

Data Science

Classification	The process of predicting the class of given data points. Classes are sometimes called as targets/ labels or categories. Classification predictive modeling is the task of approximating a mapping function (f) from input variables (X) to discrete output variables (y).
ggplot	A data visualization package for the statistical programming language R. ggplot2 can serve as a replacement for the base graphics in R and contains a number of defaults for web and print display of common scales.
Sisense	Data visualization tool, makes data easy to digest by allowing to crunch large data sets and visualize it with pictures, graphs, charts, maps and more.
LingPipe	Tool kit for processing text using computational linguistics. LingPipe is used to do tasks like: find the names of people, organizations or locations in news, automatically classify Twitter search results into categories, suggest correct spellings of queries.
CoreNLP	A Java (or at least JVM-based) annotation pipeline framework, which provides most of the common core natural language processing steps, from tokenization through to coreference resolution.
NLTK	A platform used for building Python programs that work with human language data for applying in statistical natural language processing.
Gensim	An open-source library for unsupervised topic modeling and natural language processing, using modern statistical machine learning.
SPSS	The set of software programs that are combined together in a single package. The basic application of this program is to analyze scientific data related with the social science. This data can be used for market research, surveys, data mining, etc.
Raw Data	Data that has not been processed for use. A distinction is sometimes made between data and information to the effect that information is the end product of data processing.
Gluon	Open source deep learning interface which allows developers to build machine learning models.
MXNet	An open-source deep learning software framework, used to train, and deploy deep neural networks.
Qubole	An open, simple, and secure data lake platform for machine learning, streaming and ad-hoc analytics.
Azure Databricks	A fast, easy and collaborative Apache Spark-based big data analytics service designed for data science and data engineering.
Amazon SageMaker	Amazon SageMaker is a cloud machine-learning platform. SageMaker enables developers to create, train, and deploy machine-learning models in the cloud. SageMaker also enables developers to deploy ML models on embedded systems and edge-devices.
XGBoost	XGBoost is an open-source software library which provides a gradient boosting framework for C++, Java, Python, R, Julia, Perl, and Scala.
PySpark	PySpark is the Python API written in python to support Apache Spark. Apache Spark is a distributed framework that can handle Big Data analysis. Apache Spark is written in Scala and can be integrated with Python, Scala, Java, R, SQL languages.
Apache NiFi	Apache NiFi is an integrated data logistics platform for automating the movement of data between disparate systems. It provides real-time control that makes it easy to manage the movement of data between any source and any destination.
LSTM	Long short-term memory is an artificial recurrent neural network architecture used in the field of deep learning. Unlike standard feedforward neural networks, LSTM has feedback connections. It can not only process single data points (such as images), but also entire sequences of data (such as speech or video).
Pentaho	Pentaho is business intelligence software that provides data integration, OLAP services, reporting, information dashboards, data mining and extract, transform, load capabilities.
Zoomdata	Zoomdata is a business intelligence tool that visualizes data to make exploration and analysis of big, streaming and complex data simple for the everyday user.
Azkaban	Azkaban is a batch workflow job scheduler created at LinkedIn to run Hadoop jobs. Azkaban resolves the ordering through job dependencies and provides an easy to use web user interface to maintain and track your workflows.
SnapLogic	SnapLogic encapsulates integration task complexity with "Snaps" to convert integration tasks and subtasks into modular, pluggable pieces of logic.

Visio is a program that falls under the Microsoft Office Suite of products. It is used for many things that utilize layouts, diagrams, and charts. The

This website uses cookies



We use cookies to continuously improve your experience on our site. [More info.](#)

Got it!

[Airflow](#)

Airflow is a platform to programmatically author, schedule, and monitor workflows.

1 2 3 ... 5

© Copyright 2023 Relocateme. All Rights Reserved

Development by [Synergize.digital](#)