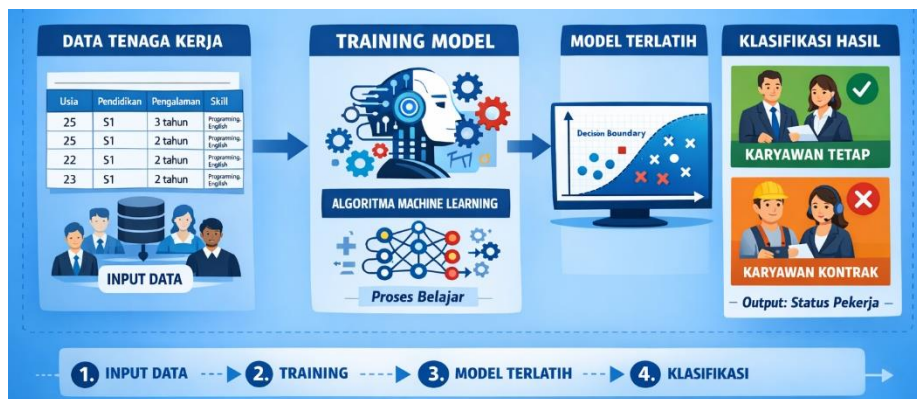


## BAB II

### LANDASAN TEORI

#### 2.1 Teori dan Konsep yang Mendasari Penelitian

Penelitian ini berlandaskan pada konsep *machine learning* yang digunakan untuk mengklasifikasikan status pengangguran berdasarkan data ketenagakerjaan. *Machine learning* memungkinkan sistem komputer untuk mempelajari pola dari data tanpa pemrograman eksplisit, kemudian digunakan untuk memprediksi atau mengklasifikasi data baru secara otomatis. Dua algoritma utama yang menjadi fokus penelitian ini adalah *Random Forest* (RF) dan *Support Vector Machine* (SVM). RF merupakan metode *ensemble learning* berbasis pohon keputusan yang menggabungkan hasil dari beberapa *decision tree* untuk menghasilkan prediksi yang lebih akurat. Sedangkan SVM berfungsi untuk memisahkan kelas data dengan menentukan *hyperplane* optimal di antara data berdimensi tinggi. Pendekatan ini relevan untuk klasifikasi status pengangguran karena mampu menangani data yang tidak linear dan memiliki variabel multidimensi (Wei et al., 2023; Annastasya et al., 2025; Ramadhanti, 2021).



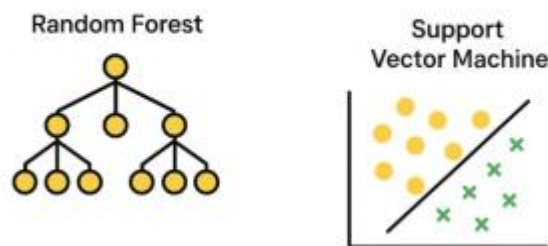
Gambar 2.1

#### Konsep dasar machine learning dalam klasifikasi data tenaga kerja

Gambar di atas menjelaskan konsep dasar dari penerapan machine learning dalam klasifikasi data ketenagakerjaan, serta peran masing - masing algoritma (*Random Forest* dan *SVM*) dalam meningkatkan akurasi klasifikasi status pengangguran.

## 2.2 Algoritma yang Digunakan

Algoritma *Random Forest* dan *Support Vector Machine* merupakan metode klasifikasi populer. RF bekerja dengan membangun banyak pohon keputusan dari subset data dan kemudian menggabungkan hasil prediksi (*bagging*) untuk mengurangi *overfitting* (Rohaeti & Andriyati, 2024; Ernawati et al., 2025; Yunizar et al., 2025). Sebaliknya, SVM mengidentifikasi *hyperplane* terbaik yang memisahkan kelas data secara maksimal (Syahputra & Wibowo, 2023; Haikal & Palupi, 2024). SVM cocok untuk dataset kecil hingga menengah dengan batas keputusan yang kompleks, sementara RF lebih unggul pada dataset besar dengan banyak variabel. Kombinasi keduanya memungkinkan analisis perbandingan yang adil terhadap performa klasifikasi status pengangguran.



Gambar 2.2

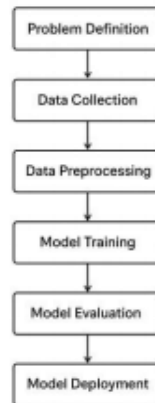
### Perbandingan visual algoritma Random Forest dan Support Vector Machine

Gambar di atas menjelaskan perbedaan antara algoritma *Random Forest* dan *Support Vector Machine* dalam konteks klasifikasi status pengangguran, serta memberikan konteks bagi perbandingan performa kedua algoritma tersebut dalam penelitian ini.

## 2.3 Langkah - Langkah *Machine Learning*

Tahapan utama *machine learning* meliputi: (1) pengumpulan data, (2) prapemrosesan data, (3) pembagian data menjadi data latih dan data uji, (4) pelatihan model, (5) evaluasi model, dan (6) penyimpulan hasil. Pada penelitian ini, data pengangguran dari Dinas Tenaga Kerja Labuhanbatu diolah melalui serangkaian langkah tersebut menggunakan RF dan SVM.

Proses pra-pemrosesan mencakup penanganan data hilang, normalisasi, dan *data balancing* untuk memastikan proporsi kelas seimbang (Maehendrayuga & Setyanto, 2024; Abidin et al., 2025).



**Gambar 2.3**

**Tahapan umum proses machine learning.**

Gambar di atas menjelaskan langkah - langkah yang diambil dalam penelitian untuk menerapkan machine learning pada data ketenagakerjaan Kabupaten Labuhanbatu, dengan tujuan untuk mengklasifikasikan status pengangguran secara akurat.

**2.4 Teknik Evaluasi Algoritma *Machine Learning***

Evaluasi model digunakan untuk menilai performa algoritma. Metrik umum yang digunakan meliputi *accuracy*, *precision*, *recall*, dan *F1-score*. Evaluasi dilakukan menggunakan *confusion matrix* untuk memvisualisasikan hasil klasifikasi antara kelas aktual dan prediksi. Pada penelitian ini, teknik *crossvalidation* digunakan untuk menghindari bias hasil (Damayanti & Baita, 2025; Maulana As'an Hamid & Subhiyakto, 2025). RF dan SVM dibandingkan berdasarkan metrik tersebut guna menentukan model terbaik dalam klasifikasi status pengangguran.

		Aktual	
		Negatif	Positif
Prediksi	Negatif	True Negative	False Positive
	Positif	False Negative	True Positive

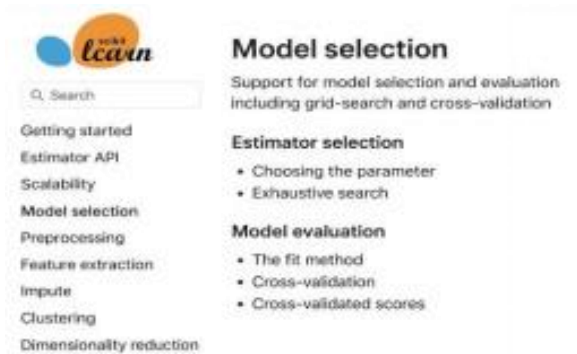
**Gambar 2.4**

### **Contoh Confusion matrix untuk evaluasi model**

Gambar di atas menunjukkan contoh confusion matrix, yang digunakan untuk mengevaluasi kinerja model klasifikasi dalam penelitian ini.

## **2.5 Alat Bantu Pemrograman dan *Tools* Pendukung**

Penelitian ini menggunakan *Python* sebagai bahasa pemrograman utama dengan pustaka pendukung seperti *Scikit-learn*, *NumPy*, dan *Pandas*. *Scikit-learn* menyediakan implementasi siap pakai untuk RF dan SVM, sedangkan *NumPy* dan *Pandas* digunakan untuk pengolahan data. Visualisasi dilakukan dengan *Matplotlib* dan *Seaborn* (Novianto et al., 2024; Elim & Utami, 2025). Selain itu, *Google Colab* dimanfaatkan sebagai platform komputasi berbasis awan (Yunizar et al., 2025).



**Gambar 2.5**

### **Tampilan antarmuka Scikit - learn untuk eksperimen machine learning**

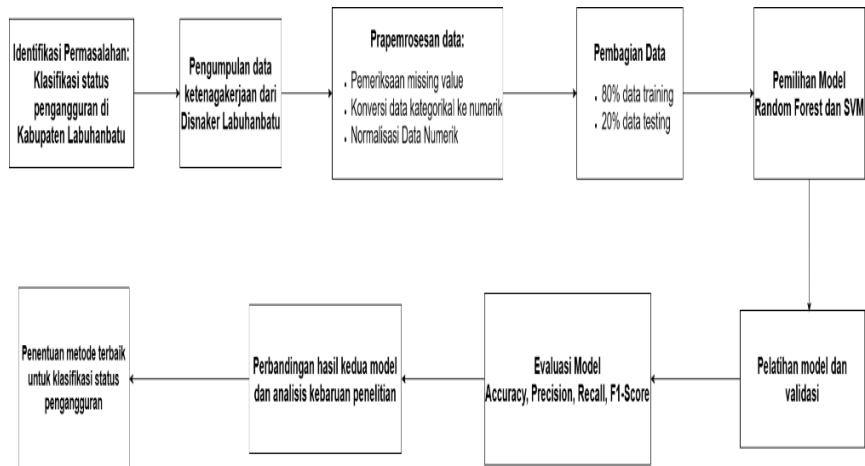
Gambar di atas menunjukkan tampilan antarmuka Scikit-learn, salah satu pustaka (library) populer untuk eksperimen machine learning di Python. Antarmuka ini menyediakan berbagai fitur untuk mendukung model selection dan evaluasi model, yang digunakan dalam penelitian untuk menguji dan membandingkan berbagai algoritma machine learning.

## 2.6 Penelitian Terdahulu dan Kelebihan Penelitian Ini

No	Peneliti	Judul Penelitian	Tahun	Data dan Algoritma	Hasil
1	E. Rohaeti dan A. Andriyati	<i>Comparative Study of Predictive Classification Models on Data with Severely Imbalanced Predictors</i>	2024	Data sosialekonomi; <i>Random Forest</i> dan <i>SVM</i>	<i>Random Forest</i> menghasilkan akurasi tertinggi (92%) dibanding <i>SVM</i> .
2	A. A. Maigur	<i>Machine Learning Algorithms for Predicting Unemployment Duration in Russia</i>	2024	Data ketenagakerjaan Rusia; <i>SVM</i> dan <i>Gradient Boosting</i>	<i>SVM</i> unggul pada data berdimensi tinggi dan linear separable.
3	D. Vallarino	<i>Decoding the Puzzle of Joblessness: Machine Learning Predicts Unemployment Trends in the Americas</i>	2025	Data tren pengangguran di Amerika; <i>Random Forest</i>	<i>Random Forest</i> memberikan interpretasi yang lebih mudah terhadap variabel sosial.

4	H. Syahputra dan A. Wibowo Wi	<i>Comparison of Support Vector Machine (SVM) and Random Forest Algorithms for Detection of Negative Content on Websites</i>	2023	Data teks daring; SVM dan Random Forest	SVM memiliki <i>precision</i> tinggi, Random Forest stabil pada data besar.
5	S. A. Utirahman dan A. M. M. Pratama	Penerapan <i>Support Vector Machine</i> dan <i>Random Forest Classifier</i> untuk Klasifikasi Tingkat Obesitas	2024	Data kesehatan; Random Forest dan SVM	Random Forest menghasilkan <i>accuracy</i> terbaik (95%).
6	Ayu Widasari Siregar	<i>Perbandingan Metode Random Forest dan Support Vector Machine untuk Klasifikasi Status Pengangguran di Dinas Tenaga Kerja Kabupaten Labuhanbatu</i>	2025	Data ketenagakerjaan Labuhanbatu; Random Forest dan SVM	Masih dalam proses penelitian

## 2.7 Kerangka Penelitian



**Gambar 2.7 Kerangka Penelitian**

Gambar di atas menunjukkan flowchart kerangka penelitian yang menggambarkan langkah - langkah utama dalam proses penelitian tentang klasifikasi status pengangguran menggunakan algoritma Random Forest dan Support Vector Machine (SVM) di Kabupaten Labuhanbatu. Flowchart ini memberikan gambaran jelas tentang tahapan - tahapan yang dilakukan untuk mencapai tujuan penelitian.