

NLP - Natural Language Processing

About me



Centrum

ZASTOSOWAŃ MATEMATYKI I INŻYNIERII SYSTEMÓW

"We solve problems"

Rafal Pronko

data scientist / machine learning enthusiast



rafalpronko

<http://piazza.com/nlprafalpronko/spring2018/nlp101>

Access code: **nlp101**

Agenda

Presentation part:

- Business problem
- Data sources
- Tools

Workshops:

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

Business problem



Kitchen & Dining

East

Shop by Department

Your Amazon.com Today's Deals Gift Cards Sell Help

Hello, Sign in Your Account Try Prime

Kitchen & Dining

Best Sellers Wedding Registry Small Appliances Kitchen Tools Cookware Bakeware Cutlery Dining & Entertaining Storage

Home & Kitchen > Kitchen & Dining > Small Appliances > Ovens & Toasters > Toasters





Proctor Silex 22611 2-Slice Toaster

by Proctor Silex

★★★★☆ 346 customer reviews | 14 answered questions

Price: **\$12.99** & **FREE Shipping** on orders over \$49. [Details](#)

In Stock.

Want it tomorrow, March 17? Order within **4 hrs 30 mins** and choose **One-Day Shipping** at checkout. [Details](#)

Ships from and sold by Amazon.com. Gift-wrap available.

- Toast boost lifts slices higher
- Auto shutoff for safety
- Shade selector
- Slide-out crumb tray
- 2 Slice capacity

41 new from **\$11.63** 2 used from **\$11.04**

☒ Buy new

☐ Q

☐ Yes, I want **Shipping**



[Turn on 1-Click](#)

Ship to:

Select a ship

☐ Buy used

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



pandas
 $y_{it} = \beta' x_{it} + \mu_i + \epsilon_{it}$



Natural Language
Analyses with NLTK

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