राष्ट्रीय प्रौद्योगिकी संस्थान हमीरपुर

हमीरपुर (हि.प्र.) — 177 005 (भारत)

(एक राष्ट्रीय महत्व का संस्थान शिक्षा मंत्रालय के तहत)

NATIONAL INSTITUTE OF TECHNOLOGY HAMIRPUR

HAMIRPUR (H.P.) - 177 005 (INDIA)

(An Institute of National Importance under Ministry of Education)

B.Tech. End Semester Examination

Roll No.

Branch : Computer Science & Engineering
Semester: 7th
Subject: Data Analytics

Subject Code: CS-432
Max Marks: 50
Time Allowed: 3 Hours

Note: All questions are compulsory.

Q1 a) A survey on cars was conducted in 2021 and determined that 60% of car owners have only one car, 28% have (5) two cars, and 12% have three or more. Supposing that you have decided to conduct your own survey and have collected the data below, determine whether your data supports the results of the study. Use a significance level of 0.05. Also, given that, out of 129 car owners, 73 had one car and 38 had two cars. (Given: the table chi-square value with 2 degrees of freedom at the given significance level of 0.05 is 5.991).

b) The following are the number of mistakes made in 5 successive days by 4 technicians working for a photographic laboratory. Test whether the difference among the four sample means can be attributed to chance. [Given: use one-way analysis of variance test and the table value at 1% level of significance is 8.7].

I	II	III	IV
6	14	10	9
14	9	12	12
10	12	7	8
8	10	15	10
11	14	11	11

Q2 a) What is the difference between ANOVA and Kruskal-Wallis tests? Three products received the following (7) performance rating by a panel of 20 customers. Use the Kruskal-Wallis test to determine whether there is a significant difference in the performance ratings for the product at a significance level of 0.05.

Product		
A	**B**	**C**
25	60	50
70	20	70
60	30	60
85	15	80
95	40	90
90	35	70
80		75

(Given: the table chi-square value with 2 degrees of freedom at the given significance level of 0.05 is 5.991).

b)   What are various clustering techniques? Discuss the Agglomerative algorithm with the following data and plot   (8)
a Dendrogram using a single link approach. The table below comprises sample data items indicating the
distance between the elements.

Item	E	A	C	B	D
E	0	1	2	2	3
A	1	0	2	5	3
C	2	2	0	1	6
B	2	5	1	0	3
D	3	3	6	3	0

Q3 a)   Consider the weather dataset given below: (10)

Day	Weather	Temperature	Humidity	Wind	Play?
1	Sunny	Hot	High	Weak	No
2	Cloudy	Hot	High	Weak	Yes
3	Sunny	Mild	Normal	Strong	Yes
4	Cloudy	Mild	High	Strong	Yes
5	Rainy	Mild	High	Strong	No
6	Rainy	Cool	Normal	Strong	No
7	Rainy	Mild	High	Weak	Yes
8	Sunny	Hot	High	Strong	No
9	Cloudy	Hot	Normal	Weak	Yes
10	Rainy	Mild	High	Strong	No

Make a classification model using the ID-3 decision tree approach to predict whether the game will be played
or not on a particular day.

b)   What are the three V's of Big Data? Give two examples of Big Data case studies. Indicate which V's are   (5)
satisfied by these case studies.

Q4 a)   Discuss the Naive Bayes classifier.   (2*5=10)

b)   Explain the difference between linear and logistics regression with an example.

c)   What do you understand by Hadoop Map Reduce? Explain Map and Reduce task.

d)   What is the primary goal of feature engineering?

e)   Explain association rules analysis.