

Review 1

Response to “not enough visualization”

In our poster, graphs are in feature selection section which is the most important part should be presented, two methods (remove highly correlated variables and feature importance), we state the heatmap of variables correlation and bar chart of normalized importance.

Model details account for nearly half of our poster, but it is necessary because all four models in our project was not included in our lecture which means the explanation is needed for providing basic information.

Response to “algorithm is common”

Our training algorithms are as follows: Naïve Bayes, Adaboost, SGD and LGB

From the survey in Kaggle 2017 (slide from lecture 6), Bayesian Techniques and GBM are less common than other models. Also, LGB model is a new model for our team that our team members have no experience of that.

Response to “the steps are standard, no unique way or method to deal with model and dataset”

In each section, we make sure the variety of configurations. For example, 5-fold validation is used instead of the traditional approach; analysis on the variables correlation and variable importance; models with different types of theory behind, naïve Bayes (probabilistic classifiers), Adaboost(statistical meta-algorithm), SGD(stochastic approximation) and LGB(leaf search algorithm). We adjust the hyperparameters in the model, such as the estimators in adaboost; loss-function and regulator in SGD; maximum leaf in LGB.

Review 2

Response to “relatively less visuals”

The number of visuals in our poster are restricted due to the information of models.

Model section account for nearly half of our poster size, but it is needed and cannot be reduced because all four models in our project was not included in our lecture which means the explanation is needed for providing basic understanding for readers.

Response to “heatmap unreadable”

The low readability of heatmap is mainly due to the huge amount of features in the dataset, which might be both infeasible and unnecessary to enlarge. The main purpose of the heatmap is to provide a general idea to the reader of how the variables related and some of the features have high correlation with each other, red for highly positive related and blue for highly negative.

Response to “quantitative result should be included”

In our report, quantitative analysis and results are mentioned all over the report. For example, Imbalanced dataset (1:9), 130 variables, 300k samples, 5-fold validation (specify the number of folds), top 10 importance features, 100 estimators in Adaboost(specify the number of estimators) and the score of results of each model. It is not ideal to put all numerical results to the poster, especially the hyperparameter configuration you suggested as we have experimented with a variety of settings. If we list out all numbers in the report that is not good practice. Therefore, we partially describe the key idea of the results.

Response to “grammatical error”

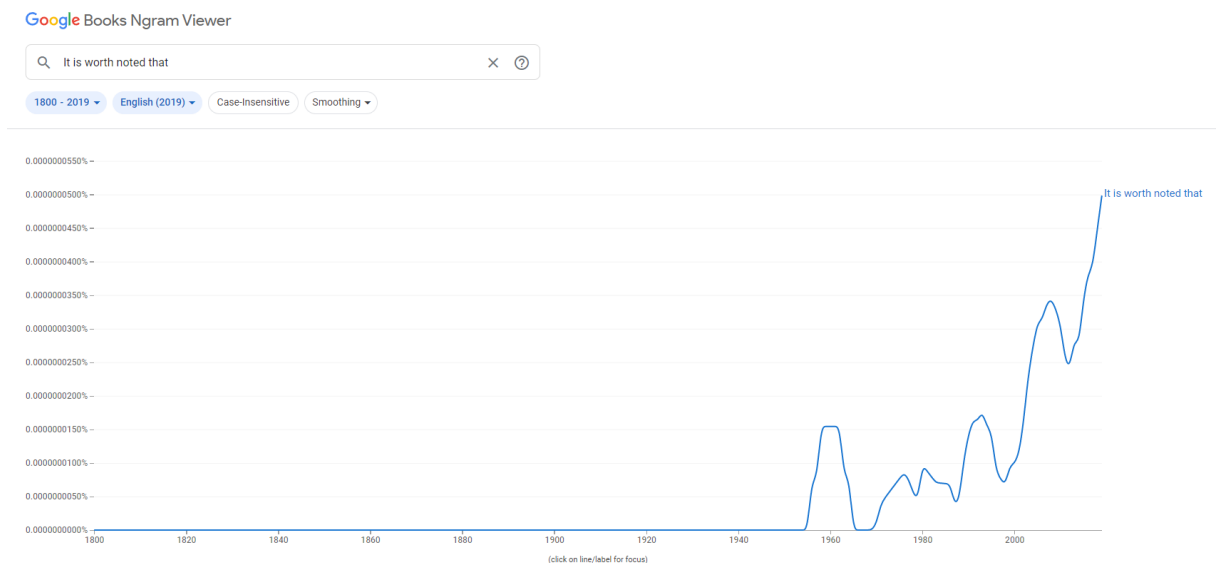
The phrase “It is worth noted that” could be seen in articles, journal and research letter. Here are the references:

Simanjuntak, F.M., Panda, D., Wei, KH. *et al.* Status and Prospects of ZnO-Based Resistive Switching Memory Devices. *Nanoscale Res Lett* **11**, 368 (2016).

Under “**ZnO-Based RRAM as Multifunctional Devices**”

Lin, T., Kuo, H.C., Jiang, X.D. *et al.* Recombination Pathways in Green InGaN/GaN Multiple Quantum Wells. *Nanoscale Res Lett* **12**, 137 (2017). <https://doi.org/10.1186/s11671-017-1922-2>

Under “**Results and Discussion**”



Response to presentation

Except for the advantages mentioned in your review, we presented more details and information for listener better understanding. For instance, we compared Adaboost and LGB model, both models have similar basis but different structure(level-growth and leaf-growth), implementation and performance(Adaboost required 5 times more runtime than LGB for same settings).

Response to creativity

In our project, 4 of our models are not mentioned in our lecture which means we have to investigate on something new. Also, we provided the feature importance with figures instead of relativity, it is not a usual analysis presentation that could be observed among 7 groups. On top of that, we normalized the importance of variables that could provide a decent scale for better understanding.

Review 4

Response to “figures are non-informative”

The dataset we have used is a tabular dataset. Going deep into exploring the data might not be efficient due to large amount of feature. Nevertheless, Heatmap and feature selection is used. The performance of different models is affected by computational time and flexibility of the model. Both figures shown on the poster describe the relationship between predictor variables. Heatmap provides a general idea to the reader how the variables related, red for highly positive related and blue for highly negative.

Response to “no novel idea”, “models are simple”

We have various configurations in each section. For example, 5-fold validation is used instead of the traditional 8:2 split method; analysis on the variables correlation and variable importance are criteria in feature engineering; model with different types of theory behind, Adaboost (classification meta-algorithm) and LGB (leaf search algorithm). We adjust the hyperparameters in the model, for instance, estimators in adaboost; loss-function and regulator in SGD; maximum leaf in LGB.

Review 5

Response to “style is plain”

The plain style of our poster may be due to the models' details.

Model section account for nearly half of our poster, but it is necessary because all four models in our project was not included in our lecture which means the explanation is needed for providing basic information.

Response to creativity

Except SGD was not introduced in the lecture, LGB was not mentioned in the lecture, it is the idea after our group search many of assessment models.