

# Perbandingan Kinerja Algoritma Support Vector Machine dan Neural Network Backpropagation untuk Mengklasifikasikan Mahasiswa Berpotensi Dropout

Virginia Tamuntuan\*, Kusrini, Kusnawi

Pascasarjana, PJJ Informatika, Universitas Amikom Yogyakarta, Yogyakarta, Indonesia

Email: virginiaawt@students.amikom.ac.id, kusrini@amikom.ac.id, kusnawi@amikom.ac.id

Email Penulis Korespondensi: virginiaawt@students.amikom.ac.id

Submitted: 22/07/2024; Accepted: 08/09/2024; Published: 09/09/2024

**Abstrak**—Pendidikan tinggi berperan penting dalam pembentukan sumber daya manusia yang berkualitas, dengan tingkat kelulusan mahasiswa sebagai salah satu indikator kesuksesannya. Namun, tantangan yang dihadapi di STMIK Multicom yaitu kasus Dropout mahasiswa yang berdampak negatif terhadap nilai akreditasi perguruan tinggi. Penelitian ini diperlukan untuk mengetahui algoritma yang lebih cocok diterapkan pada klasifikasi mahasiswa berpotensi dropout. Tujuan dari penelitian ini adalah untuk mengetahui dan membandingkan tingkat kinerja algoritma data mining Support Vector Machine dan algoritma Neural Network Backpropagation. Hasil penelitian ini diharapkan dapat memberikan wawasan tentang efektivitas algoritma dalam memprediksi dropout dan rekomendasi praktis untuk meningkatkan strategi pencegahan dropout di STMIK Multicom. Data yang digunakan adalah data lulusan dari tahun 2019 hingga 2023 di STMIK Multicom Bolaang Mongondow. Sebanyak 80% dari total 200 data digunakan sebagai data training, sedangkan 20% sisanya digunakan sebagai data testing. K-fold cross validation dilakukan dengan K sebanyak 5. Hasil penelitian menunjukkan bahwa algoritma Support Vector Machine mencapai akurasi 80%, recall 80%, dan presisi 35%, sedangkan algoritma Neural Network Backpropagation mencapai akurasi 77%, recall 63%, dan presisi 44%.

**Kata Kunci:** *Support Vector Machine; Neural Network Backpropagation; Cross validation; Data mining*

**Abstract**—Higher education plays a crucial role in shaping high-quality human resources, with student graduation rates being a key indicator of its success. However, a significant challenge faced by STMIK Multicom is student dropout, which negatively impacts the institution's accreditation rating. This research aims to identify the most suitable algorithm for classifying students at risk of dropping out. The goal of this study is to determine and compare the performance levels of the Support Vector Machine algorithm and the Neural Network Backpropagation algorithm. The findings of this research are expected to provide insights into the effectiveness of these algorithms in predicting dropout and to offer practical recommendations for improving dropout prevention strategies at STMIK Multicom. The data used in this study comprises graduation records from 2019 to 2023 at STMIK Multicom Bolaang Mongondow. A total of 80% of the 200 data entries were used as training data, while the remaining 20% were used as testing data. K-fold cross-validation was performed with K set to 5. The results of the study indicated that the Support Vector Machine algorithm achieved an accuracy of 80%, a recall of 80%, and a precision of 35%, while the Neural Network Backpropagation algorithm achieved an accuracy of 77%, a recall of 63%, and a precision of 44%.

**Keywords:** *Support Vector Machine; Neural Network Backpropagation; Cross validation; Data mining*

## 1. PENDAHULUAN

Pendidikan tinggi sebagai investasi masa depan memiliki peran penting dalam pembentukan sumber daya manusia yang berkualitas. Salah satu indikator kesuksesan pendidikan tinggi adalah tingkat kelulusan mahasiswa. Namun, pada kenyataannya, beberapa mahasiswa cenderung menghadapi tantangan yang dapat mengakibatkan potensi dropout, menghambat kemajuan akademis mereka [1]. Namun, tantangan yang dihadapi di STMIK Multicom adalah kasus dropout mahasiswa, yang tidak hanya mempengaruhi perjalanan akademis individu mahasiswa tetapi juga berdampak negatif terhadap nilai akreditasi perguruan tinggi. Penurunan nilai akreditasi ini dapat mengurangi reputasi institusi, mempengaruhi jumlah pendaftaran mahasiswa baru, dan bahkan dapat berimplikasi pada pendanaan dan dukungan dari pemerintah serta lembaga lainnya. Oleh karena itu, penelitian ini sangat diperlukan untuk mengetahui algoritma yang lebih cocok diterapkan pada klasifikasi mahasiswa berpotensi dropout. Dengan demikian, STMIK Multicom dapat mengambil langkah-langkah proaktif untuk mengidentifikasi dan membantu mahasiswa yang berisiko dropout, serta meningkatkan strategi pencegahan dropout secara keseluruhan, yang pada akhirnya akan mendukung peningkatan nilai akreditasi dan reputasi institusi.

Data mining merujuk pada konsep yang digunakan untuk mengidentifikasi atau mengurai penemuan pengetahuan dalam sejumlah data. Ini melibatkan penerapan teknik statistik, matematika, kecerdasan buatan, dan pembelajaran mesin untuk mengekstraksi dan menemukan informasi berharga serta pengetahuan terkait dari kumpulan data yang besar. Dalam konteks pertambangan data, terdapat algoritma klasifikasi yang bertujuan untuk mengelompokkan data berdasarkan variabel target yang ada atau terdefinisi [2]. Metode SVM (Support Vector Machine) adalah suatu algoritma yang mampu mengoperasikan dengan pemetaan nonlinear. Fungsinya adalah melakukan transformasi data pelatihan awal ke dalam dimensi baru yang memiliki dimensi yang lebih tinggi. Secara sederhana, konsep SVM dapat diuraikan sebagai upaya untuk menemukan hyperplane teroptimal yang dapat secara linear memisahkan dua kelas dalam ruang input. [3]. Metode Neural Network Backpropagation adalah algoritma yang dikenal sebagai aproksimator universal untuk fungsi non-linear apa pun, dan umumnya digunakan untuk memecahkan masalah yang dapat mentolerir kesalahan dengan memiliki banyak data yang terlatih dalam kebisingan [4].

Implementasi teknik pengujian Cross Validation bertujuan untuk menilai konsistensi kinerja sistem klasifikasi dengan menerapkan teknik augmentasi peta fitur yang meningkatkan invarian rotasi pada jaringan saraf konvolusi. Pengujian ini bertujuan untuk mengukur pengaruh variasi data pelatihan terhadap kinerja sistem klasifikasi [5]. Dengan memahami Precision, Recall, dan Accuracy, dapat diidentifikasi metode yang sesuai sebagai pendekatan untuk memprediksi kemungkinan mahasiswa mendapatkan status dropout (D.O). Dataset yang akan digunakan dalam penelitian ini adalah dataset data lulusan dari tahun 2019 sampai 2023 di STMIK Multicom. Tujuan dari penelitian ini adalah untuk mengevaluasi performa algoritma data mining Support Vector Machine dan algoritma Neural Network Backpropagation. Penilaian dilakukan menggunakan metode K-Fold Cross Validation dengan jumlah 'K' yang ditetapkan sebanyak 5.

Penelitian yang menggunakan metode Support vector Machine dan Neural Network Backpropagation pernah dilakukan oleh Li dkk menjelaskan BP Neural Network dan SVM diterapkan untuk mengidentifikasi kerusakan kekencangan baut berdasarkan metode gelombang piezoelektrik. Ditemukan bahwa akurasi pengenalan BP Neural Network terhadap indeks kerusakan wavelet packet node energy change rate (WPNECR) dalam sampel "kecil" dan "besar" masing-masing sebesar 75% dan 100%. Selain itu, akurasi pengenalan BP Neural Network terhadap indeks kerusakan wavelet packet node energy entropy (WPNEE) dalam sampel "kecil" dan "besar" masing-masing sebesar 92,7% dan 100%. SVM juga menunjukkan akurasi pengenalan sebesar 100% terhadap indeks kerusakan WPNEE dalam sampel "kecil" dan "besar" [6]. Selanjutnya Penelitian oleh Dewi dkk, dijelaskan Metode Neural Network (ANN) memiliki tingkat akurasi sebesar 77,60%, sedangkan metode Support Vector Machine (SVM) memiliki tingkat akurasi sebesar 65,24% [7]. Selanjutnya dijelaskan bahwa Model Neural Network (NN) memiliki RMSE sedikit lebih tinggi, yaitu 0.048, tetapi masih memberikan prediksi yang cukup akurat. Model Support Vector Machine (SVM) memiliki RMSE tertinggi, yaitu 0.075, menunjukkan tingkat kesalahan yang lebih tinggi dalam prediksinya [8]. Berdasarkan hasil evaluasi didapatkan algoritma neural network memiliki akurasi paling baik yaitu sebesar 97%, disusul algoritma SVM sebesar 93% [9]. Dijelaskan bahwa SVM telah berhasil diterapkan dalam prediksi saham, memberikan akurasi sekitar 60%-70% untuk SVM sederhana, yang lebih ditingkatkan dengan menggabungkan metode seperti Random Forest, Algoritma Genetika menghasilkan hasil yang lebih akurat [10].

Selanjutnya penelitian yang dilakukan oleh Purnama dkk terkait metode klasifikasi Support Vector Machines (SVM) dan Backpropagation Neural Network (BPNN). Hasil penelitian menunjukkan bahwa keduanya memiliki tingkat akurasi yang hampir sama, yaitu sebesar 83,33% untuk tingkat akurasi BPNN dan 83,00% untuk tingkat akurasi SVM [11]. Kemudian penelitian oleh Gunawan dan Putra, menggunakan metode klasifikasi Multi-Class SVM, Backpropagation Neural Network, K-Nearest Neighbor, dan Naïve Bayes. Dari hasil pengujian, tingkat keberhasilan yang diperoleh untuk SVM sebesar 99.3%, dan BPNN sebesar 98.28 [12]. Selanjutnya disimpulkan bahwa baik ANN tanpa detail warna maupun SVM dengan detail warna menunjukkan kinerja yang signifikan dalam peramalan penjualan di sektor ritel pakaian. ANN tanpa detail warna menunjukkan kinerja terbaik di antara semua model, sementara SVM dengan detail warna menunjukkan peningkatan kinerja yang signifikan [13].

Selanjutnya penelitian oleh Sudianto dkk membahas penerapan algoritma Support Vector Machine (SVM) dan Multi-Layer Perceptron (MLP) pada klasifikasi topik berita online. Dataset yang digunakan berasal dari Indonesian News Corpus dengan total 10.000 dataset, masing-masing berkategori Otomotif, Bisnis Ekonomi, Nasional, Olahraga, Lifestyle, Travel, Teknologi, dan Bola. Hasil pengklasifikasian menunjukkan skor akurasi sebesar 74% pada SVM [14].

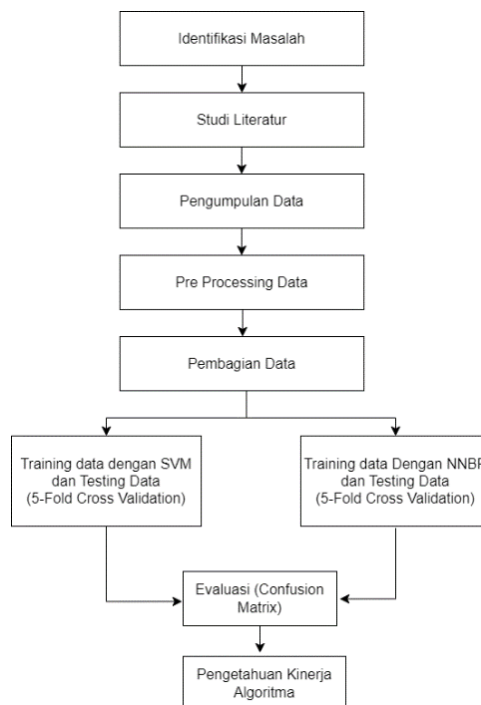
Penelitian sebelumnya telah membuktikan efektivitas berbagai metode data mining, seperti Support Vector Machine dan Neural Network Backpropagation, dalam aplikasi seperti identifikasi kerusakan, prediksi akademik, dan klasifikasi topik berita. Namun, banyak dari penelitian ini tidak secara spesifik mengeksplorasi faktor-faktor yang mempengaruhi dropout mahasiswa di institusi pendidikan tinggi, terutama di Indonesia. Selain itu, meskipun ada studi yang membandingkan akurasi antara SVM dan BPNN, masih jarang ada penelitian yang secara mendalam menilai kinerja kedua algoritma ini dalam konteks prediksi dropout mahasiswa dengan menggunakan dataset yang relevan dari institusi pendidikan tinggi lokal. Oleh karena itu, penelitian ini bertujuan untuk mengisi kekosongan ini dengan menilai performa SVM dan BPNN dalam memprediksi potensi dropout mahasiswa di STMIK Multicom menggunakan data lulusan dari tahun 2019 hingga 2023. Dengan menerapkan metode K-Fold Cross Validation, diharapkan penelitian ini dapat memberikan wawasan yang lebih mendalam mengenai efektivitas kedua algoritma tersebut dalam konteks lokal, serta memberikan rekomendasi praktis untuk meningkatkan strategi pencegahan dropout di STMIK Multicom.

## 2. METODOLOGI PENELITIAN

### 2.1 Tahapan Penelitian

Penelitian ini dilaksanakan untuk mengetahui tingkat kinerja dari algoritma SVM dan Neural Network Backpropagation. Pendekatan metodologi mencakup teknik observasi untuk mengidentifikasi masalah pada objek penelitian dan studi pustaka melalui pengumpulan dan analisis publikasi terdahulu serta referensi teori dari buku-buku yang relevan. Selanjutnya, dilakukan pengumpulan data lulusan dari tahun 2020 sampai 2023 di STMIK Multicom, pre-processing data, dan proses mining menggunakan algoritma Support Vector Machine dan Neural Network Backpropagation. Selanjutnya, dilakukan pengujian kinerja algoritma terhadap data tersebut untuk menilai hasil akhir

penelitian dengan mempertimbangkan akurasi, presisi, dan recall dari setiap pengujian algoritma. Model klasifikasi dimulai dengan mengumpulkan dataset data lulusan dari tahun 2020 sampai 2023, lalu membagi data menjadi data training dan data testing berdasarkan teknik Cross Validation dengan 5 fold Cross Validation atau 5 tahap, dengan pembagian data sebesar 20% untuk data testing dan 80% untuk data training. Evaluasi kinerja dilakukan dengan menggunakan confusion matrix. Proses data mining, pengujian, dan evaluasi menggunakan alat Google Colabs dari Google. Setelah mengevaluasi kinerja algoritma, dilakukan perbandingan akurasi, presisi, dan recall untuk menentukan algoritma yang lebih unggul dalam kasus penelitian ini. Alur penelitian dengan menggunakan metode Support Vector Machine dan Neural Network Backpropagation dapat dilihat pada Gambar 1.



**Gambar 1.** Langkah-langkah penelitian

## 2.2 Alur Klasifikasi

Proses klasifikasi dalam penelitian ini dimulai dari tahap seleksi data, preprocessing data, transformasi data, data mining, hingga evaluasi. Semua tahapan tersebut dilakukan dengan bantuan alat Google Colab. Langkah-langkahnya adalah sebagai berikut:

- a. Seleksi data  
Tahap pemilihan data untuk data mining dalam penelitian ini dilakukan melalui algoritma klasifikasi yang berperan penting dalam menentukan data yang akan digunakan.
- b. *Preprocessing* data  
Dalam penelitian ini, dilakukan pemilihan atribut yang akan digunakan, termasuk variabel dan label, yang berfungsi sebagai label target.
- c. Transformasi data  
Dalam penelitian ini, data yang telah dipilih sebelumnya disalin ke dalam format Excel yang akan digunakan untuk proses data mining menggunakan Google colab.
- d. *Data mining*  
Dalam penelitian ini, model algoritma Neural Network Backpropagation dan algoritma SVM diterapkan untuk melakukan klasifikasi terhadap data. Data ini merupakan bagian dari data training pada setiap tahap, sesuai dengan aturan k-fold cross validation.
- e. Evaluasi  
Dalam penelitian ini, dilakukan pemeriksaan hasil klasifikasi dari model algoritma data mining sesuai dengan aturan k-fold cross validation dengan  $K=5$ .

## 2.3 Model Algoritma SVM

Algoritma Support Vector Machine (SVM) merupakan metode prediksi yang menggunakan model pembelajaran dengan memanfaatkan ruang hipotesis dalam bentuk fungsi-fungsi linier di dalam ruang fitur berdimensi tinggi [15]. SVM mendasarkan diri pada prinsip dasar linear classifier, terutama pada kasus klasifikasi yang dapat dipisahkan secara linier. Meskipun demikian, SVM telah mengalami pengembangan untuk menangani problem non-linier dengan mengintegrasikan konsep kernel dalam ruang kerja berdimensi tinggi. Dalam ruang berdimensi tinggi ini, SVM

bertujuan untuk menemukan hyperplane yang dapat memaksimalkan jarak (margin) antara kelas data [16] [17]. Rumus dari algoritma Support Vector Machine adalah sebagai berikut:

$$f(x_t) = \sum_{s=1}^{ns} \alpha_1 y_s x_s \cdot x_t + b \quad (1)$$

$x^t$  merupakan data yang akan diprediksikan kelasnya (testing),  $x_s$  merupakan data support vector,  $s = 1, 2, \dots, ns$ , dan  $ns$  merupakan banyak data support vector  $ns =$  banyak data support vector.

## 2.4 Model Algoritma Neural Network Backpropagation

Neural Network Backpropagation adalah algoritma pelatihan terawasi yang menyesuaikan bobot berdasarkan kesalahan output secara mundur. Dengan tingkat pembelajaran adaptif, algoritma ini dapat meningkatkan akurasi klasifikasi [18]. Target atau keluaran referensi berfungsi sebagai peta karakter yang menunjukkan lokasi dari vektor masukan Terdapat tiga fase dalam pelatihan Neural Network Backpropagation, yaitu fase maju (feed forward), fase mundur (back propagation), dan fase modifikasi bobot [19]. Berikut merupakan langkah-langkah algoritma Neural Network Backpropagation [20] :

- Inisialisasi bobot (ambil bobot awal dengan nilai random yang cukup kecil).
- Kerjakan langkah-langkah berikut selama kondisi berhenti bernilai FALSE.
- Proses Feedforward

- Tiap-tiap unit input ( $x_i, i = 1, 2, 3, \dots, n$ ) menerima sinyal  $x_i$  dan meneruskan sinyal tersebut ke semua unit pada lapisan yang ada di atasnya (lapisan tersembunyi).

- Tiap-tiap unit tersembunyi ( $z_j, j = 1, 2, 3, \dots, p$ ) menjumlahkan sinyal-sinyal input terbobot:

$$z\_in_j = v_{0j} + \sum_{i=1}^n x_i v_{ij} \quad (1a)$$

gunakan fungsi aktivasi untuk menghitung sinyal outputnya:

$$z_j = f(z\_in_j) \quad (1b)$$

dan kirimkan sinyal tersebut ke semua unit di lapisan atasnya (unit- unit output).

Keterangan:

$z\_in_j$  = nilai unit tersembunyi ke- $j$

$v_{0j}$  = bobot layer input bias ke unit tersembunyi ke- $j$

$x_i$  = unit input ke- $i$

$v_{ij}$  = bobot unit input ke- $i$  ke layer tersembunyi ke- $j$

$z_j$  = nilai unit tersembunyi ke- $j$  menggunakan fungsi aktivasi sigmoid

$e$  = nilai konstanta = 2,718

- Tiap-tiap unit output ( $Y_k, i = 1, 2, 3, \dots, m$ ) menjumlahkan sinyal- sinyal input terbobot.

$$y\_in_k = w_{0k} + \sum_{i=1}^p z_i w_{jk} \quad (2a)$$

gunakan fungsi aktivasi untuk menghitung sinyal outputnya:

$$y_k = f(y\_in_k) \quad (2b)$$

dan kirimkan sinyal tersebut ke semua unit di lapisan atasnya (unit- unit output).

$z\_in_j$  merupakan nilai unit tersembunyi ke- $j$ ,  $v_{0j}$  merupakan bobot layer input bias ke unit tersembunyi ke- $j$ ,

$x_i$  merupakan unit input ke- $i$ ,  $v_{ij}$  adalah bobot unit input ke- $i$  ke layer tersembunyi ke- $j$ ,  $z_j$  adalah nilai unit

tersembunyi ke- $j$  menggunakan fungsi aktivasi sigmoid, dan  $e$  adalah nilai konstanta = 2,718.

- Proses Backward

- Tiap-tiap unit output ( $Y_k, i = 1, 2, 3, \dots, m$ ) menerima target pola yang berhubungan dengan pola input pembelajaran, hitung informasi errornya:

$$\delta_k = (t_k - y_k) f'(y\_in_k) \quad (3)$$

kemudian hitung koreksi bobot (yang nantinya akan digunakan untuk memperbaiki nilai  $w_{jk}$ ):

$$\Delta w_{jk} = \alpha \delta_k z_j \quad (4)$$

hitung juga koreksi bias (yang nantinya akan digunakan untuk memperbaiki nilai  $w_{0k}$ ):

$$\Delta w_{0k} = \alpha \delta_k$$

Kirimkan  $\delta_k$  ini ke unit-unit yang ada di lapisan bawahnya.

$\delta_k$  adalah nilai error unit output,  $t_k$  adalah nilai target output,  $\alpha$  adalah learning rate, dan  $\Delta w_{jk}$  adalah perubahan bobot unit tersembunyi ke- $j$  ke unit output ke- $k$ .

- Tiap-tiap unit output ( $Z_j, i = 1, 2, 3, \dots, p$ ) menjumlahkan delta inputnya (dari unit-unit yang berada pada lapisan di atasnya):

$$\delta\_in_j = \sum_{k=1}^m \delta_k w_{jk} \quad (5)$$

kalikan nilai ini dengan turunan dari fungsi aktivasinya untuk menghitung informasi error:

$$\delta_j = \delta\_in_j f'(z\_in_j) \quad (6)$$

kemudian hitung koreksi bobot (yang nantinya akan digunakan untuk memperbaiki nilai  $v_{ij}$ ):

$$\Delta v_{jk} = \alpha \delta_k x_i \quad (7)$$

hitung juga koreksi bias (yang nantinya akan digunakan untuk memperbaiki nilai  $v_{0j}$ ):

$$\Delta v_{0j} = \alpha \delta_j$$

$\delta_j$  adalah nilai error unit tersembunyi dan  $\Delta v_{jk}$  adalah perubahan bobot unit input ke-i ke unit tersembunyi ke-k

e. Update bobot dan bias:

1. Tiap-tiap unit output ( $Y_k, i = 1,2,3, \dots, m$ ) memperbaiki bias dan bobotnya ( $j=0,1,2,\dots,p$ ):  
 $w_{jk}(\text{baru}) = w_{jk}(\text{lama}) + \Delta w_{jk}$  (8)

2. Tiap-tiap unit output ( $Z_j, i = 1,2,3, \dots, p$ ) memperbaiki bias dan bobotnya ( $i=0,1,2,\dots,n$ ):  
 $v_{ij}(\text{baru}) = v_{ij}(\text{lama}) + \Delta v_{ij}$  (9)

3. Periksa kondisi berhenti.

### 2.5 Alur Validasi Pengujian

Cross validation atau estimasi rotasi adalah suatu teknik validasi model yang digunakan untuk mengevaluasi dan memahami hasil statistik analisis yang akan menggeneralisasi himpunan data independen. Pada penelitian ini menggunakan cross validation dengan  $K=5$  yaitu dataset akan dibagi menjadi 5 subset data. Pada tahap pertama subset 1 akan dijadikan sebagai data testing sedangkan subset lain menjadi data training. Proses tersebut berlanjut sampai pada penggunaan subset 5 sebagai data testing. Ilustrasi pembagian subset data dapat dilihat pada gambar 2.

s1	s2	s3	s4	s5
s1	s2	s3	s4	s5
s1	s2	s3	s4	s5
s1	s2	s3	s4	s5
s1	s2	s3	s4	s5

Gambar 2. Ilustrasi subset data

Berdasarkan Gambar 2, pada tahap pertama subset 1 sebanyak 20% data dari total data (s1) menjadi data testing sedangkan s2 sampai s5 menjadi data training. Pada tahap kedua subset 2 (s2) menjadi data testing sedangkan s3 sampai s5 dan s1 menjadi data training. Selanjutnya pada tahap ketiga subset 3 (s3) menjadi data testing sedangkan s4, s5, s1, dan s2, menjadi data training. Selanjutnya pada tahap keempat subset 4 (s4) menjadi data testing sedangkan s5, s1 sampai s3 menjadi data training. Terakhir pada tahap kelima subset 5 (s5) menjadi data testing sedangkan s1 sampai s4 menjadi data training. Kemudian untuk hasil yang akan diambil adalah rata-rata tingkat akurasi, presisi, dan recall pada seluruh tahap.

## 3. HASIL DAN PEMBAHASAN

### 3.1 Pengumpulan Data

Data yang digunakan adalah data primer yaitu data yang didapatkan dari operator kampus STMIK Multicom Bolaang Mongondow yaitu data lulusan 2019 sampai 2023 dengan total 200 record data, 7 variabel pilihan, dan 1 label target. Dataset Lulusan STMIK Multicom dapat dilihat pada Tabel 1.

Tabel 1. Dataset lulusan STMIK Multicom 2019-2023

No	NIM	Nama	Program Studi	JK	Status Pekerjaan	Pendapatan Ortu	Tempat Tinggal	Jarak Tempat Tinggal	Lulusan	Status
1	NIM001	Mhs1	S1 Teknik Informatika	L	BEKERJA	4.000.000 s.d 5.000.000	rumah sendiri	< 5 KM	SMK	Lulus
2	NIM002	Mhs2	S1 Teknik Informatika	L	BEKERJA	1.000.000 s.d 2.000.000	kos	< 5 KM	SMK	Lulus
3	NIM003	Mhs3	S1 Teknik Informatika	P	BEKERJA	1.000.000 s.d 2.000.000	rumah sendiri	5-10 KM	SMK	DROP OUT
4	NIM004	Mhs4	S1 Sistem Informasi	L	BEKERJA	2.000.000 s.d 3.000.000	rumah sendiri	< 5 KM	SMK	Lulus



5	NIM005	Mhs5	S1 Sistem Informasi	P	TIDAK	3.000.000 s.d 4.000.000	Kos	< 5 KM	SMK	Lulus
6	NIM006	Mhs6	S1 Teknik Informatika	L	BEKERJA	3.000.000 s.d 4.000.000	rumah sendiri	< 5 KM	SMK	Lulus
7	NIM007	Mhs7	S1 Sistem Informasi	L	TIDAK	3.000.000 s.d 4.000.000	rumah sendiri	< 5 KM	SMK	Lulus
8	NIM008	Mhs8	S1 Teknik Informatika	L	BEKERJA	< 1.000.000	rumah sendiri	< 5 KM	SMK	Lulus
9	NIM009	Mhs9	S1 Sistem Informasi	L	TIDAK	1.000.000 s.d 2.000.000	rumah sendiri	>10 KM	SMA	Lulus
10	NIM010	Mhs10	S1 Sistem Informasi	P	TIDAK	3.000.000 s.d 4.000.000	Kos	<5 KM	SMA	DROPOUT
...	...	...	...	...	...	...	...	...	...	...
199	NIM199	Mhs199	S1 Teknik Informatika	P	TIDAK	>5.000.000	rumah sendiri	>10 KM	SMA	Lulus
200	NIM200	Mhs200	S1 Teknik Informatika	P	TIDAK	3.000.000 s.d 4.000.000	rumah sendiri	5-10 KM	SMA	Lulus

### 3.2 Seleksi Data

Pada tahap ini dilakukan proses pemilihan variabel atribut yang akan digunakan dalam proses *data mining* menggunakan algoritma *SVM* dan *Neural Network Backpropagation*. Bentuk data hasil seleksi dapat dilihat pada Tabel 2.

**Tabel 2.** Hasil seleksi

Program Studi	JK	Status Pekerjaan	Pendapatan Ortu	Tempat Tinggal	Jarak Tempat Tinggal	Lulusan	Status
S1 Teknik Informatika	L	BEKERJA	4.000.000 s.d 5.000.000	rumah sendiri	< 5 KM	SMK	Lulus
S1 Teknik Informatika	L	BEKERJA	1.000.000 s.d 2.000.000	kos	< 5 KM	SMK	Lulus
S1 Teknik Informatika	P	BEKERJA	1.000.000 s.d 2.000.000	rumah sendiri	5-10 KM	SMK	DROP OUT
S1 Sistem Informasi	L	BEKERJA	2.000.000 s.d 3.000.000	rumah sendiri	< 5 KM	SMK	Lulus
S1 Sistem Informasi	P	TIDAK	3.000.000 s.d 4.000.000	kos	< 5 KM	SMK	Lulus
S1 Teknik Informatika	L	BEKERJA	3.000.000 s.d 4.000.000	rumah sendiri	< 5 KM	SMK	Lulus
S1 Sistem Informasi	L	TIDAK	3.000.000 s.d 4.000.000	rumah sendiri	< 5 KM	SMK	Lulus
S1 Teknik Informatika	L	BEKERJA	< 1.000.000	rumah sendiri	< 5 KM	SMK	Lulus
S1 Sistem Informasi	L	TIDAK	1.000.000 s.d 2.000.000	rumah sendiri	>10 KM	SMA	Lulus
S1 Sistem Informasi	P	TIDAK	3.000.000 s.d 4.000.000	Kos	<5 KM	SMA	DROPOUT
...	...	...	...	...	...	...	...
S1 Teknik Informatika	P	TIDAK	>5.000.000	rumah sendiri	>10 KM	SMA	Lulus
S1 Teknik Informatika	P	TIDAK	3.000.000 s.d 4.000.000	rumah sendiri	5-10 KM	SMA	Lulus

### 3.3 Preprocessing Data

#### 3.3.1 Transformasi Data

Pada tahap ini data dipastikan menggunakan format file *Microsoft excel* agar data dapat diproses sesuai dengan algoritma *SVM* dan *Neural Network Backpropagation* dengan bantuan *google colabs* dalam proses training dan testing data.

#### 3.3.2 Data Reduction

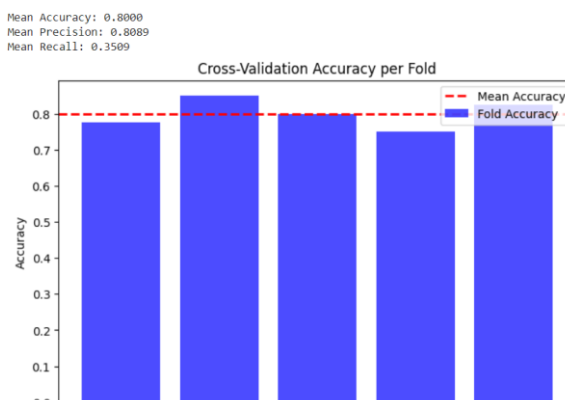
Data yang telah melalui proses seleksi dan transformasi selanjutnya akan melalui proses reduksi data. Pada tahap ini data pada variabel program studi Teknik informatika diisi dengan angka 0 sedangkan Sistem Informasi diisi dengan angka 1. Variabel JK atau jenis kelamin diisi dengan 1 untuk laki-laki dan 0 untuk perempuan. Sedangkan pada variabel status pekerjaan diisi dengan angka 1 untuk data bekerja sedangkan 0 untuk tidak bekerja. Pada variabel pendapatan ortu dibagi menjadi 6 kategori yaitu 1 untuk pendapatan lebih dari 5 juta rupiah sampai yang paling terakhir yaitu 6 untuk pendapatan kurang dari 1 juta rupiah. Bentuk dari kategorisasi data dapat dilihat pada Tabel 3.

Tabel 3. Kategorisasi data

Program_Studi	JK	Status Pekerjaan	Pendapatan Ortu	Tempat Tinggal	Jarak Tempat Tinggal	Lulusan	Status
0	1	1	2	0	1	0	0
0	1	1	5	1	1	0	0
0	0	1	5	0	2	0	1
1	1	1	4	0	1	0	0
1	0	0	3	1	1	0	0

### 3.4 Algoritma Support Vector Machine (SVM)

Data yang telah selesai dipersiapkan selanjutnya akan dilakukan mining menggunakan algoritma *data mining*. Yang pertama menggunakan algoritma SVM. Proses training dan testing data menggunakan algoritma SVM menggunakan bantuan alat google colab. Proses training dan testing data dapat dilihat pada Gambar 3.

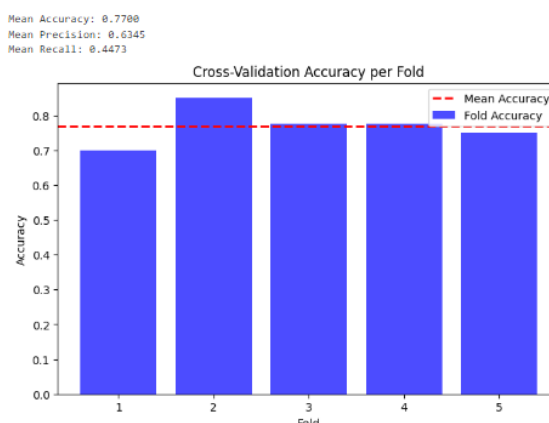


Gambar 3. Grafik pengujian algoritma SVM Cross validation K=5

Berdasarkan Gambar 3, diketahui bahwa rata-rata akurasi algoritma SVM pada data lulusan STMIK multicom untuk prediksi klasifikasi status mahasiswa berpotensi dropout yaitu: rata-rata tingkat akurasi sebesar 80%, rata-rata presisi sebesar 80%, dan rata-rata recall 35%.

### 3.5 Neural Network Backpropagation

Selanjutnya data lulusan tersebut akan digunakan untuk training dan testing menggunakan algoritma Neural Network Backpropagation. Proses training dan testing dengan menggunakan algoritma Neural Network Backpropagation dapat dilihat pada Gambar 4.



Gambar 4. Grafik pengujian algoritma Neural Network Backpropagation Cross validation K=5

Berdasarkan Gambar 4, diketahui bahwa rata-rata akurasi algoritma Neural Network Backpropagation pada data lulusan STMIK multicom untuk prediksi klasifikasi status mahasiswa berpotensi dropout yaitu: rata-rata tingkat akurasi sebesar 77%, rata-rata presisi sebesar 63%, dan rata-rata recall 44%.

Berdasarkan hasil implementasi kedua algoritma tersebut, maka didapatkan hasil perbandingan algoritma yang dapat dilihat pada Tabel 4.

**Tabel 4.** Hasil perbandingan

No	Algoritma	Akurasi	Kinerja Recall	Presisi
1	Support Vector Machine (SVM)	80%	80%	35%
2	Neural Network Backpropagation	77%	63%	44%

Berdasarkan hasil perbandingan pada Tabel 4, maka dapat diketahui tingkat akurasi algoritma SVM sebesar 80%, sedangkan algoritma NNBP sebesar 77%, tingkat recall dari algoritma SVM sebesar 80%, sedangkan NNBP sebesar 63%, tingkat Presisi algoritma SVM sebesar 35%, sedangkan NNBP sebesar 44%.

#### 4. KESIMPULAN

Berdasarkan hasil penelitian yang telah didokumentasikan pada bagian Hasil dan Pembahasan, dapat disimpulkan bahwa tingkat kinerja algoritma data mining pada data publik lulusan STMIK Multicom yang diuji dan divalidasi menggunakan metode K-Fold Cross Validation dengan jumlah 'K' sebesar 5 menunjukkan perbedaan yang signifikan. Algoritma Support Vector Machine mencapai akurasi sebesar 80%, recall sebesar 80%, dan presisi sebesar 35%. Sementara itu, algoritma Neural Network Backpropagation mencapai akurasi sebesar 77%, recall sebesar 63%, dan presisi sebesar 44%. Hasil ini mengindikasikan bahwa Support Vector Machine memiliki akurasi dan recall yang lebih tinggi dibandingkan Neural Network Backpropagation, yang berarti Support Vector Machine lebih baik dalam mengidentifikasi semua kasus dropout. Namun, Neural Network Backpropagation menunjukkan presisi yang lebih baik, menunjukkan kemampuannya yang lebih tinggi dalam mengidentifikasi kasus dropout dengan benar dari seluruh prediksi positif yang dibuat. Penelitian ini memberikan wawasan penting tentang perbandingan kinerja kedua algoritma tersebut dalam konteks prediksi dropout mahasiswa di STMIK Multicom.

#### REFERENCES

- [1] L. Setiyani, M. Wahidin, D. Awaludin, and S. Purwani, "Analisis Prediksi Kelulusan Mahasiswa Tepat Waktu Menggunakan Metode Data Mining Naïve Bayes: Systematic Review," *Fakt. Exacta*, vol. 13, no. 1, pp. 35–43, 2020.
- [2] Kusri and E. T. Lutfi, *Algoritma data mining*. Yogyakarta: ANDI, 2009.
- [3] J. J. Purnama, H. M. Nawawi, S. Rosyida, Ridwansyah, and Risnandar, "Klasifikasi Mahasiswa HER Berbasis Algoritma SVM dan Decision Tree," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 7, no. 6, pp. 1253–1260, 2020.
- [4] S. Haykin, *Neural Networks: A Comprehensive Foundation*, 2nd ed. Upper Saddle River, NJ, USA: Prentice Hall, 1999.
- [5] D. Kumar, D. Sharma, and R. Goecke, "Feature map augmentation to improve rotation invariance in convolutional neural networks," in *Proceedings of the International Conference on Advanced Concepts for Intelligent Vision Systems*, Cham, Switzerland, Feb. 2020, pp. 348–359. doi: 10.1007/978-3-030-39252-3\_32.
- [6] L. Li, P. Chen, T. Jiang, R. Tu, and M. Xu, "Research on Loose Damage Identification of High-Strength Bolts Based on Back Propagation Neural Network and Support Vector Machine," in *International Conference on Cyber-Physical Social Intelligence (ICCSI)*, 2022, pp. 185–190.
- [7] D. D. Dewi, N. Qisthi, S. S. S. Lestari, and Z. H. S. Putri, "Perbandingan Metode Neural Network Dan Support Vector Machine Dalam Klasifikasi Diagnosa Penyakit Diabetes," *Cerdika J. Ilm. Indones.*, vol. 3, no. 9, pp. 828–839, 2023.
- [8] S. Sudriyanto, F. Syahro, and N. Fitriani, "Perbandingan Performa Model Machine Learning Support Vector Machine, Neural Network, Dan K-Nearest Neighbors Dalam Prediksi Harga Saham," *J. Adv. Res. Inform.*, vol. 2, no. 1, pp. 13–21, 2023.
- [9] K. Kristiawan and A. Widjaja, "Perbandingan Algoritma Machine Learning dalam Menilai Sebuah Lokasi Toko Ritel," *J. Tek. Inform. dan Sist. Inf.*, vol. 7, no. 1, 2021.
- [10] A. Kurani, P. Doshi, and A. Vakharia, "A Comprehensive Comparative Study of Artificial Neural Network (ANN) and Support Vector Machines (SVM) on Stock Forecasting," *Ann. Data Sci.*, vol. 10, pp. 183–208, 2023, doi: 10.1007/s40745-021-00344-x.
- [11] D. I. Purnama, R. L. Islami, L. Sari, and P. R. Sihombing, "Analisis Klasifikasi Data Tracer Study Dengan Support Vector Machine Dan Neural Network," *J. SISKOM-KB (Sistem Komput. dan Kecerdasan Buatan)*, vol. 4, no. 2, pp. 46–52, 2021.
- [12] V. A. Gunawan and L. S. A. Putra, "Perbandingan Identifikasi Penggunaan American Sign Language Menggunakan Klasifikasi Multi-Class SVM, Backpropagation Neural Network, K-Nearest Neighbor dan Naive Bayes," *TEKNIK*, vol. 42, no. 2, pp. 137–148, 2021.
- [13] İ. Güven and F. Şimşir, "Demand forecasting with color parameter in retail apparel industry using artificial neural networks (ANN) and support vector machines (SVM) methods," *Comput. Ind. Eng.*, vol. 147, p. 106678, 2020, doi: 10.1016/j.cie.2020.106678.
- [14] S. Sudianto, A. D. Sripamuji, I. R. Ramadhanti, R. R. Amalia, J. Saputra, and B. Prihatnowo, "Penerapan Algoritma Support Vector Machine Dan Multi-Layer Perceptron Pada Klasifikasi Topik Berita," *J. Nas. Pendidik. Tek. Inform. JANAPATI*, vol. 11, no. 2, pp. 84–91, 2022.
- [15] N. Cristianini and J. Shawe-Taylor, *An Introduction to Support Vector Machines and Other Kernel-based Learning Methods*.



Cambridge, UK: Cambridge University Press, 2000.

- [16] B. Santosa, *Data mining teknik pemanfaatan data untuk keperluan bisnis*. Yogyakarta: Graha Ilmu, 2007.
- [17] N.-Y. Deng and Y.-J. Tian, *Support Vector Machines: Theory, Algorithms, and Extensions*. Beijing: Science Press, 2018.
- [18] R. Jullapak and A. Thammano, “Backpropagation Neural Network with Adaptive Learning Rate for Classification,” in *Advances in Natural Computation, Fuzzy Systems and Knowledge Discovery*, vol. 153, N. Xiong, M. Li, K. Li, Z. Xiao, L. Liao, and L. Wang, Eds., in *Lecture Notes on Data Engineering and Communications Technologies*, vol. 153. , Cham: Springer, 2023, pp. 130–146. doi: 10.1007/978-3-031-15345-0\_10.
- [19] R. B. Silva, J. P. Papa, and A. X. Falcao, “A Comprehensive Guide to the Backpropagation Algorithm in Neural Networks,” *Neural Networks*, vol. 123, pp. 43–59, 2022, doi: 10.1016/j.neunet.2021.12.008.
- [20] J. J. Siang, *Jaringan syaraf tiruan dan pemrogramannya menggunakan Matlab*. Yogyakarta: Penerbit Andi, 2005.