

# Maulana Azad National Urdu University

Department of Computer Science & Information Technology

M.Tech, 3<sup>rd</sup> Semester Examination, December 2023

MTCS314PET : Data Preparation and Analysis

Total Marks: 70

Time: 3 Hours

ہدایات:

یہ پڑچہ سوالات تین حصوں پر مشتمل ہیں: حصہ اول، حصہ دوم، حصہ سوم۔ ہر جواب کے لیے لفظوں کی تعداد اشارہ ہے تمام حصوں سے سوالوں کا جواب دینا لازمی ہے۔  
1. حصہ اول میں 10 لازمی سوالات ہیں جو کہ معروضی سوالات/خالی جگہوں پر کرنا/مختصر جواب والے سوالات ہیں ہر سوال کا جواب لازمی ہے ہر سوال کے لیے 1 نمبر ہے

(10x1=10 marks)

2. حصہ دوم میں 08 سوالات ہیں، اس میں سے طالب علم کو کوئی پانچ سوالوں کے جواب دینے ہیں ہر سوال کا جواب تقریباً (200) لفظوں پر مشتمل ہے

(5x6=30 marks)

ہر سوال کے لیے 6 نمبرات مختص ہیں

3. حصہ سوم میں 05 سوالات ہیں، اس میں سے طالب علم کو کوئی تین سوالوں کے جواب دینے ہیں ہر سوال کا جواب تقریباً (500) لفظوں پر مشتمل ہے

(3x10=30 marks)

ہر سوال کے لیے 10 نمبرات مختص ہیں

## حصہ اول

i. One Hot Encoding کیا ہے؟

ii. Interpretable models کیا ہے؟

iii. Ensemble model کیا ہے؟

iv. Hyperparameter کیا ہے؟

v. Underfitting کیا ہے؟

vi. Overfitting کیا ہے؟

vii. Validation-Cross کیا ہے؟

viii. Outlier کیا ہے؟

ix. Scatter Plot کیا ہے؟

x. Feature Vector کیا ہے؟

## حصہ دوم

- .2 Data Discovery اور profiling کو سمجھاؤ؟ ngData profili میں inconsistencies اور anomalies کو کیسے ٹھیک کرتے ہیں؟
- .3 Missing values کے removal سے Overall data sets کی integrity اور statistical analysis پر کیا فرق پڑیگا؟ وضاحت کریں؟
- .4 Data Transformation techniques کیا ہے؟ Data distribution کی normality پر اسکا کیا اثر پڑتا ہے؟
- .5 Data analysis میں ensemble model کیا ہے؟ Imbalance data پر یہ کیسے کام کرتا ہے؟
- .6 Random sampling اور stratified sampling میں فرق کو واضح کرو؟ Sampling کے applications کو ایک مثال سے سمجھاؤ؟
- .7 Survey، interviews اور observations کے data collection methods کے advantages اور disadvantages کی وضاحت کرو؟
- .8 Dataset میں duplicate data کو identify اور handle کرنے کے methods کو واضح کرو؟ مثال سے سمجھاؤ؟
- .9 Descriptive اور inferential statistics میں فرق کو واضح کرو؟ Regression Analysis کے common tools کو بتاؤ؟

10. Dataset میں categorical data کو کیسے handle کرتے ہیں؟ اس میں Encoding Techniques کیسے استعمال ہوتی ہے؟
11. Data Preparation اور Analysis میں scaling method کیا ہے؟ Scaling method سے clustering algorithms کی performance پر کیا فرق پڑتا ہے؟
12. Feature Engineering کیا ہے اور data preparation میں اس کا significance کیا ہے؟ Feature Engineering Techniques کے examples کی وضاحت کرو؟
13. Dimensionality Reduction میں Principal Component Analysis کے purpose کو سمجھاؤ؟ Principal Component Analysis کی working کو سمجھاؤ
14. Diverse Clustering algorithm کیا ہے؟ Linear Regression کی accuracy کو clustering کی مدد سے کیسے improve کر سکتے ہیں؟ جواب کو ایک مثال کے ساتھ سمجھاؤ؟