

**ANALISIS SENTIMEN MASYARAKAT DI MASA PANDEMI
COVID-19 TERHADAP PROGRAM BANTUAN SOSIAL
DENGAN PENDEKATAN *TEXT MINING*
SUPPORT VECTOR MACHINE LINIER DAN NON LINIER**

SKRIPSI



RIZZA SULISTIANA

**PROGRAM STUDI S-1 STATISTIKA
DEPARTEMEN MATEMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS AIRLANGGA**

2022

**ANALISIS SENTIMEN MASYARAKAT DI MASA PANDEMI
COVID-19 TERHADAP PROGRAM BANTUAN SOSIAL
DENGAN PENDEKATAN *TEXT MINING*
SUPPORT VECTOR MACHINE LINIER DAN NON LINIER**

SKRIPSI



RIZZA SULISTIANA

**PROGRAM STUDI S-1 STATISTIKA
DEPARTEMEN MATEMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS AIRLANGGA**

2022

**ANALISIS SENTIMEN MASYARAKAT DI MASA PANDEMI
COVID-19 TERHADAP PROGRAM BANTUAN SOSIAL
DENGAN PENDEKATAN *TEXT MINING*
SUPPORT VECTOR MACHINE LINIER DAN NON LINIER**

SKRIPSI

Sebagai Salah Satu Syarat untuk Memperoleh
Gelar Sarjana Statistika
pada Fakultas Sains dan Teknologi
Universitas Airlangga

Rizza Sulistiana
NIM 081811833005

Disetujui oleh :

Pembimbing I,



Dr. Nur Chamidah, S.Si., M.Si.
NIP. 197206021998022001

Pembimbing II,



Dr. Toha Saifudin, S.Si., M.Si.
NIP. 197501061999031002

LEMBAR PENGESAHAN NASKAH SKRIPSI

Judul : Analisis Sentimen Masyarakat di Masa Pandemi Covid-19 terhadap Program Bantuan Sosial dengan Pendekatan *Text Mining Support Vector Machine* Linier dan Non Linier

Penyusun : Rizza Sulistiana

NIM : 081811833005

Pembimbing I : Dr. Nur Chamidah, S.Si., M.Si.

Pembimbing II : Dr. Toha Saifudin, S.Si., M.Si.

Tanggal Seminar :

Disetujui oleh:

Pembimbing I



Dr. Nur Chamidah, S.Si., M.Si.

NIP. 197206021998022001

Pembimbing II

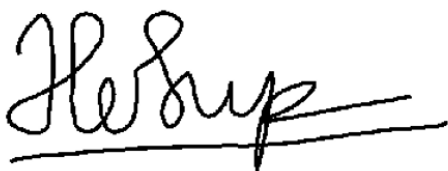


Dr. Toha Saifudin, S.Si., M.Si.

NIP. 197501061999031002

Mengetahui,

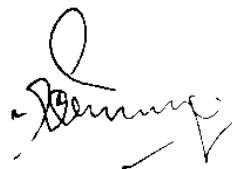
Ketua Departemen Matematika
Fakultas Sains dan Teknologi
Universitas Airlangga



Dr. Herry Suprajitno, S.Si., M.Si.

NIP. 196804041994031020

Koordinator Prodi S-1 Statistika
Fakultas Sains dan Teknologi
Universitas Airlangga



Drs. H. Sediono, M.Si.

NIP. 196107121987011001

PEDOMAN PENGGUNAAN SKRIPSI

Skripsi ini tidak dipublikasikan, namun tersedia di perpustakaan dalam lingkungan Universitas Airlangga, diperkenankan untuk dipakai sebagai referensi kepustakaan, tetapi pengutipan harus seizin penyusun dan harus menyebutkan sumbernya sesuai kebiasaan ilmiah. Dokumen skripsi ini merupakan hak milik Universitas Airlangga.

SURAT PERNYATAAN TENTANG ORISINALITAS

Yang bertanda tangan di bawah ini, saya:

Nama : Rizza Sulistiana

NIM : 081811833005

Program Studi : Statistika

Fakultas : Sains dan Teknologi Universitas Airlangga

Jenjang : Sarjana (S1)

Menyatakan bahwa saya tidak melakukan plagiat dalam penulisan skripsi saya yang berjudul :

Analisis Sentimen Masyarakat di Masa Pandemi Covid-19 Terhadap Program Bantuan Sosial dengan Pendekatan *Text Mining Support Vector Machine* Linier dan Non Linier

Apabila suatu saat nanti melakukan tindakan plagiat, maka saya akan menerima sanksi yang telah ditetapkan. Demikian surat pernyataan ini saya buat dengan sebenar-benarnya.

Surabaya, 05 April 2022



Rizza Sulistiana

NIM. 081811833005

KATA PENGANTAR

Puji dan syukur penulis panjatkan kepada Allah SWT karena telah memberikan rahmat dan hidayah-Nya sehingga penulis dapat menyelesaikan skripsi yang berjudul “Analisis Sentimen Masyarakat di Masa Pandemi Covid-19 terhadap Program Bantuan Sosial dengan Pendekatan *Text Mining Support Vector Machine* Linier dan Non Linier”. Pada kesempatan ini penulis ingin mengucapkan terima kasih kepada:

1. Prof. Dr. Mohammad Nasih, S.E., M.T., Ak., CMA. Selaku Rektor Universitas Airlangga, Prof. Dr. Moh Yasin, M. Si. Selaku Dekan Fakultas Sains dan Teknologi Universitas Airlangga, dan Drs. H. Sediono, M.Si. selaku Ketua Program Studi S-1 Statistika Universitas Airlangga yang telah memberikan kesempatan pada penulis untuk menimba ilmu di Program Studi Statistika Universitas Airlangga.
2. Kemendikbud atas beasiswa pendidikan (Bidikmisi) yang penulis dapatkan.
3. Ibu Dr. Nur Chamidah, S.Si., M.Si. dan Bapak Dr. Toha Saifudin, S.Si., M.Si. selaku dosen pembimbing penulis yang selalu memberikan bimbingan dan arahan kepada penulis dalam penyusunan skripsi.
4. Segenap dosen S1-Statistika Universitas Airlangga yang telah memberikan ilmunya selama masa perkuliahan.
5. Kedua orang tua dan adik tercinta yang telah memberikan dukungan dan doa kepada penulis.
6. Teman-teman seperjuangan di Universitas Airlangga.
7. Pak Yogi, Noviatus, Ilma, Feevrinna, Nada., Nurul A, Rahma, Frazna, Yeni, dan Aulia selaku penyemangat penulis yang telah memberikan dukungan dan kebersamaan penulis dalam penyusunan skripsi.

Penulis menyadari bahwa masih banyak kekurangan dalam penyusunan skripsi ini sehingga penulis mengharapkan kritik serta saran yang membangun dari semua pihak guna menyempurnakan skripsi ini.

Surabaya, April 2022

Rizza Sulistiana

Rizza Sulistiana, 2022. **Analisis Sentimen Masyarakat di Masa Pandemi Covid-19 terhadap Program Bantuan Sosial dengan Pendekatan *Text Mining Support Vector Machine* Linier dan Non Linier**. Skripsi di bawah Dr. Nur Chamidah, S.Si., M.Si. dan Dr. Toha Saifudin, S.Si., M.Si., Program Studi S-1 Statistika, Departemen Matematika, Fakultas Sains dan Teknologi, Universitas Airlangga, Surabaya.

ABSTRAK

Program Bantuan Sosial (Bansos) yang merupakan salah satu program pemerintah untuk meringankan perekonomian masyarakat selama masa pandemi Covid-19 menuai beragam komentar di masyarakat. Hal ini membuktikan perlunya evaluasi pelaksanaan program untuk menganalisis aspek-aspek yang perlu diperbaiki. Guna mencapai tujuan tersebut, analisis sentimen dilakukan menggunakan *text mining* metode *Support Vector Machine* (SVM) linier dan non linier kernel *polynomial*. Dalam analisisnya, penelitian ini menggunakan 800 *tweets* periode 1 s.d 5 Agustus 2021 yang dibagi menjadi 80% data *training* dan 20% data *testing* dengan 5-fold *cross validation* serta 174 *tweets* periode 14 s.d 17 April 2022 sebagai data input baru. Hasil *word cloud* sentimen negatif menunjukkan bahwa terdapat beberapa kata khusus yang muncul pada komentar tahun 2021 dan 2022, yaitu “korup”, “potong”, dan “lewat”. Hal ini menunjukkan bahwa aspek penyalahgunaan dana dan manajemen data perlu diperbaiki. Algoritma SVM linier dan non linier kernel *polynomial* menghasilkan rata-rata akurasi masing-masing sebesar 97% dan 89,13% sehingga dapat disimpulkan bahwa SVM linier lebih baik daripada SVM non linier kernel *polynomial* dalam mengklasifikasikan komentar masyarakat terhadap Bansos. Selanjutnya, algoritma SVM linier digunakan untuk mengklasifikasikan komentar terbaru (April 2022) masyarakat terkait Bansos dan menghasilkan akurasi sebesar 99,43% yang artinya bahwa algoritma SVM linier sangat sesuai dan dapat digunakan kembali untuk mengklasifikasikan komentar masyarakat di masa pandemi Covid-19 terkait Bansos. Dengan demikian, penelitian ini dapat menjadi referensi dalam upaya optimalisasi peran dan fungsi Program Bansos sehingga manfaat yang diberikan dapat dirasakan secara menyeluruh bagi masyarakat guna pemulihan sektor ekonomi nasional.

Kata Kunci : *Program Bansos, Analisis Sentimen, Text Mining, Support Vector Machine Linier, Kernel Polynomial*

Rizza Sulistiana, 2022. **Analysis of Community Sentiment in the Covid-19 Pandemic Period towards Social Assistance Programs with Linear and Non-Linear Text Mining Support Vector Machine Approaches**. The undergraduate thesis is guided by Dr. Nur Chamidah, S.Si., M.Si., and Dr. Toha Saifudin, S.Si., M.Si., Statistics Study Program, Mathematics Department, Faculty of Science and Technology, Airlangga University, Surabaya.

ABSTRACT

The Social Assistance Program (Bansos), which is one of the government's programs to ease the community's economy during the Covid-19 pandemic, has drawn various comments from the public. This proves the need to evaluate program implementation to analyze aspects that need improvement. To achieve this goal, sentiment analysis was carried out using text mining with the linear Support Vector Machine (SVM) method and non-linear kernel polynomial. In its analysis, this study uses 800 tweets for the period 1 to 5 August 2021 which is divided into 80% training data and 20% testing data with 5-fold cross-validation and 174 tweets for the period 14 to 17 April 2022 as new input data. The results of the negative sentiment word cloud show that several special words appear in the comments for 2021 and 2022, namely "korup", "potong", and "lewat". This shows that aspects of misuse of funds and data management need to be improved. The linear SVM and non-linear polynomial kernel SVM algorithms produce an average accuracy of 97% and 89.13%, respectively, so it can be concluded that linear SVM is better than non-linear polynomial kernel SVM in classifying public comments on social assistance. Furthermore, the linear SVM algorithm was used to classify the latest comments (April 2022) of the public regarding social assistance and resulted in an accuracy of 99.43% which means that the linear SVM algorithm is very suitable and can be reused to classify public comments during the Covid-19 pandemic related to social assistance. Thus, this research can be used as a reference in efforts to optimize the role and function of the Social Assistance Program so that the benefits provided can be felt as a whole for the community for the recovery of the national economic sector.

Keyword : *Bansos Program, Sentiment Analysis, Text Mining, Linear Support Vector Machine, Polynomial Kernel*

DAFTAR ISI

	Halaman
LEMBAR JUDUL	i
PEDOMAN PENGGUNAAN SKRIPSI	iv
KATA PENGANTAR	vi
ABSTRAK	vii
ABSTRACT	viii
DAFTAR ISI	ix
DAFTAR TABEL	xii
DAFTAR GAMBAR	xv
DAFTAR LAMPIRAN	xvii
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	5
1.3 Tujuan Penelitian	5
1.4 Manfaat Penelitian	6
1.5 Batasan Masalah	6
BAB II TINJAUAN PUSTAKA	7
2.1 Bantuan Sosial (Bansos)	7
2.2 Analisis Sentimen	8
2.3 <i>Text Mining</i>	9
2.4 <i>Word Cloud</i>	13
2.5 <i>Term Frequency-Inverse Document Frequency (TF-IDF)</i>	14
2.6 <i>Support Vector Machine (SVM)</i>	18
2.6.1 SVM Linier dengan <i>Hard Margin</i>	19
2.6.2 SVM Linier dengan <i>Soft Margin</i>	25
2.6.3 SVM Non Linier dengan Fungsi <i>Kernel</i>	29
2.7 Evaluasi Kinerja	33
2.7.1 <i>K-Fold Cross Validation (CV)</i>	33
2.7.2 <i>Confusion Matrix</i>	34
2.7.3 <i>PRESS' Q Test</i>	36

2.8	<i>Twitter</i>	36
2.9	<i>Software R</i>	37
BAB III METODE PENELITIAN		41
3.1	Data dan Sumber Data	41
3.2	Variabel Penelitian	42
3.3	Langkah Analisis Data	42
BAB IV HASIL DAN PEMBAHASAN		48
4.1	Analisis Deskriptif Komentar Masyarakat di Masa Pandemi Covid-19 Mengenai Program Bansos	48
4.1.1	Preprocessing Data <i>Tweets</i> Komentar Masyarakat Mengenai Program Bansos di Masa Pandemi <i>Covid-19</i>	49
4.1.2	Karakteristik Komentar Masyarakat Mengenai Program Bansos di Masa Pandemi <i>Covid-19</i>	53
4.1.3	Evaluasi Program Bantuan Sosial	56
4.2	Klasifikasi Komentar Masyarakat di Masa Pandemi Covid-19 Mengenai Program Bansos Menggunakan <i>Support Vector Machine</i> Linier	60
4.2.1	Perhitungan Pembobotan Kata TF-IDF	60
4.2.2	Pembagian Data Training dan Testing pada Data Komentar Masyarakat di Masa Pandemi Covid-19 Mengenai Program Bansos	61
4.2.3	Plot Komentar Masyarakat di Masa Pandemi Covid-19 Mengenai Program Bansos	64
4.2.4	Klasifikasi Komentar Masyarakat dengan SVM Linier pada Data <i>Training</i>	65
4.2.5	Klasifikasi Komentar Masyarakat dengan SVM Linier pada Data <i>Testing</i>	71
4.3	Klasifikasi Komentar Masyarakat di Masa Pandemi Covid-19 Mengenai Program Bansos Menggunakan <i>Support Vector Machine</i> Non Linier Kernel <i>Polynomial</i>	72
4.3.1	Klasifikasi Komentar Masyarakat dengan SVM Non Linier Kernel <i>Polynomial</i> pada Data <i>Training</i>	72
4.3.2	Klasifikasi Komentar Masyarakat dengan SVM Non Linier Kernel <i>Polynomial</i> pada Data <i>Testing</i>	76
4.4	Komparasi Performansi SVM Linier dan SVM Non Linier Kernel <i>Polynomial</i> dalam Mengklasifikasikan Komentar Masyarakat di Masa Pandemi Covid-19 Mengenai Program Bansos	78

4.5	Klasifikasi Komentar Masyarakat Terbaru terkait Bansos menggunakan Algoritma SVM dengan Kinerja Terbaik.....	79
BAB V	PENUTUP.....	86
5.1	Kesimpulan	86
5.2	Saran.....	88
DAFTAR PUSTAKA		90
LAMPIRAN		95

DAFTAR TABEL

Nomor	Judul Tabel	Halaman
2.1	Contoh Proses <i>Cleansing</i>	10
2.2	Contoh Proses <i>Case Folding</i>	11
2.3	Contoh Proses <i>Spelling Normalization</i>	11
2.4	Contoh Proses <i>Stopword Removal</i>	12
2.5	Contoh Proses <i>Tokenizing</i>	12
2.6	Contoh Proses <i>Stemming</i>	13
2.7	Hasil Kata Penting Pada Data Simulasi	15
2.8	Hasil Simulasi Perhitungan Pembobotan Kata TF-IDF	16
2.9	Data Input Simulasi Perhitungan Model Klasifikasi SVM	23
2.10	Nilai <i>Koefisien Lagrange</i> Simulasi Perhitungan Model SVM Linier	24
2.11	Nilai Fungsi Keputusan Simulasi Model Klasifikasi SVM Linier	25
2.12	Hasil Klasifikasi Sentimen pada Data Simulasi SVM Linier	25
2.13	Nilai <i>Koefisien Lagrange</i> Simulasi Perhitungan Model SVM Non Linier	32
2.14	Fungsi Keputusan Simulasi Perhitungan Model SVM Non Linier	33
2.15	Hasil Klasifikasi Sentimen pada Data Simulasi SVM Non Linier	33
2.16	<i>Confusion Matrix</i>	35
3.1	Contoh Data Penelitian	41
4.1	Hasil Filtering dan Labelling Komentar Masyarakat	48
4.2	Hasil <i>Cleansing</i> Komentar Masyarakat	49
4.3	Hasil <i>Case Folding</i> Komentar Masyarakat	50
4.4	Hasil <i>Spelling Normalization</i> Komentar Masyarakat	51
4.5	Hasil <i>Stopword Removal</i> Komentar Masyarakat	52

Nomor	Judul Tabel	Halaman
4.6	Hasil Stemming Komentar Masyarakat	52
4.7	Evaluasi Program Bantuan Sosial	57
4.8	Hasil Pembobotan Kata dengan TF-IDF	61
4.9	Pembagian Data Komentar Masyarakat Mengenai Program Bansos	62
4.10	Proporsi Data Komentar Masyarakat Mengenai Program Bansos	62
4.11	Data Input Perhitungan Model Klasifikasi SVM Linier pada Data Training	66
4.12	Nilai Fungsi Keputusan Perhitungan Model Klasifikasi SVM Linier pada Data Training untuk Fold ke-5	67
4.13	Hasil Klasifikasi Sentimen dengan SVM Linier pada Data Training untuk Fold ke-5	68
4.14	Confusion Matrix dari Data Training untuk Fold ke-1	69
4.15	Performansi Hasil Klasifikasi pada Data Training untuk Setiap Fold	70
4.16	Performansi Hasil Klasifikasi pada Data Testing pada Setiap Fold	70
4.17	Confusion Matrix Hasil Klasifikasi Data Testing untuk Fold ke-1	72
4.18	Parameter Kernel Polynomial Optimal pada Fold ke-1	73
4.19	Performansi SVM Non Linier Kernel Polynomial pada data Training	74
4.20	Performansi SVM Non Linier Kernel Polynomial Pada Data Testing	76
4.21	Confusion Matrix untuk Fold ke-1	77
4.22	Komparasi Performansi SVM Linier dan SVM Non Linier Kernel Polynomial	78
4.23	Data Tweets Komentar Masyarakat Terbaru Terkait Bansos	79

Nomor	Judul Tabel	Halaman
4.24	Data <i>Tweets</i> Komentar Terbaru Masyarakat Hasil <i>Preprocessing</i>	80
4.25	Hasil Pembobotan Kata dengan TF-IDF	83
4.26	Hasil Klasifikasi Sentimen dengan Algoritma SVM Linier	83
4.27	<i>Confusion Matrix</i>	84
4.28	Ukuran Ketepatan Klasifikasi Komentar Terbaru	84

DAFTAR GAMBAR

Nomor	Judul Gambar	Halaman
2.1	Contoh <i>word cloud</i>	14
2.2	<i>Hyperplane</i> dan <i>margin</i>	20
2.3	Data dengan <i>noise</i>	26
2.4	Data Non Linier	27
2.5	Pemetaan Data <i>Input</i> ke <i>Feature Space</i>	30
2.6	Kernel <i>Polynomial</i> dengan <i>Degree</i> 1	31
2.7	Kernel <i>Polynomial</i> dengan <i>Degree</i> 6	31
2.8	Ilustrasi <i>K-Fold Cross Validation</i>	34
2.9	Logo <i>twitter</i> dan <i>twitter API</i>	37
2.10	Logo <i>Software R</i>	38
3.1	Diagram Alir Langkah 1	43
3.2	Diagram Alir Langkah 2	44
3.3	Diagram Alir Langkah 3	45
3.4	Diagram Alir Langkah 4	46
3.5	Diagram Alir Langkah Analisis data	47
4.1	Presentase Komentar Masyarakat Berdasarkan Sentimen	53
4.2	Word Cloud Komentar Masyarakat Bersentimen Positif	54
4.3	Diagram Batang Frekuensi Kemunculan Kata Khusus pada Komentar Masyarakat Bersentimen Positif	55
4.4	Word Cloud Komentar Masyarakat Bersentimen Negatif	56
4.5	Diagram Batang Frekuensi Kemunculan Kata Khusus pada Komentar Masyarakat Bersentimen Negatif	56
4.6	Diagram Proporsi pada Data Training dan Data Testing	63
4.7	<i>Scatterplot</i> Kata "dana" dan "masyarakat"	64
4.8	<i>Scatterplot</i> Kata "uang" dan "bantu"	64
4.9	<i>Scatterplot</i> Kata Khusus Lain	65
4.10	Karakteristik Komentar Terbaru Masyarakat terkait Bansos	81

4.11	<i>Word Cloud</i> Komentar Terbaru Masyarakat terkait Bansos	82
4.12	<i>Bar Chart</i> Komentar Terbaru Masyarakat terkait Bansos	82

DAFTAR LAMPIRAN

No	Judul
1	Data Penelitian Komentar Masyarakat Mengenai Program Bansos
2	Daftar <i>slang word</i> tiap frase yang digunakan penulis
3	Hasil <i>Preprocessing</i> Dokumen Komentar Masyarakat Mengenai Program Bansos
4	Frekuensi Kata Khusus pada Dokumen Komentar Masyarakat Mengenai Program Bansos yang Bersentimen Positif
5	Frekuensi Kata Khusus pada Dokumen Komentar Masyarakat Mengenai Program Bansos yang Bersentimen Negatif
6	Hasil Pembobotan Kata dengan TF-IDF pada Dokumen Komentar Masyarakat Mengenai Program Bansos
7	Format Data Input untuk Klasifikasi SVM
8	Output Model SVM Linier pada Data Training
9	Nilai <i>Koefisien Lagrange</i> pada <i>Support Vector</i> untuk Fold ke-5 dengan SVM Linier
10	Hasil Nilai Fungsi Keputusan Klasifikasi Data <i>Training</i> untuk <i>Fold</i> ke-5 dengan SVM Linier
11	Hasil Klasifikasi Model Data <i>Training</i> untuk Fold ke-5 dengan SVM Linier
12	Output Model SVM Non Linier <i>Kernel Polynomial</i> pada Data <i>Training</i>
13	Nilai <i>Koefisien Lagrange</i> pada <i>Support Vector</i> untuk <i>Fold</i> ke-5 dengan SVM Non Linier <i>Kernel Polynomial Cost 2 Degree 2</i>
14	Hasil Nilai Fungsi Keputusan Klasifikasi Data <i>Training</i> untuk <i>Fold</i> ke-5 dengan SVM Non Linier <i>Kernel Polynomial</i>
15	Hasil Klasifikasi Model Data <i>Training</i> untuk <i>Fold</i> ke-5 dengan SVM Non Linier <i>Kernel Polynomial</i>
16	<i>Script</i> Program Analisis Data
17	Data Komentar Terbaru Masyarakat Terkait Bansos

No	Judul
18	Hasil <i>Preprocessing</i> Data Komentar Terbaru Masyarakat Terkait Bansos
19	Hasil Pembobotan Kata dengan TF-IDF pada Komentar Terbaru Masyarakat Terkait Bansos
20	Klasifikasi Komentar Terbaru Masyarakat terkait Program Bansos dengan Algoritma SVM Linier