



**DETEKSI DINI COVID-19 DARI SUARA BATUK DENGAN EKSTRAKSI
FITUR MFCC MENGGUNAKAN MACHINE LEARNING**

SKRIPSI

Untuk Memenuhi Persyaratan

Dalam Menyelesaikan Sarjana Strata – 1 Ilmu Komputer

Oleh

MUHAMMAD THORIQ HIDAYAT

NIM 1911016210010

**PROGRAM STUDI S-1 ILMU KOMPUTER
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMBUNG MANGKURAT
BANJARBARU**

JUNI 2023



**DETEKSI DINI COVID-19 DARI SUARA BATUK DENGAN EKSTRAKSI
FITUR MFCC MENGGUNAKAN MACHINE LEARNING**

SKRIPSI

Untuk Memenuhi Persyaratan

Dalam Menyelesaikan Sarjana Strata – 1 Ilmu Komputer

Oleh

MUHAMMAD THORIQ HIDAYAT

NIM 1911016210010

**PROGRAM STUDI S-1 ILMU KOMPUTER
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMBUNG MANGKURAT
BANJARBARU
JUNI 2023**

SKRIPSI
DETEKSI DINI COVID-19 DARI SUARA BATUK DENGAN EKSTRAKSI
FITUR MFCC MENGGUNAKAN MACHINE LEARNING

Oleh
MUHAMMAD THORIQ HIDAYAT
1911016210010

Telah dipertahankan di depan Dosen Penguji pada tanggal 21 Juni 2023.

Susunan Dosen Penguji :

Pembimbing I



Mohammad Reza Faisal, S.T., M.T., Ph.D.
NIP. 1976122020008121001

Dosen Penguji I



Fatma Indriani, S.T., M.I.T., Ph.D.
NIP. 198404202008122004

Pembimbing II



Dwi Kartini, S.Kom., M.Kom.
NIP. 198704212012122003

Dosen Penguji II



Irwan Budiman, S.T., M.Kom.
NIP. 197703252008121001

Banjarbaru, _____ 2023

Ketua Program Studi Ilmu Komputer,



Irwan Budiman, S.T., M.Kom.
NIP. 197703252008121001

PERNYATAAN

Dengan ini saya menyatakan bahwa dalam skripsi ini tidak terdapat karya yang pernah diajukan untuk memperoleh gelar kesarjanaan di suatu Perguruan Tinggi, dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain, kecuali yang secara tertulis diacu dalam naskah ini dan disebutkan dalam Daftar Pustaka.

Banjarbaru, 21 Juni 2023



Muhammad Thoriq Hidayat

NIM. 1911016210010

ABSTRAK

DETEKSI DINI COVID-19 DARI SUARA BATUK DENGAN EKSTRAKSI FITUR MFCC MENGGUNAKAN MACHINE LEARNING

(Oleh: Muhammad Thoriq Hidayat; Pembimbing: Mohammad Reza Faisal, S.T., M.T., PhD.; Dwi Kartini, S.Kom., M.Kom.; 2023; 66 halaman)

Penelitian ini membahas tentang pendeteksian dini COVID-19 dengan menggunakan klasifikasi audio batuk dan ekstraksi fitur MFCC. Pengujian dilakukan dengan menggunakan data batuk dari COVID-19 Cough Classification dataset, dengan total 28 fitur dan 1926 data. Penelitian ini terdiri dari beberapa tahapan, yaitu ekstraksi data audio menggunakan spectrogram dan MFCC, serta pengklasifikasian menggunakan tiga metode berbeda. Metode pertama menggunakan Support Vector Machine (SVM) dengan kernel linear dan RBF, diikuti oleh metode klasifikasi Random Forest, dan terakhir metode klasifikasi Naive Bayes. Hasil pengklasifikasian dievaluasi dan perbandingan dilakukan berdasarkan nilai AUC. Evaluasi menunjukkan bahwa metode Naive Bayes memperoleh nilai AUC sebesar 0.638428, sedangkan metode Random Forest memperoleh nilai AUC sebesar 0.611755. Metode Support Vector Machine dengan kernel Linear memperoleh nilai AUC sebesar 0.599854, sedangkan metode Support Vector Machine dengan kernel RBF memperoleh nilai AUC tertinggi sebesar 0.657715. Hasil penelitian ini menunjukkan bahwa metode klasifikasi terbaik adalah Support Vector Machine dengan kernel RBF.

Kata kunci: Batuk, *Support Vector Machine*, *Random Forest*, *Naive Bayes*.

ABSTRACT

EARLY DETECTION OF COVID-19 FROM COUGH BY EXTRACTION OF MFCC FEATURES USING MACHINE LEARNING

(By: Muhammad Thoriq Hidayat; Advisors: Mohammad Reza Faisal, S.T., M.T., PhD.; Dwi Kartini, S. Kom.; 2023; 66 pages)

This study discusses the early detection of COVID-19 using cough audio classification and MFCC feature extraction. The test was carried out using cough data from the COVID-19 Cough Classification dataset, with a total of 28 features and 1926 data. This study consisted of several stages, namely audio data extraction using a spectrogram and MFCC, as well as classification using three different methods. The first method uses a Support Vector Machine (SVM) with a linear kernel and RBF, followed by the Random Forest classification method, and finally the Naive Bayes classification method. Classification results were evaluated and comparisons were made based on AUC values. The evaluation showed that the Naive Bayes method obtained an AUC value of 0.638428, while the Random Forest method obtained an AUC value of 0.611755. The Support Vector Machine method with the Linear kernel obtained an AUC value of 0.599854, while the Support Vector Machine method with the RBF kernel obtained the highest AUC value of 0.657715. The results of this study indicate that the best classification method is the Support Vector Machine with the RBF kernel.

Keywords: *Cough, Support Vector Machine, Random Forest, Naive Bayes.*

PRAKATA

Puji dan syukur kepada ALLAH SWT atas berkat rahmat-Nya penulis dapat menyelesaikan skripsi yang berjudul **DETEKSI DINI COVID-19 DARI SUARA BATUK DENGAN EKSTRAKSI FITUR MFCC MENGGUNAKAN MACHINE LEARNING** untuk memenuhi syarat dalam menyelesaikan pendidikan program Strata-1 Ilmu Komputer, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Lambung Mangkurat.

Tidak lupa penulis menyampaikan ucapan terimakasih kepada pihak-pihak terkait yang sangat mendukung dan membantu penulis dalam penyusunan skripsi ini, adapun yang dimaksud adalah sebagai berikut:

1. Allah SWT, karena atas limpahan karunia-Nya lah penulis dapat menyelesaikan skripsi ini.
2. Abah, mama, kaka-kaka, dan seluruh keluarga yang selalu memberikan doa, semangat, dukungan, hingga kepercayaan yang membuat penulis selalu bekerja keras menyelesaikan skripsi.
3. Bapak Mohammad Reza Faisal, S.T., M.T., Ph.D. dan Ibu Dwi Kartini, S. Kom., M. Kom. selaku dosen pembimbing utama dan pendamping pada pembuatan skripsi ini.
4. Ibu Fatma Indriani, S.T, M.I.T. dan Bapak Irwan Budiman, S.T., M.Kom. selaku dosen ketua penguji dan anggota penguji dalam skripsi ini.
5. Bapak Irwan Budiman, S.T., M.Kom. selaku ketua program studi Ilmu Komputer beserta seluruh dosen dan karyawan/staff pegawai Fakultas MIPA Universitas Lambung Mangkurat atas bantuan yang diberikan selama penulis mengikuti studi.
6. Cough VID gank yang telah membantu dengan segala kritik, saran dan masukan. dalam menjalankan penelitian.
7. Falah Isnain, Muhammad Meftah Mafazy, Khairi Yahya, Muhammad Hanan Ababil, M. Fajar Aditya, Raditya Raka Alif Azizi, Bagas Rony Hidayat, Naufal Abiyyu Mayusuf, Dwa Meizadewa, Muhammad Nouval Rahman, Muhammad

Farros Zulfan, dan Muhammad Khairie. Sebagai kawan dalam berkeluh kesah, bertukar fikiran, menjalankan perkuliahan, serta tempat mengembangkan diri.

8. Intelcore 2019 yang telah membantu dalam menjalankan perkuliahan bersama dari awal sampai akhir.
9. Seluruh mahasiswa Ilmu Komputer ULM yang telah membantu dalam keseharian di lingkungan kampus ULM Banjarbaru.
10. Semua pihak yang tidak dapat disebutkan satu persatu yang telah turut membantu dalam menyelesaikan skripsi ini.

Akhir kata penulis menyadari bahwa skripsi ini jauh dari sempurna, kritik dan saran yang membangun sangat diperlukan. Semoga tulisan ini dapat bermanfaat bagi ilmu pengetahuan dan pembaca khususnya serta mendapat keridhaan Allah SWT.

Banjarbaru, juni 2023

Muhammad Thoriq Hidayat

DAFTAR ISI

HALAMAN JUDUL	i
PERNYATAAN.....	iv
ABSTRAK	v
ABSTRACT.....	vi
PRAKATA.....	vii
DAFTAR ISI.....	ix
DAFTAR TABEL	xi
BAB I PENDAHULUAN.....	1
1.1. Latar Belakang	1
1.2. Rumusan Masalah	3
1.3. Tujuan.....	3
1.4. Manfaat Penelitian.....	3
1.5. Batasan Masalah.....	3
BAB II TINJAUAN PUSTAKA.....	5
2.1 Literatur Terdahulu	5
2.2 COVID-19	6
2.3 <i>Data mining</i>	6
2.4 Preprocessing	7
2.5 <i>Spectrogram</i>	7
2.6 Mel-Frequency Cepstral Coefficient.....	8
2.7 <i>K-Fold Cross Validation</i>	10
2.8 Support vector machine.....	11
2.9 Random forest	13
2.10 Naive bayes	14
2.11 Confusion Matriks.....	15
2.12 AUC	16
2.13 Keaslian Penelitian.....	17
BAB III METODE PENELITIAN	19

3.1	Alat Penelitian	19
3.2	Bahan Penelitian	19
3.3	Prosedur Penelitian.....	20
BAB IV HASIL DAN PEMBAHASAN		23
4.1	Hasil	23
4.1.1	Pengumpulan Data	23
4.1.2	Preprocessing.....	24
4.1.3	<i>K-Fold Cross Validation</i>	28
4.1.4	<i>Klasifikasi</i>	29
4.1.5	Evaluasi	40
4.2	Pembahasan.....	51
BAB V PENUTUP.....		65
5.1	Kesimpulan.....	65
5.2	Saran.....	65
DAFTAR PUSTAKA		67
LAMPIRAN.....		69

DAFTAR TABEL

Tabel 1 Confusion Matrix	15
Tabel 2 Keaslian Penelitian.....	17
Tabel 3 Tabel Perancangan Penelitian	18
Tabel 4 dataset.....	23
Tabel 5 Tampilan data gambar.....	24
Tabel 6 Fitur pada dataset	25
Tabel 7 Daftar fitur.....	26
Tabel 8 pembagian data pada -fold cross validation.....	28
Tabel 9 Data Perhitungan Manual SVM Linear	30
Tabel 10 Hasil Pencarian Nilai Kernel Linear	31
Tabel 11 data testing hitung manual SVM Linear	31
Tabel 12 Data Perhitungan SVM RBF.....	32
Tabel 13 Hasil Pencarian Nilai Kernel RBF.....	33
Tabel 14 Tabel Data Hitung Manual Random Forest	34
Tabel 15 Tabel Pohon Keputusan Pertama	35
Tabel 16 Nilai Pohon Kedua.....	36
Tabel 17 Nilai Pohon Ketiga.....	36
Tabel 18 Data Hitung Manual Naive Bayes	37
Tabel 19 Data Uji Coba Perhitungan	39
Tabel 20 Hasil Confusion Matrix Support Vector Machine, Kernel: RBF.....	40
Tabel 21 Hasil Kinerja Parameter Terbaik Untuk SVM Kernel RBF	42
Tabel 22 K-Fold untuk SVM RBF C:0.9	43
Tabel 23 Hasil Confusion Matrix Support Vector Machine, Kernel: Linear.....	43
Tabel 24 Hasil Kinerja Parameter Terbaik Untuk SVM kernel Linear	45
Tabel 25 K-Fold Untuk SVM Linear C:0.9	46
Tabel 26 Hasil Confusion Matrix Random Forest	46
Tabel 27 Hasil Kinerja Parameter Terbaik Untuk Random Forest	48
Tabel 28 K-Fold Untuk Random Forest Estimator 100	48
Tabel 29 Hasil confusion matrix Naive Bayes.....	49

Tabel 30 Hasil Kinerja Naive Bayes Gaussian	50
Tabel 31 K-Fold Kntuk Naive Bayes Gaussian	50

DAFTAR GAMBAR

Gambar 1 Spectrogram	8
Gambar 2 MFCC.....	10
Gambar 3 Ilustrasi <i>Cross Valiation</i>	11
Gambar 4 Nilai AUC	17
Gambar 5. Alur Penelitian.....	20
Gambar 6 Gambaran data heathy.....	27
Gambar 7 Gambaran data covid-19	28
Gambar 8 Pohon Keputusan Yang Didapat	35
Gambar 9 Pohon Keputusan Yang Kedua	36
Gambar 10 Pohon Keputusan Yang Ketiga	37
Gambar 11 Confusion Matrix SVM RBF C:0.9	41
Gambar 12 Confusion Matrix SVM Linear C:0.9	44
Gambar 13 Confusion Matrix Random Forest n Estimator 100	47
Gambar 14 Confusion Matrix Naive Bayes	49
Gambar 15 Perbandingan Nilai akurasi SVM kernel RBF	52
Gambar 16 Perbandingan Nilai presisi SVM kernel RBF	53
Gambar 17 Perbandingan Nilai Sensitivity recall SVM kernel RBF	53
Gambar 18 Perbandingan Nilai Specificity SVM kernel RBF	54
Gambar 19 Perbandingan Nilai AUC SVM kernel RBF	55
Gambar 20 Hasil Evaluasi SVM RBF pada C0.9	55
Gambar 21 Perbandingan Nilai Akurasi SVM kernel Linear	56
Gambar 22 Perbandingan Nilai Presisi SVM kernel Linear	56
Gambar 23 Perbandingan Nilai Sensitivity recall SVM kernel Linear	57
Gambar 24 Perbandingan Nilai Specificity SVM kernel Linear	58
Gambar 25 Perbandingan Nilai AUC SVM kernel Linear	58
Gambar 26 Hasil Evaluasi SVM Linear pada C0.9	59
Gambar 27 Perbandingan Nilai Akurasi Random Forest.....	60
Gambar 28 Perbandingan Nilai presisi Random Forest.....	60
Gambar 29 Perbandingan Nilai Sensitivity recall Random Forest	61

Gambar 30 Perbandingan Nilai Specificity Random Forest.....	61
Gambar 31 Perbandingan Nilai AUC Random Forest.....	62
Gambar 32 Hasil Evaluasi Random Forest pada n Estimator 100	62
Gambar 33 Tampilan Data Naive Bayes.....	63
Gambar 34 Perbandingan Nilai AUC Antar Klasifikasi	64

DAFTAR LAMPIRAN

Lampiran 1 Code Untuk Mengimpor Library Yang Digunakan.....	70
Lampiran 2 Code Untuk Menampilkan Data Audio	70
Lampiran 3 Code Untuk Konversi Ke Spectogram	71
Lampiran 4 Code Untuk Inisialisasi Dataset.....	71
Lampiran 5 Code Untuk Klasifikasi Support Vector Machine	71
Lampiran 6 Code Untuk Evaluasi SVM	72
Lampiran 7 Code Untuk Klasifikasi Random Forest.....	73
Lampiran 8 Code Untuk Evaluasi Random Forest	73
Lampiran 9 Code Untuk Klasifikasi Naive Bayes	74
Lampiran 10 Code Untuk Evaluasi Naive Bayes.....	74
Lampiran 11 Riwayat Hidup Penulis.....	76