

COMP30027 Machine Learning - Project 2

Peer Review

May 25, 2021

Review 1 (392 Words)

The report clearly demonstrates what the author has done in training and evaluating classifier label. Initially, data selection is discussed, explaining the choice of text processing method and removing outliers from training data using Z-score method. Then, report proceeds to training by using three distinct machine learning models: Random Forest, K-Nearest Neighbor and Neural Networks along with three ensemble classifiers: bagging, voting, and stacking. Results and error analysis are discussed in detail with the aid of confusion matrix and various evaluation metrics along with accuracy, i.e precision and recall. Finally, potential improvements to model training are suggested including finding different base classifiers, changing splitting ratio or considering more evaluation metrics.

Good points in the report includes clear step procedures in training, occasionally with a short brief explanation of the choice selected, and conceptual explanation in error analysis. Hence, disadvantage of having few instances of label 3.0 in training dataset are realised. Additionally, outlier analysis is conducted along with its benefits and drawbacks.

However, there exist several points where the report can be improved. Firstly, in terms of formatting, the report should be divided into sections and had titles for each section. In that way, the report would seem more like a scientific report rather than an essay.

Regarding feature selection, it is mentioned that method doc2vec50 is selected due to its order preservation; doc2vec100 also possesses the same advantage, but is not investigated, even though these text processing methods can be visually compared by plots. Furthermore, there is lack of sufficient summary for data representation: which features have been selected in training, whether numerical features including `n_steps` or `n_ingredients` are included.

Also, the author uses Z-score as an outlier analysis. However, there is lack of information about why the technique or threshold is chosen. Although the advantages and disadvantages of having more or less training data are abstractly discussed, a demonstrative plot can make the point clearer.

Furthermore, model, parameters and hyper-parameters selection are stated, but not explained thoroughly. There should have been more justifications on why a model is chosen, i.e, what advantages does a model has over another, specifically in the scenario. One suggestion is to have a learning curve plot, indicating the relationship between hyper-parameters and model evaluation metrics. Lastly, results and error analysis can be more relevant to the choice of model and hyper-parameter, rather than 'randomness of sampling'.

Review 2 (387 Words)

The report shows decent effort of classifying recipes, mainly focusing on feature selection and model evaluation. The first section begins on feature selection, where the author listed out methods of choosing the optimal subset of attributes, each have its implementations explained and analysed to select the best method of choosing attribute subset, which is Random Forest. The report then moves on to model evaluation, where two models along with cross-validation are briefly described. At the end, different combinations of models and feature selection methods are evaluated using 10-Fold Cross Validation, and Support Vector Machine was chosen to be the most optimal model.

The report has a readable structure, with author's intentions and ideas pointed out in an easy-to-follow manner, yet it contains many formatting problems. There is inconsistency in the writing of algorithm names, such as 'Random Forest', 'random forest' (2.4). More formal rewriting or explanation needs to be applied to arbitrary words, such as '40K' (2.1), 'Y/N' (3.2). Section number is not clear, where it becomes 2.2.1 after Section 3.2. A number of grammar mistakes are also found, such as 'will preference features' and 'not unlike' (2.3).

Conclusions of choices are backed with logical reasoning, as well as resulting outputs at a conceptual level, which is suitable for students and academic readers. However, they can be further improved using plots and data figures to prove the key points more concisely. The report also lacks researching, with only two references in the bibliography, hence weaken the arguments' validity.

On topics, while feature selection takes a large part in the report, it is not included in the overall conclusion. It is also not clear which features are used for selection or evaluation. For model evaluation, result difference of 0.1% is too small to conclude the optimal model, given that Cross Validation results can vary for the same inputs. It is also suspicious that F1-score is identical to accuracy, given that they have different calculation methods. Error analysis is not clear in the report.

Overall, the report is adequate, but requires major revision, given the lack of content and reasoning. Consistent naming, grammar correction and appropriate wordings are needed. For concepts, more research and suitable evaluation factors are needed to improve argument strength, such as plotting, error and recall measures. Structure reorganisation is also needed to improve report cohesion.