

**ANALISIS SENTIMEN KEBIJAKAN MERDEKA
BELAJAR KAMPUS MERDEKA MENGGUNAKAN
MESIN VEKTOR PENDUKUNG DENGAN
EKSTRAKSI FITUR WORD2VEC DAN PEMODELAN
TOPIK *LATENT DIRICHLET ALLOCATION***

SKRIPSI



NURUL REZKI

H051181026

**PROGRAM STUDI STATISTIKA DEPARTEMEN STATISTIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS HASANUDDIN
MAKASSAR
AGUSTUS 2022**

**ANALISIS SENTIMEN KEBIJAKAN MERDEKA
BELAJAR KAMPUS MERDEKA MENGGUNAKAN
MESIN VEKTOR PENDUKUNG DENGAN
EKSTRAKSI FITUR WORD2VEC DAN PEMODELAN
TOPIK *LATENT DIRICHLET ALLOCATION***

SKRIPSI

**Diajukan sebagai salah satu syarat memperoleh gelar Sarjana Sains pada
Program Studi Statistika Departemen Statistika Fakultas Matematika dan
Ilmu Pengetahuan Alam Universitas Hasanuddin**

NURUL REZKI

H051181026

**PROGRAM STUDI STATISTIKA DEPARTEMEN STATISTIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS HASANUDDIN**

**MAKASSAR
AGUSTUS 2022**

LEMBAR PERNYATAAN KEOTENTIKAN

Saya yang bertanda tangan di bawah ini menyatakan dengan sungguh-sungguh bahwa skripsi yang saya buat dengan judul:

Analisis Sentimen Kebijakan Merdeka Belajar Kampus Merdeka Menggunakan Mesin Vektor Pendukung dengan Ekstraksi Fitur Word2Vec dan Pemodelan Topic *Latent Dirichlet Allocation*

adalah benar hasil karya saya sendiri, bukan hasil plagiat dan belum pernah dipublikasikan dalam bentuk apapun.

Makassar, 16 Agustus 2022


Nurul Rezki
NIM H051181026

**ANALISIS SENTIMEN KEBIJAKAN MERDEKA BELAJAR
KAMPUS MERDEKA MENGGUNAKAN MESIN VEKTOR
PENDUKUNG DENGAN EKSTRAKSI FITUR WORD2VEC
DAN PEMODELAN TOPIK *LATENT DIRICHLET*
*ALLOCATION***

Disetujui Oleh:

Pembimbing Utama,



Sri Astuti Thamrin, S.Si., M.Stat., Ph.D.

NIP. 19740713 199903 2 001

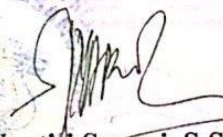
Pembimbing Pertama,



Siswanto, S.Si., M.Si.

NIP. 19920107 201903 1 012

Ketua Departemen Statistika



Dr. Nurtiti Sunusi, S.Si., M.Si.

NIP. 19720117 199703 2 002

Pada 16 Agustus 2022

HALAMAN PENGESAHAN

Skripsi ini diajukan oleh:

Nama : Nurul Rezki
NIM : H051181026
Program Studi : Statistika
Judul Skripsi : Analisis Sentimen Kebijakan Merdeka Belajar Kampus Merdeka Menggunakan Mesin Vektor Pendukung dengan Ekstraksi Fitur Word2vec dan Pemodelan Topik *Latent Dirichlet Allocation*

Telah berhasil dipertahankan di hadapan Dewan Penguji dan diterima sebagai bagian persyaratan yang diperlukan untuk memperoleh gelar Sarjana Sains pada Program Studi Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Hasanuddin.

DEWAN PENGUJI

1. Ketua : Sri Astuti Thamrin, S.Si., M.Stat., Ph.D. (.....)
2. Sekretaris : Siswanto, S.Si., M.Si. (.....)
3. Anggota : Dr. Anna Islamiyati, S.Si., M.Si. (.....)
4. Anggota : Andi Kresna Jaya, S.Si., M.Si. (.....)

Ditetapkan di : Makassar

Tanggal : 16 Agustus 2022

KATA PENGANTAR

Assalamu'alaikum Warahmatullahi Wabarakatuh

Segala puji hanya milik Allah *Subhanallahu Wa Ta'ala* atas segala limpahan rahmat dan hidayah-Nya yang telah diberikan kepada penulis sampai saat ini. Shalawat dan salam senantiasa tercurahkan kepada baginda Rasulullah *Shallallahu 'Alaihi Wa sallam. Alhamdulillahirobbil'aalamiin*, berkat rahmat dan kemudahan yang diberikan oleh Allah *Subhanallahu Wa Ta'ala*, penulis dapat menyelesaikan skripsi yang berjudul “Analisis Sentimen Kebijakan Merdeka Belajar Kampus Merdeka Menggunakan Mesin Vektor Pendukung dengan Ekstraksi Fitur Word2vec dan Pemodelan Topik *Latent Dirichlet Allocation*” sebagai salah satu syarat memperoleh gelar sarjana pada Program Studi Statistika Departemen Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Hasanuddin.

Dalam penyelesaian skripsi ini, penulis telah melewati perjuangan panjang dan pengorbanan yang tidak sedikit. Namun berkat rahmat dan izin-Nya serta dukungan dari berbagai pihak yang turut membantu dalam bentuk moril maupun material sehingga akhirnya tugas akhir ini dapat terselesaikan. Oleh karena itu, penulis menyampaikan ucapan terima kasih yang setinggi-tingginya dan penghargaan yang tak terhingga kepada Ayahanda Sunre dan Ibunda tercinta Habibah yang telah membesarkan dan mendidik penulis dengan penuh kesabaran dan dengan limpahan cinta, kasih sayang dan doa kepada penulis, saudara-saudara penulis yaitu Sappe dan Masati serta kakak ipar Siti Aminah yang selalu membantu dan menjadi penyemangat untuk segera menyelesaikan masa studi penulis, serta keponakan tersayang penulis yaitu Muhammad Nur Fajri, Muhammad Afrijal, Muhammad Nurfadilah dan Ahmad Raihan Ashap yang selalu menghibur penulis.

Penghargaan yang tulus dan ucapan terima kasih dengan penuh keikhlasan juga penulis ucapkan kepada:

1. Bapak Prof. Dr. Ir. Jamaluddin Jompa, M.Sc., selaku Rektor Universitas Hasanuddin beserta seluruh jajarannya.
2. Bapak Dr. Eng. Amiruddin, selaku Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Hasanuddin beserta seluruh jajarannya.

3. Ibu Dr. Nurtiti Sunusi S.Si., M.Si., selaku Ketua Departemen Statistika, segenap Dosen Pengajar dan Staf yang telah membekali ilmu dan kemudahan kepada penulis dalam berbagai hal selama menjadi mahasiswa di Departemen Statistika.
4. Ibu Sri Astuti Thamrin, S.Si., M.Stat., Ph.D. selaku pembimbing utama sekaligus penasehat akademik penulis yang telah ikhlas meluangkan waktu dan pemikirannya untuk memberikan arahan, pengetahuan, motivasi dan bimbingan di tengah kesibukannya.
5. Bapak Siswanto, S.Si., M.Si. selaku pembimbing pertama penulis yang telah meluangkan waktunya di tengah kesibukan untuk memberikan arahan, pengetahuan, motivasi, bimbingan untuk penulis.
6. Ibu Dr. Anna Islamiyati, S.Si., M.Si. dan Bapak Andi Kresna Jaya, S.Si., M.Si. selaku tim penguji yang telah memberikan saran dan kritikan yang membangun dalam penyempurnaan penyusunan tugas akhir ini.
7. Sahabat tercinta penulis, A. Annisa Miftahul Sakinah, Alfiana Wahyuni, Fitra Damayanti, Naura Alfatiyya Arda dan Marsya Anggun Prisila yang telah menjadi sahabat terbaik yang senantiasa mendengarkan keluhan, memberikan dorongan, semangat dan motivasi dalam setiap keadaan sehingga penulis bisa mendapatkan lebih banyak inspirasi khususnya dalam menjalani pahit manisnya kehidupan perkuliahan.
8. Sahabat penulis sejak SMA, Rabihul Fauziah dan Zahra Nur Azizah, yang senantiasa menjadi rumah kedua bagi penulis di tengah kesibukan masing-masing.
9. Sahabat penulis, Yustika dan Sri Indriani Amil yang senantiasa memberikan dorongan serta senantiasa hadir untuk berbagi suka duka masa perkuliahan.
10. Sahabat “Ukhti”, Andi Sri Yulianti, Musdalifah dan Nurhidayah L yang senantiasa memberikan energi positif dan pembelajaran hidup yang berarti bagi penulis.
11. Komunitas Data Science Indonesia khususnya kepada Alfi Fauzia Hanifah, keluarga besar Data Science Indonesia Chapter Sulsel serta teman-teman alumni program MBKM Studi Independen PT. Microsoft Indonesia *learning track Data and Artificial Intelligence* terkhusus kelas DAI-006 yang

senantiasa memberikan dukungan dan inspirasi bagi penulis dalam menyelesaikan tugas akhir ini.

12. Teman-teman Statistika 2018, terkhusus kepada Adhiyaksa Prananda RS, Fadhil Al-Anshory, La Ade, Nehemia Millenium Payung, Taufiq Akbar, Viktor Liman, Hajratul Ashwad K dan Nur Anugrah Yusuf yang selalu membantu dan menjadi sosok guru bagi penulis, terima kasih atas kebersamaan, suka dan duka selama menjalani pendidikan di Departemen Statistika.
13. Keluarga besar INTEGRAL 2018, terkhusus kepada Akidah Amaliah, Juni Wahdaniyah, Abdul Jalil Saleh, Muh. Lutfi, Ahmad Ilham B dan Ardi S, terima kasih telah memberikan pelajaran yang berharga dan arti kebersamaan selama ini kepada penulis, pengalaman yang berharga telah penulis dapatkan dari teman-teman selama berproses bersama.
14. Keluarga Mahasiswa FMIPA Unhas terkhusus anggota keluarga Himatika FMIPA Unhas dan Himastat FMIPA Unhas, terima kasih atas ilmu dan telah menjadi keluarga selama penulis kuliah di Universitas Hasanuddin.
15. Kepada seluruh pihak yang tidak dapat penulis sebutkan satu persatu, terima kasih setinggi-tingginya untuk segala dukungan dan partisipasi yang diberikan kepada penulis semoga bernilai ibadah di sisi Allah *Subhanahu Wa Ta'ala*.

Penulis menyadari bahwa masih banyak kekurangan dalam skripsi ini, untuk itu dengan segala kerendahan hati penulis memohon maaf. Akhir kata, semoga tulisan ini memberikan manfaat untuk pembaca.

Wassalamu'alaikum Warahmatullahi Wabarakatuh

Makassar, 16 Agustus 2022



Nurul Rezki

**PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR
UNTUK KEPENTINGAN AKADEMIK**

Sebagai civitas akademik Universitas Hasanuddin, saya yang bertanda tangan di bawah ini:

Nama : Nurul Rezki
NIM : H051181026
Program Studi : Statistika
Departemen : Statistika
Fakultas : Matematika dan Ilmu Pengetahuan Alam
Jenis Karya : Skripsi

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Hasanuddin **Hak Bebas Royalti Non-eksklusif (*Non-exclusive Royalty- Free Right*)** atas tugas akhir saya yang berjudul:

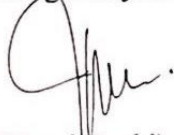
“Analisis Sentimen Kebijakan Merdeka Belajar Kampus Merdeka Menggunakan Mesin Vektor Pendukung dengan Ekstraksi Fitur Word2Vec dan Pemodelan Topik *Latent Dirichlet Allocation*”

Beserta perangkat yang ada (jika diperlukan). Terkait dengan hal di atas, maka pihak universitas berhak menyimpan, mengalih-media/format-kan, mengelola dalam bentuk pangkalan data (*database*), merawat dan memublikasikan tugas akhir saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta.

Demikian pernyataan ini saya buat dengan sebenarnya.

Dibuat di Makassar pada tanggal 16 Agustus 2022.

Yang menyatakan



(Nurul Rezki)

ABSTRAK

Analisis sentimen merupakan analisis data teks yang mengklasifikasikan data ke dalam sentimen positif dan negatif. Analisis sentimen dapat dilanjutkan dengan pemodelan topik untuk merepresentasikan topik yang dibahas pada setiap kelas sentimen. Penelitian ini bertujuan untuk memperoleh hasil klasifikasi sentimen terkait kebijakan Merdeka Belajar Kampus Merdeka di Twitter serta pemodelan topik pada kelas sentimen positif dan negatif. Algoritma klasifikasi yang digunakan adalah mesin vektor pendukung dengan ekstraksi fitur Word2Vec, sedangkan metode pemodelan topik menggunakan *Latent Dirichlet Allocation*. Data pada penelitian ini adalah *tweet* dengan kata kunci “Kampus Merdeka” yang diunggah di Twitter. Diperoleh hasil klasifikasi sentimen terdiri dari 648 *tweet* bersentimen positif dan 931 *tweet* bersentimen negatif dengan akurasi model klasifikasi 89.87%, presisi 91.20%, *recall* 84.44% dan *F-Measure* 87.68%, sedangkan pemodelan topik *Latent Dirichlet Allocation* pada kelas sentimen positif dan negatif masing-masing menghasilkan 5 topik dengan nilai *coherence* 0.44 pada pemodelan topik kelas sentimen positif dan 0.38 pada pemodelan topik kelas sentimen negatif. Klasifikasi sentimen menggunakan mesin vektor pendukung dengan ekstraksi fitur Word2Vec pada penelitian ini menghasilkan model yang baik, namun pemodelan topik *Latent Dirichlet Allocation* menghasilkan topik-topik yang sulit diinterpretasikan karena memiliki nilai *coherence* yang rendah.

Kata Kunci: Analisis Sentimen, Mesin Vektor Pendukung, Word2Vec, *Latent Dirichlet Allocation*, Merdeka Belajar Kampus Merdeka.

ABSTRACT

Sentiment analysis is a data text analysis that classifies data into positive and negative sentiments. Sentiment analysis can be continued with topic modeling to represent the topics discussed in each sentiment class. This research is to get the results of the sentiment classification Merdeka Belajar Kampus Merdeka policy on Twitter and topic modeling results in the positive and negative sentiment class using support vector machine classification algorithm with Word2Vec feature extraction and Latent Dirichlet Allocation topic modeling. Data in this research are tweets with the keyword "Kampus Merdeka" uploaded on Twitter. Sentiment classification results are 648 positive sentiment tweets and 931 negative sentiment tweets with classification model accuracy 89.87%, precision 91.20%, recall 84.44% and F-Measure 87.68%, while the Latent Dirichlet Allocation topic modeling results each 5 topics in the positive and negative sentiment class with 0.44 coherence value on the positive sentiment class topic modeling and 0.38 on the negative sentiment class topic modeling. Sentiment classification with support vector machine and Word2Vec feature extraction resulted a good classification model, but Latent Dirichlet Allocation topic modeling produced topics that were difficult to understand because they had low coherence value.

Keywords: *Sentiment Analysis, Support Vector Machine, Word2Vec, Latent Dirichlet Allocation, Merdeka Belajar Kampus Merdeka Policy.*

DAFTAR ISI

HALAMAN Sampul	i
HALAMAN Judul.....	i
HALAMAN Pernyataan Keotentikan.....	ii
HALAMAN Persetujuan Pembimbing	iv
HALAMAN Pengesahan.....	v
KATA Pengantar	vi
PERSETUJUAN Publikasi Karya Ilmiah.....	viii
ABSTRAK	x
ABSTRACT	xi
DAFTAR ISI.....	xii
DAFTAR GAMBAR	xiv
DAFTAR TABEL.....	xv
DAFTAR LAMPIRAN	xv
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	3
1.3 Tujuan Penelitian.....	3
1.4 Batasan Masalah.....	3
BAB II TINJAUAN PUSTAKA.....	4
2.1 <i>Text Mining</i>	4
2.2 Analisis Sentimen.....	4
2.3 Praproses Data Teks	4
2.4 Ekstraksi Fitur Word2Vec.....	5
2.5 <i>K-fold Cross Validation</i>	9
2.6 Mesin Vektor Pendukung.....	9
2.6.1 Klasifikasi pada <i>Linearly Separable Data</i>	10
2.6.2 Klasifikasi pada <i>Non-Linearly Separable Data</i>	14
2.7 <i>Confusion Matrix</i>	20
2.8 <i>Latent Dirichlet Allocation</i>	22
2.9 Kebijakan Merdeka Belajar Kampus Merdeka	23

2.10	Twitter.....	24
BAB III METODOLOGI PENELITIAN		25
3.1	Sumber Data	25
3.2	Struktur Data	25
3.3	Tahapan Analisis	26
BAB IV HASIL DAN PEMBAHASAN		27
4.1	Deskripsi Data	27
4.2	Praproses Data Teks	29
4.3	Ekstraksi Fitur Word2Vec.....	36
4.4	Klasifikasi Mesin Vektor Pendukung.....	37
4.4.1	<i>Data Training</i> dan <i>Data Testing</i>	37
4.4.2	Klasifikasi Mesin Vektor Pendukung Kernel <i>Radial Basis Function</i>	38
4.4.3	Model Klasifikasi Mesin Vektor Pendukung.....	38
4.4.4	<i>Confusion Matrix</i> pada Mesin Vektor Pendukung	39
4.5	Pemodelan Topik <i>Latent Dirichlet Allocation</i>	42
4.5.1	Pemodelan Topik pada Sentimen Positif.....	42
4.5.2	Pemodelan Topik pada Sentimen Negatif	45
BAB V KESIMPULAN DAN SARAN		48
5.1	Kesimpulan.....	48
5.2	Saran.....	48
DAFTAR PUSTAKA		49
LAMPIRAN.....		54

DAFTAR GAMBAR

Gambar 2.1 Arsitektur CBOW dan *Skip-Gram* 6

Gambar 2.2 Model Mesin Vektor Pendukung Linear 10

Gambar 2.3 Proses Kerja *Probabilistic Graphical Model* 22

Gambar 4.1 *Bar Chart* Kelas Sentimen Data Kebijakan MBKM 28

Gambar 4.2 Nilai *Coherence* Pemodelan Topik Sentimen Positif 44

Gambar 4.3 Nilai *Coherence* Pemodelan Topik Sentimen Negatif..... 46

DAFTAR TABEL

Tabel 2.1	Ilustrasi Pembagian Data dengan <i>K-fold Cross Validation</i>	9
Tabel 2.2	<i>Confusion Matrix</i>	21
Tabel 3.1	Contoh Struktur Data Penelitian	25
Tabel 4.1	Hasil <i>Data Crawling</i>	27
Tabel 4.2	Struktur Data Sebelum Praproses	29
Tabel 4.3	Stuktur Data Sebelum dan Setelah Penghapusan URL	30
Tabel 4.4	Stuktur Data Sebelum dan Setelah Penghapusan <i>Username</i>	30
Tabel 4.5	Stuktur Data Sebelum dan Setelah Penghapusan <i>Hashtag</i>	31
Tabel 4.6	Stuktur Data Sebelum dan Setelah Penghapusan Angka dan Tanda Baca	32
Tabel 4.7	Struktur Data Sebelum dan Setelah Proses <i>Case Folding</i>	33
Tabel 4.8	Struktur Data Sebelum dan Setelah Proses <i>Spelling Normalization</i>	33
Tabel 4.9	Struktur Data Sebelum dan Setelah Proses <i>Stemming</i>	34
Tabel 4.10	Struktur Data Sebelum dan Setelah Proses <i>Stopword Removal</i> ...	35
Tabel 4.11	Struktur Data Sebelum dan Setelah Proses <i>Tokenizing</i>	35
Tabel 4.12	Struktur Data Sebelum dan Setelah Praproses Data Teks	36
Tabel 4.13	Vektor Kata Dasar dengan Ekstraksi Fitur Word2Vec	37
Tabel 4.14	Proporsi <i>Data Training</i> dan <i>Data Testing</i>	37
Tabel 4.15	<i>Confusion Matrix</i> Model Klasifikasi Mesin Vektor Pendukung ..	39
Tabel 4.16	Data FN <i>Tweet</i> Kebijakan MBKM	40
Tabel 4.17	Data FP <i>Tweet</i> Kebijakan MBKM.....	41
Tabel 4.18	Sentimen Positif Terprediksi	43
Tabel 4.19	Pemodelan Topik LDA pada Sentimen Positif Kebijakan MBKM.....	45
Tabel 4.20	Sentimen Negatif Terprediksi	45
Tabel 4.20	Pemodelan Topik LDA pada Sentimen Negatif Kebijakan MBKM.....	47

DAFTAR LAMPIRAN

Lampiran 1	Ketetapan Model Klasifikasi Mesin Vektor Pendukung.....	54
Lampiran 2	Ketetapan Klasifikasi Mesin Vektor Pendukung Kernel RBF ..	63
Lampiran 3	Ketetapan Parameter Pada Pemodelan Topik LDA Sentimen Positif	64
Lampiran 4	Ketetapan Parameter Pada Pemodelan Topik LDA Sentimen Negatif.....	65

BAB I

PENDAHULUAN

1.1 Latar Belakang

Sentimen adalah pendapat seseorang tentang perasaan, sikap atau pikiran yang bisa diungkapkan. Sentimen individu terhadap suatu peristiwa, merek, produk atau perusahaan tertentu dapat diperoleh dari laporan berita, ulasan pengguna, pembaruan media sosial atau situs *microblogging* (Mäntylä dkk., 2018). Data terkait sentimen yang diperoleh melalui situs-situs web atau media sosial pada umumnya berbentuk teks dan tidak terstruktur.

Text mining merupakan proses ekstraksi pola dari sejumlah data tidak terstruktur dan akan diperoleh pola-pola data, *trend* serta ekstraksi pengetahuan yang potensial dari data teks (Turban, 2011). Salah satu tujuan penggunaan *text mining* adalah analisis sentimen yaitu proses memahami, mengestrak dan mengolah data tekstual untuk mendapatkan informasi sentimen yang terkandung dalam suatu kalimat opini terhadap sebuah masalah atau objek. Informasi Sentimen yang diperoleh dapat berupa sentimen positif atau negatif. (Zhang dkk., 2018).

Salah satu pendekatan untuk melakukan analisis sentimen adalah *supervised learning* yang merupakan proses *machine learning* yang membutuhkan pengawasan melalui persiapan *data training* yang telah melalui proses pelabelan sehingga performa dari pendekatan *supervised learning* sangat bergantung pada *data training* dan cara data tersebut diproses. Beberapa algoritma yang umum digunakan untuk melakukan analisis sentimen dengan pendekatan *supervised learning* adalah mesin vektor pendukung, Naive Bayes dan *Maximum Entropy*. Mesin vektor pendukung adalah salah satu algoritma yang banyak digunakan pada penelitian-penelitian terkait analisis sentimen karena menghasilkan akurasi yang baik dibandingkan algoritma atau metode klasifikasi yang lain. Salah satunya adalah analisis sentimen terhadap ulasan sebuah maskapai penerbangan menggunakan metode mesin vektor pendukung dan Naive Bayes yang menghasilkan akurasi tertinggi sebesar 82.48% untuk metode mesin vektor pendukung (Rahat dkk., 2020). Penelitian terkait lainnya yaitu “Analisis Sentimen Media Sosial Universitas Amikom Yogyakarta Sebagai Sarana Penyebaran Informasi Menggunakan Algoritma Klasifikasi SVM” (Maulana dkk., 2018) dan

“Sentimen Analisis Publik Terhadap Kebijakan *Lockdown* Pemerintah Jakarta Menggunakan Algoritma SVM” (Isnain dkk., 2021).

Data sentimen berupa data teks yang dianalisis dengan analisis sentimen perlu diubah ke bentuk numerik agar dapat dibaca oleh komputer. Mengubah data teks menjadi data numerik dapat dilakukan dengan ekstraksi fitur (Pravina dkk., 2019). Terdapat beberapa jenis ekstraksi fitur seperti *Term Frequency* (TF), *Term Frequency Inverse Document Frequency* (TF-IDF), *Word2Vec*, *Glove* dan sebagainya (Rusli dkk., 2019). Pada penelitian ini, ekstraksi fitur yang digunakan adalah *Word2Vec* yang merepresentasikan sebuah kata pada bentuk vektor (Aldiansyah dkk., 2019). Pada penelitian yang dilakukan oleh Naufal dan Setiawan (2021) tentang analisis sentimen di Twitter terkait kebijakan publik, dibandingkan dengan TF-IDF, ekstraksi fitur *Word2Vec* memberikan peningkatan akurasi pada metode SVM.

Analisis sentimen dapat dilanjutkan dengan pemodelan topik yaitu proses merepresentasikan topik yang dibahas dalam dokumen teks (Febrianta dkk., 2021). *Latent Dirichlet Allocation* (LDA) merupakan metode pemodelan topik populer yang digunakan untuk meringkas, melakukan klusterisasi, menghubungkan ataupun memproses data yang menghasilkan daftar topik (Rachman dan Pramana, 2020). Beberapa penelitian terkait LDA diantaranya adalah “*Exploring Public Response to COVID-19 on Weibo with LDA Topic Modeling and Sentiment Analysis*” (Xie dkk., 2020) dan “Analisis Sentimen Pro dan Kontra Masyarakat Indonesia tentang Vaksin COVID-19 pada Media Sosial Twitter” (Rachman dan Pramana, 2020).

Salah satu isu yang saat ini ramai dibicarakan di berbagai media sosial maupun situs-situs web adalah kebijakan Merdeka Belajar Kampus Merdeka (MBKM) oleh Kementerian Pendidikan dan Kebudayaan Republik Indonesia. Berdasarkan artikel yang dirilis pada 29 Januari 2020 oleh *tirto.id*, sejak awal diluncurkan, program ini telah banyak mendapat reaksi pro maupun kontra oleh masyarakat (Prabowo, 2020). Reaksi tersebut disampaikan melalui media sosial seperti Twitter, Instagram, Facebook dan sebagainya. Twitter merupakan salah satu media sosial yang banyak mendapatkan perhatian masyarakat Indonesia. Selain itu, Twitter menyediakan *Application Programming Interface* (API) bagi pengguna

untuk mengembangkan sebuah aplikasi dengan data yang bersumber dari Twitter (Blanchette, 2008).

Berdasarkan uraian yang telah dipaparkan, maka pada penelitian tugas akhir ini akan dilakukan analisis sentimen terkait kebijakan MBKM di Twitter dengan metode mesin vektor pendukung menggunakan ekstraksi fitur Word2Vec dan pemodelan topik LDA.

1.2 Rumusan Masalah

Berdasarkan latar belakang, maka rumusan masalah dari penelitian ini adalah sebagai berikut:

1. Bagaimana hasil klasifikasi sentimen kebijakan MBKM di Twitter menggunakan mesin vektor pendukung dengan ekstraksi fitur Word2Vec?
2. Bagaimana hasil pemodelan topik LDA pada sentimen positif dan negatif terkait kebijakan MBKM di Twitter?

1.3 Tujuan Penelitian

Berdasarkan rumusan masalah, maka tujuan penelitian ini adalah sebagai berikut:

1. Memperoleh hasil klasifikasi sentimen kebijakan MBKM di Twitter menggunakan mesin vektor pendukung dengan ekstraksi fitur Word2Vec.
2. Memperoleh hasil pemodelan topik LDA pada sentimen positif dan negatif terkait kebijakan MBKM di Twitter.

1.4 Batasan Masalah

Batasan masalah penelitian ini adalah menggunakan ekstraksi fitur Word2Vec dengan arsitektur *Skip-Gram* dan algoritma *negative sampling*.

BAB II

TINJAUAN PUSTAKA

2.1 Text Mining

Text mining adalah proses analisis terhadap data teks yang bersumber dari dokumen. Konsep *text mining* digunakan untuk melakukan klasifikasi dokumen tekstual agar dokumen-dokumen tersebut dapat diklasifikasikan sesuai dengan topik yang diinginkan. Dengan menggunakan *text mining*, kategori dokumen teks dapat diketahui melalui kata-kata yang terkandung di dalamnya (Herwijayanti dkk., 2018).

2.2 Analisis Sentimen

Analisis sentimen adalah studi komputasi opini, sentimen, emosi, penilaian dan sikap individu terhadap entitas seperti produk, layanan, organisasi, individu lain, masalah, peristiwa, topik serta atribut yang berkaitan (Zhang dkk., 2018). Tugas dasar analisis sentimen adalah mengelompokkan teks yang ada dalam sebuah kalimat atau dokumen lalu menentukan pendapat yang dikemukakan dalam kalimat atau dokumen tersebut bersifat positif atau negatif (Kurniawan, 2017).

2.3 Praproses Data Teks

Praproses merupakan proses mempersiapkan data yang belum terstruktur menjadi data yang terstruktur sehingga dapat digunakan untuk proses berikutnya (Ipmawati dkk., 2017). Berikut tahapan-tahapan dari praproses data teks:

1. *Cleaning*, yaitu membersihkan data dari *noise* seperti *hashtag*, *username*, URL, dan tanda baca (Pertiwi, 2019).
2. *Case folding*, yaitu proses penyamaan *case* dalam sebuah dokumen. Tidak semua dokumen teks konsisten dalam penggunaan huruf kapital. Oleh karena itu peran *case folding* dibutuhkan dalam mengkonversi keseluruhan teks dalam dokumen menjadi suatu bentuk standar (dalam hal ini huruf kecil atau *lowercase*).
3. *Spelling normalization*, merupakan proses perbaikan atau substitusi kata-kata yang salah eja atau disingkat dalam bentuk tertentu. Substitusi kata dilakukan

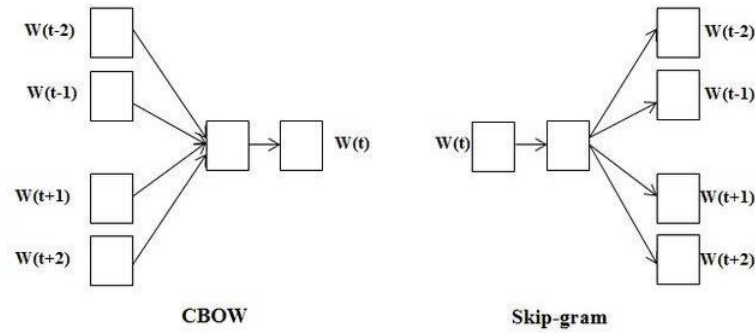
untuk menghindari jumlah perhitungan dimensi kata yang melebar (Khotimah, 2019).

4. *Stopword removal*, yaitu menghapus kata-kata yang sangat umum. Kata yang termasuk dalam *stopword* adalah yang, dan, di, itu, dengan, untuk, dari, dalam, akan, pada, ini, juga, saya, serta, adalah, bahwa, lain, kamu, dan masih banyak lagi (Jumeilah, 2017).
5. *Tokenizing*, yaitu proses memecah kalimat menjadi kata-kata. Proses tokenisasi mengandalkan karakter spasi pada dokumen teks untuk melakukan pemisahan. Hasil dari proses ini adalah kumpulan kata (Khotimah, 2019).
6. *Stemming* adalah tahapan mengubah kata menjadi kata dasar menurut kaidah Bahasa Indonesia yang benar dengan menghilangkan imbuhan, awalan serta akhiran kata (Yulita, 2021).

2.4 Ekstraksi Fitur Word2Vec

Ekstraksi fitur adalah proses mengubah kata dari suatu teks menjadi nilai numerik yang dapat dibaca oleh komputer (Pravina dkk., 2019). Word2Vec merupakan salah satu jenis ekstraksi fitur yang mengubah kata menjadi vektor. Kata-kata yang memiliki arti yang cukup mirip akan memiliki *output* yang berdekatan satu sama lain sehingga hubungan semantik antar kata dapat dipelajari (Triyantono dkk., 2021). Word2Vec merupakan model *neural network* dengan kata teks sebagai *input* dan ruang vektor sebagai *output*.

Terdapat dua jenis model arsitektur Word2Vec, yaitu *Skip-Gram* dan *Continuous Bag of Words* (CBOW) yang ditunjukkan pada Gambar 2.1. Arsitektur model *Skip-Gram* memprediksi kata konteks berdasarkan kata tengah, sedangkan model CBOW memprediksi kata tengah berdasarkan kata konteks (Rahman dkk., 2021). Kecepatan *training* CBOW lebih cepat dibandingkan dengan *Skip-Gram*, tetapi *Skip-Gram* lebih akurat dibandingkan dengan CBOW (Chang dkk., 2017).



Gambar 2.1 Arsitektur CBOW dan *Skip-Gram*

Berdasarkan Gambar 2.1, *input* dari model *Skip-Gram* adalah kata tengah dan *output* berupa kata konteks. Tujuannya adalah memaksimalkan peluang memprediksi kata konteks berdasarkan kata tengah dengan mengoptimalkan matriks pembobot. Secara matematis dapat dituliskan sebagai berikut:

$$\arg \max_{\theta} \ln P(w_c | w_t; \theta)$$

dengan θ sebagai matriks pembobot yang dioptimalkan, w_c adalah kumpulan kata konteks dan w_t adalah kata tengah. Matriks pembobot θ diinisiasi secara acak dan akan dioptimalkan pada model *neural network*. Secara matematis, optimasi matriks pembobot θ dituliskan sebagai berikut:

$$\theta^{new} = \theta^{old} - \eta \nabla J(\theta)$$

dengan θ^{new} adalah matriks pembobot baru yang telah dioptimasi, θ^{old} adalah matriks pembobot yang dioptimasi, η adalah *learning rate*, $\nabla J(\theta)$ gradien matriks pembobot dan $J(\theta)$ adalah fungsi tujuan.

Algoritma *gradien derivate* digunakan untuk mendapatkan turunan fungsi tujuan sehubungan dengan matriks pembobot θ . Secara umum, fungsi tujuan dari model *Skip-Gram* diperoleh dengan algoritma *softmax* yang secara matematis ditunjukkan pada persamaan berikut:

$$J(\theta) = -\frac{1}{T} \sum_{t=1}^T \sum_{-C \leq j \leq C, j \neq 0} \ln P(w_c | w_t; \theta)$$

dengan:

$J(\theta)$: Fungsi Tujuan

T, t : Indeks kata unik pada *corpus*

w_t : Kata tengah

w_c : Kata konteks

C : Ukuran *window*

Fungsi tujuan dengan algoritma *softmax* dinilai tidak efektif secara komputasi karena ukuran indeks T sangat besar. Untuk mengatasi hal tersebut, fungsi tujuan dapat diperoleh dengan algoritma *negative sampling* menggunakan fungsi sigmoid. Fungsi sigmoid adalah fungsi yang digunakan untuk regresi logistik biner dengan persamaan umum sebagai berikut:

$$\sigma(x) = \frac{1}{1 + \exp(-x)} \quad (2.1)$$

dengan $\sigma(x)$ adalah fungsi sigmoid dari x .

Seluruh kata dalam *corpus* akan dikategorikan menjadi dua kelas, yaitu positif dan negatif. Kata konteks akan dikategorikan sebagai kata positif dan kata lainnya akan dikategorikan sebagai kata negatif. Dengan algoritma *negative sampling*, untuk setiap kata konteks dari sebuah kata tengah, akan ditarik sebanyak K kata negatif sehingga akan lebih efektif secara komputasi karena hanya sebanyak $(K + 1)$ untuk setiap kata positif pada matriks pembobot θ yang akan dioptimasi. Model *Skip-Gram* dengan algoritma *negative sampling* akan memaksimalkan peluang kata positif dan meminimalkan peluang kata negatif untuk setiap kata tengah. Secara matematis dituliskan sebagai berikut:

$$\begin{aligned} & \arg \max_{\theta} \ln \left(P(D = 1 | w, c_{pos}; \theta) \prod_{c_{neg}} P(D = 0 | w, c_{neg}; \theta) \right) \\ & \arg \max_{\theta} \ln \left(P(D = 1 | w, c_{pos}; \theta) \prod_{c_{neg}} 1 - P(D = 1 | w, c_{neg}; \theta) \right) \\ & \arg \max_{\theta} \ln P(D = 1 | w, c_{pos}; \theta) + \prod_{c_{neg}} \ln \left(1 - P(D = 1 | w, c_{neg}; \theta) \right) \\ & \arg \max_{\theta} \ln P(D = 1 | w, c_{pos}; \theta) + \sum_{c_{neg}} \ln \left(1 - P(D = 1 | w, c_{neg}; \theta) \right) \quad (2.2) \end{aligned}$$

dengan K merupakan indeks maksimum kata negatif. Menggunakan fungsi sigmoid pada Persamaan 2.1, maka Persamaan 2.2 dapat dituliskan sebagai berikut:

$$\begin{aligned} & \arg \max_{\theta} \ln \left(\frac{1}{1 + \exp(-c_{pos}(w))} \right) + \sum_{c_{neg}} \ln \left(1 - \frac{1}{1 + \exp(-c_{neg}(w))} \right) \\ & \arg \max_{\theta} \ln \left(\frac{1}{1 + \exp(-c_{pos}(w))} \right) + \sum_{c_{neg}} \ln \left(\frac{1}{1 + \exp(c_{neg}(w))} \right) \\ & \arg \max_{\theta} \ln \sigma \left(c_{pos}(w) \right) + \sum_{c_{neg}} \ln \sigma \left(-c_{neg}(w) \right) \end{aligned}$$

Sehingga fungsi tujuan untuk algoritma *negative sampling* dengan fungsi sigmoid adalah sebagai berikut:

$$J(\boldsymbol{\theta}; \mathbf{w}, c_{pos}) = -\ln \sigma(\mathbf{c}_{pos}(\mathbf{w})) - \sum_{c_{neg}} \ln \sigma(-\mathbf{c}_{neg}(\mathbf{w}))$$

dengan \mathbf{w} adalah vektor kata input yang ekuivalen dengan \mathbf{h} yang merupakan *projection layer* pada model *neural network* dengan ukuran dimensi N yang diinisiasi. Optimasi matriks pembobot $\boldsymbol{\theta}$ pada model *Skip-Gram* dengan algoritma *negative sampling* adalah sebagai berikut:

$$\boldsymbol{\theta}^{new} = \boldsymbol{\theta}^{old} - \eta \frac{\partial J}{\partial \boldsymbol{\theta}}$$

Penurunan fungsi sigmoid adalah Nilai $\frac{\partial \sigma}{\partial x} = \sigma(x)(1 - \sigma(x))$ sehingga nilai $\frac{\partial J}{\partial \boldsymbol{\theta}}$ adalah sebagai berikut:

$$\frac{\partial J}{\partial \boldsymbol{\theta}} = \left(\sigma(\mathbf{c}_{pos}(\mathbf{h})) - 1 \right) \frac{\partial(\mathbf{c}_{pos}(\mathbf{h}))}{\partial \boldsymbol{\theta}} + \sum_{c_{neg}} 1 - \sigma(-\mathbf{c}_{neg}(\mathbf{h})) \frac{\partial(-\mathbf{c}_{neg}(\mathbf{h}))}{\partial \boldsymbol{\theta}}$$

Karena $\boldsymbol{\theta}$ terdiri dari \mathbf{W}_{input} berukuran $V \times N$ dan \mathbf{W}_{output} berukuran $N \times V$ dengan V merupakan ukuran dimensi korpus yang digunakan, maka fungsi tujuan perlu dibedakan menjadi $\frac{\partial J}{\partial \mathbf{W}_{input}}$ dan $\frac{\partial J}{\partial \mathbf{W}_{output}}$ yang diuraikan sebagai berikut (Goldberg dan Levy, (2014):

Kondisi untuk \mathbf{W}_{input}

Turunan fungsi tujuan pada \mathbf{W}_{input} dapat dituliskan sebagai berikut:

$$\frac{\partial J}{\partial \mathbf{h}} = \left(\sigma(\mathbf{c}_{pos}(\mathbf{h})) - 1 \right) \mathbf{c}_{pos} + \sum_{c_{neg}} 1 - \sigma(-\mathbf{c}_{neg}(\mathbf{h})) \mathbf{c}_{neg}$$

Maka optimasi nilai vektor kata input secara matematis ditunjukkan sebagai berikut:

$$\mathbf{w}^{(new)} = \mathbf{w}^{(old)} - \eta \left(\left(\sigma(\mathbf{c}_{pos}(\mathbf{h})) - 1 \right) \mathbf{c}_{pos} + \sum_{c_{neg}} \left(1 - \sigma(-\mathbf{c}_{neg}(\mathbf{h})) \right) \mathbf{c}_{neg} \right)$$

Kondisi untuk \mathbf{W}_{output}

Sebanyak $K + 1$ vektor kata pada \mathbf{W}_{output} akan diperbaharui yang terdiri dari K kata negatif dan 1 kata positif yaitu kata konteks. Turunan fungsi tujuan untuk kata positif dan kata negatif pada \mathbf{W}_{output} adalah sebagai berikut:

$$\frac{\partial J}{\partial \mathbf{c}_{pos}} = \left(\sigma(\mathbf{c}_{pos}(\mathbf{h})) - 1 \right) \mathbf{h}$$

$$\frac{\partial J}{\partial \mathbf{c}_{neg}} = 1 - \sigma(-\mathbf{c}_{neg}(\mathbf{h})) \mathbf{h}$$

Sehingga optimasi vektor kata pada \mathbf{W}_{output} secara matematis dituliskan sebagai berikut:

$$\mathbf{c}_{pos}^{(new)} = \mathbf{c}_{pos}^{(old)} - \eta \left(\sigma(\mathbf{c}_{pos}(\mathbf{h})) - 1 \right) \mathbf{h}$$

$$\mathbf{c}_{neg}^{(new)} = \mathbf{c}_{neg}^{(old)} - \eta \left(1 - \sigma(-\mathbf{c}_{neg}(\mathbf{h})) \right) \mathbf{h}$$

2.5 K-fold Cross Validation

K-fold cross validation adalah salah satu metode yang digunakan untuk mempartisi data menjadi *data training* dan *data testing*. *K-fold cross validation* secara berulang-ulang membagi data menjadi *data training* dan *data testing* sehingga setiap data memperoleh kesempatan menjadi *data testing* (Alkaff dkk., 2021). *K* merupakan besar angka partisi data yang digunakan untuk *data training* dan *data testing*. Menurut Max Kuhn & Kjell Johnson (2013) tidak ada aturan formal dalam penentuan nilai *K*, namun nilai *K* yang bagus digunakan adalah 5 hingga 10. Tabel 2.1 menunjukkan ilustrasi pembagian data menggunakan *K-fold cross validation*.

Tabel 2.1 Ilustrasi Pembagian Data dengan *K-fold Cross Validation*

Percobaan 1	<i>Test</i>	<i>Train</i>	<i>Train</i>	<i>Train</i>	<i>Train</i>
Percobaan 2	<i>Train</i>	<i>Test</i>	<i>Train</i>	<i>Train</i>	<i>Train</i>
Percobaan 3	<i>Train</i>	<i>Train</i>	<i>Test</i>	<i>Train</i>	<i>Train</i>
Percobaan 4	<i>Train</i>	<i>Train</i>	<i>Train</i>	<i>Test</i>	<i>Train</i>
Percobaan 5	<i>Train</i>	<i>Train</i>	<i>Train</i>	<i>Train</i>	<i>Test</i>

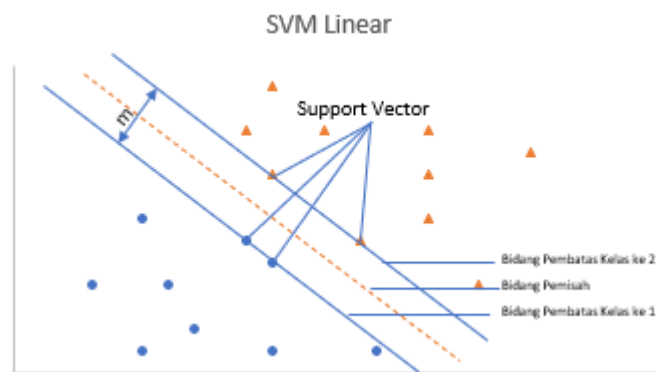
2.6 Mesin Vektor Pendukung

Mesin vektor pendukung merupakan salah satu metode klasifikasi pada *machine learning* dengan pendekatan *supervised learning* yang memprediksi kelas berdasarkan model atau pola dari hasil proses *training*. Klasifikasi dilakukan dengan mencari *hyperplane* atau garis pembatas (*decision boundary*) yang memisahkan satu kelas dengan kelas yang lain. *Hyperplane* yang optimal adalah

hyperplane yang memberikan margin optimal antara suatu sampel *data training* yang paling dekat dengan *hyperplane*. Kumpulan sampel *data training* disebut dengan *support vector* (Lutfiyanto, 2021). Mesin vektor pendukung digunakan untuk menyelesaikan masalah klasifikasi linier dan non-linier (Sari, 2017).

2.6.1 Klasifikasi pada *Linearly Separable Data*

Mesin vektor pendukung pada *linearly separable data* adalah penerapan metode mesin vektor pendukung pada data yang dapat dipisahkan secara linier. Misalkan $x_i = \{x_1, x_2, \dots, x_n\}$ adalah titik data dan $y_i \in \{-1, +1\}$ adalah label kategori atau kelas data untuk *dataset*. Penggambaran *linearly separable data* dapat dilihat pada Gambar 2.2 berikut.



Gambar 2.2 Model Mesin Vektor Pendukung Linear

Pada Gambar 2.1, kedua kelas data dapat dipisahkan oleh sepasang bidang pembatas yang sejajar (linier). Data yang berada pada bidang pembatas disebut dengan *support vector*. Persamaan *hyperplane* sebagai berikut:

$$f(x) = (\mathbf{w}^T \mathbf{x}) + b$$

$$f(x) = w_1x_1 + w_2x_2 + \dots + w_Nx_N + b$$

dengan \mathbf{w} merupakan vektor normal terhadap *hyperplane*, b merupakan suatu konstanta skalar yang menentukan lokasi fungsi *hyperplane* terhadap titik asal (Sari, 2017). Oleh karena itu *hyperplane* pemisah pada kasus ini dapat dimodelkan sebagai berikut:

$$f(x) = (\mathbf{w}^T \mathbf{x}) + b$$

Pada kasus ini akan dipisahkan dua *hyperplane* yang sejajar dengan *hyperplane* pertama akan membatasi kelas pertama sedangkan *hyperplane* kedua

akan membatasi kelas kedua, sehingga dapat dibentuk pertidaksamaan model matematika:

$$\mathbf{w}^T \mathbf{x}_i + b \geq 1, \forall_i \in \text{Kelas Positif} \quad (2.3)$$

$$\mathbf{w}^T \mathbf{x}_i + b \leq -1, \forall_i \in \text{Kelas Negatif} \quad (2.4)$$

Berdasarkan Persamaan 2.3 dan 2.4 dapat diperoleh nilai margin diantara dua *hyperplane* dengan menggunakan prinsip jarak antara dua garis sejajar pada persamaan berikut:

$$\begin{aligned} \text{margin} (m) &= \frac{|(b - 1) - (b + 1)|}{\sqrt{\mathbf{w}^T \mathbf{w}}} \\ &= \frac{|-2|}{\|\mathbf{w}\|} \\ &= \frac{2}{\|\mathbf{w}\|} \end{aligned}$$

Untuk mendapatkan dua *hyperplane* pemisah dengan jarak sejauh mungkin, nilai dari *hyperplane* optimal dapat diperoleh melalui suatu solusi optimasi.

fungsi tujuan:

$$\max \frac{2}{\|\mathbf{w}\|}$$

dengan kendala:

$$\mathbf{w}^T \mathbf{x}_i + b \geq 1, \forall_i \in \text{Kelas Positif}$$

$$\mathbf{w}^T \mathbf{x}_i + b \leq -1, \forall_i \in \text{Kelas Negatif}$$

atau dapat diringkas sebagai berikut:

fungsi tujuan:

$$\max \frac{2}{\|\mathbf{w}\|} \quad (2.5)$$

dengan kendala:

$$y_i(\mathbf{w}^T \mathbf{x}_i + b) \geq 1, i = 1, 2, 3, \dots, N$$

Menurut Sari, 2017 memaksimalkan margin $\frac{1}{\|\mathbf{w}\|}$ sama dengan meminimumkan

$\frac{1}{2} \|\mathbf{w}\|^2$, maka Persamaan 2.5 dapat ditulis sebagai berikut:

fungsi tujuan:

$$\min \frac{1}{2} \|\mathbf{w}\|^2 \quad (2.6)$$

dengan kendala:

$$y_i(\mathbf{w}^T \mathbf{x}_i + b) \geq 1, i = 1, 2, 3, \dots, N$$

Nilai *hyperplane* optimal yang memisahkan data dengan margin maksimum dapat diperoleh dengan menyelesaikan kendala optimasi kuadratik. Permasalahan optimasi kasus data yang dapat dipisahkan secara linear pada Persamaan 2.6 dapat dituliskan dalam bentuk sebagai berikut:

fungsi tujuan:

$$\min_{\mathbf{w}, b}(\mathbf{w}, b) = \frac{1}{2} \mathbf{w}^T \mathbf{w} \quad (2.7)$$

dengan kendala:

$$y_i(\mathbf{w}^T \mathbf{x}_i + b) - 1 \geq 0, i = 1, 2, 3, \dots, N$$

Untuk mendapatkan solusi dari permasalahan optimasi bentuk primal pada Persamaan 2.7 maka akan digunakan metode *Lagrange* dengan kendala pertidaksamaan atau dikenal sebagai *Karush Kuhn Tucker* (KKT). Berdasarkan Persamaan 2.7 maka diperoleh persamaan *lagrange* untuk kasus ini menjadi:

$$\begin{aligned} L_p(\mathbf{w}, b, \alpha) &= (\mathbf{w}, b) + \sum_{i=1}^n \alpha_i [1 - y_i(\mathbf{w}^T \mathbf{x}_i + b)] \\ L_p(\mathbf{w}, b, \alpha) &= (\mathbf{w}, b) - \sum_{i=1}^n \alpha_i [y_i(\mathbf{w}^T \mathbf{x}_i + b) - 1] \\ L_p(\mathbf{w}, b, \alpha) &= \left(\frac{1}{2} \mathbf{w}^T \mathbf{w}\right) - \sum_{i=1}^n \alpha_i [y_i(\mathbf{w}^T \mathbf{x}_i + b) - 1] \end{aligned} \quad (2.8)$$

Peubah α_i merupakan *Lagrange Multiplier* atau pengali *Lagrange*. Nilai dari *Lagrange Multiplier* ini adalah $\alpha_i > 0$. Fungsi Persamaan 2.8 akan diminimalkan terhadap \mathbf{w} dan \mathbf{b} serta dimaksimalkan terhadap peubah α (Sari, 2017). Turunan pertama dari fungsi pada Persamaan 2.8 adalah sebagai berikut:

Kondisi 1

$$\begin{aligned} \frac{\partial L_p}{\partial b} &= 0 \\ \frac{\partial L_p}{\partial b} &= \frac{\partial L_p(\mathbf{w}, b, \alpha)}{\partial b} \\ &= \frac{\partial \{(\mathbf{w}, b) - \sum_{i=1}^n \alpha_i [y_i(\mathbf{w}^T \mathbf{x}_i + b) - 1]\}}{\partial b} \\ &= \frac{\partial \left\{ \frac{1}{2} \mathbf{w}^T \mathbf{w} - \sum_{i=1}^n \alpha_i [y_i(\mathbf{w}^T \mathbf{x}_i + b) - 1] \right\}}{\partial b} \\ &= 0 - 0 - \sum_{i=1}^n \alpha_i y_i + 0 \end{aligned}$$

$$\sum_{i=1}^n \alpha_i y_i = 0$$

Kondisi 2

$$\frac{\partial L_p}{\partial \mathbf{w}} = 0$$

$$\frac{\partial L_p}{\partial \mathbf{w}} = \frac{\partial L_p(\mathbf{w}, b, \alpha)}{\partial b}$$

$$\frac{\partial L_p}{\partial \mathbf{w}} = \frac{\partial \{(\mathbf{w}, b) - \sum_{i=1}^n \alpha_i [y_i(\mathbf{w}^T \mathbf{x}_i + b) - 1]\}}{\partial \mathbf{w}}$$

$$\frac{\partial L_p}{\partial \mathbf{w}} = \frac{\partial \left\{ \frac{1}{2} \mathbf{w}^T \mathbf{w} - \sum_{i=1}^n \alpha_i [y_i(\mathbf{w}^T \mathbf{x}_i + b) - 1] \right\}}{\partial \mathbf{w}}$$

$$\begin{aligned} \frac{\partial L_p}{\partial \mathbf{w}} &= \frac{\partial \left\{ \frac{1}{2} \mathbf{w}^T \mathbf{w} - \sum_{i=1}^n \alpha_i y_i \mathbf{w}^T \mathbf{x}_i - b \sum_{i=1}^n \alpha_i y_i + \sum_{i=1}^n \alpha_i \right\}}{\partial \mathbf{w}} \\ &= \mathbf{w} - \sum_{i=1}^n \alpha_i y_i \mathbf{x}_i - 0 + 0 \end{aligned}$$

$$\mathbf{w} = \sum_{i=1}^n \alpha_i y_i \mathbf{x}_i \quad (2.9)$$

Tampak bahwa fungsi tujuan pada Persamaan 2.6 mengandung fungsi kuadrat pada peubah \mathbf{w} , sehingga hal tersebut akan mengakibatkan cukup sulitnya penyelesaian secara komputasi dan akan memakan waktu yang panjang. Dengan demikian, maka permasalahan tersebut akan lebih mudah dan lebih efisien jika diselesaikan dalam bentuk dual (Rashif, 2007).

Menurut teorema dualitas Keckman (2005) jika problem primal memiliki solusi optimal, maka problem dual juga akan mempunyai solusi optimal yang nilainya sama. Untuk memperoleh bentuk dual dari Persamaan 2.7, maka akan disubstitusikan Persamaan 2.9 ke dalam Persamaan 2.8 sebagai berikut:

$$\begin{aligned} L_p(\alpha) &= (\mathbf{w}, b) + \sum_{i=1}^n \alpha_i [1 - y_i(\mathbf{w}^T \mathbf{x}_i + b)] \\ &= \frac{1}{2} \mathbf{w}^T \mathbf{w} - \sum_{i=1}^n \alpha_i [y_i(\mathbf{w}^T \mathbf{x}_i + b) - 1] \\ &= \frac{1}{2} \mathbf{w}^T \mathbf{w} - \sum_{i=1}^n \alpha_i y_i \mathbf{w}^T \mathbf{x}_i - b \sum_{i=1}^n \alpha_i y_i + \sum_{i=1}^n \alpha_i \\ &= \frac{1}{2} \mathbf{w}^T \mathbf{w} - \sum_{i=1}^n \alpha_i y_i \mathbf{w}^T \mathbf{x}_i + \sum_{i=1}^n \alpha_i \\ &= \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j \mathbf{x}_i^T \mathbf{x}_j - \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j \mathbf{x}_i^T \mathbf{x}_j + \sum_{i=1}^n \alpha_i \\ &= \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j \mathbf{x}_i^T \mathbf{x}_j \end{aligned}$$

Menurut teori dualitas, meminimumkan $L_p(w)$ sama dengan memaksimumkan $L_p(\alpha)$, sehingga masalah pencarian *hyperplane* yang optimal pada kasus *linear separable* dapat dirumuskan:

fungsi tujuan:

$$\max_{\alpha} L_p(\alpha) = \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j \mathbf{x}_i^T \mathbf{x}_j \quad (2.10)$$

dengan kendala:

$$\sum_{j=1}^n \alpha_j y_j = 0, \forall_i \text{ dengan } \alpha_i > 0$$

$$y_i = \begin{cases} 1, & i > 0 \\ -1, & i < 0 \end{cases}$$

dengan demikian, nilai α_i dapat ditemukan dengan menyelesaikan kasus optimasi pada Persamaan 2.10 dan nilai w akan diperoleh dengan mensubstitusikan α_i pada persamaan 2.9. Selanjutnya fungsi *hyperplane* yang optimal pada kasus *linear separable* dapat terbentuk persaaan berikut:

$$f(x_d) = \sum_{i=1}^n \alpha_i y_i (x_i x_d) + b$$

dengan x_d merupakan data yang akan diklasifikasikan, α_i merupakan solusi optimasi dari masalah optimasi pada Persamaan 2.10 dan b diperoleh dengan cara berikut:

$$b = -\frac{1}{2} (wx^+ + wx^-)$$

dengan demikian, kelas berdasarkan *hyperplane* optimal pada kasus dataset yang *linear separable* terbentuk sebagai berikut:

$$f(x) = (\mathbf{w}^T \mathbf{x}) + b$$

2.6.2 Klasifikasi pada *Non-Linearly Separable Data*

Klasifikasi data yang tidak dapat dipisahkan secara linier memerlukan modifikasi pada formula mesin vektor pendukung agar dapat menemukan solusinya. Dalam praktiknya, jarang ditemui *data training* yang dapat dipisahkan secara linear. Modifikasi formula tersebut dilakukan pada kedua bidang *hyperplane*. Agar lebih fleksibel maka kedua *hyperplane* diberi tambahan peubah *slack* (ξ_i) dengan $\xi_i \geq 0, i = 1, 2, \dots, n$, sehingga diperoleh suatu modifikasi *hyperplane* baru pada persamaan berikut:

$$y_i \left((\mathbf{w}^T \mathbf{x}_i) + b \right) + \xi_i \geq 1$$

$$y_i(\mathbf{w}^T \mathbf{x}_i + b) \geq 1 - \xi_i$$

atau dapat dituliskan:

$$\mathbf{w}^T \mathbf{x}_i + b \geq 1 - \xi_i, \forall \in \text{Kelas Positif}$$

$$\mathbf{w}^T \mathbf{x}_i + b \leq -1 + \xi_i, \forall \in \text{Kelas Negatif}$$

Pencarian *hyperplane* optimal dengan tambahan peubah *slack* ini sering disebut sebagai *soft margin hyperplane*. Selain itu, pada modifikasi formula untuk data yang tidak dapat dipisahkan secara linier juga memerlukan tambahan parameter *penalty C* (Sari, 2017). Sehingga formula untuk mendapatkan *hyperplane* optimal menjadi:

fungsi tujuan:

$$\min \left(\frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^n \xi_i \right) \quad (2.11)$$

dengan kendala:

$$y_i(\mathbf{w}^T \mathbf{x}_i + b) \geq 1 - \xi_i, i = 1, 2, 3 \dots, n$$

dengan C merupakan parameter yang menentukan besarnya penalti akibat kesalahan dalam klasifikasi data yang ditentukan oleh pengguna. Semakin tinggi nilai C , maka kemungkinan terjadinya kesalahan dalam penentuan solusi akan semakin kecil. Sebaliknya, jika nilai C semakin rendah maka semakin tinggi proporsi kesalahan yang terjadi pada penentuan solusi.

Berdasarkan Persamaan 2.11, akan dimaksimalkan nilai margin antara 2 kelas dengan meminimalkan $\|\mathbf{w}\|^2$. Dalam formula ini, akan dicoba meminimalkan kesalahan yang dinyatakan oleh peubah ξ_i . Penggunaan peubah *slack* bertujuan untuk mengatasi kasus ketidaklayakan (*infeasibility*) dari kendala tersebut. Untuk meminimalkan nilai dari peubah *slack* maka digunakan nilai dari parameter C (Sari, 2017).

Dengan menggunakan teori optimasi, maka didapatkan:

fungsi tujuan:

$$\min_{\mathbf{w}, b, \xi} (\mathbf{w}, b, \xi) = \frac{1}{2} \mathbf{w}^T \mathbf{w} + C \sum_{i=1}^n \xi_i$$

dengan kendala:

$$y_i(\mathbf{w}^T \mathbf{x}_i + b) \geq 1 - \xi_i, \xi_i \geq 0, i = 1, 2, \dots, n$$

atau dapat dituliskan:

fungsi tujuan:

$$\min_{\mathbf{w}, b, \xi}(\mathbf{w}, b, \xi) = \frac{1}{2} \mathbf{w}^T \mathbf{w} + C \sum_{i=1}^n \xi_i$$

dengan kendala:

$$y_i(\mathbf{w}^T \mathbf{x}_i + b) - 1 + \xi_i \geq 0, \xi_i \geq 0, i = 1, 2, \dots, n$$

Untuk mendapatkan solusi dari permasalahan optimasi bentuk primal pada Persamaan 2.22, maka seperti pada kasus *linear separable* akan digunakan *KTT*. Namun pada kasus *non-linear separable* ini akan memiliki dua pengali *Lagrange* untuk kasus ini menjadi:

$$\begin{aligned} L_p(\mathbf{w}, b, \xi, \alpha, \beta) &= (\mathbf{w}, b, \xi) + \sum_{i=1}^n \alpha_i [1 - \xi_i - y_i(\mathbf{w}^T \mathbf{x}_i + b)] \sum_{i=1}^n \beta_i (-\xi_i) \\ &= (\mathbf{w}, b, \xi) - \sum_{i=1}^n \alpha_i [y_i(\mathbf{w}^T \mathbf{x}_i + b) - 1 + \xi_i] - \sum_{i=1}^n \beta_i (-\xi_i) \\ &= \left(\frac{1}{2} \mathbf{w}^T \mathbf{w} + C \sum_{i=1}^n \xi_i \right) - \sum_{i=1}^n \alpha_i [y_i(\mathbf{w}^T \mathbf{x}_i + b) - 1 + \xi_i] - \sum_{i=1}^n \beta_i (\xi_i) \\ &= \frac{1}{2} \mathbf{w}^T \mathbf{w} + C \sum_{i=1}^n \xi_i - \sum_{i=1}^n \alpha_i [y_i(\mathbf{w}^T \mathbf{x}_i + b) - 1 + \xi_i] - \sum_{i=1}^n \beta_i (\xi_i) \end{aligned} \quad (2.12)$$

Peubah α_i dan β_i merupakan *Lagrange Multiplier* dengan $\alpha_i \geq 0$ dan $\beta_i \geq 0$. Peubah α_i dan β_i dapat disebut sebagai peubah non-negatif. Fungsi Persamaan 2.9 akan diminimalkan terhadap peubah \mathbf{w} , b dan ξ serta harus dimaksimalkan terhadap peubah α dan β (Sari, 2017). Berikut turunan pertama dari fungsi Persamaan 2.12 yaitu:

Kondisi 1

$$\begin{aligned} \frac{\partial L_p}{\partial b} &= 0 \\ \frac{\partial L_p}{\partial b} &= \frac{\partial L_p(\mathbf{w}, b, \xi, \alpha, \beta)}{\partial b} \\ &= \frac{\partial \{(\mathbf{w}, b, \xi) - \sum_{i=1}^n \alpha_i [y_i(\mathbf{w}^T \mathbf{x}_i + b) - 1 + \xi_i] - \sum_{i=1}^n \beta_i (\xi_i)\}}{\partial b} \\ &= \frac{\partial \left\{ \frac{1}{2} \mathbf{w}^T \mathbf{w} + C \sum_{i=1}^n \xi_i - \sum_{i=1}^n \alpha_i [y_i(\mathbf{w}^T \mathbf{x}_i + b) - 1 + \xi_i] - \sum_{i=1}^n \beta_i (\xi_i) \right\}}{\partial b} \\ &= \frac{\partial \left\{ \frac{1}{2} \mathbf{w}^T \mathbf{w} + C \sum_{i=1}^n \xi_i - \sum_{i=1}^n \alpha_i [y_i(\mathbf{w}^T \mathbf{x}_i + b) - 1 + \xi_i] - \sum_{i=1}^n \beta_i (\xi_i) \right\}}{\partial b} \\ &= \frac{\partial \left\{ \frac{1}{2} \mathbf{w}^T \mathbf{w} + C \sum_{i=1}^n \xi_i - \sum_{i=1}^n \alpha_i y_i \mathbf{w}^T \mathbf{x}_i - b \sum_{i=1}^n \alpha_i y_i + \sum_{i=1}^n \alpha_i - \sum_{i=1}^n \alpha_i \xi_i - \sum_{i=1}^n \beta_i (\xi_i) \right\}}{\partial b} \\ &= 0 + 0 - 0 - \sum_{i=1}^n \alpha_i y_i + 0 - 0 - 0 \\ 0 &= \sum_{i=1}^n \alpha_i y_i \end{aligned} \quad (2.13)$$

Kondisi 2

$$\begin{aligned}
\frac{\partial L_p}{\partial \mathbf{w}} &= 0 \\
\frac{\partial L_p}{\partial \mathbf{w}} &= \frac{\partial L_p(\mathbf{w}, b, \xi, \alpha, \beta)}{\partial \mathbf{w}} \\
&= \frac{\partial \{(\mathbf{w}, b, \xi) - \sum_{i=1}^n \alpha_i [y_i(\mathbf{w}^T \mathbf{x}_i + b) - 1 + \xi_i] - \sum_{i=1}^n \beta_i(\xi_i)\}}{\partial \mathbf{w}} \\
&= \frac{\partial \left\{ \frac{1}{2} \mathbf{w}^T \mathbf{w} + C \sum_{i=1}^n \xi_i - \sum_{i=1}^n \alpha_i [y_i(\mathbf{w}^T \mathbf{x}_i + b) - 1 + \xi_i] - \sum_{i=1}^n \beta_i(\xi_i) \right\}}{\partial \mathbf{w}} \\
&= \frac{\partial \left\{ \frac{1}{2} \mathbf{w}^T \mathbf{w} + C \sum_{i=1}^n \xi_i - \sum_{i=1}^n \alpha_i y_i \mathbf{w}^T \mathbf{x}_i - b \sum_{i=1}^n \alpha_i y_i + \sum_{i=1}^n \alpha_i - \sum_{i=1}^n \alpha_i \xi_i - \sum_{i=1}^n \beta_i(\xi_i) \right\}}{\partial b} \\
&= \mathbf{w} + 0 - \sum_{i=1}^n \alpha_i y_i \mathbf{x}_i - 0 + 0 - 0 - 0 \\
&= \mathbf{w} - \sum_{i=1}^n \alpha_i y_i \mathbf{x}_i \\
\mathbf{w} &= \sum_{i=1}^n \alpha_i y_i \mathbf{x}_i \tag{2.14}
\end{aligned}$$

Kondisi 3

$$\begin{aligned}
\frac{\partial L_p}{\partial \xi} &= 0 \\
&= \frac{\partial \{(\mathbf{w}, b, \xi) - \sum_{i=1}^n \alpha_i [y_i(\mathbf{w}^T \mathbf{x}_i + b) - 1 + \xi_i] - \sum_{i=1}^n \beta_i(\xi_i)\}}{\partial \xi} \\
&= \frac{\partial \left\{ \frac{1}{2} \mathbf{w}^T \mathbf{w} + C \sum_{i=1}^n \xi_i - \sum_{i=1}^n \alpha_i [y_i(\mathbf{w}^T \mathbf{x}_i + b) - 1 + \xi_i] - \sum_{i=1}^n \beta_i(\xi_i) \right\}}{\partial \xi} \\
&= \frac{\partial \left\{ \frac{1}{2} \mathbf{w}^T \mathbf{w} + C \sum_{i=1}^n \xi_i - \sum_{i=1}^n \alpha_i y_i \mathbf{w}^T \mathbf{x}_i - b \sum_{i=1}^n \alpha_i y_i + \sum_{i=1}^n \alpha_i - \sum_{i=1}^n \alpha_i \xi_i - \sum_{i=1}^n \beta_i(\xi_i) \right\}}{\partial \xi} \\
&= 0 + \sum_{i=1}^n C - 0 - 0 + 0 - \sum_{i=1}^n \alpha_i - \sum_{i=1}^n \beta_i \\
&= \sum_{i=1}^n C - \sum_{i=1}^n \alpha_i - \sum_{i=1}^n \beta_i \\
\sum_{i=1}^n \alpha_i &= \sum_{i=1}^n C - \sum_{i=1}^n \beta_i \\
C &= \alpha_i + \beta_i \tag{2.15}
\end{aligned}$$

Untuk memperoleh bentuk dual, maka akan disubstitusikan Persamaan 2.13, 2.14 dan 2.15 ke dalam Persamaan 2.13.

$$\begin{aligned}
L_p(\alpha) &= \frac{1}{2} \mathbf{w}^T \mathbf{w} + C \sum_{i=1}^n \xi_i - \sum_{i=1}^n \alpha_i [y_i(\mathbf{w}^T \mathbf{x}_i + b) - 1 + \xi_i] - \sum_{i=1}^n \beta_i(\xi_i) \\
&= \frac{1}{2} \mathbf{w}^T \mathbf{w} + C \sum_{i=1}^n \xi_i - \sum_{i=1}^n \alpha_i y_i \mathbf{w}^T \mathbf{x}_i - b \sum_{i=1}^n \alpha_i y_i + \sum_{i=1}^n \alpha_i - \\
&\quad \sum_{i=1}^n \alpha_i - \sum_{i=1}^n \beta_i(\xi_i) \\
&= \frac{1}{2} \mathbf{w}^T \mathbf{w} + \sum_{i=1}^n (C - \alpha_i) \xi_i - \sum_{i=1}^n \alpha_i y_i \mathbf{w}^T \mathbf{x}_i + \sum_{i=1}^n \alpha_i -
\end{aligned}$$

$$\begin{aligned}
& \sum_{i=1}^n (C - \alpha_i) \xi_i \\
&= \frac{1}{2} \mathbf{w}^T \mathbf{w} - \sum_{i=1}^n \alpha_i y_i \mathbf{w}^T \mathbf{x}_i + \sum_{i=1}^n \alpha_i \\
&= \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j \mathbf{x}_i^T \mathbf{x}_j - \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j \mathbf{x}_i^T \mathbf{x}_j + \sum_{i=1}^n \alpha_i \\
&= \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j \mathbf{x}_i^T \mathbf{x}_j \tag{2.16}
\end{aligned}$$

Menurut teori dualitas, meminimumkan $L_p(\mathbf{w})$ sama dengan memaksimumkan $L_p(\alpha)$, sehingga masalah pencarian *hyperplane* yang optimal pada kasus *non-linear separable* dapat dirumuskan dengan:

fungsi tujuan:

$$\max_{\alpha} L_p(\alpha) = \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j \mathbf{x}_i^T \mathbf{x}_j$$

dengan kendala:

$$\sum_{i=1}^n \alpha_i y_i = 0 \text{ dan } \beta_i \xi_i = (C - \alpha_i) \xi_i = 0 \quad \forall_i \text{ dengan } \alpha_i > 0$$

Selanjutnya fungsi *hyperplane* yang optimal pada kasus *non-linear* hampir sama dengan kasus *linear separable* yaitu sebagai berikut:

$$f(x_d) = \sum_{i \in S} \alpha_i y_i \mathbf{x}_i^T \mathbf{x} + b \tag{2.17}$$

dengan S merupakan data yang akan diklasifikasikan, α_i merupakan solusi optimal dari masalah optimasi (2.5) dan b dicari dengan formula. S adalah himpunan indeks *support vector*. Karena α_i tidak nol untuk *support vector* maka penjumlahan di (2.17) ditambahkan hanya untuk *support vector*. Untuk α_i tak berhingga, maka:

$$b = y_i - \mathbf{w}^T \mathbf{x}_i$$

sudah memuaskan. Untuk memastikan ketepatan perhitungan, kita menghitung rata-rata bias (b) yang dihitung untuk *support vector* tak berhingga sebagai berikut

$$b = \frac{1}{|U|} \sum_{i \in U} (y_i - \mathbf{w}^T \mathbf{x}_i)$$

dengan U adalah himpunan indeks *support vector* tak berhingga. Dengan demikian, kelas berdasarkan *hyperplane* optimal pada kasus dataset yang *non-linear separable* terbentuk sebagai berikut:

$$f(x) = \sum_{i=1}^n \alpha_i y_i K(x, x_i) + b + b$$

Pada data pelatihan untuk kasus *non-separable*, klasifikasi yang diperoleh mungkin tidak memiliki kemampuan generalisasi yang tinggi meskipun *hyperplane* ditentukan secara optimal. Sehingga diatasi dengan cara *input space* dipetakan ke dalam *dot-product space* berdimensi tinggi yang disebut *feature space*. Fungsi

kernel merupakan suatu fungsi yang memetakan data ke ruang dimensi yang lebih tinggi dengan harapan data akan memiliki struktur yang lebih baik sehingga lebih mudah dipisahkan. Menggunakan vektor fungsi *non-linear* sebagai $\phi(x_i): R^n \rightarrow R^{n_k}$ memetakan ruang input ke ruang dimensi yang lebih tinggi. Dalam fungsi *non linier* pengklasifikasian diperoleh dengan:

$$f(x) = \text{sign} [\mathbf{w}^T \phi(x_i) + b]$$

dengan menggunakan teori optimasi, maka didapatkan:

fungsi tujuan:

$$\min_{\mathbf{w}, b, \xi} (\mathbf{w}, b, \xi) = \frac{1}{2} \mathbf{w}^T \mathbf{w} + \frac{1}{2} C \sum_{i=1}^n \xi_i^2$$

dengan kendala:

$$y_i(\mathbf{w}^T \phi(x_i) + b) = 1 - \xi_i^2$$

$$\xi_i \geq 0, i = 1, 2, \dots, n$$

Untuk menurunkan permasalahan (2.15), (2.16) dan (2.17) digunakan *Lagrange Multiplier* dengan nilai $\alpha_i \geq 0$. Optimal poin akan ada di dalam *saddle point* dari *Lagrange function* menjadi:

$$L_p(\mathbf{w}, b, \xi, \alpha, \beta) = (\mathbf{w}, b, \xi) + \sum_{i=1}^n \alpha_i [y_i(\mathbf{w}^T \phi(x_i) + b) - 1 + \xi_i]$$

Kemudian dilakukan turunan pertama yaitu sebagai berikut:

Kondisi 1

$$\begin{aligned} \frac{\partial L_p}{\partial b} &= 0 \\ &= \frac{\partial L_p(\mathbf{w}, b, \xi, \alpha, \beta)}{\partial b} \\ &= \frac{\partial \{(\mathbf{w}, b, \xi) - \sum_{i=1}^n \alpha_i [y_i(\mathbf{w}^T \phi(x_i) + b) - 1 + \xi_i]\}}{\partial b} \\ &= \frac{\partial \left\{ \frac{1}{2} \mathbf{w}^T \mathbf{w} + \frac{1}{2} C \sum_{i=1}^n \xi_i^2 - \sum_{i=1}^n \alpha_i [y_i(\mathbf{w}^T \phi(x_i) + b) - 1 + \xi_i] \right\}}{\partial b} \\ &= \frac{\partial \left\{ \frac{1}{2} \mathbf{w}^T \mathbf{w} + \frac{1}{2} C \sum_{i=1}^n \xi_i^2 - \sum_{i=1}^n \alpha_i y_i \mathbf{w}^T \phi(x_i) - b \sum_{i=1}^n \alpha_i y_i + \sum_{i=1}^n \alpha_i - \sum_{i=1}^n \alpha_i \xi_i \right\}}{\partial b} \\ &= 0 + 0 - 0 - \sum_{i=1}^n \alpha_i y_i + 0 - 0 \end{aligned}$$

$$\sum_{i=1}^n \alpha_i y_i = 0$$

Kondisi 2

$$\begin{aligned}
\frac{\partial L_p}{\partial \mathbf{w}} &= 0 \\
\frac{\partial L_p}{\partial \mathbf{w}} &= \frac{\partial L_p(\mathbf{w}, b, \xi, \alpha, \beta)}{\partial \mathbf{w}} \\
&= \frac{\partial \{(\mathbf{w}, b, \xi) - \sum_{i=1}^n \alpha_i [y_i (\mathbf{w}^T \phi(x_i) + b) - 1 + \xi_i]\}}{\partial \mathbf{w}} \\
&= \frac{\partial \left\{ \frac{1}{2} \mathbf{w}^T \mathbf{w} + \frac{1}{2} c \sum_{i=1}^n \xi_i^2 - \sum_{i=1}^n \alpha_i [y_i (\mathbf{w}^T \phi(x_i) + b) - 1 + \xi_i] \right\}}{\partial \mathbf{w}} \\
&= \frac{\partial \left\{ \frac{1}{2} \mathbf{w}^T \mathbf{w} + \frac{1}{2} c \sum_{i=1}^n \xi_i^2 - \sum_{i=1}^n \alpha_i y_i \mathbf{w}^T \phi(x_i) - b \sum_{i=1}^n \alpha_i y_i + \sum_{i=1}^n \alpha_i - \sum_{i=1}^n \alpha_i \xi_i \right\}}{\partial \mathbf{w}} \\
&= \mathbf{w} + 0 - \sum_{i=1}^n \alpha_i y_i x_i - 0 + 0 - 0 - 0 \\
&= \mathbf{w} - \sum_{i=1}^n \alpha_i y_i \phi(x_i) \\
\mathbf{w} - \sum_{i=1}^n \alpha_i y_i x_i &= 0 \\
\sum_{i=1}^n \alpha_i y_i \phi(x_i) &= \mathbf{w}
\end{aligned}$$

Dengan demikian, kelas berdasarkan *hyperplane* optimal pada kasus dataset yang *non-linear separable* terbentuk sebagai berikut:

$$f(x_d) = \sum_{i=1}^n \alpha_i y_i K(x_i, x_d) + b \quad (2.18)$$

dan b diperoleh:

$$b = y_i - \frac{1}{2} \sum_{i=1}^n \alpha_i y_i K(x_i, x)$$

K merupakan salah satu fungsi Kernel yang akan digunakan. Fungsi Kernel digunakan untuk memetakan data non linier menjadi linier. Menurut Hsu, dkk (2010) berikut ini adalah beberapa fungsi kernel yang umum digunakan yaitu:

1. Polynomial

$$K(x_i, x_j) = x_i^T x_j$$

2. RBF

$$K(x_i, x_j) = \exp \left(-\frac{(x_i - x_j)^T (x_i - x_j)}{2 \gamma^2} \right) \quad (2.19)$$

3. Sigmoid

$$K(x_i, x_j) = \tanh(\gamma x_i^T x_j + r)$$

2.7 Confusion Matrix

Confusion matrix adalah metode pengukuran kinerja yang menunjukkan ringkasan hasil prediksi pada masalah klasifikasi yang menunjukkan empat

kombinasi nilai prediksi dan aktual yang digunakan untuk menghitung kinerja sebuah model. *Confusion matrix* ditunjukkan pada Tabel 2.2.

Tabel 2.2. *Confusion Matrix*

	Prediksi Positif	Prediksi Negatif
Aktual Positif	<i>True Positive (TP)</i>	<i>False Positive (FP)</i>
Aktual Negatif	<i>False Negative (FN)</i>	<i>True Negative (TN)</i>

Berikut adalah penjelasan dari masing-masing nilai tersebut:

1. TP yaitu kondisi ketika prediksi (*predicted values*) bermakna positif dan hasilnya (*actual values*) benar.
2. TN yaitu kondisi ketika prediksi (*predicted values*) bermakna negatif dan hasilnya (*actual values*) benar.
3. FP yaitu kondisi ketika prediksi (*predicted values*) bermakna positif dan hasilnya (*actual values*) salah.
4. FN yaitu kondisi ketika prediksi (*predicted values*) bermakna negatif dan hasilnya (*actual values*) salah.

Berdasarkan nilai-nilai tersebut, dapat dihitung performa sebagai berikut (Dellia dan Tjahyanto, 2017):

1. *Accuracy*, merupakan jumlah yang diprediksi benar.

$$\text{Akurasi} = \frac{TP + TN}{TP + TN + FP + FN} \quad (2.20)$$

2. *Precision*, merupakan jumlah prediksi yang benar dari semua kelas positif yang diprediksi.

$$\text{Presisi} = \frac{TP}{TP + FP} \quad (2.21)$$

3. *Recall*, merupakan jumlah prediksi yang benar dari semua kelas positif.

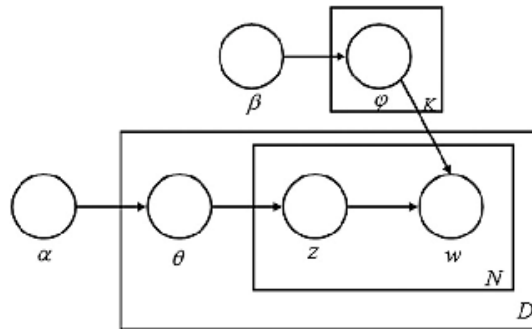
$$\text{Recall} = \frac{TP}{TP + FN} \quad (2.22)$$

4. *F-measure* untuk mengukur *recall* dan *precision* secara bersamaan.

$$F - \text{Measure} = \frac{2 \times \text{Recall} \times \text{Presicion}}{\text{Recall} + \text{Presicion}} \quad (2.23)$$

2.8 Latent Dirichlet Allocation

Ide dasar yang diusulkan metode LDA adalah setiap dokumen direpresentasikan sebagai campuran acak atas topik, setiap topik memiliki karakter yang ditentukan berdasarkan distribusi kata-kata yang terdapat di dalamnya. LDA membutuhkan pendefinisian jumlah topik yang akan dihasilkan oleh model (Rachman dan Pramana, 2020). Cara kerja LDA digambarkan menggunakan grafik yang disebut *probabilistic graphical model* yang ditunjukkan pada Gambar 2.3.



Gambar 2.3 Proses Kerja *Probabilistic Graphical Model*

Pada Gambar 2.2, α dan β merupakan parameter distribusi topik dari dokumen terlebih dahulu atau biasa disebut parameter *prior dirichlet* dan parameter distribusi kata dari sebuah topik. Nilai keduanya merupakan bilangan riil positif yang dapat dituliskan $0 \leq \alpha, \beta \leq 1$. Semakin tinggi nilai α menunjukkan bahwa setiap dokumen mengandung sebagian besar topik dan sebaliknya menunjukkan dokumen memiliki kemungkinan diwakili oleh beberapa topik. Sementara semakin tinggi nilai β menunjukkan suatu topik mengandung campuran sebagian besar kata-kata, sebaliknya suatu topik hanya mengandung campuran dari beberapa kata. Distribusi topik dari dokumen (α) mengakibatkan adanya nilai θ sebagai kumpulan campuran topik yang berbentuk matriks probabilitas topik terhadap dokumen seperti pada matriks berikut:

$$A = \begin{bmatrix} \theta_{11} & \cdots & \theta_{1K} \\ \vdots & \ddots & \vdots \\ \theta_{D1} & \cdots & \theta_{DK} \end{bmatrix}$$

θ_{DK} menunjukkan probabilitas topik ke- K pada dokumen ke- D . Dari kumpulan campuran topik (θ), dapat dipisahkan masing-masing topik (z) dari campuran topik tersebut. Sehingga diperoleh sebuah matriks baru yang berisikan nilai probabilitas kata terhadap topik untuk masing-masing dokumen sebagai berikut:

$$B = \begin{bmatrix} z_{11} & \cdots & z_{1K} \\ \vdots & \ddots & \vdots \\ z_{N1} & \cdots & z_{KN} \end{bmatrix}$$

Nilai z_{KN} menunjukkan topik ke- K pada kata ke- N . Probabilitas topik yang diperoleh (z) dan distribusi kata pada topik (β) menghasilkan probabilitas kata-kata yang muncul sebagai hasil akhir pembentukan model (w), sehingga hasil dari model satu dokumen ini akan memunculkan kata-kata dari kelompok yang terbentuk, kata-kata ini dapat membantu dalam pendefinisian kategori setiap kelompok. Sehingga total probabilitas berdasarkan grafik model LDA secara matematis dapat dituliskan sebagai berikut:

$$P(w, z, \theta | \alpha, \beta) = \prod_{d=1}^D P(\theta_j | \alpha) \prod_{k=1}^K P(\phi_k | \beta) \prod_{n=1}^N P(z_{dn} | \theta_j) P(w_{dn} | \phi, z_{dn})$$

dengan d dan D menunjukkan indeks dokumen, k dan K menunjukkan indeks topik, n dan N menunjukkan indeks kata dalam korpus serta ϕ merupakan korpus untuk kumpulan dokumen yang terjadi akibat adanya β seperti pada Gambar 2.2 (Rangkuti, 2020).

Evaluasi topik yang dihasilkan melalui LDA dilakukan dengan melihat nilai *coherence* topik yaitu seberapa mudah topik tersebut diinterpretasikan menggunakan *Point-wise Mutual Information (PMI) Topic Coherence* sebagai berikut:

$$PMI(k) = \sum_{j=2}^N \sum_{i=2}^{j-1} \log \left(\frac{P(w_i, w_j)}{P(w_i)P(w_j)} \right)$$

dengan N adalah jumlah kata teratas pada topik k , $P(w_i, w_j)$ adalah peluang munculnya kata ke- i dan kata ke- j secara bersama-sama pada dokumen serta $P(w_i)$ dan $P(w_j)$ adalah peluang munculnya kata ke- i dan kata ke- j dalam dokumen (Yao dkk., 2017).

2.9 Kebijakan Merdeka Belajar Kampus Merdeka

Merdeka Belajar Kampus Merdeka (MBKM) merupakan salah satu kebijakan Menteri Pendidikan dan Kebudayaan Nadiem Makarim. Ada dua konsep yang esensial yaitu “Merdeka Belajar” dan “Kampus Merdeka”. Konsep merdeka belajar bermakna adanya kemerdekaan berpikir. Kampus merdeka merupakan upaya untuk melepaskan belenggu untuk bisa bergerak lebih mudah. Tujuan kebijakan MBKM adalah meningkatkan kompetensi lulusan, baik *soft skills* maupun *hard skills* agar

lebih siap dan relevan dengan kebutuhan zaman. Bentuk kegiatan pembelajaran mengacu pada Permendikbud No. 3 Tahun 2020 Pasal 15 Ayat 1 bahwa dilakukan pada delapan bentuk program yang meliputi pertukaran pelajar, magang/praktik kerja, mengajar di instansi pendidikan, proyek di desa, penelitian/riset, kegiatan kewirausahaan, studi/proyek independen dan proyek kemanusiaan (Fuadi dan Aswita 2021).

2.10 Twitter

Twitter didirikan pada Maret 2006 oleh Jack Dorsey dan situs jejaring sosialnya diluncurkan pada Juli 2006. Sejak diluncurkan, Twitter telah menjadi salah satu dari 10 situs yang paling sering dikunjungi di internet (Zuhdi dkk., 2019). Twitter menyediakan *Application Programming Interface* (API) yang memungkinkan pengguna untuk mengakses dan mendapatkan informasi mengenai *tweet*, profil pengguna, data pengikut dan lainnya. Hal tersebut menjadikan Twitter sebagai *microblog* yang banyak diminati perusahaan, organisasi, maupun individu dalam mendapatkan opini publik mengenai suatu topik tertentu (Kurniawan, 2017).

BAB III

METODOLOGI PENELITIAN

3.1 Sumber Data

Data yang digunakan pada penelitian ini merupakan data primer yang diperoleh dari Twitter berupa *tweet* berbahasa Indonesia dengan kata kunci “Kampus Merdeka” yang diunggah pada tanggal 20 Januari 2020 hingga 31 Maret 2022. Data sebanyak 10000 *tweet* yang digunakan berbentuk teks.

3.2 Struktur Data

Struktur data yang digunakan dalam penelitian ini setelah praproses pada data teks *tweet* terdiri dari variabel prediktor yaitu kata dasar setiap *tweet* dan variabel respon yaitu klasifikasi sentimen *tweet* (positif dan negatif). Data klasifikasi pada metode mesin vektor pendukung dibagi menjadi *data training* dan *data testing* dengan perbandingan 80:20 menggunakan *10-fold cross validation*. Tabel 3.1 menunjukkan contoh struktur data penelitian sebelum praproses.

Tabel 3.1 Contoh Struktur Data Penelitian

No	<i>Tweet</i>	Sentimen
1	@Batutuo terus berjuang wujudkan kampus merdeka ...disain program2nya agar dihasilkan SDM yg berkualitas di Era 4.0	Positif
2	Program Kampus Merdeka ini keren banget sih, ngasih harapan buat perubahan sistem pendidikan Indonesia	Positif
3	Ngiri saya sama program kampus merdeka	Positif
⋮	⋮	⋮
1579	labelnya kampus merdeka, tapi sesungguhnya sivitasnya terjajahnya sama berlapis-lapis regulasi	Negatif

3.3 Tahapan Analisis

1. *Crawling* data *tweet* menggunakan *snsrape* yang disimpan dalam format *csv*.
2. Melakukan pelabelan manual menggunakan *doccano* yang merupakan *open source tools* untuk anotasi data.
3. Melakukan analisis sentimen dengan metode mesin vektor pendukung, adapun tahapan analisisnya adalah sebagai berikut:
 - A. Untuk *data training*:
 - a. Melakukan praproses data.
 - b. Melakukan ekstraksi fitur Word2Vec dengan arsitektur *Skip-Gram* dan algoritma *negative sampling*.
 - c. Melakukan analisis sentimen dengan metode mesin vektor pendukung.
 - d. Memperoleh model klasifikasi.
 - B. Untuk *data testing*:
 - a. Melakukan praproses data.
 - b. Melakukan ekstraksi fitur Word2Vec dengan arsitektur *Skip-Gram* dan algoritma *negative sampling*.
 - c. Melakukan analisis sentimen dengan metode mesin vektor pendukung dengan model yang diperoleh dari *data training*.
 - d. Memperoleh prediksi sentimen analisis.
4. Melakukan evaluasi model klasifikasi mesin vektor pendukung menggunakan ekstraksi fitur Word2Vec dengan *confusion matrix*.
5. Melakukan pemodelan topik LDA pada hasil prediksi sentimen positif dan negatif yang telah diperoleh melalui analisis sentimen mesin vektor pendukung menggunakan ekstraksi fitur Word2Vec.
6. Melakukan evaluasi pemodelan topik LDA menggunakan nilai *Point-wise Mutual Information (PMI) Topic Coherence*.

BAB IV

HASIL DAN PEMBAHASAN

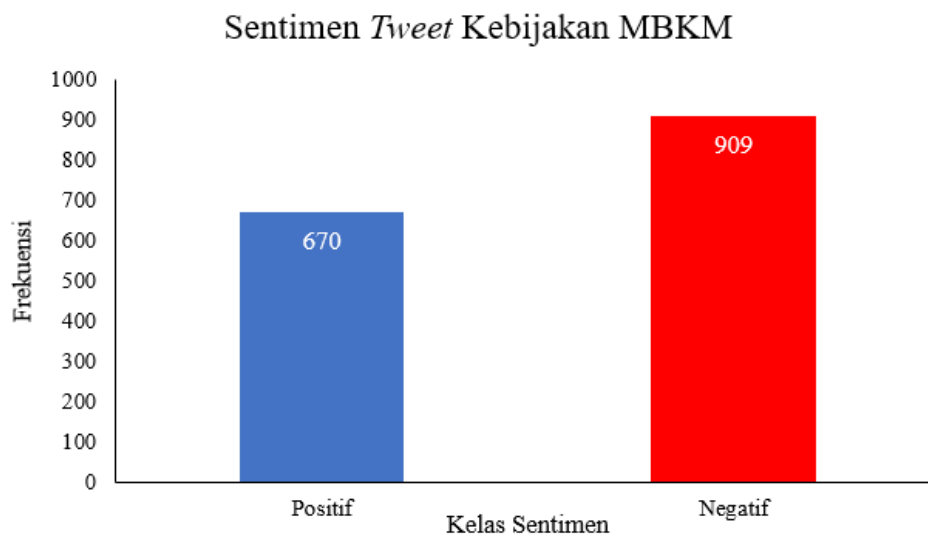
4.1 Deskripsi Data

Data diperoleh dari hasil *crawling* pada Twitter dengan kata kunci “Kampus Merdeka” yang diunggah pada 20 Januari 2020 hingga 30 Maret 2022. Hasil *crawling* diperoleh sebanyak 10000 *tweet* berbahasa Indonesia. Hasil pengumpulan data yang diperoleh ditunjukkan pada Tabel 4.1.

Tabel 4.1 Hasil *Data Crawling*

No	Tanggal Publikasi	<i>Tweet</i>	<i>Username</i>
1	2020-02-20 07:50:02+0 0:00	[MAU TAHU] Luncurkan Empat Kebijakan Merdeka Belajar, Menteri @Kemdikbud_RI Beri Tajuk Kampus Merdeka https://t.co/0ZI4ny2O3g #BersamaIndonesiaMaju #SDMUnggulIndonesiaMaju #JokowiMajukanPendidikan https://t.co/G3gaGQLrcT	tigapilarnews
2	2020-02-08 10:15:28+0 0:00	@Kemdikbud_RI. Pak Nadiem Makarim; Kampus Merdeka #MerdekaBelajar juga harus #MerdekaDariKekerasan seksual. Segera terbitkan SOP penanganan & pencegahan kekerasan seksual di instansi pendidikan. - Tandatangani Petisi! https://t.co/4MR1BB119v lewat @ChangeOrg_ID	e100ss
3	2020-02-09 04:26:27+0 0:00	@adkestwit_ugm Kampus Merdeka bikin saya merdeka dari bayar UKT ga min?	AkuSukaSambal
⋮	⋮	⋮	⋮
10000	2022-03-31 13:32:03+0 0:00	Btw Kampus Merdeka ada banyak banget programnya. Ada proyek kemanusiaan, pertukaran mahasiswa, magang, riset atau penelitian, mengajar, wirausaha. Malah katanya mau ada program Bangkit juga tuh program kesiapan kerja yg didesain Google. Mantap ya #G20BersamaIndonesia	mba_diahworo

Pelabelan manual dilakukan pada data yang terkumpul. Data dilabeli menjadi dua kelas sentimen yaitu positif dan negatif. Sentimen positif berisi tentang dukungan terhadap kebijakan MBKM, sedangkan sentimen negatif berisi tentang penolakan atau ketidakpercayaan terhadap kebijakan MBKM. Berdasarkan hasil pelabelan manual terhadap 10000 data *tweet*, diperoleh data yang terlabeli sebanyak 1579 data dan 8421 data diabaikan karena merupakan data *noise*. Data *noise* adalah data dengan sejumlah besar informasi tambahan yang tidak berarti misalnya *tweet* yang mengandung kalimat yang tidak bisa dikategorikan ke dalam sentimen positif atau negatif. Perbandingan jumlah data terlabeli dan data *noise* cukup signifikan, hal ini menggambarkan pada rentang waktu data dikumpulkan lebih banyak *tweet* yang berisi informasi tambahan yang tidak berarti seperti *tweet* yang berisi pertanyaan terkait kebijakan MBKM sehingga tidak bisa dikategorikan sebagai sentimen positif atau negatif. Gambar 4.1 menunjukkan *Bar Chart* kelas sentimen *tweet* kebijakan MBKM.



Gambar 4.1 *Bar Chart* Kelas Sentimen Data Kebijakan MBKM

Gambar 4.1 menunjukkan data sentimen negatif lebih banyak dibandingkan data sentimen positif, hal tersebut berarti bahwa pada rentang waktu data penelitian dikumpulkan, masyarakat lebih banyak mengunggah *tweet* di Twitter yang berisi tentang penolakan atau ketidakpuasan dibandingkan *tweet* yang berisi dukungan terhadap kebijakan MBKM.

4.2 Praproses Data Teks

Praproses data teks dilakukan terhadap data *tweet* terkait kebijakan MBKM yang terkumpul. Praproses data teks yang dilakukan terdiri dari *data cleansing*, *case folding*, *spelling normalization*, *stemming*, *stopword removal* dan *tokenizing*. Tabel 4.2 menunjukkan struktur data *tweet* sebelum praproses data teks dilakukan.

Tabel 4.2 Struktur Data Sebelum Praproses

No	<i>Tweet</i>	Sentimen
1	Mantap. @nadiemmakarim @pusdatin_dikbud #KampusMerdeka https://t.co/H0VsvbV0Hr	Positif
2	""@daastuff_ @collegemenfess Ikut program kampus merdeka aja kak, smt 3-5 boleh ambil prodi lain/study abroad dengan syarat dan ketentuan yang berlaku, kayak ada minimal IP""	Positif
3	Kampus Merdeka memberikan nilai plus pada kompetensi mahasiswa. https://t.co/2b64dwXXan	Positif
⋮	⋮	⋮
1579	@TaheggaAlfath Menggaungkan kampus merdeka tapi kampus masih ndk boleh diskusi yang katanya bertema sensitif.	Negatif

Data *tweet* sebelum praproses teks masih tersusun dengan struktur pada Tabel 4.2, masih terdapat tanda baca maupun kata tidak baku. Berikut merupakan tahapan dalam praproses data teks pada data *tweet* kebijakan MBKM:

1. *Data Cleansing*

Data Cleansing dilakukan untuk membersihkan data dari data *noise* yang dapat mengganggu pengolahan data. Proses *data cleansing* menghapus *data noise* berupa URL, *username*, *hashtag*, angka dan tanda baca.

- a. Menghapus URL, struktur data ditunjukkan pada Tabel 4.3

Tabel 4.3 Stuktur Data Sebelum dan Setelah Penghapusan URL

No	Teks Sebelum Penghapusan <i>URL</i>	Teks Setelah Penghapusan <i>URL</i>
1	Mantap.@nadiemmakarim @pusdatin_dikbud #KampusMerdeka https://t.co/H0VsvbV0Hr	Mantap.@nadiemmakarim @pusdatin_dikbud #KampusMerdeka
2	""@daastuff_ @collegemenfess Ikut program kampus merdeka aja kak, smt 3-5 boleh ambil prodi lain/study abroad dengan syarat dan ketentuan yang berlaku, kayak ada minimal IP""	""@daastuff_ @collegemenfess Ikut program kampus merdeka aja kak, smt 3-5 boleh ambil prodi lain/study abroad dengan syarat dan ketentuan yang berlaku, kayak ada minimal IP""
3	Kampus Merdeka memberikan nilai plus pada kompetensi mahasiswa. https://t.co/2b64dwXXan	Kampus Merdeka memberikan nilai plus pada kompetensi mahasiswa.
:	:	:
1579	@TaheggaAlfath Menggaungkan kampus merdeka tapi kampus masih ndk boleh diskusi yang katanya bertema sensitif.	@TaheggaAlfath Menggaungkan kampus merdeka tapi kampus masih ndk boleh diskusi yang katanya bertema sensitif.

- b. Menghapus *username*, struktur data ditunjukkan pada Tabel 4.4

Tabel 4.4 Stuktur Data Sebelum dan Setelah Penghapusan *Username*

No	Teks Sebelum Penghapusan <i>Username</i>	Teks Setelah Penghapusan <i>Username</i>
1	Mantap.@nadiemmakarim @pusdatin_dikbud #KampusMerdeka	Mantap. #KampusMerdeka

No	Teks Sebelum Penghapusan <i>Username</i>	Teks Setelah Penghapusan <i>Username</i>
2	""@daastuff_ @collegemenfess Ikut program kampus merdeka aja kak, smt 3-5 boleh ambil prodi lain/study abroad dengan syarat dan ketentuan yang berlaku, kayak ada minimal IP""	""Ikut program kampus merdeka aja kak, smt 3-5 boleh ambil prodi lain/study abroad dengan syarat dan ketentuan yang berlaku, kayak ada minimal IP""
3	Kampus Merdeka memberikan nilai plus pada kompetensi mahasiswa.	Kampus Merdeka memberikan nilai plus pada kompetensi mahasiswa.
⋮	⋮	⋮
1579	@TaheggaAlfath Menggaungkan kampus merdeka tapi kampus masih ndk boleh diskusi yang katanya bertema sensitif.	Menggaungkan kampus merdeka tapi kampus masih ndk boleh diskusi yang katanya bertema sensitif.

c. Menghapus *hashtag*, struktur data ditunjukkan pada Tabel 4.5

Tabel 4.5 Stuktur Data Sebelum dan Setelah Penghapusan *Hashtag*

No	Teks Sebelum Penghapusan <i>Hashtag</i>	Teks Setelah Penghapusan <i>Hashtag</i>
1	Mantap. #KampusMerdeka	Mantap.
2	""Ikut program kampus merdeka aja kak, smt 3-5 boleh ambil prodi lain/study abroad dengan syarat dan ketentuan yang berlaku, kayak ada minimal IP""	""Ikut program kampus merdeka aja kak, smt 3-5 boleh ambil prodi lain/study abroad dengan syarat dan ketentuan yang berlaku, kayak ada minimal IP""
3	Kampus Merdeka memberikan nilai plus pada kompetensi mahasiswa.	Kampus Merdeka memberikan nilai plus pada kompetensi mahasiswa.

No	Teks Sebelum Penghapusan <i>Hashtag</i>	Teks Setelah Penghapusan <i>Hashtag</i>
:	:	:
1579	Menggaungkan kampus merdeka tapi kampus masih ndk boleh diskusi yang katanya bertema sensitif.	Menggaungkan kampus merdeka tapi kampus masih ndk boleh diskusi yang katanya bertema sensitif.

d. Menghapus angka dan tanda baca, struktur data ditunjukkan pada Tabel 4.6

Tabel 4.6 Stuktur Data Sebelum dan Setelah Penghapusan Angka dan Tanda Baca

No	Teks Sebelum Penghapusan Angka dan Tanda Baca	Teks Setelah Penghapusan Angka dan Tanda Baca
1	Mantap.	Mantap
2	""Ikut program kampus merdeka aja kak, smt 3-5 boleh ambil prodi lain/study abroad dengan syarat dan ketentuan yang berlaku, kayak ada minimal IP""	Ikut program kampus merdeka aja kak smt boleh ambil prodi lain study abroad dengan syarat dan ketentuan yang berlaku kayak ada minimal IP
3	Kampus Merdeka memberikan nilai plus pada kompetensi mahasiswa.	Kampus Merdeka memberikan nilai plus pada kompetensi mahasiswa
:	:	:
1579	Menggaungkan kampus merdeka tapi kampus masih ndk boleh diskusi yang katanya bertema sensitif.	Menggaungkan kampus merdeka tapi kampus masih ndk boleh diskusi yang katanya bertema sensitif

2. *Case Folding*

Case Folding adalah proses penyamaan *case* dalam sebuah dokumen agar semua dokumen teks konsisten dalam penggunaan huruf kecil. Oleh karena itu, peran *case folding* dibutuhkan dalam mengonversi keseluruhan teks dalam dokumen menjadi suatu bentuk standar. Tabel 4.7 menunjukkan struktur data sebelum dan setelah melalui proses *case folding*.

Tabel 4.7 Struktur Data Sebelum dan Setelah Proses *Case Folding*

No	Teks Sebelum Proses <i>Case Folding</i>	Teks Setelah Proses <i>Case Folding</i>
1	Mantap	mantap
2	Ikut program kampus merdeka aja kak smt boleh ambil prodi lain study abroad dengan syarat dan ketentuan yang berlaku kayak ada minimal IP	ikut program kampus merdeka aja kak smt boleh ambil prodi lain study abroad dengan syarat dan ketentuan yang berlaku kayak ada minimal ip
3	Kampus Merdeka memberikan nilai plus pada kompetensi mahasiswa	kampus merdeka memberikan nilai plus pada kompetensi mahasiswa
⋮	⋮	⋮
1579	Menggaungkan kampus merdeka tapi kampus masih ndk boleh diskusi yang katanya bertema sensitif	menggaungkan kampus merdeka tapi kampus masih ndk boleh diskusi yang katanya bertema sensitif

3. *Spelling Normalization*

Spelling Normalization adalah proses substitusi sebuah kata-kata yang salah eja atau disingkat dalam bentuk tertentu. Tabel 4.8 menunjukkan struktur data sebelum dan setelah melalui proses *spelling normalization*.

Tabel 4.8 Struktur Data Sebelum dan Setelah Proses *Spelling Normalization*

No	Teks Sebelum Proses <i>Spelling Normalization</i>	Teks Setelah Proses <i>Spelling Normalization</i>
1	mantap	mantap
2	ikut program kampus merdeka aja kak smt boleh ambil prodi lain study abroad dengan syarat dan ketentuan yang berlaku kayak ada minimal ip	ikut program kampus merdeka saja kakak semester boleh ambil prodi lain belajar di luar negeri dengan syarat dan ketentuan yang berlaku kayak ada minimal nilai
3	kampus merdeka memberikan nilai plus pada kompetensi mahasiswa	kampus merdeka memberikan nilai tambah pada kompetensi mahasiswa
⋮	⋮	⋮

No	Teks Sebelum Proses <i>Spelling Normalization</i>	Teks Setelah Proses <i>Spelling Normalization</i>
1579	menggaungkan kampus merdeka tapi kampus masih ndk boleh diskusi yang katanya bertema sensitif	menggaungkan kampus merdeka tapi kampus masih tidak boleh diskusi yang katanya bertema sensitif

4. *Stemming*

Stemming adalah proses untuk mendapatkan kata dasar dengan cara menghilangkan awalan, akhiran, sisipan dan *confixes* (kombinasi awalan dan akhiran). Tabel 4.9 menunjukkan struktur data sebelum dan setelah melalui proses *stemming*.

Tabel 4.9 Struktur Data Sebelum dan Setelah Proses *Stemming*

No	Teks Sebelum Proses <i>Stemming</i>	Teks Setelah Proses <i>Stemming</i>
1	mantap	mantap
2	ikut program kampus merdeka saja kakak semester boleh ambil prodi lain belajar di luar negeri dengan syarat dan ketentuan yang berlaku kayak ada minimal nilai	ikut program kampus merdeka saja kakak semester boleh ambil prodi lain belajar di luar negeri dengan syarat dan tentu yang laku kayak ada minimal nilai
3	kampus merdeka memberikan nilai tambah pada kompetensi mahasiswa	kampus merdeka beri nilai tambah pada kompetensi mahasiswa
⋮	⋮	⋮
1579	menggaungkan kampus merdeka tapi kampus masih tidak boleh diskusi yang katanya bertema sensitif	gaung kampus merdeka tapi kampus masih tidak boleh diskusi yang kata tema sensitif

5. *Stopword Removal*

Stopword removal berfungsi untuk mengurangi konjungsi yang sering muncul dalam suatu dokumen teks namun kurang memberikan arti penting terhadap isi dokumen. Tabel 4.10 menunjukkan struktur data sebelum dan setelah melalui proses *stopword removal*.

Tabel 4.10 Struktur Data Sebelum dan Setelah Proses *Stopword Removal*

No	Teks Sebelum Proses <i>Stopword Removal</i>	Teks Setelah Proses <i>Stopword Removal</i>
1	mantap	mantap
2	ikut program kampus merdeka saja kakak semester boleh ambil prodi lain belajar di luar negeri dengan syarat dan ketentuan yang berlaku kayak ada minimal nilai	ikut program kampus merdeka kakak semester ambil prodi lain belajar luar negeri syarat tentu laku minimal nilai
3	kampus merdeka beri nilai tambah pada kompetensi mahasiswa	kampus merdeka beri nilai tambah kompetensi mahasiswa
⋮	⋮	⋮
1579	gaung kampus merdeka tapi kampus masih tidak boleh diskusi yang kata tema sensitif	gaung kampus merdeka kampus masih tidak boleh diskusi kata tema sensitif

6. *Tokenizing*

Tokenizing merupakan proses pemotongan teks menjadi kata, sehingga menjadi token yang dapat dianalisis. Tabel 4.11 menunjukkan struktur data sebelum dan setelah melalui proses *tokenizing*.

Tabel 4.11 Struktur Data Sebelum dan Setelah Proses *Tokenizing*

No	Teks Sebelum Proses <i>Tokenizing</i>	Teks Setelah Proses <i>Tokenizing</i>
1	mantap	[mantap]
2	ikut program kampus merdeka kakak semester ambil prodi lain belajar luar negeri syarat tentu laku minimal nilai	[ikut, program, kampus, merdeka, kakak, semester, ambil, prodi, lain, belajar, luar, negeri, syarat, tentu, laku, minimal, nilai]
3	kampus merdeka beri nilai tambah pada kompetensi mahasiswa	[kampus, merdeka, beri, nilai, tambah, kompetensi, mahasiswa]
⋮	⋮	⋮

No	Teks Sebelum Proses <i>Tokenizing</i>	Teks Setelah Proses <i>Tokenizing</i>
1579	gaung kampus merdeka tapi kampus masih tidak boleh diskusi yang kata tema sensitif	[gaung, kampus, merdeka, kampus, masih, tidak, boleh, diskusi, kata, tema, sensitif]

Tabel 4.12 menunjukkan struktur data *tweet* kebijakan MBKM sebelum dan setelah praproses data teks.

Tabel 4.12 Struktur Data Sebelum dan Setelah Praproses Data Teks

No	Data Sebelum Praproses Data Teks	Data Setelah Praproses Data Teks
1	Mantap.@nadiemmakarim @pusdatin_dikbud #KampusMerdeka https://t.co/H0VsvbV0Hr	[mantap]
2	""@daastuff_ @collegemenfess Ikut program kampus merdeka aja kak, smt 3-5 boleh ambil prodi lain/study abroad dengan syarat dan ketentuan yang berlaku, kayak ada minimal IP""	[ikut, program, kampus, merdeka, kakak, semester, ambil, prodi, lain, belajar, luar, negeri, syarat, tentu, laku, minimal, nilai]
3	Kampus Merdeka memberikan nilai plus pada kompetensi mahasiswa. https://t.co/2b64dwXXan	[kampus, merdeka, beri, nilai, tambah, kompetensi, mahasiswa]
⋮	⋮	⋮
1579	@TaheggaAlfath Menggaungkan kampus merdeka tapi kampus masih ndk boleh diskusi yang katanya bertema sensitif.	[gaung, kampus, merdeka, kampus, masih, tidak, boleh, diskusi, kata, tema, sensitif]

4.3 Ekstraksi Fitur Word2Vec

Ekstraksi fitur Word2Vec dilakukan untuk memperoleh vektor setiap kata dasar yang diperoleh dari proses *tokenizing*. Vektor kata tersebut digunakan sebagai data *input* pada komputer untuk membangun model klasifikasi mesin vektor pendukung. Kata dasar yang diperoleh dari 1579 data *tweet* adalah 490 kata. Tabel

4.13 menunjukkan vektor setiap kata dasar yang diperoleh dari proses ekstraksi fitur Word2Vec arsitektur *Skip-Gram* 100 dimensi ($h = 100$) dan *window size* 2 ($C = 2$) dengan algoritma *negative sampling* dengan jumlah sampel negatif sebanyak 17 ($k = 17$). Setiap vektor kata berukuran 1×100 dengan elemen vektor berupa angka yang merepresentasikan masing-masing kata dasar dan lebih lanjut digunakan sebagai *input data* pada analisis sentimen dengan metode mesin vektor pendukung.

Tabel 4.13 Vektor Kata Dasar dengan Ekstraksi Fitur Word2Vec

No	Kata Dasar	dim1	dim2	dim3	dim4	...	dim100
1	merdeka	-0.15	0.46	0.07	0.05	...	0.02
2	kampus	-0.16	0.47	0.08	0.05	...	0.02
3	mahasiswa	-0.16	0.49	0.06	0.04	...	0.02
4	banget	-0.17	0.46	0.05	0.03	...	0.01
5	magang	-0.17	0.46	0.05	0.04	...	0.02
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
488	beban	-0.16	0.45	0.06	0.03	...	0.02
489	kejar	-0.15	0.46	0.08	0.03	...	0.01
490	zenius	-0.17	0.47	0.06	0.03	...	0.02

4.4 Klasifikasi Mesin Vektor Pendukung

4.4.1 Data Training dan Data Testing

Klasifikasi mesin vektor pendukung menggunakan *data training* untuk membentuk model klasifikasi dan *data testing* untuk melakukan validasi terhadap model klasifikasi yang terbentuk dengan rasio 80:20. Tabel 4.14 menunjukkan jumlah *data training* dan *data testing* dengan *10-fold Cross Validation*.

Tabel 4.14 Proporsi *Data Training* dan *Data Testing*

Klasifikasi	Positif	Negatif	Jumlah
<i>Data Training</i>	535	728	1263
<i>Data Testing</i>	135	181	316
Jumlah	670	909	1579

4.4.2 Klasifikasi Mesin Vektor Pendukung Kernel *Radial Basis Function*

Klasifikasi data dengan mesin vektor pendukung terdiri dari data yang dapat dipisahkan secara linear dan non-linear. Metode *trial and error* dilakukan untuk menentukan model klasifikasi mesin vektor pendukung terbaik yang ditunjukkan pada Lampiran 1. Berdasarkan hasil *trial and error*, model klasifikasi terbaik adalah mesin vektor pendukung kernel RBF yang merupakan salah satu kernel yang digunakan untuk data yang dipisahkan secara non-linear.

Pelatihan model klasifikasi dengan mesin vektor pendukung kernel RBF mempertimbangkan dua parameter yaitu C dan γ yang ditentukan melalui *trial and error*. C merupakan parameter yang berfungsi untuk menghindari kesalahan klasifikasi setiap sampel yang diuji, sedangkan γ merupakan parameter yang menentukan *support vector* berdasarkan pengaruh jarak sampel dengan *decision boundary*. Nilai parameter C yaitu 1 dan γ sebesar 0.2875 menunjukkan performa model terbaik yaitu akurasi sebesar 89.87%. Hasil perhitungan ketetapan klasifikasi menggunakan mesin vektor pendukung kernel RBF untuk nilai C dan γ yang lain ditunjukkan pada Lampiran 2.

4.4.3 Model Klasifikasi Mesin Vektor Pendukung

Hasil ketetapan klasifikasi terbaik pada metode mesin vektor pendukung kernel RBF dengan parameter γ sebesar 0.2875 disubstitusikan pada Persamaan 2.19, sehingga diperoleh fungsi kernel RBF yaitu:

$$K(x_i, x_j) = \exp\left(-\frac{(x_i - x_j)^T (x_i - x_j)}{0.16}\right) \quad (4.1)$$

Fungsi kernel pada Persamaan 4.1 digunakan untuk membentuk fungsi *hyperplane* dengan cara mensubstitusikan nilai *support vector* kategori positif pada x_i dan nilai *support vector* negatif pada x_j . Fungsi *hyperplane* dihitung dengan mensubstitusikan Persamaan 2.24 pada Persamaan 2.18 sebagai berikut:

$$\begin{aligned} f(x_j) &= \sum_{i=1}^n \alpha_i y_i K(x_i, x_j) + b \\ &= \sum_{i=1}^{1579} \alpha_i y_i \exp\left(-\frac{(x_i - x_j)^T (x_i - x_j)}{0.16}\right) + b \end{aligned}$$

$$\begin{aligned}
&= \sum_{i=1}^{1579} \alpha_i y_i \exp\left(-\frac{(x_i - x_j)^T (x_i - x_j)}{0.16}\right) + \left(y_i - \frac{1}{2} \sum_{i=1}^{1579} \alpha_i y_i K(x_i, x_j)\right) \\
f(x_j) &= \sum_{i=1}^{1579} \alpha_i y_i \exp\left(-\frac{(x_i - x_j)^T (x_i - x_j)}{0.16}\right) + \\
&\quad \left(y_i - \frac{1}{2} \sum_{i=1}^{1579} \alpha_i y_i \exp\left(-\frac{(x_i - x_j)^T (x_i - x_j)}{0.16}\right)\right)
\end{aligned} \tag{4.2}$$

dengan:

α_i = nilai *Lagrange Multiplier* dari fungsi klasifikasi kernel non-linear

y_i = kelas data

x_j = data yang akan di klasifikasi

x_i = data yang menjadi *support vector*, $i = 1, 2, \dots, 1759$

Berdasarkan Persamaan 4.2, diperoleh prediksi kelas sentimen setiap data *tweet* yaitu adalah 648 *tweet* sentimen positif dan 931 *tweet* sentimen negatif.

4.4.4 Confusion Matrix pada Mesin Vektor Pendukung

Performa model klasifikasi diukur dengan *confusion matrix* berdasarkan hasil prediksi dari *data testing*. Tabel 4.15 menunjukkan *confusion matrix* dari model klasifikasi kebijakan MBKM menggunakan mesin vektor pendukung.

Tabel 4.15 *Confusion Matrix* Model Klasifikasi Mesin Vektor Pendukung

Kelas Aktual	Kelas Prediksi	
	Positif	Negatif
Positif	114	21
Negatif	11	170

Berdasarkan *confusion matrix* pada Tabel 4.18, dari 316 *data testing* terdapat 114 data TP atau data sentimen positif yang terprediksi dengan benar, 170 data TN atau data sentimen negatif yang terprediksi dengan benar, serta 21 data FN atau data sentimen positif yang salah terprediksi sebagai sentimen negatif serta 11 data FP atau data sentimen negatif yang salah terprediksi sebagai sentimen positif. Salah satu penyebab data FN dan FP salah terprediksi adalah karena praproses data teks yaitu *stopword removal* menghilangkan beberapa kata dengan orientasi yang

mewakili kelas sentimen aktual dari data tersebut. Data FN dan FP ditunjukkan pada Tabel 4.16 dan 4.17.

Tabel 4.16 Data FN *Tweet* Kebijakan MBKM

No	Data Sebelum Praproses Data Teks	Data Setelah Praproses Data Teks
1.	kok kampus merdeka ada lagi...damn pengen ikut lagi.....serius	[kampus, merdeka, sial, serius]
2.	Ternyata Kemensos punya program seru di kampus merdeka, namanya pejuang muda. Mahasiswa dapat langsung mengaplikasikan ilmu untuk memberi dampak sosial secara konkret, dan dinilai setara 20 SKS!Aku bersyukur, tadi pagi diberi kesempatan sharing ke 1.6K Teman-teman pejuang Muda https://t.co/NNXAiuRIrg	[kemensos, seru, kampus, merdeka, nama, juang, muda, mahasiswa, aplikasi, ilmu, dampak, sosial, konkret, nilai, tara, sks, syukur, bagi, juang, muda]
3.	Tapi harusnya bersyukur pas zaman saya belum ada tuh kampus merdeka. Kan saya pengen:') https://t.co/0oP8RJYKO3	[syukur, pas, zaman, kampus, merdeka]
⋮	⋮	⋮
21.	'@stikes_wikabali Ayo manfaatkan peluang emas kampus merdeka'	[ayo, manfaat, peluang, emas, kampus, merdeka]

Tabel 4.17 Data FP *Tweet* Kebijakan MBKM

No	Data Sebelum Praproses Data Teks	Data Setelah Praproses Data Teks
1.	'@n_riou_ Hmmm ya berarti kaga ada standarnya yang jelas kalo kayak gitu Kemarin gue juga denger ada yang kampus merdeka itu juga kayak gitu dan duitnya sih lancar2 aja juga. Tapi ya controllingnya emang ga jalan kali makanya banyak kasus2 kayak begini'	[gumam, standar, kayak, dengar, kampus, merdeka, kayak, duit, sih, lancar, kontrol, kayak]
2.	'program kampus merdeka malah bikin susah nyari intern'	[kampus, merdeka, bikin, susah, magang]
3.	Kampus merdeka program magang alias sarana tenaga kerja terdidik gratis ciat ciat ciat""CIAAATT""	[kampus, merdeka, magang, alias, sarana, tenaga, didik, gratis, ciat, ciat, ciat, ciaatt]
⋮	⋮	⋮
11.	'Kurikulum Kampus Merdeka ini tujuannya bagus, tapi sejauh ini aplikasinya fucked up banget kalau harus launching ASAP.Balik lagi, mahasiswa asal lulus mata kuliah aja. Nanti di dunia kerja belajar dari 0 lagi alias percuma belajar sekarang.	[kurikulum, kampus, merdeka, tuju, bagus, aplikasi, fucked, up, banget, luncur, mahasiswa, lulus, mata, kuliah, dunia, ajar, alias, ajar]

Performa model klasifikasi berdasarkan confusion matrix pada Tabel 4.15 diperoleh menggunakan Persamaan 2.20 sampai Persamaan 2.23 sebagai berikut:

$$\text{Akurasi} = \frac{114 + 170}{114 + 170 + 11 + 21} \times 100\% = 89.87\%$$

$$\text{Presisi} = \frac{114}{114 + 11} \times 100\% = 91.20\%$$

$$\text{Recall} = \frac{114}{114 + 21} \times 100\% = 84.44\%$$

$$F - \text{Measure} = \frac{2 \times 0.84 \times 0.91}{0.84 + 0.91} \times 100\% = 87.68\%$$

Performa klasifikasi pada mesin vektor pendukung menghasilkan akurasi sebesar 89.87% yaitu persentasi data sentimen positif dan negatif yang diprediksi dengan benar. Presisi sebesar 91.20% merupakan persentasi data sentimen positif yang diprediksi dengan benar terhadap keseluruhan data yang diprediksi bersentimen positif. *Recall* sebesar 84.44% merupakan persentasi merupakan persentasi data sentimen positif yang diprediksi dengan benar terhadap keseluruhan data aktual yang bersentimen positif. *F-Measure* sebesar 87.68% menunjukkan persentasi presisi dan *recall* secara bersamaan.

4.5 Pemodelan Topik *Latent Dirichlet Allocation*

Pemodelan topik LDA dilakukan setelah data *tweet* terkait kebijakan MBKM terprediksi berdasarkan model klasifikasi mesin vektor pendukung melalui praproses data teks yang terdiri dari *data cleansing*, *case folding*, *spelling normalization*, *stemming*, *stopword removal* dan *tokenizing* dengan struktur data yang ditunjukkan pada Tabel 4.2 hingga Tabel 4.11. Jumlah topik yang dimodelkan pada pemodelan topik LDA ditentukan berdasarkan kenaikan nilai *coherence* paling signifikan pada masing-masing jumlah topik yang diuji dengan metode *trial and error*. -

4.5.1 Pemodelan Topik pada Sentimen Positif

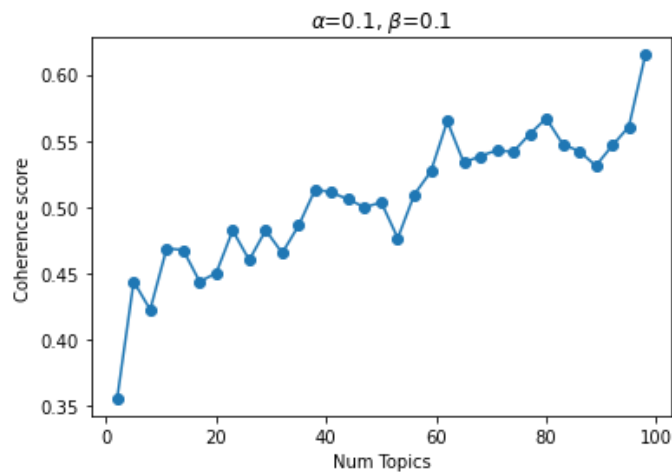
Sentimen positif terkait kebijakan MBKM yang terprediksi berdasarkan model klasifikasi mesin vektor pendukung terdiri dari 648 *tweet* yang ditunjukkan pada Tabel 4.18.

Tabel 4.18 Sentimen Positif Terprediksi

No	Tanggal Publikasi	<i>Tweet</i>	<i>Username</i>
1	2021-08-13 13:11:50+0 0:00	"Merdeka Belajar Kampus Merdeka bukan perubahan yang kecil, ini perubahan yang besar. Dengan harapan bisa mengejar ketertinggalan dan bahkan lompat melampaui negara-negara maju, ujar Mendikbudristek @nadiemmakarim	temannyahe wan
2	2021-08-19 03:07:30+0 0:00	Haduh enak banget yg belum menjadi mahasiswa akhir skrg, program kampus merdeka extremely incredible to develop our competence. udah dapet pengalaman berharga,	smwrsunny
⋮	⋮	⋮	⋮
648	2020-12-26 15:52:49+0 0:00	Paket kebijakan kampus merdeka dg fokus transformasi pendidikan tinggi dg 8 IKU akan sia sia jika pandemi ini akan berjalan lama. Seharusnya Mas Menteri fokus melakukan transformasi pendidikan ke era kebiasaan baru. Sudah 2 semester pembelajaran pandemi, tapi tidak ada perubahan	novensupra

Pemodelan topik LDA dilakukan setelah data *tweet* melalui pra proses data teks. Metode *trial and error* digunakan untuk menetapkan parameter α dan β serta jumlah topik yang dimodelkan. Gambar 4.2 menunjukkan nilai *coherence* pada jumlah topik yang diuji sebanyak 2 hingga 100 topik. Semakin banyak jumlah topik yang dimodelkan, nilai *coherence* juga cenderung meningkat. Namun, kenaikan nilai *coherence* paling signifikan terjadi pada titik jumlah topik 2 ke 5 sebesar 8.82% sehingga jumlah topik yang dimodelkan adalah 5 dengan parameter α serta

β sebesar 0.1 yang menghasilkan nilai *coherence* 0.44. Nilai *coherence* menunjukkan tingkat kemudahan interpretasi topik yang dimodelkan. Semakin tinggi nilai *coherence*, maka semakin mudah setiap topik diinterpretasikan atau dipahami. Pemodelan topik pada sentimen positif terkait kebijakan MBKM menghasilkan nilai *coherence* yang rendah sehingga topik-topik yang dimodelkan sulit diinterpretasikan atau dipahami. Perhitungan ketetapan jumlah topik serta parameter α dan β pada *tweet* terprediksi terkait kebijakan MBKM sentimen positif ditunjukkan pada Lampiran 3.



Gambar 4.2 Nilai *Coherence* Pemodelan Topik Sentimen Positif

Hasil pemodelan topik LDA pada sentimen positif terkait kebijakan MBKM sebanyak 5 topik beserta 5 kata kunci yang berkontribusi paling besar terhadap pemodelan masing-masing topik ditunjukkan pada Tabel 4.19. Kata kunci tersebut merupakan kata yang mewakili sebuah topik dengan frekuensi kemunculan terbanyak pertama hingga kelima pada keseluruhan dokumen yang dimodelkan atau kata yang berkontribusi paling besar pada pemodelan sebuah topik. Selain itu, kata kunci tersebut merepresentasikan topik yang diwakili. Jika sebuah pemodelan topik menghasilkan nilai *coherence* yang tinggi dan mudah diinterpretasikan, maka kata-kata kunci teratas akan mudah dipahami sebagai kata-kata yang berhubungan dan pada akhirnya sebuah topik dapat disimpulkan. Namun, pemodelan topik LDA pada sentimen positif terkait kebijakan MBKM menghasilkan nilai *coherence* yang rendah sehingga menyimpulkan sebuah topik berdasarkan hubungan kata-kata kunci sulit dilakukan. Tabel 4.19 menunjukkan terdapat beberapa kata dengan orientasi negatif menjadi kata kunci pada pemodelan topik kelas sentimen positif

yaitu “stres” dan “ribet”. Hal tersebut dipengaruhi oleh kesalahan prediksi kelas sentimen pada model klasifikasi yang terbangun sebelumnya.

Tabel 4.19 Pemodelan Topik LDA pada Sentimen Positif Kebijakan MBKM

Topik	Kata Kunci				
	1	2	3	4	5
Topik 1	kampus	merdeka	tertawa	saku	magang
Topik 2	merdeka	kampus	dosen	akademik	lambat
Topik 3	merdeka	kampus	mahasiswa	uang	ribet
Topik 4	anak	mendikbud	web	stres	ilmu
Topik 5	merdeka	kampus	mahasiswa	magang	kuliah

4.5.2 Pemodelan Topik pada Sentimen Negatif

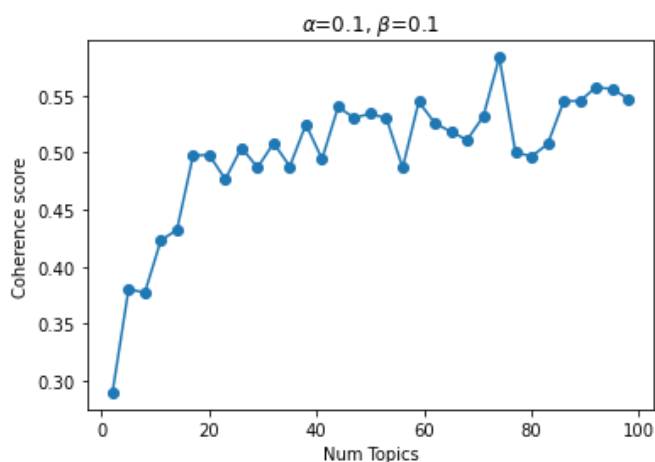
Sentimen negatif terkait kebijakan MBKM yang terprediksi berdasarkan model klasifikasi mesin vektor pendukung terdiri dari 931 *tweet* yang ditunjukkan pada Tabel 4.20.

Tabel 4.20 Sentimen Negatif Terprediksi

No	Tanggal Publikasi	<i>Tweet</i>	<i>Username</i>
1	2021-01-18 03:47:46+0 0:00	Dear people, bersyukurah kita yang sudah tamat kuliah S1. Kenapa? Sekarang kurikulumnya makin ribet! Ada kurikulum merdeka belajar kampus merdeka. Dosennya ribet, mahasiswannya juga ribet! Hayyyaaaaah!	sanddferr
2	2021-01-08 12:20:44+0 0:00	@gurulesmtk Kampus merdeka ribet wkwkw	liberalnanggun g
3	2021-04-01 11:21:48+0 0:00	@Jesse_Ath ini kampus merdeka apaan sih anjir.. ganggu ajh:')	tn_camaraderi e

No	Tanggal Publikasi	Tweet	Username
⋮	⋮	⋮	⋮
931	2020-11-29 04:43:16+0 0:00	Kampus merdeka katanya, tapi mahasiswanya dijajah tugas terus, mana merdekanya?	DraftAnakUn pad

Gambar 4.3 menunjukkan nilai *coherence* pada jumlah topik yang diuji sebanyak 2 hingga 100 topik. Semakin banyak jumlah topik yang dimodelkan, nilai *coherence* juga cenderung meningkat. Namun, kenaikan nilai *coherence* paling signifikan terjadi pada titik jumlah topik 2 ke 5 sebesar 9.09% sehingga jumlah topik yang dimodelkan adalah 5 dengan parameter α serta β sebesar 0.1 yang menghasilkan nilai *coherence* 0.38. Pemodelan topik pada sentimen negatif terkait kebijakan MBKM menghasilkan nilai *coherence* yang lebih rendah dibandingkan pemodelan topik pada sentimen positif sehingga topik-topik yang dimodelkan juga lebih sulit diinterpretasikan atau dipahami dibandingkan dengan topik-topik yang dimodelkan pada sentimen positif. Perhitungan ketetapan jumlah topik serta parameter α dan β pada *tweet* terprediksi terkait kebijakan MBKM sentimen negatif ditunjukkan pada Lampiran 4.



Gambar 4.3 Nilai *Coherence* Pemodelan Topik Sentimen Negatif

Hasil pemodelan topik LDA pada sentimen negatif terkait kebijakan MBKM sebanyak 5 topik beserta kata-kata kunci yang berkontribusi paling besar terhadap

pemodelan masing-masing topik ditunjukkan pada Tabel 4.21. Pemodelan topik LDA pada sentimen negatif terkait kebijakan MBKM menghasilkan nilai *coherence* yang lebih rendah dibandingkan dengan pemodelan topik sentimen positif sehingga menyimpulkan topik berdasarkan hubungan kata-kata kunci pada sentimen negatif lebih sulit dilakukan. Tabel 4.21 menunjukkan terdapat kata dengan orientasi positif menjadi kata kunci pada pemodelan topik kelas sentimen negatif yaitu “keren”. Hal tersebut dipengaruhi oleh kesalahan prediksi kelas sentimen pada model klasifikasi yang terbangun sebelumnya.

Tabel 4.21 Pemodelan Topik LDA pada Sentimen Negatif Kebijakan MBKM

Topik	Kata Kunci				
	1	2	3	4	5
Topik 1	merdeka	kampus	mahasiswa	bijak	sengsara
Topik 2	merdeka	kampus	mahasiswa	magang	tertawa
Topik 3	bayar	mas	susah	menteri	penuh
Topik 4	merdeka	kampus	banget	magang	keren
Topik 5	kampus	merdeka	uang	magang	kuliah

BAB V

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Berdasarkan hasil analisis dan pembahasan diperoleh kesimpulan sebagai berikut:

1. Hasil klasifikasi sentimen kebijakan Merdeka Belajar Kampus Merdeka di Twitter menggunakan mesin vektor pendukung kernel *Radial Basis Function* dengan ekstraksi fitur Word2Vec adalah 648 *tweet* sentimen positif dan 931 *tweet* sentimen negatif dengan akurasi model klasifikasi 89.87%, presisi 91.20%, *recall* 84.44% dan *F-Measure* 87.68%.
2. Pemodelan topik *Latent Dirichlet Allocation* pada sentimen positif dan negatif terkait kebijakan Merdeka Belajar Kampus Merdeka di Twitter masing-masing menghasilkan 5 topik dengan nilai *coherence* 0.44 pada pemodelan topik sentimen positif dan 0.38 pada pemodelan topik sentimen negatif.

5.2 Saran

Pada penelitian selanjutnya disarankan sebagai berikut:

1. Meningkatkan kualitas data teks dengan memperhatikan langkah-langkah pada praproses data teks untuk memperoleh hasil analisis data yang lebih baik.
2. Membandingkan penggunaan ekstraksi fitur Word2Vec dengan ekstraksi fitur yang lain terhadap performa beberapa model klasifikasi selain mesin vektor pendukung seperti Naïve Bayes, *K-Nearest Neighbor* dan sebagainya.
3. Menghubungkan pengaruh spasial pada analisis sentimen untuk memetakan sentimen masyarakat berdasarkan lokasi tertentu.
4. Melakukan estimasi parameter α dan β pada pemodelan topik LDA.

DAFTAR PUSTAKA

- Aldiansyah, G.W., Adikara, P.P. & Wihandika, R.C. (2019). Rekomendasi Lagu Cross Language Berdasarkan Lirik Menggunakan. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 3(8), 8036–8041.
- Alkaff, M., Rizky Baskara, A. & Hendro Wicaksono, Y. (2020). Sentiment Analysis of Indonesian Movie Trailer on YouTube Using Delta TF-IDF and SVM. *5th International Conference on Informatics and Computing (ICIC) 2020*, 1-5.
- Blanchette, J. (2008). *The Little Manual of API Design*. Oslo: Trolltech.
- Chang, C.Y., Lee, S.J., & Lai, C.C. (2017). Weighted Word2vec Based on The Distance of Words. *International Conference on Machine Learning and Cybernetics (ICMLC)*, 563-568.
- Dellia, P. & Tjahyanto, A. (2017). Tax Complaints Classification on Twitter Using Text Mining. *Journal of Science*, 2(1), 11-15.
- Febrianta, M.Y., Widiyanesti, S. & Ramadhan, S.R. (2021). Analisis Ulasan Indie Video Game Lokal pada Steam Menggunakan Analisis Sentimen dan Pemodelan Topik Berbasis Latent Dirichlet Allocation. *Journal of Animation and Games Studies*, 7(2), 117-144.
- Fuadi, T.M. & Aswita, D. (2021). Merdeka Belajar Kampus Merdeka (MBKM): Bagaimana Penerapan dan Kedala yang Dihadapi oleh Perguruan Tinggi Swasta di Aceh. *Jurnal Dedikasi Pendidikan*, 5(2), 603-614.
- Goldberg, Y., & Levy, O. (2014). Word2Vec Explained: Deriving Mikolov et al.'s Negative-Sampling Word-Embedding Method. *arXiv preprint arXiv:1402.3722*, 1-21.
- Herwijayanti, B., Ratnawati, D.E. & Muflikhah, L. (2018). Klasifikasi Berita Online dengan menggunakan Pembobotan TF-IDF dan Cosine Similarity. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 2(1), 306-312.
- Hsu, S. (2010). Developing a Scale for Teacher Integration of Information and Communication Technology in Grades 1–9. *Journal of Computer Assisted Learning*, 26(3), 175–189

- Ipmawati, J., Kusriani & Luthfi, E.T. (2017). Komparasi Teknik Klasifikasi Teks Mining Pada Analisis Sentimen. *Indonesian Journal on Networking and Security*, 6(1), 28–36.
- Isnain, A.R., Sakti, A.I., Alita, D. & Marga, N.S. (2021). Sentimen Analisis Publik Terhadap Kebijakan Lockdown Pemerintah Jakarta Menggunakan Algoritma Svm. *Jurnal Data Mining dan Sistem Informasi*, 2(1), 31-37.
- Juditha, C. (2018). Fenomena Trending Topic di Twitter: Analisis Wacana Twit# SaveHajiLulung. *Jurnal PIKOM (Penelitian Komunikasi dan Pembangunan)*, 16(2), 138-154.
- Jumeilah, F.S. (2017). Penerapan Support Vector Machine (SVM) untuk Pengkategorian Penelitian. *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, 1(1), 19–25.
- Khotimah, N. (2019). Analisis Sentimen Terhadap Review E-Commerce Dengan Metode Stochastic Gradient Descent. *Disertasi*. Universitas Muhammadiyah Semarang: Semarang.
- Kuhn, M. & Johnson, K. (2013). *Applied Predictive Modeling*. Springer: New York.
- Kurniawan, T. (2017). Implementasi Text mining pada Analisis Sentimen Pengguna Twitter Menggunakan Naïve Bayes Classifier dan Support Vector Machine. *Skripsi*. Institut Teknologi Sepuluh November: Surabaya.
- Lutfiyanto, M.D. & Setiawan, E.B. (2021). Expansion Feature Dengan Word2vec Untuk Analisis Sentimen Pada Opini Politik di Twitter Dengan Klasifikasi Support Vector Machine, Naïve Bayes dan Random Forest. *eProceedings of Engineering*, 8(5), 10399-10410.
- Mäntylä, M.V., Graziotin, D. & Kuutila, M. (2018). The evolution of sentiment analysis - A review of research topics, venues, and top cited papers. *Computer Science Review*, 27, 16-32.
- Maulana, M.A., Setyanto, A. & Kurniawan, M.P. (2018). Analisis Sentimen Media Sosial Universitas Amikom Yogyakarta Sebagai Sarana Penyebaran Informasi Menggunakan Algoritma Klasifikasi SVM. *Seminar Nasional Teknologi Informasi dan Multimedia*, 6(1), 7-12.

- Naufal, H.F. & Setiawan, E.B. (2021). Ekspansi Fitur Pada Analisis Sentimen Twitter Dengan Pendekatan Metode Word2vec. *eProceedings of Engineering*, 8(5), 10399-10349.
- Pertiwi, M.W. (2019). Analisis Sentimen Opini Publik Mengenai Sarana dan Transportasi Mudik Tahun 2019 Pada Twitter Menggunakan Algoritma Naïve Bayes, Neural Network, KNN Dan SVM. *Inti Nusa Mandiri*, 14(1), 27-32.
- Prabowo, H. (2020). *Pro dan Kontra atas Kebijakan 'Kampus Merdeka' Nadiem*. URL: <https://tirto.id/pro-dan-kontra-atas-kebijakan-kampus-merdeka-nadiem-evs2>. Diakses pada 24 Januari 2022.
- Pravina, A.M., Cholissodin, I. & Adikara, P.P. (2019). Analisis Sentimen Tentang Opini Maskapai Penerbangan pada Dokumen Twitter Menggunakan Algoritme Support Vector Machine (SVM). *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 3(3), 2789-2797.
- Rachman, F.F. & Pramana, S. (2020). Analisis Sentimen Pro dan Kontra Masyarakat Indonesia tentang Vaksin COVID-19 pada Media Sosial Twitter. *Indonesian of Health Information Management Journal (INOHIM)*, 8(2), 100-109.
- Rahat, A.M., Kahir, A. & Masum, A.K.M. (2020). Comparison of Naive Bayes and SVM Algorithm based on Sentiment Analysis Using Review Dataset. *International Conference on System Modeling and Advancement in Research Trends*, 8(2), 266–270.
- Rahman, M. D., Djunaidy, A., & Mahananto, F. (2021). Penerapan Weighted Word Embedding pada Pengklasifikasian Teks Berbasis Recurrent Neural Network untuk Layanan Pengaduan Perusahaan Transportasi. *Jurnal Sains dan Seni ITS*, 10(1), A1-A6.
- Rangkuti, M. D. E. (2020). Analisis Topik Komentar Video Beberapa Akun Youtube E-Commerce Indonesia Menggunakan Metode Latent Dirichlet Allocation. *Skripsi*. Universitas Islam Negeri Syarif Hidayatullah Jakarta: Jakarta.
- Rashif, M. (2007). Adaptive Credit Scoring with Kernel Learning Methods. *The European Journal of Operational Research (EJOR)*, 3(183), 1521-1536.

- Rusli, M., Faisal, M.R. & Budiman, I. (2019). Ekstraksi Fitur Menggunakan Model Word2vec untuk Analisis Sentimen pada Komentar Facebook. *Soliter*, 2, 104-109.
- Sari, P.D. (2017). Analisis Credit Scoring Menggunakan Regresi Logistik LASSO dan Support Vector Machine (SVM). *Disertasi*. Institut Pertanian Bogor: Bogor.
- Triyantono, Y.S., Al Faraby, S. & Dwifebri, M. (2021). Analisis Sentimen Terhadap Ulasan Film Menggunakan Word2vec Dan SVM. *eProceedings of Engineering*, 8(4), 4136-4144.
- Turban, E. 2011. *Decision Support and Business Intelligence Systems*. Pearson Education: London.
- Xie, R., Chu, S.K.W., Chiu, D.K.W. & Wang, Y. (2021). Exploring Public Response To COVID-19 on Weibo with LDA Topic Modeling and Sentiment Analysis. *Data and Information Management*, 5(1), 86-99.
- Yao, L., Zhang, Y., Wei, B., Jin, Z., Zhang, R., Zhang, Y., & Chen, Q. (2017) Incorporating Knowledge Graph Embeddings into Topic Modeling. *Thirty-First AAAI Conference on Artificial intelligence*, 3119-3126.
- Yulita, W. (2021). Analisis Sentimen Terhadap Opini Masyarakat Tentang Vaksin Covid-19 Menggunakan Algoritma Naïve Bayes Classifier. *Jurnal Data Mining dan Sistem Informasi*, 2(2), 1-9.
- Zhang, L., Wang, S., & Liu, B. (2018). Deep Learning for Sentiment Analysis: A Survey Wiley Interdisciplinary Reviews. *Data Mining and Knowledge Discovery*, 1-17.
- Zuhdi, A.M., Utami, E. & Raharjo, S. (2019). Analisis sentiment twitter terhadap capres Indonesia 2019 dengan metode K-NN. *Jurnal Informa: Jurnal Penelitian dan Pengabdian Masyarakat*, 5(2), 1-7.

LAMPIRAN

LAMPIRAN

Lampiran 1 Ketetapan Model Klasifikasi Mesin Vektor Pendukung

<i>Hidden Layer Size</i>	<i>Window Size</i>	<i>Negative Sample Size</i>	<i>SVM Kernel</i>	<i>Accuracy</i>	<i>Hidden Layer Size</i>	<i>Window Size</i>	<i>Negative Sample Size</i>	<i>SVM Kernel</i>	<i>Accuracy</i>
100	2	15	linear	86.61	100	3	15	linear	86.06
			rbf	88.75				rbf	88.99
			sigmoid	83.37				sigmoid	78.07
			poly	85.90				poly	85.42
		16	linear	87.17			16	linear	87.25
			rbf	88.83				rbf	88.20
			sigmoid	80.84				sigmoid	81.07
			poly	86.85				poly	82.49
		17	linear	86.54			17	linear	87.41
			rbf	89.87				rbf	88.44
			sigmoid	84.25				sigmoid	80.04
			poly	85.35				poly	84.55
		18	linear	87.17			18	linear	86.45
			rbf	89.23				rbf	88.44
			sigmoid	83.06				sigmoid	80.13
			poly	85.74				poly	83.21
		19	linear	86.54			19	linear	87.80
			rbf	88.91				rbf	87.96
			sigmoid	82.35				sigmoid	77.04
			poly	85.58				poly	85.18
		20	linear	86.69			20	linear	86.85
			rbf	88.89				rbf	88.43
			sigmoid	84.41				sigmoid	77.98
			poly	85.74				poly	83.37
<i>Hidden Layer Size</i>	<i>Window Size</i>	<i>Negative Sample Size</i>	<i>SVM Kernel</i>	<i>Accuracy</i>	<i>Hidden Layer Size</i>	<i>Window Size</i>	<i>Negative Sample Size</i>	<i>SVM Kernel</i>	<i>Accuracy</i>
100	4	15	linear	87.27	100	5	15	linear	86.93
			rbf	88.20				rbf	87.96
			sigmoid	74.82				sigmoid	77.19
			poly	83.13				poly	81.23
		16	linear	87.49			16	linear	87.25
			rbf	88.99				rbf	87.25
			sigmoid	76.40				sigmoid	74.98
			poly	82.10				poly	81.39
		17	linear	88.04			17	linear	87.33
			rbf	88.52				rbf	88.35
			sigmoid	77.43				sigmoid	76.56
			poly	82.34				poly	83.13
		18	linear	87.57			18	linear	88.12
			rbf	88.59				rbf	88.12
			sigmoid	75.61				sigmoid	75.69
			poly	81.39				poly	81.39
		19	linear	87.65			19	linear	87.41
			rbf	88.20				rbf	88.51
			sigmoid	74.66				sigmoid	76.32
			poly	83.52				poly	81.62
		20	linear	88.28			20	linear	88.04
			rbf	88.12				rbf	87.88
			sigmoid	75.45				sigmoid	75.61
			poly	82.26				poly	83.92

Lampiran 1 Ketetapan Model Klasifikasi Mesin Vektor Pendukung (Lanjutan)

<i>Hidden Layer Size</i>	<i>Window Size</i>	<i>Negative Sample Size</i>	<i>SVM Kernel</i>	<i>Accuracy</i>	<i>Hidden Layer Size</i>	<i>Window Size</i>	<i>Negative Sample Size</i>	<i>SVM Kernel</i>	<i>Accuracy</i>
100	6	15	linear	86.85	100	7	15	linear	87.25
			rbf	88.28				rbf	87.41
			sigmoid	76.33				sigmoid	76.08
			poly	80.20				poly	80.91
		16	linear	87.25			16	linear	87.56
			rbf	88.67				rbf	87.40
			sigmoid	77.11				sigmoid	76.24
			poly	81.46				poly	81.31
		17	linear	87.17			17	linear	87.33
			rbf	88.51				rbf	87.80
			sigmoid	80.68				sigmoid	77.20
			poly	81.31				poly	81.15
		18	linear	88.12			18	linear	87.25
			rbf	88.28				rbf	87.33
			sigmoid	78.14				sigmoid	75.69
			poly	81.86				poly	81.23
		19	linear	86.70			19	linear	87.80
			rbf	88.52				rbf	86.85
			sigmoid	77.59				sigmoid	75.69
			poly	82.81				poly	81.07
		20	linear	86.22			20	linear	87.17
			rbf	87.80				rbf	87.48
			sigmoid	71.73				sigmoid	76.16
			poly	85.03				poly	81.31
<i>Hidden Layer Size</i>	<i>Window Size</i>	<i>Negative Sample Size</i>	<i>SVM Kernel</i>	<i>Accuracy</i>	<i>Hidden Layer Size</i>	<i>Window Size</i>	<i>Negative Sample Size</i>	<i>SVM Kernel</i>	<i>Accuracy</i>
200	2	15	linear	87.17	200	3	15	linear	87.32
			rbf	88.83				rbf	88.36
			sigmoid	86.06				sigmoid	74.74
			poly	85.58				poly	83.76
		16	linear	87.33			16	linear	87.32
			rbf	88.83				rbf	88.20
			sigmoid	84.87				sigmoid	78.70
			poly	83.84				poly	81.94
		17	linear	87.17			17	linear	87.32
			rbf	89.39				rbf	88.36
			sigmoid	84.01				sigmoid	79.09
			poly	83.69				poly	82.18
		18	linear	87.80			18	linear	87.16
			rbf	88.99				rbf	88.36
			sigmoid	84.40				sigmoid	76.95
			poly	84.79				poly	82.81
		19	linear	87.41			19	linear	87.25
			rbf	88.75				rbf	88.44
			sigmoid	85.35				sigmoid	79.01
			poly	84.32				poly	83.45
		20	linear	87.01			20	linear	87.88
			rbf	88.91				rbf	88.20
			sigmoid	84.48				sigmoid	78.94
			poly	84.08				poly	81.86

Lampiran 1 Ketetapan Model Klasifikasi Mesin Vektor Pendukung (Lanjutan)

<i>Hidden Layer Size</i>	<i>Window Size</i>	<i>Negative Sample Size</i>	<i>SVM Kernel</i>	<i>Accuracy</i>	<i>Hidden Layer Size</i>	<i>Window Size</i>	<i>Negative Sample Size</i>	<i>SVM Kernel</i>	<i>Accuracy</i>
200	4	15	linear	88.28	200	5	15	linear	88.19
			rbf	88.12				rbf	88.04
			sigmoid	76.24				sigmoid	76.04
			poly	81.39				poly	80.67
		16	linear	87.56			16	linear	87.80
			rbf	88.20				rbf	87.96
			sigmoid	76.88				sigmoid	77.35
			poly	81.86				poly	82.49
		17	linear	87.40			17	linear	87.72
			rbf	88.44				rbf	88.04
			sigmoid	77.91				sigmoid	76.72
			poly	82.18				poly	81.39
		18	linear	87.32			18	linear	88.20
			rbf	88.44				rbf	88.12
			sigmoid	76.24				sigmoid	75.69
			poly	82.50				poly	80.20
		19	linear	87.88			19	linear	87.40
			rbf	88.28				rbf	87.80
			sigmoid	77.99				sigmoid	77.19
			poly	82.65				poly	81.78
		20	linear	87.96			20	linear	88.04
			rbf	88.28				rbf	88.12
			sigmoid	76.95				sigmoid	74.89
			poly	81.55				poly	80.59
<i>Hidden Layer Size</i>	<i>Window Size</i>	<i>Negative Sample Size</i>	<i>SVM Kernel</i>	<i>Accuracy</i>	<i>Hidden Layer Size</i>	<i>Window Size</i>	<i>Negative Sample Size</i>	<i>SVM Kernel</i>	<i>Accuracy</i>
200	6	15	linear	87.96	200	7	15	linear	87.96
			rbf	88.04				rbf	87.14
			sigmoid	76.40				sigmoid	74.18
			poly	81.07				poly	81.94
		16	linear	87.96			16	linear	87.80
			rbf	88.04				rbf	87.41
			sigmoid	78.78				sigmoid	76.00
			poly	81.94				poly	80.91
		17	linear	88.04			17	linear	88.20
			rbf	87.33				rbf	88.28
			sigmoid	79.01				sigmoid	76.01
			poly	82.34				poly	81.78
		18	linear	87.80			18	linear	88.51
			rbf	87.65				rbf	88.12
			sigmoid	77.90				sigmoid	74.97
			poly	79.40				poly	82.10
		19	linear	88.60			19	linear	88.20
			rbf	88.52				rbf	87.56
			sigmoid	79.17				sigmoid	76.88
			poly	80.12				poly	80.28
		20	linear	88.44			20	linear	88.51
			rbf	88.20				rbf	87.33
			sigmoid	79.33				sigmoid	76.40
			poly	84.16				poly	80.04

Lampiran 1 Ketetapan Model Klasifikasi Mesin Vektor Pendukung (Lanjutan)

<i>Hidden Layer Size</i>	<i>Window Size</i>	<i>Negative Sample Size</i>	<i>SVM Kernel</i>	<i>Accuracy</i>	<i>Hidden Layer Size</i>	<i>Window Size</i>	<i>Negative Sample Size</i>	<i>SVM Kernel</i>	<i>Accuracy</i>
300	2	15	linear	85.98	300	3	15	linear	86.93
			rbf	88.83				rbf	87.88
			sigmoid	84.01				sigmoid	77.27
			poly	82.18				poly	81.78
	16	linear	85.90	linear		87.01			
		rbf	88.83	rbf		87.88			
		sigmoid	81.63	sigmoid		78.86			
		poly	83.92	poly		79.32			
	17	linear	85.98	linear		86.85			
		rbf	89.23	rbf		88.20			
		sigmoid	84.64	sigmoid		77.90			
		poly	82.26	poly		81.31			
	18	linear	86.29	linear		86.77			
		rbf	88.83	rbf		88.12			
		sigmoid	84.00	sigmoid		77.27			
		poly	81.78	poly		79.96			
	19	linear	86.61	linear		86.69			
		rbf	88.83	rbf		87.64			
		sigmoid	84.96	sigmoid		77.99			
		poly	82.26	poly		83.13			
20	linear	86.61	linear	87.96					
	rbf	88.83	rbf	87.88					
	sigmoid	83.93	sigmoid	76.96					
	poly	81.86	poly	83.57					
300	4	15	linear	87.17	300	5	15	linear	86.85
			rbf	88.12				rbf	87.40
			sigmoid	76.09				sigmoid	76.40
			poly	81.70				poly	80.59
	16	linear	86.61	linear		86.61			
		rbf	88.35	rbf		87.80			
		sigmoid	76.24	sigmoid		75.06			
		poly	81.70	poly		79.80			
	17	linear	87.72	linear		86.45			
		rbf	88.12	rbf		88.20			
		sigmoid	76.17	sigmoid		77.51			
		poly	80.83	poly		82.10			
	18	linear	87.09	linear		87.17			
		rbf	88.51	rbf		88.04			
		sigmoid	76.08	sigmoid		75.13			
		poly	81.94	poly		79.49			
	19	linear	86.61	linear		87.09			
		rbf	88.20	rbf		87.56			
		sigmoid	76.00	sigmoid		77.03			
		poly	80.12	poly		80.35			
20	linear	86.85	linear	86.77					
	rbf	88.20	rbf	87.33					
	sigmoid	75.92	sigmoid	76.08					
	poly	79.96	poly	80.51					

Lampiran 1 Ketetapan Model Klasifikasi Mesin Vektor Pendukung (Lanjutan)

<i>Hidden Layer Size</i>	<i>Window Size</i>	<i>Negative Sample Size</i>	<i>SVM Kernel</i>	<i>Accuracy</i>	<i>Hidden Layer Size</i>	<i>Window Size</i>	<i>Negative Sample Size</i>	<i>SVM Kernel</i>	<i>Accuracy</i>
300	6	15	linear	86.93	300	7	15	linear	87.09
			rbf	87.48				rbf	87.41
			sigmoid	78.94				sigmoid	76.09
			poly	81.46				poly	80.75
		16	linear	87.48			16	linear	87.17
			rbf	87.80				rbf	87.72
			sigmoid	76.80				sigmoid	76.80
			poly	79.17				poly	81.07
		17	linear	87.17			17	linear	86.85
			rbf	87.65				rbf	87.64
			sigmoid	77.51				sigmoid	76.32
			poly	79.56				poly	79.96
		18	linear	87.17			18	linear	86.38
			rbf	87.96				rbf	87.88
			sigmoid	78.93				sigmoid	75.21
			poly	80.75				poly	80.75
		19	linear	87.09			19	linear	86.45
			rbf	87.96				rbf	88.12
			sigmoid	80.20				sigmoid	75.84
			poly	78.93				poly	79.88
		20	linear	87.24			20	linear	86.85
			rbf	87.80				rbf	87.88
			sigmoid	78.85				sigmoid	75.45
			poly	80.75				poly	79.80
<i>Hidden Layer Size</i>	<i>Window Size</i>	<i>Negative Sample Size</i>	<i>SVM Kernel</i>	<i>Accuracy</i>	<i>Hidden Layer Size</i>	<i>Window Size</i>	<i>Negative Sample Size</i>	<i>SVM Kernel</i>	<i>Accuracy</i>
400	2	15	linear	86.53	400	3	15	linear	86.62
			rbf	89.07				rbf	88.59
			sigmoid	83.77				sigmoid	76.48
			poly	84.00				poly	81.86
		16	linear	85.90			16	linear	86.45
			rbf	89.15				rbf	88.44
			sigmoid	82.18				sigmoid	77.59
			poly	85.19				poly	80.83
		17	linear	85.90			17	linear	86.29
			rbf	88.36				rbf	88.51
			sigmoid	82.58				sigmoid	79.88
			poly	82.58				poly	81.31
		18	linear	85.98			18	linear	86.38
			rbf	88.83				rbf	88.20
			sigmoid	82.90				sigmoid	77.67
			poly	84.39				poly	79.24
		19	linear	85.74			19	linear	86.61
			rbf	89.31				rbf	88.36
			sigmoid	82.50				sigmoid	76.88
			poly	85.35				poly	80.04
		20	linear	86.45			20	linear	86.30
			rbf	88.83				rbf	87.96
			sigmoid	83.61				sigmoid	78.14
			poly	85.50				poly	80.99

Lampiran 1 Ketetapan Model Klasifikasi Mesin Vektor Pendukung (Lanjutan)

<i>Hidden Layer Size</i>	<i>Window Size</i>	<i>Negative Sample Size</i>	<i>SVM Kernel</i>	<i>Accuracy</i>	<i>Hidden Layer Size</i>	<i>Window Size</i>	<i>Negative Sample Size</i>	<i>SVM Kernel</i>	<i>Accuracy</i>		
400	4	15	linear	86.45	400	5	15	linear	86.30		
			rbf	88.36				rbf	87.33		
			sigmoid	79.41				sigmoid	75.85		
			poly	82.57				poly	78.69		
			16	linear		86.61			16	linear	87.09
				rbf		88.59				rbf	87.25
				sigmoid		78.62				sigmoid	75.35
				poly		81.47				poly	79.17
			17	linear		86.46			17	linear	86.61
				rbf		88.20				rbf	87.56
				sigmoid		76.32				sigmoid	77.11
				poly		80.28				poly	81.15
			18	linear		87.01			18	linear	86.54
				rbf		88.75				rbf	87.80
				sigmoid		76.32				sigmoid	74.66
				poly		82.89				poly	79.64
			19	linear		86.54			19	linear	86.14
				rbf		87.48				rbf	87.49
				sigmoid		77.35				sigmoid	74.74
				poly		80.12				poly	81.39
		20	linear	87.33			20	linear	86.53		
			rbf	87.96				rbf	87.96		
			sigmoid	77.19				sigmoid	75.13		
			poly	80.12				poly	79.88		
<i>Hidden Layer Size</i>	<i>Window Size</i>	<i>Negative Sample Size</i>	<i>SVM Kernel</i>	<i>Accuracy</i>	<i>Hidden Layer Size</i>	<i>Window Size</i>	<i>Negative Sample Size</i>	<i>SVM Kernel</i>	<i>Accuracy</i>		
400	6	15	linear	87.01	400	7	15	linear	86.93		
			rbf	87.72				rbf	87.41		
			sigmoid	77.75				sigmoid	77.03		
			poly	80.44				poly	82.02		
			16	linear		86.77			16	linear	86.62
				rbf		87.80				rbf	87.33
				sigmoid		79.41				sigmoid	76.24
				poly		79.80				poly	80.44
			17	linear		86.45			17	linear	86.46
				rbf		87.33				rbf	87.01
				sigmoid		77.59				sigmoid	77.19
				poly		78.85				poly	79.64
			18	linear		86.69			18	linear	86.22
				rbf		87.88				rbf	87.01
				sigmoid		77.90				sigmoid	76.40
				poly		81.31				poly	80.44
			19	linear		87.17			19	linear	85.43
				rbf		87.88				rbf	86.69
				sigmoid		77.35				sigmoid	75.92
				poly		79.01				poly	79.88
		20	linear	87.09			20	linear	86.53		
			rbf	88.28				rbf	87.09		
			sigmoid	76.80				sigmoid	76.48		
			poly	82.97				poly	80.12		

Lampiran 1 Ketetapan Model Klasifikasi Mesin Vektor Pendukung (Lanjutan)

<i>Hidden Layer Size</i>	<i>Window Size</i>	<i>Negative Sample Size</i>	<i>SVM Kernel</i>	<i>Accuracy</i>	<i>Hidden Layer Size</i>	<i>Window Size</i>	<i>Negative Sample Size</i>	<i>SVM Kernel</i>	<i>Accuracy</i>
500	2	15	linear	85.98	500	3	15	linear	86.93
			rbf	89.23				rbf	88.91
			sigmoid	85.43				sigmoid	81.55
			poly	82.10				poly	81.15
		16	linear	86.62			16	linear	86.30
			rbf	89.15				rbf	88.20
			sigmoid	85.90				sigmoid	80.13
			poly	84.55				poly	80.04
		17	linear	86.22			17	linear	86.62
			rbf	88.99				rbf	88.75
			sigmoid	85.43				sigmoid	81.55
			poly	83.92				poly	79.72
		18	linear	86.77			18	linear	87.71
			rbf	89.15				rbf	88.44
			sigmoid	86.46				sigmoid	80.99
			poly	82.73				poly	77.58
		19	linear	87.01			19	linear	86.54
			rbf	89.15				rbf	88.83
			sigmoid	86.30				sigmoid	79.25
			poly	83.37				poly	79.88
		20	linear	86.22			20	linear	86.62
			rbf	89.07				rbf	88.75
			sigmoid	86.38				sigmoid	80.84
			poly	84.71				poly	78.29
<i>Hidden Layer Size</i>	<i>Window Size</i>	<i>Negative Sample Size</i>	<i>SVM Kernel</i>	<i>Accuracy</i>	<i>Hidden Layer Size</i>	<i>Window Size</i>	<i>Negative Sample Size</i>	<i>SVM Kernel</i>	<i>Accuracy</i>
500	4	15	linear	86.93	500	5	15	linear	86.14
			rbf	88.28				rbf	87.88
			sigmoid	77.43				sigmoid	78.14
			poly	80.67				poly	78.93
		16	linear	86.78			16	linear	86.69
			rbf	88.99				rbf	87.64
			sigmoid	78.22				sigmoid	74.98
			poly	80.67				poly	79.09
		17	linear	87.17			17	linear	86.77
			rbf	89.07				rbf	88.43
			sigmoid	79.33				sigmoid	77.67
			poly	81.07				poly	80.43
		18	linear	86.77			18	linear	86.46
			rbf	88.83				rbf	88.04
			sigmoid	79.01				sigmoid	76.88
			poly	80.04				poly	78.85
		19	linear	87.41			19	linear	85.98
			rbf	88.60				rbf	87.48
			sigmoid	79.02				sigmoid	78.86
			poly	79.96				poly	79.88
		20	linear	86.38			20	linear	86.62
			rbf	88.75				rbf	87.88
			sigmoid	78.15				sigmoid	77.12
			poly	79.64				poly	81.46

Lampiran 1 Ketetapan Model Klasifikasi Mesin Vektor Pendukung (Lanjutan)

<i>Hidden Layer Size</i>	<i>Window Size</i>	<i>Negative Sample Size</i>	<i>SVM Kernel</i>	<i>Accuracy</i>	<i>Hidden Layer Size</i>	<i>Window Size</i>	<i>Negative Sample Size</i>	<i>SVM Kernel</i>	<i>Accuracy</i>
500	6	15	linear	86.22	500	7	15	linear	86.58
			rbf	87.72				rbf	87.64
			sigmoid	78.30				sigmoid	78.14
			poly	79.32				poly	80.28
		16	linear	86.14			16	linear	86.69
			rbf	88.04				rbf	87.72
			sigmoid	79.01				sigmoid	78.15
			poly	79.64				poly	79.17
		17	linear	87.25			17	linear	86.93
			rbf	87.57				rbf	87.64
			sigmoid	78.62				sigmoid	77.19
			poly	78.85				poly	80.44
		18	linear	86.45			18	linear	87.09
			rbf	87.96				rbf	86.68
			sigmoid	77.67				sigmoid	75.76
			poly	81.47				poly	80.67
		19	linear	86.45			19	linear	86.77
			rbf	88.04				rbf	87.33
			sigmoid	79.49				sigmoid	76.00
			poly	80.12				poly	80.04
		20	linear	86.62			20	linear	86.69
			rbf	87.57				rbf	87.33
			sigmoid	80.13				sigmoid	78.46
			poly	79.88				poly	79.41
<i>Hidden Layer Size</i>	<i>Window Size</i>	<i>Negative Sample Size</i>	<i>SVM Kernel</i>	<i>Accuracy</i>	<i>Hidden Layer Size</i>	<i>Window Size</i>	<i>Negative Sample Size</i>	<i>SVM Kernel</i>	<i>Accuracy</i>
600	2	15	linear	86.46	600	3	15	linear	87.41
			rbf	88.76				rbf	88.99
			sigmoid	85.43				sigmoid	80.28
			poly	81.86				poly	78.93
		16	linear	87.25			16	linear	86.53
			rbf	89.31				rbf	88.51
			sigmoid	85.99				sigmoid	80.36
			poly	84.08				poly	80.04
		17	linear	86.14			17	linear	87.09
			rbf	89.15				rbf	88.83
			sigmoid	86.14				sigmoid	81.55
			poly	83.92				poly	80.99
		18	linear	87.33			18	linear	87.33
			rbf	89.31				rbf	88.51
			sigmoid	85.03				sigmoid	80.84
			poly	81.70				poly	80.28
		19	linear	86.77			19	linear	86.53
			rbf	88.83				rbf	88.52
			sigmoid	85.98				sigmoid	81.23
			poly	83.05				poly	80.83
		20	linear	87.25			20	linear	86.54
			rbf	88.83				rbf	88.36
			sigmoid	86.14				sigmoid	82.26
			poly	83.84				poly	77.66

Lampiran 1 Ketetapan Model Klasifikasi Mesin Vektor Pendukung (Lanjutan)

<i>Hidden Layer Size</i>	<i>Window Size</i>	<i>Negative Sample Size</i>	<i>SVM Kernel</i>	<i>Accuracy</i>	<i>Hidden Layer Size</i>	<i>Window Size</i>	<i>Negative Sample Size</i>	<i>SVM Kernel</i>	<i>Accuracy</i>
600	4	15	linear	87.48	600	5	15	linear	86.38
			rbf	88.28				rbf	87.72
			sigmoid	80.59				sigmoid	77.27
			poly	79.88				poly	78.45
		16	linear	87.56			16	linear	86.69
			rbf	88.59				rbf	88.36
			sigmoid	77.99				sigmoid	76.09
			poly	79.09				poly	78.54
		17	linear	86.69			17	linear	86.53
			rbf	88.59				rbf	88.20
			sigmoid	79.57				sigmoid	78.15
			poly	80.59				poly	80.20
		18	linear	87.25			18	linear	86.54
			rbf	88.59				rbf	88.20
			sigmoid	78.62				sigmoid	77.66
			poly	78.85				poly	78.45
		19	linear	87.33			19	linear	86.61
			rbf	88.75				rbf	88.51
			sigmoid	78.70				sigmoid	80.12
			poly	79.33				poly	78.38
		20	linear	87.49			20	linear	86.77
			rbf	88.52				rbf	87.88
			sigmoid	76.88				sigmoid	77.98
			poly	78.29				poly	79.72
<i>Hidden Layer Size</i>	<i>Window Size</i>	<i>Negative Sample Size</i>	<i>SVM Kernel</i>	<i>Accuracy</i>	<i>Hidden Layer Size</i>	<i>Window Size</i>	<i>Negative Sample Size</i>	<i>SVM Kernel</i>	<i>Accuracy</i>
600	6	15	linear	86.69		7	15	linear	86.77
			rbf	87.88				rbf	87.88
			sigmoid	81.63				sigmoid	79.25
			poly	78.37				poly	79.88
		16	linear	86.53			16	linear	86.61
			rbf	88.44				rbf	87.72
			sigmoid	81.79				sigmoid	78.38
			poly	78.06				poly	79.64
		17	linear	86.53			17	linear	87.01
			rbf	88.04				rbf	87.88
			sigmoid	81.15				sigmoid	78.94
			poly	79.40				poly	80.04
		18	linear	86.61			18	linear	87.01
			rbf	88.12				rbf	87.64
			sigmoid	81.95				sigmoid	77.19
			poly	76.63				poly	79.88
		19	linear	86.77			19	linear	87.01
			rbf	88.28				rbf	88.04
			sigmoid	80.29				sigmoid	76.48
			poly	77.03				poly	79.80
		20	linear	87.09			20	linear	87.40
			rbf	88.20				rbf	87.88
			sigmoid	78.62				sigmoid	78.62
			poly	78.69				poly	79.56

Lampiran 2 Ketetapan Klasifikasi Mesin Vektor Pendukung Kernel RBF

C	γ	Akurasi
1	0.2875	89.87
0.1	1	57.64
	0.1	64.29
	0.001	82.65
	0.0001	57.64
1	1	61.83
	0.1	85.35
	0.001	86.69
	0.0001	82.66
10	1	61.83
	0.1	85.74
	0.001	88.59
	0.0001	86.3
100	1	61.83
	0.1	85.74
	0.001	88.2
	0.0001	88.36
1000	1	61.83
	0.1	85.74
	0.001	87.57
	0.0001	87.49

Lampiran 3 Ketetapan Parameter Pada Pemodelan Topik LDA Sentimen Positif

α	β	Jumlah Topik	<i>Coherence</i>	Kenaikan Nilai <i>Coherence</i>
0.1	0.1	5	0.44375	0.08826
	0.01	17	0.39846	0.03648
	0.001	8	0.39593	0.05577
0.01	0.1	5	0.42883	0.07303
	0.01	17	0.39737	0.05529
	0.001	8	0.40134	0.04467
0.001	0.1	5	0.42327	0.06747
	0.01	8	0.39987	0.04320
	0.001	17	0.38513	0.04880

Lampiran 4 Ketetapan Parameter Pada Pemodelan Topik LDA Sentimen Negatif

α	β	Jumlah Topik	<i>Coherence</i>	Kenaikan Nilai <i>Coherence</i>
0.1	0.1	5	0.38062	0.09092
	0.01	14	0.42701	0.05481
	0.001	5	0.38657	0.05156
0.01	0.1	5	0.37304	0.09661
	0.01	14	0.39859	0.05671
	0.001	5	0.36818	0.05929
0.001	0.1	5	0.37326	0.08644
	0.01	14	0.39683	0.05359
	0.001	14	0.39678	0.05141