ML model	Assumptions	Advantages	Disadvantages	Feature Scaling	Missing Data	Outliers	Suitable for	Learning	Example Use
Naïve Bayes Classifier	Features are independent	• Performs well with categorical variables • Converges faster: less training time • Good with moderate to large training data sets • Good when dataset contains several features	• Correlated features affect performance	No	Can handle missing data (it ignores missing data)	Robust to outliers	• Classification • Multiclass classification	Supervised	• Sentiment Analysis • Document categorisation • Email Spam Filtering
Support Vector Machine (SVM)	None	• Good for datasets with more variables than observations • Good performance • Good of-the-shelf model in general for several scenarios • Can approximate complex non-linear functions	• Long training time required • Tuning is required to determine which kernel is optimal for non-linear SVMs	Yes	Sensitive	Robust to outliers	• Classification • Regression	Supervised	• Stock market forecasting • Value at risk determination
Linear Regression	• Linear relation between features and target • Residuals are normally distributed • Homoscedasticity	• Interpretability • Little tuning	• Correlated features may affect performance • Extensive feature engineering required	Yes	Sensitive	Sensitive	Regression	Supervised	• Sales forecasting • House pricing
Logistic Regression	• Linear relation between features and the log odds • Residuals are normally distributed • Homoscedasticity	• Interpretability • Little tuning	• Correlated features may affect performance • Extensive feature engineering required	Yes	Sensitive	Potentially sensitive	Classification	Supervised	• Risk Assessment • Fraud Prevention
Classification and Regression Trees	None	• Interpretability • Render feature importance • Less data pre-processing required	• Do not predict a continuous output (for regression) • It does not predict beyond the range of the response values in the training data. • Overfits	No	Some implementations do not need missing data imputation. The one in Scikit-learn does	Robust to outliers	• Classification • Regression	Supervised	• Risk Assessment • Fraud Prevention
Random Forests	None	• Interpretability • Render feature importance • Less data pre-processing required • Do not overfit (in theory) • Good performance /accuracy • Robust to noise • Little if any parameter tuning required • Apt for almost any machine learning problem	• Do not predict a continuous output (for regression) • It does not predict beyond the range of the response values in the training data • Biased towards categorical variables with several categories • Biased in multiclass problems toward more frequent classes	No	Some implementations do not need missing data imputation. The one in Scikit-learn does.	Robust to outliers	• Classification • Regression	Supervised	• Credit Risk Assessment • Predict breakdown of mechanical parts (automobile industry). • Assess probability of developing a chronic disease (healthcare) • Predicting the average number of social media shares
Gradient Boosted Trees	None	• Great performance • Apt for almost any machine learning problem • It can approximate most non-linear functions	• Prone to overfit • Needs some parameter tuning	No	Some implementations do not need missing data imputation (e.g. xgboost). The one in Scikit-learn does.	Robust to outliers	• Classification • Regression	Supervised	• Same as Random Forests
K-nearest neighbours	None	• Good performance	• Slow when predicting • Susceptible to high dimension (lots of features)	Yes	Sensitive	Robust to outliers	• Classification • Regression	Supervised	• Gene expression • Protein-protein interaction • Content retrieval (of webpages for example)
AdaBoost	None	• It doesn't overfit easily • Few parameters to tune	• Can be sensitive to noise and outliers	No	Can handle	Sensitive	• Classification • Regression	Supervised	• Same as Random Forests, less used however, as xgboost and lightGBMs are more popular implementations of gradient boosted machines
Neural Networks	None	• Can approximate any function • Great Performance	• Long training time • Several parameters to tune, including neuronal architecture • Prone to overfit • Little interpretability	Yes	Sensitive	Can handle outliers, and it affects performance if they are too many	• Classification • Regression	Supervised	• Image analysis • Forecasting • Text analysis
K-Means Clustering	• clusters are spherical • clusters are of similar size	• Fast training	• Need to determine k, the number of clusters • Sensitive to initial points and local optima	Yes	In the Scikit-learn implementation, missing data needs to be imputed	Sensitive	• Segmentation	Unsupervised	• Customer segmentation • Outlier detection
Hierarchical clustering		• No a priori information about the number of clusters required	• Final number of clusters to be decided by the scientist • Slow training	Yes	Sensitive	Sensitive	• Segmentation	Unsupervised	• Customer segmentation • Gene analyses
PCA	• Correlation among features	• Captures most of the variance in a smaller number of features	• Number of principal components that explain most of the variance to be determined by the user	Yes	Sensitive	Sensitive	Reducing feature space to train machine learning models	Unsupervised	• Creating few, informative, variables from tons of data