

Data Science Bootcamp

Duration: 5 Days

Description

Data Science is the process of using statistics, mathematics, programming and problem-solving skills to understand, analyze, interpret data, make predictions and make decisions, including discovering patterns and trends in large data sets, extracting meaningful information from data, and predicting future events. aims.

Data science processes large and complex data sets from various fields and extracts meaning from these data. This data can often be in different formats such as unstructured data, text data, images, audio recordings. Data science involves a number of disciplines such as statistics, mathematics, database management, big data technologies and machine learning.

Outline

Python Programming

Introduction to Python

Overview of Python Programming Language Development

Environments and Settings in Python

- Anaconda and Python Development Environment (Spyder, Jupyter Notebook)
- Google Colaboratory (Environment for Applications During Training)

DATA SCIENCE

- Foundations of Data Science
- What is Data Analysis? What Can Be Done with Data Analysis?
- What is Data Science?
- Elements of Data Science
- How Does the Process of Extracting Useful Information from Data Work? (Data Analytics)
 - Descriptive Analytics
 - Diagnostic Analytics
 - Predictive Analytics
 - Prescriptive Analytics
- Applications of Data Science Implementation of Data Science in the Business Process

Cycle

- CRISP-DM Methodology
 - Business Understanding
 - Data Understanding
 - Data Preparation
 - Modeling
 - Evaluation
 - Deployment Basic Stages of Application Development in Data Science Tools Used for Application Development in Data Science
- Numpy
 - Creating Arrays and Matrices
 - Formatting Operations (Reshaping, Concatenation, Splitting)
 - Index Operations
 - Mathematical Operations (Random Number Operations, etc.)
 - Statistical Operations (min, max, mean, std, etc.)
- Pandas
 - Series Operations (Creating Series, Features, etc.)
 - DataFrame Operations (Creation, Features, Element Operations, Merging, Grouping, Filtering, Apply, Pivot Tables)
 - Reading Data from Excel and CSV Files into DataFrame and Operations on Data
- Matplotlib/Seaborn
 - 2D Graphics Usage (line, bar, scatter, histogram, pie, etc.)
 - 3D Graphics Usage
 - Performing Operations on Graphics (Labeling Titles and Axes, Defining Colors, Legends, etc.)

Exploratory Data Analysis (EDA) Data Literacy

- What is Data Literacy?
- Basic Concepts of Data Literacy
 - Population and Sample
 - Unit of Observation
 - What is a Variable? What are the Types of Variables?
 - What is Scale? What are the Types of Scales?
 - Measures of Central Tendency
 - Mean, Median, Mode, Quartiles
 - Measures of Dispersion
 - Range, Standard Deviation, Variance, Skewness, Kurtosis
- Data Definition

- Organizing and Reducing Data
- Data Representation
- Data Analysis and Evaluation
- Loading/Reading Data Sets from Files
- Acquiring Information About the Size of the Data
- Sampling in Data
- Data Types in the Data

Data Preprocessing and Cleaning

- Data Definition
- Attribute Viewing and Selection
- Sorting and Grouping
- Operations on Attributes
 - Adding Attributes
 - Renaming Attribute Names
 - Deriving New Attributes from Existing Attributes
 - Using the re (regular expression) Module in Data Analysis
 - Deleting Attributes
- Operations on Observations
 - Representation of Observations (From the Beginning, From the End, Random)
 - Adding Observations
 - Deleting Observations
- Data Filtering
 - Filtering Using Dictionary and List
 - Filtering with Query
- Missing Data
 - Detecting Missing Data
 - Approaches to Deleting Missing Data
 - Approaches to Completing Missing Data
 - Complete with a Fixed Value
 - Complete with the Mean
 - Complete with Data from the Previous and Next Observation
 - Proportional Operations on Missing Data
- Duplicate Data
 - Detecting Duplicate Data
 - Cleaning Duplicate Data
- Data Transformation Operations
 - Scaling and Normalization of Data
 - Aggregation

- Categorical Data
- Detection of Outliers/Extreme Data

Statistical Operations on Numerical Data

- Distribution
- Variance Analysis
- Correlation Analysis Data Visualization
- Drawing Graphs (line, bar, pie, heat map, etc.)
- Performing Operations on Graphics

Application Studies on Ready Data Sets Tools for Creating Data Analysis Reports on Ready Data Sets

MACHINE LEARNING

Fundamentals of Machine Learning

- What is Machine Learning?
- Real-life Examples
- Basic Concepts and Terminology
 - Types of Problems (Regression, Classification)
 - Model
 - Splitting Data Set into Training and Test Sets
 - Overfitting
 - Model Validation
- Types of Learning
 - Supervised Learning
 - Unsupervised Learning
 - Reinforcement Learning
- Methods for Evaluating the Success of Machine Learning Models Tools for Implementing Machine Learning with Python
- Scikit-Learn Module and Machine Learning

Machine Learning Algorithms/Models and Application Development

- Regression Models (Theory, Model, Prediction)
 - Simple Linear
 - Multiple Linear
- Classification Models (Theory, Model, Prediction)
 - Classification with Logistic Regression
 - Classification with K-Nearest Neighbours (KNN)
 - Classification with Decision Trees (CART)

- Clustering Models (Theory, Model, Prediction) (Application of Unsupervised Learning)
 - Clustering with K-Means Algorithm.