

BTY587:DATA ANALYSIS AND SIMULATIONS

L:2 T:0 P:2 Credits:3

Course Outcomes: Through this course students should be able to

CO1 :: remember and classify the techniques for handling data in datasets

CO2 :: understand different statistical tests for data analysis and comparison

CO3 :: articulate association and correlation and classification techniques to identify frequent patterns in datasets

CO4 :: prioritize machine learning like supervised unsupervised and logistic regression to learn from data find patterns and predict outcomes

CO5 :: build machine learning models to classify data discover groups predict outcomes and evaluate performance.

CO6 :: summarize the effectiveness of artificial neural networks from other machine learning models in solving complex problems

Unit I

Data preprocessing and visualization : types of data, dealing with missing data, scatter plot, histogram, group plots, box plots, dimensionality reduction

Unit II

Data analysis : statistical analysis, hypothesis testing, significance of p-value, chi-square, t-test, ANOVA, bayesian probability

Unit III

Mining frequent patterns : associations and correlations, classification, Introduction to Association Rule Mining, Mining Correlation Patterns, Introduction to Classification, Frequent Pattern-Based Classification Methods, Pattern Evaluation in Classification (Evaluation Metrics: Precision, Recall, F1-Score)

Unit IV

Machine learning-1 : supervised, unsupervised, logistic regression

Unit V

Machine learning-2 : Support Vector Machine, decision trees, clustering, model evaluation

Unit VI

Artificial neural networks : types of ANN, case studies for the application of deep learning in biology and health care research, introduction of deep learning algorithms

List of Practicals / Experiments:

Practicals

- basic data entering and data formatting with MSExcel
- plotting histograms, group plots and pie charts with MSExcel
- plotting line charts and scatter plots with MSExcel and PRISM and its analysis
- to perform chi-square test with MSExcel and Graphpad PRISM
- to perform t-test and ANOVA with MSExcel and Graphpad PRISM
- implementation of regression based supervised machine learning algorithms
- implementation of classification based supervised machine learning algorithms
- implementation of clustering based unsupervised machine learning algorithms
- implementation of associations and correlations for data mining
- implementation of ANN for optimization of process

Text Books: 1. INTRODUCTION TO MACHINE LEARNING USING PYTHON by JEEVA JOSE, KHANNA PUBLISHERS

References:

References:

1. DEEP LEARNING by RAJIV CHOPRA, KHANNA PUBLISHERS