

**EMPLOYMENT**

---

**Principal Data Scientist, Platform**                      **WalmartLabs**                      **Fall 2012 - Present**

- Main developer of Sherlock, a real time scalable anomaly detector. Sherlock runs in Walmart's production environment, processes 1 Million events per minute, and has an F1 score of 0.973.
- Implemented a streaming platform that generates time series from massive logs data streams using Mupd8 (a @WalmartLabs open source project).
- Built a monitoring visualization platform using GWT.
- Designed and implemented Walmart's web analytics data model in Hadoop and Hive.

**Senior Manager, Analytics**                      **Walmart.com**                      **Summers 2011 – 2012**

- I was transferred to the US to implement in Walmart.com the models we had running in Chile.
- Implemented the first versions of Walmart.com's pricing, assortment and recommendation engines.

**Director, Analytics**                      **Walmart-Chile**                      **Winter 2010 – Spring 2011**

- Led a technical team responsible for leveraging Walmart's data into better customer service and increased sales.
- Deployed analytics applications for Pricing, Assortment and new stores allocation that are still being used.

**Manager, Analytics**                      **Walmart-Chile**                      **Spring 2007 – Fall 2009**

- Pioneered the use of statistical models on Walmart-Chile data.
- Developed a coupon assignation model for brick and mortar stores using Collaborative Filtering.
- Identified households with kids based on their purchases (did you know that chocolate milk is a better predictor than diapers?).
- Calculated items price elasticity using custom models.

**Analyst, Analytics**                      **LAN Airlines, Chile**                      **Winters 2006-2007**

- Generated scorecard reports
- Designed data model for analytics database.

**EDUCATION**

---

**Stanford, CA**                      **Stanford University**                      **Falls 2012 – 2013**

Mining Massive Datasets Graduate Certificate

- CS229, Machine Learning, Prof. Andrew Yng
- CS246, Mining Massive Datasets, Prof. Jure Leskovec
- CS276, Information Retrieval and Web Search, Prof. Christopher Manning
- CS224W, Social and Information Network Analysis, Prof. Jure Leskovec

**Viña del Mar, Chile**                      **Universidad Adolfo Ibanez**                      **Winter 2003 – Fall 2005**

Master in Sciences

- Stochastic processes, data mining, graphs theory.
- Focused in data mining and complex systems modeling using agent based simulation.
- Thesis title: "Modeling an Economic System with Agent Based Simulation" (A+).

**Viña del Mar, Chile**                      **Universidad Adolfo Ibanez**                      **Winter 1998 – Fall 2002**

Bachelor in Engineering Sciences

- Bioengineering, statistics, electromagnetism, electronics, thermodynamics, optimization, IT, mechanics of materials.

## TECHNICAL EXPERIENCE

---

### Projects

- **Sherlock** (2014). An implementation of the scalable anomaly detector published in Farren [2014]. Python.
- **Large-scale aggregation** (2013). A distributed data stream aggregator using mupd8. Java.
- **Typing recognition** (2013). A machine learning algorithm that classifies typing down stroke times. Used as an extra authentication layer for web apps. Python.
- **Customer Lifecycle** (2012). An analysis of Walmart.com customers. R.

### ADDITIONAL EXPERIENCE AND AWARDS

---

- **Coursera's Mining Massive Datasets TA (2014)**: Helped Professors Jure Leskovec, Anand Rajaraman, and Jeffrey Ullman running the class, creating quizzes and exams, and answering student's questions.
- **Simulation TA (2004)**: Taught weekly workshop that was a complementation to the main class.
- **Walmart's compass award (2012)**: Award given each quarter to the best performer associate.

### LANGUAGES AND TECHNOLOGIES

---

- Python; Java; SQL; PL-SQL; Objective-C; bash; R; Matlab
- Hadoop; mupd8; linux; GWT; Cassandra; Hive;

### PUBLICATIONS

---

- **Sherlock: A real time and scalable anomaly detector using parallel sparse regression** by Derek Farren. INFORMS 2014.
- **Analysis of Networks in Chess** by Derek Farren, Daniel Templeton and Meiji Wang. Stanford, 2013.
- **Predicting retail website anomalies using Twitter data** by Derek Farren. Stanford, 2012.
- **Modeling an economic system using Agent Based Simulation** by Derek Farren. Master's thesis, *Universidad Adolfo Ibañez*, 2005.
- **Economic experiments using Agent Based Simulation** by Derek Farren. *Universidad Adolfo Ibañez*, 2004.