Paul Westenthanner

Keuslinstraße 4
80798 München
Germany
Mobile: +49 1745476560
Paul.Westenthanner@web.de

Work Experience:

Freelance
Data Scientist / Data Engineer / Machine Learning Engineer
Munich, Germany
Apr. 2017 – present

I focus deliver clean-code machine learning pipelines using the Spark/Hadoop ecosystem with Scala, Python or R using agile DevOps practices. Projects include:

- *Lab Automation for Roche (18 months)*:
  I'm the only software engineer in a team of biochemists on a laboratory automation project. My responsibilities include developing applications for automating common tasks, writing interfaces to connect laboratory devices as well as advising the team on tooling decisions and architecture. Furthermore, I maintain various convenience R-Shiny apps used by the scientists.
  Technologies: Python, R-Shiny, Docker, MongoDB, various lab-specific tools.
- *Time Series Analysis Framework for Telefonica (2 years)*:
  As the lead backend developer of a time series problem detection framework, I assured high code quality, trained junior developers and suppored end users. Furthermore I advised other projects on software architecture, efficient computing and theoretical topics in machine learning.
  Technologies: PySpark, Hive, Oracle PL/SQL, Apache Kafka, Apache Hbase, R-Shiny, Tableau.
- *Recommendation system for Adidas (5 months):*
  I worked on a variety of data driven Proof-of-Concepts like improving product recommendations, click-stream analysis and purchase propensity.
  Technologies: Python, SQL (Exasol), AWS.
- *Fraud detection model for OTTO (4 months)*:
  Development and productionisation of a fraud-detection-model based on click-stream data. I translated an existing non-optimal PySpark model to Scala improving the performance by two orders of magnitude while also making it more robust. Furthermore, the model was deployed in streaming mode for real-time predictions.
  Technologies: Scala, Apache Spark, Apache Hadoop, Apache Kafka.

Vodafone Group
Data Scientist
London, UK
Mar. 2016 – Apr. 2017 (full time)

Building an end-to-end data science machine.
My responsibilities included ensuring data quality, understanding the raw data

and crafting of suitable features for predictive-analytic machine learning models. Use-cases were for example travel prediction or device recommendation. Furthermore I made key contributions to a smart Python and Scala framework to implement data flows in production on an hadoop cluster. In this context, I unit-tested the code, created and reviewed pull requests and trained colleagues and external consultants to use the tool.

Vodafone Kabel Deutschland
Working Student in Data Science and Business Intelligence
Munich, Germany
Aug. 2014 – Mar. 2016 (part time, 16h/week)

Supported BI team with multivariate statistical analysis using Python and SQL.

Education:

Oct. 2015 – Feb. 2018, Technical University of Munich
- Master of Science in Mathematics (grade 1.7)
- Thesis: Predicting Travel from Smartphone Data
- Topics in Abstract Algebra, Optimization, Statistics, Machine Learning and Computer Science

Oct. 2012 – Sept. 2015, Technical University of Munich
- Bachelor of Science in Mathematics (grade 1.7, top 15%)
- Thesis: Algorithms in the representation theory of finite groups
- Topics in Analysis, Optimization Theory, Statistics and Algebra
- Minor subject: Economics

Skills:

Programming:
- *Expert*: Python (incl. pandas, sklearn, pyspark), Scala, PL/SQL
- *Intermediate:* R, Bash
- *Basic*: Java, MATLAB

Tech Stack: Apache Spark, Hadoop, Apache Kafka, Impala, Jenkins, Keras, XGBoost, Tensorflow, CARTO, Linux, Git, OracleSQL, AWS, Jupyter Notebook, Intellij IDEA, Vim, Tableau, DevOps, GitOps

Languages: German (native), English (fluent), Latin (basic), Spanish (basic)

Other:

Speaker at @databeers London
Volunteer bike repairer at Ackermannbogen Repair Cafe
Open Source Maintainer of an categorical feature encoding package with 1M downloads/month
Interests: Chess, philosophy, cycling