

1. Model test and train accuracy 0.661
It's used a lot of models in real life

2. Logistic Regression

Probabilistic Approach, gives information about statistical significance of features.
The assumptions of logistic regression.

K - Nearest Neighbours

Simple to understand, fast and efficient. Need to manually choose the number of neighbours 'k'.

Support Vector Machine (SVM)

Performant, not biased by outliers, not sensitive to overfitting. Not appropriate for non-linear problems, not the best choice for large number of features.

Kernel SVM

High performance on non - linear 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 non - linear problems, probabilistic approach. Based in the assumption that the features have same statistical relevance.

Decision Tree Classification

Interpretability, no need for feature scaling, works on both linear / non - linear problems. Poor results on very small datasets, overfitting can easily occur.

Random Forest Classification

Powerful and accurate, good performance on many problems, including non - linear.
No interpretability, overfitting can easily occur, need to choose the number of trees manually.