

Parameters	Lin.Reg	KNN	Log.Reg	SVM	DT	RF
Def.	establishing relationships between dependent & independent variables.	It uses the distance or similarity measure in order to classify data. It uses the Euclidean distance for classifying a particular category.	Logistic regression is a statistical analysis method to predict a binary outcome, such as yes or no, based on prior observations of a data set. A logistic regression model predicts a dependent data variable by analyzing the relationship between one or more existing independent variables.	Support vectors are the coordinates or data points that are close to your hyperplane which are also used to create the boundary lines. Hyperplane is a decision boundary that separates two classes/categories.	DT will create multiple decision boundaries so that it can classify the classes, data points correctly so that there is less amount of impurities in it. The multiple decision boundaries are created on the basis of multiple features.	RF builds the multiple decision trees and merges them together to get a more accurate and stable prediction. In order to classify a new object based on an attribute each tree gives a classification and we say the tree "votes" for that class.
Aim/Goal	Lin.Reg	KNN	Log.Reg	SVM	DT	RF
	create the best fit line by using equation $y=mx+c$	to locate all of the closest neighbours around a new unknown data point in order to figure out what class it belongs to. It's a distance-based approach	plot a best-fit line which is a curve Logistic Regression Sets a default threshold of 0.5 for classifying the new observation. We apply the sigmoid function to linear regression to get the s curve & it also reduces the range to (0 to 1)	To create boundary lines, to create points in 1D subspace, line in 2D subspace, plane in 3D subspace & hyperplane in more than 3D subspace.	Creation of multiple decision boundaries/hyperplanes so that there is no misclassification. i.e. data is classified perfectly or less impurities in data.	It builds decision trees on different samples and takes their majority vote for classification and average in case of regression.

Objective	Lin.Reg	KNN	Log.Reg	SVM	DT	RF
	Establishing the relationship between x & y.	How likely a data point is to be a member of one group or another depending on what group data points are nearest to it.	to train a model that can make a binary decision about the class of a new observation. Here we introduce the sigmoid function that will help us make this decision.	is to find the best line in two dimensions or the best hyperplane in more than two dimensions in order to help us separate our space into classes.	To classify the particular category perfectly from a mixture of two or more classes with very less or no impurities.	To combine multiple algorithms for getting better predictive performance.
Advantages	Lin.Reg	KNN	Log.Reg	SVM	DT	RF
	Easy to interpret	Easy Implementation Gives High Accuracy	1.It is a Parametric model hence easier to implement, interpret, and very efficient to train. 2.Logistic regression is used when your Y variable can take only two values , and if the data is linearly separable, it is more efficient to classify it into two separate classes. 3. It is fast in training the model.	1.It is a non parametric ,non linear model. 2. It gives high performance/ Accuracy. 3.It is more productive in high dimensional spaces.	It is used to solve non-linear problem. It can efficiently work on high dimension data (i.e. data with many features/columns) Feature selection is automatic.(Select best features automatically) Easy to Visualize.	It reduces Variance and overfitting. It handles missing values on its own. It is robust to outliers. Less preprocessing is required. Efficient for both linear and nonlinear. Stable algorithm even if a new data point is introduced in the dataset the overall algorithm will

						not be affected much since the new data point will affect the one tree but the other trees will work well.
Disadvantages	Lin.Reg	KNN	Log.Reg	SVM	DT	RF
	Highly affected by outliers, missing values and skewness.	It is Computationally Expensive. Does not work well with large datasets as calculating distances between each data point would be very costly. Sensitive to Missing Data. Does not work well with high dimensionality as this will complicate the distance calculation process to calculate distance for each dimension. Gives Overfitted results	<p>1.Low performance and accuracy.</p> <p>2.If the number of observations is lesser than the number of features, Logistic Regression should not be used, otherwise, it may lead to overfitting.</p> <p>3.Non-linear problems can't be solved with logistic regression</p> <p>4.The major limitation of Logistic Regression is the assumption of linearity between the dependent variable and the independent variables.</p>	<p>1.Being non parametric model creates very weak or no assumptions.</p> <p>2. It requires more data.</p> <p>3. Slow in training the data.</p> <p>4. Gives Overfitted results.</p> <p>5. Difficult to interpret.</p>	It overfits the model. It is Non-Parametric so it takes more data and time to process it. It is computationally expensive. It is difficult to interpret.	It is a complex algorithm since it creates 100 Decision Trees at a time . It takes a long time to train the model. It is difficult to interpret . (Non Parametric Model)