Estd. 1962
"A++" Accredited by NAAC(2021)
With CGPA 3.52

SHIVAJI UNIVERSITY, KOLHAPUR - 416004, MAHARASHTRA
PHONE : EPABX – 2609000,
www.unishivaji.ac.in, **bos@unishivaji.ac.in**
शिवाजी विद्यापीठ, कोल्हापूर - ४१६००४,महाराष्ट्र
दूरध्वनी - ईपीएबीएक्स - २६०९०००, अभ्यासमंडळे विभाग दुरध्वनी विभाग २३१—२६०९०९३/९४

SU/BOS/Science/(04) 28 **Date:** 30- 09- 2022
03-10-2022

To,
The Principal,
All Affiliated Concerned Science Colleges/Institutions
Shivaji University, Kolhapur.

Subject :- Regarding syllabi of M.Sc. Part- I (NEP-2020) degree programme under the Faculty of Science and Technology as per National Education Policy, 2020 .

Sir/Madam,

With reference to the subject mentioned above, I am directed to inform you that the university authorities have accepted and granted approval to the syllabi and Nature of question paper of **M.Sc. Part- I Data Science** under the Faculty of Science and Technology **as per National Education Policy 2020 .**

Sr. No.	Faculty of Science and Technology	Programme/ Course
1	Computer Science Engineering and Technology	M. Sc. Part- I Data Science

This syllabi and nature of question paper shall be implemented from the Academic Year 2022-2023 onwards. A soft copy containing the syllabus is attached herewith and it is also available on university website www.unishivaji.ac.in (students Online Syllabus)

You are, therefore, requested to bring this to the notice of all students and teachers concerned.

Thanking you,

Yours faithfully,

Dy Registrar

Copy to:

1	The Dean, Faculty of Science & Technology	7	Appointment Section
2	Director, Board of Examinations and Evaluation	8	P.G.Seminar Section
3	The Chairman, Respective Board of Studies	9	Computer Centre (I.T.)
4	B.Sc. Exam	10	Affiliation Section (U.G.)
5	Eligibility Section	11	Affiliation Section (P.G.)
6	O.E. I Section	12	P.G.Admission Section

Shivaji University, Kolhapur

NAAC "A++" Grade with CGPA 3.52

Choice Based Credit System with Multiple Entry and Multiple Exit Option (NEP-2020)

Syllabus for

Master of Science
In
Data Science

(Under Faculty of Science and Technology)

PART–I SEMESTER- I & II

(Syllabus to be implemented from Academic Year 2022-23)

Choice Based Credit System with Multiple Entry and Multiple Exit Option (NEP-2020)
M.Sc. Program Structure
M. Sc. Part -I(Level 8)

Semester – I (Duration Six Months)											
	Sr No.	**Course Code**	**Teaching Scheme**			**Examination Scheme**					
			Theory and Practical			**University Assessment (UA)**			**Internal Assessment (IA)**		
			Lectures (Per week)	**Hours (Per week)**	**Credit**	**Maximum Marks**	**Minimum Marks**	**Exam Hours**	**Maximum Marks**	**Minimum Marks**	**Exam Hours**
CGPA	1.	CC101: Basic Statistics for Data Science	4	4	4	80	32	3	20	8	1
	2.	CC102: Feature Engineering-I	4	4	4	80	32	3	20	8	1
	3.	CC103: Data Structure using Python	4	4	4	80	32	3	20	8	1
	4.	CC104: Introduction to Database Management System	4	4	4	80	32	3	20	8	1
	5.	CCPR-105: Practical-I (Data Structure using Python)	-	6	4	80	32	3	20	8	*
	6.	CCPR106: Practical – II (Database Management System)	-	6	4	80	32	3	20	8	*
	Total (A)		**16**	**28**	**24**	**480**	**-**	**-**	**120**	**-**	**-**
Non-CGPA	**1**	**AEC-107: Communicative English-I**	**2**	**2**	**2**	**-**	**-**	**-**	**50**	**20**	**2**

Semester –II (Duration Six Months)											
CGPA	**1**	CC201: Advanced Statistics for Data Science	4	4	4	80	32	3	20	8	1
	2	CC202: Feature Engineering-II	4	4	4	80	32	3	20	8	1
	3	CC203: Machine Learning	4	4	4	80	32	3	20	8	1
	4	CC204: Design and Analysis of Algorithm	4	4	4	80	32	3	20	8	1
	5	CCPR-205: Practical – III (Machine Learning using R)	-	6	4	80	32	3	20	8	*
	6	CCPR-206: Practical – IV (Python for Data Science)	-	6	4	80	32	3	20	8	*
	Total (B)		**16**	**28**	**24**	**480**	**-**	**-**	**120**	**-**	**-**
Non-CGPA	**1**	**SEC-207**	**2**	**2**	**2**	**-**	**-**	**-**	**50**	**20**	**2**
Total (A+B)					**48**	**960**	**-**	**-**	**240**	**-**	**-**

• Students Contact Hours Per Week : **56** Hours(Min.)	• Total Marks for M.Sc-I : **960**
• Theory and Practical Lectures : **60** Minutes Each	• Total Credits for M.Sc.-I (Semester- I and II) : **48**
• CC – Core Course • CCPR – Core Course Practical • AEC- Mandatory Non-CGPA compulsory Ability Enhancement Course • SEC- Mandatory Non-CGPA compulsory Skill Enhancement Course	• Practical Examination is Semester wise after theory examination. • Examination for CCPR-105 and CCPR-106 shall be based on Semester-I Practical. • Examination for CCPR-205 and CCPR-206 shall be based on Semester-II Practical. • * Duration of Practical Examination as per respective BOS guidelines. • *Separate passing is mandatory for Theory, Internal and Practical*

	Examination.

- **Requirement for Entry at Level 8 :**
- B.Sc.(Statistics / Mathematics / Electronics / Physics / Chemistry) from recognized Indian University (Level 7).
- **B.Sc.(**Computer Science), **B.C.S., B.Sc**.(Entire Computer Science), **(BC**A)**Bachelor of Computer Application** from recognized Indian University (Level 7).
- **Bachelor of Engineering**(Computer Engineering, Electronics Engineering, Electronics and Telecommunication Engineering, Mechanical Engineering and Electrical Engineering) from recognized Indian University (Level 7).

- Exit option at Level 8 : Student can exit after successful completion of level 8 with Post Graduate Diploma in Data Science if he/she completes the courses equivalent to minimum of 48 credits.

	M.Sc. I	**M.Sc. II**	**Total**
Marks	**1200**	**1200**	**2400**
Credits	**48**	**48**	**96**

The name of the programme shall be Master of Science (**M.Sc**. in Data **Science)**

Duration of the Program

- The M.Sc. programme will be a full-time two years i.e. 4 semesters. Pattern of examination will be Semester System.

Medium of Instruction

- The medium of Instruction will be English only**.**

Admission Procedure

- **B.Sc.**(Statistics / Mathematics / Electronics / Physics / Chemistry) from recognized Indian University.
- **B.Sc.(**Computer Science), **B.C.S., B.Sc**.(Entire Computer Science), **(BC**A)**Bachelor of Computer Application** from recognized Indian University.
- **Bachelor of Engineering**(Computer Engineering, Electronics Engineering, Electronics and Telecommunication Engineering, Mechanical Engineering and Electrical Engineering) from recognized Indian University.

Project work

At the end of I, II & III semester student has to carry out a project work.

1. Project work at end of semester should be done in groups, each student must be given a responsibility for a distinct module and care should be taken to see the progress of individual modules is independent of others.

2. Students should take guidance from an internal guide and prepare a Project Report on "Project Work" to be submitted to the Department after evaluation.

3. The Project Report should contain an Introduction to Project, which should clearly explain the project scope in detail. Database/Webpage/ UI designs and a list of output reports should be included along with references.

4. The project Work should be of such a nature that it could prove useful or should be relevant from the societal/commercial/research angle.

5. The project report will be duly accessed by the internal guide of the project and internal marks will be communicated by the concerned guide.

6. Project viva-voce by the University panel will be conducted as part of Evaluation.

Research Seminar

At the end of fourth semester student shall deliver seminar on one of the advanced topic chosen in consultation with the guide after compiling the information from the latest literature and also internet. The concepts must be clearly understood and presented by student. Prior to presentation, he/she shall carry out the detailed literature survey from standard references such as International & National journals and periodicals recently published reference books etc. A hard copy of the report (A4 size, 12 fonts, Times New Roman, Single spacing both side printed) should be submitted to the Department before delivering the seminar. This seminar will be evaluated internally for 100 marks by the respective guides.

Assessment

The final total assessment of the candidate is made in terms of an internal assessment and an external assessment for each course.

1. For each theory paper, 20% marks will be based on internal assessment and 80% marks for semester examination (external assessment), unless otherwise stated.
2. Internal assessment of theory papers should be in the form of two internal tests of 10 marks each. Total 20 marks.
3. The projects will be evaluated by the university appointed panel.
4. The final practical examination will be conducted by the university appointed panel at the end of semester for each lab course and marks will be submitted to the university by the panel. The pattern of final Practical Examination will be as follows-

1	Coding and Execution of Program	60 Marks
2	Viva-voce	10 Marks
3	Journal	10 Marks
	Total	80 marks

Award of Class

There will be numerical marking on each question. At the time of declaration of the result the marks obtained by the candidate is converted into classes as per University norms.

Credit system implementation As per the University norms Clarification of Syllabus

The syllabus Committee should meet at least once in a year to study and clarify any difficulties from the Institutes. The Workshop on syllabi should be organised at the beginning of every semester.

Revision of Syllabus

As the computer technology is changing very fast, revision of the syllabus should be considered every 3 years.

PEO's for M.Sc. Program

Program Educational Objectives is to prepare graduates to:

1. Apply and continuously acquire knowledge, both theoretical and applied, related to core areas of data science.
2. To develop relevant programming abilities.
3. Demonstrate the ability to build and assess data-based models.
4. Demonstrate to apply data science concepts and methods to solve problems in real-world contexts and will communicate these solutions effectively

Program Outcomes (POs)

At the end of the Master of Science (Data Science) Program, graduating students/graduates will be able to:

1. Apply conceptual knowledge for complex computing problems and arrive at a solution based on user requirements.
2. Provide analytical skills necessary for arriving at solutions to socio economic problems.
3. Work in multi-disciplinary teams to develop interpersonal skills and project management techniques.
4. Apply mathematical foundations and computer science theory in the modeling and design of software systems.
5. Integrate the concepts of networks and security issues in the required fields

Program Specific Outcomes (PSOs)

1. Build foundational understanding of key technologies and tools driving data science to develop machine learning models for businesses at global level.
2. Inculcate research skills, develop novel ideas, and discover insights in diverse domains to develop data-driven solutions for sustainable development.
3. Apply quantitative modelling techniques and develop skills to present inferences using tools that are adaptable to evolving technologies, laying the
foundation for lifelong learning.
4. Be acquainted with the contemporary issues, latest trends in technological development and thereby innovate new ideas and solutions to existing problems.

M.Sc.-I Semester-I (Data Science)

Choice Based Credit System with Multiple Entry and Multiple Exit Option
(NEP-2020)
Course Code: CC-101
Title of Course: Basic Statistics for Data Science

Syllabus to be implemented from Academic Year 2022-23

Course Outcomes(CO)
CO1 Study of basic features of the data
CO2 Study the sample using different quantitative measures
CO3 Study different forms of probability distribution
CO4 Study how to build predictive models for sample data.

Unit No.	Title of Unit and Contents
I	**Descriptive Statistics:** 1.1 Measures of Central Tendency: Mean, Median,Mode 1.2 Partition Values: Quartiles, Percentiles, BoxPlot 1.3 Measures of Dispersion: Variance, Standard Deviation,Coefficient of variation 1.4 Skewness: Concept of skewness, measures ofskewness 1.5 Kurtosis: Concept of Kurtosis, Measures ofKurtosis (All topics to be covered for raw data using R software. Manual calculations are not expected.)
II	Introduction to Probability: 2.1 Probability - classical definition, probability models, axioms of probability, probability of an event. 2.2 Concepts and definitions of conditional probability, multiplication theorem $P(A \cap B) = P(A).\ P(B|A)$ 2.3 Bayes' theorem (without proof) 2.4 Concept of Posterior probability, problems on posterior probability. 2.5 Definition of sensitivity of a procedure, specificity of a procedure. Application of Bayes' theorem to design a procedure for false positive and false negative. 2.6 Concept and definition of independence of two events. 2.7 Numerical problems related to real life situations.
III	Introduction to Random Variables 3.1 Definition of discrete random and continuous random variable. 3.2 Concept of Discrete and Continuous probability distributions. (p.m.f. and p.d.f.) 3.3 Distribution function 3.4 Expectation and variance 3.5 Numerical problems related to real life situations
IV	Special Distributions 4.1 Binomial Distribution 4.2 Uniform Distribution 4.3 Poisson Distribution 4.4 Negative Binomial Distribution

	4.5 Geometric Distribution 4.6 Continuous Uniform Distribution 4.7 Exponential Distribution 4.8 Normal Distribution 4.9 Log Normal Distribution 4.10 Gamma Distribution 4.11 Weibull Distribution 4.12 Pareto Distribution (For all the probability distributions its pmf/pdf, p-p plot, q-q plot, generation of probabilities and random samples using R software is expected.)

References

1. Fundamentals of Applied Statistics (3rd Edition),Guptaand Kapoor, S.Chandand Sons, New Delhi,1987.
2. An Introductory Statistics, Kennedy andGentle.
3. Statistical Methods, G.W. Snedecor, W.G. Cochran, John Wiley & sons,1989.
4. Introduction to Linear Regression Analysis**,** Douglas C. Montgomery, Elizabeth A. Peck, G. Geoffrey Vining,Wiley
5. Modern Elementary Statistics, Freund J.E., Pearson Publication,2005.
6. Probability, Statistics, Design of Experiments and Queuing theory with applications Computer Science, Trivedi K.S., Prentice Hall of India, NewDelhi,2001.
7. A First course in Probability 6th Edition, Ross, Pearson Publication,2006.
8. Introduction to Discrete Probability and Probability Distributions, Kulkarni M.B., Ghatpande S.B., SIPF Academy,2007.
9. A Beginners Guide to R, Alain Zuur, Elena Leno, Erik Meesters, Springer,2009
10. Statistics Using R, SudhaPurohit, S.D.Gore, ShailajaDeshmukh, Narosa, Publishing Company

M.Sc-I Semester-I (Data Science)

Choice Based Credit System with Multiple Entry and Multiple Exit Option
(NEP-2020)
Course Code: CC-102
Title of Course: Feature Engineering-I

Syllabus to be implemented from Academic Year 2022-23

Course Outcomes(CO)

CO1 Explanation of basics of Feature engineering used for representing and generating process.
CO2 Describe features of different types of data with feature selection process.
CO3 Increases your range of techniques to preprocess data and build more powerful machine learning models.
CO4 Describes multiple techniques to deal with infrequent, rare and unseen categories.

Unit No.	Title of Unit and Contents
I	**Introduction to Feature Engineering:** Motivating example- AI powered communication, Importance of feature engineering, Introduction of feature engineering, Evaluation of machine learning algorithm and Feature engineering procedures.
II	**Feature Understanding:** Feature Improvement-cleaning datasets, Feature Selection-removing bad attributes, Feature construction, Feature Transformation, Feature learning.
III	**Basics of Feature Representation:** Scalars, Vectors and spaces, Dealing with Counts, Binarization, Quantization or Binning, Log Transformation, Feature Scaling or Normalization, Min-Max Scaling, Standardization(Variance Scaling), Normalization, interaction Features, Feature Selection
IV	**Features of Text and Categorical Data:** Bag-of- X: Turning Natural Text into Flat Vectors, Filtering for cleaner Features, Atoms of Meaning: From words to n-Grams to Phrases, TF-ldf: A simple twist on Bag-of-words, Putting it to the test, Deep Dive, Encoding Categorical Variables, Dealing with Large Categorical Variables.

References:

1. Max Kuhn , Kjell Johnson, “Feature Engineering and Selection: A Practical Approach for Predictive Models” 1st Edition, Chapman & Hall/CRC Data Science Series, ISBN 13- 978-1-138-07922-9.
2. Sinan Ozdemir, Divya Susarla, “Feature Engineering Made Easy”, Packt Publishing, ISBN 978-1-78728-760-0
3. Alice Zheng & Amanda Casari, “Feature Engineering for Machine Learning: Principles and Techniques for data scientist”, Oreilly.

M.Sc.-I Semester-I (Data Science)

Choice Based Credit System with Multiple Entry and Multiple Exit Option
(NEP-2020)
Course Code: CC-103
Title of Course: Data Structure using Python

Syllabus to be implemented from Academic Year 2022-23

Course Outcomes(CO)

CO1 Describe the basics of python programming.
CO2 Explain programming constructs and apply them to build and package python modules for reusability.
CO3 Use various data structures to gain suitable knowledge about their implementation.
CO4 Compare various file handling techniques and database interactions.
CO5 Evaluate patterns , compile expressions and write scripts to extract data.
CO6 Write an application to solve real life problems by applying Object- Oriented principles.

I	**Introduction To Python and Data Types** 1.1 Introduction 1.2 VariousIDEs 1.3 Numeric data types: int, float,complex 1.4 String, list and listslicing 1.6 Tuples
II	**Control Flow, Functions, Modules And Packages** 2.1 Control Flow Conditional blocks using if, else and elif Simple for and while loops in python, For loop using ranges, string, list and dictionaries, Loop manipulation using pass, continue, break and else 2.2 Functions Arguments, Lambda Expressions, Function Annotations 2.3 Modules Organizing python projects into modules Importing own module as well as external modules. 2.4 Packages 2.5 Programming using functions, modules and external packages
III	**Data Structures** 3.1 Lists as Stacks, Queues, Comprehensions 3.2Tuples and sequences 3.3Sets, Dictionaries
IV	**Python File Operation** 4.1Reading config files inpython 4.2Writing log files inpython 4.3Understanding read functions, read(), readline() andreadlines() 4.4Understanding write functions, write() andwritelines() 4.5Manipulating file pointer usingseek 4.6 Programming using fileoperations

Reference Books:

1. Learning Python, O'Reillypublication
2. Programming Python, O'Reillypublication
3. https://docs.python.org/3/tutorial/
4. https://realpython.com/beautiful-soup-web-scraper-python

M.Sc.-I Semester-I (Data Science)

Choice Based Credit System with Multiple Entry and Multiple Exit Option
(NEP-2020)
Course Code: CC-104
Title of Course: Introduction to Database Management System

Syllabus to be implemented from Academic Year 2022-23

Course Outcomes(CO)

CO1 Describe different concepts of database management systems.
CO2 Discuss structure of relational databases and apply relational operations on it.
CO3 Apply the basic and advanced concepts of SQL language to solve the queries in the databases.
CO4 Analyse database requirements and determine the entities involved in the system and their relationship.
CO5 Write the queries to implement different functionalities of SQL language.

I	**Introduction** 1.1 Database-system Applications 1.2 Purpose of Database Systems 1.3 View of Data-Data Abstraction, Instance and Schemas 1.4 Relational Databases: Tables, DML,DDL 1.5 Data storage and querying: Storage Manager, The query processor 1.6 Database Architecture 1.7 Speciality Databases
II	**Introduction to Relational Model and SQL** 2.1 Structure of Relational Databases 2.2 Database Schema, Keys 2.3 Relational Operations Overview of SQL query language 2.4 SQL data Definition- Basic Types, Basic schema definition, Date and Time in SQL, Default values, Index creation, Large Object types, user- defined types 2.5 Integrity constraint- Constraints on a single relation, Not Null constraint, Unique constraint, The Check clause, referential integrity 2.6 Basic structure of SQL queries- Queries on single relation, queries on multiple relations, The natural join, 2.7 Additional basic operations Set operations, Null Values 2.8 Aggregate Functions-Basic aggregation, Aggregation and grouping, The Having clause, Aggregation with Null and Boolean values 2.9 Nested subqueries- Set membership, Set comparison, Test for Empty Relations, Test for Absence of Duplicate Tuples, Subqueries in the from clause, The **with** clause, Scalar subqueries 2.10 Modification of the Database- Deletion, Insertion, Updates
III	**Intermediate and advanced SQL** 3.1 Join Expressions- Join conditions, Outer joins, Join types and Conditions 3.2 Views- View definition, using views in SQL queries,

	Materialized views, update a view 3.3 Create table extensions Schemas, Catalogs and Environments 3.5 The relational Algebra, The tuple relational calculus
IV	**Database Design and E-R model** 4.1 Overview of the Design process and Entity Relationship Model 4.2 Constraints and Removing Redundant Attributes in Entity Sets 4.3 Entity Relationship Diagrams 4.4 Introduction to UML Relational database model: Logical viewof data, keys, integrity rules 4.5 Functional Dependency 4.6 Anomalies in a Databases 4.7 The normalization process: Conversion to first normal form, Conversion to second normal form, Conversion to third normal form, The Boyce-Codd Normal Form (BCNF), Fourth Normal form and fifth normal form 4.8 Normalization and database design 4.9 Denormalization

References:

1. Abraham Silberschatz, Henry F. Korth, S. Sudarashan, Database System Concepts, McGraw-Hill International Edition, SixthEdition
2. Elmasri, Navathe, Fundamentals of Database Systems, Pearson Education, Third Education
3. Ramakrishnan, Gehrke, Database Management Systems, McGrawHill International Edition, ThirdEdition
4. Peter Rob, Carlos Coronel, Database System Concepts, Cengage Learning, India Edition
5. S.K.Singh, "Database Systems Concepts, Design and Applications", First Edition, Pearson Education,2006
6. Redmond,E. & Wilson, Seven Databases in Seven Weeks: A Guide to Modern Databases and the NoSQL Movement Edition:1st Edition.

M.Sc.-I Semester-I (Data Science)

Choice Based Credit System with Multiple Entry and Multiple Exit Option (NEP-2020)

Course Code: CCPR-105

Title of Course: Practical- I(Data Structure using Python)

Syllabus to be implemented from Academic Year 2022-23

Course outcomes:

1. Write, Test and Debug Python Programs
2. Implement Conditionals and Loops for Python Programs
3. Design and implement GUI application and how to handle exceptions and files
4. Use List, Set, Tuples and Dictionaries
5. Read and write data from & to files in Python

Lab assignments based on Data Structure Using Python course.

M.Sc.-I Semester-I (Data Science)

Choice Based Credit System with Multiple Entry and Multiple Exit Option (NEP-2020)

Course Code: CCPR-106

Title of Course: Practical- II (Database Management System)

Syllabus to be implemented from Academic Year 2022-23

Course outcomes:

1. Create, modify and apply operations on table
2. Handling different keys on database
3. Handling different types of SQL queries on database
4. Handling different types of mathematical and statistical functions
5. Handling and write code using PL/SQL
6. Handling cursor in in PL/SQL

Lab assignments based on Introduction to Database Management System course.

M.Sc.-I Semester-II (Data Science)

Choice Based Credit System with Multiple Entry and Multiple Exit Option (NEP-2020)

Course Code: CC-201

Title of the Course: Advanced Statistics for Data Science

Syllabus to be implemented from Academic Year 2022-23

Course Outcomes(CO)

CO1 Identify sampling methods from the pattern of the observed data.
CO2 Predict the future behavior of the time series data.
CO3 Predict different models of forecasting of time series data.
CO4 Analyze sample data and identify the parameters and their probability distributions.
CO5 Validate the hypothesis to ensure that the entire research process remains Scientific and reliable.
CO6 Hypothesize and test an assumption regarding population parameters using Sample data.

I	**Sampling** • Introduction to Sampling • Simple random Sampling • Stratified Random Sampling • Cluster Sampling • Concept of Sampling Error

II	**Sampling Distributions**
	• Introduction to Sampling distributions
	  o Student's t distribution
	  o Chi square distribution
	  o Snedecor's F distribution
	• Interrelations among t, chi-square and F distributions
	  o Central Limit Theorem (Various Versions) and its applications.

III	**Testing of hypothesis** • Definitions: population, statistic, parameter, standard error of estimator. • Concept of null hypothesis and alternative hypothesis, critical region, level of significance, type I and type II error, one sided and two- sided tests, p-value. • Large Sample Tests • Tests based on t, Chi-square and F-distribution **All tests to be taught using R software. Manual calculations are not expected**.
IV	**Analysis of Variance** • One Way ANOVA • Two Way ANOVA • Application of ANNOVA to test the overall significance of Regression. **All topics to be covered using R software. Manual calculations are not expected**.

References:

1. Fundamentals of Applied Statistics (3rd Edition), Gupta and Kapoor, S.Chand and Sons, New Delhi, 1987.
2. Time Series Methods, Brockell and Devis, Springer, 2006.
3. Time Series Analysis,4th Edition, Box and Jenkin, Wiley, 2008.
4. Modern Elementary Statistics, Freund J.E., Pearson Publication, 2005.
5. Probability, Statistics, Design of Experiments and Queuing theory with applications Computer Science, Trivedi K.S., Prentice Hall of India, New Delhi,2001.
6. Common Statistical Tests, Kulkarni M.B., Ghatpande S.B., Gore S.D., Satyajeet Prakashan,Pune, 1999.
7. Probability and Statistical Inference, 9th Edition, Robert Hogg, Elliot Tanis, Dale Zimmerman, Pearson education Ltd, 2015
8. A Beginners Guide to R, Alain Zuur, Elena Leno, Erik Meesters, Springer, 2009
9. Statistics Using R, Sudha Purohit, S.D.Gore, Shailaja Deshmukh, Narosa, Publishing Company

M.Sc.-I Semester-II (Data Science)

Choice Based Credit System with Multiple Entry and Multiple Exit Option
(NEP-2020)
Course Code: CC-202
Title of the Course: Feature Engineering - II

Syllabus to be implemented from Academic Year 2022-23

Course Outcomes(CO)

CO1 Describe feature transformations process for converting high dimensional features to low dimensional features.
CO2 Explanation of feature learning process from the given input.
CO3 Explain the performance of machine learning algorithm.
CO4 Helps to learn techniques used in organization worldwide and in data Competitions.

I	**Feature Selection**: Importance of Feature Selection in Machine Learning, Goals of Feature Selection, Classes of Feature Selection Methodologies.
II	**Advanced Feature Selection:** Effect of Irrelevant Feature, Over fitting to Predictors and External Validation, Greedy Search Methods- Simple Filters, Recursive Feature Elimination, Stepwise Selection
III	**Feature Transformations:** Intuition, Derivation, Linear Projection, Variance and Empirical Variance - Vector Formulation, General Solution of the Principal Components, Transforming Features, Implementing PCA, PCA in Action, Whitening and ZCA, Considerations and Limitations of PCA, Use Cases
IV	**Feature Learning**: Parametric assumptions of data, Non-parametric fallacy, feature learning algorithms, Reconstructing the data, The Bernoulli RBM, Extracting PCA components from MNIST, Extracting RBM components from MNIST, Using RBMs in a machine learning pipeline, Learning text features – word vectorizations, Word embeddings, Application of word embeddings – information Retrieval

References:

1. Max Kuhn , Kjell Johnson, "Feature Engineering and Selection: A Practical Approach for Predictive Models" 1st Edition, Chapman & Hall/CRC Data Science Series, ISBN 13- 978-1-138-07922-9.
2. Sinan Ozdemir, Divya Susarla, "Feature Engineering Made Easy", Packt Publishing, ISBN 978-1-78728-760-0
3. Alice Zheng & Amanda Casari, "Feature Engineering for Machine Learning: Principles and Techniques for data scientist", Oreilly.

M.Sc.-I Semester-II (Data Science)

Choice Based Credit System with Multiple Entry and Multiple Exit Option
(NEP-2020)
Course Code: CC-203
Title of the Course: Machine Learning

Syllabus to be implemented from Academic Year 2022-23

Course Outcomes(CO)

CO1 Define a problem to find appropriate solutions in the field of data science and other interdisciplinary areas.
CO2 Classify and apply machine learning techniques to solve real world problems.
CO3 Apply various classification algorithms and evaluate their performance.
CO4 Analyze various techniques of machine learning.
CO5 Evaluate performance of machine learning models by using various performance evaluation parameters.
CO6 Construct use case based models by analyzing datasets from various domains.

I	**Introduction to Data and Machine Learning Basics** Essentials of Data and its analysis Framework of Data Analysis History of Machine Learning Machine Learning Vs Statistical Learning Types of Machine Learning Algorithms Supervised Learning Unsupervised Learning Reinforcement Learning
II	**Understanding Regression Analysis and classification Techniques** Linear Regression Multiple Regression Logistic Regression
III	**Classification Techniques and Clustering** Decision Tree SVM Naïve Bayes KNN **Clustering** K means clustering Association Rule Mining Apriori Algorithm

IV	**Model Evaluation Ensemble Methods** Introduction Performance Measures Confusion Matrix **Ensemble Methods** Introduction Bagging, Cross Validation

Reference Books:

1. Jiawei Han, Micheline Kamber, Jian Pei, Data Mining: Concepts and Techniques, 3rd Edition
2. Margaret H. Dunham, S. Sridhar, Data Mining - Introductory and Advanced Topics, Pearson Education 5. Tom Mitchell, Machine Learning‖, McGraw-Hill, 1997
3. R.O. Duda, P.E. Hart, D.G. Stork., Pattern Classification, Second edition. John Wiley and Sons, 2000.
4. Christopher M. Bishop, Pattern Recognition and Machine Learning, Springer 2006 8. Ian H. Witten, Data Mining: Practical Machine Learning Tools and Techniques, Eibe Frank Elsevier / (Morgan Kauffman)
5. Bing Liu: Web Data Mining: Exploring Hyperlinks, Contents and Usage Data, Springer (2006).
6. Soumen Chakrabarti: Mining the Web: Discovering knowledge from hypertext data, Elsevier (2003).
7. Christopher D Manning, Prabhakar Raghavan and Hinrich Schütze: An Introduction to Information Retrieval, Cambridge University Press (2009)

M.Sc.-I Semester-II (Data Science)

Choice Based Credit System with Multiple Entry and Multiple Exit Option (NEP-2020)

Course Code: CC-204

Title of the Course: Design and Analysis of Algorithm

Syllabus to be implemented from Academic Year 2022-23

Course Outcomes(CO)

CO1 Define algorithms and its properties.
CO2 Differentiate between types of algorithms based on problem solving approach.
CO3 Demonstrate major algorithms and data structures.
CO4 Analyze the asymptotic performance of algorithms.
CO5 Evaluate and select algorithmic design paradigms and methods of analysis.
CO6 Develop analytical and problem-solving skills to design algorithms.

I	**Introduction** Definition of Algorithm & its characteristics, Recursive and Non-recursive Algorithms, Time & Space Complexity, Definitions of Asymptotic Notations, Insertion Sort (examples and time complexity), Heaps & Heap Sort (examples and time complexity)
II	**Divide and Conquer** Concept of divide and Conquer, Binary Search (recursive), Quick Sort, Merge sort **Greedy Method** Fractional Knapsack problem, Optimal Storage on Tapes, Huffman codes, Concept of Minimum Cost Spanning Tree, Prim's and Kruskal's Algorithm
III	**Dynamic Programming** The General Method, Principle of Optimality, Matrix Chain Multiplication, 0/1 Knapsack Problem, Concept of Shortest Path, Single Source shortest path, Dijkstra's Algorithm, Bellman Ford Algorithm, Floyd- Warshall Algorithm, Travelling Salesperson Problem **Branch & Bound** Introduction, Definitions of LCBB Search, Bounding Function, Ranking Function, FIFO BB Search, Traveling Salesman problem Using Variable tuple.
IV	**Decrease and conquer** Definition of Graph Representation, BFS, DFS, Topological Sort/Order, Strongly Connected Components, Biconnected Component, Articulation Point and Bridge edge **Problem Classification** Basic Concepts: Deterministic Algorithm and Non deterministic, Definitions of P, NP, NP-Hard, NP-Complete problems, Cook's Theorem (Only Statement and Significance)

REFERENCES

1. Fundamentals of Computer Algorithms, Authors - Ellis Horowitz, Sartaz Sahani, Sanguthevar Rajsekaran Publication: - Galgotia Publications
2. Introduction to Algorithms (second edition) Authors: - Thomas Cormen, Charles E Leiserson, Ronald L.Rivest ,Clifford Stein ,Publication: - PHI Publication

M.Sc.-I Semester-II (Data Science)

Choice Based Credit System with Multiple Entry and Multiple Exit Option (NEP-2020)

Course Code: CCPR-205

Title of Course: Practical – III(Machine Learning using R)

Syllabus to be implemented from Academic Year 2022-23

Course outcomes:

1. It will help students to build Reinforcement Learning.
2. It will help students to build Deep Learning
3. It will help students to build Statistical Learning Theory

---

Lab assignments based on Machine Learning

M.Sc.-I Semester-II (Data Science)

Choice Based Credit System with Multiple Entry and Multiple Exit Option (NEP-2020)

Course Code: CCPR-206

Title of Course: Practical- IV (Python for Data Science)

Syllabus to be implemented from Academic Year 2022-23

Course outcomes:

1. Design and implement efficient algorithms for a specified application.
2. Strengthen the ability to identify and apply the suitable algorithm for the given real world problem.

Lab assignments based on Design and Analysis of Algorithm