

Analisis Klasifikasi Dan Prediksi Pola Publikasi Berita Pemrov DKI Jakarta Menggunakan Machine Learning

Rifky Permana^{1*}, Febby Ariyanti Herdiana²

^{1,2}Universitas Bina Sarana Informatika
e-mail: ¹rifky.rpp@bsi.ac.id, ²febby.hrdn@gmail.com

Diterima	Direvisi	Disetujui
23-05-2025	27-05-2025	12-06-2025

Abstrak - Dalam era digital, akses informasi melalui berbagai platform online semakin meningkat, termasuk melalui situs web resmi pemerintah. Pemerintah Provinsi DKI Jakarta secara rutin menyampaikan kebijakan, program, dan berita melalui situs webnya. Namun, volume data yang besar dan keragaman topik berita menghadirkan tantangan dalam pengelolaan dan analisis informasi secara efisien. Penelitian ini mengembangkan model klasifikasi dan prediksi pola publikasi berita menggunakan tiga algoritma machine learning yaitu: *Support Vector Machine (SVM)*, *Naïve Bayes*, dan *Random Forest*. Dataset yang digunakan berupa berita yang dipublikasikan oleh Pemrov DKI Jakarta selama tahun 2023. Pra-pemrosesan data meliputi pembersihan teks, *case folding*, *tokenizing*, normalisasi, penghapusan *stopword*, dan *stemming*. Model SVM dan *Naïve Bayes* diterapkan untuk klasifikasi berita, sedangkan *Random Forest* digunakan untuk memprediksi pola publikasi berita harian. Proses *tuning hyperparameter* dilakukan menggunakan *GridSearchCV* guna meningkatkan kinerja model. Penilaian terhadap klasifikasi dilakukan dengan diukur dari beberapa parameter yaitu akurasi, *precision*, *recall*, dan *F1-score*, sedangkan evaluasi prediksi menggunakan MAE, MSE, dan R^2 . Temuan dari penelitian ini mengindikasikan bahwa SVM menunjukkan performa yang baik dengan akurasi 95%, sedangkan *Random Forest* berhasil memprediksi pola publikasi berita dengan R^2 sebesar 0,82. Penelitian ini diharapkan dapat mendukung pengelolaan informasi publik secara lebih efisien dan menyediakan wawasan mengenai pola publikasi berita oleh Pemrov DKI Jakarta.

Kata Kunci: Klasifikasi Berita, Prediksi Pola Publikasi, Machine Learning

Abstract - In the digital era, access to information through various online platforms has significantly increased, including through official government websites. The Provincial Government of DKI Jakarta routinely communicates policies, programs, and news via its website. However, the large volume of data and the diversity of news topics pose challenges for efficiently managing and analyzing information. This study develops a classification model and predicts news publication patterns using three machine learning algorithms: *Support Vector Machine (SVM)*, *Naïve Bayes*, and *Random Forest*. The dataset consists of news published by the DKI Jakarta Provincial Government during 2023. Data preprocessing includes text cleaning, *case folding*, *tokenizing*, *normalization*, *stopword removal*, and *stemming*. SVM and *Naïve Bayes* models are applied for news classification, while *Random Forest* is used to predict daily news publication patterns. Hyperparameter tuning is performed using *GridSearchCV* to enhance model performance. Classification is evaluated based on several parameters, including accuracy, precision, recall, and *F1-score*, while prediction is assessed using MAE, MSE, and R^2 . The findings of this study indicate that SVM demonstrates strong performance with an accuracy of 95%, while *Random Forest* successfully predicts news publication patterns with an R^2 of 0.82. This research is expected to support more efficient management of public information and provide insights into the news publication patterns of the DKI Jakarta Provincial Government.

Keywords: News Classification, Publication Pattern Prediction, Machine Learning

PENDAHULUAN

Berita merupakan salah satu sumber informasi esensial dalam kehidupan sehari-hari, terutama di zaman digital saat ini, akses informasi semakin mudah dijangkau. melalui platform daring (Nanda et al., 2022). Salah satu lembaga pemerintah yang aktif dalam menyebarkan informasi publik adalah

Pemerintah Provinsi DKI. Melalui situs resmi dan kanal digital lainnya, Pemrov DKI Jakarta menginformasikan berbagai kebijakan, program, dan kegiatan. Namun, volume data yang besar dan keragaman topik berita menimbulkan tantangan dalam pengelolaan dan analisis berita secara efisien (Hayami, 2023). Oleh karena itu, diperlukan pendekatan yang lebih efisien dalam pengelompokan

dan analisis berita untuk mendukung proses pengambilan keputusan berbasis data.

Seiring perkembangan teknologi, pendekatan berbasis machine learning menjadi solusi potensial dalam menangani tantangan tersebut, termasuk dalam klasifikasi dan prediksi pola publikasi berita (Kim et al., 2019; Sholih 'afif et al., 2021). Berbagai algoritma seperti SVM, Naïve Bayes, maupun Random Forest telah terbukti efektif untuk mengelompokkan teks dengan karakteristik kompleks (Setiawan et al., 2021). SVM dikenal karena kemampuannya dalam menemukan hyperplane optimal untuk memisahkan data secara efisien, sedangkan Naïve Bayes menawarkan kesederhanaan serta kinerja yang kompetitif bahkan dengan dataset besar (Kartika Sari Dewi & Purnomo Aji, 2021).

Random Forest menonjol dalam menangani data tidak seimbang dan dapat memberikan hasil prediktif yang robust dengan menggabungkan banyak pohon keputusan (Muhammad Habib et al., 2022). (Rama Bena Putra & Setya Perdana, 2023) menemukan bahwa dengan melakukan *hyperparameter tuning*, *Random Forest* tidak hanya meningkatkan akurasi prediksi tetapi juga mampu mengurangi risiko *overfitting* pada data yang memiliki banyak fitur. Keunggulan ini menjadikan *Random Forest* algoritma yang sesuai untuk prediksi pola publikasi berita, khususnya dalam memetakan tren publikasi harian dari Pemprov DKI Jakarta. Fokus utama penelitian ini adalah memanfaatkan *Random Forest* untuk memprediksi publikasi berita secara lebih akurat dan efektif.

Penelitian sebelumnya menunjukkan efektivitas SVM dan *Naïve Bayes* dalam klasifikasi berita. Contohnya, (Setiawan et al., 2021.) melaporkan akurasi 85% dengan *Naïve Bayes* menggunakan fitur N-Gram, dan (Sripamuji et al., 2022.) menemukan bahwa SVM mencapai akurasi hingga 90%. Di sisi lain, penelitian oleh (Ayuni et al., 2023) menyajikan model prediksi yang menggabungkan metode klasifikasi teks dan prediksi temporal, memberikan wawasan mendalam tentang pola publikasi berita. Sementara penelitian ini berfokus pada penerapan *Random Forest* dalam prediksi pola publikasi berita, yang masih jarang dieksplorasi dalam konteks berita lokal.

Dengan demikian, penelitian ini bertujuan untuk membangun model klasifikasi berita dengan menggunakan algoritma SVM, *Naïve Bayes*, dan *Random Forest* serta menitikberatkan pada prediksi pola publikasi dengan *Random Forest* berdasarkan data publikasi harian dari Pemprov DKI Jakarta. Diharapkan bahwa hasil penelitian ini dapat berkontribusi secara signifikan, baik dalam meningkatkan efisiensi pengelolaan berita maupun dalam memberikan wawasan yang lebih mendalam bagi pemerintah daerah untuk mendukung perencanaan dan pengambilan keputusan.

Penelitian ini menawarkan dua kontribusi

utama. Pertama, mengisi kesenjangan dalam studi terkait klasifikasi dan prediksi berita lokal dengan mengimplementasikan metode *machine learning*. Kedua, penelitian ini menyajikan model prediksi pola publikasi harian yang dapat membantu Pemprov DKI Jakarta dalam memahami tren publikasi secara lebih optimal dan terukur.

1. *Machine Learning* dalam Klasifikasi dan Prediksi Berita

Machine learning adalah pendekatan berbasis Algoritma yang memungkinkan mesin untuk memperoleh pengetahuan dari data tanpa memerlukan pengkodean yang jelas (Aditya et al., 2023). Dalam klasifikasi berita, tujuan utama adalah mengidentifikasi kategori dari teks berdasarkan fitur tertentu, seperti kata atau frase dalam judul berita. Selain klasifikasi, *machine learning* juga digunakan untuk prediksi, seperti memetakan pola publikasi dalam rentang waktu tertentu (Setiawan et al., 2021).

2. *Support Vector Machine* (SVM)

Algoritma SVM bekerja dengan mencari hyperplane paling optimal yang dapat memisahkan kelas dengan margin terbesar. Persamaan dasar hyperplane dituliskan sebagai berikut:

$$w \cdot x + b = 0 \dots\dots\dots(1)$$

Keterangan:

w = Vektor bobot

x = Vektor input atau fitur

b = Bias

Untuk memaksimalkan margin antara dua kelas, solusi SVM dapat dirumuskan sebagai:

$$\text{Maximize } \frac{2}{\|w\|} \dots\dots\dots(2)$$

3. *Naïve Bayes*

Naïve Bayes menggunakan Teorema Bayes untuk menghitung probabilitas suatu kategori berdasarkan fitur yang diamati. Persamaan dasarnya adalah:

$$P(C|X) = \frac{P(X|C).P(C)}{P(X)} \dots\dots\dots(3)$$

Keterangan:

- $P(C|X)$ = Probabilitas kategori C untuk fitur X
- $P(X|C)$ = Probabilitas fitur X muncul dalam kategori C
- $P(C)$ = Probabilitas awal kategori C
- $P(X)$ = Probabilitas dari fitur X

4. *Random Forest*

Random Forest menggabungkan beberapa pohon keputusan untuk membuat prediksi. Hasil akhir diperoleh dengan rata-rata atau mayoritas prediksi

dari setiap pohon. Persamaan prediksinya adalah:

$$\hat{y} = \frac{1}{T} \sum_{t=1}^T h_t(x) \dots \dots \dots (4)$$

Keterangan:

- T = Jumlah pohon dalam hutan
- $h_t(x)$ = Prediksi pohon ke-t untuk input x
- \hat{y} = Nilai prediksi akhir

5. Metrik Evaluasi Klasifikasi

a. Tingkat Akurasi (*Accuracy*)

Menilai bagian dari prediksi yang akurat dibandingkan dengan seluruh prediksi yang dibuat.

$$Akurasi = \frac{Jumlah\ prediksi\ benar}{Total\ prediksi} \dots \dots \dots (5)$$

b. Precision

Precision mengukur seberapa banyak prediksi positif yang benar.

$$Precision = \frac{True\ Positive\ (TP)}{True\ Positive\ (TP) + False\ Negative\ (FN)} \dots \dots \dots (6)$$

Keterangan:

- TP = True Positive (prediksi benar yang positif)
- FP = False Positive (prediksi salah yang positif)

c. Recall

Recall mengukur seberapa banyak kasus positif yang berhasil diidentifikasi dengan benar.

$$Recall = \frac{True\ Positive\ (TP)}{True\ Positive\ (TP) + False\ Negative\ (FN)} \dots \dots \dots (7)$$

Keterangan:

FN = False Negative (kasus positif yang terlewat)

d. F1-Score

F1-Score merupakan nilai rata-rata harmonis antara precision dan recall. Ini berguna jika ada ketidakseimbangan data (*data imbalanced*).

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \dots \dots \dots (8)$$

e. Confusion Matrix

Menyajikan distribusi prediksi benar dan salah dari setiap kelas. Ini memberikan gambaran lebih rinci terkait performa model untuk tiap kategori.

6. Metrik Evaluasi Prediksi

Metrik ini digunakan untuk mengukur performa algoritma seperti *Random Forest* dalam memprediksi pola publikasi berita.

a. Mean Absolute Error (MAE)

Menghitung rata-rata dari selisih absolut antara

nilai yang sebenarnya dan yang diprediksi.

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \dots \dots \dots (9)$$

Keterangan:

- n = Jumlah data
- y_i = Nilai aktual
- \hat{y}_i = Nilai prediksi

b. Mean Squared Error (MSE)

Menilai rata-rata kuadrat dari perbedaan antara nilai yang sebenarnya dan yang diprediksi.

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \dots \dots \dots (10)$$

c. R² Score (Koefisien Determinasi)

Menilai seberapa efektif model dalam menggambarkan fluktuasi dalam data yang sebenarnya.

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \dots \dots \dots (11)$$

Keterangan:

\bar{y} = Rata-rata nilai aktual

METODE PENELITIAN

Penelitian ini dilakukan secara bertahap untuk memastikan proses klasifikasi dan prediksi pola publikasi berita berjalan dengan sistematis. Setiap langkah dirancang agar data yang digunakan memiliki kualitas tinggi dan model yang dibangun dapat memberikan hasil akurat. Proses penelitian akan dijabarkan sebagai berikut:

1. Pengumpulan Data

Data yang dipakai dalam penelitian ini diambil dari situs resmi Pemprov DKI Jakarta selama tahun 2023. Data tersebut meliputi judul berita, tanggal publikasi, dan kategori berita.

2. Pra-pemrosesan Data

Pra-pemrosesan data dilaksanakan untuk membersihkan data dan menyiapkannya agar siap digunakan dalam proses pelatihan. Menurut (Kartika Sari Dewi & Purnomo Aji, 2021), teknik pra-pemrosesan ini sangat penting dalam klasifikasi teks berbahasa Indonesia, karena dapat meningkatkan akurasi model dengan menghilangkan kata-kata yang tidak relevan. Tahapan yang dilakukan meliputi:

- a. *Pembersihan (Cleaning)*: Menghapus karakter atau simbol yang tidak relevan.
- b. *Case folding*: Mengubah seluruh teks menjadi huruf kecil.
- c. *Tokenisasi*: Memisahkan kalimat menjadi kata-kata individual.
- d. *Penghapusan stopword (Stopword removal)*: Menghilangkan kata-kata umum yang tidak memberikan informasi penting.
- e. *Stemming*: Mengubah kata-kata menjadi bentuk

dasarnya.

3. Pembagian Data

Dataset dibagi secara acak menjadi 80% untuk data pelatihan dan 20% untuk data pengujian untuk memastikan representasi yang seimbang dari setiap kategori berita. Data pelatihan dapat digunakan untuk membangun model, sementara data pengujian dipakai untuk mengevaluasi performa model dalam memprediksi data baru yang belum pernah dilihat sebelumnya.

4. Pelatihan Model

Dalam penelitian ini diterapkan dua algoritma yaitu SVM dan *Naïve Bayes* untuk pengklasifikasian berita. Sementara itu, Random Forest digunakan untuk memprediksi pola publikasi berita berdasarkan tanggal dan kategori publikasi.

5. Pelatihan Model

Untuk meningkatkan kinerja model, proses *tuning hyperparameter* dilakukan dengan *GridSearchCV*. Parameter penting seperti *n_estimators*, *max_depth*, dan *max_features* pada *Random Forest* disesuaikan untuk mengurangi *overfitting*.

6. Evaluasi Model

Evaluasi model klasifikasi menggunakan metrik akurasi, *precision*, *recall*, dan *F1-score* untuk mengukur kinerja SVM dan *Naïve Bayes*. Untuk model prediksi pola publikasi menggunakan Random Forest, evaluasi dilakukan dengan MAE, MSE, dan R². Metrik ini digunakan untuk menilai kesalahan prediksi dan sejauh mana model mampu menjelaskan pola publikasi berita secara akurat.

7. Prediksi Pola Publikasi

Model *Random Forest* digunakan untuk memprediksi pola publikasi berita harian. Prediksi dilakukan dengan membedakan pola pada hari kerja dan akhir pekan agar dapat melihat perbedaan dalam intensitas publikasi.

8. Analisis dan Interpretasi

Hasil klasifikasi dan prediksi dianalisis secara mendalam untuk memahami pola publikasi dan mengidentifikasi potensi tren. Visualisasi grafik disertakan untuk membandingkan hasil prediksi dan nilai aktual.

HASIL DAN PEMBAHASAN

Penelitian ini mengaplikasikan dua algoritma pembelajaran mesin, yakni Support Vector Machine (SVM) dan *Naïve Bayes*, untuk klasifikasi berita Pemprov DKI Jakarta. Selain itu, untuk memprediksi pola publikasi harian, digunakan model Random Forest. Berdasarkan hasil eksperimen, model SVM dan Random Forest

memberikan performa yang unggul dibandingkan dengan model lainnya.

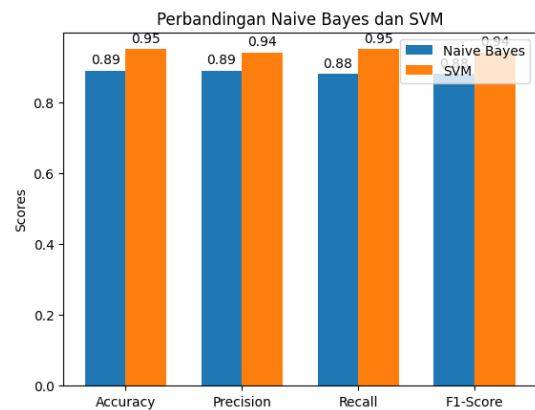
1. Hasil Klasifikasi Berita

Hasil evaluasi model klasifikasi berdasarkan akurasi, *precision*, *recall*, dan *F1-score* ditampilkan pada Tabel 1. Pada eksperimen ini, SVM menunjukkan akurasi yang lebih tinggi sebesar 0,95 dibandingkan dengan *Naïve Bayes* yang mencapai 0,89. Selain itu, *precision* dan *recall* dari SVM masing-masing mencapai 0,94 dan 0,95, menunjukkan bahwa model ini mampu melakukan klasifikasi dengan akurat dan konsisten. Sementara itu, *Naïve Bayes*, meskipun lebih cepat dan sederhana, menghasilkan skor *precision* dan *recall* sebesar 0,89 dan 0,88.

Tabel 1. Hasil evaluasi metrik antara *Naïve Bayes* dan SVM.

Metrik	<i>Naïve Bayes</i>	SVM
Akurasi	0.89	0.95
<i>Precision</i>	0.89	0.94
<i>Recall</i>	0.88	0.95
<i>F1-Score</i>	0.88	0.94

Sumber: Dokumentasi Penelitian (2024)



Sumber: Dokumentasi Penelitian (2024)

Gambar 1. Perbandingan Metrik antara *Naïve Bayes* dan SVM

Beberapa kategori berita menampilkan perbedaan signifikan dalam performa kedua model, seperti pada kategori 3 (Kebakaran) dan kategori 7 (Lintas Kota). Tabel 2 menunjukkan bahwa SVM memberikan *F1-score* lebih tinggi dibandingkan *Naïve Bayes* pada kategori ini, yaitu 0,81 dan 0,78, sementara *Naïve Bayes* hanya mencapai 0,65 dan 0,47.

Tabel 1. Hasil evaluasi metrik antara *Naïve Bayes* dan SVM.

Kategori	<i>Naïve Bayes</i> (F1-Score)	SVM (F1-Score)
Kategori 3 (Kebakaran)	0.65	0.81
Kategori 7 (Lintas Kota)	0.47	0.78

Sumber: Dokumentasi Penelitian (2024)

Dari hasil ini, dapat dilihat bahwa SVM lebih mampu menangani kategori dengan distribusi data tidak merata dan variasi kata yang kompleks. Hal ini menunjukkan keunggulan SVM dalam menangani data teks yang lebih kompleks dibandingkan *Naïve Bayes*.

2. Hasil Prediksi Pola Publikasi Berita

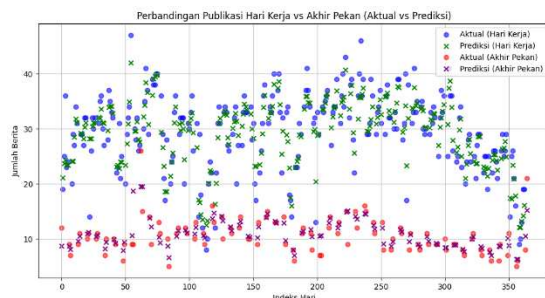
Selain klasifikasi, penelitian ini juga melakukan prediksi jumlah publikasi berita harian menggunakan model Random Forest. Model ini digunakan untuk memprediksi pola publikasi dan memberikan wawasan terkait perbedaan publikasi antara hari kerja dan akhir pekan.

Tabel 3. Performa prediksi Random Forest

Metrik	Nilai
MAE (Mean Absolute Error)	3.50
MSE (Mean Squared Error)	21.29
R ²	0.82

Sumber: Dokumentasi Penelitian (2024)

Berikut menampilkan perbandingan antara publikasi aktual dan prediksi model Random Forest untuk hari kerja dan akhir pekan. Dari grafik tersebut terlihat bahwa publikasi Lebih sering terjadi pada hari kerja jika dibandingkan dengan akhir pekan, dan model mampu memprediksi tren ini dengan cukup baik.



Sumber: Dokumentasi Penelitian (2024)

Gambar 2. Perbandingan Metrik antara *Naïve Bayes* dan SVM

Hasil ini memberikan beberapa wawasan penting. Pertama, pola publikasi yang lebih tinggi pada hari

kerja menunjukkan bahwa aktivitas pemerintahan dan informasi yang disampaikan lebih banyak terjadi selama hari kerja, sementara pada akhir pekan intensitas publikasi menurun. Kedua, model Random Forest memberikan hasil yang cukup baik dalam menangkap pola ini, meskipun performanya masih dapat ditingkatkan dengan penambahan fitur eksternal seperti data cuaca atau peristiwa penting.

KESIMPULAN

Penelitian ini berhasil mencapai tujuan yang telah dinyatakan dalam bagian pendahuluan, yaitu mengembangkan dan membandingkan model klasifikasi berita serta memprediksi pola publikasi berita harian oleh Pemprov DKI Jakarta. Berdasarkan hasil klasifikasi, Support Vector Machine (SVM) terbukti lebih unggul dibandingkan dengan *Naïve Bayes* dalam hal akurasi dan konsistensi prediksi kategori berita, terutama untuk kategori berita yang kompleks dan tidak merata distribusinya. SVM berhasil mencapai akurasi sebesar 0,95, sementara *Naïve Bayes* menghasilkan akurasi 0,89, menunjukkan bahwa SVM lebih efektif untuk tugas klasifikasi teks dalam konteks ini.

Selain klasifikasi, prediksi pola publikasi berita harian menggunakan Random Forest memberikan wawasan yang relevan mengenai tren publikasi berita antara hari kerja dan akhir pekan. Model ini menunjukkan kinerja yang baik dengan nilai MAE sebesar 3,50, MSE sebesar 21,29, dan R² sebesar 0,82, yang menunjukkan bahwa model dapat menangkap pola publikasi dengan cukup baik. Dari hasil prediksi, terlihat bahwa publikasi berita lebih banyak dilakukan pada hari kerja, sementara intensitasnya menurun selama akhir pekan. Hal ini sesuai dengan pola operasional dan aktivitas pemerintahan yang lebih aktif pada hari kerja.

Hasil penelitian ini diharapkan dapat membantu Pemprov DKI Jakarta dalam meningkatkan strategi publikasi dan penyebaran informasi publik secara lebih efektif, dengan mempertimbangkan pola waktu yang tepat. Selain itu, implementasi model Random Forest untuk prediksi dapat menjadi acuan dalam mengantisipasi volume publikasi di masa mendatang, yang berguna untuk perencanaan dan alokasi sumber daya publikasi.

Prospek pengembangan penelitian ini mencakup penambahan fitur eksternal seperti data peristiwa penting atau kondisi cuaca, yang dapat memengaruhi intensitas publikasi. Selain itu, eksplorasi algoritma prediktif lain, seperti LSTM untuk pemodelan deret waktu, dapat meningkatkan akurasi prediksi. Penelitian selanjutnya juga disarankan untuk menggunakan data lebih luas, seperti publikasi di media sosial resmi pemerintah, untuk memperluas

cakupan analisis pola publikasi.

Dengan demikian, penelitian ini telah memberikan kontribusi nyata dalam mendukung pengelolaan informasi dan publikasi berita oleh Pemprov DKI Jakarta. Hasilnya tidak hanya mendukung pengambilan keputusan berbasis data, tetapi juga membuka peluang untuk penelitian lanjutan yang lebih komprehensif dan adaptif terhadap dinamika informasi di era digital.

REFERENSI

- Aditya, A., Wahyuddin, P., Leo, S., Santoso, W., Wahyu, G., Wibowo, N., Khrisna, A., Rahmaddeni, W., Wahidin, A. J., Eka, G., Elisawati, Y., Rizqi, R., & Abdurrasyid, W. (n.d.). *MACHINE LEARNING*. www.globaleksekutifteknologi.co.id
- Ayuni, A. Q., Helen, A., Yulawati, S., Raya, J., Sumedang, B., 21, K. M., Jatinangor, K., Sumedang, K., & Barat, J. (2023). Klasifikasi Topik Berita Deutsche Welle Indonesia dengan Kata Kunci Indonesia Menggunakan Metode Multinomial Naive Bayes. In *JLK* (Vol. 6, Issue 1).
- Hayami, R. (2023). KLASIFIKASI TEKS BERITA BERBAHASA INDONESIA MENGGUNAKAN MACHINE LEARNING DAN DEEP LEARNING: STUDI LITERATUR. In *Jurnal Mahasiswa Teknik Informatika* (Vol. 7, Issue 1). <https://ieeexplore.ieee.org/>
- Kartika Sari Dewi, F., & Purnomo Aji, T. (n.d.). *KLASIFIKASI BERITA MENGGUNAKAN METODE MULTINOMIAL NAÏVE BAYES*.
- Kim, D., Seo, D., Cho, S., & Kang, P. (2019). Multi-co-training for document classification using various document representations: TF-IDF, LDA, and Doc2Vec. *Information Sciences*, 477, 15–29. <https://doi.org/10.1016/j.ins.2018.10.006>
- Muhammad Habib, S., Haerani, E., Kurnia Gusti, S., Ramadhani, S., & Informatika UIN Sultan Syarif Kasim Riau Jl Soebrantas, T. H. (2022). Klasifikasi Berita Menggunakan Metode Naïve Bayes Classifier. *Jurnal Nasional Komputasi Dan Teknologi Informasi*, 5(2).
- Nanda, R., Haerani, E., Gusti, S. K., & Ramadhani, S. (2022). Klasifikasi Berita Menggunakan Metode Support Vector Machine. *Jurnal Nasional Komputasi Dan Teknologi Informasi*, 5(2).
- Rama Bena Putra, P., & Setya Perdana, R. (2023). *Klasifikasi Judul Berita Online menggunakan Metode Support Vector Machine (SVM) dengan Seleksi Fitur Chi-square* (Vol. 7, Issue 5). <http://j-ptiik.ub.ac.id>
- Setiawan, A., Santoso, L. W., & Adipranata, R. (n.d.). *Klasifikasi Artikel Berita Bahasa Indonesia Dengan Naive Bayes Classifier*.
- Sholih 'afif, M., Muzakir, M., Al, M. I., & Al Awalien, G. (2021). TEXT MINING UNTUK MENGLASIFIKASI JUDUL BERITA ONLINE STUDI KASUS RADAR BANJARMASIN MENGGUNAKAN METODE NAÏVE BAYES. *Kumpulan Jurnal Ilmu Komputer (KLIK)*, 08(2).
- Sripamuji, Ramadhanti, I., Riski Amalia, R., Saputra, J., & Prihatnowo, B. (n.d.). *PENERAPAN ALGORITMA SUPPORT VECTOR MACHINE DAN MULTI-LAYER PERCEPTRON PADA KLASIFIKASI TOPIK BERITA*. 11. <https://doi.org/10.23887/janapati.v1i1i2.44151>