NAME	ADVANTAGES	DISADVANTAGES
CLASSIFICATION		
Decision Tree	– Inexpensive to construct. – Extremely fast at classifying unknown records. – Easy to interpret for small-sized trees. – Accuracy is comparable to other classification techniques for many simple data sets.	– May not be able to capture complex, non-linear dependencies between attributes.
Nearest Neighbor Classifier	– Easy to implement. – Incremental addition of training data trivial.	– k-NN classifiers are lazy learners, which do not build models explicitly. This can be relatively more expensive than eager learners (such as decision tree) when classifying a test/unknown record. – Unlike decision trees that attempt to find a global model that fits the entire input space, nearest neighbor classifiers make the prediction based on local information, which can be more susceptible to noise.
Naïve Bayes Classifier	– Robust to isolated noise points. – Missing values can be handled by ignoring the instances during probability estimate calculations. – Robust to irrelevant attributes.	– Independence assumption may not hold for some attributes. – Other techniques such as Bayesian Belief Networks (BBN).
Ensemble Classifier – Bagging	– Decreases variance, improves stability (tolerance to noise). – Can be parallelized.	– Reduces accuracy for stable classifiers because sample size reduced by 36%!
Ensemble Classifier – Boosting	– Because the weights of previously misclassified records are increased during training, may produce a more robust model.	– Cannot be parallelized easily.

NAME	ADVANTAGES	DISADVANTAGES
CLUSTERING		
K-Means Clustering	– Most clusterings converge in the first few iterations.	– Different initial centroids may result in very different clusterings. – Issues when clusters are of different sizes, densities and non-globular shapes. – Cannot cluster data with outliers well. – One possible solution is to create > k clusters and then merge, as necessary.
Hierarchical Clustering	– Does not assume any particular number of clusters since the dendrogram can be cut at any level to get the desired number of clusters. – May correspond to meaningful taxonomies. – $O(N^2)$ space and $O(N^3)$ or O(N lg N) time in many cases. Very computationally expensive. – Once a decision is made to combine two clusters, it cannot be undone. – No objective function is directly minimized. – MIN or Single Link can handle non-elliptical shapes but is sensitive to noise & outliers. – MAX or Complete Linkage is less susceptible to noise & outliers but tends to break large clusters and is biased towards globular clusters. Group Average has the same disadvantages, but to a lesser degree. – Ward’s Method (based on increase in squared error) also has the same advantages and disadvantages as MAX / Group Average.	
DBSCAN	– Resistant to noise. – Can handle clusters of different shapes & sizes.	– Does not handle high-dimensional data well. – Does not handle clusters of varying densities well. – Epsilon & MinPoints need to be determined empirically.
CURE	– Shrinking representative points towards the center helps avoid problems with noise & outliers. – CURE can handle clusters of arbitrary shapes & sizes.	– Cannot handle clusters of differing densities.

NAME	ADVANTAGES	DISADVANTAGES
Graph-Based Clustering	– Sparsification drastically reduces the amount of data that needs to be processed. Hence, the time needed is reduced and the problem size can be increased. – Sparsification also reduces the impact of noise & outliers since they are disconnected from the other points, which are only connected to their nearest neighbors.	
Chameleon (Graph-Based Algorithm)	– Existing merging schemes (MIN/MAX/AVG) are static in nature. – Chameleon uses a dynamic model that adapts to the characteristics of the data to find the natural clusters. – Allows clusters that vary in shape, density, form, size & orientation.	
Jarvis-Patrick Clustering (SNN Algorithm)	– Advantages of sparsification. – Can be combined with DBSCAN after the SNN graph is constructed.	– Clustering may be too brittle. – The value of the threshold can affect the clustering and must be determined empirically. – Does not cluster all the points. – Complexity is high. $O(N^2)$.