



Capstone  
Project

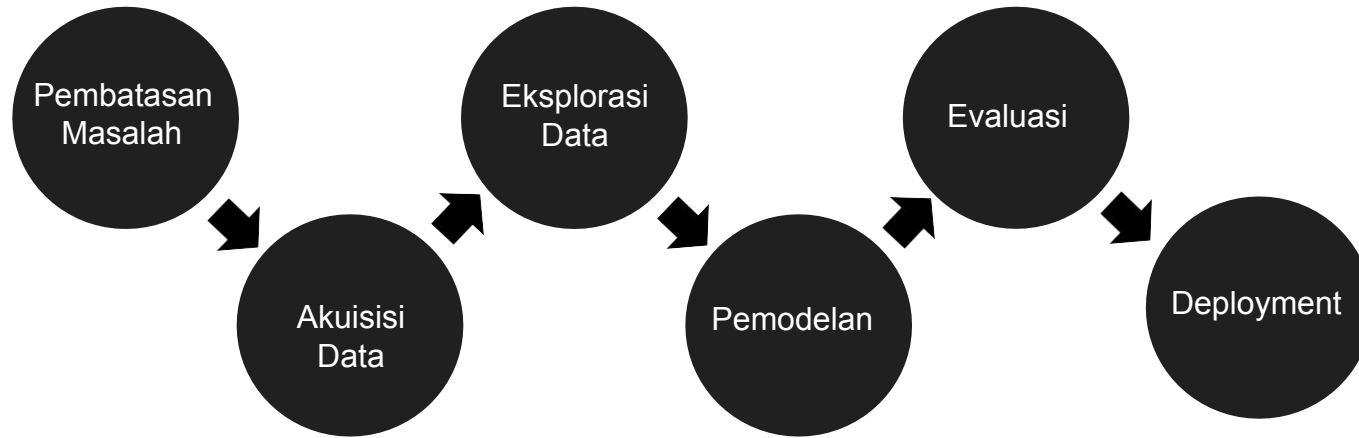
# PENDETEKSI AIR BERSIH

Elsa Grasia / 71210761

Christmasnanato I. Sayrang / 71210795

Yohanes Prasetyo Nugroho / 71210811

# AI Project Cycle



Reff: <https://medium.com/@hannnfh/ai-project-cycle-ccd67c3dd21d>

# Pembatasan Masalah

- Masalah apa yang ingin Anda cari penyelesaiannya?

Identifikasi pencemaran dalam air.

- Solusi yang ingin Anda hasilkan?

Identifikasi pencemar dalam kandungan air dengan menggunakan program pendeteksi air bersih.

# Metode 4W Canvas

|  |   |       |
|--|---|-------|
| Para<br>[stakeholders]                             | Masyarakat yang mempunyai permasalahan dalam kualitas air.  | WHO   |
| Masalah yang dihadapi<br>[isu, masalah, kebutuhan] | Analisis kualitas air untuk berbagai parameter, seperti pH, kekeruhan, kandungan logam.   | WHAT  |
| Ketika<br>[konteks dan situasi]                    | Menyediakan data kualitas air yang valid dan akurat untuk digunakan.  | WHERE |
| Solusi yang diharapkan                             | Mendeteksi dalam membantu dan menjaga kualitas air bersih, mencegah dampak negatif pada kesehatan masyarakat dan meningkatkan transparansi dan kepercayaan terhadap pasokan air yang digunakan oleh masyarakat. | WHY   |

# Akuisisi Data

- Proses mengumpulkan data-data yang dibutuhkan untuk membuat proyek AI. Hal ini merupakan dasar atau bahan yang selanjutnya diolah untuk dianalisis sesuai masalah dan diamati agar bisa menghasilkan solusi terbaik. Sering disebut dengan data collecting.
- Data yang dikumpulkan bisa berupa data yang dianalisa ataupun data berupa literatur yang mendasari solusi bisa memungkinkan untuk digunakan.
- Ada beberapa cara untuk mendapatkan data:
- Data primer:
  - Melalui Tools/Alat (IoT) : Kamera, Microphone dan Sensor
  - Observasi (formulir pengisian) : Survei, Penelitian
- Data sekunder:
  - Open Data : BPS, AWS, Kaggle, API (REST API, Twitter API, Youtube API)
  - Web Scraping/Crawling.

# Akuisisi Data

- Apakah Anda memiliki data untuk dianalisis? Ya
- Apakah Anda memiliki data primer atau sekunder atau keduanya? Sekunder
- Apakah Anda memiliki literatur yang mendukung solusi Anda? Ya
- Apakah Anda coba melengkapi data sekunder anda dari sumber data lain? Ya
- Apa yang harus disiapkan:
  - Jika ada data primer, buatlah dalam format csv atau format xlsx (excel)!
  - Jika terdapat sumber data sekunder, downloadlah dan satukan dengan data primer Anda!
  - Jika anda memiliki literatur, bacalah dan pilihlah parameter (fitur) dan class (target) yang Anda inginkan!

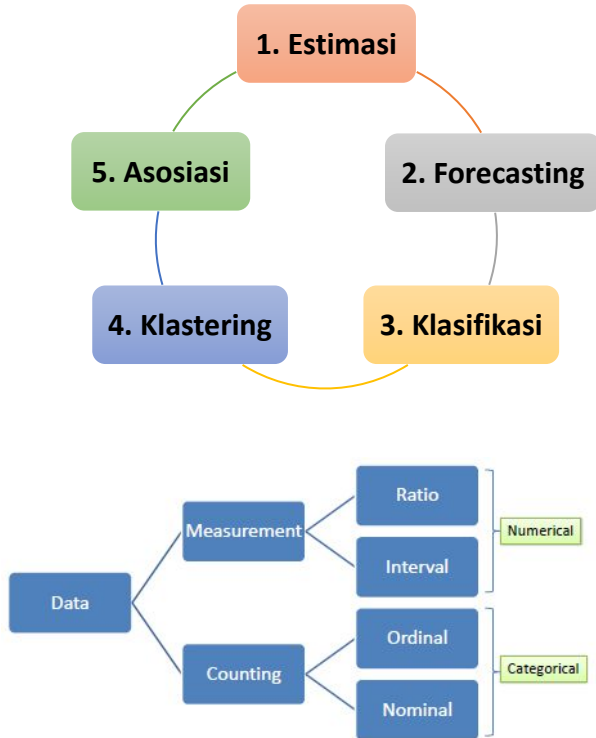
\*diisi sesuai yang anda butuhkan

# Eksplorasi

- Setelah melengkapi dan menyiapkan data, lakukan langkah selanjutnya:
- Eksplorasikan data dengan melihat jenis data yang anda miliki dan tampilkan dalam tabel berikut ini.

| No | Nama parameter                    | status (input/output) | Tipe data (nominal/numerik) |
|----|-----------------------------------|-----------------------|-----------------------------|
| 1  | PH air                            | Input                 | Nominal                     |
| 2  | Suhu air                          | Input                 | Nominal                     |
| 3  | Kekeruhan air                     | Input                 | Nominal                     |
| 4  | Kandungan logam                   | Input                 | Nominal                     |
| 5  | Air dapat digunakan<br>atau tidak | Output                | Numerik                     |
| 6  | Data Kualitas Air                 | Output                | Numerik                     |
|    |                                   |                       |                             |
|    |                                   |                       |                             |
|    |                                   |                       |                             |
|    |                                   |                       |                             |

# Metode



Attribute/Feature/Dimension

Class/Label/Target

**Nominal**

|     | Sepal Length (cm) | Sepal Width (cm) | Petal Length (cm) | Petal Width (cm) | Type        |
|-----|-------------------|------------------|-------------------|------------------|-------------|
| 1   | 5.1               | 3.5              | 1.4               | 0.2              | Iris setosa |
| 2   | 4.9               | 3.0              | 1.4               | 0.2              | Iris setosa |
| 3   | 4.7               | 3.2              | 1.3               | 0.2              | Iris setosa |
| 4   | 4.6               | 3.1              | 1.5               | 0.2              | Iris setosa |
| 5   | 5.0               | 3.6              | 1.4               | 0.2              | Iris setosa |
| ... |                   |                  |                   |                  |             |

Record/  
Object/  
Sample/  
Tuple/  
Data

**Numerik**



# Metode (2)

1. **Estimation** (Estimasi):  
Linear Regression (LR), Neural Network (NN), Deep Learning (DL), Support Vector Machine (SVM), Generalized Linear Model (GLM), etc
2. **Forecasting** (Prediksi/Peramalan):  
Linear Regression (LR), Neural Network (NN), Deep Learning (DL), Support Vector Machine (SVM), Generalized Linear Model (GLM), etc
3. **Classification** (Klasifikasi):  
Decision Tree (CART, ID3, C4.5, Credal DT, Credal C4.5, Adaptative Credal C4.5), Naive Bayes (NB), K-Nearest Neighbor (kNN), Linear Discriminant Analysis (LDA), Logistic Regression (LogR), etc
4. **Clustering** (Klastering):  
K-Means, K-Medoids, Self-Organizing Map (SOM), Fuzzy C-Means (FCM), etc
5. **Association** (Asosiasi):  
FP-Growth, A Priori, Coefficient of Correlation, Chi Square, etc

# Eksplorasi Data

- Setelah mengisi tipe data, selanjutnya:
- Silahkan menggunakan visualisasi yang anda inginkan sesuai harapan yang Anda harapkan!

| 1  |         | Lokasi                  | pH Air        | suhu Air(derajat celcius) | kekeruhan Air(NTU) | kandungan Air logam(ppm) | Keterangan                |
|----|---------|-------------------------|---------------|---------------------------|--------------------|--------------------------|---------------------------|
| 2  | Data 1  | Sungai Konteng          | 7.06 (Normal) | 29 (Normal)               | 33 (Tercemar)      | 0.59 (Tercemar)          | Air tidak dapat digunakan |
| 3  | Data 2  | Sungai Bengawan Solo    | 6.80 (Normal) | 30.5 (Normal)             | 26 (Tercemar)      | 0.20 (Normal)            | Air tidak dapat digunakan |
| 4  | Data 3  | Sungai Bedadung Jember  | 7,74 (Normal) | 28,22 (Normal)            | 12,55 (Tercemar)   | 1,13 (Tercemar)          | Air dapat digunakan       |
| 5  | Data 4  | Sungai Jawi Pontianak   | 7.2 (Normal)  | 28.8 (Normal)             | 1.77 (Normal)      | 1.59 (Tercemar)          | Air tidak dapat digunakan |
| 6  | Data 5  | Sungai Ogan Palembang   | 5.82 (Normal) | 28 (Normal)               | 38.4 (Tercemar)    | 0.3 (Normal)             | Air tidak dapat digunakan |
| 7  | Data 6  | Sungai Kapuas Pontianak | 6.32 (Normal) | 28 (Normal)               | 21 (Tercemar)      | 0.6 (Tercemar)           | Air tidak dapat digunakan |
| 8  | Data 7  | Sungai Lesti Malang     | 8.34 (normal) | 29 (Normal)               | 19.9 (Tercemar)    | 8.60 (Tercemar)          | Air dapat digunakan       |
| 9  | Data 8  | Sungai Gambut Kalsel    | 7.33 (Normal) | 30.8 (Normal)             | 54.8 (Tercemar)    | 1.71 (Tercemar)          | Air tidak dapat digunakan |
| 10 | Data 9  | Sungai BatangHari       | 5.52 (Normal) | 29 (Normal)               | 22.6 (Tercemar)    | 0.59 (Tercemar)          | Air tidak dapat digunakan |
| 11 | Data 10 | Sungai Percut           | 7.2 (Normal)  | 22 (Tercemar)             | 16.9 (Tercemar)    | 3.5 (Tercemar)           | Air tidak dapat digunakan |

- Setelah melihat fitur dan class maka pilihan Metode yang Anda gunakan adalah:
  - Klasifikasi

# Modelling

- Setelah metode sudah dipilih maka:
- Algoritma apa yang ingin Anda gunakan untuk menghasilkan model (pengetahuan):  
Regression / Decision Tree / Random Forest / ANN / K-means / BFS / DFS /dll
- Model Evaluasi apa yang dipilih:
  - Naive Bayes
    - Dalam proyek deteksi air bersih, Naive Bayes digunakan untuk mengklasifikasikan sampel air apakah bersih atau tercemar. Naive Bayes dapat memberikan hasil klasifikasi dengan cepat berdasarkan probabilitas atribut yang diamati dalam air yang diuji. Naive Bayes juga bekerja dengan baik dalam kasus jumlah atribut yang besar.

# Hasil Program

- Tampilkan hasil program/model yang anda kerjakan
- Data 1: Sungai Konteng

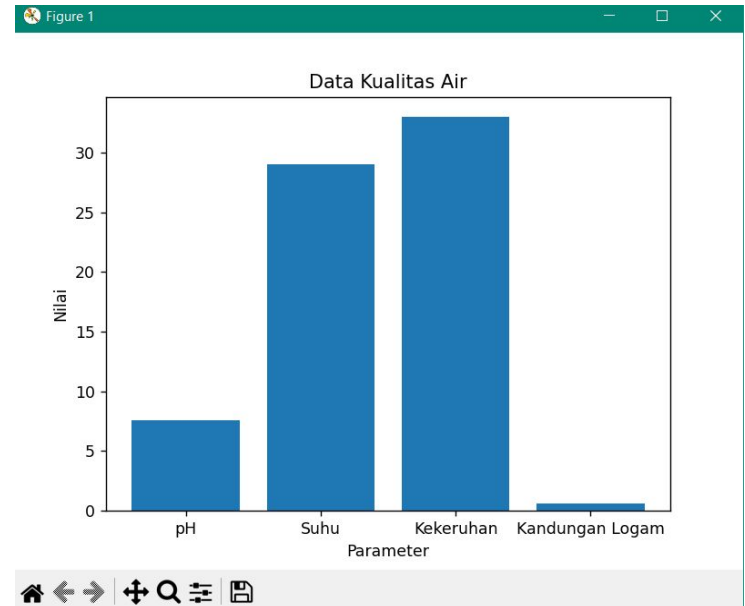
```
Program Pendeteksi Air Bersih
=====

Masukkan nilai pH air: 7.6
Masukkan suhu air (derajat Celsius): 29
Masukkan kekeruhan air (NTU): 33
Masukkan kandungan logam air (ppm): 0.59

=== Data Air ===
pH: 7.60 (Normal)
Suhu: 29.00 derajat Celsius (Normal)
Kekeruhan: 33.00 NTU (Tercemar)
Kandungan Logam: 0.59 ppm (Tercemar)

Air tidak dapat digunakan.
```

Diagram Data 1



# Hasil Program

- Tampilkan hasil program/model yang anda kerjakan  
Data 2: Sungai Bengawan Solo

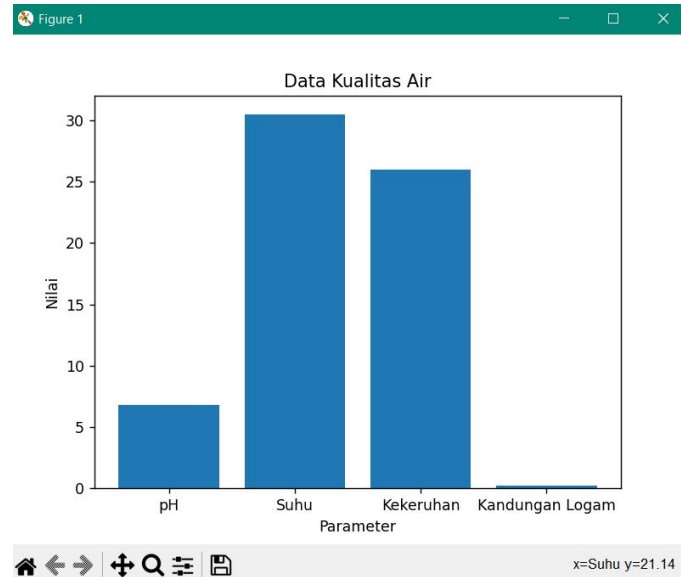
```
Program Pendeteksi Air Bersih
=====

Masukkan nilai pH air: 6.8
Masukkan suhu air (derajat Celsius): 30.5
Masukkan kekeruhan air (NTU): 26
Masukkan kandungan logam air (ppm): 0.2

=== Data Air ===
pH: 6.80 (Normal)
Suhu: 30.50 derajat Celsius (Normal)
Kekeruhan: 26.00 NTU (Tercemar)
Kandungan Logam: 0.20 ppm (Normal)

Air tidak dapat digunakan.
```

Diagram Data 2



# Kesimpulan

- Berikan kesimpulan dari kasus yang anda pilih
- Dari program pendeteksi yang kami buat program tersebut bisa berjalan dengan yang kami inginkan.
- Program ini dapat mendeteksi dan memantau kualitas air dengan efektif. Proyek ini menjaga kebersihan dan keamanan air, serta melindungi kesehatan masyarakat. Dengan adanya pendeteksi air bersih ini, masyarakat dapat dengan mudah mengidentifikasi adanya pencemaran dalam air dan mengambil langkah-langkah yang diperlukan untuk mengatasi masalah tersebut.