

**IMPLEMENTASI ALGORITMA NAIVE BAYES UNTUK
FILTRASI SPAM KOMENTAR PADA YOUTUBE**



SKRIPSI

*diajukan untuk memenuhi sebagai syarat
untuk memperoleh gelar Sarjana Bisnis Program Studi Bisnis Digital*

Oleh

Faiz Jauhari Makarim Riza

NIM 2107090

**PROGRAM STUDI S1 BISNIS DIGITAL
UNIVERSITAS PENDIDIKAN INDONESIA
KAMPUS TASIKMALAYA**

2025

IMPLEMENTASI ALGORITMA NAIVE BAYES UNTUK FILTRASI SPAM KOMENTAR PADA YOUTUBE

Oleh
Faiz Jauhari Makarim Riza

**Diajukan untuk memenuhi sebagian dari Persyaratan Memperoleh Gelar
Sarjana Bisnis Program Studi Bisnis Digital**

© Faiz Jauhari Makarim Riza
Universitas Pendidikan Indonesia
Juli 2025

**Hak Cipta dilindungi undang-undang. Skripsi ini tidak boleh diperbanyak
seluruhnya atau sebagian, Dengan dicetak ulang, difoto copy, atau cara
lainya tanpa izin dari penulis**

FAIZ JAUHARI MAKARIM RIZA

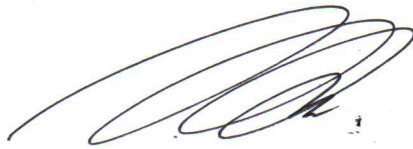
LEMBAR PENGESAHAN SKRIPSI

FAIZ JAUHARI MAKARIM RIZA

**IMPLEMENTASI ALGORITMA NAIVE BAYES UNTUK FILTRASI
SPAM KOMENTAR PADA YOUTUBE**

disetujui dan disahkan oleh dosen pembimbing:

Pembimbing I



Rangga Gelar Guntara S.Kom., M.kom.

NIP. 920200819880616101

Pembimbing II



Muhammad Rizki Nugraha, S.Pd., M.T.

NIP. 920230219930823101

Mengetahui,

Ketua Program Studi S1 Bisnis Digital UPI

Kampus Tasikmalaya



Dr. Syti Sarah Maesaroh, S.P., M.M.

NIP. 920190219900625201

IMPLEMENTASI ALGORITMA NAIVE BAYES UNTUK FILTRASI SPAM KOMENTAR PADA YOUTUBE

ABSTRAK

Oleh

Faiz Jauhari Makarim Riza

Perkembangan interaksi pengguna di platform YouTube turut memunculkan permasalahan baru, salah satunya adalah maraknya komentar spam yang mengandung promosi perjudian online. Komentar semacam berdampak negatif terhadap komunitas yang terdapat di kanal. Penelitian ini bertujuan untuk membangun sistem klasifikasi komentar spam menggunakan algoritma *Naive Bayes*. Proses pengembangan model mengikuti tahapan CRISP-DM, dimulai dari pengumpulan data komentar menggunakan YouTube API, yang kemudian dilanjutkan dengan *preprocessing* teks melalui tahapan *Unicode normalization*, *case folding*, *tokenizing*, *stopword removal*, *filtering*, dan *labelling*. Pembobotan kata dilakukan dengan metode TF-IDF untuk menghasilkan representasi fitur yang optimal. Evaluasi model dilakukan dengan teknik *K-Fold Cross Validation* dan analisis *confusion matrix*. Hasil evaluasi menunjukkan bahwa model mencapai performa yang tinggi, dengan *akurasi* sebesar 97,1%, *precision* 96,4%, *recall* 95,6%, dan *f1-score* 96%. Model yang dibangun kemudian diterapkan dalam bentuk aplikasi berbasis *Command Line Interface* (CLI) yang dapat digunakan oleh pemilik kanal untuk mendeteksi dan menghapus komentar spam secara langsung. Berdasarkan hasil pengujian, sistem menunjukkan efektivitas yang tinggi. Penelitian ini menunjukkan bahwa kombinasi *preprocessing* yang tepat dan pemilihan algoritma yang sesuai dapat menghasilkan sistem deteksi spam yang akurat.

Kata kunci: YouTube, Komentar Spam, *Naive Bayes*, TF-IDF, Klasifikasi, CLI

IMPLEMENTASI ALGORITMA NAIVE BAYES UNTUK FILTRASI SPAM KOMENTAR PADA YOUTUBE

ABSTRACT

By

Faiz Jauhari Makarim Riza

The growing interaction among users on the YouTube platform has led to new challenges, one of which is the increasing prevalence of spam comments promoting online gambling. These types of comments negatively impact the sense of community within YouTube channels. This study aims to develop a spam comment classification system using the Naive Bayes algorithm. The model development process follows the CRISP-DM framework, starting with data collection through the YouTube API, followed by a series of text preprocessing steps including Unicode normalization, case folding, tokenizing, stopword removal, filtering, and manual labelling. Word weighting is performed using the TF-IDF method to produce optimal feature representations. Model evaluation was carried out using the K-Fold Cross Validation technique and confusion matrix analysis. The results show that the model achieved strong performance, with an accuracy of 97.1%, precision of 96.4%, recall of 95.6%, and an F1-score of 96%. The trained model was then deployed as a Command Line Interface (CLI) application, allowing channel owners to directly detect and remove spam comments. Based on testing outcomes, the system demonstrated high effectiveness. This study highlights that combining proper preprocessing techniques with the appropriate classification algorithm can result in an accurate and practical spam detection system.

Keywords: YouTube, Spam Comments, Naive Bayes, TF-IDF, Classification, CLI

DAFTAR ISI

KATA PENGANTAR	ii
UCAPAN TERIMA KASIH.....	iii
ABSTRAK	v
ABSTRACT	vi
DAFTAR ISI.....	vii
DAFTAR TABEL.....	x
DAFTAR GAMBAR	xi
DAFTAR LAMPIRAN.....	xiii
BAB I PENDAHULUAN	1
1.1 Latar Belakang Penelitian	1
1.2 Rumusan Masalah Penelitian	7
1.3 Tujuan Penelitian	7
1.4 Batasan Penelitian	8
1.5 Manfaat Penelitian	8
1.5.1 Manfaat Teoritis	8
1.5.2 Manfaat Praktis	8
1.6 Struktur Organisasi Skripsi	8
BAB II TINJAUAN PUSTAKA DAN KERANGKA PEMIKIRAN	11
2.1 Tinjauan Pustaka	11
2.1.1 Spam.....	11
2.1.2 CRISP-DM.....	12
2.1.3 Google Cloud Platform	14
2.1.4 Application Programming Interface (API)	15
2.1.5 OAuth.....	15
2.1.6 Text Preprocessing.....	16
2.1.7 Python	18
2.1.8 Visual Studio Code.....	21
2.1.9 Unicode Standard	21
2.1.10 Machine Learning	22
2.1.11 Klasifikasi	24

2.1.12	<i>K-fold Cross Validation</i>	24
2.1.13	Algoritma Naive Bayes	25
2.1.14	<i>Confusion Matrix</i>	28
2.1.15	<i>Term Frequency-Inverse Document Frequency (TF-IDF)</i>	30
2.1.17	Penelitian Terdahulu	32
2.2	Kerangka Pemikiran.....	35
BAB III METODE PENELITIAN		37
3.1	Metode Penelitian	37
3.2	Desain Penelitian.....	37
3.3	Penerapan Metode CRISP-DM.....	39
3.3.1	<i>Business Understanding</i>	39
3.3.2	<i>Data Understanding</i>	40
3.3.3	<i>Data Preparation</i>	42
3.3.4	<i>Modelling</i>	43
3.3.5	<i>Evaluation</i>	43
3.3.6	<i>Deployment</i>	44
3.4	Pengujian.....	44
3.4.1	Populasi.....	45
3.4.2	Sampel.....	45
3.4.3	Teknik Pengumpulan Data.....	45
BAB IV HASIL DAN PEMBAHASAN		46
4.1	Analisis.....	46
4.1.1	Analisis Permasalahan	46
4.1.2	Analisis Prosedur yang Sedang Berjalan	47
4.2	Perancangan Sistem Menggunakan Metode CRISP-DM	48
4.2.1	<i>Business Understanding</i>	48
4.2.2	<i>Data Understanding</i>	48
4.2.3	<i>Data Preparation</i>	50
4.2.4	Modeling	55
4.2.5	Evaluation	57
4.2.6	<i>Deployment</i>	59
4.3	Tahap Implementasi.....	59

4.3.1	Persiapan Data.....	59
4.3.2	<i>Text Preprocessing</i>	62
4.3.3	Pembobotan Term Frequency-Inverse Document Frequency.....	64
4.3.4	<i>K-Fold Cross Validation</i>	65
4.3.5	Klasifikasi <i>Naive Bayes</i>	66
4.3.6	Evaluasi <i>Confusion Matrix</i>	67
4.3.7	<i>Python CLI</i>	69
4.4	Pengujian.....	72
4.5	Pembahasan.....	81
4.5.1	Hasil Penerapan Algoritma <i>Naive Bayes</i>	81
4.5.2	Perbandingan Hasil Penelitian dengan Hasil Penelitian Terdahulu..	84
4.5.3	Penggunaan Command Line Interface	85
BAB V SIMPULAN DAN SARAN		89
5.1	Simpulan.....	89
5.2	Saran.....	90
DAFTAR PUSTAKA		91
LAMPIRAN.....		98

DAFTAR TABEL

Tabel 1.1 Keluhan Para Pengguna Platform Youtube	5
Tabel 2.1 <i>Confusion Matrix</i>	29
Tabel 2.2 Perbandingan dengan Penelitian Terdahulu.....	32
Tabel 4.1 Analisis Masalah dan Solusi yang Ditawarkan.....	46
Tabel 4.2 Atribut Data.....	49
Tabel 4.3 Hasil Unicode Normalization.....	51
Tabel 4.4 Hasil Case Folding	51
Tabel 4.5 Hasil <i>Tokenizing</i>	52
Tabel 4.6 Hasil Stopwords Removal.....	52
Tabel 4.7 Hasil Filtering	53
Tabel 4.8 Proporsi Data	55
Tabel 4.9 Skenario Alpha Testing dengan Metode Black Box.....	72
Tabel 4.10 Hasil pengujian Black Box Testing	73
Tabel 4.11 Daftar Pertanyaan.....	76
Tabel 4.12 Skala Penilaian.....	77
Tabel 4.13 Presentase Skala Penilaian	77
Tabel 4.14 Hasil Kuesioner 1	78
Tabel 4.15 Hasil Kuesioner 2.....	78
Tabel 4.16 Hasil Kuesioner 3.....	79
Tabel 4.17 Hasil Kuesioner 4.....	79
Tabel 4.18 Hasil Kuesioner 5.....	80

DAFTAR GAMBAR

Gambar 1.1	Perputaran Dana Judi Online di Indonesia dari Tahun 2017 - 2024..	2
Gambar 1.2	Promosi Judi Online pada Kolom Komentar di Video YouTube.....	3
Gambar 1.3	Fitur Moderasi YouTube untuk Menyaring Kata-kata	4
Gambar 1.4	Pernyataan Pemilik Kanal YouTube	4
Gambar 2.1	Cycle CRISP-DM	13
Gambar 2.2	Tahapan Text Preprocessing	16
Gambar 2.3	Ilustrasi K-Fold Cross Validation	24
Gambar 2.4	Kerangka Pemikiran	35
Gambar 3.1	Alur Metode Penelitian	37
Gambar 3.2	Diagram Desain Penelitian dengan Metode CRISP-DM.....	38
Gambar 3.3	Teknik Pengumpulan Data	41
Gambar 4.1	Fitur Moderasi YouTube untuk Menyaring Kata-kata	47
Gambar 4.2	Fitur Moderasi YouTube untuk Menyaring Kata-kata Manual.....	47
Gambar 4.3	Data Komentar Mentah dengan Format JSON	49
Gambar 4.4	Tahapan Data Preparation.....	50
Gambar 4.5	TF-IDF Document Vectors.....	54
Gambar 4.6	Cross Validation Score Distributions	56
Gambar 4.7	Confusion Matrix Naive Bayes	57
Gambar 4.8	Import Library	60
Gambar 4.9	Data Komentar dengan Format JSON	60
Gambar 4.10	Youtube API Comment Scraper	60
Gambar 4.11	Pengelompokan Teks.....	61
Gambar 4.12	Dataset untuk Text Preprocessing	62
Gambar 4.13	Cuplikan Kode Normalisasi Unicode	62
Gambar 4.14	Cuplikan Kode untuk Case Folding.....	62
Gambar 4.15	Cuplikan Kode Tokenizing	63
Gambar 4.16	Function untuk Stopwords Removal	63
Gambar 4.17	Labelling	64
Gambar 4.18	Implementasi TF-IDF	64
Gambar 4.19	Hyperparameter TF-IDF	65
Gambar 4.20	Cuplikan Kode K-Fold Cross Validation	65
Gambar 4.21	Proses K-Fold	66
Gambar 4.22	Implementasi Algoritma Naive Bayes Menggunakan Python.....	67
Gambar 4.23	Evaluasi Kinerja Algoritma Naive Bayes menggunakan K-Fold..	67
Gambar 4.24	Cuplikan Kode Confusion Matrix	68
Gambar 4.25	Metrik Evaluasi Model	69
Gambar 4.26	Pyinstaller	69
Gambar 4.27	Proses Konversi Script.....	69
Gambar 4.28	Tampilan Awal Aplikasi.....	70
Gambar 4.29	Tampilan CLI Sistem Deteksi Spam	71
Gambar 4.30	TF-IDF Document Vectors.....	82

Gambar 4.31 Regular Expression.....	83
Gambar 4.32 TF-IDF Document Vectors (After).....	83
Gambar 4.33 Halaman Konfigurasi Credential	86
Gambar 4.34 Susunan Aplikasi di File Explorer.....	87
Gambar 4.35 Proses Autentikasi Akun.....	87
Gambar 4.36 Tahapan Interaksi Pengguna.....	88

DAFTAR LAMPIRAN

Lampiran 1. Jadwal Penelitian	98
Lampiran 2. Source Code Program.....	98
Lampiran 3. Cuplikan Dataset.....	101
Lampiran 4. Hasil Form Kuesioner.....	102

DAFTAR PUSTAKA

- Abdullah, A. O., Ali, M. A., Karabatak, M., & Sengur, A. (2018). A comparative analysis of common YouTube comment spam filtering techniques. *2018 6th International Symposium on Digital Forensic and Security (ISDFS)*, 1–5. <https://doi.org/10.1109/ISDFS.2018.8355315>
- Agustini, A., & Kurniawan, W. J. (2020). Sistem e-learning do'a dan iqro'dalam peningkatan proses pembelajaran pada TK Amal Ikhlas. *Jurnal Mahasiswa Aplikasi Teknologi Komputer dan Informasi (JMApTeKsi)*, 1(3), 154–159.
- Alias, N., Foozy, C. F. M., & Ramli, S. N. (2019). Video spam comment features selection using machine learning techniques. *Indonesian Journal of Electrical Engineering and Computer Science*, 15(2), 1046–1053. <https://doi.org/10.11591/ijeecs.v15.i2.pp1046-1053>
- Allen, J. D., Anderson, D., Becker, J., Cook, R., Davis, M., Edberg, P., Everson, M., Freytag, A., Iancu, L., Ishida, R., Jenkins, J. H., Lunde, K., McGowan, R., Moore, L., Muller, E., Phillips, A., Pournader, R., Suignard, M., & Whistler, K. (1991). *The Unicode Standard Version 7.0-Core Specification The Unicode Consortium*. <http://www.unicode.org/versions/Unicode7.0.0/>
- Aziz, A., Mohd Foozy, C. F., Shamala, P., & Suradi, Z. (2018). YouTube Spam Comment Detection Using Support Vector Machine and K-Nearest Neighbor. *Indonesian Journal of Electrical Engineering and Computer Science*, 12(2), 612. <https://doi.org/10.11591/ijeecs.v12.i2.pp612-619>
- Bahori, B. (2024). Tinjauan Hukum Terhadap Perjudian Online Menurut Undang-Undang Nomor 19 Tahun 2016 Tentang Informasi Dan Transaksi Elektronik. *SINERGI : Jurnal Riset Ilmiah*, 1(10), 964–974. <https://doi.org/10.62335/sinergi.v1i10.886>
- Bakhtiar, S. H., & Adilah, A. N. (2024). Fenomena Judi Online : Faktor, Dampak, Pertanggungjawaban Hukum. *Innovative: Journal Of Social Science Research*, 4(3), 1016–1026. <https://doi.org/10.31004/innovative.v4i3.10547>
- Christian, H., Agus, M. P., & Suhartono, D. (2016). Single Document Automatic Text Summarization using Term Frequency-Inverse Document Frequency (TF-IDF). *ComTech: Computer, Mathematics and Engineering Applications*, 7(4), 285. <https://doi.org/10.21512/comtech.v7i4.3746>
- Cooksey, B., Landers, B., & Schreiber, D. (2014). *An Introduction to APIs*.
- Cortesi, D., Bajo, G., Caban, W., & McMillan, G. (2025). *PyInstaller Manual*. <https://pyinstaller.org/en/stable/#>.

- Daniel, J., & Martin, J. H. (2024). *Naive Bayes, Text Classification, and Sentiment*.
- Denny, M., & Spirling, A. (2017). *Text Preprocessing For Unsupervised Learning: Why It Matters, When It Misleads, And What To Do About It*.
- Developer. (2025). *YouTube Data API*.
<https://developers.google.com/youtube/v3/getting-started>.
- Fathurohman, A. (2021). *Machine Learning untuk Pendidikan: Mengapa dan Bagaimana*. 1(3), 57–62.
- Fattahi, J., & Mejri, M. (2021). SpaML: A bimodal ensemble learning spam detector based on NLP techniques. *2021 IEEE 5th International Conference on Cryptography, Security and Privacy, CSP 2021*, 107–112.
<https://doi.org/10.1109/CSP51677.2021.9357595>
- Fauziyah, I. (2020). *Keputusan Menteri Ketenagakerjaan Nomor 299 Tahun 2020*. Kementerian Ketenagakerjaan.
- Fernando, D., & Guntara, R. G. (2024). Model Klasifikasi Penyebab Turnover Karyawan Menggunakan Kerangka Kerja CRISP-DM. *J-INTECH*, 12(02), 383–392. <https://doi.org/10.32664/j-intech.v12i02.1502>
- Fernando, J. R., Budiraharjo, R., & Haganusa, E. (2019). Spam Classification on 2019 Indonesian President Election Youtube Comments Using Multinomial Naïve-Bayes. *Indonesian Journal of Artificial Intelligence and Data Mining*, 2(1). <https://doi.org/10.24014/ijaidm.v2i1.6445>
- George, A. (2022). *Python Text Mining: Perform Text Processing, Word Embedding, Text Classification and Machine Translation (English Edition)*. BPB Publications.
- Ghatasheh, N., Altaharwa, I., & Aldebei, K. (2022). Modified Genetic Algorithm for Feature Selection and Hyper Parameter Optimization: Case of XGBoost in Spam Prediction. *IEEE Access*, 10, 84365–84383.
<https://doi.org/10.1109/ACCESS.2022.3196905>
- Gillam, R. (2002). *Unicode Demystified A Practical Programmer's Guide to the Encoding Standard*.
- Ginting, R., Yulistiyono, A., Rauf, A., Manullang, S. O., Siahaan, A. L. S., Kussanti, D. P., PS, T. E. A., Djaya, T. R., Ayu, A. S., & Effendy, F. (2021). *Etika komunikasi dalam media sosial: Saring sebelum sharing* (Vol. 1). Penerbit Insania.
- Guntara, R. (2023). Aplikasi Deteksi Phising Berbasis Android Menggunakan Metode Pengembangan Perangkat Lunak DSRM. *Jurnal Minfo Polgan*, 12(1), 303–310. <https://doi.org/10.33395/jmp.v12i1.12379>

Faiz Jauhari Makarim Riza, 2025

IMPLEMENTASI ALGORITMA NAIVE BAYES UNTUK FILTRASI SPAM KOMENTAR PADA YOUTUBE

Universitas Pendidikan Indonesia | repository.upi.edu | perpustakaan.upi.edu

- Hackeling, G. (2014). *Mastering machine learning with scikit-learn: apply effective learning algorithms to real-world problems using scikit-learn*. Packt Publishing.
- Hadjon, R. P., Restyandito, R., & Rahardjo, W. S. (2015). Implementasi Metode Rest Request Pada Youtube Web Service Untuk Representasi Informasi Berbasis Timeline. *Jurnal Informatika*, 10(1). <https://doi.org/10.21460/inf.2014.101.321>
- Han, J., Pei, J., & Tong, H. (2022). *Data mining: concepts and techniques*. Morgan kaufmann.
- Hani, D. (2023). *Klasifikasi Masalah Pada Komunitas Marah-Marah Di Twitter Menggunakan Bidirectional Long Short-Term Memory*.
- Hao, J., & Ho, T. K. (2019). Machine Learning Made Easy: A Review of Scikit-learn Package in Python Programming Language. *Journal of Educational and Behavioral Statistics*, 44(3), 348–361. <https://doi.org/10.3102/1076998619832248>
- Hayoung. (2021). A YouTube Spam Comments Detection Scheme Using Cascaded Ensemble Machine Learning Model. *IEEE Access*, 9, 144121–144128. <https://doi.org/10.1109/ACCESS.2021.3121508>
- Jacob, P. M., & Prasanna, M. (2016). A Comparative analysis on Black box testing strategies. *2016 International Conference on Information Science (ICIS)*, 1–6. <https://doi.org/10.1109/INFOSCI.2016.7845290>
- Jáñez-Martino, F., Alaiz-Rodríguez, R., González-Castro, V., Fidalgo, E., & Alegre, E. (2023). A review of spam email detection: analysis of spammer strategies and the dataset shift problem. *Artificial Intelligence Review*, 56(2), 1145–1173.
- Jawed, A. (2025, Juni 26). *How to Delete All Comments on YouTube? (Step-by-Step Process)*. Hollyland. https://www.hollyland.com/blog/tips/delete-all-comments-on-youtube?utm_source=chatgpt.com
- Jing, L.-P., Huang, H.-K., & Shi, H.-B. (2002). *Improved Feature Selection Approach Tf-idf In Text Mining*. <http://www.cs.cmu.edu/~mccallumhow/rainbow>
- Kudo, T., & Richardson, J. (2018). *SentencePiece: A simple and language independent subword tokenizer and detokenizer for Neural Text Processing*. <http://arxiv.org/abs/1808.06226>
- Mahmoud, T. M., El Nashar, A. I., Abd-El-Hafeez, T., & Khairy, M. (2014). An Efficient Three-phase Email Spam Filtering Technique. Dalam *British*

Journal of Mathematics & Computer Science (Vol. 4, Nomor 9).
www.sciencedomain.org

- Makarin, A. A., & Astuti, L. (2023). Faktor yang Mempengaruhi Mahasiswa Melakukan Perjudian Online. *Indonesian Journal of Criminal Law and Criminology (IJCLC)*, 3(3), 180–189. <https://doi.org/10.18196/ijclc.v3i3.17674>
- Martinez-Plumed, F., Contreras-Ochando, L., Ferri, C., Hernandez-Orallo, J., Kull, M., Lachiche, N., Ramirez-Quintana, M. J., & Flach, P. (2021). CRISP-DM Twenty Years Later: From Data Mining Processes to Data Science Trajectories. *IEEE Transactions on Knowledge and Data Engineering*, 33(8), 3048–3061. <https://doi.org/10.1109/TKDE.2019.2962680>
- Marwah, H. (2024). *Menko Polkam Sebut Perputaran Uang Judi Online Mencapai Rp900 Triliun di Tahun 2024*. <https://www.tempo.co/ekonomi/menko-polkam-sebut-perputaran-uang-judi-online-mencapai-rp900-triliun-di-tahun-2024-1171463>.
- Maulana, R. (2023). *Sistem Pendeteksi Berita Palsu Pada Konten Berita Berbahasa Indonesia Menggunakan Metode Long Short-Term Memory (Lstm)*. Universitas Pendidikan Indonesia.
- Muhamad, N. (2024). *Perkembangan Transaksi Judi Online sampai Semester I 2024*. Databoks. <https://databoks.katadata.co.id/teknologi-telekomunikasi/statistik/6732d2f4a4fae/perkembangan-transaksi-judi-online-sampai-semester-i-2024>
- Murphy, K. P. (2012). *Machine learning: a probabilistic perspective*. MIT press.
- Nainggolan, V., Rondonuwu, S. A., & Waleleng, G. J. (2018). *Peranan Media Sosial Instagram Dalam Interaksi Sosial Antar Mahasiswa Fakultas Ilmu Sosial Dan Politik Unsrat Manado*.
- Novianto, M. A., & Munir, S. (2022). Analisis Dan Implementasi Restful Api Guna Pengembangan Sistem Informasi Akademik Pada Perguruan Tinggi. *Jurnal Informatika Terpadu*, 8, 47–61. <https://journal.nurulfikri.ac.id/index.php/JIT>
- Nurdiana, M., Aisyah, N., & Ilham, S. N. (2023). *Fenomena Judi Online Di Daerah Jakarta Selatan*. 2. <https://doi.org/https://doi.org/10.53947/perspekt.v2i2.271>
- Nurdiansyah, A., & Kanda, A. (2024). Bahaya Judi Online : Dampak Sosial, Ekonomi, Dan Kesehatan. *Student Scientific Creativity Journal*, 2(1), 305–310. <https://doi.org/10.55606/sscj-amik.v2i1.2807>
- Paramartha, P., Dewi, A., & Seputra, P. (2021). Sanksi Pidana terhadap Para Pemasang dan Promosi Iklan Bermuatan Konten Judi Online. *Jurnal*

- Preferensi Hukum*, 2(1), 156–160. <https://doi.org/10.22225/jph.2.1.3062.156-160>
- Pratama, I. K. Y., Rumanti, A. A., & Arini, I. Y. (2025). Perancangan Sistem Informasi Pariwisata Berbasis Web untuk Manajemen dan Akseibilitas Informasi Wisata di Kabupaten Buleleng Menggunakan Metode Waterfall. *eProceedings of Engineering*, 12(1).
- Pressman, R. S. (2010). A practitioner's approach. *Software Engineering*, 2, 41–42.
- Purohit, N. S., Angadi, A. B., Bhat, M., & Gull, K. C. (2015). Crawling through web to extract the data from Social networking site - Twitter. *2015 National Conference on Parallel Computing Technologies (PARCOMPTECH)*, 1–6. <https://doi.org/10.1109/PARCOMPTECH.2015.7084522>
- Pusiknas. (2025). *Kecanduan Judi Online Akibatkan Frustasi, Depresi, lalu Bunuh Diri*. https://pusiknas.polri.go.id/detail_artikel/kecanduan_judi_online_akibatkan_frustasi,_depresi,_lalu_bunuh_diri.
- Putri, T. E., & Ramadhan, G. (2024). Penerapan Chatbot sebagai Alat Pembelajaran untuk Pengembangan Pendidikan Karakter. Dalam *Indonesian Journal of Computer Science and Engineering (IJCSE)* (Vol. 01).
- Ramsari, N., & Ginanjar, A. (2022). Implementasi Infrastruktur Server Berbasis Cloud Computing Untuk Web Service Berbasis Teknologi Google Cloud Platform. *Conference SENATIK STT Adisutjipto Yogyakarta*, 7. <https://doi.org/10.28989/senatik.v7i0.472>
- Ratniasih, N. L., Sudarma, M., & Gunantara, N. (2017). Penerapan Text Mining Dalam Spam Filtering Untuk Aplikasi Chat. *Majalah Ilmiah Teknologi Elektro*, 16(3), 13. <https://doi.org/10.24843/MITE.2017.v16i03p03>
- Riana, E. (2020). Implementasi Cloud Computing Technology dan Dampaknya Terhadap Kelangsungan Bisnis Perusahaan Dengan Menggunakan Metode Agile dan Studi Literatur. *JURIKOM (Jurnal Riset Komputer)*, 7(3), 439. <https://doi.org/10.30865/jurikom.v7i3.2192>
- Rish, I. (2001). *An empirical study of the naive Bayes classifier*.
- Salloum, S., Alhumaid, K., Salloum, A., & Shaalan, K. (2024). K-means Clustering of Tweet Emotions: A 2D PCA Visualization Approach. *Procedia Computer Science*, 244, 30–36.
- 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>
- Sanderson, D. (2015). *Programming Google App Engine with Python: Build and Run Scalable Python Apps on Google's Infrastructure*. “O'Reilly Media, Inc.”
- Schröer, C., Kruse, F., & Gómez, J. M. (2021). A systematic literature review on applying CRISP-DM process model. *Procedia Computer Science*, 181, 526–534. <https://doi.org/10.1016/j.procs.2021.01.199>
- Shahriar, A. (2024). *Improving Bengali and Hindi Large Language Models*.
- Sinhal, A., & Maheshwari, M. (2022). YouTube: Spam Comments Filtration using Hybrid Ensemble Machine Learning Models. *International Journal of Emerging Technology and Advanced Engineering*, 12(10), 169–182. https://doi.org/10.46338/ijetae1022_18
- Sipayung, F., & Handoyo, C. (2024). *Dampak Dalam Mempromosikan Iklan Judi Online (Studi Kasus Iklan Judi Online Indonesia)*. 1(3). <https://jicnusantara.com/index.php/jicn>
- Srinath, K. R. (2017). Python-The Fastest Growing Programming Language. *International Research Journal of Engineering and Technology*. www.irjet.net
- Sugiyono, D. (2013). *Metode penelitian pendidikan pendekatan kuantitatif, kualitatif dan R&D*.
- Surbakti, N. M., Angelyca, A., Talia, A., Perangin-Angin, C. B., Nainggolan, D. O., Friskauly, N. D., & Tumorang, S. R. B. (2024). Penggunaan Bahasa Pemrograman Python dalam Pembelajaran Kalkulus Fungsi Dua Variabel. *Algoritma: Jurnal Matematika, Ilmu pengetahuan Alam, Kebumihan dan Angkasa*, 2(3), 98–107.
- Syahril, P., Shofa, H., & Fitri, N. (2023). Analisis Sentimen Relokasi Ibukota Nusantara Menggunakan Algoritma Naïve Bayes dan KNN. *Jurnal KomtekInfo*, 1–7. <https://doi.org/10.35134/komtekinfo.v10i1.330>
- Tantithamthavorn, C., McIntosh, S., Hassan, A. E., & Matsumoto, K. (2017). An Empirical Comparison of Model Validation Techniques for Defect Prediction Models. *IEEE Transactions on Software Engineering*, 43(1), 1–18. <https://doi.org/10.1109/TSE.2016.2584050>
- Tiffany, T., Phanie, M. E., Simarmata, A., & Xu, A. (2023). Sentiment Analysis On Twitter Posts About The Russia and Ukraine War With Long Short-Term Memory. *SinkrOn*, 8(2), 789–797. <https://doi.org/10.33395/sinkron.v8i2.12235>

- Wati, R., Ernawati, S., & Rachmi, H. (2023). Pembobotan TF-IDF Menggunakan Naïve Bayes pada Sentimen Masyarakat Mengenai Isu Kenaikan BIPIH. *Jurnal Manajemen Informatika (JAMIKA)*, 13(1), 84–93. <https://doi.org/10.34010/jamika.v13i1.9424>
- Wibawa, A. P., Guntur, M., Purnama, A., Fathony Akbar, M., & Dwiyanto, F. A. (2018). Metode-metode Klasifikasi. *Prosiding Seminar Ilmu Komputer dan Teknologi Informasi*, 3(1).
- Widiyanto, A., Prameswari, M., & Latief, M. A. (2025). Gambling Comments Detection on YouTube: A Comparative Study of Tree-Based Boosting, LSTM and GRU Models. *JUTI: Jurnal Ilmiah Teknologi Informasi*, 144–160.
- Yadav, S., & Shukla, S. (2016). Analysis of k-Fold Cross-Validation over Hold-Out Validation on Colossal Datasets for Quality Classification. *Proceedings - 6th International Advanced Computing Conference, IACC 2016*, 78–83. <https://doi.org/10.1109/IACC.2016.25>
- Yudiana, Y., Agustina, A. Y., & Khofifah, N. (2023). Prediksi Customer Churn Menggunakan Metode CRISP-DM Pada Industri Telekomunikasi Sebagai Implementasi Mempertahankan Pelanggan. *Indonesian Journal of Islamic Economics and Business*, 8(1), 1–20.
- Zeng, G. (2020). On the confusion matrix in credit scoring and its analytical properties. *Communications in Statistics - Theory and Methods*, 49(9), 2080–2093. <https://doi.org/10.1080/03610926.2019.1568485>