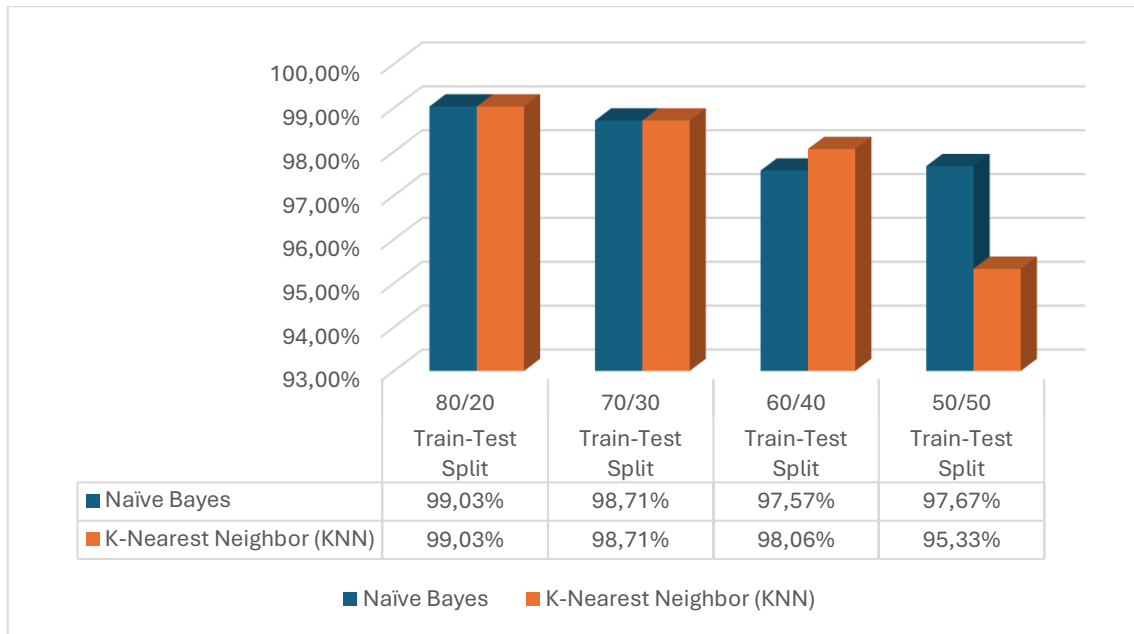


BAB V PEMBAHASAN

Berdasarkan hasil evaluasi yang telah dilakukan untuk model klasifikasi dengan berbagai rasio train-test, sebagai berikut:



Gambar 1 Hasil Akurasi

Gambar diatas menunjukkan perbandingan hasil akurasi antara dua algoritma, Naive Bayes dan K-Nearest Neighbor (KNN), dengan empat pembagian data train-test split yang berbeda: 80/20, 70/30, 60/40, dan 50/50. Berikut adalah analisis dari hasil tersebut:

1. Kinerja Naive Bayes
 - a) 80/20 Split
 - Naive Bayes mencapai akurasi tertinggi yaitu 99,03%. Ini menunjukkan bahwa dengan lebih banyak data pelatihan (80%), model Naive Bayes mampu belajar pola dengan sangat baik dan memberikan prediksi yang sangat akurat.
 - b) 70/30 Split
 - Akurasi sedikit menurun menjadi 98,71%, namun tetap sangat tinggi. Penurunan ini bisa jadi karena pengurangan data latih membuat model sedikit kurang efektif dalam mempelajari pola yang ada.
 - c) 60/40 Split
 - Akurasi lebih lanjut menurun menjadi 97,57%, yang menunjukkan bahwa model menjadi kurang efektif ketika jumlah data latih berkurang lebih jauh.
 - d) 50/50 Split
 - Meskipun data latih dan uji seimbang, Naive Bayes tetap mempertahankan akurasi yang sangat baik pada 97,67%. Hal ini menunjukkan bahwa model Naive Bayes cukup robust meskipun dengan pembagian data yang lebih seimbang.
2. Kinerja K-Nearest Neighbor (KNN)
 - a) 80/20 Split

- Sama seperti Naive Bayes, KNN juga mencapai akurasi tertinggi pada pembagian 80/20 dengan nilai 99,03%. Ini menunjukkan bahwa dengan lebih banyak data latih, KNN mampu memberikan prediksi yang sangat akurat.
- b) 70/30 Split
- Akurasi tetap tinggi pada 98,71%, menunjukkan bahwa KNN juga cukup andal dengan pembagian data ini, meskipun penurunan data latih mulai mempengaruhi performa.
- c) 60/40 Split
- Pada pembagian ini, akurasi KNN mencapai 98,06%, yang sedikit lebih baik dibandingkan Naive Bayes pada pembagian yang sama. Ini menunjukkan bahwa KNN tetap efektif meskipun data latih berkurang.
- d) 50/50 Split
- Akurasi KNN menurun lebih signifikan pada pembagian ini, dengan nilai 95,33%. Ini menunjukkan bahwa KNN mungkin lebih sensitif terhadap pengurangan data latih dibandingkan Naive Bayes, yang bisa menjadi indikasi bahwa KNN membutuhkan lebih banyak data untuk mempertahankan akurasinya.

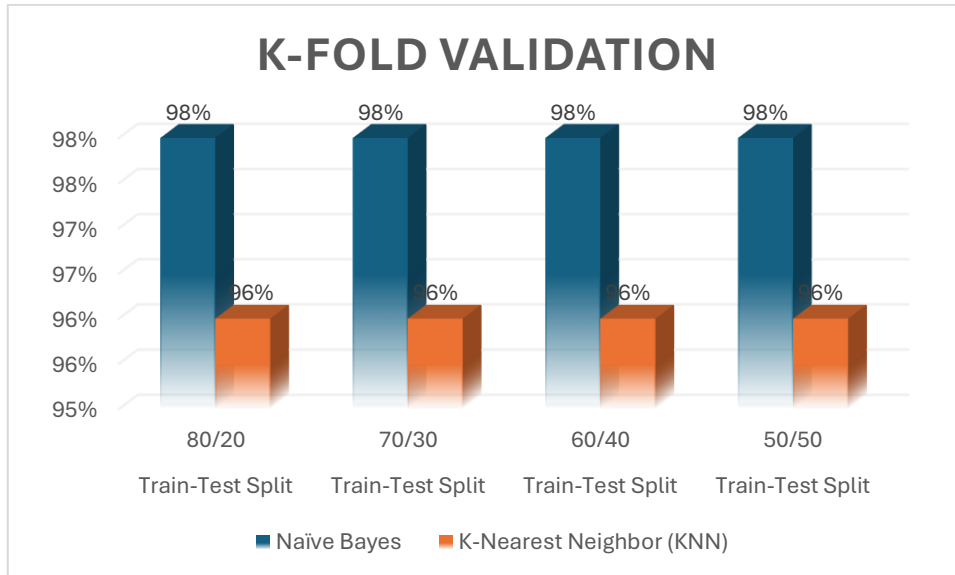
3. Analisis Perbandingan

Naive Bayes menunjukkan performa yang lebih konsisten dibandingkan KNN, terutama ketika data latih berkurang. Naive Bayes mampu mempertahankan akurasi yang lebih stabil bahkan ketika data latih berkurang secara signifikan (misalnya pada 50/50 split).

KNN tampaknya lebih sensitif terhadap perubahan dalam pembagian data, dengan penurunan akurasi yang lebih besar pada 50/50 split dibandingkan Naive Bayes. Hal ini mungkin karena KNN sangat bergantung pada jumlah data latih untuk menghitung jarak antar titik data, yang lebih banyak data latih biasanya berarti perhitungan jarak yang lebih akurat.

Kedua algoritma ini menunjukkan performa yang sangat baik dalam semua skenario pembagian data, namun Naive Bayes mungkin lebih unggul dalam situasi di mana data latih terbatas, karena kemampuannya untuk mempertahankan akurasi yang tinggi dengan data yang lebih sedikit. Sementara itu, KNN menunjukkan performa yang lebih kuat ketika data latih lebih banyak, tetapi kinerjanya menurun lebih signifikan ketika data latih berkurang.

Hasil Evaluasi K-Fold Validation.



Gambar 2 Hasil Akuras K-Fold Validation

Gambar diatas menunjukkan hasil validasi K-Fold untuk dua algoritma, yaitu Naive Bayes dan K-Nearest Neighbor (KNN), dengan empat variasi pembagian data train-test split: 80/20, 70/30, 60/40, dan 50/50.

1. Kinerja Naive Bayes:

- Naive Bayes menunjukkan performa yang sangat konsisten dengan akurasi 98% pada semua pembagian data (80/20, 70/30, 60/40, dan 50/50). Konsistensi ini menunjukkan bahwa Naive Bayes mampu memberikan hasil yang sangat andal dan stabil, terlepas dari bagaimana data latih dan uji dibagi.
- Kemampuan Naive Bayes untuk mempertahankan akurasi tinggi di berbagai skenario pembagian data menunjukkan bahwa model ini memiliki kemampuan generalisasi yang baik. Ini berarti Naive Bayes dapat bekerja dengan baik di berbagai situasi dengan variasi jumlah data latih dan uji.

2. Kinerja K-Nearest Neighbor (KNN)

- KNN menunjukkan akurasi yang lebih rendah dibandingkan Naive Bayes, dengan akurasi stabil di angka 96% untuk semua pembagian data. Meskipun lebih rendah, konsistensi ini tetap menunjukkan bahwa KNN mampu memberikan hasil yang stabil, meskipun tidak setinggi Naive Bayes.
- Meskipun KNN menunjukkan konsistensi, akurasinya tetap lebih rendah dibandingkan Naive Bayes, yang mungkin menunjukkan bahwa KNN sedikit lebih sensitif terhadap bagaimana data latih dan uji dibagi. Hal ini bisa disebabkan oleh fakta bahwa KNN sangat bergantung pada jarak antar titik data, yang bisa dipengaruhi oleh jumlah data latih.

3. Analisis Perbandingan

- Naive Bayes jelas unggul dalam semua skenario pembagian data, dengan akurasi yang lebih tinggi dan stabil. Model ini tampaknya tidak terlalu terpengaruh oleh variasi dalam jumlah data latih, yang menunjukkan bahwa Naive Bayes adalah pilihan yang sangat baik ketika diperlukan hasil yang andal di berbagai skenario data.
- Meskipun KNN memberikan hasil yang cukup baik dengan akurasi 96%, model ini berada di bawah Naive Bayes dalam hal akurasi. KNN mungkin lebih cocok digunakan dalam situasi di mana distribusi data lebih sederhana atau ketika data latih yang besar tersedia untuk memaksimalkan akurasi.

Naive Bayes menunjukkan performa yang lebih baik dibandingkan KNN dalam validasi K-Fold ini, dengan akurasi yang lebih tinggi dan lebih stabil di semua skenario pembagian data. Ini menunjukkan bahwa Naive Bayes memiliki keandalan yang lebih tinggi, terutama dalam situasi di mana hasil prediksi harus konsisten dan tidak terpengaruh oleh variasi dalam pembagian data latih dan uji. Di sisi lain, KNN tetap merupakan algoritma yang kuat, tetapi mungkin lebih cocok digunakan dalam situasi yang lebih spesifik di mana data latih yang lebih besar tersedia atau di mana sedikit penurunan akurasi bisa diterima.

