

# Ivan Pogrebnyak

---

Raleigh, NC | (919) 215-5546 | [ivan.pogrebnyak@gmail.com](mailto:ivan.pogrebnyak@gmail.com)

[linkedin.com/in/ivan-pogrebnyak](https://linkedin.com/in/ivan-pogrebnyak)

[github.com/ivankp](https://github.com/ivankp)

## Technical Skills

**PROGRAMMING LANGUAGES:** C++ | C | Python | Go | JavaScript | Java | SQL | HTML | PHP | VHDL

**DATA SCIENCE:** Bayesian and frequentist statistical methods and algorithms | Monte Carlo simulations (PyMC3, Stan) | Linear, Logistic and Non-linear Regression (Scikit-Learn) | Time series analysis, Structural Time Series (TF Probability) | Hypothesis Testing | Deep Learning/Neural Networks (TensorFlow, Keras) | Unsupervised Learning (Clustering, K-Means, PCA) | Data Processing (Pandas, Spark, Apache Arrow) | Data Mining | Data Visualization (Matplotlib, D3.js, Bokeh) | Web development (REST API Development, Web Design, Server Programming) | Distributed Computing (AWS: S3, EC2, Redshift, EMR, Lambda; HTCondor) | Databases (Redshift, Snowflake, SQLite, libpq)

**LINUX:** Version control (Git, GitHub, SVN) | Shell scripting | System management | System programming

## Education

**Data Scientist Certification – Fellowship Program | The Data Incubator | Nov 2021**

**Ph.D. in High Energy Physics | Michigan State University | May 2021**

**B.S. in Physics & B.A. in Mathematics | University of North Carolina at Chapel Hill | May 2013**

## Experience

**REVELIO LABS**

**Jan 2022 – Jan 2023**

**Data Scientist**

- Designed and implemented a regularized naive Bayes model for predicting LinkedIn users' skills.
- Reimplemented a Bayesian structured time series model from Python to C++ with a factor of 1000 speed-up.
- Developed an optimized caching textual data lookup server in Go for fast data delivery.

**MICHIGAN STATE UNIVERSITY, CERN, LARGE HADRON COLLIDER**

**Jun 2013 – Dec 2021**

**Research Assistant**

- Analyzed terabytes of experimental and simulated data to perform state-of-the-art measurements of the Higgs boson cross section in the diphoton decay channel.
- Utilized complex constrained maximum likelihood models for signal extraction and uncertainty analysis.
- Implemented algorithms to optimize model parameters and hyper-parameters using numerical methods.
- Used Monte Carlo techniques for physics simulations and statistical tests.
- Enhanced the efficiency and usability of analysis pipelines.
- Designed an idiomatic, efficient, flexible, and robust framework in C++ for distributed (500 cores) data analysis.
- Refactored data analysis software written in C++ and Python, reducing processing time by factors of 10 to 100.
- Independently learned web development using JavaScript, HTML, CSS, and server programming with PHP and C++. Constructed interactive analysis tools that automated data visualization and statistical model testing and optimization.
- Designed and instituted interactive web-based analysis tools that eliminated the need to manually perform recurring data categorization and significance assessment tasks.

**UNIVERSITY OF NORTH CAROLINA AT CHAPEL HILL**

**May 2010 – Jun 2013**

**Research Assistant**

- Consolidated data from over 50 research works and published a statistical analysis in a peer-reviewed paper.
- Successfully applied non-linear regression models using maximum likelihood and  $\chi^2$  methods.
- Designed and ran particle physics and material interaction Monte Carlo simulations.