Department of Computer Science & Engineering (Data Science)

Lesson Plan &Work-done Diary for AY: 2025-26, ODD Semester

Theory Component

Course with Code: Statistical Machine Learning For Data Science - **BAD702**				**Faculty: Dr Anitha D B**			**Semester & Section:** 7A	
Class No.	**Date planned (DD/MM)**	**Topics to be covered**	**TLP Planned**	**Class No.**	**Date of Conduction (DD/MM)**	**Topics Covered**	**TLP Executed**	**Remarks if any deviation**
MODULE 1- Exploratory Data Analysis								
1.		Estimates of locations	Chalk & Talk PPT, Jupyter					
2.		Estimates of variability	Chalk & Talk PPT, Jupyter					
3.		Exploring data distributions	Chalk & Talk PPT, Jupyter					
4.		Exploring binary data	Chalk & Talk PPT, Jupyter					
5.		Exploring categorical data	Chalk & Talk PPT, Jupyter					
6.		Exploring two variables	Chalk & Talk PPT, Jupyter					
7.		Exploring more than two variables	Chalk & Talk PPT, Jupyter					
8.		Revision	Chalk & Talk PPT, Jupyter					

Course with Code: Statistical Machine Learning For Data Science - **BAD702**				**Faculty: Dr Anitha D B**			**Semester & Section:** 7A	
Class No.	**Date planned (DD/MM)**	**Topics to be covered**	**TLP Planned**	**Class No.**	**Date of Conduction (DD/MM)**	**Topics Covered**	**TLP Executed**	**Remarks if any deviation**
MODULE 2- Data and Sampling Distribution								
9		Random sampling and bias, selection bias,	Chalk & Talk PPT, Jupyter					
10		Sampling distribution of statistic	Chalk & Talk PPT, Jupyter					
11		Bootstrap	Chalk & Talk PPT, Jupyter					
12		Confidence intervals	Chalk & Talk PPT, Jupyter					
13		Data distributions: Normal Distribution, long tailed distribution	Chalk & Talk PPT, Jupyter					
14		Student's-t distribution Binomial distribution, Chi-square distribution	Chalk & Talk PPT, Jupyter					
15		F distribution	Chalk & Talk PPT, Jupyter					
16		Poisson and related distributions.	Chalk & Talk PPT, Jupyter					

Course with Code: Statistical Machine Learning For Data Science - **BAD702**				**Faculty: Dr Anitha D B**			**Semester & Section:** 7A	
Class No.	**Date planned (DD/MM)**	**Topics to be covered**	**TLP Planned**	**Class No.**	**Date of Conduction (DD/MM)**	**Topics Covered**	**TLP Executed**	**Remarks if any deviation**
MODULE 3 - Statistical Experiments and Significance Testing								
17		A/B testing	Chalk & Talk PPT, Jupyter					
18		Hypothesis testing	Chalk & Talk PPT, Jupyter					
19		Resampling	Chalk & Talk PPT, Jupyter					
20		Statistical significance	Chalk & Talk PPT, Jupyter					
21		p-values	Chalk & Talk PPT, Jupyter					
22		t-tests	Chalk & Talk PPT, Jupyter					
23		Multiple testing	Chalk & Talk PPT, Jupyter					
24		Degrees of freedom	Chalk & Talk PPT, Jupyter					

Course with Code: Statistical Machine Learning For Data Science - **BAD702**				**Faculty: Dr Anitha D B**			**Semester & Section:** 7A	
Class No.	**Date planned (DD/MM)**	**Topics to be covered**	**TLP Planned**	**Class No.**	**Date of Conduction (DD/MM)**	**Topics Covered**	**TLP Executed**	**Remarks if any deviation**
MODULE 4 -								
25		Multi-arm bandit algorithm,	Chalk & Talk PPT, Jupyter					
26		Power and sample size,	Chalk & Talk PPT, Jupyter					
27		Factor variables in regression,	Chalk & Talk PPT, Jupyter					
28		Interpreting the regression equation,	Chalk & Talk PPT, Jupyter					
29		Regression diagnostics	Chalk & Talk PPT, Jupyter					
30		Polynomial Regression	Chalk & Talk PPT, Jupyter					
31		Spline Regression	Chalk & Talk PPT, Jupyter					
32		Revision	Chalk & Talk PPT, Jupyter					

Course with Code: Statistical Machine Learning For Data Science - **BAD702**				**Faculty: Dr Anitha D B**			**Semester & Section: 7**A	
Class No.	**Date planned (DD/MM)**	**Topics to be covered**	**TLP Planned**	**Class No.**	**Date of Conduction (DD/MM)**	**Topics Covered**	**TLP Executed**	**Remarks if any deviation**
MODULE 5 - Discriminant Analysis								
33		Covariance Matrix	Chalk & Talk PPT, Jupyter					
34		Fisher's Linear discriminant	Chalk & Talk PPT, Jupyter					
35		Generalized Linear Models,	Chalk & Talk PPT, Jupyter					
36		Interpreting the coefficients	Chalk & Talk PPT, Jupyter					
37		Interpreting odd ratios	Chalk & Talk PPT, Jupyter					
38		Strategies for Imbalanced Data.	Chalk & Talk PPT, Jupyter					
39		Revision	Chalk & Talk PPT, Jupyter					
40		Revision	Chalk & Talk PPT, Jupyter					

Course with Code: Statistical Machine Learning For Data Science - **BAD702**				**Faculty: Dr Anitha D B**			**Semester & Section:** 7A	
Class No.	**Date planned (DD/MM)**	**Topics to be covered**	**TLP Planned**	**Class No.**	**Date of Conduction (DD/MM)**	**Topics Covered**	**TLP Executed**	**Remarks if any deviation**
Practical Component								
1		**Program 1:** A dataset contains the prices of houses in a city. Find the 25th and 75th percentiles and calculate the interquartile range (IQR). How does the IQR help in understanding the price variability?	Chalk & Talk PPT, Jupyter					
2		**Program 2:** You are given a dataset with categorical variables about customer satisfaction levels (Low, Medium, High) and whether customers made repeat purchases (Yes/No). Create visualizations such as bar plots or stacked bar charts to explore the relationship between satisfaction level and repeat purchases. What can you infer from the data?	Chalk & Talk PPT, Jupyter					
3		**Program 3:** A dataset contains information about car models, including the engine size (in Liters), fuel efficiency (miles per gallon), and car price. Use a pair plot or correlation matrix to explore the relationships between these variables. Which variables seem to have the strongest relationships, and what might be the practical significance of these findings?	Chalk & Talk PPT, Jupyter					
4		**Program 4:** You want to estimate the mean salary of software engineers in a country. You take 10 different random samples, each containing 50 engineers, and calculate the sample mean for each. Plot the distribution of these sample means. How does the Central Limit Theorem explain the shape of this sampling distribution, even if the underlying salary distribution is skewed?	Chalk & Talk PPT, Jupyter					

5		**Program 5:** A researcher conducts an experiment with a sample of 20 participants to determine if a new drug affects heart rate. The sample has a mean heart rate increase of 8 beats per minute and a standard deviation of 2 beats per minute. Perform a hypothesis test using the t-distribution to determine if the mean heart rate increase is significantly different from zero at the 5% significance level.	Chalk & Talk PPT, Jupyter					
6		**Program 6:** A company is testing two versions of a webpage (A and B) to determine which version leads to more sales. Version A was shown to 1,000 users and resulted in 120 sales. Version B was shown to 1,200 users and resulted in 150 sales. Perform an A/B test to determine if there is a statistically significant difference in the conversion rates between the two versions. Use a 5% significance level	Chalk & Talk PPT, Jupyter					
7		**Program 7:** You are comparing the average daily sales between two stores. Store A has a mean daily sales value of $1,000 with a standard deviation of $100 over 30 days, and Store B has a mean daily sales value of $950 with a standard deviation of $120 over 30 days. Conduct a two-sample t-test to determine if there is a significant difference between the average sales of the two stores at the 5% significance level.	Chalk & Talk PPT, Jupyter					
8		**Program 8:** A company collects data on employees' salaries and records their education level as a categorical variable with three levels: "High School", "Bachelor's", and "Master's". Fit a multiple linear regression model to predict salary using education level (as a factor variable) and years of experience. Interpret the coefficients for the education levels in the regression model	Chalk & Talk PPT, Jupyter					

9		**Program 9:** You have data on housing prices and square footage and notice that the relationship between square footage and price is nonlinear. Fit a spline regression model to allow the relationship between square footage and price to change at 2,000 square feet. Explain how spline regression can capture different behaviours of the relationship before and after 2,000 square feet.	Chalk & Talk PPT, Jupyter					
10		**Program 10:** A hospital is using a Poisson regression model (a type of GLM) to predict the number of emergency room visits per week based on patient age and medical history. The model is given by: $\text{Log}(\lambda) = 2.5 - 0.03*\text{Age} + 0.5*\text{condition}$ where λ is the expected number of visits per week, Age is the patient's age, and condition is a binary variable (1 if the patient has a chronic condition, 0 otherwise). Interpret the coefficients of Age and condition. What is the expected number of visits per week for a 60-year-old patient with a chronic condition? How would the expected number of visits change if the patient did not have a chronic condition?	Chalk & Talk PPT, Jupyter					
11		**Program 11:** A bakery claims that its new cookie recipe is lower in calories compared to the old recipe, which had a mean calorie count of 200. You sample 40 new cookies and find a mean of 190 calories with a standard deviation of 15 calories. Perform a one-tailed t-test to determine if the new recipe has significantly fewer calories at a 5% significance level.	Chalk & Talk PPT, Jupyter					

	Activity	Planned	Actual	Remarks
1	Theory Classes + Practical Classes	40L+10P		
2	Assignments/ Quizzes/ Self-study/Programs	2		
3	Tutorials/ Extra classes	1P		
4	Internal Assessments	3		
5	ICT based Teaching (% of usage in Curriculum)	90%		
Planning			**Execution**	
Faculty Signature:			**Faculty Signature:**	
HoD Signature:			**HoD Signature:**	