



datacode360
no. 1 data science platform

DATA CODE 360
#1 Data Science Platform

Data Science & Machine Learning with AI

about course

DATA SCIENCE IS A “CONCEPT TO UNIFY STATISTICS, DATA ANALYSIS, MACHINE LEARNING, AND THEIR RELATED METHODS” IN ORDER TO “UNDERSTAND AND ANALYZE ACTUAL PHENOMENA” WITH DATA. IT EMPLOYS TECHNIQUES AND THEORIES DRAWN FROM MANY FIELDS WITHIN THE CONTEXT OF MATHEMATICS, STATISTICS, COMPUTER SCIENCE, AND INFORMATION SCIENCE. DATA SCIENCE IS AN INTERDISCIPLINARY FIELD THAT USES SCIENTIFIC METHODS, PROCESSES, ALGORITHMS, AND SYSTEMS TO EXTRACT KNOWLEDGE AND INSIGHTS FROM MANY STRUCTURAL AND UNSTRUCTURED DATA. DATA SCIENCE IS RELATED TO DATA MINING AND BIG DATA.

learning outcomes

01. LEARN TO IMPLEMENT MACHINE LEARNING TECHNIQUES USING PYTHON.
02. LEARN DATA VISUALIZATION TECHNIQUES.
03. LEARN TO ANALYZE RAW DATA.
04. LEARNING MACHINE LEARNING AND ARTIFICIAL INTELLIGENCE

PYTHON

01. INTRODUCTION TO PYTHON

1. Useful Python Resources
2. Python Tools and Utilities
3. Python Features

02. PYTHON ENVIRONMENT

1. Local Environment Setup
2. Downloads and Installation
3. Setting up Environment Path

03. EXECUTING PYTHON

1. Interactive Mode
2. Scripting Mode
3. Integrated Development Environment

04. PYTHON BASIC SYNTAX

1. Python Identifiers
2. Reserved Words
3. Lines and Indentation

05. PYTHON VARIABLE TYPES

1. Assigning Values to Variables
2. Multiple Assignment
3. Standard Data Types
4. Data Type Conversion

06. PYTHON BASIC OPERATORS

1. Arithmetic Operators
2. Membership Operators
3. Assignment Operators
4. Bitwise Operators

07. PYTHON DESIGN MAKING

1. IF statements
2. IF...ELIF...ELSE statements
3. Nested IF statements

08. PYTHON LOOPS

1. While Loop
2. For Loop
3. Nested Loop
4. Break Control Statement
5. Continue Statement
6. Pass Statement

09. PYTHON NUMBERS

1. Number Type Conversion
2. Mathematical Function
3. Random Number Function
4. Trigonometric Function

10. PYTHON STRINGS

1. String Special Operators
2. String Formatting Operator
3. Built-in String Methods

11. PYTHON LISTS

1. Basic List Operations
2. Indexing and Slicing
3. Built-in functions and methods

12. PYTHON TUPLES

1. Basic Tuple Operations
2. Indexing and Slicing
3. Built-in Functions

13. PYTHON DICTIONARY

1. Basic Dictionary Operations
2. Built-in Functions and Methods
3. Use Cases

14. PYTHON FUNCTION

1. Pass by reference and value
2. Function Arguments
3. Scope of variables
4. Default Argument Values
5. Keyword Arguments
6. Arbitrary Argument Lists
7. Unpacking Argument Lists
8. Lambda Expressions
9. Documentation Strings

15. PYTHON MODULES

1. Importing Modules
2. Namespaces and Scoping
3. Packages

16. PYTHON FILES I/O

1. Writing and Parsing Text Files

17. PYTHON EXCEPTION

1. except clause with multiple exceptions
2. try-finally clause
3. argument of an Exception
4. Raising an Exception
5. User-defined Exceptions

18. PYTHON CLASSES & OBJECTS

1. Creating Classes
2. Creating instance objects
3. Destroying Objects (garbage collection)
4. Custom Classes
5. Attributes and Methods
6. Inheritance & Polymorphism
7. Using Properties to control attribute access

19. COLLECTIONS

1. Deque
2. Counter
3. OrderedDict
4. ChainMap

20. DEBUGGING & TESTING

1. Pdb
2. Breakpoints

21. REGULAR EXPRESSIONS

1. Characters & Character Classes
2. Quantifiers
3. Grouping & Capturing
4. Assertions and Flag
5. The Regular Expression Module

22. ESSENTIAL PACKAGES

1. Operating System
2. System Specific Parameters

23. TESTING

1. Py.test Fundamentals
2. Writing and running test cases

24. DEPLOYING PYTHON APPLICATION

1. Pip
2. Virtualenv
3. The init.py files
4. The setup.py file
5. Installing the package
6. Software Deployment in Python



DATA WRANGLING

25. BLACKBOX INTRODUCTION TO MACHINE LEARNING

1. What is not Machine Learning
2. What is Machine Learning
3. Types of ML
4. Supervised, Unsupervised
5. Classification, Regression
6. Unsupervised Clustering Association
7. Machine Learning Pipeline

26. MICROSOFT EXCEL

1. Introduction - Spreadsheet
2. Functions to Organize Data
3. Filtering, Pivot Tables, Charts
4. Advanced Graphing & Charting

27. SQL

1. SQL Basics
2. SQL Joins
3. SQL Aggregations
4. Queries & Temp Tables
5. SQL Data Cleaning
6. Data Functions

28. TABLEAU

1. Tableau Functions - Bar Charts, Maps, Interactive Dashboards, Scatterplots
2. Time-series Data (Two Methods)
3. Aggregation, Granularity, Level of detail
4. Tableau to various data sets - Excel & CSV
5. Data Blending
6. Data Extracts
7. Advanced Data Preparation

29. ESSENTIAL NUMPY

1. Introduction to NumPy
2. Creation
3. Access
4. Stacking and Splitting
5. Methods
6. Broadcasting

30. PANDAS FOR MACHINE LEARNING

1. Introduction to Pandas
2. Understanding Series & Data Frames
3. Loading CSV, JSON
4. Connecting Databases
5. Descriptive Statistics
6. Accessing subsets of Data - Rows, Columns, Filters

31. UNDERSTANDING VISUALIZATIONS

1. Introduction to matplotlib & seaborn
2. Basic plotting
3. Title, Labels, Legends, Grid, Color Map, xticks, yticks
4. Color, line width
5. Sub Plotting
6. Scatter Plot
7. Histogram
8. Bar Graphs
9. Plotting Distributions
10. Plotting 3D Data Tableau
11. Fundamentals of Tableau

MATHEMATICAL FUNDAMENTALS

32. ESSENTIAL MATHS & STATISTICS

1. Introduction
2. Linear Algebra
3. Matrix Operations
4. Understanding Distributions
5. Probability
6. Concepts
7. Calculus
8. Mean, Mode, Median, Quartile
9. Statistics Concepts
10. Sampling Techniques



MACHINE LEARNING

33. LINEAR MODELS FOR CLASSIFICATION AND REGRESSION

1. Simple Linear Regression using Ordinary Least Squares
2. Gradient Descent Algorithm
3. Regularized Regression Methods - Ridge, Lasso, Elastic Net
4. Logistic Regression for Classification Online
5. Learning Methods - Stochastic Gradient Descent & Passive Aggressive
6. Robust Regression - Dealing with Outliers & Model Errors
7. Polynomial Regression
8. Bias-Variance Tradeoff Application

34. PREPROPOSING OF ML

1. Introduction to Preprocessing
2. StandardScaler
3. MinMaxScaler
4. RobustScaler
5. Normalization
6. Binarization
7. Encoding Categorical (Ordinal & Nominal) Features Imputation
8. Polynomial Features
9. Custom Transformer
10. Text Processing
11. Count Vectorizer & HashingVectorizer
12. Image using skimage

35. DECISION TREES

1. Introduction
2. The Decision Tree Algorithm
3. Decision Tree for Classification
4. Decision Tree for Regression
5. Advantages & Limitations

36. NAIVES BAYES

1. Introduction to Bayes' Theorem
2. Naive Bayes Classifier
3. Gaussian Naive Bayes
4. Multinomial Naive Bayes
5. Bernoulli's Naive Bayes
6. Naive Bayes for out-of-core

37. COMPOSITE ESTIMATORS USING PIPELINE

1. Introduction to Composite Estimators
2. Pipelines Transformed
3. Target Regressor
4. Feature Unions Column Transformer
5. Grid Search on Pipeline

38. MODEL SELECTION & EVALUATION

1. Cross-Validation
2. Hyper Parameter Tuning
3. Model Evaluation
4. Model Persistence
5. Validation Curves
6. Learning Curves

39. FEATURE SELECTION & DIMENSIONALITY REDUCTION

1. Introduction to Feature Selection
2. Variance Threshold
3. Linear Regression tests using $f_{\text{regression}}$
4. F-score vs mutual information
5. Mutual Information for Discrete Value
6. Mutual Information for Continues Value
7. SelectKBest
8. SelectPercentile
9. SelectFormModel
10. Recursive Feature Elimination
11. PCA

40. NEAREST NEIGHBOURS

1. Fundamentals of the Nearest Neighbour Algorithm Unsupervised
2. Neighbours for Classification

41. CLUSTERING TECHNIQUES

1. Introduction - Unsupervised Learning
2. Clustering Similarity
3. Clustering as an Optimization Function
4. Types of Clustering Methods
5. Partitioning Clustering - KMeans & MeanShift
6. Hierarchical Clustering - Agglomerative
7. Density Based Clustering - DBSCAN
8. Measuring Performance of Clusters
9. Comparing All Clustering Methods
10. Application Grouping Similar Customers

42. ANOMALY DETECTION

1. What are Outliers?
2. Statistical Methods for Univariate Data
3. Gaussian Mixture Models
4. Fitting an Elliptic Envelope
5. Isolation Forest
6. Local Outlier Factor
7. Using Clustering Method like DBSCAN

43. SUPPORT VECTOR MACHINES

1. Introduction
2. Maximal Margin Classifier
3. Soft Margin Classifier
4. SVM Algorithm for Classification
5. SVM for Regression

44. DEALING WITH IMBALANCED CLASSES

1. What are imbalanced classes and their impact?
2. OverSampling
3. UnderSampling
4. Connecting Sampler to Pipelines
5. Making Classification Algorithm aware of Imbalance Anomaly

45. ENSEMBLE METHODS

1. Introduction
2. RandomForest
3. AdaBoost Gradient
4. Boosting Tree
5. Voting Classifier - XGBoost
6. Application - Malicious Data Detection

46. RECOMMENDATION ENGINE

1. Understanding Distance Vector Calc. - cosine, euclidean, manhattan
2. Types of Recommendation Engines
3. Recommendation Based on Similarity

47. TIME SERIES MODELING

1. Simple Average and Moving Average
2. Single Exponential Smoothing
3. Holt's Linear Trend Method
4. Holt's Winter Seasonal Method
5. ARIMA

48. PACKAGING AND DEPLOYMENT

1. Creating Python Package
2. Deploy Trained Model Behind REST Interface
3. Deploy Model Behind API Call
4. Deploy on AWS cloud (optional)



ARTIFICIAL INTELLIGENCE

49. NEURAL NETWORK BASICS

1. Gradient Descent
2. Introduction to Perceptron and Neural Networks
3. Batch Normalization
4. Activation and Loss Functions
5. Hyper Parameter Tuning
6. Deep Neural Networks
7. Tensor Flow & Keras for Neural Networks and Deep Learning

50. COMPUTER VISION

1. Introduction to Convolutional Neural Networks
2. Forward Propagation & Back Propagation for CNNs
3. Convolution, Pooling, Padding and its mechanisms
4. CNN Architecture like AlexNet, VGGNet, InceptionNet and ResNet
5. Transfer Learning

51. STATISTICAL NATURAL LANGUAGE PROCESSING

1. Bag of Words Model
2. TF-IDF
3. POS Tagging
4. Named Entity Recognition
5. Tokenization
6. Stop Words
7. World Vectorizer

52. SEQUENTIAL NLP

1. Introduction to Sequential Data
2. RNNs and its mechanisms
3. Vanishing and Exploding Gradients in RNNs
4. Time Series Analysis
5. LSTMs
6. LSTMs with Attention mechanism
7. GRUs - Gated Recurrent Unit

53. ADVANCED COMPUTER VISION

1. Semantic Segmentation
2. YOLO
3. Siamese Networks
4. Object & Face Recognition

54. GENERATIVE ADVERSARIAL NETWORKS

1. Introduction to GANs
2. AutoEncoders
3. How GANs work?
4. Applications of GANs

55. REINFORCEMENT LEARNING

1. Value-Based Methods Q-learning
2. Policy-Based Methods

TOOLS LEARNED DURING COURSE

