Nama : A Arga Bimantara A

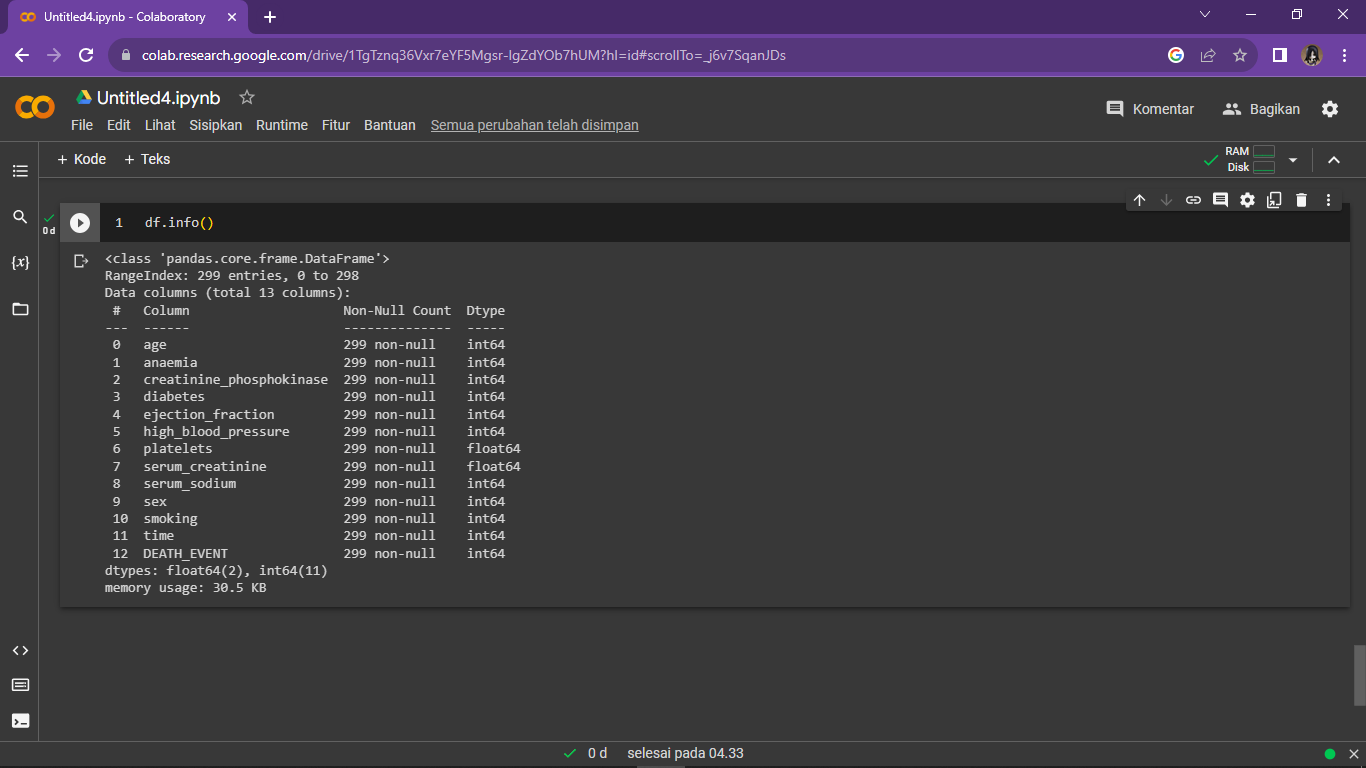
NIM : 200411100137

Kelas : PSD E

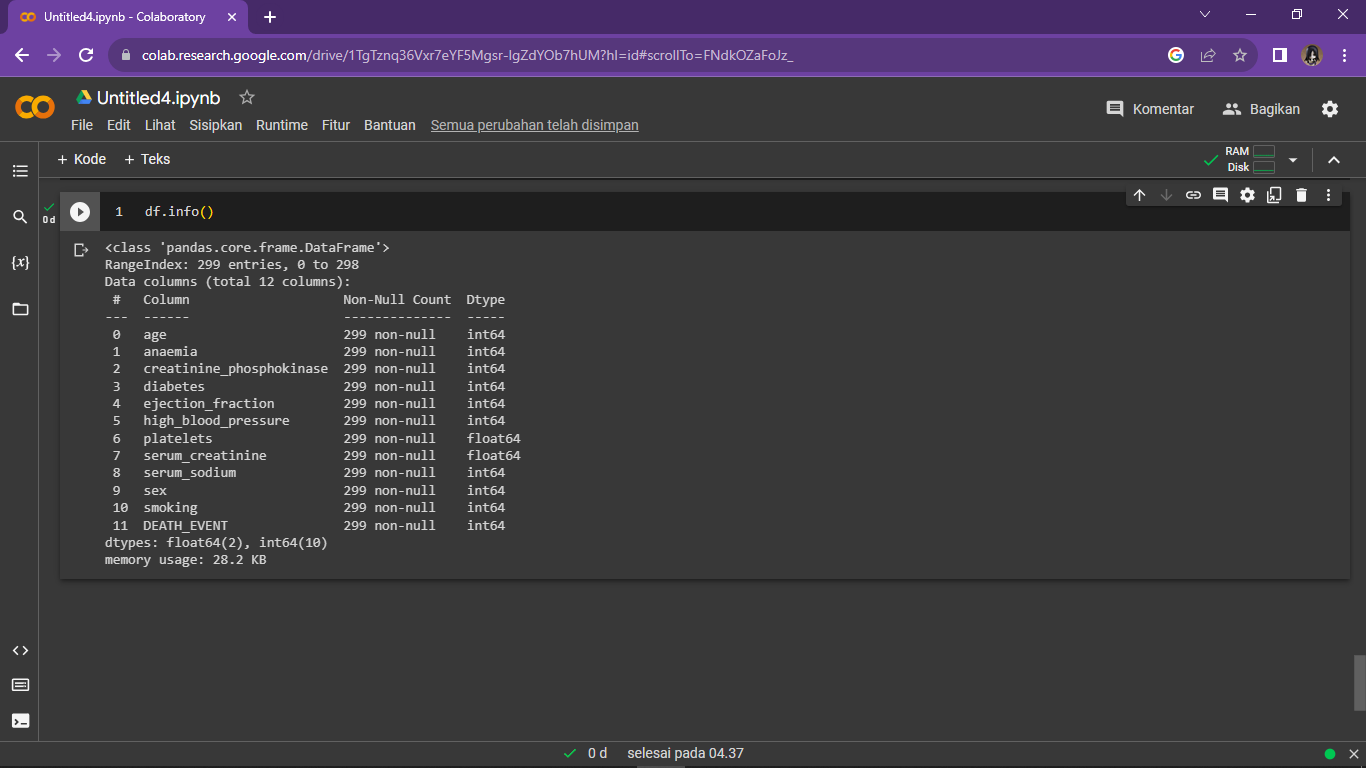
**Tugas**

Dataset : <https://www.kaggle.com/datasets/andrewmvd/heart-failure-clinical-data>

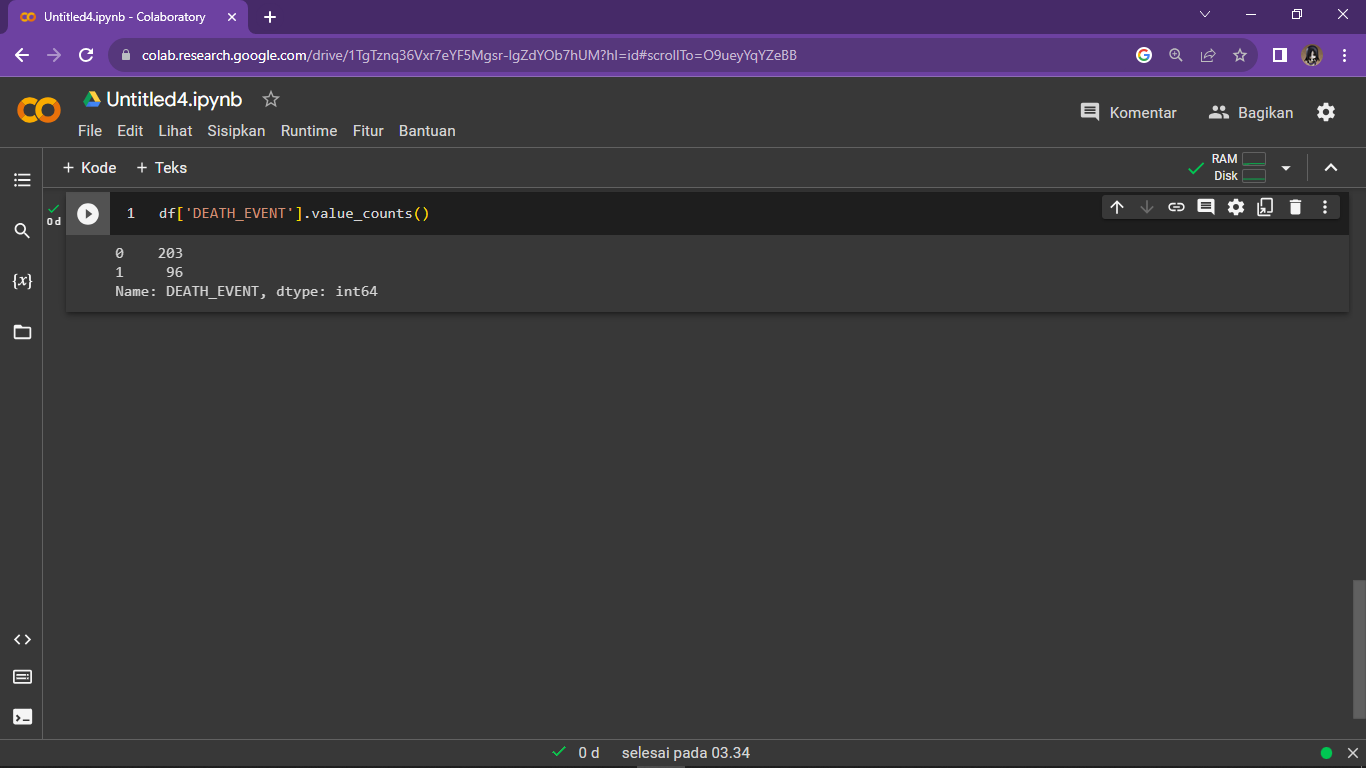
Data tentang prediksi keselamatan pasien dari penyakit jantung. Data ini memiliki 13 atribut yang terdiri dari 12 kolom fitur dan 1 kolom target dengan total data sebanyak 299



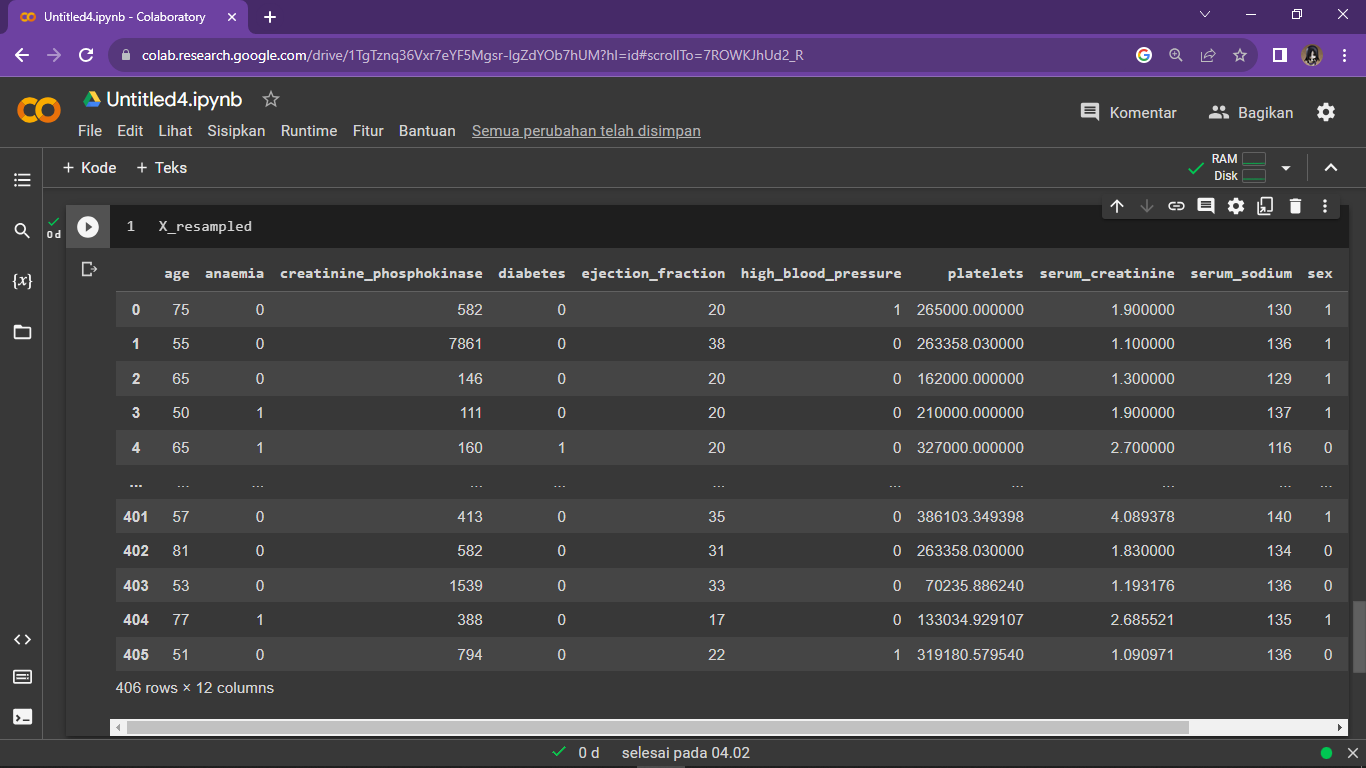
Dari 12 fitur tersebut ada 1 fitur yang tidak dipakai yaitu fitur time,fitur ini akan di drop sehingga hasilnya sebagai berikut



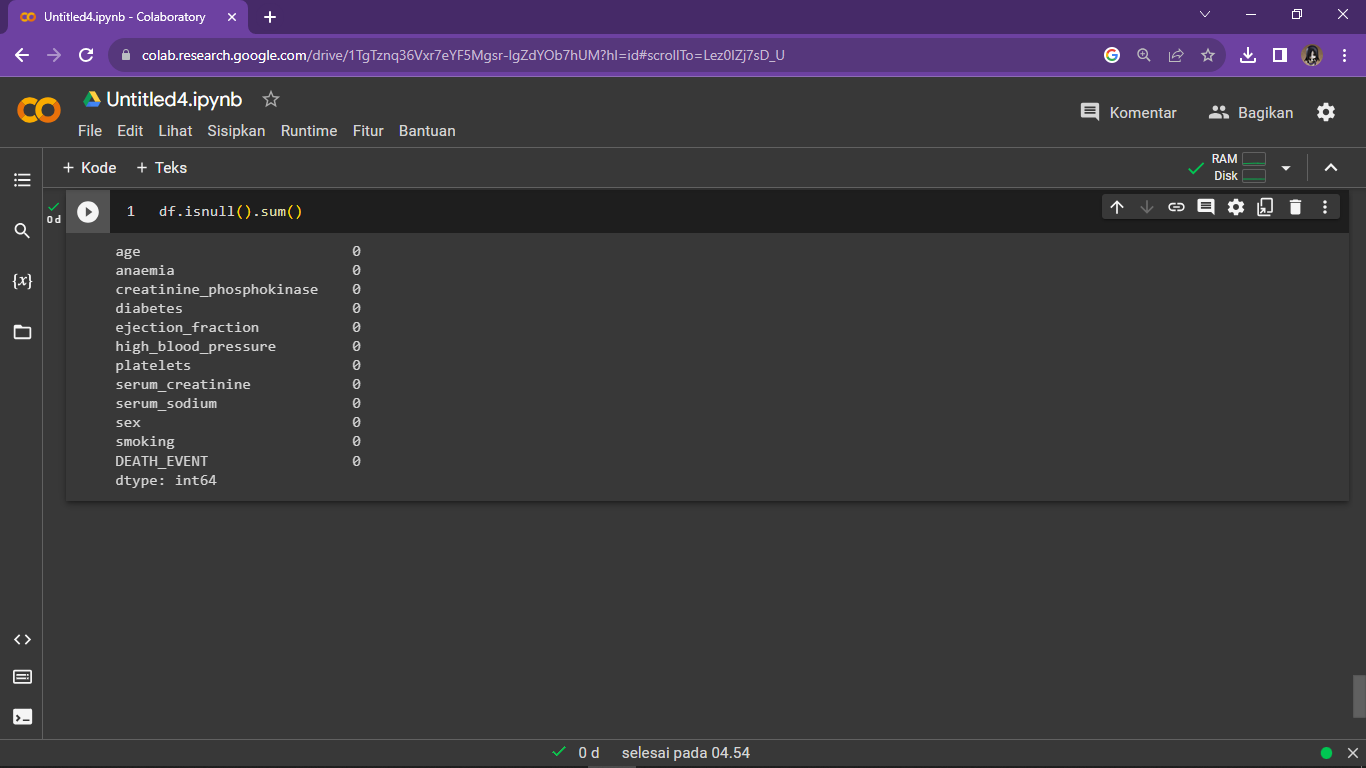
Data yang saya gunakan sudah berupa numerik dan biner, sehingga data sudah bisa di olah



Data yang tidak seimbang karena terdapat 203 kasus negative, dan 96 kasus positif.



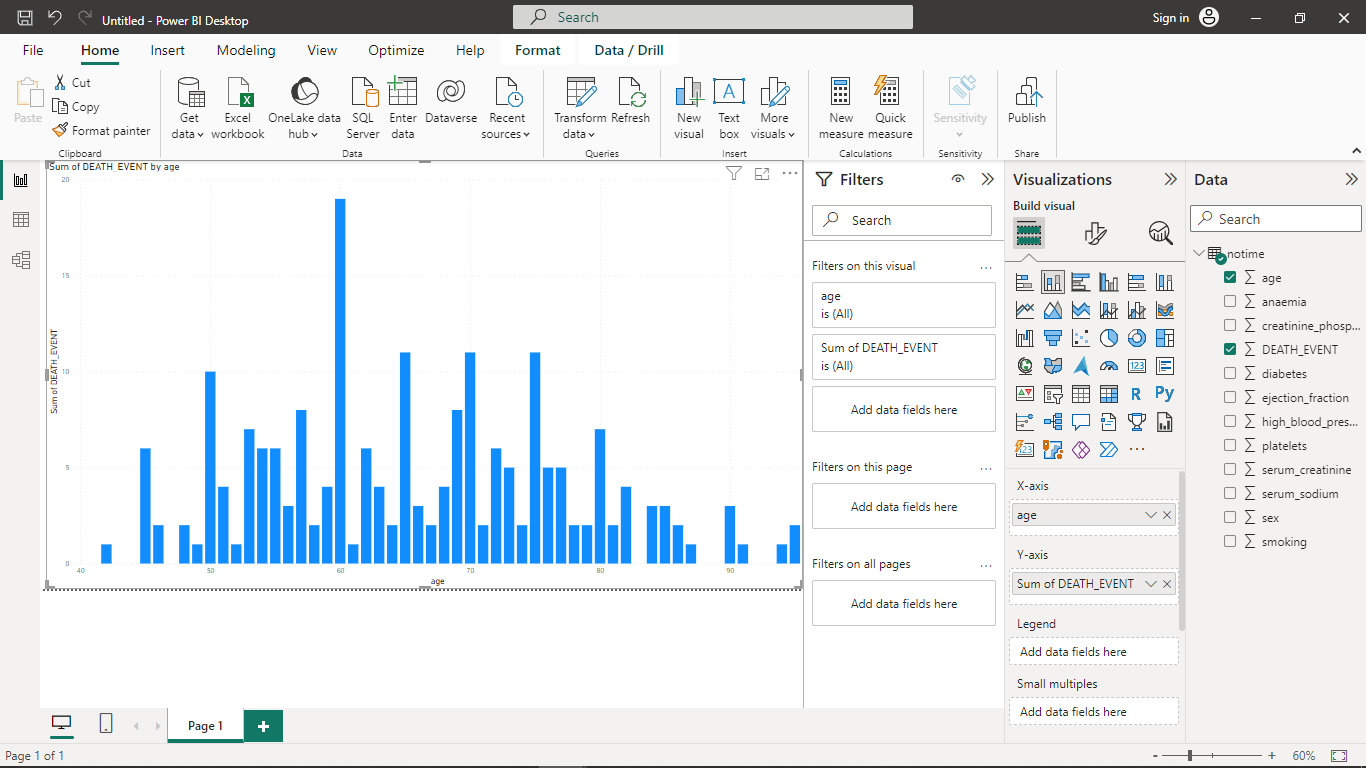
Data yang sebelumnya hanya berjumal 299 dengan 203 kasus negative dan 96 kasus positif. Dengan menggunakan metode over sampling saya membuat data sedikit diperbanyak sehingga total data menjadi 406 data



Dan masing masing kolom tidak terdapat missing value

**Pemilihan Fitur**

1. **age :** umur pasien
2. **anaemia :** apakah ada pengurangan haemoglobin
3. **creatinine\_phospoknase :** level enzim cpk dalam mcg/l
4. **diabetes :** apakah pasien punya riwayat diabetes
5. **ejection\_fraction :** persentase darah yang meninggalkan jantung
6. **high\_blood\_preasure :** apakah pasien punya darah tinggi
7. **platelets :** jumlah platelet di darah dalam kiloplatelets/ml
8. **serum\_creatinine :** level serum creatinine di darah dalam mg/dl
9. **serum\_sodium :** level serum sodium di darah dalam meq/l
10. **sex :** jenis kelamin pasien
11. **smoking :** apakah pasien merokok
12. **DEATH\_EVENT :** apakah pasien sudah meninggal saat waktu follow up



Column chart diatas menunjukkan jumlah contribusi dari variabel age dengan DEATH\_EVEN. Dimana dari chart diatas didapati usia dengan kematian tertinggi adalah 60