

# PREDIKSI PENDAPATAN PENDUDUK MENGGUNAKAN NAIVE BAYES

**Fahmi Noor Fiqri**

By [Kodesiana.com](https://kodesiana.com)



# Dataset Adult

**Dataset adult** merupakan dataset yang berisi informasi sensus penduduk dan pendapatannya.

Tujuan dari dataset ini adalah membuat model yang dapat melakukan klasifikasi apakah seorang penduduk memiliki pendapatan lebih besar atau sama dengan \$50.000 USD per tahun atau tidak.



Sumber:  
<https://archive.ics.uci.edu/ml/datasets/Adult>

# Naive Bayes

**Naive bayes** merupakan metode klasifikasi probabilistik yang berdasarkan pada teorema *Bayes*.



You Know I'm All About that Bayes:  
Crash Course Statistics #24

## Naïve Bayes Classifier

$$P(A|B) = \frac{P(B|A) P(A)}{P(B)}$$



Thomas Bayes  
1702 - 1761

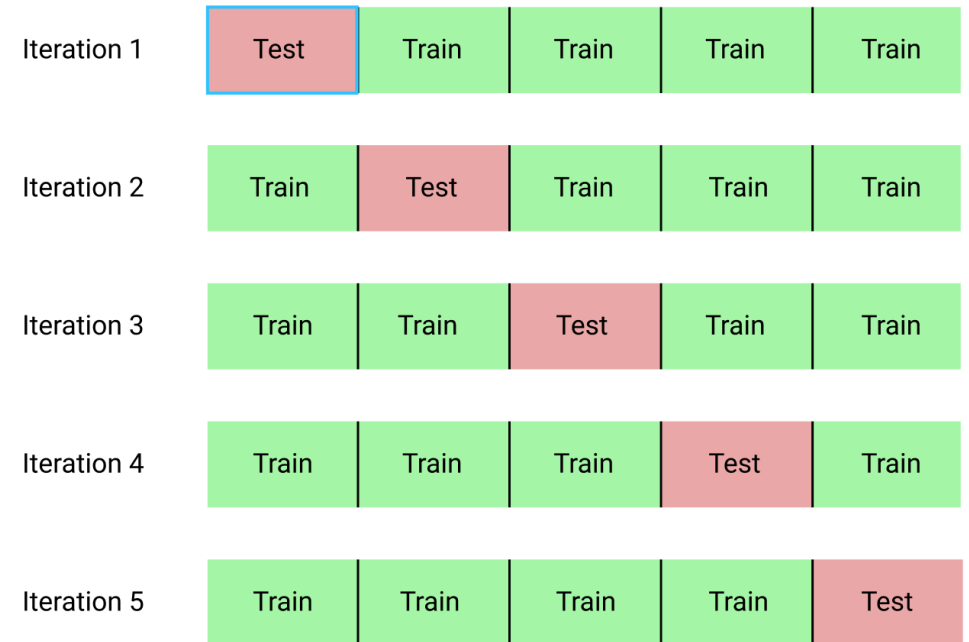
Sumber:

<https://medium.com/datadriveninvestor/a-gentle-introduction-to-naive-bayes-classifier-9d7c4256c999>

# Cross Validation

**Cross validation** merupakan metode evaluasi model dengan cara membagi data menjadi beberapa bagian untuk di latih dan diuji, disebut juga *K-fold validation*.

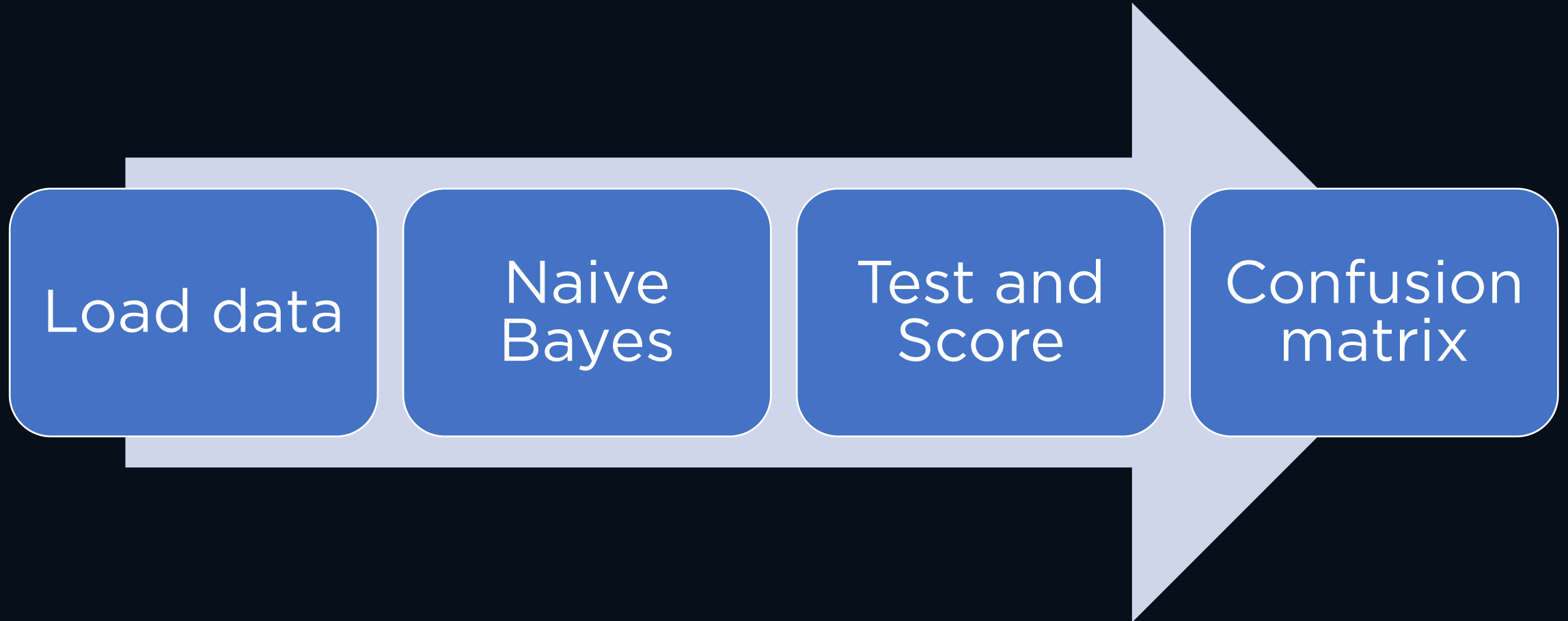
Nilai k menunjukkan berapa banyak iterasi dan berapa banyak data akan dibagi menjadi bagian-bagian kecil untuk mengevaluasi model.



Sumber:

<https://stackoverflow.com/a/58822062/5561144>

# Proses Data Mining



# PREDIKSI PENDAPATAN PENDUDUK MENGGUNAKAN NAIVE BAYES

**Fahmi Noor Fiqri**  
By [Kodesiana.com](https://kodesiana.com)

