The datasets (data features, data dimensions)

***Titanic***

12 features

892 rows

***Flight***

11 features

476,881 rows

***California Housing***

10 features

12,640 rows

***Wisconsin Breast Cancer***

30 features

569 rows

***Stocks***

2 features

649 rows

***Earthquakes***

15 features

8223 rows

***Big Mart***

12 features

14,204 rows

***Space***

16 features

1,048,576 rows

***Regression***

*Linear Regression*

Linear regression uses a linear approach to model the relationship between a dependent variable and one or more independent variables.

*Logistic Regression*

Logistic regression is used to describe data and to explain the relationship between one dependent binary variable and one or more independent variables.

*Lasso Regression*

Lasso (least absolute shrinkage and selection operator) is a regression analysis method that performs both variable selection and regularization in order to enhance the prediction accuracy and interpretability of the statistical model it produces.

***Support Vector Machine***

SVMs are supervised learning models for classification and regression analysis that maps data categories to a “hyperplane” that distinctly separates the points.

***Decision Trees***

A decision tree is a tree where each node represents a feature(attribute), each link(branch) represents a decision(rule) and each leaf represents an outcome or value.

***Random Forest***

Random forest builds multiple decision trees and merges them together to get a more accurate and stable prediction.

***K-Nearest Neighbor***

KNN is a lazy learning algorithm that does not make any assumptions on the underlying data distribution. It is used where data is separated into several classes to predict the classification of a new sample point. K is the specified number of neighbors close to the point.

***Neural Networks***

Digital brains. An interconnected group of nodes, like neurons in a brain working together.

***ElasticNet***

ElasticNet is an algorithm that builds on Lasso and Ridge, “convex sum of penalties” or something like that…

***K-Means***

K-means is unsupervised learning algorithm that is used to find groups (“clusters”) which have not been explicitly labeled in the data and to find patterns. The “K’ is the number of clusters.

***Naïve Bayes***

Naïve Bayes algorithms are "probabilistic classifiers" assuming strong independence between the features the data set. The assumption is why they are called “naive.”