Amazon Music Review Ratings Prediction

Group 19 Course Project Presentation
teamID: 6565895

Yilin Chen 1002916017
Angela HSU 1003328874
Ling Feng Zhou 1002739012
Ziqi Chen 1002576722
Yao Tong 1007410295
Xinyi Yuan 1002376519
Mina Youssef Rizkalla 1000429690

Exploratory analysis

- The purpose of this assignment is to predict the music ratings based on the user reviews.
- Dataset: Amazon Music Review.
- Preprocess:
  - Drop NaN
  - Convert ascii string into html scrip
  - Remove URLs and stopwords
- Use chi square test to determine the important features:
  - reviewText (figure 1)
  - summary

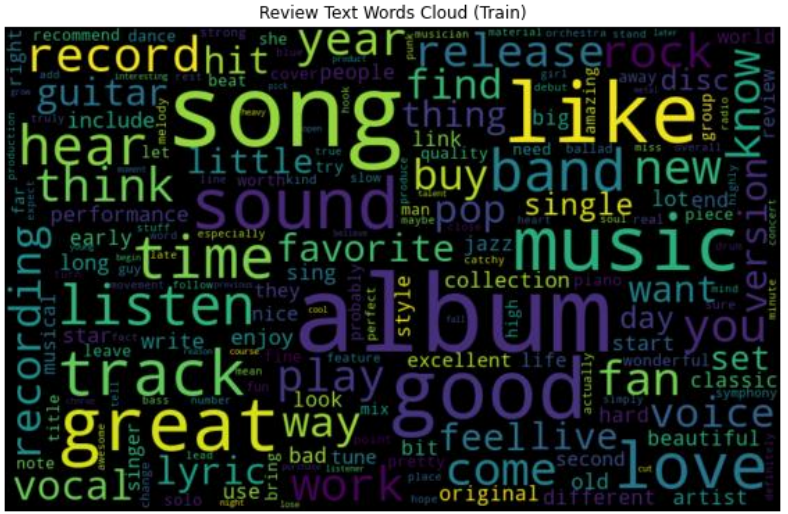

Figure 1. World cloud of reviewText

Exploratory analysis

- The purpose of this assignment is to predict the music ratings based on the user reviews.
- Dataset: Amazon Music Review.
- Preprocess:
  - Drop NaN
  - Convert ascii string into html scrip
  - Remove URLs
- Use chi square test to determine the important features:
  - **reviewText (figure 1)**
  - summary

Figure 1. World cloud of reviewText

Exploratory analysis

- The training distribution of the ratings is skewed to the right (figure 2).
- Higher ratings (5 stars or 4 stars) contribute a large proportion, which corresponds to a relatively high mean rating of 4.41 (table 1).

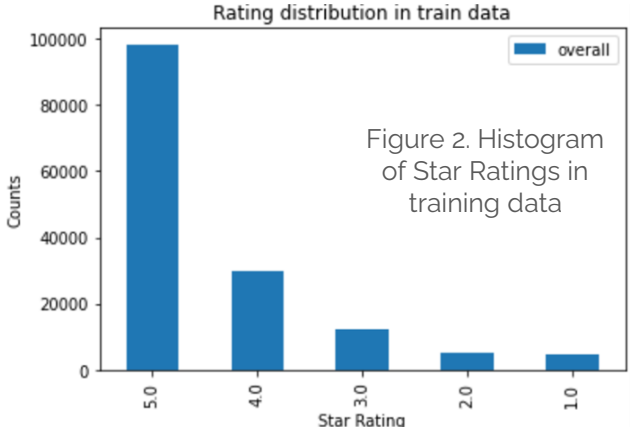

Figure 2. Histogram of Star Ratings in training data

Number of reviews	149972
Number of unique users	62998
Number of unique items	24590
Mean rating	4.408670
% 5 star rating	0.6547
% 4 star rating	0.1977
% 3 star rating	0.0810
% 2 star rating	0.0347
% 1 star rating	0.0318

Table 1. Data statistics

Model Selection

- Used Random Forest, Linear Regression and Neural Networks on the preprocessed training data.
- The Neural Networks model had the best performance with a MSE score of 0.45, which passed the strong baseline (table 1).

Model used	MSE score	Baseline
Random Forest	0.84	Naive
Linear Regression	0.66	Weak
Neural Networks	0.45	Strong

Table 2. MSE for Models

Model Tuning (NN)

- Learning rate:
  - It determines the step size that the model moves towards minimum of a loss function.
  - The NN model works the best when the learning rate is 0.005 (figure 3).

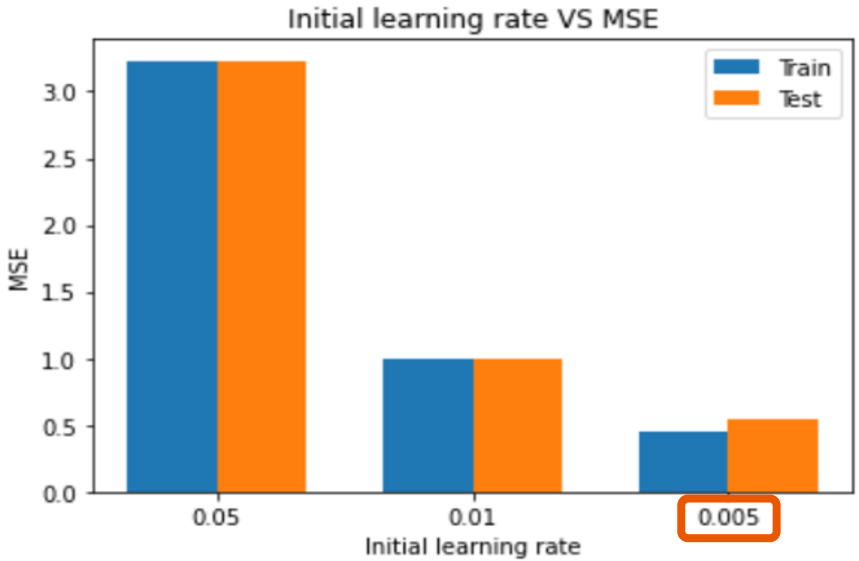

Figure 3. Initial Learning Rate vs. MSE

Model Tuning (NN)

- Number of layers:
  - It shows the complexity of the model and the model becomes more complex as the number of layers increases.
  - The NN model works the best when the number of layers is 128 (figure 4).

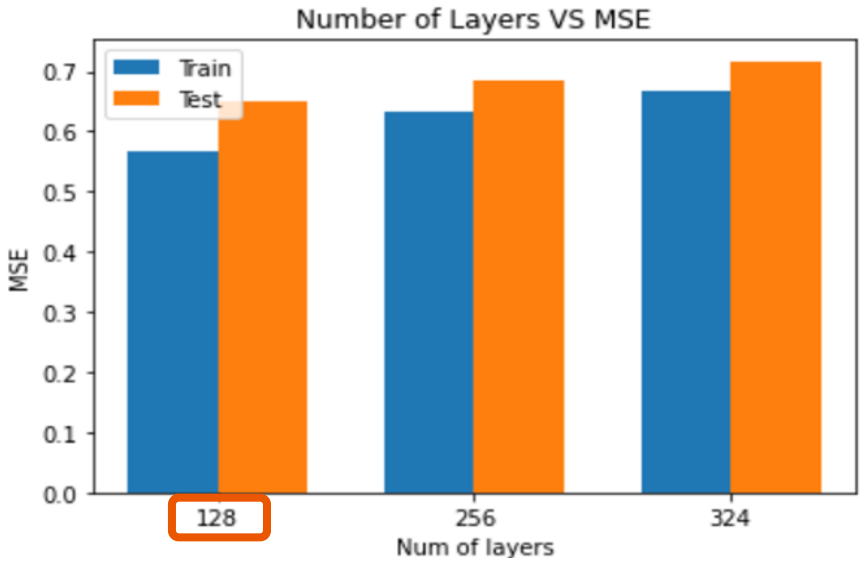

Figure 4. Number of Layers vs. MSE

Model Tuning (NN)

- Batch size:
  - It is the number of samples in one forward pass of the neural networks.
  - The optimal batch size is 128 for the NN model (figure 5).

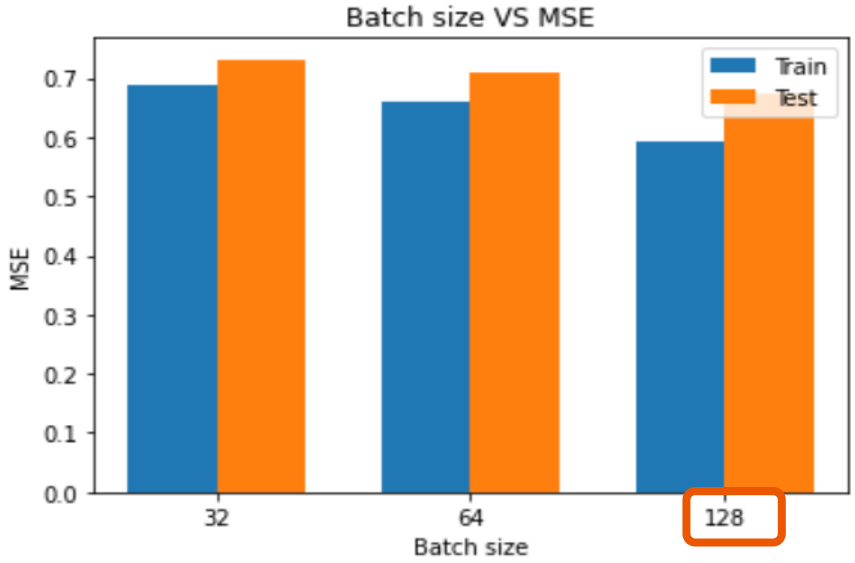

Figure 5. Batch Size vs. MSE

Result

- Compare the MSE of different features
  - Word Frequency
  - TF-IDF
  - Bigram
- TF-IDF performs the best (figure 6)

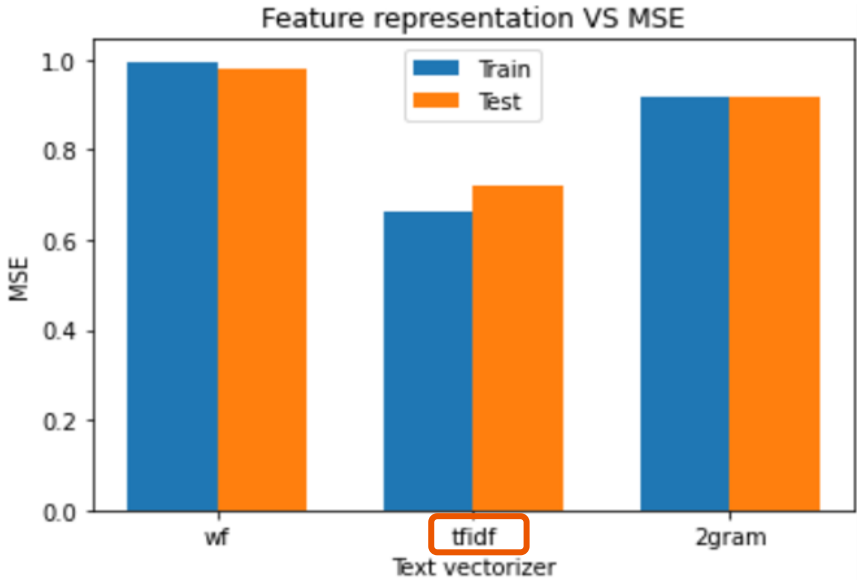

Figure 6. Feature representation vs. MSE

Discussion

- Improvement on Data
  - Use word embedding as feature representation
  - Do sentiment analysis on preprocessed data and return polarity/subjectivity
- Improvement on Model
  - Tune more parameters like momentum or activation function
  - Use ordinal classification neural networks to replace neural networks

Thank you!

Q&A