
Russell Bradt Undergraduate Colloquium

Department of Mathematics

University of Kansas



Cody E. Clifton

Marketing Specialist
Adpearance

Ph.D. 2015 KU Mathematics

After receiving his Ph.D. in Mathematics in 2015 from the University of Kansas, Cody started working this spring at Adpearance, a digital marketing agency, in Portland. At KU he received the 2015 John Bunce Memorial Award presented to an outstanding graduate student, 2014 Florence Black Award for Excellence in Teaching, and the 2013 Outstanding Graduate Teaching Assistant Award from the KU Office of Graduate Studies.

In this talk, we build up from decision trees to random forests, using the spammy link predictor as a case study. We then take a step further and discuss the future of prediction, including Google's forays into deep learning along the path towards artificial intelligence.

April 29, 2016

2:00 pm

120 Snow

Planting the Seeds of Prediction

A Random Forest Approach to Detecting Disreputable Domains

With the imminent Penguin 4.0 update to its core search algorithm, Google is expected to implement real-time ranking penalties for disreputable inbound links. It would be unrealistic for a digital marketing agency that manages the web presence of hundreds of clients to manually perform time-consuming inbound link audits for every client on a regular basis. If we could automatically detect when a website experiences an influx of spammy links pointing to, then we would be able to dedicate our time to those clients whose sites are at immediate risk of an undesirable dip in Google search rankings due to Penguin penalties.

In order to facilitate the regular monitoring of inbound links to our clients' sites, the Search Engine Optimization Team at Adpearance, Inc. has developed a machine learning algorithm that trains on a collection of URLs that have been manually categorized as either "spam" or "not spam" and then uses a random forest classifier to predict the reputability of new inbound links.