# Department of Computer Science and Engineering
# National Institute of Technology Calicut

Kozhikode - 673 601, Kerala, India

**Tentative Course Details - Winter 2023-24**
**CS4038D Data Mining**

(The instructor reserves the right to adjust the syllabus as and when required)

| | |
|---|---|
| Code | : CS4038D |
| Title | : Data Mining |
| Credit | : 4 Credits |
| Slot | : E1 and E2 |
| hours | : Tuesday 5:00 PM - 5:50 PM , Tuesday 12:00 PM - 12:50 PM (+ slot), Thursday 8:00 AM - 8:50 AM, Friday 5:00 PM - 5:50 AM |
| Venue | : ELHC 403 |

**Instructor:**


| | |
|---|---|
| Name | : Chandramani Chaudhary |
| Office | : MB 104 |
| Office Hours | : By appointment |
| Email | : chandramanic@nitc.ac.in |


**Course Objective:**
To develop a comprehensive understanding of fundamental data mining concepts, techniques, and algorithms, and acquire practical skills to apply them effectively for extracting valuable insights from large datasets and solving real-world problems.

**Course Outcomes:**

- **CO1**: Understand Core Data Mining Concepts
- **CO2**: Apply Data Mining Techniques
- **CO3**: Modify existing algorithms to suit large datasets

**Course Syllabus:**

**Module 1: (10T+7P Hours)**
Introduction to data mining-challenges and tasks Data preprocessing data analysis, measures of similarity and dissimilarity, Data visualization concepts and techniques

**Module 2: (10T+7P Hours)**
Classification- decision tree-performance evaluation of the classifier, comparison of different classifiers, Rule based classifier, Nearest-neighbor classifiers-Bayesian classifiers-support vector machines, Class imbalance problem

**Module 3: (10T+6P Hours)**
Association analysis frequent item generation rule generation, evaluation of association patterns

**Module 4: (9T+6P Hours)**
Cluster analysis,-types of clusters, K means algorithm, cluster evaluation, application of data mining to web mining and Bioinformatics

**References:**

- P. Tan, M. Steinbach, and V. Kumar, Introduction to Data Mining, Pearson Education 2006.
- J. Han and M. Kamber, Data Mining: Concepts and Techniques, 2/e, Morgan Kaufmann, 2005.
- T. Hastie, R. Tibshirani, and J. Friedman, The Elements of Statistical Learning - Data Mining, Inference, and Prediction, 2/e, Springer, California, 2008.

**Grading & Evaluation policy:**

| | Weightage | Date |
|---|---|---|
| **Mid Exam** | 20% | Feb 21 2024 - March 05 2024 |
| **Reading Assignment** | 20% | |
| **Project** | 20% | |
| **EndSem** | 40% | April 29 2024 - May 10 2024 |

**Grading Policies:**

- Grading will be relative.
- Tentative marks to grade mapping are as follows: 90-100:S; 80-89:A; 70-79:B; 60-69:C; 50-59:D; 40-49:E; <40:F.
- Absence for exams/quizzes/Assignments/Surprise tests/Programming tests without prior written permission from the respective Instructor / Faculty would be equivalent to zero marks in the corresponding exam/Quiz/Assignment/Surprise test/Programming test.
- There will be no makeup exams except in case of genuine reasons and institute rules regarding the same will be followed.
- All issues regarding the valuation of exams and quizzes must be resolved within one week after the announcement of the result.

**Attendance**:
As per institute norms.

**Standard of Conduct:**

- Each student is expected to adhere to high standards of ethical conduct, especially those related to cheating and plagiarism.
- Any work submitted MUST BE individual effort.
- Any academic dishonesty will result in zero marks in the corresponding exam or quiz and will be reported to the department council for record keeping and for permission to assign an F grade in the course.
- The department policy on academic integrity can be found at:
https://minerva.nitc.ac.in/?q=node/650