Sasho Nedelkoski

sashonedelkoski@gmail.com
snedelkoski.github.io

Summary

My interest in both research and engineering is machine learning for anomaly detection. My research focus lies in anomaly detection models for code analysis and natural language data, whereas my engineering focus lately is on development of end-to-end ML systems for software data to support various software development tasks.

Education

Berlin, Germany — Technical University of Berlin — October 2018 – April 2021

- **Ph.D. in Computer Science** (grade: summa cum laude), supervised by Prof. Dr. Odej Kao
- Thesis: "Deep Anomaly Detection in Distributed Software Systems"

Berlin, Germany — Technical University of Berlin — October 2017 – September 2018

- **M.Sc. in Computer Science** (GPA: 1.0 German grading system)
- Thesis: "Event-generated Time Series Anomaly Detection using Deep Learning"
- Best student awards from VDI & TU Berlin, and best master thesis

Skopje, Macedonia — Ss. Cyril and Methodius University — September 2013 – June 2017

- **B.Sc. in Computer Technologies and Engineering** (GPA: 9.93/10.0 Macedonian grading system)
- Thesis: "Lung Cancer Detection using Deep Learning"
- Best student award (2013/2014, 2014/2015, 2015/2016 and 2016/2017 academic year)

Work experience

Chief Data Scientist — logsight.ai — August 2021 – Present

- Developed software verification system incl. data ingestion and integrations with different logging processors and CI/CD pipelines, processing pipeline with ML models in its core, and visualization
- Developed system-agnostic NLP model for anomaly detection in application logs
- Developed NLP model for code analysis recommending "where to log?"
- Implemented data differentiation method that is primarily used for software data verification

Research Associate — Technische Universität Berlin — October 2017 – present

- Performed research and development on anomaly detection, distributed systems reliability, and learning from heterogeneous data
- Worked on various ML projects building end-to-end tools funded from Huawei, Berlin Big Data Center, and BIFOLD
- Responsible for teaching of various seminars and projects

Skills

Programming languages and frameworks

- Python, PyTorch, Python Analytics Stack, MLFlow, Kafka, Pyspark, Airflow, Spring Boot with Kotlin, Relational and NoSQL Databases, Docker

Languages

- English (full professional proficiency), Macedonian (native), German (A2), Serbo-Croatian (professional proficiency), Spanish (limited working proficiency)

Research highlights

1. Sasho Nedelkoski, Jasmin Bogatinovski, Alexander Acker, Jorge Cardoso, and Odej Kao. "Self-Attentive Classification-Based Anomaly Detection in Unstructured Logs." In Proceedings of the 20th IEEE Interna-tional Conference on Data Mining (ICDM2020). 2020.
2. Sasho Nedelkoski, Jasmin Bogatinovski, Alexander Acker, Jorge Cardoso, and Odej Kao. "Self-Supervised Log Parsing." In Proceedings of the European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML-PKDD2020). 2020.
3. Sasho Nedelkoski, Jasmin Bogatinovski, Jorge Cardoso, and Odej Kao. "Self-Supervised Anomaly Detection from Distributed Traces." In Proceedings of the 13th IEEE/ACM International Conference on Utility and Cloud Computing (UCC2020). 2020.
4. Sasho Nedelkoski, Jorge Cardoso, and Odej Kao. "Anomaly Detection and Classification using Distributed Tracing and Deep Learning." In Proceedings of the 19th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing (CCGRID). 2019.
5. Sasho Nedelkoski, Jorge Cardoso, and Odej Kao. "Anomaly Detection from System Tracing Data Using Multimodal Deep Learning." In Proceedings of the 12th IEEE International Conference on Cloud Computing (CLOUD2020). 2019.

Community service

PC & Reviewer: ICDM, ECML-PKDD, TNSM, DKE, TKDE

Competitions

- Kaggle Quora Question Pairs (May 2017) – *Gold medal, top 0.3%*. Developed complex ensemble of machine learning models (multiple deep learning and tree boosting methods)
- Kaggle Data Science Bowl 2017– Lung Cancer Detection (April 2017) – *Silver medal*, top 4%. Using thousands of high-resolution lung scans developed ensemble of deep learning models that determine when lesions in the lungs are cancerous
- Kaggle Bosch Production Line Performance (November 2016) – *Silver medal*, top 3.6%. Developed solution using ensembles and gradient boosting
- Kaggle Predicting Red Hat Business Value (September 2016) – *Silver medal*, top 1.6%. Solution using ensembles and gradient boosting.

Personal links

- Google Scholar: https://scholar.google.de/citations?user=4we2u34AAAAJ
- LinkedIn: https://www.linkedin.com/in/snedelkoski/
- Kaggle: https://kaggle.com/salkaa