

# Perancangan Aplikasi Mobile Menggunakan Machine Learning Untuk Menentukan Klasifikasi Kategori Berita

*By Hariyanti Novi Tri*

## Perancangan Aplikasi Mobile Menggunakan Machine Learning Untuk Menentukan Klasifikasi Kategori Berita

Novi Tri Hariyanti<sup>1\*</sup>, Titasari Rahmawati<sup>2</sup>, Alexander Wirapraja<sup>3</sup>

<sup>1</sup>Program Studi Manajemen Informatika, Institut Informatika Indonesia Surabaya, Surabaya, Jawa Timur

<sup>2,3</sup>Program Studi Sistem Informasi, Institut Informatika Indonesia Surabaya, Surabaya, Jawa Timur

Email: <sup>1</sup>novi@ikado.ac.id, <sup>2</sup>tita@ikado.ac.id, <sup>3</sup>alex@ikado.ac.id

(Naskah masuk: 9 Oct 2024, direvisi: 21 Oct 2024, diterima: 22 Oct 2024)

### Abstrak

Berita yang umumnya terdapat pada media publikasi baik elektronik maupun media cetak yang beredar setiap harinya dalam volume yang besar, saat ini sebagian besar telah berpindah ke media digital yang memudahkan pengguna untuk mengakses artikel berita. Jumlah peredaran berita yang besar setiap harinya inilah yang seringkali membebani kerja dari editor dan penulis berita dalam menentukan kategori dari berita yang akan dirilis. Sistem ini dirancang untuk membantu dalam melakukan klasifikasi kategori berita, pada aplikasi ini aplikasi dirancang dalam bentuk aplikasi portal berita berbasis aplikasi mobil berbasis android. Pada penelitian ini menggunakan metode *logistic regression* sebagai metode klasifikasi biner dengan dataset yang digunakan pada penelitian ini merupakan judul berita yang dipublikasikan pada tahun 2020 dengan pembagian data sekitar 2000 dataset. Hasil dari penelitian ini adalah sistem aplikasi yang membantu dalam melakukan klasifikasi kategori berita dengan tingkat akurasi diatas 85%.

**Kata Kunci:** Berita, Klasifikasi, Aplikasi *Mobile*, *Machine Learning*, *Logistic Regression*

## *Application of Machine Learning Using the Logistic Regression Algorithm in News Category Classification*

### *Abstract*

*News that is generally found in electronic and print media that circulates every day in large volumes, has now mostly moved to digital media which makes it easier for users to access news articles. The large number of news circulations every day is what often burdens the work of editors and news writers in determining the category of news to be released. This system is designed to help in classifying news categories, in this application the application is designed in the form of a news portal application based on an android-based mobile application. This study uses the logistic regression method as a binary classification method with the dataset used in this study being the title of the news published in 2020 with a data division of around 2000 datasets. The results of this study are an application system that helps in classifying news categories with an accuracy level above 85%.*

**Keywords:** News, Classification, Mobile Application, Machine Learning, Logistic Regression

### I. PENDAHULUAN

Informasi merupakan kebutuhan masyarakat yang disadari maupun tidak memiliki peranan penting dalam kehidupan masyarakat dewasa ini. Perkembangan teknologi yang cepat dan arus informasi yang cepat memberikan dampak kemudahan kepada masyarakat pengguna dalam melakukan akses informasi yang terkini dan terbaru. Sumber dari sebuah

informasi tersebut salah satunya berasal dari berita. Berita sendiri memiliki pengertian sebagai rupa laporan yang dapat berbentuk *interpretative* maupun *investigasi interpretative* sendiri adalah sajian berita yang dianggap penting dan telah melewati tahap redaksional sebelum dipublikasikan, sedangkan *investigasi* adalah berita yang menyajikan hasil

penyelidikan terhadap kajian latar belakang, tren, fakta dan analisis dimasa mendatang [1].

Berita wajib memberikan informasi yang berkualitas tentunya dengan mengacu kepada indikator dari kualitas informasi itu sendiri seperti akurat, relevan, tepat waktu, mudah didapatkan dan sesuai kebutuhan pengguna [2]. Sumber berita dapat berasal dari berbagai macam sumber seperti media sosial, situs portal berita, media massa, media elektronik konvensional atau bahkan *sharing* dari pengguna lainnya, bahkan mengutip data dari Reuters 68% dari masyarakat Indonesia pada tahun 2022 bahkan mendapatkan data berita dari media sosial, dimana angka ini mengalami kenaikan sebanyak 4% dibandingkan tahun sebelumnya.

Jumlah volume jumlah berita beredar perharinya tentu menjadi permasalahan tersendiri apabila tidak diklasifikasikan dengan baik. Jumlah ini tentu menjadi beban bagi editor dan penulis dalam menentukan kategori berita setiap harinya, apalagi saat ini media juga berkembang mengikuti dari kemajuan teknologi, wilayah, distribusi dan produksi sebagai bagian dari konsekuensi globalisasi [3]. Kategori berita sendiri secara umum dapat meliputi berita terbaru, olahraga, politik, ekonomi, sosial, kesehatan dan lain-lain.

Kategori berita dapat membantu penyelenggara portal berita dalam mengukur analisis sentimen yang mungkin didapatkan oleh pengelola portal berita. Analisis sentimen memiliki tujuan untuk melihat sudut pandang pembaca dalam bentuk opini baik positif maupun negatif. Positif merupakan bentuk tanggapan pembaca berita seperti menyukai, menyetujui, terhadap sebuah berita yang di publikasi, sebaliknya sentimen negatif merupakan bentuk ketidaksukaan, rasa kecewa, rasa tidak setuju dan sebagainya terhadap sebuah berita [4].

Berdasarkan dari latar belakang diatas maka pada penelitian ini mengambil topik mengenai penerapan pembelajaran mesin untuk melakukan klasifikasi kategori berita menggunakan metode *logistic regression*.

*Machine Learning* merupakan penggunaan metode untuk memprediksi hasil atau perilaku pada masa mendatang yang didasarkan pada pola data informasi dari input komputer. Pada teknik ini digunakan algoritma dan model statistik sebagai identifikasi pola dalam data dan menghasilkan luaran dalam bentuk prediksi atau keputusan tanpa instruksi yang eksplisit [5]. Keuntungan dan kerugian dalam mengimplementasikan model *machine learning* mengutip artikel yang dipublikasikan dari Amazon Web Service tahun 2023 [6] adalah:

Keuntungan model *machine learning*:

1. Membantu dalam melakukan identifikasi terhadap pola dan tren data
2. Model ini dapat bekerja secara otomatis tanpa adanya intervensi dari manusia, seperti melakukan pemantauan, proses monitoring, dan melakukan identifikasi terhadap penyimpangan pada sistem.
3. Hasil luaran yang lebih akurat berdasarkan satuan waktu
4. Memiliki kemampuan dalam menangani varian format data dengan volume tinggi pada lingkungan data yang dinamis dan kompleks

Kelemahan model *machine learning*:

1. Proses yang banyak menyita waktu dan mahal

2. Sulit diimplementasikan apabila tidak terdapat adanya kecukupan data
3. Memerlukan investasi awal yang besar
4. Terkadang untuk hasil yang lebih akurat dan benar memerlukan bantuan ahli

Metode Regresi Logistik (*Logistic Regression*) merupakan metode klasifikasi dalam *machine learning* yang bertujuan memprediksi hasil variabel dependen kategori serta mengukur hubungan antara variabel kategori dan variabel independent. *Logistic regression* juga dapat diartikan sebagai metode klasifikasi untuk memprediksi probabilitas diskrit dengan menggunakan fungsi logistik terhadap nilai probabilitas dengan luaran bilangan biner 0 atau 1 [7]. Implementasi *Logistic Regression* juga sebagai bagian dari upaya untuk memprediksi kemungkinan terjadinya suatu peristiwa dengan menghubungkannya dengan satu atau lebih variabel untuk prediksi (*predictor*) [8].

Pada metode ini algoritma yang digunakan menggunakan probabilitas dalam memprediksi data yang berkategori [9]. Pada *logistic regression* digunakan rumus untuk kasus dua kelas seperti terlihat pada Persamaan 1.

$$P(Y = 1|X) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n)}} \quad (1)$$

Dimana pada rumus tersebut dijelaskan bahwa  $P(Y=1|X)$  merupakan probabilitas dari peristiwa terjadi yang akan diprediksi, sedangkan  $\beta_0, \beta_1, \beta_2, \dots, \beta_n$  merupakan koefisien yang berasal dari data latih dan  $X_1, X_2, \dots, X_n$  adalah variabel data input.

Beberapa tipe atau jenis dari *logistics regression* adalah antara lain [10]:

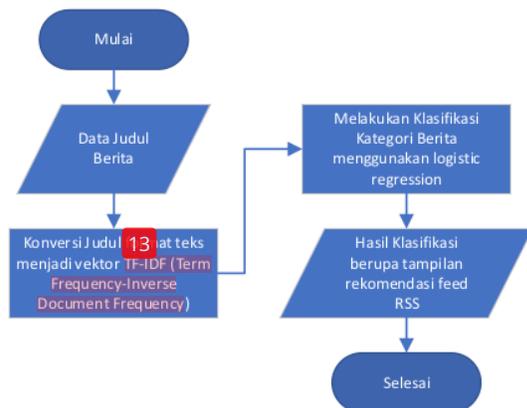
- a. *Binary logistic regression*: merupakan respon kategori yang dari dua kategori berbeda misalnya hoak atau tidak.
- b. *Multinomial logistic regression*: melakukan pengelompokkan kategori berdasarkan dari 3 atau lebih kategori berbeda tanpa urutan tertentu misalnya memprediksi masyarakat yang menyukai makanan tertentu seperti vegan, non-vegetarian dan vegetarian.
- c. *Ordinal logistic regression*: mengelompokkan kategori berdasarkan dari urutan tertentu misalnya penilaian skala layanan dari 1 hingga 5.

## II. METODOLOGI PENELITIAN

### A. Prosedur Penelitian

Target pengguna pada aplikasi ini adalah para editor dan kontributor berita agar memudahkan mereka dalam melakukan proses pengelompokkan kategori berita sesuai dengan tema berita yang mereka publikasikan.

Pada penelitian ini menggunakan prosedur penelitian dengan tahapan-tahapan pelaksanaan seperti pada Gambar 1 berikut ini:



Gambar 1. Alur Proses Klasifikasi Data

Proses awal seperti yang ditunjukkan pada Gambar 1 adalah dimana pengguna dapat melakukan prediksi klasifikasi terhadap kategori berita yang dimasukan melalui data judul berita. Proses selanjutnya adalah proses *tokenizing*, pada proses ini dilakukan pemecahan kalimat menjadi kata-kata individual secara terpisah. Pada penelitian ini proses *tokenizing*, vektorisasi kalimat dan perhitungan TF-IDF di tambahkan pada model yang dirancang menjadi satu saluran yang dapat digunakan setiap kali model digunakan.

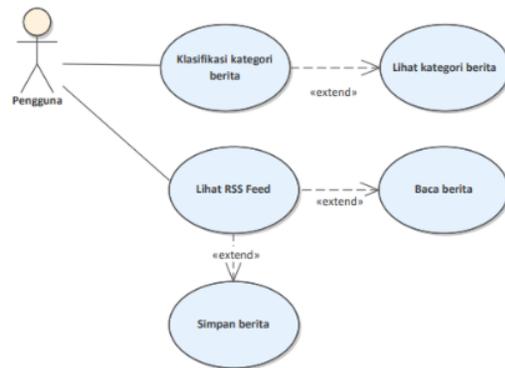
Pada proses *training* data dilakukan kombinasi menggunakan dua sumber data yang didapatkan dari SatuData Indonesia dalam bentuk rekapan judul berita yang dipublikasikan pada situs berita resmi pada tahun 2020 dengan format csv. Sistem juga menerima input dari pengguna berupa judul berita dalam bentuk teks minimal 6 kata.

Dataset yang digunakan menggunakan bahasa Indonesia dan jumlah dataset berita yang digunakan sekitar 2000 judul berita dan model yang dikembangkan akan dievaluasi akurasi menggunakan metode *confusion matrix*. Pembagian dataset dilakukan sebanyak 5 kali dengan tujuan sebagai kontrol terhadap performa model. Pada tahapan ini juga dilakukan proses pemisahan untuk mengevaluasi kinerja model.

Hasil luaran yang diterima pengguna berupa klasifikasi terhadap kategori berita dalam bentuk teks serta rekomendasi RSS feed dengan kategori yang relevan.

### B. Desain Sistem

Desain sistem yang akan dikembangkan adalah seperti pada Gambar 2 berikut ini:

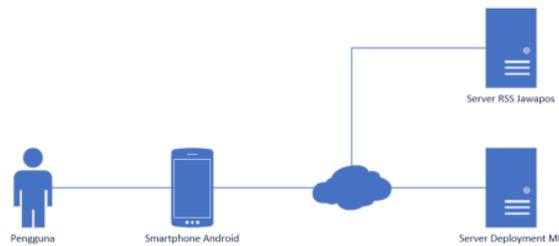


Gambar 2. Use Case Pengguna

Berdasarkan Gambar 2 dapat dijelaskan bahwa pada perancangan sistem untuk melakukan klasifikasi terhadap kategori berita, pengguna dapat melakukan prediksi klasifikasi kategori berita, dan penelusuran terhadap berita dalam kategori tersebut. Pengguna juga dapat melihat daftar berita terbaru yang ada pada *feed* RSS. Adapun fitur lainnya dalam aplikasi ini adalah dimana pengguna juga dapat melihat detail berita secara keseluruhan, dan menyimpan berita ke dalam *database* untuk dapat dibaca pada waktu yang lain.

### C. Desain Arsitektur

Desain arsitektur dari pengembangan aplikasi pada perancangan sistem untuk klasifikasi berita adalah seperti pada Gambar 3 di bawah ini.



Gambar 3. Desain Arsitektur

Pengguna dapat mengakses sistem melalui gawai berbasis Android seperti yang ditunjukkan pada Gambar 3 dan pengguna mendapatkan berita, melalui aplikasi yang berkomunikasi dengan server Jawapos melalui *service* RSS. Fungsi prediksi terhadap kategori berita dilakukan pada aplikasi dengan berkomunikasi dengan server *deployment* untuk mendapatkan hasil prediksi.

## III. HASIL DAN PEMBAHASAN

Pengambilan data sebanyak 500 entri data diambil dengan acak, dengan rasio data yang diambil dari dataset SatuData Indonesia dan data yang diambil dari dataset Kaggle sebesar 50:50. Sebelum dilakukan *pre-processing*, dilakukan

penggabungan data dari 2 dataset menjadi 1 file csv. Setelah dilakukan penggabungan kedua dataset menjadi 1, sebelum di-split menjadi *dataset training* dan *testing* terlebih dahulu dilakukan *pre-processing* data. Data yang didapatkan kemudian dilakukan proses *case folding* dimana proses ini bertujuan untuk merubah huruf kapital pada data akan diubah menjadi huruf kecil.

Pada tahapan selanjutnya dilakukan proses *stemming*. Proses ini sama seperti proses penghapusan *stopword*. Proses *stemming* adalah proses merubah sebuah kata menjadi bentuk baku, sebagai contoh adalah perubahan dari kata “belajar” menjadi “ajar”.

Pada tahapan terakhir dari penentuan kategori berita ini dilakukan *tokenizing* terhadap data-data judul berita dalam bentuk kalimat dan dikelompokkan menjadi kata-kata yang 16 diri sendiri. Pada proses ini vektorisasi kalimat dan penghitungan TF-IDF (*term frequency – inverse document frequency*) disematkan atau *embedded* pada model yang dirancang menjadi 1 *pipeline* yang akan dijalankan setiap kali model digunakan.

Proses *Tokenizing* sendiri merupakan proses yang dilakukan sebagai *preprocessing text* untuk membagi dan memisahkan teks menjadi beberapa *token*. Pada Tabel 1 berikut adalah proses sebelum dan sesudah teks pada setiap kategori yang mengalami proses *tokenizing*

Tabel 1. Hasil Sebelum dan Sesudah Proses *Tokenizing*

Teks Sebelum	Teks Sesudah	Kategori
Gibran Sebut Penyusunan Kabinet sudah Hampir 100 persen	['Gibran', 'Sebut', 'Penyusunan', 'Kabinet', 'sudah', 'Hampir', '100', 'persen']	News
Perempuan Dinilai Berpeluang Besar Menang di Pilkada 2024	['Perempuan', 'Dinilai', 'Berpeluang', 'Besar', 'Menang', 'di', 'Pilkada', '2024']	Politik
Upaya Didimax Jadi Broker Forex Lokal dengan Komisi Rendah	['Upaya', 'Didimax', 'Jadi', 'Broker', 'Forex', 'Lokal', 'dengan', 'Komisi', 'Rendah']	Ekonomi
Bahrain vs Timnas Indonesia: Mees Hilgers Beri Isyarat Bermain	['Bahrain', 'vs', 'Timnas', 'Indonesia', ':', 'Mees', 'Hilgers', 'Beri', 'Isyarat', 'Bermain']	Olahraga
PP 28 Tahun 2024, Menyisakan Tantangan Kesehatan Bayi	['PP', '28', 'Tahun', '2024', 'Menyisakan', 'Tantangan', 'Kesehatan', 'Bayi']	Kesehatan
19 WNI Dapat Diselamatkan 165 Lainnya Masih Terancam Hukuman Mati	['19', 'WNI', 'Dapat', 'Diselamatkan', '165', 'Lainnya', 'Masih', 'Terancam', 'Hukuman', 'Mati']	Lainnya

1. Implementasi Tampilan Antar Muka

Pada bagian ini membahas mengenai implementasi pada tampilan aplikasi penentuan klasifikasi kategori berita dengan tampilan sebagai berikut:

Halaman Beranda



Gambar 4. Halaman Beranda

Halaman beranda (Gambar 4) merupakan tampilan awal aplikasi yang memuat berita terbaru. Format berita dalam bentuk format *card*, dengan opsi untuk menyimpan berita sebagai favorit, membagikan berita, tombol prediksi klasifikasi kategori berita berdasarkan judul berita yang ada di kartu. Selain itu pada bagian kiri atas terdapat menu tampilan halaman untuk berpindah halaman, serta pada kanan atas terdapat tombol search untuk melakukan pencarian berita berdasarkan kata kunci.

Halaman Deteksi Kategori Berita

Pada halaman tersebut deteksi kategori berita dioperasikan pengguna dengan cara mengisikan judul berita pada kotak masukan teks yang telah disediakan. Judul berita yang 25 ah dimasukkan akan divalidasi oleh sistem, apakah telah sesuai dengan ketentuan atau tidak. Jika 10 ul berita yang dimasukkan tidak sesuai ketentuan, maka sistem akan menampilkan pesan *error*. Jika judul berita telah sesuai dengan ketentuan, maka sistem akan melakukan klasifikasi kategori berita dari judul dan menampilkan hasil klasifikasi kategori berita tersebut kepada pengguna.

Tampilan pada Gambar 5 berikut adalah halaman deteksi kategori berita, dimana pada halaman ini bekerja dengan melakukan deteksi melalui judul berita dan memberikan prediksi kategori berita.

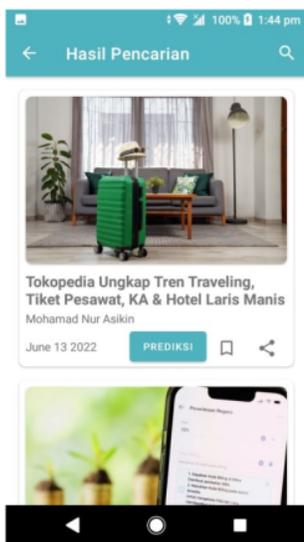


Gambar 5. Halaman Beranda

Pada Gambar 5 menampilkan halaman yang ditujukan untuk pengguna agar dapat berinteraksi dengan sistem klasifikasi kategori berita. Pada halaman ini pengguna dapat memasukkan judul berita sesuai dengan hint pada kolom isian yang tersedia. Selanjutnya Pengguna berinteraksi kepada sistem melalui menu “prediksi”, sistem akan melakukan request ke *deployment server*, lalu akan menampilkan hasil prediksi kategori berita beserta dengan daftar berita sesuai dengan kategori hasil prediksi.

**Halaman Hasil Pencarian Berita**

Halaman ini merupakan halaman hasil pencarian berita berdasarkan dari kata kunci yang terdapat pada judul berita.



Gambar 6. Halaman Beranda

Tampilan halaman seperti yang ditunjukkan pada Gambar 6 akan ditampilkan hasil pencarian berita. Jika berita yang sesuai dengan kata kunci yang dimasukkan ditemukan, maka berita akan ditampilkan dalam format *card* sama seperti halaman beranda. Apabila berita yang dicari tidak ditemukan, maka akan ditampilkan pesan pada tampilan yaitu “berita tidak ditemukan”.

Berita yang ditampilkan adalah 25 hasil pencarian diurutkan dari yang paling baru, dengan range pencarian berita adalah seluruh berita yang ada pada server RSS dengan kata kunci terkait.

**Halaman Bantuan Melalui Menu**

Halaman ini merupakan halaman bantuan yang terdapat pada menu awal pada saat pengguna pertama kali membuka aplikasi.



Gambar 7. Halaman Beranda

Pada Gambar 7 merupakan halaman yang terdapat gambar dan teks yang menjelaskan beberapa fitur dari aplikasi. Pengguna juga dapat membuka halaman ini melalui menu, dan artikel bantuan akan disajikan dalam bentuk daftar kartu.

**2. Hasil Uji Coba Aplikasi Dengan Panjang Judul Berita**

Pengujian dilakukan terhadap model diuji dengan panjang judul berita yang berbeda. Panjang judul berita yang digunakan berkisar antara satu hingga enam kata. Berita diambil secara acak dari dataset yang digunakan untuk training dengan judul berita per kategori. Metrik yang digunakan pada error analysis adalah F1 score, MSE, dan akurasi. Berikut pada Tabel 2 adalah contoh pengujian dengan menggunakan 1 kata dan 6 kata.

Tabel 2. Hasil Pengujian 1 Kata

Kategori	Accuracy	F1 Score	MSE
News	0.9	0.667	0.001
Politik	0.9	0.571	0.01
Olahraga	0.9	0.667	0.001

Kategori	Accuracy	F1 Score	MSE
Ekonomi	0.9	0.571	0.01
Kesehatan	0.93	0.8	0
Lainnya	0.73	0.556	0.071

Pada Tabel 2 menunjukkan hasil bahwa akurasi, MSE dan F1 score terlihat cukup baik pada judul berita dengan 1 kata, dimana hasil akurasi berada diatas 90% untuk kelima kategori utama dan 73% untuk kategori lainnya.

Tabel 3. Hasil Pengujian 6 Kata

Kategori	Accuracy	F1 Score	MSE
News	0.9	0.667	0.001
Politik	0.9	0.571	0.01
Olahraga	0.9	0.667	0.001
Ekonomi	0.967	0.909	0.001
Kesehatan	0.9	0.727	0.001
Lainnya	0.9	0.769	0.01

Pada Tabel 3 tersaji hasil pengujian pada panjang judul berita yang memiliki 6 kata, juga didapatkan tingg akurasi yang berada diatas 90%. Maka berdasarkan dari hasil yang ditunjukkan pada Tabel 2 dan 3 dapat terlihat bahwa semakin panjang judul berita yang dimasukkan, hasil klasifikasi juga menjadi semakin akurat. Bagi pengguna hasil ini memberikan masukan agar pengguna melakukan input pada aplikasi dengan memasukkan judul berita dengan panjang setidaknya 6 kata.

#### IV. KESIMPULAN

Kesimpulan dalam penelitian ini didapatkan kesimpulan bahwa fitur deteksi kategori berita secara otomatis, dapat membantu pengguna dalam mengetahui kategori berita dan mendapatkan berita dengan kategori yang sama dengan lebih cepat walau menggunakan judul berita yang tidak memiliki kategori yang jelas. Implementasi aplikasi berbasis *mobile* dalam penentuan klasifikasi dan kategori berita ini dapat dikatakan cukup efektif, hal ini bergantung pada dataset dimana semakin banyak dataset yang digunakan maka akan berdampak baik pada performa model.

#### UCAPAN TERIMA KASIH

Tim peneliti mengucapkan banyak terima kasih kepada Direktorat Akademik Pendidikan Tinggi Vokasi, Direktur Jenderal Pendidikan Vokasi, Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi, Republik Indonesia atas bantuan yang diberikan melalui hibah Penelitian Dosen Pemula (PDP) dengan nomor kontrak: 010/SP2H/PPKM-

PTV.3/LL7/2024 sehingga penelitian ini dapat berjalan dengan baik dan lancar baik dari awal pengumpulan data hingga pada proses dokumentasi hasil penelitian.

#### REFERENSI

- [1] E. Effendy, F. Hasugian, and M. A. Harahap, "Menulis Isi Berita Dan Feature," *J. Pendidik. dan Konseling*, vol. 4, pp. 1349–1358, 2022.
- [2] A. Wirapraja, H. Aribowo, and E. T. Setyoadi, "The Influence of E-Service Quality, and Customer Satisfaction On Go-Send Customer Loyalty In Surabaya," *Indones. J. Inf. Syst.*, vol. 3, no. 2, p. 128, 2021, doi: 10.24002/ijis.v3i2.4191.
- [3] A. A. Syaifudin and A. A. Darmaningtyas, "Strategi Manajemen Pengelolaan Media Sosial Sebagai Akun Publik: Analisis Model Komunikasi Akun Instagram @liputan.kendal.terkini," *Soc. Polit. Sci.*, vol. 1, no. 2, pp. 15–34, 2022.
- [4] A. D. Dayani, Yuhandri, and G. W. Nurcahyo, "Analisis Sentimen Terhadap Opini Publik pada Sosial Media Twitter Menggunakan Metode Support Vector Machine," *J. KomtekInfo*, vol. 11, pp. 1–10, 2024, doi: 10.35134/komtekinfo.v11i1.439.
- [5] N. F. Sahamony, T. Terttiaavini, and H. Rianto, "Analisis Perbandingan Kinerja Model Machine Learning untuk Memprediksi Risiko Stunting pada Pertumbuhan Anak," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 4, no. 2, pp. 413–422, 2024, doi: 10.57152/malcom.v4i2.1210.
- [6] AWS, "Apa itu Machine Learning?," <https://aws.amazon.com/>, 2023. .
- [7] B. B. Tangkere, "Analisis Performa Logistic Regression dan Support Vector Classification untuk Klasifikasi Email Phising," *J. Ekon. Manaj. Sist. Inf.*, vol. 5, no. 4, pp. 442–450, 2024, doi: 10.31933/jemsi.v5i4.1916.
- [8] S. Joses, D. Yulvida, and S. Rochimah, "Pendekatan Metode Ensemble Learning untuk Prakiraan Cuaca menggunakan Soft Voting Classifier," *J. Appl. Comput. Sci. Technol.*, vol. 5, no. 1, pp. 72–80, 2024, doi: 10.52158/jacost.v5i1.741.
- [9] C. Sonjaya, A. F. N. Masruriyah, D. S. Kusumaningrum, and A. R. Pratama, "The Performance Comparison of Classification Algorithm in Order to Detecting Heart Disease," *Intern. (Information Syst. Journal)*, vol. 5, no. 2, pp. 166–175, 2022, doi: 10.32627/internal.v5i2.595.
- [10] S. Swaminathan, "Logistic Regression — Detailed Overview," [towardsdatascience.com](https://towardsdatascience.com/logistic-regression-detailed-overview-46c4da4303bc), 2018. <https://towardsdatascience.com/logistic-regression-detailed-overview-46c4da4303bc>.

# Perancangan Aplikasi Mobile Menggunakan Machine Learning Untuk Menentukan Klasifikasi Kategori Berita

ORIGINALITY REPORT

# 13%

SIMILARITY INDEX

## PRIMARY SOURCES

1	<a href="http://ejournal.ikado.ac.id">ejournal.ikado.ac.id</a> Internet	58 words — 2%
2	Heri Supriyanto, Mohammad Al Hafidz, Ari Cahaya Puspitaningrum, Rayhan Abdillah Putra Firmansyah, Rafi Zuhdi. "Klasterisasi Data Obat Farmasi Berdasarkan Jumlah Persediaan Dengan Menggunakan Metode K-Means", <i>Teknika</i> , 2024 Crossref	44 words — 1%
3	Sri Rahayu, Syahrul Yanuar, Yosep Bustomi. "Optimizing Tourism Promotion for Situ Bagendit Through Innovation in a Web-Based Virtual Tour Application", <i>Teknika</i> , 2024 Crossref	44 words — 1%
4	<a href="http://journal.irpi.or.id">journal.irpi.or.id</a> Internet	34 words — 1%
5	<a href="http://ners.fkep.unand.ac.id">ners.fkep.unand.ac.id</a> Internet	22 words — 1%
6	<a href="http://www.jpnn.com">www.jpnn.com</a> Internet	20 words — 1%
7	<a href="http://pepadun.fmipa.unila.ac.id">pepadun.fmipa.unila.ac.id</a> Internet	15 words — 1%

8	<a href="http://ejurnal.itenas.ac.id">ejurnal.itenas.ac.id</a> Internet	13 words — < 1%
9	<a href="http://journal.isas.or.id">journal.isas.or.id</a> Internet	13 words — < 1%
10	Fadillah Said, Dwi Andriyanto, Retno Sari, Windu Gata. "Perancangan Validasi Permohonan Narasumber Pada Sistem Informasi Permohonan Narasumber Menggunakan Finite State Automata", Paradigma - Jurnal Komputer dan Informatika, 2020 Crossref	11 words — < 1%
11	<a href="http://issuu.com">issuu.com</a> Internet	10 words — < 1%
12	<a href="http://id.scribd.com">id.scribd.com</a> Internet	9 words — < 1%
13	<a href="http://journal.ipm2kpe.or.id">journal.ipm2kpe.or.id</a> Internet	9 words — < 1%
14	<a href="http://journal.unilak.ac.id">journal.unilak.ac.id</a> Internet	9 words — < 1%
15	<a href="http://www.slideshare.net">www.slideshare.net</a> Internet	9 words — < 1%
16	<a href="http://123dok.com">123dok.com</a> Internet	8 words — < 1%
17	Chepy Sonjaya, Anis Fitri Nur Masruriyah, Dwi Sulistya Kusumaningrum, Adi Rizky Pratama. "The Performance Comparison of Classification Algorithm in Order to Detecting Heart Disease", INTERNAL (Information System Journal), 2022 Crossref	8 words — < 1%

- 
- 18 Imam Agus Faizal, Dini Puspodewi, Ajeng Puspo Aji. "Subchronic Toxicity Test of Burdock Root (Arctium Lappa L.) Nanoemulsion Based on Histopathology on The Albino Rat Rattus noverticus (Wistar strain)", Indonesian Journal of Innovation Multidisipliner Research, 2024  
Crossref 8 words — < 1%
- 
- 19 Jasman Pardede. "Deteksi Komentar Cyberbullying Pada Media Sosial Berbahasa Inggris Menggunakan Naïve Bayes Classification", Jurnal Informatika, 2020  
Crossref 8 words — < 1%
- 
- 20 [garuda.kemdikbud.go.id](http://garuda.kemdikbud.go.id)  
Internet 8 words — < 1%
- 
- 21 [journal.ppmi.web.id](http://journal.ppmi.web.id)  
Internet 8 words — < 1%
- 
- 22 "Data-Centric Business and Applications", Springer Science and Business Media LLC, 2024  
Crossref 7 words — < 1%
- 
- 23 Kuncoro Hadi, Ema Utami. "Analysis of K-NN with the Integration of Bag of Words, TF-IDF, and N-Grams for Hate Speech Classification on Twitter", JUITA: Jurnal Informatika, 2024  
Crossref 7 words — < 1%
- 
- 24 Sally Nurlita Zaman, Nita Merlina, Nurajijah Nurajijah. "SISTEM INFORMASI KELUHAN PELANGGAN BERBASIS WEBSITE", EVOLUSI : Jurnal Sains dan Manajemen, 2021  
Crossref 7 words — < 1%
- 
- 25 [repository.ub.ac.id](http://repository.ub.ac.id)  
Internet 6 words — < 1%

---

EXCLUDE QUOTES      OFF

EXCLUDE BIBLIOGRAPHY   ON

EXCLUDE SOURCES      OFF

EXCLUDE MATCHES      OFF