

## MAKE-UP ASSIGNMENT FOR ASSIGNMENT I (CS 8803 Social Computing)

<b>Due date</b>	3:30pm, December 5
<b>Topic</b>	A write-up on the design of a data-driven empirical study that examines the relationship between Twitter use and local weather patterns
<b>What to hand in?</b>	3 page report
<b>How/where to submit?</b>	T-Square
<b>Grade</b>	10% (graded on a scale of 20 points)

**Question:** In our class, we have covered a variety of different domains and problems where social media based prediction can shed novel insights that are difficult to obtain otherwise. These include: political elections (Gayo-Avello 2012), stock market (Bollen et al., 2009), box office revenues (Asur and Huberman, 2010), personality traits (Kosinski et al., 2012), urban habits and trends (Naaman et al. 2012), and so on. This assignment assesses your understanding of a different prediction task—**predicting whether and how Twitter activity and sentiment are affected by local weather and changes in weather**. Note that the assignment tests your understanding of the subject, and how you'll design a data-driven empirical study to test hypotheses related to Twitter use and weather. *It does not require you to program or implement the actual prediction task*. Refer to the papers we read in Sep 22<sup>nd</sup> (Week 6) and Nov 10<sup>th</sup> (Week 13) classes for how to frame and present a data-driven and empirical solution to this problem.

Your task is to write a three page report with the following elements:

- 1) [3 points] Present the different hypotheses you need to test for this prediction task. Why are they reasonable hypotheses, or in other words, what is your motivation behind them? If appropriate, situate them in the context of related prior literature (you can refer below for some examples).
- 2) [4 points] How would you collect Twitter data for this problem? What is a good sample size? Which geographic locations would be of interest or appropriate? What is a time frame over which you would test your hypotheses?
- 3) [8 points] How would you frame the prediction task? Precisely discuss the different Twitter variables of interest that are likely to affect activity and sentiment other than weather, how you would measure them, and how you would measure Twitter activity and sentiment. Also discuss how you would go about collecting weather data for the regions of interest.
- 4) [5 points] What kind of prediction model would be suitable for this task? Why is that a good model? How would you evaluate if your prediction model is a good one or not? (For ideas you can consider how you evaluated your classifier in Assignment II).

### References

- [1] Bertrand, K. Z., Bialik, M., Virdee, K., Gros, A., & Bar-Yam, Y. (2013). Sentiment in New York City: A High Resolution Spatial and Temporal View. *arXiv preprint arXiv:1308.5010*.
- [2] Cox, J., & Plale, B. (2011). Improving Automatic Weather Observations with the Public Twitter Stream. *IU School of Informatics and Computing*.
- [3] Dodds, P. S., Harris, K. D., Kloumann, I. M., Bliss, C. A., & Danforth, C. M. (2011). Temporal patterns of happiness and information in a global social network: Hedonometrics and Twitter. *PloS one*, 6(12), e26752.
- [4] Hannak, A., Anderson, E., Barrett, L. F., Lehmann, S., Mislove, A., & Riedewald, M. (2012). Tweetin 'in the Rain: Exploring Societal-Scale Effects of Weather on Mood. In *ICWSM 2012*.
- [5] Kiciman, E. (2012). OMG, I have to tweet that! A Study of Factors that Influence Tweet Rates. In *ICWSM 2012*.