

# Case study 2

## FALL SEMESTER 2021

INSTRUCTOR: Sri Krishnamurthy  
s.krishnamurthy@neu.edu

---

**Due: Nov 13th 2021 12.00pm**

You have been asked by your company to create a review system for your products. "It is a serious task" your business users contend. It is our reputation at stake. We can't let spammers and bots mess up our system, your managers agree. But without letting users review the products, the sales will be affected, the sales team contends. Moreover, this is valuable data which will help us build a better recommendation system your data science team says. With all these, your team has been commissioned to prototype a review system. But you will have to develop the process and design and prove that it works. Fake reviews need to be weeded out and a pipeline needs to be developed leveraging machine learning techniques.

How have other companies done that ? Your manager asks. You and your team set out to build a prototype for a review system.

### Tasks:

#### **Part 1: Can you know if a review is fake?**

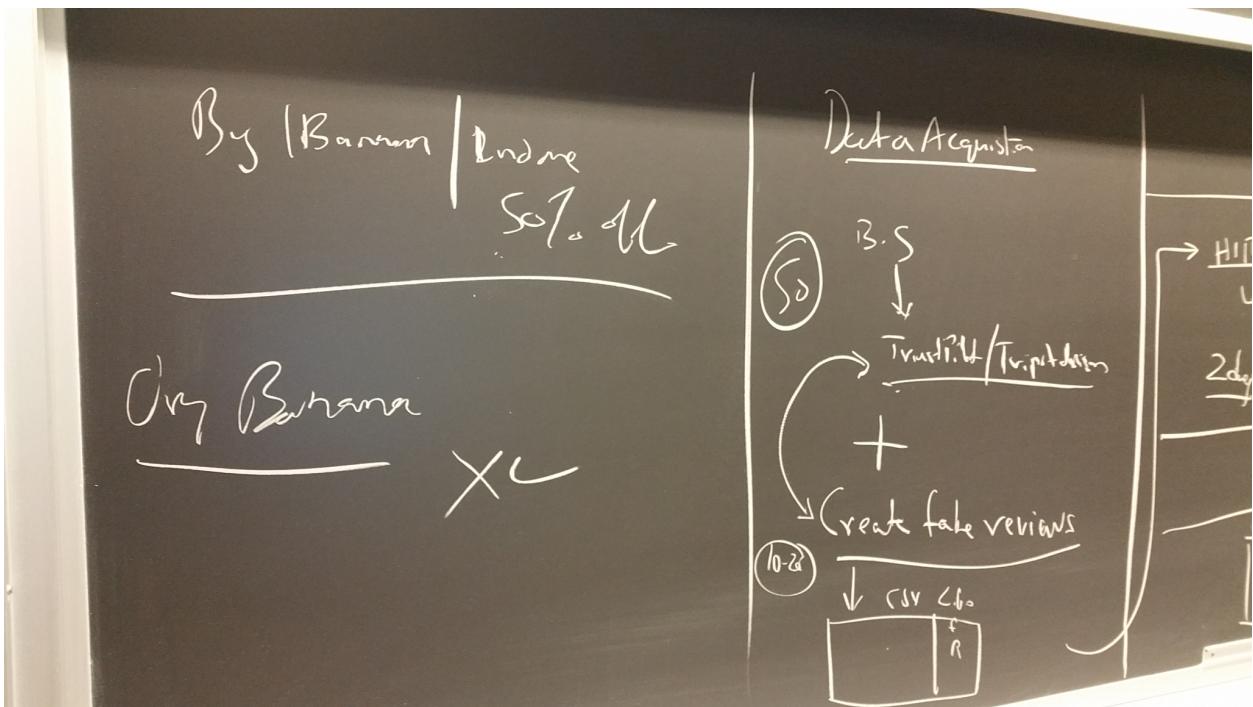
- **Ingestion, fake data creation and pre-processing**

You first try to see if humans can distinguish fake vs real reviews. To get real reviews, you first set out to use Trustpilot or Tripadvisor to scrape reviews.

Use beautifulsoup/scrapy to scape reviews.

Then you create a bunch of fake reviews by hand.

You save all in a standard CSV format.



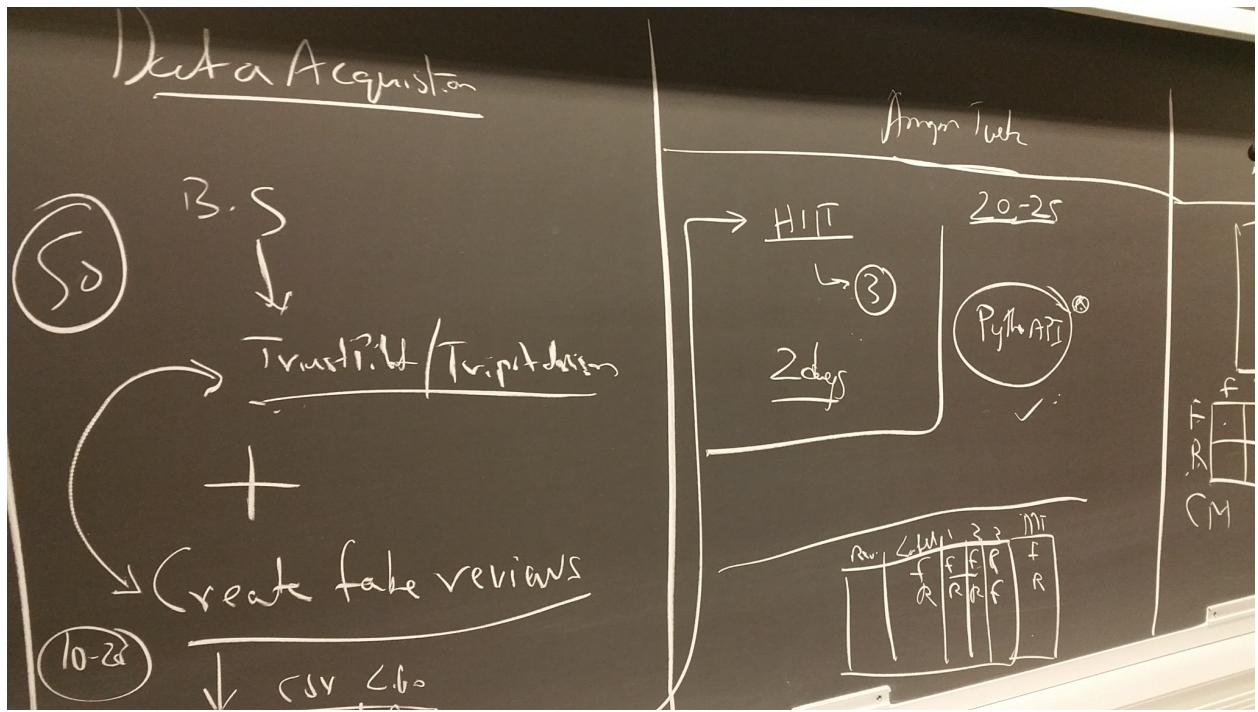
- **Can other users recognize if the data is fake?**

You set up an Amazon mechanical Turk task so that users can vote if a tweet is fake/real

You sample 10-20 reviews and setup 3 assignments at-least for the same review.

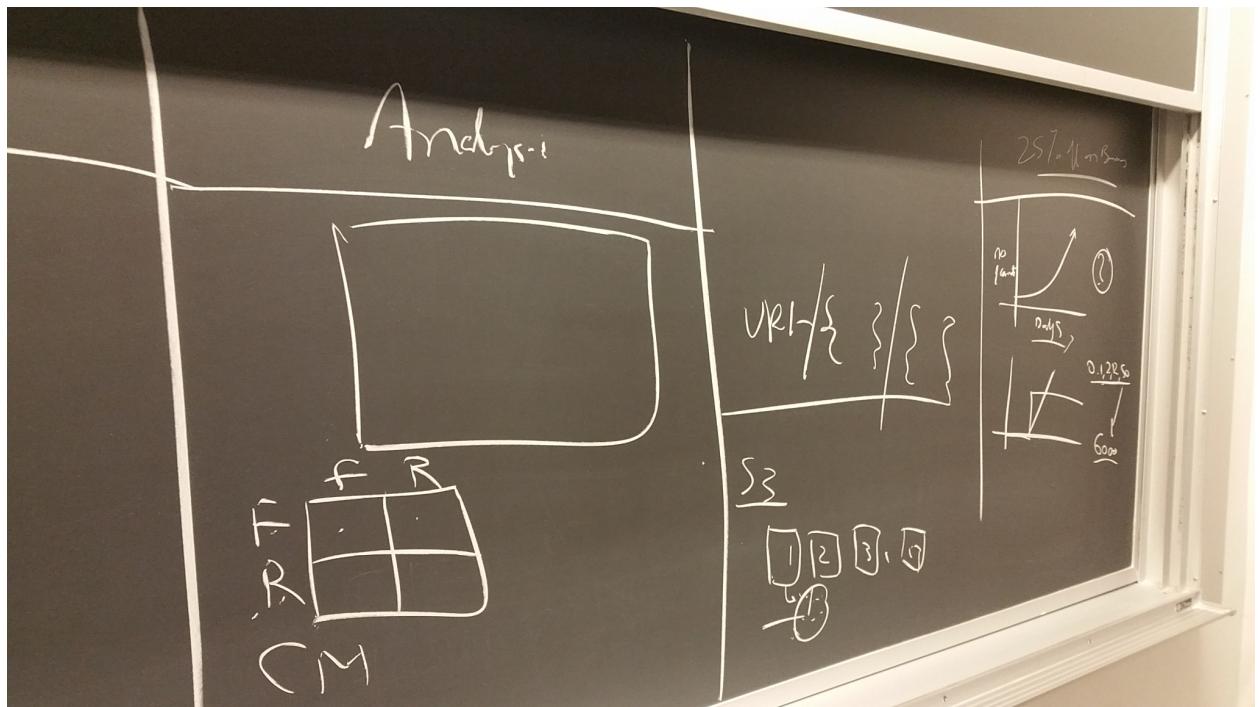
Use majority voting to determine if the review was deemed fake/real.

\*Note: If you do the Amazon mechanical turk part programmatically, you get a bonus 15% to your assignment



- **Summarize your results**

Use tableau or any other BI tools to summarize your analysis



---

## **Deliverables**

- Google Colab notebook with your Python code
- Clatt document discussing findings  
(<https://docs.google.com/document/d/1E6XMcdTexh5O8JwGy42SY3Ehzi8gOfUGiqTiUX6N04o/edit>)

### **Part 2:**

**To be discussed in class on Nov 6th**