

Nama = Aisyah

Nim = 4122003

---

## Laporan Praktikum Machine Learning Naive Bayes

### 1. Dataset

Dataset yang digunakan berjudul "**Indonesian Marketplace Product Reviews**", diperoleh dari Kaggle melalui tautan berikut:

[Indonesian Marketplace Product Reviews](#)

Dataset terdiri atas:

- reviews: teks ulasan pelanggan
- label: nilai 0 untuk ulasan negatif, dan 1 untuk ulasan positif

Jumlah total data adalah 832 record, yang kemudian dibagi menjadi:

- 70% data latih (training) → 582 data
- 30% data uji (testing) → 250 data

### 2. Pra-pemrosesan dan Ekstraksi Fitur

Langkah-langkah preprocessing meliputi:

- Menghapus karakter tidak relevan dan tanda baca

```
[6]: # 5. Bersihkan teks (lowercase, hapus simbol, dll)
df['reviews'] = df['reviews'].str.lower()
df['reviews'] = df['reviews'].str.replace(r'^a-zA-Z\s', '', regex=True)
```

- Mengubah teks menjadi lowercase (huruf kecil semua) Untuk menyamakan format teks, agar kata yang sama tapi beda penulisan tidak dianggap berbeda.
- Mengubah teks ke bentuk vektor numerik menggunakan TF-IDF (Term Frequency-Inverse Document Frequency) Transformasi ini dilakukan menggunakan TfidfVectorizer dari scikit-learn.

```
[2]: # 1. Import library yang diperlukan
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.naive_bayes import MultinomialNB
from sklearn.metrics import classification_report, accuracy_score
```

```
[8]: # 7. TF-IDF vektorisasi teks
vectorizer = TfidfVectorizer()
X_tfidf = vectorizer.fit_transform(X)
```

### 3. Evaluasi Model

Berikut hasil evaluasi model pada data uji:

```
[11]: # 10. Prediksi dan evaluasi
y_pred = model.predict(X_test)

print("\n=== Evaluasi Model ===")
print("Akurasi:", accuracy_score(y_test, y_pred))
print("\nLaporan Klasifikasi:\n", classification_report(y_test, y_pred))

=== Evaluasi Model ===
Akurasi: 0.9161676646706587

Laporan Klasifikasi:
              precision    recall  f1-score   support

     0       0.86      1.00      0.92        84
     1       1.00      0.83      0.91        83

 accuracy          0.92          167
 macro avg         0.93          167
 weighted avg      0.93          167
```

- Akurasi: 91.6%

Kelas	Precision	Recall	F1-Score	Support
Negatif (0)	0.86	1.00	0.92	84
Positif (1)	1.00	0.83	0.91	83

- Macro Avg F1-Score: 0.92
- Weighted Avg F1-Score: 0.92

Model menunjukkan performa yang sangat baik dengan presisi dan recall yang tinggi untuk kedua kelas.