

Maximization Algorithms, Simulation-Based Algorithms, Proxy-Based Algorithms, Sketch-Based Algorithms, Community-Based Influence Maximization, and Context-Aware Influence Maximization.

**Unit-V: Multilayer Social Network:** Multilayer Social Networks, Formation of Multilayer Social Networks, Heuristic-based Approaches, Greedy Approaches, Centrality-based Approaches, Meta-heuristic Approaches, Path-based Approaches, Measuring Multilayer Social Networks.

**Readings:**

1. Tanmoy Chakraborty, Social Network Analysis, Wiley India, 2021.
2. David Knoke and Song Yang. Social network analysis. SAGE publications, 2019.
3. Mark E. Dickison, Matteo Magnani and Luca Rossi, Multilayer social networks, Cambridge University Press, 2016.
4. Jennifer Golbeck, Analyzing the social web, Morgan Kaufmann, 2013.
5. Stanley Wasserman, and Katherine Faust. Social network analysis: Methods and applications, Cambridge University Press, 2012.
6. M.E.J. Newman, Networks: An introduction. Oxford University Press, 2010.
7. Wei Chen, Carlos Castillo and Laks V.S. Lakshmanan, Information and influence propagation in social networks. Springer Nature, 2014
8. Virinchi Srinivas and Pabitra Mitra, Link prediction in social networks: role of power law distribution. New York: Springer International Publishing, 2016

**MCSO301: DATA ANALYSIS AND VISUALIZATION [3-0-1]**

**Course Objectives:** The course develops students' competence in cleaning and analyzing data related to a chosen application. It also aims to develop skills in using various tools for data visualization and choosing the right tool for given data.

**Course Learning Outcomes:**

On completing the course, the students will be able to :

**CO1:** use data analysis tools with ease.

**CO2:** load, clean, transform, merge, and reshape data.

**CO3:** create informative visualisations and summarise data sets.

**CO4:** analyse and manipulate time series data.

**CO5:** solve real world data analysis problems.

**Syllabus**

**Unit 1 Introduction:** Introduction to Data Science, Exploratory Data Analysis and Data Science Process. Motivation for using Python for Data Analysis, Introduction of Python shell iPython and Jupyter Notebook. Essential Python Libraries: NumPy, pandas, matplotlib, SciPy, scikit-learn, statsmodels

**Unit 2 Getting Started with Pandas:** Arrays and vectorized computation, Introductio to pandas Data Structures, Essential Functionality, Summarizing and Computing Descriptive Statistics. Data Loading, Storage and File Formats. Reading and Writing Data in Text Format, Web Scraping, Binary Data Formats, Interacting with Web APIs, Interacting with Databases Data Cleaning and Preparation. Handling Missing Data, Data Transformation, String Manipulation

**Unit 3 Data Wrangling:** Hierarchical Indexing, Combining and Merging Data Sets Reshaping and Pivoting. Data Visualization matplotlib: Basics of matplotlib, plotting with pandas and seaborn, other python visualization tools

**Unit 4 Data Aggregation and Group operations:** Data grouping, Data aggregation, General split-apply-combine, Pivot tables and cross tabulation

**Unit 5 Time Series Data Analysis:** Date and Time Data Types and Tools, Time series Basics, Frequencies and Shifting, Time Zone Handling, Periods and Periods Arithmetic, Resampling and Frequency conversion, Moving Window Functions.

### **Readings**

1. McKinney, W.(2017). Python for Data Analysis: Data Wrangling with Pandas, NumPy and IPython. 2nd edition. O'Reilly Media.
2. O'Neil, C., & Schutt, R. (2013). Doing Data Science: Straight Talk from the Frontline, O'Reilly Media.

## **MCSO302: DATA SCIENCE [3-0-1]**

**Course Objectives:** The objective of this course is to analyze the data statistically and discover valuable insights from it. The course gives hands-on practice on predictive and descriptive modeling of the preprocessed data. In addition, the student also learns to apply mining association rules from the transactional data and mining text from the document will also be covered during the course.

### **Course Learning Outcomes:**

On completion of this course, the student will be able to:

**CO1:** demonstrate proficiency with statistical analysis of data.

**CO2:** develop the ability to build and assess data-based models.

**CO3:** execute statistical analyses and interpret outcomes.

**CO4:** apply data science concepts and methods to solve problems in real-world contexts and will communicate these solutions effectively.

### **Syllabus:**

**Unit-I Introduction:** Introduction data acquisition, data preprocessing techniques including data cleaning, selection, integration, transformation, and reduction, data mining, interpretation.

**Unit-II Statistical data modeling:** Review of basic probability theory and distributions, correlation coefficient, linear regression, statistical inference, exploratory data analysis, and visualization.

**Unit-III Predictive modeling:** Introduction to predictive modeling, decision tree, nearest neighbor classifier, and naïve Bayes classifier, classification performance evaluation, and model selection.

**Unit-IV Descriptive Modeling:** Introduction to clustering, partitional, hierarchical, and density based clustering (k-means, agglomerative, and DBSCAN), outlier detection, clustering performance evaluation.

**Unit-V Association Rule Mining:** Introduction to frequent pattern mining and association rule mining, Apriori algorithm, measures for evaluating the association patterns.

**Unit-VI Text Mining:** Introduction of the vector space model for document representation, term frequency-inverse document frequency (tf-idf) approach for term weighting, proximity measures for document comparison, document clustering, and text classification.

### **Readings:**