|  |  |
| --- | --- |
|  | The final report |
| mod1.manualgenre.tree.jmp | The recommended tree model |
|  | The recommended neural network model |
| wrangling.txt | Windows PowerShell commands and R codes I used to process the original data files. |
| tts.txt | The regular expression for extracting movie records from datasets. |

Other Models Folder:

The models that weren’t selected as recommendation

|  |  |
| --- | --- |
| mod1.tree.jmp | A tree model that used all the genres and length of movies |
| mod1.bc.log.jmp | A logistic model that used all the genres, length of movies, number of directors, and number of writers |
|  |  |
|  |  |

Data Folder:

The processed data and one original dataset

|  |  |
| --- | --- |
| 97to17.basics.jmp | A processed " title.basics.tsv.gz" file joined with " title.ratings.tsv.gz". |
| names.zip | A processed "name.basics.tsv.gz" file. It contains people who is supposed to be alive and below 79 years old. |
| crew.zip | A processed "title.crew.tsv.gz" file. It contains the directors and writers of each movie |
| \archives\tt.basics.zip | All the movie records from the "title.basics.tsv.gz" file. |
| \archives\title.ratings.tsv.gz | The rating data from IMDb as of March 28 |