

STATISTICS COLLOQUIUM

MONDAY, APRIL 8, 2013

TALK: 4:15 PM — SCIENCE CENTER RM. 705

RECEPTION: 3:50 PM — SCIENCE CENTER, 7TH FLOOR

“Big Data: Statistical and Computational Challenges and Opportunities”

**Nando de Freitas
Department of Computer Science
University of British Columbia**

ABSTRACT

I will review several successful consumer products and argue that huge datasets played a central role in their construction. Among these products, I will describe Zite, a personalized magazine built by my team and acquired by CNN, which uses massive classifiers to label nearly a billion documents and personalize content for millions of users. I will then discuss two machine learning ideas at the core of many of these products: Ensemble methods and Bayesian optimization. I will briefly present new theoretical results on online random forests, a very popular ensemble method, and point out theoretical problems and practical challenges in scaling this technique further. I will conclude with an overview of Bayesian optimization, present our theoretical advances in this area, and share our experiences in applying this technique to automatic algorithm configuration, information extraction, massive online analytics and intelligent user interfaces.