

TOPICS IN DATA MINING

PROJECT REPORT

FEATURE-BASED OPINION MINING

AND RANKING

Report:

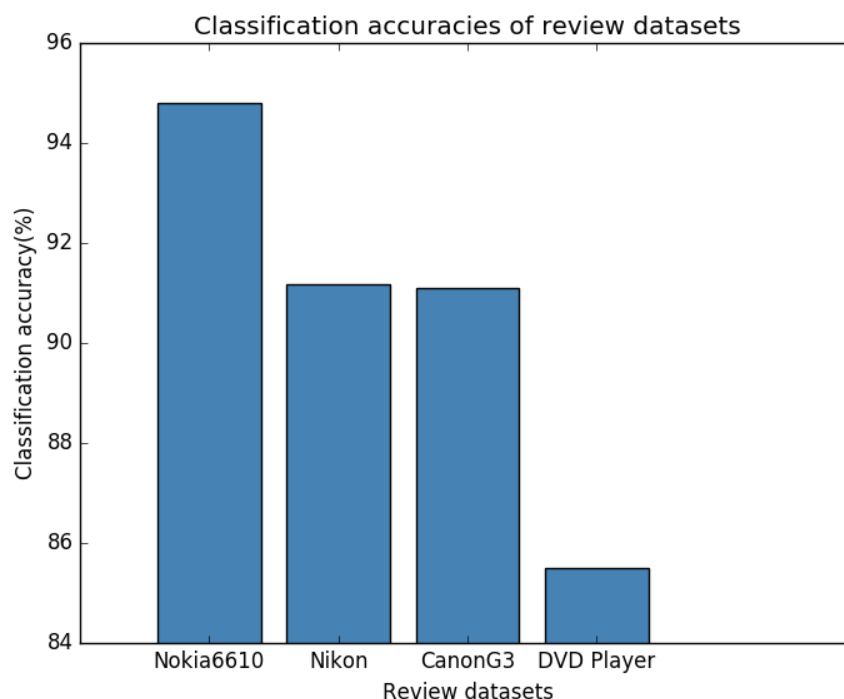
Observation 1:

The table below contains the information about the classification results.

Dataset	Total number of reviews	Actual number of positive reviews	Actual number of negative reviews	Predicted number of positive reviews	Predicted number of negative reviews	Classification Accuracy
Canon G3	45	38	7	42	3	91.11%
Nokia 6610	39	36	3	38	1	94.87%
DVD Player	76	49	27	60	16	85.5%
Nikon Coolpix	34	29	5	32	2	91.17%

Observation 2:

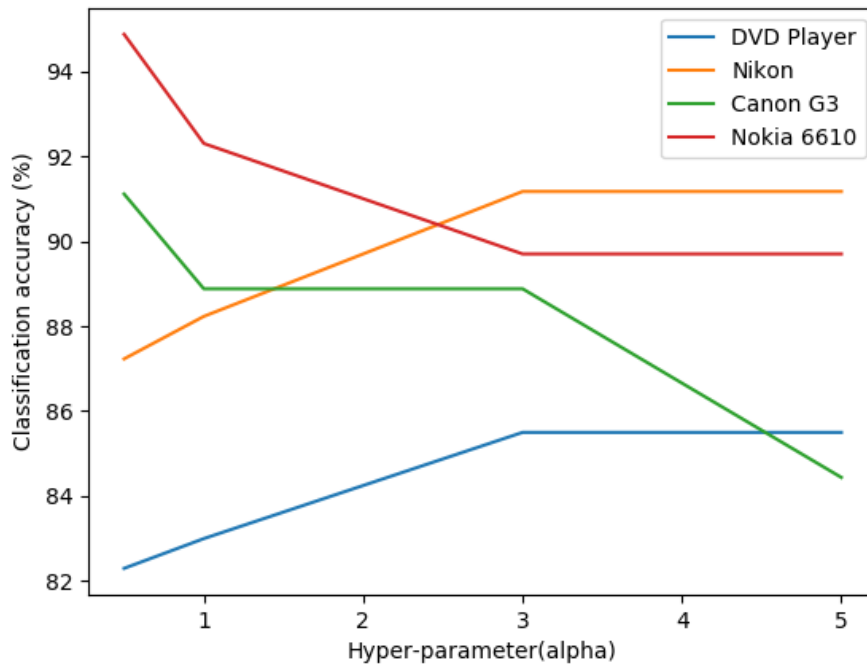
Classification accuracy of the opinion mining algorithm on various datasets:



- The classification accuracies as obtained using the HAC and MOS algorithm for different different are plotted.

Observation 3:

Variation of classification accuracy for various hyper-parameter (α) for various datasets:



- The hyper-parameter is a measure of how much does the title score contribute to the overall review score.
- For the Canon G3 and Nokia 6610 datasets, the titles were not found to contain much opinions in it. So, the classification accuracy is better when the hyper-parameter values is less (typically < 1).
- On the other hand, the titles of the DVD player and Nikon datasets were found to contain many opinions. So, the classification accuracy is better when the hyper-parameter value is greater than 1 (meaning that the title score contributes to a significant extent to the overall review score).

Review dataset	Best hyper-parameter (Estimated)
Canon G3	0.6
Nokia 6610	0.4
DVD Player	3
Nikon Coolpix	3