Dr. Babasaheb Ambedkar Technological University
(Established as a University of Technology in the State of Maharashtra)
(Under Maharashtra Act No. XXIX of 2014)
P.O. Lonere, Dist. Raigad, Pin 402 103, Maharashtra
Telephone and Fax. 02140 - 275142
www.dbatu.ac.in

Structure and Detailed Syllabus

for UG Degree

Minor in Data Science

in line with New Education Policy 2020

(Effective from Academic year 2024-25 for main campus)

Bucket for Minor in Data Science

Case I: B. Tech degree with Minor in Data Science (160-176 credits)

The Bachelor's Engineering Degree in chosen Engg./ Tech. Discipline with multidisciplinary minor (min.160-max.176 Credits) i.e. "**B. Tech in chosen Engg./ Tech. Discipline with Minor in Data Science**" (160-176 credits) enables students to take up four-six or required additional courses of 14 credits in the discipline other than **chosen Engg./ Tech. Discipline** distributed over semesters III to VIII.

Case II: Bachelor's Engineering Degree in chosen Engg./ Tech. Discipline with Double Minor (Multidisciplinary and Specialization Minor 180-194 credits)

The **Bachelor's Engineering Degree in chosen Engg./ Tech. Discipline with Double Minor** (Multidisciplinary and Specialization Minor, 180-194 credits), i.e. "**B. Tech in chosen Engg./ Tech. Discipline with minor in *other selected discipline in* Engineering (as MDM) with Specialization Minor in Data Science**" (180-194 credits) enables students to take up four-six additional courses of 14 credits in the discipline other than **chosen Engg./ Tech. Discipline** (for completion of multidisciplinary minor) and 18 to 20 extra credits in the **Data Science** distributed over semesters III to VIII. Here, the ***other selected discipline* in Engineering should be different from Specialization Minor i.e. Data Science**. This enables students to take up four-six or required additional courses of 18 to 20 credits in the discipline of **Data Science** distributed over semesters III to VIII, which are over and above the min.160-max.176 Credits. The decision regarding the mechanism of distribution of these 18-20 credits over semesters III to VIII, prescribed for the duration of four years will be taken by respective BoS. **Student must have CGPA equal to or greater than 7.5 at the end of second semester to go for this option**.

Basic Semester wise credit distribution of the syllabus is as follows as per NEP-2020.

Semester		I	II	III	IV	V	VI	VII	VIII	Total Credits
Basic Science Course	BSC/ESC	06-08	08-10		--	--	--	--	--	14-18
Engineering Science Course		10-08	06-04		--	--	--	--	--	16-12
Programme Core Course (PCC)	Program Courses	--	02	08-10	08-10	10-12	08-10	04-06	04-06	44-56
Programme Elective Course (PEC)		--	--	--	--	04	08	02	06	20
Multidisciplinary Minor (MD M)	Multidisciplinary Courses		-	02	02	04	02	02	02	14
Open Elective (OE) Other than a particular program		--	--	04	02	02	--	--	--	08
Vocational and Skill Enhancement Course (VSEC)	Skill Courses	02	02	--	02	--	02	--	--	08
Ability Enhancement Course (AEC -01, AEC-02)	Humanities Social Science and Management (HSSM)	02	--	--	02	--	--	--	--	04
Entrepreneurship/Economics/ Management Courses		--		02	02	--	--	--	--	04
Indian Knowledge System (IKS)			02		--	--	--	--	--	02
Value Education Course (VEC)		--	--	02	02	--	--	--	--	04
Research Methodology	Experiential Learning Courses	--	--	--	--	--	--		04	04
Comm. Engg. Project (CEP)/Field Project (FP)		--	--	02	--	--	--	-	-	02
Project		--	--	--	--	--	--		04	04
Internship/ OJT		--	---			--	--	12	-	12
Co-curricular Courses (CC)	Liberal Learning Courses	02	02		--	--	--	--	-	04
Total Credits (Major)		**20-22**	**20-22**	**20-22**	**20-22**	**20-22**	**20-22**	**20-22**	**20-22**	**160-176**

List of Courses for

Minor in Data Science

Sr. No.	Course Name	Teaching Scheme	Duration (Weeks)	Credits	Institute Offering Course	Name of Professor/ Resource person	Link
1	Data Science for Engineers	4 Hrs/Week	8	2	IIT Madras	Prof. Shankar Narasimhan, Prof. Ragunathan Rengasamy	https://nptel.ac.in/courses/106106179
2	Introduction to Machine Learning	4 Hrs/Week	8	2	IIT Kharagpur	Prof. S. Sarkar	https://nptel.ac.in/courses/106105152
3	Deep Learning	4 Hrs/Week	12	3	IIT Kharagpur	Prof. P.K. Biswas	https://nptel.ac.in/courses/106105215
4	Reinforcement Learning	4 Hrs/Week	12	3	IIT Madras	Prof. Balaraman Ravindran	https://onlinecourses.nptel.ac.in/noc24_cs52/preview
5	Natural Language Processing	4 Hrs/Week	12	3	IIT Kharagpur	Prof. Pawan Goyal	https://onlinecourses.nptel.ac.in/noc24_cs39/preview
6	Data Mining	4 Hrs/Week	8	2	IIT Kharagpur	Prof. Pabitra Mitra	https://nptel.ac.in/courses/106105174
7	Python for Data Science	4 Hrs/Week	4	1	IIT Madras	Prof. Ragunathan Rengasamy	https://nptel.ac.in/courses/106106212
8	Data Analytics with Python	4 Hrs/Week	12	3	IIT Roorkee	Prof. A. Ramesh	https://nptel.ac.in/courses/106107220
9	Programming and Data Structure	4 Hrs/Week	8	2	IIT Kharagpur	Dr. P.P.Chakraborty	https://nptel.ac.in/courses/106105085
10	Big Data Computing	4 Hrs/Week	8	2	IIT Patna	Dr. Rajiv Misra	https://nptel.ac.in/courses/106104189
11	Data Structures And Algorithms	4 Hrs/Week	8	2	IIT Delhi	Prof. Naveen Garg	https://nptel.ac.in/courses/106102064
12	Statistical Methods for Scientists and Engineers	4 Hrs/Week	4	1	IIT Kharagpur	Prof. Somesh Kumar	https://nptel.ac.in/courses/111105077

Data Science for Engineers

Week 1: Course philosophy and introduction to R

Week 2: Linear algebra for data science

1. Algebraic view - vectors, matrices, product of matrix & vector, rank, null space, solution of over determined set of equations and pseudo-inverse)

2. Geometric view - vectors, distance, projections, eigenvalue decomposition

Week 3: Statistics (descriptive statistics, notion of probability, distributions, mean, variance, covariance, covariance matrix, understanding univariate and multivariate normal distributions, introduction to hypothesis testing, confidence interval for estimates)

Week 4: Optimization

Week 5: 1. Optimization

2. Typology of data science problems and a solution framework

Week 6: 1. Simple linear regression and verifying assumptions used in linear regression

2. Multivariate linear regression, model assessment, assessing importance of different variables, subset selection

Week 7: Classification using logistic regression

Week 8: Classification using kNN and k-means clustering

Introduction to Machine Learning

Week 01:	Introduction: Basic definitions, types of learning, hypothesis space and inductive bias, evaluation, cross-validation.
Week 02:	Linear regression, Decision trees, over fitting.
Week 03:	Instance based learning, Feature reduction, Collaborative filtering based recommendation.
Week 04:	Probability and Bayes learning
Week 05:	Logistic Regression, Support Vector Machine, Kernel function and Kernel SVM.
Week 06:	Neural network: Perceptron, multilayer network, back propagation, introduction to deep neural network.
Week 07:	Computational learning theory, PAC learning model, Sample complexity, VC Dimension, Ensemble learning.
Week 08:	Clustering: k-means, adaptive hierarchical clustering, Gaussian mixture model.

Deep Learning

Week 1: Introduction to Deep Learning, Bayesian Learning, Decision Surfaces

Week 2: Linear Classifiers, Linear Machines with Hinge Loss

Week 3: Optimization Techniques, Gradient Descent, Batch Optimization

Week 4: Introduction to Neural Network, Multilayer Perceptron, Back Propagation Learning

Week 5: Unsupervised Learning with Deep Network, Autoencoders

Week 6: Convolutional Neural Network, Building blocks of CNN, Transfer Learning

Week 7: Revisiting Gradient Descent, Momentum Optimizer, RMSProp, Adam

Week 8: Effective training in Deep Net- early stopping, Dropout, Batch Normalization, Instance Normalization, Group Normalization

Week 9: Recent Trends in Deep Learning Architectures, Residual Network, Skip Connection Network, Fully Connected CNN etc.

Week 10: Classical Supervised Tasks with Deep Learning, Image Denoising, Semanticd Segmentation, Object Detection etc.

Week 11: LSTM Networks

Week 12: Generative Modeling with DL, Variational Autoencoder, Generative Adversarial Network Revisiting Gradient Descent, Momentum Optimizer, RMSProp, Adam

Reinforcement Learning

Week 1 : Introduction

Week 2 : Bandit algorithms – UCB, PAC

Week 3 : Bandit algorithms –Median Elimination, Policy Gradient

Week 4 : Full RL & MDPs

Week 5 : Bellman Optimality

Week 6 : Dynamic Programming & TD Methods

Week 7 : Eligibility Traces

Week 8 : Function Approximation

Week 9 : Least Squares Methods

Week 10 : Fitted Q, DQN & Policy Gradient for Full RL

Week 11 : Hierarchical RL

Week 12 : POMDPs

Natural Language Processing

Week 1: Introduction and Basic Text Processing

Week 2: Spelling Correction, Language Modeling

Week 3: Advanced smoothing for language modeling, POS tagging

Week 4: Models for Sequential tagging – MaxEnt, CRF

Week 5: Syntax – Constituency Parsing

Week 6: Dependency Parsing

Week 7: Distributional Semantics

Week 8: Lexical Semantics

Week 9: Topic Models

Week 10: Entity Linking, Information Extraction

Week 11: Text Summarization, Text Classification

Week 12: Sentiment Analysis and Opinion Mining

Data Mining

Week 1: Introduction, Data Preprocessing

Week 2: Association Rule Mining, Classification Basics

Week 3: Decision Tree, Bayes Classifier, K nearest neighbor

Week 4: Support Vector Machine, Kernel Machine

Week 5: Clustering, Outlier detection

Week 6: Sequence mining

Week 7: Evaluation, Visualization

Week 8: Case studies

Python for Data Science

Week 1:

BASICS OF PYTHON SPYDER (TOOL)

- Introduction Spyder
- Setting working Directory
- Creating and saving a script file
- File execution, clearing console, removing variables from environment, clearing environment
- Commenting script files
- Variable creation
- Arithmetic and logical operators
- Data types and associated operations

Week 2:

Sequence data types and associated operations

- Strings
- Lists
- Arrays
- Tuples
- Dictionary
- Sets
- Range
- NumPy
- ndArray

Week 3:

Pandas dataframe and dataframe related operations on Toyota Corolla dataset

- Reading files
- Exploratory data analysis
- Data preparation and preprocessing
- Data visualization on Toyoto Corolla dataset using matplotlib and seaborn libraries
- Scatter plot
- Line plot
- Bar plot
- Histogram
- Box plot
- Pair plot

Control structures using Toyota Corolla dataset

- if-else family
- for loop
- for loop with if break
- while loop
- Functions

Week 4: CASE STUDY

- Regression: Predicting price of pre-owned cars
- Classification: Classifying personal income

Data Analytics with Python

Week 1: Introduction to data analytics and Python fundamentals

Week 2: Introduction to probability

Week 3: Sampling and sampling distributions

Week 4: Hypothesis testing

Week 5: Two sample testing and introduction to ANOVA

Week 6: Two way ANOVA and linear regression

Week 7: Linear regression and multiple regression

Week 8: Concepts of MLE and Logistic regression

Week 9: ROC and Regression Analysis Model Building

Week 10: C2Test and introduction to cluster analysis

Week 11: Clustering analysis

Week 12: Classification and Regression Trees (CART)

Programming and Data Structure

1. Introduction
2. C Programming - I
3. C Programming - II
4. C Programming - III
5. Data Structuring: Case Study - I
6. Data Structuring: Case Study - II
7. Data Structuring: Case Study - III
8. Problem Decomposition By Recursion - I
9. Problem Decomposition By Recursion - II
10. Problem Decomposition By Recursion - III
11. Merge sort And Quick sort
12. Characters And Strings
13. Arrays: Addresses And Contents
14. Structures - I
15. Structures - II
16. Dynamic Allocation Part - I
17. Linked Lists - I
18. Complexity (Efficiency) of Algorithms
19. Asymptotic Growth Functions
20. Asymptotic Analysis of Algorithms
21. Data Structuring
22. Search Trees
23. Search Trees - II
24. Search Trees - III
25. 2-3 Trees
26. Algorithm Design - I
27. Algorithm Design - II
28. Algorithm Design - III
29. Graphs - I
30. Graphs - II
31. Graphs - III
32. Conclusion

Big Data Computing

Week 1: Introduction to Big Data

Week 2: Introduction to Enabling Technologies for Big Data

Week 3: Introduction to Big Data Platforms

Week 4: Introduction to Big Data Storage Platforms for Large Scale Data Storage

Week 5: Introduction to Big Data Streaming Platforms for Fast Data

Week 6: Introduction to Big Data Applications (Machine Learning)

Week 7: Introduction of Big data Machine learning with Spark

Week 8: Introduction to Big Data Applications (Graph Processing)

Data Structures and Algorithms

- Introduction to object oriented programming through stacks, queues and linked lists
- Dictionaries: skip-lists, hashing, analysis of collision resolution techniques
- Trees, traversals, binary search trees, optimal and average BST’s trees and red-black trees
- Tries and pattern matching. Priority queues and binary heaps
- Sorting: merge, quick, radix, selection, heap
- Introduction to Graphs, Breadth first search and connected components
- Depth first search in directed and undirected graphs and strongly connected components
- Spanning trees: Prim's and Kruskal's algorithm, union-find datastructure.
- Dijkstra's algorithm for shortest path. shortest path tree.
- Shortest and longest paths in directed acyclic graphs

Statistical Methods for Scientists and Engineers

- Review of Probability and Distributions
- Parametric Methods
- Multivariate Analysis
- Nonparametric Methods