KLASIFIKASI *SHORT MESSAGE SERVICE* (SMS) SPAM MENGGUNAKAN *SUPPORT VECTOR MACHINE* (SVM) DAN *ARTIFICIAL NEURAL NETWORK* (ANN)

Diajukan sebagai syarat untuk menyelesaikan
pendidikan Program Strata-1 pada
Jurusan Teknik Informatika

Oleh :

Shatia Earlangga Pratama
NIM : 09021182126017

Jurusan Teknik Informatika
FAKULTAS ILMU KOMPUTER UNIVERSITAS SRIWIJAYA
2025

HALAMAN PENGESAHAN

SKRIPSI

KLASIFIKASI SHORT MESSAGE SERVICE (SMS) SPAM MENGGUNAKAN SUPPORT VECTOR MACHINE (SVM) DAN ARTIFICIAL NEURAL NETWORK (ANN)

Sebagai salah satu syarat untuk penyelesaian studi di

Program Studi S1 Teknik Informatika

Oleh:

SHATIA EARLANGGA PRATAMA

09021182126017

Pembimbing 1	**:**	**Dian Palupi Rini, M.Kom.,Ph.D.**
		NIP. 197802232006042002
Pembimbing 2	**:**	**Desty Rodiah, S.Kom., M.T.**
		NIP. 198912212020122011

Mengetahui

Ketua Jurusan Teknik Informatika

Hadipurnawan Satria, Ph.D

198004182020121001

TANDA LULUS UJIAN KOMPREHENSIF

Pada hari kamis tangga tanggal 26 Juni 2025 telah dilaksanakan ujian komprehensif skripsi oleh Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.

Nama : Shatia Barlangga Pratama
Nim : 09021182126017
Judul : Klasifikasi *Short Message Service* (SMS) Spam Menggunakan *Support Vector Machine* (SVM) dan *Artificial Neural Network* (ANN)

dan dinyatakan LULUS.

1. Ketua Penguji
   Samsuryadi, M.Kom., Ph.D.
   NIP. 197102041997021003 .....
2. Penguji I
   Rifkie Primartha, S.T., M.T.
   NIP. 197706012009121004 .....
3. Pembimbing I
   Dian Palupi Rini, M.Kom., Ph.D.
   NIP. 197802232006042002 .....
4. Pembimbing II
   Desty Rodiah, S.Kom., M.T.
   NIP. 198912212020122011 .....

Mengetahui
Ketua Jurusan Teknik Informatika

Hadipurnawan Satria, Ph.D.
NIP. 198004182020121001

HALAMAN PERNYATAAN

Yang bertanda tangan di bawah ini :

Nama : Shatia Earlangga Pratama

NIM : 09021182126017

Program Studi : Teknik Informatika

Judul Skripsi : Klasifikasi *Short Message Service* (SMS) Spam Menggunakan *Support Vector Machine* (SVM) Dan *Artificial Neural Network* (ANN)

Hasil pengecekan Software Turnitin : 7%

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

Demikian pernyataan ini saya buat dengan sebenarnya dan tidak ada paksaan dari pihak mana pun.

Indralaya, 26 Juni 2025

Penulis,

Shatia Earlangga Pratama

NIM. 09021182126017

MOTTO DAN PERSEMBAHAN

Motto:

"*Hidup adalah sebuah film, pilihlah peranmu sendiri, panjatlah tanggamu sendiri atau galilah lubangmu sendiri*"

- J.Cole

Kupersembahkan Karya Tulis ini kepada:

- Allah SWT
- Orang Tua
- Keluarga Besar
- Fakultas Ilmu Komputer
- Universitas Sriwijaya

ABSTRACT

Short Message Service (SMS) is still used as a communication medium for promotions, notifications, and official information. However, SMS is also prone to misuse in the form of spam that can be disruptive and potentially deceive users. To address this issue, an accurate SMS spam and ham classification system is needed. This study developed an SMS spam classification system using SVM and ANN with a MLP architecture. The dataset used is a secondary dataset from the Kaggle platform, consisting of 1,143 SMS messages with a balanced class distribution of 574 spam and 569 ham messages. Feature extraction was carried out using TF-IDF and N-gram, while feature selection used Pearson Correlation. Model parameters were determined using Grid Search. The study was conducted through six testing scenarios. The results showed that the SVM model with a combination of TF-IDF and N-gram without feature selection achieved the best performance, with an accuracy of 98.25%, precision of 97.50%, recall of 99.15%, and F1-score of 98.32%. The best model used a linear kernel with a C value of 1. The findings indicate that SVM outperformed MLP-based ANN in SMS spam classification.

Keywords: *SMS Spam, Classification, SVM, ANN, MLP, TF-IDF, N-gram, Pearson Correlation, Kaggle, Grid Search*

ABSTRAK

Short Message Service (SMS) masih digunakan sebagai sarana komunikasi untuk promosi, notifikasi, dan informasi resmi. Namun, SMS juga berpotensi disalahgunakan dalam bentuk spam yang mengganggu dan dapat menipu pengguna. Mengatasi permasalahan ini, diperlukan sistem klasifikasi SMS spam dan ham secara akurat. Penelitian ini mengembangkan sistem klasifikasi SMS spam dengan menggunakan metode SVM dan ANN dengan arsitektur MLP. Dataset yang digunakan adalah data sekunder dari platform *kaggle* yang berisi 1.143 pesan SMS dengan distribusi kelas seimbang 574 pesan spam dan 569 pesan ham. Ekstraksi fitur dilakukan menggunakan TF-IDF dan *N-gram*, sedangkan seleksi fitur menggunakan *Pearson Correlation*. Parameter model ditentukan melalui *Grid Search*. Penelitian dilakukan dengan enam skenario pengujian. Hasil menunjukkan bahwa model SVM dengan kombinasi TF-IDF dan *N-gram* tanpa seleksi fitur memberikan performa terbaik dengan akurasi 98,25%, presisi 97,50%, recall 99,15% dan f1-score 98,32%. Model terbaik ini menggunakan parameter SVM berupa kernel linear dengan nilai C=1. Hasil penelitian menunjukkan bahwa SVM lebih unggul dibandingkan ANN berbasis MLP dalam klasifikasi SMS spam.

Kata Kunci: SMS Spam, Klasifikasi, SVM, ANN, MLP, TF-IDF, *N-gram*, *Pearson Correlation, Kaggle, GridSearch*

KATA PENGANTAR

Segala puji dan syukur penulis panjatkan ke hadirat Allah SWT, atas rahmat, taufik, dan hidayah-Nya, sehingga penulis dapat menyelesaikan penyusunan skripsi ini sebagai salah satu syarat untuk memperoleh gelar Sarjana pada Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Sriwijaya. Skripsi ini tidak akan terselesaikan tanpa bantuan, doa, serta dukungan dari berbagai pihak. Oleh karena itu, dengan segala kerendahan hati, penulis ingin menyampaikan ucapan terima kasih kepada:

1. Allah SWT atas rahmat dan nikmat-Nya sehingga, penulis dapat menyelesaikan skripsi ini dengan baik
2. Kedua orang tua dan keluarga tercinta yang selalu memberikan doa, motivasi, dan dukungan tanpa henti.
3. Bapak Hadipurnawan Satria, Ph.D., Ketua Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.
4. Ibu Dian Palupi Rini, M.Kom., Ph.D., dan Ibu Desty Rodiah, M.T., selaku Dosen Pembimbing Skripsi I dan II, atas bimbingan serta motivasi yang sangat berarti.
5. Seluruh dosen dan staf administrasi di Jurusan Teknik Informatika serta Fakultas Ilmu Komputer yang telah memberikan bantuan dan fasilitas selama masa studi.
6. Sahabat-sahabat penulis — Angel, Nisa, Putri, Robby, Tian dan Zaky — yang telah setia menemani, membantu, dan memberikan arahan ketika penulis merasa bingung dan kewalahan dalam menyusun penelitian ini.

Kehadiran kalian bukan hanya memberi semangat, tapi juga jadi bagian penting dalam setiap langkah yang penulis lalui.

7. Semua pihak yang tidak dapat saya sebutkan satu per satu yang telah membantu secara langsung maupun tidak langsung.

Penulis menyadari bahwa skripsi ini masih memiliki kekurangan. Oleh karena itu, penulis berharap adanya kritik dan saran yang bersifat membangun untuk penyempurnaan skripsi ini. Semoga skripsi ini dapat memberikan manfaat bagi semua pihak yang membutuhkan.

Indralaya, 26 Juni 2025

Penulis,

Shatia Earlangga Pratama

DAFTAR ISI

DAFTAR TABEL

DAFTAR GAMBAR

BAB I
PENDAHULUAN

1.1 Pendahuluan

Pada bab ini menyajikan gambaran umum dari keseluruhan penelitian yang dilakukan. Di dalamnya akan dibahas mengenai latar belakang permasalahan, perumusan masalah, tujuan yang ingin dicapai, manfaat yang diharapkan dari penelitian, serta batasan-batasan yang ditetapkan. Dengan demikian, bab ini menjadi dasar untuk memahami konteks dan arah penelitian secara menyeluruh.

1.2 Latar Belakang Masalah

Dalam era komunikasi digital yang berkembang pesat, pertukaran informasi semakin masif melalui berbagai media, termasuk layanan pesan singkat atau *Short Message Service* (SMS). Meskipun saat ini aplikasi pesan instan mendominasi komunikasi sehari-hari, SMS tetap banyak digunakan terutama untuk pengiriman informasi penting seperti notifikasi perbankan, promosi layanan, dan pemberitahuan resmi dari institusi pemerintah (Ballı & Karasoy, 2019). Tingkat keterandalan dan jangkauan yang luas membuat SMS tetap relevan, terutama di kalangan masyarakat yang tidak selalu terhubung dengan internet. Berdasarkan data yang dikutip oleh Panjaitan et al. (2023), Indonesia termasuk dalam 20 negara dengan jumlah SMS spam tertinggi pada tahun 2021, dengan rata-rata enam pesan spam diterima oleh pengguna setiap bulan menurut laporan Truecaller Insight. Data ini menunjukkan bahwa SMS spam merupakan masalah yang cukup serius di Indonesia dan dapat mengganggu kenyamanan serta keamanan komunikasi digital.

Dalam penggunaan SMS yang meluas juga memunculkan celah berupa penyalahgunaan oleh pihak-pihak tidak bertanggung jawab dalam bentuk pesan spam. Pesan semacam ini dikirimkan secara massal tanpa izin dan sering kali menyerupai pesan resmi, sehingga menimbulkan kebingungan bagi pengguna. Tidak sedikit dari pesan-pesan ini yang berisi penipuan, promosi ilegal, atau tautan berbahaya yang dapat mengarah pada pencurian data (Herwanto et al., 2021). Kondisi ini tidak semua pengguna dapat dengan mudah membedakan antara SMS spam dengan yang aslinya terutama karena teks spam sering menyerupai pesan resmi, sehingga diperlukan sistem klasifikasi SMS yang dapat membantu mengidentifikasi pesan spam secara akurat.

Dalam klasifikasi teks, diperlukan pendekatan yang mampu mengubah data teks menjadi bentuk numerik yang dapat diproses oleh algoritma *machine learning*. Kombinasi TF-IDF dan *N-gram* banyak digunakan dalam klasifikasi teks karena terbukti meningkatkan performa klasifikasi, serta dapat menangkap frekuensi kata penting (Gerliandeva et al., 2024). Selain itu, teknik seleksi fitur seperti *Pearson Correlation* (PC) dapat membantu memilih fitur paling relevan untuk meningkatkan efisiensi dan akurasi klasifikasi (Firdaus et al., 2023). Selain proses ekstraksi dan seleksi fitur, pemilihan algoritma klasifikasi juga memiliki peran krusial. SVM dan ANN merupakan dua algoritma yang umum digunakan dalam klasifikasi teks karena telah terbukti mampu menghasilkan performa yang kompetitif dalam berbagai penelitian.

Berbagai penelitian terdahulu telah membuktikan efektivitas masing-masing metode tersebut. Dwiprayoga & Raharja (2025) mencatat akurasi sebesar

93,75% dengan menggunakan TF-IDF dalam melakukan klasifikasi SMS. Gerliandeva et al. (2024) mencapai akurasi 94% dengan mengombinasikan TF-IDF dan *N-gram* dalam klasifikasi sentimen komentar online. Romadloni et al. (2024) melakukan penelitian klasifikasi SMS menggunakan algoritma *machine learning*, salah satunya adalah SVM. Hasil penelitian menunjukkan akurasi SVM mencapai 91,41%, meningkat menjadi 91,96% setelah diterapkan *N-gram*. Namun, saat diterapkan PC, akurasi model justru menurun signifikan menjadi 70,80%.

Dari sisi algoritma, Penelitian Kusuma et al. (2022) melakukan komparasi metode ANN menggunakan MLP dan SVM untuk klasifikasi kanker payudara, dimana ANN dengan arsitektur MLP sedikit lebih unggul dengan akurasi mencapai 97,7%, sedangkan SVM mendapatkan akurasi 96,2%. Di sisi lain, penelitian oleh Hartanti et al. (2023) yang membandingkan beberapa algoritma termasuk SVM dan ANN untuk klasifikasi teks dalam konteks reservasi hotel menunjukkan hasil berbeda, dimana SVM lebih unggul dengan akurasi 85,7%, sementara ANN menunjukkan akurasi lebih rendah yaitu 80,48%.

Berdasarkan uraian latar belakang di atas, meskipun telah banyak penelitian yang mengkaji efektivitas SVM, ANN, TF-IDF, *N-gram*, dan PC secara terpisah atau dalam kombinasi terbatas, masih terdapat kekurangan penelitian dalam menganalisis bagaimana kelima metode tersebut dapat diintegrasikan secara optimal untuk meningkatkan performa klasifikasi SMS spam. Penelitian ini bertujuan menyelidiki efektivitas perbandingan SVM dan ANN sebagai algoritma klasifikasi, dengan TF-IDF dan *N-gram* sebagai metode ekstraksi fitur, serta PC sebagai metode seleksi fitur dalam konteks klasifikasi SMS spam. Kombinasi

metode-metode tersebut diharapkan dapat menghasilkan model klasifikasi yang lebih akurat dalam membedakan SMS spam dan ham, sehingga dapat membantu pengguna dalam mengidentifikasi dan menghindari pesan spam, serta mendukung upaya pencegahan terhadap penipuan dan gangguan informasi yang disebarkan melalui layanan pesan singkat (SMS).

1.3 Rumusan Masalah

Rumusan masalah penelitian ini adalah:

1. Bagaimana membangun sistem klasifikasi SMS yang mampu membedakan pesan spam dan pesan ham?
2. Bagaimana pengaruh kombinasi TF-IDF, *N-gram*, dan *Pearson Correlation* dalam proses ekstraksi dan seleksi fitur untuk mendukung performa klasifikasi pesan SMS?
3. Bagaimana performa model klasifikasi SMS yang dibangun dengan kombinasi TF-IDF dan *N-gram* serta seleksi fitur *Pearson Correlation* menggunakan SVM dan ANN berdasarkan metrik evaluasi?

1.4 Tujuan Penelitian

Tujuan penelitian ini adalah:

1. Merancang sistem klasifikasi SMS yang dapat membantu pengguna dalam membedakan pesan spam dan pesan ham.
2. Menganalisis pengaruh kombinasi metode TF-IDF, *N-gram*, dan *Pearson Correlation* dalam proses ekstraksi dan seleksi fitur pada klasifikasi SMS.
3. Mengevaluasi performa model klasifikasi SMS dengan kombinasi TF-IDF dan *N-gram* serta seleksi fitur *Pearson Correlation* menggunakan SVM dan

ANN (akurasi, recall, presisi, dan f1-score).

1.5 Manfaat Penelitian

1. Penelitian ini menghasilkan sistem klasifikasi SMS yang mampu membantu pengguna dalam membedakan pesan spam dan pesan ham.
2. Mengetahui performa algoritma SVM dan ANN dengan kombinasi metode TF-IDF, *N-gram*, dan *Pearson Correlation*.
3. Menambah kontribusi penelitian klasifikasi teks berbahasa Indonesia dan menyediakan pendekatan yang dapat direplikasi untuk berbagai jenis teks lain seperti email atau komentar daring.

1.6 Batasan Masalah

Batasan masalah penelitian ini adalah:

1. Penelitian ini hanya berfokus pada klasifikasi pesan SMS berbahasa Indonesia ke dalam dua kategori, yaitu spam dan ham.
2. *Dataset* yang digunakan relatif kecil, sehingga dapat membatasi kemampuan model dalam mengenali variasi pola SMS, terutama pada SMS dengan konteks yang *ambigu*.
3. Penelitian ini menggunakan ANN dengan arsitektur MLP dan pelatihan menggunakan algoritma *Backpropagation*.
4. Ekstraksi fitur *N-gram* menggunakan kombinasi fitur *unigram* dan *bigram* sebagai representasi fitur teks.

1.7 Sistematika Penulisan

Sistematika penulisan tugas akhir ini disusun berdasarkan pedoman penulisan yang ditetapkan oleh Fakultas Ilmu Komputer Universitas Sriwijaya.

Laporan ini terdiri dari enam bab utama, yaitu.

BAB I. PENDAHULUAN

Pada bab ini membahas latar belakang, rumusan masalah, tujuan dan manfaat penelitian, batasan masalah serta sistematika penulisan. Bab ini menjadi dasar pemikirian dalam pelaksanaan penelitian.

BAB II. KAJIAN LITERATUR

Pada bab ini menguraikan landasan teori yang mendukung penelitian, antara lain definisi *Short Message Service* (SMS), Klasifikasi Teks, *Preprocessing*, TF-IDF, *N-gram*, *Pearson Correlation*, *Support Vector Machine* (SVM), *Artificial Neural Network* (ANN), *Confusion Matrix* dan serta model proses pengembangan perangkat lunak *Rational Unified Process* (RUP). Selain itu, disertakan dengan tinjauan terhadap beberapa penelitian terdahulu yang relevan.

BAB III. METODOLOGI PENELITIAN

Pada bab ini memaparkan secara sistematis tahapan-tahapan yang dilaksanakan dalam proses penelitian, yang mencakup kegiatan pengumpulan data, analisis data, serta perancangan perangkat lunak. Setiap tahapan dijabarkan secara komprehensif berdasarkan kerangka metodologis yang telah ditetapkan, agar memastikan keterpaduan antara proses penelitian dan tujuan yang ingin dicapai.

BAB IV. PENGEMBANGAN PERANGKAT LUNAK

Bab ini menjelaskan secara rinci proses pengembangan perangkat lunak yang meliputi analisis kebutuhan, perancangan sistem, tahap konstruksi perangkat lunak, serta proses pengujian untuk memastikan bahwa sistem yang dikembangkan sesuai dengan tujuan penelitian.

BAB V. HASIL DAN ANALISIS PENELITIAN

Bab ini menyajikan hasil pengujian yang telah dilakukan beserta analisis mendalam terhadap data tersebut. Analisis ini bertujuan untuk mengevaluasi efektivitas serta performa sistem klasifikasi SMS spam yang dikembangkan dalam penelitian.

BAB VI. KESIMPULAN DAN SARAN

Bab ini memuat kesimpulan berdasarkan hasil penelitian yang telah dilaksanakan, sekaligus memberikan rekomendasi untuk pengembangan lebih lanjut maupun penelitian berikutnya yang berkaitan.

1.8 Kesimpulan

Bab ini telah menguraikan latar belakang permasalahan, rumusan masalah, tujuan dan manfaat penelitian, batasan masalah, serta sistematika penulisan. Pembahasan dalam bab ini menjadi landasan fundamental dalam pelaksanaan penelitian serta pengembangan model klasifikasi SMS spam menggunakan algoritma SVM dan ANN. Bab selanjutnya akan mengkaji literatur yang relevan sebagai dasar teoritis untuk penelitian ini.

DAFTAR PUSTAKA

Arifin, N., Enri, U., & Sulistiyowati, N. (2021). Penerapan Algoritma *Support Vector Machine* (SVM) dengan TF-IDF N-Gram untuk *Text Classification*. *STRING (Satuan Tulisan Riset dan Inovasi Teknologi)*, *6*(2), 129. https://doi.org/10.30998/string.v6i2.10133

Ayu, F. (2019). Implementasi Jaringan Saraf Tiruan Untuk Menentukan Kelayakan Proposal Tugas Akhir. *IT JOURNAL RESEARCH AND DEVELOPMENT*, *3*(2), 44–53. https://doi.org/10.25299/itjrd.2019.vol3(2).2271

Ballı, S., & Karasoy, O. (2019). *Development of content-based SMS classification application by using Word2Vec-based feature extraction*. *IET Software*, *13*(4), 295–304. https://doi.org/10.1049/iet-sen.2018.5046

Cahyani, D. E., & Patasik, I. (2021). *Performance comparison of TF-IDF and Word2Vec models for emotion text classification. Bulletin of Electrical Engineering and Informatics*, *10*(5), 2780–2788. https://doi.org/10.11591/eei.v10i5.3157

Dwiprayoga, I. K., & Raharja, M. A. (2025). Komparasi Ekstraksi Fitur BoW dan TF-IDF untuk Klasifikasi SMS Menggunakan *Naive Bayes*. *3*. 3. Jurnal Nasional Teknologi Informasi dan Aplikasinya, 2025, pp. 247-254. https://paperity.org/p/362121260

Fikri, M. I., Sabrila, T. S., & Azhar, Y. (2020). Perbandingan Metode *Naïve Bayes* dan *Support Vector Machine* pada Analisis Sentimen Twitter. *SMATIKA JURNAL*, *10*(02), 71–76. https://doi.org/10.32664/smatika.v10i02.455

Firdaus, R., Mualfah, D., & Hasanah, J. S. (2023). Klasifikasi *Multi-Class* Penyakit Jantung Dengan SMOTE dan *Pearson's Correlation* menggunakan MLP. *Jurnal CoSciTech (Computer Science and Information Technology)*, *4*(1), 262–271. https://doi.org/10.37859/coscitech.v4i1.4769

García, M., Maldonado, S., & Vairetti, C. (2021). *Intelligent Data Analysis*, *25*(3), 509–525. https://doi.org/10.3233/IDA-205154

Gasparetto, A., Marcuzzo, M., Zangari, A., & Albarelli, A. (2022). A Survey on Text Classification Algorithms: From Text to Predictions. *Information*, *13*(2), 83. https://doi.org/10.3390/info13020083 *Efficient N-gram construction for text categorization using feature selection techniques.*

Gerliandeva, A., Chrisnanto, Y., & Ashaury, H. (2024). Optimasi Klasifikasi Sentimen pada Komentar Online menggunakan Multinomial Naïve Bayes dan Ekstraksi Fitur TF-IDF serta *N-grams*. *Jurnal Pekommas*, *9*(2), 260–272. https://doi.org/10.56873/jpkm.v9i2.5585

Gori, T., Sunyoto, A., & Al Fatta, H. (2024). Preprocessing Data dan Klasifikasi untuk Prediksi Kinerja Akademik Siswa. *Jurnal Teknologi Informasi dan Ilmu Komputer*, *11*(1), 215–224. https://doi.org/10.25126/jtiik.20241118074

Hanum, A. R., Zetha, I. A., Fajrina, J. N., Wulandari, R. A., Putri, C., Andina, S. P., & Yudistira, N. (2024). Analisis Kinerja Algoritma Klasifikasi Teks Bert Dalam Mendeteksi Berita Hoaks. *Jurnal Ilmiah Penelitian dan Pembelajaran Informatika*. https://doi.org/10.29100/jipi.v10i2.7811

Hartanti, D., Ichsan Pradana, A., & Lestari, S. (2023). Komprasi Algoritma Decision Tree, SVM dan ANN untuk Reservasi Hotel. *DutaCom*, *16*(1), 21–27. https://doi.org/10.47701/dutacom.v16i1.2647

Hasanah, S. H., & Permatasari, S. M. (2020). Metode Klasifikasi Jaringan Syaraf Tiruan *Backpropagation* Pada Mahasiswa Statistika Universitas Terbuka. *BAREKENG: Jurnal Ilmu Matematika dan Terapan*, *14*(2), 243–252. https://doi.org/10.30598/barekengvol14iss2pp243-252

Herwanto, H., Chusna, N. L., & Arif, M. S. (2021). Klasifikasi SMS Spam Berbahasa Indonesia Menggunakan Algoritma Multinomial Naïve Bayes. *JURNAL MEDIA INFORMATIKA BUDIDARMA*, *5*(4), 1316. https://doi.org/10.30865/mib.v5i4.3119

Irmanda, H. N., & Astriratma, R. (2020). Klasifikasi Jenis Pantun dengan Metode *Support Vector Machines (SVM)*. *4*(5). *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi),* *4*(5), 1065–1071. https://doi.org/10.29207/resti.v4i5.2313

Kurniadi, D., Setiawan, R., Adiwangsa, A. A., & Lindayani, L. (2023). *Development of Multi-Developer Housing Marketing Information System Using Rational Unified Process Method. Sinkron*, *8*(1), 348–359. https://doi.org/10.33395/sinkron.v8i1.11964

Kusuma, J., Hayadi, B. H., Wanayumini, W., & Rosnelly, R. (2022). Komparasi Metode *Multi Layer Perceptron* (MLP) dan *Support Vector Machine* (SVM) untuk Klasifikasi Kanker Payudara. *MIND Journal*, *7*(1), 51–60. https://doi.org/10.26760/mindjournal.v7i1.51-60

Milal, I. S., M. Hasanudin, M. H., Nur Azhari, M. A., Nugraha, R. A., Agustina, N., & Damayanti, S. E. (2023). Klasifikasi Teks Review Pada E-Commerce Tokopedia Menggunakan Algoritma SVM. *Naratif : Jurnal Nasional Riset, Aplikasi dan Teknik Informatika*, *5*(1), 34–45. https://doi.org/10.53580/naratif.v5i1.191

Minaee, S., Kalchbrenner, N., Cambria, E., Nikzad, N., Chenaghlu, M., & Gao, J. (2021). *Deep Learning Based Text Classification: A Comprehensive Review* (arXiv:2004.03705). arXiv. http://arxiv.org/abs/2004.03705

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

Nurhidayat, R., & Dewi, K. E. (2023). Penerapan Algoritma K-Nearest Neighbor Dan Fitur Ekstraksi N-Gram Dalam Analisis Sentimen Berbasis Aspek. *Komputa : Jurnal Ilmiah Komputer dan Informatika*, *12*(1), 91–100.

https://doi.org/10.34010/komputa.v12i1.9458

Panjaitan, M. D., Adikara, P. P., & Setiawan, B. D. (2024). Klasifikasi Spam pada *Short Message Service* (SMS) menggunakan *Support Vector Machine*. Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer, 1(1). https://repository.ub.ac.id/id/eprint/223005

Pradana, A. W., & Hayaty, M. (2019). The Effect of Stemming and Removal of Stopwords on the Accuracy of Sentiment Analysis on Indonesian-language Texts. *Kinetik: Game Technology, Information System, Computer Network, Computing, Electronics, and Control*, 375–380. https://doi.org/10.22219/kinetik.v4i4.912

Putra, M. Y., & Putri, D. I. (2022). Pemanfaatan Algoritma Naïve Bayes dan K-Nearest Neighbor Untuk Klasifikasi Jurusan Siswa Kelas XI. *Jurnal Tekno Kompak*, *16*(2), 176. https://doi.org/10.33365/jtk.v16i2.2002

Rahayu, K., Fitria, V., Septhya, D., Rahmaddeni, R., & Efrizoni, L. (2023). Klasifikasi Teks untuk Mendeteksi Depresi dan Kecemasan pada Pengguna Twitter Berbasis *Machine Learning*: *Text Classification for Detecting Depression and Anxiety among Twitter Users based on Machine Learning*. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, *3*(2), 108–114. https://doi.org/10.57152/malcom.v3i2.780

Romadloni, N. T., Septiyanti, N. D., Pratomo, C. H., Kurniawan, W., & Bintang, R. A. K. N. (2024). *Classification Of Sms Spam With N-gram And Pearson Correlation Based Using Machine Learning Techniques*. *SENTRI: Jurnal Riset Ilmiah*, *3*(2), 967–977. https://doi.org/10.55681/sentri.v3i2.2252

Roy, P. K., Singh, J. P., & Banerjee, S. (2020). *Deep learning to filter SMS Spam*. *Future Generation Computer Systems*, *102*, 524–533. https://doi.org/10.1016/j.future.2019.09.001

Samsir, S., Ambiyar, A., Verawardina, U., Edi, F., & Watrianthos, R. (2021). Analisis Sentimen Pembelajaran Daring Pada Twitter di Masa Pandemi COVID-19 Menggunakan Metode Naïve Bayes. *JURNAL MEDIA INFORMATIKA BUDIDARMA*, *5*(1), 157. https://doi.org/10.30865/mib.v5i1.2580

Septhya, D., Rahayu, K., Rabbani, S., Fitria, V., Rahmaddeni, R., Irawan, Y., & Hayami, R. (2023). Implementasi Algoritma Decision Tree dan Support Vector Machine untuk Klasifikasi Penyakit Kanker Paru: Implementation of Decision Tree Algorithm and *Support Vector Machine for Lung Cancer Classification*. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, *3*(1), 15–19. https://doi.org/10.57152/malcom.v3i1.591

Surianto, D. F., Fajar B, M., Mulia, M. R., & Indanasufya, I. (2024). *Comparative Analysis of the Performance of Hadith Text Classification Methods: A Case Study with* ANN and SVM. *Journal of Embedded Systems, Security and Intelligent Systems*, 89–98. https://doi.org/10.59562/jessi.v5i1.2942