



# Predicting Procrastination

Using machine learning to enhance productivity

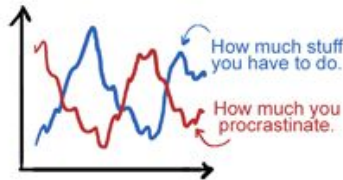
Roland Jeannier  
[rtjeannier@gmail.com](mailto:rtjeannier@gmail.com)  
[www.rolandjeannier.com](http://www.rolandjeannier.com)<sub>1</sub>

# Introduction

# Can we dynamically detect if a person is procrastinating?

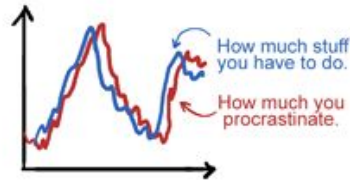
## Procrastination

Ideally:



$$\text{Procrastination} \propto \frac{1}{\text{How much stuff you have to do}}$$

In reality:



$$\text{Procrastination} \propto \text{How much stuff you have to do}$$

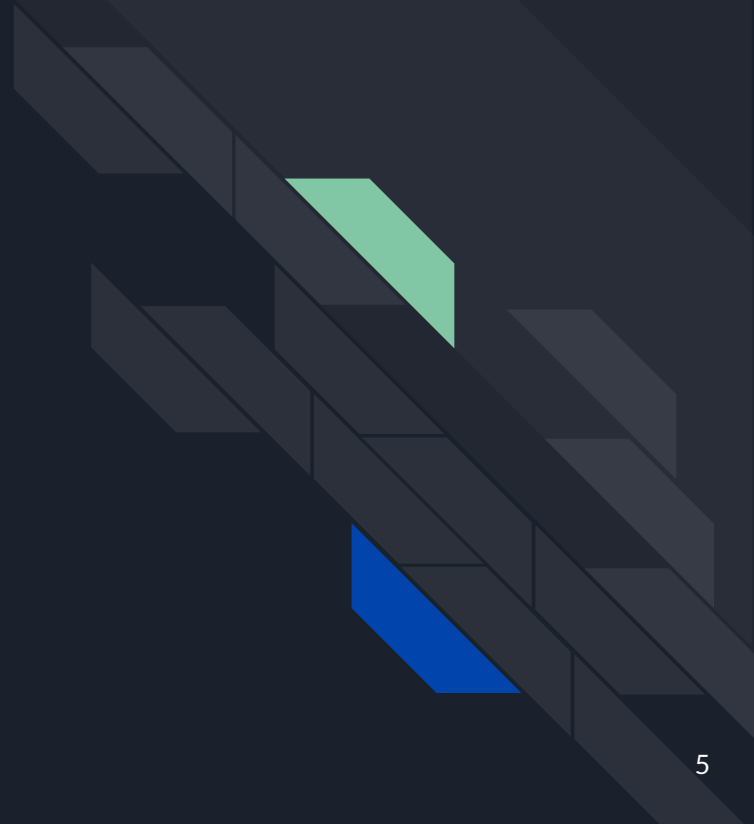
JORGE CHAM © 2010

WWW.PHDCOMICS.COM

# Overview

- Introduction
- Data Collection
- Model Building
- Interpreting the Results
- How to can this be used?
- Conclusion

# Data Collection





# Data Collection Information

- 1,088 web pages, totalling 434.8 MB
- 639 procrastinating vs 448 working
- Collected data 7/20/2017-9/06/2016
- Created a chrome extension that saved the raw HTML by click.
- The URL was embedded into the HTML.
- The filename was generated to included the classification and a timestamp.

# It wasn't pretty...

The screenshot shows the AskReddit website in a Google Chrome browser. The page title is "What did your job want you to hide from customers?". The post is by user "ActualRobot1" and has 7,378 upvotes (94% upvoted). The post text is: "I work at a movie theater. People are always trying to sneak food in. In our employee handbooks it actually says people can. Ring food in as long as it doesn't smell bad. It's funny to see customers nervously hide their bags so we don't see it." Below the post are several comments. The first comment is by "Hail\_storm4" with 4,752 points, asking "Would you give a shit even if it was not allowed?". The second comment is by "Hail\_storm4" with 2,196 points, saying "Probably not. My manager lets people bring their dogs in so she has no shits to give." The third comment is by "Dupragon" with 650 points, asking "What movie theater?". The fourth comment is by "countastrotacos" with 1,710 points, mentioning "Cinebark". The right sidebar shows the "Ask a question" button, a search bar, and the subreddit statistics: 17,871,235 subscribers and 49,395 online now. The bottom of the page shows the browser's address bar with the URL "https://www.reddit.com/r/AskReddit/comments/6yjoiw/what\_did\_your\_job\_want\_you\_to\_hide\_from\_customers/" and the page title "What did your job want you to hide from customers? : AskReddit - Google Chrome".

What did your job want you to hide from customers? : AskReddit - Google Chrome

capstone-final-not... Documents/GA-D... capstone-final-not... starter-code-comp... Predicting Procas... Boy (BeatauCue... PMCText-mining/r... Python Matplotlib... What did your job... Rel

Secure | https://www.reddit.com/r/AskReddit/comments/6yjoiw/what\_did\_your\_job\_want\_you\_to\_hide\_from\_customers/

DASHBOARD HOME POPULAR ALL RANDOM FRIENDS SAVED EDIT | add shortcuts from the my subreddits menu at left or click the button by the subreddit name, drag and drop to sort

askreddit

COMMENTS SHOW IMAGES (1)

7378

What did your job want you to hide from customers? (self AskReddit)

submitted 10 hours ago by ActualRobot1

5315 comments source share save hide give gold report crosspost hide all child comments

top 200 comments show 500 navigate by

sorted by: best

DO NOT POST PERSONAL INFORMATION, PHONE NUMBERS, EMAIL ADDRESSES, NAMES, ETC, REAL OR FAKE, OR YOU WILL BE BANNED WITHOUT WARNING.

save content policy formatting help

[-] Hail\_storm4 4752 points 6 hours ago

I work at a movie theater. People are always trying to sneak food in. In our employee handbooks it actually says people can. Ring food in as long as it doesn't smell bad. It's funny to see customers nervously hide their bags so we don't see it.

permalink source embed save save-RES report give gold reply hide child comments

[-] clockworkgir21 1860 points 5 hours ago

Would you give a shit even if it was not allowed?

permalink source embed save save-RES parent report give gold reply hide child comments

[-] Hail\_storm4 2196 points 4 hours ago

Probably not. My manager lets people bring their dogs in so she has no shits to give.

permalink source embed save save-RES parent report give gold reply hide child comments

[-] Dupragon 650 points 4 hours ago

What movie theater?

permalink source embed save save-RES parent report give gold reply hide child comments

[-] countastrotacos 1710 points 2 hours ago

Cinebark

procr\_2017-...html

Show all

Ask a question

[SERIOUS]

search

Submitted on 06 Sep 2017

7,378 (94% upvoted)

https://redd.it/6y6j6iw

Welcome to /r/AskReddit

Use subreddit style

Unsubscribe

+shortcuts +dashboard

17,871,235 subscribers 49,395 online now

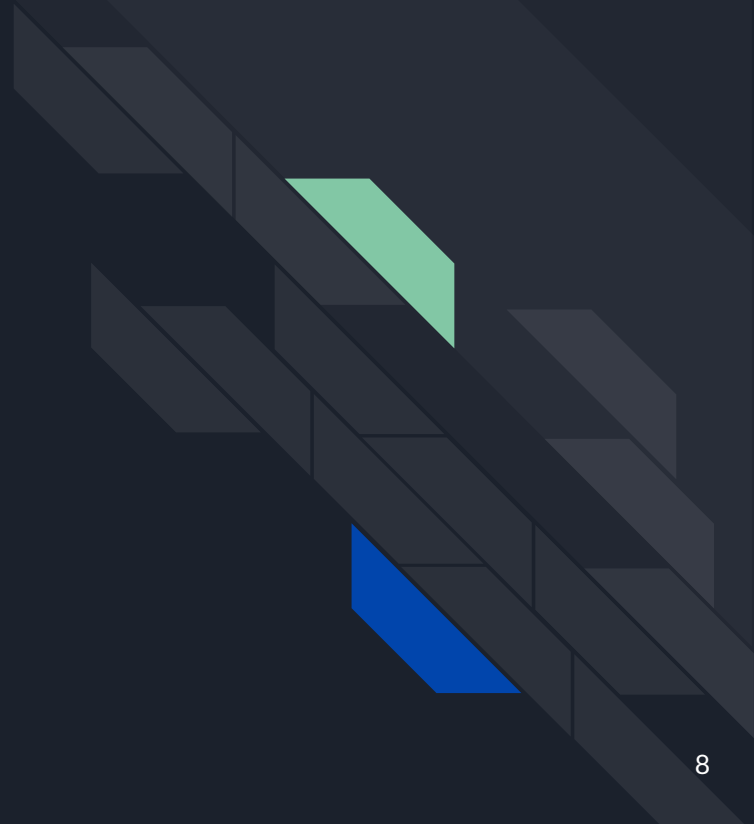
Rules:

(hover for more info)

1 You must post a clear and direct question in the title. The title may contain two, short, necessary context sentences. No text is allowed in the textbox. Your thoughts/responses to the question can go in the comments section. more >>

2 Any post asking for advice should be generic and not specific to your situation alone. more >>

# Model Building







# Natural Language Processing (NLP)

- CountVectorizer to generate features
  - Ngram with a range 1-4
- Used stop word
  - nltk default english words
  - 'http', 'https', 'com'
- TfidfTransformer to help with smoothing
- 3,104,198 unique features created
  - Lasso regularization reduced it to 163



# Logistic Regression

- Model preferred for interpretability
  - Easy to calculate odds ratio ( $e^{\text{coef}}$ )
- Lasso Regularization
  - C value grid searched to be 0.06
  - Performed excellent feature selection
- Significantly outperformed SVM,  
Random Forest, BernoulliNB



# Creating a pipeline

The overall goal of the model building process is to make the approach generalizable.

A pipeline was created streamlining the countvectorization into the logistic regression. This allows for easily modifiable code that can be updated in the future when more data is available.

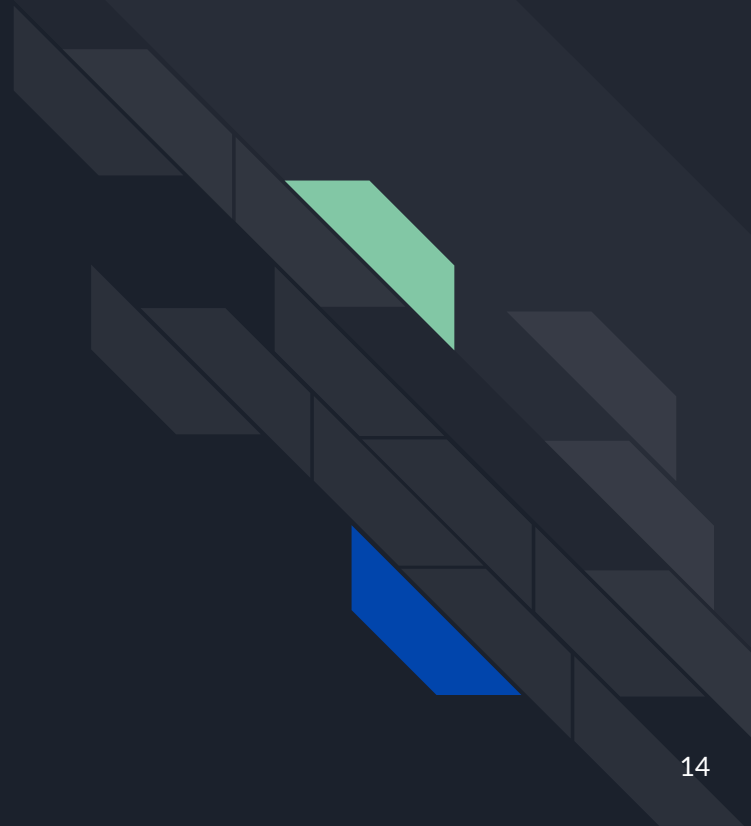
# Interpreting the Results

Accuracy

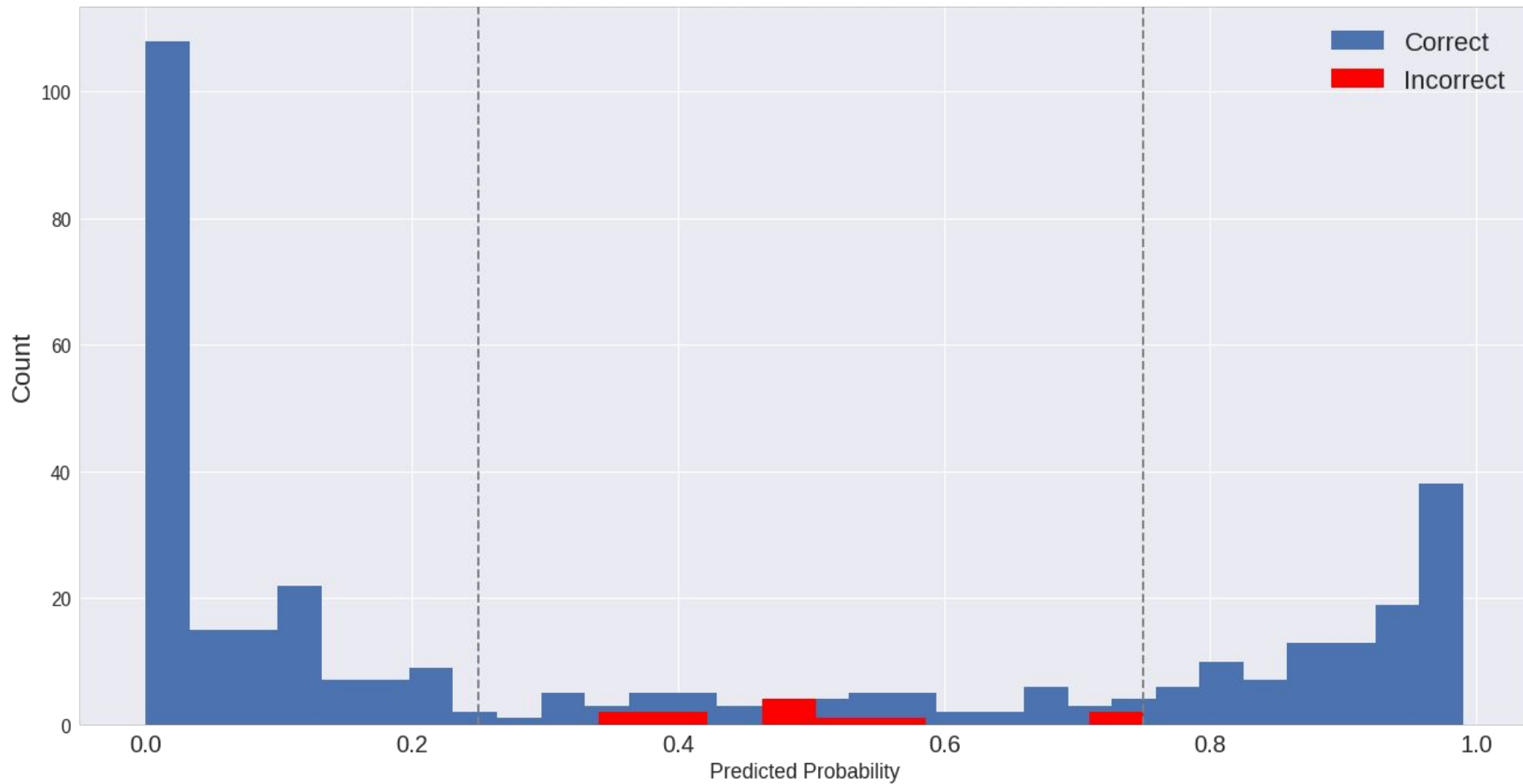
97.49%

ROC-AUC

0.992



Predicted Probabilities for Classified and Missclassified Data



How can this be used?



# Building a modern productivity application

This model has the capacity to be the backbone of a modern productivity application. Youtube can be the source of an endless amount of distraction, but it can also be a source of great amounts of useful information.





# Conclusion

In short, you can predict if you are procrastinating by viewing the content of a web page (at least I can). Using a bag of words approach to solving the problem along side logistic regression can lead to extremely accurate predictions

It is unclear if this process will work for any person. Efforts were taken to generalize the process, but without data to try we cannot extrapolate to that point yet.

This can become the backbone of a modern approach to productivity apps. Predicting when a person is procrastinating has benefits outside of just blocking web pages. You can monitor performance in real time. Record mouse movements and actions per minute. With enough participants larger generalizations can be made about productivity in general. This is a small step in the direction of utilizing machine learning to enhance routine aspects of our lives.