* **Problem:**
* **Sentiment Analysis of** [**TripAdvisor**](http://www.tripadvisor.com/) **Reviews**
* Sentiment analysis is to identify if a piece of review is expressed positively or negatively. For example: “It was a clean room, they offered an awesome service!” is a positive one, and “Staying in this hotel was a horrible experience.” is a negative one. Sentiment analysis helps to find out people's opinion of different products, experience, etc.
* In this competition, you are asked to perform sentiment analysis on reviews from [www.tripadvisor.com](http://www.tripadvisor.com/). It is believed that [TripAdvisor](http://www.tripadvisor.com/) reviews are a viable set of data to working with. The dataset has been gathered and is available at [24kData.txt](http://personal.stevens.edu/~zshao2/timeline/2015/03/24kData.txt) (Please use “right-click” “save as” to save the txt file). You will need to apply your implementation of sentiment analysis on this dataset of sampled reviews.
* The link is a single text file of 24k reviews concatenated together. Each review is as the following format:
* “rating”      “comment”.
* The rating value, taking 1 (negative review) or 5 (positive review), plays as a ground truth in the implementation. For example:
* “1      Staying in this hotel was a horrible experience”.  should be a negative review.
* And your code should take the input “string” as “Staying in this hotel was a horrible experience”, and return “negative”, to give a correct output. And the first value “rating” will be used to evaluate the results of your algorithm.
* **Submission:**
* Two files (at least) for submission:
* A Python script (you can use multiple script files of course) takes in a review and returns “positive” or “negative”. The entrance (main function) should be: def review\_sentiment\_analysis(str\_review):      …     return 'negative'     …     return 'positive' And the (main) script file should be named as “review\_sentiment\_analysis.py”
* A one page .doc file to explain all the steps included in the algorithm you employed, also all the package required for your implementation.
* **Note:**
* As for our evaluation environment, we probably would use both Windows and Mac OS X. Also, we would use [Canopy](https://www.enthought.com/products/canopy/), which is a commercial Python distribution but free to students. Here we recommend it to all the participants.
* Please use popular Python packages in your coding. Each of them should be easily installed with *pip*. If the packages you are importing is not easy to find, you will need to include that package inside your submission.
* Please mind the path for loading the dataset. You need to notice the difference on path between Windows and Unix & Linux. Please use local simple relative path in this competition, like f = open('24kData.txt', 'r'), to avoid the problem.
* **Evaluation:**
* Accuracy
* Novelty of your method

Time efficiency

* **ECE BIG DATA CHALLENGE**
* **Overview:**
* ECE department is organizing competitions designed to spur study and research interests in big-data science and engineering. The competitions are organized following the steps of "Learn-Practice-Compete". It aims to get students familiar with coding in Python and then learning how to turn big-data into valuable insights. The ECE Big Data Challenge will make available real-world big-data sets, with the goal of improving your knowledge based on the materials provided to you during the "Learn-Practice-Compete" process.
* This is to prepare you for a hot job market in big-data!
* **Eligibility:**
* All undergraduate and graduate students in the ECE department.
* **Schedule:**
* **Week 1: Wednesday 3/4/2015**
* **Self Study:** Basics of Python and Learning Materials
* - Intro to Python by Google <https://developers.google.com/edu/python/introduction>
* - Text processing using Python <http://nealcaren.web.unc.edu/an-introduction-to-text-analysis-with-python-part-1/>
* Office Hours (Q&A): 4-6PM, Wednesday 3/4/2015; B-213; Graduate TA/RA: Alireza Louni ([alouni@stevens.edu](mailto:alouni@stevens.edu)), Zongru (Doris) Shao ([zshao2@stevens.edu](mailto:zshao2@stevens.edu)), Yu Zhou ([yzhou7@stevens.edu](mailto:yzhou7@stevens.edu))
* **Week 2: Wednesday 3/11/2015**
* **Self Study:** Data Preparation and Cleaning
* - Sentimental Analysis: <http://nealcaren.web.unc.edu/an-introduction-to-text-analysis-with-python-part-2/>, <http://nealcaren.web.unc.edu/an-introduction-to-text-analysis-with-python-part-3/>
* Office Hours (Q&A): 4-6PM, Wednesday 3/11/2015; B-213; Graduate TA/RA: Alireza Louni ([alouni@stevens.edu](mailto:alouni@stevens.edu)), Zongru (Doris) Shao ([zshao2@stevens.edu](mailto:zshao2@stevens.edu)), Yu Zhou ([yzhou7@stevens.edu](mailto:yzhou7@stevens.edu))
* **Problem Release:** The problem will be published on website on March 14, 2015.
* **Week 3: Wednesday 3/18/2015**
* Spring Break
* **Week 4: Wednesday 3/25/2015**
* Challenge Opens (Details will be announced on Wednesday 3/25/2015)
* **Week 5: Wednesday 4/1/2015**
* Challenge Closes
* **Prizes:**
* **First Place**, ECE Big Data Challenge, Stevens Institute of Technology, April 2015, Certificate plus $200 gift card.
* **Second Place**, ECE Big Data Challenge, Stevens Institute of Technology, April 2015, Certificate plus $150 gift card.
* **Third Place**, ECE Big Data Challenge, Stevens Institute of Technology, April 2015, Certificate plus $100 gift card.
* **Outstanding Performance Awards**, ECE Big Data Challenge, Stevens Institute of Technology, April 2015, Certificate.
* Note:The number of awards will depend on the number of participants and quality of submissions.

For further information, contact Alireza Louni ([alouni@stevens.edu](mailto:alouni@stevens.edu)).