data iku

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.	X **Sometimes too simple to capture complex relationships between variables.** X **Tendency for the model to "overfit".**
	Logistic regression	The adaptation of **linear regression** to problems of classification (e.g., yes/no questions, groups, etc.)	Also easy to understand.	X **Sometimes too simple to capture complex relationships between variables.** X **Tendency for the model to "overfit".**
Tree-based	Decision tree	A graph that uses a **branching method** to match all possible outcomes of a decision.	Easy to understand and implement.	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 the average of many decision trees, each of which is made with a sample of the data. 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.	X **Can be slow to output predictions relative to other algorithms.** X **Not easy to understand predictions.**
	Gradient Boosting	Uses even weaker decision trees, that are increasingly **focused on "hard" examples.**	High-performing.	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	Mimics the behavior of the brain. Neural networks are interconnected neurons that pass messages to each other. Deep learning uses several **layers of neural networks put one after the other.**	Can handle extremely complex tasks - no other algorithm comes close in image recognition.	X **Very, very slow to train, because they have so many layers. Require a lot of power.** X **Almost impossible to understand predictions.**