**Name of Institute: Indus Institute of Technology & Engineering**

**Name of Faculty: Ms. Urvi Rabara, Ms. Sweta Rathod, Ms. Foram Gohel**

**Course code: CS0701**
**Course name: Machine Learning**
Prerequisites:- Python
Credit point:- 4
Offered Semester: VII

## Course Coordinator

Full Name: Ms. Urvi Rabara
Department with siting location: Computer Science & Engineering (4rd floor, Faculty room, Bhanwar building)
Telephone: 9429042918
Email: urvirabara.ce@indusuni.ac.in
Consultation times: Friday (3:00 PM to 4:00 PM)

Full name: Ms. Sweta Rathod
Department with siting location: Computer Science & Engineering (4rd floor, Faculty room, Bhanwar building)
Telephone: 8160693723
Email: swetarathod.ce@indusuni.ac.in
Consultation times: Friday (3:00 PM to 4:00 PM)

**Course lecturer**

Full name: Ms. Foram Gohel
Department with siting location: Computer Science & Engineering (4rd floor, Faculty room, Bhanwar building)
Telephone: 9104749763
Email: foramgohel.ce@indusuni.ac.in
Consultation times: Friday (3:00 PM to 4:00 PM)

Students will be contacted throughout the Session via Mail with important information relating to this Course.

## Course Objectives

By participating in and understanding all facets of this Course a student will:

1. Understand the key algorithms and theory that form the foundation of Machine Learning.
2. Understand a wide variety of learning algorithms.
3. Recognize the characteristics of machine learning that make it useful to real-world problems.
4. Understand how to perform evaluation of learning algorithms and model selection.
5. Develop skills of using recent machine learning software in order to solve practical problems.
6. Understand and learn state of the art machine learning techniques to provide employability in industry.

## Course Outcomes (CO)

After successful completion of the course, student will able:

1. Get exposure of machine learning concepts and range of problems that can be handled by machine learning
2. Compare and parameterize different learning algorithms
3. Apply the machine learning concepts in real life problems
4. Understand learning in machines with different techniques
5. Understand and apply various recognition techniques.
6. Learn about parameter selection and feature extraction.

## Course Outline

| UNIT-I [12 Hours] |
|---|
| **Introduction**<br>Learning Problems, designing a learning system, Issues with machine learning. Concept Learning, Version Spaces and Candidate Eliminations, Inductive bias, Supervised/Unsupervised Learning, Loss functions and generalization, Parametric vs Non-parametric methods, Evaluating Machine Learning algorithms and Model Selection, Introduction to Statistical Learning Theory, Ensemble Methods, Bagging, Boosting, Random Forest |
| **UNIT-II [12 Hours]** |
| **Supervised Learning (Regression/Classification)** |
| Basic methods: Distance-based methods, Nearest-Neighbors, Decision Trees, Naive Bayes, Linear models: Linear Regression, Logistic Regression, Generalized Linear Models, Support Vector Machines, Nonlinearity and Kernel Methods, Beyond Binary Classification: Multi-class/Structured Outputs, Ranking |
| **Unsupervised Learning** |
| Clustering: K-means/Kernel K-means, Dimensionality Reduction -PCA, CCA, LDA, ICA, MNF – Canonical Variates - Feature Selection vs Feature Extraction, Generative Models (mixture models and latent factor models) |
| **UNIT-III [12 Hours]** |
| **Bayesian Learning** |
| Bayes Theorem, Bayes Theorem and Concept Learning, Maximum Likelihood and Least squared Error Hypothesis, Maximum likelihood hypothesis for Predicting Probabilities, Minimum Description Length, Principle, Bayes Optimal Classifier, Gibbs Algorithm, Naïve Bayes, Classifier, Bayesian Belief Network, EM Algorithm, Case Study: Learning to classify text |
| **UNIT-IV [12 Hours]** |
| **Artificial Neural networks**<br>Neural Network Representation, Appropriate problems for Neural Network Learning, Perceptron, Multilayer Networks and Back Propagation, Algorithms, Remarks on Back Propagation Algorithms, Case Study: face Recognition |
| **Advanced topics** |
| Semi-supervised, Active Learning, Reinforcement Learning, Recent trends in various learning techniques of machine learning and classification methods, Overview of typical application areas, such as Recommender System |

### Method of delivery
Chalk and Board, PowerPoint presentation, Model generation, demonstration of devices, cables

### Study time
3 hrs theory, 2 Hrs practical

### CO-PO Mapping (PO: Program Outcomes)
Engineering Graduates will be able to:

**PO1 Engineering knowledge:** Apply the knowledge of mathematics, science, engineering fundamentals, and an engineering specialization to the solution of complex engineering problems.

**PO2 Problem analysis:** Identify, formulate, review research literature, and analyze complex engineering problems reaching substantiated conclusions using first principles of mathematics, natural sciences, and engineering sciences.

**PO3 Design/development of solutions:** Design solutions for complex engineering problems and design system components or processes that meet the specified needs with appropriate consideration for the public health and safety, and the cultural, societal, and environmental considerations.

**PO4 Conduct investigations of complex problems:** Use research-based knowledge and research methods including design of experiments, analysis and interpretation of data, and synthesis of the information to provide valid conclusions.

**PO5 Modern tool usage:** Create, select, and apply appropriate techniques, resources, and modern engineering and IT tools including prediction and modeling to complex engineering activities with an understanding of the limitations.

**PO6 The engineer and society:** Apply reasoning informed by the contextual knowledge to assess societal, health, safety, legal and cultural issues and the consequent responsibilities relevant to the professional engineering practice.

**PO7 Environment and sustainability:** Understand the impact of the professional engineering solutions in societal and environmental contexts, and demonstrate the knowledge of, and need for sustainable development.

**PO8 Ethics:** Apply ethical principles and commit to professional ethics and responsibilities and norms of the engineering practice.

**PO9 Individual and team work:** Function effectively as an individual, and as a member or leader in diverse teams, and in multidisciplinary settings.

**PO10 Communication:** Communicate effectively on complex engineering activities with the engineering community and with society at large, such as, being able to comprehend and write effective reports and design documentation, make effective presentations, and give and receive clear instructions.

**PO11 Project management and finance:** Demonstrate knowledge and understanding of the engineering and management principles and apply these to one's own work, as a member and leader in a team, to manage projects and in multidisciplinary environments.

**PO12 Life-long learning:** Recognize the need for, and have the preparation and ability to engage in independent and life-long learning in the broadest context of technological change.

## Mapping of CO with PO's

| | PO1 | PO2 | PO3 | PO4 | PO5 | PO6 | PO7 | PO8 | PO9 | PO10 | PO11 | PO12 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| CO1 | √ | √ | √ | | | | | | | | √ | √ |
| CO2 | | √ | √ | √ | √ | √ | | | √ | √ | | |
| CO3 | √ | √ | √ | √ | √ | √ | √ | | | | √ | √ |
| CO4 | √ | √ | √ | √ | √ | | | | | | √ | |
| CO5 | √ | √ | √ | | | | | | | | | √ |
| CO6 | | | √ | √ | √ | | | | | | | |

## Bloom's Taxonomy and Knowledge retention

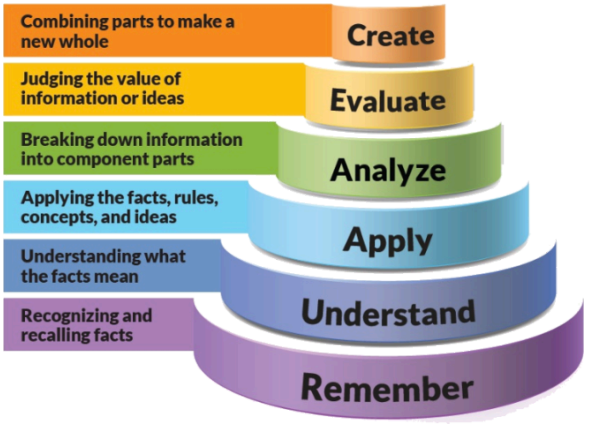


Figure 1: Bloom's Taxonomy

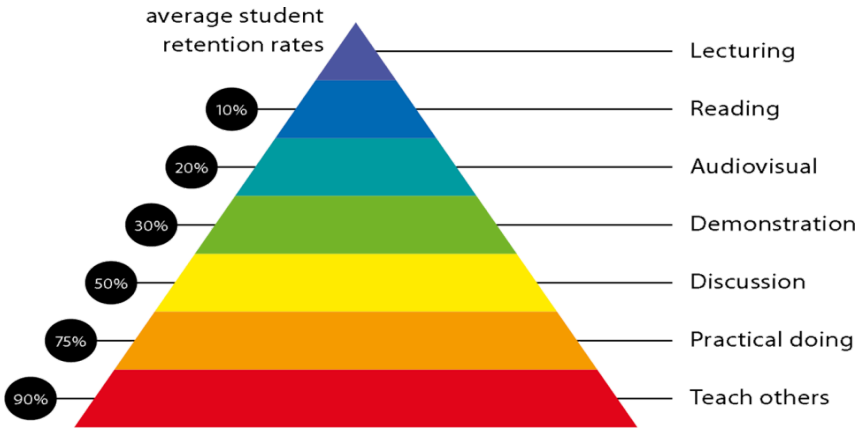


Figure 2: Knowledge retention

## Graduate Qualities and Capabilities covered

| General Graduate Qualities | Specific Department of CSE Graduate Capabilities |
|---|---|
| **Informed**<br>Have a sound knowledge of an area of study or profession and understand its current issues, locally and internationally. Know how to apply this knowledge. Understand how an area of study has developed and how it relates to other areas. | **1 Professional knowledge, grounding & awareness** |
| **Independent learners**<br>Engage with new ideas and ways of thinking and critically analyze issues. Seek to extend knowledge through ongoing research, enquiry and reflection. Find and evaluate information, using a variety of sources and technologies. Acknowledge the work and ideas of others. | **2 Information literacy, gathering & processing** |
| **Problem solvers**<br>Take on challenges and opportunities. Apply creative, logical and critical thinking skills to respond effectively. Make and implement decisions. Be flexible, thorough, innovative and aim for high standards. | **4 Problem solving skills** |
| **Effective communicators**<br>Articulate ideas and convey them effectively using a range of media. Work collaboratively and engage with people in different settings. Recognize how culture can shape communication. | **5 Written communication**<br>**6 Oral communication**<br>**7 Teamwork** |
| **Responsible**<br>Understand how decisions can affect others and make ethically informed choices. Appreciate and respect diversity. Act with integrity as part of local, national, global and professional communities. | **10 Sustainability, societal & environmental impact** |

## Practical work:

| Sr. No | Title |
|---|---|
| 1 | Study of Python Basic Libraries like a Numpy. Perform the following task using Numpy library.<br><br>- Creating blank array, with predefined data, with pattern specific data<br>- Slicing and Updating elements,<br>- Shape manipulations<br>- Looping over arrays.<br>- Reading files in numpy |
| 2 | Study of Python Libraries for ML application such as Pandas and Matplotlib Perform the following task using Pandas & Matplotlib library.<br><br>Pandas<br>- Creating data frame<br>- Reading files<br>- Slicing manipulations<br>- Exporting data to files<br>- Columns and row manipulations with loops<br>- Use pandas for masking data and reading in Boolean format.<br><br>Matplotlib<br>- Importing matplotlib<br>- Simple line chart<br>- Correlation chart<br>- Histogram<br>- Plotting of Multivariate data<br>- Plot Pi Chart |
| 3 | Implement and demonstrate the FIND-S Algorithm for finding the most specific hypothesis based on a given set of training data samples. Read the training data from a .CSV fi*le* |
| 4 | Write a Python program to implement Simple Linear Regression<br>- How many total observations in data?<br>- How many independent variables?<br>- Which is a dependent variable?<br>- Quantify the goodness of your model and discuss steps taken for improvement (RMSE, SSE, R2Score). |
| 5 | Implementation of Multiple Linear Regression for House Price Prediction using sklearn. |

| | |
|---|---|
| 6 | Two Class Classification (Logistic Regression)<br>- How many total observations in data?<br>- How many independent variables?<br>- Which is a dependent variable?<br>- Implement logistic function.<br>- Implement Log-loss function.<br>- Quantify the goodness of your model and discuss steps taken for improvement (Accuracy, Confusion matrices, F-measure). |
| 7 | Implementation of Decision tree using sklearn and its parameter tuning. |
| 8 | Write a program to implement K-mean clustering in python. |
| 9 | Write a program to implement Random Forest Algorithm |
| 10 | Write a program to implement the naïve Bayesian classifier for a sample training data set stored as a .CSV file. Compute the accuracy of the classifier |
| 11 | Write a program to implement OR, AND gate using Perceptron with learning rules. |
| 12 | Build an Artificial Neural Network by implementing the Backpropagation Algorithm |

## Attendance Requirements

The University norms state that it is the responsibility of students to attend all lectures, tutorials, seminars and practical work as stipulated in the Course outline. Minimum attendance requirement as per university norms is compulsory for being eligible for mid and end semester examinations.

## Text books

1. Tom M Mitchell, “Machine Learning”, McGraw Hill.
2. Peter Harrington, “Machine Learning in Action”, DreamTech.

## Reference Books:

1. Henrik Brink, Joseph Richards, Mark Fetherolf, “Real-World Machine Learning”, DreamTech
2. 2. Christopher Bishop, “Pattern Recognition and Machine Learning, ”Hastie, Tibshirani, and Friedman, “Elements of Statistical Learning”. Springer
3. Jiawei Han and Michelline Kamber, “Data Mining: Tools and Techniques”, 3rdEdition.
4. I H Witten, Eibe Frank, Mark A Hall, “Data Mining: A practical Machine LearningTools and techniques”, Elsevier

### Web Resources:

1. Couresera.org: Machine Learning by Andrew Ng, Stanford University

## ASSESSMENT GUIDELINES

Your final course mark will be calculated from the following:

| | |
|---|---|
| **CIE - Theory (60 Marks)**<br>**Mid Semester Exam : 40 Marks**<br>**Regularity + Class Performance : 5 Marks**<br>**Assignment/ Quiz : 15 Marks** | **CIE - Practical (60 Marks)**<br>**Lab Regularity : 5 marks**<br>**AWS (ML) certification course : 30**<br>**Lab Performance + Practical file : 15 Marks**<br>**Viva / Quiz : 10 Marks** |
| **ESE-Theory- 40 Marks** | **ESE-Practical-40 Marks** |
| **Total: 100 Marks** | **Total: 100 Marks** |

## SUPPLEMENTARY ASSESSMENT

Students who receive an overall mark less than 40% in mid semester or end semester will be considered for supplementary assessment in the respective components (i.e mid semester or end semester) of semester concerned. Students must make themselves available during the supplementary examination period to take up the respective components (mid semester or end semester) and need to obtain the required minimum 40% marks to clear the concerned components.

### Practical Work Report/Laboratory Report:

A report on the practical work is due the subsequent week after completion of the class by each group.

### Late Work

Late assignments will not be accepted without supporting documentation. Late submission of the reports will result in a deduction of -% of the maximum mark per calendar day

### Format

All assignments must be presented in a neat, legible format with all information sources correctly referenced. **Assignment material handed in throughout the session that is not neat and legible will not be marked and will be returned to the student.**

### Retention of Written Work

Written assessment work will be retained by the Course coordinator/lecturer for two weeks after marking to be collected by the students.

### University and Faculty Policies

Students should make themselves aware of the University and/or Faculty Policies

regarding plagiarism, special consideration, supplementary examinations and other educational issues and student matters.

**Plagiarism** - Plagiarism is not acceptable and may result in the imposition of severe penalties. Plagiarism is the use of another person's work, or idea, as if it is his or her own - if you have any doubts at all on what constitutes plagiarism, please consult your Course coordinator or lecturer. Plagiarism will be penalized severely.

***Do not copy the work of other students.***
***Do not share your work with other students (except where required for a group activity or assessment)***

## Course schedule (subject to change)

| Week# | Topic & contents | CO Addressed | Teaching Learning Activity (TLA) |
|---|---|---|---|
| **Week 1** | Introduction Learning Problems, designing a learning system, Issue with machine learning. Concept Learning, Version Spaces and Candidate Eliminations. | CO1 | Chalk & Board, PPT |
| **Week 2** | Inductive Bias, Supervised/Unsupervised Learning, Loss functions and generalization, Parametric vs Non-parametric methods, Evaluating Machine Learning algorithms and Model Selection. | CO1, CO2 | Chalk & Board, PPT |
| **Week 3** | Introduction to Statistical Learning Theory, Ensemble Methods, Bagging, Boosting, Random Forest | CO4 | Chalk & Board, PPT |
| **Week 4** | Supervised Learning (Regression/ Classification) basic method: Distance-based method: Nearest-Neighbour, Decision Trees, Naive Bayes, Linear models: Linear Regression, Logistic Regression, Generalized Linear Models | CO3, CO4, CO5 | Chalk & Board, PPT |
| **Week 5** | Support Vector Machines, Nonlinearity and Kernel Methods, Beyond Binary Classification: Multi class/Structured Outputs, Ranking Unsupervised Learning<br><br>Clustering: K-means/Kernel K-means Dimensionality Reduction -PCA, CCA, LDA, ICA, MNF | CO4, CO5 | Chalk & Board, PPT |

| | | | |
|---|---|---|---|
| **Week 6** | Canonical Variates - Feature Selection vs Feature Extraction, Generative Models (mixture models and latent factor models) | CO5 | Chalk & Board, PPT |
| **Week 7** | Bayesian Learning<br><br>Bayes Theorem, Bayes Theorem and Concept Learning, Maximum Likelihood and Least squared Error Hypothesis, Maximum likelihood hypothesis for Predicting Probabilities. | C02, C03, C05 | Chalk & Board, PPT |
| **Week 8** | Minimum<br><br>Description Length, Principle, Bayes Optimal Classifier, Gibbs Algorithm, Naïve Bayes, Classifier. | CO4,CO6 | Chalk & Board, PPT |
| **Week 9** | Bayesian Belief Network, EM Algorithm, Case Study: Learning to classify text. | CO5 | Chalk & Board, PPT |
| **Week 10** | Artificial Neural networks<br>Neural Network Representation, Appropriate problems for Neural Network Learning, Perceptron,<br>Multilayer Networks and Back Propagation, | C02, CO3, CO5, CO7 | Chalk & Board, PPT |
| **Week 11** | Algorithms, Remarks on Back Propagation Algorithms, Case Study: face Recognition Advanced topic Semi – Supervised,, ActiveLearning, Reinforcement Learning. | CO2, CO5, CO7 | Chalk & Board, PPT |
| **Week12** | Recent trends in various learning techniques of machine learning and classification methods, Overview of typical application areas, such Recommender System. | CO2, CO5, CO7 | Chalk & Board, PPT |

# Program Map for Bachelor of Engineering (CE/ CSE/ IT)

## COMPUTER ENGINEERING DEPARTMENT COURSE DEPENDANCY CHART