

Best Machine Learning Algorithms for Classification:

1

Naive Bayes Classifier

PROS

- Simple, easy and fast
- Not sensitive to irrelevant features
- Works great in practice
- Needs less training data
- For both multi-class and binary classification
- Works with continuous and discrete data

CONS

- Assumes every feature is independent. This isn't always the truth.

2

Decision Trees

PROS

- Easy to understand
- Easy to generate rules
- There are almost null hyper-parameters to be tuned
- Complex Decision Tree models can be significantly simplified by its visualizations

CONS

- Might suffer from overfitting
- Does not easily work with non-numerical data
- Low prediction accuracy for a dataset in comparison with other algorithms
- When there are many class labels, calculations can be complex

3

Support Vector Machines

PROS

- Fast algorithm
- Effective in high dimensional spaces
- Great accuracy
- Power and flexibility from kernels
- Works very well with a clear margin of separation
- Many applications

CONS

- Doesn't perform well with large data sets
- Not so simple to program
- Doesn't perform so well, when the data comes with more noise i.e. target classes are overlapping

4

Random Forest Classifier

PROS

- The overfitting problem does not exist
- Can be used for feature engineering i.e. for identifying the most important features among the all available features in the training dataset
- Runs very well on large databases
- Extremely flexible and have very high accuracy
- No need for preparation of the input data

CONS

- Complexity
- Requires a lot of computational resources
- Time-consuming
- Need to choose the number of trees

5

KNN Algorithm

PROS

- Simple to understand and easy to implement
- Zero to little training time
- Works easily with multiclass data sets
- Has good predictive power
- Does well in practice

CONS

- Computationally expensive testing phase
- Can have skewed class distributions
- The accuracy can be decreased when it comes to high-dimension data
- Needs to define a value for the parameter k

<http://intellspot.com>