



THE NATIONAL COLLEGE
BASAVANGUDI, BANGALORE-560004
[AUTONOMOUS]
DEPARTMENT OF MATHEMATICS

Analytical Insights: Unveiling Patterns in Data

Duration: 30 Hours

Course code: MAT-SDA

Course Description:

The Analytical Insights course delves into the methods and techniques used to uncover meaningful patterns, trends, and relationships within datasets. Through a blend of theoretical concepts, practical exercises, and real-world applications, students will gain proficiency in statistical data analysis and visualization. Topics covered include exploratory data analysis, hypothesis testing, regression analysis, and predictive modelling. Emphasis will be placed on understanding the underlying principles of data analysis and effectively communicating insights derived from data.

Course Objectives:

1. Understand the fundamentals of statistical data analysis.
2. Gain proficiency in exploratory data analysis techniques.
3. Learn hypothesis testing methodologies for making data-driven decisions.
4. Develop skills in regression analysis and predictive modelling.
5. Master data visualization techniques to communicate insights effectively.
6. Apply statistical data analysis methods to real-world datasets and problems.

Module 1:

7 Hours

Introduction to Statistical Data Analysis

- Overview of statistical analysis process
- Importance and applications of data analysis
- Ethical considerations in data analysis

Exploratory Data Analysis (EDA)

- Descriptive statistics: measures of central tendency and dispersion
- Data visualization techniques (e.g., histograms, box plots, scatter plots)
- Data cleaning and pre-processing
-

Module 2.

7 Hours

Statistical Inference

- Sampling distributions and Central Limit Theorem
- Confidence intervals and hypothesis testing
- Parametric vs. non-parametric tests

Hypothesis Testing

- Formulating null and alternative hypotheses
- One-sample and two-sample hypothesis tests
- Chi-square tests for categorical data

Module 3.

8 Hours

Regression Analysis



<ul style="list-style-type: none"> • Simple linear regression • Multiple linear regression • Model evaluation and interpretation <p>Predictive Modelling</p> <ul style="list-style-type: none"> • Introduction to machine learning algorithms • Supervised learning vs. unsupervised learning • Model training, validation, and testing 	
Module 4:	8 Hours
<p>Classification Techniques</p> <ul style="list-style-type: none"> • Logistic regression for binary classification • Decision trees and random forests • Support Vector Machines (SVM) <p>Clustering and Dimensionality Reduction</p> <ul style="list-style-type: none"> • K-means clustering • Hierarchical clustering • Principal Component Analysis (PCA) for dimensionality reduction <p>Data Visualization</p> <ul style="list-style-type: none"> • Principles of effective data visualization • Tools and libraries for data visualization (e.g., Matplotlib, Seaborn, Tableau) • Designing informative and aesthetically pleasing visualizations 	
<p>Recommended Textbooks:</p> <ul style="list-style-type: none"> • "Introduction to Statistical Learning" by Gareth James, Daniela Witten, Trevor Hastie, and Robert Tibshirani • "Data Science for Business" by Foster Provost and Tom Fawcett • "Practical Statistics for Data Scientists" by Andrew Bruce and Peter Bruce 	

