Sheena C Schier

Data Scientist/ Physicist

San Francisco Bay Area

+1 (831) 421-1608

github.com/sschier

sheena.schier@gmail.com

About me

5 years as a member of the ATLAS Experiment international collaboration of 3000+ fellows worldwide, analyzing data from high energy proton-proton collisions at the Large Hadron Collider (LHC). Collaborated in highly-dynamic, international teams on fast-paced research projects with published results. Expert in modeling backgrounds to signal using data-driven techniques. Strong programming background and passionate about big data and advancing in the field of artificial intelligence and machine learning.

Skills

C, C++,

Python

SQL

Adv. Statistical Modeling in C++ (ROOT)

Scikit-Learn, numpy, scipy

matlibplot, seaborn, pandas, jupyter

(*)[The skill scale is from 0 (Fundamental Awareness) to 6 (Expert).]

Experience

2013-2018 Graduate Research — Univ of California, Santa Cruz, CA, and The European Organization for Nuclear Research (CERN), Geneva Switzerland

Advised by Prof. Mike Hance (primary) and Prof. Bruce Schumm

- Searched for new physics in LHC proton-proton collision data using statistical and quantitative analytics in C++/Python/ROOT, leading to two peer-reviewed publications; one of which presented the first new results of its kind since 2001.
- Launched over 200,000 grid jobs on the LHC and local Condor distributed computing infrastructures using bash scripts.
- Produced novel reference datasets using advanced Monte Carlo sampling methods, resulting in the approval for official production of millions of simulated events by the ATLAS Collaboration.
- Proposed, modeled, and productionized new model-based kinematic features of detector responses in order to optimize signal to noise in detection regions previously assumed unreachable by experiments at the LHC.
- Lead developer of object-oriented framework in Python/PyROOT, generating predictive models of backgrounds arising from misidentified events in data; accounting for over 50% of the total background to signal in most of our signal regions.
- Produced 3D visualizations of signal region efficiencies and detector acceptances for publication.

2009-2012 Undergraduate Research — Univ of California, Santa Cruz, CA

- Designed common-mode noise filter board using PADS software; submitted to Sierra Proto Express for fabrication.
- Characterized performance of a 1,024 channel readout ASIC using electronics equipment and specialized data acquisition software.

Data Science Projects

July 2018 Modeling CA High School Graduation Rates — Personal Project

- Wrangled student demographics data using pandas
- Regression pipeline tuned to account for 44% of data variance

Sept 2018 Modeling Employee Attrition w/ IBM synthetic data — Personal Project

- Explored data, engineered features with numpy, seaborn, pandas
- Classification models pipeline tuned to achieve an AUC > 0.87

Skills (cont.)

General	Statistical and quantitative analysis, data visualization, data-driven modeling, data simulation, petabyte data reduction, object-oriented design, technical writing, independent learning, presenting ideas/work to large groups, advanced physics and mathematics
Software	Linux, GNU, bash, git, svn, Microsoft Excel/Powerpoint
Hobbies	Beach volleyball, writing, camping, playing craps

Education

2012 - 2018 Ph.D., M.S. Physics — Univ. of California, Santa Cruz

Dissertation Title: *Searches for Electroweak Production of Compressed Supersymmetry in Events with Soft Leptons, Missing Transverse Momentum, and a Hard Jet in the ATLAS Detector*

2008-2011 B.S. Astrophysics — Univ. of California, Santa Cruz

Thesis Title: *Effect of Channel-To-Channel Variations On Pulse Efficiency And Noise Occupancy For The Use Of KPIX ASIC For Readout Of Silicon μ-Strip Sensors*

Awards

2011 Dean's Award for Undergraduate Research