

ANALISIS KOMPARATIF ALGORITMA RANDOM FOREST, SUPPORT VECTOR CLASSIFIER, GRADIENT BOOSTING, DAN XGBOOST TERHADAP PREDIKSI SERANGAN JANTUNG DI INDONESIA MENGGUNAKAN ONE-HOT ENCODING DAN ORDINAL ENCODING

TUGAS AKHIR

Ahmad Zain Mahmud 121140232

PROGRAM STUDI TEKNIK INFORMATIKA

JURUSAN TEKNOLOGI PRODUKSI DAN INDUSTRI

INSTITUT TEKNOLOGI SUMATERA

LAMPUNG SELATAN



ANALISIS KOMPARATIF ALGORITMA RANDOM FOREST, SUPPORT VECTOR CLASSIFIER, GRADIENT BOOSTING, DAN XGBOOST TERHADAP PREDIKSI SERANGAN JANTUNG DI INDONESIA MENGGUNAKAN ONE-HOT ENCODING DAN ORDINAL ENCODING

TUGAS AKHIR

Diajukan sebagai syarat untuk memperoleh gelar sarjana

Ahmad Zain Mahmud 121140232

PROGRAM STUDI TEKNIK INFORMATIKA

JURUSAN TEKNOLOGI PRODUKSI DAN INDUSTRI

INSTITUT TEKNOLOGI SUMATERA

LAMPUNG SELATAN

LEMBAR PENGESAHAN

Tugas Akhir Sarjana dengan judul Analisis Komparatif Algoritma Random Forest, Support Vector Classifier, Gradient Boosting, Dan Xgboost Terhadap Prediksi Serangan Jantung Di Indonesia Menggunakan One-Hot Encoding Dan Ordinal Encoding adalah benar dibuat oleh saya sendiri dan belum pernah dibuat dan diserahkan sebelumnya, baik sebagian ataupun seluruhnya, baik oleh saya ataupun orang lain, baik di Institut Teknologi Sumatera maupun di institusi pendidikan lainnya.

Lampung Selatan, 8 Mei 2025

Penulis,

Ahmad Zain Mahmud NIM 121140232 Pembimbing I,

Pembimbing II,

Martin C.T. Manullang, S.T., M.T. NIP. 123456789

Angga Wijaya, S.Si., M.Si. NIP. 123456789

Disahkan oleh, Koordinator Program Studi Teknik Informatika Jurusan Teknologi Produksi dan Industri Institut Teknologi Sumatera

Ir. Hira Laksmiwati Soemitro, M.Sc. NIP 123456789

HALAMAN PERNYATAAN ORISINALITAS

Tugas Akhir ini adalah karya saya sendiri, dan semua sumber baik yang dikutip maupun dirujuk telah saya nyatakan benar

Nama	: Ahmad Zain Mahmud
NIM	: 121140232
Tanda Tangan	:
Tanggal	:

HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR UNTUK KEPENTINGAN AKADEMIS

Sebagai civitas akademik Institut Teknologi Sumatera, saya yang bertanda tangan di bawah ini:

Nama : Ahmad Zain Mahmud

NIM : 121140232

Program Studi: Teknik Informatika

Jurusan : Teknik Teknologi Produksi dan Industri

Jenis Karya : Tugas Akhir

demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Institut Teknologi Sumatera **Hak Bebas Royalti Noneksklusif (Non-exclusive Royalty Free Right)** atas karya ilmiah saya yang berjudul:

Analisis Komparatif Algoritma Random Forest, Support Vector Classifier, Gradient Boosting, Dan Xgboost Terhadap Prediksi Serangan Jantung Di Indonesia Menggunakan One-Hot Encoding Dan Ordinal Encoding

beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti Noneksklusif ini Institut Teknologi Sumatera berhak menyimpan, mengalihmedia/formatkan, mengelola dalam bentuk pangkalan data (database), merawat, dan memublikasikan tugas akhir saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta.

Demikian pernyataan ini saya buat dengan sebenarnya.

Dibuat di : Lampung Selatan Pada tanggal : 8 Mei 2025

Yang menyatakan (Ahmad Zain Mahmud)

Analisis Komparatif Algoritma Random Forest, Support Vector Classifier, Gradient Boosting, Dan Xgboost Terhadap Prediksi Serangan Jantung Di Indonesia Menggunakan One-Hot Encoding Dan Ordinal Encoding

Ahmad Zain Mahmud (121140232)

Pembimbing I Martin C.T. Manullang, S.T., M.T.

Pembimbing II Angga Wijaya, S.Si., M.Si.

ABSTRAK

Kata Kunci: Kata Kunci 1, Kata Kunci 2

Comparative Analysis of Random Forest, Support Vector Classifier, Gradient Boosting, and XGBoost Algorithms for Heart Attack Prediction in Indonesia Using One-Hot Encoding and Ordinal Encoding

Ahmad Zain Mahmud (121140232) Pembimbing I Martin C.T. Manullang, S.T., M.T. Pembimbing II Angga Wijaya, S.Si., M.Si.

ABSTRACT

Keyword: Keyword 1, Keyword 2

MOTTO

PERSEMBAHAN

KATA PENGANTAR

(Tuliskan maksud penulisan laporan, misal "Laporan penelitian ini dimaksud kan untuk memenuhi salah "........Pada halaman ini mahasiswa berkesempatan untuk menyatakan terima kasih secara tertulis kepada pembimbing dan pihak lain yang telah memberi bimbingan, nasihat, saran dan kritik, kepada mereka yang telah membantu melakukan penelitian, kepada perorangan atau lembaga yang telah memberi bantuan keuangan, materi dan/atau sarana.

Cara menulis kata pengantar beraneka ragam, tetapi hendaknya menggunakan kalimat yang baku. Ucapan terima kasih agar dibuat tidak berlebihan dan dibatasi pada pihak yang terkait secara ilmiah (berhubungan dengan subjek/materi penelitian).

Tempat penyusunan TA, tgl-bln-thn Penulis,

Ahmad Zain Mahmud

DAFTAR ISI

Ha	lama	n Judul	i
Le	mbaı	Pengesahan	ii
Ha	lama	n Pernyataan Orisinalitas	iv
Ha	lama	n Persetujuan Publikasi	V
Ab	strak		vi
Ab	strac	t	vii
Mo	otto		viii
Pe	rsem	oahan	ix
Ka	ta Pe	ngantar	X
I	1.1 1.2 1.3 1.4 1.5 1.6	Latar Belakang Rumusan Masalah Tujuan Batasan Masalah Manfaat Penelitian Sistematika Penulisan auan Pustaka Dasar Teori 2.1.1 Subbab	1 1 2 2 2 2 3 3 4 4
	2.2	Studi Terkait	5
	3.1 3.2 3.3	isis dan Perancangan Analisis Masalah	7 7 7 7
IV	4.1 4.2 4.3 4.4	uasi dan PembahasanTujuan PengujianSkenario PengujianHasil PengujianPembahasan	8 8 8 8 9
V	Pent 5.1 5.2	Kesimpulan	10 10 10

	xii
Daftar Pustaka	10
Lampiran	11
A Instrumen Pengujian	13
B Rincian Kasus Uji	14

DAFTAR GAMBAR

2.1	Contoh gambar																													5	5
	conton gamour	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	-	_

DAFTAR TABEL

DAFTAR SIMBOL

(Tuliskan maksud penulisan laporan, misal "Laporan penelitian ini dimaksud kan untuk memenuhi salah "........Pada halaman ini mahasiswa berkesempatan untuk menyatakan terima kasih secara tertulis kepada pembimbing dan pihak lain yang telah memberi bimbingan, nasihat, saran dan kritik, kepada mereka yang telah membantu melakukan penelitian, kepada perorangan atau lembaga yang telah memberi bantuan keuangan, materi dan/atau sarana.

Cara menulis kata pengantar beraneka ragam, tetapi hendaknya menggunakan kalimat yang baku. Ucapan terima kasih agar dibuat tidak berlebihan dan dibatasi pada pihak yang terkait secara ilmiah (berhubungan dengan subjek/materi penelitian).

BABI

PENDAHULUAN

1.1 Latar Belakang

Penyakit jantung merupakan salah satu penyebab utama kematian di seluruh dunia dan menjadi masalah kesehatan masyarakat yang serius, termasuk di Indonesia. Berdasarkan data dari World Health Organization (WHO), penyakit kardiovaskular, termasuk serangan jantung, menyumbang sekitar 17,9 juta kematian setiap tahun secara global. Di Indonesia sendiri, data dari Kementerian Kesehatan menunjukkan bahwa penyakit jantung menempati urutan pertama sebagai penyebab kematian tidak menular.

Serangan jantung biasanya terjadi secara mendadak dan dapat menyebabkan kematian jika tidak segera mendapatkan penanganan medis. Oleh karena itu, deteksi dini terhadap risiko serangan jantung sangat penting agar tindakan preventif dan pengobatan dapat dilakukan lebih awal. Dalam konteks inilah teknologi informasi, khususnya pemanfaatan data science dan machine learning, dapat memberikan kontribusi yang signifikan dalam upaya prediksi dan pencegahan penyakit jantung.

Perkembangan pesat dalam bidang kecerdasan buatan dan pengolahan data memungkinkan pengembangan model prediktif yang dapat menganalisis data pasien dan memberikan estimasi risiko serangan jantung secara lebih akurat. Berbagai algoritma machine learning seperti Random Forest, Support Vector Classifier (SVC), Gradient Boosting, dan XGBoost telah digunakan dalam berbagai studi untuk klasifikasi dan prediksi penyakit kronis. Keempat algoritma ini dikenal memiliki performa yang baik dalam menangani data dengan kompleksitas tinggi.

Namun, performa dari algoritma tersebut sangat bergantung pada karakteristik data dan teknik praproses yang digunakan. Salah satu tahap penting dalam praproses data adalah encoding data kategorikal. Dua metode encoding yang umum digunakan adalah One-Hot Encoding, yang mengubah variabel kategorikal menjadi representasi biner, dan Ordinal Encoding, yang mengubah kategori menjadi urutan numerik. Pemilihan teknik encoding yang tepat dapat mempengaruhi akurasi dan generalisasi model secara signifikan.

Penelitian ini menggunakan dataset Heart Attack Prediction in Indonesia dari Kaggle, yang mencakup berbagai atribut seperti usia, jenis kelamin, gaya hidup, serta riwayat kesehatan. Melalui penelitian ini, dilakukan analisis komparatif terhadap empat algoritma klasifikasi dengan menggunakan dua metode encoding yang berbeda, untuk melihat kombinasi mana yang memberikan hasil prediksi terbaik. Selain itu, penelitian ini juga bertujuan untuk mengidentifikasi fitur-fitur yang paling berpengaruh dalam menentukan risiko serangan jantung, sehingga dapat dijadikan dasar pengambilan keputusan oleh tenaga medis atau instansi terkait.

Dengan demikian, penelitian ini tidak hanya berkontribusi dalam bidang akademik dan teknologi informasi, tetapi juga memiliki nilai aplikatif dalam bidang kesehatan masyarakat, khususnya dalam mendukung sistem pendukung keputusan untuk deteksi dini penyakit jantung.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah dijelaskan sebelumnya, berikut merupakan rumusan masalah pada penelitian tugas akhir ini:

- Bagaimana performa algoritma Random Forest, Support Vector Classifier, Gradient Boosting, dan XGBoost dalam memprediksi risiko serangan jantung?
- 2. Metode encoding mana (One-Hot Encoding atau Ordinal Encoding) yang memberikan hasil prediksi terbaik?
- 3. Apa saja variabel yang paling berpengaruh dalam prediksi risiko serangan jantung berdasarkan model yang dibangun?

1.3 Tujuan

Tujuan dari penelitian ini berdasarkan rumusan masalah yang juga menjadi dasar dilakukannya penelitian ini adalah sebagai berikut:

- Menganalisis dan membandingkan performa empat algoritma machine learning (Random Forest, SVC, Gradient Boosting, dan XGBoost) dalam memprediksi serangan jantung, guna menentukan algoritma dengan performa terbaik.
- 2. Membandingkan hasil prediksi antara penggunaan metode One-Hot Encoding dan Ordinal Encoding.
- 3. Mengidentifikasi fitur-fitur yang paling berpengaruh terhadap risiko serangan jantung berdasarkan hasil pelatihan model.

1.4 Batasan Masalah

Agar penelitian ini lebih terfokus, maka batasan masalah ditentukan sebagai berikut:

- Dataset yang digunakan merupakan dataset "Heart Attack Prediction in Indonesia" dari Kaggle yang telah disampling sebanyak 10.000 data.
- Hanya digunakan empat algoritma klasifikasi yaitu Random Forest, Support Vector Classifier, Gradient Boosting, dan XGBoost.
- Evaluasi performa model menggunakan metrik akurasi, precision, recall, dan F1-score.
- Encoding data kategorikal hanya menggunakan metode One-Hot Encoding dan Ordinal Encoding.

- Data yang dipertimbangkan hanya berdasarkan data pasien yang tertera di dataset.
- Penelitian ini dilakukan menggunakan bahasa pemrograman Python dengan library scikit-learn dan XGBoost.

1.5 Manfaat Penelitian

Adapun manfaat yang diharapkan dari penelitian ini adalah:

- Memberikan referensi bagi praktisi data dan peneliti dalam memilih algoritma terbaik untuk prediksi serangan jantung.
- Memberikan wawasan mengenai pengaruh teknik encoding terhadap performa model klasifikasi.
- Meningkatkan kesadaran akan pentingnya pemanfaatan teknologi informasi dalam bidang kesehatan, khususnya dalam upaya deteksi dini penyakit kardiovaskular.

1.6 Sistematika Penulisan

Sistematika penulisan skripsi ini dibagi menjadi beberapa bab sebagai berikut:

1.6.1 Bab I: Pendahuluan

Bab ini berisi latar belakang, rumusan masalah, tujuan penelitian, batasan masalah, manfaat penelitian, dan sistematika penulisan.

1.6.2 Bab II: Tinjauan Pustaka

Bab ini menjelaskan hasil kajian pustaka terkait topik yang diteliti, serta teori-teori pendukung seperti algoritma machine learning, teknik encoding, dan metrik evaluasi model.

1.6.3 Bab III: Metode Penelitian

Bab ini membahas metode penelitian yang digunakan, termasuk desain penelitian, alur penelitian, langkah-langkah implementasi, teknik praproses data, serta metode pengujian performa model.

1.6.4 Bab IV: Hasil dan Pembahasan

Bab ini menyajikan hasil pengolahan data, evaluasi performa masing-masing model, analisis terhadap hasil yang diperoleh, serta interpretasi terhadap pentingnya fitur dalam prediksi.

1.6.5 Bab V: Kesimpulan dan Saran

Bab ini berisi kesimpulan dari hasil penelitian serta saran untuk penelitian selanjutnya yang berkaitan dengan pengembangan model prediksi serangan jantung atau penerapan di bidang lain.

BAB II

TINJAUAN PUSTAKA

Bab Studi Literatur digunakan untuk mendeskripsikan kajian literatur yang terkait dengan persoalan tugas akhir. Tujuan studi literatur adalah:

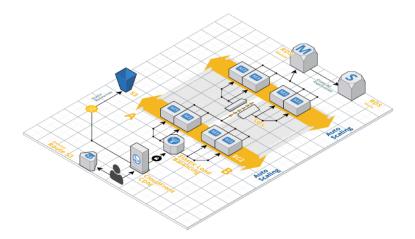
- 1. menunjukkan kepada pembaca adanya gap seperti pada rumusan masalah yang memang belum terselesaikan,
- memberikan pemahaman yang secukupnya kepada pembaca tentang teori atau pekerjaan terkait yang terkait langsung dengan penyelesaian persoalan, serta
- 3. menyampaikan informasi apa saja yang sudah ditulis/dilaporkan oleh pihak lain (peneliti/Tugas Akhir/Tesis) tentang hasil penelitian/pekerjaan mereka yang sama atau mirip kaitannya dengan persoalan tugas akhir.

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

2.1 Dasar Teori

Perujukan literatur dapat dilakukan dengan menambahkan entri baru di berkas. Tulisan ini merujuk pada [1]

2.1.1 Subbab



Gambar 2.1: Contoh gambar

2.1.1.1 Subsubbab

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

Tabel 2.1: Contoh tabel

Col1	Col2	Col2	Col3
1	6	87837	787
2	7	78	5415
3	545	778	7507
4	545	18744	7560
5	88	788	6344

2.2 Studi Terkait

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante

lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

BAB III

ANALISIS DAN PERANCANGAN

3.1 Analisis Masalah

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

3.2 Solusi Umum

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

3.3 Rancangan Solusi

BAB IV

EVALUASI DAN PEMBAHASAN

4.1 Tujuan Pengujian

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

4.2 Skenario Pengujian

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

4.3 Hasil Pengujian

4.4 Pembahasan

BAB V

PENUTUP

5.1 Kesimpulan

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

5.2 Saran

Daftar Pustaka

[1] D.E. Knuth. *The Art of Computer Programming: Fundamental Algorithms*. The Art of Computer Programming: Fundamental Algorithms v. 1. Addison-Wesley, 2001. ISBN: 9780201896831.

Lampiran

Lampiran A INSTRUMEN PENGUJIAN

Lampiran B RINCIAN KASUS UJI