



# TEXT MINING ALGORITHMS

1

## K-MEANS CLUSTERING

Aims to find groups in given data set. The number of groups is represented by a variable called K.

2

## NAIVE BAYES CLASSIFIER

A family of algorithms which assume that values of the features used in the classification are independent.

3

## K-NEAREST NEIGHBOR (KNN)

A simple algorithm that stores all existing data objects and classifies the new data objects based on a similarity measure.

4

## SUPPORT VECTOR MACHINES (SVM)

Used to sort two data sets by similar classification. Draw lines (hyperplanes) that separate the groups according to some patterns.

5

## DECISION TREE

A machine learning technique for data mining that creates classification or regression models in the shape of a tree structure.

6

## GENERALIZED LINEAR MODELS (GLM)

Combines a number of models including linear regression models, logistic regression, Poisson regression, ANOVA, log-linear models and etc.

7

## NEURAL NETWORKS

Nonlinear models which represent a metaphor for the functioning of the human brain.

8

## ASSOCIATION RULES

If/then statements that aim to uncover some relationships between unrelated data in a given database.

9

## GENETIC ALGORITHMS

A family of stochastic search algorithms with mechanism is inspired by the process of neo-Darwinian evolution.

10

## LATENT DIRICHLET ALLOCATION (LDA)

A generative probabilistic model designed for collections of discrete data.