

**IMPLEMENTASI ALGORITMA SUPPORT VECTOR MACHINE UNTUK
MENGKLASIFIKASIKAN PENYEBARAN INFORMASI JUDI DARING
DALAM PLATFORM MEDIA SOSIAL**

SKRIPSI

Diajukan untuk memenuhi sebagian dari syarat untuk memperoleh gelar Sarjana
Komputer pada program studi Rekayasa Perangkat Lunak



oleh
Faaris Muda Dwi Nugraha
2000427

**PROGRAM STUDI REKAYASA PERANGKAT LUNAK
UNIVERSITAS PENDIDIKAN INDONESIA
KAMPUS DAERAH CIBIRU
2024**

LEMBAR HAK CIPTA

**IMPLEMENTASI ALGORITMA SUPPORT VECTOR MACHINE UNTUK
MENGKLASIFIKASIKAN PENYEBARAN INFORMASI JUDI DARING
DALAM PLATFORM MEDIA SOSIAL**

Oleh

Faaris Muda Dwi Nugraha

Sebuah skripsi yang diajukan untuk memenuhi salah satu syarat memperoleh gelar
Sarjana Komputer pada program studi Rekayasa Perangkat Lunak

©Faaris Muda Dwi Nugraha 2024

Universitas Pendidikan Indonesia

Agustus 2024

Hak cipta dilindungi undang-undang.

Skripsi ini tidak boleh diperbanyak seluruhnya atau sebagian,
dengan dicetak ulang, difotokopi, atau cara lainnya tanpa izin dari penulis.

HALAMAN PENGESAHAN

FAARIS MUDA DWI NUGRAHA

IMPLEMENTASI ALGORITMA SUPPORT VECTOR MACHINE UNTUK
MENGKLASIFIKASIKAN PENYEBARAN INFORMASI JUDI DARING
DALAM PLATFORM MEDIA SOSIAL

disetujui dan disahkan oleh pembimbing:

Pembimbing I



Hendriyana, S.T., M.Kom.

NIP. 920190219870504101

Pembimbing II



Dian Anggraini, S.ST., M.T.

NIP. 920190219930526201

Mengetahui

Ketua Program Studi Rekayasa Perangkat Lunak



Mochamad Iqbal Ardimansyah, S.T., M.Kom.

NIP. 920190219910328101

LEMBAR PERNYATAAN

Dengan ini Saya menyatakan bahwa skripsi dengan judul "Implementasi Algoritma Support Vector Machine untuk Mengklasifikasikan Penyebaran Informasi Judi Daring dalam Platform Media Sosial" ini beserta seluruh isinya adalah benar-benar karya Saya sendiri. Saya tidak melakukan penjiplakan atau pengutipan dengan cara-cara yang tidak sesuai dengan etika ilmu yang berlaku dalam masyarakat keilmuan. Atas pernyataan ini, Saya siap menanggung risiko/sanksi apabila di kemudian hari ditemukan adanya pelanggaran etika keilmuan atau ada klaim dari pihak lain terhadap keaslian karya Saya ini.

Bandung, 13 Agustus 2024

Yang membuat pernyataan,



Faaris Muda Dwi Nugraha

NIM. 2000427

UCAPAN TERIMA KASIH

Alhamdulillah rabbi 'alamin, segala puji dan syukur peneliti panjatkan kepada Allah Subhanahu wa Ta'ala, yang selalu menjadi tempat peneliti untuk mencurahkan segala keluh, kesah, dan syukur selama proses penyusunan skripsi ini. Berkat rahmat dan karunia-Nya, peneliti dapat menyelesaikan penulisan dan penyusunan skripsi ini.

Peneliti menyadari bahwa skripsi ini tidak akan selesai tanpa bimbingan, saran, dan dorongan yang sangat berarti. Oleh karena itu, peneliti ingin mengucapkan terima kasih dan penghargaan yang sebesar-besarnya kepada:

1. Kedua orang tua tercinta yang selalu memberikan semangat, doa, dan kerja kerasnya yang selalu mendampingi setiap langkah peneliti dalam meraih cita-cita hingga peneliti dapat mencapai tahap ini;
2. Bapak Mochamad Iqbal Ardimansyah, S.T., M.Kom. selaku Ketua Program Studi Rekayasa Perangkat Lunak atas arahan dan dukungan yang telah diberikan selama masa studi peneliti;
3. Bapak Raditya Muhammad, S.T., M.T., selaku Dosen Wali peneliti, atas bimbingan, nasihat, dan dukungan yang telah diberikan selama masa studi peneliti, serta bantuannya dalam mengatasi berbagai tantangan akademik dan pribadi;
4. Bapak Hendriyana, S.T., M.Kom. sebagai Dosen Pembimbing 1 yang selalu memberikan nasihat, waktu, dan dukungan selama proses penyusunan skripsi ini;
5. Ibu Dian Anggraini, S.ST., M.T. sebagai Dosen Pembimbing 2 yang selalu memberikan dukungan dan waktu untuk memberikan bimbingan serta arahan sehingga skripsi ini dapat terselesaikan;
6. Seluruh dosen dan staf Program Studi Rekayasa Perangkat Lunak yang telah memberikan bantuan, dukungan, bimbingan, serta ilmu dan pengalaman selama peneliti menempuh studi;
7. Seluruh jajaran fungsionalitas di Universitas Pendidikan Indonesia, khususnya kampus Cibiru, yang telah membantu peneliti selama menempuh studi;

8. Seluruh teman-teman yang telah kebersamai peneliti selama menjadi mahasiswa UPI Kampus Cibiru, serta seluruh pihak yang terlibat dalam penelitian skripsi ini yang tidak dapat disebutkan satu per satu.

Akhir kata, peneliti berdoa agar segala kebaikan yang telah diberikan menjadi amal dan ibadah serta mendapat rida dari Allah SWT. Selanjutnya, semoga penelitian ini dapat memberikan manfaat bagi pembaca, khususnya bagi civitas akademika Universitas Pendidikan Indonesia.

Bandung, 13 Agustus 2024

Peneliti

Faaris Muda Dwi Nugraha

NIM. 2000427

IMPLEMENTASI ALGORITMA SUPPORT VECTOR MACHINE UNTUK MENGKLASIFIKASIKAN PENYEBARAN INFORMASI JUDI DARING DALAM PLATFORM MEDIA SOSIAL

Faaris Muda Dwi Nugraha

NIM. 2000427

ABSTRAK

Di era digital saat ini, media sosial menjadi platform utama untuk penyebaran berbagai jenis informasi, termasuk informasi judi daring yang merugikan. Penyebaran informasi perjudian daring melalui media sosial telah menjadi masalah serius yang memerlukan penanganan efektif karena dapat berdampak negatif pada moral dan mental masyarakat. Menanggapi tantangan ini, penelitian ini mengimplementasikan algoritma Support Vector Machine (SVM) untuk mengklasifikasikan konten terkait judi daring di platform media sosial. Tujuannya adalah mengembangkan sistem klasifikasi yang akurat serta menyajikan hasil dalam bentuk yang mudah diinterpretasikan guna mendukung pengambilan keputusan oleh pihak berwenang. Urgensi penelitian ini terletak pada kebutuhan akan alat yang efektif untuk mendeteksi dan mencegah penyebaran informasi judi daring yang semakin meresahkan. Untuk mencapai tujuan tersebut, algoritma SVM dengan *kernel* RBF diaplikasikan pada *dataset* yang terdiri atas 42.623 unggahan media sosial, dengan komposisi 20,3% konten positif (mengandung informasi judi daring) dan 79,7% negatif. Hasil evaluasi model menunjukkan performa klasifikasi dengan tingkat akurasi mencapai 96,91%. Secara rinci, model memiliki *precision* dan *recall* yang tinggi untuk kelas negatif, masing-masing sebesar 97% dan 99%, serta *F1-score* sebesar 98%, sementara untuk kelas positif, *precision* model adalah 98%, *recall* sebesar 87%, dan *F1-score* sebesar 92%. Temuan ini mengindikasikan bahwa model SVM yang dikembangkan sangat efektif dalam mengidentifikasi dan membedakan konten terkait judi daring dari konten normal. Implementasi hasil penelitian diwujudkan dalam bentuk situs web interaktif yang mampu mengumpulkan, mengklasifikasikan, dan menampilkan hasil klasifikasi beserta visualisasi data, termasuk distribusi label, frekuensi kata, dan *word cloud*.

Kata Kunci: Media Sosial, Judi Daring, Klasifikasi, Support Vector Machine, Deteksi Konten

**IMPLEMENTATION OF SUPPORT VECTOR MACHINE ALGORITHM TO
CLASSIFY THE SPREAD OF ONLINE GAMBLING INFORMATION
WITHIN SOCIAL MEDIA PLATFORMS**

Faaris Muda Dwi Nugraha

NIM. 2000427

ABSTRACT

In the current digital era, social media has become the primary platform for disseminating various types of information, including harmful online gambling information. The spread of online gambling information via social media has become a serious issue that requires effective handling because it can negatively impact the community's morals and mental health. In response to this challenge, this study implements a Support Vector Machine (SVM) algorithm to classify content related to online gambling on social media platforms. The objective is to develop an accurate classification system and present the results in an easily interpretable form to support decision-making by authorities. The urgency of this research lies in the need for an effective tool to detect and prevent the increasingly troubling spread of online gambling information. To achieve this goal, the SVM algorithm with an RBF kernel was applied to a dataset consisting of 42,623 social media posts, with 20.3% positive content (containing online gambling information) and 79.7% negative content. The model evaluation results showed a classification performance with an accuracy rate of 96.91%. Specifically, the model has high precision and recall for the negative class, each at 97% and 99%, respectively, with an F1-score of 98%, while for the positive class, the model's precision is 98%, recall is 87%, and F1-score is 92%. These findings indicate that the developed SVM model is highly effective in identifying and distinguishing online gambling-related content from normal content. The implementation of the research results is realized in the form of an interactive website capable of collecting, classifying, and displaying classification results along with data visualizations, including label distribution, word frequency, and word clouds.

Keywords: *Social Media, Online Gambling, Classification, Support Vector Machine, Content Detection*

DAFTAR ISI

LEMBAR HAK CIPTA	i
HALAMAN PENGESAHAN.....	ii
LEMBAR PERNYATAAN	iii
UCAPAN TERIMA KASIH	iv
ABSTRAK	vi
ABSTRACT.....	vii
DAFTAR ISI.....	viii
DAFTAR TABEL.....	xii
DAFTAR GAMBAR	xiii
DAFTAR LAMPIRAN	xv
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	3
1.3 Tujuan Penelitian.....	4
1.4 Manfaat Penelitian.....	4
1.5 Batasan Masalah.....	5
1.6 Sistematika Penulisan.....	5
BAB II KAJIAN PUSTAKA	7
2.1 Penelitian Terdahulu	7
2.2 Judi Daring	10
2.3 Media Sosial	13
2.4 <i>Text Mining</i>	14
2.5 Klasifikasi.....	18
2.6 <i>Web Crawling</i>	19
2.7 <i>Machine Learning</i>	19

2.7.1 Supervised Learning	21
2.7.2 Unsupervised Learning	21
2.8 Support Vector Machine	22
2.9 TF-IDF	24
2.10 <i>Confusion Matrix</i>	25
2.11 K-Fold Cross Validation	27
2.12 Google Colab	28
2.13 Python	29
2.13.1 Scikit-Learn	29
2.13.2 Pandas	30
2.13.3 Matplotlib	30
2.13.4 Natural Language Toolkit	31
2.13.5 Tweet Harvest	31
2.13.6 NumPy	31
2.13.7 Selenium	31
2.13.8 Beautiful Soup	32
2.13.9 Streamlit	32
2.13.10 Wordcloud	32
2.13.11 PySastrawi	32
BAB III METODE PENELITIAN	34
3.1 Desain Penelitian	34
3.1.1 Klarifikasi Penelitian	35
3.1.2 Studi Deskriptif I	36
3.1.3 Studi Preskriptif	36
3.1.4 Studi Deskriptif II	52
3.2 Populasi dan Sampel	55

3.3 Instrumen Penelitian.....	55
3.4 Analisis Data.....	58
BAB IV HASIL DAN PEMBAHASAN.....	60
4.1 Pengumpulan Data	60
4.2 Data <i>Pre-processing</i>	62
4.2.1 Case Folding	62
4.2.2 Cleaning	63
4.2.3 Remove Duplicate.....	65
4.2.4 Tokenization.....	65
4.2.5 Slang Normalization	66
4.2.6 Filtering.....	66
4.2.7 Stemming.....	67
4.3 Validasi <i>Dataset</i>	67
4.4 Data <i>Labeling</i>	69
4.5 <i>Modelling</i>	71
4.6 Evaluasi Model.....	72
4.7 Visualisasi dan Interpretasi Hasil	76
4.7.1 Pie Chart Distribusi Label	76
4.7.2 Word Cloud Unggahan	77
4.7.3 Bar Chart Frekuensi Kata	79
4.7.4 Histogram.....	81
4.8 Situs Web Online Gambling Promotion Detector	82
4.8.1 Pengujian Situs Web	90
BAB V SIMPULAN, IMPLIKASI, DAN REKOMENDASI	92
5.1 Simpulan.....	92
5.2 Implikasi dan Rekomendasi	93

DAFTAR PUSTAKA	95
LAMPIRAN.....	102

DAFTAR TABEL

Tabel 2.1 Perbandingan Penelitian Terdahulu	8
Tabel 2.2 Fungsi <i>Kernel</i> SVM	24
Tabel 2.3 <i>Confusion Matrix</i>	26
Tabel 3.1 Tahapan <i>Cleaning</i>	38
Tabel 3.2 Kriteria Label	45
Tabel 3.3 Contoh Dokumen untuk Ekstraksi Fitur.....	46
Tabel 3.4 <i>Term Frequency</i>	47
Tabel 3.5 <i>Inverse Document Frequency</i>	48
Tabel 3.6 Penghitungan TF-IDF.....	49
Tabel 3.7 Daftar <i>Library</i> Python yang Digunakan.....	56
Tabel 3.8 Parameter Pencarian di X.....	57
Tabel 3.9 Atribut Unggahan di X	57
Tabel 4.1 Hasil <i>Case Folding</i>	62
Tabel 4.2 Hasil <i>Cleaning</i>	63
Tabel 4.3 Hasil <i>Remove Duplicate</i>	65
Tabel 4.4 Hasil <i>Tokenization</i>	66
Tabel 4.5 Hasil <i>Slang Normalization</i>	66
Tabel 4.6 Hasil <i>Filtering</i>	67
Tabel 4.7 Hasil <i>Stemming</i>	67
Tabel 4.8 Hasil Data <i>Labeling</i>	70
Tabel 4.9 Informasi Model yang Disimpan.....	76
Tabel 4.10 Hasil Pengujian Situs Web	90

DAFTAR GAMBAR

Gambar 2.1 Support Vector Machine.....	22
Gambar 2.2 K-Fold Cross Validation.....	27
Gambar 3.1 Desain Penelitian.....	35
Gambar 3.2 Diagram Aktivitas Pengumpulan Data.....	37
Gambar 3.3 Diagram Aktivitas <i>Case Folding</i>	38
Gambar 3.4 Diagram Aktivitas <i>Cleaning</i>	40
Gambar 3.5 Diagram Aktivitas <i>Remove Duplicate</i>	41
Gambar 3.6 Diagram Aktivitas <i>Tokenization</i>	41
Gambar 3.7 Diagram Aktivitas <i>Slang Normalization</i>	42
Gambar 3.8 Diagram Aktivitas <i>Filtering</i>	43
Gambar 3.9 Diagram Aktivitas <i>Stemming</i>	44
Gambar 3.10 Tahapan Validasi <i>Dataset</i>	44
Gambar 3.11 Diagram Aktivitas Data <i>Labeling</i>	46
Gambar 3.12 Diagram Aktivitas <i>Modelling</i>	51
Gambar 3.13 Rancangan Situs Web.....	53
Gambar 3.14 <i>Flowchart</i> Online Gambling Promotion Detector.....	54
Gambar 4.1 Fitur Pencarian di X.....	60
Gambar 4.2 Unggahan yang Dikumpulkan per Tahun.....	61
Gambar 4.3 Hasil Pengumpulan Data.....	61
Gambar 4.4 Hasil Penghapusan Atribut.....	62
Gambar 4.5 Hasil Pemeriksaan Parameter Data.....	68
Gambar 4.6 Hasil Pemeriksaan Kebersihan Data.....	68
Gambar 4.7 Hasil Pemeriksaan Redundansi Data.....	69
Gambar 4.8 Hasil Pemeriksaan <i>Missing Value</i>	69
Gambar 4.9 <i>Line Chart</i> Skor Akurasi K-Fold Cross Validation.....	71
Gambar 4.10 <i>Classification Report</i>	73
Gambar 4.11 Hasil <i>Confusion Matrix</i>	75
Gambar 4.12 <i>Pie Chart</i> Distribusi Label.....	77
Gambar 4.13 <i>Word Cloud</i> Unggahan.....	78
Gambar 4.14 Frekuensi Kata Terbanyak Label Positif.....	79

Gambar 4.15 Frekuensi <i>Bigram</i> Terbanyak Label Positif.....	80
Gambar 4.16 Frekuensi Kata Terbanyak Label Negatif.....	80
Gambar 4.17 Frekuensi <i>Bigram</i> Terbanyak Label Negatif	81
Gambar 4.18 Histogram Distribusi Panjang Unggahan Berdasarkan Label.....	81
Gambar 4.19 Halaman Utama Situs Web.....	83
Gambar 4.20 Visualisasi Distribusi Label pada Situs Web	83
Gambar 4.21 Visualisasi Frekuensi Kata pada Situs Web.....	84
Gambar 4.22 Visualisasi Frekuensi <i>Bigram</i> pada Situs Web	85
Gambar 4.23 Visualisasi Distribusi Panjang Karakter pada Situs Web	85
Gambar 4.24 Visualisasi <i>Word Cloud</i> pada Situs Web.....	86
Gambar 4.25 Memilih Jenis <i>Explorer</i> pada Situs Web	87
Gambar 4.26 Mengisi <i>Field Explorer</i> pada Situs Web	88
Gambar 4.27 Rekomendasi Kata Kunci dan <i>Bigram</i> pada Situs Web	88
Gambar 4.28 Hasil <i>Crawling</i> dan Klasifikasi pada Situs Web	89
Gambar 4.29 Melihat Data dengan Lengkap pada Situs Web.....	89
Gambar 4.30 Halaman <i>Archive</i> pada Situs Web	90
Gambar 4.31 Bagian <i>About</i> pada Situs Web	90

DAFTAR LAMPIRAN

Lampiran 1 <i>Dataset</i> yang Digunakan dalam Penelitian.....	102
Lampiran 2 Hasil <i>Case Folding</i> pada <i>Dataset</i>	103
Lampiran 3 Hasil <i>Cleaning</i> pada <i>Dataset</i>	104
Lampiran 4 Hasil <i>Tokenization</i> pada <i>Dataset</i>	105
Lampiran 5 Hasil <i>Slang Normalization</i> pada <i>Dataset</i>	106
Lampiran 6 Hasil <i>Filtering</i> pada <i>Dataset</i>	107
Lampiran 7 Hasil <i>Stemming</i> pada <i>Dataset</i>	108
Lampiran 8 Hasil Data <i>Labeling</i> pada <i>Dataset</i>	109
Lampiran 9 <i>Stopwords</i> dalam Bahasa Indonesia.....	110
Lampiran 10 Kamus Kata Baku dalam Bahasa Indonesia	111
Lampiran 11 Penghitungan Metrik Penelitian	112
Lampiran 12 <i>Pseudocode</i> Pembuatan Model.....	113

DAFTAR PUSTAKA

- Ahmad, R. F. (2023). Criminal Acts for Event Organizers: Online Game Tournament Organizers Who Receive Sponsorship from Online Gambling Sites Based on Indonesian Criminal Law. *Trunojoyo Law Review*, 5(2), 161–174. <https://doi.org/10.21107/tlr.v5i2.20942>
- Aliyah Salsabila, N., Ardhito Winatmoko, Y., Akbar Septiandri, A., & Jamal, A. (2018). Colloquial Indonesian Lexicon. *2018 International Conference on Asian Language Processing (IALP)*, 226–229. <https://doi.org/10.1109/IALP.2018.8629151>
- Alkadri, A. M., Elkorany, A., & Ahmed, C. (2022). Enhancing Detection of Arabic Social Spam Using Data Augmentation and Machine Learning. *Applied Sciences*, 12(22), 11388. <https://doi.org/10.3390/app122211388>
- Amrollahi, A. (2021). A Conceptual Tool to Eliminate Filter Bubbles in Social Networks. *Australasian Journal of Information Systems*, 25. <https://doi.org/10.3127/ajis.v25i0.2867>
- Bhagya S, & Thomas, M. (2023). Social Media App. *International Journal of Advanced Research in Science, Communication and Technology*, 246–249. <https://doi.org/10.48175/IJARST-12945>
- Blessing, L. T. M., Chakrabarti, A., & Wallace, K. M. (1998). An Overview of Descriptive Studies in Relation to a General Design Research Methodology. Dalam *Designers* (hlm. 42–56). Springer London. https://doi.org/10.1007/978-1-4471-1268-6_4
- Brady, W. J., Gantman, A. P., & Van Bavel, J. J. (2020). Attentional Capture Helps Explain Why Moral and Emotional Content Go Viral. *Journal of Experimental Psychology: General*, 149(4), 746–756. <https://doi.org/10.1037/xge0000673>
- Budiman, R., Romadini, N. A., Herwandi Aziz, M. A., & Pratama, A. G. (2022). The Impact of Online Gambling Among Indonesian Teens and Technology.

- IAIC Transactions on Sustainable Digital Innovation (ITSDI)*, 3(2), 162–167.
<https://doi.org/10.34306/itsdi.v3i2.559>
- Butwall, M., Ranka, P., & Shah, S. (2019). Python in Field of Data Science: A Review. *International Journal of Computer Applications*, 178(49), 20–24.
<https://doi.org/10.5120/ijca2019919404>
- Che, X., Chen, D., & Mi, J. (2021). Feature Distribution-Based Label Correlation in Multi-Label Classification. *International Journal of Machine Learning and Cybernetics*, 12(6), 1705–1719. <https://doi.org/10.1007/s13042-020-01268-3>
- Czajkowski, M., & Kretowski, M. (2019). Decision Tree Underfitting in Mining of Gene Expression Data: An Evolutionary Multi-Test Tree Approach. *Expert Systems with Applications*, 137, 392–404.
<https://doi.org/10.1016/j.eswa.2019.07.019>
- Deebani, W., & Kachouie, N. N. (2022). Monte Carlo Ensemble Correlation Coefficient for Association Detection. *Communications in Statistics - Simulation and Computation*, 51(12), 7095–7109.
<https://doi.org/10.1080/03610918.2020.1823413>
- Duarte, V., Zuniga-Jara, S., & Contreras, S. (2022). Machine Learning and Marketing: A Systematic Literature Review. *IEEE Access*, 10, 93273–93288.
<https://doi.org/10.1109/ACCESS.2022.3202896>
- Esteva, A., Robicquet, A., Ramsundar, B., Kuleshov, V., DePristo, M., Chou, K., Cui, C., Corrado, G., Thrun, S., & Dean, J. (2019). A Guide to Deep Learning in Healthcare. *Nature Medicine*, 25(1), 24–29. <https://doi.org/10.1038/s41591-018-0316-z>
- Fahmy Amin, M. (2022). Confusion Matrix in Binary Classification Problems: A Step-by-Step Tutorial. *Journal of Engineering Research*, 6(5).
- Gholamy, A., Kreinovich, V., & Kosheleva, O. (2018). Why 70/30 or 80/20 Relation Between Training and Testing Sets: A Pedagogical Explanation.
https://scholarworks.utep.edu/cs_techrep/1209

- Grigorescu, S., Trasnea, B., Cocias, T., & Macesanu, G. (2020). A Survey of Deep Learning Techniques for Autonomous Driving. *Journal of Field Robotics*, 37(3), 362–386. <https://doi.org/10.1002/rob.21918>
- Guillou-Landreat, M., Gallopel-Morvan, K., Lever, D., Le Goff, D., & Le Reste, J.-Y. (2021). Gambling Marketing Strategies and the Internet: What Do We Know? A Systematic Review. *Frontiers in Psychiatry*, 12. <https://doi.org/10.3389/fpsy.2021.583817>
- Hannan, N., & Dagher, A. (2019). How Is Dentistry Explored Through Social Media? *BDJ Student*, 26(2), 30–31. <https://doi.org/10.1038/s41406-019-0012-6>
- Hassani, H., Beneki, C., Unger, S., Mazinani, M. T., & Yeganegi, M. R. (2020). Text Mining in Big Data Analytics. *Big Data and Cognitive Computing*, 4(1), 1. <https://doi.org/10.3390/bdcc4010001>
- Hussain, S. F. (2019). A Novel Robust Kernel for Classifying High-Dimensional Data Using Support Vector Machines. *Expert Systems with Applications*, 131, 116–131. <https://doi.org/10.1016/j.eswa.2019.04.037>
- Janiesch, C., Zschech, P., & Heinrich, K. (2021). Machine Learning and Deep Learning. *Electronic Markets*, 31(3), 685–695. <https://doi.org/10.1007/s12525-021-00475-2>
- Joseph, V. R. (2022). Optimal Ratio for Data Splitting. *Statistical Analysis and Data Mining: The ASA Data Science Journal*, 15(4), 531–538. <https://doi.org/10.1002/sam.11583>
- Kadhim, A. I. (2019). Survey on Supervised Machine Learning Techniques for Automatic Text Classification. *Artificial Intelligence Review*, 52(1), 273–292. <https://doi.org/10.1007/s10462-018-09677-1>
- Karo Karo, I. M., & Hendriyana, H. (2022). Klasifikasi Penderita Diabetes Menggunakan Algoritma Machine Learning dan Z-Score. *Jurnal Teknologi Terpadu*, 8(2), 94–99. <https://doi.org/10.54914/jtt.v8i2.564>

- Keerthi, M. P., Reddy, G. S., Raghava, V. S., & Reddy, K. B. (2023). Streamlit Interface for Multiple Disease Diagnosis. *International Journal for Research in Applied Science and Engineering Technology*, 11(2), 1159–1164. <https://doi.org/10.22214/ijraset.2023.49166>
- Khder, M. (2021). Web Scraping or Web Crawling: State of Art, Techniques, Approaches and Application. *International Journal of Advances in Soft Computing and its Applications*, 13(3), 145–168. <https://doi.org/10.15849/IJASCA.211128.11>
- Killick, E. A., & Griffiths, M. D. (2020). A Content Analysis of Gambling Operators Twitter Accounts at the Start of the English Premier League Football Season. *Journal of Gambling Studies*, 36(1), 319–341. <https://doi.org/10.1007/s10899-019-09879-4>
- Kuasa, D. A., & Jaya, F. (2022). Online Gambling Phenomenon: Law & Society. *Widya Yuridika*, 5(2), 345. <https://doi.org/10.31328/wy.v5i2.3572>
- Ling, S. (2019). Social Media: A Growing Presence in Healthcare. *Health Science Inquiry*, 10(1). <https://doi.org/10.29173/hsi20>
- Luo, F.-F., Guo, W., & Chen, G.-L. (2019). Addressing Imbalance in Weakly Supervised Multi-Label Learning. *IEEE Access*, 7, 37463–37472. <https://doi.org/10.1109/ACCESS.2019.2906409>
- Maidianti, S., & Nofianti, L. (2023). Kebijakan Penegakan Hukum Pidana dalam Rangka Penanggulangan Perjudian. *Journal of Social and Economics Research*, 5(2), 480–490. <https://doi.org/10.54783/jser.v5i2.156>
- Marcot, B. G., & Hanea, A. M. (2021). What Is an Optimal Value of K in K-Fold Cross-Validation in Discrete Bayesian Network Analysis? *Computational Statistics*, 36(3), 2009–2031. <https://doi.org/10.1007/s00180-020-00999-9>
- Nandwani, P., & Verma, R. (2021). A Review on Sentiment Analysis and Emotion Detection from Text. *Social Network Analysis and Mining*, 11(1), 81. <https://doi.org/10.1007/s13278-021-00776-6>

- Naseem, U., Razzak, I., & Eklund, P. W. (2021). A Survey of Pre-processing Techniques to Improve Short-Text Quality: A Case Study on Hate Speech Detection on Twitter. *Multimedia Tools and Applications*, 80(28–29), 35239–35266. <https://doi.org/10.1007/s11042-020-10082-6>
- Nugraha, M. R. A., & Setiawan, D. A. (2023). Penegakan Hukum Tindak Pidana Perjudian Online di Kabupaten Tasikmalaya Dihubungkan dengan Undang-Undang Nomor 19 Tahun 2016 Tentang Perubahan Atas Undang-Undang Nomor 11 Tahun 2008 Tentang Informasi dan Transaksi Elektronik. *Bandung Conference Series: Law Studies*, 3(1). <https://doi.org/10.29313/bcsls.v3i1.4916>
- Nur Fitriana, D., & Sibaroni, Y. (2020). Sentiment Analysis on KAI Twitter Post Using Multiclass Support Vector Machine (SVM). *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, 4(5), 846–853. <https://doi.org/10.29207/resti.v4i5.2231>
- Paramartha, P. P. R., Dewi, A. A. S. L., & Seputra, I. P. G. (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>
- Ponomarev, I. (2022). Development of an Automated System for Clustering Text Documents. *System Technologies*, 1(138), 115–119. <https://doi.org/10.34185/1562-9945-1-138-2022-10>
- Qaiser, S., & Ali, R. (2018). Text Mining: Use of TF-IDF to Examine the Relevance of Words to Documents. *International Journal of Computer Applications*, 181(1), 25–29. <https://doi.org/10.5120/ijca2018917395>
- Rahardja, U. (2022). Social Media Analysis as a Marketing Strategy in Online Marketing Business. *Startupreneur Business Digital (SABDA Journal)*, 1(2), 176–182. <https://doi.org/10.34306/sabda.v1i2.120>
- Raja, Dr. N. K., Bakala, Mr. N., & Suresh, Mr. S. (2019). NLP: Rule Based Name Entity Recognition. *International Journal of Innovative Technology and*

- Exploring Engineering*, 8(11), 4285–4290.
<https://doi.org/10.35940/ijitee.K2047.0981119>
- Ramli, S., Ahmad Nasir, A. N. H., Wook, M., Mat Razali, N. A., & Mohd Zainuddin, N. (2021). Classifying Fake Profile in Facebook Account Using Support Vector Machine. *Journal of Defence Science, Engineering & Technology*, 4(2).
<https://doi.org/10.58247/jdset-2021-0402-12>
- Rocks, J. W., & Mehta, P. (2022). Bias-Variance Decomposition of Overparameterized Regression with Random Linear Features. *Physical Review E*, 106(2), 025304. <https://doi.org/10.1103/PhysRevE.106.025304>
- Saigal, P., & Khanna, V. (2020). Multi-Category News Classification Using Support Vector Machine-Based Classifiers. *SN Applied Sciences*, 2(3), 458.
<https://doi.org/10.1007/s42452-020-2266-6>
- Sari, S. I. N., & Hidayah, T. N. (2023). Identifikasi Yuridis Penegakan Hukum Terhadap Perjudian Online di Indonesia. *Qaumiyyah: Jurnal Hukum Tata Negara*, 4(1), 18–36. <https://doi.org/10.24239/qaumiyyah.v4i1.58>
- Seliya, N., Abdollah Zadeh, A., & Khoshgoftaar, T. M. (2021). A Literature Review on One-Class Classification and Its Potential Applications in Big Data. *Journal of Big Data*, 8(1), 122. <https://doi.org/10.1186/s40537-021-00514-x>
- Shafira, A. (2023). Hoax COVID-19 News Detection Based on Sentiment Analysis in Indonesian using Support Vector Machine (SVM) Method. *International Journal on Information and Communication Technology (IJoICT)*, 8(2), 66–77. <https://doi.org/10.21108/ijoiect.v8i2.682>
- Shaveta. (2023). A Review on Machine Learning. *International Journal of Science and Research Archive*, 9(1), 281–285.
<https://doi.org/10.30574/ijjsra.2023.9.1.0410>
- Styawati, S., Hendrastuty, N., & Isnain, A. R. (2021). Analisis Sentimen Masyarakat Terhadap Program Kartu Prakerja Pada Twitter Dengan Metode Support Vector Machine. *Jurnal Informatika: Jurnal Pengembangan IT*, 6(3), 150–155. <https://doi.org/10.30591/jpfit.v6i3.2870>

- Takwin, B. (2020). Pesan dari Editor-in-Chief: Tantangan Psikologi Siber. *Jurnal Psikologi Sosial*, 18(1), 3–4. <https://doi.org/10.7454/jps.2020.02>
- Tsai, C.-F., Lin, W.-C., Hu, Y.-H., & Yao, G.-T. (2019). Under-Sampling Class Imbalanced Datasets by Combining Clustering Analysis and Instance Selection. *Information Sciences*, 477, 47–54. <https://doi.org/10.1016/j.ins.2018.10.029>
- Umeugo, W. (2023). Cybercrime Awareness on Social Media: A Comparison Study. *International Journal of Network Security & Its Applications*, 15(2), 23–35. <https://doi.org/10.5121/ijnsa.2023.15202>
- van Engelen, J. E., & Hoos, H. H. (2020). A Survey on Semi-Supervised Learning. *Machine Learning*, 109(2), 373–440. <https://doi.org/10.1007/s10994-019-05855-6>
- Wong, T.-T., & Yeh, P.-Y. (2020). Reliable Accuracy Estimates from K-Fold Cross Validation. *IEEE Transactions on Knowledge and Data Engineering*, 32(8), 1586–1594. <https://doi.org/10.1109/TKDE.2019.2912815>