

BAB III

METODE PENELITIAN

3.1 Desain Penelitian

Penelitian ini fokus pada prediksi preferensi etika mahasiswa menggunakan metode penelitian kuantitatif dengan pendekatan *machine learning*, yaitu pendekatan prediktif. Dalam konteks penelitian ini, pendekatan prediktif digunakan untuk menciptakan model yang dapat memproyeksikan tingkat preferensi etika mahasiswa. Proses penelitian melibatkan pengumpulan data set yang selanjutnya disiapkan sebagai *input* untuk algoritma *machine learning*. Dengan demikian, Hasil dari penelitian ini diharapkan dapat memberikan wawasan yang berharga untuk untuk lembaga pendidikan tinggi dalam mengidentifikasi pola dan tren dalam perilaku etis mahasiswa, berdasarkan berbagai konteks dan pengaruh yang ada.

3.2 Sumber Data

Penelitian ini menggunakan jenis data sekunder yang diperoleh dari survei terhadap model etika mahasiswa perguruan tinggi di Yogyakarta, Indonesia, yang dilakukan antara bulan Juli 2018 hingga Desember 2018. Data yang dimanfaatkan diperoleh dari situs web *Data in Brief*, yaitu www.sciencedirect.com, yang telah tersedia secara *open access*. Pemilihan *Data in Brief* sebagai sumber data dilakukan karena jurnal ini menyajikan informasi penelitian secara ringkas dan mudah dipahami, memudahkan pengambilan informasi yang relevan secara efisien. Keputusan ini didasari oleh dukungan terhadap prinsip-prinsip FAIR (*Findable, Accessible, Interoperable, Reusable*), yang menjamin keberlanjutan akses terbuka terhadap data penelitian. Sebagai jurnal *open access*, *Data in Brief* juga mencerminkan komitmen terhadap transparansi dan keterbukaan dalam ilmu pengetahuan. Penggunaan data dari *Data in Brief* dianggap strategis untuk meningkatkan validitas dan reproduktibilitas penelitian, sekaligus menghemat waktu yang seharusnya diperlukan untuk

pengumpulan data primer. Pemilihan sumber data ini juga konsisten dengan nilai-nilai etika berbagi data dan kontribusi terhadap *open science*. Oleh karena itu, *Data in Brief* bukan hanya menjadi sumber informasi, tetapi juga menjadi pondasi yang kokoh untuk membangun landasan data yang kuat dalam rangka pelaksanaan penelitian ini.

3.3 Partisipan

Pada **Tabel 3.1** memberikan gambaran umum tentang informasi partisipan, sebanyak 566 mahasiswa yang mengisi kuisioner. Partisipan pada penelitian ini yaitu Mahasiswa UNY, Indonesia dari berbagai fakultas, yaitu ekonomi, teknik, matematika dan ilmu pengetahuan alam, ilmu sosial, ilmu olahraga, seni, dan ilmu pendidikan.

Tabel 3.1 Data partisipan

Fakultas	Tanggapan	%
Ekonomi dan Bisnis	114	20.1
Teknik	98	17.3
Matematika dan Ilmu Pengetahuan Alam	85	15.0
Ilmu Sosial, Hukum, dan Ilmu Politik	61	10.8
Ilmu Keolahragaan dan Kesehatan	15	2.7
Bahasa, Seni, dan Budaya	120	21.2
Ilmu Pendidikan dan Psikologi	73	12.9
Total	566	100.00

3.4 Instrumen Penelitian

Penelitian ini menggunakan instrumen penelitian yang dikembangkan oleh (Indartono, 2020) untuk mengukur model perilaku etis di kalangan mahasiswa. Informasi terkait setiap pertanyaan atau pernyataan dalam masing-masing atribut dapat ditemukan dalam **Tabel 3.2** hingga **Tabel 3.7**. Adapun **Tabel 3.8** menyajikan pengukuran etika yang nantinya akan dijadikan label dalam kerangka penelitian ini.

Tabel 3.2 Pengukuran motivasi

No	Pertanyaan
Mot 1.	Secara umum, saya percaya saya bisa menyelesaikan setiap tugas dengan baik
Mot 2.	Dalam hal usaha, saya selalu mencoba yang terbaik
Mot 3.	Ketika guru saya bertanya di kelas, saya sering bersedia (mengangkat tangan) untuk menjawab
Mot 4.	Jika saya tidak mengerti sesuatu dalam pekerjaan rumah saya, hal pertama yang saya lakukan adalah mencari tahu atau terus mencoba sendiri
Mot 5.	Saya ingin nilai saya lebih baik

Tabel 3.3 Pengukuran efikasi diri

No	Pertanyaan
SE 1.	Saya merasa percaya diri dalam menganalisis masalah jangka panjang untuk menemukan solusi dalam studi saya
SE 2.	Saya merasa percaya diri mewakili departemen saya di berbagai acara
SE 3.	Saya merasa percaya diri memberikan kontribusi dalam diskusi strategi pembelajaran
SE 4.	Saya merasa percaya diri membantu mencapai target/tujuan di departemen saya
SE 5.	Saya merasa percaya diri menghubungi orang di luar departemen untuk mendiskusikan masalah pembelajaran
SE 6.	Saya merasa percaya diri menyajikan informasi kepada rekan belajar saya

Tabel 3.4 Pengukuran *resilience*

No	Pertanyaan
R1.	Ketika hal-hal yang tidak pasti terjadi pada saya di kampus, saya biasanya mencapai kesimpulan terbaik

No	Pertanyaan
R2.	Ketika kesalahan terjadi pada saya, saya menganggapnya sebagai tanda keberhasilan
R3.	Saya selalu melihat sisi positif dari pembelajaran saya
R4.	Saya optimis tentang apa yang akan terjadi pada saya di masa depan terkait dengan studi saya
R5.	Dalam mencapai tujuan pembelajaran saya, saya telah menghadapi banyak kegagalan
R6.	Dalam pembelajaran, saya selalu menghadapi berbagai rintangan

Tabel 3.5 Pengukuran *team strain*

No	Pertanyaan
TS1.	Kelompok studi saya merasa bahwa jika ada masalah di kampus (misalnya, nilai, komunikasi, dll.), maka tugas mata kuliah akan membantu menyelesaikan masalah tersebut
TS2.	Kelompok studi saya merasa bahwa masalah-masalah dalam lingkungan kampus (terkait peluang pekerjaan, harapan orang tua, atau kurikulum) dapat diatasi dengan tugas mata kuliah
TS3.	Kelompok studi saya merasa bahwa, jika ada masalah dengan ketenagakerjaan, maka tugas praktik industri dapat membantu menyelesaikan masalah tersebut

Tabel 3.6 Pengukuran *knowledge articulation*

No	Pertanyaan
KA1.	Mahasiswa yang terlibat dalam berbagai kegiatan sosial secara rutin ditanyai mengenai pengalaman kerjasama mereka
KA2.	Mahasiswa yang bertanggung jawab dalam kerjasama selalu membuat catatan (dalam bentuk memo, catatan, laporan, atau presentasi) dari semua kegiatan penting, keputusan, atau tindakan terkait kerjasama tersebut
KA3.	Mahasiswa yang mengikuti program kerjasama secara teratur melaporkan kemajuan dan kinerja terkait kerjasama mereka masing-masing

Tabel 3.7 Pengukuran *cooperative classroom environment*

No	Pertanyaan
CCE1.	Kelas lebih menyenangkan ketika saya belajar dengan teman-teman lain
CCE2.	Saya belajar paling baik ketika bersama teman-teman sekelas
CCE3.	Saya mendapatkan nilai lebih baik ketika belajar bersama teman-teman lain
CCE4.	Saya lebih suka mengambil kelas dimana siswa belajar bersama untuk menyelesaikan masalah

Tabel 3.8 Pengukuran perilaku etis

No	Pertanyaan
ET 1.	Saya memegang prinsip bahwa kejujuran lebih penting daripada mendapatkan nilai yang baik
ET 2.	Saya bertanggung jawab penuh atas setiap tindakan tidak etis yang saya lakukan (misalnya, saya akan mengakui jika dosen menemukan saya melakukan plagiarisme pada beberapa tugas)
ET 3.	Saya berperilaku secara etis dan mematuhi peraturan serta kode etik yang ditetapkan oleh universitas
ET 4.	Saya akan menerima semua pendapat/pertimbangan orang lain jika saya perlu membuat keputusan mengenai dilema etika
ET 5.	Selama studi saya di universitas, saya merujuk kepada orang lain untuk menyelesaikan dilema etika
ET 6.	Saya secara pribadi menghadapi dilema etika selama belajar di universitas
ET 7.	Saya pernah dihadapkan dengan dilema etika selama studi di universitas
ET 8.	Fakultas (yaitu, dosen, administrator) akan memberi penghargaan kepada saya ketika saya melakukan sesuatu yang etis

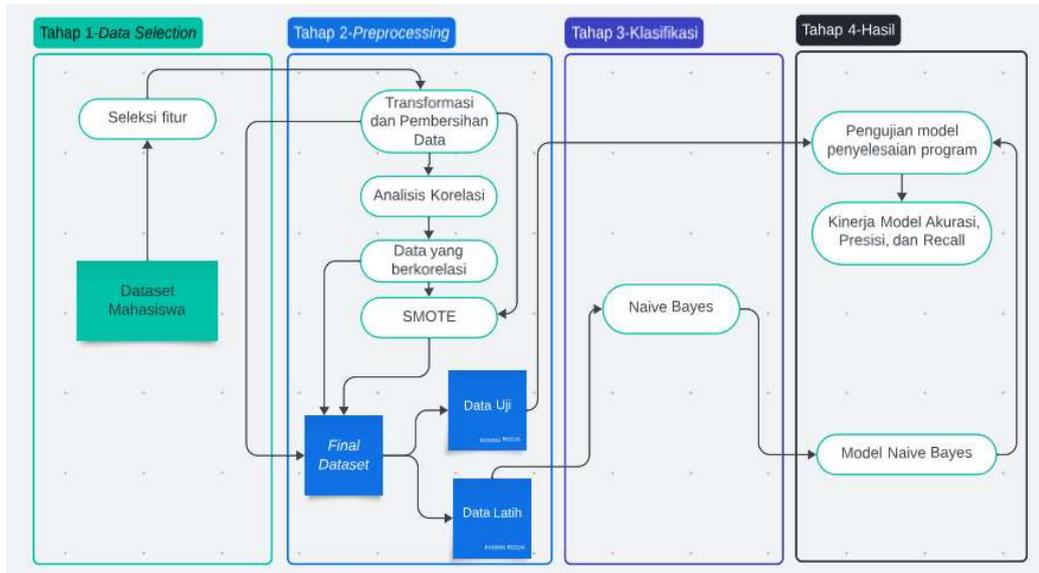
3.5 Prosedur Penelitian

Tahap awal penelitian ini melibatkan pemilihan topik, dengan fokus pada prediksi preferensi perilaku etis di kalangan mahasiswa. Setelah itu, dilakukan studi literatur untuk memperoleh pemahaman mendalam tentang algoritma-algoritma yang

digunakan dalam prediksi perilaku etis mahasiswa, dengan mengacu pada beberapa jurnal internasional, artikel, dan buku terkait. Langkah berikutnya adalah merumuskan masalah dan tujuan penelitian secara jelas. Setelah merinci masalah dan tujuan, penelitian melanjutkan dengan pemilihan pendekatan dan metode penelitian yang sesuai. Selanjutnya, instrumen-instrumen yang digunakan diambil dari jurnal dan disaring untuk keperluan penelitian. Setelah memastikan instrumen yang akan digunakan, data yang diperoleh dipindahkan ke dalam format *Microsoft Excel* guna mempermudah proses pengolahan data. Langkah terakhir melibatkan pengolahan data menggunakan perangkat lunak *RapidMiner Studio*, dengan tujuan memperoleh hasil yang kemudian dianalisis dan diintegrasikan ke dalam penelitian secara menyeluruh.

3.6 Pengolahan dan Analisis Data

Analisis data pada penelitian ini mengadopsi proses penemuan pengetahuan (*knowledge discovery process/KDP*). Rangkaian proses KDP terdiri dari *data selection*, *preprocessing*, transformasi, *data mining*, dan Interpretasi. *Data input* awalnya dipilih, dan data target diisolasi. *Preprocessing* dan transformasi dilakukan untuk memastikan keandalan *database*, dengan *Data mining* menjadi analisis inti. Proses penemuan pengetahuan diakhiri dengan menginterpretasikan hasil. Berdasarkan skema KDP ini, penelitian ini merumuskan suatu kerangka, sebagaimana terlihat dalam **Gambar 3.1**. Penelitian ini melibatkan beberapa tahap metodologi. Tahap pertama adalah *data selection*, yang melibatkan pengumpulan dan produksi data set dari *database* mahasiswa universitas. Pada tahap kedua, dilakukan *preprocessing data*, tahap ini melibatkan *data cleaning* dan transformasi data. Tahap ketiga, yaitu proses klasifikasi di mana model dibangun dengan menerapkan algoritma *Naïve Bayes* untuk memprediksi tingkat etika mahasiswa. Terakhir, tahap keempat mencakup hasil, di mana data dievaluasi dan diinterpretasi. Selain itu pola menarik diidentifikasi dan disajikan untuk merepresentasikan pengetahuan yang diperoleh dari penelitian ini. (Perez & Perez, 2021).



Gambar 3.1 Kerangka kerja penelitian (Perez & Perez, 2021)

3.6.1 Data Selection

Tahap awal penelitian ini melibatkan proses seleksi data, di mana data set yang digunakan pada penelitian ini, terdiri dari 566 data yang mencakup 6 atribut utama yaitu motivasi, efikasi diri, *resilience*, *team strain*, *knowledge articulation*, *cooperative classroom environment*. Selain itu, terdapat variabel etika yang akan dijadikan label dalam analisis. Pemilihan atribut ini didasarkan pada aspek-aspek yang dianggap paling relevan dan penting sesuai dengan tujuan penelitian.

3.6.2 Preprocessing Data

Pada tahap *preprocessing* penelitian ini, dilakukan pengolahan data yang awalnya berbentuk kuesioner menjadi format data kuantitatif dalam lembar Excel. Tujuannya adalah untuk mempermudah proses klasifikasi data. Selanjutnya, setelah data diolah, dilakukan pembersihan data untuk menangani

duplikasi. Hasil dari langkah pembersihan data ini adalah pengurangan jumlah data dari 566 menjadi 557 data.

Dalam pengolahan data hasil dari kuesioner, penelitian ini menggunakan nilai dalam skala likert 5 yang mempunyai gradasi dari sangat tidak setuju sampai sangat setuju. Pembobotan dalam skala likert dapat dilihat pada **Tabel 3.9**.

Tabel 3.9 Skala likert

Simbol	Skala Jawaban	Skor
SS	Sangat Setuju	5
S	Setuju	4
N	Netral	3
TS	Tidak Setuju	2
STS	Sangat Tidak Setuju	1

Langkah selanjutnya melibatkan pengambilan nilai rata-rata pada setiap variabel, yang kemudian dibulatkan. Contoh pengisian kuesioner dapat dilihat dalam **Tabel 3.10**.

Tabel 3.10 Pengisian kuisisioner variabel motivasi

No	Mot1	Mot2	Mot3	Mot4	Mot5
1.	4	4	3	4	4

Misalnya, jika seorang responden mengisi kuesioner seperti yang ditunjukkan pada **Tabel 3.10**, maka nilai rata-rata perlu dicari terlebih dahulu, karena pada variabel motivasi terdiri dari 5 pertanyaan, proses pencarian nilai rata-rata dilakukan dengan menjumlahkan jawaban responden pada pertanyaan motivasi, kemudian dibagi dengan jumlah pertanyaan. Dalam contoh tersebut, responden memperoleh nilai rata-rata motivasi sebesar 3,8. Nilai ini kemudian dibulatkan, dan proses serupa dilakukan untuk variabel-variabel lainnya. Hasil rata-rata keseluruhan dapat dilihat pada **Tabel 3.11**.

Tabel 3.11 Data setelah dilakukan pembulatan

No	Mot	R	SE	TS	KA	CCE	Etika
1	4	4	4	3	4	4	4
2	4	5	4	3	4	4	5
3	5	5	5	2	4	4	3
4	4	3	3	3	3	3	4
5	4	4	4	4	3	4	4
6	4	4	4	3	3	4	4
7	4	5	4	4	4	4	4
8	4	3	3	3	4	4	4
9	5	5	4	4	4	5	5
10	2	2	3	3	2	1	1
11	4	4	4	4	4	4	4
...
552	2	2	2	1	3	3	1
553	4	3	3	3	3	3	4
554	4	4	4	4	5	4	4
555	3	3	3	3	4	3	3
556	4	4	3	4	3	4	3
557	4	4	4	2	4	4	4

Selanjutnya, penelitian ini akan melibatkan pemberian label pada setiap data berdasarkan variabel “etika” dengan tujuan mengklasifikasikan tingkat etika. Proses klasifikasi ini bertujuan untuk mengelompokkan nilai-nilai etika ke dalam dua interval nilai. Interval tersebut ditentukan menggunakan **Persamaan (3.1)** (Santika et al., 2023).

$$Interval = \frac{Range}{k} \quad (3.1)$$

Dengan:

k = Jumlah kelas yang diinginkan

$Range$ = Nilai maksimum dikurangi nilai minimum

Jumlah kelas yang diinginkan dalam penelitian ini yaitu dua kelas dengan klasifikasi “rendah” dan “tinggi”, $k = 2$. Skor maksimal dan minimal pada variabel etika yaitu 5 dan 1, maka $range = 4$. Sehingga, interval tingkat etika mahasiswa dapat dilihat pada **Tabel 3.12**.

Tabel 3.12 Interval tingkat etika

Interval	Klasifikasi
1-2	Rendah
3-5	Tinggi

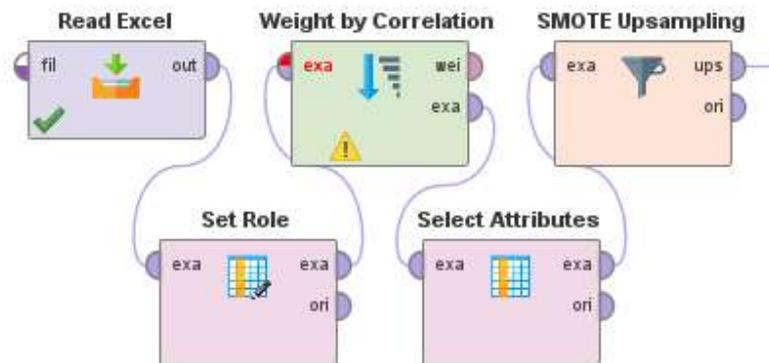
Setelah itu, dilakukan konversi data variabel etika yang akan dijadikan label, sesuai dengan interval yang telah ditentukan. Maka, data tersebut dapat dilihat pada **Tabel 3.13**.

Tabel 3.13 Data setelah dilakukan *preprocessing*

No	Mot	R	SE	TS	KA	CCE	Etika
1	4	4	4	3	4	4	Tinggi
2	4	5	4	3	4	4	Tinggi
3	5	5	5	2	4	4	Tinggi
4	4	3	3	3	3	3	Tinggi
5	4	4	4	4	3	4	Tinggi
6	4	4	4	3	3	4	Tinggi
7	4	5	4	4	4	4	Tinggi
8	4	3	3	3	4	4	Tinggi
9	5	5	4	4	4	5	Tinggi
10	2	2	3	3	2	1	Rendah
11	4	4	4	4	4	4	Tinggi
...
552	2	2	2	1	3	3	Rendah
553	4	3	3	3	3	3	Tinggi
554	4	4	4	4	5	4	Tinggi
555	3	3	3	3	4	3	Tinggi
556	4	4	3	4	3	4	Tinggi
557	4	4	4	2	4	4	Tinggi

Berdasarkan 557 data tingkat etika mahasiswa dengan menerapkan perhitungan skala *likert* pada variabel etika, diperoleh hasil bahwa 11 mahasiswa memiliki tingkat etika rendah (2%), sedangkan 546 mahasiswa memiliki tingkat etika tinggi (98%).

Selanjutnya, data yang telah disiapkan untuk dianalisis dan diproses menggunakan *software RapidMiner Studio*. **Gambar 3.2** menampilkan model pada *RapidMiner Studio* pada tahap *preprocessing*.



Gambar 3.2 Model *preprocessing*

Alur model pada **Gambar 3.2** menggambarkan rangkaian operator yang digunakan dalam pengembangan model berdasarkan data set mahasiswa. Pertama, data set mahasiswa dalam format *excel* di-*import* ke dalam perangkat lunak *RapidMiner Studio* menggunakan operator *read excel*. Selanjutnya, operator *set role* diterapkan untuk menentukan atribut label yang akan menjadi target prediksi.

Proses selanjutnya melibatkan penggunaan operator *weight by correlation*, yang berfungsi untuk menghitung bobot atribut berdasarkan analisis korelasi. Bobot ini mencerminkan tingkat keterkaitan antara atribut label dengan atribut lainnya. Semakin tinggi bobot suatu atribut, semakin

signifikan atribut tersebut dalam konteks prediksi (Perez & Perez, 2021). Dalam konteks ini, operator *weight by correlation* digunakan untuk menentukan bobot dan relevansi atribut tertentu, dengan menggunakan *Pearson r correlation*. *Pearson r correlation* merupakan statistik korelasi yang umum digunakan untuk mengukur tingkat hubungan linier antara dua variabel (Perez & Perez, 2021).

Rumus *Pearson r correlation*:

$$r = \frac{n(\sum xy - (\sum x)\sum y)}{\sqrt{[n\sum x^2 - (\sum x)^2][n\sum y^2 - (\sum y)^2]}} \quad (3.2)$$

Dengan:

r = Koefisien *Pearson r correlation*

n = Jumlah observasi

$\sum xy$ = Jumlah produk dari data yang dipasangkan

$\sum x$ = Jumlah skor x

$\sum y$ = Jumlah skor y

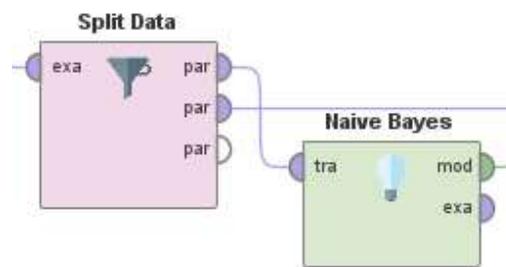
$\sum x^2$ = Jumlah kuadrat skor x

$\sum y^2$ = Jumlah kuadrat skor y

Atribut yang menunjukkan koefisien korelasi yang tinggi antar variabel menjadi input data akhir untuk merancang model klasifikasi *Naïve Bayes*. Penggunaan operator *select attributes* pada **Gambar 3.2** melibatkan penggunaan filter seleksi atribut, di mana opsi *a subset* dipilih. Opsi ini memungkinkan pemilihan beberapa atribut berdasarkan hasil dari operator *weight by correlation*. Selanjutnya, setelah pemilihan atribut berdasarkan *weight by correlation*, penelitian ini melibatkan penggunaan operator *SMOTE Upsampling*. Operator ini digunakan untuk mengatasi masalah ketidakseimbangan kelas dalam data set. Ketidakseimbangan kelas terjadi

ketika jumlah sampel antara kelas mayoritas dan minoritas sangat tidak seimbang (Thammasiri et al., 2014).

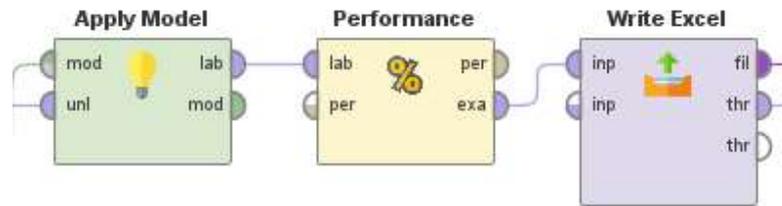
3.6.3 Proses Klasifikasi



Gambar 3.3 Proses Klasifikasi

Tahap ketiga dari penelitian ini melibatkan penggunaan operator *split data* untuk membagi data set menjadi dua subset: data pelatihan dan data pengujian. Tujuan dari tahap ini adalah melatih model menggunakan sebagian besar data dan menguji kinerjanya pada data yang tidak pernah dilihat sebelumnya, sehingga dapat mengukur kehandalan model dalam melakukan prediksi. Langkah terakhir melibatkan penerapan proses klasifikasi menggunakan algoritma *Naïve Bayes*, yang dapat memahami pola dan hubungan dalam data pelatihan untuk kemudian diaplikasikan pada data pengujian.

3.6.4 Hasil



Gambar 3.4 Model data uji dan proses evaluasi

Fase terakhir dari alur kerja penelitian ini mencakup pengujian dan evaluasi model yang telah dikembangkan. Model tersebut diterapkan pada data uji dan dievaluasi menggunakan *confusion matrix cross-validation* untuk mengukur sejauh mana akuratnya model. Penggunaan operator *apply model*, sebagaimana ditunjukkan dalam **Gambar 3.4**, memungkinkan prediksi nilai label untuk setiap data uji. Prediksi ini kemudian ditambahkan ke kolom baru yang diberi nama "*prediction* (etika)". Setelah menerapkan rumus dan menghasilkan kolom proyeksi dalam sampel uji, keluaran model dianalisis menggunakan operator *performance* untuk mendapatkan pemahaman yang lebih mendalam tentang kinerja model dan evaluasi kinerjanya. Performa model diuji dan dievaluasi melalui *confusion matrix* yang mencakup akurasi, presisi, dan *recall*. *confusion matrix* menunjukkan prediksi, baik positif maupun negatif, diatur seperti dalam matriks 2x2. Tabel *confusion matrix* dapat dilihat pada **Tabel 3.14** (Perez & Perez, 2021).

Tabel 3.14 *Confusion matrix*

	<i>Actual Positive</i>	<i>Actual Negative</i>
<i>Predicted Positive</i>	TP (<i>True Positif</i>)	FP (<i>False Positif</i>)
<i>Predicted Negative</i>	FN (<i>False Negatif</i>)	TN (<i>True Negatif</i>)

1. *True Positive*: banyak data yang aktual kelasnya positif, dan model juga memprediksi positif.

2. *True Negative*: banyak data yang aktual kelasnya negatif, dan model memprediksi negatif.
3. *False Positive*: banyak data yang aktual kelasnya negatif, namun model memprediksi positif.
4. *False Negative*: banyak data yang aktual kelasnya positif, namun model memprediksi negatif.

Akurasi mengukur persentase prediksi yang benar oleh model. Presisi mengevaluasi seberapa tepat sebuah model dalam memprediksi suatu kelas. Sedangkan *Recall* mengukur persentase suatu kelas sebenarnya yang diidentifikasi dengan benar oleh model (Perez & Perez, 2021). Berikut adalah parameter kinerja yang digunakan untuk mengevaluasi dan membandingkan model prediksi (Thammasiri et al., 2014).

1. Akurasi

$$Akurasi = \frac{TP + TN}{TP + TN + FP + FN} \quad (3.3)$$

2. Presisi kelas positif

$$Presisi = \frac{TP}{TP + FP} \quad (3.4)$$

3. Presisi kelas negatif

$$Presisi = \frac{TN}{TN + FN} \quad (3.5)$$

1. *Recall* kelas positif

$$Recall = \frac{TP}{TP + FN} \quad (3.6)$$

2. *Recall* kelas negatif

$$Recall = \frac{TN}{TN + FP} \quad (3.7)$$

Langkah selanjutnya adalah pemilihan model yang paling sesuai berdasarkan hasil evaluasi. Setelah model yang optimal dipilih, data diekspor ke dalam *file excel* menggunakan operator *write excel*. Proses ini memastikan dokumentasi yang akurat dan mudah diakses terkait hasil pengujian dan evaluasi model. Dengan demikian, fase ini merupakan tahap kritis dalam penelitian, di mana kehandalan model diukur dan hasilnya dieksplorasi dengan cermat untuk mendukung temuan dan kesimpulan penelitian secara keseluruhan.