



P.E.S. College of Engineering, Mandya - 571 401
(An Autonomous Institution affiliated to VTU, Belagavi)
Third Semester, Master of Computer Applications (MCA)
Semester End Examination; February / March - 2022
Machine Learning using Python

Time: 3 hrs

Max. Marks: 100

Course Outcomes

The Students will be able to:

CO1: Understand the basic concepts of the python programming.

CO2: Identify the appropriate dataframes, pandas and its operations.

CO3: Apply knowledge on building, diagnose and validate the linear regression models.

CO4: Distinguish between different classification techniques.

CO5: Analyse how dataset divided in unsupervised learning techniques.

Note: I) Answer any FIVE full questions, selecting ONE full question from each unit.**II) Any THREE units will have internal choice and remaining TWO unit questions are compulsory.****III) Each unit carries 20 marks.**

Q. No.	Questions	Marks	BLs	COs	POs
1 a.	Define machine learning. Explain integrated stages for ML algorithm development.	10	L1,2	CO1	PO1,2
b.	Illustrate conditional statement, generating sequence numbers and control flow statements with examples.	10	L2	CO1	
2 a.	Explain the following with examples:				
	i) Slicing and indexing of data frame	10	L2	CO2	PO1,2,3
	ii) Sorting data frame by column values				
	iii) Creating new columns				
b.	With python code explain box plot, scatter plot, density plot.	10	L2	CO2	
3 a.	Explain in detail about simple linear regression and multiple linear regressions.	10	L2	CO3	PO1,2,3
b.	What is outlier analysis? Explain any three distance measures are useful in identifying influential observations.	10	L2	CO3	PO1,2

OR

3 d. The values of independent variable x and dependent values y are given below.

x	y
1	2.4
2	3
3	3.6
4	4
6	5
8	6

10 L1,5 CO3 PO1,2,3

Find the least square regression line $y = ax + b$. Estimate the value of y when x is 5.

e.	What is multi-collinearity? How it is impact on the model.	10	L1,2	CO3	PO1,2
----	--	----	------	-----	-------

- | | | | | |
|------|--|----|------|-------------|
| 4 a. | Discuss about Variance Inflation Factor and explain how checking correlation of columns with large VIFs can be done? | 10 | L1,2 | CO4 PO1,2,3 |
| b. | Demonstrate confusion matrix with example. | 6 | L2 | CO4 PO1,2,3 |
| c. | Name the measuring accuracies used in classification for measuring model performance. | 4 | L1 | CO4 PO1,2 |

OR

- | | | | | |
|------|--|----|----|-------------|
| 4 d. | Explain about decision tree learning and write the steps used to generate classification and regression model. | 10 | L4 | CO4 PO1,2 |
| e. | Name two measures used for measuring the benefits of using the logistic regression model and explain it. | 10 | L4 | CO4 PO1,2 |
| 5 a. | Explain <i>k</i> -means clustering with an example. | 10 | L2 | CO5 PO1,2,3 |
| b. | Write note on : | | | |
| | i) Dendrogram | 10 | L2 | CO5 PO1,2 |
| | ii) Elbow method | | | |

OR

- | | | | | |
|------|--|----|----|-------------|
| 5 d. | Explain Hierarchical clustering algorithm with an example. | 10 | L2 | CO5 PO1,2,3 |
| e. | Explain Random forest in detail. | 10 | L2 | CO5 PO1,2,3 |

* * * *