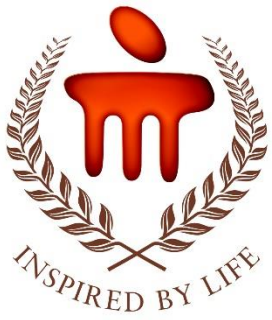

MANIPAL
ACADEMY of HIGHER EDUCATION
(Institution of Eminence Deemed to be University)

Master of Engineering - ME (Big Data Analytics)

Syllabus

July 2023 Onwards

MANIPAL SCHOOL OF INFORMATION SCIENCES
MANIPAL ACADEMY OF HIGHER EDUCATION
MANIPAL – 576104.KARNATAKA. INDIA.

Program Educational Objectives / Outcomes (PEOs)

PEO 1: Develop in depth understanding of the key technologies in data engineering, data science and business analytics.

PEO 2: Practice problem analysis and decision-making using machine learning techniques.

PEO 3: Gain practical, hands-on experience with statistics, programming languages and big data tools through coursework and applied research experiences.

Program Objectives / Outcomes (POs)

PO 1: An ability to independently carry out research /investigation and development work to solve practical problems.

PO 2: An ability to write and present a substantial technical report/document.

PO 3: An ability to demonstrate a degree of mastery over the area as per the specialization of the program. The mastery should be at a level higher than the requirements in the appropriate bachelor program.

PO 4: An ability to develop and implement big data analysis strategies based on theoretical principles, ethical considerations, and detailed knowledge of the underlying data.

P0 5: An ability to demonstrate knowledge of the underlying principles and evaluation methods for analyzing data for decision-making.

Program Structure

ME (Big Data Analytics) – I Semester									
Course Code	Course Name	No. of Hours/week				Duration of Exam in Hrs	Maximum Marks		
		Lecture	Tutorial	Practical	Credit		Internal 50	External 50	Total 100
BDA 5101	Algorithms and Data Structures for Big Data	3	-	-	3	3	50	50	100
BDA 5102	Architecture of Big Data Systems	3	-	-	3	3	50	50	100
BDA 5103	Fundamentals of Machine Learning	3	-	-	3	3	50	50	100
AML 5103	Applied Probability and Statistics	3	-	-	3	3	50	50	100
	Elective – I	3	-	-	3	3	50	50	100
BDA 5151	Algorithms and Data Structures for Big Data Lab	-	-	3	1	3	50	50	100
BDA 5152	Architecture of Big Data Systems Lab	-	-	3	1	3	50	50	100
BDA 5153	Fundamentals of Machine Learning Lab	-	-	3	1	3	50	50	100
AML 5153	Applied Probability and Statistics Lab	-	-	3	1	3	50	50	100
	Elective – I Lab	-	-	3	1	3	50	50	100
MPT 5100	Mini Project – I	-	-	-	4	-	100	-	100
PSD 5100	Professional Skill Development – I	-	-	-	1	-	100	-	100
Total		**15**	**-**	**15**	**25**				

ME (Big Data Analytics) – II Semester									
Course Code	Course Name	No. of Hrs. / week				Duration of Exam in Hrs	Maximum Marks		
		Lecture	Tutorial	Practical	Credit		Internal 50	External 50	Total 100
BDA 5201	Machine Learning for Big Data	3	-	-	3	3	50	50	100
BDA 5202	Modern Databases for Big Data	3	-	-	3	3	50	50	100
BDA 5203	Multimedia Analytics	3	-	-	3	3	50	50	100
AML 5201	Advanced Applications of Probability and Statistics	3	-	-	3	3	50	50	100
	Elective – II	3	-	-	3	3	50	50	100
BDA 5251	Machine Learning for Big Data Lab	-	-	3	1	3	50	50	100
BDA 5252	Modern Databases for Big Data Lab	-	-	3	1	3	50	50	100
BDA 5253	Multimedia Analytics Lab	-	-	3	1	3	50	50	100
AML 5251	Advanced Applications of Probability and Statistics	-	-	3	1	3	50	50	100
	Elective – II Lab	-	-	3	1	3	50	50	100
MPT 5200	Mini Project – II	-	-	-	4	-	100	-	100
PSD 5200	Professional Skill Development – II	-	-	-	1	-	100	-	100
TOTAL		**15**	**-**	**15**	**25**				

ME (Big Data Analytics) – III & IV Semesters									
BDA 6098	Project Work	-	-	-	25				
Total Number of Credits to Award Degree							75		

List of Electives (Theory)

Elective - I		Elective - II	
Course Code	**Course Name**	**Course Code**	**Course Name**
BDA 5132	Principles of Data Visualization	BDA 5231	Natural Language and Text Processing
ESD 5234	Mobile Web Application Development	CDC 5001	DevOps for Cloud
		ENP 5230	Entrepreneurship

List of Electives (Lab)

Elective - I		Elective - II	
Course Code	**Course Name**	**Course Code**	**Course Name**
BDA 5182	Principles of Data Visualization Lab	BDA 5281	Natural Language and Text Processing Lab
ESD 5284	Mobile Web Application Development Lab	CDC 5051	DevOps for Cloud Lab
		ENP 5280	Entrepreneurship Lab

SEMESTER I

BDA 5101: Algorithms and Data Structures for Big Data	L	T	P	C	Total hours
	3	0	0	3	36

Course Outcome

1. Analyse recursive programs, solve a general class of recurrence relations
2. Design programs for implementation of linked lists, stack, queues and binary search tree.
3. Design programs for dictionary, hash tables, graphs and shortest path techniques, sorting and searching.
4. Design string and text processing programs.

Unit	Topics	No. of Hours
I	**Title: Algorithm specification and analysis techniques** Analysis of recursive programs. Solving recurrence equations. General solution for a large class of recurrences.	3
II	Elementary data structures Implementation of lists, stacks, queues, Trees	10
III	**Sorting & Searching, Hash Tables, Graph** Quick sort, heap sort, merge sort. Linear search and binary search. Hashing and Dictionaries Representation of graphs. Depth First Searching. Breadth First Searching. Minimum cost spanning tree. Single source shortest paths and all-pairs shortest path	14
IV	**Title: String and text processing techniques, Data stream algorithms** Pattern-Matching Algorithms. Text Compression. Tries Sampling, Random Projections, Basic Algorithmic Techniques. Group Testing, Tree Method and Graph sketching.	9

References

1. Introduction to Algorithms - Thomas H. Cormen, Charles E. Leiserson, Ronald L. Rivest. MIT Press.
2. Data Structures and Algorithms - Aho, Hopcroft and Ulmann. Pearson Publishers.

3. Data Structures and Algorithms in Python - Michael T. Goodrich, Roberto Tamassia, and Michael H. Goldwasser. John Wiley & Sons.
4. Data Streams: Algorithms and Applications - S. Muthukrishnan. Foundations and Trends in Theoretical Computer Science archive, Volume 1 Issue 2, August 2005, Pages 117 – 236
5. https://in.coursera.org/specializations/boulder-data-structures-algorithms

BDA 5102: Architecture of Big Data Systems	L	T	P	C	Total hours
	3	0	0	3	36

Course Outcome

1. Apply various techniques to examine different types of data and understand lambda architecture.
2. Apply different tools and frameworks of Hadoop eco-system.
3. Apply Spark engine to process real-time data.
4. Design applications to handle batch and streaming data using Hadoop and Spark tools.

Unit	Topics	No. of Hours
I	**Classifying Big Data Characteristics and Big Data processing - the Lambda architecture** Analysis type, Processing methodology, Data Types, Data sources Different data storing and processing layers and architecture: Batch layer, Serving layer and Speed layer	12
II	**Batch layer, Serving layer and Speed layer** Choosing a storage solution for the batch layer: Distributed file systems, Vertical partitioning. MapReduce: a paradigm for Big Data computing. Performance metrics for the serving layer. Speed layer.	9
III	**Spark: Alternatives to MapReduce** Spark Architecture, Spark Session, DataFrame, Transformations and Actions, Spark SQL, Resilient Distributed Datasets (RDDs)	9

IV	**Stream Processing and Machine Learning using Spark** Advantages and challenges of stream processing, Stream Processing Design Points, Streaming APIs, Structured Stream Processing High level M-Lib concepts and M-Lib in Action.	6

References

1. Big Data: Principles and best practices of scalable real-time data systems - Nathan Marz and James Warren. Manning Publisher.
2. Hadoop: The Definitive Guide: Storage and Analysis at Internet Scale – Tom White, O'Reilly Publication 4th Edition.
3. Spark: The Definitive Guide: Big Data Processing Made Simple – Bill Chambers, Matei Zaharia, O'Reilly Publication 1st Edition.
4. https://learn.microsoft.com/en-us/training/modules/explore-core-data-concepts/
5. https://learn.microsoft.com/en-us/training/modules/explore-fundamentals-stream-processing/

BDA 5103: Fundamentals of Machine Learning	L	T	P	C	Total hours
	3	0	0	3	36

Course Outcome:

1. **Apply** inductive classification to obtain the hypothesis space from the training examples.
2. **Analyze** the performance of the decision tree machine learning models for classification.
3. **Evaluate** sample and computational complexities of the training samples for machine learning models.
4. **Apply** Bayesian learning and instances-based learning methods.
5. **Analyze** the performance of different Ensemble methods.

Unit	Topics	No. of Hours
I	**Introduction** - Inductive Classification - Decision Tree learning	14

	Definition of Machine Learning, Goals and applications of machine learning, Basic design issues and approaches to machine learning, Types of machine learning techniques Inductive Classification: The concept learning task, Concept learning as search through a hypothesis space, General-to-specific ordering of hypotheses, Finding maximally specific hypotheses, Version spaces and the candidate elimination algorithm, Inductive bias. Decision Tree learning: Representing concepts as decision trees, Recursive induction of decision trees, Picking the best splitting attribute, Entropy and information gain, Searching for simple trees and computational complexity.	
II	**Computational learning theory** - Bayesian learning - Instance-based learning **Computational learning theory**: Models of learnability: learning in the limit, Probably Approximately Correct (PAC) learning, Sample Complexity: quantifying the number of examples needed to PAC learn, Computational complexity of training. Sample complexity for finite hypothesis spaces, Noise Learning Multiple Classes, Bias-variance trade-off, under-fitting and over-fitting concepts **Bayesian learning:** Probability theory and Bayes rule, Naive Bayes learning algorithm - Parameter smoothing, Generative vs. discriminative training, Logistic regression, Bayes nets and Markov nets for representing dependencies **Instance-based learning:** Constructing explicit generalizations versus comparing to past specific examples, K-Nearest Neighbour learning algorithm, Case-based reasoning (CBR) learning	14
III	**Continuous Latent Variables** - Ensemble methods (bagging and boosting) **Continuous Latent Variables:** Principal Component Analysis (PCA), Applications of PCA **Ensemble methods (bagging and boosting):** Using committees of multiple hypotheses, Bagging, Boosting, DECORATE, Active learning with ensembles	8

References

1. Machine Learning, T. Mitchell, McGraw-Hill, 1997
2. Machine Learning, E. Alpaydin, MIT Press, 2010
3. Pattern Recognition and Machine Learning, C. Bishop, Springer, 2006
4. Pattern Classification, R. Duda, E. Hart, and D. Stork, Wiley-Interscience, 2000
5. T. Hastie, R. Tibshirani and J. Friedman, The Elements of Statistical Learning: Data Mining,

6. Inference and Prediction. Springer, 2nd Edition, 2009
7. Machine Learning for Big Data, Jason Bell, Wiley Big Data Series
8. Multidimensional Neural Networks Unified Theory, Rama Murthy G
9. https://www.coursera.org/specializations/machine-learning-introduction

AML 5103: Applied Probability and Statistics	L	T	P	C	Total hours
	3	0	0	3	36

Course Outcome

1. Model random phenomena using random variables.
2. Construct Bayesian models for quantifying uncertainty in practical problems.
3. Use sample information and perform hypothesis-test analysis using an appropriate statistical technique to explain attributes of a population.

Unit	Topics	No. of Hours
I	**Counting; Probability Concepts; Conditional Probability:** Multiplication rule; permutation; combination - Sampling: with/without replacement and order matters/does not matter - Binomial & multinomial coefficients - Distribution problems; Set theory; sample space; outcomes; events - Frequency based definition of probability - Equally likely vs. not equally likely outcomes - Axioms of probability Conditional probability; probability tree model; chain rule - Decomposition and the law of total probability - Bayes' rule - intuition, dependence/independence of events.	8
II	**Random Variables:** Modelling using discrete random variables: Bernoulli, geometric, binomial, negative binomial, hypergeometric, and Poisson distributions - Probability mass function and cumulative distribution function - Expectation and variance: discrete case - Modelling using continuous random variables: uniform, normal, log-normal, exponential, and beta distributions; probability density function - Expectation and variance: continuous case - Functions of random variables.	16
III	**Sampling and Parameter Estimation:** Population and sample - Statistic & sampling distribution - Sample mean and variance - Central limit theorem – intuition and applications	12

	Point estimation - Standard error - Interval estimation: interpretation of confidence interval - Hypothesis testing: p-values, significance level and their interpretations, application to analysis of one- /two-sample mean and paired data.	

References

1. Digital Dice: Computational Solutions to Practical Probability Problems, Paul Nahin, Princeton University Press
2. Applied Data Science with R Specialization from IBM (Courses 1, 3, and 4, https://www.coursera.org/specializations/applied-data-science-r)
3. Introduction to Probability, Charles M. Grinstead, American Mathematical Society; 2nd Revised Edition 1997. Available online at https://chance.dartmouth.edu/teaching_aids/books_articles/probability_book/amsbook.mac.pdf
4. A First Course in Probability, Sheldon Ross, 9th Edition, Pearson Education India; 9th Edition, 2013
5. Statistics without Tears: An Introduction for Non-Mathematicians (Paperback), Derek Rowntree, Penguin UK
6. Biostatistics Open Learning textbook – Online resource from University of Florida available at https://bolt.mph.ufl.edu/6050-6052/
7. All of Statistics: A Concise Course in Statistical Inference, Larry Wasserman – Springer

ELECTIVES - SEMESTER I

BDA 5132: Principles of Data Visualization	L	T	P	C	Total hours
	3	0	0	3	36

Course Outcome

1. Implement web scrapping techniques to extract data from websites.
2. Organize raw data for analysis using data manipulation techniques.
3. Use powerBI for preparation and modelling of data for analysis.
4. Interpret data using various data visualization techniques.
5. Report data for analytics and to manage workspace using Power BI

Unit	Topics	No. of Hours
I	**Web scraping:** Web scraping models and techniques, Usecases: BeautifulSoup, Scrapy.	6
II	**Data Analysis: Data structures for analysis:** numpy, pandas. Data Wrangling - Clean, Transform, Merge, Reshape, Data Aggregation and Group Operations. Prepare data for analysis using Power BI, Model data in Power BI	12
III	**Data Visualization:** Visualization techniques: time series, statistical distributions, maps - Data visualization for web. Visualize data in Power BI, Data analysis in Power BI, Manage workspaces and datasets in Power BI, Create and use analytics reports with Power BI, Manage workspaces and datasets in Power BI.	18

References

1. Website Scraping with Python: Using BeautifulSoup and Scrapy, Gábor & Hajba, APRESS Publications, 1st Edition, 2018.
2. Web Scraping with Python: Collecting More Data from the Modern Web, Ryan Mitchell Shroff, O'Reilly, 2nd Edition, 2018.
3. Designing Data Visualizations, Julie Steele and Noah Iliinsky; O'Reilly Media; 1st Edition, 2011.
4. Python for Data Analysis, Wes McKinney; Shroff; O'Reilly; 2nd Edition, 2018.
5. https://learn.microsoft.com/en-us/certifications/exams/pl-300/

ESD 5234: Mobile Web Application Development	L	T	P	C	Total hours
	3	0	0	3	36

Course Outcome

1. Discuss the challenges of mobile web application development.
2. Apply HTML5, CSS, javascript and DOM API's in web application development.
3. Use of programming for technologies available on smart phones.
4. Design and develop secure mobile web applications.

Unit	Topics	No. of Hours
I	**Challenges of mobile Web application development:** The limitations of mobile networks. Reducing the page weight – the amount of markup and external elements. Avoiding useless network usage. Understanding the "mobilefirst" design principles. Limitations imposed by battery life.	4
	Setting up a personal Web site: Setting free VMs - micro-instances - on AWS. Installing and configuring NGINX on AWS micro instances. Working with routing and reverse proxies HTTP and REST APIs.	5
	HTML5 and CSS for mobile devices: Media queries for handling mobile form-factors. Principles and practice of responsive design. Mobile UX, Viewport, Fluid design and responsive images.	5
II	**Programming with JavaScript and DOM APIs:** Accessing document fragments. Using jQuery and other light-weight libraries. AJAX and asynchronous programming.	5
	Architecture of Android applications: Android application framework, core libraries, android runtime, Linux kernel.	5
	Programming for technologies available on smart phones: Introduction to PhoneGap, Handling Touch events. Making use of the accelerometer and the Location APIs. Accessing camera and media devices.	4
III	**Developing offline facilities in mobile web applications:** Localstorage and IndexDB APIs	2
	Designing and developing secure mobile web applications: Understanding the single-origin policy, Dangers of Cross-site scripting Principles of the secure socket layer and HTTPS Practical encryption for client-server communication in Web applications. Best practices in developing secure client-side code.	6

References

1. Learning Web App Development (Build Quickly with Proven JavaScript Techniques) - Semmy Purewal. O'Reilly Media. 2014.
2. The Browser Security Handbook. Michal Zalewski. https://code.google.com/p/browsersec/wiki/Main
3. High Performance Responsive Design - Tom Barker. O'Reilly publisher. 2014.
4. Apple UI Design Basics.
5. Android Design Principles. https://developer.android.com/design/index.html
6. Android Application Development Reference. https://developer.android.com/develop/index.html

7. https://www.coursera.org/learn/aadcapstone

BDA 5151: Algorithms and Data structures for Big Data Lab	L	T	P	C	Total hours
	0	0	3	1	36

Course Outcome

1. Design programs for implementation of linked lists, stack and queues.
2. Design programs for implementation of binary search tree, sorting and searching, dictionary and Hash Table
3. Design programs for graphs and shortest path techniques.

Unit	Topics	No. of Hours
I	Implement single linked and double linked list Implement array based and linked list based Stack Implement array based and linked list based Queue	15
II	Implement binary tree and BST Design and Implement Hash functions, implement open and closed Hash Tables	6
III	Implement Graph algorithms and shortest path	15

References

1. Introduction to Algorithms - Thomas H. Cormen, Charles E. Leiserson, Ronald L. Rivest. MIT Press.
2. Data Structures and Algorithms - Aho, Hopcroft and Ulmann. Pearson Publishers.
3. Data Structures and Algorithms in Python - Michael T. Goodrich, Roberto Tamassia, and Michael H. Goldwasser. John Wiley & Sons.
4. Data Streams: Algorithms and Applications - S. Muthukrishnan. Foundations and Trends in Theoretical Computer Science archive, Volume 1 Issue 2, August 2005, Pages 117 – 236

BDA 5152: Architecture of Big Data Systems Lab	L	T	P	C	Total hours
	0	0	3	1	36

Course Outcome

1. Use data extraction tools to ingest various types of data into big data systems.
2. Experiment with different tools and frameworks of Hadoop eco-system.
3. Experiment with Spark Engine to process real-time data.
4. Design applications to handle batch and streaming data using Hadoop and Spark tools.

Unit	Topics	No. of Hours
I	Using HDFS commands to get familiarize with HDFS environment Transfer structured data to and from HDFS Creating Hive tables in HDFS and running SQL statements Developing and running MapReduce programs	15
II	Creating RDDs in spark Creating Spark DataFrames, loading, transforming and performing actions in Spark environment Handling streaming data using structured streaming	12
III	Developing applications using Hadoop tools Develop applications using Spark environment	9

References

1. Big Data: Principles and best practices of scalable real-time data systems - Nathan Marz and James Warren. Manning Publisher.
2. Hadoop: The Definitive Guide: Storage and Analysis at Internet Scale – Tom White, O'Reilly Publication 4th Edition.
3. Spark: The Definitive Guide: Big Data Processing Made Simple – Bill Chambers, Matei Zaharia, O'Reilly Publication 1st Edition.

BDA 5153: Fundamentals of Machine Learning Lab	L	T	P	C	Total hours
	0	0	3	1	36

Course Outcome

1. **Implement** NumPy and Pandas for data science operations with examples.
2. **Apply** Find-S and Candidate elimination algorithms to find the hypothesis space.
3. **Analyze** the performance of decision tree learning models for different data sets.
4. **Analyze** the performance of Bayesian learning, PCA and instances-based learning methods using different data sets.
5. **Evaluate** the performance of different Ensemble methods.

Unit	Topics	No. of Hours
I	Panda, Numpy.	6
II	Find S Algorithm, Candidate elimination.	6
III	Decision Tree Algorithm, Bayesian learning, PCA, Ensemble methods (bagging and boosting)	24

References

1. T. Mitchell, “Machine Learning”, McGraw-Hill, 1997.
2. E. Alpaydin, “Machine Learning”, MIT Press, 2010.
3. C. Bishop,” Pattern Recognition and Machine Learning”, Springer, 2006.
4. E. Hart, R. Duda and D. Stork, “Pattern Classification”, Wiley- Interscience, 2000.
5. T. Hastie, R. Tibshirani and J. Friedman, “The Elements of Statistical Learning: Data Mining,
6. Inference and Prediction”, Springer, 2nd Edition, 2009.
7. Jason Bell, “Machine Learning for Big Data”, Wiley Big Data Series, 2016.
8. Rama Murthy G,” Multidimensional Neural Networks Unified Theory”, New Age International, 2008

AML 5153: Applied Probability and Statistics Lab	L	T	P	C	Total hours
	0	0	3	1	36

Course Outcome

1. Visualize probability concepts through frequency-based interpretations.
2. Simulate discrete and continuous random variables for modelling random phenomena.
3. Design and apply hypothesis tests followed by interpretation of results.

Unit	Topics	No. of Hours
I	**Counting; Probability Concepts; Conditional Probability:** Understand the basic principles of the R programming language; Develop short code snippets to understand the basic principles of sampling and probability; Visualise and interpret probability concepts through a frequency-based approach; Program and analyse Bayesian models for practical problems.	8
II	**Random Variables:** Understand and apply R functions to simulate discrete and continuous random variables; Using sampling, compute and interpret different attributes of random variables; Visualise and interpret histograms and probability mass/density functions of random variables using state of the art visualisation libraries in R; Develop codes to model random phenomena using appropriate random variables.	16
III	**Sampling and Parameter Estimation:** Visualise sample data through histograms; Compute estimates of population parameters using samples and communicate the uncertainty in the estimates; Use R in-built functions for performing hypothesis tests; Interpret and communicate the results of hypothesis tests.	12

References

1. Digital Dice: Computational Solutions to Practical Probability Problems, Paul Nahin, Princeton University Press
2. Applied Data Science with R Specialization from IBM (Courses 1, 3, and 4, https://www.coursera.org/specializations/applied-data-science-r)
3. Introduction to Probability, Charles M. Grinstead, American Mathematical Society; 2nd Revised Edition 1997. Available online at https://chance.dartmouth.edu/teaching_aids/books_articles/probability_book/amsbook.mac.pdf
4. A First Course in Probability, Sheldon Ross, 9th Edition, Pearson Education India; 9th Edition, 2013

5. Statistics without Tears: An Introduction for Non-Mathematicians (Paperback), Derek Rowntree, Penguin UK
6. Biostatistics Open Learning textbook – Online resource from University of Florida available at https://bolt.mph.ufl.edu/6050-6052/
7. All of Statistics: A Concise Course in Statistical Inference, Larry Wasserman – Springer

BDA 5182: Principles of Data Visualization Lab	L	T	P	C	Total hours
	0	0	3	1	36

Course Outcome

1. Experiment web scrapping techniques to extract data from websites.
2. Implement NumPy and Pandas for data science operations with examples.
3. Organize data for visualization using data manipulation techniques.
4. Experiment different visualization techniques.
5. Use power BI for analytics and to manage workspace.

Unit	Topics	No. of Hours
I	Design programs to dynamically extract data from web.	12
II	Understand and integrate various data structures for data analysis process. Create various techniques to clean and handle missing data. Design data filtering and transformation techniques.	9
III	Describe what is the purpose of Visualization and ways of classifying visualization. Create visualization for time series data, statistical distributions. Create visualization for maps, Hierarchical data and network data. Incorporate PowerBI to plot visualization	15

References

1. Website Scraping with Python: Using BeautifulSoup and Scrapy, Gábor & Hajba, APRESS Publications, 1st Edition, 2018.

2. Web Scraping with Python: Collecting More Data from the Modern Web, Ryan Mitchell Shroff, O'Reilly, 2nd Edition, 2018.
3. Designing Data Visualizations, Julie Steele and Noah Iliinsky; O'Reilly Media; 1st Edition, 2011.
4. Python for Data Analysis, Wes McKinney; Shroff; O'Reilly; 2nd Edition, 2018.
5. https://learn.microsoft.com/en-us/certifications/exams/pl-300/

ESD 5284: Mobile Web Application Development Lab	L	T	P	C	Total hours
	0	0	3	1	36

Course Outcome

1. Implement the Web applications by applying HTML5, CSS, JavaScript and DOM API's concepts.
2. Apply the programming skills for the technologies available on smart phones.
3. Development of secure mobile web applications.

Unit	Topics	No. of Hours
I	HTML5 and CSS for mobile devices: Media queries for handling mobile form-factors. Principles and practice of responsive design. Mobile UX, Viewport, Fluid design and responsive images. Accessing document fragments. Using jQuery and other light-weight libraries. AJAX and asynchronous programming	18
II	Introduction to PhoneGap, Handling Touch events. Making use of the accelerometer and the Location APIs. Accessing camera and media devices.	12
III	Principles of the secure socket layer and HTTPS Practical encryption for client-server communication in Web applications. Best practices in developing secure client-side code	6

References

1. Learning Web App Development (Build Quickly with Proven JavaScript Techniques) - Semmy Purewal. O'Reilly Media. 2014.
2. The Browser Security Handbook. Michal Zalewski. https://code.google.com/p/browsersec/wiki/Main
3. High Performance Responsive Design - Tom Barker. O'Reilly publisher. 2014.
4. Apple UI Design Basics.

5. Android Design Principles. https://developer.android.com/design/index.html
6. Android Application Development Reference. https://developer.android.com/develop/index.html

MPT 5100: Mini Project - I	L	T	P	C	Total hours
	0	0	0	4	48

Course Outcome

1. Identify the real-world and socially relevant problems and perform feasibility analysis for finding solutions.
2. Organize work effectively as a member in a team, examine, experiment, and communicate technical information constructively.
3. Develop and implement solutions to the identified problems by applying research methodology and development life cycle with appropriate documentation by incorporating ethical standards.

Unit	Topics	No. of Hours
I	Problem identification, literature survey, formation of detailed specifications.	48
II	Design and implementation of the proposed system architecture.	
III	Demonstrate an ability to present and defend project work carried out to a panel of experts.	

References

1. Research articles and Online Resources.

PSD 5100: Professional Skill Development - I	L	T	P	C	Total hours
	0	0	0	1	12

Course Outcome

1. Identify and synthesize important themes in the field of engineering which transform socio-economic ecosystem.
2. Develop competence to communicate effectively in oral and written forms.
3. Effective management of time, involve in reflective learning and adhere to the professional code of conduct.

Unit	Topics	No. of Hours
I	Report writing involves identifying the topic of interest from current issues in the domain of engineering and technology or inter disciplinary domains, then framing the order in the report, writing abstract, deciding on the content itself, conclusion and future scope of the topic and properly citing the references from bibliography.	12
II	Presenting in classroom to audience where content spoken, the conceptual knowledge and presentation skills (like audibility, eye contact, memory) of speaker is assessed.	

References

1. Research articles and Online Resources.

SEMESTER II

BDA 5201: Machine Learning for Big Data	L	T	P	C	Total hours
	3	0	0	3	36

Course Outcome

1. Apply Artificial Neural Network, Clustering, Support Vector Machine, Deep Neural Network and Reinforcement Learning models.
2. Analyze the performance of single layer, multilayer, and deep neural networks.
3. Compare the performance of different clustering algorithms.
4. Evaluate the performance of different types of artificial neural network models, clustering models, deep neural network models, and reinforcement learning models.

Unit	Topics	No. of Hours
I	**Artificial Neural Networks:** Neurons and biological motivation, Activation functions and threshold units, Supervised and unsupervised learning, Perceptron Model: representational limitation and gradient descent training, Multilayer networks and back propagation, Overfitting.	8
	Clustering: Learning from unclassified data, Clustering. Hierarchical Agglomerative Clustering, Non- Hierarchical Clustering - k-means partitional clustering, Expectation Maximization (EM) for soft clustering, Semi-supervised learning with EM using labelled and unlabelled data.	6
II	**Kernel Methods:** Dual Representations, Design of Kernels.	4
	Support Vector Machines (SMV): Maximum margin linear separators, Quadratic programming solution to finding maximum margin separators, Kernels for learning non-linear functions, Varying length pattern classification using SVM.	4
III	**Deep Learning:**	8

	Introduction to Deep Learning, Introduction to convolutional Neural Network (CNN), CNN Architecture and layers, Building simple CNN model for classification, Training and Testing the CNN model.	
	Reinforcement Learning: Characteristics, N-arm Bandit Problem, Calculating the Value Function, Associative Learning – Adding States, The Markov Property & Markov Decision Process.	6

References

1. Machine Learning, T. Mitchell, McGraw-Hill, 1997
2. Machine Learning, E. Alpaydin, MIT Press, 2010
3. Pattern Recognition and Machine Learning, C. Bishop, Springer, 2006
4. Pattern Classification, R. Duda, E. Hart, and D. Stork, Wiley-Interscience, 2000
5. Neural Networks - A Class Room Approach, Satish Kumar, Second Edition, Tata McGraw-Hill, 2013
6. The Elements of Statistical Learning: Data Mining, Inference and Prediction, T. Hastie, R. Tibshirani and J. Friedman, Springer, 2nd Edition, 2009
7. Machine Learning for Big Data, Jason Bell, Wiley Big Data Series
8. Kernel Methods for Pattern Analysis, J. Shawe-Taylor and N. Cristianini, Cambridge University Press, 2004
9. Neural Networks and Learning Machines, S. Haykin, Prentice Hall of India, 2010
10. Multidimensional Neural Networks Unified Theory, Rama Murthy G
11. F. Camastra and A. Vinciarelli, Machine Learning for Audio, Image and Video Analysis – Theory and Applications, Springer, 2008.
12. https://www.coursera.org/specializations/deep-learning

BDA 5202: Modern Databases for Big Data	**L**	**T**	**P**	**C**	**Total hours**
	3	**0**	**0**	**3**	**36**

Course Outcome

1. Analyze different types of data and design queries to handle different data types.
2. Evaluate different data models.
3. Analyse different types of noSQL DB
4. Identify appropriate database based on the case study

Unit	Topics	No. of Hours
I	Introduction to growth of traditional and modern database systems, introduction to RDBMS	6
	SQL – Syntax and Semantics	6
	NoSQL Database - Why NoSQL? - Data Models	3
II	Distribution models for scalability: Horizontal partitioning. - Data sharding. -Master-slave replication. Peer-to-peer replication. - Version stamps – business and system transactions.	2
	Consistency Models:Update consistency, Read Consistency, CAP Theorem	2
	MapReduce: Basic MapReduce Partitioning and combining. - Composing MapReduce calculations. - Two-stage map-reduce example. Incremental MapReduce.	2
III	Case study: Key-Value Databases - Document Databases - Column-Family Stores - Graph Databases	10
	Beyond NoSQL: File systems, Event sourcing, Memory Image, Version control, XML Database, Object Database	4
	Choosing your database	1

References

1. Database System Concepts, Avi Silberschatz, Henry F. Korth, and S. Sudarshan. McGraw Hill, 6th Edition, 2010.
2. NoSQL Distilled: A Brief Guide to the Emerging World of Polyglot Persistence, Pramod J. Sadalage, Martin Fowler, Addison-Wesley, 2012.
3. Seven Databases in Seven Weeks: A Guide to Modern Databases and the NoSQL Movement, Eric Redmond, Jim R. Wilson, Pragmatic Bookshelf. 2012.

13. https://www.coursera.org/specializations/cloudera-big-data-analysis-sql

BDA 5203: Multimedia Analytics	L	T	P	C	Total hours
	3	0	0	3	36

Course Outcome

1. Identify fundament features of audio signal.
2. Demonstrate audio processing techniques in time and frequency domain.
3. Identify fundament features for image and video.
4. Demonstrate different image, video representation methods and feature extraction techniques.
5. Demonstrate image and video feature implementation in real world use-cases.

Unit	Topics	No. of Hours
I	Audio Acquisition Representation and Storage: Sound Physics, Production Perception, Audio Encoding and Storage Format, Time Domain Audio processing	6
II	Image and Video Acquisition, Representation and Storage Image Handling and Processing: Reading images from files, Simple Image transformations, Matrices, Colors and Filters, Contours and Segmentation, Object detection and recognition Video Principles, Standards, Video classification models, Motion Detection, Object Tracking in Video	12
III	Case study: Speech and hand writing recognition - Automatic Face recognition, Sign board detection, Lane change detection - Video segmentation and key frame extraction	18

References

1. Machine Learning for Audio, Image and Video Analysis, Francesco Camastra and Alessandro Vinciarelli Springer's Publication, 2nd edition. 2015.
2. Practical Python and OpenCV, An Introductory, Example Driven Guide to Image Processing and Computer Vision, Dr. Adrian Rosebrock, 4th edition, 2019
3. Computer Vision with Python Cookbook: Leverage the power of OpenCV 3 and Python to build computer vision applications, Aleksei Spizhevoi, Aleksandr Rybnikov, Packt Publishing, 1st Edition, 2018.
4. https://moocs.fcu.edu.tw/courses/course-v1:FCUx+MS133+202011_T1/about
5. https://in.coursera.org/learn/digital?action=enroll

AML 5201: Advanced Applications of Probability and Statistics	L	T	P	C	Total hours
	3	0	0	3	36

Course Outcome

1. Apply linear and logistic regression models for practical problems and assess model performance.
2. Interpret the output of principal component analysis (PCA) applied to multivariate data for dimension reduction.
3. Identify multivariate data with mixed data type features and cluster using an appropriate technique.
4. Understand the basics of time series modelling and apply to real-life problems.

Unit	Topics	No. of Hours
I	**Multivariate Distributions:** Mean vector, covariance and correlation – population vs. sample - The multivariate Gaussian – joint-, marginal-, and conditional distributions, Mahalanobis distance and outliers - Properties of the multivariate Gaussian - Parameter estimation: maximum likelihood estimation (MLE) and maximum aposteriori estimation (MAP).	6
II	**Linear and Logistic Regression:** Simple linear regression – regression model, estimating and interpreting coefficients, accuracy of coefficient estimates and model, ANOVA, R2 statistic - Multiple linear regression – estimating coefficients, qualitative predictors, interaction effects, potential problems - Logistic regression – binary and multinomial logistic regression models, estimating and interpreting coefficients, assessing model calibration and discrimination, area under the ROC.	12
III	**Principal Component Analysis; Cluster Analysis:** Geometric intuition of principal components - Maximum variance perspective – algebraic setup, eigenvectors and eigenvalues of sample correlation matrix - Interpretation and application of principal components for dimension reduction. Dissimilarity measures for mixed data types - Partition around medoids (PAM) vs. K-means algorithms - Selecting the number of clusters.	10

IV	**Bootstrapping; Time Series Analysis:** Time series concepts: stationarity, trend, seasonality, autocorrelation - Autoregressive moving average (ARMA) models - Resampling, smoothing, windowing, and rolling average - First and second order differencing - Validating time series predictions.	8

References

1. An Introduction to Statistical Learning with Applications in R, Gareth James, Daniela Witten, Trevor Hastie and Robert Tibshirani, Springer; 1st Edition, 2013, Corr. 7th printing 2017 Edition.
2. Applied Multivariate Statistical Modeling from Swayam (NPTEL, https://onlinecourses.nptel.ac.in/noc23_ma26/preview)
3. An Introduction to Applied Multivariate Analysis with R, Brian Everitt and Torsten Hothorn- Springer Publications,1st Edition, 2011.
4. Machine Learning - A Probabilistic Perspective, Kevin P. Murphy, The MIT Press; 1st Edition, 2012.
5. Mathematics for Machine Learning, Marc Peter Deisenroth, A Aldo Faisal, and Cheng Soon Ong, Cambridge University Press, 2020. – Online resource from Cambridge University Press available at https://mml-book.github.io/book/mml-book.pdf

BDA 5231: Natural Language and Text Processing	**L**	**T**	**P**	**C**	**Total hours**
	3	**0**	**0**	**3**	**36**

Course Outcome

1. Relate the syntax and semantics of text.
2. Use text processing by implementing lexical analysis, word stemming, word stop and term selection.
3. Demonstrate Categorizing, tagging of words, classification and information extraction from text.
4. Examine models for sentiment and semantic analysis from text.

Unit	Topics	No. of Hours
I	**Natural Language Basics**: Natural Language, Linguistics, Language Semantics, Text Corpora.	4
	Accessing, Processing and Understanding Text: Accessing Text Corpora, from the Web and from Disk, Conditional Frequency Distributions, Regular Expressions for Detecting Word Patterns, Tokenization.	6
	Categorizing and Tagging Words: Using a Tagger, Tagged Corpora, Automatic Tagging, N-Gram Tagging, Transformation-Based Tagging.	4
II	**Classification and Information Extraction**: Automated Text Classification, TF-IDF Model, Advanced Word Vectorization Models, Classification Algorithms - Multinomial Naïve Bayes, Support Vector Machines. Text Summarization and Information Extraction - Text Normalization, Feature Extraction, Keyphrase Extraction, Topic Modeling, Automated Document Summarization.	8
	Text Similarity and Clustering Analyzing: Term Similarity, Analyzing Document Similarity, Document Clustering.	6
III	**Semantic and Sentiment Analysis**: Exploring WordNet, Word Sense Disambiguation, Named Entity Recognition, Analyzing Semantic Representations, Sentiment Analysis.	8

References

1. Text Analytics with Python: A Practitioner's Guide to Natural Language Processing, Dipanjan Sarkar; Publisher: Apress, 2nd Edition, 2019.
2. Natural Language Processing with Python: Analyzing Text with the Natural Language Toolkit, by Steven Bird, Ewan Klein, Edward Loper, O'Reilly Media, Inc, 1st edition 2009.
3. Hands-On Natural Language Processing with Python: A practical guide to applying deep learning architectures to your NLP applications, Rajesh Arumugam, Rajalingappaa Shanmugamani, Packt Publishing Limited, 2018.
4. https://nptel.ac.in/courses/106105158

CDC 5001: DevOps for Cloud	L	T	P	C	Total hours
	3	0	0	3	36

Course Outcome

1. **Apply GIT commands to setup a public git repo for an application and explore existing Software Methodologies with DevOps Life Cycle stages.**
2. **Apply continuous integration pipeline using Jenkins master and slave architecture for web applications**
3. **Analyze containerizing of web applications using Docker containers and Setup a Kuberenetes cluster to deploy containerised application**
4. **Evaluate DevOps methodologies for Product development, release and monitor.**

Unit	Topics	No. of Hours
I	**DevOps Introduction:** Understanding Development - Development SDLC : WaterFall & Agile - Understanding Operations - Dev vs Ops - DevOps to the rescue - What is DevOps - DevOps SDLC - Continuous Delivery model - DevOps tools for DevOps SDLC - DevOps Roles & Responsibilities.	2
II	**Linux:** Linux Introduction, Principles & Linux distro – Booting - Command line utilities & Basic commands - Linux File system - Filters & I/O Redirections - Users & Group administration - File permissions & Ownerships - Sudo - Software Management - Useful tools: ssh, telnet, scp, rsync, disk utils, backups etc - Service & Process management - Shell Scripting - Systems and HW stats – Linux Containers (lxc) - Dockers – Kubernetes and Micro services.	12
III	**Networking fundamentals:** Components of computer networks - Classification: LAN, WAN, Peer to Peer network, Server based - Switches - Routers - Network Architecture - Protocols - Port numbers - DNS - DHCP - IP Addresses - Ip Addresses & Subnet Masks - IP Address Ranges - Subnetting - Private Vs Public networks - High Availability - Firewalls & NACL - Web Application Architecture - Infrastructure - Network layout - Services & Components - Architecture from a DevOps perspective.	4
IV	**Automation, Orchestration & Config Management:** Version control system with Git : What is VCS & why it is needed - DevOps use cases - Setup your own repo with git - Manage your code base/source code with GIT & GITHUB.	4

V	**Continuous Integration with Jenkins:** Introduction to continuous integration - Build & Release and relation with DevOps - Understanding development and developers - Why Continuous integration Jenkins introduction and setup - Jenkins projects/jobs - Jenkins plugins | Jenkins administration: Users - Nodes/slaves - Managing plugins - Managing software versions - Introduction - Phases - Java builds - Build and Release job/project setup | Nexus: Intro & Setup - Software versioning & Hosted repository - Integration with Jenkins - Continuous integration job/project setup | Complete Jenkins project: Packaging Artifacts - Static code Analysis - Tomcat setup Staging & productions - Artifacts deployments to webservers from Jenkins - Build Pipeline - Jenkins not just CI tool anymore - More DevOps use cases of Jenkins.	7
VI	**Ansible: Configuration Management & Automation -** What is Ansible & its features - Ansible setup on local & cloud - Understanding Ansible architecture & Execution - Inventory | Ad hoc commands: Automating change Management with Ad Hoc commands - Playbook Introduction - Ansible configuration with ansible.cfg - Ansible documentation - Modules, modules & lots of modules - Writing playbook for webserver & DB server deployments - Tasks - Variables - Templates - Loops - Handlers - Conditions - Register - Debugging - Ansible Roles - Identify server roles - Roles structure - Creating, Managing and executing roles - Ansible Galaxy - Exploring Roles from Galaxy - Download Galaxy roles and integrate with your code - Ansible Advanced Execution - Improving execution time - Limiting and selecting tasks - Troubleshooting and Testing.	7

References

1. Beginning Shell Scripting (Programmer to Programmer), Eric Foster-Johnson, John C. Welch, Micah Anderson, Wrox Publications, 2005.
2. Mastering Unix Shell Scripting: Bash, Bourne, and Korn Shell Scripting for Programmers, System Administrators, and UNIX Gurus, Randal K. Michael, 2nd Edition, Wiley Publications 2nd edition, 2008.
3. UNIX & Shell Programming, Bintu Harwani, Oxford Publications, 2013
4. Jenkins: The Definitive Guide, John Ferguson Smart, O'Reilly Media, Inc., 2011.
5. Jenkins Essentials, Mitesh Soni, Packt Publications, 2015.
6. Continuous Delivery with Docker and Jenkins, Rafal Leszko, Packt Publications, Second Edition, 2019.
7. Implementing DevOps on AWS, Veselin Kantse, Packt Publications, 2017

8. Docker Orchestration, Randall Smith, Packt Publications, 2017.
9. Jenkins Continuous Integration Cookbook, Alan Berg, Packt Publications, 2012.
10. Practical LXC and LXD Linux Containers for Virtualization and Orchestration, Kumaran S., Senthil, Apress Publications, 2017.
11. Containerization with LXC, Konstantin Ivanov,Packt Publications, Packt Publishing; 1st edition, 2017.
12. Docker: Up & Running: Shipping Reliable Containers in Production, Karl Matthias, Sean Kane, O'Reilly Media. O'Reilly Media; 1st edition, 2015
13. https://www.my-mooc.com/en/mooc/intro-to-devops--ud611/

ENP 5230: Entrepreneurship	L	T	P	C	Total hours
	3	0	0	3	36

Course Outcome

1. Explain the importance of entrepreneurship and entrepreneurial development model, social responsibilities of business.
2. Describe Entrepreneurial Traits and Factors affecting Entrepreneurship process.
3. Discuss Business Start-up Process.
4. Summarize a business and marketing plan for entrepreneurs.

Unit	Topics	No. of Hours
I	**Introduction to Entrepreneurship:** Meaning and Definition of Entrepreneurship-Employment vs Entrepreneurship, Theories of Entrepreneurship, approach to entrepreneurship, Entrepreneurs VS Manager	6
II	**Entrepreneurial Traits:** Personality of an entrepreneur, Types of Entrepreneurs	5
III	**Process of Entrepreneurship:** Factors affecting Entrepreneurship process	6

IV	**Business Start-up Process:** Idea Generation, Scanning the Environment, Macro and Micro analysis	7
V	**Business Plan writing:** Points to be considered, Model Business plan	6
VI	**Case studies:** Indian and International Entrepreneurship	6

References

1. NVR Naidu and T. Krishna Rao, "Management and Entrepreneurship", IK International Publishing House Pvt. Ltd 2008.
2. Mohanthy Sangram Keshari, "Fundamentals of Entrepreneurship", PHI Publications, 2005
3. https://nptel.ac.in/courses/110106141

BDA 5251: Machine Learning for Big Data Lab	**L**	**T**	**P**	**C**	**Total hours**
	0	**0**	**3**	**1**	**36**

Course Outcome

1. Demonstrate Artificial Neural Network, Clustering, Support Vector Machine, Deep Neural Network and Reinforcement Learning models.
2. Experiment the performance of single layer, multilayer, and deep neural networks.
3. Compare the performance of different clustering algorithms.
4. Evaluate the performance of different types of artificial neural network models, clustering models, deep neural network models, and reinforcement learning models.

Unit	**Topics**	**No. of Hours**
I	Building Artificial Neural Network and clustering model using Python.	12
II	Designing single layer and multilayer deep neural network using Python.	12
III	Constructing Support Vector Machine and building simple CNN model for classification using Python.	12

References

1. Machine Learning, T. Mitchell, McGraw-Hill, 1997
2. Machine Learning, E. Alpaydin, MIT Press, 2010
3. Pattern Recognition and Machine Learning, C. Bishop, Springer, 2006
4. Pattern Classification, R. Duda, E. Hart, and D. Stork, Wiley-Interscience, 2000
5. Neural Networks - A Class Room Approach, Satish Kumar, Second Edition, Tata McGraw-Hill, 2013
6. The Elements of Statistical Learning: Data Mining, Inference and Prediction, T. Hastie, R. Tibshirani and J. Friedman, Springer, 2nd Edition, 2009
7. Machine Learning for Big Data, Jason Bell, Wiley Big Data Series
8. Kernel Methods for Pattern Analysis, J. Shawe-Taylor and N. Cristianini, Cambridge University Press, 2004
9. Neural Networks and Learning Machines, S. Haykin, Prentice Hall of India, 2010
10. Multidimensional Neural Networks Unified Theory, Rama Murthy G
11. F. Camastra and A. Vinciarelli, Machine Learning for Audio, Image and Video Analysis – Theory and Applications, Springer, 2008.

BDA 5252: Modern Databases for Big Data Lab	L	T	P	C	Total hours
	0	0	3	1	36

Course Outcome

1. Illustrate master/slave replication by setting up the database table.
2. Illustrate CAP theory concepts.
3. Apply noSQL DB concepts on case studies

Unit	Topics	No. of Hours
I	Setup the database table to illustrate master/slave replication.	12
II	Setup the data operation models to illustrate CAP theory concepts.	12
III	Setting and operating with different types of no SQL database (Mongo DB, CouchDB)	12

References

1. Machine Learning for Audio, Image and Video Analysis, Francesco Camastra and Alessandro Vinciarelli Springer's Publication, 2nd edition. 2015.
2. Practical Python and OpenCV, An Introductory, Example Driven Guide to Image Processing and Computer Vision, Dr. Adrian Rosebrock, 4th edition, 2019
3. Computer Vision with Python Cookbook: Leverage the power of OpenCV 3 and Python to build computer vision applications, Aleksei Spizhevoi, Aleksandr Rybnikov, Packt Publishing, 1st Edition, 2018.

BDA 5253: Multimedia Analytics Lab	L	T	P	C	Total hours
	0	0	3	1	36

Course Outcome

1. Examine fundament features of audio signal.
2. Demonstrate audio processing techniques in time and frequency domain.
3. Examine fundament features for image and video.
4. Demonstrate different image, video representation methods and feature extraction techniques.
5. Demonstrate image and video feature implementation in real world use-cases.

Unit	Topics	No. of Hours
I	Time and Frequency domain audio processing in Python to extract and analyse the audio features.	12
II	Experiment various image processing techniques.	12
III	Experiment various Video processing techniques.	12

References

1. Machine Learning for Audio, Image and Video Analysis, Francesco Camastra and Alessandro Vinciarelli Springer's Publication, 2nd edition. 2015.
2. Practical Python and OpenCV, An Introductory, Example Driven Guide to Image Processing and Computer Vision, Dr. Adrian Rosebrock, 4th edition, 2019.
3. Computer Vision with Python Cookbook: Leverage the power of OpenCV 3 and Python to build computer vision applications, Aleksei Spizhevoi, Aleksandr Rybnikov, Packt Publishing, 1st Edition, 2018.

AML 5251: Advanced Applications of Probability and Statistics Lab	L	T	P	C	Total hours
	0	0	3	1	36

Course Outcome

1. Build and assess linear and logistic regression models for practical problems.
2. Perform principal component analysis (PCA) for dimension reduction in multivariate data.
3. Cluster multivariate data with mixed data types.
4. Apply time series modelling to real-life problems.

Unit	Topics	No. of Hours
I	**Multivariate Distributions:** Compute descriptive statistics of multivariate data; Perform exploratory data analysis of multivariate data; Identify outliers in multivariate data; Visualise and understand the properties of multivariate Gaussian data.	6
II	**Linear and Logistic Regression:** Use in-built functions in R to build linear models for practical problem; Compute different performance metrics to assess model performance; Interpret model coefficients and investigate the effect of input variables on output through sensitivity analysis; Use in-built functions in R to build logistic regression models for practical binary classification problems and assess model performance.	12
III	**Principal Component Analysis; Cluster Analysis:** Visualise the geometric interpretation of principal component analysis (PCA); Use in-built functions in R to perform PCA on multivariate data; Compare and contrast PCA for variance maximization vs. clustering of multivariate data; Cluster multivariate data with mixed data types using in-built functions in R.	10
IV	**Bootstrapping; Time Series Analysis:** Apply bootstrapping on a practical data set and assess performance; Understand and apply in-built functions in R for time series modelling; Apply time series modelling to practical problems; Interpret the results of times series model predictions.	8

References

1. An Introduction to Statistical Learning with Applications in R, Gareth James, Daniela Witten, Trevor Hastie and Robert Tibshirani, Springer; 1st Edition, 2013, Corr. 7th printing 2017 Edition.

2. Applied Multivariate Statistical Modeling from Swayam (NPTEL, https://onlinecourses.nptel.ac.in/noc23_ma26/preview)
3. An Introduction to Applied Multivariate Analysis with R, Brian Everitt and Torsten Hothorn– Springer Publications,1st Edition, 2011.
4. Machine Learning - A Probabilistic Perspective, Kevin P. Murphy, The MIT Press; 1st Edition, 2012.
5. Mathematics for Machine Learning, Marc Peter Deisenroth, A Aldo Faisal, and Cheng Soon Ong, Cambridge University Press, 2020. – Online resource from Cambridge University Press available at https://mml-book.github.io/book/mml-book.pdf

BDA 5281: Natural Language and Text Processing Lab	L	T	P	C	Total hours
	0	0	3	1	36

Course Outcome

1. Experiment on understanding and processing of text from corpora functionalities.
2. Implement classification and information extraction methods.
3. Examine text similarity and sentiment analysis from text corpora.

Unit	Topics	No. of Hours
I	Experimenting on Language syntax and Semantics	9
	Experimenting Accessing Text Corpora, from the Web and from Disk, and perform relevant tasks like Conditional Frequency Distributions, Regular Expressions for Detecting Word Patterns, Tokenization.	9
	Carry out experiments to categorise and tagging words	6
II	Carry out experiments to perform classification and information extraction method	6
	Carry out experiments to perform text similarity	3
III	Carry out experiments to perform sentiment analysis	3

References
1. Text Analytics with Python: A Practitioner's Guide to Natural Language Processing, Dipanjan Sarkar; Publisher: Apress, 2nd Edition, 2019. 2. Natural Language Processing with Python: Analyzing Text with the Natural Language Toolkit, by Steven Bird, Ewan Klein, Edward Loper, O'Reilly Media, Inc, 1st edition 2009. 3. Hands-On Natural Language Processing with Python: A practical guide to applying deep learning architectures to your NLP applications, Rajesh Arumugam, Rajalingappaa Shanmugamani, Packt Publishing Limited, 2018.

CDC 5051: DevOps for Cloud Lab	L	T	P	C	Total hours
	0	0	3	1	36

Course Outcome

1. Apply GIT commands to implement version control on any given application.
2. Apply continuous integration pipeline using Jenkins master and slave architecture for web applications
3. Analyze containerizing of a web applications using Docker containers and Setup a Kuberenetes cluster to deploy containerized application
4. Evaluate DevOps methodologies for Product development, release and monitor.

Unit	Topics	No. of Hours
I	Understand the lifecycle stages of DevOps product life cycle methodology.	3
II	Explore the basic Linux commands, system utilities and services.	6
III	Understand the basic architecture of software version controls like GIT, SVN. Setup a public git repo for an application explore the GIT commands to perform clone, fork, pull, push, commit etc.	3
IV	Demonstrate a continuous integration pipeline using Jenkins master and slave architecture for web applications	6

V	Explore containerizing web applications using Docker containers	6
VI	Setup a Kuberenetes cluster to deploy containerised application with high availability and load balancing.	6
VII	Create Ansible playbook for configuring Linux and window host.	3
VIII	Setup a continuous monitoring service for managing IT assets.	3

References

1. Beginning Shell Scripting (Programmer to Programmer), Eric Foster-Johnson, John C. Welch, Micah Anderson, Wrox Publications, 2005.
2. Mastering Unix Shell Scripting: Bash, Bourne, and Korn Shell Scripting for Programmers, System Administrators, and UNIX Gurus, Randal K. Michael, 2nd Edition, Wiley Publications 2nd edition, 2008.
3. UNIX & Shell Programming, Bintu Harwani, Oxford Publications, 2013
4. Jenkins: The Definitive Guide, John Ferguson Smart, O'Reilly Media, Inc., 2011.
5. Jenkins Essentials, Mitesh Soni, Packt Publications, 2015.
6. Continuous Delivery with Docker and Jenkins, Rafal Leszko, Packt Publications, Second Edition, 2019.
7. Implementing DevOps on AWS, Veselin Kantse, Packt Publications, 2017
8. Docker Orchestration, Randall Smith, Packt Publications, 2017.
9. Jenkins Continuous Integration Cookbook, Alan Berg, Packt Publications, 2012.
10. Practical LXC and LXD Linux Containers for Virtualization and Orchestration, Kumaran S., Senthil, Apress Publications, 2017.
11. Containerization with LXC, Konstantin Ivanov,Packt Publications, Packt Publishing; 1st edition, 2017.
12. Docker: Up & Running:Shipping Reliable Containers in Production, Karl Matthias, Sean Kane, O'Reilly Media. O'Reilly Media; 1st edition, 2015.

ENP 5280: Entrepreneurship Lab	L	T	P	C	Total hours
	0	0	3	1	36

Course Outcome

1. Study of prominence of entrepreneurship.
2. Develop use cases for building a business.

3. Evaluation of factors influencing business venture.

Unit	Topics	No. of Hours
I	Study of use cases for need and prominence of entrepreneurship, associated decision making process.	6
II	Study of report by the National Knowledge Commission on the importance of entrepreneurship in economic development.	6
III	Develop use cases for identifying and evaluating opportunities, developing business plan, assessment of resources, project appraisal and feasibility plan.	9
IV	creating and starting venture includes legal requirements, marketing strategies, financial plans and human resources management	9
V	Design a Case studies of Indian and International Entrepreneurship.	6

References

1. Management and Entrepreneurship, NVR Naidu and T. Krishna Rao, IK International Publishing House Pvt.Ltd, 2008.
2. Fundamentals of Entrepreneurship, Mohanthy Sangram Keshari, PHI Learning Pvt. Ltd., 2005.

MPT 5200: Mini Project - II	L	T	P	C	Total hours
	0	0	0	4	48

Course Outcome

1. Identify the real-world and socially relevant problems and perform feasibility analysis for finding solutions.
2. Organize work effectively as a member in a team, examine, experiment, and communicate technical information constructively.
3. Develop and implement solutions to the identified problems by applying research methodology and development life cycle with appropriate documentation by incorporating ethical standards.

Unit	Topics	No. of Hours
I	Problem identification, literature survey, formation of detailed specifications.	48
II	Design and implementation of the proposed system architecture.	
III	Demonstrate an ability to present and defend project work carried out to a panel of experts.	
References		
1. Research articles and Online Resources.		

PSD 5200: Professional Skill Development - II	L	T	P	C	Total hours
	0	0	0	1	12

Course Outcome

1. Develop the skills needed for approaching technical and HR interviews.
2. Use mathematical, reasoning, and domain specific skills to solve objective questionnaires in time.
3. Demonstrate depth of knowledge in the chosen field of study.

Unit	Topics	No. of Hours
I	Peer interviews, mock interviews.	12
II	Logical reasoning, mathematical aptitude, domain specific problem solving skills.	
III	Conduction of domain specific knowledge test.	

References

1. R S Aggarwal. Quantitative Aptitude for Competitive Examinations. S Chand, 2017.
2. McDowell, Gayle Laakmann. Cracking the coding interview: 189 programming questions and solutions. CareerCup, LLC, 2015.

3. Domain specific tools and online resources.

BDA 6098: Project Work	L	T	P	C	Total hours
	0	0	0	25	300

Course Outcome

1. Undertake innovative industry/research oriented projects and perform feasibility analysis for finding solutions.
2. Implement and test the proposed design using appropriate framework, programming language and tools.
3. Demonstrate an ability to present and defend project work carried out to a panel of experts.

Unit	Topics	No. of Hours
I	Problem identification, literature survey, formation of detailed requirement specification document.	300
II	Design and implementation of the proposed modules with specific test cases.	
III	Detailed report of the work carried out, present, and defend the project work carried out to a panel of experts.	

References

1. Research articles, domain specific tools and online resources.

Program Outcome and Course Outcome Mapping

Sl. No.	Course Code	Course Name	Credits	PO1	PO2	PO3	PO4	PO5
1	BDA 5101	Algorithms and Data Structures for Big Data	3			*	*	
2	BDA 5102	Architecture of Big Data Systems	3			*	*	*
3	BDA 5103	Fundamentals of Machine Learning	3			*	*	*
4	AML 5103	Applied Probability and Statistics	3			*		*
5	BDA 5132	Principles of Data Visualization	3			*		*
	ESD 5234	Mobile Web Application Development	3			*		*
6	BDA 5151	Algorithms and Data Structures for Big Data Lab	1			*	*	
7	BDA 5152	Architecture of Big Data Systems Lab	1			*	*	*
8	BDA 5153	Fundamentals of Machine Learning Lab	1			*	*	*
9	AML 5153	Applied Probability and Statistics Lab	1			*		*
10	BDA 5182	Principles of Data Visualization Lab	1			*		*
	ESD 5284	Mobile Web Application Development Lab	1			*		*
11	MPT 5100	Mini Project - I	4	*	*	*	*	*
12	PSD 5100	Professional Skill Development - I	1	*	*			
13	BDA 5201	Machine Learning for Big Data	3			*	*	*
14	BDA 5202	Modern Databases for Big Data	3			*	*	*
15	BDA 5203	Multimedia Analytics	3			*	*	*
16	AML 5201	Advanced Applications of Probability and Statistics	3			*		*
17	BDA 5231	Natural Language and Text Processing	3			*	*	*
	CDC 5001	DevOps for Cloud	3	*		*	*	*
	ENP 5230	Entrepreneurship	3			*	*	
18	BDA 5251	Machine Learning for Big Data Lab	1			*	*	*
19	BDA 5252	Modern Databases for Big Data Lab	1			*	*	*
20	BDA 5253	Multimedia Analytics Lab	1			*	*	*

21	AML 5251	Advanced Applications of Probability and Statistics Lab	1			*		*
22	BDA 5281	Natural Language and Text Processing Lab	1			*	*	*
	CDC 5051	DevOps for Cloud Lab	1	*		*	*	*
	ENP 5280	Entrepreneurship Lab	1			*	*	
23	MPT 5200	Mini Project - II	4	*	*	*	*	*
24	PSD 5200	Professional Skill Development - II	1			*	*	*
25	BDA 6098	Project Work	25	*	*	*	*	*