

**KLASIFIKASI NASKAH PUBLIKASI MENGGUNAKAN
METODE SUPPORT VECTOR MACHINE LINEAR**

SKRIPSI



disusun oleh

Muh Fierhan Hasir

14.11.8029

**PROGRAM SARJANA
PROGRAM STUDI INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2018**

**KLASIFIKASI NASKAH PUBLIKASI MENGGUNAKAN
METODE SUPPORT VECTOR MACHINE LINEAR**

SKRIPSI

untuk memenuhi sebagian persyaratan
mencapai gelar Sarjana
pada Program Studi Informatika



disusun oleh

Muh Fierhan Hasir

14.11.8458

**PROGRAM SARJANA
PROGRAM STUDI INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2018**

PERSETUJUAN

SKRIPSI

**KLASIFIKASI NASKAH PUBLIKASI MENGGUNAKAN
METODE SUPPORT VECTOR MACHINE
LINEAR**


yang dipersiapkan dan disusun oleh

Muh Fierhan Hasir

14.11.8029

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 20 Juli 2018

Dosen Pembimbing,


Heri Sismoro, M.Kom.
NIK. 190302057

PENGESAHAN

SKRIPSI

**KLASIFIKASI NASKAH PUBLIKASI MENGGUNAKAN
METODE SUPPORT VECTOR MACHINE**

LINEAR

yang dipersiapkan dan disusun oleh

Muh Fierhan Hasir

14.11.8029

telah dipertahankan di depan Dewan Penguji
pada tanggal 23 Agustus 2018

Susunan Dewan Penguji

Nama Penguji

Heri Sismoro, M.Kom
NIK. 190302057

Robert Marco, M.T
NIK. 190302228

Akhmad Dahlan, M.Kom
NIK. 190302174

Tanda Tangan



Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 27 Agustus 2018



DEKAN FAKULTAS ILMU KOMPUTER

Krisnawati, S.Si, M.T.
NIK. 190302038

PERNYATAAN

Saya yang bertandatangan di bawah ini menyatakan bahwa, skripsi ini merupakan karya saya sendiri (ASLI), dan isi dalam skripsi ini tidak terdapat karya yang pernah diajukan oleh orang lain untuk memperoleh gelar akademis di suatu institusi pendidikan tinggi manapun, dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis dan/atau diterbitkan oleh orang lain, kecuali yang secara tertulis diacu dalam naskah ini dan disebutkan dalam daftar pustaka.

Segala sesuatu yang terkait dengan naskah dan karya yang telah dibuat adalah menjadi tanggungjawab saya pribadi.

Yogyakarta, 28 Agustus 2018



Muh Fierhan Hasir

NIM. 14.11.8029

MOTTO

“Your time is limited, so don’t waste it living someone else’s life. Don’t be trapped by dogma—which is living with the results of other people’s thinking. Don’t let the noise of others’ opinions drown out your own inner voice. And most important, have the courage to follow your heart and intuition.”

- Steve Jobs –

“Setiap proses perubahan belum akan berhasil sebelum manusia berhasil memperbaiki cara berpikirnya.”

- Rhenald Kasali, Ph. D –

“Sometimes it is better to do things faster, even if they give an error, than to do nothing”

- Vladimir Putin –

“Miracles only occur for those with the determination to never stop trying”

- Emporio Ivankov, One Piece–

“Life's simple: You make choices and you don't look back.”

- Han, Tokyo Drift –

“Idealisme adalah kemewahan terakhir yang hanya dimiliki oleh pemuda”

- Tan Malaka –

PERSEMBAHAN

Sujud syukurku kusembahkan kepadamu Tuhan yang Maha Agung nan Maha Tinggi nan Maha Adil nan Maha Penyayang, atas ridho-Mu dan karunia-Mu penulis dapat menyelesaikan Skripsi dengan judul **“Klasifikasi Naskah Publikasi Menggunakan Metode Support Vector Machine Linear“**. Dengan rasa bangga dan bahagia saya khaturkan rasa syukur dan terima kasih saya kepada:

1. Allah SWT, karena atas izin dan karunia-Nya maka skripsi ini dapat dibuat dan selesai pada waktunya. Puji syukur yang tak terhingga pada Allah SWT penguasa alam yang meridhoi dan mengabulkan segala do'a.
2. Ayah dan Ibu saya (Hasir Tjenne dan Husmania serta Hendrawati Yunus), yang telah memberikan dukungan moril maupun materi serta do'a yang tiada henti untuk kesuksesan saya, karena tiada kata seindah lantunan do'a dan tiada do'a yang paling khusuk selain do'a yang terucap dari orang tua.
3. Seluruh Keluarga besar Ayah dan Ibu yang telah memberikan doa semangat dan dukungan kepada saya.
4. Teman – teman kelas serta keluarga bagi saya “14 S1 TI 07” yang sudah memberikan dukungan, semangat dan motivasi meskipun kita sering becanda dikelas tapi rasa cinta kita ke sesama begitu besar. Semoga kita selalu dalam lindungan-Nya dan diberikan kebahagiaan serta kesuksesan dunia akhirat.
5. Keluarga Besar “The Big Family of HMJTI dan HMIF” yang sudah memberikan banyak pengalaman yang manis, pahit, asam tentang kehidupan perjuangan untuk saya.

6. Reza pahlevi, Dede, Ichwan, Fikri Nudhar, Ryan, Putra, Luken, Indra, Ainun, Adit dan juga seluruh teman sekaligus keluarga seperantauan “MKS - JOG” yang sudah menjadi keluarga sejak pertama pertama kali sampai di jogja hingga sekarang.
7. Rofie, Tanjung, Leno, Alfairuz, Benny, Riki Burjo, Salim, Darmawan, Bos Adji, Gan Dedy, Latif, Oktavianus Bili dan seluruh Setan Penghuni “TI 07 Underground”, gaada kalian gak rusuh!
8. Asmo, Arman, Arsil, Chaliq serta seluruh orang di “Green Home”.
9. Tukang Gosip Sauna “Ariska Dkk”, yang sudah jadi teman serta menjadi keluarga selama saya di Jogja, Semoga kita selalu dalam lindungan-Nya dan diberikan kebahagiaan kesuksesan dunia akhirat. Sukses Selalu guys!
10. Teman Jalan, Motivator, Lawan debat, Kawan begadang Inayatul Azisah andalancu, terima kasih untuk motivasi, saran, omelan serta waktunya 😊
11. Irma Anggraini, Fandi Abdillah dan Endah Handayani selaku editor naskahku, tanpa kalian aku tak paham heading.
12. Teman – teman serta berbagai pihak yang tidak dapat disebutkan satu per satu yang juga turut membantu memberikan semangat dan motivasi dalam penyelesaian skripsi ini.

KATA PENGANTAR

Assalamualaikum Wr. Wb dan Salam Sejahtera

Puja dan puji syukur penulis panjatkan kehadirat Tuhan YME karena atas limpahan rahmat, ridho dan karunia-Nya penulis masih diberi kesempatan dan kemudahan untuk menyelesaikan skripsi yang berjudul "Klasifikasi Naskah Publikasi Menggunakan Metode Support Vector Machine Linear".

Maksud dan tujuan skripsi ini disusun dalam rangka memenuhi salah satu persyaratan kelulusan jenjang Program Sarjana Strata 1 pada Fakultas Ilmu Komputer Universitas AMIKOM Yogyakarta. Dengan selesainya skripsi ini, maka penulis tidak lupa mengucapkan terima kasih kepada:

1. Bapak Prof. Dr. M. Suyanto, M.M. , selaku Rektor Universitas AMIKOM Yogyakarta.
2. Ibu Krisnawati, S.Si., M.T. selaku Dekan Fakultas Ilmu Komputer dan Bapak Sudarmawan, S.T, M.T Selaku Kepala Program Studi Informatika Universitas AMIKOM Yogyakarta.
3. Bapak Heri Sismoro, M.Kom dan Bapak Hastari Utama M.Cs selaku Dosen yang telah memberikan banyak bimbingan, saran, dan waktunya dengan sepenuh hati.
4. Bapak Robert Marco, M.T dan Bapak Akhmad Dahlan, M.Kom selaku dosen penguji, serta semua dosen Universitas Amikom Yogyakarta, terima kasih atas ilmu yang diberikan serta pengalaman kepada penulis.

5. Orang Tua, Saudara dan segenap keluarga yang telah memberikan dukungan moril serta materi dengan tulus, ikhlas dan penuh kasih sayang.
6. Sahabat – sahabat yang telah memberikan semangat, motivasi dan membantu dalam pengerjaan skripsi ini.
7. Seluruh staff dan karyawan Universitas Amikom Yogyakarta yang telah membantu kelancaran segala aktivitas penulisan skripsi ini.
8. Seluruh pihak yang tidak dapat disebutkan satu persatu yang telah banyak membantu sehingga skripsi ini dapat terselesaikan.

Penulis menyadari bahwa dalam penyusunan skripsi ini masih terdapat banyak kekurangan dan jauh dari kata sempurna, oleh karena itu penulis selalu membuka diri untuk saran dan kritik yang membangun dari pembaca. Semoga skripsi ini dapat bermanfaat bagi penyusun sendiri pada khususnya dan bagi para pembaca pada umumnya. Terima kasih.

Wassalamualaikum Wr. Wb.

Yogyakarta, 29 Agustus 2018

Penulis

DAFTAR ISI

PENGESAHAN	i
PERNYATAAN.....	iii
MOTTO	iv
PERSEMBAHAN.....	v
KATA PENGANTAR	vii
DAFTAR ISI.....	ix
DAFTAR TABEL.....	xiii
DAFTAR GAMBAR.....	xiv
LAMPIRAN.....	xvii
INTISARI.....	xviii
<i>ABSTRACT</i>	xix
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	2
1.3 Batasan Masalah.....	3
1.4 Maksud dan Tujuan Penelitian.....	3
1.5 Manfaat Penelitian.....	3
1.6 Metode Penelitian.....	4
1.6.1 Metode Pengumpulan Data.....	4
1.6.1.1 Metode Observasi.....	4
1.6.1.2 Studi Literatur	4
1.6.2 Metode Analisis	4
1.6.3 Metode Perancangan	4
1.6.3 Metode Pengembangan	5
1.6.4 Metode Testing.....	5
1.7 Sistematika Penulisan.....	5
BAB II LANDASAN TEORI.....	8

2.1	Tinjauan Pustaka	8
2.2	Data Mining.....	10
2.2.1	Text Mining.....	11
2.2.1.1	Text Preprocessing	12
2.3	SVM	12
2.3.1	Linear SVM.....	14
2.3.2	Non-Linear SVM	14
2.3.3	Multi Class SVM	15
2.3.4	Kelebihan SVM.....	15
2.4	TF-IDF.....	16
2.4.1	Term Frequency	17
2.4.2	Inverse Document Frequency	17
2.5	MVVM	18
2.6	WPF.....	19
2.6.1	XAML.....	20
2.6.2	System Requirement	20
2.7	Python.....	21
2.8	WEKA dan Rapid Miner	22
2.8.1	Weka	22
2.8.2	Rapid Miner	22
2.9	UML (Unified Modeling Language).....	23
2.9.1	Use Case Diagram.....	23
2.9.2	Class Diagram	26
2.9.3	Sequence Diagram	30
2.9.4	Activity Diagram.....	32
BAB III ANALISIS DAN PERANCANGAN		35
3.1	Gambaran Umum	35
3.2	Pengumpulan Data	35
3.3	Perancangan Sistem dan Model	36
3.3.1	Perancangan Pembuatan Model	36
3.3.1.1	Tahap Preprocessing.....	37
3.3.1.1.1	Tokenize	37

3.3.1.1.2 Filtering.....	38
3.3.1.1.3 Stemming	38
3.3.1.2 Pembobotan TF-IDF	39
3.3.1.3 Penghitungan TF-IDF.....	40
3.3.1.4 Pembuatan Model	42
3.3.2 Perancangan Program.....	43
3.3.2.1 Use Case Diagram	44
3.3.2.2 Activity Diagram.....	46
3.3.2.3 Class Diagram.....	50
3.3.2.4 Sequence Diagram	53
3.3.3 Rancang Program Preprocessing Dokumen.....	57
3.3.3.1 Flowchart	57
3.4 Rancang Antarmuka.....	58
3.4.1 Rancangan Page About	59
3.4.2 Rancangan Page Login.....	59
3.4.3 Rancangan Page Klasifikasi.....	60
3.4.4 Rancangan Page Training	61
BAB IV IMPLEMENTASI DAN PEMBAHASAN	63
4.1 Pengumpulan Data	63
4.2 Implementasi Program Pembuatan Dataset.....	64
4.2.1 TextPreocessing	64
4.2.2 TF-IDF	66
4.2.3 Counter.....	67
4.2.4 Main	69
4.3 Pemodelan SVM.....	78
4.3.1 Pembentukan Persamaan SVM Multikelas.....	78
4.4 Pengujian Model SVM.....	84
4.5 Implementasi Aplikasi Desktop	86
4.5.1 Core.....	87
4.5.2 DoClassifier.....	91
4.5.3 Uji Coba Program	99
BAB V PENUTUP.....	101

5.1	Kesimpulan.....	101
5.2	Saran.....	101
	DAFTAR PUSTAKA	103
	LAMPIRAN.....	106



DAFTAR TABEL

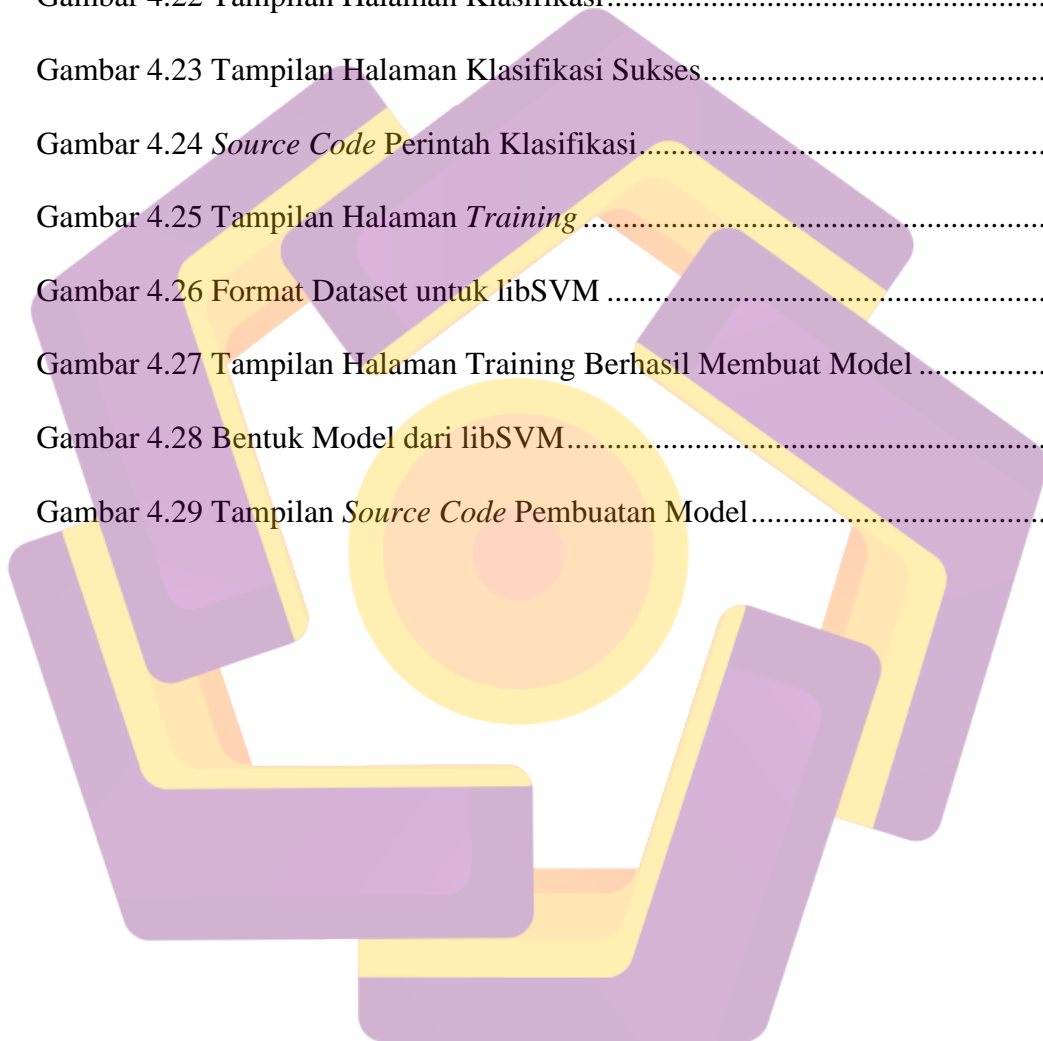
Tabel 2.1 Perbandingan Penelitian.....	9
Tabel 2.2 Notasi <i>Use Case Diagram</i>	24
Tabel 2.3 Notasi <i>Class Diagram</i>	27
Tabel 2.4 Notasi <i>Sequence Diagram</i>	31
Tabel 2.5 Notasi <i>Activity Diagram</i>	33
Tabel 3.1 Proses teks setelah <i>case folding</i>	38
Tabel 3.2 Tahap <i>tokenize, filtering, dan stemming</i>	39
Tabel 3.3 Contoh Corpus	41
Tabel 3.4 Fungsi kernel <i>support vector machine</i>	42
Tabel 4.1 Contoh Tabel Vektor.....	79
Tabel 4.2 Perbandingan C-SVC Menggunakan <i>Tools Weka 3.8.1</i>	84
Tabel 4.3 Perbandingan nu-SVC Menggunakan <i>Tools Weka 3.8.1</i>	85
Tabel 4.4 Perbandingan C-SVC Menggunakan <i>Tools Rapidminer 7.5</i>	85
Tabel 4.5 Perbandingan nu-SVC Menggunakan <i>Tools Rapidminer 7.5</i>	86
Tabel 4.6 Hasil Pengujian <i>Black Box</i>	100

DAFTAR GAMBAR

Gambar 2.1 Linear SVM.....	14
Gambar 2.2 Linear SVM (a) dan Polynomial SVM (B)	15
Gambar 2.3 Arsitektur MVVM.....	19
Gambar 3.1 <i>Flowchart</i> Pembuatan Dataset	36
Gambar 3.2 <i>Flowchart</i> Pemodelan SVM.....	36
Gambar 3.3 <i>Flowchart Preprocessing Text</i>	37
Gambar 3.4 <i>Flowchart</i> Pembuatan Corpus.....	40
Gambar 3.5 <i>Flowchart</i> Pembuatan Dataset	42
Gambar 3.6 <i>Flowchart</i> Pemodelan SVM.....	43
Gambar 3.7 <i>Flowchart</i> Alur Klasifikasi Dokumen Abstrak Naskah Publikasi	44
Gambar 3.8 <i>Use Case Diagram</i> Program DoClassifier.....	44
Gambar 3.9 <i>Activity Diagram</i> About.....	46
Gambar 3.10 <i>Activity Diagram</i> Login.....	47
Gambar 3.11 <i>Activity Diagram</i> Klasifikasi.....	48
Gambar 3.12 <i>Activity Diagram</i> Training	49
Gambar 3.13 <i>Class Diagram</i> DoClassifier	50
Gambar 3.14 <i>Detail Class Diagram</i> DoClassifier	51
Gambar 3.15 <i>Class Diagram</i> Core DoClassifier	52
Gambar 3.16 <i>Detail Class Diagram</i> Core DoClassifier	52
Gambar 3.17 <i>Sequence Diagram</i> About.....	53
Gambar 3.18 <i>Sequence Diagram</i> Login	54
Gambar 3.19 <i>Sequence Diagram</i> Klasifikasi	55

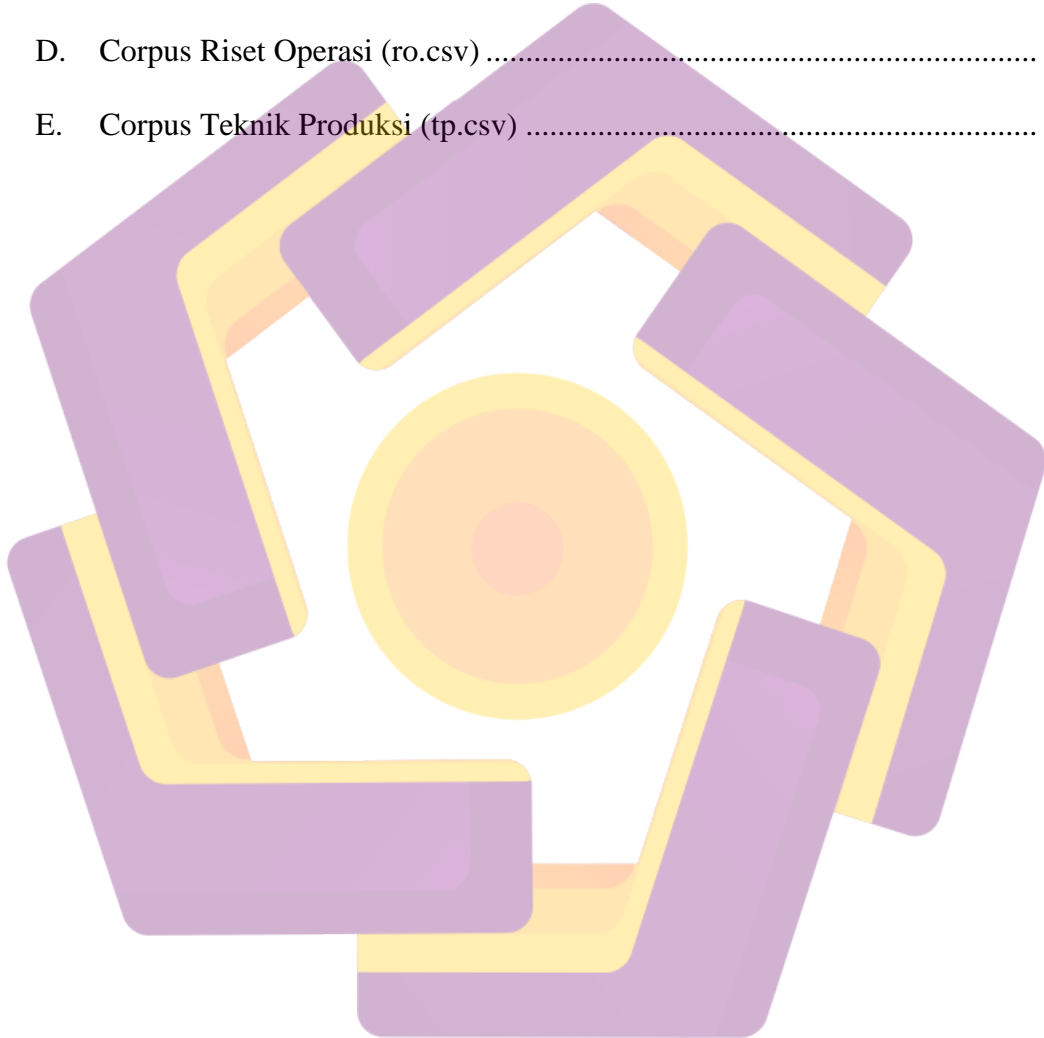
Gambar 3.20 <i>Sequence Diagram Training</i>	56
Gambar 3.21 <i>Flowchart python Dataset</i>	57
Gambar 3.22 <i>Sequence Diagram python Dataset</i>	58
Gambar 3.23 Rancang Antarmuka <i>Page About</i>	59
Gambar 3.24 Rancang Antarmuka <i>Page Login</i>	60
Gambar 3.25 Rancang Antarmuka <i>Page Klasifikasi</i>	61
Gambar 3.26 Rancang Antarmuka <i>Page Training</i>	62
Gambar 4.1 Bentuk Data Dokumen	63
Gambar 4.2 <i>Source Code Text Preprocessing</i>	65
Gambar 4.3 <i>Source Code TF-IDF</i>	67
Gambar 4.4 <i>Source Code Penjumlahan TF-IDF</i>	68
Gambar 4.5 Pemanggilan Modul pada Class Main	69
Gambar 4.6 Deklarasi <i>Variabel Class Main</i>	70
Gambar 4.7 Penulisan Fungsi <i>Class Main</i>	72
Gambar 4.8 Pembuatan Corpus Tiap Kategori	74
Gambar 4.9 Pembuatan Dataset Dengan Menghitung TF-IDF	77
Gambar 4.10 Contoh Corpus Kategori <i>Ergonomika</i>	77
Gambar 4.11 Contoh Dataset	78
Gambar 4.12 Penyebaran Data pada <i>Chart</i>	78
Gambar 4.13 Grap Data Vector Contoh	80
Gambar 4.14 Grap Data <i>Support Vector</i> Contoh	81
Gambar 4.15 Grap Data <i>Hyperplane</i> Contoh	84
Gambar 4.16 <i>Source Code Class SVMTrainer</i>	88
Gambar 4.17 <i>Source Code Class SVMPredictor</i>	89

Gambar 4.18 <i>Source Code Class SvmCore</i>	91
Gambar 4.19 Tampilan Halaman Login.....	92
Gambar 4.20 Tampilan Halaman About	92
Gambar 4.21 Tampilan <i>Source Code</i> Validasi Login	93
Gambar 4.22 Tampilan Halaman Klasifikasi.....	94
Gambar 4.23 Tampilan Halaman Klasifikasi Sukses.....	94
Gambar 4.24 <i>Source Code</i> Perintah Klasifikasi.....	95
Gambar 4.25 Tampilan Halaman <i>Training</i>	96
Gambar 4.26 Format Dataset untuk libSVM	97
Gambar 4.27 Tampilan Halaman Training Berhasil Membuat Model	97
Gambar 4.28 Bentuk Model dari libSVM.....	98
Gambar 4.29 Tampilan <i>Source Code</i> Pembuatan Model.....	99



LAMPIRAN

A. Bentuk Dataset (Dataset.txt)	106
B. Corpus Ergonomika (er.csv)	112
C. Corpus Sistem Manufaktur (sm.csv).....	134
D. Corpus Riset Operasi (ro.csv)	156
E. Corpus Teknik Produksi (tp.csv)	175



INTISARI

Latar belakang dari penelitian ini adalah karena kurang optimalnya pemanfaatan data terkhusus data teks yang disimpan dalam berbagai media repository yang dikarenakan jumlahnya yang semakin bertambah, juga karena keterbatasan kemampuan manusia dalam memilah data yang disimpan. Bidang ilmu data mining terkhusus text mining dapat dimanfaatkan dalam pengolahan data berukuran besar tersebut. Support Vector Machine adalah algoritma data mining yang akan digunakan dalam pengklasifikasian naskah publikasi pada penelitian ini.

Tujuan dari penelitian ini ialah mencari model terbaik dari 4 fungsi kernel Support Vector Machine yang dibandingkan tingkat akurasi, presisi dan recall nya, yang kemudian akan diaplikasikan pada prototipe program agar dapat mengklasifikasikan naskah publikasi. Pengklasifikasian naskah publikasi menggunakan 4 kategori yaitu Ergonomika, Riset Operasi, Teknik Produksi dan Sistem Manufaktur yang berasal dari sumber data yaitu Prosiding Seminar Nasional Teknik Industri UGM tahun 2015 sampai tahun 2017.

Hasil dari penelitian ini ialah sebuah prototipe program sudah dapat mengklasifikasikan kategori dari judul beserta abstrak yang dimasukkan berdasarkan model terbaik yang diaplikasikan. Model terbaik yang didapatkan dengan pengujian cross validation menggunakan total 171 data judul beserta abstrak dari 4 kategori mendapatkan hasil akurasi, presisi dan recall sebesar 94,71% adalah Support Vector Machine Linear.

Kata Kunci: *Data Mining, Text Mining, Klasifikasi, SVM*

ABSTRACT

Background of this study is due to the lack of optimal utilization of data, especially text data stored in various repository media due to the increasing number, also because of the limited ability of humans to sort stored data. The field of data mining technology specifically text mining can be used in processing such large data. Support Vector Machine is a data mining algorithm that will be used in classifying the publication texts in this study.

The purpose of this study was to find the best model of the 4 kernel Support Vector Machine functions compared to the level of accuracy, precision and recall, which would then be applied to the program prototype in order to classify the publication text. Classification of publication texts using 4 categories, namely Ergonomics, Operations Research, Production Engineering and Manufacturing Systems derived from data sources, namely the Proceedings of UGM Industrial Engineering National Seminar 2015 to 2017.

The results of this study are that a program prototype can classify the categories of titles along with the abstracts entered based on the best models applied. The best model obtained by cross validation testing using a total of 171 title data and abstracts from 4 categories getting the results of accuracy, precision and recall of 94.71% are Linear Support Vector Machines.

Keywords: *Data Mining, Text Mining, Classification, SVM*