KLASIFIKASI EMAIL PHISHING MENGGUNAKAN ALGORITMA K-NEAREST NEIGHBOR

Muhammad Adipa [a,1,*], Ahmad Turmudi Zy [b,2], M. Makmun Effendi [c,3]

[a] Universitas Pelita Bangsa, Jl. Inspeksi Kalimalang No.9, Cibatu, Cikarang Selatan, Bekasi 17530, Indonesia
[b] Universitas Pelita Bangsa, Jl. Inspeksi Kalimalang No.9, Cibatu, Cikarang Selatan, Bekasi 17530, Indonesia
[c] Universitas Pelita Bangsa, Jl. Inspeksi Kalimalang No.9, Cibatu, Cikarang Selatan, Bekasi 17530, Indonesia
[1] muhammadadipa06@gmail.com *; [2] turmudi@pelitabangsa.ac.id; [3] effendiyan@pelitabangsa.ac.id
* Penulis Korespondensi

Diterima : 08 Agustus 2023 | Direvisi : 10 Agustus 2023 | Diterbitkan : 13 Agustus 2023

ABSTRAK

Saat ini perkembangan teknologi informasi sangat pesat dan cepat, bahkan di Indonesia sendiri. Evolusi teknologi dunia internet terus berlanjut, dan inovasi baru terus bermunculan, membentuk masa depan yang lebih terhubung dan terintegrasi. Namun selain manfaat, muncul tantangan baru, seperti masalah terkait privasi, keamanan siber, dan pengolahan data. Pada satu sisi, perkembangan teknologi informasi yang demikian mengagumkan itu memang telah membawa manfaat yang luar biasa bagi kemajuan peradaban umat manusia. Di sisi lain, berkembangnya teknologi informasi menimbulkan pula sisi rawan yang gelap sampai tahap mencemaskan dengan kekhawatiran pada perkembangan tindak pidana di bidang teknologi informasi yang berhubungan dengan kejahatan mayantara atau "Cybercrime". Salah satu kejahatan (cybercrime) yang terjadi di Indonesia yaitu Email Phishing. Badan Siber dan Sandi Negara (BSSN) melaporkan, ada 164.131 kasus email phishing di Indonesia pada 2022. Tingginya angka kasus email phishing terus meningkat, oleh karena itu akan dilakukan pengujian untuk mengklasifikasi email phishing menggunakan algoritma K-Nearest Neighbor. Metode K-Nearest Neighbor digunakan karena bekerja dengan baik untuk pengklasifikasian data dalam bentuk objek text. Didapatkan hasil akurasi dengan nilai sebesar 84%, precision sebesar 73%, dan recall sebesar 96%. Hasil ini membuktikan bahwa algoritma K-Nearest Neighbor memberikan hasil yang cukup baik dalam mengklasifikasi email phishing.

KATA KUNCI

Cybercrime
Email Phishing
Klasifikasi
Data Mining
K-Nearest Neighbor

ABSTRACT

Currently the development of information technology is very fast and fast, even in Indonesia itself. The technological evolution of the internet world continues, and new innovations continue to emerge, shaping a more connected and integrated future. But apart from the benefits, new challenges arise, such as issues related to privacy, cyber security, and data processing. On the one hand, the development of such amazing information technology has indeed brought extraordinary benefits to the advancement of human civilization. On the other hand, the development of information technology has also created a dark vulnerable side to the point of worrying about the development of criminal acts in the field of information technology related to mayantara crime or "Cybercrime". One of the crimes (cybercrime) that occurred in Indonesia, namely Email Phishing. The National Cyber and Crypto Agency (BSSN) reported that there were 164,131 phishing email cases in Indonesia in 2022. The high number of phishing email cases continues to increase, therefore a test will be carried out to classify phishing emails using the K-Nearest Neighbor algorithm. K-Nearest Neighbor method is used because it works well for classification data in the form of text objects. Accuracy results were obtained with a value of 84%, precision of 73%, and recall of 96%. These results prove that the K-Nearest Neighbor algorithm gives good results in classifying phishing emails.

KEYWORD

Cybercrime
Phishing Email
Classification
Data Mining
K-Nearest Neighbor

1. Pendahuluan

Saat ini perkembangan teknologi informasi sangat pesat dan cepat, bahkan di Indonesia sendiri. Pada dasarnya, teknologi ada untuk mempermudah manusia[1]. Teknologi informasi telah mengubah pola hidup masyarakat di seluruh dunia dan menyebabkan perubahan yang cepat dan signifikan dalam kerangka sosial budaya, ekonomi, dan hukum[2]. Evolusi teknologi dunia internet terus berlanjut, dan inovasi baru terus bermunculan, membentuk masa depan yang lebih terhubung dan terintegrasi. Namun selain manfaat, muncul tantangan baru, seperti masalah terkait privasi, keamanan siber, dan pengolahan data. Kemajuan teknologi informasi, khususnya di bidang komputer dan internet, terbukti berdampak positif dalam memajukan kehidupan manusia tetapi terdapat sisi gelap juga yang dapat menghancurkan kehidupan manusia dan budaya itu sendiri[3].

Saat ini kebutuhan manusia akan teknologi informasi tersedia dengan bebas dan tanpa batas. Pada satu sisi, perkembangan teknologi informasi yang demikian mengagumkan itu memang telah membawa manfaat yang luar biasa bagi kemajuan peradaban umat manusia. Di sisi lain, berkembangnya teknologi informasi menimbulkan pula sisi rawan yang gelap sampai tahap mencemaskan dengan kekhawatiran pada perkembangan tindak pidana di bidang teknologi informasi yang berhubungan dengan kejahatan mayantara atau "Cybercrime"[4]. Banyak kejahatan (cybercrime) yang memanfaatkan kemajuan dari teknologi informasi, seperti kejahatan Carding (Credit Card Fraud), ATM/EDC Skimming, Hacking, Cracking, Phising (Internet Banking Fraud), Malware (Virus/Worm/Trojan/Bots), Cybersquatting, Pornografi, Perjudian Online, Transnasional Crime (Perdagangan Narkoba, Mafia, Terorisme, Money Laundering, Human Trafficking, Underground Economy)[5]. Semua kejahatan itu bisa dengan mudah dan efektif dilakukan dengan memanfaatkan kemajuan teknologi informasi itu sendiri. Selain memanfaatkan perkembangan teknologi dan informasi, tidak menutup kemungkinan untuk melakukan kejahatan (cybercrime) dengan mudah dan efektif juga di bidang pengelolaan data dan informasi, khususnya dalam hal pengelolaan data pribadi yang membutuhkan perlindungan data. Hal ini dikarenakan perkembangan teknologi informasi dan komunikasi telah mempersempit batasan privasi, sehingga berbagai data pribadi dapat lebih mudah didistribusikan[6].

Salah satu kejahatan (cybercrime) yang terjadi di Indonesia yaitu Email Phishing. Phishing adalah aktivitas kriminal yang menggunakan teknik rekayasa sosial[7]. Upaya peretasan dilakukan dengan berbagai macam cara demi mendapatkan informasi pribadi seseorang. Salah satunya adalah dengan menyamar sebagai orang atau organisasi berwenang melalui surat elektronik atau dikenal dengan email phishing. Badan Siber dan Sandi Negara (BSSN) melaporkan, ada 164.131 kasus email phishing di Indonesia pada 2022. Jumlah tersebut paling banyak berasal dari email pribadi, yakni 59.210 kasus. Sebanyak 52.744 kasus email phishing berasal dari email group. Kemudian, ada 52.177 kasus email phishing yang berasal dari email lainnya. Adapun, 93.897 kasus email phishing terjadi saat jam kerja atau pukul 09.00-17.00. Sementara, 70.234 kasus lainnya dilakukan di luar jam kerja pada pukul 17.00 hingga 09.00. Email phishing yang terjadi pada 2022 juga kerap melampirkan sebuah file. Format file yang paling mendominasi memiliki ekstensi .pdf, yakni lebih dari 100.000 kasus[8].

Perkembangan teknologi sistem informasi telah memecahkan banyak masalah di berbagai bidang, salah satunya adalah penerapan data mining. Data mining adalah proses penggalian informasi atau pola yang berguna dan bermakna dari kumpulan data yang besar dan kompleks. K-Nearest Neighbor (KNN) adalah salah satu algoritma data mining yang digunakan untuk melakukan klasifikasi[9]. Algoritma ini bekerja dengan cara mencari nilai K data terdekat dari data yang akan diklasifikasikan, kemudian data akan diklasifikasikan berdasarkan mayoritas kelas dari data-data terdekat tersebut[10].

2. Tinjauan Pustaka

Adapun beberapa referensi yang berkaitan dengan klasifikasi menggunakan algoritma K-Nearest Neighbor sebagai berikut:

(Naisah Marito Putry, Betha Nurina Sari, 2022) Komparasi Algoritma KNN Dan Naïve Bayes Untuk Klasifikasi Diagnosis Penyakit Diabetes Melitus. Penelitian ini melakukan komparansi (perbandingan) antara dua algoritma yaitu KNN dan Naive Bayes dalam mengklasifikasikan diagnosis penyakit diabetes melitus. Selain itu, penelitian ini menggunakan lima pembagian data yang dilakukan dapat dilihat bahwa nilai akurasi dari Naive Bayes lebih tinggi dibandingkan KNN. Dimana nilai akurasi paling tinggi yang didapatkan dari algoritma Naive Bayes yaitu sebesar 80%. Sedangkan algoritma KNN nilai akurasi tertinggi yaitu sebesar 75%. Selain itu, diketahui bahwa nilai recall paling tinggi dihasilkan oleh algoritma

KNN yaitu sebesar 0.92. Dan untuk nilai presisi lebih tinggi dihasilkan oleh algoritma Naive Bayes yaitu 0.86[11].

(Danny Sebastian, 2019) Implementasi Algoritma K-Nearest Neighbor untuk Melakukan Klasifikasi Produk dari beberapa E-marketplace. Berdasarkan hasil pengujian, metode K-Nearest Neighbor dapat melakukan klasifikasi produk dari e-marketplace, khususnya tokopedia dan bukalapak. Akurasi yang dihasilkan dari pengujian 1, pemilihan nilai k=1, 5, atau 10 adalah 78%, 97,33% dan 92%. Berdasarkan pengujian 1 disimpulkan nilai k yang optimal untuk kasus ini adalah 5. Pada pengujian 2, multi-brand, akurasi yang dihasilkan adalah 96.67%. Nilai akurasi ini dapat dikatakan baik karena melebihi 90%. Akurasi dari algoritma K-Nearest Neighbor sangat dipengaruhi oleh data latih. Semakin lengkap data latih, maka akurasi akan semakin baik[12].

(Green Arther Sandag, Jonathan Leopold, Vinky Fransiscus Ong, 2018) Klasifikasi Malicious Websites Menggunakan Algoritma K-NN Berdasarkan Application Layers dan Network Characteristics. Dari 4 algoritma yang telah di evaluasi yaitu K-Nearest Neighbor, Decision Tree, Logistic Regression, dan Random Forest dapat disimpulkan bahwa algoritma K-NN memiliki performa yang paling baik di antara algoritma lainnya dengan hasil 95.51% accuracy, 89.42% recall, 89.42% precision, dan 0.212 RMSE untuk hasil independent sedangkan untuk hasil 10fold cross validation memiliki hasil 93.61% accuracy, 85.05% recall, 85.25% precision, dan 0.251 RMSE dalam mendeteksi maliciousdan benign website[13].

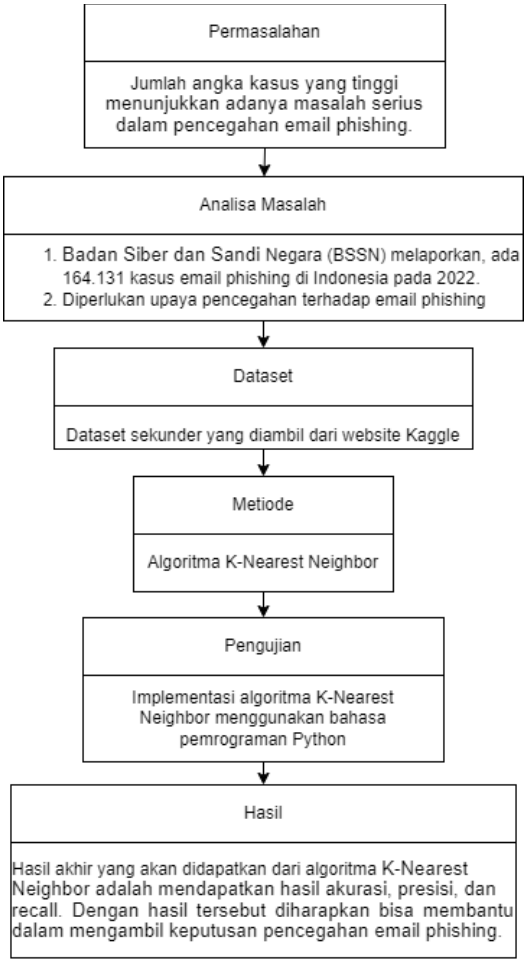
(Difari Afreyna Fauziah, Achmad Maududie, Ifrina Nuritha, 2018) Klasifikasi Berita Politik Menggunakan Algoritma K-Nearst Neighbor. Dari hasil pengujian dan analisis nilai precision, recall dan f-measure menggunakan tabel confusion matrix pada algoritma KNN untuk sistem klasifikasi konten berita politik menunjukkan nilai akurasi yang tinggi. Pada penelitian ini dilakukan tiga kali pengujian dengan menggunakan variasi jumlah dataset dan menggunakan empat nilai k. Dari hasil pengujian tersebut didapatkan nilai k terbaik yang didapatkan oleh sistem ketika sistem menggunakan nilai k=9 yang memberikan nilai precision sebesar 100%, recall sebesar 100% dan f-measure sebesar 100% pada pengujian kedua yang menggunakan data training sebanyak 210 berita dan pengujian ketiga yang menggunakan data training sebanyak 270 berita. Dengan demikian, algoritma KNN dapat bekerja dengan baik ketikamenggunakan nilai k=9. Sehingga dapat disimpulkan algoritma KNN cocok untuk diterapkan pada proses klasifikasi dengan dokumen yang memiliki similarity yang tinggi[14].

(Anis Nikmatul Kasanah, Muladi, Utomo Pujianto, 2019) Penerapan Teknik SMOTE untuk Mengatasi Imbalance Class dalam Klasifikasi Objektivitas Berita Online Menggunakan Algoritma KNN. Hasil penelitian dengan menerapkan nilai k yang bervariasi yaitu 1, 3, 5, 7 dan 9 diperoleh bahwa penerapan SMOTE dalam menangani imbalance class pada klasifikasi objektivitas berita menghasilkan performa yang kurang efektif pada nilai k 5, 7 dan 9. Hal ini ditunjukkan saat performa algoritma KNN mengalami rata-rata penurunan nilai akurasi sebesar -6,67. Sedangkan untuk nilai k=1 dan k=3 performa[15].

3. Metodologi Penelitian

3.1. Kerangka Pemikiran

Kerangka pemikiran ini dibuat untuk mempermudah memahami arah penelitian dari proses-proses hingga hasil yang akan diperoleh, kerangka pemikiran ini dapat dilihat pada gambar berikut:



Gambar 1. Kerangka Pemikiran

Menjelaskan permasalahan yang ada sebagai latar belakang penelitian dan tujuan yang akan dicapai, menganalisa masalah dan membuat ruang lingkup masalah agar lebih sederhana, data yang digunakan merupakan data sekunder yang diambil melalui pihak yang telah mengumpulkan data tersebut sebelumnya, setelah itu melakukan pengujian terdahap dataset menggunakan algoritma K-Nearest Neighbor.

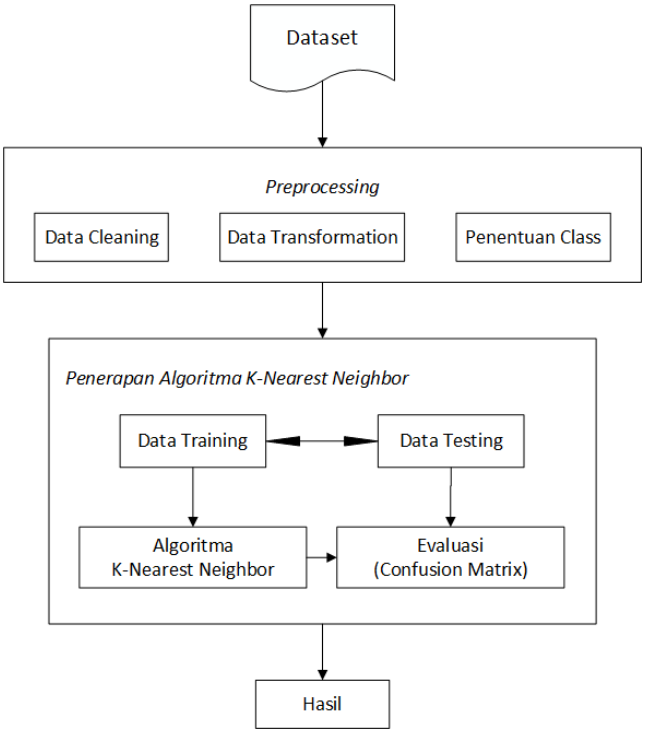
Objek Penelitian adalah permasalahan yang diteliti dan dibahas dalam penelitian. Pada penelitian ini diangkat permasalahan tentang kasus email phishing. Tingginya angka kasus email phishing terus meningkat. Oleh karena itu dilakukan pengujian terhadap dataset email phishing dengan menggunakan algoritma K-Nearest Neighbor untuk mengklasifikasi email phishing.

Subjek penelitian yang akan digunakan yaitu menggunakan dataset email phishing yang didapat dari website Kaggle untuk mengklasifikasi terkena serangan email phishing atau tidak.

Data yang digunakan merupakan dataset email phishing yang diperoleh dari website Kaggle yang berisi 18.650 data. Dataset ini berisi body email yang digunakan dalam mendeteksi email phishing dan satu atribut label yang menunjukan apakah email tersebut termasuk email phishing atau email aman. Penulis akan melakukan beberapa pengujian data sehingga akan terbentuk beberapa data, data training dan data testing.

3.2. Metode Penelitian

Penelitian ini akan melalui beberapa proses, urutan proses penelitian ini akan diuraikan pada metode penelitian ini:



Gambar 2. Metode Penelitian

Dataset yang digunakan merupakan data yang didapatkan dari website Kaggle dengan jumlah record data sebanyak 18.650 data. Setelah itu dilakukan tahap preprocessing sesuai dengan proses metode Knowledge Discovery in Database (KDD). Data dan atribut harus melalui beberapa tahap pengolahan awal data seperti pembersihan data (cleaning), filtering, transformasi data, dan lain-lain sehingga mendapatkan data yang berkualitas. Setelah itu dataset akan di bagi menjadi 2 yaitu data training dan data testing. Data training akan diterapkan pada algoritma yang akan digunakan. Pada penelitian ini algoritma yang digunakan adalah K-Nearest Neighbor. Parameter yang digunakan untuk penerapan algoritma ini menggunakan parameter default dari library scikit-learn. Evaluasi dilakukan dengan cara menganalisa hasil dari algoritma yang digunakan untuk memastikan bahwa hasil perhitungan dan pengujian sesuai dengan tujuan penelitian. Validasi dilakukan untuk mengukur hasil klasifikasi guna mengetahui tingkat accuracy, precision, dan recall.

4. Hasil dan Pembahasan

4.1 Hasil

1) Upload Dataset

Perlu mengunggah dataset yang akan digunakan ke Google Drive. Pastikan akun yang digunakan untuk meng-upload dataset ke Google Drive dan akun yang digunakan di Google Colaboratory.

```
#Menyambungkan google colab dengan google drive

from google.colab import drive
drive.mount('/content/drive')
```

```
Mounted at /content/drive
```

Gambar 3. Menghubungkan ke Google Drive

2) Memanggil Dataset

Langkah selanjutnya adalah memanggil dataset yang akan digunakan.

```
dataset = '/content/drive/MyDrive/Dataset/Phishing_Email.csv'
df = pd.read_csv(dataset)
df
```

	Unnamed: 0	Email Text	Email Type
0	0	re : 6 . 1100 , disc : uniformitarianism , re ...	Safe Email
1	1	the other side of * galicismos * * galicismo *...	Safe Email
2	2	re : equistar deal tickets are you still avail...	Safe Email
3	3	\nHello I am your hot lil horny toy.\n I am...	Phishing Email
4	4	software at incredibly low prices (86 % lower...	Phishing Email
...	...	...	...
18645	18646	date a lonely housewife always wanted to date ...	Phishing Email
18646	18647	request submitted : access request for anita	Safe Email
18647	18648	re : important - prc mtg hi dorn & john , as y...	Safe Email
18648	18649	press clippings - letter on californian utilit...	Safe Email
18649	18650	empty	Phishing Email

18650 rows × 3 columns

Gambar 4. Memanggil Dataset

3) Melihat Informasi Dataset

Langkah selanjutnya adalah mendapatkan informasi tentang dataset, termasuk jumlah data, jenis setiap atribut, dan lain-lain.

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 18650 entries, 0 to 18649
Data columns (total 3 columns):
 #   Column      Non-Null Count  Dtype
---  ------      --------------  -----
 0   Unnamed: 0  18650 non-null  int64
 1   Email Text  18634 non-null  object
 2   Email Type  18650 non-null  object
dtypes: int64(1), object(2)
memory usage: 437.2+ KB
```

Gambar 5. Melihat Informasi Dataset

4) Data Preprocessing

Pada tahap ini, data melewati proses pembersihan (data cleaning), transformasi data dan seleksi data.

```
corpus = []
for text in df['Email Text']:
    email = re.sub('[^a-zA-Z]', ' ', str(text))
    email = email.lower()
    email = email.split()
    stemmer = PorterStemmer() #stemming
    email = [stemmer.stem(word) for word in email if word not in set(stop_words)]
    email = ' '.join(email)
    corpus.append(email)
```

Gambar 6. Cleaning Data

Pada proses ini tiap kalimat dalam email akan di filter menjadi hanya berisi karakter alphabet. Kemudian semua huruf akan diubah menjadi huruf kecil dan di bagi perkata. Selanjutnya setiap kata akan di ubah agar menjadi seragam. Setelah semua proses tersebut dilakukan kemudian kata-kata tersebut akan di gabungkan kembali menjadi email yang utuh.

5) Penentuan Class dan Attributte

Langkah selanjutnya yaitu menentukan nilai atribut dan class variabel x untuk atribut dan variabel y untuk class. Pada variabel x akan di buat agar setiap email hanya akan memuat 10000 kata. Hal ini dimaksudkan agar analisis yang dilakukan lebih optimal dan tidak terjadi bias dalam proses pemodelan.

```
cv = CountVectorizer(max_features = 10000)
X = cv.fit_transform(corpus).toarray()
Y = df.iloc[:, -1].values
```

Gambar 7. Penentuan Class dan Attributte

6) Split Data Validation

Langkah selanjutnya dari penelitian ini adalah membagi data menjadi dua bagian yaitu data latih dan data uji. Data dibagi dengan dua bobot 70:30, dimana 70 digunakan sebagai data training dan 30 sebagai data testing.

```
from sklearn.model_selection import train_test_split
X_train,X_test,Y_train,Y_test = train_test_split(X, Y, test_size = 0.3, random_state = 123)
```

Gambar 8. Split Data Validation

7) Penerapan Algoritma K-Nearest Neighbor

Pada langkah ini, algoritma KNN diterapkan pada data pelatihan yang dihasilkan pada langkah sebelumnya. Langkah selanjutnya dari penelitian ini adalah menguji data uji dengan data training yang telah diterapkan pada algoritma.

```
# KNeighbors
from sklearn.neighbors import KNeighborsClassifier
knn = KNeighborsClassifier(n_neighbors = 3)
knn.fit(X_train, Y_train)

#Pengujian
Y_pred = knn.predict(X_test)
```

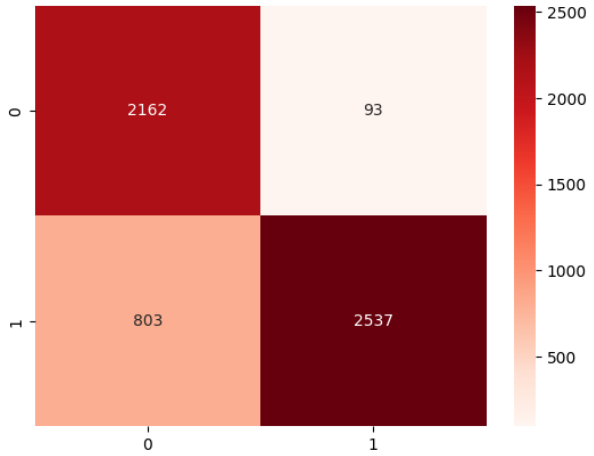
Gambar 9. Penerapan Algoritma K-Nearest Neighbor

8) Confusion Matrix

Setelah pengujian, langkah selanjutnya adalah membuat confusion matrix. Pada langkah ini, confusion matrix dibuat menggunakan library pendukung dari ScikitLearn (confusion_matrix) dan divisualisasikan dengan menggunakan library Seaborn. Pada confusion matrix dapat dilihat bahwa menggunakan algoritma K-Nearest Neighbor terdapat 2162 email aman yang terklasifikasi dengan benar dan 93 email aman yang diklasifikasikan sebagai email phishing. selain itu juga terdapat 2537 email phishing yang terdeteksi dengan baik, dan 803 email phishing yang diklasifikasikan sebagai email aman.

```
from sklearn.metrics import confusion_matrix
conf_matrix = confusion_matrix(Y_test, Y_pred)
print("Confusion Matrix:")
sns.heatmap(conf_matrix, annot=True, cmap='Reds', fmt='g')
```

```
Confusion Matrix:
<Axes: >
```



Gambar 10. Confusion Matrix

9) Evaluasi

Pada tahap ini, akan mendapatkan classification report dari kemampuan algoritma yang digunakan untuk mengklasifikasi dataset yang digunakan. Hasil yang diperoleh berupa skor accuracy, precision, dan recall. Dipatakan hasil dimana algoritma K-Nearest Neighbor memiliki akurasi sebesar 84% dalam mendeteksi email phishing, selain itu juga didapatkan nilai rata-rata presisi sebesar 85% dan recall sebesar 86%.

```
# Hasil Valuasi
acc_knn = round(knn.score(X_train, Y_train) * 100, 2)
print('KNN:'+str(acc_knn))
print(classification_report(Y_test, Y_pred))
```

```
KNN:89.53
                precision    recall  f1-score   support

Phishing Email       0.73      0.96      0.83      2255
    Safe Email       0.96      0.76      0.85      3340

      accuracy                           0.84      5595
     macro avg       0.85      0.86      0.84      5595
  weighted avg       0.87      0.84      0.84      5595
```

Gambar 11. Hasil Evaluasi

4.2 Pembahasan

Berdasarkan perhitungan dari confusion matrix data testing yang berjumlah 5595 data. 2162 data yang diklasifikasi sebagai email aman dan 803 data email aman yang diklasifikasikan sebagai email phising. Di confusion matrix yang sama terdapat 94 data email phising yang diklasifikasikan sebagi email aman dan 2537 data email phising yang diklasifikasi sebagai email phising. Berdasarkan hal tersebut didapatkanlah hasil nilai accuracy sebesar 84%, precision sebesar 85%, dan recall sebesar 86%.

5 Penutup

5.1 Kesimpulan

Berdasarkan hasil penelitian yang sudah dilakukan sebelumnya maka dapat diambil kesimpulan bahwa penerapan algoritma K-Nearest Neighbor dalam mengklasifikasi suatu email termasuk phising atau aman menggunakan bahasa pemrograman python dapat diterapkan dengan baik dan memberikan hasil yang sangat baik sehingga dapat membantu dalam mengambil keputusan untuk menghindari suatu email phising.

Dari proses pengujian yang dilakukan dengan bobot data 70:30 mendapatkan hasil akurasi dengan nilai sebesar 84%, precision sebesar 85%, dan recall sebesar 86%. Hasil model machine learning pada penelitian ini diharapkan bisa membantu mengambil keputusan dengan tepat untuk menghindari email phishing.

5.2 Saran

Berdasarkan pada hasil penelitian yang dilakukan, didapat saran untuk penelitian selanjutnya perlu dilakukan pengujian dengan menggunakan tools lainnya untuk menganalisa hasil yang didapatkan menjadi hasil yang lebih baik dari penelitian ini seperti penggunaan Bahasa pemrogaman R atau Aplikasi Rapidminer. Perlu dilakukan perbandingan dengan algoritma lain untuk menguji sejauh mana algoritma K-Nearest Neighbor dapat digunakan untuk mengklasifikasikan email phishing dan sebaiknya menggunakan dataset yang memiliki jumlah data yang lebih banyak untuk menghasilkan hasil yang lebih akurat.

Daftar Pustaka

[1] L. Yana Siregar, M. Irwan Padli Nasution Prodi Manajemen, and U. Negeri Islam Sumatera Utara, "HIRARKI Jurnal Ilmiah Manajemen dan Bisnis DEVELOPMENT OF INFORMATION TECHNOLOGY ON INCREASING BUSINESS ONLINE," vol. 2, no. 1, pp. 71–75, 2020, doi: 10.30606/hjimb.

[2] M. Melani, H. S. Disemadi, and N. S. P. Jaya, "Kebijakan Hukum Pidana Dibidang Transaksi Elektronik Sebagai Tindak Pidana Non-Konvensional," *Pandecta Research Law Journal*, vol. 15, no. 1, pp. 111–120, Jun. 2020, doi: 10.15294/pandecta.v15i1.19469.

[3] A. Turmudi Zy, A. Nugroho, A. Rivaldi, and I. Afriantoro, "Analisis Sentimen Terhadap Pembobolan Data pada Twitter dengan Algoritma Naive Bayes," *Jurnal Teknologi Informatika dan Komputer*, vol. 8, no. 2, pp. 202–213, Sep. 2022, doi: 10.37012/jtik.v8i2.1240.

[4] U. Sawerigading Makassar, "PENGARUH PERKEMBANGAN TEKNOLOGI TERHADAP TERJADINYA KEJAHATAN MAYANTARA (CYBERCRIME) Raodia," 2019.

[5] P. Aptika dan IKP, B. Litbang SDM, and K. Jl Medan Merdeka Barat No, "TANTANGAN DALAM IMPLEMENTASI STRATEGI KEAMANAN SIBER NASIONAL INDONESIA DITINJAU DARI PENILAIAN GLOBAL CYBERSECURITY INDEX Maulia Jayantina Islami TANTANGAN DALAM IMPLEMENTASI STRATEGI KEAMANAN SIBER NASIONAL INDONESIA DITINJAU DARI PENILAIAN GLOBAL CYBERSECURITY INDEX Challenges in The Implementation of National Cybersecurity Strategy of Indonesia from The Global Cybersecurity Index Point of View Maulia Jayantina Islami."

[6] M. H. Rumlus and H. Hartadi, "Kebijakan Penanggulangan Pencurian Data Pribadi dalam Media Elektronik," *Jurnal HAM*, vol. 11, no. 2, p. 285, Aug. 2020, doi: 10.30641/ham.2020.11.285-299.

[7] E. Ginting, M. Pardomuan Sinaga, M. Rizal Nurdin, M. Dimas Putra, P. Studi Sistem Informasi, and K. Kunci, "ANALISIS ANCAMAN PHISING TERHADAP LAYANAN ONLINE PERBANKAN (STUDI KASUS PADA BANK BRI) PHISING THREAT ANALYSIS OF ONLINE BANKING SERVICES (CASE STUDY ON BANK BRI)," 2023. [Online]. Available: https://ojs.ekasakti.org/i

[8] Sarnita Sadya, "Ada 164.131 Kasus Email Phising di Indonesia pada 2022," *dataindonesia.id*, Mar. 28, 2023.

[9] A. M. Argina, "Penerapan Metode Klasifikasi K-Nearest Neigbor pada Dataset Penderita Penyakit Diabetes," *Indonesian Journal of Data and Science*, vol. 1, no. 2, pp. 29–33, Jul. 2020, doi: 10.33096/ijodas.v1i2.11.

[10] D. Cahyanti, A. Rahmayani, and S. A. Husniar, "Analisis performa metode Knn pada Dataset pasien pengidap Kanker Payudara," *Indonesian Journal of Data and Science*, vol. 1, no. 2, pp. 39–43, Jul. 2020, doi: 10.33096/ijodas.v1i2.13.

[11] N. M. Putry, "KOMPARASI ALGORITMA KNN DAN NAÏVE BAYES UNTUK KLASIFIKASI DIAGNOSIS PENYAKIT DIABETES MELLITUS," *EVOLUSI : Jurnal Sains dan Manajemen*, vol. 10, no. 1, Apr. 2022, doi: 10.31294/evolusi.v10i1.12514.

[12] D. Sebastian, "Implementasi Algoritma K-Nearest Neighbor untuk Melakukan Klasifikasi Produk dari beberapa E-marketplace," *Jurnal Teknik Informatika dan Sistem Informasi*, vol. 5, no. 1, May 2019, doi: 10.28932/jutisi.v5i1.1581.

[13] G. A. Sandag, J. Leopold, and V. F. Ong, "Klasifikasi Malicious Websites Menggunakan Algoritma K-NN Berdasarkan Application Layers dan Network Characteristics," *CogITo Smart Journal*, vol. 4, no. 1, pp. 37–45, Jun. 2018, doi: 10.31154/cogito.v4i1.100.37-45.

[14] D. A. Fauziah, A. Maududie, and I. Nuritha, "Klasifikasi Berita Politik Menggunakan Algoritma K-nearst Neighbor," *BERKALA SAINSTEK*, vol. 6, no. 2, p. 106, Dec. 2018, doi: 10.19184/bst.v6i2.9256.

[15] A. N. Kasanah, M. Muladi, and U. Pujianto, "Penerapan Teknik SMOTE untuk Mengatasi Imbalance Class dalam Klasifikasi Objektivitas Berita Online Menggunakan Algoritma KNN," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 3, no. 2, pp. 196–201, Aug. 2019, doi: 10.29207/resti.v3i2.945.

[16] Sujjada, A., Ramdani, A. R., Kibtiyah, K., Utami, M. P., & Nullah, M. R. (2023). Prediksi Nilai Ujian Sekolah Siswa SMK Plus Padjadjaran Berbasis Web Menggunakan Jaringan Syaraf Tiruan Backpropagation. *Jurnal Informasi dan Teknologi*, 151-158.