

Szabo Roland Teodor

CONTACT

Iosif Silviu Fogarasi 33
Oradea, 410283
Bihor, Romania

E-mail: consulting@rolisz.ro

Website: <https://rolisz.com>

Phone: +40741767719

WORK EXPERIENCE

Gradient Ascent SRL

December 2020 — Present

Machine Learning Consultant

I am an independent machine learning consultant, trainer and advisor looking to help companies do machine learning fast, securely and in a privacy aware manner.

I can help companies that are interested in ML projects by advising them what projects are feasible and worth pursuing.

I can offer consultancy in what approaches work for different machine learning problems, how to formulate questions and figuring out if there is enough signal in the data to extract useful information.

I offer machine learning trainings for software engineers.

Some of the projects I've worked on:

- A project to extract structured data (domain, required experience, required skills) from job ads in multiple languages. I trained Transformer models to classify ads, perform NER from ads. I used multilingual models to generate embeddings to cluster the extracted skills in order to be able to perform matches across different languages.
- A project to index and categorize articles written by thought leaders in the ESG domain based on a given taxonomy. The articles were gathered by scraping from Twitter. One of the challenges was identifying authors even when different variations of their name were used.
- A project to extract information from financial documents. The documents were in different languages, from different countries and used different names for the same concepts. I used Document Understanding Models (Donut) to do key/value pair extraction and question answering on the documents. The extracted values were normalized using a combination of dictionaries and embeddings.
- A chatbot that can answer questions employees have about the different policies the company has. The project was done on the company's hardware, using open source LLMs, so that the data didn't leave the company infrastructure.

Archive 360

July 2019 — November 2020

Tech Lead

I was the tech lead for a team of 5 people, working on a project to extract and archive data from cloud based SaaS (Office 365 and related products). We extracted the data with high fidelity, so that customers can then save money by deactivating the unneeded licenses, without losing any data. The binary data is stored in cloud blob storage, while metadata about the files (the users to whom it belongs or was shared with, relations to other files) are stored in CosmosDB (graph database). Our tech stack was based on Python, Spark, Kafka and Kubernetes.

Laif Science

March 2018 — July 2019

Software Engineer

I lead a team of 5 to build a real time surveillance camera processing system to recognize faces, identify intruders and detect dangerous objects such as guns and knives. The ML models were built using Tensorflow, taking pretrained models and doing fine tuning using our own data. The models had to be tuned in order to achieve the necessary low latency to

enable doing this in real time.

I designed the architecture and lead a team that builds a GUI system to enable non technical people to create ML pipelines without having to write code. This pipeline enables the creation of scikit-learn and Tensorflow pipelines by dragging and connecting blocks that do preprocessing and apply ML algorithms. Many best practices are built in and the UI helps choose the best model.

I worked on a data analysis project to understand, forecast and optimize the logistics and supply chain for an oil and gas company. The project required analyzing the historical data of orders and transportation routes, making predictions for future demand and identifying delivery routers that were optimal from a cost perspective.

Google July 2016 — January 2018
Software Engineer

I was part of the Google Calendar team. I was building a recomender system for rooms, which is used in companies that use Google Calendar to manage their meetings.

I developed pipelines to process historical event data, join it with other data sources and then extract useful signals from them. I trained machine learning models on those signals and I ran experiments to see which ones performed better in production.

Google September 2014 — June 2016
Site Reliability Engineer

I was part of the team that was responsible for the storage layer of GMail and Google Drive. I was working on automation workflows for deployments, capacity planning and anomaly detection on monitoring data.

EDUCATION Computer Science October 2011 — June 2014
Babes-Bolyai University, Cluj-Napoca

PROGRAMMING
LANGUAGES

- Python - advanced
- Java - advanced
- Rust - medium
- Javascript - medium

TOOLS

- Machine learning libraries: Tensorflow, scikit-learn, Keras, FastText, HuggingFace, Spacy
- Distributed processing: Apache Spark, Apache Beam, Kafka
- Cloud vendors: Google Cloud Platform, Azure, AWS
- Web apps: Flask, FastAPI
- NoSQL: CosmosDB, Gremlin
- Containers and Orchestration: Docker, Kubernetes
- Data collection: Scrapy, Doccano, Prodigy

ADDITIONAL
INFORMATION

Spoken Languages:

- Native proficiency in Romanian and Hungarian
- Professional in English