https://doi.org/10.48047/AFJBS.6.14.2024.4777-4796

AFJBS
African Journal of Biological Sciences
Journal homepage: http://www.afjbs.com
Research Paper Open Access

Data Augmentation with DistilBERT to categorize Patronizing and Condescending Language

Bolukonda Prashanth[1,a)] and Dr P Vijaya Pal Reddy[2]

[1]Research Scholar, Department of CSE, University College of Engineering, Osmania University, Hyderabad, Telangana, India.
[2]Professor, Department of CSE, Matrusri Engineering College, Hyderabad, Telangana, India.

[a)]Corresponding author : boluk.cse@gmail.com

Volume 6, Issue Si4, Aug 2024

Received: 09 June 2024

Accepted: 19 July 2024

Published: 08 Aug 2024

doi: 10.48047/AFJBS.6.Si4.2024.4777-4796

Abstract
Condescending and patronizing language can occasionally be interpreted positively or badly. Patronizing someone can imply "supporting." One definition of condescension is "superior attitude towards others." The suggested method analyses news stories from different nations and determines whether or not they use patronizing or condescending language and categorizes the detected PCL into various groups like 1) Unbalanced power relations(UPR), 2) Shallow solution(SS), 3) Presupposition(PS), 4) Authority voice(AV), 5) Metaphor(MP), 6) Compassion(CP) and 7) The poorer, the merrier(TPTM). This work use deep learning (DL) techniques to address this issue, approaching it like a typical multi label text classification problem. It is suggested to use a pre-trained model Distil-Bert with Data Augmentation methods to achieve best classification. It is considered macro f1 as the metric, The model that has been suggested DistilBERT attained a score of 47.01. The utilization of back translation in data augmentation resulted in a score of 47.34. The utilization of contextual word embedding for data augmentation resulted in a score of 47.80, while the implementation of synonym replacement for data augmentation yielded a score of 49.26.

1. Introduction

The use of patronizing and condescending language has increased in recent years due to the rapid rise in social media usage (PCL). Condescending and condescending language can mask acts that seem helpful or compassionate, but really show a sense of superiority toward others. Harmful language behaviour, including hate speech[1], offensive language[2], fake news[3], rumor propagation, disinformation[4], and many others, has been thoroughly explored in NLP, despite the fact that PCL was up until now an understudied field of study.

PCL can be difficult to identify, even for humans, because to its subjectivity and intricacy. Some individuals may not perceive any issue with expressing how individuals in a position of privilege contribute the remainder of their resources to those who are in need, or something that one person finds condescending may be perceived by another as an objective description of the circumstances. An individual belonging to a community that is commonly referred to as vulnerable. may also anticipate feeling more patronized than an outsider when reading how others refer to them.

With in the domain of natural language processing, the identification of patronizing and condescending language (PCL) is an open, difficult, and little-studied topic [5,6]. A person in a position of leadership can have a helpful and friendly attitude toward others, who are usually portrayed as having a subtle sense of compassion, as opposed to being arrogant and condescending [5]. PCL is regarded as a rather real occurrence. It is frequently motivated by good intentions, unconscious, and conveyed in poetic language [7,8]. Because PCL cannot be connected to particular words, it is challenging for NLP systems and human annotators to identify and classify (see Figure 1). However, it creates prejudices, a sense of superiority, and discrimination, among other undesirable messages [9]. When employed by media outlets with a wide audience, this is especially harmful because it increases the marginalization of already vulnerable populations [5]. The ability to automatically identify PCL may open up a variety of applications and research avenues, including studies on the relationship between condescending and sociodemographic variables and recommendation tools for news editors to reduce condescension in writing prior to publication.

The recent proposal of the SemEval-2022 Task 4 aims to facilitate research on the detection of patronizing and condescending language. The objective of the collaborative task is to examine methods for detecting PCL and categorizing the language techniques employed to convey it in English news items pertaining to marginalized communities. The purpose of SemEval 2022-Task 4[10] is to develop a system capable of discerning whether a given text contains PCL and accurately identifying its presence or absence. The PCL category expresses condescension. The organizers supplied 2 sets of data: one containing PCL classifications and the other containing annotations based on PCL intensity categories. Several authors have developed alternative models in the area and achieved significant outcomes. The objective of our study was to utilize a pre-trained model called Distil-Bert to classify PCL (Prescriptive, Conceptual, and Linguistic) in a given text. Data augmentation is also suggested as a means to address the issue of class disparity. For this study, This work focus on the PCL categorization task. Table 1 displays the textual examples corresponding to each category of PCL.

Example	Category of PCL
“We can be extremely proud of the current women winemakers”	Unbalanced power relations
“The inclusion of a refugee team”	Shallow Solution
“An immigrant to a developed country lives in two worlds”	Presupposition
“women must wake up”	Authority voice
“trapped in the prison of poverty”	Metaphor
“more than 400 suspected asylum seekers are awaiting their fate”	Compassion
“how talented disabled people can be”	The poorer, the merrier

Table 1: Examples PCL for each category.

The remainder of the article is arranged as follows. Section 2 provides an overview of the related work. The dataset is described in Section 3. The experimental configuration of our suggested model is outlined in Section 4. Pre-processing and implementation specifics for the suggested model are involved. The outcomes and discussion are covered in Section 5. Finally, this work wrapped up the report and offered some suggestions for more research in the Section number 6.

2. Related Work

Many discourse forms, including hate speech[11], offensive language[12], intended sarcasm [13], fake news[14], and rumors[15,16], have been the subject of extensive research on harmful language detection and recognition. The study of PCL [16,8,17,18] has been largely disregarded in the NLP community until recently. In contrast, the majority of these research have concentrated on explicit, aggressive, and flagrant phenomena.

In order to promote more investigation into the PCL language [6], introduce a task focused on detecting condescension. They also offer a dataset called TALKDOWN, which consists of pairs of comments and replies from Reddit. In addition, Pérez et. al. [5] present a dataset called "Don't Patronize Me!" and discuss the problem of detecting Patronizing Communication Language (PCL) in relation to vulnerable communities such as refugees, homeless individuals, and impoverished families. These studies set up multiple advanced benchmarks by utilizing pre-trained language models[19,20]. They indicate that identifying this type of linguistic is a difficult task for both NLP systems and humans because of its nuanced and subjective characteristics.

Language detection for identifying hate speech While the field of NLP has not extensively studied the identification of patronizing and condescending language, significant research has been conducted on various types of harmful language detection. These include automated hate speech detection[21], rumor propagation[22], fake news detection[23], and trustworthiness prediction[24]. The Social Bias Inference corpus[25], these were created to examine the unequal power dynamics inherent in condescending language.

Text classification with multiple labels, There are three primary strategies for solving a multi label text classification problem. They are, Binary Relevance, Classifier Chains. This is studied by Dembczy´nski et al. in 2010 [26] and the Label Powerset approach developed by Boutell et al. in 2004 [27]. Binary Relevance is a method that treats each class separately and does not take into account the relationship between labels. Label Powerset technique treats every possible combination of labels as a separate class, By efficiently turning a problem of classifying multiple labels into a problem of classifying a single label [28] suggest a method of document transformation where label weights are determined by considering label entropy. The study in 2012 [29] the objective is to incorporate label dependence into binary relevance technique. Spolaor in 2013 [30] studied the relevance of each label is evaluated by combining Relief and Information Gain with label power set and binary relevance methods. Wang studied in 2017 [31] about incorporating regularization techniques throughout the period of training and utilizing support inference during prediction, in addition to applying optimizer, enhances the F1 score of the multi-label issue.

Pre-trained transformer based language models like RoBERTa, BERT, transfer learning and others frequently surpass the performance of numerous classical models that are trained from the beginning. Their complex contextual embedding's are responsible for this achievement. Consequently, such models are employed for numerous subsequent tasks. Li and Xiao in 2020 employ SpanBERT[32] to identify propaganda strategies in news stories. Ranasinghe [33] utilized BERT based multilingual models for the purpose of identifying offensive words in social media.

The conventional classification approach using SciBERT is substituted with a label ranking model that relies on a Bi-Encoder and a Cross-Encoder. This allows for effective management of extensive labels in Multi label text classification. In addition, Dan Li [34] introduced an active learning-based process to handle the limited availability of new labels when updating a classification model. The application of DistilBERT and Data Augmentation approaches [35] enhances the accuracy and other metrics of text classification. Data augmentation with back translation, synonym replacement and Contextual word embedding have provided good accuracy on various sizes of datasets. These metrics may vary depending on the size of datasets and number of classes.

3. Dataset

Primary source material in this project is a dataset named "Don't Patronize Me!" (DPM). The collection of words is annotated and exhibits both patronizing and condescending attitudes towards oppressed individuals. The dataset was originally introduced in 2020 by Perez-Almendros [5]. The dataset comprises 10,469 paragraphs and was utilized as the dataset for training at the Semantic Evaluation assignment. In order to generate the test set for this study, the authors carefully annotated an extra 3,898 paragraphs, following the same approach precisely. The news articles were obtained from media sources in twenty nations where English is the primary language. The News onWeb (NoW) corpus, as provided by Davies in 2013, is the primary source of these stories. This research utilizing this dataset for research purposes under the authorization of the authors [5].

4. Implementation

4.1 Pre processing

The fundamental data pre-processing procedures that were implemented for proposed experiments are outlined in this section.

It has been dropped all null value attributes from the dataset, Next, the dataset was partitioned into training, testing and validation sets using an 80:10:10 ratio. Tokenization is applied on the split datasets then created as batches. Pre-trained models are not compatible with unprocessed text. Therefore, it has been transformed the text into encoding and added with 2 more columns input_ids and attention masks to extract the features of datasets. Subsequently, the encoded sequence is sent into the model to execute the classification process.

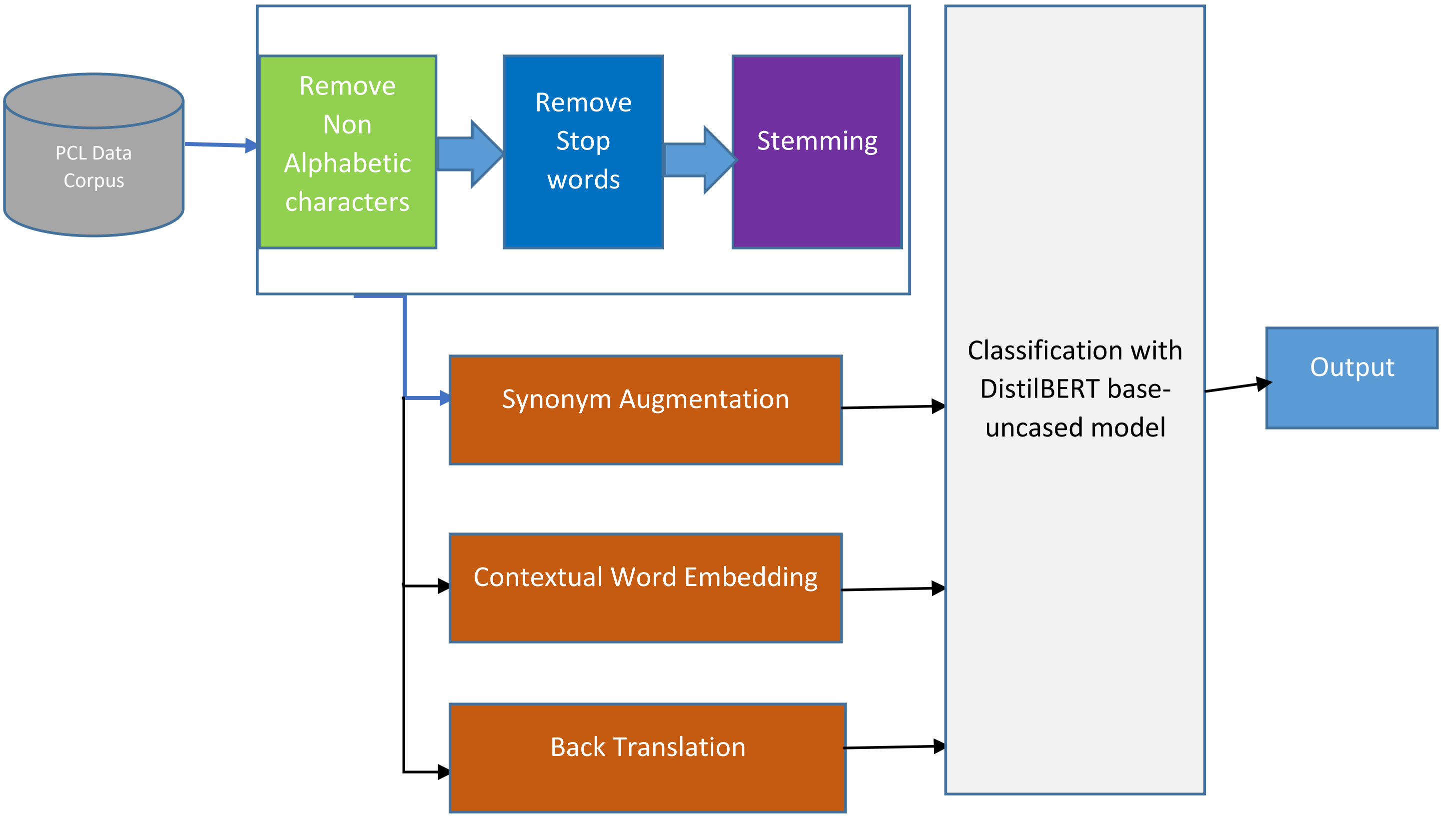

Figure 1: Proposed Model

4.2 Multilabel Classification

4.2.1 DistilBERT: The process of categorization is executed by utilizing a meticulously adjusted pre-existing distilBERT model with base uncased. The DistilBERT model with base uncased is a popular variant of the BERT (Bidirectional Encoder Representations from Transformers) family. It is specifically designed to be smaller, faster, and more efficient, while yet keeping a substantial percentage of BERT's performance. Understand the complexities of this paradigm and grasp its characteristics and use. Hugging Face's team introduced DistilBERT as a method to create more efficient and effective versions of the BERT model. The goal is to attain a balanced equilibrium between the model's performance and its computing efficiency.

DistilBERT achieves this by reducing the number of layers and parameters compared to BERT, while employing a technique known as "knowledge distillation". The uncased technique is typically preferred when case sensitivity is not essential, as it simplifies the tokenization process. DistilBERT consists of 6 transformer layers and utilizes 66 million parameters, which is a smaller number compared to BERT's 110 million parameters. Figure 2 depicts the architectural design of the DistilBERT model, which incorporates Data Augmentation.

Despite its modest size, DistilBERT demonstrates competitive performance across a variety of NLP benchmarks. It achieves a performance level of around 97% compared to BERT on tasks such as the GLUE benchmark, making it an appealing option for resource-constrained applications. The training procedure entails the utilization of knowledge distillation, a technique in which a smaller model called DistilBERT is trained to mimic a larger model known as BERT. This method allows the student model to capture a substantial part of the teacher's knowledge in a more concise form. The proposed model is shown in the Figure 1. The model was trained using the following hyper parameters, they are, learning rate of 2e-5, batch size=8 and epoch=3. Which are provided with a comprehensive outcome.

4.2.2 Data Augmentation: Data augmentation is a potent method that improves training datasets and boosts the performance of models across different data sources. In the domain of Natural Language Processing, methods such as substituting synonyms, translating in reverse, and expanding context are crucial for generating varied and resilient training datasets. Implementing data augmentation in an efficient manner can greatly enhance the generalization and performance of models, especially in situations where there is a scarcity of data.

In order to resolve the issue of class disparity in the dataset, it is implemented that the technique of data augmentation to enhance the results. This research has implemented three distinct data augmentation methodologies as described below.

4.2.2.1 Synonym Replacement: Synonym replacement is a straightforward yet potent strategy for enhancing written content. The process is substituting words in a phrase along with their synonyms to produce new phrases that possess both grammatical and semantic similarity. This enhances the variability of the training data and enhances the model's capacity to generalize.

4.2.2.2 Contextual Word Embedding: Contextual word embedding is a sophisticated technique for enhancing text data that leverages the context within phrases to generate more pertinent and grammatically precise variations. This method typically employs pre-trained language models like BERT, GPT, or their derivatives. These models possess the capacity to understand and generate words by utilizing the context offered by the surrounding text.

4.2.2.3 Back Translation: Back translation is a technique used in NLP to augment data. It involves translating text into another language and then translating back into the original language. This approach is extremely successful in producing a wide range of excellent training data while preserving the semantic significance of the original text. This work utilized the procedure of English-to-German translation followed by German-to-English translation.

5 Results and Discussion

Proposed model DistilBERT-base-uncased with data augmentation for categorization PCL has performed well. The metric macro F1-score used to judge proposed model performance. Proposed model achieved the macro F1-score 49.26, precision score 52.95 and recall score of 46.10 at testing phase which given the increase in the metrics with comparing models. This work compares proposed model with prompt based models [36] has achieved Macro F1 Score 46.90 and Prompt training and label attention mechanism [37] has achieved Macro F1 score 43.90. Proposed model received a favourable score when compared to other criteria. Transformer models of larger size generally exhibit slower inference speeds compared to DistilBERT, which is particularly crucial for processing real-time data. During the training phase, larger models generally undergo fine-tuning at a slower pace compared to DistilBERT, resulting in greater benefits. DistilBERT effectively transfers acquired information throughout the distillation process across larger transformer models [38].

DistilBERT is typically fine-tuned on specific datasets for text classification tasks, enabling it to capture task-specific patterns and nuances. When comparing these fine tuning approaches to all purpose techniques. In prompt-based models, the level of accuracy is often higher. DistilBERT-base-uncased is significantly more resource-efficient than large prompt-based models. Modifying DistilBERT does not necessitate intricate rapid engineering or dynamic alterations. Instead, it employs a straightforward approach to alter the weights of the model by taking into account task-specific inputs. DistilBERT can outperform prompt-based model and prompt training and label attention mechanisms in terms of training and data requirements [38]. It has been proposed data augmentation with synonym replacement, and obtained an Macro F1 score 49.26, Macro recall score of 46.10 and Macro precision score of 52.95. When comparing prompt-based learning with transformer-based models, data augmentation with synonym replacement emerges as the most successful strategy for text categorization across many significant factors. This comprehensive comparison highlights the benefits of using synonym replacement in the process of text classification. The training dataset effectively increases in both size and diversity without requiring more labelled data by replacing terms with their equivalents. Enhancing the model's capacity to generalize to novel instances may lead to a decrease in overfitting. Synonym replacement [39] is an uncomplicated technique, which may be easily included into the data pre-processing pipeline and requires minimal computational resources. This stands in sharp opposition to the computational load associated with using large prompt-based models for generating responses, or training enormous transformer models. Synonym substitution augmentation of data does not necessitate pre-existing models. This method can be applied to a diverse array of languages and subjects. When compared to large transformer models or systems that rely on prompts, using augmented datasets with synonym substitution is more convenient for scaling and can be used with smaller models that are easier to construct.

It has been proposed data augmentation with contextual word embedding, and obtained Macro F1 score 47.80, a Macro recall score of 44.74 and Macro precision score of 51.38. It is conducted an investigation on several significant attributes of data augmentation using contextual word embedding [40]. This research's findings indicate that this approach may offer superior performance for text classification compared to prompt-based learning and transformer-based models. Models like BERT or similar ones are used to generate contextual word embedding, which enhance the realism and semantic richness of text data changes. Consequently, less complex models may achieve superior performance and rival more advanced tactics. Contextual word embedding, when employed for data augmentation, provides comprehensive and meaningful representations of words within their specific contexts, hence enhancing generalization and performance. Contextual embedding can be enhanced to capture certain nuances in domain-specific data, leading to more effective data augmentation. This is particularly useful for specialist applications such as detecting patronizing and condescending language. By including contextual embedding into the data, the model gains increased resistance to variations in language and adversarial inputs, leading to improved performance.

It has been proposed another model data augmentation with back translation, and obtained Macro F1 score 47.34, a Macro recall score of 44.16 and Macro precision score of 51.06. Data Augmentation with Back Translation [41,42,43] is considered superior to Transformer-based models and prompt-based learning for text classification due to several crucial properties. Back translation generates diverse and linguistically nuanced data variations. This strategy would enhance the resilience and efficiency of the model without increasing the computational burden of more complex models. Back translation involves translating text into another language and then translating back into the original language and generates paraphrases that incorporate a wide range of vocabulary and sentence constructions. This enhances the model's capacity to apply its knowledge to other linguistic patterns. It is particularly advantageous for capturing a broader spectrum of phrases, as it maintains the basic sense of the text while retaining natural variations in language, as opposed to mere substitution of synonyms.

The suggested models showed improved performance compared to standard transformer-based models and prompt-based models, as well as Prompt training with label attention mechanism, when the DistilBERT-base-uncased model was combined with varied data augmentation strategies. The metrics considered in the comparing models are Macro F1 Score. This work also considers the same along with Macro Precision and Recall Score. Specifically, macro F1 score improved by 5.03%, indicating a reduction in false positives. Macro Precision and Macro Recall saw a good score, showcasing better detection of true positives. Consequently, the Macro F1-score rose to 5.03% significantly, reflecting an overall balanced and enhanced performance. The improvements can be attributed to the model's efficient architecture and enhanced ability to generalize from enriched data, resulting in more dependable and precise predictions.

Table 2, 3, 4, 5 and 6 shows the score of proposed and comparing models. Figure 2, 3, 4, 5 and 6 compare the precision, recall, and F1-score of all the proposed models and comparing models that are being presented.

Metric	Patronizing and Condescending Language Categories							
	UPR	**SS**	**PS**	**AV**	**MP**	**CP**	**TPTM**	**Macro Avg**
F1	65.01	53.48	37.15	41.11	34.01	53.18	45.15	47.01
Precision	68.77	55.16	39.22	46.22	39.11	57.01	49.83	50.76
Recall	61.64	51.90	35.29	37.01	30.09	49.84	41.28	43.86

Table 2: Scores of proposed Fine-tuned DistilBERT model with base uncased

Figure 2: Metrics score of Fine-tuned DistilBERT

Metric	Patronizing and Condescending Language Categories							
	UPR	SS	PS	AV	MP	CP	TPTM	Macro Avg
F1	65.75	53.92	37.98	41.86	34.97	52.98	47.12	47.80
Precision	69.99	55.84	40.33	47.25	39.39	56.47	50.36	51.38
Recall	61.99	52.13	35.89	37.58	31.44	49.90	44.28	44.74

Table 3: Scores of proposed Data Augmentation with Text Embedding

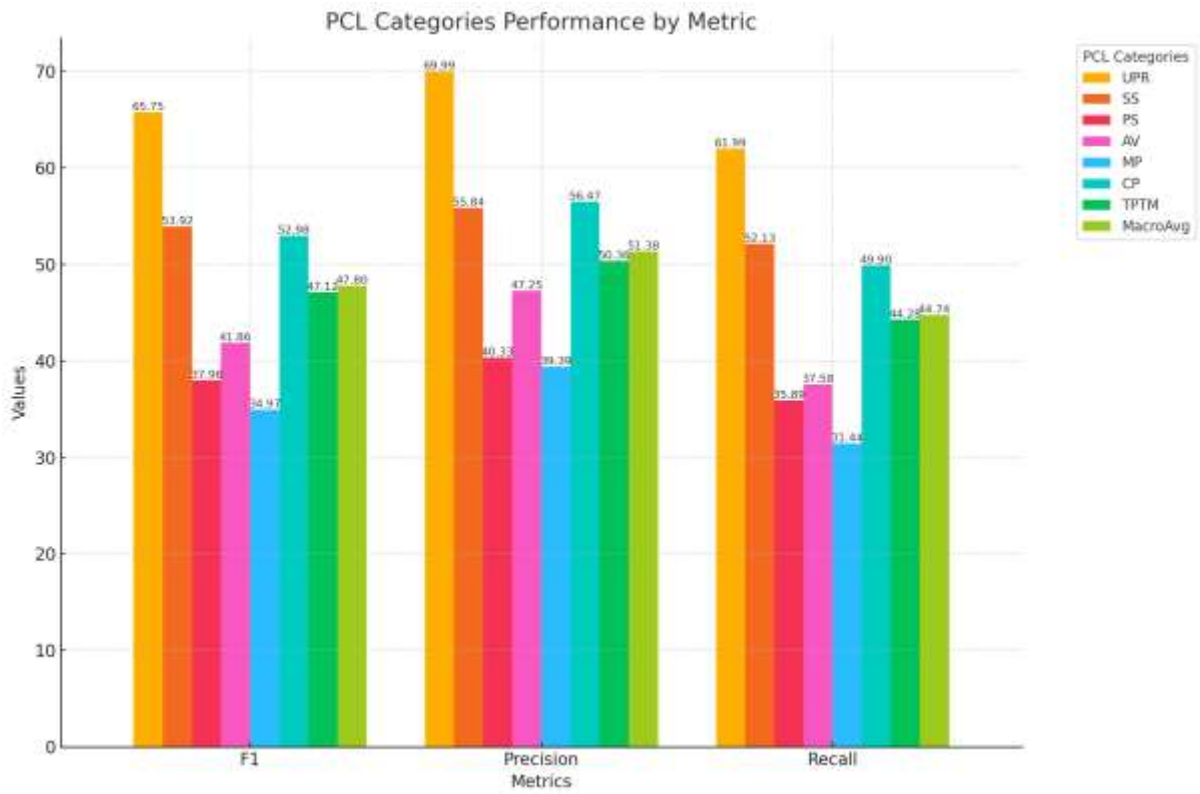

Figure 3: Metrics score of Fine-tuned DistilBERT with Data Augmentation Contextual Word Embedding

Metric	Patronizing and Condescending Language Categories							
	UPR	**SS**	**PS**	**AV**	**MP**	**CP**	**TPTM**	**Macro Avg**
F1	65.98	54.21	38.45	42.58	42.87	53.48	47.24	**49.26**
Precision	70.49	56.14	41.40	48.60	46.81	56.56	50.62	**52.95**
Recall	62.01	52.41	35.89	37.88	39.54	50.71	44.28	**46.10**

Table 4: Scores of proposed Data Augmentation with Synonym Replacement

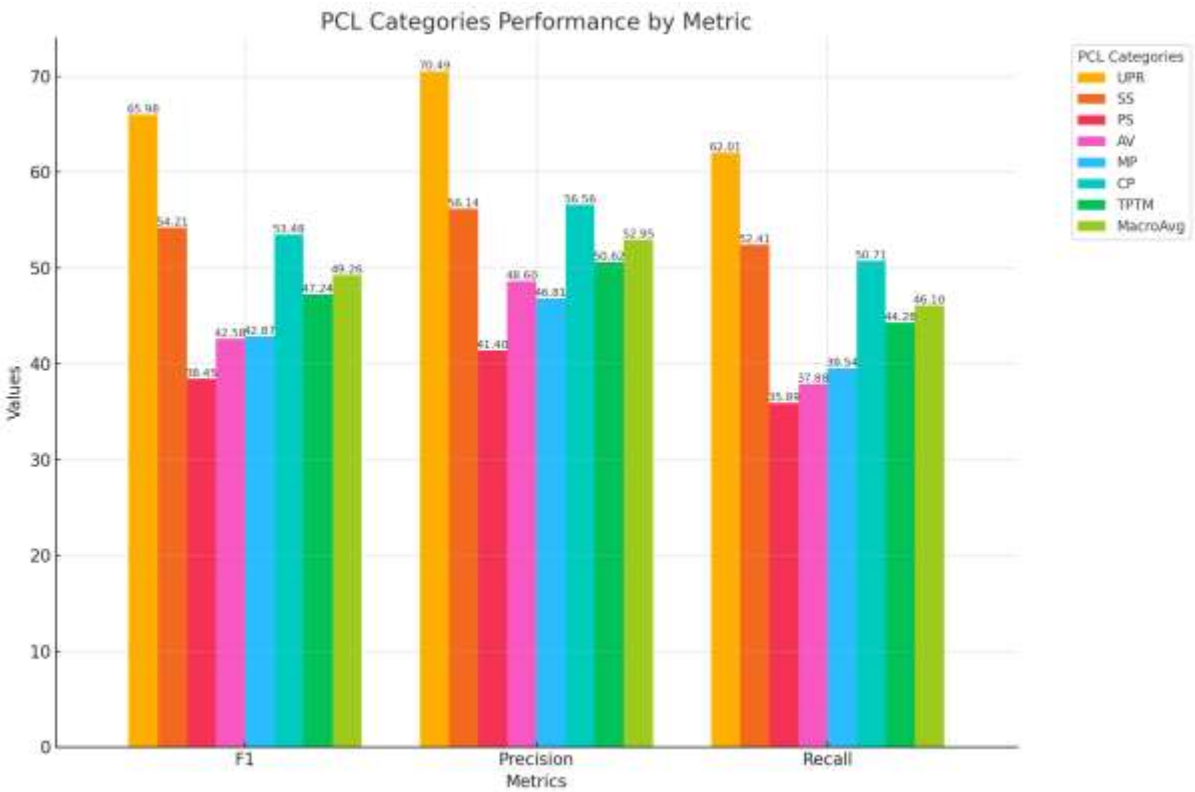

Figure 4: Metrics score of Fine-tuned DistilBERT with Data Augmentation Synonym Replacement

Metric	Patronizing and Condescending Language Categories							
	UPR	**SS**	**PS**	**AV**	**MP**	**CP**	**TPTM**	**Macro Avg**
F1	64.74	55.48	37.15	41.47	34.22	53.18	45.15	47.34
Precision	68.75	59.14	40.40	46.60	38.11	56.27	48.18	51.06
Recall	61.17	52.25	34.39	37.36	31.05	50.41	42.48	44.16

Table 5: Scores of proposed Data Augmentation with Back Translation

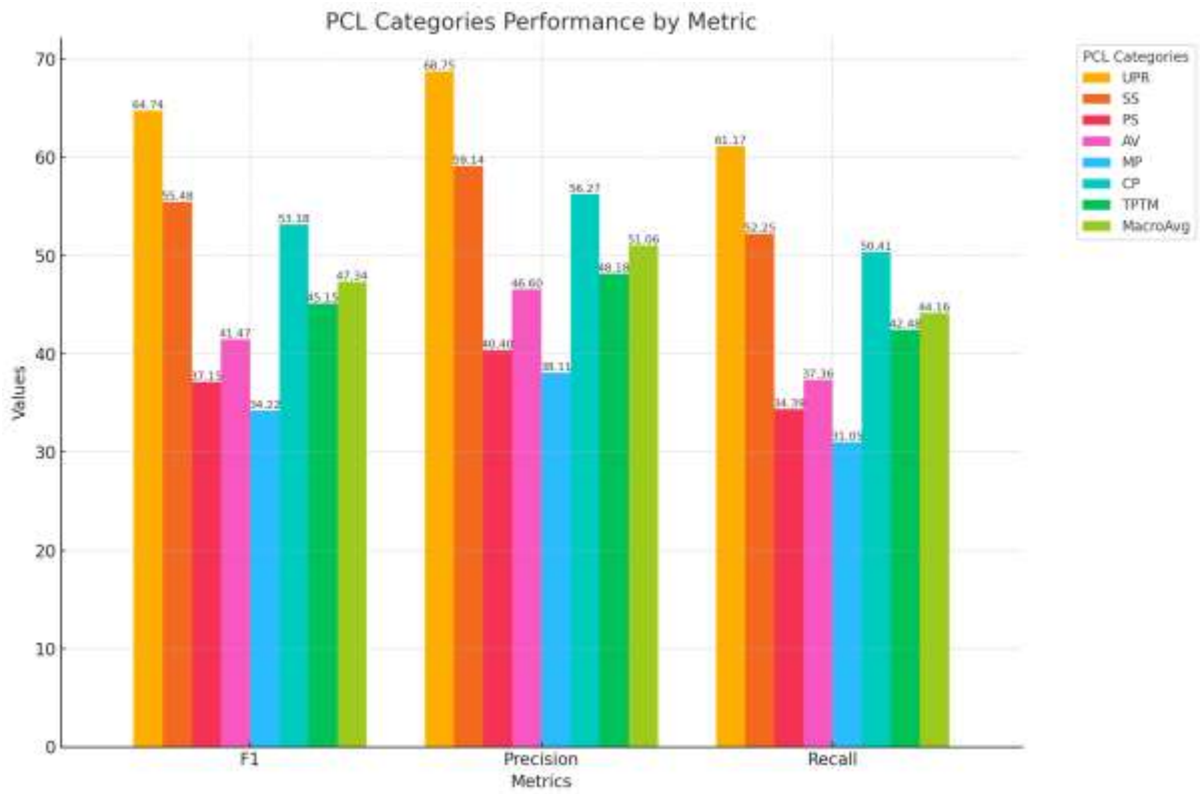

Figure 5: Metrics score of Fine-tuned DistilBERT with Data Augmentation back translation

Models	**Model Name**	**UPR**	**SS**	**PS**	**AV**	**MP**	**CP**	**TPTM**	**Macro Avg**
Proposed Models	Data Augmentation Synonym Replacement	65.98	54.21	38.45	42.58	42.87	53.48	47.24	**49.26**
	Data Augmentation Contextual Word Embedding	65.75	53.92	37.98	41.86	34.97	52.98	47.12	47.80
	Data Augmentation Back Translation	64.74	55.48	37.15	41.47	34.22	53.18	45.15	47.34
	DistilBERT Base Uncased	65.01	53.48	37.15	41.11	34.01	53.18	45.15	47.01
Compared Models	Prompt based model	65.60	52.90	36.90	40.70	35.90	49.20	47.10	46.90
	Prompt training and Label attention mechanism	59.70	53.10	41.70	43.40	42.70	51.30	15.40	43.90

Table 6: Scores(macro F1) of Proposed and Comparing Models for all the categories

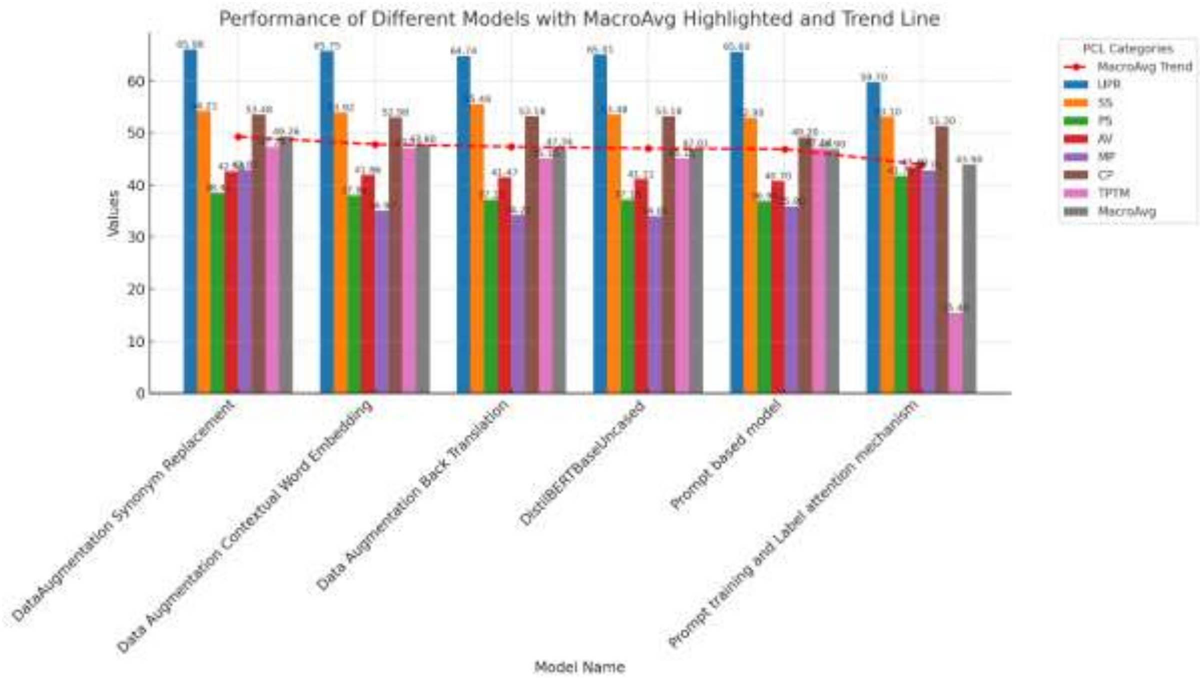

Figure 6: Macro F1 score of proposed and compared models

4 Conclusion and Future Work

PCL Detection is a Binary Classification problem. Detected PCL can be categorised into various groups. Categorizing into groups is a multi label classification problem. This research proposed a fine-tuned DistilBERT pre trained model and Data Augmentation for PCL categorization. Proposed models have performed well and achieved the best score than the previous models. Data augmentation with synonym replacement method achieved a good macro f1 score of 49.26. This work has used back translation English to German and German to English. Data augmentation with Contextual Word Embedding also shown a better performance. In future large case pre-trained models can be applied on this dataset, also a different back translation language can be applied. In the proposed model due to high threshold it is able to detect true positives, there is a chance to miss some of the true positives due to high sensitivity. This also can be addressed in the further improvements.

References

1. Paula Fortuna and Sergio Nunes. 2018. A survey on automatic detection of hate speech in text. ACM Computing Surveys (CSUR), 51(4):1–30.
2. Amir H Razavi, Diana Inkpen, Sasha Uritsky, and Stan Matwin. 2010. Offensive language detection using multi-level classification. In Canadian Conference on Artificial Intelligence, pages 16–27. Springer.

3. Ray Oshikawa, Jing Qian, and William Yang Wang.2018. A survey on natural language processing for fake news detection. arXiv preprint arXiv:1811.00770.
4. Xinyi Zhou and Reza Zafarani. 2020. A survey of fake news: Fundamental theories, detection methods, and opportunities. ACM Computing Surveys (CSUR), 53(5):1–40.
5. Carla Perez-Almendros, Luis Espinosa-Anke, and Steven Schockaert. 2020. Don't patronize me! An annotated dataset with patronizing and condescending language towards vulnerable communities. In Proceedings of the 28th International Conference on Computational Linguistics, pages 5891–5902.
6. Zijian Wang and Christopher Potts. 2019. Talkdown: A corpus for condescension detection in context. In Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing.
7. Gloria Wong, Annie O Derthick, EJR David, Anne Saw, and Sumie Okazaki. 2014. The hat, the why, and the how: A review of racial microaggressions research in psychology. Race and social problems, 6(2):181–200.
8. Thomas Huckin. 2002. Critical discourse analysis and the discourse of condescension. Discourse studies in composition, 155:176.
9. Susan T Fiske. 1993. Controlling other people: The impact of power on stereotyping. American psychologist, 48(6):621.
10. Carla Perez-Almendros, Luis Espinosa-Anke, and Steven Schockaert. 2022. SemEval-2022 Task 4: Patronizing and Condescending Language Detection. In Proceedings of the 16th InternationalWorkshop on Semantic Evaluation (SemEval-2022). Association for Computational Linguistics.
11. Valerio Basile, Cristina Bosco, Elisabetta Fersini, Debora Nozza, Viviana Patti, Francisco Manuel Rangel Pardo, Paolo Rosso, and Manuela Sanguinetti. 2019. Semeval-2019 task 5: Multilingual detection of hate speech against immigrants and women in twitter. In Proceedings of the 13th International Workshop on Semantic Evaluation, pages 54–63.
12. Marcos Zampieri, Shervin Malmasi, Preslav Nakov, Sara Rosenthal, Noura Farra, and Ritesh Kumar. 2019. Semeval-2019 task 6: Identifying and categorizing offensive language in social media (offenseval). In Proceedings of the 13th International Workshop on Semantic Evaluation, pages 75–86.
13. Dou Hu, Mangyuan Zhou, Xiyang Du, Mengfei Yuan, Meizhi Jin, Lianxin Jiang, Yang Mo and Xiaofeng Shie. 2022. PALI-NLP at SemEval-2022 Task 4: Discriminative

Fine-tuning of Transformers for Patronizing and Condescending Language Detection. In Proceedings of 16th International Workshop on Semantic Evaluation (SemEval-2022). Pages 335-343.

14. Xueyao Zhang, Juan Cao, Xirong Li, Qiang Sheng, Lei Zhong, and Kai Shu. 2021b. Mining dual emotion for fake news detection. In WWW, pages 3465–3476. ACM / IW3C2.
15. Lingwei Wei, Dou Hu, Wei Zhou, Zhaojuan Yue, and Songlin Hu. 2021. Towards propagation uncertainty: Edge-enhanced bayesian graph convolutional networks for rumor detection. In ACL/IJCNLP (1), pages 3845–3854. Association for Computational Linguistics.
16. Howard Giles, Susan Fox, and Elisa Smith. 1993. Patronizing the elderly: Intergenerational evaluations. Research on Language and Social Interaction, 26(2):129–149.
17. Lilie Chouliaraki. 2006. The spectatorship of suffering. Sage.
18. Branka Drljaˇca Margi´c. 2017. Communication courtesy or condescension? linguistic ccommodation of native to non-native speakers of english. Journal of English as a lingua franca, 6(1):29–55.
19. Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: pre-training of deep bidirectional transformers for language understanding. In NAACL-HLT (1), pages 4171–4186. Association for Computational Linguistics.
20. Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. Roberta: A robustly optimized BERT pretraining approach. CoRR, abs/1907.11692.
21. Thomas Davidson, Dana Warmsley, Michael W. Macy, and Ingmar Weber. 2017. Automated hate speech detection and the problem of offensive language. In ICWSM.
22. Genevieve Gorrell, Elena Kochkina, Maria Liakata, Ahmet Aker, Arkaitz Zubiaga, Kalina Bontcheva, and Leon Derczynski. 2019. SemEval-2019 task 7: RumourEval, determining rumour veracity and support for rumours. In Proceedings of the 13th International Workshop on Semantic Evaluation, pages 845–854, Minneapolis, Minnesota, USA. Association for Computational Linguistics.
23. Niall J Conroy, Victoria L Rubin, and Yimin Chen. 2015. Automatic deception detection: Methods for finding fake news. Proceedings of the Association for Information Science and Technology, 52(1):1–4.

24. Alberto Barron-Cedeno, T. Elsayed, Reem Suwaileh, Lluis Marquez i Villodre, Pepa Atanasova, Wajdi Zaghouani, Spas Kyuchukov, Giovanni Da San Martino, and Preslav Nakov. 2018. Overview of the clef-2018 checkthat! lab on automatic identification and verification of political claims. task 2: Factuality. In CLEF.
25. Maarten Sap, Saadia Gabriel, Lianhui Qin, Dan Jurafsky, Noah A Smith, and Yejin Choi. 2020. Social bias frames: Reasoning about social and power implications of language. In Association for Computational Linguistics.
26. Krzysztof Dembczy´nski, Weiwei Cheng, and Eyke Hullermeier. 2010. Bayes optimal multilabel classification via probabilistic classifier chains. In Proceedings of the 27th International Conference on International Conference on Machine Learning, ICML'10, page 279–286, Madison, WI, USA. Omnipress.
27. Matthew R. Boutell, Jiebo Luo, Xipeng Shen, and Christopher M. Brown. 2004. Learning multilabel scene classification. Pattern Recognition, 37(9):1757–1771.
28. Weizhu Chen, Jun Yan, Benyu Zhang, Zheng Chen, and Qiang Yang. 2007. Document transformation for multi-label feature selection in text categorization. In Seventh IEEE International Conference on Data Mining (ICDM 2007), pages 451–456.
29. Everton Alvares-Cherman, Jean Metz, and Maria Carolina Monard. 2012. Incorporating label dependency into the binary relevance framework for multi-label classification. Expert Syst. Appl., 39(2):1647–1655.
30. Newton Spolaor, Everton Alvares Cherman, Maria Carolina Monard, and Huei Diana Lee. 2013. A comparison of multi-label feature selection methods using the problem transformation approach. Electronic Notes in Theoretical Computer Science, 292:135–151. Proceedings of the XXXVIII Latin American Conference in Informatics (CLEI).
31. Bingyu Wang, Cheng Li, Virgil Pavlu, and Javed A. Aslam. 2017. Regularizing model complexity and label structure for multi-label text classification. CoRR,abs/1705.00740.
32. Jinfen Li and Lu Xiao. 2020. syrapropa at semeval-2020 task 11: Bert-based models design for propagandistic technique and span detection.
33. Tharindu Ranasinghe and Hansi Hettiarachchi. 2020. BRUMS at SemEval-2020 task 12: Transformer based multilingual offensive language identification in social media. In Proceedings of the Fourteenth Workshop on Semantic Evaluation, pages 1906–1915, Barcelona (online). International Committee for Computational Linguistics.
34. Dan Li, Zi Long Zhu, Janneke van de Loo, Agnes Masip Gomez, Vikrant Yadav, Georgios Tsatsaronis, Zubair Afzal. (2023). Enhancing Extreme Multi-Label Text

Classification: Addressing Challenges in Model, Data, and Evaluation. Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing: Industry Track, pages 313–321

35. Aarathi Rajagopalan Nair, Rimjhim Padam Singh, Deepa Gupta, Priyanka Kumar. (2024). Evaluating the Impact of Text Data Augmentation on Text Classification Tasks using DistilBERT. International Conference on Machine Learning and Data Engineering(ICMLDE 2023). Procedia Computer Science 235 (2024) 102–111. 10.1016/j.procs.2024.04.013.
36. Yong Dong, Chenxiao Dou, Liangyu Chen, Deqiang Miao, Xianghui Sun, Baochang Ma and Xiangang Li. 2022. BEIKE NLP at SemEval-2022 Task 4: Prompt-Based Paragraph Classification for Patronizing and Condescending Language Detection. In Proceedings of 16th International Workshop on Semantic Evaluation (SemEval-2022). Pages 319-323.
37. Ye Wang, YanmengWang, Baishun Ling, Zexiang Liao, Shaojun Wang and Jing Xiao. (2022). PINGAN Omini-Sinitic at SemEval-2022 Task 4: Multi-prompt Training for Patronizing and Condescending Language Detection. In Proceedings of 16th International Workshop on Semantic Evaluation (SemEval-2022). Pages 319-323.
38. Sanh, V., Debut, L., Chaumond, J., & Wolf, T. (2019). DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter. arXiv preprint arXiv:1910.01108.
39. Wei, J., & Zou, K. (2019). EDA: Easy Data Augmentation Techniques for Boosting Performance on Text Classification Tasks. arXiv preprint arXiv:1901.11196.
40. Kobayashi, S. (2018). Contextual Augmentation: Data Augmentation by Words with Paradigmatic Relations. arXiv preprint arXiv:1805.06201.
41. Sennrich, R., Haddow, B., & Birch, A. (2016). Improving Neural Machine Translation Models with Monolingual Data. arXiv preprint arXiv:1606.04164.
42. Bolukonda Prashanth, Dr P Vijaya Pal Reddy. (2024). "Use of a refined Distil-BERT model and Data Augmentation to Identify Patronizing and Condescending Language", African Journal of Biological Sciences, vol.6, no. 10, pp.575805768, June, 2024. *https://doi.org/10.48047/AFJBS.6.10.2024.5758-5768*
43. Nandula Anuradha, Dr P Vijaya Pal Reddy. (2024). "Deep Hybrid models with BERT for Cross Domain Suggestion Classification", African Journal of Biological Sciences, vol.6, no. 4, pp.1076-1087, June, 2024. https://doi.org/ 10.33472/AFJBS.6.4.2024.1076-1086.