**LAPORAN TUGAS**

**ANALISIS DATA HEPATITIS MENGGUNAKAN POWER BI**

Disusun untuk memenuhi tugas Mata Kuliah Proyek Sains Data

Dosen Pengampu: EKA MALA SARI ROCHMAN, S.Kom.M.Kom



**DISUSUN OLEH:**

SYAYID AL AZIZ 210411100008

**PRODI TEKNIK INFORMATIKA**

**JURUSAN TEKNIK INFORMATIKA**

**FAKULTAS TEKNIK**

**UNIVERSITAS TRUNOJOYO MADURA**

**2023**

1. Latar Belakang

Hepatitis adalah istilah umum yang digunakan untuk merujuk pada peradangan hati. Ini bisa disebabkan oleh berbagai faktor, termasuk infeksi virus, konsumsi alkohol berlebihan, penyakit autoimun, paparan zat beracun, atau kondisi medis lainnya. Hepatitis dapat memiliki berbagai latar belakang tergantung pada jenisnya. Hepatitis A, B, C, D, dan E. Gejala hepatitis dapat berupa kelelahan, penyakit kuning (kulit dan mata menguning), urin berwarna gelap, sakit perut, kehilangan nafsu makan, dan gejala mirip flu. Namun, beberapa orang dengan hepatitis mungkin hanya memiliki gejala ringan atau tanpa gejala sama sekali, sehingga deteksi dini dan diagnosis menjadi sulit. Langkah-langkah pencegahan hepatitis termasuk vaksinasi (tersedia untuk hepatitis A dan B), menjaga kebersihan, menghindari perilaku berisiko tinggi seperti berbagi jarum suntik atau melakukan hubungan seks bebas, dan berhati-hati saat bepergian ke daerah dengan prevalensi hepatitis tinggi. Data yang digunakan pada analisis ini berupa data hepatitis dengan beberapa gejala yang umumnya dialami oleh penderitanya.

1. Analisis Data

Data yang digunakan pada analisis ini merupakan data sekunder berupa data hepatitis dengan record data sebanyak 155, data diperoleh pada url halaman ini: <https://www.kaggle.com/datasets/codebreaker619/hepatitis-data>. Berikut merupakan hasil analisis dari fitur atau kolom yang terdapat pada dataset di atas antara lain:

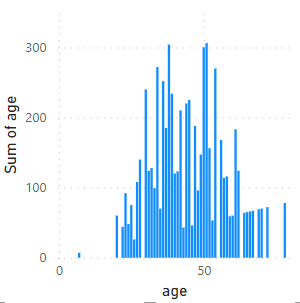
1. age (usia): usia dari penderita hepatitis dengan rentang 7-78 tahun.
2. gender (jenis kelamin): jenis kelamin yaitu male atau female.
3. steroid: penggunaan obat jenis steroid bernilai false atau true.
4. antivirals: penggunaan obat jenis antivirals bernilai false atau true.
5. fatigue: gejala kelelahan bernilai false atau true.
6. malaise: gejala kurang enak badan bernilai false atau true.
7. anorexia: gangguan makan ditandai dengan berat badan menurun bernilai false atau true.
8. liver\_big: pembesaran ulu hati bernilai false atau true.
9. liver\_firm: gejala kondisi hati yang terasa keras atau kaku bernilai true atau false.
10. spleen\_palpable: gangguan pada limpa yang menyentuh area perut.
11. spiders: gejala peningkatan pembuluh darah terjadi pada permukaan kulit bernilai false atau true.
12. ascites: penumpukan cairan berlebih di rongga perut bernilai false atau true.
13. varices: pembengkakan pembuluh darah pada tubuh bernilai false atau true.
14. bilirubin: gejala pemecahan hemoglobin dari sel darah merah bernilai 0.39, 0.80, 1.20, 2.00, 3.00, 4.00.
15. alk\_phosphate: gejala kadar alp yang tinggi pada darah bernilai 33, 80, 120, 160, 200, 250.
16. sgot: gejala kadar sgot yang tinggi pada darah bernilai false atau true.
17. albumin: kadar albumin dalam darah bernilai 2.1, 3.0, 3.8, 4.5, 5.0, 6.0
18. protime: tes lab yang menunjukan seberapa cepat darah dapat membeku bernilai 10, 20, 30, 40, 50, 60, 70, 80, 90
19. histology: pengambilan sampel jaringan hati bernilai false atau true.
20. class: bernilai die atau live.

Hasil analisis di atas kolom-kolom tersebut mencakup beberapa gejala yang telah diketahui dari pasien yang mengidap penyakit hepatitis serta dari beberapa fitur tersebut juga mencakup hasil tes yang telah dilakukan oleh pasien seperti bilirubin, alk phosphate, sgot dan protime. Data tersebut akan menggambarkan berdasarkan gejala yang dialami oleh pasien yang mengidap penyakit hepatitis tersebut apakah mereka dapat bertahan hidup atau meninggal.

1. Visualisasi Data

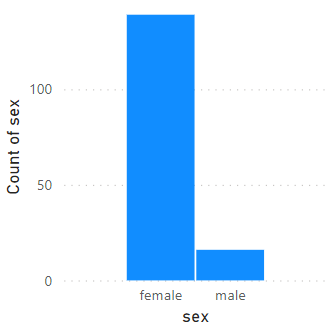
Berdasarkan hasil analisis yang telah dilakukan terdapat 20 fitur atau kolom pada data hepatitis yang akan digunakan berikut merupakan visualisasi data dari masing-masing fitur antara lain:

1. age



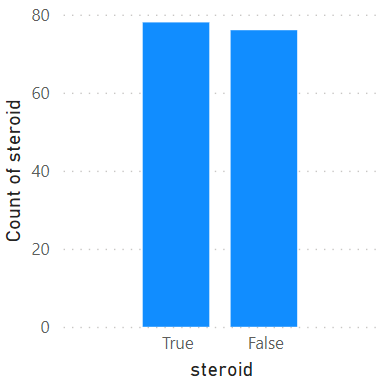
*Gambar 1. Grafik Data Usia*

1. gender



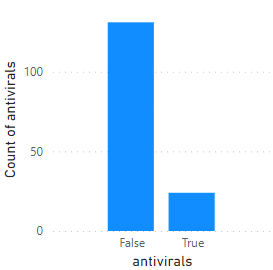
*Gambar 2. Grafik Data Gender*

1. steroid



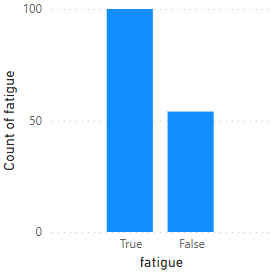
*Gambar 3. Grafik Data Steroid*

1. antiviral



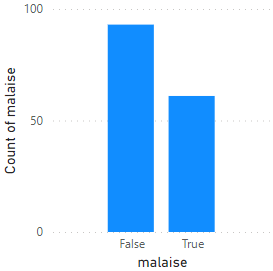
*Gambar 4. Grafik Data Antiviral*

1. fatigue



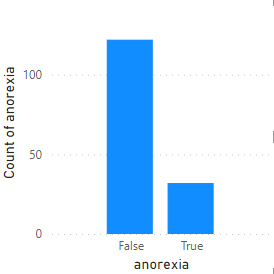
*Gambar 5. Grafik Data Fatigue*

1. malaise



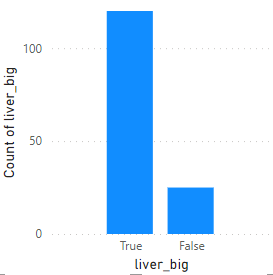
*Gambar 6. Grafik Data Malaise*

1. anorexia



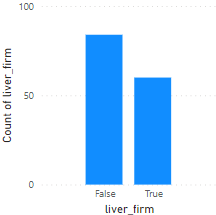
*Gambar 7. Grafik Data Anorexia*

1. liver\_big



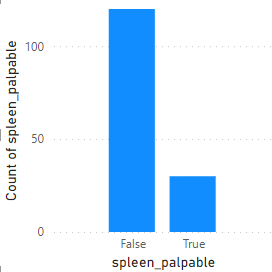
*Gambar 8. Grafik Data Liver Big*

1. liver\_firm



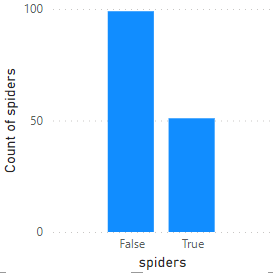
*Gambar 9. Grafik Data Liver Firm*

1. spleen\_palpable



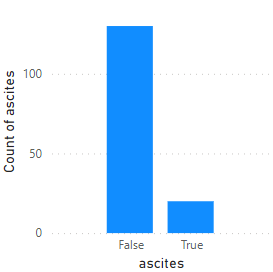
*Gambar 10. Grafik Data Spleen Palpable*

1. spiders



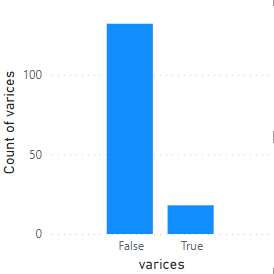
*Gambar 11. Grafik Data Spiders*

1. ascites



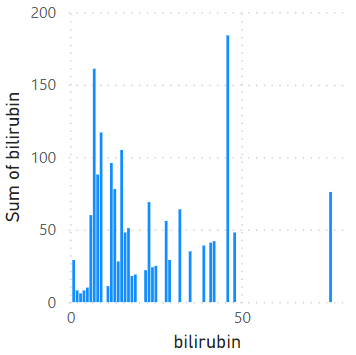
*Gambar 12. Grafik Data Ascites*

1. varices



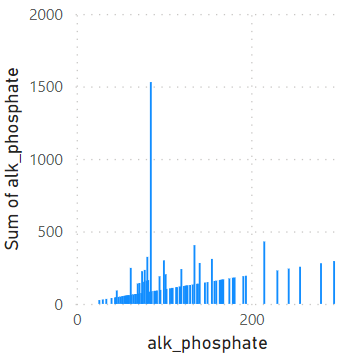
*Gambar 13. Grafik Data Varices*

1. bilirubin



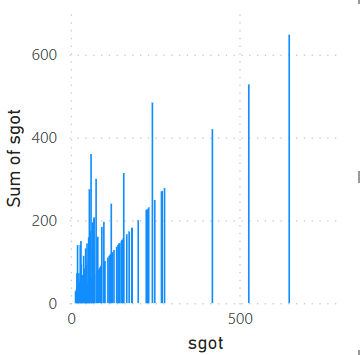
*Gambar 14. Grafik Data Bilirubin*

1. alk\_phospate



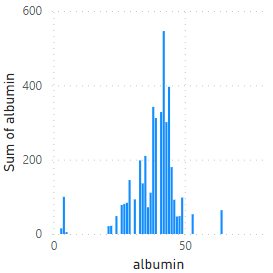
*Gambar 15. Grafik Data Alk Phosphate*

1. sgot



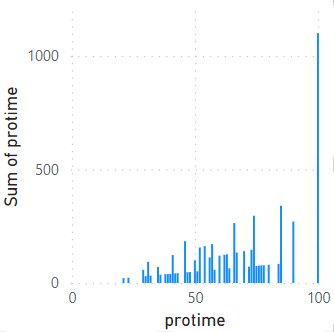
*Gambar 16. Grafik Data Sgot*

1. albumin



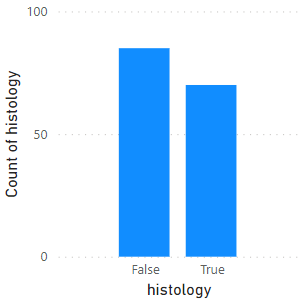
*Gambar 17. Grafik Data Albumin*

1. protime



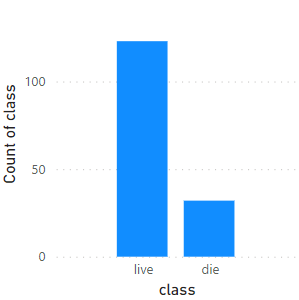
*Gambar 18. Grafik Data Protime*

1. histology



*Gambar 19. Grafik Data Histology*

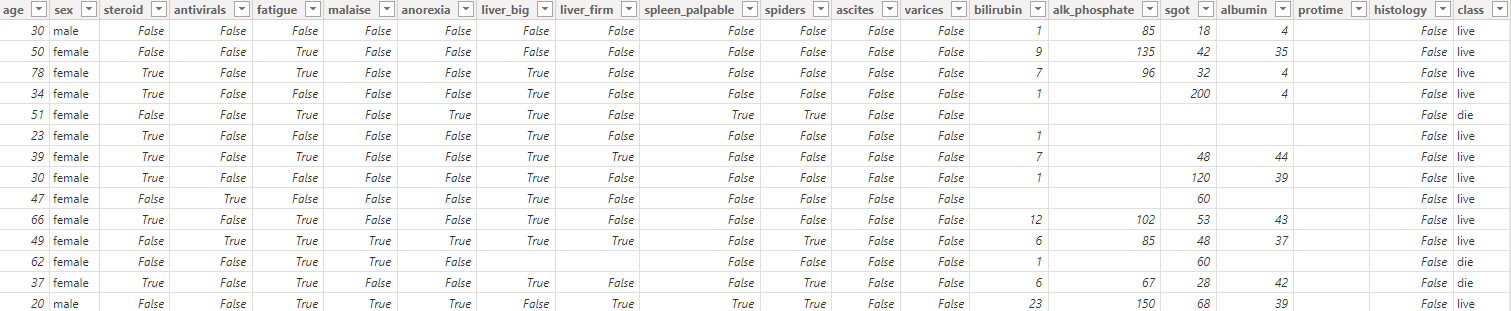
1. class



*Gambar 20. Grafik Data Class*

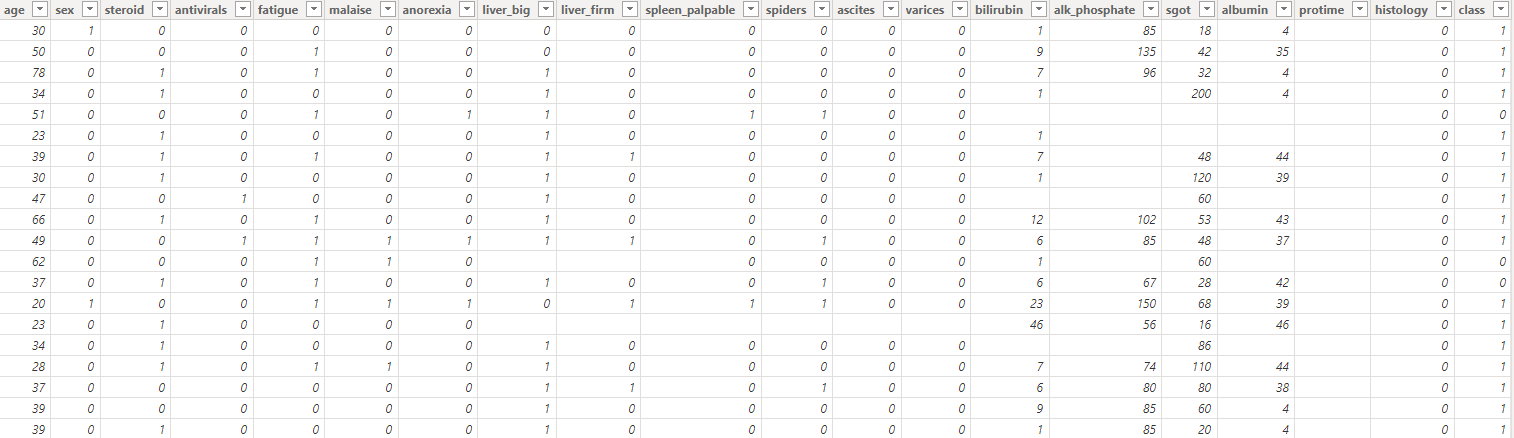
1. Preprocessing Data
2. Data Transformation

Data transformation ini digunakan untuk mengubah atau mentransformasi data dari satu bentuk atau format ke bentuk atau format lainnya. Dataset yang digunakan ini terdapat beberapa data kategorikal yang berupa teks agar dapat tersebut dapat dibaca oleh komputer maka kita ganti data kategorikal tersebut dengan metode biner yaitu 0 dan 1. Berikut merupakan gambar tabel yang belum dilakukan data transformation:



*Gambar 21. Tabel Data Hepatitis Sebelum Transformasi*

Terdapat data kategorik dengan jenis yang sama kecuali pada gender dan class, pada gender nilai female akan diganti dengan 0 dan male diganti dengan 1 sementara pada class die akan diganti dengan 0 dan live diganti dengan dan pada kolom yang memiliki nilai false diganti dengan 0 dan true diganti dengan 1. Berikut merupakan gambar tabel setelah dilakukan data transformation:



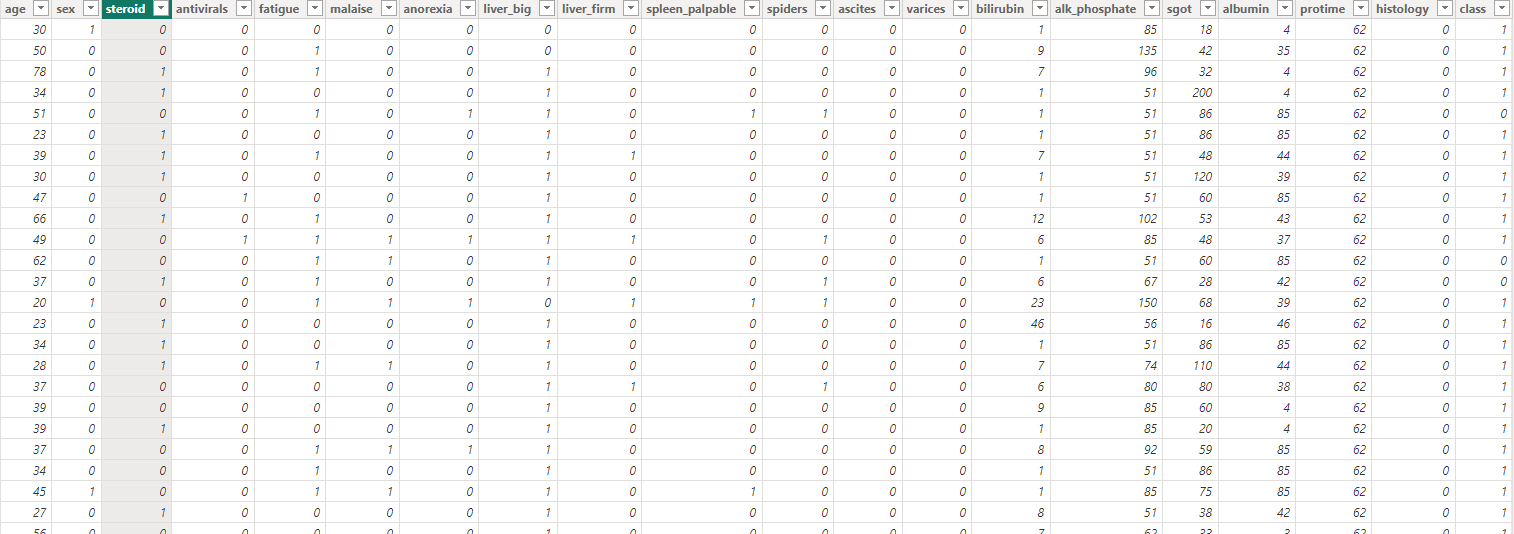
*Gambar 22. Tabel Data Hepatitis Setelah Transformasi*

1. Missing Value

Missing value adalah istilah yang digunakan dalam analisis data untuk merujuk kepada nilai yang hilang atau tidak ada dalam suatu dataset. Missing value terjadi ketika informasi atau data yang seharusnya ada untuk suatu observasi atau variabel tidak ada atau tidak dapat diukur dengan benar. Data yang digunakan ini terdapat beberapa *missing value* yang harus diisikan berdasarkan kriteria data yang ada. Berikut merupakan tabel kolom dan jumlah nilai *missing value* yang ada:

| NO | Kolom | Jumlah Nilai Kosong | Tipe data | Mean atau Mode |
| --- | --- | --- | --- | --- |
| 1 | steroid | 1 | kategorikal | 1 |
| 2 | fatigue | 1 | kategorikal | 1 |
| 3 | malaise | 1 | kategorikal | 0 |
| 4 | anorexia | 1 | kategorikal | 0 |
| 5 | liver\_big | 10 | kategorikal | 1 |
| 6 | liver\_firm | 11 | kategorikal | 0 |
| 7 | spleen\_palpable | 5 | kategorikal | 0 |
| 8 | spiders | 5 | kategorikal | 0 |
| 9 | ascites | 5 | kategorikal | 0 |
| 10 | varices | 5 | kategorikal | 0 |
| 11 | bilirubin | 6 | numerik | 1.43 |
| 12 | alk\_phospate | 29 | numerik | 51.3 |
| 12 | sgot | 4 | numerik | 85.9 |
| 13 | albumin | 16 | numerik | 3.82 |
| 14 | protime | 67 | numerik | 61.9 |

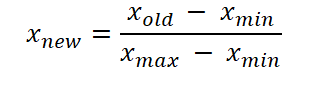
Setelah diketahui nilai *missing value* dari setiap kolom maka data yang termasuk ke dalam kategorikal akan diisikan dengan nilai mode (nilai yang sering muncul) sedangkan data numerik akan diisikan dengan nilai mean (nilai rata-rata) untuk mengisi baris yang terdapat *missing value* tersebut seperti pada rekap tabel diatas. Berikut merupakan gambar tabel setelah dilakukan proses pengisian data pada *missing value*:



*Gambar 22. Tabel Data Hepatitis Setelah Transformasi*

1. Normalisasi (*Min Max Scaler*)

Normalisasi Min-Max adalah metode normalisasi data yang digunakan dalam statistik dan analisis data untuk mengubah rentang nilai dari suatu variabel ke dalam rentang tertentu, biasanya antara 0 dan 1. Tujuan normalisasi Min-Max adalah untuk membuat data memiliki skala yang seragam sehingga memudahkan perbandingan antar variabel yang memiliki unit atau rentang nilai yang berbeda-beda. Terdapat beberapa kolom yang harus dilakukan normalisasi data yaitu age, bilirubin, alk\_phospate, sgot, albumin, protime. Berikut merupakan rumus yang akan digunakan untuk melakukan normalisasi menggunakan *min max* antara lain:



*Gambar 23. Rumus Normalisasi Min Max Scaler*

Keterangan:

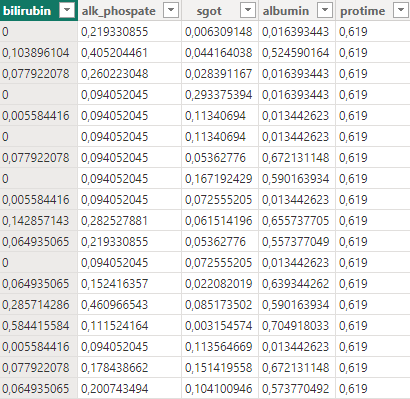
*xnew*: nilai baru yang akan ditambahkan

*xold*: nilai asli pada kolom.

*xmin*: nilai paling rendah pada kolom.

*xmax*: nilai paling tinggi pada kolom.

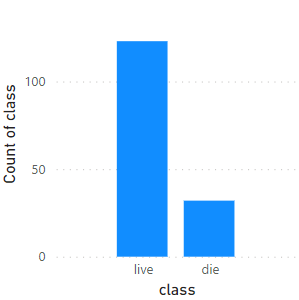
Berikut merupakan gambar tabel hasil normalisasi pada beberapa kolom diatas menggunakan *min max scaler*:



*Gambar 24. Hasil Normalisasi Min Max Scaler*

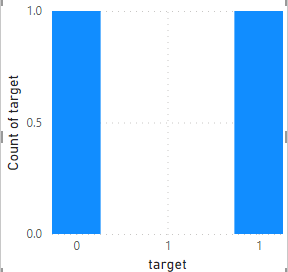
1. Balancing Data

*Balancing data* merupakan tindakan untuk mengatasi ketidakseimbangan dalam jumlah sampel antara kelas atau kategori yang berbeda dalam dataset Anda. Ketidakseimbangan sering terjadi dalam situasi di mana satu kelas atau kategori memiliki jumlah sampel yang jauh lebih banyak daripada yang lain. Ketidakseimbangan ini dapat menyebabkan model pembelajaran mesin kurang mampu mengidentifikasi kelas minoritas. Dataset yang digunakan pada analisis kali ini belum *balance* dikarenakan jumlah data class yaitu live dan die tidak seimbang seperti pada gambar di bawah ini:



*Gambar 25. Grafik Sebaran Data Pada Fitur Class*

Teknik *balancing* data pada kali ini menggunakan algoritma SMOTE (Synthetic Minority Over-sampling Technique), algoritma ini akan bekerja untuk menambahkan sample data yang serupa dengan class die yang termasuk ke dalam minoritas. Berikut merupakan hasil sebaran data ketika sudah dilakukan *balancing data*:



*Gambar 25. Grafik Sebaran Data Balance Fitur Class*

REFERENSI:

Alasadi, S. A., & Bhaya, W. S. (2017). Review of data preprocessing techniques in data mining. Journal of Engineering and Applied Sciences, 12(16), 4102-4107.

ilmudatapy. 3 Metode Normalisasi Data (Feature Scaling) di Python. Diakses pada 14 September 2023 dari https://ilmudatapy.com/metode-normalisasi-data/