👤 PROFILE

I am a machine learning enthusiast, Data Scientist, Python programmer, and former network engineer. I focus on applied science, sustainability and open source solutions.

🎓 EDUCATION

PhD   2013 - 2021
Telecommunications, FEE, CTU, Prague, Czech rep.

Hierarchical density-based clustering and interpretation for network measurements.

- The aim was to detect and interpret unknown patterns in passive/active network/Cloud measurements by hierarchical density-based unsupervised machine learning techniques.

Bc/Ing (Masters)   2004 - 2009
Telecommunications, FEE, STU, Bratislava, Slovakia

- Bc.: Measurement of glottal period of human voice
  - Measurement, automatic marking and analysis of glottal period of human voice
- Ing.(Masters): Classificators for identification of the speaker
  - Building on top of the Bc. results by identification of speakers by glottal period and other aux. features

👑 OPEN SOURCE CONTRIBUTIONS

Collaborator   01/07/2021 - now
pytorch-widedeep

Collaboration on open source deep learning library pytorch-widedeep:

- Deep Imbalanced Regression
- New loss functions
- Custom Imbalanced DataLoader

External collaborator/Volunteer   01/04/2020 - 01/03/2021
Wikimedia Scoring platform team

Collaboration on Data Science related tasks under the username Pavol86, please see the Phabricator tickets and related git merges:

- Compress Gensim models
- python-mwtext
- Tokenization of "word" things for CJK
- deltas
- revscoring

Pavol Mulinka

Data Scientist / Engineer / Machine learning Enthusiast

- ✉ mulinka.pavol@gmail.com
- 🌐 Slovak
- 🪪 driving lic. A+B
- 🪪 diving lic. OWD
- 🕐 CET (UTC+1)
- Google scholar zsJ4nfoAAAAJ
- Ⓘ Orcid 0000-0002-9394-8794
- Scopus 53980138500
- WoS ABG-8213-2020
- 🔗 mulinka
- ⬛ 5uperpalo
- 📄 CV - full
- 📄 CV - WTJ

LANGUAGES

English (Professional)

Spanish (Intermediate)

Slovak (Native)

Czech (Fluent)

INTERESTS

Climbing/Bouldering

Motorcycles

Ceramics/Pottery

Nature (Hiking/Trail running)

💼 RECENT EXPERIENCE - WORK

Cybersecurity Data Analyst (freelancer) - hybrid 17/09/2024 - now
Forescout, Eindhoven, Netherlands

Analysis and classification of data from diverse security tools and systems. Pattern, anomaly, and potential security threats detection. Research and application of machine learning methods (NLP, LLM, statistics) to enhance and enrich solution outputs.

Data scientist (Researcher) - onsite 23/11/2020 - 30/09/2024
CTTC, Barcelona/Castelldefels, Spain

Design and analysis of machine learning approaches for pattern and anomaly detection in synthetic & real-world data. For related projects see the *Projects* section. Related projects:

- SUCCESS-6G: Towards robust, secure and computationally efficient vehicular services in 6G
    - git repo - https://5uperpalo.github.io/success6g-edge/
    - description - data analysis, design, and implementation of the solution for AIoT in V2X
    - used technologies - Python(numpy, pandas, matplotlib, etc.), MySQL, MLflow, Kubeflow, Kubernetes(Minio, Prometheus, Zero-to-Jupyterhub, Flower, Kserve, Knative, Istio, Kepler, Grafana, InfluxDB), streamlit, models(LLMs/Huggingface, Transfrormers, GBMs, sklearn, torch, pytorch-widedeep)
- FIREMAN (Framework for the Identification of Rare Events via MAchine learning and IoT Networks)
    - git repo - https://github.com/5uperpalo/FIREMAN-project
    - description - data analysis and implementation of the solution for classification of rare events in IIoT
    - used technologies - Python(numpy, pandas, matplotlib, etc.), models(LLMs/Huggingface, Transfrormers, GBMs, sklearn, torch, pytorch-widedeep, HDBSCAN, OPTICS, DBSCAN, GANs, GAIN, scikit-multiflow/river, deep-river), MOA, Kafka, Airflow, MLflow, KSQL, Faust, InfluxDB, Flask

Python programmer (freelancer) - hybrid 01/03/2021 - 26/01/2024
Assetario, Bratislava, Slovakia

Design, analysis and impl. of machine learning approaches for (i) prediction of customer lifetime value in mobile apps and (ii) in-app purchase recommendation system. Related projects:

- IAP (In-App-Purchases)
    - description - data analysis, design and implementation of recommendation system for In-App-Purchases on mobile phones
    - used technologies - Python(numpy, pandas, matplotlib, etc.), Athena/SQL, AWS, models(LLMs/Huggingface, Transfrormers, GBMs, sklearn), Weights&Biases
- PLTV (Predicted LifeTime Value)
    - description - data analysis, design and implementation of solution to predict life time value of customers from aggregated data from their early in-app behavior, and additional auxiliary data

- used technologies - Python(numpy, pandas, matplotlib, etc.), Athena/SQL, AWS, models(LLMs/Huggingface, Transfrormers, GBMs, sklearn, torch, pytorch-widedeep, H20, DBSCAN, OPTICS), Weights&Biases,

💼 EXPERIENCE - WORK

Python specialist (freelancer)
01/06/2020 - 01/10/2020
SEAS, Bratislava, Slovakia

Design and development of communication interface between Slovak Electricity Hydro optimization model and user GUI. Docker containerized solution utilizing Redis for in-memory database, Flask as web framework and Celery for multi-processing. Design of code unit testing. Transformation of procedural code to object oriented. Code refactorization.

F5 loadbalancer specialist (freelancer)
01/12/2021 - 01/10/2022
Oksystems, Prague, Czech rep.

Migration of existing Apache XML firewall/loadbalancer solution to F5 loadbalancer for a Ministry of agriculture project

Previous positions
01/08/2007 - 20/11/2018
in descending order

- Network Engineer (VSHosting, Prague, CZ)
- Network Consulting Engineer (Verizon, Prague, CZ)
- Senior System Engineer (AT&ampT, Bratislava, SK)
- HP Radia Specialist (Soitron, Bratislava, SK)
- HP Monitoring Support Specialist (Soitron, Bratislava, SK)
- IT VoIP support specialist (Soitron, Bratislava, SK)

🎓 EXPERIENCE - INTERNSHIPS

Data scientist
08/03/2019 - 03/09/2019
NII, Tokyo, Japan

Application of the unsupervised machine learning (ML) approaches to network (NW) traces (MAWI, Darknet). Generalization and improvement of the hierarchical density-based clustering approach to NW measurements interpretation proposed during AIT Vienna internship. Improvement of PySpark ML scripts running in distributed UX server environment. Results were summarized in conference papers.

Data scientist
21/11/2018 - 20/02/2019
O2 Telefonica, Barcelona, Spain

Analysis of the relations between socioeconomic status of customers and network performance, and investigation of potential discrimination in network deployment and management. Correlating LSOA database (Lower-layer Super Output Areas) and operator measurements by Geographic Information System (QGIS, ArcGIS, GeoPandas) in distributed computing env. (PySpark).

Data scientist
01/03/2018 - 31/08/2018

AIT, Vienna, Austria

Cybersecurity and network performance analysis, anomaly detection and diagnosis. Application of supervised, unsupervised, batch and stream-based machine learning techniques on big network measurement datasets (MAWI and Cloud latency). Integration of machine learning approaches into big data analytics platforms - in particular, working on a distributed computing environment within the BIG-DAMA project. Utilization of distributed computing tools and platforms such as Cloudera, PySpark, Apache Pig, Hive, Kafka, Elasticsearch etc.. Running and configuration of machine learning bash script on linux server. Results were summarized in conference papers.

EXPERIENCE - PEDAGOGICAL

Mentor
(Sakura Science Plan)
2019

NII, Tokyo, Japan

Mentoring, academical support and provisioning of computing environment for under-grad intern visiting NII Tokyo for three weeks supported by Sakura Science Plan internship.

Teaching assistant
(Network Operating Systems)
2015, 2016 winter sem.

Czech Technical University, FEE, Prague, Czech republic

Network operating systems, Linux, Unix. Administration and network tools, managing and administration of documentation. Basic concepts, configuration and procedures in operating systems administration (UNIX).

Teaching assistant
(Digital Engineering)
2014 winter sem.

Czech Technical University, FEE, Prague, Czech republic

Basic principles of both classical and programmable logic devices and their practical use in the design of digital systems. Design and implementation of digital circuits VHDL language. Implementation of logic gates, measurement of their static and dynamic properties. Verification of digital circuits in the simulator.

Teaching assistant
(Communication Processes Control)
2014, 2015 summer sem.

Czech Technical University, FEE, Prague, Czech republic

Review of switching systems solution principles, i.e. (i) switching fields, (ii) control systems and (iii) signalization for switching control (in central office as well in networks). Focus on digital switching systems with circuit commutation as well as transport of IP packets. Basic review and consideration about convergence of voice and data services and networks including functional principles of new generation networks with respect to philosophy and services of intelligent network.

🎓 CERTIFICATIONS (EXPIRED)

Cisco

- Cisco Certified Network Associate (CCNA, 640-802)
    - (640-553) Implementing Cisco IOS Network Security
    - (640-460) Implementing Cisco IOS Unified Communications
    - (640-721) Implementing Cisco Unified Wireless Net. Essentials
- Cisco Certified Design Associate (CCDA)
    - (640-863) Designing for Cisco Internetwork Solutions
- Cisco Certified Network Professional (CCNP)
    - (642-901) Building Scalable Cisco Internetworks
    - (642-812) Building Cisco Multilayer Switched Networks
    - (642-825) Implementing Secure Converged Wide Area Networks
    - (642-845) Optimizing Converged Cisco Networks
- Cisco Certified Internetwork Professional (CCIP)
    - (642-642) Quality of Service
    - (642-611) Multiprotocol Label Switching
    - (642-661) Border Gateway Protocol
- Cisco Certified Design Professional (CCDP)
    - (300-320) Designing Cisco Network Service
    - (642-873) Designing Cisco Network Service Architectures
- Conducting Cisco Unified Wireless Site Survey (CUWSS, 642-731)
- Implementing Cisco Edge Network Security Solutions (SENSS, 300-206)

F5

- F5 Certified Product Consultant for LTM (F5-PCL, F50-531)
- F5 Certified Administrator
    - (101) Application Delivery Fundamentals
    - (201) TMOS Administration

Juniper

- Juniper Networks Certified Internet Associate EX (JNCIA-EX, JN0-400)

Other

- Information Technology Infrastructure Library Foundation in IT Service Management (ITILv3, Foundation)
- The Open Group Architecture Framework (TOGAF 9)
- ArchiMate 3

🗄 OTHER PROJECTS

PhD

- Metrics for Automated Detection of Cloud Anomalous Behavior} focused on automated detection and interpretation of suspicious events in active Cloud latency measurements
- Practical Privacy-Preserving Data Collection and Utilization using Provable Cryptographic Tools
- Privacy Protection and Machine Learning Utilization of IoT Data in Cloud
- Cloud Performance Analysis and Improvement
- Smart-home IoT and Cloud Telemetry Datamining
- Methods Enhancing Work with Cloud Data

Hobby coding

- SUHEC (SUrname HEritage Classifier)

- git repo - https://5uperpalo.github.io/surname_heritage_classifier/
- description - A hobby project to classify surnames to countries and areas of the world. An attempt for an open source alternative to paid services.
  - CHURNPRED (CHURN PREDictor)
    - git repo - https://github.com/5uperpalo/churnpred
    - description - Customer churn predictor.

📄 PUBLICATIONS

- pytorch-widedeep A flexible package for multimodal deep learning
Zaurin, Javier Rodriguez and Mulinka, Pavol
Journal of Open Source Software, 2023

- A Robust and Explainable Data-Driven Anomaly Detection Approach For Power Electronics
Beattie, Alexander and Mulink, Pavol and Sahoo, Subham and Christou, Ioannis T. and Kalalas, Charalampos and Gutierrez-Rojas, Daniel and Nardelli, Pedro H. J.
2022 IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids (SmartGridComm)

- Optimizing a Digital Twin for Fault Diagnosis in Grid Connected Inverters - A Bayesian Approach
Mulinka, Pavol and Sahoo, Subham and Kalalas, Charalampos and Nardelli, Pedro H. J.
2022 IEEE Energy Conversion Congress and Exposition (ECCE)

- A Large-Scale Examination of "Socioeconomic" Fairness in Mobile Networks
Park, Souneil and Mulinka, Pavol and Perino, Diego
Proceedings of the 5th ACM SIGCAS/SIGCHI Conference on Computing and Sustainable Societies

- Information processing and data visualization in networked industrial systems
Mulinka, Pavol and Kalalas, Charalampos and Dzaferagic, Merim and Macaluso, Irene and Rojas, Daniel Gutierrez and Nardelli, Pedro Juliano and Marchetti, Nicola
2021 IEEE 32nd Annual International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)

- NetSEC at High-Speed Distributed Stream Learning for Security in Big Networking Data
Casas, Pedro and Mulinka, Pavol and Vanerio, Juan
Data Science-Analytics and Applications, 2021, Springer

- Adaptive and Reinforcement Learning Approaches for Online Network Monitoring and Analysis
Wassermann, Sarah and Cuvelier, Thibaut and Mulinka, Pavol and Casas, Pedro
2020 IEEE Transactions on Network and Service Management

- HUMAN - Hierarchical Clustering for Unsupervised Anomaly Detection $\&$ Interpretation
Mulinka, Pavol and Casas, Pedro and Fukuda, Kensuke and Kencl, Lukas
11th international Conferece on Network of the Future IEEE 2020

- WhatsThat? On the Usage of Hierarchical Clustering for Unsupervised Detection & Interpretation of Network Attacks
Mulinka, Pavol and Fukuda, Kensuke and Casas, Pedro and Kencl, Lukas
The 5th International Workshop on Traffic Measurements for Cybersecurity 2020 IEEE

- Should I (re)Learn or Should I Go(on)? Stream Machine Learning for Adaptive Defense against Network Attacks

  Casas, Pedro and Mulinka, Pavol and Vanerio, Juan

  The 6th ACM Workshop on Moving Target Defense (MTD 2019)

- Continuous and Adaptive Learning over Big Streaming Data for Network Security

  Mulinka, Pavol and Casas, Pedro and Vanerio, Juan

  IEEE International Conference on Cloud Networking CLOUDNET, 2019 International Conference on

- ADAM & RAL Adaptive Memory Learning and Reinforcement Active Learning for Network Monitoring

  Wassermann, Sarah and Cuvelier, Thibaut and Mulinka, Pavol and Casas, Pedro

  15th International Conference on Network and Service Management 2019 IEEE

- Remember the Good, Forget the Bad, do it Fast Continuous Learning over Streaming Data

  Mulinka, Pavol and Wassermann, Sarah and Marín, Gonzalo and Casas, Pedro

  NeurIPS 2018 Workshops, Workshop on Continual Learning

- Hi-Clust Unsupervised Analysis of Cloud Latency Measurements through Hierarchical Clustering

  Mulinka, Pavol and Casas, Pedro and Kencl, Lukas

  IEEE International Conference on Cloud Networking CLOUDNET, 2018 International Conference on

- Stream-based Machine Learning for Network Security and Anomaly Detection

  Mulinka, Pavol and Casas, Pedro

  Proc. of the Workshop on Big Data Analytics and ML for Data Comm. Net., Big-DAMA@SIGCOMM

- Adaptive Network Security through Stream Machine Learning

  Mulinka, Pavol and Casas, Pedro

  Proceedings of the ACM SIGCOMM '18 Posters and Demos

- Multidimensional cloud latency monitoring and evaluation

  Tomanek, Ondrej and Mulinka, Pavol and Kencl, Lukas

  Computer Networks, Elsevier

- Learning from Cloud latency measurements

  Mulinka, Pavol and Kencl, Lukas

  Communication Workshop (ICCW), 2015 IEEE International Conference on

- Speaker identification by K-Nearest Neighbors Application of PCA and LDA prior to KNN

  Kacur, Juraj and Vargic, Radoslav and Mulinka, Pavol

  Systems, Signals and Image Processing (IWSSIP), 2011 18th International Conference on
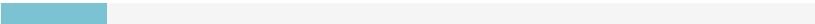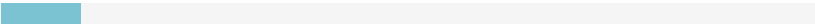
🔧 SKILLS & PROFICIENCY

Python

Kubernetes/Helm/Docker

Machine learning

Linux/Unix

Networking