Jansons Institute of Technology

Karumathampatti, Coimbatore – 641 659
Approved by AICTE and Affiliated to Anna University
An ISO 9001:2015 certified institution

Industry Readiness Course - Syllabus

Department of Computer Science and Engineering

Data Analysis & Visualization in R

2021- 2022 ODDSemester

Learning Objectives

The objective of this module to make students exercise the fundamentals of statistical analysis in R environment.

They would be able to analysis data for the purpose of exploration using Descriptive and Inferential Statistics.

Students will understand Probability and Sampling Distributions and learn the creative application of Linear Regression in multivariate context for predictive purpose.

Learning Outcomes

After the successful completion of this module, students will be able to:

• Install, Code and Use R Programming Language in R Studio IDE to perform basic tasks on Vectors, Matrices and Data frames.

• Describe key terminologies, concepts and techniques employed in Data Analysis.

• Define, Calculate, Implement Probability and Probability Distributions to solve a wide variety of problems.

• Conduct and Interpret a variety of Hypothesis Tests to aid Decision Making.

• Understand, Analyse, Interpret Correlation and Regression to analyse the underlying relationships between different variables

Module 1:

Overview of Data Analytics and R Programming

Introduction to Data Analytics - Data Analytics Model - Data Analytics Methods – Classification – Regression – Decision Trees - Clustering - Frequent Patterns - Introduction to R and RStudio - R tools and their uses in Business Analytics - Data Frame and write file.

Module 2:

Data Manipulation in R

Data cleaning - Data inspection - Use of functions grep() - grepl() - sub() - apply() –Data import techniques - csv files, spreadsheets and text files -Data Wrangling - Understanding the Exploratory Data Analysis(EDA) - Implementation of EDA on various datasets.

Module 3:

Data Visualization in R

Understanding on Data visualization: Basic principles - categorical and continuous variables -Graphical functions present in R - Plot various graphs -tableplot – Histogram - Box Plot - Customizing Graphical Parameters to improvise the plots - Exploratory graphical analysis - Creating static graphs, animated visualizations.

Books and references

1. Hastie, T., Tibshirani, R.,, Friedman, J. (2009). The elements of statistical learning: data mining, inference and prediction. Springer. Textbook.
2. Richard O. Duda, Peter E. Hart, and David G. Stork. 2000. Pattern Classification (2nd Edition). Wiley-Interscience, New York, NY, USA.

PRINCIPAL
JANSONS INSTITUTE OF TECHNOLOGY
KARUMATHAMPATTI
COIMBATORE - 641 659.

3. Christopher M. Bishop. 2006. Pattern Recognition and Machine Learning (Information Science and Statistics). Springer-Verlag, Berlin, Heidelberg.
4. Teetor, P. (2011). R cookbook. Sebastopol, CA: O'Reilly. ISBN 9780596809157.
5. Chang, W. (2013). R graphics cookbook. Sebastopol, CA: O'Reilly. ISBN 9781449316952.
6. Andy Field, Jeremy Miles and Zoe Field. (2012) Discovering Statistics Using R. Publisher: SAGE Publications Ltd. ISBN-13: 978-1446200469.
7. Gareth James, Daniela Witten, Trevor Hastie and Robert Tibshirani. (2013) An Introduction to Statistical Learning with Applications in R. Springer.
8. https://www.openintro.org/stat/ Free PDF for download & R tutorials and codes.

Course Designed By

Dr A Velayudham Professor / CSE Jansons Institute of Technology	Ms M Pavithra AP / CSE Jansons Institute of Technology	Dr A Dinesh Kumar Professor, CSE KL University

Approved By	**Approved By**	**Approved By**
Dr E S Shamila, Prof / CSE Industry Readiness Coordinator	Dr A Velayudham, Prof & Head /CSE Head of Department	Principal

Dr. A. VELAYUDHAM, M.E., Ph.D.,
Professor & Head,
Dept. of Computer Science and Engineering,
Jansons Institute of Technology,
Karumathampatti, Coimbatore - 641 659.

PRINCIPAL
JANSONS INSTITUTE OF TECHNOLOGY
KARUMATHAMPATTI
COIMBATORE - 641 659.

JANSONS INSTITUTE OF TECHNOLOGY ★ JIT