Metode Klasifikasi Gejala Penyakit *Coronavirus Disease 19* (COVID-19) Menggunakan Algoritma *Neural Network*

**Rahmi[1], Darius Antoni[2], Hadi Syaputra[3], Fatoni[4] , Tri Basuki Kurniawan[5]**

Program Studi Magister Tehnik Informatika [1],[2],[3],[4],[5]

Fakultas Ilmu Komputer

Universitas Bina Darma Palembang

E-mail:[puspitarahmi9@gmail.com](mailto:puspitarahmi9@gmail.com), [darius.antoni@binadarma.ac.id,](mailto:darius.antoni@binadarma.ac.id2) [hadisyaputra@binadarma.ac.id,](mailto:hadisyaputra@binadarma.ac.id,)

fatoni@binadarma.ac.id, [tribasukikurniawan@binadarma.ac.id](mailto:tribasukikurniawan@binadarma.ac.id) [1],[2],[3],[4],[5]

***Abstrack*­— Coronavirus Disease 19 (COVID-19) is a different virus from Severe Acute Respiratory Syndrome Associated Coronavirus (SARS-COV2), Coronavirus Disease 19 (COVID-19) is a new virus that causes infections of the respiratory tract. Covid-19 can occur due to transmission of the virus from animals to humans by direct contact with infected animals or referred to as zoonotic transmission and can be transmitted from humans to humans also by direct contact or exposure to saliva splashes. According to data from epimology, the average age of those infected with this virus is at the age of 15-89 years, symptoms of COVID-19 occur in someone who has unstable immunity, generally showing symptoms of shortness of breath, sore throat, and coughing runny nose to fever. Indonesia with the first positive case with 2 people in March 2020. In this study the author raised the issue of how to classify the risk of contracting the Covid-19 virus from the symptoms caused. The purpose of this study was to determine the level of risk of contracting the covid-19 virus using the Cross Industry Standard Process for Data Mining (CRISP-DM) method. In this study, 2 comparison algorithms were used, namely Neural Network (NN) and Logistic Regression using Python tools. The accuracy value for the dataset with 278.848 data records using the Neural Network (NN) algorithm is 95%, while the accuracy value for the dataset using the Logistic Regression algorithm is 94%. In this study, classification using the Neural Network algorithm provides a higher aura value than using the Logistic Regression algorithm.**

***Keyword: Covid-19, Neural Network, Logistic regression, classification, CRISP-DM***

***Abstrak*— *Coronavirus Disease* 19 (COVID-19) merupakan virus yang berbeda dengan *Severe Acute Respiratory Syndrome Associated Coronavirus* (SARS-COV2), *Coronavirus Disease* 19 (COVID-19) merupakan virus baru yang menyebabkan infeksi pada saluran pernapasan. Covid-19 ini dapat terjadi karena adanya penularan virus dari hewan ke manusia dengan kontak langsung pada hewan yang sudah terinfeksi atau disebut sebagai *Transmisi Zoonosis* dan dapat tertular dari manusia ke manusia juga dengan kontak langsung atau terkena percikan liurnya. Menurut data dari epimologi rata-rata usia yang terjangkit virus ini yaitu pada usia 15-89 tahun, gejala covid-19 terjadi pada seseorang yang memiliki imunitas tidak stabil umumnya menunjukkan gejala sesak nafas, sakit tenggorokan, dan batuk pilek hingga demam. Indonesia dengan kasus positif pertama kali dengan 2 orang pada maret 2020. Dalam penelitian ini penulis mengangkat masalah bagaimana pengklasifikasian resiko terjangkit virus covid-19 dari gejala yang ditimbulkan. Tujuan penelitian ini yaitu untuk mengetahui tingkat resiko terjangkit virus covid-19 dengan menggunakan metode *Cross Industry Standard Process for Data Mining* (CRISP-DM). Dalam penelitian ini digunakan 2 algoritma pembanding yaitu *Neural Network (NN)* dan *Logistic Regression* dengan menggunakan *tools Phyton.* Nilai akurasi pada dataset dengan 278.848 record data menggunakan algoritma *Neural Network (NN)* memperoleh 95%, sedangkan nilai akurasi pada dataset menggunakan algoritma *Logistic Regression* memperoleh 94%. Pada penelitian ini klasifikasi menggunakan algoritma *Neural Network (NN)* memberikan nilai akurasi yang lebih tinggi dibandingkan dengan menggunakan algoritma *Logistic Regression.***

**Kata kunci: Covid-19,*Neural Network, Logistic Regression,* klasifikasi, CRISP-DM**

# PENDAHULUAN

*World Health Organization* (WHO) atau biasa dikenal dengan Organisasi Kesehatan Dunia memberitahukan kasus baru Pneumonia di Kota Wuhan, Hubei, China yang mengidentifikasi jenis baru *novel coronavirus*. Nama *Coronavirus Disease 2019* (COVID-19) resmi ditetapkan pada 12 Februari 2020 oleh WHO (WHO, 2020). Menurut ahli virologi dari China, virus covid-19 ini merupakan virus yang berbeda dengan *Severe Acute Respiratory Syndrome Associated Coronavirus* (SARS-COV2) yang muncul di Guangdong, China Tahun 2003 (Peiris et al., 2003), tetapi memiliki gejala yang sama (WHO,2020). Tingkat penyebaran Covid-19 lebih luas dibandingkan dengan SARS namun tingkat kematian SARS mencapai 9,6% dibanding tingkat kematian Covid -19 yang masih di bawah 5%.

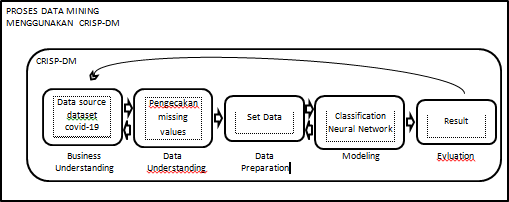
[1]Ada 4 (empat) kelompok dalam penyebutan Covid-19, dalam pedoman WHO dan Stigma Orang Positif Covid-19 yaitu Orang Dalam Pantauan (ODP), Pasien Dalam Pengawasan (PDP), Orang Tanpa Gejala (OTG), sedangkan Positif adalah pasien yang telah melakukan test menunjukan hasil positif covid-19. Kasus di Indonesia hingga 28 Desember 2020 sebanyak 719.219 ribu orang dengan kasus per 14 hari sebanyak 95 rbu dan untuk kasus meninggal mencapai 21 ribu jiwa. Di Indonesia telah melakukan upaya untuk memperlambat penyebaran covid-19 dengan istilah 3M yaitu menjaga jarak, memakai masker, dan mencuci tangan serta melakukan social distancing. Dalam penelitian ini menggunakan 2 (dua) algoritma yang yaitu [2]*Neural Networ*k dan *Logistic Regression* dengan menggunakan *software* phyton untuk mempermudah dalam menganalisis.

# METODOLOGI PENELITIAN

Metode pengumpulan data dalam penelitian ini ialah data gejala penyakit Coronavirus Disease 2019 (COVID-19) yang diambil dari salah satu artikel yang juga membahas tentang prediksi diagnosis covid-19 yang juga diambil dari situs web Kementrian Kesehatan Israel. Dengan kolom gejala 0 (not present) 1 (present), untuk kolom resiko 2 kelas yaitu positif dan negatif dengan menggunakan tools Phyton. Dataset ini digunakan sebagai bahan penelitian klasifikasi untuk mengetahui risiko seseorang terjangkit covid-19.

## Alur Penelitian

## Alur dari penelitian yang dilakukan pada penelitian ini dapat dilihat seperti pada Gbr. 1 berikut ini.

**

Gbr. 1. Alur Penelitian

* 1. Metodologi *Cross-Industry Standard Process for Data Mining* (CRISP-DM)

Pada penelitian ini disesuaikan dengan tahapan pada metode CRISP-DM. Tahapan yang digunakan mengacu pada penelitian ini, terdapat 6 tahapan pada metodologi CRISP-DM[3] yaitu :

1. *Business Understanding*

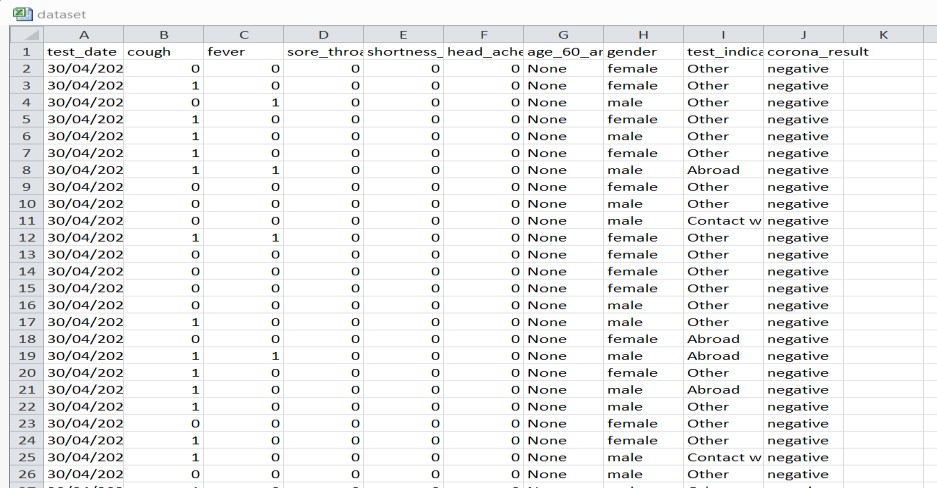
*Business understanding,* tahap pertama dalam proses CRISP-DM juga dapat disebut sebagai tahap pemahaman penelitian. Pada tahap ini dibutuhkan pemahaman tentang substansi dari kegiatan *data mining* yang akan dilakukan serta kebutuhan dari perspektif bisnis. Dalam rangka untuk memahami data yang kemudian dianalisis, sangat penting untuk sepenuhnya mempelajari objek bisnis (objek penelitian).

Penerapan data mining pada penelitian ini berhubungan langsung dengan data gejala covid- 19 yang diperoleh dari kaggle dataset untuk mengklasifikasikan resiko terjangkit virus covid- 19 menggunakan algoritma Neural Network. Ada 3 atribut yang akan digunakan untuk melaksanakan klasifikasi, yaitu tingkta resiko rendah, tingkat resiko sedang dan tingkat resiko tinggi.

1. *Data Understanding*

*Data Understanding* atau pemahaman data adalah tahapan mengumpulkan data awal dan mempelajari data tersebut untuk bisa mengenal dan memahami apa saja yang bisa dilakukan pada data-data itu. Pemahaman data mengacu pada database gejala covid-19, tahap memahami format data secara permukaan *(format form* dan *report)* dan secara lebih mendalam (bentuk fisik data).

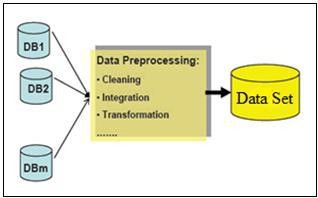
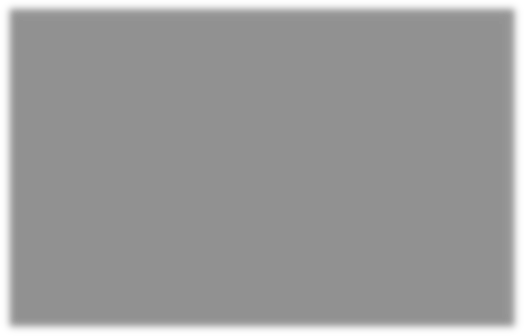
Dataset yang akan digunakan yaitu diambil dari situs resmi kaggle. Ada 17 atribut yang akan digunakan untuk melaksanakan klasifikasi, yaitu suhu tubuh, batuk kering, sakit tenggorokan, kelelahan, masalah pernapasan, kantuk, nyeri dada, riwayat perjalanan ke Negara terinfeksi, diabetes, penyakit jantung, penyakit paru-paru, stroke atau penurunan kekebalan tubuh, perkembangan gejala, tekanan darah tinggi, penyakit ginjal, perbahan nafsu makan, kehilangan indra penciuman. Beberapa atribut tersebut merupakan parameter yang akan digunakan untuk melakukan klasifikasi resiko terinfeksi gejala covid-19, seperti pada Gbr. 2:

 Gbr. 2. Data Gejala Civid-19

1. *Data Preparation*

Referensi[4] *Data preparation* mencakup semua kegiatan untuk membangun *data set* yang akan dimasukkan ke dalam alat pemodelan dari data mentah awal atau membuat *database* baru untuk *set up data mining*. *Database* ini bersifat independen atau terpisah dari *database* operasional. Fungsi utamanya khusus untuk alat pemodelan *data clustering*. Persiapan data merupakan tahap yang padat karya. Tahap membangun *database* baru sebagai set data akhir untuk pemodelan *data mining clustering*.

Gbr. 3. Ilustrasi Proses *Data Preparation*



* 1. *Data set Description*

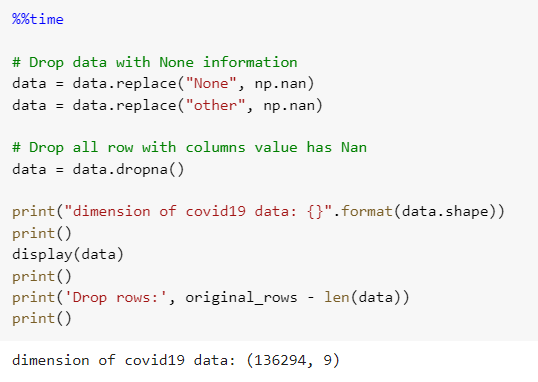
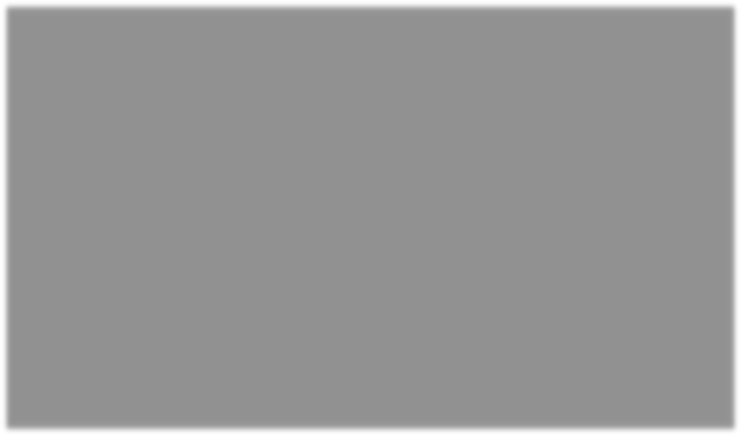
Tahap merancang format set data mentah pada database parsial yang digunakan sebagai gudang data yang digunakan untuk pemodelan data mining. Rancangan set data harus menyesuaikan dengan apa yang telah dirumuskan pada tahapan business unserstanding terutama pada perumasan tujuan data mining.

TABLE I.DATASET COVID-19

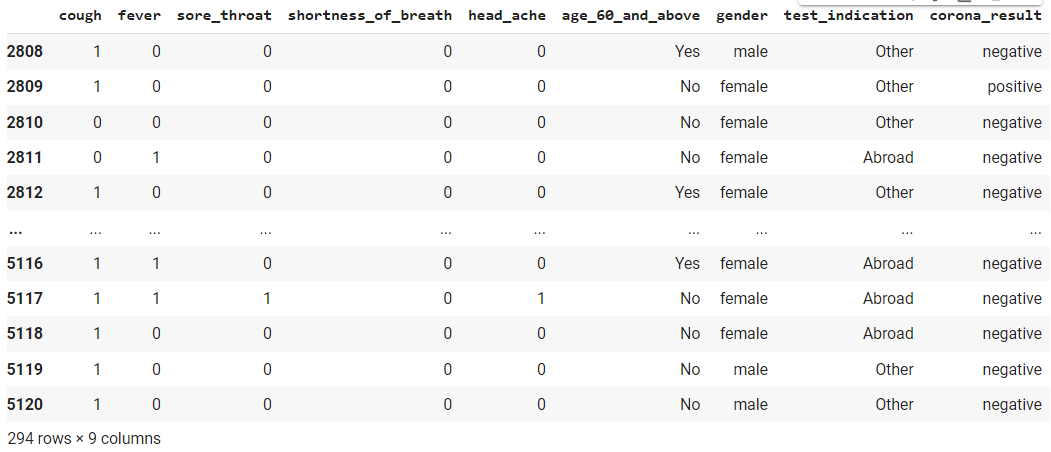
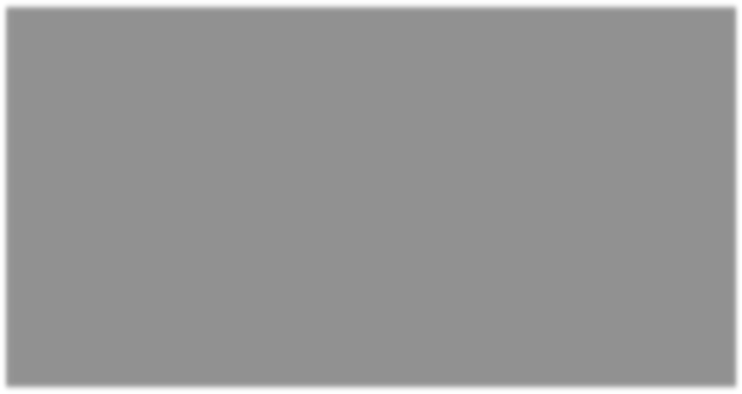
|  |  |
| --- | --- |
| No | Gejala Penyakit Covid-19 |
| 1 | Batuk |
| 2 | Demam |
| 3 | Sakit Tenggorokan |
| 4 | Sesak napas |
| 5 | Sakit kepala |
| 6 | Kontak langsung dengan seseorang yang dikonfirasi Covid-19 |

*a) Clean Data*

Tahap ini memastikan data yang dipilih, dikonstruksi dan diintegrasikan telah layak untuk ditambang. Kegiatannya antara lain membersihkan dan memperbaiki data yang rusak, menghapus data yang tidak di perlukan, menyeragamkan data yang dianggap sama namun memiliki nilai yang berbeda atau membuatnya menjadi konsisten. Pada tahap *data cleaning*, dataset yang terdiri dari 278.848data tidak terdapat *missing values*, dapat dilihat pada Gbr. 4:



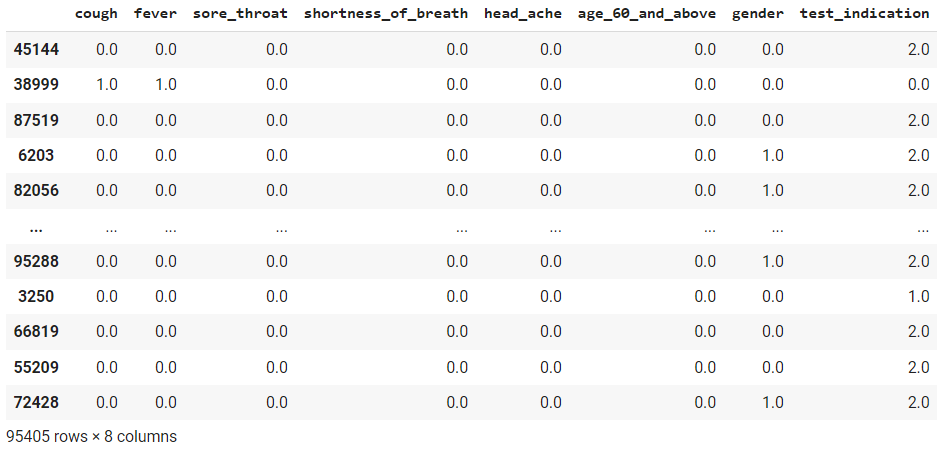
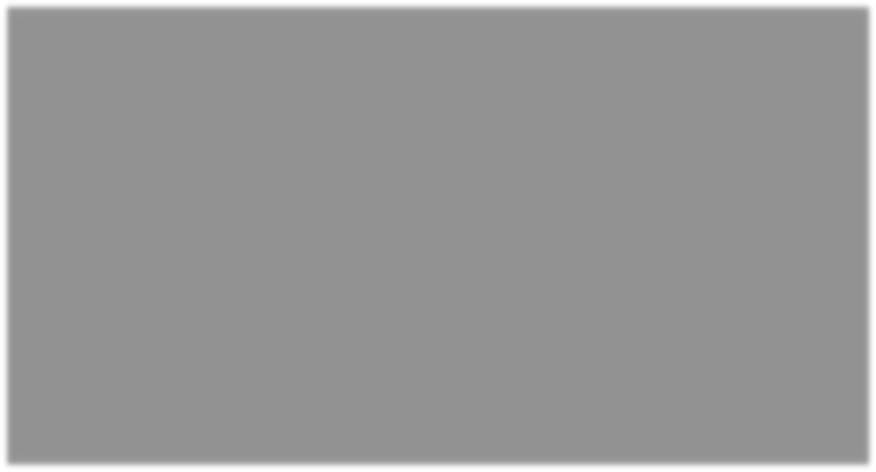
Gbr. 4. *Cleaning Data*



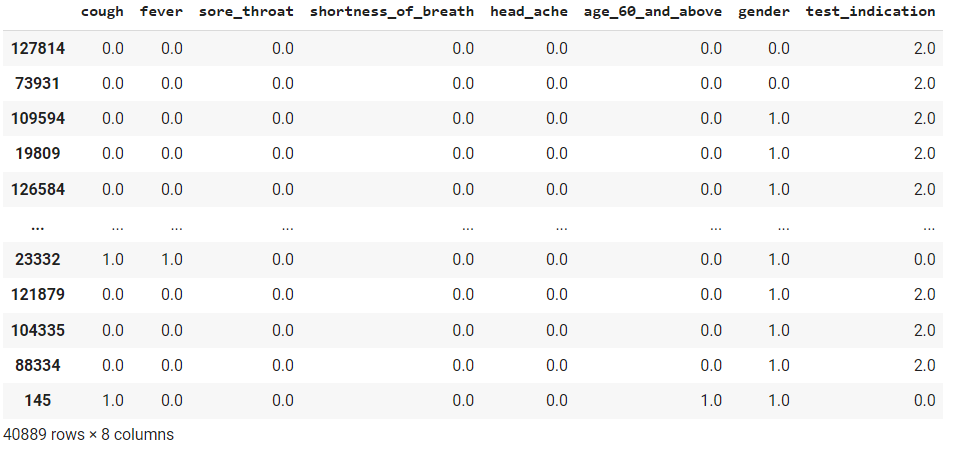
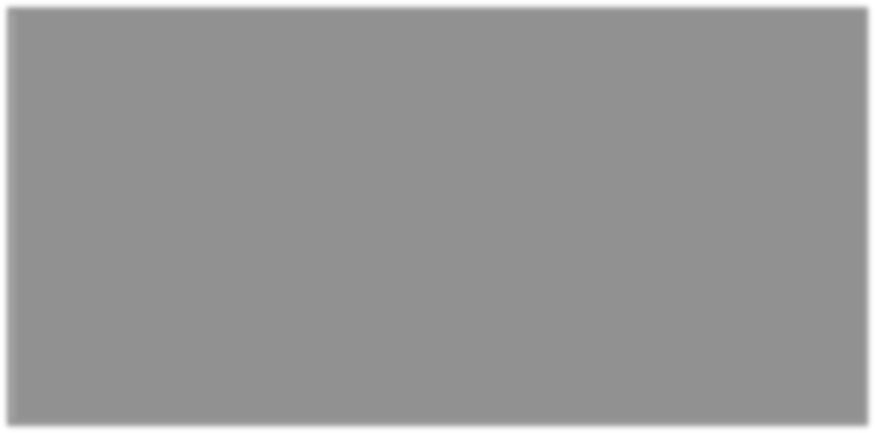
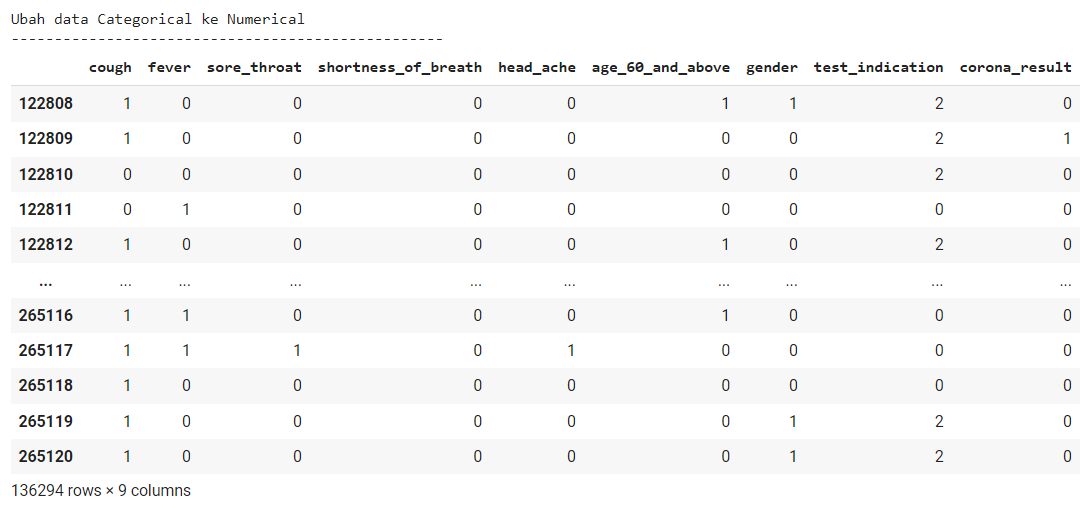
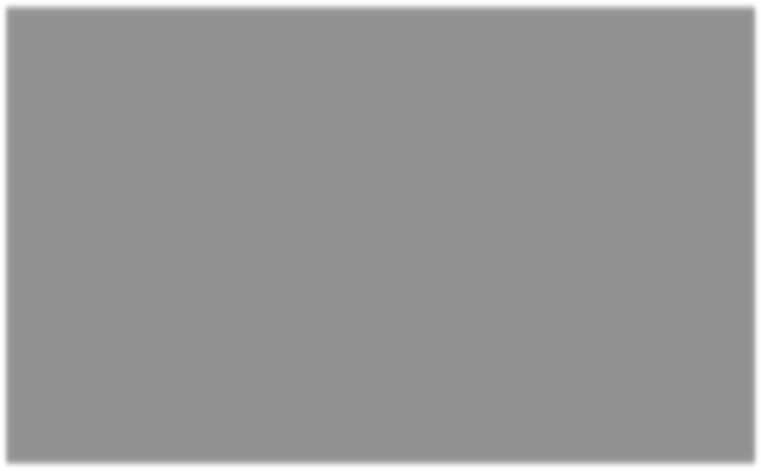
Gbr. 5. Hasil setelah dilakukan cleaning data

Setelah di analisis dan dilakukan proses cleaning data pada dataset, yang awalnya terdapat 278.847 data ternyata setelah dilakukan tahap cleaning data jumlah data menjadi 136.294 data dan 9 kolom.

### Convert Data



Dalam nelakukan pemodelan menggunakan Multi-layer Perceptron (MLP) data yang digunakan untuk melakukan pemodelan yaitu berupa data numeric, sehingga data yang masih berupa categoricsl diubah menjadi data numerical seperti gambar 6:



Gbr. 6. Data setelah dilakukan ubah data categprical menjadi numerical

* 1. *Data Splitting*

Adapun data yang digunakan yaitu data gejala-gejala pasien resiko covid-19, dalam penelitian ini data yang digunakan sebanyak 136.294 record data dengan 8 kolom. Evaluasi model machine learning dengan train dan test split cocok digunakan untuk dataset yang berukuran besar, train/test split membagi dataset menjadi *train set* dan *test set* atau data yang digunakan untukk proses training dan testing merupakan kumpulan data yang berbeda. Dalam penelitian ini penulis membagi data train set dan test set menjadi 30% untuk data test set dan 70% untuk data train set. Dapat dilihat pada Gbr. 7:

Gbr. 7. Split data Train dan Data Test



Dapat dilihat pada Gbr. 8 ditunjukkan bahwa unutk training set terdapat 95.405 baris dan 8 kolom, sedangkan untuk testing set terdapat 40.889 baris dan 8 kolom yang akan digunakan untuk pemodelan menggunakkan algoritma neural network dan logistic regression

Gambar 8. Informasi jumlah data training

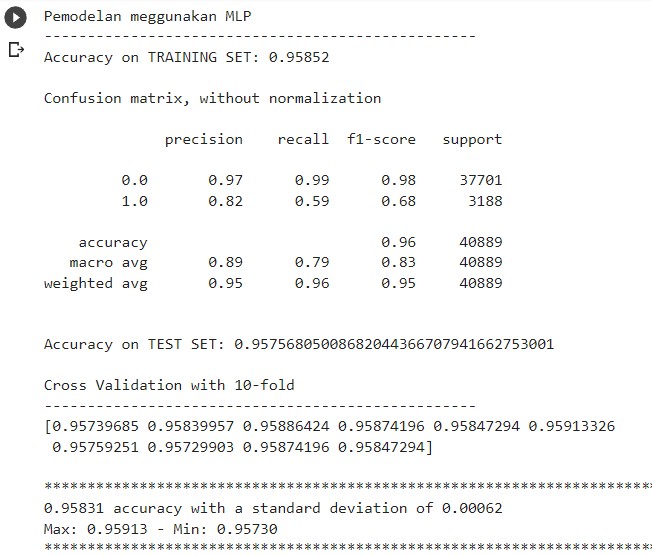
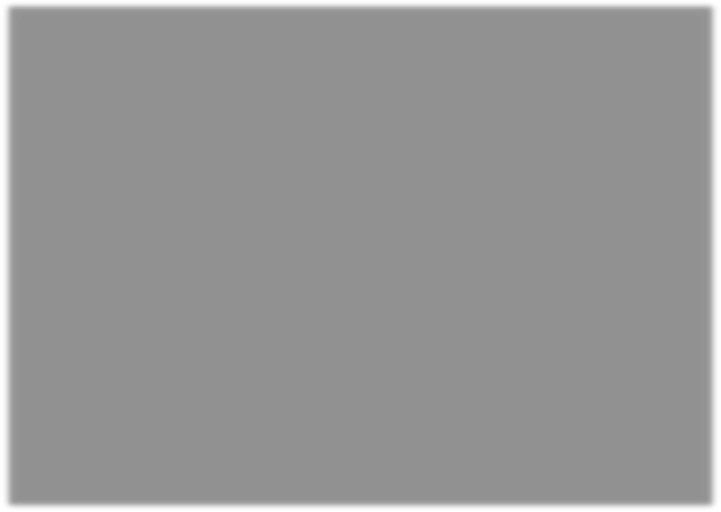
Gbr 8. Informasi jumlah data training dan testing

# hasil dan pembahasan

Setelah melakukan analisis terhadap dataset gejala Covid-19dengan dua label klasifikasi gejala yang ditimbulkan oleh pasien Covid-19 dengan menggunakan Algoritma *Neural Network* dan *Logistic regression*. Hasil yang dicapai oleh peneliti adalah mengetahui klasifikasi gejala penyakit Covid-19 berdasarkan dataset yang telah diunduh dari salah satu website resmi pada artikel yang meneliti tentang kasus covid-19. Peneliti menggunakan tools Phyton untuk melakukan proses *data science*, yang memberikan hasil informasi klasifikasi gejala Covid-19. Dengan menggunakan dataset sebanyak 136.294 data dengan training dan testing 70:30 telah dilakukan pengujian masing-masing algoritma. Dari pengujian 2 algoritma ini didapatkan hasil algoritma klasifikasi gejala penyakit covid-19 dengan akurasi yaitu untuk algoritma neural network dengan model MLP didapatkan jumlah 95.8% untuk training set akurasu dan 95.7% untuk test set akurasi sedangkan neural network setelah dilakukan normalisasi menggunakan normalisasi dengan *MinMaxScaller* 95.8% untuk training set *accuracy* dan 93.7% untuk test set *accuracy*. Sedangkan untuk algoritma *Logistic Regression* didapatkan akurasi sebesar 94.48% untuk training set dan 94.49% untuk test set akurasi, sedangkan metode *logistic regression* setelah dilakukan normalisasi menggunakan normalisasi *MinMaxScaller* didapatkan 94.38% untuk training set akurasi dan 94.39% untuk test set akurasi.

## Neural Network (NN)

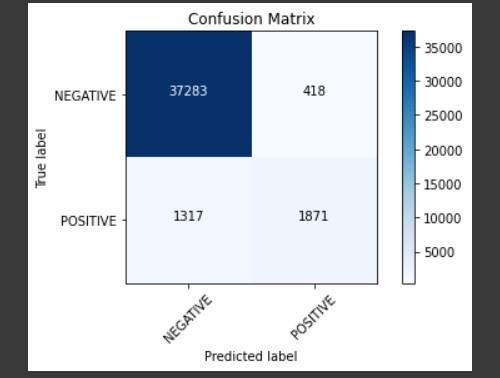
Pengujian menggunakan data gejala covid-19 sebanyak 136.294 record data dengan 6 kolom, dengan *split data* yang ada pada *machine learning* digunakan untuk membagi antara *data training* (data latih) dan *data testing* (data uji). Penulis membagi data menjadi 70% (0.7) untuk *data training* (data latih) dan 30% (0.3) untuk *data testing* (data uji). Kemudian digunakan *Algoritma neural network* sebagai metode klasifikasiyang digunakan, setelah itu hubungkan split data dan algoritma yang digunakan pada *apply model* untuk mengetahui data mana yang tidak memiliki label pada data testing. Pada penelitian ini penulis menggunakan *Multi-Layer Perceptron (MLP)* sebagai model jaringan saraf tiruan (JST) dengan karakteristik memiliki nilai bobot yang lebih baik untuk menghasilkan klasifikasi yang lebih akurat[5]. Pada gambar 10 menunjukkanpemodelan neural network menggunakan MLPClassifier sebelum dilakukan proses klasifikasi menggunakan algoritma *neural network* menggunakan model MLP. Setelah dilakukan pemodelan pada algoritma *neural network* dengan menggunakan model MLP Classifier maka akan dilakukan klasifikasi dataset dengan algoritma *neural network* menggunakan model MLP, proses klasifikasi dapat dilihat pada Gbr. 9:



Gbr. 9. Model klasifikasi Neural network dengan model MLP

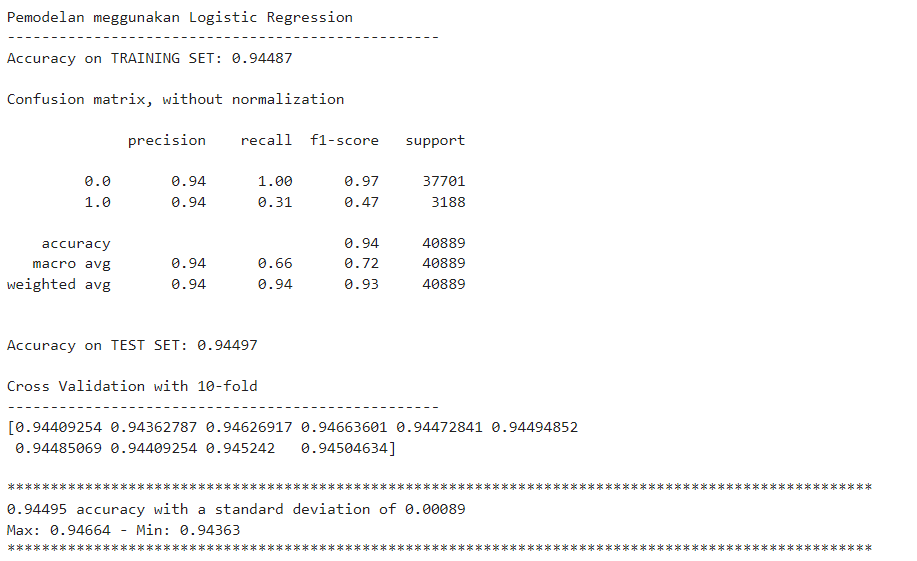
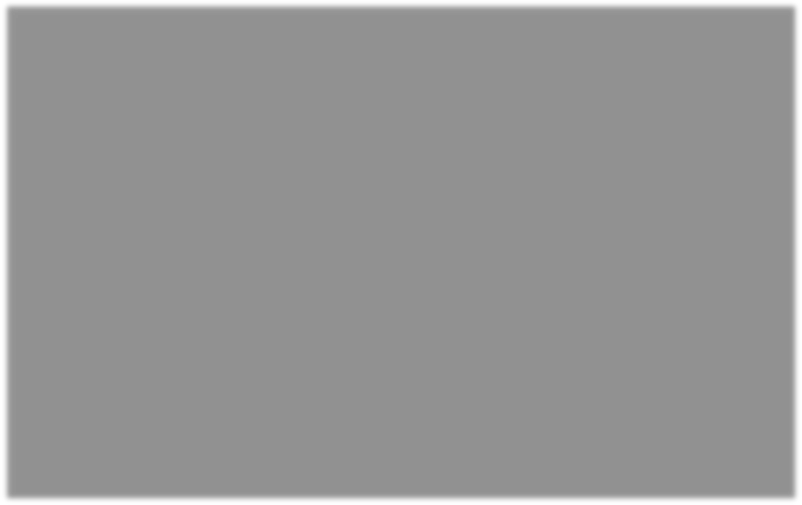
Hasil akurasi pada data training sebesar 95.8% dan pada data testing sebesar 95.7% dari algoritma neural network menggunakan model multi-layer perceptron (MLP) dan dengan melakukan proses validasi data dengan *k-fold cross validation* Sebanyak 10 kali dari data gejala covid-19.

Gambar 10. Confusion Matrix MLP

Pada gambar 10 dapat dilihat nilai dari confusin matrix data gejala covid-19 dimana nilai aktual untuk label negative sebanyak 37.283 data dan nilai prediksi sebanyak 418, sedangkan nilai aktual untuk label positif sebanyak 1.871 data dan nilai prediksi sebanyak 1.317 data.

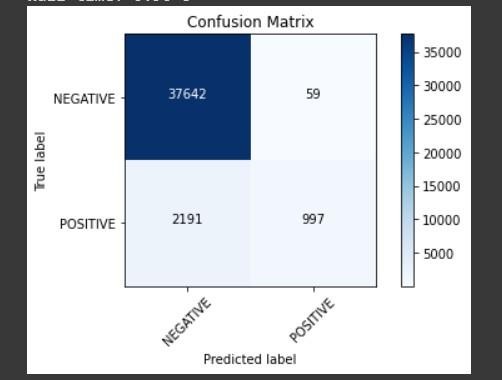
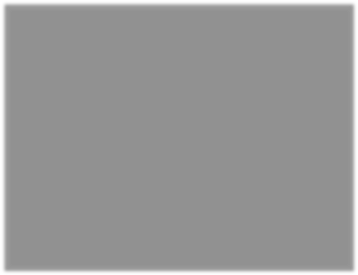
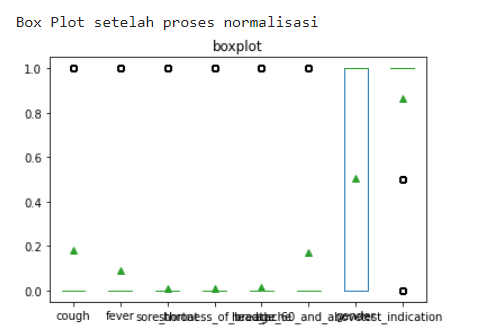
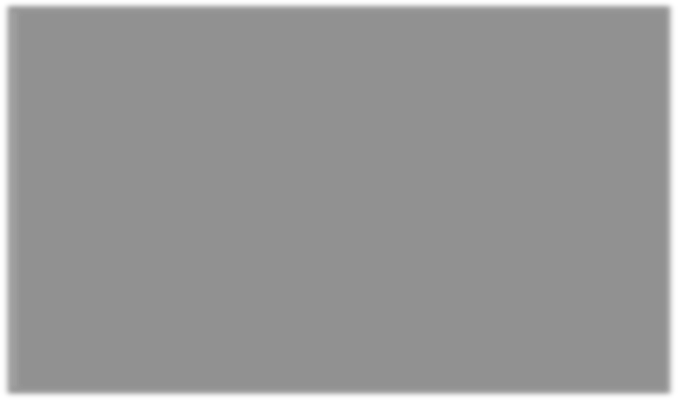
## Logistic Regression

Pengujian menggunakan data gejala covid-19 sebanyak 136.294 record data dengan 8 kolom, dengan *split data* yang ada pada *machine learning* digunakan untuk membagi antara *data training* (data latih) dan *data testing* (data uji). Penulis membagi data menjadi 70% (0.7) untuk *data training* (data latih) dan 30% (0.3) untuk *data testing* (data uji). Kemudian digunakan *Algoritma Logistic Regression* sebagai metode klasifikasi yang digunakan, *Logistic Regressio*n adalah salah satu algoritma klasifiikasi yang paling umum. Pada Gbr. 11 menunjukkan model proses klasifikasi menggunakan algoritma *Logistic Regression.*



Gbr. 11. Klasifikasi Logistic Regression

Hasil akurasi dari algoritma logistic regression dengan dengan pembagian 70:30 data training dan data testing dan telah dilakukan proses validasi data dengan *k-fold cross validation* sebanyak 10 kali, maka didapatkan nilai untuk training set sebesar 94% sedangkan untuk tsting data didapatkan nilai sebesar 94%.

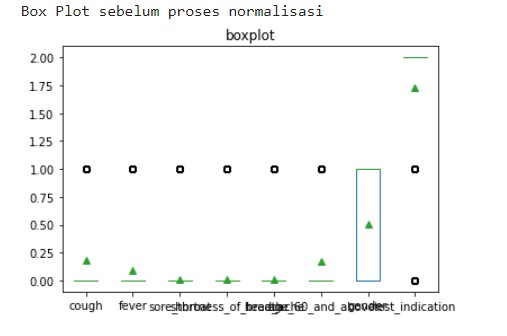
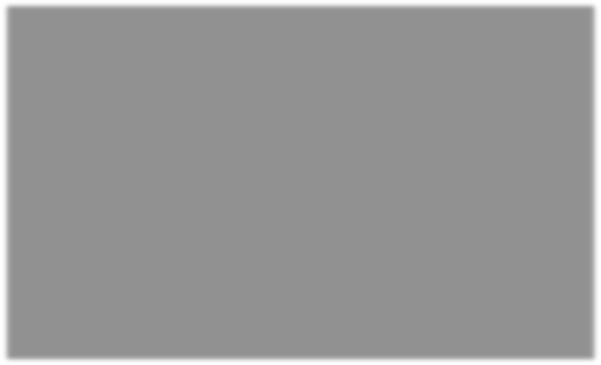


Gbr. 12. Confusion Matrix Logistic Regression

Pada Gbr. 12 dapat dilihat nilai dari confusin matrix data gejalacovid-19 dimana nilai aktual untuk label negative sebanyak 37.642 data dan nilai prediksi sebanyak 59, sedangkan nilai aktual untuk label positif sebanyak 997 data dan nilai prediksi sebanyak 2.191 data.

* 1. Pembahasan Algoritma dengan Normalisasi

Normlisasi data adalah proses membuat beberapa variabel memiliki rentang nilai yang sama, tidak ada yang terlalu besar maupun terlalu kecil sehingga dapat membuat analisis statistik menjadi juh lebih mudah[6]. Normalisasi data pada penelitian ini menggunakan *Min-Max Scaller*, *Min-Max Scaller* data atau menyesuikan data dalam rentang/range nilai minimum hingga nilai maksimum, dengan rentang yang biasa digunakan adalah 0 hingga 1.



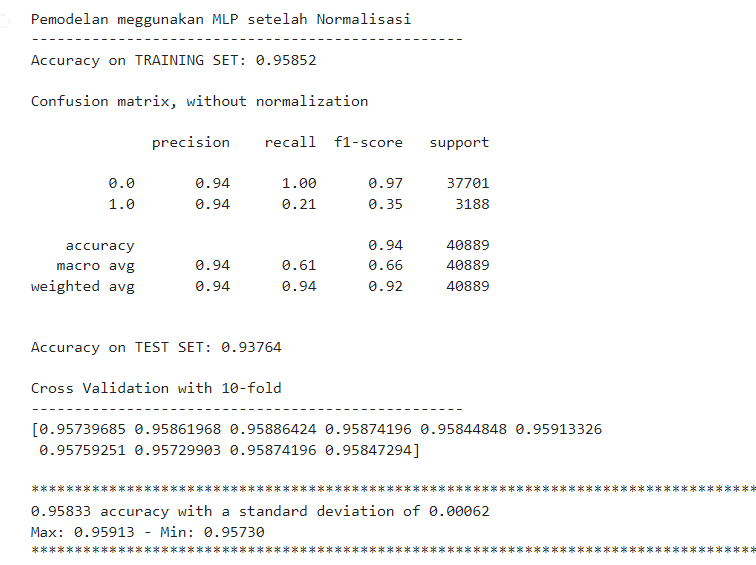
Gbr. 13. Boxplot sebelum proses Normalisasi

Dapat dilihat pada gambar 14 diatas menunjukan rentang nilai yang belum normal pada atribut test indication dimana masih terdapat nilai yang melebihi nilai maksimum yang dimana nilai maksimum pada data tersebut mencapai 2.

Gbr.14. Boxplot setelah dilakukan normalisasi

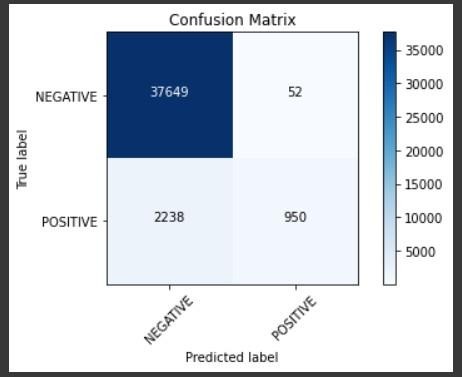
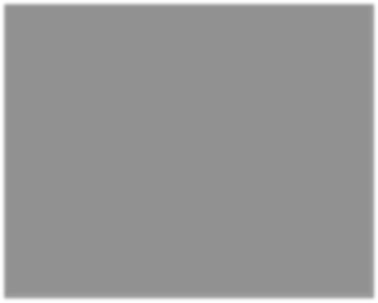
Pada gambar 14 diatas menunjukan data yang telah di normalisasidimana data telah sesuai dengan rentang data minimun dan maksimum yaitu 0 dan 1. Pada gambar diatas dapat dilihat nilai rentang telah normal antara rentang 0 dan 1.

## Pemodelan MLP setelah Normalisasi



Gbr. 15. Model setelah di Normalisasi

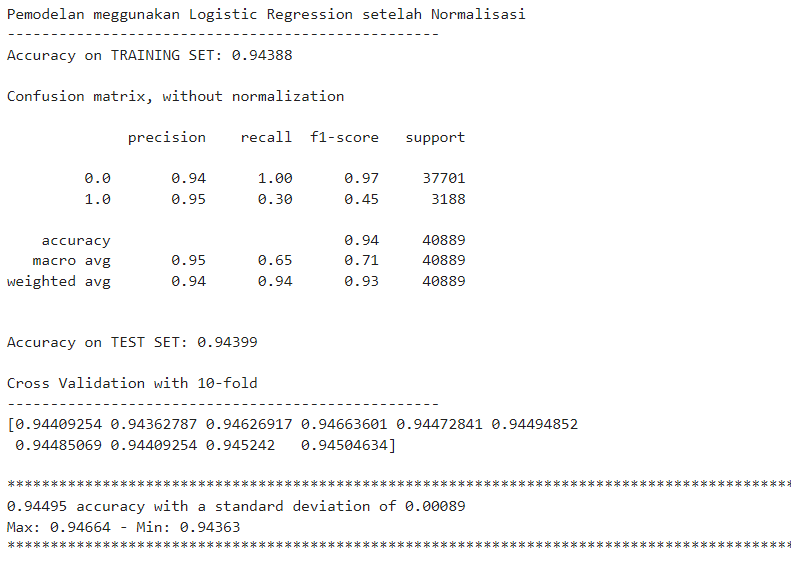
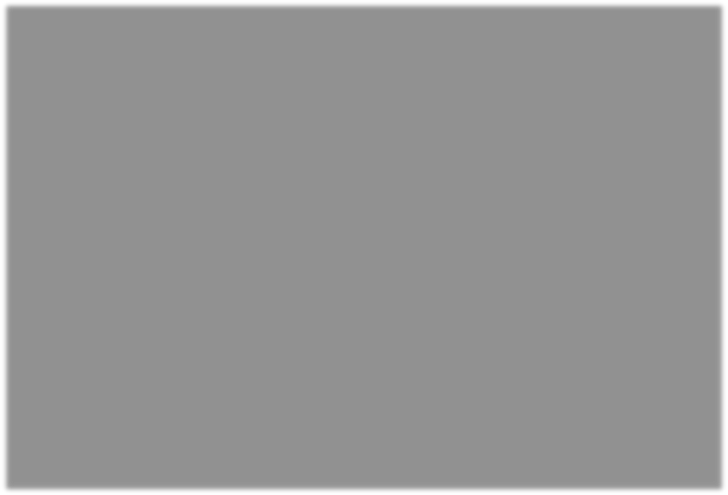
Hasil akurasi setelah dilakukan normalisasi data pada dataset gejala didapatkan nilai akurasi data training sebesar 95.8% dan pada data testing sebesar 93.7% dari algoritma neural network menggunakan model *multi-layer perceptron* (MLP) dan dengan melakukan proses validasi data dengan *k-fold cross validation* sebanyak 10 kali dari data gejala covid-19.



Gbr. 16. Confusion matrix MLP setelah di Normalisasi

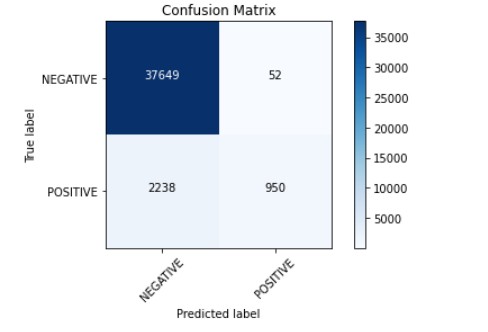
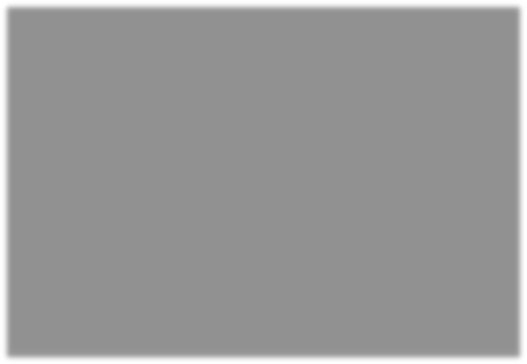
Pada Gbr. 16 dapat dilihat nilai dari confusin matrix yang telah dilakukan normalisasi data pada dataset gejala covid-19 dimana nilai aktual untuk label negative sebanyak 37.656 data dan nilai prediksi sebanyak 45, sedangkan nilai aktual untuk label positif sebanyak 683 data dan nilai prediksi sebanyak 2.505 data.

## Pemodelan Logistic Regression setelah Nirmalisasi



Gbr. 17. Model logistic regression setelah dinormalisasi

Hasil akurasi dari algoritma logistic regression setelah dilakukan normalisasi dan telah dilakukan proses validasi data dengan *k-fold cross validation* sebanyak 10 kali, maka didapatkan nilai untuk training set sebesar 94.3% sedangkan untuk tsting data didapatkan nilai sebesar 94.3%.



Gbr. 18. Confusion matrix logistic regression setelah dinormalisasikan

Pada Gbr. 18 dapat dilihat nilai dari confusin matrix data gejala covid-19 setelah dilakukan normalisasi data dimana nilai aktual untuk label negative sebanyak 37.649 data dan nilai prediksi sebanyak 52, sedangkan nilai aktual untuk label positif sebanyak 950 data dan nilai prediksi sebanyak 2.238 data.

## Hasil Pengujian Sistem

Setelah melakukan pengujian akurasi terhadap gejala penyakit covid-19 pada tools phyton, didepat hasil akurasi perhitungan dengan 2 algoritma yaitu algoritma *neural network* dan *logistic regression* pada table 2.

TABLE II. Hasil akurasi data gejala covid-19

|  |  |  |
| --- | --- | --- |
| **No** | **Algoritma** | **Akurasi (%)** |
| **1** | Neural Network | 95% |
| **2** | Logistic Regression | 94% |

Seperti yang dapat dilihat dari table diatas, dengan menggunakan data gejala covid-19 dimana data latih dan data uji yang berbeda dan dengan menggunakan algoritma dan model yang berbeda memperoleh hasil akurasi yang lebih tinggi yaitu sampai 95%. Dataset Covid- 19 yang diperoleh dari *website* berjumlah 136.294 record data. Dari uji coba yang dilakukan dengan menggunakan algoritma *Neural Network* dan *Logistic Regression* yang digunakan, diperoleh akurasi klasifikasi tertinggi 95%. Hasil akurasi tertinggi diperoleh dengan menggunakan algoritma *Neural Network* dengan menggunakan model *Multi-layer Perceptron*. Algoritma ini dapat menghasilkan akurasi terbaik dikarenakan tidak bergantung dengan model sebaran data sehingga dapat dikatakan tidak mempunyai parameter dan informasi sebelumnya dan dapat menjadi solusi dari data yang bersifat kompleks atau memiliki *noise*.

## Evaluation

Evaluasi adalah fase interpretasi terhadap hasil *data mining*. Evaluasi dilakukan secara mendalam dengan tujuan agar hasil pada tahap modeling sesuasi dengan sasaran yang ingin dicapai dalam tahap *business understanding*[7]*.*

1. Evaluate Result

Tahap ini menilai sejauh mana hasil pemodelan data mining memenuhi tujuan data mining yang ditentukan pada tahap business understanding. Jika diaplikasikan dalam dunia nyata, tahap evaluation sebaiknya melibatkan pihak kesehatan yang kompeten misalnya pihak rumah sakit atau dokter.

1. Review Process

Tahap memeriksa kembali tahapan dari awal untuk memastikan bahwa tidak ada faktor penting dalam proses tersebut yang terabaikan atau terlewatkan.

1. Determine Next Steps

Tahap ini menentukan langkah apa yang diambil. Ada 2 pilihan, kembali ke tahap awal

*(business understanding)* atau melanjutkan ke tahap akhir *(deployment).*

##### Kesimpulan

Coronavirus Disease 19 adalh virus yang berasal dari keluarga SARS- Co-V2, yang memiliki perbedaan dalam virusnya namun memiliki gejala yang sama. Dataset yang sudah diklasifikasi pada Phyton menunjukan pada tingkat negative memiki jumlah data 125.668 record data sedangkan tingkat positif memiliki jumlah data 10.626 record data. Dari klasifikasi yang telah dilakukan menggunakan *tools Phyton* sehingga didapatkan kesimpulan Algoritma *Neural Network (NN)* memiliki hasil yang lebih tinggi dibandingkan Algoritma klasifikasi menggunakan *Logistic Regression* dengan akurasi 95%. Algoritma *Logistic Regression* memiliki hasil yang sedikit lebih rendah dibandingkan Algoritma klsifikasi *Neural Network* dengan akurasi 94%. Berdasarkan penelitian yang sudah disimpulkan, maka peneliti dapat rekomendasi untuk melakukan penelitian yang lebih baik lagi, antara lain diharapkan hasil penelitian ini menjadi masukan bagi pengembangan ilmu komputer khususnya pada Phyton agar dapat digunakan untuk pembelajaran dan edukasi. Selain hasil penelitian ini dapat dijadikan salah satu acuan maupun litelature dan diharapkan untuk peneliti selanjutnya dapat mengembangkan penelitian ini lebih lanjut dengan metode dan rancangan yang berbeda, cakupan responden yang lebih luas, penelitian dengan penambahan kelompok berbeda yang dapat menghasilkan lebih variatif dan mendatangkan informasi yang lebih tinggi.

##### Daftar pustaka

[1] L. Abdillah, “Stigma Terhadap Orang Positif COVID-19 (Stigma on Positive People COVID-19),” *Pandemik COVID-19 Antara Persoalan Dan Refleks. Di Indones. Forthcom.*, 2020.

[2] B. Rifai, “Algoritma Neural Network Untuk Prediksi Penyakit Jantung,” *Techno Nusa Mandiri J. Comput. Inf. Technol.*, vol. 10, no. 1, pp. 1–9, 2013.

[3] A. P. Fadillah, “Penerapan Metode CRISP-DM untuk Prediksi Kelulusan Studi Mahasiswa Menempuh Mata Kuliah (Studi Kasus Universitas XYZ),” *J. Tek. Inform. Dan Sist. Inf.*, vol. 1, no. 3, 2015.

[4] M. A. Hasanah, S. Soim, and A. S. Handayani, “Implementasi CRISP-DM Model Menggunakan Metode Decision Tree dengan Algoritma CART untuk Prediksi Curah Hujan Berpotensi Banjir,” 2021. [Online]. Available: http://jurnal.polibatam.ac.id/index.php/JAIC.

[5] A. C. Siregar and B. C. Octariadi, “Perbandingan Metode Jaringan Syaraf Tiruan Pada Klasifikasi Motif Kain Tenun Sambas,” *Cybernetics*, vol. 4, no. 02, 2021, doi: 10.29406/cbn.v4i02.2489.

[6] H. Junaedi, H. Budianto, I. Maryati, and Y. Melani, “Data Transformation pada Data Mining,” *Pros. Konf. Nas. Inov. dalam Desain dan Teknol.*, vol. 7, pp. 93–99, 2011.

[7] A. P. Windarto, M. R. Lubis, and S. Solikhun, “Model Arsitektur Neural Network Dengan Backpropogation Pada Prediksi Total Laba Rugi Komprehensif Bank Umum Konvensional,” *Klik-Kumpulan J. Ilmu Komput.*, vol. 5, no. 2, pp. 147–158, 2018.