Predictive Text Mining

School of Information Studies
Syracuse University

Predictive Text Mining
Text Categorization

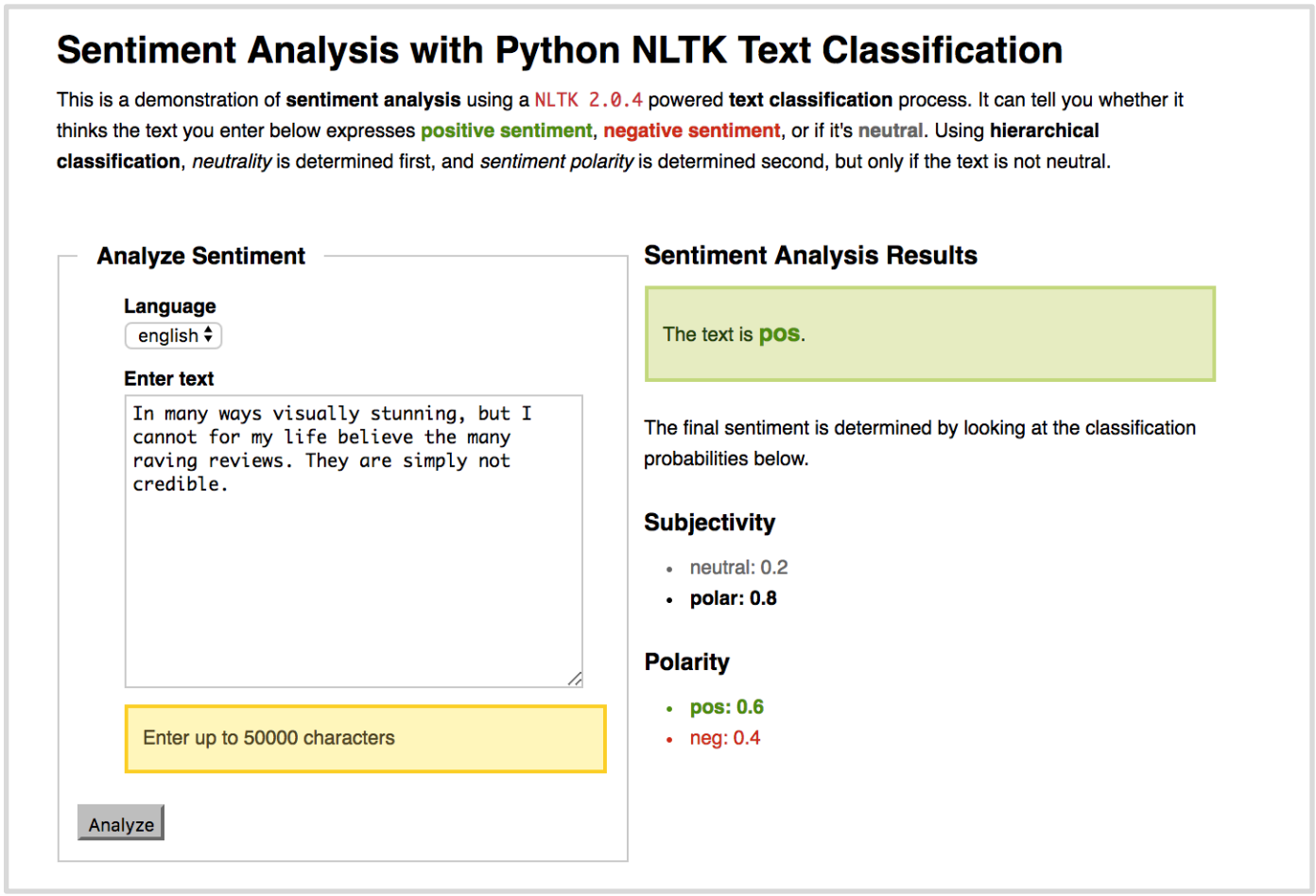
Given pre-defined categories and some training examples, automatically assign documents to categories

Applications

- Sentiment classification
- News topic classification
- Genre classification

Naïve Bayes and SVMs algorithms

Sentiment Analysis



What to Categorize?

Topics

- Categorize news articles to pre-defined topic categories: politics, finance, sports, science

Genres, styles, authors

Or just anything you can define with training examples

- Quality of writing
- Readability
- Ideology

Annotating Training Examples

Example: sentiment annotation

- Positive, negative, neutral
- Or more granularity

Quality of annotations