







TOP PREDICTION ALGORITHMS

	TYPE	NAME	DESCRIPTION	ADVANTAGES	DISADVANTAGES
Linear		Linear regression	The "best fit" line through all data points. Predictions are numerical.	Easy to understand -- you clearly see what the biggest drivers of the model are.	<ul style="list-style-type: none"> X Sometimes too simple to capture complex relationships between variables. X Does poorly with correlated features.
		Logistic regression	The adaptation of linear regression to problems of classification (e.g., yes/no questions, groups, etc.)	Also easy to understand.	<ul style="list-style-type: none"> X Sometimes too simple to capture complex relationships between variables. X Does poorly with correlated features.
Tree-based		Decision tree	A series of yes/no rules based on the features, forming a tree, to match all possible outcomes of a decision.	Easy to understand.	<ul style="list-style-type: none"> X Not often used on its own for prediction because it's also often too simple and not powerful enough for complex data.
		Random Forest	Takes advantage of many decision trees, with rules created from subsamples of features. Each tree is weaker than a full decision tree, but by combining them we get better overall performance.	A sort of "wisdom of the crowd". Tends to result in very high quality models. Fast to train.	<ul style="list-style-type: none"> X Models can get very large. X Not easy to understand predictions.
		Gradient Boosting	Uses even weaker decision trees, that are increasingly focused on "hard" examples.	High-performing.	<ul style="list-style-type: none"> X A small change in the feature set or training set can create radical changes in the model. X Not easy to understand predictions.
Neural networks		Neural networks	Interconnected «neurons» that pass messages to each other. Deep learning uses several layers of neural networks stacked on top of one another.	Can handle extremely complex tasks - no other algorithm comes close in image recognition.	<ul style="list-style-type: none"> X Very slow to train, because they often have a very complex architecture. X Almost impossible to understand predictions.