

survival8

Job opening for freshers

HGS is hiring Healthcare Associates in Bangalore. Immediate joining.

Hinduja Global Solutions

BITS WILP Data Mining Mid-Sem Exam 2016-H2

[Index](#) [Subjects ▼](#) [Mail Us](#)

Birla Institute of Technology & Science, Pilani
Work-Integrated Learning Programmes Division
First 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 : 24/09/2016 (AN)
 No. of page: 2
 No. of questions: 6

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.

Q1. Given the following set of numbers that represents the percentages achieved by ten students.

63, 81, 64, 70, 73, 64, 77, 76, 81, 42

- a) What is the IQR of above data? [2]
- b) Draw boxplot from the above data. Mention any outlier if present. [3]

Q2. Given the following marks scored by a student in two subjects, compute z-scores to find out in which subject the student has done better comparatively.

	Mark obtained by the student	Mean mark of the class	Standard deviation of marks of the class
Subject 1	70	60	15
Subject 2	65	60	6

[3]

Q3. Consider the set of data below:

5, 10, 11, 13, 15, 35, 50, 55, 72, 150, 204, 215.

- a) Partition it into two bins using equal-width partitioning. [2]
- b) Perform smoothing by bin boundary.

Q4. Consider the training examples shown in table below for a binary classification problem

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
 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)
 Coronavirus' 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...

Instance	a_1	a_2	Target Class
1	T	T	+
2	T	T	+
3	T	F	-
4	F	F	+
5	F	T	-
6	F	T	-
7	F	F	-
8	T	F	+
9	F	T	-

- a) What is the entropy of this collection of training examples? [2]
b) What are the information gains(entropy based) of splitting by a_1 and a_2 relative to these training examples? Compute separately. [4]

Q5. Consider the training examples shown in table below for a binary classification problem.

Customer ID	Gender	Car Type	Shirt Size	Class
1	M	Family	Small	C0
2	M	Sports	Medium	C0
3	M	Sports	Medium	C0
4	M	Sports	Large	C0
5	M	Sports	Extra Large	C0
6	M	Sports	Extra Large	C0
7	F	Sports	Small	C0
8	F	Sports	Small	C0
9	F	Sports	Medium	C0
10	F	Luxury	Large	C0
11	M	Family	Large	C1
12	M	Family	Extra Large	C1
13	M	Family	Medium	C1
14	M	Luxury	Extra Large	C1
15	F	Luxury	Small	C1
16	F	Luxury	Small	C1
17	F	Luxury	Medium	C1
18	F	Luxury	Medium	C1
19	F	Luxury	Medium	C1
20	F	Luxury	Large	C1

- a) Compute the Gini index for the overall collection of training examples. [2]
b) Compute the Gini index for the Gender attribute. [2]
c) Compute the Gini index for the Car Type attribute using multiway split. [2]
d) From your answers b. and c., which attribute is better, Gender or Car Type? [1]
e) Explain why Customer ID should not be used as the attribute test condition even though it has the lowest Gini. [1]

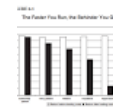
Q6. In the following table, the third column is the predicted probability (posterior) for the positive class in a binary classification problem.

Instance	True Class	$P(+ A, \dots, Z, M_1)$
1	+	0.73
2	+	0.69
3	-	0.44
4	-	0.55
5	+	0.67
6	+	0.47
7	-	0.08
8	-	0.15
9	+	0.45
10	-	0.35

Assume that any test instances whose posterior probability is greater than Threshold=0.5 will be classified as a positive example. Compute the Precision, Recall, and F-measure for the model at this threshold value.

[4]

No comments:



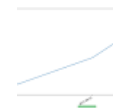
Intelligent investor (Ben Graham & Jason Zweig, 4e)

Reading from "A Note About Benjamin Graham by Jason Zweig" Here

are Graham...

Life Lessons By Steve Jobs

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...



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...

Register & get

₹310

Assured Bonus

+

₹5250

Welcome Bonus

REGISTER NOW

About Me




Ashish Jain

View my complete profile

Post a Comment

Enter your comment...

 Comment as: Narendran (Go ▾)

Sign out

Publish

Preview

☐ Notify me

[Home](#)

Subscribe to: [Posts \(Atom\)](#)


**Register & get
₹310 Assured Bonus
+ ₹5250 Welcome Bonus**


REGISTER NOW

Followers

Followers (0)

Follow

Study science/maths


Study Science & Maths at home with en Aakash faculties.


digital.aakash.ac.in

Top Data Science Course - With Placement Assistance

Great Learning Bangalore - Campus

Learn Data Science in a Classroom. India's best analytics sch... faculty.

 WEBSITE

 CALL

Simple theme. Powered by Blogger.

survival8.blogspot.com/p/blog-page_28.html

3/3