

Anna Orosz

<http://anna.orosz.pro> | +36 (30) 588-4410 | aorosz@sas.upenn.edu | [linkedin.com/in/anorosz](https://www.linkedin.com/in/anorosz)

Education

University of Pennsylvania | **Master of Science in Engineering in Data Science**

December 2021

Concentration: Machine Learning | Master Thesis: Computational Linguistics

University of Pennsylvania | **Bachelor of Arts in Mathematics and Computer Science**

May 2021

Relevant Coursework: Artificial Intelligence, Machine Learning, Computational Linguistics, Databases, Statistics for Data Science, Algorithms & Data Structures, Unix, Python, Computer Architecture, Graph Theory & Algorithms, Analysis, Abstract & Linear Algebra, Game Theory, Calculus, Language and Automata, Micro- and Macroeconomics

Professional Experience

MBition – Mercedes-Benz Innovation Lab | **Artificial Intelligence Intern** | Berlin, Germany

June 2020 – August 2020

- Built time series model to detect & classify data outages among millions of Mercedes-Benz data across Europe
- Evaluated state-of-the-art approaches (SARIMAX, LSTM, PROPHET) in Azure Databricks environment with *PySpark*
- Innovated on frontline of automotive world to develop next generation of self-driving Mercedes-Benz automobiles

LogMeIn | **Machine Learning Intern** | Budapest, Hungary

January 2018 – July 2019

- Developed award-winning NLP Hackathon project for training QA systems using state-of-the-art Deep Neural Networks
 - Won 1st prize Tech Innovation + Audience favorite prize + 2nd place Best Business Value Innovation prizes
- Researched term weighting methods, class hierarchy models facilitating email automation for Top 5 Indian Bank
- Enriched Q&A text corpus in languages (VN, GE, IN, etc.) by building sophisticated web-scraping services for Bold360 AI
- Specialized in Applied Research for **Natural Language Processing**, worked in *Python* with *Keras*, *Tensorflow*, *scikit-learn*

Facebook | **Software Engineering Intern** | Menlo Park, California

May 2017 – August 2017

- Modernized internal client-library tool *Hyperloop* completing bulk data transfers for Facebook's Data Scientists
- Overhauled internal tool *COPTA* to copy directories across *Hadoop Distributed File System* clusters and FB's data centers
- Engineered detailed *Scuba* tables for *Hyperloop* and *COPTA* as (100+ petabytes data)
- Operated with multi-tenancy, network utilization, scheduling, cross-dc connection-pooling in *C++*

RapidMiner | **Software Developer Intern** | Budapest, Hungary

May 2016 – August 2016

- Used **Data Science** to program operators in Studio designed to process, analyze and alter data locally
- Designed Operators in *Radoop* - in *Spark* scripts - to process data on large scale with Apache Hadoop
- Repaired bugs and designed new features in *Java* while maintaining a stable infrastructure

Projects

Time series model

June 2020 – August 2020

Create seasonal and trend-sensitive time series model assessing Mercedes Benz system failures.

[Python]

Standardized Testing AI

March 2020 – May 2020

Built DistillROBERTa model for answering standardized testing Science questions in multiple-choice style.

[Python, TensorFlow]

Chatbot

November 2018 – March 2019

Established Question Answering system using BERT to serve Bold360 AI's clients. (3x Hackathon winner)

[Python, TensorFlow]

Youtube Spam Comments Detector

November 2017

Classified comments as ham or spam with Naïve Bayes and SVM models by using TF-IDF transformation.

[Python, scikit-Learn]

County Election Predictor

September 2017

Developed and compared NN & SVM models predicting 2016 election results by county. [Python, Keras, scikit-Learn, TensorFlow]

Image Classification

May 2017

Fine-tuned NN model trained on the ImageNet dataset to classify images of UPenn/non-UPenn logo.

[Python, TensorFlow]

Leadership Experience

CIS 530 - NLP class:

Teacher's Assistant in graduate-level Natural Language Processing class at Penn in Fall 2020

The Daily Pennsylvanian: *Machine Learning Engineer* at the UPenn school newspaper's Analytics department

Proficiencies and Passions

Prog. Languages: Java, Python, SQL, C++, C, OCaml, Bash/Shell, HTML, Assembly Language

Platforms/Modules: scikit-learn, Keras, TensorFlow, PySpark, Linux/Unix, Git/Mercurial, JupyterHub/Lab, Azure Databricks

Languages: German (Fluent), Hungarian (Native Speaker), French (Intermediate)

Interests: Natural Language Processing, Autonomous Vehicles, Machine Learning, AI & Data Science, Cloud services

Activities & Hobbies: riding my Piaggio, horseback-riding, bouldering, salsa-dancing, ice-skating & skiing, cinematography