INOTEK, Vol.7 Agustus 2023
ISSN: 2580-3336 (Print) / 2549-7952 (Online)
Url: https://proceeding.unpkediri.ac.id/index.php/inotek/

Implementasi Algoritma Regresi Logistik untuk Binary Classification dalam Spam SMS dan WhatsApp

Diterima:
10 Mei 2023
Revisi:
10 Juli 2023
Terbit:
1 Agustus 2023

[1*]Ana Nur Rani Hasanah, [2]Rr. Artiana Krestianti, [3]Sutresna Wati
[1-3]Universitas Gunadarma

Abstrak— WhatsApp merupakan layanan seperti pengiriman pesan; panggilan suara dan video; serta pengiriman dokumen dan media; telah menjadi pilihan yang sangat digemari oleh masyarakat di seluruh dunia. Meskipun demikian; aplikasi ini juga seringkali menjadi tempat penyebaran spam; yang dapat berupa penipuan; promosi; atau bentuk negatif lainnya. Meskipun terdapat berbagai upaya untuk mengklasifikasikan spam SMS berbahasa Inggris dengan menggunakan algoritma regresi logistik; namun masih sangat jarang ditemukan SMS dan WA dalam bahasa Indonesia. Oleh karena itu; penelitian ini bertujuan untuk melakukan klasifikasi biner pada data ham dan spam menggunakan metode algoritma regresi logistik pada 10793 data; dengan 10038 ham dan 775 spam. Penelitian ini juga menghasilkan lima matriks evaluasi yang dapat membantu memvisualisasikan hasil model yang telah dihasilkan; serta beberapa dekomposisi data untuk menemukan hasil terbaik selama proses pembuatan model; berdasarkan penelitian sebelumnya.; menghasilkan Accuracy = 7525 (0;9703593923675435); Precision = 5050 (0;9753694581280288); Recall = 8020 (0;6385542168674698); F1-score = 8020 (0;7653429602888085) dan ROC dengan nilai AUC = 7525 (0;98716810090769 8).

Kata Kunci— Algoritma Logistik Regresi; Klasifikasi Biner; Pembelajaran Mesin; Pesan Spam dan Ham; Python

Abstract— *WhatsApp is a popular communication application that provides services such as sending messages; voice and video calls; as well as sending documents and media; has become a very popular choice for people around the world. However; these applications are also often a place for spreading spam; which can be in the form of fraud; promotions; or other negative forms. Although there have been various attempts to classify SMS spam in English using a logistic regression algorithm; it is still very rare to find SMS and WA in Indonesian. Therefore; this study aims to carry out a binary classification of ham and spam data using the logistic regression algorithm method on 10793 data; with 10038 ham and 775 spam. This study also produces five evaluation matrices that can help visualize the results of the model that has been produced; as well as some data decomposition to find the best results during the model building process; based on previous research; resulting in Accuracy = 7525 (0.9703593923675435); Precision = 5050 (0 .9753694581280288); Recall = 8020 (0.6385542168674698); F1-score = 8020 (0.7653429602888085) and ROC with AUC value = 7525 (0.98716810090769 8).*

Keywords— *Binary Classification; Machine Learning; Regression Logistics Algorithm; Spam and Ham Chat; Python*

Penulis Korespondensi:

Ana Nur Rani Hasanah;
Sistem Informasi;
Universitas Gunadarma;
Email: ananur.rani@gmail.com

ISSN: 2580-3336 (Print) / 2549-7952 (Online)
Url: https://proceeding.unpkediri.ac.id/index.php/inotek/

I. PENDAHULUAN

SMS (Short Message Service) merupakan salah satu media komunikasi nirkabel yang masih digunakan oleh masyarakat dengan biaya yang lebih murah sehingga cepat dan mudah bagi pengirim dan penerima SMS. Pesatnya perkembangan teknologi SMS dipengaruhi oleh banyaknya penyedia jasa telekomunikasi yang menawarkan layanannya dengan harga yang cukup terjangkau kepada seluruh masyarakat. Semakin banyak pengguna SMS di masyarakat yang disalahgunakan oleh pihak yang tidak bertanggung jawab untuk melakukan kejahatan dengan menyebarkan spam SMS yang tidak diinginkan dan tidak diinginkan seperti promosi; penipuan; pesan pornografi; dll [1]. Meskipun SMS jarang digunakan untuk komunikasi di ponsel cerdas; SMS saat ini sangat penting untuk pendaftaran; pembuatan akun; dan konfirmasi untuk berbagai aplikasi; dll. Oleh karena itu; semakin banyak orang yang memanfaatkan situasi ini. WhatsApp (WA) adalah aplikasi panggilan video dan per pesanan gratis. Aplikasi ini telah digunakan oleh lebih dari 2 miliar orang di lebih dari 180 negara dengan peringkat teratas sebagai aplikasi obrolan paling populer dan 500 juta unduhan di PlayStore. WhatsApp sederhana; andal; dan pribadi sehingga Anda dapat dengan mudah tetap berhubungan dengan teman dan keluarga. WhatsApp dapat digunakan di smartphone dan komputer desktop; bahkan dengan koneksi internet yang lambat; tanpa biaya langganan bulanan atau tahunan seperti aplikasi lainnya. Per pesanan global pribadi; koneksi instan aman dan sederhana; panggilan suara dan video berkualitas tinggi; obrolan grup untuk tetap terhubung secara waktu nyata dan berbagi momen sehari-hari melalui Status [2]. Sama halnya dengan aplikasi WhatsApp; karena WhatsApp sekarang menjadi aplikasi chatting yang sangat populer untuk komunitas besar di seluruh dunia; banyak akun WhatsApp yang digunakan untuk spam seperti penipuan; promosi; dll. Banyak orang telah mencoba mengklasifikasikan spam SMS berbahasa Inggris menggunakan algoritma regresi logistik; namun masih sangat jarang ditemukan SMS dan WA berbahasa Indonesia.

Berdasarkan latar belakang di atas; penelitian ini bertujuan untuk mengevaluasi 10.793 pesan SMS dan WA Indonesia yang diperoleh dari data private message; baik spam maupun ham. Di mana pesan spam adalah pesan dengan konten iklan; lelucon; dll. sedangkan pesan ham adalah pesan yang biasanya kita terima dari seseorang dengan kebutuhan pribadi.

Machine learning merupakan salah satu cabang kecerdasan buatan (Artificial Intelligence) yang memungkinkan otomatis untuk belajar sendiri tanpa harus diprogram ulang oleh manusia. Proses belajar pada machine learning data pelatihan dalam jumlah tertentu dan hasil pembelajaran ini akan diuji terhadap tipe data yang sama atau tipe data yang berbeda. Jenis pembelajaran dalam machine learning antara lain Supervised Learning; di mana pengetahuan membangun model menggunakan input berupa data berlabel dan tes untuk membuat prediksi untuk data yang tidak

berlabel dan jenis Unsupervised Learning; di mana pengetahuan tidak diberi label dan data dikelompokkan berdasarkan fitur yang ditemukan [3].

Metode TF-IDF menggunakan 2 parameter bobot; yaitu bobot lokal dengan $tf_{i;j}$ dan bobot global dengan menggunakan $idf_{i;} \cdot tf_{i;j}$ adalah bobot yang diperoleh dari frekuensi kemunculan kata i pada dokumen j. Sedangkan idf_i adalah bobot yang diperoleh dengan memperhatikan banyaknya kemunculan kata i (DF_i) pada keseluruhan dokumen N. Selanjutnya; bobot total diperoleh dengan mengalikan nilai bobot dengan angka lokal dan nilai berat keseluruhan [4]. Persamaan menunjukkan persamaan matematis dari metode TF-IDF yang digunakan pada rumus dibawah ini:

$$idf_{i;} = log\left(\frac{N}{DF_i}\right) \text{ dan } W_{i;j} = tf_{i;j} x (log\left(\frac{N}{DF_i}\right)) \quad (1)$$

Binary Classification adalah mengklasifikasikan nilai yang tidak diketahui ke dalam salah satu dari dua kelas. Teknik Machine Learning yang berhubungan dengan klasifikasi biner adalah regresi logistik dimana memberikan nilai yang tidak diketahui 0 atau 1 berdasarkan fitur dari nilai lain yang diketahui [5]. Masalah klasifikasi dengan dua label kelas disebut klasifikasi biner. Dalam kebanyakan masalah klasifikasi biner; satu kelas mewakili kondisi normal dan kelas lainnya mewakili kondisi abnormal [6].

Regresi logistik biner adalah metode Pemodelan matematika yang digunakan untuk menganalisis hubungan banyak faktor dengan variabel biner; yaitu 1 untuk berhasil dan 0 untuk gagal. Dalam regresi logistik biner; digunakan beberapa variabel prediktor untuk memprediksi kemungkinan terjadinya keberhasilan atau kegagalan dalam variabel responnya [7].

Regresi logistik adalah algoritma klasifikasi berbasis probabilitas yang sangat populer yang menggunakan pembelajaran terawasi. Ini dikembangkan pada tahun 1940-an untuk melengkapi metode regresi linier dan analisis diskriminan linier.

Meskipun bentuk asli regresi logistik dikembangkan untuk variabel keluaran biner (misalnya 1/0; ya/tidak; lulus/gagal; menerima/menolak); versi modifikasi saat ini mampu memprediksi variabel keluaran multi-kelas (yaitu regresi logistik multinomial).

Evaluation Metrics digunakan untuk mengembangkan dan mengukur performa penyajian model dalam program machine learning yang sedang dibuat. Pada Binary Classification; pada penelitian ini akan menggunakan metode confusion matrix; accuracy; precision; recall; f1-score/f1-measure dan ROC [8].

Tabel 1. Rumus Accuracy; Precision; Recall dan F1-Score/ F1-Measure

Matriks	Formula	Keterangan
Accuracy (acc)	$\frac{tp + tn}{tp + fp + tn}$ -	Secara umum; dalam evaluasi kinerja model klasifikasi; terdapat beberapa beberapa metrik dapat digunakan untuk mengukur kinerja model; salah satunya adalah indeks akurasi yang mengukur persentase prediksi yang benar terhadap jumlah kasus.
Precision (p)	$\frac{tp}{tp + fp}$	Metrik precision yang digunakan untuk mengukur model positif yang diprediksi dengan benar dari total model prediksi kelas positif.
Recall (r)	$\frac{tp}{tp + tn}$	*recall* yang digunakan untuk mengukur fraksi model positif yang diklasifikasi-kan dengan benar.
F1-Score / F1-Measure (FM)	$\frac{2 * p * r}{p + r}$	Matriks ini secara khusus dikenal sebagai rata-rata yang diselaraskan antara ingatan dan akurasi dan merupakan salah satu metode yang umum digunakan untuk mengevaluasi model klasifikasi.

Plot Receiver Operating Characteristic (ROC) merupakan teknik yang digunakan untuk mengilustrasikan; memberikan peringkat; dan memilih pengklasifikasian berdasarkan kinerjanya secara visual [8]. Metode ini sudah lama diterapkan dalam teori deteksi sinyal untuk menggambarkan pertukaran antara keberhasilan klasifikasi dan tingkat alarm palsu. Analisis ROC telah diperluas untuk visualisasi dan analisis perilaku sistem diagnostic [9]. Area di bawah kurva ROC (AUC) adalah ukuran kinerja klasifikasi yang terkenal dan biasa digunakan sebagai ukuran kinerja klasifikasi yang mencakup ambang keputusan dan kemiringan kelas dan biaya [10]. Namun; AUC secara inheren tidak konsisten sebagai ukuran kinerja keseluruhan pengklasifikasian dan menyarankan penggunaan metrik alternatif.

II. METODE

Metode penelitian yang digunakan dalam penyusunan penulisan ini terdiri dari 5 langkah. Langkah-langkah untuk melakukan penelitian ini adalah sebagai berikut:

1. Studi Literatur

Pada fase ini dikumpulkan beberapa bahan referensi yang berkaitan dengan algoritma regresi logistik; diklasifikasikan pesan spam; baik SMS; email; WA dari berbagai sumber skripsi; majalah; buku; artikel dan sumber lainnya.

2. Analisis Masalah

Pada tahap ini menganalisis permasalahan pesan SMS dan WA yaitu tidak terorganisasinya pesan yang ada. Sehingga pesan menjadi bercampur antara yang spam dengan yang no spam (ham).

ISSN: 2580-3336 (Print) / 2549-7952 (Online)
Url: https://proceeding.unpkediri.ac.id/index.php/inotek/

Solusi dari permasalahan tersebut adalah perlunya teknologi berupa message filtering untuk mengatur pesan SMS dan WA secara terorganisir dan efisien.

3. Desain Sistem

Langkah ini menjelaskan alur dari pemodelan sistem yang akan dibuat dan bagaimana proses algoritma yang digunakan agar dapat bekerja sesuai dengan yang diharapkan. Gambaran sistem dan algoritmanya digambarkan dengan flowchart.

4. Tahap Implementasi

Pada tahap ini; implementasi dari desain sistem pemodelan yang dibuat dengan bahasa pemrograman Python dijelaskan; dan dilakukan pelatihan dan pengujian pemodelan yang dibuat dengan data yang telah disiapkan.

5. Tahap Uji Coba

Pada fase ini; pengujian dilakukan dengan membandingkan common train dan data uji sambil memodelkan akurasi.

2.1 Analisa Kebutuhan Penelitian

Pada tahap ini dilakukan Analisa kebutuhan yang dijadikan sebagai acuan dalam pembuatan sistem pada program ini meliputi perangkat lunak dan perangkat keras. Dengan spesifikasi yang digunakan; sebagai berikut :

- HP Pavilion dv4 Notebook
- Prosesor Intel(R) Core (TM) i5
- Intel(R) HD Graphics
- 8GB RAM
- SSD 120GB
- Hard Disk 290GB
- Windows 10 Pro 64-bit

Perangkat lunak yang digunakan adalah:

- Browser Internet (Google Chrome dan Microsoft Edge)
- Google Collaboratory
- Library Machine Learning

III. HASIL DAN PEMBAHASAN

3.1 Pengumpulan Data

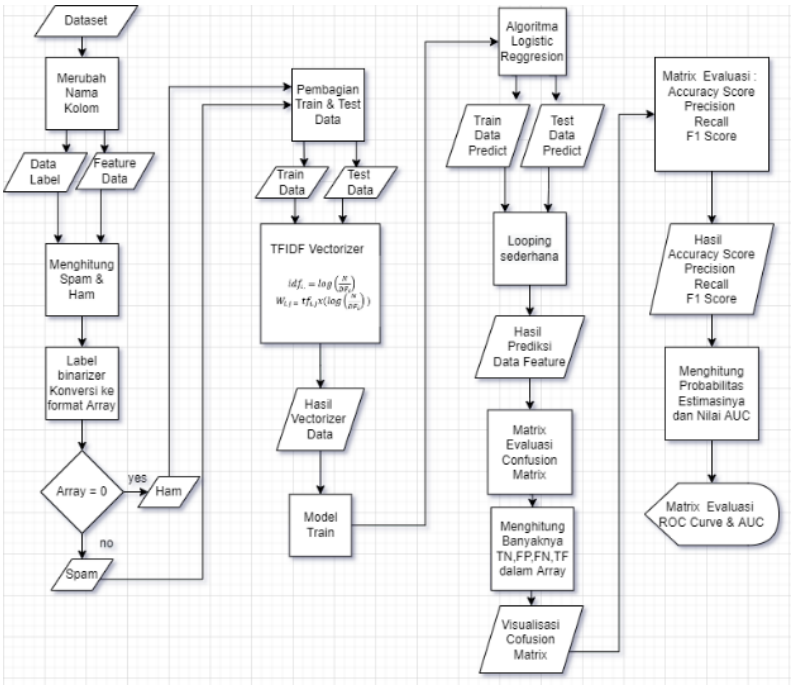
Penelitian ini menggunakan data dari pesan SMS dan WA yang digabungkan dan diperoleh data dari 10.793 pesan berbahasa Indonesia yang ditandai sebagai spam dan spam. Kategori spam menunjukkan bahwa pesan tersebut adalah spam dan ham menunjukkan bahwa pesan tersebut bukan spam. 10.038 data pesan ham dan 755 data pesan spam digunakan untuk membangun model. Spam adalah cara mengirim pesan berulang kali tanpa diketahui pemiliknya. Orang yang mengirim spam disebut spammer; sedangkan proses spamming sendiri disebut spamming [11]. Spam biasanya datang dalam berbagai bentuk; antara lain: Spam email; spam pesan instan; spam mesin pencari; spam blog; spam media sosial; dan iklan online.

3.2 Tahap Penelitian

Pada poin ini terdapat beberapa diagram yang menjelaskan proses penelitian ini; yaitu:

3.2.1 Kerangka penelitian

Penelitian ini dilakukan dengan merancang suatu kerangka penelitian seperti yang terdapat pada gambar 1; yang menjelaskan alur proses yang digunakan untuk menggambarkan keseluruhan penelitian.



Gambar 1 Diagram Alur Program

Gambar 1 menunjukkan sebuah diagram alur program yang akan dilaksanakan selanjutnya. Di bawah ini adalah penjelasan dari gambar di atas.

3.2.2 Preprocessing Data

Dalam hal ini; dilakukan preprocessing data; yang melibatkan penginputan data; pemberian nama pada data dengan memanipulasi data; dan penghitungan jumlah data spam dan ham .

3.2.2.1 Penginputan Data dan Manipulasi Kolom

Dalam penginputan data dan manipulasi kolom; dilakukan entri data; penamaan data dengan manipulasi data; spamming dan perhitungan data. Setelah itu; dilakukan proses pemberian nama pada data tersebut; menggunakan kolom pertama berisi data ham dan spam sebagai pengenal data dan data kolom kedua berisi SMS dan WA sebagai data karakteristik. Hasil dari labeling data tersebut dapat dilihat pada Gambar 2.

	label	sms
0	ham	marhaban ya ramadhan
1	ham	sebelum menjalankan ibadah puasa . mohon maaf ...
2	ham	Assalamualaikum warahmatullahi wabarakaatuuh
3	ham	Assallammu'alaikum wr.wb !!
4	ham	Izinkan saya utk mengajak kluarga utk berpatis...

Gambar 2 Hasil Labeling Data

3.2.2.2 Menghitung Banyaknya Data Spam dan Ham

Selanjutnya; dilakukan prose penghitungan banyaknya data spam dan ham dari data yang tersimpan dalam tabel. Hasil dari perhitungan tersebut terlihat pada Gambar 3.

```
ham     10038
spam      755
Name: label, dtype: int64
```

Gambar 3 Hasil Count Ham dan Spam

3.2.3 Pembagian Data dan Konversi

3.2.3.1 Label Binarizer / Konversi ke Array

Pada tahap pembagian data dan konversi; langkah pertama adalah mengubah data yang sebelumnya ditandai sebagai spam 1 dan ham 0 menjadi array satu dimensi [12]. Berikut cuplikan program dan outputnya seperti pada Gambar 4

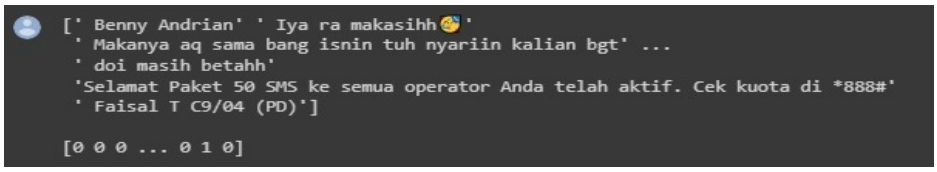
```
array(['ham', 'spam'], dtype='<U4')
```

Gambar 4 Konklusi Konversi ke Array

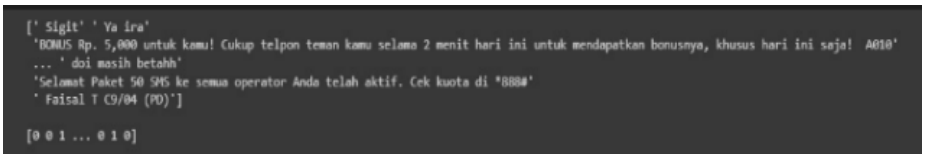
3.2.3.2 Pembagian Data Training dan Testing

Setelah itu; dilakukan pembagian data untuk training dan testing. Ada beberapa percobaan pembagian data yang dilakukan; yaitu 50% data training dan 50% data testing; 60% data training

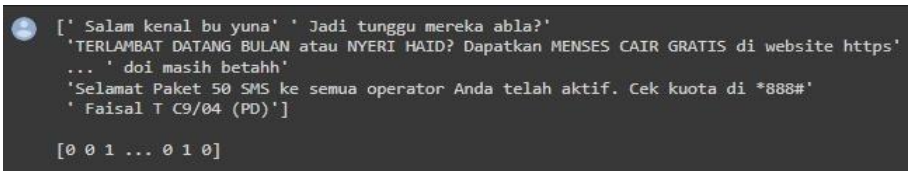
dan 40% data testing; 75% data training dan 25% data testing; serta 80% data training dan 20% data testing [13]. Hasil dari pembagian data tersebut terlihat pada Gambar 5-8.



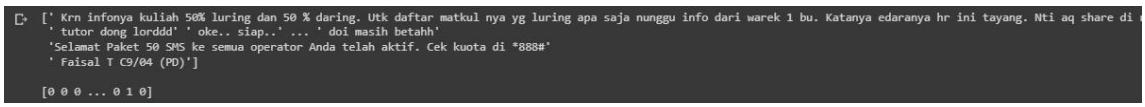
Gambar.5 Hasil Split Data (5050)



Gambar 6 Hasil Split Data (6040)



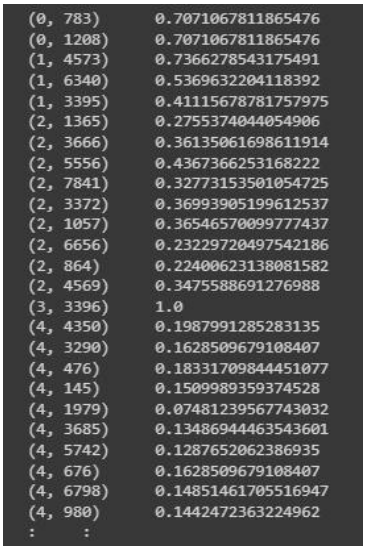
Gambar 7 Hasil Split Data (7525)



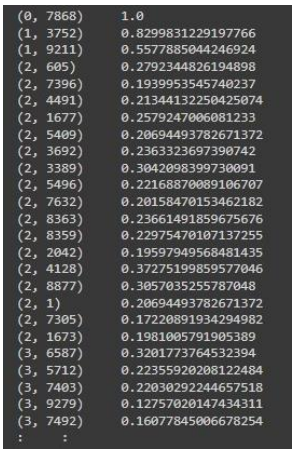
Gambar 8 Hasil Split Data (8020)

3.2.4 Count Vectorization TF-IDF

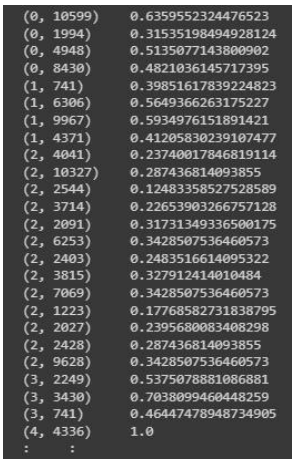
Langkah selanjutnya adalah melakukan count vectorization TF-IDF pada data karakteristik yang diperoleh; yang mengubah data berupa teks menjadi nilai vektor berupa angka. Hasil dari count vectorizer tersebut terlihat pada Gambar 9-12.



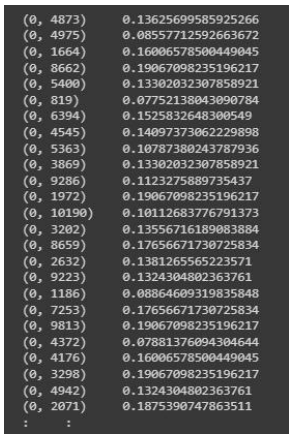
Gambar 9 Count Vectorizer (5050)



Gambar 10 Count Vectorizer (6040)



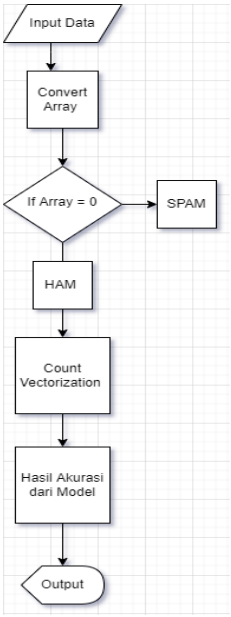
Gambar 11 Count Vectorizer (7525)



Gambar 12 Count Vectorizer (8020)

3.2.5 Testing

Pada tahap testing; dibuat suatu model train yang menggunakan algoritma regresi logistik; yang kemudian dijalankan melalui loop sederhana untuk memeriksa hasil akurasi model. Hasil dari pemodelan regresi logistik tersebut terlihat pada Gambar 13.



Gambar 13 Diagram Alur Prediksi

3.2.6 Klasifikasi Model train menggunakan Algoritma Logistik Regresi

Pada tahap klasifikasi model train dengan algoritma regresi logistik; terlebih dahulu dilakukan prediksi berdasarkan data fitur menggunakan model train.

3.2.6.1 Looping sederhana

Selanjutnya; dilakukan looping sederhana untuk memverifikasi prediksi dan mencetak hasil prediksi.

3.2.6.2 Hasil Prediksi

Terakhir; cetak hasil prediksi; tiga langkah akan menampilkan multiple split test yaitu 50 training dan 50 test; 60 training dan 40 test; 75 training dan 25 test; 80 train dan 20 test bertujuan untuk menunjukkan hasil terbaik. Berikut cuplikan program dan outputnya seperti pada Gambar 14.

```
PRED: 0 - SMS:  Cakep gak nya aq ga tau aq ga beli😡
PRED: 0 - SMS:  bolehhh kan lu satpam nya dull wkwkwk canda, kalo ga boleh tenang ada maya, dikedipin sama maya nanti yg ngusir pasti kelepek2
PRED: 0 - SMS:  Manut saja mbak
PRED: 0 - SMS:  Jam 18.30 🙏
PRED: 0 - SMS:  Bahan 1 roll mahal ya?
```

Gambar 14 Pemodelan Regresi Logistik

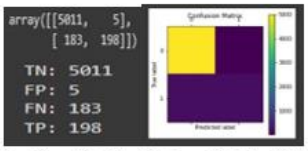
3.2.7 Evaluasi Matriks

Tahap evaluasi matriks dilakukan untuk menentukan model yang sesuai dengan kumpulan data yang digunakan. Evaluasi matriks ini meliputi Confusion Matrix Accuracy Score; Precision;

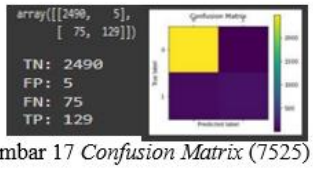
Recall; F1 Score; dan ROC Curve dan AUC pada binary classification. Pada tahap evaluasi ini; dilakukan beberapa percobaan pembagian data; 50 training dan 50 test; 60 training dan 40 test; 75 training dan 25 test; 80 train dan 20 test. Berikut beberapa model penilaian matriks; yaitu:

3.2.7.1 Confusion Matrix

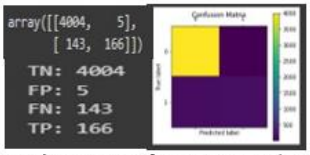
Hasil dari beberapa model penilaian matriks tersebut ditampilkan dalam bentuk visualisasi [14] nilai TP; TN; FP; dan FN pada suatu array 2 dimensi dan hasil colorbar dalam Confusion Matrix; berikut hasilnya :



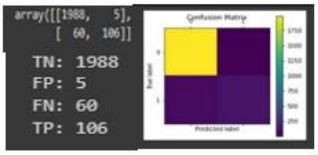
Gambar 15 *Confusion Matrix* (5050)



Gambar 17 *Confusion Matrix* (7525)



Gambar 16 *Confusion Matrix* (6040)



Gambar 18 *Confusion Matrix* (8020)

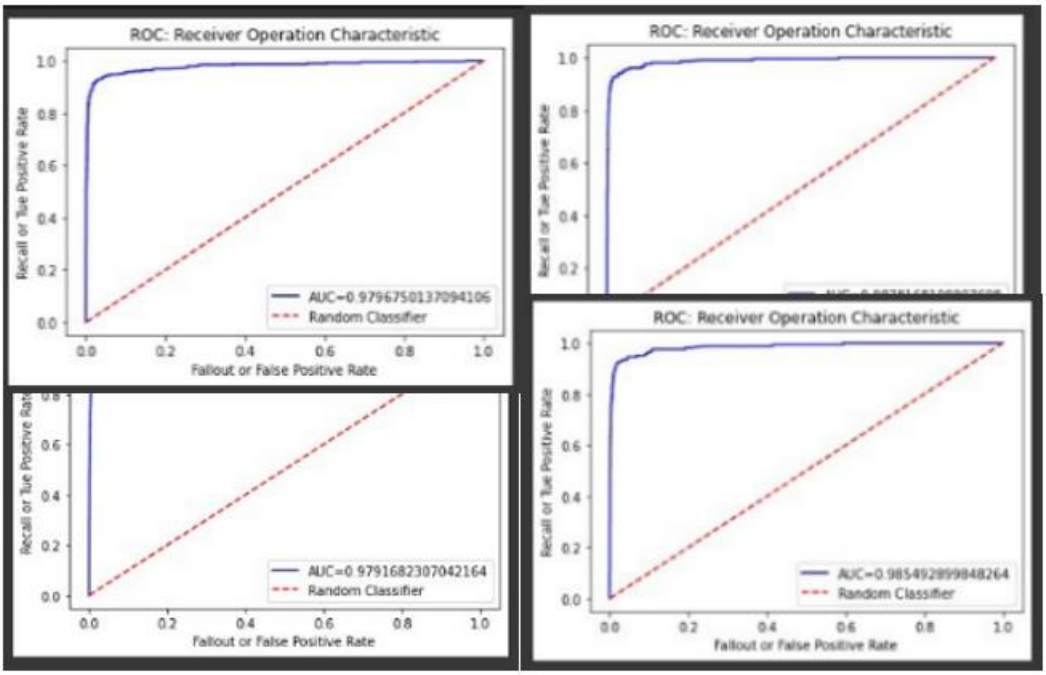
3.2.7.2 Accuracy Score; Precision; Recall & F1 Score

Pada evaluasi berikutnya itu Accuracy Score; Precision; Recall & F1 Score yang akan menampilkan nilai desimal atau persentase nya; beserta hasilnya; yaitu :

Gambar 19 Hasil Accuracy Score; Precision; Recall & F1-Score/F1-Measure

3.2.7.3 ROC Curve & AUC

Pada tahap evaluasi terakhir kali ini yaitu ROC Curve & AUC yang memvisualisasikan kinerja classifier dan menunjukkan hasil perbandingan antara nilai Recall (TPR) dan Fallout (FPR) [15]; hasilnya adalah sebagai berikut:



Gambar 20 Hasil Visualisasi ROC Curve dan AUC

IV. KESIMPULAN

Berdasarkan penelitian ini; dari 10793 data dengan 10038 ham dan 775 spam menggunakan metode algoritma regresi logistik untuk melakukan klasifikasi biner pada data ham dan spam. Studi ini menyediakan lima matriks evaluasi yang dapat membantu memvisualisasikan hasil model yang telah dihasilkan menggunakan algoritma regresi logistik dan beberapa dekomposisi data; yang akan menemukan hasil terbaik selama proses pembuatan model. Berdasarkan kesimpulan yang telah diuraikan diatas; maka ada beberapa saran yang bisa disampaikan yaitu dari model yang sudah dibuat dapat diimplementasikan dalam sebuah aplikasi pengelompokkan atau deteksi sebuah pesan tersebut apakah berupa spam atau ham. Pada penelitian selanjutnya untuk prediksi bagiian kllasfikasi dapat dikembangkan dengan menggunakan jenis data yang sama dengan menggunakan metode algoritma yang lain seperti Decion Tree; Neural Network; Support Vector Machines; KNN; GA dan ANN.

DAFTAR PUSTAKA

[1] H. Yandhi; "Prototype E-Polling Berbasis Sms Gateway Pada Pemilihan Ketua Rw. 06 Perum. Bugel Mas Indah;" Ict Learning; vol. 3; no. 1; pp. 45-64; 2017.

[2] A. Febriyanti; "Analisis Sentimen Persepsi Pengguna Jne Menggunakan Algoritma Naïve Bayes Classifier;" 16522259; 2018.

[3] R. Kumari and S. K. Srivastava; "Machine Learning: A Review On Binary Classification;" International Journal Of Computer Applications; vol. 160; no. 7; 2017.

[4] F. Syadid; "Analisis Sentimen Komentar Netizen Terhadap Calon Presiden Indonesia 2019 Dari Twitter Menggunakan Algoritma Term Frequency-Invers Document Frequency (Tf-Idf) Dan Metode Multi Layer Perceptron (Mlp) Neural Network;" BS thesis; Fakultas Sains Dan Teknologi Universitas Islam Negeri Syarif Hidayatullah Jakarta; 2019.

[5] T. E. Sutanto; "Deteksi Berita Hoax Pada Website Turnbackhoax Dengan Menggunakan Machine Learning;" BS thesis; Fakultas Sains Dan Teknologi UIN Syarif Hidayatullah Jakarta.

[6] S. S. Septiani; "Klasifikasi Mengkudu Berdasarkan Warna Dan Tekstur Menggunakan Metode Support Vector Machine (SVM);" dissertation; Universitas Muhammadiyah Gresik; 2016.

[7] K. W. Patunduk et al.; "Pemodelan Pasien Covid-19 Di Kota Palopo Dengan Regresi Logistik (Studi Perbandingan Regresi Logistik Dan Analisis Survival);" Proximal: Jurnal Penelitian Matematika Dan Pendidikan Matematika; vol. 5; no. 2; pp. 260-269; 2022.

[8] H. Yan et al.; "Mfe-Net: Multi-Type Feature Enhancement Net For Retinal Blood Vessel Segmentation;" in 2022 5th International Conference On Artificial Intelligence And Big Data (Icaibd); 2022.

[9] N. A. Gusti; "Analisis Sentimen Terhadap Perkuliahan Jarak Jauh Di Masa Pandemi Covid-19 Pada Media Sosial Twitter Menggunakan Textblob Dan Algoritma Support Vector Machine (SVM);" BS thesis; Fakultas Sains Dan Teknologi UIN Syarif Hidayatullah Jakarta.

[10] S. W. Sidehabi; "Implementasi Mesin Pemilah Buah Markisa Berdasarkan Tingkat Kematangan Berbasis Visi Komputer;" dissertation; Universitas Hasanuddin; 2019.

[11] F. W. Giffary; "Text Classification Untuk Mendeteksi Spam Di Media Sosial Twitter Menggunakan Tf-Idf Dan Algoritma Multilayer Perceptron;" dissertation; Universitas Pembangunan Nasional "Veteran" Yogyakarta; 2022.

[12] Z. M. Yusuf and R. M. Awangga; "Deteksi Spam Sms Menggunakan Naive Bayes;" Penerbit Buku Pedia; 2023.

[13] B. N. Azmi; A. Hermawan; and D. Avianto; "Analisis Pengaruh Komposisi Data Training Dan Data Testing Pada Penggunaan Pca Dan Algoritma Decision Tree Untuk Klasifikasi Penderita Penyakit Liver;" JTIM: Jurnal Teknologi Informasi Dan Multimedia; vol. 4; no. 4; pp. 281-290; 2023.

[14] R. M. Pradhana; "Analisis Sentimen Publik Terhadap Kebijakan Pemberlakuan Pembatasan Kegiatan Masyarakat Skala Mikro Menggunakan Algoritma Support Vector Machine Studi Kasus Twitter;" dissertation; Universitas Dinamika; 2021.

[15] N. I. Putri; "Deep Learning Dan Teknologi Big Data Untuk Keamanan Iot;" Computing| Jurnal Informatika; vol. 7; no. 1; pp. 48-73; 2020.