KDnuggets

search KDnuggets   [Search]

- SOFTWARE
- News/Blog
- Top stories
- Opinions
- Tutorials
- JOBS
- Companies
- Courses
- Datasets
- EDUCATION
- Certificates
- Meetings
- Webinars

PAW Healthcare: Improve Patient Care with Predictive Analytics, Las Vegas, June 3-7

5 Things You Need to Know about Sentiment Analysis and Classification

◀ **Previous post**
Next post ▶

[Like 221] [Share 221]   Tweet [G+]   Share   **49**

Tags: Classification, Sentiment Analysis

We take a look at the important things you need to know about sentiment analysis, including social media, classification, evaluation metrics and how to visualise the results.

comments

By Symeon Symeonidis, Democritus University of Thrace

In the last years, Sentiment Analysis has become a hot-trend topic of scientific and market research in the field of Natural Language Processing (NLP) and Machine Learning. Below, you can find 5 useful things you need to know about Sentiment Analysis that are connected to Social Media, Datasets, Machine Learning, Visualizations, and Evaluation Methods applied by researchers and market experts. Let's get started!

1. Social Media are the main resource

Sentiment Analysis examines the problem of studying texts, like posts and reviews, uploaded by users on microblogging platforms, forums, and electronic businesses, regarding the opinions they have about a product, service, event, person or idea.
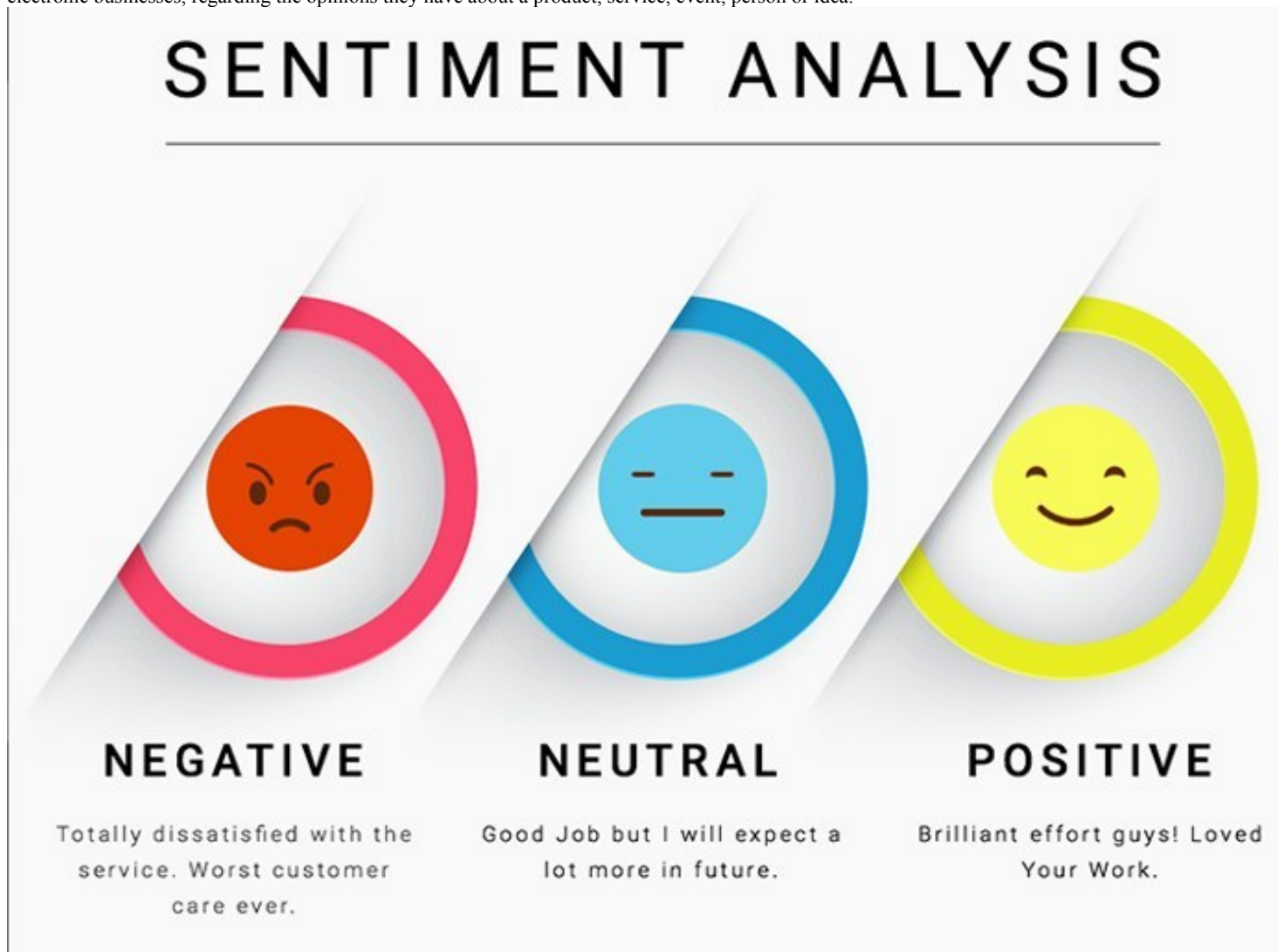
Figure 1. 3-Classes Sentiment Analysis **[1]**

The most common use of Sentiment Analysis is this of classifying a text to a class. Depending on the dataset and the reason, Sentiment Classification can be binary (positive or negative) or multi-class (3 or more classes) problem.

In addition, among researchers and stakeholders, you can find either similar or completely different opinions concerning the relation between emotion detection and sentiment analysis, depending on their perspective. However, regardless the result or approach, they all adopt the same techniques.

2. Before starting the Sentiment Analysis

Datasets

Many evaluations and labeled sentiment datasets have been created, especially for Twitter posts and Amazon product reviews.

The most popular and widespread are:

- Stanford Twitter Sentiment
- Sentiment Strength Twitter Dataset
- Amazon Reviews for Sentiment Analysis
- Large Movie Review Dataset
- Sanders Corpus
- SemEval (Semantic Evaluation) dataset

Also, anyone using the APIs provided by many platforms and forums can crawl and collect data. The most famous API is that of Twitter.

Pre-processing

An initial step in text and sentiment classification is pre-processing. A significant amount of techniques is applied to data in order to reduce the noise of text, reduce dimensionality, and assist in the improvement of classification effectiveness. The most popular techniques include:

- Remove numbers
- Stemming
- Part of speech tagging
- Remove punctuation
- Lowercase
- Remove stopwords

3. How to classify Sentiment?

Machine Learning

This approach, employs a machine-learning technique and diverse features to construct a classifier that can identify text that expresses sentiment. Nowadays, deep-learning methods are popular because they fit on data learning representations.

Lexicon-Based

This method uses a variety of words annotated by polarity score, to decide the general assessment score of a given content. The strongest asset of this technique is that it does not require any training data, while its weakest point is that a large number of words and expressions are not included in sentiment lexicons.

Hybrid

The combination of machine learning and lexicon-based approaches to address Sentiment Analysis is called Hybrid. Though not commonly used, this method usually produces more promising results than the approaches mentioned above.
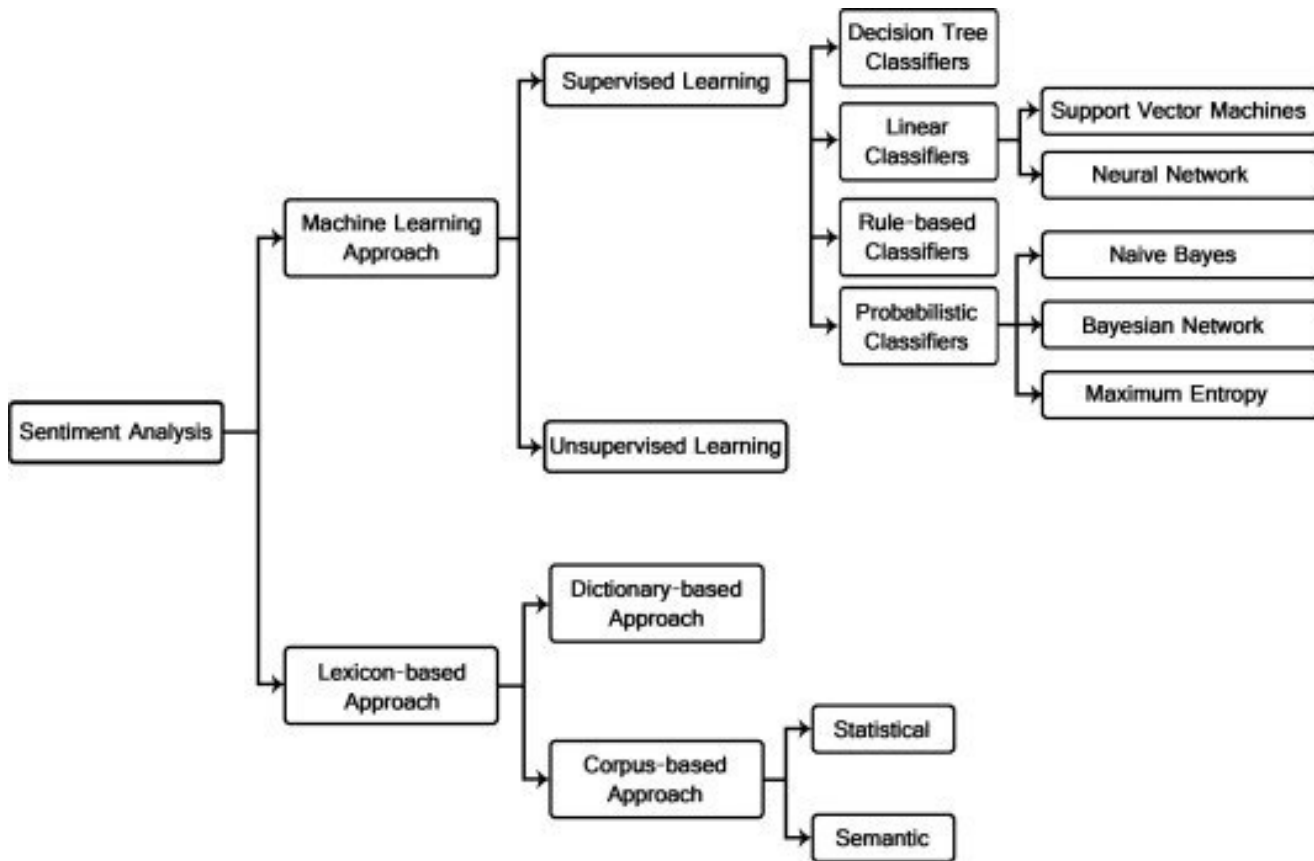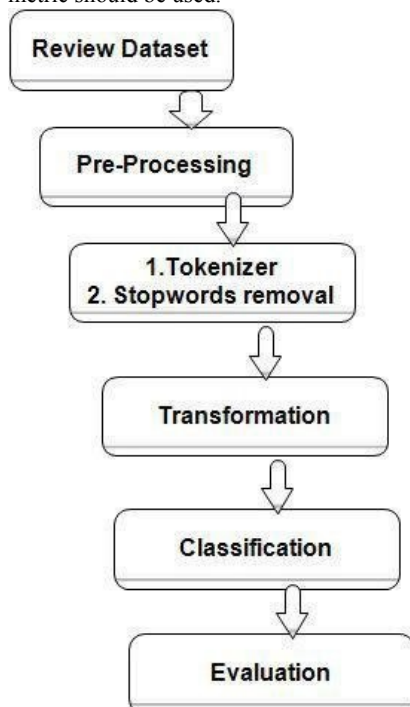
Figure 2. Sentiment classification techniques [2]

4. Evaluation metrics

As a classification problem, Sentiment Analysis uses the evaluation metrics of Precision, Recall, F-score, and Accuracy. Also, average measures like macro, micro, and weighted F1-scores are useful for multi-class problems. Depending on the balance of classes of the dataset the most appropriate metric should be used.

Figure 3. Steps-to-Evaluate-Sentiment-Analysis **[3]**

5. Visualise Results

To visualize the results of Sentiment Analysis, many people employ well-known techniques, such as graphs, histograms, and confusion matrices. Because of present multiple data domains and tasks, visualizations approaches like wordcloud, interactive maps, sparkline-style plots are also very popular.

Figure 4. Sentiment Word Cloud **[4]**

To dive deeper into the fascinating world of Sentiment Analysis, we recommend you to follow some posts from KDnuggets:

- SlangSD: A Sentiment Dictionary for Slang Words
- Mining Twitter Data with Python Part 6: Sentiment Analysis Basics
- Political Data Science: Analyzing Trump, Clinton, and Sanders Tweets and Sentiment
- Sentiment Analysis & Predictive Analytics for trading, Avoid this systematic mistake
- Tutorial: Building a Twitter Sentiment Analysis Process

[1] https://www.newgenapps.com/blog/the-secret-way-of-measuring-customer-emotions-social-media-sentiment-analysis

[2] https://www.sciencedirect.com/science/article/pii/S2090447914000550

[3] https://www.researchgate.net/figure/Steps-to-Evaluate-Sentiment-Analysis_fig1_321479545

[4] https://www.displayr.com/sentiment-analysis-simple

Bio: Symeon Symeonidis is a PhD candidate in the area of intention and sentiment mining, at Democritus University of Thrace.

Related:

- Hierarchical Classification – a useful approach for predicting thousands of possible categories
- The Machine Learning Abstracts: Classification
- AI and Sentiment Analysis to help you move ahead of the competition

2 Comments   **KDnuggets**   ● **Phuc Coi**   ▾

♡ **Recommend** 1   ↰ **Share**   Sort by Best ▾

┌───┐
│ **Join the discussion…**   │
└───┘

Phuc Coi • 3 days ago

how to do sentiment analysis on facebook?

⌃ | ⌄ • Edit • Reply • Share ›

Dimitrios Pesios • a month ago

What about exploratory data analysis as a first step?

⌃ | ⌄ • Reply • Share ›

✉ **Subscribe**   ⓓ **Add Disqus to your siteAdd DisqusAdd**   🔒 **Privacy**

◄ **Previous post**
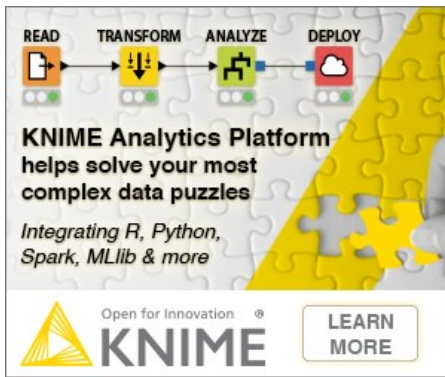**Next post** ►

Top Stories Past 30 Days

Most Popular

1. **Why so many data scientists are leaving their jobs**
2. **Key Algorithms and Statistical Models for Aspiring Data Scientists**
3. **Blockchain Explained in 7 Python Functions**
4. **Data Science vs Machine Learning vs Data Analytics vs Business Analytics**
5. **Python Regular Expressions Cheat Sheet**
6. **8 Useful Advices for Aspiring Data Scientists**
7. **7 Books to Grasp Mathematical Foundations of Data Science and Machine Learning**

Most Shared

1. **7 Books to Grasp Mathematical Foundations of Data Science and Machine Learning**
2. **Python Regular Expressions Cheat Sheet**
3. **Blockchain Explained in 7 Python Functions**
4. **2018 KDnuggets Poll: What software you used for Analytics, Data Mining, Data Science, Machine Learning projects in the past 12 months?**
5. **Choosing the Right Metric for Evaluating Machine Learning Models – Part 1**
6. **Boost your data science skills. Learn linear algebra.**
7. **Key Algorithms and Statistical Models for Aspiring Data Scientists**

Latest News

- DSTI: Applied MSc in Data Engineering, Advanced MSc in ...
- PAW for Industry 4.0 – Take a Look at the Agenda
- Worcester Polytechnic Institute: Post-Doctoral Fellow
- Data Engineer vs Data Scientist: the evolution of aggre...
- Top Stories, May 7-13: 2018 KDnuggets Analytics, Data M...
- Simple Derivatives with PyTorch

**Top Stories
Last Week**

Most Popular

1. `NEW` **2018 KDnuggets Poll: What software you used for Analytics, Data Mining, Data Science, Machine Learning projects in the past 12 months?**
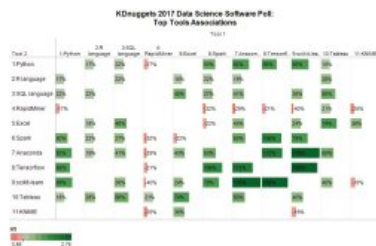
2. `NEW` **WTF is a Tensor?!?**
3. `NEW` **Data Science vs Machine Learning vs Data Analytics vs Business Analytics**
4. **Why so many data scientists are leaving their jobs**
5. `NEW` **Boost your data science skills. Learn linear algebra.**
6. `NEW` **8 Useful Advices for Aspiring Data Scientists**
7. `NEW` **Key Algorithms and Statistical Models for Aspiring Data Scientists**

Most Shared

1. **2018 KDnuggets Poll: What software you used for Analytics, Data Mining, Data Science, Machine Learning projects in the past 12 months?**
2. **WTF is a Tensor?!?**

3. **5 Reasons "Logistic Regression" should be the first thing you learn when becoming a Data Scientist**
4. **PyTorch Tensor Basics**
5. **Top 7 Data Science Use Cases in Finance**
6. **Detecting Breast Cancer with Deep Learning**
7. **To SQL or not To SQL: that is the question!**

KDnuggets Home » News » 2018 » Mar » Opinions, Interviews » 5 Things You Need to Know about Sentiment Analysis and Classification (18:n13)

© 2018 KDnuggets. About KDnuggets. Privacy policy.

Subscribe: Email [Your email]   Name [your name]   [Subscribe]

X