

BOW

Digital Humanities: Hands on Data
LHU002P05.2020-2021.1B

Gaurika Chaturvedi
S3803252

Word Count: 885

November 23, 2020

1 Introduction

As Ted Underwood says, (2014) digital methods such as distant reading allow for fields that were originally dialectical to also become discovery-based. Various endeavours are being made in literary studies to train classifiers to understand what linguistic patterns (Ashok, Feng, & Choi, 2013), readability levels (Ashok et al., 2013), and cultural capital (Piper & Portelance, 2016) make a book successful. Efforts are also being made to understand how standards of success change over time (Underwood & Sellers, 2016). However, I want to focus on whether these standards are universal across genres. Thus, my research question is, “What makes the Science-Fiction genre difficult to classify by a classifier trained on diverse genres?” I chose the science-fiction genre because. after a preliminary round of analysis, I found it to be the most difficult genre to classify.

2 Methodology

The corpus used for this analysis was retrieved from the Gutenberg Project (<https://www.gutenberg.org/>, n.d.) along with relevant metadata including genre

classification and download count. The corpus consisted of 801 texts from diverse genres, including science-fiction, historical books, poetry, love, mystery, adventure, and fiction. In determining whether or not a text was successful, Ashok et. al's (2013) approach was followed. Specifically, the success of a text was determined by the download count of the text as reported by project Gutenberg. If a text had been downloaded more than 100 times, the text was classified as successful.

The various texts were then read, cleaned, and converted to a bag-of-words table which contained the words and their counts. This table excluded the top and bottom 20% of ngrams to exclude very rare and very common ngrams from the analysis. This was then used by the classifier to determine whether or not a book was successful.

For analysis, the corpus was divided into two datasets - training and testing. The training data was further divided into 5 folds for five-fold cross-validation. The classifier was programmed in Python by Dr. Andreas van Cranenburgh. The classifier used the linear model logistic regression because it is a useful model for binary dependent variables (Kessler, Nunberg, & Schuetze, 1997). Logistic regression models the probability of a situation occurring. Unlike, linear regression, which models data points against a straight line, logistic regression models data points against an asymptote-like graph known as the logistic regression graph. The classifier further uses Limited Memory Broyden-Fletcher-Goldfarb-Shanno algorithm to to minimise error.

I then ran the classifier by tweaking two parameters in the application of the bag-of-words model- ngram, range which controls how many words should be taken together; and term frequency-inverse document frequency (tf-idf) which determines the weight of a word for a specific text. I ran the model 6 times with following parameters:

Run No.	IDF	ngram Range
1	True	2,2
2	False	2,2
3	True	2,3
4	False	2,3
5	True	1,3
6	False	1,3

The complete results of these runs are in the appendix. Here I will discuss

two noteworthy runs -3 and 6.

3 Results

Science-Fiction consistently had the lowest scores, the highest being 0.6364 and lowest being 0.4545, lower than 0.50 meaning that the classifier was worse than random in this regard. The genres poetry and love performed well reaching highs such as 0.9 for love and 0.9048 for poetry. It appeared that the classifiers was somehow more attuned to the words that determined success and failure for these genres.

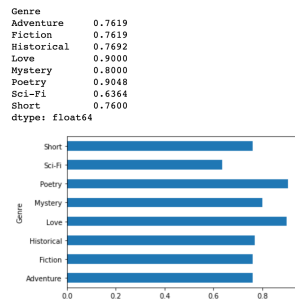


Figure 1: Precision scores after test 3 divided by genre

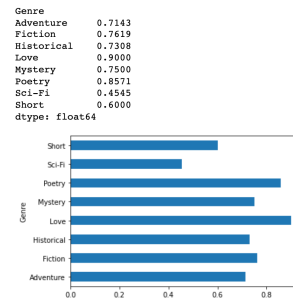


Figure 2: Precision scores after test 6 divided by genre

Underwood and Sellers’ (2016) analysis of the genre of poetry found that their model identifies “concrete language and a relatively dark tone” (Underwood & Sellers, 2016, p. 332) as indicators of success. On reviewing the words that are firmly associated with success and failure in our case, such a theme does seem present, if subtle.

FAILURE			SUCCESS		neutral	
	word	score	word	score	word	weight
1	she said	-0.0261	she had	0.0305	and red	1.1673e-05
2	the girl	-0.0226	that she	0.0118	to touch	1.4418e-05
3	as she	-0.0164	looked at	0.0115	telling me	-1.6980e-05
4	the young	-0.0158	the house	0.0098	looked so	2.1419e-05
5	the boys	-0.0151	you re	0.0094	than one	2.5847e-05
6	the doctor	-0.0112	the wind	0.0090	of the most	2.8971e-05
7	we ll	-0.0083	kind of	0.0089	take his	-3.4013e-05
8	and she	-0.0082	so that	0.0088	as this	3.8476e-05
9	her eyes	-0.0079	said he	0.0086	his former	-4.2399e-05
10	she is	-0.0075	which was	0.0085	if to	4.4434e-05
11	his father	-0.0061	as though	0.0081	of flowers	-5.0435e-05
12	her father	-0.0055	the sea	0.0080	make no	5.5209e-05
13	my father	-0.0052	in which	0.0078	fresh and	5.5904e-05
14	the boy	-0.0049	among the	0.0071	will say	6.3527e-05
15	we can	-0.0046	of which	0.0070	you at	6.8035e-05
16	her mother	-0.0045	the window	0.0069	up to him	-9.4428e-05
17	he has	-0.0044	and said	0.0063	door to	1.0309e-04
18	any one	-0.0043	the very	0.0063	all that was	1.1153e-04
19	the young man	-0.0043	let us	0.0059	have told you	1.1267e-04
20	the river	-0.0043	the earth	0.0058	last words	1.1382e-04

Figure 3: The top twenty words most strongly associated with each label and words with the least weight for run 3.

FAILURE			SUCCESS		neutral	
word	score	word	score	word	weight	
1	girl	-0.0361	mr	0.0356	how it	1.4028e-05
2	miss	-0.0203	she had	0.0240	left it	3.3228e-05
3	she said	-0.0157	re	0.0205	kissed	-4.5259e-05
4	boys	-0.0156	thou	0.0201	and the old	4.5884e-05
5	exclaimed	-0.0149	sir	0.0198	up by	4.6189e-05
6	as she	-0.0110	de	0.0171	me like	-4.9817e-05
7	the young	-0.0106	wind	0.0150	was silent	5.0203e-05
8	sweet	-0.0104	letter	0.0132	she would not	5.1019e-05
9	the girl	-0.0103	that she	0.0100	well to	-7.2602e-05
10	tom	-0.0094	sea	0.0098	into that	7.7401e-05
11	chapter	-0.0088	window	0.0097	hand upon	7.9259e-05
12	doctor	-0.0087	lady	0.0095	ll take	-7.9640e-05
13	laughed	-0.0072	person	0.0091	so often	-8.4637e-05
14	er	-0.0070	mrs	0.0087	was heard	-1.0645e-04
15	jack	-0.0067	thus	0.0083	all for	1.1720e-04
16	uncle	-0.0057	looked at	0.0074	act of	1.2243e-04
17	frank	-0.0056	clock	0.0071	shout	-1.2502e-04
18	party	-0.0055	the house	0.0071	curse	-1.3108e-04
19	safe	-0.0053	idea	0.0069	man whom	-1.3157e-04
20	sister	-0.0051	thee	0.0068	more than one	1.3436e-04

Figure 4: The top twenty words most strongly associated with each label and words with the least weight for run 6.

Positive unigrams like “exclaimed”, “laughed”, “sweet” seemed to be present more on the side of failure. However, the label of success did not show a clear bias towards dark or negative sentiment. Rather, it appeared to be indifferent.

The science-fiction books present in the corpus were mainly dark. For ex-

ample, *The Wolf’s Long Howl* follows an insomniac protagonist kept awake by the howl of a wolf which only they can hear (*The Wolf’s Long Howl by Stanley Waterloo*, n.d.). *2BR02B* follows a citizen of a society where no one ages and for the birth of a child, an adult must sacrifice their life (*Kurt Vonnegut’s Short Stories “2BR02B” Summary and Analysis*, n.d.). And yet, science fiction books were not easy to classify as successful or unsuccessful.

Gender also played an important role in the classification of success and failure. Feminine pronouns were associated more with failure than with success. When feminine pronouns were mentioned on the side of success, they were descriptive, “she had”. Whereas on the side of failure, they played a more active role, “she said”. Masculine pronouns when associated with youth (“the boys”) or old-age (“father”) were associated with failure. When masculine pronouns were associated with respect (“sir”, ‘Mr.’), they were on the side of success. As discussed by Piper and Portelance (2016) gender was represented in a conservative manner. However, success in general has less pronouns than failure.

This was also odd, since most science-fiction books in the corpus followed a male protagonist.

4 Conclusion

Using a classifier trained on diverse genres still performs better on certain genres as compared to others. Thus it might be more useful to create genre specific classifiers which can be attuned to the various linguistic patterns and literary devices unique to that genre. For example, clues for the genre of mystery (Moretti, 2000). This analysis does not provide any conclusive answers to the research question, “What makes the Science-Fiction genre difficult to classify by a classifier trained on diverse genres?”. However, it has explored some aspects that may help answer the question, such as the role of gender and sentiment.

References

- Ashok, V. G., Feng, S., & Choi, Y. (2013). Success with style: Using writing style to predict the success of novels. In *Proceedings of the 2013 conference on empirical methods in natural language processing* (pp. 1753–1764). (n.d.). Retrieved from <https://www.gutenberg.org/>
- Kessler, B., Nunberg, G., & Schuetze, H. (1997). Automatic detection of text genre. *Annual Meeting- Association for Computational Linguistics*, 35, 32–38.
- Kurt vonnegut's short stories "2br02b" summary and analysis.* (n.d.). Retrieved from <https://www.gradesaver.com/kurt-vonnegut-short-stories/study-guide/summary-2br02b>
- Moretti, F. (2000). The slaughterhouse of literature. *MLQ: Modern Language Quarterly*, 61(1), 207–227.
- Piper, A., & Portelance, E. (2016, May). *How cultural capital works: Prizewinning novels, bestsellers, and the time of reading.* Retrieved from <http://post45.org/2016/05/how-cultural-capital-works-prizewinning-novels-bestsellers-and-the-time-of-reading/>
- underwood, T. (2014, May). *We don't already understand the broad outlines of literary history.* Retrieved from <https://tedunderwood.com/2013/02/08/we-dont-already-know-the-broad-outlines-of-literary-history/>
- Underwood, T., & Sellers, J. (2016). The longue durée of literary prestige. *Modern Language Quarterly*, 77(3), 321–344.
- The wolf's long howl by stanley waterloo.* (n.d.). Goodreads. Retrieved from <https://www.goodreads.com/book/show/10373421-the-wolf-s-long-howl>

Appendix A

Results Run 1

	precision	recall	f1-score	support
FAILURE	0.71	0.86	0.78	76
SUCCESS	0.82	0.65	0.73	75
accuracy			0.75	151
macro avg	0.77	0.75	0.75	151
weighted avg	0.77	0.75	0.75	151

	FAILURE	SUCCESS
FAILURE	65	11
SUCCESS	26	49

Figure 5: Results of evaluating the classifier after training.

	precision	recall	f1-score	support
FAILURE	0.76	0.78	0.77	80
SUCCESS	0.77	0.75	0.76	79
accuracy			0.76	159
macro avg	0.76	0.76	0.76	159
weighted avg	0.76	0.76	0.76	159

8]:

	FAILURE	SUCCESS
FAILURE	62	18
SUCCESS	20	59

Figure 6: Test performance of the classifier.


```

Genre
Adventure    0.7619
Fiction       0.8095
Historical    0.7692
Love          0.9000
Mystery       0.7500
Poetry        0.8571
Sci-Fi        0.6364
Short         0.7600
dtype: float64

```

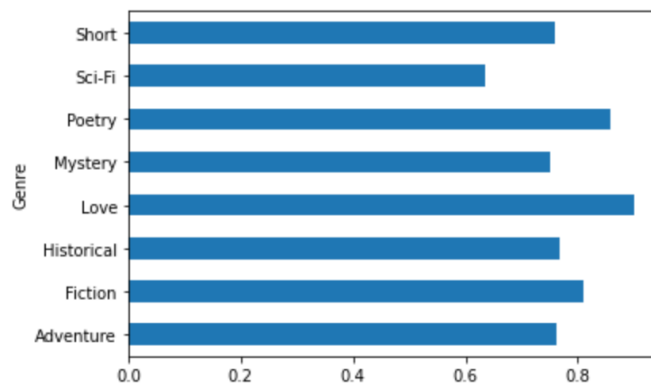


Figure 7: Precision scores after test divided by genre.

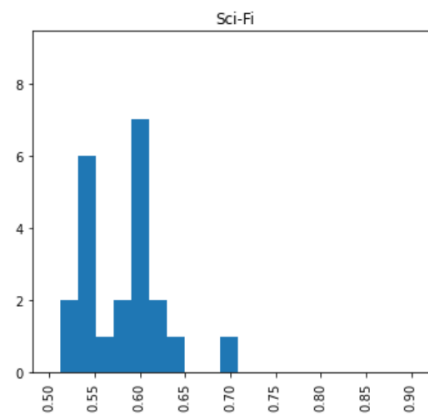


Figure 8: Bar graph showing the distribution of the classifier's confidence of a science-fiction book having the most likely label.

Filename	Label	Genre	Fold	Success	Title	Author	Language	DownloadCount	predicted	prob
10391.txt	Waterloo_TheWolfSLongHowl	Sci-Fi	1	FAILURE	the wolf's long howl	waterloo, stanley, 1846-1913	en	30	FAILURE	0.5129
21279.txt	Vonnegut_2BR02B	Sci-Fi	3	SUCCESS	2 b r 0 2 b	vonnegut, kurt, 1922-2007	en	2811	SUCCESS	0.5223
19029.txt	Norton_TheGiftsOfAsti	Sci-Fi	1	SUCCESS	the gifts of asti	norton, andre, 1912-2005	en	144	FAILURE	0.5403
29601.txt	Robles_See	Sci-Fi	1	FAILURE	see?	robles, edward g.	en	29	FAILURE	0.5405
30199.txt	Derleth_McIlvaineSStar	Sci-Fi	1	SUCCESS	mcilvaine's star	derleth, august william, 1...	en	196	SUCCESS	0.5408
30680.txt	Olin_AllDayWednesday	Sci-Fi	1	FAILURE	all day wednesday	olin, richard	en	24	FAILURE	0.5442
18342.txt	Piper_TheAnswer	Sci-Fi	1	SUCCESS	the answer	piper, h. beam, 1904-1964	en	131	SUCCESS	0.5461
28156.txt	Sharkey_MinorDetail	Sci-Fi	1	FAILURE	minor detail	sharkey, jack, 1931-	en	22	SUCCESS	0.5505
24517.txt	Baily_AccidentalDeath	Sci-Fi	1	SUCCESS	accidental death	baily, peter	en	171	FAILURE	0.5622
24927.txt	Sevcik_AIMatterOfMagnitude	Sci-Fi	1	FAILURE	a matter of magnitude	sevcik, al	en	29	FAILURE	0.5737
29548.txt	Sheckley_WarriorRace	Sci-Fi	1	SUCCESS	warrior race	sheckley, robert, 1928-2005	en	108	FAILURE	0.5882
26955.txt	Huekels_AdvancedChemistry	Sci-Fi	1	SUCCESS	advanced chemistry	huekels, jack g.	en	137	FAILURE	0.5956
30386.txt	Marlowe_TheOneAndTheMany	Sci-Fi	1	FAILURE	the one and the many	marlowe, stephen, 1928-2008	en	23	FAILURE	0.5971
29559.txt	Farmer_TheyTwinkledLikeJewels	Sci-Fi	1	SUCCESS	they twinkled like jewels	farmer, philip José, 1918-...	en	285	SUCCESS	0.6050
30528.txt	Horne_TheDayOfTheDog	Sci-Fi	1	FAILURE	the day of the dog	horne, anderson	en	21	FAILURE	0.6057
29019.txt	Norton_AllCatsAreGray	Sci-Fi	1	SUCCESS	all cats are gray	norton, andre, 1912-2005	en	376	FAILURE	0.6066
18109.txt	Piper_GraveyardOfDreams	Sci-Fi	1	SUCCESS	graveyard of dreams	piper, h. beam, 1904-1964	en	115	FAILURE	0.6079
30438.txt	Mckimmey_TheEyesHaveIt	Sci-Fi	1	FAILURE	the eyes have it	mckimmey, james, 1923-	en	27	FAILURE	0.6097
30240.txt	Vonnegut_TheBigTripUpYonder	Sci-Fi	3	SUCCESS	the big trip up yonder	vonnegut, kurt, 1922-2007	en	972	SUCCESS	0.6220
29458.txt	Sheckley_CostOfLiving	Sci-Fi	1	SUCCESS	cost of living	sheckley, robert, 1928-2005	en	106	FAILURE	0.6259
30382.txt	Coggins_SayHelloForMe	Sci-Fi	1	FAILURE	say "hello" for me	coggins, frank w.	en	12	FAILURE	0.6469
30329.txt	Marlowe_BlackEyesAndTheDai...	Sci-Fi	1	FAILURE	black eyes and the dai...	marlowe, stephen, 1928-2008	en	23	FAILURE	0.7081

Figure 9: Title-wise break down of probability values.

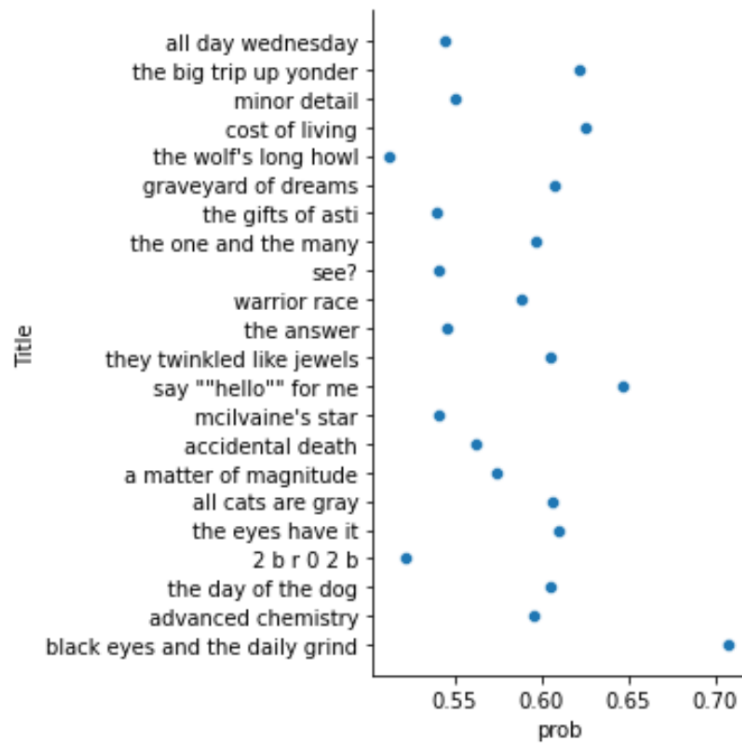


Figure 10: Scatter plot of probability values of science fiction titles.

FAILURE			SUCCESS		neutral	
	word	score	word	score	word	weight
1	she said	-0.0266	she had	0.0346	so and	6.7685e-06
2	the girl	-0.0242	that she	0.0136	her again	-1.5205e-05
3	the young	-0.0176	looked at	0.0127	found myself	2.1110e-05
4	as she	-0.0174	the house	0.0106	them as	2.4321e-05
5	the boys	-0.0157	you re	0.0097	as mr	-3.2707e-05
6	the doctor	-0.0117	so that	0.0095	that when	4.4922e-05
7	we ll	-0.0088	the wind	0.0094	engaged in	5.8030e-05
8	her eyes	-0.0084	kind of	0.0094	how she	6.0897e-05
9	and she	-0.0081	said he	0.0093	than we	6.9892e-05
10	she is	-0.0080	which was	0.0091	love in	-7.1252e-05
11	his father	-0.0065	as though	0.0090	yes he	-7.1701e-05
12	her father	-0.0059	in which	0.0086	anything that	9.6106e-05
13	my father	-0.0055	the sea	0.0085	drew the	-9.9934e-05
14	the boy	-0.0051	among the	0.0075	have thought	-1.0478e-04
15	we can	-0.0049	of which	0.0074	see he	-1.0721e-04
16	he has	-0.0048	the window	0.0074	you at	-1.0918e-04
17	her mother	-0.0047	and said	0.0068	the arrival	1.1423e-04
18	any one	-0.0046	the very	0.0067	give up	1.2557e-04
19	the river	-0.0045	let us	0.0061	opened it	-1.4024e-04
20	with me	-0.0043	was that	0.0060	more or	-1.4789e-04

Figure 11: The top twenty words most strongly associated with each label and words with the least weight.

Appendix B

Results Run 2

	precision	recall	f1-score	support
FAILURE	0.70	0.82	0.76	76
SUCCESS	0.78	0.65	0.71	75
accuracy			0.74	151
macro avg	0.74	0.73	0.73	151
weighted avg	0.74	0.74	0.73	151

	FAILURE	SUCCESS
FAILURE	62	14
SUCCESS	26	49

Figure 12: Results of evaluating the classifier after training.

	precision	recall	f1-score	support
FAILURE	0.75	0.75	0.75	80
SUCCESS	0.75	0.75	0.75	79
accuracy			0.75	159
macro avg	0.75	0.75	0.75	159
weighted avg	0.75	0.75	0.75	159

	FAILURE	SUCCESS
FAILURE	60	20
SUCCESS	20	59

Figure 13: Test performance of the classifier.

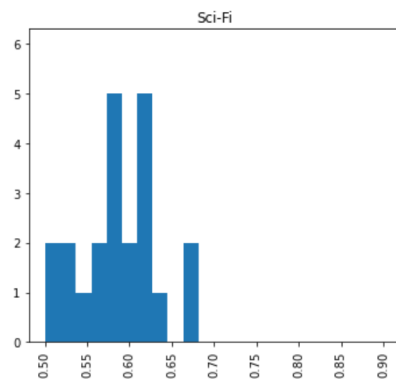


Figure 14: Bar graph showing the distribution of the classifier's confidence of a science-fiction book having the most likely label.

```

Genre
Adventure    0.7143
Fiction      0.8095
Historical   0.7308
Love         0.9000
Mystery      0.8000
Poetry       0.9048
Sci-Fi       0.5455
Short        0.7200
dtype: float64

```

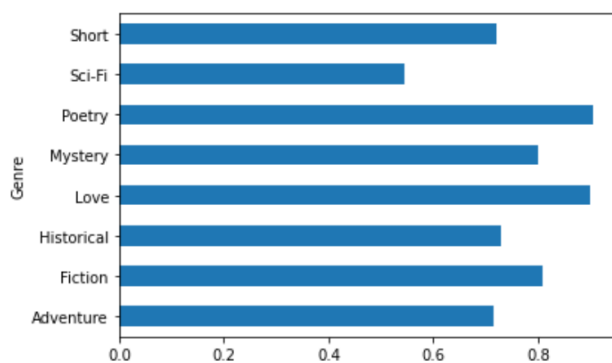


Figure 15: Precision scores after test divided by genre.

Filename	Label	Genre	Fold	Success	Title	Author	Language	DownloadCount	predicted	prob
21279.txt	Vonnegut_2BR02B	Sci-Fi	3	SUCCESS	2 b r 0 2 b	vonnegut, kurt, 1922-2007	en	2811	FAILURE	0.5007
10391.txt	Waterloo_TheWolfSLongHowl	Sci-Fi	1	FAILURE	the wolf's long howl	waterloo, stanley, 1846-1913	en	30	SUCCESS	0.5106
28156.txt	Sharkey_MinorDetail	Sci-Fi	1	FAILURE	minor detail	sharkey, jack, 1931-	en	22	SUCCESS	0.5294
19029.txt	Norton_TheGiftsOfAsti	Sci-Fi	1	SUCCESS	the gifts of asti	norton, andre, 1912-2005	en	144	FAILURE	0.5316
29601.txt	Robles_See	Sci-Fi	1	FAILURE	see?	robles, edward g.	en	29	FAILURE	0.5396
30680.txt	Olin_AllDayWednesday	Sci-Fi	1	FAILURE	all day wednesday	olin, richard	en	24	FAILURE	0.5578
18342.txt	Piper_TheAnswer	Sci-Fi	1	SUCCESS	the answer	piper, h. beam, 1904-1964	en	131	SUCCESS	0.5687
30199.txt	Derleth_McIlvaineSStar	Sci-Fi	1	SUCCESS	mcilvaine's star	derleth, august william, 1...	en	196	SUCCESS	0.5781
29559.txt	Farmer_TheyTwinkledLikeJewels	Sci-Fi	1	SUCCESS	they twinkled like jewels	farmer, philip José, 1918-...	en	285	SUCCESS	0.5817
24517.txt	Baily_AccidentalDeath	Sci-Fi	1	SUCCESS	accidental death	baily, peter	en	171	FAILURE	0.5835
30438.txt	Mckimmey_TheEyesHaveIt	Sci-Fi	1	FAILURE	the eyes have it	mckimmey, james, 1923-	en	27	FAILURE	0.5835
30528.txt	Horne_TheDayOfTheDog	Sci-Fi	1	FAILURE	the day of the dog	horne, anderson	en	21	FAILURE	0.5905
30386.txt	Marlowe_TheOneAndTheMany	Sci-Fi	1	FAILURE	the one and the many	marlowe, stephen, 1928-2008	en	23	FAILURE	0.5919
29548.txt	Sheckley_WarriorRace	Sci-Fi	1	SUCCESS	warrior race	sheckley, robert, 1928-2005	en	108	FAILURE	0.5964
30240.txt	Vonnegut_TheBigTripUpYonder	Sci-Fi	3	SUCCESS	the big trip up yonder	vonnegut, kurt, 1922-2007	en	972	SUCCESS	0.6100
24927.txt	Sevcik_AMatterOfMagnitude	Sci-Fi	1	FAILURE	a matter of magnitude	sevcik, al	en	29	FAILURE	0.6184
26955.txt	Huekels_AdvancedChemistry	Sci-Fi	1	SUCCESS	advanced chemistry	huekels, jack g.	en	137	FAILURE	0.6216
29019.txt	Norton_AllCatsAreGray	Sci-Fi	1	SUCCESS	all cats are gray	norton, andre, 1912-2005	en	376	FAILURE	0.6219
18109.txt	Piper_GraveyardOfDreams	Sci-Fi	1	SUCCESS	graveyard of dreams	piper, h. beam, 1904-1964	en	115	FAILURE	0.6244
29458.txt	Sheckley_CostOfLiving	Sci-Fi	1	SUCCESS	cost of living	sheckley, robert, 1928-2005	en	106	FAILURE	0.6296
30382.txt	Coggins_SayHelloForMe	Sci-Fi	1	FAILURE	say ""hello"" for me	coggins, frank w.	en	12	FAILURE	0.6660
30329.txt	Marlowe_BlackEyesAndTheDai...	Sci-Fi	1	FAILURE	black eyes and the daily g...	marlowe, stephen, 1928-2008	en	23	FAILURE	0.6814

Figure 16: Title-wise break down of probability values.

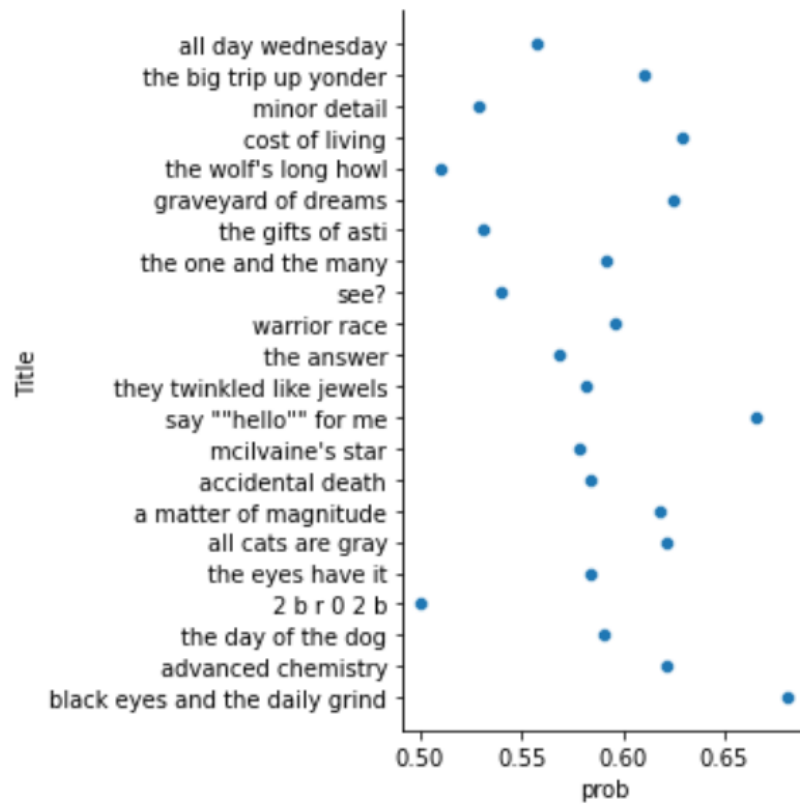


Figure 17: Scatter plot of probability values of science fiction titles.

FAILURE			SUCCESS		neutral	
	word	score	word	score	word	weight
1	she said	-0.0346	she had	0.0483	same moment	-5.8915e-06
2	as she	-0.0265	looked at	0.0192	and covered	1.1131e-05
3	the young	-0.0255	that she	0.0188	be little	1.1963e-05
4	the girl	-0.0249	the house	0.0158	carried the	-1.2197e-05
5	and she	-0.0130	so that	0.0142	they gave	-2.4314e-05
6	the boys	-0.0119	which was	0.0134	me like	2.7034e-05
7	her eyes	-0.0117	kind of	0.0131	eyes at	3.2994e-05
8	we ll	-0.0113	said he	0.0129	but once	3.9566e-05
9	she is	-0.0103	in which	0.0127	me when	4.2873e-05
10	the doctor	-0.0100	you re	0.0126	to view	-5.2764e-05
11	we can	-0.0079	the wind	0.0121	the youngest	-6.3457e-05
12	his father	-0.0074	among the	0.0108	kept him	6.9546e-05
13	with me	-0.0070	of which	0.0108	them she	7.1049e-05
14	he has	-0.0069	the very	0.0105	the fireplace	7.2807e-05
15	went on	-0.0063	the window	0.0095	whether the	-1.0138e-04
16	like to	-0.0059	and said	0.0094	contrary to	1.0303e-04
17	my father	-0.0058	the sea	0.0091	or had	1.0344e-04
18	you were	-0.0058	was that	0.0091	reach of	-1.1001e-04
19	the boy	-0.0058	let us	0.0089	how she	-1.1832e-04
20	any one	-0.0056	and what	0.0088	could she	-1.2163e-04

Figure 18: The top twenty words most strongly associated with each label and words with the least weight.

Appendix C

Results Run 3

	precision	recall	f1-score	support
FAILURE	0.70	0.86	0.77	76
SUCCESS	0.81	0.63	0.71	75
accuracy			0.74	151
macro avg	0.75	0.74	0.74	151
weighted avg	0.75	0.74	0.74	151

	FAILURE	SUCCESS
FAILURE	65	11
SUCCESS	28	47

Figure 19: Results of evaluating the classifier after training.

	precision	recall	f1-score	support
FAILURE	0.77	0.76	0.77	80
SUCCESS	0.76	0.77	0.77	79
accuracy			0.77	159
macro avg	0.77	0.77	0.77	159
weighted avg	0.77	0.77	0.77	159

	FAILURE	SUCCESS
FAILURE	61	19
SUCCESS	18	61

Figure 20: Test performance of the classifier.

```

Genre
Adventure    0.7619
Fiction       0.7619
Historical    0.7692
Love          0.9000
Mystery       0.8000
Poetry        0.9048
Sci-Fi        0.6364
Short         0.7600
dtype: float64

```

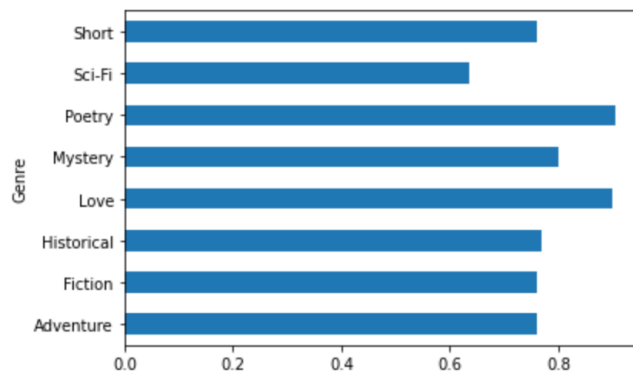


Figure 21: Precision scores after test divided by genre.

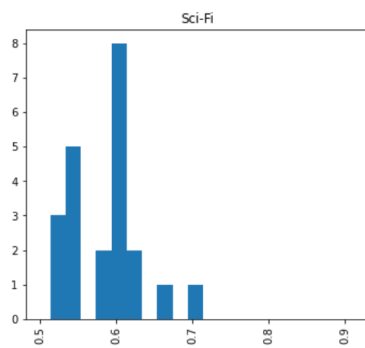


Figure 22: Bar graph showing the distribution of the classifier's confidence of a science-fiction book having the most likely label.

Label	Genre	Fold	Success	Title	Author	Language	DownloadCount	predicted	prob
Filename									
10391.txt	Waterloo_TheWolfSLongHowl	Sci-Fi	1	FAILURE	the wolf's long howl	waterloo, stanley, 1846-1913	en	30	FAILURE 0.5132
30199.txt	Derleth_McIlvaineSStar	Sci-Fi	1	SUCCESS	mcilvaine's star	derleth, august william, 1...	en	196	SUCCESS 0.5135
21279.txt	Vonnegut_2BR02B	Sci-Fi	3	SUCCESS	2 b r 0 2 b	vonnegut, kurt, 1922-2007	en	2811	SUCCESS 0.5277
28156.txt	Sharkey_MinorDetail	Sci-Fi	1	FAILURE	minor detail	sharkey, jack, 1931-	en	22	SUCCESS 0.5392
30680.txt	Olin_AllDayWednesday	Sci-Fi	1	FAILURE	all day wednesday	olin, richard	en	24	FAILURE 0.5398
19029.txt	Norton_TheGiftsOfAsti	Sci-Fi	1	SUCCESS	the gifts of asti	norton, andre, 1912-2005	en	144	FAILURE 0.5433
18342.txt	Piper_TheAnswer	Sci-Fi	1	SUCCESS	the answer	piper, h. beam, 1904-1964	en	131	SUCCESS 0.5445
29601.txt	Robles_See	Sci-Fi	1	FAILURE	see?	robles, edward g.	en	29	FAILURE 0.5480
30438.txt	Mckimmey_TheEyesHaveIt	Sci-Fi	1	FAILURE	the eyes have it	mckimmey, james, 1923-	en	27	FAILURE 0.5922
29548.txt	Sheckley_WarriorRace	Sci-Fi	1	SUCCESS	warrior race	sheckley, robert, 1928-2005	en	108	FAILURE 0.5924
29019.txt	Norton_AllCatsAreGray	Sci-Fi	1	SUCCESS	all cats are gray	norton, andre, 1912-2005	en	376	FAILURE 0.5948
18109.txt	Piper_GraveyardOfDreams	Sci-Fi	1	SUCCESS	graveyard of dreams	piper, h. beam, 1904-1964	en	115	FAILURE 0.5996
24517.txt	Baily_AccidentalDeath	Sci-Fi	1	SUCCESS	accidental death	baily, peter	en	171	FAILURE 0.5999
24927.txt	Sevcik_AMatterOfMagnitude	Sci-Fi	1	FAILURE	a matter of magnitude	sevcik, al	en	29	FAILURE 0.6044
29458.txt	Sheckley_CostOfLiving	Sci-Fi	1	SUCCESS	cost of living	sheckley, robert, 1928-2005	en	106	FAILURE 0.6073
30386.txt	Marlowe_TheOneAndTheMany	Sci-Fi	1	FAILURE	the one and the many	marlowe, stephen, 1928-2008	en	23	FAILURE 0.6075
26955.txt	Huekels_AdvancedChemistry	Sci-Fi	1	SUCCESS	advanced chemistry	huekels, jack g.	en	137	FAILURE 0.6109
30528.txt	Horne_TheDayOfTheDog	Sci-Fi	1	FAILURE	the day of the dog	horne, anderson	en	21	FAILURE 0.6135
29559.txt	Farmer_TheyTwinkledLikeJewels	Sci-Fi	1	SUCCESS	they twinkled like jewels	farmer, philip josé, 1918-...	en	285	SUCCESS 0.6173
30240.txt	Vonnegut_TheBigTripUpYonder	Sci-Fi	3	SUCCESS	the big trip up yonder	vonnegut, kurt, 1922-2007	en	972	SUCCESS 0.6283
30382.txt	Coggins_SayHelloForMe	Sci-Fi	1	FAILURE	say "hello" for me	coggins, frank w.	en	12	FAILURE 0.6651
30329.txt	Marlowe_BlackEyesAndTheDai...	Sci-Fi	1	FAILURE	black eyes and the daily a...	marlowe, stephen, 1928-2008	en	23	FAILURE 0.7138

Figure 23: Title-wise break down of probability values.

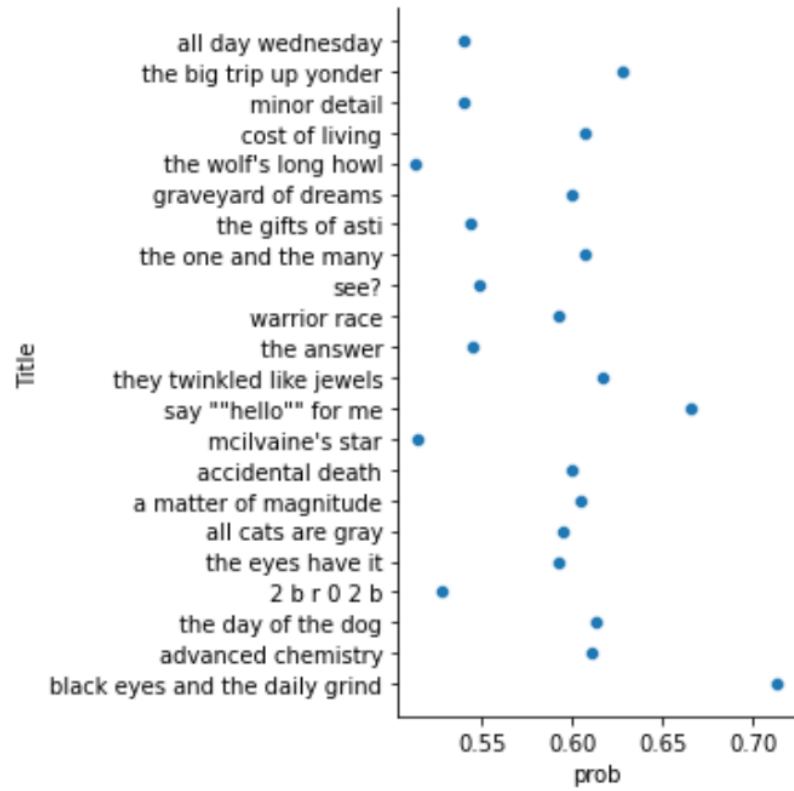


Figure 24: Scatter plot of probability values of science fiction titles.

FAILURE			SUCCESS		neutral	
	word	score	word	score	word	weight
1	she said	-0.0261	she had	0.0305	and red	1.1673e-05
2	the girl	-0.0226	that she	0.0118	to touch	1.4418e-05
3	as she	-0.0164	looked at	0.0115	telling me	-1.6980e-05
4	the young	-0.0158	the house	0.0098	looked so	2.1419e-05
5	the boys	-0.0151	you re	0.0094	than one	2.5847e-05
6	the doctor	-0.0112	the wind	0.0090	of the most	2.8971e-05
7	we ll	-0.0083	kind of	0.0089	take his	-3.4013e-05
8	and she	-0.0082	so that	0.0088	as this	3.8476e-05
9	her eyes	-0.0079	said he	0.0086	his former	-4.2399e-05
10	she is	-0.0075	which was	0.0085	if to	4.4434e-05
11	his father	-0.0061	as though	0.0081	of flowers	-5.0435e-05
12	her father	-0.0055	the sea	0.0080	make no	5.5209e-05
13	my father	-0.0052	in which	0.0078	fresh and	5.5904e-05
14	the boy	-0.0049	among the	0.0071	will say	6.3527e-05
15	we can	-0.0046	of which	0.0070	you at	6.8035e-05
16	her mother	-0.0045	the window	0.0069	up to him	-9.4428e-05
17	he has	-0.0044	and said	0.0063	door to	1.0309e-04
18	any one	-0.0043	the very	0.0063	all that was	1.1153e-04
19	the young man	-0.0043	let us	0.0059	have told you	1.1267e-04
20	the river	-0.0043	the earth	0.0058	last words	1.1382e-04

Figure 25: The top twenty words most strongly associated with each label and words with the least weight.

Appendix D

Results Run 4

	precision	recall	f1-score	support
FAILURE	0.70	0.82	0.75	76
SUCCESS	0.77	0.64	0.70	75
accuracy			0.73	151
macro avg	0.74	0.73	0.73	151
weighted avg	0.74	0.73	0.73	151

	FAILURE	SUCCESS
FAILURE	62	14
SUCCESS	27	48

Figure 26: Results of evaluating the classifier after training.

	precision	recall	f1-score	support
FAILURE	0.76	0.75	0.75	80
SUCCESS	0.75	0.76	0.75	79
accuracy			0.75	159
macro avg	0.75	0.75	0.75	159
weighted avg	0.75	0.75	0.75	159

	FAILURE	SUCCESS
FAILURE	60	20
SUCCESS	19	60

Figure 27: Test performance of the classifier.

```

Genre
Adventure    0.7143
Fiction       0.8095
Historical    0.7308
Love          0.9000
Mystery       0.8000
Poetry        0.9048
Sci-Fi        0.5909
Short         0.7600
dtype: float64

```

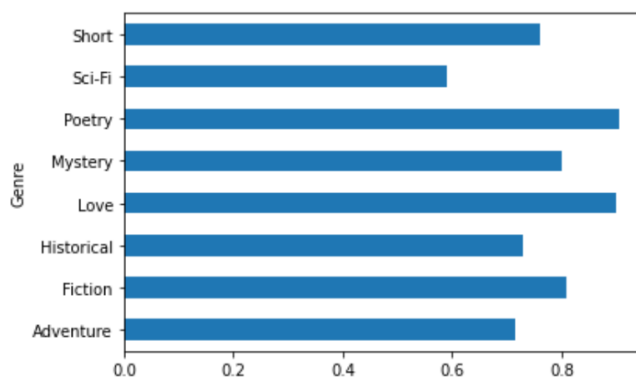


Figure 28: Precision scores after test divided by genre.

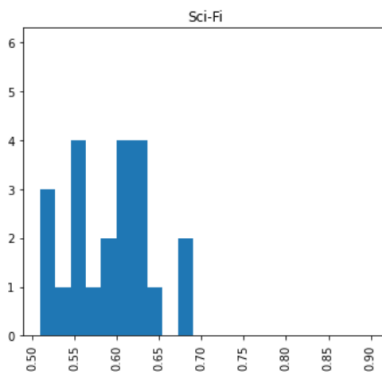


Figure 29: Bar graph showing the distribution of the classifier's confidence of a science-fiction book having the most likely label.

	Label	Genre	Fold	Success	Title	Author	Language	DownloadCount	predicted	prob
Filename										
10391.txt	Waterloo_TheWolfSLongHowl	Sci-Fi	1	FAILURE	the wolf's long howl	waterloo, stanley, 1846-1913	en	30	SUCCESS	0.5092
21279.txt	Vonnegut_2BR02B	Sci-Fi	3	SUCCESS	2 b r 0 2 b	vonnegut, kurt, 1922-2007	en	2811	SUCCESS	0.5103
28156.txt	Sharkey_MinorDetail	Sci-Fi	1	FAILURE	minor detail	sharkey, jack, 1931-	en	22	SUCCESS	0.5180
19029.txt	Norton_TheGiftsOfAsti	Sci-Fi	1	SUCCESS	the gifts of asti	norton, andre, 1912-2005	en	144	FAILURE	0.5387
30680.txt	Olin_AllDayWednesday	Sci-Fi	1	FAILURE	all day wednesday	olin, richard	en	24	FAILURE	0.5462
29601.txt	Robles_See	Sci-Fi	1	FAILURE	see?	robles, edward g.	en	29	FAILURE	0.5471
30199.txt	Derleth_McIlvaineSStar	Sci-Fi	1	SUCCESS	mcilvaine's star	derleth, august william, 1...	en	196	SUCCESS	0.5495
18342.txt	Piper_TheAnswer	Sci-Fi	1	SUCCESS	the answer	piper, h. beam, 1904-1964	en	131	SUCCESS	0.5594
30438.txt	Mckimmey_TheEyesHaveIt	Sci-Fi	1	FAILURE	the eyes have it	mckimmey, james, 1923-	en	27	FAILURE	0.5725
30528.txt	Horne_TheDayOfTheDog	Sci-Fi	1	FAILURE	the day of the dog	horne, anderson	en	21	FAILURE	0.5923
29559.txt	Farmer_TheyTwinkledLikeJewels	Sci-Fi	1	SUCCESS	they twinkled like jewels	farmer, philip José, 1918-...	en	285	SUCCESS	0.5954
30386.txt	Marlowe_TheOneAndTheMany	Sci-Fi	1	FAILURE	the one and the many	marlowe, stephen, 1928-2008	en	23	FAILURE	0.6026
29548.txt	Sheckley_WarriorRace	Sci-Fi	1	SUCCESS	warrior race	sheckley, robert, 1928-2005	en	108	FAILURE	0.6028
29019.txt	Norton_AllCatsAreGray	Sci-Fi	1	SUCCESS	all cats are gray	norton, andre, 1912-2005	en	376	FAILURE	0.6039
24517.txt	Baily_AccidentalDeath	Sci-Fi	1	SUCCESS	accidental death	baily, peter	en	171	FAILURE	0.6086
29458.txt	Sheckley_CostOfLiving	Sci-Fi	1	SUCCESS	cost of living	sheckley, robert, 1928-2005	en	106	FAILURE	0.6180
18109.txt	Piper_GraveyardOfDreams	Sci-Fi	1	SUCCESS	graveyard of dreams	piper, h. beam, 1904-1964	en	115	FAILURE	0.6182
30240.txt	Vonnegut_TheBigTripUpYonder	Sci-Fi	3	SUCCESS	the big trip up yonder	vonnegut, kurt, 1922-2007	en	972	SUCCESS	0.6212
24927.txt	Sevcik_AMatterOfMagnitude	Sci-Fi	1	FAILURE	a matter of magnitude	sevcik, al	en	29	FAILURE	0.6328
26955.txt	Huekels_AdvancedChemistry	Sci-Fi	1	SUCCESS	advanced chemistry	huekels, jack g.	en	137	FAILURE	0.6359
30382.txt	Coggins_SayHelloForMe	Sci-Fi	1	FAILURE	say ""hello"" for me	coggins, frank w.	en	12	FAILURE	0.6858
30329.txt	Marlowe_BlackEyesAndTheDal...	Sci-Fi	1	FAILURE	black eyes and the daily g...	marlowe, stephen, 1928-2008	en	23	FAILURE	0.6901

Figure 30: Title-wise break down of probability values.

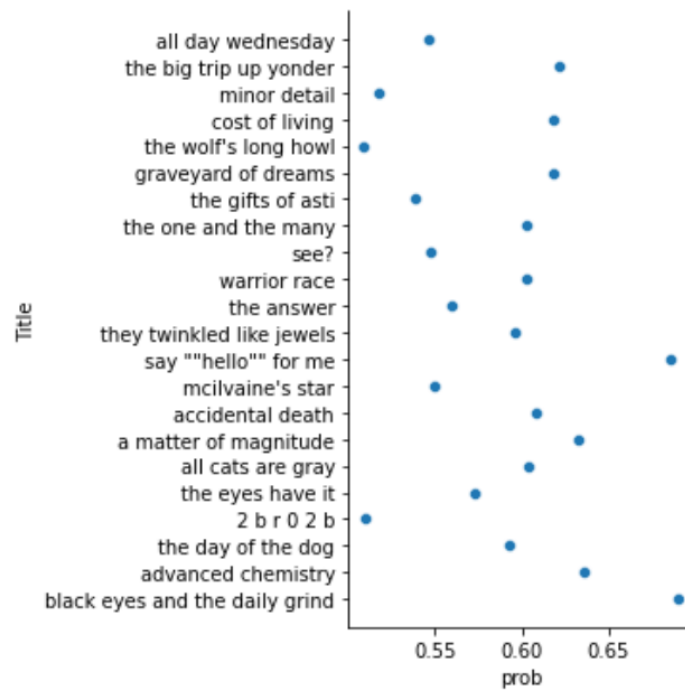


Figure 31: Scatter plot of probability values of science fiction titles.

FAILURE			SUCCESS		neutral	
word	score	word	score	word	weight	
1	she said	-0.0337	she had	0.0436	the change	3.2941e-07
2	as she	-0.0248	looked at	0.0175	is over	6.4965e-07
3	the girl	-0.0233	that she	0.0165	same moment	-2.0614e-05
4	the young	-0.0230	the house	0.0146	shall never	2.3549e-05
5	and she	-0.0125	so that	0.0131	door to	-3.8898e-05
6	the boys	-0.0112	which was	0.0124	some reason	-4.1019e-05
7	her eyes	-0.0110	kind of	0.0122	the arrival	-6.2583e-05
8	we ll	-0.0106	said he	0.0120	with many	6.7069e-05
9	she is	-0.0097	you re	0.0119	stared at	-7.5644e-05
10	the doctor	-0.0095	in which	0.0117	here they	7.6510e-05
11	we can	-0.0073	the wind	0.0114	face that	7.6732e-05
12	his father	-0.0070	among the	0.0101	it really	7.7318e-05
13	with me	-0.0065	of which	0.0101	man would	-9.2491e-05
14	he has	-0.0064	the very	0.0099	remainder of	-9.9653e-05
15	went on	-0.0059	the window	0.0088	each other in	-1.0110e-04
16	the boy	-0.0056	and said	0.0088	the same moment	-1.0249e-04
17	like to	-0.0055	the sea	0.0086	glass and	-1.1009e-04
18	my father	-0.0054	let us	0.0085	his former	-1.1193e-04
19	you were	-0.0053	and what	0.0085	absorbed in	1.2375e-04
20	the heart	-0.0053	was that	0.0084	so but	-1.2635e-04

Figure 32: The top twenty words most strongly associated with each label and words with the least weight.

Appendix E

Results Run 5

	precision	recall	f1-score	support
FAILURE	0.69	0.74	0.71	76
SUCCESS	0.71	0.67	0.69	75
accuracy			0.70	151
macro avg	0.70	0.70	0.70	151
weighted avg	0.70	0.70	0.70	151

	FAILURE	SUCCESS
FAILURE	56	20
SUCCESS	25	50

Figure 33: Results of evaluating the classifier after training.

	precision	recall	f1-score	support
FAILURE	0.72	0.68	0.70	80
SUCCESS	0.69	0.73	0.71	79
accuracy			0.70	159
macro avg	0.71	0.70	0.70	159
weighted avg	0.71	0.70	0.70	159

	FAILURE	SUCCESS
FAILURE	54	26
SUCCESS	21	58

Figure 34: Test performance of the classifier.

```

Genre
Adventure    0.7143
Fiction       0.7143
Historical    0.6923
Love          0.9000
Mystery       0.7500
Poetry        0.8571
Sci-Fi        0.5000
Short         0.6800
dtype: float64

```

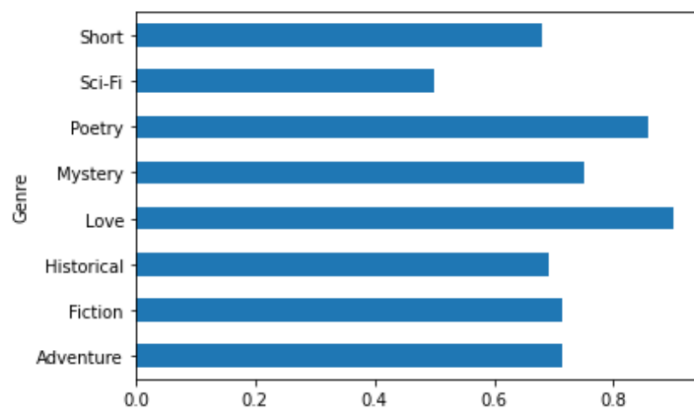


Figure 35: Precision scores after test divided by genre.

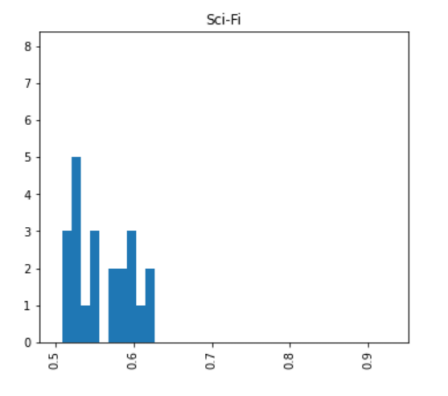


Figure 36: Bar graph showing the distribution of the classifier's confidence of a science-fiction book having the most likely label.

Filename	Label	Genre	Fold	Success	Title	Author	Language	DownloadCount	predicted	prob
29458.txt	Sheckley_CostOfLiving	Sci-Fi	1	SUCCESS	cost of living	sheckley, robert, 1928-2005	en	106	SUCCESS	0.5097
18109.txt	Piper_GraveyardOfDreams	Sci-Fi	1	SUCCESS	graveyard of dreams	piper, h. beam, 1904-1964	en	115	FAILURE	0.5103
24927.txt	Sevcik_AMatterOfMagnitude	Sci-Fi	1	FAILURE	a matter of magnitude	sevcik, al	en	29	SUCCESS	0.5186
30438.txt	Mckimmey_TheEyesHaveIt	Sci-Fi	1	FAILURE	the eyes have it	mckimmey, james, 1923-	en	27	SUCCESS	0.5240
24517.txt	Baily_AccidentalDeath	Sci-Fi	1	SUCCESS	accidental death	baily, peter	en	171	FAILURE	0.5243
30329.txt	Marlowe_BlackEyesAndTheDai...	Sci-Fi	1	FAILURE	black eyes and the dai...	marlowe, stephen, 1928-2008	en	23	FAILURE	0.5252
29601.txt	Robles_See	Sci-Fi	1	FAILURE	see?	robles, edward g.	en	29	SUCCESS	0.5264
26955.txt	Huekels_AdvancedChemistry	Sci-Fi	1	SUCCESS	advanced chemistry	huekels, jack g.	en	137	FAILURE	0.5319
21279.txt	Vonnegut_2BR02B	Sci-Fi	3	SUCCESS	2 b r 0 2 b	vonnegut, kurt, 1922-2007	en	2811	SUCCESS	0.5404
30680.txt	Olin_AllDayWednesday	Sci-Fi	1	FAILURE	all day wednesday	olin, richard	en	24	SUCCESS	0.5461
19029.txt	Norton_TheGiftsOfAsti	Sci-Fi	1	SUCCESS	the gifts of asti	norton, andre, 1912-2005	en	144	SUCCESS	0.5527
10391.txt	Waterloo_TheWolfSLongHowl	Sci-Fi	1	FAILURE	the wolf's long howl	waterloo, stanley, 1846-1913	en	30	FAILURE	0.5544
28156.txt	Sharkey_MinorDetail	Sci-Fi	1	FAILURE	minor detail	sharkey, jack, 1931-	en	22	SUCCESS	0.5711
30382.txt	Coggins_SayHelloForMe	Sci-Fi	1	FAILURE	say ""hello"" for me	coggins, frank w.	en	12	FAILURE	0.5769
29019.txt	Norton_AllCatsAreGray	Sci-Fi	1	SUCCESS	all cats are gray	norton, andre, 1912-2005	en	376	FAILURE	0.5868
29548.txt	Sheckley_WarriorRace	Sci-Fi	1	SUCCESS	warrior race	sheckley, robert, 1928-2005	en	108	FAILURE	0.5900
30199.txt	Derleth_McIlvaineSStar	Sci-Fi	1	SUCCESS	mcilvaine's star	derleth, august william, 1...	en	196	SUCCESS	0.5949
29559.txt	Farmer_TheyTwinkledLikeJewels	Sci-Fi	1	SUCCESS	they twinkled like jewels	farmer, philip José, 1918-...	en	285	FAILURE	0.5987
18342.txt	Piper_TheAnswer	Sci-Fi	1	SUCCESS	the answer	piper, h. beam, 1904-1964	en	131	SUCCESS	0.5994
30240.txt	Vonnegut_TheBigTripUpYonder	Sci-Fi	3	SUCCESS	the big trip up yonder	vonnegut, kurt, 1922-2007	en	972	SUCCESS	0.6073
30528.txt	Home_TheDayOfTheDog	Sci-Fi	1	FAILURE	the day of the dog	home, anderson	en	21	FAILURE	0.6165
30386.txt	Marlowe_TheOneAndTheMany	Sci-Fi	1	FAILURE	the one and the many	marlowe, stephen, 1928-2008	en	23	FAILURE	0.6276

Figure 37: Title-wise break down of probability values.

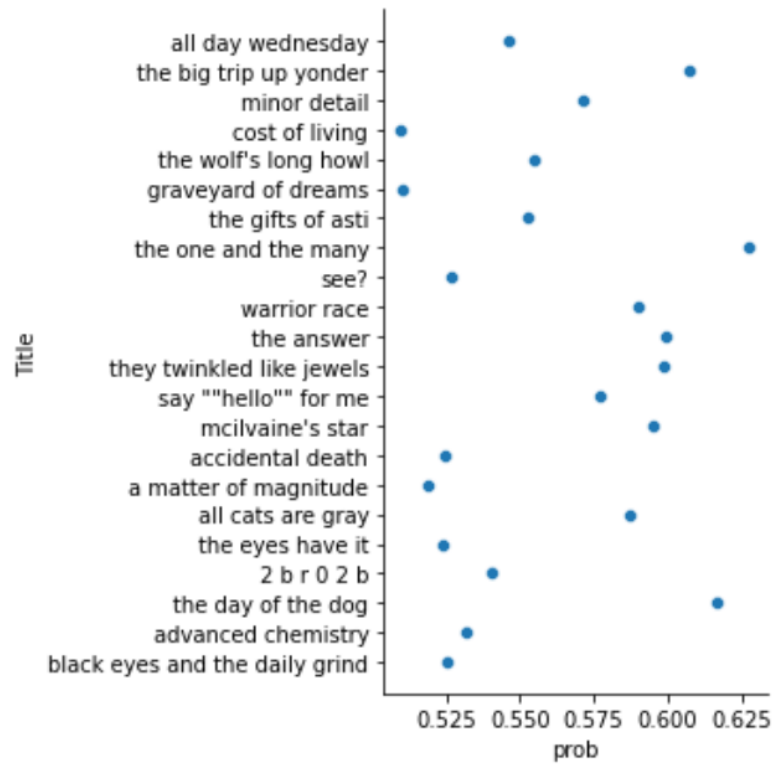


Figure 38: Scatter plot of probability values of science fiction titles.

FAILURE			SUCCESS		neutral	
	word	score	word	score	word	weight
1	girl	-0.0271	mr	0.0350	never had	2.4359e-06
2	miss	-0.0179	thou	0.0220	how the	-5.2547e-06
3	exclaimed	-0.0130	de	0.0186	remarks	8.2004e-06
4	she said	-0.0122	sir	0.0166	put her	-3.2703e-05
5	tom	-0.0119	she had	0.0159	crack	-3.5018e-05
6	boys	-0.0116	re	0.0150	ideal	3.5043e-05
7	the girl	-0.0108	wind	0.0104	that it would	4.5372e-05
8	er	-0.0098	letter	0.0092	not leave	-5.5000e-05
9	doctor	-0.0081	mrs	0.0084	set up	6.9378e-05
10	chapter	-0.0079	sea	0.0080	it was his	7.2849e-05
11	the young	-0.0076	thee	0.0076	far more	-7.2921e-05
12	as she	-0.0075	that she	0.0067	suppose	-7.5321e-05
13	jack	-0.0074	person	0.0066	in new	7.9550e-05
14	sweet	-0.0072	window	0.0065	praise	-9.7403e-05
15	frank	-0.0060	lady	0.0058	to be seen	1.0096e-04
16	uncle	-0.0060	madame	0.0057	would do	-1.0454e-04
17	the boys	-0.0050	mary	0.0056	the door of	1.0795e-04
18	laughed	-0.0048	clock	0.0055	proposed	-1.1019e-04
19	sister	-0.0045	thus	0.0054	from it	1.1817e-04
20	the doctor	-0.0044	towards	0.0054	attached to	-1.2240e-04

Figure 39: The top twenty words most strongly associated with each label and words with the least weight.

Appendix F

Results Run 6

	precision	recall	f1-score	support
FAILURE	0.70	0.75	0.72	76
SUCCESS	0.72	0.67	0.69	75
accuracy			0.71	151
macro avg	0.71	0.71	0.71	151
weighted avg	0.71	0.71	0.71	151

	FAILURE	SUCCESS
FAILURE	57	19
SUCCESS	25	50

Figure 40: Results of evaluating the classifier after training.

	precision	recall	f1-score	support
FAILURE	0.71	0.68	0.69	80
SUCCESS	0.69	0.72	0.70	79
accuracy			0.70	159
macro avg	0.70	0.70	0.70	159
weighted avg	0.70	0.70	0.70	159

	FAILURE	SUCCESS
FAILURE	54	26
SUCCESS	22	57

Figure 41: Test performance of the classifier.

```

Genre
Adventure    0.7143
Fiction      0.7619
Historical   0.7308
Love         0.9000
Mystery      0.7500
Poetry       0.8571
Sci-Fi       0.4545
Short        0.6000
dtype: float64

```

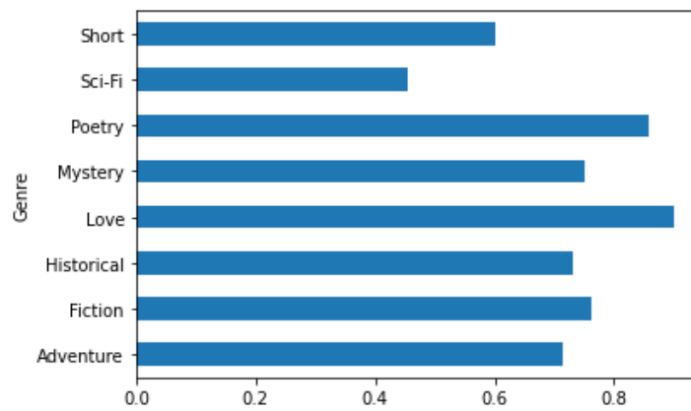


Figure 42: Precision scores after test divided by genre.

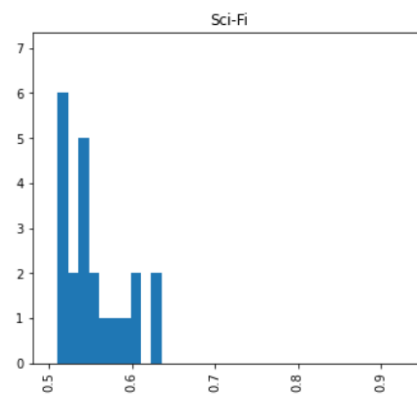


Figure 43: Bar graph showing the distribution of the classifier's confidence of a science-fiction book having the most likely label.

Filename	Label	Genre	Fold	Success	Title	Author	Language	DownloadCount	predicted	prob
24927.txt	Sevcik_AMatterOfMagnitude	Sci-Fi	1	FAILURE	a matter of magnitude	sevcik, al	en	29	SUCCESS	0.5106
18109.txt	Piper_GraveyardOfDreams	Sci-Fi	1	SUCCESS	graveyard of dreams	piper, h. beam, 1904-1964	en	115	FAILURE	0.5121
29601.txt	Robles_See	Sci-Fi	1	FAILURE	see?	robles, edward g.	en	29	SUCCESS	0.5145
30329.txt	Marlowe_BlackEyesAndTheDai...	Sci-Fi	1	FAILURE	black eyes and the daily g...	marlowe, stephen, 1928-2008	en	23	FAILURE	0.5146
24517.txt	Baily_AccidentalDeath	Sci-Fi	1	SUCCESS	accidental death	baily, peter	en	171	FAILURE	0.5173
29458.txt	Sheckley_CostOfLiving	Sci-Fi	1	SUCCESS	cost of living	sheckley, robert, 1928-2005	en	106	FAILURE	0.5200
30680.txt	Olin_AllDayWednesday	Sci-Fi	1	FAILURE	all day wednesday	olin, richard	en	24	SUCCESS	0.5264
30438.txt	Mckimmey_TheEyesHaveIt	Sci-Fi	1	FAILURE	the eyes have it	mckimmey, james, 1923-	en	27	SUCCESS	0.5265
29559.txt	Farmer_TheyTwinkledLikeJewels	Sci-Fi	1	SUCCESS	they twinkled like jewels	farmer, philip José, 1918-...	en	285	FAILURE	0.5375
29019.txt	Norton_AllCatsAreGray	Sci-Fi	1	SUCCESS	all cats are gray	norton, andre, 1912-2005	en	376	FAILURE	0.5392
21279.txt	Vonnegut_2BR02B	Sci-Fi	3	SUCCESS	2 b r 0 2 b	vonnegut, kurt, 1922-2007	en	2811	SUCCESS	0.5467
10391.txt	Waterloo_TheWolfsLongHowl	Sci-Fi	1	FAILURE	the wolf's long howl	waterloo, stanley, 1846-1913	en	30	FAILURE	0.5471
26955.txt	Huekels_AdvancedChemistry	Sci-Fi	1	SUCCESS	advanced chemistry	huekels, jack g.	en	137	FAILURE	0.5473
19029.txt	Norton_TheGiftsOfAsti	Sci-Fi	1	SUCCESS	the gifts of asti	norton, andre, 1912-2005	en	144	SUCCESS	0.5502
29548.txt	Sheckley_WarriorRace	Sci-Fi	1	SUCCESS	warrior race	sheckley, robert, 1928-2005	en	108	FAILURE	0.5589
28156.txt	Sharkey_MinorDetail	Sci-Fi	1	FAILURE	minor detail	sharkey, jack, 1931-	en	22	SUCCESS	0.5679
30240.txt	Vonnegut_TheBigTripUpYonder	Sci-Fi	3	SUCCESS	the big trip up yonder	vonnegut, kurt, 1922-2007	en	972	SUCCESS	0.5851
30199.txt	Derleth_McIlvaineSStar	Sci-Fi	1	SUCCESS	mcilvaine's star	derleth, august william, 1...	en	196	SUCCESS	0.5928
30382.txt	Coggins_SayHelloForMe	Sci-Fi	1	FAILURE	say "hello" for me	coggins, frank w.	en	12	FAILURE	0.6021
18342.txt	Piper_TheAnswer	Sci-Fi	1	SUCCESS	the answer	piper, h. beam, 1904-1964	en	131	SUCCESS	0.6063
30528.txt	Horne_TheDayOfTheDog	Sci-Fi	1	FAILURE	the day of the dog	horne, anderson	en	21	FAILURE	0.6289
30386.txt	Marlowe_TheOneAndTheMany	Sci-Fi	1	FAILURE	the one and the many	marlowe, stephen, 1928-2008	en	23	FAILURE	0.6361

Figure 44: Title-wise break down of probability values.

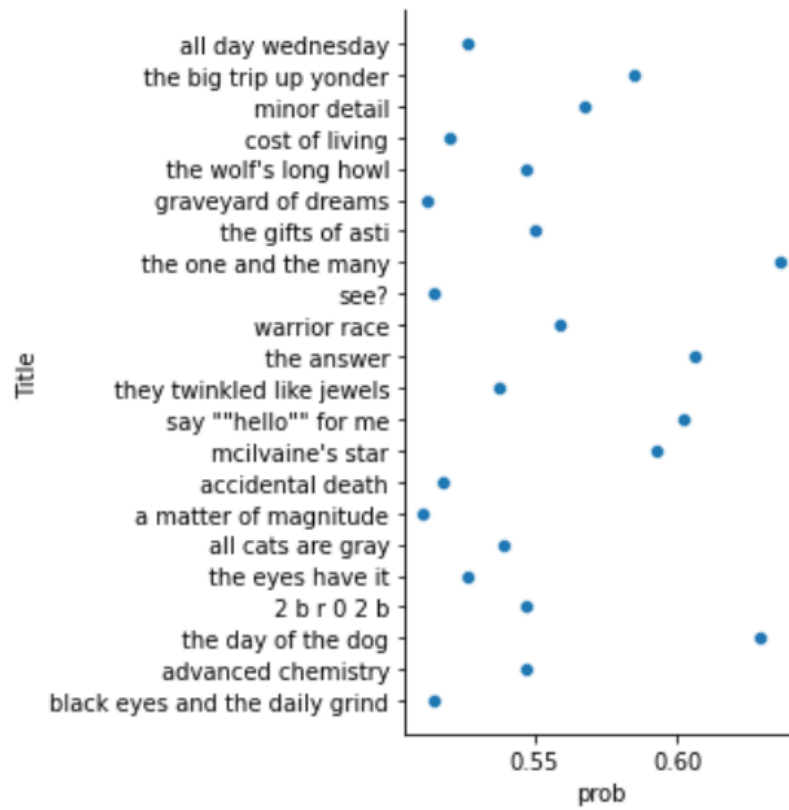


Figure 45: Scatter plot of probability values of science fiction titles.

FAILURE			SUCCESS		neutral	
word	score	word	score	word	weight	
1	girl	-0.0361	mr	0.0356	how it	1.4028e-05
2	miss	-0.0203	she had	0.0240	left it	3.3228e-05
3	she said	-0.0157	re	0.0205	kissed	-4.5259e-05
4	boys	-0.0156	thou	0.0201	and the old	4.5884e-05
5	exclaimed	-0.0149	sir	0.0198	up by	4.6189e-05
6	as she	-0.0110	de	0.0171	me like	-4.9817e-05
7	the young	-0.0106	wind	0.0150	was silent	5.0203e-05
8	sweet	-0.0104	letter	0.0132	she would not	5.1019e-05
9	the girl	-0.0103	that she	0.0100	well to	-7.2602e-05
10	tom	-0.0094	sea	0.0098	into that	7.7401e-05
11	chapter	-0.0088	window	0.0097	hand upon	7.9259e-05
12	doctor	-0.0087	lady	0.0095	ll take	-7.9640e-05
13	laughed	-0.0072	person	0.0091	so often	-8.4637e-05
14	er	-0.0070	mrs	0.0087	was heard	-1.0645e-04
15	jack	-0.0067	thus	0.0083	all for	1.1720e-04
16	uncle	-0.0057	looked at	0.0074	act of	1.2243e-04
17	frank	-0.0056	clock	0.0071	shout	-1.2502e-04
18	party	-0.0055	the house	0.0071	curse	-1.3108e-04
19	safe	-0.0053	idea	0.0069	man whom	-1.3157e-04
20	sister	-0.0051	thee	0.0068	more than one	1.3436e-04

Figure 46: The top twenty words most strongly associated with each label and words with the least weight.