Sam Stites

Somerville, MA
703.399.1812

sam@stites.io
stites.io

github : stites
linkedin : stites
stackoverflow : stites

RESEARCH AND PUBLICATIONS

Research — CSSR *An Algorithm for Blind Construction of Markov Models from Time Series* *2015-present*

- In the process of writing an upgrade from v0.1.0 to v0.2.0 in collaboration with authors
- Ported the original C++, untouched from 2003, to Scala with prototypes in Haskell and Python
- Researching of automaton, epsilon machines, and monadic structures for a second publication

Publications — Beginning Neo4j, Apress *Technical reviewer* *2015*

- Provided feedback, verification, and fact-checking on the technical content of Neo4j

Research — Low and High Pressure Study of Langmuir-Blodgett Films *NSF grant: PHY-0851599* *Summer 2010*

- A comparative study of the ideal gas law applied to Langmuir-Blodgett monolayer Films of DPPC, TOPO, and colloidal gold nanoparticles coated in Dodecanethiol (DDT)

EXPERIENCE

Sentenai — Machine Learning Engineer *2016 – present*

- Built online bayesian networks (tree-augmented naive bayes) for index prediction of heterogeneous data streams
- Wrote binary compression formats for TAN-bayes nets to run them in our high-performance, real-time systems
- Constructed proofs for various type classes in a Mealy- and Moore- Machine stream transducer library
- Spearheading development of hierarchical policy gradients-based reinforcement learning algorithm for schema prediction of unstructured sensor data. Libraries created are to be open-sourced

Bina Technologies, acquired by Roche Sequencing — Software Engineer *2014 – 2016*

- Built the RDF-compliant server layer in Postgres and Spring for Bina's circulating tumor DNA diagnostics tool
- Designed an aspect-oriented, stream based framework with heavy use of reflection and java 8 features
- Worked with Zookeeper primitives in building out features for our in-house, distributed eventing system
- Implemented custom visualizations and sequencing quality reports in d3 and Ramda.js
- Architected and programmed the UI-layer of Bina's flagship product, RAVE, in Angular with Ramda.js + Typescript

Keychain Logistics — Data Engineering Consultant *2014*

- Researched and developed a pricing model using Random Forest regression using scikit-learn
- Built an algorithm agnostic framework for regressors to automate optimizing and benchmarking models
- Deployed production Flask-server and built out API for a clean user experience on Digital Ocean

Hack Reactor — Full Stack Instructor and Mentor *2014*

- Guided students through a dedicated, 90hr+ per week curriculum in Angular, Backbone, d3, and Node
- Code-reviewed 110+ repositories to evaluating technical ability and provide feedback to Hack Reactor

TurboSquid — Data Scientist in Credit Fraud *2012 – 2013*

- Lead a two-man team to build a Random Forest credit fraud detection classifier with scikit-learn which alleviates the processing of $1.2M per year at 98% accuracy and saves explicit overhead of $40k/year
- Restructured search and categorized assets to improve the user experience of customers on the site
- Wrote unsupervised learning algorithms for market segmentation with k-means clustering and DBSCAN

Physics Department, Vassar College — Academic Intern *2011 – 2012*

- Initiated and coordinated dialogue between students and faculty members, adding efficiency to the tutoring structure. Events in 2011 doubled the participation in Physics committee meetings
- Taught a variety of physics topics from introductory- to advanced- level material

Physics Research Experience for Undergraduates, NSF — Condensed Matter Physics Researcher *Summer 2010*

- Conducted independent research on graduate level material, which required fast-pace comprehension
- Presented a final product of research in a campus-wide lecture, attended by researchers from varying fields

INVITED TALKS

Life After Vassar (LAVA) — Vassar Mathematics and Statistics Department *2017*

Life After Vassar is a series of talks to introduce Math majors to possible opportunities after graduation

Colloquium Series — Vassar Mathematics and Statistics Department *2017*

An ongoing series of research talks. I will be speaking about my work on CSSR

PROJECTS

Raft *(In progress) An implementation of the raft consensus algorithm in Haskell* *2016-present*

- a Haskell implementation of the Raft Consensus Algorithm with experimental branch using the Free Monad

KafkaSphere *An Apache Kafka monitor for DemandCube's open source data pipeline.* *2014*

- Wrote the Sparkjava server with Kafka endpoints for both Producers and multithreaded Consumers
- Developed client-side Angular/d3 frontend to display information on the Consumer Groups and Topics

Satellite.io *The first browser-based, multiplayer, WebGL space warfare game optimized for the Oculus Rift* *2013*

- Spearheaded, project managed, and developed code for all layers of the stack, guiding team
- Wrote a physics-based network syncing system, reducing network calls by 80% while keeping high user experience
- Set up an asynchronous Redis API maintaining complex queries using Gang of Four patterns

---

EDUCATION

Vassar College *BA in Mathematics, BA in Physics.* *2008 – 2012*

Coursework included programming languages: Java, C++, Scheme, Racket, SML, R

Relevant math work: applied linear algebra, abstract algebra, applied statistics and statistical theory

---

Semi-professional juggler, Eagle Scout, Venture for America fellow, improv comedy performer, Mini-Maker Faire organizer