

**ANALISIS DAN RANCANG BANGUN SISTEM PREDIKSI KEBERTAHANAN HIDUP  
MENGUNAKAN ALGORITMA SUPPORT VECTOR MACHINE (SVM) DAN DATASET TITANIC**

**ANALYSIS AND DESIGN OF A SURVIVAL PREDICTION SYSTEM USING THE SUPPORT VECTOR  
MACHINE (SVM) ALGORITHM AND TITANIC DATASET**

**Julianti Yakin<sup>1</sup>, Aryo Nur Utomo<sup>2</sup>**

<sup>1,2</sup>Program Studi Teknik Informatika, Fakultas Sains dan Teknologi Informasi  
Institut Sains dan Teknologi Nasional

[juliantiyakin@gmail.com](mailto:juliantiyakin@gmail.com)<sup>1</sup>, [aryo.nurutomo@istn.ac.id](mailto:aryo.nurutomo@istn.ac.id)<sup>2</sup>

**ABSTRAKS**

Pada 15 April 2022, satu abad telah berlalu sejak kapal Titanic tenggelam. Bencana kapal Titanic telah membangkitkan spekulasi serta minat publik dan ilmuwan terhadap kecelakaan kapal. Pada penelitian ini bertujuan untuk mengetahui tingkat kebertahanan hidup penumpang kapal dengan menggunakan algoritma Support Vector Machine. SVM dijadikan sebagai model dengan membagi dataset yang digunakan menjadi data training sebanyak 80% dan data testing sebanyak 20% yang kemudian akan di hitung tingkat akurasi model menggunakan metode Holdout yang menghasilkan 81.12% dan K-Fold Cross Validation yang kemudian rata-rata hasil yang didapatkan sebesar 81.03%. Hasil dari penelitian ini merupakan sistem prediksi sederhana kebertahanan hidup penumpang kapal.

**Kata Kunci :** Titanic, Support Vector Machine, Metode Hold Out, K-Fold Cross Validation, Prediksi Kebertahanan Hidup.

**ABSTRACT**

*A century has passed since the sinking of Titanic on April 15, 2022. The disaster has sparked speculation and public and scientific interest in the shipwreck. This study aims to determine the survival rate of Titanic passengers using the Support Vector Machine algorithm. SVM is used as the model by dividing the dataset into 80% training data and 20% testing data, and then calculating the model's accuracy rate using the Holdout method, which produces 81.12%, and K-Fold Cross Validation, which produces an average score of 81.03%. The result of this study is a simple survival prediction system for ship passengers.*

**Keywords :** Titanic, Support Vector Machine, Hold Out Method, K-Fold Cross Validation, Survival Prediction

**1. PENDAHULUAN**

Pada 15 April 2022, satu abad telah berlalu sejak Royal Mail Ship Titanic tenggelam di Samudera Atlantik Utara. Adanya banyak spekulasi dalam literatur tentang kecelakaan kapal Titanic, dan penelitian tentang kecelakaan tersebut masih berlangsung.

Dataset Titanic merupakan kumpulan data yang berisikan data penumpang yang selamat dan tidak selamat, dataset ini akan digunakan untuk mengkarakterisasi dengan permodelan dalam machine learning. Tugas dari machine learning adalah untuk memprediksi apakah penumpang akan selamat pada tenggelam nya kapal dengan menggunakan model machine learning, dengan melihat beberapa faktor yang mengakibatkan keselamatan dari penumpang itu sendiri dalam tenggelamnya kapal.

Pada penelitian ini bertujuan untuk mengetahui tingkat kebertahanan hidup

penumpang kapal dengan menggunakan algoritma Support Vector Machine. Algoritma Support Vector Machine (SVM) dapat mengklasifikasikan data secara linier dan non-linier dengan mencari hyperplane yang digunakan untuk jarak maksimal antar kelas data.

Pada penelitian sebelum nya yang dilakukan oleh (Wafa et al., 2022) dengan judul Prediksi Penyakit Diabetes Menggunakan Algoritma Support Vector Machine (SVM). Dari penelitian yang dilakukan didapatkan hasil yaitu tingkat akurasi yang dihasilkan oleh performa yang didapatkan dari metode Support Vector Machine dengan kernel radial basis function dan forward selection memperoleh hasil sebesar 91.2% untuk accuracy, 93.0% untuk precision, 94.3% untuk recall, dan 93.7% untuk f1-score.

Lalu penelitian yang dilakukan oleh (Haque et al., 2021) melakukan penelitian dengan judul Passenger data analysis of

Titanic using machine learning approach in the context of chances of surviving the disaster mendapatkan hasil yaitu teknik klasifikasi dan pengelompokan digunakan untuk menganalisis data tersebut, dan ditemukan bahwa penggunaan pohon keputusan memiliki akurasi sebesar 85% dalam menganalisis data tersebut. Dari penelitian sebelumnya maka dilakukan penelitian untuk mengetahui tingkat akurasi algoritma Support Vector Machine pada dataset Titanic dan kemudian membuat aplikasi prediksi sederhana mengenai keberuntungan hidup penumpang kapal.

## 2. METODOLOGI PENELITIAN

Agar proses penelitian berjalan dengan terstruktur dan rapih maka dibuatlah tahapan penelitian. Adapun tahapan-tahapan yang dilakukan dalam penelitian ini adalah studi literatur, klasifikasi, dan pembuatan laporan.

### Studi Literatur

Tahap studi literatur merupakan tahap penting dalam penelitian karena membantu dalam memahami konsep dan teori yang berkaitan dengan masalah yang akan diteliti. Pada tahap ini juga melakukan observasi terhadap penelitian yang sebelumnya telah dilakukan seperti jurnal penelitian hingga karya ilmiah akademis. Tujuannya untuk memperoleh gambaran umum tentang topik penelitian, menemukan keterkaitan antar studi dan menentukan arah penelitian selanjutnya yang kemudian akan dituangkan ke dalam bab 2 yang nantinya akan menjadi landasan teori dari penelitian ini.

### Klasifikasi

Tahap klasifikasi merupakan tahap dimana diaplikasikannya teori yang diperoleh dari tahap studi literatur ke dalam teknik machine learning dan mengimplementasikannya. Tujuannya untuk memperoleh hasil klasifikasi yang dapat menjawab rumusan masalah penelitian dan membantu menentukan kevalidan hipotesis.

- Data Loading : Data loading merupakan tahap pertama dalam proses klasifikasi, data loading ialah proses memuat data dari sumber yang berbeda ke dalam sistem untuk diolah dan dianalisis.
- Data understanding : Data understanding adalah tahap memahami data yang sudah di-load, meliputi mengetahui jumlah observasi, tipe data, statistik sederhana, mengecek data yang hilang, hingga menghapus missing value pada kolom yang tidak relevan.

- Exploratory Data Analysis (EDA) : Exploratory Data Analysis (EDA) adalah tahap mengeksplorasi dan memahami data secara lebih mendalam melalui visualisasi dan analisis statistic tujuannya untuk menemukan pola, hubungan, dan anomali dalam data. Pada proses EDA juga terdapat proses data encoding yaitu proses mengubah data kategorikal menjadi data numerik, Tujuannya adalah agar dapat diproses oleh algoritma machine learning.
- Model Implementation : Model implementation adalah tahap di mana model machine learning dipilih dan diterapkan pada data yang sudah disiapkan sebelumnya. Dalam hal ini, digunakan model machine learning Support Vector Machine. Pada tahap ini, dilakukan nya pelatihan model dengan data training dan mengevaluasi performa model dengan data testing.
- Model Validation : Metode validasi model pada penelitian ini adalah metode Holdout dan metode K-Fold Cross Validation. Tujuannya adalah untuk mengatasi masalah overfitting dan memperoleh suatu estimasi yang lebih akurat dari kinerja model.
- GUI Prediction Testing : GUI Prediction Testing adalah tahap di mana akan dilakukan pengujian hasil klasifikasi model machine learning yang telah dipilih. Dalam hal ini, dibuatlah antarmuka pengguna (GUI) untuk mengimplementasikan model machine learning dan melakukan tes prediksi pada data baru yang tidak termasuk dalam data training dan data testing. Tujuannya untuk memastikan bahwa model machine learning dapat bekerja dengan baik dan dapat digunakan oleh pengguna untuk melakukan klasifikasi data jaringan secara real-time. Pada tahap ini, dipastikan bahwa GUI mudah digunakan dan hasil klasifikasi dari model machine learning sesuai dengan harapan.

### Pembuatan Laporan

Pembuatan Laporan adalah tahap terakhir dalam proses penelitian. Dalam tahap ini, dilakukan penyusunan laporan yang berisi seluruh teori, hasil pengujian, dan kesimpulan yang telah didapatkan selama proses penelitian. Laporan tersebut menjadi bukti bahwa penelitian telah selesai dan dapat digunakan sebagai sumber informasi bagi penelitian selanjutnya. Tujuannya untuk memenuhi syarat program sarjana dan memberikan manfaat bagi pengembangan ilmu pengetahuan.

### 3. HASIL DAN PEMBAHASAN

Pembahasan berikut adalah uraian hasil analisis data dan bangun sistem prediksi keberuntungan hidup pada penumpang kapal *Titanic*.

#### Hasil Data Loading

Hasil dari data loading yang didapatkan bahwa *dataset Titanic* terdapat 891 baris data dan 12 kolom. 12 kolom yang terdapat pada *dataset Titanic* adalah kolom *PassengerId*, *Survived*, *Pclass*, *Name*, *Sex*, *Age*, *Sibsp*, *Parch*, *Ticket*, *Fare*, *Cabin* dan *Embarked*. Dapat dilihat pada Tabel Hasil Dari Proses Menampilkan Data.

Tabel Hasil Dari Proses Menampilkan Data

PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked	
476	477	0	2	Renouf, Mr. Peter Henry	male	34.0000	1	0	53027	21.0000	'0A'	S
638	639	0	3	Planitz, Mrs. Julia Maria Emilia (Johal)	female	41.0000	0	5	330225	29.6875	'0A'	S
536	537	0	1	Butt, Major. Archibald William	male	45.0000	0	0	112050	26.5500	B38	S
565	566	0	3	Davies, Mr. Alfred J	male	24.0000	2	0	A/4 48871	24.1500	'0A'	S
136	137	1	1	Newson, Miss. Helen Marypeny	female	33.0000	0	2	11752	26.2833	D47	S
124	125	0	2	Leyson, Mr. Robert William Norman	male	34.0000	0	0	C.A. 39546	10.5000	'0A'	S
820	821	1	1	Hays, Mrs. Charles Melville (Cara Jennings Gregg)	female	52.0000	1	1	12749	93.5000	B69	S
397	398	0	2	Nicholas, Mr. Peter David	male	46.0000	0	0	28403	26.0000	'0A'	S
532	534	1	3	Peter, Mrs. Catherine (Catherine Rick)	female	'0A'	0	2	2968	22.3500	'0A'	C
26	27	0	3	Emir, Mr. Fared Dahab	male	'0A'	0	0	2631	7.2250	'0A'	C

#### Hasil Data Understanding

Hasil dari tahap *data understanding* yaitu menampilkan informasi *dataset*, *missing value* dan deskripsi statistik *dataset*. Bahwa *dataset Titanic* memiliki 891 baris data dan 12 kolom. 2 fiturnya adalah float, 5 fiturnya adalah bilangan bulat atau int dan 5 lainnya adalah objek. Dapat dilihat pada Tabel Hasil Informasi Dataset.

Tabel Hasil Informasi Dataset

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 891 entries, 0 to 890
Data columns (total 12 columns):
#   Column      Non-Null Count  Dtype
---  -
0   PassengerId  891 non-null   int64
1   Survived     891 non-null   int64
2   Pclass       891 non-null   int64
3   Name         891 non-null   object
4   Sex          891 non-null   object
5   Age          714 non-null   float64
6   SibSp        891 non-null   int64
7   Parch        891 non-null   int64
8   Ticket       891 non-null   object
9   Fare         891 non-null   float64
10  Cabin        204 non-null   object
11  Embarked     889 non-null   object
dtypes: float64(2), int64(5), object(5)
```

Hasil dari *missing value* adalah *Dataset Titanic* terdapat *missing value* sebanyak 177 pada kolom *Age*, 687 pada kolom *Cabin* dan 2 pada kolom *Embarked*. Dapat dilihat pada Tabel Hasil Missing value.

Tabel Hasil Missing value

```
PassengerId    0
Survived        0
Pclass          0
Name            0
Sex             0
Age             177
SibSp           0
Parch           0
Ticket          0
Fare            0
Cabin           687
Embarked        2
dtype: int64
```

Hasil dari deskripsi statistik *dataset* menunjukkan nilai count, mean, standard deviation, minimum, quantile 25-75, dan maximum. Dapat dilihat pada Tabel Hasil Deskripsi Statistik Dataset.

Tabel Hasil Deskripsi Statistik Dataset

PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
count	891.000000	891.000000	891.000000	891	891	714.000000	891.000000	891	891.000000	204	889
unique	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
top	NaN	NaN	NaN	Brandt, Mr. Owen Harris	male	NaN	NaN	NaN	34.7000	NaN	B96 B106
freq	NaN	NaN	NaN	1	237	NaN	NaN	NaN	7	NaN	4
mean	480.000000	0.383838	2.03842	NaN	NaN	29.691118	0.523208	0.315184	NaN	32.264208	NaN
std	257.232042	0.486162	0.819871	NaN	NaN	14.524487	1.102743	0.816057	NaN	49.883428	NaN
min	1.000000	0.000000	1.000000	NaN	NaN	0.420000	0.000000	0.000000	NaN	0.000000	NaN
25%	222.500000	0.000000	2.000000	NaN	NaN	20.125000	0.000000	0.000000	NaN	7.915000	NaN
50%	480.000000	0.000000	3.000000	NaN	NaN	26.000000	0.000000	0.000000	NaN	14.454208	NaN
75%	664.500000	1.000000	3.000000	NaN	NaN	38.000000	1.000000	0.000000	NaN	31.000000	NaN
max	891.000000	1.000000	3.000000	NaN	NaN	61.000000	8.000000	6.000000	NaN	512.320000	NaN

#### Hasil Exploratory Data Analysis (EDA)

Setelah pada tahap sebelumnya yaitu tahap *Data Understanding* dilakukan nya penghapusan terhadap kolom yang tidak memiliki relevansi terhadap penelitian dan juga penghapusan terhadap *row missing value*. Maka akan dilakukan kembali penampilan data terbaru menggunakan sintaks *read.csv* Hasilnya adalah data yang akan diolah pada tahap *exploratory data analysis* berjumlah 712 baris data dan 7 kolom, 7 kolom yang dimaksud adalah *Survived*, *Pclass*, *Sex*, *Age*, *Sibsp*, *Parch*, dan *Embarked*. Dapat dilihat pada Tabel Hasil Read Data.

Tabel Hasil Read Data

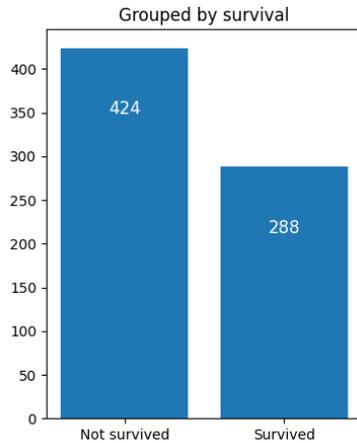
Survived	Pclass	Sex	Age	SibSp	Parch	Embarked	
0	0	3	male	22.0	1	0	S
1	1	1	female	38.0	1	0	C
2	1	3	female	26.0	0	0	S
3	1	1	female	35.0	1	0	S
4	0	3	male	35.0	0	0	S
...	...	...	...	...	...	...	...
707	0	3	female	39.0	0	5	Q
708	0	2	male	27.0	0	0	S
709	1	1	female	19.0	0	0	S
710	1	1	male	26.0	0	0	C
711	0	3	male	32.0	0	0	Q

[712 rows x 7 columns]

Kemudian dilakukannya visualisasi data terhadap beberapa kolom yang tersedia pada *dataset Titanic*.

- Grouped by Survival (dikelompokkan berdasarkan yang selamat dan tidak)

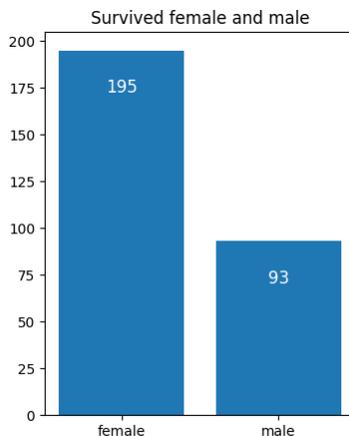
Diketahui bahwa sebanyak 424 orang dari 712 penumpang kapal tidak selamat sedangkan 288 orang lainnya dinyatakan selamat. Berdasarkan Gambar dapat ditarik kesimpulan bahwa sebanyak 60% penumpang kapal Titanic tidak selamat dan 40% sisanya adalah penumpang yang selamat. Dapat dilihat pada Gambar Hasil Visualisasi Kelangsungan Hidup.



**Gambar** Hasil Visualisasi Kelangsungan Hidup

- *Survival number according to gender or Sex* (jumlah kelangsungan hidup menurut jenis kelamin)

Pengelompokkan data pada jumlah kelangsungan hidup berdasarkan jenis kelaminnya diketahui dari 288 penumpang yang selamat sebanyak 195 orang berjenis kelamin perempuan dan sisanya yaitu sebanyak 93 orang berjenis kelamin laki-laki selamat. Berdasarkan Gambar dapat ditarik kesimpulan bahwa sebanyak 68% penumpang kapal Titanic dengan jenis kelamin perempuan selamat dan 32% sisanya adalah penumpang yang berjenis kelamin laki-laki. Dapat dilihat pada Gambar Hasil Visualisasi Penumpang Selamat Berdasarkan Jenis Kelamin.



**Gambar** Hasil Visualisasi Penumpang Selamat Berdasarkan Jenis Kelamin

Dilakukan pengelompokkan kumpulan data berdasarkan kolom 'Survived' dan 'Sex'. Hasil yang didapatkan adalah diketahui bahwa pada kolom *Survived* terdapat angka 0 yang berarti penumpang kapal tidak selamat sedangkan angka 1 menunjukkan bahwa penumpang kapal selamat, setelah itu dilakukan lagi pengelompokkan berdasarkan jenis kelamin, pada penumpang yang tidak selamat terdapat 64 penumpang berjenis kelamin perempuan dan 360 penumpang berjenis kelamin laki-laki, sedangkan pada penumpang selamat terdapat 195 penumpang berjenis kelamin perempuan dan 93 penumpang berjenis kelamin laki-laki. Dapat dilihat pada Tabel Hasil Menghitung Data Penumpang Berdasarkan Kolom 'Survived' dan 'Sex'.

**Tabel** Hasil Menghitung Data Penumpang Berdasarkan Kolom 'Survived' dan 'Sex'

Survived	Sex	Count
0	female	64
	male	360
1	female	195
	male	93

- *Survival number according on Pclass* (jumlah kelangsungan hidup menurut Pclass)

Dilakukan pengelompokkan kumpulan data berdasarkan kolom 'Pclass', 'Survived' dan 'Sex',

Hasil yang didapatkan adalah ditunjukkan bahwa total penumpang Pclass 1 yang tidak selamat sebanyak 64 orang, diantaranya sebanyak 3 orang dengan jenis kelamin perempuan dan 61 orang dengan jenis kelamin laki-laki. Sedangkan total penumpang Pclass 1 yang selamat sebanyak 120 orang, diantaranya sebanyak 80 orang dengan jenis kelamin perempuan dan 40 orang dengan jenis kelamin laki-laki.

Pada Pclass 2, ditunjukkan bahwa total penumpang yang tidak selamat sebanyak 90 orang, diantaranya sebanyak 6 orang dengan jenis kelamin perempuan dan 84 orang dengan jenis kelamin laki-laki. Sedangkan total penumpang Pclass 2 yang selamat sebanyak 83 orang, diantaranya sebanyak 68 orang dengan jenis kelamin perempuan dan 15 orang dengan jenis kelamin laki-laki.

Pada Pclass 3, ditunjukkan bahwa total penumpang yang tidak selamat sebanyak 270 orang, diantaranya sebanyak 55 orang dengan jenis kelamin perempuan dan 215 orang dengan jenis kelamin laki-laki.

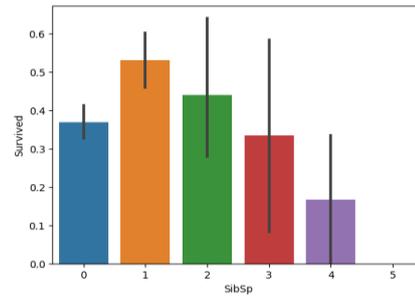
Sedangkan total penumpang Pclass 2 yang selamat sebanyak 85 orang, diantaranya sebanyak 47 orang dengan jenis kelamin perempuan dan 38 orang dengan jenis kelamin laki-laki.

Berdasarkan Gambar sebanyak 42% dari 288 orang yang selamat, penumpang Pclass 1 selamat dengan presentase jenis kelamin perempuan sebanyak 28% dan jenis kelamin laki-laki sebanyak 14%. Pada Pclass 2 sebanyak 29% dari 288 orang yang selamat, penumpang Pclass 2 selamat dengan presentase jenis kelamin perempuan sebanyak 24% dan jenis kelamin laki-laki sebanyak 5%. Pada Pclass 3 sebanyak 29% dari 288 orang yang selamat, penumpang Pclass 3 selamat dengan presentase jenis kelamin perempuan sebanyak 16% dan jenis kelamin laki-laki sebanyak 13%. Dapat ditarik kesimpulan bahwa penumpang terbanyak yang selamat berada di Pclass 1. Dapat dilihat pada Tabel Hasil Menghitung Data Penumpang Berdasarkan Kolom 'Pclass', 'Survived' dan 'Sex'.

**Tabel Hasil Menghitung Data Penumpang Berdasarkan Kolom 'Pclass', 'Survived' dan 'Sex'**

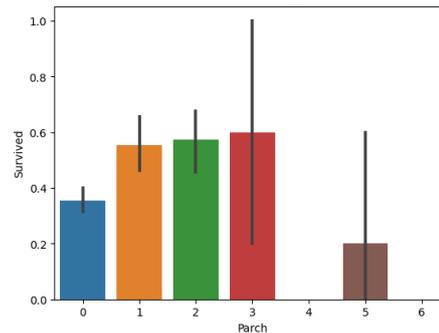
Pclass	Survived	Sex	
1	0	female	3
		male	61
	1	female	80
		male	40
2	0	female	6
		male	84
	1	female	68
		male	15
3	0	female	55
		male	215
	1	female	47
		male	38

- *Survival based on Sibsp* (kelangsungan hidup berdasarkan Sibsp)  
Dilakukan pengelompokan data berdasarkan kelangsungan hidup penumpang kapal berdasarkan kolom 'Sibsp'. Hasil didapatkan adalah, dapat ditarik kesimpulan bahwa penumpang yang memiliki banyak Sibsp memiliki sedikit peluang untuk bertahan hidup jika Sibsp == 0 atau 1 atau 2, penumpang memiliki lebih banyak peluang untuk selamat. Dapat dilihat pada Gambar Hasil Menvisualisasikan Kelangsungan Hidup Berdasarkan Kolom 'Sibsp'.



**Gambar Hasil Menvisualisasikan Kelangsungan Hidup Berdasarkan Kolom 'Sibsp'**

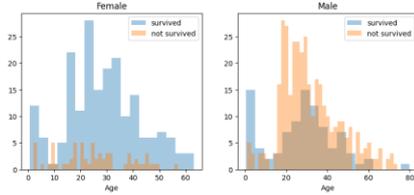
- *Survival based on Parch* (kelangsungan hidup berdasarkan Parch)  
Dilakukan pengelompokan data berdasarkan kelangsungan hidup penumpang kapal berdasarkan kolom 'Sibsp'. Hasil yang didapatkan adalah dapat ditarik kesimpulan bahwa keluarga kecil memiliki lebih banyak peluang untuk bertahan hidup. Dapat dilihat pada Gambar Hasil Menvisualisasikan Kelangsungan Hidup Berdasarkan Kolom 'Sibsp'.



**Gambar Hasil Menvisualisasikan Kelangsungan Hidup Berdasarkan Kolom 'Sibsp'**

- *Survival based on Age* (kelangsungan hidup berdasarkan usia)  
Dilakukan pengelompokan data berdasarkan kelangsungan hidup penumpang kapal berdasarkan kolom 'Age'. Hasil yang didapatkan adalah, dapat disimpulkan bahwa penumpang dengan jenis kelamin perempuan memiliki kemungkinann bertahan hidup lebih tinggi diantar usia 14-40 tahun sedangkan penumpang laki-laki memiliki kemungkinan bertahan hidup yang tinggi ketika mereka berusia antara 18 dan 30 tahun. Untuk penumpang berjenis kelamin laki-laki kemungkinan untuk bertahan hidup sangat rendah antara usia 5 dan 18 tahun, tapi itu tidak berlaku untuk penumpang berjenis kelamin perempuan. Hal lain yang

perlu diperhatikan adalah bahwa bayi juga memiliki kemungkinan bertahan hidup yang sedikit lebih tinggi. Dapat dilihat pada Gambar Hasil Menvisualisasikan Kelangsungan Hidup Berdasarkan Kolom 'Age' dan 'Sex'.

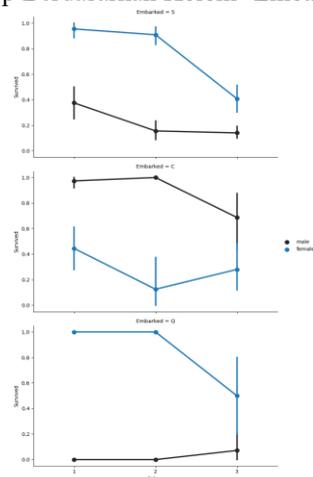


**Gambar Hasil Menvisualisasikan Kelangsungan Hidup Berdasarkan Kolom 'Age' dan 'Sex'**

- *Survival based on Embarked* (keberlangsungan hidup berdasarkan embarked)

Dilakukan pengelompokkan data berdasarkan kelangsungan hidup penumpang kapal berdasarkan kolom 'Embarked'.

Hasil yang didapatkan adalah diketahui bahwa penumpang dengan jenis kelamin perempuan di port Q dan di port S memiliki peluang lebih tinggi untuk bertahan hidup. Sebaliknya penumpang dengan jenis kelamin perempuan jika mereka berada di port C maka peluang kebertahan hidupnya sangat rendah. Penumpang dengan jenis kelamin laki-laki memiliki probabilitas bertahan hidup yang tinggi jika mereka berada di port C, tetapi probabilitas rendah jika mereka berada di port Q atau S. Dapat dilihat pada Gambar Hasil Menvisualisasikan Kelangsungan Hidup Berdasarkan Kolom 'Embarked'.



**Gambar Hasil Menvisualisasikan Kelangsungan Hidup Berdasarkan Kolom 'Embarked'**

Setelah dilakukan visualisasi maka akan dilakukannya tahap data encoding. Pada tahap

data encoding dilakukan proses pengubahan data kategorikal atau teksual menjadi bentuk numerik yang dapat diterima oleh model statistik atau *machine learning*.

Hasil yang didapatkan adalah telah terlihat perubahan berkat proses data encoding, pada kolom *Sex* telah berubah menjadi 0 dan 1 dan kolom *Embarked* telah berubah menjadi 1, 2 dan 3. Ditunjukkan pada Tabel 8.

**Tabel 1. Hasil Data Encoding**

	Survived	Pclass	Sex	Age	SibSp	Parch	Embarked
0	0	3	1	22.0	1	0	2
1	1	1	0	38.0	1	0	1
2	1	3	0	26.0	0	0	2
3	1	1	0	35.0	1	0	2
4	0	3	1	35.0	0	0	2
...	...	...	...	...	...	...	...
707	0	3	0	39.0	0	5	3
708	0	2	1	27.0	0	0	2
709	1	1	0	19.0	0	0	2
710	1	1	1	26.0	0	0	1
711	0	3	1	32.0	0	0	3

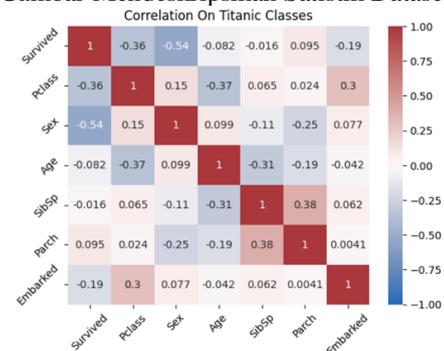
Maka setelahnya akan dilakukan deskripsi *statistic dataset* kembali terhadap *dataset* yang telah di encoding. Hasil yang didapatkan adalah data merupakan hasil deskripsi statistic terhadap data set, terdapat count, mean, standard deviation, minimum, quantile 25-75, dan maximum. dapat dilihat pada Tabel 9.

**Tabel 2 Mendeskripsikan Statistik Dataset**

	Survived	Pclass	Sex	Age	SibSp	Parch	Embarked
count	712.000000	712.000000	712.000000	712.000000	712.000000	712.000000	712.000000
mean	0.404494	2.240169	0.636236	29.642093	0.514045	0.432584	1.856742
std	0.491139	0.836854	0.481420	14.492933	0.930692	0.854181	0.449077
min	0.000000	1.000000	0.000000	0.420000	0.000000	0.000000	1.000000
25%	0.000000	1.000000	0.000000	20.000000	0.000000	0.000000	2.000000
50%	0.000000	2.000000	1.000000	28.000000	0.000000	0.000000	2.000000
75%	1.000000	3.000000	1.000000	38.000000	1.000000	1.000000	2.000000
max	1.000000	3.000000	1.000000	80.000000	5.000000	6.000000	3.000000

Setelah tahap diatas maka dilakukan visualisasi *heatmap*.

Hasil yang didapatkan adalah dapat disimpulkan bahwa korelasi *Survived* dan *Pclass* sebesar -0.36, korelasi *Survived* dan *Sex* sebesar -0.54, korelasi *Survived* dan *Age* sebesar -0.082, korelasi *Survived* dan *Sibsp* sebesar -0.016, korelasi *Survived* dan *Parch* sebesar 0.095 dan yang terakhir korelasi *Survived* dan *Embarked* sebesar -0.19. Visualisasi *heatmap* dapat di lihat pada Gambar Mendeskripsikan Statistik Dataset.



**Gambar Hasil Visualisasi Heatmap**

### Pembahasan Model Implementation Support Vector Machine

Untuk mendapatkan nilai akurasi dari model SVM yang lebih akurat, maka pengujian terhadap kernel *Support Vector Machine*, pengujian dilakukan untuk mengetahui model *Support Vector Machine* mana yang memiliki tingkat akurasi terbaik. Hasil uji dari model *Support Vector Machine* dapat dilihat pada Tabel hasil uji model support vector machine.

Tabel Hasil Uji Model Support Vector Machine

No	Support Vector Machine				HASIL	
	Kernel	C	Random state	Test size	Hold Out	10-Fold Cross Validation
1.	Linear	1.0	42	0.3	76.64%	77.94%
			1		76.64%	77.94%
			42	0.2	79.02%	77.94%
			1		79.02%	77.94%
			42	0.1	70.83%	77.94%
			1		70.83%	77.94%
		10.0	42	0.3	76.64%	77.94%
			1		76.64%	77.94%
			42	0.2	79.02%	77.94%
			1		79.02%	77.94%
			42	0.1	70.83%	77.94%
			1		70.83%	77.94%
		100.0	42	0.3	76.64%	-
			1		76.64%	-
			42	0.2	79.02%	-
			1		79.02%	-
			42	0.1	-	-
			1		-	-
2.	RBF (Radial Basis Function)	1.0	42	0.3	60.75%	62.20%
			1		60.75%	62.20%
			42	0.2	60.84%	62.20%
			1		60.84%	62.20%
			42	0.1	62.50%	62.20%
			1		62.50%	62.20%
		10.0	42	0.3	73.83%	80.75%
			1		73.83%	80.75%
			42	0.2	78.32%	80.75%
			1		78.32%	80.75%
			42	0.1	73.61%	80.75%
			1		73.61%	80.75%
		100.0	42	0.3	78.50%	81.03%
			1		78.50%	81.03%
			42	0.2	81.12%	81.03%
			1		81.12%	81.03%
			42	0.1	75.00%	81.03%
			1		75.00%	81.03%

Berdasarkan Tabel Hasil Uji Model Support Vector Machine, dapat ditarik kesimpulan bahwa kernel RBF (*Radial Basis Function*) dengan parameter C =100 dan test size sebesar 20% memiliki tingkat akurasi lebih tinggi dibandingkan yang lain nya, maka dari itu kernel RBF (*Radial Basis Function*) yang akan digunakan dalam melatih model dalam penelitian ini.

### Hasil Model Implementation dan Validation

Pada tahap ini dilakukan model implementation dan Validation. Metode validasi yang digunakan adalah *Holdout* dan *K-Fold Cross Validation* Hasil akurasi metode *Holdout* menghasilkan akurasi sebesar 81.12% yang berarti memiliki performa cukup baik dalam melakukan klasifikasi. Dapat dilihat Pada Gambar Hasil Skor *Holdout*.

Hold-out evaluation accuracy: 81.12%

Gambar Hasil Skor Holdout

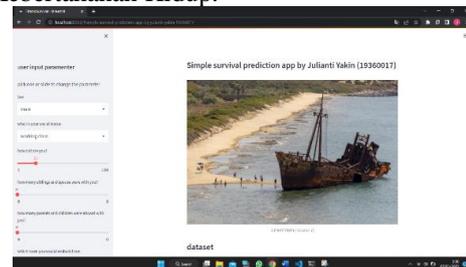
Hasil dari sintaks pada Gambar 4.26 diatas adalah rata-rata yang dihasilkan *K-Fold Cross Validation* adalah 0.8103 atau 81.03% yang berarti memiliki performa cukup baik dalam melakukan klasifikasi. Dapat dilihat Pada Gambar Hasil Skor *K-Fold Cross Validation*.

K-Fold Cross Validation scores: [0.8333333 0.8472222 0.77464789 0.87323944 0.78873239 0.78873239 0.74647887 0.74647887 0.83898592 0.87323944]  
Mean KFCV score: 0.8103890766823161

Gambar Hasil Skor K-Fold Cross Validation

### Hasil Graphical User Interface (GUI)

GUI sederhana untuk melakukan prediksi keberuntungan hidup yang dihasilkan dapat ditunjukkan pada Gambar Hasil GUI Prediksi Keberuntungan Hidup.

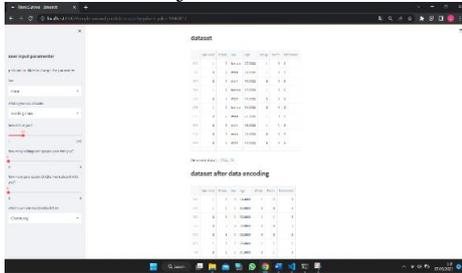


Gambar Hasil GUI Prediksi Keberuntungan Hidup

Pada Gambar Hasil GUI Prediksi Keberuntungan Hidup, terdapat *user input parameter* yang berfungsi untuk menentukan hasil prediksi berdasarkan *input* yang dimasukkan oleh *user*. *User input parameter* terdiri atas :

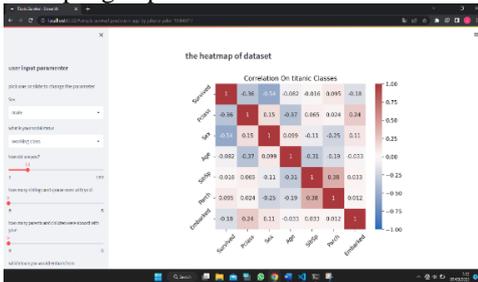
- *Selectbox Sex* yang berlandaskan atas kolom *Sex*, yang memungkinkan *user* untuk memilih jenis kelamin antara *female* dan *male*.
- *Selectbox social status* yang berlandaskan atas kolom *Pclass*, yang memungkinkan *user* untuk memilih antara *working class* yang sama dengan *Pclass 3*, *middle class* yang sama dengan *Pclass 2* dan *upper class* yang sama dengan *Pclass 1*.
- *Slider Siblings* dan *Spouse* yang berlandaskan atas kolom *Sibsp*, yang memungkinkan *user* untuk memilih berapa banyak saudara dan pasangan yang naik ke atas kapal bersama dimulai dari 0 hingga 8 orang.
- *Slider Parents* dan *Children* yang berlandaskan atas kolom *Parch*, yang memungkinkan *user* untuk memilih berapa banyak orangtua dan anak yang naik ke atas kapal bersama dimulai dari 0 hingga 6 orang.
- *Selectbox Embark* yang berlandaskan atas kolom *Embarked*, yang memungkinkan *user* untuk memilih Pelabuhan keberangkatan.

Pada Gambar Tampilan Dataset dan Hasil Data Encoding Pada GUI, terdapat sisi sebelah kanan GUI terdapat tampilan *dataset* sebelum dilakukan data *encoding* dan setelah dilakukannya data *encoding* maka data kolom *Sex* akan berubah menjadi 0 dan 1, kolom *Embarked* akan berubah menjadi 1, 2 dan 3.



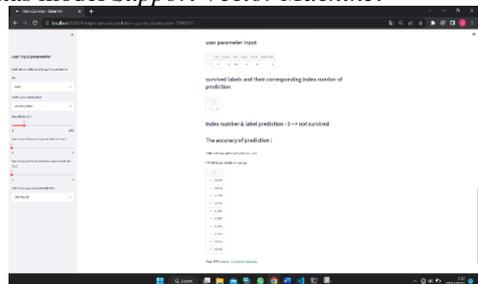
**Gambar** Tampilan Dataset dan Hasil Data Encoding Pada GUI

Pada Gambar Tampilan Hasil Heatmap Pada GUI, terdapat tampilan *heatmap* pada GUI yang memudahkan *user* untuk mengetahui korelasi antara kolom-kolom pada *dataset* terhadap keberlangsungan hidup penumpang kapal.



**Gambar** Tampilan Hasil Heatmap Pada GUI

Lalu yang terakhir pada Gambar Hasil Input Parameter, Prediksi dan Skor K-Fold Cross Validation, terdapat hasil *input parameter* yang dilakukan oleh *user*, label selamat dan nomor indeks prediksi yang sesuai dengan *input*, *output* prediksi atas *user parameter input* yang telah dimasukkan oleh *user* dan *output* skor *K-Fold Cross Validation* atas model *Support Vector Machine*.



**Gambar** Hasil Input Parameter, Prediksi dan Skor K-Fold Cross Validation

**Pembahasan Penggunaan Graphical User Interface (GUI)**

Yang harus dilakukan *user* adalah melakukan perubahan pada *user input parameter* yang terletak pada sebelah kiri tampilan GUI, setelah itu pada bagian bawah GUI akan muncul hasil prediksi berdasarkan *input* yang telah dimasukkan, jika hasil dari *input* yang telah dimasukkan *user* dinyatakan selamat maka akan muncul angka 1 dengan keterangan *Survived*, jika tidak maka akan muncul angka 0 dengan keterangan *not Survived*. Berikut merupakan contoh-contoh *output* :

**Output 1**

Pada Gambar *Output 1*, akan dimasukkan data pada *user parameter input* seperti yang dibawah ini.



**Gambar Output 1**

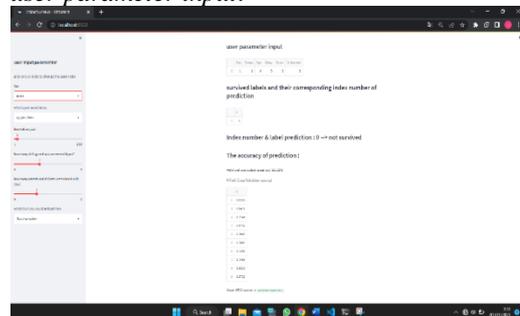
Pada Gambar *Output 1*, diatas menunjukkan bahwa berdasarkan *input* yang dimasukkan berupa :

- Sex = female
- Social status = upper class (Pclass 1)
- Age = 6
- Sibsp = 3
- Parch = 2
- Embarked = Southampton

maka hasil *output* prediksinya dinyatakan selamat.

**Output 2**

Pada Gambar *Output 2* akan dimasukkan data sama seperti Gambar *Output 1*, tetapi hanya *input Sex* nya yang dirubah menjadi *male* pada *user parameter input*.



**Gambar Output 2**

Pada Gambar Output 2 diatas menunjukkan bahwa berdasarkan *input* yang dimasukkan berupa :

- *Sex = male*
- *Social status = upper class (Pclass 1)*
- *Age = 6*
- *Sibsp = 3*
- *Parch = 2*
- *Embarked = Southampton*

maka hasil *output* prediksinya dinyatakan tidak selamat.

#### 4. SIMPULAN

Berdasarkan hasil penelitian yang telah dijalani dapat diambil beberapa kesimpulan sebagai berikut:

- Model machine learning dengan algoritma Support Vector Machine memiliki performa yang cukup baik dalam melakukan klasifikasi kebertahanan hidup.
- Hasil akurasi Algoritma Support Vector Machine menggunakan metode Holdout adalah sebesar 81.12%.
- Hasil akurasi Algoritma Support Vector Machine menggunakan K-Fold Cross Validation sebanyak 10 kali berturut-turut adalah 0.83, 0.84, 0.77, 0.87, 0.77, 0.78, 0.74, 0.74, 0.85, 0.8. Dan rata-rata yang didapat adalah sebesar 0.8103 atau sebesar 81.03%.
- Model yang telah dibuat kemudian dilakukan deployment dan diimplementasikan ke dalam sebuah sistem yang hasilnya sistem sederhana berbasis web dan berjalan dengan baik sesuai dengan keluaran yang diinginkan.

#### Saran

Adapun saran yang menjadi bahan masukan yang dapat dijadikan patokan untuk melaksanakan penelitian selanjutnya adalah sebagai berikut:

- Dalam Penelitian ini dilakukan menggunakan hitungan rasio/interval. Diharapkan kedepannya dapat mencoba algoritma lain yang menggunakan hitungan nominal/kategorikal.
- Sistem prediksi yang dibangun masih sangat sederhana, diharapkan kedepannya dapat ditingkatkan kembali.

#### 5. DAFTAR PUSTAKA

Aditya Nugroho, P., Saptono, R., & Eko Sulistyono, M. (2016). Perbandingan Metode Probabilistik Naive Bayesian

- Classifier dan Jaringan Syaraf Tiruan Learning Vector Quantization dalam Kasus Klasifikasi Penyakit Kandungan. *Jurnal Teknologi & Informasi ITSmart*, 2(2), 21. <https://doi.org/10.20961/its.v2i2.628>
- Apriani, D., Aan, M., & Saputra, W. E. (2022). Data Visualization Using Google Data Studio. *International Journal of Cyber and IT Service Management*, 2(1), 11–19. <https://doi.org/10.34306/ijcitsm.v2i1.68>
- Buulolo, E. (2020). Data Mining Untuk Perguruan Tinggi. In DeePublish.
- Cahyanti, D., Rahmayani, A., & Husniar, S. A. (2020). Analisis performa metode Knn pada Dataset pasien pengidap Kanker Payudara. *Indonesian Journal of Data and Science*, 1(2). <https://doi.org/10.33096/ijodas.v1i2.13>
- Dwiasnati, S., & Devianto, Y. (2021). Optimasi Prediksi Bencana Banjir menggunakan Algoritma SVM untuk penentuan Daerah Rawan Bencana Banjir. *Prosiding SISFOTEK, 202–207*. <http://seminar.iaii.or.id/index.php/SISFOTEK/article/view/283>
- Evi Fitria Umi Latifah. (2018). Perbandingan Kinerja Machine Learning Berbasis Algoritma Support Vector Machine dan Naive Bayes [Universitas Islam Indonesia]. [https://dspace.uui.ac.id/bitstream/handle/123456789/7996/LAPTA\\_14611064\\_Evi\\_Fitria\\_Umi\\_Latifah.pdf?sequence=1&isAllowed=y](https://dspace.uui.ac.id/bitstream/handle/123456789/7996/LAPTA_14611064_Evi_Fitria_Umi_Latifah.pdf?sequence=1&isAllowed=y)
- Feizal, F. (2022). Analisis dan Rancang Bangun Sistem Klasifikasi dan Seleksi Fitur Menggunakan Random Forest pada Studi Kasus Intrusi jaringan.
- Haque, M. A., Shivaprasad, G., & Guruprasad, G. (2021). Passenger Data Analysis of Titanic Using Machine Learning Approach in the Context Of Chances of Surviving the Disaster. *IOP Conference Series: Materials Science and Engineering*, 1065(1). <https://doi.org/10.1088/1757-899X/1065/1/012042>
- Kadarina, T. M., & Ibnu Fajar, M. H. (2019). Pengenalan Bahasa Pemrograman Python Menggunakan Aplikasi Games Untuk Siswa/I Di Wilayah Kembangan Utara. *Jurnal Abdi Masyarakat (JAM)*, 5(1), 11. <https://doi.org/10.22441/jam.2019.v5.i1.003>
- Miller, I. to B. D. V. J. D. (2017). Big Data Visualization.
- Nugroho. (2015). Penerapan Algoritma

- Support Vector Machine untuk Prediksi Harga Emas. *Jurnal Informatika Upgris*, 1(1), 10–19.
- Nugroho, A. M. (2021). Keterangan Gambar Otomatis Berbahasa Indonesia Menggunakan CNN dan LSTM Dengan Attention. [https://dspace.uui.ac.id/bitstream/handle/123456789/36118/17523105\\_AmiinMajiidNugroho.pdf?sequence=1](https://dspace.uui.ac.id/bitstream/handle/123456789/36118/17523105_AmiinMajiidNugroho.pdf?sequence=1)
- Nugroho, K. S. (2020). Validasi Model Klasifikasi Machine Learning pada Rapid Miner. <https://ksnugroho.medium.com/validasi-model-machine-learning-pada-rapidminer-50be0080df14>
- Nur Utomo, A. (2020). Analisa Data Ekstraksi Ciri Citra Momen Histogram dan Perbandingan Model Algoritma Klasifikasi Naive Bayes, Nearest Neighbor, Support Vector Machine, dan Decision Tree pada Studi Kasus Citra Jalan Aspal Rusak dan Jalan Aspal Tidak Rusak. *Incomtech*, 9(2), 8–18.
- Permana, A. Y., & Romadlon, P. (2019). Perancangan Sistem Informasi Penjualan Perumahan Menggunakan Metode SDLC pada PT. Mandiri Land Prosperous Berbasis Mobile. 10, 153–167.
- Putra, I. Y. (2020). Klasifikasi Tumor Otak Meningioma, Glioma dan Pituitari Dengan Menggunakan Convolutional Neural Network.
- Setiabudidaya, D. (2018). Penggunaan Piranti Lunak Jupyter Notebook dalam Upaya Mensosialisasikan Open Science. <https://doi.org/10.31227/osf.io/2h7q4>
- Wafa, H. S., Hadiana, A. I., & Umbara, F. R. (2022). Prediksi Penyakit Diabetes Menggunakan Algoritma Support Vector Machine (SVM). 4(1), 40–45. <https://e-journal.unper.ac.id/index.php/informati>  
s
- Wijoyo, R. H. S. (2021). Exploratory Data Analysis (EDA) Tinggi Muka Air di Jakarta.
- Wisdayani, Nur, I. M., & Wasono, R. (2019). Penerapan Algoritma K-Nearest Neighbor dalam Klasifikasi Tingkat Keparahan Korban Kecelakaan Lalu Lintas di Kabupaten Jawa Tengah. *Prosiding Mahasiswa Seminar Nasional Unimus*, 2(1), 373–380.