NLP - Natural Language Processing

About me

Rafal Pronko

data scientist / machine learning enthusiast

rafalpronko

Access code: **nlp101**

Agenda

Presentation part:

- Business problem
- Data sources
- Tools

Workshops:

- NLP - introduction
- Spark - introduction
- Basic NLP@Spark
- Basic ML@Spark

Business problem

Data sources

- R. He, J. McAuley. Modeling the visual evolution of fashion trends with one-class collaborative filtering. WWW, 2016
- J. McAuley, C. Targett, J. Shi, A. van den Hengel. Image-based recommendations on styles and substitutes. SIGIR, 2015

Tools url.getindata.com/text-mining

NLP - short introduction

Natural-language processing (**NLP**) is a field of computer science, artificial intelligence concerned with the interactions between computers and human (natural) languages, and, in particular, concerned with programming computers to fruitfully process large natural language data.

Challenges in natural-language processing frequently involve speech recognition, natural-language understanding, and natural-language generation.

NLP - short introduction - key application areas

Automatic summarizer

Sentiment analysis

Text classification

Information extraction

Notebooks introduction

Block of texts

Code to run (import pandas as pd)

Short exercises to do (# TODO - show how many items left)

Biggest exercises to do

Bonuses