ANDHRA PRADESH STATE COUNCIL OF HIGHER EDUCATION

Programme: B.Sc. Honours Data Science (Major)

w.e.f. AY 2023-24

COURSE STRUCTURE

<table>
<thead>
<tr>
<th>Year</th>
<th>Semester</th>
<th>Course</th>
<th>Title of the Course</th>
<th>No. of Hrs /Week</th>
<th>No. of Credits</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>I</td>
<td>1</td>
<td>Essentials and Applications of Mathematical, Physical and Chemical Sciences</td>
<td>3+2</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2</td>
<td>Advances in Mathematical, Physical and Chemical Sciences</td>
<td>3+2</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>II</td>
<td>3</td>
<td>Introduction to Data Science and R Programming</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Introduction to Data Science and R Programming Practical Course</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4</td>
<td>Descriptive Statistics</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Descriptive Statistics Practical Course</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>III</td>
<td>5</td>
<td>Python Programming for Data Analysis</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Python Programming for Data Analysis Practical Course</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>6</td>
<td>Inferential and applied statistics</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Inferential and applied statistics Practical Course</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>7</td>
<td>Data mining techniques using R</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Data mining techniques using R Practical Course</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>8</td>
<td>Web technologies</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Web technologies Practical Course</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>IV</td>
<td>9</td>
<td>Data visualization using Tableau</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Data visualization using Tableau Practical Course</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>10</td>
<td>Data visualization using python</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Data visualization using python Practical Course</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>11</td>
<td>Introduction to SQL &amp; Advanced Tableau</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Introduction to SQL &amp; Advanced Tableau Practical Course</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>III</td>
<td>12</td>
<td>Supervised Machine Learning with Python</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Supervised Machine Learning with Python Practical Course</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Year</td>
<td>Semester</td>
<td>Course</td>
<td>Title of the Course</td>
<td>No. of Hrs/Week</td>
<td>No. of Credits</td>
</tr>
<tr>
<td>------</td>
<td>----------</td>
<td>--------</td>
<td>---------------------</td>
<td>-----------------</td>
<td>---------------</td>
</tr>
<tr>
<td>13</td>
<td></td>
<td></td>
<td>Unsupervised Machine Learning with Python</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Unsupervised Machine Learning with Python Practical Course</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>14 A</td>
<td></td>
<td></td>
<td>Web Scraping with Python</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Web Scraping with Python Practical Course</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>OR</td>
<td></td>
<td></td>
<td>Predictive &amp; Advanced Analytics using R</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Predictive &amp; Advanced Analytics using R Practical Course</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>15 A</td>
<td></td>
<td></td>
<td>Advanced Data Analysis Using Python</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Advanced Data Analysis Using Python Practical Course</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>OR</td>
<td></td>
<td></td>
<td>Data Wrangling with Java Script</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Data Wrangling with Java Script Practical Course</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>VI</td>
<td>Semester Internship/Apprenticeship with 12 Credits</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16 A</td>
<td></td>
<td></td>
<td>Big Data Analytics Using Spark &amp; Hadoop</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Big Data Analytics Using Spark &amp; Hadoop Practical Course</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>OR</td>
<td></td>
<td></td>
<td>Big Data security</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Big Data security Practical Course</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>17 A</td>
<td></td>
<td></td>
<td>Introduction to Deep Learning</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Introduction to Deep Learning Practical Course</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>OR</td>
<td></td>
<td></td>
<td>Deep Learning with Pytorth</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Deep Learning with Pytorth Practical Course</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>18 A</td>
<td></td>
<td></td>
<td>AI Concepts and Techniques With Python</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>AI Concepts and Techniques With Python Practical Course</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>OR</td>
<td></td>
<td></td>
<td>Data and Information Security</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Data and Information Security Practical Course</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>SEC</td>
<td></td>
<td></td>
<td>Introduction to Neural Networks</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Introduction to Neural Networks Practical Course</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>20</td>
<td></td>
<td></td>
<td>Natural Language Processing</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Natural Language Processing Practical</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Year</td>
<td>Semester</td>
<td>Course</td>
<td>Title of the Course</td>
<td>No. of Hrs /Week</td>
<td>No. of Credits</td>
</tr>
<tr>
<td>------</td>
<td>----------</td>
<td>--------</td>
<td>---------------------</td>
<td>------------------</td>
<td>---------------</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Course</td>
<td></td>
<td></td>
</tr>
<tr>
<td>VIII</td>
<td>21 A</td>
<td></td>
<td>Research Exploration</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Research Exploration Practical Course</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>OR</td>
<td>21 B</td>
<td></td>
<td>Computational Data Science</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Computational Data Science Practical Course</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>22 A</td>
<td></td>
<td>Computer Vision with Python</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Computer Vision with Python Practical Course</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>OR</td>
<td>22 B</td>
<td></td>
<td>Data Wrangling with Java Script</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Data Wrangling with Java Script Practical Course</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>23 A</td>
<td></td>
<td>Social Media Analytics</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Social Media Analytics Practical Course</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>OR</td>
<td>23 B</td>
<td></td>
<td>Pyspark Essentials For Data Science</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Pyspark Essentials For Data Science Practical Course</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>SEC</td>
<td>24</td>
<td></td>
<td>Business Intelligence and Visualization</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Business Intelligence and Visualization Practical Course</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>25</td>
<td></td>
<td>Data Visualization using JavaScript</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Data Visualization using JavaScript Practical Course</td>
<td>2</td>
<td>1</td>
</tr>
</tbody>
</table>
SEMESTER-I

COURSE 1: ESSENTIALS AND APPLICATIONS OF MATHEMATICAL, PHYSICAL AND CHEMICAL SCIENCES

Theory Credits: 4 5 hrs/week

Course Objective:

The objective of this course is to provide students with a comprehensive understanding of the essential concepts and applications of mathematical, physical, and chemical sciences. The course aims to develop students’ critical thinking, problem-solving, and analytical skills in these areas, enabling them to apply scientific principles to real-world situations.

Learning outcomes:

1. Apply critical thinking skills to solve complex problems involving complex numbers, trigonometric ratios, vectors, and statistical measures.
2. To Explain the basic principles and concepts underlying a broad range of fundamental areas of physics and to connect their knowledge of physics to everyday situations.
3. To Explain the basic principles and concepts underlying a broad range of fundamental areas of chemistry and to connect their knowledge of chemistry to daily life.
4. Understand the interplay and connections between mathematics, physics, and chemistry in various applications. Recognize how mathematical models and physical and chemical principles can be used to explain and predict phenomena in different contexts.
5. To explore the history and evolution of the Internet and to gain an understanding of network security concepts, including threats, vulnerabilities, and countermeasures.

UNIT I: ESSENTIALS OF MATHEMATICS:

Complex Numbers: Introduction of the new symbol i – General form of a complex number – Modulus-Amplitude form and conversions

Trigonometric Ratios: Trigonometric Ratios and their relations – Problems on calculation of angles

Vectors: Definition of vector addition – Cartesian form – Scalar and vector product and problems

Statistical Measures: Mean, Median, Mode of a data and problems

UNIT II: ESSENTIALS OF PHYSICS:

Definition and Scope of Physics- Measurements and Units - Motion of objects: Newtonian Mechanics and relativistic mechanics perspective - Laws of Thermodynamics and Significance- Acoustic waves and electromagnetic waves- Electric and Magnetic fields and their interactions- Behaviour of atomic and nuclear particles- Wave-particle duality, the uncertainty principle- Theories and understanding of universe

UNIT III: ESSENTIALS OF CHEMISTRY:

Definition and Scope of Chemistry- Importance of Chemistry in daily life - Branches of chemistry and significance- Periodic Table-
Electronic Configuration, chemical changes, classification of matter, Biomolecules- carbohydrates, proteins, fats and vitamins.

**UNIT IV: APPLICATIONS OF MATHEMATICS, PHYSICS & CHEMISTRY:**

**Applications of Mathematics in Physics & Chemistry:** Calculus, Differential Equations & Complex Analysis

**Application of Physics in Industry and Technology:** Electronics and Semiconductor Industry, Robotics and Automation, Automotive and Aerospace Industries, Quality Control and Instrumentation, Environmental Monitoring and Sustainable Technologies.

**Application of Chemistry in Industry and Technology:** Chemical Manufacturing, Pharmaceuticals and Drug Discovery, Materials Science, Food and Beverage Industry.

**UNIT V: ESSENTIALS OF COMPUTER SCIENCE:**

Milestones of computer evolution - Internet, history, Internet Service Providers, Types of Networks, IP, Domain Name Services, applications.

**Ethical and social implications:** Network and security concepts- Information Assurance Fundamentals, Cryptography-Symmetric and Asymmetric, Malware, Firewalls, Fraud Techniques- Privacy and Data Protection

**Recommended books:**

1. Functions of one complex variable by John B. Conway, Springer-Verlag
2. Elementary Trigonometry by H.S. Hall and S.R. Knight
4. Basic Statistics by B.L. Agarwal, New age international Publishers
5. University Physics with Modern Physics by Hugh D. Young and Roger A. Freedman
6. Fundamentals of Physics by David Halliday, Robert Resnick, and Jearl Walker
7. Physics for Scientists and Engineers with Modern Physics" by Raymond A. Serway and John W. Jewett Jr.
8. Physics for Technology and Engineering" by John Bird
9. Chemistry in daily life by Kirpal Singh
10. Chemistry of bio molecules by S. P. Bhutan
11. Fundamentals of Computers by V. Raja Raman
STUDENT ACTIVITIES

UNIT I: ESSENTIALS OF MATHEMATICS:

1: Complex Number Exploration
Provide students with a set of complex numbers in both rectangular and polar forms.
They will plot the complex numbers on the complex plane and identify their properties.

2: Trigonometric Ratios Problem Solving
Give students a set of problems that require the calculation of trigonometric ratios and their relations.
Students will solve the problems using the appropriate trigonometric functions (sine, cosine, tangent, etc.) and trigonometric identities.

3: Vector Operations and Applications
Provide students with a set of vectors in Cartesian form.
Students will perform vector addition and subtraction operations to find the resultant vectors.
They will also calculate the scalar and vector products of given vectors.

4: Statistical Measures and Data Analysis
Give students a dataset containing numerical values.
Students will calculate the mean, median, and mode of the data, as well as other statistical measures if appropriate (e.g., range, standard deviation).
They will interpret the results and analyze the central tendencies and distribution of the data.

UNIT II: ESSENTIALS OF PHYSICS:

1. Concept Mapping
Divide students into groups and assign each group one of the topics.
Students will create a concept map illustrating the key concepts, relationships, and applications related to their assigned topic.
Encourage students to use visual elements, arrows, and labels to represent connections and interdependencies between concepts.

2. Laboratory Experiment
Select a laboratory experiment related to one of the topics, such as motion of objects or electric and magnetic fields.
Provide the necessary materials, instructions, and safety guidelines for conducting the experiment.
Students will work in small groups to carry out the experiment, collect data, and analyze the results.
After the experiment, students will write a lab report summarizing their findings, observations, and conclusions.

UNIT III: ESSENTIALS OF CHEMISTRY

1: Chemistry in Daily Life Presentation

Divide students into groups and assign each group a specific aspect of daily life where chemistry plays a significant role, such as food and nutrition, household products, medicine, or environmental issues.

Students will research and create a presentation (e.g., PowerPoint, poster, or video) that showcases the importance of chemistry in their assigned aspect.

2: Periodic Table Exploration

Provide students with a copy of the periodic table.

Students will explore the periodic table and its significance in organizing elements based on their properties. They will identify and analyze trends in atomic structure, such as electronic configuration, atomic size, and ionization energy.

3: Chemical Changes and Classification of Matter

Provide students with various substances and chemical reactions, such as mixing acids and bases or observing a combustion reaction. Students will observe and describe the chemical changes that occur, including changes in color, temperature, or the formation of new substances.

4: Biomolecules Investigation

Assign each student or group a specific biomolecule category, such as carbohydrates, proteins, fats, or vitamins.

Students will research and gather information about their assigned biomolecule category, including its structure, functions, sources, and importance in the human body.

They can create informative posters or presentations to present their findings to the class.

UNIT IV: APPLICATIONS OF MATHEMATICS, PHYSICS & CHEMISTRY

1: Interdisciplinary Case Studies

Divide students into small groups and provide them with interdisciplinary case studies that involve the interdisciplinary application of mathematics, physics, and chemistry.

Each case study should present a real-world problem or scenario that requires the integration of concepts from all three disciplines.

2: Design and Innovation Project

Challenge students to design and develop a practical solution or innovation that integrates mathematics, physics, and chemistry principles.

Students can choose a specific problem or area of interest, such as renewable energy, environmental conservation, or materials science.

3: Laboratory Experiments
Assign students laboratory experiments that demonstrate the practical applications of mathematics, physics, and chemistry.

Examples include investigating the relationship between concentration and reaction rate, analyzing the behavior of electrical circuits, or measuring the properties of materials.

.4: Mathematical Modeling

Present students with real-world problems that require mathematical modeling and analysis.

UNIT V: ESSENTIALS OF COMPUTER SCIENCE:

1. Identifying the attributes of network (Topology, service provider, IP address and bandwidth of your college network) and prepare a report covering network architecture.
2. Identify the types of malwares and required firewalls to provide security.
3. Latest Fraud techniques used by hackers.
SEMESTER-I

COURSE 2: ADVANCES IN MATHEMATICAL, PHYSICAL AND CHEMICAL SCIENCES

Theory Credits: 4 5 hrs/week

Course Objective:

The objective of this course is to provide students with an in-depth understanding of the recent advances and cutting-edge research in mathematical, physical, and chemical sciences. The course aims to broaden students' knowledge beyond the foundational concepts and expose them to the latest developments in these disciplines, fostering critical thinking, research skills, and the ability to contribute to scientific advancements.

Learning outcomes:

1. Explore the applications of mathematics in various fields of physics and chemistry, to understand how mathematical concepts are used to model and solve real-world problems.
2. To Explain the basic principles and concepts underlying a broad range of fundamental areas of physics and to Connect their knowledge of physics to everyday situations.
3. Understand the different sources of renewable energy and their generation processes and advances in nanomaterials and their properties, with a focus on quantum dots. To study the emerging field of quantum communication and its potential applications. To gain an understanding of the principles of biophysics in studying biological systems. Explore the properties and applications of shape memory materials.
4. Understand the interplay and connections between mathematics, physics, and chemistry in various advanced applications. Recognize how mathematical models and physical and chemical principles can be used to explain and predict phenomena in different contexts.
5. Understand and convert between different number systems, such as binary, octal, decimal, and hexadecimal. Differentiate between analog and digital signals and understand their characteristics. Gain knowledge of different types of transmission media, such as wired (e.g., copper cables, fiber optics) and wireless (e.g., radio waves, microwave, satellite).

UNIT I: ADVANCES IN BASICS MATHEMATICS

Straight Lines: Different forms – Reduction of general equation into various forms – Point of intersection of two straight lines

Limits and Differentiation: Standard limits – Derivative of a function – Problems on product rule and quotient rule

Integration: Integration as a reverse process of differentiation – Basic methods of integration
Matrices: Types of matrices – Scalar multiple of a matrix – Multiplication of matrices – Transpose ofa matrix and determinants

UNIT II: ADVANCES IN PHYSICS:

Renewable energy: Generation, energy storage, and energy-efficient materials and devices. Recent advances in the field of nanotechnology: Quantum dots, Quantum Communication-recent advances in biophysics-recent advances in medical physics- Shape Memory Materials.

UNIT III: ADVANCES IN CHEMISTRY:

Computer aided drug design and delivery, nano sensors, Chemical Biology, impact of chemical pollutants on ecosystems and human health, Dye removal - Catalysis method

UNIT IV: ADVANCED APPLICATIONS OF MATHEMATICS, PHYSICS & CHEMISTRY

Mathematical Modelling applications in physics and chemistry
Application of Renewable energy: Grid Integration and Smart Grids,
Application of nanotechnology: Nanomedicine,
Application of biophysics: Biophysical Imaging, Biomechanics, Neurophysics,
Application of medical physics: Radiation Therapy, Nuclear medicine

UNIT V: Advanced Applications of computer Science

Number System-Binary, Octal, decimal, and Hexadecimal, Signals-Analog, Digital, Modem, Codec, Multiplexing, Transmission media, error detection and correction- Parity check and CRC, Networking devices- Repeater, hub, bridge, switch, router, gateway.

Recommended books:
2. Calculus by Thomas and Finny, Pearson Publications
4. "Renewable Energy: Power for a Sustainable Future" by Godfrey Boyle
6. "Nanotechnology: Principles and Applications" by Sulabha K. Kulkarni and Raghvendra A. Bohara
7. "Biophysics: An Introduction" by Rodney Cotterill
8. "Medical Physics: Imaging" by James G. Webster
9. "Shape Memory Alloys: Properties and Applications" by Dimitris C. Lagoudas
10. Nano materials and applications by M.N.Borah
11. Environmental Chemistry by Anil K. D. E.
12. Digital Logic Design by Morris Mano

STUDENT ACTIVITIES

UNIT I: ADVANCES IN BASIC MATHEMATICS

1: Straight Lines Exploration
Provide students with a set of equations representing straight lines in different forms, such as slope-intercept form, point-slope form, or general form.

Students will explore the properties and characteristics of straight lines, including their slopes, intercepts, and point of intersection.

2: Limits and Differentiation Problem Solving
Students will apply the concept of limits to solve various problems using standard limits.

Encourage students to interpret the results and make connections to real-world applications, such as analyzing rates of change or optimizing functions.

3: Integration Exploration
Students will explore the concept of integration as a reverse process of differentiation and apply basic methods of integration, such as the product rule, substitution method, or integration by parts.

Students can discuss the significance of integration in various fields, such as physics and chemistry

4: Matrices Manipulation
Students will perform operations on matrices, including scalar multiplication, matrix multiplication, and matrix transpose.

Students can apply their knowledge of matrices to real-world applications, such as solving systems of equations or representing transformations in geometry.

UNIT II: ADVANCES IN PHYSICS:

1: Case Studies
Provide students with real-world case studies related to renewable energy, nanotechnology, biophysics, medical physics, or shape memory materials.

Students will analyze the case studies, identify the challenges or problems presented, and propose innovative solutions based on the recent advances in the respective field.

They will consider factors such as energy generation, energy storage, efficiency, sustainability, materials design, biomedical applications, or technological advancements.

2: Experimental Design
Assign students to design and conduct experiments related to one of the topics: renewable
energy, nanotechnology, biophysics, medical physics, or shape memory materials. They will identify a specific research question or problem to investigate and design an experiment accordingly.

Students will collect and analyze data, interpret the results, and draw conclusions based on their findings. They will discuss the implications of their experimental results in the context of recent advances in the field.

3: Group Discussion and Debate

Organize a group discussion or debate session where students will discuss the ethical, social, and environmental implications of the recent advances in renewable energy, nanotechnology, biophysics, medical physics, and shape memory materials.

Assign students specific roles, such as proponent, opponent, or moderator, and provide them with key points and arguments to support their positions.

UNIT III: ADVANCES IN CHEMISTRY:

1. Experimental Design and Simulation

In small groups, students will design experiments or simulations related to the assigned topic. For example, in the context of computer-aided drug design, students could design a virtual screening experiment to identify potential drug candidates for a specific disease target.

For nano sensors, students could design an experiment to demonstrate the sensitivity and selectivity of nano sensors in detecting specific analytes. Chemical biology-related activities could involve designing experiments to study enzyme-substrate interactions or molecular interactions in biological systems. Students will perform their experiments or simulations, collect data, analyze the results, and draw conclusions based on their findings.

2. Case Studies and Discussion

Provide students with real-world case studies related to the impact of chemical pollutants on ecosystems and human health. Students will analyze the case studies, identify the sources and effects of chemical pollutants, and propose mitigation strategies to minimize their impact.

Encourage discussions on the ethical and environmental considerations when dealing with chemical pollutants. For the dye removal using the catalysis method, students can explore case studies where catalytic processes are used to degrade or remove dyes from wastewater. Students will discuss the principles of catalysis, the advantages and limitations of the catalysis method, and its applications in environmental remediation.

3: Group Project
Assign students to work in groups to develop a project related to one of the topics.
The project could involve designing a computer-aided drug delivery system, developing an nano sensor for a specific application, or proposing strategies to mitigate the impact of chemical pollutants on ecosystems. 
Students will develop a detailed project plan, conduct experiments or simulations, analyze data, and present their findings and recommendations.
Encourage creativity, critical thinking, and collaboration throughout the project.

UNIT IV: ADVANCED APPLICATIONS OF MATHEMATICS, PHYSICS & CHEMISTRY

1: Mathematical Modelling Experiment

Provide students with a mathematical modelling experiment related to one of the topics. For example, in the context of renewable energy, students can develop a mathematical model to optimize the placement and configuration of solar panels in a solar farm. 
Students will work in teams to design and conduct the experiment, collect data, and analyze the results using mathematical models and statistical techniques. 
They will discuss the accuracy and limitations of their model, propose improvements, and interpret the implications of their findings in the context of renewable energy or the specific application area.

2: Case Studies and Group Discussions

Assign students to analyze case studies related to the applications of mathematical modelling in nanotechnology, biophysics, medical physics, solid waste management, environmental remediation, or water treatment. 
Students will discuss the mathematical models and computational methods used in the case studies, analyze the outcomes, and evaluate the effectiveness of the modelling approach. 
Encourage group discussions on the challenges, ethical considerations, and potential advancements in the field.

Students will present their findings and engage in critical discussions on the advantages and limitations of mathematical modelling in solving complex problems in these areas.

3. Group Project

Assign students to work in groups to develop a group project that integrates mathematical modelling with one of the application areas: renewable energy, nanotechnology, biophysics, medical physics, solid waste management, environmental remediation, or water treatment. 
The project could involve developing a mathematical model to optimize the delivery of radiation therapy in medical physics or designing a mathematical model to optimize waste management practices.
Students will plan and execute their project, apply mathematical modelling techniques, analyze the results, and present their findings and recommendations. 
Encourage creativity, critical thinking, and collaboration throughout the project.
UNIT V: Advanced Applications of computer Science
Students must be able to convert numbers from other number system to binary numbersystems

1. Identify the networking media used for your college network

2. Identify all the networking devices used in your college premises.
Aim and objectives of Course:

Data Science is a fast-growing interdisciplinary field, focusing on the analysis of data to extract knowledge and insight. This course will introduce students to the collection, preparation, analysis, modelling and visualization of data, covering both conceptual and practical issues. Examples and case studies from diverse fields will be presented, and hands-on use of statistical and data manipulation software will be included.

Learning outcomes of Course:
- Recognize the various disciplines that contribute to a successful data science effort.
- Understand the processes of data science identifying the problem to be solved, data collection, preparation, modeling, evaluation, and visualization.
- Be aware of the challenges that arise in Data Sciences.
- Be able to identify the application of the type of algorithm based on the type of the problem.
- Be comfortable using commercial and open source tools such as the R/Python language and its associated libraries for data analytics and visualization.

UNIT I:
Defining Data Science and Big data, Benefits and Uses, facets of Data, Data Science Process. History and Overview of R, Getting Started with R, R Nuts and Bolts

UNIT II:
The Data Science Process: Overview of the Data Science Process - Setting the research goal, Retrieving Data, Data Preparation, Exploration, Modeling, data Presentation and Automation. Getting Data in and out of R, Using reader package, Interfaces to the outside world.

UNIT III:
Machine Learning: Understanding why data scientists use machine learning - What is machine learning and why we should care about, Applications of machine learning in data science, Where it is used in data science, The modeling process, Types of Machine Learning - Supervised and Unsupervised.

UNIT IV:
Handling large Data on a Single Computer: The problems we face when handling large data, General Techniques for handling large volumes of data, Generating programming tips for dealing with large datasets.
UNIT V:
Sub setting R objects, Vectorised Operations, Managing Data Frames with the dplyr, Control structures, functions, Scoping rules of R, Coding Standards in R, Loop Functions, Debugging, Simulation. Case studies on preliminary data analysis.

TEXT BOOKS:

REFERENCE BOOKS:

Web References for case studies:
1. https://www.kaggle.com/datasets
2. https://github.com/
SEMESTER-II

COURSE 3: INTRODUCTION TO DATA SCIENCE AND R PROGRAMMING

Practical Credits: 1 2 hrs/week

Lab/Practical/Experiments/Tutorials syllabus:

1. Installing R and R studio, with proper notes on version management, cosmetic settings and different libraries.
2. Basic operations in r with arithmetic and statistics.
3. Getting data into R, Basic data manipulation, Loading Data into R
4. Basic plotting
5. Loops and functions
6. Create Vectors, Lists, Arrays, Matrices, Data frames and operations on them.
7. Demonstrate the visualization and graphics using visualization packages like ggplot2.
8. Implement Loop functions with lapply(), sapply(), tapply(), apply(), mapply().
10. Explore data using two Variables: Line plots, Scatter Plots, smoothing curves, Bar charts
11. Explore and implement commands using dplyr package
12. Download a dataset and work on basic data manipulation followed by inferential statistics.

RECOMMENDED TEXT BOOKS:

Recommended Reference books:

Recommended Co-curricular activities: (Co-curricular Activities should not promote copying from text book or from others’ work and shall encourage self/independent and group learning)

A. Measurable:
1. Assignments on:
2. Student seminars (Individual presentation of papers) on topics relating to:
3. Quiz Programmes on:
4. Individual Field Studies/projects:
5. Group discussion on:
6. Group/Team Projects on:
B. General
1. Collection of news reports and maintaining a record of paper-cuttings relating to topics covered in syllabus
2. Group Discussions on:
3. Watching TV discussions and preparing summary points recording personal observations etc., under guidance from the Lecturers
4. Any similar activities with imaginative thinking.
5. Recommended Continuous Assessment methods:
SEMESTER-II

COURSE 4: DESCRIPTIVE STATISTICS

Theory: 3  
Credits: 3  
3 hrs/week

Course Learning Outcomes: Students will acquire:
• knowledge of Statistics and its implementation through practical understanding for various domains related to data science.
• knowledge of various types of data, their organization and evaluation of summary measures such as measures of central tendency and dispersion etc.
• knowledge of other types of data reflecting quality characteristics including concepts of independence and association between two attributes,
• insights into preliminary exploration of different types of data.
• Knowledge of correlation, regression analysis, regression diagnostics, partial and multiple correlations.

UNIT I:

UNIT II:

UNIT III:
Curve fitting: Bi-variate data, Principle of least squares, fitting of degree polynomial. Fitting of straight line, Fitting of Second degree polynomial or parabola, Fitting of power curve and exponential curves.

Correlation: Meaning, Types of Correlation, Measures of Correlation: Scatter diagram, Karl Pearson’s Coefficient of Correlation, Rank Correlation Coefficient (with and without ties), Bi-variate frequency distribution, correlation coefficient for bi-variate data and simple problems. Concept of multiple and partial correlation coefficients (three variables only) and properties

UNIT IV:
Regression: Concept of Regression, Linear Regression: Regression lines, Regression coefficients and it’s properties, Regressions lines for bi-variate data and simple problems. Correlation vs regression, sigmoid curve, derivation from linear regression to logistic regression.
UNIT-V

**Attributes**: Notations, Class, Order of class frequencies, Ultimate class frequencies, Consistency of data, Conditions for consistency of data for 2 and 3 attributes only, Independence of attributes, Association of attributes and its measures, Relationship between association and colligation of attributes, Contingency table: Square contingency, Mean square contingency, Coefficient of mean square contingency,

**TEXT BOOKS**: 


2. BA/BSc I year statistics - descriptive statistics, probability distribution - Telugu Academy - Dr M. Jaganmohan Rao, Dr N. Srinivasa Rao, Dr P. Tirupathi Rao, Smt. D. Vijayalakshmi.


**REFERENCE BOOKS**: 


SEMESTER-II

COURSE 4: DESCRIPTIVE STATISTICS

Practical

Credits: 1

List of the experiments:

1. Graphical presentation of data (Histogram, frequency polygon).
2. Diagrammatic presentation of data (Bar and Pie).
3. Computation of measures of central tendency (Mean, Median and Mode).
5. Computation of non-central, central moments, $\sigma^1$ and $\sigma^2$ for ungrouped data.
6. Computation of Karl Pearson’s coefficients of Skewness and Bowley’s coefficients of Skewness.
7. Fitting of straight line by the method of least squares.
8. Fitting of parabola by the method of least squares.
9. Fitting of power curve of the type by the method of least squares.
10. Fitting of exponential curve of the type and by the method of least squares.
11. Computation of correlation coefficient and regression lines for ungrouped data.
12. Computation of correlation coefficient, forming regression lines for grouped data.
SEMESTER-III

COURSE 5: PYTHON PROGRAMMING FOR DATA ANALYSIS

Theory Credits: 3 3 hrs/week

Aim and objectives of Course:
- To be able to Program in Python
- To know and understand the data Analysis phases
- To know the usage of all libraries

Learning outcomes of Course:
- Understands and learn all basic concepts of
- Python Program Data Analysis methods in Python
- Get used with Python Programming environments

UNIT I:

UNIT II:

UNIT III:
**Getting Started with Pandas**: Introduction to Pandas Data Structures, Essential Functionality, Summarizing and Computing Descriptive Statistics

Data Loading, Storage and File Formats: Reading and Writing Data in TextFormat, Binary Data Formats, Interacting with Web APIs, Interacting with Databases.

UNIT IV:
**Data Cleaning and Preparation**: Handling Missing Data, Data Transformation, String Manipulation.

**Data Wrangling**: Join, Combine and Reshape: Hierarchical Indexing, Combining and Merging Datasets, Reshaping and Pivoting.

UNIT V:
**Introduction to Modeling Libraries in Python**: Interfacing between pandas and Model code, Creating model descriptions with Patsy, Introduction to stats models.

**Plotting and Visualization**: A brief matplotlib API Primer, Plotting with Pandas and Seaborn, Other Python visualization tools.

TEXT BOOKS:
2. Charles R Suverance “Python for Everybody” Exploring data using Python 3

REFERENCE BOOKS:

Co-curricular Activities
Take up any application which involves the python coding. Example Case studies/Simulators: (https://knightlab.northwestern.edu/2014/06/05/five-mini-programming-projects-for-the-python-beginner/)

- Dice Rolling Simulator
- Guess the number
- Text based adventure game
- Hangman

Continuous assessment:
Let the students be tested in the following questions from each unit

1. What is Data Analysis. List out the differences between data analysis and data analytics
2. What is Python? Explain Python basics
3. Explain NumPy Basics
4. What is Data Loading. Explain Pandas Data Structures
5. What is Data Cleaning. Explain different phases in it
6. Explain Plotting and Visualization in Python
SEMMESTER-III

COURSE 5: PYTHON PROGRAMMING FOR DATA ANALYSIS

Practical Credits: 1 2 hrs/week

1. Use matplotlib and plot an inline in Jupyter.
2. Implement commands of Python Language basics
3. Create Tuples, Lists and illustrate slicing conventions.
4. Create built-in sequence functions.
5. Clean the elements and transform them by using List, Set and DictComprehensions.
6. Create a functional pattern to modify the strings in a high level.
7. Write a Python Program to cast a string to a floating-point number but fails with Value Error on improper inputs using Errors and Exception handling.
8. Create an n array object and use operations on it.
9. Use arithmetic operations on Numpy Arrays
10. Using Numpy array perform Indexing and Slicing Boolean Indexing, FancyIndexing operations
11. Create an image plot from a two-dimensional array of function values.
12. Implement some basic array statistical methods (sum, mean, std, var, min,max, argmin, argmax, cumsum and cumprod) and sorting with sortmethod.
13. Implement numpy.random functions.
14. Plot the first 100 values on the values obtained from random walks.
15. Create a data frame using pandas and retrieve the rows and columns in it by performing some indexing options and transpose it.
16. Implement the methods of descriptive and summary statistics
17. Load and write the data from and to different file formats including WebAPIs.
18. Implement the data Cleaning and Filtering methods(Use NA handling methods, fillna function arguments)
19. Transform the data using function or mapping
20. Rearrange the data using unstack method of hierarchical Indexing
21. Implement the methods that summarize the statistics by levels.
22. Use different Join types with how argument and merge data with keys and multiple keys.
SEMESTER-III

COURSE 6: INFERENTIAL AND APPLIED STATISTICS

Course Learning Outcomes
After completion of this course, the students will know about

• Concept of law large numbers and their uses
• Knowledge about important inferential aspects such as point estimation, test of hypotheses and associated concepts,
• Knowledge about inferences from Binomial, Poisson and Normal distributions as illustrations,
• Concept about non-parametric method and some important non-parametric tests.

• Time series data, its applications to various fields and components of time series,
• Various data collection methods enabling to have a better insight in policy making, planning and systematic implementation, Construction and implementation of life tables, Population growth curves, population estimates and projections,
• Real data implementation of various demographic concepts as outlined above through practical assignments.

UNIT I:
Concepts: Population, Sample, Parameter, statistic, Sampling distribution, Standard error. convergence in probability and convergence in distribution, law of large numbers, central limit theorem (statements only). Student’s t-distribution, F-Distribution, $\chi^2$-Distribution: Definitions, properties and their applications.

UNIT II:

UNIT III:
Sample tests: t-test for single mean, difference of means and paired t-test. 2. confidence intervals for mean(s). standard deviation(s) and correlation coefficient(s). Test for goodness of fit and independence of attributes. F-test for equality of variances.
Non-parametric tests- their advantages and disadvantages, comparison with parametric tests. Measurement scale- nominal, ordinal, interval and ratio.

UNIT IV:
UNIT V:

TEXT BOOKS:
1. BA/BSc II year statistics - statistical methods and inference - Telugu Academy by A.Mohanrao, N.Srinivasa Rao, Dr R.Sudhakar Reddy, Dr T.C. RavichandraKumar.
2. K.V.S. Sarma: Statistics Made Simple: Do it yourself on PC.PHI.

REFERENCE BOOKS:

**CO-CURRICULAR ACTIVITIES:**
- Quiz Competition
- Expert Lectures
- Seminars

**EXTRA CURRICULAR ACTIVITIES:**
- Formal Examination
- Lab Practical
- Presentation
- Simple Projects
SEMESTER-III

COURSE 6: INFERENTIAL AND APPLIED STATISTICS

<table>
<thead>
<tr>
<th>Theory</th>
<th>Credits: 3</th>
<th>3 hrs/week</th>
</tr>
</thead>
</table>

List of Experiments:

1. Large sample test for difference of means.
2. Large sample test for single proportion
3. Large sample test for difference of proportions, standard deviations, correlation coefficient.
4. Small sample test for single mean, difference of means and correlation coefficient
5. Paired t-test(paired samples).
6. Small sample test for single variance(χ²-test)
   Time Series:
7. Measurement of trend by method of moving averages (odd and even period)
8. Measurement of trend by method of Least squares (linear and parabola)
9. Determination of seasonal indices by method simple averages
10. Determination of seasonal indices by method of Ratio to moving averages

Vital Statistics:

11. Computation of various Mortality rates
12. Computation of various Fertility rates
14. Construction of Life Tables
SEMMESTER-III

COURSE 7: DATA MINING TECHNIQUES USING R

Theory

Credits: 3

3 hrs/week

Aim and objectives of Course:
• To understand Data mining techniques and algorithms.
• Comprehend the data mining environments and application.

Learning outcomes of Course:
Students who complete this course will be able to
• Compare various conceptions of data mining as evidenced in both research and application.
• Evaluate mathematical methods underlying the effective application of data mining.
• Should be able to apply the type of techniques based on the problems considered.
• Can find out the market patterns and association amongst different products.

UNIT I:
An idea on Data Warehouse, Data mining-KDD versus data mining, Stages of the Data Mining Process-Task primitives., Data Mining Techniques – Data mining knowledge representation.

UNIT II
Data mining query languages- Integration of Data Mining System with a Data Warehouse- Issues, Data pre-processing – Data Cleaning, Data transformation – Feature selection – Dimensionality reduction

UNIT III
Concept Description: Characterization and comparison What is Concept Description, Data Generalization by Attribute-Oriented Induction(AOI), AOI for Data Characterization, Efficient Implementation of AOI.

Mining Frequent Patterns, Associations and Correlations: Basic Concepts, FrequentItemset Mining Methods: Apriori method, generating Association Rules, Improving the Efficiency of Apriori, Pattern-Growth Approach for mining Frequent Item sets.

UNIT-IV
Classification Basic Concepts: Basic Concepts, Decision Tree Induction: Decision Tree Induction Algorithm, Attribute Selection Measures, Tree Pruning. Bayes Classification Methods.

UNIT-V
Association rule mining: Antecedent, consequent, muti-relational association rules, ECLAT. Case study on Market Basket Analysis.

Cluster Analysis: Cluster Analysis, Partitioning Methods, Hierarchical methods, Density based methods-DBSCAN.
TEXT BOOKS:

REFERENCES BOOKS:

Student Activities:
1. Students should be able to implement Data Mining algorithms provided the relevant data
2. Given the data, students can visualize all statistical measures
3. Differentiate the types of mining problems and identify what type of algorithms are to be implemented.

Continuous assessment:
Let the students be tested in the following questions from each unit
1. What is Data Mining and KDD? Where Data Mining fits in KDD Process
2. Describe all Preprocessing methods
3. Explain Data Description and AOI Algorithm
4. Explain Classification and Write any Decision tree induction algorithm
5. Explain the concept of clustering and write any algorithm to form clusters.
2. Visualize all Statistical measures (Mean, Mode, Median, Range, InterQuartile Range etc., using Histograms, Boxplots and Scatter Plots).
3. Create a data frame with at least 10 entries of columns EMPID, EMPNAME, SALARY, STARTDATE
   a. Extract two column names using column name.
   b. Extract the first two rows and then all columns.
   c. Extract 3rd and 5th row with 2nd and 4th column.
4. Create a data frame with 10 observations and 3 variables and add new rows and columns to it using ‘rbind’ and ‘cbind’ function.
5. Create a function to discretize a numeric variable into 3 quantiles and label them as low, medium, and high. Apply it on each attribute of any dataset to create a new data frame. ‘discrete’ with Categorical variables and the class label.
6. Create a simple scatter plot using any dataset using ‘dplyr’ library. Use the same data to indicate distribution densities using box whiskers.
7. Write R Programs to implement k-means clustering, k-medoids clustering and density based clustering on any datasets.
8. Write a R Program to implement decision trees using ‘reading Skills’ dataset.
10. Generate top 5 association rules using apriori.
11. Generate top 5 association rules using ECLAT.
12. Write an R program to implement Naïve bayes Classification.
SEMESTER-III

COURSE 8: WEB TECHNOLOGIES

Theory Credtis: 3 3 hrs/week

COURSE OBJECTIVES: This subject enables the student to create flexible, attractive, user-friendly web sites comprised of both static and dynamic web pages. Along with that students will also learn about interactions with web pages through JavaScript and host own web site on internet.

LEARNING OUTCOMES: After Studying this subject students would have capability to make their own web site and host on internet. Also students would have enough knowledge about the technologies used in internet.


UNIT II: Cascading Style Sheets: Introduction, Syntax, Selectors, Background Cursors, Text Fonts, Lists, Tables, Box Model, Using Styles, Simple Examples, Creation of Own Styles, Properties And Values In Styles, Formatting Blocks of Information, Layers.


UNIT IV: DHTML with JavaScript: Data Validation, Opening A New Window, Messages and Confirmations, Status Bar, Different Frames, Rollover Buttons, Moving Images.


Co-curricular Activities:

- We for Web – Students with right mix of skills are formed as groups to develop websites.
- Web Ninja- A platform to showcase creative websites developed by students to their peers.

Assessment Methods:

- Formal Examinations.
- Lab Practical Examination.
- Presentations.
- Simple Project.
SEMESTER-III
COURSE 8: WEB TECHNOLOGIES

Practical

Credits: 1

2 hrs/week

1. Design web pages for your college containing a description of the courses, departments, faculties, library etc, use href, list tags.
2. Create your class timetable using table tag.
3. Create a feedback form for your curriculum. Use textbox, text area, checkbox, radio button etc.
4. Create a web page using frame. Divide the page into two parts with Navigation links on lefthand side of page (width=20%) and content page on right hand side of page (width = 80%). On clicking the navigation Links corresponding content must be shown on the right hand side.
5. Write html code to develop a webpage having two frames that divide the webpage into two equal rows and then divide the row into equal columns fill each frame with a different background colour.
6. Create your resume using HTML tags. Experiment with colours, text, link, size and also other tags you studied.
7. Design a web page of your College Day Celebrations with an attractive background colour, text colour, images, font etc. Use CSS.
8. Use Inline CSS to format your resume that you created.
9. Use External CSS to format your class timetable as you created.
10. Use External, Internal, and Inline CSS to format web page of your start up.
11. Develop a JavaScript to display your admission details in the college.
12. Develop simple calculator for addition, subtraction, multiplication and division operation using JavaScript.
13. Create HTML page with JavaScript which takes integer number as input and tells whether the number is odd or even.
14. Create HTML page that contains form for registration of your participation in a hackathon. Use relevant fields for input data. Write a JavaScript code to combine and display the input information when the button is clicked.
15. Create a login form with id and password. Perform input validation.
SEMESTER-IV

COURSE 9: DATA VISUALIZATION

Aim and objectives of Course:

- To know the importance of data Visualization in the world of Data Analytics and Prediction
- To know the important libraries in Tableau
- To get equipped with Tableau Tool

Learning outcomes of Course:

- Students should be able to visualize data through seven stages of data analysis process
- Should be able to do explanatory and hybrid types of data visualization
- Should be able to understand various stages of visualizing data

UNIT I:
Creating Visual Analytics with tableau desktop, connecting to your data- How to Connect to your data, What are generated Values? Knowing when to use a direct connection, Joining tables with tableau, blending different data sources in a single worksheet.

UNIT II:
**Building your first Visualization**- How Me works- Chart types, Text Tables, Maps, bar chart, Line charts, Area Fill charts and Pie charts, scatter plot, Bullet graph, Gantt charts, Sorting data in tableau, Enhancing Views with filters, sets groups and hierarchies.

UNIT III:
**Creating calculations to enhance your data**- What is aggregation, what are calculated values and table calculations, Using the calculation dialog box to create, Building formulas using table calculation functions

UNIT IV:
**Using maps to improve insights**- Create a Standard Map View, Plotting your own locations on a map, Replace Tableau’s standard maps, Shaping data to enable Point-to-Point mapping.

UNIT V:
**Developing an Adhoc analysis environment**- generating new data with forecasts, providing self evidence adhoc analysis with parameters, Editing views in tableau Server.

TEXT BOOKS:
1. Tableau your data-Daniel G. Murray and the Inter works BI team, Wiley Publications
2. Tableau Data Visualizaton Cookbook, AshutoshNandeshwar, PACKT publishing.
4. ggplot2: Elegant Graphics for Data Analysis by Hadley Wickham (2009)

REFERENCE BOOKS:
1. Designing Data Visualizations: Representing Informational Relationshipsby Noah Iliinsky, Julie Steele (2011)
Student Activity
Create a sample super store data set and visualize the following requirements

General Requirements
1. Dashboard size is 1250px wide by 750px tall.
2. Prefer using containers
3. The dashboard has a total of 5 containers (no more, no less)
4. The Filter Pane
5. Each filter has some padding

1. Charts Pane Requirement
1. All 3 charts must be in one vertical container
2. Do proper formatting
3. Each chart has some padding between them and other objects
4. Each chart has a grey border, slightly darker than the Pane background color.
5. The Pane under the Title has a border
2. The second graph should have the title as “Sales” and should show monthly sales per year. Make sure it is an area chart with proper formatting.
3. The third graph should the title as “Profit” and should show monthly profit per year. Make sure it is an area chart with proper formatting.

Continuous assessment:
Let the students be tested in the following questions from each unit
1. What are generated values? Join tables using Tableau
2. Create any visualization charts using Chart types, Text Tables, Maps, barchart, Line charts, Area Fill charts and Pie charts, scatter plot etc.,
3. What is aggregation, what are calculated values and table calculations?
4. Using Standard Map View, Plot your own locations on a map
5. Develop an Adhoc analysis environment.
SEMESTER-IV

COURSE 9: DATA VISUALIZATION

Practical Credits: 1 2 hrs/week

1. Connect to data Sources
2. Create Univariate Charts
3. Create Bivariate and Multivariate charts
4. Create Maps
5. Calculate user-defined fields
6. Create a workbook data extract
7. Save a workbook on a Tableau server and web
8. Export images, data.
SEMESTER-IV
COURSE 10: DATA VISUALIZATION USING PYTHON

| Theory | Credits: 3 | 3 hrs/week |

Course Objective:

This course introduces students to data analysis and visualization in the field of exploratory data science using Python.

Course Learning Outcomes: On successful completion of the course, the students will be able to

1. Use data analysis tools in the pandas library.
2. Load, clean, transform, merge and reshape data.
3. Create informative visualization and summarize data sets.
4. Analyze and manipulate time series data.
5. Solve real world data analysis problems.

Unit 1


Unit 2

Getting Started with Pandas: Arrays and vectorized computation, Introduction to pandas Data Structures, Essential Functionality, Summarizing and Computing Descriptive Statistics. Data Loading, Storage and File Formats. Reading and Writing Data in Text Format, Web Scraping, Binary Data Formats, Interacting with Web APIs,

Interacting with Databases Data Cleaning and Preparation. Handling Missing Data, Data Transformation, String Manipulation

Unit 3

Data Wrangling: Hierarchical Indexing, Combining and Merging Data Sets Reshaping and Pivoting. Data Visualization matplotlib: Basics of matplotlib, plotting with pandas and seaborn, other python visualization tools. Advanced categorical and numeric plots.
Unit 4

Data Aggregation and Group operations: Group by Mechanics, Data aggregation, General split-apply-combine, Pivot tables and cross tabulation

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

Unit 5 Advanced Pandas:

Categorical Data: cleaning data and visualization techniques, Advanced GroupBy methods, Use Techniques for Method Chaining.

Textbook:


Reference:
SEMESTER-IV
COURSE 10: DATA VISUALIZATION USING PYTHON

Practical Credits: 1 2 hrs/week

1. Practical based on NumPy ndarray
2. Practical based on Pandas Data Structures
3. Practical based on Data Loading, Storage and File Formats
4. Practical based on Interacting with Web APIs
5. Practical based on Data Cleaning and Preparation
6. Practical based on Data Wrangling
7. Practical based on Data Visualization using matplotlib
8. Practical based on Data Aggregation
9. Practical based on Time Series Data Analysis
SEMESTER-IV

COURSE 11: INTRODUCTION TO SQL & ADVANCED TABLEAU

Theory Credits: 3 3 hrs/week

Learning Objectives:

✓ Design a database using DBMS softwares.
✓ Perform SQL queries on database.
✓ Use Tableau’s visualization tools to conduct data analysis, especially exploration of an unfamiliar dataset.

Course Outcomes:

✓ Design a database by their own and perform simple and adhoc queries.
✓ Employ best practices in data visualization to develop charts, maps, tables, and other visual representations of data.
✓ Employ best practices in data visualization to develop charts, maps, tables, and other visual representations of data.
✓ Create compelling, interactive dashboards to combine several visualizations into a cohesive and functional whole.
✓ Utilize advanced Tableau features including parameters, data blending, custom SQL, very large data

UNIT 1:
Overview of Database Management System: Introduction to data, information, database, database management system, DBMS software’s,

, keys in DBMS. the building blocks of an entity relationship diagram, classification of entity sets, attribute classification, relationship degree, relationship classification, reducing ER diagram to tables,

UNIT 2:
Structured Query Language: Introduction, History of SQL Standard, Commands in SQL, Data Types in SQL, Data Definition Language, Data Manipulation Language, database constraints, Aggregate functions, Join Operation, Set Operations, Views. SQL queries, sub queries and corelated queries,
Unit 3: Optimal visualization types – bar chart, pie chart, gantt chart, bubble chart, bullet chart, scatter plot, line chart, heat map, tree map Maps- geographical locational plotting, Binning values, Calculated fields, Table calculations, Level of Detail calculations.

Unit 4: Dashboard development, Dashboard design principles, dashboard interactivity, Connected “drill-down” dashboards Best Practices, Creating visualizations with Tableau.

Unit 5: Advanced Tableau, Large datasets, Fiscal Year Calculations, Parameters, tableau scripting, tableau server, integration of tableau with R programming.

Textbooks:
1. Show me the Numbers: Designing Tables and Graphs to Enlighten by Stephen Few
2. The Data Loom: Weaving Understanding by Thinking Critically and Scientifically with Data by Stephen Few

Reference Books:
1. The Big Book of Dashboards: Visualizing your Data using Real-World Business Scenarios by Steve Wexler, Jeffrey Shaffer, and Andy Cotgreave
DATABASE MANAGEMENT SYSTEM LAB

Consider following databases convert entities and relationships to relation table for a given scenario.

1. COLLEGE DATABASE:
   STUDENT (stno, SName, Address, Phone, Gender) course(courseid, Sem, Sec)

   CLASS (stno, courseid)

   SUBJECT (Subcode, Title, Sem, Credits)

   MARKS (stno, Subcode, courseid, Test1, Test2, Test3, total)

2. COMPANY DATABASE:
   EMPLOYEE (SSN, Name, Address, Sex, Salary, SuperSSN, DNo) DEPARTMENT (DNo, DName, MgrSSN, MgrStartDate) DLOCATION (DNo, DLoc)

   PROJECT (PNo, PName, PLocation, DNo) WORKS_ON (SSN, PNo, Hours)

3. Consider a college database schema
a. Create above tables with relevant Primary Key, Foreign Key and other constraints

b. Populate the tables with data

4. Perform queries to generate outputs:

1. Display all the details of all employees working in the company.

2. Display ssn, lname, fname, address of employees who work in department no 7.

3. Retrieve the Birthdate and Address of the employee whose name is 'Franklin T. Wong'

4. Retrieve the name and salary of every employee.

5. Retrieve all distinct salary values

6. Retrieve all employee names whose address is in ‘Bellaire’

7. Retrieve all employees who were born during the 1950s

8. Retrieve all employees in department 5 whose salary is between 50,000 and 60,000 (inclusive)

5. Perform the following queries

1. Retrieve the names of all employees who do not have supervisors

2. Retrieve SSN and department name for all employees

3. Retrieve the name and address of all employees who work for the 'Research' department
4. For every project located in 'Stafford', list the project number, the controlling department number, and the department manager's last name, address, and birth date.

5. For each employee, retrieve the employee's name, and the name of his or her immediate supervisor.

6. Retrieve all combinations of Employee Name and Department Name.

7. Make a list of all project numbers for projects that involve an employee whose last name is 'Narayan' either as a worker or as a manager of the department that controls the project.

8. Increase the salary of all employees working on the 'ProductX' project by 15%. Retrieve employee name and increased salary of these employees.

9. Retrieve a list of employees and the project name each works in, ordered by the employee's department, and within each department ordered alphabetically by employee first name.

10. Select the names of employees whose salary does not match with the salary of any employee in the same department.

6. Perform following queries:

1. Retrieve the employee numbers of all employees who work on project located in Bellaire, Houston, or Stafford.

2. Find the sum of the salaries of all employees, the maximum salary, the minimum salary, and the average salary. Display with proper headings.

3. Find the sum of the salaries and number of employees of all employees of the ‘Marketing’ department, as well as the maximum salary, the minimum salary, and the average salary in this department.

4. Select the names of employees whose salary is greater than the average salary of all employees in department 10.

5. Delete all dependents of employee whose ssn is ‘123456789’.

6. Perform a query using alter command to drop/add field and a constraint in Employee table.

7. Format your data using filters with colors.
8. create dashboards and stories.

9. Distribute and publish your visualization.

10. create advanced mapping –
    1. point-to-point map
    2. Dual axis map

11. Calculate distance between two points on a map.
SEMESTER-V

COURSE 12: SUPERVISED ML WITH PYTHON

Aim and objectives of Course:

- The purpose of this course is to serve as an introduction to Supervised machine learning with Python.
- We will explore several classifications, regression algorithms and see how they can help us perform a variety of Supervised machine learning tasks.

Learning outcomes of Course:

- Able to understand introduction to machine learning concepts.
- Able to Loading datasets, build models and model persistence.
- Understand Feature extraction from data sets.
- Able to do Regression & Classification.
- Able to compare SVM with other classifiers.

UNIT I:
Machine Learning Basics: What is machine learning? Key terminology, Key tasks of machine learning, How to choose right algorithm, steps in developing a machine learning, why python?
Getting started with Numpy library Classifying with k- Nearest Neighbors: The k-Nearest Neighbors classification algorithm, Parsing and importing data from a text file, Creating scatter plots with Matplotlib, Normalizing numeric values

UNIT II:
Splitting datasets one feature at a time- Decision trees: Introducing decision trees, measuring consistency in a dataset, using recursion to construct a decision tree, plotting trees in Matplotlib

UNIT III:
Classifying with probability theory - Naïve Bayes: Using probability distributions for classification, learning the naïve Bayes classifier, Parsing data from RSS feeds, using naïve Bayes to reveal regional attitudes

UNIT IV:
Logistic regression: Classification with logistic regression and the sigmoid function, Using optimization to find the best regression coefficients, the gradient descent optimization algorithm, Dealing with missing values in our data

UNIT V:
Support vector machines: Introducing support vector machines, using the SMO algorithm for optimization, using kernels to “transform” data, Comparing support vector machines with other classifiers

TEXT BOOK:
1. Machine learning in action, Peter Harrington by Manning publications
2. Supervised ML with Python Lab
SEMESTER-V

COURSE 12: SUPERVISED ML WITH PYTHON

Practical ........................................ Credits: 1 ........................................ 2 hrs/week

Details of Lab/Practical/Experiments/Tutorials syllabus:

1. Implement and demonstrate the FIND-S algorithm for finding the most specific hypothesis based on a given set of training data samples. Read the training data from a CSV file. For a given set of training data examples stored in a .CSV file, implement and demonstrate the Candidate-Elimination algorithm to output a description of the set of all hypotheses consistent with the training examples.

2. Write a program to demonstrate the working of the decision tree based ID3 algorithm.

3. Write a program to implement the naïve Bayesian classifier for a sample training data set stored as a CSV file.

4. Assuming a set of documents that need to be classified, use the naïve BayesianClassifier model to perform this task. Built-in Java classes/API can be used to write the program. Calculate the accuracy, precision, and recall for your dataset.
SEMESTER-V

COURSE 13: UNSUPERVISED ML WITH PYTHON

<table>
<thead>
<tr>
<th>Theory</th>
<th>Credits: 3</th>
<th>3 hrs/week</th>
</tr>
</thead>
</table>

Aim and objectives of Course (Unsupervised ML with Python):
- Unsupervised Machine Learning involves finding patterns in datasets.
- The core of this course involves study of Clustering, feature extraction and optimization algorithms.
- The purpose of this course is to serve as an introduction to machine learning with Python.

Learning outcomes of Course:
- Able to do Clustering, feature extraction and optimization.
- Students will be able to understand and implement in Python algorithms of Unsupervised Machine Learning and apply them to real-world datasets.

Syllabus: (Total Hours: 90 including Teaching, Lab and internal exams, etc.)

UNIT I:
Unsupervised Learning: Clustering: k-means clustering algorithm, Improving cluster performance with post processing, Bisecting k-means, Example: clustering points on a map

UNIT II:
Association analysis: Apriori algorithm: Association analysis, The Apriori principle, Finding frequent item sets with the Apriori algorithm, Mining association rules from frequent item sets, uncovering patterns in congressional voting

UNIT III:
Finding frequent item sets: FP-growth–FP trees, Build FP-tree, mining frequent from an FP-tree, finding co-occurring words in a Twitter feed, mining a click stream from a news site.

UNIT IV:
Principal component analysis: Dimensionality reduction techniques, using PCA to reduce the dimensionality of semiconductor manufacturing data

UNIT V:
Singular value decomposition: Applications of the SVD, Matrix factorization, SVD in Python, Collaborative filtering–based recommendation engines, a restaurant dish recommendation engine

TEXT BOOK:
1. Machine learning in action, Peter Harrington by Manning publications
Unsupervised ML with Python Lab

SEMESTER-V

COURSE 13: UNSUPERVISED ML WITH PYTHON

<table>
<thead>
<tr>
<th>Theory</th>
<th>Credits: 3</th>
<th>3 hrs/week</th>
</tr>
</thead>
</table>

SEMESTER-V
COURSE 13: UNSUPERVISED ML WITH PYTHON

Practical ___________________ Credits: 1 ___________________ 2 hrs/week

Details of Lab/Practical/Experiments/Tutorials syllabus:

1. Implementation of K-Means Clustering
2. Implement the bisecting k-means clustering algorithm
3. Implement Apriori algorithm
4. Implement Association rule-generation functions
5. Implement FP-tree creation
6. Write a function to find all paths ending with a given item.
7. Implement Code to access the Twitter Python library
8. Implement the PCA algorithm
9. Write a program to find Rating estimation by using the SVD
10. Implement Image-compression functions using SVD.
SEMESTER-V

COURSE 14 A: WEB SCRAPING WITH PYTHON

Theory

| Credits: | 3        | 3 hrs/week |

1. Learning Outcomes

   Students at the successful completion of the course will be able to:

1. Parse complicated HTML pages
2. Develop crawlers with the Scrapy framework
3. Learn methods to store data you scrape
4. Read and extract data from documents
5. Clean and normalize badly formatted data
6. Read and write natural languages
7. Crawl through forms and logins
8. Scrape JavaScript and crawl through APIs
9. Use and write image-to-text software
10. Avoid scraping traps and bot blockers
11. Use scrapers to test your website

Detailed Syllabus: (Five units with each unit having 12 hours of class work) UNIT-1

**Building Scrapers: First Web Scraper:** Connecting, An Introduction to BeautifulSoup,

**Advanced HTML Parsing:** Another Serving of BeautifulSoup, Regular Expressions, Regular Expressions and BeautifulSoup, Accessing Attributes, Lambda Expressions, **Writing Web Crawlers:** Traversing a Single Domain, Crawling an Entire Site, Crawling Across the Internet.

UNIT-II

**Web Crawling Models:** Planning and Defining Objects, Dealing with Different Website Layouts, Structuring Crawlers. **Scrapy:** Installing Scrapy, Initializing a New Spider, Writing a Simple Scraper, Spidering with Rules, Creating Items, Outputting Items, The Item Pipeline, Logging with Scrapy. **Storing Data:** Media Files, Storing Data to CSV, MySQL: Integrating with Python, Database Techniques and Good Practice, Six Degrees in MySQL, Email

UNIT-III

**Advanced Scraping:Reading Documents:** Document Encoding, Text, CSV, PDF, Microsoft Word and .docx. **Reading and Writing Natural Languages:** Summarizing Data, Markov Models, Natural Language Toolkit. **Crawling Through Forms and Logins:** Python Requests
Library, Submitting a Basic Form, Radio Buttons, Checkboxes, and Other Inputs, Submitting Files and Images, Handling Logins and Cookies.

UNIT-IV

**Crawling Through APIs:** A Brief Introduction to APIs, Parsing JSON, Undocumented APIs, Finding Undocumented APIs, Documenting Undocumented APIs, Finding and Documenting APIs Automatically.

**Image Processing and Text Recognition:** Overview of Libraries, Pillow, Processing Well-Formatted Text, Reading CAPTCHAs and Training Tesseract, Retrieving CAPTCHAs and Submitting Solutions

UNIT-V


Text Books:


References:


**CO-CURRICULAR ACTIVITIES:**

- Quiz Competition
- Expert Lectures
- Seminars

**EXTRA CURRICULAR ACTIVITIES:**

- Formal Examination
- Lab Practical
- Presentation
- Simple Projects
SEMESTER-V

COURSE 14 A: WEB SCRAPING WITH PYTHON

Practical Credits: 1 2 hrs/week

1. Write a Python program to test if a given page is found or not on the server.

2. Write a Python program to download and display the content of robot.txt for en.wikipedia.org.

3. Write a Python program to get the number of datasets currently listed on data.gov.

4. Write a Python program to convert an address into geographic coordinates (like latitude and longitude).

5. Write a Python program to display the name of the most recently added dataset on data.gov.

6. Write a Python program to extract h1 tag from example.com.

7. Write a Python program to extract and display all the header tags from en.wikipedia.org/wiki/Main_Page.

8. Write a Python program to extract and display all the image links from a website.

9. Write a Python program to get 90 days of visits broken down by browser for all sites on data.gov.

10. Write a Python program to that retrieves an arbitrary Wikipedia page of "Python" and creates a list of links on that page.

11. Write a Python program to check whether a page contains a title or not.

12. Write a Python program to list all language names and number of related articles in the order they appear in wikipedia.org.

13. Write a Python program to get the number of followers of a given twitter account.

14. Write a Python program to find the live weather report (temperature, wind speed, description and weather) of a given city.

15. Write a Python program to display the date, days, title, city, country of next 25 events.

16. Write a Python program to download IMDB’s Top 250 data (movie name, Initial release, director name and stars).

17. Write a Python program to get movie name, year and a brief summary of the top 10 random movies.
SEMESTER-V

COURSE 14 B: PREDICTIVE AND ADVANCED ANALYTICS USING R

Theory Credits: 3
3 hrs/week

Aim and objectives of Course (Predictive and Advanced Analytics):

The course enables students to:

- To learn, how to develop models to predict categorical and continuous outcomes, using such techniques as neural networks, decision trees, logistic regression, support vector machines and Bayesian network models.
- To know the use of the binary classifier and numeric predictor nodes to automate model selection.
- To advice on when and how to use each model. Also learn how to combine two or more models to improve prediction.

Learning outcomes of Course (In consonance with the Bloom’s Taxonomy): The students will be able to:

- Understand the process of formulating business objectives, data selection/collection, preparation and process to successfully design, build, evaluate and implement predictive models for a various business application.
- Compare the underlying predictive modeling techniques.
- Select appropriate predictive modeling approaches to identify cases to progress with.
- Apply predictive modeling approaches using a suitable package such as SPSSModeler.

Unit-I

Introduction to Data Mining Introduction, what is Data Mining? Concepts of Data mining, Technologies Used, Data Mining Process, KDD Process Model, CRISP – DM, Mining on various kinds of data, Applications of Data Mining, Challenges of Data Mining.

Unit II: Data Understanding and Preparation Introduction, Reading data from various sources, Data visualization, Distributions and summary statistics, Relationships among variables, Extent of Missing Data. Segmentation, Outlier detection.


Unit IV: Automated Data Preparation, Combining data files, Aggregate Data, Duplicate Removal, Sampling DATA, Data Caching, Partitioning data, Missing Values. Model Evaluation and Deployment Introduction, Model Validation, Rule Induction Using CHAID.
Unit V: Automating Models for Categorical and Continuous targets, Comparing and Combining Models, Evaluation Charts for Model Comparison, Deploying Model, Assessing Model Performance, Updating a Model.

Recommended Text Book:

1. Predictive & Advanced Analytics (IBM ICE Publication)

Cocurricular Activities:

- Quiz Competition
- Expert Lectures
- Seminars

Assessment Methods:

- Formal Examination
- Lab Practical
- Presentation
- Simple Projects
SEMESTER-V

COURSE 14 B: PREDICTIVE AND ADVANCED ANALYTICS USING R

Practical Credits: 1 2 hrs/week

Predictive And Advanced Analytics Using R Lab

- Implementation of following methods using R or Matlab
- 1. Simple and multiple linear regression
- 2. Logistic regression
- 3. Linear discriminant analysis
- 4. Ridge regression
- 5. Cross-validation and boot strap
- 6. Fitting classification and regression trees
- 7. K-nearest neighbors
- 8. Principal component analysis
SEMESTER-V

COURSE 15 A: ADVANCED DATA ANALYSIS USING

Course outcomes:

This course will enable the student to:

- Present an overview data science and applications.
- Plan the methods of data collection.
- Describe the statistical methods in EDA.
- Apply statistical methods to develop and evaluate the models.
- Becoming an expert in decision making for complex projects.

UNIT I – INTRODUCTION
Introduction to Data Science – Evolution of Data Science – Data Science Roles – Stages in a Data Science Project – Applications of Data Science in various fields – Data Security Issues.

UNIT II – DATA COLLECTION AND PRE-PROCESSING

UNIT III – EXPLORATORY DATA ANALYTICS
Descriptive Statistics – Mean, Standard Deviation, Skewness and Kurtosis – Box Plots – Pivot Table – Heat Map – Correlation Statistics – ANOVA.

UNIT IV – MODEL DEVELOPMENT

UNIT V – MODEL EVALUATION

Text Books:

3. David Dietrich, Barry Heller, Beibei Yang, “Data Science and Big data Analytics”, EMC 2013
1. Creating a Data Frame and Matrix-like Operations on a Data Frame, Merging two Data Frames
2. Applying functions to Data Frames, import of external data in various file formats, statistical functions, compilation of data.
3. Using Functions with Factors
4. Accessing the Internet
5. Visualization Effects
6. Plotting with Layers
7. Overriding Aesthetics
8. Histograms and Density Charts
9. Simple Linear Regression – Fitting, Evaluation and Visualization
10. Multiple Linear Regression, Lasso and Ridge Regression
11. Use the following scenarios:
12. Use the Diabetes data set from UCI and Pima Indians Diabetes data set for performing the following:
   Univariate Analysis: Frequency, Mean, Median, Mode, Variance, Standard Deviation, Skewness and Kurtosis.
   i. Bivariate Analysis: Linear and logistic regression modeling.
   ii. Multiple Regression Analysis
   iii. Also Compare the results of the above analysis for the two data sets.
   b. Data Modelling
      i. Apply Bayesian and SVM techniques on Iris and Diabetes data set.
      ii. Apply and explore various plotting functions on UCI datasets.
SEMESTER-V
COURSE 15 B: DATA WRANGLING WITH JAVA SCRIPT

Theory Credits: 3 3 hrs/week

UNIT I

Getting started: establishing your data pipeline - Why data wrangling - What’s data wrangling - Why use JavaScript for data wrangling - Is JavaScript appropriate for data analysis? Navigating the JavaScript ecosystem - Establishing your data pipeline

UNIT II


UNIT III

Acquisition, storage, and retrieval - Getting the code and data - The core data representation - Loading data from text files - Loading data from a REST API - Parsing JSON text data - Parsing CSV text data - Importing data from databases - Importing data from MongoDB - Importing data from MySQL - Exporting data - Exporting data to text files - Exporting data to JSON text files - Exporting data to CSV text files - Exporting data to a database - Exporting data to MongoDB - Exporting data to MySQL.

UNIT IV

Exploratory coding - Iteration and your feedback loop - A first pass at understanding your data - Working with a reduced data sample - Prototyping with Excel - Exploratory coding with Node.js - Using Nodemon - Exploring your data - Using Data-Forge - Computing the trend column - Outputting a new CSV file - Exploratory coding in the browser.

UNIT 5

Clean and prepare - The need for data cleanup and preparation - Where does broken data come from? - How does data cleanup fit into the pipeline - Identifying bad data - Techniques for fixing bad data - Cleaning our data set - Preparing our data for effective use.

TEXTBOOK

Data Wrangling with JavaScript - Ashley Davis - Manning Publication
**REFERENCE TEXTBOOKS:**
Principles of Data Wrangling - Practical Techniques for Data Preparation
- Tye Rattenbury, Joseph M. Hellerstein, Jeffrey Heer, Sean Kandel, Connor Carreras - Oreilly Publication.

Data Wrangling with Python - Jacqueline Kazil, Katharine Jarmul 2016-Oreilly Publication.
1. Installing steps for Node.js, Installing npm dependencies, printing version of Node.js. Developing a Simple program to print “Hello World” on console.
2. Create a Code library to a command-line application in Node.js to produce a report from data.
3. Install “Express Node.js framework”. Create a simple Web application to print your college name on Web Page.
4. Develop an Web application to add static files to web server.
5. Develop an application to import & export data using MongoDB.
6. Develop an application to import & export data using MYSQL.
7. Loading your input CSV file and printing its contents to the console. Apply slicing operation on data and print on Console. Print datatype of each column (attribute) of CSV file.
8. List column names and data types of column names of a CSV file using Data-Forge (An open source data-wrangling toolkit for JavaScript).
10. Develop an application to do Excel data analysis in Node.js using Formulajs.
11. Develop a program to clean your data by rewriting rows to fix bad data.
SEMESTER-VII
COURSE 16 A: BIG DATA ANALYTICS USING SPARK & HADOOP

Theory Credits: 3 3 hrs/week

Aim and objectives of Course:

- To Understand the Complete Architecture of Spark
- To know the differences between Hadoop and Spark
- To know the concepts of Spark Programming

Learning outcomes of Course:

- Students will get well knowledge of what is Big Data
- Knowledge in Spark Eco System
- Mapping of Data Analytics techniques in Spark
- Application of Spark Programming to Analytics problems

UNIT I:
Introduction to Big Data: What is Big Data-Characteristics, Data in the Warehouse and Data in Hadoop, Why is Big Data Important- When to consider Big Data Solution, Applications. Introduction to Hadoop: Hadoop- definition, Application development in Hadoop. The building blocks of Hadoop, Name Node, Data Node, Secondary Name Node, Job Tracker and Task Tracker.

UNIT II:
Introduction to Spark: What is Apache Spark, Why Spark when Hadoop is there, Spark Features, Spark components, Spark program flow, Spark Eco System. Differences between implementation of programs in Hadoop and Spark Programming environments.

UNIT III:
Spark Fundamentals- Using spark in action VM, Using Spark Shell and writing first spark program, Basic RDD actions and transformations. Spark SQL-Working with Data Frames, Using SQL Commands, Saving and loading Data Frame.

UNIT IV:

UNIT V:
Graph Representation in MapReduce: Graph Processing with Spark, Spark GraphX, GraphX features, Graph Examples, Graph algorithms-Shortest Path Algorithm.
TEXT BOOKS:

REFERENCE BOOKS:

Student Activities:
1. Use basic Operations with PySpark (Spark with Python)
2. Data Pre-processing
3. Build a data processing pipeline
4. Build the classifier
5. Train and evaluate the model
6. Tune the hyper parameter

Continuous assessment:
1. Let the students be tested in the following questions from each unit
2. What is Big Data? Explain the characteristics of it
3. What is Spark? What are the advantages of it over Hadoop
4. Explain Spark SQL
5. Explain Spark Streaming
SEMESTER-VII
COURSE 16 A: BIG DATA ANALYTICS USING SPARK & HADOOP

Practical Credits: 1 2 hrs/week

1. Using Python Implement the following Programs
   a) Write Program to implement arithmetic operations
   b) Write Program to find the biggest of two numbers
   c) Write a program to find the matrix multiplication
2. Install Hadoop
3. Install Spark on top of Hadoop
4. Create and Implement the transformations in RDDs
5. Create a data frame from an existing RDD using Spark Session
6. Execute a Word Count example in Spark Shell by creating RDDs.
7. Implement Spark SQL Queries in Python.
8. Write a Program to implement maximum temperature give the recordings of one year.
9. Write a Program to implement the Pie estimation
10. Write a User Defined Function to convert a given text to Uppercase.
SEMESTER-VII

COURSE 16 B: BIG DATA SECURITY

Theory Credits: 3 3 hrs/week

Course Objective

With the data generated from electronic devices growing exponentially, the need to analyse data on a large scale is important. Such data are of many types like financial, personal etc. Big data environment also created significant security challenges. When trying to make quick decisions. Data breach poses many complications. This course aims at introducing concepts related to big data security.

Learning Outcomes:

1. Understanding significance of privacy, ethics in big data environment.
2. Analyzing the steps to secure big data.
3. Analyzing data security and event logging.

Unit-I


Unit-II


Unit-III


Unit-IV

HADOOP ECOSYSTEM SECURITY: Configuring Kerberos for Hadoop ecosystem components – Pig, Hive, Oozie, Flume, HBase, Scoop.

Unit-V

SEMESTER-VII

COURSE 16 B: BIG DATA SECURITY

Practical Credits: 1 2 hrs/week

1. Learn the basics of Mongo DB.
2. Installation steps for Mongo DB.
3. Use the following commands
   (a) DATABASE_NAME.
   (b) Drop Database( )
   (c) create Collection
   (d) insert( )
   (e) drop( )
   (f) find( )
4. Differentiate between SQL and Mongo DB.
5. Write a program to update a collection in Mongo DB
6. Write a program to remove specific document from Mongo DB.
7. Write a program to implement aggregate function in Mongo DB
   8. Apply the Map reduce operation to find total salary of each department assuming employee collection is already exists.
COURSE 17 A: INTRODUCTION TO DEEP LEARNING

Theory          Credits: 3          3 hrs/week

Course Objectives:
1. To introduce basics of linear algebra and probability theory
2. To introduce the fundamental techniques and principles of Neural Networks
3. To familiarize different models in Artificial Neural Networks (ANN) and their applications
4. To familiarize deep learning concepts with Convolutional Neural Network case studies
5. To explain functioning of deep neural networks

Course Outcomes:
After learning the course, the students will be able to:
1. Discuss feed forward networks and their training issues
2. Distinguish different types of ANN architectures
3. Design Feed Forward Neural Network architecture for research problems
4. Apply mathematical concepts such as linear algebra, calculus to solve the research problems.
5. Apply deep learning techniques to practical problems
6. Evaluate model performance and interpret results

Unit 1

Unit 2
Fundamentals of Neural Networks: Introduction to Neural Network, Model of Artificial Neuron, Learning rules and various activation functions.

Unit 3

Unit 4

Unit 5
Text Books:

Reference Books:
SEMESTER-VII

COURSE 17 A: INTRODUCTION TO DEEP LEARNING

Practical Credits: 1 2 hrs/week

1. Design and implement a CNN for Image Classification:
   a) Select a suitable image classification dataset (medical imaging, agricultural, etc.).
   b) Optimized with different hyper-parameters including learning rate, filter size, no. of layers, optimizers, dropouts, etc.

2. Apply a pre-trained network and apply it to a new task using transfer learning
   a) Use any three pre-trained models including AlexNet, GoogleNet, VGGNet, MobileNet, ResNet, DenseNet, etc.
   b) Fine-tune the hyper-parameters and compare their performance for a suitable application.

3. Design RNN or its variant including LSTM or GRU
   a) Select a suitable time series dataset. Example – predict sentiments based on product reviews
   b) Apply for prediction

4. build a word2vec model for unstructured data
   a) Use any unstructured text dataset
   b) Convert words into a representative vector of numerical values

5. Implement an artificial neural network on GPUs
   a) Implement ANN on GPUs.
Aim and objectives of Course (Predictive and Advanced Analytics):
The course enables students to:
- To learn, PyTorch and relates PyTorch to deep learning.
- To know about pretrained models to work with data sets
- To learn how to work with neural networks in deep learning.

Learning outcomes of Course (In consonance with the Bloom’s Taxonomy):
The students will be able to:
- Learn about deep learning, PyTorch library and their applications.
- How to use Tensors in deep learning Tensor-API
- To work with time series in deep learning.
- To develop a PyTorch Neural Network model using “torch.nn” module.

UNIT I

UNIT II
Pretrained Networks - A pretorined that reconizes the subject of an image - A pertained network that describes scenes - Starting with Tensor- Tensors.

UNIT III
Multidimensional arrays - Indexing tensors - Named Tensors - Tensor element types - The tensor API - Tensors: Scenic views of storage.

UNIT IV
Real workd data representation using tensors - Working with images - Representing tabular data - Working with time series - Representing text - The mechanics of Learning - Learning is just parame-ter estimation - Less loss is what we want.

UNIT V
Using a nearly network to fit the data - Artificial neutrons - The PyTorch nn module - Finally a neural network.

TEXTBOOK:
Deep Learning with PyTorch - Eli Stevens, Luca Antiga, Thomas Viehmann - Manning Publica-tions

REFERENCES:
SEMESTER-VII

COURSE 17 B: DEEP LEARNING WITH PYTORCH

Practical Credits: 1 2 hrs/week

2. Program On CycleGAN network: Use CycleGAN network model to turn one image into another image. Feed the image of the golden retriever into the horse-to-zebra model. (Refer Prescribed Textbook)
3. Develop a program to convert integers into floating point numbers in pyTorch using tensors.
4. Develop a program on multi-dimensional arrays using TensorFlow.
5. Develop a PyTorch program to perform different Tensor Indexing operations.
6. Develop a PyTorch program on “named tensors”.
7. Develop a program in TensorFlow for representing text. Define a function clean_words, which takes text as input and returns it in lowercase and print individual words in given text.
8. Perform timeseries operations on bikes data. Use following url for reference (code/p1ch4/4_time_series_bikes.ipynb)
9. Working with images: Take several pictures of red, blue, and green items with your phone or other digital camera. Load each image, and convert it to a tensor. For each imagetensor, use the .mean() method to get a sense of how bright the image is.
SEMESTER-VII

COURSE 18 A: AI CONCEPTS AND TECHNIQUES WITH PYTHON

Theory                                                        Credits: 3

Aim and objectives of Course:
• This course provides an introduction to the fundamentals of artificial intelligence.
• Demonstrates fundamental understanding of the history of artificial intelligence (AI) and its foundations.
• Apply basic principles of AI in solutions that require problem solving, inference, perception, knowledge representation, and learning.
• Demonstrates awareness and a fundamental understanding of various applications of AI techniques in intelligent Agents.

Learning outcomes of Course:
• List the objectives and functions of modern Artificial Intelligence.
• Categorize an AI problem based on its characteristics and its constraints.
• Understand and implement search algorithms.
• Learn how to analyze the complexity of a given problem and come with suitable optimizations.
• Demonstrate practical experience by implementing and experimenting with the learnt Algorithms

UNIT I:
Problems and Search: What is Artificial Intelligence, The AI Problems, and Underlying Assumption, what is an AI Technique. Problems, Problems Spaces, and Search: Defining the problem as a state space search, production systems, problems characteristics, issues in the design of search programs.

UNIT II:
Heuristic Search Techniques: Generate-and-test, Hill Climbing, Best-First Search, Problem Reduction, Constraint Satisfaction, Means-Ends Analysis

UNIT III:

UNIT IV:
Representing Knowledge using Rules: Procedural Vs Declarative knowledge, Logic Programming, Forward Vs Backward Reasoning, Matching, Control Knowledge

UNIT V:
TEXT BOOK:

REFERENCES BOOK:
SEMESTER-VII

COURSE 18 A: AI CONCEPTS AND TECHNIQUES WITH PYTHON

Practical Credits: 1 2 hrs/week

1. Write a Program to Implement Breadth First Search using Python.
2. Write a Program to Implement Depth First Search using Python.
3. Write a Program to Implement Tic-Tac-Toe game using Python.
4. Write a Program to implement 8-Puzzle problem using Python.
5. Write a Program to Implement Water-Jug problem using Python.
6. Write a Program to Implement Travelling Salesman problem using Python.
8. Write a Program to implement 8-Queens problem using Python.
SEMESTER-VII

COURSE 18 B: DATA & INFORMATION SECURITY

Theory Credits: 3 3 hrs/week

Overview of Security: Protection versus security; aspects of security – data integrity, data availability, privacy; security problems, user authentication, Orange Book.

Unit -II

Security Threats: Program threats, worms, viruses, Trojan horse, trap door, stack and buffer overflow; system threats- intrudes; communication threats- tapping and piracy.

Unit -III

Cryptography: Substitution, transposition ciphers, symmetric-key algorithms – Data Encryption Standard, advanced encryption standards, public key encryption – RSA; DiffieHellman key exchange, ECC cryptography, Message Authentication – MAC, has functions.

Unit -IV

Digital Signatures: Symmetric key signatures, public key signatures, message digests, public key infrastructures.

Unit -V

Security Mechanism: Intrusion detection, auditing and logging, tripwire, system –call monitoring.

Text Books:

Reference Books:
2. J. Piwprzyk, T. Hardjono and J. Seberry, Fundamentals of Computer Security,
Student Activity

Case Study I: Transform Data from one format to another format using Cryptography.
Case Study II: How mails are hacked.
1. Implement Ceiser Cipher encryption in Python.
2. Implement Ceiser Cipher decryption in Python.
3. Implement Transposition technique encryption in Python.
4. Implement Substitution cipher encryption in Python.
5. Implement Substitution cipher decryption in Python.
6. Implement One time Pad cipher in Python.
7. Implement DES encryption in Python.
8. Implement RSA Public Key encryption in Python
SEMESTER-VII

COURSE 19: INTRODUCTION TO NEURAL NETWORKS

Theory Credits: 3 3 hrs/week

Learning Objectives:

1. The main objective of Neural Network Techniques to Improve Data Analysis Solutions is to strengthen the dialogue between the statistics and soft computing research communities in order to cross-pollinate both fields and generate mutual improvement activities.

2. Also introduce the neural networks for classification, regression and to give design methodologies for artificial neural networks.

Learning Outcome:

On successful completion of this course, student will be able to

1. Obtain the fundamentals and types of neural networks. The student will have a broad knowledge in developing the different algorithms for neural networks.

2. Analyze neural controllers

3. Have a broad knowledge in Fuzzy logic principles and will be able to determine different methods of Defuzzification

UNIT 1


Unit 2


Unit 3


Unit 4

Integration of Neural Network, Fuzzy logic and Genetic Algorithm: Hybrid system. Neural Networks, Fuzzy logic, and Genetic Algorithm Hybrids.
Unit 5

Associative Memory: Autocorrelators, Heterocorrelators, Wang et al’s Multiple Training Encoding Strategy, Exponential BAM, Associative Memory for Real coded pattern pairs, Applications.

Text book:
1. Neural Networks a Comprehensive Foundations, Simon S Haykin, PHI Ed.,
3. Neural Networks in Computer Inteligance, Li Min Fu TMH 2003 Neuro-Fuzzy Systems, Chin Teng Lin, C. S. George Lee, PHI.
4. Build_Neural_Network_With_MS_Excel_sample by Joe choong.
SEMESTER-VII

COURSE 19: INTRODUCTION TO NEURAL NETWORKS

Practical Credits: 1 2 hrs/week

1. Create a perceptron with appropriate no. of inputs and outputs. Train it using fixed increment learning algorithm until no change in weights is required. Output the final weights.

2. Create a simple ADALINE network with appropriate no. of input and output nodes. Train it using delta learning rule until no change in weights is required. Output the final weights.

3. Train the autocorrelator by given patterns: $A_1=(-1,1,-1,1)$, $A_2=(1,1,1,-1)$, $A_3=(-1,-1,-1,1)$. Test it using patterns: $Ax=(-1,1,-1,1)$, $Ay=(1,1,1,1)$, $Az=(-1,-1,-1,1)$.

4. To Implement Convolution Neural network for Text classification or Image Classification

5. Implementation of Naïve Bayes/SVM/SGD/SVM classifier on text and image

6. To study Word Embedding techniques : Word2vec,doc2vec,Glove

7. Implement Linear/Logistic regression
SEMESTER-VII
COURSE 20: NATURAL LANGUAGE PROCESSING

Theory Credits: 3 3 hrs/week

COURSE OBJECTIVE:

The basic objectives of natural language processing course are the following:

1. Learn the basics of natural language processing and understand various steps in it.
2. To introduce the fundamentals of language processing from the algorithmic viewpoint.
3. To discuss various issues that make natural language processing a hard task.
4. To discuss some well-known applications of natural language processing

LEARNING OUTCOME:

At the end of the course, the student should be able to:

1. Appreciate the fundamental concepts of natural language processing.
2. Design algorithms for natural language processing tasks.
3. Develop useful systems for language processing and related tasks involving text processing.

UNIT I:


UNIT II:


UNIT III:

UNIT IV:

UNIT V:

Text Books:
1. Natural Language Processing with Python. Steven Bird, Ewan Klein, and Edward Lope, O’Reily, 2009

Suggested Reading:
References Books:
1. Text segmentation: Segment a text into linguistically meaningful units, such as paragraphs, sentences, or words. Write programs to segment text (in different formats) into tokens (words and word-like units) using regular expressions. Compare an automatic tokenization with a gold standard.

2. Part-of-speech tagging: Label words (tokens) with parts of speech such as noun, adjective, and verb using a variety of tagging methods, e.g., default tagger, regular expression tagger, unigram tagger, and n-gram taggers.

3. Text classification: Categorize text documents into predefined classes using Naïve BayesClassifier and the Perceptron model.

4. Chunk extraction, or partial parsing: Extract short phrases from a part-of-speech tagged sentence. This is different from full parsing in that we're interested in standalone chunks, or phrases, instead of full parse trees.

5. Parsing: parsing specific kinds of data, focusing primarily on dates, times, and HTML. Make use of the following preprocessing libraries:
   - dateutil which provides datetime parsing and timezone conversion
   - lxml and BeautifulSoup which can parse, clean, and convert HTML
   - charade and UnicodeDammit which can detect and convert text character encoding

Course Objective: The purpose of a research methodology is to explain the reasoning behind your approach to your research - you'll need to support your collection methods, methods of analysis, and other key points of your work. Think of it like writing a plan or an outline for what you intend to do.


UNIT II – Problem Identification & Formulation – Research Question – Investigation Question


UNIT IV – Qualitative and Quantitative Research: Qualitative research – Quantitative research
– Concept of measurement, causality, generalization, replication. Merging the two approaches. Measurement: Concept of measurement– what is measured? Problems in measurement in research – Validity and Reliability. Levels of measurement – Nominal, Ordinal, Interval, Ratio.


Text Books Recommended:-
Business Research Methods – Donald Cooper & Pamela Schindler, TMGH, 9th edition
Business Research Methods – Alan Bryman & Emma Bell, Oxford University Press.
Research Methodology – C.R.Kothari.
Research Exploration: Case Studies

The purpose of a research methodology is to explain the reasoning behind your approach to your research - you'll need to support your collection methods, methods of analysis, and other key points of your work. Think of it like writing a plan or an outline for what you intend to do.

1. Apply the research concepts and tools to identify a research problem, sub-statements of the problem to clarify the direction of research effort?
2. Construct a research framework showing the variables of the study relevant for a specific area of inquiry
3. Design a research methodology by classifying the research design, population and sampling techniques suitable for a research problem.
4. Develop a research proposal by using principles and frameworks of research methodology to justify an ethical, practical and effective research project.
SEMESTER-VIII
COURSE 21 B: COMPUTATIONAL DATA SCIENCE

Theory                                      Credits: 3                        3 hrs/week

Aim and objectives of Course (Predictive and Advanced Analytics):
The course enables students to:
1. To learn descriptive statistics using computational tools
2. To learn distributions in probability.
3. To work with different types of dispersion measures.

Learning outcomes of Course (In consonance with the Bloom’s Taxonomy): The students will be able to:

- To understand different computational tools and statistical techniques.
- To learn about curve fitting in data science.
- Able to perform visual diagnostics.

UNIT I

UNIT II

UNIT III
Comparing Two Samples - A/B Testing - Deflategate - Causality - Estimation – Importance of Mean - Properties of the Mean - Variability - The SD and the Normal Curve.

UNIT IV
The Central Limit Theorem - The Variability of the Sample Mean - Choosing a Sample Size - Prediction - Correlation - The Regression Line - The Method of Least Squares - Least Squares Regression

UNIT V
Visual Diagnostics - Numerical Diagnostics. Inference for Regression - A Regression Model - Inference for the True Slope - Prediction Intervals

TEXTBOOK:
2. Web link: https://inferentialthinking.com/chapters/intro.html
REFERENCES:
1. Project Gutenberg is a website that publishes public domain books online. Using Python, load the text of two classic novels directly from the web.

2. Create a Table, add new columns, add data, move data to a table using cvs file, access the data in a column and choose sets of columns.


4. Write a Program that simulates Multiple values

5. Based on the top_movies_2017.csv data set, draw some samples, create a deterministic sample, find probability sample and find systematic sample (https://inferentialthinking.com/chapters/10/Sampling_and_Empirical_Distributions.html)

6. Simulate one value, multiple value of the statistic and visualise the results in an empirical history ram circuits (https://inferentialthinking.com/chapters/11/1/Assessing_a_Model.html)

7. Under A/B Testing, observed difference, Predict the statistic under null hypothesis, and perform permutations test.

8. Using Central Limit theorem, find out the average flight delay from the database ‘unit-ed_summer2015.csv’.
SEMESTER-VIII

COURSE 22 A: COMPUTER VISION WITH PYTHON

Theory  Credits: 3  3 hrs/week

Course Outcomes:

□ This course enables the learners to understand the advanced concepts in computer vision.
□ The course covers the basics of image processing, imaging geometry, image segmentation, feature extraction, object recognition and classification and common applications of computer vision.
□ This course helps the students to design solutions for complex real-life problems.

Unit – 1 (Image Formation and Processing) Image formation and Image model- Components of a vision system- Cameras- camera model and camera calibration-Radiometry- Light in space- Light insurface - Sources, shadows and shading. Fundamentals of Image processing: Basic steps of Image processing system sampling and quantization of an Image – Basic relationship between pixels.


Unit - 3 (Image Segmentation) Classification of segmentation techniques, Edge detection, Edge linking, Thresholding, Region growing, Region splitting and merging, Watershed based segmentation. Shadow detection and removal. Image processing using OpenCV - blending, smoothing, and reshaping.

Unit - 4 (Image Recognition and Classification) Shape based object classification, Motion based object classification, Viola Jones Object Detection Framework, Object classification using CNNs, use of RCNN for object classification.

Unit - 5 (Applications) Speech and Handwriting Recognition, Automatic Face Recognition, Video Segmentation and Keyframe Extraction, Real-Time Hand Pose Recognition.

Text Books:


Reference Books
1. Reducing the Number of Intensity Levels in an Image.
2. Zooming and Shrinking Images by Pixel Replication.
3. Zooming and Shrinking Images by Bilinear Interpolation.
7. Spatial Filtering.
9. Unsharp Masking
SEMESTER-VIII
COURSE 22 B: DATA WRANGLING WITH JAVA SCRIPT

Course Outcomes:
Aim and objectives of Course (Predictive and Advanced Analytics): The course enables students to:
- To learn how to use REST APIs.
- To learn different types of data formats.
- To learn connectivity of different databases like MYSQL, MongoDB.

Learning outcomes of Course (In consonance with the Bloom’s Taxonomy): The students will be able to:

- To learn working with NodeJS
- To learn how to host a server and run a server in local host.
- Importing and exporting data from different data formats.
- Develop a web visualization application

UNIT I
Getting started: establishing your data pipeline - Why data wrangling - What’s data wrangling - Why use JavaScript for data wrangling- Is JavaScript appropriate for data analysis?
Navigating the JavaScript ecosystem - Establishing your data pipeline

UNIT II

UNIT III
Acquisition, storage, and retrieval - Getting the code and data - The core data representation - Loading data from text files - Loading data from a REST API - Parsing JSON text data - Parsing CSV text data - Importing data from databases - Importing data from MongoDB - Importing data from MySQL - Exporting data - Exporting data to text files - Exporting data to JSON text files - Exporting data to CSV text files - Exporting data to a database - Exporting data to MongoDB - Exporting data to MySQL.

UNIT IV
Exploratory coding - Iteration and your feedback loop - A first pass at understanding your data - Working with a reduced data sample - Prototyping with Excel - Exploratory coding with Node.js
Using NodeMon

- Exploring your data
- Using Data-Forge
- Computing the trend column
- Output-ting a new CSV file

UNIT V
Exploratory coding in the browser - Clean and prepare - The need for data cleanup and preparation
- Where does broken data come from? - How does data cleanup fit into the pipeline-
Identifying bad data - Techniques for fixing bad data - Cleaning our data set - Preparing our
data for effective use.

TEXTBOOK:
1. Data Wrangling with JavaScript - Ashley Davis - Manning Publication

REFERENCE TEXTBOOKS:
2. Data Wrangling with Python - Jacqueline Kazil, Katharine Jarmul 2016 - Oreilly Publication
SEMESTER-VIII
COURSE 22 B: DATA WRANGLING WITH JAVA SCRIPT

Practical Credits: 1 2 hrs/week

1. Install Node.js and install dependencies
2. Run one web application and get the data
3. Install “Express Node.js framework” . Create a simple Web application to print your college name on Web Page.
4. Create a simple web server and add static files to it
5. Add a REST API to web server to dynamically generate a report.
6. Loading your input CSV file and printing its contents to the console. Apply slicing operation on data and print on Console. Print datatype of each column(attribute) of CSV file.
7. Retrieve data from REST API
8. Import Earthquakes data from REST API
9. Import data from MySql
10. Use HTTP GET to retrieve the data from your CSV file.
11. Develop a program to clean your data by rewriting rows to fix bad data.
COURSE 23A: SOCIAL MEDIA ANALYTICS

OBJECTIVES

To Understand the Complete Architecture of Social Media Analytics

- To know Web analytics tools
- To know the social media of ongoing campaigns

OUTCOMES

- Students will get well knowledge of what is Social Media Analytics
- Knowledge in Web analytics tools
- Mapping of Network Analysis. (LinkedIn, Instagram, YouTube Twitter etc.)

UNIT-I

Introduction to Social Media Analytics (SMA): Social media landscape, Need for SMA; SMA in Small organizations; SMA in large organizations; Application of SMA in different areas

Network fundamentals and models: The social networks perspective - nodes, ties and influencers, Social network and web data and methods.

UNIT-II

Graphs and Matrices- Basic measures for individuals and networks. Information visualization


UNIT-III

Web crawling and Indexing; Natural Language Processing Techniques for Micro-text


UNIT-IV

Processing and Visualizing Data, Influence Maximization, Link Prediction, Collective Classification. Applications in Advertising and Game Analytics (Use of tools like Unity30 / PyCharm). Introduction to Python Programming, Collecting and analyzing social media data; visualization and exploration

UNIT-V

Social campaigns. Measuring and Analyzing social campaigns, defining goals and evaluating outcomes, Network Analysis. (LinkedIn, Instagram, YouTube Twitter etc.) Use of Google Analytics. Post Performance on FB, Use of Facebook Business Manager
Text Books:
1. Mathew Ganis, Avinash Koihrkar Social Media Analytics IBM Press 2015 / 1st
2. Jim Sterne Social Media Metrics Wiley Latest
3. Oliver Blanchard Social Media ROI QuePublishing Latest

Reference books:
2. Tracy L. Tuten, Michael R. Solomon Social Media Marketing Sage 2018 / 3rd
3. Gohar F. Khan Creating Value With Social Media Analytics CreateSpace Independent Publishing 2018 1st
4. Alex Gonsalves Social Media Analytics Strategy Appress 2017 / 1st
SEMMESTER-VIII

COURSE 23 A: SOCIAL MEDIA ANALYTICS

Practical Credits: 1 2 hrs/week

Using Python Implement the following Programs

1. Managing text data
2. Syntactical analysis

3. Vector semantics and latent semantic analysis (LSA)
4. Clustering and topic modeling:
   PLSA, LDA, SLDA, ...

5. Text classification: naive Bayesian, maximum entropy, SVM, ...

6. Clustering Physician Reviews

7. Discovering Topics on Twitter
8. Deep Learning with Text

9. Students should analyses the social media of any ongoing campaigns and present the findings.
SEMESTER-VIII

COURSE 23 B: MAJOR PYSPARK ESSENTIALS FOR DATA

Course Outcomes:

- To learn PYSPARK, and to develop models
- To know the use of PYSPARK to automate model selection.
- To advice on when and how to use each model. Also learn how to combine two or more Models.

UNIT-I
Data Engineering
Distributed Computing Primer: Technical requirements, Distributed Computing, Distributed Computing with Apache Spark, Big data processing with Spark SQL and Data Frames.
Data Ingestion: Technical requirements, Introduction to Enterprise Decision Support Systems, ingesting data from data sources, ingesting data into data sinks, Using file formats for data storage in data lakes, Building data ingestion pipelines in batch and real time, Unifying batch and real time using Lambda Architecture.

UNIT-II
Data Cleansing and Integration: Technical requirements, transforming raw data into enriched meaningful data, building analytical data stores using cloud data lakes, consolidating data using data integration, Making raw data analytics-ready using data cleansing.
Real-Time Data Analytics: Technical requirements, Real-time analytics systems architecture, Stream processing engines, Real-time analytics industry use cases, Simplifying the Lambda Architecture using Delta Lake, Change Data Capture, Handling late-arriving data, Multi-hop pipelines.

UNIT-III
Data Science
Scalable Machine Learning with PySpark: Technical requirements, ML overview, Scaling out machine learning, Data wrangling with Apache Spark and MLlib.
Feature Engineering – Extraction, Transformation, and Selection: Technical requirements, The machine learning process, Feature extraction, Feature transformation, Feature selection, Feature store as a central feature repository, Delta Lake as an offline feature store

UNIT-IV
Data Analysis
Data Visualization with PySpark: Technical requirements, Importance of data visualization, Techniques for visualizing data using PySpark, Considerations for PySpark to pandas conversion.

UNIT-V
Integrating External Tools with Spark SQL: Technical requirements, Apache Spark as a distributed SQL engine, Spark connectivity to SQL analysis tools, Spark connectivity to BI tools, Connecting Python applications to Spark SQL using Pyodbc The Data Lakehouse: Moving from BI to AI, The data lakehouse paradigm, Advantages of data lakehouses.
Text Books:
2. Applied Data Science Using Pyspark, by Ramcharan Kakarla, Sundar Krishnan, Sridhar Alla, Apress, Springer India

Reference Books:
1. Machine Learning with Pyspark, by Pramod Singh, Apress India
1. Installing PySpark
   2. Demonstrate on Big Data concepts in Python Lambda Functions
      Filter(), map(), and reduce() Sets.
3. Hello World in PySpark
4. PySpark API and Data Structures
5. Running PySpark Programs in
   • Jupyter Notebook
   • Command-Line Interface
   • Cluster
   • PySpark Shell
6. Combining PySpark with other Tools
7. Demonstrate on Data Cleaning with PySpark
8. Demonstrate on Data wrangling with Apache Spark and MLlib
9. Demonstrate Data Visualization with PySpark
10. Demonstrate Connecting Python applications to Spark SQL using Pyodbc
OBJECTIVES:

The student should be made to be familiar with the Business intelligence architectures most fundamental Graphs and results.

Be exposed to the techniques of proofs and analysis.

Detailed Syllabus: (Five units with each unit having 12 hours of class work)

UNIT-I
BUSINESS INTELLIGENCE: Effective and timely decisions - Data, information, and knowledge - Role of mathematical models - Business intelligence architectures: Cycle of a business intelligence analysis - Enabling factors in business intelligence projects

UNIT-II

UNIT- III
EFFICIENCY: Efficiency measures -The CCR model: Definition of target objectives- peer groups - Identification of good operating practices; cross-efficiency analysis - virtual inputs and outputs - Other models. Pattern matching -cluster analysis, outlier analysis

UNIT-IV

UNIT- V

TEXT BOOK:
REFERENCES:


SEMESTER-VIII
COURSE 24: BUSINESS INTELLIGENCE AND VISUALIZATION

<table>
<thead>
<tr>
<th>Practical</th>
<th>Credits</th>
<th>2 hrs/week</th>
</tr>
</thead>
<tbody>
<tr>
<td>Practical 1: Import the legacy data from different sources such as (Excel, SqlServer, Oracle etc.) and load in the target system.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Practical 2: Perform the Extraction Transformation and Loading (ETL) process to construct the database in the Sqlserver / Power BI.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Practical 3: Data Visualization from ETL Process Power BI Desktop</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Practical 4: Creating a Cube in SQL server 2012</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Practical 5: Apply the what – if Analysis for data visualization. Design and generate necessary reports based on the data warehouse data.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Practical 6: Implementation of Classification algorithm in R Programming.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Practical 7: Practical Implementation of Decision Tree using R Tool</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Practical 8: k-means clustering using R</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Practical 9: Prediction Using Linear Regression</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Practical 10: Data Analysis using Time Series Analysis</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Practical 11: Data Modelling and Analytics with Pivot Table in Excel</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Practical 12: Data Analysis and Visualization using Advanced Excel</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Theory</td>
<td>Credits: 3</td>
<td>3 hrs/week</td>
</tr>
<tr>
<td>--------</td>
<td>------------</td>
<td>-------------</td>
</tr>
</tbody>
</table>

**SEMESTER-VIII**

**COURSE 25: DATA VISUALIZATION USING JAVASCRIPT**