Yaroslav Lyutvinskiy, Ph.D. – Curriculum Vitae

Personal data:

Address: Makedoniya str. 61b, apt. 8, Varna, 9002, Bulgaria

Mobile: +(359)877-917 638

E-mail: yar_spb@yahoo.com

EU work permit: Sweden citizenship

Languages: English, Russian

Summary:

Experienced engineer with expertise in data science, data processing, statistical analysis, and software development, particularly, for mass spectrometry and proteomics. Proven track record of developing ranking methods for recommender systems, creating open-source web-service software for remote data access, and implementing label-free quantitative proteomics workflow and many other. Published research findings, presented at international conferences. Skilled in many programming languages, data science model training, data pipeline development, machine learning, and data visualization. Collaborative team player with experience in stakeholder coordination and team leadership.

Technical skills:

Programing Languages: Python, SQL, C#, C++, Wolfram Mathematica, Java, Delphi

Technical stacks:

- Python data science stack: numpy, pandas, matplotlib, scikit-learn, tensorflow, keras, lightgbm, xgboost
- Data science in production stack: docker, airflow, github, FastAPI
- Data engineering stack: Databricks, PySpark, AWS Sagemaker, BigQuery, Vertex AI
- Windows application stack: Win32, .Net, SOAP, STL, COM, VCL
- Mass spectrometry stack: Mascot, PyOpenMS, mzML, mass spectrometry APIs and libraries (Thermo, Agilent)

Software: Jupyter & Jupyter lab, AWS Sagemaker, Databricks, BigQuery, GCP, Microsoft Visual Studio 4.2-16.9, Borland Delphi 1.0-7.0, Borland CBuilder 5.0-6.0, Wolfram Mathematica 9.0-11.0, MS SQL Server, SQLite, MySQL

Certificates: Azure Data Scientist Associate, Tensorflow Certified Developer.

Professional experience:

02/2024 - present, AI/ML engineer - remote

Employee: Tallarna ltd. (UK),

Responsibility: Development machine learning method for retrofit opportunity analysis at real estate and EaaS (Energy-as-a-Service) market.

Projects: The building properties by pictures (CV). Market analysis and bridging to retrofit opportunities application for New York state LMI (low-to-middle income) NYSERDA dataset. The core modelling engine refactoring to fit ISO13790 implementation to ASHRAE 140 standard tests.

08/2023 - 11/2023, Data Scientist - remote

Employee: Turing inc. (USA), data science contractor

Responsibility: Development machine learning method for treatment outcome prediction.

Project: The pilot project for one of the major USA healthcare providers on Opioid Usage Disorder treatment outcome prediction. This provider is interested in predicting treatment outcomes at the patient admission time to make the therapy more targeted. The provided data includes a diversity of factors, both clinical and demographic. I ran a full cycle of data science development from problem formulation to the model in production.

04/2023 - 09/2023, Data Scientist - remote

Employee: Playgenes Int. Ltd. (Cyprus)

Responsibility: Game bot development and game analytics.

Project: Playgenes ltd. contracted me to develop a game bot for their casual game The Match Cafe. I developed such a bot based on residual convolutional neural net. Later, I navigated Playgenes to use such a game bot for the game analytics.

12/2021 - 12/2022, Data Scientist - remote

Employee: Turing inc. (USA), Internal core data science team

Responsibility: Development of ranking methods for developer profiles in core recommender system.

Projects: The "Feedback Features" project for Turing core data science team. I have developed an implicit feedback-driven ranking approach based on user interface signals from different stages of the developer matching funnel at Turing. The matching process at Turing is highly automated, and actors working with the matching system leave numerous signals that provide implicit feedback on the experience of actors. I performed the full cycle of development, including data collection, modeling, and productionization of the data pipeline and models.

09/2019 - 08/2021, Research engineer - remote

Employee: Karolinska Institutet, Sweden, supervised by Prof. Roman Zubarev

Responsibility: Processing methods development for native FTMS mass spectrometry data.

Main project: CHARDA - CHARge Determination Analysis[1]. A method of processing for native transient (before Fourier Transform) FTMS data for determination of charge for single signals. Logistic regression made on the set of signal parameter allows about 75% accuracy in charge determination. The resolution of method is about 6, which is better then for preceding methods. The method is implemented as Python Jupyter notebook. Detailed method description has being selected for publication.

08/2014 - 05/2019, Research engineer

Employee: Karolinska Institutet, Sweden, supervised by Dr. Roland Nilsson

Responsibility: Processing methods development for mass spectrometry data in metabolomics experiments.

Main project: mzAccess [2] is an open source web-service software for remote interactive access to the large collections of mass spectrometry data. That service hosts multi-terabyte collections of native raw data and provides on demand with millisecond scale access time any fragment of hosted data to many users on network.

mzAccess centered metabolomics processing workflow includes, particularly, interactive Inspector software for data curation and IsoTrack software for untargeted metabolomics.

09/2009 – 07/2014, Post-doc, Research engineer

Employee: Karolinska Institutet, Sweden, supervised by Prof. Roman Zubarev

Responsibility: Data processing methods development for high-throughput, high-resolution mass spectrometry.

Main project: Software workflow for label-free quantitative proteomics called Quanti for processing of Orbitrap high resolution mass spectrometry data.

Characteristic features of this software are:

- Accuracy (Coefficient of variance for major proteins in replicates – 1-3%, for the rest of proteins – 5-7%);
- Large data processing (>100 samples, >200 Gb as an integrated dataset);
- High throughput ~ 10 GB/hour on regular PC;
- Easy to use

Description of the workflow has been published in [3]. Some principles were applied for patenting with Thermo Scientific Corp, application number PCT/EP2013/056808. The work has been reported in oral presentations at IMSC2012 (Kyoto, Japan) and ASMS2013 (Minneapolis, USA). Workflow has been successfully used in a range of biological research.

As a core facility staff member, I performed support and consulting of customers with bioinformatics processing and interpretation of their MS data with Quanti workflow and statistical analysis including such methods as PCA, OPLS and correlation analysis.

11/2002 – 08/2009, Lead Software Engineer, Scientist

Employee: Laboratory of Biomedical Mass Spectrometry, Institute for Analytical Instrumentation, St. Petersburg, Russia

Position: Lead Software Engineer, later Scientist. Team lead for two engineers.

Responsibility: As a lead software engineer of the laboratory and team lead for two engineers I was responsible for all aspects of laboratory software needs, developments, and research including:

- supervise team of two engineers for software development of home-build mass spectrometers (ESI-TOF-MS) control, data acquisition and data processing system
- research in data processing and interpretation including proteomics
- coordination with physics, electronic engineers and chemists divisions to fulfill their software requirements

Main projects:

Two generations of LC-ESI-TOF-MS control, data acquisition and data processing software has been developed. First – rapidly developed to fulfill primary physics needs. Second – well-designed, reliable software assigned to a wide range of end-users [5]. The software has been used for prototypes of LECO multi-reflecting LC-HRT system gained PITCON gold award in 2011.

CrystallTag - proteomics workflow based on sequence tag search approach for protein identification and PTM discovery. Software implements naïve Bayes classifiers and graph theory approach for identification and scoring of peptide sequence tags and modified Aho-Corasick algorithm for database search. Fully functional proof-of-concept application was used for research activity and subjected for the Ph.D. thesis defense [4].

10/1996-10/2002 early software engineer experience

These years working as software engineer I completed several project in banking automation, game development, office acces system, etc. As a trainer I delivered a set of courses on Microsoft technologies for those who aimed to pass Microsoft certification. I have got MCSD and MCAD certifications from Microsoft.

Resources:

https://github.com/Yaroslav-Lyutvinskiy,

https://www.linkedin.com/in/yaroslav-lyutvinskiy-68514993/

Education:

1990-1996: Graduated specialist in software engineering
St. Petersburg State University of Telecommunications, St. Petersburg, Russia

2007: Ph.D. in engineering, Institute for Analytical Instrumentation, St. Petersburg, Russia

Main Publications:

1. **Lyutvinskiy Y, et all,** Adding colour to mass spectra: Charge Determination Analysis (CHARDA) assigns charge state to every ion peak. // JASMS, 2024 Apr 12
2. **Lyutvinskiy Y, Watrous JD, Jain M, Nilsson R**. A Web Service Framework for Interactive Analysis of Metabolomics Data.// Anal. Chem. 2017 Jun 6
3. **Lyutvinskiy Y, Yang H, Rutishauser D, Zubarev R.** In silico instrumental response correction improves precision of label-free proteomics and accuracy of proteomics-based predictive models.// Mol. Cell. Proteomics. 2013 Apr 15.
4. **Y. I. Lutvinsky, V. V. Makarov, N. V. Krasnov, E. P. Podolskaya, A. N. Verentchikov** Partial Sequencing of Peptides using Fragment Mass Spectra – Algorithm and Testing (in Russian) // Nauchnoye Priborostroyeniye. 2006. V. 16., N 3., pp. 122-131.
5. **Y. I. Lutvinskiy, D. M. Petrov, A. N. Verenchikov, Y. I. Khasin, M. A. Gavrik** Data Acquisition System for TOF-MS and TOF-MS Tandem (in Russian) // Nauchnoye Priborostroyeniye. 2004. V. 14., N 2., pp. 80-91.