Experiment 3: Classification Algorithm Comparisons

Name: Tarun Suresh

Register Number: 3122237001055

Date: 12/08/25

Department of Computer Science and Engineering
SSN College of Engineering

Academic Year: 2024–2025

Naïve Bayes Variant Comparison

Table 1: Performance Comparison of Naïve Bayes Variants

Metric	Gaussian NB	Multinomial NB	Bernoulli NB
Accuracy	0.8219	0.8706	0.8573
Precision	0.7233	0.8326	0.8004
Recall	0.9385	0.8547	0.8128
F1 Score	0.8170	0.8435	0.8065

KNN: Varying k Values

Table 2: KNN Performance for Different k Values

k	Accuracy	Precision	Recall	F1 Score
1	0.8958	0.8769	0.8769	0.8769
3	0.8936	0.8883	0.8564	0.8721
5	0.8958	0.9016	0.8462	0.8730
7	0.8921	0.8965	0.8423	0.8685

KNN: KDTree vs BallTree

Table 3: KNN Comparison: KDTree vs BallTree

Metric	KDTree	BallTree
Accuracy	0.8958	0.8958
Precision	0.9016	0.9016
Recall	0.8462	0.8462
F1 Score	0.8730	0.8730
Training Time (s)	0.012	0.015

SVM Performance

Table 4: SVM Performance with Different Kernels and Parameters

Kernel	Hyperparameters	Accuracy	F1 Score	Training Time (s)
Linear	C = 100	0.9444	0.8000	0.031
Polynomial	C = 0.1, degree = 1, gamma = auto	0.8930	0.8930	0.042
RBF	C = 100, gamma = 0.1	0.9290	0.9123	0.050
Sigmoid	C = 100, gamma = 0.1	0.8570	0.8502	0.047

K-Fold Cross-Validation Results

Table 5: Cross-Validation Scores for Each Model

Fold	Gaussian NB Acc.	Bernoulli NB Acc.	KNN Acc.	SVM Acc.
Fold 1	0.8261	0.8586	0.8967	0.9444
Fold 2	0.8282	0.8597	0.8956	0.9423
Fold 3	0.8173	0.8575	0.8956	0.9444
Fold 4	0.8204	0.8564	0.8945	0.9423
Fold 5	0.8204	0.8575	0.8945	0.9444
Average	0.8225	0.8579	0.8954	0.9436

Observations & Conclusions

- **Best overall accuracy:** SVM with Linear kernel (C = 100) achieved the highest average accuracy in K-Fold validation.
- **Best Naïve Bayes variant:** MultinomialNB had the highest single-split accuracy, though BernoulliNB was also competitive.
- **KNN trend:** Accuracy remained high for k = 1 to 7, with best stability at k = 5.
- **KDTree vs BallTree:** Both yielded identical accuracy, with KDTree slightly faster in training.
- **SVM kernels:** Linear kernel outperformed others; polynomial kernel lagged behind despite parameter tuning.