



### UJIAN AKHIR SEMESTER GANJIL 2024/2025

Nama Mata Kuliah : Bengkel Koding Sifat : Take Home  
Hari / Tanggal : Jumat, 3 Januari 2024 Waktu : -  
Pukul : - Dosen : Tim Data Science

Masa Pengerjaan : **Tanggal 2 Desember s.d 3 Januari 2025 23.59 WIB**

### KETENTUAN CAPSTONE PROJECT BENGKEL KODING DATA SCIENCE

Bagi mahasiswa Bengkel Koding Data Science diwajibkan untuk mengerjakan capstone berikut untuk tugas akhir dari mata kuliah Bengkel Koding ini. Ketentuan project sebagai berikut :

#### **Dataset**

Dataset diberikan oleh asisten Data Science Bengkel Koding sebagai berikut:

Sumber: Klasifikasi Data Kualitas Air dan Kelayakan Minum

<https://drive.google.com/dataset-water-potability>

#### **Tahapan Pengerjaan Capstone**

##### **1. Pengumpulan Data**

Berdasarkan dataset kualitas air yang telah disediakan untuk melakukan klasifikasi, gunakanlah kode di bawah ini untuk membaca file dataset.

```
file_path = '/content/water_potability.csv'
water_data = pd.read_csv(file_path)

water_data.head()
```

Output:

	ph	Hardness	Solids	Chloramines	Sulfate	Conductivity	Organic_carbon	Trihalomethanes	Turbidity	Potability
0	NaN	204.890455	20791.318981	7.300212	368.516441	564.308654	10.379783	86.990970	2.963135	0
1	3.716080	129.422921	18630.057858	6.635246	NaN	592.885359	15.180013	56.329076	4.500656	0
2	8.099124	224.236259	19909.541732	9.275884	NaN	418.606213	16.868637	66.420093	3.055934	0
3	8.316766	214.373394	22018.417441	8.059332	356.886136	363.266516	18.436524	100.341674	4.628771	0
4	9.092223	181.101509	17978.986339	6.546600	310.135738	398.410813	11.558279	31.997993	4.075075	0

##### **2. Menelaah Data**

Lakukan analisis awal terhadap 10 kolom dari dataset, yang terdiri dari: Ph, Hardness, Solids, Chloramines, Sulfate, Conductivity, Organic\_carbon, Trihalomethanes, Turbidity, dan

Potability. Tampilkan informasi dari dataset, seperti jumlah baris, tipe data tiap kolom, dan nilai unik.

### 3. Validasi dan Visualisasi Data

Tahap ini bertujuan untuk mengetahui dan memahami isi dari dataset agar dapat dilakukan penanganan yang sesuai dengan kondisinya, yang meliputi:

- Cek missing values dan outlier di setiap kolom. Gunakan teknik imputasi rata-rata (mean) untuk mengganti nilai dari kondisi tersebut.
- Visualisasi distribusi data kualitas air sebelum dan setelah melakukan resampling melalui bar plot.

### 4. Menentukan Objek Data

Dalam dataset kualitas air, fitur yang digunakan terdapat 9 fitur, yaitu:

- Ph
- Conductivity
- Hardness
- Organic\_carbon
- Solids
- Trihalomethanes
- Chloramines
- Turbidity
- Sulfate

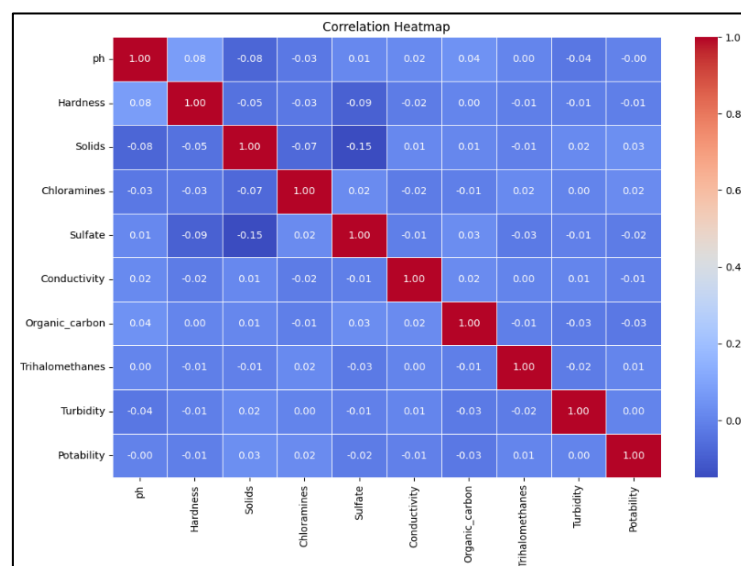
Target yang ingin diprediksi adalah Potability.

### 5. Membersihkan Data

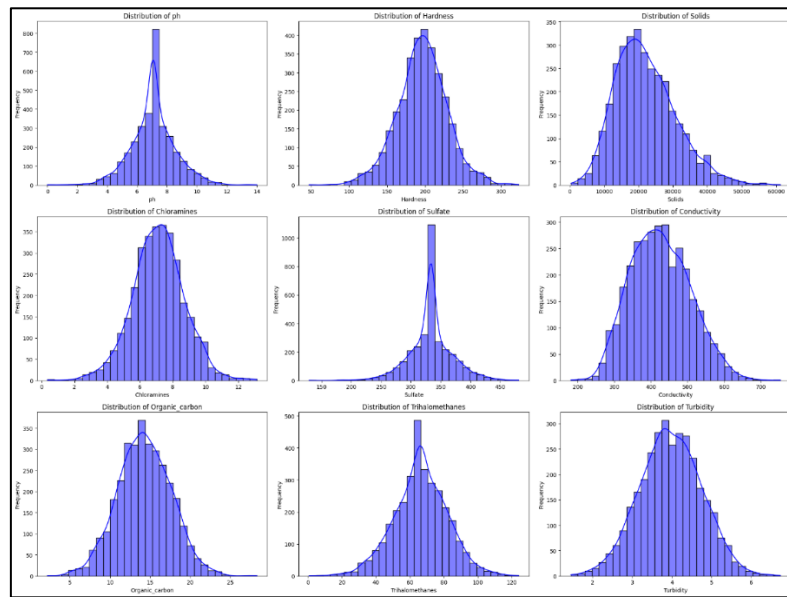
Sebelum melakukan pemodelan dilakukan pembersihan data agar model yang dihasilkan lebih akurat dengan menampilkan visualisasi korelasi heatmap dan distribusi atribut menggunakan histogram plot.

Output:

- Korelasi Heatmap



- Distribusi Histogram Plot



## 6. Konstruksi Data

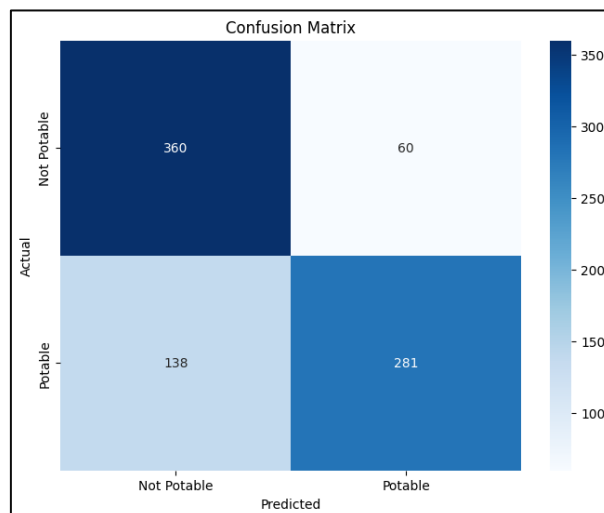
Salah satu tujuan dari tahapan konstruksi data yaitu untuk menyesuaikan semua tipe data yang terdapat dalam dataset. Namun, pada tahap ini dataset sudah memiliki tipe data yang sesuai sehingga tidak perlu dilakukan penyesuaian kembali.

## 7. Pemodelan

Project **wajib** menggunakan **minimal 3 algoritma klasifikasi** yang kemudian dikomparasi hasil akurasi satu sama lain.

Output:

Berikan confusion matrix untuk setiap model contoh:



## 8. Evaluasi

Setelah melakukan pemodelan harus dilakukan evaluasi model yang telah dibuat dengan menampilkan hasil akurasi sebelum dinormalisasikan dan setelah dinormalisasikan.

## 9. Deployment

- Deployment dapat menggunakan **satu model terbaik** atau **semua model sekaligus**.
- Deployment dilakukan menggunakan Streamlit, ketentuan template deployment berdasarkan kreativitas mahasiswa masing-masing.

### Tambahan:

Akan ada nilai tambahan jika project telah deploy secara **online** (contoh: Streamlit Cloud). Tutorial deployment ke streamlit cloud:

[How to Deploy Your App to Streamlit Community Cloud](#)

## 10. Kesimpulan

Buatlah kesimpulan dari hasil analisis perbandingan algoritma yang digunakan, meliputi:

- Tingkat akurasi setiap algoritma,
- Keunggulan dan keterbatasannya,
- Rekomendasi algoritma yang paling efektif untuk kasus ini beserta alasan mengapa algoritma tersebut optimal.

## **Presentasi Capstone Project**

Jenis Presentasi dapat dipilih sesuai dengan kesepakatan kelas

**Pilihan 1:** Mahasiswa dapat datang ke Bengkel Koding dan bertemu secara langsung dengan Mentor untuk melakukan presentasi (**Offline**)

**Pilihan 2:** Mahasiswa dapat melakukan presentasi secara **online** dengan syarat sebagai berikut:

- (1) Mahasiswa **Wajib mengupload video secara public di Youtube.**
- (2) Mahasiswa **Wajib mengupload code di github.**
- (3) Durasi presentasi minimal 10 menit, dengan ketentuan:
  - a. Perkenalan diri (nama dan NIM)
  - b. Penjelasan setiap step dari code hasil program
  - c. Menjelaskan Hasil Evaluasi

## **Ketentuan Pengumpulan Capstone Project**

Pengumpulan project ini dikumpulkan di website Bengkel Koding menggunakan format file pdf, berupa:

- Link Github source code project
- Link Streamlit yang sudah di deploy (jika ada)
- Link Video Presentasi
- Bahan untuk presentasi (sesuai kesepakatan masing-masing kelas)

**Capstone Project ini BERSIFAT WAJIB,**  
**Diharapkan mahasiswa mengumpulkan tepat waktu project capstone ini, bila ada keterlambatan dimohon untuk konfirmasi terlebih dahulu ke asisten Bengkel Koding masing-masing sebelum waktu tenggat pengumpulan. Jika tidak atau melanggar ketentuan ada sanksi pengurangan nilai.**

Sekian, terima kasih.

Mengetahui,

**Asisten**

**Bengkel Koding Data Science**

**Dosen Pengampu**

**Bengkel Koding Data Science**