



ANALISIS PENUMPANG KAPAL TITANIC MENGGUNAKAN *TITANIC DATASET* DENGAN BANTUAN PEMROGRAMAN PYTHON

Alfian Fakhri

20032010097@student.upnjatim.ac.id

Universitas Pembangunan Nasional “Veteran” Jawa Timur

Yekti Condro Winursito

yekti.condro.ti@upnjatim.ac.id

Universitas Pembangunan Nasional “Veteran” Jawa Timur

Program Studi Teknik Industri, Fakultas Teknik & Sains,

Jl. Raya Rungkut Madya, Gunung Anyar, Surabaya 60294

ABSTRACT. *Titanic was a popular luxury passenger ship and sank on April 15, 1912. The Titanic accident was one of the most tragic events. As a result, the Titanic continues to be an object of study, research and discussion by various groups. One of the results of research into the sinking of the Titanic is a sample dataset of Titanic passengers. The primary goal of this research is to uncover insights, patterns, and relationships in the data set to better understand the factors that determined the survival of Titanic passengers. The research method used uses a quantitative approach. The quantitative analysis method focuses more on proving something through mathematical and statistical analysis. Data collection is carried out by taking data from the Titanic dataset. The results of this research are that the majority of passengers are in their twenties, with the highest concentration of passengers being in that age group. Unfortunately, more passengers did not survive the tragedy, with the proportion of male passengers who did not survive being greater than female passengers. There is a clear correlation between ticket class and passenger survival rate.*

Keywords: *Python, Titanic Dataset, Visualization*

ABSTRAK *Titanic adalah sebuah kapal penumpang mewah yang populer dan tenggelam pada tanggal 15 April 1912. Kecelakaan Titanic merupakan salah satu kejadian paling tragis. Akibatnya, Titanic terus menjadi objek studi, penelitian, dan perbincangan oleh berbagai kalangan. Salah satu hasil dari penelitian terhadap tenggelamnya Titanic adalah dataset sampel penumpang Titanic. Tujuan utama dari penelitian ini adalah untuk mengungkap wawasan, pola, dan hubungan dalam kumpulan data untuk lebih memahami faktor-faktor yang menentukan kelangsungan hidup penumpang Titanic. Metode penelitian yang dilakukan menggunakan pendekatan kuantitatif. Metode analisis kuantitatif lebih berfokus pada pembuktian suatu hal melalui analisis matematika dan statistika. Pengumpulan data dilakukan dengan cara mengambil data dari titanic dataset. Adapun hasil dari penelitian ini adalah mayoritas penumpang berusia dua puluhan, dengan konsentrasi penumpang tertinggi berada pada kelompok usia tersebut. Sayangnya, lebih banyak penumpang yang tidak selamat dari tragedi tersebut, dengan proporsi penumpang laki-laki yang tidak selamat lebih besar dibandingkan penumpang perempuan. Terdapat korelasi yang jelas antara kelas tiket dan tingkat kelangsungan hidup penumpang.*

Kata Kunci: *Python, Titanic Dataset, Visualisasi*

PENDAHULUAN

Titanic adalah sebuah kapal penumpang mewah yang populer dan tenggelam pada tanggal 15 April 1912. Titanic dibangun oleh perusahaan kapal Inggris bernama White Star Line. Pada saat itu, Titanic dianggap sebagai salah satu prestasi teknik yang luar biasa karena merupakan salah satu kapal penumpang terbesar dan termewah pada masanya. Banyak orang yang menunggu peluncuran dan pelayaran perdananya dengan antusias.

Pada tanggal 10 April 1912, Titanic memulai perjalanan perdananya dari Southampton. Kemudian, Titanic sampai di Cherbourg pada hari yang sama dan Queenstown pada keesokan harinya. Titanic dijadwalkan untuk sampai di New York pada 17 April 1912, tetapi kapal tersebut tidak pernah mencapai New York (Kompas, 2021). Kecelakaan Titanic merupakan salah satu kejadian paling tragis. Hal ini dikarenakan Titanic telah dipromosikan sebagai kapal yang tidak bisa tenggelam, dan tidak lama setelah pelayaran perdananya kapal tersebut tenggelam (Novanra, 2019). Pada tanggal 15 April 1912 pukul 2:20 pagi, Titanic tenggelam di Samudra Atlantik Utara. Kurangnya jumlah sekoci penyelamat untuk semua penumpang dan kru menjadi factor yang memperparah situasi. Meskipun usaha penyelamatan dilakukan dan beberapa sekoci penyelamat berhasil diluncurkan, banyak orang terpaksa tinggal di kapal yang tenggelam atau terjatuh ke dalam air yang dingin (Lim, et al., 2023).

Titanic tetap menjadi topik yang hangat dan menarik minat publik dari dulu sampai sekarang. Salah satu faktor yang mempengaruhi popularitasnya adalah film yang terkenal dengan judul "Titanic" yang dirilis pada tahun 1997. Akibatnya, Titanic terus menjadi objek studi, penelitian, dan perbincangan oleh berbagai kalangan. Salah satu hasil dari penelitian terhadap tenggelamnya Titanic adalah dataset sampel penumpang Titanic (Lim, et al., 2023).

Python merupakan sebuah bahasa pemrograman interpretatif yang memiliki banyak fungsi, dan didesain dengan fokus pada kejelasan dan kemudahan pemahaman kode. Python dianggap sebagai bahasa yang menggabungkan kemampuan dan kejelasan sintaks kode. Bahasa pemrograman Python dirancang khusus untuk memudahkan *programmer* dalam membuat program dengan efisiensi waktu, kemudahan pengembangan, dan kompatibilitas dengan system (Triono, et al., 2023). Python juga menyediakan berbagai *library* dan alat bantu yang kuat untuk analisis data, seperti Pandas, NumPy, dan Matplotlib (Susanto, et al., 2023).

Pengumpulan data yang akurat sangat penting dalam menjaga integritas penelitian. *Data cleaning* proses identifikasi informasi yang tidak lengkap, atau tidak relevan dan kemudian memodifikasi atau menghapus data yang kotor. Library python untuk memproses data tekstual. Menyediakan API yang konsisten untuk menganalisa ke dalam tugas-tugas pemrosesan bahasa alami (NLP) yang umum seperti *part-of-speech tagging*, ekstraksi kata benda, analisis sentimen (Fauziyyah & Gautama, 2020). Proses *cleaning data* yakni dengan cara membuang duplikasi data, memeriksa data yang inkonsisten (*missing value*), dan memperbaiki kesalahan data (Ginting, et al., 2020).

Tujuan utama dari penelitian ini adalah untuk mengungkap wawasan, pola, dan hubungan dalam kumpulan data untuk lebih memahami faktor-faktor yang menentukan kelangsungan hidup penumpang Titanic. Dengan mengeksplorasi atribut-atribut ini dan dampaknya terhadap kelangsungan hidup, kita dapat memperoleh wawasan berharga mengenai peristiwa tragis tersebut dan berpotensi membangun model prediktif untuk memperkirakan kelangsungan hidup berdasarkan karakteristik penumpang.

METODE PENELITIAN

Penelitian ini menggunakan metode kuantitatif. Analisis kuantitatif adalah proses mengumpulkan dan mengevaluasi data yang dapat diukur dan bernilai. Metode analisis kuantitatif lebih berfokus pada pembuktian suatu hal melalui analisis matematika dan statistika. Pengumpulan data dilakukan dengan cara mengambil data dari *titanic dataset*.

Dataset merupakan kumpulan data yang telah disusun supaya lebih mudah untuk di analisa. Dataset dapat digunakan untuk menjawab pertanyaan, membuat prediksi, dan membuat model. Kualitas kumpulan data sangat penting untuk keberhasilan *machine learning* atau ilmu data apa pun. Titanic dataset adalah kumpulan data yang terkenal dan sering dianalisis di bidang ilmu data dan pembelajaran mesin. Ini memberikan catatan komprehensif tentang penumpang di kapal naas RMS Titanic, yang terkenal tenggelam pada pelayaran perdananya pada bulan April 1912. Kumpulan data tersebut berisi berbagai informasi tentang setiap penumpang, termasuk detail pribadi mereka, informasi tiket, dan status kelangsungan hidup.

HASIL DAN PEMBAHASAN

Masalah utama yang ingin dipecahkan melalui analisis data ini adalah memahami faktor-faktor yang mempengaruhi kelangsungan hidup penumpang Titanic. Dengan banyaknya kolom dataset, sehingga dapat mengeksplorasi dan menganalisis berbagai aspek yang terkait dengan masalah ini. Dataset berisi 891 data yang terdiri dari *passenger_id*, *survived*, *pclass*, *name*, *sex*, *age*, *sibsp*, *parch*, *ticket*, *fare*, *cabin*, *embarked*. Kelas kabin terdiri dari tiga kelas, yaitu kelas satu, kelas dua, dan kelas tiga. Status kehidupan penumpang terdiri dari dua jenis, yaitu nol dan satu. Angka nol menyatakan bahwa penumpang tersebut meninggal, sedangkan angka satu menyatakan bahwa penumpang tersebut selamat dari kecelakaan Titanic.

- *Data Cleaning*

1. *FILL IN THE MISSING VALUES IN THE 'AGE' AND 'FARE' COLUMNS USING TITANIC_TEST DATASET*



```
# mengisi nilai missing values dalam kolom 'Age' dan 'Fare' (titanic_test)
# before
test_df.isnull().sum()

PassengerId    0
Survived       0
Pclass         0
Name           0
Sex            0
Age            86
Sibsp          0
Parch          0
Ticket         0
Fare           1
Cabin          327
Embarked       0
dtype: int64

# mengisi nilai missing values dalam kolom 'Age' dan 'Embarked' (titanic_test)
# after
test_df['Age'].fillna(test_df['Age'].median(), inplace = True)
test_df['Fare'].fillna(test_df['Fare'].mode()[0], inplace = True)
test_df.isnull().sum()

PassengerId    0
Survived       0
Pclass         0
Name           0
Sex            0
Age            0
Sibsp          0
Parch          0
Ticket         0
Fare           0
Cabin          327
Embarked       0
dtype: int64
```

Gambar 1. *Filling in missing values in the 'age' and 'fare' columns pada test dataset*

Dari gambar 1 diketahui bahwa sebelum dilakukan pengisian kolom, kolom *age* terdapat 86 rows yang kosong, kemudian 86 rows tersebut diisi dengan median dari kolom *age*. Dan kolom *fare* terdapat 1 row yang kosong, kemudian 1 row tersebut diisi dengan 0.

2. *REMOVE UNNECESSARY COLUMN: CABIN*



```
#menghapus kolom yang tidak diperlukan (titanic_train)
#before
train_df.columns

Index(['PassengerId', 'Survived', 'Pclass', 'Name', 'Sex', 'Age', 'SibSp',
       'Parch', 'Ticket', 'Fare', 'Cabin', 'Embarked'],
      dtype='object')

#menghapus kolom yang tidak diperlukan (titanic_train)
#after
train_df.drop(['Cabin'], axis=1, inplace=True)
train_df.columns

Index(['PassengerId', 'Survived', 'Pclass', 'Name', 'Sex', 'Age', 'SibSp',
       'Parch', 'Ticket', 'Fare', 'Embarked'],
      dtype='object')
```

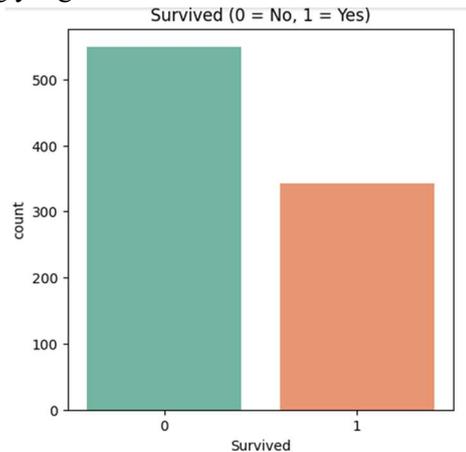
Gambar 2. *Remove unnecessary column: cabin*

Dari gambar 2 diketahui bahwa menghapus kolom yang tidak diperlukan. Kolom yang dihapus adalah kolom *cabin*. Kolom *cabin* dihapus karena *cabin* dianggap variabel yang tidak berhubungan dengan keselamatan penumpang.

- **EXPLORATORY DATA ANALYSIS & VISUALIZATION**

1. Jumlah Penumpang yang Selamat dan Penumpang yang Tidak Berhasil Selamat

```
# jumlah penumpang selamat vs tidak selamat (titanic_train)
plt.figure(figsize= (5, 5))
sns.countplot(data=train_df, x = 'Survived', palette= 'Set2')
plt.title("Survived (0 = No, 1 = Yes)")
plt.show()
```

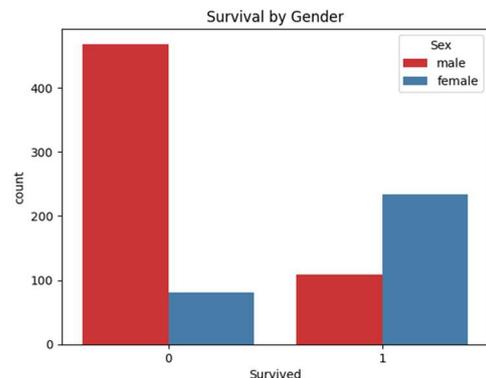


Gambar 3. Visualisasi jumlah penumpang yang selamat dan tidak selamat

Pada gambar 3 diketahui bahwa penumpang yang selamat berjumlah 342 orang dan penumpang yang tidak berhasil selamat berjumlah 549 orang. Sehingga jumlah penumpang yang tidak selamat lebih banyak dari pada jumlah penumpang yang selamat.

2. Jumlah Penumpang yang Selamat Berdasarkan Jenis Kelamin

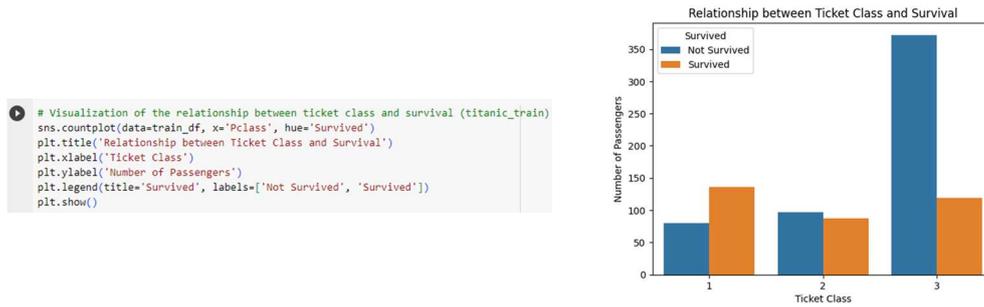
```
#survival by Gender (titanic_train)
sns.countplot(x='Survived', hue='Sex', data=train_df, palette='Set1')
plt.title("Survival by Gender")
plt.show()
```



Gambar 4. Visualisasi persebaran penumpang yang selamat dan tidak selamat berdasarkan *gender*

Pada gambar 4 diketahui bahwa penumpang laki-laki yang selamat berjumlah 109 orang dan penumpang perempuan yang selamat berjumlah 233 orang. Sedangkan penumpang laki-laki yang tidak berhasil selamat berjumlah 468 orang dan penumpang perempuan yang tidak berhasil selamat berjumlah 81 orang. Sehingga jumlah penumpang laki-laki yang tidak berhasil selamat lebih banyak dari pada jumlah penumpang laki-laki yang selamat. Sedangkan untuk Perempuan kebalikannya jumlah penumpang perempuan yang selamat lebih banyak dari pada jumlah penumpang Perempuan yang tidak berhasil selamat.

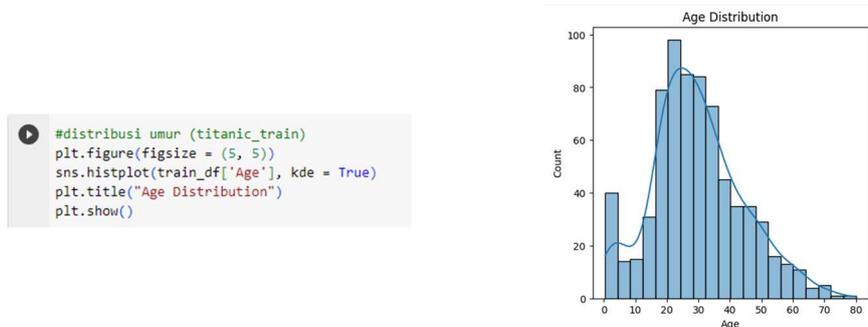
3. Hubungan tiket kelas dan keberhasilan untuk selamat



Gambar 5. Visualisasi hubungan *ticket class* dan *survival*

Pada gambar 5 diketahui bahwa tiket kelas tiga yaitu kelas terendah dipilih banyak penumpang. Sehingga terdapat korelasi yang jelas antara kelas tiket dan tingkat kelangsungan hidup penumpang. Semakin tinggi kelas tiketnya, semakin tinggi kemungkinan untuk bertahan hidup. Hubungan ini berbanding lurus, dengan angka kematian yang jauh lebih tinggi pada penumpang kategori tiket kelas tiga.

4. Persebaran Usia Penumpang



Gambar 6. Visualisasi usia penumpang

Pada gambar 6 diketahui bahwa usia penumpang bervariasi dengan rentang 0 hingga 80 tahun.

Dengan rincian penumpang dengan usia 0 – 10 tahun berjumlah 64 anak, penumpang dengan usia 11 – 20 tahun berjumlah 115 orang, penumpang dengan usia 21 – 30 tahun berjumlah 229 orang, penumpang dengan usia 31 – 40 tahun berjumlah 153 orang, penumpang dengan usia 41 – 50 tahun berjumlah 84 orang, penumpang dengan usia 51 – 60 tahun berjumlah 42 orang, penumpang dengan usia 61 – 80 tahun berjumlah 21 orang, dan terdapat 183 orang yang tidak diketahui usianya.

KESIMPULAN

Distribusi usia penumpang Titanic bervariasi, mulai dari 0 hingga 80 tahun. Mayoritas penumpang berusia dua puluhan, dengan konsentrasi penumpang tertinggi berada pada kelompok usia tersebut. Sayangnya, lebih banyak penumpang yang tidak selamat dari tragedi tersebut, dengan proporsi penumpang laki-laki yang tidak selamat lebih besar dibandingkan penumpang perempuan. Terdapat korelasi yang jelas antara kelas tiket dan tingkat kelangsungan hidup penumpang. Semakin tinggi kelas tiketnya, semakin tinggi kemungkinan untuk bertahan hidup. Hubungan ini berbanding lurus, dengan angka kematian yang jauh lebih tinggi pada penumpang kategori tiket kelas tiga.

DAFTAR PUSTAKA

- Fauziyyah, A. K. & Gautama, D. H., 2020. Analisis Sentimen Pandemi Covid-19 Pada Streaming Twitter Dengan Text Mining Python. *Jurnal Ilmiah Sinus (JIS)*, 18(2), pp. 31 - 42.
- Ginting, V. S., Kusriani & Taufiq, E., 2020. Implementasi Algoritma C4.5 untuk Memprediksi Keterlambatan Pembayaran Sumbangan Pembangunan Pendidikan Sekolah Menggunakan Python. *Jurnal Teknologi Informasi dan Komunikasi*, 10(2), pp. 36 - 44.
- Lim, K. J. et al., 2023. Penggunaan Bahasa Pemrograman Python Untuk Memvisualisasikan Data Peluang Selamat Dari Kecelakaan Titanic. *Jurnal Publikasi Teknik Informatika (JUPTI)*, 2(2), pp. 66 - 79.
- Novanra, Fahira, Rushariyanto, Rasyifa, dkk. 2019. Analisis Film Titanic Sebagai Media Kampanye Mitigasi Bencana Kapal Tenggelam. *SEMIOSTIKA: Jurnal Komunikasi*, 13(02), 1-8.
- Susanto, K. A. et al., 2023. Implementasi Bahasa Python Dalam Menganalisis Pengaruh Rokok Terhadap Risiko Pasien Terkena Penyakit Stroke. *Jurnal Publikasi Teknik Informatika (JUPTI)*, pp. 48 - 58.
- Tim Kompas. 2021. *Mengapa Titanic Bisa Tenggelam? Ini Kronologi dan Kisahnya*. <https://internasional.kompas.com/read/2021/09/30/112221170/mengapa-titanic-bisa-tenggelam-ini-kronologi-dan-kisahny?page=all>. Diakses tanggal 9 Januari 2024.
- Triono, A., Budi, A. S., Abdillah, R. & Wahyudi, 2023. Implementasi Peretasan Sandi Vigenere Chipper Menggunakan Bahasa Pemrograman Python. *Jurnal JOCOTIS - Journal Science Informatica and Robotics*, 1(1), pp. 1 -9.