Program Name : **Diploma in Artificial Intelligence and Machine Learning**

Program Code : **AN**

Semester : **Third**

Course Title : **Mathematics for Machine Learning-I**

Course Code : **22397**

1. RATIONALE

Machine Learning refers to the automated identification of patterns in data. This course is included in curriculum as Mathematics is foundation for AI and ML. Statistic and Probability are the core components of AI/ML. This course deals with concept of R-Programming to enhance the knowledge of statistics and Probability. This course will enable the students to inculcate programming concepts and methodology to solve AI/ML based Eengineering problems.

2. COMPETENCY

The aim of this course is to help the student to attain the following industry identified competency through various teaching learning experiences:

- **Write Programs using R-Programming to solve Statistics and Probability based problems.**

3. COURSE OUTCOMES (COs)

The theory, practical experiences and relevant soft skills associated with this course are to be taught and implemented, so that the student demonstrates the following industry oriented COs associated with the above mentioned competency:

a. Solve the given problem based on Statistic Techniques using R-Programming.
b. Implement Statistic methods using R-Programming.
c. Use Probability Theorems to solve Real-Time Problems.
d. Apply Probability Distribution in R-Programming.
e. Apply Sampling Methods in R-Programming.

4. TEACHING AND EXAMINATION SCHEME

Teaching Scheme			Credit (L+T+P)	Examination Scheme											
				Theory							Practical				
L	T	P		Paper Hrs.	ESE		PA		Total		ESE		PA		Total	
					Max	Min	Max	Min	Max	Min	Max	Min	Max	Min	Max	Min
3	2	2	7	3	70	28	30*	00	100	40	25@	10	25	10	50	20

(**) *marks should be awarded on the basis of internal end semester theory exam of 50 marks based on the specification table given in S. No. 9.*

*(~2): For the **practical only courses,** the PA has two components under practical marks i.e. the assessment of practicals (seen in section 6) has a weightage of 60% (i.e.30 marks) and micro-project assessment (seen in section 12) has a weightage of 40% (i.e.20 marks). This is designed to facilitate attainment of COs holistically, as there is no theory ESE.*

Legends: **L**-Lecture; **T** – Tutorial/Teacher Guided Theory Practice; **P** - Practical; **C** – Credit, **ESE** - End Semester Examination; **PA** - Progressive Assessment, **'#':** No Theory Examination

5. COURSE MAP (with sample COs, Learning Outcomes i.e. LOs and topics)

This course map illustrates an overview of the flow and linkages of the topics at various levels of outcomes (details in subsequent sections) to be attained by the student by the end of the course, in all domains of learning in terms of the industry/employer identified competency depicted at the centre of this map.

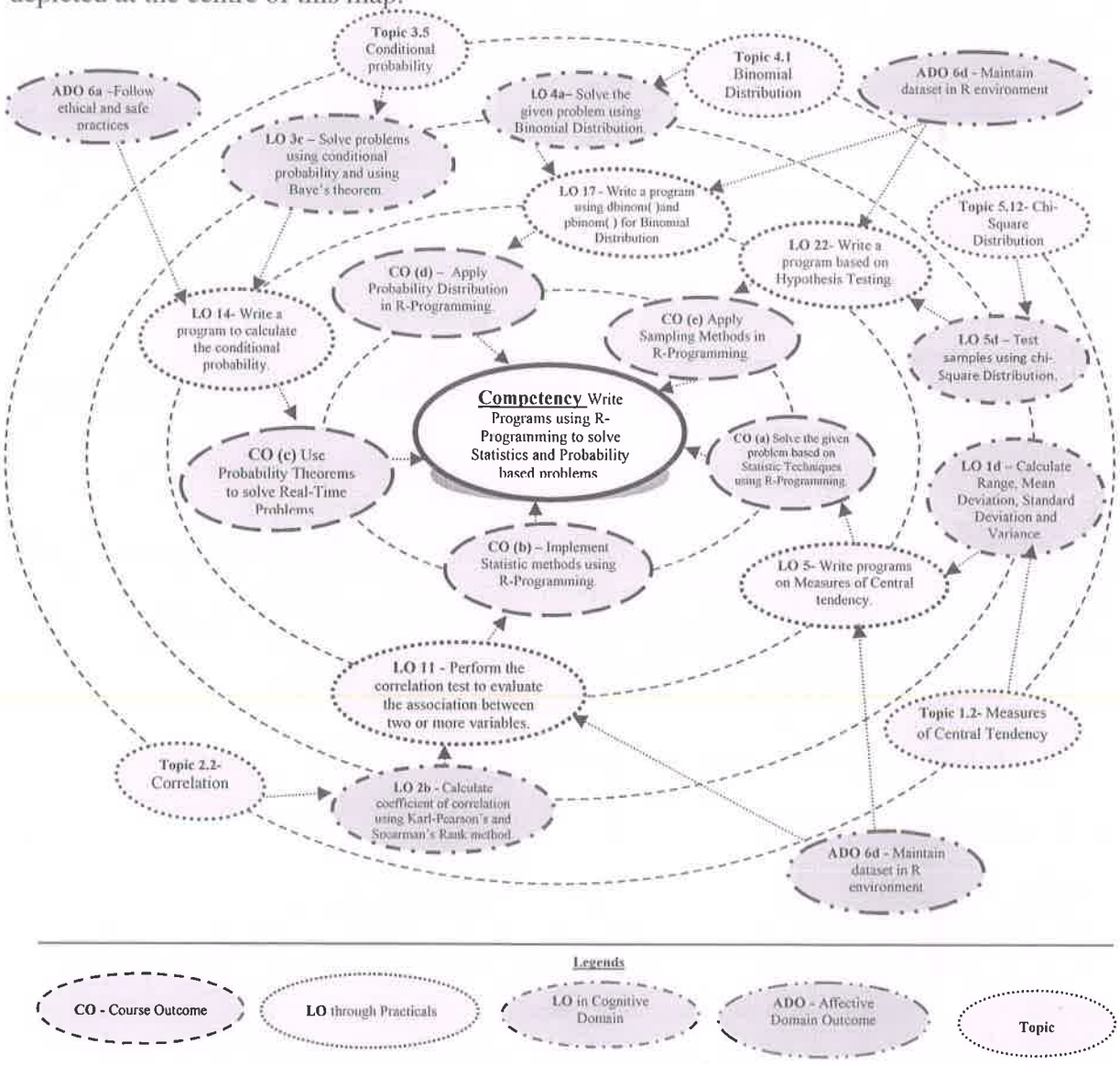

Figure 1 - Course Map

6. SUGGESTED PRACTICALS/ EXERCISES

The practicals/exercises/tutorials in this section are psychomotor domain LOs (i.e. sub-components of the COs) are to be developed and assessed in the student to lead to the attainment of the competency.

S. No.	Practical Exercises (Learning Outcomes in Psychomotor Domain)	Unit No.	Approx. Hrs. Required
1	Install R studio and R Packages.	I	02*
2	a. Write simple programs using R-Numbers. (numeric, integer, complex). b. Write simple programs to convert from one type to other using functions. c. Write simple programs to perform following operations. i. Addition and Subtraction ii. Find Square root using built-in function.	I	02*
3	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.	I	02*
4	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).	I	02*
5	Write programs on 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.		02*
6	Write a program to make a frequency distribution table in R. 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.	I	02*
7	Write programs to calculate Variance, Standard Deviation, Range, Mean Deviation for the given data.	I	02
8	a. Draw Histogram for the given data. b. Draw Ogive Curve for the given data.	I	02*
9	Write a Program to calculate Skewness for the given data.	I	02
10	Write a Program to draw a scatterplot for two variables for the given dataset.	I	02*
11	Perform the correlation test to evaluate the association between two or more variables. a. Install and load required R packages	II	02*

S. No.	Practical Exercises (Learning Outcomes in Psychomotor Domain)	Unit No.	Approx. Hrs. Required
	b. Compute correlation in R c. Import your data into R d. Visualize your data using scatter plots e. Preliminary test to check the test assumptions		
12	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.	II	02*
13	Write a Program based on Line of Regression.	II	02*
14	Write a program to calculate the conditional probability.	III	02
15	Write a program to compute probability using prod( ) function.	III	02
16	Write a program to use Bayes' Theorem in R-Programming.	III	02*
17	Write a program using dbinom( )and pbinom( ) for Binomial Distribution.	IV	02*
18	Write a program using dpois( ) and ppois( ) for Poisson Distribution.	IV	02
19	Write a program using dnorm ( ) and pnorm ( ) for Normal Distribution.	IV	02
20	Write a program to generate a Samples Using the Sampling Functions in R.	V	02*
21	Write a program to generate a Sampling Distribution in R.	V	02
22	Write a program based on Hypothesis Testing.	V	02*
23	Write a program based on t-Distribution using dt, pt, qt & rt functions.	V	02*
24	Write a program based on Chi-Square Distribution using dchisq, pchisq, qchisq & rchisq functions.	V	02
		Total	**48**

Note

i. *Given in above tables is suggestive list of practical exercises. Teachers can design other similar exercises.*

ii. *Pracicals Marked as * are Compulsory. Other practicals can be given to fast learners/Bright students.*

iii. *The faculty from Computer/Information Technology Department teaching listed practicals should co-ordinate with the Math's faculty teaching theory contents to formulate the problem Statement to perform above practicals,*

iv. *To attain the COs and competency, a judicial mix of practical from above listed Learning Outcomes (LOs) need to be undertaken to achieve the 'Applying Level' of Blooms's 'Cognitive Domain Taxonomy'.*

Assessment of the 'Process' and 'Product' related skills in the laboratory/workshop/field work should be done as per suggested sample below:

S.No.	Performance Indicators	Weightage in %
1	Import Packages and use libraries.	20
2	Use R-Studio/R-Environment to create, edit, compile the 'R' programs/applications.	40
3	Debug, test and execute the programs/applications.	20
4	Able to answer oral questions.	10
5	Submission of written assignment in time.	10
	Total	**100**

Additionally, the following affective domain LOs (social skills/attitudes), are also important constituents of the competency which can be best developed through the above mentioned laboratory/field based experiences:

a. Handle R-Programming environment.
b. Experiment with R-Programming environment.
c. Import, construct, compile, debug, test and execute R programs.
d. Maintain dataset in R environment.
e. Manage R packages.
f. Follow ethical and safe practices.

The development of the attitude related LOs of Krathwohl's 'Affective Domain Taxonomy', the achievement level may reach:

- 'Valuing Level' in 1st year
- 'Organising Level' in 2nd year and
- 'Characterising Level' in 3rd year.

7. MAJOR EQUIPMENT/ INSTRUMENTS REQUIRED

The major equipment with broad specification mentioned here will usher in uniformity in conduct of experiments, as well as aid to procure equipment by authorities concerned.

S. No.	Equipment Name with Broad Specifications	Expt. S. No.
1	Hardware: Personal computer, (i3-i5 preferable), RAM minimum 2 GB onwards.	For all Experiments
2	Operating system: Windows XP/Windows 7/LINUX onwards.	
3	Software: R-Studio	

8. UNDERPINNING THEORY COMPONENTS

The following topics/subtopics should be taught and assessed in order to develop LOs in cognitive domain for achieving the COs to attain the identified competency.

Unit	Major Learning Outcomes (in cognitive domain)	Topics and Sub-topics
Unit – I Statistical Techniques	1a. Solve problems based on Frequency Distribution. 1b. Calculate Mean, Median and Mode for all types of data. 1c. Calculate Geometric, Harmonic and Combined	1.1. Frequency Distribution 1.1.1. Definition- Basic terms 1.1.2. Classification of Data-Raw, Ungroup and Group data 1.1.3. Frequency Distribution for all types of data

Unit	Major Learning Outcomes (in cognitive domain)	Topics and Sub-topics
	Mean for raw data. 1d. Calculate Range, Mean Deviation, Standard Deviation and Variance. 1e. Find Mode and Median using Graphical methods. 1f. Find Co-efficient of Skewness for given data.	1.2. Measures of Central Tendency 1.2.1. Mean, Median, mode for all types of Data 1.2.2. Concept of quartiles, deciles, percentiles. 1.2.3. Geometric mean and Harmonic mean 1.2.4. Combined mean 1.3. Measures of Dispersion 1.3.1. Range 1.3.2. Mean Deviation 1.3.3. Standard Deviation 1.3.4. Variance 1.3.5. Graphical Representation of mode and median 1.3.6. Find mode of data using Histogram 1.3.7. Find median using ogive curve 1.4. Skewness for given data 1.4.1. Types of skewness 1.4.2. Test of skewness 1.4.3. Measures of Skewness 1.4.4. Karl Perason's coefficient of skewness 1.4.5. Bowley's coefficient of skewnes 1.4.6. Types of skewness in terms of Mean and Mode 1.4.7. Measures of Kurtosis
Unit– II Statistical methods	2a. Fit straight line and parabolic curve using Least Square method. 2b. Calculate coefficient of correlation using Karl-Pearson's and Spearman's Rank method. 2c. Obtain the equation to the Line of Regression.	2.1 Method of Least Square 2.1.1. Fitting of Straight Line $y=a+bx$ 2.1.2. Fitting of Parabolic $y=a+bx+cx^2$ 2.2 Correlation 2.2.1. Covariance 2.2.2. Correlations, types of Correlations 2.2.3. Scatter or Dot-Diagram 2.2.4. Karl-Pearson's Co-efficient of Correlation 2.2.5. Spearman's Rank Co-efficient of Correlation 2.3 Regression 2.3.1. Line of Regression 2.3.2. Equations to Line of Regression
Unit– III Probability of Random Variable	3a. Solve simple problems on probability. 3b. Solve problems using addition and multiplication probability	3.1. Basic Definition 3.2. Definition of Probability 3.3. Addition theorem of Probability 3.4. Multiplication Theorem of

Unit	Major Learning Outcomes (in cognitive domain)	Topics and Sub-topics
	theorem. 3c. Solve problems using conditional probability and using Baye's theorem.	Probability 3.5. Conditional probability 3.6. Baye's theorem
Unit-IV Probability Distribution	4a. Solve the given problem using Binomial Distribution. 4b. Solve the given problem using Poisson Distribution. 4c. Solve the given problem using Normal Distribution.	4.1. Binomial Distribution 4.2. Poisson Distribution 4.3. Normal Distribution
Unit –V Sampling methods	5a. Solve the given problem using Sampling methods. 5b. Solve the given problem using Sampling distribution. 5c. Test samples using t-Distribution. 5d. Test samples using chi-Square Distribution. 5e. Use Chi-Square test to test the independence.	5.1. Population, Sampling: Aim of Sampling, Types of Sampling, Parameters and Statistic. 5.2. Sampling Distribution 5.3. Standard Error. 5.4. Hypothesis: Testing Hypothesis, Null Hypothesis, 5.5. Errors 5.6. Level of Significance, Test of Significance 5.7. Confidence Limits 5.8. Test of Significance of Large Samples (N>30) 5.9. Sample Distribution of proportion 5.10. Comparison of Large Samples. 5.11. The t-Distribution (For Small Samples) 5.12. Chi-Square Distribution 5.13. Condition for Chi Square test 5.14. Degree of freedom 5.15. Chi -Square Test of Goodness of fit 5.16. Chi – Square test as a test of Independence.

***Note**: To attain the COs and competency, above listed Learning Outcomes (LOs) need to be undertaken to achieve the 'Application Level' of Bloom's 'Cognitive Domain Taxonomy'.*

9. SUGGESTED SPECIFICATION TABLE FOR QUESTION PAPER DESIGN

Unit No.	Unit Title	Teaching Hours	Distribution of Theory Marks			
			R Level	U Level	A Level	Total Marks
I	Statistical Techniques	14	02	06	12	20
II	Statistical methods	10	02	04	08	14
III	Probability of Random Variable	08	02	04	04	10
IV	Probability Distribution	08	02	02	08	12
V	Sampling methods	08	02	04	08	14
	Total	**48**	**10**	**20**	**40**	**70**

Legends: *R=Remember, U=Understand, A=Apply and above (Bloom's Revised taxonomy)*
Note*: This specification table provides general guidelines to assist student for their learning and to teachers to teach and assess students with respect to attainment of LOs. The actual distribution of marks at different taxonomy levels (of R, U and A) in the question paper may vary from above table.*
This specification table also provides a general guideline for teachers to frame internal end semester practical theory exam paper which students have to undertake.

10. SUGGESTED STUDENT ACTIVITIES

Other than the classroom and laboratory learning, following are the suggested student-related ***co-curricular*** activities which can be undertaken to accelerate the attainment of the various outcomes in this course:

a. Prepare journals based on practical performed in laboratory.
b. Execute programs using different data sets in R-Programming.
c. Library/E-Book survey regarding 'Mathematics' used in Artificial Intelligence and Machine Learning.

11. SUGGESTED SPECIAL INSTRUCTIONAL STRATEGIES (if any)

These are sample strategies, which the teacher can use to accelerate the attainment of the various outcomes in this course:

a. Massive open online courses (***MOOCs***) may be used to teach various topics/sub topics.
b. ***'L' in item No. 4*** does not mean only the traditional lecture method, but different types of teaching methods and media that are to be employed to develop the outcomes.
c. About ***15-20% of the topics/sub-topics*** which is relatively simpler or descriptive in nature is to be given to the students for ***self-directed learning*** and assess the development of the LOs/COs through classroom presentations (see implementation guideline for details).
d. With respect to item No.10, teachers need to ensure to create opportunities and provisions for ***co-curricular activities***.
e. Guide student(s) in undertaking micro-projects.
f. No. of practical's selection to be performed should cover all units.

12. SUGGESTED MICRO-PROJECTS

Only one micro-project is planned to be undertaken by a student assigned to him/her in the Only one micro-project is planned to be undertaken by a student assigned to him/her in the beginning of the semester. S/he ought to submit it by the end of the semester to develop the industry oriented COs. Each micro-project should encompass two or more COs which are in fact, an integration of practicals, cognitive domain and affective domain LOs. The micro-project could be industry application based, internet-based, workshop-based, laboratory-based or field-based. Each student will have to maintain dated work diary consisting of individual contribution in the project work and give a seminar presentation of it before submission. The total duration of the micro-project should not be less than ***16 (sixteen) student engagement hours*** during the course.

In the first four semesters, the micro-project could be group-based. However, in higher semesters, it should be individually undertaken to build up the skill and confidence in every student to become problem solver so that s/he contributes to the projects of the industry. A suggestive list is given here. Similar micro-projects could be added by the concerned faculty.

a. Write a program to apply Marginal frequency distribution of predictors and response variable.
b. Write Code for ATM Card Fraud Detection System using Machine Learning Techniques using R-Programming.
c. Create a probabilistic model for credit card fraud detection
d. For a payroll dataset create Measure of central tendency and its measure of dispersion for statistical analysis of given data.
e. Case Study on probabilistic model for predicting relations in social websites system.
f. Demonstrate the case study of real direct online real estate using R language.
g. Design a microproject to analyze Uber Data.
    - Design data analysis using the R libraries like ggplot2.
    - Create a precise prediction of customers who will avail Uber trips and rides.
    - 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, the number of trips found maximum on which day of the month, etc.

13. SUGGESTED LEARNING RESOURCES

S. No.	Title of Book	Author	Publication
1	Higher Engineering Mathematics	H.K.Dass, Er.Rajnish Verma	S. Chand Technical ISBN: 9788121938907, 9788121938907
2	Statistical Methods	S. P. Gupta	Sultan Chand and sons Education Publisher, New Delhi. ISBN13: 978-93-5161-176-9
3	Higher Engineering Mathematics	B. V. Ramana	Tata Mcgraw Hill Education private limited, New Delhi. ISBN: 9780070634190, 9780070634190
4	Higher Engineering Mathematics	B. S Grewal	Khanna Publishers ISBN-10. 9788193328491 ISBN-13. 978-8193328491
5	R Programming For Dummies	Andrie de Vries, Joris Meys	Wiley india ISBN: 9788126562183
6	Introduction to R Programming	Hicham and Mohamed Ibnalkadi	(202 Non Fiction Book 4) Kindle Edition

14. SOFTWARE/LEARNING WEBSITES

a. https://www.w3resource.com/r-programming-exercises/basic/r-programming-basic-exercise-3.php
b. www.datamentor.io/r-programming/examples/
c. https://www.tutorialspoint.com/r_programming_language/index.asp
d. https://www.freecodecamp.org/news/all-the-math-you-need-in-artificial-intelligence/
e. https://byjus.com/maths/least-square-method/
f. https://www.coursera.org/learn/r-programming