

Klasifikasi Text Dokumen Web Berbasis Supervised Learning Sebagai Pemodelan Aplikasi Pembelajaran Kebudayaan Melayu di Indonesia

Mustakim*, Febi Nur Salisah, Suryani

Fakultas Sains dan Teknologi, Program Studi Sistem Informasi, Universitas Islam Negeri Sultan Syarif Kasim Riau, Pekanbaru, Indonesia

Email: ^{1,*}mustakim@uin-suska.ac.id, ²febinursalisah@uin-suska.ac.id, ³suryani@gmail.com

Email Penulis Korespondensi: mustakim@uin-suska.ac.id

Submitted: 09/10/2025; Accepted: 31/12/2025; Published: 31/12/2025

Abstrak—Indonesia, sebagai negara kepulauan terbesar, memiliki keberagaman budaya, salah satunya kebudayaan Melayu di Provinsi Riau. Terdapat banyak dokumen text pada website baik dalam bentuk artikel, berita maupun dokumen unggahan pribadi masyarakat pelaku budaya. Penelitian ini bertujuan untuk mendukung pelestarian kebudayaan Melayu melalui teknologi dengan mengimplementasikan sistem pembelajaran digital berbasis Machine Learning. Penelitian sebelumnya mengidentifikasi kelemahan dalam penerapan sistem cerdas dan algoritma pembelajaran mesin. Penelitian ini menguji lima algoritma klasifikasi Random Forest, SVM, Naïve Bayes, KNN, dan PNN untuk meningkatkan akurasi dan performa sistem. Hasil penelitian menunjukkan bahwa Random Forest mencapai akurasi tertinggi 91,17%, diikuti KNN 88,23%, SVM dan NBC 82,35%, dan PNN 76,47%. Sistem Digital Learning System (DLS) yang dikembangkan mendapatkan umpan balik positif dengan skor User Acceptance Test (UAT) 86% dan keberhasilan 100% pada pengujian Blackbox, menunjukkan kinerja yang stabil di berbagai perangkat. Penelitian ini memperkenalkan inovasi baru dalam aplikasi pelestarian kebudayaan Melayu menggunakan algoritma Machine Learning untuk meningkatkan akurasi dan fungsionalitas.

Kata Kunci: KNN; NBC; PNN; Random Forest, SVM

Abstract—Indonesia, as the largest archipelagic country, is home to diverse cultures, including Malay culture in Riau Province. The website features numerous text documents, including articles, news, and personal documents, uploaded by members of the cultural community. This study aims to support the preservation of Malay culture through technology by implementing a digital learning system based on Machine Learning. Previous research has identified weaknesses in the application of intelligent systems and machine learning algorithms. This study tests five classification algorithms Random Forest, SVM, Naïve Bayes, KNN, and PNN to improve the system's accuracy and performance. The results show that Random Forest achieved the highest accuracy of 91.17%, followed by KNN at 88.23%, SVM and NBC at 82.35%, and PNN at 76.47%. The developed Digital Learning System (DLS) received positive feedback, with a User Acceptance Test (UAT) score of 86% and a 100% success rate in Blackbox testing, demonstrating stable performance across various devices. This research introduces a new innovation in Malay cultural preservation applications, utilizing Machine Learning algorithms to enhance both accuracy and functionality.

Keywords: KNN; NBC; PNN; Random Forest, SVM

1. PENDAHULUAN

Indonesia merupakan negara kepulauan terbesar di dunia, yang memiliki keaneka ragam suku dan budaya didalamnya. Berbagai macam suku budaya dan ciri khas di Indonesia disetiap daerahnya memiliki perbedaan. Ciri khas budaya dari tiap daerah meliputi adat istiadat, pakaian daerah, rumah adat, kesenian daerah, bahasa daerah hingga masakan khas daerah [1]. Pada era globalisasi saat ini nilai-nilai kebudayaan dalam masyarakat seakan semakin punah tergerus oleh perkembangan sosial media yang memiliki dampak besar bagi tatanan kebudayaan dan agama. Kondisi ini mengakibatkan warisan budaya dan kearifan lokal di Indonesia nantinya semakin punah [2]. Salah satunya budaya yang menjadi ciri khas Islam di Indonesia adalah kebudayaan Melayu.

Provinsi Riau atau yang dikenal dengan sebutan Bumi Lancang Kuning, memiliki sejarah kerajaan Melayu terbesar di Indonesia dan merupakan daerah yang memiliki beraneka ragam kebudayaan yang lekat dengan kebudayaan Islam [2]. Kebudayaan Melayu Riau merupakan hasil pemikiran, rasa dan kerja orang Melayu yang berbahasa Melayu, beradat istiadat Melayu, dan beragama Islam [1]. Undang-undang Republik Indonesia Nomor 5 Tahun 2017 menyatakan bahwa kebudayaan Melayu yang berada di Provinsi Riau memberikan kontribusi sangat besar terhadap kebudayaan Nasional. Budaya Nusantara dengan nilai religius menjadi kontradiktif dengan keadaan saat ini yang sarat dengan pengaruh barat. Dengan kondisi tersebut, selayaknya Generasi Emas Indonesia perlu mendukung dan mengkampanyekan kembali kebudayaan yang bersinggungan langsung dengan budaya Islam tersebut melalui pengembangan teknologi.

Beberapa penelitian telah menyatakan bahwa hal tersebut diperlukannya pengembangan teknologi informasi untuk menyebar luaskan informasi terkait kebudayaan di Indonesia. Seperti halnya penelitian yang dilakukan oleh Muhammad Juanda Saputra untuk kebudayaan Aceh [3], Abdurahman Dayat untuk kebudayaan Papua [4], Yulisman untuk kebudayaan Riau [2] dan Ahmad Suryadi untuk Pengenalan suku di Indonesia [4]. Kelemahan dari penelitian-penelitian tersebut terletak pada dua hal penting dalam menerapkan aspek teknologi yaitu Sistem Cerdas (Intelligence System) dan Mesin Pembelajaran (Machine Learning) yang tidak diimplementasikan dalam penelitian tersebut sehingga cenderung hanya sebatas aplikasi biasa yang tidak merekomendasikan apapun. Penelitian pada tahun 2023 dilakukan oleh Mustakim dan Febi Nur Salisah, mengangkat kebudayaan Melayu di Indonesia yang diterapkan dalam platform Digital Learning System (DLS) telah sukses diimplementasikan [5]. Dimana dalam pengembangannya,

menggunakan 2 (dua) algoritma yaitu Density-Based Spatial Clustering of Application Noise (DBSCAN) dan Frequent Pattern Growth (FP-Growth). Dari penelitian yang dilakukan tersebut, terdapat satu model yang masih belum secara powerfull mengatasi keterbatasan dokumen teks dalam jumlah besar, sehingga menghasilkan akurasi dan validitas yang sangat rendah. Hal ini menyebabkan pembelajaran mesin pada aplikasi tidak secara menyeluruh powerfull dalam mengatasi beberapa kondisi tertentu dan akurasi dalam mengelompokkan data. Namun dari sisi penggunaan aplikasi DLS sudah sesuai yang diharapkan, yaitu dengan akurasi uji sebesar 86% dan 100% untuk kehandalan aplikasi [5]. Kelemahan dari algoritma DBSCAN dari penelitian tersebut tentunya menjadi bagian penting yang perlu diteliti kembali terhadap aplikasi Kebudayaan Melayu yang telah dibangun. Hal ini bertujuan untuk mencapai akurasi dan performa yang tinggi dari sisi pemodelannya. Sebagai bahan pembandingan dari penelitian tersebut dengan beberapa kendala, maka dalam penelitian ini akan mencobakan konsep Supervised Learning atau yang sering dikenal dengan teknik Klasifikasi Data. Penelitian ini selain dari memperbaiki kelemahan-kelemahan dari penelitian sebelumnya, juga akan melakukan pemodelan konsep baru dengan melakukan pembelajaran mesin yang lain dalam aplikasi dengan menyisipkan model matematis, tentunya belum pernah dilakukan oleh peneliti-peneliti sebelumnya khusus untuk aplikasi Kebudayaan di Indonesia.

Teknologi Machine Learning hingga saat ini berkembang pesat dan telah diimplementasikan di banyak platform salah satunya adalah mobile based [5]. Machine Learning adalah satu bidang ilmu yang menerapkan beberapa konsep data dan matematika dalam melakukan pemodelan, penemuan pola hingga menentukan konsep baru dalam penyelesaian persoalan [6]. Salah satu bagian dari Machine Learning adalah teknologi learning berbasis text, atau sering dikenal dengan Text Mining [7]. Pemodelan yang dilakukan dalam pembelajarannya dengan menghitung beberapa dokumen text berdasarkan jarak dan aturan yang dikenal dengan TF-IDF [8], pemrosesan dengan menerapkan suatu algoritma tertentu untuk memberikan validitas atau akurasi yang lebih baik [9] dan menghasilkan simpulan sesuai penghitungan suatu metode tertentu [10]. Implementasinya dalam kasus text mining untuk Kebudayaan Melayu dalam penelitian ini akan mencobakan beberapa algoritma klasifikasi seperti Random Forest, Support Vector Machine (SVM), Naïve Bayes Classifier (NBC), K-Nearest Neighbor (K-NN) dan Probabilistic Neural Network (PNN). Seperti halnya teknik klasifikasi yang lain, beberapa algoritma tersebut memerlukan support dengan kemampuan mesin yang tinggi [11], melakukan pelatihan dengan baik [12], membagi data menjadi dua bagian training dan testing [13] dan melakukan proses untuk menghasilkan akurasi tinggi [14]. Penelitian-penelitian terdahulu berbasis text mining telah banyak menggunakan algoritma-algoritma tersebut dengan akurasi terbaik. Diantaranya adalah Random Forest oleh Nasir Jalal dan kawan-kawan menyatakan bahwa kehandalan Random Forest terbukti baik untuk klasifikasi teks [15], SVM oleh Xiaofang Zheng dan Richard Schweickert mengatakan bahwa SVM dapat mengatasi klasifikasi text dan biner [12]. NBC memiliki ketetapan terdekat yang dapat mengasumsikan independensi lebih mungkin untuk menjadi valid [16] sedangkan K-NN merupakan algoritma dengan sensitifitas baik untuk kasus text classification [14]. Algoritma PNN pernah digunakan oleh Mustakim dan Rice Novita pada penelitian 2023 menyatakan bahwa keterbatasan beberapa klasifikasi dapat diatasi dengan model PNN [17]. Algoritma-algoritma ini memiliki kelebihan-kelebihan dan performa tinggi dapat dimanfaatkan untuk pengembangan aplikasi mobile dengan hasil model terbaik.

Penelitian ini memanfaatkan data teks yang diperoleh dari berbagai sumber di web, mencakup artikel, berita, serta dokumen unggahan pribadi masyarakat pelaku budaya yang berkaitan dengan kebudayaan Melayu di Indonesia. Seluruh data teks tersebut akan diolah menggunakan pendekatan text mining untuk mengekstraksi informasi penting dan melakukan proses klasifikasi. Dalam tahap analisis, penelitian ini menerapkan beberapa algoritma supervised learning yaitu Random Forest, Support Vector Machine (SVM), Naïve Bayes, K-Nearest Neighbor (KNN), dan Probabilistic Neural Network (PNN). Setiap algoritma akan dibandingkan berdasarkan tingkat akurasi, presisi, dan f-measure untuk menentukan model terbaik dalam mengklasifikasi teks kebudayaan Melayu. Hasil penelitian diharapkan menghasilkan algoritma paling optimal yang dapat menjadi dasar pengembangan sistem pembelajaran digital berbasis kebudayaan Melayu.

Keberhasilan algoritma Machine Learning pada aplikasi mobile diterapkan pada penelitian ini dengan melakukan percobaan dengan membandingkan 5 (lima) algoritma untuk memperoleh performa terbaik. Akurasi dari pemodelan yang dilakukan mampu merekomendasikan aplikasi yang memiliki tingkat kesamaan terdekat untuk dapat mengakses dokumen secara baik. Sehingga user dapat dengan mudah melakukan identifikasi secara langsung baik dari sisi artikel maupun dari sub bagian kebudayaan Melayu sebagai objek utamanya. Kebaruan dari penelitian ini meliputi aspek percobaan yang mensubstitusikan sebuah model matematika kedalam suatu algoritma algoritma Machine Learning dengan formulasi terbaik diterapkan pada aplikasi DLS, sebagai mesin cerdas dalam pembelajaran kebudayaan Melayu di Indonesia yang sampai saat ini belum pernah dilakukan oleh peneliti lain. Selain itu penelitian ini akan memperbaiki kelemahan penelitian terdahulu untuk meningkatkan performa yang lebih baik dalam aplikasi DLS.

2. METODOLOGI PENELITIAN

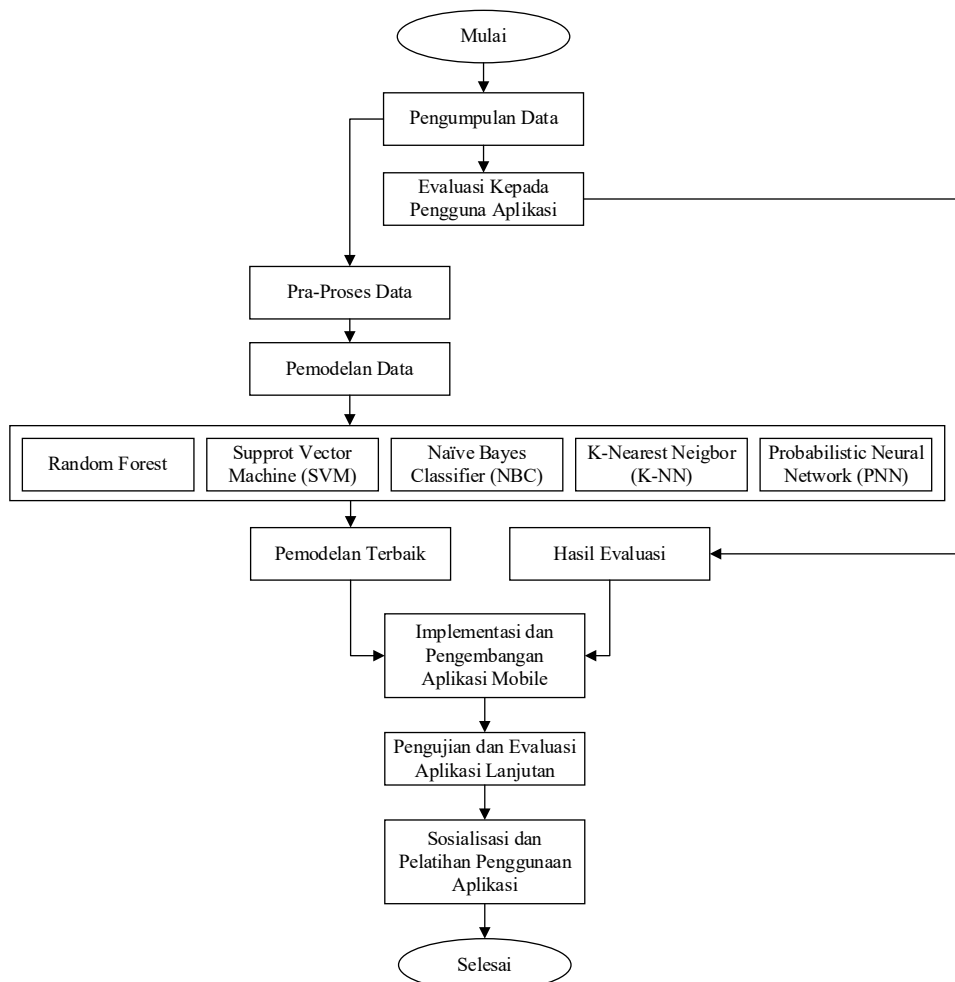
2.1 Alur dan Proses Penelitian

Proses penelitian ini dibagi menjadi empat tahapan utama yaitu (1) Pengumpulan Data dan Evaluasi User; (2) pra proses data menggunakan Feature Extraction Term Frequency-Inverse Document Frequency (TF-IDF), percobaan

dan pemodelan data menggunakan 5 (lima) algoritma Machine Learning yaitu Random Forest, Support Vector Machine (SVM), Naïve Bayes Classifier (NBC), K-Nearest Neighbor (K-NN) dan Probabilistic Neural Network (PNN); (3) Pengembangan dan optimalisasi aplikasi mobile dengan menggunakan platform android dan web, dibangun dengan menerapkan konsep Object Oriented Programming (OOP) dengan pendekatan waterfall model serta memanfaatkan framework Android Studio dan database MySQL sebagai basis penyimpanan data; (4) pengujian aplikasi dan pelatihan kepada pengguna lanjutan.

Atribut dan kategori utama yang digunakan dalam penelitian ini adalah (1) Tradisi Lisan, (2) Manuskrip, (3) Adat Istiadat, (4) Ritus, (5) Pengetahuan Tradisional, (6) Teknologi Tradisional, (7) Seni, (8) Bahasa, (9) Permainan Rakyat dan (10) Olahraga Tradisional. Atribut ini dijadikan sebagai pedoman dan rujukan untuk pengumpulan data text dan pengembangan struktur aplikasi mobile. Secara umum metodologi penelitian dapat dilihat pada Gambar 1.

Model terbaik dari pemodelan akan diterapkan pada bahasa pemrograman berbasis mobile dengan merekomendasikan beberapa fitur diantaranya adalah menentukan wilayah berdasarkan dari cerita rakyat berdasarkan radius terdekat dengan tempat dimana aplikasi aktif. Kedua, aplikasi dapat secara otomatis memberikan informasi seputar cerita rakyat berdasarkan atribut, demografi dan kata kunci, yang dapat diakses dalam aplikasi.



Gambar 1. Metodologi Penelitian

2.2 Kajian Pustaka

2.2.1 Algoritma Random Forest

Random Forest merupakan algoritma Machine Learning yang digunakan untuk klasifikasi dan prediksi [18]. RF menggunakan konsep bagging untuk memperbaiki proses prediksi yang dilakukan [19]. Algoritma ini berhasil menyelesaikan masalah regresi dan klasifikasi selama beberapa tahun terakhir [20]. Hasil klasifikasi algoritma ini bergantung pada pohon keputusan [21]. Persamaan 1 digunakan untuk menghitung nilai entropy yaitu.

$$Entropy(Y) = - \sum_i p(c|Y) \log_2 p(c|Y) \tag{1}$$

2.2.2 Algoritma Support Vector Machine (SVM)

Support Vector Machine (SVM) diperkenalkan sebagai model Machine Learning berbasis kernel dengan pengenalan pola untuk klasifikasi dan regresi [22]. Algoritma SVM secara linear memisahkan kelas dengan menampilkan fitur ke

dalam kernel. Secara umum, SVM bekerja dengan menambahkan ruang fitur dengan dimensi tambahan [23]. SVM dapat menguji tingkat keakuratan data dalam sebuah sistem. Pembagian data yang dilakukan SVM dilakukan dengan cara mencari titik vector terdekat yang mendekati hyperline dari tiap kelompok [24]. Persamaan Linear pada SVM ditunjukkan pada persamaan 2.

$$f(x) = \text{sign}(w \cdot x + b) \tag{2}$$

2.2.3 Algoritma Naïve Bayes Classifier (NBC)

Naïve Bayes Classifier merupakan teori Bayesian yang digunakan untuk pengambilan keputusan suatu informasi [25]. Algoritma ini merupakan aplikasi dari model pembelajaran bayes [26]. Model Naive Bayes kompatibel untuk membangun kumpulan data besar sebelum dilakukan analisis lanjutan [12]. NBC merupakan model klasifikasi yang sederhana namun memiliki keakuratan dalam kasus yang rumit. Naive Bayes dapat dilihat pada persamaan 3.

$$Y = w^T x_i + b, \quad i = 1, 2, \dots, l \tag{3}$$

2.2.4 Algoritma K-Nearest Neighbor (K-NN)

K Nearest Neighbor (KNN) merupakan algoritma Supervised Learning yang mengandalkan data input berlabel untuk mempelajari fungsi dengan output yang sesuai ketika berlabel baru dan tanpa label [27]. Algoritma ini bekerja dengan mengklasifikasikan objek berdasarkan data pembelajaran yang memiliki kedekatan minimum [28], di mana hasil query instance yang baru diklasifikasikan berdasarkan nilai mayoritas kedekatan jarak dari kategori yang ada [29] [30]. Rumus umum dari KNN dapat dilihat pada persamaan 4.

$$P(Y|X) = \frac{P(X|Y) \cdot P(Y)}{P(X)} \tag{4}$$

2.2.5 Algoritma Probabilistic Neural Network (PNN)

Salah satu jaringan syaraf tiruan yang cepat dan efisien dalam menyelesaikan masalah pengenalan pola adalah Probabilistic Neural Network (PNN) [31]. PNN merupakan jaringan syaraf tiruan yang bekerja berdasarkan algoritma statistic dan yang banyak digunakan dan dikembangkan berbagai kepentingan manusia [17]. Keuntungan menerapkan algoritma PNN adalah kecepatan proses, mudah, dan instan [10]. Arsitektur PNN terdiri atas 4 lapisan yaitu Input layer, Pattern Layer, Summation Layer dan Output Layer. Persamaan umum PNN dapat dilihat pada persamaan 5-7.

$$W_{i,j}(x) = \frac{1}{2\pi^{d/2} \sigma^d} \exp[-|x - x_{ij}|^2 / (2\sigma^2)] \tag{5}$$

$$g_i(x) = \frac{1}{2\pi^{d/2} \pi^d} \frac{1}{N_i} \sum_{j=1}^{N_i} N_i \exp[-|x - w_i|^2 / (2\sigma_k^2)] \tag{6}$$

$$G_i(x) = \text{argmax}\{g_i(x)\}, \quad i = 1, 2, \dots, m \tag{7}$$

4. HASIL DAN PEMBAHASAN

4.1 Dataset Penelitian

Proses yang pertama dilakukan adalah pengumpulan dataset, dalam hal ini dataset yang digunakan adalah data tidak terstruktur dari internet berupa data-data yang berkaitan dengan kebudayaan melayu, jumlah data yang digunakan terdiri dari 14 kategori dan 156 data diambil dengan cara scrapping web menggunakan python programming. Dari 14 kelas disederhanakan menjadi 6 kelas, penyederhanaan ini berdasarkan hasil penelitian dan wawancara dengan pihak LAM Riau. Dataset atau corpus yang digunakan sebagai sampel dapat dilihat pada Tabel 1.

Tabel 1. Sampel Dataset Final Proses Machine Learning

No Urut Data	Kelas	ID Kelas	Text
1	Legenda, Adat dan Budaya	1	Siak Sri Indrapura adalah sebuah kerajaan di Riau yang memiliki banyak cerita rakyat. Salah satu legenda terkenal menceritakan tentang pendirian kerajaan ini oleh Raja Kecil, seorang pemuda yang bijaksana dan berani.
2	Legenda, Adat dan Budaya	1	Sang Nila Utama adalah seorang pahlawan legendaris yang konon merupakan pendiri Singapura. Menurut legenda, ia datang ke pulau tersebut dan melihat seekor singa, yang kemudian menjadi nama Singapura. Kisah ini memiliki keterkaitan dengan wilayah Riau karena terdapat hubungan historis antara Riau dan Singapura.
9	Bahasa, Tulisan dan Lisan	2	Gurindam adalah bentuk puisi lisan yang terdiri dari dua larik, sedangkan pantun adalah bentuk puisi dengan empat larik. Kedua jenis puisi ini sering digunakan untuk mengungkapkan perasaan, nasihat, atau humor dalam



No Urut Data	Kelas	ID Kelas	Text
10	Bahasa, Tulisan dan Lisan	2	bahasa Melayu. Masyarakat Melayu Riau sangat terampil dalam menciptakan gurindam dan pantun. Lagu-lagu tradisional Melayu Riau merupakan bagian penting dari tradisi lisan mereka. Lagu-lagu ini sering dinyanyikan dalam berbagai acara, seperti upacara adat, pernikahan, atau perayaan budaya. Mereka sering diiringi oleh alat musik tradisional seperti gambus dan gendang.
29	Legenda, Adat dan Budaya	1	Adat Perkawinan: Upacara perkawinan di masyarakat Melayu Riau adalah salah satu bagian yang paling penting dari adat istiadat mereka. Proses perkawinan melibatkan serangkaian langkah dan ritual, termasuk lamaran, pertukaran cincin, dan upacara akad nikah. Adat perkawinan ini sering melibatkan banyak anggota keluarga dan menjadi momen yang sangat penting dalam kehidupan masyarakat Melayu Riau.
57	Ilmu, Teknologi dan Seni	3	Pengetahuan Navigasi Laut: Masyarakat Melayu Riau, terutama yang tinggal di daerah pesisir, memiliki pengetahuan yang mendalam tentang navigasi laut. Mereka tahu bagaimana membaca tanda-tanda alam, seperti arah angin, gelombang laut, dan bintang, untuk berlayar dan menangkap ikan dengan sukses.
99	Permainan dan Olahraga	4	Sepak Takraw: Sepak takraw adalah permainan tradisional yang mirip dengan voli, tetapi menggunakan bola bulu yang dipukul dengan kaki atau kepala. Ini adalah permainan yang sangat populer di seluruh wilayah Melayu, termasuk Riau.
100	Permainan dan Olahraga	4	Gasing: Gasing adalah permainan tradisional yang melibatkan pelemparan gasing (papan berputar) ke tanah. Tujuan permainan ini adalah untuk membuat gasing berputar sekuat mungkin dan bertahan sebanyak mungkin. Gasing adalah permainan yang sering dimainkan di Riau dan memiliki turnamen resmi.
123	Wilayah dan Kerajaan	5	Kerajaan Kampar: Kerajaan Kampar terletak di wilayah kabupaten Kampar, Riau. Ini adalah salah satu kerajaan yang cukup kuat pada masanya dan mengendalikan jalur perdagangan penting di wilayah tersebut.
...
166	Raja dan Tokoh	6	Raja Ali Haji: Raja Ali Haji adalah seorang ulama, sastrawan, dan cendekiawan terkemuka Melayu pada abad ke-19. Ia terkenal karena karyanya yang penting dalam sastra Melayu, termasuk "Gurindam Dua Belas."

4.2 Preprocessing

Tahap preprocessing text merupakan langkah awal yang sangat penting dalam proses text mining untuk memastikan data teks berada dalam bentuk yang bersih, terstruktur, dan siap diolah secara komputasional. Proses ini meliputi beberapa tahapan seperti cleaning, tokenizing, filtering, dan stemming yang bertujuan untuk menghilangkan noise serta menyederhanakan kata ke bentuk dasarnya. Dengan menerapkan preprocessing pada teks bertema kebudayaan Melayu, seperti contoh kalimat tentang Kerajaan Siak Sri Indrapura, sistem dapat mengenali pola linguistik dengan lebih akurat. Hasil akhir dari tahap ini menjadi fondasi penting bagi analisis lanjutan seperti klasifikasi, klusterisasi, dan ekstraksi informasi, secara detail dapat ditunjukkan pada Tabel 2.

Tabel 2. Preprocessing Text Dataset

Tahapan	Hasil
Data Asli	Siak Sri Indrapura adalah sebuah kerajaan di Riau yang memiliki banyak cerita rakyat. Salah satu legenda terkenal menceritakan tentang pendirian kerajaan ini oleh Raja Kecil, seorang pemuda yang bijaksana dan berani.
Cleaning	siak sri indrapura adalah sebuah kerajaan di riau yang memiliki banyak cerita rakyat salah satu legenda terkenal menceritakan tentang pendirian kerajaan ini oleh raja kecil seorang pemuda yang bijaksana dan berani
Tokenizing	[siak, sri, indrapura, adalah, sebuah, kerajaan, di, riau, yang, memiliki, banyak, cerita, rakyat, salah, satu, legenda, terkenal, menceritakan, tentang, pendirian, kerajaan, ini, oleh, raja, kecil, seorang, pemuda, yang, bijaksana, dan, berani]
Filtering	[siak, sri, indrapura, kerajaan, riau, cerita, rakyat, legenda, terkenal, pendirian, kerajaan, raja, kecil, pemuda, bijaksana, berani]
Stemming	[siak, sri, indrapura, raja, kecil, kerajaan, riau, cerita, rakyat, legenda, terkenal, diri, muda, bijak, berani]

4.3 TF-IDF

Untuk menggambarkan proses pembobotan kata pada data text Kebudayaan Melayu di Indonesia, dilakukan perhitungan menggunakan metode Term Frequency–Inverse Document Frequency (TF-IDF). Metode ini berfungsi untuk menilai tingkat kepentingan suatu kata dalam satu dokumen relatif terhadap keseluruhan kumpulan dokumen. Semakin tinggi nilai TF-IDF, semakin besar peran kata tersebut dalam merepresentasikan isi dokumen. Berikut contoh Tabel hasil perhitungan TF-IDF yang memuat kata-kata bertema kebudayaan Melayu berawalan huruf A, seperti adat, aksara, dan agama, hingga zapin pada sepuluh dokumen teks yang dikumpulkan dari berbagai sumber web. Hasil penghitungan TF-IDF dapat ditunjukkan pada Tabel 3.

Tabel 3. TF-IDF Kebudayaan Melayu di Indonesia

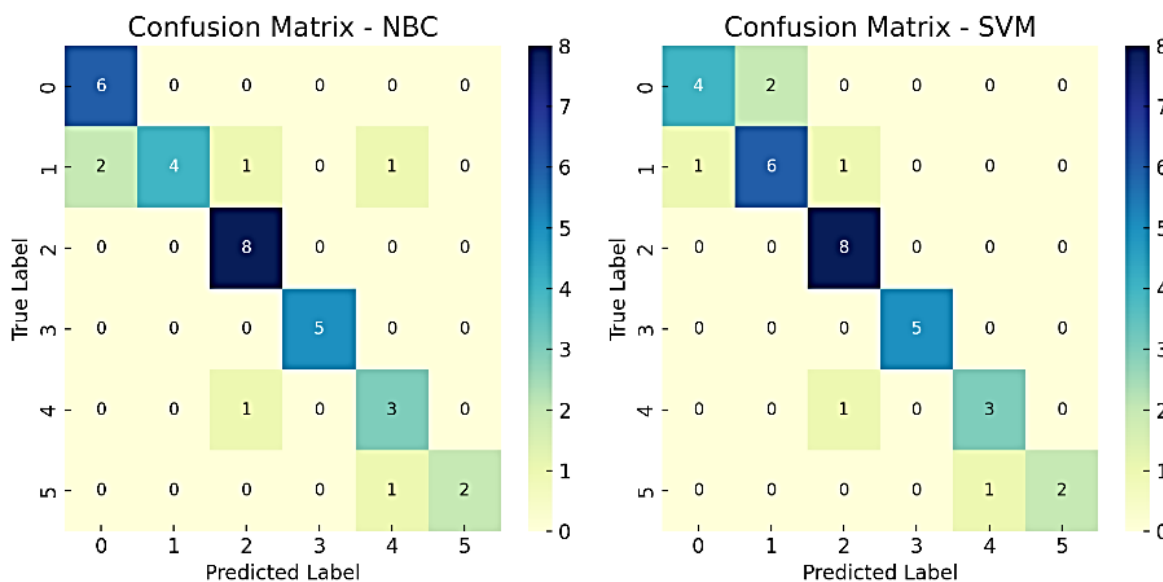
Dokumen	adat	aksara	alat	alam	arsitektur	agama	aksentuasi	...	zapin
D1	0.42	0.00	0.18	0.27	0.10	0.33	0.00	...	0.20
D2	0.36	0.25	0.00	0.31	0.28	0.14	0.00	...	0.22
D3	0.00	0.41	0.33	0.12	0.00	0.27	0.10	...	0.16
D4	0.29	0.32	0.00	0.24	0.15	0.20	0.11	...	0.00
D5	0.40	0.28	0.22	0.00	0.18	0.25	0.00	...	0.09
D6	0.00	0.19	0.35	0.29	0.21	0.00	0.14	...	0.11
D7	0.38	0.00	0.27	0.33	0.24	0.19	0.00	...	0.20
D8	0.31	0.26	0.00	0.20	0.19	0.23	0.13	...	0.18
D9	0.00	0.36	0.29	0.15	0.00	0.28	0.11	...	0.10
...
D164	0.27	0.22	0.00	0.30	0.16	0.24	0.12	...	0.19

4.2 Confusion Matrix

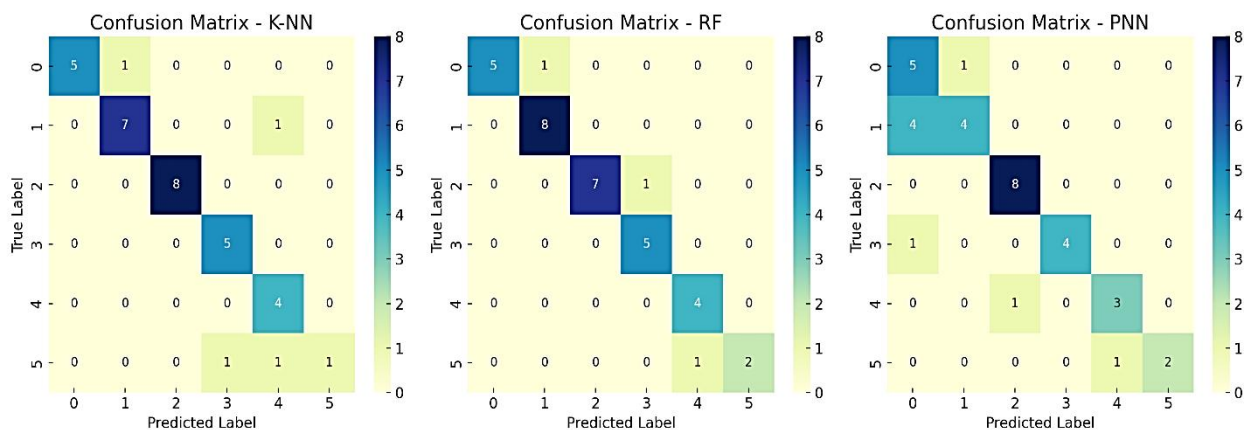
Confusion matrix digunakan untuk mengevaluasi kinerja model klasifikasi dengan membandingkan hasil prediksi model terhadap label asli. Ini adalah Tabel yang menunjukkan jumlah True Positives (TP), True Negatives (TN), False Positives (FP), dan False Negatives (FN). Confusion Matrix dari kelima algoritma yang dilakukan dapat dilihat pada Gambar 2.

Hasil visualisasi heatmap pada Gambar 2 menunjukkan bahwa Random Forest (RF) memiliki pola diagonal paling dominan dibanding algoritma lain, menandakan tingkat akurasi dan konsistensi klasifikasi tertinggi. RF mampu memprediksi label dengan benar pada sebagian besar data, sedangkan SVM dan Naïve Bayes (NBC) masih menunjukkan kesalahan klasifikasi antar kelas. K-NN menampilkan performa yang cukup baik, namun sensitif terhadap kemiripan antar data, sementara PNN menghasilkan distribusi prediksi yang tidak stabil. Secara umum, RF terbukti paling unggul dalam mengenali pola teks kebudayaan Melayu dari berbagai dokumen.

Secara analitis, keunggulan Random Forest disebabkan oleh mekanisme ensemble learning yang menggabungkan banyak pohon keputusan sehingga mengurangi overfitting dan meningkatkan generalisasi model. Pola diagonal yang kuat pada heatmap RF memperlihatkan kestabilan model dalam mengidentifikasi fitur penting dari data teks. Sementara itu, algoritma lain seperti PNN dan NBC lebih sensitif terhadap ketidakseimbangan data dan variasi kata, menyebabkan akurasi lebih rendah dalam klasifikasi teks kebudayaan.



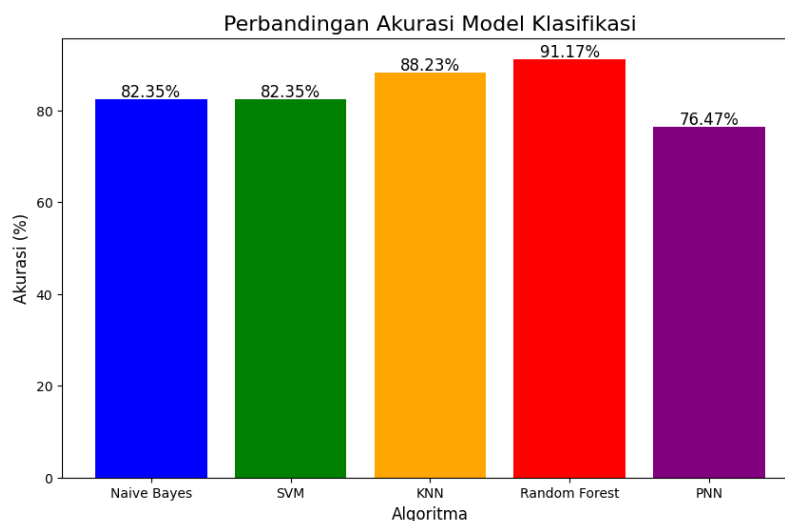
Gambar 2. (a) Confusion Matrix Perbandingan Algoritma



Gambar 2. (b) Confusion Matrix Perbandingan Algoritma

4.3 Perbandingan Akurasi

Setelah pemodelan selesai dilakukan menggunakan algoritma Naïve Bayes, K-NN, SVM, Random Forest, dan PNN dilakukan visualisasi perbandingan untuk melihat model mana yang menghasilkan akurasi tertinggi. Perbandingan akurasi model klasifikasi dapat dilihat pada Gambar 3.



Gambar 3. Perbandingan Akurasi Model Klasifikasi

Berdasarkan Gambar 3, terlihat bahwa Random Forest (RF) merupakan algoritma dengan tingkat akurasi tertinggi yaitu 91,17%, menunjukkan bahwa model ini paling efektif dalam mengklasifikasikan data teks kebudayaan Melayu di Indonesia. Hal ini menandakan bahwa RF mampu menangani kompleksitas dan variasi data dengan baik melalui mekanisme ensemble learning yang menggabungkan banyak pohon keputusan untuk menghasilkan hasil yang lebih stabil dan akurat. Di posisi kedua terdapat K-Nearest Neighbor (KNN) dengan akurasi 88,23%, yang cukup baik meskipun cenderung sensitif terhadap jumlah tetangga dan kedekatan data antar kelas. Sementara itu, Naïve Bayes (NBC) dan Support Vector Machine (SVM) memiliki akurasi sama yaitu 82,35%, menunjukkan performa yang cukup stabil namun belum optimal dalam mengenali fitur semantik teks yang kompleks.

Adapun algoritma dengan akurasi terendah adalah Probabilistic Neural Network (PNN) dengan nilai 76,47%, menandakan bahwa model ini kurang mampu beradaptasi terhadap variasi kosakata dan konteks pada data teks yang digunakan. Perbedaan akurasi antar algoritma juga menunjukkan bahwa karakteristik data teks kebudayaan cenderung lebih cocok diproses dengan metode berbasis pohon keputusan seperti Random Forest yang mampu menangkap hubungan non-linear antar fitur. Secara keseluruhan, grafik ini memperlihatkan bahwa RF merupakan model paling andal dan konsisten untuk implementasi klasifikasi teks berbasis text mining, sedangkan model lainnya memerlukan optimasi tambahan seperti parameter tuning atau feature selection agar performanya dapat meningkat.

4.4 Use Case Diagram

Use case diagram berguna untuk menggambarkan kegiatan-kegiatan yang dapat dilakukan oleh pengguna terhadap aplikasi yang sedang berjalan. Penggambaran use case terlihat pada Gambar 4. Deskripsi dari masing masing use case yang berada dalam aplikasi tersebut, dapat dilihat pada Tabel 3.



Gambar 4. Use Case Diagram

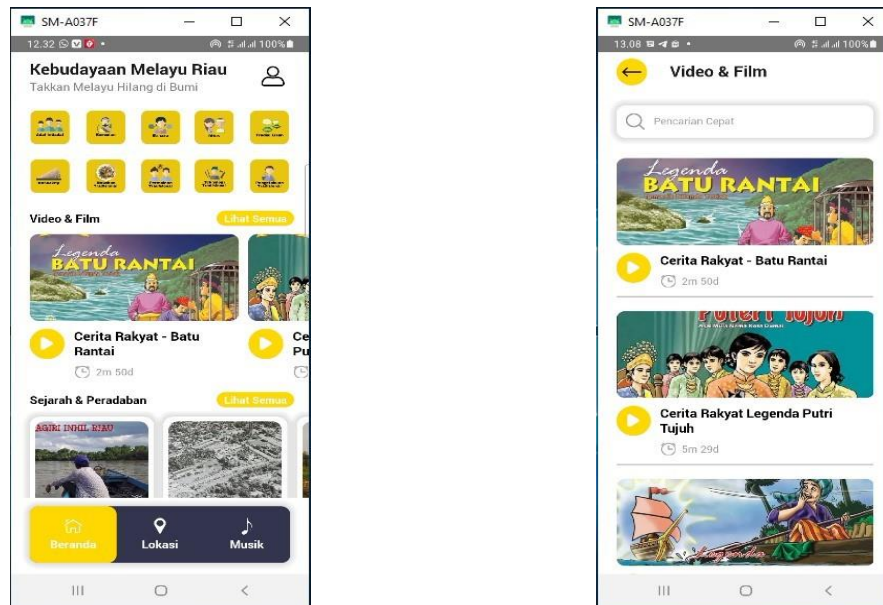
Tabel 3. Deskripsi Use Case

ID	Usecase	Deskripsi
UC-01	Homepage	Usecase ini mengGambarkan tampilan halaman utama pada aplikasi. Setiap pengguna dapat melihat halaman utama dari aplikasi dengan login terlebih dahulu
UC-02	Melihat Informasi Adat Istiadat	Usecase ini mengGambarkan tampilan halaman adat istiadat, pada halaman ini user dapat melihat informasi adat istiadat Melayu Riau.
UC-03	Melihat Informasi Tradisi Lisan	Usecase ini mengGambarkan tampilan halaman tradisi lisan, pada halaman ini user dapat melihat informasi tradisi lisan Melayu Riau.
UC-04	Melihat Informasi Manuskrip	Usecase ini mengGambarkan tampilan halaman manuskrip, pada halaman ini user dapat melihat informasi manuskrip Melayu Riau.
UC-05	Melihat Informasi Seni	Usecase ini mengGambarkan tampilan halaman seni, pada halaman ini user dapat melihat informasi seni Melayu Riau.
UC-06	Melihat Informasi Ritus	Usecase ini mengGambarkan tampilan halaman ritus, pada halaman ini user dapat melihat informasi ritus Melayu Riau.
UC-07	Melihat Informasi Bahasa	Usecase ini mengGambarkan tampilan halaman Bahasa, pada halaman ini user dapat melihat informasi Bahasa Melayu Riau.
UC-08	Melihat Informasi Teknologi Tradisional	Usecase ini mengGambarkan tampilan halaman teknologi tradisional, pada halaman ini user dapat melihat informasi teknologi tradisional Melayu Riau.
UC-09	Melihat Informasi Makanan Tradisional	Usecase ini mengGambarkan tampilan halaman makanan tradisional, pada halaman ini user dapat melihat informasi makanan tradisional Melayu Riau.
UC-10	Melihat Informasi Olahraga Tradisional	Usecase ini mengGambarkan tampilan halaman olahraga tradisional, pada halaman ini user dapat melihat informasi olahraga tradisional Melayu Riau.
UC-11	Melihat Informasi Pengetahuan Tradisional	Usecase ini mengGambarkan tampilan halaman pengetahuan tradisional, pada halaman ini user dapat melihat informasi pengetahuan tradisional Melayu Riau.

4.5 Tampilan Sistem

Aplikasi Kebudayaan Melayu Riau berbasis Android ini dirancang untuk memudahkan masyarakat mengenal kebudayaan Melayu Riau secara digital dan interaktif. Dengan antarmuka yang sederhana, aplikasi ini menampilkan berbagai kategori budaya, seperti adat istiadat, tradisi lisan, manuskrip, seni, ritus, bahasa, teknologi tradisional, makanan, olahraga, dan pengetahuan tradisional. Aplikasi ini juga menyediakan konten multimedia dalam bentuk video dan musik tradisional yang dapat diakses melalui menu navigasi, sehingga pengguna dapat memahami dan

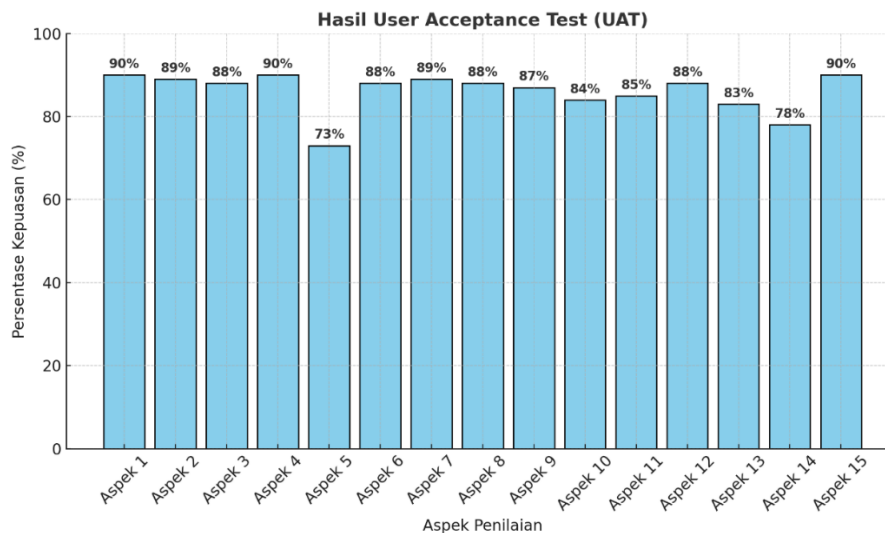
merasakan budaya Melayu secara lebih mendalam. Selain menyajikan informasi budaya yang komprehensif, aplikasi ini diharapkan berperan dalam melestarikan budaya Melayu Riau agar tetap dikenal dan diapresiasi oleh masyarakat luas. Tampilan system Aplikasi Kebudayaan Melayu Riau berbasis Android dapat dilihat pada Gambar 5.



Gambar 5. Tampilan Sistem

4.6 Pengujian Aplikasi dan Pelatihan Kepada Pengguna

Bagian akhir dari implementasi aplikasi adalah pengujian. Pengujian dilakukan dengan menggunakan 2 jenis uji yaitu Blackbox dan User Acceptance Test (UAT). Pengujian UAT merupakan salah satu pengujian sistem yang langsung diuji oleh pengguna, hal ini bertujuan untuk mengetahui apakah sistem yang dibangun dapat menjadi solusi bagi sasaran sistem, dan dapat mempermudah pekerjaan sebelumnya. Pengujian UAT adalah suatu pengujian oleh pengguna untuk menghasilkan dokumen yang dijadikan bukti bahwa sistem yang dikembangkan dapat diterima atau tidak oleh pengguna. Apabila hasil pengujian dianggap memenuhi kebutuhan pengguna, maka aplikasi dapat diterapkan. Pengujian dengan UAT dilakukan dengan mengajukan beberapa pertanyaan terhadap Pegawai Dinas Kebudayaan Provinsi Riau, Budayawan Riau, dan mahasiswa yang bertindak sebagai pengguna. Pengujian ini melibatkan 5 pegawai Dinas Kebudayaan Provinsi Riau, 5 Budayawan Riau, 5 pelaku seni budaya Melayu, dan 5 mahasiswa sebagai pengguna. Hasil UAT dinilai dengan 5 kategori, yaitu SS (Sangat Sesuai), S (Sesuai), KS (Kurang Sesuai), TS (Tidak Sesuai), dan STS (Sangat Tidak Sesuai). Hasil perhitungan UAT dapat dilihat pada Gambar 6.



Gambar 6. Hasil Uji UAT

Gambar 6 menunjukkan hasil *User Acceptance Test (UAT)* terhadap 15 aspek penilaian sistem. Berdasarkan grafik tersebut, terlihat bahwa tingkat kepuasan pengguna berada pada kisaran 73% hingga 90%, dengan sebagian besar aspek menunjukkan nilai di atas 85%. Aspek dengan nilai tertinggi mencapai 90%, menunjukkan bahwa mayoritas pengguna menilai sistem sudah sangat baik dari sisi fungsi, tampilan, dan kemudahan penggunaan.

Sementara itu, aspek dengan nilai 73% menunjukkan area yang perlu diperbaiki, kemungkinan terkait respons sistem atau fitur tambahan yang belum sepenuhnya optimal.

Secara keseluruhan, hasil ini mengindikasikan bahwa sistem yang dikembangkan telah diterima dengan baik oleh pengguna dengan rata-rata kepuasan mencapai 86%, yang termasuk dalam kategori sangat layak digunakan. Nilai ini juga memperkuat hasil pengujian fungsional (blackbox) yang menunjukkan tingkat keberhasilan 100%, sehingga sistem dapat dinyatakan berjalan stabil dan siap diimplementasikan lebih luas untuk mendukung pembelajaran kebudayaan Melayu di Indonesia secara digital. Hal ini menyatakan bahwa aplikasi pengenalan kebudayaan Melayu Riau berbasis Android dapat digunakan dengan baik oleh user.

4.7 Pembahasan

Hasil penelitian ini menunjukkan bahwa penerapan beberapa algoritma supervised learning pada klasifikasi teks kebudayaan Melayu memberikan variasi performa yang signifikan. Algoritma Random Forest (RF) menempati posisi tertinggi dengan akurasi 91,17%, mengungguli KNN, SVM, NBC, dan PNN. Keunggulan RF secara ilmiah disebabkan oleh sifatnya sebagai metode ensemble yang menggabungkan beberapa pohon keputusan untuk memperkuat generalisasi model serta mengurangi risiko overfitting [32]. Temuan ini sejalan dengan penelitian Nasir Jalal dkk. yang menyatakan bahwa RF memiliki ketahanan tinggi terhadap data kompleks dan bervariasi dalam klasifikasi teks [15]. Dengan kemampuan tersebut, RF terbukti lebih adaptif terhadap data teks kebudayaan Melayu yang memiliki konteks linguistik kaya dan heterogen. Hal ini menegaskan bahwa RF adalah algoritma yang sesuai untuk menangani data teks berbahasa alami dalam konteks budaya daerah di Indonesia.

Selain itu, K-Nearest Neighbor (KNN) menunjukkan performa cukup baik dengan akurasi 88,23%, menandakan bahwa metode berbasis jarak masih relevan dalam klasifikasi teks berukuran terbatas. Namun, sensitivitas KNN terhadap data berdekatan dapat menurunkan akurasi jika terdapat kemiripan konteks antar dokumen [33]. Sementara SVM dan Naïve Bayes Classifier (NBC) memiliki akurasi yang sama yaitu 82,35%, menandakan bahwa keduanya masih mampu memberikan hasil cukup stabil dalam kondisi data yang relatif kecil. Penelitian oleh Xiaofang Zheng dan Richard Schweickert juga menegaskan bahwa SVM unggul pada data biner namun cenderung kurang optimal dalam multiclass classification tanpa penyesuaian kernel yang tepat [12]. Adapun Probabilistic Neural Network (PNN) memiliki akurasi terendah (76,47%), kemungkinan akibat jumlah data terbatas yang menyebabkan jaringan tidak dapat melakukan generalization dengan baik [34].

Secara keseluruhan, hasil penelitian memperlihatkan bahwa penggunaan pendekatan text mining berbasis machine learning efektif dalam mendukung pelestarian kebudayaan Melayu melalui sistem pembelajaran digital. Model yang dikembangkan dalam aplikasi Digital Learning System (DLS) tidak hanya berhasil meningkatkan akurasi klasifikasi teks, tetapi juga mendapatkan penilaian tinggi dari pengguna dengan skor UAT 86% dan pengujian blackbox 100%. Hal ini membuktikan bahwa integrasi machine learning dengan sistem edukasi berbasis budaya dapat meningkatkan efektivitas pembelajaran berbasis digital [35]. Ke depan, penelitian lanjutan disarankan menggunakan dataset yang lebih besar dan beragam, serta menambahkan algoritma berbasis deep learning seperti CNN atau BERT untuk meningkatkan performa dan kemampuan generalisasi model pada teks kebudayaan yang lebih kompleks.

4. KESIMPULAN

Berdasarkan hasil penelitian, algoritma Random Forest (RF) memberikan kinerja terbaik dalam pemodelan dengan akurasi sebesar 91,17%, diikuti oleh K-Nearest Neighbor (KNN) dengan akurasi 88,23%, sedangkan Support Vector Machine (SVM) dan Naïve Bayes Classifier (NBC) memiliki akurasi yang sama yaitu 82,35%, serta Probabilistic Neural Network (PNN) memiliki akurasi terendah sebesar 76,47% dengan menggunakan dataset dokumen text kebudayaan melayu. Selain itu, pengembangan Digital Learning System (DLS) dengan konsep end-user information one-way berhasil diterapkan dengan baik, terbukti dari nilai User Acceptance Test (UAT) sebesar 86% dan pengujian Blackbox yang menunjukkan keberhasilan 100%. Sistem ini juga berhasil diuji pada berbagai perangkat, menunjukkan performa yang stabil dan berjalan lancar. Algoritma Random Forest (RF) memberikan akurasi tertinggi karena kemampuannya dalam menggabungkan hasil dari banyak pohon keputusan (decision trees), sehingga mampu mengurangi risiko overfitting dan meningkatkan kemampuan generalisasi model. Secara ilmiah, RF bekerja dengan prinsip ensemble learning, yaitu menggabungkan prediksi dari berbagai model lemah untuk menghasilkan keputusan yang lebih stabil dan akurat. Mekanisme bagging pada RF memungkinkan variasi data pelatihan, sehingga model dapat menangkap pola kompleks dan non-linear pada teks. Hal ini menjadikan RF lebih efektif dibanding algoritma tunggal seperti SVM, KNN, NBC, dan PNN dalam klasifikasi teks kebudayaan. Kedepan sebagai evaluasi dan riset lanjutan, penelitian ini dapat menggunakan dataset yang lebih besar, total data 156 untuk 6 kelas klasifikasi (multiclass) memiliki keterbatasan sebagai penyebab utama akurasi dan keterbatasan generalisasi model.

REFERENCES

- [1] M. P. Sari and A. R. Hidayatulloh, "Pengenalan Kebudayaan Indonesia melalui Fotografi pada Akun Instagram 'KWODOKIJO,'" *Edsence J. Pendidik. Multimed.*, vol. 2, no. 2, pp. 111–120, 2020, doi: 10.17509/edsence.v2i2.27460.
- [2] N. D. Budi Setyaningrum, "Local Culture in the Global Era," *Eksprei Seni*, vol. 20, no. 2, p. 102, 2018.



- [3] M. J. Saputra and N. Hamdi, “Rancang Bangun Aplikasi Sejarah Kebudayaan Aceh Berbasis Android Studi Kasus Dinas Kebudayaan Dan Pariwisata Aceh,” *J. Informatics Comput. Sci.*, vol. 5, no. 2, pp. 147–158, 2019, [Online]. Available: <http://www.jurnal.uui.ac.id/index.php/jics/article/view/555>
- [4] A. Suryadi, N. M. Rosa, and E. Subandriyo, “Perancangan Aplikasi Pengenalan Suku Dan Kebudayaan Berbasis Android,” *Semin. Nas. Ris. dan Teknol. (SEMNAS RISTEK)*, vol. 4, no. 1, pp. 186–192, 2020, [Online]. Available: <http://www.proceeding.unindra.ac.id/index.php/semnasristek/article/view/2497>
- [5] M. Mustakim and F. N. Salisah, “Density-Based Spatial Clustering, K-Means and Frequent Pattern Growth for Clustering and Association of Malay Cultural Text Data in Indonesia”, *bits*, vol. 7, no. 1, pp. 884–895, Jun. 2025.
- [6] A. C. Benabdellah, A. Benghabrit, and I. Bouhaddou, “A survey of clustering algorithms for an industrial context,” *Procedia Comput. Sci.*, vol. 148, pp. 291–302, 2019, doi: 10.1016/j.procs.2019.01.022.
- [7] S. Reddy et al., “Use and validation of text mining and cluster algorithms to derive insights from Corona Virus Disease-2019 (COVID-19) medical literature,” *Comput. Methods Programs Biomed. Updat.*, vol. 1, no. February, p. 100010, 2021, doi: 10.1016/j.cmpbup.2021.100010.
- [8] S. Jun, S. S. Park, and D. S. Jang, “Document clustering method using dimension reduction and support vector clustering to overcome sparseness,” *Expert Syst. Appl.*, vol. 41, no. 7, pp. 3204–3212, 2014, doi: 10.1016/j.eswa.2013.11.018.
- [9] J. Rejito, A. Atthariq, and A. S. Abdullah, “Application of text mining employing k-means algorithms for clustering tweets of Tokopedia,” *J. Phys. Conf. Ser.*, vol. 1722, no. 1, 2021, doi: 10.1088/1742-6596/1722/1/012019.
- [10] Mustakim, M. Z. Fauzi, Mustafa, A. Abdullah, and Rohayati, “Clustering of Public Opinion on Natural Disasters in Indonesia Using DBSCAN and K-Medoids Algorithms,” *J. Phys. Conf. Ser.*, vol. 1783, no. 1, 2021, doi: 10.1088/1742-6596/1783/1/012016.
- [11] A. Patel, P. Oza, and S. Agrawal, “Sentiment Analysis of Customer Feedback and Reviews for Airline Services using Language Representation Model,” *Procedia Comput. Sci.*, vol. 218, pp. 2459–2467, 2022, doi: 10.1016/j.procs.2023.01.221.
- [12] V. Jackins, S. Vimal, M. Kaliappan, and M. Y. Lee, “AI-based smart prediction of clinical disease using random forest classifier and Naive Bayes,” *J. Supercomput.*, vol. 77, no. 5, pp. 5198–5219, 2021, doi: 10.1007/s11227-020-03481-x.
- [13] L. C. Lee, C. Y. Liong, and A. A. Jemain, “Validity of the best practice in splitting data for hold-out validation strategy as performed on the ink strokes in the context of forensic science,” *Microchem. J.*, vol. 139, no. 2017, pp. 125–133, 2018, doi: 10.1016/j.microc.2018.02.009.
- [14] I. Ho, H. N. Goh, and Y. F. Tan, “Preprocessing Impact on Sentiment Analysis Performance on Malay Social Media Text,” *J. Syst. Manag. Sci.*, vol. 12, no. 5, pp. 73–90, 2022, doi: 10.33168/JSMS.2022.0505.
- [15] N. Jalal, A. Mehmood, G. S. Choi, and I. Ashraf, “A novel improved random forest for text classification using feature ranking and optimal number of trees,” *J. King Saud Univ. - Comput. Inf. Sci.*, vol. 34, no. 6, pp. 2733–2742, 2022, doi: 10.1016/j.jksuci.2022.03.012.
- [16] S. G. Delis et al., “Hepatic resection for large hepatocellular carcinoma in the era of UCSF criteria,” *Hpb*, vol. 11, no. 7, pp. 551–558, 2009, doi: 10.1111/j.1477-2574.2009.00084.x.
- [17] Mustakim and R. Novita, “The Implementation of Probabilistic Neural Networks to Sentiment Analysis of National Principle and Religion Issues in Indonesia,” *J. Syst. Manag. Sci.*, vol. 13, no. 5, pp. 311–321, 2023, doi: 10.33168/JSMS.2023.0520.
- [18] J. L. Speiser, M. E. Miller, J. Tooze, and E. Ip, “A comparison of random forest variable selection methods for classification prediction modeling,” *Expert Syst. Appl.*, vol. 134, pp. 93–101, 2019, doi: 10.1016/j.eswa.2019.05.028.
- [19] M. Azhari, Z. Situmorang, and R. Rosnelly, “Perbandingan Akurasi, Recall, dan Presisi Klasifikasi pada Algoritma C4.5, Random Forest, SVM dan Naive Bayes,” *J. Media Inform. Budidarma*, vol. 5, no. 2, p. 640, 2021, doi: 10.30865/mib.v5i2.2937.
- [20] W. Apriliah, I. Kurniawan, M. Baydhowi, and T. Haryati, “Prediksi Kemungkinan Diabetes pada Tahap Awal Menggunakan Algoritma Klasifikasi Random Forest,” *Sistemasi*, vol. 10, no. 1, p. 163, 2021, doi: 10.32520/stmsi.v10i1.1129.
- [21] G. A. Sandag, “Prediksi Rating Aplikasi App Store Menggunakan Algoritma Random Forest,” *CogITo Smart J.*, vol. 6, no. 2, pp. 167–178, 2020, doi: 10.31154/cogito.v6i2.270.167-178.
- [22] J. Cervantes, F. Garcia-Lamont, L. Rodríguez-Mazahua, and A. Lopez, “A comprehensive survey on support vector machine classification: Applications, challenges and trends,” *Neurocomputing*, vol. 408, no. xxxx, pp. 189–215, 2020, doi: 10.1016/j.neucom.2019.10.118.
- [23] M. Muhathir, M. H. Santoso, and D. A. Larasati, “Wayang Image Classification Using SVM Method and GLCM Feature Extraction,” *J. Informatics Telecommun. Eng.*, vol. 4, no. 2, pp. 373–382, 2021, doi: 10.31289/jite.v4i2.4524.
- [24] D. Tuhenay and E. Mailoa, “Perbandingan Klasifikasi Bahasa Menggunakan Metode Naïve Bayes Classifier (Nbc) Dan Support Vector Machine (Svm) Comparison of Language Classification Using Naive Bayes Classifier (Nbc) and Support Vector Machine (Svm) Method,” *JIKO (Jurnal Inform. dan Komputer)*, vol. 4, no. 2, pp. 105–111, 2021, doi: 10.33387/jiko.
- [25] F. Paquin, J. Rivnay, A. Salleo, N. Stingelin, and C. Silva, “Multi-phase semicrystalline microstructures drive exciton dissociation in neat plastic semiconductors,” *J. Mater. Chem. C*, vol. 3, pp. 10715–10722, 2015, doi: 10.1039/b000000x.
- [26] H. Chen, S. Hu, R. Hua, and X. Zhao, “Improved naive Bayes classification algorithm for traffic risk management,” *EURASIP J. Adv. Signal Process.*, vol. 2021, no. 1, 2021, doi: 10.1186/s13634-021-00742-6.
- [27] A. R. Lubis, M. Lubis, and Al-Khowarizmi, “Optimization of distance formula in k-nearest neighbor method,” *Bull. Electr. Eng. Informatics*, vol. 9, no. 1, pp. 326–338, 2020, doi: 10.11591/eei.v9i1.1464.
- [28] D. Prasetyawan and R. Gatra, “Algoritma K-Nearest Neighbor untuk Memprediksi Prestasi Mahasiswa Berdasarkan Latar Belakang Pendidikan dan Ekonomi,” *JISKA (Jurnal Inform. Sunan Kalijaga)*, vol. 7, no. 1, pp. 56–67, 2022, doi: 10.14421/jiska.2022.7.1.56-67.
- [29] M. Laia, R. K. Hondro, and T. Zebua, “Implementasi Pengolahan Citra dengan Menggunakan Metode K-Nearest Neighbor Untuk Mengetahui Daging Ayam Busuk dan Daging Ayam Segar,” *JURIKOM (Jurnal Ris. Komputer)*, vol. 8, no. 2, pp. 39–49, 2021, doi: 10.30865/jurikom.v8i2.2818.
- [30] S. Ramadona, M. Diono, M. Susantok, and S. Ahdan, “Indoor location tracking pegawai berbasis Android menggunakan algoritma k-nearest neighbor,” *JITEL (Jurnal Ilm. Telekomun. Elektron. dan List. Tenaga)*, vol. 1, no. 1, pp. 51–58, 2021, doi: 10.35313/jitel.v1.i1.2021.51-58.



- [31] Y. Bodyanskiy, A. Deineko, I. Pliss, O. Chala, and A. Nortsova, "Matrix fuzzy-probabilistic neural network in image recognition task," Proc. 2020 IEEE 3rd Int. Conf. Data Stream Min. Process. DSMP 2020, pp. 33–36, 2020, doi: 10.1109/DSMP47368.2020.9204236.
- [32] L. Breiman, "Random Forests," Machine Learning, vol. 45, no. 1, pp. 5–32, 2001.
- [33] T. Cover and P. Hart, "Nearest neighbor pattern classification," IEEE Transactions on Information Theory, vol. 13, no. 1, pp. 21–27, 1967.
- [34] D. F. Specht, "Probabilistic Neural Networks for Classification, Mapping, or Associative Memory," IEEE Transactions on Neural Networks, vol. 1, no. 3, pp. 525–532, 1990.
- [35] A. Suryadi, M. Juanda, and Yulisman, "Integrasi Machine Learning dalam Sistem Pembelajaran Kebudayaan Digital," Jurnal Teknologi Informasi dan Pembelajaran, vol. 6, no. 2, pp. 115–124, 2024.