Jurnal Informatika dan Sistem Informasi (JIFoSI)
Vol. 6, No. 3. Desember 2025
e-ISSN : 2722-130
DOI: https://doi.org/10.33005/jifosi.v6i3.615

ANALISIS PREDIKSI KELANGSUNGAN HIDUP PENUMPANG TITANIC MENGGUNAKAN ALGORITMA DECISION TREE PADA DATASET KAGGLE

Raja Samudra Tandepadang[1)], Tryana Bagus Saputra[2)], Moch. Dicky Ferdyansyah[3)] , Clarisa Tita Agnevia[4)] , Rangga Razzaq Firdhaus[5)]
E-mail : [1)]24081010063.student@upnjatim.ac.id , [2)]24081010086.student@upnjatim.ac.id , [3)] 24081010121.student@upnjatim.ac.id , [4)] 24081010122.student@upnjatim.ac.id , [5)] 24081010277.student@upnjatim.ac.id
[1,2,3,4,5]Program Studi Informatika, Fakultas Ilmu Komputer, Universitas Pembangunan Nasional "Veteran" Jawa Timur

(Naskah masuk: 22 Desember 2025, diterima untuk diterbitkan: 31 Desember 2025)

Abstrak

Penelitian ini bertujuan menganalisis penerapan algoritma Decision Tree dalam memprediksi kelangsungan hidup penumpang Titanic menggunakan dataset yang diperoleh dari platform Kaggle. Dataset yang digunakan terdiri atas beberapa atribut input, yaitu Pclass, Sex, Age, SibSp, Parch, Fare, dan Embarked. Tahapan penelitian mencakup praproses data, transformasi atribut kategorikal, pembagian data latih dan data uji, pembangunan model Decision Tree, serta evaluasi performa model menggunakan confusion matrix. Hasil penelitian menunjukkan bahwa model Decision Tree mampu mencapai akurasi sebesar 90%, dengan nilai precision 83%, recall 82%, dan F1-score 82%. Selain itu, atribut gender, pclass, dan age terbukti menjadi faktor paling berpengaruh dalam prediksi kelangsungan hidup penumpang. Temuan ini membuktikan bahwa algoritma Decision Tree memiliki kemampuan klasifikasi yang andal pada dataset Titanic.

Kata Kunci: Decision Tree, Titanic, Confusion Matrix

1. PENDAHULUAN

Perkembangan teknologi informasi dan komputer telah mendorong kemajuan signifikan dalam bidang kecerdasan buatan atau Artificial Intelligence (AI). Teknologi AI kini banyak dimanfaatkan untuk membantu manusia dalam berbagai pekerjaan, seperti keperluan pembelajaran [1], hingga keperluan menganalisis data, mengenali pola, dan mengambil keputusan secara otomatis [2]. Salah satu cabang AI yang berkembang paling pesat adalah machine learning, yakni metode yang memungkinkan komputer mempelajari pola dari data dan menghasilkan prediksi tanpa perlu diprogram secara eksplisit untuk setiap tugas [3].

Dalam machine learning, klasifikasi merupakan salah satu teknik yang paling banyak diterapkan[4]. Algoritma Decision Tree menjadi salah satu metode klasifikasi yang populer karena memiliki kemampuan interpretasi yang tinggi serta struktur model yang intuitif dan mudah dipahami [5][6]. Algoritma ini telah diterapkan secara luas di berbagai bidang, seperti kesehatan, pendidikan, bisnis, dan transportasi [2],[7]. Berbagai penelitian telah membuktikan bahwa metode klasifikasi berbasis Decision Tree mampu menghasilkan prediksi dengan tingkat akurasi yang kompetitif.

Salah satu dataset yang paling populer dan sering digunakan dalam penelitian machine learning adalah Titanic dataset yang tersedia pada platform Kaggle [8]. Dataset ini memuat informasi penumpang kapal Titanic, meliputi PassengerId, Pclass, Sex, Age, SibSp, Parch, Fare, Embarked, dan status keselamatan penumpang (Survived). Dataset Titanic menjadi pilihan menarik karena memiliki label klasifikasi yang jelas, yakni

penumpang selamat (survived = 1) dan tidak selamat (survived = 0), sehingga sangat representatif untuk keperluan penelitian klasifikasi biner [8].

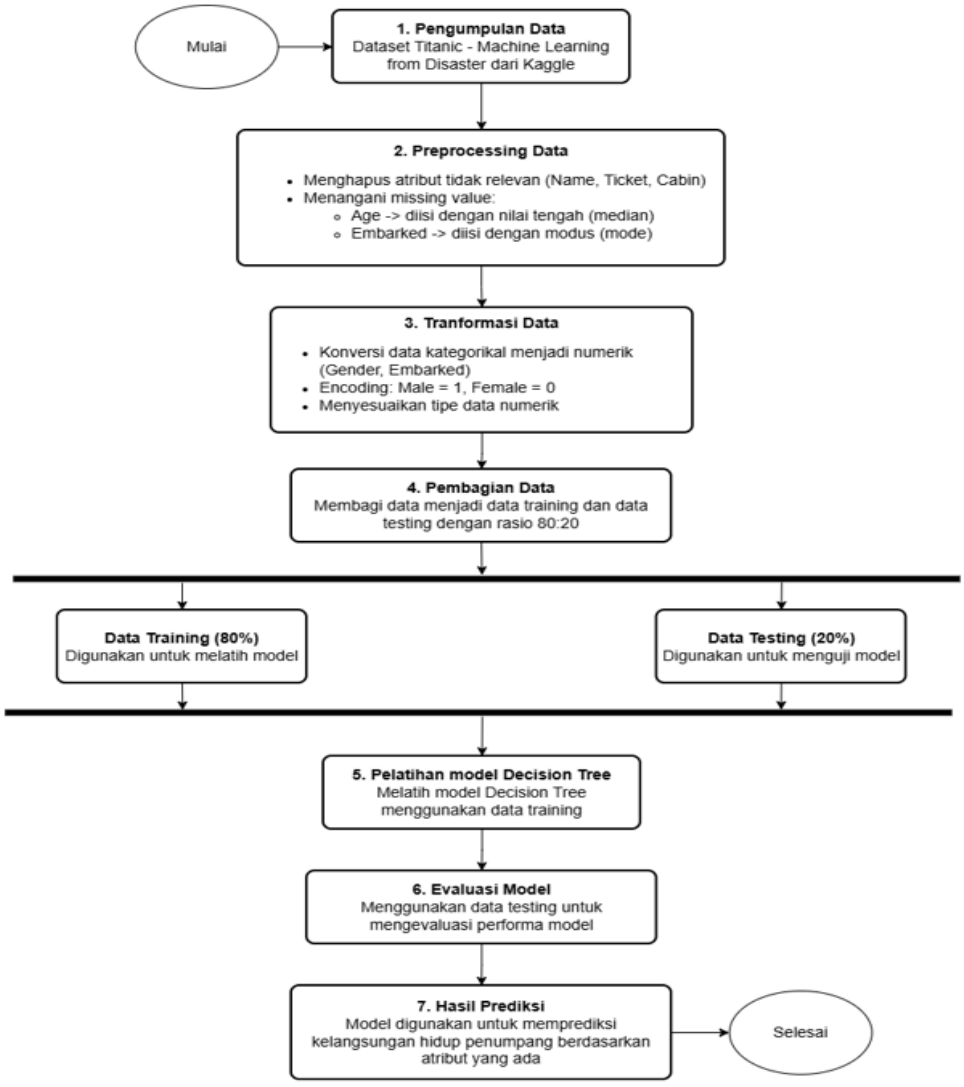
Penelitian mengenai klasifikasi menggunakan algoritma berbasis pohon keputusan telah banyak dilakukan sebelumnya [2]. Algoritma Decision Tree dikenal unggul dalam hal interpretabilitas karena menghasilkan struktur keputusan berbentuk pohon yang mudah dipahami. Di samping itu, algoritma ini mampu menangani data numerik maupun kategorikal sekaligus, dengan proses klasifikasi yang relatif efisien dan cepat [7], [9].

Berdasarkan uraian di atas, penelitian ini bertujuan menerapkan algoritma Decision Tree untuk memprediksi kelangsungan hidup penumpang Titanic menggunakan dataset dari Kaggle. Hasil penelitian ini diharapkan dapat memberikan gambaran yang komprehensif mengenai performa algoritma Decision Tree dalam menyelesaikan permasalahan klasifikasi, serta menjadi referensi bagi penelitian selanjutnya di bidang machine learning.

2. METODOLOGI

2.1 Jenis Penelitian

Penelitian ini menggunakan pendekatan kuantitatif dengan metode eksperimen [10] untuk komputasional untuk membangun model klasifikasi berbasis algoritma Decision Tree. Pendekatan kuantitatif dipilih karena penelitian berfokus pada pengolahan data numerik serta pengukuran performa model secara objektif berdasarkan metrik evaluasi yang terstandar.



Gambar 1. Flowchart Penelitian

2.2 Sumber Data

Data yang digunakan merupakan data sekunder yang bersumber dari dataset Titanic – Machine Learning from Disaster yang tersedia secara publik pada platform Kaggle [8]. Dataset tersebut memuat informasi penumpang yang meliputi PassengerId, Pclass, Gender, Age, SibSp, Parch, Fare, Embarked, serta label status keselamatan penumpang (Survived).

Dalam penelitian ini, dataset train dan test dari Kaggle digabungkan bersama file gender_submission.csv sebagai label kelas, kemudian diproses secara terpadu untuk membangun model klasifikasi yang mampu memprediksi probabilitas keselamatan penumpang.

Table 1. Contoh Dataset Titanic

ID	Survived	Pclass	Gender	Age	SibSp	Parch	Fare	Embarked	Family_Size
1	0	3	Male	22	1	0	7.25	S	2
2	1	1	Female	38	1	0	71.2833	C	2
3	1	3	Female	26	0	0	7.925	S	1
4	1	1	Female	35	1	0	53.1	S	2
5	0	3	Male	35	0	0	8.05	S	1
6	0	3	Male		0	0	8.4583	Q	1
7	0	1	Male	54	0	0	51.8625	S	1
8	0	3	Male	2	3	1	21.075	S	5
9	1	3	Female	27	0	2	11.1333	S	3
10	1	2	Female	14	1	0	30.0708	C	2
..	..	1	Female	..	..	..	..	..	..
..	..	3	Female	..	..	..	..	..	..
890	1	1	Male	26	0	0	30	C	1
891	0	3	male	32	0	0	7.75	Q	1

Keterangan:

- Survived = 0 menunjukkan penumpang tidak selamat
- Survived = 1 menunjukkan penumpang selamat

2.3 Tahapan Penelitian

Tahapan penelitian dilaksanakan secara sistematis, dimulai dari pengumpulan dataset Titanic melalui platform Kaggle. Selanjutnya, dilakukan praproses data yang mencakup penanganan missing value, penghapusan atribut yang tidak relevan, serta transformasi data kategorikal menjadi representasi numerik.

Dataset kemudian dibagi menjadi data latih dan data uji dengan rasio 80:20. Model klasifikasi dibangun menggunakan algoritma Decision Tree dengan bantuan library Scikit-learn pada bahasa pemrograman Python. Setelah proses pelatihan selesai, performa model dievaluasi menggunakan confusion matrix yang menghasilkan metrik accuracy, precision, recall, dan F1-score.

2.4 Algoritma yang Digunakan

Algoritma yang digunakan dalam penelitian ini adalah Decision Tree. Decision Tree merupakan salah satu algoritma supervised learning yang umum digunakan untuk tugas klasifikasi maupun prediksi [2]. Algoritma ini bekerja dengan membangun struktur pohon keputusan secara rekursif berdasarkan atribut-atribut data yang tersedia.

Pada struktur Decision Tree, setiap node internal merepresentasikan suatu atribut pengujian, setiap cabang menunjukkan hasil keputusan dari pengujian tersebut, dan setiap leaf node menunjukkan kelas hasil akhir klasifikasi. Proses pembentukan pohon dilakukan secara rekursif dengan memilih atribut terbaik berdasarkan nilai information gain tertinggi pada setiap tahap pemisahan.

Dua konsep utama yang mendasari cara kerja Decision Tree adalah entropy dan information gain. Entropy digunakan untuk mengukur tingkat ketidakpastian atau impuritas dalam suatu kumpulan data, sedangkan information gain digunakan untuk menentukan atribut yang paling efektif dalam memisahkan data ke dalam kelas-kelas yang lebih homogen.

Entropy mengukur tingkat ketidakpastian atau impuritas pada suatu himpunan data. Semakin kecil nilai entropy, semakin homogen kelas dalam data tersebut. Persamaan entropy ditunjukkan sebagai berikut:

$$\text{Entropy}(S) = -\sum_{i=1}^{n} p_i \log_2 p_i$$

Keterangan:

- S adalah himpunan data,
- p_i adalah probabilitas kemunculan kelas ke-i.

Nilai entropy bernilai 0 apabila seluruh data berada dalam satu kelas yang sama.

Information Gain mengukur pengurangan entropy setelah dataset dipisahkan berdasarkan suatu atribut tertentu. Atribut dengan nilai information gain tertinggi dipilih sebagai node pemisah pada setiap level pohon. Persamaan information gain ditunjukkan sebagai berikut:

$$\text{Gain}(S, A) = \text{Entropy}(S) - \sum_{v \in \text{Values}(A)} \frac{|S_v|}{|S|} \text{Entropy}(S_v)$$

Atribut dengan nilai gain tertinggi dipilih sebagai node pemisah utama, sehingga pemisahan data menghasilkan subset yang paling murni secara kelas.

Dalam penelitian ini, tujuh atribut digunakan sebagai input model, yaitu Pclass, Sex, Age, SibSp, Parch, Fare, dan Embarked. Model Decision Tree mempelajari pola hubungan antaratribut tersebut untuk menentukan probabilitas keselamatan seorang penumpang, dengan Sex, Pclass, dan Age sebagai faktor yang paling dominan.

Decision Tree dipilih karena memiliki sejumlah keunggulan, antara lain mudah diinterpretasikan, mampu menangani data numerik maupun kategorikal, serta dapat mengidentifikasi atribut yang paling berpengaruh terhadap hasil klasifikasi. Namun demikian, algoritma ini memiliki kelemahan berupa kecenderungan mengalami overfitting apabila struktur pohon terlalu kompleks dan dalam. Oleh karena itu, pengaturan parameter seperti max_depth dan min_samples_split menjadi faktor krusial dalam mengoptimalkan kemampuan generalisasi model.

2.5 Praproses Data

Pada tahap praproses, dilakukan pembersihan dan persiapan data untuk meningkatkan kualitas dataset sebelum digunakan dalam pelatihan model. Atribut PassengerId, Name, dan Ticket dihapus terlebih dahulu karena dianggap tidak relevan terhadap proses prediksi. Selanjutnya, atribut Cabin dihapus secara terpisah karena memiliki proporsi missing value yang sangat tinggi, yaitu lebih dari 25% dari total data.

Penanganan missing value dilakukan setelah penghapusan atribut tidak relevan. Nilai yang hilang pada seluruh atribut numerik diimputasi menggunakan nilai tengah (median) untuk meminimalkan pengaruh nilai ekstrem, sedangkan nilai yang hilang pada atribut kategorikal diisi menggunakan nilai modus (nilai yang paling sering muncul).

Sebelum encoding, dilakukan standardisasi pada seluruh atribut numerik menggunakan StandardScaler (metode Z-score) agar seluruh atribut berada pada skala

yang seragam. Selanjutnya, data kategorikal dikonversi menjadi representasi numerik menggunakan LabelEncoder dari library Scikit-learn, yang diterapkan pada seluruh atribut bertipe kategorikal, yaitu Sex dan Embarked. Setelah seluruh tahapan transformasi selesai, atribut yang digunakan dalam penelitian ini adalah Pclass, Sex, Age, SibSp, Parch, Fare, dan Embarked.

2.6 Praproses Data

Dataset dibagi menjadi dua subset, yaitu data training dan data testing dengan rasio 80:20. Sebanyak 80% data digunakan untuk melatih model, sedangkan 20% sisanya digunakan untuk menguji kemampuan generalisasi model.

Pembagian data ini bertujuan agar model dapat mempelajari pola dari data training dan dievaluasi secara objektif menggunakan data testing yang tidak pernah dilihat sebelumnya oleh model.

Table 2. Pembagian Data

Jenis Data	Persentase
Data Training	80%
Data Testing	20%

2.7 Pembangunan Model Decision Tree

Model klasifikasi dibangun menggunakan DecisionTreeClassifier dari library Scikit-learn dalam bahasa pemrograman Python. Algoritma ini membangun struktur pohon keputusan secara rekursif dengan memilih atribut yang memiliki nilai information gain tertinggi pada setiap node pemisahan.

Model dilatih menggunakan data training untuk mempelajari hubungan antara atribut-atribut penumpang dengan status keselamatannya. Hasil pelatihan berupa seperangkat aturan keputusan yang selanjutnya digunakan untuk mengklasifikasikan data testing.

2.8 Evaluasi Model

Evaluasi model dilakukan menggunakan confusion matrix untuk mengukur performa klasifikasi model Decision Tree secara komprehensif. Berdasarkan confusion matrix, dihitung empat metrik evaluasi utama, yaitu accuracy, precision, recall, dan F1-score.

Accuracy mengukur proporsi prediksi yang benar dari keseluruhan data uji. Precision mengukur ketepatan prediksi positif model, yakni seberapa besar prediksi selamat yang memang benar-benar selamat. Recall mengukur kemampuan model dalam mendeteksi seluruh sampel kelas positif yang sesungguhnya. Adapun F1-score merupakan rata-rata harmonik antara precision dan recall, yang memberikan gambaran keseimbangan kinerja model secara menyeluruh.

3. HASIL DAN PEMBAHASAN

3.1 Hasil Preprocessing Data

Tahap praproses memiliki peranan krusial dalam meningkatkan kualitas data sebelum proses pelatihan model dilakukan. Tahap pertama adalah penghapusan atribut yang tidak relevan, yaitu PassengerId, Name, dan Ticket. Kemudian, atribut Cabin dihapus secara terpisah karena memiliki proporsi missing value yang sangat tinggi (lebih dari 25% dari total data), sehingga imputasi tidak disarankan.

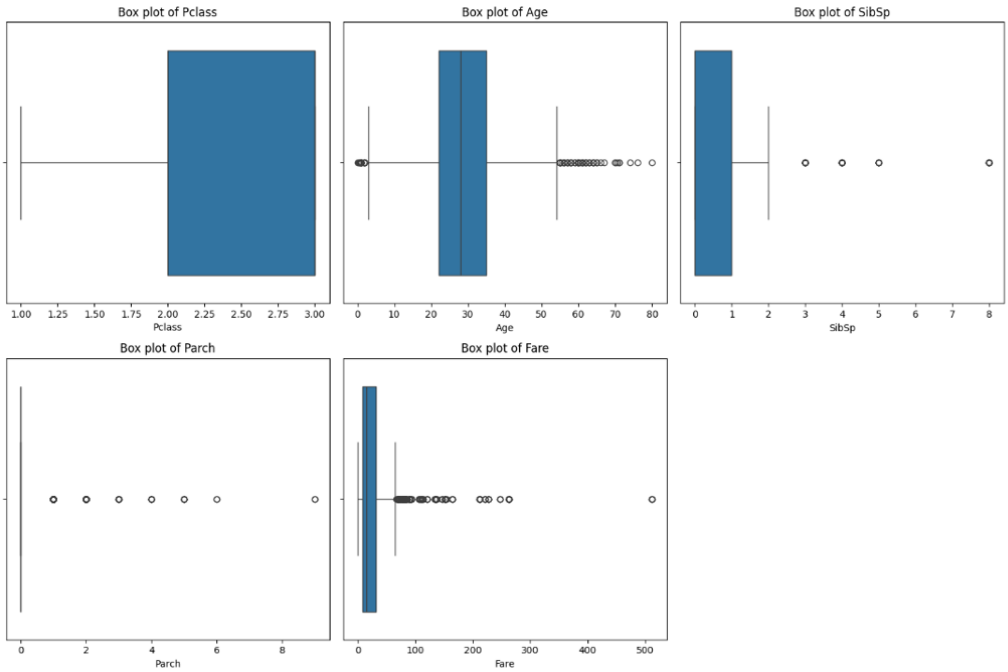
Selanjutnya, penanganan missing value dilakukan pada dua atribut utama: atribut Age diimputasi menggunakan nilai tengah (median) untuk menjaga ketahanan terhadap

outlier, sedangkan atribut Embarked diisi menggunakan nilai yang paling sering muncul (modus).

Standardisasi data numerik dilakukan menggunakan metode Z-score (StandardScaler) sebelum proses encoding. Selanjutnya, transformasi data kategorikal menjadi representasi numerik dilakukan menggunakan LabelEncoder, yang diterapkan pada seluruh atribut kategorikal (Sex dan Embarked) secara otomatis.

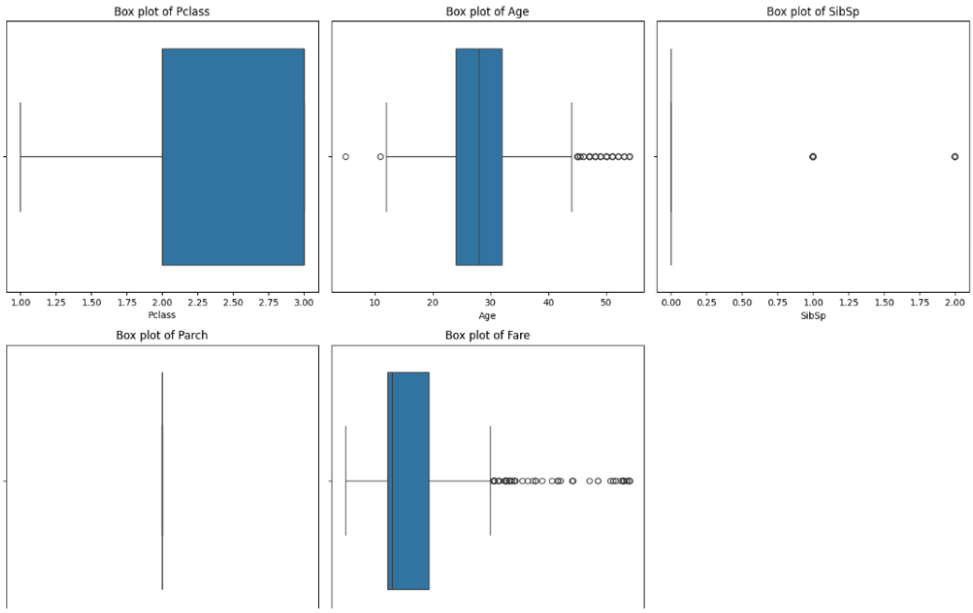
Setelah seluruh tahapan praproses selesai dilaksanakan, dataset berada dalam kondisi yang bersih, konsisten, dan siap digunakan untuk tahap pelatihan serta pengujian model.

Sebelum penghapusan outlier, visualisasi box plot menunjukkan keberadaan sejumlah nilai ekstrem pada beberapa atribut. Atribut Age memiliki beberapa outlier di sisi kanan distribusi yang mengindikasikan adanya penumpang berusia sangat tua. Atribut SibSp dan Fare juga menunjukkan outlier yang jauh melampaui batas whisker atas, khususnya Fare yang mencapai nilai lebih dari 500. Kondisi ini mengindikasikan perlunya penanganan outlier agar performa model tidak terdegradasi akibat pengaruh nilai-nilai ekstrem tersebut.



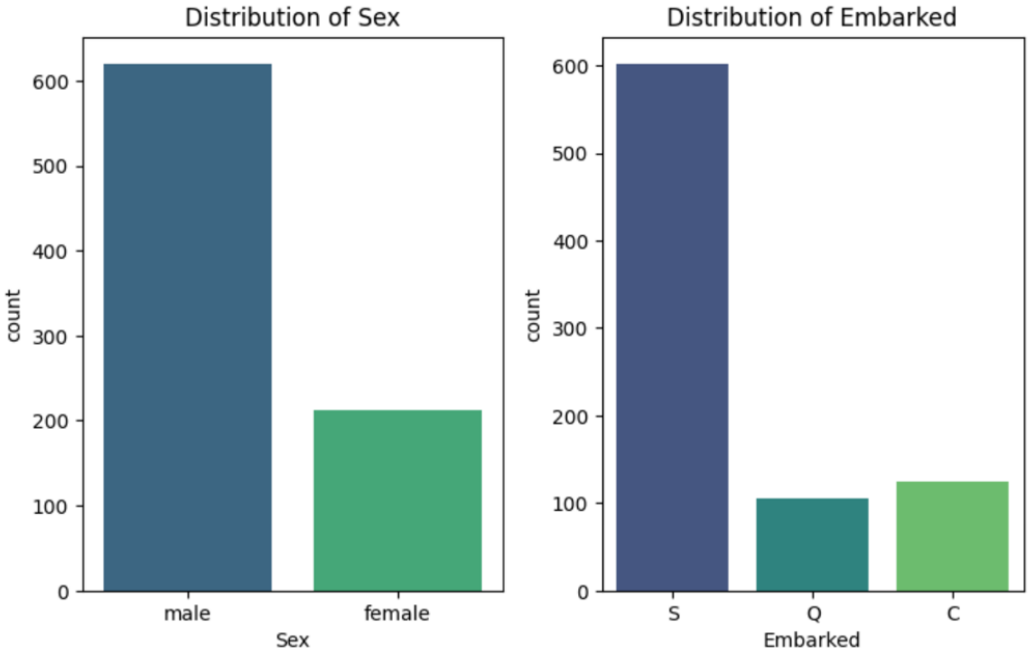
Gambar 2. Box Plot Atribut Numerik Sebelum Penghapusan Outlier

Setelah penghapusan outlier menggunakan metode Interquartile Range (IQR), distribusi data pada setiap atribut menjadi lebih terpusat dan representatif. Rentang nilai pada atribut Age, SibSp, dan Fare tampak jauh lebih sempit dibandingkan sebelumnya. Atribut Fare yang sebelumnya mencapai nilai ekstrem di atas 500 kini dibatasi hingga sekitar 55, sehingga distribusi data lebih seimbang dan tidak lagi bias terhadap nilai-nilai ekstrem.



Gambar 3. Box Plot Atribut Numerik Setelah Penghapusan Outlier

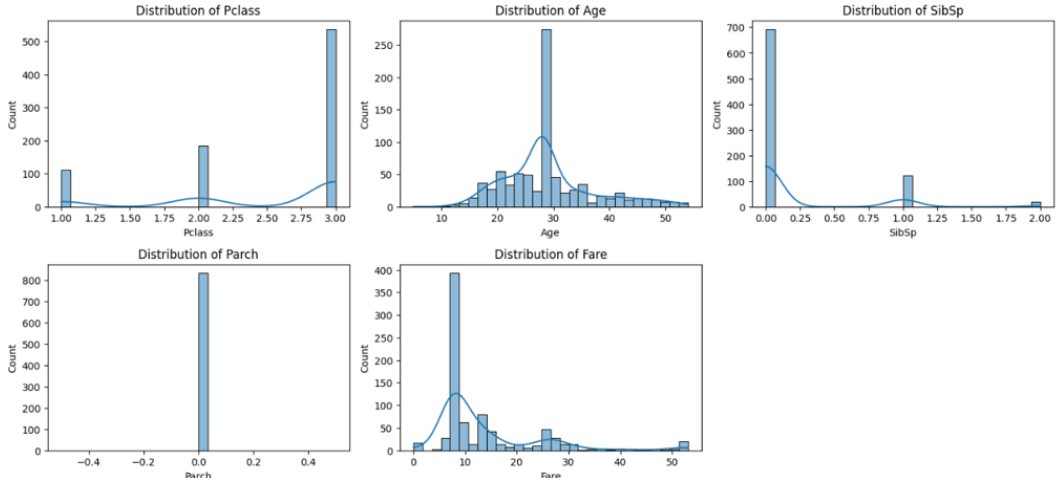
Visualisasi distribusi atribut kategorikal menunjukkan bahwa penumpang berjenis kelamin laki-laki (male) mendominasi dengan jumlah sekitar 620 penumpang, dibandingkan perempuan (female) yang berjumlah sekitar 215 penumpang setelah penghapusan outlier. Sementara itu, distribusi atribut Embarked menunjukkan bahwa mayoritas penumpang naik dari Southampton (S) dengan jumlah lebih dari 600 penumpang, diikuti oleh Cherbourg (C) sekitar 130 penumpang, dan Queenstown (Q) sekitar 110 penumpang. Ketidakseimbangan distribusi pada kedua atribut kategorikal ini perlu diperhatikan dalam proses analisis dan interpretasi hasil model.



Gambar 4. Distribusi Atribut

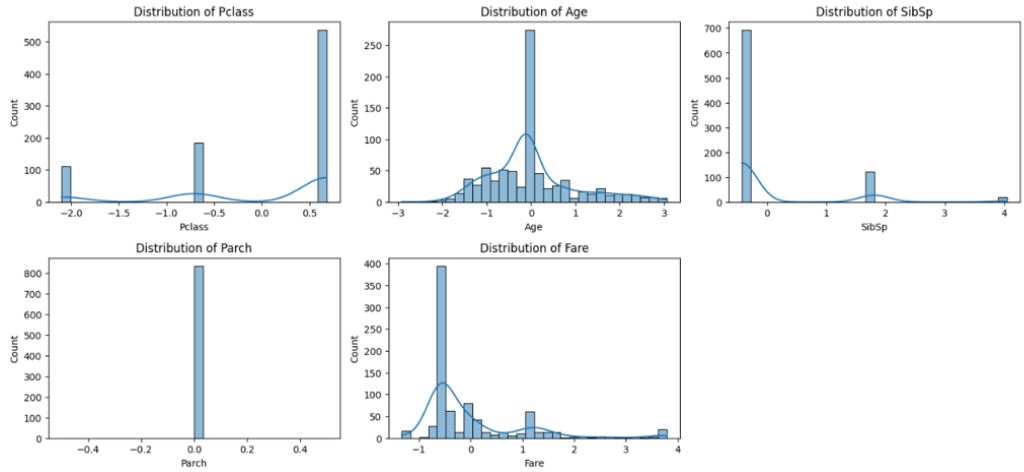
Sebelum standardisasi dilakukan, atribut-atribut numerik seperti Pclass, Age, SibSp, Parch, dan Fare berada pada skala yang sangat berbeda satu sama lain. Atribut Age memiliki rentang nilai sekitar 0 hingga 55, sedangkan Fare juga berkisar antara 0 hingga 55 setelah penghapusan outlier. Perbedaan skala antaratribut ini berpotensi

memengaruhi performa model secara negatif, sehingga proses standardisasi diperlukan untuk menyeragamkan skala seluruh atribut numerik.



Gambar 5. Distribusi Atribut Numerik Sebelum Standarisasi

Setelah dilakukan standardisasi menggunakan metode Z-score, seluruh atribut numerik telah berada pada skala yang seragam dengan rata-rata mendekati 0 dan standar deviasi 1. Distribusi atribut Age tampak mendekati distribusi normal dengan puncak di sekitar nilai 0. Atribut Pclass, SibSp, Parch, dan Fare juga menunjukkan distribusi yang lebih terstandarisasi. Proses standardisasi ini memastikan bahwa tidak ada atribut yang mendominasi proses pembelajaran model akibat perbedaan skala nilai.



Gambar 6. Distribusi Atribut Numerik Setelah Standarisasi

Hasil preprocessing juga menunjukkan bahwa dataset memiliki distribusi kelas yang tidak seimbang, di mana jumlah penumpang yang tidak selamat lebih banyak dibandingkan penumpang yang selamat. Kondisi ini berpotensi memengaruhi performa model, khususnya pada kemampuan model dalam mengenali kelas minoritas.

Gambar 7. Pie Chart Penumpang Selamat vs Tidak Selamat

3.2 Hasil Pembagian Data

Setelah tahap praproses selesai, dataset dibagi menjadi dua subset dengan rasio 80:20. Data training digunakan untuk melatih model Decision Tree agar dapat mempelajari pola hubungan antaratribut, sedangkan data testing digunakan untuk mengevaluasi kemampuan generalisasi model terhadap data yang belum pernah dilihat sebelumnya.

Table 3. Hasil Pembagian Data

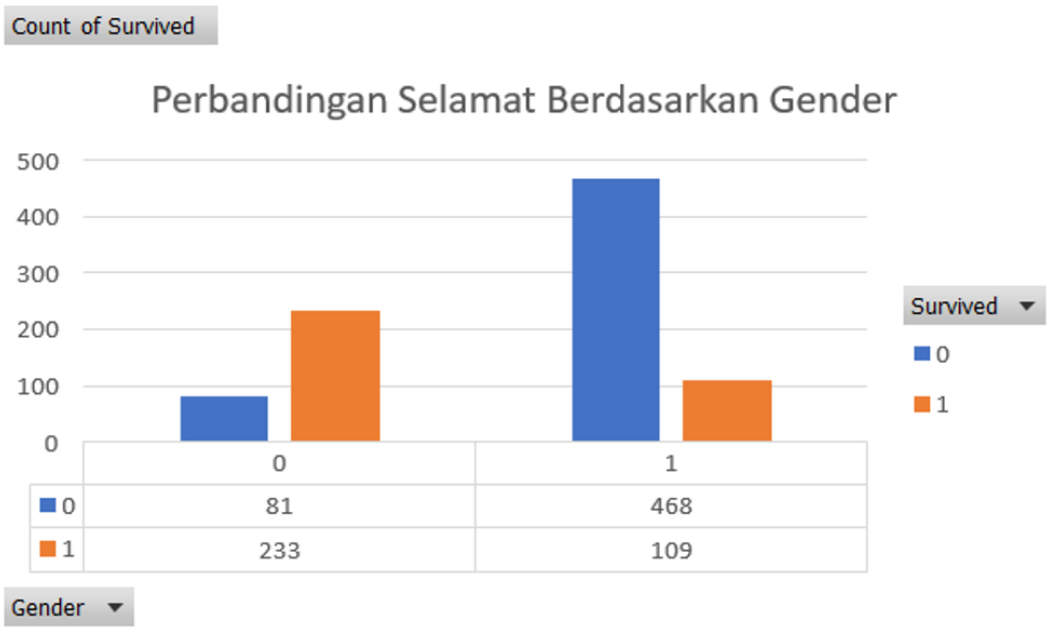
Jenis Data	Jumlah Data	Persentase
Data Training	718	80%
Data Testing	173	20%

Strategi pembagian ini dilakukan untuk mencegah overfitting sekaligus memastikan bahwa evaluasi performa model dilakukan secara objektif dan representatif.

3.3 Hasil Pelatihan Model Decision Tree

Model klasifikasi dibangun menggunakan algoritma Decision Tree dengan memanfaatkan tujuh atribut input, yaitu Pclass, Sex, Age, SibSp, Parch, Fare, dan Embarked. Algoritma membangun struktur pohon keputusan secara rekursif dengan memilih atribut yang memiliki nilai information gain tertinggi pada setiap node pemisahan.

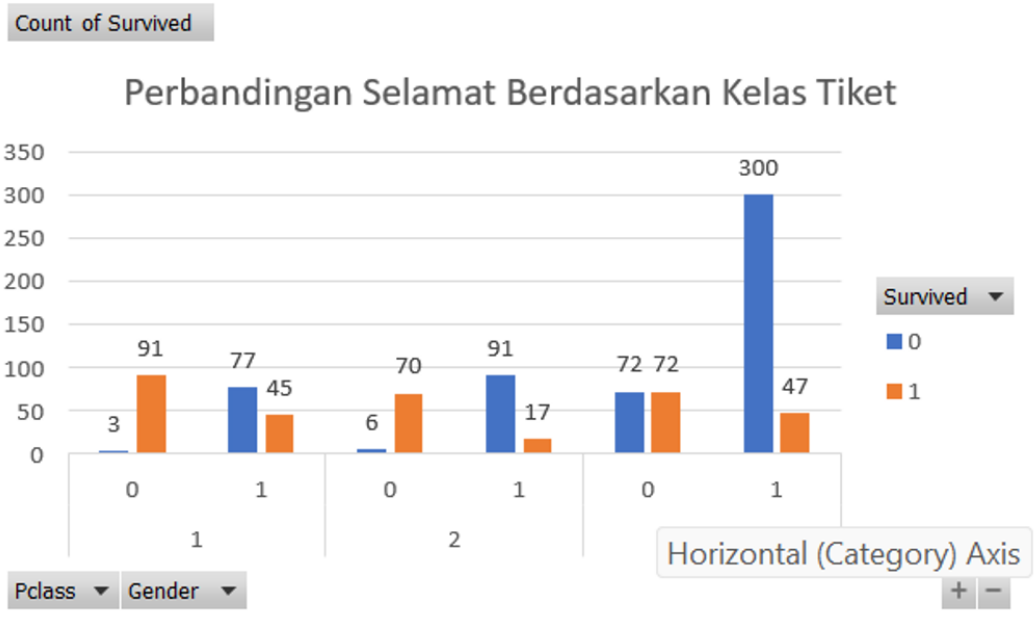
Hasil pelatihan menunjukkan bahwa tiga atribut yang paling berpengaruh terhadap prediksi keselamatan penumpang adalah Gender, Pclass, dan Age. Dalam struktur pohon keputusan yang terbentuk, atribut Gender menjadi simpul utama (root node) yang digunakan model sebagai kriteria pemisahan pertama dalam menentukan klasifikasi keselamatan penumpang.



Gambar 8. Diagram Batang Penumpang Selamat vs Tidak Selamat

Berdasarkan hasil pelatihan, atribut Gender terbukti menjadi faktor yang paling dominan dalam proses klasifikasi keselamatan penumpang Titanic. Temuan ini konsisten dengan data historis yang menunjukkan bahwa gender memiliki pengaruh signifikan terhadap peluang keselamatan penumpang.

Tingginya tingkat keselamatan penumpang perempuan selaras dengan kondisi historis tragedi Titanic, di mana protokol evakuasi secara eksplisit memprioritaskan perempuan dan anak-anak di atas laki-laki dewasa. Model Decision Tree berhasil mengenali dan mengabstraksikan pola ini sebagai aturan keputusan yang dominan dalam proses klasifikasi.



Gambar 9. Pie Chart Penumpang Selamat vs Tidak Selamat

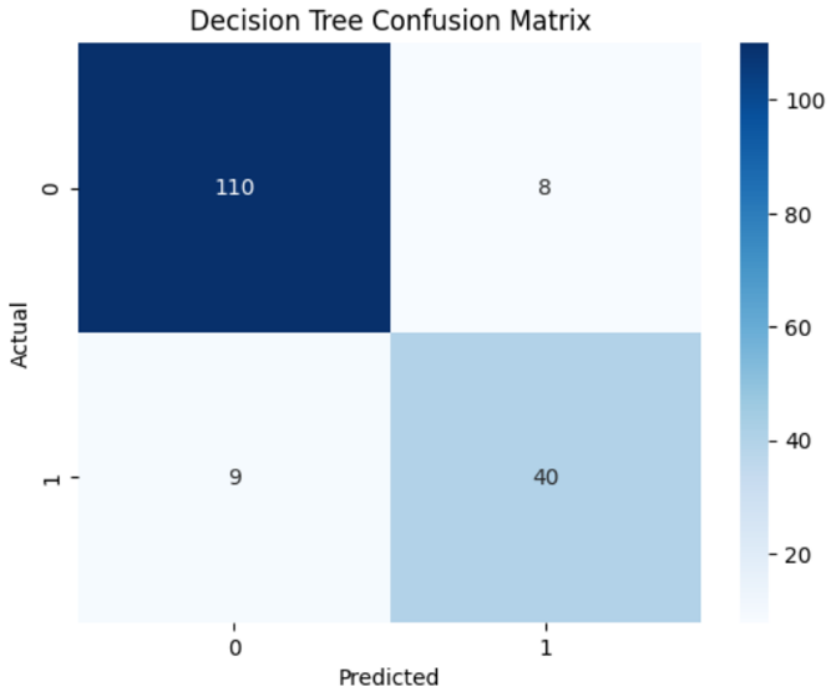
Selain Gender, atribut Pclass juga memberikan kontribusi yang signifikan terhadap hasil prediksi. Penumpang kelas 1 memiliki peluang keselamatan yang jauh lebih tinggi dibandingkan penumpang kelas 2 dan kelas 3. Hal ini disebabkan oleh lokasi kabin yang

lebih dekat dengan geladak atas, kemudahan akses ke sekoci penyelamat, serta prioritas penyelamatan yang lebih baik bagi penumpang kelas atas.

Secara keseluruhan, hasil ini mengonfirmasi bahwa algoritma Decision Tree mampu mempelajari hubungan antaratribut secara efektif, sehingga menghasilkan pola klasifikasi yang relevan dan sesuai dengan kondisi data historis yang sesungguhnya.

3.4 Evaluasi Model

Evaluasi model dilakukan menggunakan confusion matrix beserta empat metrik evaluasi turunannya, yaitu accuracy, precision, recall, dan F1-score. Penggunaan metrik-metrik ini bertujuan untuk memperoleh gambaran yang komprehensif mengenai performa model dalam melakukan klasifikasi keselamatan penumpang Titanic.



Gambar 10. Confusion Matrix

Table 4. Hasil Pembagian Data

Metrik Evaluasi	Nilai
Accuracy	90%
Precision	83%
Recall	82%
F1-score	82%

Berdasarkan hasil evaluasi, model Decision Tree memperoleh accuracy sebesar 90%, yang menunjukkan bahwa model mampu mengklasifikasikan status keselamatan penumpang dengan tingkat ketepatan yang tinggi.

Nilai precision sebesar 83% menunjukkan bahwa sebagian besar penumpang yang diprediksi selamat oleh model memang benar-benar selamat, sehingga tingkat false positive relatif rendah. Sementara itu, nilai recall sebesar 82% menunjukkan bahwa model mampu mendeteksi sebagian besar penumpang yang sesungguhnya selamat, dengan tingkat false negative yang terkendali.

Nilai F1-score sebesar 82% mengindikasikan bahwa model memiliki keseimbangan performa yang baik antara precision dan recall. Hal ini menunjukkan bahwa model tidak hanya menghasilkan prediksi yang akurat, tetapi juga konsisten dalam mengidentifikasi kedua kelas selamat dan tidak selamat secara berimbang.

Untuk meningkatkan performa model, dilakukan proses hyperparameter tuning menggunakan RandomizedSearchCV dengan pencarian pada kombinasi parameter max_depth, min_samples_split, min_samples_leaf, dan criterion (gini/entropy). Proses ini menggunakan 3-fold cross-validation dengan 20 iterasi pencarian acak. Hasilnya adalah model yang lebih optimal dengan kemampuan generalisasi yang lebih baik terhadap data yang belum pernah dilihat sebelumnya.

Selain itu, ketidakseimbangan distribusi kelas pada dataset Titanic juga berpotensi memengaruhi hasil klasifikasi. Dalam kondisi ini, model cenderung lebih dominan mempelajari pola kelas mayoritas (tidak selamat) dibandingkan kelas minoritas (selamat), sehingga teknik penyeimbangan data seperti oversampling atau class weighting dapat dipertimbangkan pada penelitian selanjutnya.

4. KESIMPULAN DAN SARAN

Berdasarkan hasil penelitian, algoritma Decision Tree mampu digunakan untuk memprediksi kelangsungan hidup penumpang Titanic dengan tingkat akurasi yang cukup baik. Faktor yang paling berpengaruh terhadap hasil prediksi adalah gender, pclass, dan age. Hasil evaluasi menunjukkan bahwa model memiliki performa klasifikasi yang efektif dalam mengidentifikasi penumpang yang selamat maupun tidak selamat.

Penelitian ini juga menerapkan hyperparameter tuning menggunakan RandomizedSearchCV untuk mengoptimalkan parameter max_depth, min_samples_split, min_samples_leaf, dan criterion, sehingga diperoleh model yang lebih tahan terhadap overfitting. Sebagai rekomendasi untuk penelitian selanjutnya, perlu dilakukan perbandingan dengan algoritma lain seperti Random Forest, Support Vector Machine, atau Logistic Regression. Selain itu, penerapan teknik penyeimbangan kelas seperti SMOTE dapat menjadi langkah lanjutan untuk meningkatkan kemampuan model dalam mengenali kelas minoritas.

5. DAFTAR RUJUKAN

[1] S.V. Hedi , “Interactive NPC for Interactive Learning Media Of Indonesian Traditional Music,” Journal of Creative Industries (MERAKI), vol. 1, no. 2, pp 12-23, June. 2024, doi: https://doi.org/10.24123/meraki.v1i02

[2] D. Rahajoe, M. R. A. Subekti, dan M. Suriansyah, “Real-Time Drowsy Face Detection for Online Learning Based on Random Forest and Decision Tree Algorithms,” Jurnal Teknik Informatika (JUTIF), vol. 4, no. 6, pp. 1549–1554, Dec. 2023, doi: 10.52436/1.jutif.2023.4.6.1554.

[3] Barbierato, E., dan Gatti, Alice, “The Challenges of Machine Learning: A Critical Review,” electronic MDPI, vol. 13, no. 2, pp. 1-30, January.2024, doi: https://doi.org/10.3390/electronics13020416

[4] Alnuaimi. A. F. A. H, Albaldawi. T. H. K. , “An overview of machine learning classification techniques,” Fifth International Scientific Conference of Alkafeel University, vol. 97, pp 45-62, doi: https://doi.org/10.1051/bioconf/20249700133

[5] Lee. Jeongeon, Sim. Min Kyu, Hong. Jung Sik, “Assessing Decision Tree Stability: A Comprehensive Method for Generating a Stable Decision Tree,” IEEE Access, vol. 12, pp 90061 – 90072, June. 2024, doi: 10.1109/ACCESS.2024.3419228

[6] A. A. Hozan, M. A. Adnan, “A Review on Decision Tree Algorithm in Healthcare Applications,” The Indonesian Journal of Computer Science, vol. 13, no. 3, pp 3863 – 3881, June. 2024, doi: https://doi.org/10.33022/ijcs.v13i3.4026

[7] M. D. Alifiansyah, A. D. Rahajoe, dan E. Y. Puspaningrum, “Klasifikasi Kualitas Udara menggunakan Algoritma LightGBM berdasarkan ISPU,” Prosiding SNATIF Ke-8 Tahun 2025, Fakultas Teknik Universitas Muria Kudus, 2025.

[8] Kaggle, “Titanic - Machine learning from Disaster,” Available: https://www.Kaggle.com/c/Titanic, accessed: 2026.

[9] M. D. Alifiansyah, A. D. Rahajoe, dan E. Y. Puspaningrum, “Implementation of PSO Optimization on the LightGBM Algorithm for Air Pollution Classification,” Bit-Tech (Binary Digital - Technology), vol. 8, no. 2, Dec. 2025, doi: 10.32877/bt.v8i2.3243.

[10] O. Michael, K. A. Putra, P. Bayu, “Studi Empiris Penerapan Traditional Design pada Antarmuka Web Profil Perusahaan PT Benteng Celebes dengan Metode Eksperimen,” Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer, vol. 9, no. 6, 1-9, Juni. 2025, doi: https://j-ptiik.ub.ac.id/index.php/j-ptiik/article/download/14929/6641