E-ISSN : 2686-5238
P-ISSN : 2686-4916
JURNAL EKONOMI MANAJEMEN SISTEM INFORMASI (JEMSI)
https://dinastipub.org/DIJMS dinasti.info@gmail.com +62 811 7404 455
DOI: https://doi.org/10.38035/jemsi.v5i4
Received: 3 April 2024, **Revised:** 17 April 2024, **Publish:** 19 April 2024

Analisis Performa Logistic Regression dan Support Vector Classification untuk Klasifikasi Email Phising

Brury Barth Tangkere[1]
[1] President University, Bekasi, Indonesia, brurytangkere@gmail.com

Corresponding Author: brurytangkere@gmail.com

Abstract: *Cyber security is something that is very important to pay attention to, especially with the rapid development of information and communication technology now. With the increasing development of the information and communications sector and easy access to information, it is important to be able to safeguard personal data so that it is not misused by irresponsible parties. Therefore, in this research, we will carry out a phishing email classification process to find out whether the email received is a safe email or not. In this research, a total of 18,650 data will be used, consisting of 11,322 secure email data and 7,328 phishing email data. To carry out the classification process, this research will use the Logistic Regression algorithm and Support Vector Machine. The purpose of using these two algorithms is to find which algorithm can carry out the phishing email classification process well. After carrying out classification testing, the results were that the classification process using Logistic Regression got an accuracy of 96.5% and classification with Support Vector Classification got an accuracy of 97.4%.*

Keyword: *Email Phishing, Machine Learning, Logistic Regression, Support Vector Classification.*

Abstrak: Keamanan siber merupakan suatu hal yang sangat penting diperhatikan terutama dengan maraknya perkembangan teknologi informasi dan komunikasi sekarang. Dengan semakin berkembangnya sektor informasi dan komunikasi serta mudahnya akses informasi, maka penting untuk dapat menjaga data pribadi sehingga tidak disalahgunakan oleh pihak tidak bertanggung jawab. Oleh karena itu, pada penelitian ini, akan melakukan proses klasifikasi email phising untuk mengetahui bahwa email yang diterima merupakan email yang aman atau tidak. Pada penelitian ini, akan menggunakan data dengan total sebanyak 18650 data yang dimana terdiri dari 11322 data email aman dan 7328 data email phising. Untuk melakukan proses klasifikasi, pada penelitian ini akan menggunakan algoritma Logistic Regression dan Support Vector Machine. Tujuan digunakannya kedua algoritma ini yaitu untuk menemukan mana algoritma yang dapat melakukan proses klasfikasi email phising dengan baik. Setelah dilakukannya pengujian klasifikasi, mendapatkan hasil bahwa proses klasifikasi dengan Logistic Regression mendapatkan akurasi sebesar 96.5% dan klasfikasi dengan Support Vector Classification mendapatkan akurasi sebesar 97.4%.

Kata Kunci: *Email Phishing, Machine Learning, Logistic Regression, Support Vector Classification.*

PENDAHULUAN

Zaman sekarang, teknologi informasi dan komunikasi sudah berkembang sangat pesat. Dengan berkembangnya teknologi informasi dan komunikasi, maka kita dapat dengan mudah untuk mendapatkan informasi dari sumber mana saja di internet [1]. Dengan mudahnya akses informasi dari internet, maka kita harus dapat juga menjaga data pribadi kita sehingga dapat aman dari kejahatan siber yang dapat terjadi internet dan dapat membahayakan data pribadi kita [2][3]. Karena dengan adanya kejahatan siber yang terjadi, hal tersebut dapat menyebabkan data diri kita disalahgunakan oleh pihak yang tidak bertanggung jawab, sehingga kita dapat mengalami kerugian dan pemerasan [4]. Salah satu kejahatan siber yang dapat terjadi yaitu melalui sarana email. Banyaknya kehajatan dengan menggunakan email yaitu karena email merupakan sarana berkomunikasi secara formal dan penting, sehingga dengan melakukan phising email, pihak pelaku dapat meraup informasi penting bahkan dari Perusahaan besar, dengan proses penyebaran menggunakan bentuk pengiriman dokumen atau link berbahaya [5]. Oleh karena itu, penting untuk kita dapat menjaga informasi kita dan mengantisipasi email phising yang dapat terjadi.

Regresi logistik adalah metode analisis statistik untuk memprediksi hasil biner, seperti ya atau tidak, berdasarkan pengamatan sebelumnya dari kumpulan data [6]. Model regresi logistik menggunakan fungsi logistik, atau fungsi logit, dalam matematika sebagai persamaan antara x dan y. Algoritma SVC (Support Vector Classification) merupakan pengembangan dari algoritma SVM (Support Vector Machine) yang digunakan untuk klasifikasi data. Dalam algoritma SVC, data dikelompokkan berdasarkan jarak antara data dengan centroid atau pusat kelompok. Algoritma ini menggunakan teknik kernel untuk mengubah data ke dalam dimensi yang lebih tinggi sehingga data dapat dipisahkan dengan lebih baik [7]. Information Retrieval (IR) merupakan suatu sistem atau metode yang digunakan untuk mengidentifikasi dan mengambil informasi yang sesuai dengan kebutuhan pengguna dari suatu set data yang besar, seperti dokumen, halaman web, atau objek multimedia [8]. IR melibatkan serangkaian langkah, termasuk representasi, penyimpanan, pengaturan, dan akses terhadap berbagai item informasi tersebut. Proses IR dapat dilakukan melalui penerapan teknik-teknik khusus, seperti tokenisasi, indexing, dan weighting.

Pada penelitian ini, akan dilakukan proses klasifikasi email phising berdasarkan data yang didapatkan pada body email. Tujuan dilakukannya penelitian ini, yaitu untuk membangun model yang dapat melakukan proses klasifikasi email phising, sehingga dengan adanya penelitian ini, maka dapat membantu dalam melakukan proses klasifikasi email, sehingga pengguna dapat mengetahui email yang diterima merupakan email yang aman atau email phishing. Pada penelitian ini, proses klasifikasi akan menggunakan algoritma Logistic Regression dan Support Vector Classification, tujuan digunakannya kedua algoritma ini yaitu untuk menemukan mana algoritma yang paling baik untuk melakukan proses klasifikasi email phishing. Tujuan digunakannnya logistic regression yaitu karena bagus digunakan untuk melakukan proses klasifikasi biner, sedangkan digunakannya Support Vector Classification yaitu karena algoritma SVC ini merupakan bentuk implementasi dari algoritma Support Vector Machine yang dimana khsusus digunakan untuk melakukan tugas klasifikasi data, sehingga diharapkan menghasilkan akurasi yang maksimal.

Penelitian terdahulu yang dilakukan oleh aufar et. al [9] pada tahun 2020 membahas mengenai proses analisis sentiment berdasarkan komentar pada social media youtube dengan menggunakan algoritma decision tree dan random forest. Tujuan dari penelitian ini yaitu untuk mempermudah melakukan analisis komentar positif atau negative dari suatu komentar yang ada pada media sosial youtube. Hasil dari penelitian ini dengan menggunakan split data

70% training dan 30% testing, mendapatkan hasil pengujian yaitu Algoritma Decision Tree mendapatkan akurasi sedikit lebih baik daripada Algoritma Random Forest. Algoritma Decision Tree mendapatkan akurasi sebesar 89,4%, sedangkan Random Forest mendapatkan akurasi sebesar 88,2%. Penelitian yang dilakukan oleh Jehad et. al [10] pada tahun 2020 membahas mengenai proses klasifikasi berita hoax dengan menggunakan random forest dan decision tree. Tujuan dilakukannya penelitian ini yaitu karena Banyaknya berita hoax yang tersebar, sehingga dengan dilakukannya penelitian ini diharapkan dapat membantu pengguna dalam idenfitikasi berita hoax. Hasil dari penelitian ini adalah algoritma decision tree bekerja lebih baik daripada random forest pada segi akurasi, dimana akurasi dari decision tree mendapatkan akurasi sebesar 89,11% sedangkan pada algoritma random forest mendapatkan akurasi sebesar 84,97%.

METODE

Dataset

Dataset yang akan digunakan pada penelitian ini yaitu merupakan dataset email phising yang didapatkan dari website kaggle.com dengan judul Phising Email Detection. Dengan total data yang digunakan yaitu sebanyak 18650 data email phising yang terdiri dari 11322 data safe email dan 7328 data phising email. Untuk visualisasi dataset diberikan pada gambar

	Email Text	Email Type
0	re : 6 . 1100 , disc : uniformitarianism , re ...	Safe Email
1	the other side of * galicismos * * galicismo *...	Safe Email
2	re : equistar deal tickets are you still avail...	Safe Email
3	\nHello I am your hot lil horny toy.\n I am...	Phishing Email
4	software at incredibly low prices (86 % lower...	Phishing Email

Gambar 1. Dataset Penelitian

Dari total 18650 data email tersebut, akan dibagi menjadi 70% data pelatihan dan 30% data pengujian. Yang dimana data pelatihan digunakan agar model dapat berlatih mengenali pola dari data, sehingga dapat menjadi sebuah model yang prediktif. Sedangkan, data pengujian berguna untuk melakukan proses evaluasi performa model dalam melakukan proses klasifikasi.

TF-IDF Vectorizer

TF-IDF adalah metode penentuan nilai frekuensi kata dalam dokumen dengan memberikan bobot pada kata kunci di setiap kategori [11]. Prosesnya melibatkan lima tahap preprocessing, seperti pemecahan kalimat, case folding, tokenizing, filtering, dan stemming. Jadi, TF-IDF merupakan metode pembobotan kata berdasarkan statistik kemunculan kata dan tingkat kepentingan dokumen yang mengandungnya, sementara TF-IDF Vectorizer menghitung bobot frekuensi kata dalam dokumen dan mengonversinya menjadi bobot TF-IDF. Tujuan dari proses konversi data menjadi bentuk vector space model yaitu agar data teks yang diberikan dapat diproses dengan menggunakan algoritma machine learning. TF atau Term Frecuency merupakan metode yang digunakan untuk melihat seberapa sering data muncul pada dokumen dan IDF digunakan untuk mencari seberapa penting kata secara global dalam dokumen. Untuk rumus perhitungan matematis TF-IDF diberikan pada poin 1, 2 dan 3.

$$TF(k, dok) = \frac{jml\_kata\_dlm\_dok}{total\_kata\_dlm\_dok} \quad (1)$$

$$IDF(k, korpus) = log\left(\frac{total\_dok\_dlm\_korpus}{jml\_dok\_ada\_korpus + 1}\right) + 1 \quad (2)$$

$$TF_{IDF(k,dok,korpus)} = TF(k, dok) * IDF(k, korpus) \quad (3)$$

Logistic Regression

Logistic Regression merupakan metode klasifikasi terpandu yang menonjol dalam memprediksi probabilitas diskrit dengan kinerja yang unggul. Dalam implementasinya, regresi logistik menggunakan fungsi logistik untuk mengukur nilai probabilitas suatu kejadian, menghasilkan output biner 0 atau 1 [12]. Fungsi logistik, atau sigmoid function, mengubah nilai dari rentang minus tak hingga hingga plus tak hingga menjadi rentang antara 0 dan 1, memungkinkan interpretasi output sebagai probabilitas kejadian positif. Regresi logistik digunakan ketika model perlu memprediksi kemungkinan kejadian dua kelas yang berbeda.

Pentingnya regresi logistik tidak hanya terletak pada kemampuannya memprediksi, tetapi juga dalam memberikan wawasan tentang kontribusi setiap fitur terhadap probabilitas hasil positif. Variabel dependen yang bersifat biner dalam regresi logistik memungkinkan analisis tentang kekuatan dan arah hubungan antara variabel independen dan variabel dependen. Untuk psudocode dari logistic regression diberikan dibawah ini.

Prudocode:

Define param ; Define w (weight), b (bias), ∝ (learn rate), iteration
Logistic Function ; 1. σ(x)= 1/((1+e^(-z)))
iteration training start ; 2. Repeat for iteration
Calculate combination ; (1) x = w * feature + b
apply function ; (2) prob = σ(x)
calculate loss function ; (3) loss = -(target*log(prob)+(1-target)*log(1-prob))
Calculate gradients ; (4) dw = 1/(total data)*∑((prob-target)*feature) ; (5) db = 1/(total data)*∑(prob-target)
Update weight ; (6) w = w * α * dw
Update bias ; (7) b = b * α * db
Prediction ; 3, if prob > 0.5 then 1 else 0

Support Vector Classification

Support Vector Classification (SVC) merupakan algoritma di bidang machine learning yang digunakan untuk melakukan tugas klasifikasi [13]. Cara kerja algoritma ini adalah dengan mencari hyperplane terbaik yang dapat memisahkan dua kelas dalam ruang fitur. Hyperplane ini dipilih sedemikian rupa sehingga jarak minimum antara hyperplane dan support vectors (titik-titik terdekat dari kedua kelas) maksimum. Dengan kata lain, tujuan SVC adalah menciptakan batas keputusan optimal untuk memisahkan antara berbagai kelas.

SVC menggunakan konsep fungsi kernel untuk mentransformasi data ke dimensi yang lebih tinggi, di mana pola-pola yang kompleks dapat lebih mudah dipisahkan. Fungsi kernel memungkinkan SVC untuk menangani data yang memiliki hubungan antar fitur yang tidak linear. Selain itu, SVC memiliki parameter penalti (C) yang mengendalikan trade-off antara mencapai margin maksimum dan mengizinkan kesalahan klasifikasi. Penyetelan parameter ini perlu dilakukan dengan hati-hati karena dapat memengaruhi performa model SVC. Untuk psudocode SVC diberikan dibawah ini:

Psudocode:

Define param ; Define w (weight), b (bias), ∝ (learn rate), Z (regularization param)
Iteration training start ; (1) Repeat until convergence (i)
Calculate margin ; (1) Margin = m(i) * (w * n(i) * b) ; (2) if Margin < 1 then
update w ; (1) w = w+ α*(m(i)*n(i)-2*Z*w)
update b ; (2) b = b+ α*m(i) ; (3) else
update w ; (1) w = w+ α*(-2*Z*w)

```
# update b ; (2) b = b
# predict new x ; (2) f(x) = w * x * b
# prediction ; (3) if f(x) > 0 then 1 else 0
```

Confusion Matrix

Confusion matrix adalah metrik evaluasi yang digunakan untuk mengukur performa model machine learning yang telah dikembangkan [14]. Output dari confusion matrix meliputi variabel precision, recall, f1-score, dan support untuk setiap kelas setelah melalui proses pelatihan dan pengujian dengan algoritma yang telah ditentukan. Confusion matrix memiliki empat nilai utama, yaitu True Positive, True Negative, False Positive, dan False Negative. Dari kombinasi nilai-nilai ini, kita dapat menghitung performa model dengan mengukur precision, yaitu kemampuan model untuk mengidentifikasi kasus positif secara akurat; recall, yang menunjukkan seberapa baik model dalam menemukan semua kasus positif dari seluruh iterasi yang dilakukan; serta f1-score, yang mewakili nilai harmonik atau rata-rata perhitungan antara recall dan precision. Rincian perhitungan untuk precision, recall, dan f1-score dijelaskan pada poin 4, 5, dan 6.

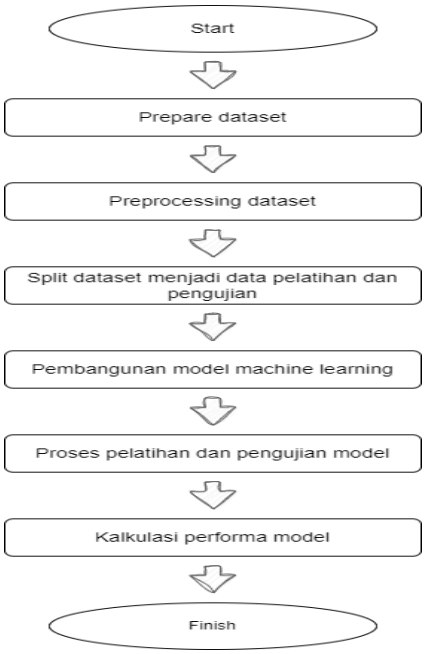
$$Precision = \frac{TP}{(TP + FP)} \quad (4)$$

$$Recall = \frac{TP}{(TP + FN)} \quad (5)$$

$$F1 - score = \frac{2 * (Precision * Recall)}{(Precision + Recall)} \quad (6)$$

Alur Kerja Proses

Dalam penelitian ini, proses implementasi dan pengembangan model untuk sistem klasifikasi email phishing akan menggunakan bahasa pemrograman python dan menggunakan Jupyter notebook untuk melakukan penulisan code. Tujuan digunakannya python untuk implementasi sistem yaitu karena python sangat mendukung dalam pembangunan machine learning melalui library library yang telah disediakan, sehingga dapat mempersingkat dan effisiensi waktu pelatihan dan juga pengujian model. Untuk alur proses yang dilakukan dalam penelitian ini yaitu diberikan pada gambar 2.



Gambar 2. Alur kerja proses klasifikasi

Gambar 2 menunjukkan alur kerja yang dilakukan dalam proses klasifikasi email phising. Untuk penjelasan detail langkah langkah klasifikasi diberikan dibawah ini.

Pertama tama, akan dilakukan proses menyiapkan dataset yang akan digunakan untuk proses klasifikasi. Pada penelitian ini, akan menggunakan dataset publik yang didapatkan dari kaggle.com.

Selanjutnya, setelah menyiapkan dataset, maka akan melakukan proses menghilangkan stop words dan mengubah dataset menjadi bentuk vektor. Tujuan dihilangkannya stops words yaitu untuk membantu model dapat bekerja lebih efisien. Sedangkan tujuan dari pengubahan dataset teks menjadi bentuk vector space model yaitu karena model machine learning yang dibangun hanya dapat memproses data data numerik, sehingga penting untuk diubah dari dataset yang berbentuk dokumen teks menjadi bentuk vektor.

Setelah dilakukan preprocessing dataset, maka langkah selanjutnya yang dilakukan yaitu membagi dataset menjadi data pelatihan dan data pengujan. Pada proses penelitian ini, data yang digunakan akan dibagi menjadi 70% data pelatihan dan 30% data pengujian.

Lalu setelah dilakukan pembagian dataset, maka selanjutnya akan melakukan proses pembangunan model machine learning yang akan dibangun, yaitu Logistic Regression dan Support Vector Classification.

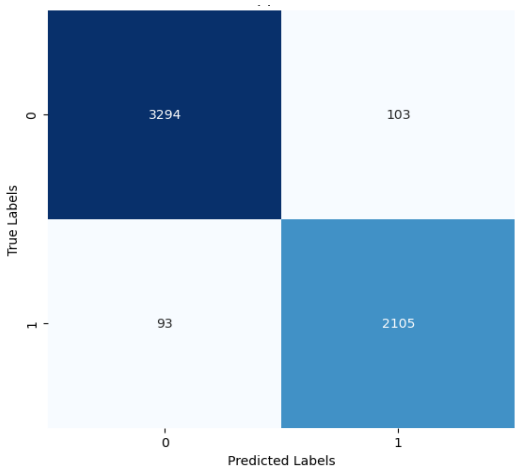
Setelah dilakukan preprocessing dataset dan pembangunan model machine learning yang akan digunakan, maka selanjutnya akan melakukan proses pelatihan untuk melatih model mengenali pola dari dataset dan pengujian model untuk menguji performa model dalam melakukan proses klasifikasi data.

Lalu selanjutnya, setelah melakukan proses pengujian model, maka akan melakukan proses kalkulasi performa model dengan menggunakan confusion matrix, sehingga dapat dilakukan evaluasi performa model dalam melakukan proses klasifikasi data.

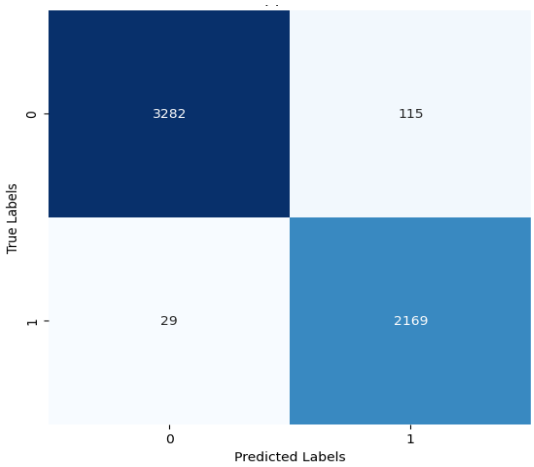
HASIL DAN PEMBAHASAN

Hasil

Pada penelitian ini, akan menggunakan bahasa pemrograman python dan IDE Jupyter notebook untuk implementasi dan modelling sistem klasifikasi email phishing. Setelah dataset disiapkan, maka akan dilakukan proses menghilangkan stop words atau kata kata yang kurang berguna untuk pembangunan suatu kalimat. Tujuan dihilangkan stop words ini yaitu untuk menghemat waktu proses pelatihan dan pengujian model. Karena, dengan dihilangkannya stop words, maka akan mengurangi jumlah kata yang nantinya akan di proses dari teks email. Setelah dilakukan proses menghilangkan stop words, maka selanjutnya akan melakukan proses mengubah data teks menjadi bentuk vektor. Tujuan dilakukan proses pengubahan data menjadi bentuk vektor yaitu agar dataset yang sudah diproses sebelumnya dapat digunakan untuk proses pelatihan dan pengujian machine learning, karena model machine learning yang dibangun hanya dapat melakukan proses pelatihan pengenalan pola dengan menggunakan data numerik, sehingga penting untuk dilakukan proses mengubah data teks menjadi data vektor. Selanjutnya setelah selesai melakukan proses preprocessing data, maka akan melakukan pembagian dataset menjadi 70% data pelatihan dan 30% data pengujian. Tujuan dilakukan pembagian data yaitu agar model dapat melakukan proses latihan mengenali pola yang ada pada data dengan menggunakan data pelatihan dan kita dapat melakukan proses evaluasi performa model dengan menggunakan data pengujian sehingga dapat melihat performa model dalam melakukan proses klasifikasi data. Setelah dilakukan proses pengujian, mendapatkan hasil berupa confusion matrix, yang dimana diberikan pada gambar 3 dan 4.



Gambar 3. Confusion Matrix hasil Logistic Regression



Gambar 4. Confusion Matrix hasil SVC

Gambar 3 dan 4 menunjukkan confusion matrix hasil pengujian yang sudah dilakukan dengan menggunakan algoritma Logistic Regression dan Support Vector Classification. Dapat dilihat pada gambar 3, merupakan confusion matrix hasil proses pengujian dengan menggunakan logistic regression. Dapat dilihat, bahwa pada gambar 3 menunjukkan model logistic regression dapat melakukan proses klasifikasi email phishing dengan baik. Hal tersebut dapat diketahui dari hasil prediksi salah yang hanya sedikit apabila dibandingkan dengan prediksi tepat model. Sedangkan, pada gambar 4 menunjukkan confusion matrix hasil proses pengujian dengan menggunakan Support Vector Classification. Dapat dilihat pada gambar 4, bahwa model klasifikasi yang dibangun dapat melakukan proses klasifikasi email phishing dengan optimal, yang dimana apabila dibandingkan dengan model Logistic Regression yang sebelumnya dibangun, model Support Vector Classification dapat menebak dan melakukan klasifikasi benar pada lebih banyak data, sehingga model Support Vector Classification yang dibangun hanya mendapatkan sedikit tebakan yang salah apabila dibandingkan hasil pengujian dengan model Logistic Reggression yang sebelumnya sudah diuji.

Pembahasan

Setelah dilakukan proses pengujian model, maka dapat dilihat bahwa model Logistic Regression dan Support Vector Classification yang dibangun dapat melakukan proses klasifikasi email phishing dengan optimal. Hal tersebut dapat dilihat dari hasil akurasi pengujian yang didapatkan setelah melakukan pelatihan model. Untuk hasil akurasi pengujian yang didapatkan diberikan pada tabel 1.

Tabel 1. Hasil Akurasi Pengujian

Model	Akurasi
Logistic Regression	96.5%
Support Vector Classification	97.4%

Tabel 1 menunjukkan hasil akurasi yang didapatkan setelah dilakukannya proses pengujian model machine learning dengan menggunakan model Logistic Regression dan Support Vector Classification yang dibangun pada penelitian ini. Dapat dilihat dari tabel 1, menunjukkan bahwa model Support Vector Classification yang dibangun dapat lebih optimal untuk melakukan proses klasifikasi email phishing apabila dibandingkan dengan model Logistic Regression yang dibangun. Hal tersebut dapat dikatakan karena model Support Vector Classification yang dibangun mendapatkan akurasi yang lebih baik 0.9% apabila dibandingkan dengan model Logistic Regression yaitu sebesar 97.4%. Selain dengan melihat nilai akurasi performa model, maka kita juga dapat mengukur performa klasifikasi model dengan menggunakan nilai presisi, recall dan juga f1-score yang didapatkan dari confusion matrix hasil pengujian. Untuk nilai presisi, recall dan f1-score yang didapatkan diberikan pada tabel 2.

Tabel 2. Hasil presisi, recall dan f1-score pengujian

Model	Presisi	Recall	F1-Score
LR	97.0%	96.0%	96.0%
SVC	97.0%	97.0%	97.0%

Tabel 2 menunjukkan nilai recall, presisi dan juga f1-score dari model Logistic Regression dan Support Vector Classification. Dapat dilihat pada tabel 2, bahwa model Support Vector Classification yang dibangun memiliki nilai presisi, recall dan f1-score yang paling optimal yaitu dengan masing masing nilai sebanyak 97.0%. Hal tersebut menunjukkan bahwa model Support Vector Classification yang dibangun memiliki keakuratan yang baik untuk dapat menebak data, memiliki performa yang baik untuk melakukan proses menebak semua kelas dengan benar, dan juga model Support Vector Classification yang dibangun memiliki nilai harmonik atau nilai keseimbangan yang baik antara nilai presisi dan juga recall.

KESIMPULAN

Setelah melalui proses pelatihan dan pengujian model dengan menggunakan algoritma Logistic Regression dan Support Vector Classification, maka dapat disimpulkan bahwa kedua model tersebut dapat melakukan proses klasifikasi dengan baik. Hal tersebut ditunjukkan bahwa dengan menggunakan model Logistic Regression mendapatkan akurasi pengujian sebesar 96.5%, sedangkan dengan menggunakan model Support Vector Classification mendapatkan akurasi pengujian sebesar 97.4%. Dari nilai nilai tersebut, maka dapat dilihat bahwa model Support Vector Classification mendapatkan akurasi yang paling optimal apabila dibandingkan dengan menggunakan algoritma Logistic Regression. Sehingga dengan hal tersebut, dapat disimpulkan bahwa model Suport Vector Classification dapat melakukan proses klasifikasi email phishing dengan lebih optimal. Untuk penelitian selanjutnya, diharapkan untuk dapat menggunakan algoritma lain untuk melakukan proses klasifikasi data, sehingga semakin banyak komparasi yang dapat dilakukan sehingga proses klasifikasi email phishing dapat lebih optimal. Untuk penelitian selanjutnya, diharapkan juga agar dapat melakukan proses penambahan parameter yang digunakan sehingga diharapkan proses klasifikasi yang dilakukan dapat lebih optimal.

REFERENSI

Azzani, I. K., Purwantoro, S. A., & Almubaroq, H. Z. (2023). Urgensi Peningkatan Kesadaran Masyarakat Tentang Kasus Penipuan Online Berkedok Kerja Paruh Waktu Sebagai Ancaman Negara. NUSANTARA: Jurnal Ilmu Pengetahuan Sosial, 10(7), 3556-3568. http://dx.doi.org/10.31604/jips.v10i7.2023.3556-3568

Arrasuli, B. K., & Fahmi, K. (2023). PERLINDUNGAN HUKUM POSITIF INDONESIA TERHADAP KEJAHATAN PENYALAHGUNAAN DATA PRIBADI. Unes Journal of Swara Justisia, 7(2), 369–392. https://doi.org/10.31933/ujsj.v7i2.351

Andriyanto, T. (2022). Komunikasi Termediasi Penipuan dengan Modus Business Email Compromise. Jurnal Riset Komunikasi, 5(2), 220-243. https://doi.org/10.38194/jurkom.v5i2.627

Wahyudi, W. R., Adriko, S. A., Firdaust, M. I., Harits, M. H. A., & Hapsari, D. P. (2023, April). Perbandingan Kinerja Algoritma Klasifikasi Naive Bayes, k-Nearest Neighbor dan Logistic Regression pada Dataset Multiclass. In Prosiding Seminar Nasional Teknik Elektro, Sistem Informasi, dan Teknik Informatika (SNESTIK) (Vol. 1, No. 1, pp. 380-386). https://doi.org/10.31284/p.snestik.2023.4157

Listanto, F. ., Fatchan, M. ., & Hadikristanto, W. (2023). Prediksi Defect Produk Casting Dengan Algoritma SVM Berbasis RBF dan Linier. Jurnal Ilmiah Intech : Information Technology Journal of UMUS, 5(2), 109–119. https://doi.org/10.46772/intech.v5i2.1376

AHMAD, N. ., PRASETYO, A. A. ., & MASRURI, A. . (2021). PENERAPAN INFORMATION RETRIEVAL PADA SEARCH ENGINE. KNOWLEDGE: Jurnal Inovasi Hasil Penelitian Dan Pengembangan, 1(1), 15-23. Retrieved from https://www.jurnalp4i.com/index.php/knowledge/article/view/771

M. Aufar, R. Andreswari and D. Pramesti, "Sentiment Analysis on Youtube Social Media Using Decision Tree and Random Forest Algorithm: A Case Study," 2020 International Conference on Data Science and Its Applications (ICoDSA), Bandung, Indonesia, 2020, pp. 1-7, doi: 10.1109/ICoDSA50139.2020.9213078.

Jehad, R., & A. Yousif, S. (2020). Fake News Classification Using Random Forest and Decision Tree (J48). Al-Nahrain Journal of Science, 23(4), 49–55. Retrieved from https://anjs.edu.iq/index.php/anjs/article/view/2306

Luthfiyah Annisa, & Anna Dina Kalifia. (2024). Analisis Teknik TF-IDF Dalam Identifikasi Faktor-Faktor Penyebab Depresi Pada Individu. Gudang Jurnal Multidisiplin Ilmu, 2(1), 302–307. https://doi.org/10.59435/gjmi.v2i1.249

Strzelecka, A., Kurdyś-Kujawska, A., & Zawadzka, D. (2020). Application of logistic regression models to assess household financial decisions regarding debt. Procedia Computer Science, 176, 3418–3427. doi:10.1016/j.procs.2020.09.055

Savitri, N. L. P. C., Rahman, R. A., Venyutzky, R., & Rakhmawati, N. A. (2021). Analisis Klasifikasi Sentimen Terhadap Sekolah Daring pada Twitter Menggunakan Supervised Machine Learning. Jurnal Teknik Informatika Dan Sistem Informasi, 7(1). https://doi.org/10.28932/jutisi.v7i1.3216

Siregar, Z., & Marpaung, T. B. (2020). Pemanfaatan Teknologi Informasi dan Komunikasi (TIK) Dalam Pembelajaran di Sekolah. BEST Journal (Biology Education, Sains and Technology), 3(1), 61-69. https://doi.org/10.30743/best.v3i1.2437

Saly, J. N., & Sulthanah, L. T. (2023). Pelindungan Data Pribadi dalam Tindakan Doxing Berdasarkan Undang-Undang Nomor 27 Tahun 2022. Jurnal Kewarganegaraan, 7(2), 1708-1713. https://doi.org/10.31316/jk.v7i2.5413

Uly, N., Hendry, H., & Iriani, A. (2023). CNN-RNN Hybrid Model for Diagnosis of COVID-19 on X-Ray Imagery: Hybrid Model CNN-RNN untuk Diagnosis COVID-19 pada Citra X-Ray. Digital Zone: Jurnal Teknologi Informasi Dan Komunikasi, 14(1), 57-67. https://doi.org/10.31849/digitalzone.v14i1.13668