

**Computer Science Courses for Undergraduate Programme of study with **Computer Science** discipline Elective**

**DISCIPLINE SPECIFIC ELECTIVE COURSE: Data Mining-II**

**Credit distribution, Eligibility and Pre-requisites of the Course**

Course title & Code	Credits	Credit distribution of the course			Eligibility criteria	Pre-requisite of the course
		Lecture	Tutorial	Practical/ Practice		
<b>Data Mining-II</b>	<b>4</b>	<b>3</b>	<b>0</b>	<b>1</b>	Pass in Class XII	DSC01 Programming using Python, / GE1b Programming with Python / A1 Programming Fundamentals using Python ,Data Mining-I

**Learning Objectives**

The course introduces the students to the important supervised and unsupervised learning techniques. Students will learn about the importance of ensemble methods, cluster analysis, anomaly detection and their applicability in mining patterns in real applications. At the end students will be exposed to two advanced topics: text mining and time-series mining. Students will use the learned topics in solving real applications using Open-source software.

**Learning outcomes**

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

- Differentiate between partition-based, density-based and hierarchical clustering.
- Build ensemble models to improve predictive performance of the classifier.
- Identify anomalies and outliers using supervised and unsupervised techniques.
- Analyze time-series data and extract patterns from the stamped data.
- Mine textual data and do topic modelling.

**SYLLABUS OF DSE**

### **Unit 1 (10 hours)**

**Clustering:** Partitioning Methods, Hierarchical Methods, Density-Based Methods, Comparison of different methods

### **Unit 2 (8 hours)**

**Ensemble Methods:** Need of ensemble, Random Forests, Bagging and Boosting

### **Unit 3 (10 hours)**

**Anomaly Detection:** Outliers and Outlier Analysis, Outlier Detection Methods, Statistical Approaches, Proximity-based and density-based outlier detection, Clustering-based approaches

### **Unit 4 (7 hours)**

**Mining Text Data:** Document Preparation and Similarity, Clustering Methods for Text, Topic Modeling

### **Unit 5 (10 hours)**

**Stream Mining:** Time series basics, Date Ranges, Frequencies, and Shifting, Resampling and moving windows functions, Decay function, Clustering stamped data: STREAM and CluStream

### **Essential/recommended readings**

1. Tan P.N., Steinbach M, Karpatne A. and Kumar V. *Introduction to Data Mining*, 2<sup>nd</sup> edition, Pearson, 2019.
2. Zaki M. J. and Meira J. Jr. *Data Mining and Machine Learning: Fundamental Concepts and Algorithms*, 2<sup>nd</sup> edition, Cambridge University Press, 2020.
3. Aggarwal C. C. *Data Mining: The Textbook*, Springer, 2015.

### **Additional References**

1. Han J. Kamber M. and Pei J. *Data Mining: Concepts and Techniques*, Morgan Kaufmann Publishers, 2011.
2. Dunham M. *Data Mining: Introductory and Advanced Topics*, Pearson, 2006.

### **Suggested Practical List : (30 Hours)**

#### **Practical exercises such as**

Use a dataset of your choice from Open Data Portal ([https:// data.gov.in/](https://data.gov.in/), UCI repository or a dataset verified by the teacher) or load from scikit, seaborn library for the following exercises to practice the concepts learnt.

1. Apply Partitioning Methods, Hierarchical Methods, Density-Based Methods for clustering on a data set and compare the performance of the obtained results using different metrics.
2. Create an ensemble using Random Forest and show the impact of bagging and boosting on the performance.

3. Apply different outlier-detection methods on a noisy dataset and compare their effectiveness in terms of outliers reported.
4. Compute similarity between two documents after required document preparation.
5. Considering a time-stamped data (sales data/weather data), compare the aggregate values visually using different moving windows function.
6. Write a program to find the latent topics in a document using any topic modeling method and display top 5 terms that contribute to each topic along with their strength. Also, visualize the distribution of terms contributing to the topics.

**Project:** Students are encouraged to work on a good dataset in consultation with their faculty and apply the concepts learned in the course.

### DISCIPLINE SPECIFIC ELECTIVE COURSE: Data Privacy

#### Credit distribution, Eligibility and Pre-requisites of the Course

Course title & Code	Credits	Credit distribution of the course			Eligibility criteria	Pre-requisite of the course
		Lecture	Tutorial	Practical/ Practice		
<b>Data Privacy</b>	<b>4</b>	<b>3</b>	<b>0</b>	<b>1</b>	Pass in Class XII	NIL

#### Learning Objectives

This course aims to provide students with an ability to identify privacy related aspects of data uses (including attacks on data privacy), evaluate proposed mechanisms for privacy protection and relate to ethical issues related to data privacy.

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

- Understand the basic principles of data privacy and the implications of data breaches.
- Identify and evaluate different methods of protecting sensitive data.
- Explain the role of privacy regulations in safeguarding personal information.
- Implement basic cryptographic techniques to secure data.
- Apply data anonymization techniques to protect personal information.
- Analyze the ethical considerations in data privacy

## SYLLABUS OF DSE

### Unit 1 (10 hours)

**Introduction to Data Privacy and Privacy Regulations:** Notion of data privacy, Historical context of data privacy, Types of sensitive data, Privacy laws and regulations.

### Unit 2 (15 hours)

**Data Privacy Attacks, Cryptography and Data Protection:** Type of Attacks/ Data Breaches on Data Privacy, Impact of Data Breaches / Attacks, Introduction to cryptography, Symmetric and asymmetric encryption, Hashing and digital signatures.

### Unit 3 (10 hours)

**Data Collection, Use and Reuse:** Harms Associated with Data collections, use and reuse, Introduction to data anonymization, Data Anonymization Techniques for anonymizing data, Challenges in anonymizing data

### Unit 4 (10 hours)

**Ethical considerations in Data Privacy:** Privacy and Surveillance, Ethics of Data Collection and Use, Bias and discrimination in data analysis

### Essential/recommended readings

1. Ronald Leenes, Rosamunde van Brakel, and Serge Gutwirth: *Data Protection and Privacy: The Age of Intelligent Machines*, Hart Publishing, 2017.
2. Naavi: *Personal Data Protection Act of India (PDP Act 2019): Be Aware, Be Ready and Be Compliant*, Notion Press, 2020.
3. Ravinder Kumar Gaurav Goyal, *The Right to Privacy in India: Concept and Evolution*, Publisher: Lightning Source, 2016.

### Additional References

1. [https://onlinecourses.nptel.ac.in/noc22\\_cs37/preview](https://onlinecourses.nptel.ac.in/noc22_cs37/preview)
2. <https://www.coursera.org/learn/northeastern-data-privacy/home/info>

### Suggested Practical List : (30 Hours)

#### Practical exercises such as

1. Data Privacy Audit: Students can conduct a data privacy audit of a company or organization to identify potential vulnerabilities and risks in their data privacy practices.

2. Privacy Impact Assessment: Students can conduct a privacy impact assessment (PIA) of a new technology or system to identify potential privacy risks and develop strategies to mitigate them.
3. Regulation Compliance: Students can explore the requirements of the Data Protection Regulations and develop a plan for ensuring compliance with the regulation.
4. Cryptography: Students can learn about different cryptographic techniques and tools, such as encryption, hashing, and digital signatures, and implement them in practice.
5. Anonymization Techniques: Students can learn about data anonymization techniques, such as k-anonymity, differential privacy, and data masking, and apply them to a real-world dataset.
6. Privacy Policy Analysis: Students can analyze the privacy policies of different companies and identify gaps or areas for improvement.
7. Privacy-Enhancing Technologies: Students can explore privacy-enhancing technologies (PETs), such as virtual private networks (VPNs), Tor, and secure messaging apps, and evaluate their effectiveness in protecting privacy.
8. Privacy Breach Response Plan: Students can develop a privacy breach response plan for a company or organization, including steps to take in the event of a data breach and strategies for communicating with affected parties.
9. Ethical Considerations: Students can explore ethical considerations in data privacy, such as the balance between privacy and security, the impact of data collection and analysis on marginalized communities, and the role of data ethics in technology development.
10. Case Studies: Students can analyze case studies of privacy breaches or successful privacy protection strategies, and identify key lessons and takeaways.

### DISCIPLINE SPECIFIC ELECTIVE COURSE: Unix Network Programming

#### Credit distribution, Eligibility and Pre-requisites of the Course

Course title & Code	Credits	Credit distribution of the course			Eligibility criteria	
		Lecture	Tutorial	Practical/ Practice		
<b>Unix Network Programming</b>	<b>4</b>	<b>3</b>	<b>0</b>	<b>1</b>	Pass in Class XII	DSC 04 Object Oriented Programming with C++/ GE 1a Programming using C++ / GE1b Programming with Python/ DSC 01 Programming using Python/ GE 3b: Java Programming

## Learning Objectives

This course introduces the concepts of Internet protocols, ports used during communication, Client/Server concepts and various transport protocols used in computer network applications and services. The objective is to equip the students with technical knowledge of it comprises of the study of the sockets used with TCP and UDP.

## Learning outcomes

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

- Describe and analyze the various Internet Transport layer protocols used in TCP AND UDP.
- Comprehend the concepts and structures of both TCP based connection-oriented and UDP based connectionless client server applications.
- Write various real-life client-server applications using socket programming.
- Modify, maintain and extend the present internet client-server applications and write any new type of internet applications to suit the current needs of Internet users.

## SYLLABUS OF DSE

### Unit 1 (6 hours)

**Introduction** Basics of Client Server applications, Example of day time client server, concurrent servers, protocols, sockets, port numbers.

### Unit 2 (17 hours)

**Connection-oriented Socket Applications:** Elementary TCP sockets – Socket, connect, bind, listen, accept, fork and exec function, close function, Socket Address Structures, Byte Ordering and Manipulation Functions, TCP Client and Server for Echo, Signal Handling in case of crashing and rebooting of server, Shutdown process function. Socket Options: Getsockopt and stockpot functions, Socket states, Generic socket option.

### Unit 3 (15 hours)

**Connectionless Socket Applications:** TCP-oriented basic concurrent client server applications, UDP oriented Echo client and server application, Handling of errors like lost datagram, Lack of flow control with UDP, determining outgoing interface with UDP.

### Unit 4 (7 hours)

**Elementary name and Address conversions:** Domain Name System, socket functions like gethostbyname, gethostbyname2, gethostbyaddr function, uname function, gethostname function, getservbyname and getservbyport functions.

## Essential/recommended readings

1. W. Richard Stevens, Bill Fenner, Andrew M. Rudoff, *Unix Network Programming*, The sockets Networking API, Vol. 1, 3rd Edition, PHI.
2. B. A. Forouzan: *Data Communications and Networking*, THM Publishing Company Ltd
3. R. Stevens, *Unix Network Programming*, PHI 2nd Edition

### **Suggested Practical List : (30 Hours)**

#### **Practical exercises such as**

1. Implement TCP Echo client and TCP Echo server (Iterative).
2. Implement TCP Echo client and TCP Echo server (Concurrent).
3. Implement TCP daytime client and TCP daytime server (Iterative).
4. Implement TCP daytime client and TCP daytime server (concurrent).
5. Implement UDP Echo Client and UDP Echo Server.
6. Implement UDP daytime Client and UDP daytime server.
7. Implement TCP client and server (concurrent) where client gets input from the user and sends it to server. Server displays it on the screen. Server then gets another input from the user and sends it to client. Client displays it on the screen. The process continues till server or client sends "bye" to the other party.
8. Implement TCP client and server (concurrent) where client requests server to transfer a file. Assume file is smaller than 1K size. If the file is present on the server, it is sent to the client otherwise an error message is sent to client. Client copies the file on the hard disk and disconnects.
9. Implement UDP client and UDP server where server displays the IP address and port number of the client sending the datagram. Client sends a datagram (size 64 bytes) three times to the same server. Server sends the message back to client. Client reports the time elapsed in sending and receiving of the message. Use connected UDP sockets.
10. Write a program to
  - a) display name of the host
  - b) all IP addresses of the host.
  - c) Check whether FTP and HTTP services are running on the system.
  - d) Display the name of the service running on port number specified by user.