Register Number:……..

KUMARAGURU
college of technology
character is life

B.E/B.TECH DEGREE EXAMINATIONS: NOV/DEC 2023

(Regulation 2018)

Third Semester

ARTIFICIAL INTELLIGENCE AND DATA SCIENCE

U18AII3204: Applied Machine Learning

COURSE OUTCOMES

CO1: Understand different methodologies to create application using statistical models.

CO2: Design the test procedures to assess the efficacy of the developed model.

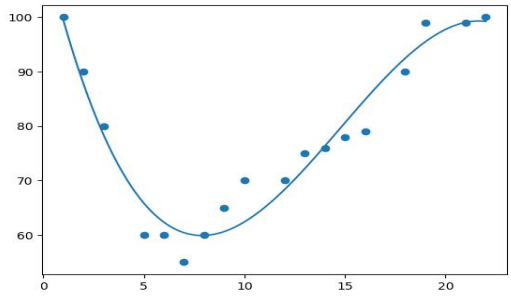
CO3: Identify and apply appropriate machine learning models for analyzing the data for a variety of problems.

CO4: Implement different algorithms for business intelligence

Time: Three Hours **Maximum Marks: 100**

Answer all the Questions:-
PART A (10 x 2 = 20 Marks)
(Answer not more than 40 words)

1.	Define supervised learning.	CO1	[K_1]
2.	Apply regression analysis between sales (y in Rs) and advertising cost (x in Rs) across all the branches of the company ABC. The data analyst discovered the regression as $\hat{y} = 300 + 20.5x$. If the advertising budgets of two branches of the company ABC differ by Rs 2000, then what will be the predicted difference in their sales?	CO1	[K_3]
3.	Explain Receiver Operating Characteristic curve in evaluating a model.	CO2	[K_2]
4.	Identify the type of regression given below and discuss the pros and cons of the model.	CO2	[K_2]



5. Discuss the impacts of models with high bias and low variance? CO2 [K_2]

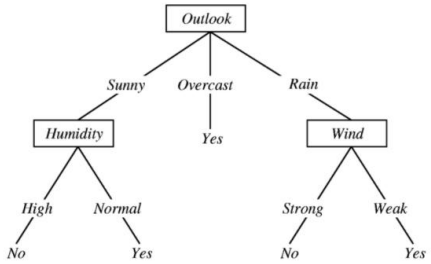
6. Interpret how AdaBoost and Gradient Boosting models are built for analytics. CO3 [K_2]

7. Describe how cross-validation is applied to select a machine learning model. CO3 [K_2]

8. Compare samples with and without replacement in decision making. CO4 [K_2]

9. Outline Classification and Regression Trees. CO4 [K_2]

10. Formulate the hypothesis for the given diagram IF (Outlook = _____) ∧ (Humidity = High) THEN Play Tennis = _____. CO4 [K_2]



Answer any FIVE Questions:-
PART B (5 x 16 = 80 Marks)
(Answer not more than 400 words)

11. a) Implement in python to fit an appropriate model for the following dataset to predict Weight based on linear combination of the given features (Age and Height) and evaluate the model. 10 CO1 [K_2]

Age	Height	Weight
45	150	56
34	157	45
27	140	60
65	151	58

b) Explain the machine learning life cycle. 6 CO1 [K_2]

12. a) Elaborate on the significance of the hyperparameter 'learning rate' in the context of gradient descent. 10 CO2 [K_3]

b) Describe the Support Vector Machine (SVM) model and demonstrate the application of the kernel trick using scikit. 6 CO2 [K_2]

13. a) Illustrate the techniques used in regression analysis to deal with overfitting. 8 CO2 [K3]

b) Develop a plan for a linear model be adapted into a classification model by utilizing suitable activation functions. 8 CO2 [K3]

14. a) A regression analysis done for the relationship between hours spent (x) in study and the scores obtained (y) in Applied Machine Learning course for 12 students yielded the following output in OLS statsmodel. 8 CO3 [K3]

OLS Regression Results

Dep. Variable:	y	R-squared:	0.338
Model:	OLS	Adj. R-squared:	0.272
Method:	Least Squares	F-statistic:	5.113
Date:	Tue, 30 Jan 2018	Prob (F-statistic):	0.0473
Time:	14:21:22	Log-Likelihood:	-41.442
No. Observations:	12	AIC:	86.88
Df Residuals:	10	BIC:	87.85
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[95.0% Conf.	Int.]
const	176.6364	20.546	8.597	0.000	130.858	222.415
x	-0.3572	0.158	-2.261	0.047	-0.709	-0.005

Omnibus:	1.934	Durbin-Watson:	1.182
Prob(Omnibus):	0.380	Jarque-Bera (JB):	1.010
Skew:	-0.331	Prob(JB):	0.603
Kurtosis:	1.742	Cond. No.	1.10e+03

A. What is the regression equation to predict score (y) as a function of hours spent (x)?

B. Briefly interpret the coefficients of this regression model.

C. Describe the evidence for a relationship between hours spent and score obtained?

D. What does the R-squared value indicate?

b) Analyze the confusion matrix of binary classification problem to distinguish between spam and non-spam emails given below and calculate the following metrics: 8 CO3 [K3]

1. Accuracy
2. Precision for the 'Spam' class
3. Recall for the 'Non-Spam' class
4. F1-score for the 'Spam' class

	Predicted	
Actual	Spam	Non-Spam
Actual Spam	450	50
Non-Spam	30	720

15. a) Given the probability of dangerous fires are rare and it is 1%, but smoke is common (10%) due to barbecue and 90% of dangerous fires make smoke. Find the probability of fire when there is a smoke. 6 CO4 [K2]

b) Consider the training examples shown in table for a binary classification problem. 10 CO4 [K3]

a. What is the entropy of this collection of training examples with respect to the positive class?

b. What are the information gains of a1 and a2 relative to these training examples?

Instance	a_1	a_2	a_3	Target Class
1	T	T	1.0	+
2	T	T	6.0	+
3	T	F	5.0	−
4	F	F	4.0	+
5	F	T	7.0	−
6	F	T	3.0	−
7	F	F	8.0	−
8	T	F	7.0	+
9	F	T	5.0	−

16. a) Summarize how AdaBoost operate in the context of a basic binary classification? 12 CO4 [K2]

b) Compare Random Forest and Aba Boost classifiers in the context of learning technique and handling overfitting. 4 CO4 [K2]

****