

Horror vs Sci Fi



A classification of posts from r/nosleep and r/shortscifistories

Problem statement

Are there a set of words that can help us define a genre?



r/nosleep

Original horror stories

VS



r/shortscifistories

Thought provoking
short sci fi stories

Methodology

Scrape + Clean

Scraped 700+ posts from both subreddits

Removed null values, duplicate, and non text posts

Preprocess

Removed digits, punctuations and symbols

Split texts into tokens

Lemmatized tokens

Modelling

Vectorized tokens using CountVectorizer / TfidfVectorizer

Used Logistic Regression, Naive Bayes and Random Forests as models

Used GridSearchCV to find best estimators and parameters

Evaluation

Decided on best model based on a set of metrics:

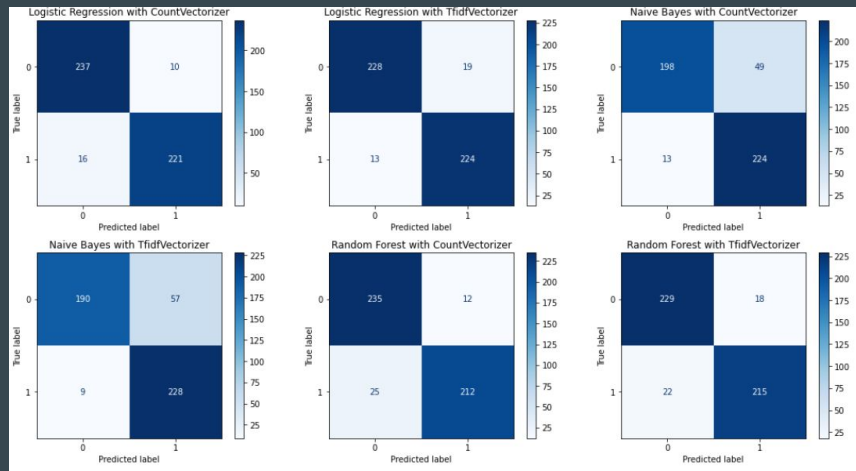
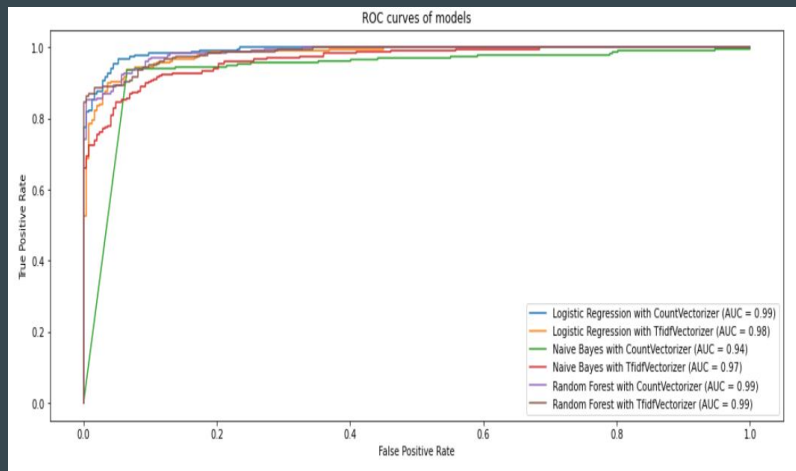
- 1) Best CV score
- 2) Delta (difference between train test scores)
- 3) ROC AUC
- 4) Confusion Matrix
- 5) Precision
- 6) Recall

Final Analysis

Analysed misclassified posts (false positives and false negatives).

Best model - Logistic Regression with TfidfVectorizer

	model	train score	test score	best score	precision	recall	f score	delta
1	Logistic Regression with TfidfVectorizer	0.966395	0.933884	0.939931	0.933934	0.934112	0.933880	0.032511
0	Logistic Regression with CountVectorizer	1.000000	0.946281	0.960282	0.946734	0.946002	0.946222	0.053719
3	Naive Bayes with TfidfVectorizer	0.920570	0.863636	0.893101	0.877387	0.865628	0.862791	0.056934
4	Random Forest with CountVectorizer	0.985743	0.923554	0.943976	0.925137	0.922966	0.923381	0.062190
2	Naive Bayes with CountVectorizer	0.938900	0.871901	0.895136	0.879451	0.873384	0.871530	0.066999
5	Random Forest with TfidfVectorizer	0.993890	0.917355	0.944996	0.917549	0.917149	0.917286	0.076535



Top 20 tokens for each class based on coefficients

coefficient	
token	
earth	-1.185525
planet	-0.969874
ai	-0.961754
world	-0.919119
ship	-0.850411
artifact	-0.826702
system	-0.826652
universe	-0.786065
human	-0.767545
robot	-0.763940
scientist	-0.740005
alien	-0.732413
daryl	-0.642421
sir	-0.631317
humanity	-0.612606
joe	-0.587793
susan	-0.567483
captain	-0.557041
smith	-0.543953
space	-0.542307

r/shortscifistories (class 0)

Most tokens related to
planets, technology,
space

coefficient	
token	
go	1.072944
even	1.079735
look	1.093486
could	1.109990
went	1.145653
friend	1.153134
room	1.190734
eye	1.191086
car	1.192582
started	1.257356
get	1.279465
something	1.301522
door	1.310932
night	1.326798
know	1.351832
like	1.455514
looked	1.462452
house	1.628495
back	1.701222
said	2.085166

r/nosleep (class 1)

Most tokens related to
senses, body parts,
houses

Misclassified posts - Analysing a false positive post

32 misclassified posts. 19 false positives, 13 false negatives.

'drip drop part see powerless lost home devoured spare pain embrace darkness gonna lie considering half body fire half invaded poisonous chill idea attractive lid eye slowly descend muscle relax resigned fate make peace end force eye open bully suffering limb raising foot stand sway cannot keep much longer see shining salvation beam bright gold light shine right side room light half face beam feel warmth righteousness cold creeping driven mad anguish pain focus see emitted ornate spear hovering rotating slowly stone floor warm antidote creature cool posison reach use put end monster calculate around metre away reach creature focused least one piece look floor see battered canteen weathered hide strap ripped buckle spread across floor like tired tadpole grab canteen thrust air left arm scream protest bone grate lightning spark pain pulse arm canteen serf purpose eye peel away zone new enemy take chance force battered body limbing jog golden spear whole frame shake step every footfall come surety next last must keep going eternity spear nearly grasp reach broken hand towards warm glow hush pain cry index finger brush smooth perfect gold fill determination win win left leg ripped feel familiar prickly thorn deathly cold invades vein time know reversing cancerous spread creature lift hold centimetre away pulsing blue orb look moment seeming feast fear hopelessness dark maw open greeted imperfect row yellowing teeth thrust smell rotting noxious interior beast begin meal feel slice fire rip string flesh bone grind cannot swallow say goodbye done quest failed game really shit pissboot jake threw weathered game controller floor digu st disappeared mountain clothes fuck sake always last sprint get stupid tentacle swipe need health potion man tom turned head look friend one eyebrow cocked dude fuck health potion never gonna reach spear red herring need new approach jake snorted golden spear final weapon cumsock else supposed beat hentai fucker last time checked schoolgirl occupy tentacle tom turn snort added chuckle boot kiss mother mouth shook head nah man told trapdoor eye need go slam sword hilt stone get use anti gravs rip kraken apart jake rolled eye yeah yeah know little theory care cousin said get anti gravs medieval fantasy game plus try kraken head tom gave knowing look sometimes gotta take leap faith trust friend mon frere either let boy two teenager looked moment fear freezing face rewind half hour mind see done anything wrong jake tentatively answered yeah mum hope going let tom keep staying going sleep time jake eye widened snatched look watch holy shit right heard jake sighed called downstairs sorry mum tom unimpressed fine man orchard like minute away think sleep time co house wall painted semen tom chuckled grabbed bag made door shut pissboat two boy left room headed downstairs forgetting turn console mr dean definitely would approved screen jake high score flashed white letter writhing green mass game final boss stared kraken roared sickly cry crushed metal noone particular crackling tv mod est speaker like ghost whispering megaphone able speak without laughed superstitious weirdo might said room changed drastically much slightly colder part series calling drip drop part linked part soon part http www reddit com r comment ww utm medium android app amp utm source share thank reading story'

```
{ 'said': 1,
  'house': 1,
  'looked': 1,
  'like': 1,
  'know': 1,
  'door': 1,
  'get': 1,
  'car': 1,
  'eye': 1,
  'room': 1,
  'friend': 1,
  'look': 1,
  'go': 1 }
```

```
{ 'ai': 1 }
```

False positive text - wrongly classified as class 1 (r/nosleep)

Count of top 20 class 1 tokens that appear in text

Count of top 20 class 0 tokens that appear in text

coefficient	
token	
earth	-1.185525
planet	-0.969874
ai	-0.961754
world	-0.919119
ship	-0.850411
artifact	-0.826702
system	-0.826652
universe	-0.786065
human	-0.767545
robot	-0.763940
scientist	-0.740005
alien	-0.732413
daryl	-0.642421
sir	-0.631317
humanity	-0.612606
joe	-0.587793
susan	-0.567483
captain	-0.557041
smith	-0.543953
space	-0.542307

Side quest - who or what is Daryl?

A protagonist created by an active redditor on r/shortscifistories

↑ 15 ↓	[micro] Daryl Soon Disappeared r/shortscifistories · Posted by u/BobiOSGamer 5 months ago 2 Comments Award Share Save ...
↑ 13 ↓	[micro] Alas, None of It Ever Happened r/shortscifistories · Posted by u/BobiOSGamer 10 days ago 4 Comments Award Share Save ...
↑ 15 ↓	[micro] Hung by a Rope r/shortscifistories · Posted by u/BobiOSGamer 7 days ago 2 Comments Award Share Save ...
↑ 9 ↓	[micro] Alone at Last! r/shortscifistories · Posted by u/BobiOSGamer 29 days ago 8 Comments Award Share Save ...
↑ 14 ↓	[micro] The Fate of the Universe Remained Sealed r/shortscifistories · Posted by u/BobiOSGamer 23 days ago 2 Comments Award Share Save ...
↑ 24 ↓	[micro] What About the Duplicates? r/shortscifistories · Posted by u/BobiOSGamer 1 month ago 4 Comments Award Share Save ...
↑ 22 ↓	[micro] Grandfather, Again r/shortscifistories · Posted by u/BobiOSGamer 1 month ago 4 Comments Award Share Save ...
↑ 14 ↓	[micro] Wait, I Know How To Swim! r/shortscifistories · Posted by u/BobiOSGamer 2 months ago 10 Comments Award Share Save ...
↑ 9 ↓	[micro] Soulmate r/shortscifistories · Posted by u/BobiOSGamer 1 month ago 2 Comments Award Share Save ...
↑ 11 ↓	[micro] 'Justice' r/shortscifistories · Posted by u/BobiOSGamer 1 month ago 2 Comments Award Share Save ...

Posted by u/BobiOSGamer 2 months ago

↑
14
↓

[micro]

Daryl invented the 'Inspector,' an anti-paradox time travel device. His consciousness transferred into a host's body to share the same experience, non-interactively.

Daryl's first goal was to find who discovered fire.

He unraveled millions of years of historical links in the chain of causality.

A group of early humans huddled in a cave for warmth; then lightning caught a tree on fire.

Daryl's caveman grabbed a scorchingly hot burning stick, sprinted to a lake, and soaked his hands. But then he saw a reflection of Daryl and attacked the lake.

"Wait, I know how to swim!" said Daryl while drowning.


10 Comments Award Share Save Hide Report 94% Upvoted

Comment as [kevowez](#)

What are your thoughts?

B i ↻ ↺ </> A⁶ ! T ⋮ ... Markdown Mode Comment

SORT BY BEST

 yaxriifgyn Time traveler 2 months ago

LOL. Daryl is as dumb as the caveman. Discovering fire is not the same as using fire!

↑ 4 ↓ Reply Give Award Share Report Save

Conclusion

Certain sets of words define the respective genres and set them apart from each other.

Some of the words that define the scifi genre are words related to technology, the galaxy and planets, e.g. ai, system, robot, earth etc.

Words that define horror genres seem to be related to either the senses, e.g. seeing or hearing something, or to home related words like room, house, and door.

Misclassifications occur sometimes because tokens from one class occurs more frequently than tokens from the other class.

Thank
you