

LAPORAN SKRIPSI
PREDIKSI JUMLAH KASUS PENYAKIT DBD, TBC, DAN DIARE
BERDASARKAN DATA PER KECAMATAN DAN TAHUN DI
KABUPATEN BEKASI MENGGUNAKAN ALGORITMA DECISION
TREE REGRESSION



Disusun Oleh:

Muhammad Dzaki Zahirsyah

2021230043

PROGRAM STUDI TEKNOLOGI INFORMASI
FAKULTAS TEKNIK
UNIVERSITAS DARMA PERSADA
JAKARTA
2025

LEMBAR MONITORING BIMBINGAN



UNIVERSITAS DARMA PERSADA

Jl. Taman Malaka Selatan, Pondok Kelapa, Jakarta Timur, Indonesia 13450
Telp. (021) 8649051, 8649053, 8649057 Fax. (021) 8649052
E-mail : humas@unsada.ac.id Home page : <http://www.unsada.ac.id>

Instrumen Monitoring Bimbingan Skripsi Program Studi Teknologi Informasi

Tahun Akademik : 2024/2025 Genap

NIM>Nama Mhs : 2021230043/ Muhammad Dzaki Zahirsyah

Judul Skripsi : Prediksi Jumlah Kasus Penyakit DBD, TBC, dan Diare Berdasarkan Data Per Kecamatan dan Waktu di Kabupaten Bekasi Menggunakan Algoritma Decision Tree Regression

Dosen Pembimbing : Timor Setiyaningsih, St., Mt

No	BAB Utama Skripsi dan BATAS WAKTU Bimbingan	Materi Yang dibahas saat Konsultasi	Tanggal Bimbingan	TTD Dosen
1	BAB I PENDAHULUAN Paling lama upload: 9 Mei 2025	Latar belakang		
2				
3				
		Tanggal BAB I di ACC pembimbing =>		<i>[Signature]</i> 9 Mei 2025
4	BAB II LANDASAN TEORI Paling lama upload: 9 Mei 2025	Data base dutambahkan teorinya		
5		Tambahkan algoritma pembanding		
6				
		Tanggal BAB II di ACC pembimbing =>		<i>[Signature]</i> 9 Mei 2025
7	BAB III METODOLOGI Paling lama upload : 16 Mei 2025	Perhatikan kelengkapan data peleaskan secara rinci langkah		
8		langkah pada flowchart		
9				
		Tanggal BAB III di ACC pembimbing =>		<i>[Signature]</i> 23 Mei 2025



UNIVERSITAS DARMA PERSADA

Jl. Taman Malaka Selatan, Pondok Kelapa, Jakarta Timur, Indonesia 13450
 Telp. (021) 8649051, 8649053, 8649057 Fax. (021) 8649052
 E-mail : humas@unsada.ac.id Home page : <http://www.unsada.ac.id>

10	Percobaan/Demo Aplikasi atau Sistem Paling lama upload : 23 Mei 2025	Prediksi tampil semua jenis penyakit		
11		- penghapusan data prediksi hiper baik		
12		- dan hasil prediksi ditambah dengan rekomendasi		
13		- dan admin & user bisa beda tampilan		
		Tanggal Aplikasi/Sistem ACC pembimbing =>	11 Juni 2025	
14	BAB IV HASIL DAN PEMBAHASAN Paling lama upload : 13 Juni 2025	Laporan pada aplikasi threshold, diperjelas lebih 3		
15		top feature importance diperjelas, proses memperoleh hasilnya.		
16				
		Tanggal BAB IV di ACC pembimbing =>	17 Juni 2025	
17	BAB V PENUTUP Paling lama upload : 18 Juni 2025	Kesimpulan menjelaskan jawaban dari rumusan masalah		
18				
		Tanggal BAB V di ACC pembimbing =>	26 Juni 2025	

Catatan :

- Mahasiswa harus konsultasi jauh-jauh hari sebelum batas akhir tanggal per BAB nya.
- Tanggal Bimbingan dan ACC per BAB **HARUS** sebelum batas tanggal maksimum, tetapi boleh sebelum tanggalnya jika bisa lebih cepat
- Dokumen ini WAJIB diupload ke gform yang ditentukan pada range tanggal setiap BAB
- Ujian Seminar ISI akan diadakan pada range tanggal : 21 s.d 27 Juni 2025

ACC Mengikuti Seminar dari Pembimbing :

Jenis ACC	Tanggal	TTD Pembimbing
ACC Mendaftar Seminar ^{leji} Judat	26 Juni 2025	
ACC Mendaftar Sidang Skripsi	29 Juli 2025	

LEMBAR PERBAIKAN REVISI



UNIVERSITAS DARMA PERSADA

Jl. Taman Malaka Selatan, Pondok Kelapa, Jakarta Timur, Indonesia 13450
Telp. (021) 8649051, 8649053, 8649057 Fax. (021) 8649052
E-mail : humas@unsada.ac.id Home page : <http://www.unsada.ac.id>

LEMBAR REVISI - SIDANG SKRIPSI

NIM>Nama : 2021230043 – MUHAMMAD DZAKI ZAHIRSYAH
Fakultas/Prodi : Teknik / Teknologi Informasi

No.	Keterangan Revisi	Dosen
	Uji Coba pada lapang. /ohari. Perbaikan & dinasideln dh lapang ul keripulan	Dr. Bayu S. Accepted

Mengetahui,
Kaprodi Teknologi Informasi

Herianto S. MT

LEMBAR PERNYATAAN

Saya yang bertanda tangan dibawah ini :

Nama : Muhammad Dzaki Zahirsyah

NIM : 2021230043

Fakultas : Teknik

Jurusan : Teknologi Informasi

Dengan ini menyatakan bahwa Skripsi yang berjudul:

“Prediksi Jumlah Kasus Penyakit DBD, TBC, dan Diare Berdasarkan Data Per Kecamatan dan Tahun di Kabupaten Bekasi Menggunakan Algoritma Decision Tree Regression”

adalah benar-benar hasil karya saya sendiri dan bukan merupakan plagiat dari karya orang lain yang diajukan untuk memenuhi salah satu persyaratan memperoleh gelar Sarjana (S1) pada Program Studi Teknologi Informasi di Universitas Darma Persada . Semua sumber yang saya gunakan dalam penulisan skripsi ini telah saya cantumkan sesuai dengan kaidah penulisan ilmiah yang berlaku di Universitas Darma Persada.

Apabila di kemudian hari ditemukan adanya ketidakbenaran dalam pernyataan ini, maka saya bersedia menerima segala sanksi akademik sesuai dengan peraturan yang berlaku, termasuk pencabutan gelar kesarjanaan yang telah saya peroleh.

Demikian pernyataan ini saya buat dengan sesungguhnya untuk dapat dipergunakan sebagaimana mestinya.

Bekasi, 24 September 2025

A 10000 Rupiah postage stamp with a signature over it. The stamp features the Garuda Pancasila emblem and the text '10000', 'METRAL TEMPEL', and 'FBALX291987301'. The signature is in black ink and appears to be 'Muhammad Dzaki Zahirsyah'.

Muhammad Dzaki Zahirsyah

LEMBAR PENGUJI SKRIPSI

Skripsi yang berjudul:

“PREDIKSI JUMLAH KASUS PENYAKIT DBD, TBC, DAN DIARE
BERDASARKAN DATA PER KECAMATAN DAN TAHUN DI KABUPATEN
BEKASI MENGGUNAKAN ALGORITMA DECISION TREE REGRESSION”

ini telah diujikan pada tanggal

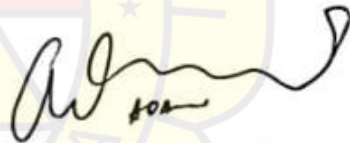
“4 Agustus 2025”

Penguji I



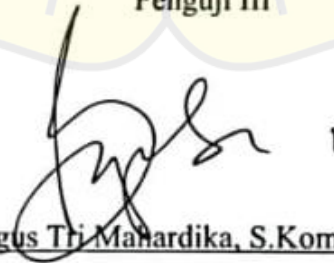
Timor Setyaningsih, ST, MTI

Penguji II



Adam Arif Budiman, ST, M.Kom

Penguji III



Bagus Tri Mahardika, S.Kom, MMSI

LEMBAR PENGESAHAN

PREDIKSI JUMLAH KASUS PENYAKIT DBD, TBC, DAN DIARE
BERDASARKAN DATA PER KECAMATAN DAN TAHUN DI KABUPATEN
BEKASI MENGGUNAKAN ALGORITMA DECISION TREE REGRESSION


Disusun Oleh :

Nama : Muhammad Dzaki Zahirsyah

Nim : 2021230043

Pembimbing Skripsi

Kajur Teknologi Informasi


Timot Setiyaningsih, ST, MTI


Herianto S.Pd, M.T

KATA PENGANTAR

Alhamdulillah, puji syukur saya panjatkan ke hadirat Allah ﷻ atas rahmat dan hidayah-Nya sehingga saya dapat menyelesaikan skripsi berjudul “Prediksi Jumlah Kasus Penyakit DBD, TBC, dan Diare Berdasarkan Data per Kecamatan dan Tahun di Kabupaten Bekasi Menggunakan Algoritma Decision Tree Regression”. Skripsi ini disusun sebagai salah satu syarat akademik untuk memperoleh gelar Sarjana (S1) pada Program Studi Teknologi Informasi, Universitas Darma Persada.

Dalam proses penyusunan Skripsi ini, saya memperoleh bimbingan, dukungan, dan bantuan dari berbagai pihak. Oleh karena itu saya mengucapkan terimakasih yang sebesar-besarnya kepada :

1. Orang tua yang telah memberikan do'a serta dukungan moril dan materil.
2. Ibu Timor Setiyaningsih, ST, MTI selaku pembimbing skripsi.
3. Ibu Astriarini selaku pegawai dinas kesehatan, yang bersedia memberikan informasi dan wawancara sebagai bagian dari data penelitian.
4. Serta rekan-rekan sesama mahasiswa Universitas Darma Persada yang telah banyak memberi bantuan dan dukungan.

Saya menyadari skripsi ini masih memiliki kekurangan, untuk itu kritik dan saran sangat saya harapkan. Semoga skripsi ini bermanfaat bagi semua pihak yang membutuhkan.

Bekasi, 29 Juni 2025



Muhammad Dzaki Zahirsyah
Penulis

ABSTRAK

Penyakit menular seperti Demam Berdarah Dengue (DBD), Tuberkulosis (TBC), dan Diare menjadi tantangan serius bagi kesehatan masyarakat di Kabupaten Bekasi, dengan jumlah kasus yang cenderung meningkat setiap tahun. Peningkatan kasus ini memerlukan sistem prediksi berbasis data untuk mengantisipasi lonjakan di masa mendatang dan mendukung pengambilan keputusan yang efektif oleh Dinas Kesehatan. Penelitian ini menjawab kebutuhan tersebut dengan membangun model prediksi jumlah kasus per kecamatan menggunakan algoritma *Decision Tree Regression* berdasarkan data sekunder dari tahun 2020 hingga 2023 yang diperoleh dari Dinas Kesehatan Kabupaten Bekasi, Badan Pusat Statistik (BPS), dan Badan Meteorologi, Klimatologi, dan Geofisika (BMKG), yang mencakup variabel kasus penyakit, demografi, dan lingkungan. Metodologi penelitian yang digunakan adalah Cross Industry Standard Process for Data Mining (CRISP-DM), Model ini dibangun menggunakan library Scikit-learn dalam bahasa pemrograman Python dan diimplementasikan dalam sebuah aplikasi web dengan arsitektur hibrida (PHP & Python/Flask). Sistem ini menyediakan antarmuka bagi admin untuk mengelola dataset, melakukan pelatihan ulang model, serta memperoleh hasil prediksi secara interaktif, lengkap dengan klasifikasi risiko dan rekomendasi penanganan. Hasil pengujian dengan metode *Hold-out Validation* menunjukkan kinerja model yang sangat akurat. Hal ini dibuktikan dengan perolehan nilai metrik *R-squared* (R^2) untuk model TBC sebesar 0.9130, model DBD sebesar 0.8805, dan model Diare sebesar 0.8228. Nilai-nilai ini membuktikan keandalan model dalam menjelaskan variasi data aktual. Kesimpulannya, sistem prediksi yang dihasilkan menawarkan keunggulan signifikan dengan menyediakan prediksi kuantitatif yang objektif, menjadi alat bantu yang lebih cepat dan terukur bagi Dinas Kesehatan Kabupaten Bekasi dalam menyusun kebijakan preventif.

Kata Kunci: *Decision Tree Regression*, Prediksi Penyakit, *Machine Learning*, Kesehatan Masyarakat, Kabupaten Bekasi.

DAFTAR ISI

LEMBAR MONITORING BIMBINGAN	ii
LEMBAR PERBAIKAN REVISI	iv
LEMBAR PERNYATAAN	v
LEMBAR PENGUJI SKRIPSI	vi
LEMBAR PENGESAHAN	vii
KATA PENGANTAR	viii
ABSTRAK	ix
DAFTAR ISI	x
DAFTAR TABEL	xvii
DAFTAR GAMBAR	xviii
DAFTAR LAMPIRAN	xx
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	4
1.3 Batasan Masalah	5
1.4 Tujuan Dan Manfaat Penelitian	5
1.4.1 Tujuan Penelitian	5
1.4.2 Manfaat Penelitian	6
1.5 Metode Penelitian	6
1.5.1 Metode Pengumpulan Data	6
1.5.2 Metode Pengembangan Sistem	7

1.5.3 Metode Analisis Data.....	8
BAB II LANDASAN TEORI	12
2.1 Tinjauan Pustaka	12
2.1.1 Kesehatan Masyarakat dan Penyakit Menular	12
2.1.1.1 Definisi dan Karakteristik Demam Berdarah Dengue (DBD)	13
2.1.1.2 Definisi dan Karakteristik TBC	14
2.1.2 Machine Learning	15
2.1.3 Data Mining	16
2.1.4 Prediksi.....	17
2.1.5 Algoritma Decision Tree Regression	17
2.1.6 CRISP-DM (Cross Industry Standard Process for Data Mining)	19
2.1.7 Perbedaan Konseptual Regresi dan Decision Tree Regression	20
2.1.8 Pemodelan UML	21
2.1.8.1 Use Case Diagram.....	21
2.1.8.2 Diagram Activity	22
2.1.8.3 Sequence Diagram	23
2.1.9 Database dan DBMS	24
2.1.9.1 SQL (Structured Query Language)	24
2.1.9.2 MySQL.....	25
2.1.9.3 XAMPP	25
2.1.10 Software Dan Pemrograman Terkait.....	26
2.1.10.1 Editor Jupyter	26
2.1.10.2 Bahasa Pemrograman Python dan Library Terkait	27

2.1.10.3 Library Scikit-learn	28
2.1.10.4 Bahasa Pemrograman PHP	29
2.2 Kajian Penelitian Terdahulu.....	29
2.2.1 Paper 1.....	29
2.2.1.1 Tujuan Penelitian	30
2.2.1.2 Metodologi Yang Digunakan.....	31
2.2.1.3 Temuan Utama	31
2.2.1.4 Kesimpulan penelitian.....	32
2.2.2 Paper 2.....	32
2.2.2.1 Tujuan Penelitian	33
2.2.2.2 Metodologi Yang Digunakan.....	33
2.2.2.3 Temuan Utama	34
2.2.2.4 Kesimpulan Penelitian	34
2.2.3 Paper 3.....	35
2.2.3.1 Tujuan Penelitian	35
2.2.3.2 Metodologi Yang Digunakan.....	36
2.2.3.3 Temuan Utama	36
2.2.3.4 Kesimpulan Penelitian	37
2.2.4 Paper 4.....	37
2.2.4.1 Tujuan Penelitian	38
2.2.4.2 Metodologi Yang Digunakan.....	38
2.2.4.3 Temuan Utama	38
2.2.4.4 Kesimpulan	39

2.2.5 Paper 5.....	39
2.2.5.1 Tujuan Penelitian	40
2.2.5.2 Metodologi Yang Digunakan.....	40
2.2.5.3 Temuan Utama	41
2.2.5.4 Kesimpulan Penelitian	42
BAB III METODOLOGI PENELITIAN.....	43
3.1 Rancangan Dasar Penelitian.....	43
3.1.1 Bidang dan Jenis Penelitian	43
3.1.2 Lokasi Penelitian.....	43
3.1.3 Jadwal Tahapan Penelitian.....	44
3.2 Rancangan Metodologi Penelitian	44
3.2.1 Perancangan UML	45
3.2.1.1 Use Case Diagram.....	45
3.2.1.2 Aktiviti Diagram	48
3.2.2 Perancangan Struktur Database.....	55
3.2.2.2 Tabel Historical_Data	56
3.2.2.3 Tabel Data Predictions_Summary.....	57
3.2.2.4 Tabel Prediction_Details.....	57
3.2.2.5 Tabel Data Recommendation_rules	58
3.2.2.6 Tabel Hasil_Pelatihan	59
3.2.2.10 Struktur Relasi Database	59
3.2.3 Perancangan Interface Aplikasi.....	61
3.2.3.1 Rancangan Halaman Login	61

3.2.3.2 Rancangan Halaman Dashboard	62
3.2.3.3 Rancangan Halaman Upload Data	62
3.2.3.4 Rancangan Halaman Prediksi	63
3.2.3.5 Rancangan Halaman Form input Prediksi.....	63
3.2.3.6 Rancangan Halaman Klasifikasi Tingkat Kasus	64
3.2.3.7 Rancangan Halaman Rekomendasi Penanganan	64
3.2.3.8 Rancangan Halaman Cetak Hasil Prediksi.....	65
3.2.4 Perancangan flowchart algoritma.....	65
3.2.5 Analisa Tahap Business understanding.....	71
3.2.5.1 Determine Business Objectives.....	71
3.2.5.2 Assess Situation	71
3.2.5.3 Determine Data Mining Goals	72
3.2.5.4 Produce Project Plan	72
3.2.6 Analisa Tahap Data understanding	73
3.2.6.1 Collect Initial Data	74
3.2.7 Rancangan Tahap Data Preparation.....	74
3.2.7.1 Pembersihan Data.....	75
3.2.7.2 Feature Encoding	75
3.2.8 Rancangan Tahap Pemodelan	76
3.2.8.1 Pemodelan Decision Tree	76
3.2.8.2 Perancangan Aturan Klasifikasi (Threshold).....	84
3.2.9 Rancangan Tahap Pengujian (Testing)	85
3.2.9.1 Pengujian Fungsional.....	85

3.2.9.2 Pengujian Kinerja Model (Model Evaluation).....	86
3.2.10 Rancangan Tahap Deployment	87
3.2.10.1 Arsitektur Sistem.....	87
3.2.10.2 Rancangan Hosting dan Deployment.....	88
BAB IV HASIL DAN PEMBAHASAN	90
4.1 Hasil Penelitian	90
4.1.1 Spesifikasi Hardware dan Software Yang Digunakan	90
4.1.2 Tampilan Interface Hasil Deploy	91
4.1.2.1 Halaman Login.....	92
4.1.2.2 Halaman Dashboard Admin Dinas Kesehatan.....	92
4.1.2.3 Halaman Manajemen Dataset	93
4.1.2.4 Halaman Pelatihan Model.....	94
4.1.2.5 Tampilan Halaman Prediksi Kasus	95
4.1.2.6 Tampilan Halaman Manajemen Penanganan.....	96
4.1.2.7 Tampilan Halaman Laporan prediksi.....	97
4.1.3 Struktur Database	98
4.1.3.1 Tabel users	98
4.1.3.2 Tabel historical_data	99
4.1.3.3 Tabel hasil_pelatihan	100
4.1.3.4 Tabel predictions_summary	101
4.1.3.5 Tabel prediction_details	102
4.1.3.6 Tabel recommendation_rules	103
4.2 Analisa Hasil	103

4.2.1 Percobaan Input – Output	104
4.2.1.1 Input pada Halaman Manajemen Dataset:	104
4.2.1.2 Input pada Halaman Pelatihan Model:.....	105
4.2.1.3 Input pada Halaman Prediksi Kasus:	106
4.2.2 Testing Hasil	108
4.2.2.1 Metode Pengujian: Hold-out Validation	112
4.2.2.2 Evaluasi Kinerja Model Decision Tree Regression	113
4.2.2.3 Analisis Kepentingan Fitur (Feature Importance).....	131
4.2.3 Keunggulan dan Optimalisasi Sistem	135
4.2.4 Proses Deploy Sistem Aplikasi	138
BAB V PENUTUP.....	147
5.1 Kesimpulan	147
5.2 Saran.....	148
DAFTAR PUSTAKA	150
LAMPIRAN.....	156

DAFTAR TABEL

Table 2.1 Komponen komponen Use Case	20
Table 2.2 Komponen dalam activity diagram	21
Table 3.1 Jadwal Penelitian.....	43
Table 3.2 tb_pengguna	54
Table 3.3 historical_data	55
Table 3.4 predictions_summary	56
Table 3.5 prediction_details	57
Table 3.6 recommendation_rules	57
Table 3.7 hasil_pelatihan	58
Table 3.8 Dataset Dinas Kesehatan. BMKG, BPS	74
Table 3.9 Hasil Penerapan Encoding	76
Table 3.10 Contoh Data Training (Decision Tree Regression).....	77
Table 3.11 Hasil Perhitungan Variance Reduction pada Node Akar.....	83
Table 4.1 users	90
Table 4.2 historical_data	91
Table 4.3 hasil_pelatihan	92
Table 4.4 predictions_summary	93
Table 4.5 prediction_details	94
Table 4.6 rekomendasi_rules	95
Table 4.7 Dataset Kasus Penyakit.....	100
Table 4.8 Dataset Faktor Lingkungan.....	102
Table 4.9 Tabel Data Aktual vs Prediksi Model DBD.....	107
Table 4.10 Tabel Data Aktual vs Prediksi Model TBC	112
Table 4.11 Tabel Data Aktual vs Prediksi Model Diare	118
Table 4.12 Kinerja Metrik Model	122

DAFTAR GAMBAR

Gambar 1.1 Tahapan Metode Waterfall (Harjono et al., 2022)	7
Gambar 1.2 Metode CRISP-DM.....	9
Gambar 3.1 Use case Diagram Sistem Prediksi Penyakit.....	44
Gambar 3.2 Diagram Aktiviti Login Dashboard Admin	48
Gambar 3.3 Diagram Aktiviti Mengelola Data	49
Gambar 3.4 Diagram Aktiviti prediksi Jumlah kasus Penyakit	50
Gambar 3.5 Diagram Aktiviti Klasifikasi Tingkat kasus.....	51
Gambar 3.6 Diagram Aktiviti Rekomendasi Penanganan	52
Gambar 3.7 Diagram Aktiviti Mencetak hasil Prediksi	53
Gambar 3.8 Struktur Relasi Database	59
Gambar 3.9 Rancangan Tampilan Halaman Login.....	60
Gambar 3.10 Rancangan Halaman Dashboard	61
Gambar 3.11 Rancangan Halaman Upload Data	61
Gambar 3.12 Rancangan Halaman Prediksi.....	62
Gambar 3.13 Rancangan Halaman Input Prediksi	62
Gambar 3.14 Rancangan Halaman Klasifikasi	63
Gambar 3.15 Rancangan Halaman Penanganan	63
Gambar 3.16 Rancangan Halaman Cetak Hasil.....	64
Gambar 3.17 Flowchart Algoritma	65
Gambar 4.1 Tampilan Login.....	84

Gambar 4.2 Halaman Dashboard Admin Dinkes.....	85
Gambar 4.3 Halaman Manajemen Dataset.....	86
Gambar 4.4 Halaman Pelatihan & Kinerja Model.....	87
Gambar 4.5 Halaman Prediksi Kasus.....	87
Gambar 4.6 Halaman Hasil Prediksi.....	88
Gambar 4.6 Halaman Manajemen Penanganan.....	89
Gambar 4.7 Tampilan Halaman Laporan Prediksi.....	89
Gambar 4.8 Tampilan Output Hasil Pelatihan Model Pada Sistem.....	97
Gambar 4.9 pengisian data faktor penyebab.....	98
Gambar 4.10 Hasil Output Prediksi.....	99
Gambar 4.11 Plot Aktual vs Prediksi DBD.....	106
Gambar 4.12 Plot Aktual vs Prediksi TBC.....	112
Gambar 4.13 Plot Aktual vs Prediksi Diare.....	117
Gambar 4.14 Top Fitur Penting – DBD.....	125
Gambar 4.15 Top Fitur Penting – TBC.....	126
Gambar 4.16 Top Fitur Penting – Diare.....	127

DAFTAR LAMPIRAN

Lampiran 1 Surat Keterangan Bebas Plagiat	156
Lampiran 2 Hasil Turnitin.....	157
Lampiran 3 Source Code App.py.....	158
Lampiran 4 Source Code Process_Prediction.php.....	167

