About the Course

Data Analytics & Business Intelligence is the need of the hour. Today, huge amounts of data are being generated without knowing how to make beneficial use of it. In order to utilize this, machine learning and statistical techniques are being used to develop predictive models from existing data to forecast future outcomes.

Rationale

Data analytics and business intelligence are of great importance in today's world. Data analysis is required to understand organizational problems and to explore data while business intelligence helps companies to make better decisions by showing current and historical data within their business context. To make the organization run smoother and more efficiently, analysts use BI to provide performance and competitor benchmarks.

Aims

The fundamental aim of leveraging Data Analysis and Business Intelligence skills is to help understand trends and derive actionable insights from data, thus allowing us to make data-driven, strategic and tactical business decisions.

Learning Outcomes

Students will be able to apply principles of statistics, python programming, machine learning, probability and decision making in the context of data analysis. Moreover, they should design tested and effective advanced analytics models for decision making and communicate effectively in a variety of modes and contexts.

Objectives

Expecting to build a solid business analytics foundation, this course has been designed to impart knowledge of machine learning and statistical methods for data analysis. The course shall also provide sufficient knowledge of python programming language for machine learning algorithms and python/ R programming for statistical methods. A brief introduction to neural networks and deep learning will also be covered.

Target Audience

Those who are interested in developing a strong foundation in business analytics and have graduated or are pursuing graduation (studied mathematics till class 12th level).

Future Prospects

Upon completion of the course, the students will be able to enhance their skills in data analysis,

python programming for machine learning and python/ R programming for statistical methods. They will also be able to find answers to the questions they don't know the answers to. This course will help them to adapt themselves to the automated future of business intelligence.

Course Duration: 125 Hours [January, 2025 - June, 2025]

Course Evaluation: Evaluation of the course will be done based on Class Assignments, End term Examination, and Project Work.

Fees: Rs. 40,200/-

(Course Fees: Rs. 40000/- plus taxes if any. Application Fees: Rs. 200/-)

Resource Persons
1. Eminent resource persons from reputed institutes will be invited.
2. The experts from various industries will also be invited for delivering lectures and hands-on training.

Course Contents

Module 1: Foundation of Data Analytics & Python Programming   **(20 hours)**

Foundation of Data Analytics: - Introduction, Evolution , Concept and Scopes, Data, Big Data, Metrics and Data classification, Data Reliability & Validity, Problem Solving with Analytics, Different phases of Analytics in the business and Data science domain, Descriptive Analytics, Predictive Analytics and Prescriptive Analytics, Different Applications of Analytics in Business, Text Analytics and Web Analytics, Skills for Business Analytics, Concepts of Data Science, Basic skills required for understanding Data Science.

Python Programming: - Introduction to Python Editors & IDE's (Jupyter, Spyder, pycharm, etc.), custom environment settings, basic data types-numeric, string, float, tuples, list, dictionary, sets and their operations, control flow (if-elif-else), loops (for, while), inbuilt functions for data conversion, writing user defined functions.

Concepts of packages/libraries – important packages like NumPy, SciPy, scikit-learn, Pandas, Matplotlib, seaborn, etc., installing and loading packages, reading and writing data from/to different formats, simple plotting, functions, list comprehensions, database connectivity, Playing with Date Format.

Module 2: Probability & Statistics   **(30 hours)**

Descriptive Analytics: Introduction to Descriptive Analytics, Measures of central tendency: mean, median, mode. Measures of dispersion: range, variance, standard deviation. Understanding skewness and kurtosis and their implications in data analysis, Five number summary, data visualization: Bar diagrams, Histograms, box plots, and scatter plots, Understanding how to interpret and draw insights from these visualizations, Correlation Analysis, Outlier detection: Z-score method, IQR method, box plots, Introduction to Exploratory Data Analysis

Probability: Measures of probability, conditional probability, independent event, Bayes' theorem, random variable, discrete (binomial, Poisson, geometric, hypergeometric, negative binomial) and continuous (uniform, exponential, normal, gamma). Expectation and variance, markov inequality, chebyshev's inequality, central limit theorem.

Inferential Statistics: Sampling & Confidence Interval, Inference & Significance. Estimation and Hypothesis Testing, t- test/z-test (one sample, independent, paired), Goodness of fit test, Test of Independence,

Module 3: Data Munging & Data Visualization   **(10 hours)**

Relevance in industry, Statistical learning vs machine learning, types and phases of analytics.

Data pre-processing and cleaning: data manipulation steps (sorting, filtering, duplicates, merging, appending, subsetting, derived variables, data type conversions, renaming, formatting, etc.), normalizing data, sampling, missing value treatment, outliers detection and treatment.

Exploratory data analysis and Data Visualization : Data visualization using matplotlib, seaborn libraries, creating graphs (bar/line/pie/boxplot/histogram, heatmap etc.), bivariate analysis (cross tabs, pivot table, distributions and relationships, graphical analysis).

Module 4: Machine learning – Part 1   **(20 hours)**

Introduction, Applications of Machine Learning, Key elements of Machine Learning, Supervised vs.Unsupervised Learning.

Supervised Machine Learning: Linear Regression, Multiple Linear Regression Polynomial Regression.

Classification: Using Logistic Regression, Logistic Regression vs. Linear Regression, Logistic

Regression with one variable and with multiple variables, Application to multi-class classification. The problem of Overfitting, Application of Regularization in Linear and Logistic Regression. Regularization and Bias/Variance. Classification using K-NN, Naive Bayes classifier, Classification & Regression Tree (CART) for Decision Tree classification, Random Forest, Support Vector Machines.

Natural Language Processing (NLP): Definition and scope of NLP, Applications of NLP in data analytics, Text classification, sentiment analysis

Model Evaluation: Cross validation types (train & test, bootstrapping, k-fold validation), parameter tuning, confusion matrices, basic evaluation metrics, precision-recall, ROC curves.

Case study with critical thinking

Module 5: Machine learning – Part 2 (15 hours)

Neural Networks: Introduction, Model Representation, Gradient Descent vs. Perceptron Training, Stochastic Gradient Descent, Multiclass Representation, Multilayer Perceptrons, Backpropagation Algorithm for Learning, Introduction to Deep Learning.

Association Rule Mining: Mining frequent item sets, Apriori algorithm, market basket analysis.

Case study with critical thinking

Unsupervised Machine Learning: Introduction, Clustering, K-Means algorithm, Affinity Propagation, Agglomerative Hierarchical, DBSCAN, Dimensionality Reduction using Principal Component Analysis.

Case study with critical thinking

Time Series Forecasting: Trends and seasonality in time series data, identifying trends, seasonal patterns, first order differencing, periodicity and autocorrelation, rolling window estimations, stationarity vs. non-stationarity, ARIMA modeling, time series forecasting using XGBoost

Case study with critical thinking

Module 6: Optimization in Analytics (5 hours)

Introduction to Operations Research (OR), Linear Programming Problems (LPP), Geometry of linear programming, Sensitivity and Post-optimality analysis, Duality, and its economic interpretation. Introduction to Non-linear Programming

Module 7: Introduction to SQL & No-SQL (10 hours)

Introduction to DBMS. Learning SQL query structure with examples, Operators in SQL, SQL vs No-SQL, CRUD operations using SQL and No-SQL using Python.

Introduction to Data management and query system OLTP and OLAP and Their data models, Data warehousing, ETL and data integration

Module 8: Business Intelligence through Excel & Power BI (15 hours)

Advanced Excel for Analytics: Data Cleaning and Processing, Vlookup, Pivot table, Pivot Charts and Dashboards, Date functions, Conditional Formatting and Data Validation, VBA Functions and Procedures

Power BI: Concepts of Business intelligence (BI), Relevance of BI in business analytics, industry and different domains, Report Generation and Dashboard creation using Power BI.

Course Co-coordinator:

Dr. Rishi Rajan Sahay (Mob: 9818011766, Email: rajansahay@sscbsdu.ac.in),

Course email id: dataanalytics@sscbsdu.ac.in