Todd Perry

Email: todd.perry@myport.ac.uk • Phone: +447508313890 • http://toddperry.ninja
Linkedin: https://www.linkedin.com/in/toddperry93 • GitHub: https://github.com/djentleman
Flat 19 Sebright House, London, UK, E2 9AD

Summary

Software Engineer based in London. Passionate about engineering and technology, pursuing a career in machine learning and data analytics.

Employment History

Huq Industries — http://huq.io

Senior Software Engineer — *Nov '16 – Current, London, UK*

Senior Software Engineer at a market intelligence company based in London - focused on generating rich location based consumer behavior datasets collected via a mobile SDK installed on over 45 million devices world wide. Notable work included:

- Rebuilding a monolithic data processing pipeline as a serverless architecture using **AWS Lambda**.
- Building various machine learning models with **Sklearn** and **Keras** for consumer behavior analytics.
- Designing and implementing an architecture to process upwards of 1 billion datapoints a day.
- Utilizing **Ansible** and **Docker** to both automatically provision infrastructure and deploy code.
- Developing both customer facing dashboards and data processing pipelines using **Python**.
- Analysing over 30 billion rows of geospacial data in **BigQuery** using **GIS**.
- Writing **Logstash** preprocessing scripts in **Ruby** before loading into **Elasticsearch**.
- Implementing a Genetic Algorithm to optimize the quality of the data submitted by the mobile SDK.
- Using **Apache Beam** and **Google Dataflow** to process massive amounts of data.

Lockheed Martin — http://www.lockheedmartin.co.uk/uk.html

Software Engineer — *Sept '15 – Nov '16, Farnborough, UK*

Worked on a number of Internal R&D projects as well as intelligence platform being sold to a UK government customer. Responsibilities included:

- Using **Apache Storm** and **Apache Spark** to build real-time analytics pipelines.
- Refactoring monolithic systems into micro-service architectures.
- Utilizing **Docker** and **Docker Swarm** to provide modular, containerized architectures.
- Writing a responsive web UI using **AngularJS** and **Material Design**.
- Building machie learning models to solve problems in transport and cyber monitoring.
- Collaborating with other LM R&D efforts in the US.
- Representing the company at various conferences and events.
- Working with customers to improve system UX.

While at Lockheed I was also contracted as a Web Developer at a customer site, developing a system that monitors the state of global airspace. The system had an established (**MEAN stack**) codebase.

Universitá Di Salerno — http://neuronelab.unisa.it/

Research Intern (Erasmus+ Work Placement) — *Jun '15 – Aug '15, Fisciano, Italy*

3 Month Erasmus+ placement at the University Of Salerno in Italy, working with researchers on a number of bioinformatics problems. Work included:

- Using Gene-Expression and Protein Expression datasets to search for breast cancer biomarkers.
- Visualizing high dimensional data using a combination of Clustering algorithms, **PCA** and **MDS**.
- Applying computational intelligence techniques to the Bioinformatics Data.
- Writing scripts to preprocess the Gene Expression data in **R**.
- Clustering similar Genes using **KMeans**, **DBSCAN** and **Hierarchical Clustering**.

Seagate (Formerly Xyratex) — https://www.seagate.com/

Product Engineer (Industrial Placement) — *Jul '13 – Sep '14, Havant, UK*

Industrial placement via University, part of a team of around 6 engineers working in manufacturing test development. Responsibilities included:

- Analysing manufacturing data to calculate and predict yield impacts.
- Capturing and analysing SAS and SATA protocol traces.
- Debugging faulty storage products, writing scripts to trigger hardware failures.
- Developing software tools to be used by the product engineering team.

Education

University Of Portsmouth — Portsmouth, UK

BSc (Hons) Computer Science w/ Year In Industry — *2011 - 2015*

First Class: 93%

Final Year Project:

My final year project was focused on failure prediction of Seagate storage systems using classification. Using around 18 months (over 40000 systems) of manufacturing test data, the goal was to predict if a given system would fail within the warranty period after being shipped to a customer. After investigating the effectiveness of different classifiers on the data, I implemented a cost-sensitive Random Forest in **Java** (**Weka**), and used a genetic algorithm to optimize it's performance. Using this I could predict over 50% of storage controller failures up to 2 years in advance, with a precision of 73%. The results were published at the IEEE CEC in 2015.

Course Contents:

- Data Structures & Algorithms
- Database design and development
- Neural Networks, Genetic Algorithms, Data Warehousing & Data Mining
- Parallel Programming In **Java** and **C**
- Programming in **Python**, **Java** and **Haskell**
- Software Engineering Methodologies & Tools
- Theoretical Computer Science & Discrete Mathematics
- Web application design and development

Brockenhurst College — Brockenhurst, UK

A - Levels — *2009 – 2011*

A-Level Mathematics: A • A-Level Physics: B • A-Level Computing: B

Applemore Technology College — Southampton, UK

GCSEs — *2004 – 2009*

9 A-C, including English (B), Maths (A) and Science (A)

Skills

Programming & Software Skills: Python (Numpy, Scipy, Sklearn, Pandas, Keras, Django, Flask) • Java (Spark, Storm, Weka) • Ruby • Javascript (Node) • Bash • C • R • SQL • MongoDB • Docker • Vagrant • Ansible • Git • Jira • Bamboo • MQTT • Kafka • ELK Stack • GIS • LaTeX

Languages: English: Native Proficiency • Japanese: JLPT N2 • Russian: Elementary

Technical Specialities: Strong knowledge of machine learning and evolutionary computation, experience applying computational intelligence algorithms to problems in both industry and academia • Experiance designing and developing systems on AWS (EC2, Elasticache, Lambda, RDS/Aurora, Route53) & GCP (Dataflow, BigQuery) • Experience using various databases (Postgresql, MySQL, MongoDB, Redis, Cassandra) • Experience provisioning and using virtualized environments, and working as part of an agile team • Extensive knowledge of machine learning & statistical analysis algorithms, such as decision trees, Bayesian networks, neural networks, DBSCAN, hierarchical clustering, GAs, PSO, PCA, RBMs, autoencoders, etc... • Development experience using Windows, Linux (Ubuntu, CentOS) and OSX systems

---

Publications

Biased random forest for dealing with the class imbalance problem Theory — IEEE Neural Networks & Learning Systems (Nov' 18)
w/ Eleman Teitei & Mohamed Bader

Mathematical Function Optimization Using A Novel Algorithm Based On Newtonian Field Theory — IEEE WCCI ('Jul 16)
w/ Mohamed Bader

Imbalanced Classification Using Genetically Optimized Random Forests — ACM GECCO (Jul '15)
w/ Mohamed Bader

Imbalanced Classification Using Genetically Optimized Cost Sensitive Classifiers — IEEE CEC (May '15)
w/ Mohamed Bader & Steve Cooper

Hobbies & Interests

Airsofting • Cycling • Electronics • Machine Learning • Travelling • Languages • Computing

---

References Can Be Given On Request.