Klasifikasi SMS *Spam* Berbahasa Indonesia Menggunakan K-Nearest Neighbor dan Chi-Square

Diajukan Untuk Menyusun Skripsi
di Jurusan Teknik Informatika Fakultas Ilmu Komputer UNSRI

Oleh :

Adisti Kusumawardhani
09021381924151

Jurusan Teknik Informatika
FAKULTAS ILMU KOMPUTER UNIVERSITAS SRIWIJAYA
2024

LEMBAR PENGESAHAN SKRIPSI

Klasifikasi SMS *Spam* Berbahasa Indonesia Menggunakan K-Nearest Neighbor dan Chi-Square

Oleh :

Adisti Kusumawardhani
NIM : 09021381924151

Palembang, 26 Agustus 2024

Pembimbing I,

Novi Yusliani, M.T.
NIP. 19821082012122001

Pembimbing II,

M.Naufal Rachmatullah, M.T.
NIP. 199212012022031008

Mengetahui,
Ketua Jurusan Teknik Informatika

Hadipurnawan Satria, Ph.D.
NIP. 198004182020121001

TANDA LULUS UJIAN SIDANG SKRIPSI

Pada hari Senin tanggal 29 Juli 2024 telah dilaksanakan ujian sidang skripsi oleh Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.

Nama : Adisti Kusumawardhani

NIM : 09021381924151

Judul : Klasifikasi SMS *Spam* Berbahasa Indonesia Menggunakan K-Nearest Neighbor dan Chi-Square

1. Ketua Penguji
   Samsuryadi, S.Kom., Ph.D.
   NIP. 197102041997021003 .....
2. Pembimbing I
   Novi Yusliani, S.Kom., M. T.
   NIP. 198211082012122001 .....
3. Pembimbing II
   M. Naufal Rachmatullah, S.Kom., M. T.
   NIP. 199212012022031008 .....
4. Penguji I
   Dian Palupi Rini, S.Kom., Ph.D.
   NIP. 197802232006042002 .....

Mengetahui,
Ketua Jurusan Teknik Informatika

Hadipurnawan Satria, Ph.D.
NIP. 198004182020121001

HALAMAN PERNYATAAN BEBAS PLAGIAT

Yang bertanda tangan dibawah ini :

Nama	: Adisti Kusumawardhani
NIM	: 09021381924151
Program Studi	: Teknik Informatika Bilingual
Judul	: Klasifikasi SMS *Spam* Berbahasa Indonesia Menggunakan K-Nearest Neighbor dan Chi-Square

Hasil Pengecekan Software iThenticate/Turnitin : 9%

Menyatakan bahwa laporan skripsi 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, 26 Agustus 2024

Penulis

Adisti Kusumawardhani

09021381924151

MOTTO DAN PERSEMBAHAN

Motto :

"Life is tough, and things don't always work out well, but we should be brave and go on with our lives"

Skripsi ini persembahkan kepada :

- Kedua Orang Tua, Saudara, serta Keluarga Besar
- Dosen - dosen
- Fakultas Ilmu Komputer
- Universitas Sriwijaya

Classification of Indonesian Spam SMS Using K-Nearest Neighbor and Chi-Square

By :

Adisti Kusumawardhani (09021381924151)
Department of Informatics, Faculty of Computer Science, Sriwijaya University
Email : adisti.kusumawardhani@gmail.com

ABSTRACT

Spam SMS is so dangerous that it can cause losses to SMS service users. To overcome this, a method is needed that can help classify SMS according to its category, namely Spam and non-spam. Classification is one of the processes of grouping data into predetermined classes. Classification goes through several stages, namely, pre-processing, feature selection, weighting, and classification. The method used in this research is K-Nearest Neighbor for the classification process, and Chi-Square for the feature selection process. The results of classification research using a value of N = 100 show that by applying a combination of chi-square feature selection with the KNN algorithm, the classification accuracy rate decreases by about 10% compared to without feature selection. This research shows that chi-square feature selection is less effective with the KNN algorithm for classification results.

Keywords:Spam Classification, K-Nearest Neighbor, Feature Selection, Chi Square

Palembang, 26 Agustus 2024

Supervisor I,

Novi Yusliani, M. T.
NIP. 198211082012122001

Supervisor II,

M.Naufal Rachmatullah, M. T.
NIP. 199212012022031008

Approved By,
Head of Department of Informatics

Hadipurnawan Satria, Ph.D.
NIP. 198004182020121001

Klasifikasi SMS *Spam* Berbahasa Indonesia Menggunakan K-Nearest Neighbor dan Chi-Square

By :

Adisti Kusumawardhani (09021381924151)
Department of Informatics, Faculty of Computer Science, Sriwijaya University
Email : adisti.kusumawardhani@gmail.com

ABSTRAK

SMS *Spam* sangat membahayakan sehingga dapat menyebabkan kerugian bagi pengguna layanan SMS. Untuk mengatasi hal tersebut, dibutuhkan metode yang dapat membantu mengelompokan SMS sesuai dengan ketegorinya, yaitu *Spam* dan *non spam*. Klasifikasi merupakan salah satu proses mengelompokkan data kedalam kelas yang telah ditentukan sebelumnya. Pengklasifikasian melewati beberapa tahapan yaitu, pra pengolahan, seleksi fitur, pembobotan, dan pengklasifikasian. Metode yang di gunakan pada penelitian ini adalah K-Nearest Neighbor untuk proses pengklasifikasian, dan Chi-Square untuk proses seleksi fitur. Hasil penelitian klasifikasi menggunakan nilai N = 100 menunjukkan bahwa dengan menerapkan kombinasi seleksi fitur chi-square dengan algoritma KNN justru mengalami penurunan tingkat akurasi pengklasifikasian sekitar 10% dibandingkan tanpa seleksi fitur. Penelitian ini menunjukkan bahwa kurang efektif nya seleksi fitur chi-square dengan algoritma KNN untuk hasil klasifikasi.

Kata Kunci : Klasifikasi Spam, K-Nearest Neighbor, Seleksi Fitur, Chi Square

Palembang, 26 Agustus 2024

Pembimbing I,

Novi Yusliani, M. T.
NIP. 198211082012122001

Pembimbing II,

M.Naufal Rachmatullah, M. T.
NIP. 199212012022031008

Mengetahui,
Ketua Jurusan Teknik Informatika

Hadipurnawan Satria, Ph.D.
NIP. 198004182020121001

KATA PENGANTAR

Puji dan syukur kepada Allah SWT atas rahmat dan nikmat Nya yang lebih diberikan kepada penulis sehingga dapat menyelesaikan skripsi ini dengan baik. Skripsi ini disusun sebagai salah satu syarat menyelesaikan Pendidikan program Strata-1 di Fakultas Ilmu Komputer Universitas Sriwijaya. Dalam menyelesaikan skripsi ini penulis menerima bantuan, bimbingan dan dukungan dari banyak pihak, baik secara langsung maupun tidak langsung. Oleh karena itu, penulis ingin menyampaikan terima kasih kepada :

1. Allah SWT atas rahmat dan nikmat-Nya penulis dapat menyelesaikan skripsi ini dengan baik.
2. Kedua orang tua penulis, ibunda Henny dan ayahanda Ismed yang telah mendukung, mendoakan, memberi semangat, memotivasi dan memberi nasihat.
3. Ibu Ellyn Saputra selaku sosok ibu yang telah mendukung, mendoakan, memberi semangat, memotivasi dan memberi nasihat.
4. Saudara dan saudari penulis, kakak Emir Kusuma, kakak Edwin Krisnaputra, kakak Aurelia Ananda yang telah memberi semangat, memotivasi dan memberi nasihat.
5. Bapak Muhammad Qurhanul Rizqie, S.Kom,. M.T., Ph.D. selaku Dosen dan sekaligus pembimbing akademik.
6. Bapak Hadipurnawan Satria, Ph.D. selaku Ketua Jurusan Teknik Informatika Universitas Sriwijaya.

7. Ibu Novi Yusliani, M.T. selaku Dosen Pembimbing I dan Bapak Muhammad Naufal Rachmatullah, M. T. selaku dosen Pembimbing II yang telah membimbing, memberikan motivasi serta arahan kepada penulis dalam proses pengerjaan skripsi.
8. Seluruh dosen program studi serta admin Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.
9. Bintang Dwitama yang telah memberikan semangat, saran dan motivasi selama mengerjakan skripsi ini.
10. Rosa Mulyani selaku sahabat penulis yang memberikan semangat, saran dan menemani penulis menyelesaikan skripsi.
11. Karina Nur Aliyyah, Tiara Larasati, dan Fitra Aliya Rahma sahabat lama penulis yang semangat dan saran untuk penulis menyelesaikan skripsi.
12. Pihak – pihak lain yang tidak dapat penulis sebutkan satu-persatu.

Penulis menyadari bahwa dalam penyusunan skripsi ini masih banyak sekali kekurangan dikarenakan kurangnya pengalaman dan pengetahuan penulis. Oleh karena itu penulis mengharapkan saran dan kritik yang membangun guna kemajuan penelitian selanjutnya. Semoga tugas akhir ini dapat bermanfaat. Terima kasih.

Palembang, 26 Agustus 2024

Adisti Kusumawardhani
09021381924151

DAFTAR ISI

DAFTAR TABEL

DAFTAR GAMBAR

DAFTAR LAMPIRAN

BAB I

PENDAHULUAN

1. 1 Pendahuluan

Pada bab ini akan diuraikan tentang latar belakang penelitian klasifikasi dokumen berbahasa Indonesia menggunakan K-Nearest Neighbor dan chi square, rumusan masalah, tujuan dan manfaat penelitian, batasan atau ruang lingkup masalah, sistematika penulis, dan kesimpulan.

1. 2 Latar Belakang Masalah

Perkembangan teknologi yang pesat dalam informasi digital memudahkan pengguna untuk melakukan pertukaran informasi. Hal ini disebabkan karena adanya teknologi dibidang komunikasi. Salah satu teknologi komunikasi tersebut adalah SMS. SMS (*Short Message Service*) merupakan layanan pengiriman pesan berupa teks singkat antar perangkat telepon seluler. SMS merupakan salah satu saluran komunikasi yang masih banyak digunakan, selain harga yang terjangkau SMS juga sederhana dalam layanan komunikasi, SMS tidak banyak membutuhkan media sehingga pesan teks dapat terkirim dengan cepat. Oleh karena itu, banyaknya pengiriman massal pesan teks yang diterima tanpa ada permintaan atau persetujuan dari penerima sebelumnya, seperti salah satunya yaitu pesan teks promosi, penipuan atau *phishing*, serta link yang mencurigakan yang bisa membawa penerima ke situs yang berbahaya atau penuh dengan malware.

Hal tersebut mengakibatkan pesan teks yang tidak dikehendaki atau dapat diistilahkan sebagai *spam*. SMS *spam* sangat menggangu bahkan sampai membahayakan sehingga dapat menyebabkan kerugian bagi pengguna layanan SMS dan spam merupakan masalah yang akan terus berkembang. Untuk mengatasi hal tersebut, dibutuhkan metode yang dapat membantu mengelompokan SMS sesuai dengan ketegorinya. Salah satu metode tersebut adalah klasifikasi.

Klasifikasi merupakan proses mengelompokkan data kedalam kelas yang telah ditentukan sebelumnya. Hal ini membuat klasifikasi dapat membantu dalam pergorganisasian dan pengelompokan data yang akan digunakan oleh pengguna. Untuk mengelompokkan data-data tersebut dibutuhkanlah sistem untuk membantu mengklasifikasikan SMS.

Salah satu metode yang dapat digunakan dalam proses pengklasifikasian adalah *K-Nearest Neighbor*. *K-Nearest Neighbor* atau KNN merupakan metode non-parametrik yang melakukan klasifikasi berdasarkan kelas tetangga terdekat. Pada penelitian (Perdana & Fauzi, 2018) KNN melakukan klasifikasi dengan menghasilkan akurasi yang baik karena melihat jarak antara objek dengan kelas. Berdasarkan hasil penelitian tersebut, menunjukkan bahwa algoritma KNN dapat diimplementasikan pada sistem klasifikasi dokumen. Pada bagian penelitian klasifikasi dokumen tersebut, algoritme KNN berhasil menghasilkan nilai akurasi sebesar 66% dengan jumlah data yang tersedia.

KNN juga memiliki kekurangan yaitu bergantung pada pengukuran jarak, dimana semakin besar nilai K hasil akurasi justru menurun. Hal ini karena semakin besar nilai K (jarak), semakin banyak tetangga yang tidak relevan (Bagaskoro et al.,

2018). Maka dari itu penggunaan seleksi fitur dapat membantu mengurangi atau menyeleksi fitur-fitur yang dianggap penting dan membuang fitur- fitur yang tidak diperlukan.

Chi-Square adalah salah satu metode seleksi fitur yang relatif sederhana dan Chi-Square umumnya digunakan untuk data kategorikal atau data yang dapat dikelompokkan menjadi kategori. Metode Chi-Square dapat digunakan untuk memilih fitur yang dianggap penting untuk digunakan pada proses klasifikasi dan dapat menghapus fitur yang tidak berpengaruh terhadap kelas target (Taufiqurrahman et al., 2021). *Chi Square* merupakan salah satu seleksi fitur yang dapat menghilangkan beberapa fitur tanpa beresiko mengurangi tingkat akurasi. *Chi-square* adalah metode seleksi fitur yang dalam perhitungannya memanfaatkan distribusi statistika dengan mengukur nilai ketergantungan antara *term* dan kategori (Harish et al., 2021). Pada penelitian (Suharno et al., 2017) membuktikan seleksi fitur *Chi Square* membantu metode *K-Nearest Neighbor* dalam mengklasifikasi dokumen pengaduan sambat online. Berdasarkan penelitian tersebut dengan menggunakan seleksi fitur menghasilkan *F-measure* terbaik sebesar 78% dengan seleksi fitur sebesar 25%. Dari penelitian ini menunjukkan bahwa hasil yang didapatkan dengan menggunakan seleksi fitur lebih baik daripada tanpa adanya proses seleksi fitur dalam klasifikasi dokumen teks. Maka dari itu, pada penelitian ini peneliti akan menggunakan metode *K-Nearest Neighbor* dengan menggunakan seleksi fitur yaitu *Chi Square* untuk pengklasifikasian dokumen.

1. 3 Rumusan Masalah

Rumusan masalah dalam penelitian ini adalah :

1. Bagaimana mengembangkan sistem pengklasifikasian SMS berbahasa Indonesia menggunakan *K-Nearest neighbor* dan *Chi square* ?
2. Bagaimana tingkat akurasi pengklasifikasian SMS berbahasa Indonesia menggunakan metode *K-Nearest Neighbor* dan *Chi square* ?

1. 4 Tujuan Penelitian

Tujuan dari penelitian ini adalah sebagai berikut.

1. Mengklasifikasikan SMS menggunakan metode *K-Nearest Neighbor* dan *Chi square* pada SMS berbahasa Indonesia.
2. Mengetahui tingkat akurasi metode *K-Nearest Neighbor* dan *Chi square* pada pengklasifikasian SMS berbahasa Indonesia.

1. 5 Manfaat Penelitian

Manfaat dari penelitian ini adalah sebagai berikut.

1. Perangkat lunak yang dihasilkan dapat digunakan untuk mengklasifikasikan SMS berbahasa Indonesia.
2. Penelitian ini dapat digunakan pada penelitian selanjutnya di bidang terkait.

1. 6 Batasan Masalah

Batasan masalah dari penelitian ini adalah sebagai berikut.

1. Data yang digunakan merupakan data yang sudah diberi label.

2. Jenis data yang digunakan dalam penelitian ini adalah 1143 data SMS berbahasa Indonesia.
3. Label yang digunakan yaitu *Spam* dan non *Spam*

1. 7 Sistematika Penulisan

BAB I. PENDAHULUAN

Bab ini membahas tentang latar belakang, rumusan masalah, tujuan penelitian, manfaat penelitian, batasan masalah, dan sistematika penulisan yang digunakan dalam menyusun laporan akhir ini.

BAB II. TINJAUAN PUSTAKA

Bab ini menjelaskan landasan teori yang digunakan dalam penelitian. Bab ini memuat penjelasan tentang penelitian terdahulu yang relevan dengan penelitian ini, penjelasan tentang *K-Nearest Neighbor* dengan *Chi Square*, serta penjelasan lain yang berkaitan dengan penelitian ini.

BAB III. METODOLOGI PENELITIAN

Bab ini membahas tahapan-tahapan yang akan dilakukan dalam penelitian ini. Setiap rencana tahapan penelitian dijelaskan secara rinci dengan mengacu pada kerangka kerja. Pada bagian akhir bab ini berisi tentang perancangan manajemen proyek dalam melakukan penelitian.

BAB IV. PENGEMBANGAN PERANGKAT LUNAK

Pada bab ini akan dibahas mengenai perancangan perangkat lunak yang akan dibangun pada penelitian ini.

BAB V. HASIL DAN ANALISIS PENELITIAN

Bab ini akan menampilkan hasil pengujian berdasarkan langkah-langkah yang telah direncanakan. Analisis tersebut diberikan sebagai dasar kesimpulan yang ditarik dalam penelitian ini.

BAB VI. KESIMPULAN DAN SARAN

Bab ini berisi kesimpulan dari semua uraian pada bab-bab sebelumnya dan juga berisi saran-saran yang diharapkan dapat berguna dalam pengembangan perangkat lunak ini selanjutnya.

1.8 Kesimpulan

Penelitian akan membahas tentang klasifikasi dokumen berbahasa Indonesia menggunakan *K-Nearest Neighbor* dengan *Chi Squar*e.

Daftar Pustaka

Amrizal, V. (2018). Penerapan Metode Term Frequency Inverse Document Frequency (Tf-Idf) dan Cosine Similarity Pada Sistem Temu Kembali Informasi Untuk Mengetahui Syarah Hadits berbasis web (Studi Kasus: Hadits Shahih Bukhari-Muslim). *Jurnal Teknik Informatika*, *11*(2), 149–164. https://doi.org/10.15408/jti.v11i2.8623

Azis, H., Purnawansyah, P., Fattah, F., & Putri, I. P. (2020). Performa Klasifikasi K-NN dan Cross Validation Pada Data Pasien Pengidap Penyakit Jantung. *ILKOM Jurnal Ilmiah*, *12*(2), 81–86. https://doi.org/10.33096/ilkom.v12i2.507.81-86

Bagaskoro, G. N., Fauzi, M. A., & Adikara, P. P. (2018). *Penerapan Klasifikasi Tweets Pada Berita Twitter Menggunakan Metode K-Nearest Neighbor Dan Query Expansion Berbasis Distributional Semantic* (Vol. 2, Issue 10). http://j-ptiik.ub.ac.id

Cindy Chairunnisa, Iin Ernawati, & Mayanda Mega Santoni. (2022). Klasifikasi Sentimen Ulasan Pengguna Aplikasi PeduliLindungi di Google Play Menggunakan Algoritma Support Vector Machine dengan Seleksi Fitur Chi-Square. *JURNAL INFORMATIK*, *18*(1).

Dwiyansaputra, R., Satya Nugraha, G., Bimantoro, F., & Aranta, A. (2021a). *Deteksi SMS Spam Berbahasa Indonesia Menggunakan Tf-Idf dan Stochastic Gradient Descent Classifier (Indonesian SMS Spam Detection using TF-IDF and Stochastic Gradient Descent Classifier)*. http://jtika.if.unram.ac.id/index.php/JTIKA/

Dwiyansaputra, R., Satya Nugraha, G., Bimantoro, F., & Aranta, A. (2021b). *Deteksi SMS Spam Berbahasa Indonesia Menggunakan Tf-Idf dan Stochastic Gradient Descent Classifier (Indonesian SMS Spam Detection using TF-IDF and Stochastic Gradient Descent Classifier)*. http://jtika.if.unram.ac.id/index.php/JTIKA/

Geraldy, H., & Rahmatuti Maghfiroh, L. (2021). Penerapan Klasifikasi Kueri untuk Meningkatkan Efektivitas Mesin Pencari (Implementation of Query Classification to Improve Effectiveness of Search Engine). In *Seminar Nasional Official Statistics*. www.bps.go.id

Harish, H. N., Al Faraby, S., & Dwifebri, M. (2021). Klasifikasi Multi Label Pada Hadis Bukhari Terjemahan Bahasa Indonesia menggunakan Random Forest, Mutual Information, dan Chi-Square. *E-Proceeding of Engineering*, *8*(5), 10583.

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

Istighfarizky, F., ER Sanjaya A, N., Widiartha M, I., Astuti G, L., Putra Cahyadi Anom Ngurah G, I., & Suhartana Gede K, I. (2022). Klasifikasi Jurnal menggunakan Metode KNN dengan Mengimplementasikan Perbandingan Seleksi Fitur. *Jurnal Elektronik Ilmu Komputer Udayana* , *11*(1). https://scholar.google.com

Laksono, E. P., & Wicaksono, A. (2022). Penyaringan Spam email menggunakan K-Means. *Jurnal Spektro*, *5*(2).

Listiowarni, I., & Setyaningsih, E. R. (2018). Feature Selection Chi-Square dan K-NN pada Pengkategorian Soal Ujian Berdasarkan Cognitive Domain Taksonomi Bloom. In *Jurnal Komputer Terapan* (Vol. 4, Issue 1). http://jurnal.pcr.ac.id

Meriohengki, & Wahyudi, M. (2020). Klasifikasi Algoritma Naïve Bayes dan SVM Berbasis PSO Dalam Memprediksi Spam Email Pada Hotline-Sapto. *Paradigma – Jurnal Informatika Dan Komputer*, *22*(1). https://doi.org/10.31294/p.v21i2

Negara C, I., & Prabowo, A. (2018). *Penggunaan Uji Chi–Square Untuk Mengetahui Pengaruh Tingkat Pendidikan dan Umur Terhadap Pengetahuan Penasun Mengenai Hiv–Aids di Provinsi DKI Jakarta*.

Nisa, A., Darwiyanto, E., & Asror, I. (2019). *Analisis Sentimen Menggunakan Naive Bayes Classifier dengan Chi-Square Feature Selection Terhadap Penyedia Layanan Telekomunikasi*.

Noviana, D., Susanti, Y., & Susanto, I. (2019). *Analisis Rekomendasi Penerima Beasiswa Menggunakan Algoritma K-Nearest Neighbor (K-NN) dan Algoritma c4.5*.

Nurul Chasanah, D., & Mutoi Siregar, A. (2022). Klasifikasi Kelayakan Siswa dalam Menentukan Kelas Unggulan Menggunakan Algoritma K-Nearest Neighbor. *Scientific Student Journal for Information, Technology and Science*, *III*(1), 51.

Perdana, R. S., & Fauzi, M. A. (2018). *Klasifikasi Dokumen Twitter Untuk Mengetahui Karakter Calon Karyawan Menggunakan Algoritme K-Nearest Neighbor (KNN)*. https://www.researchgate.net/publication/322959490

Puspita Hidayanti, W., & Yahya. (2020). Penerapan Algoritma K-Nearest Neighbor Untuk Klasifikasi Efektivitas Penjualan Vape (Rokok Elektrik) pada "Lombok Vape On." *Jurnal Informatika Dan Teknologi*, *3*(2).

Randhika, M. N., Young, J. C., Suryadibrata, A., & Mandala, H. (2021). Implementasi Algoritma Complement dan Multinomial Naïve Bayes Classifier Pada Klasifikasi Kategori Berita Media Online. *Ultimatics : Jurnal Teknik Informatika*, *13*(1).

Reviantika, F., Azhar, Y., & Marthasari, G. I. (2021). Analisis Klasifikasi SMS Spam Menggunakan Logistic Regression. In *Jurnal Sistem Cerdas*. APIC. https://www.cnbcindonesia.com

Saleh, H., & Hamria. (2023). K-Nearest Neighbor Berbasis Seleksi Atribut Chi Square Untuk Klasifikasi Penerima Beasiswa Kurang Mampu. *Jurnal SIMETRIS*, *14*(1).

Salma, Dewanta, F., & Abdillah, M. (2022). Klasifikasi Beban Listrik dengan Machine Learning Menggunakan Metode K-Nearest Neighbor. *RESISTOR (Elektronika Kendali Telekomunikasi Tenaga Listrik Komputer)*, *5*(2).

Samsudin, N. M., Mohd Foozy, C. F. B., Alias, N., Shamala, P., Othman, N. F., & Wan Din, W. I. S. (2019). Youtube spam detection framework using naïve bayes and logistic regression. *Indonesian Journal of Electrical Engineering and Computer Science*, *14*(3), 1508–1517. https://doi.org/10.11591/ijeecs.v14.i3.pp1508-1517

Shinta Sari, W., & Atika Sari, C. (2022). Klasifikasi Bunga Mawar Menggunakan KNN dan Ekstraksi Fitur GLCM dan HSV. *SKANIKA: Sistem Komputer Dan Teknik Informatika*, *5*(2), 145–156.

Suharno, C. F., Fauzi, M. A., & Perdana, R. S. (2017). *Klasifikasi Teks Bahasa Indonesia pada Dokumen Pengaduan Sambat Online menggunakan Metode K-Nearest Neighbors (K-NN) dan Chi-Square* (Vol. 1, Issue 10). http://j-ptiik.ub.ac.id

Taufiqurrahman, F., Al Faraby, S., & Purbolaksono, M. D. (2021). *Klasifikasi Teks Multi Label pada Hadis Terjemahan Bahasa Indonesia Menggunakan Chi-Square dan SVM*.