GITAA
AI TRAINING & AI POWERED EDTECH SOLUTIONS

AI/ML using Python

gitaa.in

Agenda

- Course Objectives
- Overview of the Course (Weekly Wise)
- Briefing about the final exam
- Q & A

Course objective

- Understand foundational concepts in Data Science and Machine Learning.
- Gain hands-on experience with Python programming and development environments like Spyder
- Learn to work with data structures such as variables, data types, and sequence data types.
- Develop proficiency in using Python libraries such as NumPy and Pandas for data manipulation.
- Analyze data using descriptive statistics to derive meaningful insights.
- Build a base for advanced machine learning and data analytics applications.

Week-1

GITAA
AI TRAINING & AI POWERED EDTECH SOLUTIONS

- Introduction to Data science
- Why Python for Data science?
- Introduction to Spyder

Week-2

GITAA
AI TRAINING & AI POWERED EDTECH SOLUTIONS

Sequence data types and associated operations

- Strings
- Lists
- Arrays
- Tuples
- Dictionary
- Set
- Range

Week-3

GITAA
AI TRAINING & AI POWERED EDTECH SOLUTIONS

- Descriptive statistics
    - Measures of central tendency
    - Measures of spread
    - Distribution of mean and variance
    - Sampling basics
- Pandas dataframe and dataframes related operations
    - Reading files
    - Comma separated value files.
    - Tab-delimited files
    - Excel files

Week-4

- Data visualization using matplotlib and seaborn libraries
- Scatter Plot
- Bar Plot
- Histogram
- Box plot
- Pair plot

Week-5

- Purposes of Statistics: Introduction to descriptive, inferential, and predictive statistics, and their practical roles.
- Sampling and Sampling Distribution: Concepts of population vs. sample, sampling distribution of the mean, and variance.
- Central Limit Theorem (CLT): Explanation of how sample means approximate a normal distribution as sample size increases.
- Confidence Intervals: Construction and interpretation of confidence intervals for population parameters.
- Hypothesis Testing: Steps involved including null and alternate hypotheses, significance level (α), p-values, and decision-making.
- Types of Statistical Tests: Overview of z-tests, t-tests, F-tests, and tests for proportions, including one-sample, two-sample (independent and paired) scenarios.

Week-6-Linear Algebra

- Representation of data using matrices (rows as samples, columns as attributes)
- Matrix Properties and Operations (Rank of Matrix,Null space and nullity,Basic vectors)
- Vectors, distances, and unit vectors
- Eigenvalues and Eigenvectors
- Orthogonal and orthonormal vectors

Week-7

- Role of optimization in machine learning and data science
- Components of an optimization problem (objective function, variables, constraints)
- Directional search: gradient descent and step size selection
- Learning rule in machine learning (e.g., backpropagation in neural networks)

Week-8 & 9

- Introduction to Machine Learning
- Types of ML
- Simple Linear Regression
- Multiple Linear Regression

Week-10

- Logistics regression
- KNN
- K Means clustering

Week-11 & 12

- PCA
- SVD
- Neural Network

Final Exam

- The final exam consist of multiple-choice questions (MCQs), covering a wide range of topics in course

- These will include questions about statistical methods, machine learning, data manipulation, and data visualization

Final Assessment Details

- Exam Date: Feb 8,15,22 And Mar 1
- Time : 10 AM
- Assessment Type: Multiple Choice Questions
- Number of Questions: 50
- Number of Attempts Allowed: 1
- Duration: 1 hour 30 minutes

1.Basic Python

Key Areas:

- Basic Data types
- Variables
- Control structures
- Basic Functions in Python

2.Data Cleaning and Manipulation with Pandas

Key Areas:

- Handling Missing Data: Understand how to handle missing values using imputation methods like replacing with mean, median, or mode.
- Data Wrangling: Know how to filter, merge, and transform data using Pandas.
- Dealing with Duplicates: Know how to identify and handle duplicate rows in a dataset.

  Preparation Tips:

- Review Pandas functions for handling missing data (e.g., fillna(), dropna()).
- Practice cleaning cricket datasets (like the GSW dataset) by removing duplicates, handling missing values, and preparing data for analysis.

3.Visualization and Interpretation

Key Areas:

- Data Visualization: Know how to use visualization tools (like Tableau or Python libraries) to plot data and interpret trends.
- Box Plots, Histograms, and Pie Charts: Understand how to create and interpret these visualizations to analyze data distributions
  Preparation Tips:
- Familiarize yourself with different chart types like line plots, box plots, and histograms.
- Practice analyzing the data distributions and drawing insights from these visualizations.

4.Machine learning

Key Areas:

- Feature Extraction and Model Training: Understand how to use machine learning algorithms for predicting match outcomes.
- Model Evaluation: Evaluate classification models using metrics such as accuracy, confusion matrices, and precision/recall.

Preparation Tips:

- Review the key concepts of classification vs. regression in machine learning
- Understand how to interpret model performance metrics and evaluate classification results (e.g., using a confusion matrix)

5.Statistical Analysis and Interpretation

- Key Areas:
    a. Correlation: Understand how to calculate and interpret correlation between two variables (e.g., runs scored and batting average).
    b. Outliers and Anomalies: Know how to identify outliers in the data and understand their impact on analysis
- Preparation Tips:
    c. Practice correlation analysis and interpretation of the correlation coefficient
    d. Familiarize yourself with box plots and other methods to detect outliers in data

5.Regression and Predictive Modeling

Key Areas:

- Simple & Multiple Linear Regression: Be familiar with regression analysis to predict the output
- Model Interpretation: Interpret the results of regression models and understand how to evaluate them

- Machine Learning Concepts: Understand the types of machine learning problems like regression vs. classification

  Preparation Tips:

- Practice solving problems involving regression equations.

- Understand the differences between supervised learning (e.g., regression) and unsupervised learning.

- Review how to use regression results to make predictions of the outcome