30

MACHINE LEARNING ALGORITHMS USED BY DATA SCIENTISTS

SWIPE NEXT

CLOUDYML

1. LINEAR REGRESSION

Used for predicting a continuous output variable based on one or more input variables

2. LOGISTIC REGRESSION

Used for binary classification problems

3. POLYNOMIAL REGRESSION

Used for modeling relationships between input and output variables that are not linear

Akash Raj
Data Scientist

SWIPE NEXT

CLOUDYML

4. RIDGE REGRESSION

Used to prevent overfitting in linear regression by adding a penalty term to the loss function

5. LASSO REGRESSION

Used to perform variable selection in linear regression by shrinking coefficients towards zero

6. ELASTICNET REGRESSION

Combines the Ridge and Lasso regularization techniques to handle both multicollinearity and variable selection

Akash Raj
Data Scientist

SWIPE NEXT

CLOUDYML

7. DECISION TREES

Used for both classification and regression problems by recursively partitioning the input space into smaller regions

8. RANDOM FORESTS

An ensemble of decision trees that reduces variance and improves generalization by averaging the results of multiple trees

9. GRADIENT BOOSTING MACHINES

An ensemble of decision trees that improves performance by fitting each subsequent tree to the residual errors of the previous trees

Akash Raj
Data Scientist

SWIPE NEXT

CLOUDYML

10. EXTREME GRADIENT BOOSTING (XGBOOST)

A highly optimized implementation of gradient boosting that includes many advanced features

11. LIGHT GRADIENT BOOSTING MACHINE (LIGHTGBM)

Another optimized implementation of gradient boosting that can handle large datasets and high-dimensional features

12. CATBOOST

A gradient boosting algorithm that can handle categorical features and missing values

Akash Raj
Data Scientist

SWIPE NEXT

CLOUDYML

13. K-NEAREST NEIGHBORS (KNN)

A non-parametric algorithm that predicts the output of a new example by finding the K closest examples in the training data and taking their average

14. SUPPORT VECTOR MACHINES (SVM)

A parametric algorithm that finds the hyperplane that separates two classes with the maximum margin

15. NAIVE BAYES

A probabilistic algorithm that uses Bayes' theorem to calculate the probability of a new example belonging to each class

Akash Raj
Data Scientist

SWIPE NEXT

CLOUDYML

16. ARTIFICIAL NEURAL NETWORKS

A set of algorithms that attempt to mimic the structure and function of the human brain

17. CONVOLUTIONAL NEURAL NETWORKS (CNN)

A type of neural network that is particularly good at image recognition and other tasks that involve spatial data

18. RECURRENT NEURAL NETWORKS (RNN)

A type of neural network that is designed for processing sequential data, such as time series or natural language

Akash Raj
Data Scientist

SWIPE NEXT

CLOUDYML

19. LONG SHORT-TERM MEMORY (LSTM)

A type of RNN that can remember information over longer periods of time and avoid the vanishing gradient problem

20. AUTOENCODERS

A type of neural network that is used for dimensionality reduction, data compression, and unsupervised feature learning

21. GENERATIVE ADVERSARIAL NETWORKS (GAN)

A type of neural network that can generate new examples that are similar to the training data by learning to generate realistic examples while being trained to distinguish between real and fake examples

Akash Raj
Data Scientist

SWIPE NEXT

CLOUDYML

22. PRINCIPAL COMPONENT ANALYSIS (PCA)

A technique for reducing the dimensionality of high-dimensional datasets by projecting them onto a lower-dimensional space that captures most of the variation

23. T-SNE

t-Distributed Stochastic Neighbor Embedding (t-SNE) is a technique for visualizing high-dimensional datasets in two or three dimensions by mapping similar examples to nearby points and dissimilar examples to distant points.

24. K-MEANS CLUSTERING

A unsupervised learning algorithm that partitions a dataset into K clusters by minimizing the distance between each example and the centroid of its cluster

Akash Raj
Data Scientist

SWIPE NEXT

CLOUDYML

25. HIERARCHICAL CLUSTERING

A unsupervised learning algorithm that builds a hierarchy of clusters by recursively merging the closest pairs of clusters

26. GAUSSIAN MIXTURE MODELS (GMM)

A probabilistic clustering algorithm that models each cluster as a Gaussian distribution and estimates the parameters using the expectation-maximization algorithm

27. HIDDEN MARKOV MODELS (HMM)

A type of generative model that is used for sequence modeling, such as speech recognition, natural language processing, and bioinformatics

Akash Raj
Data Scientist

SWIPE NEXT

CLOUDYML

28. LINEAR DISCRIMINANT ANALYSIS (LDA)

A supervised learning algorithm that finds a linear combination of features that maximizes the separation between two or more classes

29. QUADRATIC DISCRIMINANT ANALYSIS (QDA)

Similar to LDA, but allows for quadratic decision boundaries

30. ENSEMBLE LEARNING

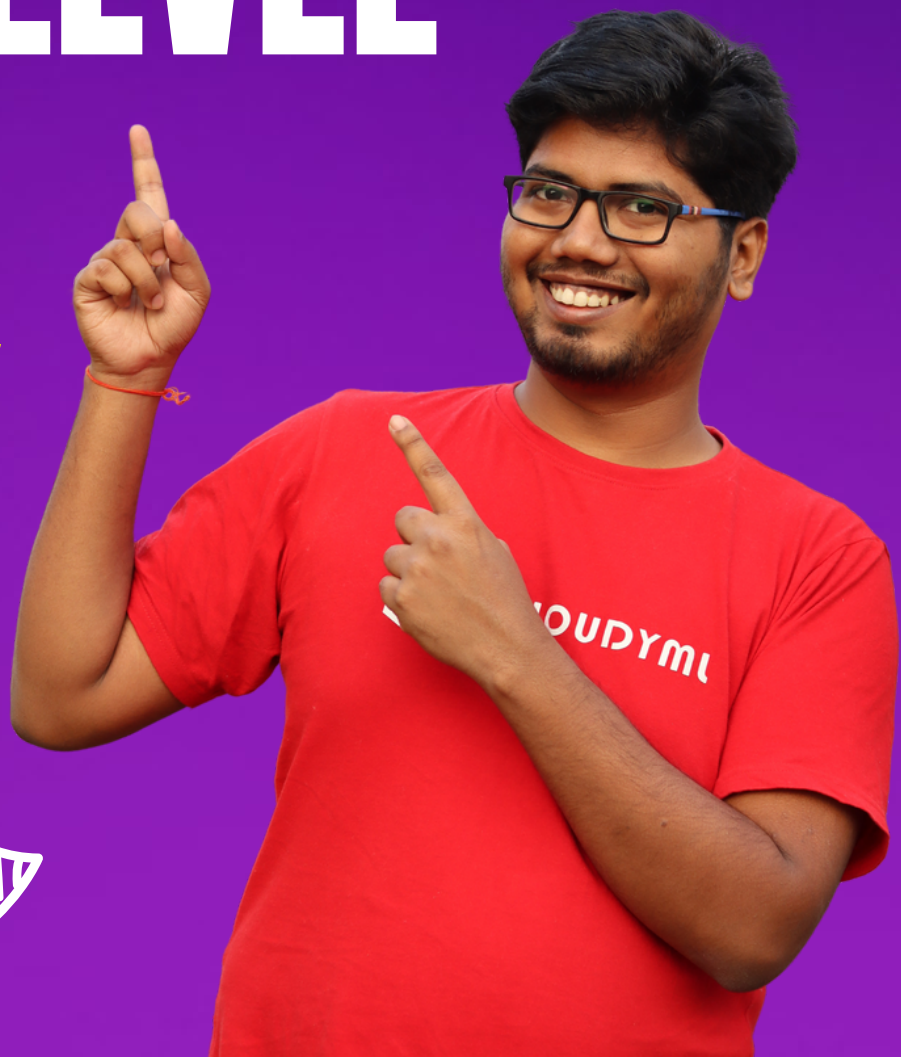
A technique for combining multiple models to improve performance, reduce variance, and increase robustness. Examples include bagging, boosting, and stacking

Akash Raj
Data Scientist

SWIPE NEXT

FROM BEGINNER TO PRO

LET'S TAKE YOUR DATA SCIENCE SKILLS TO THE NEXT LEVEL



FOLLOW ME TODAY

Akash Raj
Data Scientist