OPEN ACCESS
Indo-JC
Indonesia Journal of Computing
ISSN 2460-9056
journals.telkomuniversity.ac.id/indojc

Ind. Journal on Computing
Vol. 10, Issue. 1, August 2025. pp. 23-36
doi:10.21108/indojc.v10i1.9100

Hyperparameter Optimization Analysis of MultinomialNB and Logistic Regression in Multi-Feature Text-Based Film Genre Classification

Shabrio Cahyo Wardoyo[1], Umniy Salamah[2]

Informatics, Faculty of Computer Science, Mercu Buana University
Meruya Selatan Street, Kembangan District, West Jakarta, DKI Jakarta 11650, Indonesia

[1] 41521010025@student.mercubuana.ac.id
[2] umniy.salamah@mercubuana.ac.id

Abstract

This study aims to analyze and compare the performance of two text classification algorithms—Multinomial Naive Bayes (MNB) and Logistic Regression (LR)—for film genre classification using multi-feature text data, both with and without hyperparameter optimization. Film genres play a crucial role in digital content recommendation systems; however, manual classification is subjective and time-consuming. The dataset, obtained from Letterboxd via Kaggle, includes film titles, descriptions, and themes. After preprocessing and text normalization (tokenization, lemmatization, and stemming), the text data were transformed into numerical features using the TF-IDF method. Two modeling scenarios were applied: the first using default parameters, and the second employing GridSearchCV to find the optimal hyperparameter settings. Model performance was evaluated using accuracy, precision, recall, and F1-score. The results indicate that the optimized LR model achieved the highest accuracy of 0.847, followed by the optimized MNB model with an accuracy of 0.837. This study concludes that hyperparameter optimization significantly improves model performance and that LR outperforms MNB in the context of multi-feature text-based genre classification.

Keywords: Multinomial Naive Bayes, Logistic Regression, TF-IDF, Hyperparameter Optimization, Film Genre Classification

Abstrak

Penelitian ini bertujuan untuk menganalisis dan membandingkan kinerja dua algoritma klasifikasi teks—Multinomial Naive Bayes (MNB) dan Logistic Regression (LR)—dalam klasifikasi genre film berbasis teks multi-fitur, dengan dan tanpa optimasi hyperparameter. Genre film dipilih karena memiliki peran penting dalam sistem rekomendasi konten digital, namun klasifikasi manual bersifat subjektif dan memakan waktu. Data penelitian diperoleh melalui scraping dari platform Letterboxd dan diakses melalui Kaggle, terdiri dari judul, deskripsi, dan tema film. Setelah melalui proses pembersihan dan normalisasi teks (tokenisasi, lemmatization, dan stemming), fitur teks diekstraksi menggunakan metode TF-IDF. Penelitian ini menerapkan dua skenario pemodelan: pertama tanpa optimasi hyperparameter, dan kedua menggunakan GridSearchCV untuk menemukan kombinasi parameter terbaik. Evaluasi dilakukan menggunakan metrik akurasi, presisi, recall, dan F1-score. Hasil menunjukkan bahwa LR dengan optimasi menghasilkan akurasi tertinggi sebesar 0.847, diikuti oleh MNB dengan optimasi sebesar 0.837. Penelitian ini menyimpulkan bahwa optimasi hyperparameter secara signifikan meningkatkan performa klasifikasi, dan LR lebih unggul dibandingkan MNB dalam skenario ini.

Kata Kunci: Multinomial Naive Bayes, Logistic Regression, TF-IDF, Optimasi Hyperparameter, Klasifikasi Genre Film

Received on January, 2025. Accepted on August, 2025

I. Introduction

THE film industry—both globally and in Indonesia—has been undergoing rapid transformation with the advent of streaming services such as Netflix, Disney+, and various local platforms. Recent analyses show that streaming services have permanently altered distribution economics, making robust metadata indispensable for discovery [1]. Film genre plays a pivotal role in recommendation mechanisms by signalling a work's thematic focus and narrative tone; unfortunately, manual genre assignment is often subjective and inconsistent [2]. Consequently, data-driven approaches—particularly those based on machine learning—are gaining wider adoption.

Lightweight text-classification algorithms such as Multinomial Naïve Bayes (MNB) and Logistic Regression (LR) are frequently selected for processing film synopses because they remain efficient even with high-dimensional data [3]. Their performance, however, is highly sensitive to hyperparameter choices—for example, the alpha value in MNB or the regularisation strength (C) in LR—making hyperparameter optimization critical. Moreover, most studies in Indonesia restrict textual features to a single source (the synopsis alone) and rarely compare MNB and LR before and after tuning, especially in multi-feature scenarios that combine a film's title, thematic tags, and synopsis [4].

Addressing this gap, the present study poses three research questions: (1) How does MNB perform in classifying film genres using multi-feature text, with and without hyperparameter optimization? (2) How does LR perform under the same scenario? and (3) Which algorithm delivers the best performance after tuning? Answers to these questions are expected to deepen our understanding of the effectiveness of these two classic algorithms for film-genre classification.

Accordingly, this research aims to: (a) evaluate MNB performance before and after hyperparameter optimization using accuracy, precision, recall, and F1-score; (b) assess LR under identical conditions; and (c) compare the two algorithms post-optimization to determine the most reliable model for multi-feature, text-based genre classification.

The study's contributions span both academic and practical realms. Academically, it expands the literature on hyperparameter optimization for MNB and LR within natural language processing (NLP). Practically, the resulting automatic classification system can accelerate genre tagging for film studios and distributors, assist user-generated content platforms in filtering submissions, and and systems that generate genre tags automatically can raise recommendation accuracy while reducing manual effort [5].

This investigation is delimited as follows: (i) only MNB and LR are analyzed, both using TF-IDF feature representations; (ii) textual features are confined to the film's title, thematic tags, and synopsis; (iii) after extensive cleaning, the Letterboxd dataset is reduced to 7,635 entries released up to 2031; (iv) classification focuses on three dominant genres—Drama, Comedy, and Horror—to maintain balanced data; (v) evaluation metrics include accuracy, precision, recall, F1-score, and processing time; and (vi) the research concludes at model evaluation without full deployment into a production platform. These boundaries were set to maintain a clear focus and allow in-depth analysis before extending the work in future studies.

II. Literature Review

Over the past five years, research in Indonesia on film-text classification—especially on optimising Multinomial Naïve Bayes (MNB) and Logistic Regression (LR)—has expanded rapidly. Toyibah et al. (2024) compared the two algorithms on 40,000 Indonesian-language IMDb reviews; LR proved more reliable with an accuracy of 89.32 %, whereas MNB achieved only 85.28 % [3].

Earlier, Ramdan et al. (2023) evaluated 2,000 film reviews in the Journal of Applied Intelligent System and showed that LR remained competitive (80.61 % accuracy), although it ranked slightly below SVC [6]. Focusing on Naïve Bayes, Rifki et al. (2024) applied text-mining techniques to 1,500 IMDb comments and attained 87.2 % accuracy, highlighting the importance of TF-IDF feature selection and α-tuning for MNB [7].

For multi-label tasks, Akbar et al. (ICITISEE 2022) processed 10,432 film synopses covering 18 genres and employed GridSearch; the best MNB model (F1 = 0.48) narrowly outperformed LR (F1 = 0.43), although both were surpassed by SVM (F1 = 0.58) [4].

Most recently, Sintawati et al. (2025) analyzed 8,888 YouTube comments on the documentary Dirty Vote; LR achieved an AUC of 0.955, overtaking SVM in an imbalanced-data setting and underscoring the effectiveness of class-weight tuning for LR in multi-class problems [8].

These five studies demonstrate that hyperparameter optimization (e.g., class weight, solver choice, and C in LR, or α in MNB) directly affects model accuracy. They also reveal that data sources are still dominated by IMDb and YouTube, leaving a research gap: no previous work has employed Letterboxd or multi-feature datasets (e.g., title, synopsis, and reviews) as planned in the present study. Overall, LR tends to remain stable and superior on large Indonesian-language datasets, whereas MNB remains competitive in short-text or multi-label scenarios given appropriate feature selection.

A. Multinomial Naïve Bayes

Multinomial Naïve Bayes is a generative model that estimates the posterior probability that a document d belongs to class c from word-frequency statistics:

$$P(c|d) \propto P(c) \prod_{i=1}^{|V|} P(w_i|c)^{n_i} \tag{1}$$

where n_i is the frequency of term w_i and $|V|$ is the vocabulary size. Because its parameters are obtained simply by counting frequencies—without any iterative optimization—the model is extremely fast and memory-efficient even for high-dimensional text data.

Performance is controlled chiefly by the hyper-parameter α (Laplace/Lidstone smoothing) and the fit_prior option. A small α preserves the influence of rare words. A large α dampens the dominance of very common words.

B. Logistic Regression

Logistic Regression is a discriminative model that learns weight vector w directly by maximising P(c | x) via the sigmoid or soft-max function. For the binary case:

$$P(y = 1|x) = \sigma(w^\top x + b) \tag{2}$$

Logistic Regression is widely used as an NLP baseline because it handles sparse text representations effectively, supports both L_1 and L_2 regularisation to mitigate overfitting, and scales well to large feature spaces while remaining conceptually simple.

C. Hyperparameter

In the field of machine learning, hyperparameters refer to fixed values determined by the researcher before the training process begins and they are not updated by the algorithm during optimization.

In this experiment, there are two layers of hyperparameters. The first layer is at the feature-extraction stage; an example is max_features, the number of TF-IDF n-grams considered. Setting the limit at 5,000 produces vectors that are compact and quick to compute, whereas increasing it to 10,000 enriches the vocabulary representation at the cost of greater memory requirements. The second layer operates at the level of the classification algorithm. For Multinomial Naïve Bayes, the key hyperparameter is α (Lidstone smoothing). A small α (e.g., 0.5) makes the model sensitive to rare words—potentially improving its ability to distinguish genres with distinctive terminology—but it is prone to high variance. A large α (1.5) smooths the distribution, making the model more stable but potentially losing specific details. Logistic Regression has the hyperparameter C, the inverse of the regularisation strength. A small C drives the weights toward zero, reducing over-fitting, whereas a large C allows the weights to adapt freely to complex patterns.

III. Research Method

Fig. 1. Research Cycles

This study is quantitative in nature and employs a comparative experimental design. The experimental framework was chosen because the main objective is to measure—objectively—the performance differences between two text-classification algorithms, Multinomial Naïve Bayes (MNB) and Logistic Regression (LR), on a multi-feature corpus (title, theme, synopsis) that has been represented as TF-IDF vectors. Performance validation is carried out using accuracy, precision, recall, and F1-score, so that each treatment (hyperparameter setting) can be compared statistically [2].

A. Data Collection

At this stage, data collection is conducted through web-scraping techniques from the Letterboxd.com site, a social platform that focuses on the film-fan community and provides various information on films of different genres, release years, and user reviews. The scraped data are then compiled in a dataset that can be accessed and downloaded via the Kaggle platform.

The dataset used in this study consists of three main files—genres.csv, movies.csv, and themes.csv—each with the following structure and amount of data:

1) genres.csv: This dataset contains information on the genre of each film. The dataset consists of more than one million rows, with two main attributes:

- id: a unique identifier for each film.
- genre: the film's genre (e.g., Action, Comedy, Drama, etc.)

2) movies.csv: This file stores the primary film metadata. The dataset consists of 941,597 rows and includes the following attributes:

- id: a unique film identifier
- name: the film title
- date: the film's release date
- tagline: the film's promotional slogan or tagline
- description: a short synopsis or description of the film
- minute: the film's duration (in minutes)
- rating: the film's rating score (typically on a 1–10 scale)

3) themes.csv: This file records various themes related to each film. This dataset also consists of more than one million rows, with two main attributes:

- id: a unique film identifier
- theme: the film's main narrative element or theme (e.g., Friendship, Survival, Love, Revenge, etc.)

These three datasets are used as the primary data source in the research and will be further processed in the subsequent stages, namely preprocessing, exploration, feature engineering, modeling, and evaluation. With data obtained directly from Letterboxd.com, the study gains very broad and representative coverage of various film genres and descriptions, which strongly supports the automatic genre-classification process.

B. Data Processing

After the data have been collected, the next stage is data processing, whose purpose is to clean and organise the data so that they can be used in the film-genre prediction modelling process. This procedure is carried out in the Python programming language with a variety of libraries, including pandas, nltk, re, langid, and concurrent.futures.

1) Merging Theme Data: The themes.csv dataset initially has a format in which each row contains only one theme for a single film id, as shown in Table I. To simplify analysis, all themes belonging to the same id must be combined into a single row, and the resulting structure is shown in Table II.

TABLE I
Sample of Theme Data Before Merged

id	theme
1000001	Quirky and endearing relationships
1000001	Amusing jokes and witty satire
1000002	Humanity and the world around us
1000002	Twisted dark psychological thriller
1000003	Humanity and the world around us

TABLE II
Sample of Theme Data After Merged

id	theme
1000001	Quirky and endearing relationships, Amusing jokes and witty satire
1000002	Humanity and the world around us, Twisted dark psychological thriller
1000003	Humanity and the world around us, Crude humor and satire
1000004	Humanity and the world around us, Surreal and thought-provoking visions of life and death
1000005	Humanity and the world around us, Moving relationship stories

2) Merging Genre Data: Just like the themes, in Tables III and IV the genre data are also merged by id so that each film has only one row of data even though it may belong to more than one genre.

TABLE III
Sample of Genre Data Before Merged

id	genre
1000001	Comedy
1000001	Adventure
1000002	Thriller
1000002	Drama
1000003	Comedy

TABLE IV
Sample of Genre Data After Merged

id	genre
1000001	Comedy, Adventure

1000002	Thriller, Drama
1000003	Comedy, Action
1000004	Drama, Romance
1000005	Drama, Comedy

3) Merging All Attributes into a Single Dataset: After the themes, genres, and movies data have been processed, the three are combined into a single dataset based on the film ID. An example of the main dataset after the themes, genres, and movies attributes have been merged is presented in Table V.

TABLE V
Sample of Main Dataset After Merging All Attributes

id	name	date	tagline	description	minute	rating	theme	genre
1000001	Barbie	2023	She's everything. He's just Ken.	Barbie and Ken are having the time of their lives in the colorful and seemingly perfect world of Barbie Land. However, when they get a chance to go to the real world, they soon discover the joys and perils of living among humans.	114	3.86	Humanity and the world around us, Crude humor and satire, Moving relationship stories, Emotional and captivating fantasy storytelling, Surreal and thought-provoking visions of life and death, Quirky and endearing relationships, Amusing jokes and witty satire, Laugh-out-loud relationship entanglements	Comedy, Adventure
1000002	Parasite	2019	Act like you own the place.	All unemployed, Ki-taek's family takes peculiar interest in the wealthy and glamorous Parks for their livelihood until they get entangled in an unexpected incident.	133	4.56	Humanity and the world around us, Moving relationship stories, Crude humor and satire, Surreal and thought-provoking visions of life and death, Touching and sentimental family stories, Powerful stories of heartbreak and suffering, Quirky and endearing relationships, Dreamlike, quirky, and surreal storytelling	Thriller, Drama

4) Removing unused features from the main dataset: The first step in dataset cleansing is to eliminate features that are not required. Attributes such as date, minute, rating, and tagline are removed because they do not contribute directly to text-based film-genre classification. Moreover, these features contain a high proportion of null values, so including them in modelling could impair the classifier's performance and accuracy.

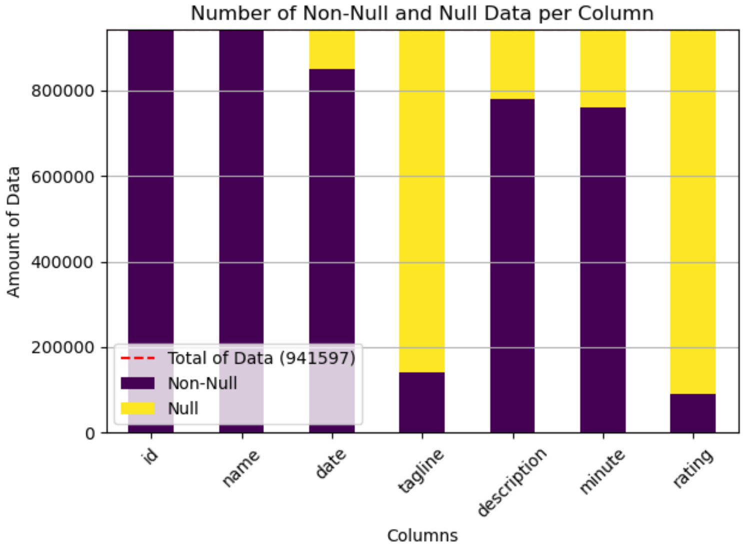

Fig. 2. Number of Non-Null and Null Data per Column

Fig.2 shows the distribution of non-null and null values for each feature in the dataset. It is clear that several columns—date, tagline, description, minute, and rating—contain a relatively high proportion of null values. This observation is one of the main reasons for removing these features, so as to maintain the quality of the data that will be used for model training.

After the data have been cleaned, the only remaining features in the dataset are the columns id, name, description, themes, and genre. Table VI provides two sample rows of data after the cleaning step.

TABLE VI
SAMPLE DATASET AFTER FEATURE DROP

id	name	description	themes	genre
1000001	Barbie	Barbie and Ken are having the time of their lives in the colorful and seemingly perfect world of Barbie Land. However, when they get a chance to go to the real world, they soon discover the joys and perils of living among humans.	Humanity and the world around us, Crude humor and satire, Moving relationship stories, Emotional and captivating fantasy storytelling, Surreal and thought-provoking visions of life and death, Quirky and endearing relationships, Amusing jokes and witty satire, Laugh-out-loud relationship entanglements	Comedy, Adventure
1000002	Parasite	All unemployed, Ki-taek's family takes peculiar interest in the wealthy and glamorous Parks for their livelihood until they get entangled in an unexpected incident.	Humanity and the world around us, Intense violence and sexual transgression, Twisted dark psychological thriller, Heartbreaking and moving family drama, Enduring stories of family and marital drama, Touching and sentimental family stories, Intense political and terrorist thrillers	Comedy, Thriller, Drama

5) Removing non-English data using the langid library: In text-based classification, language consistency within the data is a critical factor for ensuring model accuracy. Because this classification model is designed and trained on English-language text, a filtering process is required to eliminate data written in languages other than English. This step ensures that the extracted text features are not mixed with sentence structures or vocabulary from other languages that could impair model performance.

6) Text cleansing and normalization with tokenization, lemmatization, and stemming techniques: An essential stage in text-based data processing is to perform cleansing and normalization so that the data fed into the model become more consistent and intelligible to the classification algorithm. This process is applied to text features such as name, description, and themes. Several steps carried out in this stage include:

- Lowercasing: Converting all letters to lowercase to avoid duplication of identical words in uppercase form.
- Special-character removal: Deleting non-alphanumeric characters (except basic punctuation) so the text is free of meaningless symbols.
- Tokenization: Splitting the text into words (tokens) so that each can be processed individually.
- Lemmatization and Stemming: These two methods are used to reduce words to their root forms. Lemmatization preserves the grammatically correct base form of a word, whereas stemming cuts the word down to its simplest form.

Table VII shows a sample of the dataset after data-cleansing steps ranging from lowercasing through lemmatization and stemming.

TABLE VII
Sample of Data After Cleansing and Normalization

id	name	description	themes	genre
1000001	barbie	barbie and ken are having the time of their life in the colorful and seemingly perfect world of barbie land . however , when they get a chance to go to the real world , they soon discover the joy and peril of living among human	Humanity and the world around us, Crude humor and satire, Moving relationship stories, Emotional and captivating fantasy storytelling, Surreal and thought-provoking visions of life and death, Quirky and endearing relationships, Amusing jokes and witty satire, Laugh-out-loud relationship entanglements	[Comedy, Adventure]
1000002	parasite	all unemployed , kitaeks family take peculiar interest in the wealthy and glamorous park for their livelihood until they get entangled in an unexpected incident	Humanity and the world around us, Intense violence and sexual transgression, Twisted dark psychological thriller, Heartbreaking and moving family drama, Enduring stories of family and marital drama, Touching and sentimental family stories, Intense political and terrorist thrillers	[Comedy, Thriller, Drama]

7) Simplifying the classification by storing a single genre in each row: Each film was initially listed with multiple genres (e.g., [“Action”, “Adventure”]). Because this study employs single-label classification using Multinomial Naïve Bayes and Logistic Regression, every entry was modified so that it contains only one genre. When a film has more than one genre, the genre appearing last is chosen, as its position typically indicates the most specific or dominant category. This decision is also influenced by the data’s class imbalance and the limited number of examples for most of the other genres, conditions that could reduce the model's overall accuracy. Table VIII shows sample data after storing a single genre in each row.

TABLE VIII
SAMPLE OF DATA AFTER STORING ONE GENRE

id	name	description	themes	genre
1000001	barbie	barbie and ken are having the time of their life in the colorful and seemingly perfect world of barbie land . however , when they get a chance to go to the real world , they soon discover the joy and peril of living among human	humanity and the world around u , crude humor and satire , moving relationship story , emotional and captivating fantasy storytelling , surreal and thoughtprovoking vision of life and death , quirky and endearing relationship , amusing joke and witty satire , laughoutloud relationship entanglement	Adventure
1000002	parasite	all unemployed , kitaeks family take peculiar interest in the wealthy and glamorous park for their livelihood until they get entangled in an unexpected incident	humanity and the world around u , intense violence and sexual transgression , twisted dark psychological thriller , heartbreaking and moving family drama , enduring story of family and marital drama , touching and sentimental family story , intense political and terrorist thriller	Drama

8) Selecting the top three genres and balancing the dataset to prevent model bias: After the data were cleaned and simplified, this study was directed toward the three genres with the highest frequencies to narrow the classification scope while minimising class imbalance, which can cause the model to favor genres that have larger data counts; choosing these top three genres—identified by their most frequent occurrences in the dataset as shown in Fig. 3—is expected to keep the analysis focused and to prevent the model from biasing toward dominant categories even when a film's description does not truly support such predictions.

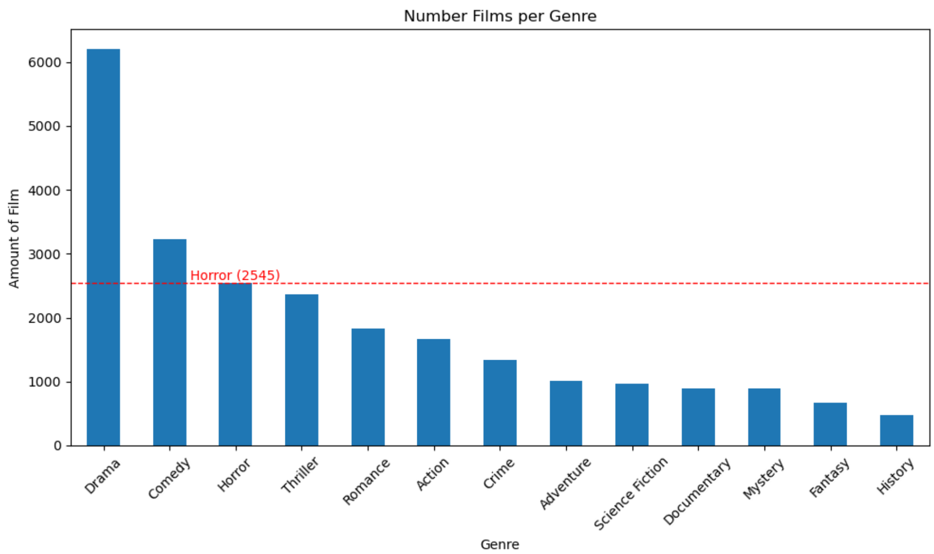

Fig. 3. Number Films per Genre

C. Feature Engineering

Text features are the primary components in this modelling task. The columns used to build the text features are name (title), description (synopsis), and themes. These three columns are concatenated into a single text string for each film entry to represent a broader semantic context.

Next, feature extraction is performed using the Term Frequency–Inverse Document Frequency (TF-IDF) vectorisation method, which converts text into a numerical representation. The parameter ngram_range = (1, 2) is applied to capture unigrams and bigrams, and the maximum-feature limit is adjusted during hyperparameter search (the top 5 000 and 10 000 features). The classification target (y) is taken from the genre column, which has been encoded with LabelEncoder.

D. Modeling

1) Multinomial Naïve Bayes (Without Optimization): The first model is built as a two-stage scikit-learn pipeline—a unigram-based TF-IDF vectoriser with all default parameters, followed by Multinomial Naïve Bayes with *alpha = 1.0*—and is trained directly with *fit()* without further hyperparameter tuning.

2) Logistic Regression (Without Optimization): The second model adopts a two-stage pipeline with default settings: an initial unigram TF-IDF vectoriser without vocabulary restriction, followed by Logistic Regression with C = *1.0* and *max_iter = 1000*. This pipeline is trained and evaluated directly through *fit–predict* without additional hyperparameter tuning.

3) Multinomial Naïve Bayes (With GridSearchCV): To maximise performance, Grid Search is applied to test various combinations of TF-IDF feature counts (max_features) and MultinomialNB smoothing values (alpha); each configuration is evaluated through five-fold cross-validation, and the parameter set with the highest mean accuracy is selected as the final configuration.

4) Logistic Regression (With GridSearchCV): Grid Search tunes two main hyperparameters: *tfidf__max_features*, the maximum number of features produced by TF-IDF, and *clf__C*, the regularisation coefficient in Logistic Regression. The value of C controls regularisation strength—the smaller the C, the stronger the penalty applied to the model weights—so the optimal combination balances model complexity and generalisation ability.

IV. Results And Discussion

Evaluation is conducted to measure the performance of the Multinomial Naïve Bayes and Logistic Regression algorithms in classifying film genres based on text features. The evaluation is carried out under two conditions—before and after optimization using GridSearchCV. The evaluation parameters include accuracy, precision, recall, F1-score, and model processing time.

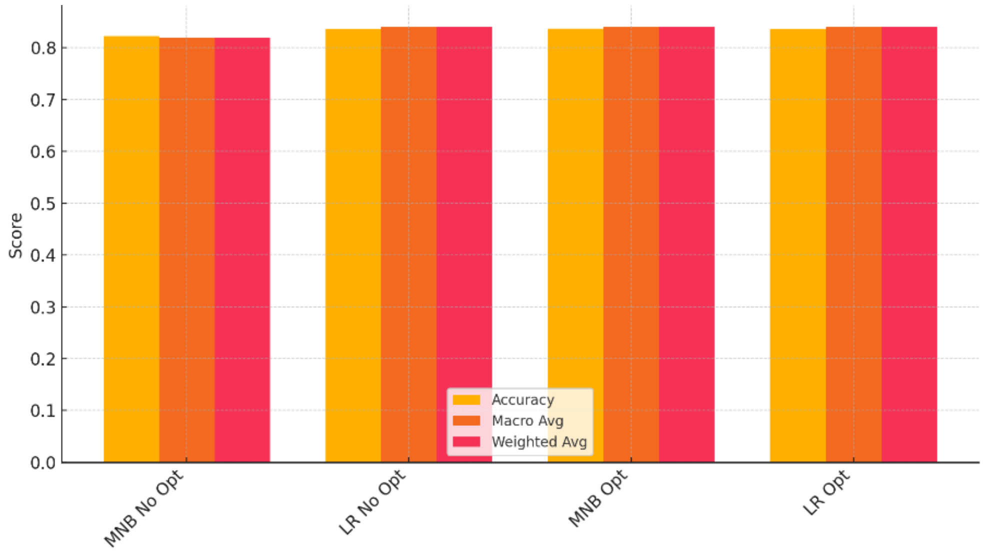

Fig. 4. Comparison of Overall Metrics per Model

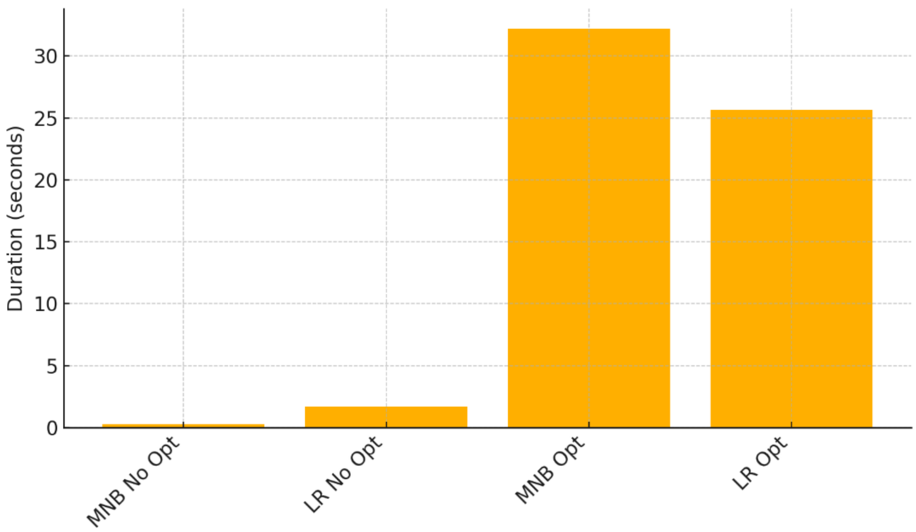

Fig. 5. Processing Time per Model

Figures 4 and 5 together illustrate that although hyperparameter tuning via GridSearchCV produces only marginal improvements in overall performance metrics (accuracy, macro-average F1 and weighted-average F1 rise by merely 0.001–0.014), it incurs a dramatic increase in training time—from under two seconds for the base MNB and LR models to over 25 seconds for their optimized counterparts—highlighting the clear trade-off between slight accuracy gains and substantially higher computational cost.

A. Multinomial Naïve Bayes (Without Optimization)

- Processing Time fit(): 0.278 seconds
- Accuracy: 0.823
- Macro Avg: 0.82
- Weighted Avg: 0.82

TABLE IX
Classification Report MNB Without Optimization

Genre	Precision	Recall	F1-Score	Support
Comedy	0.80	0.77	0.78	509
Drama	0.81	0.75	0.78	509
Horror	0.86	0.94	0.90	509

B. Logistic Regression (Without Optimization)

- Processing Time fit(): 1.689 seconds
- Accuracy: 0.836
- Macro Avg: 0.84
- Weighted Avg: 0.84

TABLE X
Classification Report LR Without Optimization

Genre	Precision	Recall	F1-Score	Support
Comedy	0.80	0.78	0.79	509
Drama	0.80	0.81	0.80	509
Horror	0.90	0.92	0.91	509

C. Multinomial Naïve Bayes (With GridSearchCV)

- Processing Time Grid Search: 32.223 seconds
- Best Parameters: {'clf__alpha': 1.5, 'tfidf__max_features': 10000}
- Accuracy: 0.837
- Macro Avg: 0.84
- Weighted Avg: 0.84

TABLE XI
Classification Report MNB With Optimization

Genre	Precision	Recall	F1-Score	Support
Comedy	0.81	0.78	0.80	509
Drama	0.82	0.79	0.80	509
Horror	0.88	0.94	0.91	509

D. Logistic Regression (With GridSearchCV)

- Processing Time Grid Search: 25.630 seconds
- Best Parameters: {'clf__C': 1, 'tfidf__max_features': 10000}
- Accuracy: 0.836
- Macro Avg: 0.84
- Weighted Avg: 0.84

TABLE XII
Classification Report LR With Optimization

Genre	Precision	Recall	F1-Score	Support
Comedy	0.81	0.78	0.80	509
Drama	0.80	0.81	0.80	509
Horror	0.90	0.92	0.91	509

V. Conclusion

This study evaluates hyperparameter optimization for Multinomial Naïve Bayes (MNB) and Logistic Regression (LR) in a text-based film-genre classification task that leverages three complementary information sources—title, description, and theme—to enrich semantic context beyond what any single text feature can provide. Under default settings, LR achieves an accuracy of 83.6 %, marginally outperforming MNB's 82.3 %, which suggests that LR is inherently more resilient at mapping diverse textual representations without additional tuning. However, after performing a five-fold GridSearchCV, MNB's accuracy increases to 83.7 %, slightly surpassing LR, which remains at 83.6 %. This shift underscores the critical role of the smoothing parameter α in MNB and the regularization constant C in LR, both of which directly impact classification quality.

Post-tuning, MNB emerges as the preferred model: not only does it deliver the highest accuracy, but it also requires shorter training time and entails lower computational complexity. Such characteristics make it especially suitable for deployment in resource-constrained environments or real-time systems [9]. Moreover, the success of combining title, description, and theme confirms the value of multi-feature approaches: enriching the input with multiple semantic cues consistently outperforms single-feature baselines and provides empirical support for feature-fusion strategies in text classification research [4], [10].

The achieved accuracy of 83.7 % offers a practical benchmark for various stakeholders. Film studios and distributors can accelerate post-production genre tagging and tailor marketing strategies more effectively. User-generated content platforms (e.g., YouTube, Vimeo) may use it to automate content grouping and recommendation, while family-safe services gain more precise content-filtering tools. Digital libraries and OTT providers stand to reduce metadata-management time and cost. Academically, our findings extend the literature on MNB and LR optimization and lay an empirical foundation for future studies on feature combination and hyperparameter tuning in other text domains. Socially, a robust classification system enhances viewers' ability to discover content aligned with their preferences and supports a more structured film-distribution ecosystem.

References

[1] Song, J. The Evolution and Impact of Streaming Services: Changing the Media Landscape. Global Media Journal, 22(72), Article 470, 2024. doi:10.36648/1550-7521.22.72.470.

[2] Rajput, N. K. & Grover, B. A. (2022). "A Multi-Label Movie Genre Classification Scheme Based on the Movie's Subtitles." Multimedia Tools and Applications, 81(22), 32469–32490. doi:10.1007/s11042-022-12961-6

[3] Z. B. Toyibah et al., "Performance Comparison of Multinomial Naïve Bayes and Logistic Regression Algorithms in Sentiment Analysis of IMDb Movie Ratings," Edutic Scientific Journal, vol. 10, no. 2, pp. 181–189, 2024, doi: 10.21107/edutic.v10i2.28150.

[4] J. Akbar, E. Utami, and A. Yaqin, "Multi-Label Classification of Film Genres Based on Synopses Using SVM, Logistic Regression, and Naïve Bayes," in Proceedings of ICITISEE 2022, 2022, doi: 10.1109/ICITISEE57756.2022.10057828.

[5] Park, H.; Yong, S.; You, Y.; Lee, S.; Moon, I.-Y. Automatic Movie Tag Generation System for Improving the Recommendation System. Applied Sciences, 12(21), 10777, 2022. doi:10.3390/app122110777.

[6] D. S. R. Ramdan et al., "Film Review Sentiment Analysis: A Comparison of Logistic Regression and Support Vector Classification Performance Using TF-IDF," Journal of Applied Intelligent Systems, vol. 8, no. 3, 2023, doi: 10.33633/jais.v8i3.9090.

[7] M. H. Rifki et al., "Text Mining for Sentiment Analysis of Film Reviews Using the Naïve Bayes Algorithm," JuSiTik, vol. 7, no. 2, 2024, doi: 10.32524/jusitik.v7i2.1168.

[8] A. Sintawati et al., "Performance Evaluation of SVM and Logistic Regression on Multiclass Data in the Sentiment Analysis of the Film Dirty Vote," RIGGS – Journal of Data Science, vol. 4, no. 2, pp. 821–832, 2025, doi: 10.31004/riggs.v4i2.576.

[9] S. U. Hasan, J. Ahamed, and K. Ahmad, "Analytics of Machine-Learning-Based Algorithms for Text Classification," Sustainable Operations and Computers, vol. 3, pp. 238–248, 2022, doi: 10.1016/j.susoc.2022.03.001.

[10] D. A. Moreno-Galván et al., "Automatic Movie Genre Classification and Emotion Recognition via a Bi-Projection Multimodal Transformer," Information Fusion, vol. 113, Art. no. 102641, 2025, doi: 10.1016/j.inffus.2024.102641.

[11] N. Pavitha et al., "Movie Recommendation and Sentiment Analysis Using Machine Learning," Global Transitions Proceedings, vol. 3, no. 1, pp. 279–284, Jun. 2022, doi: 10.1016/j.gltp.2022.03.012.

[12] H. Asyraf and M. E. Prasetya, "Implementation of the CRISP-DM Method and Decision-Tree Algorithm for the Handicraft Production Strategy at UMKM A," Budidarma Journal of Media Informatics, vol. 8, no. 1, pp. 94–105, Jan. 2024, doi: 10.30865/mib.v8i1.7050.

[13] M. T. Razaq, D. Nurjanah, and H. Nurrahmi, "Sentiment Analysis of Film Reviews Using a Naïve Bayes Classifier with TF-IDF Features," e-Proceeding of Engineering, vol. 10, p. 1698, 2023.

[14] N. Zelina and A. Afiyati, "Sentiment Analysis of Mobile-Banking User Reviews Using Support Vector Machine and Decision Tree Algorithms," Journal of Computational Linguistics (JLK), vol. 7, no. 1, pp. 31–37, Mar. 2024

[15] N. J. Ramadhan, V. A. P. Putri, and D. A. Riyadi, “Exploring Sentiment Analysis of IMDb Film Ratings: A Comparative Approach Using Decision Tree and Naïve Bayes,” INNOVATIVE: Journal of Social Science Research, vol. 4, no. 3, pp. 7273–7286, 2024.

[16] S. Navisa, L. Hakim, and A. Nabilah, “Comparison of Music-Genre Classification Algorithms on Spotify Using CRISP-DM,” Journal of Intelligent Systems, vol. 4, no. 1, pp. 114–125, Aug. 2021.

[17] V. Indumathi, S. Santhanamegala, and R. Padmapriya, “Classifying Bullying Text with an Improved Model Using Grid-Search Hyperparameter Tuning,” Advances and Applications in Mathematical Sciences, vol. 21, no. 9, pp. 4973–4980, Jul. 2022.

[18] R. K. Pratama and P. W. Setyaningsih, “Analysis of Twitter Comments on Employment Opportunities Using the Naïve Bayes Method,” Journal of Information System and Artificial Intelligence, vol. 3, no. 2, p. 217, May 2023.

[19] A. Kurniasih and C. N. Previana, “Using GridSearchCV to Obtain Optimal Hyperparameters for Classification Models in Predicting Water Potability,” Journal of Artificial Intelligence and Engineering Applications, vol. 4, no. 2, pp. 2808–4519, Feb. 2025.

[20] D. A. Mukhsinin, M. Rafliansyah, S. A. Ibrahim, R. Rahmaddeni, and D. Wulandari, “Implementation of the Decision Tree Algorithm for Film Recommendation and Rating Classification on the Netflix Platform,” MALCOM: Indonesian Journal of Machine Learning and Computer Science, vol. 4, no. 2, pp. 570–579, Mar. 2024, doi: 10.57152/malcom.v4i2.1255.

[21] Hidayat, E., Susilo, D., & Mujiono (2023). “Are Audience Preferences for Electronic Cinema Affected by Preferences in Indonesian Movie Choices?” Nyimak Journal of Communication, 7(2), 297–320.

[22] A. N. K. Krisna and U. Salamah, “Comparison of Naïve Bayes and K-Nearest Neighbor Algorithms for Classifying Health-Hoax News on Twitter,” KAPUTAMA Journal of Informatics Engineering (JTIK), vol. 6, no. 2, pp. 836–845, Jul. 2022.

[23] P. Alkhairi, A. P. Windarto, M. Kom, and M. M. Efendi, “Optimising LSTM to Reduce Overfitting for Text Classification Using the Kaggle IMDb Film-Review Dataset,” Building of Informatics, Technology and Science (BITS), vol. 6, no. 2, pp. 1142–1150, 2024.

[24] N. Khoirunnisaa, K. N. N. Kesuma, S. Setiawan, and A. Y. P. Yusuf, “Text Classification of Netflix Application Reviews on Google Play Store Using Naïve Bayes and SVM Algorithms,” SKANIKA, vol. 7, no. 1, pp. 64–73, Jan. 2024.

[25] D. I. Gunawan, A. Rahardjo, and D. Pratiwi, “Optimising Sentiment-Analysis Performance with Grid Search on Shopee Reviews,” Indonesian Journal of Computer Science, vol. 8, no. 1, pp. 15–22, 2023.