

Attempt 1 Feedback:

May 27, 2021 at 6:25

This report focuses on steps and ingredients as main features. In pre-processing part, the author chooses N-gram and TF-IDF, with the combining of removing punctuation and stop words. This can increase the difference between instances and improve the meaningful data's influence on the classifier. And the two main pre-processing tools filter the dataset to make them more suitable for the classifier. Moreover, this report implements 3 different types of classifiers: logistic regression, decision tree, and linear SVM. Each classifier's analysis contains one figure to show the changing variance's effect on the model's accuracy. The addition of graph really exhibits the result clearly. In the evaluation part, combining with 3 different confusion matrix graphs, the author shows predicted labels' distribution in the classifiers. In my opinion, the author shows excellent ability on representation, the balance between figures and words motivates the report to be vivid and pellucid. However, there is a lack of thinking. For example, why different classifiers own different results? Why choosing the steps and ingredients as the main features? Moreover, further discussion on contrast is expected. The author only compares accuracy in the comparison part, it is better to position discussion and guess in terms of reasons for the formation of these results.

- Anonymous User

May 28, 2021 at 17:04

project-2-report.pdf- This report was put together very well; each section was well defined, with sub-sections being relevant and informative. Simple definitions were provided for text classification terms, which would make the report easy to grasp for beginners / those who haven't learnt about ML. The report structure allowed the author to build upon each concept they talked about in the previous section. The author included many useful, relevant visualisations in the report, perfectly illustrating the decision making process behind parameters / model selection. The error analysis section was short, but included good information, lending to the model's behaviour and why it might've struggled with classifying certain instances / why there was a disparity in prediction. Referencing was also done quite well, being in-text citations, which made the report very easy to read. There were some minor inconsistencies throughout the report, but nothing major. These included grammatical errors, capitalisation and the author included a quick sentence or two at the start of each model to give a basic understanding of how classification worked. This was a particularly good inclusion, however was missed for Logistic Regression. Contextualisation of model behaviour was not as in detail as it could have been, as the error analysis talked more to issues with prediction due to class bias in training data, rather than explaining poor (or good) performance behind each specific model. There was no reference to other works (other than with the citations), so whether or not there was a requirement for an explicit Literature Review is up to the assessors.

- Anonymous User