

Reyhan Suisanto

2301872980 – LD01

1. **Esai**
2. Jelaskan pemahaman mengenai Discrete Random Variables dan Continiuous Random Variables ! Berikan Contoh.

* Discrete Random Variables

Discrete random variables merupakan sebuah variabel random yang pengambilan variabelnya mengambil salah satu daftar nilai pasti yang ditentukan. Artinya, discrete random variables berarti merupakan sebuah angka random yang pasti jumlahnya. Discrete random variables dapat mengambil nilai diskrit terbatas atau paling banyak yang tidak terhitung.

Contoh = Jumlah mobil yang terjual di showroom dalam 1 bulan (angkanya pasti)

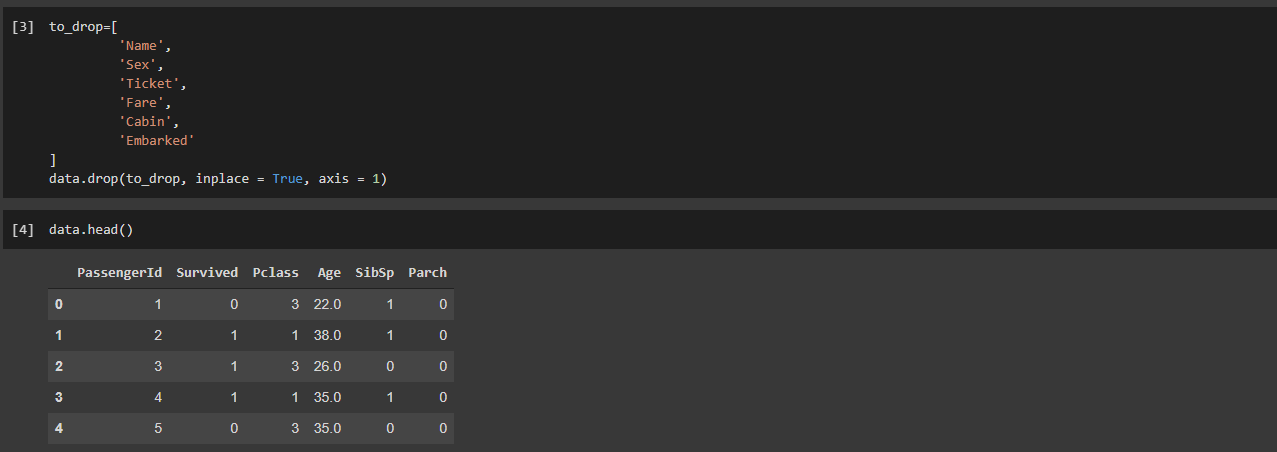
* Continiuous Random Variables

Sedangkan continuous random variables merupakan variabel random yang mengambil nilai numerik bentuk apapun dalam sebuah interval atau sekumpulan interval. Artinya, continuous random variables merupakan angka yang bisa berubah-ubah.

Contoh = kedalaman sebuah pengeboran oli. (Berubah-ubah)

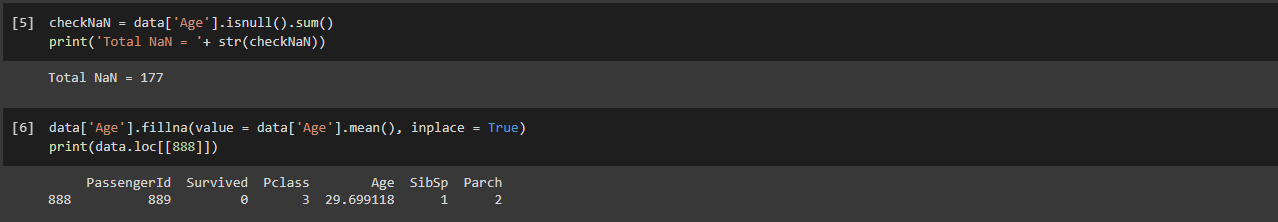
1. **Kasus**
2. Buka file data “titanic.txt” yang diberikan dan lakukan Data Cleaning dan Visual Exploratory Data Analysis, dengan detail yang dimaksud sebagai berikut:
3. Siapkan Variable yang akan dipakai Survived, PassengerId, PClass, Age, SibSp & Parch

* Saya menggunakan fungsi to\_drop untuk membuang variabel-variabel yang tidak diperlukan. Kemudian saya print datasetnya sehingga hanya menyisakan variabel yang dibutuhkan.



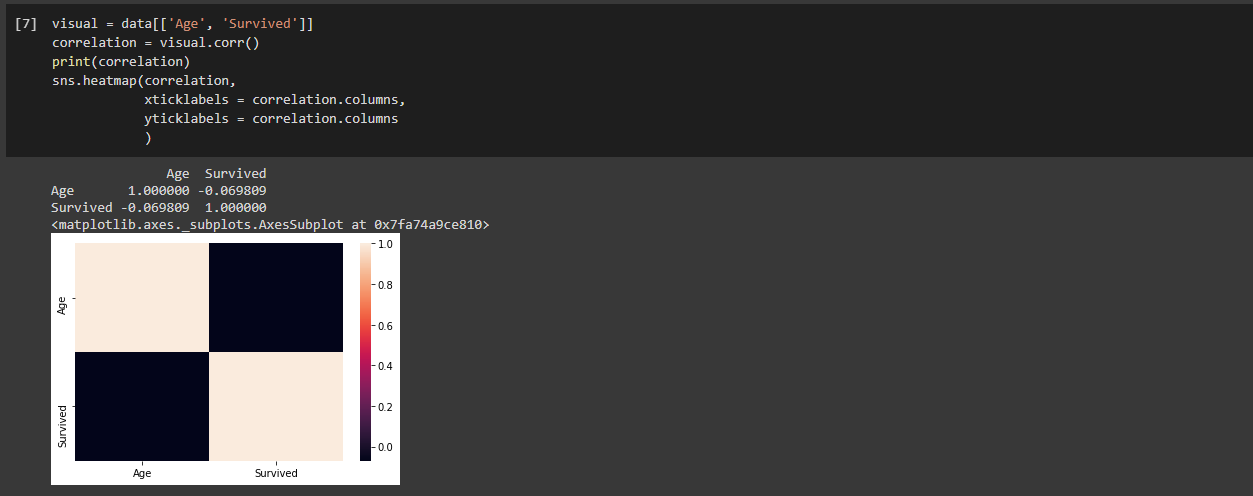
1. Periksa missing values dan fill NA dengan nilai rata-rata untuk continiuous variable

* Kemudian untuk memeriksa NaN saya menggunakan fungsi ‘checkNaN’ dan saya print jumlah NaN yang ada. Saya juga menemukan bahwa NaN hanya terdapat pada kolom Age. Jadi, saya fillNaN dengan rata-rata dari nilai kolom Age menggunakan fungsi fillna.

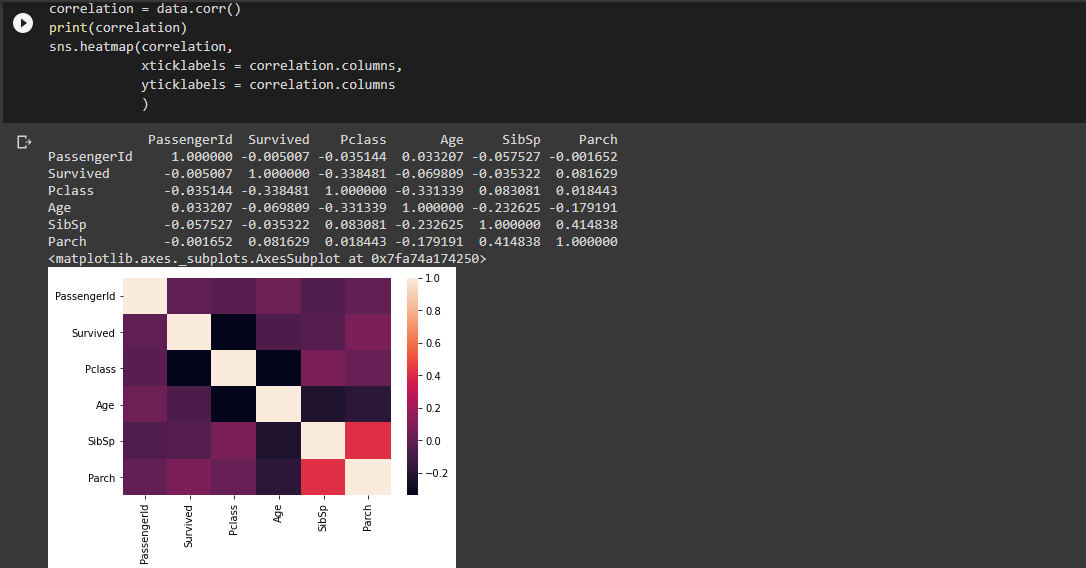


1. Visualisasikan dan jelaskan hubungan 2 atau 3 variabel

* Saya mengambil variabel Age dan Survived

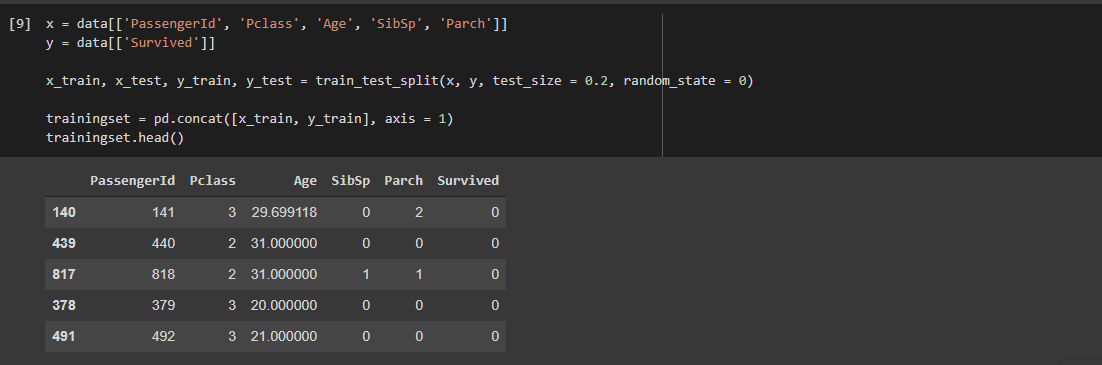


1. Periksa Correlations dari independent variable dengan dependent variable



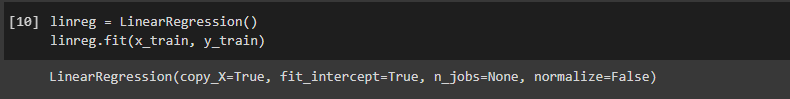
1. Buatlah Training Set dan Test Set, lalu print Training Set.

* Pertama saya bagi dataset menjadi 2 bagian yaitu x(feature) dan juga y(target). Kemudian saya lakukan train test split menggunakan library dengan aturan 80% training dan 20% testing. Lalu saya satukan kedua training set dan saya print

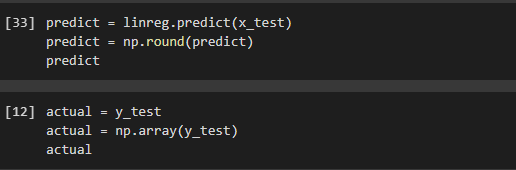


1. Lakukan Pemodelan Linear Regression
2. Prediksi Penumpang yang Survived

* Untuk linear regression saya menggunakan library sklearn.linear\_model. Saya membuat variabel linreg dan memasukkannya dengan linear regression. Lalu saya menggunakan data dari x\_train dan y\_train untuk melakukan linear regression

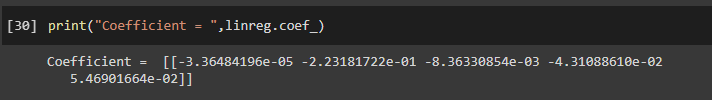


1. Print Data Prediksi dan Aktual.
   * Lalu saya melakukan prediksi untuk membandingkan hasil prediksi dengan data aktualnya. Untuk prediksi saya menggunakan fungsi linreg.predict dan menggunakan data dari x\_test untuk di prediksi. Kemudian data aktual saya ambil dari y\_test karena berisi data Survived. Hasilnya dapat dilihat di .ipynb yang saya sertakan dalam file.

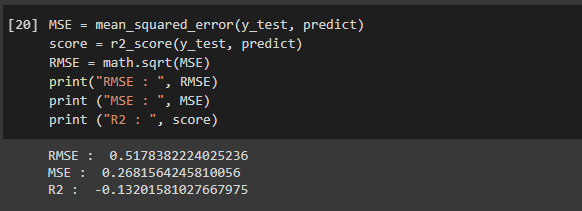


1. Lakukan Evaluasi Pemodelan Linear Regression tersebut dengan cara berikut:
2. Evaluasi Coefficient untuk setiap variable

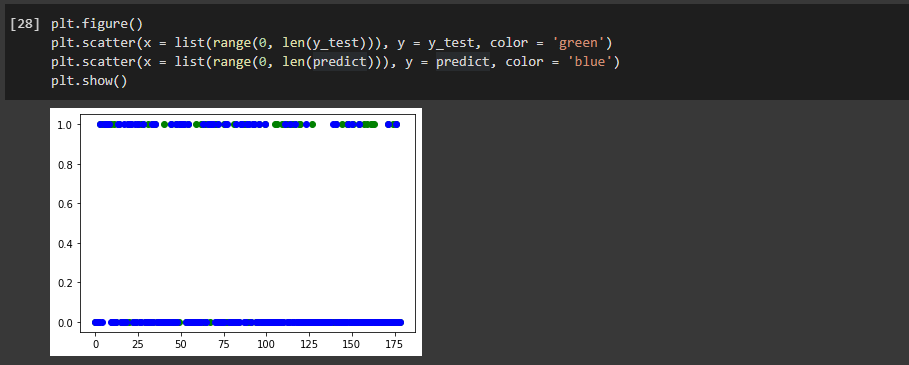
* Untuk coefficient saya menggunakan .coef\_ dan mendapatkan hasil sebagai berikut.



1. Evaluasi menggunakan Metrics (MAE,MSE, RMSE, dsb)

* Lalu saya juga mencari RMSE, MSE, dan R2 untuk model kali ini. Saya mendapatkan hasil sebagai berikut. R2 nya minus menandakan bahwa data aktual tidak cocok dengan data prediksi seperti yang tadi sudah saya print.

1. Lakukan Plot Predicted Value vs Measure

* Terakhir, untuk membandingkan antara prediksi dan juga data aktual saya menggunakan plt.figure. Berikut adalah hasil dari kedua data yang dibandingkan
* Jadi saya memiliki kesimpulan bahwa data aktual dengan data prediksi lumayan berbeda di mana data prediksi menunjukkan jumlah yang tidak selamat lebih banyak dari data aktual dan jumlah yang selamat lebih sedikit dari data aktual.