

Maulana Azad National Urdu University
M.C.A, III Semester Examination, February 2022
Paper : MMCA311PCT : Data Science

پرچہ : ڈیٹا سائنس

Time : 3 hrs

Marks : 70

ہدایات:

یہ پرچہ سوالات دو حصوں پر مشتمل ہے: حصہ اول اور حصہ دوم۔ ہر جواب کے لیے لفظوں کی تعداد اشارہ ہے۔ تمام حصوں سے سوالوں کا جواب دینا لازمی ہے۔

1. حصہ اول میں 10 سوالات ہیں، اس میں سے طالب علم کو کوئی 08 سوالوں کے جواب دینے ہیں۔ ہر سوال کا جواب تقریباً سو (100) لفظوں پر مشتمل ہے ہر سوال کے لیے 05 نمبرات مختص ہیں۔
(8 x 5 = 40 Marks)
2. حصہ دوم میں 05 سوالات ہیں۔ اس میں سے طالب علم کو کوئی 03 سوال کا جواب دینا ہے۔ سوال کا جواب تقریباً ڈھائی سو (250) لفظوں پر مشتمل ہے سوال کے لیے 10 نمبرات مختص ہیں۔
(10x3 = 30 Marks)

حصہ اول

1. MapReduce سے آپ کیا سمجھتے ہیں؟ Big Data کے ساتھ MapReduce کب استعمال کرتے ہیں۔ MapReduce کا استعمال کرتے ہوئے Word Count Problem کو حل کریں۔
2. Hadoop سے آپ کیا سمجھتے ہیں؟ موزوں مثالوں کے ساتھ Hadoop کے مختلف Components اور Features کی وضاحت کریں۔
3. Big Data کیا ہے؟ اور یہ کہاں سے آتا ہے؟ یہ کیسے کام کرتا ہے؟ Big Data میں V5 کیا ہیں؟ وضاحت کریں۔
4. HDFS کی وضاحت کریں۔ HDFS روایتی NFS سے کیسے مختلف ہے؟
5. Data Science کیا ہے؟ تفصیلات میں Data Science کے Process کے کردار کی وضاحت کریں۔
6. BigData کیا ہے؟ Bigdata میں چار V کی اہمیت کی وضاحت کیجیے۔
7. NoSQL سے کیا مراد ہے؟ NoSQL کی خصوصیات کی وضاحت کیجیے۔
8. Hadoop Architecture کے بارے میں مختصراً بیان کریں۔
9. Data Science میں Data کی مختلف اقسام کی وضاحت کریں۔ مناسب مثال کے ساتھ وضاحت کریں۔
10. ایک Python Script لکھیں جس میں متعلقہ Packages اور CSV اور EXL فائلوں تک Access کا طریقہ کار شامل ہو۔ ضروری مثال کے ساتھ وضاحت کریں۔

حصہ دوم

11. Data Cleaning اور Sampling کے بارے میں وضاحت کریں۔ Data Cleaning تجزیہ میں کس طرح اہم کردار ادا کرتی ہے؟
12. Logistics Regression کیا ہے؟ ایک مثال بیان کریں جب آپ نے حال ہی میں Logistic Regression کا استعمال کیا ہو۔
13. Machine Learning کے Regression Techniques اور Classification میں فرق واضح کیجیے۔
14. Machine Learning سے آپ کیا سمجھتے ہیں؟ Data Science میں اس کا کیا کردار ہے؟ Supervise Learning اور Unsupervise Learning کے درمیان فرق کی فہرست بنائیں۔
15. Python میں Multiple Variable کے درمیان Regression Analysis کا تجزیہ کیسے کریں۔ CATS ڈیٹا سیٹ میں Heart کے Weight اور Body Weight سے Linear Equation $y=mx+c$ حاصل کرنے کے لیے ایک Python کوڈ لکھیں۔