1

Briefly summarise what the author has done in one paragraph (50-100 words)

The author has only using meta and audio features to investigate whether using only one type of features with balanced training data would increase the overall performance in terms of accuracy for classifiers (K-Nearest Neighbour (Knn) and Multinomial Logistic Regression (Mlr)). The author found that having continuous features will increase the accuracy score compared to only using discrete features or a combination of both. The author found that the under-sampling method failed to improve the model's performance.

Indicate what you think that the author has done well, and why in one paragraph (100-200 words)

For method, the author has a clear hypothesis of why textual features are not considered. The author has applied standardization to magnitude sensitive model such as Knn otherwise Knn's prediction will be dominated by features have large values or high proportions. For critical analysis, effect comparison is applied as two models are compared with each other as well the baseline (One-R). It is good to see that the author has noticed the imbalanced class distribution in the train and valid dataset because the imbalanced class distribution can influence the performance of models such as naïve Bayes and Knn. To solve this issue, the author has attempted the under-sampling method. I did not realize this approach. For report quality, the reference is in the correct format and peer-reviewed. The report has a moderate length. The report is structured in a sensible way for me because my report also has a similar structure.

Indicate what you think could have been improved, and why in one paragraph (50-100 words)

The caption in table 2 is too small. It is difficult for me to read and it appeared to be pixels when I try to zoom in. For the error analysis, I think it is better to draw a learning curve to see how models preform when altering the training size. Thus, to conclude that the model is overfitting because it has a higher training score than a validation score when training size is large. The report has figure 1 and 4, but where are figure 2 and 3? Hyper-parameter tuning for models seems missing.

2

Briefly summarise what the author has done in one paragraph (50-100 words)

The author investigates that whether a combination of audio and text features is more predictable in terms of accuracy than using them separately for models such as Gaussian naïve Bayes (Gnb), K-Nearest Neighbors (Knn,), Decision Tree (Dt) and Multi-layer Perceptron (Mlp). The author has found that the combination of audio and text features performs best across all candidate models. The author has found that Mlp performs the best among all models because it has the feature engineering ability which can generate useful features based on the original features for predicting.

Indicate what you think that the author has done well, and why in one paragraph (100-200

words)

For method, the author has stated the used evaluation metric clearly with appropriate reasoning. The author has applied standardization to magnitude sensitive model such as Knn otherwise Knn's prediction will be dominated by features have large values or high proportions. For critical analysis, a lot of models are compared which makes reasoning more reliable and trustable. For report quality, the reference is in the correct format and peer-reviewed. There is some discussion about why models perform good or bad such as some data-set's features are not Gaussian distributed. This shows the author has thinking the theory behind rather than looking at accuracy and give a superficial conclusion. The report has a moderate length. The report is structured in a sensible way for me because my report also has a similar structure. The accuracies are listed in tables rather than figures which are clear and intuitive.

Indicate what you think could have been improved, and why in one paragraph (50-100 words)

The tuning of Dt's hyper-parameter max_depth. The distance metric for Knn is missing. I cannot see whether it is Euclidean or Manhattan. It's better to mention them otherwise this issue makes the report less convincing and models are not performing optimally. The author has mentioned the noise in data-set can influence the performance of Knn. However, it's better to investigate it further to see what the noise are. For example, finding the closest neighbours for wrongly predicted valid instances. The author chooses not to use meta-features. But I didn't see reasoning which makes hard for me to follow the discussions.