

Proposal Dataset Titanic

Dibuat Oleh: Stefannus Christian 202000138

Masalah yang Coba Diselesaikan

Masalah yang saya coba selesaikan adalah masalah Titanic yang didasarkan pada tenggelamnya kapal Titanic pada awal 1912. Terdapat informasi tentang penumpang – penumpang titanic saat itu seperti *age*, *gender*, *number of siblings*, *embarked*, dan apakah mereka selamat atau tidak dari bencana ini. Berdasarkan fitur-fitur ini, saya akan membuat model Logistic Regression untuk memprediksi jika penumpang di Titanic akan selamat dari tenggelam.

Deskripsi Singkat Data

Terdapat 12 fitur dari dataset titanic yang saya ambil. Terdapat 11 fitur input dan 1 fitur output.

1. PassengerID: Passenger ke i dari dataset (*Unique Value*).
2. Survived: Menandakan apakah suatu penumpang selamat atau tidak. 0 = tidak selamat, 1 = selamat (Merupakan fitur output).
3. Pclass (Passenger Class): Divisi dari penumpang. Terdiri dari 3 kelas yaitu kelas 1, 2, dan 3. Deskripsi Kelas: 1 → Upper Class, 2 → Middle Class, 3 → Lower Class.
4. Name: Nama dari penumpang.
5. Sex: Jenis kelamin penumpang.
6. Age: Umur dari penumpang.
7. SibSp (Sibling Spouse): Jumlah saudara kandung atau pasangan dari seorang penumpang di dalam kapal. Note: Mistresses and Fiances Ignored.
8. Parch: Jumlah orangtua atau anak dari seorang penumpang di dalam kapal.
9. Ticket: Nomor ticket dari penumpang.
10. Fare: Harga tiket kapal yang dibayar penumpang dalam *British Pound* pada saat itu.
11. Cabin: Nomor Cabin penumpang.
12. Embarked: Dari mana penumpang naik kapal. Ada tiga kemungkinan nilai untuk fitur ini yaitu Southampton, Cherbourg, dan Quennstown.

Sumber Data

https://github.com/krishnaik06/EDA1/blob/master/titanic_train.csv

Cara Memperoleh Data

Saya mengambil data ini dari github krishnaik06 sehingga data yang saya ambil sudah dalam keadaan *Raw Data Ready*.