RAJIV GANDHI PROUDYOGIKI VISHWAVIDYALAYA, BHOPAL

New Scheme Based On AICTE Flexible Curricula

Electrical & Electronics Engineering, VIII-Semester

Open Elective EX- 803 (D) Data Analytics

Course Objectives:

Data Analytics is the science of analyzing data to convert information to useful knowledge. This knowledge could help us understand our world better, and in many contexts enable us to make better decisions. While this is broad and grand objective, the last 20 years has seen steeply decreasing costs to gather, store, and process data, creating an even stronger motivation for the use of empirical approaches to problem solving.

This course will enable you with a wide range of data analytic techniques and is structured around the broad contours of the different types of data analytics, namely, descriptive, inferential, predictive, and prescriptive analytics.

Pre-requisites:

This course requires that you are familiar with high-school level linear algebra, and calculus.Knowledge of probability theory, statistics, and programming is desirable

UNIT-I
DESCRIPTIVE STATISTICS :Probability Distributions, Inferential Statistics ,Inferential Statistics through hypothesis tests Regression & ANOVA ,Regression ANOVA(Analysis of Variance).

UNIT-II
INTRODUCTION TO BIG DATA: Big Data and its Importance, Four V's of Big Data, Drivers for Big Data, Introduction to Big Data Analytics, Big Data Analytics applications.

BIG DATA TECHNOLOGIES: Hadoop's Parallel World, Data discovery, Open source technology for Big Data Analytics, cloud and Big Data, Predictive Analytics, Mobile Business Intelligence and Big Data, Crowd Sourcing Analytics, Inter- and Trans-Firewall Analytics, Information Management.

UNIT-III
PROCESSING BIG DATA: Integrating disparate data stores, Mapping data to the programming framework, Connecting and extracting data from storage, Transforming data for processing, subdividing data in preparation for Hadoop Map Reduce.

UNIT-IV

HADOOP MAPREDUCE: Employing Hadoop Map Reduce, Creating the components of Hadoop Map Reduce jobs, Distributing data processing across server farms, Executing Hadoop Map Reduce jobs, monitoring the progress of job flows, The Building Blocks of Hadoop Map Reduce Distinguishing Hadoop daemons, Investigating the Hadoop Distributed File System Selecting appropriate execution modes: local, pseudo-distributed, fully distributed.

UNIT-V
BIG DATA TOOLS AND TECHNIQUES: Installing and Running Pig, Comparison with Databases, Pig Latin, User- Define Functions, Data Processing Operators, Installing and Running Hive, Hive QL, Querying Data, User-Defined Functions, Oracle Big Data.

Reference Books and Study Materials:

1. Hastie, Trevor, et al.�The elements of statistical learning. Vol. 2. No. 1. New York: springer, 2009.

2. Montgomery, Douglas C., and George C. Runger.�Applied statistics and probability for engineers. John Wiley & Sons, 2010

3. NPTEL Video Course :Introduction to Data Analytics by Dr. Balaraman Ravindran Department of Computer Science and Engineering IIT Madras and   Dr. Nandan Sudarsanam Department of Management Studies IIT Madras.