*READ ME*

*Location in cgupta4:* **C:/share/repo/MAHALIA\_MILLER\_FINAL\_SUMMER\_2011/:**

*Date: September 15, 2011*

*Author: Mahalia Miller, mahalia@alum.mit.edu*

*Project: An incident just occurred. What will be its impact?*

**Data in a folder in cgupta4 at C:/share/repo/MAHALIA\_MILLER\_FINAL\_SUMMER\_2011/:**

Weather data: wwp\_accumulated\_rain.txt

Free text CHP reports: accident\_details\_FREE\_TEXT\_janfeb\_2009csv.txt

Database: d4Huge.db. Relevant tables: sensors, devices, incidents, vStarValsMedianWeek. Also, sensorsi5manual and sensorsi605manual are two tables used by the current codes that are subsets of the table sensors.

Sensor metadata: d07\_stations\_2008\_11\_26.txt

**Data in a folder in cgupta4 at C:/share/janFeb2009/:**

Sensor info: files of the form: d07\_text\_station\_YYYY\_MM\_DD.txt.gz

**Work flow:**

1. Calculating v\*. Reads from sensors and devices (only to know which subset to use) tables, writes to a table.
2. Create graph of sensors and store it as a python pickle.
3. Compute impact region and delay. Reads from graph, rain, and tables sensors, incidents, vStar, devices. Writes a text file that is the raw feature vector YYYYMMDD\_featureVector\_...
4. Post processes feature vector. Reads raw feature vector, rain, CHP raw free text. Writes a processed final feature vector as a csv.
5. Weka reads processed final feature vector csv and does machine learning.