



**ANALISIS TINGKAT AKURASI PREDIKSI GEJALA *COVID-19*
DENGAN MENGGUNAKAN METODE *LOGISTIC REGRESSION*
DAN *SUPPORT VECTOR MACHINE***

LAPORAN PENELITIAN

BRIANDY TRI PUTRA

191410085

PROGRAM STUDI SISTEM INFORMASI

FAKULTAS SAINS TEKNOLOGI

UNIVERSITAS BINA DARMA

PALEMBANG

2023



**ANALISIS TINGKAT AKURASI PREDIKSI GEJALA *COVID-19*
DENGAN MENGGUNAKAN METODE *LOGISTIC REGRESSION*
DAN *SUPPORT VECTOR MACHINE***

BRIANDY TRI PUTRA

191410085

**Laporan Penelitian ini diajukan sebagai syarat memperoleh gelar
Sarjana Komputer**

PROGRAM STUDI SISTEM INFORMASI

FAKULTAS SAINS TEKNOLOGI

UNIVERSITAS BINA DARMA

PALEMBANG

2023

HALAMAN PENGESAHAN

Analisis Tingkat Akurasi Prediksi Gejala COVID-19 Menggunakan Metode Logistic Regression Dan Support Vector Machine

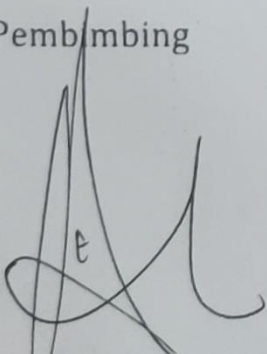
BRIANDY TRI PUTRA

191410085

Telah diterima sebagai salah satu syarat untuk memperoleh gelar
Sarjana Komputer pada Program Studi Sistem Informasi

Palembang, 09 Agustus 2023
Fakultas Sains Teknologi
Universitas Bina Darma
Dekan,

Pembimbing



Evi Yulianingsih, M.M., M.Kom.



Universitas Bina Darma
Fakultas Sains Teknologi

Dr. Tata Sutabri, S.Kom., MMSI., MKM.

HALAMAN PERSETUJUAN

Skripsi Berjudul "Analisis Tingkat Akurasi Prediksi Gejala COVID-19 Menggunakan Metode Logistic Regression Dan Support Vector Machine" Oleh "Briandy Tri Putra", telah dipertahankan di depan komisi penguji pada hari Rabu tanggal 09 Agustus 2023.

Komisi Penguji

1. Ketua : Evi Yulianingsih, M.M., M.Kom.

(.....)

2. Anggota : Fatmasari, M.M., M.Kom

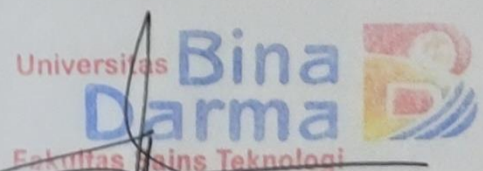
(.....)

3. Anggota : Ferdiansyah, M.Kom., CEH

(.....)

Mengetahui,

Program Studi Sistem Informasi
Fakultas Sains Teknologi
Universitas Bina Darma
Ketua,



Nita Rosa Damayanti, M.Kom., Ph.D.

SURAT PERNYATAAN

Saya yang bertanda tangan di bawah ini :

Nama : Briandy Tri Putra

NIM : 191410085

Dengan ini menyatakan bahwa :

1. Karya tulis saya adalah asli dan belum pernah diajukan untuk mendapatkan gelar akademik (Sarjana) di Universitas Bina Darma atau perguruan tinggi lainnya.
2. Karya tulis ini murni gagasan, rumusan dan penelitian saya dengan arahan dari tim pembimbing;
3. Di dalam karya tulis ini tidak terdapat karya atau pendapat yang telah ditulis atau di publikasikan orang lain, kecuali secara tertulis dengan jelas dikutip dengan mencantumkan nama pengarang dan memasukkan ke dalam daftar rujukan;
4. Saya bersedia karya tulis ini di cek keasliannya menggunakan plagiarism checker serta di unggah ke internet, sehingga dapat diakses secara daring;
5. Surat pernyataan ini saya tulis dengan sungguh-sungguh dan apabila terbukti melakukan penyimpangan atau ketidakbenaran dalam pernyataan ini maka saya bersedia menerima sanksi dengan peraturan dan perundang-undangan yang berlaku;

Demikian surat pernyataan ini saya buat agar dapat dipergunakan sebagaimana mestinya.

Palembang, 09 Agustus 2023

uat pernyataan,



Briandy Tri Putra
NIM. 191410085

ABSTRAK

Kecerdasan buatan merupakan salah satu teknologi ilmu komputer yang diprogram untuk mempelajari dan melakukan aktivitas seperti manusia. Teknologi kecerdasan buatan telah dipakai pada beberapa bidang, salah satunya di bidang kesehatan. Dibidang kesehatan, kecerdasan buatan digunakan sebagai alat untuk mendeteksi penyakit pada manusia, salah satu contohnya adalah memprediksi gejala awal *COVID-19* merupakan salah satu penyakit menular *SARS-CoV2* yang menyebabkan pandemi di seluruh dunia, dan virus tersebut terdeteksi pertama kali dari hewan-hewan liar di pasar Kota Wuhan, China pada akhir 2019. Pada penelitian sebelumnya yang berjudul Metode Klasifikasi Gejala Penyakit *Coronavirus Disease 19 (COVID-19)* Menggunakan Algoritma *Neural Network*" oleh Rahmi, dkk, menggunakan data gejala-gejala *COVID-19* untuk mendapatkan tingkat akurasi dalam prediksi *COVID-19* dengan menggunakan metode *Neural Network* dan *Logistic Regression*. Hasil penelitian tersebut mendapatkan tingkat akurasi sebesar 95% dengan metode *Neural Network*, dan 94% dengan metode *Logistic Regression*. Pada penelitian ini, penulis ingin membandingkan metode *Logistic Regression* dengan *Support Vector Machine* dalam memprediksi gejala awal *COVID-19*. Hasil dari penelitian ini adalah mendapatkan akurasi dengan tingkat yang tertinggi dari kedua metode tersebut.

Kata kunci: *Covid19, Logistic Regression, SVM*



ABSTRACT

Artificial intelligence is a computer science technology that is programmed to learn and perform human-like activities. Artificial intelligence technology has been used in several fields, one of which is in the health sector. In the health sector, artificial intelligence is used as a tool to detect diseases in humans, one example is predicting the early symptoms of COVID-19 is one of the SARS-CoV2 infectious diseases that caused a worldwide pandemic, and the virus was first detected from wild animals in the Wuhan City market, China in late 2019. In previous research entitled "Classification Method of Coronavirus Disease 19 (COVID-19) Symptoms Using Neural Network Algorithm" by Rahmi, et al, using data on COVID-19 symptoms to get an accuracy rate in predicting COVID-19 using the Neural Network and Logistic Regression methods. The results of this study obtained an accuracy rate of 95% with the Neural Network method, and 94% with the Logistic Regression method. In this study, the authors want to compare the Logistic Regression method with Support Vector Machine in predicting the early symptoms of COVID-19. The result of this research is to get the highest level of accuracy of the two methods.

Keywords: Covid19, Logistic Regression, SVM

KATA PENGANTAR

Puji dan syukur kepada Tuhan Yang Maha Esa, karena berkat dan rahmat yang diberikan oleh-Nya, sehingga saya telah menyelesaikan laporan penelitian ini. Laporan penelitian ini dilakukan sebagai syarat untuk memperoleh gelar Sarjana Komputer pada Program Studi Sistem Informasi. Dalam penyelesaian penyusunan laporan penelitian ini, penulis ingin mengucapkan banyak terima kasih kepada berbagai pihak yang telah membantu:

1. Ibu Dr. Sunda Ariana M.Pd., M.M., Selaku Rektor Universitas Bina Darma.
2. Bapak Dr. Tata Sutabri, S.Kom., MMSI., MKM., Selaku Dekan Fakultas Ilmu Komputer Universitas Bina Darma.
3. Ibu Nita Rosa Damayanti, M.Kom., Ph.D., Selaku Ketua Program Studi Sistem Informasi Universitas Bina Darma.
4. Ibu Evi Yulianingsih, M.M., M.Kom., Selaku Dosen Pembimbing Penelitian
5. Coach Deni Suprihadi, S.T., M.Kom., Selaku Pembimbing di Synthesis Academy.
6. Ibu Fatmasari, M.M., M.Kom., Selaku Dosen Penguji Penelitian.
7. Bapak Ferdiansyah, M.Kom., Ph.D., Selaku Dosen Penguji Penelitian.
8. Orang tua dan keluarga yang telah memberikan do'a dan dukungan secara moral maupun materi.
9. Teman seperjuangan Program Studi Sistem Informasi Angkatan 2019 yang telah memberikan dukungannya.

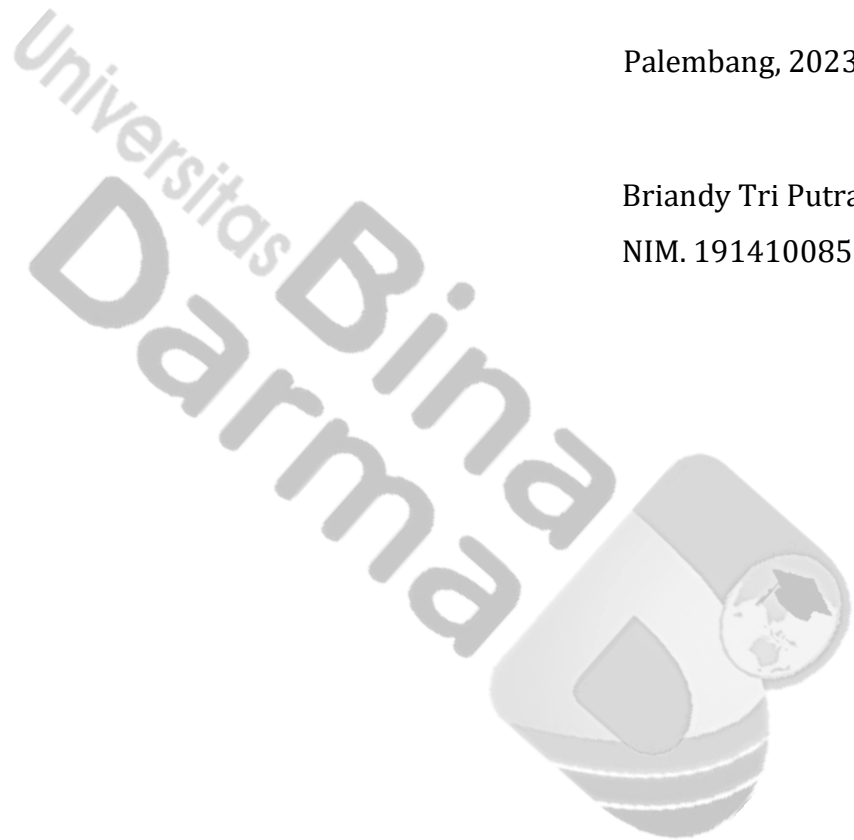
Penulis menyadari bahwa masih ada kekurangan dalam penulisan hasil laporan penelitian ini, baik dalam berbahasa hingga cara pemaparannya. Oleh sebab itu, penulis berharap saran dan tanggapannya dari berbagai pihak demi kesempurnaan laporan penelitian ini.

Semoga laporan penelitian ini bermanfaat kepada berbagai pihak yang membacanya, khususnya untuk mahasiswa Program Studi Sistem Informasi.

Palembang, 2023

Briandy Tri Putra

NIM. 191410085



DAFTAR ISI

HALAMAN PENGESAHAN	i
HALAMAN PERSETUJUAN	ii
SURAT PERNYATAAN	iii
ABSTRAK	iv
KATA PENGANTAR	vi
DAFTAR ISI	viii
DAFTAR GAMBAR	xi
DAFTAR TABEL	xiii
BAB I PENDAHULUAN	1
1.1. Latar Belakang	1
1.2. Perumusan Masalah	3
1.2.1. Identifikasi Masalah	3
1.2.2. Rumusan Masalah	4
1.3. Tujuan Penelitian.....	4
1.4. Perumusan Masalah	4
1.5. Manfaat Penelitian	4
BAB II KAJIAN TEORI	6
2.1. Kecerdasan Buatan	6
2.2. <i>Dataset</i>	6
2.3. <i>Israeli Minsitry of Health</i>	6
2.4. <i>Google Colaboratory</i>	7
2.5. <i>Scikit-Learn</i>	7
2.6. <i>Confusion Matrix</i>	7
2.7. <i>Logistic Regression</i>	8
2.8. <i>Support Vector Machine</i>	9
2.9. Penelitian Terdahulu	10
BAB III METODOLOGI PENELITIAN	12

3.1. Metode Penelitian.....	12
3.1.1. <i>Logistic Regression</i>	12
3.1.2. <i>Support Vector Machine</i>	13
3.1.3. <i>Dataset</i>	14
3.2. Tahapan Penelitian	15
1. <i>Import Data</i>	15
1.1. <i>Isi Dataset</i>	16
1.2. Tipe Data Pada Kolom	17
1.3. Total Data Pasien Pada Setiap Kolom	18
1.4. Data Kosong dan Duplikasi Data Dalam <i>Dataset</i> Pasien Berdasarkan Gejala <i>COVID-19</i>	20
2. <i>Data Preprocessing</i>	21
2.1. <i>Data Cleaning</i>	21
2.2. <i>Labeling Data String Menjadi Integer</i>	24
2.3. <i>Splitting Data</i>	26
2.4. Standarisasi Fitur	27
3. <i>Modeling</i>	27
3.1. <i>Logistic Regression</i>	27
3.2. <i>Support Vector Macine</i>	28
4. <i>Model Evaluation</i>	30
3.3. Metode Pengumpulan Data.....	31
1. Studi Literatur	31
2. <i>Dataset</i>	31
BAB IV HASIL DAN PEMBAHASAN.....	32
4.1. Hasil.....	32
1. Hasil <i>Dataset</i>	32
2. Hasil Tingkat Akurasi Pada Penelitian Terdahulu	33
3. Hasil Tingkat Akurasi Dengan Menggunakan Metode <i>Logistic</i> <i>Regression</i>	35
4. Hasil Menggunakan Metode <i>Support Vector Machine</i>	37

5.	Hasil Menggunakan Metode <i>Support Vector Machine</i> Dengan <i>Kernel Linear</i>	37
6.	Hasil Menggunakan Metode <i>Support Vector Machine</i> Dengan <i>Kernel Sigmoid</i>	39
7.	Hasil Menggunakan Metode <i>Support Vector Machine</i> Dengan <i>Kernel RBF</i>	41
4.2.	Pembahasan	43
1.	Hasil <i>Dataset</i>	39
2.	Tingkat Akurasi Dengan Metode <i>Logistic Regression</i>	44
3.	Tingkat Akurasi Dengan Metode <i>Support Vector Machine</i> <i>Kernel Linear</i>	46
4.	Tingkat Akurasi Dengan Metode <i>Support Vector Machine</i> <i>Kernel Sigmoid</i>	49
5.	Tingkat Akurasi Dengan Metode <i>Support Vector Machine</i> <i>Kernel RBF</i>	51
BAB V KESIMPULAN DAN SARAN		56
5.1.	Kesimpulan	56
5.2.	Saran	57

DAFTAR GAMBAR

Gambar 1. <i>Confusion Matrix</i>	8
Gambar 2. Grafik <i>Support Vector Machine</i>	13
Gambar 3. Tahapan Penelitian.....	15
Gambar 4. <i>Dataset</i> Pasien Berdasarkan Gejala-Gejala <i>COVID-19</i>	16
Gambar 5. Tipe Data Pada <i>Dataset</i> Pasien Berdasarkan Gejala- Gejala <i>COVID-19</i>	17
Gambar 6. Total Pasien Pada Kolom <i>Cough, Fever, Sore Throat,</i> <i>Shortness Of Breath,</i> dan <i>Head Ache</i>	18
Gambar 7. Total Pasien Pada Kolom <i>Corona Result, Age 60 And</i> <i>Above, Gender,</i> dan <i>Test Indication</i>	19
Gambar 8. Data Kosong dan Duplikasi Data.....	21
Gambar 9. <i>Code</i> Penghapusan Kolom <i>test_date</i>	22
Gambar 10. <i>Dataset</i> Setelah Dilakukan Penghapusan Kolom <i>test_date</i>	22
Gambar 11. <i>Code</i> Penghapusan Data “ <i>other</i> ” dan “ <i>None</i> ”.....	23
Gambar 12. Total Data Setelah Dilakukan Penghapusan Data Kosong.....	24
Gambar 13. <i>Code</i> Labelling Kolom <i>Age 60 and Above, Gender, Test</i> <i>Indication,</i> dan <i>Corona Result</i>	24
Gambar 14. Kolom <i>Age 60 And Above, Gender, Test Indication,</i> dan <i>Corona Result</i> Setelah Dilabeling.....	25
Gambar 15. Proses <i>Splitting Dataset</i>	26
Gambar 16. <i>Code</i> <i>Splitting Train and Test Data</i>	27
Gambar 17. <i>Code</i> Library <i>Scikit-Learn Logistic Regression</i>	27
Gambar 18. <i>Code</i> <i>Logistic Regression</i>	28
Gambar 19. <i>Code</i> Library <i>Scikit-Learn Support Vector Machine</i>	29
Gambar 20. <i>Code</i> Library <i>Scikit-Learn Support Vector Machine</i>	29
Gambar 21. <i>Confusion Matrix</i>	30

Gambar 22. <i>Code Confusion Matrix Dari Library Scikit-Learn</i>	31
Gambar 23. <i>Confusion Matrix Neural Network</i>	33
Gambar 24. Hasil Tingkat Akurasi <i>Neural Network Pada Google Colaboratory</i>	34
Gambar 25. <i>Confusion Matrix Logistic Regression</i>	35
Gambar 26. Hasil Tingkat Akurasi <i>Logistic Regression Pada Google Colaboratory</i>	36
Gambar 27. <i>Confusion Matrix Support Vector Machine Dengan Kernel Linear</i>	37
Gambar 28. Hasil Tingkat Akurasi <i>Support Vector Machine Kernel Linear Pada Google Colaboratory</i>	39
Gambar 29. <i>Confusion Matrix Support Vector Machine Dengan Kernel Sigmoid</i>	39
Gambar 30. Hasil Tingkat Akurasi <i>Support Vector Machine Kernel Sigmoid Pada Google Colaboratory</i>	41
Gambar 31. <i>Confusion Matrix Support Vector Machine Dengan Kernel RBF</i>	41
Gambar 32. Hasil Tingkat Akurasi <i>Support Vector Machine Kernel RBF Pada Google Colaboratory</i>	43

DAFTAR TABEL

Tabel 1. Formula <i>Logistic Regression</i>	12
Tabel 2. <i>Dataset</i> Pasien Berdasarkan Gejala-Gejala <i>COVID-19</i>	14
Tabel 3. <i>Train Data</i> dan <i>Test Data</i>	26
Tabel 4. <i>Dataset</i> Pasien Berdasarkan <i>Gejala-Gejala COVID-19</i> Setelah Dilakukan <i>Data Preprocessing</i>	32
Tabel 5. <i>Dataset</i> Pasien Berdasarkan Gejala-Gejala <i>COVID-19</i>	44
Tabel 6. <i>Dataset</i> Pasien Berdasarkan Gejala-Gejala <i>COVID-19</i> Sebelum dan Sesudah Dilakukan Tahapan <i>Data Preprocessing</i>	46
Tabel 7. Nilai <i>Confusion Matrix</i> Menggunakan Metode <i>Logistic Regression</i>	48
Tabel 8. Nilai <i>Confusion Matrix</i> Menggunakan Metode <i>Support Vector Machine Kernel Linear</i>	51
Tabel 9. Nilai <i>Confusion Matrix</i> Menggunakan Metode <i>Support Vector Machine Kernel Sigmoid</i>	53
Tabel 10. Nilai <i>Confusion Matrix</i> Menggunakan Metode <i>Support Vector Machine Kernel RBF</i>	56
Tabel 11. Hasil Tingkat Akurasi.....	58