Courses for minor degree on ‘Data Science’

<table>
<thead>
<tr>
<th>SL. NO</th>
<th>COURSE CODE</th>
<th>COURSE TITLE</th>
<th>CONTACT PERIODS</th>
<th>L</th>
<th>T</th>
<th>P</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>CSM507</td>
<td>Foundations of Data Science with Python</td>
<td>3</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>2.</td>
<td>CSM508</td>
<td>Machine Learning for Data Science</td>
<td>3</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>3.</td>
<td>CSM509</td>
<td>Data Visualization</td>
<td>3</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>4.</td>
<td>CSM510</td>
<td>Data Security &amp; Privacy</td>
<td>3</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>5.</td>
<td>CSM511</td>
<td>Big data analytics</td>
<td>3</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>6.</td>
<td>CSM512</td>
<td>Exploratory Analysis</td>
<td>3</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>3</td>
</tr>
</tbody>
</table>

Total Credits 18
# Course Objectives:
- To understand fundamentals and the process of data science.
- To comprehend different types and representation of data and analyze them.
- To apply inferential techniques to extrapolate information from the available data.
- To utilize the Python libraries for Data Wrangling.
- To interpret data and present it using visualization libraries in Python.

## Unit I: INTRODUCTION

## Unit II: DESCRIBING DATA
Types of Data – Types of Variables – Basic Statistical descriptions of Data – Describing Data with Tables and Graphs – Describing Data with Averages – Describing Variability – Normal Distributions and Standard (z) Scores

## Unit III: PROBABILITY & STATISTICS

## Unit IV: PYTHON LIBRARIES FOR DATA WRANGLING

## Unit V: DATA VISUALIZATION

### Total: 45 Periods

### References
**Course Outcomes:**

**Upon completion of the course, the students will be able to**

- Understand data science fundamental and follow the correct process for applying data science.
- Represent and understand data in different formats and analyse it.
- Infer new information from the data using different analysis techniques.
- Gather, collect, and transform raw data into useful formats with Python libraries.
- Apply Python libraries to visualize and study data.
Course Objectives:
- To understand the basic concepts of machine learning.
- To understand and build supervised learning models.
- To understand neural network and learn combination of classifiers.
- To understand and build unsupervised learning models.
- To design and analysis of probabilistic graphical models.

<table>
<thead>
<tr>
<th>Unit I</th>
<th>INTRODUCTION TO MACHINE LEARNING</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Unit II</th>
<th>SUPERVISED LEARNING</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Linear Regression Models: Multiple regression – Logistic regression, Naïve Bayes classifier, Nearest Neighbour and KNN Algorithm, Decision Trees, Support Vector Machines, Kernel functions</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Unit III</th>
<th>NEURAL NETWORKS, ENSEMBLE TECHNIQUES</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Artificial Neural Network(ANN), perceptron, multilayer perceptron, Back propogation network(BPN) activation functions, gradient descent optimization, error back propagation, Unit saturation (vanishing gradient problem) - ReLU, hyperparameter tuning, batch normalization, regularization, Ensemble Methods – Bagging, Boosting</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Unit IV</th>
<th>UNSUPERVISORY &amp; REINFORCEMENT LEARNING</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Clustering – Distance Function, Minimum, maximum &amp; average connection, Hierarchical Clustering, agglomerative – K Means clustering, Self-organizing Map, Reinforcement Learning overview</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Unit V</th>
<th>GRAPHICAL MODELS &amp; DIMENSION REDUCTION</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Directed Graphical Models, Bayesian Networks, Markov Models, Hidden Markov Models, Inference- Learning Generalization, Dimension reduction-Curse of Dimensinality, PCA</td>
</tr>
</tbody>
</table>

**Total: 45 Periods**

COURSE OUTCOMES:

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

- **CO1:** Explain the basic concepts of machine learning.
- **CO2:** Construct supervised learning models.
- **CO3:** Construct unsupervised learning algorithms.
- **CO4:** Evaluate and compare different models.
- **CO5:** Design of experiments using machine learning.
<table>
<thead>
<tr>
<th>References:</th>
</tr>
</thead>
</table>
Course Objectives:
- To understand the fundamentals of data visualization.
- To know the working principles of various information visualization depth tools.
- To acquire knowledge about the issues in data representation.
- To visualize the Data using tools Tableau
- To gain skill in designing real time interactive information visualization system.

Unit I | INTRODUCTION

Unit II | VISUALIZATION TECHNIQUES FOR TIME-SERIES, TREES & GRAPHS
Mapping - Time series - Connections and correlations – Indicator-Area chart-Pivot table-Scatter charts, Scatter maps - Tree maps, Space filling and non-space filling methods- Hierarchies and Recursion - Networks and Graphs-Displaying Arbitrary Graphs-node link graph-Matrix representation for graphs- Info graphics

Unit III | TEXT AND DOCUMENT VISUALIZATION

Unit IV | INTERACTIVE DATA VISUALIZATION

Unit V | SECURITY IN DATA VISUALIZATION
Port scan visualization - Vulnerability assessment and exploitation - Firewall log visualization - Intrusion detection log visualization -Attacking and defending visualization systems – Creating secured visualization system..

Total: 45 Periods
Course Outcomes:
Upon completion of the course, the students will be able to
- Apply mathematics and basic science knowledge for designing information visualizing System.
- Collect data ethically and solve engineering problem in visualizing the information.
- Implement algorithms and techniques for interactive information visualization.
- Conduct experiments by applying various modern visualization tool and solve the space layout problem.
- Analyze and design system to visualize multidisciplinary multivariate Data individually or in teams.

Develop a cost effective and a scalable information visualization system.

References

<table>
<thead>
<tr>
<th></th>
<th>Author/Title</th>
</tr>
</thead>
</table>
DATA SECURITY AND PRIVACY

<table>
<thead>
<tr>
<th>Unit</th>
<th>Course Title</th>
<th>Credits</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unit I</td>
<td>ATTACKS AND PRIVACY</td>
<td>9</td>
</tr>
<tr>
<td>Unit II</td>
<td>ENCRYPTED COMPUTATION</td>
<td>9</td>
</tr>
<tr>
<td>Unit III</td>
<td>DATA GOVERNANCE AND PRIVACY APPROACHES</td>
<td>9</td>
</tr>
<tr>
<td>Unit IV</td>
<td>FEDERATED LEARNING AND DATA SCIENCE</td>
<td>9</td>
</tr>
<tr>
<td>Unit V</td>
<td>LEGALITY OF PRIVACY</td>
<td>9</td>
</tr>
</tbody>
</table>

Course Outcomes:
- Gain knowledge on the nature of attacks and threats and security management goals and framework
- Knowledge on the landscape of hacking and defense mechanisms
- Able to differentiate and integrate strategies for data security and protecting critical infrastructure
- Able to understand policies to mitigate data security breaching
- Knowledge on IT Act, and amendments, copyright, IPR and cyber law to deal with offenses.

References

1. Katharine Jarmul, Practical Data Privacy, O'Reilly Media, Inc, 2023
2. David Evans, Vladimir Kolesnikov and Mike Rosulek, A Pragmatic Introduction to Secure Multi-Party Computation, NOW Publishers, 2022 (Free access at https://securecomputation.org/)
**Course Objectives:**
- To understand big data.
- To learn and use NoSQL big data management.
- To learn MapReduce analytics using Hadoop and related tools.
- To work with MapReduce applications.
- To understand the usage of Hadoop related tools for Big Data Analytics.

<table>
<thead>
<tr>
<th>Unit I</th>
<th>UNDERSTANDING BIG DATA</th>
<th>5</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Unit II</th>
<th>NOSQL DATA MANAGEMENT</th>
<th>7</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Unit III</th>
<th>MAP REDUCE APPLICATIONS</th>
<th>6</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Unit IV</th>
<th>BASICS OF HADOOP</th>
<th>6</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Unit V</th>
<th>HADOOP RELATED TOOLS</th>
<th>6</th>
</tr>
</thead>
</table>

**Total:** 30 Periods
Course Outcomes:
After the completion of this course, students will be able to:

- Describe the big data and use cases from selected business domains.
- Explain NoSQL big data management.
- Install, configure and run Hadoop and HDFS.
- Perform map-reduce analytics using Hadoop.
- Use Hadoop-related tools such as HBase, Cassandra, Pig and Hive for big data analytics.

References

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>Sadalage, Pramod J. “NoSQL distilled”, 2013</td>
</tr>
</tbody>
</table>
Course Objectives:
- To outline an overview of exploratory data analysis.
- To implement data visualization using Matplotlib.
- To perform univariate data exploration and analysis.
- To apply bivariate data exploration and analysis.
- To use Data exploration and visualization techniques for multivariate and time series data.

Unit I EXPLORATORY DATA ANALYSIS

Unit II EDA USING PYTHON

Unit III UNIVARIATE ANALYSIS

Unit IV BIVARIATE ANALYSIS
Relationships between Two Variables – Percentage Tables – Analysis Contingency Tables – Handling Several Batches – Scatterplots and Resistant Lines.

Unit V MULTIVARIATE AND TIME SERIES ANALYSIS

Total: 45 Periods
### COURSE OUTCOMES:

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

<table>
<thead>
<tr>
<th>CO1</th>
<th>Understand the fundamentals of exploratory data analysis.</th>
</tr>
</thead>
<tbody>
<tr>
<td>CO2</td>
<td>Implement the data visualization using Matplotlib.</td>
</tr>
<tr>
<td>CO3</td>
<td>Perform univariate data exploration and analysis.</td>
</tr>
<tr>
<td>CO4</td>
<td>Apply bivariate data exploration and analysis.</td>
</tr>
<tr>
<td>CO5</td>
<td>Use Data exploration and visualization techniques for multivariate and time series data.</td>
</tr>
</tbody>
</table>

Total: 60 Periods

### REFERENCES:

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
</table>