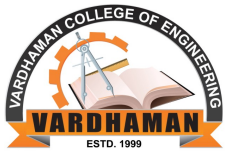

VARDHAMAN COLLEGE OF ENGINEERING

(AUTONOMOUS)

Affiliated to **JNTUH,** Approved by **AICTE,** Accredited by **NAAC** with **A++** Grade, **ISO 9001:2015** Certified

Kacharam, Shamshabad, Hyderabad - 501218, Telangana, India

www.vardhaman.org

CURRICULUM

For

Bachelor of Technology (Minors)

In

Data Science

Department of Computer Science and Engineering

Programme Curriculum Structure
B.Tech Minors in Data Science

S.No.	Yr/ Sem	Course Code	Course Name	Type	Credits
1	III/I	M1517	Foundations of Data Science	Theory	3
2	III/I	M1502	Data Analysis with Python	Practice	2
3	III/II	M1501	Big Data Tools	Theory	3
4	III/II	M1518	Business Analytics	Theory	3
5	IV/I	M1519	Data Visualization	Practice	2
6	IV/I	M1520	Web Analytics	Theory	3
7	IV/II	M1543	Mini Project in Minor Specialization	Project Work	2
Total Credits					18

Course Structure
M1517 - Foundations of Data Science

Hours Per Week			Hours Per Semester			Credits	Assessment Marks		
L	T	P	L	T	P	C	CIE	SEE	Total
3	0	0	42	0	0	3	30	70	100

1. Course Description

Course Overview

Data Science is exponentially growing field which consists of a set of tools and techniques used to extract useful information from data.This Course will cover the fundamentals of Data Science, Data formats available and Techniques to analyze data. Here students will understand various pre-processing methods to transform the raw data into a useful format and use exploratory data analytics.The course also enables to understand the R programming for reading data and apply various statistis on it.

Course Pre/co-requisites

The course has no specific prerequisite and co requisite.

2. Course Outcomes (COs)

After the completion of the course, the student will be able to:

M1517.1 Identify the design issues and challanges of data science in real time.
M1517.2 Choose data pre-processing techniques to extract useful data
M1517.3 Select a data analytical techniques on given real time data.
M1517.4 Use Model Development strategies for a given data set.
M1517.5 Choose various statistical methods in R for data analytics.

3. Course Syllabus

Data Science: Introduction to Data Science, Evolution of Data Science, Data Science Roles, Stages in a Data Science Project, Applications of Data Science in various fields, Data Security Issues.

Data Collection and Data Pre-Processing: Data Collection Strategies, Data PreProcessing Overview, Data Cleaning, Data Integration and Transformation, Data Reduction, Data Discretization.

Data Analysis Techniques: Introduction, Data Analysis and Data Analytics, Descriptive Analysis: Variables, Frequency Distribution, Measures of Centrality, Dispersion of a Distribution, Diagnostic Analytics: Correlations, Predictive Analytics, Prescriptive Analytics,Exploratory Analysis, Mechanistic Analysis.

Model Development: Simple and Multiple Regression, Model Evaluation using Visualization, Residual Plot, Distribution Plot, Polynomial Regression and Pipelines, Measures for In-sample Evaluation, Prediction and Decision Making.

Data Analytics using R: Data types in R, Vectors, Matrices and Reading data, Using Statistics, Percentiles and Quartiles, Measures of Central Tendency, Measures of Variability, Methods of Displaying Data- Grouped Data and the Histogram, Skewness and Kurtosis,Box Plots, Pivot Table, Heat Map, Exploratory Data Analysis.

4. Books and Materials

Text Books:

1. Chirag Shah ., A Hands-On Introduction to Data Science, Cambridge University Press, 2020
2. David Dietrich, Barry Heller, Beibei Yang, Data Science and Big data Analytics, EMC, 2013
3. Seema Acharya.,Data Analytics Using R, McGrwahill India, 2018.

Reference Books:

1. Cathy O'Neil and Rachel Schutt , Doing Data Science, O'Reilly, 2015.

Course Structure
M1502 - Data Analysis with Python

Hours Per Week			Hours Per Semester			Credits	Assessment Marks		
L	T	P	L	T	P	C	CIE	SEE	Total
0	0	4	0	0	56	2	30	70	100

1. Course Description

Course Overview

Python is a powerful programming language used for Artificial Intelligence Machine Learning applications. In this course learners will be able to implement various python libraries Numpy and Pandas to work with multi dimensional arrays and data sets of real time problems. NumPy and Pandas libraries are essential and fundamental requirement for building machine learning models.

Course Pre/co-requisites

The course has no specific prerequisite and co requisite.

2. Course Outcomes (COs)

After the completion of the course, the student will be able to:

M1502.1 Identify the process of data analysis for a given data set.
M1502.2 Make use of Numpy functions to perform various mathematical operations on arrays.
M1502.3 Use Pandas to perform data analysis in machine learning applications.
M1502.4 Construct various plots on real data to identify useful insights.

3. Course Syllabus

Introduction to Data Analysis: Understanding the Nature of the Data, The Data Analysis Process, Data Extraction, Data Preparation, Data Exploration/Visualization, Predictive Modeling, Model Validation, Deployment, Quantitative and Qualitative Data Analysis, Open Data, Python and Data Analysis.

Introduction to NumPy: NumPy arrays, special numeric values, creating numpy arrays, creating ndarray, Numpy operators. Operations on NumPy Arrays: Selecting elements explicitly,Slicing arrays,Expanding arrays,Arithmetic and linear algebra with arrays,Employing array methods and functions.

NumPy Arrays: Computation on NumPy Arrays, Universal Functions, Aggregations,Min, Max, and Everything in Between, Computation on Arrays: Broadcasting Comparisons, Masks, and Boolean Logic, Fancy Indexing, Sorting Arrays,Structured Data: NumPy's Structured Arrays,

Pandas and Opearations: Exploring series and Dataframe objects, creating series, creating dataframes, adding data, saving data frames, subsetting your data, indexing methods, Arithmetic, Function Application and Mapping with pandas: arithmetic operations with dataframes,vectorization with dataframes,Dataframe function application,handling missing data in pandas DataFrame.

Managing, Indexing and Plotting: Using pandas for Data Analysis, Index sorting, Hierarchial indexing,plotting with pandas.

List of Programs for Practice

1. Program on Numpy Aggregations: Min, Max, and etc.. Example What Is the Average Height of prime ministers of india?
2. Program using NumpyComparisons, Masks, and Boolean Logic example: Counting Rainy Days
3. Program using Numpy Fancy Indexing example: Selecting Random Points.
4. Write a NumPy program to create a 3x3 identity matrix.
5. Write a NumPy program to create a vector of length 10 with values evenly distributed between 5 and 50.
6. Program using Pandas to Combining Datasets: Join.
7. Program using Pandas on Pivot Tables.
8. Program using Pandas to Vectorized String Operations.
9. Program using Pandas to Working with Time Series Example: Visualizing Seattle Bicycle Counts.
10. Write a NumPy program to swap rows and columns of a given array in reverse order.
11. Write a NumPy program to compute the mean, standard deviation, and variance of a given array along the second axis.
12. Write a NumPy program to sort the student id with increasing height of the students from given students id and height. Print the integer indices that describes the sort order by multiple columns and the sorted data.

4. Laboratory Equipment/Software/Tools Required

1. A computer System with Ubuntu Operating System.
2. Python 3.x (Any Latest Version).
3. Anakonda , Jupyter Notebook

5. Books and Materials

Text Books:

1. Jake VanderPlas. Python Data Science Handbook, 1st Edition, Published by O'Reilly Media, 2017.
2. Curtis Miller., Hands-On Data Analysis with NumPy and pandas, Packt Publishing, India,2018.

Reference Books:

1. Jake VanderPlas., Python Data Science Handbook,1st *Edition*,O'Reilly Publications Media, 2017.
2. Fabio Nelli., Python Data Analytics with Pandas, NumPy, and Matplotlib, 2nd Edition, Apress, 2018.

Course Structure
M1501 - Big Data Tools

Hours Per Week			Hours Per Semester			Credits	Assessment Marks		
L	T	P	L	T	P	C	CIE	SEE	Total
3	0	0	42	0	0	3	30	70	100

1. Course Description

Course Overview

It serves as an introductory course for graduate students who are expecting to face Big Data storage, processing, analysis, visualization, and application issues on both workplaces and research environments.Big data analytics is the use of analytic techniques against very large, diverse data sets that include structured, semi-structured and unstructured data, from different sources and in different sizes. Analysis of big data allows users to make better and faster decisions using data that was previously inaccessible or unusable. Students will gain knowledge on analysing Big Data.

Course Pre/co-requisites

The course has no specific prerequisite and co requisite.

2. Course Outcomes (COs)

After the completion of the course, the student will be able to:

M1501.1 Identify the fundamental concepts of big data analytics.
M1501.2 Select Hadoop environment and apply HDFS commands on file management tasks.
M1501.3 Make use of NoSQL databases like MangoDB to stock log data to be pulled for analysis.
M1501.4 Choose modern tools Pig and Hive for complex data flow and analysis.

3. Course Syllabus

Big Data: Classification of Digital Data, Characteristics of Data, Def-inition of Big Data, Challenges with Big Data, Traditional Business Intelligence (BI) versus Big Data, Realms of Big Data,Classification of Analytics,Few Top Analytics Tools. NoSQL:NoSQL (Not Only SQL), Types of NoSQL Databases, SQL versus NoSQL, RDBMS versus Hadoop, Distributed Computing Challenges

HDFS: Introduction to Hadoop, Hadoop Overview, Hadoop Distributors, HDFS (Hadoop Distributed File System),Working with HDFS commands,Interacting with Hadoop Ecosystem.

MONGODB: Features of MongoDB, RDBMS vs MongoDB, Data Types in MongoDB, MongoDB Query Language, CRUD operations,Count, Limit, Sort, and Skip.

Hive: Saying Hello to Hive, Seeing How the Hive is Put Together, Getting Started with Apache Hive,Examining the Hive Clients, Working with Hive Data Types, Creating and Managing Databases and Tables, Seeing How the Hive Data Manipulation Language

Works, Querying and Analyzing Data.

Pig: Hadoop Programming Made Easier Admiring the Pig Architecture, Going with the Pig Latin Application Flow, Working through the ABCs of Pig Latin, Evaluating Local and Distributed Modes of Running Pig Scripts, Checking out the Pig Script Interfaces, Scripting with Pig Latin.

4. Books and Materials

Text Books:

1. Seema Acharya, Subhashini Chellappan. Big Data and Analytics, 2nd Edition,Wiley IndiaPrivate Limited, New Delhi, 2019.

Reference Books:

1. Tom White. Hadoop - The Definitive Guide, 4th Edition, O'Reilly Publications, India,2015.
2. Judith Hurwitz, Alan Nugent, Dr. Fern Halper, Marcia Kaufman. Big Data for Dummies, John Wiley & Sons, Inc., 2013.

Course Structure
M1518 - Business Analytics

Hours Per Week			Hours Per Semester			Credits	Assessment Marks		
L	T	P	L	T	P	C	CIE	SEE	Total
3	0	0	42	0	0	3	30	70	100

1. Course Description

Course Overview

This course gives the fundamental concepts of business analytics and gives insights about tools in understanding how data analysis works in today's organizations. This course helps to apply basic business analytics principles and effectively use them in decision making. Also this course gives knowledge of predictive analytics in business forecasting. The course enables to identify the need of data analytics in HR , Supply Chain Management Marketing and Sales.

Course Pre/co-requisites

M1501 - Big Data Analytics

2. Course Outcomes (COs)

After the completion of the course, the student will be able to:

M1518.1 Identify the need of data analytics in business.
M1518.2 Make use of data analytics tools for business intelligence.
M1518.3 Choose predictive analytics for forecasting business.
M1518.4 Identify the data analytics process for HR and Supply chain management.
M1518.5 Use data analytics on marketing and sales data.

3. Course Syllabus

Introduction to Business Analytics: Analytics and Data Science, Analytics Life Cycle, Types of Analytics,Business Problem Definition, Data Collection , Data Preparation , Hypothesis Generation ,Modeling ,Validation and Evaluation , Interpretation , Deployment and Iteration

Business Intelligence: Data Warehouses and Data Mart , Knowledge Management ,Types of Decisions, Decision Making Process , Decision Support Systems , Business Intelligence , OLAP , Analytic functions

Business Forecasting: Introduction to Business Forecasting and Predictive analytics , Logic and Data Driven Models , Data Mining and Predictive Analysis Modeling , Machine Learning for Predictive analytics.

HR & Supply Chain Analytics: Human Resources , Planning and Recruitment , Training and Development , Supply chain network , Planning Demand, Inventory and Supply , Logistics , Analytics applications in HR & Supply Chain.

Marketing & Sales Analytics: Marketing Strategy, Marketing Mix, Customer Behavior , selling Process , Sales Planning , Analytics applications in Marketing and Sales.

4. Books and Materials

Text Books:

1. James R. Evans, Business Analytics Methods, Models, and Decisions, 3rd Edition,Pearson Publications, New Delhi, 2020.

Reference Books:

1. R N Prasad, Seema Acharya., Fundamentals of Business Analytics, 2nd Edition, Wiley Publications, 2015.

Course Structure
M1519 - Data Visualization

Hours Per Week			Hours Per Semester			Credits	Assessment Marks		
L	T	P	L	T	P	C	CIE	SEE	Total
0	0	4	0	0	56	2	30	70	100

1. Course Description

Course Overview

This course is all about addressing today's data driven world, in visualizing the data graphically. Though its foundations are rooted in statistics, many advanced fields of science are practicing it to explore their large volumes of data and to throw deeper insights of the data. The main aim of this course is to give better understandings of data and making sense of hidden information. Visualizing enormous data using graphics can run all possible unknown stories about data. The course uses some of the open source data Visualization tools.

Course Pre/co-requisites

M1502 - Data Analysis with Python

2. Course Outcomes (COs)

After the completion of the course, the student will be able to:

M1519.1 Examine the various graphs for a given data set.
M1519.2 Make use of python libraries in visualizing data.
M1519.3 Design visualizations to extract useful insights from data.

3. Course Syllabus

Introduction: Importance of Data Visualization, Visualization plots-Bar, Pie, Histogram, Box, Line,scatter plots.

Correlation: Importance of each plot, Correlation-Importance of correlation of variables

Working with data: creating Python-Lists, Tuples, Data Frames. Creating a CSV file, loading a csv file into python Data Frame. Analyzation of Datasets- IRIS, TITANIC

Matplotlib plotting: Visualization of IRIS data-line,Bar, Pie,Histogram, Box, Scatter plots, correlation plots-Heatmaps

Seaborn plotting: Visualization of Titanic data-Strip, Box, Swarm, Joint plots, correlation plots-Heatmaps.

Managing, Indexing and Plotting: Using pandas for Data Analysis, Index sorting, Hierarchial indexing,plotting with pandas.

List of Programs

Plot Various graphs on the following data sets.

1. Students Performance in Exams.
2. 2018 Airplane Flights – Predicting prices of airline flights.
3. E-Learning Student Reactions.
4. Uber Traffic Data Visualization.
5. factors-affecting-campus-placement.
6. Human Resources Data Set.

4. Laboratory Equipment/Software/Tools Required

1. A Computer with Ubuntu Operating System
2. Anakonda, Jupyter Notebook
3. Python 3.X or above version

5. Books and Materials

Text Books:

1. Swapnil Saurav , Data Visualization using Python, Learn and Practice, 2020.

Reference Books:

1. Purna Chandar Rao Kathula., Hands-On Data Analysis And Visualization With Pandas,BPB Publication,2020.
2. https://mksaad.wordpress.com/2020/06/30/datasets-for-visualization/

Course Structure
M1520 - Web Analytics

Hours Per Week			Hours Per Semester			Credits	Assessment Marks		
L	T	P	L	T	P	C	CIE	SEE	Total
3	0	0	42	0	0	3	30	70	100

1. Course Description

Course Overview

Web analytics ia a way of learning how users interact with websites by automatically recording aspects of user's behaviour then combining and transforming the behaviour into data that can be analyzed.This course allows to learn need , working of web analytics , goals and also enables to implement traffic analysis and click path analysis.The student can also learn how google data analytics work in web data analytics.

Course Pre/co-requisites

M1501 - Big Data Analytics

2. Course Outcomes (COs)

After the completion of the course, the student will be able to:

M1520.1 Select the web analytics approach for extracting data from websites and mobiles.
M1520.2 Identify the various metrics for web data aalytics.
M1520.3 Make use of goals in web data for desired action on website.
M1520.4 Choose traffic analysis for collecting and interpreting key data points
M1520.5 Examine the users move in web sites with Click-Path Analysis:

3. Course Syllabus

Introduction to Web Analytics: Definition, User Experience and Web Analytics Questions. Web Analytics Approach: Introduction, a model of Analysis, Show casing the work, Context Matters, Contradicting the data.

Working of Web Analytics: Introduction, Log File Analysis, Page Tagging, Metrics and Dimensions, Interacting with data in Google Analytics.

Goals: Definition of Goals and Conversions, Conversion Rate, Goal Reports in Google Analytics, Finding the right things to measure as key, Performance Indicators, Measure on a website that can constitute a goal. Learning about users: Introduction, Visitor Analysis.

Traffic Analysis: Introduction, Source and medium, Organic Search, Search Query Analysis, Referral Traffic, Direct Traffic, Paid Search Keyword. Analyzing usage of content: introduction, Website content Reports.

Click-Path Analysis: Introduction, Focus on Relationships between pages, Navigation Summary, Visitors Flow Report, analyzing how users move from one-page type to another.

4. Books and Materials

Text Books:

1. Avinash Kaushik., Web Analytics An Hour a Day, Wiley Publishing, Inc, 2007
2. Michael Beasley, Morgan Kaufmann.,Practical Web Analytics for User Experience , 1st Edition, MK Publishing, 2013.

Reference Books:

1. Clifton B., Advanced Web Metrics with Google Analytics, 2nd Edition,Wiley Publishing, Inc.2017.