Greg Lever | Ph.D. (Cantab.)

82 Martin Way, Morden, London SM4 4AG

📱 +44 (0) 7814 676 385 • ✉ greglever@cantab.net • 🌐 greglever.github.io
in levergreg • ⓞ greglever

I am a Software and Data professional with extensive Python, Java, SQL and applied Machine Learning experience. I gained my PhD in Computational Enzymology at the Cavendish Laboratory, University of Cambridge, where my thesis was nominated as outstanding and later published by Springer. I then worked as an independent Postdoctoral Research Associate at MIT before transitioning into commercial roles in startups building Machine Learning products and am now looking to move back into a scientifically-orientated role. I am passionate about continual and lifelong learning.

Current Position

Data Scientist (Data Science Developer), Arkera - `http://arkera.ai` **Westminster, London**

A fintech startup at the intersection of Data Science and Wealth Management (established March 2015) Dec 2015 – present

- o I was Arkera's first Data hire and was instrumental in building the Data team, adding an additional three members.
- o Python (Flask/SQLAlchemy) and SQL (complex queries for PostgreSQL, migrated from MySQL) for our backend and web scraping, along with the design of our database schema and dimensional data modelling.
- o I am a strong advocate of Test-Driven Development and implemented the company's unit and functional testing framework.
- o I played a key role in moving the Data team to an agile development process (stand-ups, sprint planning, code reviews, retrospectives, continuous integration) initially through Trello and then to JIRA.
- o I design, prototype, implement and maintain our proprietary Machine Learning (Scikit-Learn/NumPy/SciPy) and Natural Language Processing algorithms, along with ETL processes in production and our data models.
- o Organised the code base into an ecosystem of microservices providing HTTP endpoints for our iOS and web applications.
- o RESTful APIs, Amazon SQS event messaging, Docker and continuous integration (Jenkins) and exposure to iOS for the app.

Computation and Modelling Skills

Languages: *Strong in* Python (Scikit-Learn/Pandas/NumPy/SciPy/Matplotlib/Natural Language Toolkit/Flask/ SQLAlchemy/pytest/unittest/pdb/cython), SQL (PostgreSQL/MySQL/BigQuery), Java (Spring), Linux (BASH scripting along with awk, sed and grep), AWS, Docker, Git (I encourage rebasing and gladly help others with merge conflicts), NoSQL (MongoDB) and LATEX typesetting. *Confident with* R, Javascript, CSS, HTML, Mathematica and the Google App Engine. *Experience of* Neo4j, Cypher Query, Perl, Fortran, C++ and Matlab.

Operating Systems: Strong capabilities in UNIX/Linux and High Performance Computing clusters, Mac OS X, and Windows-based environments.

Certifications: Medicinal Chemistry (EdX), R Programming (Coursera) and Data Scientist's Toolbox (Coursera).

Experience

Senior Data Scientist, Stylect **WeWork, Liverpool St**

An e-commerce startup with an app and website designed to enable women to find their perfect shoes April 2015 – Dec 2015

- o I headed the Data Science efforts at Stylect, utilising my skills in Python, SQL and Machine Learning.
- o I led the design, engineering and deployment of production-worthy Machine Learning architectures for Recommender Systems to improve the user experience and ultimately increase sales.
- o Involved extensive Python development (web scraping, pytest) with Pandas (including NumPy, SciPy and Matplotlib) and Scikit-Learn along with maintaining and adding new HTTP endpoints serving web, iOS and Android applications.
- o Required SQL to use services from the Google Cloud Platform including Big Query and the App Engine.
- o Utilised Natural Language Processing (Natural Language Toolkit) to perform Feature Engineering for Machine Learning (Scikit-Learn) approaches for classifying retail products.
- o Devised, implemented and interpreted A/B tests to assess the accuracy of the predictive models.

Data Science Fellow, The ASI **London**

The ASI Fellowship enables academic scientists to become data scientists and engineers February 2015 – April 2015

- o Built a prototype recommender system for an e-commerce startup (Stylect) to improve their user experience and increase sales.
- o Developed strategies for growth hacking and user retention for a community-building startup (Quiet Riots).
- o Expert-led training and practical workshops in Hadoop, Spark and MapReduce as well as Statistics and Business-specific concepts.

Postdoctoral Associate, MIT and University of Cambridge **Cambridge, MA (USA)**

MIT Department of Chemical Engineering and Cavendish Laboratory March 2014 – February 2015

- o Further developed and galvanised existing skills in Python gained from my PhD.
- o Built upon my knowledge of R from my MSci, learning new concepts and tools for analysing results of simulations.
- o Mentoring of graduate students, taught undergraduates to use linux and High Performance Computing for the very first time.
- o Independent exploratory analysis using Python/Pandas (including NumPy, SciPy and Matplotlib) of an existing body of data.
- o Collaborated closely with teams based in the UK (Cambridge, Bristol and Southampton) and the USA (Yale).
- o The work is currently being prepared to submit for publication in the *American Chemical Society* journal *Catalysis*.

Education

PhD, University of Cambridge   **Cambridge**

Computational Enzymology, Cavendish Laboratory   *September 2010 – March 2014*

- Long-term independent research and analysis from self-driven projects that delivered scientific results and expanded my field.
- Applied mathematical theories to complex real-world problems and generated new methodologies to improve existing approaches.
- Extensive Python (Pandas, NumPy, SciPy) development for data analysis of computational chemistry simulations.
- Immersed in a High Performance Computing Environment utilising linux, BASH, perl, awk, sed and grep.
- Winner of thesis prize "Recognising Outstanding Research", *Springer* subsequently bought my intellectual property from me.
- Published first-name author papers in the *Journal of Chemical Physics* and the *Journal of Physics: Condensed Matter*.

MSci (First Class Honours), University College London   **London**

Theoretical Physics   *September 2006 – July 2010*

- Through summer projects in 2009 and '10, extended self-taught Linux and Python programming knowledge through real research.
- Contributed to Bayesian Inference software using R and Java and developed extensions in C++.
- Resulted in a publication in *Monthly Notices of the Royal Astronomical Society*, summer work published in *Physical Review B*.
- Courses Highlight: Java and C++, Mathematica, Advanced Quantum Theory, Stochastic Processes, Group Project.

Extracurricular Activities

Supervisor for MIT Undergraduate Research Opportunities Program (UROP)   **Cambridge, MA (USA)**

MIT Department of Chemical Engineering   *July 2014 to Oct. 2014*

Provided modelling & simulation training for a candidate who had never previously run any simulations or used the linux terminal.

Co-supervised Cambridge MPhil Student in Scientific Computing   **Cambridge**

Cavendish Laboratory, University of Cambridge   *October 2012 to Oct. 2013*

Provided electronic structure expertise and biological expertise, to complement the MPhil candidate's additional co-supervisor who provided physics and analytical mathematical modelling expertise. Helped the candidate to prepare for PhD applications.

Physics At Work - Outreach Event for the Cavendish Laboratory   **Cambridge**

Cavendish Laboratory, University of Cambridge   *September 2011/'12/'13*

Organised exhibitors for and participated in a three day outreach exhibition in Cambridge aimed at inspiring 14-17 year olds to become the next generation of scientists, engineers and technology specialists by showcasing the many and varied ways computational physics is used in the real world.

Procter & Gamble R&D European PhD Seminar 2013   **Brussels**

Procter & Gamble (P&G) Brussels Innovation Centre   *April 2013*

- Applied technical knowledge to solve real problems found in the typical day of a Scientist or Engineer in R&D at P&G.
- Gained insight into daily business challenges, including how to make difficult and tough decisions, set priorities and communicate.
- Over the week, led a team of PhD students, practising the skills required to motivate people, via exercises and case studies.

i-Teams: Commercialising Creativity   **Cambridge**

Finding real-world applications for a synthetic reversible molecular lego system   *October 2011 – December 2011*

- Led a team of 6 PhD students and 1 MBA student asked to investigate an invention that came out of a Cambridge laboratory.
- Furthered our ideas for real-world applications by contacting external industry scientists and harnessing their expertise.
- Understood the practical processes needed to turn a new technology into a commercially-viable product.
- Learned how to work with sensitive intellectual property and to bring it to the attention of an existing market.

Business Challenge hosted by Cambridge University Technology and Enterprise Club   **Cambridge**

Building a Business Case for more in silico *modelling in the Pharmaceutical Industry*   *March 2011*

- Led a team of three fellow PhD students working in close collaboration with the Director of Clinical Pharmacology at Medimmune.
- Had to think about real-world research within a business context
- Gained a firm appreciation for what is expected from data-driven scientists within commercial organisations

Researcher, London Centre for Nanotechnology   **London**

Successfully applied for funding from the Engineering and Physical Sciences Research Council   *June 2010 – September 2010*

Researcher, University College London   **London**

Successful in funding competition for all department undergraduates   *June 2009 – September 2009*

References: Prof. Mike Payne (mcp1@cam.ac.uk), Dr. Danny Cole (daniel.cole@ncl.ac.uk)

Greg Lever | Ph.D. (Cantab.)

82 Martin Way, Morden, London SM4 4AG

 +44 (0) 7814 676 385 •   greglever@cantab.net •   greglever.github.io

 levergreg •   greglever

To whom it may concern,

I am a Software and Data professional with extensive Python, Java, SQL and applied Machine Learning experience (5+ years academic/scientific and 2 years commercial). Following my first-class honours MSci in Theoretical Physics from UCL I gained my PhD in Computational Enzymology at the Cavendish Laboratory [1], University of Cambridge, where my thesis [2] was nominated as outstanding by the University and later published by Springer [3], now with over 1600 combined chapter downloads. My first-name author publications concerning Computational Enzymology [4] and Protein Modelling [5] including my previous collaborations on Computational Predictions in Silicon Materials Physics [6] and Computational Predictions of Extrasolar Planets [7] in peer-reviewed journals have received 50 citations so far.

After leaving Cambridge I worked as an independent Postdoctoral Research Associate at MIT [8] before transitioning into commercial roles [9] within technology startups where I have been building Machine Learning products. I have also spoken at commercial conferences such as Big Data Week London where I hosted a workshop on How to Become an Effective Data Scientist [10].

I am now looking to move into a more scientifically-orientated role. I am passionate about continual and lifelong learning and always seek to develop and improve my professional skills, where examples of online courses I have taken include R Programming (Coursera), Medicinal Chemistry (edX) and Data Scientist's Toolbox (Coursera). More recently I have been pursuing the Algorithms for DNA Sequencing course from Coursera allowing me to stay up to date with Whole Genome Sequencing data manipulation and analysis techniques.

I am currently working as a Data Science Developer at a FinTech start-up named Arkera where I was their first Data Science hire. Within the team we always try our best to adhere to good SOLID object-oriented practices including YAGNI and loose-coupling but also understand when it can be appropriate to add 'good' technical debt into a project, in order to speed up time to production, with the knowledge and understanding of how to remove it (ideally in the next sprint !). This then aligns with our policies of iterative improvement through continual refactoring.

I was instrumental in adding three additional members to the Data Team and we now stay in close communication between the team and the other members of the company via Slack, JIRA and Confluence. At Arkera I play a key role in the team responsible for the backend and web scraping as well as designing, prototyping and implementing proprietary Machine Learning and Natural Language Processing algorithms. For the server code we predominantly use Python (numpy, scipy, scikit-learn, pdb, nosetest, unittest, flask, sqlalchemy) talking to a PostgreSQL database where I had hands-on experience with the schema design and dimensional data modelling and am writing efficient SQL statements day-to-day. We host it on GitHub repositories where we practice Git Flow and use Jenkins for continuous integration. We run on AWS with Docker (on Linux with a nice combination of bash, ssh, sed and awk) to serve iOS and web applications. I play a mainly supportive role in the iOS application as my knowledge of Swift is not as great as other languages but is continually improving. The web application has a CSS/HTML/JavaScript and Angular2 frontend where I again provide a supportive role.

The Python backend code base is organised into an ecosystem of microservices so I'm happy writing HTTP endpoints providing efficient and authenticated communication while developing a secure linked data platform. At Arkera I have performed integrations with third-party API providers and created and now maintain in-house tools for our Content Team (in Javascript). I introduced code review at Arkera, ensuring the junior developers had an open forum within which to improve their skills. I implemented Arkera's first set of coding standards and documentation and integrated Test-Driven Development within the team, ensuring all code deployed into production has good test coverage. I am a strong advocate of Test Driven Development (although understand where it can be more or less appropriate) and in my current role I was responsible for implementing Agile development processes within the team such as stand-ups, sprint planning, code reviews, retrospectives and continuous integration.

I was previously the first Data hire at Stylect where I built their first recommender system for their app and website, utilising my expertise in Machine Learning and integrating ML capabilities within the backend code, hosted on BitBucket. At Stylect I was responsible for creating and maintaining a unified and coherent set of APIs (with OAuth2

authentication) that served 100,000 visits each month on their e-commerce website and also served an app which had one million downloads. To scale to that size we used Google BigQuery and MongoDB and I would enjoy improving my skills in NoSQL technologies in the future.

As mentioned above, before transitioning into Stylect via a Data Science & Business Analytics Fellowship at The ASI where I had the opportunity to learn neo4j and apply the techniques for a startup. Before the ASI I was a Postdoctoral Research Associate at MIT. Part of my role as a post doc was providing documentation and tutorials to more junior members of the research group. Part of this included the integration and loading of public biological databases into the workflow of the group from sources such as the National Center for Biotechnology Information for Next Generation Sequence data and Protein Data Bank. In addition I generated educational material for delivering lectures to students taking the graduate computational chemistry course provided by the Chemical Engineering Department.

Since leaving academia I have been working at technology startups officially as a Data Scientist, but with a very hands-on development role productionising software. During this time I have come to realise that I really enjoy building tangible applications with real-world use. As such I am looking to build my career within a development-focused role.

Yours sincerely,

Greg Lever

Links

[1] http://www.tcm.phy.cam.ac.uk/~gl319/.

[2] https://www.repository.cam.ac.uk/handle/1810/246261?

[3] http://www.springer.com/gp/book/9783319193502.

[4] https://goo.gl/KrI2qW.

[5] https://goo.gl/IpWH5G.

[6] https://goo.gl/KYO9BT.

[7] https://goo.gl/qmDhTV https://goo.gl/w8Pglt.

[8] https://engineeringphds.mit.edu/profile/greg-lever?

[9] http://www.asidatascience.com/fellowship/alumni/.

[10] http://london.bigdataweek.com/speaker/greg-lever/
https://www.youtube.com/watch?v=-8-WglavoIw?
http://blog.bigdataweek.com/2015/11/04/become-effective-data-scientist-answers-bests/?