CLASSIFICATION MODEL

For subreddit "Flair" suggestion

FLAIR

"TAG" of posts in subreddit

---

r/productivity · Posted by u/productivitydatabase 3 hours ago

1

Have you guys heard of the Temporal Motivation Theory? I think I have found a scientific approach to overcoming procrastination `Technique`

I'm not sure if this is common knowledge but I have stumbled upon an academic model for measuring and managing procrastination. I believe many books (e.g. Atomic Habits) may have used principles of this model to guide their message, and I was quite impressed with the research I found. What do you guys think of this model?

💬 1 Comment   ↗ Share   🔖 Save   ...

---

r/productivity · Posted by u/Apptubrutae 3 hours ago

1

Strategies for dealing with incoming paper? `General Advice`

Anyone have any tips for keeping incoming papers from accumulating in an office environment?

I don't get a lot of physical paper across my desk, but when it does happen, I never have figured out how to sort and store them so they just sort of pile up.

It's a variety of things like digitally deposited checks, tax records, receipts, etc. All things that I want to keep but not revisit with minimal exceptions (like say some of those tax documents I might need months down the road).

I'd break it down like so:

- Documents I want to hold to do something with later that I don't want to make hard to find (5%)

💬 0 Comments   ↗ Share   🔖 Save   ...

---

r/productivity ⤴ Crossposted by u/taskade-narek 3 hours ago

1

Automation: Triggers, Actions, & Use Cases (Mods Please Remove if not allowed) `Question`

reddit.com/r/Task... ⎘

THE PROBLEM

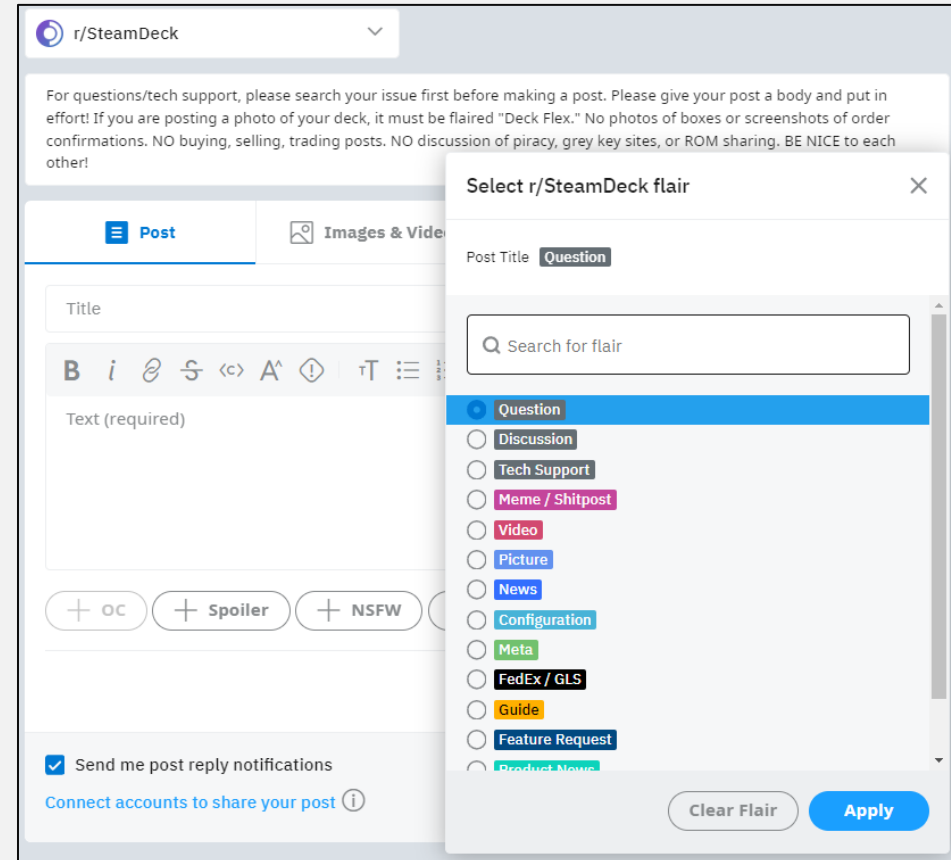Creating new post in subreddit

Pick one flair incorrectly

Reader confused

THE DATA

Column	Description
Title	Title of post on subreddit
Selftext	Text body of the post
Tag	"Flair" name of subreddit

From Steam Deck subreddit

TAG TO BE CLASSIFIED

Configuration

Discussion

Question

Tech Support

MEGATHREAD
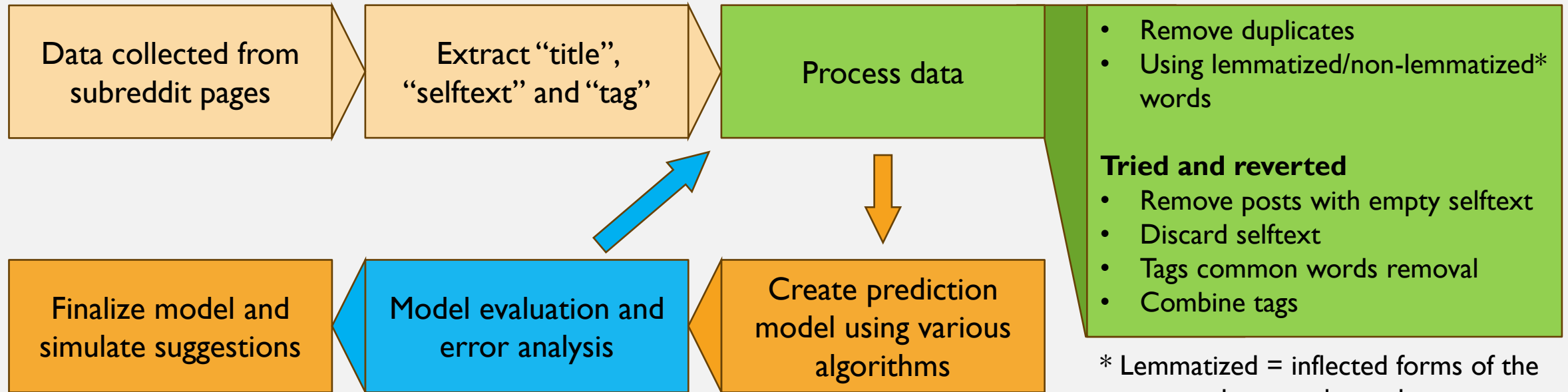
Feature Request

Guide

Hot Wasabi

Meme / Shitpost
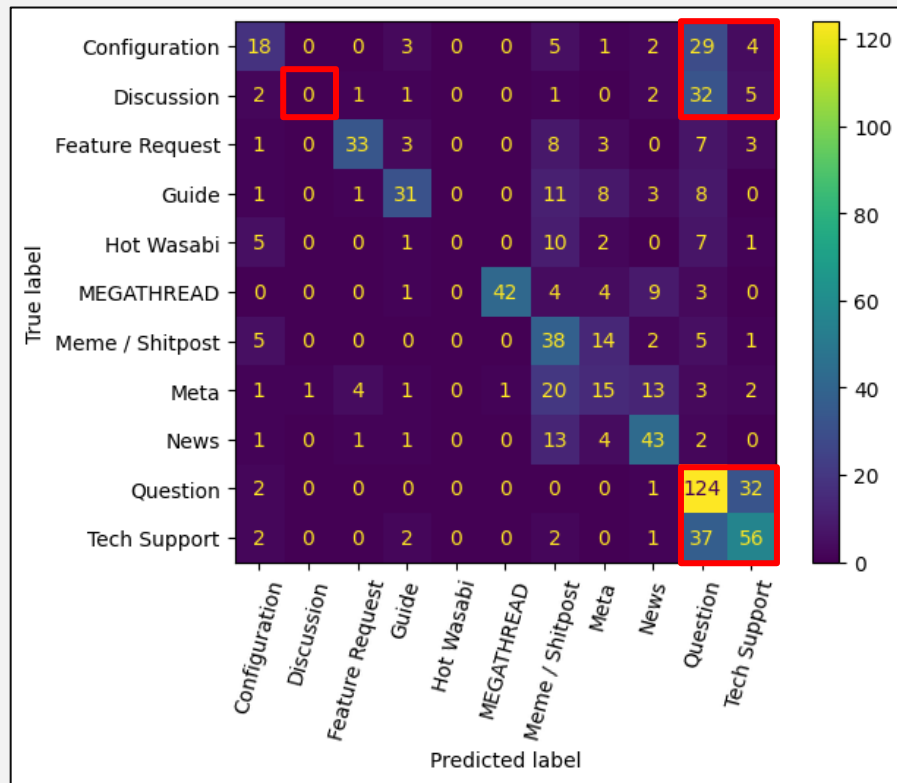
Meta

News

Picture

Video

PROCESS

```
Data collected from     Extract "title",          Process data
subreddit pages         "selftext" and "tag"
```

- Remove duplicates
- Using lemmatized/non-lemmatized* words

Tried and reverted
- Remove posts with empty selftext
- Discard selftext
- Tags common words removal
- Combine tags

```
Finalize model and      Model evaluation and      Create prediction
simulate suggestions    error analysis            model using various
                                                  algorithms
```

* Lemmatized = inflected forms of the same word grouped together

MODEL SELECTION

Precision + Correctness Score (%)

Algorithm	Data	Train data	Test data
Random Forest	Over 20,000 non-lemmatized words from title and selftext	99.7%	50.3%
Naïve Bayes	6,000 non-lemmatized words from title and selftext	86%	51.1%
Logistic Regression	2,000 lemmatized words from title and selftext	73.6%	52.2%
Logistic Regression	1,000 lemmatized words from title and selftext	68%	49.9%

*Words from title and selftext are considered to be different, even if it is the same word

THE ERRORS

- Most errors on
  - Configuration
  - Discussion
  - Question
  - Tech Support

- Some are mostly the same things.

- Many off-topic posts in a tag

FIXING THE PROBLEM

- The model will be used to generate top N tags for new post instead if it doesn't confident enough to pick only one tag.

Actual	Suggestion (suggest 1 tag if confident >= 50%)	Explanation
Guide	• Meme / Shitpost (30%) • News (12%) • Guide (11%)	Model cannot judge, suggests top 3 possible tags, one matched
Meme / Shitpost	• Question (13%) • Configuration (13%) • Tech Support (11%)	Model cannot judge, suggests top 3 possible tags, no match
MEGATHREAD	MEGATHREAD (71%)	Model select MEGATHREAD as a predicted result

RESULT

By setting confident percentage = model performance + some margin

(52.2% + 5% = 57.2%)

Percentage of relevant suggestion

goes up to **78.2%**, which should be enough for the problem

```python
# Compare with actual
for i in range(len(y_test)):
    print('Actual:', list(y_test)[i], '--- Suggest:', predictions[i])
```
✓ 0.0s

```
Actual: Hot Wasabi --- Suggest: ['Meme / Shitpost', 'Meta', 'Question']
Actual: Hot Wasabi --- Suggest: ['Tech Support', 'Configuration', 'Question']
Actual: Guide --- Suggest: ['Meme / Shitpost', 'News', 'Guide']
Actual: Question --- Suggest: ['Question', 'Tech Support', 'Feature Request']
Actual: Guide --- Suggest: ['Question', 'Tech Support', 'Configuration']
Actual: Meme / Shitpost --- Suggest: ['Question', 'Configuration', 'Tech Support']
Actual: Tech Support --- Suggest: ['Tech Support', 'Question', 'Discussion']
Actual: Hot Wasabi --- Suggest: ['Question', 'Meme / Shitpost', 'Configuration']
Actual: Question --- Suggest: ['Question', 'Discussion', 'Tech Support']
Actual: Configuration --- Suggest: ['Guide', 'Feature Request', 'Configuration']
Actual: Feature Request --- Suggest: ['Feature Request', 'Meme / Shitpost', 'Meta']
Actual: MEGATHREAD --- Suggest: ['MEGATHREAD', 'Feature Request', 'Tech Support']
Actual: Question --- Suggest: ['Question', 'Discussion', 'Tech Support']
Actual: MEGATHREAD --- Suggest: ['MEGATHREAD', 'Meme / Shitpost', 'News']
Actual: News --- Suggest: ['News', 'Meme / Shitpost', 'Meta']
Actual: Tech Support --- Suggest: ['Meme / Shitpost', 'Question', 'Meta']
Actual: Discussion --- Suggest: ['Question', 'Meta', 'Meme / Shitpost']
Actual: Question --- Suggest: ['Question', 'Tech Support', 'Meta']
Actual: Guide --- Suggest: ['Guide', 'News', 'Meme / Shitpost']
Actual: Question --- Suggest: ['Question', 'Tech Support', 'Configuration']
Actual: Question --- Suggest: ['Tech Support', 'Question', 'Configuration']
Actual: News --- Suggest: ['News', 'Meme / Shitpost', 'Meta']
Actual: Question --- Suggest: ['Tech Support', 'Question', 'Configuration']
Actual: Question --- Suggest: ['Configuration', 'Question', 'Tech Support']
Actual: Meme / Shitpost --- Suggest: ['Meme / Shitpost', 'Meta', 'News']
...
```

Confident percentage and number of suggested tags can also be adjusted to achieve even better result.

LIMITATIONS

- The model is trained on only one subreddit, it will not work with the other subreddit. To create model for another subreddit, we may have to process the data from other subreddit differently.

- The model will perform worse on tag (flair) with many off-topic posts.

- To maintain the performance of tag suggestion, the model need to be retrain periodically as posts in subreddit should be continuously growing with a little bit different of language used.