

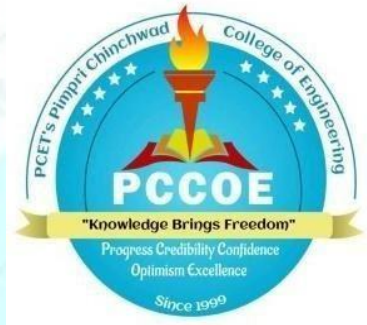
Pimpri Chinchwad Education Trust's

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 2023)



Academic Year 2026-28

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.

EOMS Policy

“We at PCCOE are committed to offer exemplarily Ethical, Sustainable and Value Added Quality Education to satisfy the applicable requirements, needs and expectations of the Students and Stakeholders.

We shall strive for technical development of students by creating globally competent and sensible engineers, researchers, and entrepreneurs through Quality Education.

We are committed for Institute’s social responsibilities and managing Intellectual property.

We shall achieve this by establishing and strengthening state-of-the-art Engineering Institute through continual improvement in effective implementation of Educational Organizations Management Systems (EOMS).”

"Knowledge Brings Freedom"

Progress Credibility Confidence

Optimism Excellence

Since 1979

	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	BCE25MN01	5	
2	Programming with Python Laboratory	BCE25MN02	7	
3	Data Science and Machine Learning	BCE26MN03	10	
4	Data Science Laboratory	BCE26MN02	12	
5	Data Analytics & Visualization	BCE27MN04	15	
7	Minor Project	BCE28MN04	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 18 credits in multidisciplinary areas of other domains.

Objectives of Honors 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 that 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

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

Learning Outcomes:

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

1. Analyze the data by applying statistical methods.
2. Hypothesize various advanced statistical techniques for modelling and exploring practical situations.
3. Apply Exploratory Data Analysis on various data sets to solve real life applications.
4. Apply machine learning algorithms and Implement data science concepts using various tools and techniques for real world applications.
5. Apply Data Analytics Life Cycle to business problem.
6. 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	4
4	Course Syllabus of Semester – VI Courses	9
5	Course Syllabus of Semester – VII Courses	14
6	Course Syllabus of Semester – VIII Courses	19
7	Vision and Mission of Computer Engineering Department	21

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.	FA1	Formative Assessment 1
7.	FA2	Formative Assessment 2
8.	SA	Summative Assessment
9.	TW	Term Work
10.	OR	Oral
11.	PR	Practical
12.	PROJ	Project

Curriculum Structure

Minors in Data Science

"Knowledge Brings Freedom"

Progress Credibility Confidence
Optimism Excellence

Since 1978

CURRICULUM STRUCTURE**Structure for Minors in Data Science****(Computer Engineering)**

Semester	Coursecode	Course Name	Teaching Scheme				CR	Evaluation Scheme						
			L	P	T	H		FA1	FA2	SA	TW	PR	OR	Total
V	BCE25MN01	Probability and Statistics for Data Science	3	-	-	3	3	20	20	60	-	-	-	100
V	BCE25MN02	Programming with Python Laboratory	-	4	-	4	2	-	-	-	50	-	-	50
VI	BCE26MN03	Data Science and Machine Learning	3	-	-	3	3	20	20	60	-	-	-	100
VI	BCE26MN02	Data Science Laboratory	-	4	-	4	2	-	-	-	50	-	-	50
VII	BCE27MN04	Data Analytics & Visualization	2	4			4	10	10	30	50			100
VIII	BCE28MN04	Minor Project	-	10	-	10	4	-	-	-	100	-	50	150
	Total		9	22	-	31	18	50	50	150	250	-	50	550

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

The logo of PCCOE (Pimpri Chinchwad College of Engineering) is a circular emblem. It features a central torch with a flame, set against a background of a gear-like border. The text 'PCCOE' is prominently displayed in the center of the emblem. Below the emblem, a banner reads 'Knowledge Brings Freedom'. Further down, the words 'Progress, Credibility, Confidence' and 'Optimism, Excellence' are written, followed by 'Since 1979' at the bottom.

Course Syllabus
Semester – V
Minors in Data Science

"Knowledge Brings Freedom"

Progress, Credibility, Confidence
Optimism, Excellence

Since 1979

Program:	Minors in Data Science					Semester: V	
Course:	Probability and Statistics for Data Science					Code: BCE25MN01	
Credit	Teaching Scheme (Hrs./Week)			Evaluation Scheme and Marks			
	Lecture	Practical	Tutorial	FA		SA	Total
				FA1	FA2		
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 and statistics to solve problems. 3. Analyze statistical data using measures of central tendency and dispersion. 4. Hypothesize various advanced statistical techniques for modelling and exploring practical situations 5. 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 Self Learning Spam detection using Bayes, Reliability of systems (AND/OR models) Simple game probability analysis						9
II	Random Variables and Probability Distribution Concept of Random Variables, Discrete Probability Distributions (Binomial, Poisson, Normal), Continuous Probability Distribution. Mean, Variance, Standard deviation, Moments (central and about the origin), Skewness and kurtosis, Self-Learning						9

	Bernoulli, Binomial, Poisson Distribution	
III	Mathematical Expectation Mean of Random variable, Variance of random variable, covariance of random variable with examples. Self-Learning : Correlation and covariance	9
IV	Descriptive Statistics Introduction: Statistics. Describing Data with Tables and Graphs: Tables (Frequency Distributions), Graphs. Measures of Central Tendency, Measures of Variability. Self Learning Position Measures Quartiles, deciles, percentiles Z-scores (standard scores)	9
V	Statistical Inference Population and sample, sampling distributions of the mean, Introduction to hypothesis testing: z-test, null hypothesis, chi-square test. Correlation and regression Self Learning Type I error and Type II error, Critical region approach, p-value approach	9
Total		45
Text Books: 1. Morris H. DeGroot, Mark J. Schervish, " Probability and Statistics", Fourth Edition, ISBN 978-0-321-50046-5 by Pearson Education, 2012.		
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.		

"Knowledge Brings Freedom"

Progress, Creativity, Confidence

Optimism, Excellence

Since 1978

Program:	Minors in Data Science			Semester: V			
Course:	Programming with Python Laboratory			Code: BCE25MN02			
Credit	Teaching Scheme (Hrs./Week)			Evaluation Scheme and Marks			
	Lecture	Practical	Tutorial	TW	OR	PR	Total
2	-	4	-	50	-	-	50
Prior knowledge of Basic mathematics is essential							
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							
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							
Detailed Syllabus							
Suggested List of Assignments							
Assignment No.	Assignment Title						

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.
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.
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.
11.	Write a Python program to create a dataframe by importing CSV, JSON files and perform various operations.
Text Book 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) Reference Book 1. Alex Martelli, Anna Martelli Ravenscroft, David Ascher, "Python Cookbook", Oreilly publications 2. Zed A. Shaw, "Learn PYTHON The Hard Way", Pearson, ISBN: 978-93-325-8210-1 3. Kenneth A Lambert and B L Juneja, "Fundamentals of PYTHON", CENGAGE Learning, ISBN:978-81-3152903-4	

Course Syllabus
Semester – VI
Minors in Data Science

Program:	Minors in Data Science					Semester: VI	
Course:	Data Science and Machine Learning					Code: BCE26MN03	
Credit	Teaching Scheme (Hrs./Week)			Evaluation Scheme and Marks			
	Lecture	Practical	Tutorial	FA		SA	Total
				FA1	FA2		
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 Association Rule mining algorithm for market basket analysis 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. Data-driven decision making Self Learning : Emerging Big Data Ecosystem and new approach						7
II	Data Collection and Data Pre-Processing Data Collection Strategies, Structured, Semi-structured, Unstructured data, Open datasets (government portals, Kaggle, UCI) , Enterprise databases, Data Pre-Processing Overview, Data Cleaning, Data Integration and Transformation, Data Reduction, Data Discretization. Self Learning: Data Splitting, Train-test split Validation sets Cross-validation basics						7
III	Exploratory Data Analysis (EDA) Statistical measures, Basic tools (plots, graphs and summary statistics) of EDA, Data Analytics Lifecycle, Discovery Developing Initial Hypotheses,						7

	Identifying Potential Data Sources, EDA case study, testing hypotheses on means, proportions and variances Self Learning : Use any open source tool like Power Bi, KNIME and analyze data of any company/ sales etc	
IV	Supervised Learning : Regression Linear regression, A bi-dimensional Example, Linear Regression and higher dimensionality, Regularization Ridge, Lasso ,Logistics Regression, Linear Classification Classification : Naïve Bayes Classifier, Confusion Matrix Self Learning Case study of Application of house price prediction, share market etc.	8
V	Unsupervised Learning : Association Rule Mining : Market Basket Analysis, Frequent Item Set, Apriori Algorithm Clustering- Kmeans, Self Learning Learn application of shopping mall for market basket Analysis	8
VI	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. Self Learning Build any Application in domain like health care or sales or agriculture etc domain using analytical tools like python/ WEKA/ KNIME/R	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/ 3. https://www.geeksforgeeks.org/machine-learning		

Program:	Minors in Data Science			Semester: VI			
Course:	Data Science Laboratory			Code: BCE26MN02			
Credit	Teaching Scheme (Hrs./Week)			Evaluation Scheme and Marks			
	Lecture	Practical	Tutorial	TW	OR	PR	Total
2	-	4	-	50	-	-	50
Prior knowledge of Machine Learning basic concepts is essential							
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: 1. The laboratory assignments are to be submitted by students in the form of a journal. 2. 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: 1. Continuous assessment of laboratory work is done based on overall performance and Laboratory performance of students. 2. Each Laboratory assignment assessment should assign grade/marks based on parameters with appropriate weightage. 3. 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: 1. Assignments on all concepts covered in the following list are mandatory. 2. Students are suggested to take suitable dataset of their interest from Kaggle, UCI or any standard repository. 3. Use of open-source software is to be encouraged. (Python, R Studio, Weka etc.)							
Detailed Syllabus							
Suggested List of Assignments							
Assignment No.	Assignment Title						
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 plot the 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	Download Diabetic data from kaggle/ UCI / any other source, Build classification model for diabetic prediction and Evaluate the model using accuracy, precision, Recall etc.,. Example Dataset: (https://www.kaggle.com/datasets/uciml/pima-indians-diabetes-database) (Note: Use sci-kit learn Library)
8	Select a suitable real-world dataset and apply K-Means clustering to group similar data points. Use clustering results to derive meaningful insights for decision-making.
9	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. <u>Hadley Wickham</u>, Garrett Golemund, " R for Data Science" , O'Reilly Media, 12 December 2016, ISBN:9781491910344, 1491910348	
Web Reference: 1. http://www.datasciencecentral.com/ 2. https://towardsdatascience.com	

"Knowledge Brings Freedom"

Progress Creativity Confidence

Optimism Excellence

Since 1979

Course Syllabus
Semester – VII
Minors in Data Science

"Knowledge Brings Freedom"

Progress Credibility Confidence
Optimism Excellence

Since 1979

Program:	Minors in Data Science					Semester: VII				
Course:	Data Analytics and Visualization					Code: BCE27MN04				
Credit	Teaching Scheme (Hrs./Week)			Evaluation Scheme and Marks						
	Lecture	Practical	Tutorial	FA		SA	TW	PR	OR	Total
				FA1	FA2					
4	2	4	-	10	10	30	50	-	-	150
Prior knowledge of Basics of Data Science is essential										
Course Objectives: This course aims at enabling students: 1. To understand the basics of data analytic concepts. /Develop a thorough understanding of each stage of the data analytics lifecycle, from data discovery and preparation to modeling, evaluation, and operationalization. 2. To apply data analytic life cycle to business problems. 3. To analyze and interpret datasets using exploratory data analysis and visualization techniques. 4. To visualize big data using visualization tools and techniques.										
Course Outcomes: After learning the course, the students will be able to: 1. Apply the Data Analytics Lifecycle to business problems. 2. Analyze and interpret the business data using basic statistical methods. 3. Apply appropriate data analysis techniques to solve real-world business problems. 4. Visualize data using modern visualization tools.										
Guidelines for Laboratory/Term Work Assessment: 1. Continuous assessment of laboratory work is done based on overall performance and lab assignments performance of students. Each lab assignment assessment will assign marks based on parameters with appropriate weightage. 2. 50% weightage of term work is based on suggested parameters for lab assignment includes- understanding and participation, timely completion, punctuality and neatness, overall professional behavior. 3. 50% weightage of term work is based on internal oral/practical examination. 4. All assignments are compulsory.										
Guidelines for Laboratory/Tutorial Conduction: Tools: Python, Pandas , Matplotlib , Seaborn , Tableau / Power BI Best Practices: Use of Github, Use of open datasets (Kaggle, UCI repository)										
Detailed Syllabus										
Unit	Description									Duration (H)
I	Classroom Instructions: Introduction to Data Analytics Lifecycle and Data Handling:									7

	Data analytics lifecycle: data discovery, preparation, modeling, evaluation and deployment, Data acquisition methods and data sources (structured, semi-structured, unstructured data) , Data storage formats (CSV, JSON, databases) , Data preparation and feature understanding Self-Learning : Study real-world case studies where data analytics is used for decision making in domains such as healthcare, manufacturing, and finance.	
II	Classroom Instructions: Data Preprocessing and Exploratory Data Analysis: Data preprocessing techniques , Data transformation and feature engineering , Exploratory Data Analysis (EDA) techniques, Statistical methods for evaluation: Hypothesis testing , Difference of means , Wilcoxon Rank–Sum Test , Type I and Type II errors, Power and sample size , ANOVA Case study: Perform EDA to explore Online Retail dataset Case study: Analyze mental health of IT professionals Self-Learning : Study feature engineering techniques used in real-world analytics workflows.	8
III	Data Analysis Techniques: Association and Clustering : Clustering : Overview of clustering , K-Means clustering and use cases , Determining number of clusters, Evaluation and diagnostics of clustering results, Association Analysis: Overview of association rule mining , Apriori algorithm, Evaluation of candidate rules (support, confidence, lift) Case study: grocery store transaction analysis Self-Learning : Study real-world applications of clustering and association analysis in recommendation systems and retail analytics.	7
IV	Classroom Instructions: Data Visualization Tools: Principles of effective data visualization, Visualization techniques for comparison, distribution, and relationships , Statistical visualizations (histograms, box plots, correlation plots) , Visualization using Matplotlib and Seaborn , Interactive visualization concepts , Dashboard design and visual storytelling Visualization tools: Tableau , Power BI , Plotly , Google Looker Studio Principles of effective data visualization and visual encoding, Interactive data visualization concepts ,Dashboard design principles and data storytelling Visualization using Python tools: Matplotlib and Seaborn for statistical visualization, Plotly for interactive visualizations Modern data visualization platforms: Tableau, Power BI, Google Looker Studio Self-Learning : Explore modern visual analytics tools and develop interactive dashboards for effective data-driven decision making.	8
Total		30

Suggested List of Assignment

1. Introduction to Python libraries such as NumPy, Pandas, Matplotlib, Seaborn, and Scikit-learn.

Perform basic operations on DataFrames, data handling, and visualization of simple datasets.

2. Perform Exploratory Data Analysis (EDA) on any suitable dataset such as IRIS / Titanic:

a. Identify number of features and their types (numeric, categorical)

b. Compute summary statistics (minimum, maximum, mean, standard deviation, variance, percentiles)

c. Detect missing values and outliers

3. Perform statistical analysis on a suitable dataset by applying hypothesis testing, difference of means test, and ANOVA. Interpret the results with respect to Type I and Type II errors.

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. Consider a suitable dataset and apply different clustering techniques:

a. Compare clustering results (e.g., K-Means, Hierarchical)

b. Visualize clusters using Python / Tableau / Power BI

c. Interpret grouping patterns

6. Develop an interactive dashboard using Tableau/Power BI:

a. Include multiple visualizations

b. Apply filters, slicers, or drill-down features

c. Ensure effective layout and usability

7. Perform data analytics and visualization using AI-powered tools like Microsoft PowerBI, Tableau, Google Looker Studio, Pandas AI on a suitable dataset:

a. Use an AI tool to automatically generate visualizations and insights from the dataset

b. Create interactive charts and dashboards using AI assistance

c. Analyze patterns, trends, and key insights generated by the tool

d. Compare AI-generated insights with manually created visualizations

8. Develop a real-time data visualization dashboard using live/streaming data:

a. Collect real-time data (e.g., stock/weather)

b. Create dynamic visualizations

c. Analyze real-time trends and insights

Self Learning:

1. Study modern data analytics and visualization tools and their importance in decision making.
2. Apply clustering and association techniques on suitable datasets and analyze the results for real-world applications.
3. Study AI-assisted analytics tools (such as Power BI/Tableau/AI-based platforms) and prepare interactive dashboards, visual reports, and data storytelling interfaces for the selected dataset.

Text Books:

1. David Dietrich, Barry Hiller, "Data Science and Big Data Analytics", EMC education services, Wiley publications, 2012, ISBN 0-07-120413-X.
2. Wes McKinney, "Python for Data Analysis", O'Reilly Media, 2nd Edition, 2017, ISBN 978-1491957660.

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
3. Core Python Programming, "R. Nageswara Rao", dreamtech Maheshwari Anil, Rakshit, Acharya, "Data Analytics", McGraw Hill, ISBN: 789353160258.
4. David Dietrich, Barry Hiller, "Data Science and Big Data Analytics", EMC education services, Wiley publications, 2012, ISBN 0-07-120413-X
5. Martin C. Brown, "Python: The Complete Reference", McGraw-Hill/Osborne Media , September 21st 2001 , ISBN 007212718X (ISBN13:9780072127188)

PCCOE
"Knowledge Brings Freedom"

Progress Credibility Confidence
Optimism Excellence
Since 1979

Course Syllabus
Semester – VIII
Minors in Data Science

Program:	Minors in Data Science			Semester:VIII			
Course:	Minor Project			Code: BCE28MN04			
Credit	Teaching Scheme (Hrs./Week)			Evaluation Scheme and Marks			
	Lecture	Practical	Tutorial	TW	OR	PR	Total
4	-	8	-	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

Vision and Mission of Computer Department

Department Vision

To be a premier Computer Engineering Department 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.