

Menerapkan Analisis Sentimen Metode Naïve Bayes dan svm

Arja Guswandri¹⁾, Lili Andraini²⁾

¹ Sistem Informasi Akutansi, ² Teknik Komputer

*⁾ Email : lilimesuji02mesuji@gmail.com

Abstrak

Karena pandemi coronavirus, rapat cloud Zoom baru-baru ini mulai digunakan dan semua aktivitas telah divirtualkan. Zoom Cloud Meetings, aplikasi yang menawarkan penggunaan seperti konferensi video dan audio. Penelitian ini menggunakan metode Naive Bayes dan Support Vector Machine untuk menganalisis opini review pengguna Zoom Cloud Meetings di Google Play. Evaluasi model dengan validasi silang 10 kali lipat memberikan nilai presisi dan AUC untuk setiap algoritma. Artinya, nilai AUC = 0.886 dan nilai AUC = 0.886 dengan akurasi = 81,22% untuk algoritma SVM. Sedangkan untuk NB, nilai AUC = 0,659 dan nilai Accuracy = 74,37%. Setelah dilakukan preprocessing record-record tersebut menghasilkan 1.007 record. Situasi data cukup berimbang dengan 546 peringkat label positif dan 461 peringkat label negatif.

Kata Kunci: *Zoom Cloud Meeting*, Naïve Bayes, SVM, Google Play Store dan Virus Corona.

PENDAHULUAN

Penggunaan Zoom Cloud Meetings mulai meledek jumlah pemakaiannya ini dikarenakan pandemi virus corona yang melanda hampir seluruh negara termasuknya Negara indonesia(Gumantan, Nugroho, and Yuliandra 2021). Salah satu kebijakan pemerintah menekankan bahwa penduduk indonesia dikarantina selama adanya pandemi ini(Saputra and Pasha 2021), para pelajar dan mahasiswa/i supaya belajar pada rumah selama pandemi atau mampu dibilang (SFH) Study From Home(Yuliansyah and Ayu 2021), Sehingga menciptakan seluruh aktivitas dilakukan secara impian dikarenakan adanya kebijakan ini(Styawati, StyawatiStyawati, S., & Ariany, F. (2021). Sistem Monitoring Tumbuh Kembang Balita/Batita di Tengah Covid-19 Berbasis Mobile. J. Inform. Univ. Pamulang, 5(4) and Ariany 2021).

Kegiatan belajar ketika ini nir mampu dilakukan secara tatap muka (Offline) misalnya sebelumnya(Arpiansah, Fernando, and Fakhrurozi 2021), ketika ini aktivitas beralih sebagai e-learning yg didukung sang poly sekali pelaksanaan(Sari and Oktaviani 2021), galat satu pelaksanaan yang dipakai yaitu zoom cloud meetings yang paling acapkalikali dipakai sang pengajar juga dosen pada melakukan aktivitas mengajar, sebagai akibatnya dibuatlah impian class(Hendra Saputra and Pasha 2021).

Analisis Sentimen merupakan proses komputasi menggunakan cara memahami, memasak dan mengekstrak data tekstual secara otomatis buat menerima fakta sentimen yg masih ada pada suatu kalimat pendapat atau opini (Alita and Isnain 2020), konduite dan emosi seorang terhadap entitas (mendeskripsikan individu, insiden atau topik) (Rahman Isnain et al. 2021), tujuannya menyediakan fakta berharga bagi seorang yg terkandung berdasarkan sebuah dataset yang nir terstruktur dampak (Rahman Isnain et al. 2021) dan manfaat berdasarkan analisis sentimen mengakibatkan penelitian dan apliasi berbasis sentimen berkembang pesat (Alita, Fernando, and Sulistiani 2020), poly perusahaan yang memfokuskan dalam layanan analisis sentimen (Aldino, Saputra, and Nurkholis 2021).

KAJIAN PUSTAKA

Penulis dalam penelitian ini menggunakan metode Naive Bayes dan Support Vector Machine untuk mendapatkan hasil akurasi tepat dalam menganalisa label sentimen pada ulasan pada penggunaan aplikasi zoom cloud meetings di Google play store. ulasan yang diberikan agar dapat menghasilkan kekuatan terbaik adalah sebagai berikut :

Pengambilan Data

Pengumpulan data penelitian ini diperoleh dari review pengguna aplikasi Zoom cloud meeting di Google Play store (Syah and Witanti 2022). Teknik pengumpulan data yang digunakan adalah web scraping (data miner) (Styawati et al. 2021). Ini adalah teknik yang digunakan untuk mengekstrak data dalam jumlah besar dan menyimpannya dalam file komputer (Isnain et al. 2021). Atau bisa juga database bergaya spreadsheet (Styawati and Mustofa 2019). Proses kemudian menggunakan alat Rapidminer dan kemudian memproses data (Pratama and Surahman 2021).

Pemrosesan Data

Proses ini mempersiapkan data yang dikumpulkan untuk kompilasi pada waktu struktur (Wibowo and Priandika 2021). Pengolahan data meliputi kegiatan menyiapkan data sehingga dapat dikelola pada tahap pemodelan data (Pasha 2020). Berikut ini adalah pemrosesan data antara lain:

1. Remove Number : Menghapus data yang terdapat pada suatu teks.
2. Remove URL : Menghapus URL yang terdapat pada suatu teks.

3. **Regex Removal** : Menghilangkan suatu regular expression yang terdapat pada suatu teks.

Pemodelan Data

Metode yang digunakan pada tahap ini adalah Naive Bayes dan Support Vector Machines(Neneng, Adi, and Isnanto 2016). Pemodelan ini menetapkan validasi silang 10 kali lipat(Neneng, Putri, and Susanto 2021). Yaitu membagi data menjadi 10 bagian(Series 2021), salah satunya adalah data uji dan yang lainnya adalah data latih(Permata and Abidin 2020). Pengujian model yang dilakukan adalah klasifikasi label positif dan negatif saat melakukan scoring dengan Zoom Cloud Meetings di Google Play Store menggunakan metode Naive Bayes dan Support Vector Machines untuk mendapatkan skor tertinggi(Ahmad, Prasetyawan, and Sari 2019). Alat yang digunakan adalah RapidMiner(Purnama, Megawaty, and Fernando 2018).

Evaluasi

Setelah menyelesaikan tahap pemodelan, evaluasi hasil pemodelan(Rahmanto et al. 2021). Bandingkan dua hasil pemodelan dalam hal area di bawah kurva (AUC) antara model(Ahmad et al. 2021). Tujuan dari tahap evaluasi adalah untuk mengetahui kegunaan model yang berhasil dibangun pada langkah sebelumnya(Nabila et al. 2021). Evaluasi menggunakan validasi silang 10 kali lipat(Pamungkas et al. 2020). Mengevaluasi kinerja akurasi AUC dalam eksperimen yang ditentukan.

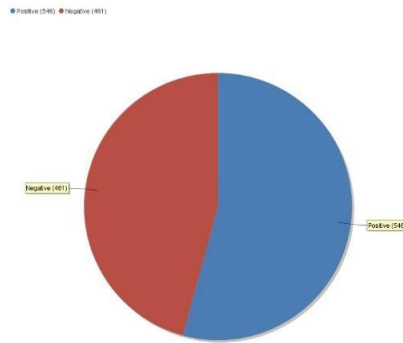
HASIL DAN PEMBAHASAN

Penulis menjelaskan bagaimana proses yang akan dijalankan dalam penelitian ini, tools yang akan digunakan adalah RapidMiner dalam menerapkan metode algoritma yang telah ditentukan ada beberapa proses yang diterapkan pada penelitian ini, yaitu:

A. Pengambilan Data

Data yang diperoleh dalam penelitian ini diperoleh dari review pengguna Zoom Cloud Meetings di Google Play menggunakan teknik web scraping(Handoko and Neneng 2021). Dataset yang diekstraksi adalah 1.158 peringkat dan diproses dengan alat RapidMiner(Pohan, Budi, and Suryono 2019). Setelah mengabaikan prapemrosesan label netral (nilai = 0), kumpulan data sekarang memiliki 1.007 peringkat, dan volume data kira-kira seimbang per label(Parinata and Puspaningtyas 2021). Karena data yang tidak

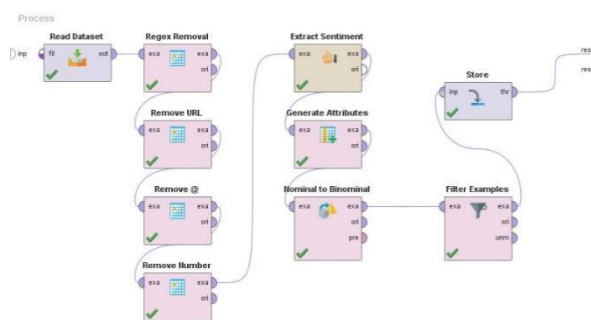
seimbang, klasifikasi yang dilakukan cenderung mengabaikan kelas minoritas (Susanto and Puspaningrum 2019). Data berlabel negatif adalah 461 ulasan dan data berlabel positif adalah 546 ulasan.



Gambar. 1 Perbandingan label positif dan negatif

B. Pemrosesan Data

Sesudah mengambil data dari website, proses selanjutnya adalah mempersiapkan data yang telah didapatkan supaya terolah pada saat melakukan pemodelan proses preprocessing kegiatan membangun data dan juga membersihkan data supaya siap untuk dikelola ke tahap modeling (Megawaty 2020), preprocessing mempunyai beberapa langkah yaitu : Regex Removal, Remove Number dan Remove URL. Labelling (Megawati 2017). Gambar 2 dan 3 merupakan pemrosesan data menggunakan RapidMiner.



Gambar.2 Pemrosesan Data awal

C. Pemodelan Data

Pada hasil pengujian model yang dilakukan mengklasifikasi label positif dan label negatif pada ulasan penggunaan aplikasi Zoom Cloud Meetings di Google Play Store

menggunakan algoritma Naïve Bayes dan Support Vector Machine untuk mendapatkan nilai akurasi terbaik (Riskiono and Pasha 2020). Gambar 3 menunjukkan pemodelan data Naïve Bayes dan Support Vector Machine.

D. Evaluasi

Data yang diperoleh dalam penelitian ini diperoleh dari review pengguna Zoom Cloud Meetings di Google Play menggunakan teknik web scraping (Yulianti, Damayanti, and Prastowo 2021). Dataset yang diekstraksi adalah 1.158 peringkat dan diproses dengan alat RapidMiner (Darwis, Siskawati, and Abidin 2021). Setelah mengabaikan prapemrosesan label netral (nilai = 0), kumpulan data sekarang memiliki 1.007 peringkat, dan volume data kira-kira seimbang per label (Ade and Novri 2019). Karena data yang tidak seimbang, klasifikasi yang dilakukan cenderung mengabaikan kelas minoritas (Melinda, Borman, and Susanto 2018). Data berlabel negatif adalah 461 ulasan dan data berlabel positif adalah 546 ulasan (Anggraini, Pasha, and Damayanti 2020).

SIMPULAN

Hasil analisis saya masing-masing metode yakni naive bayes dan support vector machine, saya merangkum hasil penelitian hasilnya dari evaluasi dinyatakan bahwa untuk mengklarifikasi label positif dan label negatif dari ulasan penggunaan aplikasi zoom cloud meetings di google play, dapat dipastikan nilai akurasi dan nilai AUC pada masing-masing algoritma yaitu untuk NB nilai akurasi = 74,37% dan nilai AUC = 0,659. Sedangkan untuk algoritma SVM nilai akurasi = 81,22% dan nilai AUC = 0,886. Dalam penelitian ini dapat diketahui bahwa tingkat akurasi yang didapatkan algoritma Support Vector Machine (SVM) lebih unggul 6,85% dibandingkan algoritma Naïve Bayes (NB) dengan 1.007 record dataset. Banyak terdapat penelitian sebelumnya yang menggunakan algoritma Support Vector Machine dengan dataset berbeda menghasilkan akurasi yang baik ketika dibandingkan dengan algoritma lainnya.

REFERENSI

Ade, Ade Putra, and Novri Hadinata Novri. 2019. "APLIKASI SIMPAN PINJAM PADA KOPERASI PT. TELKOM PALEMBANG (KOPEGTEL) MENGGUNA Andrian, D. (2021). Penerapan Metode Waterfall Dalam Perancangan Sistem Informasi Pengawasan Proyek Berbasis Web. *Jurnal Informatika Dan Rekayasa Perangkat Lunak (JATIKA)*, 2(1), ." *Jurnal Informanika* 5(2).

Ahmad, Imam, Rohmat Indra Borman, Gavan Gorbi Caksana, and Jafar Fakhrurozi. 2021.

“IMPLEMENTASI STRING MATCHING DENGAN ALGORITMA BOYER-MOORE UNTUK MENENTUKAN TINGKAT KEMIRIPAN PADA PENGAJUAN JUDUL SKRIPSI/TA MAHASISWA (STUDI KASUS: UNIVERSITAS XYZ).” *SINTECH (Science and Information Technology) Journal* 4(1): 53–58.

Ahmad, Imam, Purwono Prasetyawan, and Tri Darma Rosmala Sari. 2019. “Penerapan Algoritma Rekomendasi Pada Aplikasi Rumah Madu Untuk Perhitungan Akuntansi Sederhana Dan Marketing Digital.” In *Seminar Nasional Hasil Penelitian Dan Pengabdian*, , 38–45.

Aldino, Ahmad Ari, Alvin Saputra, and Andi Nurkholis. 2021. “Application of Support Vector Machine (SVM) Algorithm in Classification of Low-Cape Communities in Lampung Timur.” 3(3): 325–30.

Alita, Debby, Yusra Fernando, and Heni Sulistiani. 2020. “Implementasi Algoritma Multiclass SVM Pada Opini Publik Berbahasa Indonesia Di Twitter.” *Jurnal Tekno Kompak* 14(2): 86–91.

Alita, Debby, and Auliya Rahman Isnain. 2020. “Pendeteksian Sarkasme Pada Proses Analisis Sentimen Menggunakan Random Forest Classifier.” *Jurnal Komputasi* 8(2): 50–58.

Anggraini, Yeni, Donaya Pasha, and Damayanti Damayanti. 2020. “SISTEM INFORMASI PENJUALAN SEPEDA BERBASIS WEB MENGGUNAKAN FRAMEWORK CODEIGNITER.” *Jurnal Teknologi dan Sistem Informasi* 1(2): 64–70.

Arpiansah, Rega, Yusra Fernando, and Jafar Fakhrurozi. 2021. “Game Edukasi VR Pengenalan Dan Pencegahan Virus Covid-19 Menggunakan Metode MDLC Untuk Anak Usia Dini.” *Jurnal Teknologi dan Sistem Informasi* 2(2): 88–93.

Darwis, Dedi, Nery Siskawati, and Zaenal Abidin. 2021. “Penerapan Algoritma Naive Bayes Untuk Analisis Sentimen Review Data Twitter Bmkg Nasional.” *Jurnal Tekno Kompak* 15(1): 131–45.

Gumantan, Aditya, Reza Adhi Nugroho, and Rizki Yuliandra. 2021. “Learning During the

- Covid-19 Pandemic: Analysis of E-Learning on Sports Education Students.” *Journal Sport Area* 6(1): 66–75.
- Handoko, Muhammad Ridho, and Neneng Neneng. 2021. “SISTEM PAKAR DIAGNOSA PENYAKIT SELAMA KEHAMILAN MENGGUNAKAN METODE NAIVE BAYES BERBASIS WEB.” *Jurnal Teknologi dan Sistem Informasi* 2(1): 50–58.
- Hendra Saputra, Very, and Donaya Pasha. 2021. “Komik Berbasis Scientific Sebagai Media Pembelajaran Di Masa Pandemi Covid-19.” *SJME (Supremum Journal of Mathematics Education)* 5(1): 85–96.
- Isnain, Auliya Rahman et al. 2021. *6 Comparison of Support Vector Machine and Naïve Bayes on Twitter Data Sentiment Analysis*.
- Megawati, Dyah Ayu. 2017. “Analisis Perbandingan Social Commerce Dari Sudut Pengguna Website.” *Jurnal Teknoinfo* 11(1): 10–13.
- Megawaty, Dyah Ayu. 2020. “Sistem Monitoring Kegiatan Akademik Siswa Menggunakan Website.” *Jurnal Tekno Kompak* 14(2): 98–101.
- Melinda, Mia, Rohmat Indra Borman, and Erliyan Redy Susanto. 2018. “Rancang Bangun Sistem Informasi Publik Berbasis Web (Studi Kasus: Desa Durian Kecamatan Padang Cermin Kabupaten Pesawaran).” *Jurnal Tekno Kompak* 11(1): 1–4.
- Nabila, Zulfa, Auliya Rahman Isnain, Permata Permata, and Zaenal Abidin. 2021. “ANALISIS DATA MINING UNTUK CLUSTERING KASUS COVID-19 DI PROVINSI LAMPUNG DENGAN ALGORITMA K-MEANS.” *Jurnal Teknologi dan Sistem Informasi* 2(2): 100–108.
- Neneng, Neneng, Kusworo Adi, and Rizal Isnanto. 2016. “Support Vector Machine Untuk Klasifikasi Citra Jenis Daging Berdasarkan Tekstur Menggunakan Ekstraksi Ciri Gray Level Co-Occurrence Matrices (GLCM).” *JSINBIS (Jurnal Sistem Informasi Bisnis)* 6(1): 1–10.
- Neneng, Neneng, Novia Utami Putri, and Erliyan Redi Susanto. 2021. “Klasifikasi Jenis Kayu Menggunakan Support Vector Machine Berdasarkan Ciri Tekstur Local Binary Pattern.” *CYBERNETICS* 4(02): 93–100.

- Pamungkas, Nurhuda Budi, Dedi Darwis, Ditha Nurjayanti, and Agung Tri Prastowo. 2020. "Perbandingan Algoritma Pixel Value Differencing Dan Modulus Function Pada Steganografi Untuk Mengukur Kualitas Citra Dan Kapasitas Penyimpanan." *Jurnal Informatika* 20(1): 67–77.
- Parinata, Dwi, and Nicky Dwi Puspaningtyas. 2021. "Optimalisasi Penggunaan Google Form Terhadap Pembelajaran Matematika." *MATHEMA: JURNAL PENDIDIKAN MATEMATIKA* 3(1): 56–65.
- Pasha, Donaya. 2020. "SISTEM PENGOLAHAN DATA PENILAIAN BERBASIS WEB MENGGUNAKAN METODE PIECIES." *Jurnal Teknologi dan Sistem Informasi* 1(1): 97–104.
- Permata, Permata, and Zaenal Abidin. 2020. "Statistical Machine Translation Pada Bahasa Lampung Dialek Api Ke Bahasa Indonesia." *JURNAL MEDIA INFORMATIKA BUDIDARMA* 4(3): 519–28.
- Pohan, Nur Wulan Adhani, I Budi, and Ryan Randy Suryono. 2019. "Borrower Sentiment on P2P Lending in Indonesia Based on Google Playstore Reviews." In *Proceedings of the Sriwijaya International Conference on Information Technology and Its Applications (SICONIAN 2019), Palembang, Indonesia, , 17–23.*
- Pratama, Reno Renaldi, and Ade Surahman. 2021. 1 Jurnal Informatika dan Rekayasa Perangkat Lunak *Perancangan Aplikasi Game Fighting 2 Dimensi Dengan Tema Karakter Nusantara Berbasis Android Menggunakan Construct 2.*
- Purnama, Sandy, Dyah Ayu Megawaty, and Yusra Fernando. 2018. "Penerapan Algoritma A Star Untuk Penentuan Jarak Terdekat Wisata Kuliner Di Kota Bandarlampung." *Jurnal teknoinfo* 12(1): 28–32.
- Rahman Isnain, Auliya, Adam Indra Sakti, Debby Alita, and Nurman Satya Marga. 2021. "Sentimen Analisis Publik Terhadap Kebijakan Lockdown Pemerintah Jakarta Menggunakan Algoritma Svm." *Jdmsi* 2(1): 31–37. <https://t.co/NfhnfMjtXw>.
- Rahmanto, Yuri, Joni Alfian, Damayanti Damayanti, and Rohmat Indra Borman. 2021. "Penerapan Algoritma Sequential Search Pada Aplikasi Kamus Bahasa Ilmiah

- Tumbuhan.” *Jurnal Buana Informatika* 12(1): 21.
- Riskiono, Sampurna Dadi, and Donaya Pasha. 2020. “Analisis Metode Load Balancing Dalam Meningkatkan Kinerja Website E-Learning.” *Jurnal TeknoInfo* 14(1): 22–26.
- Saputra, Very Hendra, and Donaya Pasha. 2021. “Comics as Learning Medium During the Covid-19 Pandemic.” In *Proceeding International Conference on Science and Engineering*, , 330–34.
- Sari, Fatimah Mulya, and Lulud Oktaviani. 2021. “Undergraduate Students’ Views on the Use of Online Learning Platform during COVID-19 Pandemic.” *Teknosastik* 19(1): 41.
- Series, Conference. 2021. “Effect of Mono Corpus Quantity on Statistical Machine Translation Indonesian – Lampung Dialect of Nyo Effect of Mono Corpus Quantity on Statistical Machine Translation Indonesian – Lampung Dialect of Nyo.”
- Styawati, StyawatiStyawati, S., & Ariany, F. (2021). Sistem Monitoring Tumbuh Kembang Balita/Batita di Tengah Covid-19 Berbasis Mobile. *J. Inform. Univ. Pamulang*, 5(4), 490., and Fenty Ariany. 2021. “Sistem Monitoring Tumbuh Kembang Balita/Batita Di Tengah Covid-19 Berbasis Mobile.” *J. Inform. Univ. Pamulang* 5(4): 490.
- Styawati, Andi Nurkholis, Zaenal Abidin, and Heni Sulistiani. 2021. “Optimasi Parameter Support Vector Machine Berbasis Algoritma Firefly Pada Data Opini Film.” *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)* 5(5): 904–10.
- Styawati, Styawati, and Khabib Mustofa. 2019. “A Support Vector Machine-Firefly Algorithm for Movie Opinion Data Classification.” *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)* 13(3): 219–30.
- Susanto, Erliyan Redy, and Ajeng Savitri Puspaningrum. 2019. “Rancang Bangun Rekomendasi Penerima Bantuan Sosial Berdasarkan Data Kesejahteraan Rakyat.” 15(1): 1–12.
- Syah, Herwin, and Arita Witanti. 2022. “Analisis Sentimen Masyarakat Terhadap Vaksinasi Covid-19 Pada Media Sosial Twitter Menggunakan Algoritma Support

Vector Machine (Svm).” *Jurnal Sistem Informasi dan Informatika (Simika)* 5(1): 59–67.

Wibowo, Dellys Okta, and Adhie Thyo Priandika. 2021. “SISTEM PENDUKUNG KEPUTUSAN PEMILIHAN GEDUNG PERNIKAHAN PADA WILAYAH BANDAR LAMPUNG MENGGUNAKAN METODE TOPSIS.” *Jurnal Informatika dan Rekayasa Perangkat Lunak* 2(1): 73–84.

Yuliansyah, Aldi, and Mutiara Ayu. 2021. “The Implementation of Project-Based Assignment in Online Learning during Covid-19.” *Journal of English Language Teaching and Learning* 2(1): 32–38.

Yulianti, Devita Tri, Damayanti Damayanti, and Agung Tri Prastowo. 2021. “Pengembangan Digitalisasi Perawatan Kesehatan Pada Klink Pratama Sumber Mitra Bandar Lampung.” *Jurnal Teknologi dan Sistem Informasi* 2(2): 32–39.