Klasifikasi Deteksi *Link Phising* DANA Kaget Menggunakan Metode *Support Vector Machine* Berbasis *Website*

Mutiara Vebriani[1], Wiyli Yustanti[2]

[1,2] Teknik Informatika, Universitas Negeri Surabaya
[1]mutiara.20014@mhs.unesa.ac.id
[2]wiyliyustanti@unesa.ac.id

***Abstrak*—Penelitian ini bertujuan untuk mengembangkan model klasifikasi deteksi Link phishing DANA Kaget menggunakan metode *Support Vector Machine (SVM)* berbasis websitesite. Tujuan utama dari model ini adalah memberikan perlindungan kepada pengguna layanan "DANA Kaget" dengan memberi peringatan ketika mereka mengunjungi tautan yang mencurigakan atau situs website yang berpotensi phishing. Data Link diperoleh melalui observasi manual pada media sosial seperti Facebook, Telegram, dan WhatsApp, dengan total 865 Link phishing dan non-phishing yang dikumpulkan. Metode SVM yang diterapkan pada model ini Berdasarkan hasil *cross-validation* dengan fold 3, 5, dan 10, dapat disimpulkan bahwa model memiliki kinerja yang stabil dengan peningkatan akurasi pada data train dan test seiring dengan peningkatan jumlah fold. Pada fold 3, model mencapai akurasi train sebesar 88% dan test sebesar 84%, menunjukkan kemampuan model dalam mempelajari pola pada data training. Dengan penambahan fold menjadi 5, akurasi train dan test meningkat menjadi 89% dan 85% secara berturut-turut, menandakan peningkatan kemampuan generalisasi model. Pada fold 10, akurasi train meningkat menjadi 90% dan akurasi test mencapai 88%, menunjukkan bahwa model ini memiliki kemampuan yang baik dalam menggeneralisasi pada data baru. Dengan demikian, dapat disimpulkan bahwa model ini konsisten dalam performanya dan memiliki potensi untuk diterapkan pada data baru. Penelitian ini memberikan kontribusi penting dalam meningkatkan keamanan pengguna layanan DANA Kaget dengan memberikan peringatan dini terhadap tautan yang berpotensi phishing. Dengan adanya model ini, diharapkan dapat mengurangi risiko kebocoran informasi dan penipuan yang dilakukan melalui tautan phishing.**

Kata Kunci*— *SVM, Phishing*, Dana, *CrossValidation, Website

I. PENDAHULUAN

Dalam era digital saat ini pembayaran digital sangat diminati, Pembayaran digital merujuk pada proses pembayaran yang dilakukan tanpa menggunakan uang tunai fisik, seperti uang kertas atau koin, dan seringkali tidak melibatkan kartu kredit atau debit secara fisik dengan kemudahan pembayaran tersebut pembayaran digital sangat diminati. Bank Indonesia terus menggalakkan upaya untuk mempercepat transformasi digital dalam sistem pembayaran dan memperluas kerja sama lintas negara dalam hal pembayaran. Tujuannya adalah untuk mendorong inklusi ekonomi keuangan, serta memperluas cakupan ekonomi dan sektor keuangan digital [1].

Salah satu platform pembayaran digital adalah DANA adalah platform pembayaran digital yang resmi dan diawasi oleh Bank Indonesia (BI). Platform ini memiliki pendekatan open platform dan dapat digunakan di berbagai aplikasi, toko online, dan transaksi konvensional. DANA merupakan produk dari perusahaan startup yang beroperasi di bawah hukum Indonesia. Pada tahun 2023, DANA mencapai pertumbuhan yang sangat signifikan dengan jumlah pengguna mencapai 150 juta. Selain itu, terdapat peningkatan rata-rata sebesar 137% dalam jumlah transaksi [2].

Banyak fitur-fitur yang ada pada DANA salah satu yang akhir-akhir ini banyak digunakan adalah fitur Dana Kaget, DANA Kaget adalah salah satu fitur yang menarik dalam aplikasi DANA, karena tidak hanya memungkinkan transfer uang antar pengguna DANA, tetapi juga memberikan peluang mendapatkan saldo DANA secara gratis hanya dengan menekan Link yang diberikan. Fitur DANA Kaget yang banyak digunakan oleh pengguna tidak menutup kemungkinan adanya banyak Link phishing yang mencoba memanfaatkannya. Dengan popularitasnya, DANA Kaget telah menjadi sasaran bagi penjahat cyber yang mencoba menipu pengguna dengan mengirimkan tautan palsu yang mengklaim hadiah atau saldo gratis. Ini dapat menjadi ancaman serius karena pengguna yang tidak waspada dapat terjebak dalam skema phishing dan mengungkapkan informasi pribadi mereka. Keamanan informasi pengguna Dana Kaget menjadi sangat penting karena banyaknya transaksi keuangan yang dilakukan melalui platform ini Jumlah serangan phishing unik yang dilaporkan sebanyak 7.988 [3].

Tercatat bahwa lembaga keuangan menjadi target utama serangan phishing, dengan persentase mencapai 23,5%. Serangan phishing pada media sosial berada di posisi berikutnya dengan 22,3%. Sebanyak 16,3% serangan phishing menyasar SAAS/Websitemail, sedangkan hanya 2,2% serangan phishing yang ditujukan pada pencurian data di game online dan akun aset kripto. Banyaknya laporan phishing ini juga dipengaruhi oleh rendahnya tingkat kesadaran masyarakat. Selain itu, pelaku phishing kini dapat menggunakan lebih dari satu nama domain, yang menyebabkan peningkatan jumlah laporan yang masuk. [4].

Terlebih lagi, pengguna seringkali tidak memiliki sumber daya teknis untuk mendeteksi tautan phising dengan akurat. Penelitian ini akan memberikan kontribusi signifikan dengan menyediakan alat otomatis untuk mengidentifikasi tautan phising dengan keandalan tinggi. Phising adalah upaya untuk

mendapatkan informasi seperti User name, password, dan rincian kartu kredit dengan menyamar sebagai orang lain dalam pesan elektronik. Untuk memikat orang yang tidak berhati-hati, komunikasi yang mengaku berasal dari situs website sosial yang populer, situs lelang, prosesor pembayaran online, atau administrator IT biasanya digunakan. [5].

Berdasarkan penelitian terdahulu, teknik deteksi yang paling banyak digunakan adalah analisis situs phising yang mengambil data pada phisnktank dan kaggle. Didalam Jurnal disebutkan bahwa Berdasarkan jurnal yang disebutkan, model WebsitePhish memiliki akurasi sebesar 98.1% dalam mendeteksi halaman website phishing[6].

Di dalam analisis situs disebutkan bahwa terdapat beberapa pendekatan untuk mendeteksi situs phising antara lain pendekatan Suport Vector Machine, Regresion Linier. Didalam jurnal tersebut penulis memberikan beberapa saran untuk penelitian selanjutnya yaitu penggunaan metode machine learning lainnya serta data yang lebih bervariasi untuk meningkatkan tingkat akurasi model. Metode SVM memiliki Akurasi klasifikasi terbaik diperoleh dengan menggunakan algoritma SVM, dengan akurasi sebesar 0.9419. Namun algoritma SVM memiliki waktu berjalan paling lama yaitu rata-rata 186,12 detik. Di sisi lain, algoritma hutan acak mencapai keseimbangan yang baik antara waktu dan kualitas klasifikasi, dengan akurasi 0,9269 dan waktu berjalan 7,82 detik [7].

Pada Jurnal tersebut peneliti selanjutnya dapat memandingkan metode yang lain untuk mengetahui metode mana yang paling efektif dalam menganalisa berita hoax. Perbandingan Metode Naïve bayes, LTMS dan SVM menunjukan bahwa pada klasifikasi berita hoax menunjukan bahwa Metode SVM unggul dengan akurasi sebesar 74% dari 2000 data [8].

SVM (Support Vector Machine) adalah sebuah algoritma dalam machine learning yang digunakan untuk klasifikasi dan regresi. Algoritma SVM bekerja dengan cara menemukan hyperplane (bidang pemisah) yang dapat memisahkan dua set data atau kelas yang berbeda secara optimal. SVM memiliki kelebihan dalam menangani data yang kompleks dan dapat mengatasi masalah klasifikasi dengan baik, terutama ketika data tidak linear terpisah. SVM menggunakan konsep "support vectors" yang merupakan titik-titik data yang paling dekat dengan hyperplane, sehingga memungkinkan SVM untuk mencari pemisah yang paling baik. Selain itu, SVM juga memiliki parameter kernel yang memungkinkan transformasi data ke dalam dimensi yang lebih tinggi, sehingga dapat menangani data yang tidak dapat dipisahkan secara linear di dimensi rendah [9].

SVM telah digunakan dalam berbagai aplikasi, termasuk klasifikasi teks, pengenalan wajah, analisis citra medis, dan deteksi anomali. Algoritma ini telah terbukti efektif dalam banyak kasus dan menjadi salah satu alat penting dalam machine learning. Pada penelitian ini peneliti menggunakan berbagai metode diantaranya Suport Vector Machine, Naïve bayes, KNN, Random Forest untuk mengetahui metode mana yang memiliki akurasi tinggi untuk di implementasiken pada website yang akan di buat [10].

Sebelum menentukan metode yang pilih peneliti melakukan Preprosesing data menggunakan URL yang dibersihkan diubah menjadi vektor numerik untuk pengenalan pola dan penemuan pengetahuan yang efektif sebelum algoritma pembelajaran diterapkan. Ini adalah fase vektorisasi. CountVectorizer adalah teknik vektorisasi yang digunakan untuk percobaan. Dengan menggunakan SVM, model ini dapat memisahkan data antara situs website yang aman dan situs website yang berpotensi phishing dengan lebih efisien. SVM bekerja dengan cara mencari garis pemisah terbaik antara dua kelas data, sehingga dapat dengan akurat mengklasifikasikan tautan atau situs website apakah masuk ke dalam kategori aman atau mencurigakan [11].

Tujuan utama dari model ini adalah memberikan perlindungan kepada pengguna layanan "DANA Kaget" dengan memberi peringatan ketika mereka mengunjungi tautan yang mencurigakan atau situs website yang berpotensi phishing. Dengan demikian, korban kejahatan phishing dapat dihindari, dan kerugian finansial serta pencurian data pribadi dapat diminimalkan. Hasil dari model ini dapat diukur secara matematis dan validasinya dapat diandalkan, sehingga dapat digunakan sebagai dasar untuk penelitian serupa di masa depan. Jika model klasifikasi ini diimplementasikan, dapat membantu mengurangi risiko pengguna "DANA Kaget" terkena serangan phishing.

II. METODOLOGI PENELITIAN

Metode penelitian adalah serangkaian kegiatan yang dilakukan untuk penelitian dalam skripsi ini guna mendapatkan hasil sesuai yang diharapkan dalam tujuan penelitian tentunya dengan batasan yang sudah ditentukan, dengan langkah langkahnya. Pada bab ini akan membahas tahapan-tahapan yang digunakan dalam menyelesaikan penelitian ini. Adapun tahapan tersebut disini peneliti menggunakan Framework Data Mining CRIPS DM merupakan singkatan dari Cross-Industry Standard Process for Data Mining, adalah model proses yang telah diadopsi secara luas untuk memandu proyek ilmu data. Model ini terdiri dari enam tahap, yaitu Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation, dan Deployment. Tahap pertama, Business Understanding, melibatkan definisi masalah bisnis dengan jelas, sementara tahap-tahap berikutnya membawa data dari pemahaman awal hingga pengembangan model dan implementasi [12]. Dimana tahapanya adalah sebagai berikut

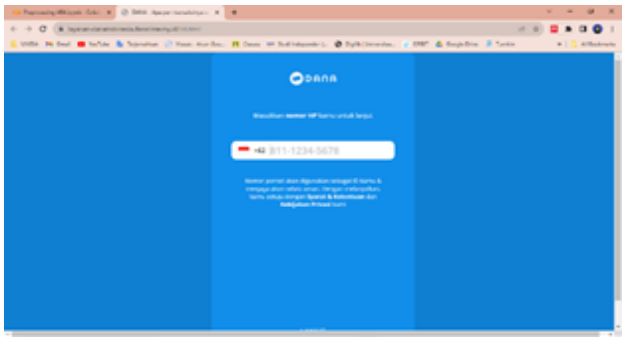
A. Business Understanding

Pada Tahap ini penelitian Business Understanding adalah tahap kunci dalam menghadapi masalah penyebaran banyaknya tautan phishing "Dana Kaget" di media sosial. Tahap ini mencakup pemahaman mendalam tentang masalah ini dari perspektif bisnis [13]. Pada tahap ini peneliti menemukan banyak Link dana kaget phising yang disebarkan oknum tidak bertanggung jawab di berbagai social media seperti telegram, Whatsapp Grup dan Facebook, Kurangnya

pengetahuan mengenai Link phising banyak korban yang tertipu dengan memasukan informasi pribadi yang dapat digunakan penipu untuk login akun Dana mereka.

B. *Understanding The Data*

Data yang digunakan merupakan data primer yang peneliti observasi secara manual mencari Link dana phising pada platform sosial media seperti telegram, Whatsapp Grup dan Facebook. Untuk Link dana non phising dilakukan observasi secara manual mengumpulkan data pada Link Whatsapp Grup dan banyak artikel berita yang membagikan Link dana kaget di artikelnya. Tampilan Link dana Phising dibuat semirip mungkin dengan Link dana yang asli berikut tampilan Link dana phising dan Link dana non phising:



Gbr. 1 Tampilan Dana Phishing

Gbr. 2 Tampilan Dana Non Phising pada website

Gbr. 3 Tampilan Artikel penyebar Dana Kaget Non Phishing

Berikut jumlah data yang berhasil dikumpulkan oleh peneliti:

TABEL I
JUMLAH DATA

Label	Jumlah
Non Phishing	314
Phishing	314
Total	638

C. *Data Preparation*

Data yang telah dikumpulkan kemudian di labeli sesuai dengan label masing-masing. Terdapat 2 label yaitu Phising dan Non Phising. Setelah data terkumpul data kemudian di bersihkan menggunakan Bahasa pemrograman python dan menggunakan google colab sebagai pengeksekusi kode python [14]. Pada tahap data preparation Langkah pertama yang dilakukan dalam penelitian ini adalah:

1. Membersihkan Data Ganda

   Dalam proses pengumpulan data Link dana kaget phising tidak menutup kemmungkinan bahwa data yang dikumpulkan terdapat data yang ganda oleh sebab itu tahap awal penelitian ini adalah membersihkan data yang ganda agar pada saat evaluasi tidak overfitting.

2. Membersihkan *Sub Domain*

   Dalam pengumpulan Link terdapat sub domain yang sangat Panjang hal tersebut tentu dapat mempengaruhi kinerja evaluasi metode dalam tahap ini menggunakan Teknik URL parsing [15]. URL parsing adalah proses analisis dan pemrosesan URL (Uniform Resource Locator) untuk mendapatkan informasi seperti protokol, domain, path, parameter, dan fragment. Hal ini umumnya digunakan dalam pengembangan website dan aplikasi untuk memahami dan memanfaatkan informasi yang terkandung dalam URL dengan menggunakan library python URLlib dapat mengubah Link dana phising menjadi berikut:

TABEL II
PARSING DOMAIN

Link Sebelum di *Parsing*	Link Setelah di Parsing
https://claim-danakaget.pictte.com/kaget=s6j7sxzeazg&r=iVlQUs/	https://claim-danakaget.pictte.com
https://danakagetvezridd.vercel.app/kaget=s6j7sxzeazg&riVlQUs.HTML	https://danakagetvezridd.vercel.app
https://danakaget-3.id-com.tech/kaget=s6j7sxzeazg&r=iVlQUs/	https://danakaget-3.id-com.tech

3. Transformasi data ke vector menggunakan *CountVectorizer*

   CountVectorizer melakukan tokenisasi, yang merupakan proses membagi teks menjadi kata-kata atau frasa. Kemudian, untuk setiap kata yang ditemukan dalam teks, CountVectorizer menghitung berapa kali kata tersebut muncul dalam teks tersebut [16].

   Jumlah dimensi CountVectorizer (485, 632) dari data X yang berisi kumpulan URL menunjukkan bahwa terdapat total 485 sampel (baris) dan 632 fitur (kolom) yang dihasilkan setelah proses vektorisasi. Setiap fitur mewakili

satu kata unik yang muncul dalam kumpulan URL tersebut. Dengan demikian, masing-masing URL diwakili oleh vektor dengan panjang 632, di mana setiap elemen vektor menunjukkan frekuensi kemunculan kata tertentu dalam URL tersebut. Hasil ini memungkinkan analisis lebih lanjut terhadap pola atau karakteristik yang terkandung dalam kumpulan URL tersebut, seperti frekuensi kata kunci, topik, atau tren tertentu. Hasilnya adalah vektor di mana setiap elemen mewakili frekuensi kata dalam teks. Sebagai contoh dari Link phising yang semula seperti berikut

https://claim-danakaget.pictte.com.
https://danakagetvezridd.vercel.app.
https://danakaget-3.id-com.tech.

Setelah melakukan proses Vectorisasi data diatas diubah menjadi sebuah vector seperti data dibawah ini

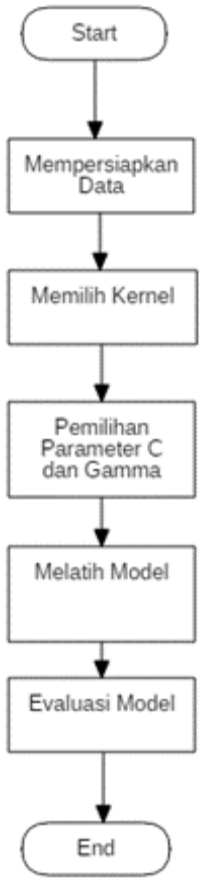
TABEL III
VEKTORISASI DATA

app	claim	com	danakaget	danakagetvezridd	https	id	pictte	tech	vercel
0	1	1	1	0	1	0	1	0	0
1	0	0	0	1	1	0	0	0	1
0	0	1	1	0	1	1	0	1	0

D. Modelling

Tahap modeling dalam pemrosesan data menggunakan Support Vector Machine (SVM) dimulai dengan proses training pada data latih. Pada tahap ini, dilakukan pembelajaran model SVM untuk membangun sebuah model yang dapat mengklasifikasikan data dengan akurasi tinggi. Langkah pertama yang harus dilakukan adalah menentukan tipe kernel yang akan digunakan, seperti linear, polynomial, radial basis function (RBF), atau sigmoid. Pemilihan kernel sangat penting karena mempengaruhi kemampuan model dalam menangani data yang tidak linear. Selain itu, nilai parameter seperti C (regularization parameter) dan gamma (untuk kernel RBF) juga harus ditentukan. Parameter-parameter ini berperan dalam mengatur keseimbangan antara kompleksitas model dan kinerjanya terhadap data latih.

Setelah menentukan tipe kernel dan nilai parameter, dilakukan proses pencarian parameter terbaik menggunakan algoritma gridsearch. Gridsearch adalah teknik yang memungkinkan pencarian kombinasi parameter yang menghasilkan model terbaik berdasarkan kinerja pada data validasi. Dengan mencoba berbagai kombinasi parameter, gridsearch memastikan bahwa model yang dipilih adalah yang memiliki performa optimal. Setelah menemukan parameter yang tepat, model terbaik ini kemudian digunakan untuk melakukan peramalan atau prediksi pada data baru. Dengan demikian, tahap modeling yang melibatkan training SVM, penentuan kernel dan parameter, serta optimasi melalui gridsearch menjadi kunci dalam membangun model SVM yang robust dan akurat untuk aplikasi prediksi.



Gbr. 4 Tahap Modelling SVM

E. Evaluasi

Evaluasi performa klasifikasi dilakukan dengan menggunakan metrik akurasi, presisi, recall, dan F1-score. Akurasi mengukur tingkat keberhasilan model dalam memprediksi secara keseluruhan dengan membandingkan jumlah prediksi benar dengan total prediksi. Namun, untuk mendapatkan pemahaman yang lebih komprehensif, digunakan pula presisi, recall, dan F1-score. Presisi mengukur proporsi prediksi positif yang benar, sementara recall mengukur proporsi hasil positif yang benar terhadap total hasil yang seharusnya positif. F1-score, yang merupakan rata-rata harmonik dari presisi dan recall, memberikan evaluasi yang seimbang terutama pada dataset yang tidak seimbang dalam distribusi kelasnya. Metrik-metrik ini membantu memahami kemampuan model dalam menangani tantangan klasifikasi dan mengoptimalkan keseimbangan antara prediksi positif dan negatif, memastikan prediksi yang akurat dan relevan untuk aplikasi praktis [17].

F. Deployment

Penelitian ini mengadopsi Metode Rapid Application Development (RAD) sebagai pendekatan utama dalam

pengembangan perangkat lunak. RAD adalah metode pengembangan perangkat lunak yang menekankan proses iteratif dan inkremental, memungkinkan pengembang untuk menciptakan prototipe dengan cepat dan melibatkan pengguna akhir secara intensif [18]. Metode RAD fokus pada pemahaman yang mendalam terhadap kebutuhan pengguna dengan melibatkan mereka dalam seluruh siklus pengembangan. Pendekatan ini memungkinkan perubahan cepat berdasarkan umpan balik dari pengguna, yang berkontribusi pada akurasi implementasi yang lebih tinggi dan kepuasan pengguna yang lebih baik. Selain itu, RAD juga dikenal karena kemampuannya dalam merespons perubahan kebutuhan dengan fleksibilitas tinggi. Dengan memecah proyek menjadi iterasi kecil, pengembang dapat dengan cepat menyesuaikan perangkat lunak dengan perubahan kebutuhan bisnis atau pemangku kepentingan. Kecepatan pengembangan yang tinggi dan keterlibatan pengguna yang aktif membuat Metode RAD menjadi pilihan yang sesuai untuk proyek-proyek yang membutuhkan perubahan cepat dan umpan balik langsung dari pengguna.

III. *HASIL DAN PEMBAHASAN*

A. *Crossvalidation* dan *Tunning Hyperparameter*

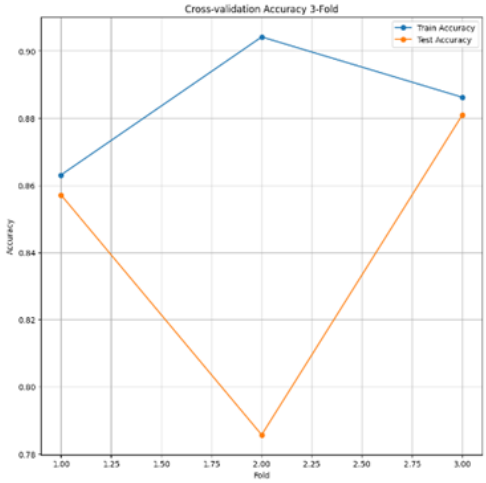
Pada tahap ini, dilakukan pengecekan akurasi menggunakan metode cross-validation (CV) pada model Support Vector Machine (SVM) dengan menggunakan 3, 5, dan 10 fold pada data phishing dan non-phishing. Proses ini menggunakan parameter terbaik yang telah ditemukan sebelumnya melalui grid search CV. Tujuan dari cross-validation ini adalah untuk memastikan bahwa model yang dikembangkan mampu memprediksi dengan konsisten dan efektif dalam memisahkan data phishing dari non-phishing. Dalam setiap iterasi, satu fold dari dataset digunakan untuk pengujian, sementara sisa fold digunakan untuk pelatihan model. Fold yang menghasilkan akurasi tertinggi akan dipilih untuk digunakan dalam proses selanjutnya. Hal ini penting untuk menjamin bahwa model memiliki performa terbaik dan mampu menangani berbagai variasi dalam data.

Dalam cross-validation, "fold" adalah bagian dari dataset yang digunakan secara bergantian untuk melatih dan menguji model. Dalam k-fold cross-validation, dataset dibagi menjadi k bagian yang sama besar. Pada setiap iterasi, satu fold digunakan sebagai data uji, sementara k-1 fold lainnya digunakan sebagai data latih. Proses ini diulang k kali sehingga setiap fold digunakan sebagai data uji satu kali. Hasil dari k iterasi ini kemudian dirata-rata untuk memberikan estimasi kinerja model yang lebih stabil dan akurat. Penggunaan fold dalam cross-validation membantu mengurangi risiko overfitting, di mana model terlalu sesuai dengan data latih sehingga kurang efektif pada data baru, dan underfitting, di mana model terlalu sederhana dan tidak menangkap pola dalam data. Dengan demikian, cross-validation menjadi metode yang efektif untuk mengevaluasi dan memastikan kemampuan generalisasi model pada data baru.

Gbr. 5 Tahap Crossvalidation

Dari implementasi tahapan pada Gambar 5 berikut hasil yang didapatkan.

1. Fold 3



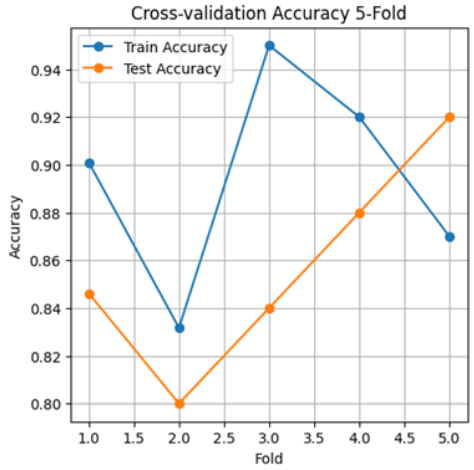
Gbr. 6 Crossvalidation Fold 3

Dari Gambar 6, dapat dilihat rincian hasil akurasi dari cross-validation dengan fold 3. Proses cross-validation ini menunjukkan bagaimana dataset dibagi menjadi tiga bagian yang masing-masing digunakan secara bergantian sebagai data uji dan data latih.

TABEL III
CROSSVALIDATION FOLD 3

Fold ke	Train	Test
1	0.85	0.85
2	0.90	0.78
3	0.88	0.88
Mean	**0.88**	**0.84**

2. Fold 5



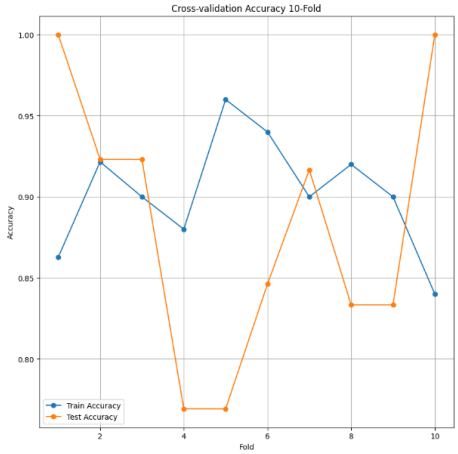
Gbr. 7 Crossvalidation Fold 5

Dari Gambar 7 tersebut berikut rincian hasil akurasi dari Crossvalidation fold 5

TABEL V
CROSSVALIDATION FOLD 5

Fold ke	Train	Test
1	0.90	0.84
2	0.83	0.80
3	0.95	0.84
4	0.92	0.88
5	0.87	0.92
Mean	**0.89**	**0.85**

3. Fold 10



Gbr. 8 Crossvalidation Fold 10

Dari Gambar 8 tersebut berikut rincian hasil akurasi dari Crossvalidation fold 10

TABEL VI
CROSSVALIDATION FOLD 10

Fold ke	Train	Test
1	0.86	1.00
2	0.92	0.92
3	0.90	0.92
4	0.88	0.76
5	0.96	0.76
6	0.88	0.84
7	0.96	0.91
8	0.94	0.83
9	0.92	0.83
10	0.80	1.00
Mean	**0.90**	**0.87**

Hasil cross-validation 10 fold menunjukkan parameter terbaik dari grid search: {'clf__C': 10, 'clf__kernel': 'linear', 'vect__max_features': 1000, 'vect__ngram_range': (1, 1)}. Dengan parameter ini, akurasi rata-rata pada data training adalah 90.24%, dan pada data uji adalah 88.14%. Meskipun ada variasi dalam akurasi di setiap lipatan, namun model menunjukkan performa yang baik dalam menggeneralisasi pada data baru. Dari cross-validation dengan fold 3, 5, dan 10, disimpulkan bahwa model stabil dengan kenaikan akurasi seiring peningkatan jumlah fold. Pada fold 3, akurasi train adalah 88% dan test 84%, menunjukkan kemampuan model dalam mempelajari pola pada data training. Dengan fold 5, akurasi meningkat menjadi 89% (train) dan 85% (test), menandakan peningkatan generalisasi. Pada fold 10, akurasi train mencapai 90% dan test 88%, menunjukkan kemampuan model dalam menggeneralisasi pada data baru. Dengan demikian, model ini konsisten dan memiliki potensi untuk diterapkan pada data baru.

B. Menyimpan Model dan menyambungkan model ke *Framework*

Menyimpan model machine learning adalah tahap penting dalam pengembangan model. Setelah pelatihan dan evaluasi, model disimpan untuk penggunaan ulang tanpa pelatihan ulang. Proses ini melibatkan serialisasi model dan parameter ke format berkas seperti .pkl, .h5, atau .joblib. Model disimpan sebagai berkas pickle/pkl, dan untuk menyambungkannya ke Flask, model dimuat menggunakan pickle.load(). Kemudian, model diintegrasikan ke aplikasi Flask dengan menentukan endpoint untuk menerima permintaan prediksi dari pengguna. Fungsi view Flask menggunakan model untuk membuat prediksi berdasarkan input pengguna, dan hasilnya dikembalikan sebagai respons HTTP kepada pengguna.

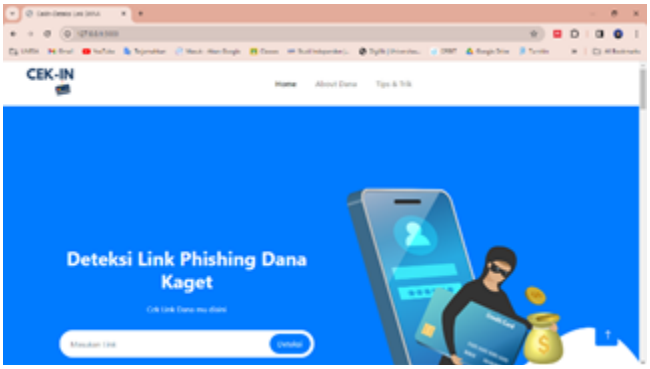
C. Membuat *Route Website*

Dalam pengembangan aplikasi website dengan Flask, pembuatan rute-rute untuk menangani permintaan HTTP dari pengguna sangat penting. Ini melibatkan pembuatan rute khusus untuk berbagai fitur aplikasi, seperti halaman beranda,

formulir, dan lainnya. Setiap rute akan terhubung ke fungsi view yang akan dijalankan saat rute tersebut diakses oleh pengguna. Fungsi-fungsi view ini akan memproses permintaan pengguna, melakukan operasi tertentu seperti mengambil data dari database, memproses formulir, atau menampilkan halaman website tertentu. Setelah semua rute dan fungsi view telah ditentukan, server Flask diaktifkan dengan menggunakan app.run(), sehingga aplikasi siap menerima permintaan HTTP dari pengguna dan memberikan respons sesuai dengan logika yang telah diimplementasikan dalam fungsi-fungsi view tersebut. Berikut algoritma yang digunakan.

D. Membuat Tampilan Home/Prediksi

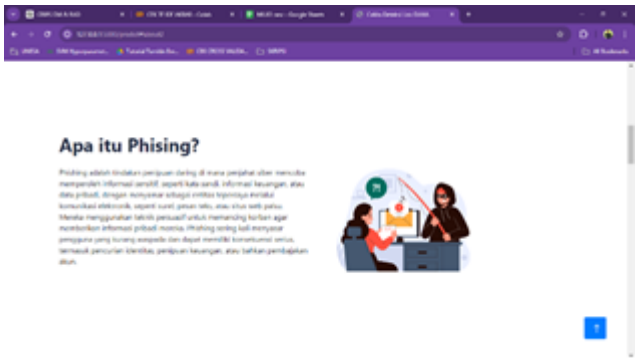
Tampilan website deteksi Link phishing yang memungkinkan pengguna untuk memasukkan Link dan kemudian mendapatkan prediksi phishing atau non-phishing menggunakan metode Support Vector Machine (SVM) adalah sebuah antarmuka yang sederhana namun informatif. Pengguna akan disambut dengan sebuah kotak input di mana mereka dapat menyalin atau memasukkan Link yang ingin mereka periksa. Setelah memasukkan Link, ada tombol "Deteksi" yang memicu proses analisis. Setelah pengguna mengklik tombol tersebut, mereka akan melihat hasil prediksi di layar, yang mungkin berupa label "Phishing" atau "Non-Phishing" Desain yang responsif dan User -friendly adalah kunci untuk memastikan pengalaman pengguna yang optimal, dengan informasi yang disajikan dengan jelas dan ringkas. Dengan demikian, tampilan website ini memberikan alat yang berguna bagi pengguna untuk menguji dan mengkonfirmasi keamanan Link yang mereka hadapi secara online menggunakan metode SVM. Berikut merupakan tampilan awal atau tampilan home dari Websitesite Deteksi Link Phising Dana Kaget ini.



Gbr. 9 Tampilan Home/Prediksi

E. Membuat Tampilan About Phising

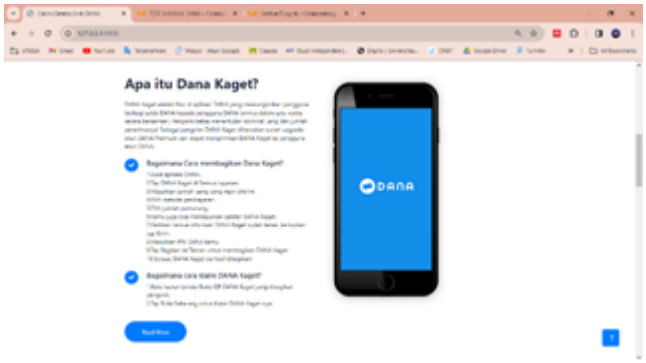
Pada sub bab ini, "About Phishing" disusun untuk memberikan pemahaman yang jelas tentang konsep phishing. Tampilan ini bertujuan sebagai sarana untuk memberikan pemahaman kepada pengguna tentang ancaman serius terkait keamanan digital. Pengertian tentang phishing akan dijelaskan secara singkat dan mudah dipahami, dengan fokus pada praktik penipuan daring di mana pelaku mencoba memperoleh informasi pribadi atau sensitif seperti kata sandi, nomor kartu kredit, atau data keuangan lainnya dengan menyamar sebagai entitas yang dapat dipercaya melalui komunikasi elektronik. Informasi yang terdapat dalam "About Phishing" dirancang untuk meningkatkan kesadaran pengguna tentang risiko phishing serta mendorong adopsi tindakan pencegahan yang efektif untuk melindungi diri dari potensi serangan phishing. Berikut penjelasan dari "About Phishing".



Gbr. 10 Tampilan About Phising

F. Membuat Tampilan About Dana

Halaman ini memberikan penjelasan menyeluruh tentang DANA Kaget, sebuah fitur di aplikasi DANA yang memungkinkan pengguna untuk berbagi saldo dengan pengguna lain secara bersamaan. Penjelasan tentang DANA Kaget meliputi definisi fitur tersebut, yaitu kemampuannya untuk mengirim saldo DANA kepada pengguna lain dalam jumlah yang ditentukan. Selain itu, halaman ini juga menyediakan langkah-langkah yang jelas tentang cara membagikan Link DANA Kaget, mulai dari membuka aplikasi DANA hingga memilih jumlah uang yang ingin dikirim dan metode pembayaran yang diinginkan . Selain itu, halaman ini juga memberikan instruksi tentang cara klaim saldo DANA Kaget dengan langkah-langkah yang terperinci, termasuk membuka tautan atau kode QR DANA Kaget yang dibagikan oleh pengirim, kemudian menekan tombol "Klaim Sekarang" untuk mendapatkan saldo yang dikirimkan . Berikut tampilan dari halaman About Dana Kaget.



Gbr. 11 Tampilan About Dana

G. Membuat Tampilan Tips and Trik

Halaman Tips and Trik memberikan panduan penting tentang cara menghindari jebakan phishing atau Link DANA Kaget palsu yang dapat mengancam keamanan finansial dan privasi pengguna. Salah satu tip yang disorot adalah berhati-hati saat menerima pesan atau email yang mencurigakan, terutama yang meminta informasi pribadi atau mengarahkan untuk mengklik tautan yang tidak dikenal. Selain itu, halaman tersebut juga mungkin menekankan pentingnya memverifikasi keaslian situs atau tautan sebelum memasukkan informasi sensitif atau melakukan transaksi keuangan secara online. Dengan memperhatikan tips dan trik yang diberikan dalam halaman tersebut, pengguna dapat meningkatkan kewaspadaan mereka terhadap ancaman phishing dan mengurangi risiko jatuh ke dalam perangkap penipuan finansial online. Berikut tampilan tips & trik nya.

Gbr. 12 Tampilan tips and trik

H. Akurasi SVM

Pada tahap evaluasi ini, dilakukan penilaian kinerja klasifikasi menggunakan metrik umum seperti akurasi, presisi, recall, dan F1-score. Metrik-metrik ini digunakan untuk mengukur seberapa baik model klasifikasi dapat memprediksi Link dana kaget. Salah satu metrik yang sering digunakan adalah akurasi, yang menghitung proporsi prediksi yang benar dibandingkan dengan total prediksi. Meskipun akurasi memberikan gambaran umum tentang performa model, terutama pada dataset yang seimbang, metrik lain seperti presisi, recall, dan F1-score memberikan pemahaman yang lebih mendalam tentang kemampuan model dalam mengklasifikasikan data. Berikut adalah hasil evaluasi kinerja menggunakan metode SVM:

```
Classification Report:
              precision    recall  f1-score   support

 Non Phising       0.90      0.95      0.93        20
     Phising       0.98      0.95      0.96        43

    accuracy                           0.95        63
   macro avg       0.94      0.95      0.95        63
weighted avg       0.95      0.95      0.95        63
```

Gbr. 13 Akurasi SVM

Evaluasi model menunjukkan kinerja yang baik dengan akurasi 0.95, artinya 95% prediksi model benar. Presisi untuk kelas Non Phising adalah 0.90 dan untuk kelas Phising 0.98, menunjukkan bahwa 90% dari prediksi Non Phising dan 98% dari prediksi Phising adalah benar. Recall untuk kedua kelas adalah 0.95, menunjukkan model mengidentifikasi 95% kasus dengan benar. F1-Score, yang menyeimbangkan presisi dan recall, adalah 0.93 untuk Non Phising dan 0.96 untuk Phising. Support menunjukkan ada 20 kasus Non Phising dan 43 kasus Phising dalam dataset. Rata-rata makro untuk precision, recall, dan F1-Score adalah 0.94, sedangkan rata-rata terbobot adalah 0.95, mencerminkan distribusi kelas dalam evaluasi metrik.

IV. KESIMPULAN

Model klasifikasi menggunakan Support Vector Machine (SVM) menunjukkan kinerja stabil dengan peningkatan akurasi pada data train dan test seiring bertambahnya jumlah fold dalam cross-validation. Pada fold 3, akurasi train mencapai 88% dan test 84%; dengan fold 5, akurasi train dan test meningkat menjadi 89% dan 85%; dan pada fold 10, akurasi train mencapai 90% dan test 88%, menandakan kemampuan generalisasi yang baik. Implementasi deteksi phishing menggunakan Flask menunjukkan solusi modern dan efisien untuk keamanan online, dengan pengujian black box memastikan akurasi deteksi serangan phishing, sehingga melindungi pengguna dari ancaman siber.

REFERENSI

[1] Bank Indonesia. (2023). Tinjauan Kebijakan Monoter.
[2] Sharoon. (2023). Punya Potensi Besar, DANA Sebut Pengguna dan Transaksinya Tumbuh pada 2023. Https://Keuangan.Kontan.Co.Id/News/Punya-Potensi-Besar-Dana-Sebut-Pengguna-Dan-Transaksinya-Tumbuh-Pada-2023
[3] Eshan, & Hasan. (n.d.). ICCIT : 2017 20th International Conference of Computer and Information Technology : 22-24 December 2017.
[4] Yudho, & Teddy. (2022). Ada 34.622 Kasus Phising di Indonesia Selama 5 Tahun Terakhir. Https://Www.Detik.Com/Jatim/Berita/d-6483650/Ada-34622-Kasus-Phising-Di-Indonesia-Selama-5-Tahun-Terakhir
[5] Maskun. (2017). Aspek Hukum penipuan berbasis internet. Keni Media.

[6] Opara, C., Chen, Y., & Wei, B. (2023). Look before you leap: Detecting phishing Website pages by exploiting raw URL and HTML characteristics. Expert Systems with Applications, 236. https://doi.org/10.1016/j.eswa.2023.121183

[7] Probierz, B., Stefanski, P., & Kozak, J. (2021). Rapid detection of fake news based on machine learning methods. Procedia Computer Science, 192, 2893–2902. https://doi.org/10.1016/j.procs.2021.09.060

[8] Winar Cahyo, P., Saidata Aesyi, U., Yani Yogyakarta Jl Siliwangi Ringroad Barat Banyuraden Gamping, A., & Siliwangi Ringroad Barat Banyuraden Gamping, J. (2023). Perbandingan LSTM dengan Support Vector Machine dan Multinomial Naïve Bayes pada Klasifikasi Kategori Hoax. 20(2), 23–29. https://doi.org/10.26623/transformatika.v20i2.5880 //

[9] Ritonga, A. S., & Purwaningsih, E. S. (2018). PENERAPAN METODE SUPPORT VECTOR MACHINE (SVM) DALAM KLASIFIKASI KUALITAS PENGELASAN SMAW (SHIELD METAL ARC WELDING). In Jurnal Ilmiah Edutic (Vol. 5, Issue 1).

[10] Ganesan. (2019). 10+ Examples for Using CountVectorizer. Https://Kavita-Ganesan.Com/How-toUseCountVectorizer/

[11] Han, J., Kamber, M., & Pei, J. (2011). Data Mining. Concepts and Techniques, 3rd Edition (The Morgan Kaufmann Series in Data Management Systems).

[12] Yudho, & Teddy. (2022). Ada 34.622 Kasus Phising di Indonesia Selama 5 Tahun Terakhir. Https://Www.Detik.Com/Jatim/Berita/d-6483650/Ada-34622-Kasus-Phising-Di-Indonesia-Selama-5-Tahun-Terakhir

[13] Suryadi. (2020). Phising: Pengertian, Cara Kerja dan Langkah Mengatasinya. Https://Www.Niagahoster.Co.Id/Blog/Mengatasi-Phishing/

[14] Winar Cahyo, P., Saidata Aesyi, U., Yani Yogyakarta Jl Siliwangi Ringroad Barat Banyuraden Gamping, A., & Siliwangi Ringroad Barat Banyuraden Gamping, J. (2023). Perbandingan LSTM dengan Support Vector Machine dan Multinomial Naïve Bayes pada Klasifikasi Kategori Hoax. 20(2), 23–29. https://doi.org/10.26623/transformatika.v20i2.5880

[15] Hermant. (2022). Node.js URL.parse(URLString, parseQueryString, slashesDenoteHost) API. Https://Www.Geeksforgeeks.Org/Node-Js-URL-ParseURLstring-Parsequerystring-Slashesdenotehost-Api/

[16] Kołcz, A., Prabakarmurthi, V., & Kalita, J. (2001). Summarization as feature selection for text categorization. International Conference on Information and Knowledge Management, Proceedings, 365–370. https://doi.org/10.1145/502585.502647

[17] Pupale Rushikesh. (2018). Support Vector Machines(SVM) — An Overview. Https://Towardsdatascience.Com/Https-Medium-Com-Pupalerushikesh-SVM-F4b42800e989.

[18] Rouse Margaret. (2020). What Does WEBSITE Mean? Https://Www.Techopedia.Com/Definition/5411/WEBSITE.