# Model Performance Comparison and Analysis

This report compares three classification models — Gaussian Naive Bayes, K-Nearest Neighbors (KNN), and Decision Tree — on a binary classification task using metrics such as Accuracy, Precision, Recall, F1 Score, and Confusion Matrix.

## Summary of Results

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Model** | **Accuracy** | **Precision** | **Recall** | **F1 Score** | **TP** | **FP** | **FN** | **TN** |
| **GaussianNB** | 0.93 | 0.94 | 0.86 | 0.9 | 32 | 2 | 5 | 61 |
| **KNN (k=3)** | 0.92 | 0.89 | 0.89 | 0.89 | 33 | 4 | 4 | 59 |
| **KNN (k=5)** | 0.93 | 0.89 | 0.92 | 0.91 | 34 | 4 | 3 | 59 |
| **KNN (k=7)** | 0.93 | 0.88 | 0.95 | 0.91 | 35 | 5 | 2 | 58 |
| **Decision Tree (Gini)** | 0.83 | 0.78 | 0.76 | 0.77 | 28 | 8 | 9 | 55 |
| **Decision Tree (Entropy)** | 0.83 | 0.78 | 0.76 | 0.77 | 28 | 8 | 9 | 55 |

## Key Observations

* **Best Performing Model (Overall):** KNN (k=5) and KNN (k=7) both perform best in terms of F1 Score (0.91), showing a strong balance between precision and recall.
* **GaussianNB Strength**: GaussianNB has the highest precision (0.94), making it suitable for cases where false positives are costly. It has high accuracy and F1 Score but a slightly lower recall.
* **KNN Trend**: KNN improves recall as k increases (k=3 to k=7), but precision drops slightly, with F1 Score remaining steady.
* **Decision Tree Limitation**: Decision Trees underperform, with lower accuracy and F1 scores. Both Gini and Entropy produce identical results and may be overfitting.

## Conclusion

The best-performing models in this comparison are **KNN with k=5 or k=7**, due to their balanced and high scores across all metrics.

* Choose **KNN (k=5)** if you prefer **slightly better precision**.
* Choose **KNN (k=7)** if **recall is more critical** (e.g., you want to minimize false negatives).

**Gaussian Naive Bayes** is a close second, especially if **model simplicity** and **speed** are required.

**Decision Trees** underperform across the board and are not recommended for this dataset.