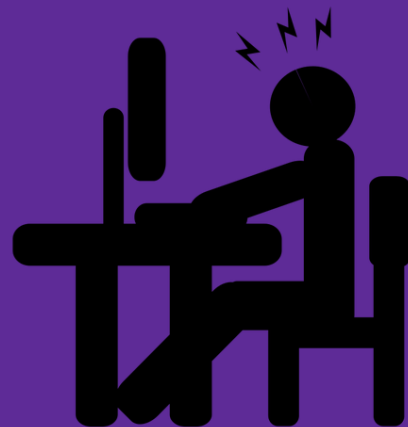


Analyzing Hate Speech

Menace vs. noise



Russell Blanchard

- Data Scientist
- Logistics Background
- PLU Psychology
- Flatiron school



An aerial photograph of a river flowing through a valley. The river is light green and meanders through a landscape of colorful terraced fields in shades of yellow, orange, and green. The fields are arranged in a patchwork pattern. In the background, there are steep, green karst mountains with some rocky peaks. A small village with red-roofed houses is visible on the right side of the river. The sky is overcast with grey clouds.

Background

Agenda

1. Business Understanding
2. Data Overview
3. Modeling Process
4. Hate Classification
5. Next Steps

Hate Speech vs. Regular Speech

- Critical Differentiation
- Legal Consequences
- Toxic Platforms



A photograph showing the silhouettes of seven sailors in uniform standing in a line on the deck of a ship. They are facing right, looking out at the ocean under a dramatic sunset sky with orange and blue hues. The text 'Data Analysis' is overlaid in the bottom left corner.

Data Analysis

Vietnamese Social Media

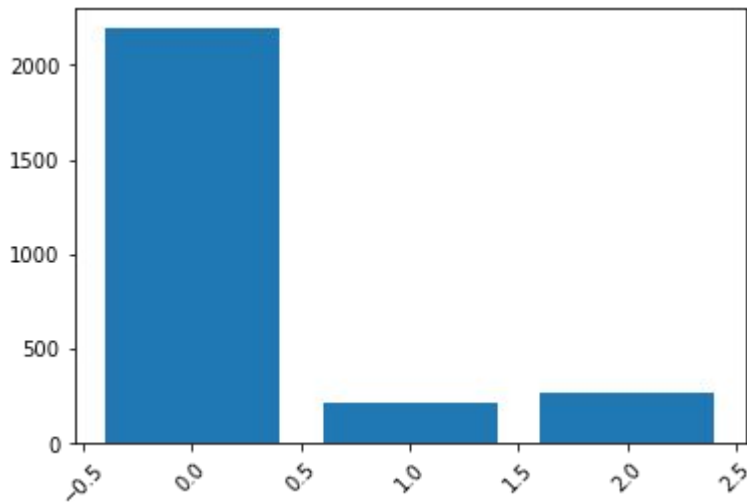
- 30000 comments
- 3 categories
- Critical differentiation between offensive and hate speech

Common terms and phrases

- Consonants and adverbs
- Epithets and threats
- Ambiguous shorthand



Number of Comments



0 = Clear

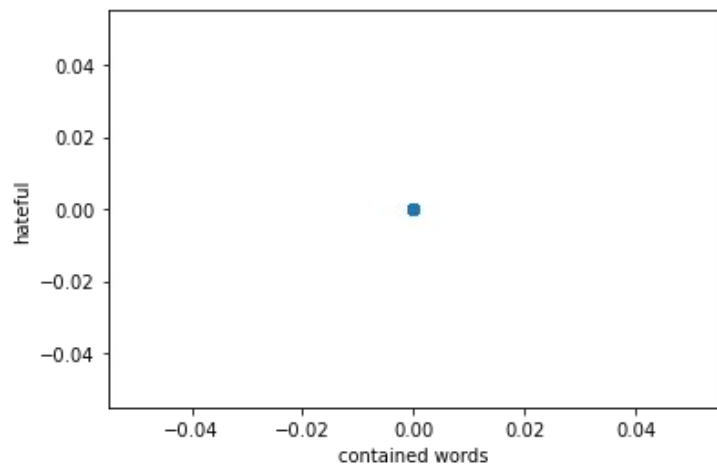
1 = Offensive

2 = Hateful

Vast Majority of Speech Benign

Modelling Process





Models

Conclusion



Significant overlap between Hate and Offense

- Contextual Determinations
- Common phrases
- Degrees of offense

Next Steps

- Incorporate AI
- Translation
- Identifying repeat bad actors

Questions?



Contact Information

LinkedIn: [linkedin.com/in/russell-blanchard-285b80271](https://www.linkedin.com/in/russell-blanchard-285b80271)

GitHub: russelltblanch1993

Citations

```
@InProceedings{10.1007/978-3-030-79457-6_35,author="Luu, Son T. and Nguyen, Kiet Van and  
Nguyen, Ngan Luu-Thuy", editor="Fujita, Hamido and Selamat, Ali and Lin, Jerry Chun-Wei and  
Ali, Moonis", title="A Large-Scale Dataset for Hate Speech Detection on Vietnamese Social  
Media Texts", booktitle="Advances and Trends in Artificial Intelligence. Artificial  
Intelligence Practices", year="2021", publisher="Springer International Publishing",  
address="Cham", pages="415--426",
```

Images credit Google