Are You Afraid Of The Story?
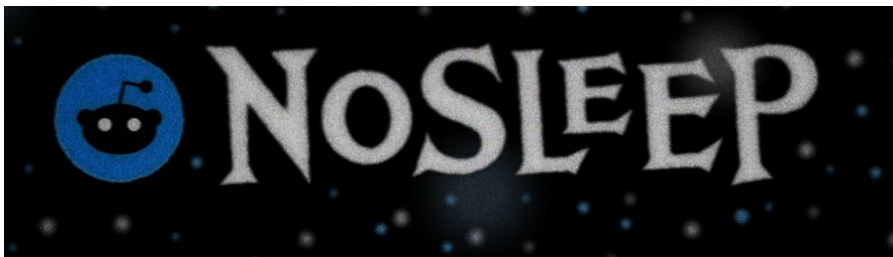
By Chuck Dye

Is it possible to tell whether a "scary story" is true?

- Historically considered easy
- Recent changes in the 'meta' of scary stories
- Telling the difference is more difficult than ever

/r/NoSleep

Nosleep is a subreddit for realistic horror stories.
Everything is true here, even if it's not.

/r/LetsNotMeet

A place to read true stories about people you never want to meet again.

My Process?

- **Gather**

- **Clean**

- **Vectorize**

- **Model**

- **Test**

Gathering

Web Scraping Function

Almost 2500 posts per subreddit

More info than just the title and text

Cleaning

Data came very clean

While the DataFrame was low on Null values The text was dirty

RegEx and string functions

Vectorizing

Countvectorizer

'Bag of Words'

Extremely easy to fit the model

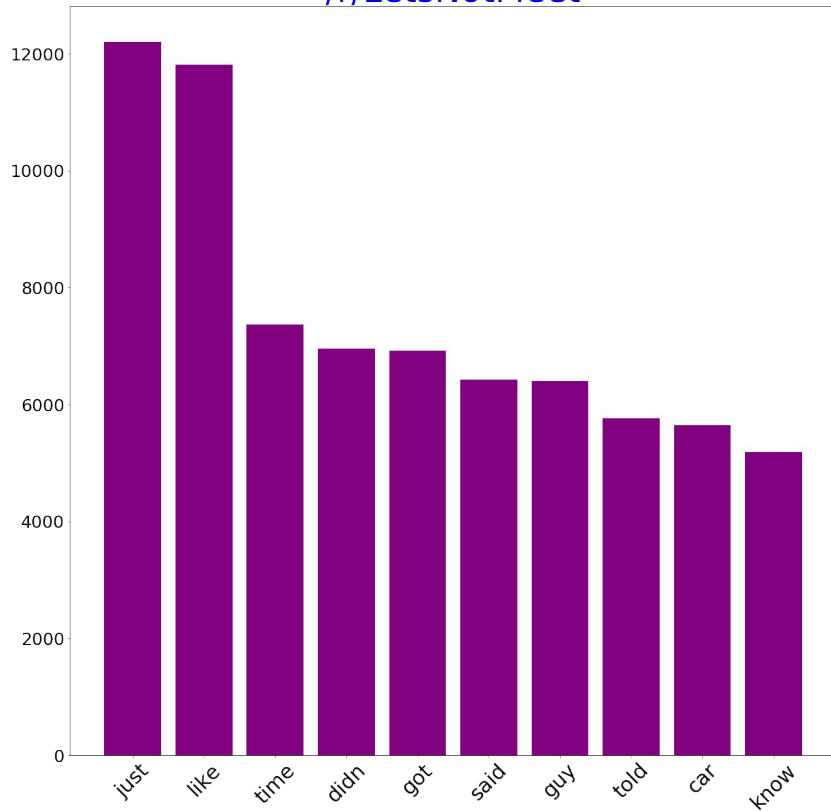# Modeling

Fit all the above steps into a pipe line

Gridsearch to find optimal model
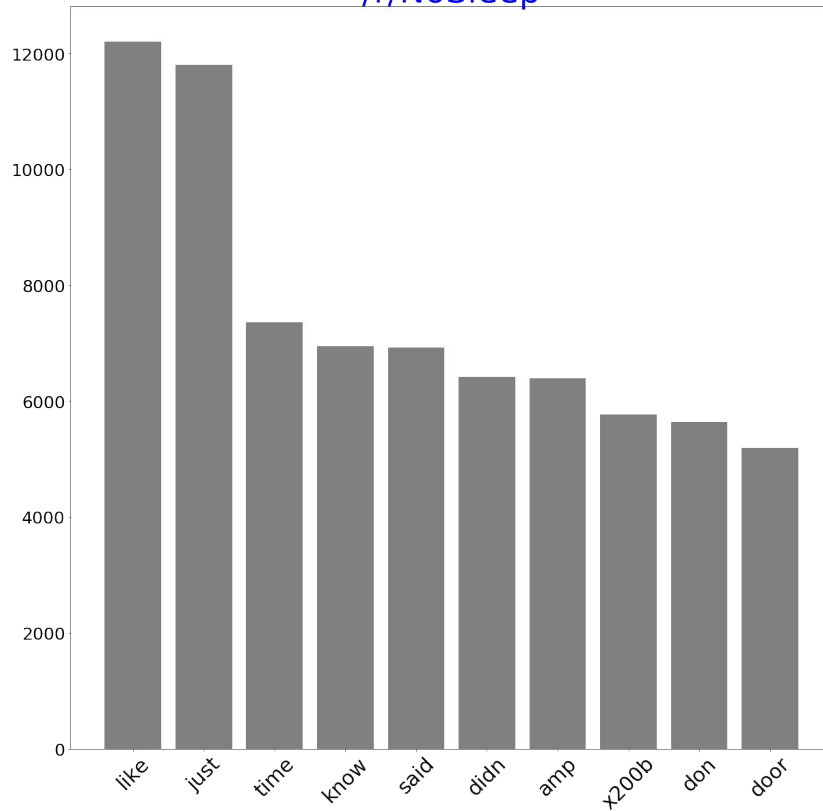
Strongest model: LogReg

/r/LetsNotMeet

Top words: just, like, time, didn, got, said, guy, told, car, know

/r/NoSleep

Top words: like, just, time, know, said, didn, amp, x200b, don, door

Top words for each Subreddit

↑
69
↓

r/nosleep · Posted by u/sortof_haunted 2 months ago

Yeah I'm definitely undead [Part2]

Series

Hardships & Challenges

- RegEx functions
- Subreddits too similar
- Visualizing Gridsearch params

What did we learn?

Machine Learning as a fact or fiction classifier Is absolutely a viable process

Over many models Logistic Regression was by far the strongest

Using Gridsearch we found that the regularization strength that's optimized at a value of 10.0

People are getting better at making fake stories seem more real

Even with a large similarity between classes There is still a decent opportunity to classify correctly

Thank You

By Chuck Dye