

**DESAIN SENTIMEN ANALISIS DENGAN *SELF-ORGANIZING*
MAP UNTUK PEMILIHAN PRESIDEN 2024**

THESIS

Oleh :
DWI PURBO YUWONO
NIM. 200605210011



**PROGRAM STUDI MAGISTER INFORMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM
MALANG
2024**

**DESAIN SENTIMEN ANALISIS DENGAN *SELF-ORGANIZING MAP*
UNTUK PEMILIHAN PRESIDEN 2024**

THESIS

Diajukan kepada:

**Universitas Islam Negeri Maulana Malik Ibrahim Malang
Untuk memenuhi Salah Satu Persyaratan dalam
Memperoleh Gelar Magister Komputer (M.Kom)**

Oleh :

**Dwi Purbo Yuwono
200605210011**

**PROGRAM STUDI MAGISTER INFORMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM
MALANG
2024**

DESAIN SENTIMEN ANALISIS DENGAN *SELF-ORGANIZING MAP*
UNTUK PEMILIHAN PRESIDEN 2024

THESIS

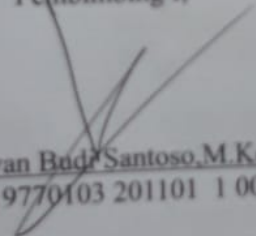
Oleh :

Dwi Purbo Yuwono
200605210011

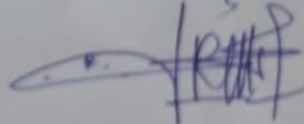
Telah diperiksa dan disetujui untuk diuji:

Tanggal:

Pembimbing I,



Dr. Irwan Budi Santoso, M.Kom
NIP. 19770103 201101 1 004

Pembimbing II,


Dr. Ririen Kusumawati, S.Si., M.M.Kom
NIP. 19720309 200501 2 002

Mengetahui,
Ketua Program Studi Magister Informatika
Fakultas Sains dan Teknologi
Universitas Islam Negeri Maulana Malik Ibrahim Malang




M. Cs
NIP. 19740424 200901 1 008

DESAIN SENTIMEN ANALISIS DENGAN *SELF-ORGANIZING MAP*
UNTUK PEMILIHAN PRESIDEN 2024

THESIS

Oleh :
Dwi Purbo Yuwono
200605210011

Telah dipertahankan di Depan Dewan Penguji Thesis
Dan Dinyatakan Diterima Sebagai Salah Satu Persyaratan
Untuk memperoleh Gelar Magister Komputer (M.Kom)
Tanggal:

Susunan Dewan Penguji:


Penguji 1 : Dr. Cahyo Crysdian, M. Cs
NIP.19740424 200901 1 008

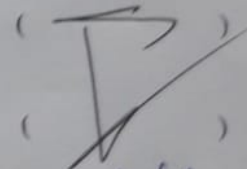
Penguji 2 : Dr. Usman Pagalay
NIP. 19650414 200312 1 001

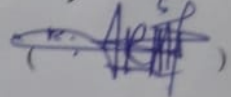
Pembimbing 1 : Dr. Irwan Budi Santoso, M.Kom
NIP 19770103 201101 1 004

Pembimbing 2 : Dr. Ririen Kusumawati, S.Si., M.M.Kom
NIP. 19720309 200501 2 002

Tanda Tangan

()

()

()

Mengetahui,
Ketua Program Studi Magister Informatika
Fakultas Sains dan Teknologi
Universitas Islam Negeri Maulana Malik Ibrahim Malang



Dr. Cahyo Crysdian, M. Cs
NIP. 19740424 200901 1 008

PERNYATAAN KEASLIAN TULISAN

Yang bertanda tangan di bawah ini saya:

Nama : Dwi Purbo Yuwono
NIM : 200605210011
Program : Magister Informatika
Fakultas : Sains dan Teknologi

Menyatakan dengan sebenarnya bahwa Thesis yang saya tulis ini benar-banar merupakan hasil karya saya sendiri, bukan merupakan pengambilalihan data, tulisan atau pikiran orang lain yang saya akui sebagai hasil tulisan atau pikiran saya sendiri, kecuali dengan mencantumkan sumber cuplikan pada daftar pustaka.

Apabila dikemudian hari terbukti atau dapat dibuktikan Thesis ini hasil jiplakan, maka saya bersedia menerima sanksi atas perbuatan tersebut.

Malang, 12 Juni 2024

Saya yang menyatakan,



Dwi Purbo Yuwono
NIM. 200605210011

MOTTO

"Pengetahuan adalah dasar dari semua kebajikan, dan kebodohan adalah akar dari semua kejahatan"

Ibnu Rusyd dari kitab "Tahafut al-Tahafut".

PERSEMBAHAN

Dengan mengucapkan syukur Alhamdulillah rabbil alamin, Thesis ini saya persembahkan untuk :

1. Allah SWT yang memberikan rahmat, hidayah, kesehatan, rejeki, serta semua yang saya butuhkan sampai selesainya tesis ini.
2. Rosulullah Muhammad SAW yang telah menerangi alam ini dengan *nur tauhidnya*. Berkat ajaran beliau melalui Al-Quran dan sunnah, membuka era ilmu pengetahuan, cahaya dan sains.
3. Seluruh keluarga tercinta (istri Siti Robi'ah, anak perempuan Khadijah Aqifa Naila, anak laki-laki Hasan Abdullah Sattar), yang selalu memberikan dukungan, doa dan semangat yang tiada henti.
4. Seluruh orang tua tersayang (bapak Tri Wahyu Busono dan ibu Nyuwarsini tercinta, serta bapak Subakir, dan ibu Alm. Munzaiyah).
5. Bapak Dr. Irwan Budi Santoso, M.Kom. dan Ibu Dr. Ririen Kusumawati, S.Si., M.M.Kom. selaku dosen pembimbing yang sudah banyak memberikan ilmu, arahan dan bimbingannya.
6. Bapak Dr. Cahyo Crysdiand dan Bapak Dr. Usman Pagalay, selaku tim penguji tesis yang sudah banyak membantu terselesainya tesis saya.
7. Seluruh civitas akademika Universitas Islam Negeri Maulana Malik Ibrahim Malang untuk dapat menimba ilmu, pengalaman dan khususnya tentang ilmu agama.
8. Kakak Muchlis Fanani yang memberikan dukungan moril, materiil dan kesempatan untuk dapat studi lanjut.
9. Rekan-rekan sekalian yang tidak bisa saya sebutkan satu persatu yang telah mendukung penyelesaian tesis ini.

KATA PENGANTAR

Assalamu 'alaikum Wr. Wb.

Syukur alhamdulillah penulis hanturkan kehadiran Allah SWT yang telah melimpahkan Rahmat dan Hidayah-Nya. Serta sholawat dan salam selalu tercurahkan kepada Rosulullah Muhammad SAW yang telah menerangi alam ini dengan *nur tauhidnya*. Berkat ajaran beliau melalui Al-Quran dan sunnah, membuka era ilmu pengetahuan, cahaya dan sains. Sehingga penulis dapat menyelesaikan studi di Program Studi Magister Informatika Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang sekaligus menyelesaikan tesis ini dengan baik.

Selanjutnya penulis haturkan ucapan terima kasih seiring do'a dan harapan *jazakumullah ahsanal jaza'* kepada semua pihak yang telah membantu terselesaikannya tesis ini. Ucapan terima kasih ini penulis sampaikan kepada:

1. Bapak Dr. Irwan Budi Santoso, M.Kom. dan Ibu Dr. Ririen Kusumawati, S.Si., M.M.Kom. selaku dosen pembimbing yang sudah banyak memberikan ilmu, arahan dan bimbingannya.
2. Bapak Dr. Cahyo Crysdiyan dan Bapak Dr. Usman Pagalay, selaku tim penguji tesis yang sudah banyak membantu terselesainya tesis saya.
3. Seluruh civitas akademika Universitas Islam Negeri Maulana Malik Ibrahim Malang untuk dapat menimba ilmu, pengalaman dan khususnya tentang ilmu agama.
4. Ayahanda dan Ibunda tercinta yang senantiasa memberikan doa dan restunya kepada penulis dalam menuntut ilmu.
5. Keluarga penulis yang selalu memberikan semangat kepada penulis untuk menyelesaikan tesis ini.
6. Semua pihak yang ikut membantu dalam menyelesaikan tesis ini yang tidak bisa penulis sebutkan satu persatu.

Penulis menyadari bahwa dalam penyusunan tesis ini masih terdapat kekurangan dan penulis berharap semoga tesis ini bias memberikan manfaat kepada para pembaca khususnya bagi penulis secara pribadi. Amin Ya Rabbal Alamin.

Wassalamu 'alaikum Wr. Wb.

Malang, Juni 2024

Penulis

DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PENGAJUAN	ii
HALAMAN PERSETUJUAN	iii
HALAMAN PENGESAHAN	iii
HALAMAN PERNYATAAN KEASLIAN TULISAN	v
MOTTO	vi
HALAMAN PERSEMBAHAN	vii
KATA PENGANTAR	viii
DAFTAR ISI	ix
DAFTAR GAMBAR	xi
DAFTAR TABEL	xii
ABSTRAK	xiii
ABSTRACT	xiv
ملخص	xv
BAB I PENDAHULUAN	1
1.1. Latar Belakang	1
1.2. Pernyataan Masalah	7
1.3. Tujuan Penelitian	8
1.4. Batasan Penelitian	8
1.5. Manfaat Penelitian	8
1.6. Sistematika Penulisan	9
BAB II STUDI PUSTAKA	10
2.1. Sentimen Masyarakat	10
2.2. Metode <i>Extraction feature</i>	12
2.3. Metode Klusterisasi	15
2.4. Kerangka Teori	18
BAB III METODE PENELITIAN	25
3.1. Lokasi dan Waktu Penelitian	25

3.2.	Alur Penelitian.....	25
3.2.1.	Pengumpulan Data	26
3.2.2.	Pengembangan Sistem.....	30
3.3.	<i>Extraction Fearure</i>	41
3.3.1.	TF-IDF (<i>Term Frequency-Inverse Document Frequency</i>).....	41
3.3.2.	<i>Latent Semantic Analysis (LSA)</i>	43
3.4.	Permodelan <i>Self-Organizing Maps (SOM)</i>	46
3.5.	Uji Coba	48
3.6.	Konseptual Model	48
BAB IV TERM FREQUENCY-INVERSE DOCUMENT FREQUENCY (TF-IDF) DAN SELF-ORGANIZING MAPS (SOM)		53
4.1.	Desain.....	53
4.2.	Implementasi	54
BAB V LATENT SEMANTIC ANALYSIS (LSA) DAN SELF-ORGANIZING MAPS (SOM)		61
5.1.	Desain.....	61
5.2.	Implementasi	62
BAB VI PEMBAHASAN		69
6.1.	Pembahasan Komparasi Performa Algoritma	69
6.2.	Sentimen Analisis Dalam Al-Qur'an	72
BAB VII KESIMPULAM		77
7.1.	Kesimpulan.....	77
7.2.	Saran.....	77
DAFTAR PUSTAKA		79

DAFTAR GAMBAR

Gambar 1.1 Jumlah Pengguna Media Sosial Twitter Secara Global	3
Gambar 1.2 Gambaran opini para pengguna media sosial di Twitter.....	4
Gambar 2.1. Kerangka Teori.....	23
Gambar 3.1 Desain Sistem Penelitian yang diusulkan.....	26
Gambar 3.2 Pengembangan Sistem.....	30
Gambar 3.4 Ilustrasi struktur layer pada SOM.....	46
Gambar 3.5 Konseptual Model Penilitan.....	50
Gambar 4.1. Alur Proses Analisis Sentimen TF-IDF + SOM.....	54
Gambar 4.2. Data Perbandingan.....	55
Gambar 4.3. Confusion Matrix TF-IDF + SOM.....	57
Gambar 4.4. Sentiment Clustering – TF-IDF + SOM.....	58
Gambar 5.1. alur proses analisis sentimen LSA + SOM.....	62
Gambar 5.2. Distribution of Labels in LSA Data.....	63
Gambar 5.3. Confusion matrix LSA+SOM.....	66
Gambar 5.4. Sentiment Clustering – LSA+SOM.....	67
Gambar 6.1. Perbandingan Antara TF-IDF+SOM dan LSA+SOM.....	71

DAFTAR TABEL

Tabel 2.1. Literature Review	18
Tabel 3.1. Hasil Pengumpulan Data Twitter.....	27
Tabel 3.2. Hasil Cleaning.....	30
Tabel 3.3. Hasil Case Folding.....	32
Tabel 3.4. Hasil Tokenizing.....	34
Tabel 3.5. Cuitan Hasil Normalisasi.....	35
Tabel 3.6. Hasil Filtering Stopwords.....	37
Tabel 3.7. Cuitan Hasil Stemming.....	39
Tabel 3.8. Pelebelan data	40
Tabel 4.1. Data perbandingan.....	55
Tabel 4.2. Akurasi Keseluruhan.....	56
Tabel 5.1. Tabel Hasil Pembagian.....	63
Tabel 5.2. Evaluasi Performa Model.....	64
Tabel 6.1. Tabel Komparasi TF-IDF+LSOM & LSA+SOM.....	69

ABSTRAK

Yuwono, Dwi Purbo, 2024, **Desain Sentimen Analisis Dengan Self-Organizing Map Untuk Pemilihan Presiden 2024**, Program Magister Informatika Universitas Islam Negeri Maulana Malik Ibrahim, Pembimbing: (1) Dr. Irwan Budi Santoso, M.Kom (2) Dr. Ririen Kusumawati, S.Si., M.M.Kom

Kata Kunci: *Sentimen, Analisis, Pemilu 2024, SOM.*

Pemilihan Umum (PEMILU) adalah salah satu cara untuk memilih presiden, kepala daerah, dan anggota parlemen yang berlangsung setiap lima tahun sekali. Dalam memasuki tahun-tahun politik saat ini akan banyak bertebaran informasi dan komentar dari masyarakat terhadap pelaksanaan pemilu, komentar atau pendapat yang disampaikan akan sangat beragam dimulai dari dukungan terhadap pelaksanaan pemilu, penggiringan opini publik, ujaran kebencian dan komentar-komentar lainnya. Kemajuan teknologi saat ini mengakibatkan penyampaian pendapat dapat dengan mudah dipublikasikan melalui media sosial, salah satunya adalah melalui media twitter, twitter menjadi salah satu media sosial yang paling sering digunakan masyarakat dalam mengemukakan pendapatnya karena dianggap bebas. Oleh karena itu, pada penelitian ini diusulkan desain sentimen analisis tentang Pilpres melalui X-Twitter, baik itu positif, negatif, atau netral dengan menggunakan dua metode yaitu *Term Frequency-Inverse Document Frequency* (TF-IDF) dan *Latent Semantic Analysis* (LSA) yang kemudian dianalisis dengan algoritma *Self-Organizing Maps* (SOM). Dari hasil penelitian didapatkan bahwa 1) Model LSA+SOM memiliki akurasi yang lebih tinggi (0.569) dibandingkan dengan model TF-IDF+SOM (0.543), menunjukkan bahwa LSA memberikan peningkatan performa dalam analisis teks. 2) Perbedaan akurasi antara kedua model hanya sebesar 0.026, yang menunjukkan bahwa kedua metode memiliki performa yang relatif dekat. 3) LSA mampu menangkap hubungan laten antara kata-kata dan dokumen lebih baik daripada TF-IDF, yang mungkin berkontribusi pada peningkatan akurasi dalam model LSA+SOM. 4) Meskipun LSA+SOM menunjukkan akurasi yang lebih tinggi, TF-IDF+ SOM tetap merupakan alternatif yang kompetitif, tergantung pada kebutuhan spesifik dari analisis teks yang dilakukan. 5) Selain akurasi, faktor lain seperti waktu pemrosesan, kompleksitas model, dan interpretabilitas hasil juga penting untuk dipertimbangkan dalam pemilihan model.

ABSTRACT

Yuwono, Dwi Purbo, 2024, **Design Sentiment Analysis With a Self-Organizing Map for 2024 Presidential Election**, Master Program in Informatics, Maulana Malik Ibrahim State Islamic University, Promotor: (1) Dr. Irwan Budi Santoso, M.Kom (2) Dr. Ririen Kusumawati, S.Si., M.M.Kom

Keywords: *Sentiment, Analysis, Election 2024, SOM.*

General Elections (PEMILU) are one way to elect the president, regional heads and members of parliament which take place every five years. As we enter the current political years, there will be a lot of information and comments from the public regarding the implementation of elections, the comments or opinions expressed will be very diverse, starting from support for the implementation of elections, leading public opinion, hate speech and other comments. Current technological advances mean that expressing opinions can be easily published via social media, one of which is via Twitter. Twitter is one of the social media that is most often used by people to express their opinions because it is considered free. Therefore, in this research, we propose a sentiment analysis design regarding the Presidential Election via then analyzed using the Self-Organizing Maps (SOM) algorithm. From the research results, it was found that 1) The LSA + SOM model had higher accuracy (0.569) compared to the TF-IDF + SOM model (0.543), indicating that LSA provides increased performance in text analysis. 2) The difference in accuracy between the two models is only 0.026, which shows that the two methods have relatively close performance. 3) LSA is able to capture latent relationships between words and documents better than TF-IDF, which may contribute to the increased accuracy in the LSA + SOM model. 4) Although LSA+ SOM shows higher accuracy, TF-IDF + SOM remains a competitive alternative, depending on the specific needs of the text analysis being performed. 5) Apart from accuracy, other factors such as processing time, model complexity, and interpretability of results are also important to consider in model selection.

ملخص

يونو، دوي بوربو، 2024، تصميم تحليل المشاعر بخريطة التنظيم الذاتي للانتخابات الرئاسية 2024، برنامج ماجستير المعلوماتية، جامعة مولانا مالك إبراهيم الإسلامية الحكومية، المشرف: (1) د. إروان بودي سانتوسو. M.Kom. (2) د. ريرين كوسوماواي، S.Si، M.Kom.

الكلمات المفتاحية: انتخابات 2024، تويتر، TF-IDF، SOM.

الانتخابات العامة هي إحدى الطرق لانتخاب الرئيس ورؤساء المناطق وأعضاء البرلمان والتي تجرى كل خمس سنوات. مع دخولنا السنوات السياسية الحالية، سيكون هناك الكثير من المعلومات والتعليقات من الجمهور فيما يتعلق بتنفيذ الانتخابات، وستكون التعليقات أو الآراء المعبر عنها متنوعة للغاية، بدءًا من دعم تنفيذ الانتخابات، وقيادة الرأي العام، والكراهية الكلام والتعليقات الأخرى. إن التقدم التكنولوجي الحالي يعني أن التعبير عن الآراء يمكن نشره بسهولة عبر وسائل التواصل الاجتماعي، وأحدها عبر تويتر هو أحد وسائل التواصل الاجتماعي التي يستخدمها الناس في كثير من الأحيان للتعبير عن آرائهم لأنه يعتبر مجانيًا. لذلك، في هذا البحث، نقترح تصميم تحليل المشاعر تجاه الانتخابات الرئاسية من خلال تحليلها باستخدام خوارزمية خرائط التنظيم الذاتي (SOM). ومن نتائج البحث وجد أن (1) يتمتع نموذج LSA + SOM بدقة أعلى (0.569) مقارنة بنموذج (0.543 TF-IDF + SOM)، مما يشير إلى أن LSA يوفر أداءً متزايدًا في تحليل النص. (2) الفرق في الدقة بين النموذجين هو 0.026 فقط، مما يدل على أن الطريقتين لهما أداء متقارب نسبيًا. (3) LSA قادر على التقاط العلاقات الكامنة بين الكلمات والمستندات بشكل أفضل من TF-IDF، مما قد يساهم في زيادة الدقة في نموذج LSA + SOM. (4) على الرغم من أن LSA + SOM يُظهر دقة أعلى، إلا أن TF-IDF + SOM يظل بديلًا تنافسيًا، اعتمادًا على الاحتياجات المحددة لتحليل النص الذي يتم إجراؤه. (5) وبصرف النظر عن الدقة، فإن هناك عوامل أخرى مثل وقت المعالجة، وتعقيد النموذج، وقابلية تفسير النتائج، من المهم أيضًا أخذها في الاعتبار عند اختيار النموذج.

BAB I

PENDAHULUAN

1.1. Latar Belakang

Pemilihan Umum (PEMILU) merupakan salah satu cara untuk memilih Presiden, Pimpinan Daerah, dan Anggota DPR. PEMILU dilaksanakan setiap lima tahun sekali. Tahun 2024 merupakan tahun politik bagi Indonesia, karena pada tahun 2024 KPU telah menyelenggarakan pemilu legislatif seperti Dewan Perwakilan Rakyat (DPR RI) atau Dewan Perwakilan Rakyat Daerah (DPRD) baik tingkat provinsi, kabupaten, maupun kota, serta pemilihan presiden dan wakil presiden, untuk periode 2024-2029. Kampanye politik yang dilakukan oleh tim sukses serta elemen pendukung salah satu partai atau aktor politik merupakan salah satu untuk meraih suara terbanyak saat pemilihan (Junaidi dkk, 2015:6–8).

Pemilu dilaksanakan oleh penyelenggara pemilu yang jujur, profesional dan bertanggung jawab. Pemilu dilaksanakan secara bermutu, sistematis, sah secara hukum, dan dapat dipertanggungjawabkan serta mencakup partisipasi masyarakat luas. Penyelenggara pemilu, pejabat pemerintah, peserta pemilu, pengawas pemilu, pemantau pemilu, pemilih, dan semua pihak yang terkait harus berperilaku dan bertindak jujur sesuai dengan peraturan perundang-undangan yang berlaku (Ardiansyah, 2019:13). Penerapannya kerap memunculkan persoalan terkait kecurangan, polarisasi politik, dan perbedaan opini masyarakat terhadap kandidat atau partai tertentu (Yusuf 2022:34). Oleh karena itu, penting agar pemilih dan peserta pemilu diperlakukan secara adil dan tanpa adanya kecurangan dan perlakuan tidak adil dari partai politik mana pun. Pemilu harus diselenggarakan

dengan kualitas yang lebih tinggi untuk memastikan persaingan yang sehat, inklusif, keterwakilan yang lebih besar, dan mekanisme akuntabilitas yang jelas.

Dalam hal ini, tahun politik merupakan tahun yang sering dinanti-nantikan oleh sebagian masyarakat, karena setiap mendekati tahun politik, setiap peristiwa pastinya mempunyai banyak fenomena politik yang unik atau peristiwa yang terjadi serupa dengan tahun-tahun sebelumnya, salah satunya berbeda. komentar masyarakat khususnya mengenai pemilu presiden dan wakil presiden tahun 2024. Untuk tahun-tahun politik yang akan datang masyarakat akan banyak menerima informasi dan komentar mengenai pelaksanaan pemilu, komentar atau pendapat yang dipublikasikan sangat beragam, mulai dari dukungan terhadap pemilu. pelaksanaan pemilu, pembentukan opini publik, ujaran kebencian, dll. komentar (Manullang dan Prianto, 2023:12). Perkembangan teknologi modern semakin memudahkan dalam mengutarakan pendapat di jejaring sosial. Salah satu caranya adalah melalui Twitter. Twitter merupakan salah satu media sosial yang banyak digunakan masyarakat untuk menyampaikan pendapatnya karena dianggap gratis. (Buntoro, 2017:57).

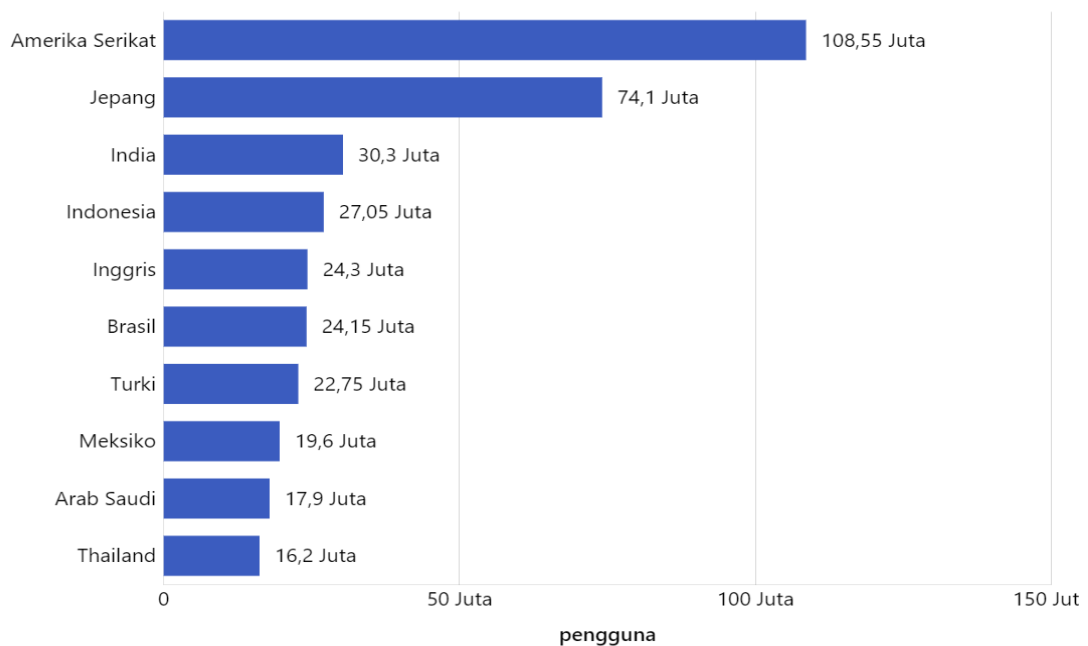
Hal ini sejalan dengan penyampaian suatu berita yang bersifat opini publik juga dijelaskan dalam ayat Al-Quran pada surat An-Nissa:83, sebagai berikut;

وَإِذَا جَاءَهُمْ أَمْرٌ مِّنَ الْأَمْنِ أَوْ الْخَوْفِ أَذَاعُوا بِهِ ۖ وَلَوْ رَدُّوهُ إِلَى الرَّسُولِ وَإِلَىٰ أُولِي الْأَمْرِ
 مِنْهُمْ لَعَلِمَهُ الَّذِينَ يَسْتَنْبِطُونَهُ مِنْهُمْ ۗ وَلَوْلَا فَضْلُ اللَّهِ عَلَيْكُمْ وَرَحْمَتُهُ لَاتَّبَعْتُمُ الشَّيْطَانَ إِلَّا
 قَلِيلًا

Artinya: “Dan apabila datang kepada mereka suatu berita tentang keamanan ataupun ketakutan, mereka lalu menyiarkannya. Dan kalau mereka menyerahkannya kepada Rasul dan Ulil Amri di antara mereka, tentulah

orang-orang yang ingin mengetahui kebenarannya (akan dapat) mengetahuinya dari mereka (Rasul dan Ulil Amri). Kalau tidaklah karena karunia dan rahmat Allah kepada kamu, tentulah kamu mengikut syaitan, kecuali sebahagian kecil saja (di antaramu).” (QS. An-Nissa:83)

Berdasarkan data yang diperoleh peneliti dari databook.katadata.co.id, pada gambar 1.1 per bulan Januari 2023, negara Indonesia menduduki peringkat ke-3 dari rangking 10 negara pengguna aplikasi media sosial Twitter terbanyak dengan 27,05 Juta pengguna aktif. Hal ini menunjukkan bahwa banyaknya masyarakat Indonesia yang aktif menggunakan media sosial Twitter.



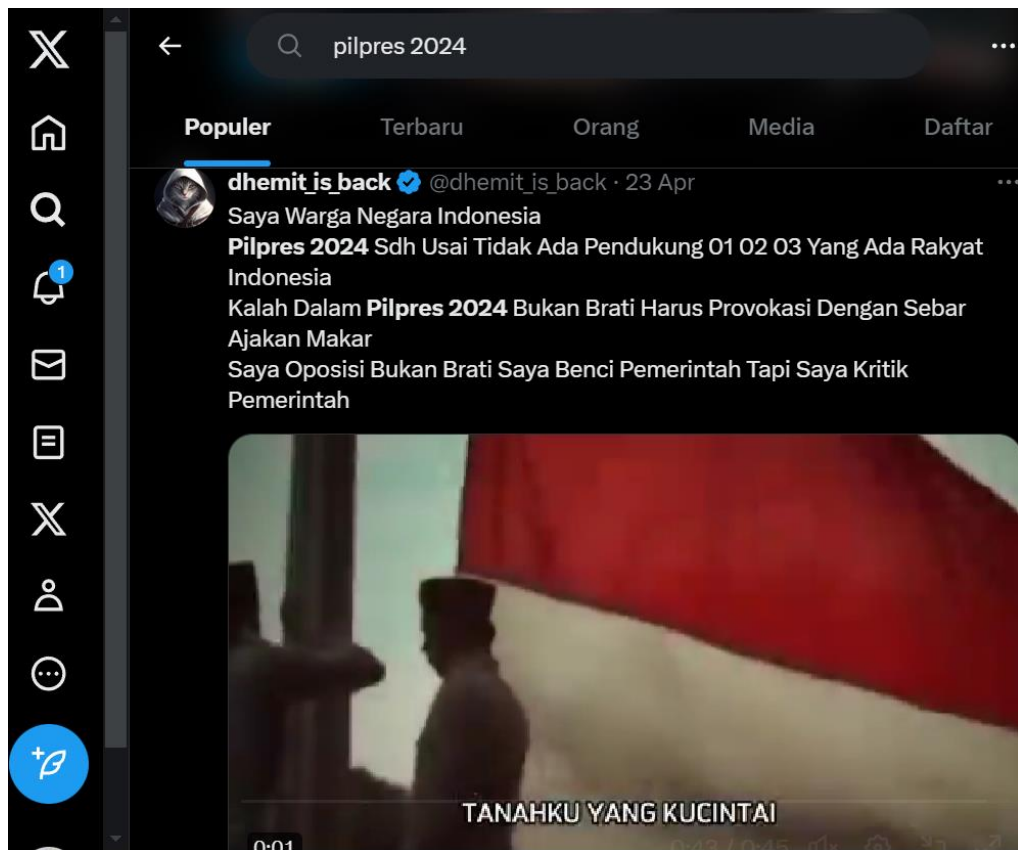
Sumber:
We Are Social

Informasi Lain:

Gambar 1.1 Jumlah Pengguna Media Sosial Twitter Secara Global

Twitter menjadi sarana komunikasi yang sangat bebas dan luas sehingga memasuki tahun politik ini banyak sekali spekulasi masyarakat mengenai pemilu presiden 2024, komentar yang diberikan pun sangat beragam sehingga menimbulkan fenomena politik, karena informasi yang diberikan oleh masing-

masing pihak sangat beragam. Seseorang itu unik dan berbeda, sehingga perlu adanya analisis masyarakat terhadap perilaku Pilpres 2024. Dengan melakukan analisis opini publik yang menunjukkan opini masyarakat tentang Pilpres 2024 di jejaring sosial Twitter, kita dapat menemukan untuk mengetahui reaksi masyarakat Indonesia terhadap peristiwa tersebut, selain itu dapat dijadikan acuan untuk menghindari berita bohong dan opini negatif serta ujaran kebencian dari salah satu pihak. Selain itu, banyak opini publik yang netral, seperti yang terlihat pada kutipan gambar di bawah ini.



Gambar 1.2 Gambaran opini para pengguna media sosial di Twitter

Penentuan polaritas positif atau negatif suatu opini dapat dilakukan secara manual, namun seiring dengan bertambahnya jumlah sumber opini, waktu dan tenaga yang diperlukan untuk mengklasifikasikan polaritas opini tersebut tentu saja

semakin bertambah. Gambar 1.2 menunjukkan bahwa terdapat banyak jajak pendapat dan sentimen masyarakat yang mengutarakan pendapat dan pemikirannya. Oleh karena itu, perlu diterapkan metode pembelajaran mesin untuk mengklasifikasikan polaritas opini dari sumber data yang besar tersebut. Untuk ini, beberapa fungsi penambangan teks dapat digunakan, dalam hal ini klasifikasi dokumen (Dewi dkk, 2023:12).

Rancangan analisis sentimen dilakukan untuk mengkategorikan berbagai opini masyarakat terkait pemilu presiden tahun 2024. Analisis sentimen diartikan sebagai proses sistematis dalam mendeskripsikan masyarakat, yang meliputi pengumpulan informasi mengenai profil, ide, opini, perasaan, dan kebutuhan masyarakat. masyarakat ini (Hasim & Remiswal, 2009:159). Sedangkan perasaan itu sendiri adalah suatu pendapat atau pandangan yang didasarkan pada perasaan yang berlebihan terhadap suatu hal (berlawanan dengan pendapat pikiran).

Analisis sentimen, atau penambangan opini, adalah kombinasi dari penambangan data dan penambangan teks, suatu teknik yang digunakan untuk menganalisis opini, perasaan, penilaian, sikap, penilaian, perasaan dan emosi seseorang. Terlepas dari apakah pembicara atau peneliti tertarik pada hal tema, produk, layanan, organisasi, aktivitas individu atau hal-hal lain secara spesifik (Hakimi 2018:34). Dengan bantuan analisis sentimen, peneliti dapat mengetahui apakah isi teks tersebut positif atau negatif ataupun netral. Ambience mengacu pada fokus suatu topik tertentu, suatu pernyataan tentang suatu topik dapat mempunyai arti yang berbeda dengan pernyataan yang sama tentang topik yang berbeda,

sehingga beberapa penelitian mengawalinya dengan mendefinisikan unsur-unsur produk yang sedang dibahas sebelum memulai (Abdillah dan Hasan 2023:56).

Ada beberapa jenis algoritma untuk melakukan analisis sentimen, salah satunya adalah *Self-Organizing Maps* (SOM). SOM merupakan perpanjangan dari jaringan permusuhan, sering disebut jaringan Kohonen, dengan metode pembelajaran tanpa pengawasan. Algoritma ini juga mempunyai kelebihan dalam pelatihannya yang sangat cepat dan mudah sehingga dapat memudahkan (Rahmah 2022:56). Secara garis besar sistem yang diterapkan dimulai dengan pengumpulan dataset, setelah itu dataset yang terkumpul diberi tanda dan kemudian dilakukan tahap pra-pemrosesan. Setelah itu, data diperoleh dari proses *preprocessing* yang diberi bobot dengan *Term-Frequency Inverse Document Frequency* (TF-IDF). Hasil dari ekstraksi TF-IDF merupakan ukuran statistik berupa matrik yang menggambarkan pentingnya suatu istilah dari suatu dokumen dalam suatu koleksi atau korpus. Metrik ini sering digunakan untuk faktor pembobotan sebagai pengambilan informasi, *text mining*, dan pemodelan *user*. Nilai dari TF-IDF meningkat secara proporsional dengan banyaknya kemunculan ekspresi dan bergantung pada banyaknya dokumen di korpus tempat ekspresi tersebut berada (Rajaraman, A.; Ullman, J. D, 2011:1–17).

Dalam hal ini, TF-IDF memasuki proses klasifikasi menggunakan algoritma SOM, yaitu teknik pembelajaran mesin tanpa pengawasan yang digunakan untuk memperoleh representasi berdimensi rendah (biasanya dua dimensi) dari kumpulan data berdimensi lebih tinggi dengan tetap mempertahankan struktur topologinya. Kemudian memberikan penilaian terhadap sikap komunitas pengguna Twitter

terhadap pemilu 2024, yang kemudian menunjukkan apakah informasi tersebut masuk dalam kategori positif, negatif, atau netral.

Alasan penggunaan model analisis sentimen TF-IDF adalah proses yang dijalankan guna untuk mengekstrak fitur dari teks dengan memberikan nilai di setiap kata pada dataset sentimen tweet Twitter. Penghitungan kemudian dilakukan untuk menentukan bobot teks setiap kata. TF-IDF bertujuan untuk melihat seberapa penting per-kata dalam kalimat. Penilaian atau pembobotan pada TF-IDF didasarkan pada frekuensi kata dalam suatu dokumen (Yusuf 2022:45).

Dilihat dari opini masyarakat, pemilu dapat memberikan tanggapan positif dan negatif maupun netral. Dalam konteks ini, analisis sentimen melalui media sosial merupakan cara yang efektif untuk mengetahui opini dan pandangan masyarakat terhadap penyelenggaraan pemilu 2024 (Afrizal dkk, 2019:43). Analisis sentimen media sosial memungkinkan untuk mengidentifikasi dan menganalisis ekspresi emosional, opini, atau tanggapan seseorang terhadap topik tertentu yang dibagikan melalui media sosial (Yusuf, Fauzi, dan Brata, 2018:13). Oleh karena itu, penelitian ini mengusulkan untuk menganalisis opini pemilu presiden melalui X-Twitter, baik positif, negatif atau netral, dengan menggunakan dua metode ekstraksi fitur yaitu TF-IDF dan LSA. Hasil dari kedua pembobotan tersebut dianalisis menggunakan algoritma Self-Organizing Maps (SOM).

1.2. Pernyataan Masalah

Mengacu pada latar belakang di atas maka pernyataan masalah yang dinyatakan dalam penelitian ini adalah bagaimana cara membuat sistem analisis

sentimen yang efektif pada data teks terkait Pilpres 2024 dengan menggunakan metode *Self-Organizing Maps* (SOM)?

1.3. Tujuan Penelitian

Tujuan dari penelitian ini adalah mengembangkan model analisis sentimen dengan menggunakan *self-organizing map* (SOM) yang berguna untuk:

1. Memahami pola dan variasi opini publik terkait Pilpres 2024.
2. Menyajikan visualisasi yang intuitif untuk memperlihatkan distribusi sentimen secara spasial.
3. Mengidentifikasi kluster atau kelompok opini yang dominan dalam data sentimen terkait Pilpres 2024.

1.4. Batasan Penelitian

Penelitian ini mempunyai beberapa batasan pada ruang lingkup pelaksanaan penelitian, antara lain:

1. Penelitian dilakukan dengan menggunakan data tweet media sosial Twitter yang diperoleh dari Twitter dengan cara *crawling*.
2. *Dataset* yang digunakan adalah data yang diambil pada *tweets* yang terkumpul dalam jangka bulan 14 Februari sampai 30 April 2024, dan *tweets* berbahasa Indonesia dengan *query* pencarian yaitu “pilpres 2024”.

1.5. Manfaat Penelitian

1. Manfaat Bagi Masyarakat Umum

Mendapatkan informasi tentang pendapat komunitas pengguna Twitter tentang pemilu Republik Indonesia tahun 2024 dan mengetahui seberapa besar partisipasi masyarakat pasca pemilu.

2. Manfaat Bagi Akademik

Ikut serta dalam pengembangan bidang ilmu *text mining* dan penerapannya pada *sentiment analysis*. Pelajari lebih lanjut tentang menggunakan Twitter sebagai alat penelitian untuk menganalisis sentimen terhadap isu tertentu.

1.6. Sistematika Penulisan

Peneliti membagi tesis menjadi tiga bab, yaitu bab pendahuluan; Bab ini menjelaskan tentang latar belakang, identifikasi masalah, rumusan masalah, batasan masalah, tujuan masalah, kelebihan penelitian dan sistematika penelitian. Bab 2 Tinjauan Pustaka; Bab ini menyajikan teori-teori pendukung dari buku, jurnal, e-book, website dan penelitian sebelumnya mengenai sentimen publik, algoritma analisis teks dan metode clustering sebagai landasan penelitian ini, serta gambaran umum perangkat lunak yang dikembangkan. Bab 3 Metode Penelitian; Bab ini membahas tentang metode penelitian yang digunakan, alat penelitian dan langkah-langkah penelitian. Bab 4; Hasil penelitian dibahas; pembobotan menggunakan metode Tf-Idf (*Term-Frequency Inverse Document Frequency*) dan metode *self-organizing map* (SOM). Bab 5; Metode Latent Semantic Analysis (LSA) dan *Self-Organizing Map* (SOM). Bab 6 adalah pembahasan; Diskusi kinerja komparatif dari algoritma penugasan tugas pelengkap dan sistem rekomendasi berdasarkan metode klasifikasi Al-Quran. Dan Bab 7 berisi kesimpulan dan saran.

BAB II

STUDI PUSTAKA

2.1. Sentimen Masyarakat

Terdapat beragam penelitian terkait sentimen publik, dalam hal ini menjadi landasan peneliti untuk mengembangkan dan memperdalamnya diantaranya.

Kajian Damayanti dkk. (2024) tentang transformasi kebijakan pembayaran BPJS Kesehatan sering menjadi perbincangan populer di internet. Tujuan penelitian ini adalah melakukan analisis sentimen terhadap transformasi iuran BPJS Kesehatan dengan algoritma SVM (*Support Vector Machine*). Fokus utamanya adalah memahami tanggapan masyarakat yang beradar pada komentar-komentar di media sosial online, dimana SVM digunakan mengklasifikasikan teks, menyortir pandangan masyarakat dan mengidentifikasi potensi implikasi kebijakan.

Kajian tersebut membahas aspek-aspek penting khususnya analisis sentimen, penerapan SVM serta kontribusi penelitian terhadap pengembangan metode analisis sentimen, memahami respon masyarakat terhadap transformasi kebijakan BPJS kesehatan. SVM berjalan baik dengan akurasi 94,28 persen. Saat mengevaluasi opini negatif, model SVM menunjukkan skor *presisi*, *recall*, dan F1 mencapai 97%. Sedangkan untuk opini positif akurasi mencapai 35%, mengingat 42%, dan skor F1 39%. Penelitian ini menyimpulkan bahwa, SVM memberikan peran penting dalam analisis opini tentang kebijakan BPJS kesehatan. Dampaknya adalah meningkatkan operasional kinerja BPJS Kesehatan berdasarkan penilaian respon masyarakat, dengan fokus pada peningkatan komunikasi dan pelayanan untuk mengurangi dampak negatif dan meningkatkan kepuasan pengguna layanan.

Penelitian ini akan menjadi landasan bagi pihak terkait, terutama petugas BPJS Kesehatan untuk mengambil keputusan yang responsif dan memenuhi harapan masyarakat. Konteks permasalahan penelitian ini menyoroti kompleksitas pemahaman dan permasalahan mendalam mengenai pandangan masyarakat terhadap transformasi kebijakan BPJS kesehatan.

Grisna dkk (2021) mengkaji Instagram sebagai salah satu media sosial yang sering digunakan oleh para manajer daerah. Media sosial digunakan untuk berkomunikasi dengan masyarakat, apalagi di masa pandemi Covid-19 yang membatasi aktivitas manusia dan berdampak pada banyak sektor, salah satunya adalah sektor UMKM yang berperan sebagai penopang perekonomian negara. Secara khusus, para pemimpin daerah menginformasikan kepada UKM tentang bidang-bidang yang didukung oleh negara melalui Instagram. Tujuan dari penelitian ini adalah untuk mengetahui rencana analisis opini Walikota Kediri menggunakan media sosial Instagram, mengidentifikasi bidang prioritas dukungan pemerintah terhadap usaha kecil dan menengah, dan memetakan respon masyarakat terhadap dukungan pemerintah. diumumkan oleh Abu Bakar sebagai walikota . Kediri Penelitian ini menggunakan metode *mixed text mining* dan data yang digunakan adalah data caption dan bahan komentar dari akun Instagram Abu Bakar pada tanggal 20 Agustus 2019 sampai dengan 20 November 2020 dengan total 17 *caption* dan 203 komentar. Hasil survei menunjukkan total 4 dari lima bidang dukungan pemerintah mendapat respon positif, bidang dukungan *real estate* mendapat skala prioritas tertinggi dan bidang dukungan modal usaha mendapat skala prioritas terendah.

Kajian Hidayat dkk (2021) mempertanyakan bagaimana halal menjadi tren global sehingga halal menjadi standar kualitas suatu produk khususnya makanan dan minuman bagi konsumen muslim dan non-muslim. Oleh karena itu, tujuan penelitian ini adalah memetakan kata kunci dan polaritas opini pengguna Twitter tentang produk halal di seluruh dunia. Data Twitter diekstraksi dan disiapkan menggunakan penambangan teks di Rstudio, yang kemudian dianalisis polaritasnya menggunakan pengklasifikasi Naïve Bayesian. Pengolahan data meliputi penghapusan duplikat tweet, penandaan, penghapusan kata akhir, dan pencarian akar kata. Desain busana Berdasarkan hasil diagram batang yang diperoleh, dilakukan analisis kata kunci. Kata kunci tersebut memberikan gambaran apa saja yang menjadi perhatian pengguna Twitter terkait status kehalalan suatu produk, khususnya makanan. Hasil survei menunjukkan kata kunci terpenting di Twitter untuk opini terkait produk halal adalah produk, makanan, minuman, jaminan dan lingkungan. Selain itu, polaritas pendapat menunjukkan mayoritas pengguna Twitter menilai positif terhadap tren produk halal yaitu 70%, 16% netral dan sisanya negatif. Hal ini memberikan peluang bisnis di bidang makanan halal di pasar global dan diharapkan dapat mendorong para pengusaha untuk melakukan sertifikasi halal terhadap produknya.

2.2. Metode *Extraction feature*

Literature review dalam penelitian ini menyoal tentang metode TF-IDF dan LSA yang digunakan dalam permasalahan penelitian terhadap teks analisis sentimen masyarakat, dengan demikian arah penelitian ini menjadikan landasan penelitian yang sudah dilakukan sebagai berikut.

Anggraini dkk (2021) melakukan penelitian yang bermula dari masalah pengurangan penyebaran COVID-19. Saat itu, pemerintah menggalakkan program vaksinasi untuk mencapai kekebalan kelompok. Karena kegagalan vaksinasi sebelumnya, sebagian besar masyarakat menolak keras vaksinasi, hal ini sangat menakutkan karena terjadi keributan di masyarakat. Untuk mengembalikan kepercayaan masyarakat, pemerintah berupaya menyebarkan informasi mengenai vaksinasi di media sosial (Instagram) dan hal ini akan menarik para peneliti untuk mengkaji lebih lanjut proses vaksinasi. Namun, ada beberapa hal dalam opini publik yang sulit ditemukan dan memerlukan analisis tekstual. Tujuan dari analisis teks adalah untuk menguji istilah pemeringkatan dan informasi lainnya dengan menggunakan metode analisis sentimen berbasis aturan. TF-IDF dan LSI/LSA adalah metode penambangan aturan yang digunakan untuk ekstraksi data. Hasil analisis penelitian ini mungkin mempengaruhi data lainnya. Misalnya analisis persepsi pengguna untuk melihat gambaran lebih besar mengenai isu atau topik pembahasan penting, serta titik temu untuk menyelesaikan permasalahan terkait vaksinasi COVID-19.

Tujuan dari kajian ilmiah yang dilakukan oleh Situmorang dkk (2023) adalah menganalisis sentimen tempat wisata di Jawa Barat di Twitter dengan menggunakan metode TF-IDF dan LSA. Berdasarkan data 5000 tweet, mayoritas masyarakat berpandangan positif terhadap destinasi wisata tersebut (42,24%), disusul opini netral (31,74%) dan opini negatif (26,02%). Algoritma Naive Bayes mempunyai presisi 0,332, presisi 0,658, dan recall 0,332.

Penelitian lain yang dilakukan Virgelus dkk (2023) menunjukkan bahwa nilai penilaian yang diperoleh dari hasil pencitraan adalah: prediksi negatif (sentimen) = 1,404 dan prediksi positif = 1,129, dengan akurasi penilaian sebesar 98,20%. Oleh karena itu, dapat disimpulkan bahwa berdasarkan cuitan tersebut, semakin banyak pengguna Twitter yang mengungkapkan perasaan negatif terhadap tragedi Kanjuruhan. Perasaan negatif yang besar seputar tragedi ini dapat dijadikan patokan atau patokan bagi pihak-pihak yang bertanggung jawab atas tindakan tersebut untuk mengetahui opini masyarakat terhadap tindakan yang dilakukan agar dapat mengambil hikmah di kemudian hari, serta ketika menyelenggarakan kegiatan serupa.

Studi yang dilakukan Miftahul Fikri dkk mengamati platform media sosial yang paling banyak digunakan, Twitter. Twitter dapat memberikan wawasan mengenai perubahan perilaku manusia terkait kontrasepsi, vaksinasi, dan perubahan adaptasi sosial dari waktu ke waktu selama periode endemi. Untuk mengklasifikasikan data yang diperoleh dari Twitter secara efektif, perlu mengembangkan model klasifikasi yang kuat dan menerapkan metode analisis data tingkat lanjut. Pendekatan yang digunakan dalam penelitian ini adalah penggunaan teknik *Latent Semantic Analysis (LSA)* yang bertujuan untuk mengungkap struktur semantik atau makna tersembunyi dalam kalimat.

Fokus utama penelitian ini adalah mengevaluasi pengaruh pengklasifikasian data Twitter dari fase pandemi hingga endemik terhadap kinerja analisis semantik laten yang diterapkan dengan metode TF-IDF. Hasil penelitian menunjukkan bahwa pada skenario klasifikasi, tingkat akurasi yang dicapai metode TF-IDF tanpa

pemeriksaan proses LSA adalah sebesar 22,00%. Dalam konteks yang sama, ketika LSA digunakan sebelum menerapkan metode TF-IDF, akurasi meningkat secara signifikan menjadi 24,00%. Dari hasil tersebut dapat disimpulkan bahwa penggunaan LSA memberikan efek positif dalam mencapai hasil yang lebih memuaskan. Selanjutnya, dalam upaya mengklasifikasikan data dengan menggunakan metode Naive Bayesian, penelitian dilakukan dengan dua skenario, yaitu menggunakan data berimbang dan tidak berimbang. Menggunakan data seimbang dengan metode TF-IDF, akurasi yang dicapai setelah langkah LSA adalah 16,00%. Sebaliknya, ketika data yang tidak seimbang digunakan dalam skenario yang sama, akurasi meningkat secara signifikan hingga 24,00%. Dari hasil tersebut dapat disimpulkan bahwa penggunaan informasi yang tidak seimbang berpengaruh positif terhadap kinerja klasifikasi.

2.3. Metode Klasterisasi

Berangkat dari penelitian terdahulu yang membahas tentang metode klasterisasi dengan *Self-Organizing Maps* untuk membongkar permasalahan analisis sentimen, maka peneliti menjadikan landasan penelitian untuk mengembangkan dan memperdalamnya diantaranya.

Kajian Danajaya dkk. (2022) yang diterbitkan pada jurnal bereputasi berfokus pada metode text mining yaitu *self-organizing map* yang belajar menambahkan komentar tanpa pengawasan, dan Kamus Sentimen Indonesia yang menganalisis setiap komentar sentimen. dari *cluster*. Vaksinasi dan persyaratan perjalanan bagi anak mendominasi komentar semua moda transportasi, sedangkan sentimen negatif mendominasi komentar kebijakan pariwisata dalam negeri.

Ghozali dkk. (2021) melakukan penelitian yang menunjukkan bahwa diperlukan upaya dalam bisnis untuk memaksimalkan keuntungan. Masalahnya adalah aturan penggabungan mengabaikan kumpulan data yang besar. Untuk mengatasi hal tersebut dibuat klasifikasi barang yang dibeli dan tidak dibeli. Algoritma *Self Organizing Map* (SOM) dan K-Medoids cocok untuk mengelompokkan kumpulan data besar. Penelitian ini menguji validitas dan kecepatan algoritma Self-Organizing Map (SOM) dan K-Medoids dengan *Frequent Pattern Growth* (FP-Growth).

Penelitian Overbeek (2019) mengembangkan pembelajaran mesin berbasis clustering menggunakan gambar digital. Ekstraksi orde kedua, yaitu *Gray Level Co-occurrence Matrix* menggunakan empat fitur Haralick, digunakan untuk ekstraksi fitur. Pada saat yang sama digunakan teknik kelompok *Self Organizing Map* (SOM). Hasil yang diperoleh pada penelitian ini memberikan rata-rata error sebesar 25% untuk setiap fitur Haralick, dan waktu pelatihan tercepat adalah 12 detik untuk fitur kontras.

Artikel Alkhalid dkk (2020) membahas tentang *self-organizing map* (SOM) sebagai metode untuk menentukan pengelompokan data pada domain kecanduan narkoba. Peta pengorganisasian mandiri (SOM) berisi lapisan yang mengatur dirinya sendiri ke dalam kelompok yang disebut cluster berdasarkan beberapa masukan nilai. Dengan demikian, algoritma pembelajaran SOM membentuk 4 cluster dengan karakteristik yang berbeda-beda pada setiap cluster, dan setiap cluster mempunyai warna yang berbeda. Selanjutnya implementasi sistem menggunakan ArcGIS 10.1. Hasil penelitian ini berupa gambaran pemetaan

wilayah kecanduan narkoba di wilayah Aceh Tenggara yang dapat memberikan informasi mengenai wilayah kecanduan narkoba di wilayah administratif Aceh Tenggara.

Penelitian selanjutnya dilakukan oleh Muh. Mujib (2019) dilakukan dengan tujuan untuk mengetahui secara pasti karakteristik emosional/alasan ulama memilih perguruan tinggi sebagai tujuan akademik. Metode yang digunakan dalam penelitian ini adalah *self-organizing map* (SOM) dengan algoritma Kohonen yang merupakan metode klasifikasi. SOM dengan learning rate yang digunakan 0:05, 0.25, 0:50, 0.75, 0.95 dan mengembalikan nilai bobot awal serta nilai titik tengah dan 500 iterasi dengan output terbentuk 3 *cluster*. Indeks Davies-Bouldin digunakan untuk memvalidasi hasil *clustering* SOM dengan hasil clustering terbaik yaitu DBI minimum (1.7802) dengan learning rate 0.95 dan *cluster* terbentuk tiga *cluster*. Untuk *cluster* pertama sebanyak enam anggota, *cluster-2* sebanyak 9 anggota dan *cluster* ke 3 sebanyak 5 anggota hasil *clustering* dengan prioritas utama terdapat pada *cluster* ke-2 dengan mean (7,434). Dimana karakteristik masing-masing anggota menjadi alasan emosional untuk memilih jurusan.

Santosa (2021) memetakan profil wilayah kecamatan menggunakan algoritma *Self-Organizing Maps* (SOM), karena SOM dapat memetakan data berdimensi tinggi pada bidang dua dimensi. Hasil pemetaan profil wilayah berupa data klaster subwilayah yang menggambarkan persamaan dan perbedaan karakteristik masing-masing subwilayah. Keunggulan dari pemetaan profil subwilayah Kabupaten Sragen adalah memberikan informasi persamaan dan perbedaan antar profil subwilayah berdasarkan hasil pemetaan profil subwilayah

Kabupaten Sragen. Sehingga profil subwilayah dapat diperlakukan sama dengan sifat serupa. Dengan demikian diharapkan akan membawa nilai tambah terhadap kebijakan yang diambil.

2.4. Kerangka Teori

Penelitian-penelitian terdahulu dapat dijadikan acuan bagi penulis untuk memperkaya teori dan juga dalam evaluasi penelitian yang sedang dilakukan. Berikut ini adalah beberapa penelitian terdahulu yang dapat dijadikan referensi.

Tabel 2.1 *Literature Review*

No	Judul	Penulis/Tahun	Hasil
1.	Desain sentimen analisis Opini Publik Terhadap Perubahan Kebijakan BPJS Kesehatan Dengan Pendekatan <i>Support Vector Machine</i> (Svm) Dalam Analisis Sentimen	Damayanti Damayanti dkk, (2018)	Hasil penelitian ini menunjukkan bahwa SVM mempunyai kontribusi penting dalam analisis opini terkait kebijakan BPJS kesehatan, terutama dalam menghadapi opini negatif. Dampak dari magang adalah meningkatkan operasional BPJS Kesehatan berdasarkan penilaian respon masyarakat, dengan fokus pada peningkatan pelayanan dan komunikasi untuk mengurangi dampak negatif dan meningkatkan kepuasan pengguna. Penelitian ini akan menjadi landasan bagi pengambil keputusan untuk mengambil keputusan yang responsif dan memenuhi harapan masyarakat. Latar belakang topik menyoroti kompleksitas dan pemahaman mendalam pandangan masyarakat terhadap perubahan kebijakan BPJS kesehatan.
2.	Desain sentimen analisis Komunikasi Walikota Kediri Melalui Instagram Terkait Dukungan Pemerintah Terhadap UMKM (Periode 20 Agustus 2019-20 November 2020)	Grisna Anggadwita, dkk, (2019)	Penelitian ini menggunakan metode <i>text mining</i> , dan data yang digunakan adalah data caption dan bahan komentar pada akun Instagram Abu Bakar pada tanggal 20 Agustus 2019 hingga 20 November 2020 dengan total 17 caption dan 203 komentar. Hasil survei menunjukkan bahwa total 4 dari lima bidang dukungan di negara ini mendapat tanggapan positif, bidang dukungan real estate mendapat tanggapan tertinggi dan bidang dukungan modal usaha mendapat skala prioritas terendah.
3.	Desain sentimen analisis Kata Kunci dan Polaritas Sentimen Pengguna Twitter Terhadap Kehalalan Produk	Hety Handayani Hidayat dkk, (2022)	Hasil penelitian menunjukkan bahwa kata kunci terpenting opini Twitter tentang produk halal adalah produk, makanan, minuman, jaminan dan lingkungan. Selain itu, polaritas pendapat menunjukkan mayoritas pengguna

Tabel 2.1 (Lanjutan) *Literature Review*

No	Judul	Penulis/Tahun	Hasil
			Twitter menilai positif terhadap tren produk halal yaitu 70%, 16% netral dan sisanya negatif. Hal ini memberikan peluang bisnis di bidang makanan halal di pasar global dan diharapkan dapat mendorong para pengusaha untuk melakukan sertifikasi halal terhadap produknya.
4.	Analisis Sentimen pada Komentar terhadap Kebijakan Perjalanan Domestik yang Dikelompokkan Menggunakan Metode <i>Self-Organizing Maps</i>	Adrianus Bagas Tanyo Dananjaya, Mohammad Isa Irawan, (2022)	Penelitian ini berfokus pada metode text mining yaitu <i>self-organizing map</i> , <i>unsupervised learning</i> untuk pengelompokan komentar dan Kamus Sentimen Indonesia untuk menganalisis komentar-komentar di setiap <i>cluster</i> .
5.	Analisa Pola Belanja Menggunakan Algoritma <i>Fp Growth</i> , <i>Self Organizing Map (Som)</i> Dan <i>K Medoids</i>	Muhammad Imam Ghozali Rif'an Zaenal Ehwan Wibowo Harry Sugiharto, (2021)	Penelitian ini menunjukkan bahwa diperlukan upaya dalam bisnis untuk memaksimalkan keuntungan. Masalahnya adalah aturan penggabungan mengabaikan kumpulan data yang besar. Untuk mengatasi hal tersebut dibuat klasifikasi barang yang dibeli dan tidak dibeli. Algoritma <i>Self Organizing Map (SOM)</i> dan <i>K-Medoids</i> cocok untuk mengelompokkan kumpulan data besar. Penelitian ini menguji validitas dan kecepatan algoritma <i>Self-Organizing Map (SOM)</i> dan <i>K-Medoids</i> dengan <i>Frequent Pattern-Growth (FP-Growth)</i> .
6.	Sortasi Tekstur Biji Jagung Sebagai Benih Tanam Menggunakan Self Organizing Map (Studi Kasus: Desa Bismarak Kabupaten Kupang Timur Provinsi Nusa Tenggara Timur)	Marlinda Vasty Overbeek, (2019)	Penelitian ini mengembangkan pembelajaran mesin berdasarkan pengelompokan gambar digital. Ekstraksi orde kedua, yaitu matriks kejadian bersama tingkat abu-abu menggunakan empat fitur Haralick, digunakan untuk ekstraksi fitur. Pada saat yang sama digunakan teknik kelompok <i>Self Organizing Map (SOM)</i> . Hasil yang diperoleh pada penelitian ini memberikan rata-rata error sebesar 25% untuk setiap fitur Haralick, dan waktu pelatihan tercepat adalah 12 detik untuk fitur kontras.
7.	Sistem Informasi Geografis Pemetaan Wilayah Penyalahgunaan Narkoba Menggunakan Metode SOM (<i>Self-Organizing Map</i>) Studi Kasus: Kabupaten Aceh Tenggara	Muhammad Waly Alkhalidi, (2020)	Pada penelitian ini membahas <i>self-organizing map (SOM)</i> sebagai metode untuk menentukan pengelompokan spasial data kecanduan narkoba. Peta pengorganisasian mandiri (<i>SOM</i>) berisi lapisan yang mengatur dirinya sendiri ke dalam kelompok yang disebut cluster berdasarkan beberapa masukan nilai. Dengan demikian, algoritma pembelajaran <i>Self Organizing Map</i> membentuk 4 <i>cluster</i> dengan karakteristik yang berbeda-beda pada setiap <i>cluster</i> , dan

Tabel 2.1 (Lanjutan) *Literature Review*

No	Judul	Penulis/Tahun	Hasil
			setiap cluster mempunyai warna yang berbeda. Selanjutnya implementasi sistem menggunakan ArcGIS 10.1. Hasil penelitian ini berupa gambaran pemetaan wilayah kecanduan narkoba di wilayah Aceh Tenggara yang dapat memberikan informasi mengenai wilayah kecanduan narkoba di wilayah administratif Aceh Tenggara.
8.	Pemetaan Preferensi Mahasiswa Baru Dalam Memilih Jurusan Menggunakan Artificial Neural Network (Ann) Dengan Algoritma Self Organizing Maps (Som)	Muh, Mujib, (2019)	Penelitian ini dilakukan untuk menyelidiki secara akurat karakteristik/alasan emosional peneliti dalam memilih pendidikan tinggi sebagai tujuan akademik. Metode yang digunakan dalam penelitian ini adalah self-organizing map dengan algoritma Kohonen yang merupakan metode klasifikasi. SOM learning rate yang digunakan adalah 0:05, 0.25, 0:50, 0.75, 0.95 dan nilai bobot awal dan nilai titik tengah dan 500 sehingga dihasilkan 3 <i>cluster</i> . Indeks Davies-Bouldin digunakan untuk memvalidasi hasil <i>clustering</i> SOM dengan hasil <i>clustering</i> terbaik yaitu DBI minimum (1.7802) dengan learning rate 0.95 dan terbentuk tiga <i>cluster</i> . <i>Cluster</i> pertama beranggotakan enam orang, <i>cluster</i> kedua beranggotakan sembilan orang, dan <i>cluster</i> ketiga beranggotakan lima orang. Hasil <i>cluster</i> pertama berada pada <i>cluster</i> kedua dengan mean karakteristik masing-masing <i>cluster</i> (7,434). Hasil tersebut adalah alasan emosional untuk memilih jurusan.
9.	Perancangan Sistem Pendukung Keputusan Pemetaan Profil Kecamatan Di Kabupaten Sragen Berdasarkan Indeks Pembangunan Manusia Menggunakan Algoritma Self Organizing Maps (Som)	Agung Dwi Santosa, (2021)	Pada penelitian ini pemetaan profil subangkaian dilakukan dengan menggunakan algoritma <i>Self-Organizing Maps</i> (SOM) karena SOM dapat memetakan data berdimensi tinggi pada bidang dua dimensi. Hasil pemetaan profil wilayah berupa data klaster subwilayah yang menggambarkan persamaan dan perbedaan karakteristik masing-masing subwilayah. Keunggulan dari pemetaan profil subwilayah Kabupaten Sragen adalah memberikan informasi persamaan dan perbedaan antar profil subwilayah berdasarkan hasil pemetaan profil subwilayah Kabupaten Sragen, sehingga profil subwilayah dapat diperlakukan sama dengan sifat serupa. Dengan demikian, diharapkan dapat membawa nilai tambah terhadap kebijakan yang disusun.

Tabel 2.1 (Lanjutan) *Literature Review*

No	Judul	Penulis/Tahun	Hasil
10	Text Mining - Analisis Teks Terkait Isu Vaksinasi COVID-19 (Text Mining - Analisis Teks Terkait Masalah Vaksinasi COVID-19)	Novita Anggraini, 2021	Hasil analisis penelitian ini mungkin mempengaruhi data lainnya. Misalnya analisis persepsi pengguna untuk melihat gambaran lebih besar mengenai topik-topik penting atau topik pembahasan, serta titik temu untuk mengatasi permasalahan terkait vaksinasi COVID-19.
11	Analisis Sentimen Destinasi Wisata Di Jawa Barat Pada Twitter Menggunakan Algoritma Naive Bayes Classifier	Riwanti Situmorang, 2023	Tujuan dari penelitian ini adalah menganalisis sentimen destinasi wisata di Jawa Barat di Twitter dengan menggunakan metode SOM. Berdasarkan data 5000 tweet, mayoritas masyarakat berpandangan positif terhadap destinasi wisata tersebut (42,24%), disusul opini netral (31,74%) dan opini negatif (26,02%). Algoritma Naive Bayes mempunyai presisi sebesar 0,332, presisi sebesar 0,658, dan recall sebesar 0,332.
12	Analisis Sentimen Media Sosial Twitter Terhadap Tragedi Kanjuruhan	Virgelius Hendrawan Taralandu, 2021	Dari penelitian tersebut, dapat disimpulkan bahwa semakin banyak pengguna Twitter yang berpandangan negatif terhadap tragedi Kanjuruhan. Banyaknya emosi negatif yang terkait dengan tragedi ini dapat dijadikan patokan atau patokan bagi pihak-pihak yang bertanggung jawab atas tindakan tersebut untuk mengetahui pendapat masyarakat terhadap tindakan yang dilakukan sehingga dapat menjadi pembelajaran untuk pengaturan ke depan dengan aktivitas serupa.
13	Klasifikasi Data Twitter pada Masa Transisi Pandemi menuju Endemi menggunakan Latent Semantic Analysis (LSA)	Miftahul Fiqri, 2023	Hasil penelitian menunjukkan bahwa dalam skenario klasifikasi menggunakan metode TF-IDF tanpa melalui proses LSA, tingkat akurasi yang berhasil dicapai adalah 22,00%. Akan tetapi, dalam konteks yang sama, saat LSA diaplikasikan sebelum menerapkan metode TF-IDF, terjadi peningkatan signifikan dalam tingkat akurasi menjadi 24,00%. Dari hasil ini, dapat ditarik kesimpulan bahwa penggunaan LSA memberikan sumbangan positif dalam mencapai hasil yang lebih memuaskan. Selanjutnya, dalam upaya mengklasifikasikan data menggunakan metode TF-IDF, penelitian dilakukan dalam dua skenario, yaitu menggunakan data yang seimbang dan tidak seimbang. Ketika data seimbang digunakan bersamaan dengan penerapan metode TF-IDF

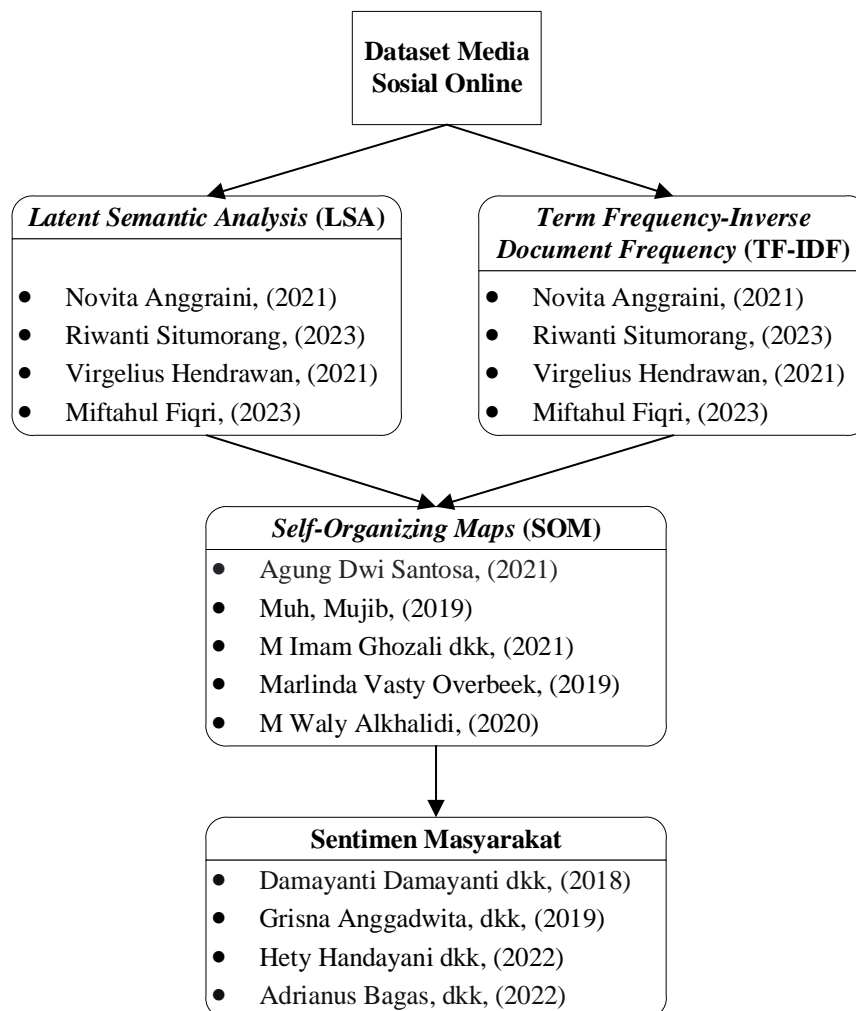
Tabel 2.1 (Lanjutan) *Literature Review*

No	Judul	Penulis/Tahun	Hasil
			setelah tahap LSA, tingkat akurasi yang tercapai adalah 16,00%. Sebaliknya, ketika data tidak seimbang digunakan dalam skenario yang sama, terjadi peningkatan yang signifikan dalam tingkat akurasi, mencapai angka 24,00%. Dari hasil ini, dapat disimpulkan bahwa penggunaan data yang tidak seimbang memiliki dampak positif pada performa klasifikasi.

Terdapat beberapa perbedaan berdasarkan penelitian sebelumnya yang disajikan pada Tabel 2.1. Perbedaan penelitiannya terletak pada materi yang digunakan yaitu dalam hal penelitian yang dilakukan diambil objeknya pada pemilu presiden tahun 2024 yang kemudian dipertimbangkan dengan mengintegrasikan dua metode yaitu *Term Frequency-Inverse Document Frequency* (TF-IDF) dan Analisis Semantik Laten (LSA). Keduanya memiliki tujuan yang sama yaitu untuk mengungkap struktur makna yang tersembunyi dalam kalimat. Hasil dari kedua metode tersebut yang kemudian dianalisis dengan algoritma *Self-Organizing Maps* (SOM). Dari beberapa *literature review* yang telah dipaparkan menunjukkan bahwa tidak terdapat kesamaan dalam hal objek material maupun objek formal penelitian. Hal inilah yang menjadi ketertarikan peneliti untuk membongkar lebih dalam terkait dengan sentimen masyarakat terhadap Pilpres 2024. Dengan demikian penelitian yang akan dilakukan ini berfokus pada Desain Sentimen Analisis Dengan *Self-Organizing Map* Untuk Pemilihan Presiden 2024.

Pada gambar 2.1 di bawah ini menggambarkan alur proses analisis dataset media sosial online yang melalui beberapa tahap utama. Proses dimulai dengan pemilahan data menggunakan dua metode analisis teks: Latent Semantic Analysis

(LSA) dan Term Frequency-Inverse Document Frequency (TF-IDF). Kedua metode ini digunakan oleh peneliti untuk mengidentifikasi dan mengekstrak informasi yang relevan dari data teks. Pada tahap LSA, penelitian dilakukan oleh Novita Anggraini (2021), Riwanti Situmorang (2023), Virgelius Hendrawan (2023), dan Mifthaul Fiqri (2023). Metode TF-IDF juga melibatkan penelitian dari individu yang sama pada tahun yang sama, menunjukkan penggunaan kedua teknik untuk memastikan analisis yang komprehensif.



Gambar 2.1. Kerangka Teori

Selanjutnya, hasil dari tahap pertama dianalisis lebih lanjut menggunakan Self-Organizing Maps (SOM), sebuah metode pembelajaran mesin yang digunakan

untuk pengelompokan data. Penelitian pada tahap ini dilakukan oleh Agung Dwi Santosa (2021), Muh. Mujib (2019), M. Imam Ghozali (2018), Marlinda Vasty Overbeek (2019), dan M. Waly Alkhaidi (2020). Data yang telah dikelompokkan kemudian digunakan untuk memahami Sentimen Masyarakat. Analisis sentimen ini melibatkan penelitian dari Damayanti Damayanti dkk. (2018), Grisan Anggaddwi dkk. (2019), Hety Handayani dkk. (2022), dan Adrianus Bagas dkk. (2023). Alur proses ini menunjukkan pendekatan sistematis dari pemrosesan data teks hingga analisis sentimen untuk memahami persepsi publik terhadap topik tertentu di media sosial.

BAB III

METODE PENELITIAN

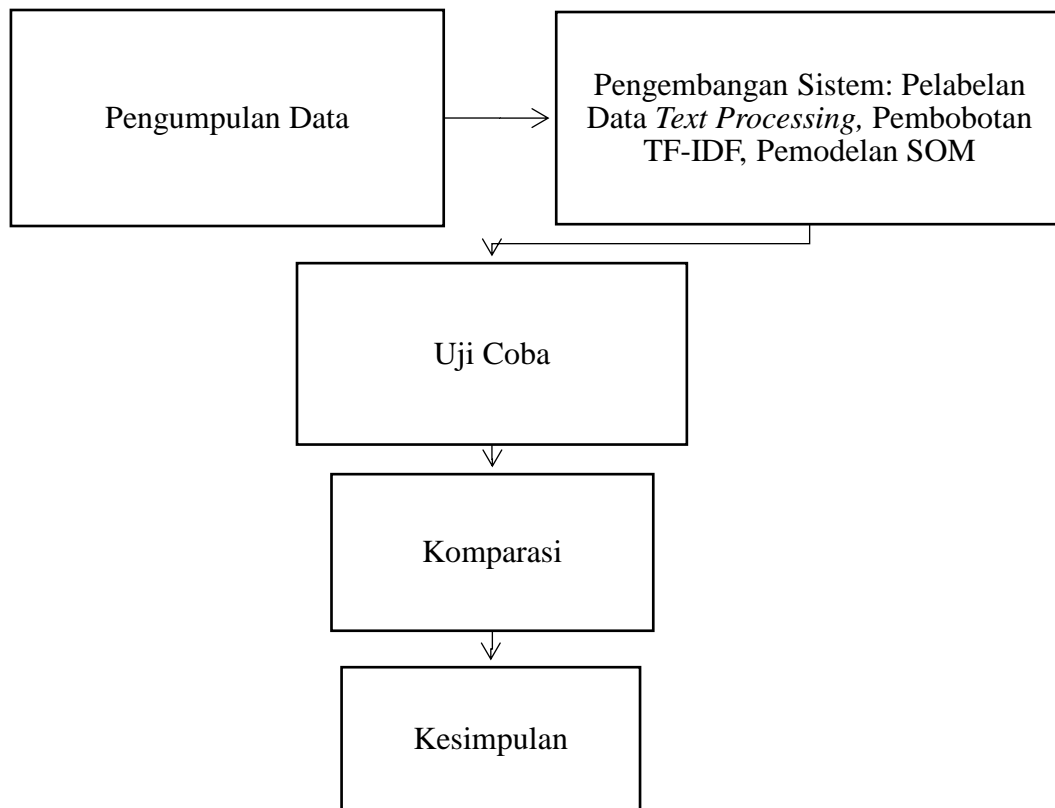
3.1. Lokasi dan Waktu Penelitian

Dalam penelitian ini lokasi pengambilan datanya melalui media sosial online Twitter yang memiliki otorisasi API (*Application Programming Interface*). Sedangkan penelitian ini dimulai dari 15 Februari 2024 sampai 24 April 2024, waktu penelitian dilakukan pasca Pemilihan Umum 2024 dalam rangka menggali data terkait sentimen masyarakat terhadap Pilpres 2024.

3.2. Alur Penelitian

Penelitian diawali dengan pengumpulan data dengan mengindeks data dari jejaring sosial Twitter untuk mendapatkan dataset berbasis kata kunci, *pilpres2024*, yang berupa cuitan dari pengguna atau media massa yang menyoal tentang sentimen dan tanggapan terhadap pilpres 2024. Dataset yang terkumpul diberi label terlebih dahulu agar dapat mengetahui polaritas tweet tersebut, apakah tweet bersifat positif, netral, atau negatif. Proses selanjutnya yang dilakukan adalah tahap *pre-processing* yang terdiri dari *cleaning*, *case folding*, *tokenizing*, *normalization*, *filtering*, dan *stemming*. Data *clean* tweet selanjutnya digunakan pada proses ekstraksi fitur TF-IDF untuk mendapatkan bobot pada setiap kata. Tahapan berikutnya akan berlatih membangun model analisis sentimen dengan algoritma SOM. Model yang dilatih berhasil dievaluasi untuk mendapatkan model terbaik.

Jalannya tahapan penelitian dapat dilihat pada Gambar 3.1.



Gambar 3.1. Desain Sistem Penelitian yang diusulkan

3.2.1. Pengumpulan Data

Menilik gambar desain sistem penelitian diatas merepresentasikan alur penelitian yang diawali dengan pengumpul-an data twitter. Data Tweet diperoleh melalui proses pengindeksan menggunakan *Application Programming Interface* (API) Twitter. Pengindeksan data Tweet dilakukan dengan menggunakan tweepy library, yaitu perpustakaan yang dikembangkan dengan bahasa pemrograman Python. Hasil dari proses tersebut menghasilkan data dengan format JSON (*Javascript Object Notation*), yang selanjutnya diubah menjadi CSV dengan pengolahan menggunakan data frame library Pandas (Mitchell, 2018:15). Berikut hasil pengumpulan data Twitter.

Tabel 3.1 hasil pengumpulan data Twitter

no	cre- ated_at	favorite_count	full_text	reply_count	re- tweet_count	username
1	Fri Apr 05 18:10:33 +0000 2024	0	@tempodotco Mungkin ini jk capres tdk peduli norma2 Etik. Rakyat fokus sidang sengketa pilpres 2024 di MK ramai demo pemilu curang....ybs spt bodoh buta tuli tdk mengubris keadaan.	0	0	03_malikka
2	Fri Apr 05 15:21:58 +0000 2024	62	@tempodotco Dana Operasional Presiden lebih banyak utk bansos yg tujuannya menaikkan popularitas jokowi (?) sehingga bisa cawe-cawe selama pilpres 2024 artinya ini bisa melanggar undang-undang / etika/ asas pemilu jurdil dugaan pelanggaran ini harus digul	2	3	sayno_KKN
3	Fri Apr 05 16:59:56 +0000 2024	0	Jika Perpres penetapan APBN sejak 29 November untuk pencairan dana 2024 oleh Menkeu?. Jelas itu sudah masuk tahapan Pilpres. Kampanye. Kenapa dadakan yg tersistem mulai tgl 29 sudah bisa disebarkan Bansos?. Termasuk Presiden bagi2 sendiri setiap kunjungan.	1	0	Mas_Say_Fh
4	Fri Apr 05 16:27:24 +0000 2024	2	Mmg menkeu Sri mulyani pakar dlm mengamankan Presiden Jokowi terkait Aliran Dana bansos yg dilakukan Presiden Jokowi dlm menghadapi elnino bertepatan masa kampanye Pemilu Pilpres 2024 Berarti AMAN Oresuden Jpkpwi Good job menkeu Sri mulyani https://t.co/4INN3kj526	0	1	ENS53311139

Tabel Lanjutan 3.1 hasil pengumpulan data Twitter

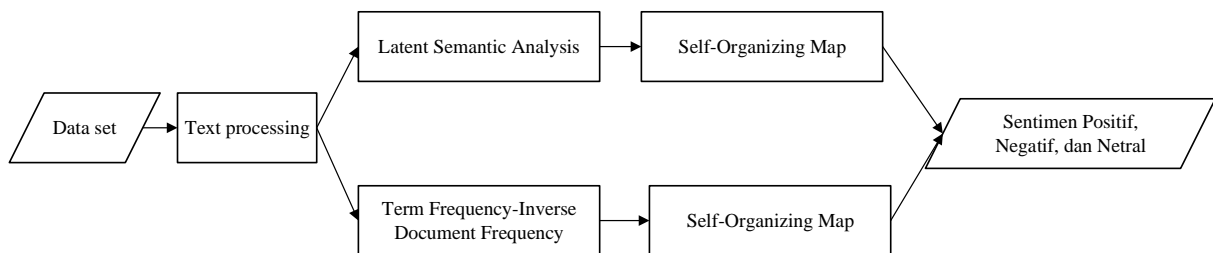
no	cre- ated_at	favor- ite_count	full_text	re- ply_count	re- tweet_count	username
5	Fri Apr 05 17:13:54 +0000 2024	2	@fajarnugros 2020 s.d Februari angkanya lebih tinggi dibanding yg 2024. Padahal 1. Covid mulai rame di bulan maret dan kabinet blm ada aksi atau instruksi yg ex- traordinary 2. Pilpresnya dah lewat yakni 2019 Yg pu- nya ide munculkan isu bansos pada pilpres i	0	0	akoromu936
.....
6412	Fri Apr 05 10:30:00 +0000 2024	0	Gugatan dari 01 03 terkait sengketa Pilpres 2024 di Mahkamah Konstitusi sejak awal memang banyak menimbulkan masalah. @prabowo @Gerindra @MataNajwa #GugatanNgawur	0	0	oktavianim5929 3
6413	Fri Apr 05 15:15:11 +0000 2024	0	'CAWE CAWE' MERUPAKAN BUKTI ADANYA NIAT ITIKAD JAHAT RENCANA DAN AKSI TSM DARI SI TUKANG NGIBUL NEOPKI KOKOWI DKK BANDITNYA ATAS KECURANGAN DAN KEJAHATAN PEMILU PILPRES2024 DEMI PEMENANGAN PASLON 02 DALAM SATU PUTARAN	0	0	penjagadapur

Tabel Lanjutan 3.1 hasil pengumpulan data Twitter

no	created_at	favorite_count	full_text	reply_count	retweet_count	username
			#JOKOWIOPERATOR DANDALANGKE CURANGANDANK EJAHATANPEMILU 2024.			
6414	Fri Apr 05 10:30:00 +0000 2024	0	yakni dua propinsi DI Aceh dan Sumatera Barat dua titik di mana bansos diberikan untuk rakyat oleh pemerintah demi meringankan beban akibat pandemi tapi kenyataannya pemenang Pilpres 2024 di dua propinsi itu adalah paslon 01. @prabowo @Gerindra @MataNajwa #GugatanNgawur	0	0	sintanura60
6415	Fri Apr 05 13:31:39 +0000 2024	2	Kalau silent majority (rakyat bawah) tdk masa bodoh terhadap politik mk tentu di Pilpres 2024 mereka gak akan pilih capres yg melanggar etika buat curang merusak demokrasi terancam kembali otoriter Jadi rakyat bawah juga bersalah shg LAYAK menanggung keburukan di hari ini	2	0	MbulGemmy
6416	Fri Apr 05 11:58:40 +0000 2024	1	@AntonKu08408962 itu sudah kuncinya. dan sejatinya pilpres sejak th 2009 2014 2019 2024 adalah sama. People Behaviour nya sama. Rakyat lapar 60%.	1	0	FajriGrahaArtha

3.2.2. Pengembangan Sistem

Pengembangan sistem dalam konteks pengolahan kata melibatkan beberapa langkah penting, antara lain pelabelan data, pembobotan *Term Frequency-Inverse Document Frequency* (TF-IDF), dan pemodelan menerapkan algoritma *Self-Organizing Maps* (SOM).



Gambar. 3.2. Pengembangan Sistem

Berikut adalah penjelasan lebih mendetail mengenai setiap langkah tersebut:

3.2.2.1. *Text Preprocessing*

Proses selanjutnya adalah *Text Preprocessing* dengan berbagai tahapan yaitu.

a. *Cleaning*

Pembersihan menghapus data teks untuk diproses. Data teks yang telah dikumpulkan masih berisi banyak emoji dan @namapengguna_pengguna, untuk itu data teks yang mengandung emoji dan @username_pengguna akan dibersihkan.

Tabel 3.2. Hasil Cleaning

no	full_text	tweet_cleaning
1	@tempodotco Mungkin ini jk capres tdk peduli norma2 Etik. Rakyat fokus sidang sengketa pilpres 2024 di MK ramai demo pemilu curang....ybs spt bodoh buta tuli tdk mengubris keadaan.	Mungkin ini jk capres tdk peduli norma Etik Rakyat fokus sidang sengketa pilpres di MK ramai demo pemilu curangybs spt bodoh buta tuli tdk mengubris keadaan

Tabel Lanjutan 3.2. Hasil Cleaning

no	full_text	tweet_cleaning
2	@tempodotco Dana Operasional Presiden lebih banyak utk bansos yg tujuannya menaikkan popularitas jokowi (?) sehingga bisa cawe-cawe selama pilpres 2024 artinya ini bisa melanggar undang- undang / etika/ asas pemilu jurdil dugaan pelanggaran ini harus digul	Dana Operasional Presiden lebih banyak utk bansos yg tujuannya menaikkan popularitas jokowi sehingga bisa cawecawe selama pilpres artinya ini bisa melanggar undangundang etika asas pemilu jurdil dugaan pelanggaran ini harus digulirkan hak angket
3	Jika Perpres penetapan APBN sejak 29 November untuk pencairan dana 2024 oleh Menkeu?. Jelas itu sudah masuk tahapan Pilpres. Kampanye. Kenapa dadakan yg tersistem mulai tgl 29 sudah bisa disebarkan Bansos?. Termasuk Presiden bagi2 sendiri setiap kunjungan.	Jika Perpres penetapan APBN sejak November untuk pencairan dana oleh Menkeu Jelas itu sudah masuk tahapan Pilpres Kampanye Kenapa dadakan yg tersistem mulai tgl sudah bisa disebarkan Bansos Termasuk Presiden bagi sendiri setiap kunjungan
4	Mmg menkeu Sri mulyani pakar dlm mengamankan Presiden Jokowi terkait Aliran Dana bansos yg dilakukan Presiden Jokowi dlm menghadapi elnino bertepatan masa kampanye Pemilu Pilpres 2024 Berarti AMAN Oresuden Jpkpwi Good job menkeu Sri mulyani https://t.co/4INN3kj526	Mmg menkeu Sri mulyani pakar dlm mengamankan Presiden Jokowi terkait Aliran Dana bansos yg dilakukan Presiden Jokowi dlm menghadapi elnino bertepatan masa kampanye Pemilu Pilpres Berarti AMAN Oresuden Jpkpwi Good job menkeu Sri mulyani
5	@fajarnugros 2020 s.d Februari angkanya lebih tinggi dibanding yg 2024. Padahal 1. Covid mulai rame di bulan maret dan kabinet blm ada aksi atau instruksi yg extraordinary 2. Pilpresnya dah lewat yakni 2019 Yg punya ide munculkan isu bansos pada pilpres i	sd Februari angkanya lebih tinggi dibanding yg Padahal Covid mulai rame di bulan maret dan kabinet blm ada aksi atau instruksi yg extraordinary Pilpresnya dah lewat yakni Yg punya ide munculkan isu bansos pada pilpres ini unik juga
.....
6412	Gugatan dari 01 03 terkait sengketa Pilpres 2024 di Mahkamah Konstitusi sejak awal memang banyak menimbulkan masalah. @prabowo @Gerindra @MataNajwa #GugatanNgawur	Gugatan dari terkait sengketa Pilpres di Mahkamah Konstitusi sejak awal memang banyak menimbulkan masalah
6413	'CAWE CAWE' MERUPAKAN BUKTI ADANYA NIAT ITIKAD JAHAT RENCANA DAN AKSI TSM DARI SI TUKANG NGIBUL NEOPKI KOKOWI DKK BANDITNYA ATAS KECURANGAN DAN KEJAHATAN PEMILU PILPRES 2024 DEMI PEMENANGAN PASLON 02 DALAM SATU PUTARAN #JOKOWIOPERATORDANDALANGKECURANGANDANKEJAHATANPEMILU2024.	CAWE CAWE MERUPAKAN BUKTI ADANYA NIAT ITIKAD JAHAT RENCANA DAN AKSI TSM DARI SI TUKANG NGIBUL NEOPKI KOKOWI DKK BANDITNYA ATAS KECURANGAN DAN KEJAHATAN PEMILU PILPRES DEMI PEMENANGAN PASLON DALAM SATU PUTARAN
6414	yakni dua propinsi DI Aceh dan Sumatera Barat dua titik di mana bansos diberikan untuk rakyat oleh pemerintah demi meringankan beban akibat pandemi tapi kenyataannya pemenang	yakni dua propinsi DI Aceh dan Sumatera Barat dua titik di mana bansos diberikan untuk rakyat oleh pemerintah demi meringankan beban

Tabel Lanjutan 3.2. Hasil Cleaning

no	full_text	tweet_cleaning
	Pilpres 2024 di dua propinsi itu adalah paslon 01. @prabowo @Gerindra @MataNajwa #GugatanNgawur	akibat pandemi tapi kenyataannya pemenang Pilpres di dua propinsi itu adalah paslon
6415	Kalau silent majority (rakyat bawah) tdk masa bodoh terhadap politik mk tentu di Pilpres 2024 mereka gak akan pilih capres yg melanggar etika buat curang merusak demokrasi terancam kembali otoriter Jadi rakyat bawah juga bersalah shg LAYAK menanggung keburukan di hari ini	Kalau silent majority rakyat bawah tdk masa bodoh terhadap politik mk tentu di Pilpres mereka gak akan pilih capres yg melanggar etika buat curang merusak demokrasi terancam kembali otoriter Jadi rakyat bawah juga bersalah shg LAYAK menanggung keburukan di hari ini
6416	@AntonKu08408962 itu sudah kuncinya. dan sejatinya pilpres sejak th 2009 2014 2019 2024 adalah sama. Peopple Behaviour nya sama. Rakyat lapar 60%.	itu sudah kuncinya dan sejatinya pilpres sejak th adalah sama Peopple Behaviour nya sama Rakyat lapar

b. Case Folding

Case folding adalah langkah untuk membuat kalimat menjadi huruf kecil saja, sehingga memudahkan proses selanjutnya. Hal ini akan memudahkan analisis teks di kemudian hari, karena menghilangkan perbedaan antara huruf besar dan huruf kecil.

Tabel 3.3. Case Folding

no	tweet_cleaning	tweet_case_folding
1	Mungkin ini jk capres tdk perduli norma Etik Rakyat fokus sidang sengketa pilpres di MK ramai demo pemilu	mungkin ini jk capres tdk perduli norma etik rakyat fokus sidang sengketa pilpres di mk
	curangybs spt bodoh buta tuli tdk mengubris keadaan	ramai demo pemilu curangybs spt bodoh buta tuli tdk mengubris keadaan
2	Dana Operasional Presiden lebih banyak utk bansos yg tujuannya menaikkan popularitas jokowi sehingga bisa cawecawe selama pilpres artinya ini bisa melanggar undangundang etika asas pemilu jurdil dugaan pelanggaran ini harus digulirkan hak angket	dana operasional presiden lebih banyak utk bansos yg tujuannya menaikkan popularitas jokowi sehingga bisa cawecawe selama pilpres artinya ini bisa melanggar undangundang etika asas pemilu jurdil dugaan pelanggaran ini harus digulirkan hak angket
3	Jika Perpres penetapan APBN sejak November untuk pencairan dana oleh Menkeu Jelas itu sudah masuk tahapan Pilpres Kampanye Kenapa dadakan yg tersistem mulai tgl sudah bisa disebarakan Bansos Termasuk Presiden bagi sendiri setiap kunjungan	jika perpres penetapan apbn sejak november untuk pencairan dana oleh menkeu jelas itu sudah masuk tahapan pilpres kampanye kenapa dadakan yg tersistem mulai tgl sudah bisa disebarakan bansos termasuk

Tabel Lanjutan 3.3. Case Folding

no	tweet_cleaning	tweet_case_folding
		presiden bagi sendiri setiap kunjungan
4	Mmg menkeu Sri mulyani pakar dlm mengamankan Presiden Jokowi terkait Aliran Dana bansos yg dilakukan Presiden Jokowi dlm menghadapi elnino bertepatan masa kampanye Pemilu Pilpres Berarti AMAN Oresuden Jpkpwi Good job menkeu Sri mulyani	mmg menkeu sri mulyani pakar dlm mengamankan presiden jokowi terkait aliran dana bansos yg dilakukan presiden jokowi dlm menghadapi elnino bertepatan masa kampanye pemilu pilpres berarti aman oresuden jpkpwi good job menkeu sri mulyani
5	sd Februari angkanya lebih tinggi dibanding yg Padahal Covid mulai rame di bulan maret dan kabinet blm ada aksi atau instruksi yg extraordinary Pilpresnya dah lewat yakni Yg punya ide munculkan isu bansos pada pilpres ini unik juga	sd februari angkanya lebih tinggi dibanding yg padahal covid mulai rame di bulan maret dan kabinet blm ada aksi atau instruksi yg extraordinary pilpresnya dah lewat yakni yg punya ide munculkan isu bansos pada pilpres ini unik juga
.....
6412	Gugatan dari terkait sengketa Pilpres di Mahkamah Konstitusi sejak awal memang banyak menimbulkan masalah	gugatan dari terkait sengketa pilpres di mahkamah konstitusi sejak awal memang banyak menimbulkan masalah
6413	CAWE CAWE MERUPAKAN BUKTI ADANYA NIAT ITIKAD JAHAT RENCANA DAN AKSI TSM DARI SI TUKANG NGIBUL NEOPKI KOKOWI DKK BANDITNYA ATAS KECURANGAN DAN KEJAHATAN PEMILU PILPRES DEMI PEMENANGAN PASLON DALAM SATU PUTARAN	cawe cawe merupakan bukti adanya niat itikad jahat rencana dan aksi tsm dari si tukang ngibul neopki kokowi dkk banditnya atas kecurangan dan kejahatan pemilu pilpres demi kemenangan paslon dalam satu putaran
6414	yakni dua propinsi DI Aceh dan Sumatera Barat dua titik di mana bansos diberikan untuk rakyat oleh pemerintah demi meringankan beban akibat pandemi tapi kenyataannya pemenang Pilpres di dua propinsi itu adalah paslon	yakni dua propinsi di aceh dan sumatera barat dua titik di mana bansos diberikan untuk rakyat oleh pemerintah demi meringankan beban akibat pandemi tapi kenyataannya pemenang pilpres di dua propinsi itu adalah paslon
6415	Kalau silent majority rakyat bawah tdk masa bodoh terhadap politik mk tentu di Pilpres mereka gak akan pilih capres yg melanggar etika buat curang merusak demokrasi terancam kembali otoriter Jadi rakyat bawah juga bersalah shg LAYAK menanggung keburukan di hari ini	kalau silent majority rakyat bawah tdk masa bodoh terhadap politik mk tentu di pilpres mereka gak akan pilih capres yg melanggar etika buat curang merusak demokrasi terancam kembali otoriter jadi rakyat bawah juga bersalah shg layak menanggung keburukan di hari ini
6416	itu sudah kuncinya dan sejatinya pilpres sejak th adalah sama Peolple Behaviour nya sama Rakyat lapar	itu sudah kuncinya dan sejatinya pilpres sejak th adalah sama peolple behaviour nya sama rakyat lapar

c. Tokenizing

Tokenisasi adalah proses mengurai kalimat menjadi array yang berisi kata-kata. Proses tokenisasi bersamaan dengan penghapusan tanda baca, angka, simbol,

dan spasi yang tidak digunakan pada proses nantinya. Tokenisasi bertujuan untuk memecah kalimat menjadi token atau kata-kata.

Tabel 3.4. Hasil Tokezing

no	tweet_case_folding	tokenisasi
1	mungkin ini jk capres tdk perduli norma etik rakyat fokus sidang sengketa pilpres di mk ramai demo pemilu curangybs spt bodoh buta tuli tdk mengubris keadaan	['mungkin', 'ini', 'jk', 'capres', 'tdk', 'perduli', 'norma', 'etik', 'rakyat', 'fokus', 'sidang', 'sengketa', 'pilpres', 'di', 'mk', 'ramai', 'demo', 'pemilu', 'curangybs', 'spt', 'bodoh', 'buta', 'tuli', 'tdk', 'mengubris', 'keadaan']
2	dana operasional presiden lebih banyak utk bansos yg tujuannya menaikkan popularitas jokowi sehingga bisa cawecawe selama pilpres artinya ini bisa melanggar undangundang etika asas pemilu jurdil dugaan pelanggaran ini harus digulirkan hak angket	['dana', 'operasional', 'presiden', 'lebih', 'banyak', 'utk', 'bansos', 'yg', 'tujuannya', 'menaikkan', 'popularitas', 'jokowi', 'sehingga', 'bisa', 'cawecawe', 'selama', 'pilpres', 'artinya', 'ini', 'bisa', 'melanggar', 'undangundang', 'etika', 'asas', 'pemilu', 'jurdil', 'dugaan', 'pelanggaran', 'ini', 'harus', 'digulirkan', 'hak', 'angket']
3	jika perpres penetapan apbn sejak november untuk pencairan dana oleh menkeu jelas itu sudah masuk tahapan pilpres kampanye kenapa dadakan yg tersistem mulai tgl sudah bisa disebarkan bansos termasuk presiden bagi sendiri setiap kunjungan	['jika', 'perpres', 'penetapan', 'apbn', 'sejak', 'november', 'untuk', 'pencairan', 'dana', 'oleh', 'menkeu', 'jelas', 'itu', 'sudah', 'masuk', 'tahapan', 'pilpres', 'kampanye', 'kenapa', 'dadakan', 'yg', 'tersistem', 'mulai', 'tgl', 'sudah', 'bisa', 'disebarkan', 'bansos', 'termasuk', 'presiden', 'bagi', 'sendiri', 'setiap', 'kunjungan']
4	mmg menkeu sri mulyani pakar dlm mengamankan presiden jokowi terkait aliran dana bansos yg dilakukan presiden jokowi dlm menghadapi elnino bertepatan masa kampanye pemilu pilpres berarti aman oresuden jpkpwi good job menkeu sri mulyani	['mmg', 'menkeu', 'sri', 'mulyani', 'pakar', 'dlm', 'mengamankan', 'presiden', 'jokowi', 'terkait', 'aliran', 'dana', 'bansos', 'yg', 'dilakukan', 'presiden', 'jokowi', 'dlm', 'menghadapi', 'elnino', 'bertepatan', 'masa', 'kampanye', 'pemilu', 'pilpres', 'berarti', 'aman', 'oresuden', 'jpkpwi', 'good', 'job', 'menkeu', 'sri', 'mulyani']
5	sd februari angkanya lebih tinggi dibanding yg padahal covid mulai rame di bulan maret dan kabinet blm ada aksi atau instruksi yg extraordinary pilpresnya dah lewat yakni yg punya ide munculkan isu bansos pada pilpres ini unik juga	['sd', 'februari', 'angkanya', 'lebih', 'tinggi', 'dibanding', 'yg', 'padahal', 'covid', 'mulai', 'rame', 'di', 'bulan', 'maret', 'dan', 'kabinet', 'blm', 'ada', 'aksi', 'atau', 'instruksi', 'yg', 'extraordinary', 'pilpresnya', 'dah', 'lewat', 'yakni', 'yg', 'punya', 'ide', 'munculkan', 'isu', 'bansos', 'pada', 'pilpres', 'ini', 'unik', 'juga']
.....
6412	gugatan dari terkait sengketa pilpres di mahkamah konstitusi sejak awal memang banyak menimbulkan masalah	['gugatan', 'dari', 'terkait', 'sengketa', 'pilpres', 'di', 'mahkamah', 'konstitusi', 'sejak', 'awal', 'memang', 'banyak', 'menimbulkan', 'masalah']
6413	cawe cawe merupakan bukti adanya niat itikad jahat rencana dan aksi tsm dari si tukang ngibul neopki kokowi dkk banditnya atas kecurangan dan kejahatan pemilu pilpres demi kemenangan paslon dalam satu putaran	['cawe', 'cawe', 'merupakan', 'bukti', 'adanya', 'niat', 'itikad', 'jahat', 'rencana', 'dan', 'aksi', 'tsm', 'dari', 'si', 'tukang', 'ngibul', 'neopki', 'kokowi', 'dkk', 'banditnya', 'atas', 'kecurangan', 'dan', 'kejahatan', 'pemilu', 'pilpres', 'demi']

Tabel Lanjutan 3.4. Hasil Tokezing

no	tweet_case_folding	Tokenisasi
		'pemenangan', 'paslon', 'dalam', 'satu', 'putaran']
6414	yakni dua propinsi di aceh dan sumatera barat dua titik di mana bansos diberikan untuk rakyat oleh pemerintah demi meringankan beban akibat pandemi tapi kenyataannya pemenang pilpres di dua propinsi itu adalah paslon	['yakni', 'dua', 'propinsi', 'di', 'aceh', 'dan', 'sumatera', 'barat', 'dua', 'titik', 'di', 'mana', 'bansos', 'diberikan', 'untuk', 'rakyat', 'oleh', 'pemerintah', 'demi', 'meringankan', 'beban', 'akibat', 'pandemi', 'tapi', 'kenyataannya', 'pemenang', 'pilpres', 'di', 'dua', 'propinsi', 'itu', 'adalah', 'paslon']
6415	kalau silent majority rakyat bawah tdk masa bodoh terhadap politik mk tentu di pilpres mereka gak akan pilih capres yg melanggar etika buat curang merusak demokrasi terancam kembali otoriter jadi rakyat bawah juga bersalah shg layak menanggung keburukan di hari ini	['kalau', 'silent', 'majority', 'rakyat', 'bawah', 'tdk', 'masa', 'bodoh', 'terhadap', 'politik', 'mk', 'tentu', 'di', 'pilpres', 'mereka', 'gak', 'akan', 'pilih', 'capres', 'yg', 'melanggar', 'etika', 'buat', 'curang', 'merusak', 'demokrasi', 'terancam', 'kembali', 'otoriter', 'jadi', 'rakyat', 'bawah', 'juga', 'bersalah', 'shg', 'layak', 'menanggung', 'keburukan', 'di', 'hari', 'ini']
6416	itu sudah kuncinya dan sejatinya pilpres sejak th adalah sama peolple behaviour nya sama rakyat lapar	['itu', 'sudah', 'kuncinya', 'dan', 'sejatinya', 'pilpres', 'sejak', 'th', 'adalah', 'sama', 'peolple', 'behaviour', 'nya', 'sama', 'rakyat', 'lapar']

d. Normalization

Normalisasi adalah proses mengubah kata-kata yang melanggar aturan baku menjadi bentuk baku sesuai ejaan bahasa Indonesia. Tujuan dari normalisasi teks adalah untuk mengubah kata-kata yang melanggar aturan baku menjadi bentuk baku sesuai ejaan bahasa Indonesia. Ini termasuk mengubah kata-kata menjadi bentuk dasarnya, mengoreksi kesalahan ejaan, dan menstandarisasi penggunaan kata-kata yang serupa.

Tabel 3.5. Cuitan Hasil Normalisasi

no	tokenisasi	tweet_normalisasi
1	['mungkin', 'ini', 'jk', 'capres', 'tdk', 'perduli', 'norma', 'etik', 'rakyat', 'fokus', 'sidang', 'sengketa', 'pilpres', 'di', 'mk', 'ramai', 'demo', 'pemilu', 'curangybs', 'spt', 'bodoh', 'buta', 'tuli', 'tdk', 'mengubris', 'keadaan']	['mungkin', 'ini', 'jika', 'capres', 'tidak', 'perduli', 'norma', 'etik', 'rakyat', 'fokus', 'sidang', 'sengketa', 'pilpres', 'di', 'maka', 'ramai', 'demo', 'pemilu', 'curangybs', 'seperti', 'bodoh', 'buta', 'tuli', 'tidak', 'mengubris', 'keadaan']
2	['dana', 'operasional', 'presiden', 'lebih', 'banyak', 'utk', 'bansos', 'yg', 'tujuannya', 'menaikkan', 'popularitas', 'jokowi',	['dana', 'operasional', 'presiden', 'lebih', 'banyak', 'untuk', 'bansos', 'yang', 'tujuannya', 'menaikkan', 'popularitas', 'jokowi', 'sehingga',

Tabel Lanjutan 3.5. Cuitan Hasil Normalisasi

no	Tokenisasi	tweet_normalisasi
	'sehingga', 'bisa', 'cawecawe', 'selama', 'pilpres', 'artinya', 'ini', 'bisa', 'melanggar', 'undangundang', 'etika', 'asas', 'pemilu', 'jurdil', 'dugaan', 'pelanggaran', 'ini', 'harus', 'digulirkan', 'hak', 'angket']	'bisa', 'cawecawe', 'selama', 'pilpres', 'artinya', 'ini', 'bisa', 'melanggar', 'undangundang', 'etika', 'asas', 'pemilu', 'jurdil', 'dugaan', 'pelanggaran', 'ini', 'harus', 'digulirkan', 'hak', 'angket']
3	['jika', 'perpres', 'penetapan', 'apbn', 'sejak', 'november', 'untuk', 'pencairan', 'dana', 'oleh', 'menkeu', 'jelas', 'itu', 'sudah', 'masuk', 'tahap', 'pilpres', 'kampanye', 'kenapa', 'dadakan', 'yg', 'tersistem', 'mulai', 'tgl', 'sudah', 'bisa', 'disebarkan', 'bansos', 'termasuk', 'presiden', 'bagi', 'sendiri', 'setiap', 'kunjungan']	['jika', 'perpres', 'penetapan', 'apbn', 'sejak', 'november', 'untuk', 'pencairan', 'dana', 'oleh', 'menkeu', 'jelas', 'itu', 'sudah', 'masuk', 'tahap', 'pilpres', 'kampanye', 'kenapa', 'dadakan', 'yang', 'tersistem', 'mulai', 'tanggal', 'sudah', 'bisa', 'disebarkan', 'bansos', 'termasuk', 'presiden', 'bagi', 'sendiri', 'setiap', 'kunjungan']
4	['mmg', 'menkeu', 'sri', 'mulyani', 'pakar', 'dlm', 'mengamankan', 'presiden', 'jokowi', 'terkait', 'aliran', 'dana', 'bansos', 'yg', 'dilakukan', 'presiden', 'jokowi', 'dlm', 'menghadapi', 'elnino', 'bertepatan', 'masa', 'kampanyeu', 'pemilu', 'pilpres', 'berarti', 'aman', 'oresuden', 'jpkpwi', 'good', 'job', 'menkeu', 'sri', 'mulyani']	['memang', 'menkeu', 'sri', 'mulyani', 'pakar', 'dalam', 'mengamankan', 'presiden', 'jokowi', 'terkait', 'aliran', 'dana', 'bansos', 'yang', 'dilakukan', 'presiden', 'jokowi', 'dalam', 'menghadapi', 'elnino', 'bertepatan', 'masa', 'kampanyeu', 'pemilu', 'pilpres', 'berarti', 'aman', 'oresuden', 'jpkpwi', 'good', 'job', 'menkeu', 'sri', 'mulyani']
5	['sd', 'februari', 'angkanya', 'lebih', 'tinggi', 'dibanding', 'yg', 'padahal', 'covid', 'mulai', 'rame', 'di', 'bulan', 'maret', 'dan', 'kabinet', 'blm', 'ada', 'aksi', 'atau', 'instruksi', 'yg', 'extraordinary', 'pilpresnya', 'dah', 'lewat', 'yakni', 'yg', 'punya', 'ide', 'munculkan', 'isu', 'bansos', 'pada', 'pilpres', 'ini', 'unik', 'juga']	['sd', 'februari', 'angkanya', 'lebih', 'tinggi', 'dibanding', 'yang', 'padahal', 'covid', 'mulai', 'ramai', 'di', 'bulan', 'maret', 'dan', 'kabinet', 'belum', 'ada', 'aksi', 'atau', 'instruksi', 'yang', 'extraordinary', 'pilpresnya', 'sudah', 'lewat', 'yakni', 'yang', 'punya', 'ide', 'munculkan', 'isu', 'bansos', 'pada', 'pilpres', 'ini', 'unik', 'juga']
.....
6412	['gugatan', 'dari', 'terkait', 'sengketa', 'pilpres', 'di', 'mahkamah', 'konstitusi', 'sejak', 'awal', 'memang', 'banyak', 'menimbulkan', 'masalah']	['gugatan', 'dari', 'terkait', 'sengketa', 'pilpres', 'di', 'mahkamah', 'konstitusi', 'sejak', 'awal', 'memang', 'banyak', 'menimbulkan', 'masalah']
6413	['cawe', 'cawe', 'merupakan', 'bukti', 'adanya', 'niat', 'itikad', 'jahat', 'rencana', 'dan', 'aksi', 'tsm', 'dari', 'si', 'tukang', 'ngibul', 'neopki', 'kokowi', 'dkk', 'banditnya', 'atas', 'kecurangan', 'dan', 'kejahatan', 'pemilu', 'pilpres', 'demi', 'pemenangan', 'paslon', 'dalam', 'satu', 'putaran']	['cawe', 'cawe', 'merupakan', 'bukti', 'adanya', 'niat', 'itikad', 'jahat', 'rencana', 'dan', 'aksi', 'tsm', 'dari', 'si', 'tukang', 'berbohong', 'neopki', 'kokowi', 'dkk', 'banditnya', 'atas', 'kecurangan', 'dan', 'kejahatan', 'pemilu', 'pilpres', 'demi', 'pemenangan', 'paslon', 'dalam', 'satu', 'putaran']
6414	['yakni', 'dua', 'propinsi', 'di', 'aceh', 'dan', 'sumatera', 'barat', 'dua', 'titik', 'di', 'mana', 'bansos', 'diberikan', 'untuk', 'rakyat', 'oleh', 'pemerintah', 'demi', 'meringankan', 'beban', 'akibat', 'pandemi', 'tapi', 'kenyataannya', 'pemenang', 'pilpres', 'di', 'dua', 'propinsi', 'itu', 'adalah', 'paslon']	['yakni', 'dua', 'propinsi', 'di', 'aceh', 'dan', 'sumatera', 'barat', 'dua', 'titik', 'di', 'mana', 'bansos', 'diberikan', 'untuk', 'rakyat', 'oleh', 'pemerintah', 'demi', 'meringankan', 'beban', 'akibat', 'pandemi', 'tapi', 'kenyataannya', 'pemenang', 'pilpres', 'di', 'dua', 'propinsi', 'itu', 'adalah', 'paslon']
6415	['kalau', 'silent', 'majority', 'rakyat', 'bawah', 'tdk', 'masa', 'bodoh', 'terhadap', 'politik', 'mk', 'tentu', 'di', 'pilpres', 'mereka', 'gak', 'akan', 'pilih', 'capres', 'yg', 'melanggar', 'etika', 'buat', 'curang', 'merusak']	['kalau', 'silent', 'majority', 'rakyat', 'bawah', 'tidak', 'masa', 'bodoh', 'terhadap', 'politik', 'maka', 'tentu', 'di', 'pilpres', 'mereka', 'tidak', 'akan', 'pilih', 'capres', 'yang', 'melanggar', 'etika', 'buat', 'curang', 'merusak', 'demokrasi']

Tabel Lanjutan 3.5. Cuitan Hasil Normalisasi

no	Tokenisasi	tweet_normalisasi
	'demokrasi', 'terancam', 'kembali', 'otoriter', 'jadi', 'rakyat', 'bawah', 'juga', 'bersalah', 'shg', 'layak', 'menanggung', 'keburukan', 'di', 'hari', 'ini']	'terancam', 'kembali', 'otoriter', 'jadi', 'rakyat', 'bawah', 'juga', 'bersalah', 'shg', 'layak', 'menanggung', 'keburukan', 'di', 'hari', 'ini']
6416	['itu', 'sudah', 'kuncinya', 'dan', 'sejatinya', 'pilpres', 'sejak', 'th', 'adalah', 'sama', 'people', 'behaviour', 'nya', 'sama', 'rakyat', 'lapar']	['itu', 'sudah', 'kuncinya', 'dan', 'sejatinya', 'pilpres', 'sejak', 'th', 'adalah', 'sama', 'people', 'behaviour', 'nya', 'sama', 'rakyat', 'lapar']

e. Filtering

Tujuan dari *filtering* adalah untuk memisahkan konjungsi dan kata benda, atau biasa disebut dengan menghilangkan akhiran. Dengan menyaring atau menghilangkan kata-kata akhir, tujuannya adalah untuk membedakan kata-kata yang tidak terlalu mempengaruhi pemahaman teks, misalnya kata penghubung dan kata bantu. Ini menggunakan daftar kata penyelesaian yang telah ditentukan sebelumnya dalam proses ini.

Tabel 3.6 Hasil Filtering Stopwords

no	tweet_normalisasi	tweet_tokens_stopwords
1	['mungkin', 'ini', 'jika', 'capres', 'tidak', 'perduli', 'norma', 'etik', 'rakyat', 'fokus', 'sidang', 'sengketa', 'pilpres', 'di', 'maka', 'ramai', 'demo', 'pemilu', 'curangybs', 'seperti', 'bodoh', 'buta', 'tuli', 'tidak', 'mengubris', 'keadaan']	['jika', 'capres', 'tidak', 'perduli', 'norma', 'etik', 'rakyat', 'fokus', 'sidang', 'sengketa', 'pilpres', 'maka', 'ramai', 'demo', 'pemilu', 'curangybs', 'seperti', 'bodoh', 'buta', 'tuli', 'tidak', 'mengubris']
2	['dana', 'operasional', 'presiden', 'lebih', 'banyak', 'untuk', 'bansos', 'yang', 'tujuannya', 'menaikkan', 'popularitas', 'jokowi', 'sehingga', 'bisa', 'cawecawe', 'selama', 'pilpres', 'artinya', 'ini', 'bisa', 'melanggar', 'undangundang', 'etika', 'asas', 'pemilu', 'jurdil', 'dugaan', 'pelanggaran', 'ini', 'harus', 'digulirkan', 'hak', 'angket']	['dana', 'operasional', 'presiden', 'untuk', 'bansos', 'yang', 'tujuannya', 'menaikkan', 'popularitas', 'jokowi', 'cawecawe', 'pilpres', 'melanggar', 'undangundang', 'etika', 'asas', 'pemilu', 'jurdil', 'dugaan', 'pelanggaran', 'digulirkan', 'hak', 'angket']
3	['jika', 'perpres', 'penetapan', 'apbn', 'sejak', 'november', 'untuk', 'pencairan', 'dana', 'oleh', 'menkeu', 'jelas', 'itu', 'sudah', 'masuk', 'tahap', 'pilpres', 'kampanye', 'kenapa', 'dadakan', 'yang', 'tersistem', 'mulai', 'tanggal', 'sudah', 'bisa', 'disebarkan', 'bansos', 'termasuk', 'presiden', 'bagi', 'sendiri', 'setiap', 'kunjungan']	['perpres', 'penetapan', 'apbn', 'november', 'pencairan', 'dana', 'menkeu', 'masuk', 'tahap', 'pilpres', 'kampanye', 'dadakan', 'yang', 'tersistem', 'tanggal', 'disebarkan', 'bansos', 'presiden', 'kunjungan']

Tabel Lanjutan 3.6 Hasil Filtering Stopwords

no	tweet_normalisasi	tweet_tokens_stopwords
4	['memang', 'menkeu', 'sri', 'mulyani', 'pakar', 'dalam', 'mengamankan', 'presiden', 'jokowi', 'terkait', 'aliran', 'dana', 'bansos', 'yang', 'dilakukan', 'presiden', 'jokowi', 'dalam', 'menghadapi', 'elnino', 'bertepatan', 'masa', 'kampanyeu', 'pemilu', 'pilpres', 'berarti', 'aman', 'oresuden', 'jpkpwi', 'good', 'job', 'menkeu', 'sri', 'mulyani']	['memang', 'menkeu', 'sri', 'mulyani', 'pakar', 'dalam', 'mengamankan', 'presiden', 'jokowi', 'terkait', 'aliran', 'dana', 'bansos', 'yang', 'presiden', 'jokowi', 'dalam', 'menghadapi', 'elnino', 'bertepatan', 'kampanyeu', 'pemilu', 'pilpres', 'aman', 'oresuden', 'jpkpwi', 'good', 'job', 'menkeu', 'sri', 'mulyani']
5	['sd', 'februari', 'angkanya', 'lebih', 'tinggi', 'dibanding', 'yang', 'padahal', 'covid', 'mulai', 'ramai', 'di', 'bulan', 'maret', 'dan', 'kabinet', 'belum', 'ada', 'aksi', 'atau', 'instruksi', 'yang', 'extraordinary', 'pilpresnya', 'sudah', 'lewat', 'yakni', 'yang', 'punya', 'ide', 'munculkan', 'isu', 'bansos', 'pada', 'pilpres', 'ini', 'unik', 'juga']	['sd', 'februari', 'angkanya', 'dibanding', 'yang', 'covid', 'ramai', 'maret', 'kabinet', 'belum', 'aksi', 'instruksi', 'yang', 'extraordinary', 'pilpresnya', 'sudah', 'yang', 'ide', 'munculkan', 'isu', 'bansos', 'pilpres', 'unik']
.....
6412	['gugatan', 'dari', 'terkait', 'sengketa', 'pilpres', 'di', 'mahkamah', 'konstitusi', 'sejak', 'awal', 'memang', 'banyak', 'menimbulkan', 'masalah']	['gugatan', 'terkait', 'sengketa', 'pilpres', 'mahkamah', 'konstitusi', 'menimbulkan']
6413	['cawe', 'cawe', 'merupakan', 'bukti', 'adanya', 'niat', 'itikad', 'jahat', 'rencana', 'dan', 'aksi', 'tsm', 'dari', 'si', 'tukang', 'berbohong', 'neopki', 'kokowi', 'dkk', 'banditnya', 'atas', 'kecurangan', 'dan', 'kejahatan', 'pemilu', 'pilpres', 'demi', 'pemenangan', 'paslon', 'dalam', 'satu', 'putaran']	['cawe', 'cawe', 'bukti', 'niat', 'itikad', 'jahat', 'rencana', 'aksi', 'tsm', 'si', 'tukang', 'berbohong', 'neopki', 'kokowi', 'dkk', 'banditnya', 'kecurangan', 'kejahatan', 'pemilu', 'pilpres', 'pemenangan', 'paslon', 'putaran']
6414	['yakni', 'dua', 'propinsi', 'di', 'aceh', 'dan', 'sumatera', 'barat', 'dua', 'titik', 'di', 'mana', 'bansos', 'diberikan', 'untuk', 'rakyat', 'oleh', 'pemerintah', 'demi', 'meringankan', 'beban', 'akibat', 'pandemi', 'tapi', 'kenyataannya', 'pemenang', 'pilpres', 'di', 'dua', 'propinsi', 'itu', 'adalah', 'paslon']	['propinsi', 'aceh', 'sumatera', 'barat', 'titik', 'bansos', 'rakyat', 'pemerintah', 'meringankan', 'beban', 'akibat', 'pandemi', 'kenyataannya', 'pemenang', 'pilpres', 'propinsi', 'paslon']
6415	['kalau', 'silent', 'majority', 'rakyat', 'bawah', 'tidak', 'masa', 'bodoh', 'terhadap', 'politik', 'maka', 'tentu', 'di', 'pilpres', 'mereka', 'tidak', 'akan', 'pilih', 'capres', 'yang', 'melanggar', 'etika', 'buat', 'curang', 'merusak', 'demokrasi', 'terancam', 'kembali', 'otoriter', 'jadi', 'rakyat', 'bawah', 'juga', 'bersalah', 'shg', 'layak', 'menanggung', 'keburukan', 'di', 'hari', 'ini']	['silent', 'majority', 'rakyat', 'tidak', 'bodoh', 'politik', 'maka', 'pilpres', 'tidak', 'pilih', 'capres', 'yang', 'melanggar', 'etika', 'curang', 'merusak', 'demokrasi', 'terancam', 'otoriter', 'rakyat', 'bersalah', 'shg', 'layak', 'menanggung', 'keburukan']
6416	['itu', 'sudah', 'kuncinya', 'dan', 'sejatinya', 'pilpres', 'sejak', 'th', 'adalah', 'sama', 'people', 'behaviour', 'nya', 'sama', 'rakyat', 'lapar']	['kuncinya', 'sejatinya', 'pilpres', 'th', 'people', 'behaviour', 'nya', 'rakyat', 'lapar']

f. Stemming

Kata sebenarnya dengan sufiks diubah menjadi kata dasar. Pada umumnya

verba yang mempunyai imbuhan seperti *take* diubah menjadi kata pokok yaitu *take*. Dalam penyusunan kata sebenarnya, tujuannya adalah mengubah kata menjadi bentuk dasar atau kata dasar. Hal ini membantu mengurangi dimensi data dan meningkatkan konsistensi, terutama dengan menghilangkan sufiks atau akhiran kata.

Tabel 3.7 Cuitan Hasil Stemming

no	tweet_tokens_stopwords	tweet_tokens_stemmed
1	['jika', 'capres', 'tidak', 'perduli', 'norma', 'etik', 'rakyat', 'fokus', 'sidang', 'sengketa', 'pilpres', 'maka', 'ramai', 'demo', 'pemilu', 'curangybs', 'seperti', 'bodoh', 'buta', 'tuli', 'tidak', 'mengubris']	['jika', 'capres', 'tidak', 'duli', 'norma', 'etik', 'rakyat', 'fokus', 'sidang', 'sengketa', 'pilpres', 'maka', 'ramai', 'demo', 'milu', 'curangybs', 'seperti', 'bodoh', 'buta', 'tuli', 'tidak', 'mengubris']
2	['dana', 'operasional', 'presiden', 'untuk', 'bansos', 'yang', 'tujuannya', 'menaikkan', 'popularitas', 'jokowi', 'cawecawe', 'pilpres', 'melanggar', 'undangundang', 'etika', 'asas', 'pemilu', 'jurdil', 'dugaan', 'pelanggaran', 'digulirkan', 'hak', 'angket']	['dana', 'operasional', 'presiden', 'untuk', 'bansos', 'yang', 'tuju', 'naik', 'popularitas', 'jokowi', 'cawecawe', 'pilpres', 'langgar', 'undangundang', 'etika', 'asas', 'milu', 'jurdil', 'duga', 'langgar', 'gulir', 'hak', 'angket']
3	['perpres', 'penetapan', 'apbn', 'november', 'pencairan', 'dana', 'menkeu', 'masuk', 'tahapan', 'pilpres', 'kampanye', 'dadakan', 'yang', 'tersistem', 'tanggal', 'disebarkan', 'bansos', 'presiden', 'kunjungan']	['pres', 'tetap', 'apbn', 'november', 'cair', 'dana', 'menkeu', 'masuk', 'tahap', 'pilpres', 'kampanye', 'dada', 'yang', 'sistem', 'tanggal', 'sebar', 'bansos', 'presiden', 'kunjung']
4	['memang', 'menkeu', 'sri', 'mulyani', 'pakar', 'dalam', 'mengamankan', 'presiden', 'jokowi', 'terkait', 'aliran', 'dana', 'bansos', 'yang', 'presiden', 'jokowi', 'dalam', 'menghadapi', 'elnino', 'bertepatan', 'kampanyeu', 'pemilu', 'pilpres', 'aman', 'oresuden', 'jpkpwi', 'good', 'job', 'menkeu', 'sri', 'mulyani']	['memang', 'menkeu', 'sri', 'mulyani', 'pakar', 'dalam', 'aman', 'presiden', 'jokowi', 'kait', 'alir', 'dana', 'bansos', 'yang', 'presiden', 'jokowi', 'dalam', 'hadap', 'elnino', 'tepat', 'kampanyeu', 'milu', 'pilpres', 'aman', 'oresuden', 'jpkpwi', 'good', 'job', 'menkeu', 'sri', 'mulyani']
5	['sd', 'februari', 'angkanya', 'dibanding', 'yang', 'covid', 'ramai', 'maret', 'kabinet', 'belum', 'aksi', 'instruksi', 'yang', 'extraordinary']	['sd', 'februari', 'angka', 'banding', 'yang', 'covid', 'ramai', 'maret', 'kabinet', 'belum', 'aksi', 'instruksi', 'yang', 'extraordinary']
.....
6412	['gugatan', 'terkait', 'sengketa', 'pilpres', 'mahkamah', 'konstitusi', 'menimbulkan']	['gugat', 'kait', 'sengketa', 'pilpres', 'mahkamah', 'konstitusi', 'timbul']
6413	['cawe', 'cawe', 'bukti', 'niat', 'itikad', 'jahat', 'rencana', 'aksi', 'tsm', 'si', 'tukang', 'berbohong', 'neopki', 'kokowi', 'dkk', 'banditnya', 'kecurangan', 'kejahatan', 'pemilu', 'pilpres', 'pemenangan', 'paslon', 'putaran']	['cawe', 'cawe', 'bukti', 'niat', 'itikad', 'jahat', 'rencana', 'aksi', 'tsm', 'si', 'tukang', 'bohong', 'neopki', 'kokowi', 'dkk', 'bandit', 'curang', 'jahat', 'milu', 'pilpres', 'menang', 'paslon', 'putar']
6414	['propinsi', 'aceh', 'sumatera', 'barat', 'titik', 'bansos', 'rakyat', 'pemerintah', 'meringankan', 'beban', 'akibat', 'pandemi', 'kenyataannya', 'pemenang', 'pilpres', 'propinsi', 'paslon']	['propinsi', 'aceh', 'sumatera', 'barat', 'titik', 'bansos', 'rakyat', 'perintah', 'ringan', 'beban', 'akibat', 'pandemi', 'nyata', 'menang', 'pilpres', 'propinsi', 'paslon']

Tabel Lanjutan 3.7 Cuitan Hasil Stemming

no	tweet_tokens_stopwords	tweet_tokens_stemmed
6415	['silent', 'majority', 'rakyat', 'tidak ', 'bodoh', 'politik', 'maka ', 'pilpres', 'tidak ', 'pilih', 'capres', 'yang ', 'melanggar', 'etika', 'curang', 'merusak', 'demokrasi', 'terancam', 'otoriter', 'rakyat', 'bersalah', 'shg', 'layak', 'menanggung', 'keburukan']	['silent', 'majority', 'rakyat', 'tidak', 'bodoh', 'politik', 'maka', 'pilpres', 'tidak', 'pilih', 'capres', 'yang', 'langgar', 'etika', 'curang', 'rusak', 'demokrasi', 'ancam', 'otoriter', 'rakyat', 'salah', 'shg', 'layak', 'tanggung', 'keburu']
6416	['kuncinya', 'sejatinya', 'pilpres', 'th', 'peolple', 'behaviour', 'nya', 'rakyat', 'lapar']	['kunci', 'sejati', 'pilpres', 'th', 'peolple', 'behaviour', 'nya', 'rakyat', 'lapar']

g. Pelabelan Data

Data mentah yang dikumpulkan dari Twitter mencakup informasi tentang pengguna, teks tweet, timestamp, jumlah retweet, jumlah favorit, dan daftar hashtag. Data ini dapat digunakan sebagai dasar untuk analisis sentimen menggunakan algoritma Self-Organizing Maps (SOM) dalam penelitian tentang pemetaan sentimen masyarakat terhadap Pemilihan Presiden 2024.

Setelah melakukan proses *stemming*, proses selanjutnya adalah pelabelan data. Proses pelabelan dilakukan dengan menggunakan pustaka `textblob` pada `python`. Berfungsi untuk mengidentifikasi sebuah kata dalam kalimat yang memiliki sifat positif, negatif atau netral.

Tabel 3.8 Pelebelan Data

no	tweet_tokens_stemmed	sentimen
1	['jika', 'capres', 'tidak', 'duli', 'norma', 'etik', 'rakyat', 'fokus', 'sidang', 'sengketa', 'pilpres', 'maka', 'ramai', 'demo', 'milu', 'curangybs', 'seperti', 'bodoh', 'buta', 'tuli', 'tidak', 'mengubris']	negatif
2	['dana', 'operasional', 'presiden', 'untuk', 'bansos', 'yang', 'tujuan', 'naik', 'popularitas', 'jokowi', 'cawecawe', 'pilpres', 'langgar', 'undangundang', 'etika', 'asas', 'milu', 'jurdil', 'duga', 'langgar', 'gulir', 'hak', 'angket']	positif
3	['pres', 'tetap', 'apbn', 'november', 'cair', 'dana', 'menkeu', 'masuk', 'tahap', 'pilpres', 'kampanye', 'dada', 'yang', 'sistem', 'tanggal', 'sebar', 'bansos', 'presiden', 'kunjung']	positif
4	['memang', 'menkeu', 'sri', 'mulyani', 'pakar', 'dalam', 'aman', 'presiden', 'jokowi', 'kait', 'alir', 'dana', 'bansos', 'yang', 'presiden', 'jokowi', 'dalam', 'hadap', 'elnino', 'tepat', 'kampanyeu', 'milu', 'pilpres', 'aman', 'oresuden', 'jpkpwi', 'good', 'job', 'menkeu', 'sri', 'mulyani']	positif
5	['sd', 'februari', 'angka', 'banding', 'yang', 'covid', 'ramai', 'maret', 'kabinet', 'belum', 'aksi', 'instruksi', 'yang', 'extraordinary', 'pilpresnya', 'sudah', 'yang', 'ide', 'muncul', 'isu', 'bansos', 'pilpres', 'unik']	positif

Tabel Lanjutan 3.8 Pelebelan Data

no	tweet_tokens_stemmed	sentimen
.....
6412	['caprescawapres', 'terima', 'hasil', 'pilpres', 'pks', 'terima', 'rakyat', 'rakyat', 'putus', 'mas', 'anies', 'mas', 'ganjar', 'prabowo', 'benarbenar', 'rendah', 'hati', 'terima', 'rakyat', 'ketua', 'dpp', 'pks', 'mardani', 'ali', 'sera']	negatif
6413	['cawe', 'cawe', 'bukti', 'niat', 'itikad', 'jahat', 'rencana', 'aksi', 'tsm', 'si', 'tukang', 'bohong', 'neopki', 'kokowi', 'dkk', 'bandit', 'curang', 'jahat', 'milu', 'pilpres', 'menang', 'paslon', 'putar']	negatif
6414	['kader', 'pks', 'capres', 'terima', 'hasil', 'pilpres']	netral
6415	['silent', 'majority', 'rakyat', 'tidak', 'bodoh', 'politik', 'maka', 'pilpres', 'tidak', 'pilih', 'capres', 'yang', 'langgar', 'etika', 'curang', 'rusak', 'demokrasi', 'ancam', 'otoriter', 'rakyat', 'salah', 'shg', 'layak', 'tanggung', 'keburu']	negatif
6416	['kunci', 'sejati', 'pilpres', 'th', 'peopple', 'behaviour', 'nya', 'rakyat', 'lapar']	netral

3.3. *Extraction Fearure*

Metode *extraction fearure* yang umum digunakan dalam pemrosesan teks untuk mengubah teks mentah menjadi representasi numerik yang dapat digunakan dalam analisis lebih lanjut, seperti klasifikasi, clustering, atau pembelajaran mesin adalah TF-IDF dan LSA. Berikut ini merupakan penjelasan singkat mengenai masing-masing metode.

3.3.1. TF-IDF (*Term Frequency-Inverse Document Frequency*)

Extraction fearure TF-IDF yang dilakukan setelah data tweet telah melewati *pre-prosessing*. TF-IDF akan menghitung bobot setiap kata. Bobot dihitung mengacu pada nilai frekuensi kata pada setiap dokumen. Pembobotan TF-IDF adalah teknik yang digunakan dalam pengolah kata untuk mengevaluasi pentingnya sebuah kata dalam dokumen dalam korpus teks. Hal ini sangat sering digunakan dalam analisis teks, pengambilan informasi dan pemahaman bahasa alami. Berikut adalah langkah-langkah umum dalam menerapkan pembobotan TF-IDF:

1. Data Cuitan Pengguna (*Tokenisasi*), dokumen teks dibagi menjadi token atau kata-kata individual. Ini bisa melibatkan langkah-langkah seperti menghapus tanda baca, memisahkan kata-kata, dan mengabaikan kata-kata penghubung. Kata-kata dalam fokus penelitian dipecah menjadi token atau kata-kata individual. Dalam hal ini, kata-kata yang relevan adalah, "Sentimen Masyarakat", dan "Pilpres 2024".
2. *Term Frequency* (TF) dalam hal ini, mengukur seberapa sering sebuah kata muncul dalam sebuah dokumen. Kata tersebut dihitung dengan cara yang relatif sederhana, yaitu dengan menghitung jumlah kemunculan suatu kata dalam dokumen tersebut.

Rumus TF:

$$TF_{(t,d)} = \frac{\text{jumlah kemunculan kata } t \text{ dalam dokumen } d}{\text{jumlah total kata dalam dokumen } d} \dots \dots \dots (3.1)$$

3. *Inverse Document Frequency* (IDF), di sini mengukur pentingnya kata dalam keseluruhan korpus. Kata-kata yang muncul di banyak dokumen dianggap kurang penting, sedangkan kata-kata yang muncul di sedikit dokumen dianggap lebih penting.

$$IDF_{(t,d)} = \text{Log} \left(\frac{\text{jumlah total dokumen dalam korpus } D}{\text{jumlah dokumen yang mengandung kata } t} \right) \dots \dots \dots (3.2)$$

4. *TF-IDF Weighting*: Hasil perkalian TF dan IDF yang memberi bobot lebih pada kata-kata yang sering muncul dalam suatu dokumen namun jarang muncul di dokumen lain.

Rumus TF-IDF:

$$TF - IDF_{(t,d,D)} = TF_{(t,d)} \times IDF_{(t,D)} \dots \dots \dots (3.3)$$

5. Normalisasi, hasil TF-IDF dapat dinormalisasi untuk memastikan bahwa panjang dokumen tidak mempengaruhi bobot. Salah satu metode normalisasi yang umum adalah dengan menggunakan TF-IDF *Euclidean Normalization* atau *Cosine Normalization*.

Proses ini akan memberikan representasi numerik dari setiap kata dalam fokus penelitian, dengan bobot yang menunjukkan pentingnya kata tersebut dalam konteks korpus fokus penelitian. Dengan demikian, dapat melihat kata-kata mana yang paling relevan dalam fokus penelitian tersebut dan bagaimana kata-kata tersebut dapat mempengaruhi analisis sentimen dan pemetaan masyarakat terhadap Pilpres 2024.

3.3.2. *Latent Semantic Analysis (LSA)*

Setelah dataset lolos proses pembobotan dengan metode TF-IDF, maka dilakukan pengujian dengan metode lain yaitu *Latent Semantic Analysis (LSA)*, adalah teknik statistik dan matematika yang digunakan dalam pemrosesan bahasa alami dan analisis teks untuk mengidentifikasi hubungan antara kata dan konsep dalam dokumen (Fiqri dkk. 2023:45). Proses LSA dimulai dengan pengumpulan data teks yang akan dianalisis, seperti artikel, buku, atau kumpulan dokumen lainnya. Setelah data terkumpul, langkah pertama adalah preprocessing teks yang melibatkan beberapa tahap penting. Teks dibersihkan dari karakter khusus dan format yang tidak diperlukan, kemudian dilakukan tokenisasi untuk memecah teks menjadi unit-unit kata atau token. Penghapusan stop words, kata-kata umum tanpa

arti penting, seperti "dan", "atau", "tetapi", dihilangkan untuk mengurangi gangguan data. Selanjutnya, kata-kata diubah ke bentuk dasar menggunakan stemming atau lemmatization, dan seluruh teks diubah menjadi huruf kecil untuk konsistensi.

Langkah berikutnya adalah pembentukan matriks term-document. Dalam matriks ini, baris mewakili kata-kata unik (terms) yang muncul dalam korpus, dan kolom mewakili dokumen. Setiap sel dalam matriks menunjukkan frekuensi kemunculan setiap kata dalam dokumen tertentu. Untuk meningkatkan kualitas matriks, bobot frekuensi kata sering kali diubah menggunakan *Term Frequency-Inverse Document Frequency* (TF-IDF), yang mengukur pentingnya kata dalam sebuah dokumen relatif terhadap seluruh korpus (Yusuf 2022:12). Hal ini membantu mengurangi dampak kata-kata yang umum namun tidak bermakna.

Setelah matriks term-document atau TF-IDF terbentuk, langkah selanjutnya adalah penerapan Singular Value Decomposition (SVD). SVD dapat menerapkan rumus di bawah ini.

$$A_{mn} = U_{mn} \times S_{mn} \times V_{nn}^t \dots \dots \dots (3.4)$$

Dimana:

- A_{mn} : matriks awal.
- U_{mn} : matriks orthogonal U.
- S_{mn} : matriks diagonal s.
- V_{nn}^t : transpose matriks orthogonal V.

Melalui SVD, LSA mampu mengurangi dimensi data dengan menangkap hubungan laten antara kata-kata dan dokumen. Dengan memilih sejumlah kecil

singular values dari matriks Σ , dimensi data dapat direduksi secara signifikan, sehingga mengungkap struktur laten dalam data tanpa noise yang tidak diinginkan (Nuansa 2017:45).

Data teks sekarang direpresentasikan dalam ruang laten berdimensi lebih rendah, di mana setiap dokumen dan kata memiliki representasi vektor yang lebih kompak dan bermakna. Analisis lebih lanjut dilakukan dengan mengevaluasi vektor-vektor dalam ruang laten ini. Dokumen-dokumen yang memiliki vektor yang dekat dalam ruang laten cenderung membahas topik yang sama, memungkinkan peneliti untuk mengidentifikasi topik utama atau konsep tersembunyi dalam kumpulan dokumen.

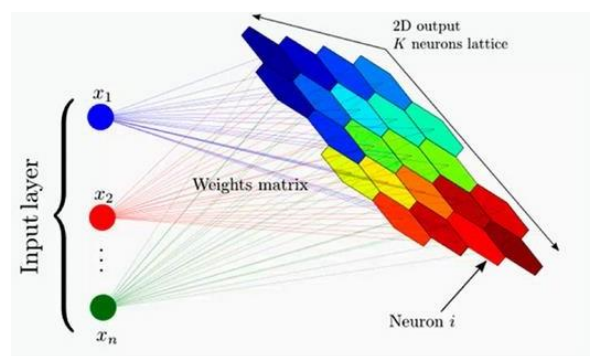
Representasi data dalam ruang laten ini dapat digunakan untuk berbagai tugas lanjutan seperti clustering atau klasifikasi teks. Clustering mengelompokkan dokumen ke dalam kelompok yang serupa berdasarkan kedekatan vektor dalam ruang laten, sementara klasifikasi dapat mengkategorikan dokumen berdasarkan label yang telah ditentukan (Firdlous 2023:12). Hasil analisis kemudian dievaluasi menggunakan metrik yang sesuai untuk memastikan model bekerja dengan baik. Dalam clustering, metrik seperti Silhouette Score atau Purity dapat digunakan, sedangkan dalam klasifikasi, metrik seperti akurasi, precision, recall, dan F1 score menjadi pilihan.

Hasil LSA kemudian diinterpretasikan untuk memberikan wawasan atau informasi yang dapat digunakan dalam penelitian lebih lanjut atau aplikasi praktis. Sebagai contoh, dalam analisis sentimen, hasil LSA dapat menunjukkan pola umum dari sentimen positif atau negatif dalam kumpulan ulasan produk. Semua langkah,

temuan, dan interpretasi hasil dicatat dan didokumentasikan dengan baik dalam laporan akhir yang mencakup penjelasan metodologi, hasil analisis, evaluasi, dan kesimpulan yang diambil dari penelitian (Fathir 2023:16). Dengan mengikuti proses ini secara sistematis, LSA menjadi alat yang kuat untuk mengungkap struktur laten dan hubungan dalam data teks.

3.4. Permodelan *Self-Organizing Maps* (SOM)

Setelah dataset mengalami proses pembobotan dengan menggunakan dua metode berbeda yaitu metode TF-IDF dan LSA, maka dataset tersebut mengalami proses pelatihan menggunakan algoritma SOM. Proses pelatihan menggunakan algoritma SOM membagi kelas menjadi tiga kelas polar yaitu positif, netral dan negatif. Alur pelatihan SOM memiliki beberapa tahapan yaitu inialisasi, pelatihan, dan pematkhiran.



Gambar 3.4 Ilustrasi struktur layer pada SOM

Dalam hal ini, SOM merupakan salah satu jenis algoritma jaringan syaraf tiruan yang digunakan untuk analisis data dan pemetaan data multidimensi ke dalam ruang berdimensi rendah. Ini adalah alat yang berguna untuk menemukan pola tersembunyi dalam data yang kompleks. SOM bekerja dengan menempatkan neuron dalam jaringan 2D atau 3D yang mewakili data masukan. Selama pelatihan, bobot setiap neuron diperbarui secara berulang untuk mencocokkan pola dalam

data. Neuron dengan bobot serupa atau serupa terletak berdekatan satu sama lain dalam jaringan dan dengan demikian mewakili kelompok atau kelas informasi yang serupa. Proses pelatihan SOM biasanya melibatkan langkah-langkah berikut:

1. Inisialisasi

Langkah awal dari algoritma ini adalah inisialisasi, dimana setiap node data menerima vektor bobot acak. Nilai vektor bobot setiap neuron berbeda satu sama lain. Vektor bobot dapat dilambangkan dengan W_{ij} , radius tetangga, serta parameter *learning rate* (α).

2. Training

Untuk setiap neuron, proses pelatihan dilakukan sebagai langkah awal berupa sampling, dimana langkah ini mengambil sampel secara acak dari kumpulan data masukan, dimana vektor x mewakili pola aktivasi yang diterapkan pada jaringan.

Selanjutnya, kita beralih ke fase mencari unit pencocokan terbaik (BMU), di mana neuron dengan perbedaan terkecil dari data yang diberikan menang. BMU diperoleh dengan memilih vektor neuron dan bobot yang memiliki jarak terkecil dari vektor masukan menggunakan jarak *Euclidean* sesuai persamaan (3.5)

$$[(x, y), (a, b)] = \sqrt{(x - a)^2 + (y - b)^2} \dots \dots \dots (3.5)$$

3. Updating

Proses ini mengubah atau mengoreksi nilai bobot pada data berikut menggunakan persamaan berikut:

$$W_{ij}(new) = W_{ij}(old) + \alpha[x_i - W_{ij}(old)] \dots \dots \dots (3.6)$$

Dimana:

$W_{ij}(new)$: adalah bobot baru yang akan dicari atau dimodifikasi.

$W_{ij}(old)$: adalah bobot sebelumnya.

α : adalah *learning rate*.

x_i : adalah data atribut ke i berdasarkan bobot yang dicari.

Setelah itu, learning rate (α) untuk proses iterasi atau epoch berikutnya dihitung dengan mengalikan learning rate sebelumnya dengan 0,6. Proses kemudian dilanjutkan dengan penghitungan ulang jarak seperti pada tahap latihan hingga pengulangan yang diperlukan selesai.

3.5. Uji Coba

Pada tahap ini, peneliti mencoba melakukan uji coba terkait dengan data yang sudah dikumpulkan diberikan pelabelan. Hal ini, bertujuan untuk mengembangkan sistem yang dapat mengidentifikasi dan menganalisis sentimen masyarakat terhadap Pilpres 2024. Sistem ini menggunakan beberapa teknik dalam pemrosesan teks dan pembelajaran mesin, termasuk pelabelan data, pembobotan TF-IDF, dan pemodelan Self-Organizing Map (SOM). Dengan demikian, dalam uji coba, diharapkan sistem dapat membantu dalam memahami dan menganalisis sentimen masyarakat terhadap Pilpres 2024. Dengan memahami pola sentimen ini, mengambil kebijakan dapat membuat keputusan yang lebih baik dan strategi yang lebih efektif dalam pemilihan presiden.

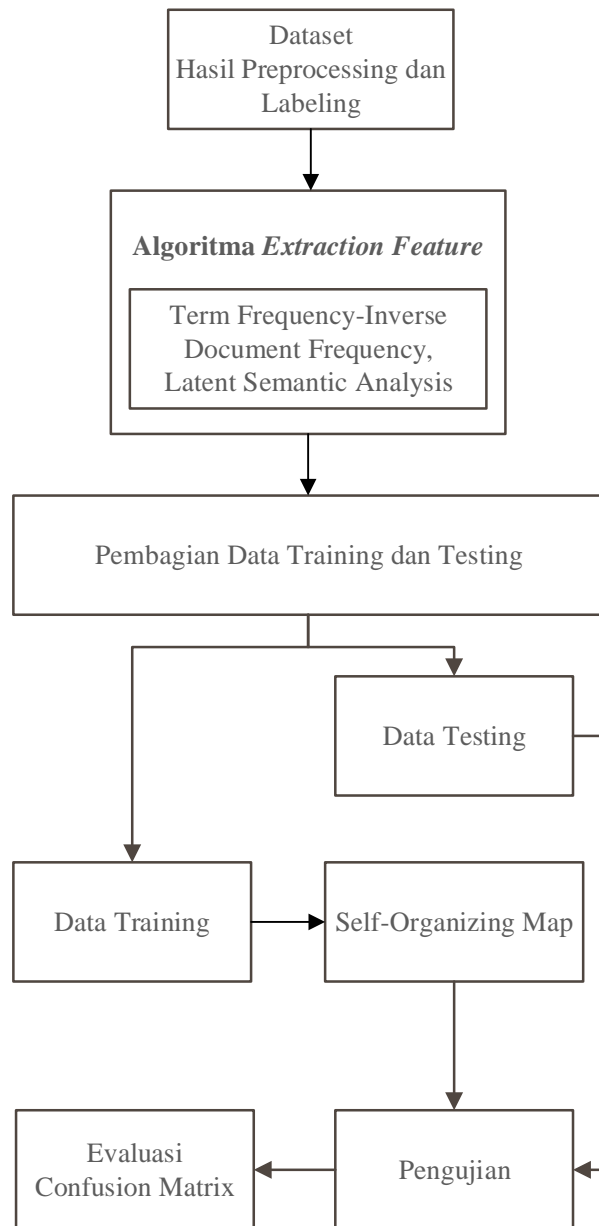
3.6. Konseptual Model

Berdasarkan penjelasan di atas, peneliti membuat model konseptual dalam penelitian ini. Tujuan dari penelitian ini adalah untuk menciptakan sistem yang mampu menganalisis sikap masyarakat terhadap Pilpres 2024. Model ini terdiri dari

tiga tahap utama: pelabelan data, pembobotan TF-IDF, dan pemodelan SOM. Dengan pendekatan ini, sistem dapat membantu dalam memahami opini publik terhadap Pilpres 2024 dan memberikan wawasan yang berharga bagi peneliti dan pembuat kebijakan.

Langkah model konteks *pertama*, pengumpulan data: Data mengenai pemilu presiden 2024 dikumpulkan dari berbagai sumber, antara lain media sosial, berita, dan forum diskusi. *Kedua*, pelabelan data: data tekstual yang dikumpulkan diberi label emosi (positif, negatif, netral) baik secara manual maupun menggunakan model pelabelan otomatis. *Ketiga*, pra-pemrosesan teks: data teks yang diberi tag dibersihkan dan diproses, termasuk menghilangkan karakter non-abjad, penandaan, menghilangkan kata-kata berhenti dan kata-kata sebenarnya.

Keempat, pembobotan TF-IDF: Bobot TF-IDF dihitung untuk setiap teks, menghasilkan vektor numerik yang mewakili dokumen dalam bentuk yang dapat diolah oleh algoritma pembelajaran mesin. *Kelima*, pemodelan SOM: Model SOM dilatih dengan vektor TF-IDF untuk mengelompokkan dokumen berdasarkan sentimen. Hasil klusterisasi divisualisasikan untuk melihat distribusi sentimen dalam data. Dan *keenam*, Evaluasi Hasil: Kualitas klusterisasi dievaluasi menggunakan Silhouette Score dan interpretasi visualisasi untuk memahami pola sentimen yang muncul. Tahapan terakhir, hasil dan diskusi: Analisis hasil klusterisasi memberikan wawasan mengenai opini publik terhadap Pilpres 2024, membantu dalam memahami sentimen masyarakat dan pengaruhnya terhadap calon presiden. Berikut gambar konseptual model dari penelitian ini.



Gambar 3.5 Konseptual Model

Berdasarkan konseptual model diagram alir proses analisis teks menggunakan algoritma Self-Organizing Map (SOM). Berikut adalah penjelasan alur dari diagram tersebut diawali dengan dataset Hasil Preprocessing dan Labeling yang merupakan proses dimulai dengan dataset yang telah melalui tahap *preprocessing* (pemrosesan awal) dan labeling (pemberian label). *Preprocessing*

mungkin mencakup pembersihan teks, tokenisasi, penghapusan stop words, stemming, dan lain-lain.

Algoritma *Extraction Feature*, Setelah preprocessing, data dianalisis menggunakan teknik analisis teks. Pada diagram ini, teknik yang digunakan adalah TF-IDF dan LSA. TF-IDF digunakan untuk mengukur pentingnya kata-kata dalam dokumen, sedangkan LSA digunakan untuk menemukan hubungan semantik antar kata. Ketika data pelatihan dan pengujian dipisah, data yang dianalisis dibagi menjadi dua bagian: data pelatihan dan data pengujian. Data pelatihan digunakan untuk melatih model, sedangkan data pengujian digunakan untuk menguji performa model yang dilatih.

Data pelatihan, data pelatihan digunakan untuk melatih algoritma SOM. SOM adalah jenis jaringan saraf tiruan yang digunakan untuk mengelompokkan data berdasarkan kesamaan fitur. Algoritma SOM diterapkan pada data pelatihan untuk membuat peta topologi data. Peta ini menggambarkan hubungan dan pola dalam data yang dapat digunakan untuk klasifikasi atau pengelompokan. Dalam pengujian, model SOM yang telah dilatih kemudian diuji menggunakan data uji. Tujuan dari pengujian ini adalah untuk mengevaluasi seberapa baik model yang dilatih dapat mengklasifikasikan atau mengelompokkan data yang belum pernah dilihat sebelumnya.

Evaluasi Confusion Matrix, hasil dari pengujian kemudian dievaluasi menggunakan *confusion matrix*. *Confusion matrix* merupakan alat evaluasi yang menunjukkan kinerja model klasifikasi, dengan memperlihatkan jumlah prediksi yang benar dan salah untuk masing-masing kelas. Dengan demikian, keseluruhan

proses ini bertujuan untuk membuat model yang dapat menganalisis teks dan mengelompokkan atau mengklasifikasikan data dengan akurasi yang tinggi. Diagram ini menggambarkan alur kerja dari pengumpulan data hingga evaluasi akhir model.

BAB IV

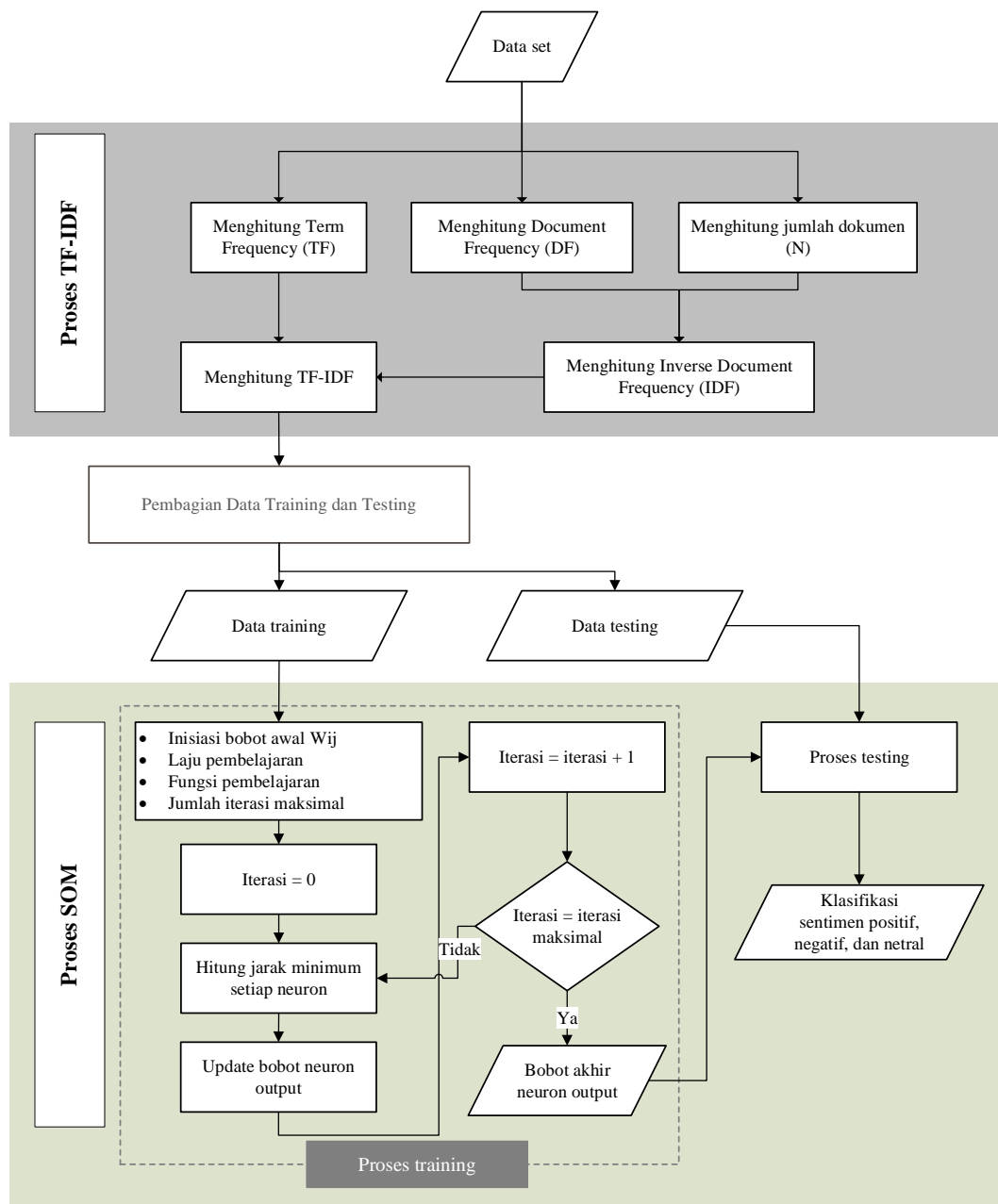
TERM FREQUENCY-INVERSE DOCUMENT FREQUENCY (TF-IDF) DAN SELF-ORGANIZING MAPS (SOM)

4.1. Desain

Flowchart di bawah ini menggambarkan alur proses analisis sentimen yang melibatkan dua metode utama: TF-IDF dan SOM. Proses dimulai dengan pengumpulan data set yang akan dianalisis. Selanjutnya, dilakukan perhitungan *Term Frequency* (TF) untuk setiap dokumen dalam data set, yang menentukan frekuensi kemunculan setiap istilah (term) dalam dokumen tersebut. Kemudian, dihitung *Document Frequency* (DF) yang menunjukkan jumlah dokumen dalam data set yang mengandung setiap istilah. Setelah itu, total jumlah dokumen (N) dalam data set dihitung.

Dengan nilai DF dan N, perhitungan *Inverse Document Frequency* (IDF) dilakukan untuk setiap istilah, dan hasilnya digunakan untuk menghitung nilai TF-IDF bagi setiap istilah dalam setiap dokumen. Setelah nilai TF-IDF diperoleh, dataset dibagi menjadi dua bagian yaitu data latih dan data uji. Data latih digunakan untuk melatih model, sedangkan data uji digunakan untuk menguji model. Data latih dan uji kemudian diolah menggunakan SOM.

Pada langkah terakhir, model yang dilatih diuji terhadap data pengujian. Berdasarkan hasil pengujian, model mengklasifikasikan sentimen data pengujian ke dalam kategori positif, negatif, atau netral. Dengan mengikuti langkah-langkah tersebut, proses analisis sentimen dapat dilakukan secara sistematis menggunakan kombinasi metode TF-IDF untuk ekstraksi fitur dan SOM untuk klasifikasi.



Gambar 4.1. Alur Proses Analisis Sentimen

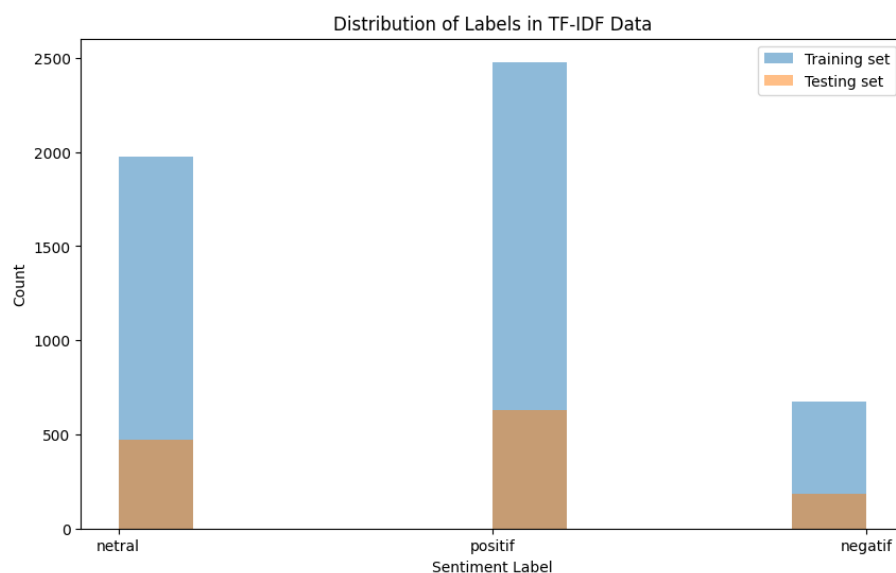
4.2. Implementasi

Proses klasifikasi sentimen menggunakan TF-IDF dan SOM, dimulai dari pembobotan TF-IDF kemudian diakhiri proses SOM. Data yang digunakan merupakan data yang telah dilakukan preprocessing dan pelabelan. Setelah dilakukan proses pembobotan TF-IDF data didistribusikan kedalam dataset training

dan testing. Dari hasil distribusi, training set terdapat data berlabel negatif: 676 data, berlabel netral: 1.977 data, dan berlabel positif: 2.479 data. Sementara testing set terdapat data berlabel negatif: 185 data, berlabel netral: 472 data, dan berlabel positif: 627 data. Data perbandingan dapat dilihat pada table dan grafik di bawah ini.

Tabel 4.1. Data perbandingan

	Negatif	Positif	Netral	Total
Training	676	1.977	2.479	5.132
Testing	185	472	627	1.284



Gambar 4.2. Data Perbandingan

Dari hasil percobaan yang dilakukan, algoritma SOM dengan TF-IDF memperoleh hasil akurasi yang kurang baik, yaitu 0.54. Hal ini terlihat dari precision, recall, dan f1-score yang bervariasi untuk setiap kelas. Precision untuk kelas negatif hanya mencapai 0.43, dengan recall yang sangat rendah yaitu 0.05, dan f1-score sebesar 0.09, menunjukkan bahwa model kesulitan dalam mendeteksi dan mengklasifikasikan data negatif. Untuk kelas netral, precision adalah 0.52,

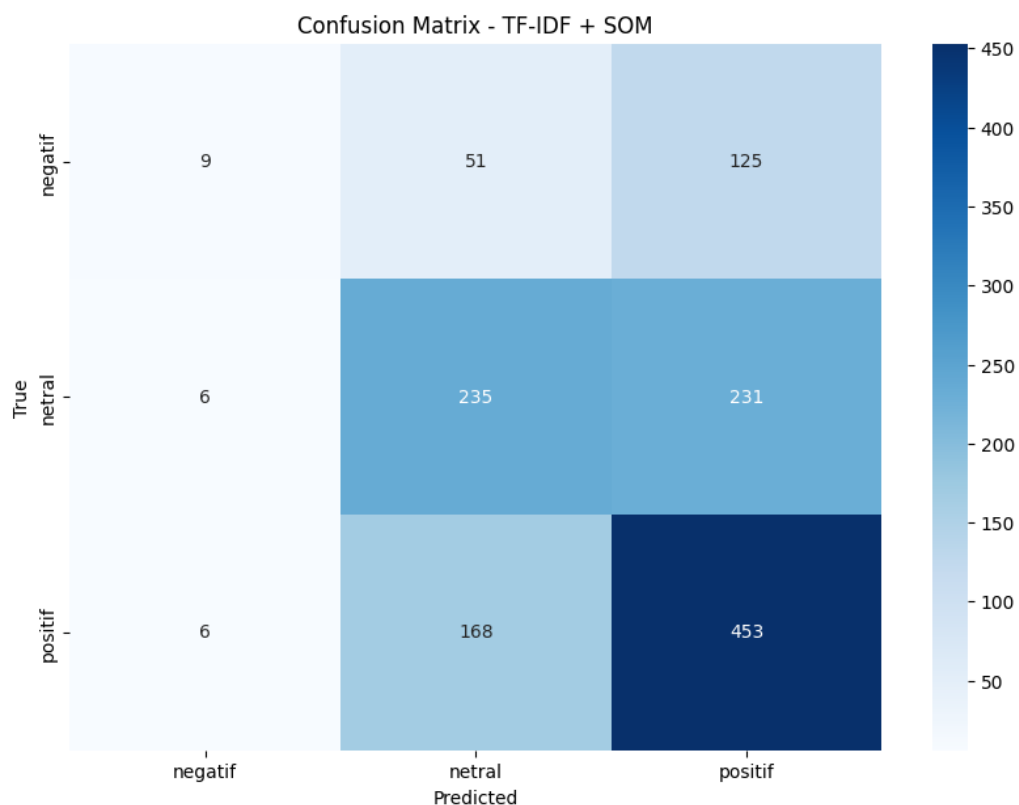
recall 0.50, dan f1-score 0.51, yang menunjukkan performa yang sedikit lebih baik namun masih di bawah standar yang diharapkan. Kelas positif memiliki performa terbaik dengan precision 0.56, recall 0.72, dan f1-score 0.63, namun tetap menunjukkan bahwa ada ruang untuk perbaikan, khususnya dalam meningkatkan recall untuk kelas negatif dan netral. Rata-rata makro dan rata-rata tertimbang juga menunjukkan bahwa model ini memiliki performa yang konsisten rendah di seluruh metrik, mengindikasikan perlunya pendekatan atau penyesuaian lebih lanjut untuk meningkatkan akurasi keseluruhan.

Tabel 4.2. Akurasi Keseluruhan

	precision	recall	f1-score	support
negative	0.43	0.05	0.09	185
netral	0.52	0.50	0.51	472
positif	0.56	0.72	0.63	627
accuracy			0.54	1284
macro avg	0.50	0.42	0.41	1284
weighted avg	0.53	0.54	0.51	1284

Selain itu, distribusi support yang tidak merata dengan jumlah data terbesar pada kelas positif (627), diikuti oleh netral (472), dan negatif (185) juga dapat mempengaruhi hasil evaluasi. Ketidakseimbangan ini mungkin menyebabkan bias dalam model, di mana kelas dengan lebih banyak data (positif dan netral) mendapatkan performa yang lebih baik dibandingkan kelas negatif. Rata-rata makro, yang memberikan bobot yang sama untuk setiap kelas, menunjukkan performa yang lebih rendah dengan precision 0.50, recall 0.42, dan f1-score 0.41. Ini menegaskan bahwa model cenderung tidak berkinerja baik pada kelas yang lebih kecil. Sebaliknya, rata-rata tertimbang, yang memperhitungkan jumlah data dalam

setiap kelas, memberikan hasil yang sedikit lebih baik dengan precision 0.53, recall 0.54, dan f1-score 0.51, tetapi tetap mengindikasikan bahwa model ini belum optimal. Oleh karena itu, penting untuk mempertimbangkan teknik penyeimbangan data atau metode augmentasi untuk meningkatkan representasi kelas yang kurang dalam dataset.

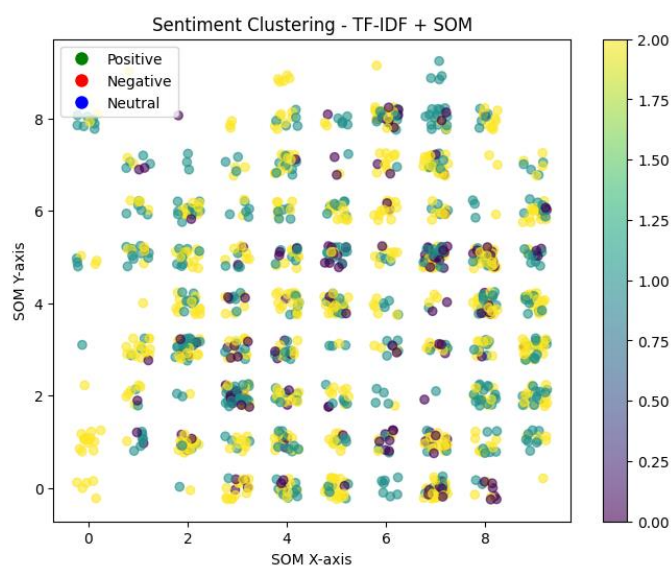


Gambar 4.3. Confusion Matrix

Pada confusion matrix diatas menunjukkan model klasifikasi yang menggunakan TF-IDF dan SOM. Matriks ini menunjukkan kinerja model dalam memprediksi tiga kelas: "negatif", "netral", dan "positif". Pada sumbu Y (label asli), kita melihat label sebenarnya dari data uji yang terdiri dari negatif, netral, dan positif. Sementara pada sumbu X (label prediksi), kita melihat label yang diprediksi oleh model yang juga terdiri dari negatif, netral, dan positif.

Nilai dalam matriks menunjukkan jumlah prediksi yang benar dan salah yang dibuat oleh model untuk masing-masing kelas. Sebagai contoh, untuk kelas negatif, model dengan benar memprediksi 9 sampel sebagai negatif, namun salah memprediksi 51 sampel sebagai netral dan 125 sampel sebagai positif. Untuk kelas netral, model dengan benar memprediksi 235 sampel, tetapi salah memprediksi 6 sampel sebagai negatif dan 231 sampel sebagai positif. Sementara itu, untuk kelas positif, model dengan benar memprediksi 453 sampel, tetapi salah memprediksi 6 sampel sebagai negatif dan 168 sampel sebagai netral.

Dari confusion matrix ini, kita dapat menganalisis bahwa model memiliki kinerja terbaik dalam memprediksi kelas positif, yang ditunjukkan dengan jumlah prediksi benar yang paling banyak (453). Namun, model kesulitan dalam memprediksi kelas negatif, yang terlihat dari tingginya jumlah kesalahan prediksi di kelas ini (125 sampel diprediksi sebagai positif). Secara keseluruhan, analisis ini menunjukkan area di mana model berkinerja baik dan area yang memerlukan perbaikan lebih lanjut.



Gambar 4.4. Sentiment Clustering – TF-IDF + SOM

Di sisi lain grafik clustering sentimen menggunakan metode TF-IDF dan SOM memvisualisasikan distribusi data sentimen pada peta dua dimensi yang dihasilkan oleh SOM. Sumbu X dan Y mewakili koordinat pada peta SOM, sementara warna dan posisi titik-titik pada grafik menunjukkan distribusi sentimen yang berbeda.

Grafik ini menunjukkan bahwa sentimen positif (ditandai dengan titik hijau) tersebar di berbagai lokasi pada peta SOM, namun terdapat beberapa area di mana titik-titik hijau lebih terkonsentrasi, menunjukkan adanya kelompok data dengan sentimen positif yang kuat. Sentimen netral (ditandai dengan titik biru) juga tersebar di seluruh peta, dengan beberapa kluster terlihat lebih padat. Sentimen negatif (ditandai dengan titik merah) terlihat lebih sedikit dibandingkan dengan dua kategori lainnya, dan juga tersebar di seluruh peta dengan beberapa konsentrasi kecil.

Skala warna di sisi kanan grafik menunjukkan densitas atau kepadatan data dalam setiap kluster. Warna kuning menunjukkan kepadatan tertinggi, sedangkan warna ungu menunjukkan kepadatan terendah. Dari skala ini, kita dapat melihat bahwa ada beberapa area dengan kepadatan tinggi, di mana berbagai jenis sentimen (positif, netral, dan negatif) tumpang tindih.

Analisis grafik ini menunjukkan bahwa metode TF-IDF dan SOM berhasil mengelompokkan data sentimen ke dalam berbagai kluster berdasarkan kesamaan fitur. Namun, terdapat tumpang tindih antara berbagai jenis sentimen di beberapa kluster, yang mengindikasikan bahwa model ini mungkin masih perlu penyempurnaan lebih lanjut untuk meningkatkan ketepatan klasifikasi sentimen.

Secara keseluruhan, distribusi data sentimen pada peta SOM ini memberikan wawasan tentang bagaimana sentimen tersebar dan mengelompok dalam dataset, serta area yang mungkin memerlukan perhatian lebih lanjut dalam model klasifikasi.

BAB V

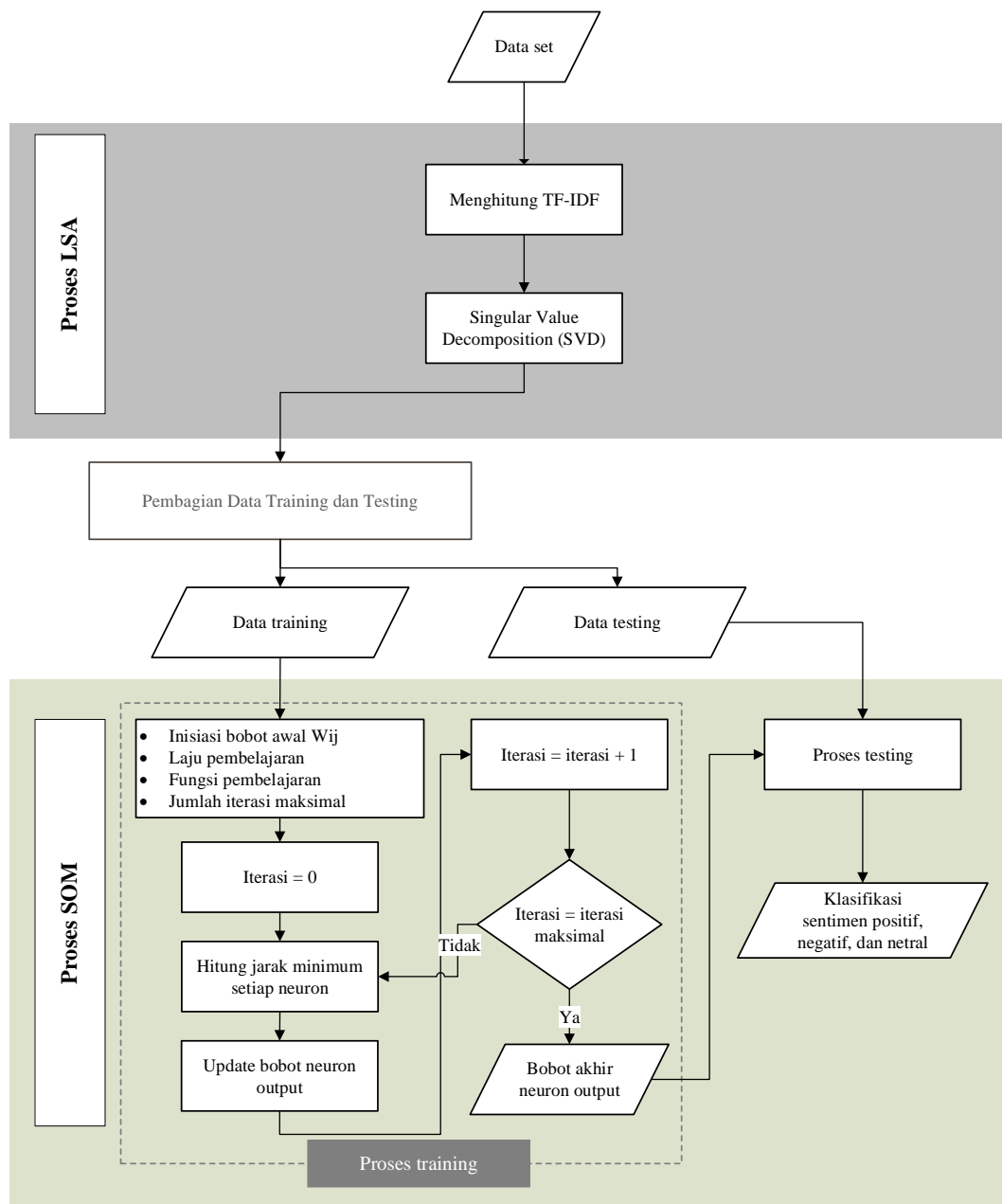
LATENT SEMANTIC ANALYSIS (LSA) DAN SELF-ORGANIZING MAPS (SOM)

5.1. Desain

Flowchart di bawah ini menggambarkan alur proses analisis sentimen yang melibatkan metode LSA dan SOM. Proses dimulai dengan pengumpulan data set yang akan dianalisis. Langkah pertama dalam proses LSA adalah menghitung TF-IDF (*Term Frequency-Inverse Document Frequency*) dari data set.

Setelah nilai TF-IDF diperoleh, dilakukan *Singular Value Decomposition* (SVD) untuk mengurangi dimensi data dan menemukan struktur latent dalam data tersebut. Data yang telah diproses dengan SVD kemudian dibagi menjadi dua bagian: data training dan data testing. Kemudian, data yang telah dibagi dua akan dilanjutkan proses pemodelan SOM.

Tahap akhir melibatkan proses testing di mana data testing digunakan untuk menguji model yang telah dilatih. Berdasarkan hasil pengujian, model mengklasifikasikan sentimen dari data testing ke dalam kategori positif, negatif, atau netral. Dengan mengikuti langkah-langkah tersebut, proses analisis sentimen dapat dilakukan secara sistematis menggunakan kombinasi metode LSA untuk ekstraksi fitur dan SOM untuk klasifikasi.



Gambar 5.1. alur proses analisis sentimen LSA + SOM

5.2. Implementasi

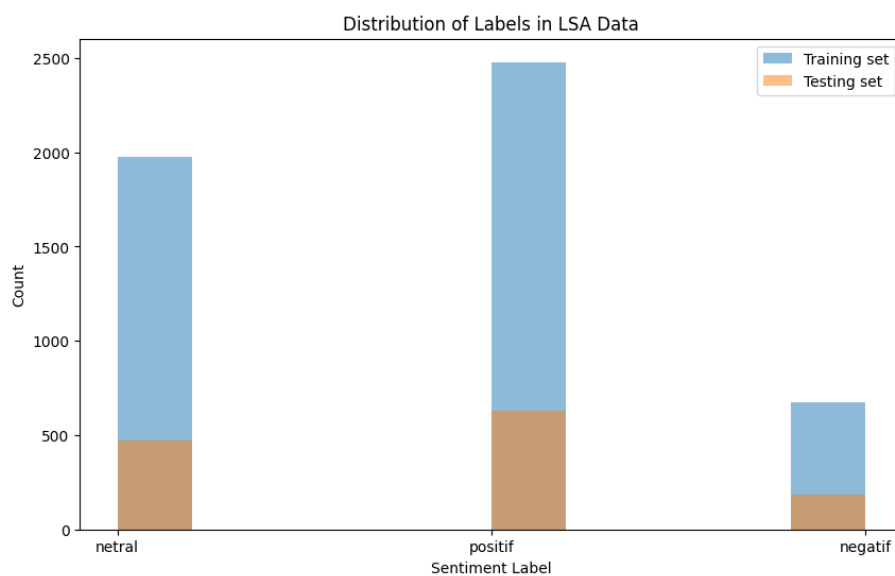
Pada tahap pembagian data, data yang telah diproses menggunakan SVD dibagi menjadi dua bagian utama: data training dan data testing. Data training terdiri dari 5.132 sampel dengan rincian 676 sampel berlabel negatif, 1.977 sampel berlabel positif, dan 2.479 sampel berlabel netral. Data training ini digunakan untuk

melatih model SOM agar dapat mengenali pola dan hubungan dalam data yang digunakan untuk klasifikasi sentimen.

Setelah proses pelatihan, data testing digunakan untuk menguji keakuratan model yang telah dilatih. Data testing terdiri dari 1.284 sampel dengan rincian 185 sampel berlabel negatif, 472 sampel berlabel positif, dan 627 sampel berlabel netral. Penggunaan data testing ini penting untuk mengevaluasi performa model dan memastikan bahwa model dapat mengklasifikasikan sentimen secara akurat pada data yang belum pernah dilihat sebelumnya. Dengan demikian, model yang dihasilkan diharapkan mampu memberikan prediksi sentimen yang andal dan konsisten. Hasil pembagian data dapat dilihat pada table dan grafik di bawah ini.

Tabel 5.1. Tabel Hasil pembagian

	Negatif	Positif	Netral	Total
Training	676	1.977	2.479	5.132
Testing	185	472	627	1.284



Gambar 5.2. Distribution of Labels in LSA Data

Setelah melalui tahap pemodelan dengan SOM, hasil evaluasi performa model diperoleh dan dirangkum dalam tabel berikut.

Tabel 5.2. Evaluasi Performa Model

	precision	recall	f1-score	support
negative	0.17	0.01	0.01	185
netral	0.57	0.51	0.54	472
positif	0.57	0.78	0.66	627
accuracy			0.53	1284
macro avg	0.44	0.43	0.40	1284
weighted avg	0.51	0.57	0.52	1284

Untuk kategori sentimen negatif, model menunjukkan precision sebesar 0.17, recall sebesar 0.01, dan f1-score sebesar 0.01 dengan dukungan 185 sampel. Ini menunjukkan bahwa model kesulitan dalam mengidentifikasi dan mengklasifikasikan sentimen negatif secara akurat, dengan sebagian besar sampel negatif tidak dikenali dengan baik oleh model.

Di sisi lain, untuk kategori sentimen netral, model menunjukkan hasil yang lebih baik dengan precision sebesar 0.57, recall sebesar 0.51, dan f1-score sebesar 0.54 berdasarkan 472 sampel. Hasil ini menunjukkan bahwa model mampu mengidentifikasi sebagian besar sampel netral dengan tingkat akurasi yang lebih baik dibandingkan kategori negatif. Meski demikian, ada ruang untuk peningkatan dalam hal recall agar lebih banyak sampel netral dapat dikenali dengan benar.

Untuk kategori sentimen positif, model menunjukkan performa yang paling baik dengan precision sebesar 0.57, recall sebesar 0.78, dan f1-score sebesar 0.66 dari 627 sampel. Ini menunjukkan bahwa model sangat efektif dalam mengklasifikasikan sentimen positif, dengan sebagian besar sampel positif dikenali

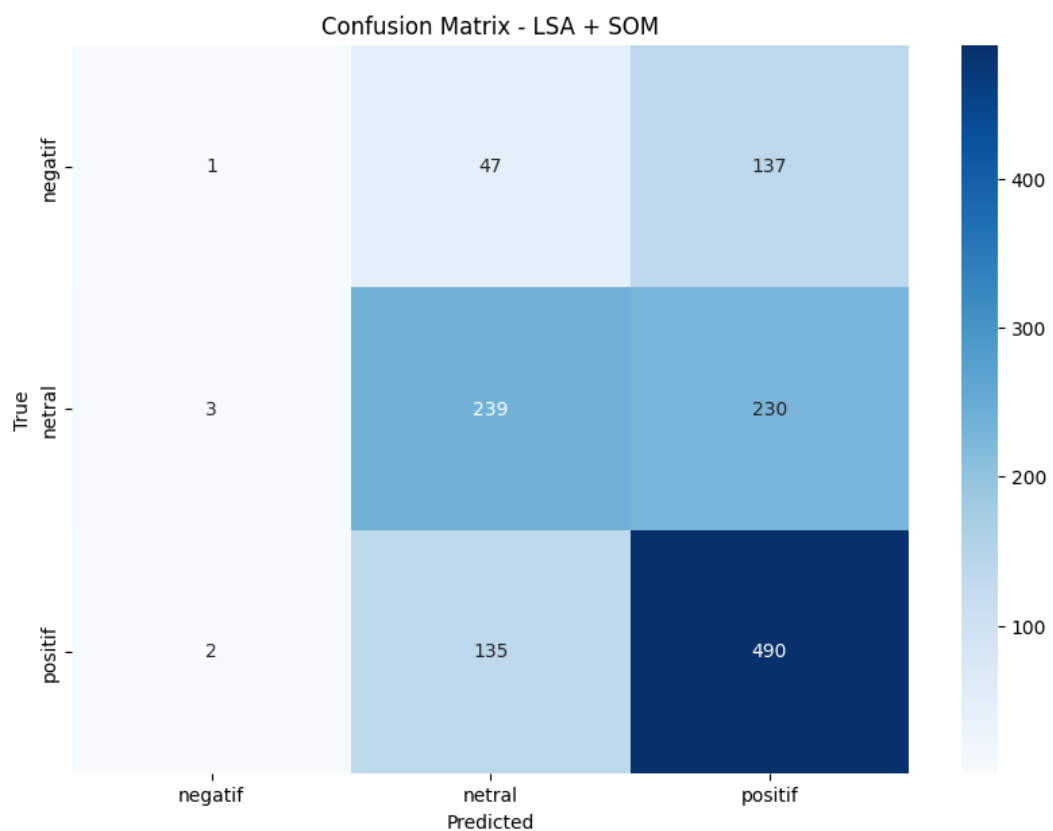
dengan benar. Tingginya nilai recall mengindikasikan bahwa model jarang melewatkan sampel positif, meskipun precision menunjukkan ada beberapa kesalahan dalam klasifikasi.

Secara keseluruhan, akurasi model mencapai 0.53, menunjukkan bahwa model berhasil mengklasifikasikan sentimen dengan benar dalam lebih dari setengah jumlah total sampel (1.284). Namun, ini juga mengindikasikan bahwa masih ada ruang yang signifikan untuk peningkatan dalam hal akurasi keseluruhan. Melihat nilai macro average, model memiliki precision sebesar 0.44, recall sebesar 0.43, dan f1-score sebesar 0.40, yang mencerminkan kinerja rata-rata model untuk setiap kategori sentimen tanpa mempertimbangkan distribusi kelas.

Terakhir, nilai weighted average menunjukkan precision sebesar 0.51, recall sebesar 0.57, dan f1-score sebesar 0.52. Ini memberikan gambaran yang lebih realistis tentang performa model dengan mempertimbangkan distribusi kelas yang tidak seimbang dalam data testing. Nilai weighted average yang lebih tinggi dibandingkan macro average mengindikasikan bahwa model performa lebih baik pada kategori dengan jumlah sampel lebih besar, seperti kategori sentimen positif dan netral. Secara keseluruhan, hasil ini menunjukkan potensi model SOM dalam analisis sentimen, namun juga menggarisbawahi perlunya perbaikan, terutama dalam mengidentifikasi sentimen negatif.

Confusion matrix di bawah ini memberikan gambaran mendetail tentang bagaimana model mengklasifikasikan setiap kategori sentimen untuk data testing. Untuk kategori negatif, dari total 185 sampel, model hanya berhasil mengklasifikasikan 1 sampel dengan benar sebagai negatif. Sebanyak 47 sampel

negatif salah diklasifikasikan sebagai netral, dan mayoritas, yakni 137 sampel negatif, salah diklasifikasikan sebagai positif. Ini menunjukkan kelemahan model dalam mengenali dan mengklasifikasikan sentimen negatif dengan benar, yang tercermin juga dalam rendahnya nilai precision, recall, dan f1-score untuk kategori ini.

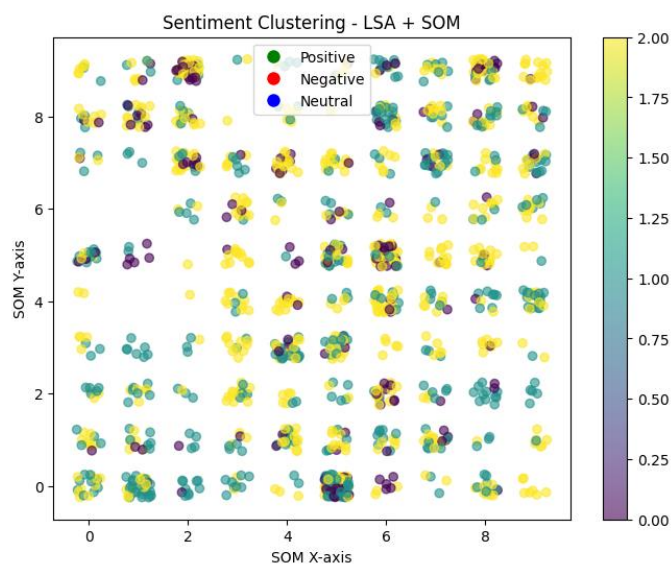


Gambar 5.3. Confusion matrix LSA+SOM

Untuk kategori netral, dari 472 sampel, 239 sampel diklasifikasikan dengan benar sebagai netral. Namun, ada 3 sampel netral yang salah diklasifikasikan sebagai negatif, dan 230 sampel netral yang salah diklasifikasikan sebagai positif. Meskipun model lebih akurat dalam mengenali sentimen netral dibandingkan sentimen negatif, masih ada tingkat kesalahan yang signifikan di mana banyak sampel netral dianggap sebagai positif. Pada kategori positif, dari 627 sampel,

model berhasil mengklasifikasikan 490 sampel dengan benar sebagai positif. Namun, 2 sampel positif salah diklasifikasikan sebagai negatif dan 135 sampel positif salah diklasifikasikan sebagai netral. Tingginya angka sampel yang diklasifikasikan dengan benar menunjukkan kemampuan model yang lebih baik dalam mengenali sentimen positif, meskipun ada kesalahan klasifikasi yang perlu diperbaiki untuk meningkatkan akurasi keseluruhan.

Di samping itu, pada grafik di bawah ini menggambarkan distribusi sampel berdasarkan sentimen positif, negatif, dan netral yang diwakili oleh titik-titik dengan warna berbeda: hijau untuk positif, merah untuk negatif, dan biru untuk netral. Sumbu X dan Y masing-masing menunjukkan koordinat pada peta SOM, sedangkan bilah warna di sisi kanan grafik menunjukkan kepadatan atau intensitas kluster dengan skala dari 0 hingga 2.



Gambar 5.4. Sentiment Clustering – LSA+SOM

Dari grafik tersebut, terlihat bahwa sampel sentimen tidak terdistribusi secara merata di seluruh peta SOM. Kluster-kluster yang mengandung sentimen

positif (hijau) cenderung lebih banyak tersebar di berbagai bagian peta, menunjukkan bahwa model mampu mengidentifikasi dan mengelompokkan sebagian besar sampel positif secara efisien. Sebaliknya, sentimen netral (biru) dan negatif (merah) terlihat lebih tersebar secara sporadis, dengan beberapa area yang menunjukkan campuran ketiga jenis sentimen. Hal ini dapat menunjukkan bahwa meskipun model mampu mengenali beberapa sentimen netral dan negatif, terdapat tingkat ketidakakuratan yang signifikan dalam mengklasifikasikan beberapa sampel ini.

Dari Gambar 5.4 juga memperlihatkan bahwa model memiliki kesulitan dalam membedakan antara sentimen netral dan positif. Hal ini sesuai dengan data yang telah disebutkan sebelumnya, di mana sejumlah besar sampel netral salah diklasifikasikan sebagai positif. Ini tercermin dalam area peta yang menunjukkan kluster campuran antara warna hijau dan biru. Untuk meningkatkan akurasi model, perlu dilakukan penyempurnaan dalam pemrosesan data atau penyesuaian algoritma klasifikasi untuk lebih baik dalam membedakan antara sentimen-sentimen yang berbeda, terutama untuk mengurangi kesalahan dalam mengklasifikasikan sampel netral sebagai positif.

BAB VI

PEMBAHASAN

6.1. Pembahasan Komparasi Performa Algoritma

Hasil perbandingan komparasi dari TF-IDF+SOM dan LSA+SOM dapat dilihat pada table berikut ini:

Tabel 6.1 Tabel Komparasi TF-IDF+LSOM & LSA+SOM

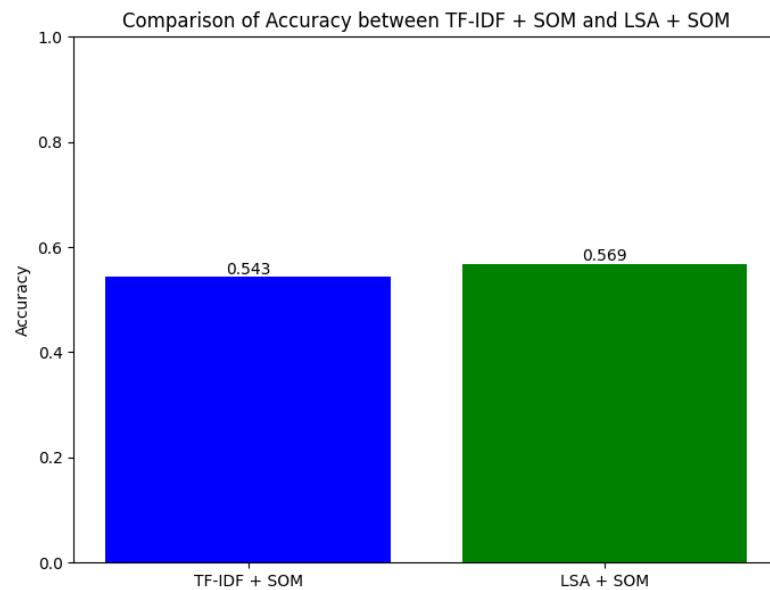
Model	Precision	Recall	F1-Score	Support
TF-IDF+SOM				
negative	0.43	0.05	0.09	185
netral	0.52	0.50	0.51	472
positif	0.56	0.72	0.63	627
accuracy			0.54	1284
macro avg	0.50	0.42	0.41	1284
weighted avg	0.53	0.54	0.51	1284
Model	Precision	Recall	F1-Score	Support
LSA+SOM				
negative	0.17	0.01	0.01	185
netral	0.57	0.51	0.54	472
positif	0.57	0.78	0.66	627
accuracy			0.53	1284
macro avg	0.44	0.43	0.40	1284
weighted avg	0.51	0.57	0.52	1284

Analisis tabel menunjukkan performa dua model teks analisis yang menggunakan kombinasi TF-IDF dan Self-Organizing Map (SOM), serta Latent Semantic Analysis (LSA) dan SOM. Model pertama, TF-IDF+SOM, menunjukkan bahwa precision untuk kelas negatif adalah 0.43, recall 0.05, dan F1-Score 0.09 dengan support 185. Untuk kelas netral, precision adalah 0.52, recall 0.50, dan F1-Score 0.51 dengan support 472. Kelas positif memiliki precision 0.56, recall 0.72, dan F1-Score 0.63 dengan support 627. Model ini memiliki akurasi keseluruhan

0.54. Rata-rata makro untuk precision, recall, dan F1-Score masing-masing adalah 0.50, 0.42, dan 0.41, sementara rata-rata tertimbang adalah 0.53, 0.54, dan 0.51.

Model kedua, LSA+SOM, menunjukkan bahwa precision untuk kelas negatif adalah 0.17, recall 0.01, dan F1-Score 0.01 dengan support 185. Untuk kelas netral, precision adalah 0.57, recall 0.51, dan F1-Score 0.54 dengan support 472. Kelas positif memiliki precision 0.57, recall 0.78, dan F1-Score 0.66 dengan support 627. Model ini memiliki akurasi keseluruhan 0.53. Rata-rata makro untuk precision, recall, dan F1-Score masing-masing adalah 0.44, 0.43, dan 0.40, sementara rata-rata tertimbang adalah 0.51, 0.57, dan 0.52.

Dari hasil analisis tersebut, terlihat bahwa kedua model memiliki kinerja yang lebih baik pada kelas positif dibandingkan dengan kelas negatif dan netral. Namun, TF-IDF+SOM menunjukkan precision yang lebih tinggi untuk kelas negatif dibandingkan dengan LSA+SOM, meskipun recall dan F1-Score untuk kelas ini sangat rendah pada kedua model. Secara keseluruhan, kedua model memiliki performa yang mirip dalam hal akurasi, namun LSA+SOM sedikit lebih baik dalam recall dan F1-Score tertimbang. Kedua model menunjukkan adanya tantangan dalam mengklasifikasikan kelas negatif dengan baik. Grafik perbandingan antara TF-IDF+SOM dan LSA+SOM dapat dilihat pada grafik berikut:



Gambar 6.1 Perbandingan Antara TF-IDF+SOM dan LSA+SOM

Gambar tersebut menampilkan perbandingan akurasi antara dua model analisis teks, yaitu TF-IDF + SOM dan LSA + SOM. Pada sumbu vertikal terdapat nilai akurasi, sedangkan pada sumbu horizontal terdapat dua kategori model yang dibandingkan. Model TF-IDF + SOM memiliki akurasi sebesar 0.543, sementara model LSA + SOM memiliki akurasi yang sedikit lebih tinggi yaitu 0.569. Perbandingan ini menunjukkan bahwa kedua model memiliki performa yang cukup dekat, namun LSA + SOM sedikit lebih unggul dalam hal akurasi.

Akurasi yang lebih tinggi pada LSA + SOM menunjukkan bahwa penggunaan LSA dapat memberikan peningkatan dalam mengidentifikasi pola dan makna dari data teks dibandingkan dengan penggunaan TF-IDF. LSA mampu menangkap hubungan laten antara kata-kata dan dokumen, yang mungkin membantu dalam meningkatkan kemampuan model untuk mengelompokkan data teks dengan lebih akurat. Hal ini terlihat dari akurasi LSA + SOM yang lebih tinggi dibandingkan dengan TF-IDF + SOM.

Namun, perbedaan akurasi antara kedua model ini tidak terlalu signifikan, yaitu hanya sekitar 0.026. Ini menunjukkan bahwa meskipun LSA + SOM memiliki keunggulan dalam hal akurasi, TF-IDF + SOM masih cukup kompetitif dan dapat digunakan sebagai alternatif tergantung pada konteks dan kebutuhan spesifik dari analisis teks yang dilakukan. Dalam pemilihan model, faktor lain seperti waktu pemrosesan, kompleksitas model, dan interpretabilitas hasil juga perlu dipertimbangkan selain dari nilai akurasi saja.

6.2. Sentimen Analisis Dalam Al-Qur'an

Berdasarkan hasil pembahasan di atas dapat dijadikan pedoman bagi pemangku kebijakan untuk menentukan sistem dekomarasi seadil mungkin dan hasil dari pengumpulan dataset twitter tersebut dapat dijadikan informasi kepada masyarakat atau publik bahwa menerima opini perlu adanya pemetaan sentimen masyarakat karena untuk memilih dan memilah berita yang bertebaran di media sosial. Hal ini sejalan dengan penyampaian suatu berita yang bersifat opini publik juga dijelaskan dalam ayat Al-Quran pada surat An-Nissa:83, sebagai berikut;

وَإِذَا جَاءَهُمْ أَمْرٌ مِّنَ الْأَمْنِ أَوْ الْخَوْفِ أَدَّعَوْا بِهٖ ۗ وَلَوْ رَدُّوهُ إِلَى الرَّسُولِ وَإِلَىٰ أُولَى الْأَمْرِ
مِنْهُمْ لَعَلِمَهُ الَّذِينَ يَسْتَنْبِطُونَهُ مِنْهُمْ ۗ وَلَوْلَا فَضْلُ اللَّهِ عَلَيْكُمْ وَرَحْمَتُهُ لَاتَّبَعْتُمُ الشَّيْطَانَ إِلَّا
قَلِيلًا

Artinya: Dan apabila datang kepada mereka suatu berita tentang keamanan ataupun ketakutan, mereka lalu menyiarkannya. Dan kalau mereka menyerahkannya kepada Rasul dan Ulil Amri di antara mereka, tentulah orang-orang yang ingin mengetahui kebenarannya (akan dapat) mengetahuinya dari mereka (Rasul dan Ulil Amri). Kalau tidaklah karena karunia dan rahmat Allah kepada kamu, tentulah kamu mengikut syaitan, kecuali sebahagian kecil saja (di antaramu). (QS. An-Nissa:83)

Sistem informasi merupakan suatu ilmu yang mempelajari bagaimana merancang, mengumpulkan, menganalisis, menginterpretasikan dan menyajikan data sehingga terdapat suatu konsep pemetaan opini masyarakat. Dengan kata lain statistika adalah suatu metode ilmiah dalam memetakan, mengumpulkan, mengelola, menganalisis, menyajikan, menafsirkan, dan menyajikan data. Sistem informasi ada untuk mendapatkan wawasan tentang kumpulan data yang sedang dipelajari sehingga dapat dipetakan dan diambil kesimpulan dari data tersebut.

Sebagai seorang mahasiswa, kita perlu melakukan pengolahan data untuk mencapai suatu kesimpulan, seperti memverifikasi hipotesis, membuat rencana atau prediksi, mengambil keputusan yang lebih baik, menghadapi berbagai perubahan, dll, dan kesimpulan tersebut perlu diambil. tepatnya agar kebenarannya dapat dikonfirmasi.

Dalam kehidupan sehari-hari, seperti sistem teknologi informasi, ekonom, manajer bisnis, dan selain bidang keuangan, manajemen, akuntansi dan bidang-bidang tersebut, teknologi informasi juga dapat digunakan dalam bidang lain, seperti pendidikan, pertanian, psikologi, teknik, politik dan lain sebagainya, dan masih banyak materi informatika yang berkaitan dengan Al-Qur'an, hal-hal yang masih banyak siswa yang belum mengetahuinya, seperti masalah pendataan yaitu pencatatan atau pencatatan data, Al-Qur'an juga berbicara . tentang hal ini dalam surat Al-Kahf ayat 49, Az-Zukhruf ayat 80, Al-Jaatsiyah ayat 29, Al-Qamar ayat 52 dan lain-lain. Selain pengumpulan data, statistik juga banyak memperhatikan keakuratan. Dalam Al-Qur'an surat Maryam ayat 94.

لَقَدْ أَخْصَلْتَهُمْ وَعَدَّاهُمْ عَدًّا ۙ

Artinya: *“Sesungguhnya Allah telah menentukan jumlah mereka dengan hitungan yang teliti...”* (QS. Maryam:94)

Al-Qur'an sendiri telah memberikan bukti nyata tentang informatika. Terdapat ketelitian dan keseimbangan dalam Al-Qur'an dalam hal jumlah penyebutan suatu kata yang berkaitan dengan sinonim, antonim, sebab, akibat atau bahkan realitas kehidupan kita sehari-hari. Banyak ulama Islam yang memelopori bahkan menciptakan banyak prinsip teknologi informasi yang masih digunakan dan diajarkan oleh setiap peradaban manusia.

Bahkan Al-Quran memperoleh keajaiban, keagungan dan keasliannya dari perhitungan dan klasifikasi yang dikandungnya. Unsur kepastian yang merupakan logika susunan dan susunan Al-Qur'an pada dasarnya menjadi bukti bahwa Al-Qur'an adalah kitab suci yang benar-benar berasal dari Tuhan dan bukan buatan manusia.

Sistem teknologi informasi merupakan ilmu yang patut kita kembangkan dan pelajari dalam rangka mengembangkan peradaban serta menumbuhkan dan meningkatkan keimanan kita terhadap kebesaran dan keagungan Allah SWT. Al-Qur'an juga mencerminkan keagungan setiap ajarannya. Dari situ kita bisa bersemangat mendalami sistem informasi. Jika kita mempunyai tujuan, mempelajari sistem informasi tentu saja merupakan salah satu bentuk ibadah, dan ada gunanya jika dilakukan. Dan, tentu saja, anggapan bahwa penelitian sistem informasi adalah ajaran sesat telah terbantahkan.

Mempelajari Al-Qur'an ibarat mempelajari alam semesta beserta isinya, tiada habisnya dan selalu memunculkan hal-hal baru. Inilah keajaiban Al-Quran, semakin dalam anda mempelajarinya maka semakin terbuka pula tabir ilmunya, membuka

cakrawala yang sangat luas bagi anda. Karena Allah SWT telah memberikan kepada kita sebagai makhluk yang sempurna alasan untuk memikirkan tanda-tanda yang ada, baik yang konkrit maupun yang abstrak, seperti dalam Al-Quran surat Ali Imron (3) ayat 190 yang berbunyi:

إِنَّ فِي خَلْقِ السَّمَوَاتِ وَالْأَرْضِ وَاخْتِلَافِ اللَّيْلِ وَالنَّهَارِ لَآيَاتٍ لِّأُولِي الْأَبْصَارِ

Artinya: “*Sesungguhnya dalam penciptaan langit dan bumi, dan silih bergantinya malam dan siang terdapat tanda-tanda bagi orang-orang yang berakal*” (QS. Ali Imron:190)

Orang memang seperti itu, mempunyai pikiran adalah suatu kerugian yang besar jika tidak dimanfaatkan dengan kemampuan terbaiknya. Oleh karena itu, mempelajari Al-Qur'an merupakan suatu keharusan untuk memahami dan menemukan hikmah serta ajaran baik yang tersirat maupun tertulis. Hal ini sesuai dengan ayat 29 Al-Qur'an Surat Shaad (38) yang berbunyi:

كِتَابٌ أَنْزَلْنَاهُ إِلَيْكَ مُبَارَكٌ لِّيَدَّبَّرُوا آيَاتِهِ وَلِيَتَذَكَّرَ أُولُوا الْأَبْصَارِ

Artinya: “*Ini adalah sebuah kitab yang Kami turunkan kepadamu penuh dengan berkah supaya mereka memperhatikan ayat-ayatnya dan supaya mendapat pelajaran orang-orang yang mempunyai fikiran.*” (QS. Shaad:29)

Jelas sekali bagi masyarakat bahwa pemahaman Al-Qur'an diperlukan untuk menyebarkan ilmu pengetahuan yang terkandung dalam isi Al-Qur'an itu sendiri. Bukti di atas dapat menggugah minat para ulama untuk meningkatkan ghiroh (semangat) mempelajari Al-Qur'an hingga Allah SWT berfirman dalam surat Al-Hijr (15) ayat 21:

وَأَنْ مِنْ شَيْءٍ إِلَّا عِنْدَنَا خَزَائِنُهُ وَمَا نُنزِّلُهُ إِلَّا بِقَدَرٍ مَّعْلُومٍ ﴿٢١﴾

Artinya: “Dan tidak ada sesuatupun melainkan pada sisi Kami-lah khazanahnya; dan Kami tidak menurunkannya melainkan dengan ukuran yang tertentu.” (QS. Al-Hijr:21)

Dari ayat di atas jelas bahwa Al-Quran harus dipelajari dari sudut pandang sistem informasi karena berkaitan dengan pengukuran tertentu. Hal ini sesuai dengan firman Allah SWT dalam surat Al-Qomar (54) ayat 49 yang berbunyi:

إِنَّا كُلَّ شَيْءٍ خَلَقْنَاهُ بِقَدَرٍ ﴿٤٩﴾

Artinya: “Sesungguhnya Kami menciptakan segala sesuatu menurut ukuran.” (QS. Al-Qomar:49)

Tidak dapat dipungkiri lagi, bahwa Al-Qur’an mempunyai pesan-pesan yang perlu dikaji dalam berbagai aspek keilmuan, khususnya sistem informasi.

BAB VII

KESIMPULAN

7.1. Kesimpulan

1. Model LSA + SOM memiliki akurasi yang lebih tinggi (0.569) dibandingkan dengan model TF-IDF + SOM (0.543), menunjukkan bahwa LSA memberikan peningkatan performa dalam analisis teks.
2. Perbedaan akurasi antara kedua model hanya sebesar 0.026, yang menunjukkan bahwa kedua metode memiliki performa yang relatif dekat.
3. LSA mampu menangkap hubungan laten antara kata-kata dan dokumen lebih baik daripada TF-IDF, yang mungkin berkontribusi pada peningkatan akurasi dalam model LSA + SOM.
4. Meskipun LSA + SOM menunjukkan akurasi yang lebih tinggi, TF-IDF + SOM tetap merupakan alternatif yang kompetitif, tergantung pada kebutuhan spesifik dari analisis teks yang dilakukan.
5. Selain akurasi, faktor lain seperti waktu pemrosesan, kompleksitas model, dan interpretabilitas hasil juga penting untuk dipertimbangkan dalam pemilihan model.

7.2. Saran

Berdasarkan hasil analisis dan kesimpulan yang diperoleh, model LSA + SOM direkomendasikan untuk dipilih sebagai analisis tekstual, karena akurasinya sedikit lebih tinggi dibandingkan TF-IDF + SOM. Namun, mengingat perbedaan akurasi yang tidak signifikan, pengguna juga dapat mempertimbangkan TF-IDF + SOM sebagai alternatif, terutama jika ada kendala terkait waktu pemrosesan atau

kompleksitas implementasi. Selain itu, penting untuk melakukan eksperimen lebih lanjut dengan berbagai parameter dan dataset yang lebih besar dan beragam untuk memastikan generalisasi hasil yang lebih baik. Evaluasi tambahan yang mencakup metrik lain seperti precision, recall, dan F1-Score untuk masing-masing kelas juga akan memberikan gambaran yang lebih lengkap tentang performa model.

DAFTAR PUSTAKA

- Abdillah, Allif Rizki, dan Firman Noor Hasan. 2023. "Analisis Sentimen Terhadap Kandidat Calon Presiden Berdasarkan Tweets Di Sosial Media Menggunakan Naive Bayes Classifier." *SMATIKA JURNAL: STIKI Informatika Jurnal* 13(01):117–30.
- Afrizal, Sarika, Helena Nurramdhani Irmanda, Noor Falih, dan Ika Nurlaili Isnainiyah. 2019. "Implementasi Metode Naive Bayes untuk Analisis Sentimen Warga Jakarta Terhadap." *Informatik: Jurnal Ilmu Komputer* 15(3):157–66.
- Alkhalidi, Muhammad Waly, Berto Nadeak, dan Muhammad Sayuthi. 2020. "Sistem informasi geografis pemetaan wilayah penyalahgunaan narkoba menggunakan metode SOM (Self-Organizing Map) studi kasus: kabupaten aceh tenggara." *Building of Informatics, Technology and Science (BITS)* 2(1):1–9.
- Anggadwita, Grisna, Nikita Aida Alfisyahrina, dan Astri Ghina. 2021. "Pemetaan Komunikasi Walikota Kediri Melalui Instagram Terkait Dukungan Pemerintah Terhadap UMKM (Periode 20 Agustus 2019–20 November 2020)." *Journal of Integrated System* 4(2):115–32.
- Anggraini, Novita, Edi Surya Negara Harahap, dan Tri Basuki Kurniawan. 2021. "Text Mining-Analisis Teks Terkait Isu Vaksinasi COVID-19 (Text Mining-Text Analysis Related to COVID-19 Vaccination Issues)." *JURNAL IPTEKKOM Jurnal Ilmu Pengetahuan & Teknologi Informasi* 23(2):141–53.
- Ardiansyah, Rizka. 2019. "Analisis sentimen calon presiden dan wakil presiden periode 2019-2024 pasca debat pilpres di Twitter." *ScientiCO: Computer Science and Informatics Journal* 2(1):21–28.
- Buntoro, Ghulam Asrofi. 2017. "Analisis Sentimen Calon Gubernur DKI Jakarta 2017 Di Twitter." *INTEGER: Journal of Information Technology* 2(1).
- Damayanti, Damayanti, Dendy Indriya Efendi, Dodi Solihudin, Cep Lukman Rohmat, dan Sandy Eka Permana. 2024. "PEMETAAN OPINI PUBLIK TERHADAP PERUBAHAN KEBIJAKAN BPJS KESEHATAN DENGAN PENDEKATAN SUPPORT VECTOR MACHINE (SVM) DALAM ANALISIS SENTIMEN." *JATI (Jurnal Mahasiswa Teknik Informatika)* 8(1):88–94.
- Dananjaya, Adrianus Bagas Tanty, dan Mohammad Isa Irawan. 2023. "Analisis Sentimen pada Komentar terhadap Kebijakan Perjalanan Domestik yang Dikelompokkan Menggunakan Metode Self-Organizing Maps." *Jurnal Sains dan Seni ITS* 12(1):A54–59.

- Dewi, Anggi Riantika, Sri Diana, Moch Alvi Fakhrezi, Nana Awang, Helmi Ma'arif, dan Dedi Dwi Saputra. 2023. "Sentimen Analisis Terhadap Puan Maharani Sebagai Kandidat Calon Presiden 2024 Berdasarkan Opini Twitter Menggunakan Metode Naive Bayes Dan Adaboost." *JSiI (Jurnal Sistem Informasi)* 10(1):75–80.
- Fathir, Fathir. 2023. "Analisis sentimen artikel berita pemilu berbasis metode klasifikasi."
- Fiqri, Miftahul, Indriati Indriati, dan Rizal Setya Perdana. 2023. "Klasifikasi Data Twitter pada Masa Transisi Pandemi menuju Endemi menggunakan Latent Semantic Analysis (LSA)." *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer* 7(6):2736–42.
- Firdlous, Dimas Anugrah. 2023. "Analisis sentimen publik twitter terhadap pemilu 2024 menggunakan model long short term memory (lstm)."
- Ghozali, Muhammad Imam, dan Wibowo Harry Sugiharto. 2017. "Analisa Pola Belanja Menggunakan Algoritma Fp Growth, Self Organizing Map (Som) Dan K Medoids." *Simetris: Jurnal Teknik Mesin, Elektro dan Ilmu Komputer* 8(1):317–26.
- Hakimi, F. Darwis Dzikril. 2018. "Sistem Analisis Sentimen Publik Tentang Opini Pemilihan Kepala Daerah Jawa Timur 2018 Pada Dokumen Twitter Menggunakan Naive Bayes Classifier." *Universitas Islam Negeri Sunan Ampel Surabaya*.
- Hidayat, Hety Handayani, Ardiansyah Ardiansyah, Poppy Arsil, dan Laras Isna Rahmawati. 2021. "Pemetaan Kata Kunci dan Polaritas Sentimen Pengguna Twitter Terhadap Kehalalan Produk." *MATRIK: Jurnal Manajemen, Teknik Informatika dan Rekayasa Komputer* 21(1):1–10.
- Hilmi, Muh Najib, Yuciana Wilandari, dan Hasbi Yasin. 2015. "Pemetaan Preferensi Mahasiswa Baru dalam Memilih Jurusan Menggunakan Artificial Neural Network (ANN) dengan Algoritma Self Organizing Maps (SOM)." *Jurnal Gaussian* 4(1):53–60.
- Junaidi, Junaidi, Fifit Alfiah, Ely Susanti, Juita Kristinna, Olga Riski Ardiansyah, Dhimas Pradipta, dan Wulaningsih Wulaningsih. 2015. "Manfaat Menganalisis Pengaruh Sosial Media Facebook Terhadap Kampanye Partai Politik Di Indonesia." *SEMNASSTEKNOMEDIA ONLINE* 3(1):4–5.
- Manullang, Oktaviami, dan Cahyo Prianto. 2023. "Analisis Sentimen dalam Memprediksi Hasil Pemilu Presiden dan Wakil Presiden: Systematic Literature Review." *Jurnal Informatika dan Teknologi Komputer (J-ICOM)* 4(2):104–13.

- Nuansa, Eza Putra. 2017. "Analsis Sentimen Pengguna Twitter Terhadap Pemilihan Gubernur Dki Jakarta Dengan Metode Naïve Bayesian Classification Dan Support Vector Machine." *Surabaya: Institut Teknologi Sepuluh Nopember*.
- Overbeek, Marlinda Vasty. 2016. "SORTASI TEKSTUR BIJI JAGUNG SEBAGAI BENIH TANAM MENGGUNAKAN SELF ORGANIZING MAP (Studi Kasus: Desa Bismarak Kabupaten Kupang Timur Provinsi Nusa Tenggara Timur)." Hlm. 284–88 dalam.
- Rahmah, Devi Amelia. 2022. "Analisis klaster berdasarkan indikator kesejahteraan rakyat menggunakan metode Self Organizing Maps (SOM)."
- Santosa, Agung Dwi. 2013. "Perancangan Sistem Pendukung Keputusan Pemetaan Profil Kecamatan di Kabupaten Sragen Berdasarkan Indeks Pembangunan Manusia Menggunakan Algoritma Self Organizing Maps (SOM)."
- Situmorang, Riwanti, Uus Muhammad Husni Tamyis, dan Lise Sri Andar Muni. 2023. "ANALISIS SENTIMEN DESTINASI WISATA DI JAWABARAT PADA TWITTER MENGGUNAKAN ALGORITMA NAIVE BAYES CLASSIFIER." *Simtek: jurnal sistem informasi dan teknik komputer* 8(2):339–42.
- Taralandu, Virgelius Hendrawan, dan Magdalena A. Ineke Pakereng. 2023. "Analisis Sentimen Media Sosial Twitter Terhadap Tragedi Kanjuruhan." *Jutisi: Jurnal Ilmiah Teknik Informatika dan Sistem Informasi* 12(1).
- Yusuf, Muhammad. 2022. "Analisis Sentimen Data Twitter Terhadap Bakal Calon Presiden Republik Indonesia 2024 Dengan Metode Backpropagation."
- Yusuf, Safier, Mochammad Ali Fauzi, dan Komang Candra Brata. 2018. "Sistem Temu Kembali Informasi Pasal-Pasal KUHP (Kitab Undang-Undang Hukum Pidana) Berbasis Android Menggunakan Metode Synonym Recognition dan Cosine Similarity." *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer* 2(2):838–47.