MANIPAL UNIVERSITY JAIPUR
School Of COMPUTER SCIENCE AND ENGINEERING

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

Course Hand-out

Data Science and Machine Learning | CS 3203 | 3 Credits | 3 0 0 3

Session: Jan 24-May 24 | Faculty: Dr. Sandeep Chaurasia/ Dr. Neha Chaudhary/ Mr Vivek Singh Sikarwar/ Mr Abhay Bisht/ Dr Atul Kumar Verma/ Dr Arvind Kumar/ Mr Mahesh Kr Joshi/ Ms. Shikha Mundra/ Mr. Tarun Jain

A. Introduction: This course is offered by Computer Science and Engineering, targeting students who wish to pursue development and research in industries or higher studies in field of Computer Science, Information Technology and Communication Engineering. This course will form the concept of data science and machine learning hence this course is introduced at this level to make the students understand concept of Probability Distribution, ANOVA, neural networks, Supervised learning, unsupervised learning and use the type depending upon the application.

B. Course Outcomes: At the end of the course, students will be able to:

[CS 3203.1] Define Data Science techniques to analyse data and create statistical models that can lead to statistical results.

[CS 3203.2] Illustrate recent tools and algorithms used in Data Science & Machine Learning for solving real-world problems.

[CS 3203.3] Identify the use of supervised and unsupervised learning in machine learning domain.

[CS 3203.4] Illustrate and Compare various feature extraction models to analyse the performance.

[CS 3203.5] Evaluate machine learning algorithms on various hyper-parameters.

C. PROGRAM OUTCOMES AND PROGRAM SPECIFIC OUTCOMES

[PO.1] Engineering knowledge: Apply the knowledge of mathematics, science, engineering fundamentals, and an engineering specialization to the solution of complex engineering problems.

[PO.2] Problem analysis: Identify, formulate, research literature, and analyze complex engineering problems reaching substantiated conclusions using first principles of mathematics, natural sciences, and engineering sciences.

[PO.3] 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.

[PO.4] 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.

[PO.5] 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.

[PO.6] 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.

[PO.7] 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.

[PO.8] Ethics: Apply ethical principles and commit to professional ethics and responsibilities and norms of the engineering practices.

[PO.9] Individual and team work: Function effectively as an individual, and as a member or leader in diverse teams, and in multidisciplinary settings.

[PO.10] 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.

[PO.11] 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.

[PO.12] 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.

D. PROGRAM SPECIFIC OUTCOMES

At the end of the B Tech CSE program, the student:

[PSO.1] Will be able to identify the existing open problems in the field of computing Science and propose the best possible solutions.

[PSO.2] Will be able to apply knowledge of AI, Machine Learning and Data Mining in analyzing big data for extracting useful information from it and for performing predictive analysis.

[PSO.3] Will be able to design, manage and secure wired/ wireless computer networks for transfer and sharing of information.

E. Assessment Plan:

Criteria	Description	Maximum Marks
Internal Assessment (Summative)	Sessional Exam I (Closed Book)	30
	In class Quizzes and Project	30
End Term Exam (Summative)	End Term Exam (Closed Book)	40
	Total	100
Attendance (Formative)	A minimum of 75% Attendance is required to be maintained by a student to be qualified for taking up the End Semester examination. The allowance of 25% includes all types of leaves including medical leaves.	
Make up Assignments (Formative)	Students who misses a class will have to report to the teacher about the absence. A makeup assignment on the topic taught on the day of absence will be given which has to be submitted within a week from the date of absence. No extensions will be given on this. The attendance for that particular day of absence will be marked blank, so that the student is not accounted for absence. These assignments are limited to a maximum of 5 throughout the entire semester.	
Homework/ Home Assignment/ Activity Assignment (Formative)	There are situations where a student may have to work at home, especially before a flipped classroom. Although these works are not graded with marks. However, a student is expected to participate and perform these assignments with full zeal since the activity/ flipped classroom participation by a student will be assessed and marks will be awarded.	

F. SYLLABUS

Data Science: Descriptive Statistics, Probability Distribution, regression analysis, ANOVA. **Machine Learning**: Goals, Applications of ML, developing a learning system, training data, concept representation,

function approximation. Decision Tree Learning: Representing concepts as decision trees, Recursive induction of decision trees, best splitting attribute, entropy, information gain., Occam's razor, Overfitting, noisy data, and pruning. **Artificial Neural Networks**: Neurons and biological motivation. Linear threshold units, Perceptron, representational limitation and gradient descent training, Multilayer networks and backpropagation. Hidden layers and constructing intermediate, distributed representations, Overfitting, learning network structure, recurrent networks. Comparing learning algorithms: cross-validation, learning curves, and statistical hypothesis testing. Support Vector Machines: Maximum margin linear separators. Kernels for learning non-linear functions. Bayesian Learning: Probability theory and Bayes rule. Naive Bayes learning algorithm, Logistic regression, Bayes nets and Markov nets for representing dependencies. **Instance-Based Learning:** k-Nearest-neighbor algorithm, Case-based learning, Relevance feedback and Rocchio algorithm. Naive Bayes for text. Clustering and **Unsupervised Learning**: Hierarchical Agglomerative Clustering, k-means partitioned clustering, expectation maximization (EM) for soft clustering. **Ensemble Learning**: Bagging, boosting, and Decorate. Active learning with ensembles.

G. REFERENCE BOOKS

1. G. James, D. Witten, T Hastie, R Tibshirani, An introduction to statistical learning with applications in R, Springer, 2013.

2. J. Han, M. Kamber, J. Pei, Data Mining concepts and techniques, (2e), Morgan Kaufmann- Elsevier, 2011.

3. T. Hastie, R. Tibshirani, J. Friedman, The Elements of Statistical Learning, (2e), Springer, 2009.

4. K. Murphy, Machine Learning: A Probabilistic Perspective, MIT Press, 2012.

5. T. M. Mitchell, Machine Learning, (Indian Edition), MacGraw Hill, 2017.

H. Lecture Plan:

Class Number	Topics	Mode of Delivery	Corresponding Course Outcome	Mode of Assessing the Outcome
1	Introduction and Course Hand-out briefing	Lecture	NA	NA
2	Overview of Data Science and its applications	Lecture	CO 1	Mid Term, Quiz & End Term
3-4	Descriptive Statistics	Lecture	CO 1,2	Mid Term, Quiz & End Term
5-6	Probability Distribution	Lecture	CO 1	Mid Term, Quiz & End Term
7-8	Linear and Multiple regression analysis,	Lecture	CO 1	Mid Term, Quiz & End Term
9-10	Analysis of variance (ANOVA)	Lecture	CO 1,2	Mid Term, Quiz & End Term
11	Overview of Machine Learning, its goal and applications	Lecture	CO 2,3	Mid Term, Quiz & End Term
12	Logistic Regression	Lecture	CO 2,3	Mid Term, Quiz & End
13-14	Naive Bayes learning algorithm	Lecture	CO 2,3	Mid Term, Quiz & End
15-17	Decision Tree Learning (CART , pruning, bagging, boosting)	Lecture	CO 2,3	Mid Term, Quiz & End Term

18-19	Support Vector Machine with examples	Lecture	CO 2,3	Mid Term, Quiz & End Term
20-22	Artificial Neural Network , Perceptron Network with examples , Backpropagation network (Revision) & MacCulloch Pitts Model	Lecture	CO 3	Mid Term, Quiz & End Term
23-24	Statistical Text Representation: Bag of Words, TF-IDF, N-gram	Lecture	CO 4	Mid Term, Quiz & End Term
25-26	Distributed representation(Word2vec [CBOW. Skip gram]) (Application of NN)	Lecture	CO 3,4	Mid Term, Quiz & End Term
27-28	CNN, LSTM (Application of NN)	Lecture	CO 4	Mid Term, Quiz & End Term
29-30	Instance-Based Learning (KNN)	Lecture	CO 3	Mid Term, Quiz & End Term
31	Overview of Clustering and Unsupervised Learning	Lecture	CO 3,5	End Term Exam & Quiz
32	k-means partitioned clustering	Lecture	CO 3,4,5	End Term Exam & Quiz
33-34	Hierarchical Agglomerative Clustering	Lecture	CO 3,4,5	End Term Exam & Quiz
35-36	Kohonen Self Organizing maps	Lecture	CO 3,4,5	End Term Exam & Quiz
37	Ensemble Learning	Lecture	CO 4,5	End Term Exam & Quiz

I. Course Articulation Matrix: (Mapping of COs with POs & PSOs)

CO	STATEMENT	Correlation with Program Outcomes(POs)												Correlation with Program Specific Outcomes (**PSOs**)		
		PO 1	PO 2	PO 3	PO 4	PO 5	PO 6	PO 7	PO 8	PO 9	PO 10	PO 11	PO 12	PSO 1	PSO 2	PSO 3
[CS 3203.1]	Define Data Science techniques to analyse data and create statistical models that can lead to statistical results.	**2**	**2**		**1**									**2**		
[CS 3203.2]	Illustrate recent tools and algorithms used in Data Science & Machine Learning for solving real-world problems.	**2**	**3**			**2**						**1**		**2**	**1**	
[CS 3203.3]	Identify the use of supervised and unsupervised learning in machine learning domain.	**2**	**1**	**3**		**1**	**1**								**2**	
[CS 3203.4]	Compare various machine learning models to analyse the performance.		**3**	**2**		**1**						**1**				**2**
[CS 3203.5]	Evaluate machine learning algorithms on various hyper-parameters.		**2**	**2**	**2**		**1**						**2**		**2**	

1- Low Correlation; 2- Moderate Correlation; 3- Substantial Correlation