



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

Sentiment Analysis with Python NLTK Text Classification

This is a demonstration of **sentiment analysis** using a **NLTK 2.0.4** powered **text classification** process. It can tell you whether it thinks the text you enter below expresses **positive sentiment**, **negative sentiment**, or if it's **neutral**. Using **hierarchical classification**, *neutrality* is determined first, and *sentiment polarity* is determined second, but only if the text is not neutral.

Analyze Sentiment

Language

english ↕

Enter text

In many ways visually stunning, but I cannot for my life believe the many raving reviews. They are simply not credible.

Enter up to 50000 characters

Analyze

Sentiment Analysis Results

The text is **pos**.

The final sentiment is determined by looking at the classification probabilities below.

Subjectivity

- neutral: 0.2
- polar: 0.8

Polarity

- pos: 0.6
- neg: 0.4

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

