

**CETPA INFOTECH PVT. LTD.**  
**CURRICULUM OF DATA SCIENCE PROGRAM USING JULIA**

➤ **Introduction to Data Science**

- What is Data and it's use in present world ?
- What are the sources of Data ?
- Exponential Growth in Data.
- Different types of Data.
- Structured, Semi-structured and Unstructured Data.
- What is Data Science?
- Need of Data Science in present world.
- What is Data Mining?
- What is Data Analytics?

➤ **Processes of Data Analytics**

- Data collection from different sources.
- What is data sampling?
- Data Processing and preprocessing.
- Data processing based on different kinds of data.
- Data Cleaning before processing.
- Handling the missing values in Data.
- Different algorithm implementation on data
- Different tools available for data analytics
- Introduction to and hands on to GitHub.

➤ **Basic Concepts in Julia programming language**

- Introduction to Julia language
- History of Julia language
- Why to Learn Julia Programming Language?
- Datatypes in Julia language
- Operators in Julia language
- Introduction to loops
- Introduction to Conditions
- Mathematical Operators in Julia
  - Arithmetic Operators
  - Bitwise operators
  - Numeric Comparisons
  - Operator Precedence
  - Numeric Conversions
- Strings in Julia
  - String basics
- Trigonometric Plots in Julia
- Customizing the graphs with Julia

- Whiskers graph in Julia
- Line Graphs in Julia
- Pie Chart using Julia
- Funnel Chart with Julia
- Relative Frequency Histogram
- Segmented Bar Graph

➤ **Different Packages in Julia**

- Introduction to basic plot package 'Plots'
- Introduction to Gadfly
- Working with IJulia package
- Working in notebook with Julia
- Working with Clustering Package
- Working with Plotly package
- Working with Winston Package
- Introduction to Dataframe package
- Working with dataframes
- Introduction to GLM Package
- Implementing the Linux commands in Julia
- Implementing python and R packages in Julia
- Using Calculus package in Julia

➤ **Statistics used in Data Analytics**

- What is statistics?
- Usefulness of statistics in Data Science
- Understanding the use of mean, median and mode in real dataset
- Interpretation of variance and standard deviation on dataset
- Understanding the meaning of co-variance
- Understanding and Implementation of Binomial Distribution
- Implementation of Linear Regression on dataset
- Implementation of Multiple Regression on dataset
- Implementation of Logistic Regression
- Understanding and Implementation of Poisson Distribution
- Understanding and Implementing Time-Series
- Conducting the Chi- Square Test
- Meaning of R square

- Meaning of Z score
- Survival analysis using Julia
- **Machine Learning with Julia**
  - What is machine learning?
  - Supervised Learning
  - Unsupervised Learning
  - Reinforced Learning
  - K-Means Clustering
  - KNN Algorithm
  - Naïve Bayes Algorithm
  - Support Vector Machine
  - Principle Component Analysis
  - Implementing Decision Tree
- **Introduction to Big Data**
  - What is Big Data ?
  - Sources of Big Data in present world.
  - Features of Big Data.
  - Importance of Big Data in real world.
  - What are the problems of Big Data?
  - Why conventional methods can't process the Big Data?
  - Tools available to process the Big Data
  - Data Handling using conventional tools(oracle, mysql, postgre)
  - Hands on to the core Java

- **Introduction to Hadoop**
  - What is Hadoop?
  - History of Hadoop
  - Why Hadoop was needed to handle bigdata?
  - Architecture of Hadoop
  - Components of Hadoop and their working
  - Algorithm behind the working of Hadoop
  - Hadoop Distributed File System
  - Map Reduce Algorithm in Hadoop and it's dissection
  - Map Reduce Programming hands on in Hadoop
- **A mini Project on Data Science using learnt tools**