



Using Machine Learning to Automate Reddit Posts

Uri Eckmann

Issues With Moderating



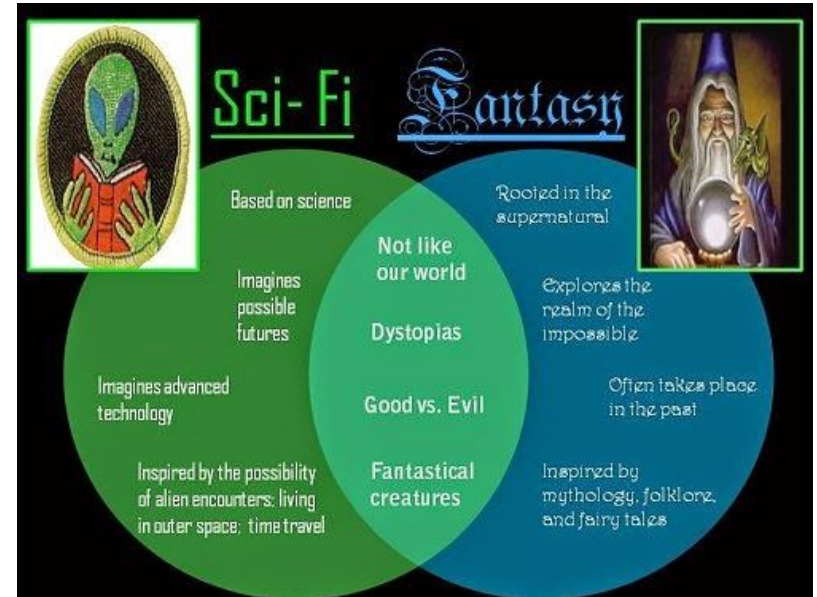
- Too Much Crossposting
- Too Many Posts
- Subject to Human Error/Bias
- Can be Automated
- Save Money, and Users



Fantasy vs. Scifi



- Pulling from r/geekdom
- Splitting between r/fantasy and r/scifi



Workflow

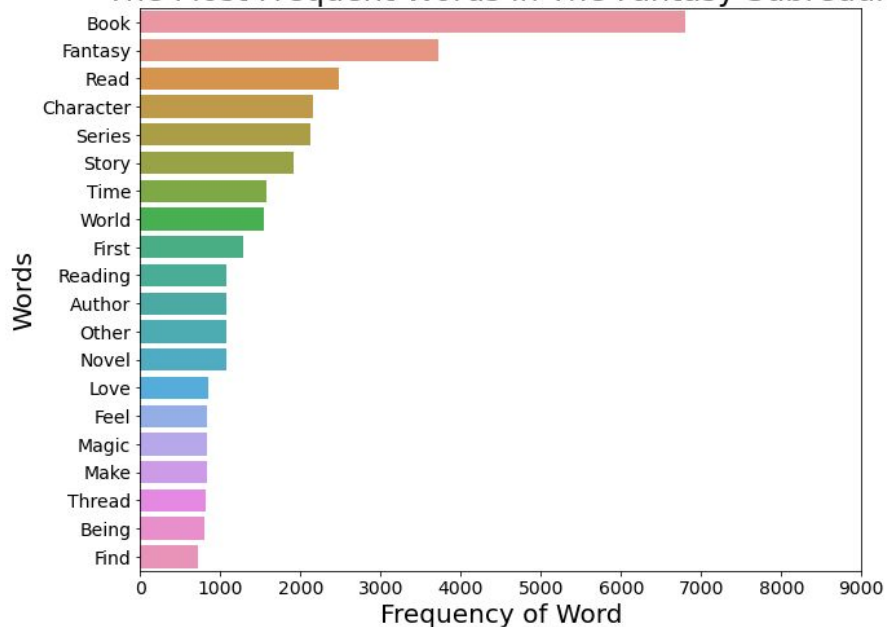


- Extracted Data from the Reddit API
- Cleaned the Data
- Utilized Natural Language Processors
- Exploratory Analysis
- Created Models
- Evaluated Models

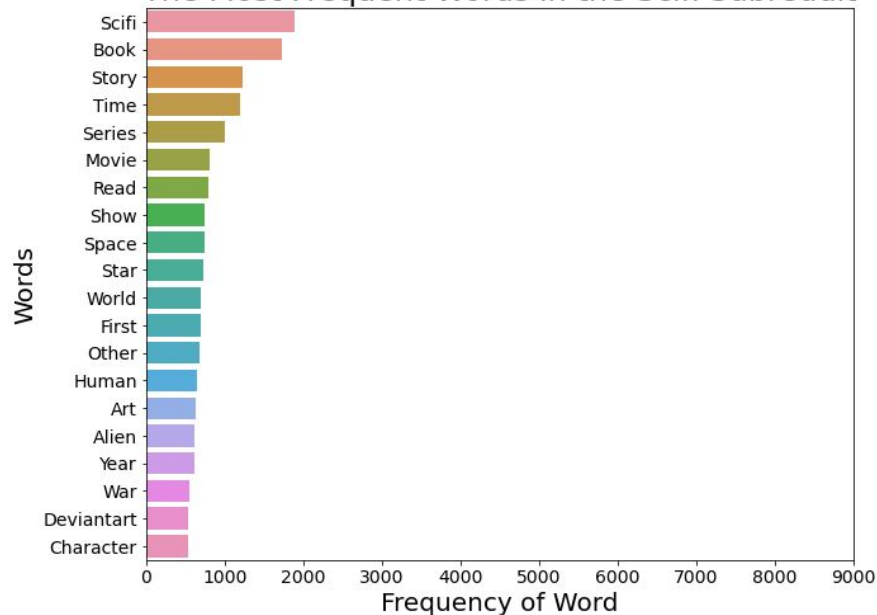
Data Exploration



The Most Frequent Words in The Fantasy Subreddit



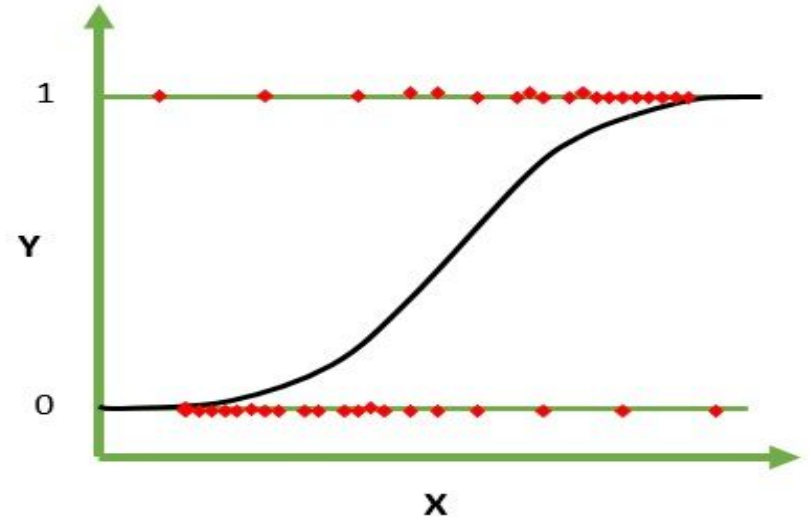
The Most Frequent Words in the Scifi Subreddit



Models



- Baseline
- Logistic Regression
- Decision Tree
- Bagging
- Random Forest
- Gradient Boost



Model Evaluation



- Logistic Regression
- Training Data Accuracy: 98.8%
- Testing Data Accuracy : 89.0%
- Cross Validation Mean : 89.6%
- Model is Overfit

Model Coefficients



	Variable	Logistic Regression Coefficient
2844	fantasy	1.831267
4513	magic	0.720911
933	bingo	0.669893
6395	sanderson	0.653388
8111	wheel	0.626592
4172	king	0.593631
2309	dream	0.565807
7139	style	0.541990
1828	creature	0.540269
2031	demon	0.508513

	Variable	Logistic Regression Coefficient
6446	scifi	-1.892536
6442	science	-0.911147
399	alien	-0.894448
3177	future	-0.878793
6887	space	-0.730566
5479	planet	-0.686241
2728	expanse	-0.685375
2601	episode	-0.656116
7611	travel	-0.640754
5542	poll	-0.625749

Model Misclassification



- 139 Misclassified

	Post	Predicted Subreddit	Actual Subreddit
3856	A question about Looper (spoilers). I get the ending, but there's one thing I didn't get. How is the Gat guy able to catch Old Joe at the second kids house? I believe he tells Jeff Daniels that...	1	0
1566	The Prophecy ~ A sci fi / fanatasy Story Hi \n\n\nThis is my first post here on reddit. I have wrote a sci fi fantasy story and have created it as an audiobook. My aspiration would be for it ...	0	1
2033	The Long Way to a Small Angry Planet is \$2.99 Today on Kindle! [https://www.amazon.com/Long-Small-Angry-Planet-Wayfarers-ebook/dp/B00ZP64F28/ref=tmm_kin_swatch_0?_encoding=UTF8&qid=159039...	0	1
4746	Is "Vacuum tube punk" an existing term ? I watched a Chinese video which introduced the game "the Outer Worlds" and it mentioned the term "Vacuum tube punk".\nAfter searching with Google I found ...	1	0
455	Merlin: Starring Sam Neil Does anybody remember the mini-series 'Merlin' starring Sam Neil? Apparently it aired on NBC, but over this side of the Atlantic, I think I watched it on Channel 4.\n\nA...	0	1

Conclusions



- Collect More Data
- Try More Models
- Work With More Hyperparameters