مولانا آزاد نیشنل اردو یونیورسٹی

Semester Examinationسمسٹر امتحانات Nov 2025

Programme: B.Tech. (Computer Science) :پروگرام

Semester: V :سمسٹر

Title & Paper Code: Data science (BTCS514PET) :مضمون مع کوڈ

گھنٹے 3 Hrs :وقت Time   Maximum Marks 70 :جملہ نشانات

ہدایات:

یہ پرچہ تین حصوں پر مشتمل ہے: حصہ اول، حصہ دوم، حصہ سوم. ہر جواب کے لئے لفظوں کی تعداد اشارةً ہے. تمام حصوں سے سوالوں کا جواب دینا لازمی ہے۔

1. حصہ اول میں 10 لازمی سوالات ہیں جو کہ مختصر جواب والے سوالات ہیں. ہر سوال کا جواب دینا لازمی ہے ہر سوال کے لئے 1 نمبر مختص ہے۔ **(Marks 10 =1 x 1)**

2. حصہ دوم میں 8 سوالات ہیں، اس میں طالب علم کو کوئی پانچ سوالوں کے جواب دینے ہیں ہر سوال کا جواب تقریباً دو سو (200) پر مشتمل ہے ہر سوال کے لئے 6 نمبرات مختص ہیں۔ **(Marks 30 =6 x 5)**

3. حصہ سوم میں پانچ سوالات ہیں۔ اس میں سے طالب علم کو کوئی تین سوالوں کے جواب دینے ہیں. ہر سوال کا جواب تقریباً پانچ سو (500) لفظوں پر مشتمل ہے. ہر سوال کے لئے 10 نمبرات مختص ہیں۔ **(Marks 30 =10 x 3)**

حصہ اول

1. مندرجہ ذیل سوال کے جواب لکھیں ۔

(i) sample کے mean اور population کے mean میں فرق کریں۔

(ii) ایک Python library کا نام بتائیں جو عام طور پر plotting data کے لیے استعمال ہوتی ہے۔

(iii) کونسی Python لائبریری بنیادی طور پر data manipulation اور analysis کے لیے استعمال ہوتی ہے؟

(iv) Central tendency کا کون سا measure آؤٹ لیر (outliers) سے سب سے زیادہ affected ہوتا ہے؟

(v) Supervised learning میں استعمال ہونے والے تین main datasets کون سے ہیں؟

(vi) کون سا algorithm عام طور پر unsupervised learning میں clustering کے لیے استعمال ہوتا ہے؟

(vii) Pandas میں کون سا فنکشن DataFrame کی first five rows کو display کرنے کے لیے استعمال ہوتا ہے؟

(viii) Data distribution میں skewness کا کیا مطلب ہے؟

(ix) Variance اور standard deviation کے درمیان کیا relationship ہے؟

(x) machine learning model میں overfitting کیا ہے؟

حصہ دوم

2. 10 طلباء کے نمبر 8، 12، 15، 17، 19، 21، 24، 26، 28، اور 30 ہیں۔ meanاور standard deviation کا حساب لگائیں، پھر نیا mean اور standard deviation تلاش کریں اگر ہر mark کو 2 سے multiplied کیا جائے۔

3. مندرجہ ذیل جدول (table) میں 6 طلباء کے مطالعے کے گھنٹوں کی تعداد (number of hours studied) اور حاصل کردہ نمبروں (marks obtained) کو دکھاتا ہے:

Student	Hours Studied (X)	Marks Obtained (Y)
1	1	45
2	3	55
3	5	60
4	6	70
5	8	80
6	9	88

a) linear regression equation (Y=a+bX) تلاش کریں۔

b) 7 گھنٹے study کرنے والے طالب علم کے marks obtained کو Predict کریں۔

4. مناسب مثالوں کے ساتھ data Preparation اور Exploratory Data Analysis (EDA) کے processکی وضاحت کریں۔

5. Data Science میں Pythonکی اہمیت کی وضاحت کریں۔ مثالوں کے ساتھ کسی بھی تین fundamental Python libraries (NumPy, Pandas, Matplotlib)کے استعمال پر بحث کریں۔

6. Descriptive analyticsمیں استعمال ہونے والے five-number summary statistics کی وضاحت کریں، اور بیان کریں کہ وہ datasetکی distributionکی understanding میں کس طرح help کرتے ہیں۔

7. ایک dataset میں 12 نمبر ہیں:

35, 42, 48, 50, 55, x, 60, 62, 65, 68, 70, 72

تمام 12 نمبروں کا mean 57.25 ہے۔ اگر x کی value 5 سے decreased کیا جائے تو نیا mean 56.83 بن جاتا ہے۔

x کی original value تلاش کریں۔

8. machine learning میں feature selection کے concept کی وضاحت کریں۔ model performance اور generalization کے لیے یہ کیوں important ہے؟

9. Data Science کو Define کریں اور مناسب مثالوں کے ساتھ data science life cycle میں اس کے key stagesکی وضاحت کریں۔

حصہ سوم

10. طلباء کے ناموں اور ان کے اسکور کے درج ذیل ڈیٹاسیٹ کو دیکھتے ہوئے:

Student	Aisha	Rahul	Neha	Imran	Zoya	Arjun
Score	78	92	85	67	90	82

اس dataset کو پڑھنے کے لیے Pandas کا استعمال کرتے ہوئے Python کوڈ لکھیں، 80 سے زیادہ score کرنے والے تمام طلبہ کو filter کریں، اور پھر Matplotlib کا استعمال کرتے ہوئے ان کے scores bar chart بنائیں۔

11. Machine learning میں overfitting اور underfitting کے concepts کی وضاحت کریں ۔ مثال کے ساتھ بتائے کی model کے performance پر کس تارہ affects کرتا ہے۔

12. Mean، variance، اور standard deviation کی وضاحت کریں۔ وضاحت کریں کہ وہ ڈیٹا spread اور central tendency کو سمجھنے میں کس طرح مدد کرتے ہیں۔

13. Machine learning کی وضاحت کریں اور اس کی major types کی وضاحت کریں (suoervised,unsupervised and reinforcement learning)۔ ہر ایک کے لیے ایک real time مثال لکھے۔

14. آپ کو درج ذیل موسم کا dataset دیا گیا ہے۔

The target is **Play?** (Yes/No)

Outlook	Temperature	Humidity	Windy	Play?
Sunny	Hot	High	False	No
Sunny	Hot	High	True	No
Overcast	Hot	High	False	Yes
Rainy	Mild	High	False	Yes
Rainy	Cool	Normal	False	Yes
Rainy	Cool	Normal	True	No
Overcast	Cool	Normal	True	Yes
Sunny	Mild	High	False	No

a) "Outlook" attribute کا استعمال کرتے ہوئے، Information Gain نیکالے۔

b) Based on your result, should the decision tree split on *Outlook* first

S-17