Maulana Azad National Urdu University

कंप्यूटर विज्ञान और सूचना प्रौद्योगिकी विभाग شعبہ کمپیوٹر سائنس و انفارمیشن ٹکنالوجی

Department of Computer Science & Information Technology

Semester Examination, Nov – Dec 2025

Branch/Year/Sem: B.Tech / IV/ VII
Duration: 03:00:00 hrs

Paper Code: BTCS716PET
Max Marks: 70

Paper Title: Machine Learning

ہدایات:

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

1. حصہ اوّل میں 10 لازمی سوالات ہیں جو کہ مختصر جواب والے سوالات ہیں۔ ہر سوال کا جواب دینا لازمی ہے۔ ہر سوال کے لیے 1 نمبر مختص ہے۔(10 x 1 = 10 Marks)
2. حصہ دوم میں 8 سوالات ہیں۔ اس میں طالب علم کو کوئی ۵ سوالوں کے جواب دینے ہیں۔ ہر سوال کا جواب تقریباً دو سو پچاس (250-300) پر مشتمل ہے۔ ہر سوال کے لئے 6 نمبرات مختص ہیں۔ (5x6 = 30 Marks)
3. حصہ سوم میں 5 سوالات ہیں۔ اس میں سے طالب علم کو کوئی 3 سوال کے جواب دیں۔ ہر سوال کے جواب تقریباً پانچ سو (500) لفظوں پر مشتمل ہے۔ ہر سوال کے لئے 10 نمبرات مختص ہے۔ (3 x 10 = 30 Marks)

حصہ اول

سوال نمبر 1:

۱۔ مشین لرننگ (Machine Learning) کی تعریف کریں۔

۲۔ مشین لرننگ ایک ایسا مطالعے کا میدان (field of study) ہے جو کمپیوٹرز کو بغیر __________ سیکھنے کی صلاحیت دیتا ہے۔

۳۔ درج ذیل میں سے کون سا سپروائزڈ لرننگ (Supervised Learning) الگورتھم ہے؟

a) K-Means b) Linear Regression c) PCA d) Aprior

۴۔ ڈیٹاسیٹ کو ٹریننگ اور ٹیسٹنگ ذیلی سیٹوں (Subset) میں تقسیم کرنے کے عمل (process of dividing) کو __________ کہا جاتا ہے۔

۵۔ کلاسفکیشن مسائل (classification problems) کے لیے عام طور پر کون سا تشخیصی میٹرک (evaluation metric) استعمال ہوتا ہے؟

a) Mean Squared Error b) Accuracy c) Silhouette Score d) Euclidean Distance

۶۔ درج ذیل میں سے کون سی تکنیک ڈائمنشنلیٹی ریڈکشن (Dimensionality Reduction) کے لیے استعمال ہوتی ہے؟

a) PCA b) SVM c) KNN d) Decision Tree

۷۔ عام تقسیم (Normal Distribution) میں ڈیٹا کا کتنا فیصد IQR میں آتا ہے۔

a) 25% b) 50% c) 75% d) 100%

۸۔ سیگموئڈ (Sigmoid) فنکشن کی مساوات (equation) لکھیں اور بتائیں کہ اس فنکشن کا آوٹ پٹ کس حد (range) میں ہوتا ہے؟

۹۔ اگر کسی ماڈل میں Bias زیادہ اور Variance کم ہو تو اس کا نتیجہ کیا ہوگا؟

۱۰۔ اگر کسی ڈیٹاسیٹ میں ایک ویلیو (value) باقی تمام ویلیوز سے بہت زیادہ مختلف ہو تو اُسے کیا کہتے ہیں؟

24/11/2025

حصہ دوم

2. مثل کے ساتھ ذیل میں دی گئی terms کے بارے میں وضاحت کریں۔

i. Bias, Variance

ii. Overfitting, Underfitting, Best fit

iii. Interquartile Range (IQR)

iv. Measure of Central Tendency

3. آؤٹ لائر (Outlier) کیا ہے؟ اسے IQR طریقہ استعمال کرتے ہوئے outlier کا شناخت کرے اور ہینڈل کرنے کا طریقہ بیان کریں۔

دیا ہوا dataset: [50, 52, 49, 51, 48, 47, 120, 52]

4. لینئر ریگریشن (Linear Regression) کی وضاحت کریں اور دیے گئے مسئلے کو حل کریں (using tabular method)۔

I. find the best fit line equation.

II. find β1(Slope) value.

III. Get the value of Standard Error of the Estimate (SES).

S.No	X(Independent)	Y(Dependent)
1	2	3
2	4	1
3	3	2
4	1	5
5	5	0
6	2.5	?
7	1.5	?

5. کنفیوژن میٹرکس (Confusion Matrix) کو اس کے فہر مثلاً کے ساتھ وضاحت کریں اور دیے گے مسئلے کو حل کرے۔

		Predicted	
Actual		Yes	No
	Yes	45	50
	No	36	35

I. Find accuracy

II. Precision

III. Recall

IV. Error Rate

6. دو متغیری تجزیہ (Bivariate Analysis) کی تعریف کریں اور اس کا مقصد بیان کریں۔ نیچے دیے گئے اعداد و شمار کی مدد سے مطالعہ کے گھنٹوں (X) اور امتحانی نمبروں (Y) کے درمیان تعلقی عدد (Correlation Coefficient, r) معلوم کریں اور تعلق کی نوعیت بیان کریں۔

X (Hours)	2	3	4	5	6	7	8
Y (Scores)	60	85	60	65	80	70	45

7. (a) Z-Test اور T-Test میں فرق بیان کریں۔ یہ بھی وضاحت کریں کہ ہر ٹیسٹ کو کب استعمال کیا جاتا ہے۔(2M)

(b) ایک استاد کا دعویٰ ہے کہ طلباء کے ٹیسٹ کے اوسط نمبر 70 ہیں۔ 10 طلباء کے نمونے میں درج ذیل نمبر حاصل ہوئے:

72, 68, 75, 70, 74, 69, 71, 73, 70, 68.

آبادی (Population) کا معیاری انحراف (Standard Deviation) نامعلوم ہے۔ 5% معنی خیزی کی سطح (Significance Level) پر یہ پرکھیں کہ کیا استاد (Teacher) کا دعویٰ (claim) درست ہے یا نہیں۔(4)

8. ANN کی وضاحت کریں اور Perceptron کے کام کرنے کا طریقہ ایک صاف ستھرے خاکے (Diagram) کی مدد سے بیان کریں۔ ساتھ ہی Activation Function اور اس کی اقسام (types) (Tanh, Sigmoid,ReLU, Leaky ReLU) کی وضاحت مناسب خاکوں کے ساتھ کریں۔

9. ایک Python پروگرام لکھیں جو Pandas لائبریری کا استعمال کرتے ہوئے درج ذیل کام انجام دے:

a. اپنی مقامی ڈائریکٹری (Local System) سے کوئی ڈیٹا فائل مثلاً (*data.csv*) لوڈ کر کے ایک DataFrame میں محفوظ کریں۔

b. مندرجہ ذیل بنیادی آپریشنز (Basic Operations) انجام دیں اور نتائج ظاہر کریں:

o head() اور tail()

24/11/2025

- ○ describe() اور Info ()
- ○ dtypes اور shape
- ○ مخصوص کالمز یا قطاریں (Rows/Columns) حاصل کریں
- ○ isnull().sum() کے ذریعے غائب قدریں (Missing Values) معلوم کریں

ایک Dummy Dataset استعمال کریں اور مکمل کوڈ بمعہ آؤٹ پٹ کی وضاحت لکھیں۔

حصہ سوم

---

10. Clustering کی وضاحت کریں اور مشین لرننگ میں اس کی اہمیت بیان کریں۔ کلسٹرنگ کی مختلف اقسام - Hierarchical based Partition- based (K-Means), اور Density-based (DBSCAN) کو واضح خاکوں (Diagrams) اور مناسب مثالوں کے ساتھ بیان کریں۔

11. Ensemble Learning کی وضاحت کریں اور بتائیں کہ یہ مشین لرننگ میں ماڈل کی درستگی (Accuracy) کو کیسے بہتر بناتا ہے۔ اس کی بنیادی اقسام — Bagging, Boosting اور Stacking — کو واضح خاکوں (Diagrams)، کام کرنے کے اصول (Working Principles) اور مناسب مثالوں کے ساتھ تفصیل سے بیان کریں۔

12. (a). Hypothesis Testing, Population اور Sample کی تعریف کریں۔ ایک مثال کے ساتھ Hypothesis Testing کے بنیادی مراحل لکھیں۔
(b). Type-I اور Type-II Errors کی وضاحت کریں اور انہیں ایک جدول (Table) کی شکل میں پیش کریں جس میں فیصلہ (Decision) اور حقیقی حالت (Actual Condition) ظاہر ہو۔

13. (a) Overfitting، Underfitting اور Best Fitting کی وضاحت کریں ایک موزوں خاکے (Diagram) کے ساتھ۔ ساتھ ہی Global Minima اور Local Minima کے تصورات بیان کریں۔
(b). Confusion Matrix کی تعریف کریں اور اس کے اجزاء (TP, TN, FP, FN) کی وضاحت کریں۔ مزید یہ کہ Accuracy، Precision، Recall اور F1-Score کے فارمولے ایک مثال کے ساتھ لکھیں۔

14. ذیل میں دیے گئے ڈیٹاسیٹ کی مدد سے ایک Decision Tree بنائیں۔ ہر مرحلے میں Entropy اور Information Gain (IG) کا حساب لگائیں اور بہترین فیچر (feature) کو Root Node کے طور پر منتخب کریں۔ اس کے بعد درخت کو دوسرے لیول تک بنائیں۔ تمام مراحل کے حساب کے ساتھ نتیجہ اور درخت کا خاکہ واضح طور پر لکھیں۔

#	Weather	Temperature	Humidity	Wind	Play
1	Sunny	Hot	High	Weak	No
2	Sunny	Hot	High	Strong	No
3	Overcast	Hot	High	Weak	Yes
4	Rain	Mild	High	Weak	Yes
5	Rain	Cool	Normal	Weak	Yes
6	Overcast	Cool	Normal	Strong	Yes
7	Sunny	Mild	High	Weak	No
8	Rain	Mild	Normal	Strong	No

-------{ Dataset for creating Decission Tree}

#	Weather	Temperature	Humidity	Wind	Play (Predicted?)
1	Sunny	Cool	Normal	Weak	?
2	Overcast	Mild	High	Strong	?
3	Rain	Cool	Normal	Strong	?
4	Sunny	Mild	High	Strong	?

------{ Dataset for validation for the feature (Play) }

24/11/2025