



Le Nguyen The Dat

DATA SCIENCE AND ENGINEERING AT UBER

KAGGLE MASTER

🌐 <http://lenguyenthedat.com>

✉ resume@lenguyenthedat.com

🏠 Tanah Merah, Singapore



WORK EXPERIENCE

Uber

Data Scientist

May 2017 - Present

Utilize data science and engineering to drive marketing efficiency in APACx

Recruit and mentor a team of 4 data analysts and engineers

In charge of APACx Engucation (Uber's Engineering Onboarding and Education) for Data Analysts and Scientists

In spare time, I help colleagues around the world with data pipelines / python / sql / data science / data engineering related challenges

Technologies: Python, SQL, Docker, Uber's Piper (a fork from Airbnb's Airflow), Hive, Presto, Vertica, and so on

Honestbee

Data Science and Engineering Lead

Jan 2016 - May 2017

Recruit, mentor, and [lead a team of 3 data engineers, 2 data scientists, 1 product manager](#) to build [production-grade data warehouse ecosystem, data applications](#) such as fraud detection systems, predictive models, [dynamic time slot pricing](#), and recommendation systems from the ground up

Oversee honestbee's data-related challenges from the technical point of view

In spare time, I help taking care of honestbee's cyber security matters, as well as internal employees access policies to various systems and data sources

Technologies: Amazon Redshift, Amazon RDS and other Amazon Web Services, Python, Scikit-learn, XGBoost, Docker, Airbnb's Airflow, Periscope, and so on.

Commercialize TV

Senior Data Technologist

Mar 2015 - Dec 2015

One-man team: full-stack data technologist - engineer, [from DevOps with AWS distributed cloud infrastructure and Docker to programatic ETL and data visualization](#)

Build data warehouse, analytics and data science infrastructure from the ground up

Integrate and analyze data from Google Analytics, Google Ads (DFP, AdEx, AdSense), AOL, Youtube, DailyMotion, Baidu, TenCent, and various other sources

Technologies: Amazon Redshift, Amazon RDS and other Amazon Web Services, Python, Docker, and Tableau.

Zalora Group

Head of Data Warehouse & Data Science

Jun 2013 - Dec 2014

May 2014: Head of Data Warehouse & Data Science

Build and manage the data science and data engineering team

Manage the company's data warehouse, data marts, and data science infrastructure.

June 2013: Data Scientist

Zalora's first Data Scientist. One-man engineer team who designed, built, and was known for:

- CRM data platform, fully managed and integrated with Oracle Responsys
- [Product recommendation engines](#) for Zalora Website and Personalized Email Marketing Campaigns
- Zalora Data Warehouse infrastructure - hosted on Amazon Web Service: from storage, ETL to data management
- Various internal decision support tools, automated reports, data marts, that were deployed in Rocket Internet ventures globally

A-STAR Data Storage Institute

Research Engineer

Jul 2011 - Jun 2013

Conduct researches, design algorithms and develop full system for multiple projects:

- Advanced Smart Caching logics dedicated for travel booking industry. Fully implemented, integrated, and deployed, potentially save more than 5% of A-STAR's Business Partner's annual cost
- [Network Virtualization](#) and [Software Defined Network with OpenFlow](#). Develop and apply new algorithms for efficient traffic routing, load balancing, virtual machine placement techniques in scalable data center network
- Knowledge based data classification and recommendation system, built on top of Apache Mahout's collaborative filtering engine. Implemented functional and scalable Java / Python prototypes for travel industry with social network's data set



EDUCATION

University of Oxford, United Kingdom

Computer Science

Master of Science

Sep 2009 - Oct 2010

Focus on Data Mining, Automata Theory, and Functional Programming Language

University of Nottingham, United Kingdom

Computer Science

Bachelor of Science

Sep 2007 - Jul 2009

Focus on Functional Programming Language, Mathematics, Algorithms and Data Structures



AWARDS

Kaggle Inc

Kaggle Master - [More Info](#)

Feb 2016

Dextra: Rakuten Viki Challenge

Finalist - 4th Prize (\$600) - [More Info](#)

Sep 2015

Dextra: Singapore Ministry of Defense Data Analytics Challenge

Winner - 1st Prize (\$5000) - [More Info](#)

Aug 2015



PUBLICATIONS

Data Analytics Blogs

Towards Data Science - [More Info](#)

1 Jan 2018

Kaggle Ensemble Guide

KDnuggets / MLWave - [More Info](#)

11 Jun 2015

Virtual machine placement with two-path traffic routing for reduced congestion in data center networks

Computer Communications Vol.53 - [More Info](#)

Software defined network based adaptive routing for data replication in Data Centers

IEEE International Conference on Networks, ICON 2013, Singapore - [More Info](#)

11 Dec 2013



SKILLS

Programming

- Python
- Shell Scripts
- SQL
- Haskell
- Java
- R

Data Science

- Recommendation System
- Predictive Modeling
- Scikit-Learn - XGBoost - Neural Networks

Others

- Amazon Web Services
- Data Infrastructure
- Distributed System
- Docker
- Agile Development



LANGUAGES

English

Full professional proficiency

Vietnamese

Native



INTERESTS

Photography

Hackathons

Poker and eSports



REFERENCES

Dat was our first hire into Data Science. His interview task code set the tone: he first submitted solutions in Haskell, then taught himself Python in 3 days and resubmitted in Python.

He was incredibly productive from the outset. In 2 hours, he taught himself Amazon Redshift then built a data warehouse which allowed the email team to integrate their mailing solution provider seamlessly in a matter of days. He still manages this data warehouse on his own, in his spare time; rebranded Redcat, it feeds all the products built by Data Science as well as other analysts in the company and contains a copy of practically all the company's data, all the way to customer service call recordings.

He also wrote the recommendation algorithms currently running on the site, once again having not done this kind of work before. I learnt a lot just from listening to him explaining his R&D approach for that project.

Working with Dat is always the same experience: you give him a high level spec, sit back, and the product comes back quickly, showing an understanding of your actual problem instead of the one you thought you had. I've also noticed him fixing import issues due to third parties (since his own code is rock solid) at 3 in the morning, and turning up cheerfully the next day as if nothing had happened. There's evidence of his work in all departments from Marketing to Buying, and everybody agrees that he has become a key part of the company's success and growth. It is not often that you can say that you feel privileged to have worked with somebody.

It was most definitely a real, rare honour to have had the chance to work with Dat, and I always saw our relationship less as employee/employer and more as having the chance to rent extraordinary talent at cut throat pricing and getting a fantastic return on investment. Even if he studied at the Other Place.

Dat set up and managed the AWS Redshift DWH at Zalora, which allowed my team to restructure and put all reporting and planning processes on a much better foundation.

He cares a lot about his work, products, and technology and possesses a high level of intellectual curiosity. Deeply committed, Dat will get things done on time, no matter what.

Working with him on the DWH was one of the most enjoyable and educational projects for me at Zalora. Dat will be a great addition to any Data Team or the guy you want to hire to set one up.

— Christian Eik, Associate Director - Buying Budgeting & Data Analytics, Zalora Group

At honestbee, Dat started the Data Science chapter as a one man show, and has since then developed it into an extremely efficient team that supports all departments across our company.

As a team leader and as an individual contributor, Dat has always struck me as an exceptional performer. He is known for his speed of execution and for always being proactive on all fronts.

In many situations, Dat also impressed me by his capacity to grasp the business issues at stake and to come up with solutions to address them. He is one of the most brilliant guys I have had the opportunity to work with.

— François Picard, AVP Operations - Groceries Marketplace, honestbee

Dat is one of the folks I will always trust to watch my back. He was hired as a lead for the Data Science team but grew to be so much more. He is a champion of shipping and a champion of delivering the best work possible. Always hungry to do more, in a smarter way, Dat constantly pushes all of us to be accountable to each other and have a sense of ownership in the work we do.

In another life, Dat would have been a most excellent librarian. He took it upon himself to constantly organize our Confluence spaces, ensuring that knowledge and information is constantly shared and available. His fastidious and disciplined nature is one that I highly respect.

— Isaac Tay, VP of People & Special Projects, co-founder, honestbee

Dat is the perfect role model for what it means to take ownership. At honestbee, Dat built the data team and infrastructure from scratch, taking us from zero to being best in class. Dat leads by example and with firmness. He adopts a direct, no-nonsense approach and is results driven. Colleagues that work with Dat have a deep respect for his work ethics and abilities.

Dat acts selflessly in the best interest for the organisation and treats the company as if it was his own. He would go above and beyond his scope to contribute in any other ways that he found possible to. He is deeply thoughtful and regularly provides constructive feedback both downwards and upwards.

What Dat has accomplished at honestbee is nothing short of exceptional.

— Jonathan Low, VP of Engineering, co-founder, honestbee

Dat was a brilliant colleague to work with. His understanding of machine learning and data engineering was top notch. What surprised me the most is that he is humble and willing to teach people more junior than him in any capacity. Definitely would hope to work with him again in the future.

— Aaron Mak, Data Analyst, Uber

Dat is one of the data scientists in Uber that I respect greatly. Throughout my time with Uber, Dat has constantly provided coaching sessions in data engineering and data science. Some of these sessions include coding best practices, data science flow, git 101, time series prediction, and the list goes on. With his passion to coach and contribute to the community, Dat is certainly someone that will bring immense value to any organisation as a data science manager. I highly recommend Dat.

— Walter Ngaw, Data Analyst, Uber