

## FR.IA.02. TUGAS PRAKTIK DEMONSTRASI

<b>Skema Sertifikasi</b> (KKNI/Okupasi/Klaster)	Judul	:	<b>Associate Data Scientist</b>
	Nomor	:	<b>SKM/0317/00010/2/2022/33</b>
TUK	:	<input type="checkbox"/> Sewaktu <input type="checkbox"/> Tempat Kerja <input type="checkbox"/> Mandiri	
Nama Asesor	:		
Nama Asesi	:		
Tanggal	:		

☒ yang perlu

### A. Petunjuk

1. Baca dan pelajari setiap instruksi kerja di bawah ini dengan cermat sebelum melaksanakan praktek
2. Klarifikasi kepada Asesor apabila ada hal-hal yang belum jelas
3. Laksanakan pekerjaan sesuai dengan urutan proses yang sudah ditetapkan
4. Seluruh proses kerja mengacu kepada SOP/WI yang dipersyaratkan
5. Waktu pengerjaan tugas praktek adalah **180 menit / 3 jam**

### B. Alat dan Bahan

- a. Perangkat komputer dengan spesifikasi minimum:
  - CPU Intel Pentium Core i3 3.1 GHz atau sederajat
  - RAM 4 GB
  - Free space 10 GB
  - Resolusi layar 1024 x 768 pixel
- b. Operating System minimum:
  - Windows 8
  - 64-bit macOS 10.13+
  - Linux, including Ubuntu, RedHat, CentOS 7+
- c. Perangkat lunak:
  - Tools Pengolah Kata dan Spreadsheet
  - Jupyter Notebook
  - Google Colaboratory
  - R
  - RapidMiner
  - Weka
  - *Data science tools* lainnya
- d. Alat tulis

### C. Skenario

Untuk memenuhi kualifikasi sebagai seorang Associate Data Scientist, anda diminta untuk mendemonstrasikan pekerjaan berikut:

- 1) Mengumpulkan data
- 2) Menelaah data
- 3) Memvalidasi data
- 4) Menentukan objek data
- 5) Membersihkan data
- 6) Mengkonstruksi data
- 7) Menentukan label data
- 8) Membangun model
- 9) Mengevaluasi hasil pemodelan

Lakukan pekerjaan berdasarkan skenario tugas praktek (**Terlampir**). Unduh/download **Buku Kerja/Job Sheet** pada link berikut:

[https://drive.google.com/file/d/1okVXXIIIbToqe0oQsKw3\\_cUE6MxrVarW/view?usp=sharing](https://drive.google.com/file/d/1okVXXIIIbToqe0oQsKw3_cUE6MxrVarW/view?usp=sharing)

Dokumentasikan hasil pekerjaan pada Buku Kerja tersebut.

### D. Langkah Kerja

Berdasarkan skenario tugas praktek, lakukan langkah-langkah kerja berikut:

#### 1) Mengumpulkan Data

Peralatan: Komputer

Perlengkapan: Aplikasi pengolah kata, spreadsheet

Durasi Waktu: 20 menit

No.	Langkah Kerja	Instruksi Kerja
1.	Menentukan kebutuhan data	1.1 Identifikasi kebutuhan data sesuai tujuan teknis data science. 1.2 Periksa kebutuhan data diperiksa ketersediaannya sesuai aturan yang berlaku. 1.3 Tentukan kebutuhan data volumenya sesuai tujuan teknis data science.
2.	Mengambil data	2.1 Identifikasi metode dan tools pengambilan data sesuai tujuan teknis data science. 2.2 Tentukan tools pengambilan data sesuai tujuan teknis data science. 2.3 Siapkan tools pengambilan data sesuai tujuan teknis data science. 2.4 Jalankan proses pengambilan data sesuai dengan tools yang telah disiapkan.
3.	Mengintegrasikan data	3.1 Periksa integritas data sesuai tujuan teknis data science. 3.2 Integrasikan data sesuai tujuan teknis data science.

#### 2) Menelaah Data

Peralatan: Komputer

Perlengkapan: Aplikasi pengolah kata, tools pengolahan data, tools pembuat grafik

Durasi Waktu: 20 menit

No.	Langkah Kerja	Instruksi Kerja
1.	Menganalisis tipe dan relasi data	1.1 Identifikasi tipe data yang terkumpul sesuai tujuan teknis. 1.2 Uraikan nilai atribut data yang terkumpul sesuai dengan batasan konteks bisnisnya.

		1.3 Identifikasi relasi antar data yang terkumpul sesuai dengan tujuan teknis.
2.	Menganalisis karakteristik data	2.1 Sajikan karakteristik data yang terkumpul dengan deskripsi statistik dasar. 2.2 Sajikan karakteristik data yang terkumpul dengan visualisasi grafik. 2.3 Analisis hasil penyajian data karakteristiknya untuk telaah data.
3.	Membuat laporan telaah data	3.1 Dokumentasikan hasil analisis dalam bentuk laporan sesuai dengan tujuan teknis. 3.2 Susun hipotesis berdasar hasil analisis sesuai tujuan teknis data science.

### 3) Memvalidasi Data

Peralatan: Komputer

Perlengkapan: Aplikasi pengolah kata, tools pengolahan data, tools pembuat grafik

Durasi Waktu: 20 menit

No.	Langkah Kerja	Instruksi Kerja
1.	Melakukan pengecekan kelengkapan data	1.1 Sajikan penilaian kualitas data dari hasil telaah sesuai tujuan teknis data science. 1.2 Sajikan penilaian tingkat kecukupan data dari hasil telaah sesuai tujuan teknis data science.
2.	Membuat rekomendasi kelengkapan data	2.1 Susun rekomendasi hasil penilaian kualitas sesuai tujuan teknis data science. 2.2 Susun rekomendasi hasil penilaian kecukupan data sesuai tujuan teknis data science.

### 4) Menentukan Objek Data

Peralatan: Komputer

Perlengkapan: Aplikasi pengolah kata, spreadsheet, text editor, SQL

Durasi Waktu: 15 menit

No.	Langkah Kerja	Instruksi Kerja
1.	Memutuskan kriteria dan teknik pemilihan data	1.1 Identifikasi kriteria pemilihan data sesuai dengan tujuan teknis dan aturan yang berlaku. 1.2 Tetapkan teknik pemilihan data sesuai dengan kriteria pemilihan data.
2.	Menentukan attributes (columns) dan records (row) data	2.1 Identifikasi attributes (columns) data sesuai dengan kriteria pemilihan data. 2.2 Identifikasi records (row) data sesuai dengan kriteria pemilihan data.

### 5) Membersihkan Data

Peralatan: Komputer

Perlengkapan: Aplikasi pengolah kata, spreadsheet, text editor, SQL

Durasi Waktu: 15 menit

No.	Langkah Kerja	Instruksi Kerja
1.	Melakukan pembersihan data yang kotor	1.1 Tentukan strategi pembersihan data berdasarkan hasil telaah data 1.2 Koreksi data yang kotor berdasarkan strategi pembersihan data
2.	Membuat laporan dan rekomendasi hasil membersihkan data	2.1 Deskripsikan masalah dan teknis koreksi data sesuai dengan kondisi data dan strategi pembersihan data 2.2 Lakukan evaluasi berdasarkan analisis koreksi yang telah dilakukan 2.3 Dokumentasikan evaluasi proses dan hasil pembersihan data kotor

## 6) Mengkonstruksi Data

Peralatan: Komputer

Perlengkapan: Aplikasi pengolah kata, tools pengolah data

Durasi Waktu: 20 menit

No.	Langkah Kerja	Instruksi Kerja
1.	Menganalisis teknik transformasi data	1.1 Analisis data untuk menentukan representasi fitur data awal. 1.2 Analisis representasi fitur data awal untuk menentukan teknik rekayasa fitur yang diperlukan untuk pembangunan model data science.
2.	Melakukan transformasi data	2.1 Lakukan transformasi untuk mendapatkan fitur data awal. 2.2 Lakukan rekayasa fitur data untuk mendapatkan fitur baru yang diperlukan untuk pembangunan model data science.
3.	Membuat dokumentasi konstruksi data	3.1 Jabarkan teknis transformasi data dalam bentuk tertulis. 3.2 Tuangkan hasil transformasi data dan rekomendasi hasil transformasi dalam bentuk tertulis.

## 7) Menentukan Label Data

Peralatan: Komputer

Perlengkapan: Aplikasi pengolah kata, aplikasi pelabelan data

Durasi Waktu: 20 menit

No.	Langkah Kerja	Instruksi Kerja
1.	Melakukan pelabelan data	1.1 Uraikan kesesuaian antara analisis hasil pelabelan data sejenis yang sudah ada dengan Standard Operating Procedure (SOP) pelabelan 1.2 Lakukan pelabelan data sesuai dengan SOP pelabelan
2.	Membuat laporan hasil pelabelan data	2.1 Uraikan statistik hasil pelabelan pada laporan 2.2 Uraikan evaluasi proses pelabelan pada laporan

## 8) Membangun Model

Peralatan: Komputer

Perlengkapan: Perangkat lunak data science di antaranya: rapid miner, weka, atau development untuk bahasa pemrograman tertentu seperti python atau R.

Durasi Waktu: 30 menit

No.	Langkah Kerja	Instruksi Kerja
1.	Menyiapkan parameter model	1.1 Identifikasi parameter-parameter yang sesuai dengan model. 1.2 Tetapkan nilai toleransi parameter evaluasi pengujian sesuai dengan tujuan teknis.
2.	Menggunakan tools pemodelan	2.1 Identifikasi tools untuk membuat model sesuai dengan tujuan teknis data science. 2.2 Bangun algoritma untuk teknik pemodelan yang ditentukan menggunakan tools yang dipilih. 2.3 Eksekusi algoritma pemodelan sesuai dengan skenario pengujian dan tools untuk membuat model yang telah ditetapkan. 2.4 Optimasi parameter model algoritma untuk menghasilkan nilai parameter evaluasi yang sesuai dengan skenario pengujian.

## 9) Mengevaluasi Hasil Pemodelan

Peralatan: Komputer

Perlengkapan: Tools untuk mengeksekusi model, tools untuk pengumpulan data riil

Durasi Waktu: 20 menit

No.	Langkah Kerja	Instruksi Kerja
1.	Menggunakan model dengan data riil	1.1 Kumpulkan data baru untuk evaluasi pemodelan sesuai kebutuhan yang mengacu kepada parameter evaluasi 1.2 Uji model dengan menggunakan data riil yang telah dikumpulkan
2.	Menilai hasil pemodelan	2.1 Nilai keluaran pengujian model berdasarkan metrik kesuksesan 2.2 Dokumentasikan hasil penilaian sesuai standar yang berlaku

-- oOo --

## FR.IA.02A. SKENARIO TUGAS PRAKTIK DEMONSTRASI

<b>Skema Sertifikasi</b> (KKM/Okupasi/Klaster)	Judul	:	<b>Associate Data Scientist</b>
	Nomor	:	<b>SKM/0317/00010/2/2022/33</b>
TUK	:	<input type="checkbox"/> Sewaktu <input type="checkbox"/> Tempat Kerja <input type="checkbox"/> Mandiri	
Nama Asesor	:		
Nama Asesi	:		
Tanggal	:		

☒ yang perlu

### A. Tujuan Bisnis Data Science

*Cardiovascular disease* (CVD) atau penyakit kardiovaskuler merupakan salah satu penyebab kematian nomor satu secara global. Diestimasi telah merenggut nyawa sekitar 18 juta jiwa setiap tahunnya atau 31% dari seluruh penyebab kematian secara global. Empat dari lima kematian CVD diakibatkan oleh serangan jantung dan stroke, dan sepertiganya dialami secara prematur oleh seseorang dengan usia di bawah 70 tahun.

Kegagalan jantung merupakan salah satu penyebab umum dari penyakit kardiovaskuler. Sebagai seorang associate data scientist, Anda ditugaskan untuk melakukan prediksi terhadap kemungkinan terjadinya penyakit jantung. Diharapkan pemodelan yang dihasilkan dapat membantu kinerja dokter dalam mendeteksi pasien dengan penyakit jantung, sehingga dapat menolong lebih banyak orang yang berisiko tinggi terkena penyakit kardiovaskuler.

### B. Tujuan Teknis Data Science

Membuat model klasifikasi menggunakan **decision tree** dan **satu model pilihan lainnya**, untuk memprediksi apakah seorang pasien mengidap penyakit jantung atau tidak. Berdasarkan hasil pemodelan data, anda diminta untuk **mengevaluasi kinerja model klasifikasi** dan **menginterpretasikan fitur-fitur mana yang paling berpengaruh** terhadap hasil prediksi fitur target.

Kinerja model diukur menggunakan metrik berikut:

- Confusion matrix
- Accuracy
- Precision
- Recall
- F1-score

### C. Metrik Kesuksesan

Ukuran keberhasilan pengembangan model klasifikasi: **accuracy > 70%**

## D. Sumber Dataset

Dataset dapat diakses pada link github: <https://github.com/arubhasy/dataset/blob/main/heart.csv>

## E. Informasi Fitur Data

No	Attribute	Description	Role
1.	Age	age of the patient [years]	<i>predictor (independent) variables</i>
2.	Sex	sex of the patient [M: Male, F: Female]	<i>predictor (independent) variables</i>
3.	ChestPainType	chest pain type [TA: Typical Angina, ATA: Atypical Angina, NAP: Non-Anginal Pain, ASY: Asymptomatic]	<i>predictor (independent) variables</i>
4.	RestingBP	resting blood pressure [mm Hg]	<i>predictor (independent) variables</i>
5.	Cholesterol	serum cholesterol [mm/dl]	<i>predictor (independent) variables</i>
6.	FastingBS	[1: if FastingBS > 120 mg/dl, 0: otherwise]	<i>predictor (independent) variables</i>
7.	RestingECG	fasting blood sugar [1: if FastingBS > 120 mg/dl, 0: otherwise]	<i>predictor (independent) variables</i>
8.	MaxHR	resting electrocardiogram results [Normal: Normal, ST: having ST-T wave abnormality (T wave inversions and/or ST elevation or depression of > 0.05 mV), LVH: showing probable or definite left ventricular hypertrophy by Estes' criteria]	<i>predictor (independent) variables</i>
9.	ExerciseAngina	maximum heart rate achieved [Numeric value between 60 and 202]	<i>predictor (independent) variables</i>
10.	Oldpeak	oldpeak = ST [Numeric value measured in depression]	<i>predictor (independent) variables</i>
11.	ST_Slope	the slope of the peak exercise ST segment [Up: upsloping, Flat: flat, Down: downsloping]	<i>predictor (independent) variables</i>
12.	HartDisease	output class [1: heart disease, 0: Normal]	<i>target (dependent) variable</i>

-- oOo --