

Certificate Course in Data Science

Duration: 60 hours (learning) + 10 hours (project) + 57 hours (self-study)

Pre-Requisites: Knowledge of mathematics and basic statistics with sound knowledge of computer

Qualification: Any graduate with at least second class OR CA intermediate pass out

Fees: 6000/-

Institutes Name: Department of Computer Science, Gujarat Vidyapith,

Language: English + Hindi

Strength: 15

Starting Date: First week of February, 2021

Schedule: Saturday – Sunday

Admission details will be provided soon on the website or contact to the department of Computer Science.

Course goal

The main goal of the certificate course is to help students learn about data science that includes application of various methods on different kind of data like day to day data, scientific data, industrial data or big data; with application of techniques like formatting data, cleaning up data, summarizing data, analyzing data and plotting data with relevant methods.

Module 1

Introduction to Data Science

- Overview
- Data Science and Analytics
- Types of Business Analytics
- Applications of Data Science
- Tools of Data Science

Learning outcomes:

Ability to identify characteristics of datasets and according to the types of data select best analysis methods. Apart from that selection of best applications for the given data using tool (here R)

Module 2

Introduction to R Programming

- Overview
- Importance of R
- Data Types and Variables in R
- Operators in R
- Conditional Statements in R
- Loops in R
- R script
- Functions in R

Learning outcomes:

Ability to work with different datasets using R and managing data and apply control statements and control loop to the datasets.

Module 3

Data Structures

- Overview
- Identifying Data Structures
- Hands-on for Data Structures Identification
- Assigning Values to Data Structures
- Data Manipulation
- Hands-on for assigning values and applying functions

Learning outcomes:

Ability to identify data structures of the datasets and assigning values to the array and list data structures. It also helps to apply different data manipulation techniques

Module 4

Statistics

- Important statistical concepts used in data science
- Difference between population and sample
- Types of variables
- Measures of central tendency
- Measures of Dispersion
- Measures of variability
- Coefficient of variance
- Skewness and Kurtosis

Learning outcomes:

Ability to apply various statistical concepts like central tendency of data, its variability, variance, mean, median, skewness and kurtosis.

Module 5

Inferential Statistics

- Normal distribution
- Types of Hypothesis
- Test hypotheses
- Parametric test
- Non-parametric test
- Hypothesis Tests about Population Means
- Hypothesis Tests about Population Variance
- Hypothesis Tests about Population Proportions
- Data Sampling
- Sampling distribution
- Spread of sampling distribution
- Central limit theorem
- Sampling errors
- Confidence interval
- T-test, chi-square distribution
- Type I and II errors

Learning outcomes:

Ability to understand different distribution techniques, sampling techniques, selection of hypothesis and finding errors of the data.

Module 6

Regression

- Overview
- Introduction to Regression Analysis
- Types of Regression Analysis Models
- Linear Regression
- Non-Linear Regression
- Cross Validation
- Logistic regression

Learning outcomes:

Ability to integrate data various library of R and apply statistical concepts of regression analysis on various models

Detail Schedule

| Sr. No | Topics | Total hours | In class hours | | Self-study |
|--------|--|-------------|----------------|------------|------------|
| | | | Class room | Laboratory | |
| 1. | Introduction to data science and R | 6 | 1 | 2 | 3 |
| 2. | Overview of R | 6 | 1 | 2 | 3 |
| 3. | Control loop and statements | 6 | 1 | 2 | 3 |
| 4. | R scripts, functions and plot | 6 | 1 | 2 | 3 |
| 5. | Data structure | 6 | 1 | 2 | 3 |
| 6. | Data manipulation | 6 | 1 | 2 | 3 |
| 7. | Basic Statistical concepts | 5 | 1 | 2 | 2 |
| 8. | Measures of central tendency | 5 | 1 | 2 | 3 |
| 9. | Measures of dispersion, coefficient, skewness | 6 | 1 | 2 | 3 |
| 10. | Probability | 5 | 1 | 2 | 2 |
| 11. | Distribution and types of distribution | 6 | 1 | 2 | 3 |
| 12. | Normal Distribution | 5 | 1 | 2 | 2 |
| 13. | Data sampling, sampling distribution and sample error | 6 | 1 | 2 | 3 |
| 14. | Hypothesis and test of hypothesis | 6 | 1 | 2 | 3 |
| 15. | Non central distributions with chi- square and t- test | 6 | 1 | 2 | 3 |
| 16. | Central limit theorem and confidence interval | 6 | 1 | 2 | 3 |
| 17. | T-test and types I and II errors | 6 | 1 | 2 | 3 |

| | | | | | |
|-----|---|-----|----|----|----|
| 18. | Introduction to Regression analysis and types of regression | 6 | 1 | 2 | 3 |
| 19. | Linear regression | 6 | 1 | 2 | 3 |
| 20. | Logistic regression | 6 | 1 | 2 | 3 |
| 21. | Project on Data Science | 10 | | | 10 |
| | Total | 127 | 20 | 40 | 67 |