

Klasifikasi Postingan Pengguna Facebook Untuk Deteksi Phising Menggunakan Naive Bayes

Muhammad Fahmi Fadhilah¹, Fikki Arsyi Nur Fadlillah²

¹Informatics, Bhayangkara Jakarta Raya University, Bekasi, INDONESIA

Email : fahvvif@gmail.com¹, fikkiarsyi@gmail.com²

ABSTRAKSI

Phishing merupakan penipuan digital yang umum dilakukan oleh penjahat siber dengan tujuan untuk mengambil data informasi pribadi pengguna dengan cara memanipulasi. Facebook adalah platform media social yang sangat populer di dunia sehingga bisa menjadi tempat yang basah bagi penjahat siber phishing. Pada penelitian kali ini, kami membangun model Klasifikasi untuk mengidentifikasi dan mencegah upaya phishing pada postingan facebook. Dataset yang digunakan dalam penelitian ini diperoleh dari posting pengguna facebook yang dikumpulkan. Pengolahan data dilakukan dengan melakukan preprocessing pada teks posting, termasuk penghapusan tanda baca dan kata kata yang tidak penting. Metode yang digunakan adalah Naïve Bayes untuk mengklasifikasikan posting kedalam kategori phishing atau tidak phishing. Metode Naïve Bayes digunakan karena kemampuannya dalam mengklasifikasikan data dengan tingkat Akurasi yang baik. Hal ini menunjukkan bahwa fitur fitur yang dipilih dalam penelitian ini dapat menjadi indicator yang kuat untuk mendeteksi phishing pada postingan pengguna facebook. Hasil penelitian menunjukkan Naïve Bayes dapat menjadi solusi yang efektif untuk deteksi phishing pada postingan pengguna facebook. Selain itu, hasil dari penelitian ini dapat memberikan wawasan yang berharga tentang ciri ciri umum dari postingan phishing pada Facebook. Dengan nilai akurasi sebesar 99,01% diharapkan penelitian ini dapat membantu meningkatkan kesadaran dan keamanan Pengguna Facebook terhadap postingan phishing.

Kata kunci: Klasifikasi, Postingan Pengguna, Deteksi Phising, Naïve Bayes.

ABSTRACT

Phishing is a digital fraud that is commonly carried out by cybercriminals with the aim of taking user's personal information data by manipulating it. Facebook is a very popular social media platform in the world so it can be a wet place for phishing criminals. In this research, we built a Classification model to identify and prevent phishing attempts on Facebook posts. The dataset used in this study was obtained from Facebook user posts collected. Data processing is done by preprocessing the post text, including removing punctuation marks and words that are not important. The method used is Naïve Bayes to classify posts into phishing or not phishing categories. The Naïve Bayes method is used because of its ability to classify data with a good level of accuracy. This shows that the features selected in this study can be a strong indicator for detecting phishing on Facebook user posts. The results of the study show that Naïve Bayes can be an effective solution for phishing detection on Facebook user posts. In addition, the results of this research can provide valuable insight into the common characteristics of phishing posts on Facebook. With an accuracy value of 99.01%, it is hoped that this research can help increase awareness and security of Facebook users against phishing posts.

Keywords: Classification, User Posting, Phishing Detection, Naïve Bayes.

Penulis Korespondensi

Muhammad Fahmi

Tanggal Submit : 02/04/2023

Tanggal Diterima : 04/05/2023

Tanggal Terbit : 12/07/2023

This is an open access article under the [CC-BY-NC-SA](https://creativecommons.org/licenses/by-nc-sa/4.0/) license



Copyright: © 2023 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 (CC BY-NC-SA 4.0) International License (<https://creativecommons.org/licenses/by-nc-sa/4.0/>).

Publisher's Note: JPPM stays neutral with regard to jurisdictional claims in published maps and institutional affiliations

I. PENDAHULUAN

Meningkatnya penggunaan sosial media untuk berkomunikasi secara online melalui pesan, postingan serta komentar terutama di Facebook, para pengguna facebook ini pun tidak semua mengerti betapa berbahayanya apabila kita terkena jebakan phishing yang di sebar melalui pesan, postingan maupun komentar, hal ini merupakan celah yang sangat besar dan dapat dengan mudah dimanfaatkan oleh para penjahat siber untuk mendapatkan informasi pribadi seperti email, kata sandi, nomer kartu kredit tanpa disadari oleh pengguna-nya karena salah memasuki website atau asal tekan link phishing yang muncul di beranda facebook-nya. Phishing merupakan serangan siber yang dilakukan oleh penipu dengan tujuan mencuri informasi pribadi pengguna, seperti kata sandi atau data keuangan, dan Facebook sering menjadi target utama serangan ini. Untuk melawan phishing di Facebook, perlu adanya metode deteksi yang efektif. Salah satu pendekatan yang populer adalah menggunakan metode klasifikasi Naive Bayes. .

Dengan menggunakan metode Naive Bayes, sistem dapat mempelajari pola-pola dalam postingan phishing dan non-phishing yang ada. Setelah melalui proses pelatihan, sistem dapat mengklasifikasikan postingan baru berdasarkan probabilitas menggunakan rumus Bayes. Hal ini membantu mengidentifikasi dan menghapus postingan phishing dengan lebih efisien, melindungi pengguna Facebook.

Dalam pengembangan metode ini, penting untuk terus meningkatkan akurasi dan kemampuan adaptasi sistem dalam mendeteksi serangan phishing baru yang muncul. Kecepatan dan efisiensi juga harus diperhatikan agar deteksi dapat dilakukan secara real-time dan tanggap terhadap serangan phishing di Facebook.

Tujuan penelitian ini yaitu untuk menentukan jenis dan sumber data yang tepat untuk melatih dan menguji model klasifikasi. Data posting pengguna Facebook yang telah diberi label phishing atau non-phishing digunakan sebagai sumber data. Label-label tersebut dapat diberikan oleh ahli keamanan siber atau dengan melakukan analisis manual terhadap posting yang ada. Data dapat diambil dari akun-akun Facebook pengguna yang berbeda. Selain itu, penelitian ini juga bertujuan untuk mengevaluasi metode analisis yang tepat untuk melakukan klasifikasi pada posting pengguna Facebook dengan algoritma Naive Bayes. Kata-kata kunci yang berkaitan dengan phishing dapat diidentifikasi dan diekstraksi menggunakan metode Bag of Words. Model Naive Bayes kemudian dilatih dengan data latih dan fitur-fitur yang telah diekstraksi. Akurasi, presisi, recall, dan F1-score digunakan untuk menguji keakuratan model tersebut dalam melakukan klasifikasi pada posting pengguna Facebook.

II. PENELITIAN YANG TERKAIT

Berdasarkan permasalahan ini, sangat diperlukan metode penyelesaian yang efektif dalam proses pengklasifikasian agar dapat memilih dan memilah berdasarkan jenis dan kategori yang ada. Solusi dari permasalahan ini adalah dibutuhkan sebuah machine learning yang mampu memahami dan melakukan proses pengelompokkan dari hasil data yang didapat agar nantinya dapat menjadi alat bantu dalam mendeteksi dan mengklasifikasikan jenis URLs berdasarkan serangannya. [1], Karakteristik phishing tersebut digolongkan menjadi empat golongan utama yaitu, Address Bar based Feature, Abnormal based Feature, HTML and Javascript based Features dan Domain based Feature [2], fitur dengan menggunakan metode MICe TICE memiliki nilai akurasi Naive Bayes, MICe TICE lebih tinggi 4.33% dari korelasi Spearman yaitu sebesar 92,98% [3], Metode yang saya gunakan adalah menggunakan Naive Bayes Classifier Algoritma.

Metode yang digunakan dalam penelitian ini adalah dengan melakukan percobaan terhadap data website yang akan segera diuji menggunakan Naive Bayes Classifier Algorithm yang baik, sehingga tingkat keamanan website dapat Dikenal [4], Proses pengumpulan data dilakukan dengan proses crawling data yang menggunakan perangkat lunak RapidMiner [5]. Pre-Processing merupakan tahapan yang bertujuan untuk menyeleksi data dari missing values sehingga mendapatkan data yang bersih dan siap digunakan [6],[8], Pada tahap ini dilakukan implementasi algoritma ke sistem yang dibuat. Algoritma yang diterapkan yaitu algoritma yang memiliki nilai akurasi terbaik [7],[9], Pada penelitian ini, dataset merupakan data publik yang menggunakan tiga kategori dalam penentuan website yaitu legitimate, suspicious, dan phishing [10], data scaling perlu dilakukan untuk memastikan validitas pemodelan prediktif, terutama ketika variabel atau fitur memiliki skala yang berbeda-beda. Salah satunya yaitu metode data standardization [11]. Serangan phishing seringkali dilakukan melalui email, email ini berisi tautan URL yang mengarahkan pengguna ke situs web lain [12], permasalahan ini dapat diminimalisir dengan membuat sebuah model anti spam yang bertujuan untuk mengklasifikasikan surel dan memberikan informasi terhadap pengguna surel apabila terdapat pesan yang diprediksi sebagai pesan spam [13],[14]. Dari hasil penelitian dengan dataset website phishing sebanyak 1.353 data algoritma Naive Bayes

menghasilkan model yang memiliki nilai akurasi yaitu sebesar 82,31% [15].

III. METODE PENELITIAN

Pada penelitian ini, dataset yang akan digunakan bersumber dari postingan pada laman Facebook baik dari laman pribadi ataupun laman grup umum melalui proses manual dan data yang akan diambil berupa postingan yang mengandung *Link*. Data postingan diambil sebanyak 277 kemudian dilakukan proses menggunakan software excel dan rapidminer. Data yang terkumpul nantinya akan dibagi dua menjadi data training dan data testing. Pada penelitian klasifikasi posting pada pengguna facebook ini menggunakan metode Naive Bayes Classifier untuk mengetahui postingan mana yang mengandung link phising.

Pada tahap awal, data akan diolah melalui proses preprocessing kemudian diberi label terhadap setiap postingan dalam dataset. Selanjutnya, data yang telah diberi label akan dikumpulkan menjadi data training dan data testing. Selanjutnya, metode klasifikasi Naive Bayes akan diterapkan pada data tersebut dan akan menghasilkan Confusion Matrix, setelah itu nilai akurasi dari metode Naive Bayes akan dihitung berdasarkan hasil yang diperoleh. Berikut ini tahapan riset yang akan dijalankan pada penelitian Klasifikasi Postingan Pengguna Facebook untuk mendukung kelancaran penelitian. Penjelasan yang sudah disampaikan akan menggabungkan antara teori dengan masalah yang dibahas pada penelitian ini.

Gambar 1 menjelaskan urutan langkah-langkah riset yang dilakukan dalam penelitian ini. Kemudian untuk melakukan pengolahan datanya akan menggunakan

algoritma naïve bayes agar dapat menghasilkan kesimpulan yang baik Metode Naive Bayes adalah algoritma klasifikasi yang memanfaatkan teori probabilitas untuk memprediksi kategori atau label dari sebuah data. Algoritma ini memerlukan data latih (training data) yang terdiri dari sejumlah objek yang telah diklasifikasikan ke dalam kategori atau label yang telah ditentukan sebelumnya. Naive Bayes mengasumsikan bahwa setiap fitur pada data latih adalah independen satu sama lain, sehingga memudahkan perhitungan probabilitas. Dengan menggunakan teori probabilitas Bayes, algoritma Naive Bayes akan menghitung probabilitas setiap kategori atau label untuk sebuah data baru yang belum diketahui kategorinya.



Gambar 1. Tahapan Riset

IV. HASIL DAN PEMBAHASAN

Data awal yang akan digunakan dan diambil dari laman facebook untuk penelitian sebanyak 300 lalu setelah melalui proses pre-processing atau menyeleksi data serta memberikan label baru yaitu *status* dengan manual menggunakan software Excel, data akhir yang akan digunakan hanya 277 Postingan Facebook baik dari laman pribadi maupun laman grup, berikut ini preview data yang sudah di seleksi:

Tabel 1. Dataset Penelitian

No	Id Postingan	Tanggal	Sumber	Link yang terdapat pada postingan	Status
271	967337600481351	6 Desember 2022	(1) MLBB GROUP Facebook	https://mobile-egends-free-skin2503.unpkg.my.id	Suspect
272	705413279981807	11 November 2021	(1) Ako Si DoGiE (GROUP) Facebook	https://lynk.id/eventcoda.1	Suspect
273	466273230635819	12 Desember 2022	(1) Mobile Legends North America Server Facebook	https://mobilelegends-v3.com/p/livop	Suspect
274	774219676119921	12 Desember 2022	https://www.facebook.com/photo/?fbid=661001746074031&set=a.562379039269636	https://mobilelegends-v3.com/p/livop	Non Suspect
275	321491802886914	10 Juni 2022	(1) MLBB FREE SKIN GIVEAWAY Facebook	https://mlbb-moonton.com/p/p3TePxEluP	Suspect
276	208298131214373	21 Desember 2022	(1) HRJ HaruJar International Fans Facebook	https://mobilelegendss-s22.xyz/p/EJ0anv	Suspect
277	1355298581553390	10 Desember 2022	(1) Mobile Legends Cambodia Facebook	https://mobile-egends-free-skin2503.unpkg.my.id	Suspect

Selanjutnya, data akan dibagi menjadi data training dan data testing dengan rasio 70:30 serta memberikan role

label pada tabel status dengan metode split data dan change role pada Rapidminer.

Langkah selanjutnya adalah menghitung tingkat akurasi, presisi, dan recall terhadap data testing menggunakan software *Rapidminer* untuk mengetahui seberapa akurat hasil pengujian data tersebut. Dalam penelitian ini, hanya dilakukan satu kali pengujian untuk menghitung tingkat akurasi, presisi, dan recall.

Akurasi adalah ukuran seberapa dekat hasil prediksi dengan kenyataan yang sebenarnya. Prediksi merujuk pada tingkat kesesuaian antara permintaan informasi dari pengguna dengan jawaban yang diberikan oleh sistem. Setelah data dibagi, nilai prediksi kemudian ditentukan

Tabel 2. Accuracy Data Testing

Accuracy: 90,36%			
	True Non Suspect	True Suspect	Class Precision
Pred. Non Suspect	38	7	84.44%
Pred. Suspect	1	37	97.37%
Class recall	97.44%	84.09%	

Prediksi dari data postingan yang terdeteksi mengandung phising pada data testing adalah 37 dan yang salah adalah 1, sedangkan data postingan yang tidak mengandung phising adalah 38 dan yang salah adalah 7. Hasil dari accuracy Naïve Bayes dengan perhitungan Rapid Miner adalah 90.36%.

Tabel 3. Precision data testing

Precision: 97,37% (positive class: Suspect)			
	True Non Suspect	True Suspect	Class Precision
Pred. Non Suspect	38	7	84.44%
Pred. Suspect	1	37	97.37%
Class recall	97.44%	84.09%	

Presisi adalah ukuran seberapa akurat informasi yang diberikan oleh sistem. Dalam pengujian ini, kami akan menghitung presisi terhadap data pengujian menggunakan Rapid Miner. Hasil pengujian presisi dari algoritma Naïve Bayes dengan data testing yaitu sebesar 97.37%.

Recall merupakan ukuran keberhasilan dalam mengambil data yang relevan. Dalam pengujian ini, recall akan dihitung terhadap data pengujian menggunakan Rapid Miner menghasilkan Recall sebesar 84,09%.

Tabel 4. Recall Data testing

Recall: 84,09% (positive class: Suspect)			
	True Non Suspect	True Suspect	Class Precision
Pred. Non Suspect	38	7	84.44%
Pred. Suspect	1	37	97.37%
Class recall	97.44%	84.09%	

V. KESIMPULAN

Berdasarkan penelitian yang telah dilakukan maka dapat disimpulkan penggunaan algoritma Naive bayes sangat tepat untuk memperhitungkan deteksi phising klasifikasi posting pengguna facebook. Dari dataset yang telah di dapat, dapat di tarik kesimpulan dengan hasil sebagai berikut, Hasil dari pengujian algoritma Naive bayes diperoleh nilai rata-rata akurasi sebesar 90,36%. Precision sebesar 97,37% dan Recall sebesar 84,09%. Dengan demikian hasil penerapan algoritma Naive bayes tersebut untuk mengetahui deteksi phising pada potingan pengguna facebook dikatakan sangat baik, dan penggunaan algoritma tersebut sudah tepat jika digunakan untuk mencegah pencurian data dari sebuah ancaman URL Phising pada postingan Facebook.

DAFTAR PUSTAKA

- [1] Moh Yunus, Dwi Widiastuti, Hasma Rasjid dan Yulia Chalri. (2019). Metode Klasifikasi Untuk Deteksi Uniform Resource Locator (URL) Berdasarkan Jenis Serangan Menggunakan Algoritma Naive Bayes, C4.5 dan K-Nearest Neighbor
- [2] Agus Fatkhurohman , Eli Pujastuti. (2019). Penerapan Algoritma Naïve Bayes Classifier Untuk Meningkatkan Keamanan Data Dari Website Phising
- [3] Jimmy H. Moedjahedy , Arief Setyanto , Komang Aryasa. (2020). ANALISIS PERBANDINGAN KORELASI SPEARMAN DAN MAXIMAL INFORMATION COEFFICIENT DALAM SELEKSI FITUR WEBSITE PHISHING MENGGUNAKAN ALGORITMA MACHINE LEARNING
- [4] Roni Anagora, Rudini, Rohmat Taufiq, Ahmad Dedi Jubaedi, Rio Wirawan, Arman Syah Putra. (2022). The Classification of Phishing Websites using Naive Bayes Classifier Algorithm
- [5] Ahmad Turmudi Zy, Agung Nugroho, Ahmad Rivaldi, Irfan Afriantoro. (2022). Analisis

- Sentimen Terhadap Pembobolan Data pada Twitter dengan Algoritma Naive Bayes
- [6] Pungkas Subarkah, Ali Nur Ikhsan. (2021). IDENTIFIKASI WEBSITE PHISHING MENGGUNAKAN ALGORITMA CLASSIFICATION AND REGRESSION TREES (CART)
- [7] Farida, Ali Mustopa. (2023). Perbandingan Logistic Regression dan Random Forest menggunakan Correlation-based Feature Selection untuk Deteksi Website Phishing
- [8] Anggit Ferdita Nugraha, Rifda Faticha, Alfa Aziza, Yoga Pristyanto. (2022). Penerapan metode Stacking dan Random Forest untuk Meningkatkan Kinerja Klasifikasi pada Proses Deteksi Web Phishing
- [9] Sunardi, Abdul Fadlil, Nur Makkie Perdana Kusuma. (2022). Implementasi Data Mining dengan Algoritma Naïve Bayes untuk Profiling Korban Penipuan Online di Indonesia
- [10] Agung Susilo Yuda Irawan, Nono Heryana, Hopi Siti Hopipah, Dyas Rahma. (2021). Identifikasi Website Phishing dengan Perbandingan Algoritma Klasifikasi
- [11] YUSUP MIFTAHUDDIN, MOHAMAD MUQIIT FATURRAHMAN. (2022). Penerapan Data Standardization dan Multilayer Perceptron pada Identifikasi Website Phishing
- [12] Michael Jonathan, Silvia Rostianingsih, Henry Novianus Palit. (2020). Pengaruh Feature Selection terhadap Kinerja C5.0, XGBoost, dan Random Forest dalam Mengklasifikasikan Website Phishing
- [13] APWG. (2019). Phishing Activity Report Quarter 4
- [14] Fayruz Rahma, Azmiardhy Zulkifli Farmadiansyah, Ahmad Fathan Hidayatullah. (2019). Deteksi Surel Spam dan Non Spam Bahasa Indonesia Menggunakan Metode Naïve Bayes
- [15] Nabila Bianca Putri, Arie Wahyu Wijayanto. (2019). Analisis Komparasi Algoritma Klasifikasi Data Mining Dalam Klasifikasi Website Phishing