

ANALISIS SENTIMEN KEBIJAKAN JAMINAN HARI TUA (JHT) PADA TWITTER MENGGUNAKAN NAÏVE BAYES

Sentiment Analysis of Old Age Guarantee Policy (JHT) On Twitter Using Naïve Bayes

Hendra Wijaya, hendra.wijaya095@gmail.com¹⁾, Lukman Hakim,
lhakim2710@gmail.com^{2)*}

^{1) 2)}Program Studi Informatika/Fakultas Teknologi dan Desain, Universitas Bunda Mulia

Diajukan 7 Juli 2022 / Disetujui 13 Maret 2023

ABSTRACT

This research is based on the policy of disbursement of JHT funds, which at this time has changed, because in the Regulation of the Minister of Manpower Number 2 of 2022 concerning Procedures and Requirements for Payment of Old Age Guarantee Benefits, workers cannot disburse JHT directly after experiencing termination of employment (LAYOFFS) or resigning from the job. This research uses the naïve bayes method to determine the sentiment of the public towards the policy. The goal is to find out the public opinion whether the public is pro and contra with the policy regarding the revision of the JHT disbursement rules. The results of this study received dataset input obtained from Twitter using the web scraping method, then the data went through the preprocessing stage. The results of this study received dataset input obtained from Twitter using the web scraping method, then the data went through the preprocessing stage. The data from preprocessing is divided into training data and testing data, after which the data is then analyzed using the naïve bayes method so as to produce positive sentiment and negative sentiment. From the test results using the confusion matrix, three tests were carried out using different amounts of test data, the best accuracy results were obtained using the number of test data of 10%, namely the accuracy value of 91.46%.

Keywords: *Confusion Matrix, JHT, Naïve Bayes, Twitter*

ABSTRAK

Penelitian ini dilatar belakangi kebijakan pencairan dana JHT yang pada saat ini terjadi perubahan, pasalnya dalam Peraturan Menteri Ketenagakerjaan Nomor 2 Tahun 2022 tentang Tata Cara dan Persyaratan Pembayaran Manfaat Jaminan Hari Tua itu, pekerja tidak bisa mencairkan JHT langsung setelah mengalami Pemutusan Hubungan Kerja (PHK) atau mengundurkan diri dari pekerjaan tersebut. Penelitian ini menggunakan metode *naïve bayes* untuk mengetahui sentiment dari masyarakat terhadap kebijakan tersebut. Tujuannya diharapkan dapat mengetahui opini public apakah masyarakat pro dan kontra dengan kebijakan mengenai revisi aturan pencairan dana JHT tersebut. Hasil dari penelitian ini menerima input dataset yang diperoleh dari twitter menggunakan metode *web scraping*, lalu data tersebut melalui tahapan *preprocessing*. Data hasil dari *preprocessing* dibagi menjadi data *training* dan data *testing*, setelah itu data kemudian dianalisis menggunakan metode *naïve bayes* sehingga menghasilkan sentimen positif dan sentimen negatif. Dari hasil pengujian menggunakan *confusion matrix*, dilakukan tiga kali pengujian menggunakan jumlah data *test* yang berbeda, hasil akurasi terbaik di dapatkan menggunakan jumlah data *test* 10% yaitu nilai akurasi 91,46%

Kata Kunci: *Confusion Matrix, JHT, Naïve Bayes, Twitter*

PENDAHULUAN

Kebijakan pencairan dana JHT pada saat ini terjadi perubahan, pasalnya dalam Peraturan Menteri Ketenagakerjaan Nomor 2 Tahun 2022 tentang Tata Cara dan Persyaratan Pembayaran Manfaat Jaminan Hari Tua itu, pekerja tidak bisa mencairkan JHT langsung setelah mengalami Pemutusan Hubungan Kerja (PHK) atau mengundurkan diri dari pekerjaan tersebut. Dalam aturan baru, peserta JHT hanya bisa mencairkan tabungannya ketika mencapai usia 56 tahun[1]. Dari

terbitnya peraturan tersebut mendapatkan respon dari berbagai pihak, misalnya buruh, anggota DPR (Dewan Perwakilan Rakyat), hingga masyarakat umum dengan menandatangani petisi pada [change.org](https://www.change.org) serta sosial media Twitter. Sosial media Twitter dijadikan tempat untuk memberikan pendapat masyarakat dalam bentuk *tweet* terkait kebijakan pemerintah tersebut. Banyak masyarakat yang tidak setuju terhadap kebijakan tersebut, namun ada juga masyarakat yang setuju karena kebijakan tersebut dirasa tepat. Sekitar 424.272 menolak revisi aturan pencairan dana JHT dengan menandatangani petisi dihalaman [change.org](https://www.change.org), Meski terdapat peraturan tersebut ada pihak yang mendukung peraturan terbaru, yaitu Kementerian Ketenagakerjaan (Kemnaker) dikarenakan peraturan tersebut dinilai dapat memperbaiki aturan Jaminan Hari Tua (CNN Indonesia, 2022).

Dari uraian diatas masalah yang dapat diidentifikasi yaitu banyaknya reaksi masyarakat yang dicurahkan dalam media sosial Twitter sulit untuk mengelompokan kecenderungan opini yang ada. Akan memakan waktu bila harus membaca satu persatu opini yang ditulis masyarakat.

Tujuan dari penelitian ini adalah untuk mengetahui pendapat masyarakat apakah pro dan kontra terhadap kebijakan tersebut dengan melakukan Analisis Sentimen, membuat aplikasi yang dapat melakukan analisis sentimen masyarakat terkait revisi kebijakan jaminan hari tua (JHT) pada Twitter serta mengimplementasikan metode *Naïve Bayes* dan mengetahui tingkat akurasi dalam sistem analisis sentimen masyarakat terkait revisi kebijakan pencairan dana jaminan hari tua (JHT) pada Twitter. Manfaat yang diharapkan dari penelitian ini yaitu sistem analisis sentimen ini akan digunakan pribadi sebagai sumber pengetahuan baru bagi peneliti. Dari hasil analisis sentimen ini dapat diketahui opini masyarakat terkait revisi kebijakan pencairan dana jaminan hari tua (JHT) pada Twitter, serta dapat mengetahui hasil akurasi dari metode *Naïve Bayes* dalam melakukan analisis sentimen masyarakat terhadap revisi kebijakan pencairan dana jaminan hari tua (JHT) pada Twitter berupa sentimen positif dan negatif.

METODOLOGI PENELITIAN

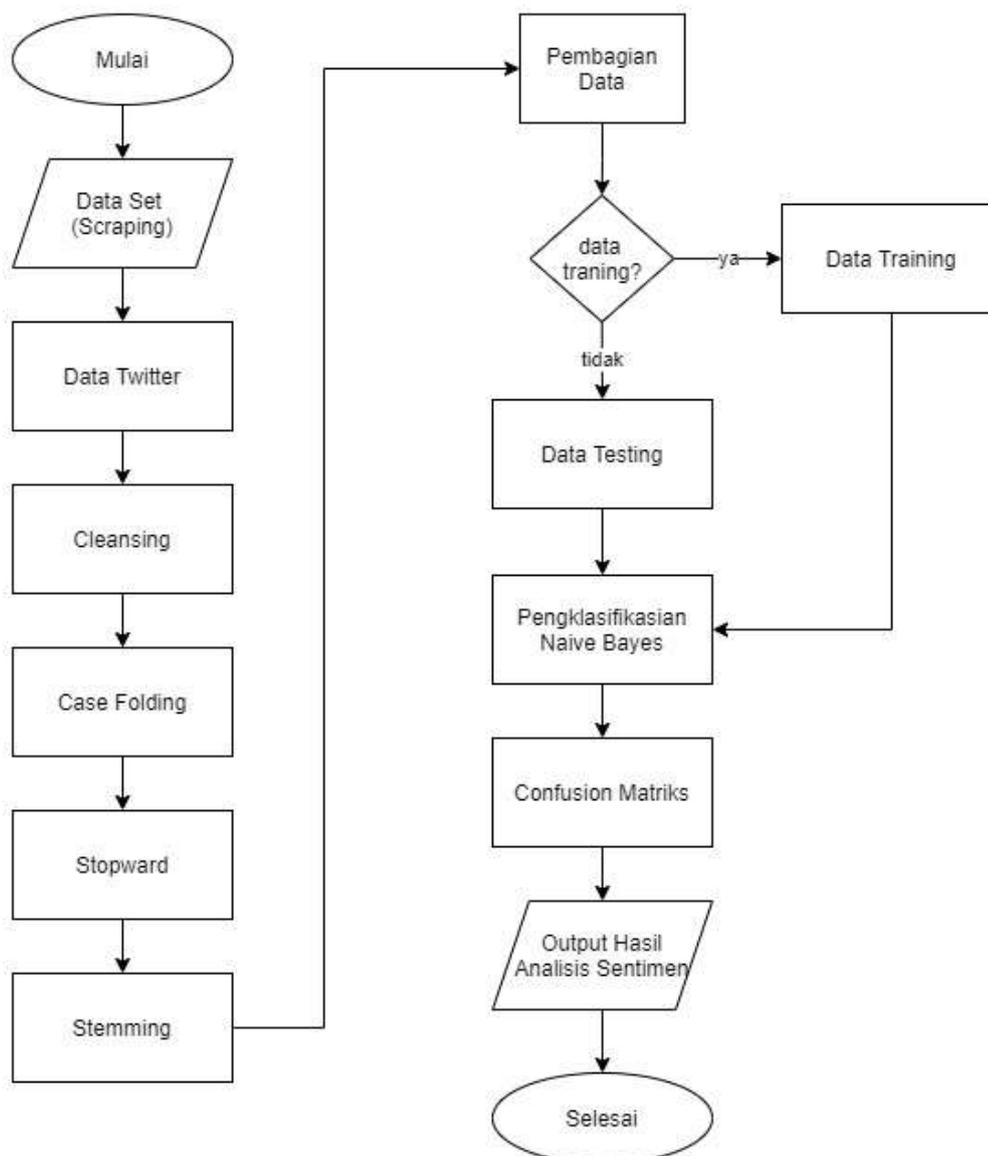
Metode penelitian yang diterapkan pada penelitian ini adalah dengan pengembangan metode *waterfall*. Alasan menggunakan metode *waterfall* adalah karena dalam pengembangannya bersifat sistematis dan sekuensial. Selain itu metode *waterfall* dilakukan secara berurutan dan berkelanjutan. Berikut beberapa tahapan-tahapan yang dilakukan dalam penelitian ini: *requirement* (analisis kebutuhan), desain sistem (*system design*), *Coding & Testing*, Penerapan Program, pemeliharaan.

Pemilihan algoritma untuk memfokuskan pemecahan masalah pada analisis sentimen twitter terkait Jaminan Hari Tua dengan menggunakan *Naive Bayes Classifier*. Dengan menggunakan *Naive Bayes Classifier* dapat menentukan jenis kata yang bersifat positif, atau negatif.. [2], [3]

Perancangan Proses Metode *Naïve Bayes* terdiri dari

1. Melakukan *Pre-Processing* menggunakan *Term Frequency*. *Term frequency* (TF) adalah metode pernghitungan jumlah kata kunci yang digunakan dengan token dari setiap kelas (klasifikasi).[4] Selain itu juga menggunakan *Document Frequency*. *Document frequency* adalah banyaknya kata kunci yang muncul dari dokumen.[4], [5] Berikutnya dilakukan *Inverse Document Frequency (IDF)*, dimana IDF merupakan pembobotan untuk setiap kata kunci yang dihasilkan.[5]–[7]
2. Setelah dilakukan tahapan TF-IDF maka dilakukan proses klasifikasi menggunakan metode *Naïve Bayes*. [8], [9]
3. Tahapan terakhir adalah melakukan evaluasi terhadap hasil klasifikasi dari *Naïve Bayes*, evaluasi dilakukan dengan menggunakan *confusion matrix*. [10] [11], [12]

Perancangan sistem dilakukan sebagai acuan pada saat proses implementasi baik implementasi metode *Naïve Bayes* maupun implementasi pembuatan perangkat lunak. Gambar 1 menjelaskan alur dari proses analisis sentimen menggunakan metode *Naïve Bayes*



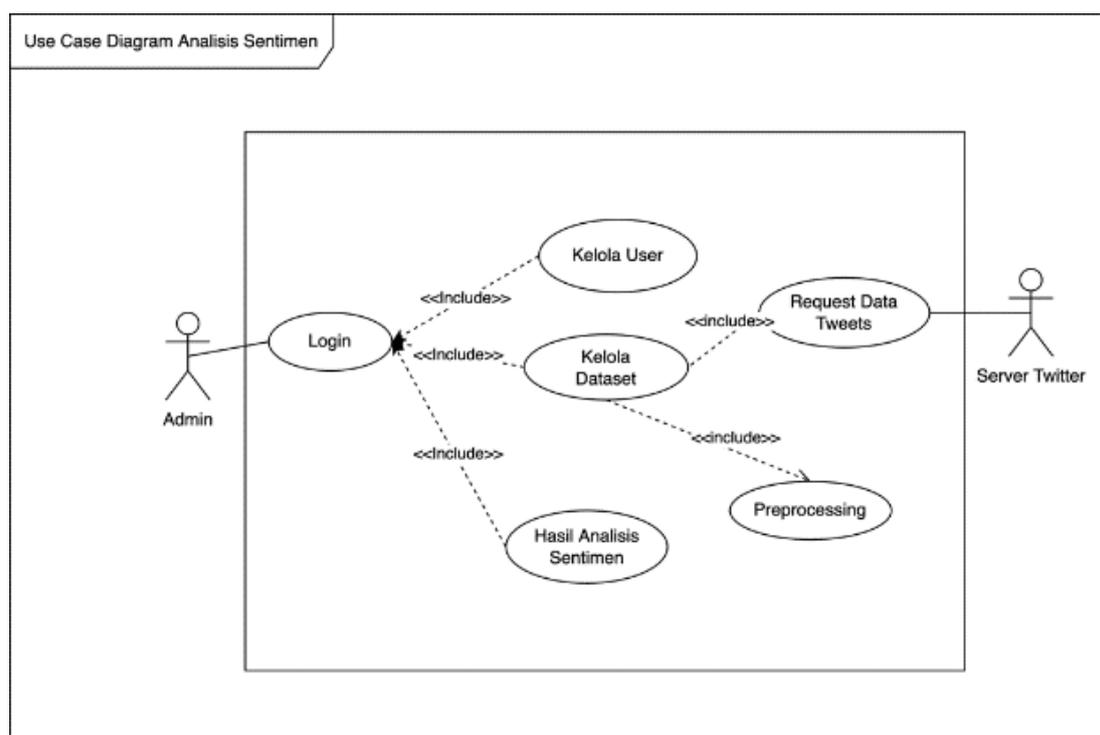
Gambar 1 Flowchart Sentimen Naïve Bayes

Berdasarkan alur sentimen untuk melakukan proses pengelompokkan menggunakan *Naïve Bayes*, hal pertama yang dilakukan adalah melakukan *data scraping* untuk mengumpulkan atau mengekstraksi informasi atau data semi-terstruktur dari website, yang menghasilkan data twitter mengenai kebijakan jaminan hari tua, lalu dilakukan proses *preprocessing* (*cleansing, case folding, stopword, stemming*), kemudian data dibagi dua menjadi *data testing*, dan *data training*, setelah data dibagi dua maka dilakukanlah pengklasifikasian *Naïve Bayes*, diikuti dengan *Confusion matrix* yang digunakan untuk menghitung nilai *accuracy* dan *error rate*, dan menampilkan hasil analisis sentimen.

Use Case Diagram ditunjukkan pada gambar 2. Terdapat dua aktor dalam penelitian ini yaitu admin dan juga server twiiter. Aktor user dan twitter memiliki karakteristik interaksi saat mengelola sistem. Adapun penjabarannya sebagai berikut:

1. Langkah pertama admin dapat melakukan terlebih dahulu login setelah itu memasukkan *username* dan *password*, kemudian setelah itu jika valid maka sistem dapat menampilkan halaman utama aplikasi.

2. Kemudian admin dapat melakukan kelola *user* untuk mengelola data *user* sesuai dengan hak aksesnya masing – masing.
3. Setelah itu admin dapat menambahkan dataset yang akan digunakan untuk proses klasifikasi, dataset yang digunakan berasal dari data twitter. Pada saat setelah penambahan dataset sistem akan otomatis melakukan pengolahan kata yaitu tahap *preprocessing* sebelum dilakukan klasifikasi.
4. Kemudian sistem secara otomatis melakukan perhitungan menggunakan *Naïve Bayes*, dan admin dapat melihat hasil analisis sentimennya.
5. Sedangkan server twitter, *usecase* untuk server twitter adalah data twiiter, data twitter ini juga dapat diakses oleh admin untuk kepentingan sistem yang akan dibuat.



Gambar 2. Use Case Diagram

HASIL DAN PEMBAHASAN

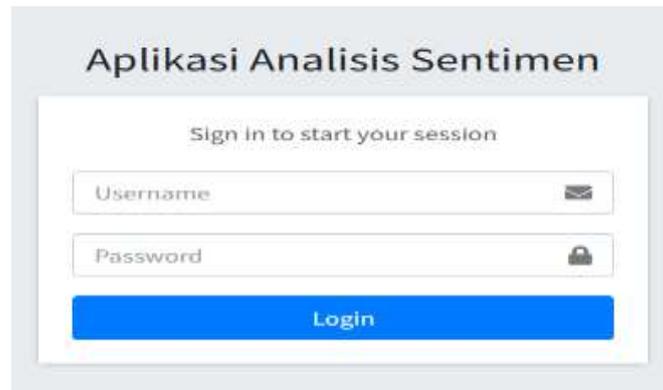
Adapun hasil implementasi sistem pada penelitian ini sebagai berikut:

1. Halaman *Login*

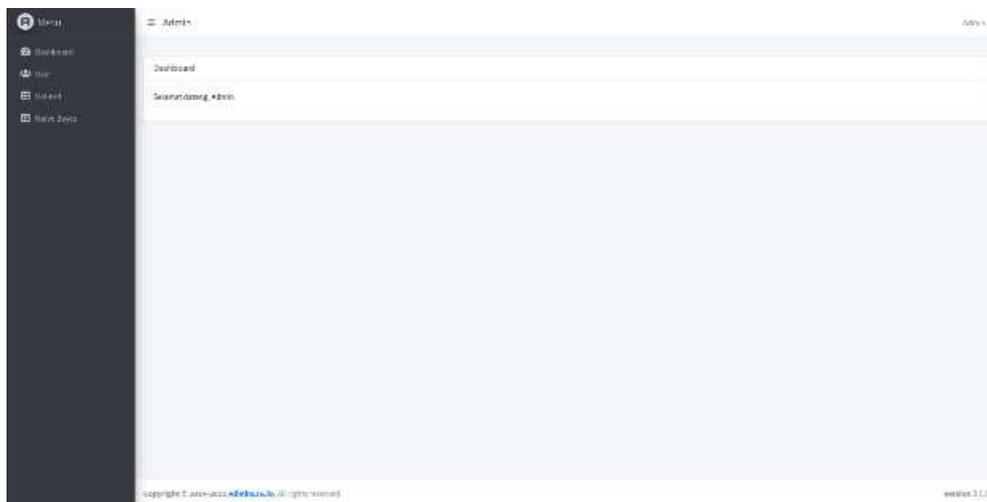
Halaman *login* merupakan halaman yang digunakan untuk masuk ke dalam sistem. Pada halaman ini pengguna pertama kali sebelum melakukan proses *login* ke dalam sistem terlebih dahulu menginputkan *username* dan *password*. Berikut tampilan halaman *login* dapat dilihat pada Gambar 3.

2. Halaman *Dashboard*

Halaman *dashboard* merupakan menu tampilan utama ketika pengguna pertama kali masuk kedalam sistem. Pada halaman *dashboard* terdapat beberapa menu yang dapat digunakan oleh pengguna. Adapun tampilan halaman *dashboard* dapat dilihat pada Gambar 4.



Gambar 3. Halaman Login



Gambar 4. Halaman Dashboard

3. Halaman User

Halaman *user* merupakan halaman yang dapat digunakan oleh admin untuk mengelola data pengguna. Adapun tampilan halaman *user* dapat dilihat pada Gambar 5



Gambar 5. Halaman User

4. Halaman Dataset

Halaman dataset menampilkan halaman yang dapat digunakan oleh pengguna untuk melakukan kelola dataset. Adapun tampilan halaman dataset dapat dilihat pada Gambar 6

No	Username	Nama	Tweet	Status
1	ScrapTwee	ScrapTwee	RT @Gawir_Melusi: Persepsi... (text truncated)	1
2	ScrapTwee	ScrapTwee	RT @Gawir_Melusi: Persepsi... (text truncated)	1
3	ScrapTwee	ScrapTwee	RT @Gawir_Melusi: Persepsi... (text truncated)	1
4	ScrapTwee	ScrapTwee	RT @Gawir_Melusi: Persepsi... (text truncated)	1
5	ScrapTwee	ScrapTwee	RT @Gawir_Melusi: Persepsi... (text truncated)	1
6	ScrapTwee	ScrapTwee	RT @Gawir_Melusi: Persepsi... (text truncated)	1
7	ScrapTwee	ScrapTwee	RT @Gawir_Melusi: Persepsi... (text truncated)	1
8	ScrapTwee	ScrapTwee	RT @Gawir_Melusi: Persepsi... (text truncated)	1
9	ScrapTwee	ScrapTwee	RT @Gawir_Melusi: Persepsi... (text truncated)	1
10	ScrapTwee	ScrapTwee	RT @Gawir_Melusi: Persepsi... (text truncated)	1

Gambar 6. Halaman Dataset

Pada halaman dataset terdapat *icon* tambah dataset. *Icon* ini berfungsi untuk menampilkan *form* yang digunakan untuk mengupload dataset. Berikut tampilan *form* tambah dataset dapat dilihat pada Gambar 7

Gambar 7 Tambah Dataset

5. Halaman Naïve Bayes

Halaman *Naïve Bayes* berfungsi untuk menampilkan perhitungan sentimen dari dataset yang digunakan. Pada halaman ini terdapat beberapa tampilan seperti data *training* dan data *testing*. Adapun tampilan halaman pembagian data *training* dan *testing* dapat dilihat pada Gambar 8 dan Gambar 9.

Pada halaman ini, terdapat tampilan hasil klasifikasi dari perhitungan menggunakan *Naïve Bayes*. Adapun tampilan halaman perhitungan *Naïve Bayes* dapat dilihat pada Gambar 10

Hasil dan pembahasan didukung dengan data dan menggunakan uji dan indikator yang umum digunakan di bidang keilmuan masing-masing. Pada dalam pembahasan dikemukakan keterkaitan antara hasil penelitian dengan teori, tujuan penelitian dan perbandingan dengan penelitian lain yang telah dipublikasikan. Pembahasan juga menjelaskan implikasi/kontribusi temuan bagi ilmu pengetahuan. Hasil dan pembahasan diungkapkan secara jelas dan lugas menggunakan kalimat sederhana. Hasil dapat ditampilkan dalam grafik ataupun tabel.

Hasil klasifikasi *Naïve Bayes* dilakukan pengujian dengan jumlah data test yang digunakan berbeda yaitu data test 30%, data test 20% dan data test 10%. Hal ini dilakukan untuk mengetahui apakah jumlah data test akan mempengaruhi akurasi yang di dapatkan

Tabel 1. Hasil Uji

	Data Test 30%	Data test 20%	Data Test 10%
Akurasi	88,94%	90,95%	91,46%
<i>Precision</i>	96,88%	97,96%	97,08%
<i>Recall</i>	90,86%	92,05%	93,26%
<i>Error rate</i>	10,06%	8,63%	7,23%
<i>F-Measure</i>	93,77%	94,92%	95,13%

Berdasarkan dari tabel 1 dapat dilihat bahwa penggunaan jumlah data test mempengaruhi nilai akurasi yang didapatkan. Semakin banyak jumlah data test yang digunakan maka semakin rendah tingkat akurasi yang didapatkan, hal ini disebabkan oleh adanya ketidakcocokan data test dengan data train yang ada dalam sistem. Semakin banyak data test yang memiliki kecocokan dengan data train maka tingkat akurasinya akan semakin tinggi.

SIMPULAN

Simpulan

Berdasarkan perancangan, hasil implementasi dan pengujian yang dilakukan pada penelitian ini, maka dapat disimpulkan sebagai berikut:

- Aplikasi analisis sentimen kebijakan Jaminan Hari Tua (JHT) pada twitter berhasil dibuat menggunakan bahasa pemrograman python dan MySQL sebagai databasenya.
- Metode *Naïve Bayes* berhasil diimplementasikan pada analisis sentiment terkait dengan kebijakan Jaminan Hari Tua (JHT) dimana metode ini menerima input dari dataset yang diperoleh dari twitter menggunakan metode *web scraping*, lalu data tersebut melalui tahapan *preprocessing*. Data hasil dari *preprocessing* dibagi menjadi data *training* dan data *testing*, setelah itu data kemudian dianalisis menggunakan metode *naïve bayes* sehingga menghasilkan klasifikasi sentimen positif dan negatif. Kemudian data yang diperoleh dari hasil metode *naïve bayes* dihitung kembali menggunakan perhitungan *confusion matrix*, dilakukan tiga kali

pengujian menggunakan jumlah data *test* yang berbeda, hasil akurasi terbaik di dapatkan menggunakan jumlah data *test* 10% yaitu nilai akurasi 91,46%.

Saran

Adapun saran yang diberikan untuk melakukan tahapan pengembangan selanjutnya adalah sistem dapat dikembangkan dengan menerapkan metode lain kemudian dihitung nilai akurasi sebagai perbandingan metode mana yang lebih baik dalam melakukan analisis sentimen.

DAFTAR PUSTAKA

- [1] A. I. Falah, Rachmat Febfauza, and Noor Kumala Sari, "Polemik Penetapan Kebijakan Baru mengenai Jaminan Hari Tua (JHT) di Indonesia Tahun 2022," *Jurnal Administrasi Publik*, vol. 18, no. 1, pp. 69–96, Jun. 2022, doi: 10.52316/jap.v18i1.90.
- [2] F. S. Pattiiha, "Perbandingan Metode K-NN , Naïve Bayes , Decision Tree untuk Analisis Sentimen Tweet Twitter Terkait Opini Terhadap PT PAL Indonesia," *JURIKOM (Jurnal Riset Komputer)*, vol. 9, no. 2, pp. 506–514, 2022, doi: 10.30865/jurikom.v9i2.4016.
- [3] Edwin, "APLIKASI ASSESMENT KEBIJAKAN PEMERINTAH TERKAIT OPERASI OJEK ONLINE DI MASA PANDEMI (COVID-19) MENGGUNAKAN ALGORITMA NAÏVE BAYES CLASSIFIER Application Of Government Policy Assessment Related To Online Oject Operations In The Pandemic Time (COVID-19)," *Jurnal Algoritma, Logika dan Komunikasi*, vol. III, no. 2, pp. 299–307, 2020, [Online]. Available: <https://journal.ubm.ac.id/index.php/alu/article/view/2651>
- [4] Ida Widaningrum, Dyah Mustikasari, Rizal Arifin, Siti Lathifah Tsaqila, and Dwiyunia Fatmawati, "Algoritma Term Frequency – Inverse Document Frequency (TF-IDF) dan K-Means Clustering Untuk Menentukan Kategori Dokumen," *Prosiding SISFOTEK*, vol. 6, no. 1, pp. 145–149, Oct. 2022, [Online]. Available: <https://seminar.iaii.or.id/index.php/SISFOTEK/article/view/349>
- [5] M. P. Simatupang and D. P. Utomo, "ANALISA TESTIMONIAL DENGAN MENGGUNAKAN ALGORITMA TEXT MINING DAN TERM FREQUENCY-INVERSE DOCUMENT FREQUENCE (TF-IDF) PADA TOKO ALLMEEART," *KOMIK (Konferensi Nasional Teknologi Informasi dan Komputer)*, vol. 3, no. 1, Dec. 2019, doi: 10.30865/komik.v3i1.1697.
- [6] A. N. Rohman, R. D. Handayani, R. D. Y. P., and K. Kusriani, "DETEKSI EMOSI MEDIA SOSIAL MENGGUNAKAN TERM FREQUENCY- INVERSE DOCUMENT FREQUENCY," *CSRID (Computer Science Research and Its Development Journal)*, vol. 11, no. 3, p. 140, Mar. 2021, doi: 10.22303/csrid.11.3.2019.140-148.
- [7] T. Tan, S. Gao, W. Yang, Y. Song, and C. Lin, "Two New Term Weighting Methods for Router Syslogs Anomaly Detection," in *2016 IEEE 18th International Conference on High Performance Computing and Communications; IEEE 14th International Conference on Smart City; IEEE 2nd International Conference on Data Science and Systems (HPCC/SmartCity/DSS)*, IEEE, Dec. 2016, pp. 1454–1460. doi: 10.1109/HPCC-SmartCity-DSS.2016.0207.

- [8] S. Styawati and K. Mustofa, "A Support Vector Machine-Firefly Algorithm for Movie Opinion Data Classification," *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, vol. 13, no. 3, p. 219, Jul. 2019, doi: 10.22146/ijccs.41302.
- [9] R. A. Rizal, M. Susanto, and A. Chandra, "Classification Of Borax Content In Tomato Sauce Through Images Using GLCM," *Sinkron*, vol. 4, no. 2, p. 6, Mar. 2020, doi: 10.33395/sinkron.v4i2.10508.
- [10] T. M. S. Mulyana, D. Widyaningrum, and H. Herlina, "OCR HURUF JAWA DENGAN FITUR KODE RANTAI DAN LEVENSHTTEIN DISTANCE," *Network Engineering Research Operation*, vol. 6, no. 1, p. 67, Apr. 2021, doi: 10.21107/nero.v6i1.217.
- [11] C. Manliguez, "Generalized Confusion Matrix for Multiple Classes." 2016. doi: 10.13140/RG.2.2.31150.51523.
- [12] A. M. Glazer, "Confusion over the description of the quartz structure yet again," *J Appl Crystallogr*, vol. 51, pp. 915–918, 2018, doi: 10.1107/S160057671800434X.