Purvesh Desai
Master in Data Analysis and Visualization

Comment's Sentiment Analyzing Model using BERT

Introduction:

LLMs are fascinating and truly remarkable in their capabilities. The level of sophistication and natural language understanding they exhibit is a testament to the rapid progress in artificial intelligence. These models have the potential to revolutionize how we interact with technology and understand and generate text, opening up new possibilities for communication, creativity, and problem-solving. As we continue to explore and harness the power of LLMs, it's important to approach their development and use with a thoughtful and responsible mindset, ensuring that they benefit society in ways that promote fairness, inclusivity, and ethical considerations.

In the present context, social media has seamlessly woven itself into people's lives, deepening their connections. However, exposure to negativity on these platforms can lead to stress, impacting mental health. The objective of my project is to foster positivity in an individual's environment, shielding them from negativity and mitigating mental stress. Through this initiative, I aim to create a digital space that contributes positively to individuals' well-being, ensuring that their online interactions promote a healthier mental state and a more uplifting social experience.

My aim to conduct sentiment analysis on Twitter comments. With the vast amount of information shared on Twitter daily, it will be interesting to understand the sentiments expressed in user comments. This project aims to train the model based sentiments as positive or negative. By harnessing the power of natural language processing and machine learning, I aspire to train my model using BERT technique which can help me achieve my goal.

This model can extend its utility as an application, serving as a screening tool for browsers, social media platforms, or devices. It aims to inform users about the content they are about to post, assisting them in refining and improving their content before sharing it on their platforms.

<u>Data Collection:</u>

I selected my dataset from Kaggle, a well-known platform for data science and machine learning resources. This dataset specifically features user comments and their corresponding sentiments as its key attributes. The volume of this dataset is 1.7 million. In this case, the dataset I've chosen will serve as the foundation for my sentiment analysis project, as it contains the comments and associated sentiments that are essential for training and evaluating sentiment analysis on my model.

In my dataset, the columns are labeled "Comments," containing different text-based comments, and "Sentiments," which contain values of 0.0 and 1.0 indicating negative and positive sentiments, respectively.

<u>Analysis Plan:</u>

Cleaning the Data:

I cleaned the text data by removing null or 'NA' values. Tokenize the comments to break them into smaller units for analysis. Split the dataset into training, validation, and test sets for model training and evaluation.

Choosing Model:

Choose BERT (Bidirectional Encoder Representations from Transformers) because of its good at understanding words in sentences and as it is the underlying model for sentiment analysis due to its ability to capture contextual information in text.

Training the Model:

Teach BERT to understand if a comment is happy or sad using the training data. Train the BERT-based model on the training dataset, adjusting hyper-parameters such as learning rate, batch size, and epochs.

Out of 1.7 million records, I trained the model using a dataset of 10,000 records and tested its performance on 1,000 records. This ensured a focused and efficient training process for the model's optimal performance.

Initially, after training the model for five epochs, the accuracy plateaued at about 57%. However, the results lacked precision, prompting an adjustment. To enhance accuracy, I increased the epoch count to 50. This modification significantly improved performance, yielding an accuracy of around 87%. The refined model demonstrated a more accurate representation of the data, indicating a substantial enhancement in its ability to correctly classify information. This iterative process underscores the importance of modifying the hyper-parameters to optimize model performance.

Checking How Well It Learned:

Evaluate the final model on passing some new comments it hasn't seen before to ensure generalizability and robustness. Check for potential overfitting or under-fitting issues and make necessary adjustments.

Giving It a Real Test:

Test BERT model on new comments to make sure it works in the real world. Check for any big mistakes and make BERT model better.

This exercise helps understand the emotions behind statements, categorizing them as positive or negative based on their tone and context. I gathered sentences from friends to analyze their sentiments. Statements like 'Good time to have a Taco' and 'Finally, the semester is about to get over' were labeled as 'Positive', while 'Why does it always rain on weekends' was labeled as ‘Negative'.

Accessibility and Limitations:

The Comment's Sentiment Analyzing Model using BERT aims to be accessible to a wide range of users. This model facilitates the monitoring and screening of all data and statements shared on social media, effectively curbing negativity and promoting positivity across platforms. By analyzing sentiments, it contributes to fostering a positive online environment. This proactive approach not only enhances the overall social media experience but also cultivates a more uplifting atmosphere for individuals. The model's ability to sift through content aligns with the goal of creating a supportive and constructive online community, where users can engage in a positive exchange of ideas and information.

Despite efforts to enhance accessibility, limitations exist. Challenges may arise in scenarios where input data includes complex language structures or in cases of rare sentiments not adequately covered during training, also personal views from an individual can be different from others. The model's effectiveness could be compromised if it encounters unfamiliar expressions or context. Additionally, accessibility may be affected if there are disparities in the availability of resources or technological infrastructure. Continuous refinement, broader training datasets, and feedback mechanisms are essential for mitigating these limitations and improving the overall accessibility of the sentiment analysis model.

User Expectations:

Users anticipate accurate identification and categorization of sentiments in comments, providing insightful feedback on whether the expressed emotions are positive or negative. Users also expect the model to handle a diverse range of language structures and nuances, ensuring robust performance across various contexts. Additionally, they look for user-friendly interfaces and efficient processing, allowing seamless integration into their applications or platforms. Overall, user expectations center around the model's accuracy, versatility, and ease of use in enhancing the understanding and analysis of sentiments in comments.

Social and Ethical Implications:

The implementation of a Comment's Sentiment Analyzing Model using BERT carries both social and ethical implications. On the positive side, it can foster a more positive online environment by identifying and mitigating negative sentiments, promoting healthy digital interactions. However, concerns arise regarding privacy, as the model involves the analysis of user-generated content. Ethical considerations include ensuring transparency about data usage, obtaining informed consent, and preventing potential misuse, such as bias in sentiment categorization. Additionally, the model's impact on free expression and diverse perspectives must be monitored to avoid unintentional censorship. Striking a balance between sentiment analysis benefits and ethical considerations is crucial for responsible deployment in the ever-evolving landscape of digital communication.

Significance of the Project:

The project of developing a Comment's Sentiment Analyzing Model using BERT holds significant importance in several key aspects. Firstly, it addresses the growing need for a more positive and constructive online environment. By accurately identifying sentiments in comments, the model contributes to reducing online negativity and fostering healthier digital interactions. This directly impacts users' mental well-being, providing a tool to navigate through a more supportive online space.

Moreover, the project has broader implications for responsible AI deployment. It highlights the ethical considerations involved in sentiment analysis, emphasizing transparency, privacy, and the prevention of biases. As technology continues to shape the digital landscape, the significance of this project lies in its potential to contribute to a more ethical, empathetic, and inclusive online community.

Conclusion:

In conclusion, the development and application of the Comment's Sentiment Analyzing Model using BERT represent a significant step towards improving online discourse. By effectively categorizing sentiments in comments, the model contributes to cultivating a more positive and supportive digital environment. Its potential impact on mental well-being and the reduction of online negativity is noteworthy. As technology continues to shape our digital interactions, this model stands as a valuable tool for enhancing the quality of online conversations and fostering a more constructive and empathetic online community.