Classification Model Comparison

Classification Model	Pros	Cons
Logistic Regression	Probabilistic approach, interpretability through statistical perspective	Need to choose the right polynomial degree for a good bias/variance tradeoff
K-NN	Simple to understand, fast and efficient	Need to choose the number of neighbors k
SVM	Performant, not biased by outliers, not sensitive to overfitting	Not appropriate for nonlinear problems, not the best choice for large number of features
Kernel SVM	High performance on nonlinear problems, not biased by outliers, not sensitive to overfitting	Not the best choice for large number of features, more complex
Naive Bayes	Efficient, not biased by outliers, works on nonlinear problems, probabilistic approach	Based on the assumption that features have same statistical relevance
Decision Tree Classification	Interpretability, works on both linear / nonlinear problems	Poor results on too small datasets, overfitting can easily occur
Random Forest Classification	Powerful and accurate, good performance on many problems, including nonlinear	Low interpretability, need to choose the number of trees