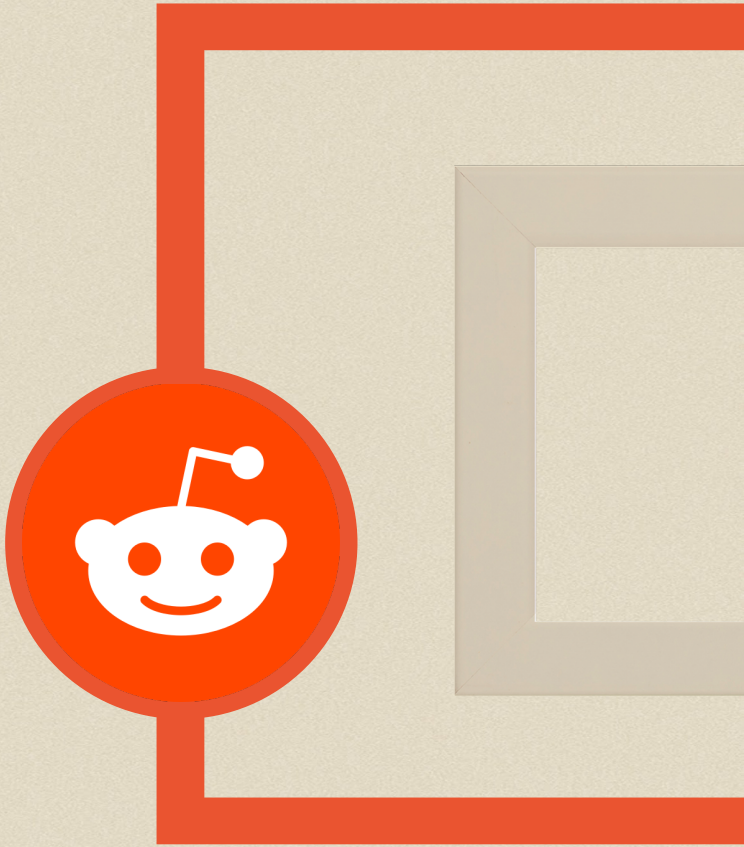
Is it Reddit's genuine advice or fantasy?

A model that can discern the language between fiction and fantasy



Subreddits

r/LifeProTips

Tips shared on any topics that may improve someone's life.

"LPT: wear earplugs to loud concert venues. Tinnitus is real and not fun."

r/godtiersuperpowers

Users sharing their fantasy superpowers.

"You instantly get 10 years of experience in whatever skill you want"

Overview of the dataset

Subreddit	Class Balance	Binarized:
LifeProTips	50.47%	1
godtiersuperpowers	49.52%	0

Subreddits sorted by:

‘hot’

‘top’

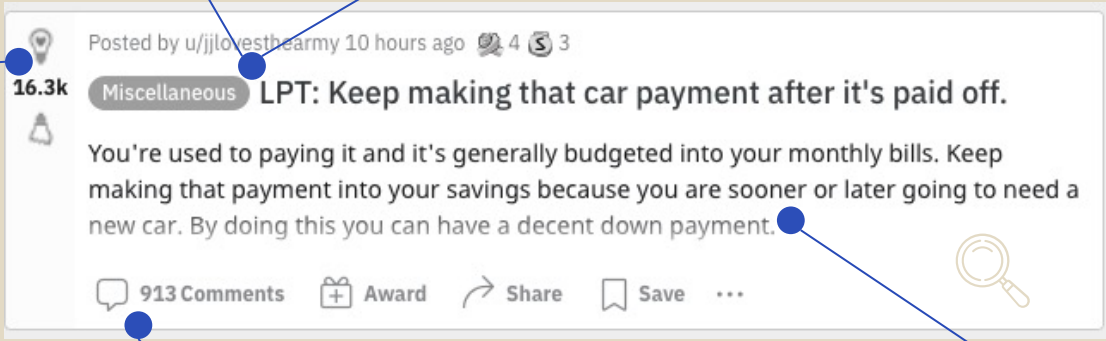
‘new’

Features

Title
Post ‘language’

Title word count
Post ‘length’

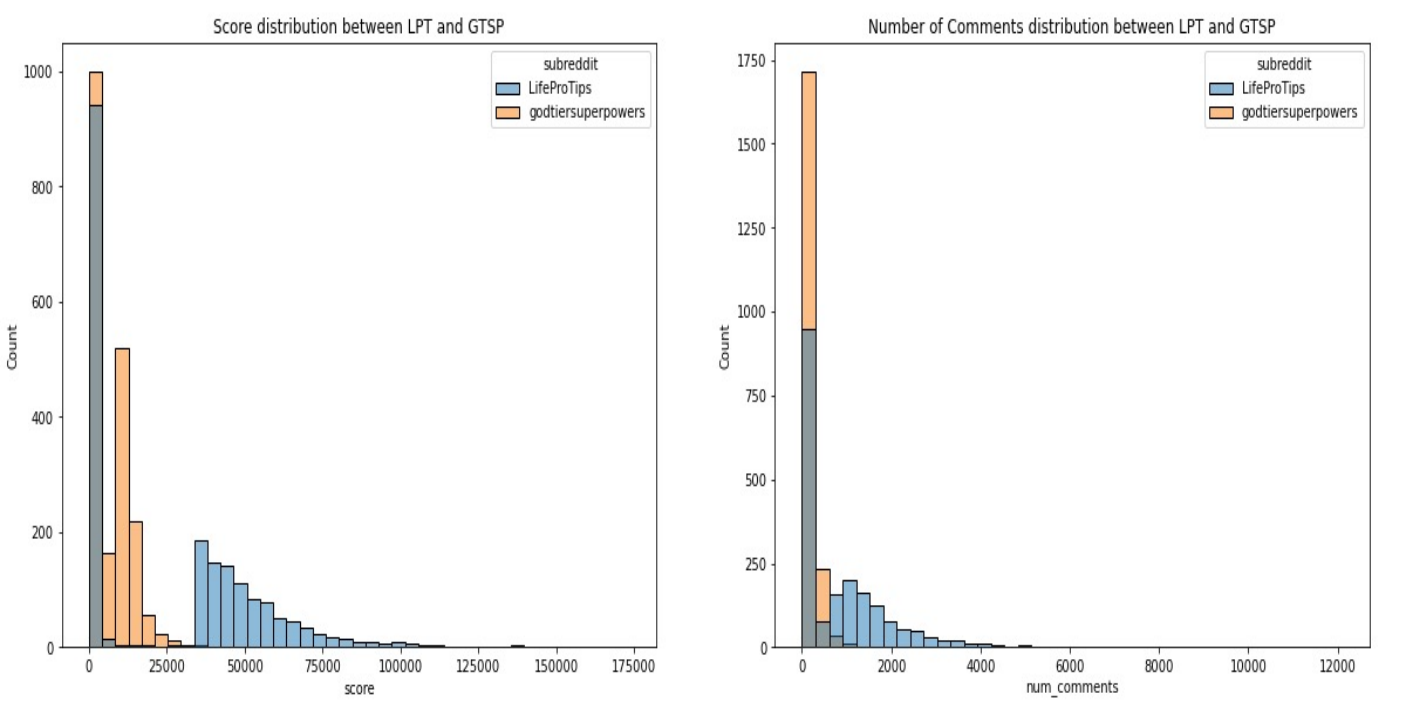
Score
Post ‘popularity’



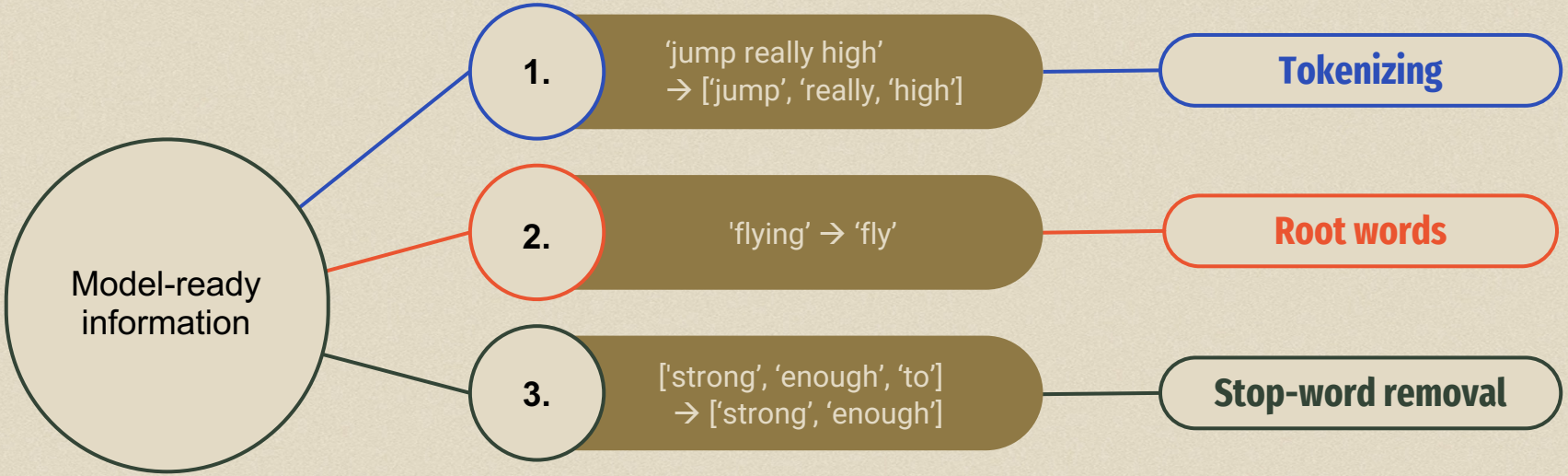
Self text
Post ‘detail’

Number of comments
Post ‘engagement’

Characteristics of the subreddits:



Pre-processing ‘Titles’



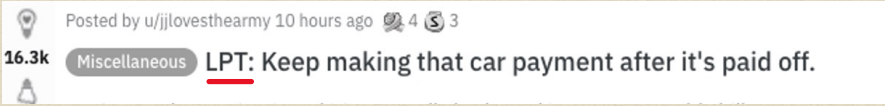
Most popular words in subreddit titles

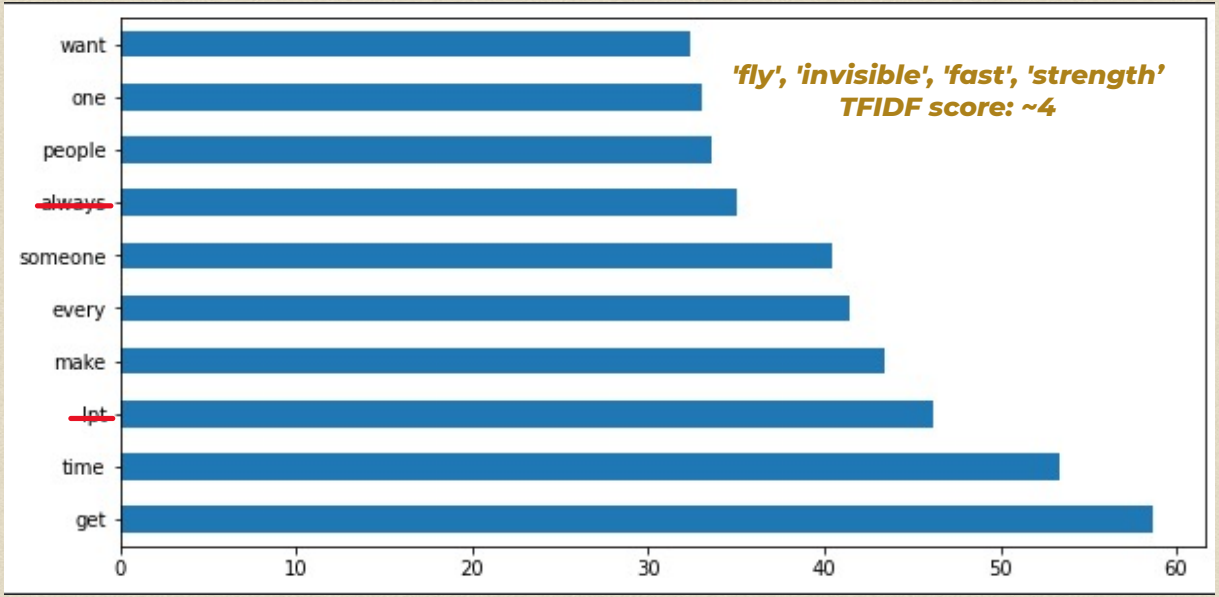
'always'

LPT: 51%
gottiersuperpowers: 49%

'people'

LPT: 63%
gottiersuperpowers: 37%





Chosen model:

- ForestTree Classifier

ForestTree Classifier metrics

0.91

Recall

Around 1 in 10 of actual LifeProTips entries were misclassified as godtiersuperpowers

Accuracy

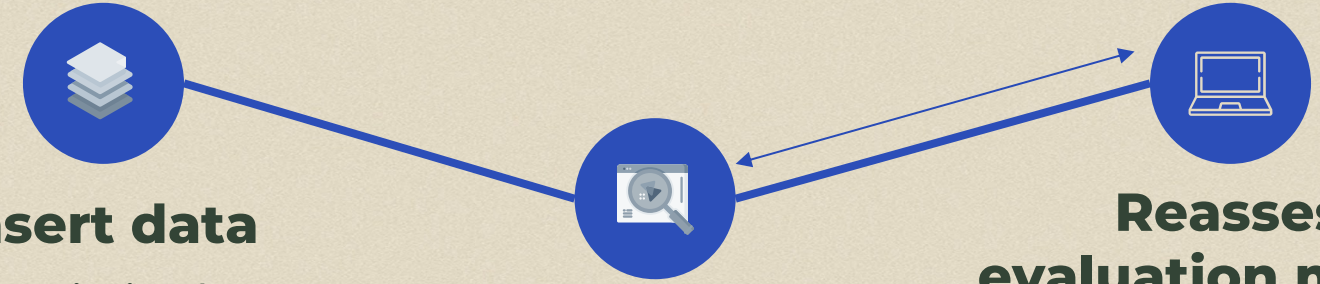
A test accuracy of 0.91 means that 9% of the variance cannot be explained by our model.

Specificity

0.93

93% of our actual godtiersuperpower entries were classified correctly

Model improvement



Insert data

Feed optimized pre-processed language using NLP and feature selection

Hyperparameter Tuning

Adjust NLP and model hyperparameters

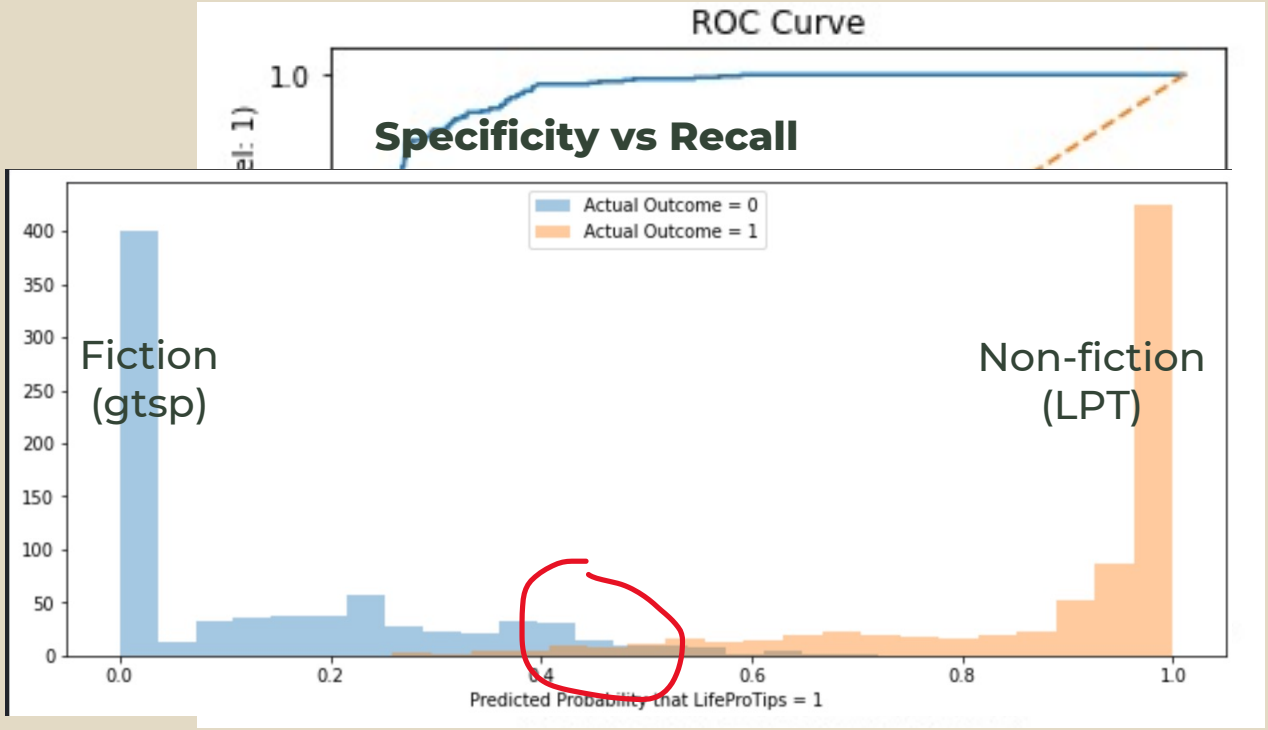
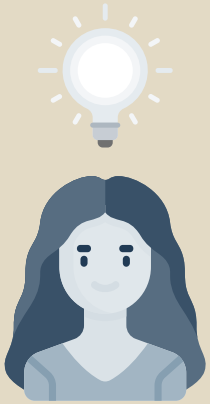
Reassess evaluation metrics

1. Reduce overfitting
2. Achieve highest recall and specificity scores

ROC curve

Area under curve (AUC)

0.97



Conclusion and Recommendations