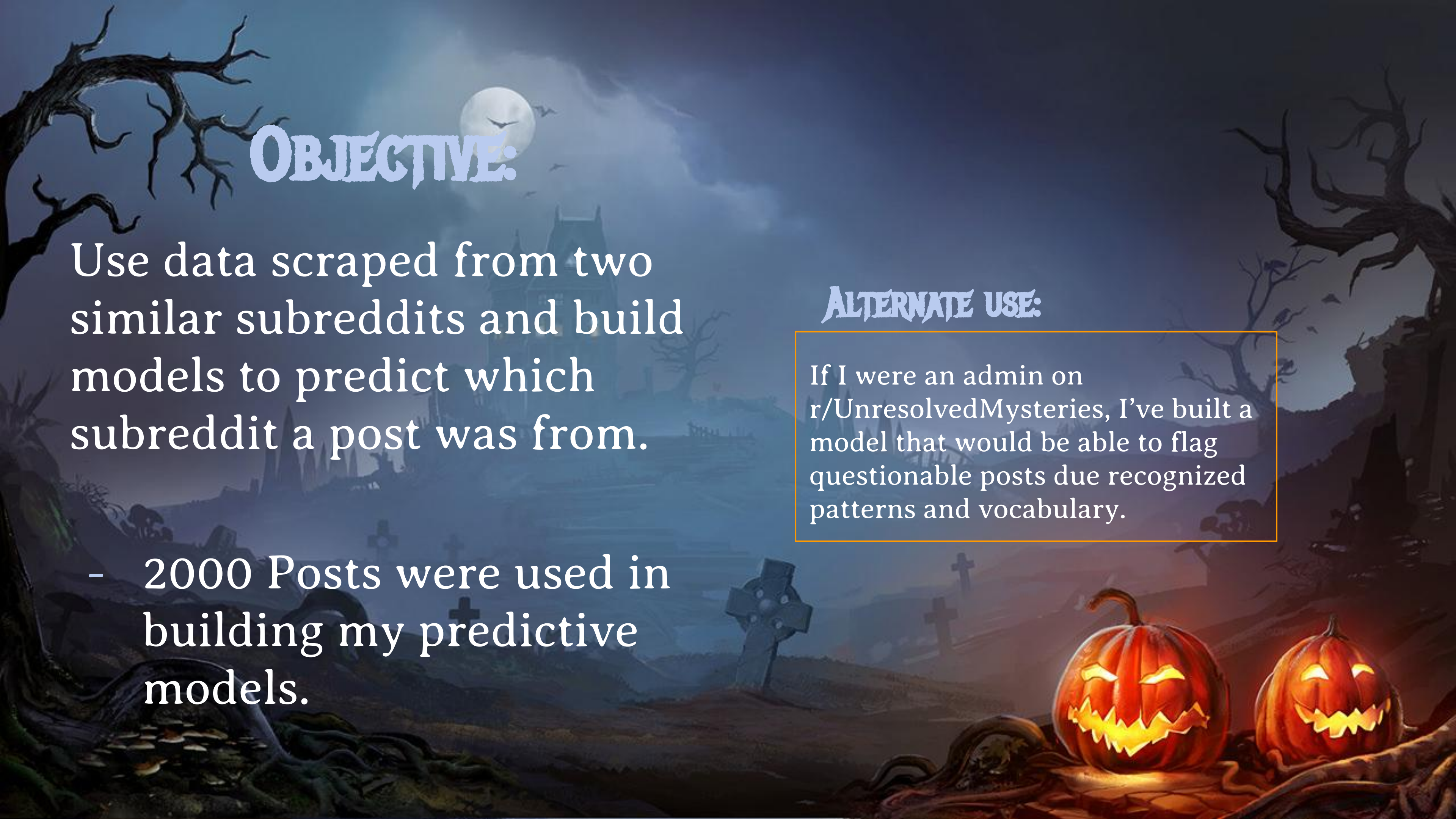
R/NOSLEEP VS. R/UNRESOLVEDMYSTERIES

Spoopy!



OBJECTIVE:

Use data scraped from two similar subreddits and build models to predict which subreddit a post was from.

- 2000 Posts were used in building my predictive models.

ALTERNATE USE:

If I were an admin on r/UnresolvedMysteries, I've built a model that would be able to flag questionable posts due recognized patterns and vocabulary.

Explaining the Subreddits.

r/nosleep :

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

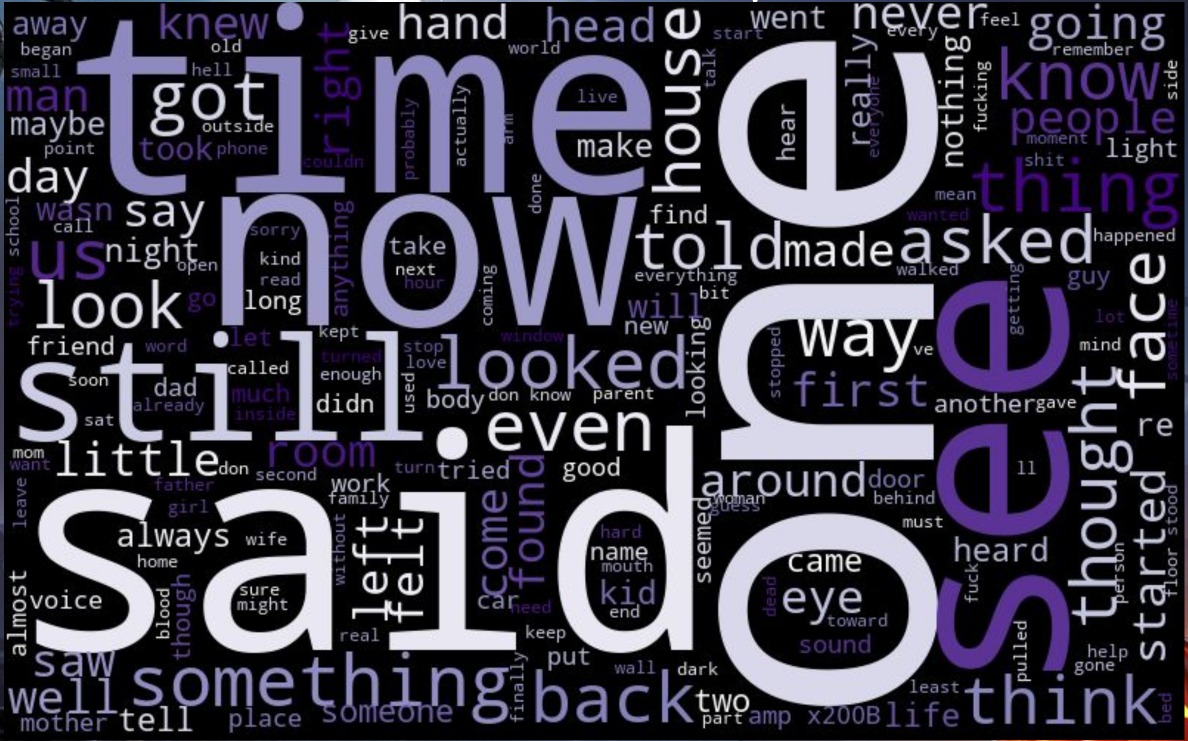
Fictional Stories

r/UnresolvedMysteries:

“A subreddit dedicated to the unresolved mysteries of the world. Submissions should outline a mystery and provide a link to a more detailed review of the case such as a Wiki article or news report.”

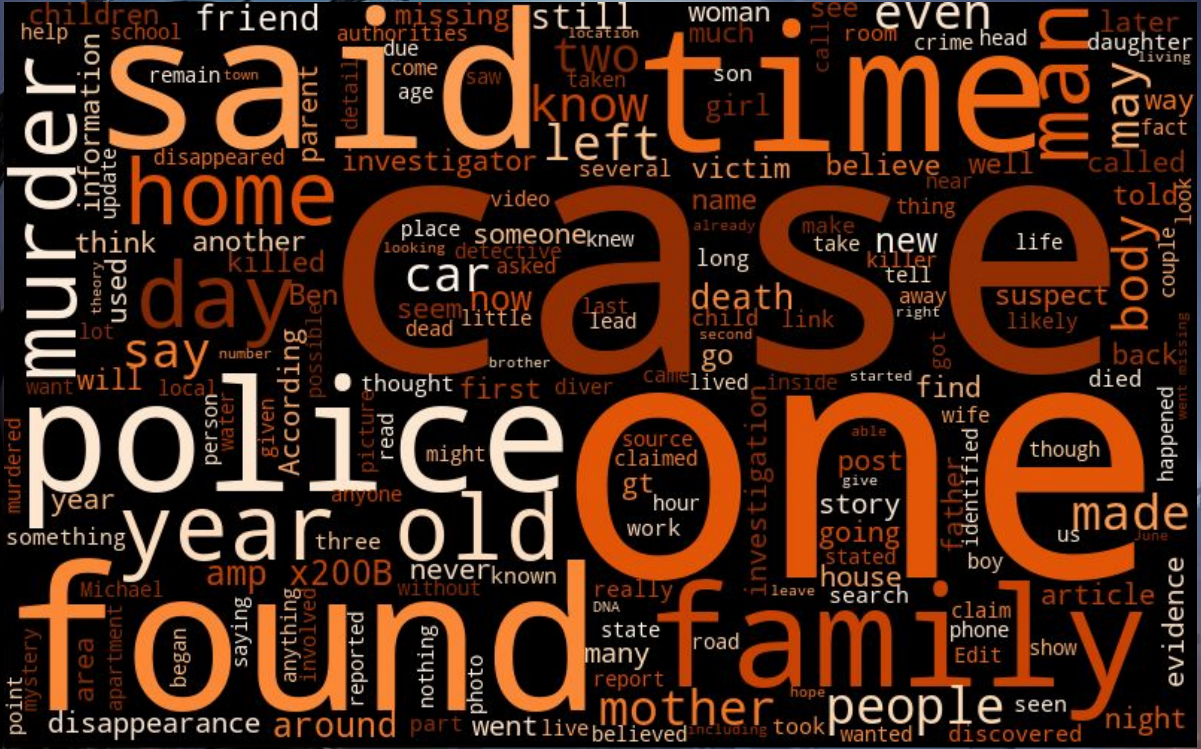
~ Mostly Non-fictional stories/news

R/NOSLEEP



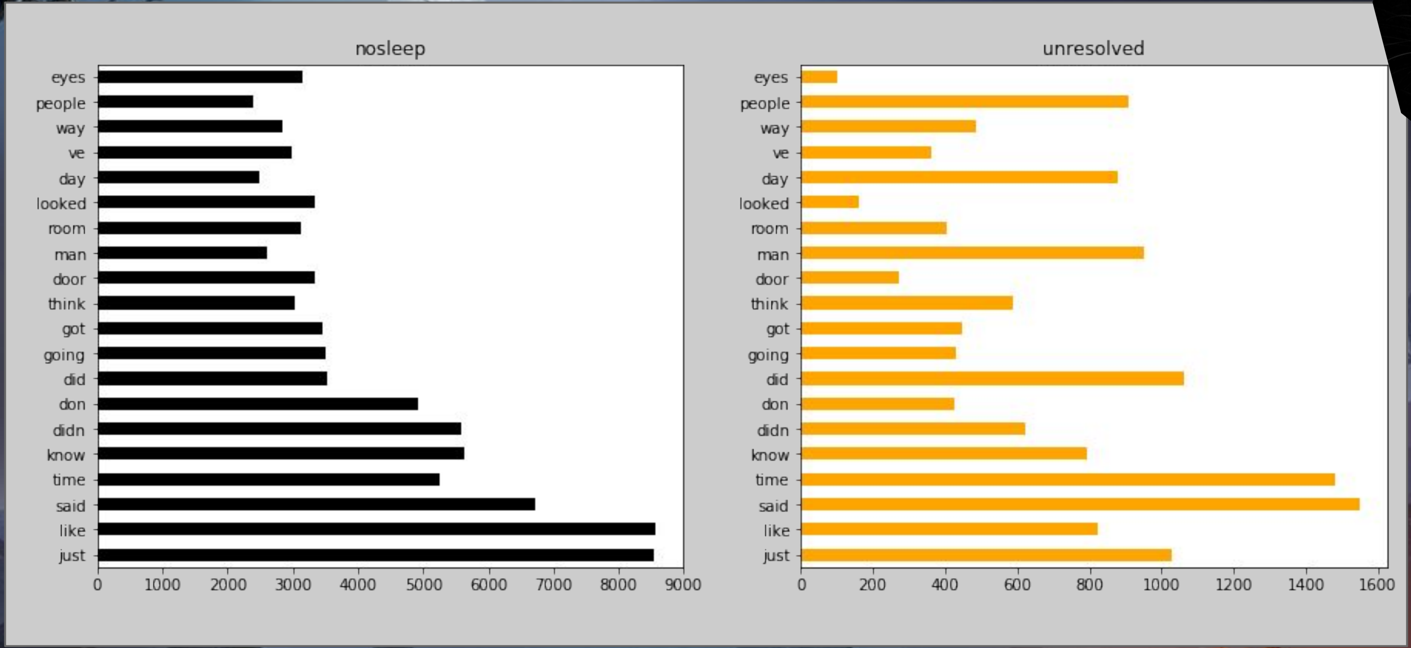
Most common words used in the r/Nosleep subreddit. Nothing particularly interesting...

R/Unresolved Mysteries

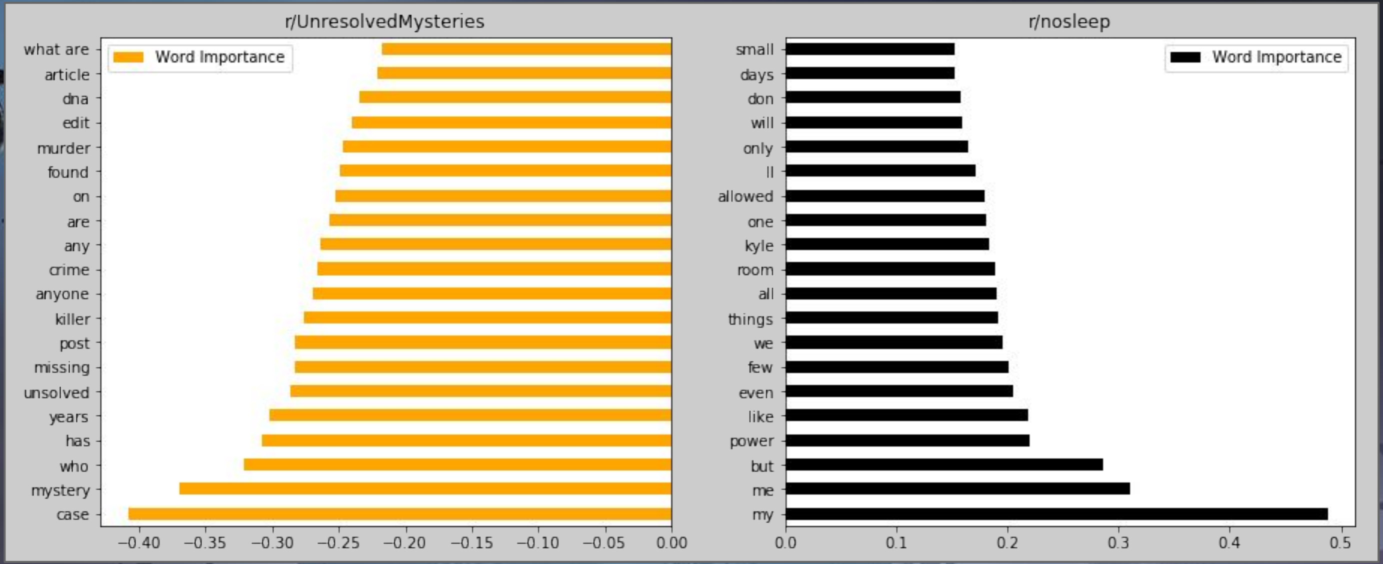


Most common words used in the r/UnresolvedMysteries subreddit. These were much more interesting… and more spooky.

Top 20 most common words for each subreddit



The most used words for both subreddits were generally very common. I didn't see much value in these, but there were some interesting finds. Apparently time is a very spooky word.



I immediately notice a couple things about the models highest calculated important vocabulary associated with r/UnresolvedMysteries. The words used were definitely more unique and… more spooky

r/nosleep associated words were much less complex. I didn't see any words or phrases that were particularly notable. Apparently "Kyle" has a rough time in the r/nosleep subreddit.

Top 20 most important words for each subreddit

Best Model

Logistic Regression Model using CountVectorizer

‘max_features': 2500,

'ngram_range': (1, 2)}

Train Accuracy Score : 1.0

Test Accuracy Score : 0.9817905918057663

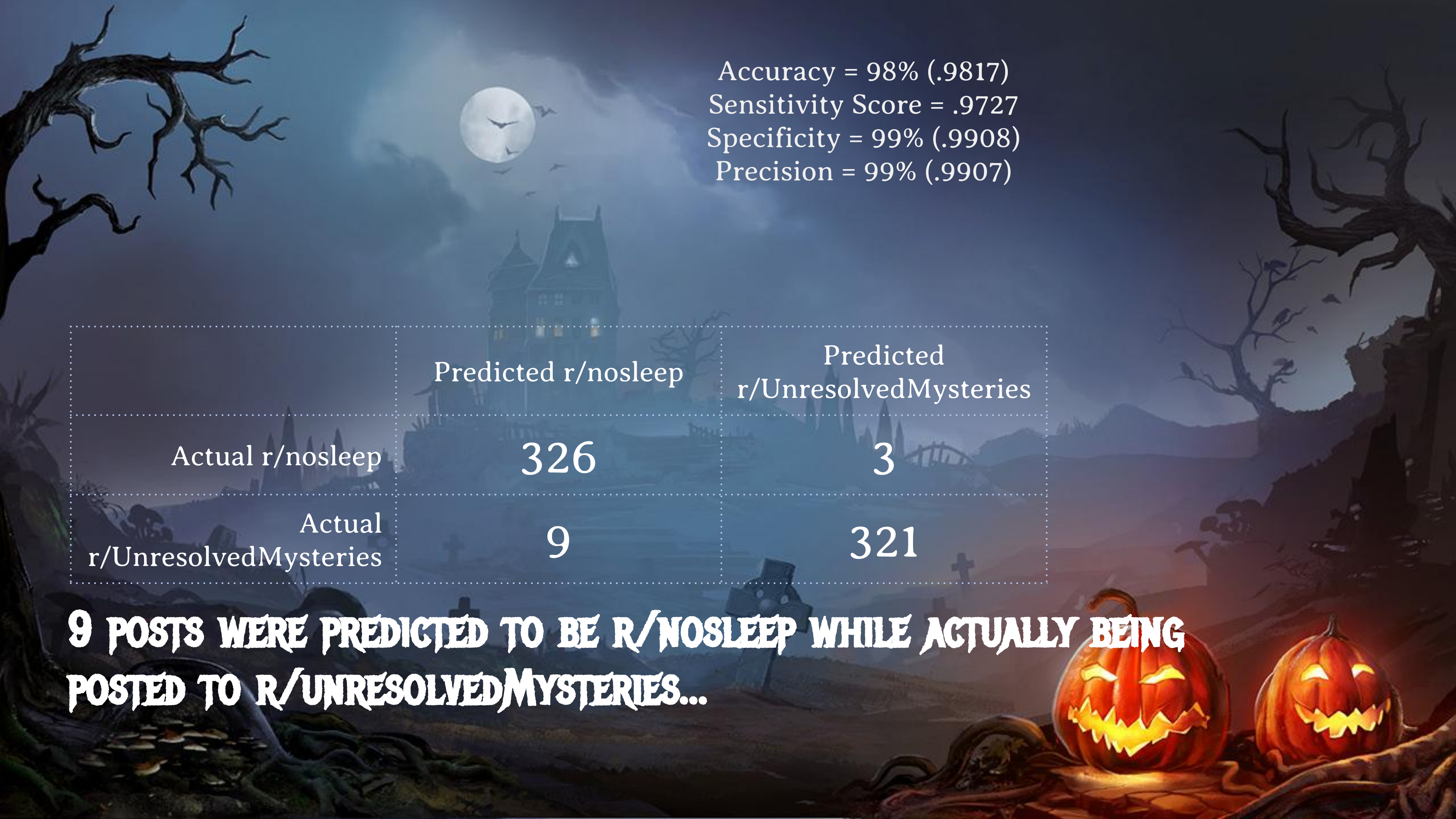
Honorable Mention :

MultiNomial Naive Bayes using TfidfVectorizer

Train Accuracy Score : 0.9827973074046372

Test Accuracy Score : 0.9666160849772383

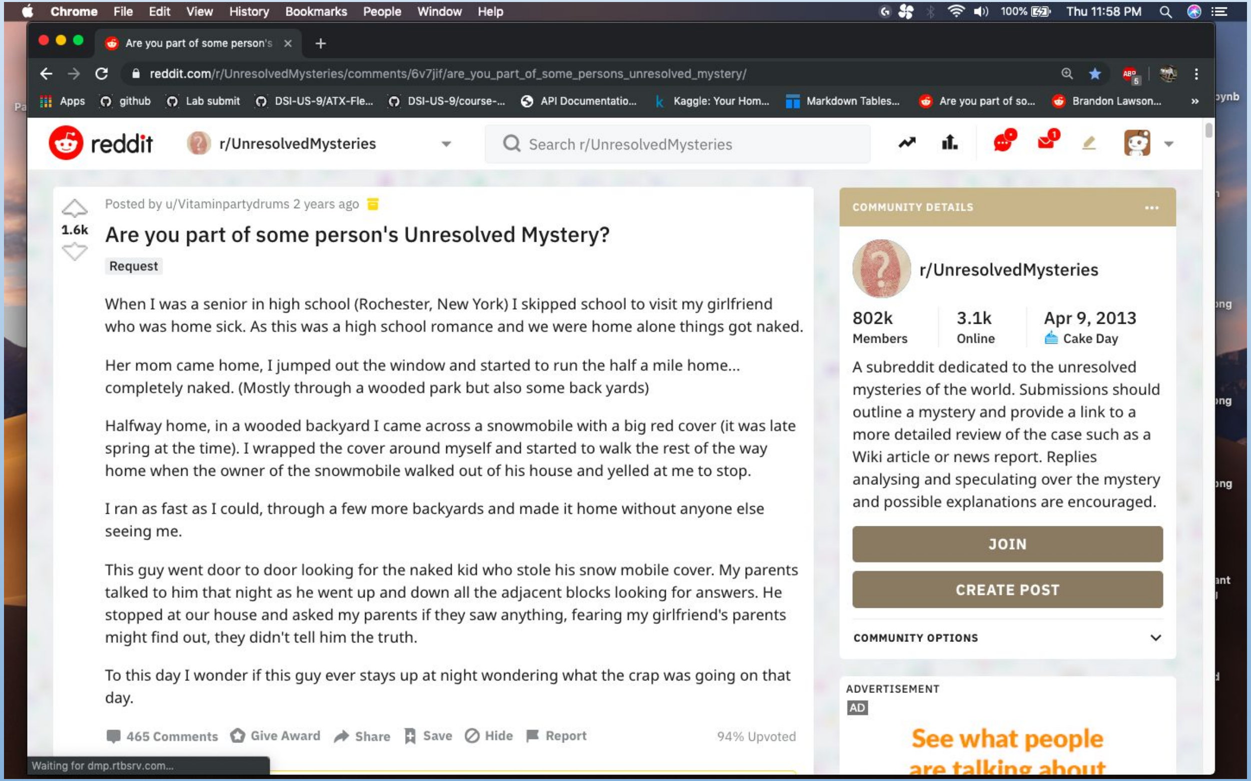
Spooky good results



Accuracy = 98% (.9817)
Sensitivity Score = .9727
Specificity = 99% (.9908)
Precision = 99% (.9907)

	Predicted r/nosleep	Predicted r/UnresolvedMysteries
Actual r/nosleep	326	3
Actual r/UnresolvedMysteries	9	321

9 posts were predicted to be r/nosleep while actually being posted to r/unresolvedMysteries...


Chrome File Edit View History Bookmarks People Window Help
100% Thu 11:58 PM
Are you part of some person's
reddit.com/r/UnresolvedMysteries/comments/6v7jif/are_you_part_of_some_persons_unresolved_mystery/
Apps github Lab submit DSI-US-9/ATX-Fle... DSI-US-9/course-... API Documentatio... Kaggle: Your Hom... Markdown Tables... Are you part of so... Brandon Lawson...
reddit
r/UnresolvedMysteries
Search r/UnresolvedMysteries
1.6k

Posted by u/Vitaminpartydrums 2 years ago

Are you part of some person's Unresolved Mystery?

Request

When I was a senior in high school (Rochester, New York) I skipped school to visit my girlfriend who was home sick. As this was a high school romance and we were home alone things got naked.

Her mom came home, I jumped out the window and started to run the half a mile home... completely naked. (Mostly through a wooded park but also some back yards)

Halfway home, in a wooded backyard I came across a snowmobile with a big red cover (it was late spring at the time). I wrapped the cover around myself and started to walk the rest of the way home when the owner of the snowmobile walked out of his house and yelled at me to stop.

I ran as fast as I could, through a few more backyards and made it home without anyone else seeing me.

This guy went door to door looking for the naked kid who stole his snow mobile cover. My parents talked to him that night as he went up and down all the adjacent blocks looking for answers. He stopped at our house and asked my parents if they saw anything, fearing my girlfriend's parents might find out, they didn't tell him the truth.

To this day I wonder if this guy ever stays up at night wondering what the crap was going on that day.

465 Comments Give Award Share Save Hide Report
94% Upvoted
Waiting for dmp.rtbsrv.com...

COMMUNITY DETAILS

r/UnresolvedMysteries

802k	3.1k	Apr 9, 2013
Members	Online	Cake Day

A subreddit dedicated to the unresolved mysteries of the world. Submissions should outline a mystery and provide a link to a more detailed review of the case such as a Wiki article or news report. Replies analysing and speculating over the mystery and possible explanations are encouraged.

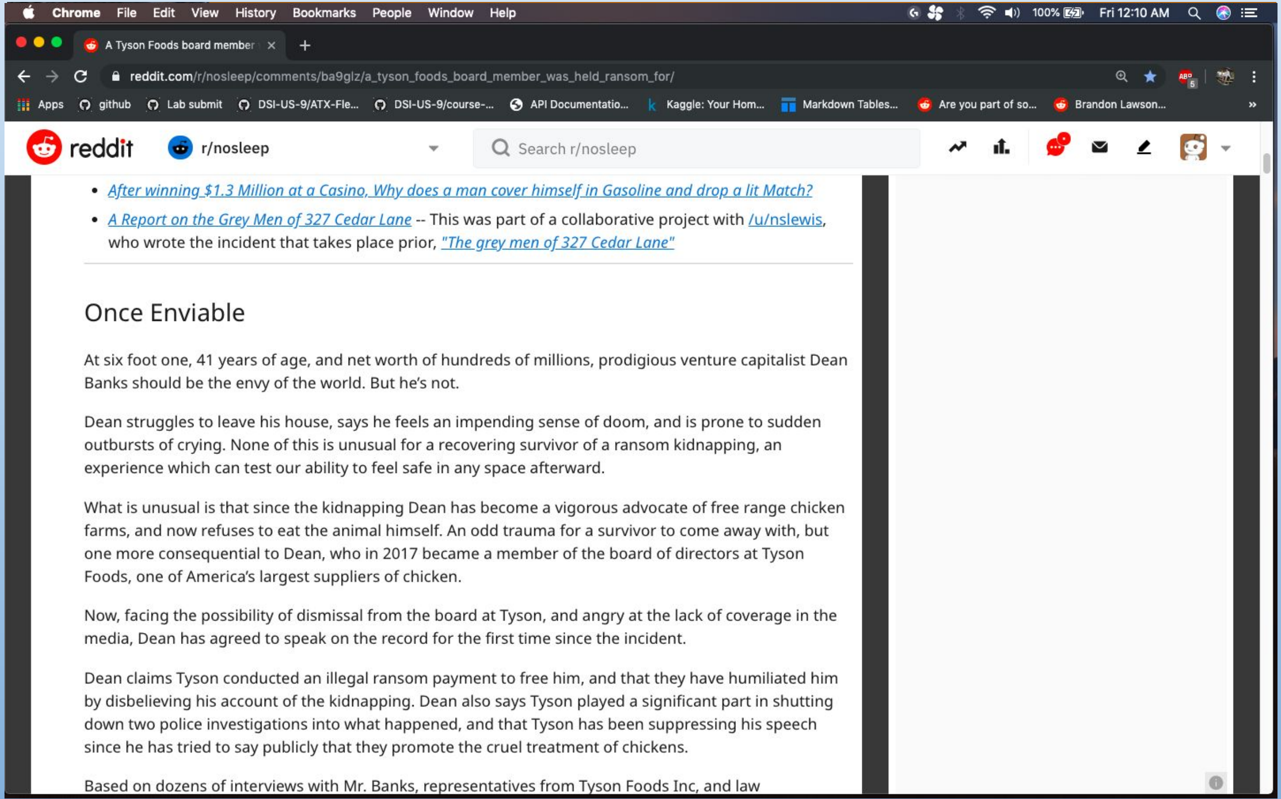
9 BANNED USERS.

My model incorrectly predicted this post to be r/nosleep.

You can immediately tell this story is mostly lacking any factual data. It’s written very casually and through the perspective of the user instead of official reports.

As an admin I don’t think this post fits the subreddit.

3 Scary good stories.


Chrome File Edit View History Bookmarks People Window Help
100% Fri 12:10 AM
A Tyson Foods board member
reddit.com/r/nosleep/comments/ba9glz/a_tyson_foods_board_member_was_held_ransom_for/
Apps github Lab submit DSI-US-9/ATX-Fle... DSI-US-9/course-... API Documentatio... Kaggle: Your Hom... Markdown Tables... Are you part of so... Brandon Lawson...
reddit
r/nosleep
Search r/nosleep

- *After winning $1.3 Million at a Casino, Why does a man cover himself in Gasoline and drop a lit Match?*
- *A Report on the Grey Men of 327 Cedar Lane* -- This was part of a collaborative project with /u/nslewis, who wrote the incident that takes place prior, *"The grey men of 327 Cedar Lane"*

Once Enviable

At six foot one, 41 years of age, and net worth of hundreds of millions, prodigious venture capitalist Dean Banks should be the envy of the world. But he's not.

Dean struggles to leave his house, says he feels an impending sense of doom, and is prone to sudden outbursts of crying. None of this is unusual for a recovering survivor of a ransom kidnapping, an experience which can test our ability to feel safe in any space afterward.

What is unusual is that since the kidnapping Dean has become a vigorous advocate of free range chicken farms, and now refuses to eat the animal himself. An odd trauma for a survivor to come away with, but one more consequential to Dean, who in 2017 became a member of the board of directors at Tyson Foods, one of America's largest suppliers of chicken.

Now, facing the possibility of dismissal from the board at Tyson, and angry at the lack of coverage in the media, Dean has agreed to speak on the record for the first time since the incident.

Dean claims Tyson conducted an illegal ransom payment to free him, and that they have humiliated him by disbelieving his account of the kidnapping. Dean also says Tyson played a significant part in shutting down two police investigations into what happened, and that Tyson has been suppressing his speech since he has tried to say publicly that they promote the cruel treatment of chickens.

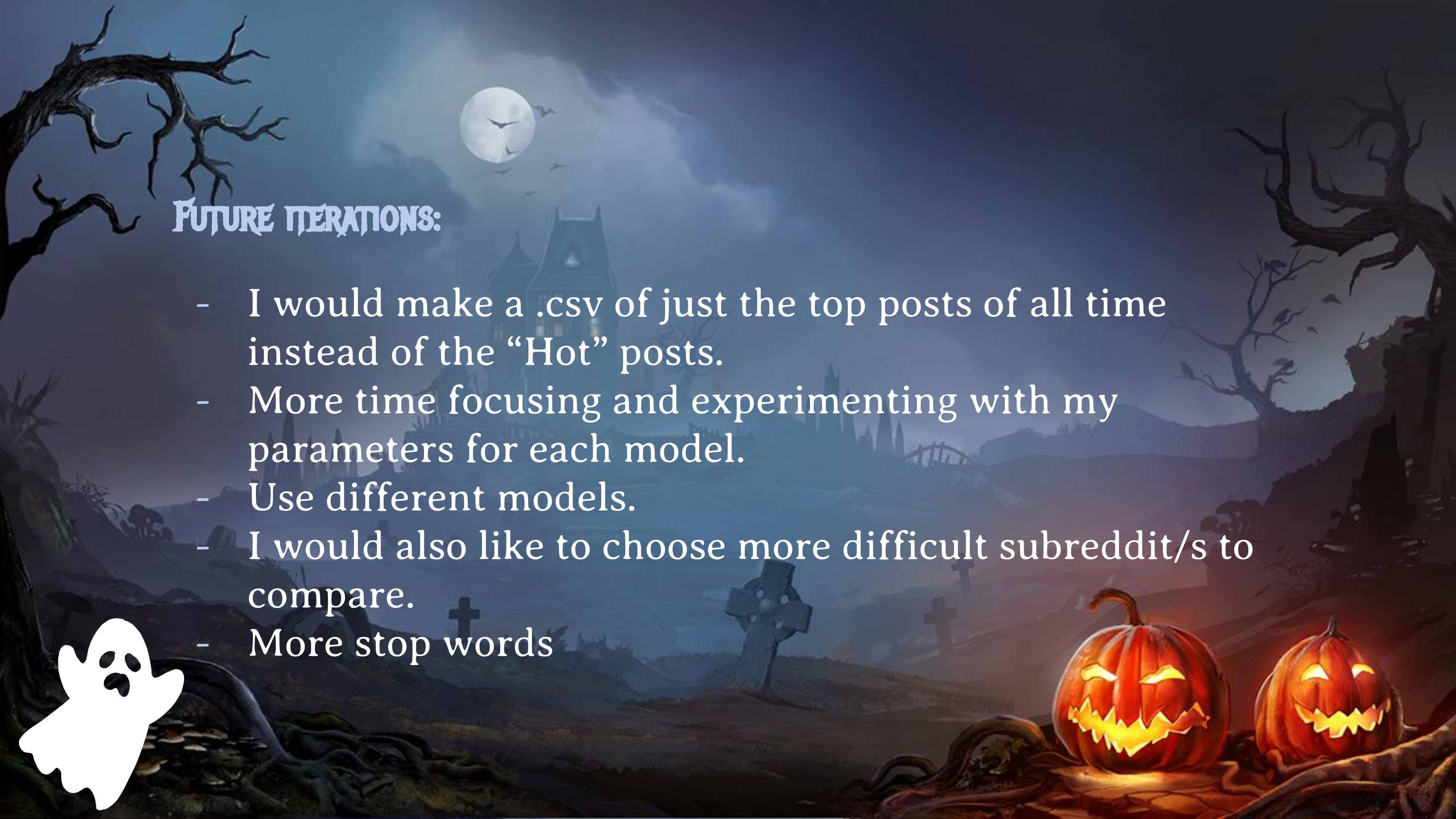
Based on dozens of interviews with Mr. Banks, representatives from Tyson Foods Inc, and law

My model predicted this post to be a from r/UnresolvedMysteries, but it was actually a fictional story written for the r/nosleep subreddit.

The story was written well and included links and dates to all events. Very detail oriented.

Conclusions:

- People aren’t very good at writing up scary stories for r/nosleep.
- Natural language processing is scary powerful and has increasing opportunity in today's growing online communities.
- If I were an admin monitoring posts for any forum or subreddit I could make great use of a model like this.



Future iterations:

- I would make a .csv of just the top posts of all time instead of the “Hot” posts.
- More time focusing and experimenting with my parameters for each model.
- Use different models.
- I would also like to choose more difficult subreddit/s to compare.
- More stop words

Questions