

Analisis Perbandingan Kernel Algoritma *Support Vector Machine* dalam Mengklasifikasikan Skripsi Teknik Informatika Berdasarkan Abstrak

Anggri Liani¹, Ultach Enri², Yuyun Umaidah³

Universitas Singaperbangsa Karawang

¹anggri.16040@student.unsika.ac.id, ²ultach@staff.unsika.ac.id,

³yuyun.umaidah@staff.unsika.ac.id

Diterima: 25 Juni 2020; Direvisi: 17 Oktober 2020; Disetujui: 30 Oktober 2020

Abstrak

Mahasiswa memiliki kewajiban menyelesaikan skripsi untuk menyelesaikan pendidikan jenjang S-1, namun justru menentukan topik skripsi adalah kesulitan pertama mahasiswa dalam pembuatan skripsi yang menjadi salah satu faktor mahasiswa lulus terlambat, Dengan melakukan pengklasifikasian skripsi berdasarkan abstrak dapat membantu mahasiswa dalam mencari referensi untuk menentukan topik skripsi. Metodologi yang digunakan adalah proses *text Mining* dengan proses *case folding*, *tekonizing*, *filtering*, *stemming*, *TF-IDF*, *data mining* dan *evaluation*. Pembagian data menggunakan rasio 80% data latih dan 20% data uji. Pengklasifikasian menggunakan algoritma *Support Vector Machine* (SVM). Algoritma *support vector machine* (SVM) adalah salah satu algoritma pengklasifikasian yang memiliki beberapa kernel yaitu linier dan 3 kernel yang paling dipertimbangkan. Validasi data menggunakan *cross validation* dengan *10-fold*. Tingkat akurasi didapatkan 81%, presisi 82% dan *recall* 81% pada kernel linier.

Kata kunci: *klasisfikasi, text mining, skripsi, support vector machine, kernel.*

Abstract

Students have the obligation to complete a thesis to complete their undergraduate education, but instead determining the topic of the thesis is the first difficulty of students in making a thesis which is one of the factors of students graduating late. By doing thesis classification based on abstracts can help students in finding references to determine thesis topics. The methodology used is the text mining process with process case folding, tekonizing, filtering, stemming, TF-IDF, data mining and evaluation. Data separation using a ratio of 80% training data and 20% testing data. Classification uses the Support Vector Machine (SVM) algorithm. The support vector machine (SVM) algorithm is a classification algorithm that has several kernels which are linear and the 3 most considered kernels. Data validation uses 10-fold cross validation. The accuracy rate is 81%, 82% precision and 81% recall on the Linearkernel.

Keywords: *klasifikasi, text mining, thesis, support vector machine, kernel.*

1. PENDAHULUAN

Mahasiswa lulus tidak tepat waktu atau terlambat adalah mahasiswa yang lulus tidak sesuai dengan peraturan akademik. Pendidikan jenjang S1 pada dasarnya dapat ditempuh dalam waktu 4 tahun dan pendidikan D3 dapat ditempuh dalam waktu 3 tahun [1]. Untuk menyelesaikan pendidikan jenjang S1 mahasiswa memiliki kewajiban akhir menyelesaikan skripsi sebagai tugas akhir, namun terdapat faktor eksternal yang dapat menjadi pengaruh mahasiswa terlambat

menyelesaikan skripsi yaitu meliputi ketersediaan motivasi dari keluarga, referensi, relasi antara dosen dengan mahasiswa, motivasi dari keluarga dan teman sebaya [2]. Skripsi adalah salah satu karya tulis ilmiah yang ditulis mahasiswa dimana mengandung topik didalamnya yang dapat dimanfaatkan sebagai referensi yang dapat membantu penelitian selanjutnya dan didalam skripsi terdapat abstrak yang berisi rangkuman singkat mengenai penelitian dalam suatu skripsi.

Text mining adalah proses mengolah sekumpulan teks atau dokumen untuk mendapatkan informasi dengan tugas khusus dari data mining yang di antaranya adalah pengelompokan dan peengkategorian [3]. *Support Vector machine* merupakan algoritma *klasifikasi supervised learning* yang diciptakan oleh Vladimir Vapnik dimana memprediksi kelas berdasarkan pola yang dihasilkan dari proses *training* [3]. Klasifikasi *Support Vector Machine* (SVM) menciptakan garis pembatas (*hyperlane*) untuk memisahkan antar kelas. SVM adalah algoritma dengan metode pembelajaran dalam menyelesaikan masalah klasifikasi dan regresi [4]. Pada awalnya *Support Vector Machine* dirancang untuk menyelesaikan masalah untuk klasifikasi linier, oleh karena itu, untuk dapat menyelesaikan permasalahan non-linear digunakan kernel *trick*. SVM memiliki 3 kernel yang paling dipertimbangkan yaitu polynomial, RBF, dan sigmoid [5].

Dalam penelitian yang dilakukan oleh Hidayattullah & Ma'ari (2016) yang berjudul Penerapan Text Mining dalam Klasifikasi Judul Skripsi menerapkan algoritma *Support Vector Machine* (SVM) dalam klasifikasi skripsi dengan 3 percobaan 3 fitur *n-gram* dan menghasilkan akurasi terbaik sebesar 73% pada fitur *n-gram unigram*[6]. Pada penelitian oleh (Inge Santoso, et al., 2017) Penerapan 3 Sentiment Analysis Pada Hasil Evaluasi Dosen Dengan *Support Vector Machine* didapatkan bahwa Algoritma *Support Vector Mechine* (SVM) dapat melakukan sentiment analysis dengan menggunakan 3 kelas sentimen. Akurasi tertinggi SVM pada sistem ini yaitu 67,83% [7].

Dalam upaya untuk mengatasi peningkatan jumlah mahasiswa terlambat Program Studi Teknik Informatika UNSIKA, perlu dilakukan pengklasifikasian skripsi berdasarkan topik sebagai salah satu upaya untuk memudahkan mahasiswa dalam mendapatkan referensi. Pada penelitian ini pengklasifikasian dilakukan menggunakan algoritma *Support Vector Machine* berdasarkan abstrak skripsi.

2. METODE PENELITIAN

Pada penelitian ini menggunakan metode penelitian terapan deskriptif dengan tujuan menerapkan, menguji, dan mengevaluasi suatu teori dalam memecahkan suatu masalah. Rancangan penelitian ini mengacu pada [8]. Metodologi yang digunakan dalam proses penelitian ini adalah proses *text mining* (*Text Processing*). Rancangan penelitian ini mengacu pada metode dan metodologi di atas. Rancangan penelitian terdapat pada Gambar 1, dimana alur rancangan penelitian disesuaikan dari penelitian yang dilakukan.

2.1 Pengumpulan Data

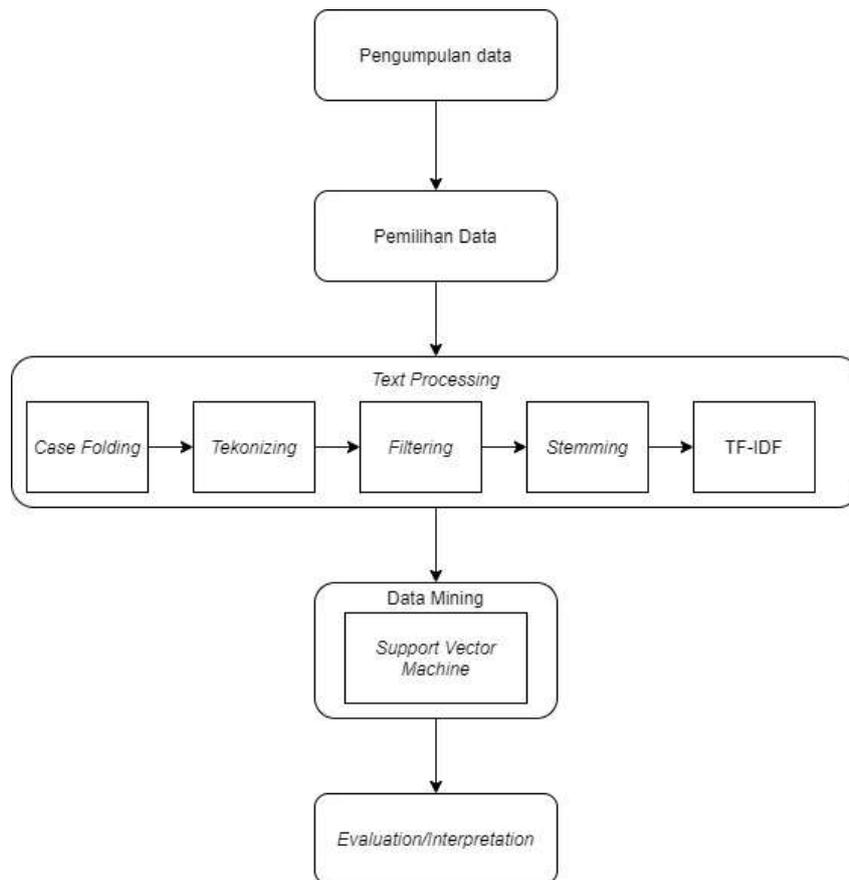
Pada tahap ini adalah melakukan studi dokumen abstrak skripsi atau mengumpulkan data abstrak skripsi dari tahun 2015-2019.

2.2 Pemilihan Data

Pada tahap pemilihan data adalah melakukan proses *labelling* data agar data dapat digunakan sebagai data pembelajaran.

2.3 Text Processing

Pada tahapan ini dilakukan proses *text mining*. Proses utama *text mining* sendiri adalah *preprocessing*, *pattern discovery algortitems* dan *presentastation or interpretation*. *Text preprocessing* memiliki tahap *case folding*, *tekonezing*, *filtering*, *steaming*, dan *term frequency*[4].



Gambar 1. Rancangan penelitian

A. Case Folding

Pada tahapan ini pada gambar 2 semua data diubah menjadi huruf kecil. Hanya huruf “a” sampai dengan “z” yang diterima. Karakter selain huruf dihilangkan dan dianggap delimiter.



Gambar 2. Proses case folding

B. Tekonizing

Pada tahap ini pada gambar 3 data hasil *case folding* diubah kalimat menjadi kata-kata, tahap ini dilakukan setelah proses *case folding*.



Gambar 3. Proses tekonizing

C. *Filtering*

Tahap ini pada gambar 2 dengan menghapus kata-kata yang kurang penting (*stoplist*) atau menyimpan kata penting (*wordlist*). *Stoplist / Stopword* adalah kata - kata yang tidak deskriptif yang dapat dibuang dalam pendekatan *bag-of-words*.



Gambar 4. Proses filtering

D. *Stemming*

Pada tahap ini pada gambar 5 data hasil *filtering* dilakukan proses pengembalian berbagai bentukan kata ke dalam suatu representasi yang sama seperti mengembalikan kata berimbuhan menjadi kata baku atau dasar.



Gambar 5. Proses steaming

E. TF-IDF

Pada tahap ini adalah tahap pembobotan kata pada dokumen menggunakan metode TF-IDF. menggabungkan dua konsep yaitu frekuensi kemunculan sebuah kata di dalam sebuah dokumen dan *inverse frekuensi* dokumen yang mengandung kata tersebut.

Term Frequency (TF) adalah pembobotan kata pada suatu dokumen yang didasarkan pada frekuensi jumlah kemunculan kata pada dokumen tersebut, semakin besar frekuensi kemunculan suatu kata maka semakin besar pula bobotnya. Persamaan TF pada persamaan (1) sebagai berikut [9]

$$d_{ij} = \log(d_{ij} + 1) \quad (1)$$

Dimana: d_{ij} = jumlah kemunculan kata i dalam dokumen j .

Invers document frequency (IDF) berfungsi untuk menghitung kata-kata yang sering muncul atau hadir di semua dokumen atau secara keseluruhan data. Secara umum dapat digunakan persamaan (2) sebagai berikut[9]:

$$d_{ij} = \log(f + 1) \times \log\left(\frac{D}{df}\right) \quad (2)$$

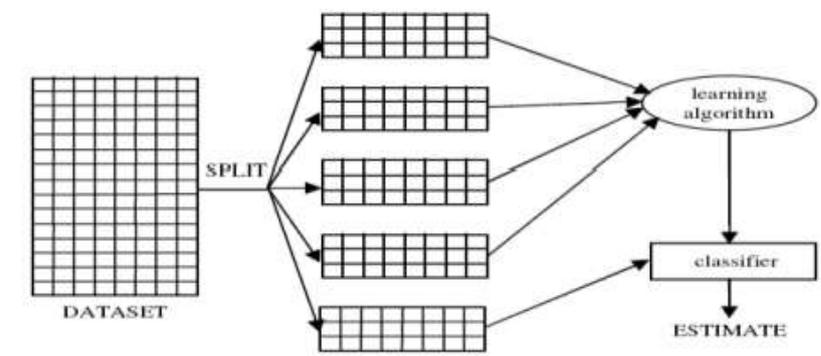
Dimana: D = jumlah seluruh dokumen.

df = jumlah dokumen yang mengandung kata i .

f = dij nilai TF transform.

2.4 Data Mining

Pada tahap ini hal pertama yang dilakukan adalah membagi data latih dan data uji. Kemudian dilanjutkan dengan melakukan pengklasifikasian dengan melakukan *K-Fold validation* menggunakan algoritma *Support Vector Machine* (SVM) dengan kernel berbeda yaitu: *Linier*, *Polinomial*, *Gaussian RBF*, dan *Sigmoid*. *K-fold validation* yang paling direkomendasikan adalah *10 fold* karena yang paling relative dan bervariasi [10].



Gambar 6. Cross validation. Sumber [11]

K-Fold Cross Validation adalah metode yang direkomendasikan *accuracy estimation* [9] seperti pada gambar 6. Pada *10-Fold validation* maka data latih akan dibagi menjadi 10 subset, dimana setiap subset akan disebut dengan model 1 sampai model 10. Model 1 yaitu jika subset pertama dijadikan data uji dan subset lainnya menjadi data latih. Model 2 yaitu jika subset kedua menjadi data uji dan sisanya menjadi data latih. Iterasi dilakukan sampai dengan model 10.

Penelitian ini dilakukan dengan menggunakan 2 skenario berbeda yaitu dengan *default parameter* dan parameter yang ditentukan (*free parameter*) yang dirujuk dari [9]. Parameter pada *default parameter* ditunjukkan pada tabel 1 dimana *tools* yang digunakan pada penelitian ini adalah SCIKIT LEARN dan untuk *free parameter* menggunakan parameter seperti pada tabel 2.

Tabel 1. Parameter *default parameter*

λ	0.001
γ	<i>scale</i>
C	1
ϵ	<i>false</i>
<i>iterasi</i>	-1

Tabel 2. Parameter *free parameter*

λ	0.5
γ	0.001
C	1
ϵ	0.000001
<i>iterasi</i>	20

2.5 Evaluation dan Interpretation

Pada tahap ini adalah tahap untuk mengetahui tingkat kinerja dari hasil pengklasifikasian menggunakan model yang telah dibuat. Kinerja hasil klasifikasi harus dilakukan evaluasi dengan cara menghitung nilai *recall*, *precision*, dan akurasi. Untuk memudahkan melakukan perhitungan, maka digunakan tabel *confusion matrix*. *Confusion matrix* atau disebut juga matriks klasifikasi adalah suatu alat visual yang biasanya digunakan dalam *supervised learning* [12]. Hasil dari evaluasi harus diinterpretasikan dengan baik dalam hal ini penelitian ini akan diinterpretasikan hasil evaluasi menggunakan tabel.

3. HASIL DAN PEMBAHASAN

3.1 Hasil Penelitian

Pada penelitian ini menggunakan 616 data abstrak skripsi yang dikumpulkan dari data skripsi Teknik Informatika UNSIKA tahun 2015-2019 yang tampak pada gambar 7. Yang kemudian dilakukan pemelihan data dengan melakukan labbling data abstrak skripsi seperti pada tabel 3 dengan 4 kategori topik bidang ilmu yang diambil dari penelitian sebelumnya oleh [13]. 4 kategori bidang ilmu tersebut yaitu *Computer Science Software Engineering (SE)*, *Information System (SI)*, *Artificial Intelligence (AI)*.

```

0 Bimbingan konseling merupakan metode yang dibu...
1 Saat ini jumlah Sekolah Menengah Atas dan seko...
2 Seiring dengan perkembangan teknologi yang sem...
3 Kemajuan teknologi semakin berkembang sangat p...
4 PT. Miyuki Indonesia yang bergerak di bidang M...
5 Jumlah kehadiran setiap peserta praktikum di l...
6 penelitian yang berjudul analisa metode penyim...
7 Data merupakan bagian yang sangat penting dala...
8 Kehidupan manusia yang tidak terlepas dari lis...
9 perkembangan teknologi saat ini mengalami perk...
10 Kemiskinan merupakan permasalahan yang komplek...
11 CV Putra Mandiri ceria sebagai perusahaan jasa...
12 perguruan tinggi merupakan suatu wadah bagi ma...
13 rumah merupakan tempat tinggal tempat berteduh...
14 buku merupakan sarana baca bagi setiap orang s...
.. ..

```

Gambar 7. Data Abstrak Skripsi 2015-2019 [data ke-0 sampai ke-14]

Tabel 3. Jumlah data abstrak skripsi yang telah di labelling

Kategori	Simbol	Jumlah
<i>Artificial Intelligence(AI)</i>	0	45
<i>Computer Science(CS)</i>	1	245
<i>Software Engineering(SE)</i>	2	142
<i>Information System(SI)</i>	3	184

A. Hasil Case Folding

```

0 bimbingan konseling merupakan metode yang dibu...
1 saat ini jumlah sekolah menengah atas dan seko...
2 seiring dengan perkembangan teknologi yang sem...
3 kemajuan teknologi semakin berkembang sangat p...
4 pt miyuki indonesia yang bergerak di bidang ma...
5 jumlah kehadiran setiap peserta praktikum di l...
6 penelitian yang berjudul analisa metode penyim...
7 data merupakan bagian yang sangat penting dala...
8 kehidupan manusia yang tidak terlepas dari lis...
9 perkembangan teknologi saat ini mengalami perk...
10 kemiskinan merupakan permasalahan yang komplek...
11 cv putra mandiri ceria sebagai perusahaan jasa...
12 perguruan tinggi merupakan suatu wadah bagi ma...
13 rumah merupakan tempat tinggal tempat berteduh...
14 buku merupakan sarana baca bagi setiap orang s...
15 informasi merupakan hal yang sangat penting se...

```

Gambar 8. Hasil *case folding* [data ke-0 sampai ke-15]

Pada gambar 8 adalah hasil dari data yang telah dihapus angka, spasi, dan tanda bacanya serta semua huruf telah diubah menjadi huruf kecil.

B. Hasil *Tekonizing*

Berikut adalah data yang telah dilakukan *tekonizing* atau telah dilakukan pemisahan kata, hasil dapat dilihat pada gambar 9:

```

0 [bimbingan, konseling, merupakan, metode, yang...
1 [saat, ini, jumlah, sekolah, menengah, atas, d...
2 [seiring, dengan, perkembangan, teknologi, yan...
3 [kemajuan, teknologi, semakin, berkembang, san...
4 [pt, miyuki, indonesia, yang, bergerak, di, bi...
5 [jumlah, kehadiran, setiap, peserta, praktikum...
6 [penelitian, yang, berjudul, analisa, metode, ...
7 [data, merupakan, bagian, yang, sangat, pentin...
8 [kehidupan, manusia, yang, tidak, terlepas, da...
9 [perkembangan, teknologi, saat, ini, mengalami...
10 [kemiskinan, merupakan, permasalahan, yang, ko...
11 [cv, putra, mandiri, ceria, sebagai, perusahaa...
12 [perguruan, tinggi, merupakan, suatu, wadah, b...
13 [rumah, merupakan, tempat, tinggal, tempat, be...
14 [buku, merupakan, sarana, baca, bagi, setiap, ...

```

Gambar 9. Hasil *tekonizing* [data ke-0 sampai ke-14]

C. Hasil *Filtering*

Berikut adalah hasil dari *filtering* atau data yang telah dihapus semua kata sambung menggunakan teknik stopword Bahasa Indonesia, hasil dapat dilihat pada gambar 10:

```

0 [bimbingan, konseling, merupakan, metode, dibu...
1 [jumlah, sekolah, menengah, atas, sekolah, men...
2 [seiring, perkembangan, teknologi, semakin, ca...
3 [kemajuan, teknologi, semakin, berkembang, san...
4 [pt, miyuki, indonesia, bergerak, bidang, manu...
5 [jumlah, kehadiran, peserta, praktikum, labora...
6 [penelitian, berjudul, analisa, metode, penyim...
7 [data, merupakan, bagian, sangat, penting, mel...
8 [kehidupan, manusia, terlepas, listrik, sering...
9 [perkembangan, teknologi, mengalami, perkemban...
10 [kemiskinan, merupakan, permasalahan, kompleks...
11 [cv, putra, mandiri, ceria, perusahaan, jasa, ...
12 [perguruan, tinggi, merupakan, suatu, wadah, m...
13 [rumah, merupakan, tempat, tinggal, tempat, be...
14 [buku, merupakan, sarana, baca, orang, sangat...

```

Gambar 10. Hasil *tekonizing* [data ke-0 sampai ke-14]

D. Hasil *Stemming*

Berikut adalah hasil dari *stemming* dimana semua imbuhan setiap kata dihapus , hasil dapat dilihat pada gambar 11:

```

0 [bimbing, konseling, rupa, metode, buat, bantu...
1 [jumlah, sekolah, tengah, atas, sekolah, tenga...
2 [iring, kembang, teknologi, makin, canggih, mu...
3 [maju, teknologi, makin, kembang, sangat, pesa...
4 [pt, miyuki, indonesia, gerak, bidang, manufac...
5 [jumlah, hadir, serta, praktikum, laboratorium...
6 [teliti, judul, analisa, metode, simpan, data,...
7 [data, rupa, bagi, sangat, penting, laku, suat...
8 [hidup, manusia, lepas, listrik, seringkali, t...
9 [kembang, teknologi, alami, kembang, pesat, te...
10 [miskin, rupa, masalah, kompleks, perintah, in...
11 [cv, putra, mandiri, ceria, usaha, jasa, renta...
12 [guru, tinggi, rupa, suatu, wadah, masyarakat,...
13 [rumah, rupa, tempat, tinggal, tempat, teduh, ...
14 [buku, rupa, sarana, baca, orang, sangat, pent...
    
```

Gambar 11. Hasil *stemming* [data ke-0 sampai ke-14]

E. Hasil TF-IDF

Berikut adalah hasil dari TF-IDF dimana data dihitung berdasarkan probabilitas frekuensi kemunculan kata pada setiap data dan seluruh data abstrak, hasil dapat dilihat pada gambar 12:

	tfidf
konseling	0.646670
bimbing	0.546409
hypertext	0.171286
sekolah	0.124050
basis	0.121803
...	...
ilmiah	0.000000
illustrator	0.000000
ikut	0.000000
iklim	0.000000
zoom	0.000000

5719 rows × 1 columns

Gambar 11. Hasil TF-IDF

F. *Data Mining*

Pada penelitian ini menggunakan algoritma *Support Vector Machine* (SVM) dan dilakukan dengan menggunakan 2 skenario berbeda yaitu dengan default parameter dan parameter yang ditentukan (*free parameter*) yang dirujuk dari [12]. Pembagian data menggunakan rasio 80% data latih dengan 492 data dan 20% data uji dengan 124 data.

a) *Default Parameter*

Berikut ini adalah hasil dari uji model dengan menggunakan algoritma *Support Vector Machine* (SVM) dengan *default parameter*. Hasil dapat dilihat pada tabel 4:

Tabel 4. Tabel Hasil *default parameter*

	Linier	Polynomial	RBF	Sigmoid
Akurasi	81%	57%	73%	78%
<i>Precision</i>	82%	64%	71%	78%
<i>Recall</i>	81%	57%	73%	78%

b) *Free Parameter*

Berikut ini adalah hasil dari uji model dengan menggunakan algoritma *Support Vector Machine* (SVM) dengan *default parameter*. Hasil dapat dilihat pada tabel 5:

Tabel 5 Tabel Hasil *free parameter*

	Linear	Polynomial	RBF	Sigmoid
Akurasi	26%	26%	22%	30%
<i>Precision</i>	26%	21%	17%	25%
<i>Recall</i>	26%	26%	22%	30%

3.2 Pembahasan

Pengimplementasian algoritma *Support Vector Machine* (SVM) dalam mengklasifikasikan abstrak skripsi dengan metodologi proses *text mining*. Pada pemodelan klasifikasi *Support Vector Machine* (SVM) dengan menggunakan default parameter menghasilkan akurasi terbaik pada *kernel linier* dengan akurasi 81%, *precision* 82%, dan *recall* 81%, dengan data yang berhasil diprediksi 101 data dari 124 data uji.

Hal ini menunjukkan bahwa klasifikasi abstrak skripsi dapat diklasifikasikan menggunakan algoritma SVM *kernel linier* dengan *default parameter*.

Pada pemodelan klasifikasi *Support Vector Machine* (SVM) dengan menggunakan *free parameter* menghasilkan akurasi terbaik pada *kernel Sigmoid* dengan akurasi 30%, *precision* 25%, dan *recall* 30%, dengan data yang berhasil diprediksi 37 data dari 124 data uji.

Hal ini menunjukkan bahwa model ini hanya mampu memprediksi 30% dari data testing menandakan bahwa hasil prediksi oleh model tidak akurat.

4. KESIMPULAN

Dari hasil penelitian ini disimpulkan untuk mengatasi peningkatan mahasiswa lulus terlambat dilakukan pengklasifikasian skripsi berdasarkan abstrak dalam beberapa topik bidang ilmu agar dapat memudahkan mahasiswa dalam mencari referensi. Pengklasifikasian pengklasifikasian abstrak skripsi Teknik Informatika dilakukan beberapa proses seperti *case folding*, *tokenizing*, *filtering*, *stemming*, *term weight* (TF-IDF) dan melakukan pemodelan dengan menggunakan algoritma *Support Vector Machine* dan melakukan pemilihan kernel maupun parameter sequential lainnya untuk mendapatkan model terbaik. Berdasarkan hasil uji coba parameter pada model pengklasifikasian dengan menggunakan algoritma *Support Vector Machine* (SVM) menunjukkan bahwa model pengklasifikasian terbaik adalah model dengan *default parameter* dan kernel linear yang menghasilkan akurasi 81%.

5. SARAN

Berdasarkan hasil penelitian yang telah dilakukan, disarankan untuk penelitian selanjutnya dilakukan penambahan dataset dengan rasio label data sama dan memungkinkan peningkatan akurasi dan pembelajaran model setiap modelnya seimbang. Melakukan variasi skenario pengujian yang lebih banyak untuk mendapatkan performa terbaik model dari algoritma yang

digunakan. Menggunakan algoritma lainya untuk mendapatkan model pengklasifikasian terbaik yang dapat memungkinkan peningkatkan akurasi.

DAFTAR PUSTAKA

- [1] Widarto. (2017). Faktor Penghambat Studi Mahasiswa Yang Tidak Lulus Tepat Waktu Di Jurusan Pendidikan Teknik Mesin FT UNY. *Jurnal Dinamika Vokasional Teknik Mesin, Vol.2, No.2, Oktober-2017, 127-138*
- [2] Aslinawati, E. N., & Mintarwati, S. U. (2017) Keterlambatan Penyelesaian Skripsi Mahasiswa Angkatan 2012 (Studi Kasus di Jurusan Ekonomi Pembangunan Fakultas Ekonomi Universitas Negeri Malang). *Jurnal Pendidikan Ekonomi, Vol.10, No.1. 23-33.*
- [3] Ernawati I. (2019). Naïve Bayes Classifier Dan Support Vector Machine Sebagai Alternatif Solusi Untuk Text Mining. *Jurnal Teknologi Informasi dan Pendidikan. Vol. 12, No. 2, 32-38.*
- [4] Musyaffa N & Rifai B. (2018). Model Support Vector Machine Berbasis Particle Swarm Optimization Untuk Prediksi Penyakit Liver. *Jurnal Ilmu Pengetahuan Dan Teknologi Komputer, e-ISSN: 2527-4864, Vol. 3. No. 2, 189-193.*
- [5] Gorunescu, F. (2011). *Data Mining Concepts, Models, Concepts.* Verlag Berlin Heidelberg: Springer.
- [6] Hidayatullah, A. F. & Ma'arif, M. R. (2016). Penerapan *Text Mining* dalam Klasifikasi Judul Skripsi. *Seminar Nasional Aplikasi Teknologi Informasi (SNATi), 33-36.*
- [7] Inge Santoso. V., Virginia. G. & Lukito. Y. (2017). Penerapan Sentiment Analysis Pada Hasil Evaluasi Dosen Dengan *Metode Support Vector Machine.* *Jurnal Transformatika, Vol.14, No.2, 79-83.*
- [8] R. Feldman dan J. Sanger, (2007), *The Text Mining Handbook*, New York: Cambridge University Press.
- [9] Tapikap H. F, Djahi B. S & Widiastuti T. (2019). Klasifikasi Spam E-Mail Menggunakan Metode Transformed Complement Naïve Bayes (TNCB). *J-ICON, Vol. 7 No. 1, 21-26.*
- [10] Han, Jiawei. (2007). *Data Mining: Concepts and Techniques.* United States of America: Diane Cerra.
- [11] Pudjiarti E. (2016). Prediksi Spam Email Menggunakan Metode Support Vector Machine Dan Particle Swarm Optimization. *Jurnal Pilar Nusa Mandiri Vol.XII, No.2, 171-181.*
- [12] Puspitasari, A. M., Ratnawati, D. K. & Widodo, A. W. (2018). Klasifikasi Penyakit Gigi Dan Mulut Menggunakan Metode *Support Vector Machine.* *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer, e-ISSN: 2548-964, Vol. 2, No. 2, 802-810.*
- [13] Hamidi, M. N. (2019). *Text Mining Klasifikasi Skripsi Berdasarkan Abstrak Menggunakan Metode Naïve Bayes.* (Skripsi). Universitas Singaperbangsa Karawang, Karawang.