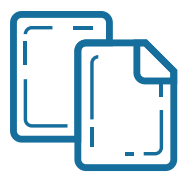


# Choosing your machine learning algorithm



## Is the data labeled?

Data contains ground truth about patients' disease status

Yes

No

Make predictions

Explore variation

Find groups

Continuous outcomes

Ex: Cost of care in \$

Discrete outcomes

Ex: Diabetes Y/N

Regression

Classification

Listed by increasing sophistication

- |  |  |
|--|--|
| <div>Linear Model</div> <div><div>Lasso</div><div>Ridge</div></div> <div>Random Forest</div> <div>k-Nearest Neighbors</div> <div>Linear Mixed Model</div> <div>XGBoost</div> <div>Neural Network</div> | <div>Linear Model</div> <div><div>Lasso</div><div>Ridge</div></div> <div>Random Forest</div> <div>k-Nearest Neighbors</div> <div>Support Vector Machine</div> <div>Linear Mixed Model</div> <div>XGBoost</div> <div>Neural Network</div> |
|--|--|

Pick a measure

Ex: Cost of care in \$

Compare within categories

Find Variation

Visualize difference between groups

Variation Across Groups

Need to group the data?

Clustering

K-means

Need to reduce dimensions?

Dimension Reduction

Principle Component Analysis