|  |
| --- |
| KF5012 |
| SOLUTION TESTING DOCUMENTATION |
| Group: Only fools and horses |

|  |
| --- |
| Michael Anderson  W18032122 |

Testing

Master Test Plan

The purpose of these tests is to identify any problems with our system and model and to help in the development of an accurate and robust model by creating recommendations that address any found weaknesses.

The model will be tested at each iteration to guide development of the next iteration. Tests will be created by looking at the dataset used to train that version of the model to expose weaknesses in the model’s predictions.

Genre prediction testing

The model’s purpose is to take a movie plot or summary and make a prediction of its genre, this test aims to test this capability of making an accurate prediction.

A number of genres will be selected from the dataset used to train the selected model, the movie plots for these genres will be sanitised, which involves removing stop words, word contractions, any nonalphabetical characters and making all words lower case. This sanitation is done to create normalised and uniform version of the movie plots with an accurate representation of the words within the plot, for example “can’t” and “can not” have the same contextual meaning but to the model these would be handled as different and possibly lead to inaccurate predictions, removing all contractions between words prevents this. similarly making all words lowercase ensures that the model only sees 1 form of a word ensuring consistency amongst predictions.

Stop words and nonalphabetic characters are removed from the movie plots with the aim to leave only the words that are significantly relevant to each genre.

After the movie plots are sanitised a function will be used to list the most frequent words in each genre and the 3 most frequent in each genre will be chosen and used as a string to test the model. Predicting against a string of the 3 most frequent words in a genre should give the model the best possibility of correctly predicting that genre and any incorrect predictions could signify an issue with the model or data used to train the model.

If a top 3 word is frequent amongst multiple genres it will not be used in the test as to give the model the best chance to make an accurate prediction. Where a word is common amongst another genre, another word will be selected for this test as to avoid confusing the model.

Genre Confusion Test

The aim of this test is to try and confuse the model, the method is like the Genre prediction test but instead of taking the 3 most popular words for 1 genre, 3 frequent words will be combined from all the selected genres.

The model will then make predictions on these, it is expected that the model could predict any one of the chosen genres that these words were taken from but if the model predicts any other genre it could signify a bias towards a genre or an imbalance in the training data.

Empty string and Whitespace test

The model will make predictions on a string of whitespace and on an empty string. The results of this test could help to identify any biases the model has towards genres as the white space and empty string should provide no significance towards any genre unless the data for that genre was to contain a disproportionately high amount of white space, which could be corrected and balanced through prepossessing before training the next iteration of the model.

Number and Symbol test

The model will make predictions on a list of numbers and a string of symbols, again these should hold very little significance to any genre so the predictions against these may show biases in the model or issues with the dataset used to train the model such as the presence of these characters in the movie plots which could be removed.

Stop words

The model will make a prediction on a collection of stop words, these words hold very little significance for any genre and the predictions made on this test could help to identify issues with the training data that may need to be corrected to help improve the model.

Unknown genre test

Within the datasets used to train the models there are a lot of entries that have an unknown genre, these are valid complete movie plots that are uncategorised. This test aims to use a number of these movie plots for the model to make predictions on, because the genre is unknown the accuracy of the models predictions cannot be measured but by looking at a sequence of predictions made on these movie plots it is possible to identify potential issues or trends in the models predictions. If the model is consistently predicting a small number of genres on most of the unknown movie plots, then it could signify a bias in the model towards those genres which would need to be explored further.

Baseline implementation test

Baseline implementation Test Plan

The aim of this test is to test the model created in the baseline implementation and to identify any weaknesses or issues with the model or the dataset used to train the model in order to help in the development of the next iteration and to ensure a more robust solution. The data for these tests will be taken from the dataset used to train the baseline implementation.

Below is a summary of each test.

Genre prediction testing

Three genres have been selected from the dataset, these are ‘western’, ‘drama’ and ‘romance’. The movie plots for each of these genres has been sanitised and the most frequent words have been listed for each. A string has been created for each genre that consists of its top 3 frequent words, these are:

Western: “town men man”

Drama: “mother home tells”

Romance: “love father family”

These were the most frequent words for each genre that were not also frequent in other genres.

Genre confusion testing

From the most frequent words for the 3 selected genres, a word was selected from each and combined into a string. These strings are:

Confusion test1: “town father home”

Confusion test2: “men love mother”

Confusion test3: “man family tells”

Empty string and whitespace testing

Whitespace: “ “

Empty string: “”

Whitespace, empty string, symbol and number tests

The model will make predictions against various strings of nonalphabetic characters and symbols, the chosen data for each of these tests is as follows:

Whitespace: “ “

Empty string: “”

Symbols: “%&^$\*?@:>”

Numbers: '1','2','3','4','5'

Stop word test

Two strings of stop words will be used to test the model, these two strings are:

Stop word test 1: “this is the”

Stop word test 2: “that there are”

Unknown tests

Three movie plots have been selected from the data with a genre of ‘unknow’ for the model to make predictions on, after sanitisation these three movie plots are:

Unknown plot 1: “bartender working saloon serving drinks customers fills stereotypically irish man bucket beer carrie nation followers burst inside assault irish man pulling hat eyes dumping beer head group begin wrecking bar smashing fixtures mirrors breaking cash register bartender sprays seltzer water nation face group policemen appear order everybody”

Unknown plot 2: “lasting seconds consisting shots shot set wood winter actor representing vice president theodore roosevelt enthusiastically hurries hillside tree foreground falls rights cocks rifle men bearing signs reading photographer press agent respectively follow shot photographer sets camera teddy aims rifle upward tree fells appears common house cat proceeds stab teddy holds prize aloft press agent takes notes second shot taken slightly different wood path teddy rides path horse camera left shot followed closely press agent photographer dutifully holding signs”

Unknown plot 3: “story concerns soldier deserts regiment encounters wagon train settlers finding attack american indians eminent returns unit order elicit help”

Baseline implementation Test Cases

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Test Case | Test Case Description | Test Data | Predicted output | Actual output | Notes |
| Genre prediction test 1 | A string of the top 3 most common words for the movie genre western | "town men man" | ‘western’ | ‘drama’ | Model incorrectly predicted drama over western |
| Genre prediction test 2 | A string of the top 3 most common words for the movie genre romance | "love father family" | ‘romance’ | ‘unknown’ | Model incorrectly predicted unknown instead of romance |
| Genre prediction test 3 | A string of the top 3 most common words for the movie genre drama | "mother home tells" | ‘drama’ | ‘drama’ | Model correctly predicted drama |
| Genre confusion test 1 | A string consisting of a combination of the top words from each of the selected genres, western, drama and romance | “town father home” | “western”, “drama” or “romance” | ‘drama’ | Model predicted drama. |
| Genre confusion test 2 | A string consisting of a combination of the top words from each of the selected genres, western, drama and romance | "men love mother" | “western”, “drama” or “romance” | ‘unknown’ | Model incorrectly predicted unknown. |
| Genre confusion test 3 | A string consisting of a combination of the top words from each of the selected genres, western, drama and romance | "man family tells"- men love mother | “western”, “drama” or “romance” | ‘drama’ | The model predicted drama again, this could indicate a bias towards drama |
| White space test | Model will be tested with only white space | ‘ ‘ | ‘unknown’ | ‘unknown’ | Model correctly predicted unknown |
| Empty string test | Model will be tested with an empty string | ‘’ | ‘unknown’ | ‘unknown’ | Model correctly predicted unknown |
| Stop words test 1 | Model will be tested with a string of stop words | ‘this is the’ | ‘unknown’ | ‘unknown’ | Model correctly predicted unknown |
| Stop words test 2 | Model will be tested with a string of stop words | ‘that there are’ | ‘unknown’ | ‘unknown’ |  |
| Number test | Model will be tested with a list of numbers | ['1','2','3','4','5'] | ‘unknown’  For all | ‘unknown’ for all | Model correctly predicted unknown for all numbers |
| Symbol test | Model will be tested with a string consisting of symbols | '%&^$\*?@:>' | ‘unknown’ | ‘unknown’ | Model correctly predicted unknown |
| Unknown genre prediction test 1 | Model will be tested using movie plots that have an unknown genre | “bartender working saloon serving drinks customers fills stereotypically irish man bucket beer carrie nation followers burst inside assault irish man pulling hat eyes dumping beer head group begin wrecking bar smashing fixtures mirrors breaking cash register bartender sprays seltzer water nation face group policemen appear order everybody” | Any movie genre except ‘unknown’ | ‘unknown’ | Model incorrectly predicted unknown |
| Unknown genre prediction test 2 | Model will be tested using movie plots that have an unknown genre | “lasting seconds consisting shots shot set wood winter actor representing vice president theodore roosevelt enthusiastically hurries hillside tree foreground falls rights cocks rifle men bearing signs reading photographer press agent respectively follow shot photographer sets camera teddy aims rifle upward tree fells appears common house cat proceeds stab teddy holds prize aloft press agent takes notes second shot taken slightly different wood path teddy rides path horse camera left shot followed closely press agent photographer dutifully holding signs” | ‘unknown’ | ‘drama' | Model predicted drama |
| Unknown genre prediction test 3 | Model will be tested using movie plots that have an unknown genre | “story concerns soldier deserts regiment encounters wagon train settlers finding attack american indians eminent returns unit order elicit help” | ‘unknown’ | ‘drama’ | Model predicted drama again, also hinting at a bias towards drama |

Baseline implementation Recommendations

Recommendation 1

The dataset has 2265 unique genre labels many being multi-labels where more than one genre is given to a movie plot. The format of these multi-label genres is not standardised as some use a comma to separate the genres, some use a hyphen and others use a back slash making the model recognise 2 identical genres with different separators as different labels for example ‘romantic-comedy’ and ‘romantic comedy’ both appear in the dataset but are treat as different labels. Since the model currently uses single label prediction it is recommended that these multiple genre labels be replaced with single labels, possibly using the first occurring genre as the new label, as this would be the genre the movie plot is most strongly associated, taking care to remove any separating characters. This would help to balance out the data by removing many of the extra genres and improve the model’s ability to make accurate predictions.

Recommendation 2

The model predicted ‘unknown’ instead of ‘romance’ during the genre prediction test and again during the genre confusion test but correctly predicted ‘unknown’ for the empty string, whitespace, symbol, number and stop word tests. To prevent the model predicting ‘unknown’ for genuine movie plots it is recommended that movie plots with the unknown genre be removed and replaced with entries with invalid movie plots such as collections of white space, symbols and numbers to preserve the models ability to recognise this invalid data.

Note:

The model may be overfitting on drama as this was the main genre that was predicted but it is difficult to tell at this time.

Version 2

Version 2 test plan

The aim of these tests is to test the second version of our model and to identify any weaknesses or issues that it may have and to develop recommendations that will improve the model and guide the development of the next iteration to ensure an accurate and robust model is created.

Since the baseline implementation the dataset used to train the model has been changed, the labels have been separated leaving only single genre labels for each movie plot. Data from this new version of the dataset will be used for testing.

Genre prediction testing

Four genres have been selected from the dataset, these are ‘western’, ‘drama’ and ‘romance’ and ‘family’. After sanitisation, strings consisting of the most frequent words for each are:

Western: “town men man”

Drama: “father tells family”

Romance: “love life day”

Family: “mother dog house”

Genre confusion testing

From the most frequent words for the selected genres, a word was selected from each and combined into a string. These strings are:

Confusion test 1: “town father love”

Confusion test 2: “tells men life”

Confusion test 3: “man family tells”

Confusion test 4: “mother men town”

Empty string and whitespace testing

Whitespace: “ “

Empty string: “”

Whitespace, empty string, symbol, and number tests

The model will make predictions against various strings of nonalphabetic characters and symbols, the chosen data for each of these tests is as follows:

Whitespace: “ “

Empty string: “”

Symbols: “%&^$\*?@:>”

Numbers: '1','2','3','4','5'

Stop word test

Two strings of stop words will be used to test the model, these two strings are:

Stop word test 1: “this is the”

Stop word test 2: “that there are”

Unknown tests

Three movie plots have been selected from the data with a genre of ‘unknow’ for the model to make predictions on, after sanitisation these three movie plots are:

Unknown plot 1: “bartender working saloon serving drinks customers fills stereotypically irish man bucket beer carrie nation followers burst inside assault irish man pulling hat eyes dumping beer head group begin wrecking bar smashing fixtures mirrors breaking cash register bartender sprays seltzer water nation face group policemen appear order everybody”

Unknown plot 2: “lasting seconds consisting shots shot set wood winter actor representing vice president theodore roosevelt enthusiastically hurries hillside tree foreground falls rights cocks rifle men bearing signs reading photographer press agent respectively follow shot photographer sets camera teddy aims rifle upward tree fells appears common house cat proceeds stab teddy holds prize aloft press agent takes notes second shot taken slightly different wood path teddy rides path horse camera left shot followed closely press agent photographer dutifully holding signs”

Unknown plot 3: “story concerns soldier deserts regiment encounters wagon train settlers finding attack american indians eminent returns unit order elicit help”

Version 2 Test Cases

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Test Case | Test Case Description | Test Data | Predicted output | Actual output | Notes |
| Genre prediction test 1 | A string of the top 3 most common words for the movie genre drama | “father tells family” | ‘drama’ | ‘drama’ | Model correctly predicted drama |
| Genre prediction test 2 | A string of the top 3 most common words for the movie genre western | "town men man" | ‘western’ | ‘drama’ | Model incorrectly predicted drama |
| Genre prediction test 3 | A string of the top 3 most common words for the movie genre romance | "love life day" | ‘romance’ | ‘drama’ | Model incorrectly predicted drama again |
| Genre prediction test 4 | A string of the top 3 most common words for the movie genre family | "mother dog house" | ‘family’ | ‘drama’ | Model incorrectly predicted drama  again |
| Genre confusion test 1 | A string consisting of a combination of the top words from each of the selected genres | "town father love " | ‘drama’,  ’western’,  ’romance’  Or  ‘family’ | ‘drama’, | Model predicted drama |
| Genre confusion test 2 | A string consisting of a combination of the top words from each of the selected genres | "tells men life " | ‘drama’,  ’western’,  ’romance’  Or  ‘family’ | ‘drama’, | Model predicted drama |
| Genre confusion test 3 | A string consisting of a combination of the top words from each of the selected genres | "man family tells" | ‘drama’,  ’western’,  ’romance’  Or  ‘family’ | ‘drama’, | Model predicted drama |
| Genre confusion test 4 | A string consisting of a combination of the top words from each of the selected genres | "mother men town" | ‘drama’,  ’western’,  ’romance’  Or  ‘family’ | ‘drama’ | Model predicted drama, model  Is biased towards drama |
| White space test | Model will be tested with only white space | ‘ ‘ | ‘unknown’ | ‘drama’ | Model predicted drama |
| Empty string test | Model will be tested with an empty string | ‘’ | ‘unknown’ | ‘drama’ | Model predicted drama |
| Stop words test 1 | Model will be tested with a string of stop words | ‘this is the’ | ‘unknown’ | ‘drama’ | Model predicted drama |
| Stop words test 2 | Model will be tested with a string of stop words | ‘that there are’ | ‘unknown’ | ‘drama’ | Model predicted drama |
| Number test | Model will be tested with a list of numbers | ['1','2','3','4','5'] | ‘unknown’  For each number in list | ‘drama’  For each number in list | Model predicted drama for all numbers |
| Symbol test | Model will be tested with a string consisting of symbols | '%&^$\*?@:>' | ‘unknown’ | ‘drama’ | Model predicted drama |
| Unknown genre prediction test 1 | Model will be tested using movie plots that have an unknown genre | “bartender working saloon serving drinks customers fills stereotypically irish man bucket beer carrie nation followers burst inside assault irish man pulling hat eyes dumping beer head group begin wrecking bar smashing fixtures mirrors breaking cash register bartender sprays seltzer water nation face group policemen appear order everybody” | Any movie genre except ‘unknown’ | ‘drama’ | Model predicted drama |
| Unknown genre prediction test 2 | Model will be tested using movie plots that have an unknown genre | “lasting seconds consisting shots shot set wood winter actor representing vice president theodore roosevelt enthusiastically hurries hillside tree foreground falls rights cocks rifle men bearing signs reading photographer press agent respectively follow shot photographer sets camera teddy aims rifle upward tree fells appears common house cat proceeds stab teddy holds prize aloft press agent takes notes second shot taken slightly different wood path teddy rides path horse camera left shot followed closely press agent photographer dutifully holding signs” | ‘unknown’ | ‘drama' | Model predicted drama |
| Unknown genre prediction test 3 | Model will be tested using movie plots that have an unknown genre | “story concerns soldier deserts regiment encounters wagon train settlers finding attack american indians eminent returns unit order elicit help” | ‘unknown’ | ‘drama’ | Model predicted drama  again |

Version2 Recommendations

Recommendation 1

The model is heavily overfitting on the ‘drama’ genre. This may possibly be due to the amount of movie plots in each genre being unbalanced for example the top 5 genres with the most movie plots are ‘drama’ with 9397, comedy with 6468, unknown with 6063, action with 1355 and horror with 1209, this clearly shows how massively unbalanced the data is and gives us a clear reason as to why the model is overfitting on drama.

For this reason, it is recommended that the dataset be balanced with a similar number of movie plots for each genre to prevent overfitting. It is also recommended that the ‘unknown’ genre be removed to prevent the model incorrectly predicting that label, since it is not a valid movie genre.

Recommendation 2

The multi-genre labels have been separated into single labels but now many miscellaneous genres are present in the data for example some of the genres present are ‘of’, ‘in’, ‘&’, ‘fo’, ‘co’ and ‘\’. These labels will negatively affect the model’s ability to make accurate predictions, so it is recommended that all the movies with miscellaneous genres be removed from the dataset before retraining the new model.

Version 3

Version 3 Test plan

The aim of these tests is to test the third version of our model and to identify any weaknesses or issues that it may have and to develop recommendations to further improve the model.

Many of the recommendations from the previous test have been applied to the dataset used to train this model. The dataset has been balanced, now all movie genres have between 500 and 1000 movie entries. Genres with less than 500 movies were removed or had their numbers increased using data from other datasets and all movies with an unknown or miscellaneous genre have been removed from the dataset.

This new dataset has been used to test this model.

Genre prediction testing

Four genres have been selected from the dataset, these are ‘western’, ‘drama’ and ‘romance’ and ‘adventure. The ‘family’ genre from the previous tests has been removed from the dataset as it had less than 500 movies and no other data could be found to increase its numbers.

The frequent word strings for these genres are:

Western: “town men man”

Drama: “life man new”

Romance: “sebastian luisa friedrich”

adventure: “new world help”

Genre confusion testing

From the most frequent words for the selected genres, a word was selected from each and combined into a string. These strings are:

Confusion test 1: “sebastian men life”

Confusion test 2: “town luisa help”

Confusion test 3: “man friedrich world”

Confusion test 4: “time ranch man”

Whitespace, empty string, symbol, and number tests

The test data for these tests is the same as previous tests.

Stop word test

The stop words tests are the same as in previous tests.

Unknown tests

The three unknown movie plots used in this test are the same as in previous tests.

Version 3 Test Cases

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Test Case | Test Case Description | Test Data | Predicted output | Actual output | Notes |
| Genre prediction test 1 | A string of the top 3 most common words for the movie genre drama | "life man new" | ‘drama’ | ‘drama’ | Model  correctly predicted drama |
| Genre prediction test 2 | A string of the top 3 most common words for the movie genre western | "town men ranch" | ‘western’ | ‘western’ | Model  correctly predicted western |
| Genre prediction test 3 | A string of the top 3 most common words for the movie genre romance | "sebastian luisa friedrich" | ‘romance’ | ‘comedy’ | Model  incorrectly predicted comedy |
| Genre prediction test 4 | A string of the top 3 most common words for the movie genre adventure | "new world help" | ‘adventure’ | ‘adventure’ | Model  correctly predicted adventure |
| Genre confusion test 1 | A string consisting of a combination of the top words from each of the selected genres | "town father love " | ‘drama’,  ’western’,  ’romance’  Or  ‘adventure’ | ‘western’ | Model  predicted western |
| Genre confusion test 2 | A string consisting of a combination of the top words from each of the selected genres | "tells men life " | ‘drama’,  ’western’,  ’romance’  Or  ‘adventure’ | ‘western’ | Model  predicted western |
| Genre confusion test 3 | A string consisting of a combination of the top words from each of the selected genres | "man family tells" | ‘drama’,  ’western’,  ’romance’  Or  ‘adventure’ | ‘drama’, | Model  predicted western |
| Genre confusion test 4 | A string consisting of a combination of the top words from each of the selected genres | "mother men town" | ‘drama’,  ’western’,  ’romance’  Or  ‘adventure’ | ‘western’ | Model  predicted western |
| White space test | Model will be tested with only white space | ‘ ‘ | No valid genre | ‘comedy’ | Model  predicted comedy |
| Empty string test | Model will be tested with an empty string | ‘’ | No valid genre | ‘comedy’ | Model  predicted comedy |
| Stop words test 1 | Model will be tested with a string of stop words | ‘this is the’ | No valid genre | ‘comedy’ | Model  predicted comedy |
| Stop words test 2 | Model will be tested with a string of stop words | ‘that there are’ | No valid genre | ‘comedy’ | Model  predicted comedy |
| Number test | Model will be tested with a list of numbers | ['1','2','3','4','5'] | No valid genre | ‘comedy’  For each number in list | Model  predicted comedy |
| Symbol test | Model will be tested with a string consisting of symbols | '%&^$\*?@:>' | No valid genre | ‘comedy’ | Model  predicted comedy, may be overfitting |
| Unknown genre prediction test 1 | Model will be tested using movie plots that have an unknown genre | “bartender working saloon serving drinks customers fills stereotypically irish man bucket beer carrie nation followers burst inside assault irish man pulling hat eyes dumping beer head group begin wrecking bar smashing fixtures mirrors breaking cash register bartender sprays seltzer water nation face group policemen appear order everybody” | Any movie genre except ‘unknown’ | ‘comedy’ | Model  predicted comedy |
| Unknown genre prediction test 2 | Model will be tested using movie plots that have an unknown genre | “lasting seconds consisting shots shot set wood winter actor representing vice president theodore roosevelt enthusiastically hurries hillside tree foreground falls rights cocks rifle men bearing signs reading photographer press agent respectively follow shot photographer sets camera teddy aims rifle upward tree fells appears common house cat proceeds stab teddy holds prize aloft press agent takes notes second shot taken slightly different wood path teddy rides path horse camera left shot followed closely press agent photographer dutifully holding signs” | Any movie genre except ‘unknown’ | ‘comedy’ | Model  predicted comedy  again |
| Unknown genre prediction test 3 | Model will be tested using movie plots that have an unknown genre | “story concerns soldier deserts regiment encounters wagon train settlers finding attack american indians eminent returns unit order elicit help” | Any movie genre except ‘unknown’ | ‘comedy’ | Model  predicted comedy again |

Version2 Recommendations

The model falsely made predictions on the none-alphabetic data tests such as the white space and symbols. To prevent this and restore the models ability to detect invalid movie plots it may be worth reintroducing the ‘unknown’ genre label with the move plots for that genre be a combination of white space, empty strings, numbers and special characters.

Recommendation 1

The ‘unknown’ genre was removed from the dataset to help with the models accuracy when making predictions but since this genre was removed the model was forced to predict another genre for the whitespace, empty string, number, stop words and symbol tests. To restore the model’s ability to detect invalid movie plots it is recommended that the ‘unknown’ genre be replaced but with movie plots consisting of symbols, white space, stop words and numbers.

Recommendation 2

The model repeatedly predicted ‘comedy’ for the unknown genre tests and for the various none-alphabetic tests, this may indicate the model is overfitting on the ‘comedy’ genre. A possible caused of this may be the length of the comedy movie plots, if these movie plots are longer than other genres and have more words then the model will associate

Since the number of movie plots for each genre is balanced the cause may be the length of the movie plots in the comedy genre. If the movie plots are longer and contain more words|| then the model will associate more words to the comedy genre, the more words the movie plot has the more white space it will contain a spaces between words, this may explain why the

If the movie plots are longer and contain more words, then this would explain why the model is overpredicting on comedy because more words are associated with that genre. This may also explain why the model predicted comedy on the white space test as a longer movie plot will contain more white space.

Version 4

Version 4 test plan

The aim of these tests is to test the fourth version of our model to identify any weaknesses or issues that it may have and to create recommendations to further improve the model.

No changes to the dataset have been made since the last version, all changes have been made to the model.

To test the new changes to the model the same tests and test data has been used from the version 3 tests.

Version 4 Test Cases

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Test Case | Test Case Description | Test Data | Predicted output | Actual output | Notes |
| Genre prediction test 1 | A string of the top 3 most common words for the movie genre drama | "life man new" | ‘drama’ | ‘drama’ | Model correctly predicted drama |
| Genre prediction test 2 | A string of the top 3 most common words for the movie genre western | "town men ranch" | ‘western’ | ‘western’ | Model correctly predicted western |
| Genre prediction test 3 | A string of the top 3 most common words for the movie genre romance | "sebastian luisa friedrich" | ‘romance’ | ‘romance’ | Model correctly predicted romance |
| Genre prediction test 4 | A string of the top 3 most common words for the movie genre adventure | "new world help" | ‘adventure’ | ‘SciFi’ | Model incorrectly predicted SciFi instead of adventure |
| Genre confusion test 1 | A string consisting of a combination of the top words from each of the selected genres | "town father love " | ‘drama’,  ’western’,  ’romance’  Or  ‘adventure’ | ‘western’ | Model predicted western |
| Genre confusion test 2 | A string consisting of a combination of the top words from each of the selected genres | "tells men life " | ‘drama’,  ’western’,  ’romance’  Or  ‘adventure’ | ‘drama’ | Model predicted drama |
| Genre confusion test 3 | A string consisting of a combination of the top words from each of the selected genres | "man family tells" | ‘drama’,  ’western’,  ’romance’  Or  ‘adventure’ | ‘drama’, | Model predicted drama |
| Genre confusion test 4 | A string consisting of a combination of the top words from each of the selected genres | "mother men town" | ‘drama’,  ’western’,  ’romance’  Or  ‘adventure’ | ‘western’ | Model predicted western |
| White space test | Model will be tested with only white space | ‘ ‘ | No prediction to be made | ‘comedy’ | Model predicted comedy |
| Empty string test | Model will be tested with an empty string | ‘’ | No prediction to be made | ‘comedy’ | Model predicted comedy  again |
| Stop words test 1 | Model will be tested with a string of stop words | ‘this is the’ | No prediction to be made | ‘adventure’ | Model predicted adventure |
| Stop words test 2 | Model will be tested with a string of stop words | ‘that there are’ | ‘No prediction to be made’ | ‘adventure’ | Model predicted adventure test without sanitising |
| Number test | Model will be tested with a list of numbers | ['1','2','3','4','5'] | ‘unknown’  For each number in list | ‘comedy’  For each number in list | Model predicted comedy |
| Symbol test | Model will be tested with a string consisting of symbols | '%&^$\*?@:>' | ‘unknown’ | ‘comedy’ | Model predicted comedy |
| Unknown genre prediction test 1 | Model will be tested using movie plots that have an unknown genre | “bartender working saloon serving drinks customers fills stereotypically irish man bucket beer carrie nation followers burst inside assault irish man pulling hat eyes dumping beer head group begin wrecking bar smashing fixtures mirrors breaking cash register bartender sprays seltzer water nation face group policemen appear order everybody” | Any movie genre except ‘unknown’ | ‘comedy’ | Model predicted comedy |
| Unknown genre prediction test 2 | Model will be tested using movie plots that have an unknown genre | “lasting seconds consisting shots shot set wood winter actor representing vice president theodore roosevelt enthusiastically hurries hillside tree foreground falls rights cocks rifle men bearing signs reading photographer press agent respectively follow shot photographer sets camera teddy aims rifle upward tree fells appears common house cat proceeds stab teddy holds prize aloft press agent takes notes second shot taken slightly different wood path teddy rides path horse camera left shot followed closely press agent photographer dutifully holding signs” | Any movie genre except ‘unknown’ | ‘animation | Model predicted animation |
| Unknown genre prediction test 3 | Model will be tested using movie plots that have an unknown genre | “story concerns soldier deserts regiment encounters wagon train settlers finding attack american indians eminent returns unit order elicit help” | Any movie genre except ‘unknown’ | ‘history | Model predicted history |

Version4 Recommendations

Recommendation

The model makes predictions on none valid movie plots such as stop words or symbols, in order to prevent this the API or GUI may have some form of data validation in place such as possibly filtering out stop words and symbols before they are passed to the model.