

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

Rahmi^[1], Darius Antoni^{*[2]}, Hadi Syaputra^[3], Fatonj^[4], Tri Basuki Kurniawan^[5]
^{[1],[3],[4],[5]} Magister Teknik Informatika, Fakultas Ilmu Komputer Universitas Bina Darma Palembang
^{*[2]} Sistem Informasi, Fakultas Ilmu Komputer Universitas Indo Global Mandiri Palembang
E-mail : puspitarahmi9@gmail.com^[1], darius.antoni@uigm.ac.id^[2], hadisyaputra@binadarma.ac.id^[3],
fatoni@binadarma.ac.id^[4], tribasukikurniawan@binadarma.ac.id^[5]

Abstrack— *Coronavirus Disease 19 (COVID-19) is a new virus that can cause respiratory infections. This virus comes from animals that can be transmitted to humans through splashes of their saliva. According to epidemiological data, the average age of sufferers of this virus is 15-80 years. This virus has an incubation period of 3-14 days which has initial symptoms of high fever, shortness of breath, cough and runny nose. Indonesia had the first 2 cases on March 2 2020, Covid-19 increased regularly on December 29 2020, data shows 719,219 thousand people are confirmed to have contracted Covid-19. The problem raised in this study is how to classify the risk of contracting the Covid-19 virus from the symptoms it causes. The purpose of this study was to determine the accuracy value of the risk classification of contracting the Covid-19 virus based on the instrument used from the Cross Industry Standard Process for Data Mining (CRISP-DM) method. The dataset used by researchers was taken from the website <http://github.com/nshomron/covidpred>. This study used the Neural Network (NN) Algorithm with the help of the Python tool, the accuracy of the Neural Network Algorithm (NN) obtained a value of 95%, meaning that it has shown good classification results. Researchers also tested the Logistic Regression Algorithm but the accuracy value obtained was not much different from the NN Algorithm, the Logistic Regression Algorithm obtained an accuracy value of 94%.*

Keywords: *Coronavirus Disease 19, Covid-19, Neural Network, Logistic Regression, classification, CRISP-DM*

Abstrak— *Coronavirus Disease 19 (COVID-19) adalah virus baru yang dapat menyebabkan infeksi saluran pernafasan. Virus ini berasal dari hewan yang dapat menular ke manusia melalui percikan ludahnya. Menurut data epidemiologis, rata-rata penderita virus ini berusia 15-80 tahun. Virus ini memiliki masa inkubasi 3-14 hari yang memiliki gejala awal yaitu demam tinggi, sesak napas, batuk dan pilek. Indonesia memiliki 2 kasus pertama pada 2 Maret 2020, Covid-19 meningkat secara teratur pada 29 Desember 2020 data menunjukkan 719.219 ribu orang*

dipastikan terjangkit Covid-19. Masalah yang diangkat dalam penelitian ini adalah bagaimana mengklasifikasikan risiko tertular virus Covid-19 dari gejala yang ditimbulkan. Tujuan dari penelitian ini adalah untuk mengetahui nilai akurasi dari klasifikasi risiko tertular virus Covid-19 berdasarkan instrument yang digunakan dari metode *Cross Industry Standard Process for Data Mining (CRISP-DM)*. Dataset yang digunakan peneliti diambil dari website <http://github.com/nshomron/covidpred>. Penelitian ini menggunakan Algoritma Neural Network (NN) dengan bantuan alat Python, akurasi Algoritma Neural Network (NN) diperoleh nilai sebesar 95%, artinya telah menunjukkan hasil klasifikasi yang baik. Peneliti juga menguji dengan Algoritma Logistic Regression namun nilai akurasi yang diperoleh tidak jauh berbeda dengan Algoritma NN, Algoritma Logistic Regression diperoleh akurasi nilai sebesar 94%.

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

I. 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[1] 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[2], tetapi memiliki gejala yang sama[1]. 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%.

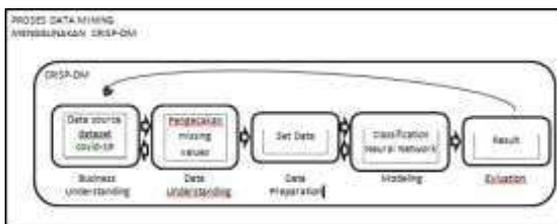
Ada 4 (empat) kelompok dalam penyebutan Covid-19, dalam pedoman WHO dan Stigma Orang Positif Covid-19 yaitu Orang Dalam Pantauan (ODP), Pasien Dalam Pengawasan (PDP), Orang Tanpa Gejala (OTG), sedangkan Positif adalah pasien yang telah melakukan test menunjukkan hasil positif covid-19. Kasus di Indonesia 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 Neural Network dan Logistic Regression dengan menggunakan *software* phyton untuk mempermudah dalam menganalisis.

II. 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 Kementerian Kesehatan Israel. Dengan kolom gejala 0 (not present) 1 (present), untuk kolom resiko 2 kelas yaitu positif dan negatif dengan menggunakan tools Phyton. Dataset yang digunakan sebagai bahan penelitian klasifikasi untuk mengetahui risiko seseorang terjangkit covid-19 diambil dari website <http://github.com/nshomron/covidpred>

A. Alur Penelitian

Alur dari penelitian yang dilakukan pada penelitian ini dapat dilihat seperti pada gambar 1, berikut ini.



Gambar 1. Alur Penelitian

B. 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 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 tingkat resiko rendah, tingkat resiko sedang dan tingkat resiko tinggi.

2) 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).

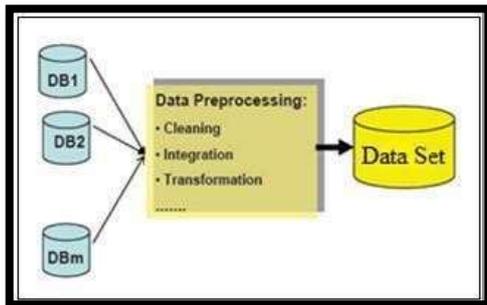
Dataset yang akan digunakan yaitu diambil dari situs resmi kaggle. Ada 17 atribut yang akan digunakan untuk melaksanakan klasifikasi, yaitu suhu tubuh, batuk kering, sakit tenggorokan, kelelahan, masalah pernapasan, kantuk, nyeri dada, riwayat perjalanan ke Negara terinfeksi, diabetes, penyakit jantung,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

Gambar 2. Data Gejala Covid-19

3	Sakit Tenggorokan
4	Sesak napas
5	Sakit kepala
6	Kontak langsung dengan seseorang yang dikonfirmasi Covid-19

3) Data Preparation

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.



Gambar 3. Ilustrasi Proses Data Preparation

a) 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 understanding terutama pada perumusan tujuan data mining.

TABEL 1. DATASET COVID-19

No	Gejala Penyakit Covid-19
1	Batuk
2	Demam

a) Clean Data

Tahap ini memastikan data yang dipilih, dikonstruksi dan diintegrasikan telah layak untuk ditambah. 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 127 data tidak terdapat missing values, dapat dilihat pada gambar 4:

```

%%Time
# Drop data with None information
data = data.replace("None", np.nan)
data = data.replace("other", np.nan)

# Drop all row with columns value has Nan
data = data.dropna()

print("dimension of covid19 data: {}".format(data.shape))
print()
display(data)
print()
print('Drop rows:', original_rows - len(data))
print()

dimension of covid19 data: (136294, 9)
    
```

Gambar 4. Cleaning Data

Gambar 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.

C. Convert Data

Dalam melakukan pemodelan menggunakan Multi-layer Perceptron (MLP) data yang digunakan untuk melakukan pemodelan yaitu berupa data numeric, sehingga data yang masih berupa categorical diubah menjadi data numerical

seperti gambar 6:

Gambar 6. Data setelah dilakukan ubah data categorical menjadi numerical

Gambar 8. Informasi jumlah data training

D. 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 untuk 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 gambar 7:

Gambar 9. Informasi jumlah data testing

```

In [10]:
from sklearn.model_selection import train_test_split

X = data.drop('corona_result', axis=1)
y = data['corona_result']

X_train, X_test, y_train, y_test = train_test_split(
    X,
    stratify=data['corona_result'],
    test_size=0.3,
    random_state=0)

print(f'Informasi training set: {len(X_train)}, {len(y_train)}')
print(f'Informasi testing set: {len(X_test)}, {len(y_test)}')

print(f'Informasi training set: {len(X_train)}, {len(y_train)}')
print(f'Informasi testing set: {len(X_test)}, {len(y_test)}')

display(X_train)
display(X_test)
print()

Out[10]:
Informasi training set: 95405 baris, 8 kolom
Informasi testing set: 40889 baris, 8 kolom
    
```

Gambar 7. Split data Train dan Data Test

Dapat dilihat pada gambar 8 ditunjukkan bahwa untuk 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 menggunakan algoritma neural network dan logistic regression

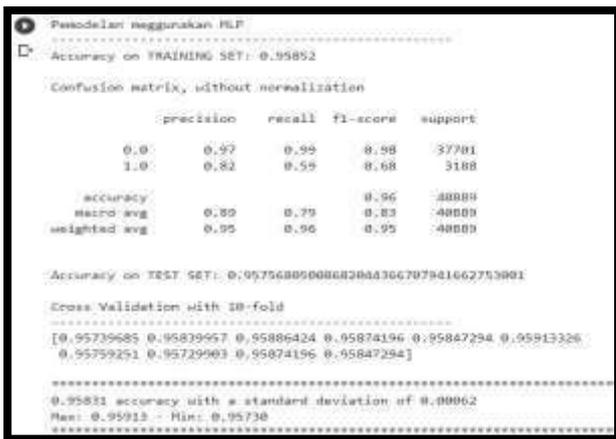
III. HASIL DAN PEMBAHASAN

Setelah melakukan analisis terhadap dataset gejala Covid-19 dengan 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 Python 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 akurasi dan 95.7% untuk test set akurasi sedangkan neural network setelah dilakukan normalisasi menggunakan normalisasi dengan *MinMaxScaler* 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 *MinMaxScaler* didapatkan 94.38% untuk training set akurasi dan 94.39% untuk test set akurasi.

A. Neural Network (NN)

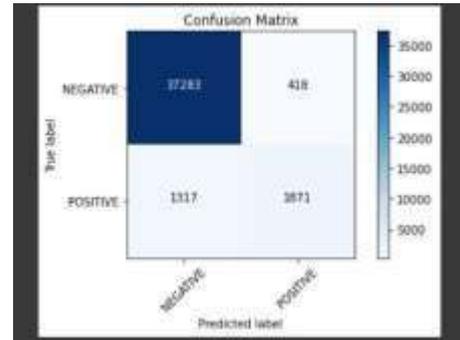
Pengujian menggunakan data gejala covid-19 sebanyak 136.294 record data dengan 6 kolom, dengan *split data* yang ada pada *machine learning* digunakan untuk membagi antara *data training* (data latih) dan *data testing* (data uji). Penulis membagi data menjadi 70% (0.7) untuk *data training* (data latih) dan 30% (0.3) untuk *data testing* (data uji). Kemudian digunakan *Algoritma neural network* sebagai metode klasifikasi yang 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. Pada gambar 10 menunjukkan pemodelan 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 gambar 10:



Gambar 10. 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.

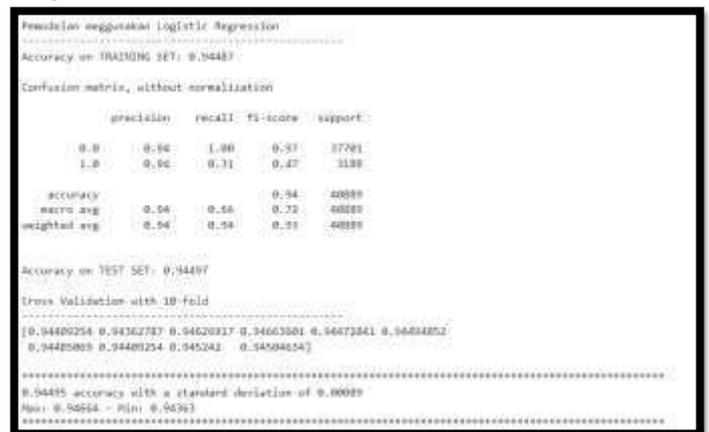


Gambar 11. Confusion Matrix MLP

Pada gambar 11 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.

B. Logistic Regression

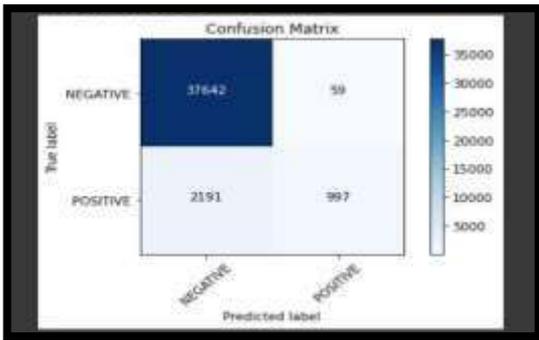
Pengujian menggunakan data gejala covid-19 sebanyak 136.294 record data dengan 8 kolom, dengan *split data* yang ada pada *machine learning* digunakan untuk membagi antara *data training* (data latih) dan *data testing* (data uji). Penulis membagi data menjadi 70% (0.7) untuk *data training* (data latih) dan 30% (0.3) untuk *data testing* (data uji). Kemudian digunakan *Algoritma Logistic Regression* sebagai metode klasifikasi yang digunakan, *Logistic Regression* adalah salah satu algoritma klasifikasi yang paling umum. Pada gambar 12 menunjukkan model proses klasifikasi menggunakan algoritma *Logistic Regression*.



Gambar 12. 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%.

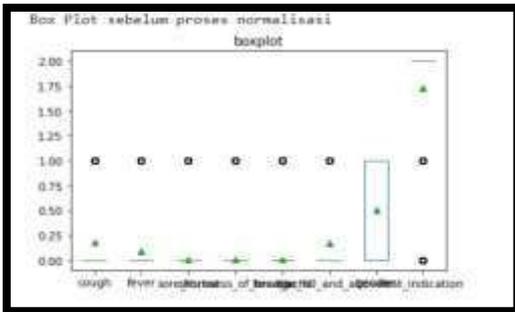


Gambar 13. Confusion Matrix Logistic Regression

Pada gambar 13 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.

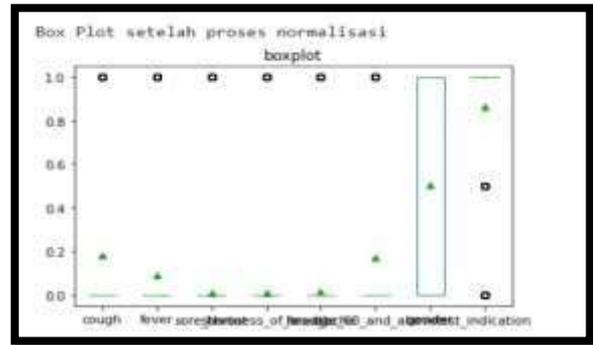
C. Pembahasan Algoritma dengan Normalisasi

Normalisasi 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 jauh lebih mudah. Normalisasi data pada penelitian ini menggunakan *Min-Max Scaller*, *Min-Max Scaller* data atau menyesuaikan data dalam rentang/range nilai minimum hingga nilai maksimum, dengan rentang yang biasa digunakan adalah 0 hingga 1.



Gambar 14. Boxplot sebelum proses Normalisasi

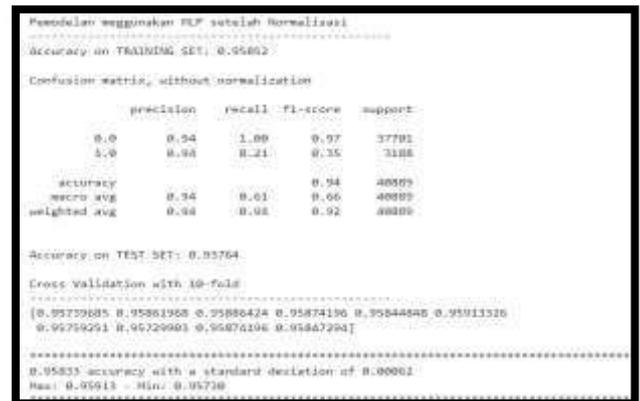
Dapat dilihat pada gambar 14 diatas menunjukkan rentang nilai yang belum normal pada atribut test indication dimana masih terdapat nilai yang melebihi nilai maksimum yang dimana nilai maksimum pada data tersebut mencapai 2.



Gambar 15. Boxplot setelah dilakukan normalisasi

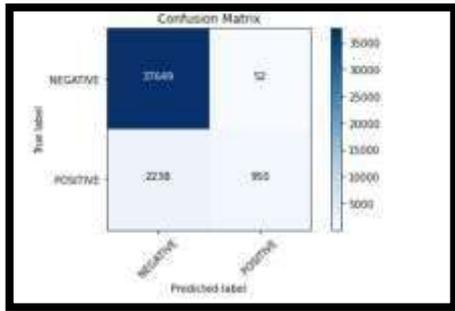
Pada gambar 15 diatas menunjukkan data yang telah dinormalisasidimana data telah sesuai dengan rentang data minimum dan maksimum yaitu 0 dan 1. Pada gambar diatas dapat dilihat nilai rentang telah normal antara rentang 0 dan 1.

D. Pemodelan MLP setelah Normalisasi



Gambar 16. Model setelah di Normalisasi

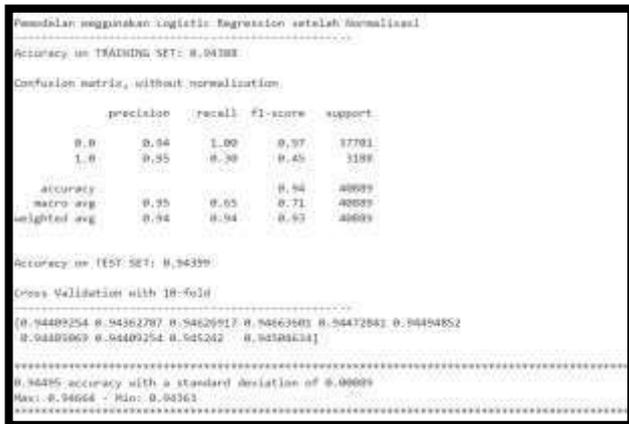
Hasil akurasi setelah dilakukan normalisasi data pada dataset gejala didapatkan nilai akurasi data training sebesar 95.8% dan pada data testing sebesar 93.7% dari algoritma neural network menggunakan model *multi-layer perceptron* (MLP) dan dengan melakukan proses validasi data dengan *k-fold cross validation*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



Gambar 17. Confusion matrix MLP setelah di Normalisasi

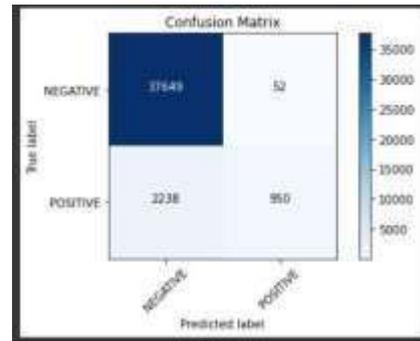
Pada gambar 17 dapat dilihat nilai dari confusion matrix yang telah dilakukan normalisasi data pada dataset gejala covid-19 dimana nilai aktual untuk label negative sebanyak 37.656 data dan nilai prediksi sebanyak 45, sedangkan nilai aktual untuk label positif sebanyak 683 data dan nilai prediksi sebanyak 2.505 data.

E. Pemodelan Logistic Regression setelah Nirmalisasi



Gambar 18. 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 testing data didapatkan nilai sebesar 94.3%.



Gambar 19. Confusion matrix logistic regression setelah dinormalisasikan

Pada gambar 19 dapat dilihat nilai dari confusion 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.

F. Hasil Pengujian Sistem

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

TABEL 2. 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*.

G. Evaluation

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

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.

2) Review Process

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

3) 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*).

IV. KESIMPULAN

Coronavirus Disease 19 adalah virus yang berasal dari keluarga SARS- Co-V2, yang memiliki perbedaan dalam virusnya namun memiliki gejala yang sama. Dataset yang sudah diklasifikasi pada Phyton menunjukkan pada tingkat negative memiliki 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 klasifikasi *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 literatur 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] Abdillah, G., Putra, F. A., Renaldi, F., Informatika, P. S., Jenderal, U., & Yani, A. (2016). Penerapan Data Mining Pemakaian Air Pelanggan Untuk Menentukan Klasifikasi Potensi Pemakaian Air Pelanggan Baru Di Pdam Tirta Raharja Menggunakan Algoritma K-Means. Seminar Nasional Teknologi Informasi Dan Komunikasi, 2016, 18–19.
- [2] Abdillah, L. (2020). Stigma Terhadap Orang Positif COVID-19 (Stigma on Positive People COVID-19). Pandemi COVID-19: Antara Persoalan Dan Refleksi Di Indonesia, Forthcoming.
- [3] Ahmad, A. (2017). Mengenal artificial intelligence, machine learning, neural network, dan deep learning. J. Teknol. Indones, 3.
- [4] Ayumi, V., & Nurhaida, I. (2021). Klasifikasi Chest X-Ray Images Berdasarkan Kriteria Gejala Covid-19 Menggunakan Convolutional Neural Network. JSAI (Journal Scientific and Applied Informatics), 4(2), 147–153.
- [5] Behrens, J. T. (1997). Principles and procedures of exploratory data analysis. Psychological Methods, 2(2), 131. Fadillah, A. P. (2015). Penerapan Metode CRISP-DM untuk Prediksi Kelulusan Studi Mahasiswa Menempuh Mata Kuliah (Studi Kasus Universitas XYZ). Jurnal Teknik Informatika Dan Sistem Informasi, 1(3).
- [6] Rifai, B. (2013). Algoritma Neural Network Untuk Prediksi Penyakit Jantung. Techno Nusa Mandiri: Journal of Computing and Information Technology, 10(1), 1–9.
- [7] Saifudin, A. (2018). Metode Data Mining untuk Seleksi Calon Mahasiswa pada Penerimaan Mahasiswa Baru di Universitas Pamulang. Jurnal Teknologi, 10(1), 25–36.
- [8] Saputro, W. T., & Jumasa, H. M. (2018). Memprediksi Daftar Ulang Mahasiswa Baru Menggunakan Algoritma Bayesian Classification Di Universitas XYZ. INTEK: Jurnal Informatika Dan Teknologi Informasi, 1(2), 73–82.
- [9] Sutoyo, I. (2018). Implementasi Algoritma Decision Tree Untuk Klasifikasi Data Peserta Didik. Pilar Nusa Mandiri: Journal of Computing and Information System, 14(2), 217–224.
- [10] Windarto, A. P., Lubis, M. R., & Solikhun, S. (2018). Model Arsitektur Neural Network Dengan Backpropogation Pada Prediksi Total Laba Rugi Komprehensif Bank Umum Konvensional. Klik-Kumpulan Jurnal Ilmu Komputer, 5(2), 147–158.
- [11] Yuliana, D., Purwanto, P., & Supriyanto, C. (2018). Klasifikasi Teks Pengaduan Masyarakat Dengan Menggunakan Algoritma Neural Network. Jurnal KomTekInfo, 5(3), 92–116.