Analysis of Terms & Conditions

by Tomaso Leoni

About Me

› Tomaso Leoni

› IMS Intern at the IDAS

› NLP Projects & Web applications

The Background

Idea brought forward by Mascha Kurpicz-Briki

Part of IDPA (*interdisciplinary project, 3rd year school project IMS*)

Focus on usage of NLP to analyse terms & conditions

Before we start – what is NLP?

Describes technical ways of analysing natural language (articles, voice, books, reviews)

Used for:

Document scanning

Voice recognition

Voice assistants

Interface between humans and machines

How does NLP work?

› 2 main approaches
   › Rule based
   › Algorithm based
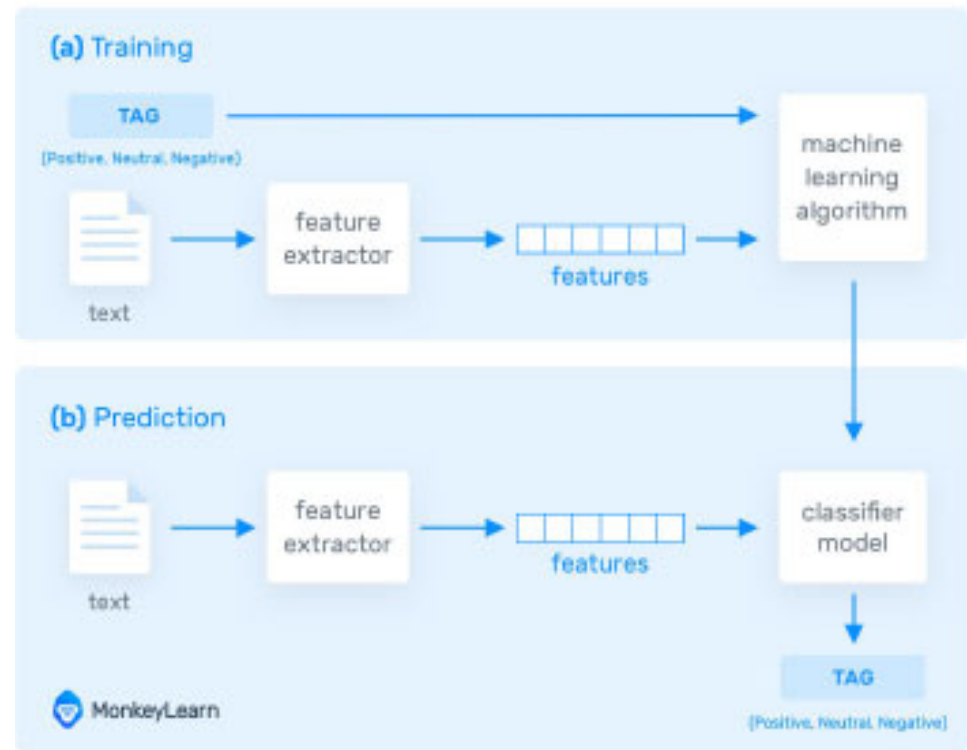
Rule based

› Done by assigning human selected rules to a text
  › Check grammar
  › Filter questions
  › Recognize different, fixed patterns

Algorithm based

› A model is trained on existing data (newspaper articles, reviews for example)

› Guesses based on that training

What's the goal?

› Distinguish especially negative words from the rest of the text

› Drive attention towards those items

› Quickly analyse long, generic items of text

Demo

Technologies used

› spaCy & nltk

  › Natural Language Processing Frameworks

› SentiWS & SentiWordNet

  › Sentimental Corpora for the German and English language respectively

› React

  › JavaScript Framework, used in the front end

› Flask

  › Python Framework used for the back end

spaCy & nltk

- › Tool widely used for natural language processing (NLP)
- › Utilities used in this project:
  - › **Sentencizer** – allows user to split text into sentences
  - › **German language model** – allows analysis of German texts
    - › Grammar, structure, spelling, etc.

SentiWS & SentiWordNet

› **SentiWordNet** – a lexical resource for opinion mining
  › Words are assigned a negative, positive and objective value

› **SentiWS** – a German corpora for sentimental analysis
  › Same principle as SentiWordNet, applied to the German language

React & Flask

› **React** – a JavaScript based Framework for Web interfaces

  › Used to design the landing & results page, user interaction

› **Flask** – a lightweight framework for web applications

  › Allows easy interaction between the analysis and the user interaction (both based on python)

Analysis of this project – how it works

1. The text is split into sentences

2. All unnecessary words (prepositions, pronouns) are removed

3. The word's negative score (if available) is evaluated

4. If the word has a particularly negative score, it gets highlighted

Thank you for your interest

› Feedback / Questions
  › tomaso.leoni@bfh.ch