Comparison of Different Classifiers on Different Datasets Using Evaluation Metrics

Datasets:

 (from Scikit Learn and UCI ML repository)

- Iris
- Breast Cancer
- Heart
- Digits
- Wine

Classifiers:

- Decision Tree
- SVM
- KNN
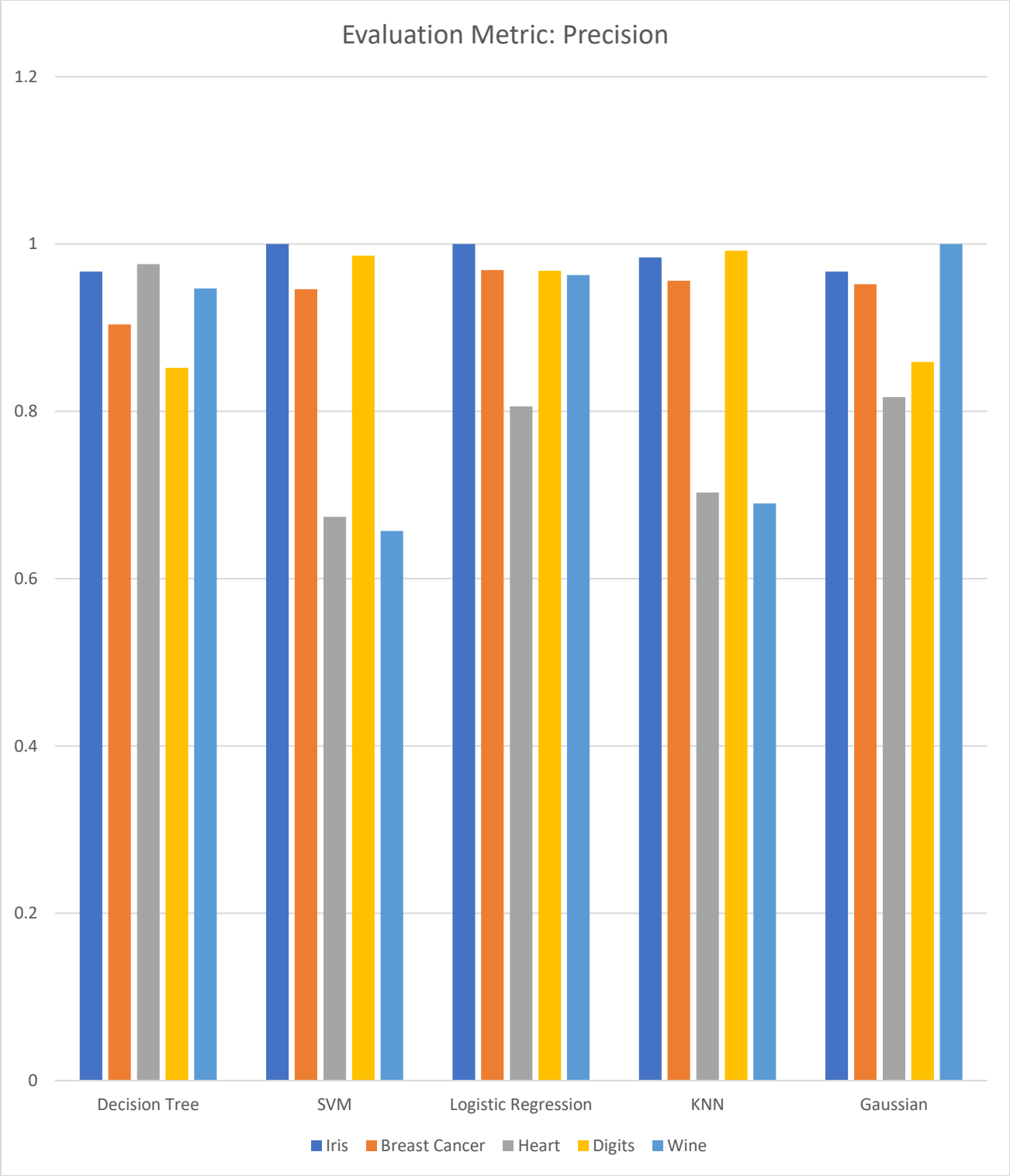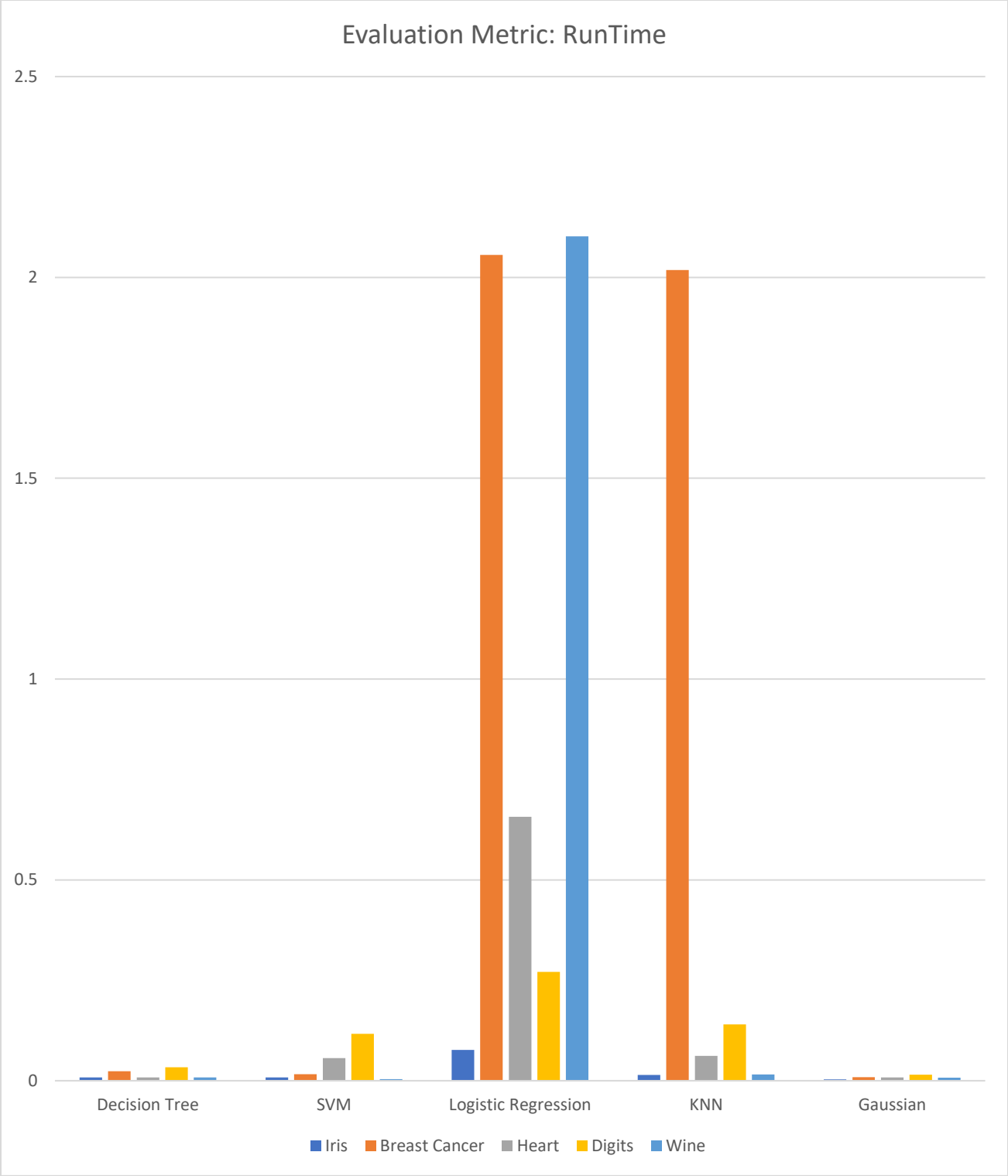- Logistic Regression
- Gaussian

Evaluation Metrics:

- precision
- Accuracy
- Runtime

All 5 algorithms were performed on each dataset and were evaluated based on their accuracy, precision, and runtime. The results are given in bar charts below for each evaluation metric.

Tools used: Jupyter Notebook

Evaluation Metric: Accuracy

Evaluation Metric: Precision

Evaluation Metric: RunTime
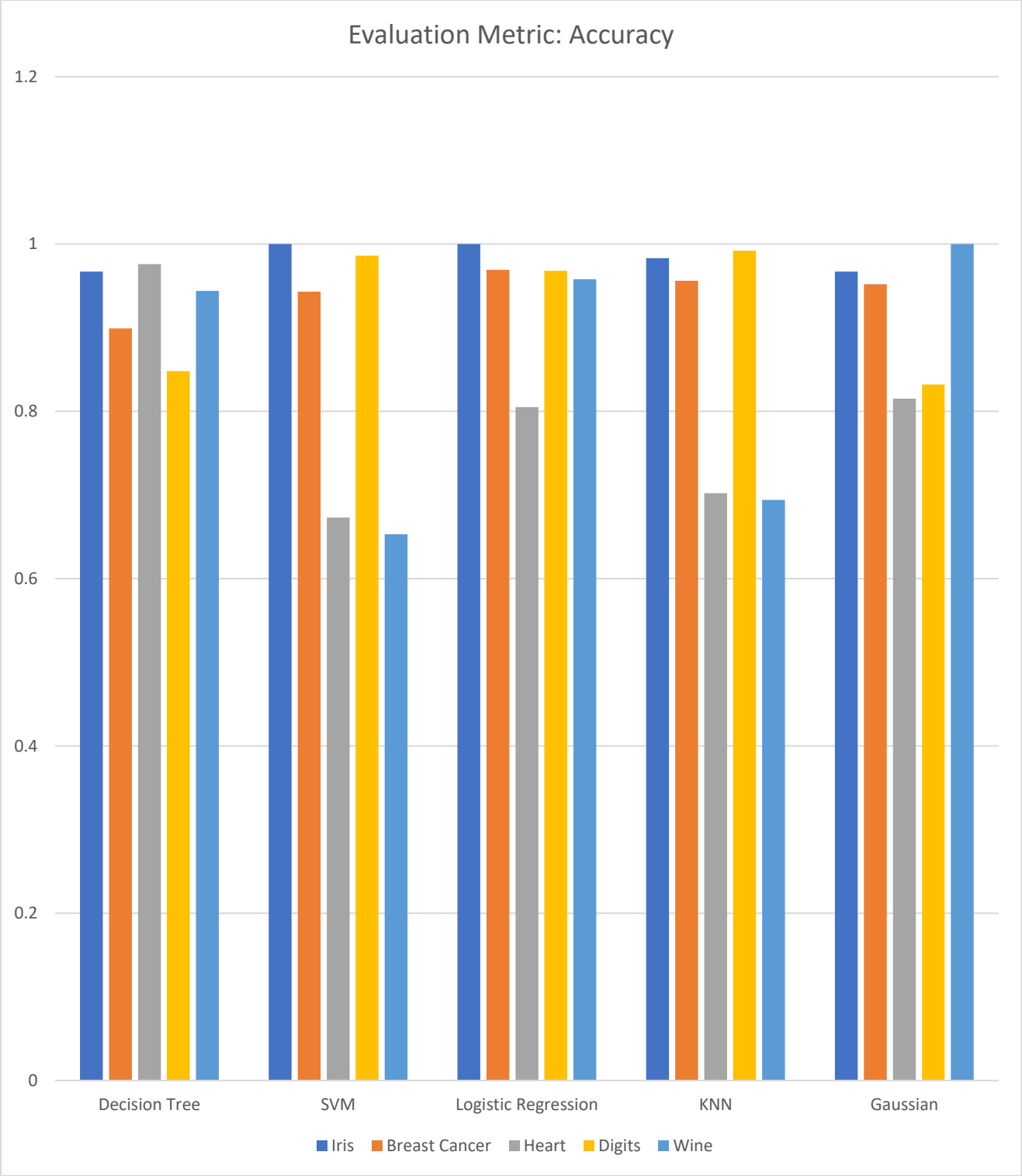
Conclusion:

For the Iris dataset, based on its accuracy, precision, and runtime, it can be said that SVM and Logistic Regression are the best two classifying algorithms for Iris. Whereas for the Breast Cancer dataset, even though Logistic Regression and KNN have the highest accuracy rates, they fall far behind in terms of Runtime. Overall, Gaussian can be considered the best one for Breast cancer. The Decision Tree algorithm is seen to be performing the best across all the metrics for the Heart dataset. For the Digits dataset, the k-nearest neighbor algorithm seems to be performing the best, and SVM the second best, as its accuracy and precision for this dataset are almost close to KNNs but with lower Runtime. Finally, for the Wine dataset, the Gaussian algorithm performs the best across all the evaluation metrics specified above, having 100% accuracy and precision and almost zero runtime.

Datasets	Preferred algorithms
Iris	SVM, Logistic Regression
Breast Cancer	Gaussian
Heart	Decision Tree
Digits	KNN, SVM
Wine	Gaussian