Hypothesis
The research hypothesis posits that the utilization of advanced natural language processing models like XLNet, ALBERT, and BERT can effectively assist in categorizing memes as either troll or non-troll. The study suggests that discerning the intention behind memes, especially troll memes, is challenging due to their subjective nature, which often involves offensive language, colloquial expressions, emojis, references, symbols, and images without accompanying text. The research also proposes that the widespread dissemination of memes through social media, combined with the internet's role in facilitating their propagation, amplifies the difficulty of monitoring and mitigating the negative mental impacts of troll memes on individuals. By employing machine learning models, the study aims to contribute to the identification and labeling of troll memes, thereby enhancing the ability to regulate and manage offensive content on the internet. Additionally, the research acknowledges the unique linguistic and cultural context of Tamil, one of the world's oldest classical languages, as it investigates the challenges and opportunities in classifying troll memes within this linguistic landscape

Contribution
The research provides practical implications for the field of troll meme classification in Tamil by offering valuable insights into the performance of transformer models. Notably, the study finds that the XLNet model surpasses BERT and ALBERT in accuracy, establishing it as a promising choice for effective troll meme classification in the Tamil language.
The paper underscores the importance of data preprocessing using the Natural Language Toolkit (NLTK) as a critical factor in enhancing classification accuracy. This highlights the practical utility of careful data preprocessing in contributing to improved performance in the task of troll meme classification in Tamil.

The paper underscores the importance of data preprocessing using the Natural Language Toolkit (NLTK) as a critical factor in enhancing classification accuracy. This highlights the practical utility of careful data preprocessing in contributing to improved performance in the task of troll meme classification in Tamil
The study extends its contribution to the broader field of Natural Language Processing (NLP) by showcasing the performance of transformer models (BERT, ALBERT, and XLNet) in the specific context of Tamil troll meme classification. This demonstration not only advances the understanding of NLP in the context of Dravidian languages but also serves as an inspiration for future research and development of NLP models tailored for other Dravidian languages or similar linguistic tasks. The research thus contributes to the expanding landscape of NLP applications and methodologies

Methodology
In the endeavor to enhance the categorization of troll memes in the Tamil language, this research paper utilized a meticulously designed dataset specifically tailored for this particular objective.The dataset, featuring memes annotated as either troll or not-troll, strategically embedded label information within the file names, ensuring a structured foundation for subsequent analyses.

To prepare the dataset for robust modeling, a comprehensive preprocessing phase was initiated, leveraging methodologies from the Natural Language Toolkit (NLTK) and spacy toolkits. This encompassed a spectrum of transformative measures, including the conversion of uppercase letters to lowercase, expansion of contractions, elimination of URLs, removal of accented characters, extraction of stopwords, and lemmatization of the textual content. These preprocessing steps collectively aimed at standardizing and refining the textual data, laying the groundwork for effective model training and evaluation.

While the paper did not explicitly disclose the size or origin of the utilized dataset, it elucidated the evaluation process wherein the trained models were rigorously tested on a dedicated set of test data, accompanied by provided labels. This methodological choice ensured a thorough assessment of model performance within the defined scope of troll meme classification.

In terms of modeling, the research incorporated three transformer models, namely BERT, ALBERT, and XLNet, renowned for their prowess in natural language processing tasks. The BERT model, characterized by its ability to comprehend the entire sequence of words simultaneously, enabling contextual learning, underwent training for four epochs. ALBERT, a streamlined variant of BERT designed to conserve computational resources while preserving performance, was trained for five epochs. Lastly, XLNet, a BERT-like pretrained model with a focus on capturing bi-directional context through permutation language modeling, underwent a four-epoch training regimen. These distinct models were strategically selected to provide a nuanced understanding of their individual contributions to the task at hand.

In addition to the overarching dataset preprocessing strategy, the paper also highlighted the specific application of NLTK and spacy toolkits for additional tasks, including the annotation of emojis and emoticons. This nuanced approach further underscored the commitment to a comprehensive preprocessing pipeline tailored to the intricacies of troll meme classification in the Tamil language.

In essence, this research methodology encompassed a judicious combination of dataset curation, preprocessing intricacies, and model selection, all orchestrated to advance the understanding and effectiveness of troll meme classification in the Tamil language. The deliberative choices made at each juncture of the methodology underscored a commitment to methodological rigor and nuanced analysis within the domain of computer science research.

Limitation

- The paper acknowledges that classifying memes based solely on text can lead to reduced accuracy, as the perception of a meme is influenced by multiple factors that cannot be captured by simple conventional models.
- The intuitive nature of the words present in memes makes it challenging for models to accurately detect their meaning.

- The paper does not provide detailed information about the size or origin of the dataset used, which limits the generalizability of the findings.
- The paper does not provide extensive information about the methodology or steps followed in the offensive text classification, making it difficult to assess the robustness of the approach.**[1]**

Future Plan

The paper suggests further work on the troll meme classification task by adding multiple hidden layers and building a complex network structure.**[1]**

- The paper also acknowledges the limitations of classifying memes based solely on text and suggests exploring other factors that influence meme perception, such as visual and contextual features.**[2]**
- The paper mentions the importance of considering the distribution and types of images in the test set, as it can affect the performance of models trained on ImageNet. Further investigation into the impact of test set characteristics on model performance is recommended.**[3]**