



Analisis Sentimen di Media Online menggunakan Metode Naive Bayes

Elizabeth Graseylla Ruus¹, Luther Alexander Latumakulita^{1*}, Jantje Denny Prang¹

¹Jurusan Matematika–Fakultas Matematika dan Ilmu Pengetahuan Alam–Universitas Sam Ratulangi Manado, Indonesia

*Corresponding author : latumakulitala@unsrat.ac.id

ABSTRAK

Telah dilakukan penelitian untuk menghasilkan sistem analisis sentimen pada judul berita tentang COVID-19 di media online dengan parameter positif (klasifikasi berita baik) dan negatif (klasifikasi berita buruk). Proses analisis dan klasifikasi menggunakan algoritma Naive Bayes dengan 50 Data berita Training dan 10 Data berita Testing. Hasil eksperimen menunjukkan bahwa performa dari penerapan algoritma Naive Bayes menghasilkan nilai accuracy sebesar 0,7 , nilai precision sebesar 0,6 , dan nilai recall sebesar 0,75.

INFO ARTIKEL

Diterima :
Diterima setelah revisi :
Tersedia online :

Kata Kunci:
COVID-19
Media Online
Naive Bayes

ABSTRACT

Research has been carried out to produce a sentiment analysis system for news headlines about COVID-19 in online media with positive parameters (classification of good news) and negative (classification of bad news). The analysis and classification process uses the Naive Bayes algorithm with 50 Training news data and 10 Testing news data. The experimental results show that the performance of the application of the Naive Bayes algorithm produces an accuracy value of 0.7, a precision value of 0.6, and a recall value of 0.75.

ARTICLE INFO

Accepted :
Accepted after revision :
Available online :

Keywords:
COVID-19
Online Media
Naive Bayes

1. PENDAHULUAN

Coronavirus Disease-2019 (COVID-19) ditemukan pada manusia sejak kejadian muncul di Wuhan Cina, pada Desember 2019. Kebanyakan orang yang terinfeksi virus COVID-19 akan mengalami penyakit pernapasan ringan hingga sedang dan sembuh tanpa memerlukan perawatan khusus [1]. Pemberitaan mengenai COVID-19 di media online yang bersifat sensasional, menimbulkan kehebohan bahkan berpotensi menyebabkan kepanikan karena begitu banyak judul berita yang memiliki kandungan sentimen, baik positif ataupun negatif. Membutuhkan waktu yang lama jika berita-berita tersebut dianalisis secara manual. Maka dari itu perlu dibuat sistem analisis sentimen yang mampu menganalisis berita yang ada dalam media online tersebut apakah termasuk ke dalam berita yang mengandung sentimen positif (berita baik) atau negatif (berita buruk).

Analisis sentimen merupakan kombinasi antara text mining dan natural language processing, dimana proses memahami, mengekstrak, dan mengolah data tekstual secara otomatis untuk mendapatkan informasi sentimen yang terkandung dalam suatu kalimat [2]. Untuk membuat analisis sentimen, yang harus dipersiapkan adalah memilih metode klasifikasi yang akan digunakan. Naive Bayes merupakan metode klasifikasi yang dikembangkan berdasarkan aturan

Bayes dengan melihat kondisi-kondisi yang ada dan peluang-peluang setiap kondisinya [3].

Beberapa Penelitian tentang analisis sentimen menggunakan metode Naive Bayes juga telah pernah dilakukan, membuat aplikasi analisis sentimen berbasis website untuk memberikan informasi mengenai klasifikasi opini yang ada pada tempat wisata Taman Mini Indonesia apakah opini tersebut tergolong opini positif atau opini negatif. Selain itu juga, ada penelitian yang pernah dilakukan oleh [4], analisis sentimen berita artis dengan menggunakan penggabungan metode pemilihan fitur yaitu Particle Swarm Optimization (PSO) agar bisa meningkatkan akurasi pada Support Vector Machine, sehingga menghasilkan akurasi 76%.

Berdasarkan contoh dari penelitian-penelitian sebelumnya, maka penulis membuat sebuah sistem analisis sentimen pada judul berita yang ada di media online, dengan menggunakan Naive Bayes sebagai metode klasifikasi.

2. METODE PENELITIAN

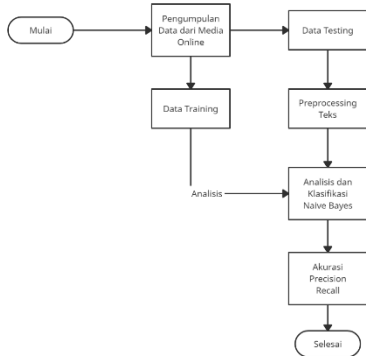
Waktu dan Tempat Penelitian

Dalam penulisan skripsi ini, peneliti melakukan pengolahan data dirumah secara daring/online, dikarenakan situasi dan kondisi adanya virus Covid-19 yang mengharuskan untuk tinggal dirumah. Untuk waktu penyusunan proposal skripsi dilakukan dari awal Maret 2021 sampai September 2021.

Data Penelitian

Data yang digunakan pada penelitian ini berupa judul berita tentang Covid-19. Untuk data berita positif merujuk pada berita baik, dan data berita negatif pada berita buruk. Data tersebut diambil secara langsung dari media Online seperti: kompas.com dan tribunnews.com.

Tahapan Penelitian



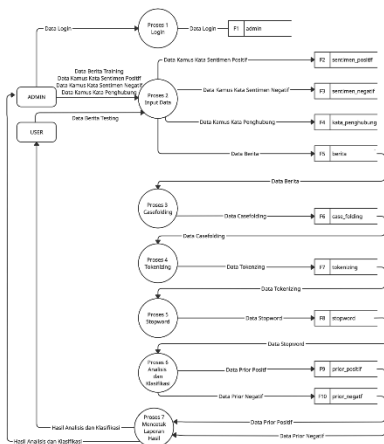
Gambar 1. Diagram Tahapan Penelitian

Perancangan Sistem



Gambar 2. Data Flow Diagram (DFD) Level 0

Alur sistem dimulai dengan aktivitas Admin yang akan melakukan proses login dengan memasukkan username dan password terlebih dahulu. Setelah berhasil login, Admin dapat melakukan input, update dan hapus data berita, kamus kata sentimen positif, kamus kata sentimen negatif, dan kamus kata penghubung. Kemudian untuk User sebagai pengguna untuk melakukan Analisis Sentimen di Media Online dapat melakukan input data berita (data berita testing/uji) yang akan diproses kedalam sistem. Hasil keluaran atau output dari sistem berupa hasil analisis sentimen positif atau negatif berdasarkan klasifikasi berita baik atau berita buruk.



Gambar 3. Data Flow Diagram (DFD) Level 1

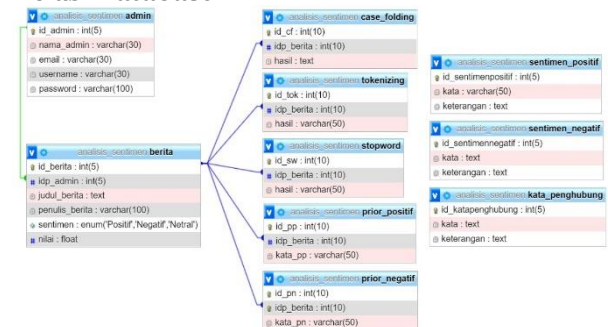
Pada Gambar 3 DFD Level 1, terdapat 7 proses yaitu login, input data, casefolding, tokenizing, stopword, analisis dan klasifikasi & hasil analisis dan mencetak laporan hasil. Proses pertama login, admin akan melakukan login atau masuk terlebih dahulu

dengan memasukkan username dan password yang akan tersimpan dalam tabel admin. Setelah berhasil login, Admin bisa melakukan proses input data kamus kata sentimen positif yang akan tersimpan dalam tabel sentimen_positif, data kamus kata sentimen negatif yang akan tersimpan dalam tabel sentimen_negatif, data kamus kata penghubung yang akan tersimpan dalam tabel kata_penghubung. Namun untuk data berita training, admin akan langsung meng-input dalam database tabel berita. Untuk user atau pengguna yang akan melakukan analisis sentimen bisa langsung meng-input data berita testing sebagai data yang akan diproses untuk analisis sentimen.

Data berita testing akan masuk pada proses Casefolding, yang merupakan tahapan dari preprocessing teks dimana mengubah semua huruf pada judul berita menjadi huruf kecil, hasil dari tahapan casefolding akan tersimpan dalam tabel case_folding. Setelah itu akan masuk pada proses Tokenizing yang merupakan proses pemisahan teks menjadi potongan-potongan yang disebut sebagai token untuk kemudian dianalisa, hasil dari tahapan tokenizing akan tersimpan dalam tabel tokenizing. Kemudian masuk pada proses Stopword dimana proses menghapus kata-kata yang dianggap tidak penting dan tidak berpengaruh terhadap proses analisis, hasil dari proses stopword akan tersimpan dalam tabel stopword.

Hasil dari proses stopword akan berupa kata yang akan masuk pada proses analisis dan klasifikasi. Sistem akan menganalisis dan melakukan klasifikasi menggunakan algoritma Naive Bayes dengan menghitung nilai kemunculan kata dari setiap kelas positif akan tersimpan dalam tabel prior_positif dan negatif akan tersimpan dalam tabel prior_negatif. Setelah berhasil melakukan analisis dan klasifikasi, sistem akan memberikan hasil analisis dan klasifikasi kepada Admin dan User.

Relasi Database



Gambar 4. Desain Relasi antar tabel

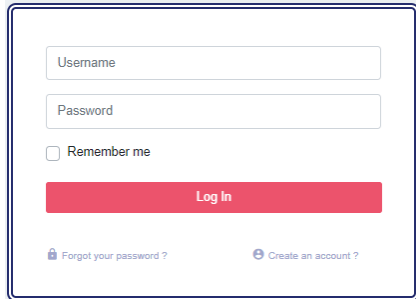
Gambar 4 merupakan perancangan desain antar tabel pada database. Peneliti menggunakan software MySQL dalam merancang database untuk sistem ini. Terdapat 7 tabel, yaitu tabel admin yang memiliki relasi dengan tabel berita dan tabel berita memiliki relasi dengan tabel case_folding, tabel tokenizing, tabel stopword, tabel prior_positif, tabel prior_negatif. Selain itu juga ada tiga tabel pendukung yaitu tabel sentimen_positif, tabel sentimen_negatif dan tabel kata_penghubung.

3. HASIL DAN PEMBAHASAN

Implementasi Antar Muka

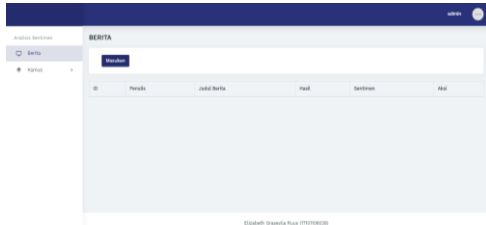
Aplikasi Analisis Sentimen ini memiliki antarmuka berupa tampilan web. Berikut ini adalah tampilan-tampilan dari aplikasi ini :

- Tampilan Admin untuk Login



Gambar 5. Tampilan Admin untuk Login

- Tampilan berita Admin



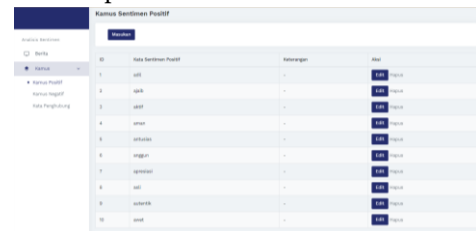
Gambar 6. Tampilan Berita

- Tampilan berita User



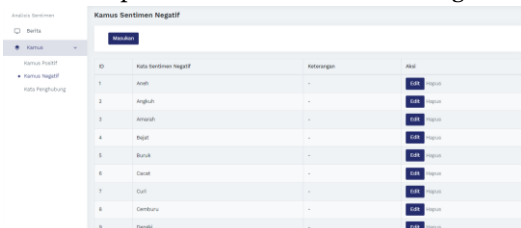
Gambar 7. Tampilan Form Input berita User

- Tampilan Kamus Kata Sentimen Positif



Gambar 8. Tampilan Kamus Kata Sentimen Positif

- Tampilan Kamus Kata Sentimen Negatif



Gambar 9. Tampilan Kamus Kata Sentimen Negatif

- Tampilan Kamus Kata Penghubung



Gambar 10. Tampilan Kamus Kata Penghubung

Proses Analisis dan Klasifikasi

Proses analisis dan klasifikasi data berita dibagi menjadi 2 data yaitu data training dan data testing. Data training akan menjadi sumber informasi bagi data testing untuk melakukan proses analisis. Untuk jumlah data training sebanyak 50 data berita dan data testing sebanyak 10 data berita.

Data Testing	Label
Vaksin AstraZeneca Aman dan Efektif	Tidak diketahui

Gambar 11. Data Testing

Sebelum melakukan analisis sentimen, Data Testing akan masuk pada proses Preprocessing. Terdapat 3 tahapan dalam preprocessing, antara lain:

1. *Casefolding* yaitu mengubah semua huruf pada judul berita menjadi huruf kecil.

Casefolding	
Sebelum	Sesudah
Vaksin AstraZeneca Aman dan Efektif	vaksin astrazeneca aman dan efektif

Gambar 12. Tahapan Casefolding

2. *Tokenizing* dimana proses pemisahan teks menjadi potongan-potongan yang disebut sebagai token untuk kemudian di analisa.

Tokenizing	
Sebelum	Sesudah
vaksin astrazeneca aman dan efektif	vaksin astrazeneca aman dan efektif

Gambar 13. Tahapan Tokenizing

3. *Stopword* yaitu proses menghapus kata-kata yang di anggap tidak penting dan tidak berpengaruh terhadap proses analisis.

Stopword	
Sebelum	Sesudah
vaksin astrazeneca aman dan efektif	vaksin astrazeneca aman efektif

Gambar 14. Tahapan Stopword

Setelah selesai Preprocessing Teks, Sistem akan menganalisis token kata berita testing yang muncul pada berita training.

Data Training	Vaksin	Astrazeneca	Aman	Efektif
Pos4	0	1	0	0
Pos7	1	0	0	0
Pos10	1	0	0	0
Pos11	0	1	0	0
Pos15	1	0	0	1
Pos16	0	1	1	0
Pos17	1	0	0	0
Pos19	0	0	0	1
Pos22	0	1	0	0
Pos23	0	1	0	0
Neg3	0	1	0	0
Neg4	0	1	0	0
Neg6	1	0	0	0
Neg9	1	0	0	0
Neg11	0	0	0	1
Neg12	0	0	1	0
Neg13	1	1	0	0
Neg24	1	0	0	0

Gambar 15. Atribut Data Testing

Atribut kemunculan token kata data testing pada data training di setiap kelas positif dan negatif. Dengan nilai

1 = kata muncul, nilai 0 = kata tidak muncul. Kemudian sistem akan melakukan analisis dan klasifikasi pada token kata menggunakan algoritma Naive Bayes, sebagai berikut:

- a. Sistem akan melakukan perhitungan peluang prior dari data testing berdasarkan data training :

$$P'(K) = \frac{N_1}{n}$$

Keterangan :

$P'(K)$ = Peluang Hipotesis K

N_1 = Jumlah Total pada suatu *Class*

N = Jumlah Total Data

$$P(\text{Class} = \text{"Positif"}) = \frac{10}{18} = 0,555555556$$

$$P(\text{Class} = \text{"Negatif"}) = \frac{8}{18} = 0,444444444$$

- b. Kemudian sistem akan menghitung peluang atribut kemunculan token kata terhadap masing-masing kelas pada data testing berdasarkan data training: Kelas *Positif* $P(X|Positif)$

$$= P(\text{Vaksin}=4 | \text{Positif}) * P(\text{Astrazeneca}=5 | \text{Positif}) * P(\text{Ama} \\ n=1 | \text{Positif}) * P(\text{Efektif}=2 | \text{Positif}) \\ = 0,4 * 0,5 * 0,1 * 0,2 = 0,004$$

Kelas *Negatif* $P(X31|Negatif)$

$$= P(\text{Vaksin}=4 | \text{Positif}) * P(\text{Astrazeneca}=3 | \text{Positif}) * P(\text{Ama} \\ n=1 | \text{Positif}) * P(\text{Efektif}=1 | \text{Positif}) \\ = 0,5 * 0,375 * 0,125 * 0,125 = 0,00293$$

- c. Kemudian Sistem akan menghitung perkalian peluang prior dengan peluang atribut pada masing-masing kelas, yaitu Kelas *Positif* dan Kelas *Negatif* .

Untuk Kelas Positif :	Untuk Kelas Negatif:
$= P(X Positif) * P(\text{Class} = \text{"Positif"})$	$= P(X Negatif) * P(\text{Class} = \text{"Negatif"})$
$= 0,004 * 0,555555556$	$= 0,002929688 * 0,444444444$
$= 0,002222222$	$= 0,001302083$

Gambar 16. Atribut Data Testing

- d. Setelah sistem mendapatkan nilai dari masing-masing kelas, nilai terbesar dari perhitungan yang merupakan hasil prediksi klasifikasi. Nilai terbesar dari perhitungan diatas berada di kelas positif, maka untuk klasifikasi berita testing pada contoh ini termasuk pada kelas Positif.

Proses Analisis dan Klasifikasi menggunakan Program

Berikut ini merupakan Analisis dan Klasifikasi antarmuka tampilan web :

- Pengguna akan melakukan Input Penulis serta Judul Berita yang akan diklasifikasi kedalam menu "Masukan" :

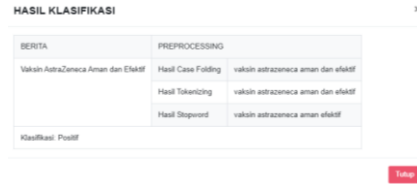


Gambar 17. Menu Input topik berita

- Tekan tombol "Simpan" untuk diproses



Gambar 18. Form input topik berita



Gambar 19. Hasil Analisis dan Klasifikasi topik berita

Pengukuran Kinerja Sistem

Peneliti melakukan uji coba pada sistem untuk mengukur kinerja sistem menggunakan confusion matrix. Dengan 10 data berita testing, diantaranya 5 berita testing positif dan 5 berita testing negatif.

Gambar 20. Hasil Analisis dan Klasifikasi topik berita

Prediksi	Aktual	
	Negatif	Positif
Negatif	4 (True Negative)	2 (False Positive)
Positif	1 (False Negative)	3 (True Positive)

Gambar 21. Confusion Matrix Data Testing

Terdapat sebanyak 5 data testing negatif yang terdiri dari 4 data diprediksi negatif dan 1 data diprediksi positif. Sebaliknya, untuk data testing positif sebanyak 5 data yang terdiri dari 3 data diprediksi positif dan 2 data diprediksi negatif. Dari data tersebut dapat dihitung nilai accuracy, precision, dan recall.

- Accuracy untuk mengukur tingkat kualitas keberhasilan klasifikasi berita pada tahap testing [5].
- Precision merupakan tingkat ketepatan antara informasi yang diminta dan jawaban yang diberikan oleh sistem (Anindya et al., 2020).
- Recall merupakan tingkat keberhasilan sistem dalam menemukan kembali suatu informasi (Anindya et al., 2020).

Berikut ini merupakan hasil Perhitungan dari Nilai Accuracy, Precision dan Recall :

Accuracy	Precision	Recall
0.7	0.6	0.75

4. PENUTUP Kesimpulan

Aplikasi sistem analisis sentimen pada media online berhasil melakukan analisis dan klasifikasi berita mengenai COVID-19 di media online. Dilakukan pengukuran kinerja sistem untuk mengukur performa dari penerapan algoritma naive bayes dengan 50 data

berita training dan 10 data berita testing. Sistem berhasil mengklasifikasikan 7 data testing diprediksi benar, dan 3 data testing diprediksi salah. Maka didapatkan Hasil accuracy metode klasifikasi sebesar 0,7, precision sebesar 0,6, dan recall sebesar 0,75. Hal ini menunjukkan bahwa metode klasifikasi baik untuk menganalisis dan mengklasifikasi judul berita tentang COVID-19 yang ada di media online.

REFERENSI

- [1] Kementerian Kesehatan Republik Indonesia. 2020. Coronavirus Disease 2019(COVID-19). <https://www.who.int/indonesia/news/novel-coronavirus/qa-for-public> [29 Juli 2021]
- [2] Rozi, I. F., Hadi, S., & Achmad, E. 2012. Implementasi Opinion Mining untuk Ekstraksi Data Opini Publik pada Perguruan Tinggi. Jurnal EECCIS. 6(1): 37-43.
- [3] Fanissa,S., Fauzi, M.A., & Adinugroho,S. 2018. Analisis Sentimen Pariwisata di Kota Malang menggunakan Metode Naive Bayes dan Seleksi fitur Query Expansion Ranking. Jurnal Pengembangan Teknologi informasi dan Ilmu Komputer. 2(2): 2766-2770.
- [4] Yunita, N. 2016. Analisis Sentimen Berita Artis dengan Menggunakan Algoritma Support Vector Machine dan Particle Swarm Optimization. Jurnal Sistem Informasi STMIK Antar Bangsa. 5(2): 104-112.
- [5] Sokolova, M., & Lapalme, G. 2009. A Systematic analysis of performance measures for classification tasks. Information Processing and Management. vol. 45, no. 4. 427-437.

Elizabeth Graseylla Ruus

17101106038@student.unsrat.ac.id



Lahir di Manado, 31 Januari 2000. Menempuh pendidikan tinggi Prodi Sistem Informasi Jurusan Matematika, FMIPA, Universitas Sam Ratulangi Manado. Tahun 2021 adalah tahun terakhir ia menempuh studi. Makalah ini merupakan hasil penelitian skripsinya yang dipublikasikan.

Luther Alexander Latumakulita

latumakulita@unsrat.ac.id



Pada tahun 1997, ia memperoleh gelar Sarjana dari departemen Komputer Ilmu Universitas Gadjah Mada Yogyakarta. Master Ilmu Komputer didapat dari Departemen Ilmu Komputer Universitas Gadjah Mada Yogyakarta pada tahun 2008. Pada tahun 2018, ia memperoleh Gelar Doktor dari Departemen Ilmu Komputer dan Teknik Elektro, Universitas Kumamoto Jepang. Menjadi dosen Matematika Jurusan, FMIPA, Universitas Sam Ratulangi Manado dari tahun 2008 sampai sekarang. Ahli dalam Sistem *Fuzzy*, *Artificial Neural Network* dan *Machine learning*.

Jantje Denny Prang (jantjeprang@gmail.com)



Lahir pada tanggal 20 Desember 1958. Gelar Master Sains (M.Si) diperoleh dari Institut Pertanian Bogor. Ia bekerja di UNSRAT di Program Studi Matematika sebagai pengajar akademik tetap UNSRAT.