

# Capstone Proposals

N Butterworth - DSI Ida Wells

# Option 1: Kaggle Ad Click Competition

## Description

Predict click-through rates on ads

## Data Sources

Kaggle:

<https://www.kaggle.com/c/avazu-ctr-prediction>

## Resources

<https://www.kaggle.com/c/avazu-ctr-prediction>

## Methods

Mass quantities of Neural Networks, bin counting to probabilize categorical features, maybe try Naive Bayes

# Option 2: Predict Tomorrow's News Today

## Description

Look at velocity of topics from top news sources

## Data Sources

Scrape APIs and/or Google News

## Resources

Starting set of topics - FreeBase(?)

## Methods

NLP to extract topics (Latent Dirichlet Allocation), time series analysis to predict velocity

# Option 3: Kermit/not Kermit

## Description

Detect whether a video has a green frog. If so, is it Kermit?

## Data Sources

YouTube videos

## Resources

Video convolution NN stuff



## Methods

Convolution on video, frame by frame

