Savitribai Phule Pune University
(Formerly University of Pune)

Department of Technology
Board of Studies, M. Tech Data Science (DS)
Curriculum Structure for M.Tech Program

Sr. No.	Subject Code	Subject Name	Credits	Teaching Scheme (Theory)	Teaching Scheme (Lab)
Semester (I)					
1	DSC1	Python Programming	4	√	
2	DSC2	Applied Statistics	4	√	
3	DSC3	Data Mining & Predictive Modeling	4	√	
4	DSE1*	Elective-1	4	√	
5	DSLP1	Lab Practice – 1	4		√
6	DSS1	Seminar – 1	2		√
		Total	**22**		
Semester (II)					
7	DSC4	Business Intelligence	4	√	
8	DSC5	Machine Learning	4	√	
09	DSC6	Deep Learning	4	√	
10	DSE2*	Elective-2	4	√	
11	DSLP2	Lab Practice – 2	4		√
12	DSS2*	Internship\Seminar – 2	2		√
		Total	**22**		
Semester (III)					
13	DSC7	Research Methodology	4	√	
14	DSE3*	Elective-3	4	√	
15	DSS3*	MOOC\Internship	4	√	√
16	DSIntProj	Interim Project	10		√
		Total	**22**		
Semester (IV)					
17	DSS4*	Open Elective\MOOC\Module	4	√	

18	DSS5*	Journal\Conference\Internship\MOOC	2		√
19	DSFinProj	Final Project (Dissertation Submission)	16		√
		Total	**22**		
		TOTAL CREDITS	**88**		

*-Detailed codes are given in given in Table A and Table-B

Table –A LIST OF ELECTIVES AND DIRECTED STUDY SUBJECTS FOR MTECH DATA SCIENCE

Sr. No.	Subject Code	Subject Name
		DSE1*
1	DSE1	Mathematics for Technology
2	DSE2	Introduction to Data Analytics
3	DSE3	Data Warehousing and Multidimensional Modelling
		DSE2*
5	DSE4	Big Data Analytics
6	DSE5	Application of Machine Learning in Industries
7	DSE6	Social and Web Analytics
		DSE3*
9	DSE7	Computational Linguistics and Natural Language Processing
10	DSE8	Analytics for Industries
11	DSE9	Pattern recognition and Anomaly Detection

Table –B LIST OF CODES FOR SUBJECT SPECIFIC, MOOCs AND OPEN ELECTIVE SUBJECTS

Sr. No.	Subject Code	Subject Name
		DSS2*
12	DSSI1	Internship
13	DSSS2	Seminar-2
		DSS3*

15	OOC1	MOOC
16	DSSI2	Internship
DSS4*		
17	DSS-BM	Banking Module
18	DSS-IM	Insurance Module
23	OE**	Open Elective
24	OOC2	MOOC
DSS5*		
25	DSSI2	Internship
26	DSSJ	Journal
27	DSSC	Conference
28	OOC3	MOOC

**-Indicates Subject code are aligned from the codes of the respective Department

Table C- AUDIT COURSES (Minimum 12 audit Credits should earn in a Post Graduate Course)				
Sr. No.	**Subject Code**	**Subject Name**	**Credits**	**Compulsory/ Optional**
1	## CBCS - CS	Cyber Security	4	Compulsory
2	## CBCS – HR	Human Rights	2	Compulsory
3	## CBCS – IIC	Introduction to the Indian Constitution	2	Compulsory
4	## SDC01 – BC	Business Communication Skills	4	Optional
5	## SDC02 – FP	Personal Financial Planning	2	Optional
6	## SDC03 – PLC	Programmable Logic Controller	2	Optional
7	## SDC04 – WD	Web Designing	2	Optional

Audit Courses Credits are not counted towards the CGPA calculation; whereas these courses are compulsorily to be completed towards the award of the Degree.

Notes:

1) Electives can also be Open Electives in spirit of DSCS.
2) Maximum 25% Open Electives are allowed.
3) Candidates are expected to perform minimum eight (8) assignments for every Lab Practice, and submit report as a bona fide document to course instructor. The assignment may be in the form of modeling/ simulation/ programming/ experimental investigation/ fieldwork.

4) MOOC Courses should be Board Specific.

5) **MOOC-**Open Online Course- Student is required to complete Online course through Coursera\NPTEL and other standard open Online Platforms. The course has to be pre- approved from the Department by the Course-Coordinator.

6) **Internship:** Students are encouraged to do Internship which will enable them towards state-of-art technologies and best practices followed by Industries. Internship Letter has to submitted to the course coordinator. Post Internship Presentation and report is to be submitted.

7) **Journal\Conference-** Students are expected to present their research findings in standard Research conferences and encouraged to publish in reputed Journals approved by Course Coordinator and Research Guide.

8) Students can do their Projects either in Industry or in academic Institution's\Research Lab. Students pursuing Projects in Industry cannot earn credits towards DSS3, DSS5 through Internship, they are encouraged to earn those credits through MOOCs\Journal\Conference.

9) **Exit Norms:** Student can exit the programme after one year of completion of earning 44 credits, and can opt for Post Graduate Diploma (PG Diploma) and exit the Degree Program as per NEP 2022.

Savitribai Phule Pune University

Department of Technology

SYLLABUS FOR M. Tech (Data Science)

DSC1: Python Programming

Course Description -

The course "Python Programming" provides a comprehensive introduction to the Python programming language and covers various topics, including Python basics, language concepts, object-oriented programming (OOP), exception handling, input/output (I/O) operations, regular expressions, data structures, GUI programming, CGI, and Python applications. Students will learn how to work with variables, data types, operators, and collections, as well as how to implement control statements and loops. The course also delves into OOP principles, file handling, modules, and regular expressions. Additionally, students will explore GUI development, CGI programming, networking fundamentals, and data processing with NumPy and pandas.

Course Objectives -

- Understand the fundamentals of the Python programming language and its historical development.
- Master Python basics, including data types, operators, collections, and string manipulation.
- Implement control statements and loops to make decisions and perform iterative tasks.
- Explore object-oriented programming (OOP) concepts and apply them to create reusable and modular code.
- Handle exceptions and errors effectively to ensure robustness in Python programs.
- Work with files and perform input/output (I/O) operations to read and write data.
- Learn regular expressions and use them for text processing and pattern matching.
- Gain proficiency in working with modules to modularize Python code and enhance reusability.
- Develop graphical user interfaces (GUI) using various widgets and events.
- Understand the basics of Common Gateway Interface (CGI) and create interactive web applications.
- Explore Python applications, including networking, serialization, and data processing with NumPy and pandas.

Course Outcomes -

Upon completion of the course, learners will be able to:

- Demonstrate a solid understanding of Python programming language and its syntax.
- Implement Python basics and effectively work with variables, data types, and collections.
- Apply control statements and loops to solve complex problems and execute iterative tasks.
- Develop object-oriented Python code, encapsulating data and behavior in classes and objects.
- Handle exceptions and errors gracefully to ensure smooth program execution.
- Perform input/output (I/O) operations to read from and write to files and streams.
- Utilize regular expressions for efficient text processing and pattern matching.

- Modularize Python code using modules and libraries to enhance code maintainability and reusability.
- Create graphical user interfaces (GUI) with various widgets and handle user events.
- Build interactive web applications using CGI in Python for data processing and form handling.
- Implement Python applications, including networking, serialization, and data processing with NumPy and pandas.

Unit 1. Python Basics

Introduction to programming languages, Python as a programming language, History of python, Python versions, Python installation, Environmental variables, Environmental variables in Windows operating system, Add python to Windows path, Executing python from the command line, Invoking python IDLE, Python documentation, Getting help, Dynamic types, Python reserved words, Naming conventions, Character set, Comments, Identifiers, Data types, Operators, Assigning values to variables, Type conversions, String methods, Simple output, Output formatting with "format", Simple input: Input function, Mutable vs immutable objects in python, Lists: Create and access, Lists: Modify and slice, Lists: Operations, Lists: Methods, Sets: Create and operations, Sets: Operators and methods, Sets: Frozenset, Sets: Methods, Tuples: Create and access, Tuples: Slice and alteration, Dictionaries: Create and access, Dictionaries: Modify and delete, Dictionaries: Methods, Copying collections: Shallow, Copying collections: Deep.

Unit 2. Language Concepts

Indenting requirements, Control statements, Decision making statement: If statement, If statement, If-else statement, Implementation, If-elif-else statement, Implementation, Decision making statement: Nested if-else statement, Nested if-else statement, Implementation, Iteration statements: While loop, While loop, Implementation, Iteration statements: While loop with else, While loop with else, Implementation, Iteration statements: For loop, For loop, Implementation, For loop range function with else, Implementation, For loop with object sequences, Nested for loops, Implementation, Break statement, Implementation, Continue statement, Implementation, Functions, Functions: Built in functions, Functions: User-defined functions, Calling a function, Returning a value from the function, Implementation, Scope or lifetime of variables, Creation and usage of global variables, Creation and usage of nonlocal variables, Passing collections to a function, Variable number of arguments, Implementation, Keyword arguments, Optional parameters, Default parameters, Nested functions, Recursive functions, Advantage and disadvantage, Passing functions to a function, map() function, filter() function, Lambda functions.

Unit 3. OOP, Exceptions and I/O

Object-oriented programming concepts, Class and object, Abstraction and encapsulation, Inheritance, Polymorphism, Classes in Python, Creating objects: Instance methods, Implementation, Memory management, Constructors, Constructors with parameters, Optional parameters in constructor, Deleting attributes and objects, Special methods, Class variables, Inheritance, Accessing base class element in derived class, Single inheritance, Multi-level inheritance, Multiple inheritance, Method Resolution Order (MRO), Access modifiers, Polymorphism, Operator overloading, Method overloading, Method overriding, Python errors: Syntax errors, Built-in exceptions, Exception handling, Simple exceptions, Multiple exceptions, Using else and finally, Raise an exception, Assert statement, Data streaming and buffering:

Serial data, I/O streams and buffers, Access modes, File open, File close, Exceptions in file, Writing to files, Reading from files, seek() and tell() methods, readline() and readlines() methods, Renaming and deleting files.

Unit 4. Modules and Regular Expressions

Modules, Modularization, Abstraction versus modularization, Modules in Python, Using modules in Python code, Import statement variances, Module search path, Loading and reloading of modules, dir() function, Python built-in modules: sys, Python built-in modules: math, Methods in module math, Python built-in modules: datetime, Python built-in modules: random, Regular expressions, Special sequences, Character classes in RegEx, Regular expression methods, Implementation, re.split(), Implementation, re.sub(), Implementation, re.subn(), re.search(), re.compile(), Implementation, Match object, Implementation, Raw string with “r” or “R” prefix, RegEx quantifiers, RegEx greedy match, RegEx non-greedy match, Greedy vs non-greedy match, RegEx flags.

Unit 5. Data Structures, GUI and CGI

Abstract data structures, Primitive data structure, Non-primitive data structures, List comprehensions, Accessing elements, Performing operations, Comprehension using If, Comprehension If-else, Nested list comprehensions, Dictionary comprehensions, Accessing elements, Performing operations, Comprehension using zip(), Comprehension for lambda functions, Nested dictionary comprehensions or dictionaries with compound values, Processing lists in parallel, Time functionality: Big O notation, Case scenarios, Time complexity in python collections, GUI in Python, Components and events, GUI example, Widgets, Attributes for widgets, Label widget, Button widget, Image button, Entry widget, Combo box widget, Check button widget, Radio button widget, Canvas widget, Frame widget, Message box widget, Message widget, Menu widget, Methods, Menu button, Implementation, CGI basics, Configuring CGI, HTML form elements, Form structure, Python and CGI program, CGI scripts, HTTP headers and environment variables, GET method, POST method, Complete CGI code, Advantages and disadvantages.

Unit 6. Python Applications

OS methods, Environment methods, Directory methods, File methods, Implementation, Implementation, Serialization and deserialization, The pickle module, Pickling and unpickling: Dictionary, Protocol formats, Picklable vs unpicklable types, Serializing attribute connections, Pickle exceptions, Networking fundamentals, Basic communication model, Network topologies, RING topology, STAR topology, MESH topology, TREE topology, Transmission mediums and modes, The client/server model, Client interaction with server, The socket module, The server-side, The client-side, Threaded server, Numpy: Overview, Numpy: Setup, Datatypes, Numpy: Basic operations, Binary operations, NumPy operations, Slicing and indexing, Broadcasting, Matrix operators, Iteration order in multi-dimensional array, Array values modification, External loop in multi-dimensional array, Broadcast iteration, Matplotlib, Plot methods in Matplotlib, Image design functions, Axis functions, Figure functions, Simple plots, Basic plots, Matplotlib options, Matplotlib: Multi plots, 3D plots, Data processing with pandas, Invalid value, Processing strings, Indexing and selecting data, iloc(): Integer based, Column name access, Group-by operations, Iterating through groups, Aggregation, Transformations, Filtration.

Reference books -

1. "Python Crash Course, 2nd Edition: A Hands-On, Project-Based Introduction to Programming" by Eric Matthes
2. "Learning Python, 5th Edition" by Mark Lutz
3. "Python Programming: An Introduction to Computer Science, 3rd Edition" by John Zelle
4. "Python Cookbook, 3rd Edition" by David Beazley and Brian K. Jones
5. "Automate the Boring Stuff with Python: Practical Programming for Total Beginners" by Al Sweigart
6. "Python for Data Analysis" by Wes McKinney
7. "Python GUI Programming Cookbook" by Burkhard A. Meier

DSC2: Applied Statistics

Course Description -

This course provides an in-depth exploration of advanced statistical analysis and data examination techniques. Students will gain a comprehensive understanding of measures of variability, probability distributions, hypothesis testing, relationship examination, and advanced data analysis methods. The course aims to equip students with the necessary skills to apply these advanced statistical techniques to real-world data and draw meaningful conclusions from the results.

Course Objectives -

- To familiarize students with advanced statistical measures, including skewness, variance, standard deviation, quartiles, percentiles, and their applications in data analysis.
- To introduce students to various probability distributions, such as the binomial, Poisson, normal, and standard normal distributions, and their significance in modeling real-world phenomena.
- To enable students to perform hypothesis tests and interpret the results to make data-driven decisions with confidence.
- To develop students' proficiency in examining relationships between variables using correlation analysis and regression models.
- To provide an in-depth understanding of advanced statistical techniques, including non-parametric tests, structural equation modeling, cluster analysis, factor analysis, and multidimensional scaling.
- To empower students with the ability to apply these advanced techniques to diverse data analysis scenarios and enhance their analytical skills.

Course Outcomes -

- Demonstrate a clear understanding of advanced statistical measures and their interpretation in data analysis.
- Apply probability distributions to model and analyze various real-world phenomena.
- Perform hypothesis tests using appropriate statistical methods and draw valid conclusions from the results.
- Analyze relationships between variables using correlation analysis and interpret the findings.
- Utilize advanced statistical techniques such as non-parametric tests, structural equation modeling, cluster analysis, factor analysis, and multidimensional scaling to solve complex data analysis challenges.

- Apply the acquired knowledge and skills to real-world datasets, drawing meaningful insights and making data-driven decisions.
- Communicate effectively about advanced statistical concepts and their applications to technical and non-technical stakeholders.

Unit 1. Measures of Variability

Skewness, Variance and Standard Deviation, Coefficient of Variation, Range, Quartiles and Interquartile Range, Percentiles and Percentile Rank.

Unit 2. Probability Distributions

Discrete Random Variables, The Binomial Distribution, The Poisson Distribution, Continuous Random Variables, The Normal Distribution, The Standard Normal Distribution, Finding Areas under a Normal Curve.

Unit 3. Testing Hypothesis

Hypothesis Testing, Errors arising in Hypothesis Testing, Steps to Perform a Test of Hypothesis using the p-value approach, The Chi-Square Distribution, The F – Distribution.

Unit 4. Examining Relationship

Pearson Correlation Coefficient, Hypothesis Test for a Correlation, Correlation Analysis, ANOVA Overview Simple Linear Regression Analysis, Error of Prediction, Limitations of Dummy & Interaction Terms.

Unit 5. Advanced Techniques

Non-Parametric Tests - Chi-Squared Goodness-of-Fit Test, Chi-Square Test of Independence, The Sign Test, Mann-Whitney Test, Kruskal-Wallis H-Test. Structural Equation Modeling (SEM), Cluster Analysis Background, Cluster Analysis, Hierarchical Clustering, Non-Hierarchical Method, Factor Analysis (FA), Centroid Method, Principal Components Method, Multidimensional Scaling (MDS).

Reference Books -

1. "Statistical Inference" by George Casella and Roger L. Berger
2. "Probability and Statistics for Engineering and the Sciences" by Jay L. Devore
3. "Introduction to the Practice of Statistics" by David S. Moore, George P. McCabe, and Bruce A. Craig
4. "Applied Multivariate Statistical Analysis" by Richard A. Johnson and Dean W. Wichern
5. "Statistics for Business and Economics" by Paul Newbold, William L. Carlson, and Betty Thorne
6. "Introduction to Mathematical Statistics" by Robert V. Hogg, Joseph W. McKean, and Allen T. Craig
7. "Discovering Statistics Using SPSS" by Andy Field, Jeremy Miles, and Zoe Field

DSC3: Data Mining & Predictive Modeling

Course Description -

This course focuses on the essential concepts and techniques of data understanding, preparation, and data mining in the context of business intelligence (BI). Students will learn how to extract valuable insights and knowledge from data to support decision-making processes. The course covers data exploration, cleaning, and transformation techniques, as well as various data mining methods, including clustering, classification, regression, and association rules. Students will also gain an understanding of web usage mining and its applications in e-business analytics. The course emphasizes hands-on experience with real-world datasets and data mining tools.

Course Objectives -

- Understand the importance of data understanding and preparation in the data mining process.
- Learn techniques for data exploration and verification of data quality, including outlier detection.
- Acquire knowledge of data cleaning methods and handling categorical variables.
- Explore various data transformation techniques and conduct univariate data analysis.
- Gain insights into the concepts and applications of different data mining methods, such as clustering, classification, and association rules.
- Familiarize with model evaluation methods, meta-level modeling, and model deployment in a business intelligence context.
- Develop skills in using data mining tools and applying advanced data mining techniques to solve real-world business problems.

Course Outcomes -

By the end of this course, students will be able to:

- Effectively prepare and clean data for data mining tasks.
- Analyze and explore data to identify patterns and outliers.
- Apply different data transformation techniques to enhance data quality and usefulness.
- Implement various data mining algorithms, such as decision trees, regression, and neural networks, to make informed business decisions.
- Evaluate the performance of data mining models using appropriate metrics and charts.
- Deploy data mining models in a business setting to support decision-making processes.
- Understand the applications of web usage mining and its relevance in e-business analytics.

Unit 1. Data Understanding and Preparation

Data preparation, Data understanding: Quality, Data exploration, Verification of data quality: Outlier detection, Data cleaning, Categorical variables: Ordinal variables, Data transformation, Univariate data analysis, Analysis of categorical data.

Unit 2. Model Development and Techniques

Data partitioning, Model selection, Simple Linear Regression (SLR) model, Clustering, Classification, Decision trees, Rule induction, Logistic regression, Support Vector Machine (SVM), Association rules, Bayesian networks, Neural network,

Unit 3. Model Evaluation and Deployment

Model evaluation, Automating models for categorical and continuous targets, Model evaluation methods: Confusion matrix, Lift charts, ROC, Area Under the Curve (AUC), Regression evaluation. Meta-level modeling, Model deployment. Bias, variance, and model complexity, Cross-validation, Bootstrap methods, Model selection criteria, Regression model building, Model validation.

Unit 4. Web Mining and Usage

Web usage mining: Simplified view, Web usage mining process, Sessionization strategies, User identification, Session uncertainty: Evaluate real vs. Re-constructed sessions, Pageview identification, Path completion, Data modeling for web usage mining, Integration with e-commerce events, Web usage and E-business analytics, OnLine Analytical Processing (OLAP), Data Mining purpose: Personalization, Standard approaches for personalizing/recommending, Suggest: Online recommender system, How does suggest work, Recommendation building.

Unit 5. Advanced Data Mining Techniques for BI

Data mining process and benefits, Data mining challenges and disadvantages, Data mining examples, Advanced data mining methods.

Reference Books -

1. "Data Mining: Concepts and Techniques" by Jiawei Han, Micheline Kamber, and Jian Pei.
2. "Introduction to Data Mining" by Pang-Ning Tan, Michael Steinbach, and Vipin Kumar.
3. "The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling" by Ralph Kimball and Margy Ross.
4. "Data Mining for Business Analytics: Concepts, Techniques, and Applications in R" by Galit Shmueli, Peter C. Bruce, Inbal Yahav, and Nitin R. Patel.
5. "Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data" by Bing Liu.
6. "Data Science for Business: What You Need to Know about Data Mining and Data-Analytic Thinking" by Foster Provost and Tom Fawcett.
7. "Machine Learning: A Probabilistic Perspective" by Kevin P. Murphy.

DSC4: Business Intelligence

Course Description -

This course provides an in-depth exploration of Business Intelligence (BI) and its various components. Students will learn about BI architecture, functional areas of BI tools, and their applications in different industries. The course delves into data warehousing and decision support, covering topics such as data warehouse properties, schemas, OLAP server architectures, and real-time monitoring. Students will gain practical knowledge in building BI projects, including the planning process, engagement activities, and proof of concept development. Additionally, the course introduces advanced data mining techniques and their applications in BI.

Course Objectives:

- Understand the fundamental concepts of Business Intelligence (BI) and its architecture, including the components that make up a BI system.
- Explore the various functional areas where BI tools can be applied and identify their specific applications in different industries and domains.
- Analyze the capabilities of enterprise analytics and discern between complex and volume operations in data processing.
- Recognize the importance of fostering a fact-based culture and collaboration between IT and business teams for effective BI implementation.
- Evaluate the significance of a robust data infrastructure in supporting BI initiatives and enabling data-driven decision-making.
- Gain insights into the role of the analytics workforce and its impact on driving successful BI projects.
- Examine the properties of data warehouses, including multidimensional data and different schema types (Star, Snowflake, Constellation).
- Develop an understanding of the data warehouse design process and the architectures of OLAP servers, along with their application in business queries, dashboards, and scorecards development.
- Learn about advanced data mining techniques used in BI, including the data mining process, challenges, benefits, and practical examples of applying data mining methods to extract valuable insights from data.

Course Outcomes -

By the end of the course, students will be able to:

- Comprehend the architecture and components of Business Intelligence systems.
- Apply BI tools and techniques in different functional areas and industry contexts.
- Analyze complex vs. volume operations and make data-driven decisions based on analytics.
- Design and develop data warehouses using various schema models.
- Implement OLAP server architectures and build interactive dashboards and scorecards.
- Execute BI projects using best practices and measure their success and value.
- Utilize advanced data mining methods to discover patterns and trends in large datasets.

Unit 1. Business Intelligence Architecture and Components

BI architecture & components, Functional areas of BI tool, Applications of BI tool, Enterprise analytics capabilities, Complex vs volume operations, IT vs business and fact-based culture, A strong data infrastructure, Analytics workforce.

Unit 2. Data Warehousing and Decision Support

Data warehouse properties, Multidimensional data, Schemas (Star, Snowflake, Constellation), Conceptual modeling of data warehouses, Data warehouse design process, OLAP server architectures, Business query, Dashboards, and scorecards development, Metadata model, Mobile BI, Disconnected BI, Collaborative BI. Real-time monitoring, Making BI easy to consume.

Unit 3. Building a BI Project

BI maturity model, Measuring BI success and value, Five key areas of strategy, Planning a BI Project, Engagement process, Project tasks and milestones, Proof of Concept (POC), Table creation, OLAP creation, Server administration data measurement, Knowledge Discovery in a database (KDD), The Cross Industry Standard Process for Data Mining (CRISP-DM)

Unit 4. Building Reports and Dashboards

Metrics and their impact on behavior, Balanced scorecard and performance dashboards, Dashboard design principles, Sales simulation and what-if scenarios, Report building techniques, Charts and chart types, Drill-down and drill-through in reports, Report scheduling and on-demand running.

Unit 5. Advanced Data Mining Techniques for BI

Data mining process and benefits, Data mining challenges and disadvantages, Data mining examples, Advanced data mining methods.

Reference Books -

1. "Business Intelligence Guidebook: From Data Integration to Analytics" by Rick Sherman
2. "Data Warehousing, Data Mining, and OLAP" by Alex Berson and Stephen J. Smith
3. "Business Intelligence Roadmap: The Complete Project Lifecycle for Decision-Support Applications" by Larissa T. Moss and Shaku Atre
4. "The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling" by Ralph Kimball and Margy Ross
5. "Performance Dashboards: Measuring, Monitoring, and Managing Your Business" by Wayne W. Eckerson
6. "Data Mining: Concepts and Techniques" by Jiawei Han, Micheline Kamber, and Jian Pei
7. "Big Data, Data Mining, and Machine Learning: Value Creation for Business Leaders and Practitioners" by Jared Dean and Paul C. Evans

DSC5: MACHINE LEARNING

Course Description -

Machine Learning course provides a comprehensive exploration of advanced topics in statistical analysis, and data analysis techniques. It covers measures of variability, probability distributions, hypothesis testing, correlation and regression analysis, classification algorithms, clustering techniques, and information retrieval. Students will gain hands-on experience in applying these methods to real-world datasets and understanding their significance in solving complex problems. The course also focuses on model building, model validation, and selecting appropriate machine learning algorithms for different types of data.

Course Objectives -

- Understand the measures of variability and their significance in data analysis.
- Explore various probability distributions and apply them to model real-world data.
- Gain proficiency in hypothesis testing and use it to make data-driven decisions.
- Develop an in-depth understanding of correlation and regression analysis for predictive modeling.
- Learn about different classification algorithms and ensemble techniques for solving classification problems.
- Master clustering techniques and their applications in data grouping and pattern recognition.
- Familiarize themselves with information retrieval methods and text document clustering.

Course Outcomes -

- Apply measures of variability to analyze and interpret data distributions.
- Model and analyze real-world data using probability distributions.
- Perform hypothesis tests and draw conclusions based on statistical significance.
- Build and validate regression models to make accurate predictions.
- Implement classification algorithms and evaluate their performance for different applications.
- Utilize clustering techniques to group data and identify patterns in diverse datasets.
- Demonstrate proficiency in information retrieval methods and text document clustering for text analysis.

Unit 1. Simple Linear Regression

Supervised Learning, Regression and its Examples, Simple Linear Regression, Least Squares Estimation, Coefficient of Determination (R-squared), Testing Hypotheses in Simple Linear Regression.

Unit 2. Multiple Regression and Model Building

Ordinary Least Squares Estimation for Multiple Linear Regression, Multiple Linear Regression Model Building, Partial Correlation and Regression Model Building, Interpretation of Multiple Linear Regression Coefficients, Coefficient of Multiple Determination (R-Squared), Adjusted R-Squared, Statistical Significance of Individual Variables in Multiple Linear Regression.

Unit 3. Classification & Classification Algorithms

Machine Learning Classifiers, Classification Algorithms, Instance-Based Learning and K-Nearest Neighbor (KNN), Decision Trees and ID3 Algorithm, Bayesian Algorithms and Ensemble Techniques, Neural Networks and Support Vector Machine (SVM), Classification Model Evaluation and Selection: ROC Curves, Cost-Benefit Analysis.

Unit 4. Clustering Techniques

Clustering Algorithms and General Applications, Hierarchical Clustering, K-Means Clustering and Other Partitioning Methods, Density-Based Clustering: DBSCAN, External Criteria for Clustering Quality, Internal Measures for Cluster Validation, Information Retrieval: Introduction, Models, Term Weighting, Retrieval in Vector Space Model.

Unit 5. Text Document Clustering

Text Document Clustering: Agglomerative vs. Divisive, Cluster Distance Measures and Impact on Clustering, Internal Measures for Cluster Validity: SSE, Cohesion, Separation, Silhouette Coefficient. Construction of Inverted Index for Information Retrieval, Text Preprocessing: Stop Words Removal and Stemming.

Reference Books -

1. "Introduction to Statistical Learning" by Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani
2. "Pattern Recognition and Machine Learning" by Christopher M. Bishop
3. "The Elements of Statistical Learning: Data Mining, Inference, and Prediction" by Trevor Hastie, Robert Tibshirani, Jerome Friedman
4. "Data Science for Business" by Foster Provost, Tom Fawcett
5. "Machine Learning: A Probabilistic Perspective" by Kevin P. Murphy
6. "Introduction to Data Mining" by Pang-Ning Tan, Michael Steinbach, Vipin Kumar
7. "Information Retrieval: Implementing and Evaluating Search Engines" by Stefan Büttcher, Charles L. A. Clarke, and Gordon V. Cormack

DSC6: Deep Learning

Course Description -

This course provides an in-depth exploration of deep learning and neural networks, focusing on their capabilities and relevance in various applications. Students will learn about different learning tasks in neural networks, including supervised and unsupervised learning, and delve into memory-based learning techniques. The backpropagation algorithm and learning curves will be covered, along with the fundamental concepts of Perceptron and Multi-Layer Perceptron (MLP). The course also investigates generalization and approximations of functions in deep learning models. Additionally, students will gain hands-on experience with Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs) with Long Short-Term Memory (LSTM), and Generative Deep Learning techniques, including Generative Adversarial Networks (GANs).

Course Objectives -

- Understand the capabilities and significance of deep learning in various real-world applications.
- Identify and differentiate between different learning tasks in neural networks, such as supervised and unsupervised learning.
- Comprehend memory-based learning techniques and their relevance in deep learning models.
- Implement and analyze the backpropagation algorithm and learning curves for neural networks.
- Grasp the fundamentals of Perceptron, MLP, and heuristics for enhancing backpropagation.
- Investigate generalization and approximations of functions within deep learning models.
- Explore Convolutional Neural Networks (CNNs) and their applications in image processing and computer vision.
- Develop expertise in Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) for sequence modeling tasks.
- Apply Generative Deep Learning techniques, including text synthesis, image synthesis, and neural style transfer.

Course Outcomes -

By the end of this course, students will be able to:

- Apply deep learning algorithms and neural network models to solve various real-world problems.
- Evaluate and select appropriate learning tasks for specific applications.
- Implement memory-based learning techniques to enhance model performance.
- Analyze and optimize the backpropagation algorithm for efficient neural network training.
- Design and build complex neural network architectures, including Perceptron and MLP.
- Evaluate the generalization and approximation capabilities of deep learning models.
- Develop and train Convolutional Neural Networks (CNNs) for image recognition tasks.
- Create and deploy Recurrent Neural Networks (RNNs) and LSTMs for sequential data analysis.
- Utilize Generative Deep Learning techniques to generate text, images, and perform style transfer.

Unit 1. Deep Learning and Neural Networks

Deep learning capabilities and relevance, Learning tasks in neural networks (Supervised, Unsupervised, etc.), Memory-based learning techniques, Backpropagation algorithm and learning curves, Perceptron, MLP concepts, and Heuristics for backpropagation. Generalization and Approximations of functions.

Unit 2. Convolutional Neural Networks (CNN)

Introduction to Convolutional Neural Networks, Visualization of 2D and 3D convolutions, Training CNN on datasets and data preprocessing, Data augmentation and feature extraction in CNN, Visualizing CNN filters and intermediate representations.

Unit 3. Recurrent Neural Networks (RNN) and LSTM

Introduction to Recurrent Neural Networks and LSTMs, Core idea and working of LSTMs and GRUs, Backpropagation through timeline in RNN, Over-fitting and prevention techniques, Stacked LSTM and Multi-directional RNNs.

Unit 4. Generative Deep Learning

Introduction to Generative Deep Learning, Using LSTMs for text synthesis, Neural style transfer and image synthesis with VAE, Generative Adversarial Networks (GANs) and working principle, Training GAN and GAN applications.

Unit 5. Advanced Topics in Deep Learning

Auto-Encoder with SISO modeling, Reinforcement learning and Q learning, Directed acyclic graphs and cyclic graphs in neural networks, Hyperparameter tuning and optimization, Ensemble modeling techniques (Bagging, Boosting) and their advantages.

Reference Books -

1. "Deep Learning" by Ian Goodfellow, Yoshua Bengio, and Aaron Courville
2. "Neural Networks and Deep Learning: A Textbook" by Charu Aggarwal
3. "Python Deep Learning" by Ivan Vasilev, Daniel Slater, and Gianmario Spacagna
4. "Recurrent Neural Networks with Python Quick Start Guide" by N.D Lewis
5. "Generative Deep Learning: Teaching Machines to Paint, Write, Compose, and Play" by David Foster
6. "Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow" by Aurélien Géron
7. "Deep Learning for Computer Vision" by Rajalingappaa Shanmugamani

DSE1: MATHEMATICS FOR TECHNOLOGY

Unit 1: Linear Algebra

Matrices, Properties of matrix addition, Various types of matrices, Property of adjoint matrix, Inverse of a matrix, Solution of simultaneous equations, Inverse of a matrix by elementary transformations, Rank of a marix, Consistency of linear equations, Homogeneous equations, Linear dependence and independence of vectors, Eigen values and eigen vectors, Cayley Hamilton Theorem, Properties of eigenvalues and eigen vectors, Non-symmetric matrices with non-repeated eigen values, Non-symmetric matrices with repeated eigen values, Diagonalization of matrix, Powers of matrix, Sylvester's Theorem, Quadratic form, Differentiation and Integration of Matrices.

Solution of Nonlinear Algebraic Equations: Relative advantages of methods and Engineering Applications

Unit 2 : Numerical methods

Newton-Raphson method or Succesive substitution method, Rule of False position (Regula False method), Iteration method, Solution of linear system, Crout's method, Jacobi's method, Gauss-Seidal method, Solution of ordinary differential equations, Taylor series method, Picard's method for successive approximation, Euler's method, Euler's modified formula., Higher order differential equations.

Solution of Industry relevant ODEs both in IVP and BVP domains, Solution of Partial Differential Equations, Finite Difference Technique.

Unit 3: Transforms Techniques

Laplace Transform definition, First shifting theorem, second shifting theorem, change of scale property, Laplace transform of the derivative, Initial value theorem, Final value theorem, Laplace transform of integrals, Multiplication by t, Division by t, Inverse Laplace transform definition, First shifting theorem, second shifting theorem, change of scale property, Use of partial fractions, Inverse Laplace transform of derivatives, Inverse Laplace transform of integrals, Multiplication by powers of s, division by powers of s, Convolution theorem, Heaviside's expansion formula, Applications of Laplace transform to solutions of differential equations.

Fourier Transform and their practical applications in Engineering, Z – Transforms and its applications in Control Aspects and Electronics

Reference Books

1. Higher Engineering Mathematics by B. S. Grewal (Khanna Publication, Delhi).

2. Venkatraman, M.K., Numerical Methods in Science and Engineering, National Publisher Company.

3. Shanthakumar M.S., Numerical Methods & Analysis

4. Gupta, S.C. and Kapur, V.K., "Fundamentals of Mathematical Statistics ", Sultan Chand & Sons, New Delhi, 1999.

DSE2: Introduction to Data Analytics

Course Description -

The Dat Analytics course offers a dynamic exploration of the field, providing valuable insights into its applications and benefits in today's business landscape. Students will understand the challenges of data creation and the pivotal role Business Analytics plays in maximizing profitability and informed decision-making. Through in-depth study of data warehousing, BI reporting, and data mining, students will learn to harness valuable information for strategic business strategies. The course also focuses on leveraging powerful data visualization tools like Tableau to create compelling data-driven stories and insightful dashboards. Additionally, students will uncover the fundamentals of Big Data Analytics and its growing importance in the interconnected world. Industry use cases and case studies will be explored to develop practical knowledge and proficiency in various statistical analysis techniques. By the end of this comprehensive course, students will be equipped to excel in the field of Business Analytics, gaining a competitive edge in the world of data-driven decision-making.

Course Objectives -

- Introduce students to the concepts and principles of Business Analytics, its historical evolution, and its significance in modern organizations.
- Familiarize students with data warehousing, business intelligence, and data mining techniques to derive meaningful insights from large datasets.
- Develop students' skills in using data visualization tools like Tableau to create impactful dashboards and reports for effective data storytelling.
- Provide an overview of Big Data Analytics, its characteristics, benefits, and key application frameworks in today's data-rich environment.
- Enable students to apply statistical analysis techniques, such as inferential statistics, cluster analysis, and market basket analysis, to real-world business scenarios.
- Present industry use cases and case studies to demonstrate the practical applications of Business Analytics across diverse sectors.

Course Outcomes -

- Gain a comprehensive understanding of Business Analytics and its role in enhancing business performance and decision-making.
- Demonstrate proficiency in data warehousing, BI reporting, and data mining techniques to extract valuable insights from complex datasets.
- Develop the ability to create visually appealing and informative dashboards using Tableau for effective data storytelling.
- Acquire insights into the field of Big Data Analytics and its growing significance in managing and analyzing large volumes of data.
- Master various statistical analysis techniques to draw meaningful conclusions and identify patterns in business data.
- Apply Business Analytics concepts and methods to industry use cases, providing valuable solutions and recommendations for real-world challenges.

Unit 1. Introduction to Business Analytics

Introduction to Business Analytics and Optimization, Challenges of data creation, Approaches to help maximize profitability and returns, what are Business Analytics? Business Analytics capabilities, Enterprise Analytics capabilities, Business Analytics technologies, Predictive Analytics, Prescriptive Analytics, A fact-based decision-making culture, A strong data infrastructure, the right analytical tools, New tools and architectures may be needed, Analytics workforce, Knowledge requirements, Business Analyst, Data Scientist, where to put the Analytics team? IBM Business Analytics maturity model, Optimization, Key BAO concepts, The need for BAO now, Essential capabilities in BAO, BAO capabilities: Business performance management, BAO capabilities: Predictive analysis and mining, Value of BAO to business organization, Impact of BAO on diverse industries, Advantages to implementing BAO solutions, BAO capabilities: real-time Analytics: Data in motion, BAO support for decision-making, High level architecture of BAO, Importance of reference architecture, BAO reference architecture, BAO reference architecture to BAO architects, IBM technology portfolio for BAO.

Unit 2. Data Warehouse

Decision support, Three-tier decision support systems, Exploring and analyzing data, what is a data warehouse? Data warehouse architecture choices, Enterprise data warehouse, Independent data mart architecture, Dependent data mart architecture, Data warehouse, Multidimensional data, Conceptual modeling of data warehouses, Data warehouse design process, Single-layer architecture, Two-layer architecture, Three-tier data warehouse architecture, Data warehouse development, Multi-tiered architecture, Information pyramid, BI reporting tool architectures, Multidimensional analysis techniques, Data analysis and OLAP, OLAP server architectures, Data cube, OLTP vs OLAP, Business query, Dashboards and scorecards development, Metadata model, Automated tasks and events, Mobile BI, Software development kit (SDK), Setting up data for BI, Making BI easy to consume.

Unit 3. Business Intelligence

What is business intelligence? Definitions of business intelligence, Sample BI architecture, Things are getting more complex, BI components and architecture, Scope and fit of BI solutions within existing infrastructure, High level BI process, A single or a few applications, Benefits of BI, Maximize value from BI systems, Strategy and business intelligence, Business transformation projects, Business role of BI (TWDI), ASUG business intelligence maturity model, Why act? BI effectiveness scorecard, BI value scorecard, five key areas of strategy, Planning a BI project, Pre-engagement activities, Engagement activities and process, BI design and development, Business environment, Project tasks, Task 1: Knowledge capture goals
Discuss business objectives & prior learning, Interview key stakeholders, Project planning
Task 2: Consolidate findings
Create logical design
Task 3: Map the customer situation
Current environment, Business/functional requirements sample diagram, Logical BI diagram
Task 4: Methodology & approach
Task 5: Standards & governance
Task 6: Sections, milestones, and tasks
Task 7: Proof of Concept (POC)
Task 8: Table creation

Task 9: OLAP creation
Task 10: Final deliverables
Risk management and mitigation, Cost justification and measuring success

Unit 4. Data Mining

What is Data Mining? Evolution of Data Mining, Why Data Mining and data Analytics? Knowledge based system, Data Mining process, Phases of Data Mining process, Data Mining process, KDD process model, CRISP – DM, CRISP DM, CRISP-DM elaborate view, Data Mining on what kinds of data? DM tasks and components of DM methods, Data Mining operations, Data Mining techniques, Data Mining techniques, Industry examples of application of DM, Challenges of Data Mining, why machine should "learn"? What is machine learning? Growth of machine learning, Machine learning types, Unsupervised learning, Reinforcement learning.

Unit 5. Data Storytelling with Tableau

Definition of dashboard, Dashboard types, Layers of information, Evolution of dashboards, Dashboard design, Dashboard design principles, Chart overview, Singular components, Metrics, Kaplan-Norton balanced scorecard, The Ray port-Jaworski performance dashboard and strategy framework, Introducing the R-J performance dashboard, Blueprint to the R-J performance dashboard, Building reports, List report, Data group, sort and filters, Add calculations to report, Conditions and aggregations in report, Drilling in report, Run report: On demand or schedule, Charts, Chart type: Bar chart, What is data visualization? Data visualization tools, what is tableau? Tableau features, Tableau products, Tableau architecture, an introduction to tableau workspace, Tableau charts, How to post tableau service dashboards? Table in tableau computation, Basic table calculation functions, Quick filters.

Unit 6. Introduction to Big Data Analytics

What is Big Data? Intrinsic property of data it grows, A growing interconnected and instrumental world, Need for Big Data, Characteristics of Big Data, Structure of Big Data and need for standards, Big Data Analytics adoption, Benefits & barrier of Big Data Analytics, Trends for Big Data Analytics, Commoditization of hardware enabling new Analytics, The 5 key Big Data use cases, More ways wide ranging Analytics and techniques, Big Data platform and application frameworks, A Big Data platform manifesto, Use cases for a Big Data platform, Data science.

Unit 7. Industry Use Cases for Business Analytics

Descriptive statistics, Inferential statistics, Univariate analysis, Factor analysis, Cluster analysis, Density-based clustering, Variance analysis, Discriminatory analysis, Multidimensional scaling, Collaboration methods, K-means grouping, Redundancy analysis, Market basket analysis, Introduction to RFM and Capstone project, Recency, frequency, and monetary analysis.

Reference books -

1. "Business Analytics: Data Analysis & Decision Making" by S. Christian Albright and Wayne L. Winston
2. "Data Warehousing, Data Mining, and OLAP" by Alex Berson, Stephen J. Smith, and Kurt Thearling

3. "Business Intelligence Guidebook: From Data Integration to Analytics" by Rick Sherman
4. "Data Science for Business" by Foster Provost and Tom Fawcett
5. "Big Data: A Revolution That Will Transform How We Live, Work, and Think" by Viktor Mayer-Schönberger and Kenneth Cukier
6. "Tableau Your Data!: Fast and Easy Visual Analysis with Tableau Software" by Daniel G. Murray
7. "Predictive Analytics: The Power to Predict Who Will Click, Buy, Lie, or Die" by Eric Siegel

DSE3: Data Warehousing and Multidimensional Modelling

Course Description -

The course on "Data Warehousing and Multidimensional Modeling" provides a comprehensive understanding of data warehousing concepts, methodologies, and techniques for modeling data in multidimensional structures. Students will learn about data warehouse architectures, ETL (Extract, Transform, Load) processes, and decision support applications. The course explores various data warehouse modeling approaches, including ER modeling and normalization techniques. Additionally, students will gain insights into multi-dimensional modeling methodologies, such as star and snowflake schemas, OLAP (OnLine Analytical Processing) techniques, and R-OLAP (Relational OLAP) and MOLAP (Multidimensional OLAP) architectures. The course also covers temporal modeling and how to design historical data tracking in a data warehouse.

Course Objectives -

- To introduce students to the fundamental concepts and principles of data warehousing and its importance in decision support systems.
- To provide students with a comprehensive understanding of data warehouse modeling techniques, including multi-dimensional modeling and normalization.
- To familiarize students with different data warehouse architectures, such as two-layer and three-layer architectures, and their implementation.
- To explore R-OLAP and MOLAP architectures, including their characteristics and applications in online data analysis.
- To understand temporal modeling and the methods for keeping historical data in data warehousing.
- To provide students with practical knowledge of data warehouse tools and data mining and reporting tools used in the industry.

Course Outcomes -

By the end of the course, students should be able to:

- Explain the concepts and significance of data warehousing and its role in decision support systems.
- Develop data warehouse models using ER modeling, normalization techniques, and multi-dimensional modeling approaches.
- Design and implement various data warehouse architectures, such as centralized, distributed, and peer-to-peer architectures.

- Analyze and compare R-OLAP and MOLAP architectures and understand their applications in online data analysis.
- Apply temporal modeling techniques to track historical data and manage changes in dimensions over time.
- Utilize data warehouse tools and data mining/reporting tools effectively for data analysis and reporting.
- Apply the knowledge gained to real-world scenarios and case studies in data warehousing, such as the IBM data warehouse case study for American Airlines.

Unit 1. Introduction to Data Warehousing

An introduction to data warehousing, Data warehouse architectures, Two layer architecture, Three layer architecture, Data staging and ETL, Decision support applications, Data warehouse functional definitions.

Unit 2. Data Warehousing and Modeling

Data warehouse & modeling, Data warehouse modeling, Types of data warehouse models, Data warehouse implementation, Normal system for metadata sharing, Data warehouse environment, Data warehouse and data mart, Operational data store, Data warehouse data modeling styles, OLTP data model considerations, Data warehouse usage, Data mart, Warehouse modeling approaches, Global data warehouse architecture, Independent data mart architecture, Approaches to implement the architecture, Data warehouse delivery process, Data warehouse modelling: Techniques and guidelines, ER modeling, The ER model, Normalization: First normal form, Normalization: Second normal form, Normalization: Third normal form, OLAP (OnLine Analytical Processing), Multi-dimensional data analysis, Drill-down and roll-up.

Unit 3. Multi - Dimensional Modeling - Methodology

Requirements analysis: Base activities, Representation of the query as a cube, Requirements modeling, Measures, Candidate measures, Dimensions, The grain of a dimension, Granularity of a measure, Facts, Terminologies in a multi-dimension model, Advantage of multi-dimensional modelling, Multi-dimensional queries, Dimension entities and attributes, Dimension hierarchies, Aggregation levels, MD model structures, Star schema, Snowflake schema, Example for snowflake schema, Considerations star and snowflake, Hints, tips, and guidelines, Business-related facts, What is fact constellation schema? Applications for data warehouse, Fact identifiers, dimension keys and uniqueness, Identifiers in an MD model, Dimension roles, Requirements analysis, Solution validation, Base activities, Detailed dimension modeling, Why needed? Detailed dimension modeling– approach, Detailed dimension modeling– considerations.

Unit 4. Non-Temporal Design - R-OLAP

R-OLAP, OLAP query against RDBMS (ROLAP), R-OLAP design techniques, OLAP, Fundamental OLAP analytical practices, Types of OLAP systems, Architecture of ROLAP, Star schema and star join queries, Star join support, Efficient star join, Dynamic bitmap index ANDing –DB2 UDB for distributed platforms, Star join: DB2 UDB for distributed platforms, Star join: Snowflake schema context, Collapse dimensions, Effects of dimension collapse, Dimension split, Redundant attributes in facts, Fact-oriented design techniques, Fact aggregation, Consolidated facts, Utilize cubing services to improve R-OLAP and M-OLAP performance, Cubing services architecture, Cubing services performance and scalability, Scalability, Cubing services

security, Benefits of cubing services: Improved data mart performance, Data marts, Data marts: Materialized query tables, Materialized query table example, Performance with MQT refresh options, Implementation example: MQT, Multi-dimension cluster tables, MDC table example, Terminology-dimension, Terminology -slice, Terminology-cell, Block based index: Dynamic bitmap index ANDing, Query processing examples, MDC performance example, Example: Object size comparisons, Power of block prefetch coupled, Performance results for a query, Performance results from a query, Example: Index oring of block and RID indexes, Example: Joins with block index.

Unit 5. Non-Temporal Design - MOLAP

Introduction to MOLAP, MOLAP framework, M-OLAP architecture, OLAP schemes types, Cognos functional architecture, Cognos Transformer: Functions, Storage method, Hyperion, Hyperion data load, Hyperion: data load rules, Hyperion calculation storage types, MOLAP characteristics, ROLAP characteristics, Online data analysis.

Unit 6. Temporal Modeling and Design

Keeping history about dimensions, Customer attrition: Business requirements, Customer attrition: Exercise, Customer attrition: Base analysis, Customer attrition: Solution, Primary key change, The historical customer, Customer address: Business requirement, Customer address: Exercise, Customer address: Solution, Issues, Time normalization, Customer address: Time normalized, Neighborhoods: Business requirements, Neighborhoods: Exercise, Neighborhoods: Base solution, Neighborhoods: Review, Neighborhoods: HTK solution, Customer classes: Business requirement, Customer classes: Exercise, Customer classes: Proposed solution.

Unit 7. Process Architecture: Data Warehouse

Process architecture: Data warehouse, Contrasting OLTP and data warehousing environments, Common data warehouse tasks, Architectures of the data warehouse, Centralized process architecture, Distributed process architecture, Peer-to-Peer architecture, Data warehouse tools, 25 BEST data mining tools in 2020, 18 BEST reporting tools in 2020, IBM data warehouse case study: American Airlines.

Reference books -

1. "The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling" by Ralph Kimball and Margy Ross
2. "Building the Data Warehouse" by W. H. Inmon, Claudia Imhoff, and Ryan Sousa
3. "Data Warehousing Fundamentals: A Comprehensive Guide for IT Professionals" by Paulraj Ponniah
4. "The Data Warehouse ETL Toolkit: Practical Techniques for Extracting, Cleaning, Conforming, and Delivering Data" by Ralph Kimball, Joe Caserta, and Bill Inmon
5. "Data Warehouse Design: Modern Principles and Methodologies" by Matteo Golfarelli and Stefano Rizzi
6. "OLAP Solutions: Building Multidimensional Information Systems" by Erik Thomsen and Christopher Adamson
7. "Data Warehousing in the Age of Big Data" by Krish Krishnan

DSE4: Big Data Analytics

Course Description -

The course on "Big Data Analytics" provides an in-depth understanding of the concepts, tools, and technologies associated with handling and analyzing large volumes of data known as Big Data. The course covers various aspects, including the introduction to Big Data, Hadoop fundamentals, query languages like Hive and Pig for Hadoop reporting and analysis, data visualizations, NoSQL databases, and the integration of R programming language with Hadoop for scalable analytics. Students will gain practical knowledge of using Hadoop's ecosystem, exploring big data sources, implementing data processing and analysis techniques, and visualizing data effectively.

Course Objective -

The main objective of the course is to equip students with the essential skills and knowledge required to work with Big Data effectively. By the end of the course, students will:

- Understand the fundamentals of Big Data, its characteristics, and its significance in various industries.
- Gain proficiency in Hadoop, including HDFS, MapReduce, HBase, and the integration of Hadoop with Hive and Pig for reporting and analysis.
- Learn to write and execute queries in query languages like Hive and Pig for data processing and manipulation in Hadoop.
- Acquire knowledge of NoSQL databases, with a focus on MongoDB, and its advantages in handling unstructured data.
- Develop skills in data visualization techniques, understanding various chart types, and creating compelling visualizations for Big Data analysis.
- Explore the integration of open-source R programming with Hadoop to perform scalable machine learning and analytics tasks on large datasets.

Course Outcome -

Upon successful completion of the course, students will be able to:

- Analyze and manage large-scale datasets using Hadoop and its ecosystem tools.
- Write and execute queries in query languages like Hive and Pig for data processing and analysis in Hadoop.
- Apply NoSQL databases like MongoDB to handle unstructured data and perform efficient data modeling.
- Create effective data visualizations using various chart types to present insights from Big Data.
- Utilize open-source R programming for scalable machine learning and analytics on big datasets.
- Apply the knowledge gained to solve real-world Big Data challenges in different domains such as healthcare, finance, and social media.

Unit 1. Introduction to Big Data Analytics

Big Data overview, Structures of data, Big Data growth story, Big Data sources, Big Data adoption drivers, Need of Big Data, Growth drivers for IT industry, Big Data: Definition, Characteristics of Big Data, Units to measure Big Data, Big Data types, Benefits & barrier of Big Data analytics, Need of Big Data, Big Data process, Big Data framework, Big Data platform and application frameworks, An example of Big Data platform in practice, A Big Data platform manifesto, Big Data technologies, Big Data tools, Big Data &

analytics, Merging the traditional and Big Data approaches, More ways: Wide ranging analytics and techniques, The 5 key Big Data use cases, Big Data usage, Use cases, Big Data and complexity in health care, Use cases: Healthcare and life sciences, Use cases: Transportation services, Use cases: Life insurance, IBM's Big Data success story, Data repository analyst view, Business drivers with examples, BI versus data studies, Architecture of modern analytics, Large data drivers, Method to study the evolving Big Data environment, Latest Big Data ecosystem, Large data research explanations, Overview of lifecycle data processing, Major functions for a good research, Information analytics history and summary, Big Data resources, Tackling the question, Primary stakeholders recognition, Analytical supporter interview, Original assumptions creation, Potential databases detection, Preparing evidence for Big Databases, Analytical sandbox planning, ETL execution, The data learning, Data conditioning, Visualize and verification, Popular data preparation instruments, Step 3: Planning of models, Research on model planning in industry verticals, Data exploration, Big Data: Choice of model, Popular model of Big Data, Step 4: Model design, Popular model building tools, Step 5: Results contact, Step 6: Consumption, Key outputs from a successful analytics project, Case study: Digital network for creativity and research (GINA), Step 1: Searching, Step 2: Preparing evidence, Step 3: Planning of models, Step 4: Model design, Step 5: Results contact, Step 6: Consumption.

Unit 2. Hadoop Fundamentals

What is Hadoop? Examples of Hadoop in action: IBM Watson, Examples of Hadoop in action, Introduction to Hadoop, Data distribution, Flat scalability, HDFS (Hadoop Distributed File System), Name nodes, Data nodes, Data nodes with blocks of multiple files with a replica of 2, The data is distributed across nodes at the time of loading, MapReduce, An SQL example of MapReduce, The map function, Sort phase, The reduce function, Combiner and partition functions, Streaming and pipes, MapReduce example: Wordcount, MapReduce co-locating with HDFS, MapReduce processing, Speculative execution, MapReduce: A tale of two APIs, MapReduce anatomy, What is HBase? NoSQL technology, CAP theorem, ACID properties, Why HBase? Important things to keep in mind, HBase vs. RDBMS, For example, Physical view in HBase, Logical to physical view, HBase components, HBase components definitions and roles, Characteristics of HBase tables, HBase is a sorted multidimensional map, Row key design considerations, Column family design considerations, Cluster configuration, HDFS configurations settings, Hadoop site.xml for a single-node configuration, HDFS start, Interact with HDFS, Example of put command, Retrieve data from HDFS, HDFS command reference, HDFS permissions and security, HDFS additional tasks: Rebalancing blocks, Closing the nodes, Check health file system, Load scenarios, Load solution using Flume, How Flume works, Consolidation, Replicating and multiplexing, Apache Hadoop core components, Why Hadoop Apache? Why is it easier for Hadoop than other distributed computing systems? Where is Hadoop ideal for computing? Daemons of HDFS, Secondary name node, Check-pointing by secondary name node, HDFS architecture, Daemons MapReduce, YARN capital assignments, Node manager with three containers, Resource allocation on another node, Running tasks on m multiple containers, Resource allocation, Fair scheduler per-queue properties, The map reducing job workflow, HDFS daemons with strong disponibility, Height of the block, A file stored in a single block, Abstract block, Name node keeps the block locations, Data flow in replication, Data pipeline in creating block replicas, Under reproduction, Name node metadata, Single namespace HDFS architecture, HDFS federation architecture, Place of data, Replica placement on two racks, HDFS network topology, Table: Block location class methods, How does HDFS store, read, and write files? Data node pipeline in writing a file, Verification of checksum, Data collection

and data analysis Hadoop cluster, Hadoop cluster in data storage and processing, Master protocol application, MRv2 cluster operation, Current and existing APIs, Data serialization options, Apache Avro, Sequence and Avro reference files, Apache Thrift, Thrift and protocol buffers comparison, Commands for HDFS shell file system, Select MapReduce work main and importance forms, A mapper's development cycle and areducer's function, The mapper's life cycle in the latest API, A reducer's life-cycle in the old API, The reducer lifecycle in the latest API, Input clues to output clues link, Input/Output Mapper sort, Key/Value types, Input Formats in the old API, Mapper Key/Values Input/Output number, Reducer Input/Output number of K values, Keys and attributes sorting, Combiners, Shuffle, Table: Parameters in compare methods, Table: Configuration properties to tune the sort and shuffle process, MapReduce with shuffle and sort, Settings and submissions for MapReduce job, Table: FileInputFormat<K,V> class Static methods using JobConf, Settings and submissions for MapReduce job, Combiner on reducer, Shuffle transfer number, Speculative performance, Data paths in MapReduce task input and output, Data movement in a MapReduce job in the reduce phase, Data flow provided by the InputFormat, Data flow provided by the OutputFormat, InputFormats for File-Based Input formats, Table: InputFormats for File-Based Input formats, RecordReader, compression andsequence files, LineRecordReader example, RecordReader with FileSplit, Built-In RecordReaders, Sequence files, SequenceFileInputFormat classsubclasses, SequenceFile class nested classes and interfaces, Sequence file header, Compression, Configuration properties for configuring, Codecs supported by Hadoop, Commands for HDFS Shell file system, Administration commands.

Unit 3. Query Languages for Hadoop

What is JAQL? JSON: JavaScript Object Notation, JSON format, where does JAQL fit? MapReduce overview, MapReduce and Hadoop, Starting up the JAQL server & entering JAQL in command line mode, JAQL and MapReduce, Let's do this step by step, JAQL and MapReduce: The rewrite engine and explain, JAQL schema, Data types, JAQL basics, Arrays, Records, Operators, Lazy/late evaluation, Why materialized assignment (:=) ? The -> operator, Expressions, Functions, Why JAQL core operators, Core operators: Expand, Core operators: Group, Core operators: Group format (single), Core operators: Group (single), Core operators: Understanding grouping, Core operators: Co-groups, Core operators: Join outer joins, Core operators: Sort, JAQL SQL, JAQL SQL: Case-sensitivity, JAQL and MapReduce basics, JAQL and MapReduce: Explain, JAQL and MapReduce: Map, MapReduce: Job configuration, JAQL and MapReduce: Native MR jobs, JAQL I/O, JAQL I/O adapter operations, JAQL I/O: I/O adapters, JAQL I/O: I/O adapters arguments, JAQL I/O: Delimited files, JAQL I/O: Binary sequence files, JAQL I/O: Text sequence files, JAQL I/O: Other adapters.

Unit 4. Hive: Hadoop Reporting and Analysis

History of Hive, Hive components, Hive directory structure, Physical layout: Data in Hive, Database use/drop/alter, Primitive data types, Complex data types, Creating a table, Table partitioning, Managed Vs external tables, Indexes, Drop/alter table, Loading data into Hive: From a file, Loading data from a directory, Select from, Selecting from partitions, Joins, Order by/sort by, Views, Order, CLI (Command Line Interface), Metastore, Real world use cases, Hive use case, Hive command line, Language of Hive Query (HQL), Creating tables, Hive primitive data types, Hive data review, Aggregations and affiliation, HBase, HBase schema, Social media events with sparse columns, HBase timestamp versioning, Importing from

MySQL to Hive, Import to HBase from MySQL, Intake of Flume streaming data, Multi-agent Flume data flow, Fan-in Flume data flow, Log ingestion into HDFS.

Unit 5. Pig: Hadoop Reporting and Analysis

What is Pig? Pig versus other tools, Executing Pig, First look at Pig data, Pig Latin statement basics, Input, LOAD operator continued, Accessing data, Case sensitivity, Field reference, Pig data types, Operators, Parameter substitution, Output, MapReduce in Pig, Cascading in Pig, Apache hive and Pig, Pig data form, Complex data types, Map, Schema, Casting, Casting error, Comparison operators, Identifiers, Boolean operators, Invoking the grunt shell, Auto completion, Grunt shell flow, Pig operators and commands, Regex in the file path, Store, Dump, Foreach generate, Flatten, New schema, Nested block, Null, Comparison operators, Assert, SPLIT, Flatten, RANK, Order by, Using the partitioner, Using a shell program, MapReduce program, CUBE, Rollup, Parameter substitution, Advanced JOIN, Equi Joins, Inner Joins, Left outer join, Cogroup, CROSS Join, Functions, Pig storage, HBase Storage, Apache Oozie, Types of Oozie jobs, Set a value to a property, Scheduling a Pig script, Integrating with the workflow, Upload Files to HDFS, Bundle, Oozie user interface.

Unit 6. Data Visualizations

Visualization, Visualization in Big Data? Visualization value of Big Data, Large data visualization issues, Analysis of diagrams, Graphs and network organization, Algorithms for graph analytics and solutions, Dedicated appliances for graph analytics, how to select among different chart types? Pie chart, Doughnut chart, Line chart, Map chart, Tree map chart, Waterfall chart, Scatter plot, Histogram chart.

Unit 7. NoSQL

What is NoSQL? NoSQL and SQL, Brief history of NoSQL databases, Features of NoSQL, Types of NoSQL databases, Relational Vs Document database, Graph-based, Advantages and disadvantagesof NOSQL, NoSQL's benefits, MongoDB, Features of MongoDB, Why use MongoDB? MongoDB data modeling, MongoDB: Create database, MongoDB drop database, MongoDB create collection, MongoDB drop collection, MongoDB update documents, MongoDB delete documents, MongoDB query documents, MongoDB and SQL similarity, MongoDB text search, Text index, MongoDB shell, How to run the shell, MongoDB shell collection methods, Specifying the collation.

Unit 8. R: Hadoop Reporting and Analysis

What is open-source R? The R appeal: What attracts users? Companies currently using R, what is the R programming language? Limitations of open-source R, Open source R packages to boost performance, Challenges with running large-scale analytics, 3 key capabilities in big R, Big R architecture, User experience for big R, what's behind running big r's scalable algorithms? Big R machine learning: Scalability and performance, Simple Big R example.

Reference books -

1. "Big Data Analytics: Turning Big Data into Big Money" by Frank J. Ohlhorst
2. "Hadoop: The Definitive Guide" by Tom White
3. "Programming Pig: Dataflow Scripting with Hadoop" by Alan Gates
4. "Hive: The Definitive Guide" by Edward Capriolo, Dean Wampler, and Jason Rutherglen

5. "Learning Spark: Lightning-Fast Data Analytics" by Holden Karau, Andy Konwinski, Patrick Wendell, and Matei Zaharia
6. "Data Science for Business: What You Need to Know about Data Mining and Data-Analytic Thinking" by Foster Provost and Tom Fawcett
7. "MongoDB: The Definitive Guide" by Kristina Chodorow and Michael Dirolf

DSE5: Application of Machine Learning in Industries

Course Description -

The course on "Applications of Machine Learning in Industries" provides an in-depth exploration of how machine learning techniques are applied in various sectors to address complex challenges and drive innovation. The course covers a wide range of industries, including banking and securities, communication and media, healthcare and life sciences, education, manufacturing and petroleum, government administration, insurance, retail and supply chain, transportation and logistics, and energy and utilities. Students will gain a comprehensive understanding of the use of machine learning algorithms and models to solve real-world problems and improve operational efficiency in different industrial domains. The course also emphasizes the importance of data preparation, model selection, and evaluation methods specific to each industry.

Course Objectives -

1. To familiarize students with the application of machine learning in diverse industries and its significance in solving complex problems and driving business growth.
2. To explore various machine learning techniques and algorithms used in different industrial domains, including banking, healthcare, retail, transportation, energy, and more.
3. To provide hands-on experience in implementing machine learning models to solve industry-specific challenges through case studies and practical assignments.
4. To understand the challenges and limitations of applying machine learning in different industries and develop strategies to overcome them.
5. To develop critical thinking and analytical skills to identify suitable machine learning solutions for specific industry problems.

Course Outcomes -

By the end of the course, students will be able to:

1. Demonstrate a comprehensive understanding of the application of machine learning in various industries, including banking, healthcare, manufacturing, logistics, and more.
2. Apply machine learning algorithms and techniques to address specific challenges in different industrial domains.
3. Evaluate and compare the performance of machine learning models in industry-specific applications and make data-driven decisions.
4. Design and implement machine learning-based solutions to improve operational efficiency and decision-making in different industries.

5. Identify potential use cases for machine learning in various industrial sectors and propose innovative solutions.
6. Analyze the challenges and limitations of applying machine learning in industries and propose strategies to enhance model performance and overcome obstacles.
7. Communicate effectively about machine learning concepts, applications, and results to both technical and non-technical stakeholders in industrial settings.

Unit 1. Machines Learning in Banking and Securities

Why machine learning in banking sector, Use of AI in banking and finance, Fraud detection, Tough competition in banking industry, Risk modeling and investment banks, Customer data management, Decreased customer experience and loyalty, Personalized marketing, Role of machine learning: Challenges of banking sector and securities, Widely used machine learning algorithm in banking and security, Fraud prevention and detection systems, Rule based and machine learning based approach in fraud detection, Anomaly detection: Ways to expose suspicious transactions in banks, Advanced fraud detection systems, Risk management systems, Case study: Application of machine learning for financial risk management, Credit risk analysis using machine learning classifier, Investment prediction systems, Portfolio management systems, Objectives of portfolio management, Algorithmic trading, Deep learning for customer services, Chatbot: Deep learning approach, AI powered marketing systems, Deep learning in cyber security, Types of cyber-attacks in banks, Deep learning methods used in cyber security, Deep learning v/s restricted Boltzmann machines, Convolution Neural Networks (CNNs), Recurrent neural networks, Machine learning techniques: Loan underwriting & sentiment/news analysis, Sentiment or news analysis, Current challenges and opportunities: Banking and security domain.

Unit 2. Machine Learning in Communication, Media and Entertainment

Machine learning in communication, media and entertainment, Usage of machine learning in media and entertainment industry, Machine learning techniques for customer sentiment analysis, World embedding's, Sentiment analysis with long short term memory networks, Real-time analytics in communication, media and entertainment industries, Real time analytics and social media, Deep learning for social media analytics, Recommendations engines, Collaborative filtering, Memory based collaborative filtering, Model based collaborative filtering, Content based filtering, Hybrid recommendation systems, Summary of recommendation systems, Deep learning techniques on recommender systems.

Unit 3. Machine Learning in Healthcare and Life Science

Applications of machine learning in health and life sciences, The most important applications of machine learning in healthcare, Role of machine learning in drug discovery, Machine learning approaches in drug discovery, Medical image analysis, Why deep learning for medical image analysis, Neural network and deep learning architecture, Comparisons between architecture of different types of deep learning models, Machine learning in genetics and genomics, Genomics and AI background, Two category of genomics, How to use deep learning effectively, Interpreting deep learning models, Predictive medicine: Prognosis and diagnostics accuracy, Predictive medicine: Examples, ML applications in breast cancer diagnosis and prognosis.

Unit 4. Machines Learning in Education

Machine learning in education, Advantages of machine learning in education, learning analytics, Academic analytics, Action research, Educational data mining, Recommender system, Personalized adaptive learning, Learning analytics process, Data environment: What? Stakeholders: Who? Methods: How? Case study: Sentimental analysis for student's feedback using ML, Recommender systems in education, Domain model, Learner model, Students classification algorithm, Recommendation model, Case study: Application of ML in predicting students' performance, Proposed methodology, Data description, Sample data sets, Visualization, Selection of machine learning technique.

Unit 5. Machine Learning in Manufacturing and Petroleum Industries

Introduction, Applications of machine learning in manufacturing industry, Deep learning for smart manufacturing, Machine learning for quality control in manufacturing, Case study, Construction of CNN, Experimental results, Efficiency of CNN for defect detection, Comparative experiments, Machine learning for fault assessment, Time frequency methods, Spectrograms: Short-Time Fourier Transform (STFT), Scalograms: Wavelet transform, Hilbert-Huang transform, Proposed CNN architecture for fault classification based on vibration signals, Case study, Machinery failure prevention technology, Conclusion.

Unit 6. Applications of Machine Learning in Government Administration

Introduction, Risk and compliance, Type of government problems appropriate for AI applications, AI for citizen services use cases, Answering questions, Routing requests, Translation, Drafting documents, Chat bots for communication between citizen and government, Media richness theory, Chatbots in the public sector, Case study, Data management services, Knowledge processing services, Application services, An application scenario, Classifications of citizen complaints using ML, Case study Step 1: Document collection, Step 2: Prepossessing, Step 3: Feature extraction, Term frequency-Inverse document frequency, Step 4: Feature selection, Step 5: Classification, How to implement, Result.

Unit 7. Machine Learning in Insurance Industry

Importance of machine learning in insurance, Potential use cases of machine leaning in insurance industry, Case study on insurance climb analysis using machine learning algorithms, Case study on using machine learning for insurance pricing optimization, Personalized marketing in insurance industry, Predictive model for insurance underwriting, Case study: Risk prediction in life insurance industry.

Unit 8. Applications of Machine learning in Retail Industry and Supply Chain

Introduction, Inventory management, Few use case examples, Benefits of predictive analytics to retailers, Robots-seeing to customer satisfaction, IoT: Prevention first, Predictive analytics: Weathering demand, Analysing buying patterns, Analysing traffic patterns, Assortment planning, Eliminate guess work, Feed the right stores, Get better information, Assortment planning to drive supply chain, Retail analytics, Domestic forecasting, Case study: Forecasting seasonal footwear demand using ML, Demand forecasting methods, Predictor variables in demand forecasting, Traditional techniques v/s machine learning techniques, Methodology, Machine learning techniques used, List of attributes from the aggregated data by month at the style level, Feature selection and engineering, List of attributes for feature selection, Dataset partitioning, Model building, Three step model, K-means clustering, Three steps followed in classification, Three sub-steps in prediction, Performance measurement, Results, Three step model,

Machine learning for supply chain management, Recommended architecture for machine learning models, Machine learning models use case.

Unit 9. Machine Learning in Transportation and Logistics

Introduction, Applications of ML and artificial intelligence in transportation, Applications of machine learning in transport, Incident detection, Predictive models, Application of AI in aviation and public transportation, Aviation, Shared mobility, Buses, Intelligent urban mobility, Autonomous vehicles, Autonomous transportation, Artificial intelligence use cases in logistics, Back office AI, Cognitive customs, Predictive logistics, Predictive risk management, Seeing thinking and speaking logistics operations, ML powered customer experience, Limitations of AI techniques in transportation, Computation complexity of AI algorithms.

Unit 10. Machine Learning in Energy and Utilities

Introduction, Smart grid, Smart grid technologies, Key characteristics of smart grid, Machine learning applications in smart grid, Machine learning techniques for renewable energy generation, Forecasting renewable energy generation, Wind power generation, Solar energy generation, Hydro power generation, Determining plant location, size and configuration, Managing renewable energy-integrated smart grid, Machine learning applications in wind energy forecasting, Case study: Wind power forecasting based on daily wind speed data, Wind energy output calculations based on hourly wind speed, Machine learning techniques used, LASSO regression, KNN regression, xGBoost regression, Random forest regression, Support vector regression, Wind power forecasting method using machine learning algorithm, About data set, Case studies, Case 1: Wind power forecasting based on daily mean wind speed and standard deviation, Forecasting accuracy of algorithms, Case 2: Wind power forecasting based on only daily mean wind speed, Case 3: Wind power forecasting for a different region.

Reference books -

1. "Machine Learning in Finance: From Theory to Practice" by Bob Hunt, Yanbo Wang, and David Wong
2. "Machine Learning for Dummies" by John Paul Mueller and Luca Massaron
3. "Machine Learning for Healthcare" by Stephen P. King
4. "Machine Learning in Action" by Peter Harrington
5. "Predictive Analytics for Business: Using Data Mining for Business Advantage" by Galit Shmueli, Nitin R. Patel, and Peter C. Bruce
6. "Machine Learning for the Web" by Andrea Isoni
7. "Artificial Intelligence for Supply Chain Management: Smart Manufacturing, Digital Twins, and Industry 4.0" by Chung-Yee Lee and Tsan-Ming Choi

DSE6: Social and Web Analytics

Course Description -

The course on Social and Web Analytics provides an in-depth understanding of the use of data analytics in the context of social media and web platforms. It covers the fundamental concepts and tools for collecting, analyzing, and interpreting data from various social media platforms and websites. Students will learn how to leverage social media data to gain insights into user behavior, measure the effectiveness of marketing strategies, and make data-driven decisions. The course also explores different types of analytics platforms and techniques, as well as the future trends in social media analytics and monitoring.

Course Objectives -

- To introduce students to the significance of social media and web analytics in the modern business landscape.
- To familiarize students with the tools and methodologies for collecting and analyzing social media data.
- To enable students to understand and interpret key performance indicators (KPIs) and metrics relevant to web and social analytics.
- To provide insights into managing web and social media analytics through the use of dashboards and reports.
- To explore the future trends and potential advancements in social media analytics and monitoring.

Course Outcomes -

By the end of the course, students will be able to:

- Demonstrate a comprehensive understanding of social media platforms, their impact on businesses, and the need for data analytics in this domain.
- Apply data collection strategies and effectively analyze data from various social media sources.
- Identify and define relevant KPIs and metrics for measuring the performance of websites and social media campaigns.
- Utilize dashboards and reports to monitor and manage web and social media analytics effectively.
- Gain insights into future trends and advancements in the field of social media analytics and monitoring.
- Apply social media analytics to improve decision-making and develop data-driven strategies for businesses.

Unit 1. Introduction to Web and Social Analytics

Social media, Importance of social media, Social media popularity, Benefits of social media, Tools available, Need of using analytics, Social analytics vs web analytics, Web analytics, Terms used by web analytics tools, Types of web analytics, Social and web analytics technical requirements, Social analytics main basic activities, Social media environment, Impact of social media on business, Leverage social media for better services, Current analytics platforms, Open-source vs licensed platforms, Google analytics, IBM social media analytics, Other social and web analytics tools, Choosing right specifications and optimal

solution, Measuring variations in user behaviour, The diversity of user activities, The origin of the user activity distribution, Long tails everywhere: The 80/20 rule (p/q Rule), User activities on twitter.

Unit 2. Relevant Data and its Collection & KPIs/ metrics

Data collection strategy, Content sharing, Challenges in the social media data collection, Characteristics of people-centric approach, Social graph, Influencers, Community managers, Organizing for social media, Factors that help in applying the model to an organization, Measures to organize the social media for success, Choosing focused data sources and social networks, Collecting and understanding social media data, Social data collection strategy, Facebook social APIs, Best practices for collecting data from social networks, Leverage qualitative data: What, why and how much, How to conduct a lab usability testing, Usability alternatives, Recruiting user research participants with social media, Outsourced online usability, Web-enabled emerging user research options, Online surveys, Types of surveys, Eight tips to choose online survey providers, Types and Properties of Social Networks, Directed Versus Undirected Graphs, Node and Edge Properties, Visualization Network, The discipline of social media analytics, Collecting data, Performing analysis, Reporting results, Aligning social objectives with business goals, Social media objectives, Developing a waterfall strategy, A SMART methodology, Identify common social business objectives, Developing KPIs, Creating KPI's, Developing KPI's, Basic KPIs, Standard vs critical metrics , Critical web metrics: Bounce rate, Critical web metrics: Exit rate, Critical web metrics: Conversion rate, Critical web metrics: Engagement, Strategically aligned KPIS and tactics to find best web and social metrics, Specific KPIs, Tactics to find the best web and social media metrics, Business goals, Visualizing the social analytics framework, Build scorecards and dashboards to track KPIs, Measuring macro and micro conversions, Quantify economic value, Measuring success for non-ecommerce and B2B websites, Measuring B2B websites.

Unit 3. Manage Web and social media with Analytics.

Managing web and social analytics, Social media analytics, Dashboard, Dashboard: Drill down capabilities, Dashboards in IBM Cognos Consumer Insight (CCI), Types of charts, Drill down to view more details, Sharing insights with dashboards, Relationships, Viewing relationships, Sentiments, Sentiment analysis, Sentiment terms, Evolving topics, Evolving topics: IBM social media analytics, Reports, Social media impact, Segmentation, Discovery, Content creation and tracking, Social media content creation, Issues in content creation, Competitive intelligence analysis, CI data sources and types, Competitor intelligence studies, Website traffic analysis, Analysing referrals and destinations, Top keywords performance trend, Audience identification and segment analysis, Segmentation analysis, Social media optimization and benefits, Social media enablement audit, Understand signals and potential, Improving social signals, Focus on text and unstructured data, The basics of natural language processing, Modelling: Unsupervised vs supervised.

Unit 4. Future of Social Media Analytics and Monitoring

Future of social media analytics and monitoring, , Large volumes of social media, Future of social analytics, Future social media analytics and monitoring, Mashing up data from disparate sources, Integrating customer profile data, True profile, Colliding data sets for big bang ideas, Purchase funnel, Social interactions and web visits, Server response time and help desk tickets, Social media integration and benefits, Integrate solution to share outcome with others, Strategy planning, Benefits of social media

integration, Integrating social media into the organization, Integration mode, Social media integration, Goals of integrating social media, Social media integration, Social media services online, Search engines, Content engagement, Interactions: Real world & people, Learning and mapping, Under and overfitting, Regularizing in matrix factorization, Exploratory analysis, Prediction and recommendation.

Reference books -

1. "Social Media Analytics: Effective Tools for Building, Interpreting, and Using Metrics" by Marshall Sponder
2. "Web Analytics 2.0: The Art of Online Accountability and Science of Customer Centricity" by Avinash Kaushik
3. "Social Media Metrics: How to Measure and Optimize Your Marketing Investment" by Jim Sterne
4. "Data Science for Business: What You Need to Know about Data Mining and Data-Analytic Thinking" by Foster Provost and Tom Fawcett
5. "Web and Network Data Science: Modeling Techniques in Predictive Analytics" by Thomas W. Miller
6. "Social Media Analytics Strategy: Using Data to Optimize Business Performance" by Alex Gonçalves
7. "Social Media Analytics and Data Mining" by Gabor Szabo and Oscar Boykin

DSE7: Computational Linguistics and Natural Language Processing

Course Description -

This course on Computational Linguistics and Natural Language Processing (NLP) provides a comprehensive introduction to the fundamental concepts, approaches, and applications of NLP. Students will learn about classical approaches to NLP, including linguistic levels of analysis (morphology, syntax, semantics, pragmatics), and text processing techniques such as statistical methods and text classification. The course explores empirical and statistical approaches to NLP, covering topics like corpus creation, annotations, and statistical parsing. Students will also delve into the applications of NLP, including information retrieval, question-answering systems, information extraction, report generation, and ontology construction. Additionally, emerging applications of natural language generation in information visualization, education, and healthcare are explored.

Course Objectives -

- To introduce students to the foundations of natural language processing and computational linguistics.
- To familiarize students with classical, empirical, and statistical approaches to NLP and their respective use cases.
- To provide students with a practical understanding of text processing techniques, including statistical methods and text classification.
- To enable students to work with linguistic corpora, perform annotations, and understand the significance of corpus linguistics.
- To explore various applications of NLP, such as information retrieval, question-answering systems, and report generation.

- To introduce students to emerging applications of natural language generation in information visualization, education, and healthcare.

Course Outcomes -

By the end of this course, students should be able to:

- Understand the linguistic levels of analysis in NLP (morphology, syntax, semantics, pragmatics) and apply them in practical scenarios.
- Analyze and process text data using statistical methods and text classification techniques.
- Create and work with linguistic corpora, perform annotations, and interpret the results.
- Apply empirical and statistical approaches to NLP tasks such as parsing and word sense disambiguation.
- Develop applications in information retrieval, question-answering systems, information extraction, and report generation using NLP techniques.
- Explore the use of NLP in emerging fields, including information visualization, education, and healthcare, for multimedia presentation generation, language interfaces, and sentiment analysis.

Unit 1: Classical Approaches of NLP

Natural Language Processing (NLP) and its significance in understanding human language. Overview of classical approaches used in NLP, focusing on linguistic levels: Morphology, Syntax, Semantics, and Pragmatics. Traditional and empirical approaches to NLP, with an example of Automatic Summarization using NLP. Text processing techniques, including text analysis, text mining, and text analytics. Tools and methodologies employed in NLP, such as statistical methods, text classification, and text extraction. The scope and importance of text analysis, along with the working principles and evaluation of text classification and extraction processes. Overview of NLP APIs and levels of NLP, including lexical analysis, pre-processing, POS tagging, syntactic parsing, and semantic analysis. Introduction to natural language generation (NLG) and its working principle, along with limitations.

Unit 2: Empirical Approaches

Creating and utilizing corpora in NLP, focusing on corpus linguistics and different types of corpora. The role of annotations in text corpora, along with NLP task-specific training corpora. Usage of data sets for NLP and treebank annotation for linguistic description. Introduction to statistical techniques in NLP, such as Hidden Markov Models, Maximum Entropy Markov Models, Conditional Random Field Models, and Support Vector Machines. POS tagging and Word Sense Disambiguation (WSD) using statistical models.

Unit 3: Statistical Approaches

Statistical parsing in NLP, including lexicalized statistical parsing and different parsing approaches. Introduction to multiword expressions and word similarity measures. Utilizing pre-trained sentence encoders and BERT for text similarity. Challenges and methods in Word Sense Disambiguation, along with evaluation techniques. History and working principle of speech recognition technology, and major leaders in the field. Introduction to machine translation and its various approaches.

Unit 4: Applications of Natural Language Processing

Understanding the role of NLP in Information Retrieval (IR) and its model types. Design features and evaluation metrics of IR systems. Introduction to Question Answering (QA) systems and their architecture.

Information extraction process and applications of chunking. Report generation using NLP text, and the construction and advantages of ontologies.

Unit 5: Emerging Applications of Natural Language Generation

Exploring emerging applications of NLG in various domains, such as multimedia presentation generation, education, and healthcare. Utilizing NLP in scripted dialogue and language interfaces for intelligent tutoring systems. Applications of NLP in healthcare, including sentiment analysis and opinion summarization.

Reference books -

1. "Speech and Language Processing" by Daniel Jurafsky and James H. Martin
2. "Foundations of Statistical Natural Language Processing" by Christopher D. Manning and Hinrich Schütze
3. "Natural Language Processing in Action" by Lane, Howard, and Hapke
4. "An Introduction to Language Processing with Perl and Prolog: An Outline of Theories, Implementation, and Application with Special Consideration of English, French, and German" by Pierre Nugues
5. "Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition" by Daniel Jurafsky and James H. Martin
6. "Natural Language Processing: A Concise Introduction" by Jacob Eisenstein
7. "Statistical Language Learning" by Eugene Charniak

DSE8: Analytics for Industries

Course Description -

This course on "Analytics for Industries" provides an in-depth understanding of business analytics and optimization, with a focus on its application in specific industries such as banking, insurance, telecom, and healthcare. Students will learn about various types of analytics, including predictive analytics, prescriptive analytics, and cognitive analytics, and how these techniques can be utilized to drive data-driven decision making and enhance profitability in different industry contexts. The course will cover essential concepts, capabilities, and technologies in business analytics, with real-world examples and case studies from the respective industries.

Course Objectives -

- To introduce students to the fundamental concepts of business analytics and optimization and its relevance in various industries.
- To explore different types of analytics and their applications in specific industry domains, such as banking, insurance, telecom, and healthcare.
- To understand the challenges and opportunities in implementing business analytics solutions in diverse industry contexts.
- To familiarize students with essential capabilities and tools in business analytics, including predictive analysis, real-time analytics, and customer insights.
- To provide insights into the impact of business analytics on key areas, such as customer care, risk management, and operational efficiency in different industries.
-

Course Outcomes -

By the end of this course, students will be able to:

- Demonstrate a comprehensive understanding of business analytics and its role in various industries.
- Apply different types of analytics techniques to solve industry-specific challenges and make data-driven decisions.
- Evaluate the impact of business analytics on business performance, customer retention, and risk management in different industries.
- Develop strategies to leverage business analytics for competitive advantage and improved operational efficiency in specific industry domains.
- Analyze real-world case studies and industry examples to identify best practices and successful implementations of business analytics.

Unit No. 1: Introduction to Business Analytics and Optimization

Introduction to Business Analytics and Optimization, Types of Analytics, Business Analytics value model - e.g. Higher education, Challenges - Volume, Variety (Diversity), and speed of Data Creation (and needed decisions), Approaches to help maximize profitability and returns, Business Analytics, Business Analytics Capabilities, Enterprise Analytics Capabilities, Business Analytics Technologies, Predictive Analytics, Prescriptive Analytics, Cognitive analytics – Watson breaks the glass ceiling, A fact-based decision making culture, A strong data infrastructure, The Right Analytical Tools, New tools and architectures may be needed, Analytics Workforce, Knowledge Requirements, Business Analyst, Data Scientist, Where to put the analytics team, IBM Business Analytics Maturity Model, Optimization, Key BAO Concepts, The need for BAO now, Essential Capabilities In BAO, BAO Capabilities: Business Performance Management, BAO Capabilities: Predictive Analysis and Mining, Value of BAO to Business Organization, Impact of BAO on diverse industries, Advantages to implementing BAO solutions, BAO Capabilities: Real-time Analytics: Data In Motion, BAO support for decision-making, High level architecture of BAO, Importance of reference architecture, BAO reference architecture, BAO reference architecture to BAO architects, IBM Technology Portfolio for BAO.

Unit No. 2. Business Analytics for the Banking Industry

New approaches, Public Focus as The Banking Industry, An Example Of Possible Prediction (?), Priorities for Enhancing Competitiveness, Challenges, Capability Map, Transition Towards BAO With Actionable Insights, Example: Imagine If A Bank Could…, IBM Business Analytics Maturity Model, Essential capabilities in BAO for Banking, BAO capabilities: Conceptual Model, BAO capabilities: Analysis and Reporting, BAO capabilities: Analysis And Reporting, BAO Capabilities: BPM, BAO Capabilities: Predictive Analysis and Mining, BAO Capabilities: Example Of Simplified "Decision Tree" – Credit Risk Rating, BAO Capabilities: Real-time Analytics: Data In Motion, Business Analytics For Banking, Operational Efficiency, Examples of BAO for Operational Efficiency, Integrated Risk Management, Examples Of BAO for Integrated Risk Management (IRM), IBM Solution Architecture - IRM, (IRM) Screenshot: Enterprise Risk Dashboard, Customer Care and Insight, Examples Of BAO for Customer Care and Insight, Payments and Securities, Example of BAO for Payments and Securities, Business Analytics Maturity Model (Reminder), IBM Technology Portfolio (Excerpt) for BAO for banking, Big Data analytics to create predictive models, Analyzing Customer Behavior for Insights, Cloud based solution for acquiring new customers,

Data consolidation for profit, Near Real-Time Insights Driving Growth, BBVA - Monitors social media to improve online reputation, BBVA - Banco Bilbao Vizcaya Argentaria, Analytical insights to drive operational and marketing decisions, Integrated risk management for competitiveness, Big data analytics for insights from customer transactions, Pioneer West Virginia Federal Credit Union, Summary: Banks Are Becoming Smarter With Business Analytics.

Unit No. 3. Business Analytics for the Insurance Industry

Industry changes mean “business as usual” is no longer an option, Big Data presents a huge opportunity for insurers—if they can harness it, Analytics has evolved from a business initiative to a business imperative, Analytic-driven organizations are distinguished by their ability to leverage…, Analytics can deliver value in all parts of your insurance organization, And can improve outcomes in four key areas, The smarter insurance company: business requirements for success, Creating a customer-focused enterprise is difficult, Customer insight is vital for retaining and growing customers, Getting a complete view of customer data with analytics, Capture: Creating a 360 degree customer view, Predict: Predictive Analytics uses the evidence in your data to…, Predict: Some factors which affect retention, Predict: Example retention model, Act upon the predictions and insights, Close the loop: ongoing refinement, Customer retention and growth, Insurance customer retention and growth, In Closing…, Corona Direct, Managing risk across the enterprise is no simple task, IBM Risk Analytics helps you manage risk enterprise wide, IBM solutions help address today’s risk management challenges, IBM Risk Management Framework, Example : Solvency II Pillar 3 Reporting, Allianz Group, Creating an optimized multichannel enterprise is difficult, IBM sales performance management solutions enable insurers to create an integrated channel enterprise, Sales performance management overview, Primerica, Insurers need to find a smarter way to manage claims, IBM Analytics Solutions enable insurers to optimize the Claims Process, Predictive analytics and reporting for claims business scenario, Predictive analytics and reporting for claims, Business Scenario, Santam, Analytics in Insurance, IBM Business Analytics addresses key insurance imperatives, Risk, customer, channel optimization, and operational effectiveness.

Unit No. 4. Business Analytics for the Telecom Industry

Challenges facing communications service providers in a connected world, Information explosion with Big Data—Variety, Velocity, Volume…, Analytics in Telecommunications, Commoditization is relentless; business as usual is not an option, CSP’s are exploiting opportunities to become the middleware for a ‘Smarter Planet’, Turning large data into usable and actionable insight… , Analytic-driven organizations are distinguished by their ability to leverage…, Analytics can deliver value in all parts of your organization, Smarter Analytics for Telecommunications, Smarter Analytics solutions for CSPs, Customer Analytics Solutions: Identifying and retaining at-risk customers, IBM Churn Prediction Analytics, Customer retention strategy is an iterative process, Client success: proactive intervention versus costly reactive saves at XO Communications, Marketing Optimization solutions: determining what the ‘next best offer’ is as well as when and how to deliver it, Using Predictive Models to drive retention…, Next best action capabilities, Sentiment derived from text analytics, Leveraging social networks analytics, Social media and customer advocacy solutions: Identifying and rewarding happy customers is differentiating CSPs, Unleashing the power of IBM SPSS Social Media Analytics, Harnessing traditional enterprise and external social data, Social Media Analytics dashboards deliver segmentation and auto-discovery of evolving and associated topics, Network analytics solutions: Vast volumes of data exist, but insights are fragmented and inconsistent, Leveraging network analytics to gain deeper insight into the actual customer experience and network performance, Client Success: A leading US wireless provider is exposing insights across the entire enterprise, Management dashboards to identify network performance issues.

Unit No. 5. Business Analytics for Healthcare

What are the key healthcare trends & business challenges? Global Industry Crisis: Driving Analytics and Performance Management in Healthcare, Key Pain Points, How do we solve these challenges?, Selected Cognos Customers in Healthcare, IBM Cognos Performance Management Framework, Identifying BA Opportunities in Healthcare: Target with an entry point strategy, Performance Management, Key Healthcare Performance Management Initiatives, Exploit Core Competencies and Plan for Strategic Growth, The Provider Planning Performance Blueprint, Support quality initiatives and modifications to care practices, Supporting and Enhancing Physician Productivity and Satisfaction, Customer example -

SHARP, Customer example - SETMA, Customer example – Martin's Point, SPSS example – Sequoia Hospital, SPSS example – CRI, SPSS example – Independent Health, Business Analytics and Optimization (BAO) is a strategic growth priority for IBM, IBM Business Analytics solutions are integral within the Health Framework and developing analytics solutions, Analytics will progress along a solid infrastructure, Benefits…What if?, Key capabilities that drive performance, Customer Success Stories – VITAS Innovative Hospice Care, Customer Success Stories – Cardiff and Vale University Health Board, Customer Success Stories - Sharp HealthCare, Customer Success Stories - SETMA, Customer Success Stories – Martin's Point Healthcare.

Unit No. 6. BA for Sales & Marketing

Analytics in sales and marketing – Introduction, Key questions by marketing functions, Marketing imperatives, Empowered customer challenge, CXO focus and views surveyed, ROI as a measure, Empowered consumers challenge, Customer analytics linkage to business applications, Analytics drive business applications,

Analytics for the results-driven marketer, Customer insight, Smart marketing functions through analytics, Demand generation by smart analytics, Business analytics for demand generation solutions, Demand generation – Response driven by analytics, Business analytics for customer acquisition, IBM solution: Customer acquisition for marketing, Acquiring the ideal customer, Segmentation strategies, Targeting those most likely to respond, Online customers acquisition – Delivering a smarter shopping experience, Align and understand the data, anticipate and predict, Case study: Resource allocation, Case study 2: Effective targeting, Benefits of customer analytics solutions, Smart marketing operations through analytics, Maximizing ROI and identifying best performing offers, Business analytics for marketing operations, Marketing performance analytics, Marketing performance indicators, "What if" scenario modeling for optimized resource allocation, Case study: Marketing performance analytics.

Unit No. 7. Human Resource Analytics

Human resources and challenges, Human resource analytics, HR analytics – Some of the key areas & insights, Solution area 1: HR analytics – Talent management, Human capital source of economic value, Effectiveness of building capabilities, Challenges of infrastructure, Usage of workforce analytics, Analytics is a key enabler, Use analytics to acquire, grow and retain, HR information in analytical process, Solution area 2: HR analytics – compensation & benefits, Compensation and benefits- Challenges, Alignment with key business processes, Analytics enable compensation management, Analytics in compensation process, Case study – Getty images, Case study – Mitel networks, Solution area 3: HR analytics – Workforce attrition, Human capital management, Workforce Attrition insights, HR survey– Data collection IBM SPSS, Cognos integration – Publishing in one step, Data Mining – Combine data sources, Data mining – Clustering, Data mining – Predict attrition, Data mining – Publish results to Cognos, Cognos integration – Creating a report, The dashboard, Predictive analytics – Anticipating potential attrition scenarios.

Unit No. 8. Financial Analytics

Role of CFOs to make the company smarter, Factors influencing the challenges faced by the companies, Role of CFOs in company, Agenda of CFO, Need for the core finance capabilities, Four finance profiles, Three key themes emerged from examining the four profiles, Challenges related to the automation and standards, Advantages of standard adoption, Steps to improve the corporate performance, Second key themes emerged from examining the four profiles, Challenges faced due to business insights, Business insight contribution, Business insight accelerators, Third key themes emerged from examining the four profiles, Value integrators effectiveness, Business insight leading to higher value, Characteristics of value integrators, Financial Analytics: Solutions Landscape, Importance of financial and predictive analytics, IBM solution to integrate the CFO performance insights, CFO performance insight – Solution architecture, CFO performance insight-Finance scorecard metrics, Metrics coverage of value drivers, CFO performance

insight-executive dashboard, Standard GL/AR/AP reports, Standard financial reports, CFO performance insight- Risk status indicator, Key risk indicators provide early warnings to potential weaknesses, CFO performance insight – "What-If" analysis, CFO performance insight- predictive analysis, CFO performance insight– Temporal Causal Modeling (TCM), Case study – BBVA, Case study – Muller Inc.

Unit No. 9. Case Studies

Case study – Cross-selling in the call center, Case study – Generating higher returns on marketing campaigns, Case study – XO communications, Case study – Telecom churn prediction, Case study – FBTO, Case study – WAZ Media Group, Case study – Home Trust, Fortis bank of Turkey, Case study – A large commercial bank in Japan, Case study – Large life insurance provider in Korea, Case study – Home furnishing retailer, Case study – Vimpel Com, Case study – KPN, Case study – Redcats, Case study – Marketing performance analytics, Customer profitability analytics for banking, Case study – Banco Itau, Case study – Mater, Case study – European government agency, Case study – AUVA, Case study – Mitel networks, Case study – Getty images, Case study – Elavon, Case study – Convenience store chain, Case study – PepsiCo, Case study – Chickasaw nation division of commerce, Case study – ANCAP, Case study – Jabil circuit, Inc., Case study – US food flavoring manufacturer.

Reference Books -

1. "Big Data Analytics in Banking and Finance" by Gloria Phillips-Wren and Stephanie Teufel
2. "Predictive Analytics for Insurance" by Eric Siegel
3. "Business Analytics for Telecoms: Building a Better Customer Experience" by Tony Costa and Yves de Montcheuil
4. "Healthcare Analytics for Quality and Performance Improvement" by Trevor L. Strome
5. "Healthcare Analytics: From Data to Knowledge to Healthcare Improvement" by Hui Yang and Walter W. Piegorsch
6. "The Predictive Healthcare: How Data Analytics Can Save Lives" by Brian T. O'Neill and Rory C. Sutherland
7. "Predictive Modeling Applications in Actuarial Science: Volume 1" by Edward W. Frees and Glenn Meyers

DSE9: Pattern Recognition & Anomaly Detection

Course Description -

The course on "Pattern Recognition & Anomaly Detection" introduces students to the concepts and techniques used in identifying patterns in data and detecting anomalies. Pattern recognition is a fundamental aspect of machine learning and artificial intelligence, where algorithms learn to recognize and classify patterns in various forms of data. Anomaly detection, on the other hand, focuses on identifying abnormal or unusual patterns in data that deviate significantly from the expected behavior. The course covers statistical and machine learning approaches for both pattern recognition and anomaly detection, providing students with a comprehensive understanding of these essential concepts in data analysis.

Course Objective -

The primary objective of this course is to equip students with the knowledge and skills to apply pattern recognition and anomaly detection techniques to real-world data problems. By the end of the course, students will be able to:

- Understand the fundamentals of pattern recognition and its applications in various domains.
- Gain insight into statistical approaches for analyzing and interpreting patterns in data.
- Familiarize themselves with machine learning algorithms and neural networks for pattern recognition tasks.
- Develop an understanding of anomaly detection and its practical use cases.
- Explore different anomaly detection approaches and evaluate the performance of anomaly detectors.
- Apply pattern recognition and anomaly detection techniques to real-world problems, such as network intrusion detection, IoT data analysis, time series forecasting, and more.
- Analyze and interpret the results of anomaly detection algorithms to make informed decisions in diverse applications.

Course Outcome -

Upon successful completion of the course, students will be able to:

- Identify patterns in data and utilize statistical techniques to analyze and interpret patterns effectively.
- Apply various machine learning algorithms, including neural networks, for pattern recognition tasks.
- Implement anomaly detection techniques and evaluate the performance of anomaly detectors on different types of data.
- Solve real-world problems related to anomaly detection, such as network intrusion detection, IoT data analysis, and time series forecasting.
- Demonstrate a comprehensive understanding of the principles and applications of pattern recognition and anomaly detection in diverse domains.
- Interpret the results of pattern recognition and anomaly detection algorithms and communicate their findings effectively to stakeholders.
- Apply the knowledge gained from the course to address complex data analysis challenges and make data-driven decisions.

Unit 1. Pattern Recognition and Anomaly Detection

Pattern and pattern recognition, Training and learning in pattern recognition, - Anomaly detection and its applications, Practical uses of anomaly detection, Anomaly detection over time, Key points for AI and ML in anomaly detection, AI system learning process, Case study: Anomaly detection with IBM Watson, Probability theory and maximum likelihood estimation ,Model selection and uncertainty matrices (confusion matrices), Loss of logging (log-loss) and F1 score, Hyperparameter selection and the problem with high dimensionality, Information theory.

Unit 2. Statistical Approaches for Pattern Recognition

Understanding statistics in pattern recognition, T-test and Z-test for hypothesis testing, Type I and Type II errors in hypothesis testing, Null hypothesis and statistical significance, Regression models and model selection, Linear models for classification.

Unit 3. Machine Learning Approaches for Pattern Recognition

Neural networks and their learning process, Kernel methods and sparse kernel machines, Graphical models, mixture models, and EM algorithm, Bayesian networks and sampling methods, Combining models for pattern recognition, Applications of K-means clustering.

Unit 4. Anomaly Detection Approaches

Anomaly detection and its applications, Types of anomalies and input data, Evaluation of anomaly detectors, Taxonomy of anomaly detection approaches, Classification-based anomaly detection, Contextual, collective, and online anomaly detection, Distributed anomaly detection and IDS analysis strategy.

Unit 5. Real-world Applications of Anomaly Detection

Anomaly detection in network intrusion detection systems, Anomaly detection in big data and 5G networks, Real-time anomaly detection in Docker and Hadoop clusters, Anomaly detection in IoT devices, Anomaly detection in time series data and forecasting, Anomaly detection using machine learning and deep learning, Anomaly detection in e-commerce pricing systems, IBM Watson AIOps for IT anomaly detection and remediation.

Reference books -

1. "Pattern Recognition and Machine Learning" by Christopher M. Bishop
2. "Anomaly Detection Principles and Algorithms" by Kishan G. Mehrotra, Chilukuri K. Mohan, and HuaMing Huang
3. "Introduction to Statistical Pattern Recognition" by Keinosuke Fukunaga
4. "Pattern Recognition: A Machine Learning Approach" by Christos Aggelos and Dimitrios Zevgolis
5. "Machine Learning: A Probabilistic Perspective" by Kevin P. Murphy
6. "Anomaly Detection for the Internet of Things" by Salvatore Venticinque and Michele Nitti
7. "Intrusion Detection: A Machine Learning Approach" by Rebecca Montanari