Course Code: DSC 4804	*Course Name:* *Statistical Learning*	
Credits: 3-0-0	***Contact hours per week*** = 3 hours	
Instructor-in-charge: Dr. Arijit Maitra Email: arijit.maitra@bmu.edu.in	***Pre-requisite:*** Basic knowledge of programming	

Aim of the course: The aim of the course is to utilize methods of statistical learning such as regression, classification, discriminant analysis, tree-based methods, support vector machines and unsupervised learning for data mining, modeling, and analyses.

Course Overview and Context: *Statistical Learning* is an open elective course for the B.Tech undergraduate program. This course deals with the principles and methods of statistical learning, and computational application of statistical methods to analyze and mine various types of data. At first, general principles of statistical learning methods will be discussed, and then applications on problems of simple and moderate complexities will be practiced. The course will follow a project-centric approach to reveal the underlying principles of data driven model development.

Topics of the Course

- Principles of statistical learning: Trade-off between accuracy and interpretability of a model, supervised vs. unsupervised learning, regression vs. classification problems
- Linear Regression: Simple Linear Regression, Multiple Linear Regression
- Classification: Logistic Regression, Multivariate Logistic Regression, Discriminant Analysis, Gaussian Discriminant Analysis, Naïve Bayes method.
- Resampling and validation methods
- Linear Model Selection and Regularization
- Tree-Based Methods
- Support Vector Machines
- Unsupervised Learning
- Applications of Neural Networks and Sequence Modelling

Schedule

Week 1: Principles of statistical learning: Trade-off between accuracy and interpretability of a model, supervised vs. unsupervised learning, regression vs. classification

Week 2-3: Linear Regression: Simple Linear Regression, Multiple Linear Regression

Week 4-5: Classification: Logistic Regression, Multivariate Logistic Regression, Discriminant Analysis, Gaussian Discriminant Analysis, Naïve Bayes method

Week 6: Resampling and Validation method

Week 7: Linear Model Selection and Regularization

Week 8-9: Tree-Based Methods

Week 10: Support Vector Machines

Week 11: Unsupervised Learning

Week 12-16: Neural Networks and Sequence Models

Course Outcomes: By the end of the course, the student will be able to:

CO1: Understand basic concepts of statistical learning methods, supervised and unsupervised learning.
CO2: Apply techniques of linear regression, logistic regression, discriminant analysis, tree based methods, SVM, neural networks
CO3: Analyse and compare different methods of statistical learning on diverse datasets.

CO-PO Mapping

PO and PSO →	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12	PSO1	PSO2	PSO3	PSO4
↓ CO Mapping Course Mapping →																
CO1			2	2				2	1	3		2				
CO2			2	2				2	1	3		2	3			
CO3			3	3				2	1	3		2	3	3		

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

Course Competencies: (Course Outcomes further elaborated) and Instruction schedule:

Module	**Competency**	**CO**	**No of sessions**
C1	Differentiate between accuracy and interpretability of a model, supervised vs. unsupervised learning, regression vs. classification problems, assess model accuracy	CO1	3
C2	Apply techniques of simple linear regression and multiple linear regression on data sets; analyze models using hypothesis testing and confidence intervals	CO1, CO2	6
C3	Apply classification algorithms on data sets using logistic regression and multivariate analyses; Gaussian discriminant analysis: single and many variables, Naïve Bayes and Quadratic Discriminant Analysis	CO3	6
C4	Delineate commonly used types of resampling methods for model validation, K-fold Cross validation, Bootstrap	CO1	3
C5	Apply techniques of Linear Model Selection and Regularization: Ridge regression, Lasso, Dimension Reduction Methods, Principal Components Regression	CO2, CO3	6
C6	Apply tree-based methods, Classification trees, Bagging and Random forests, Boosting to solve classification problems.	CO2, CO3	8
C7	Apply support vector machines on datasets for classification.	CO2, CO3	4
C8	Differentiate between supervised vs. unsupervised learning; apply techniques of unsupervised learning such as principal component analyses, k-means clustering, Hierarchical clustering.	CO2, CO3	3
C9	Introduction to neural networks and sequence models	CO2, CO3	9

Learning Resources:

1. Textbook: An Introduction to Statistical Learning with Applications in Python by Gareth James, Daniela Witten, Trevor Hastie, and Robert Tibshirani. Download: https://www.statlearning.com

2. Datasets: https://archive.ics.uci.edu/ml/datasets.html

3. Python tutorial:
https://www.datacamp.com/community/tutorials/matplotlib-tutorial-python/
http://cs231n.github.io/python-numpy-tutorial/
https://data36.com/pandas-tutorial-1-basics-reading-data-files-dataframes-data-selection/
https://data36.com/pandas-tutorial-2-aggregation-and-grouping/

4. Internet Links (annotated): Following are the online available resources:

5. Stanford University course (Statistical Learning): https://www.edx.org/course/statistical-learning

Note: Some portions of the syllabus may be delivered in the *flipped classroom* mode, in which students will learn specific portions of the syllabus from *MOOCs* followed by classroom discussion and/or project work. Instructor will regularly post information on learning resources such as lectures, tutorials, and assignments to the online course management portal i.e. LMS. Additional resources related to software, etc. would be provided on a timely basis.

Assessment Pattern: The final grade will be based on the marks or grades obtained in the mid-semester and end-semester evaluation along with other assessments defined in the assessment table. Relative grading method defined in the academic regulations of the university will be followed to grade the students. A student has to secure a minimum 40% of marks after completing all the assessments given in the following table to become eligible for grading.

Component	**Weightage (%)**	**Evaluation Schedule**	**Remarks**
Paper Selection & Abstract	5	26 Aug 2024	Abstract Submission
Report of "Literature Review"	15	22 Sept. 2024	Report Submission: to be formatted as a research article preferably using latex; *Turnitin* submit AI report and
Presentation of "Literature Review"	20	23/24 Sept. 2024	Oral Presentation, Equivalent to Mid Sem Exam
Project Report	20	24 Nov. 2024	To be formatted as a research article preferably using latex; submit *Turnitin* AI report and plagiarism checked
Presentation of Project	40	25/26 Nov. 2024	Oral Presentation, Equivalent to End Sem Exam

Experiential Learning: Students shall design and develop models or prototype software tools and/or applications to address data driven problems.

Attendance Policy: Students are expected to attend classes regularly. Failure to attend classes regularly and adhere to the expected attendance percentage will result in a reduction of the grade as per the University's grading policy. As specified in the Academic Regulations, if the overall attendance is less than 75% in a course, a student's grade will be reduced by one grade. For example, if the student's grade is A+ at the end of the semester it will be reduced to A and so on. For attendance less than 60%, the final grade for the student in the course will be R, where student has to repeat the course.

Assignments: All students in the class will be divided into teams. Team wise practice problems will be posted in the LMS as assignments. Each assignment will have a deadline for submission and will carry a weightage in final evaluation. Every student should submit his/her assignment on or before the deadline specified.
Late assignment submission policy: Late submission of assignments is not allowed, and any late submission will be awarded "0" marks.
Recourse examination policy: If a student fails or wishes to improve his/her grade in the course, recourse is permitted as per university academic regulations. Recourse is allowed only for end examination (2 hours) with 40% weightage.
Make-up policy: No make-up work will be given for unexcused absences. The faculty needs to be informed in advance in case the student is not going to be able to submit an assignment or take any evaluation component, and it is at the discretion of the faculty to sanction make-up test for an evaluation component.
Behavior expectations: Use of mobile phones and other distractive gadgets are not permitted in the class.
Academic dishonesty/cheating/plagiarism: Plagiarism and dishonesty in any form in any evaluation component will lead to disciplinary action according to the University's rules and regulations.