

# Klasifikasi Buku Menggunakan Metode *Support Vector Machine* pada *Digital Library*

Dhiany Hana Amalia<sup>1</sup>, Wiyli Yustanti<sup>2</sup>

<sup>1,2</sup> Sistem Informasi, Teknik Informatika, Universitas Negeri Surabaya

<sup>1</sup>[dhianyamalial16051214019@mhs.unesa.ac.id](mailto:dhianyamalial16051214019@mhs.unesa.ac.id)

<sup>2</sup>[wiyliyustanti@unesa.ac.id](mailto:wiyliyustanti@unesa.ac.id)

**Abstrak**— Perpustakaan merupakan suatu tempat dimana memiliki berbagai koleksi buku hingga jurnal dan majalah. Namun dengan seiringnya perkembangan teknologi, kini perpustakaan dapat mengembangkan pelayanannya kedalam digital atau disebut dengan *e-library*. Hal ini sangat memudahkan pemustaka dalam pencarian data buku. Dalam penelitian ini bertujuan untuk menganalisis pada klasifikasi teks, dimana untuk mengetahui kualitas pada metode *Support Vector Machine* (SVM) dengan menggunakan kernel *Linear*. *Dataset* yang digunakan diambil dari <https://opac.unesa.ac.id/>, dimana terdiri dari sepuluh kategori buku yang disesuaikan dengan *Dewey Decimal Classification* (DDC). Setelah data terkumpul kemudian memasuki tahapan *pre-processing*, dimana data terlebih dahulu dilakukan penghapusan terhadap data ganda dan melalui *case folding*, *tokenizing*, *stopwords*. Pada penelitian ini menggunakan data 1.000 records sebagai uji data serta melakukan pengambilan data secara acak, hal ini dilakukan guna menyeimbangkan pada antar data. Selanjutnya dilakukan tahapan ekstraksi fitur dimana pada tahapan ini dilakukannya pengolahan teks menjadi numerik dengan menggunakan pembobotan pada *Term Frequency-Inverse Document Frequency* (TF-IDF). Kemudian data memasuki tahap pemodelan klasifikasi dengan menggunakan metode SVM. Dalam penggunaan metode ini terdapat empat jenis fungsi kernel, yaitu kernel *Linear*, *RBF*, *Polynomial*, dan *Sigmoid* dimana memiliki karakteristik masing-masing pada setiap fungsinya. Tujuan dari penelitian ini, yaitu untuk mengetahui pengaruh pada jumlah data latih terhadap kualitas dan efektivitas metode SVM, dimana dilakukan tiga kali pengujian pada model kombinasi data latih dan data uji, yaitu 60:40, 70:30, dan 80:20. Setelah ketiga pengujian telah memasuki tahap proses klasifikasi, maka dapat diambil kesimpulan bahwa metode SVM dalam klasifikasi teks pada *digital library* dapat digunakan dan menghasilkan akurasi tertinggi yang didapat dengan nilai 69,24% pada penggunaan kernel *Linear* dibandingkan dengan kernel lainnya, dimana memiliki nilai *precision* 71%, *recall* 61%, dan *f1-score* 64%.

**Kata Kunci**— *Digital Library*, Klasifikasi, Python, *Support Vector Machine*, *Text Mining*.

## I. PENDAHULUAN

Kemajuan teknologi kini tak diragukan lagi di era revolusi industri 4.0. Sebagian industri telah mengembangkan kemajuan teknologi tersebut dengan pengembangan sistem informasi. Salah satunya yaitu perpustakaan, dimana merupakan suatu tempat yang memiliki berbagai koleksi buku, mulai dari sekumpulan macam jurnal hingga majalah. Namun dengan berkembangnya teknologi kini perpustakaan dapat mengembangkan pelayanannya secara digital atau disebut dengan *e-library*, sehingga mampu memudahkan pemustaka

dalam mencari buku yang dicari secara online tanpa mencarinya secara manual satu per satu dari rak ke rak lainnya.

Berbagai macam koleksi buku yang ada di perpustakaan membuat pustakawan harus mengelompokkan buku sesuai dengan kebutuhan. Dalam pengelompokan tersebut dibutuhkan klasifikasi, dimana hal ini dapat membantu pustakawan dalam pengelolaan perpustakaan. Sehingga digunakanlah salah satu metode pada *text mining*, yaitu klasifikasi teks.

*Support Vector Machine* atau disebut dengan SVM merupakan salah satu model klasifikasi teks yang memiliki nilai akurasi dan durasi yang sesuai dengan data yang diolah. Kelebihan pada metode ini yaitu mampu menentukan *hyperplane* dengan memilih bidang dengan margin yang optimal, sehingga SVM mampu tergeneralisasi dan tingkat generalisasi pada SVM tidak mempengaruhi jumlah data latih dengan menentukan parameter *soft margin* serta *noise* yang dapat dikontrol, sehingga membuat parameter *soft margin* semakin besar dan semakin besar pula pinalti yang didapat pada klasifikasi apabila terdapat kesalahan sehingga mengakibatkan proses *training* yang semakin ketat [1]. Dalam penelitian ini dilakukan klasifikasi teks guna untuk mengetahui kualitas pada metode dengan menentukan hasil paling akurat untuk menentukan klasifikasi buku yang sesuai dengan pemustaka.

Penelitian mengenai klasifikasi teks sebelumnya telah dilakukan oleh ShIELA Novelia dan Brodjol Sutijo [2] yang dilakukan pada tahun 2016 dengan judul “Klasifikasi Email Spam dengan Menggunakan Metode *Support Vector Machine* dan *K-Nearest Neighbor*”. Pada penelitian ini menerapkan sistem kombinasi dengan teknik partisi data *K-Fold Cross Validation* dimana k yang digunakan adalah 10, sehingga metode *K-Nearest Neighbor* menghasilkan klasifikasi terbaik pada k=3 dengan ketepatan klasifikasi sebesar 92,28% dengan error 7,72% sedangkan metode *Support Vector Machine* menghasilkan klasifikasi terbaik pada kernel linier dengan ketepatan klasifikasi sebesar 96,6% dengan error 3,4%, sehingga membuktikan bahwa metode *Support Vector Machine* lebih unggul.

Selanjutnya yaitu penelitian yang dilakukan oleh Sindy Caesar Rusnanda Himatul Haliza, dan Anita Qoiriah yang dilakukan pada tahun 2021 dengan judul “*Predictive Maintenance* untuk Kendaraan Bermotor dengan Menggunakan *Support Vector Machine* (SVM)”[3]. Pada penelitian ini menerapkan algoritma SVM dengan menggunakan metode *one-againts-all* dan metode *one-*

*agents-one* yang dilakukan untuk *predictive maintenance* pada kendaraan dinas roda-4 dari DPU Bina Marga Jatim. Hasil dari pengujian penelitian ini mendapatkan nilai akurasi yang tinggi hingga mencapai 92,92% pada penggunaan kernel *polynomial* dan kernel *Radian Basis Function* (RBF). Akan tetapi kernel *Polynomial* menghasilkan nilai akurasi tinggi yang lebih banyak dibandingkan dengan kernel RBF pada beberapa percobaan.

Selain itu penelitian lain yang juga dilakukan oleh Arif Pratama, Randy Cahya Wihandika, dan Dian Eka Ratnawati yang dilakukan pada tahun 2018 dengan judul “Implementasi Algoritme *Support Vector Machine* (SVM) untuk Prediksi Ketepatan Waktu Kelulusan Mahasiswa”[4]. Pada penelitian ini menggunakan metode SVM dengan menerapkan kernel *Gaussian RBF*, *Polynomial*, dan *Linier* dengan kombinasi nilai parameter  $\lambda = 0,5$ , konstanta  $\gamma = 0,01$ , dan  $\epsilon$  (epsilon) = 0,001 itermax = 100, c = 1. Dibandingkan dari ketiga jenis kernel tersebut rata-rata akurasi tertinggi terdapat pada kernel *Gaussian RBF* yaitu sebesar 80,55%.

Kemudian penelitian lain juga dilakukan oleh Debby Alita, Yusra Fernando, Heni Sulistani yang dilakukan pada tahun 2020 dengan judul “Implementasi Algoritma *Multiclass SVM* Pada Opini Publik Berbahasa Indonesia di Twitter”[5]. Pada penelitian ini menggunakan metode *Multiclass SVM*, dimana dilakukan dengan metode SVM *One Againts Rest* dan SVM *One Againts One* untuk proses klasifikasi pada tiga kelas yaitu positif, negatif, dan netral. Untuk hasil *presisi*, *recall*, dan *f1-score* pada metode SVM *One Againts One* memiliki hasil yang lebih unggul sedangkan metode SVM *One Againts Rest* memiliki nilai unggul dengan perbedaan 0,06. Apabila dilakukan perbandingan dengan metode *Naïve Bayes Classifier*, maka tetap metode *Multiclass SVM* memiliki nilai lebih unggul.

Dari beberapa penelitian sebelumnya, maka penelitian ini mengambil topik “Klasifikasi Buku dengan Menggunakan Metode *Support Vector Machine* pada *Digital Library*”, sehingga pada penelitian ini dirancang untuk mengetahui kualitas kemampuan pada metode *Support Vector Machine* dalam menghasilkan suatu klasifikasi buku dengan lebih baik.

## II. METODOLOGI PENELITIAN

### A. Sumber Data

Pada penelitian ini data yang digunakan merupakan data buku yang diambil dari opac.unesa.ac.id atau *digital library* pada perpustakaan Universitas Negeri Surabaya. Dimana data berjumlah 1.000 *record* untuk dilakukan uji data dan memiliki sepuluh kategori yang disesuaikan dengan *Dewey Decimal Classification* (DDC), yaitu dengan kode :

TABEL I  
KATEGORI BUKU

Kode	Kategori
000-099	Komputer, informasi, dan referensi umum
100-199	Filsaat dan Psikologi
200-299	Agama
300-399	Ilmu Sosial
400-499	Bahasa

500-599	Sains dan Matematika
600-699	Teknologi
700-799	Kesenian dan Rekreasi
800-899	Sastra
900-999	Sejarah dan Geografi

### B. Pengolahan Data

Pada penelitian ini menerapkan salah satu metode pada data mining, yaitu dengan menggunakan *Support Vector Machine*. Berikut merupakan alur pengklasifikasian SVM yang bisa dilihat pada Gbr. 1.



Gbr. 1 Alur Pengklasifikasi SVM

#### 1) Text Preprocessing

*Text Preprocessing* merupakan salah satu kunci komponen dalam berbagai algoritma *text mining*, dimana data yang pada awalnya tidak terstruktur menjadi terstruktur. Berikut merupakan tahapan yang digunakan dalam *text preprocessing* :

##### a. Case Folding

Pada tahap ini mengubah huruf abjad pada teks yang semula berbentuk *uppercase* diubah kedalam bentuk *lowercase*. Kemudian karakter selain huruf akan dihilangkan, seperti penulisan tanda baca dan angka. Berikut merupakan contoh proses *case folding* yang dapat dilihat pada tabel II.

TABEL II  
CASE FOLDING

Sebelum	Sesudah
Inovasi Pendidikan : Pendekatan Praktik Teknologi Multimedia dan Pembelajaran Online	inovasi pendidikan pendekatan praktik teknologi multimedia dan pembelajaran online
Teknologi Beton Dari Material, Pembuatan, Ke Beton Kinerja Tinggi	teknologi beton dari material pembuatan ke beton kinerja tinggi

b. *Tokenizing*

Pada tahap ini merupakan proses yang mampu memecahkan suatu teks menjadi satu kata atau frasa. Dimana elemen yang digunakan adalah selain huruf abjad dan tanda penghubung (-). Berikut merupakan contoh proses *tokenizing* yang dapat dilihat pada tabel III.

TABEL III  
TOKENIZING

Sebelum	Sesudah
inovasi pendidikan pendekatan praktik teknologi multimedia dan pembelajaran online	inovasi pendidikan pendekatan praktik teknologi multimedia dan pembelajaran online
teknologi beton dari material pembuatan ke beton kinerja tinggi	teknologi beton dari material pembuatan ke beton kinerja tinggi

c. *Stopwords*

Pada tahap ini merupakan proses yang dilakukan setelah melakukan proses *tokenizing* di setiap teksnya. Dimana pada setiap token akan disesuaikan dengan kata-kata yang termasuk pada kata dalam *stopwords*, jika token sesuai dengan salah satu kata pada *stopwords*, maka token akan dihapus. Berikut merupakan beberapa kata yang termasuk dalam *stopwords* [6] yang dapat dilihat pada tabel IV. Dan pada tabel V merupakan contoh dari hasil proses *stopwords*.

TABEL IV  
STOPWORDS

Ada	Bagian	berakhir
Agar	Bagi	berupa
Akan	begitu	dikira
Atau	berada	hal
Harus	jadi	jangan

TABEL V  
CONTOH HASIL STOPWORDS

Sebelum	Sesudah
inovasi pendidikan pendekatan praktik teknologi multimedia dan pembelajaran online	inovasi pendidikan praktik teknologi multimedia pembelajaran online
teknologi beton dari material pembuatan ke beton kinerja tinggi	teknologi beton material pembuatan beton kinerja

2) *Ekstraksi Fitur*

Setelah melalui tahapan *text preprocessing*, maka selanjutnya yaitu melalui tahapan ekstraksi fitur, dimana pada tahapan ini perlu dilakukannya pengolahan teks menjadi numerik. Hal ini dikarenakan prinsip pada komputer yang tidak mampu mengolah data selain data numerik. Selain itu ekstraksi fitur ini digunakan untuk menggali informasi dalam mempresentasikan kata-kata sebagai vektor. Salah satu teknik yang digunakan dalam ekstraksi fitur yaitu, dengan pembobotan TF-IDF atau disebut dengan *Term Frequency-Inverse Document Frequency*. TF-IDF merupakan metode pembobotan yang digunakan untuk perhitungan jumlah kecocokan dari kata pada setiap dokumen atau disebut dengan *Term Frequency* (TF), sedangkan *Inverse Document Frequency* (IDF) adalah nilai invers dari *Document Frequency* (DF). Dan hasil akhir dari metode ini adalah matriks *output* yang meliputi kata-kata unik dan nilai-nilai yang dihasilkan pada TF-IDF dari setiap kata pada seluruh data. Berikut merupakan nilai persamaan untuk menghitung TF :

$$TF_{(t,d)} = f_{(t,d)} \quad (1)$$

Dimana :

$f_{(t,d)}$  = kemunculan pada kata  $t$  dalam dokumen  $d$

Sedangkan DF atau *Document Frequency* merupakan banyaknya dokumen yang dimiliki oleh *term t*. Berikut merupakan nilai persamaan untuk menghitung DF :

$$IDF_{(t)} = \log \left( \frac{N}{nt} \right) \quad (2)$$

Dimana :

N = jumlah dokumen

Nt = jumlah dokumen yang mengandung kata t

Sehingga TF-IDF akan memberikan bobot dari suatu kata yang diambil dari nilai TF dan nilai invers DF. Berikut merupakan nilai persamaan untuk mencari bobot menggunakan TF-IDF :

$$TF-IDF = f_{(t,d)} * \log \left( \frac{N}{nt} \right) \quad (3)$$

TF-IDF dilakukan setelah data telah melalui tahap *preprocessing*, dimana bentuk data harus berbetuk numerik. Maka untuk mengubah data tersebut digunakanlah metode TF-IDF agar data dapat dianalisis dengan menggunakan metode SVM.

### 3) Modeling (Pemodelan)

Selanjutnya yaitu memasuki pada tahap pemodelan, dimana data yang telah dibersihkan dan melalui proses *preprocessing* akan dimasukkan ke dalam model. Model yang digunakan dalam penelitian ini yaitu *Support Vector Machine* (SVM). Pada tahap ini data yang telah diolah akan dibangun dengan menggunakan model SVM, dimana proses ini membutuhkan penambahan *package*, yaitu menggunakan *library NLTK (Natural Language Toolkit)*. Kemudian membuat model SVM dan memasukkan data uji yang telah diambil sebelumnya. Sehingga data yang telah dimuat tersebut akan menampilkan hasil nilai pada setiap prediksi yang dibuat pada model tersebut. Model ini merupakan bentuk awal, dimana ditentukan tipe kernel dan nilai pada parameternya. Untuk mendapatkan hasil yang maksimal maka selanjutnya dibutuhkan untuk melakukan pelatihan pada model tersebut.

Kernel pada SVM memiliki banyak fungsi. Namun pemilihan kernel tergantung pada masalah yang dihadapi karena mengikuti dengan yang dimodelkan. Misalnya, kernel polynomial memungkinkan untuk memodelkan fitur konjungsi hingga urutan ke polynomial. *Radial Basis Function* memungkinkan untuk memilih kelompok yang berbeda dengan kernel linear dimana hanya memungkinkan untuk memilih garis (atau *hyperplane*). Pada pemetaan fungsi kernel dapat digunakan Polynomial Normalisasi, RBF, Linear, dan Sigmoid dimana digunakan berdasarkan kebutuhan aplikasi. Tetapi beberapa fungsi kernel telah terbukti bekerja dengan baik untuk berbagai macam aplikasi. Berikut penjelasan pada setiap kernel :

#### 1) Kernel Linear

*Kernel Linear merupakan fungsi kernel yang paling sederhana. Pada kernel ini inner product berupa <x.y> dan ditambahkan dengan pilihan konstanta c.*

#### 2) Kernel RBF

Kernel RBF atau disebut dengan *Radial Basis Function* merupakan kernel utama. Pada kernel ini disebut utama karena kernel RBF merupakan kernel nonlinear yang mampu memetakan sampel ke dalam ruang dimensi tertinggi dimana tidak seperti dengan kernel Linear. Kernel RBF memiliki *hyper parameter* yang lebih sedikit dibandingkan dengan kernel Polynomial dan memiliki sedikit kesulitan pada bagian perhitungan (*numerical*).

#### 3) Kernel Polynomial

Kernel Polynomial merupakan kernel non-stasioner dimana kernel ini cocok untuk masalah yang semua data latihnya telah dinormalisasi.

#### 4) Kernel Sigmoid

Kernel Sigmoid atau disebut dengan *fuzzy sigmoid kernel* merupakan fungsi kernel yang memodelkan fungsi *hyperbolic tangent* melalui *variable linguistic*.

Ketika pengklasifikasian data tidak mampu dipisahkan dengan cara *linier*, maka metode lain yang dapat digunakan Fungsi pada kernel mampu mengubah data menjadi ruang dimensi tertinggi untuk memungkinkan dilakukannya pemisahan. Fungsi kernel adalah kelas dalam algoritma yang ditujukan untuk menganalisis atau *recognition*, dimana elemen yang paling terkenal adalah SVM. *Training vectors* pada  $x_i$  dipetakan ke dalam ruang dimensi tertinggi (bisa jadi tak terhingga) oleh fungsi  $\Phi$ . Kemudian SVM menemukan *hyperplane* pemisah *linier* dengan margin yang maksimal dalam ruang dimensi tertinggi.  $C > 0$  merupakan parameter penalti.

#### 4) Evaluasi

Pada tahap ini memasuki tahapan pengujian evaluasi, dimana hal ini ditujukan untuk mengetahui kualitas serta nilai akurasi yang telah dihasilkan pada metode tersebut. Sehingga yang dihasilkan pada tahapan ini merupakan perhitungan nilai akurasi, *precision*, dan *recall* dengan menggunakan *library scikit-learn*.

## III. HASIL DAN PEMBAHASAN

### A. Data

Data yang telah dikumpulkan dari portal *Integrated Online Public Access Catalogue Unesa* ([opac.unesa.ac.id](http://opac.unesa.ac.id)) merupakan data awal, dimana terdiri dari beberapa banyak kolom yang dapat dilihat pada Gbr. 1. Kemudian data yang dibutuhkan dalam penelitian adalah kolom "*title*" dan "*classification*". Selanjutnya, data tersebut melalui tahapan *preprocessing*, dimana data terlebih dahulu telah dilakukan penghapusan pada data ganda. Pada penelitian ini menggunakan 1.000 *record* sebagai uji data. Dimana data tersebut akan dipisahkan sebagai data latih dan data uji. Penentuan data latih dan data uji dilakukan secara acak, sehingga mampu membuatnya tetap seimbang pada antar data. Kemudian dilakukan pembagian tiap proporsi antara data latih dengan data uji. Hal ini dilakukan untuk mengetahui kualitas pada metode SVM.

Untuk mengetahui pengaruh pada jumlah data latih terhadap kualitas dan efektivitas metode SVM, maka dilakukan beberapa kombinasi pada jumlah data latih dan data uji. Kombinasi pada data latih dan data uji antara lain yaitu 60:40, 70:30, 80:20, dimana dapat dilihat pemagian data pada Tabel VI.

TABEL VI  
KOMBINASI DATA LATIH DAN DATA UJI

No.	Data Latih	Data Uji
1	600	400
2	700	300
3	800	200

A	B	C	D	E	F	G	H	I	J	K	L	N	O	
1	biblio_id	gmd_id	title	sor	edition	isbn_isbn	publisher	publish_y	collation	series_titl	call_numt	language	publish_p	classificati
2	1	1	Menulis artikel dan tajuk rencana	979-3782-4	1	2004 xiv, 162 hal.; 24 cm	411 SUM	n id	1	411 SUM	n id	1	411 SUM	n id
3	2	1	Menulis artikel dan tajuk rencana	979-3782-4	1	2004 xiv, 162 hal.; 24 cm	411 SUM	n id	1	411 SUM	n id	1	411 SUM	n id
4	3	1	Dasar-dasar manajemen keuangan	979-691-3	2	2006 xii, 672 hal.; ilus.; 20x25,5 cm	15 BR1	id	2	658,15 BR1	id	2	658,15 BR1	id
5	8	1	Renaissans Islam Asia Tenggara	979-514-8	3	1999 xxii, 180 hal.; ind.; 24x29,63 A	AZ1	id	1	294,63 A	AZ1	1	294,63 A	AZ1
6	9	1	Renaissans Islam Asia Tenggara	979-514-8	3	1999 xxii, 180 hal.; ind.; 24x29,63 A	AZ1	id	1	294,63 A	AZ1	1	294,63 A	AZ1
7	10	1	Renaissans Islam Asia Tenggara	979-514-8	3	1999 xxii, 180 hal.; ind.; 24x29,63 A	AZ1	id	1	294,63 A	AZ1	1	294,63 A	AZ1
8	11	1	Renaissans Islam Asia Tenggara	979-514-8	3	1999 xxii, 180 hal.; ind.; 24x29,63 A	AZ1	id	1	294,63 A	AZ1	1	294,63 A	AZ1
9	12	1	Renaissans Islam Asia Tenggara	979-514-8	3	1999 xxii, 180 hal.; ind.; 24x29,63 A	AZ1	id	1	294,63 A	AZ1	1	294,63 A	AZ1
10	13	1	Speak out: panduan praktis dai	979-22-07	4	2004 x, 156 hal.; ilus.; 23 c 621.384	1 id	2	621.384	1 id	2	621.384	1 id	
11	14	1	Speak out: panduan praktis dai	979-22-07	4	2004 x, 156 hal.; ilus.; 23 c 621.384	1 id	2	621.384	1 id	2	621.384	1 id	
12	15	1	Speak out: panduan praktis dai	979-22-07	4	2004 x, 156 hal.; ilus.; 23 c 621.384	1 id	2	621.384	1 id	2	621.384	1 id	
13	16	1	Speak out: panduan praktis dai	979-22-07	4	2004 x, 156 hal.; ilus.; 23 c 621.384	1 id	2	621.384	1 id	2	621.384	1 id	

Gbr. 2 Data awal sebelum di komputasi

Pada kolom "classification" ini merupakan kolom yang menentukan hasil kategori buku, dimana telah disesuaikan dengan Dewey Decimal Classification (DDC) yang bisa dilihat pada tabel VII.

TABEL VII  
CONTOH KOLOM DATA YANG DIAMBIL UNTUK UJI DATA, DIMANA KODE BUKU DIAMBIL DARI TIGA ANGKA TERDEPAN YANG DISESUAIKAN DENGAN DDC

No.	Judul Buku	Klasifikasi Buku	Kode Buku	Kategori Buku
1	Menulis artikel dan tajuk rencana: panduan praktis penulis & jurnalis profesional	411 SUM	4	Bahasa
2	Renaissans Islam Asia Tenggara: sejarah wacana & kekuasaan	294.63 AZR	2	Agama

Setelah seluruh data telah mengetahui kategori pada setiap buku, maka selanjutnya disimpan dalam bentuk file berformat .csv dan kemudian data di proses pada tahapan preprocessing dimana melalui proses cleaning, case folding, tokenizing, stopwords, dan pembobotan pada TF-IDF.

### B. Modeling

Pada tahapan ini data telah memasuki tahap proses klasifikasi, dimana pada penelitian ini digunakan empat kernel untuk pengujian dengan menggunakan beberapa model kombinasi data. Hal ini ditujukan untuk mendapatkan model dengan kualitas serta nilai akurasi yang paling tinggi. Berikut merupakan beberapa hasil uji coba yang dilakukan:

### 1) Pengujian Pertama

Pada pengujian pertama ini menggunakan kombinasi data sebesar 60:40. Dimana data latih yang dimasukkan sebanyak 60% sedangkan data uji 40% dari jumlah dataset yang telah ditentukan dengan menggunakan nilai parameter yaitu  $\sigma = 3$ , nilai parameter  $C = 1$ , dan  $d = 3$ . Hasil akurasi klasifikasi pada pengujian pertama dapat dilihat pada tabel VII, dimana kernel linear lebih unggul dibandingkan kernel lainnya, yaitu sebesar 68,24 %.

TABEL VIII  
AKURASI MODEL PADA PENGUJIAN PERTAMA

Linear	RBF	Polynomial	Sigmoid
68,24%	66,57%	55,77%	67,14%

### 2) Pengujian Kedua

Selanjutnya pada pengujian kedua yaitu dengan menggunakan kombinasi data sebesar 70:30. Dimana data latih yang dimasukkan sebanyak 70% sedangkan data uji 30% dari jumlah dataset yang telah ditentukan dengan menggunakan nilai parameter yaitu  $\sigma = 3$ , nilai parameter  $C = 1$ , dan  $d = 3$ . Hasil akurasi klasifikasi pada pengujian pertama dapat dilihat pada tabel IX, dimana kernel linear lebih unggul dibandingkan kernel lainnya, yaitu sebesar 68,77 %.

TABEL IX  
AKURASI MODEL PADA PENGUJIAN KEDUA

Linear	RBF	Polynomial	Sigmoid
68,77%	67,57%	57,85%	67,51%

### 3) Pengujian Ketiga

Berikutnya yaitu pengujian yang ketiga yaitu menggunakan kombinasi data sebesar 80:20. Dimana data latih yang dimasukkan sebanyak 80% sedangkan data uji 20% dari jumlah dataset yang telah ditentukan dengan menggunakan nilai parameter yaitu  $\sigma = 3$ , nilai parameter  $C = 1$ , dan  $d = 3$ . Hasil akurasi klasifikasi pada pengujian pertama dapat dilihat pada tabel X, dimana kernel linear lebih unggul dibandingkan kernel lainnya, yaitu sebesar 69,24 %.

TABEL X  
AKURASI MODEL PADA PENGUJIAN KETIGA

Linear	RBF	Polynomial	Sigmoid
69,24%	68,46%	59,70%	67,91%

### C. Evaluasi

Pada tahapan evaluasi dilakukan perhitungan pada nilai akurasi, precision, dan recall, dimana hal ini ditujukan untuk mengetahui kualitas dari model yang dibuat dalam metode SVM. Terdapat tiga percobaan pada pengujian model tersebut dengan mendapatkan hasil rata-rata yaitu:

#### 1) Performa Pengujian Pertama

Berdasarkan pada tabel XI, telah didapatkan rata-rata hasil akurasi, precision, recall, F1-score pada keempat kernel dalam pengujian pertama di kombinasi data 60:40.

TABEL XI  
PERFORMA PENGUJIAN PERTAMA

Kernel	Akurasi	Precision	Recall	F1 Score
Linear	68,24%	0,70	0,59	0,63
RBF	66,57%	0,73	0,54	0,60
Polynomial	55,77%	0,76	0,37	0,43
Sigmoid	67,14%	0,69	0,57	0,61

2) *Performa Pengujian Kedua*

Berdasarkan pada tabel XII, telah didapatkan rata-rata hasil akurasi, *precision*, *recall*, *F1-score* pada keempat kernel dalam pengujian pertama di kombinasi data 70:30.

TABEL XII  
PERFORMA MODEL PADA PENGUJIAN KEDUA

Kernel	Akurasi	Precision	Recall	F1 Score
Linear	68,77%	0,71	0,60	0,64
RBF	67,5%	0,74	0,56	0,61
Polynomial	57,85%	0,78	0,39	0,46
Sigmoid	67,51%	0,70	0,58	0,62

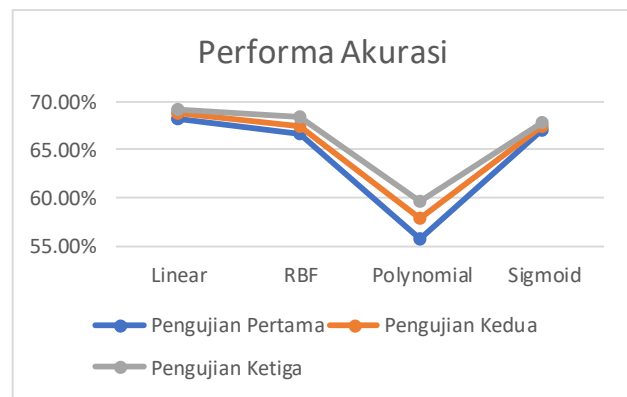
3) *Performa Pengujian Ketiga*

Berdasarkan pada tabel XIII, telah didapatkan rata-rata hasil akurasi, *precision*, *recall*, *F1-score* pada keempat kernel dalam pengujian pertama di kombinasi data 80:20.

TABEL XIII  
PERFORMA MODEL PADA PENGUJIAN KETIGA

Kernel	Akurasi	Precision	Recall	F1 Score
Linear	69,24%	0,71	0,61	0,64
RBF	68,46%	0,75	0,57	0,62
Polynomial	59,70%	0,78	0,41	0,48
Sigmoid	67,91%	0,70	0,59	0,63

Untuk memudahkan dalam hasil performa pada tabel pengujian diatas, maka akan dijelaskan dalam bentuk grafik untuk hasil akurasi, dimana bisa dilihat pada Gbr. 3. Sehingga dapat diartikan bahwa hasil akurasi pada kernel Linear lebih unggul diantara kernel lainnya. Di samping itu dapat dilihat pada hasil *precision* dan *recall* yang terdapat pada tabel XI, XII, dan XIII, bahwa ketepatan pada model klasifikasi dalam memberikan informasi memiliki nilai yang rendah, dimana belum sesuai dengan permintaan dan keberhasilan model dalam proses klasifikasi.



Gbr. 3 Grafik akurasi selama pengujian pertama, kedua, dan ketiga.

IV. KESIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan, penelitian ini menggunakan *dataset* sebesar 1000 *records* data buku sebagai uji data, dimana terdiri dari sepuluh kategori yang disesuaikan dengan DDC yaitu komputer, informasi, dan referensi umum; Filsafat dan psikologi; Agama; Ilmu sosial; Bahasa; Sains dan matematika; Teknologi; Kesenian dan rekreasi; Sastra; Sejarah dan geografi. Tujuan dari penelitian ini, yaitu untuk mengetahui pengaruh pada jumlah data latih terhadap kualitas dan efektivitas metode SVM, dimana dilakukan tiga kali pengujian pada model kombinasi data latih dan data uji, yaitu 60:40, 70:30, dan 80:20. Setelah ketiga pengujian telah memasuki tahap proses klasifikasi, maka dapat diambil kesimpulan bahwa metode SVM dalam klasifikasi teks pada *digital library* dapat digunakan dan menghasilkan akurasi tertinggi yang didapat dengan nilai 69,24% pada penggunaan kernel Linear dibandingkan dengan kernel lainnya, dimana memiliki nilai *precision* 71%, *recall* 61%, dan *f1-score* 64%.

UCAPAN TERIMA KASIH

Dengan mengucapkan puji syukur kepada Allah SWT yang telah memberikan kemudahan dan kelancaran dalam menyelesaikan jurnal dan penelitian ini. Penulis mengucapkan terima kasih kepada kedua orang tua serta semua pihak yang telah membantu dan memberikan do'a serta memberi dukungan, sehingga jurnal dan penelitian ini dapat diselesaikan dengan baik.

REFERENSI

- [1] Vapnik, V. N. (1999). *The Nature of Statistical Learning Theory*. New York: Springer-Verlag.
- [2] Shiela Novelia Dharma Pratiwi, B. S. (2016). Klasifikasi Email Spam dengan Menggunakan Metode Support Vector Machine dan K-Nearest Neighbor. *Jurnal Sains dan Seni ITS*.
- [3] Sindy Caesar Rusnanda Himatul Haliza, A. Q. (2021). Predictive Maintenance untuk Kendaraan Bermotor dengan Menggunakan Support Vector Machine (SVM). *Journal of Informatics and Computer Science (JINACS)*.
- [4] Arif Pratama, R. C. (2018). Implementasi Algoritme Support Vector Machine (SVM) untuk Prediksi Ketepatan Waktu Kelulusan Mahasiswa . *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 1704-1708.

- [5] Debby Alita, Y. F. (2020). Implementasi Algoritma Multiclass SVM Pada Opini Publik Berbahasa Indonesia di Twitter. *Jurnal TEKNOKOMPAK*, 86-91.
- [6] Tala, F. Z. (2003). *A Study of Stemming Effects on Information Retrieval in Bahasa Indonesia*. Netherlands: Institute for Logic, Language and Computation, Universitetivan Amsterdam.
- [7] *Integrated Online Public Access Catalogue (IOPAC) Universitas Negeri Surabaya*. (2021, Mei 6). Retrieved from Integrated Online Public Access Catalogue (IOPAC): <https://opac.unesa.ac.id/>
- [8] Aggarwal, C. C. (2015). *Data Mining : The Textbook*. New York: Springer International Publishing Switzerland.
- [9] Jiawei Han, M. K. (2006). *Data Mining: Concepts and Techniques, Second Edition*. San Francisco:Diane Cerra: Elsevier Science Ltd.