

Sistem Analisa Sentiment Bakal Calon Presiden 2024 Menggunakan Metode NLP Berbasis Web

The Sentiment Analysis System For the 2024 Presidential Candidates Uses Web-Based NLP Method

Abdul Mukti¹, Anggun Dwi Hadiyanti², Airlangga Nurlaela³, Jeremy Panjaitan⁴

¹Magister Sistem Informasi
STMIK Likmi, Bandung,
Indonesia
moektie@gmail.com

²Magister Sistem Informasi
STMIK Likmi, Bandung,
Indonesia
anggundh@gmail.com

³Magister Sistem Informasi
STMIK Likmi, Bandung,
Indonesia
airlangga.elanx@gmail.com

⁴Magister Sistem Informasi
STMIK Likmi, Bandung,
Indonesia
jeremypanjaitan@gmail.com

ABSTRACT

The spread of forecasts for presidential candidates in digital media, especially articles, has created public interest in the popularity, politics and performance of the presidential candidate, which has the effect of leading public opinion to positive or negative sides depending on people's assessments, arguments and information conveyed. In increasing sources of accurate information on the election of presidential candidates, an information system will be built that will assist the public in objectively selecting presidential candidates. The word processing method will be processed and managed by Natural Language Processing (NLP) using the Naïve Baiyes Classifier Algorithm by comparing the KBBI (Big Indonesian Dictionary) vocabulary so that conclusions can be drawn about people's emotions (positive, negative and even neutral) and will present information in the form of charts, trends and content.

Keywords : *sentiment analysis, natural language processing (NLP) and naïve baiyes classifier*

1. Pendahuluan

Pada Undang-Undang Republik Indonesia No. 7 Tahun 2017 tentang Pemilihan Umum BAB II pasal 169-170 menyatakan bahwa salah satu persyaratan untuk mencalonkan diri menjadi calon presiden dan calon wakil presiden, salah satunya adalah jika seseorang yang memiliki jabatan sebagai Pejabat negara yang dicalonkan oleh partai politik peserta Pemilu atau Gabungan Partai Politik sebagai calon Presiden atau calon Wakil Presiden harus mengundurkan diri dari jabatannya, kecuali Presiden, Wakil Presiden, Pimpinan dan anggota MPR, Pimpinan dan anggota DPR, pimpinan dan anggota DPD, gubernur, wakil gubernur, bupati, wakil bupati, walikota, dan wakil walikota.

Dilansir dari artikel kompas.com dengan judul Survei Nama-nama Capres Potensial di 2024 sebagai berikut lima nama teratas berdasarkan nilai elektabilitas, yaitu:

1. Ganjar Pranowo: 26,6 persen
2. Prabowo Subianto: 19,7 persen
3. Anies Baswedan: 17,7 persen

Tersebar nya peramalan akan bakal calon Presiden di digital media khususnya artikel, menyebabkan menimbulkan daya tarik masyarakat terhadap popularitas, politik dan kinerja dari calon Presiden tersebut, yang berimbas menggiring opini publik terhadap sisi positif maupun negatif tergantung dari penilaian masyarakat, argumen dan informasi yang disampaikan.

Dalam meningkatkan sumber informasi yang akurat terhadap pemilihan calon Presiden, akan dibangun sebuah sistem informasi yang akan membantu masyarakat dalam pemilihan calon presiden secara objektif. Maka dari itu akan dilakukan *data crawling* dari beberapa artikel dengan proses

pengambilan data selama 3 bulan dan menggunakan *sentiment analysis* untuk dapat memahami dan mengelompokkan emosi, baik positif, negatif maupun netral dengan menganalisa teks menggunakan kosa kata. Bahasa dalam kosa kata yang terkandung dalam teks, akan diolah dan dikelola oleh *Natural Language Processing* (NLP) untuk di dapatkan kesimpulan tentang emosi masyarakat (positif, negatif bahkan netral) dengan membandingkan dengan kosa kata KBBI (Kamus Besar Bahasa Indonesia) dan akan menyajikan informasi berupa *chart*, tren maupun konten.

Berdasarkan uraian tersebut, maka akan dibahas permasalahan yang akan dituangkan dalam bentuk penelitian yang berjudul “Sistem Analisa sentiment bakal calon presiden 2024 menggunakan metode NLP berbasis web”.

2. Kajian Pustaka

Analisis Sentimen

Analisis sentimen adalah proses untuk mengidentifikasi, mengekstrak, dan menilai sentimen atau opini yang terkandung dalam teks. Tujuan dari analisis sentimen adalah untuk memahami perasaan, sikap, atau evaluasi yang dinyatakan oleh penulis atau pengguna teks, terutama dalam konteks sosial media, ulasan produk, komentar pelanggan, dan sumber teks lainnya.

Berikut adalah beberapa teori dan metode yang digunakan dalam analisis sentimen:

1. Analisis Sentimen Berbasis Aturan (Rule-Based Sentiment Analysis):

Metode ini menggunakan kumpulan aturan linguistik atau leksikon yang telah ditentukan sebelumnya untuk mengklasifikasikan sentimen. Aturan-aturan ini mengidentifikasi kata-kata kunci atau frase yang berhubungan dengan sentimen positif atau negatif. Contoh leksikon yang sering digunakan adalah SentiWordNet dan Linguistic Inquiry and Word Count (LIWC).

2. Analisis Sentimen Berbasis Mesin Pembelajaran (Machine Learning-Based Sentiment Analysis):

Metode ini melibatkan pembelajaran mesin untuk mengklasifikasikan teks ke dalam kategori sentimen tertentu (misalnya, positif, negatif, atau netral). Biasanya, pendekatan ini melibatkan teknik seperti pembelajaran berbasis aturan, analisis teks, dan pemodelan bahasa. Algoritma pembelajaran mesin yang sering digunakan dalam analisis sentimen termasuk Naive Bayes, Support Vector Machines (SVM), dan algoritma pengklasifikasi berbasis jaringan saraf seperti Jaringan Saraf Konvolusional (CNN) dan Jaringan Saraf Rekurensi (RNN).

3. Analisis Sentimen Berbasis Aspek (Aspect-Based Sentiment Analysis):

Metode ini fokus pada pemahaman sentimen terkait dengan aspek atau entitas tertentu dalam teks, seperti fitur produk atau topik khusus. Pendekatan ini memungkinkan analisis yang lebih mendalam dan detail tentang sentimen terkait aspek tertentu dalam teks. Teknik-teknik seperti pemisahan kalimat, klasifikasi berbasis aturan, dan pengenalan entitas bernama (NER) dapat digunakan dalam analisis sentimen berbasis aspek.

Natural Language Processing

Natural Language Processing (NLP) adalah cabang ilmu komputer yang berkaitan dengan pemahaman dan pengolahan bahasa manusia oleh mesin. Tujuan utama NLP adalah memungkinkan komputer untuk berinteraksi dengan manusia dalam bahasa alami dan memproses, menganalisis, dan menghasilkan teks atau ujaran dalam bahasa manusia. Berikut adalah beberapa teori dan pendekatan yang digunakan dalam Natural Language Processing:

1. Pemrosesan Bahasa Alami Berbasis Aturan (Rule-Based Natural Language Processing):

Metode ini melibatkan pembuatan dan penggunaan aturan linguistik atau tata bahasa untuk memahami dan memproses teks. Aturan ini ditentukan sebelumnya dan mencakup sintaksis, semantik, dan struktur bahasa. Beberapa teknik yang digunakan dalam pemrosesan bahasa alami berbasis aturan termasuk pengurai sintaksis (parser) dan generator bahasa alami.

2. Pembelajaran Mesin dalam Pemrosesan Bahasa Alami (Machine Learning in Natural Language Processing):

Pendekatan ini melibatkan penggunaan algoritma pembelajaran mesin untuk mempelajari pola dari data bahasa alami yang telah diberikan. Algoritma ini dapat digunakan untuk tugas-tugas seperti klasifikasi teks, ekstraksi informasi, pemodelan topik, dan terjemahan mesin. Metode pembelajaran mesin yang sering digunakan dalam NLP termasuk Naive Bayes, Support Vector Machines (SVM), Jaringan Saraf, dan Model Bahasa.

3. Representasi Bahasa Alami (Natural Language Representation):

Representasi bahasa alami adalah cara mengkodekan dan mengubah teks menjadi bentuk yang dapat dipahami oleh komputer. Representasi ini mencakup representasi kata, frase, dan dokumen, serta representasi grafis untuk analisis sintaksis dan semantik. Beberapa metode populer dalam representasi bahasa alami termasuk vektor kata (word embeddings), model bahasa berbasis Transformer, dan grafik dependensi.

Aplikasi Berbasis Web

Aplikasi berbasis web adalah jenis aplikasi perangkat lunak yang dijalankan melalui browser web dan diakses melalui internet. Aplikasi ini dapat digunakan oleh pengguna dari berbagai perangkat dan lokasi, dengan mengaksesnya melalui URL yang ditentukan. Aplikasi web dapat menampilkan informasi, mengumpulkan data dari pengguna, berinteraksi dengan pengguna, dan melakukan berbagai tugas sesuai dengan kebutuhan. Berikut adalah beberapa teori dan konsep yang berkaitan dengan aplikasi berbasis web:

1. Arsitektur Client-Server:

Arsitektur client-server adalah pendekatan di mana aplikasi terbagi menjadi dua bagian: sisi klien (client-side) dan sisi server (server-side). Sisi klien berjalan di perangkat pengguna dan bertanggung jawab untuk tampilan dan interaksi antarmuka pengguna, sedangkan sisi server berjalan di server dan bertanggung jawab untuk pemrosesan logika bisnis, pengolahan data, dan penyimpanan.

2. Teknologi Web:

Teknologi web mencakup bahasa pemrograman, protokol, dan standar yang digunakan untuk mengembangkan aplikasi web. Ini termasuk HTML (Hypertext Markup Language) untuk struktur halaman web, CSS (Cascading Style Sheets) untuk tampilan dan desain, JavaScript untuk interaksi dan logika di sisi klien, dan berbagai protokol seperti HTTP (Hypertext Transfer Protocol) untuk komunikasi antara klien dan server.

3. Basis Data dan Penyimpanan Data:

Aplikasi web sering membutuhkan penyimpanan data yang dapat diakses dan dikelola secara efisien. Basis data relasional (misalnya, MySQL, PostgreSQL) atau basis data NoSQL (misalnya, MongoDB, Cassandra) sering digunakan untuk menyimpan data aplikasi web. Konsep-konsep seperti pengaturan skema basis data, kueri data, dan integrasi basis data dengan aplikasi web penting dalam pengembangan aplikasi berbasis web.

4. Keamanan dan Autentikasi:

Keamanan aplikasi web adalah faktor penting untuk melindungi data pengguna dan menjaga kerahasiaan serta integritas aplikasi. Autentikasi pengguna, otorisasi akses, enkripsi data, dan perlindungan terhadap serangan keamanan seperti serangan injeksi (injection attacks) adalah beberapa konsep penting dalam pengembangan aplikasi berbasis web yang aman.

Algoritma Naïve Baiyes

Algoritma Naive Bayes adalah sebuah algoritma klasifikasi yang didasarkan pada Teorema Bayes dengan asumsi bahwa fitur-fitur (variabel independen) dalam data pelatihan saling independen. Algoritma ini populer dalam analisis teks dan pengenalan pola, serta digunakan secara luas dalam analisis sentimen, klasifikasi dokumen, deteksi spam, dan banyak aplikasi lainnya. Berikut adalah teori dasar tentang algoritma Naive Bayes:

1. Teorema Bayes:

Algoritma Naive Bayes didasarkan pada Teorema Bayes, yang menyediakan cara untuk memperbarui probabilitas suatu kejadian dengan memperhitungkan bukti baru. Dalam konteks klasifikasi, Teorema Bayes digunakan untuk menghitung probabilitas kelas target (output) berdasarkan fitur-fitur (input) yang diamati.

2. Asumsi Naive Bayes:

Algoritma Naive Bayes mengasumsikan bahwa semua fitur yang diamati dalam data pelatihan saling independen, artinya tidak ada ketergantungan langsung antara fitur-fitur tersebut. Meskipun asumsi ini jarang terpenuhi sepenuhnya dalam data nyata, algoritma Naive Bayes masih sering memberikan hasil yang baik dan efisien dalam banyak kasus.

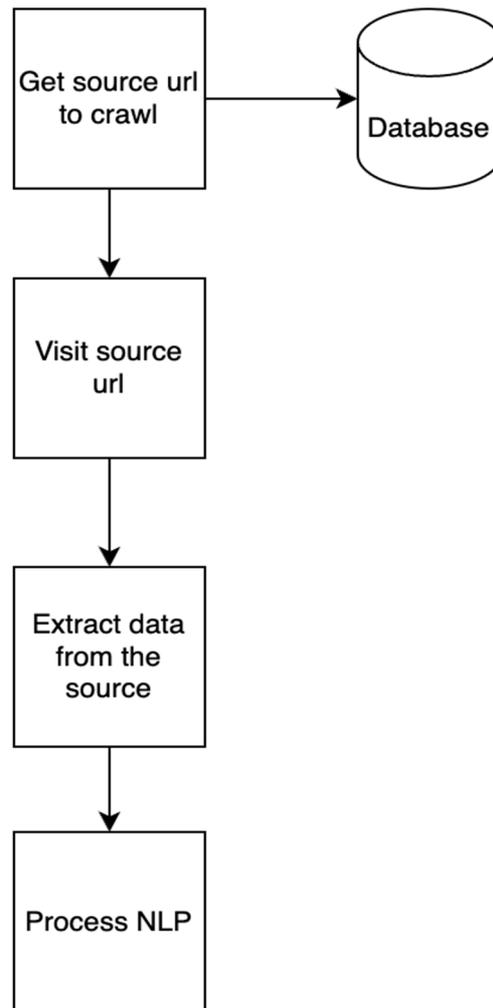
3. Probabilitas Kondisional:

Algoritma Naive Bayes menggunakan probabilitas kondisional untuk menghitung probabilitas kelas target berdasarkan fitur-fitur yang diamati. Probabilitas kondisional adalah probabilitas suatu kejadian terjadi, dengan mempertimbangkan kejadian lain yang terjadi bersamaan.

4. Model Multinomial dan Model Gaussian:

Algoritma Naive Bayes umumnya mengadopsi dua model probabilitas: Model Multinomial dan Model Gaussian. Model Multinomial digunakan ketika fitur-fitur yang diamati adalah diskrit atau terhitung, seperti dalam analisis sentimen. Model Gaussian digunakan ketika fitur-fitur yang diamati memiliki distribusi Gaussian (normal), seperti dalam klasifikasi berbasis angka atau data kontinu.

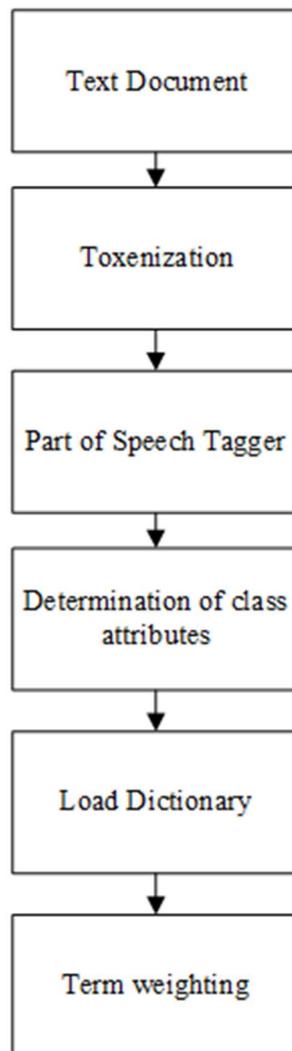
3. Metode Penelitian Data Crawling



Gambar 1. Data Crawling

Dari 2 sumber artikel seperti kompas dan detik.com, proses *crawling* dimulai dengan memberikan daftar URL yang sudah diketahui *crawler*, akan mengambil halaman web dari daftar URL. Selama proses ini, biasanya *crawler* menemukan tautan ke URL lain. *Crawler* akan berhenti untuk menemukan tautan ke URL yang lain apabila sudah mencapai kedalaman yang telah ditentukan. Setelah proses sebelumnya selesai, maka akan masuk ke daftar halaman lain untuk di *crawling* setelahnya. Data *crawling* tersebut akan dikelola oleh *database* untuk mengetahui apakah tautan sudah pernah dikunjungi atau belum dikunjungi. Kemudian, jika cocok dengan ekstraksi tautan akan diambil URL dan kontennya. Maka akan menyaring catatan yang akan digunakan maupun dihapus. Proses ini akan berlangsung secara berulang, terus menerus, dan terjadwal sesuai dengan pengaturan di *windows task scheduler*.

NLP (*Natural Language Processing*)



Gambar 2. NLP (*Natural Language Processing*)

Di dalam proses NLP (*Natural Language Processing*), untuk dapat menghasilkan *content data* yang diberi nilai *sentiment*, dilakukan beberapa proses, yaitu :

Tokenizing

Proses NLP (*Natural Language Processing*) yang selanjutnya adalah *Tokenizing*, yaitu proses memotong setiap kata didalam teks dan mengkonversi huruf didalam dokumen menjadi huruf kecil. Didalam proses ini, hanya huruf yang dapat diterima, karakter khusus atau tanda baca dihilangkan. Oleh karena itu hasil dari proses *tokenization* adalah kata-kata yang membentuk sebuah kalimat atau *string* tanpa ada tanda baca.

Part of speech tagger

POS tagger adalah sebuah proses yang dapat memberikan kelas pada sebuah kata. Dalam proses POS tagger dilakukan dengan cara *parsing*, setelah itu ditentukan kelas setiap kata dengan menggunakan bantuan kamus yang dibuat sendiri berdasarkan Kamus Besar Bahasa Indonesia (KBBI).

Proses POS tagging terdapat tiga proses yaitu pemisahan, pengecekan setiap kata dalam dokumen, dan melakukan identifikasi bentuk imbuhan dan akhiran yang menghasilkan kata dasar. Penentuan *sentiment* dengan *polarity positif* maupun *negative* dari artikel yang sudah diberikan label kata. Kelas kata yang dipilih menggunakan kata predikat.

Determination of Class Attributes

Setelah dilakukan *Preprocessing*, dari data artikel akan ditentukan *class attribute*. *Class attribute* yang dibahas pada penelitian ini, diantaranya positif, negatif, dan netral. Dengan adanya *class attribute* ini diharapkan dapat memberikan penilaian kepada masyarakat secara akurat terhadap suatu objek tertentu.

Load Dictionary

Load dictionary merupakan proses pemanggilan kamus, dengan menggunakan kata kunci yang dapat menunjukkan sentimen positif, negatif, netral, atau menelusuri bahasa gaul pada frasa. Proses *load dictionary* dilakukan agar dapat melakukan klasifikasi kata sesuai dengan makna aslinya dari kata yang dimaksud.

1. Kamus kata kunci positif: cerdas, baik, rajin.
2. Kamus kata kunci negatif: penipu, korupsi, lambat.
3. Kamus netral: sama, tidak, bukan.
4. Kamus konversi bahasa gaul ke KBBI: dmn = dimana, spt = seperti

Term Weighting

Setelah *semua* kata diklasifikasikan ke dalam kategori kata positif, negatif, dan netral. Kemudian akan dilakukan proses penghitungan dengan melakukan akumulasi bobot setiap kata dari sebuah kalimat opini. Saat nilai yang didapatkan dari kalimat opini tersebut = 1, maka nilai *sentiment* dari kalimatnya adalah positif. Ketika nilai opini = 0, maka nilai *sentimen* dari kalimatnya adalah netral. Dan nilai opini dalam kalimat tersebut adalah -1, maka nilai *sentiment* dari kalimatnya adalah negatif.

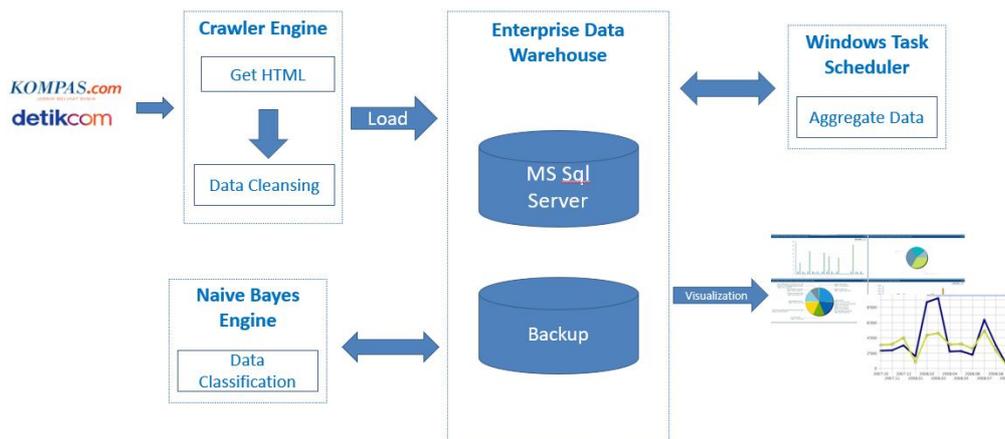
Naïve Baiyes Clasifier

Naïve Bayes Classifier adalah metode klasifikasi dengan probabilitas sederhana yang mengaplikasikan Teorema Bayes dengan asumsi ketidaktergantungan (independen) yang tinggi. Penggunaan metode Naive Bayes Classifier, berdasarkan pada jumlah dataset yang digunakan sehingga memerlukan sebuah metode yang memiliki performansi yang cepat dalam pengklasifikasian dan keakuratan. Kelebihan dari penggunaan Naive Bayes Classifier adalah metode ini menggunakan jumlah data pelatihan (*training data*) yang kecil untuk menentukan estimasi parameter yang diperlukan dalam proses pengklasifikasian.

4. Hasil dan Pembahasan

Crawling artikel dilakukan terhadap 2 sumber digital media mainstream di Indonesia, kompas.com dan detik.com. Mesin yang digunakan untuk crawling data artikel, adalah dengan sistem custom yang telah dibentuk khusus untuk melakukan pengambilan data artikel tersebut.

Adapun tahapan crawling mulai dari pencarian artikel hingga mengklasifikasikan polaritas artikel tersebut dengan metode naive bayes serta visualisasi hasil yang diberikan:



Gambar 3. Crawling artikel

1. Crawler Engine

Pada bagian ini, aplikasi custom mengambil data publik 2 digital media mainstream di Indonesia, kompas.com dan detik.com. Data HTML yang diperoleh memiliki banyak distorsi seperti iklan, layanan berbaris, banner pemasaran dan lain sebagainya. Pembersihan pun dilakukan agar mendapatkan konten utama dalam badan (*body*) html tersebut. Pengambilan tanggal diterbitkannya artikel tersebut, diekstraksi dari lokasi yang sudah ditentukan oleh struktur *website digital media* bersangkutan. Setiap *website digital media* memiliki lokasi penempatan tanggal penerbitan yang berbeda-beda. Ada yang ditanam dalam URL adapun yang disisipkan pada konten artikel dalam *body* html. Sehingga, perilaku, struktur dan karakteristik setiap digital media ditangani dengan proses ekstraksi data yang spesifik. Waktu pengambilan crawling dilakukan dalam selang waktu 3 bulan mulai dari 1 Maret 2023 hingga 31 Mei 2023.

2. Enterprise Data Warehouse

Penyimpanan data *crawling* menggunakan *Microsoft SQL server* dengan sistem *backup data* setiap harinya pada jam 23:59 setiap malamnya. Data yang disimpan mencakup *field* seperti judul artikel, konten artikel, tanggal penerbitan, tanggal *crawling* dan alamat URL sumber artikel. Setiap data artikel yang disimpan, akan dikodekan dengan sistem *hashtext* dimana setiap artikelnnya unik antara satu dengan yang lainnya. Pada saat mesin *crawling* mendapati data artikel yang sama dengan hasil *hashtext* yang diberikan, *database* akan menolak menyimpan data artikel tersebut. Hal ini dilakukan untuk menghindari redundansi data.

3. Naive Bayes Engine

Pada tahap ini dilakukan 2 tugas utama untuk menghasilkan polaritas suatu artikel:

- Text Extraction*, ekstraksi dari konten artikel berdasarkan klasifikasi kata kunci yang diberikan. Dalam penelitian ini kata kunci tersebut adalah 3 calon kandidat Ganjar Pranowo, Prabowo Subianto dan Anies Baswedan. *Text extraction* dijalankan menggunakan salah satu fitur *SQL Server Database Engine* yang dinamakan *Full-Text Search*. *Full-Text Search* akan melakukan *query* terhadap data artikel yang sudah tersimpan pada *database* berdasarkan klasifikasi kata kunci yang telah ditentukan dengan memanfaatkan memory RAM pada komputer kita maupun *server* sehingga terjadi *scanning* dan *indexing data* yang lebih cepat dan hasil yang akurat.
- NLP, dengan metode *naive bayes* dalam menentukan polaritas data yang telah di ekstrasi sebelumnya. Nilai polaritas terdiri dari sentimen *negative*, *positive* maupun netral. Setiap artikel yang melalui proses *text extraction* dan NLP akan ditandai dan di *update* datanya dengan nilai polaritas yang diberikan.

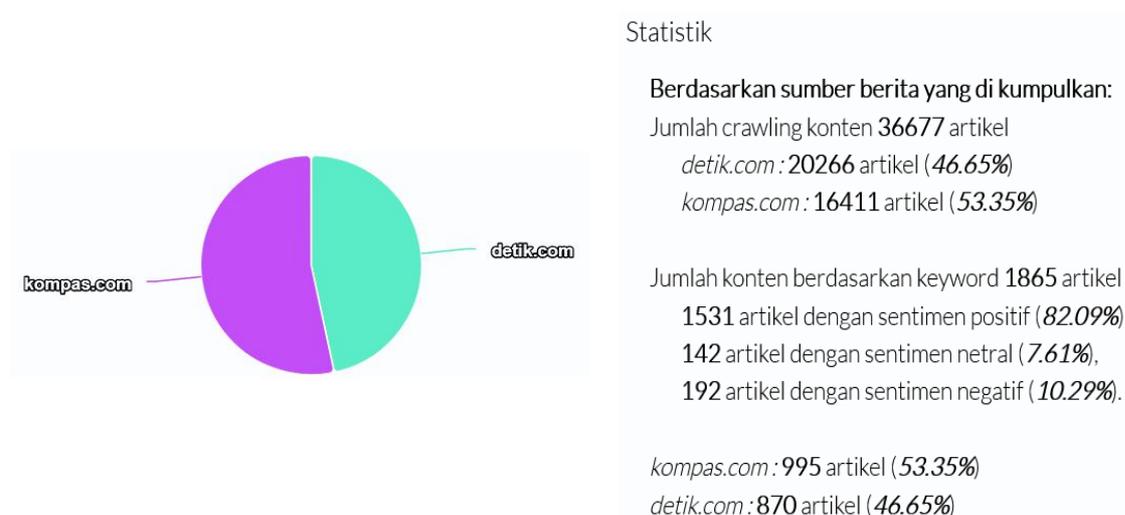
4. *Windows Task Scheduler*, dilakukan untuk melakukan agregasi data sebagai pembacaan analisa pada visualisasi data nantinya. Sistem *custom* yang dibentuk tidak didesain dan tidak disarankan untuk melakukan agregasi pada ribuan data *on-the-fly* atau dibentuk saat suatu halaman visualisasi ditampilkan yang mengakibatkan render halaman mengalami error. Tugas agregasi yang dilakukan pada *windows task scheduler*, dieksekusi setiap 1 atau 2 jam sekali. Hasil agregasi data disimpan di tabel yang berbeda pada *database* penyimpanan yang sama. Hasil agregasi inilah yang nantinya akan dibaca dan ditampilkan pada grafik pembacaan atau dashboard.
5. *Visualization*, ditampilkan dalam bentuk grafik *bar chart*, *pie chart*, *stacked bar chart*, *donut chart* dan *trend chart*. Hasil pembacaan ini digunakan untuk menganalisa pola sentimen yang diberikan pada setiap calon kandidat terhadap sumber digital media. Hal ini dapat memberikan gambaran kepada kita semua trend yang diberikan kepada bakal calon presiden 2024 tanpa harus membaca semua artikel yang dipublikasikan. Hasil pola sentimen ini hanya menjadi salah satu faktor penentu sebagai pengambil suara, bukan menjadi satu-satunya faktor penentu nantinya.

Kelima tahapan diatas berjalan secara terus menerus tanpa terjadinya suatu bentrokan antara satu tahap dengan yang lainnya. Sistem *custom* yang dibentuk memastikan setiap tahap berjalan secara independen tanpa bergantung pada tahap yang lainnya. Berikut hasil dan analisa yang ditampilkan pada visualisasi sistem custom tersebut.

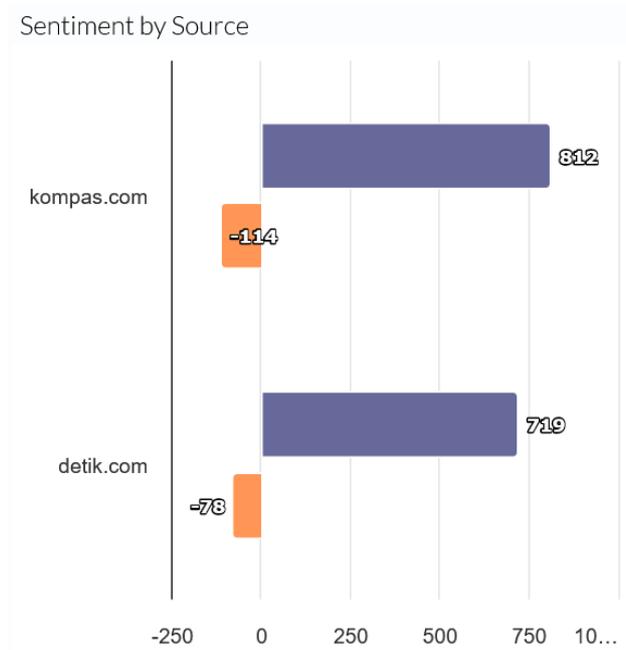


Gambar 4. Hasil analisa sistem *custom*

Jumlah total konten crawling selama periode 3 bulan antara 1 Maret 2023 dan 31 Mei 2023 adalah 36677 artikel dari sumber digital media *kompas.com* dan *detik.com*. Dari total tersebut, hasil *text extraction* terhadap 3 kata kunci, Ganjar Pranowo, Prabowo Subianto dan Anies Baswedan, menggunakan fitur *Full-Text Search* pada *SQL Server Database Engine* adalah 1865 artikel. Statistik yang dihasilkan pada pemberitaan sebagai berikut:



Gambar 5. Statistik *Text Extraction*



Gambar 6. Statistik *Sentiment*

Pemberitaan paling positif terdapat pada media kompas.com sebanyak 812 artikel (43.54%) dari total 1865 artikel. Pemberitaan paling negatif pun terdapat pada media kompas.com sebanyak 114 artikel (6.11%) dari total 1865 artikel.



Gambar 7. Statistik Bakal Calon Presiden

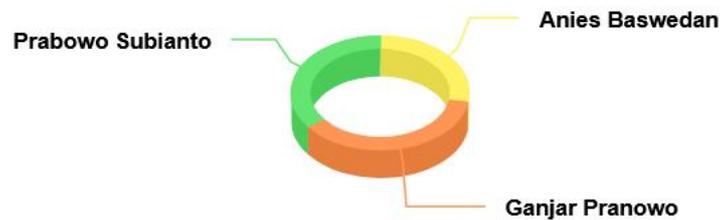
Secara total, Ganjar Pranowo menghasilkan 843 artikel dengan 91.22% berpolaritas positif dan 8.78% berpolaritas negatif. Prabowo Subianto menghasilkan 792 artikel dengan 92.17%

berpolaritas positif dan 7.83% berpolaritas negatif. Anies Baswedan menghasil 620 artikel dengan 81.45% berpolaritas positif dan 18.58% berpolaritas negatif.

Berikut hasil yang diberikan berdasarkan polaritas terhadap kandidat calon Presiden Indonesia 2024:

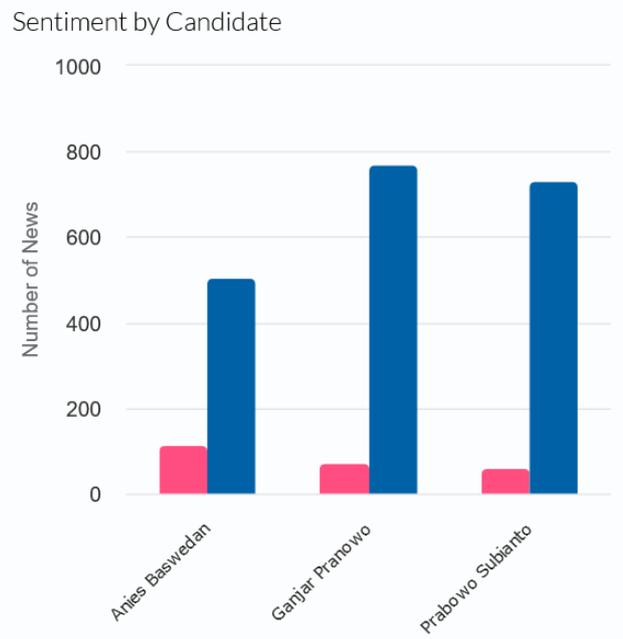
Kandidat	Jumlah Artikel	Sentimen Positif	Sentimen Netral	Sentimen Negatif
Ganjar Pranowo	907	769 (84.79%)	64 (7.06%)	74 (8.16%)
Prabowo Subianto	849	730 (85.98%)	57 (6.71%)	62 (7.30%)
Anies Baswedan	675	505 (74.81%)	55 (8.15%)	115 (17.04%)

Popularity Chart



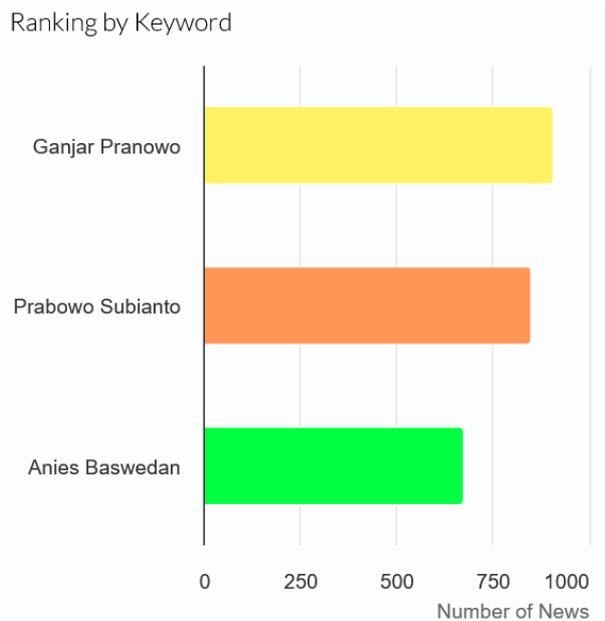
Gambar 8. Popularity Chart

Berdasarkan popularitas, pemberitaan seputar Ganjar Pranowo menempati posisi pertama dengan jumlah artikel sebanyak 907 atau 48.63% dari total 1865 artikel berdasarkan *keyword*.



Gambar 9. Sentiment by candidate

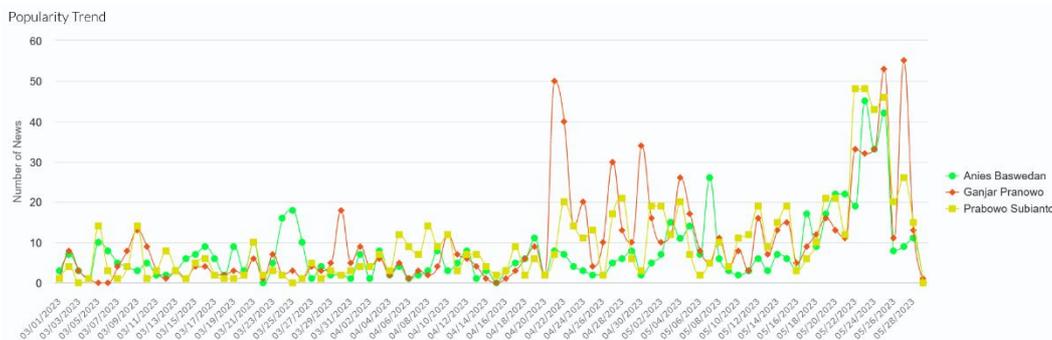
Sentimen pemberitaan paling positif terlihat *keyword* Ganjar Pranowo sebanyak 769 artikel (31.63%) dan pemberitaan paling negatif terlihat pada *keyword* Anies Baswedan sebanyak 115 artikel (4.73%).



Gambar 10. *Ranking by keyword*

Popularitas pemberitaan terbanyak pada digital media berdasarkan keyword adalah Ganjar Pranowo sebanyak 907 artikel (37.31%) dan pemberitaan paling sedikit adalah Anies Baswedan sebanyak 675 artikel (27.77%).

Berikut hasil trend popularitas pemberitaan terhadap 3 calon kandidat Presiden Indonesia 2024 dalam kurun waktu 3 bulan mulai dari 1 Maret 2023 hingga 31 Mei 2023.



Gambar 11. *Popularity Trend*

Spike pemberitaan Ganjar Pranowo terjadi pada tanggal 27 Mei 2023 sebanyak 55 artikel, spike pemberitaan Prabowo Subianto terjadi pada tanggal 22 Mei 2023 sebanyak 48 artikel dan spike pemberitaan Anies Baswedan terjadi pada tanggal 23 Mei 2023 sebanyak 45 artikel.

5. Kesimpulan

Dari penelitian yang telah dilakukan untuk menghasilkan pola sentimen terhadap 3 calon kandidat Presiden Indonesia 2024 dengan keyword Ganjar Pranowo, Prabowo Subianto dan Anies Baswedan terhadap 2 sumber digital media, kompas.com dan detik.com, terlihat calon kandidat Ganjar Pranowo unggul untuk sementara ini dari sisi popularitas sebanyak 907 artikel (37.31%), jumlah

sentimen positif sebanyak 769 artikel (31.63%) dan spike jumlah pemberitaan terbanyak pada tanggal 27 Mei 2023 sebanyak 55 artikel terhadap 2 sumber digital media tersebut.

Perlu diingat bahwa analisa pola hasil sentimen yang ditunjukkan hanya sebagai salah satu faktor petunjuk dan pertimbangan bagi para pemungut suara, bukan menjadi salah satu faktor penentu baik buruknya seorang calon kandidat Presiden Indonesia 2024 nanti.

Adapun pengembangan penelitian berikutnya yang dapat dilakukan untuk mendapatkan hasil yang lebih akurat menggunakan metode deeplearning yang merupakan subset dari machine learning. Jumlah artikel dapat ditambah dengan periode rentang waktu yang lebih lama. Hal ini memberikan hasil yang jauh lebih akurat dan pembentukan pola yang lebih sempurna dengan margin error yang rendah.

DAFTAR PUSTAKA

- [1] I. Budi and R. R. Suryono, "Application of named entity recognition method for Indonesian datasets: a review," *Bull. Electr. Eng. Informatics*, vol. 12, no. 2, pp. 969–978, 2023, doi: 10.11591/eei.v12i2.4529.
- [2] S. Fu, N. Lin, G. Zhu, and S. Jiang, "Towards Indonesian Part-of-Speech Tagging: Corpus and Models," *2018 Int. Conf. Asian Lang. Process.*, vol. 1, pp. 303–307, 2018, [Online]. Available: <http://universaldependencies.org/>
- [3] A. A. Jalal and B. H. Ali, "Text documents clustering using data mining techniques," *Int. J. Electr. Comput. Eng.*, vol. 11, no. 1, pp. 664–670, 2021, doi: 10.11591/ijece.v11i1.pp664-670.
- [4] S. K. Shinde, V. Bhojane, and P. Mahajan, "NLP based Object Oriented Analysis and Design from Requirement Specification," *Int. J. Comput. Appl.*, vol. 47, no. 21, pp. 30–34, 2012, doi: 10.5120/7475-0574.
- [5] H. Gohil, "a Review on a Emotion Detection and Recognition From Text a Review on a Emotion Detection and Recognition From Text Using Natural," vol. 13, no. April, pp. 6745–6750, 2018.
- [6] V. S and J. R, "Text Mining: open Source Tokenization Tools – An Analysis," *Adv. Comput. Intell. An Int. J.*, vol. 3, no. 1, pp. 37–47, 2016, doi: 10.5121/acii.2016.3104.
- [7] Das Deepak, "Social Media Sentiment Analysis using Machine Learning : Part — I ,," *Towar. Data Sci.*, vol. 6, no. 6, pp. 465–472, 2019, [Online]. Available: <https://towardsdatascience.com/social-media-sentiment-analysis-49b395771197>
- [8] A. Romanov, A. Kurtukova, M. Vasilieva, and R. Meshcheryakov, "Sentiment analysis of text using machine learning techniques," *CEUR Workshop Proc.*, vol. 2233, no. 05, pp. 2952–2956, 2017.
- [9] D. Jurafsky and J. Martin, "Naive bayes and sentiment classification," *Speech Lang. Process.*, p. 1024, 2019, [Online]. Available: <https://web.stanford.edu/~jurafsky/slp3/6.pdf>
- [10] M. H. S. Quadri* and D. R. K. Selvakumar, "Performance of Naïve Bayes in Sentiment Analysis of User Reviews Online," *Int. J. Innov. Technol. Explor. Eng.*, vol. 10, no. 2, pp. 64–68, 2020, doi: 10.35940/ijitee.a8198.1210220.