



Flair-It

...

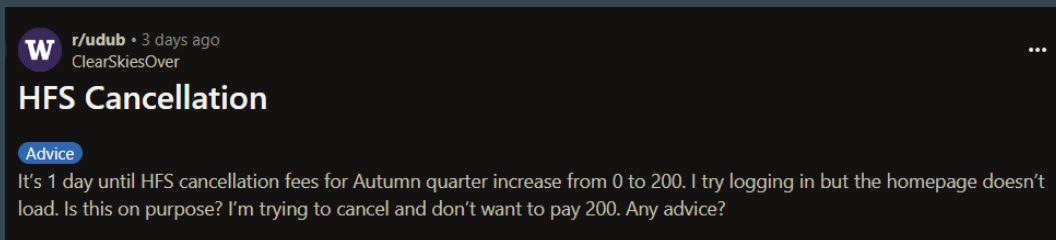
Alex, Cade, & Tyler

Team Introduction

- Alex
 - 3rd year, Informatics, transformer model lead and demo web dev
- Cade
 - 3rd year, Informatics(Data Science), Data curation and project management/scope
- Tyler
 - 4th year, Informatics(Data Science) and data cleaning lead/probabilistic model trainer

Problem Context

Reddit is not just a social media platform. It brings people to have discussions, share their opinions, give and get advice.. Depending on the subreddit of your interest, the posts can be tagged for filtering purposes called “flairs”. Flairs can be helpful when searching through the subreddit to find the most relevant questions/answers and information. However, it currently requires a manual input by the users and some may not know which flair their content fits most into when making a post on Reddit.



Problem Statement

How can we automate post flair classification on Reddit using NLP and text classification techniques to accurately assign pre-established flairs, given the current challenges of manual flair assignment and the ambiguity of many posts?

Key Research Insights

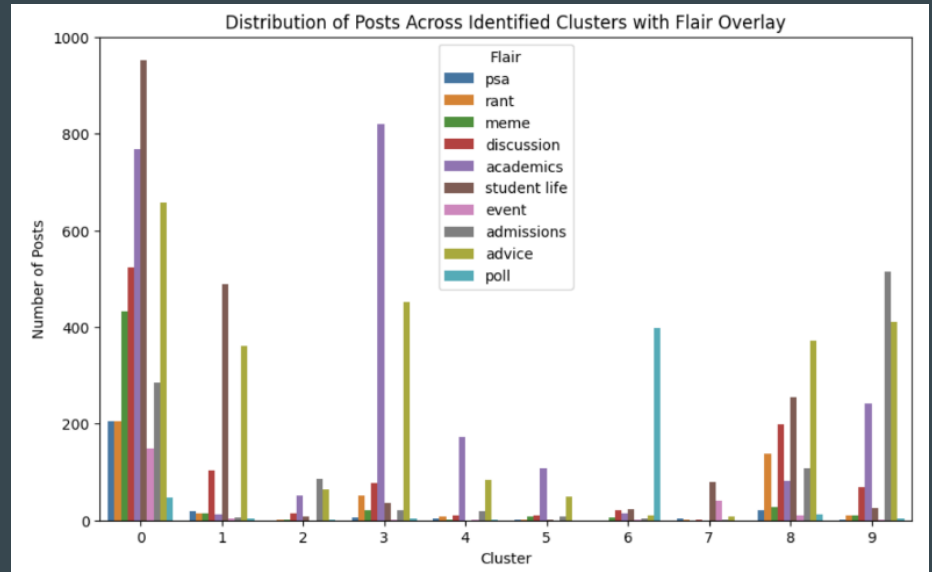
- Transformer models are good for text classification
- Data cleaning is very important/necessary to get good results
- Post flairs can be very ambiguous

Model results

Model	Baseline	Naive bayes	SVC	Logistic	Roberta
Accuracy	23.32%	54.23%	53.05%	53.09%	59.07%

Ethical Considerations

- Data privacy issues
- Model bias and potential to be controversially flaired
- Data collection could go against Reddit's policies



Next Steps Beyond Capstone

- Building product using our model
- Enhancing our model using other transformers
- Experimenting with other subreddits

Post Flair Classifier

visiting for the summer

hi! basically the title i'll be staying on campus for 6 weeks for an internship. i just wanna know some fun spots i should check out from a students pov (on campus or off)! i also wanted to know what the rock climbing scene is like if anyone who sees this knows anything about it! ty for any suggestions ❤️👍📱

Predicted label: ['advice', 'student life']