

CS753 Assignment 6: Text Mining

Due on December 6, 2018

Notes:

- This is an independent, individual assignment. Group work is prohibited.
- Copying others' work is considered cheating.
- The homework is worth 34 points toward the final grade.
- The homework will be due at the beginning of the class on the due date. No late submission is accepted.
- Submit the printout of your homework. Do NOT submit it via email.
- Make sure you have your name on the printout.

Sentiment Analysis of Customer Reviews at Amazon

The dataset Amazon_Instant_Videos.xls includes 2000 customer reviews of instant videos sold at Amazon during 2013-2014.

- Data preprocessing
 - a. Randomly select 500 records out of this set (using the Sample operator).
 - b. Use the Nominal to Text operator to change the data type of reviewText from nominal to Text.
- Sentiment Analysis
 - a. Use the Analyze Sentiment operator, and select reviewText as the input attribute.
 - b. Copy and paste the pie chart for Polarity.
 - c. Copy and paste the bar chart for Subjectivity.
 - d. Have a closer look at the sentiment analysis results. Do you think the classification result is very accurate? Include a few examples which you think are misclassified. Explain the possible reasons for these misclassification errors and how you think can fix them.