

**PREDIKSI KELANGSUNGAN HIDUP PENUMPANG
TITANIC MENGGUNAKAN ALGORITMA PEMBELAJARAN
MESIN**

SKRIPSI SARJANA REKAYASA TEKNOLOGI INFORMATIKA



Oleh:

I Gede Agus Subawa

207064516134

**FAKULTAS TEKNOLOGI KOMUNIKASI DAN INFORMATIKA
UNIVERSITAS NASIONAL
JAKARTA
2024**

HALAMAN PENGESAHAN
TUGAS SARJANA
PREDIKSI KELANGSUNGAN HIDUP PENUMPANG TITANIC
MENGGUNAKAN ALGORITMA PEMBELAJARAN MESIN

Oleh:
I Gede Agus Subawa
207064516134

Program Studi Informatika
Fakultas Teknologi Komunikasi dan Informatika

Disetujui pada tanggal: 24 Agustus 2024

Pembimbing 1



Dr. Moh. Iwan Wahyuddin, S.T., M.T.
NIDN. 0318047004

Pembimbing 2



Aris Gunaryati, S.Si., M.M.S.I
NIDN: 0313087705

Penguji 1



Dr. Andrianingsih, S.Kom.,MMSI.
NIDN. 0303097902

Penguji 2



Agus Iskandar, S.Kom., M.Kom.
NIDN. 0310087503

HALAMAN PENGESAHAN

TUGAS AKHIR

PREDIKSI KELANGSUNGAN HIDUP PENUMPANG TITANIC
MENGUNAKAN ALGORITMA PEMBELAJARAN MESIN



PERNYATAAN KEASLIAN TUGAS AKHIR

Saya menyatakan dengan sesungguhnya bahwa Tugas Akhir dengan judul :

Prediksi Kelangsungan Hidup Penumpang Titanic Menggunakan Algoritma Pembelajaran Mesin

Yang dibuat untuk melengkapi salah satu persyaratan menjadi Sarjana Komputer pada Program Studi Informatika Fakultas Teknologi Komunikasi dan Informatika Universitas Nasional, sebagaimana yang saya ketahui adalah bukan merupakan tiruan atau publikasi dari Tugas Akhir yang pernah diajukan atau dipakai untuk mendapatkan gelar di lingkungan Universitas Nasional maupun perguruan tinggi atau instansi lainnya, kecuali pada bagian – bagian tertentu yang menjadi sumber informasi atau acuan yang dicantumkan sebagaimana mestinya.

Jakarta, 27 Agustus 2024



I Gede Agus Subawa

207064516134

LEMBAR PERSETUJUAN REVIEW AKHIR

Tugas Akhir dengan judul :

**PREDIKSI KELANGSUNGAN HIDUP PENUMPANG
TITANIC MENGGUNAKAN ALGORITMA**

PEMBELAJARAN MESIN

Dibuat untuk melengkapi salah satu persyaratan menjadi Sarjana Komputer pada Program Studi Informatika, Fakultas Teknologi Komunikasi dan Informatika Universitas Nasional. Tugas Akhir ini diujikan pada Sidang Review Akhir Semester Genap 2023/2024 pada tanggal 21 Agustus Tahun 2024

Dosen Pembimbing 1



Dr. Moh Iwan Wahyuddin,
S.T., M.T.
NIDN: 0318047004

Ketua Program Studi


Ratih Titi Komala Sari, ST
MM, MMSI
NID: 0103150850

UNIVERSITAS NASIONAL

LEMBAR PERSETUJUAN JUDUL YANG TIDAK ATAU YANG DIREVISI

Nama : I Gede Agus Subawa

NPM : 207064516134

Fakultas/Akademi : Fakultas Teknologi Komunikasi dan Informatika

Program Studi : Informatika

Tanggal Sidang : 21 Agustus 2024



JUDUL DALAM BAHASA INDONESIA :

PREDIKSI KELANGSUNGAN HIDUP PENUMPANG TITANIC
MENGUNAKAN ALGORITMA PEMBELAJARAN MESIN

JUDUL DALAM BAHASA INGGRIS :

PREDICTION OF TITANIC PASSENGER SURVIVAL USING MACHINE
LEARNING ALGORITHMS

TANDA TANGAN DAN TANGGAL

Pembimbing I	Ka. Prodi	Mahasiswa
TGL: 29/8/2024	TGL: 27/8/2024	TGL: 28/8/2024
		 I Gede Agus Subawa

LEMBAR PERSETUJUAN JUDUL YANG TIDAK ATAU YANG DIREVISI

Nama : I Gede Agus Subawa

NPM : 207064516134

Fakultas/Akademi : Fakultas Teknologi Komunikasi dan Informatika

Program Studi : Informatika

Tanggal Sidang : 21 Agustus 2024




JUDUL DALAM BAHASA INDONESIA :

PREDIKSI KELANGSUNGAN HIDUP PENUMPANG TITANIC
MENGUNAKAN ALGORITMA PEMBELAJARAN MESIN

JUDUL DALAM BAHASA INGGRIS :

PREDICTION OF TITANIC PASSENGER SURVIVAL USING MACHINE
LEARNING ALGORITHMS

TANDA TANGAN DAN TANGGAL

Pembimbing 2	Ka. Prodi	Mahasiswa
TGL: 29 Agustus 2024	TGL: 27/9/2024	TGL: 28/8/2024
 Aris Gunaryati, S.Si, MMSI		 I Gede Agus Subawa

ABSTRAK

Tenggelamnya kapal Titanic menjadi perhatian dunia serta menjadi subjek pemberitaan yang luas. Terdapat banyak buku, film, dan dokumenter yang menceritakan tragedi tersebut, salah satunya film berjudul "Titanic" yang dirilis tahun 1998. Penggunaan machine learning dalam memprediksi kelangsungan hidup penumpang Titanic memiliki potensi untuk memberikan wawasan tentang faktor-faktor yang berpengaruh. Dalam penelitian ini menggunakan model Logistic Regression, Decision Tree, Random Forest, dan Naive Bayes untuk dibandingkan dan memilih model terbaik dalam prediksi kelangsungan hidup penumpang Titanic. Evaluasi model dilakukan dengan metrik akurasi, presisi, F1-Score, serta cross validation untuk memastikan keakuratan dan generalisasi model. Dataset Titanic yang digunakan dieksplorasi menggunakan berbagai teknik statistika deskriptif dan perhitungan peluang selamat untuk setiap variabel. Hasil penelitian menunjukkan bahwa model Decision Tree memiliki kinerja yang baik dengan nilai tertinggi pada metrik akurasi, presisi, dan cross validation sehingga menjadikannya sebagai model terpilih untuk prediksi kelangsungan hidup penumpang Titanic.

Kata kunci: Titanic, prediksi kelangsungan hidup, *machine learning*, Decision Tree, Logistic Regression, Random Forest, Naive Bayes, akurasi, presisi, F1-Score, *cross validation*.

ABSTRACT

The sinking of the Titanic became the world's attention and the subject of extensive news coverage. There are many books, movies, and documentaries that recount the tragedy, one of which is the movie titled "Titanic" released in 1998. The use of machine learning in predicting the survival of Titanic passengers has the potential to provide insight into influential factors. In this study, Logistic Regression, Decision Tree, Random Forest, and Naive Bayes models were used to compare as the best models for survival prediction. Model evaluation is performed with accuracy, precision, F1-Score, and cross validation metrics to ensure model accuracy and generalization. The Titanic dataset used was explored using various descriptive statistical techniques and calculation of the chance of survival for each variable. The results showed that the Decision Tree model performed well with the highest scores on accuracy, precision, and cross validation metrics, making it the preferred model.

Keywords: Titanic, survival prediction, machine learning, Decision Tree, Logistic Regression, Random Forest, Naive Bayes, accuracy, precision, F1-Score, cross validation.



KATA PENGANTAR

Puji dan syukur penulis panjatkan kepada Tuhan Yang Maha Esa yang telah memberikan rahmat dan karunia sehingga penulis dapat menyelesaikan skripsi dengan judul “PREDIKSI KELANGSUNGAN HIDUP PENUMPANG TITANIC MENGGUNAKAN ALGORITMA PEMBELAJARAN MESIN”. Selama proses pengerjaan skripsi tidak terlepas dari bantuan berbagai pihak, oleh karena itu penulis menyampaikan banyak terima kasih kepada serta doa yang tulus kepada:

1. Bapak Dr. Moh. Iwan Wahyuddin, S.T., M.T. dan Ibu Aris Gunaryati, S.Si., M.M.S.I selaku dosen pembimbing 1 dan dosen pembimbing 2 yang telah memberikan saran dan arahan selama pengerjaan skripsi.
2. Ibu Rima Tamara Aldisa, S.Kom., M.Kom selaku dosen pembimbing akademik
3. Seluruh jajaran staf Program Studi Informatika secara langsung maupun tidak langsung terlibat dalam proses pengerjaan skripsi
4. Dosen pengajar di program studi Informatika yang telah memberikan banyak ilmu.
5. Orang tua dan keluarga yang memberikan nasehat, semangat, dan motivasi untuk menyelesaikan skripsi ini
6. Teman-teman satu angkatan yang telah membantu dan mendukung selama proses pengerjaan skripsi

Penulis menyadari bahwa penelitian ini masih jauh dari kata sempurna dan karenanya penulis membuka kesempatan untuk setiap saran dan kritik yang membangun demi penyempurnaan skripsi agar bermanfaat bagi penulis dan pihak yang membutuhkan.

Jakarta, 19 Juni 2024

Penulis

I Gede Agus Subawa

DAFTAR ISI

KATA PENGANTAR	i
ABSTRAK	i
ABSTRACT	viii
DAFTAR ISI	x
DAFTAR GAMBAR	xii
DAFTAR TABEL	xiv
BAB I PENDAHULUAN	1
1.1. Latar Belakang	1
1.2. Identifikasi Masalah	3
1.3. Batasan Masalah.....	3
1.4. Tujuan Penelitian.....	3
1.5. Manfaat Penelitian.....	4
BAB II TINJAUAN PUSTAKA	5
2.1. Dataset Titanic.....	5
2.2. <i>Machine Learning</i>	5
2.3. <i>Confusion Matrix</i>	6
2.4. Logistic Regression	7
2.5. Decision Tree.....	8
2.6. Random Forest	8
2.7. Naive Bayes.....	9
2.8. Studi Literatur.....	10
BAB III METODE PENELITIAN	14
3.1. Penentuan Objek Penelitian.....	14
3.2. Fokus Penelitian	14
3.3. Sumber Data	14
3.4. Desain Penelitian	15
3.4.1 Desain Tahapan Penelitian.....	15
3.4.2 Variabel Penelitian.....	17
3.4.3 Pemilihan Algoritma	17

3.4.4 Pengolahan Data	18
3.4.5 Evaluasi Model	19
BAB IV HASIL DAN PEMBAHASAN	22
4.1. Hasil Eksplorasi Dataset Titanic	22
4.1.1 Survived	23
4.1.2 Pclass (Kelas Penumpang)	23
4.1.3 Sex (Jenis Kelamin)	25
4.1.4 Embarked (Pelabuhan Keberangkatan / Embarkasi)	29
4.1.5 Cabin (Kabin)	32
4.1.6 Usia	37
4.1.7 SibSp dan Parch (Jumlah Anggota Keluarga)	41
4.1.8 Fare (Harga)	45
4.1.9 Korelasi Variabel Survived dengan Variabel Lainnya	48
4.2. Hasil Analisis Faktor Eksternal	49
4.2.1 Material Kapal	49
4.2.2 Jumlah Sekoci (Perahu Penyelamat)	50
4.2.3 Faktor Cuaca	51
4.3. Hasil Eksplorasi Data Setelah Imputasi Nilai yang Hilang	51
4.3.1 Usia	52
4.3.2 Cabin	56
4.3.3 Embarked	59
4.4. Penyajian Data dengan <i>Dashboard</i>	63
4.5. Persiapan Data untuk Model	64
4.6. Evaluasi dan Perbandingan Model	65
4.7. Implementasi Model Terpilih	68
BAB V KESIMPULAN DAN SARAN	70
5.1. Kesimpulan	70
5.2. Saran	71
DAFTAR PUSTAKA	72
LAMPIRAN	75

DAFTAR GAMBAR

Gambar 2. 1 Confusion Matrix	7
Gambar 2. 2 Ilustrasi Model Decision Tree	8
Gambar 2. 3 Ilustrasi Klasifikasi dengan Random Forest	9
Gambar 3. 1 Alur Tahapan CRISP-DM.....	15
Gambar 3. 2 Ilustrasi Cross Validation	20
Gambar 4. 1 Visualisasi jumlah penumpang selamat dan tidak serta proporsinya	23
Gambar 4. 2 Visualisasi distribusi dan proporsi kelas penumpang	23
Gambar 4. 3 Visualisasi Jumlah Penumpang dan Proporsi Berdasarkan Jenis Kelamin.....	25
Gambar 4. 4 Visualisasi Jumlah Selamat dan Tidak Berdasarkan Jenis Kelamin	26
Gambar 4. 5 Visualisasi Jumlah Jenis Kelamin di Setiap Kelas Penumpang.....	26
Gambar 4. 6 Visualisasi Jumlah dan Proporsi Penumpang Berdasarkan Embarkasi.....	29
Gambar 4. 7 Visualisasi Jumlah Penumpang Selamat dan Tidak Berdasarkan Embarkasi.....	29
Gambar 4. 8 Distribusi Jumlah Penumpang Berdasarkan Embarkasi dan Kelas Penumpang.....	30
Gambar 4. 9 Distribusi Jumlah Penumpang yang Selamat dan Tidak Berdasarkan Deck	33
Gambar 4. 10 Distribusi Jumlah Penumpang Berdasarkan Dek dan Jenis Kelamin.....	33
Gambar 4. 11 Distribusi Usia Penumpang.....	37
Gambar 4. 12 Distribusi Usia dan Survived	37
Gambar 4. 13 Distribusi Usia dengan Boxplot	38
Gambar 4. 14 Distribusi Kelompok Usia.....	38
Gambar 4. 15 Distribusi Parch dan SibSp.....	41
Gambar 4. 16 Distribusi Penumpang Sendirian dan Berkeluarga	42
Gambar 4. 17 Distribusi Penumpang Sendirian dan Berkeluarga Berdasarkan Survived, Kelas, dan Jenis Kelamin.....	42
Gambar 4. 18 Distribusi Harga Berdasarkan Kelas Penumpang	45
Gambar 4. 19 Distribusi Harga Berdasarkan Survived.....	45
Gambar 4. 20 Distribusi Harga Berdasarkan Jenis Kelamin.....	46
Gambar 4. 21 Scatterplot Harga dan Usia.....	46
Gambar 4. 22 Distribusi Usia Setelah dilakukan Imputasi dengan Median	52
Gambar 4. 23 Distribusi Usia dengan Boxplot	53
Gambar 4. 24 Distribusi Berdasarkan Kelompok Usia Setelah Imputasi	53
Gambar 4. 25 Distribusi Dek dan Survived Setelah Imputasi	56
Gambar 4. 26 Distribusi Embarkasi Setelah Imputasi	59
Gambar 4. 27 Distribusi Penumpang Selamat dan Tidak Berdasarkan Embarkasi.....	60
Gambar 4. 28 Distribusi Jenis Kelamin Berdasarkan Embarkasi	61

Gambar 4. 29 Dashboard Dataset Titanic 63
Gambar 4. 30 Confusion matrix dari model Decision Tree 66



DAFTAR TABEL

Tabel 2. 1 Jurnal Acuan	10
Tabel 4. 1 Distribusi Keselamatan Penumpang Berdasarkan Kelas Penumpang .	24
Tabel 4. 2 Ringkasan Harga Tiket Berdasarkan Kelas Penumpang.....	24
Tabel 4. 3 Peluang Selamat Berdasarkan Jenis Kelamin	27
Tabel 4. 4 Peluang Selamat Berdasarkan Jenis Kelamin dan Kelas Penumpang .	27
Tabel 4. 5 Peluang Selamat Berdasarkan Embarkasi dan Kelas Penumpang	30
Tabel 4. 6 Peluang Selamat Berdasarkan Embarkasi dan Jenis Kelamin	31
Tabel 4. 7 Distribusi Jumlah Penumpang Berdasarkan Dek dan Kelas Penumpang	32
Tabel 4. 8 Peluang Selamat Berdasarkan Dek	34
Tabel 4. 9 Peluang Selamat Berdasarkan Dek dan Jenis Kelamin.....	34
Tabel 4. 10 Peluang Selamat Berdasarkan Dek dan Kelas Penumpang	35
Tabel 4. 11 Distribusi Kelompok Usia Berdasarkan Peluang, Jumlah Selamat dan Tidak Selamat.....	39
Tabel 4. 12 Distribusi Peluang Selamat Berdasarkan Jenis Kelamin dan Peluang Selamat	39
Tabel 4. 13 Distribusi Selamat dan Tidak Berdasarkan Parch.....	41
Tabel 4. 14 Distribusi Selamat dan Tidak Berdasarkan SibSp	41
Tabel 4. 15 Nilai Korelasi Variabel Survived dan Variabel Lainnya	48
Tabel 4. 16 Distribusi Kelompok Usia Berdasarkan Peluang, Jumlah Selamat dan Tidak Selamat (Imputasi).....	54
Tabel 4. 17 Distribusi Peluang Selamat Berdasarkan Jenis Kelamin dan Peluang Selamat (Imputasi)	54
Tabel 4. 18 Distribusi Jumlah Penumpang Berdasarkan Dek dan Kelas Penumpang	56
Tabel 4. 19 Peluang Selamat Berdasarkan Dek	57
Tabel 4. 20 Peluang Selamat Berdasarkan Dek dan Jenis Kelamin.....	57
Tabel 4. 21 Peluang Selamat Berdasarkan Dek dan Kelas Penumpang	58
Tabel 4. 22 Distribusi Embarkasi dan Kelas Penumpang serta Peluang Selamat Setelah Imputasi	60
Tabel 4. 23 Peluang Selamat Berdasarkan Embarkasi dan Kelas Penumpang	61
Tabel 4. 24 Hasil Evaluasi Tiap Model.....	66