

Hyperparameter Tuning and Model Accuracy Analysis

1 Hyperparameter Tuning Comparison

The table below compares Grid Search and Bayesian Search results across models:

Model	Grid Best Score	Bayesian Best Score	Key Observations
Naive Bayes	79.78%	79.78%	Similar 'var_smoothing', no difference.
Logistic Regression	81.74%	82.16%	Bayesian found better 'C=0.49', L1 penalty.
KNN	83.95%	83.95%	Same optimal parameters ('n_neighbors=9', Manhattan metric).
Linear SVM	82.86%	82.86%	Bayesian chose smaller 'C=0.44', no accuracy change.
RBF SVM	83.01%	83.15%	Bayesian improved performance with 'C=8.93', 'gamma=0.01'.

Table 1: Comparison of Hyperparameter Tuning Methods

2 Final Model Performance

Validation and test accuracies of the best-tuned models:

Model	Validation Accuracy	Test Accuracy	Best Search	Key Hyperparameters
Naive Bayes	82.02%	79.33%	Grid	'var_smoothing=1e-09'
Logistic Regression	85.39%	82.68%	Bayesian	'C=0.49', L1 penalty
KNN	86.52%	80.45%	Grid	'n_neighbors=9', Manhattan metric
Linear SVM	86.52%	84.36%	Grid	'C=1'
RBF SVM	86.52%	84.36%	Bayesian	'C=8.93', 'gamma=0.01'

Table 2: Final Model Accuracy and Hyperparameter Summary

3 Critical Analysis

3.1 Hyperparameter Tuning Insights

- Bayesian search outperformed Grid Search in 3/5 models.
- RBF SVM saw a 0.35% accuracy gain using Bayesian search.
- KNN and Naive Bayes showed no benefit from Bayesian tuning.

3.2 Accuracy Trends

- **Overfitting in KNN:** Validation accuracy (86.52%) was much higher than test accuracy (80.45%).
- **Best Models:** SVM models performed best, both achieving 84.36% test accuracy.
- **Naive Bayes Limitation:** Lowest test accuracy (79.33%), likely due to simplistic assumptions.

4 Conclusion

- Bayesian search is optimal for complex models like SVM and Logistic Regression.
- Best model: **RBF SVM** (Test Accuracy: 84.36%, Parameters: 'C=8.93', 'gamma=0.01').
- KNN overfitting suggests the need for stricter cross-validation.