Phuoc Le, Gil Rabara, Dilnoza Saidova, Vivian Tran
456: Natural Language Processing
April 23, 2023

Revised Project Proposal: Toxic Comment Classification

Introduction

The project proposal for the Toxic Comment Classification aims to address the challenges of online etiquette and the misinterpretation of text-based communication. In the current digital age, promoting positive and respectful communication in professional and academic settings is crucial. The project will focus on detecting and classifying negative and hateful comments in plain text, while also providing users with alternative suggestions that adhere to online etiquette guidelines.

Target Users

The project aims to benefit a wide range of users, with a focus on employees in a professional business setting and students at a university. By providing a baseline for online etiquette, the project can encourage respectful communication among coworkers and help create a positive and welcoming online environment. It can also assist students in drafting emails to professors, ensuring clarity and avoiding redundancy. Additionally, the project will support users applying to programs via email, helping them maintain formality and professionalism throughout their correspondence.

Implementation Approach

The project will utilize Natural Language Processing (NLP) techniques to detect and classify negative and hateful comments within emails. It will recommend replacements for words or phrases that don't meet online etiquette criteria. Additionally, the system will identify hurtful comments, such as bullying, harassment, or hate speech, and provide suggestions to rephrase or remove them. The implementation process involves collecting datasets of words and phrases that oppose online etiquette rules and those indicating hateful speech. These datasets will be used to train the model for accurate detection and suggestion generation. The model will be trained to handle emails of varying lengths to ensure efficient performance. The goal is to integrate the application into different email platforms, allowing users to benefit from its features while composing emails.

Impact and Benefits

The Toxic Comment Classification project aims to assist users in sending messages that are more likely to be interpreted positively. By replacing negative words and suggesting appropriate alternatives, the project can help prevent misinterpretation and potentially harmful comments. The application will support users in maintaining professionalism and adhering to online etiquette guidelines, ensuring that their emails are well-written and suitable for the intended setting.

Conclusion

The Toxic Comment Classification project intends to leverage NLP techniques to improve online etiquette and promote positive communication. By detecting and suggesting replacements for negative words and phrases, the project aims to mitigate misinterpretation and encourage respectful dialogue. With its integration into email platforms, the project can provide valuable support to users, facilitating a more positive and welcoming online environment.

Helpful Sources

Microsoft 365 Team. “Your Guide to Chat Etiquette in the Workplace.” *Microsoft.Com*, Microsoft 365, 17 June 2021, https://www.microsoft.com/en-us/microsoft-365/business-insights-ideas/resources/your-guide-to-chat-etiquette-in-the-workplace. Accessed 22 April 2023.

Microsoft 365. "What Is Nettiquette?" *Microsoft.Com*, Microsoft 365, 20 Jul. 2022, www.microsoft.com/en-us/microsoft-365-life-hacks/privacy-and-safety/what-is-netiquette. Accessed 22 April 2023.