

Professional summary:

I am a data scientist and software engineer with a substantial experience in big (petabyte-scale) data handling and processing. My work relies heavily on implementing efficient program algorithms, running Monte Carlo modeling, and performing statistical inference. I routinely use machine learning techniques, architect data analysis pipelines, and publish in the major peer-reviewed journals.

Highlights:

Data mining, cleaning, and imputing, principal component analysis, statistical data modeling
Pattern recognition, forecasting, classification, clustering, regression/fitting
Greedy algorithms, efficient data structures, dynamic programming
Parallel computing, High-Throughput Computing, computing clusters and grids

Technical skills:

Full proficiency in C/C++/STL programming
Strong background in Python, SQL, R (including writing packages), regular expressions
Experience with Apache Spark, Linux administration

Experience:

2015 Aug – present: Scientific software engineer at the University of Florida (UF)

Took over an orphan database logging system recording hardware configuration and running conditions and extended its functionality to address new requirements of the hardware upgrade

2010 Oct – 2015 July: Data Scientist at The Ohio State University (formally Ph.D. Research Associate)

Proposed, designed, performed, and published topical physics analyses at LHC CERN [1,2,3,4]

- conducted exploratory analysis, identified relevant classifiers
- designed workflows for processing large datasets on LHC computing grid and clusters
- performed statistical modeling, drawn quantitative conclusions, wrote scientific publications

2004 Jan – 2010 Sep: Ph.D. Research Assistant at the University of Florida

Assisted production and commissioning of the custom-built electronic system [5] for fast online pattern recognition. Prepared and performed physics analyses with the very first data from LHC.

- designed suites for the data quality monitoring [6,7]
- evaluated performance of the system by analyzing data from the dedicated tests
- measured the underlying event model parameters in the new energy regime of the LHC

1998 Sep – 2003 Dec: Research Assistant at Budker Institute of Nuclear Physics

Developed software tools for data analysis at the detector KEDR; here I implemented:

- analysis framework structuring numerous loosely connected Fortran and C/C++ libraries
- machine learning based particle identification (decision trees, neural networks [rus])
- algorithms for energy measurement calibration [rus] (corrected 30% error at the start-up)

Education:

Ph.D. in Physics, University of Florida, Gainesville, USA, 2010 (4.0 GPA)

M.S. (2001) and B.S. (1999) in Physics with honor, Novosibirsk State University, Russia

Coursera certificates in:

Data Science [1,2,3,4], Statistical Inference, Machine Learning [1,2], Algorithms, Bioinformatics [1,2,3,4]