LEMBAR PORTOFOLIO UAS

DATA MINING TAHUN AJARAN 2022/2023

NAMA KELOMPOK:

|  |  |
| --- | --- |
| NAMA:  NPM:  FOTO: | NAMA:  NPM:  FOTO: |
| NAMA:  NPM:  FOTO: | NAMA:  NPM:  FOTO: |

1. DATA YANG SAYA BERIKAN DATA APA SAJA?

|  |  |  |
| --- | --- | --- |
|  | **Data 1** | **Data 2** |
| Nama Dataset |  | CreditScore |
| Deskripsikan Data |  | Dataset CreditScore berisi informasi yang dikumpulkan dari lembaga keuangan atau lembaga kredit, seperti bank atau perusahaan kartu kredit.  Data dalam dataset dapat mencakup informasi seperti riwayat pembayaran tagihan, jumlah utang, batas kredit, data akun bank yang dimiliki, riwayat kredit, penghasilan, dan faktor lain yang relevan untuk menghitung skor kredit seseorang (individu). |
| Deskripsikan Atribut data nya |  | **Id** : Kode Unik untuk setiap transaksi, |
| **Customer id** : Kode unik yang merujuk identitas setiap customer |
| **Month** : Menunjukan bulan terjadinya Transaksi atau input data baru |
| **Name** : Nama customer |
| **Age** : Umur |
| **SSN** : angka yang catatan kredit individu. |
| **Occupation** : Profesi seseorang |
| **Annual\_Income** : Pendapatan Tahunan Individu |
| **Monthly\_Inhand\_Salary** : Gaji bersih bulanan individu |
| **Num\_Bank\_Accounts** : Jumlah akun bank yang dimiliki |
| **Num\_Credit\_Card** : Jumlah kartu Kredit (Individu) |
| **Interest\_Rate** : Tingkat suku bunga |
| **Num\_of\_Loan** : Jumlah pengajuan / Penggunaan pinjaman |
| **Type\_of\_Loan** : Jenis Pinjaman yang di berikan |
| **Delay\_from\_due\_date** : Telat membayar tagihan |
| **Num\_of\_Delayed\_Payment** : Jumlah pembayaran yang tertunda |
| **Changed\_Credit\_Limit** : Perubahan batas kredit |
| **Num\_Credit\_Inquiries** : jumlah permintaan pengajuan pinjaman |
| **Credit\_Mix** : Indikator Baik atau bermasalah |
| **Outstanding\_Debt** : Jumlah utang yang belum di lunasi |
| **Credit\_Utilization\_Ratio** : Credit Rasio |
| **Credit\_History\_Age** : Riwayat usia peminjam |
| **Payment\_of\_Min\_Amount** : Pembayaran minimum |
| **Total\_EMI\_per\_month** : jumlah total angsuran bulanan yang harus dibayarkan |
| **Amount\_invested\_monthly** : uang yang di investasikan tiap bulan |
| **Payment\_Behaviour** : kostintensi pembayaran |
| **Monthly\_Balance** : Saldo terakhir yang di miliki customer |
| **Credit\_Score** : Kesimpulan akhir tentang customer |
| Jumlah baris data |  | 100000 Baris |
| Jumlah atribut data |  | 28 columns |
| Jumlah atribut numerik |  | 8 Kolom |
| Jumlah atribut kategorikal |  | 10 Kolom |
| Jelaskan cara penanganan data Kategorikal pada masing-masing atribut kategorikal |  |  |

**PREPROCESSING**

1. Apakah ada missing value pada data tersebut?jika ada silahkan isi tabel berikut!

|  |  |  |
| --- | --- | --- |
|  | **Data 1 (nama\_data)** | **Data 2(CreditScore)** |
| Ada berapa atribut yang memiliki missing value? |  | 8 Atribut |
| Masing-masing atribut ada berapa missing value? |  | 1. Name 9985 2. Monthly\_Inhand\_Salary 15002 3. Type\_of\_Loan 11408 4. Num\_of\_Delayed\_Payment 7002 5. Num\_Credit\_Inquiries 1965 6. Credit\_History\_Age 9030 7. Amount\_invested\_monthly 4479 8. Monthly\_Balance 1200 |
| Apakah ada atribut yang dihapus karena missing value? |  | 1. Name (9.99%) 2. Monthly\_Inhand\_Salary (15.00%) 3. Type\_of\_Loan (11.41%) 4. Num\_of\_Delayed\_Payment (7.00%) 5. Num\_Credit\_Inquiries (1.97%) 6. Credit\_History\_Age (9.03%) 7. Amount\_invested\_monthly (4.48%)   Monthly\_Balance (1.20%)  di bawah 10% maka bisa di isi dengan metode yang tepat, jika di atas 10 harus kita pertimbangkan uulang |

Tangani missing value dengan beberapa metode serta catat performa masing-masing untuk setiap kasus!

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Penanganan | **Performa Data 1** | | | | | **Performa Data 2** | | | | |
| Akurasi | Precision | Recall | Specificity | F score | Akurasi | Precision | Recall | Specificity | F score |
| Mean |  |  |  |  |  |  |  |  |  |  |
| Median |  |  |  |  |  |  |  |  |  |  |
| Modus |  |  |  |  |  |  |  |  |  |  |
| Menghapus baris data yang mengandung missing value |  |  |  |  |  |  |  |  |  |  |
| Penanganganan lain yang anda kerjakan(tuliskan)  ……………………… |  |  |  |  |  |  |  |  |  |  |

1. Apakah ada duplikat data pa data anda?jika ada silahkan isi tabel berikut!

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Penanganan | **Performa Data 1** | | | | | **Performa Data 2** | | | | |
| Akurasi | Precision | Recall | Specificity | F score | Akurasi | Precision | Recall | Specificity | F score |
| Biarkan data duplikat |  |  |  |  |  |  |  |  |  |  |
| Hapus data duplikat |  |  |  |  |  |  |  |  |  |  |

1. Apakah ada outliers data pada data anda?jika ada silahkan isi tabel berikut!

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Penanganan | **Performa Data 1** | | | | | **Performa Data 2** | | | | |
| Akurasi | Precision | Recall | Specificity | F score | Akurasi | Precision | Recall | Specificity | F score |
| Biarkan data outliers |  |  |  |  |  |  |  |  |  |  |
| Hapus outliers |  |  |  |  |  |  |  |  |  |  |

1. Lakukan beberapa percobaan untuk transformasi fitu, selanjutnya cek masing-masing performanya

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Penanganan | **Performa Data 1** | | | | | **Performa Data 2** | | | | |
| Akurasi | Precision | Recall | Specificity | F score | Akurasi | Precision | Recall | Specificity | F score |
| minmax |  |  |  |  |  |  |  |  |  |  |
| desimal scaling |  |  |  |  |  |  |  |  |  |  |
| Z score |  |  |  |  |  |  |  |  |  |  |

1. Apakah ada kasus Multicollinearity pada data yang anda dapat, deskripsikan singkat dalam bentuk paragraf, apa saja atribut yang berpotensi multicollinearity dan bagaimana cara cek serta penanganannya?

|  |  |  |
| --- | --- | --- |
|  | **Data 1** | **Data 2** |
| Nama Dataset |  |  |
| Uraian |  |  |
| PERFORMA SETELAH DILAKUKAN PENANGANAN MULTICOLLINIEARITY JIKA ADA | Akurasi:  Precision:  Recall:  F.Score: | Akurasi:  Precision:  Recall:  F.Score: |

SELANJUTNYA DARI DATA YANG SAMA LAKUKAN CLUSTERING (TARGET DIHAPUS) DENGAN EKSPERIMEN HAMPIR SAMA DENGAN SEBELUMNYA, SELANJUTNYA CEK PERFORMA MASING-MASING HASIL EKSPERIMEN!

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Penanganan | **Performa Data 1** | | **Performa Data 2** | |
| DBI | silhouette score | DBI | silhouette score |
| Mean |  |  |  |  |
| Median |  |  |  |  |
| Modus |  |  |  |  |
| Menghapus baris data yang mengandung missing value |  |  |  |  |
| Penanganganan lain yang anda kerjakan(tuliskan)  ……………………… |  |  |  |  |

1. Apakah ada duplikat data pa data anda?jika ada silahkan isi tabel berikut!

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Penanganan | **Performa Data 1** | | **Performa Data 2** | |
| DBI | silhouette score | DBI | silhouette score |
| Biarkan data duplikat |  |  |  |  |
| Hapus data duplikat |  |  |  |  |

1. Apakah ada outliers data pada data anda?jika ada silahkan isi tabel berikut!

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Penanganan | **Performa Data 1** | | **Performa Data 2** | |
| DBI | silhouette score | DBI | silhouette score |
| Biarkan data outliers |  |  |  |  |
| Hapus outliers |  |  |  |  |

1. Lakukan beberapa percobaan untuk transformasi fitu, selanjutnya cek masing-masing performanya

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Performa Data 1** | | **Performa Data 2** | |
| minmax | DBI | silhouette score | DBI | silhouette score |
| desimal scaling |  |  |  |  |
| Z score |  |  |  |  |
|  | | | | |

1. Apakah ada kasus Multicollinearity pada data yang anda dapat,

|  |  |  |
| --- | --- | --- |
|  | **Data 1** | **Data 2** |
| PERFORMA SETELAH DILAKUKAN PENANGANAN MULTICOLLINIEARITY JIKA ADA | DBI:  Silhouette score: | DBI:  Silhouette score: |