

Implementasi Metode Support Vector Machine (SVM) Pada Klasifikasi Drop Out (DO) Mahasiswa

Rahmad Hidayat*¹, Mifzal Haris¹, Zulfan Khairil Simbolon¹

¹*Jurusan Teknologi Informasi dan Komputer Politeknik Negeri Lhokseumawe, Lhokseumawe 24351, Indonesia*

*¹*corresponding author: rahmad_hidayat@pnl.ac.id*

Abstrak

Politeknik Negeri Lhokseumawe merupakan salah satu perguruan tinggi vokasi di Indonesia yang berusaha mengurangi angka putus kuliah. Drop out (DO) di lingkungan kampus ini dapat terjadi karena beberapa faktor, seperti dua kali berturut-turut tidak lulus masa percobaan meskipun sudah mendapat peringatan dan kesempatan perbaikan, IPK yang kurang dari 2.00, tidak lulus mata kuliah tertentu di akhir semester, serta melebihi batas maksimal cuti yang diperbolehkan, yaitu dua semester. Faktor lain yang mempengaruhi adalah kehadiran yang kurang dari 50% dari total pertemuan setiap semester. Menghadapi tantangan ini, penerapan model klasifikasi berbasis machine learning, khususnya dengan algoritma Support Vector Machine (SVM), dapat membantu mengidentifikasi mahasiswa yang berpotensi DO dengan akurat. Model SVM ini mengolah dataset yang mencakup nilai IPS, IPK, tingkat kehadiran, dan jumlah cuti mahasiswa. Dalam penelitian ini, data dibagi menjadi data pelatihan dan pengujian dengan perbandingan 80:20. Hasil uji metode SVM menunjukkan akurasi sebesar 86%. Pada kelas non-drop out, precision mencapai 0.95 dan recall 0.86, sementara pada kelas drop out, precision sebesar 0.71 dan recall 0.89. Sistem klasifikasi yang diusulkan ini diharapkan mampu mengidentifikasi status potensi drop out mahasiswa secara lebih tepat, sehingga dapat mendukung upaya pencegahan DO di Politeknik Negeri Lhokseumawe.

Kata Kunci: Drop Out, Sistem Klasifikasi, SVM

Abstract

Politeknik Negeri Lhokseumawe is one of the vocational universities in Indonesia that is trying to reduce the dropout rate. Dropping out (DO) in this campus environment can occur due to several factors, such as not passing the probationary period twice in a row despite warnings and opportunities for improvement, a GPA of less than 2.00, not passing certain courses at the end of the semester, and exceeding the maximum allowable leave limit, which is two semesters. Another influencing factor is attendance that is less than 50% of the total meetings each semester. Facing this challenge, the application of machine learning-based classification models, especially with the Support Vector Machine (SVM) algorithm, can help identify students who have the potential to drop out accurately. This SVM model processes a dataset that includes social studies scores, GPA, attendance rates, and the number of student leaves. In this research, the data is divided into training and testing data with a ratio of 80:20. The SVM method test results show an accuracy of 86%. In the non-drop out class, precision reached 0.95 and recall 0.86, while in the drop out class, precision was 0.71 and recall 0.89. The proposed classification system is expected to be able to identify the potential dropout status of students more precisely, so that it can support dropout prevention efforts at Politeknik Negeri Lhokseumawe.

Keywords: Drop Out, Classification System, SVM

PENDAHULUAN

Perguruan tinggi di Indonesia menghadapi tantangan dalam mengurangi angka drop out (DO), yang berdampak negatif pada mahasiswa, lembaga pendidikan, dan masyarakat [1] [2]. DO, yang berarti tidak menyelesaikan studi, sering kali disebabkan oleh faktor-faktor seperti rendahnya kehadiran dan nilai IPK yang tidak memenuhi standar [3].

Politeknik Negeri Lhokseumawe menghadapi permasalahan DO dengan peraturan khusus yang mengatur DO berdasarkan ketentuan akademis, termasuk kegagalan dalam masa percobaan, $IPK < 2.00$, tidak lulus mata kuliah tertentu, serta batas masa studi [4] [5].

Untuk mengatasi hal ini, penelitian ini menerapkan algoritma Support Vector Machine (SVM) yang telah terbukti efektif dalam klasifikasi [6] [7] [8]. Metode SVM digunakan untuk memisahkan mahasiswa yang berisiko DO dan tidak, dengan rasio data pelatihan dan pengujian 80:20 [9]. Diharapkan, hasil penelitian ini dapat membantu institusi

mengidentifikasi mahasiswa berisiko tinggi dan mencegah DO secara dini [10] [11].

METODE PENELITIAN

Pengumpulan Data

Pengumpulan data dilakukan untuk mendapatkan informasi yang dibutuhkan dalam mencapai tujuan penelitian. Penelitian melakukan proses pengumpulan data dengan cara mengajukan permintaan data kepada UPT TIK dan administrasi setiap jurusan pada Politeknik Negeri Lhokseumawe mengenai faktor-faktor penyebab drop out. Data yang digunakan dalam penelitian ini merujuk pada Peraturan Direktur Politeknik Negeri Lhokseumawe Nomor 1 Tahun 2022, bagian ke-6 tentang putus studi, Pasal 40 poin 2.

Analisis Kebutuhan Data

Analisis kebutuhan data pada penelitian ini terdiri dari analisis data absensi kehadiran mahasiswa yang berasal dari rekapan administrasi setiap jurusan di Politeknik Negeri Lhokseumawe, serta data Indeks Prestasi Semester (IPS), Indeks Prestasi Kumulatif (IPK), dan jumlah cuti mahasiswa yang diperoleh dari Sistem Informasi Akademik (SIKAD) Politeknik Negeri Lhokseumawe. Faktor mahasiswa di-drop out adalah dua kali berturut-turut tidak lulus percobaan pada setiap akhir semester, $IP < 2,00$, tidak lulus pada setiap akhir semester, atau melewati batas akhir masa studi.

Analisis Kebutuhan Fungsional

Kebutuhan fungsional pada sistem hanya melibatkan satu entitas eksternal, yaitu operator. Operator dapat melakukan login, input data dan klasifikasi, lihat hasil klasifikasi, dan log out.

Metode Support Vector Machine (SVM)

Support Vector Machine adalah salah satu algoritma dalam bidang Machine Learning yang menggunakan ruang hipotesis berupa fungsi-fungsi linier dalam sebuah ruang fitur (feature space) yang berdimensi tinggi dan mengimplementasikan learning bias yang berasal dari teori pembelajaran statistik yang dilatih dengan algoritma pembelajaran. Teori yang mendasari SVM telah berkembang sejak tahun 1960-an, tetapi baru diperkenalkan oleh Vapnik, Boser dan Guyon pada tahun 1992 [8].

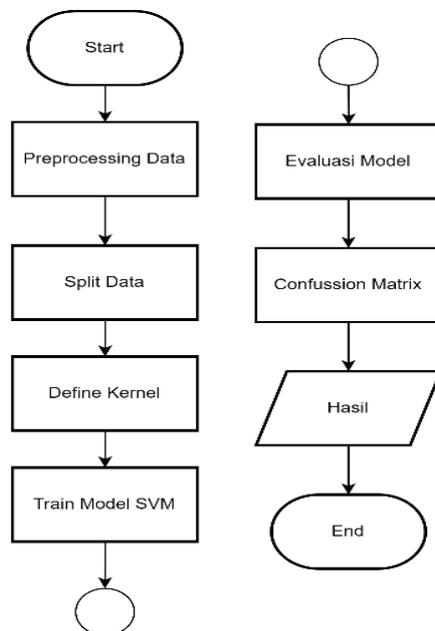
Langkah-langkah untuk melakukan proses klasifikasi dengan metode Support Vector Machine (SVM) adalah sebagai berikut:

1. Langkah awal dalam proses ini adalah melakukan preprocessing data, yang mencakup pembersihan data seperti menangani data yang hilang, mengubah data kategorikal menjadi numerik, serta melakukan normalisasi atau standarisasi fitur jika diperlukan. Tujuan utama dari preprocessing adalah memastikan data berada dalam format yang bersih dan siap digunakan oleh algoritma machine learning, sehingga model dapat mempelajari pola yang relevan tanpa terganggu oleh data yang tidak terstruktur atau tidak relevan. Dalam penelitian ini, parameter data yang digunakan meliputi IPS (Indeks Prestasi Semester), IPK (Indeks Prestasi Kumulatif), jumlah cuti, dan jumlah kehadiran, yang kesemuanya penting dalam analisis lebih lanjut.
2. Split data. Setelah preprocessing, data dibagi menjadi dua set, yaitu data training dan data testing. Data training digunakan untuk melatih model, sementara data testing digunakan untuk mengukur kinerja model. Pembagian ini biasanya dilakukan dengan rasio seperti 70:30 atau 80:20, di mana sebagian besar data dialokasikan untuk pelatihan dan sebagian kecil untuk pengujian. Pembagian ini sangat penting untuk menghindari overfitting, yaitu kondisi di mana model terlalu baik dalam mengingat data pelatihan tetapi gagal melakukan generalisasi pada data baru.
3. Train model SVM. Model Support Vector Machine (SVM) dilatih menggunakan data pelatihan, di mana algoritma SVM mempelajari batasan keputusan (decision boundary) antara dua kelas, "DO" dan "Tidak DO". Proses ini berfokus pada pencarian hyperplane optimal yang memisahkan kedua kelas dengan margin terbesar, sehingga model dapat memprediksi dengan lebih akurat saat diterapkan pada data baru. Dengan demikian, SVM dapat membedakan dua kelas secara efektif dan menghindari kesalahan prediksi pada data yang belum pernah dilihat sebelumnya.
4. Evaluasi Model. Setelah model dilatih, kinerjanya dievaluasi menggunakan data testing untuk mengukur seberapa baik model mengklasifikasikan data yang belum pernah dilihat sebelumnya. Evaluasi matrix seperti akurasi, precision, recall, dan F1-score digunakan untuk menilai kinerja model. Akurasi menunjukkan persentase prediksi yang benar dari total prediksi, precision mengukur seberapa banyak prediksi positif (DO) yang benar, dan recall menunjukkan seberapa banyak kasus positif (DO) yang berhasil dideteksi oleh model. F1-score menggabungkan precision dan recall dalam satu metrik yang seimbang untuk memberikan evaluasi yang lebih komprehensif.
5. Klasifikasi data baru. Setelah model berhasil dievaluasi, langkah selanjutnya adalah menggunakannya untuk mengklasifikasikan data baru yang belum pernah dihadapi sebelumnya. Model memprediksi apakah mahasiswa pada data baru tersebut akan "DO" atau "Tidak DO." Hasil klasifikasi ini berupa keluaran yang menunjukkan apakah data baru termasuk dalam kategori "DO" (positif) atau "Tidak DO" (negatif), sehingga model dapat digunakan untuk membuat keputusan berdasarkan data yang belum pernah dilihat.
6. Hasil akhir. Proses ini diakhiri dengan tampilan hasil klasifikasi dari model, yang menunjukkan apakah model memprediksi seorang mahasiswa akan drop out atau tidak. Berdasarkan hasil klasifikasi ini, pengguna dapat mengambil keputusan atau tindakan lebih lanjut, seperti melakukan intervensi untuk mahasiswa yang diprediksi akan drop out, menggunakan hasil klasifikasi dari model SVM sebagai dasar pertimbangan.

Rancangan Flowchart

Flowchart pada Gambar 1 menggambarkan proses klasifikasi menggunakan metode SVM. Proses dimulai dengan preprocessing data lalu membagi data menjadi dua bagian yaitu data training dan data testing. Data pelatihan digunakan untuk membangun model, sementara data pengujian digunakan untuk mengevaluasi kinerja model.

Model SVM kemudian dilatih menggunakan data pelatihan. Model ini belajar dari data untuk mengklasifikasikan data DO. Setelah itu, model yang telah dilatih dievaluasi menggunakan data pengujian untuk mengukur kinerjanya, dengan meTI evaluasi seperti akurasi, presisi, recall, dan F1-score.



Gambar 1. Metode SVM

Model yang telah dievaluasi kemudian digunakan untuk mengklasifikasi data DO dari data baru. Hasil klasifikasi ini ditampilkan, menunjukkan apakah data DO dari data baru bersifat positif atau negatif. Proses ini diakhiri dengan tampilan hasil akhir klasifikasi, yang menandakan selesainya klasifikasi drop out menggunakan metode SVM.

HASIL DAN PEMBAHASAN

Rancangan Flowchart

Pada Tabel 1 menunjukkan data yang telah melewati proses preprocessing untuk program studi Teknologi Informatika (TI).

Tabel 1. Hasil Data Preprocessing

Nama	Prodi	NIM	Jenjang	IP	IPK	Jumlah Cuti	Jumlah Semester	Jumlah Kehadiran
M. AL ISRAQ SYARIF	TI	1520301001	D4	3,2	3	2	8	32
Rahima Diniati	TI	1520301002	D4	2,63	2,87	0	8	45
Salamah	TI	1520301003	D4	2,79	2,91	0	8	41
ITA ZAHARA	TI	1520301005	D4	2,89	2,89	0	8	33
Nolarita	TI	1520301006	D4	2,95	3,23	0	8	69
Rinaldi	TI	1520301007	D4	2,74	2,92	0	8	74
Putriani	TI	1520301008	D4	2,78	2,78	0	8	93
RAJA MULKAN AZHARI	TI	1520301009	D4	2,79	2,91	0	8	51
Rohimah	TI	1520301010	D4	2,95	3,13	0	8	68
Wayuda	TI	1520301011	D4	3	3,21	1	8	61
Salman Al Farisi	TI	1520301012	D4	4	3,54	0	8	80
RACHMAT FARHAN	TI	1520301015	D4	2,89	3,24	0	8	62
Nurul Hasanah	TI	1520301016	D4	3,81	3,84	0	8	71
Ilham Akbal	TI	1520301017	D4	3,26	3,46	0	8	87

Tabel 1 menunjukkan data yang sudah siap untuk digunakan, dengan variabel yang akan digunakan pada penelitian ini telah terpenuhi, seperti nama, program studi, NIM, jenjang, IPS, IPK, jumlah cuti, jumlah semester, dan jumlah kehadiran. Data yang ditampilkan merupakan data dari Program Studi TI, yang merupakan kepanjangan dari Teknologi Rekayasa Instrumentasi Kontrol.

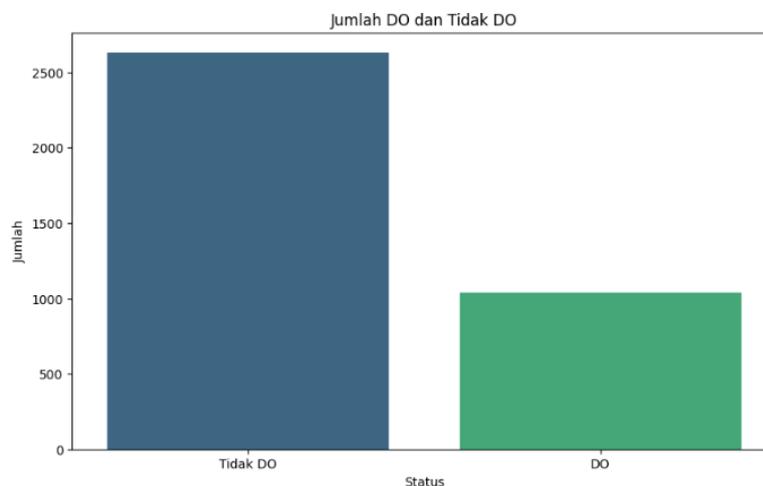
Hasil Labelling Data Drop Out

Pada Tabel 2 menunjukkan data yang telah dilakukan proses labelling. Tabel 2 menunjukkan telah terdapat status atau label dari setiap data yang sebelumnya telah melewati proses preprocessing.

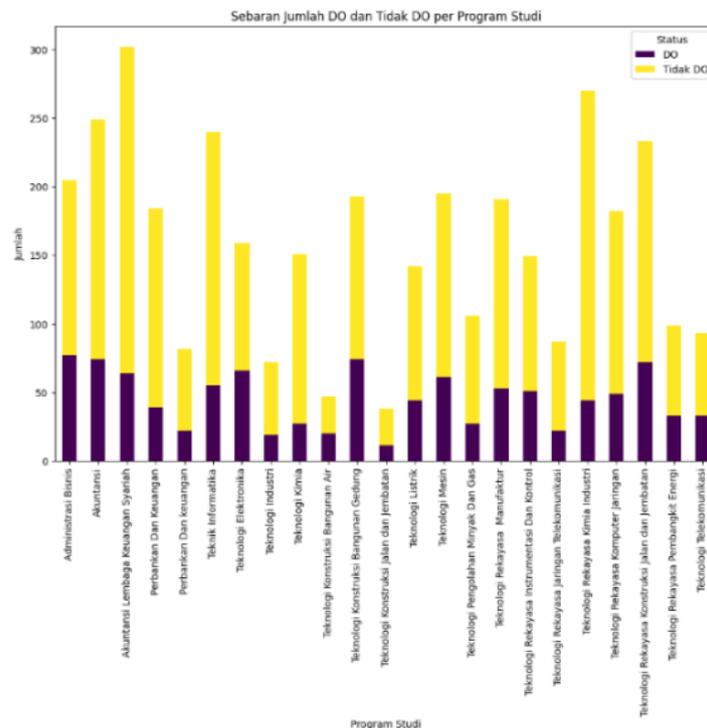
Tabel 2. Hasil Labelling Data

Nama	Prodi	NIM	Jenjang	IP	IPK	Jumlah Cuti	Jumlah Semester	Jumlah Kehadiran	Status
M.Al Israq Syarif	TI	1520301001	D4	3,2	3	2	8	32	DO
Rahima Diniati	TI	1520301002	D4	2,63	2,87	0	8	45	Tidak DO
Salamah	TI	1520301003	D4	3,0	2,2	2	8	40	DO
Ita Zahara	TI	1520301005	D4	2,7	2,5	2	8	35	DO
Nolarita	TI	1520301006	D4	2,6	2,1	2	8	37	DO
Rinaldi	TI	1520301007	D4	2,74	2,92	0	8	74	Tidak DO
Putriani	TI	1520301008	D4	2,78	2,78	0	8	93	Tidak DO
Raja M Azhari	TI	1520301009	D4	2,79	2,91	0	8	51	Tidak DO
Rohimah	TI	1520301010	D4	1,95	2,13	0	8	68	DO
Wayuda	TI	1520301011	D4	2,0	2,21	1	8	61	DO
Salman Al Farisi	TI	1520301012	D4	4	3,54	0	8	80	Tidak DO
Rachmat Farhan	TI	1520301015	D4	2,89	3,24	0	8	62	Tidak DO
Nurul Hasanah	TI	1520301016	D4	3,81	3,84	0	8	71	Tidak DO
Ilham Akbal	TI	1520301017	D4	3,26	3,46	0	8	87	Tidak DO

Label pada Tabel 2 adalah positif dan negatif, pada kasus penelitian ini berupa status DO atau Tidak DO.



a. Visualisasi Labelling Data DO/Tidak DO Umum



b. Visualisasi Data DO/Tidak DO Jurusan

Gambar 2. Visualisasi Labelling Data DO/Tidak DO umum dan jurusan

Gambar 2(a) memperlihatkan perbandingan jumlah mahasiswa yang DO (Drop Out) dan tidak DO di seluruh program studi. Terlihat bahwa jumlah mahasiswa yang tidak DO jauh lebih tinggi dibandingkan dengan mahasiswa yang DO, dengan selisih yang signifikan. Ini menunjukkan bahwa sebagian besar mahasiswa berhasil menyelesaikan studi tanpa mengalami DO.

Gambar 2(b) menampilkan distribusi jumlah mahasiswa yang DO dan tidak DO per program studi setelah proses labelling data. Berdasarkan visualisasi ini, terlihat bahwa Program Studi Tata Niaga memiliki jumlah mahasiswa yang DO terbanyak, diikuti oleh Program Studi Teknik Sipil. Sementara itu, program studi lainnya memiliki jumlah mahasiswa DO yang lebih sedikit. Program studi dengan jumlah mahasiswa tidak DO tertinggi terlihat pada beberapa bidang seperti Administrasi Bisnis dan Akuntansi. Visualisasi ini memberikan gambaran yang jelas mengenai program studi dengan tingkat DO yang lebih tinggi, yang dapat menjadi fokus perbaikan untuk menurunkan angka DO pada program studi tersebut.

Implementasi Metode SVM

```
import pandas as pd

from sklearn.model_selection import train_test_split

from sklearn.preprocessing import StandardScaler

from sklearn.svm import SVC

from sklearn.metrics import accuracy_score, classification_report, confusion_matrix
```

Gambar 3. Import Library

Gambar 3 merupakan proses import library yang dilakukan sebelum melakukan klasifikasi menggunakan metode SVM.

```
# Split the data into training and testing sets

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2,
random_state=42)
```

Gambar 4. Split Data

Gambar 4 dilakukan proses splitting data pada data yang sudah di label, data latih sebanyak 80% dan data uji sebanyak 20% "test_size = 0.2".

```

pipeline = Pipeline([
    ('scaler', StandardScaler()),
    ('svm', SVC(kernel='linear', class_weight='balanced', random_state=42))
])

# Train the model
pipeline.fit(X_train, y_train)

# Save the model to a file
joblib.dump(pipeline, 'svm_model.pkl')

# Predict on the test data
y_pred = pipeline.predict(X_test)

# Evaluate the model
print("Accuracy:", accuracy_score(y_test, y_pred))

```

Gambar 5. Proses Klasifikasi SVM

Gambar 5 dilakukan penginisialisasian model SVM dengan menggunakan kernel Linear, Setelah itu model akan dilatih dan dilakukan klasifikasi.

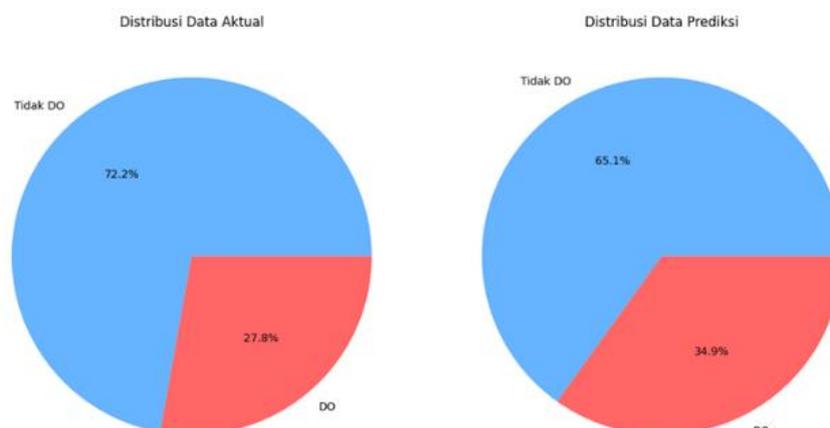
```

Class distribution in the dataset:
  Status
0    2630
1    1039
Name: count, dtype: int64
Class distribution in the training set:
  Status
0    2100
1     835
Name: count, dtype: int64
Class distribution in the test set:
  Status
0     530
1     204
Name: count, dtype: int64
Accuracy: 0.86928988092643851

```

Gambar 6. Distribusi Kelas Dataset

Gambar 6 menampilkan distribusi kelas dalam dataset secara keseluruhan serta pembagian data antara set pelatihan dan set pengujian. Dalam dataset utama, terdapat dua kelas yang diwakili oleh kolom "Status", yaitu "0" dan "1", dengan masing-masing jumlah data sebanyak 2630 dan 1039. Saat data dibagi menjadi set pelatihan dan set pengujian, distribusi ini tetap dipertahankan. Pada set pelatihan, terdapat 2100 data dengan status "0" dan 835 data dengan status "1", sementara pada set pengujian, terdapat 530 data dengan status "0" dan 204 data dengan status "1". Pembagian ini menunjukkan bahwa proporsi antara kelas "0" dan "1" dipertahankan dalam kedua set, memastikan bahwa model yang dilatih akan memiliki eksposur yang proporsional terhadap setiap kelas, yang penting untuk akurasi dalam prediksi. Hasil dari proses Support Vector Machine (SVM) menunjukkan akurasi sebesar 86%, yang mengindikasikan bahwa model ini mampu memprediksi dengan cukup baik terhadap data uji yang diberikan.



Gambar 7. Actual and Prediction

Pada Gambar 7 menampilkan visualisasi data setelah melakukan klasifikasi menggunakan metode SVM. Terdapat 2 buah diagram lingkaran, yaitu distribusi data aktual dan data prediksi. Pada distribusi aktual menjelaskan bahwa jumlah data Tidak DO memiliki persentase sebanyak 72.2 %, jumlah data DO memiliki persentase sebanyak 27.8%. Sedangkan untuk distribusi prediksi menjelaskan bahwa jumlah data Tidak DO memiliki persentase sebanyak 65.1%, jumlah data DO memiliki persentase 34,9%.

Pengujian Confusion Matrix

Gambar 8 merupakan pengujian sistem menggunakan confusion matrix.

```

pipeline = Pipeline([
    ('scaler', StandardScaler()),
    ('svm', SVC(kernel='linear', class_weight='balanced', random_state=42))
])

# Train the model
pipeline.fit(X_train, y_train)

# Save the model to a file
joblib.dump(pipeline, 'svm_model.pkl')

# Predict on the test data
y_pred = pipeline.predict(X_test)

# Evaluate the model
print("Accuracy:", accuracy_score(y_test, y_pred))
print("Confusion Matrix:\n", confusion_matrix(y_test, y_pred))
print("Classification Report:\n", classification_report(y_test, y_pred,
zero_division=1))

```

Gambar 8. Pengujian Confusion Matrix

Pada Gambar 8 proses pengujian Confusion Matrix dilakukan dengan menghitung hasil akurasi dan report dengan menggunakan variabel `y_test` dan prediction. Hasil dari pengujian confusion matrix dapat dilihat pada gambar 9.

```

Accuracy: 0.8692098092643051
Confusion Matrix:
[[456 74]
 [ 22 182]]
Classification Report:

```

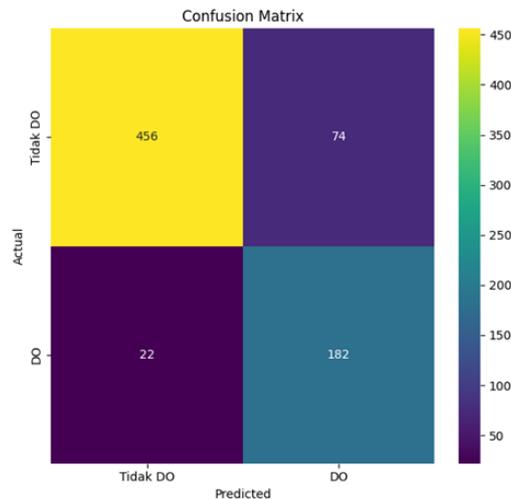
	precision	recall	f1-score	support
0	0.95	0.86	0.90	530
1	0.71	0.89	0.79	204
accuracy			0.87	734
macro avg	0.83	0.88	0.85	734
weighted avg	0.89	0.87	0.87	734

Gambar 9. Hasil Evaluasi Model Klasifikasi

Gambar 11 merupakan hasil evaluasi model klasifikasi yang ditampilkan menunjukkan bahwa model memiliki tingkat akurasi sebesar 86,92%, yang berarti sekitar 87% prediksi yang dihasilkan oleh model sudah benar. Confusion matrix memperlihatkan bahwa model mampu memprediksi kelas 0 dengan benar sebanyak 456 kali dan salah sebanyak 74 kali, sedangkan untuk kelas 1, model memprediksi dengan benar sebanyak 182 kali dan melakukan kesalahan sebanyak 22 kali.

Dalam laporan klasifikasi, terlihat bahwa untuk kelas 0, model memiliki precision sebesar 95%, yang berarti 95% dari semua prediksi kelas 0 adalah benar. Recall untuk kelas ini adalah 86%, menunjukkan bahwa dari semua sampel aktual kelas 0, 86% berhasil diidentifikasi dengan benar. F1-score untuk kelas ini adalah 0,90, yang menggabungkan precision dan recall. Sebaliknya, untuk kelas 1, precision model adalah 71%, yang berarti model sering kali salah memprediksi kelas 1. Namun, recall untuk kelas ini cukup tinggi yaitu 89%, menunjukkan bahwa sebagian besar kelas 1 berhasil diidentifikasi dengan benar, meskipun F1-score hanya 0,79.

Secara keseluruhan, model lebih unggul dalam mengidentifikasi kelas 0 dibandingkan dengan kelas 1, namun masih menunjukkan kinerja yang baik dengan nilai akurasi total 87%. Rata-rata makro dari precision, recall, dan F1-score masing-masing adalah 0,83, 0,88, dan 0,85, sedangkan rata-rata tertimbang, yang memperhitungkan jumlah sampel di setiap kelas, menunjukkan performa keseluruhan model yang seimbang.



Gambar 10. Representasi Grafik Matrix

Gambar 10 adalah representasi grafis dari confusion matrix yang digunakan untuk mengevaluasi performa sebuah model klasifikasi, dalam hal ini terkait dengan prediksi "DO" dan "Tidak DO". MATIs ini membagi hasil prediksi menjadi empat kategori utama berdasarkan prediksi model terhadap data aktual.

Pada sumbu vertikal (Actual) terlihat label sebenarnya, yaitu "Tidak DO" dan "DO", sementara pada sumbu horizontal (Predicted) menunjukkan hasil prediksi model. Untuk kelas "Tidak DO", model berhasil memprediksi dengan benar sebanyak 456 sampel, yang direpresentasikan dengan warna kuning cerah. Namun, sebanyak 74 sampel yang sebenarnya "Tidak DO" salah diprediksi sebagai "DO", ditampilkan dalam warna ungu tua.

Untuk kelas "DO", model memprediksi dengan benar sebanyak 182 sampel, terlihat dengan warna biru tua. Hanya 22 sampel yang seharusnya "DO" salah diprediksi sebagai "Tidak DO", ditunjukkan oleh kotak ungu di bagian bawah kiri. Warna-warna dalam grafik ini mewakili jumlah prediksi, dengan skala warna di sisi kanan menunjukkan intensitas warna berdasarkan jumlah prediksi.

Secara keseluruhan, MATIs kebingungan ini menunjukkan bahwa model memiliki kinerja yang baik dalam memprediksi kedua kelas, terutama kelas "Tidak DO", meskipun terdapat kesalahan yang lebih banyak saat memprediksi kelas "Tidak DO" sebagai "DO".

KESIMPULAN

Berdasarkan hasil evaluasi model SVM yang dilatih menggunakan data mahasiswa di Politeknik Negeri Lhokseumawe, beberapa kesimpulan dapat diambil mengenai keakuratan model tersebut. Model SVM ini menunjukkan tingkat akurasi sebesar 86,92%, yang berarti model mampu mengklasifikasikan status drop out dengan tingkat kesalahan yang relatif rendah. Pada kelas "Tidak DO" (Tidak Drop Out), precision mencapai 0,95, menunjukkan bahwa ketika model mengklasifikasikan seorang mahasiswa tidak akan drop out, tingkat kebenarannya sangat tinggi. Recall untuk kelas ini mencapai 0,86, menandakan bahwa model dapat mendeteksi 86% dari semua kasus "Tidak DO". Sementara itu, untuk kelas "DO" (Drop Out), precision adalah 0,71 dan recall adalah 0,89. Ini berarti bahwa model cukup baik dalam mendeteksi mahasiswa yang akan drop out, meskipun terkadang masih ada kesalahan dalam mengklasifikasikannya.

Confusion matrix menggambarkan distribusi klasifikasi yang benar dan salah untuk masing-masing kelas. Dari 734 kasus yang diuji, 456 dari 530 mahasiswa yang tidak drop out terdeteksi dengan benar oleh model, sedangkan 182 dari 204 mahasiswa yang drop out juga berhasil diprediksi dengan tepat. MATI F1-score, yang menggabungkan precision dan recall, menunjukkan hasil yang baik. Untuk kelas "Tidak DO", F1-score adalah 0,90, sedangkan untuk kelas "DO" adalah 0,79, menandakan keseimbangan yang cukup antara precision dan recall di kedua kelas. Hal ini menunjukkan bahwa model SVM memiliki performa yang cukup andal dalam memprediksi status drop out mahasiswa.

REFERENSI

- [1] Ahmad Taufik, Pengantar Teknologi Informasi, Jawa Tengah: CV. PENA PERSADA, 2022.
- [2] Anbuselvan Sangodiah, "Minimizing Student Attrition In Higher Learning Institutions In Malaysia Using Support Vector Machine," *Journal of Theoretical and Applied Information Technology Vol 71 No 3 2015*, vol. LXXI, pp. 377-385, 2015.
- [3] I. S. Annisa Fitri, "Implementasi Standar Pelayanan Minimal (SPM) Penanggulangan Bencana Kebakaran Pada Pemadam Kebakaran Kabupaten Pesisir Selatan," *Jurnal Ilmiah Ekotrans & Erudisi*, vol. 2, pp. 55-65, 2022.
- [4] M. S. Barham Siregar, "Pemanfaatan SIAKAD dalam Menunjang Pelaksanaan Pendidikan serta Manfaatnya bagi Institusi dan Mahasiswa," *AFoSJ-LAS, Vol.2, No.4, 30 Desember 2022*, vol. 2, pp. 210-216, 2022.
- [5] Eliezer M Putra Sianturi, "Rancang Bangun Rest API Aplikasi Blues Untuk Mempermudah Bisnis Food & Beverage," *Prosiding Seminar Nasional Teknologi dan Sistem Informasi (SITASI) 2023*, pp. 68-77, 2023.
- [6] K. Faozi, "Optimasi Algoritma C4.5 dengan Fuzzy Inference System Mamdani dalam Memprediksi Mahasiswa Berpotensi Dropout," *Scientia Sacra: Jurnal Sains, Teknologi dan Masyarakat Vol. 2 No. 3, September 2022*, vol.

II, pp. 272-280, 2022.

- [7] Hasanuddin, "Rancang Bangun REST API Aplikasi WeShare sebagai Upaya Mempermudah Pelayanan Donasi Kemanusiaan," *JINTEKS (Jurnal Informatika Teknologi dan Sains)*, vol. 4, pp. 8-14, 2022.
- [8] Laksamana Rajendra Haidar, "Analisa Prediksi Mahasiswa Drop Out Menggunakan Metode Decision Tree Dengan Algoritma ID3 dan C4.5," *TRANSFORMATIKA, Vol.17, 2019*, vol. XIVII, pp. 97-106, 2019.
- [9] A. T. P. Prasetyo Bella Ramadhanu, "Rancang Bangun Web Service API Aplikasi Sentralisasi Produk UMKM pada UPTD PLUT KUMKM Provinsi Lampung," *Jurnal Teknologi dan Sistem Informasi (JTSI)*, vol. 2, pp. 59-64, 2021.
- [10] A. Putra, "Solusi Prediksi Mahasiswa Drop Out Pada Program Studi Sistem Informasi Fakultas Ilmu Komputer Universitas Bina Darma," *Jurnal Simetris, Vol 8 No 1 April 2017*, Vol. VIII, P. 177, 2017.
- [11] W. Rendra Nasrul Rifai, "Penggunaan Metode Cobit Framework 4.1 Dalam Mengaudit Sistem Informasi Akademik (Siakad) Pada Iain Raden Intan Lampung," *Jurnal TIM Darmajaya Vol. 01 No. 01 Mei 2015*, vol. 01, pp. 83-91, 2015.
- [12] Sartika Dewi Purba, "Prediction Of Students Drop Out With Support Vector Machine Algorithm," *Jurnal Mantik*, vol. VI, pp. 582-586, 2022.
- [13] K. D. Sistem Informasi, "Sistem Informasi Konsep Dasar," pp. 1-9, 1996.
- [14] H. S. Sonny Nugroho Aji, "Usulan Perancangan Sistem Informasi Gudang Sparepart dengan Menggunakan Metode UML (Unified Modelling Language)," *Jurnal RISTEK Fakultas Teknik Universitas Bhayangkara Jakarta Raya*, vol. 4, pp. 140-146, 2017.
- [15] Y. Surono, "Data Flow Diagram (DFD) pada Apotek Candra Kota Jambi," *Jurnal Ilmiah Universitas Batanghari Jambi Vol.14 No.4 Tahun 2014*, vol. 14, pp. 56-64, 2014.
- [16] C. U. M. U. R. Ummiy Fauziyah Laili, "Analisis Potensial Drop Out Mahasiswa Dengan K-Means++ Clustering Dalam Upaya Peningkatan Kualitas Iain Kediri," *Paedagoria : Jurnal Kajian, Penelitian dan Pengembangan Kependidikan Vol. 14, No. 2, April 2023*, vol. XIIIV, pp. 145-153, 2023.
- [17] A. I. Yunus, "Perancangan Desain User Interface dan User Experience pada Aplikasi SIAKAD dengan Menggunakan Metode User Centered Design (UCD) pada Universitas Islam Negeri Sunan Ampel Surabaya," *Tugas Akhir Fakultas Teknologi dan Informatika, Institut Bisnis dan Informatika STIKOM*, 2018.