|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | | | | **Course Title: DATA SCIENCE** | | | | | | | | | | | | | | |
| **Course Code: 18CS654** | | | | | **No. of Credits: 3: 0: 0**  **(L-T-P)** | | | | | **No. of lecture hours/week: 3** | | | | |
| **Exam Duration: 3 hours** | | | | | **CIE + Assignment + SEE = 45 + 5 + 50 = 100** | | | | | **Total No. of Contact Hours: 42** | | | | |
|  | | | | | | | | | | | | | | | | | | |
| **Course Objectives:** | | | | **Description** | | | | | | | | | | | | | | |
| 1. Determine the appropriate natural language processing, machine learning and deep learning models to solve the business-related challenges. 2. Indicate proficiency with statistical analysis of data to derive insight from results and interpret the data findings visually. 3. Demonstrate skills in data management by obtaining, cleaning and transforming the data. 4. Discuss how social networks appraise the ways in which the social clustering shape individuals and groups in contemporary society. | | | | | | | | | | | | | | |
|  | | | | | | | | | | | | | | | | | | |
| **Unit No** | **Syllabus Content** | | | | | | | | | | | | | | | | **No of Hours** | |
| **1.** | **Visualizing Data,** matplotlib, Bar Charts, Line Charts, Scatterplots, **Linear Algebra,** Vectors, Matrices, **Statistics,** Describing a Single Set of Data, Correlation, Simpson’s Paradox, Some Other Correlational Caveats, Correlation and Causation, **Probability,** Dependence and Independence, Conditional Probability, Bayes’s Theorem, Random Variables, Continuous Distributions, The Normal Distribution, The Central Limit Theorem. | | | | | | | | | | | | | | | | **08** | |
| **2.** | **Hypothesis and Inference,** Statistical Hypothesis Testing, Example: Flipping a Coin, p-Values, Confidence Intervals, p-Hacking, Example: Running an A/B Test, Bayesian Inference, **Gradient Descent,** The Idea Behind Gradient Descent Estimating the Gradient, Using the Gradient, Choosing the Right Step Size, Using Gradient Descent to Fit Models, Minibatch and Stochastic Gradient Descent, **Getting Data,** stdin and stdout, Reading Files, Scraping the Web, Using APIs, Example: Using the Twitter APIs, **Working with Data,** Exploring Your Data, Using NamedTuples, Dataclasses, Cleaning and Munging, Manipulating Data, Rescaling, An Aside: tqdm, Dimensionality Reduction. | | | | | | | | | | | | | | | | **08** | |
| **3.** | **Machine Learning,** Modeling, What Is Machine Learning?, Overfitting and Underfitting, Correctness, The Bias-Variance Tradeoff, Feature Extraction and Selection, **k-Nearest Neighbors,** The Model, Example: The Iris Dataset, The Curse of Dimensionality, **Naive Bayes,** A Really Dumb Spam Filter, A More Sophisticated Spam Filter, Implementation, Testing Our Model, Using Our Model, **Simple Linear Regression,** The Model, Using Gradient Descent, Maximum Likelihood Estimation, **Multiple Regression,** The Model, Further Assumptions of the Least Squares Model, Fitting the Model, Interpreting the Model, Goodness of Fit, Digression: The Bootstrap, Standard Errors of Regression Coefficients, Regularization, **Logistic Regression,** The Problem, The Logistic Function, Applying the Model, Goodness of Fit, Support Vector Machines. | | | | | | | | | | | | | | | | **09** | |
| **4.** | **Decision Trees,** What Is a Decision Tree?, Entropy, The Entropy of a Partition, Creating a Decision Tree, Putting It All Together, Random Forests, **Neural Networks,** Perceptrons, Feed-Forward Neural Networks, Backpropagation, Example: Fizz Buzz,  **Deep Learning,** The Tensor, The Layer Abstraction, The Linear Layer, Neural Networks as a Sequence of Layers, Loss and Optimization, Example: XOR Revisited, Other Activation Functions, Example: FizzBuzz Revisited, Softmaxes and Cross-Entropy, Dropout, Example: MNIST, Saving and Loading Models, **Clustering,** The Idea, The Model, Example: Meetups, Choosing k, Example: Clustering Colors, Bottom-Up Hierarchical Clustering. | | | | | | | | | | | | | | | | **09** | |
| **5.** | **SELF-STUDY**  **Natural Language Processing,** Word Clouds, n-Gram Language Models, Grammars, An Aside: Gibbs Sampling, Topic Modeling, Word Vectors, Recurrent Neural Networks, Example: Using a Character-Level RNN, **Network Analysis,** Betweenness Centrality, Eigenvector Centrality, Directed Graphs and PageRank, **Recommender Systems,** Manual Curation, Recommending What’s Popular, User-Based Collaborative Filtering, Item-Based Collaborative Filtering, Matrix Factorization. | | | | | | | | | | | | | | | | **08** | |
| **Course Outcomes** | | | **Description** | | | | | | | | | | | | | | **RBT Levels** | |
| **CO1** | | | Interpret the concepts and methods of mathematical disciplines relevant to data analytics and statistical modeling. | | | | | | | | | | | | | | **L3** | |
| **CO2** | | | Examine, visualize, curate, and prepare data and recognize how the quality of the data and the means of data collection may affect interpretation. | | | | | | | | | | | | | | **L3** | |
| **CO3** | | | Determine the machine learning, deep learning and natural language processing skills to design and implement efficient, data-driven solutions for real world problems. | | | | | | | | | | | | | | **L3** | |
| **CO4** | | | Illustrate how network analysis and recommender systems can contribute to increasing knowledge about diverse aspects of societal clustering. | | | | | | | | | | | | | | **L3** | |
| **CO-PO Mapping** | | **PO1** | | | **PO2** | | **PO3** | **PO4** | | **PO5** | **PO6** | **PO7** | **PO8** | **PO9** | **PO10** | **PO11** | | **PO12** |
| **CO1** | | 3 | | | 3 | | 2 | 2 | | 3 | - | - | - | - | - | - | | - |
| **CO2** | | 2 | | | 2 | | 2 | 2 | | 3 | - | - | - | - | - | - | | - |
| **CO3** | | 3 | | | 3 | | 3 | 3 | | 3 | - | - | - | - | - | - | | - |
| **CO4** | | 3 | | | 2 | | 2 | 2 | | 3 | - | - | - | - | - | - | | - |
| **Strong -3 Medium -2 Weak -1** | | | | | | | | | | | | | | | | | | |
| **TEXT BOOKS:** | | | | | | | | | | | | | | | | | | |
| 1. Joel Grus, **“Data Science from Scratch”,** 2nd Edition, O’Reilly Publications/Shroff Publishers and Distributors Pvt. Ltd., 2019. ISBN-13: 978-9352138326. | | | | | | | | | | | | | | | | | | |
| **REFERENCE BOOKS:** | | | | | | | | | | | | | | | | | | |
| 1. Emily Robinson and Jacqueline Nolis, **“Build a Career in Data Science”,** 1st Edition, Manning Publications, 2020. ISBN: 978-1617296246. 2. Aurélien Géron, **“Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems”,** 2nd Edition, O'Reilly Publications/Shroff Publishers and Distributors Pvt. Ltd., 2019. ISBN-13: 978-1492032649. 3. François Chollet, **“Deep Learning with Python**”, 1st Edition, Manning Publications, 2017. ISBN-13: 978-1617294433 4. Jeremy Howard and Sylvain Gugger, **“Deep Learning for Coders with fastai and PyTorch”,** 1st Edition, O'Reilly Publications/Shroff Publishers and Distributors Pvt. Ltd., 2020. ISBN-13: 978-1492045526. 5. Sebastian Raschka and Vahid Mirjalili, **“Python Machine Learning: Machine Learning and Deep Learning with Python, scikit-learn, and TensorFlow 2”**, 3rd Edition, Packt Publishing Limited, 2019. ISBN-13: 978-1789955750 | | | | | | | | | | | | | | | | | | |
| **SELF-STUDY REFERENCES/WEBLINKS:** | | | | | | | | | | | | | | | | | | |
| 1. **Natural Language Processing**   https://www.youtube.com/watch?v=xvqsFTUsOmc   1. **Network Analysis**   https://www.youtube.com/watch?v=K5xiFDClgjo   1. **Recommender Systems**   https://www.youtube.com/watch?v=39vJRxIPSxw | | | | | | | | | | | | | | | | | | |
| **COURSE COORDINATOR:** | | | | | | **Dr.Gowrishankar S.** | | | | | | | | | | | | |