Digital Receipt

This receipt acknowledges that Turnitin received your paper. Below you will find the receipt information regarding your submission.

The first page of your submissions is displayed below.

Submission author:	Wyatt Alcala
Assignment title:	NO REPOSITORY 102
Submission title:	Jurnal Data Mining.docx
File name:	Jurnal_Data_Mining.docx
File size:	289.73K
Page count:	6
Word count:	3,059
Character count:	20,275
Submission date:	22-Jul-2025 08:46PM (UTC+0200)
Submission ID:	2719098103

Analisis Kompleksitas Password Dengan Metode KNN, Naïve Bayes, Random Forest, SVM, Linear Regression

1st Melco Lauwento
Sistem Informasi
Universitas Komputer Indoensia
Bandung, Indonesia
melco05lauwento.1@gmail.com

2nd Daffa Yudis Pratama
Sistem Informasi
Universitas Komputer Indoensia
Bandung, Indonesia
melco05lauwento.1@gmail.com

3th Muhammad Naufal Sandu
Sistem Informasi
Universitas Komputer Indoensia
Bandung, Indonesia
melco05lauwento.1@gmail.com

***Abstrak*— Keamanan kata sandi telah menjadi askpek kritis dalam sistem proteksi Informasi digital, dengan penilaian kekuatan kata sandi yang bergantung pada kompleksitasnya. Penelitian ini bertujuan untuk menganalisis kompleksitas kata sandi Menggunakan berbagai algoritma machine learning, termasuk K-Nearest Neighbor (KNN), Naïve Bayes, Decision tree,Linear Regression, dan Support Vector Machine (SVM), Dengan menggunakan dataset dari Kaggle yang berisi berbagai kata sandi dengan tingkat kekuatan yang berbeda, model-model ini diuji untuk memprediksi tingkat keamanan kata sandi. Untuk mengilustrasikan data dengan dapa dimengerti, salah satu alat bantu yang dapa diunakan adalah Goggle Colabs menggunakan Bahasa Python. Dalam proses yang dilakukan adalah menganalisi bagaimana prediksi antara password dan tingkat kesulitannya menggunakan lima Machine Learning. Analisis data mining ini melakukan dengan penerapan Teknik klasifikasi dengan memanfaatkan lima metode algoritma yang berbeda.**

Kata Kunci —Machine Learning, Password Strength, K-Nearest Neighbor, Naïve Bayess, Random Forest, Linear Regression, Support Vector Machine.

I. Pendahuluan

Keamanan Kata sandi menjadi aspek krusial dalam sistem perlindungan data di dunia digital saat ini. Di banyak sistem keamanan, kata sandi berfungsi sebagai lapisan pertahanan pertama yang melindungi Informasi sensitif dari akses yang tidak sah. Namun, meskipun keberadaanya sangat penting, bantak pengguna yang memilih kata sandi yang lemah dan mudah ditebak, seperti menggunakan kata yang ada dalam kamu atau pola keyboard yang sederhana. Keadaan ini menciptakan kerentanannya terhadap serangan seperti brute-force atau dictionary attack. Oleh karena itu, penting untuk memiliki sistem yang dapat scara efisien menilai kekuatan kata sandi, untuk memastikan bahwa pengguna membuat kata sandi yang cukup kuat guna melindungi data pribadi mereka dari potensi ancaman. Beberapa alat pengukur kekuatan kata sandi berbasis statistic dan model Meachine Learning dikembangkan, namun mereka masih terbatas dalam hal penerapan di berbagai Bahasa dan konteks, khususnya dalam menganalisis kata sandi yang lebih kompleks atau berbassis Bahasa tertentu.

Penelitian ini bertujuan untuk mengesskplorasi penerapan Machine Learning dalam menilai kekuatan kata sandi dengan menggunakan berbagai model, seperti K-Nearest Neighbor(KKN), Naïve Bayes, Random Forest, Linear Regression, dan Support Vector Machine(SVM). Motivasi utama dibalik penelitian ini adalah untuk meningkatkan akurasi dan efisien dalam mengklasifikasikan tingkat kekuatan kata sandi yang dapat digunakan dalam berbagai sistem keamanan. Melalui pendekatan ini, kami mencoba untuk menggali potensi teknik Machine Learning yang lebih canggih dalam menganalisis dan memberikan umpan balik yang lebih akurat terkait dengan kekuatan kata sandi. Dengan model-model ini, diharapkan dapat diperoleh prediksi yang lebih presisi dan dapat diadaptasi untuk berbagai dataset dan konteks, bauk untuk kata sandi berbahasa inggris maupun Bahasa lainnya.

Lingkup penelitian ini mencakup penggunaan lima algortima Machine Learning yang berbeda untuk menilai kekuatan kata sandi K-Nearest Neighbors(KKN), Naïve Bayes, Random Forest, Linear Regression, dan Support Vector Machine(SVM). Setiap model ini diuji menggunakan dataset yang berisi kata sandi dengan label kekuatan yang telah ditentukan, berdasarkan fitur-fitur seperti Panjang kata sandi, jumlah symbol, dan keberadaan huruf kapital serta angka. Kontribusi utama dari penelitian ini adalah perbandingan kinerja dari masing-masing model dalam hal akurasi prediksi dan waktu komputasi,serta penyesuaian model-model tersebut untuk meningkatkan kemampuan dalam mengklasifikasikan kata sandi yang lebih kompleks, seperti yang ditemukan dalam data dunia nyata, Inovasi lain yang dihasilkan adalah pemanfaatan berbagai ukuran kesamaan teks dalam meningkatkan akurasi model, serta penerapan ensemble methods seperti Random Forest yang terbukti memberikan hasil yang lebih baik dibandingkan dengan model yang lebih sederhana. Penelitian ini juga memberikan wawasan terkait dengan adaptasi model Machine Learning untuk analisis kekuatan kata sandi dalam berbagai konteks dan Bahasa.

II. Metodologi

A. Pengumpulan dan Preprocessing Data

Dataset yang digunakan dalam penelitian ini diperoleh dari platfrom Kaggle, yang menyediakan data mengenai kekuatan password yang mencakup kategori-kategori seperti 0 = week, 1 = Medium, 2 = Strong. Data set ini lebih dari 600.000 password yang sudah diberi label kekuatannya, yang memungkinkan kami untuk melatih model Machine Learning dan memprediksi tingkat kekuatan password. Sumber dataset ini di ambil dari kumpulan data yang telah dipublikasikan oleh peneliti sebelumnya, dengan beberapa variasi kata sandi dari berbagai konteks, baik internasional maupun lokal, yang memastikan keberagaman data.

Langkah pertama dalam pengolahan data adalah membersihkan dataset untuk memastikan bahwa data yang