D. Y. Patil College of Engineering and Technology

Kasaba Bawada, Kolhapur

(An Autonomous Institute)

Accredited by NAAC with 'A' Grade

Structure and Syllabus of T. Y. B. Tech in Computer Science and Engineering (Data Science) 2022-23

D. Y. PATIL COLLEGE OF ENGINEERING & TECHNOLOGY, KOLHAPUR

Teaching and Evaluation Scheme from Year 2022-23

Third Year B. Tech. Computer Science & Engineering (Data Science)

SEMESTER-V

Sr. No.	Course Code	Course Type	Name of the Course	Teaching Scheme per Week				Total Marks	Evaluation Scheme			
				Lecture	Tutorial	Practical	Credits		Type	Max. Marks	Min. Marks for Passing	
1	201DSL301	PCC	System Programming and Compilers	3	-	-	3	100	**ISE**	20	20	40
									MSE	30		
									ESE	50	20	
2	201DSL302	PCC	Exploratory Data Analysis and Visualization	3	-	-	3	100	**ISE**	20	20	40
									MSE	30		
									ESE	50	20	
3	201DSL303	PCC	Introduction to Machine Learning	3	-	-	3	100	**ISE**	20	20	40
									MSE	30		
									ESE	50	20	
4	201DSL304	PCC	Database Engineering	3	-	-	3	100	**ISE**	20	20	40
									MSE	30		
									ESE	50	20	
5	201DSL305	ESC	Software Engineering	3	-	-	3	100	**ISE**	20	20	40
									MSE	30		
									ESE	50	20	
6	201DSP306	PCC	Java Programming	2	-	2	3	50	**ISE**	25	10	20
									POE	25	10	
7	201DSP307	PCC	EDA and Visualization Laboratory	-	-	2	1	50	**ISE**	25	10	20
									POE	25	10	
8	201DSP308	PCC	Machine Learning Laboratory	-	-	2	1	25	**ISE**	25	10	10
9	201DSP309	PCC	Database Engineering Laboratory	-	-	2	1	50	**ISE**	25	10	20
									POE	25	10	
10	201DSMC310	MC	Economics and Management for IT(Mandatory Course-II)	2	-	-	-	50	**ESE**	50	20	20
	Total			**19**	**0**	**8**	**21**	**725**	**-**	**-**	**-**	**290**

ISE: In Semester Evaluation MSE: Mid Semester Examination ESE: End Semester Examination

Note 1: Tutorials and practical shall be conducted in batches with batch strength not exceeding 15 students.

D. Y. PATIL COLLEGE OF ENGINEERING & TECHNOLOGY, KOLHAPUR

Teaching and Evaluation Scheme from Year 2022-23

Third Year B. Tech. Computer Science & Engineering (Data Science)

SEMESTER-VI

Sr. No.	Course Code	Course Type	Name of the Course	Teaching Scheme per Week				Total Marks	Evaluation Scheme			
				Lecture	Tutorial	Practical	Credits		Type	Max. Marks	Min. Marks for Passing	
1	201DSL311	PCC	Advanced Database Systems	3	-	-	3	100	ISE	20	20	40
									MSE	30		
									ESE	50	20	
2	201DSL312	PCC	Optimization for DataScience	3	-	-	3	100	ISE	20	20	40
									MSE	30		
									ESE	50	20	
3	201DSL313	PCC	Information Security	3	-	-	3	100	ISE	20	20	40
									MSE	30		
									ESE	50	20	
4	201DSL3PX	PEC	Professional Elective-I	3	1	-	4	100	ISE	20	20	40
									MSE	30		
									ESE	50	20	
5	201DSL3OX	OEC	Open Elective-I	3	1	-	4	100	ISE	20	20	40
									MSE	30		
									ESE	50	20	
6	201DSP319	PCC	Advanced Database System Laboratory	-	-	2	1	75	ISE	25	10	30
									POE	50	20	
7	201DSP320	PROJ	Project-II	-	-	4	2	75	ISE	25	10	30
									POE	50	20	
8	201DSMC321	MC	Human Values and Ethics (Mandatory Course-III)	2	-	-	-	50	ESE	50	20	20
Total				**17**	**2**	**6**	**20**	**700**				**280**

Professional Elective I

201DSL3P14 Block Chain Technology
201DSL3P15 High Performance Computing.
201DSL3P16 E-Commerce and Recommendation System

Open Elective-I

- **List Attached**

Open Elective:

Open elective courses are offered to gain the knowledge of multidisciplinary areas. Students must choose one open elective course from the list of courses offered by other departments (excluding open elective courses offered by their department). Following is the list of open elective courses. The detailed syllabus is available on to the college website under academic tab.

Sr. No.	Department	Course Code	Open Elective-I Course
1	Chemical	201CHL318	Industrial Safety and Act
		201CHL319	Energy Conservation and Audit
2	Mechanical	201MEL313	Human Resource Management
		201MEL314	Electric Vehicle
3	Civil	201CEL330	Disaster Management
		201CEL331	Green Building
4	Architecture	201ARL318	Residential Gardening
		201ARL319	Role of Art & Technology in Interior Design
5	Electronics and Telecommunication	201ETL314	Sensor Technology
		201ETL315	Electronic Instrumentation
6	Computer Science & Engineering (Artificial Intelligent & Machine Learning)	201AIML320	Applications of AI ML
		201AIML321	Augmented Reality and Virtual Reality
7	Computer Science & Engineering (Data Science)	201DSL3O17	Basics of Data Science
		201DSL3O18	Basics of Database

DR. G. V. Patil
HOD (Data Science)

DR. B. D. Jitkar
Dean Academics

DR. S. D. Chede
Principal

SEMESTER –V

D.Y. PATIL COLLEGE OF ENGINEERING & TECHNOLOGY KASABA BAWADA KOLHAPUR-416006

(An Autonomous Institute) B. Tech. Curriculum

T. Y. B. Tech. CSE (Data Science) Sem. V

(AcademicYear-2022-23)

Course Plan

Course Title: System Programming and Compilers	
Course Code: 201DSL301	**Semester: V**
Teaching Scheme: L-T-P: 3-0-0	**Credits: 3**
Evaluation Scheme: ISE+MSE Marks: 20+30	**ESE Marks: 50**

Course Description:

The compiler is the programmer's primary tool. Understanding the compiler is therefore critical for programmers, even if they never build one. This course introduces students to the essential elements of building a compiler. It deals with the basic concepts of system programs as well as provides deeper insights into Compiler and its phases. This course will help students to learn about lexical analysis, parsing, semantic analysis, intermediate code generation, code optimization, and code generation.

Course Objectives:

1. To expose the students to the fundamentals of various system programs.
2. To introduce the fundamentals of Compiler and its phases.
3. To expose the students to various compiler construction tools.

Course Outcomes (COs):

Upon successful completion of this course, the students will be able to:

C301.1	Understand the basics of system programs, Assemblers, Macros, Linkers, Loaders	Understand
C301.2	Recall the compiler phases and compiler construction tools.	Understand
C301.3	Learn Lexical analysis and various parsing techniques.	Apply
C301.4	Understand syntax-directed translation, intermediate code generation, and target code generation.	Understand
C301.5	Identify and apply appropriate code optimizing transformation for given code.	Apply

Prerequisite:	Theory of Computation, Assembly language.

Course Articulation Matrix: Mapping of Course Outcomes (COs) with Program Outcomes(POs) and Program Specific Outcomes (PSOs):

COs	POs												PSOs		BTL
	1	2	3	4	5	6	7	8	9	10	11	12	1	2	
C301.1	2											1	1		2
C301.2	2											1	1		2
C301.3	2	1	1									1	1		3
C301.4	1														2
C301.5	2	2													3

Content	Hours
Unit 1: Language Processors: Introduction, Language processing activities, Fundamentals of language Processing, Language processor development tools: LEX, YACC. Compiler construction tools, cousins of the compiler.	5
Unit 2: Assemblers, Linkers, and Loaders: Elements of assembly language programming, A simple assembly scheme, pass structure of assemblers, and Design of a two-pass assembler. **Macros and Macro Pre-Processors**: Macro definition and call, Macro expansion, Nested macro calls. **Linkers & Loaders**: Introduction.	6
Unit 3: Compilers: Phases of a compiler, Role of a Lexical analyzer, input buffering, specification and recognition of tokens, finite automata implications.	5
Unit 4: Syntax Analysis: Role of Parser, Recursive descent and predictive parsers (LL), Operator precedence parsing, Working of LR Parser and introduction to its types SLR, canonical LR, and LALR.	7
Unit 5: Syntax Directed Translation and Intermediate Code Generation: Syntax-directed definitions, construction of syntax tree, S-attributed definitions, L-attributed definitions, Intermediate languages, assignment statements, back patching.	6

Unit 6: Code Generation and Code Optimization: Issues in the design of a code generator and target machine, Basic blocks and flow graphs, Next use information and simple code generator, Issues of register allocation, Principal sources of optimization, optimization of Basic Blocks, Peephole optimization, example on how to optimize python code for data science.	7

Text Books:

1. System Programming and operating systems, D. M. Dhamdhere, 2nd Edition (TMGH) (Unit 1,2).
2. Compilers - Principles, Techniques, and Tools A. V. Aho, R. Sethi and J. D. Ullman Pearson Education (Unit 3,4,5,6).

Reference Books:

1. Compiler Construction 2nd Edition, Dhamdhere, Binding Paperback Publisher Macmillan 2003.

Online Resources:

1. https://onlinecourses.nptel.ac.in/noc20_cs13/preview
2. https://nptel.ac.in/courses/106108113
3. https://nptel.ac.in/courses/106105190
4. https://www.coursera.org/lecture/nand2tetris2/unit-4-1-syntax-analysis-5pC2Z
5. https://www.analyticsvidhya.com/blog/2019/09/4-methods-optimize-python-code-data-science/

Course Plan

Course Title: Exploratory Data Analysis and Visualization	
Course Code :201DSL302	Semester :V
Teaching Scheme : L-T-P : 3-0-0	Credits :3
Evaluation Scheme: ISE+MSE Marks: 20+30	ESE Marks : 50

Course Description: Exploratory Data Analysis (EDA) is an approach to data analysis that involves the application of diverse techniques to gain insights into a dataset. The course contains fundamental concept of Exploratory data analysis. The focus is given on analysis and illustration of data.

Course Objectives:

1. Understand the fundamental concepts of exploratory data analysis using Python.
2. Find missing values in data and identify the correlation between different variables.
3. Understand the hypothesis testing and advance data visualization tools.

Course Outcomes (COs):

Upon successful completion of this course, the students will be able to:

C302.1	Import, clean, and explore data to perform preliminary analysis	Apply
C302.2	Identify and transform erroneous data using different data wrangling techniques	Analyzing
C302.3	Understand and interpret results obtained from graphical analysis	Understand
C302.4	Explore hypothesis testing and explore techniques of time-series analysis	Apply

Prerequisite:	Python, Fundamentals of Data Science

Course Articulation Matrix: Mapping of Course Outcomes (COs) with Program Outcomes (POs) and Program Specific Outcomes (PSOs):

COs	POs												PSOs		BTL
	1	2	3	4	5	6	7	8	9	10	11	12	1	2	
C302.1	2	-	-	-	-	-	-	-	-	-	-	-	-	-	3
C302.2	3	2	2	-	2	-	-	-	-	-	-	2	1	-	4
C302.3	3	-	2	2	2	-	-	-	-	-	-	2	1	-	2
C302.4	3	2	2	-	-	-	-	-	-	-	-	-	-	-	3

Content	Hours
Unit 1: Exploratory Data Analysis Fundamentals: Understanding data science, Steps in EDA, Numerical data, categorical data, measurement scale, software tools for EDA.	4
Unit 2: Visual Aids For EDA: Study of different charts and plots, EDA with personal email-loading dataset, data transformation, data analysis.	8
Unit 3: Descriptive Statistics: Understanding statistics, Measures of central tendency:- mean, mode, median, Measures of dispersion:- Standard deviation, variance, skewness, kurtosis, percentile, quartile.	6
Unit 4: Grouping dataset: Understanding groupby(), groupby mechanics, data aggregation-groupwise operation, groupwise transformation, pivot tables and cross tabulation	6
Unit 5: Correlation and Time series analysis: Types of analysis, Multivariate analysis using Titanic dataset, Simpsons Paradox, correlation vs. causation, fundamentals of TSA, Characteristic of Time series data, TSA with Open Power System Data.	6
Unit 6: Hypothesis Testing and advance data visualization tool: Hypothesis testing, p-hacking, Understanding regression, Types of regression, Model development and evaluation, Inferential Statistics. **Advance data visualization tool: Tableau-**Connecting to data, Creating Sheets and Dashboards, Publishing to the web	6

Text Books:

1. "Hands-On Exploratory Data Analysis with Python", Suresh Kumar Mukhiya, Usman Ahmed, Packt Publication.

Reference Books:

1. "Python for Data Analysis", Wes Mckinney, O'REILLY Publication. 2017

Online Resources:

1) https://www.coursera.org/lecture/data-analysis-with-python/exploratory-data-analysis-iNeWs

Course Plan

Course Title: Introduction to Machine Learning	
Course Code: 201DSL303	**Semester: V**
Teaching Scheme: L-T-P: 3-0-0	**Credits: 3**
Evaluation Scheme: ISE+MSE Marks: 20+30	**ESE Marks: 50**

Course Description:
This course is an introduction to the theoretical aspects of design and implementation of algorithms that enable machines to "learn" from experience. Course will provide in-depth knowledge to the areas of Supervised and Unsupervised Machine Learning. The course cover core Machine Learning algorithms for classification, regression and clustering.

Course Objectives:
1. To understand pattern classification algorithms to classify multivariate data.
2. To understand and implement standard machine learning Algorithms.
3. To understand text mining and its applications.

Course Outcomes (COs):

Upon successful completion of this course, the students will be able to:

C303.1	Understand the paradigms of supervised and unsupervised machine learning.	Understand
C303.2	Identify the strengths and weaknesses of multiple machine learning approaches.	Understand
C303.3	Formalize a task as a machine learning problem.	Apply
C303.4	Identify suitable algorithms to tackle different machine learning problems.	Understand

Prerequisite:	Linear Algebra, Statistics, Probability theory.

Course Articulation Matrix: Mapping of Course Outcomes (COs) with Program Outcomes (POs) and Program Specific Outcomes (PSOs):

COs	POs												PSOs		BTL
	1	2	3	4	5	6	7	8	9	10	11	12	1	2	
C303.1	1	2	1	-	-	-	-	-	-	-	-	2	-	-	2
C303.2	2	2	2	-	-	-	-	-	-	-	-	2	-	-	2
C303.3	2	3	2	2	1	-	-	-	-	-	-	2	2	-	3
C 303.4	2	3	2	2	1	-	-	-	-	-	-	2	2	-	2

D.Y. PATIL COLLEGE OF ENGINEERING & TECHNOLOGY KASABA BAWADA KOLHAPUR-416006

(An Autonomous Institute) B. Tech. Curriculum

T. Y. B. Tech. CSE (Data Science) Sem. V

(AcademicYear-2022-23)

Content	Hours
Unit 1: Introduction to Machine Learning: Machine Learning: Definition, Terminology, Types of learning, Machine Learning Problem categories, Machine process, Lifecycle, Performance measures, tools and framework and data visualization.	06
Unit 2: Regression: Simple regression – Hypothesis, cost function, parameter learning with Gradient Descent, learning rate, Gradient Descent for linear regression. **Multivariate Linear Regression** – Multiple features, Hypothesis functions, Gradient Descent for multiple variables and Feature scaling.	06
Unit 3: Classification- Logistic Regression: Definition, Hypothesis representation, decision boundary, cost function, Gradient Descent for Logistic Regression. Multiclass Classification, **Regularization** - Over fitting & Under fitting, cost function, Regularized Linear Regression, Regularized Logistic Regression and K- Nearest Neighbors Classifier.	05
Unit 4: Bayesian Learning: Introduction to Bayes theorem, Naïve Bayes theorem, Naïve Bayes Classifier, Bayesian Belief Networks, Introduction to Hidden Markov Model and Issues in Hidden Markov Model.	06
Unit 5: **Decision Trees and SVM:** Definition, Terminology, the need, Advantages, and Limitations. Constructing and Understanding Decision Trees, common problems with Decision Trees, Decision Tree algorithms, random forest and examples. Introduction to Support Vector Machines, Linear Support Vector Machines soft margin SVM, hard margin SVM, Kernel Tricks, Primal and Dual Form, Cost Function.	07
Unit 6: Clustering: Introduction to Clustering, Types of Clustering, Partitioning Methods of Clustering, Hierarchical Methods. Introduction to text mining, Methods and techniques of text mining, Application of text mining.	06

Text Books:

1. Machine Learning, Anuradha Srinivasara ghavan, and Vincy Joseph, Kindle Edition, 2020, WILEY.
2. Introduction to Machine Learning, Ethem Alpaydin, Second Edition, 2010, Prentice Hall of India.
3. Practical Machine Learning Sunila Gollapudi Packt Publishing Ltd.
4. Fundamentals of Data Science, BySanjeev J. Wagh, Manisha S. Bhende, Anuradha D. Thakare, 1st Edition, New York.

Reference Books:

1. Machine Learning by Tom M. Mitchell, International Edition 1997, McGraw Hill Education.
2. Data mining and analysis Mohammed J.ZakiWagnerMeiraJr.,Cambridge UniversityPress.

Online Resources:

1. https://livebook.manning.com/book/machine-learning-in-action/about-this-book/ https://www.coursera.org/learn/machine-learning
2. https://nptel.ac.in/courses/106106139

D.Y. PATIL COLLEGE OF ENGINEERING & TECHNOLOGY KASABA BAWADA KOLHAPUR-416006

(An Autonomous Institute) B. Tech. Curriculum

T. Y. B. Tech. CSE (Data Science) Sem. V

(AcademicYear-2022-23)

Course Plan

Course Title: Database Engineering	
Course Code:201DSL304	**Semester: V**
Teaching Scheme: L-T-P: 3-0-0	**Credits: 3**
Evaluation Scheme: ISE+MSE Marks: 20+30	**ESE Marks: 50**

Course Description:

This course introduces the fundamental concepts, principles and tools of database system. The course includes relational data model and languages, database design techniques, SQL, data storage and indexing techniques. Also, the focus is given on concurrency control and recovery techniques.

Course Objectives:

1. To understand fundamental concepts of database systems.
2. To gain familiarity with SQL, PLSQL and DBMS.
3. To learn database design techniques.
4. To understand indexing, transaction management, recovery and security techniques.

Course Outcomes (COs):

Upon successful completion of this course, the students will be able to:

C304.1	Understand fundamental concepts of database systems.	Understand
C304.2	Construct logical design of database using E-R Diagram.	Evaluating
C304.3	Study and apply SQL queries , PLSQL procedures to design & manage the database.	Apply
C304.4	Analyze & construct good database design.	Analyzing
C304.5	Understand transaction concepts and concurrency control techniques.	Apply
C304.6	Understand failures in database, appropriate recovery and security techniques.	Understand

Prerequisite:	Set Theory and Data Structures

Course Articulation Matrix: Mapping of Course Outcomes (COs) with Program Outcomes (POs) and Program Specific Outcomes (PSOs):

COs	**POs**												**PSOs**		**BTL**
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**1**	**2**	
C304.1	2	2													2
C304.2		2	2												5
C304.3		3	3	1								2	2		3
C304.4		2	2									2	2		4
C304.5	2	2													3
C304.6	2		2												2

Content	Hours
Unit 1: Introduction to databases and E-R model Purpose of Database Systems, View of data, Database architecture, Database users and administrator, E-R model**:** Entity sets, Relationship sets, Mapping Constraints, Keys, E-R Diagram, Reducing E-R Diagrams to relational schemas, Extended E-R features: Specialization, Generalization, and Aggregation	6
Unit 2: Relational Model, SQL and PLSQL Relational Model: Structure of Relational Databases, Database Schema, Keys, Schema Diagram, Relational Algebra. SQL: Overview of the SQL Query Language, Basic Structure of SQL Queries, Additional Basic Operations, Set Operations, Aggregate Functions, Nested Sub queries, Modification of the Database, Join Expressions, Views. PLSQL: Triggers, Stored Procedures, PL/SQL Processing with Cursors, PL/SQL Stored Functions	9
Unit 3: Relational Database Design Referential Integrity, features of good relational designs, functional dependency, closure of a set of functional dependencies and Canonical cover. Normalization: Purpose of normalization, First Normal Form (1NF), Second Normal Form (2NF), Third Normal Form(3NF), Boyce-Codd Normal Form(BCNF).	7
Unit 4: Data Storage and Indexing Storage and File structure: Overview of physical storage media, RAID, File Organization, Organization of Records in Files, Data Dictionary Storage, Database Buffer. Indexing and Hashing: Basic Concepts, Ordered Indices, B+ Tree Index Files, Multiple Key Access. Static Hashing, Dynamic Hashing, Index definition in SQL.	8
Unit 5: Concurrency Control and Crash Recovery Transaction concept, Transaction state, Concurrent Executions, Serializability, Recoverability, testing for Serializability, Lock-Based Protocols, Graph based Protocols, Timestamp Based Protocols, Validation based protocols, Failure Classification, Recovery and Atomicity, Log-Based Recovery, Checkpoints, Shadow Paging, Buffer Management	7
Unit 6: Database Security and Authorization: Introduction to Database Security Issues, Discretionary Access Control Based on Granting and Revoking Privileges, Mandatory Access Control and Role- Based Access Control for Multilevel Security, Introduction to Statistical Database Security	3

Text Books:

1. A. Silberschatz, H.F. Korth, S. Sudarshan, "Database System Concepts", 6th Edition, McGraw Hill Education. (Unit 1,2, 3,4,5)
2. Thomos Connolly, Carolyn Begg, "Database Systems - A practical approach to Design, Implementation and Management", 3rd Edition, Pearson Education. (Unit 3- Normalization)
3. Coronel, Morris, Rob, "Database Systems, Design, Implementation and Management", Ninth Edition, Cengage Learning, (Unit 2- PLSQL)
4. Ramez Elmasri and Shamkant Navathe, "Fundamentals of Database Systems", Pearson Education,Fifth Edition (Unit 6)

Reference Books:

1. Raghu Ramkrishnan, Johannes Gehrke, "Database Management System", Fourth Edition, McGrawHill Education.

Online Resources:

1. https://onlinecourses.nptel.ac.in/noc19_cs46/preview

Course Plan

Course Title: Software Engineering	
Course Code:201DSL305	Semester: V
Teaching Scheme: L-T-P: 3-0-0	Credits: 3
Evaluation Scheme: ISE+MSE Marks: 20+30	ESE Marks: 50

Course Description:

This course act as guidelines provided for software development. It commences the main methodologies of software engineering and Software Development Life Cycle (SDLC). It also contains exposure to different tools and models which play important role in SDLC.

Course Objectives:

1. To expose the students to basic concepts and principles of software engineering.
2. To make the students aware of the importance of SDLC in their project development.
3. To expose the students to agile processes.

Course Outcomes (COs):

Upon successful completion of this course, the students will be able to:

C305.1	Understand systematic methodologies of SDLC.	Understand
C305.2	State SRS for their problem domain.	Remembering
C305.3	Use UML for Object Oriented Modelling.	Apply
C305.4	Understand coding, testing methods and importance of software maintenance.	Understand
C305.5	Understand in detail agile processes.	Understand

Prerequisite:	C and C++ Languages

Course Articulation Matrix: Mapping of Course Outcomes (COs) with Program Outcomes (POs) and Program Specific Outcomes (PSOs)

COs	POs												PSOs		BTL
	1	2	3	4	5	6	7	8	9	10	11	12	1	2	
C305.1	2	-	-	-	-	-	-	-	-	-	-	-	-	-	2
C305.2	2	2	-	-	-	-	-	-	-	-	-	2	-	-	1
C305.3	2	2	2	-	-	-	-	-	-	-	-	2	-	-	3
C305.4	2	-	-	-	-	-	-	-	-	-	-	-	-	-	2
C305.5	2	-	-	-	-	-	-	-	-	-	-	-	-	-	2

Content	Hours
Unit 1. Introduction to Software and Processes Software Problem- Cost, Schedule & Quality, Scale and change, Software Processes- Process and Project, Component Software Processes, SDLC, Software Development Process Modules, Project Management Process. Agile Development- XP, other Agile Process Models, Tool Set For Agile Process.	6
Unit 2: Software Requirements Analysis Introduction to Requirements Engineering. Value of a good SRS, Requirements Process, Requirements Specifications, Other Approaches for Analysis, Validation. Case study on Software requirements.	5
Unit 3: Software Planning & Risk Management Responsibilities of Software Project Manager, Project Planning, Project Scheduling, Project Staffing, People CMM, Risk Management, Case study on COCOMO.	6
Unit 4: Software Design Basics of Software Design, Function Oriented Design - DFD, Object Oriented Design Class diagram, Sequence & collaboration diagram, Detailed Design, Case Study for Software Design.	6
Unit 5: Object Modelling Using UML and OO Software Development Basic OO Concepts, UML, UML Diagrams, Use Case Model, Interaction Diagram, Activity Diagram, State Chart Diagram, Patterns and Common Design Patterns, OO Analysis and Design Methodology, Interaction Modelling. Hands-on for StarUML.	7
Unit 6: Coding & Testing Features of Software Code, Coding Guidelines, Coding Methodology, Programming Practice, Code Verification Techniques, Coding Tools, Code Documentation, Software Testing Basics, Test plan, Test case Design, Software Testing Strategies, Level of Testing, Testing Techniques, Software testing Tools, Debugging, Software Test Report, and Introduction to Software deployment & Maintenance. Hands-on for Selenium open source tool.	6

Text Books:

1. Pankaj Jalote, "Software Engineering- A precise approach", Wiley India, (Unit 1,2,4)
2. Roger S. Pressman, "Software Engineering – A Practitioner's Approach", 7th Edition, Mc Graw Hill, (Unit 1,2)
3. Rohit Khurana, "Software Engineering Principles and Practices", Vikas Publication.
4. Ugrasen Suman, "Software Engineering concept & Practices", CENANGE Learning, (Unit 6)
5. Rajib Mall, "Fundamentals of Software Engineering", PHI, Third Edition (Unit5)
6. Grady Booch, James Rambaugh, Lvar Jacobson ,"The Unified Modeling Language User Guide", Addison Wesley (Unit 5).

Reference Books:

1. Hansvan Vliet, "Software Engineering Principles and Practice", Willey-India Edition.
2. Sommer Ville, "Software Engineering", Pearson Education, India.
3. P Fleeger, "Software Engineering", Pearson Education, India.

Online Resources:

1. https://onlinecourses.nptel.ac.in/noc21_cs13/course

- **ISE 1: -**

1. Each student must be given a case study to write SRS.
2. Each student must study, design and draw diagram for given case study as in 1 above using appropriate Object Oriented Design Tool.

- **ISE 2: -**

1. Students must be asked to study, demonstrate and prepare report on tools of following types
   a. Code Editing Tools
      Notepad++, Sublime Text, Jedit, Atom, Visual Studio Code, Brackets, NetBeans, Eclipe, Bluefish,VIM.
   b. Android development Tools
      Android Studio, Android Asset Studio, AIDE, PhoneGap, Flutter, Stetho.
   c. API Tools
      Swagger, Postman, REST, SOAP.
   d. Testing Related Tools
      Jmeter, Seleniom, Cucumber, TestNG, JUnit.
   e. Repository Maintaining Tools
      Tortoise, Git, GitHub,Bitbucket.

Course Plan

Course Title :Java Programming	
Course Code :201DSP306	**Semester :V**
Teaching Scheme : L-T-P: 2-0-2	**Credits :3**
Evaluation Scheme : ISE Marks : 25	**ESE-POE Marks :25**

Course Description:

This course provides the object-oriented approach using Java programming constructs. The course includesbasics of Java language programming, the different object-oriented features, packages, file handling and multithreading. This course enables the students to develop the GUI based applications using advanced features such as swing, database handling, networking and collection. This course provides the basics for developing android applications, games, and many more programming language applications in the differentfields.

Course Objective:

1. To introduce the concepts of object-oriented programming using JAVA programming constructs.
2. To expose the students with the JAVA concepts using inheritance, interface, package, I/O and exception handling mechanisms.
3. To develop the problem-solving ability using GUI designing components.
4. To build the foundations of advanced java programming for application development.

Course Outcomes (COs):

At the end of the course the student should be able to:

C306.1	Use the java programming concepts for solving the problems with object-oriented approach.	Apply
C306.2	Develop the reliable and user -friendly application using inheritance, interface, package, I/O and exception handling mechanisms.	Apply
C306.3	Create the applications using the GUI designing components with the use of modern tools.	Apply
C306.4	Apply the knowledge of the advanced java programming concepts for developing the applications from different domains.	Apply

Course Articulation Matrix

Mapping of Course Outcomes (COs) with Program Outcomes (POs)

COs	POs												PSOs		BTL
	1	2	3	4	5	6	7	8	9	10	11	12	1	2	
C306.1	2	-	-	-	-	-	-	-	-	-	-	-	-	-	3
C306.2	2	2	2	-	-	-	-	-	-	-	-	-	-	-	3
C306.3	2	2	3	-	1	-	-	-	2	-	-	-	2	-	3
C306.4	3	2	2	-	-	-	-	-	2	-	-	-	1	-	3

Content	Hours
Unit 1: Fundamental Programming in Java: The Java Buzzwords, The Java Programming Environment- JVM, JIT Compiler, Byte Code Concept, A Simple Java Program, Source File Declaration Rules, Comments, Data Types, Variables, Operators, Strings, Input and Output, Control Flow, Array. **Objects and Classes:** Object-Oriented Programming Concepts, Declaring Classes, Declaring Member Variables, Defining Methods, Constructor, Passing Information to a Method or a Constructor, Creating and using objects, Controlling Access to Class Members, Static Fields and Methods, this keyword	3
Unit 2: Inheritance, Interface and Package: **Inheritance:** Definition, Super classes, and Subclasses, Overriding and Hiding Methods, Polymorphism, Inheritance Hierarchies, Super keyword, Final Classes and Methods, Abstract Classes and Methods, casting, Design Hints for Inheritance, Inner Classes, garbage collection**.** **Interfaces:** Defining an Interface, implementing an Interface, using an Interface as a Type, Evolving Interfaces, and Default Methods. **Packages:** Class importing, creating a Package, naming a Package, Using Package Members, Managing Source and Class Files.	4
Unit 3: Exception and I/O Streams **Exception:** Definition, dealing with Errors, The Classification of Exceptions, Declaring Checked Exceptions, Throw an Exception, Creating Exception Classes, Catching Exceptions, Catching Multiple Exceptions, Re-throwing and Chaining Exceptions, finally clause. **I/O Streams:** Byte Stream – Input Stream, Output Stream, Data Input Stream, Data Output Stream, File Input Stream, File Output Stream, Character Streams, Buffered Stream, Scanner class	4

Unit 4: Graphical User Interfaces using AWT and Swing **Introduction to AWT components**, Frame, Applet, Introduction to the Swing, Swing components. **Layout Management:** Introduction to Layout Management, APIs for Border Layout, Flow Layout, Grid Layout **Event Handling:** Basics of Event Handling, The AWT Event Hierarchy, Semantic and Low-Level Events in the AWT, Low-Level Event Types	6
Unit 5: Multithreading, Collections **Multithreading:** Processes and Threads, Runnable Interface and Thread Class, Thread Objects, Defining and Starting a Thread, Pausing Execution with Sleep, Thread States, Thread Properties **Collections:** Collection Interfaces, Concrete Collections-List, Queue, Set, Map, the Collections Framework	4
Unit 6: Database Programming and Networking **Database Programming:** The Design of JDBC, The Structured Query Language, Basic JDBC Programming Concepts, Query Execution, Scrollable and Updatable Result Sets. **Networking:** Overview of Networking, Networking Basics, Sockets, reading from and Writing to a Socket, Writing the Server Side of a Socket	3

List of Experiments			
Experiment No.	**Name of Experiments**	**S/O**	**Hours**
1	Study of JAVA basics.	S	2
2	Implementation of a problem statement using class and object.	O	2
3	Design and develop the programs for different types of inheritance	O	2
4	Implementation of stack/queue operations using Interface	O	2
5	Implementation of user defined package.	O	2
6	Implementation of any type of Exception Handling	O	2
7	Implementation of different I/O operations using console and file.	O	2
8	Implementation of program for designing the GUI using swing components and demonstrating layout managers.	O	4
9	Implementation of different types of event handling.	O	2

10	Design and develop an application for demonstration of multithreading	O	2
11	Implementation of any program using collections.	O	2
12	Implementation of different database operations using JDBC	O	2
13	Develop any application using networking.	O	2
14	Design an application using any **modern tools available for java programming** such as Eclipse IDE, NetBeans, Oracle JDeveloper, IntelliJ etc.	O	2

S-STUDY, O-OPERATIONAL

Note: The instructor may choose minimum 8 to 10 assignments from assignment no. 1 to 13 & Assignment no. 14is mandatory

Text Books:

1. Herbert Schildt, JAVA-The Complete Reference, Mcgraw Hill, Ninth edition.

Reference Books:

1. Cay Horstmann and Gary Cornell Core Java- Volume I Fundamentals Pearson, Eight edition (Unit 1 to Unit 4).
2. Cay Horstmann and Gary Cornell, Core Java- Volume II Advanced Features ,Pearson,Eight edition(Unit 5 and Unit 6).

Online Resources –

1. https://nptel.ac.in/courses/106/105/106105191/
2. https://java-iitd.vlabs.ac.in/List%20of%20experiments.html

Course Plan

Course Title: Exploratory Data Analysis and Visualization Laboratory	
Course Code :201DSP307	**Semester: V**
Teaching Scheme: L-T-P: 0-0-2	**Credits :1**
Evaluation Scheme: ISE Marks : 25	**ESE-POE Marks :25**

Course Description: Exploratory Data Analysis (EDA) is an approach to data analysis that involves the application of diverse techniques to gain insights into a dataset. The course contains fundamental concept of exploratory data analysis. The focus is given on analysis and illustration of data.

Course Objectives:

1. Understand the fundamental concepts of EDA using Python and its libraries
2. Find missing values in data and identify the correlation between different variables
3. Understand different distributions and descriptive statistics

Course Outcomes (COs):

Upon successful completion of this course, the students will be able to:

C307.1	Import, clean, and explore data to perform preliminary analysis	Apply
C307.2	Identify and transform erroneous data using different data wrangling techniques	Analyzing
C307.3	Understand and interpret results obtained from graphical analysis	Understand
C307.4	Explore hypothesis testing and explore techniques of time-series analysis	Apply

Prerequisite:	Python Programming Language, Fundamentals of Data Science

Course Articulation Matrix: Mapping of Course Outcomes (COs) with Program Outcomes (POs) and Program Specific Outcomes (PSOs):

COs	**POs**												**PSOs**		**BTL**
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**1**	**2**	
C307.1	2	-	-	-	-	-	-	-	-	-	-	-	-	-	3
C307.2	3	2	2	-	2	-	-	-	-	-	-	2	1	-	4
C307.3	3	-	2	2	2	-	-	-	-	-	-	2	1	-	2
C307.4	3	2	2	-	-	-	-	-	-	-	-	-	-	-	3

List of A Experiments:

Sr. No.	Experiments	S/O
1	Program to get statistical characteristics of dataset using pandas.	O
2	Programs for analysis of data through different plots(scatter, bubble, area, stacked) and charts(line, bar, table, pie, histogram).	O
3	Implementation of data transformation-reshaping and deduplication of data.	O
4	Implementation of data transformation-Handling missing data ,filling missing data.	O
5	Program on Discretization and binning of data.	O
6	Program on Handing dummy variables.	O
7	Implementation of different distributions(normal, poisson, uniform, gamma).	O
8	Program on Data cleaning.	O
9	Implementation of descriptive statistics(variance, skewness, kurtosis, percentile).	O
10	Implementation of grouping and group by.	O
11	Implementation of hypothesis testing -T Test.	O
12	Create simple dashboard using tableau.	O

Text Books:

1. **"**Hands-On Exploratory Data Analysis with Python", Suresh Kumar Mukhiya, Usman Ahmed, Packet Publication.

Online Resources

1.https://www.tableau.com/learn/tutorials/on-demand/getting-started

Course Plan

Course Title: Machine Learning Laboratory.	
Course Code: 201DSP308	**Semester: V**
Teaching Scheme: L-T-P: 0-0-2	**Credits: 1**
Evaluation Scheme: ISE: 25	**ESE-POE Marks: -**

Course Description:

This course provides knowledge of Supervised and Unsupervised Machine Learning techniques and their implementation and also covers core Machine Learning algorithms like classification, regression, clustering.

Course Objectives:

1. To implement standard machine learning algorithms
2. To implement supervised and unsupervised learning techniques
3. To implement different text mining techniques.

Course Outcomes (COs):

C308.1	Understand the paradigms of supervised and unsupervised machine learning	Understand
C308.2	Identify the strengths and weaknesses of multiple machine learning approaches.	Understand
C308.3	Formalize a task as a machine learning problem.	Apply
C308.4	Identify suitable algorithms to tackle different machine learning problems.	Understand

Prerequisite:	Python Programming language

Course Articulation Matrix: Mapping of Course Outcomes (COs) with Program Outcomes (POs) and Program Specific Outcomes (PSOs):

COs	**POs**												**PSOs**		**BTL**
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**1**	**2**	
C308.1	1	2	1	-	-	-	-	-	-	-	-	2	-	-	2
C308.2	2	2	2	-	-	-	-	-	-	-	-	2	-	-	2
C308.3	2	3	2	2	1	-	-	-	-	-	-	2	2	-	3
C308.4	2	3	2	2	1	-	-	-	-	-	-	2	2	-	2

List of Experiments			
Exp. No.	**Name of Experiment**	**S/O**	**Hours**
1	Implementation of Linear Regression	O	2
2	Implementation of Multivariate Linear Regression	O	2
3	Implementation of Logistic Regression for Binary Classification	O	2
4	Implementation of Multiclass Classification	O	2
5	Implementation of KNN Classifier	O	2
6	Implementation of Naïve Bayes Classifier	O	2
7	Implementation of Bayesian Network	O	2
8	Implementation of Decision Tree	O	2
9	Implementation of SVM	O	2
10	Implementation of K-means Clustering	O	2
11	Implementation of Agglomerative Clustering	O	2
12	Design and analysis of sentiment analysis model	O	2

- **S-STUDY, O-OPERATIONAL**
- **Note:**

Text Books:

1. Machine Learning, Anuradha Srinivasaraghavan, and Vincy Joseph, Kindle Edition, 2020, WILEY.
2. Machine Learning – An Algorithmic Perspective by Stephen Marsland.

Reference Books:

1. Machine Learning by Tom M. Mitchell, International Edition 1997, McGraw Hill Education

Online Resources

1. https://livebook.manning.com/book/machine-learning-in-action/about-this-book/
2. https://www.coursera.org/learn/machine-learning
3. https://nptel.ac.in/courses/106106139

Course Plan

Course Title: Database Engineering Laboratory	
Course Code: 201DSP309	**Semester: V**
Teaching Scheme: L-T-P: 0-0-2	**Credits :1**
Evaluation Scheme: ISE Marks:25	**ESE-POE Marks: 25**

Course Description:

This course focuses on implementation of the fundamental concepts and principles of database engineering. Focus is given on hands-on practical's considering SQL-DDL, DML commands, database connectivity and implementation of views.

Course Objectives:

1. To demonstrate fundamental concepts of database systems.
2. To gain familiarity with SQL, PLSQL and DBMS.
3. To construct the database for a given application.

Course Outcomes (COs):

Upon successful completion of this course, the students will be able to:

C309.1	Install and use database management systems.	Understand
C309.2	Represent logical design of database using E-R Diagram.	Apply
C309.3	Apply and demonstrate SQL queries, PLSQL Procedures to design and manage the database.	Apply
C309.4	Analyze and construct good database design.	Analyze

Prerequisite:	Set Theory and Data Structures

Course Articulation Matrix: Mapping of Course Outcomes (COs) with Program Outcomes (POs) and Program Specific Outcomes (PSOs):

COs	**POs**												**PSOs**		**BTL**
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**1**	**2**	
C309.1	2	2	-	-	-	-	-	-	-	-	-	-	-	-	2
C309.2	-	2	2	-	-	-	-	-	-	-	-	-	-	-	3
C309.3	-	3	3	1	-	-	-	-	-	-	-	2	2	-	3
C309.4	-	2	2	-	-	-	-	-	-	-	-	2	2	-	4

Minimum 8 to 10 experiments to be performed based on following guidelines:

Sr. No.	List of Experiments	S/O	Hours
1	ER Diagram of an Organization- Draw an E-R Diagram for any organization like Insurance Company, Library systems, College Management systems, Hospital Management systems etc. Use data modelling tools like Oracle SQL developer, Tode etc. to draw ER diagrams.	S	2
2	Conversion of ER Diagram to Tables- Convert the Above mentioned E-R Diagram in Relational Tables.	S	2
3	DDL Statements- Execute DDL commands to create, alter, rename, truncate and drop tables in SQL. Apply all types of constraints such as primary key, foreign key, not null, unique, check.	O	2
4	DML Statements- Use DML Queries to insert, delete, update & display records of the tables.	O	2
5	SQL character functions, String functions- Display the results using String operations.	O	2
6	Aggregate functions- Display the records using Aggregate functions and Group by, having, between, Order by clauses.	O	2
7	Join operations and set operations- Display the results of union, intersection, set difference, Cartesian product and Join operations of two different tables.	O	2
8	Views, Subqueries- Create Views for the table. Solve subqueries for given questions	O	2
9	Demonstrate PLSQL Functions and Procedures.	O	2
10	Demonstrate Cursors, and triggers using PL/SQL.	O	2
11	Database Connectivity- Write a program of Database connectivity with any object oriented language.	O	2
12	Write a program to implement Static Hashing.	O	2
13	Study of DCL commands (Grant, Revoke)	O	2

Text Books:

1. A. Silberschatz, H.F. Korth, S. Sudarshan, “Database System Concepts”, 6th Edition, McGraw Hill Education.
2. Thomos Connolly, Carolyn Begg, “Database Systems - A practical approach to Design, Implementation and Management”, 3rd Edition, Pearson Education.
3. Coronel, Morris, Rob, “Database Systems, Design, Implementation and Management”, Ninth Edition, Cengage Learning
4. Ramez Elmasri and Shamkant Navathe, “Fundamentals of Database Systems”, PearsonEducation, Fifth Edition

Reference Books:

1. Raghu Ramkrishnan, Johannes Gehrke, "Database Management System", Fourth Edition,
    a. McGrawHill Education.

Online Resources:

1. https://onlinecourses.nptel.ac.in/noc19_cs46/preview

Course Plan

Course Title: Economics and Management for IT	
Course Code :201DSMC310	**Semester: V**
Teaching Scheme: L-T-P : 2-0-0	**Credits: NA**
Evaluation Scheme: NA	**ESE Marks :50**

Course Description:
The course is intended to provide basic understanding of Economics and Management to engineering students with following aspects:
1. To impart knowledge, with respect to concepts of management information system,
2. To expose the students to the characteristic and applications of DSS.
3. To help the students to understand different trends in current information system technology and also IT Tools & Techniques for Business

Course Objectives:
1. To get the overview of system development management life cycle
2. To understand scope and objective of management information system.
3. To Enhancing management decision making
4.To understand Trends in information technology
5. To get the knowledge about IT Tools & Techniques for Business
6. To make the engineering students aware different corporate case studies.

Course Outcomes (COs):
Upon successful completion of this course, the students will be able to:

C310.1	Understand concept of system development management life cycle	Understand
C310.2	scope and objective of management information system	Remembering
C310.3	Enhancing management decision making	Remembering
C310.4	understand Trends in information technology	Understand
C310.5	get the knowledge about IT Tools & Techniques for Business	Appling
C310.6	make the engineering students aware different corporate case studies.	Analyzing

Course Articulation Matrix: Mapping of Course Outcomes (COs) with Program Outcomes (POs) and Program Specific Outcomes (PSOs):

COs	**POs**												**PSOs**		**BTL**
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**1**	**2**	
C310.1	1	2	2	-	-	-	-	-	-	-	-	-	2	-	**2**
C310.2	2	1	3	-	-	-	-	-	-	-	-	-	-	-	**1**
C310.3	2	2	2	-	-	-	-	-	-	-	-	-	-	-	**1**
C310.4	3	2	2	-	-	-	-	-	-	-	-	-	2	-	**2**
C310.5	2	2	2	-	-	-	-	-	-	-	-	-	1	-	**3**
C310.6	3	2	1	1	-	-	-	-	-	-	-	-	1	1	**4**

Content	Hours
Unit 1: Management Information System Conceptual foundations of information systems; Information theory; Information resource management; Types of information systems; Systems development - Overview of systems and design; System development management life-cycle, designing for online and distributed environments; Implementation and control of project;	**5**
Unit 2: Scope and Objectives of MIS MIS meaning and role, MIS concepts, Management science structure, Information flow in management, MIS for management support, Planning with MIS, control with MIS. Problem solving & decision making, Development of MIS, strategic & project planning for MIS.	**4**
Unit 3: Enhancing management decision making Decision support systems (DSS) – understanding DSS, characteristics components, major DSS applications. Group decision support systems (GDSS), - elements, characteristics, how GDSS can enhance group decision - making? Executive support systems (ESS) – role of ESS in the organization, developing ESS, benefits of ESS	**5**
Unit 4: Trends in information technology Managing data resources - Organizing data; DSS and RDBMS; Enterprise Resource Planning (ERP), Expert systems, e-Business architecture, e-Governance;	**4**
Unit 5:IT Tools & Techniques for Business Different IT Tools & Techniques for Business economics, Basic business statistics for data analysis, global economics, economics management used in project development	**4**
Unit 6: Case studies Web Publishing: types of websites, Web surfing, E-commerce, B2B, B2C, C2C, E-commerce security issues, Ethical issues.	**4**

Text Books:

1. Management of Information systems, Gordon B. Davis &Margreth H. Olson, Pearson Edition, (Unit I to III)

2. Information Systems for Management, Robert C. Murdik, PHI. 2nd ed, Unit –IV andV

Reference Books:

1. MIS Concepts & Design by Robert C. Murdik. PHI 2nd Ed.
2. Information system by H.F. & Abraham, S., Database System Concepts, McGraw Hill.
3. Engineering Economics, R.Paneerselvam, PHI publication
4. Economics: Principles of Economics, N Gregory Mankiw, Cengage Learning
5. Modern Economic Theory, By Dr. K. K. Dewett& M. H. Navalur, S. Chand Publications

Online recourses:

1. https://onlinecourses.swayam2.ac.in/cec21_ge05/preview
2. https://www.coursera.org/courses?query=management%20information%20systems
3. https://www.coursera.org/specializations/managerial-economics-business-analysis
4. https://www.mbaknol.com/management-information-systems/case-study-on-mis-information-system-in-restaurant/

D.Y. PATIL COLLEGE OF ENGINEERING & TECHNOLOGY

KASABA BAWADA KOLHAPUR-416006

(An Autonomous Institute) B. Tech. Curriculum

T. Y. B. Tech. Data Science (SEM-VI)

(AcademicYear-2022-23)

SEMESTER - VI

D.Y. PATIL COLLEGE OF ENGINEERING & TECHNOLOGY
KASABA BAWADA KOLHAPUR-416006
(An Autonomous Institute) B. Tech. Curriculum

T. Y. B. Tech. Data Science (SEM-VI)

(AcademicYear-2022-23)

Course Plan

Course Title: Advanced Database Systems	
Course Code: 201DSL311	**Semester: VI**
Teaching Scheme: L-T-P: 3-0-0	**Credits: 3**
Evaluation Scheme: ISE+MSE Marks: 20+30	**ESE Marks: 50**

Course Description:

This course focuses on different database systems like parallel databases, distributed databases and object relational databases. It also focuses on NoSQL, Data warehousing, Data mining and Web mining.

Course Objectives:

1. To acquire knowledge on parallel and distributed databases and its applications.
2. To understand the fundamentals of object-oriented databases.
3. To study the usage and applications of SQL and NOSQL databases.
4. To understand the usage of data warehousing, data mining, web mining techniques.

Course Outcomes (COs):

Upon successful completion of this course, the students will be able to:

C311.1	Understand and identify issues arising from parallel and distributed processing of data.	Understand
C311.2	Demonstrate the usage of object oriented databases.	Apply
C311.3	Compare and Contrast NoSQL databases with each other and Relational Database Systems.	Analyze
C311.4	Make use of data mining, web mining techniques and business intelligence to solve problems.	Apply

Prerequisite:	Database engineering

Course Articulation Matrix: Mapping of Course Outcomes (COs) with Program Outcomes (POs) and Program Specific Outcomes (PSOs):

COs	POs												PSOs		BTL
	1	2	3	4	5	6	7	8	9	10	11	12	1	2	
C311.1	3	-	1	-	-	-	-	-	-	-	-	-	-	-	2
C311.2	3	-	-	-	-	-	-	-	-	-	-	-	-	-	3
C311.3	2	-	2	-	-	-	-	-	-	-	-	-	2	-	4
C311.4	-	2	2	-	-	-	-	-	-	-	-	-	-	-	3

Content	Hours
Unit I: Parallel and Distributed Databases Parallel Systems, Parallel Database Architectures, Parallel Databases --I/O Parallelism, Design of Parallel Systems, Distributed Systems, Distributed Database Concepts, Distributed Data Storage, Distributed Transactions, Commit Protocols, Distributed Query Processing, Case Study-Distributed Databases in Oracle.	6
Unit 2: Object Relational Databases Motivating example, Structured data types, Operations on structured data, Encapsulation and ADTs, Inheritance, Objects, OIDS and Reference types, Database design for an ORDBMS, Object identity, Nested collections, Storage and access methods, Query processing and optimization, Comparison of RDBMS and ORDBMS. Case Study: Multimedia databases, spatial databases.	5
Unit 3: NoSQL Database Introduction, Need, Features. Types of NoSQL Databases: Key-value store, document store, graph, wide column stores, BASE Properties, Data Consistency model, ACID Vs BASE, Comparative study of RDBMS and NoSQL. MongoDB (with syntax and usage): CRUD Operations, Indexing, Aggregation, MapReduce, Replication, Sharding. **Case Study:** Cassandra, DynamoDB	7
Unit 4: Business Intelligence and Decision Support The Need for Data Analysis, Business Intelligence, Business Intelligence Architecture, Introduction to decision support, Data Warehousing, OLAP, Implementation Techniques for OLAP, Star Schemas, Views and decision support, View materialization, Maintaining materialized views. Case Study: Introduction to Business Intelligence tool- Power BI.	5
Unit 5: Data Mining and Graph Pattern Mining Introduction, Basic Data Mining Tasks, Data Mining Versus Knowledge Discovery in Databases, Data Mining Issues, Counting Co-occurrences, Mining for rules, Tree structured rules, Clustering: BIRCH algorithm, Similarity search over sequences, Graph Data- Graph Concepts, Topological Attributes, Graph Pattern Mining-Isomorphism and Support, Candidate Generation, The gSpan Algorithm.	8

Unit 6: Web Mining Introduction, Web Content Mining, Crawlers, Harvest System, Virtual Web View, Personalization, Web Structure Mining, PageRank, HITS algorithm, Clever, Web Usage Mining, Pre-processing, Data Structures, Pattern Discovery, Pattern Analysis.	5

Text Books:

1. A. Silberschatz, H.F. Korth, S. Sudarshan, “Database System Concepts”, 6th Edition, McGraw Hill Education. (Unit 1)
2. Raghu Ramkrishnan, Johannes Gehrke, “Database Management System”, Fourth Edition, McGraw Hill Education.(Unit 2, 4,5)
3. Pramod J. Sadalage and Marin Fowler “NoSQL Distilled: A brief guide to merging world of Polyglot persistence”, Addison Wesley, 2012. (Unit 3)
4. NoSQL for Mere Mortals- Dan Sullivan- 1st Edition, Pearson Education (Unit 3)
5. Mohammed J. Zaki, Wagner Meira Jr, “Data Mining and Analysis- Fundamental Concepts and Algorithms”, Cambridge University Press (Unit 5)
6. Margaret H. Dunham “Data Mining” Pearson Education (Unit 5, 6)

Reference Books:

1. Kevin Roebuck, "Storing and Managing Big Data - NoSQL, HADOOP and More", Emereo Pty Limited, 2011,
2. Ralph Kimball, "The Data Warehouse Lifecycle toolkit', 2nd edition, Wiley India.

Online Resources:

1. NoSQL-https://nptel.ac.in/courses/106104189

D.Y. PATIL COLLEGE OF ENGINEERING & TECHNOLOGY

KASABA BAWADA KOLHAPUR-416006

(An Autonomous Institute) B. Tech. Curriculum

T. Y. B. Tech. Data Science (SEM-VI)

(AcademicYear-2022-23)

Course Plan

Course Title: Optimization for Data Science	
Course Code: 201DSL312	**Semester: VI**
Teaching Scheme: L-T-P : 3-0-0	**Credits: 3**
Evaluation Scheme: ISE+MSE Marks: 20+30	**ESE Marks:50**

Course Description:

Techniques for formulating data science models as optimization problems. Algorithms for solving data science problems with an emphasis on scalability, efficiency and parallelizability including gradient- descent based algorithms, derivative-free algorithms, and randomized algorithms. Theoretical analyses of algorithms and approaches for recognizing the most suitable algorithm for solving a particular problem.

Course Objectives:

1. To get knowledge of several fundamental problems in machine learning.
2. To become familiar with basic linear-algebra- and optimization-centric algorithms in machine learning
3. To understand classic learning enabled optimization method and algorithm configuration.

Course Outcomes COs:

At the end of this course students will be able to

C312.1	Understand basics of optimization	Understand
C312.2	Optimize gradient-based algorithms, deterministic and stochastic first-order methods.	Apply
C312.3	Evaluate convex functions and optimality condition	Evaluate
C312.4	Perform model selection and hyperparameter optimization	Apply

Prerequisite	Basic linear algebra, probability, Introduction to Machine learning

Course Articulation Matrix: Mapping of Course Outcomes (COs) with Program Outcomes (POs) and Program Specific Outcomes (PSOs)

COs	**POs**												**PSOs**		**BTL**
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**1**	**2**	
C312.1	3	-	-	-	-	-	-	-	-	-	-	2	-	-	2
C312.2	3	2	2	2	-	-	-	-	-	-	-	2	-	-	3
C312.3	3	2	2	2	-	-	-	-	-	-	-	2	-	-	5
C312.4	3	2	2	2	-	-	-	-	-	-	-	2	-	-	3

Contents	Hours.
Unit 1: Linear Programming: Introduction to linear and nonlinear programming, Geometry of linear programming, Graphical method, Linear programming (LP) in standard form, Solution of LP by simplex method, Duality theory, Dual simplex method.	6
Unit 2: **Nonlinear Programming:** Unconstrained Optimization: unimodal functions, Fibonacci search method, Steepest Descent method. Constrained Optimization: Concept of convexity and concavity, Maxima and minima of functions of n-variables, Lagrange multipliers, Karush-Kuhn-Tucker conditions for constrained optimization.	6
Unit 3: Optimization with Numerical and binary targets: Linear regression, Tikhonov regression, stochastic gradient descent ,use of bias, least square classification,SVM, Logistic regression, How LR is parent problem in ML.	6
Unit 4: Convex Optimization Theory: Convex set-Definition, projection, separation theorem, Convex function-definition, differentiable and non-differentiable convex functions, optimality condition for convex optimization	6
Unit 5: Hyper parameter Optimization in ML: Introduction, understanding hyperparameters, need of hyperparameter optimization, ML algorithms and their hyperparameters, distribution of possible hyperparameters values, Hyperparameter Optimization Using Scikit-Learn.	6
Unit 6: Automated ML and Bayesian optimization: Automated ML-model selection, hyperparameter optimization, combined algorithm selection, Bayesian optimization: basic structure, sequential model base optimization, surrogate model, acquisition function, Automated ML for predictive analysis.	6

Text Books:

1) Chandra, S., Jayadeva, Mehra, A.,"Numerical Optimization and Applications", Narosa Publishing House[Unit 1,2]
2) Charu C. Aggarwal ,"Linear Algebra and Optimization for Machine Learning-A Textbook",Springer Publication[Unit 3]
3) Guanghui Lan,"First-order and Stochastic Optimization Methods for Machine Learning",Springer Series in the Data Sciences[Unit 4]
4) Tanay Agrawal,"Hyperparameter Optimization in Machine Learning",Apress Publication[Unit 5]
5) Francesco Archetti ,Antonio Candelier ,"Bayesian Optimization and Data Science" Springer Briefs inOptimization[Unit 6]

Reference Books:

1) Jean Gallier and Jocelyn Quaintance," Fundamentals of Optimization Theory with Applications to MachineLearning*"*

References:

1. https://onlinecourses.nptel.ac.in/noc21_me10/preview
2. https://www.coursera.org/learn/optimize-machine-learning-model-performance
3. https://machinelearningmastery.com/optimization-for-machine-learning-crash-course/

Course Plan

Course Title: Information Security	
Course Code: 201DSL313	**Semester: VI**
Teaching Scheme: L-T-P: 3-0-0	**Credits: 3**
Evaluation Scheme: ISE+MSE Marks: 20+30	**ESE Marks:50**

Course Description:

This course aims to explore information security through some introductory contents, defining the key terms, explaining essential concepts and goes through a review of the origins of the field. It presents several legal and ethical issues that are commonly found in today's organizations. The key topics explored include cryptography, access control, protocols, software, software security, OS security, data security and Software Vulnerabilities.

Course Objectives:

1. To introduce the principles, components of information security, legal & ethical issues.
2. To expose students to various cipher techniques and cryptographic algorithms, tools.
3. To understand ways to provide access control like authorization and authentication.
4. To make the students understand the working of several real-world security protocols.
5. To analyze the security provisions in software, operating system and data.
6. To make students to explore non-cryptographic and software vulnerabilities.

Course Outcomes (COs):

Upon successful completion of this course, the students will be able to:

C313.1	Describe the major components of information security, security model, legal and ethical issues.	Understand
C313.2	Apply cipher techniques , different cryptographic algorithms and tools.	Apply
C313.3	Present the ways to provide access control using authorization and authentication.	Analyze
C313.4	Illustrate the working of various real-world security protocols.	Apply
C313.5	Analyze the security provisions in software, operating system and data.	Analyze
C313.6	Explore newer vulnerabilities and provide to take precautionary measures	Understand

Prerequisite:	Computer Networks

Course Articulation Matrix: Mapping of Course Outcomes (COs) with Program Outcomes

(POs) and Program Specific Outcomes (PSOs):

COs	POs												PSOs		BTL
	1	2	3	4	5	6	7	8	9	10	11	12	1	2	
C313.1	-	-	3	1	-	-	-	1	-	-	-	-	1	-	2
C313.2	3	-	1	-	2	-	-	-	-	-	-	-	1	-	3
C313.3	-	2	1	-	1	-	-	-	-	-	-	-	1	-	4
C313.4	1	-	3	-	-	-	-	-	-	-	-	-	1	-	3
C313.5	-	3	-	1	-	-	-	-	-	-	-	-	1	-	4
C313.6	-	-	2	-	-	-	2	1	-	-	-	-	1	-	2

Content	Hours
Unit 1: Introduction to Information Security. **Introduction:** Components of an Information System, The Security Systems Development Life Cycle, Information Security: The Need for Security, Legal, Ethical and Professional Issues. **Cryptosystems:** Foundations of Cryptology, The OSI Security Architecture, Principles of Public-Key Cryptosystems, Attacks on Cryptosystems.	6
Unit 2: Cryptography Symmetric Cipher Models: Substitution Techniques, Transposition Techniques, Block Cipher Principles, The Data Encryption Standard. Cryptographic Algorithms-The RSA Algorithm, Key Management, Diffie-Hellman Key Exchange, MAC and Hash functions and their requirements.	7
Unit 3: Protocols Simple Security Protocols, Authentication Protocols, Digital Signature, Digital Signature Standard, Authentication applications - Kerberos, X.509 Authentication service. Email Security - PGP, S/MIME, IP Security - IP Security Architecture, Authentication, Header and Encapsulating Security Payload.	7
Unit 4: Access Control: Authentication & Authorization Authentication Methods, Passwords, Biometrics, Two-Factor Authentication, Single Sign- On and Web Cookies, Authorization, Access Control Matrix, CAPTCHA, Firewalls, Intrusion Detection Systems. Web and System Security - Secure Socket Layer and Transport Layer Security, Secure Electronic Transaction.	5
Unit 5: Security in Software, Operating System and Data Security Software Flaws and malware, Insecurity in Software: Software, Reverse Engineering,	6

Digital Rights Management, Software Development. Operating Systems and Security: OS Security Functions, Trusted Operating System, Next Generation Secure Computing Base, Introduction to data Protectionby patents, copyrights and trademarks, IT security acts.	
Unit 6: key cybersecurity compliance &Vulnerabilities key cybersecurity compliance and industry Standards-Introduction to GTA, Zero Trust architecture. Software Security Software Vulnerabilities: Buffer Overflow, Salami Attack, Format string, cross-site scripting, SQL injection, Malware: Viruses, Worms, Trojans, Logic Bomb, Bots, Rootkits.	5

Text Books:

1. Michael E. Whitman and Herbert J. Mattord, “Principles of Information Security”, Course Technology Cengage Learning, Fourth Edition.
2. Mark Stamp, “Information Security Principles and Practice”, John Wiley & Sons Publications, Second Edition.

Reference Books:

1. Atul Kahate, “Cryptography and Network Security”, TMGH, Third Edition.
2. Williams Stallings, “Cryptography and Network Security Principles and Practices”, Pearson Education (LPE), Seventh Edition.

Online Resources:

https://nptel.ac.in/courses/106106129

Professional Elective I

Course Plan

Course Title: Blockchain Technology	
Course Code: 201DSL3P14	Semester: VI
Teaching Scheme: L-T-P: 3-1-0	Credits: 4
Evaluation Scheme: ISE+MSE Marks: 20+30	ESE Marks: 50

Course Description:
This course gives deep understanding of basics and design principles of Blockchain systems. Various payment verification protocols that are used in practice are included in this course. Design, building and deployment of distributed applications are covered in this course.

Course Objectives:

1. Understand how Blockchain systems (mainly Bitcoin and Ethereal) work.
2. To securely interact with bitcoin and ethereum.
3. Design, build, and deploy smart contracts and distributed applications.
4. Integrate ideas from Blockchain technology into their own projects.

Course Outcomes (COs):

Upon successful completion of this course, the students will be able to:

C314.1	Explain design principles of Bitcoin and Ethereum.	Understand
C314.2	Explain Nakamoto consensus.	Understand
C314.3	Explain the Simplified Payment Verification protocol	Apply
C314.4	List and describe differences between proof-of-work and proof-of-stake consensus.	Analyze
C314.5	Interact with a blockchain system by sending and reading transactions.	Evaluate
C314.6	Design, build, and deploy a distributed application.	Creating

Prerequisite:	Expertise In Programming, Basic Knowledge of Computer Security, Cryptography, Networking, Concurrent Or Parallel Programming

Course Articulation Matrix: Mapping of Course Outcomes (COs) with Program Outcomes(POs) and Program Specific Outcomes (PSOs):

CO	PO												PSO		BTL
	1	2	3	4	5	6	7	8	9	10	11	12	1	2	
C314.1	3	3	-	-	-	-	-	-	-	-	-	-	-	-	2
C314.2	2	2	-	2	-	-	-	-	-	-	-	-	-	-	2
C314.3	2	2	2	-	-	-	-	-	-	-	-	-	-	-	2
C314.4	2	2	-	2	-	-	-	-	-	-	-	-	-	-	4
C314.5	3	3	2	3	-	-	-	-	-	-	-	-	-	-	5
C314.6	2	2	-	2	-	-	-	-	-	-	-	-	-	-	6

Content	Hours
Unit I – Introduction Distributed Database, Two General Problem, Byzantine General problem and Fault Tolerance, HadoopDistributed File System, Distributed Hash Table, ASIC resistance, Turing Complete. Cryptography: Hash function, Digital Signature - ECDSA, Memory Hard Algorithm, Zero Knowledge Proof.	8
Unit II – Blockchain Introduction, Advantage over conventional distributed database, Blockchain Network, MiningMechanism, Distributed Consensus, Merkle Patricia Tree, Gas Limit, Transactions and Fee,Anonymity, Reward, Chain Policy, Life of Blockchain application, Soft & Hard Fork, Private and Public blockchain.	5
Unit III - Distributed Consensus Nakamoto consensus, Proof of Work, Proof of Stake, Proof of Burn, Difficulty Level, SybilAttack, Energy utilization and alternate.	5
Unit IV – Crypto Currency History, Distributed Ledger, Bitcoin protocols - Mining strategy and rewards, Ethereum - Construction,DAO, Smart Contract, GHOST, Vulnerability, Attacks, Sidechain, Namecoin.	5
Unit V - Crypto Currency Regulation Stakeholders, Roots of Bit coin, Legal Aspects-Crypto currency Exchange, Black Market and GlobalEconomy.	8
Unit VI - Crypto currency Applications Internet of Things, Medical Record Management System, Domain Name Service and future ofBlockchain.	5

List of Assignments

Ass. No.	Name of Assignment	S/O	Hours
1	Naive Blockchain construction.	S	1
2	Memory Hard algorithm – Hashcash implementation.	S	1
3	Direct Acyclic Graph.	S	1
4	Study of Go-ethereum.	S	1
5	Smart Contract Construction.	S	1
6	Toy application using Blockchain.	O	1
7	Mining puzzles.	S	1
8	List the use of public blockchain platforms.	S	1
9	Study of Blockchain protocols.	S	1
10	Case study: Secure sharing of medical data using Blockchain technology.	S	1

Text Books:

1. Arvind Narayanan, Joseph Bonneau, Edward Felten, Andrew Miller and Steven Goldfeder, “Bitcoin and Cryptocurrency Technologies: A Comprehensive Introduction” Princeton University Press (July19, 2016).

Reference Books:

1. S. Shukla, M. Dhawan, S. Sharma, S. Venkatesan, 'Blockchain Technology: Cryptocurrency andApplications, Oxford University Press, 2019.
2. Josh Thompson, "Blockchain: The Blockchain for Beginnings, Guild to Blockchain Technology andBlockchain Programming", Create Space Independent Publishing Platform, 2016

Online Resources:

1. https://www.youtube.com/watch?v=PPFsG92-HiI&list=PLbRMhDVUMngfxxyVLh2t2gKDUfsOdGn56
2. https://www.youtube.com/watch?v=9ce1FbFO_YQ&list=PLFW6lRTa1g80R3O_5-jsNVhDFAeGaABo9
3. https://www.youtube.com/watch?v=mzPoUjOC4WU&list=PLEAYkSg4uSQ2x4I7ASRHlraNxSwf8xOAB
4. https://www.youtube.com/watch?v=TOlv1nRDdfs&list=PLyqSpQzTE6M8wy_JBTgplS_HGuOYU1qkm
5. https://www.youtube.com/watch?v=QCvL-DWcoj

D.Y. PATIL COLLEGE OF ENGINEERING & TECHNOLOGY

KASABA BAWADA KOLHAPUR-416006

(An Autonomous Institute) B. Tech. Curriculum

T. Y. B. Tech. Data Science (SEM-VI)

(AcademicYear-2022-23)

Course Plan

Course Title: High Performance Computing	
Course Code: 201DSL3P15	**Semester: VI**
Teaching Scheme: L-T-P :3-1-0	**Credits: 4**
Evaluation Scheme: ISE +MSE Marks: 20+30	**ESE Marks:50**

Course Description:

This course is an introduction to the architecture of and software techniques for parallel and high performance computing systems. The content includes fundamental aspects of vector processing, shared-memory, and distributed-memory systems. Specific applications in parallel processing paradigms will be covered.

Course Objectives:

1. To Study various computing technology architectures.
2. To know Emerging trends in Parallel computing technology.
3. To highlight the performance of computing.

Course Outcomes COs:

At the end of this course students will be able to

C315.1	Identify the key factors affecting performance of computing systems.	Understand
C315.2	Illustrate mapping of applications to high-performance computing systems	Apply
C315.3	Apply hardware/software co-design for achieving performance on real-world applications	Evaluate

Prerequisite	Fundamentals of Computer Architecture

Course Articulation Matrix: Mapping of Course Outcomes (COs) with Program Outcomes (POs)and Program Specific Outcomes (PSOs)

Cos	Pos												PSOs		BTL
	1	2	3	4	5	6	7	8	9	10	11	12	1	2	
C315.1	2	-	-	-	-	-	-	-	-	-	-	2	-	-	2
C315.2	2	2	2	2	-	-	-	-	-	-	-	2	-	-	3
C315.3	2	2	2	2	-	-	-	-	-	-	-	2	-	-	5

Contents	Hours.
Unit 1: Introduction: Computational Science and Engineering: Computational Science and Engineering Applications, characteristics and requirements, Review of Computational Complexity, Performance: metrics and measurements, Granularity and Partitioning, Locality: temporal / spatial /stream / kernel, Basic methods for parallel programming, Real-world case studies (drawn from multiscale, multi-discipline applications)	6
Unit 2: High-End Computer Systems: Memory Hierarchies, Multi-core Processors: Homogeneous and Heterogeneous, Shared-memory Symmetric Multiprocessors, Vector Computers, Distributed Memory Computers, Supercomputers and Petascale Systems, Application Accelerators / Reconfigurable Computing, Novel computers: Stream, multithreaded, and purpose-built	8

Unit 3: Parallel Algorithms: Parallel models: ideal and real frameworks, Basic Techniques: Balanced Trees, Pointer Jumping, Divide and Conquer, Partitioning, Regular Algorithms: Matrix operations and Linear Algebra, Irregular Algorithms: Lists, Trees, Graphs, Randomization: Parallel Pseudo-Random Number Generators, Sorting, Monte Carlo techniques	8
Unit 4: Parallel Programming: Revealing concurrency in applications, Task and Functional Parallelism, Task Scheduling, Synchronization Methods, Parallel Primitives (collective operations), SPMD Programming (threads, OpenMP, MPI), I/O and File Systems, Parallel Matlabs (Parallel Matlab, Star-P, Matlab MPI), Partitioning Global Address Space (PGAS) languages (UPC, Titanium, Global Arrays)	8
Unit 5: Achieving Performance: Measuring performance, identifying performance bottlenecks, Restructuring applications for deep memory hierarchies, Partitioning applications for heterogeneous resources, using existing libraries, tools, and framework	6

List of Assignments

Ass. No.	Name of Assignment	S/O	Hours
1	Computational Complexity and performance matrices	S	1
2	Memory hierarchy	S	1
3	Multicore distributed memory computer systems	S	1
4	Study of Multithreaded architecture	S	1
5	Study of Parallel algorithms	S	1
6	Parallel programming and task scheduling	S	1
7	Parallelizing algorithms	S	1
8	Deep memories hierarchies	S	1
9	Case study :Performance measuring	S	1
10	Study of Performance bottleneck	S	1

Text Books:

1. Introduction to Parallel Computing, Ananth Grama, Anshul Gupta, George Karypis, and Vipin Kumar, 2nd edition, Addison-Welsey, 2003
2. Petascale Computing: Algorithms and Applications, David A. Bader (Ed.), Chapman & Hall/CRC Computational Science Series, 2007Optimization and Data Science" Springer Briefs in Optimization

Reference Books:

1. Grama, A. Gupta, G. Karypis, V. Kumar, An Introduction to Parallel Computing, Design andAnalysis of Algorithms: 2/e, Addison-Wesley, 2003
2. G.E. Karniadakis, R.M. Kirby II, Parallel Scientific Computing in C++ and MPI: A SeamlessApproach to Parallel Algorithms and their Implementation, Cambridge University Press,2003.
3. Wilkinson and M. Allen, Parallel Programming: Techniques and Applications Using NetworkedWorkstations and Parallel Computers, 2/E, Prentice Hall, 2005.

4. M.J. Quinn, Parallel Programming in C with MPI and OpenMP, McGraw-Hill, 2004.
5. G.S. Almasi and A. Gottlieb, Highly Parallel Computing, 2/E, Addison-Wesley, 1994.

Course Plan

Course Title: E-Commerce and Recommendation Systems	
Course Code: 201DSL3P16	**Semester: VI**
Teaching Scheme: L-T-P: 3-1-0	**Credits: 4**
Evaluation Scheme: ISE+MSE Marks: 20+30	**ESE Marks: 50**

Course Description:
The course will cover fundamental aspects of Recommendation systems, focusing on the use and applications of Recommender systems. Recommender systems are around us and are encountered on multiple domains such as e-commerce, content and media distribution, social media and so on. The course aims to explain both basics and advanced topics and concepts for recommendation systems.

Course Objectives:

1. To become familiar with concept of E commerce.
2. To get exposed basic concepts of recommendation system.
3. To classify different recommendation systems.

Course Outcomes (COs):

Upon successful completion of this course, the students will be able to:

C316.1	Understand the fundamentals of E- commerce and recommendation systems.	Understand
C316.2	Explain different collaborative recommendation systems.	Understand
C316.3	Describe content based, knowledge-based recommendation systems.	Understand
C316.4	Carry out performance evaluation of recommender systems based on various metrics.	Understand

Prerequisite:	Basic calculus and algorithms background.

Course Articulation Matrix: Mapping of Course Outcomes (COs) with Program Outcomes (POs) and Program Specific Outcomes (PSOs):

Cos	POs												PSOs		BTL
	1	2	3	4	5	6	7	8	9	10	11	12	1	2	
C316.1	2	2	-	2	-	2	-	-	-	-	-	2	-	-	2
C316.2	2	2	-	2	-	2	-	-	-	-	-	2	-	-	2
C316.3	2	2	-	2	-	2	-	-	-	-	-	2	2	2	2
C316.4	2	2	-	2	-	2	-	-	-	-	-	2	2	2	2

Content	Hours
Unit 1: Introduction to Electronic Commerce Introduction to e-commerce, electronic commerce frame work, Anatomy of e-commerce application, electronic commerce and World Wide Web, Architectural frame work of e-commerce, electronic payment system and its types.	4
Unit 2: Introduction to Recommendation Systems Introduction and basic taxonomy of recommendation systems, Traditional and non-personalized recommendation systems, Overview of data mining methods for recommendation systems, Overview of convex and linear optimization principles, Applications of recommendation systems, Issues with recommender system.	6
Unit 3: Collaborative Recommendation Systems User-based Nearest Neighbor recommendation, Item-based nearest neighbor recommendation, Model based and pre-processing-based approaches, Attacks on collaborative recommender systems. Collaborative Filtering as a Generalization of Classification and Regression Modelling.	8
Unit 4: Content-Based Recommendation Systems Introduction, Basic components of content based recommendation systems, Advantages and drawbacks of content based filtering, Pre-processing and Feature Extraction, Item representation Methods for learning user profiles.	7
Unit 5: Knowledge Based Recommendation Systems Introduction, Knowledge representation and reasoning, Constraint based recommenders, Case based recommenders: Similarity Metrics, Critiquing methods, Explanation in critiques, Persistent Personalization in Knowledge-Based Systems	5
Unit 6: Hybrid Recommendation Systems Opportunities for hybridization, Monolithic hybridization design: Feature combination, Feature augmentation, Parallelized hybridization design: Weighted, Switching, Mixed, Pipelined hybridization design: Cascade Meta-level, Limitations of hybridization strategies.	7

List of Tutorials			
Ass. No.	**Name of Tutorial**	**S/O**	**Hours**
1	Fundamental concepts of E Commerce.	S	1
2	Architecture of Electronic commerce.	S	1
3	Applications of recommendation systems.	S	1
4	Basic taxonomy of recommendation systems.	S	1
5	Data mining methods for recommendation systems	S	1
6	Collaborative recommendation systems	S	1
7	Basic components of content based recommendation systems.	S	1
8	Constraint based recommenders	S	1
9	Hybridization design	S	1

Text Books:

1. C.C. Aggarwal, Recommender Systems: The Textbook, Springer, 2016.
2. Jannach D., Zanker M. and FelFering A., Recommender Systems: An Introduction, Cambridge University Press(2011), 1st ed.

Reference Books:

1. Ricci F., Rokach L., Shapira D., Kantor B.P., Recommender Systems Handbook, Springer(2011) 1st edition.
2. M. Chiang, Networking Life, Cambridge, 2010

Online Resources

1. http://www.digimat.in/nptel/courses/video/110105083/L52.html
2. https://www.youtube.com/watch?v=giIXNoiqO_U&list=PL3ZVX5cUMdLbiFgitZszhnMUZHDDEL0rS
3. https://www.coursera.org/specializations/recommender-systems

Course Plan

Course Title : Advanced Database Systems Laboratory	
Course Code: 201DSP319	**Semester: VI**
Teaching Scheme: L-T-P : 0-0-2	**Credits: 1**
Evaluation Scheme: ISE Marks:25	**ESE-POE: 50**

Course Description:

This course focuses on different database systems like parallel databases, distributed databases and object relational databases. It also focuses on NoSQL, Data warehousing, Data mining and Web mining.

Course Objectives:

1. To acquire knowledge on parallel and distributed databases and its applications.
2. To understand the fundamentals of object oriented databases.
3. To study the usage and applications of SQL and NOSQL databases.
4. To understand the usage of data warehousing, data mining, web mining techniques.

Course Outcomes (COs):

Upon successful completion of this course, the students will be able to:

C319.1	Understand and identify issues arising from parallel and distributed processing of data.	Understand
C319.2	Demonstrate the usage of object oriented databases.	Apply
C319.3	Compare and Contrast NoSQL databases with each other and Relational Database Systems.	Evaluate
C319.4	Make use of data mining, web mining techniques and business intelligence to solve problems.	Apply

Prerequisite:	Database engineering

Course Articulation Matrix: Mapping of Course Outcomes (COs) with Program Outcomes (POs) and Program Specific Outcomes (PSOs):

COs	POs												PSOs		BTL
	1	2	3	4	5	6	7	8	9	10	11	12	1	2	
C319.1	3	-	1	-	-	-	-	-	-	-	-	-	-	-	2
C319.2	3	-	-	-	-	-	-	-	-	-	-	-	-	-	3
C319.3	2	-	2	-	-	-	-	-	-	-	-	-	2	-	4
C319.4	-	2	2	-	-	-	-	-	-	-	-	-	-	-	3

Minimum 8 to 10 experiments to be performed based on following guidelines:

Sr. No.	List of Experiments	S/O	Hours
1	Implement partitioning techniques on parallel databases.	O	2
2	Implement vertical or horizontal fragmentation in distributed DBMS.	O	2
3	Implement semi join in distributed DBMS.	O	2
4	Implementation of 2 Phase Commit protocol for distributed databases.	O	2
5	Create structured data types of ORDBMS and perform operations-create tableusing structured data types, insert data and solve queries.	O	2
6	Study of Open Source NOSQL Database: MongoDB (Installation, BasicCRUD operations, Execution)	O	2
7	Design and Develop MongoDB Queries using CRUD operations. (UseCRUD operations, SAVE method, logical operators)	O	2
8	Implement aggregation with suitable example using MongoDB	O	2
9	Implement Map Reduce operation with suitable example using MongoDB	O	2
10	Demonstrate all OLAP operations and cube operator in OLAP.	O	2
11	Implement A-priori algorithm in data mining.	O	2

Text Books:

1. Silberschatz, H.F. Korth, S. Sudarshan, “Database System Concepts”, 6th Edition, McGraw HillEducation.
2. Raghu Ramkrishnan, Johannes Gehrke, “Database Management System”, Fourth Edition, McGraw Hill Education.
3. Pramod J. Sadalage and Marin Fowler “NoSQL Distilled: A brief guide to merging world ofPolyglot persistence”, Addison Wesley, 2012.
4. NoSQL for Mere Mortals- Dan Sullivan- 1st Edition, Pearson Education

Reference Books:

1. Kevin Roebuck, "Storing and Managing Big Data - NoSQL, HADOOP and More", Emereo Pty Limited, 2011,ISBN 1743045743, 9781743045749
2. Ralph Kimball, "The Data Warehouse Lifecycle toolkit', 2nd edition, Wiley India.

Online Resources:

1. NoSQL-https://nptel.ac.in/courses/106104189

Course Title: Project II	
Course Code:201DSP320	Semester: VI
Teaching Scheme: L-T-P:0-0-4	Credits:2
Evaluation Scheme: ISE Marks: 25	ESE(POE)Marks :50

Course Description

This course emphasizes on a Project-based learning approach. It is a group activity / work where students have to present an idea/ solution for the problem chosen. The project should enable the students to combine the theoretical and practical concept applied in his / her academics. The project work should enable the students to exhibit their ability to work in a team, develop planning and execute skills and perform analyzing and troubleshooting of their respective problem chosen for the project. Ultimately this course enhances students programming skills and enable them to learn how to perform requirements analysis, system designing, testing, coding and report writing and to find scope for further development of project. Over the course of the work, students should use modern techniques, methods and approaches for the development of systems and software.

Course Objectives

1. To expose the students to use engineering approaches to solve problems in real time.
2. To use modern techniques, methods and approaches for the development of the project.
3. To learn the techniques of working together in a team.

Course Outcomes (COs)

At the end of the course the student should be able to:

C320.1	Identify specific problem statement from a selected domain.	Understand
C320.2	Analyze the problem and prepare SRS and design document.	Analyze
C320.3	Write code and carry out testing.	Apply
C320.4	Write a report covering details of the project and present it.	Apply

Prerequisite:	Engineering Mathematics, Data Structures, Software Engineering and knowledge of Programming language.

Course Articulation Matrix: Mapping of Course Outcomes (COs) with Program Outcomes(POs)and Program Specific Outcomes (PSOs):

	POs												PSOs		BTL
Cos	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**1**	**2**	
C320.1	2	3	-	-	-	2	2	-	-	-	-	-	-	2	2
C320.2	2	3	3	3	2	-	-	-	-	-	-	-	-	2	4
C320.3	-	-	3	-	2	-	-	-	-	3	2	-	3	3	3
C320.4	3	-	3	-	3	-	-	-	3	3	2	3	-	-	3

Course Contents

The Project should be undertaken by a group of 2 to 3 students and every group supposed to select the domain (Preferably problem statement from Smart India Hackathon / Industry Hackathons/ Government Hackathons etc.). The group should identify the relevant problems in the selected domain, define & analyze the problem, design the solution and implementing it using suitable technology. A synopsis approval presentation should be conducted where the student shall propose the project work. The Project work should be evaluated by a panel of teachers appointed by the department based on a minimum of two reviews. Review I & II should be conducted for 25 Marks using rubrics I & II. Final ISE marks should be the average of review I & II. Final ESE exam should be conducted for 50 marks using rubrics III. It is desirable that the student participates in project competitions, hackathon and paper presentation. Student should not involve in out-sourcing of the project work.

Course Plan

Course Title : Human Values and Ethics	
Course Code : 201DSMC321	**Semester : VI**
Teaching Scheme : L-T-P : 2-0-0	**Credits : --**
Evaluation Scheme : ISE Marks : -	**ESE Marks : 50**

Course Description:

The methodology of this course is universally adaptable, involving a systematic and rational study of the human being vis-à-vis the rest of existence. It is free from any dogma or value prescriptions. This process of self-exploration takes the form of a dialogue between the teacher and the students to begin with and within the student himself/herself finally.

Course Objectives:

1. To create an awareness on Engineering Ethics and Human Values.
2. To understand social responsibility of an engineer.
3. To appreciate ethical dilemma while discharging duties in professional life.

Course Outcomes (COs):

Upon successful completion of this course, the students will be able to:

C321.1	Understand the significance of value inputs in a classroom and start applying them in their life and profession	Understand
C321.2	Distinguish between values and skills, happiness and accumulation of physical facilities, the Self and the Body, Intention and Competence of an individual, etc.	Evaluate
C321.3	Understand the role of a human being in ensuring harmony in society and nature.	Understand
C321.4	Distinguish between ethical and unethical practices, and start working out the strategy to actualize a harmonious environment wherever they work.	Evaluate

Course Articulation Matrix: Mapping of Course Outcomes (COs) with Program Outcomes (POs) and Program Specific Outcomes (PSOs):

COs	**POs**												**PSOs**		**BTL**
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**1**	**2**	
C321.1	-	-	-	-	-	1	-	2	-	-	-	2	-	2	2
C321.2	-	-	-	-	-	2	-	2	-	-	-	2	-	2	4
C321.3	-	-	-	-	-	2	2	2	2	-	-	2	-	2	2
C321.4	-	-	-	-	-	2	2	2	2	-	2	2	-	2	4

Content	Hours
UNIT I: Introduction to Value Education Value Education, Definition, Concept and Need for Value Education. The Content and Process of Value Education. Basic Guidelines for Value Education. Self-exploration as a means of Value Education. Happiness and Prosperity as parts of Value Education.	4
UNIT II: Harmony in the Human Being Human Being is more than just the Body. Harmony of the Self ('I') with the Body. Understanding Myself as Co-existence of the Self and the Body. Understanding Needs of the Self and the needs of the Body. Understanding the activities in the Self and the Activities in the Body.	4
UNIT III: Harmony in the Family, Society and in the Nature Family as a basic unit of Human Interaction and Values in Relationships. The Basics for Respect and today's Crisis: Affection, e, Guidance, Reverence, Glory, Gratitude and Love. Comprehensive Human Goal: The Five Dimensions of Human Endeavour. Harmony in Nature: The Four Orders in Nature. The Holistic Perception of Harmonyin Existence.	5
UNIT IV: Social Ethics The Basics for Ethical Human Conduct. Defects in Ethical Human Conduct. Holistic Alternative and Universal Order. Universal Human Order and Ethical Conduct. Human Rights violation and Social Disparities.	5
UNIT V: Professional Ethics Value based Life and Profession. Professional Ethics and Right Understanding. Competence in Professional Ethics. Issues in Professional Ethics – The Current Scenario. Vision for Holistic Technologies, Production System and Management Models.	4
Unit VI: code of ethics for computer engineers: fundamental principles, human welfare, employers and clients, engineering profession and society.	4

Text Books:

1.A.N.Tripathy, New Age International Publishers, 2003.
2.Bajpai. B. L. , New Royal Book Co, Lucknow, Reprinted, 2004
3.Bertrand Russell Human Society in Ethics & Politics

Reference Books:

1. Corliss Lamont, Philosophy of Humanism
2. Gaur. R.R. ,Sangal. R, Bagaria. G.P, A Foundation Course in Value Education, Excel Books,2009.
3. Gaur. R.R. ,Sangal. R ,Bagaria. G.P, Teachers Manual Excel Books, 2009.
4.I.C. Sharma . Ethical Philosophy of India Nagin& co Julundhar
5.Mortimer. J. Adler, – Whatman has made of man
6. William Lilly Introduction to Ethic Allied Publisher

Open Elective –I

Course Plan

Course Title: Basics of Data Science	
Course Code: 201DSL3O17	**Semester: VI**
Teaching Scheme: L-T-P: 3-1-0	**Credits: 4**
Evaluation Scheme : ISE + MSE = 20 + 30 = 50	**ESE Marks : 50**

Course Description:
The course on Basics of Data Science provides an overview of Data Science, covering a broad selection of key challenges in and methodologies for working with big data. Topics to be covered include data collection, integration, management, modeling, analysis, visualization, prediction and informed decision making, as well as data security and data privacy.

Course Objectives:
To provide strong foundation for data science and application area related to it and understand the underlying core concepts and emerging technologies in data science.

Course Outcomes (COs):

Upon successful completion of this course, the students will be able to:

C317.1	Explore the fundamental concepts of data science.	Understand
C317.2	Understand data analysis techniques for applications handling large data.	Understand
C317.3	Understand various machine learning algorithms used in data science process.	Apply
C317.4	Visualize and present the inference using various tools.	Analyze
C317.5	Learn to think through the ethics surrounding privacy, data sharing and algorithmic decision-making	Evaluate

Prerequisite:	Knowledge of social media channels.

Course Articulation Matrix: Mapping of Course Outcomes (COs) with Program Outcomes (POs) and Program Specific Outcomes (PSOs):

COs	**POs**												**PSOs**		**BTL**
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**1**	**2**	
C317.1	2	2	-	-	-	-	-	-	-	-	-	-	-	-	2
C317.2	3	3	2	-	-	-	-	-	-	-	-	-	-	-	2
C317.3	2	2	-	-	-	-	-	-	-	-	-	-	-	-	3
C317.4	2	2	-	2	2	-	-	-	-	-	-	-	-	-	4
C317.5	2	2	2	2	2	-	-	2	-	-	-	2	-	-	5

Content	Hours
Unit 1: Introduction to core concepts and technologies Introduction, Terminology, data science process, data science toolkit, Types of data, Example applications, Mathematical Foundations for Data Science: linear algebra; Analytical and numerical solutions of linear equations; Mathematical structures, concepts and notations used in discrete mathematics.	**6**
Unit 2: Data collection and management Introduction, Sources of data, Data collection and APIs, Exploring and fixing data, Data storage and management, using multiple data sources	**4**
Unit 3: Data analysis Introduction, Terminology and concepts, Introduction to statistics, Central tendencies and distributions, Variance, Distribution properties and arithmetic, Samples/CLT, Basic machine learning algorithms, Linear regression, SVM, Naive Bayes.	**7**
Unit 4: Data visualization Introduction, Types of data visualization, Data for visualization: Data types, Data encodings, Retinal variables, mapping variables to encodings, Visual encodings.	**6**
Unit 5: Technologies in Data Science Computer science and engineering applications Data mining, Network protocols, analysis of Web traffic, Computer security, Software engineering, Computer architecture, operating systems, distributed systems, Bioinformatics, Machine learning	**7**
Unit 6: Applications of Data Science Technologies for visualization, Bokeh (Python), recent trends in various data collection and analysis techniques, various visualization techniques, application development methods of used in data science.	**6**

Text Books:

1. Jure Leskovek, Anand Rajaraman, Jeffrey Ullman, Mining of Massive Datasets. v2.1, Cambridge University Press, 2014.
2. Data Science from Scratch: First Principles with Python, Joel Grus, O'Reilly, 1st edition, 2015.

Reference Books:

1. Doing Data Science, Straight Talk from the Frontline, Cathy O'Neil, Rachel Schutt, O' Reilly, 1st edition, 2013.
2. Mining of Massive Datasets, Jure Leskovec, Anand Rajaraman, Jeffrey David Ullman, Cambridge University Press, 2nd edition, 2014.

Online Resources:

1. https://www.youtube.com/watch?v=XohgKT13FKY&list=PLqICp9VkFcbEWeZ0Q-6gsHCRaqe5eyf
2. https://www.youtube.com/watch?v=fn1rKKNLuzk&list=PL15FRvx6P0OWTlNBS_93NHG2hIn9cy

Course Plan

Course Title: Basics of Databases	
Course Code:201DSL3O18	**Semester: VI**
Teaching Scheme: L-T-P:3-1-0	**Credits:4**
Evaluation Scheme: ISE + MSE Marks: 20 + 30	**ESE Marks:50**

Course Description:

In computing, a database is an organized collection of data stored and accessed electronically. A relational database, more restrictively, is a collection of schemas, tables, queries, views, and other elements. It defines the data models, relational models, and constraints that can be used in designing the relational database. The course also providesinsight into SQL that is used for the implementation of relational databases. The course also covers transaction processing and data recovery.

Course Objectives:

1. To develop conceptual understanding of fundamentals of database management systems.
2. To get familiar with different data models.
3. To model the database by query writing.
4. To apply functional dependency and normalization.
5. To expose the transaction processing and data recovery mechanisms.

Course Outcomes (COs):

Upon successful completion of this course, the students will be able to:

C318.1	Describe the fundamental elements of relational database management systems.	Understand
C318.2	Draw ER-models to explain the logical structure of databases for applications.	Apply
C318.3	Execute queries using SQL.	Apply
C318.4	Use the principles and practices of good database design like functional dependency and normalization.	Apply
C318.5	List different concurrency control mechanisms.	Remember
C318.6	Explain different data recovery techniques	Understand

Prerequisite:	**Data Structure**

Course Articulation Matrix: Mapping of Course Outcomes (COs) with Program Outcomes (POs) and Program Specific Outcomes (PSOs):

COs	**POs**												**PSOs**		**BTL**
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**1**	**2**	
C318.1	**2**	**2**	**-**	**-**	**-**	**-**	**-**	**-**	**-**	**-**	**-**	**-**	**-**	**-**	**2**
C318.2	**-**	**2**	**2**	**2**	**-**	**-**	**-**	**-**	**-**	**-**	**-**	**-**	**2**	**-**	**3**
C318.3	**-**	**2**	**2**	**2**	**2**	**-**	**-**	**-**	**-**	**-**	**-**	**-**	**1**	**-**	**3**
C318.4	**2**	**2**	**-**	**-**	**-**	**-**	**-**	**-**	**-**	**-**	**-**	**-**	**2**	**-**	**3**
C318.5	**2**	**-**	**-**	**-**	**-**	**-**	**-**	**-**	**-**	**-**	**-**	**-**	**1**	**-**	**1**
C318.6	**2**	**-**	**-**	**-**	**-**	**-**	**-**	**-**	**-**	**-**	**-**	**-**	**2**	**-**	**2**

Content	Hours
Unit 1: Introduction to Database System Introduction, what is Database System, Purpose of Database System, View of data, Data Models, Relational Database, Data-Manipulation Language, Data-Definition Language, Database Architecture, Transaction Management	6
Unit 2: Database Design and E-R Model Introduction, ER Relationship Model, Constraints, ER Diagram, ERD issues, Relational Schema, Logical view of data, Keys, Integrity rules, Relational database design, Features of good relational database design	6
Unit 3: Structured Query Language (SQL) Introduction to SQL, data types, DDL Statements, DML Statements, DCL Statements, Aggregate functions, Group by clause, having clause, order by clause, set operations, Joins, Nested Queries, Views	7
Unit 4: Functional Dependency & Normalization **Integrity constraints** – domain constraints, referential integrity, Pitfalls in Relational-Database Design, Functional dependency, types of functional dependency **Normalization** – Purpose of normalization, First Normal Form (1NF), Second Normal Form (2NF), Third Normal Form (3NF), Boyce-Codd Normal Form (BCNF), Fourth Normal Form (4NF), Fifth Normal Form (5NF)	6
Unit 5: Transaction Management and Concurrency Control **Transaction Processing** – Concept, ACID properties, Transaction model, Schedule, Serializability – conflict and view Serializability, Recoverable schedule. **Concurrency Control Mechanisms** – Lock based protocols, Multiple Granularity, Timestamp based protocols, Thomas's Write Rule, Validation based protocols	6
Unit 6: Deadlock Handling & Data Recovery **Deadlock Handling** – Deadlock prevention, deadlock detection and deadlock recovery. **Data Recovery** – Failure Classification, Storage, Log based recovery, checkpoints, Shadow paging	5

Text Books:

1. "Database System Concepts", Abraham Silberschatz, Henry F. Korth, S. Sudarshan, 6th edition, McGraw- Hill.
2. "Database Systems - A Practical Approach to Design, Implementation and Management", Thomas Connolly, Carolyn Begg, 4th Edition, Addison Wesley.
3. "MySQL Cookbook", Paul DuBois, 3rd edition, O'REILLY.

Reference Books

1. "Fundamentals of Database Systems", Ramez, Elmasri, Shamkant B. Navathe, 6th Edition, Addison Wesley
2. "Database Systems – Design, Implementation and Management", Rob & Coronel, 5th Edition, Thomson Course Technology

Online Resources:

https://nptel.ac.in/courses/106104135

HEAD OF DEPARTMENT
CSE - DATA SCIENCE DEPT.
D. Y Patil College of Engg. & Tech
Kasaba Bawada, Kolhapur