Predicting Which Recommended Content Users Click

Pramod Varma (github@**pramodv79** )

# Abstract

In this paper I explore application of data science to perform predictive analysis.

# Introduction

The internet is a stimulating treasure trove of possibility. Every day we stumble on news stories relevant to our communities or experience the serendipity of finding an article covering our next travel destination. [Outbrain](http://www.outbrain.com/), the web’s leading content discovery platform, delivers these moments while we surf our favorite sites.

Currently, Outbrain pairs relevant content with curious readers in about 250 billion personalized recommendations every month across many thousands of sites. In this paper, I explore to predict which pieces of content outbrain’s global base of users are likely to click on. Improving Outbrain’s recommendation algorithm will mean more users uncover stories that satisfy their individual tastes.[1]

# Data Exploration

The dataset for this project was provided by outbrain as part of a Kaggle competition. It contains a sample of users’ page views and clicks, as observed on multiple publisher sites in the United States between 14-June-2016 and 28-June-2016. Each viewed page or clicked recommendation is further accompanied by some semantic attributes of those documents. For full details, see data specifications below.

The dataset contains numerous sets of content recommendations served to a specific user in a specific context. Each context (i.e. a set of recommendations) is given a display\_id. In each such set, the user has clicked on at least one recommendation. The identities of the clicked recommendations in the test set are not revealed. The task is to rank the recommendations in each group by decreasing predicted likelihood of being clicked.

## Pre-processing

Outbrain provided multiple normalized dataset.

# Feature Selection

# Deciding on estimator

# Model Training

# Model Prediction

# Model Evaluation

# Challenges and Success stories

# Conclusion and Learnings

# Future Work

# Reference

[1] https://www.kaggle.com/c/outbrain-click-prediction/data