

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:

- 80% data latih (training) → 582 data
- 20% data uji (testing) → 250 data

2. Pra-pemrosesan dan Ekstraksi Fitur

Langkah-langkah preprocessing meliputi:

- Mengubah ke lowercase: semua huruf diubah menjadi kecil agar konsisten.

```
[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)
```

- Menghapus karakter tidak relevan: seperti angka, tanda baca, dan karakter spesial dihapus menggunakan regex.
- TF-IDF Vectorization: data teks diubah ke bentuk numerik menggunakan TfidfVectorizer dari library scikit-learn.

```
[1]: # 1. Import library yang diperlukan
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
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, confusion_matrix
```

```
[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       0.92       0.92         167
 weighted avg              0.93       0.92       0.92         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

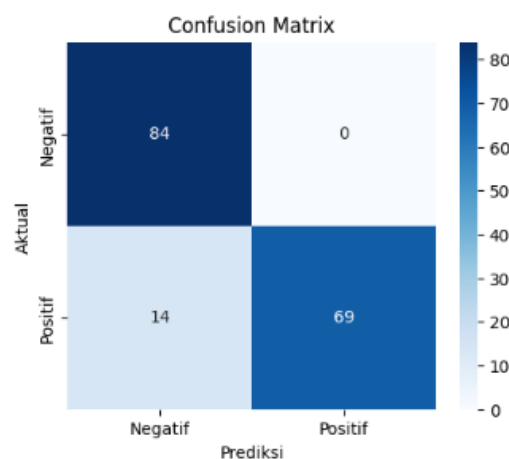
- Macro Avg F1-Score: 0.92
- Weighted Avg F1-Score: 0.92

Confusion Matrix

```
[11]: # Tambahan: Confusion Matrix (angka)
cm = confusion_matrix(y_test, y_pred)
print("Confusion Matrix:\n", cm)

# Tambahan: Visualisasi Confusion Matrix
plt.figure(figsize=(5, 4))
sns.heatmap(cm, annot=True, fmt='d', cmap='Blues', xticklabels=['Negatif', 'Positif'], yticklabels=['Negatif', 'Positif'])
plt.xlabel('Prediksi')
plt.ylabel('Aktual')
plt.title('Confusion Matrix')
plt.show()

Confusion Matrix:
[[84  0]
 [14 69]]
```



4. Kesimpulan Akhir

Berdasarkan hasil klasifikasi ulasan produk menggunakan model Multinomial Naive Bayes, diperoleh beberapa poin penting:

1. Akurasi Model sebesar 91.6%, yang menunjukkan bahwa model mampu mengklasifikasikan ulasan sebagai positif atau negatif dengan sangat baik.
2. Confusion Matrix menunjukkan:
 - 84 data ulasan negatif berhasil diprediksi dengan benar.
 - 69 data ulasan positif juga berhasil diprediksi dengan benar.
 - Terdapat 14 kesalahan prediksi di mana ulasan positif diklasifikasikan sebagai negatif.
 - Tidak terdapat kesalahan prediksi pada ulasan negatif (False Positive = 0).
3. Nilai Precision, Recall, dan F1-Score:
 - Untuk kelas negatif: Precision = 0.86, Recall = 1.00
 - Untuk kelas positif: Precision = 1.00, Recall = 0.83
 - F1-score rata-rata = 0.92
4. Hal ini mengindikasikan bahwa model sangat bagus dalam mengenali ulasan negatif, dan cukup baik dalam mengenali ulasan positif, meskipun masih ada beberapa kesalahan klasifikasi.