THIRD-YEAR OF BACHELOR OF COMPUTER SCIENCE REVISED SYLLABUS ACCORDING TO CBCS

COURSE TITLE: DATA SCIENCE

SEMESTER-V, W.E.F. 2021-2022

Recommended by the Board of Studies in Computer Science
And
Approved by the Academic Council
Devrukh Shikshan Prasarak Mandal's
Nya. Tatyasaheb Athalye Arts, Ved. S. R. Sapre Commerce, and
Vid. Dadasaheb Pitre Science College (Autonomous), Devrukh.
Tal. Sangameshwar, Dist. Ratnagiri-415804, Maharashtra,
India

Name of the Implementing Institute	:	Nya. Tatyasaheb Athalye Arts, Ved. S. R. Sapre Commerce, and Vid. Dadasaheb Pitre Science College (Autonomous), Devrukh. Tal. Sangameshwar, Dist. Ratnagiri-415804,
Name of the Parent University	:	University of Mumbai
Name of the Programme	:	Bachelor of Science
Name of the Department	:	Computer Science
Name of the Class	:	Third Year
Semester	:	Six
No. of Credits	:	03
Title of the Course	:	Data Science
Course Code	:	USCST66
Name of the Vertical	:	Elective II
Eligibility for Admission	:	Any 12th Pass seeking Admission to Degree Programme in adherence to Rules and Regulations of the University of Mumbai and Government of Maharashtra
Passing Marks	:	40%
Mode of Assessment	:	Formative and Summative
Level	:	UG
Pattern of Marks Distribution for TE and CIA	:	70:30
Status	:	CBCS
To be implemented from Academic Year	:	2021-2022
Ordinances /Regulations (if any)		

Syllabus for Third Year of Bachelor of Science in Computer Science

(With effect from the academic year 2021-2022)

SEMESTER-VI   **Paper No.– 6**

Course Title: Data Science   **No. of Credits - 03**

Type of Vertical: Elective II   **COURSE CODE: USCST66**

Learning Outcomes Based on BLOOM's Taxonomy:

After completing the course, the learner will be able to…		
Course Learning Outcome No.	Blooms Taxonomy	Course Learning Outcome
CO-01	Understand	After completion of this course, the students should be able to understand & comprehend the problem
CO-02	Remember	After completion of this course, the students should be able to define suitable statistical method to be adopted..

Syllabus for Third Year of Bachelor of Science in Computer Science

(With effect from the academic year 2021-2022)

SEMESTER-VI   **Paper No.–6**

Course Title: Data Science   **No. of Credits - 03**

Type of Vertical: Elective II   **COURSE CODE: USCST66**

COURSE CONTENT			
Unit No.	**Content**	**Credits**	**No. of Lectures**
I	Introduction to Data Science: What is Data? Different kinds of data, Introduction to high level programming language + Integrated Development Environment (IDE), Exploratory Data Analysis (EDA) + Data Visualization, Different types of data sources, Data Management: Data Collection, Data cleaning/extraction, Data analysis & Modeling.	01	15
II	Data Curation: Query languages and Operations to specify and transform data, Structured/schema based systems as users and acquirers of data Semi-structured systems as users and acquirers of data, Unstructured systems in the acquisition and structuring of data, Security and ethical considerations in relation to authenticating and authorizing access to data on remote systems, Software development tools, Large scale data systems, Amazon Web Services (AWS)	01	15
III	Statistical Modelling and Machine Learning: Introduction to model selection: Regularization, bias/variance tradeoff e.g. parsimony, AIC, BIC, Cross validation, Ridge regressions and penalized regression e.g. LASSO Data transformations: Dimension reduction, Feature extraction, Smoothing and aggregating	01	15

	Supervised Learning: Regression, linear models, Regression trees, Time-series Analysis, Forecasting, Classification: classification trees, Logistic regression, separating hyperplanes, k-NN Unsupervised Learning: Principal Components Analysis (PCA), k-means clustering, Hierarchical clustering, Ensemble methods		
	Total	03	45

Required Previous Knowledge

Students should know basic concepts related to computer and computer handling

Access to the Course

The course is available for all the students admitted for Bachelor of Science (Computer Science).

Forms of Assessment

The assessment of the course will be of Diagnostic, Formative and Summative type. At the beginning of the course diagnostic assessment will be carried out. The formative assessment will be used for the Continuous Internal Evaluation whereas the summative assessment will be conducted at the end of the term. The weightage for formative and summative assessment will be 60:40. The detailed pattern is as given below.

Semester End Evaluation (70 Marks)
Question Paper Pattern
Time: 2:30 hours

Question No.	Unit/s	Question Pattern	Marks
Q.1	I ,II &III	MCQ/Fill in the blanks/One line sentence	10
Q.2	I	Descriptive Questions	20
Q.3	II	Descriptive Questions	20
Q4.	III	Descriptive Questions	20
		Total	**70**

Internal evaluation (30 Marks)

Sr. No.	Description	Marks
1	Classroom Tests	10
2	Project/ Viva/ Presentations/ Assignments	10
3	Attendance	10
	Total	**30**

Grading Scale

10 points grading scale will be used. The grading scale used is O to F. Grade O is the highest passing grade on the grading scale, and grade F is a fail. The Board of Examinations of the college reserves the right to change the grading scale.

Reference book:
- Doing Data Science, Rachel Schutt and Cathy O'Neil, O'Reilly,2013
- Mastering Machine Learning with R, Cory Lesmeister, PACKT
Publication,2015
Text book:
- Techmax publication book

Additional References:
- Hands-On Programming with R, Garrett Grolemund,1st Edition, 2014
- An Introduction to Statistical Learning, James, G., Witten, D., Hastie, T., Tibshirani,R.,Springer,2015)

Course: USCSP69	Practical of USCST66 (Credits : 1, Lectures/Week: 3)
USCSP69	Practical shall be performed using R 1. Practical of Data collection, Data curation and management for Unstructured data (NoSQL) 2. Practical of Data collection, Data curation and management for Large-scale Data system (such as MongoDB) 3. Practical of Principal Component Analysis 4. Practical of Clustering 5. Practical of Time-series forecasting 6. Practical of Simple/Multiple Linear Regression 7. Practical of Logistics Regression 8. Practical of Hypothesis testing 9. Practical of Analysis of Variance 10. Practical of Decision Tree