IMPORTANT

MACHINE LEARNING ALGORITHMS - BY CLOUDYML

* Linear Regression

→ It is used to predict real values. A relationship is established between independent and dependent variables by fitting a best fit line. This best fit line is known as regression line and represented as $Y = a * X + b$

Y - Dependent variable a - slope
X - Independent variable b - intercept

Akash Raj
(Data Scientist)
35K+
33K+
46K+
Swipe Next

Date / /

* Logistic Regression

→ It is a classification algorithm used to estimate discrete values (values like 0/1, yes/no) based on given set of independent variables. Its output value lies between 0 and 1. Uses the sigmoid function.

* Decision Tree

→ It is a supervised learning algorithm used for classification problems which works for both categorical and continuous dependent variables. In this algorithm, the population is split into two or more homogenous sets. It uses different techniques like Gini Impurity, Information Gain, Chi-square entropy.


Akash Raj
(Data Scientist)
35K+
33K+
46K+
Swipe Next

* SVM (Support Vector Machine)

→ In this algorithm, each data item is plotted as a point in a n-dimensional space (n- numbers of features) with the value of each feature being the value of a particular co-ordinate. A classifier line is then found between the features. Depending on where the testing data lands, the data is classified accordingly.

* Random Forest

→ A collection of decision trees is called as Random Forest. To classify a new object based on its attributes, each tree is classified and the tree votes for that class. The forest chooses the classification having the most votes.

Akash Raj
(Data Scientist)
35K+
33K+
46K+
Swipe Next

Date / /

* KNN (K Nearest Neighbors)

→ K Nearest Neighbors is a simple algorithm that stores all available cases and classifies new cases by a majority vote of its K neighbors. The case being assigned to the class is most common amongst its K nearest neighbors measured by a distance function.

* K-Means

→ It is a type of unsupervised algorithm which solves the clustering problem. Its procedure follows a simple way to classify a given data set into certain numbes of clusters (K clusters).


Akash Raj
(Data Scientist)
35K+
33K+
46K+
Swipe Next

* Naive Bayes Algorithm

→ It is a supervised learning algorithm which is based on Bayes theorem and used for solving classification problems. Bayes theorem is used to determine the probability of a hypothesis with prior knowledge.

$$P(A|B) = \frac{P(B|A)\,P(A)}{P(B)}$$

* Gradient Boosting Algorithm

→ The main idea behind this algorithm is to build models sequentially and these subsequent models try to reduce the errors of the previous models. Then these models are combined to improve the accuracy score.


Akash Raj
(Data Scientist)
35K+
33K+
46K+
Swipe Next

CLOUDYML

MAKE YOUR CAREER IN
DATA SCIENCE & ANALYTICS

HAND-ON LEARNING

1-1 DOUBT CLEARANCE SUPPORT

CAPSTONE END-TO-END PROJECTS

- Python
- SQL
- Statistics
- Machine Learning
- Deep Learning
- ML Projects
- DL Projects
- AWS Sagemaker
- Tableau
- Excel
- PowerBi
- Amazon QuickSight
- Google Data Studio
- Interview QnA PDF

VISIT OUR WEBSITE FOR MORE DETAILS

WWW.CLOUDYML.COM