Jurnal Ilmiah KURSOR
Menuju Solusi Teknologi Informasi
Vol. 13, No. 2, Desember 2025

ISSN 0216 – 0544
e-ISSN 2301– 6914

Multi-label book genre classification: Comparison of machine learning techniques and problem transformation methods

Eka Mira Novita Subroto[a], Muhammad Faisal[b]

[a, b] Departement of Informatics, Faculty of Science and Mathematics, Universitas Islam Negeri Maulana Malik Ibrahim Malang, Jl. Gajayana, Malang 65144, Indonesia
E-mail: ekamiranovitas@gmail.com, mfaisal@ti.uin-malang.ac.id

Abstract

Books play an essential role in life as a source of knowledge and information. The increasing number of books published makes classification more complex, especially in a multi-label context where a book may belong to more than one genre. Furthermore, automatic classification of book genres is required due to the transition of books to e-book and audiobook formats. This research analyzes the application of machine learning techniques using Support Vector Machine (SVM), Logistic Regression (LR), and Multinomial Naive Bayes (MNB) for multi-label book genre classification by comparing their performance through stemming and unstemming process in text preprocessing with TF-IDF and K-Fold cross-validation (k = 10). In addition, two problem transformation methods, Binary Relevance (BR) and Label Powerset (LP), are evaluated. The results show that SVM combined with stemming outperforms other models across all metrics of accuracy, precision, recall, and F1-score. SVM is effective in handling complex and imbalanced data distributions, resulting in more accurate and consistent predictions. The stemming process positively contributes by reducing word variation and allowing the model to focus on word meanings. Among problem transformation methods, LP yields better results because it can capture relationships between labels more effectively than BR.

Key words: Binary Relevance, Label Powerset, Logistic Regression, Multi-label classification, Multinomial Naïve Bayes, SVM.

INTRODUCTION

Books have a significant role in life as a source of knowledge, education, and entertainment. In addition, books also function as the main media for disseminating information and creative ideas. Perpunas statistical data shows that the number of book titles that received ISBNs from 2019 to 2024 was 693,734 titles [1]. In this digitalization era, printed books have shifted to e-books and audiobooks. The number of publishers turning to e-books increased by 20% in 2019, coupled with the number of e-book apps such as Goodreads, Wattpad, Google Play Books, Kindle, and others [2]. Most universities also switched to audiobooks because they are more affordable and practical. This growth shows that the interest in writing and reading is also increasing.

With the increasing number of books published each year, the classification process is also becoming more difficult because many books have the same title but different genres. Therefore, an appropriate automatic classification is needed to help readers find books according to their interests and increase the efficiency of book recommendation systems both in print and digital. Previous research has examined the classification of novel genres using the Naive Bayes method, resulting in an accuracy of 80.5% [3]. Then, other researchers examined using multi-class SVM and Chi-Square, with the highest accuracy achieved by SVM being 94.58% [4]. However, the above research only categorizes books into one type of genre (single-label), whereas books can often fall into one or more genre categories (multi-label).

Multi-label classification is categorizing a text into several categories [5]. One method that is often used to handle multi-label problems is the problem transformation method [6]. Binary Relevance (BR) and Label Powerset (LP) methods are commonly used problem transformation methods for multi-label classification. BR transforms a multi-label task into several single-label tasks, while LP considers possible combinations of labels.

Several studies have demonstrated the effectiveness of BR and LP in multi-label classification. For example, research that BR combined with CNN into BR-CNN can improve the performance of multi-label models [7]. Another study comparing four problem transformation models, BR, CC, CR, and LP, with SVM, KNN, and Random Forest classification algorithms showed that LP + SVM is the most optimal model [8]. Other researchers also tried to compare BR and LP. The results proved BR is better with a smaller hamming loss of 0.072 [9].

The appropriate machine learning technique is very influential in text classification. Many studies have been conducted to find the best machine learning model for text classification. Research comparing SVM, Naive Bayes, and k-NN in aspect-based gadget sentiment analysis shows that SVM has the highest average accuracy of 96.43% [10]. The SVM method also performs better than Pre-trained Language Models (PLMs) [11]. Furthermore, the Logistic Regression (LR) method showed the highest accuracy performance of 97% compared to Random Forest and k-NN in various comparative studies. [12]. Research related to chatbot development shows that LR achieves the highest accuracy and f1-score results [13]. In addition, research that examines the type of Naive Bayes obtained that Multinomial gets a high result of 98.20% [14]. Other studies also assess that Multinomial Naive Bayes (MNB) is a fast, easy-to-implement, and effective algorithm for categorizing text [15]. The above studies show that SVM, LR, and MNB are suitable methods for text classification.

In addition, data preprocessing significantly affects model performance, one of which is stemming [16]. Stemming can improve accuracy because it reduces the dimensionality of features and unifies word variations with the same meaning, resulting in quality data [17], [18]. Stemming converts words into their basic form by removing suffixes, prefixes, or both.

This research compares SVM, LR, and MNB in classifying multi-label genre books, focusing on the impact of stemming and non-stemming in text preprocessing to identify the most effective classification technique. Then, it evaluates two problem transformation approaches, Binary Relevance (BR) and Label Powerset (LP), to determine the most effective method in multi-label classification. Through this research, we will significantly contribute to multi-label book genre classification and provide new insights that can improve the accuracy and efficiency of book recommendation systems in print and digital formats.

MATERIAL AND METHODS

Research Flowchart

The research flowchart in this study is shown in Fig. 1.

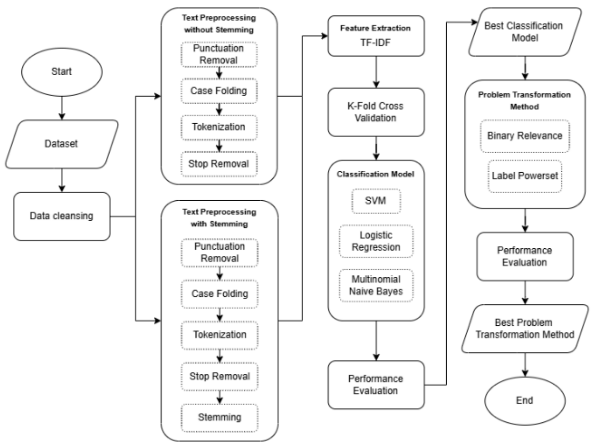

Fig. 1. System design of research

The research begins by collecting data from various relevant sources and cleaning up the data. Next, comparing three classification methods, namely SVM, LR, and MNB using stemming and without stemming with BR approach for handling multi-label classification. TF-IDF is used for feature extraction and K-Fold Cross-Validation is used for model validation. Model performance evaluation uses accuracy, recall, precision, and F1-score metrics to determine the best model. The best model is then used to compare the effectiveness of two problem transformation methods: BR an LP. The evaluation results will determine the most effective problem transformation method for multi-label classification of book genres.

Dataset

The dataset was collected from Goodreads from the list of *Books That Everyone Should Read At Least Once* available at https://www.kaggle.com/datasets/ishikajohari/best-books-10k-multi-genre-data. The dataset consists of 10,000 books grouped into one or more genres based on their descriptions.

Data Cleaning

The data cleansing process includes removing unused columns and only selecting description and genre columns for the book genre classification process. Next, empty and duplicate data were removed. From a total of 617 genres, the top 50 genres were selected based on their frequency of occurrence for multi-label classification. Thus, the total number of book data is 8814 books. The genre distribution can be seen in Fig. 2.

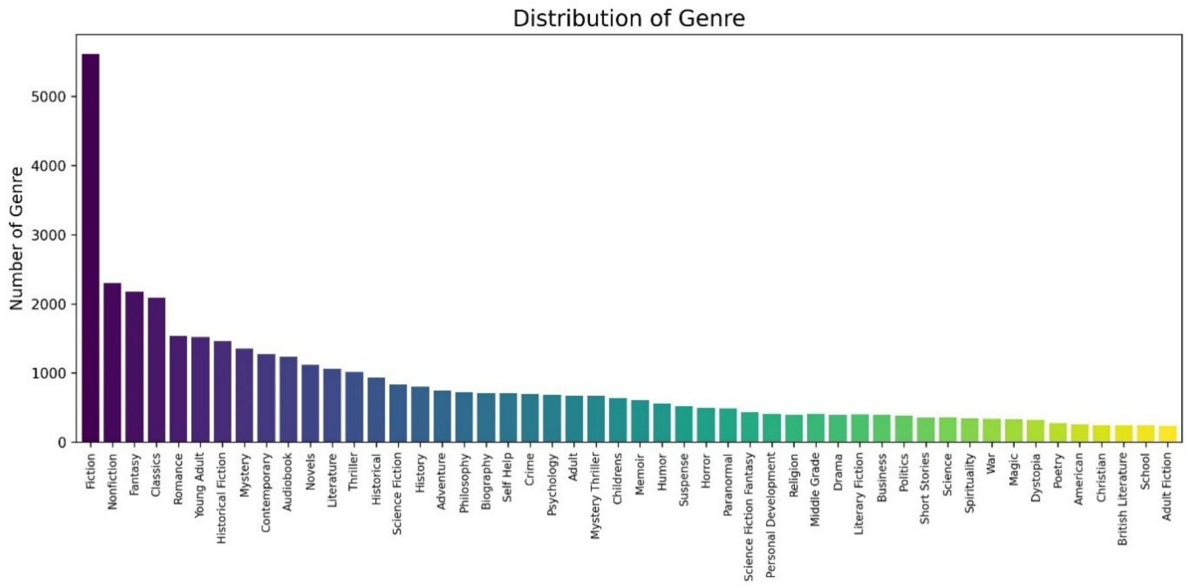

Fig. 2. Number of samples per genre

Table 1. Text preprocessing results

Process	Result
Description	The unforgettable novel of a childhood in a sleepy Southern town and the crisis of conscience that rocked it. "To Kill A Mockingbird" became both an instant bestseller and a critical success when it was first published in 1960. It went on to win the Pulitzer Prize in 1961 and was later made into an Academy Award-winning film, also a classic. Compassionate, dramatic, and deeply moving, "To Kill A Mockingbird" takes readers to the roots of human behavior - to innocence and experience, kindness and cruelty, love and hatred, humor and pathos. Now with over 18 million copies in print and translated into forty languages, this regional story by a young Alabama woman claims universal appeal. Harper Lee always considered her book to be a simple love story. Today it is regarded as a masterpiece of American literature.
Punctuation Removal	The unforgettable novel of a childhood in a sleepy Southern town and the crisis of conscience that rocked it To Kill A Mockingbird became both an instant bestseller and a critical success when it was first published in 1960 It went on to win the Pulitzer Prize in 1961 and was later made into an Academy Award winning film as well as a classic Compassionate dramatic and deeply moving To Kill A Mockingbird takes readers to the roots of human behavior to innocence and experience kindness and cruelty love and hatred humor humor and pathos takes readers to the roots of human behavior to innocence and experience kindness and cruelty love and hatred humor and pathos Now with over 18 million

Process	Result
	copies in print and translated into forty languages this regional story by a young Alabama woman claims universal appeal Harper Lee always considered her book to be a simple love story Today it is regarded as a masterpiece of American literature
Case Folding	the unforgettable novel of a childhood in a sleepy southern town and the crisis of conscience that rocked it to kill a mockingbird became both an instant bestseller and a critical success when it was first published in 1960 it went on to win the pulitzer prize in 1961 and was later made into an academy award winning film also a classic compassionate dramatic and deeply moving to kill a mockingbird takes readers to the roots of human behavior to innocence and experience kindness and cruelty love and hatred humor and pathos now with over 18 million copies in print and translated into forty languages this regional story by a young alabama woman claims universal appeal harper lee always considered her book to be a simple love story today it is regarded as a masterpiece of american literature
Tokenization	['the', 'unforgettable', 'novel', 'of', 'a', 'childhood', 'in', 'a', 'sleepy', 'southern', 'town', 'and', 'the', 'crisis', 'of', 'conscience', 'that', 'rocked', 'it', 'to', 'kill', 'a', 'mockingbird', 'became', 'both', 'an', 'instant', 'bestseller', 'and', 'a', 'critical', 'success', 'when', 'it', 'was', 'first', 'published', 'in', '1960', 'it', 'went', 'on', 'to', 'won', 'the', 'pulitzer', 'prize', 'in', '1961', 'and', 'was', 'later', 'made', 'into', 'an', 'academy', 'awardwinning', 'movie', 'also', 'a', 'classiccompassionate', 'dramatic', 'and', 'deeply', 'moving', 'to', 'kill', 'a', 'mockingbird', 'takes', 'readers', 'to', 'the', 'roots', 'of', 'human', 'behavior', 'to', 'innocence', 'and', 'experience', 'kindness', 'and', 'cruelty', 'love', 'and', 'hatred', 'humor', 'and', 'pathos', 'now', 'with', 'over', '18', 'million', 'copies', 'in', 'print', 'and', 'translated', 'into', 'forty', 'languages', 'this', 'regional', 'story', 'by', 'a', 'young', 'alabama', 'woman', 'claims', 'universal', 'appeal', 'harper', 'lee', 'always', 'considered', 'her', 'book', 'to', 'be', 'a', 'simple', 'love', 'story', 'today', 'it', 'is', 'regarded', 'as', 'a', 'masterpiece', 'of', 'american', 'literature']
Stop Removal	['unforgettable', 'novel', 'childhood', 'sleepy', 'southern', 'town', 'crisis', 'conscience', 'rocked', 'killed', 'mockingbird', 'became', 'instant', 'bestseller', 'critical', 'success', 'first', 'published', '1960', 'went', 'won', 'pulitzer', 'prize', '1961', 'later', 'made', 'academy', 'awardwinning', 'movie', 'also', 'classiccompassionate', 'dramatic', 'deeply', 'moving', 'kill', 'mockingbird', 'takes', 'readers', 'roots', 'human', 'behavior', 'innocence', 'experience', 'kindness', 'cruelty', 'love', 'hatred', 'humor', 'pathos', '18', 'million', 'copies', 'print', 'translated', 'forty', 'languages', 'regional', 'story', 'young', 'alabama', 'woman', 'claims', 'universal', 'appeal', 'harper', 'lee', 'always', 'considered', 'book', 'simple', 'love', 'story', 'today', 'regarded', 'masterpiece', 'american', 'literature']
Stemming	['unforgett', 'novel', 'childhood', 'sleepi', 'southern', 'town', 'crisi', 'conscienc', 'rock', 'kill', 'mockingbird', 'becam', 'instant', 'bestsel', 'critic', 'success', 'first', 'publish', '1960', 'went', 'won', 'pulitz', 'prize', '1961', 'later', 'made', 'academy', 'awardwin', 'film', 'also', 'classiccompassion', 'drama', 'deepli', 'move', 'kill', 'mockingbird', 'take', 'reader', 'root', 'human', 'behavior', 'innoc', 'experi', 'kind', 'cruelti', 'love', 'hatr', 'humor', 'patho', '18', 'million', 'copy', 'print', 'translat', 'forti', 'languag', 'region', 'stori', 'young', 'alabama', 'woman', 'claim', 'univers', 'appeal', 'harper', 'lee', 'alway', 'consid', 'book', 'simpl', 'love', 'stori', 'today', 'regard', 'masterpiec', 'american', 'literature']

A one-hot encoding technique is used to map genres. One-hot encoding converts each genre into a binary vector of 1 and 0, where a value of 1 indicates that the novel falls into that genre category, while a value of 0 indicates otherwise.

Text Preprocessing

In the preprocessing stage, five steps are performed. First is punctuation removal, which removes punctuation marks in the text. Second, case folding is changing all capital letters to lowercase letters to ensure consistency. Third, tokenization is the process of breaking the text into tokens in the form of words or phrases. Fourth, stopword removal by removing words that are not meaningful in the context of text analysis. Finally, stemming returns the word to its basic form [19] [20]. The stemming stage can be used or not in text processing, depending on the needs of the analysis. This study will compare texts without stemming and by using stemming. The results of this preprocessing process can be seen in Table 1.

TF-IDF

TF-IDF is a feature extraction method that looks at the relationship of terms in documents. Two approaches affect the value of TF-IDF: Term Frequency (TF) measures how often a word appears in a document, and Inverse Document Frequency (IDF) measures how important a word is in a document. The higher the IDF value, the less often the word appears, which means the more important the word is, and vice versa. Calculating the TF-IDF weight in equation 1 [19].

$$TF.IDF_{std}(t) = tf_d^t x log \frac{N}{df^t} \quad (1)$$

Where tf_d^t is the number of times the term t appears in the document d. N is the total number of documents, and df^t is the number of documents where t occurs.

K-Fold Cross Validation

This method is often used to reduce bias or overfitting in data [21]. In addition, K-Fold Validation is also good for imbalanced data. As in Figure 3, the Fiction genre is too dominating compared to other genres. K-Fold Cross-Validation works by dividing the data into train and test as many times as the number k or fold [22] [23]. This study used k = 10 to divide the data into ten folds, one of which is the test data. Iteration will be done ten times, alternating each fold with the test data. That way, the estimated performance of the model will increase.

Classification Model

Support Vector Machine (SVM)

SVM is a machine learning technique that aims to find the optimal hyperlane with the highest margin that can divide classes linearly. Hyperlane and margin are two very important aspects of SVM that determine whether it is successful in dividing the classes. In SVM, the outermost data points (support vector) are used as a reference in forming class boundaries [24].

In finding the optimal hyperlane, SVM is influenced by the type of kernel used. If the type of kernel used is right, SVM will work optimally [23] [16]. Commonly used kernels are linear, polynomial, and radial basis functions (RBF). Fig. 3. shows the general architecture of the SVM algorithm.

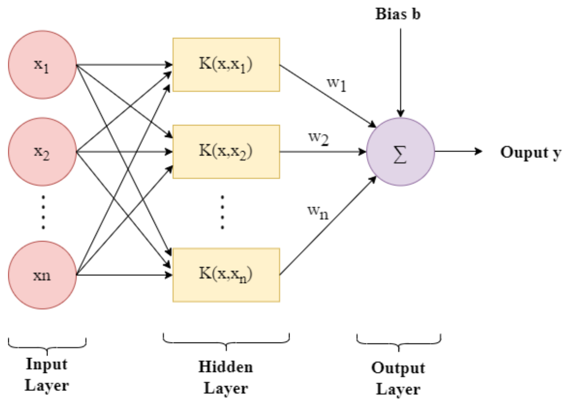

Fig. 3. SVM model architecture

In Fig. 4, SVM has three layers:

1) Input layer
   This layer receives the vector $(x_1, x_2, \ldots, x_n)$ resulting from the TF-IDF feature extraction as a vector for one document.
2) Hidden layer
   The hidden layer contains the kernel function $K(x, x_i)$ to calculate the similarity between input x and training data x_i. This study uses the Radial Basis Function (RBF) kernel type with the equation:

   $$K(x, x_i) = \exp\left(-\gamma \|x - x_i\|^2\right) \quad (2)$$

   where γ is a kernel parameter that determines the influence of a single training example. The closer the input vector x is to a support vector x_i, the higher the similarity value, and thus the more it influences the classification.
3) Output layer
   This layer determines the predicted class of input x by summing the product of the learned weight $\propto_i$, class label y_i, the kernel function $K(x, x_i)$ and bias b, which can be seen in equation 3.

   $$f(x) = \sum_{i=1}^{n} \propto_i y_i K(x, x_i) + b \quad (3)$$

   Since the kernel used is RBF, the decision function becomes non-linear in the original input space.

For multi-label classification, the function $f(x)$ will use the One-vs-Rest approach, where the predicted class is selected based on the highest value. The strenght of the SVM algorithm lies in its ability to handle highdimensional inputs, making it suitable for text classification.

Logistic Regression (LR)

LR is a machine learning technique used to predict the probability of target classes. There are several functions used by LR in determining the target class, namely the net input function to calculate the combination of input features and weights that produce logit values, the sigmoid function to convert logit values into a range of 0 and 1 as a probability scale, and threshold function to determine the final decision in classification by comparing the probability with a certain threshold [25]. Fig. 4. displays the architecture of the Logistic Regression model.

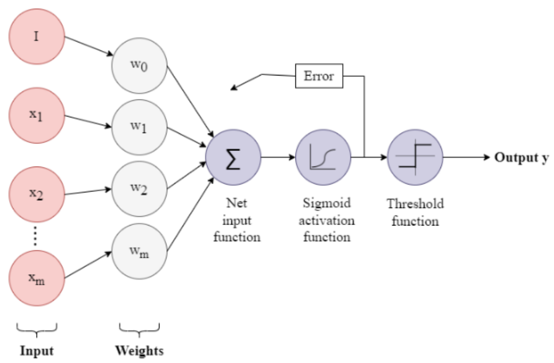

Fig. 4. Logistic Regression model architecture

From Fig. 4, Logistic Regression is implemented in the following steps.

1) Net function input

$$z = \sum_{i=0}^{m} w_i x_i = w^T x \quad (4)$$

Where z indicates the net input value, x is the TF-IDF result vector, and w^T is the transposed weight.

2) Sigmoid activation function
The z value is then put into the sigmoid function to convert it into a probability between 0 and 1. The formula for the sigmoid function:

$$\sigma(z) = \frac{1}{1 + e^{-z}} \quad (5)$$

so that

$$P = \sigma(z) \quad (6)$$

Where P indicates the probability of document x being included in a class.

3) Threshold function
Decides whether a label is active (1) or not (0).

$$\hat{y} = \begin{cases} 1 \; jika \; \sigma(z) \geq \tau \\ 0 \; jika \; \sigma(z) < \tau \end{cases} \quad (7)$$

Where τ is the threshold value, $\hat{y}$ is the predicted label.

Multinomial Naive Bayes (MNB)

MNB is a variant of the Naive Bayes algorithm that is suitable for multi-label text classification [24]. MNB works on the concept of term frequency or the number of occurrences of a word in a document to predict the class [26]. MNB assumes that features (words) in a document appear independently and calculates the probability of a document belonging to a certain class based on the frequency of words in the document.

The probability calculation formula in MNB can be seen in equation 8.

$$P(c|d) \propto P(c) \prod_{i=1}^{n} P(\omega_i|c) \quad (8)$$

Where $P(c|d)$ is the probability of the document d belongs to class c, $P(c)$ is the prior probability of class c or how often the class c appears in training, $P(\omega_i|c)$ is the probability of the word ω_i appears in the class c.

The document d is represented as a set of words $\{\omega_1, \omega_2, \dots, \omega_n\}$ obtained through TF-IDF feature extraction. To determine the class, MNB computes the posterior probability $P(c|d)$, which reflects the likelihood that document d belongs to class c, based on the word frequencies in the document and the word distributions across each class. For example, consider a book synopsis containing *"magic"* and *"wizard"*. If, in the training data, these words frequently occur in the *Fantasy* and *Fiction* genres, then a new document containing these words will have a higher probability of being classified under those genres.

Problem Transformation Methods

Binary Relevance (BR)

Binary Relevance (BR) method is an approach in multi-label classification that splits a multi-label dataset into several single-label datasets. If there are n labels in the dataset, then BR will generate n binary classification models trained independently for each label [27]. The steps of the BR method are as follows: First, a multi-label dataset is converted into several single-label datasets according to the number of labels. Next, a binary model is trained for each of these single-label datasets. Finally, to predict new samples, each binary classifier is used, and the prediction results of all classifiers are combined to form a multi-label prediction. For example, there are five genres in this study, so there will be 5 BR models trained. Table 2 is an example of applying BR on three initial datasets, as shown in Fig. 3.

Table 2. Application of BR method

ID	Fiction
Book1	1
Book2	1
Book3	1

ID	Nonfiction
Book1	0
Book2	0
Book3	0

ID	Fantasy
Book1	0
Book2	1
Book3	0

ID	Mystery
Book1	0
Book2	0
Book3	0

ID	Romance
Book1	0
Book2	0
Book3	1

Label Powerset (LP)

Label Powerset (LP) method transforms multilabel classification into multi-class (single-label) classification by considering each unique combination of labels as a separate class. Instead of predicting multiple labels independently, LP groups label combinations and treats each group as a single label in a new classification task. LP effectively handles the relationship between labels because all labels are treated as combined [9], [24]. The application of LP on the three initial datasets that can be seen in Fig. 3. is shown in Table 3.

Table 3. Application of LP method

ID	Label Combination
Book1	10000
Book2	10100
Book3	10001

The label combination above shows label categories such as fiction, nonfiction, fantasy, mystery, and romance so Book1 indicates the book is only in the fiction category indicated by the number 1, the other labels are 0 (nonfiction = 0, fantasy = 0, fantasy = 0, and mystery = 0).

Performance Evaluation

In evaluating the performance of machine learning models in the multi-label classification of book genres and determining the most effective problem transformation method, four evaluation metrics were used: accuracy, recall, precision, and F1-score.

Accuracy measures the proportion of correct predictions out of all predictions made by the model. The accuracy formula is as follows.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{9}$$

Recall measures the proportion of positive examples that the model identifies out of all positive examples. The recall formula is as follows.

$$Recall = \frac{TP}{TP + FN} \tag{10}$$

Precision measures the proportion of positive examples identified by the model out of all predicted positive examples. The precision formula is as follows.

$$Precision = \frac{TP}{TP + FP} \tag{11}$$

F1-Score is the harmonic mean of precision and recall, providing a metric that balances both aspects. The formula is as follows.

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{12}$$

RESULT AND DISCUSSION

Evaluation of the Best Machine Learning Techniques

In the first stage, a multi-label classification of book genres was tested using three machine learning algorithms to determine which algorithm was the best. The three algorithms tested were Support Machine Learning (SVM), Logistic Regression (LR), and Multinomial

Naive Bayes (MNB). Testing was carried out using two different preprocessing methods, with stemming and without stemming. All algorithms tested use Binary Relevance (BR) as a problem transformation to handle multi-label classification with validation using the K-Fold Cross-Validation method (k = 10). The model evaluation uses accuracy, precision, recall and F1-score metrics. The evaluation results are shown in Table 4. and it illustrated in the Fig. 5.

Table 4. Evaluation of ML algorithm test matrix

Model	Matrix	Preprocessing	
		Stemmed	Unstemmed
SVM	Accuracy	0.519744	0.513845
	Precision	0.836861	0.837768
	Recall	0.711777	0.705934
	F1-score	0.769169	0.766130
LR	Accuracy	0.502498	0.500113
	Precision	0.833857	0.834551
	Recall	0.697354	0.694714
	F1-score	0.759451	0.758121
MNB	Accuracy	0.453368	0.455298
	Precision	0.811317	0.812115
	Recall	0.654424	0.658680
	F1-score	0.724394	0.727280

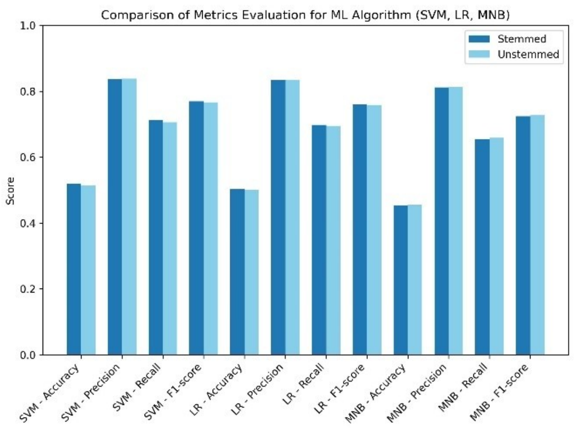

Fig 5. Metrics evaluation for ML algorithm

Based on Fig. 5, the results show that SVM through the stemming process is the best algorithm for multi-label classification of book genres. This is demonstrated by SVM's superior performance in all evaluation metrics: accuracy, precision, recall and F1-score. On data that has gone through the stemming process, SVM achieves an accuracy of 0.519744, a precision of 0.836861, a recall of 0.711777, and an F1-score of 0.769169. This value is higher than LR and MNB on both data types, with and without stemming.

In addition, the stemming process has been proven to contribute positively to model performance. Stemming helps reduce word variability by converting words to their basic form so the model can better focus on the core meaning of each word. This is proven by the higher performance of data that has gone through the stemming process compared to data that has not been stemmed.

Although SVM accuracy is below 80%, this algorithm is still the best choice for this study's multi-label classification of book genres. An accuracy rate of less than 80% can be caused by uneven data distribution, where some genres have much more data than others. This uneven distribution can affect model performance, but SVM still shows superior performance in terms of precision, recall, and F1-score. This shows that SVM can produce more precise and consistent predictions despite the data's complexity and an unbalanced distribution.

Evaluation of the Best Transformation Method

In the second stage, the SVM algorithm with the best algorithm in the first stage is used to test two problem transformation methods: BR and LP with the same K-Fold Cross-Validation of k = 10. The test results are measured using accuracy, precision, recall, and F1-score metrics. The results of BR and LP testing can be seen in Table 5 and it illustrated in the Fig. 6.

The test results show that the LP method has a higher accuracy of 0.542317 than the BR method, which has an accuracy of 0.519744. Precision, recall, and F1-score show slight differences in results between LP and BR. From the results, the best problem transformation method is achieved by LP with a higher accuracy value than BR. Although BR's precision and F1 score are slightly better, LP excels in marginally higher accuracy and recall. This shows that LP is more reliable in classifying data in general, although the differences between other metrics are minimal. LP is superior because it considers the combination of labels present in each instance, thus capturing the relationship between labels better than BR, which treats each label independently. This makes LP more effective in handling label correlations, improving the model's accuracy.

Table 5. Evaluation of BR and LR test matrix

Model	Matrix	Problem Transformation Methods	
		BR	LR
SVM	Accuracy	0.519744	0.542317
	Precision	0.836861	0.814992
	Recall	0.711777	0.712971
	F1-score	0.769169	0.760500

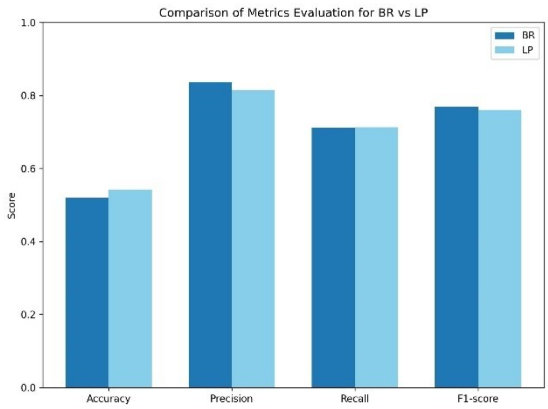

Fig. 6. Metrics evaluation for BR vs LP

CONCLUSION

The following conclusions can be drawn from the multi-label book genre classification research: First, the SVM algorithm with stemming process is the best technique for multi-label book genre classification as it shows superior performance in all evaluation metrics. The advantage of SVM lies in handling data with complex and uneven distributions and producing more precise and consistent predictions. Second, the stemming process contributes positively by reducing word variability and helping the model focus on the meaning of each word. Third, the LP is superior to BR with better accuracy of 0.542317 and better recall, as LP can more effectively capture the relationship between labels. Then, for future research, it is recommended to try data balancing techniques such as oversampling or undersampling and explore additional features or other models to improve accuracy. Overall, SVM + LP method is the best combination in this research.

REFERENCES

[1] P. N. R. Indonesia, ‘Analisis Statistik ISBN per Tahun’, *https://isbn.perpusnas.go.id/Home/Statistik#intro*. 2024.

[2] D. Nurbaiti, ‘Perkembangan Ebook Dalam Industri Penerbitan Buku Fisik Serta Pertumbuhan Minat Menulis Buku’, *IKRAITH Ekon.*, vol. 2, no. 2, pp. 11–20, Jul. 2019.

[3] V. R. S. Nastiti, S. Basuki, and Hilman, ‘Klasifikasi Sinopsis Novel Menggunakan Metode Naïve Bayes Classifier’, *REPOSITOR*, vol. 1, no. 2, pp. 125–130, Dec. 2019. https://doi.org/10.22219/repositor.v1i2.799

[4] B. Falakhi, I. Cholissodin, and R. S. Perdana, ‘Klasifikasi Sinopsis Novel berdasarkan Jenis Genre menggunakan Multi-class Support Vector Machine dan Chi-square’, *J. Pengemb. Teknol. Inf. Dan Ilmu Komput.*, vol. 7, no. 1, pp. 192–202, Jan. 2023.

[5] W. K. Sari, D. P. Rini, R. F. Malik, and I. S. B. Azhar, ‘Multilabel Text Classification in News Articles Using Long-Term Memory with Word2Vec’, *J. RESTI Rekayasa Sist. Dan Teknol. Inf.*, vol. 4, no. 2, pp. 276–285, Apr. 2020, https://doi.org/10.29207/resti.v4i2.1655 .

[6] R. Wang, R. Ridley, X. Su, W. Qu, and X. Dai, ‘A novel reasoning mechanism for multi-label text classification’, *Inf. Process. Manag.*, vol. 58, no. 2, p. 102441, Mar. 2021, https://doi.org/10.1016/j.ipm.2020.102441

[7] Z. Yang and F. Emmert-Streib, ‘Optimal performance of Binary Relevance CNN in targeted multi-label text classification’, *Knowl.-Based Syst.*, vol. 284, p. 111286, Jan. 2024, https://doi.org/10.1016/j.knosys.2023.111286 .

[8] B. Al-Salemi, M. Ayob, G. Kendall, and S. A. M. Noah, ‘Multi-label Arabic text categorization: A benchmark and baseline comparison of multi-label learning algorithms’, *Inf. Process.*

Manag., vol. 56, no. 1, pp. 212–227, Jan. 2019, https://doi.org/10.1016/j.ipm.2018.09.008 .

[9] A. Wiraguna, S. A. Faraby, and Adiwijaya, 'Klasifikasi Topik Multi Label pada Hadis Bukhari dalam Terjemahan Bahasa Indonesia Menggunakan Random Forest', *E-Proceeding Eng.*, vol. 6, no. 1, pp. 2144–2153, Apr. 2019.

[10] J. W. Iskandar and Y. Nataliani, 'Perbandingan Naïve Bayes, SVM, dan k-NN untuk Analisis Sentimen Gadget Berbasis Aspek', *J. RESTI Rekayasa Sist. Dan Teknol. Inf.*, vol. 5, no. 6, pp. 1120–1126, Dec. 2021, https://doi.org/10.29207/resti.v5i6.3588 .

[11] Y. Wahba, N. Madhavji, and J. Steinbacher, 'A Comparison of SVM Against Pre-trained Language Models (PLMs) for Text Classification Tasks', 2023, pp. 304–313. https://doi.org/10.1007/978-3-031-25891-6_23 .

[12] K. Shah, H. Patel, D. Sanghvi, and M. Shah, 'A Comparative Analysis of Logistic Regression, Random Forest and KNN Models for the Text Classification', *Augment. Hum. Res.*, vol. 5, no. 1, p. 12, Dec. 2020, https://doi.org/10.1007/s41133-020-00032-0 .

[13] M. Y. H. Setyawan, R. M. Awangga, and S. R. Efendi, 'Comparison Of Multinomial Naive Bayes Algorithm And Logistic Regression For Intent Classification In Chatbot', in *2018 International Conference on Applied Engineering (ICAE)*, IEEE, Oct. 2018, pp. 1–5. https://doi.org/10.1109/INCAE.2018.8579372 .

[14] M. Veziroğlu, E. Veziroğlu, and İ. Ö. Bucak, 'Performance Comparison between Naive Bayes and Machine Learning Algorithms for News Classification', in *Bayesian Inference - Recent Trends*, IntechOpen, 2024. https://doi.org/10.5772/intechopen.1002778

[15] M. Abbas, K. A. Memon, A. A. Jamali, S. Memon, and A. Ahmed, 'Multinomial Naive Bayes Classification Model for Sentiment Analysis', *IJCSNS Int. J. Comput. Sci. Netw. Secur.*, vol. 19, no. 3, pp. 62–67, Mar. 2019.

[16] K. Khadijah, N. Sabilly, and F. A. Nugroho, 'Sentiment Analysis of League of Legends: Wild Rift Reviews on Google Play Using Naive Bayes Classifier', *J. Ilm. Kursor*, vol. 12, no. 1, pp. 23–30, Jul. 2023, https://doi.org/10.21107/kursor.v12i01.328 .

[17] I. Z. Amalia, A. N. Ponco Bimantoro, A. Z. Arifin, M. Faisol, R. Indraswari, and R. W. Sholikah, 'Indonesian-Translated Hadith Content Weighting in Pseudo-Relevance Feedback Query Expansion', *J. Ilm. Kursor*, vol. 11, no. 1, Jul. 2021, https://doi.org/10.21107/kursor.v11i1.249 .

[18] Rianto, A. B. Mutiara, E. P. Wibowo, and P. I. Santosa, 'Improving the accuracy of text classification using stemming method, a case of non-formal Indonesian conversation', *J. Big Data*, vol. 8, no. 1, p. 26, Dec. 2021, https://doi.org/10.1186/s40537-021-00413-1 .

[19] M. Lestandy, A. Abdurrahim, and L. Syafa'ah, 'Analisis Sentimen Tweet Vaksin COVID-19 Menggunakan Recurrent Neural Network dan Naïve Bayes', *J. RESTI Rekayasa Sist. Dan Teknol. Inf.*, vol. 5, no. 4, pp. 802–808, Aug. 2021, https://doi.org/10.29207/resti.v5i4.3308 .

[20] U. L. Yuhana, I. Imamah, C. Fatichah, and B. J. Santoso, 'Effectiveness of Deep Learning Approach For Text Classification in Adaptive Learning', *J. Ilm. Kursor*, vol. 11, no. 3, pp. 137–144, Jul. 2022, https://doi.org/10.21107/kursor.v11i3.285 .

[21] V. N. Wijayaningrum, A. P. Kirana, and I. K. Putri, 'Student Academic Performance Prediction Framework with Feature Selection and Imbalanced Data Handling', *J. Ilm. Kursor*, vol. 12, no. 3, pp. 123–134, May 2024,

https://doi.org/10.21107/kursor.v12i3.356 .

[22] T. Ridwansyah, 'Implementasi Text Mining Terhadap Analisis Sentimen Masyarakat Dunia Di Twitter Terhadap Kota Medan Menggunakan K-Fold Cross Validation Dan Naïve Bayes Classifier', *KLIK Kaji. Ilm. Inform. Dan Komput.*, vol. 2, no. 5, pp. 178–185, Apr. 2022, https://doi.org/10.30865/klik.v2i5.362 .

[23] D. M. Abdullah and A. M. Abdulazeez, 'Machine Learning Applications based on SVM Classification A Review', *Qubahan Acad. J.*, vol. 1, no. 2, pp. 81–90, Apr. 2021, https://doi.org/10.48161/qaj.v1n2a50 .

[24] W. Fernando, D. Jollyta, D. Priyanto, and D. Oktarina, 'The Influence Of Data Categorization And Attribute Instances Reduction Using The Gini Index On The Accuracy Of The Classification Algorithm Model', *J. Ilm. Kursor*, vol. 12, no. 3, pp. 111–122, May 2024, https://doi.org/10.21107/kursor.v12i3.372 .

[25] R. Torres, O. Ohashi, and G. Pessin, 'A Machine-Learning Approach to Distinguish Passengers and Drivers Reading While Driving', *Sensors*, vol. 19, no. 14, p. 3174, Jul. 2019, https://doi.org/10.3390/s19143174 .

[26] Yuyun, N. Hidayah, and S. Sahibu, 'Algoritma Multinomial Naïve Bayes Untuk Klasifikasi Sentimen Pemerintah Terhadap Penanganan Covid-19 Menggunakan Data Twitter', *J. RESTI Rekayasa Sist. Dan Teknol. Inf.*, vol. 5, no. 4, pp. 820–826, Aug. 2021, https://doi.org/10.29207/resti.v5i4.3146 .

[27] Nofriani and N. B. Kurniawan, 'Harnessing Multi-label Classification Approaches for Economic Phenomena Categorization', *ASEAN J. Sci. Technol. Dev.*, vol. 38, no. 2, Aug. 2021, https://doi.org/10.29037/ajstd.680 .