

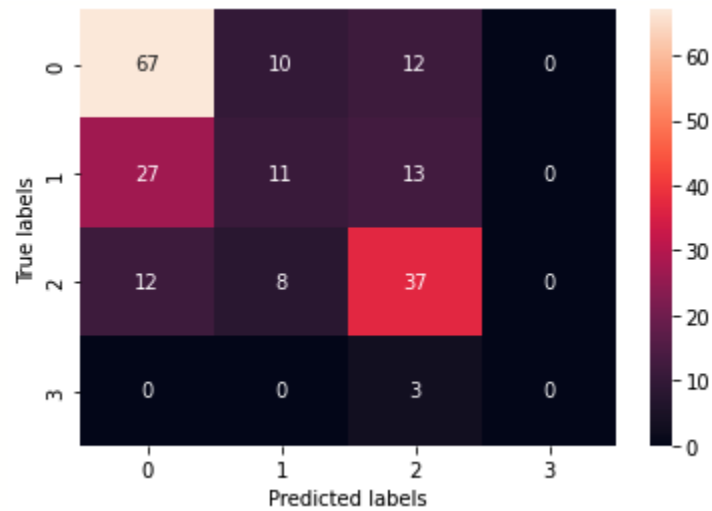
AP4 Analysis Report

Casualness-Formalness Classification

Mingyuan Ma, Shuyao Zhou, Yuanrui Zhu

*** For keeping consistency with the Logistic Regression Jupyter Notebook, we will use Class 0 as Casual, Class 1 as Ordinary, Class 2 as Formal, and Class 3 as Artistic throughout this report. Note that this is one index off with our guidelines. See the notebook for more detailed information.

Part 1: Label Mistaken Identification Using Confusion Matrix & Rethinking Category Boundaries (Logistic Regression Model)



We plotted the confusion matrix comparing each class of actual and predicted test results, and accordingly calculated the **Precision, Recall, and F-1** score for each of the four classes.

- For Class 0 (Casual), precision = 0.632, recall = 0.753, F1 = 0.687
- For Class 1 (Ordinary), precision = 0.379, recall = 0.2, F1 = 0.262
- For Class 2 (Formal), precision = 0.569, recall = 0.649, F1 = 0.607
- For Class 3 (Artistic), precision = recall = F1 = 0

Analysis of statistics:

- For each class, its precision and recall are close. That is, for a class with high precision, its recall is also high. So, clearly, there is a performance gap among the classes.
- The performance of Casual and Formal is much better than Ordinary, which follows the general case that the Ordinary class is a lot more difficult to classify than the other two classes. However, when putting a text to be in Class 1 (Ordinary), we may involve our personal judgment based on our unique educational and multilingual background. Hence, we're not classifying based on the same metrics as machines, and it's normal for algorithms not to capture these factors and performing badly.
- The score of Casual is the highest among all the classes, which is largely due to the apparent feature it demonstrates to readers (as we stated in the guideline).
- We observed that the logistic model tends to classify real Class 1 to Class 0/Class 2 since the precision for Class 1 is greater than its recall, which is the opposite case for Class 0 and Class 2. Given for the recall, real Y is in the denominator position, we can infer **Class 1 has nearly no apparent features that can be relied on.**
- Because of the uneven nature of our dataset, precision and recall are all zero for Class 3, which we will conduct further discussion in the following parts.

From the data set, we can clearly conclude that Class 1 (Ordinary) is often mistaken for Class 0 (Casual). Let's see why this would happen by looking at the example texts:

1. This is a children's book and it's self explanatory rules that you are looking for an exact match. "This" does in fact imply that it is that exact picture you must find in the bigger picture. There are multiple anteaters and you can't just say any one of them counts because they are all anteaters. I've said in other comments as well that I've done tons of these books with my kids and that's how it works. You may find something that looks incredibly similar but if you keep looking you find the exact picture. Every single time. My 5 year old was smart enough to notice it was not the same mongoose but I guess a bunch of Redditors weren't.

2. This was another The Alchemist for me. Everyone in my bookclub loved it and found it deeply insightful. I really wanted to like it but I just found it cliché. I agree with others that the protagonist wasn't relatable... things just kept happening to her. It was a frustrating read.

3. This is probably an unpopular opinion but I feel like r/books focuses an ordinate amount of time ranting about things they don't like in physical books rather than discussing what's in them.

4. I use a combination of Excel and Goodreads.

5. Definitely! Here's my list...
 -only read what you enjoy. If a books not doing it for you put it down. You can always give it another chance later.
 -find a quite dim spot with minimal distractions. I always read in bed at night because that's when everyone else is asleep and won't interrupt me.
 -cut out tv/streaming. Seems weird at first but once you get into a good book tv loses its appeal.
 -get rid of social media.

Especially FB. It's just designed to bait you into arguments anyway. I make exceptions for book related platforms like Reddit or Bookstagram. It's been 3 years and I don't miss a single thing about FB.-buy cheap used books or use the library. That way you can sift through a wide variety of books for minimal monies. Some used book shops have mint condition books for \$4.-join in on the r/52book subreddit. That's what motivated me to start reading more and it's been a lot of fun over the years. I found great recommendations and had some really good interactions with other Redditors -always carry a book with you. You never know when you'll find an extra 15 minutes to read. And all those little chunks really add up. I did that less in 2021 and you can see the above results. Those are my main bookish habits that keep me reading at a high rate. Even at 100+ books it doesn't feel like I'm forcing myself. It's just reading for pleasure. I hope that helps!

6. Maybe try r/52book? I use Goodreads to set my reading goal for the year and track everything. Other than that, I make time to read every day. Usually at night before bed and on lunch breaks. More on the weekends. I read 81 books this year. I was thinking of trying for 100 but that might be a big stretch and I'm afraid I'd try to read only short books, which would limit the books I want to read that are long. I think maybe having a goal based on number of pages read would be neat too. I read less pages in 2021 than I did in 2020, despite that I read two books more in 2021 than I did the previous year.

7. This is why I joined Reddit, /r/booksuggestions, /r/suggestmeabook, /r/52book. After I spent some time learning there, a visit to the library or a real bookstore feels like a trip to Disneyland.

8. I actually just finished reading this one as part of my reading challenge this year [link] and I loved it! And yet I completely agree with you. The lesson/s in it are nothing new or overly enlightening by the time you are in your 30s if you have a modicum of self awareness

9. Truelit and truefilm are good though.

10. Next level reminds me of extreme movie.

(27 results omitted)

There are fewer cases of the opposite, where the prediction is Class 1 when the actual category is

Class 0:

1. It's really too extensive to list. I'll see if I can find some literature on it. There are plenty of options out there, I'm just really surprised at how many people just blindly take all of these supplements (a lot of shitty brands I might add) without knowing whether or not you need it. Everything is synergistic, so if you over do it on one thing, you can throw off a serious of other metabolic processes.

2. I don't think there's a good catch-all term here... though you could check with r/askliterarystudies as it's possible there's something more defined in that world that could be applied here.I've seen people use "Thematic Shift" to discuss the technique you seem to be describing. Though, I think that term applies pretty well to *Hereditary* and less so to something like *Mulholland Drive.*

3. I'll have to try Story Graph.Zlib has two separate features for keep track of books, Favorites and Booklists. They're both great and really well coded. I much prefer the layout of the booklists to the favorites and I'm trying to migrate all my favorites over there, but you have to do it one book at a time and I have 35 pages of them :(

4. AFAIK UK and European PhD applicants typically already have a masters level degree in the field, and you'll be expected to have a thesis project basically ready to go, and you should really have a specific faculty member in mind to work with.

5. The bookclub. There's 10 of them I think? You gotta sleep a few times in-between a few to get the next. So have fun making that run 10x lmao. I did em all in a row and it was draining... Idk if it's an unmarked quest you gotta stumble upon or not.

6. The one with the book glances to you, then asks if your ok.

7. If you look at Jinger's Instagram and scroll way back to May 30th 2017 (about the tenth oldest photo on her Insta) you will find the legendary photo of Jeremy in a library, with the caption, "Jeremy loves books. I love Jeremy. Now I love books. 🥰📚"

8. If you're unfamiliar with Zen Pencils [link], I strongly recommend reading more about those comics and their background - [link]There is a lot of good material on there.

9. Additionally, he beat the shit out of a guy in ring after he found out he was gayTotal lie. Kanyon said in the same book that accuses WWE of having Undertaker hurt him with chairshots that Undertaker was supportive of his struggles.

10. Volume, direction, ema, sma. That's, quite literally, all you need if you are able to think fast and react quickly. Day trading is meant to be in and out with 0.5-3%gains (rarely more), and cutting your losses extremely quickly (within 0-0.5%). On average, I'd say I make about 10-20 trades a day. The first 3/4 are in and out because price action isn't there yet but volume is high and there could possibly be a trend so I try and price in accordingly. I'm a market open and near market close trader because I prefer to play with price action and volume(so 9:30-11am and 2-4pm generally).EMA/SMA for support/resistance levelsVolume, candles, order books, and ticker list to include QQQ, SPY, etc to gauge general direction. Ultimately, you have to be quick to multitask and assess all this information within seconds to get in early and ride a small trend. Otherwise, it's going to be tough.

Looking closely at the misclassification cases, we can tell that they're actually closer to Class 1 than we expected them to be. From the perspective of speakers, the person who says *'Next level reminds me of extreme movies'* and *'Truelit and Truefilm are good though'* may very likely be in a casual state. We may have the wrong sense of those texts at the very beginning since we didn't quite get what the writer really wanted to express.

Beyond that, we would modify our guidelines boundary based on our observations. There're some sentences we would identify as "plain text", such as *"I use a combination of Excel and Goodreads."* We originally defined such "plain text" to be ordinary, but now we would choose to categorize them as Casual rather than Ordinary. We believe that **besides "sentence structure", "sentence content" is another critical measurement of casualness**. If the author explains a preference or a quick fact without any detailed discussion, this should be categorized as casualness of "sentence content", and thus it is casual overall.

Now let's look at other boundaries between Class 1 and Class 2:

True Label - 2, Predicted Label - 1

1. If you read my post you'd see that I had already started one of the books for a book club I joined, so I was definitely not looking for recommendations in that regard. I was looking for comments that would inspire good discussion, like the first comment saying that I should definitely read Ready Player One prior to watching the movie.

2. I think like with anything there's good and bad. The issue is there's maybe 70-80% bad books out there peddling not just shitty ideas, but sometimes dangerous ones. Anyone who reads these types of books go through a journey where they may discover better and better books. If you want a life hack I'd say the best starting point to void the fluff is to go read some of the classics like Meditations by Marcus Aurelius and a few other stoic classics. You'll

realise that most modern self help is a rehash of these classics. There's still a few other great ones out there though - I find GoodReads helps weed out a few

3. How does r/TrueLit actually think of J. R. R. Tolkien's work? I thought it's the most bizarre thing to discuss here, so I've been hesitant to ask this. Of course, I'll never put Tolkien as the same level as writers as Pynchon, Faulkner or McCarthy in any way. But whenever I read Edmund Wilson's terrible review rampant with ad hominem and even spelling errors for character names, I grimace a lot.

4. There is a genetic insensitivity to gut THC adsorption, if I understand the literature correctly. Both my mother and I have this insensitivity and have taken both edibles and sublingual tinctures without any effect even at high doses, while others experienced highness from much smaller doses.

5. If this doesn't make sense, please reply with more than a one-word question so I have a better idea of what is confusing you about this.

6. Have a search and ask around at r/AskLiteraryStudies.

7. Audra Simpson's *Mohawk Interruptus* You might also want to ask over at /r/AskLiteraryStudies too.

8. I should have said at the outset that I will conduct this discussion like a university seminar. Questions of belief or appeals to authority or dogma will not enter in. The basis will be historical evidence and analysis and academic literature.

True Label - 1, Predicted Label - 2:

1. To be fair, the only poets he thinks are good are ["Homer, Virgil, Dante, Chaucer, Shakespeare, Milton, Goethe and maybe Dickinson"][\[link\]](#), so it's a pretty high bar (especially since you assume he must have read them in their original languages to make such a pronouncement). [\[He's also 19\]](#)[\[link\]](#), so there is still a little hope that he'll grow out of it.

2. There are cutscenes that came out today in game that are covered in the novel. There are no other spoilers and nothing for future content, it appears.

3. I'm not sure if such a website or program exists and - as a literary scholar by training - I am somewhat skeptical that such an approach would do a good job at teaching the necessary skills to critically engage with literature. I could very much see it run the risk of boiling literary analysis down to nothing but a checklist of actionable steps and removing any kind of ambiguities (and thus potential) from literary texts. While there are certain protocols to follow in literary studies, the craft is much more free form than, say, running a cost-benefit analysis. Instead, it is a matter of being receptive to the nuances of a text - and to bring one's understanding of (literary) theory and history into dialogue with these nuances. (I am, of course, oversimplifying.) Thus, building a personal repertoire of literary theories and a solid understanding of (literary) histories might be one of the best bets to become better at critically engaging with literature. Building this repertoire, in turn, is probably best achieved through reading - and discussing what one has read. When it comes to the former, you can find a few recommendations in the following: If you want to dip your toes into some pop-literary studies first, you might want to try something like Thomas C. Foster's *How to Read Literature Like A Prof*, which [u/pm_me_poems](#) please recommended in a similar context. Another great way to get a first feel for almost any academic field are Oxford University Press's *Very Short Introductions* - in this case Jonathan Culler's *Literary Theory: A Very Short Introduction*. You could also have a look at general introductory textbooks, like Mario Klarer's *An Introduction to Literary Studies*,* or at textbooks focusing specifically on literary theory, such as Jerry Eagleton's *Literary Theory: An Introduction** and Peter Barry's *Beginning Theory: An Introduction to Literary and Cultural Theory**. In case you would prefer to watch a lecture on the topic, there is always [Yale's Open Course on Literary Theory](https://oyc.yale.edu/english/engl-300) (<https://oyc.yale.edu/english/engl-300>). Finally, if you want to dive straight into the deep end, you could have a look at Cambridge University Press's extensive list of *Cambridge Companions to X** and *Cambridge Histories of Y**, where X and Y are something like "the Victorian novel" and "American Literature." I hope some of this helps and I am sorry I couldn't offer any more open-source resources. If you've got any questions, feel free to ask! EDIT: You might also want to check out r/AskLiteraryStudies. :)

4. The subreddit is titled 'AskLiteraryStudies'. That's what the OP did. If you want, start a subreddit called 'Don'tAskResearchQuestionsYouCouldGoogle'. Your indignation when confronted with a similar polite request will then be appropriate.
5. Our kids prefer to own most of their books so they can share and re-read. Libraries have book sales. Try booksalefinder.com. The Mensa website has good booklists broken down by age group. You can also google top 100 kids chapter books of all time and start there.
6. This is a sub for pictures of people's bookshelves. /r/books, /r/bookdiscussion, /r/paranormalromance, and /r/YALit would be better places for this.
7. Just for fun, I gave a relevant prompt to an [outside link] to see what it would come up with:
 ** (Prompt) An open letter by Franz Kafka about the dearth of novelty in /r/TrueLit:
 ** I have just read your statement on the subject of novelty. You are correct, I do not know how to express the thing, because I myself do not know how to express the thing. Yet the thing is the thing. The thing is what the thing is. The thing is what the thing is. I will not insult your intelligence by saying that the thing is the thing. You know it is the thing. I know it is the thing. It is the thing. The thing. The thing. The thing. The thing. The thing. The thing. The thing. The thing. I tried the same prompt again, just with a different writer.
 ** (Prompt) An open letter by Fyodor Dostoevsky about the dearth of novelty in /r/TrueLit:
 ** We have been reading and rereading the article in which you have described the monotonous state of your life and the "dearth of novelty" in it. It has a "sad and gloomy" look. Your heart aches as you think of it. But the article is not gloomy, not sad. It is, rather, quite the opposite. The last one I tried was strangely ... sensible. Not what I was expecting.
8. Who's we? He obviously wanted to, and since it's his novel, I don't see why not. Also, this is truelit, I come here for in-depth literature discussion, not to be reminded of reddit's tendency to review books they don't plan on reading.
9. Check out the weekly suggestions thread (should be stickied), the mod recommendation in the sidebar, /r/booksuggestions, or /r/whattoreadwhen
10. This is my favorite book list: [link]
11. I know that you're specifically asking for the opinion of a _historian_, but your question lies at the intersection of literature and history. An academic trained in language and literature might find some parts of your question easier to weigh in on with authority. You might benefit from posing this to r/AskLiteraryStudies, too.
12. Written in rich, seductive prose, this novel contains senses and the experience.
13. There's also the [weekly recommendation thread] (http://www.reddit.com/r/books/comments/2mgd3p/weekly_recommendation_thread_for_the_week_of/) and the various recommendation subs pulled from this [List of book related subs] (<http://www.reddit.com/r/BookRelatedSubs/wiki/sublist>)

The boundary between Class 1 (ordinary) and Class 2 (formal) is vaguer than the boundary between casual and ordinary since there are no apparent signs/features for formality that can make us believe that one document is formal. Unlike Class 0 we have some clear signs, we need to consider a document holistically and apply our prior experience which is subjective and different among people.

We can still find some cases of misclassification. **The first observation we have from misclassification is that several words (which appear in an informal document frequently) may lead a formal document to be classified as informal.** As we mentioned above, we need to view the document holistically, including words, tone, emotion, etc. rather than basing it on a

particular word. **However, we can see that the machine may classify one document based on some typical words.** It is not wrong, but it is restricted by some edge cases: for example, when the majority of one document is not informal from the machine's perspective, but as long as it sees an informal word, it would classify the document as informal. **Thus, it is crucial for viewing the document holistically.** For example, given the sentence "*Audra Simpson's Mohawk Interrupts you might also want to ask over at r/AskLiteraryStudies too*", it should be classified as 2 since the author is giving advice from its content, but in a likely scenario, the word "*too*" presenting in the document leads it to be classified as 1.

The second observation from misclassification is that **we feel the model ignores the dependency or correlation between one word (token) with the whole document**; since we use unigram, bigram, and trigram (word combinations) as features, we cannot capture one combination's information perfectly. Since a word pair (maybe 0, 1, or 2) may still have correlation or dependency with the whole document. However, if we are using finite gram, we may not be able to capture it or in some cases may overfit the model. Hence, we can notice that "70-80%" may be regarded as a sign to be Class 1 because other documents may use 70-80% in a different way. However, 70-80% is statistics that back up the main idea with evidence and thus should optimally make the document more formal. The machine cannot capture the weightage of each feature (which is hard even for us humans since language is subjective, so there's no doubt that n-gram does worse). For example, "*just for fun*", and "*To be fair*", with other obvious formal/professional words would lead it to be Class 1, but the machine classifies them as Class 2.

The third observation from misclassification is **how to deal with those abbreviated (misspelled) words and the position of the words**. We are analyzing posts from Reddit, and some documents may contain an abbreviation of words, which is a feature of casual; however, those topics are ranging more different books, products, daily language use, so it is hard to include all the potential abbreviations or internet slangs, especially some of them are not common (which possibly belong to a particular topic or book). More importantly, it is hard for us to capture misspelled words since there are uncountable ways to misspell a word. Unfortunately, it is important to capture them as features.

However, the observations from misclassification also help us a lot in clarifying the boundary. For example, some words containing emotions may be captured by the model, which also makes us realize that we can clarify the boundary through different dimensions toward each

document: not only the words themselves and the content, but also the tone and the degree of emotion.

Part 2: Level of Balance

Firstly, we get the percentage of each Class in the training set:

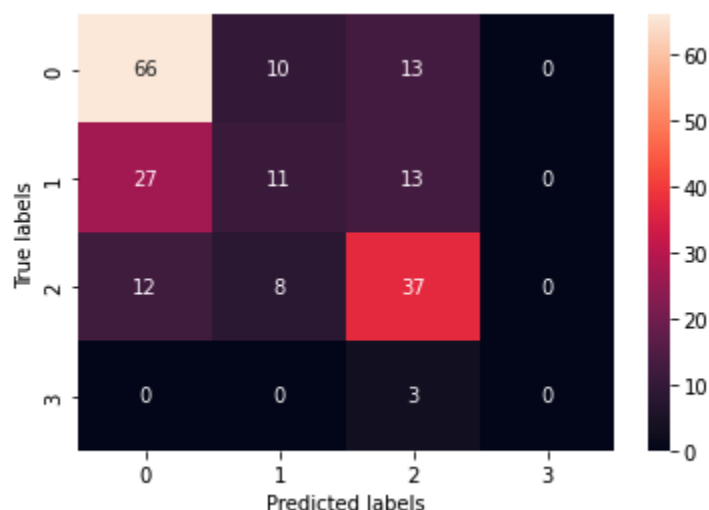
```
Class 0 percentage: 40.67%
Class 1 percentage: 26.83%
Class 2 percentage: 31.33%
Class 3 percentage: 1.17%
```

One flaw of the training set and testing set is the imbalance of our data. Class 3 (Artistic) is rare, which puts some limitations on training, thus pushing the algorithm to ignore the underrepresented class entirely. Also, literary devices are very literal and are hard to learn from a small set of data. This explains why none of the three artistic labels in the test set are classified correctly. We should include more data in the training set to capture the internal patterns and rhetorical devices of the artistic paragraphs.

At this point, one way to deal with this problem is **randomly oversampling Class 3**. We modified the classifier so it over-samples the underrepresented class and makes it approximately equal to the size of the most prevalent class. Now we reach the following accuracy:

```
Training set has 837 data now.
Class 0 percentage: 29.15%
Class 1 percentage: 19.24%
Class 2 percentage: 22.46%
Class 3 percentage: 29.15%
```


Test accuracy for best dev model: 0.570, 95% CIs: [0.501 0.639]



Here is the misclassified Class 3 (Artistic):

1. True love or not, the game must play out, and the fates of everyone involved, from the cast of extraordinary circus performers to the patrons, hang in the balance, suspended as precariously as the daring acrobats overhead.
2. To them good ol days exist in pre industrial natural selection world or in some fictional novel.Worst ideologies are those that promises Utopian world. They roads leads to nowhere and are paved with blood."Nirvana is not for this world"
3. In the story, Kate Daniels is a down-on-her-luck mercenary who makes her living cleaning up these magical problems. But when Kate's guardian is murdered, her quest for justice draws her into a power struggle between two strong factions within Atlanta's magic

However, even though we increase the sampling of underrepresented classes, we haven't seen any changes in the prediction. We would infer that in our case, resampling cannot promote the performance of our classifier. Apparently, the artistic class has no strong feature, and we couldn't simply perform oversampling to enhance the model performance in this specific class. Given we're using an n-gram model, the rhythm and vocabulary of artistic text cannot be well captured. Also, the special meaning of one word under a specific context cannot be modeled, although being an important part of defining the text to be artistic. We may solve this by expanding the n-gram model, but we may have the problem of overfitting. We didn't find it to be useful. Another way that may help is to include more words that often appear in artistic paragraphs, but the versatility of artistic text makes it hard for us to do so.

Part 3: Features Learned to most define the Class (logistic regression)

The Strongest coefficients for each class are (class-score-feature:

0 0.753 !	1 0.236 if	2 0.233 as	3 0.151 but
0 0.409 i	1 0.176 If you	2 0.228 by	3 0.150 . But
0 0.244 ?	1 0.176 If you	2 0.228 be	3 0.150 . But
0 0.232 ...	1 0.175 r/52book	2 0.222 r/truelit	3 0.146 love
0 0.195 me	1 0.154 literature	2 0.210 's	3 0.142 of
0 0.170 my	1 0.151 bookclub	2 0.194 .	3 0.098 one
0 0.169 :)	1 0.147 maybe	2 0.182 are	3 0.093 dark
0 0.161 'm	1 0.144 found	2 0.179 on	3 0.091 world
0 0.154 with	1 0.140 then	2 0.167 /r/truelit	3 0.087 diaries
0 0.146 na	1 0.139 have	2 0.164 /r/askliterarystudies	3 0.085 her ,

We will explain the strength of some features. For example, "!" / "?" are signs of strong sentiment, and "me" / "my" are signs of personal expression, which as we defined in the guideline: Are indicators of casualness. Furthermore, "as" can be interpreted as "reasoning" and make sense to be a sign of Class 2. More "." in a document indicates that it is a longer text, which would be more likely to have detailed explanations and thus higher probability of formality. Although Class 2 has the worst performance among all the classes, the defining features of Class 2 are still quite intuitive. "Dark", "world", and "love" are general intangible words that aren't usually used in formal expressions.

However, there are plenty of features that we wouldn't intuitively interpret as their strength. "Are" is a singular plural present tense of "be" which seems very prevalent in every kind of text in the real world, but here we only have it as a strong sign of formality.

From the feature above, we also see something we didn't carefully consider when selecting the data. We notice that for Class 1, "book", "literature", and "r/book" are among the top features of predictive power. This means there is a bias that the post we selected from the

r/book thread is full of ordinary posts. Similarly, "r/truelit" and "r/askliterary studies" have a large proportion of formal posts, which makes them the defining feature of Class 2. This phenomenon is against our initial idea of only focusing on the text itself in order to grasp the feature of casualness-formalness. We believe changing the span of posts we analyze (also maybe removing any tags that exist in the post) can solve this issue to some extent.

We also tried to run data on BERT and made some observations:

*** For keeping consistency with the BERT Jupyter Notebook, we will use Class 1 as Casual, Class 2 as Ordinary, Class 0 as Formal, and Class 3 as Artistic throughout this part. Note that this is different from our guidelines.

Not satisfying with the result from logistic regression with self-defined feature above, we want to see if the deep learning model, BERT, considering the dependency of each words in one document and view the document as a whole, would perform better than the logistic regression which is feature-based using n-grams. Due to the GPU capacity issue of Google Colab, we changed the BERT model to "google/bert_uncased_L-2_H-128_A-2". After 25 Epoch, BERT achieved the test accuracy for best dev model: 0.610 with 95% Confidence Interval [0.542 0.678]. We can find that BERT performs better than the previous logistic regression. Although our self-defined features in the logistic regression model are the optimal ones after doing several trials, the BERT model can perform better than the logistic regression. This outstanding performance makes us understand the importance of viewing the document holistically when doing subjective judgment tasks. We should not emphasize certain words or certain pairs of words. Instead, as humans, we judge a sentence after reading it from start to the end. BoW or other n-grams cannot capture such nuance in sentences; otherwise, as what we mentioned above, it may overfit the data. So it is not surprising that BERT performs better: it works more similarly to how human perform classification tasks than n-grams. Although there is no feature in the deep learning model, BERT, we still want to delve into those misclassifications to see if we can observe any pattern.

We then generated a dictionary (key is the text, value is the (test_label, our_label) tuple pair) to compare the discrepancy between the test labels and our labels. Most of the differences are (1, 2) and (2, 0). We can know that Class 1 (Casual) is often mistaken for Class 2 (Ordinary). Also, Class 2 is often mistaken for Class 0 (Formal). This is expected because the boundaries between Class 2 and other classes were not strict. After examining the texts, we can tell that BERT has some biases. BERT would surprisingly classify some documents incorrectly.

For example, *"Always happy to hear more people chime in! I know a big thing for all the contributors is knowing that people are actually reading and using posts/booklists/work. A little thank you goes a long way! Always happy to hear suggestions too. *Generally* we try to keep the bulk of recommendations from flaired members because it gives us a good idea of who's expertise is suggesting it. Not *super* sure what other avenues for volunteer work there is, other than as your using it let us know if you come across any problems. Broken links, typos, stuff like that.\n"*. BERT classifies it as Class 2 whereas we classified it as Class 1. It is surprising that BERT would misclassify this text. Obviously it would be Class 1 as there we can find informal words and exclamations which show the casual emotions. Such a basic classification task would be successfully done by a simple model, such as logistic regression whereas BERT may fail classifying it. Another phenomenon that BERT fails to consider is the group of internet slangs. This is also expected because internet slangs update very fast and hard to predict one from others. Also, BERT may not capture it correctly as a sign/feature of one class. However, we still believe that BERT would perform better if we have a more complex structure to capture those details and gain more useful information from documents.