

MARTIN POLACEK

martin.polacek@stonybrook.edu
408-368-4250
New York, NY

github.com/KvitnucaZahradka
linkedin.com/in/martin-polacek1024

EXPERIENCE

Fellow at Insight Data Science (2017)

- ◇ **Fake News Checker:** A predictive model which assesses whether a new article is real or fake.
 - Collected, cleaned and stored (PostgreSQL) data from several sources, including New York Times, Wall Street Journal and Kaggle. Tools primarily used: Beautiful Soup, New York Times API, Kaggle data dump and URL metadata obtained by querying whois server.
 - Processed text using TextBlob, Spacy and NLTK.
 - Trained AdaBoosted and Random Forest classifiers using Python libraries: Pandas, NumPy and scikit-learn.
 - Developed front-end app using Flask and D3.js. Employed as instance on Amazon Web Services.
website: crownlies.com

Data Science Consultant for medical startup REVON (2016)

- ◇ **Consulting, pulmonary disease prediction:** Built a predictive model which assesses whether patient should visit doctor based on current health state.
 - Analyzed data set containing predictions (made by several doctors) for chronic obstructive pulmonary disease patients.
 - Trained AdaBoosted decision tree classifier to estimate severeness of patient's state.

Data Science Projects (2014 - present)

- ◇ **Early Trendsetters Detector:** Built an application predicting early trendsetters among papers posted on [arXiv.org](https://arxiv.org).
 - Estimated innovation and trendsetting-ness of given new paper. The estimation was assessed by examining lexical and contextual (dis-)similarities to known trendsetters.
 - Trained Random Forest classifier to predict whether new paper will be a trendsetter.
website: github.com/KvitnucaZahradka/lazy_ions
- ◇ **Yelp Dataset Challenge Part 6:** Built analysis that identifies and predicts bad business practices and their future extent.
 - Cleaned and analyzed text in Yelp - provided data set (users reviews and metadata). Used R libraries: wordcloud, coreNLP, openNLP, syuzhet.
 - Developed algorithm to assess factual information about bad business practices and provide quantitative measure of their extent.
 - Trained time series prediction algorithm: nnetar for neural network based time prediction, Arima - custom Arima model for time series prediction.
website: rpubs.com/kvitnuca_zahradka/139697
- ◇ **Random Brain:** Built web application in which player produces as random sequence as possible.
 - Application was built using R Shiny. The algorithm estimates the randomness (Kolmogorov complexity) of the human brain produced sequence.

EDUCATION and RESEARCH (2010 - 2017)

- ◇ **Stony Brook University:** PhD in Physics (2011 - expected May 2017).
- ◇ **Comenius University:** MSc & BSc in Physics (2005 - 2010).
- ◇ **Research:** Numerical analysis (C++) of probabilities in particle physics models beyond the Standard Model. Developed analytical techniques to compute probabilities in particle physics using recently developed correspondence (AdS/CFT).

COMPUTER SKILLS

- ◇ Java, C++, Python, R, shell script, SQL (PostgreSQL, LiteSQL), Matlab, Octave, Mathematica.
- ◇ Ruby, Ruby on Rails, Jekyll, HTML, CSS, R Shiny, Flask.

OTHER EXPERIENCE

- ◇ Teaching Assistant (Stony Brook University, Comenius University): Mathematical Analysis, Statistical Physics, Mechanics.