**Detailed Syllabus**

Lecture-wise Breakup

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Course Code** | **20B12CS333** | **Semester ODD** | | **Semester**   **Session** 2021 -2022 Month from JUL-DEC | |
| **Course Name** | **Introduction to Big Data & Data Analytics** | | | | |
| **Credits** | 3 | | **Contact Hours** | | 3-1-0 (4 hrs per week) |

|  |  |  |
| --- | --- | --- |
| **Faculty (Names)** | **Coordinator(s)** | **Dr. Bharat Gupta (62), Dr. Neeraj Jain (128)** |
| **Teacher(s) (Alphabetically)** |  |

|  |  |  |
| --- | --- | --- |
| **COURSE OUTCOMES** | | **COGNITIVE LEVELS** |
| C330-3.1 | Understand the fundamental concepts of an exciting growing field of big data analytics | Understanding [Level 2] |
| C330-3.2 | Demonstrate the tools required to manage and analyze big data like Hadoop, NoSql, MapReduce | Apply [Level 3] |
| C330-3.3 | Apply predictive models and advanced computing paradigms for big data analytics | Apply [Level 3] |
| C330-3.4 | Analyze the big data using intelligent & visualization techniques | Analyze [Level 5] |
| C330-3.5 | Design and create predictive and mathematical model to solve complex real-world problems for decision making. | Create [Level 6] |

|  |  |  |  |
| --- | --- | --- | --- |
| **Module No.** | **Title of the Module** | **Topics in the Module** | **No. of Lectures for the module** |
| **1.** | Introduction to Big Data | Introduction to Big Data landscape, Big Data: Why and where, Characteristics of Big Data (V’s of Big Data (volume, velocity, variety, veracity, valence, and value) and Dimensions of Scalability, Data Models for Big Data Products(NOSQL, NEWSQL,HADOOP),Data Science and Analytics. | 7 |
| **2.** | Data Visualization Techniques | Introduction to Python or R, Understanding and Visualizing Data, Data Visualization R/Python | 5 |
| **3.** | Data Modeling and Optimization | Modeling Uncertainty and Risk, Optimization and Modeling Simultaneous Decisions, Case Study | 5 |
| **4.** | Decision Making and Predictive Analytics-1 | Data exploration, Evaluation methods, Regression Techniques, Classification Techniques, Case Study | 9 |
| **5.** | Decision Making and Predictive Analytics-2 | Clustering Techniques, Anomaly Detection, Dimensionality Reduction, Neural networks for deep learning, Hands-on using Python/R, Case Study | 9 |
| **6.** | Big Data Technologies | Using Hadoop to store data(HDFS, HBASE), Process Data using Map Reduce, Testing and Debugging Map Reduce Applications | 7 |
| **Total number of Lectures** | | | **42** |
| **Evaluation Criteria**  **Components Maximum Marks**   |  |  | | --- | --- | | T1 | 20 | | T2 | 20 | | End Semester Examination | 35 | | TA | 25 (**Attendance-07, Class Test/ Quizzes-07, Internal assessment-05,** **PBL mode-06**) |     **Total 100** | | | |

Project based learning: The students are grouped into groups of size 5-6 and will be implementing a decision making and predictive analytics techniques for big data. The student will analyze the big data and select appropriate technique for processing. This will help in the employability of students in the data science and big data sector.

|  |  |
| --- | --- |
| **Recommended Reading material:** Author(s), Title, Edition, Publisher, Year of Publication etc. ( Text books, Reference Books, Journals, Reports, Websites etc. in the IEEE format) | |
| Reference Books: | |
| **1.** | Dey, N., Hassanien, A. E., Bhatt, C., Ashour, A., & Satapathy, S. C. (Eds.). (2018). Internet of things and big data analytics toward next-generation intelligence (pp. 3-549). Berlin: Springer. |
| **2.** | Marz, N., & Warren, J. (2015). Big Data: Principles and best practices of scalable real time data systems. Manning Publications Co. |
| **3.** | Grover, M., Malaska, T., Seidman, J., & Shapira, G. (2015). Hadoop Application Architectures: Designing Real-World Big Data Applications. " O'Reilly Media, Inc.". |
| **4.** | Covington, D. (2016). Analytics: Data Science, Data Analysis, and Predictive Analytics for Business. CreateSpace Independent Publishing Platform. |
| **Text Books:** | |
| **5.** | EMC Education Services. (2015). Data Science and Big Data Analytics: Discovering, Analyzing, Visualizing and Presenting Data. Wiley. |
| **6.** | Nelli, F. (2018). Python data analytics: with pandas, numpy, and matplotlib. Apress. |
| **7.** | Sedkaoui, S. (2018). Data analytics and big data. John Wiley & Sons. |
| **8.** | Erl, T., Khattak, W., & Buhler, P. (2016). Big data fundamentals: concepts, drivers & techniques. Prentice Hall Press. |
| **9.** | Dasgupta, N. (2018). Practical big data analytics: Hands-on techniques to implement enterprise analytics and machine learning using Hadoop, Spark, NoSQL and R. Packt Publishing Ltd. |
| **10.** | Kumar, V. N., & Shindgikar, P. (2018). Modern Big Data processing with Hadoop: Expert techniques for architecting end-to-end Big Data solutions to get valuable insights. Packt Publishing Ltd. |