

INT234:PREDICTIVE ANALYTICS

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

Course Outcomes: Through this course students should be able to

CO1 :: explain the basics of data preprocessing and its implementation by using R programming Language.

CO2 :: discuss the basics of classification by using Supervised Learning Algorithms

CO3 :: use of different Supervised learning techniques to predict numeric values

CO4 :: demonstrate the predictive models by using Neural networks and Support vector machines

CO5 :: categorize the data by implementing unsupervised learning algorithms

CO6 :: illustrate the techniques to evaluate the model performance and various methods to improve it

Unit I

DATA PREPROCESSING : Managing data with R, Exploring and understanding data, Exploring the structure of data, Exploring numeric variables, Exploring categorical variables, Exploring relationships between variables

Unit II

SUPERVISED LEARNING: CLASSIFICATION : Lazy learning:Nearest neighbors, Probabilistic Learning: Using Naive Bayes, Divide and Conquer: Decision Trees and Rules

Unit III

SUPERVISED LEARNING : NUMERIC PREDICTION : Forecasting Numeric Data, Simple Linear Regression, Polynomial Regression, Ordinary least squares estimation, Correlations

Unit IV

SUPERVISED LEARNING:DUAL USE : Black Box Methods, Neural Networks, Support Vector Machines

Unit V

UNSUPERVISED LEARNING: CLUSTERING AND PATTERN DETECTION : K-Means Clustering, K-means clustering intuition, K-means random initialization trap, K-means selecting number of clusters, Dataset gathering, Hierarchical Clustering, Association Rules, Finding Patterns, Market Basket Analysis Using Association Rules

Unit VI

MODEL PERFORMANCE : Evaluation Model Performance, Improving Model Performance, Bagging, Boosting, Random forests

List of Practicals / Experiments:

Practical 1: Managing Data with R

- Exploring and understanding the data and loading it into different data structures.

Practical 2: Basics Of Data Preprocessing

- Exploring numeric and categorical variables, and finding relationships between different variables.

Practical 3: Implementation of Lazy and Probabilistic learning algorithms.

- Classification based on Nearest Neighbor and Naïve Bayes.

Practical 4: Implementation of Divide and Conquer Algorithms.

- Classification using Decision Tree and Rules.

Practical 5: Implementation of Regression Algorithms.

- Forecasting using Simple Linear Regression, Polynomial Regression and Multiple Linear Regression Algorithms.

Practical 6: Defining Relationship between Numeric Values.

- Implementation of Ordinary least squares estimation and Correlation algorithms.

Practical 7: Implementation of Dual Supervised Learning Algorithms.

- Black Box Methods, Neural Networks and Support Vector Machines.

Practical 8: Implementation of Clustering Algorithms.

- K-Means Clustering and Hierarchical Clustering.

Practical 9: Implementation of Association Rules.

- Market Basket Analysis Using Association Rules

Practical 10: Model Performance Testing.

- Evaluation Model Performance, Improving Model Performance, Bagging, Boosting, and Random forests

References:

1. APPLIED PREDICTIVE ANALYTICS: PRINCIPLES AND TECHNIQUES FOR THE PROFESSIONAL DATA ANALYST by DEAN ABBOTT, WILEY, 4th Edition, (2012)
2. PREDICTIVE ANALYTICS: THE POWER TO PREDICT by ERIC SIEGEL, WILEY
3. EFFICIENT R PROGRAMMING: A PRACTICAL GUIDE TO SMARTER PROGRAMMING by COLIN GILLESPIE AND ROBIN LOVELACE, SHROFF/O'REILLY
4. PYTHON AND R FOR THE MODERN DATA SCIENTIST: THE BEST OF BOTH WORLDS by RICK J. SCAVETTA AND BOYAN ANGELOV, SHROFF/O'REILLY
5. INTRODUCTION TO MACHINE LEARNING WITH R: RIGOROUS MATHEMATICAL ANALYSIS by SCOTT BURGER, SHROFF/O'REILLY
6. R IN A NUTSHELL 2E by JOSEPH ADLER, O'REILLY

INT312:BIG DATA FUNDAMENTALS

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

Course Outcomes: Through this course students should be able to

CO1 :: restate the need and importance of fundamental concepts and principles of Big Data

CO2 :: examine internal functioning of different modules of Big Data and Hadoop

CO3 :: understand the big data ecosystem and appreciate its key components

CO4 :: deduce how Hadoop solves those Big Data problems

CO5 :: apply tools and techniques to analyze Big Data

CO6 :: explore solution for a given problem using suitable Big Data Techniques

Unit I

Introduction to Hadoop : Introduction to Big Data, Types of Data, V's of Big Data, Introduction to Hadoop, Components of Hadoop, Installation of Apache Hadoop

Unit II

Hadoop Architecture : Hadoop Architecture, Hadoop Storage: HDFS, Hadoop MapReduce paradigm, MapReduce Terminology, Hadoop - Namenode, DataNode, Job Tracker and TaskTracker

Unit III

Map Reduce and YARN : the MapReduce model v1, limitations of Hadoop 1 and MapReduce 1, review of the Java code required to handle the Mapper class, the Reducer class, and the program driver needed to access MapReduce, the YARN model, comparison of YARN / Hadoop 2 / MR2 vs Hadoop 1 / MR1

Unit IV

Introduction to Apache Hive : hive installation, hive data types, hive bucketing, hive partitioning, hiveql operations, hive operators

Unit V

Introduction to Apache HBase : Installation of Apache Hbase, Hbase Fundamentals and Hbase Data Model, Hbase Architecture, General Commands in Apache Hbase, Interacting with Hbase using Java API, MapReduce with HBase

Unit VI

Introduction to Apache Pig : apache pig installation, pig latin basics, load and store operations in apache pig, pig diagnostic operators, grouping and joining in pig, combining and splitting in pig, pig filtering, pig sorting, pig latin built in functions

List of Practicals / Experiments:

Practical List

- Installation of apache hadoop
- Execution of hdfs commands
- Wordcount problem solving using apache hadoop
- Partitioning and Bucketing in Hive
- Interacting with Hbase using Java API
- Data Analysis using Apache Pig
- Java code required to handle the Mapper class, the Reducer class, and the program driver needed to access MapReduce

Text Books: 1. BIG DATA 2E by ANIL MAHESHWARI, Tata McGraw Hill, India

References: 1. BIG DATA ANALYTICS by RAJ KAMAL, PREETI SAXENA, Tata McGraw Hill, India

INT405:COMPUTING PRACTICUM-III

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

Course Outcomes: Through this course students should be able to

- CO1 :: understand the basic concept of swift language
- CO2 :: understand MVC architecture for simulating
- CO3 :: demonstrate different UI Kit controls to develop application
- CO4 :: categorize the concept of various Table controls view
- CO5 :: illustrate project template to develop app using Cocoa frameworks
- CO6 :: demonstrate auto- layout and webservices

List of Practicals / Experiments:

MVC architecture

- Introduction to Xcode
- Workspace window
- Interface-builder
- Attribute inspector
- Simulator & creating project
- Looking at view-controller
- Understanding outlets
- Actions
- Designing UI
- Application-delegate
- Working with navigator pane
- Utility pane

UIKit controls

- Implementing UIButton
- Creating small application with these controls
- Understanding & working with Views
- Multi-view applications
- Calling another view-controller using navigation-controller
- Passing & receiving data in them
- UITextField and customizing inputs
- UITextView & keyboard handling
- UISwitch
- UISegmented control
- UISlider
- UIImageView
- UIAlertView
- UIActionSheet

- Concept of segue

TableViews

- Implementing simple table
- Customizing TableView cells
- Grouped & indexed sections
- Adding header, footer & image
- Displaying item-selected
- Navigating to another view
- UITableView basics

iPhone Project templates

- Using navigation controllers
- Concept of auto-layout
- Playing audio & video files
- Implementing UICollectionView
- Implementing UIToolbar and UITabBar in applications

Web Services

- Displaying maps & monitoring changes using MapKitFramework
- UIWebView

Introduction to iPhone and iOS platform with Swift

- Difference between iOS and MAC OS
- Object-oriented programming
- Declaring & defining classes
- Variables
- Arrays
- Methods & messages
- Closure
- Dictionary

Text Books: 1. SWIFT IOS PROGRAMMING (24-HOUR TRAINER) by ABHISHEK MISHRA, WILEY

References: 1. MASTERING SWIFT 5 by JON HOFFMAN, PACKT PUBLISHING

MGN909:WORKSHOP ON STATISTICAL ANALYSIS

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

Course Outcomes: Through this course students should be able to

CO1 :: devise ability to create, edit and manage data sets using SPSS windows

CO2 :: develop ability to create data visualisation using appropriate graphs and charts.

CO3 :: practice effective use of cross tabulation, frequency distribution to gain insights from data

CO4 :: develop skills to use hypotheses testing procedure as per the data type/research question.

CO5 :: compute and interpret data using parametric and non parametric tests to draw the inference

List of Practicals / Experiments:

Creating and managing data with SPSS

- SPSS windows processes
- creating data file
- editing a data file
- managing data

Creating graphs and charts

- creating and editing graphs
- creating and editing charts

Frequencies and descriptive statistics

- frequencies
- bar charts
- histograms
- percentiles
- measures of central tendency

Cross tabulation and means procedure

- computing cross tabulation
- computing means

Bivariate correlation

- bivariate correlation and matrix

Testing of hypothesis

- one sample hypothesis tests
- two sample hypothesis tests
- linear regression

Multivariate analysis

- one way ANOVA
- two way ANOVA
- multiple regression
- factor analysis

Non-parametric procedures

- Chi Square test (Test of Independence)
- Mann Whitney U-test
- Wilcoxon Sign rank test
- Kruskal-Wallis h-test
- Spearman's rank order-correlation

Text Books:

1. DOING DATA ANALYSIS WITH SPSS by ROBERT H. CARVER AND JANE GRADWOHL NASH, CENGAGE LEARNING
2. DATA ANALYSIS USING SPSS by LOKESH JASRAI, SAGE PUBLICATIONS

References:

1. SPSS FOR WINDOWS STEP BY STEP by DARREN GEORGE AND PAUL MALLERY, PEARSON