

# survival8

## BITS WILP Data Mining Assignment 2017-H1

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

### Download data files required for assignment

BIRLA INSTITUTE OF TECHNOLOGY & SCIENCE, PILANI

Work Integrated Learning Programme

Data Mining IS ZC415

EC-1 (Lab Component) Weightage 10%

Due date of submission: 22/April/2017

This evaluation component comprises two lab activities. Use either scikit-learn or Weka to complete this assignment. In case you think more information is required to solve any problem, make a suitable assumption and proceed.

1. Please download lab.zip from the course page and extract it. There is a file dataset.csv. You will use it for this problem. First field of this file is binary label 1/0. Remaining 17 fields are features. Apply the following classification algorithms on the dataset.

- ☐ Naïve Bayes
- ☐ Random Forest
- ☐ AdaBoost
- ☐ SVM (linear)

Use 10-fold cross validation and experiment with various parameters of the classifiers to achieve the best performance out of them.

To submit: Precision and Recall of all the classifiers on the given dataset. [6]

2. There are files input\_file\_1 and input\_file\_2 in lab.zip. Run the following Unix/Linux command to create a single file out of these:

```
cat input_file_1 input_file_2 > input_file
```

You need to use input\_file to complete the following problem. The file contains partly pre-processed social media messages in English over a period of last year's leap day (UCT).

The messages in input\_file are roughly sorted in time and there are one or multiple messages every minute except in some cases. In such cases, use linear regression of the past 2 minutes of the missing minute and the next 2 minutes of received messages to fill the missing text of the minute. You need to sort the lines in time order and fill the following text as many times as the number you get from regression:

"missing text here"

For example:

...

Mon Feb 29 08:21:07 leonardo dicaprio deserved the award after many years of awesome movies and no oscar

Mon Feb 29 08:22:45 leonardo dicaprio's performance has won him an oscar.

Mon Feb 29 08:23:56 so, i can't compare myself of not getting girlfriend with leonardo dicaprio not get oscar.

Mon Feb 29 08:24:19 ladies and gentlemen....simply and elegant leonardo dicaprio

Mon Feb 29 08:26:02 the amount of leonardo dicaprio tweets is too high rn

Mon Feb 29 08:26:02 leonardo di caprio wins an oscar. i can die happy now that

### 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)**

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...

**Life Lessons By Steve Jobs**

everyone will shut up about him not winning one!

Mon Feb 29 08:27:06 so happy for leonardo dicaprio. looks like his career is finally taking off.

Mon Feb 29 08:27:11 finally leonardo dicaprio! the crowd were cheering for you.

Mon Feb 29 08:27:27 and you finally received the award, congratulations leonardo!

Mon Feb 29 08:27:52 congratulations to leonardo for finally winning an oscar.

Mon Feb 29 08:28:09 so buzzing for leonardo di caprio!

Mon Feb 29 08:28:30 i liked a video best actor oscars 2016 leonardo dicaprio full speech hd

...

In the above snippet, messages for the minute Mon Feb 29 08:25 are missing, so you will

use regression on minutes from Mon Feb 29 08:23 - Mon Feb 29 08:27. The updated text

would now be:

...

Mon Feb 29 08:21:07 leonardo dicaprio deserved the award after many years of awesome

movies and no oscar

Mon Feb 29 08:22:45 leonardo dicaprio's performance has won him an oscar.

Mon Feb 29 08:23:56 so, i can't compare myself of not getting girlfriend with leonardo dicaprio

not get oscar.

Mon Feb 29 08:24:19 ladies and gentlemen....simply and elegant leonardo dicaprio

Mon Feb 29 08:24:19 ladies and gentlemen....simply and elegant leonardo dicaprio

Mon Feb 29 08:25:00 missing text here

Mon Feb 29 08:25:00 missing text here

Mon Feb 29 08:26:02 leonardo di caprio wins an oscar. i can die happy now that everyone will

shut up about him not winning one!

Mon Feb 29 08:27:06 so happy for leonardo dicaprio. looks like his career is finally taking off.

Mon Feb 29 08:27:11 finally leonardo dicaprio! the crowd were cheering for you.

Mon Feb 29 08:27:27 and you finally received the award, congratulations leonardo!

Mon Feb 29 08:27:52 congratulations to leonardo for finally winning an oscar.

Mon Feb 29 08:28:09 so buzzing for leonardo di caprio!

Mon Feb 29 08:28:30 i liked a video best actor oscars 2016 leonardo dicaprio full speech hd

...

In the input file, there may be more than one continuously missing minutes. You need to fill

them all. For simplicity, as has been done above, filled seconds can be put as 00.


To submit: the sorted file with filled missing values. [4]

You need to upload a zipped file having your ID and name as filename. It should contain all the solutions. The location to upload this file will be communicated later.

No comments:

Post a Comment

Enter your comment...


Comment as: Narendran (Go ▼)

Sign out

Publish

Preview

☐ Notify me

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

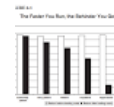
Day: 10th Aug 2017  
Sensex tanks 267 points.  
Nifty hits one-month

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...



Register & get

₹310

Assured Bonus

+

₹5250

Welcome Bonus

REGISTER NOW

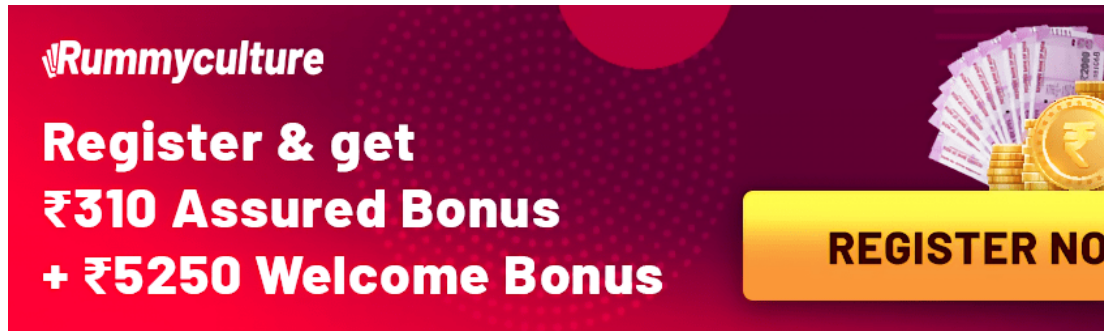
About Me



Ashish Jain

View my complete profile

Home



**Rummyculture**

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

**REGISTER NOW**

The banner features a dark red background with a subtle pattern of small white dots. On the right side, there is an illustration of a fan of Indian Rupee banknotes and several gold coins. The text is in a bold, white, sans-serif font.

Followers

Followers (0)

[Follow](#)

**LAST DAY TODAY**

**NMIMS-NPAT**  
Admission Test for NMIMS Program

**APPLY NOW**

**#ExamFr**

The banner has a dark red background. On the left is the SVKM'S NMIMS Deemed to be UNIVERSITY logo. In the center is a large graphic of an hourglass with the words 'LAST DAY' and 'TODAY' in bold white letters. On the right, the text 'NMIMS-NPAT' is in large white letters, with 'Admission Test for NMIMS Program' in smaller white letters below it. A yellow button with 'APPLY NOW' in red text is at the bottom right. The hashtag '#ExamFr' is in the top right corner.

Simple theme. Powered by Blogger.