survival8

BITS WILP Data Mining Mid-Sem Exam 2017-H1 (Regular)

Index Subjects Mail Us

Download solutions

Birla Institute of Technology & Science, Pilani
Work-Integrated Learning Programmes Division
Second Semester 2016-2017
Mid-Semester Test (EC-2 Regular)

Course No. : IS ZC415
Course Title : DATA MINING
Nature of Exam : Closed Book
Weightage : 30%
Duration : 2 Hours
Date of Exam : 25/02/2017 (AN)
No. of pages: 2
No. of questions: 5
Note:

1. Please follow all the *Instructions to Candidates* given on the cover page of the answer book.
2. All parts of a question should be answered consecutively. Each answer should start from a fresh page.
3. Assumptions made if any, should be stated clearly at the beginning of your answer.

Q.1 (a) What is mode of the following data?
10, 2, 30, 14,
50 [1]

Q.1 (b) Eleven students were asked to measure their pulses for 30 seconds and multiply by two to get their one minute pulse rates. The results were: 62, 32, 60, 66, 70, 72, 74,
74, 78, 80, 84. Create five-number summary for the pulse rates and draw boxplot. [3]

Q.1 (c) Students admitted for a certain course have mean score of 560 and a standard deviation of 60. Calculate the z-score of a student having a score of 500. [1]

Q.1 (d) Calculate the cosine similarity between the two phrases below. Feature vector of a word occurring multiple times is greater than 1. Clearly show steps of your calculations.
mid term regular exam
regular exam mid term mid term regular exam [2]

Q.2. You are given 10 training samples. They are divided into four classes: A, B, C, and D.
One sample belongs to A, two belong to B, three belong to C, and four to D. Use the
following $\log_2$ table to answer the questions:

Pages

- Postings Index
- Index of BITS WILP Exam Papers and Content
- Index of Lessons in Technology
- Index of Guest Interviews
- Downloads
- Book Requests

Blog Archive

▼ 2020 (31)
▼ May (1)
Covid-19 and response of IT companies (by Divjot S...
► April (6)
► March (12)
► February (6)
► January (6)
► 2019 (48)
► 2018 (31)
► 2017 (15)
► 2016 (6)

Popular Posts

You Are a Badass. How to stop doubting your greatness and start living an awesome life (Jen Sincero, 2013)
INTRODUCTION The language used in the book extremely funny and Jen Sincero still makes sure that she m...

Covid-19 and response of IT companies (by Divjot Singh)
As the Covid-19 pandemic ravages the world, many domains like airlines, tourism and services...

Innovation to beat the Coronavirus (Covid19)
Coronovirus' Exponential growth and decline In the first phase of the pandemic, we saw a...

Download fiction books (March 2018)
Download fiction books for free: Link for Google Dr...

Life Lessons By Steve Jobs

p	$\log_2(p)$
0.1	-3.32
0.2	-2.32
0.3	-1.74
0.4	-1.32
0.5	-1.0
0.6	-0.74

(a) What is the total information contained in the samples? [2]

(b) What is the total Gini index? [2]

Q.3 (a) Given below is a database of flight delays over a period and under various conditions. We Want to create a decision tree classifier with information gain(entropy) as the attribute splitting criterion.

Feature	Value = Yes	Value = no
Rain	Delayed=30, not Delayed=10	Delayed=10, not Delayed=30
Fog	Delayed=25, not Delayed=15	Delayed=15, not Delayed=25
Summer	Delayed=5, not Delayed=35	Delayed=35, not Delayed=5
Winter	Delayed=20, not Delayed=10	Delayed=20, not Delayed=30
Day	Delayed=20, not Delayed=20	Delayed=20, not Delayed=20
Night	Delayed=15, not Delayed=10	Delayed=25, not Delayed=30

Which feature should be at the root of decision tree? [2]

Q.3 (b) Given the following training documents and their classes:

Document#	Content of document	Class
1	good	Ham
2	very good	Ham
3	bad	Spam
4	very bad	Spam
5	very bad very bad	Spam

Use Naïve Bayes classifier with Laplace (+1) smoothing to find the class of a document with the following contents:

very good bad very very bad [5]

Q.4. Suppose you have the following candidate itemsets of length 4:
{1 2 3 5}, {1 2 4 7}, {1 2 5 6}, {1 3 5 9}, {1 4 5 7}, {1 5 6 9}, {2 3 5 9}, {3 4 5 9},
{4 5 6 8}, {5 6 7 9}

(a) Use hash function k mod 5 to create a hash tree of the itemsets. Assume that each leaf node can store a maximum of three itemsets. [4]

(b) Given transaction {1, 2, 3, 5, 7, 9}, which leaf nodes of the hash tree will be visited for support-counting? Clearly show the visited leaf nodes in the hash tree. [2]

Q.5. Given that min support is 2, and min confidence is 70%, find all association rules from the following market basket dataset using Apriori: [6]

Transaction ID	Items
1	a, b, c
2	b, c, d, e
3	c, d
4	a, b, d
5	a, b, c

****

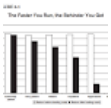
No comments:

Post a Comment

Steve Jobs' last words will change your views on life. The billionaire passed away at the ...

Effects of news and world events on Nifty50 and stock market
Day: 10th Aug 2017 Sensex tanks 267 points. Nifty hits one-month low. 1. Market outlook: ...

Why Bill Gates would raise chickens
I'm excited about the poverty-fighting power of poultry. If you were living on $2 a day, wh...


Intelligent investor (Ben Graham & Jason Zweig, 4e)
Reading from "A Note About Benjamin Graham by Jason Zweig" Here are Graham...

The Essays Of Warren Buffett (Lessons For Corporate America)
INTRODUCTION Buffett has applied the traditional principles as chief executive officer of Berkshi...

How To Talk TO Anyone (92 Little Tricks For Big Success In Relationships, by Leil Lowndes) - Book Summary
There are two kinds of people in this life: Those who walk into a room and say, "Well, here I...

About Me

Ashish Jain

View my complete profile

Enter your comment...

Comment as: Narendran (Go ▾ Sign out

Publish Preview Notify me

Home

Subscribe to: Posts (Atom)

Followers

Followers (0)

Follow

Simple theme. Powered by Blogger.