Dr David **Voong**

voong.david@gmail.com
www.davidvoong.co.uk
Telephone: +447795990807

PROFILE

DATA SCIENCE AND DATA DRIVEN APPLICATIONS

I am a data scientist with a strong background in statistical and computational data analysis techniques and their use within distributed systems technologies. Business oriented, with excellent communication, leadership, mentoring skills and extensive expereience across multiple disciplines.

I am looking for a challenging and rewarding work environment with a strong ethic in team work and the opportunity to help build and support a team in their long term development.

SKILLS

DATA ANALYSIS

Statistics: *forecasting* and *anomaly Detection* models, *hypothesis and AB testing*, *Monte Carlo simulation modelling*, *linear and logistic Regression*, *error propagation*, *probability distributions*, *bayesian statistics*, *confidence intervals*, *parameter Estimation & model Fitting*, *goodness of fit tests*, *ANOVA*, etc.
Data Analysis Frameworks: *NumPy, SciPy, scikit-learn, pandas* and *ROOT*.
Databases: *Amazon Redshift, MySQL, PostgreSQL, SQLite* and *MongoDB*.
Visualisations: *matplotlib, d3.js & Bokeh*.

ENGINEERING AND BIG DATA

Programming: 8+ years of *Python* and *C++* object orientated programming, *Jupyter Notebooks*, *TDD*, *git* and *JIRA*.
Big Data Architectures and Cloud Platforms: *AWS Redshift, S3, Kinesis. DigitalOcean, Bluemix, Microsoft Azure, DIRAC, Hadoop, Spark, PySpark*.
DevOps: *AWS Elastic Beanstalk, Linux Fabric, Ansible* and *Docker* deployment tools and *Airflow* job orchestration.

WEB APPLICATIONS

Frontend:*HTML, Javascript, JQuery, d3, CSS* and *Bootstrap*. **Backend**: *Django* and *Flask* Frameworks, *Apache HTTP Server* and *gunicorn* WSGI. **Testing**: *sinon, QUnit* and *Selenium*. **Other**: *Jupyter Web Apps*.

EMPLOYMENT HISTORY

PRODUCT MADNESS Senior Data Scientist

January 2016 - Present | London UK
Product Madness is a digital social gaming company which produces some of the highest grossing apps across multiple platforms. They are a tech company with a strong philosophy of data driven decision making. The data analytics is built on high frequency, high volume and highly granular data with millions of daily users logging hundreds of millions of events each day. My role here is to provide data insights, support decision making processes and to build and maintain a platform in which to do this.

Lead Data Scientist for the Cashman Casino App

Cashman Casino is the latest game from Product Madness. It is a US centric app that was launched on Android in December 2016. After a successful launch it is currently ranked at number 9 in the Social Casino game category.

- Planning - Configuring the economy and simulating the user journey.
- Launch management - Validating technical status and game configuration. Testing initial strategic assumptions.
- Benchmarking - Identifying strengths and weaknesses vs other applications. Investigation the impact across the entrire portfolio i.e. cannibalisation vs incremental revenue.
- Mid to Long Term Strategy - Identifying content preferences and determining the strategy for the next year, how best to optimise & scale.

AB Testing and Cohort Analysis Tool Project Lead

Developed a production ready tool which included the statistical framework, data provider service, UI, search functionality, slides generation, test result archival and API.

Anomaly Detection System - Project Lead

This involved developing the forecasting models, incorporating them into our dashboarding tool and into a notification system via slack. This allowed the company to identify trends in time series data e.g. from successful campaigns as well as identify technical, economic or configuration issues.

Player 360 Framework Project Lead

Development of new player data models that describe their interactions and behaviours as a function of time and platform. This involved developing a new framework for ETL processes with full test coverage. The player 360 database is used as the main source of input for powering user segmentation systems.

Other Contributions

Lead Scientist for the Fafafa app, Product Madness' main Asian oriented app. Successfully identified new market segments based on content type. **Loyalty Programme** - helped to redesign, develop and monitor a loyalty programme focused on personalised player bonuses. **Loading and Payment Funnel Tools**: Built a set of tools for identifying and debugging technical issues. **Content Preference and Value Tool** - A tool used for providing insight into game popularity (within an app) and attributing real money value to each.

EDUCATION AND RESEARCH

UNIVERSITY OF BRISTOL PhD in Particle Physics

2009 - 2014 | Bristol, UK

My research involved analysis of particle production rates from proton-proton collision data from the Large Hadron Collider Beauty (LHCb) experiment at the CERN Laboratory in Switzerland. Using this data I was able to successfully model, verify and improve upon existing soft QCD models for hadron production rates and distributions using Monte Carlo simulation methods. This is important for understanding background processes in rare events which enables the detection of more exotic processes such as the production and decay of Higg's particles.

I also used image detection and processing methods to implement a software driven detector alignment in the LHCb Ring Imaging Cerenkov detectors. This enabled the group to improve the resolution in their particle identification systems and successfully distinguish between particles such as pions, muons and kaons.

UNIVERSITY OF COLLEGE LONDON Msc in Particle Physics

2006 - 2007 | London, UK

Comparison and Tuning of Monte Carlo Simulation Phenomenological Models of Tevatron Collision Data using the RIVET (https://rivet.hepforge.org) Tool.

UNIVERSITY OF LEICESTER Bsc in Physics with Space Science

2002 - 2005 | Leicester, UK

Modelling Surface growth simulation using accretion models and Monte Carlo driven methods.

Jackpot Behaviour Analysis into how short, mid and long term engagement and monetisation metrics are affected after a big win. This involved matching players who had won a jackpot with those who had not via a matching pairs algorithm. **Lobby Personalisation**: This involved using a linear regression model to predict the next game a user is most likely to play. **Team Management**: Successfully hired and mentored 3 new members to the data science team and organised team participation in conferences such as the Games Industry Analytics Forum and Scipy 2017 in Austin as well as an in house international data summit. **Organisational Outreach**: Developed and instructed a SQL course for members outside of the data science group.

OPENTRV Data Scientist | Software Engineer

July 2015 - January 2016 | London, UK

OpenTRV are an IoT company that specialise in the production of intelligent radiator valves that optimise comfort and efficiency. At OpenTRV I was responsible for building the UDP server, data pipeline as well as their web data platform, data visualisation interface and energy efficiency optimisation system.

LEARNING SCIENCE LTD Scientific Designer

August 2014 - April 2015 | Bristol, UK

Learning Science is a specialist in using innovation in technology to improve and modernise education. At Learning Science I worked as a scientific designer and developer of interactive physics simulations for use as learning resources and tutorials for university level education.

PROJECTS AND HACKATHONS

COMMONPLACE Team Lead

September 2015 - March 2016 | London, UK

Commonplace is a web and mobile platform aimed at building communities in London in order to tackle alienation, gang culture and violent extremism. The project is a collaboration between International Alert and peace building community. http://www.international-alert.org/news/peacehack-2015

POLLUPLA Co Founder

July 2015 - January 2016 | London, UK

Pollupla is a property search engine which incorporates air pollution data. We aim to provide users with insight into where they choose to live, businesses with market insight, and new air pollution standards relevant to the everyday person. I lead a team of three and was responsible for building the backend, pollution model and data pipeline as well as for putting together the business plan and pitching to investors. We successfully secured funding from the Climate KIC incubator in September 2015.

TFL DATA SCIENCE INNOVATION AWARD

January 2016 | London, UK

The TfL's urban data science (http://urbandatascience.com/) hackathon was an opportunity for TfL to explore how the field of data science can be used to optimise the service they provide to millions of customers each day.
Our project involved using data collected from tap ins at ticket barriers to predict in real-time the occupancy of a TfL stations allowing passengers to better plan alternative routes or departure times to ensure their arrival on time and in comfort.

REFERENCES

Daniel Freireich, Managing Director Digital, Product Madness
daniel.freireich@productmadness.com

Volodymyr Kazantsev, Head of Data Science, Wooga
v.kazantsev@gmail.com

Mark Hill, OpenTRV Co Founder
Mark@opentrv.uk

Nick Brook, Dean of Science, UCL
Nick.Brook@cern.ch