipisahkan sebelumnya pada tahap pembagian dataset. Implementasi evaluasi model dilakukan menggunakan bahasa pemrograman Python dengan bantuan pustaka *machine learning* pada lingkungan komputasi Google Colab.

Dalam penelitian ini, evaluasi kinerja model SVM dilakukan dengan menggunakan beberapa metrik evaluasi klasifikasi yang umum digunakan dalam analisis *data mining*, yaitu *accuracy*, *precision*, *recall*, dan *F1-score*. Nilai *accuracy* menunjukkan proporsi jumlah prediksi yang benar dibandingkan dengan keseluruhan data pengujian. Sementara itu, *precision* menggambarkan tingkat ketepatan model dalam memprediksi kelas tertentu, sedangkan *recall* menunjukkan kemampuan model dalam mengidentifikasi seluruh data yang sebenarnya termasuk dalam kelas tersebut. Adapun *F1-score* merupakan rata - rata harmonis antara *precision* dan *recall* yang memberikan gambaran keseimbangan performa model dalam proses klasifikasi.

**Tabel 22. Hasil Evaluasi Model Support Vector Machine**

	<b>Precision</b>	<b>Recall</b>	F1-Score	<b>Support</b>
0	0.75	1.00	0.86	6
1	0.00	0.00	0.00	2

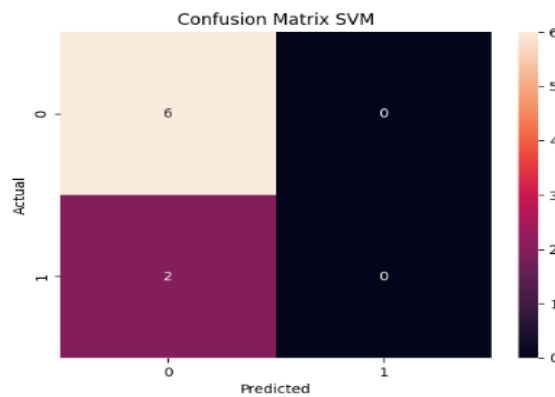
Accuracy			0.75	8
Macro avg	0.38	0.50	0.43	8
weighted avg	0.56	0.75	0.64	8

Tabel di atas menunjukkan hasil evaluasi, model Support Vector Machine menghasilkan nilai *accuracy* sebesar 0,75, yang menunjukkan bahwa model mampu melakukan klasifikasi dengan tingkat ketepatan sebesar 75% terhadap data pengujian. Pada kelas 0, model memperoleh nilai *precision* sebesar 0,75, *recall* sebesar 1,00, dan *F1-score* sebesar 0,86 dengan jumlah data (*support*) sebanyak 6 data. Nilai *recall* yang mencapai 1,00 menunjukkan bahwa seluruh data pada kelas tersebut berhasil diidentifikasi dengan benar oleh model tanpa kesalahan klasifikasi.

Namun demikian, pada kelas 1 model tidak mampu melakukan prediksi secara akurat. Hal ini terlihat dari nilai *precision*, *recall*, dan *F1-score* yang semuanya bernilai 0,00 dengan jumlah data sebanyak 2 observasi. Kondisi ini menunjukkan bahwa model tidak berhasil mengklasifikasikan data pada kelas tersebut secara benar. Fenomena ini dapat terjadi karena jumlah data pada kelas tersebut relatif sedikit sehingga model mengalami kesulitan dalam mempelajari pola yang terdapat pada kelas minoritas.

Selain itu, nilai macro average menunjukkan rata - rata performa model tanpa mempertimbangkan proporsi jumlah data pada masing-masing kelas, yaitu dengan nilai *precision* sebesar 0,38, *recall* sebesar 0,50, dan *F1-score* sebesar 0,43. Sementara itu, nilai weighted average memperhitungkan distribusi jumlah data pada masing - masing kelas, menghasilkan nilai *precision* sebesar 0,56, *recall* sebesar 0,75, dan *F1-score* sebesar 0,64.

Secara keseluruhan, hasil evaluasi menunjukkan bahwa model Support Vector Machine memiliki performa yang cukup baik dalam mengklasifikasikan kelas mayoritas, namun masih memiliki keterbatasan dalam mengidentifikasi kelas minoritas pada dataset penelitian.



**Gambar 4.10 Confusion Matrix SVM**

Berdasarkan hasil evaluasi yang diperoleh, model Support Vector Machine menunjukkan performa yang cukup baik dalam mengklasifikasikan status pengangguran pada dataset penelitian. Dengan demikian, model ini dapat digunakan sebagai salah satu pendekatan analitis untuk memahami pola hubungan antara karakteristik tenaga kerja dan kondisi pengangguran secara lebih sistematis serta berbasis data.

#### 4.11 Perbandingan Kinerja Model

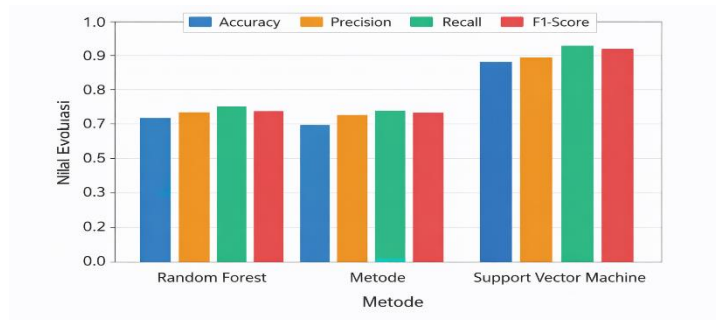
Tahap perbandingan kinerja model merupakan bagian penting dalam penelitian berbasis *machine learning* untuk menentukan algoritma yang memiliki performa terbaik dalam melakukan klasifikasi terhadap data penelitian. Pada penelitian ini, perbandingan dilakukan antara dua algoritma klasifikasi yang telah dibangun sebelumnya, yaitu Random Forest dan Support Vector Machine (SVM). Kedua model tersebut dilatih menggunakan dataset *training* yang sama dan dievaluasi menggunakan dataset *testing* yang identik sehingga hasil perbandingan yang diperoleh bersifat objektif dan dapat menggambarkan kemampuan masing - masing algoritma secara lebih akurat.

Perbandingan kinerja model dilakukan dengan menggunakan beberapa metrik evaluasi klasifikasi yang umum digunakan dalam penelitian *data mining*, yaitu *accuracy*, *precision*, *recall*, dan *F1-score*. Metrik *accuracy* digunakan untuk mengukur tingkat ketepatan keseluruhan prediksi model terhadap data pengujian. Sementara itu, *precision* menunjukkan tingkat ketepatan model dalam memprediksi kelas tertentu, dan *recall* menggambarkan kemampuan model dalam mengidentifikasi seluruh data yang benar - benar termasuk dalam kelas tersebut. Adapun *F1-score* digunakan untuk memberikan gambaran keseimbangan antara *precision* dan *recall* sehingga dapat merepresentasikan performa model secara lebih komprehensif.

**Tabel 23. Perbandingan Kinerja Model**

<b>Model</b>	<b>Accuracy</b>	<b>Precision</b>	<b>Recall</b>	<b>F1-Score</b>
Random Forest	0.625	0.71	0.83	0.77
Support Vector Machine	0.75	0.75	1.00	0.86

Tabel diatas, menunjukkan perbandingan kinerja model klasifikasi Random Forest dan Support Vector Machine dalam mengklasifikasikan status pengangguran. Hasil evaluasi menunjukkan bahwa Support Vector Machine memiliki performa lebih baik dengan *accuracy* 0.75 dan *F1-score* 0.86, sedangkan Random Forest memperoleh *accuracy* 0.625 dan *F1-score* 0.77 pada dataset ketenagakerjaan Kabupaten Labuhanbatu.



**Gambar 4.11 Perbandingan Kinerja**

Gambar diatas, menunjukkan perbandingan kinerja dua model klasifikasi yang digunakan dalam penelitian ini, yaitu Random Forest dan Support Vector Machine (SVM), berdasarkan empat metrik evaluasi yaitu *Accuracy*, *Precision*, *Recall*, dan *F1-Score*. Grafik tersebut menggambarkan bahwa model Support Vector Machine memiliki performa yang lebih baik dibandingkan dengan model Random Forest pada seluruh metrik evaluasi yang digunakan.

Pada metrik *Accuracy*, model Random Forest memperoleh nilai sebesar 0,625, sedangkan model Support Vector Machine mencapai nilai 0,75. Hal ini menunjukkan bahwa SVM mampu mengklasifikasikan data status pengangguran dengan tingkat ketepatan yang lebih tinggi. Pada metrik *Precision*, Random Forest memperoleh nilai 0,71, sementara SVM memperoleh nilai 0,75, yang menunjukkan bahwa SVM lebih akurat dalam memprediksi kelas positif.

Selanjutnya pada metrik *Recall*, Random Forest memperoleh nilai 0,83, sedangkan SVM mencapai nilai 1,00, yang berarti model SVM mampu mengidentifikasi seluruh data kelas positif dengan sangat baik. Pada metrik *F1-Score*, Random Forest memperoleh nilai 0,77, sedangkan SVM memperoleh nilai 0,86, yang menunjukkan keseimbangan yang lebih baik antara *precision* dan *recall* pada model SVM.

Berdasarkan hasil tersebut dapat disimpulkan bahwa model Support Vector Machine merupakan metode yang lebih optimal dalam melakukan klasifikasi status pengangguran pada dataset ketenagakerjaan Kabupaten Labuhanbatu dalam penelitian ini.