Piotr **Bielak**

Graph Machine Learning Researcher & Engineer

I am a graph machine learning specialist with a recently obtained PhD, possessing over 4 years of industrial experience. My research expertise centers on graph representation learning in terms of unsupervised and self-supervised learning, yielding over 100 citations for my work. As an accomplished author of both conference and journal articles, I've introduced several pioneering methods in this field, such as GBT, AttrE2vec, and FILDNE. I am a dedicated Python enthusiast and a well-rounded practitioner, proficient in full-stack machine learning development, including DevOps/MLOps, as well as model implementation and evaluation.

contact

Wrocław, Poland

+ 48 503 510 155

piotrbielak2@gmail.com

piotrbielak.com
pbielak@GitHub
pbielak@GitLab
LinkedIn
Google Scholar

languages

polish native
german fluency
english advanced

programming

Advanced
Python | Bash

Intermediate
Java | C++/C | SQL | ASM x86

technologies

Virtualization
Docker | Kubernetes | Openstack

Machine learning
PyTorch
Pytorch-Geometric | networkX
Pandas | Scikit-learn | NumPy
Matplotlib | Seaborn

experience

Visiting scholar

University of Notre Dame, IN, USA
08.2023 - 09.2023

During the research visit at the Lucy Institute for Data and Society (prof. Nitesh Chawla), two projects were undertaken: (1) building representations of neural networks based on their weights and training dynamics (**weight-space models**), (2) development of a novel **self-supervised graph representation learning** method, founded on the Joint Embedding Predictive Architecture. Responsibilities span across the **full research stack**, i.e., problem definition, model implementation and experimental evaluation.

Python | PyTorch | DVC | Scikit-learn | Hydra | PyTorch-Geometric | GNN

ML Ops Developer

Debster.AI (Remote)
09.2022 - 06.2023

Development of machine learning solutions tailored for **debt collection processes**. This involved the **creation and deployment of predictive models**, along with **automation of data pipelines** to enhance the efficiency and effectiveness of debt collection operations.

Python | PyTorch | DVC | Scikit-learn | Pandas | XGBoost

Senior Machine Learning Developer

Growbots (Remote)
09.2020 - 12.2022

Development of machine learning-based **recommendation** solutions for **company-company interactions**. The role encompassed comprehensive responsibilities throughout the **entire project pipeline**, i.e., from the initial data preprocessing and feature engineering stages (**text representations** and **graph building**), through model development (**GNN** and **recommendation**) to the final **deployment** of these models, ensuring that the recommendations were fine-tuned for maximum effectiveness and tailored to the specific needs of a company.

Python | PyTorch | DVC | Pandas | Jupyter | PyTorch-Geometric | GNN | Scikit-learn | Sentence-Transformers | Docker | Google Cloud | MLFlow | Weaviate | AirFlow | Streamlit

Machine Learning Developer

SecurionPay (Wrocław)
06.2019 - 08.2020

Development of a **user behavior prediction** model based on **clickstream data** using gradient boosting trees classification. Shared responsibilities across the **full pipeline**, from data cleaning and feature extraction to model training and evaluation, as well as demos preparation and a **production-level PoC implementation**.

Python | PyTorch | DVC | AWS | Docker | Jupyter | XGBoost | Redis

Machine Learning Developer

Tradeteq via Wrocław University of Science and Technology (Wrocław)

06.2019 - 12.2019

Development of a **financial transactions overdue prediction** model based on a **transaction graph**. Contributing to various stages of the **whole project pipeline**, with responsibilities in data preprocessing, feature extraction, model training, and evaluation.

Python PyTorch DVC Docker Jupyter FeatureTools NumPy GNN XGBoost

Research assistant

Wrocław University of Science and Technology

01.2019 - now

Recently finished **Ph.D. studies** at the Department of Artificial Intelligence have been accompanied by research in various areas, with a primary focus on **graph representation learning**, complemented by expertise in **self-supervised** and **unsupervised** learning. Responsible of **leading a research group** dedicated to graph representation learning. Additional practical experience in didactics, including active involvement in the **development of educational materials** for the Artificial Intelligence master's degree program.

graph representation learning self-supervised learning Probabilistic Machine Learning

Representation Learning Large-Scale Data Processing

Junior DevOps

OVH (Wrocław)

01.2018 - 12.2018

In the role within the **Public Cloud** team, responsibilities encompassed the maintenance and development of the **OpenStack** cloud infrastructure. Key tasks included implementing **test automation** using **Jenkins** and **Gerrit**, streamlining the testing process for enhanced efficiency. The notable achievement of presenting "From messy XML to wonderful YAML and pretty JobDSL – an in-Jenkins migration story" at the **OpenStack Summit Berlin 2018** underscored the commitment to improving and innovating cloud operations.

Python Bash Openstack Jenkins Gerrit

Software Developer Intern (Cloud Computing)

Intel (Gdańsk)

07.2017 - 09.2017

The role involved dedicated efforts in the **research and development** of a **machine learning**-based **resource manager** designed for modern **cluster schedulers**. This engagement contributed to the advancement of resource allocation methodologies, leveraging machine learning techniques (state-of-the-art **reinforcement learning**) to optimize the efficiency and scalability of cluster management systems. Responsibilities across the **entire project pipeline**, including environment preparation, model implementation and result analysis.

Python Tensorflow Keras Reinforcement Learning

Junior Java & Javascript Developer

Capgemini Software Solution Center (Wrocław)

10.2016 - 03.2017

Fullstack web development of an application dedicated to **staff room allocation**. The responsibilities encompassed active involvement in both frontend and backend aspects of the project, including the creation of **interactive room maps** and the development of the backend **REST API**. A pivotal role was played in **bugfixing** and the implementation of **new features**, thus making significant contributions to the overall application enhancement and functionality.

Java Spring Boot MongoDB AngularJS

QA Test Automation Engineer
Capgemini Software Solution Center (Wrocław)
02.2016 - 10.2016
Automation of integration tests for a management application for **logistics companies**. Responsibilities encompassed various aspects, including **specification analysis**, **defect reporting**, and the creation and review of test scripts. Additionally, a secondary project was undertaken involving the **development of a test script crawler** and **result analyzer** in Python, contributing to more efficient testing processes and quality assurance.

Python BeautifulSoup

education

2019 - 2023 **PhD, Machine Learning**
Wrocław University of Science and Technology
Methods for selected problems in unsupervised graph representation learning

2018 - 2019 **Master of Science, Data Science**
Wrocław University of Science and Technology
Incremental learning techniques for embedding of temporal graphs
Best Masters' thesis in Computer Science in Poland (`https://kpm.pti.org.pl/archive/winners`)

2014 - 2018 **Bachelor of Engineering, Computer Science**
Wrocław University of Science and Technology
Implementation of a neural network based process scheduler

selected publications

PyTorch-Geometric Edge – a Library for Learning Representations of Graph Edges
P. Bielak, T. Kajdanowicz
The First Learning on Graphs Conference (LoG 2022)
edge representation learning PyTorch-Geometric GNN

Graph Barlow Twins: A self-supervised representation learning framework for graphs
P. Bielak, T. Kajdanowicz, N.V. Chawla
Knowledge-Based Systems
node representation learning self-supervised learning GNN

AttrE2vec: Unsupervised Attributed Edge Representation Learning
P. Bielak, T. Kajdanowicz, N.V. Chawla
Information Sciences
edge representation learning unsupervised learning random walk

FILDNE: A Framework for Incremental Learning of Dynamic Networks Embeddings
P. Bielak, K. Tagowski, M. Falkiewicz, T. Kajdanowicz, N.V. Chawla
Knowledge-Based Systems
dynamic graphs incremental learning unsupervised learning