

Machine Learning Algorithms and their Applications

This whitepaper covers various ML algorithms and their applications in various areas.

Naive Bayes Classifier Algorithm :

It classifies words based on the popular Bayes Theorem of probability and is used in applications related to document classification, spam filters and sentiment analysis.

Applications:

- Text Classification: Naive Bayes classifiers mostly used in text classification - Automatically classify web pages, forum posts, blog snippets and tweets without manually going through them.
- Ranking Pages: We can use the Naive Bayes Classifier Algorithm for ranking pages, indexing relevancy scores and classifying data categorically.
- Recommendation System: Naive Bayes Classifier and Collaborative Filtering together builds a Recommendation System that leverages machine learning and data mining to filter unseen information and predict whether a user would like a given resource (item, movie) or not

K-Means Clustering Algorithm

It is an unsupervised Machine Learning Algorithm that is used in cluster analysis. It works by categorizing unstructured data into several different groups 'k' being the number of groups. Each dataset contains a collection of features and the algorithm classifies unstructured data and categorizes them based on specific features.

Applications:

- Data Segmentation: Leveraged for segmenting data by purchase history, classifying persons based on different interests, grouping inventories by manufacturing and sales metrics, etc.
- Customer Segmentation: Clustering helps marketers improve their customer base by segmenting customers based on purchase history, interests, or activity monitoring
- Insurance Fraud Detection: Fraud detection has numerous applications in automobile, healthcare, and insurance. Utilizing past historical data on fraudulent claims, it is possible to isolate new claims based on its proximity to clusters that indicate fraudulent patterns.

Support Vector Machine (SVM) Learning Algorithm

SVM Algorithm is a supervised learning algorithm, that classifies data sets into different classes through a hyperplane. It marginalizes the classes and maximizes the distances between them to provide unique distinctions.

Applications:

- Face detection – SVM classify parts of the image as a face and non-face and create a square boundary around the face.
- Comparing stocks performance: It is used in applications such as comparing the relative performance of stocks over a period.
- Classification of images: Use of SVMs provides better search accuracy for image classification

Recommender System Algorithm

The Recommender Algorithm works by filtering and predicting user ratings and preferences for items by using collaborative and content-based techniques. The algorithm filters information and identifies groups with similar tastes to a target user and combines the ratings of that group for making recommendations to that user

Applications:

- Product Recommendation: Personalize the online store on an e-commerce application based on user preferences.
- Movie Recommendations: Recommend shows of a similar genre to the user

Linear Regression

It is basically used to showcase the relationship between dependent and independent variables and show what happens to the dependent variables when changes are made to independent variables.

Applications:

- Sales forecasting
- Risk assessment analysis in health insurance companies.
- Predict real estate value based on location, size and other factors
- Prediction of ETA's in traffic

Logistic Regression

Logistic Regression is a statistical analysis technique which is used for predictive analysis. It uses binary classification to reach specific outcomes and models the probabilities of default classes.

Applications:

- To Identifying risk factors for diseases and planning preventive measures
- Classifying words as nouns, pronouns, and verbs

- Weather forecasting applications for predicting rainfall and weather conditions
- Leveraged by credit card companies to develop models which decide whether a customer will default on their loan EMIs or not.

Decision Tree Machine Learning Algorithm

Its model is operational in nature, and depending on our answers, the algorithm will use forward, and backward calculation steps to arrive at different conclusions.

Applications:

- Determination of likely buyers of a product using demographic data to enable targeting of limited advertisement budget
- Evaluation of brand expansion opportunities for a business using historical sales data

Random Forest ML Algorithm

The Random Forest ML Algorithm is a versatile supervised learning algorithm that's used for both classification and regression analysis tasks. It creates a forest with several trees and makes them random. Although like the decision trees algorithm, the key difference is that it runs processes related to finding root nodes and splitting feature nodes randomly.

Applications:

- Finding whether a loan applicant is low-risk or high-risk
- Predicting the failure of mechanical parts in automobile engines
- Predicting social media share scores

Principal Component Analysis (PCA) Algorithm

The Principal Component Analysis (PCA) is a dimensionality reduction algorithm, used for speeding up learning algorithms and can be used for making visualizations of complex datasets. It identifies patterns in data and aims to make correlations of variables in them. Whatever correlations the PCA finds is projected on a similar (but smaller) dimensional subspace.

Applications:

- Gene expression analysis
- Stock market predictions

Artificial Neural Networks

Artificial Neural Network algorithms consist of different layers which analyze data. There are hidden layers which detect patterns in data and the greater the number of layers, the more accurate the outcomes are. Neural networks learn on their own and assign weights to neurons

every time their networks process data. Convolutional Neural Networks and Recurrent Neural Networks are two popular Artificial Neural Network Algorithms.

Applications:

- Handwriting analysis, colorization of black and white images, computer vision processes and describing or captioning photos based on visual features.
- Convolutional Neural Networks: Image feature classification and video processing tasks.
- Recurrent Neural Networks: Language processing tasks and text and speech analysis

K-Nearest Neighbors Algorithm

The K-Nearest Neighbors Algorithm is an algorithm that takes a non-parametric approach to predictive analysis. If we have unstructured data or lack knowledge regarding the distribution data, then the K-Nearest Neighbors Algorithm can be leveraged. The algorithm works by finding similar examples to our unknown example, and using the properties of those neighboring examples to estimate the properties of our unknown examples

Applications:

- Used in industrial applications in tasks such as when a user wants to look for similar items in comparison to others.
- Used in handwriting detection applications.