

Comment Classification Project

Kai Uechi



Summary

- Create model for automated content moderation
- Final model has 91% accuracy
- Explainable model gives insight into moderation

Outline

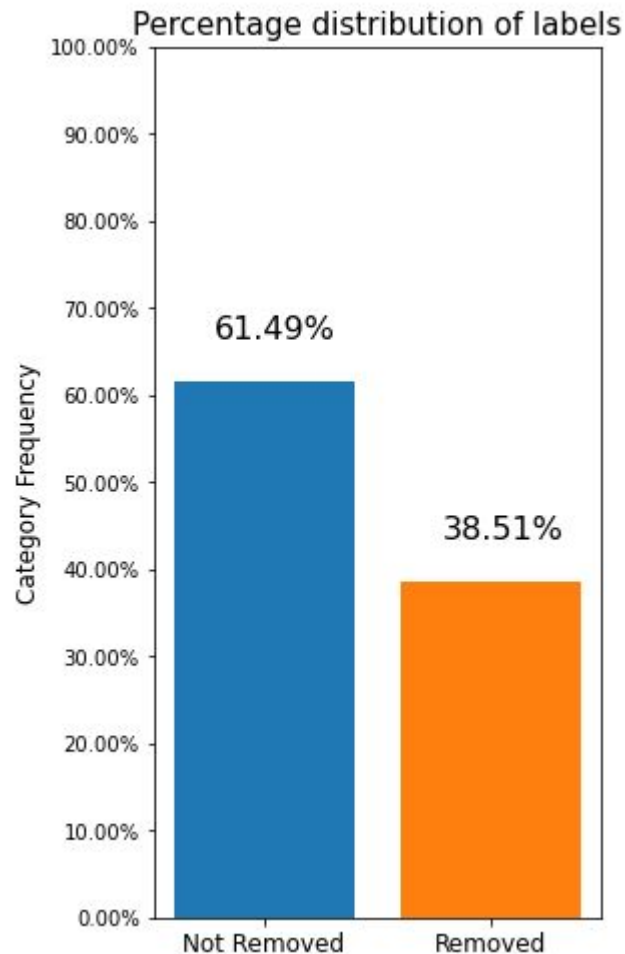
- Business Problem
- Data
- Methods
- Results
- Conclusions

Business Problem

- Looking to expand
- More users would overwhelm moderation
- Automate moderation using machine learning

Data

- 223,372 comments from r/science
- Length and content vary greatly
- 38.5% removed by moderators



Methods

Data

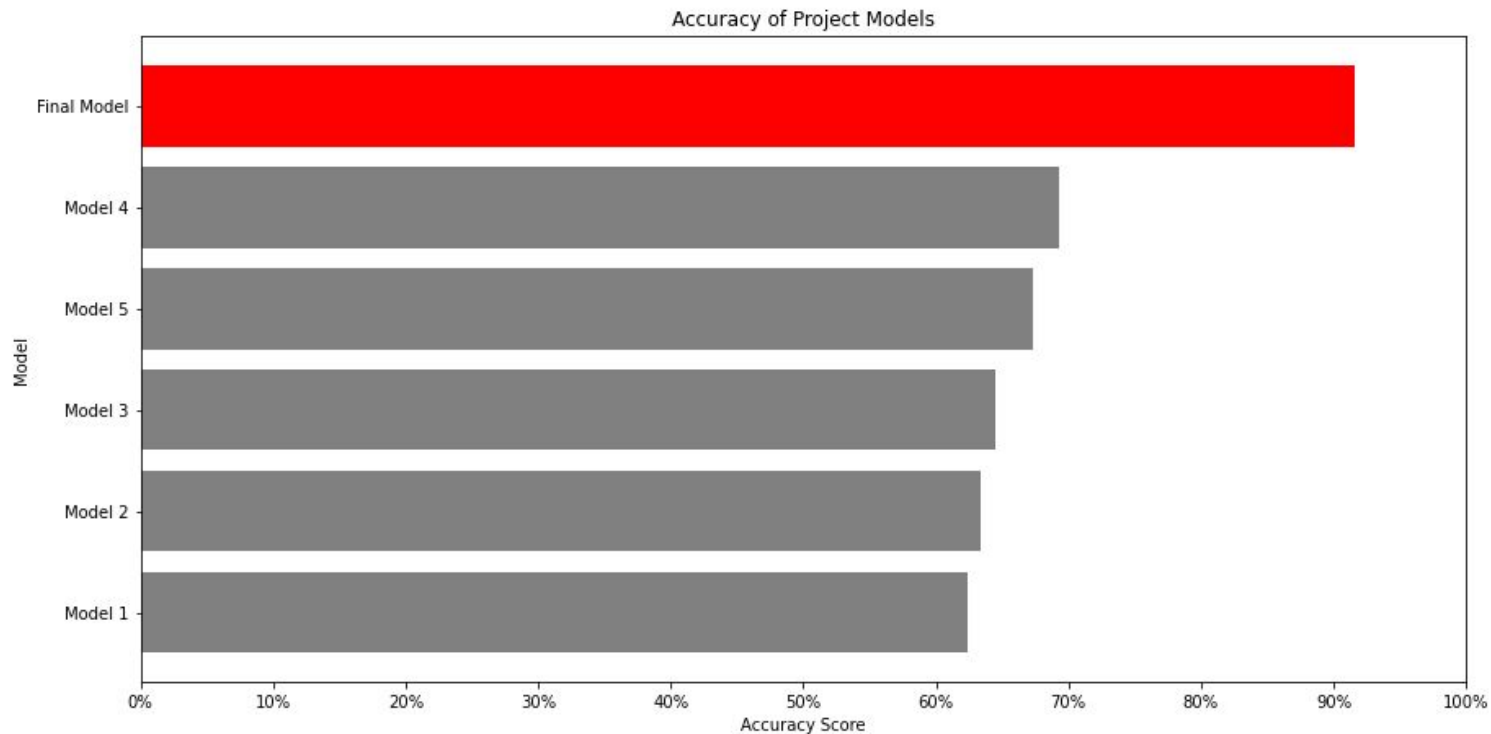
- Clean text

Modeling

- Naive Bayes
- Long Short-Term Memory
 - Word2Vec, GloVe
- Transformer

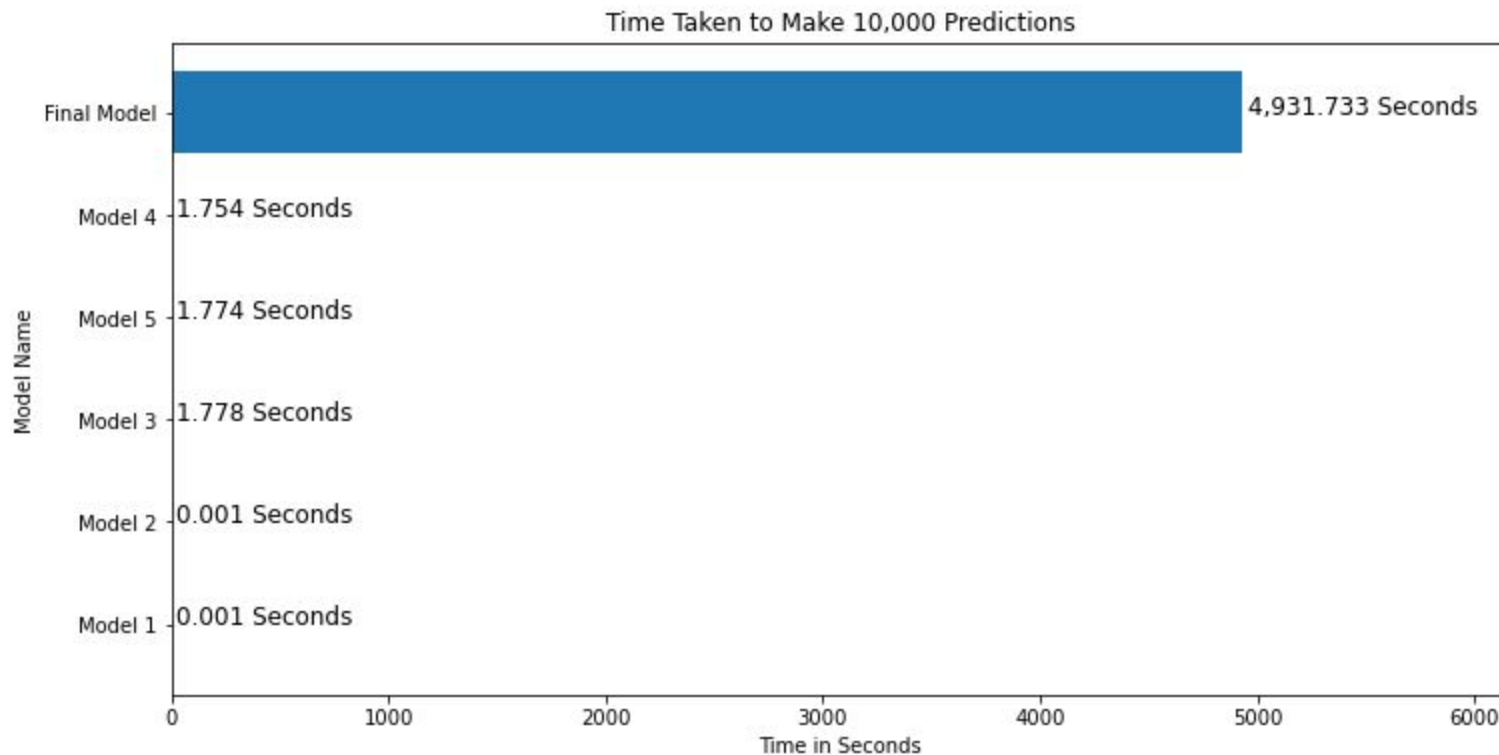
Results

- Final model vastly outperforms all others



Results

- Final model has slow speed
- Must use other model for demonstration
- Slow speed shouldn't be a major problem



Conclusions

Explainability Benefits

- Ensure model makes decisions for the correct reasons
- Deeper understanding of moderation policies

Text Explanation

i've used marijuana on and off for 25-30 years. one thing i've noticed that there is no specific pot-smoker type. i've smoked with young people, senior citizens, hillbillies, lawyers, teachers, engineers, conservatives, liberals, organized people, messy people, atheists, christians, vegetarians, meatheads, etc.

i live in a very conservative southern usa state, the vast majority of people i know support medical marijuana at least if not full legalization.

Conclusions

Why not leave it up to the model?

- Accuracy very good, but not perfect
- Unfair removal is a major inconvenience to user
- Not worth deploying the model in this way

Conclusions

Tag-Hold System

- Model tags comments likely to be removed
- Comments hidden until human moderator makes final decision
- Regular users see significantly less unwanted content
- Better user experience leads to higher user engagement

Conclusions

Looking Forward

- Begin with limited deployment to specific boards
- Could eventually be offered to all boards

Conclusions

Next Steps

- Transformer architecture has more potential
- Investigate more advanced model explanation

Thank You!

Email: kaiuechi@gmail.com

GitHub: [@kaiuechi](https://github.com/kaiuechi)

LinkedIn: [kai-uechi](https://www.linkedin.com/in/kai-uechi)