Tel +44 20 4571 2395
info@siliconbeachtraining.co.uk

The Machine Learning with Mahout Certification Training program introduces delegates to the concepts of big data, algorithms, machine learning, Apache Mahout, collaborative filtering and many more advanced topics. The demand of Apache Mahout Specialists is very high as they help their organisations to make intelligent business decisions with their data sets. The robustness and scalability features of Apache Mahout are useful in creating machine learning algorithms. The course teaches how machine learning helps in rapid processing, predictions and analysis. The delegates will also learn how Apache Mahout is used for building and customising machine-learning algorithms. The three basic approaches to machine learning carried out by Mahout are Clusters, Collaborative filters and Categories.

Machine learning can be used in different fields; the entertainment industry, the stock market and even in detecting frauds. Machine learning algorithm is being used globally by the large companies like Amazon and Yahoo in apps. The two main approaches to machine learning are Supervised and Unsupervised learning which are given support by Mahout. The Machine Learning with Mahout Certification Training is conducted by industry expert instructors. After the completion of the Machine Learning with Mahout Certification course, the delegates will be able to implement Mahout at their workplace confidently and efficiently. The course enables the delegates to find solutions to all data-related problems and manage the huge amount of data easily as Mahout processes vast amounts of data at a rapid rate. The delegates will gain an understanding of various machine learning methods like clustering, collaborative filtering, classification and categorisation. The Machine Learning with Mahout Certification training program can empower everyone from developers to engineers.

Prerequisites

The delegates must have basic knowledge of Java programming. An understanding of Hadoop terms and concepts would be beneficial.

Course Objectives

- Learn Machine-learning techniques with Mahout
- Use MapReduce for implementing a Recommender
- Learn the concept of Classifiable Data
- Perform Collaborative Filtering and Pearson Coefficient Algorithm
- Gain insight into Logistic Regression
- Understand the use of Vectorizing and Clustering documents

Introduction to Machine Learning and Apache Mahout

Tel +44 20 4571 2395
info@siliconbeachtraining.co.uk

- Fundamentals of Machine Learning
- Basics of Apache Mahout
- History of Mahout
- Describe Supervised and Unsupervised Learning techniques
- Mahout and Hadoop
- Describe Clustering
- Classification

Mahout and Hadoop

The module teaches delegates how to set up Mahout on Apache Hadoop. Myrrix Machine Learning Platform will also be discussed.

- Mahout and Myrrix
- Mahout on Apache Hadoop setup

Recommendation Engine

The module will give an insight into recommendation system in Mahout. The delegates will also get an understanding of different filtering methods.

- Overview of Recommendation systems
- Recommendations using Mahout
- Content Based
    - Nearest N Users
    - Collaborative filtering
    - Item based
    - User based
    - Threshold
- Mahout Optimizations

Implementing a recommender and recommendation platform

The module introduces delegates to the Recommendation platforms. Also, learn how to use MapReduce for implementing a Recommender.

- Item based Recommendation
- User based recommendation
- User Neighbourhood
- Use MapReduce for implementing a Recommender
- Assessing Recommendation Engines (Online and Offline)
- Platforms
    - Manhattan Distance
    - Cosine Similarity
    - Similarity Measures
    - Tanimoto
    - Euclidean Distance
    - Loglikihood Similarity
    - Pearson's Correlation Similarity
- Recommenders in Production

Clustering

The module provides an overview of 'Clustering' in Mahout and delegates get an understanding

Tel +44 20 4571 2395
info@siliconbeachtraining.co.uk

- Common Clustering Algorithms
- Fuzzy K-means and Mean Shift
- Vectorization
- Canopy Clustering
- K-means
- Feature Selection
- Implementing clustering in Hadoop
- TF-IDF
- Representing Data
- Representing Vectors
- Clustering documents through example
- Classification

Classification

The module will give insight about Classifier and explains common Classifier Algorithms.

- Common Algorithms
- Developing a Classifier
- Basics
- Examples
- SVM
- SGD
- Predictor variables and Target variables
- Random Forests
- Navie Bayes
- Training and evaluating a Classifier

Mahout and Amazon EMR

The module explains the use of Mahout on Amazon EMR Hadoop distribution.

- Explain tools like Weka
- Mahout on Amazon EMR
- Octave, Matlab, SAS
- Mahout Vs R

The Machine Learning with Mahout Certification Training program introduces delegates to the concepts of big data, algorithms, machine learning, Apache Mahout, collaborative filtering and many more advanced topics.