PERBANDINGAN METODE SELEKSI FITUR UNTUK KLASIFIKASI PERTANYAAN BERBAHASA INDONESIA MENGGUNAKAN ALGORITMA SUPPORT VECTOR MACHINE (SVM)

Diajukan Sebagai Syarat Untuk Menyelesaikan
Pendidikan Program Strata-1 Pada
Jurusan Teknik Informatika

Oleh:

SYECHKY AL QODRIN ARUDA
NIM: 09021381823120

Jurusan Teknik Informatika

FAKULTAS ILMU KOMPUTER UNIVERSITAS SRIWIJAYA

2022

LEMBAR PENGESAHAN SKRIPSI

PERBANDINGAN METODE SELEKSI FITUR UNTUK KLASIFIKASI PERTANYAAN BERBAHASA INDONESIA MENGGUNAKAN ALGORTIMA *SUPPORT VECTOR MACHINE* (SVM)

Oleh :

SYECHKY AL QODRIN ARUDA
NIM: 09021381823120

Palembang, 17 Juni 2022

Pembimbing I,

Novi Yusliani, M.T.
NIP. 19821082012122001

Pembimbing II,

Alvi Syahrini Utami, M.Kom.
NIP. 197812222006042003

Mengetahui,
Ketua Jurusan Teknik Informatika

Alvi Syahrini Utami, M.Kom.
NIP. 197812222006042003

TANDA LULUS UJIAN KOMPREHENSIF SKRIPSI

Pada hari Rabu tanggal 17 Juni 2022 telah dilaksanakan ujian komprehensif skripsi oleh Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.

Nama : Syechky Al Qodrin Aruda
NIM : 09021381823120
Judul : Perbandingan Metode Seleksi Fitur untuk Klasifikasi Pertanyaan Berbahasa Indonesia Menggunakan Algortima *Support Vector Machine* (SVM)

dan dinyatakan **LULUS.**

1. Ketua

   Yunita, M.Cs.
   NIP. 198306062015042002 .....

2. Penguji I

   Dr. Abdiansyah, S. Kom., M.Cs.
   NIP. 198410012009121005 .....

3. Penguji II

   Mastura Diana Marieska, S.T., M.T.
   NIP. 198603212018032001 .....

4. Pembimbing I

   Novi Yusliani, M.T.
   NIP. 198211082012122001 .....

5. Pembimbing II

   Alvi Syahrini Utami, M.Kom
   NIP. 197812222006042003 .....

Mengetahui,
Ketua Jurusan Teknik Informatika

Alvi Syahrini Utami, M.Kom.
NIP. 197812222006042003

HALAMAN PERNYATAAN

Yang bertanda tangan dibawah ini:

Nama	:	Syechky Al Qodrin Aruda
NIM	:	09021381823120
Program Studi	:	Teknik Informatika
Judul	:	Perbandingan Metode Seleksi Fitur untuk Klasifikasi Pertanyaan Berbahasa Indonesia Menggunakan Algoritma *Support Vector Machine* (SVM)

Hasil Pengecekan *Software Ithenticate/Turnitin* : 15 %

Menyatakan bahwa laporan tugas akhir saya merupakan hasil karya sendiri dan bukan hasil penjiplakan/plagiat. Apabila ditemukan unsur penjiplakan atau plagiat dalam laporan tugas akhir ini, maka saya bersedia menerima sanksi akademik dari Universitas Sriwijaya.

Demikian, pernyataan ini saya buat dengan sebenarnya dan tidak ada paksaan dari siapapun.

Palembang, 24 Mei 2022

Syechky Al Qodrin Aruda
NIM. 09021381823120

ABSTRACT

Most texts have a large number of features. However, the features contained in the text mostly have a low level of relevance and even caontain noise which can later reduce the accuracy of the results. Feature selection is used to reduce the dimensions of feature space by weighting all features then features with lower weights than treshold will be eliminated. It aims to improve the accuracy and efficiency of computational time in the text classification process. In this research, selection method Information Gain, Chi Square, Mutual Information were used in the text classification process in the form of Indonesian questions using the Support Vector Machine (SVM) algorithm. Then, a comparative analysis will be carried out on each classification model based on the evaluation results obtained. The results showed that the use of the feature selection method was able to increase accuracy and reduce computation time.The use of the Chi Square feature selection method on the SVM algorithm with a linear kernel and parameter C:1 give the best performance with average of accuracy 0.92, precision 0.93, recall 0.89, f-measure 0.91 and computation time 8 seconds.

Key Word : Text Classification, Number of Features, Feature Selection, Support Vector Machine (SVM)

ABSTRAK

Sebagian besar teks memiliki jumlah fitur yang banyak. Namun, fitur yang terdapat pada teks sebagian besar memiliki tingkat relevansi yang kurang bahkan mengandung *noise* yang nantinya dapat mengurangi hasil akurasi. Seleksi fitur digunakan untuk mengurangi dimensi ruang fitur dengan cara melakukan pembobotan pada semua fitur kemudian fitur dengan bobot yang kurang dari ambang batas akan dieliminasi. Hal ini bertujuan untuk meningkatkan akurasi serta efisiensi waktu komputasi pada proses klasifikasi teks. Pada penelitian ini, metode seleksi *Information Gain, Chi Square* dan *Mutual Information* digunakan pada proses klasifikasi teks berupa pertanyaan berbahasa Indonesia menggunakan algoritma *Support Vector Machine* (SVM). Kemudian, akan dilakukan analisis perbandingan pada setiap model klasifikasi berdasarkan hasil evaluasi yang didapat. Hasil penelitian menunjukan penggunaan metode seleksi fitur mampu memberikan peningkatan akurasi serta mengurangi waktu komputasi. Penggunaan metode seleksi fitur *Chi Square* pada algortima SVM dengan kernel linear dan parameter C: 1 menghasilkan kinerja terbaik dengan rata-rata *accuracy* 0.92, *precision* 0.93, *recall* 0.89, *f-measure* 0.91 dan waktu komputasi 8 detik.

Kata Kunci : Klasifikasi Teks, Jumlah Fitur, Seleksi Fitur, *Support Vector Machine*

KATA PENGANTAR

Segala puji bagi Allah Subhanahu Wa Ta'ala karna atas karunia dan rahmat-nya sehingga penulis dapat menyelesaikan Skripsi dengan judul "**Perbandingan Metode Seleksi Fitur Untuk Klasifikasi Pertanyaan Berbahasa Indonesia Menggunakan Algoritma *Support Vector Machine* (SVM)**". Penulisan Skripsi ini ditujukkan untuk melengkapi salah satu syarat dalam menyelesaikan pendidikan dan mendapatkan gelar Sarjana Komputer di Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.

Atas selesainya skripsi ini, penulis mengucapkan rasa syukur kepada Allah SWT. Dan penulis menyampaikan rasa terima kasih kepada yang terhormat :

1. Kedua Orang Tua saya yaitu Rudi Hartono dan Ida Hariani, saudari dan saudara saya yaitu Oktefvia Aruda Lisjana dan Ghaniyah Al Aflah Aruda, dan semua keluarga besar penulis yang sangat saya cintai. Terima kasih untuk semua doa yang telah dipanjatkan dan terimakasih juga untuk semua dukungan dan bantuan yang telah diberikan.
2. Bapak Jaidan Jauhari, S.Pd., M.T., selaku Dekan Fakultas Ilmu Komputer Universitas Sriwijaya.
3. Ibu Alvi Syahrini Utami, M.Kom, selaku Ketua Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.
4. Bapak Danny Matthew, S.T., M.Sc., selaku Dosen Pembimbing Akademik di Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.
5. Ibu Novi Yusliani, M.T. dan Ibu Alvi Syahrini Utami, M.Kom, selaku Dosen Pembimbing Skripsi di Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.
6. Ibu Lestarini, S.SI., M.T.selaku ketua lab Basis Data dan Big Data dan kak yogi selaku admin lab fasilkom unsri bukit yang telah memberi saya fasillitas untuk membantu pengerjaan skripsi saya.
7. Mba Wiwin selaku admin Jurusan Teknik Informatika yang telah membantu mengurus berkas administrasi penulis.

8. Seluruh dosen dan staff Fakultas Ilmu Komputer Universitas Sriwijaya.
9. Kak Prayogi, Kak Cesil dan Kak Saniyah yang telah membantu dalam pengerjaan Skripsi ini.
10. Raisha Fatiya, Rafliandi Ardana, Muhammad Yasykur Lutfi, Sultan Alfarid, Muhammad Febriansyah, Altundri Wahyu, Muhammad Farhan, Kgs. M. Rusdiansyah Muharrom, Nisa Auli, Nopriyansyah, Acilla, Arry Erpapalemlah, Aqil Citrayasa, Tyansyah, Muhammad Raihan, Viva Andharsyah dan teman-teman IFBILB serta seluruh teman-teman seperjuangan angkatan 2018 Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.
11. Kak Egi, Kak Salim, Kak Eka, Yuni, Amel, Kak Kia dan teman-teman seperjuangan saya di Generasi Baru Indonesia yang sangat menginspirasi saya untuk selalu menjadi yang terbaik.

Penulis menyadari bahwa masih banyak keterbatasan dan kekurangan yang ada dalam penulisan Skripsi ini. Maka dari itu, segala kritik dan saran sangat penulis butuhkan agar dapat menghasikan karya tulis yang lebih baik. Sekian yang dapat penulis sampaikan, semoga skripsi ini dapat berguna dan bermanfaat bagi banyak orang.

DAFTAR ISI

Halaman

DAFTAR TABEL

DAFTAR GAMBAR

DAFTAR LAMPIRAN

BAB I
PENDAHULUAN

1.1 Pendahuluan

Pada bab ini akan membahas mengenai latar belakang, rumusan masalah, tujuan dan manfaat penelitian, batasan masalah serta sistematika penulisan skripsi. Bab ini akan memberikan penjelasan umum mengenai pokok pikiran dari keseluruhan penelitian.

1.2 Latar Belakang

Kalimat tanya merupakan serangkaian kata yang ditujukan untuk memperoleh informasi atau jawaban yang ditujukan kepada seseorang. Untuk mendapatkan suatu jawaban atau informasi yang tepat maka diperlukan suatu analisa dalam memahami kalimat tanya tersebut. Proses klasifikasi dokumen pertanyaan secara otomatis dapat dipergunakan untuk mempermudah melakukan analisa terkait kategori suatu konteks pertanyaan berbahasa Indonesia. Berdasarkan pembagian pada analisa pertanyaan, kalimat tanya sendiri terbagi menjadi 3 jenis pertanyaan yaitu pertanyaan *factoid, non-factoid* dan *other*. Untuk pertanyaan *factoid* didefinisikan sebagai pertanyaan yang hanya membutuhkan jawaban singkat berupa kata atau frasa, seperti nama orang, lokasi, organisasi, jumlah dan tanggal. Untuk pertanyaan *non-factoid* didefinisikan sebagai pertanyaan yang membutuhkan penjelasan seperti pertanyaan yang terkait dengan definisi, alasan, metode dan tatacara (Purwarianti & Yusliani, 2011). Sedangkan untuk pertanyaan diluar dari *factoid* dan *non-factoid* dibuat kategori tersendiri yaitu *other*.

Masalah utama pada klasifikasi teks adalah sebagian besar teks memiliki dimensi fitur yang banyak, sedangkan sebagian besar fitur ini tidak relevan dan bahkan mengandung *noise* yang dapat mengurangi tingkat akurasi klasifikasi (Chandra, 2019). Oleh karena itu *feature selection* umumnya digunakan dalam klasifikasi teks untuk mengurangi dimensi ruang fitur dan meningkatkan efisiensi dan akurasi pengklasifikasi. Terdapat beberapa metrik seleksi fitur yang terkenal seperti *Information Gain* (IG), *Chi-Square*, dan *Mutual Information* (Khan et al., 2010).

Tentunya masing-masing metode seleksi fitur tersebut memiliki kelebihan dan kekurangan tersendiri. Menurut (Rahmad & Pribadi, 2015), Metode *Chi-Square* dapat meningkatkan hasil kinerja klasifikasi seperti *Recall, Precision, F-measure*, dan *Accuracy*. Metode ini bekerja dengan melakukan pehitungan statistic dalam pemilihan seleksi fitur. Penggunaan *treshold* sangat berpengaruh terhadap jumlah fitur yang didapat. (Maulida et al., 2016) dalam penelitiannya menjelaskan, metode *Information Gain* dapat mengurangi dimensi fitur pada dokumen berbahasa Indonesia dan metode ini menerapkan teknik *scoring* dalam melakukan pembobotan menggunakan maksimal *entropy*. Selain itu, (Irham et al., 2019) dalam penelitiannya menjelaskan metode *Mutual Information* membantu mengeliminasi fitur yang tidak menginterpretasikan sebuah kelas. Sehingga metode ini dapat meningkatkan kecepatan dan efektifitas hasil kinerja dalam klasifikasi.

Terdapat beberapa penelitian yang telah dilakukan sebelumnya berkaitan dengan komparasi fitur seleksi pada klasifkasi dokumen. Dalam penelitian (Chandani, 2015) peneliti melakukan perbandingan algoritma *machine learning*

(*Artificial Neural Network* (ANN), *Support Vector Machine* (SVM), *Naïve Bayes* (NB)) dan juga membandingkan seleksi fitur (*Information Gain, Chi-Square, Forward Selection dan Backward Selection*). Hasil dari penelitian tersebut didapati bahwa SVM sebagai algoritma *machine learning* terbaik dan *Information Gain* sebagai metode seleksi fitur terbaik. Selain itu, (Bahassine et al., 2020) juga melakukan perbandingan terhadap algoritma *machine learning* (*Decision Tree* dan SVM) dan metode seleksi fitur (ImpCHI, *Chi-Square*, IG, MI) untuk klasifikasi dokumen berbahasa arab. Hasil penelitian menunjukkan bahwa penggunaan seleksi fitur *Improvement Chi-Square* menunjukkan hasil kinerja terbaik diikuti *Information Gain* dan *Chi-Square*. Kemudian juga didapati bahwa algoritma SVM menunjukkan hasil kerja yang lebih baik dibandingkan dengan algoritma *Decision Tree*.

Berdasarkan uraian dan referensi penelitian sebelumnya, maka dari itu, akan dilakukan perbandingan terhadap 3 metode seleksi fitur yaitu *Information Gain, Chi-Square,* dan *Mutual Information* pada klasifikasi pertanyaan berbahasa Indonesia menggunakan algoritma *Support Vector Machine* untuk mengetahui perbandingan hasil kinerja tiga metode seleksi fitur tersebut. Hasil penelitian ini diharapkan dapat menjadi rujukan bagi pengembang sistem pertanyaan berbahasa Indonesia dan memberikan solusi dalam melakuan optimalisasi hasil kinerja pada klasifikasi pertanyaan berbahasa Indonesia menggunakan algoritma klasifikasi *Support Vector Machine*.

1.3 Rumusan Masalah

Berdasarkan latar belakang permasalahan yang telah dijabarkan maka

rumusan masalah dari penelitian ini:

1. Bagaimana mengembangkan sistem klasifikasi pertanyaan berbahasa Indonesia menggunakan algoritma SVM ?
2. Bagaimana kinerja metode seleksi fitur *Information Gain, Chi-Square*, dan *Mutual Information* dalam melakukan pemilihan fitur pada sistem klasifikasi pertanyaan berbahasa Indonesia menggunakan algoritma SVM ?

1.4 Tujuan Penelitian

Adapun tujuan dari penelitian ini adalah sebagai berikut :

1. Menghasilkan perangkat lunak sistem klasifikasi pertanyaan Berbahasa Indonesia menggunakan algoritma SVM.
2. Mengetahui kinerja *Information Gain, Chi-Square* dan *Mutual Information* dalam melakukan seleksi fitur pada sistem klasifikasi pertanyaan berbahasa Indonesia menggunakan algoritma SVM.

1.5 Manfaat Penelitian

Adapun manfaat dari penelitian ini adalah sebagai berikut :

1. Mendapatkan informasi mengenai kinerja dan hasil klasifikasi pertanyaan berbahasa Indonesia dengan membandingkan seleksi fitur *Information Gain, Chi-Square, Mutual Information* pada metode klasifikasi *Support Vector Machine*.
2. Mengetahui metode seleksi fitur yang paling optimal dalam melakukan klasifikasi pertanyaan berbahasa Indonesia.
3. Hasil penelitan dapat menjadi rujukan penelitian yang relevan.

1.6 Batasan Masalah

Adapun batasan masalah pada penelitian ini adalah sebagai berikut :

1. Data yang digunakan merupakan data pertanyaan berbahasa Indonesia.
2. Data pertanyaan berbahasa Indonesia yang telah dikumpulkan berjumlah 1.195 pertanyaan berbahasa Indonesia yang terbagi menjadi 553 data berlabel '*factoid*', 451 data berlabel '*non-factoid*', dan 185 data berlabel '*others*'.
3. Klasifikasi terdiri dari 3 kelas, yaitu *factoid, non-factoid* dan *others.*
4. Metode seleksi fitur yang dibandingkan terdiri dari 3 metode seleksi fitur bertipe *filtering* yaitu *Information Gain, Chi Square* dan *Mutual Information.*

1.7 Sistematika Penulisan

Adapun sistematika penulisan pada penelitian ini adalah sebagai berikut:

BAB I. PENDAHULUAN

Bab ini akan membahas landasan dari penelitian, seperti latar belakang, rumusan masalah, tujuan dan manfaat penelitian, batasan masalah serta sistematika penulisan.

BAB II. KAJIAN LITERATUR

Bab ini membahas dasar-dasar teori yang berkaitan dengan penelitian ini, seperti pertanyaan Berbahasa Indonesia, jenis pertanyaan *factoid* dan *non-factoid*, seleksi fitur *Information gain*, seleksi fitur *Chi-square*, seleksi fitur *Mutual Information* dan algoritma *Support Vector Machine*, serta membahas beberapa

penelitian yang relevan.

BAB III. METODOLOGI PENELITIAN

Bab ini membahas mengenai tahapan alur penelitian. Diantaranya pengumpulan data dan perancangan perangkat lunak yang akan dibangun.

BAB IV. PENGEMBANGAN PERANGKAT LUNAK

Bab ini membahas mengenai analisa dan rancangan pengembangan sistem perangkat lunak. Diawali dengan kebutuhan analisis, perancangan dan konstruksi, kemudian diakhiri dengan melakukan pengujian.

BAB V. HASIL DAN ANALISA SARAN

Bab ini menguraikan hasil pengujian berdasarkan perancangan. Tabel hasil evaluasi pengujian dan analisis serta grafik menjadi patokan dari kesimpulan yang akan diambil dalam penelitian

BAB VI. KESIMPULAN DAN SARAN

Bab ini membahas mengenai kesimpulan berdasarkan semua uraian pada bab sebelumnya dan juga saran yang diberikan dari hasil penelitian.

1.8 Kesimpulan

Pada Bab ini menguraikan dasar dan patokan pada penelitian, seperti latar belakang, rumusan masalah, tujuan penelitian, manfaat penelitian, batasan masalah dan sistematika penulisan.

DAFTAR PUSTAKA

Alita, D., Fernando, Y., & Sulistiani, H. (2020). Implementasi Algoritma Multiclass Svm Pada Opini Publik Berbahasa Indonesia Di Twitter. *Jurnal Tekno Kompak*, *14*(2), 86. https://doi.org/10.33365/jtk.v14i2.792

Amrullah, A. Z., Anas, S. A., Hidayat, & Muh. Adrian. (2020). Analisis Sentimen Movie Review Menggunakan Naive Bayes Classifier Dengan Seleksi Fitur Chi Square. *Jurnal*, *2*(1), 40–44. https://doi.org/10.30812/bite.v2i1.804

Bahassine, S., Madani, A., Al-sarem, M., & Kissi, M. (2020). Feature selection using an improved Chi-square for Arabic text classification. *Journal of King Saud University - Computer and Information Sciences*, *32*(2), 225–231. https://doi.org/10.1016/j.jksuci.2018.05.010

Bramer, M. (2007). Principles of Data Mining. In *Principles of Data Mining* (Issue January 2007). https://doi.org/10.1007/978-1-84628-766-4

Buani, D. (2021). *Penerapan Algoritma Naïve Bayes dengan Seleksi Fitur Algoritma Genetika Untuk Prediksi Gagal Jantung*. *9*(2), 43–48.

Chandani, V. (2015). *Komparasi Algoritma Klasifikasi Machine Learning Dan Feature Selection pada Analisis Sentimen Review Film*. *1*(1), 56–60.

Chandra, A. (2019). Comparison of Feature Selection for Imbalance Text Datasets. *Proceedings of 2019 International Conference on Information Management and Technology, ICIMTech 2019, August*, 68–72. https://doi.org/10.1109/ICIMTech.2019.8843773

Februariyanti, H., & Zuliarso, E. (2012). Klasifikasi Dokumen Berita Teks Bahasa

Indonesia menggunakan Ontologi. *Teknologi Informasi DINAMIK*, *17*(1), 14–23. http://www.unisbank.ac.id/ojs/index.php/fti1/article/view/1612/594

Hanafi, A., Adiwijaya, A., & Astuti, W. (2020). Klasifikasi Multi Label pada Hadis Bukhari Terjemahan Bahasa Indonesia Menggunakan Mutual Information dan k-Nearest Neighbor. *Jurnal Sisfokom (Sistem Informasi Dan Komputer)*, *9*(3), 357–364. https://doi.org/10.32736/sisfokom.v9i3.980

Hanati, H., & Sari, K. (2021). *Perbandingan Metode Support Vector Machine (SVM) dan Artificial Neural Network (ANN) pada Klasifikasi Gizi Balita*. 1036–1043.

Irham, L. G., Adiwijaya, A., & Wisesty, U. N. (2019). Klasifikasi Berita Bahasa Indonesia Menggunakan Mutual Information dan Support Vector Machine. *Jurnal Media Informatika Budidarma*, *3*(4), 284. https://doi.org/10.30865/mib.v3i4.1410

Irmanda, H., & Astriratma, R. (2020). Klasifikasi Jenis Pantun Dengan Metode Support Vector Machines (SVM). *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, *4*(5), 915–922. https://doi.org/10.29207/resti.v4i5.2313

Khan, A., Baharudin, B., & Lee, L. H. (2010). *A Review of Machine Learning Algorithms for Text-Documents Classification*. *May 2014*. https://doi.org/10.4304/jait.1.1.4-20

Lucia, G., & Londo, Y. (2019). *A Study of Text Classification for Indonesian News Article*. 2019–2022.

Luthfiana, L., Young, J. C., & Rusli, A. (2020). *Implementasi Algoritma Support Vector Machine dan Chi Square untuk Analisis Sentimen User Feedback*

Aplikasi. *XII*(2), 125–128.

Made, N., Dwi, G., Fauzi, M. A., & Dewi, L. S. (2018). *Identifikasi Tweet Cyberbullying pada Aplikasi Twitter menggunakan Metode Support Vector Machine (SVM) dan Information Gain (IG) sebagai Seleksi Fitur*. *2*(11), 5326–5332.

Maulida, I., Suyatno, A., Rahmania Hatta, H., & Mulawarman, U. (2016). Seleksi Fitur Pada Dokumen Abstrak Teks Bahasa Indonesia Menggunakan Metode Information Gain. *JSM STMIK Mikroskil*, *17*(2), 249–258.

Mutawalli, L., Zaen, M. T. A., & Bagye, W. (2019). KLASIFIKASI TEKS SOSIAL MEDIA TWITTER MENGGUNAKAN SUPPORT VECTOR MACHINE (Studi Kasus Penusukan Wiranto). *Jurnal Informatika Dan Rekayasa Elektronik*, *2*(2), 43. https://doi.org/10.36595/jire.v2i2.117

Nasution, M. R. A., & Hayaty, M. (2019). Perbandingan Akurasi dan Waktu Proses Algoritma K-NN dan SVM dalam Analisis Sentimen Twitter. *Jurnal Informatika*, *6*(2), 226–235. https://doi.org/10.31311/ji.v6i2.5129

Pessoa, M., & Anwar, A. (2014). *A Review of RUP (Rational Unified Process) - 2014*.

Prakoso, B. S., Rosiyadi, D., Aridarma, D., Utama, H. S., Fauzi, F., & Qhomar, M. A. N. (2019). Optimalisasi Klasifikasi Berita Menggunakan Feature Information Gain Untuk Algoritma Naive Bayes Terhubung Random Forest. *Jurnal Pilar Nusa Mandiri*, *15*(2), 211–218. https://doi.org/10.33480/pilar.v15i2.684

Purwarianti, A., & Yusliani, N. (2011). Sistem Question Answering Bahasa

Indonesia Untuk Pertanyaan Non-Factoid. *Jurnal Ilmu Komputer Dan Informasi*, *4*(1), 10–14. https://doi.org/10.21609/jiki.v4i1.151

Rahmad, A., & Pribadi, F. (2015). *Edu Komputika Journal*. *2*(1), 13–21.

Rahman, O. H., Abdillah, G., & Komarudin, A. (2021). Classification of Hate Speech on Social Media Twitter Using Support Vector Machine. *RESTI Journal (Systems Engineering and Information Technology)*, *5*(1), 17–23.

Ridok, A., & Latifah, R. (2015). *Klasifikasi Teks Bahasa Indonesia Pada Corpus Tak Seimbang Menggunakan NWKNN*. 9–10.

Sahuri, G. (2018). *Studi Perbandingan Penggabungan Metode Pemilihan Fitur dengan Metode Klasifikasi dalam Klasifikasi Teks*. *01*(02), 1–5.

Saniyah. (2019). *Named Entity Recognition pada Teks Berita menggunakan Support Vector Machine*.

Sudin, S., Junaedi, H., & Santosa, J. (2019). *Analisis Jenis Pertanyaan Berbahasa Indonesia pada Question and Answering System Menggunakan Metode Support Vector Machine (SVM)*. *12*, 72–80.

Suharno, C., Fauzi, A., & Perdana, R. (2017). *KLASIFIKASI TEKS BAHASA INDONESIA PADA DOKUMEN PENGADUAN SAMBAT ONLINE MENGGUNAKAN METODE K-*. *03*(01), 25–32.

SUPARTINI, I. A. M., SUKARSA, I. K. G., & SRINADI, I. G. A. M. (2017). Analisis Diskriminan Pada Klasifikasi Desa Di Kabupaten Tabanan Menggunakan Metode K-Fold Cross Validation. *E-Jurnal Matematika*, *6*(2), 106. https://doi.org/10.24843/mtk.2017.v06.i02.p154

Tanti, Sirait, P., & Andri. (2021). *Optimalisasi Kinerja Klasifikasi Melalui Seleksi*

Fitur dan AdaBoost dalam Penanganan Ketidakseimbangan Kelas. *5*, 1377–1385. https://doi.org/10.30865/mib.v5i4.3280

Tuhenay, D. (2021). Perbandingan Klasifikasi Bahasa Menggunakan Metode Naïve Bayes Classifier (NBC) Dan Support Vector Machine (SVM). *JIKO (Jurnal Informatika Dan Komputer)*, *4*(2), 105–111. https://doi.org/10.33387/jiko.v4i2.2958

Yulietha, I., Faraby, S., & Adiwijaya. (2017). *Klasifikasi Sentimen Review Film Menggunakan Algoritma Support Vector Machine Sentiment Classification of Movie Reviews Using Algorithm Support Vector Machine*. *4*(3), 4740–4750.

Yusliani, N., Marieska, M., & Saputra, D. (2021). *Sistem Pengklasifikasian Pertanyaan untuk Kalimat Tanya*. Laporan Akhir Penelitian LPPM Universitas Sriwijaya.