

Certificate/ Diploma Course in Artificial Intelligence and Data Science

Duration of the Course:

- 1) Certificate course: 6 months
- 2) Diploma: 12 months

Eligibility: Any student enrolled in the degree program of the college and having knowledge about the basics of Computers.

Sr.No.	Topics	
Topics covered under Certificate Course are Sr No.1, 2 and 3		
1	CCAD-1-01T Problem Solving using Python CCAD-1-01P Problem Solving using Python Lab	Credits: 6(4 Th. 2 Lab)
2	CCAD-1-02T Introduction to AI and Data Science	Credits: 6
3	CCAD-1-03T Statistical Foundation CCAD-1-03P Statistical Foundation Lab	Credits: 6(4Th. 2 Lab)
Topics covered under Diploma Course are Sr No. 1, 2, 3, 4, 5 and 6		
4	DAD-1-04T Data Mining and Visualization DAD-1-04P Data Mining and Visualization Lab	Credits: 6 (4Th. 2 Lab)
5	DAD-1-05T Machine Learning DAD-1-05P Machine Learning Lab	Credits: 6 (4 Th. 2 Lab)
6	DAD-1-06P e- platforms for machine learning	Credits: 6

**Certificate/ Diploma Course in Artificial Intelligence and Data
Science
Semester I
CCAD-1-01T: Problem Solving using Python**

**Total Marks: 100
External Marks: 70
Internal Marks: 30
Credits: 4
Pass Percentage: 40%**

Objective

Objective of this paper is to explain the basic of Python concepts objects, data structures and concepts related to Methods and Functions in python. Paper explicate Object Oriented Programming with Python and comprehend the concepts related to Python Generators and file handling.

Section A

Unit I: Introduction: Python installation and setup, Command line Basics; Python Objects and Data Structures Basics: Introduction to Python data types, Variable assignments, Numbers, String, String methods, Lists

Unit II: Python Comparison Operators: Chaining comparison operators with logical operators, Pass Break and continue.

Unit III: Program Flow control in Python: If Elif and Else statements in python, for loops, While loops

Unit IV: Methods and Functions in python: Introduction to functions, Def keyword, User defined functions, arguments and parameters, Parameter naming in python

Section B

Unit V: Object Oriented Programming: Introduction, Classes and objects, attributes and methods, Inheritance and polymorphism, Special methods; Modules and Packages: Pip install and PyPi.

Unit VI: Errors and Exception Handling: Introduction to errors, Built-in errors, raising errors in python, Pylint overview

Unit VII: Python Generators: Yielding and Generator function, Making an iterable from a generator, Generator expressions and performance.

Unit VIII: File handling in Python: Files in python, importing own files, Read and writing text files, working with CSV, XML and JSON files.

Suggested Readings

1. Timothy Budd, Exploring Python, TMH, 1st Ed, 2011
3. Allen Downey, Jeffrey Elkner, Chris Meyers , How to think like a computer scientist : learning with Python , Green Tea Pr, 2002
4. Paul Barry, Head First Python: A Brain-Friendly Guide, O'Reilly, 2nd ed. 2016
5. Udemy, <https://www.udemy.com/course/complete-python-bootcamp/>
6. Udemy, <https://www.udemy.com/course/python-the-complete-python-developer-course/>

CCAD-1-01P: Problem Solving using Python Lab

Total Marks: 100

External Marks: 70

Internal Marks: 30

Credits: 4

Pass Percentage: 40%

Besides below given practical, other List of Lab Assignments – Session wise will be provided to the students.

- 1) Find the sum of two numbers. (Python basics)
- 2) Find the maximum of three numbers in python (Comparison Operators)
- 3) Print all the prime numbers between two numbers. (Loops)
Define a function to return the number of vowels in a string. (Methods and functions in python)
- 5) Delete an item from a tuple (List, tuple, Type conversions)
- 6) Demonstrate the use of polymorphism by creating different functions for sum. (Polymorphism)
- 7) Suppose you are given two strings. Compare if the two strings are good or not. Strings will be good if they contain same words but the order and case of the words can vary.
For example:
 “Samsung galaxy s10” and “galaxy SAMSUNG s10” are both good strings.
 “Punjab State Open University Patiala” and “Punjab State Open University Chandigarh” are not good strings. (Concept: Strings)
- 8) Create Fibonacci series using Generators (Python generators)
- 9) Write an exception for divisibility of a number by 0. (Exception handling)
- 10) Print the first 10 lines in text file using python (File handling)
- 11) Find the words in a string that are greater than k length, k is the user’s input.
- 12) Find the number of pairs in a given list.
 e.g. [10,20,20,10,2,20,7] has 3
- 13) Take a list of integers. Find the pairs which give the minimum difference
- 14) Take a string as user’s input and print the words from the string that doesn’t contain any letter between 0 to 9 between them.
- 15) Copy content from a file in a computer and paste it into another file.
- 16) Find out all the numbers from a list in python that has sum k, where k is the user’s input.
- 17) Create a linked list using python
- 18) Find out the sums of all the “strictly decreasing” numbers from a list of numbers.
- 19) A strictly decreasing number is the one in which all the numbers from the left to right are decreasing. e.g. 542 and 431 are strictly decreasing but 411 and 424 are not.
- 20) Rearrange words of a string, according to the sum of ASCII values of the letters of the words, the answer should be a single string of the words.

**Certificate/ Diploma Course in Artificial Intelligence and Data
Science
Semester I**

CCAD-1-02T: Introduction to AI and DS

Total Marks: 100

External Marks: 70

Internal Marks: 30

Credits: 6

Pass Percentage: 40%

Objective: This course will enable students to understand the fundamentals of logic. Students will be able to infer the concept Logic and inference.

SECTION A:

Unit I: Introduction to Artificial Intelligence: Definitions of AI, Intelligent Agents, Problem solving. Knowledge, Reasoning and Planning: Logical Agents, Classical Planning, Knowledge Representation and Reasoning. Learning: Learning from examples, Knowledge in learning.

Unit II: Communicating, Perceiving and Acting: Communication, Natural Language Processing, Perception, Computer Vision, Robotics.

Unit III: AI Applications (General): Speech Recognition, Image Recognition, Natural Language Processing, Autonomous Transportation. Natural Language understanding, Recognizing objects and describing images, Dimensionality reduction, feature selection and feature extraction.

Unit IV: AI Applications (Specific): Virtual Personal Assistants/ Chatbots, Gaming, Smart Cars, Drones, Fraud Detection, Software Testing and Development, Business, Health Care, Education, Finance.

SECTION B

Unit I: Introduction to Data Science: Data Science-a discipline, Landscape-Data to Data science, Data Growth-issues and challenges, data science process. Foundations of data science.

Unit II: Data Exploration and Preparation: Structured vs unstructured data, Quantitative vs qualitative data. Four levels of data – nominal, ordinal, interval, ration. Messy data, Anomalies and artifacts in datasets.Cleaning data.

Unit III: Data Representation and Transformation: Forms of data-tabular, text data, graph-based data. Modern databases- text files, spreadsheets, SQL databases, NoSQL databases, distributed databases, live data streams. Representation of data of special types-acoustic, image, sensor and network data.

Unit IV: Computing with Data: Overview of various tools Data Modeling: Basics of Generative modeling and Predictive modeling. Data Visualization and Presentation: Charts-histograms, scatter plots, time series plots etc. Applications of Data Science in Business, Insurance, Energy, Health care, Biotechnology, Manufacturing, Utilities, Telecommunication, Travel, Governance, Gaming, Pharmaceuticals, Geospatial analytics and modeling

Suggested Readings:

1. S.J. Russell and P.Norving: "Artificial Intelligence: A Modern Approach", Pearson.

2. SinanOzdemir, "Principles of Data Science", Packt Publishing.

E.Rich, K.Knight, S.B. Nair: "Artificial Intelligence", Tata McGraw Hill Ed Pvt Ltd.

Joel Grus: "Data Science from Scratch", O'Reilly. 3. Foster Provost & Tom Fawcett: "Data Science for Business" O'Reilly (SPD)

Certificate/ Diploma Course in Artificial Intelligence and Data Science

Semester I

CCAD-1-03T: Statistical Foundation

Total Marks: 100

External Marks: 70

Internal Marks: 30

Credits: 4

Pass Percentage: 40%

Objective: This course will enable students to understand the fundamentals of statistics to apply descriptive measures and probability for data analysis. Students will be able to infer the concept of correlation and regression for relating two or more related variables and probabilities for various events.

Section A

Unit I: Origin and development of Statistics, Scope, limitation and misuse of statistics. Types of data: primary, secondary, quantitative and qualitative data. Types of Measurements: nominal, ordinal, discrete and continuous data.

Unit II: Presentation of data by tables: construction of frequency distributions for discrete and continuous data, graphical representation of a frequency distribution by histogram and frequency polygon, cumulative frequency distributions. Classification and Graphical representation of data (Pie Chart, Bar Diagram, Histogram, Frequency Polygon, Ogive Curve, etc.).

Unit III: Measures of Central Tendency – Arithmetic Mean, Median and Mode and its Graphical representation, Measures of dispersion – range, variance, mean deviation, standard deviation and coeff. of variation, Concepts and Measures of Skewness and Kurtosis.

Unit IV: Descriptive and Exploratory Analysis: Descriptive Statistics, Exploratory data analysis, Coefficient of variation, Data **visualization**, Scatter diagram, Grouped data,

Section B

Unit V: Correlation: Scatter plot, Karl Pearson coefficient of correlation, Spearman's rank correlation coefficient, multiple and partial correlations (for 3 variates only). Regression: Introduction to regression analysis: Modelling a response, overview and applications of regression analysis, Simple linear regression (Two variables)

Unit VI: Mathematical and Statistical probability: Introduction, Elementary events, Sample space, Compound events, Types of events, Random experiment, sample point and sample space, event, algebra of events.

Unit VII: Definition of Probability: Introduction, classical, empirical and axiomatic approaches to probability, properties of probability. Theorems on probability, conditional probability and independent events

Unit VIII: Statistical inference: Introduction, Concept of Random Variable, Probability Mass Function & Density Function, Mathematical Expectation (meaning and properties), Moments, Moment Generating Function and Characteristic Function

Suggested Readings

1. Gupta, S.C. and Kapoor, V.K.: Fundamentals of Mathematical Statistics, Sultan & Chand & Sons, New Delhi, 11th Ed,
2. Hastie, Trevor, et al. The elements of Statistical Learning, Springer,
3. Ross, S.M., Introduction to Probability and Statistics, Academic Foundation,
4. Papoulis, A. and Pillai, S.U. , Probability, Random Variables and Stochastic Processes, TMH

CCAD-1-03P: Statistical Foundation Lab

Total Marks: 100

External Marks: 70

Internal Marks: 30

Credits: 2

Pass Percentage: 40%

Besides below given practicals List of Lab Assignments – Session wise will be given to the student

1. Performing Data exploration and preprocessing in Python
2. Classification and Graphical representation of data
3. Measurement of Central Tendencies
4. Measurement of dispersion
5. Implementing correlation and regression analysis
6. Implementing normal and binominal distribution
7. Implementing conditional Probability
8. Drawing random samples using random number tables.
9. Implementing Density function

**Certificate/ Diploma Course in Artificial Intelligence and Data
Science
Semester II**

DAD-1-04T: Data Mining and Visualization

Total Marks: 100

External Marks: 70

Internal Marks: 30

Credits: 4

Pass Percentage: 40%

Objective

This course helps student understand the need for Data Mining and advantages to the business world. They will get a clear idea of various classes of Data Mining techniques, their need, scenarios (situations), scope of their applicability and use of data visualization techniques

Section-A

Unit -1: Data Mining: Introduction, Scope, What is Data Mining; How does Data Mining Works, Predictive Modeling: Data Mining and Data Warehousing: Architecture for Data Mining: Profitable Applications: Data Mining Tools

Unit -II: Data Pre-processing: Overview, Data Cleaning, Data Integration and Transformation, Data Reduction, Discretization and Concept Hierarchy Generation.

Unit- III: Data Mining Techniques: An Overview, Data Mining Versus Database Management System, Data Mining Techniques- Association rules, Classification, Regression, Clustering, Neural networks.

Unit –IV: Clustering: Introduction, Cluster Analysis, Clustering Methods- K means, Hierarchical clustering, Agglomerative clustering, Divisive clustering, evaluating clusters.

Section-B

Unit –V: Applications of Data Mining: Introduction, Business Applications Using Data Mining- Risk management and targeted marketing, Customer profiles and feature construction, Medical applications (diabetic screening), Scientific Applications using Data Mining, Other Applications.

Unit –VI: Data Visualization: Introduction, Acquiring and Visualizing Data, Simultaneous acquisition and visualization, Applications of Data Visualization, Keys factors of Data Visualization (Control of Presentation, Faster and Better JavaScript processing, Rise of HTML5, Lowering the implementation Bar)

Unit –VII: Exploring the Visual Data Spectrum: charting Primitives (Data Points, Line Charts, Bar Charts, Pie Charts, Area Charts), Exploring advanced Visualizations (Candlestick Charts, Bubble Charts, Surface Charts, Map Charts, Infographics).

Unit –VIII: Visualizing data Programmatically, Starting with Google charts (Google Charts API Basics, A Basic bar chart, A basic Pie chart, Working with Chart Animations).

Suggested Readings

1. Jiawei Han, MichelineKamber and Jian Pei, Data Mining Concepts and Techniques, Third Edition, 2000
2. Pang-Ning Tan, Michael Steinbach, and Vipin Kumar, Introduction to Data Mining, Pearson 2005,
3. M. Kantardzic, “Data Mining: Concepts, Models, Methods, and Algorithms”, 2nd edition, Wiley-IEEE Press, 2011
4. Jon Raasch, Graham Murray, VadimOgievetsky, Joseph Lowery, JavaScript and jQuery for Data Analysis and Visualization, 2014
5. Ben Fry, "Visualizing data: Exploring and explaining data with the processing environment", O'Reilly, 2007

**Certificate/ Diploma Course in Artificial Intelligence and Data
Science
Semester II**

DAD-1-04P: Data Mining and Visualization Lab

Total Marks: 100

External Marks: 70

Internal Marks: 30

Credits: 2

Pass Percentage: 40%

Besides below given practical, List of Lab Assignments – Session wise will be given to student

1. Explore WEKA Data Mining/Machine Learning Toolkit

- Downloading and/or installation of WEKA data mining toolkit,
- Understand the features of WEKA toolkit such as Explorer, Knowledge Flow interface,
- Experimenter, command-line interface.
- Navigate the options available in the WEKA
- (ex. Select attributes panel, Preprocess panel, Classify panel, Cluster panel, Associate panel
- and Visualize panel)
- Study the arff file format
- Explore the available data sets in WEKA.
- Load a data set (ex. Weather dataset, Iris dataset, etc.)
- Load each dataset and observe the following:
 1. List the attribute names and they types
 2. Number of records in each dataset
 3. Identify the class attribute (if any)
 4. Plot Histogram
 5. Determine the number of records for each class.
 6. Visualize the data in various dimensions

2. Perform data pre-processing tasks and Demonstrate performing association rule mining on data sets.

3. Demonstrate performing classification on data sets

- Load each dataset into Weka and run Id3, J48 classification algorithm. Study the classifier output. Compute entropy values, Kappa statistic.

- Extract if-then rules from the decision tree generated by the classifier, Observe the confusion matrix.
- Load each dataset into Weka and perform Naïve-bayes classification and k-Nearest Neighbour classification. Interpret the results obtained.
- Plot RoC Curves
- Compare classification results of ID3, J48, Naïve-Bayes and k-NN classifiers for each dataset, and deduce which classifier is performing best and poor for each dataset and justify.

4. Demonstrate performing clustering of data sets

- Load each dataset into Weka and run simple k-means clustering algorithm with different values of k (number of desired clusters). Study the clusters formed. Observe the sum of squared errors and centroids, and derive insights.
- Explore other clustering techniques available in Weka.

Explore visualization features of Weka to visualize the clusters. Derive interesting insights and explain.

**Certificate/ Diploma Course in Artificial Intelligence and Data
Science
Semester II
DAD-1-05T: Machine Learning**

**Total Marks: 100
External Marks: 70
Internal Marks: 30
Credits: 4
Pass Percentage: 40%**

Objectives

This course aims to provide sound foundation to fundamental concepts of machine learning and its application and prepare students for advanced research and real time problem solving in machine learning and related fields. This course will help students to understand the concepts related to regression, classification, clustering, concept clarity on deep learning, dimensionality reduction, model selection and boosting.

Section A

Unit I: Introduction to ML: Applications of Machine learning, machine learning as a future; Data Pre-processing: Importing the libraries, Importing the dataset, taking care of missing data, encoding categorical data, Splitting the dataset into training set and test set, Feature scaling.

Unit II: Regression: Simple linear regression, Multiple linear regression, Understanding the P-value, Polynomial regression

Unit III: Classification: Logistic Regression, K-Nearest Neighbors, Support vector machine, Naïve Bayes, Decision tree classification, Random forest classification.

Unit IV: Clustering: k-means clustering, k means random initialization trap, selecting the number of clusters, Hierarchical clustering.

Section B

Unit V: ANN: Introduction to ANNs, Biological Neural Networks; Usefulness and Applications of ANNs; Architectures of ANNs: Single layer, Multi layer, Competitive layer; Learning: Supervised and Unsupervised; Activation functions; Linear and Non-linear Separability

Unit VI: Supervised Models: Hebb Net: introduction, algorithm, application for AND problem; **Perceptron:** architecture, algorithm, application for OR Problem; ADALINE: architecture, algorithm, application for XOR problem;

Unit VII: MADALINE: architecture, algorithm, application for XOR problem; Back propagation Neural Network: architecture, parameters, algorithm, applications, different issues regarding convergence

Unit VIII Unsupervised Models: Kohonen Self –Organizing Maps: architecture, algorithm, application, Adaptive Resonance Theory: introduction, basic architecture, basic operation, ART1 and ART2

Suggested Readings

1. Andreas C. Müller, Introduction to Machine Learning with Python: A Guide for Data Scientists, 2016 ,Sarah Guido
2. E. Alpaydin, Introduction to Machine Learning, 3rd Edition, PHI Learning, 2015
3. K. P. Murphy, Machine Learning:A Probabilistic Perspective, MIT Press, 2012
4. <https://www.udemy.com/course/machinelearning/>

**Certificate/ Diploma Course in Artificial Intelligence and Data
Science
Semester II
DAD-1-05P: Machine Learning Lab**

Total Marks: 100

External Marks: 70

Internal Marks: 30

Credits: 2

Pass Percentage: 40%

Besides below given assignment List of Lab Assignments – Session wise will be given to student

- 1) Write a program to demonstrate the working of decision tree based ID3 algorithm. Use an appropriate data set for building the decision tree and apply this knowledge to classify a new sample.
- 2) Build an Artificial Neural Network by implementing the back propagation algorithm and test the same using appropriate data set.
- 3) Write a program to implement naïve Bayesian classifier for a sample training data set stored as a .CSV file. Compute the accuracy of the classifier, considering few test data sets.
- 4) Write a program to construct a Bayesian network considering any medical related dataset at UCI.
- 5) Apply EM algorithm to cluster a set of data stored in a .CSV file. Use the same data set for clustering using k-means algorithm. Compare the results of these two algorithms and comment the quality of clustering.
- 6) Write a program (using Python ML library classes) to implement k-Nearest Neighbor algorithm to classify the iris data set. Print both the correct and wrong predictions.
- 7) Implement the non-parametric Locally weighted regression algorithm in order to fit data points. Select the appropriate data set for your experiment and draw graphs

**Certificate/ Diploma Course in Artificial Intelligence and Data
Science
Project
Semester II**

DAD-1-06P: e-Platforms for the Machine Learning

Total Marks: 100

External Marks: 70

Internal Marks: 30

Credits: 6

Pass Percentage: 40%

Objective:

To make student familiar with online platform for executing machine learning models. Students will improve their skill in machine learning and data science

Instruction

Following are the famous online platforms for executing the machine learning models. This has been recommended to the learner to use at-least one platform and run the assignments given by the university using Python and Deep learning. External Examination will be conducted through quiz/viva-vice or any other relevant method.

Google Colab

Colab is a free Jupyter notebook environment that runs entirely in the cloud. Most importantly, it does not require a setup and the notebooks that you create can be simultaneously edited by your team members - just the way you edit documents in Google Docs. Colab supports many popular machine learning libraries which can be easily loaded in your notebook.

Kaggle

Kaggle, a subsidiary of Google LLC, is an online community of data scientists and machine learning practitioners. Kaggle allows users to find and publish data sets, explore and build models in a web-based data-science environment, work with other data scientists and machine