

## SELWYN McCracken, Ph.D.

I am a Data Scientist with extensive international statistical experience across the Media, Financial Services, Government, Pharma, Health, Automotive & Consumer Goods sectors.

A portfolio of examples is available on my website [www.selwyn-mccracken.com](http://www.selwyn-mccracken.com)



### PROFESSIONAL EXPERIENCE

#### **Freelance Data Scientist, October 2012 – Present** (*currently based in New Zealand*)

My principal client has been Guardian News & Media UK ([www.theguardian.com](http://www.theguardian.com)) where I was the lead Data Scientist responsible for:

- Behaviour profiling of their 5-million website visitors/day (e.g. user retention, preferred visit time/article type);
- Development of analytical tables based on 5-Billion page view records (with a 24 node Redshift cluster);
- Article and section similarity clustering;
- Monthly website traffic forecasting;
- Daily end-of day traffic prediction & identification of early signals that can be acted upon when a poor day is ahead.

#### **Ernst and Young - Fraud Investigation and Dispute Services (London)**

##### **Data Analytics Manager, October 2011 – September 2012.**

I co-managed a team of 6 - 10 analysts for Ernst & Young's largest global project, which sought to improve the public finances of a sub-Saharan African Government.

I led and coordinated the analytical support my team gave to the 200+ E&Y accountants that were involved across a range of Government activities such as Budget, Revenue, Expenditure, Debt Management, IT & HR.

Examples of specific activities that I directed include:

- Implementation of an interactive web-based visualisation and financial reporting portal;
- Diagnosis of system integration issues across the key Government databases;
- Design and validation of predictive models (e.g. identifying Ministries that will overspend their budget);
- Reconciliation of external debt disbursements and repayments, across the various debt monitoring systems;
- Tax payment and fine monitoring of companies granted tax breaks and exemptions;
- Quality assurance reviews and code version control processes for all requests.

#### **Ernst and Young - Fraud Investigation and Dispute Services (London)**

##### **Data Analytics Executive, March 2010 – September 2011**

Completed several pieces of work on E&Y's largest project such as:

- Tracing and visualisation of \$8.5 Billion of tax revenue through more than 200 taxation accounts;
- Confirmation of \$5 Billion of suspected outstanding payments to Government vendors;
- Account reconciliation and fuzzy transaction matching between the Government's financial system and commercial bank account statements;
- Clustering of Ministries according to budget and expenditure behaviour to identify high variance outliers;
- Hadoop-based large-scale parallel computation analysis proof of concept trial;
- Automatic classification of financial system transactions from free-text descriptions in SQL.

#### **Vanguard Strategy (London)**

##### **Senior Analyst, January 2008 – February 2010**

Provided statistical support on client engagements, such as:

*Continental tyres (Germany, UK, Czech Republic) – Tyre market consumer dynamics:*

- Established the main factors of tyre purchases for the client brand using logistic regression models;
- Estimated customer lifetime value via survival analysis;
- Forecasted market share equilibria of tyre channels.

*Wyeth Pharmaceuticals (UK) – Effexor (antidepressant) end-of-life sales force reallocation*

- Saved £1.75m in costs by optimising the size of a Pharmaceutical sales force across regions & product line, after consideration of drug patent lifecycle, promotion effectiveness & intensity.

*Coca Cola (Japan) – Georgia coffee:*

- Established that Coke Japan were incorrectly targeting canned coffee consumers that were 10-15 years older than the bulk of their customer base.

## **Vanguard Strategy (London)**

### **Analyst-Statistician, June 2006 – December 2007:**

- Identified high-value Physician segments within the US Anti-Depressant market;
- Identified early adopters of a new Anti-Depressant based on past drug-analogue uptake;
- Estimated the level of influence that a Specialist has upon their linked GPs prescribing behaviour.

## **Dept. Preventive & Social Medicine, University of Otago (New Zealand)**

### **Health Research Council Training Fellow, April 2001-December 2005:**

- Undertook Ph.D. Research, epidemiology tutoring, supervisor research assistance, and Public Health training
- Ph.D. research focused on the design and analysis of a national injury statistics database for the New Zealand construction industry, requiring:
  - Regression based analysis of construction injury risk factors and injury rate fluctuations over time;
  - Machine-learning classification of free-text injury descriptions.

## **Dept. Preventive & Social Medicine, University of Otago (New Zealand)**

### **Junior Research Fellow, 1999-2001**

- Established that work fatality rate differences for Maori & Non-Maori are due to occupational concentration.

## **TECHNICAL SKILLS**

Expert user of the **R** programming language:

- Multivariate regression modelling and prediction (linear, logistic, GAM regression);
- Tree-based segmentation and classification (e.g. random-forests);
- Loyalty/Survival/Time-to-Event modelling;
- Machine Learning classification & pattern recognition (SVMs, kNN, kmeans);
- Data cleaning, missing-data interpolation and imputation;
- Monte-Carlo and bootstrap analyses, kernel density estimation;
- Advanced graphics – e.g. multi-dimensional plots, map making and plot animation;
- Time series analysis and forecasting;
- Model cross-validation, regularization and sensitivity testing.

Seasoned **SQL** database analyst:

- Amazon Redshift, PostgreSQL, MS SQL Server, MySQL, MS-Access, SQLite.

‘Big Data’ tools experience with **Hadoop**:

- Hadoop cluster jobs on Amazon’s EMR /EC2 with Apache Whirr.
- Pig, Hive

Interactive web-based **data visualisation** skills:

- D3.js. + JQuery + Javascript.

Considerable **Python** programming experience:

- Scripting, analysis automation, database ETL, data cleaning, & text processing.

Capable **Linux** user and administrator (primarily the Debian and Ubuntu distros):

- Shell scripting and task automation
- Remote access and cluster computing configuration (e.g. initiating a Hadoop cluster via SSH on Amazon EC2)
- Git version control

## **EDUCATION**

2007 Ph.D. – Occupational Epidemiology. University of Otago (New Zealand)

2001 MSc – Psychology, University of Otago

1999 BSc – Psychology, University of Otago

## **PERSONAL DETAILS**

Nationality: New Zealand

UK Work Visa: Indefinite Leave to Remain granted February 2013

Email: selwyn.mccracken@gmail.com

Mobile: (+44) 789-686-9183 / (+61) 478-685-903

Website: <http://selwyn-mccracken.com>