



Jiann-Liang Tsay &lt;jlt245@nyu.edu&gt;

---

## more datasets

---

**Raman Kannan** <rk2153@gmail.com>  
To: Jiann-Liang Tsay <jlt245@nyu.edu>

Thu, Oct 27, 2016 at 7:57 PM

sure -- i think u r doing lot of work and learning

am satisfied and please do continue the work into p-2.

also make sure u r crediting ibm cloud computing even though it has not been very useful  
and that you are doing this work as partial requirement for  
M/L course at nyu tandon school of engineering -- fall -- 2016

thanks  
raman

On Thu, Oct 27, 2016 at 4:34 PM, Jiann-Liang Tsay <jlt245@nyu.edu> wrote:

Hi Raman,

I have checked in the project-1 to github as below:

<https://github.com/billtsay/project-1>

The presentation is attached here.

There are two things that I am not doing very well:

- 1 topic modeling that I plan to trace the events for a period of time and see the changes of topics among the tweets. But then I need more data collections for a few days.
2. clustering, I need some skill to come up better interpretation of data.

In fact there are lots of things can improve this project. Just wonder if I can move them into project-2 considering the amount of work in this topic?  
in project-2, I will add sentiment analysis as well.

JLT (jlt245)

On Tue, Oct 25, 2016 at 8:49 PM, Raman Kannan <rk2153@gmail.com> wrote:

ok let me know when you are ready

but there are zillion datasets

On Tue, Oct 25, 2016 at 11:48 PM, Jiann-Liang Tsay <jlt245@nyu.edu> wrote:

I will certainly think about the project of college enrollment. Let me quickly finalize my project-1 and submit it to you within this week. Then we can go further the other possible project.

JLT

On Tue, Oct 25, 2016 at 8:42 PM, Raman Kannan <rk2153@gmail.com> wrote:

I have not seen your work yet but I like the way you work and think.

May be you should look at the image segmentation and segment classification...

identifying natural vs man-made images in the image.

Looks like an interesting problem. I have not done research ...

I have another problem

if you are willing to scrape data I can share some data on the problem of college enrollment.

You have to gather data integrating demographics and economic conditions to come up with a thesis. And as you scrape the data you must agree to provide that data so I can also do the analysis...

Let me know...

R

On Tue, Oct 25, 2016 at 11:37 PM, Jiann-Liang Tsay <jlt245@nyu.edu> wrote:

Yes, I did go there. Then I need to fill in form about my job ... etc.

then I got a message saying some one will contact me for ....

It seems in order to download a "free" dataset, they are seeking some business opportunity from me...

I have not yet particular idea about project-2, although it looks like my project-1 has grown big enough for two... but I am interested in doing another topic: supervised SVM classification on images or some data set...

JLT

On Tue, Oct 25, 2016 at 8:13 PM, Raman Kannan <rk2153@gmail.com> wrote:

not sure if you went here <https://discoverorg.com/get-free-data/?CPN=7011600000sZeM>

JLT, this is not the only site with data... what type of data are you looking for...

On Tue, Oct 25, 2016 at 11:09 PM, Jiann-Liang Tsay <jlt245@nyu.edu> wrote:

I am seeking dataset for my project-2. but the link you provided "[discoverorg.com](https://discoverorg.com)" seems not the dataset that we can download?

JLT

On Fri, Oct 21, 2016 at 6:29 PM, Raman Kannan <rk2153@gmail.com> wrote:

<https://discoverorg.com/discoverorg-sales-intelligence-platform/data-sets-for-marketing-sales-and-staffing/>

for the 9223 stream

Guys, I am researching topicModeling as part of 6923 stream.

I realized topicModeling is fertile field to experiment with the tools and techniques you are learning to speedup processing.

Let me share some links which would drive home this point...

bing Bayesian Inference with Tears: A tutorial workbook for natural language researchers

<http://www.bigdatanews.com/profiles/blogs/topic-modeling-in-r> (this has R code)

<https://cran.r-project.org/web/views/NaturalLanguageProcessing.html>

<http://onepager.togaware.com/TextMiningO.pdf>

<http://www.isi.edu/natural-language/people/bayes-with-tears.pdf>

<http://www.umiacs.umd.edu/~resnik/pubs/LAMP-TR-153.pdf>

Again, I appreciate those of you taking risk to attack significant and challenging problems instead of simple no risk pursuits....

regards

Raman

--

You received this message because you are subscribed to the Google Groups "NYU-CS-6923-fall-2016" group.

To unsubscribe from this group and stop receiving emails from it, send an email to [nyu-cs-6923-fall-2016+unsubscribe@googlegroups.com](mailto:nyu-cs-6923-fall-2016+unsubscribe@googlegroups.com).

To post to this group, send email to [nyu-cs-6923-fall-2016@googlegroups.com](mailto:nyu-cs-6923-fall-2016@googlegroups.com).

To view this discussion on the web visit [https://groups.google.com/d/msgid/nyu-cs-6923-fall-2016/CAC3pmRZQuGpypDP%2BiNXpCR1QF5LyUm0h6\\_xJB5TAq609RbbGJw%40mail.gmail.com](https://groups.google.com/d/msgid/nyu-cs-6923-fall-2016/CAC3pmRZQuGpypDP%2BiNXpCR1QF5LyUm0h6_xJB5TAq609RbbGJw%40mail.gmail.com).

For more options, visit <https://groups.google.com/d/optout>.