Mekhti Musaev

Data Scientist/Research Engineer (Machine Learning)

Summary

Over 5 years of experience in (research) software engineering and machine learning projects. I came to machine learning with software engineering and research background. Since 2017 I have been working on different machine learning projects including NLP, Bayesian Data Analysis, Neural Networks. In my current working place I have been involving in building ML pipelines on AWS cloud service.

1. **Nominated** to FIRST+ PROGRAMME (Finnish–Russian Student and Teacher Exchange Programme) SCHOLARSHIP as an exceptional Master's student
2. **Open source projects:** There are many of open source projects from different organizations (**including Google and Aalto University**) where I have been contributed for more than 4 years
3. **Tech stack:** AWS services, Python, Tensorflow, LightGBM, GPOpt (bayesian optimization for hyperparameters tuning), Scala, Apache Spark, Linux, bash, Git, jupyter notebook, pandas, numpy, Apache Zeppelin, Jira, Confluence, scikit-learn, matplotlib, API, Flask, pytest, Jenkins, redmine

Experience

2019-05 - present

Data Scientist / Machine Learning Engineer

CGI Suomi Oy

Working as a Machine Learning Consultant for clients in different fields. Building and tuning data pipelines from scratch with a large amount of data (from 100 million up to 9 billion samples). Incorporating training and prediction workflows into data pipelines.

Technology stack : GCP, AWS EMR, AWS Lambda, AWS EC2, AWS S3, AWS Step Functions, AWS CodeBuild, AWS CodeCommit, Amazon Sagemaker, AWS SQS, AWS SNS, AWS Athena, AWS Glue, Apache Spark, Apache Kafka, Apache Zeppelin, Python (pyspark), Scala, Helium, Jenkins, Kubernetes, BigQuery, A/B testing

Projects/clients (in chronological order):

1) Building recommendation system for one of the largest video-streaming companies. In this project I was part of a team that built a recommendation system for the SHAHID service (owned by The MBC Group (Middle East Broadcasting Center) - the largest television company in the Middle East and North Africa). More info can be found ***here (in Finnish)*** . **Tech stack** : Python (pyspark), Scala, Apache Spark, Apache Zeppelin, AWS EMR, AWS Lambda, AWS EC2, AWS S3, AWS Step Functions, AWS CodeBuild, AWS CodeCommit, Amazon Sagemaker, AWS SQS, AWS SNS, AWS Athena, AWS Glue

2) Optimizing data pre-processing project (as a part of the ETL pipeline). The final result helped to decrease EMR cluster consumption for almost 50% for the same amount of data. **Tech Stack** : Python, Apache Spark,

Personal Info

Address

Espoo, Finland

Phone

+358406838452

E-mail

mehty.musaev@gmail.com

WWW

https://musayev.me

GitHub

https://github.com/b5y

LinkedIn

https://www.linkedin.com/in/mehtimusayev/

Skills

Python, PyTorch, TensorFlow, LightGBM, Linux / Mac OS, bash, Git, CI, CD, numpy, scipy, jupyter notebook, IPython, Zeppelin, scikit-learn, matplotlib, pytest, Flask, Jenkins, Jira, Confluence, Redmine, web crawling/scraping (scrapy), REST API

Proficient

seaborn, Scala, Hadoop, Spark, PostgreSQL, MySQL, NoSQL, mock, integration testing, TDD, gerrit

Competent

Software

PyCharm, Intellij IDEA, atom, sublime, vim, emacs

Languages

English

C1

Russian/Azerbaijani (Bilingual)

C2

Turkish

B2

Apache Kafka, Scala, AWS EMR cluster, AWS S3, Jenkins, Git

3) Accumulating data from IoT devices across Europe. **Tech Stack** : AWS API Gateway, AWS Lambda, AWS RDS, Python

4) Data engineering project for the company in the energy industry. Writing tests, fixing bugs. **Tech stack** : Scala, Apache Spark, Microsoft Azure

5) Machine Learning project for one of the biggest telecommunication companies in Europe. Building a smart ticketing system using NLP (German language). ML model to predict products that are likely to fail in next time. **Tech stack:** AWS S3, AWS EKS, AWS Redshift, Tensorflow, Python, pandas, matplotlib,numpy, Optuna (hyper-parameter optimization for neural network)

German

A2

2018-03 - 2018-12

Machine Learning Research Assistant

Probabilistic Machine Learning Group, Aalto University

Worked on **ELFI (Engine for Likelihood-Free Inference,** clickable) project.

Research group home page: https://research.cs.aalto.fi/pml/

Supervisor: Samuel Kaski, Professor, PI.

Technologies: jupyter notebook, scikit-learn, numpy, matplotlib, GPy, ipywidgets, IPython, pytorch

2018-09 - 2018-10

Teaching Assistant

Machine Learning for Big Data Group, Aalto University

CS-E3210 Machine Learning: Basic Principles course (clickable).

Responsibilities: helping students with assignments, explaining lectures with more details, giving more information about python, numpy, scipy, pandas and scikit-learn.

Teacher in Charge: Alex Jung, Assistant Professor of Computer Science.

2016-10 - 2017-09

Software Backend Engineer

Selectel Ltd., selectel.com

Implemented new and fixed old backend applications in Go and Python.

Participated in implementing architectural solutions (following design patterns) in products. Code review.

Implemented functional and unit-tests using mocks. Found and fixed bugs during going into projects. Wrote code documentation.

Finished projects: Rewriting traffic counter from Python to Go.

Created debian packages from projects.

Technologies: OpenStack, Python, pytest, Flask, Ubuntu, Go, Jenkins, Git, bash, MySQL, PostgreSQL, Akamai, Jira, Confluence, gerrit

2015-12 - 2016-07

Research Software Engineer

Digital Security Ltd., https://dsec.ru

Implemented static code analyzer for obfuscating algorithms (intellectual property protection), stylometry (project which collects unique style of specific developer) and transpilers (converting code from JS to Python and vice versa).

Completed projects: stylometry and partly obfuscation algorithms

Open source projects which were used (all of them are clickable): **RedBaron** , **baron** and **yapf** .

Technologies: Debian, Python Internals, Git (Continuous Integration),

Redmine, MongoDB

2016-04 - present

Open source projects

RedBaron (open source project from Python Code Quality Authority): Fixed bug with rendering in full syntax tree, updated TODO list with some improvement ideas, removed dead code in the project, Added at method to insert code to specific place in the syntax tree
GitHub page: https://github.com/PyCQA/redbaron
yapf (open source project from Google): Fixed bug with formatting unique style. Implemented Unit Tests.
GitHub page: https://github.com/google/yapf
ELFI (open source project from Aalto University): Added new example Lorenz model, fixed bug with plotting pairs for 1-D, fixed processing multivariate distributions, added progress bar for inference methods, added functionality to save samples for inference methods, fixed bug with shared axes, added functionality for pairwise contour plots of GP fit with more than 2 parameters, fixed bug in plot_discrepancy for more than 6 parameters.
GitHub page: https://github.com/elfi-dev/elfi

2012-03 - 2012-05

C# .NET developer (Internship)

Evrograndsoft LLC, http://cofite.com
Responsibilities: supporting closed part of website for schedulers of Russian Railways in the Northwestern Federal District.

Education

2018-01 - 2018-12

Aalto University, Machine Learning and Data Mining (Macadamia), Exchange student

Courses: Machine Learning: Basic Principles, Bayesian Data Analysis, Big Data Platforms (Scala), Concurrent Programming (Scala), Algorithmic Methods of Data Mining, Introduction to Analytics and Data Science, Machine Learning: Advanced Probabilistic Methods, Artificial Intelligence, Special Course on Supervised Learning with Large Label Sets, Special Course on Latent Variable Modeling and Bayesian Matrix Factorization, Mathematical Methods for Network Science.

2017-09 - 2018-10

ITMO University, Urban Supercomputing (Data Science), Master's Degree

The Faculty of Information Technologies and Programming.
Department of High Performance Computing

2009-09 - 2015-06

Saint Petersburg State University, Mathematics and Mechanics Faculty (Mathematics and Computer Science)

Thesis topic: Scala language implementation for Microsoft .NET on the basis of Common Compiler Infrastructure (CCI)
Research supervisor: Doctor of Engineering, Professor Vladimir O. Safonov

2008-09 - 2009-06

Saint Petersburg Presidential Physics and Mathematics Lyceum №239, high school

Studying mathematics and physics in a strong academic program.

Conferences

2016-12 — IT Global Meetup, Saint Petersburg, Russia. Speaker. RedBaron: The new Era in code manipulation

2017-02 — SPb Python Interest Group, Saint Petersburg, Russia. Speaker. Code metrics: why do we need them?

2017-03 — Python Users Berlin (PUB), Berlin, Germany. Speaker. RedBaron: The new Era in code manipulation

Courses

2014-09 - 2014-11 — Functional Programming Principles in Scala from École Polytechnique Fédérale de Lausanne, Switzerland (December 2014)

2015-09 - 2015-11 — Cybersecurity Specialization from University of Maryland, USA (September - November 2015)

2020-06 - 2020-06 — DeepLearning.AI TensorFlow Developer Spezialization (June 2020)

2020-07 - 2020-08 — Advanced Machine Learning with TensorFlow on Google Cloud Platform Specialization from Google Training (July-August 2020)