

# THESIS LOG

STEFAN SABEV

## 1. 14TH OCTOBER 2014

What's been tried so far:

- I have used smoothing to smooth the weekly seasonality component out of both the twitter and searches data. That has yielded small improvement in the correlation coefficients.
- I have also used a very basic method of prediction which works as follows:  
Calculate the mean and the standard deviation of the searches. If the standard deviation is more than the mean, then there has been a spike which has pushed it higher.
- I've also used that to determine which destinations should have a classifier built. That has yielded very small improvements.

The fact that the simple combination of LASSO + Ridge regression does not perform miraculously well has led my supervisor and I to believe that perhaps we should investigate more sophisticated models that will perhaps model the problem better.

I am currently doing:

- Reading the Adams & MacKay paper on Bayesian change point models
- Will look into the matlab implementation and try to port it to Python.