Satyarth Mishra Sharma

satyarth.me |✉ mail@satyarth.me | satyarth.mishra@gmail.com | ☎ +7 915 260 3318

Education

University of Cambridge — **Cambridge, UK**
B.A. in Natural Sciences (Experimental and Theoretical Physics) — October 2012 - June 2015

Skolkovo Institute of Science and Technology — **Moscow, Russia**
MSc in Data Science — September 2017 - June 2019

Skolkovo Institute of Science and Technology — **Moscow, Russia**
PhD in Machine Learning — November 2019 - October 2023
Thesis title: *Predicting Phenotype from Genotype with Machine Learning* — Defended December 2024

Work Experience

Raspberry Pi Foundation — **Computer Laboratory, Cambridge, UK**
Undergraduate research student — June - September 2013

Institute of Astronomy — **Cambridge, UK**
Undergraduate research student — June - September 2014

Tsuru Robotics — **Moscow, Russia**
Researcher/developer — June - September 2018

GENXT LTD — **Hinxton, UK**
Machine learning researcher — January 2022 - March 2023

Publications

- E. Nabieva, S. Mishra Sharma, et al. *"Accurate fetal variant calling in the presence of maternal cell contamination."* European Journal of Human Genetics 28, no. 11 (2020): 1615-1623.
- A. Medvedev, S. Mishra Sharma, et al. *"Human genotype-to-phenotype predictions: Boosting accuracy with nonlinear models."* PloS one 17.8 (2022): e0273293.
- D. Kolobkov, S. Mishra Sharma, et al. *"Efficacy of federated learning on genomic data: a study on the UK Biobank and the 1000 Genomes Project."* Frontiers in big data 7 (2024): 1266031.

Skills

Machine Learning in Python
- Machine learning and deep learning libraries (pytorch, scikit-learn, numpy, pandas)
- MLops, experiment tracking (mlflow) and automated hyperparamter optimization (optuna)
- Computer Vision (opencv) and Natural Language Processing (NLTK)
- Large datasets (dask) and Federated Learning (flower)

General Computing
- Environment management and containerization (singularity, docker), running reproducible experiments on large clusters (SLURM)
- DBMS (PostgreSQL, redis)
- Embedded systems and signal processing in C, microcontollers (ESP32, arduino)
- Functional programming in Haskell, Clojure
- Basic webdev (django, flask), server administration and self/cloud hosting
- Exclusively GNU/Linux (I use arch btw)
- GitHub profile: `github.com/satyarth`

Research
- Most experienced in applying deep learning to large genomics datasets
- Effective at visualizing and communicating data (plots for publication, interactive visualizations)
- Interested in the applications of machine learning in modeling complex systems, especially in the fields of ecology and food/water/energy security
- Academic writing in LaTeX
- Your jack of all trades!

Misc
- Fluent in English, Hindi, and Russian
- Check out my other projects at `http://satyarth.me/`