ICSOFT 2023 - Paper #62

Paper Title: *Effectiveness of Data Augmentation and Ensembling using Transformer-based Models for Sentiment Analysis: Software Engineering Perspective*

Reviewer *#1*

General Assessment (Please assign scores using the following criteria (1=weakest; 6=strongest))

Relevance (Paper fits one or more of the topic areas?): 4
Originality (Newness of the ideas expressed): 3
Technical Quality (Theoretical soundness/methodology): 3
Significance (Is the problem worth the given attention?): 4
Presentation (Structure/Length/English): 4
Overall Rating (Weighted value of above items): 4

Improvement Suggestions (for authors to consider in the camera-ready version. Additional detail in "Observations")

Abstract and Introduction are adequate? Yes
Needs more experimental results? No
Needs comparative evaluation? No
Improve critical discussion ? (validation): No
Figures are adequate ? (in number and quality): Yes
Conclusions/Future Work are convincing? Yes
References are up-to-date and appropriate? Yes
Paper formatting needs adjustment? No
Improve English? No

Detailed comments to authors, including aspects that must be improved in the camera-ready version of the paper:

The authors used context−specific data augmentation using SE−specific vocabularies and ensembled multiple models to classify sentiment polarity.
They presented the improved performance compared with the conventional approaches.

The result is good enough.
However, the novel points are not clearly described.
For example, the method applied in the authors approach to resolve bias problem caused by class imbalance is generally accepted approach.
So, it may not be the novel point of this research.
Furthermore, the authors augmented text to ensure equal distribution in every class.
This approach may not be the novel point of their research.

When the authors submit their final version for the conference, the reviewer strongly encouraged to focus on the novel point of the authors' work.
The reviewer is convinced that the ensemble model is the core part of this paper.

The improved results were obtained by the ensemble approach.
As the authors mentioned that the results depend on the characteristics of data set, please summarize the dependencies in the final part of this paper.
This information can be a valuable for attendees and readers of the conference.

ICSOFT 2023 - Paper #62

Paper Title: *Effectiveness of Data Augmentation and Ensembling using Transformer-based Models for Sentiment Analysis: Software Engineering Perspective*

Reviewer *#2*

General Assessment (Please assign scores using the following criteria (1=weakest; 6=strongest))

Relevance (Paper fits one or more of the topic areas?): 5
Originality (Newness of the ideas expressed): 5
Technical Quality (Theoretical soundness/methodology): 5
Significance (Is the problem worth the given attention?): 5
Presentation (Structure/Length/English): 5
Overall Rating (Weighted value of above items): 5

Improvement Suggestions (for authors to consider in the camera-ready version. Additional detail in "Observations")

Abstract and Introduction are adequate? Yes
Needs more experimental results? No
Needs comparative evaluation? No
Improve critical discussion ? (validation): No
Figures are adequate ? (in number and quality): No
Conclusions/Future Work are convincing? Yes
References are up-to-date and appropriate? Yes
Paper formatting needs adjustment? No
Improve English? No

Detailed comments to authors, including aspects that must be improved in the camera-ready version of the paper:

The paper entitled: "Effectiveness of Data Augmentation and Ensembling using Transformer−based Models for Sentiment Analysis: Software Engineering Perspective" presents a ccontext−specific data augmentation using SE−specific vocabularies and multiple ensembled models to classify sentiment polarity to consider the efficiency of ensembling multiple pre−trained models on SE−specific datasets.

In general, the paper is well−written and covers an interesting topic. We enjoyed reading the paper. Congratulations to the authors.

I only have one minor comment to improve the paper since it is well contextualized, proposes a method that is well applied, and the results illustrate improvements in the state of the art. We recommend the authors include a new figure (if there is enough space), instantiating the method of Figure 1 with the experimentation results. In this context, readers may have a complete view of the paper by simply looking at two figures.

We also recommend the authors to extend the limitations and future work sections/paragraphs to reinforce their

conclusions.

ICSOFT 2023 - Paper #62

Paper Title: *Effectiveness of Data Augmentation and Ensembling using Transformer-based Models for Sentiment Analysis: Software Engineering Perspective*

Reviewer *#3*

General Assessment (Please assign scores using the following criteria (1=weakest; 6=strongest))

Relevance (Paper fits one or more of the topic areas?): 2
Originality (Newness of the ideas expressed): 3
Technical Quality (Theoretical soundness/methodology): 4
Significance (Is the problem worth the given attention?): 3
Presentation (Structure/Length/English): 5
Overall Rating (Weighted value of above items): 2

Improvement Suggestions (for authors to consider in the camera-ready version. Additional detail in "Observations")

Abstract and Introduction are adequate? Yes
Needs more experimental results? Yes
Needs comparative evaluation? No
Improve critical discussion ? (validation): Yes
Figures are adequate ? (in number and quality): No
Conclusions/Future Work are convincing? No
References are up-to-date and appropriate? Yes
Paper formatting needs adjustment? No
Improve English? No

Detailed comments to authors, including aspects that must be improved in the camera-ready version of the paper:

First, I am not convinced that sentiment analysis (that is, estimating the emotional state of a writer) benefits software engineering.

Second, the paper is about building a better tool for sentiment analysis, which has only very indirect connection to software engineering at all.

Third, I am concerned about creating a *useful* larger training set by mutating (augmenting) text samples is valid. I did not see evidence that the mutations yield text similar to an actual larger sample. The mutations may simply amplify data skews that are already present. It may have the effect of simply training on the same samples many times. It may result in overtraining.

I do compliment you on using reasonable numbers, that is, two decimal places. Too many papers use three (or more!) decimal places with no justification.

That said, fourth, your statistical analysis lacks summarizing numbers. There are many tables with lots of numbers, but no computations of statistical significance or acceptance (or rejection) of a hypothesis.

RQ1 says, "The ensemble models outperform [other] models by a significant margin for three out of the four datasets showing ..." I did not see computation of significance. I am especially concerned since no benefit is found for one−quarter of datasets!

The paper could be more convincing (and shorter!) if all the tables were replaced with a few summarizing statistics.

ICSOFT 2023 - Paper #62

Paper Title: *Effectiveness of Data Augmentation and Ensembling using Transformer-based Models for Sentiment Analysis: Software Engineering Perspective*

Reviewer *#4*

General Assessment (Please assign scores using the following criteria (1=weakest; 6=strongest))

Relevance (Paper fits one or more of the topic areas?): 5
Originality (Newness of the ideas expressed): 3
Technical Quality (Theoretical soundness/methodology): 4
Significance (Is the problem worth the given attention?): 4
Presentation (Structure/Length/English): 4
Overall Rating (Weighted value of above items): 4

Improvement Suggestions (for authors to consider in the camera-ready version. Additional detail in "Observations")

Abstract and Introduction are adequate? Yes
Needs more experimental results? No
Needs comparative evaluation? No
Improve critical discussion ? (validation): No
Figures are adequate ? (in number and quality): Yes
Conclusions/Future Work are convincing? Yes
References are up-to-date and appropriate? Yes
Paper formatting needs adjustment? Yes
Improve English? No

Detailed comments to authors, including aspects that must be improved in the camera-ready version of the paper:

his paper presents a study to compare the performances of an SE−specific data augmentation technique in ensemble models to those of pre−trained models to classify sentiment polarity.
I am not an expert of the field and my analysis could be biased.
The topic is specific and narrow in scope. However, the study (as far as I can judge) is carefully and seriously conducted. Experiments are based on well−established data sets. The paper is well structured and clearly written. The results are various and informative.
Globally, the paper presents an interest by reporting on different variations of established techniques to detect sentiment polarity. It offers a concrete adaptation technique to mitigate the issue of small data set size and class imbalance over its samples.