**Name:** Abel Jara

**PID:** A10584942

**Display Name:** Pusheen-The-Cat

**Date:** May 21, 2015

* **Task 1:**

In task 1, (the purchase prediction) we were supposed to predict given a (user, item) pair whether the item was purchased by a user. One way I went about this aside from the given find most popular item algorithm, I looked for the most active users by inserting the users into a list and sorting it. Another feature that I used was categorizing the different genders clothing to say males or females. A last algorithm I used which was extremely useful was Pearson's algorithm which compared the user's items and the item's users and rating. I found Pearson's to be the most difficult to understand and implement. It took the most time out of the whole project. In addition to all these major tweaks I had minor tweaks by multiplying/dividing constants to the lists to get better results.

* **Task 2:**

In task 2, (the helpfulness prediction) we were supposed to predict if a user's review of an item is considered helpful. First, I got a list of all the reviews, call it "allHelpful". Then I sorted it in order to get the most helpful instances at the end of the list (so I accessed say the last 50%). This way I could neglect the data with a high "out of" value (since a high "out of" value tends to get buried and unseen). Another thing I did was to make the "average rate" weigh more than the "user rate" by multiplying/dividing it by certain constants. In addition to these weightings, I made separate average ratings pertaining to certain portions of the data in the sorted list (so one would have more helpful data than the others). Another significant feature was when writing to the file I made sure the "out of" values were very low and that no high values were being considered. One challenge I had was implementing linear regression successfully, for some reason it just wasn't working out for me so I dropped it.