**UJIAN AKHIR SEMESTER**

**BIG DATA 2021/2022**

**Nomor 1. Klasifikasi Data**

Unduh data dari <https://www.kaggle.com/fedesoriano/heart-failure-prediction>.

1. Buatlah EDA menggunakan teknik visualisasi data. Kemudian jelaskan hasil dari EDA tersebut.
2. Apa pra-proses yg cocok dilakukan untuk dataset tersebut.
3. Pilih dua metode pembagian data. Kemudian jelaskan alasan menggunakan metode tersebut.
4. Pilih dua metode klasifikasi data. Kemudian jelaskan alasan menggunakan metode tersebut.
5. Hitung nilai akurasi, presisi, recall.

Template jawaban.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Klasifikasi Data | Pembagian data | Confusion Matrix | | |
| Akurasi | Presisi | Recall |
| Metode klasifikasi 1 | Metode Pembagian data 1 | ? | ? | ? |
| Metode Pembagian data 2 | ? | ? | ? |
| Metode klasifikasi 2 | Metode Pembagian data 1 | ? | ? | ? |
| Metode Pembagian data 2 | ? | ? | ? |

**Nomor 2. Klastering Data**

Unduh data dari <https://www.kaggle.com/uciml/iris>.

1. Apa pra-proses yg cocok dilakukan untuk dataset tersebut.
2. Pilih dua metode pembagian data. Kemudian jelaskan alasan menggunakan metode tersebut.
3. Pilih dua metode menghitung jarak antar data. Kemudian jelaskan alasan menggunakan metode tersebut.
4. Pilih dua metode klasifikasi data. Kemudian jelaskan alasan menggunakan metode tersebut.
5. Hitung nilai SSE dan Centroid.

Template jawaban.

|  |  |  |
| --- | --- | --- |
| Percobaan 1. Metode Klastering ke-1 | | |
|  | Metode jarak antar data 1 | Metode jarak antar data 2 |
| Metode Pembagian data 1 | SSE dan Centroid …? | SSE dan Centroid …? |
| Metode Pembagian data 2 | SSE dan Centroid …? | SSE dan Centroid …? |
|  |  |  |
| Percobaan 2. Metode Klastering ke-2 | | |
|  | Metode jarak antar data 1 | Metode jarak antar data 2 |
| Metode Pembagian data 1 | SSE dan Centroid …? | SSE dan Centroid …? |
| Metode Pembagian data 2 | SSE dan Centroid …? | SSE dan Centroid …? |

**Nomor 3. Menggunakan dataset dari tugas 3 “Praproses Data "Model Prediksi Deret Waktu Titik Panas dengan Memperhatikan Faktor Iklim".**

Dimana percobaan yang dilakukan dari sisi datatset.

* Dataset pertama, var terikat dari FIRMS NASA. Var bebas dari BMKG. (Sesuai dengan dataset tugas 3)
* Dataset kedua, var terikat dari FIRMS NASA. Var bebas dari CRU TS v4.05 (https://catalogue.ceda.ac.uk/uuid/c26a65020a5e4b80b20018f148556681)

Nomor 3. Prediksi Deep Learning

1. Apa pra-proses yg cocok dilakukan untuk dataset tersebut.
2. Pilih satu metode pembagian data. Kemudian jelaskan alasan menggunakan metode tersebut.
3. Tentukan nilai arsitektur NN. Kemudian jelaskan alasan mengapa arsitektur NN dibuat seperti itu. Adapun arsitektur NN terdiri dari jumlah input layer, jumlah hidden layer, jumlah output layer, jumlah neuron, epoch, batch size, learning rate, momentum, activasi function, optimizers. Minimal jumlah kombinasi dari arsitektur NN akan menghasilkan 480 model prediksi. Kemudian pilih salah satu dari 480 model tersebut menggunakan metode hyperparameter tuning grid serach.
4. Metode Deep Learning menggunakann LSTM-RNN.
5. Hitung nilai RMSE dan waktu komputasi.

Template jawaban.

Dataset 1.

* Metode pembagian data x. Persentasi data latih x% dan data uji x%
* Aristektur NN
  + …
  + …
  + …
* Nilai RMSE dan Waktu komputasi (Saat memilih model terbaik dari 480 model)

Dataset 2.

* Metode pembagian data x. Persentasi data latih x% dan data uji x%
* Aristektur NN
  + …
  + …
  + …
* Nilai RMSE dan Waktu komputasi (Saat memilih model terbaik dari 480 model)

Nomor 1 dan 2. Yg dibandingkan adalah metode

Nomor 3 yg dibandingkan adalah dataset nya.