Shirpur Education Society's
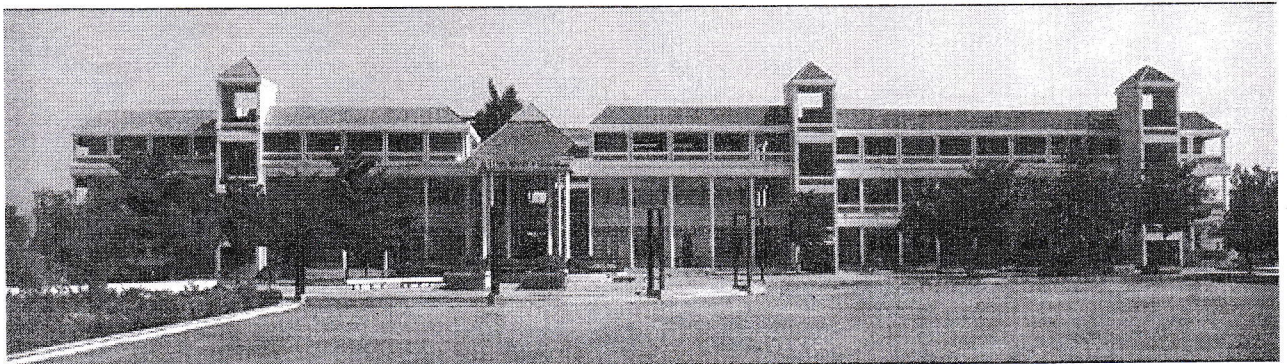
R. C. Patel Institute of Technology, Shirpur

(An Autonomous Institute)

Course Structure and Syllabus

Honors Program in Data Science

Computer Engineering

With effect from Year 2024-25

Shahada Road, Near Nimzari Naka, Shirpur, Maharashtra 425405
Ph: 02563 259 802, Web: www.rcpit.ac.in

Honors Program in Data Science offered by Computer Engineering (w.e.f. 2024-25)

Sr. No.	Course Category	Course Code	Course Title	Teaching Scheme L	Teaching Scheme T	Teaching Scheme P	CA TA [A]	CA Term Test 1 (TT1)	CA Term Test 2 (TT2)	CA Average of (TT1 & TT2) [B]	ESE [C]	Total [A+B+C]	Credit	Credit
Sem-III														
1	H1	RCP23CCH1301	Statistics for Data Science	3			25	15	15	15	60	100	3	3
Sem-IV														
2	H1	RCP23CCH1401	Visualization in Data Science	3			25	15	15	15	60	100	3	4
	H1	RCP23CLH1401	Visualization in Data Science Laboratory			2	25				25	50	1	
Sem-V														
3	H1	RCP23CCH1501	Graph Data Science	3			25	15	15	15	60	100	3	4
	H1	RCP23CLH1501	Graph Data Science Laboratory			2	25				25	50	1	
Sem-VI														
4	H1	RCP23CCH1601	Essentials of Generative AI and Prompt Engineering	3			25	15	15	15	60	100	3	3
Sem-VII														
5	H1	RCP23CCH1701	Cloud Computing for Data Analysis	3			25	15	15	15	60	100	3	4
	H1	RCP23CLH1701	Cloud Computing for Data Analysis Laboratory			2	25				25	50	1	
			Total	15		6	200			75	375	650		18

Prepared by:

Ms. J. S. Sonawane

Checked by:

Mr. S. M. Pardeshi

Prof. Dr. R. B. Wagh
BOS Chairman

Prof. S. P. Shukla
C.O.E.

Prof. Dr. P. J. Deore
Dean Academics/Dy. Director

Prof. Dr. J. B. Patil
Director

Honors Program in Data Science offered by Computer Engineering (w.e.f. 2024-25)

Sr. No.	Course Category	Course Code	Course Title	Teaching Scheme			Evaluation Scheme							Credit
							Continuous Assessment (CA)					ESE	Total	
				L	T	P	TA	Term Test. 1 (TT1)	Term Test 2 (TT2)	Average of (TT1 & TT2)				
							[A]			[B]	[C]	[A+B+C]	
Sem-III														
1	H1	RCP23CCH1301	Statistics for Data Science	3			25	15	15	15	60	100	3	3

Prepared by:

Ms. J. S. Sonawane

Checked by:

Mr. S. M. Pardeshi

Prof. Dr. R. B. Wagh

BOS Chairman

Prof. S. P. Shukla

C.O.E.

Prof. Dr. P. J. Deore

Dean Academics/Dy. Director

Prof. Dr. J. B. Patil

Director

Semester - III

Statistics for Data Science (RCP23CCH1301)

Teaching Scheme	Examination Scheme
Lectures : 03 Hrs./week	Term Test : 15 Marks
Credits : 03	Teacher Assessment : 25 Marks
	End Sem Exam : 60 Marks
	Total Marks : 100 Marks

Prerequisite: Knowledge of Basics of probability.

Course Objectives:

1. To provide students with a solid understanding of statistical concepts and techniques essential for data science.

2. To develop students' ability to apply statistical methods to real-world data sets and draw meaningful insights.

3. To equip students with the skills to use statistical software tools for data analysis and interpretation.

4. To foster critical thinking and problem-solving skills in the context of statistical analysis for data science.

COs	Course Outcomes	Blooms Level	Blooms Description
CO1	Describe and summarize data using appropriate statistical measures.	L2	Understand
CO2	Apply correlation analysis and regression to real-world problems.	L3	Apply
CO3	Apply regression techniques to model and forecast time series data.	L3	Apply
CO4	Apply nonparametric test and draw appropriate conclusions for data analysis.	L3	Apply

1

Course Contents

Unit-I Introduction to Data and Statistics   08 Hrs.

Elements, Variables, and Observations, Scales of Measurement, Qualitative and Quantitative Data, Cross-Sectional and Time Series Data, Descriptive Statistics, Statistical Inference, Summarizing Qualitative and Quantitative Data using Tables and Graphs, Cross Tabulations and Scatter Diagram, Measures of Central Tendency, Measures of Dispersion, Skewness, Moments and Kurtosis

Unit-II Correlation Analysis   06 Hrs.

Types of correlation, Karl Pearson's Coefficient of Correlation, Coefficient of Determination, Rank Correlation Coefficient, Coefficient of Concurrent Deviation

Unit-III Regression   08 Hrs.

Simple Linear Regression: Concept of linear relationship between variables Least squares method, Estimating the coefficients (slope and intercept), Assessing the goodness of fit (R-squared), Interpretation of coefficients

Multiple Linear Regression: Extending simple linear regression to multiple predictors, Estimating the coefficients, Interpreting the coefficients, Assessing the model fit (adjusted R-squared), Multicollinearity and its implications, Heteroscedaticity.

Logistic Regression: Modeling binary outcomes, Odds ratios and log odds, Interpreting the coefficients, Assessing model fit (deviance, likelihood ratio test).

Unit-IV Hypothesis Testing in Regression   04 Hrs.

Significance testing for regression coefficients, t-tests and p-values, Confidence intervals for coefficients, F-test for overall model significance.

Unit-V Regression for Time Series Data   05 Hrs.

Autocorrelation and its impact on regression, Autoregressive models (AR), Moving average models (MA), Autoregressive integrated moving average models (ARIMA).

Unit-VI Non-Parametric Statistics   08 Hrs.

Nonparametric Methods, Sign Test, Wilcoxon Signed-Rank Test, MannWhitney-Wilcoxon Test, Kruskal Wallis Test, Spearman's Rank Correlation, Runs Test.

Text Books:

1. Dr. S. P. Gupta, "Statistical Methods", 46th Edition, S. Chand and Sons, 2021.

2. James T. McClave, P. George Benson, Terry T Sincich, "Statistics for Business and Economics". 14th Edition, Pearson, 2021.

Reference Books:

1. Maurits Kaptein, Edwin van den Heuvel, "Statistics for Data Scientists: An Introduction to Probability, Statistics, and Data Analysis", Springer, 2022.

2. Peter Bruce, Andrew Bruce, "Practical Statistics for Data Scientists", O'Reilly, 2017.

3. C.B. Gupta & Vijay Gupta, "An Introduction to Statistical Methods", 23rd Edition, S. Chand and Sons, 2004.

Evaluation Scheme:

Theory :

Continuous Assessment (A):

Subject teacher will declare Teacher Assessment criteria at the start of semester.

Continuous Assessment (B):

Two term tests of 15 marks each will be conducted during the semester. Average of the marks scored in both the tests will be considered for final grading.

End Semester Examination (C):

1. Question paper based on the entire syllabus, summing up to 60 marks.

2. Total duration allotted for writing the paper is 2 hrs.