Jurnal SAINTIKOM (Jurnal Sains Manajemen Informatika dan Komputer)
Volume 24 ; Nomor 2 ; Agustus 2025 ; Page 193-202
E-ISSN : 2615-3475; P-ISSN : 1978-6603
https://ojs.trigunadharma.ac.id/index.php/jis/index

Implementasi *Explainable AI* dalam Klasifikasi Kekuatan *Password*

Ahmad Subhan Yazid [1], Yanuar Wicaksono[2], Eko Setiawan[3], Riski Nurhadi[4]

[1,4] Program Studi Informatika, Universitas Alma Ata, Yogyakarta, Indonesia
[2,3] Program Studi Sistem Informasi, Universitas Alma Ata, Yogyakarta, Indonesia
Email: [1]subhan@almaata.ac.id, [2]yanuar@almaata.ac.id, [3]eko@almaata.ac.id, [4]223200259@almaata.ac.id
Email Penulis Korespondensi: subhan@almaata.ac.id

Article History:
Received Jul 25th, 2025
Revised Aug 14th, 2025
Accepted Aug 30th, 2025

Abstrak

Keamanan kata sandi (*password*) merupakan aspek krusial dalam menjaga integritas sistem informasi digital. Namun, banyak pengguna masih menggunakan *password* yang lemah dan mudah ditebak, sehingga rentan terhadap serangan siber. Penelitian ini bertujuan untuk mengklasifikasikan kekuatan *password* dan menjelaskan hasil prediksi dengan pendekatan *Explainable Artificial Intelligence* (XAI), khususnya SHAP (*SHapley Additive exPlanations*) dan LIME (*Local Interpretable Model-agnostic Explanations*). Dataset yang digunakan berisi 100.000 *password* yang telah dilabeli dalam tiga kelas kekuatan *password*. Proses pra-pemrosesan mencakup ekstraksi fitur berbasis struktur karakteristik *password*, seperti panjang, huruf besar, angka, dan karakter spesial. Model Naive Bayes yang dibangun menunjukkan performa klasifikasi yang sangat baik dengan skor akurasi: 97,66%, Precision: 94,51%, Recall: 98,23%, dan F1-score: 96,22%. Selanjutnya, analisis XAI dilakukan untuk mengungkap kontribusi fitur terhadap keputusan model, baik secara global maupun lokal. Hasil visualisasi menggunakan SHAP dan LIME menunjukkan bahwa fitur panjang dan keberadaan karakter kapital memberikan pengaruh signifikan terhadap kekuatan *password*. Selain itu, juga dilakukan interpretasi terhadap prediksi lokal (analisis per-*password*) sehingga memberikan gambaran kontribusi fitur terhadap setiap kekuatan *password*. Sebagai implementasi akhir, sebuah antarmuka interaktif berbasis Streamlit dikembangkan untuk memungkinkan pengguna melakukan prediksi dan interpretasi kekuatan *password* secara *real-time*.

Kata Kunci : *password*, Naive Bayes, *explainable AI*, SHAP, LIME.

Abstract

Password security is a crucial aspect of maintaining the integrity of digital information systems. However, many users still use passwords that are weak and easy to guess, making them vulnerable to cyberattacks. This research aims to classify password strength and explain the prediction results with Explainable Artificial Intelligence(XAI) approaches, specifically SHAP (SHapley Additive exPlanations) and LIME (Local Interpretable Model-agnostic Explanations). The dataset used contains 100,000 passwords that have been labeled in three password strength classes. The pre-processing process includes structure-based feature extraction of password characteristics, such as length, capitalization, numbers, and special characters. The constructed Naive Bayes model shows excellent classification performance with accuracy score: 97.66%, Precision: 94.51%, Recall: 98.23%, and F1-score: 96,22%. Furthermore, XAI analysis was conducted to reveal the contribution of features to the model's decision, both globally and locally. Visualization results using SHAP and LIME show that the length feature and the presence of capital characters have a significant influence on password strength. In addition, local predictions (per-password analysis) were also interpreted to illustrate the contribution of features to each password strength. As a final implementation, a Streamlit-based interactive interface was developed to allow users to predict and interpret password strength in real-time.

Keyword : *password, Naive Bayes, explainable AI, SHAP, LIME.*

Jurnal SAINTIKOM (Jurnal Sains Manajemen Informatika dan Komputer)
Volume 24 ; Nomor 2 ; Agustus 2025 ; Page 193-202
E-ISSN : 2615-3475; P-ISSN : 1978-6603
https://ojs.trigunadharma.ac.id/index.php/jis/index

1. PENDAHULUAN

Kata sandi atau *password* merupakan salah satu mekanisme autentikasi tertua dan paling luas digunakan dalam sistem keamanan digital. *Password* digunakan untuk membatasi akses ke informasi atau sistem, memastikan bahwa hanya pihak yang berwenang yang dapat mengaksesnya [1]. Dalam praktiknya, *password* menjadi pengenal utama dalam berbagai layanan daring seperti email, media sosial, perbankan digital, hingga sistem informasi organisasi. Meskipun teknologi keamanan telah berkembang, seperti dengan hadirnya otentikasi dua faktor dan biometrik, *password* tetap menjadi lapisan pertahanan utama yang tidak tergantikan di banyak konteks [2].

Seiring dengan meningkatnya kompleksitas ancaman siber, perhatian terhadap kekuatan *password* pun semakin menguat. *Password* yang lemah atau mudah ditebak menjadi titik masuk favorit bagi pelaku kejahatan siber [3]. Hal ini mendorong berbagai pihak untuk mengembangkan penilaian kekuatan *password* secara otomatis, agar pengguna mendapatkan umpan balik saat membuat *password* baru. Penilaian kekuatan *password* secara konvensional biasanya dilakukan menggunakan aturan heuristik [4], misalnya *password* harus memiliki panjang minimal tertentu, mengandung huruf kapital, angka, dan karakter khusus. Namun, pendekatan ini tidak sepenuhnya terjamin keandalannya karena tidak mempertimbangkan konteks pola yang kompleks [5]. Sebagai contoh, *password* seperti "P@ssw0rd" mungkin lolos uji heuristik, tetapi tetap mudah ditebak. Oleh sebab itu, perkembangan penelitian mulai mengarah ke pendekatan berbasis pembelajaran mesin (*machine learning*) untuk memodelkan kekuatan *password* secara lebih cerdas berdasarkan karakteristik statistik atau fitur tertentu yang dapat dipelajari dari data.

Dalam klasifikasi kekuatan *password*, telah dilakukan berbagai studi untuk mengembangkan model yang dapat mengelompokkan *password* ke dalam kategori seperti lemah, sedang, atau kuat. Beberapa studi memanfaatkan teknik *supervised learning* dengan menggunakan dataset *password* yang telah diberi label kekuatannya [6], [7]. Fitur yang digunakan umumnya mencakup panjang *password*, keberadaan huruf kapital, angka, simbol, serta pola karakter lainnya. Model-model ini dilatih untuk mengenali pola dalam data, lalu digunakan untuk memberikan prediksi terhadap *password* baru yang diuji. Meski menunjukkan performa klasifikasi yang baik, pendekatan ini masih memiliki kelemahan berupa tidak adanya penjelasan yang transparan tentang bagaimana model mengambil keputusan. Kelemahan ini yang menjadi pembahasan utama dalam penelitian ini.

Keterbatasan interpretabilitas menjadi tantangan tersendiri, khususnya dalam konteks keamanan informasi yang sangat bergantung pada kepercayaan pengguna. Jika pengguna tidak memahami alasan mengapa *password*-nya dianggap lemah atau kuat oleh sistem, maka pengguna tidak dapat menentukan dengan pasti *password* seperti apa yang diperlukan. Untuk menjembatani kesenjangan ini, pendekatan *Explainable Artificial Intelligence* (XAI) mulai diperkenalkan. XAI adalah cabang dari kecerdasan buatan yang berupaya menjelaskan bagaimana sebuah model pembelajaran mesin menghasilkan prediksinya secara dapat dipahami manusia [8].

Dua metode XAI yang populer adalah SHAP (*SHapley Additive exPlanations*) dan LIME (*Local Interpretable Model-agnostic Explanations*). SHAP menggunakan teori nilai Shapley dari teori permainan untuk menghitung kontribusi masing-masing fitur dalam prediksi secara global dan lokal [9]. Sementara itu, LIME memberikan penjelasan lokal dengan membangun model linier sederhana di sekitar instance yang diuji [10], sehingga dapat menggambarkan perilaku lokal model utama. Kedua pendekatan ini telah banyak digunakan dalam berbagai domain seperti kesehatan [11], keuangan [12], industri [13], pendidikan [14], dan hukum [15]. Akan tetapi, penerapannya dalam konteks kekuatan *password* masih terbatas.

Penelitian ini bertujuan untuk mengembangkan sistem klasifikasi kekuatan *password* menggunakan algoritma Naive Bayes. Algoritma ini dipilih karena kesederhanaannya dalam penerapan klasifikasi berbasis probabilistik serta kemampuannya menangani fitur diskrit [16]. Sistem tersebut dilengkapi dengan fitur penjelasan menggunakan XAI untuk memperjelas kontribusi fitur terhadap setiap prediksi. Sistem ini tidak hanya memberikan prediksi kekuatan *password*, tetapi juga menyajikan visualisasi penjelasan yang menunjukkan alasan di balik keputusan tersebut. Dengan pendekatan ini, pengguna diharapkan tidak hanya mengetahui apakah *password* mereka kuat, tetapi juga memahami mengapa, sehingga dapat belajar membentuk *password* yang lebih aman ke depannya.

2. METODOLOGI PENELITIAN

Penelitian ini bertujuan untuk mengembangkan sistem klasifikasi kekuatan password berbasis pembelajaran mesin dengan dukungan Explainable AI (XAI) menggunakan SHAP dan LIME. Secara garis besar, metodologi yang diterapkan mencakup tahapan: 1) Akuisisi dan eksplorasi data, 2) Pra-pemrosesan dan ekstraksi fitur, 3) Pelatihan model klasifikasi menggunakan Naive Bayes, 4) Evaluasi dan interpretasi model dengan XAI, serta 5) Implementasi sistem melalui antarmuka Streamlit. Diagram alur metodologi dapat dilihat pada Gambar 1.

Gambar 1 menunjukkan Diagram alur metodologi penelitian terdiri atas lima tahap utama yang terhubung secara sekuensial. Dimulai dengan akuisisi dataset dari sumber terbuka Kaggle yang berisi data password dan kelas kekuatannya. Dataset ini kemudian disimpan dan dipersiapkan untuk proses analisis lebih lanjut. Selanjutnya, dilakukan pra-pemrosesan mencakup pengecekan adanyaa nilai kosong dan duplikat. Setelah itu, dilakukan ekstraksi fitur dari teks

password, seperti panjang karakter, keberadaan huruf kapital, angka, dan simbol. Data hasil pra-pemrosesan kemudian dibagi menjadi data latih dan data uji. Data latih dilatih dengan Algoritma Naive Bayes .

Gambar 1. Alur Penelitian

Evaluasi dilakukan berdasarkan akurasi serta metrik lain yang relevan untuk menilai performa model. Setelah model berhasil dibuat, tahap berikutnya adalah memberikan penjelasan terhadap hasil klasifikasi menggunakan dua teknik *Explainable AI*: SHAP dan LIME. Agar memudahkan akses oleh pengguna, sistem diimplementasikan dalam aplikasi berbasis web dengan Streamlit. Streamlit merupakan kerangka kerja (*framework*) sumber terbuka (*open-source*) berbasis Python yang dirancang untuk memudahkan pembuatan aplikasi web interaktif, terutama untuk bidang ilmu data dan pembelajaran mesin [17]. Pengguna dapat memasukkan password, melihat hasil klasifikasi secara langsung, serta menerima penjelasan visual terhadap keputusan model.

2.1 Akuisisi dan Eksplorasi Data

Dataset yang digunakan dalam penelitian ini bersumber dari platform Kaggle dengan judul "*Password Security: Sber Dataset*" [18]. Dataset tersebut terdiri atas 100.000 data *password* dengan label kekuatan (*strength*) yang diklasifikasikan ke dalam tiga kelas, yakni 0 = lemah, 1 = sedang, 2 = kuat. Tabel 1 menunjukkan lima baris pertama dataset yang dihasilkan dari fungsi *head()*. Setiap baris data memuat dua kolom utama: *password* (dalam bentuk teks) dan *strength*/kelas kekuatan (angka).

Tabel 1. Lima Baris Pertama Dataset Password

	Password	Strength
0	yrtzuab476	1
1	yEdnN9jc1NgzkkBP	2
2	sarita99	1
3	Suramerica2015	2
4	PPRbMvDIxMQ19TMo	2

2.2 Pra-pemrosesan dan Ekstraksi Fitur

Pra-pemrosesan mencakup penghapusan nilai kosong dan data duplikat. Selanjutnya, dilakukan proses ekstraksi fitur dari kolom *password* menjadi beberapa atribut numerik, antara lain:

a. *length*: panjang karakter dalam *password*,
b. *contains_uppercase*: indikator keberadaan huruf kapital,
c. *contains_specialcharacter*: indikator keberadaan karakter khusus,
d. *contains_number*: indikator keberadaan angka.

Seluruh fitur ini direpresentasikan dalam bentuk numerik (0 atau 1) untuk memudahkan pemrosesan oleh algoritma pembelajaran mesin. Eksplorasi awal juga dilakukan untuk mengamati distribusi kelas, panjang karakter, serta penggunaan karakter spesial dan angka yang digunakan. Visualisasi juga dilakukan untuk memahami dominasi kelas serta potensi ketidakseimbangan data [19].

2.3 Pelatihan Model Klasifikasi

Algoritma Naive Bayes dipilih karena kesederhanaannya serta kemampuannya untuk menangani data dengan fitur yang bersifat independen. Model dilatih menggunakan 80% data, sementara 20% sisanya digunakan untuk pengujian. Proses pelatihan dilakukan dengan pendekatan *supervised learning*, di mana model belajar mengenali pola pada fitur-fitur input untuk memprediksi kelas kekuatan *password*. Hasil prediksi dibandingkan dengan label aktual untuk menghitung akurasi dan metrik evaluasi lainnya seperti *precision* dan *recall* melalu bantuan tabel *confusion matrix*.

2.4 Interpretasi Model dengan Explainable AI

Untuk memahami kontribusi masing-masing fitur terhadap hasil klasifikasi, dua pendekatan XAI diterapkan:

1. SHAP (SHapley Additive exPlanations): digunakan untuk menghasilkan nilai kontribusi tiap fitur terhadap prediksi, baik secara global (untuk keseluruhan data) maupun lokal (untuk *password* tertentu).
2. LIME (Local Interpretable Model-agnostic Explanations): digunakan untuk menganalisis pengaruh fitur terhadap prediksi model secara lokal dengan membangun model sederhana di sekitar prediksi spesifik.

Hasil interpretasi ditampilkan dalam bentuk visualisasi untuk memudahkan pemahaman oleh pengguna akhir.

2.5 Implementasi Sistem dalam Streamlit

Sebagai bentuk penerapan praktis, sistem dikembangkan dalam bentuk aplikasi berbasis web menggunakan framework Streamlit. Aplikasi memungkinkan pengguna untuk:

a. memasukkan *password* secara langsung,
b. melihat prediksi kelas kekuatan *password*,
c. serta mendapatkan penjelasan visual dari SHAP dan LIME.

Dengan pendekatan ini, sistem tidak hanya mampu memberikan hasil klasifikasi, tetapi juga memberikan wawasan yang dapat dimengerti tentang alasan di balik prediksi yang diberikan.

3. HASIL DAN PEMBAHASAN

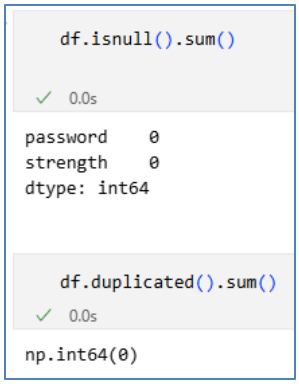
3.1 Analisis Dataset dan Pra-pemrosesan

Dataset yang digunakan dalam penelitian ini terdiri atas 100.000 baris data yang masing-masing memuat dua atribut utama, yaitu *password* dan *strength*. Kolom *password* berisi string kata sandi yang beragam dari segi panjang, karakteristik huruf, angka, serta simbol. Sementara itu, kolom *strength* merupakan label kelas yang menunjukkan tingkat kekuatan *password*, dikategorikan menjadi tiga kelas yaitu 0 (lemah), 1 (sedang), dan 2 (kuat). Dari analisis statistik awal, *password* yang berlabel 0 berjumlah 13428, berlabel 1 berjumlah 74278, dan yang berlabel 2 berjumlah 12294.

Sebelum dilakukan proses klasifikasi, data perlu melalui tahapan pra-pemrosesan. Tahapan ini bertujuan untuk menyiapkan data dalam bentuk yang dapat diterima oleh algoritma pembelajaran mesin. Adapun tahapan-tahapan yang dilakukan meliputi:

3.1.1 Pembersihan Data dan Ekstraksi Fitur

Data dianalisis untuk mendeteksi keberadaan nilai kosong (*missing values*) atau entri duplikat. Berdasarkan hasil eksplorasi awal, tidak ditemukan entri kosong pada kolom *password* maupun kolom *Strength* (gambar 2).



Gambar 2. Hasil Pengecekan Nilai Kosong dan Duplikat

Karena data *password* merupakan teks mentah, maka diperlukan ekstraksi fitur numerik agar dapat digunakan sebagai input model klasifikasi. Fitur-fitur yang diekstraksi dari setiap *password* meliputi:

a. Panjang *password* (*length*)
b. Keberadaan huruf kapital (*contains_uppercase*)
c. Keberadaan karakter spesial (*contains_specialcharacter*)
d. Keberadaan angka (*contains_number*)

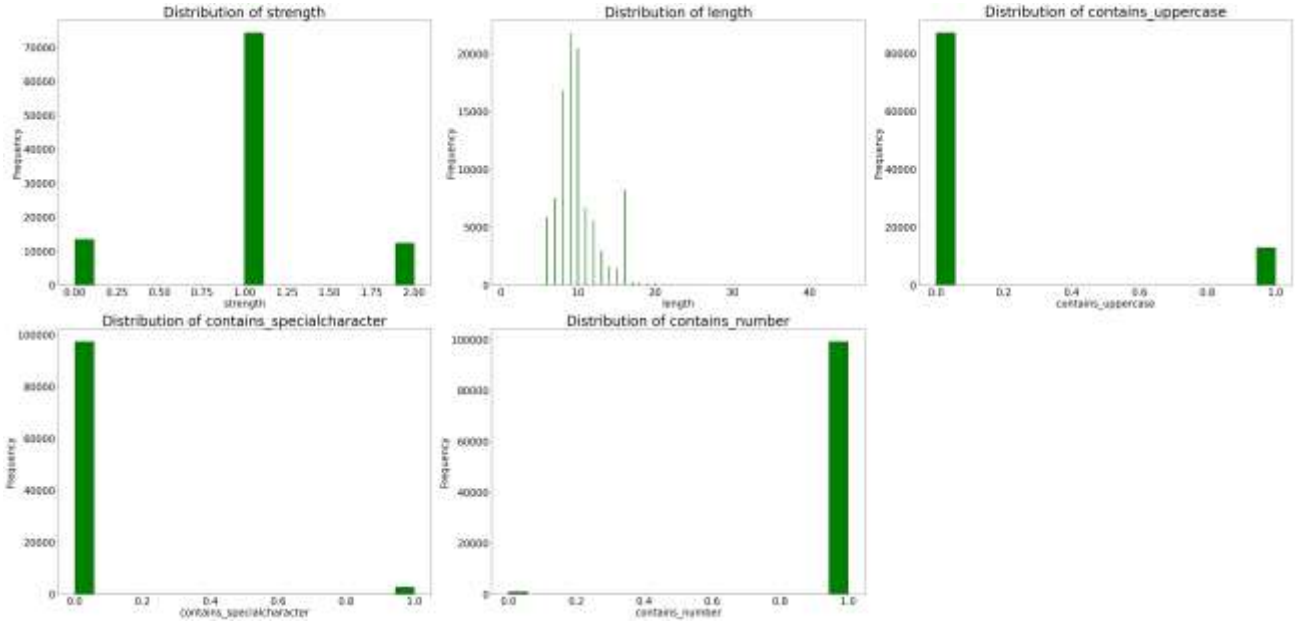
Setiap fitur ini dihitung menggunakan teknik pemrosesan string dan direpresentasikan dalam bentuk boolean (*True*, *False*) yang kemudian diubah menjadi nilai biner (1,0). Langkah ini membuat jumlah kolom bertambah dari yang awalnya terdiri dari 2 kolom menjadi 6 kolom (Tabel 2).

Jurnal SAINTIKOM (Jurnal Sains Manajemen Informatika dan Komputer)
Volume 24 ; Nomor 2 ; Agustus 2025 ; Page 193-202
E-ISSN : 2615-3475; P-ISSN : 1978-6603
https://ojs.trigunadharma.ac.id/index.php/jis/index

Tabel 2. Hasil Ekstraksi Fitur Dataset

Password	Strength	Length	Contains_uppercase	Contains_specialcharacter	Contains_number
yrtzuab476	1	10	0	0	1
yEdnN9jc1NgzkkBP	2	16	1	0	1
sarita99	1	8	0	0	1
Suramerica2015	2	14	1	0	1
PPRbMvDIxMQ19TMo	2	16	1	0	1

Setelah proses ekstraksi fitur, didapatkan data *password* yang mengandung huruf kapital (*contains-uppercase*) sebanyak 12869 data, sementara yang tidak mengandung huruf kapital sebanyak 87131. Adapun *password* yang mengandung karakter spesial (*contains_spesialcharacter*) berjumlah 2677 dan yang tidak mengandung karakter spesial sejumlah 97323. Selain itu, ektraksi fitur juga mendapati bahwa terdapat 99130 *password* yang mengandung angka (*contains_number*) dan 870 *password* yang tidak mengandung angka. Untuk mempertajam intuisi, visualisasi histogram dari kolom numerik dilakukan untuk melihat persebaran data sebagaimana ditampilkan pada gambar 3.



Gambar 3. Distribusi Data Numerik dari Dataset Hasil Ekstraksi Fitur

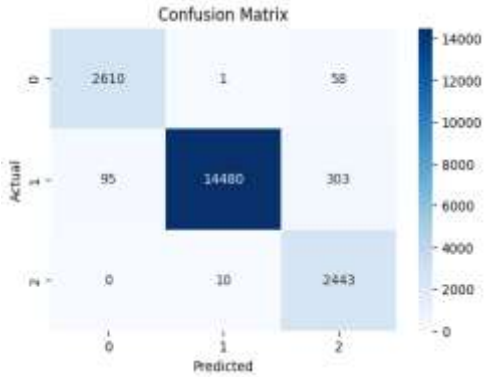
3.1.2 Pemisahan Data

Setelah fitur diekstrak, dataset dibagi menjadi data pelatihan (*train*) dan data pengujian (*test*) dengan rasio 80:20 menggunakan *fungsi train_test_split()*. Hal ini dimaksudkan agar model dapat belajar dari mayoritas data, namun tetap diuji akurasinya terhadap data yang belum pernah dilihat sebelumnya. Dengan tahapan pra-pemrosesan ini, data yang semula berupa teks mentah telah dikonversi menjadi format numerik terstruktur yang siap digunakan dalam tahap klasifikasi dan interpretasi.

3.2 Kinerja Model Klasifikasi Naive Bayes

Data pelatihan berjumlah 80.000 yang sebelumnya telah dibagi kemudian dilatih menggunakan algoritma *Naive Bayes*. Dari proses pelatihan ini, dihasilkan sebuah model klasifikasi yang digunakan dalam prediksi kekuatan *password*. Selanjutnya, pengujian dilakukan untuk menilai performa model. Metrik *confusion matrix* digunakan untuk mengevaluasi kesalahan klasifikasi antar kelas (gambar 4). Dari *confusion matrix* dapat dilihat bahwa sebagian besar kesalahan terjadi antara kelas 1 (sedang) dan kelas 2 (kuat), yang secara semantik memang memiliki batas yang lebih kabur dibanding kelas 0 (lemah). Selain *confusion matrix*, didapatkan pula metrik-metrik umum evaluasi model klasifikasi, yaitu *accuracy*, *precision*, *recall*, dan *f1-score*.

Jurnal SAINTIKOM (Jurnal Sains Manajemen Informatika dan Komputer)
Volume 24 ; Nomor 2 ; Agustus 2025 ; Page 193-202
E-ISSN : 2615-3475; P-ISSN : 1978-6603
https://ojs.trigunadharma.ac.id/index.php/jis/index



Gambar 4. *Confusion Matrix* Hasil Pengujian Model

Hasil evaluasi model pada data uji menunjukkan sebagai berikut:

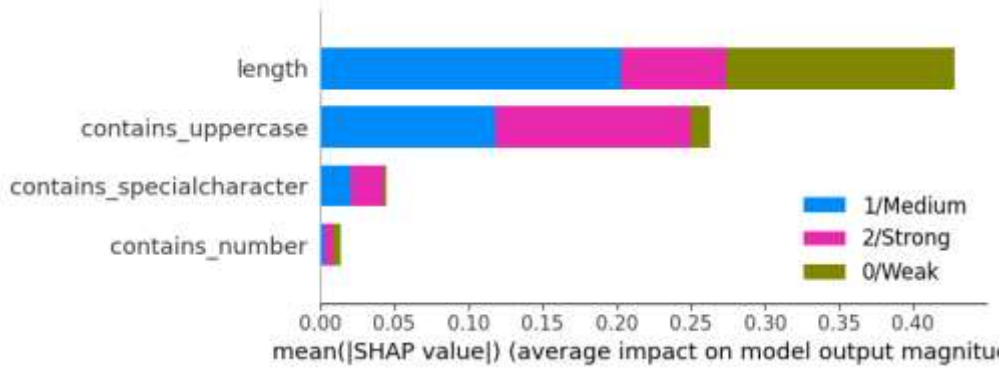
a. Akurasi**:** 97,66%
b. Precision (rata-rata makro)**:** 94,51%
c. Recall (rata-rata makro)**:** 98,23%
d. F1-score (rata-rata makro)**:** 96,22%

Hasil tersebut menunjukkan bahwa model cukup andal dalam mengklasifikasikan kekuatan *password* ke dalam tiga kelas (0/lemah, 1/sedang, 2/kuat). Secara khusus, algoritma ini berhasil menangkap pola dari fitur-fitur seperti panjang *password*, keberadaan huruf kapital, karakter numerik, dan simbol khusus dalam membedakan tingkat kekuatan *password*.

3.3 Implementasi Explainable AI dengan SHAP dan LIME

Untuk mengimplementasikan transparansi model dan memberikan penjelasan atas hasil prediksi, penelitian ini mengimplementasikan dua pendekatan *Explainable Artificial Intelligence* (XAI), yaitu SHAP (*SHapley Additive exPlanations*) dan LIME (*Local Interpretable Model-agnostic Explanations*)**.** Keduanya diterapkan untuk menganalisis alasan di balik klasifikasi kekuatan *password* yang dilakukan oleh model *Naive Bayes*.

3.3.1 Penjelasan Global dengan SHAP



Gambar 5. Hasil SHAP (Rata-rata Kontribusi Fitur terhadap Output Hasil Prediksi)

SHAP digunakan untuk menganalisis kontribusi masing-masing fitur secara global terhadap keputusan model. Hasil analisis SHAP global (terhadap seluruh data uji) ditunjukkan pada gambar 5. Hasil visualisasi *summary plot* menunjukkan bahwa fitur *length* (panjang *password*) memiliki kontribusi paling besar terhadap prediksi kelas (ditunjukkan dengan posisi paling atas), disusul oleh *contains_uppercase*, *contains_specialcharacter*, dan *contains_number*. *Password* yang lebih panjang dan mengandung karakter kapital diklasifikasikan ke kelas dengan kekuatan yang lebih tinggi. Sementara itu, fitur contains_number memiliki kontribusi yang mendekati 0, artinya tidak memiliki kontribusi signifikan terhadap hasil prediksi. Detail rata-rata kontribusi fitur ditampilkan dalam tabel 3.

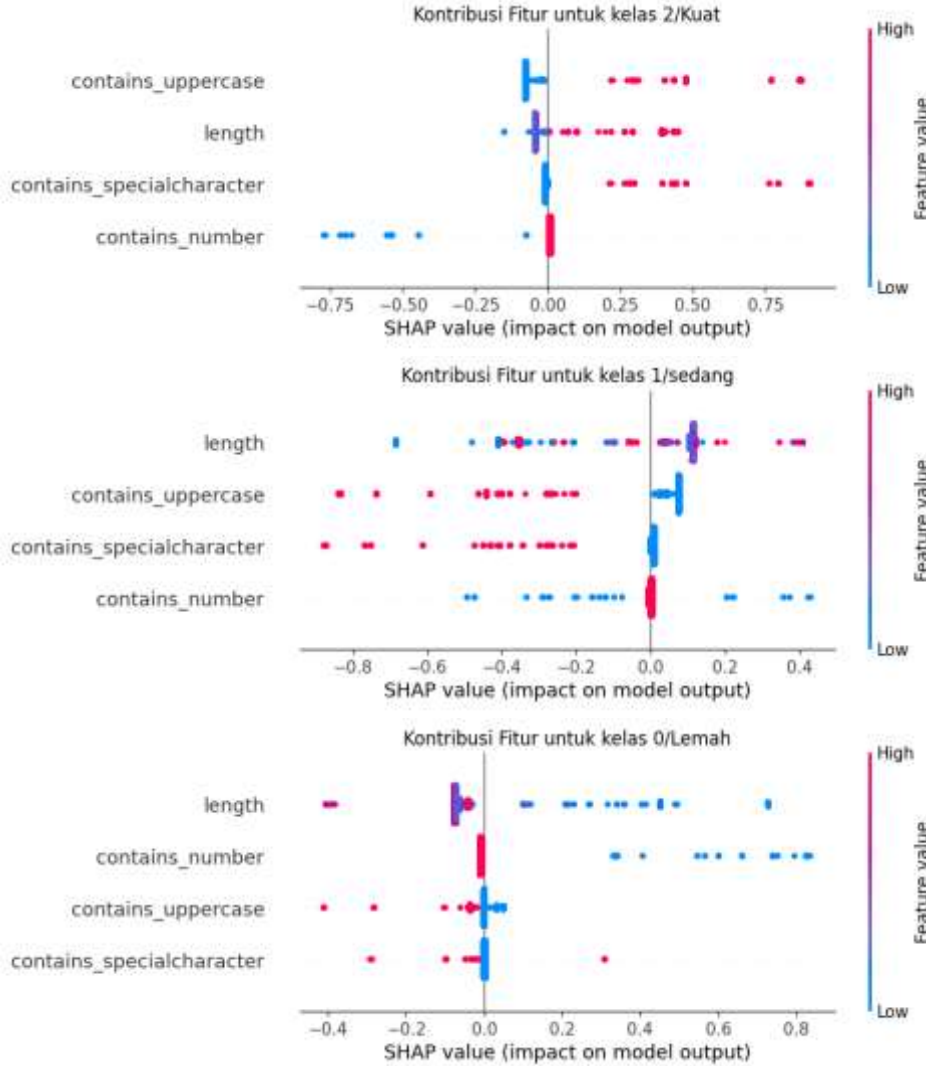
Tabel 3. Rata-rata Kontribusi Fitur terhadap Hasil Prediksi

Fitur	0 / lemah	1 / sedang	2 / kuat
length	0.15408423	0.20296148	0.07086858
contains_uppercase	0.0131368	0.11830527	0.13144217
contains_specialcharacter	0.00164566	0.02071282	0.02224594

Jurnal SAINTIKOM (Jurnal Sains Manajemen Informatika dan Komputer)
Volume 24 ; Nomor 2 ; Agustus 2025 ; Page 193-202
E-ISSN : 2615-3475; P-ISSN : 1978-6603
https://ojs.trigunadharma.ac.id/index.php/jis/index

contains_number	0.00569254	0.00227335	0.00619331

Nilai SHAP rata-rata pada tabel 3 menunjukkan bahwa model mengandalkan pola umum yang sejalan dengan prinsip keamanan *password*, yakni semakin kompleks dan panjang sebuah *password*, maka semakin kuat klasifikasinya. SHAP juga memberi kemudahan untuk manganalisis secara spesifik kontribusi setiap fitur terhadap dari setiap klasifikasi yang ada (Gambar 6). Sejalan dengan hasil ringkasan di gambar 5, *password* dengan label kelas 2/kuat sangat dipengaruhi oleh adanya karakter kapital. Sementara, untuk kelas 1/sedang dan 0/lemah sangat dipengaruhi oleh panjang *password*. Ini terlihat dari posisi paling atas dari setiap fitur pada visualisasi SHAP.

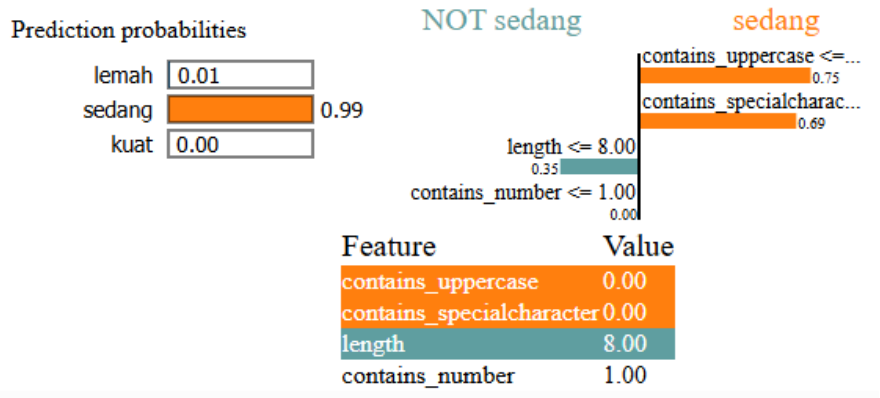


Gambar 6. Diagram *Beeswarm* SHAP untuk Kontribusi Fitur di Setiap Kelas Prediksi

3.3.2 Penjelasan Lokal dengan LIME

LIME digunakan untuk memberikan penjelasan lokal terhadap prediksi satu *password*. Pendekatan ini membangun model lokal sederhana di sekitar data uji yang ingin dijelaskan. Misalnya, untuk *password* "hello123", model memilih prediksi kelas 1/sedang dengan nilai probabilitas tertinggi 0.99 (gambar 7). LIME memberikan visualisasi dari kontribusi setiap fitur terhadap *password* yang diuji.

Jurnal SAINTIKOM (Jurnal Sains Manajemen Informatika dan Komputer)
Volume 24 ; Nomor 2 ; Agustus 2025 ; Page 193-202
E-ISSN : 2615-3475; P-ISSN : 1978-6603
https://ojs.trigunadharma.ac.id/index.php/jis/index



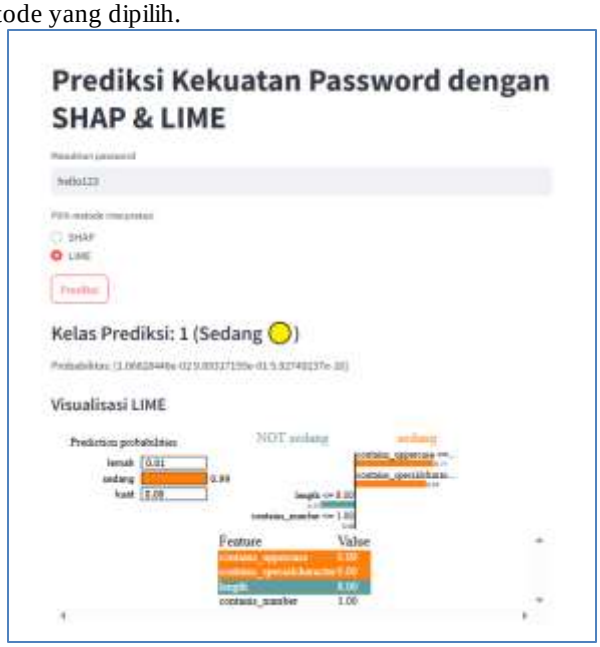
Gambar 7. Hasil LIME untuk *Password* "hello123"

Dari gambar 7, didapatkan penjelasan lebih rinci sebagai berikut:

a. contains_uppercase <= 0.00 (+0.75). *Password* ini tidak memiliki huruf kapital (contains_uppercase = 0). Hal ini sangat mendorong model untuk memprediksi kelas yang lebih lemah, karena nilai positif pada LIME di sini justru mengarah ke kelas yang diprediksi saat ini (kemungkinan besar kelas 0 atau 1). Jadi tidak adanya huruf kapital adalah salah satu alasan utama *password* diklasifikasikan sebagai tidak kuat.
b. contains_specialcharacter <= 0.00 (+0.69). *Password* tidak memiliki karakter spesial (@, #, dll). Ini juga menjadi alasan kuat mengapa model tidak menganggap *password* ini kuat. Mirip dengan poin sebelumnya, tidak adanya karakter spesial menyebabkan prediksi diarahkan ke kelas yang lebih rendah.
c. length <= 8.00 (−0.35). *Password* ini memiliki panjang kurang dari atau sama dengan 8 karakter. Namun kontribusinya negatif, artinya kondisi ini sebenarnya sedikit menahan model untuk tidak memprediksi kelas yang lebih lemah. Bisa jadi karena ada fitur lain (seperti kehadiran angka) yang membuat *password* tidak sepenuhnya buruk.
d. contains_number <= 1.00 (0.000). Fitur ini tidak berkontribusi dalam keputusan model untuk *password* yang diuji. Artinya kehadiran angka tidak terlalu berdampak karena mungkin fitur ini umum di banyak kelas.

3.4 Implementasi Aplikasi Interaktif dengan Streamlit

Untuk mendemonstrasikan hasil klasifikasi kekuatan *password* secara lebih praktis dan mudah diakses serta analisis SHAP maupun LIME, penelitian ini juga mengembangkan sebuah aplikasi interaktif berbasis Streamlit (gambar 8). Aplikasi ini memiliki antarmuka sederhana yang memungkinkan pengguna untuk memasukkan sebuah *password* melalui form input. Setelah itu, pengguna dapat memilih tipe penjelasan yang ingin didapatkan, yakni antara SHAP dan LIME. Setelah klik tombol 'Prediksi', pengguna dapat melihat hasil prediksi kekuatan *password* (lemah, sedang, kuat) berdasarkan model *Naive Bayes* yang telah diseimpan dari tahap pelatihan dan melihat visualisasi dan penjelasan tentang kontribusi fitur berdasarkan metode yang dipilih.



Gambar 8. Aplikasi Klasifikasi Kekuatan *Password* dan XAI Berbasis Streamlit

Aplikasi ini tidak hanya berguna untuk pengguna individu dalam menilai kekuatan *password*, tetapi juga dapat berfungsi sebagai alat edukasi dalam konteks pelatihan keamanan siber. Dengan adanya penjelasan visual berbasis XAI, pengguna dapat memahami alasan di balik klasifikasi dan belajar membuat *password* yang lebih kuat berdasarkan elemen-elemen penting yang diidentifikasi oleh model. Jika dibandingkan dengan penelitian sebelumnya tentang klasifikasi kekuatan *password*, seperti [20]–[22], penelitian ini memiliki keunggulan di aspek interpretasi hasil prediksi dari model klasifikasi. Ketiga penelitian tersebut hanya menyoroti performa model machine learning terbaik yang didapatkan dari beberapa algoritma klasifikasi, tanpa menjelaskan bagaimana hasil prediksi didapatkan. Penelitian ini menutup celah tersebut dengan menerapkan XAI sebagai alat bantu interpretasi hasil prediksi model *machine learning*.

4. KESIMPULAN

Penelitian ini telah berhasil mengimplementasikan Explainable Artificial Intelligence (XAI) dalam klasifikasi kekuatan password menggunakan model Naive Bayes. Dengan memanfaatkan dataset dari Kaggle yang terdiri dari 100.000 password dengan tiga label kelas kekuatan, model dilatih berdasarkan fitur-fitur sederhana namun representatif, seperti panjang password, keberadaan huruf kapital, angka, dan karakter khusus. Model klasifikasi yang dibangun menunjukkan hasil yang baik dengan nilai matrik akurasi: 97,66%, Precision: 94,51%, Recall: 98,23%, dan F1-score: 96,22%.

Untuk memberikan interpretasi terhadap hasil klasifikasi, digunakan dua metode XAI, yaitu SHAP dan LIME. Keduanya mampu memberikan wawasan mendalam tentang kontribusi masing-masing fitur terhadap prediksi yang dihasilkan. SHAP memberikan visualisasi berbasis kontribusi fitur secara kumulatif/global, sedangkan LIME menyoroti pengaruh fitur secara lokal dalam ruang data di sekitar instance tertentu. Untuk memudahkan interpretasi bagi pengguna, dikembangkan pula aplikasi interaktif berbasis Streamlit yang memungkinkan pengguna memasukkan password dan mendapatkan evaluasi langsung atas kekuatannya beserta visualisasi interpretatif. Implementasi ini menunjukkan potensi dari integrasi XAI dalam sistem klasifikasi untuk meningkatkan transparansi, akuntabilitas, dan edukasi pengguna.

DAFTAR PUSTAKA

[1] M. Yamin, T. T. Malethi, Monica, Jodhika, and S. Natali, "Evaluasi Risiko Pada Penggunaan Password Yang Lemah: Analisis Kasus Penggunaan Password Umum," *J. Ilm. Multidisiplin Ilmu Komput.*, vol. 1, no. 1, pp. 41–48, 2023, doi: 10.61674/jimik.v1i1.112.

[2] V. Zimmermann, "From the Quest to Replace Passwords towards Supporting Secure and Usable Password Creation," Technische Universität Darmstadt, Darmstadt, 2021.

[3] Arnah Ritonga *et al.*, "Analisis Kombinatorik Dalam Menentukan Keamanan dan Kompleksitas Password dengan Penerapan Teori Kombinatorik," *Katalis Pendidik. J. Ilmu Pendidik. dan Mat.*, vol. 2, no. 2 SE-Articles, pp. 49–64, Apr. 2025, doi: 10.62383/katalis.v2i2.1463.

[4] W. Y. Saputra, S. Sugiarti, H. Junianto, and D. Suhartono, "Password Strength Study Using The Zxcvbn Algorithm And Brute-Force Time Estimation To Strengthen Cybersecurity," *J. Pilar Nusa Mandiri*, vol. 21, no. 1, pp. 52–59, 2025, doi: 10.33480/pilar.v21i1.6119.

[5] T. Rochmadi, A. Fadlil, and I. Riadi, "Tinjauan Pustaka Sistematis: Tantangan Dan Faktor-Faktor Pengembangan Kesiapan Forensik Digital ," *Cyber Secur. dan Forensik Digit.*, vol. 7, no. 2 SE-Articles, pp. 81–89, Dec. 2024, doi: 10.14421/csecurity.2024.7.2.4861.

[6] A. Asaduzzaman, D. D'Souza, M. R. Uddin, and Y. Woldeyes, "Increase Security by Analyzing Password Strength using Machine Learning," in *2024 Joint International Conference on Digital Arts, Media and Technology with ECTI Northern Section Conference on Electrical, Electronics, Computer and Telecommunications Engineering (ECTI DAMT & NCON)*, 2024, pp. 32–37, doi: 10.1109/ECTIDAMTNCON60518.2024.10479995.

[7] E. Darbutaitė, P. Stefanovič, and S. Ramanauskaitė, "Machine-Learning-Based Password-Strength-Estimation Approach for Passwords of Lithuanian Context," *Appl. Sci.*, vol. 13, no. 13, 2023, doi: 10.3390/app13137811.

[8] R. Dwivedi *et al.*, "Explainable AI (XAI): Core Ideas, Techniques, and Solutions," *ACM Comput. Surv.*, vol. 55, no. 9, Jan. 2023, doi: 10.1145/3561048.

[9] S. M. Lundberg and S.-I. Lee, "A Unified Approach to Interpreting Model Predictions," in *Advances in Neural Information Processing Systems*, 2017, vol. 30.

[10] M. T. Ribeiro, S. Singh, and C. Guestrin, "'Why Should I Trust You?': Explaining the Predictions of Any Classifier," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2016, pp. 1135–1144, doi: 10.1145/2939672.2939778.

[11] T. Hulsen, "Explainable Artificial Intelligence (XAI): Concepts and Challenges in Healthcare," *AI*, vol. 4, no. 3, pp. 652–666, 2023, doi: 10.3390/ai4030034.

[12] J. Černevičienė and A. Kabašinskas, "Explainable artificial intelligence (XAI) in finance: a systematic literature review," *Artif. Intell. Rev.*, vol. 57, no. 8, p. 216, 2024, doi: 10.1007/s10462-024-10854-8.

[13] C. Trivedi *et al.*, "Explainable AI for Industry 5.0: Vision, Architecture, and Potential Directions," *IEEE Open J. Ind. Appl.*, vol. 5, pp. 177–208, 2024, doi: 10.1109/OJIA.2024.3399057.

[14] Q. Liu, J. D. Pinto, and L. Paquette, "Applications of Explainable AI (XAI) in Education," in *Trust and Inclusion in AI-Mediated Education: Where Human Learning Meets Learning Machines*, D. Kourkoulou, A.-O. (Olnancy) Tzirides, B. Cope, and M. Kalantzis, Eds. Cham: Springer Nature Switzerland, 2024, pp. 93–109.

[15] A. Dovier, T. Dreossi, and A. Formisano, "XAI-LAW Towards a logic programming tool for taking and explaining legal

decisions," *CEUR Workshop Proc.*, vol. 3733, 2024.

[16] A. S. Yazid, "Eksplorasi Data Akademik untuk Memprediksi Ketepatan Waktu Lulus Mahasiswa Menggunakan Algoritma Naive Bayes," *Jatisi*, vol. 11, no. 4, pp. 558–568, 2024.

[17] M. Khorasani, M. Abdou, and J. Hernández Fernández, "Getting Started with Streamlit BT - Web Application Development with Streamlit: Develop and Deploy Secure and Scalable Web Applications to the Cloud Using a Pure Python Framework," M. Khorasani, M. Abdou, and J. Hernández Fernández, Eds. Berkeley, CA: Apress, 2022, pp. 1–30.

[18] K. Maxim, "Password Security: Sber Dataset," *Kaggle*, 2022. [Online]. Available: https://www.kaggle.com/datasets/morph1max/password-security-sber-dataset/. [Accessed: 19-Apr-2025].

[19] N. R. Dzakiyullah, M. A. Burhanuddin, R. R. Raja Ikram, N. Yudistira, M. R. Fauzi, and D. Purbohadi, "Multi-Label Risk Prediction Diabetes Complication Using Machine Learning Models," *Int. J. Online Biomed. Eng.*, vol. 20, no. 16 SE-Papers, pp. 66–88, Dec. 2024, doi: 10.3991/ijoe.v20i16.51643.

[20] S. Sarkar and M. Nandan, "Password Strength Analysis and its Classification by Applying Machine Learning Based Techniques," in *2022 Second International Conference on Computer Science, Engineering and Applications (ICCSEA)*, 2022, pp. 1–5, doi: 10.1109/ICCSEA54677.2022.9936117.

[21] S. J. Kim and B. M. Lee, "Multi-Class Classification Prediction Model for Password Strength Based on Deep Learning," *J. Multimed. Inf. Syst.*, vol. 10, no. 1, pp. 45–52, Mar. 2023, doi: 10.33851/JMIS.2023.10.1.45.

[22] H. Rehman *et al.*, "Password Strength Classification Using Machine Learning Methods," in *2024 Global Conference on Wireless and Optical Technologies (GCWOT)*, 2024, pp. 1–7, doi: 10.1109/GCWOT63882.2024.10805622.