**Review 1.**

Please describe what the author has done, in a couple of sentences:

First, the author indicated the method and dataset he used. In the String Classification section, he introduced his method of classification and provided the reason why this operation should be considered with the strong reference. Then he illustrated detailed evaluation and matrix. Finally, the features he used, methodology, result and conclusion, including various tables and concepts, formulas and improvement, are presented. Overall, the author demonstrated six supervised machine learning methods, which provided significant stringency to this report.

Please indicate what you think the author has done well, and why:

In my opinion, the most appreciative aspect is the clear and well-organised statement. With each heading, it showed the content clearly. The author also referred some formulas and expressions to state the academic concepts. Another highlight is the tables the author used. Although there are six classifiers which are required to illustrate and compare, the charts showed the contrast and result clearly, which avoid the chaos of text description. In the end, the bibliographies and citations displayed that the author has done sufficient research works.

Please indicate what you think could have been improved, and why (we would prefer you to focus on the content, rather than the language, structure, or style):

First, some of the sections are a little bit redundant. In my opinion, some sections could be integrated, such as introduction, dataset, model background. I think these sections should be included in an Introduction part with subheadings. Meanwhile, it showed two evaluation sections, which may lead to the confusion. Moreover, some concepts were not demonstrated well. For instance, the author stated the new features were added without stop words. Although I know what the stop words mean, as a report, the author should interpret this concept to the public in one or two sentences. And lastly, the matrix section is missing. The author mentioned the confusion matrix in the previous section, but in the result and evaluation part, it just contains accuracy, precision and root mean square error in the three tables. Some concepts and data in the matrix are not shown in the report, such as TP, TN…etc.

**Review 2.**

Please describe what the author has done, in a couple of sentences:

First, the author indicated the background, the method and dataset he used. Then the author demonstrated the methodology he applied with the attributes he added. In the last content, he illustrated the evaluation, analysis, improvement and conclusion with the precise interpretation of various concepts.

Please indicate what you think the author has done well, and why:

Personally, I like the structure and organisation of the article. The author interpreted every operation and each file he used clearly. And the analysis part is also detailed and comprehensive. In addition, some important concepts are explained well, such as TP, TN, precision, confusion matrix. Lastly, in the explore section, the author improved the selection of attributes after the first trial and did the contrastive analysis, which supports the thought of the author strongly.

Please indicate what you think could have been improved, and why (we would prefer you to focus on the content, rather than the language, structure, or style):

In my opinion, the methodology (add features) the author used should be improved. His method is to find some sets of specific words, such as ‘but I’ and ‘sleep paralysis’, which is not universal to operate on the entire tweet dataset. If there is a tweet with other symptoms and not included these words, it is possible to be ignored in the evaluation process. Moreover, the critical thinking is not sufficient. In the explore section, the author just analysed the possible reasons from the surface of the dataset and attributes. For instance, the order of attribute, interference of other attributes. I think he needs to consider the problem deep and did more researches. Actually, the variation of precision and accuracy is also related to many factors, like the chosen classifier, Weka’s overfitting and selected attributes. Meanwhile, the conclusion is hasty, just one sentence to repeat the conclusion and consideration in the explore section. Last but not least, the number of references is only two, which is insufficient to support the report.