Vivekanand Education Society's Institute of Technology

Approved by AICTE & Affiliated to University of Mumbai

AI and Data Science Department

Data Analytics and Visualization (DAV)

Year: 2023-24 (Even)

Sr.	Module	Weightage (Topics to be considered)	Hours allocated
1	**Module - 1 : Introduction to Data analytics and life cycle** Data Analytics Lifecycle overview: Key Roles for a Successful Analytics, Background and Overview of Data Analytics Lifecycle Project. Phase 1: Discovery: Learning the Business Domain, Resources Framing the Problem, Identifying Key Stakeholders. Interviewing the Analytics Sponsor, Developing Initial Hypotheses Identifying Potential Data Sources. Phase 2: Data Preparation: Preparing the Analytic Sandbox, Performing ETLT, Learning About the Data, Data Conditioning, Survey and visualize, Common Tools for the Data Preparation Phase. Phase 3: Model Planning: Data Exploration and Variable Selection, Model Selection , Common Tools for the Model Planning Phase. Phase 4: Model Building: Common Tools for the Model Building Phase. Phase 5: Communicate Results. Phase 6: Operationalize	Approximately **15 Marks** ● Life cycle with detailed steps ● Case Study Based on Lifecycle	5
2	**Module - 2: Regression Models** Introduction to Regression, Types of Regression: Simple Linear Regression, Multiple Linear Regression, Polynomial Regression, Interaction Regression, Weighted least squares Regression, Ridge Regression, Loess Regression, Bootstrapping Regression. Qualitative predictor variables, Model Evaluation Measures, Model selection procedures, Leverage, Influence measures, Diagnostics. Logistic Regression: Logistic Response function and logit, Predicted values from Logistic Regression, Interpreting the coefficients and odds ratios Generalized Linear model, Logistic Regression Vs GLM, Linear Regression Vs Logistic Regression, Assessing the models.	Approximately **25 Marks** ● Types of Regression ● Problems on SLR and MLR ● Evaluation Measures ● Types of Diagnostic Plots ● Comparison between Logistic and Linear Regression ● Comparision btw SLR and MLR ● Interpreting the model summary	6

3	**Module - 3 : Time Series** Definition of time series, Times series forecasting. Time series components, Decomposition – additive and multiplicative. Exponential smoothing, Holt winters method. Time Series Analysis - Box-Jenkins Methodology, ARIMA Model Autocorrelation Function (ACF, PACF) Autoregressive Models ,Moving Average Models ,ARMA and ARIMA Models , Building and Evaluating an ARIMA Model.	Approximately **20 Marks** ● Components of TS ● Decomposition methods ● Box-Jenkins Methodology ● ARMA Model ● ARIMA Model ● PACF and ACF Plots	7
4	**Module - 4 : Introduction to Data Visualization** Acquiring and Visualizing Data, Simultaneous acquisition and visualization, Applications of Data Visualization, Keys factors of Data Visualization , Exploring the Visual Data Spectrum: charting Primitives (Data Points, Line Charts, Bar Charts, Pie Charts, Area Charts), Exploring advanced Visualizations (Candlestick Charts, Bubble Charts, Surface Charts, Map Charts) ; Narrative visualization and digital story Telling ,infographics and interactive dashboards	Approximately **20 Marks** ● Life cycle of Data Viz ● Charting Primitives ● Advanced Viz ● Comparison between basic charts ● Tools used for Data Viz	8
5	**Module - 5 : Introduction to D3.js:** Getting setup with D3, Making selections, changing selection's attribute, Loading and filtering External data. Building a graphic that uses all of the population distribution data, Data formats you can use with D3, Creating a server to upload your data, D3's function for loading data, Dealing with Asynchronous requests, Loading and formatting Large Data Sets	Approximately **10 Marks** ● Functions used for selection ● Simple program for transformations ● Steps to Viz data using D3.js	7
6	**Module - 6 : Data analytics and Visualization with Python** Essential Data Libraries for data analytics: Pandas, NumPy, SciPy. Plotting and visualization with python: Introduction to Matplotlib, Basic Plotting with Matplotlib, Create Histogram, Bar Chart, Pie Chart, Box Plot, violin plot using Matplotlib, Matrix charts and heat maps. Introduction to seaborne Library, Multiple Plots, Regression plot, replot. Discover and visualize the data to gain insights, Feature scaling and Transformation pipelines	Approximately **10 Marks** ● Short notes on Pandas, NumPy and SciPy ● Comparison between matplotlib and seaborne	6