

**TUGAS AKHIR SEMESTER MATA KULIAH  
PENAMBANGAN DATA - KELOMPOK: A11.4701**

**Tujuan:**

---

Mahasiswa mampu mengimplementasikan satu atau lebih teknik penambangan untuk menangani tugas-tugas komputasi seperti estimasi, klasifikasi, klasterisasi, peramalan, atau asosiasi yang bermanfaat untuk pengambilan keputusan.

**Deskripsi:**

---

- Tugas dilaksanakan secara kelompok dan setiap kelompok terdiri maksimal 5 mahasiswa.
- Masing-masing mahasiswa dapat membentuk kelompok sendiri.
- Setiap kelompok wajib memilih 1 (satu) topik proyek dari topik-topik penambangan berikut: estimasi, peramalan, klasterisasi, klasifikasi atau asosiasi, termasuk teknik/metode/algoritma yang akan digunakan.
- Dataset/obyek penelitian dapat dipilih sendiri secara bebas dan disarankan menggunakan dataset publik.
- Implementasi penambangan data menggunakan bahasa pemrograman Python.
- Tiap kelompok wajib mempresentasikan pekerjaannya kepada Dosen Pengampu untuk menggantikan Ujian Akhir.

**Apa yang harus dilakukan:**

---

- Lakukan pembentukan kelompok, maksimal 5 orang dan lakukan pembagian tugas masing-masing.
- Lakukan studi literatur dari sumber-sumber referensi untuk mencari ide/inspirasi.
- Tentukan topik proyek, cari dataset publik, tentukan 1 atau lebih metode/algoritma yang akan digunakan (luaran: **dataset**).
- Implementasikan dengan coding dengan Python (luaran: **Python code**), gunakan IDE Google Colab dan simpan coding Anda dalam ekstensi .ipynb)
- Dokumentasikan hasil pekerjaan Anda di Laporan sesuai dengan lampiran contoh template laporan yang disediakan (luaran: **laporan**). Sistematika laporan sesuai dengan template, namun isi untuk butir-butir sistematika dapat divariasi sendiri.
- Buat slide presentasi (power point) untuk dipresentasikan pada saat ujian (luaran: **slide presentasi**).
- Buat folder di google drive dan upload ke-4 file tersebut (**dataset, Python code, laporan, slide presentasi**) dalam satu folder.
- Share link folder Anda (without restrictions) melalui WA ke Dosen pengampu.

**Kapan deadline tugas:**

---

- Deadline upload : 19 Januari 2023.
- Deadline presentasi : 26 Januari 2023 (jadwal menyusul). Kelompok yang sudah menyelesaikan tugas sebelum tanggal tersebut dapat mengajukan jadwal presentasi lebih awal ke Dosen. Pada saat jadwal UAS, masing-masing mahasiswa tinggal melakukan absensi di Ruang Dosen.

## JUDUL PROYEK

# KLASIFIKASI PASIEN BERPOTENSI STROKE MENGGUNAKAN METODE LINEAR REGRESSION CLASSIFIER (HANYA CONTOH)



Proyek ini disusun untuk memenuhi tugas akhir mata kuliah Penambangan Data pada Program Sarjana Teknik Informatika Fakultas Ilmu Komputer Universitas Dian Nuswantoro Semester Ganjil Tahun Akademik 2023/2023

Dosen Pengampu  
Dr. Aris Marjuni, S. Si, M. Kom

TIM PROYEK:

No.	NIM	NAMA
1.		
2.		
3.		
4.		
5.		

## ABSTRAKSI PROYEK (HANYA CONTOH)

Menurut Organisasi Kesehatan Dunia (WHO), stroke adalah penyebab kematian nomor dua secara global dan memberikan kontribusi sekitar 11% dari total kematian. Proyek ini dilakukan dengan tujuan untuk mengklasifikasi status seseorang terhadap kemungkinan apakah berpotensi mengalami stroke atau tidak berdasarkan data dari orang-orang yang pernah mengalami stroke sebelumnya. Klasifikasi dilakukan menggunakan pendekatan penambangan data dengan model Linear Regression Model. Dataset yang digunakan pada klasifikasi ini adalah dataset publik yang bersumber dari Kaggle.com. Hasil klasifikasi menunjukkan bahwa Linear Regression Model mampu mengidentifikasi potensi seseorang terkena stroke sebesar 95%.

---

### A. DATASET (HANYA CONTOH)

Nama dataset:

healthcare-dataset-stroke-data.csv

Sumber:

<https://www.kaggle.com/datasets/fedesoriano/stroke-prediction-dataset/data>

Informasi dataset:

5110 observations / records

12 attributes

Informasi Atribut:

- 1) id: unique identifier
- 2) gender: "Male", "Female" or "Other"
- 3) age: age of the patient
- 4) hypertension: 0 if the patient doesn't have hypertension, 1 if the patient has hypertension
- 5) heart\_disease: 0 if the patient doesn't have any heart diseases, 1 if the patient has a heart disease
- 6) ever\_married: "No" or "Yes"
- 7) work\_type: "children", "Govt\_jov", "Never\_worked", "Private" or "Self-employed"
- 8) Residence\_type: "Rural" or "Urban"
- 9) avg\_glucose\_level: average glucose level in blood
- 10) bmi: body mass index
- 11) smoking\_status: "formerly smoked", "never smoked", "smokes" or "Unknown"\*
- 12) stroke: 1 if the patient had a stroke or 0 if not

Contoh instances: Tampilkan 5 data instances pertama

	id	gender	age	hypertension	heart_disease	ever_married	work_type	Residence_type	avg_glucose_level	bmi	smoking_status	stroke
0	9046	Male	67.0	0	1	Yes	Private	Urban	228.69	36.6	formerly smoked	1
1	51676	Female	61.0	0	0	Yes	Self-employed	Rural	202.21	NaN	never smoked	1
2	31112	Male	80.0	0	1	Yes	Private	Rural	105.92	32.5	never smoked	1
3	60182	Female	49.0	0	0	Yes	Private	Urban	171.23	34.4	smokes	1
4	1665	Female	79.0	1	0	Yes	Self-employed	Rural	174.12	24.0	never smoked	1

---

### B. DATA PREPROCESSING

Data preprocessing apa saja yang telah Anda lakukan terhadap dataset yang Anda gunakan (jika ada). Misalnya: pembersihan data, transformasi data, reduksi data, penanganan missing value, diskritisasi data/label encoding, normalisasi, dan sebagainya. Tampilkan capture di bagian ini data sebelum dan sesudah data preprocessing untuk menunjukkan perubahan yang terjadi atas data preprocessing tersebut. Jika Anda tidak melakukan data preprocessing, bagian ini dapat dikosongi.

---

## C. EXPLORATORY DATA ANALYSIS (EDA)

Lakukan EDA terhadap data-data yang sudah Anda miliki dan capture hasilnya di bagian ini, meliputi:

- Deskripsi/informasi dataset : gunakan referensi `info()`
- Capture 10 data secara random : gunakan referensi `sample()`
- Periksa apakah ada data yang kosong (missing) : gunakan referensi `isnull()`
- Tampilkan informasi statistik dari dataset tersebut : gunakan referensi `describe()`
- Plot correlogram (heatmap) : gunakan referensi `heatmap()`

Anda juga dapat berkreasi untuk menambahkan EDA dan visualisasi data yang lain.

---

## D. PEMODELAN

Bagian ini menjelaskan implementasi model yang Anda gunakan, seperti:

- Pendefinisian data variabel bebas dan tidak bebas
- Pendefinisian data training dan testing
- Hasil eksperimen dan pengukuran/evaluasi model

Anda dapat berkreasi untuk menampilkan hasil implementasi pemodelan baik secara tekstual, tabular maupun visual.

---

## E. KESIMPULAN

Tuliskan secara singkat kesimpulan Anda berdasarkan hasil eksperimen yang sudah Anda lakukan.

---

## F. KONTRIBUSI ANGGOTA

Tuliskan persentase kontribusi masing-masing anggota dalam pekerjaan tugas ini. Kontribusi diukur dari peran aktif anggota, khususnya dalam kerja sama tim untuk menyelesaikan proyek ini.

No.	NIM	NAMA	KONTRIBUSI DAN KEAKTIFAN (%)
1.			
2.			
3.			
4.			
5.			