Pimpri Chinchwad Education Trust's

PIMPRI CHINCHWAD COLLEGE OF ENGINEERING

SECTOR NO. 26, PRADHIKARAN, NIGDI, PUNE 411044

An Autonomous Institute Approved by AICTE and Affiliated to SPPU, Pune

DEPARTMENT OF COMPUTER ENGINEERING

Curriculum Structure and Syllabus

of

Minors in Data Science

(Regulations 2020)

Effective from Academic Year 2024-25

Institute Vision

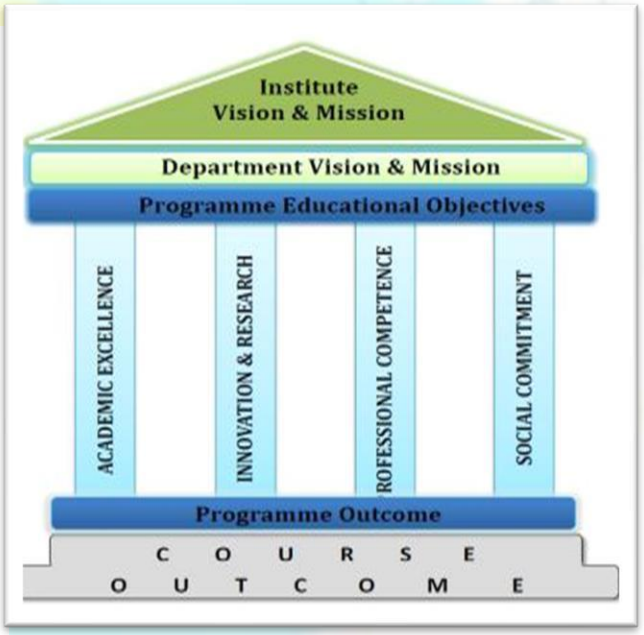
To be one of the top 100 Engineering Institutes of India in coming five years by offering exemplarily Ethical, Sustainable and Value-Added Quality Education through a matching ecosystem for building successful careers.

Institute Mission

1. Serving the needs of the society at large through establishment of a state-of-art Engineering Institute
2. Imparting right Attitude, Skills, Knowledge for self-sustenance through Quality Education
3. Creating globally competent and Sensible engineers, researchers and entrepreneurs with an ability to think and act independently in demanding situations.

Quality Policy

We at PCCOE are committed to impart Value Added Quality Education to satisfy the applicable requirements, needs and expectations of the Students and Stakeholders. We shall strive for academic excellence, professional competence and social commitment in fine blend with innovation and research. We shall achieve this by establishing and strengthening state-of- the-art Engineering and Management Institute through continual improvement in effective implementation of Quality Management System.



	Pimpri Chinchwad Education Trust's **Pimpri Chinchwad College of Engineering**	
Course Approval Summary		

A) Board of study - Department of Computer Engineering

Sr. No.	Name of the Course	Course Code	Page number	Signature and stamp of BoS
1	Probability and Statistics for Data Science	MCE5991	6	
2	Programming with Python Laboratory	MCE5992	8	
3	Basics of Data Science	MCE6991	11	
4	Data Science Laboratory	MCE6992	13	
5	Data Analytics & Visualization	MCE7991 / MCE8991	16	
6	Data Visualization Laboratory	MCE7992 / MCE8992	18	
7	Minor Project	MCE7993 / MCE8993	20	

Approved by Academic Council:

Chairman, Academic Council
Pimpri Chinchwad College of Engineering

Preface

Looking at Global Scenario to enhance the employability skills and impart deep knowledge in emerging/ multidisciplinary areas, an additional avenue is provided to passionate learners through the Minors and Honors Degree Scheme in academic structure.

For **Minors degree** program, student has to earn additional 20 credits in multidisciplinary areas of other domains.

Objectives of Minors Degree

- To impart knowledge in multidisciplinary areas.
- To provide effective yet flexible options for students to achieve basic to intermediate level competence in the multidisciplinary area.
- To enhance the employability skills through different combinations in the diverse fields of engineering.
- To provide an academic mechanism for fulfilling multidisciplinary demands of industries.
- To provide a strong foundation to students aiming to pursue research/ higher studies in an interdisciplinary field of study.

Preface of Minors in Data Science

Data science is an interdisciplinary field deals with a very large amount of data that uses scientific methods, processes, algorithms, modern tools and techniques to find various unseen patterns, derive meaningful information, and make business decisions using knowledge of machine learning, deep learning algorithms, artificial intelligence to build predictive models. The data used for analysis can be noisy, structured and unstructured come from many different sources and different formats.

Students in the Data Science minor gain literacy and fluency in statistics, data science methods machine learning algorithms, data visualization techniques and understand their implications for society and the world.
Earning a minor in data science will prepare them for careers as a Data Scientist, Business analyst, data engineer, risk analyst and many more or postgraduate work using data skills.

Objectives:

This program aims to

- Analyze and apply various data preprocessing techniques, statistical methods.
- Derive meaningful information from any raw data.
- Impart knowledge of Machine learning, deep learning, Artificial intelligence, data visualization to prepare data scientist, business analyst etc.
- Expose students to real world problems through experiential learning.

Learning Outcomes:

At the successful completion of this Minor program, students will able to

- Analyze the data by applying statistical methods.
- Hypothesize various advanced statistical techniques for modelling and exploring practical situations.
- Apply Exploratory Data Analysis on various data sets to solve real life applications.
- Apply machine learning algorithms and Implement data science concepts using various tools and techniques for real world applications.
- Apply Data Analytics Life Cycle to business problem.
- Visualize Big data using appropriate tools.

INDEX

Sr. No.	Content	Page No.
1	List of Abbreviations in Curriculum Structure	1
2	Curriculum Structure	2
3	Course Syllabus of Semester – V Courses	5
4	Course Syllabus of Semester – VI Courses	10
5	Course Syllabus of Semester – VII/VIII Courses	15
6	Vision and Mission of Computer Engineering Department	22

LIST OF ABBREVIATIONS IN CURRICULUM STRUCTURE

Sr. No.	Abbreviation	Expansion
1.	L	Lecture
2.	P	Practical
3.	T	Tutorial
4.	H	Hours
5.	CR	Credits
6.	FA	Formative Assessment
8.	SA	Summative Assessment
9.	TW	Term Work
10.	OR	Oral
11.	PR	Practical
12.	PROJ	Project

Curriculum Structure

Minors in Data Science

CURRICULUM STRUCTURE

Structure for Minors in Data Science (Computer Engineering) for Scheme A and C

Semester	Course code	Course Name	Teaching Scheme				CR	Evaluation Scheme						
			L	P	T	H		FA1	FA2	SA	TW	PR	OR	Total
V	MCE5991	Probability and Statistics for Data Science	3	-	-	3	3	20	20	60	-	-	-	100
V	MCE5992	Programming with Python Laboratory	-	4	-	4	2	-	-	-	50	-	-	50
VI	MCE6991	Basics of Data Science	3	-	-	3	3	20	20	60	-	-	-	100
VI	MCE6992	Data Science Laboratory	-	4	-	4	2	-	-	-	50	-	-	50
VII	MCE7991	Data Analytics & Visualization	3	-	-	3	3	20	20	60	-	-	-	100
VII	MCE7992	Data Visualization Laboratory	-	4	-	4	2	-	-	-	50	-	-	50
VIII	MCE8993	Minor Project	-	10	-	10	5	-	-	-	100	-	50	150
Total			**9**	**22**	**-**	**31**	**20**	**60**	**60**	**180**	**250**	**-**	**50**	**600**

L-Lecture, **P**-Practical, **T**-Tutorial, **H**-Hours, **Cr**-Credits, **FA**-Formative Assessment, **SA**-Summative Assessment, **TW**-Term Work, **OR**-Oral, **PR**-Practical

Structure for Minors in Data Science (Computer Engineering) for Scheme B

Semester	Course code	Course Name	Teaching Scheme				CR	Evaluation Scheme						
			L	P	T	H		FA1	FA2	SA	TW	PR	OR	Total
V	MCE5991	Probability and Statistics for Data Science	3	-	-	3	3	20	20	60	-	-	-	100
V	MCE5992	Programming with Python Laboratory	-	4	-	4	2	-	-	-	50	-	-	50
VI	MCE6991	Basics of Data Science	3	-	-	3	3	20	20	60	-	-	-	100
VI	MCE6992	Data Science Laboratory	-	4	-	4	2	-	-	-	50	-	-	50
VII	MCE7993	Minor Project	-	10	-	10	5	-	-	-	100	-	50	150
VIII	MCE8991	Data Analytics & Visualization	3	-	-	3	3	20	20	60	-	-	-	100
VIII	MCE8992	Data Visualization Laboratory	-	4	-	4	2	-	-	-	50	-	-	50
Total			**9**	**22**	**-**	**31**	**20**	**60**	**60**	**180**	**250**	**-**	**50**	**600**

L-Lecture, **P**-Practical, **T**-Tutorial, **H**-Hours, **Cr**-Credits, **FA**-Formative Assessment, **SA**-Summative Assessment, **TW**-Term Work, **OR**-Oral, **PR**-Practical

Course Syllabus

Semester – V

Minors in Data Science

Program:	Minors in Data Science			Semester:			V
Course:	Probability and Statistics for Data Science			Code:			MCE5991
Teaching Scheme				Evaluation Scheme			
Lecture	Practical	Tutorial	Credit	FA1	FA2	SA	Total
3	-	-	3	20	20	60	100

Prior Knowledge of:
Linear Algebra, Basic Mathematics
is essential.

Course Objectives:

1. To provide foundations of probabilistic and statistical analysis.
2. To learn concept of testing of hypothesis.
3. To formulate problems precisely and to solve the problems.
4. To enable the use of statistical, graphical and algebraic techniques wherever relevant.

Course Outcomes:
After learning the course, the students will be able to:

1. Comprehend the concept of probability and its features.
2. Apply laws of probability to concrete problems.
3. Solve real life problems Using the concept of probability and statistics.
4. Analyze statistical data using measures of central tendency and dispersion.
5. Hypothesize various advanced statistical techniques for modelling and exploring practical situations.
6. Determine the line of best fit for a set of data.

Detailed Syllabus:

Unit	Description	Duration (H)
I	**Introduction to Probability** Basics: sample space, outcomes, probability. Events: mutually exclusive, independent. Counting sample points, Conditional probability, Bayes' rule.	7
II	**Random Variables and Probability Distribution** Concept of Random Variables, Discrete Probability Distributions (Binomial, Poisson, Normal), Continuous Probability Distribution.	7
III	**Mathematical Expectation** Mean of Random variable, Variance of random variable, covariance of random variable with examples.	7
IV	**Descriptive Statistics** Introduction: Statistics. Describing Data with Tables and Graphs: Tables (Frequency Distributions), Graphs. Measures of Central Tendency, Measures of Variability.	8
V	**Statistical Inference** Population and sample, sampling distributions of the mean, Introduction to hypothesis testing: z-test, null hypothesis, chi-square test.	8
VI	**Statistical Models**	8

	Regression and correlation, Basics of Regression, Simple Linear Regression, Least Squares Method, Prediction.	
Total		**45**

Text Books:

1. Morris H. DeGroot, Mark J. Schervish, " Probability and Statistics", Fourth Edition, ISBN 978-0-321-50046-5.

Reference Books:

1. Forsyth, David. Probability and Statistics for Computer Science. Germany, Springer International Publishing, 2017.
2. Ross, Sheldon M. Introduction to probability and statistics for engineers and scientists. Academic press, 2020.

Program:	Minors in Data Science			Semester:			V
Course:	Programming with Python Laboratory			Code:			MCE5992
Teaching Scheme				Evaluation Scheme			
Practical	Tutorial	Credit	Hours	TW	PR	OR	Total
4	-	2	4	50	-	-	50

Course Objectives:

1. To get comfortable with the main elements of Python programming
2. Expose students to application development and prototyping using Python
3. Learn to Apply Fundamental Problem-Solving Techniques using Python

Course Outcomes:

After learning the course, students will be able to:

1. Define and demonstrate the use of built-in data structures "lists" and "dictionary".
2. Implement a given algorithm as a computer program (in Python)
3. Design and implement a program to solve a real-world problem.
4. Read and write data from & to files in Python and develop Application using it
5. Use Python Libraries –NumPy, panda for representing data
6. Develop algorithmic solutions to simple computational problems using python

Guidelines for Students:

- The laboratory assignments are to be submitted by students in the form of a journal.
- Journal consists of a prologue, certificate, table of contents and handwritten write-up of each assignment (Title, Objectives, Problem Statement, Outcomes, Software & Hardware requirements, Date of Completion, Assessment grade/marks and, assessor's sign, Theory- Concept, conclusion/analysis).

Guidelines for Laboratory /TW Assessment:

- Continuous assessment of laboratory work is done based on overall performance and Laboratory performance of students.
- Each Laboratory assignment assessment should assign grades/marks based on parameters with appropriate weightage.
- Suggested parameters for overall assessment as well as each Laboratory assignment assessment include- timely completion, performance, innovation, efficiency, punctuality, and neatness.

Guidelines for Laboratory Conduction

- Assignments on all concepts covered in Group A are mandatory.
- Any Two Assignments from Each Group B & Group C respectively
- Use of open-source software is to be encouraged.
- Operating System recommended: - 64-bit Open-source Linux or its derivative.
- Programming tools recommended: - python 3.0 and above, Sublime/Atom/PyCharm

Assignment No.	Suggested List of Assignments
	Group A
1	Write a program to create, concatenate, and print string and accessing sub-string from a given string.
2	Write a program to create, append, and remove lists in python.
3	Write a program to demonstrate working with dictionaries in python

4	Write a program that inputs a text file. Program should print all of the unique words in the file in alphabetical order.
Group B	
5	Make a Binomial Random variable XX and compute its probability mass function (PMF) or cumulative density function (CDF)
6	Write a NumPy program to calculate the difference between the maximum and the minimum values of a given array along the second axis. Expected Output: Original array: [[0 1 2 3 4 5] [6 7 8 9 10 11]] Difference between the maximum and the minimum values of the said array: [5 5]
7	Write a Pandas program to find the sum, mean, max, min value from the excel sheet.
Group C	
8	Write a program to calculate the correlation between variables in python.
9	Conduct and interpret the Chi-square test of Independence using Python
10	Write a Pandas program to convert a Panda module Series to a Python list and its type.

Reference Books:

1. Allen B. Downey, Jeffrey Elkner, Peter Wentworth, and Chris Meyers, "How to think like a computer scientist: Learning with Python", Green Tea Press (2002), eBook (GNU Edition, Read TheDocs.org, Apr 17, 2020), ISBN-10:0971677506, ISBN-13**:**978-0971677500
2. Martin C. Brown, "Python: The Complete Reference", McGraw-Hill/Osborne Media, September 21st 2001, ISBN 007212718X(ISBN13:9780072127188)
3. Alex Martelli, Anna Martelli Ravenscroft, David Ascher, "Python Cookbook", Oreilly publications
4. Zed A. Shaw, "Learn PYTHON The Hard Way", Pearson, ISBN: 978-93-325-8210-1
5. Kenneth A Lambert and B L Juneja, "Fundamentals of PYTHON", CENGAGE Learning, ISBN:978-81-3152903-4

Web reference:

1. https://www.python.org/about/gettingstarted/
2. https://www.udacity.com/course/introduction-to-python--ud1110

Course Syllabus

Semester – VI

Minors in Data Science

Program:	Minors in Data Science			Semester:			VI
Course:	Basics of Data Science			Code:			MCE6991
Teaching Scheme				Evaluation Scheme			
Lecture	Practical	Tutorial	Credit	FA1	FA2	SA	Total
3	-	-	3	20	20	60	100

Prior knowledge of
Basic concepts of Mathematics and Statistics (Linear Algebra, Calculus, Probability)
is essential.

Course Objectives:

1. To understand basics of data science concepts.
2. To apply data preprocessing techniques.
3. To apply machine learning algorithms.
4. To explore different tools of data science.

Course Outcomes:

After learning the course, the students should be able to:
1. Differentiate between Data Science and Big Data Concepts.
2. Apply Data Preprocessing techniques on real databases.
3. Apply Exploratory Data Analysis on various data sets to solve real life applications.
4. Solve real life problems using regression techniques.
5. Apply machine learning algorithms for real world applications.
6. Implement data science concepts using various tools and techniques.

Detailed Syllabus:

Unit	Description	Duration (H)
I	**Introduction to Data Science** Evolution of Data Science, Data Science Roles, Stages in a Data Science Project, Applications of Data Science in various fields, Introduction to Big Data and Big Data Analytics.	7
II	**Data Collection and Data Pre-Processing** Data Collection Strategies, Data Pre-Processing Overview, Data Cleaning, Data Integration and Transformation, Data Reduction, Data Discretization.	7
III	**Exploratory Data Analysis (EDA)** Statistical measures, Basic tools (plots, graphs and summary statistics) of EDA, Data Analytics Lifecycle, Discovery Developing Initial Hypotheses, Identifying Potential Data Sources, EDA case study, testing hypotheses on means, proportions and variances	7
IV	**Regression models** Simple linear regression, least-squares principle, MLR, logistic regression, Multiple correlation, Partial correlation	8
V	**Model Evaluation** Generalization Error, Out-of-Sample Evaluation Metrics, Cross Validation, Overfitting, Under Fitting and Model Selection, Prediction by using Ridge Regression, Testing Multiple Parameters by using Grid Search.	8

VI	**Introduction to Machine Learning** Supervised learning, Unsupervised learning, Association Rule mining, Naive Bayes Classifier, k-means clustering.	**8**
Total		**45**

Text Books:

1. Jiawei Han, Micheline Kamber and Jian Pei, "Data Mining: Concepts and Techniques", Third Edition, ISBN 0123814790, (2011).
2. Cathy O'Neil and Rachel Schutt, "Doing Data Science", O'Reilly, 2015, ISBN: 9781449358655

Reference Books:

1. Seema Acharya, Subhasini Chellappan, "Big Data Analytics", paperback 2nd ed., Wiley (2019), ISBN:9788126579518
2. Rajan Chattamvelli, Narosa Publication, "Data mining methods",2nd edition, 2016, ISBN: 978-81-8487-510-2.

Web references:

1. http://www.analyticsvidhya.com/
2. http://www.datasciencecentral.com/

Program:	Minors in Data Science			Semester:			VI
Course:	Data Science Laboratory			Code:			MCE6992
Teaching Scheme				Evaluation Scheme			
Practical	Tutorial	Credit	Hours	TW	PR	OR	Total
4	-	2	4	50	-	-	50

Course Objectives:

1. To analyze data using statistical methods.
2. To apply data cleaning techniques.
3. To apply machine learning algorithms.
4. To apply regression techniques.
5. Analyze various data sets from standard repositories.

Course Outcomes:

After learning the course, students will be able to:

1. Analyze the data by applying statistical methods.
2. Apply Data Preprocessing techniques on real databases.
3. Apply Exploratory Data Analysis on various data sets to solve real life applications.
4. Solve real life problems using regression techniques.
5. Apply machine learning algorithms for real world applications.
6. Implement data science concepts using various tools and techniques.

Guidelines for Students:

- The laboratory assignments are to be submitted by students in the form of a journal.
- Journal consists of prologue, certificate, table of contents and handwritten write-up of each assignment (Title, Objectives, Problem Statement, Outcomes, Software & Hardware requirements, Date of Completion, Assessment grade/marks and assessor's sign, Theory- Concept, conclusion/analysis).

Guidelines for Laboratory /TW Assessment:

- Continuous assessment of laboratory work is done based on overall performance and Laboratory performance of students.
- Each Laboratory assignment assessment should assign grade/marks based on parameters with appropriate weightage.
- Suggested parameters for overall assessment as well as each Laboratory assignment assessment include- timely completion, performance, innovation, efficiency, punctuality, and neatness.

Guidelines for Laboratory Conduction:

- Assignments on all concepts covered in the following list are mandatory.
- Students are suggested to take suitable dataset of their interest from Kaggle, UCI or any standard repository.
- Use of open-source software is to be encouraged. (Python, R Studio, Weka etc.)

Assignment No.	Suggested List of Assignments
1	Apply basic statistical techniques to analyze and understand properties of the data on any standard dataset downloaded from Kaggle, UCI repository.
2	Write python code that loads any standard dataset downloaded from Kaggle, UCI repository, and plotthe graph.
3	Perform some basic data cleaning operations on standard dataset using Python / R / Weka etc. (Dataset should contain minimum 1000 records)
4	Perform data preprocessing, exploratory analysis on any standard dataset (dataset should contain minimum 1000 records) Using Python /R.
5	Perform correlation and regression analysis for any standard dataset downloaded from Kaggle, UCI, Government databases etc using Python / R.

6	Perform Simple Linear regression on for any standard data set downloaded from Kaggle, UCI, Government databases Using Python / R
7	Perform data analysis using Python / R / Weka for readily available data set using any one machine learning algorithm.

Reference Books:

1. Jiawei Han, Micheline Kamber and Jian Pei, "Data Mining: Concepts and Techniques", Third Edition, ISBN0123814790, (2011).
2. Martin C. Brown, "Python: The Complete Reference", McGraw-Hill/Osborne Media , September 21st 2001 , ISBN 007212718X(ISBN13:9780072127188)
3. Hadley Wickham,Garrett Grolemund, " R for Data Science" , O'Reilly Media, 12 December 2016, ISBN:9781491910344, 1491910348

Web Reference:

4. http://www.datasciencecentral.com/
5. https://towardsdatascience.com

Course Syllabus
Semester – VII/VIII
Minors in Data Science

Program:	Minors in Data Science			Semester:	VII / VIII		
Course:	Data Analytics and Visualization			Code:	MCE7991/ MCE8991		
Teaching Scheme				Evaluation Scheme			
Lecture	Practical	Tutorial	Credit	FA1	FA2	SA	Total
3	-	-	3	20	20	60	100

Prior knowledge of
Basics of Data Science
is essential.

Course Objectives:
1. To understand basics of data analytic concepts.
2. To apply data analytic life cycle to business problems.
3. To apply machine learning algorithms.
4. To visualize big data using visualization tools and techniques.

Course Outcomes:
After learning the course, the students should be able to:
1. Apply Data Analytics Life Cycle to business problem.
2. Analyze data using basic statistical methods.
3. Analyze and interpret the business data.
4. Apply data analysis methods to evaluate the business model.
5. Build and use appropriate models of data analysis to answer business related questions.
6. Visualize Big data using appropriate tools.

Detailed Syllabus

Unit	Description	Duration (H)
I	**Introduction to Data Analytic** Big data overview, BI Vs Data Analytics Current Analytical Architecture, drivers of Big Data, Emerging Big Data Ecosystem and new approach. Data Analytical Life cycle: Overview, phase 1- Discovery, Phase 2- Data preparation, Phase 3- Model Planning, Phase 4- Model Building, Phase 5- Communicate Results, Phase 6- Operationalize.	7
II	**Basic Data Analytical Methods** Statistical Methods for Evaluation- Hypothesis testing, difference of means, wilcoxon rank–sum test, type 1 type 2 errors, power and sample size, ANNOVA.	7
III	**Advanced Data Analytical Methods -Clustering** Clustering- Overview, K means- Use cases, Overview of methods, determining number of clusters, diagnostics, reasons to choose and cautions.	7
IV	**Advance Analytical Methods-Association** A Overview, a-priori algorithm, evaluation of candidate rules, case study-transactions in grocery store, validation and testing, diagnostics.	8
V	**Data Visualization Techniques** Introduction to Data visualization, Correlation of Data, Challenges to Big data visualization, Conventional data visualization tools, Techniques for visual data representations,	8

VI	**Data Visualization Tools** Visualizing using Excel or Tableau or R or Python, Visualization Before Analysis, Dirty data, Examining single and Multiple Variables, Data Exploration Versus Presentation	**8**
	Total	**45**

Text Books:

1. David Dietrich, Barry Hiller, "Data Science and Big Data Analytics", EMC education services, Wiley publications, 2012, ISBN0-07-120413-X.

Reference Books:

1. Maheshwari Anil, Rakshit, Acharya, "Data Analytics", McGraw Hill, ISBN: 789353160258.
2. Mark Gardner, "Beginning R: The Statistical Programming Language", Wrox Publication, ISBN: 978-1-118-16430-3

Web references:

1. http://www.datasciencecentral.com/
2. https://towardsdatascience.com

Program:	Minors in Data Science			Semester:			VII / VIII
Course:	Data Visualization Laboratory			Code:			MCE7992 / MCE8992
Teaching Scheme				Evaluation Scheme			
Practical	Tutorial	Credit	Hours	TW	PR	OR	Total
4	-	2	4	50	-	-	50

Course Objectives:

1. To understand basics of data analytic concepts.
2. To apply data analytic life cycle to business problems.
3. To apply machine learning algorithms.
4. To visualize big data using visualization tools and techniques.

Course Outcomes:

After learning the course, students will be able to:

1. Apply Data Analytics Life Cycle to business problem.
2. Analyze data using basic statistical methods.
3. Analyze and interpret the business data.
4. Apply data analysis methods to evaluate the business model.
5. Build and use appropriate models of data analysis to answer business related questions.
6. Visualize Big data using appropriate tools.

Guidelines for Students:

- The laboratory assignments are to be submitted by students in the form of a journal.
- Journal consists of prologue, certificate, table of contents and handwritten write-up of each assignment (Title, Objectives, Problem Statement, Outcomes, Software & Hardware requirements, Date of Completion, Assessment grade/marks and assessor's sign, Theory- Concept, conclusion/analysis).

Guidelines for Laboratory /TW Assessment:

- Continuous assessment of laboratory work is done based on overall performance and Laboratory performance of students.
- Each Laboratory assignment assessment should assign grade/marks based on parameters with appropriate weightage.
- Suggested parameters for overall assessment as well as each Laboratory assignment assessment include- timely completion, performance, innovation, efficiency, punctuality, and neatness.

Guidelines for Laboratory Conduction:

- Assignments on all concepts covered in the following list are mandatory.
- Students are suggested to take suitable dataset of their interest from Kaggle, UCI or any standard repository.
- Use of open-source software is to be encouraged. (Python, R Studio, etc.)

Assignment No.	Suggested List of Assignments
1	Introduction to python library such as scikit, sklearn Library, Pandas, dataframe, etc.
2	Analysis of any suitable dataset like IRIS a. How many features are there and what are their types (e.g., numeric, nominal)? b. Compute and display summary statistics for each feature available in the dataset. (e.g. minimum value, maximum value, mean, range, standard deviation, variance and percentiles Please Visit UCI repository or Kaggle for standard dataset.
3	Using clustering form the clusters of Trip History data. Use trip history dataset that is from a bike sharing service in the United States. Predict the class of user. Sample Test data set available here. Students are encouraged to use any other suitable dataset. https://www.capitalbikeshare.com/trip-history-data

4	Find the strong rules and frequent itemset from given Grocery store datasets. https://www.kaggle.com/ekrembayar/apriori-association-rules-grocery-store/data or use any suitable dataset.
5	Use any visualization tool (Excel/Python /R/ Tableau) and visualize any standard dataset of interest and analyze it.
6	Consider a suitable dataset for clustering of data instances in different groups, apply different clustering techniques (minimum 2). Visualize the clusters using suitable tool.
7	Analysis and prediction of stock market data.

Reference Books:

1. Core Python Programming, "R. Nageswara Rao", dreamtech Maheshwari Anil, Rakshit, Acharya, "Data Analytics", McGraw Hill, ISBN: 789353160258.
2. David Dietrich, Barry Hiller, "Data Science and Big Data Analytics", EMC education services, Wiley publications, 2012, ISBN0-07-120413-X
3. Martin C. Brown, "Python: The Complete Reference", McGraw-Hill/Osborne Media , September 21st 2001 , ISBN 007212718X(ISBN13:9780072127188)
4. Allen B. Downey, Jeffrey Elkner, Peter Wentworth, and Chris Meyers, "How to think like a computer scientist: Learning with Python", Green Tea Press (2002);, eBook (GNU Edition, Read TheDocs.org, Apr 17, 2020) , ISBN-10:0971677506, ISBN-13:978-0971677500

Web reference:

1. http://www.datasciencecentral.com/
2. https://towardsdatascience.com

Program:	Minors in Data Science			Semester:			VII / VIII
Course:	Minor Project			Code:			MCE7993 / MCE8991
Teaching Scheme				Evaluation Scheme			
Practical	Tutorial	Credit	Hours	TW	PR	OR	Total
10	-	5	10	100	-	50	**150**

Course Objectives:

1. To apply the knowledge for solving realistic problem.
2. To develop problem solving ability.
3. To organize, sustain and report on a substantial piece of team work over a period of several months
4. To evaluate alternative approaches, and justify the use of selected tools and methods,
5. To follow SDLC meticulously and meet the objectives of proposed work
6. To test rigorously before deployment of system
7. To validate the work undertaken
8. To consolidate the work as furnished report

Course Outcomes:

After learning the course, students will be able to:

1. Solve real life problems by applying knowledge of data science.
2. Analyze alternative approaches and apply appropriate solution.
3. Prepare to solve multi-disciplinary and heterogeneous business problems.
4. Apply modern tools and techniques to solve real life applications.
5. Develop Inter-personal communication, management and leadership quality to work effectively in team.
6. Prepare report and present the results in an orderly way.

Guidelines for Students:

- Minor Project is one of the significant contributions of a student that has to be completed during 8th semester in group of 2 to 4 students based on the required number of credits as per academic regulations.
- Explore the current research trends / Thrust areas / society needs for problem identification.
- Guideline sessions to explain project work will be arranged by project coordinator, Minor Coordinator with HOD and panel of experts.
- **Types of projects:** Projects may be in-house, Sponsored or Multidisciplinary. Projects can be carried out inside or outside the institute, in any relevant industry or research institution.

1. Selection of Project:

- Students shall identify the area or topics in recent trends and developments in consultation with institute guide or industry or any research organization.

2. Guide Allotment:

- Considering registered student's area and expertise of guide, the Project coordinator in consultation with a panel of experts allots Project guides.
- Guide should be allotted from the same program.
- In case of interdisciplinary project, along with the guide from same program, co-guide should be allotted from the other program

- In case of a sponsored project (with a reputed industry or any research organization) the external guide should be from the sponsored company/Industry along with the internal guide from the program.
- Students in consultation with guides will prepare project Synopsis.
- Project synopsis must include project title, sponsorship details (if any), detailed problem definition, area, type of project [Sponsored/Non-Sponsored/Interdisciplinary], abstract, scope of the project and software-hardware requirements. [Sponsorship details include name of sponsoring authority, address, name of external guide, sponsorship terms and conditions and respective documents certifying the same from authorities].

3. Project Review:

- The Project coordinator with the Head of the department shall constitute a review committee for project groups.
- A review committee will approve the project title. Discussion / presentation may be arranged covering topics listed in the synopsis.
- The Project Review committee will be responsible for evaluating the timely progress of the projects.
- Students are expected to appear for all three reviews as per the project calendar.

4. Report:

- Report should be prepared as per the template provided by the program.
- Project reports shall be submitted in softcopy form/ (Hard bound reports could be avoided).
- In the case of Sponsored project, students must submit a Completion certificate with the signature of an external guide from the sponsored company.

5. Expected Project Deliverables/Outcomes:

- Students are encouraged to do at least one activity from the list below.
    1. Paper publication in quality journal/conference (such as SCOPUS / UGC approved)
    2. Consultancy Work
    3. IPR (Patent / Copyright)
    4. Participation in various Project Competitions/Awards

Project Work syllabus guidelines:

- Students in group of 2 to 4 are expected to initiate the minor project work at the start of 8th semester, clearly identify the Problem, objectives of the study, the scope which indicates the usefulness of the project, Do complete Literature review, Data Collection, Analysis of data. By selecting appropriate methodology, tools complete the implementation, testing and results of problem statement in stipulated time period.
- The student shall prepare and submit the report of Project work in standard format for satisfactory completion of the work that is the duly certified by the concerned guide and head of the Department/Institute.

Vision and Mission of Computer Department

Department Vision

To be a premier Computer Engineering program by achieving excellence in Academics and Research for creating globally competent and ethical professionals.

Department Mission

M1: To develop technologically competent and self-sustained professionals through contemporary curriculum.

M2: To nurture innovative thinking and collaborative research, making a positive impact on society.

M3: To provide state-of-the art computing environment and learning opportunities through Center of Excellence.

M4: To foster leadership skills and ethics with holistic development.