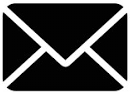
**Yuriy V. Sereda, PhD RESEARCHER & DATA SCIENTIST**

[sereda\_yu@yahoo.com](mailto:sereda_yu@yahoo.com) phone2215-253-4337 locationBloomington, IN, USA [ysereda](file:///C:\Sereda\Job\CV_SeredaY\linkedin.com\in\ysereda\)  [ysereda](file:///C:\Sereda\Job\CV_SeredaY\github.com\ysereda) [location2Yuriy\_Sereda](file:///C:\Sereda\Job\CV_SeredaY\researchgate.net\profile\Yuriy_Sereda)

|  |  |  |
| --- | --- | --- |
| **Education** |  | **Summary** |
| **Springboard**  Data Science 2022 |  | I am a data scientist specializing in classification, regression, optimization, business analytics, natural language processing, and theoretical models. |
| **Tulane University**  Ph.D. Chemistry 2007 |  | **Core Qualifications**   * Twenty-four years of experience transforming data analysis into computational and theoretical models, developing and deploying algorithms. * Managed methodology development projects, helped secure external funding, trained employees in alignment with the organization's objectives, taught and mentored students. * Experienced in publication and presenting scientific data to stake holders, scientists, and the general public in national and international meetings. * **T**rained in machine learning, deep learning, natural language processing. |
| **Institute of Electrophysics**  Ph.D. Physics & Mathematics 2001 |  |
| **Kharkiv University**  M.S. Engineering & Theoretical Physics 1997 |  |

**Skills**

**Programming Languages:** C++, Assembly, Python (Pandas, NumPy, Scipy), R, SQL, Bash, Tcl/Tk.

**Data Collection:** JSON, CSV, API, Web-Scraping.

**Data Visualization:** MatplotLib, Seaborn, Origin, VMD.

**Machine Learning and Deep Learning:** Statistics, Hypothesis Testing, Statsmodels, A/B Testing, Feature Engineering, Supervised Learning, Unsupervised Learning, Linear Regression, Logistic Regression, Classification, Decision Tree, Random Forest, XGBoost, Naive Bayes, kNN, K-Means Clustering, PCA, Time Series Analysis, ARIMA, SARIMAX, Neural Networks, Keras, Tensorflow, Natural Language Processing, Sentiment Analysis, PolyAnalyst, discovering patterns in unstructured texts.

**Analytical Derivations, Numerical Analysis, and Prototyping:** Maple, Mathematica, MathCAD, Sage.

**Modeling:** MatLab; traditional and multiscale Molecular Dynamics (NAMD, GROMACS).

**Industrial Optimization:** Linear Programming, Mixed-Integer LP, Excel QM, Lingo, CPLEX, Gurobi, LKH.

**High Performance Computing:** Linux, Version Control (Git, SVN), job schedulers (Torque, Slurm).

**Theory:** Theoretical Physics, Quantum Mechanics, Mathematics, Linear Algebra, Discrete Math, Biophysics of DNA, proteins, and viruses, Physical Chemistry.

**Data Science Projects**

* **IMAGE RECOGNITION USING NEURAL NETWORKS** Feb 2022

Trained convolutional neural networks (CNNs) for recognizing handwritten images and enhanced the performance of NN models by creating a graphical interface in Tkinter. This GUI allows user to draw an image, see its classification as one of the digits, allows specifying a correct label in case it is misclassified, saves the misclassified image and its label for model re-training, and keeps track of prediction accuracy.

* [**PREDICT CUSTOMER CHURN USING MACHINE LEARNING**](https://github.com/ysereda/Springboard/blob/main/Capstone2_CustomerChurn.ipynb) Nov 2021 - Feb 2022

Cleaned, merged, aggregated and analyzed demographic, transaction, and product usage data of KKBOX music streaming service. Applied machine learning classification methods (Logistic Regression, Decision Tree, Random Forest, and XGBoost) to develop a model for predicting if a customer will continue subscription. Achieved **AUC = 0.94 and accuracy = 93.5%**, which is 6.5% better than typical approaches.

* [**SHORTENING THE LONGEST SUPERPERMUTATION USING TSP FORMULATION**](https://www.kaggle.com/c/santa-2021) Dec 2021 - Jan 2022

Participated in Kaggle Santa-2021 machine learning competition, aimed at minimizing the longest of three strings collectively containing all permutations of 7 symbols, with each string containing a certain mandatory subset of permutations, and two wildcards allowed per string. Mapped to Traveling Salesman routing problem, used optimization OR-Tools and Lin-Kernighan heuristics (LKH) for asymmetric multi-vehicle task.

* **INFERENTIAL STATISTICS OF APPAREL RETAILER TRANSACTIONS** Jan 2022

Formulated and answered questions to understand the clustering of customers and identify promising areas for expanding the range of products sold.

* **IMPROVING PRICING AND ALLOCATION OF INVESTMENT** Jul – Aug 2021

Helped a ski resort get the most out of their facilities: increase the distribution of visitors across the mountain, decrease operating costs, guide on how to adjust their pricing and investment into existing and new facilities.

* [**FORECAST TIME SERIES USING SYMBOLIC COMPUTING**](https://pubs.acs.org/doi/10.1021/acs.jpcb.9b10735) Nov 2015 - Jan 2020

Developed an All-Atom Padé-Ito Molecular Dynamics approach based on a temporal extrapolation of atomic coordinates using stochastic theory and Padé approximants. Three of the four parameters in this non-linear model were optimized analytically, and one parameter numerically. The advantage of this perfectly parallelizable approach over the spatial coarse-graining was demonstrated by forecasting nearly 3 million atomic coordinates in a virus over hundreds of millions of time steps. This is the first practical approach to accelerating the modeling of nanosystems that avoids the introduction of phenomenological equations for coarse-grained descriptors, while providing the exchange of information between space-time scales.

**Employment**

**Indiana University, Department of Chemistry** Bloomington, IN  
Visiting Scientist, **Center for Theoretical & Computational Nanoscience** 2019  
Visiting Assistant Professor, **Nuclear Chemistry Group** 2018  
Postdoctoral Researcher, **Center for Cell and Virus Theory** 2010-2018

* Developed several multi-scale simulation methods that allow up to 20-fold speedup over the traditional molecular dynamics approach to computer-aided design of therapeutic strategies and nanomaterials.
* Developed the Free Energy Basin method that improves the sampling capability of dynamics simulators and thus enables discovery of new states of nanostructures, and deployed it via Deductive Multiscale Simulator.
* Performed statistical analysis and visualization of nuclear chemistry data in ROOT data analysis framework.
* Wrote grant proposals; provided guidance to 16 undergraduate and graduate students during their research.

**Tulane University, Department of Mathematics** New Orleans, LA  
Postdoctoral Fellow, **Center for Computational Science** 2007-2009

* Co-developed a multiscale model of DNA, nucleosomes and chromatin based on the theory of elastic rods and statistical analysis of large-scale datasets (> 500 TB) of simulated atomic coordinates and energy.
* Performed a comprehensive assessment of the available elastic models of DNA, optimized model parameters, and generalized the elastic-rod model by inclusion of electrostatic interactions.
* Designed and implemented the Virtual DNA plugin in VMD software to visualize DNA.
* Developed an ultra-fast global energy optimization algorithm to optimize nucleosome positioning in various DNA base pair sequences and then construct and visualize the spatial model of chromosomal DNA.

**Tulane University**, **Department of Chemistry**Research Assistant, Teaching Assistant 2004-2007

* Analyzed extensive experimental data on quantum tunneling in amorphous solids at ultra-low temperatures and converted into a theoretical model whose predictions enabled calibration of millikelvin thermometers.
* Performed scientific programming in C and Assembly, and symbolic calculations in Maple of nonstandard integrals describing the response of quantum tunneling systems to magnetic fields through Zeeman splitting.
* Contributed to the research of DNA as a molecular wire by developing an iterative solver that allowed finding all ground and excited states of a charge for a tight-binding Hamiltonian describing DNA-solvent interaction.

**Institute of Electrophysics and Radiation Technologies** Kharkiv, Ukraine  
Post-Doctoral Research Associate 2001-2004  
Engineer, Research Assistant 1997-2001

* Developed a novel analytical method for finding unknown and more general exact and approximate solutions of ordinary differential equations with high-order gradient and power nonlinearities, based on the expansion of the squared gradient in powers of the solution and leading to various combinations of elliptic functions.
* Co-implemented a two-level genetic optimization algorithm to optimize both parameters and hyperparameters of non-linear polarization models in the theory of phase transitions.

**Other Experience**

**varsitytutors.com** and **skooli.com**  Online

* Tutored data science, statistics, mathematics, operations research and optimization (linear, mixed integer, and nonlinear programming), computer science, numerical methods, physics. 2018-present