Roll No.

ANNA UNIVERSITY (UNIVERSITY DEPARTMENTS)

B.E. /B. Tech / B. Arch (Full Time) - END SEMESTER EXAMINATIONS, APR / MAY 2024

MINOR DEGREE IN AI ML
VI SEMESTER
ITM505 – MLOPS
(Regulation 2019)

Time:3 hrs

Max. Marks: 100

CO1	Train, validate and Test the Learning models
CO2	Understand the workflow of MLOPS
CO3	Develop a user interface with user defined parameters for ML algorithms
CO4	Deploy ML as a web application in local server/ cloud
CO5	Create a Docker image for ML
CO6	Deploy ML models in real world environment

BL – Bloom's Taxonomy Levels
(L1-Remembering, L2-Understanding, L3-Applying, L4-Analysing, L5-Evaluating, L6-Creating)

PART- A (10x2=20Marks)
(Answer all Questions)

Q. No.	Questions	Marks	CO	BL
1	Name any four Python libraries that you use for MLOPS	2	1	2
2	Assume you have a data with text and data. How will you preprocess the text?	2	1	3
3	What is the need for MLOPS?	2	2	3
4	Are there any challenges in MLOPS? Explain	2	2	3
5	Assume you are using a Decision tree for classification purposes. Which hyper-parameters will you use?	2	3	2
6	What are the metrics used for Regression problems?	2	3	2
7	Name any two user-interface tools that you would use for development of ML application.	2	6	3
8	How is CI/CD carried out in MLOPS?	2	5	2
9	What are the advantages of using Containers in MLOPS	2	4	3
10	Write any four functions of mlflow in MLOPS			

PART- B (5x 13=65Marks)
(Restrict to a maximum of 2 subdivisions)

Q. No.	Questions	Marks	CO	BL
11 (a)	(I) Assume you have the data of 100 students marks with the columns of Maths, English and Science courses. (a) Write a program to take these data and convert to a dataframe (b) Using the require python library, extend the program to compute the statistics for these data.	8	1	3
	(ii) Write a Pandas program to add, subtract, multiple and divide two Pandas Series	5	1	3
	OR			
11 (b)	Consider this dataset . Write python program for exploratory data analysis and preprocessing according to these questions	13	1	3

ID	Name	Age	Gender	Occupation	Salary
1	Alice	28	Female	Data Scientist	$85,000
2	Bob	35	Male	Software Engineer	$95,000
3	Charlie	40	Male	Project Manager	$105,000
4	Diana	32	Female	UX Designer	$78,000
5	Ethan	45	Male	CTO	$150,000

(I) What is the overall structure and shape of the dataset? How many rows and columns are there?
(ii) Convert the categorical variable to numeric form.
(iii) How will you treat missing values in the dataset?
(iv) How will you check the pattern of input variables?
(v) How will you prepare the data without outliers?

12 (a)	You have been given a dataset that captures the growth of a certain species of plant over time. The dataset includes the number of days since planting and the corresponding height of the plant in centimeters. Preliminary analysis suggests that the relationship between the days since planting and the height of the plant is non-linear. Your task is to build a I regression model to better capture this relationship. Day — Height 1 — 2.1 2 — 4.2 3 — 6.3 4 — 8.3 5 — 10.6 (i) Develop regression model for this dataset. (ii) Compute the metrics of RMSE or MAE. (iii) What will you do to increase the accuracy?	13	1	4
	OR			
12 (b)	How is logistic regression different from linear and polynomial regression? With an example of your choice, provide the steps for fitting a logistic regression model and provide the appropriate metrics.	13	1	4
13 (a)	With an example, explain the classification algorithm behind KNN. What are the different metrics available for the classification? Explain.	13	1	4
	OR			
13 (b)	(I) Consider this dataset on diabetes detection. You have to find whether the particular person is diagnosed with diabetes or not based on the factors such as Age, BMI, Blood Pressure and Glucose level. How will you use a decision tree for this problem? Mention the steps of training, validation and testing.	13	1	4

Patient_ID	Age	BMI	Blood Pressure	Glucose Level	Diagnosed
1	45	28.5	130	160	Yes
2	34	24.0	120	140	No
3	50	30.0	135	180	Yes
4	29	22.5	115	130	No
5	60	27.0	145	200	Yes

14 (a)	(I) Differentiate supervised and unsupervised algorithm and mention examples for each. (ii) A new medical test has been developed to diagnose a particular disease. The test was evaluated on a sample of 1000 patients, and the results are summarized in the confusion matrix below: a. How many true positives are there? b. How many true negatives are there c. Calculate Precision d. Calculate Recall e. Find F1 score What do these values mean? How does it impact the workflow of MLOPS?	3 10	2	3

	Disease Positive	Disease Negative
Test Positive	150	30
Test Negative	50	770

OR

14 (b)	Consider the dataset of penguin to determine the species given the other parameters. This has already been trained and tested and stored as .'pkl' file using scikit learn library. You are now given the task to deploy it on to the internet with a simple user-interface. How will you do that? Mention the detailed steps. Are there any disadvantages with this approach?	13	2	3

	species	island	culmen_length_mm	culmen_depth_mm	flipper_length_mm	body_mass_g	sex
0	Adelie	Torgersen	39.1	18.7	181.0	3750.0	MALE
1	Adelie	Torgersen	39.5	17.4	186.0	3800.0	FEMALE
2	Adelie	Torgersen	40.3	18.0	195.0	3250.0	FEMALE
4	Adelie	Torgersen	36.7	19.3	193.0	3450.0	FEMALE
5	Adelie	Torgersen	39.3	20.6	190.0	3650.0	MALE

15 (a)	(I) How are the traditional software developed in windows or Linux Operating Systems? What are the disadvantages? (ii) What is a docker? How is it advantageous over the traditional methods? (iii) Mention the steps to create a docker for a simple python application with the required files.	4 4 5	3,6	4

OR

15 (b)	Assume you have a trained ML model in the extension of .pkl in the iris dataset consisting of Sepal width, Sepal length, Petal	13	3, 6	4

	(I) Write a python program to classify the type of flowers given the four parameters as user input. (ii) Write a Dockerfile with the required packages and commands (iii) Mention the steps to build the docker image and run it. (iv) Assume that you have made a mistake in the program and correcting it. Is it required to rebuild and run the docker again? Why or why not?			

PART- C (1x 15=15Marks)

(Q.No.16 is compulsory)

Q. No.	Questions	Marks	CO	BL
16.	(I) Explain the MLOPS workflow in detail with an example (ii) Consider this code snippet. What is the essence of this code and how is it useful in MLOPS. Explain each line of the code in a sentence or two (You can write this explanation against the line of code).	8 7	2, 6	4

```
mlflow.set_tracking_uri(uri="http://127.0.0.1:8080")
mlflow.set_experiment("MLflow test")
with mlflow.start_run():
    mlflow.log_params(params)
     mlflow.log_metric("accuracy", accuracy)
     mlflow.set_tag("Training Info", "Basic LR model for
iris data")
    signature =
infer_signature(X_train,lr.predict(X_train))
       model_info = mlflow.sklearn.log_model(
        sk_model=lr,
        artifact_path="iris_model",
        signature=signature,
        input_example=X_train,
        registered_model_name="tracking-test",
    )
```