STATISTICAL MODELLING FOR MACHINE LEARNING **Course Code : 313307**

Programme Name/s : **Artificial Intelligence/ Artificial Intelligence and Machine Learning/ Data Sciences**

Programme Code : **AI/ AN/ DS**

Semester : **Third**

Course Title : **STATISTICAL MODELLING FOR MACHINE LEARNING**

Course Code : **313307**

I. RATIONALE

Machine Learning refers to the automated identification of patterns in data. This course is included in curriculum to establish foundation for Artificial Intelligence and Machine Learning. Statistic, Probability, Interpolation and sampling methods are the core components of AI/ML. This course will enable students to implement mathematical concepts using R-Programming which will enhance the knowledge and skills to use the methodology for solving AI/ML based problems of various domains.

II. INDUSTRY / EMPLOYER EXPECTED OUTCOME

Apply Mathematics to solve real-world problems using AI/ML concept and principles to enhance decision-making, design and innovation with precision and efficiency.

III. COURSE LEVEL LEARNING OUTCOMES (COS)

Students will be able to achieve & demonstrate the following COs on completion of course based learning

- CO1 - Solve the given problem based on Statistic Techniques using R-Programming.
- CO2 - Implement Statistic methods using R-Programming.
- CO3 - Use Principles of Probability to solve given Problem.
- CO4 - Implement appropriate method based on the Interpolation.
- CO5 - Apply Sampling Methods to solve given problem using R-Programming.

IV. TEACHING-LEARNING & ASSESSMENT SCHEME

Course Code	Course Title	Abbr	Course Category/s	Learning Scheme					Credits	Assessment Scheme											
				Actual Contact Hrs./Week			SLH	NLH		Paper Duration	Theory				Based on LL & TL				Based on SL		Total Marks
															Practical						
				CL	TL	LL					FA-TH	SA-TH	Total		FA-PR		SA-PR		SLA		
											Max	Max	Max	Min	Max	Min	Max	Min	Max	Min	
313307	STATISTICAL MODELLING FOR MACHINE LEARNING	SML	DSC	3	1	4	2	10	5	3	30	70	100	40	25	10	50@	20	25	10	200

Total IKS Hrs for Sem. : Hrs
Abbreviations: CL- ClassRoom Learning , TL- Tutorial Learning, LL-Laboratory Learning, SLH-Self Learning Hours, NLH-Notional Learning Hours, FA - Formative Assessment, SA -Summative assessment, IKS - Indian Knowledge System, SLA - Self Learning Assessment
Legends: @ Internal Assessment, # External Assessment, *# On Line Examination , @$ Internal Online Examination
Note :

1. FA-TH represents average of two class tests of 30 marks each conducted during the semester.
2. If candidate is not securing minimum passing marks in FA-PR of any course then the candidate shall be declared as "Detained" in that semester.
3. If candidate is not securing minimum passing marks in SLA of any course then the candidate shall be declared as fail and will have to repeat and resubmit SLA work.
4. Notional Learning hours for the semester are (CL+LL+TL+SL)hrs.* 15 Weeks
5. 1 credit is equivalent to 30 Notional hrs.
6. * Self learning hours shall not be reflected in the Time Table.
7. * Self learning includes micro project / assignment / other activities.

V. THEORY LEARNING OUTCOMES AND ALIGNED COURSE CONTENT

Sr.No	Theory Learning Outcomes (TLO's)aligned to CO's.	Learning content mapped with Theory Learning Outcomes (TLO's) and CO's.	Suggested Learning Pedagogies.
1	TLO 1.1 Solve problems based on Frequency Distribution. TLO 1.2 Calculate Mean, Median and Mode for all types of data. TLO 1.3 Find Mode and Median using Graphical method. TLO 1.4 Find Karl Pearson's and Bowley's Co-efficient of Skewness for the given data. TLO 1.5 Calculate the Measures of Kurtosis based on moment for given data.	**Unit - I Statistical Techniques** 1.1 Frequency Distribution: Definition, Basic terms. 1.2 Classification of Data: Raw, Ungroup and Group data. 1.3 Measures of Central Tendency: Mean, Median and Mode for all types of data. 1.4 Concept of Quartiles, Deciles and Percentiles for all types of data. 1.5 Geometric mean and Harmonic mean and Combined mean for given data. 1.6 Graphical Representation to find Mode (Histogram) and Median (Ogive curve). 1.7 Measures of Dispersion: Range, Mean Deviation, Standard Deviation, Variance. 1.8 Skewness: Types of skewness, Test of skewness, Co-efficient of skewness-Karl Pearson's and Bowley's coefficient. 1.9 Types of skewness in terms of Mean and Mode. 1.10 Measures of Kurtosis using central moment.	Classroom Lecture Flipped Classroom Demonstration
2	TLO 2.1 Fit straight line and second-degree polynomial using Least Square method. TLO 2.2 Calculate Coefficient of correlation using Karl-Pearson's and Spearman's Rank methods. TLO 2.3 Obtain the equation to the Line of Regression for the given data.	**Unit - II Statistical Methods** 2.1 Fitting of Straight Line y = a+bx using method of least square. 2.2 Fitting of second degree polynomial $y=a+bx+cx^2$ using method of least square. 2.3 Covariance of the data. 2.4 Correlations, Types of Correlations. 2.5 Scatter Or Dot-Diagram. 2.6 Karl-Pearson's and Spearman's Rank Coefficient of Correlation. 2.7 Regression Equation of line in two variables.	Classroom Lecture Flipped Classroom Demonstration

Sr.No	Theory Learning Outcomes (TLO's)aligned to CO's.	Learning content mapped with Theory Learning Outcomes (TLO's) and CO's.	Suggested Learning Pedagogies.
3	TLO 3.1 Solve problems using addition and multiplication probability theorem. TLO 3.2 Solve problems using Conditional Probability. TLO 3.3 Solve problems using Bayes' theorem.	**Unit - III Probability of Random Variable** 3.1 Probability : Definition, Terminologies. 3.2 Theorem of Probability: Addition, Multiplication. 3.3 Conditional probability. 3.4 Bayes' theorem.	Classroom Lecture Flipped Classroom Demonstration
4	TLO 4.1 Solve problems using Lagrange's interpolation formula. TLO 4.2 Construct forward and backward difference table. TLO 4.3 Solve problems using Forward, Backward, Shift, Inverse shift operators. TLO 4.4 Solve problems on Forward and Backward Interpolation. TLO 4.5 Solve problems on Extrapolation.	**Unit - IV Interpolation** 4.1 Introduction. 4.2 Lagrange's Interpolation formula. 4.3 Finite Differences: Forward difference, Backward difference, Shift operator, Inverse shift operator. 4.4 Relation between forward, backward, shift and inverse shift operator. 4.5 Newton's Gregory forward and backward difference Interpolation Formula. 4.6 Concept of Extrapolation.	Classroom Lecture Flipped Classroom Presentations
5	TLO 5.1 Solve the given problem using Sampling distribution proportion. TLO 5.2 Test samples using t-Distribution. TLO 5.3 Test samples using chi-Square Distribution. TLO 5.4 Use Chi-Square test to test the Independence.	**Unit - V Sampling Methods** 5.1 Population, Sampling, Aim of Sampling, Parameters and Statistic, Types of Sampling, Standard Error. 5.2 Hypothesis: Null Hypothesis and Alternating Hypothesis. 5.3 Level of Significance, Test of Significance and Confidence Limits. 5.4 Test of Significance of Large Samples (N>30). 5.5 Sample Distribution of the proportion. 5.6 Comparison of Large Samples. 5.7 The t-Distribution (For Small Samples). 5.8 Chi-Square Distribution, Condition for Chi Square test. 5.9 Degree of freedom. 5.10 Chi-Square Test of Goodness of fit and Chi-Square test as a test of Independence.	Classroom Lecture Presentations Demonstration

VI. LABORATORY LEARNING OUTCOME AND ALIGNED PRACTICAL / TUTORIAL EXPERIENCES.

Practical / Tutorial / Laboratory Learning Outcome (LLO)	Sr No	Laboratory Experiment / Practical Titles / Tutorial Titles	Number of hrs.	Relevant COs
LLO 1.1 Download and Install R studio and R Packages.	1	* Download R Studio.Install and configure R studio and R Packages.	2	CO1
LLO 2.1 Implement basic mathematical operations in R programming.	2	Write Program to- a. Demonstrate the use of R-Numbers. (numeric, integer, complex). b. Convert number from one type to other using functions. c. Perform following operations. i. Addition and Subtraction on numbers. ii. Find Square root using of number using built-in function.	2	CO1

STATISTICAL MODELLING FOR MACHINE LEARNING **Course Code : 313307**

Practical / Tutorial / Laboratory Learning Outcome (LLO)	Sr No	Laboratory Experiment / Practical Titles / Tutorial Titles	Number of hrs.	Relevant COs
LLO 3.1 Perform given operation on Datasets using R programming.	3	Write Program to- a. Print any built-in data set of R. b. Get information about the data set. c. Find the dimensions of the data set and view the names of the variables. Hint: Use dim( ) and names( ) function. d. Find the name of each row in the first column. Hint: Use the rownames( ) function. e. Print all values that belong to a variable. f. Sort the values of variable. g. Get the statistical summary of the data.	2	CO1
LLO 4.1 Find lowest and highest values and index position of max and min value from dataset using R Program.	4	Write a program to- a. Find the lowest or highest value in a data set. Hint: Use min( ) and max( ) functions. b. Find the index position of the max and min value in the table. Hint: use which.max( ) and which.min( ) functions.	2	CO1
LLO 5.1 Calculate Measures of Central tendency including Quartiles using R Program.. LLO 5.2 Calculate Deciles and Percentiles using R Program.	5	* Write programs to calculate Measures of Central tendency. a. Import data into R. b. Calculate the Mean (Average value) of a variable from the given data set. c. Find the Median (Mid-Point value) of the variable from the given data set. d. Calculate the mode for the variable from the given data set.(by sorting the column of the dataframe and by using the 'modest' package). e. Calculate the Percentile of the variable from the given data set.	4	CO1
LLO 6.1 Create Frequency Distribution table using R Programming.	6	* Write programs to- a. Print Original Data Frame, Modified Frequency Table, Cumulative Frequency Table, Relative Frequency Table. b. Create the Frequency Table by using multiple arguments. c. Plot the frequency table using ggplot function.	2	CO1
LLO 7.1 Calculate measures of Dispersion using R Program.	7	* Write programs to calculate-Variance, Standard Deviation, Range, Mean Deviation for the given data.	2	CO1
LLO 8.1 Draw Histogram and Ogive Curve to graphically represent Mode and Median for the given data.	8	* Write Programs to graphically represent mode and median of the given data. a. Draw Histogram for the given data. b. Draw Ogive Curve for the given data.	2	CO1
LLO 9.1 Calculate Skewness for the given data using R Program.	9	* Write a Program to calculate Skewness for the given data.	2	CO1
LLO 10.1 Draw a scatterplot for the given data using R Program.	10	Write a Program to draw a scatterplot for two variables for the given dataset.	2	CO1

Practical / Tutorial / Laboratory Learning Outcome (LLO)	Sr No	Laboratory Experiment / Practical Titles / Tutorial Titles	Number of hrs.	Relevant COs
LLO 11.1 Perform the correlation test to evaluate the association between two or more variables using R Program.	11	Write Program to perform the correlation test to evaluate the association between two or more variables. a. Install and load required R packages. b. Compute correlation in R. c. Visualize your data using scatter plots. d. Perform Preliminary test to check the test assumptions.	2	CO2
LLO 12.1 Perform the correlation test to evaluate the association between two or more variables using R Program.	12	* Write Program to perform the correlation test to evaluate the association between two or more variables. a. Pearson correlation test. b. Interpretation of the result. c. Use Spearman rank correlation coefficient to estimate a rank-based measure of association.	4	CO2
LLO 13.1 Find the Line of Regression for two variables using R Program.	13	* Write a Program based on Line of Regression using two variables.	2	CO2
LLO 14.1 Implement Classical Probability using R Program.	14	* Write Programs to- a. Calculate the probability of getting heads when flipping a fair coin. b. Calculate the probability of drawing a spade from a standard deck of 52 cards.	4	CO3
LLO 15.1 Implement Classical Probability using R Program.	15	Write a program to calculate the probability of rolling a 3 on a fair six-sided die.	2	CO3
LLO 16.1 Compute probability using R Program.	16	Write a program to compute probability using prob( ) function.	2	CO3
LLO 17.1 Calculate the conditional probability using R Program.	17	Write a program to calculate the conditional probability.	2	CO3
LLO 18.1 Calculate the probability of an event based on given information using R Program.	18	* Write a program to use Bayes' Theorem in R-Programming.	4	CO3
LLO 19.1 Find forward Interpolation using R Program.	19	* Write a Program to interpolate using newton forward interpolation.	2	CO4
LLO 20.1 Find backward Interpolation using R Program.	20	* Write a Program to interpolate using newton backward interpolation.	2	CO4
LLO 21.1 Find Extrapolation using R Program.	21	* Write a program for the implementation of extrapolation.	2	CO4
LLO 22.1 Generate Samples for the given dataset using R Program.	22	Write a program to generate Samples using the Sampling Functions.	2	CO5
LLO 23.1 Perform the given type of sampling using R program.	23	Write programs to perform following types of sampling- a. Simple Random Sampling. b. Stratified Sampling. c. Systematic Sampling. d. Biased Sampling.	2	CO5
LLO 24.1 Generate Sampling Distribution proportion using R program.	24	* Write a program to generate a Sampling Distribution proportion.	2	CO5
LLO 25.1 Perform Hypothesis Testing by t-Distribution using R program.	25	* Write a program based on t-Distribution using dt, pt, qt & rt functions.	2	CO5

Practical / Tutorial / Laboratory Learning Outcome (LLO)	Sr No	Laboratory Experiment / Practical Titles / Tutorial Titles	Number of hrs.	Relevant COs
LLO 26.1 Perform Hypothesis Testing by Chi-Square Distribution using R program.	26	* Write a program based on Chi-Square Distribution using dchisq, pchisq, qchisq & rchisq functions.	2	CO5

Note : Out of above suggestive LLOs -

- '*' Marked Practicals (LLOs) Are mandatory.
- Minimum 80% of above list of lab experiment are to be performed.
- Judicial mix of LLOs are to be performed to achieve desired outcomes.

VII. SUGGESTED MICRO PROJECT / ASSIGNMENT/ ACTIVITIES FOR SPECIFIC LEARNING / SKILLS DEVELOPMENT (SELF LEARNING)

Assignment

- Collect data of at least 05 real world examples and test the Hypothesis of sampling distribution.
- Collect data of at least 05 real world examples and calculate Measures of skewness and kurtosis and prepare the document.
- Collect data of at least 05 real world examples and draw/fit straight line and second-degree polynomial.
- Collect data of at least 05 real world examples and calculate probability using Bayes' theorem.
- Collect data of at least 03 city like cost of living and temperature data etc. and interpolate the missing index number for it and prepare the document.

Micro project

- Analyze Uber Data: Analyze different parameters like the number of trips made in a day, the number of trips during a particular month, average passenger that uber can have in a day, the peak hours where more customers are available, maximum number of trips found on day of the month, etc.
- Implement each least squares regression technique using a programming language such as Python or R. Utilize libraries like scikit-learn or stats models for implementation, ensuring proper parameter tuning and regularization settings for each technique.
- Collect temperature data from different locations at various times of the day. Use interpolation techniques such as linear interpolation or spline interpolation to estimate the temperature at specific times and locations where data is not available.
- Design a simple hypothesis testing scenario where you simulate data under different conditions and perform chi-square tests to assess the significance of observed difference.
- Perform case Study on probabilistic model for predicting relations in social websites system.

Note :

- Above is just a suggestive list of microprojects and assignments; faculty must prepare their own bank of microprojects, assignments, and activities in a similar way.
- The faculty must allocate judicial mix of tasks, considering the weaknesses and / strengths of the student in acquiring the desired skills.
- If a microproject is assigned, it is expected to be completed as a group activity.
- SLA marks shall be awarded as per the continuous assessment record.
- For courses with no SLA component the list of suggestive microprojects / assignments/ activities are optional, faculty may encourage students to perform these tasks for enhanced learning experiences.
- If the course does not have associated SLA component, above suggestive listings is applicable to Tutorials and maybe considered for FA-PR evaluations.

VIII. LABORATORY EQUIPMENT / INSTRUMENTS / TOOLS / SOFTWARE REQUIRED

Sr.No	Equipment Name with Broad Specifications	Relevant LLO Number

Sr.No	Equipment Name with Broad Specifications	Relevant LLO Number
1	Open-source software like SageMaths, MATHS3D, GeoGebra, Graph, DPLOT, and Graphing Calculator (Graph Eq 2.13), ORANGE can be used for Graph theory and tree, Statistics respectively.	All
2	Minimum Hardware requirement: Personal computer, (i3-i5 preferable), RAM minimum 4 GB onwards.	All
3	Minimum system requirement: 64-bit operating system such as Windows 10, macOS 10.13, or a recent version of Linux.	All
4	Software: R-Studio	All

IX. SUGGESTED WEIGHTAGE TO LEARNING EFFORTS & ASSESSMENT PURPOSE (Specification Table)

Sr.No	Unit	Unit Title	Aligned COs	Learning Hours	R-Level	U-Level	A-Level	Total Marks
1	I	Statistical Techniques	CO1	10	2	6	12	20
2	II	Statistical Methods	CO2	10	2	4	8	14
3	III	Probability of Random Variable	CO3	7	2	2	4	8
4	IV	Interpolation	CO4	10	2	4	8	14
5	V	Sampling Methods	CO5	8	2	4	8	14
Grand Total				**45**	**10**	**20**	**40**	**70**

X. ASSESSMENT METHODOLOGIES/TOOLS

Formative assessment (Assessment for Learning)

- Laboratory Performance, Unit Tests , Midterm Exam, Self-learning, Term Work, Seminar/Presentations.
- Continuous assessment based on process and product related performance indicators.
- Each practical will be assessed considering 60% weightage to process and 40% weightage to product.

Summative Assessment (Assessment of Learning)

- End Semester Exam, Practical exam, viva voce.

XI. SUGGESTED COS - POS MATRIX FORM

Course Outcomes (COs)	Programme Outcomes (POs)							Programme Specific Outcomes* (PSOs)		
	PO-1 Basic and Discipline Specific Knowledge	PO-2 Problem Analysis	PO-3 Design/ Development of Solutions	PO-4 Engineering Tools	PO-5 Engineering Practices for Society, Sustainability and Environment	PO-6 Project Management	PO-7 Life Long Learning	PSO-1	PSO-2	PSO-3
CO1	3	1	1	-	-	-	1			
CO2	3	2	1	1	1	-	1			
CO3	3	2	2	1	-	1	2			
CO4	3	2	2	1	-	1	1			
CO5	3	2	2	1	1	1	2			

Legends :- High:03, Medium:02,Low:01, No Mapping: -
*PSOs are to be formulated at institute level

XII. SUGGESTED LEARNING MATERIALS / BOOKS

Sr.No	Author	Title	Publisher with ISBN Number
1	H.K.Dass, Er.Rajnish Verma	Higher Engineering Mathematics	S. Chand Technical ISBN: 9788121938907,9788121938907
2	A.C. Srivastava, P.K. Srivastava	Engineering Mathematics	PHI Publication, New Delhi ISBN: 9788120342934
3	Grewal B. S.	Higher Engineering Mathematics	Khanna Publication New Delhi, 2013 ISBN: 8174091955
4	B. V. Ramana	Higher Engineering Mathematics	Tata Mcgraw Hill Education private limited, New Delhi. ISBN: 9780070634190, 9780070634190
5	Marvin L. Bittinger David J. Ellenbogen Scott A. Surgent	Calculus and Its Applications	Addison-Wesley 10th Edition ISBN-13: 978-0-321-69433-1
6	Andrie de Vries and Joris Meys	R For Dummies	Wiley Publisher ISBN-13: 978-8126562183
7	Jeeva Jose	Beginner's Guide for Data Analysis using R Programming	Khanna Book Publishing, First Edition ISBN-13: 978-9386173454

XIII . LEARNING WEBSITES & PORTALS

Sr.No	Link / Portal	Description
1	http://nptel.ac.in/courses/106102064/1	Online Learning Initiatives by IITs and IISc
2	https://ocw.mit.edu/	MIT Open courseware
3	https://www.khanacademy.org/math	Concept of Mathematics through video lectures and notes
4	www.mathworks.com/	Applications of concepts of Mathematics to coding.
5	https://amser.org/b920509/science--mathematics	AMSER (Interpolation. Extrapolation
6	https://www.coursera.org/learn/r-programming	R Programming: Free online Course
7	https://libguides.furman.edu/oer/subject/mathematics	Open Education Resources (OER) in Mathematics. (Interpolation. Extrapolation
8	https://www.wolframalpha.com/	Solving Mathematical Problems, performing calculations, visualizing mathematical concepts.
9	https://brilliant.org/	Interactive Learning in Mathematics
10	https://www.w3resource.com/r-programming-exercises/basic/	R Programming Basic, Exercises, Practice, Solution
11	www.datamentor.io/r-programming/examples/	R Programming Examples
12	https://www.tutorialspoint.com/r_programming_language/index.asp	R-Programming Online Course
13	https://www.freecodecamp.org/news/all-the-math-you-need-in-artificial-intelligence/	Mathematics in AI
14	https://byjus.com/maths/least-square-method/	Least Square Method
15	https://www.w3resource.com/r-programming-exercises/basic/r-programming-basic-exercise-3.php	R-Programming: Basic Exercises with Solution

STATISTICAL MODELLING FOR MACHINE LEARNING **Course Code : 313307**

Sr.No	Link / Portal	Description

Note :

- Teachers are requested to check the creative common license status/financial implications of the suggested online educational resources before use by the students