EDUCATION

2012-2017 **University of Manchester, PhD Bioinformatics:**

- Building supervised (Naive Bayes - based) and unsupervised (Latent Variable Allocation - based) machine learning models of gene regulation from ChIP-seq and RNA-seq time course data. The project involved writing fast, computationally efficient models from scratch in Python and some use of R for clustering and visualisation of data (please see my GitHub). The models were ran on CSF (high performance computing cluster at the University of Manchester) utilising multiple processors.

  Attended "Gaussian Process Models Summer School (Sheffield, June 2013)" workshop.

  Attended "Rapid Development and Distribution of Statistical Tools for High-Throughput Sequencing Data (RADIANT) (Naples, 2-3 July 2015)" workshop:

2008-2011 **Manchester Metropolitan University, BSc Mathematics (first class)**

- Modules taken: Operational Research, Statistics (classification, linear regression, time series data analysis), Probability Theory, Dynamical systems and Chaos, Advanced numerical methods for Ordinary and Partial Differential Equations (ODEs and PDEs), Financial Mathematics, Linear Algebra, Analysis, Vision and Sound Processing.
- Extensive use of Matlab (numerical ODEs, PDEs, image and sound processing).

2007-2008 **Manchester Metropolitan University, Chemistry (grades above 80%)**

2004-2007 **High school No. 1, Mikolow, Poland, New Matura (A levels equivalent) in:**

- Biology, Mathematics, English, Polish

WORK EXPERIENCE

May 2022 - **Lead Data Scientist at GynCentrum (contract/remote), Poland**

- Developed cutting-edge CNN models to evaluate the viability of developing embryos from video data, effectively predicting their potential for successful pregnancies. Employed advanced frameworks such as PyTorch, Detectron2, LSTM-CNN, and 3D-CNN models, among others.

- Using Gitlab CI/CDfor deployment.

September 2022 - June 2023

Data Scientist at MediaMath/Varwise (contract/remote), US branch

- Developed and deployed machine learning models in the programmatic advertising domain, leveraging DataBricks, AWS, PySpark, Facebook Prophet, unittest, Github, and CircleCI. Collaborated closely with software developers to ensure seamless integration and adherence to industry best practices, guaranteeing production-ready code.

Jan 2020 - May 2022

Expert Data Scientist at Warsaw Stock Exchange, Warsaw (remote)

- Simultaneously managed a team of data scientists while actively contributing to the coding efforts. Collaborated closely with the team in developing financial models using PySpark, Dask, Kafka, and incorporating some JavaScript. These models were instrumental in accurately predicting trends and prices of financial instruments by leveraging high frequency tick and end-of-session data. Utilised RNN, LSTM, Transformer architecture, and continual Gaussian Process.

Aug 2019 - March 2020

Scientific Software Developer at DNA electronics, London

- Developing sequence calling algorithms using scientific Python (Numpy, Pandas, Numba).

- Utilising Batch K-means and DBSCAN clustering algorithms from Scikit-learn for clustering of large volumes of time course data and OpenCV for image segmentation.

- Preparing visualisations using Matplotlib.

Jul, 2018 - Jul 2019

Senior/Lead Data Scientist at ST-YL.com, Dubai (remote)

- Built a novel visually-aware recommender system based on Bayesian statistics.

- Developed deep neural networks for classification and search of visually similar items – utilised Mask-RCNN, Faster-RCNN, CNN-classifiers with hybrid loss functions, CNN-auto-encoders.

- Administered its own Linux-based, dual-GPU server and remote cloud/production server – AWS.

- Converted business needs to machine learning solutions – design and delivery of products.

- Built communication of the team's APIs with other APIs in the REST-based architecture.

Feb-Jul, 2018

Data Scientist at ING BANK, Katowice, Poland

- Developed novel attribution models for marketing based on Markov chains and their stationary distributions. Utilised SQL (HIVE), Spark (PySpark) and Hadoop for distributed processing and data storage.

Mar-Sep, 2017

Machine Learning Engineer at Amplyfi LTD, Cardiff – Natural Language Processing, meaning extraction from large volumes of text documents.

- Applied Random Forest and Logistic Regression to filter out data.
- Applied likelihood ratio tests for hypothesis testing on bigrams using data from Wikipedia.
- Topic modelling.
- Extensive use of Python: Numpy, Pandas, NetworkX, Scikit- learn and PyTables.
- Attended "Advances in Data Science 2017, 15-16 May 2017", a conference organised by the Data Science Institute of the University of Manchester, with a focus on recent developments in data science, advanced analytics (machine learning, Bayesian statistics, scalable algorithms), privacy, visualisation, software and diverse applications.

2013-2016 **Secretary of the Polish Student Society at University of Manchester**

Responsible for communication with the Consulate General of Poland in Manchester.

Responsible for social events.

2008-2012 **Employed in Retail**

COMPUTATIONAL/DATA SKILLS

- Proficient Linux and AWS user
- Programming languages:

  Proficient - Python: Numpy, PySpark, Pandas, Scipy, Cython, Pytables, Scikit-learn, Numexp, Multiprocessing, Matplotlib, NetworkX, Keras, TensorFlow (Basic), GPflow (Basic), Edward (Basic), PyMongo;

  Intermediate – Matlab;

  Basic - R, C#, Java, C/C++, SQL, Hadoop, docker, git, bash
- Knowledge of machine learning techniques:

  Naive Bayes, Random Forest, Logistic Regression, Kernel Density Estimation, Non-negative Matrix Factorisation, Gaussian Processes (Intermediate), Neural Networks (Intermediate).
- Knowledge of statistical modelling:

  Bayesian statistics, mixture models (LDA), inference (MCMC).
- Proficient at analysis of Next Generation Sequencing (NGS) datasets:

  ChIP-seq, RNA-seq, ChIA-PET, mapping and alignment (quality checks), peak-calling (MACS2), Galaxy

ADDITIONAL SKILLS

- Driving licence – categories A2 and B
- Languages: English (fluent), Polish (native), German (intermediate)

INTERESTS & HOBBIES

Sports (cycling, gym, hiking), computer games (RPG, strategies, adventure), science, coding, travelling.