Using natural language processing and machine learning algorithm for book categorization

Danrui Wang[1], Bowen Tan[2,6], Muchen Wei[3], Xuhao Cui[4] and Xingru Huang[5]

[1]University of California San Diego, La Jolla, California, 92093, USA
[2]University of Science and Technology Beijing, Beijing, 100083, China
[3]The Ohio State University, Columbus, Ohio 43210, USA
[4]North Carolina State University, Raleigh, NC 27695, USA
[5]North Carolina State University, Raleigh, NC 27695, USA

[6]tambowen2001@gmail.com

Abstract. This research analyzed the relationship between multiple elements of a book's classification through natural language processing and machine learning. This paper used SVM and KNN to classify books according to the titles and author respectively. Also, books are categorized by summary through Decision Tree, Naïve Bayes and BERT. In the end, this work compared effects of these methods. Our findings show that 1) books have different levels of categorical characteristics in various parts of the book 2) through combining the title, author, and summary of the book, more accurate classification results were obtained 3) BERT achieved more accurate recognition compared to a variety of other algorithms used.

Keywords: natural language processing, machine learning, book classification

1. Introduction

Literature, as an art form, can be traced back to ancient Mesopotamia [1]. Literature is not only a combination of words but also one of the important carriers for recording social forms and historical changes [2,3]. Literature as an expression with subjective factors like other art forms, the genre classification of an article cannot be summarized with a few keywords as it is also subjective. Stylistics, as a branch of applied linguistics, studies various styles of literary genres [4]. Literature has four main genres, fiction, non-fiction, poetry, and drama. The genres cover the basic styles of the literature, and the subgenres and cross-genres can clarify specific styles [5]. The term book is a general term that refers to works of literature. Books have three parts: front matter, body matter, and back matter. The front matter has a title page containing the title of the book, the subtitle, the author or authors, and the publisher. The title means the genre or style of the book, which gives the reader expectation. The book's genre is important because readers often know what they are looking for ahead of time and are browsing for something in particular through the genres. However, book categorization has always been very tedious and complicated. The genres of books are usually identified by publishers, booksellers, critics, and readers, where they would have to read the entire book to properly categorize the genre of the book [6]. Sometimes one literature could fit many different genres, and the genre itself might have different definitions under different cultures [7]. Even though genre classification is very subjective, as it is based on the philosophy, understanding, and ideology of the one who categorizes, and due to the abstract nature, art is difficult to categorize [5,8]. The genre is a very good way to quickly filter for marketing purposes and for readers to choose their books. Thus, this paper would like to try to use natural language processing (NLP) and machine learning (ML) algorithms to help identify the genres of the books, which might both improve the efficiency and objectivity of the classification work. Also, this technique could be used to sort unmarked text collections.

As the carrier of information, words have helped human beings transmit information for thousands of years. In the information age, the emergence of natural language processing has come up with new ways for the collection, sorting, and dissemination of information. Natural language processing can analyze the structure of the sentence and the order of the words in the sentence, or it can be based on the grammatical category of the words instead of their meaning [9]. Alhuqail used BERT, Latent Semantic Analysis (LSA), Bag of Words (BOW), and other natural language processing methods to infer and identify the authors of anonymous articles, and believes that merged features will get more accurate results than individuals [10]. Ferrario and Naegelin used a variety of machine learning methods such as NLP pipelines to classify text documents [11]. Krishna et al. has used the method of removing the special characters and the punctuations, stop-words, and then converting the original words into root words (stemming), in order to obtain a more accurate NLP model [12]. Sel and Hanbay used the Word2Vec algorithm and k-means algorithm to detect the subject of the email and judged the classification of the email successfully [13].

NLP is good at text clarification by automatically analyzing text and assigning predefined categories based on the context [14]. Yoon Kim first used neural network implementations to classify text [15]. Recently, the hierarchy of long short term memory (LSTM) and bidirectional Gated Neural Networks are also used to find the natural structure of language for sentiment classifications [16]. Liu, Loh, and Sun have used Naive Bayes and support vector machine (SVM) and compared different NLP representations, such as bag of words (BOW) and custom weighting schemes, to classify text [17]. Random forests and KNN are also widely used in text classification [18]. Santini has found ways to classify unlabeled text from the web to group unknown web pages [8]. This was performed by using a very flexible genre classification scheme, stating that each web page could have zero, one, or multi genres. Santini's clarifies that the definition of the genre could affect the classification result significantly, and the experiments get impediments because the genre itself is an art that is defined by an individual [8]. Jordan has done thorough research on the complexity of genre prediction and the overlap of different genres [19]. Worsham and Kalita used the whole text of the book to make genre classification [20]. Chiang, Ge, and Wu used the cover image to make predictions on book genres and they found this image-based transfer learning has similar precision as the title-based NLP approach [21]. Koppel showed that automated text categorization techniques can explore combinations of easy lexical and grammatical characteristics to speculate the gender of the author of an invisible written document with roughly 80% accuracy [22].

In this paper, it will utilize three models including Decision Tree, Bert, and Naive Bayes to classify books by their summaries respectively. Then, this work will compare three individual prediction results and obtain the one which shows up most often and is called genre 1. In addition, SVM and KNN will be used to classify books by their titles and by their authors, then this work will get genre 2 and genre 3. Finally, this work will get three prediction results and choose the one which shows up most often among them to label the books.

2. Related works

There are many attempts at categorizing books by chapters, titles, even the covers of the book. Worsham and Kalita use different methods in NLP like CNN and LSTM to classify books by the chapters of every single book [20]. Although they compared these main algorithms' ability in classifying books, they were still working on how to raise the accuracy in classifying books by chapters instead of changing the way to classify books. In light of this paper, this work decided to search whether different ways of classification will influence model performance.

Later, Ganeshprasad et al. utilized titles and covers of the books to classify them [23]. They used the logistic regression model to train and predict the genres of the books. Then they obtained an accuracy up to 87.2% when combining titles and the covers of books, which is better than only using the title features or cover image features. However, when combining images and titles, a small 1% accuracy improvement was achieved. This highlighted that these titles have the highest propensity in classification. According to this paper, this paper intends to use multiple algorithms in machine learning to classify books by titles, authors, and summaries.

Naive Bayes is a popular algorithm in classification, which yields good results when the independence between attributes is not obvious. Zhang et al. demonstrated the strong competitiveness and advantages of Naive Bayes in processing classification [24]. The characteristics of Naive Bayes enable us to analyze the summary in the book. Naive Bayes can also maintain high efficiency and high classification accuracy while processing a large number of data sets [25].

Decision Tree (DT) is a widely used model in classification, which was proposed by Hunt. Patil and Umakant concluded that the DT algorithm possesses a high accuracy and performs well on small to medium sized numerical and nominal datasets [26]. At the same time, the algorithm is easy to explain by graph [27]. So, this paper intends to use it to predict the books' genres.

Support Vector Machine is an algorithm used for classification. Vladimir and Alexey invented the original algorithm of SVM in 1963. Bernhard Boser, Isabelle Guyon, and Vladimir applied the kernel trick to maximum-margin hyperplanes to create nonlinear classifiers. Corinna Cortes and Vapnik published the "soft margin" incarnation as software packages that are often used [28,29].

Over the years, developing a universal representation of text has been fundamental to NLP. The development of pre-trained text attachments like word2vec and GloVe was a big breakthrough in this area. Even though supervised learning has generally shown better results than unsupervised learning within NLP, unsupervised learning is more widely adopted as it can analyze the text without complex preparation and learn a larger corpora of texts [30]. BERT was introduced in 2019 as a new pre-trained text attachment model based on the text from Wikipedia [30]. BERT was designed to help the computers to understand the meaning of the text based on the context of the surrounding text. The algorithm is based on Transformers, which is a deep learning model that connects every output element to every input element; the weightings between them are dynamically calculated based on their connection [31]. The property is given by Transformers also makes BERT capable of reading the text from both left-to-right and right-to-left directions, and this bidirectional property helps BERT pre-trained on both Masked Language Modeling (MLM) and Next Sentence Prediction [31]. MLM is able to predict the hidden word based on the context, and Next Sentence Prediction is used to check if there is a logical and sequential relationship between two sentences [31].

3. Dataset

This paper uses the CMU Book Summary Dataset collected by David Bamman and Noah Smith (2013). The dataset contains 16,559 books, and each book has 7 attributes, including Wikipedia ID, Freebase ID, Book title, Book author, Publication date, genres, and summaries which are extracted from Wikipedia and Freebase metadata. The basic descriptions of each of these attributes are illustrated in the first column in Table 1. Since this work only considers Book Title, Book Author, genre, and plot summary in this research, it provides statistics and more information under these 4 categories in Table 1. The second column in Table 1 gives an example of how George Orwell's Animal Farm is illustrated in the original dataset.

Table 1. Data attributes description.

Attributes	**Description**	**Sample**
Wikipedia ID	ID of the data stored in wikidata, which is an open and collaborative database.	620
Freebase ID	Identifier for a page in the Freebase database. The format is "/m/0" followed by 2 to 7 characters.	/m/0hhy
Book Title	Name of the book, which is usually chosen by the author.	Animal Farm
Book Author	Name of the primary author of the book. The dataset includes 3101 authors.	George Orwell
Publication date	The date on which a book is published.	1945-08-17
Genres	The style or category of the book. The dataset contains 227 genres in total.	["Roman", "Satire", "Children's Literature", "Speculative fiction", "Fiction"]
Summary	A brief description of the content and main points of the book. The plot summaries of this dataset were extracted from the November 2, 2012 dump of English-language Wikipedia.	Old Major, the old boar on the Manor Farm, calls the animals on the farm for a meeting, where he compares the humans to parasites and teaches the animals a revolutionary song, 'Beasts of England'. When Major dies, two young pigs, Snowball and Napoleon, assume command and turn his dream into a philosophy. The animals…[continue]

4. Methodology

4.1. General outline

A human typically would look at a book's title to determine the genre. For example, when we read a book named A Wizard of Earthsea we would probably assume it is Children's literature or fantasy fiction. Then, we might check the author, because as for authors like Agatha Christie and Conan Doyle, their books most likely fall under the detective genre. If we still cannot distinguish the genre, we might check the brief introduction of the book. In this paper, we are going to classify books in the same order, which is the title, author, and summary (Figure 1).

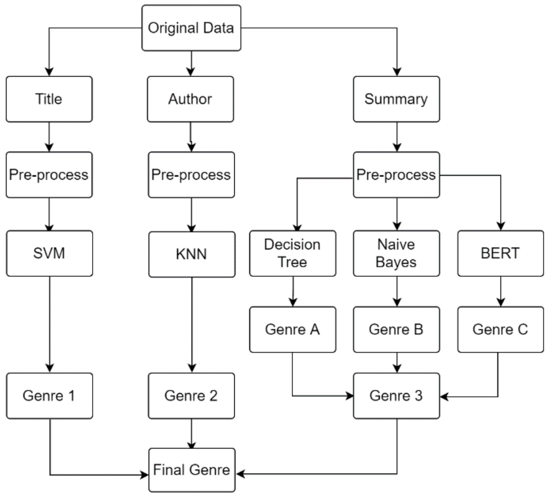

Figure 1. Research Process.
Note: This is the overall research process flow diagram used in this study.)

For each section, this work pre-processes the data, and it makes one genre prediction for each book. SVM is used in the title section to make predictions, and the precision is marked by score 1. KNN is used to predict based on authors, whose precision is marked by score 2. In the summary section, this paper is going to use three different models to make predictions, including Decision Tree, BERT, and Naive Bayes. this work first checks the accuracies of each of the models, which are marked by Score 3, Score 4, and Score 5. It takes a majority vote over the results given by Decision Tree, BERT, and Naive Bayes. For example, if the predictions are adventure novels, horror, and horror, then the final answer would be horror. If all three predictions are different, the final answer would be the one predicted by the model that gives the highest accuracy score. Throughout this process, all scores are compared to see whether title, author, or summary is better to be used for genre prediction, which methods (Decision Tree, BERT, and Naive Bayes) is a better model for book classification based on the summary, and whether taking the mode of the answers given by the title, author and summary would increase the accuracy.

4.2. Pre-processing

This paper first takes the genres whose frequencies are over 500 in the original dataset, which results in 11 genres. Since this work wants to do single-label text classification whereas almost all books have more than one genre, it removes the redundant genres by taking the first genre that appears in the genre list as the target to ensure that each book has a defined classification label. After cleaning, 10,871 books and 8 genres are obtained in the experiment (Figure 2).

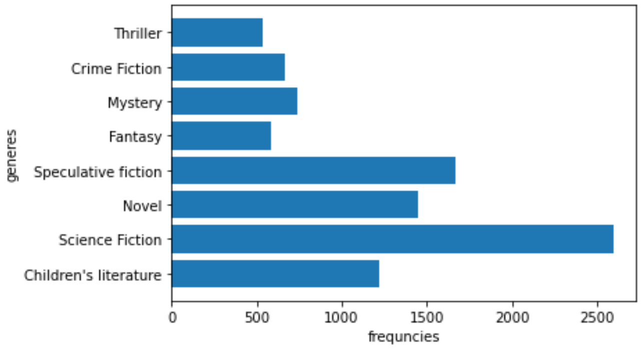

Figure 2. Eight Genres after data cleaning.

In the summary section, this work uses Python to delete the books whose genres don't belong to the 8 genres. Then it uses Natural Language Toolkit to remove the stop words and stem every word in the summary. Finally, this paper uses Python to remove duplicate stemmed words in it and use TF-IDF to get the matrix which it uses to train our model (Figure 3)

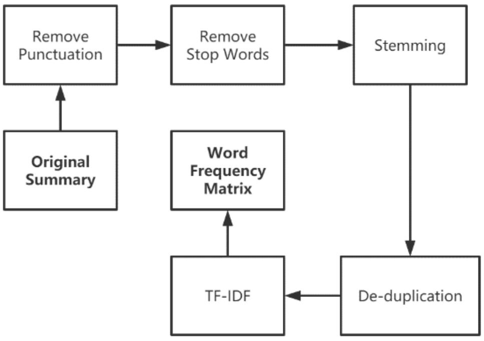

Figure 3. Summary data processing flow.
Note: This diagram shows the flow of Pre-processing.

4.3. Model

4.3.1. SVM. In this study, SVM is used to classify the titles of books. After encoding non-numeric attributes, the titles become vectors. Next, some vectors Xi are classified into a training set.

$$A\ (x_1, y_1), \ldots., (x_\mathrm{n}, x_\mathrm{n}) \tag{1}$$

In the SVM model, the goal is to find the best hyperplane that separates the data. The processed training set is linearly separable, which means that a pair (w, b) can be found. The best hyperplane is between 1 and -1.

$$W^T x_i - b \leq -1, \ \mathrm{if} y_i = -1 \tag{2}$$

$$W^T x_i - b \geq 1, \ \mathrm{if} y_i = 1 \tag{3}$$

Through the model, the best hyperplane (Formula 1) and decision function (Formula 2) can be attained.

$$w^* \cdot x + b^* = 0 \tag{4}$$

$$f(x) = sign(w^* \cdot x + b^*) \tag{5}$$

Finally, the accuracy rate of predicting the genres of books is generated by using the best hyperplane and decision function.

4.3.2. KNN. In this research, it classifies the genres of the books based on KNN. In this study, it chose each author as a point and chose one of the author's books as the points around the author.

$$A\ (x_1, y_1), B\ (x_2, y_2) \tag{6}$$

The distance is calculated between the two points to find the distance between them.

$$d_{A,B} = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2} \tag{7}$$

Since it needs multi-dimensional, each point is given many variables to make it more accurate.

$$A\ (x_1, x_2, \ldots, x_\mathrm{n}), B\ (y_1, y_2, \ldots, x_\mathrm{n}) \tag{8}$$

For the KNN, this work is practiced and tested. The program would calculate the distance between the author and his books separately. As a result, the program uses the shortest distance to determine the genre.

$$d_{A,B} = \sqrt{(y_1 - x_1)^2 + (y_2 - x_2)^2 + \cdots + (y_n - x_n)^2} \tag{9}$$

4.3.3. Decision tree. Decision Tree (DT) is a basic classification method, and here this paper focuses on DT for classification. In the process of classification of DT, it signifies the process of classifying samples based on features, which is regarded as a series of IF-ELSE.

Here this paper uses the ID3 algorithm to split the dataset, and it contains three steps: 1. Feature selection 2. Decision tree generation 3. Decision tree pruning [32].

4.3.3.1. Feature selection. Entropy is defined (Formula 10) as the expected value of the information and is used to measure the purity or impurity of the information. If the entropy is higher, the more impure the dataset is. p(xi) is the probability of the i-th class. n is the number of categories.

$$H = -\sum_{i=1}^{n} p(x_i)\, log_2\, p\,(x_i) \tag{10}$$

Conditional entropy is defined (Formula 11) as the mathematical expectation of the entropy of the conditional probability distribution of Y for a given condition of X.

$$H(Y|X) = \sum_{i=1}^{n} p_i\, H(Y|X = x_i) \tag{11}$$

Information Gain (Formula 12) is related to the features. Therefore, the information gain g (D, A) of feature A on the training dataset D, defined as the difference between the entropy H(D) of the set D and the conditional entropy H(D|A) of D under the given conditions of feature A.

$$g(D, A) = H(D) - H(D|A) \tag{12}$$

Information gain ratio (Formula 13) is the ratio of information gain to entropy.

$$g_R(D, A) = \frac{g(D,A)}{H(D)} \tag{13}$$

The larger the information gain ratio of a feature, the purer the sample it gets after classifying by the feature. After recursively calculating and sifting out all the features, this work can classify all the samples.

4.3.3.2. Decision tree generation. Starting from the root node, a feature of the sample is tested, and the sample is assigned to its child nodes, at which time each child node comparative to a value of the feature, and so on recursively, the sample is tested and assigned until it reaches the leaf node, and finally the sample is assigned to the class of the leaf node. Figure 4 shows a simple decision tree .

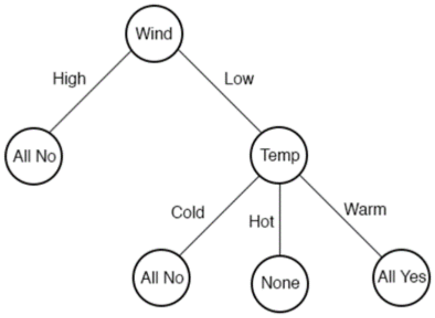

Figure 4. A Simple Decision Tree [26].

4.3.3.3. Decision tree pruning. The decision tree algorithm recursively generates decision trees until it comes to the end node. The resulting trees are often accurate in classifying the training data, but not as accurate in classifying the unknown test data, overfitting can occur. The reason for overfitting is that too much consideration is given to increasing the correct classification of the training data during learning, and thus overly complex decision trees are constructed.

So, the classification tree model is simplified by pruning some subtrees or leaf nodes from the already generated tree and using its root or parent node as the new leaf node. It is implemented by minimizing the loss function of the decision tree to achieve a higher performance as a whole.

4.3.3.4. How this paper uses DT on our dataset. This paper used feature words from each pre-processed text as features for each text, and it finally took 10,000 words as features for each text in order to prevent too many features from degrading the prediction accuracy. This work used 80% of the dataset as our training set and limited the maximum depth of the decision tree from 15-20 to prevent the model from overfitting, and then set the minimum number of leaf nodes to 50 to prevent the model from running too slow.

4.3.4. Naive bayes. Naive Bayes is a kind of generative model. In this study, Naive Bayes was used to judging the most likely classification of books based on the summary of the book. The Naive Bayes algorithm is to establish the joint probability distribution P(XY) between feature X and output Y. When predicting a given feature, use Bayes' theorem to find all possible outputs P(XY), and take the largest one as the prediction result.

$$P(Y|X) = \frac{P(XY)}{P(X)} \tag{14}$$

$$P(x_1, x_2, \cdots, x_n) = P(x_1) \prod_{i=2}^{n} P(x_i | x_1, \cdots, x_{i-1}) \tag{15}$$

Conditional probability & Chain rule is the basic principle of Naive Bayes, which refers to the probability of event A occurring under the condition that event B occurs.

$$P(Y_n|X) = \frac{P(X|Y_n)P(Y_n)}{\sum_n P(X|Y_n)P(Y_n)} \tag{16}$$

P(Y) is the prior probability or marginal probability of Y, and any factor in X is not considered.

P(Y|X) is the conditional probability of Y after X occurs, and it is also called the posterior probability of Y.

P(X|Y) is the conditional probability of X after Y occurs, and is also called the posterior probability of B.

P(X) is the prior probability or marginal probability of X, and is also used as the normalized constant.

$$P(Y = c_k|X) = \frac{P(Y=c_k) \prod_{i=1}^{n} P(x_i|Y=c_k)}{\sum_{j=1}^{K} P(X|Y=c_j)P(Y=c_j)} \tag{17}$$

After introducing a strong assumption that all features are conditionally independent, the principle of the Naive Bayes algorithm can be simplified to Formula 4.

$$argmax_{c_k} = P(Y = c_k) \prod_{i=1}^{n} P(x_i = f_i|Y = c_k) \tag{18}$$

The data features this work needs to predict are eight categories, and it needs to calculate the probability of them belonging to each category. After excluding all the same denominators, compare the maximum value by Formula 18.

4.3.5. BERT. BERT is a bidirectional transformer block connection which uses P $(\mathrm{w_i}|\mathrm{w_1}, \dots, \mathrm{w_n})$as a target function to train language modeling (LM). The encoder unit of the transformer is constructed by multi-head-attention, layer normalization, feedforward, and layer normalization, and this unit makes up every layer of BERT. In this experiment, it uses 12 layers and each layer contains 12 attention mechanisms, so the dimension of the word vector is 768. BERT takes the sum of three embeddings, including token embeddings, segment embeddings, and position embeddings. Token embeddings is a word vector, where the first word in a sentence is a CLS sign which can be used for later classification. Segment embedding is used to classify two sentences, and it is very efficient for distinguishing questioning, answering, and asymmetric sentences. Position embedding encodes the position of the word as a feature vector. The pre-training task includes masked LM and Next Sentence Prediction. In Masked LM, it randomly takes 15% of the token in the training set as the mask target. Within these selected tokens, there are 80% of chances that the mask would replace the token, 10% of chances that a random word would replace the token, and 10% of chances that the token is unchanged. The transformer encoder would not know what words need to be predicted or what words have been changed or masked. Thus, the model would learn the token from both directions of the text. Next sentence prediction is used to learn the relationship between sentences. All sentences in the training set would be selected, where 50% of sentences are the next sentences of a sentence and 50% of sentences are randomly selected. In this experiment, BERT first automatically transforms summary into tokens. Then, this work makes slight adjustments on epochs and the learning rate. Then, the encoded data is transformed into vectors, taken cls-token in BERT as the representation of summary, and use the attention mechanism of the transformer to predict the genre.

4.3.6. Accuracy. The percentage of predictions that are correct.

4.3.7. F1 score. Measure of accuracy calculated from precision and recall. Precision gives the percentage of positive predictions that are correct, and recall measures the proportion of true positive results that are identified by the model. F1 takes the harmonic mean of precision and recall.

5. Result

This paper takes 80% of the data for the training set and 20% for the test set to do genre prediction by SVM, KNN, DT, and NB. As for predictions including BERT, it takes 90% of data for the training set and 10% for the test set because the summaries in the dataset are not very long, BERT needs more data to learn the relationship between words and sentences. Since BERT has the highest accuracy score, the majority vote would take the predictions of BERT if all models give completely different results.

Table 2. Result of five method.

Prediction Basis	**Model**	**Train%**	**Test%**	**Accuracy%**	**F1 score**
Title	SVM	80	20	14.37	0.04
Author	KNN	80	20	34.50	0.13
Summary	Decision Tree (DT)	80	20	42.46.	0.32
Summary	Naive Bayes (NB)	80	20	40.73	0.39
Summary	BERT	90	10	57.80	0.45
Summary	DT+NB+BERT	90	10	48.45	0.41

According to table 2, it’s obvious that book genre prediction based on title with SVM model has the lowest accuracy and f1 score. This might be because most of the books have very abstract titles; only few books like Love Story by Erich Segal have very obvious genre belonging names. Certainly, it is still impressive that titles could reach 14.37% accuracy. This might suggest that the titles of certain book genres have similar patterns. As for genre prediction based on the author by KNN, the accuracy 34.50% is much lower than the result have expected. This might be because most authors write various kinds of books, instead of focusing on one genre. Book classification based on Summary generally has higher accuracy scores. Decision Tree is slightly better than Naive Bayes, and BERT has the highest accuracy overall. This is not unexpected, but this paper does expect BERT to achieve a higher accuracy score. It’s believed BERT would behave much better if more data is possessed or 3 or 4 genres are used. Also, it’s expected a majority vote over DT, NB and BERT would perform much better. However, the accuracy score is actually near to the average of that of DT, NB and BERT. First of all, the combination does not beat BERT alone may suggest that DT and NB cannot be complementary to each other; they do right or wrong together for most of the cases. Also, there are 30.75% of the cases where all three models made wrong predictions, so a majority vote could not raise the accuracy score either.

6. Conclusion

Books and natural language are closely related. NLP provides efficient and objective methods in the processing and understanding of text information. The research direction of natural language processing on classification has made rapid progress in the past decades. This paper aims to study the relationship between multiple elements in a book and the classification of this book through natural language processing. This will help improve the accuracy of book classification and contribute to future predictability. In this research, it used five machine learning classification methods to classify and predict different parts of the book. This paper uses SVM on the title of the book, KNN on the author of the book, and Decision Tree, Naïve Bayes, and BERT on the summary of the book. After comparing the three methods to get the classification orientation of the summary, this work reached the most accurate result by combining it with the classification orientation of the title and the author. Our results show that different parts in books have different levels of categorical characteristics. Our findings suggest that basically book classification based on summary has higher accuracy than that of titles and authors. Title has the lowest accuracy score as most of the books do not have very obvious names that could suggest the genre. Author also has a relatively low accuracy score as most of the authors could write books of more than one genre. Certainly, as for titles and authors, it might use different models to redo and see if it could achieve better performance. As for the summary section, BERT achieved the best and the majority vote did not outperform BERT. Around 30% of times that all BERT, Decision Tree and Naive Bayes made wrong predictions. This may suggest that Decision Tree and Naive Bayes might not be perfect choices for book classification based on summaries. In the future, this work might use more advanced models like Gradient Boosting Decision Tree (GBDT), where each of the trees could learn the residuals of the conclusions and sums of all the previous trees, to do the classification based on summaries. Also, this paper did a single-labeled classification, but most of the books belong to more than one genre. Our models might predict right as one of the genres the book belongs to, but the prediction just didn’t follow the primary genre. In this case, it might use the same models to do multi-labeled classification to compare if it could achieve better results.

References

[1] Black, J., Cunningham, G., Robson, E. and Zólyomi, G. (2006) The literature of ancient Sumer. Oxford University Press, Oxford.

[2] Albrecht, M. C. (1954) The Relationship of Literature and Society. American Journal of Sociology, 59(5), 425–436.

[3] John Tosh. (1984) The Pursuit of History: Aims, Methods and New Directions in the Study of History, 58-59.

[4] Simpson, P. (2004) Stylistics: A resource book for students. Psychology Press, Hove.

[5] Wood, Robert. (2019) How to Figure Out the Genre of Your Book. https://www.standoutbooks.com/what-genre-book-genres/

[6] Strathy, Glen C. (2008) The Genres of Books: 7 Ways to Categorize and Identify Fiction. https://www.how-to-write-a-book-now.com/genres-of-books.html

[7] Chandler,Daniel. (1997) An introduction to genre theory. http://www.aber.ac.uk/media/Documents/intgenre/chandler_genre_theory.pdf

[8] Santini, Marina. (2007) Automatic genre identification: Towards a flexible classification scheme. In: BCS IRSG Symposium: Future Directions in Information Access 2007 (FDIA). Brighton. pp. 1-6

[9] Joseph, S. R., Hlomani, H., Letsholo, K., Kaniwa, F., & Sedimo, K. (2016). Natural language processing: A review. International Journal of Research in Engineering and Applied Sciences, 6: 207-212.

[10] Alhuqail, Noura Khalid. (2021). Author Identification Based on NLP. European Journal of Computer Science and Information Technology, 9: 1-26.

[11] Ferrario, Andrea and Naegelin, Mara. (2020). The Art of Natural Language Processing: Classical, Modern and Contemporary Approaches to Text Document Classification. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3547887

[12] Krishna, Akshay and Aich, Animikh and V, Akhilesh and Hegde, Chetana. (2018). Analysis of Customer Opinion Using Machine Learning and NLP Techniques. International Journal of Advanced Studies of Scientific Research, 3: 128-132.

[13] Sel, Slhami and Hanbay, Davut. (2019). E-Mail Classification Using Natural Language Processing. In: 27th Signal Processing and Communications Applications Conference (SIU). Sivas, Turkey. pp. 1-4

[14] Neupane, Parlad (2020). Understanding text classification in NLP with Movie Review Example. https://www.analyticsvidhya.com/blog/2020/12/understanding-text-classification-in-nlp-with-movie-review-example-example/

[15] Kim, Yoon. (2014). Convolutional neural networks for sentence classification. https://arxiv.org/abs/1408.5882

[16] Tang, Duyu; Qin, Bing; Liu, Ting; (2015) Document modeling with gated recurrent neural network for sentiment classification. In: Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, Lisbon, Portugal. pp. 1422–1432,

[17] Liu, Ying; Loh, Hantong; Sun, Aixin. (2009). Imbalanced text classification: A term weighting approach. Expert systems with Applications. Expert Systems with Applications, 36: 690-701.

[18] Xu, Baoxun; Guo, Xiufeng; Ye, Yunming; Cheng, Jiefeng. (2012). An improved random forest classifier for text categorization. J. Comput, 7.12: 2913-2920.

[19] Jordan, Emily. (2012). AUTOMATED GENRE CLASSIFICATION IN LITERATURE. http://hdl.handle.net/2097/17578

[20] Joseph Worsham and Jugal Kalita (2018). Genre Identification and the Compositional Effect of Genre in Literature. In: Proceedings of the 27th international conference on computational linguistics. Santa Fe, New Mexico, USA. pp. 1963-1973.

[21] Chiang, Holly; Ge, Yifan; Wu, Connie. (2015). Classification of Book Genres By Cover and Title. https://www.semanticscholar.org/paper/Classification-of-Book-Genres-By-Cover-and-Title-Chiang-Ge/d0d0096d307a6da1332153b9cb8a72c29df38f87#citing-papers

[22] Koppel, M., Argamon, S., Shinobi, A, R. (2002). Automatically Categorizing Written Texts by Author Gender. Literary and linguistic computing, 17(4): 401-412.

[23] Ganeshprasad R Biradar, Raagini JM, Aravind Varier and Manisha Sudhir (2019) Classification of Book Genres using Book Cover and Title. In: 2019 IEEE International Conference on Intelligent Systems and Green Technology (ICISGT). Visakhapatnam, India. pp. 72-723.

[24] Kang, D. K., Silvescu, A., Zhang, J., & Honavar, V. (2004). Generation of attribute value taxonomies from data and their use in data-driven construction of accurate and compact naive bayes classifiers. In: Proceedings of the ECML/PKDD Workshop on Knowledge Discovery and Ontologies. Pisa, Italy.

[25] Belik, I. (2018). A Comparative Analysis of the Neural Network and Naïve Bayes Classifiers. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3371889

[26] Patil, S., & Kulkarni, U. (2019). Accuracy prediction for distributed decision tree using machine learning approach. In: 2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI). Tirunelveli, India. pp. 1365-1371.

[27] L. Breiman, J. Friedman, R. Olshen, and C. Stone. (1984) Classification and Regression Trees. Routledge, New York.

[28] Boser, B. E., Guyon, I. M., & Vapnik, V. N. (1992, July). A training algorithm for optimal margin classifiers. In: Proceedings of the fifth annual workshop on Computational learning theory. Pennsylvania, Pittsburgh, USA. pp. 144-152.

[29] Cortes, Corinna; Vapnik, Vladimir N. (1995). Support-vector networks. Machine Learning.

CiteSeerX, 20.3: 273–297.

[30] Koroteev, M. V. (2021). BERT: a review of applications in natural language processing and understanding. arXiv preprint arXiv:2103.11943.

[31] Lutkevich, Ben. (2020), BERT language model. https://www.techtarget.com/searchenterpriseai/definition/language-modeling

[32] QUINLAN, J.. Ross .(1986). Induction of decision trees. Machine learning, 1.1: 81-106.