Lucian Istrati

Curriculum Vitae / Resume

Cologne, North Rhine-Westphalia
50823
Germany
✆ *+49 151 2001 6748*
✉ *lucianistrati@yahoo.com*

Work experience

- **Head of Data Science and AI** at **apadua** (March 2023 - Present) - tech procurement start-up where I led an AI and Data Science team that:
  - finetuned and integrated **LLMs** in a generative AI text functionality;
  - created several **information extraction** pipelines that drastically reduced the number of missing values for some of the fields in our database;
  - developed a mechanism that calibrates the number of retrieved results on a **search functionality** by reducing them when there were too many and enriching them when there are too few, increasing clients satisfaction with the search product.

- **CEO, Co-Founder and Machine Learning / Natural Language Processing Engineer** at **Verbosity** (May 2022 - February 2023) - **eLearning** platform customed for learning foreign languages where I managed to:
  - create a **recommender system** that used a **curriculum based approach** in order to recommend content to the users adequate to their level of understanding of a language, thus ensuring that a user has a progressive learning trajectory;
  - design a pipeline where teachers could upload raw data like texts or audio records and they would then received **AI generated exercises** for learning foreign languages (of several kinds) for their students.

- **AI Scientist I and Deliverable Lead** at **Ordaos Bio** (Oct 2021 - May 2022) - human-enabled, machine-driven drug design company where I was able to achieve:
  - development **ELT** and **ETL** pipelines able to load protein data from various formats and then transform that data and make it ready for training;
  - training of various **GAN, CycleGANS, Perceivers** and other transformers architectures in a **multimodal** setting (where the modalities were the primary, secondary and tertiary structures of proteins) having as downstream **classification** and **regression** tasks for certain proteins properties, as well as cross-generating between the input modalities;
  - organizing and leading teams of up to 4-6 developers for almost 3 months as a **Deliverable Lead** by organizing the tasks for the incoming sprints, deciding priorities and other architecture / system design choices.

- **CEO, Co-Founder and Machine Learning / Backend Engineer** at **DataZard** (Mar 2021 - Dec 2021) - social media monitoring start-up in which I:
  - lead and was responsible of the AI and part of the backend side of the **MVP** that we built (8 people were working in this venture at its peak);
  - contributed to our qualification in **"Innovation Labs 2021"** Semifinal and accepted in the **"The No Code AI Peltarion"** and **"Bucharest AI"** accelerators.

- **Machine Learning Engineer** at **EmailTreeAI** (Aug 2021 - Sep 2021) - Luxembourg based start-up specializing in automatic email answering where I developed:
  - an **Information Extraction** component able to continuously create rules for event extraction as live time examples were provided to the pipeline - this was then helpful as a feature extraction mechanism for generating better informed replies to emails;
- **Machine Learning Engineer** at **Metaminds** (Oct 2020 - Jul 2021) - AI Romanian based start-up included in FT1000 where I:
  - created a live video stream action classifier receiving multiple frames as input;
  - developed an object detection model with inference in live video stream data;
  - helped in having both models succesfully integrated in a functional mobile app.
- **Junior Machine Learning Intern** at **Institute for Advanced Technologies (ITA)** (Jun 2020 - Sep 2020) - research flagship unit of the **Romanian Intelligence Service (SRI)** where I was responsible for:
  - collecting and leading the labelling of a novel sentiment analysis dataset;
  - training a sentiment analysis model which was then integrated;
  - building a a social media monitoring tool that scrapped Twitter and summarised the public perception of a brand in visual reports in real time;
- **Tutor** at **Faculty of Mathematics and Computer Science, University of Bucharest** (Oct 2019 - May 2020) - where for students in their first year of Undergraduate Studies of Computer Science I co-held the following tutorials:
  - Programming of Algorithms - backracking, greedy, dynamic programming etc;
  - Data Structures - stacks, queues, BSTs, trees, hash tables, Huffman coding etc.

Research experience

- **First running author** of a published paper entitled **"Automatic monitoring and analysis of brands using data extracted from Twitter in Romanian"**. The paper has been accepted at **IntelliSys 2021 Conference** and will be published in the Springer series "Advances in Intelligent Systems and Computing"

Education

2018–2021 **Faculty**, *University of Bucharest, Faculty of Mathematics and Computer Science*, Bucharest, Ranked 16th out of 1019 candidates in the admission contest, student in the "Performance Group", Weighted average GPA - 8.92/10.00, BSc. thesis grade - 10.00/10.00.
Undegraduate field: Bachelor of Science in Computer Science

2021–2023 **Master**, *University of Bucharest, Faculty of Mathematics and Computer Science*, Bucharest, Master of Sience in Data Science.
Ranked 1st out of 44 candidates in the admission contest, Admission mark - 9.84, MSc. thesis grade - 10.00/10.00

July 2022 **Summer School**, *Eastern European Machine Learning Summer School - EEML 2022*, University of Vilnius, Faculty of Mathematics and Informatics (backed by **DeepMind**), Vilnius.
Attended Machine Learning, Explainability, Causality, Bayesian Learning, Diffusion Models, Interpretability, Bayesian Learning, GraphNets, Reinforcement and Deep Learning courses

August 2022 **Summer School**, *Three Seas Leadership School - TSLS 2022*, SGH Warsaw School of Economics (backed by **Polish National Agency for Academic Exchange**), Warsaw.

Attended Business, Marketing, Fundraising, Negotiation, PR, Financing & Leadership courses and lead a team designing a strategy of development for the field of **new technologies** in the member countries of the **3 Seas Initiative**

Programming languages and Technologies

Experimented	Python, Numpy, Pandas, Sklearn, PyTorch, Django, transformers	**Advanced**	NLTK, SpaCy, Imgaug, Fast API, Tensorflow, Keras, gensim, git
Average	SQL, C++, JavaScript, OpenCV	**Begginer**	R, Flask, Tornado

Courses and Accreditations

- Completed Udemy Course **"Natural Language Processing: NLP With Transformers in Python"** (September 2021 - October 2021)
- Completed in the Udemy Course **"Complete Data Science Bootcamp"** provided by 365 Careers (June 2020 - July 2020)
- Completed **"Machine Learning Fundamentals CS302"** - machine learning Course in Python - TelAcad. (Jul 2019 - Sep 2019)

Contests and Competitions

- **"Medical Image Multi - Label Classification - Classify medical images in a multi-label setup"** Kaggle Competition - ranked 17/44 (Nov 2021 - Dec 2021)
- **"Word Complexity Estimation - Predict the probability of a word being complex"** Kaggle Competition - ranked 13/89 (Nov 2021 - Dec 2021)
- **SEMEVAL 2022** - "Multimedia Automatic Misogyny Identification" Competition - ranked 32/83 on Task B and 52/83 on Task A (Jan 2022)
- **Textract NLP Hackathon** - Classifying news articles in automatically generated or written by human author - ranked 7/16 (07.12.2019 - 08.12.2019)

Projects and Volunteering

- **Co-Founder and Co-organizer** of a Machine Learning hackathon - "Nitro Language Processing Hackathon" for 2 editions (Jan 2022 - March 2022; Jan 2023 - March 2023)
  - chose **Named Entity Recognition** as task for the 1st edition, with 2nd edition having **Sexism Detection** as task;
  - scrapped several datasources in order to create a **novel corpuses** for both editions and then organized the **adnotation process** of these datasets by teams of up to **10-20 students**;
  - convinced **Bitdefender, EEML and Sparktech** to sponsor us and thus we were able to reward the winners with prizes of up to 1.500 euros and free of charge participation at **EEML 2023** in Kosice, Slovakia;
  - hosted the **Kaggle competitions** themselves in which 150 (240 in the 2nd edition) students from different universities and highschools from across the country competed.

- Member of a **machine translation** team project where I collected a dataset of **lexical, grammatical and phonetical** varieties that exist in various dialects of Romanian and created two approaches for translating between these dialects a **rule-based** one and a deep learning **encoder-decoder** based one.
- Applied several **clustering** and **dimensionality reduction** techniques in a **data analysis & visualization** team project based on Beverages & Breweries dataset
- Lead a team project called "WhatsApp Group Chat Data Analysis" team project involving text analysis of messages posted in a whatsapp group (applied **language detection, text classification, machine translation and topic modelling** (Dec 2021 - Jan 2022)
- Developed a Stock Price prediction series of backtested models that use **Deep Learning** and NLP techniques / algorithms in order to trade stocks based previous **stock prices, financial reports and news sentiment / emotion analysis** (Jan 2021 - May 2021)
- Developed an adapted algorithm for **topic modelling** that uses **Sentence Embeddings** in order to filter out noisy clusters (Apr 2021 - May 2021)
- Lead a team project on "Hate Speech Author Profilling" with an ablation study comprised of several DL architectures, some of them were: **BERT, TextRNN, Seq2Seq Attention, CharCNN and TextCNN** (Feb 2021 - Mar 2021)
- Built a **CoReference Resolution** ML model where NLP features are derived from a text and fed as input to a model able to predict whether a pair of tokens is a coreference or not (May 2021)
- Lead a Computer Vision team project for the "University of Bucharest Software Development Center" using **Handwritten Recognition** and **Optical Character Recognition** techniques with the aim to extract data from the documents High-school students send in order to participate in the Admission Contest of the Faculty of Mathematics and Computer Science, University of Bucharest. (Mar 2020 - Jun 2020)
- Performed a human organ segmentation by clusterizing pixels from images in **DICOM** format in order to better localize tumors patients (Mar 2020)
- Created a model able to **classify** tweets in hate/non-hate classses and predict the topic of StackOverflow Questions using Python and ML (Sep 2019)