**Introduction**

**an introduction/overview/executive summary section that describes the dataset and summarizes the goal of the project and key steps that were performed**

In this report, our goal is to predict the edibility (class: edible / poisonous) of mushroom basing on attribution information. Data set includes descriptions of hypothetical samples corresponding to 23 species of gilled mushrooms in the Agaricus and Lepiota Family (pp. 500-525). The reason of selecting this dataset is that this problem is related to classification which is a large part application in data science. And, it is also a complement to project – MovieLens that we can cover each part of what we have learnt from the course.

The mushroom dataset has already been well formatted from the source already. Data cleaning is only applied by removing 2 attributions prior to splitting the data to training set and test set. 10 algorithms are applied and an ensemble model combining the prior 10 different algorithms to see if it can provide improvement to our predictions.

1. glm
2. lda
3. Naïve Bayes
4. svmLinear
5. classification
6. knn
7. gamLoess
8. multinom
9. rf
10. adaboost
11. ensemble

[Talk about the results here]

The evaluation of algorithm is based on root mean squared error (RMSE) of predicted rating against actual rating. Algorithm is trained on train set and being test on test set. Final RMSE is presented basing the on the final hold-out validation set with result in the tier of “RMSE < 0.86490”.

**Method**

**a methods/analysis section that explains the process and techniques used, including data cleaning, data exploration and visualization, insights gained, and your modeling approach**

**1. Data Cleaning**

edx data set contains 6 columns (userId, movieId, rating, timestamp, title and genres).

**2. Data Exploration**

**3. Modeling Approach**

**Result**

**a results section that presents the modeling results and discusses the model performance**

**Conclusion**

**a conclusion section that gives a brief summary of the report, its limitations and future work**