LEMBAR PORTOFOLIO UAS

DATA MINING TAHUN AJARAN 2022/2023

NAMA KELOMPOK:

|  |  |
| --- | --- |
| NAMA: RICKY AGUNG SUMIRANTO  NPM: 2013020101  FOTO: | NAMA:IKA MARIA DANIATI  NPM:2013020085  FOTO: |
| NAMA:ANAS TASIA  NPM:  FOTO: | NAMA:MUHAMMAD YUSUF EFENDI  NPM:  FOTO: |

1. DATA YANG SAYA BERIKAN DATA APA SAJA?

|  |  |  |
| --- | --- | --- |
|  | **Data 1** | **Data 2** |
| Nama Dataset | Mobil Prize | CreditScore |
| Deskripsikan Data | Data ini berisi daftar spesifikasi handphone.Data ini bertujuan untuk mengetahui hubungan antara fitur-fitur ponsel(misalnya RAM, Memori internal dll)dan harga jualnya. | 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 | 1. Battery power : total energy batrai yang diukur dalam satuan mAh. 2. Blue:Bluethooth. 3. Clock speed : kecepatan prossesor 4. Dual sim : ketersediaan fitur dual sim pada handphone. 5. FC :ukuran kamera depan dalam mega pixel. 6. Four-G: ketersediaan fitur 4G pada handphone. 7. Int-memory :besarnya memori internal (GB). 8. M-dep: Mobile depth(CM) 9. Mobile\_wt: Berat hp. 10. N-cores:jumlah core pada prossesor 11. Pc: Primary Camera Megapixels: Merepresentasikan resolusi kamera utama ponsel dalam megapiksel (MP). 12. Px\_height: Pixel Resolution Height Merupakan tinggi resolusi layar ponsel dalam piksel. 13. Px\_width: Pixel Resolution Width: Merupakan lebar resolusi layar ponsel dalam piksel. 14. Ram: RAM: Merupakan kapasitas memori RAM ponsel dalam megabita (MB). 15. Sc\_h: Screen Height (cm): Merupakan tinggi layar ponsel dalam sentimeter (cm). 16. Sc\_w: Screen Width (cm): Merupakan lebar layar ponsel dalam sentimeter (cm). 17. Talk\_time: Talk Time (hours): Mewakili estimasi waktu bicara yang dapat dihasilkan oleh ponsel dalam jam. 18. Tree\_g: Menunjukkan apakah ponsel mendukung jaringan 3G atau tidak. Nilai 1 menunjukkan adanya dukungan, sementara nilai 0 menunjukkan tidak adanya dukungan. 19. Touch\_screen: Mengindikasikan apakah ponsel memiliki layar sentuh atau tidak. Nilai 1 menunjukkan adanya layar sentuh, sementara nilai 0 menunjukkan tidak adanya layar sentuh. 20. Wifi: Atribut ini menunjukkan apakah ponsel memiliki fitur Wi-Fi atau tidak. Jika nilainya 1, berarti ponsel memiliki Wi-Fi; jika 0, berarti tidak ada fitur Wi-Fi. 21. Price\_range: ni adalah atribut target yang menunjukkan kategori harga ponsel. Nilai atribut ini dapat berupa 0, 1, 2, atau 3, yang mewakili rentang harga yang berbeda untuk ponsel (misalnya, 0 untuk harga rendah, 3 untuk harga tinggi). | **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 | 2000 | 100000 Baris |
| Jumlah atribut data | 21 | 28 columns |
| Jumlah atribut numerik | 21 | 8 Kolom |
| Jumlah atribut kategorikal | 0 | 10 Kolom |
| Jelaskan cara penanganan data Kategorikal pada masing-masing atribut kategorikal | Tidak ditangani karena pada data ini tidak terdapat data 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? | 0 | 8 Atribut |
| Masing-masing atribut ada berapa missing value? | 0 | 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? | Tidak ada | 1. Name (9.99%) 2. Type\_of\_Loan (11.41%) |

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 | - | - | - | - | - | 69.92% | 69.90% | 69.92% | 70.38% | 69.91% |
| Median | - | - | - | - | - | 69.87% | 69.89% | 69.87% | 70.51% | 69.88% |
| Modus | - | - | - | - | - | 71.41% | 71.23% | 71.41% | 71.59% | 71.20% |
| Menghapus baris data yang mengandung missing value | - | - | - | - | - | 70.65% | 70.49% | 70.65% | 68.84% | 70.40% |
| Dibiarkan Saja | - | - | - | - | - | 71.62% | 71.45% | 71.62% | 71.78% | 71.43% |

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 | 0.665 | 0.6729189742989287 | 0.665 | Kelas 0: 0.08143322475570032 Kelas 1: 0.12786885245901639 Kelas 2: 0.14473684210526316 Kelas 3: 0.09154929577464789 | 0.6680003918825577 | 74.97% | 74.74% | 74.97% | 80.33% | 74.79% |
| Hapus outliers | 0.696969696969697 | 0.7104752837683274 | 0.696969696969697 | Kelas 0: 0.06040268456375839 Kelas 1: 0.1342281879194631 Kelas 2: 0.12794612794612795 Kelas 3: 0.08135593220338982 | 0.696969696969697 | 75.07% | 74.84% | 75.07% | 80.14% | 74.79% |

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 | 0.9090909090909091 | 0.9076286292025971 | 0.9090909090909091 | Kelas 0: 0.06040268456375839  Kelas 1: 0.1342281879194631  Kelas 2: 0.12837837837837837  Kelas 3: 0.08445945945945946 | 0.9083591806512736 | 0.5901639344262295 | 0.4844997898276587 | 0.5901639344262295 | 0.20898136471906967 | 0.5114314559052793 |
| desimal scaling | 0.9090909090909091 | 0.9076286292025971 | 0.9090909090909091 | Kelas 0: 0.06040268456375839 Kelas 1: 0.1342281879194631 Kelas 2: 0.12837837837837837 Kelas 3: 0.08445945945945946 | 0.9083591806512736 | 0.5901639344262295 | 0.47363817810768993 | 0.5901639344262295 | 0.21582491582491584 | 0.5192818110850898 |
| Z score | 0.9595959595959596 | 0.9595894872490618 | 0.9595959595959596 | Spesifisitas: Kelas 0: 0.06040268456375839 Kelas 1: 0.1342281879194631 Kelas 2: 0.12837837837837837 Kelas 3: 0.08445945945945946 F1-Score: 0.9595927234115968 | 0.9802194484562156 | 0.6721311475409836 | 0.6635893011216566 | 0.6721311475409836 | 0.16434235712332515 | 0.6528460528949885 |

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 | Mobile prize |  |
| Uraian | Dari hasil pengecekan multicollinearity menggunakan vif menujukkan bahwa **tidak** terdapat nilai vif yang tinggi ,nilai vif yang tinggi biasanya diatas 5 atau 10. |  |
| 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 | 4.708494751646101 | 0.045035546576833625 |  |  |
| Z score | 4.708494751646101 | 0.045035546576833625 |  |  |
|  | | | | |

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: |