Hierarchical Topic Mining and Multi-label Classification on Online News in Bahasa

Ratih Nur Esti Anggraini
Informatics Department
Institut Teknologi Sepuluh Nopember
Surabaya, Indonesia
ratih_nea@if.its.ac.id

Hana Machmudah
Informatics Department
Institut Teknologi Sepuluh Nopember
Surabaya, Indonesia
machmudahhana@gmail.com

Riyanarto Sarno
Informatics Department
Institut Teknologi Sepuluh Nopember
Surabaya, Indonesia
riyanarto@if.its.ac.id

2023 International Conference on Advanced Mechatronics, Intelligent Manufacture and Industrial Automation (ICAMIMIA) | 979-8-3503-0922-5/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICAMIMIA60881.2023.10427844

***Abstract*—The proliferation of online news media has resulted in a surge of online news. Consequently, properly categorizing news based on their respective topics becomes crucial, enabling readers to analyze and comprehend the information more easily. In this study, the researchers conducted topic mining using the Joint Spherical Tree and Text Embedding (JOSH) method, a technique that utilizes spherical space for topic modeling. This method involves embedding a tree structure into the spherical space, with each category being surrounded by its representative terms in the same direction within that space. The resulting topic hierarchy was then employed for multi-label classification of Bahasa news texts acquired through web scraping from various news sites. The most successful topic mining outcomes exhibited a topic coherence value of 0.7882. Subsequently, a Long Short Term Memory (LSTM) model was developed for multi-label classification, yielding good results, including a hamming loss value of 0.9229 and a recall score of 0.9982.**

Keywords—topic mining, topic hierarchy, multi-label classification, online news

I. Introduction

The escalating volume of news can pose challenges for users seeking information. To address this, news can be organized into topic-based groups, facilitating easier analysis and comprehension of related information. This process is referred to as topic mining, also known as topic modeling, which is a text mining technique used for data mining, latent data discovery, and identifying relationships between data and text documents [1].

Numerous research studies have been conducted on topic mining, aiming to analyze social media data, research publications, and various other texts [2]. Among the commonly utilized methods in topic modeling is Latent Dirichlet Allocation (LDA) [3]. However, LDA suffers from certain limitations, such as instability and the necessity for extensive text preprocessing to achieve desirable results [4].

Another popular approach in topic mining is hierarchical topic mining, which allows for multiple paths through the tree by leveraging information at the sentence level [5]. Meng et al. introduced the Joint Spherical Tree and Text Embedding (JOSH) method, which builds a topic tree based on topic names and explores representative terms for each topic using the text corpus, primarily composed of English articles [6].

On the other hand, it is important to categorize online news to enhance readers' comprehension and analysis. In the context of online news classification, it is essential to recognize that a single article can encompass multiple topics [7]. Consequently, the objective is to forecast the various topics to which a news article is relevant. This stands in contrast to traditional single-label classification, where each instance is assigned to a singular category.

Among the well-known techniques for multi-label classification is the employment of Long Short-Term Memory (LSTM) [8]. LSTM has gained popularity as a favored method due to its proficiency in handling multi-label assignments. Unlike traditional models, LSTM is specifically designed to process sequential data, making it highly suitable for analyzing text and natural language tasks.

The focus of this study is on using the JOSH method [6] to construct a hierarchy of Bahasa news topics, which will be evaluated using topic coherence. Additionally, this topic hierarchy will serve as the basis for conducting multi-label classification on the collected news dataset. The classification is conducted using LSTM.

The remainder of the paper is organized as follows: Section 2 provides a comprehensive literature review on related topics, followed by a detailed description of the method used in this research in Section 3. Section 4 presents and discusses the results, and finally, Section 5 concludes the research with a summary of the findings. The proposed approach has the potential to enhance the understanding and classification of news articles in the Bahasa, contributing to more efficient information retrieval and analysis.

II. Literature Review

A. Hierarchical Topic Mining via Joint Spherical Tree and Text Embedding (JOSH)

Hierarchical topic modeling is a statistical approach used to uncover abstract "topics" within a collection of documents and subsequently organize them in a hierarchical manner [9]. In this modeling process, the more abstract topics are positioned closer to the root of the hierarchy, while the more specific and concrete topics are located near the leaves. The hierarchical structure derived from this modeling can be harnessed for a multitude of applications, including document classification, information retrieval, and recommendation systems.

The Joint Spherical Tree and Text Embedding technique represents a novel model that combines text embedding and tree embedding in spherical space, aiming to facilitate Hierarchical Topic Mining [6]. By concurrently modeling the structure of the topic tree and the corpus text within the spherical space, this model capitalizes on the similarity of directions to establish semantic connections between words, documents, and categories, ultimately enabling the extraction of representative terms for various categories. Maintaining a hierarchical category structure within the spherical embedding space fosters inter-category specificity, ensuring that terms possess distinctive and clear descriptions. Additionally, the intra-category coherence highlights the semantic relevance between the representative terms of each category, further enhancing the overall efficiency and accuracy of the method.

A notable aspect of this approach is its recursive embedding of the local category tree into the sphere, which effectively preserves the relative relationships between categories. By adhering to the principle of maintaining relative tree spacing within the local tree, each category topic remains closer to its parent topic than to its sister topics within the embedding space. This mechanism ensures a coherent and organized representation of category relationships, facilitating more effective retrieval and analysis of topics in the hierarchical structure. Through the Joint Spherical Tree and Text Embedding method, Hierarchical Topic Mining gains valuable insights into the underlying semantic correlations within the corpus, paving the way for improved information retrieval and knowledge discovery.

B. *Multi-label Classification*

Multi-label classification expands upon the conventional single-label classification by allowing each data instance to be associated with multiple labels simultaneously [10]. In this supervised learning approach, data points can possess more than one label, leading to the task of assigning a set of labels to each specific data instance [11], [12]. The primary objective of multi-label classification is to accurately predict and assign the relevant combination of labels that best represent the characteristics or attributes of the given data. By embracing the complexity of multiple labels, this approach caters to scenarios where data points can belong to diverse and overlapping categories, resulting in a more comprehensive and nuanced classification system.

Some machine learning methods have been employed to perform multi-label classification, including SVM [13] or KNN [14]. Another effective method for multi-label classification is LSTM [15], which proves particularly useful for multi-label classification tasks involving text data due to its ability to retain information over extended period. LSTMs are composed of gated structures where data are selectively forgotten, updated, stored, and outputted. They have a special cell state called Ct, where information flows and three special gates: the forget gate, the input gate, and the output gate. LSTMs gates are continually updating information in the cell state.

III. PROPOSED APPROACH

A. *System Architecture*

Fig 1 illustrates the system architecture, while Fig 2 outlines the proposed method. The news dataset used in this study was collected through web scraping from various online news sites. The dataset underwent preprocessing steps, including case folding, normalization, stopword removal, and stemming. Furthermore, Named Entity Recognition (NER) was applied to cleanse person and location information.

The subsequent stage involves topic mining using the Joint Spherical Tree and Text Embedding technique. The topic hierarchy is assessed using topic coherence, and the best outcome is combined with the New York Times (NYT) topic hierarchy. This combined hierarchy is then utilized for multi-label classification of online news in Bahasa. The topic mining and multi-label classification process will be explained in the next section in more detail.

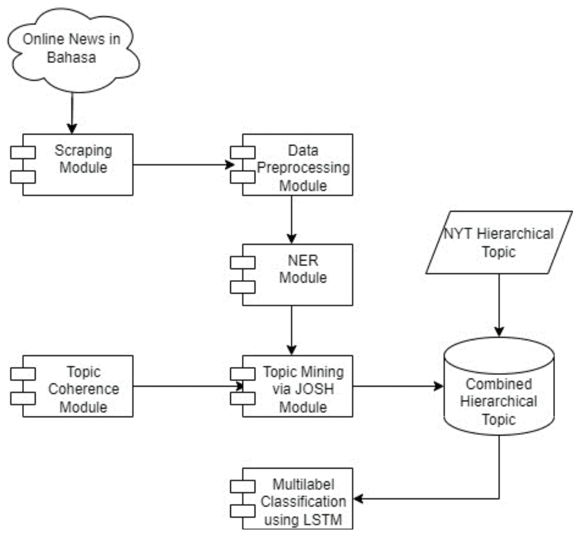

Fig 1. System Architecture

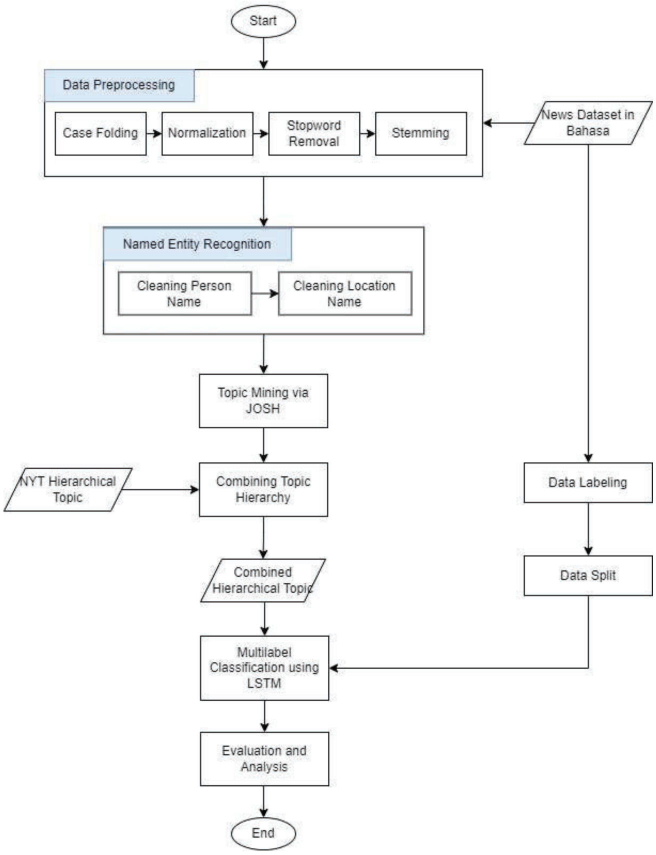

Fig 2. Proposed Approach

B. Topic Mining

The Topic Mining process using the Joint Spherical Tree and Text Embedding (JOSH) involved several steps, beginning with data preparation as the input model, leading to the generation of results in the form of a topic mining hierarchy containing ten representative terms for each category. Fig 3 visually depicts the entire process of topic mining using the Joint Spherical Tree and Text Embedding. The input data consists of three components: news description data, news taxonomy, and news topic names. The outcome of this process is the topic mining hierarchy accompanied by ten representative terms associated with each topic.

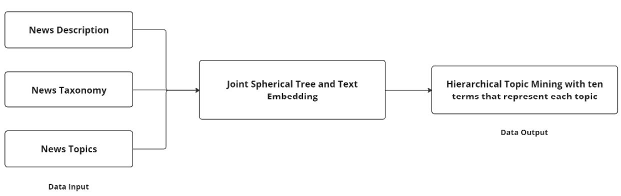

Fig 3. Topic Mining Process

The corpus used in the model consists of news description data extracted from the scraping process. Prior to applying the JOSH method for data modeling, preprocessing is performed, including the cleaning of people's and location names derived from the NER results, ensuring improved outcomes. The processed data is then stored in a txt file, with each news description occupying a single line.

The subsequent input data comprises news taxonomy and topic names, encompassing six main topics in Bahasa news: *Bencana* (disaster), *Ekonomi* (economy), *Kecelakaan* (accident), *Kesehatan* (health), *Kriminalitas* (criminals), and *Olahraga* (sports). The topic hierarchy is subsequently merged with the New York Times topic hierarchy. The hierarchy is further broken down into sub-topics to facilitate topic mining in this study, resulting in 20 distinct topics. Table I presents the names of these 20 topics alongside the taxonomy data, which serves as the input for JOSH mining in this research.

The outcome of the topic mining process is assessed using topic coherence, which serves as a framework to evaluate the quality of the generated topics. A reliable model is expected to produce topics with high coherence scores. To calculate the topic coherence, a systematic search of coherence measures is performed using all available topics to evaluate the data's relevance [16]. The equation for measuring the confirmation measure, which determines the proximity relationship between two sets of topic terms, can be observed in Equation 1.

$$NPMI(w_i, w_j)^{\gamma} = \left(\frac{log \frac{P(w_i, w_j) + e}{P(w_i).P(w_j)}}{-log(P(w_i).P(w_j) + e)} \right)^{\gamma} \quad (1)$$

Where w_i is a topic mining term and w_j is another topic mining term. The value of *e* is used to calculate the logarithm of zero. γ is the excess weight at a higher NPMI value, and NPMI is normalized pointwise mutual information.

C. Multi-label Classification

Fig 4 illustrates the multi-label classification process, which begins with dataset preparation. This study uses a combined dataset of Bahasa news and news from the New York Times (NYT). To begin with, the NYT news is translated into Bahasa. Before classifying the merged news data, dataset labeling is performed based on the topic mining hierarchies' outcomes. The labeling is conducted by examining the news descriptions, and if a description contains three or more terms from a single topic, it will be labeled accordingly with that topic.

Subsequently, preprocessing is carried out, involving case folding and stopword removal stages, before proceeding with the multi-label classification. The classification results will be evaluated using a hamming loss matrix and a recall score obtained from the sklearn library.

To evaluate the results of multi-label classification, we employed Hamming Loss and Recall. Hamming Loss serves as an evaluation metric used to quantify the prediction error or dissimilarity in each label between the predicted and actual labels [17]. It yields values ranging from 0 to 1, with lower values indicating higher model prediction accuracy in multi-label data. The calculation for Hamming Loss (*HL*) is defined by Equation 2.

$$HL = \frac{1}{N.L} \sum_{i=1}^{N} \sum_{j=1}^{L} [\hat{Y}_j^i \oplus Y_j^i] \quad (2)$$

Where *N* is the number of data and *L* is the number of bit representations or the number of labels. Y_j^i is prediction label and $\hat{Y}_j^i$ is the actual label.

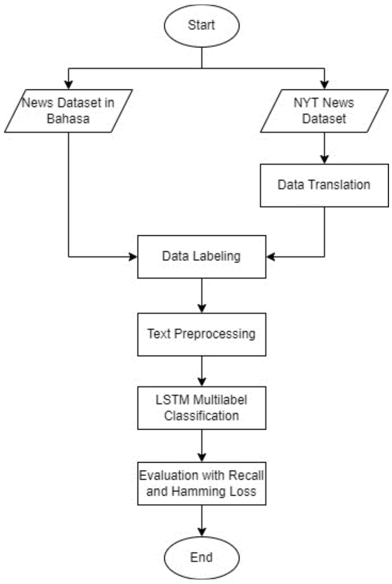

Fig 4. Multi-label Classification Process

In addition, we also calculate the recall. The Recall Score measures the model's ability to identify the positive class correctly from the total data including the positive class identified by the model [18]. Equation 3 represents recall calculation.

$$RC = \frac{TP}{TP + FN} \quad (3)$$

Where TP is True Positive and FN is False Negative.

IV. Experiment and Evaluation

In this study, the experimental process involves the use of two datasets. These datasets are categorized into the topic mining dataset (Table II) and the multi-label classification dataset (Table III).

TABLE 1. Topic Hierarchy and Taxonomy

Topics	Taxonomy
0 ROOT	0 1
1 olahraga	0 2
2 ekonomi	0 3
3 kecelakaan	0 4
4 kriminalitas	0 5
5 bencana	1 6
6 bulutangkis	1 7
7 voli	1 8
8 basket	1 9
9 tenis	4 10
10 pembunuhan	4 11
11 pencurian	5 12
12 gempa	5 13
13 kebakaran	5 14
14 tsunami	5 15
15 gunung_meletus	5 16
16 banjir	5 17
17 putting_beliung	5 18
18 kekeringan	5 19
19 abrasi	5 20
20 longsor	

TABLE II. Topic Mining Dataset

Bahasa News Dataset	45731
Number of Topics	20

TABLE III. Multi-label Classification Dataset

News in Bahasa Dataset	1883
NYT Dataset	1644
Number of Topics	39
Total Data	3527
Training dataset	2821
Testing dataset	706

A. *Topic Mining Results*

Various experiment scenarios were conducted for topic mining in this research. The dataset conditions play a significant role in shaping the topic mining model. The scenarios were derived from preprocessing outcomes to identify the best model, characterized by a topic coherence evaluation value approaching one. The most successful scenario obtained from topic mining will be employed to label the data in the subsequent multi-label classification process. For this study, topic mining was applied with 20 topics, as listed in Table 1. The experiment scenarios are presented as follows.

1. Scenario 1: Preprocessing used includes case folding, normalization, dan cleansing person name and location.
2. Scenario 2: Preprocessing used includes case folding, normalization, stopword removal, dan cleansing person name and location.
3. Scenario 3: Preprocessing used includes case folding, normalization, stopword removal, stemming dan cleansing person name and location.

TABLE IV. Topic mining scenario

Model	**Topic Coherence**
Scenario 1	0,7709
Scenario 2	**0,7882**
Scenario 3	0,6826

As shown in Table IV, scenario 2 emerged as the most optimal model out of the three scenarios, as indicated by its topic coherence evaluation score of 0.7882. Consequently, the topic mining results from scenario 2 were selected for integration with the New York Times hierarchy. The combined outcomes are presented in Table V.

TABLE V. Topic mining and its terms

Topic	**Term**
olahraga	futsal, tenis, bulutangkis, olimpiade, menpora, voli, bola, basket, cabor, kemenpora
olahraga. bulutangkis	olimpiade, pebulutangkis, prestasi, pbsi, futsal, atlet, bwf, ganda_putri, thomas_cup, ganda_campuran
olahraga. voli	pon, pbvsi, olahraga, kegiatan, cabang, bola, proliga, asian_games, sea_games, timnas
olahraga. basket	league, liga, pebasket, ibl, kompetisi, proliga, klub, perbasi, ibl, tim
olahraga. tenis	as_terbuka, australia_terbuka, raket, grand_slam, wimbledon, petenis, prancis_terbuka, wta, atp, titel
olahraga. golf	pegolf, sembilan lubang, fairways, pemain golf, golf_courses, rumput gergaji, klub golf, bermain golf, 18 lubang, tiger_woods

olahraga. sepakbola	piala, turnamen, final, semifinal, perempat final, stiker, soccer_federation, juara, kejuaraan, kejuaraan
olahraga. bisbol	yankee, bertemu, pelaut, red_sox, yankee, anggukan, dodgers, kendi, pemain luar, pemain bola
olahraga. hoki	mogilny, pemain_bertahan, olahraga, tim hoki, n.h.l, atletik, atletis, collegiate, lindros, canucks
ekonomi	penurunan, dampak, tumbuh, pemulihan, meningkat, pertumbuhan, memitigasi, resesi, inflasi, global
kecelakaan	kapolres, polres, polisi, kejadian, lokasi, ditlantas, maut, beruntun, jalan, solo
kriminalitas	tindakan, kepolisian, kejahatan, mencegah, terorisme, aksi, hukum, kekerasan, wartawan, tindak
Kriminalitas. pembunuhan	Membunuh, dibunuh, merencanakan, motif, tega, menghabisi, keji, pembunuh, terungkap, fakta
kriminalitas. pencurian	dicuri, maling, pencurian, pemberatan, digondol, curat, curanmor, beraksi, dibobol, pencurian
bencana. gempa	bermagnitudo, diguncang, magnitudo, mengguncang, berpusat, berkekuatan, bumi, guncangan, gempabumi, susulan
bencana. kebakaran	korsleting, terbakar, damkar, gulkarmat, hangus, memadamkan, pemadam, pemadaman, dipadamkan, api
bencana. tsunami	hidrometeorologi, inatews, early_warning_system, peringatan_dini, bencana, peringatan, gunung_meletus, bmkg, perairan, gelombang
bencana. gunung meletus	gunungnya, magmatik, vulaknik, erupsi, merapi, letusan, freatik, meletus, sinabung, semeru
bencana. banjir	meluapnya, merendam, terendam, luapan, bandang, menggenangi, meluap, kebanjiran, surut, rob
bencana. puting beliung	pusaran, hujan_es, ainginnya, berputar, kencang, pohon, angin, menerbangkan, petir, beterbangan
bencana. kekeringan	air_bersih, dropping, kemarau, tangki, pdam, air_bersihnya, truk_tangki, rit, menyuplai, sumur
bencana. abrasi	mangrove, bakau, ditanam, biota_laut, bibir_pantai, menanam, penanaman, penghijauan, ditebang, pengikisan
bencana. longsor	menimbun, longsornya, longsoran, tertimbun, tertutup, material, tebing, bawahnya, timbunan, tempursari
pendidikan	kurikulum, keterlibatan orang tua, petunjuk, berbasis sekolah, kurikulum, mendidik, instruksional, sekolah, dasar
teknologi	komputer, teknologi, desktop, komputer, perangkat_lunak, digital, komputer_pribadi, elektronik, pc, chip
politik	demokratis, ideologi, republik, berpesta, ideologis, partisan, politik, liberal, konservatif, demokrat
kesehatan	peduli, insurance_coverage, perawatan kesehatan, kesehatan, aids, kesehatan_mental, pasien, rumah jompo, anak, medicaid
sains	fisika, biologi, peneliti, kimia, universitas, ilmuwan, profesor, ilmu, astronomi, sosiologi
bisnis	karyawan, perusahaan, firma, bisnis, broker, investasi, grosir, industri, pekerjaan
bisnis.bisnis kecil	bisnis_kecil, pensiunan, wiraswasta, upah_rendah, pengusaha, pemberi_kerja, berpenghasilan_rendah, upah minimum, usaha_kecil, keuntungan
bisnis.media	editor, publikasi, koran, penerbit, majalah, surat_kabar, televisi, kolumnis, siaran, radio
bisnis.pasar	saham, bursa, pasar, jual beli, sekuritas, pedagang, investor, mata uang, obligasi_kota, pasar obligasi
seni	seni, kontemporer, galeri, pameran, lukisan, artis, museum, teater, klasik
bisnis.desain	arsitektur, desainer, arsitek, modernis, desain, bangunan, pematung, interior
bisnis.musik	chamber_orchestra, orkestra, simfoni, orkes simfoni, ansambel, lagu, gitar, melodi, jazz
bisnis.tari	balet, penari, tari_modern, penari, tarian, koreografer, tarian, rombongan, grup_dansa
bisnis.film	film, pembuat_film, dokumenter, komedi, hollywood, blockbuster
bisnis.teater	roundabout, teater, shubert, ruang_teater, broadway, produksi, musikal, mccarter, lortel, studio_teater

B. *Multi-label Classification Results*

During the multi-label classification, an experiment was conducted to classify the dataset independently for both the news in Bahasa and the New York Times dataset. The results for the New York Times dataset are provided in Table 6, where the model achieved the best hamming loss value of 0.4717 and the highest recall score of 0.9982. On the other hand, the test results for the Bahasa news dataset can be found in Table 7, with the model demonstrating the best hamming loss value of 0.4785 and the top recall score of 0.9111.

TABLE VI. Multi-label Classification result on NYT dataset

Batch Size	Epoch	Hamming Loss	Recall Score
32	20	0.5329	0.9982
64	50	0.7109	0.9484
64	20	0.4717	0.9542
64	10	0.4717	0.9552
64	100	0.8143	0.9724

TABLE VII. Multi-label classification result on news in Bahasa

Batch Size	Epoch	Hamming Loss	Recall Score
32	20	0.4785	0.9111
64	50	0.4786	0.9111
64	20	0.4785	0.9111
64	10	0.4785	0.9111
64	100	0.4785	0.9111

Subsequently, an experiment was conducted on the combined dataset, and the results are presented in Table VIII. In the multi-label classification of the combined dataset, the model with the best hamming loss value achieved 0.9229, and the highest recall score value was 0.9982. Despite performing experiments on the combined dataset, the model's performance still falls short as the hamming loss value remains high, exceeding 0.9. It is suspected that the data labeling process might have introduced errors. This could be attributed to the dataset from the New York Times, which was translated into Bahasa, containing terms from in Bahasa topic mining, leading to the dataset being labeled with Bahasa news topics. Conversely, the Bahasa news dataset may include terms from New York Times topic mining, resulting in the dataset being labeled with New York Times topics.

TABLE VIII. Multi-label Classification on Combined Dataset

Embedding Dim	LSTM Units	Batch Size	Epoch	Hamming Loss	Recall Score
100	100	32	10	0.9541	0.9390
100	100	32	20	0.9541	0.9390
200	100	64	20	0.9501	0.9640
200	128	64	20	0.9541	0.9390
100	100	64	100	0.9541	0.9544
100	100	64	200	0.9488	**0.9829**
100	128	64	300	0.9414	0.9585
100	100	64	10	0.9389	0.8506
200	100	64	300	0.9280	0.9528
200	100	64	100	0.9433	0.9633
200	100	64	50	0.9510	0.9701
200	100	64	20	0.9528	0.9709
200	100	64	10	0.9277	0.8764

200	128	64	20	0.9571	0.9578
200	128	64	50	0.9540	0.9724
200	128	64	10	0.9316	0.9458
200	128	64	50	0.9454	0.9636
200	128	64	50	**0.9229**	0.9316

CONCLUSION AND FUTURE WORK

Based on the implementation of the research, the following conclusions can be drawn. Preprocessing or text preprocessing is an essential step before modeling the data. The preprocessing process involves several stages, including case folding, normalization, stopword removal, and stemming, which are applied to the topic mining datasets derived from scraping Bahasa online news.

The topic mining hierarchy using the Joint Spherical Tree and Text Embedding (JOSH) method is obtained by collecting news datasets, followed by data preprocessing and cleaning of people's and location names using NER. The topic mining results are then evaluated using topic coherence, with the best evaluation score being 0.7882.

The Long Short Term Memory (LSTM) method is utilized to obtain a multi-label classification model, and the datasets are labeled based on the topic mining hierarchical terms. Additionally, the English dataset is translated into Bahasa, and preprocessing is conducted using case folding and stopword removal. The experiment involves cross-validation and hyperparameter tuning, resulting in the model with the best hamming loss value of 0.9229 and the top recall score of 0.9982 from the combined dataset.

For future research, a focus on reducing the high hamming loss value is recommended. The data labeling process based on topic mining results can be further explored. Furthermore, experiments could involve incorporating News in Bahasa datasets related to New York Time topics or exploring alternative methods apart from Long Short Term Memory.

ACKNOWLEDGMENT

This research is funded by the Institut Teknologi Sepuluh Nopember (ITS) through the Department Research grant 2023.

REFERENCES

[1] H. Jelodar *et al.*, “Latent Dirichlet allocation (LDA) and topic modeling: models, applications, a survey,” *Multimed Tools Appl*, vol. 78, pp. 15169–15211, 2019.

[2] C. D. P. Laureate, W. Buntine, and H. Linger, “A systematic review of the use of topic models for short text social media analysis,” *Artif Intell Rev*, pp. 1–33, 2023.

[3] J. W. Lee, Y. Kim, and D. H. Han, “LDA-based topic modeling for COVID-19-related sports research trends,” *Front Psychol*, vol. 13, p. 1033872, 2022.

[4] A. Agrawal, W. Fu, and T. Menzies, “What is wrong with topic modeling? And how to fix it using search-based software engineering,” *Inf Softw Technol*, vol. 98, pp. 74–88, 2018.

[5] L. Liu, L. Tang, W. Dong, S. Yao, and W. Zhou, “An overview of topic modeling and its current applications in bioinformatics,” *Springerplus*, vol. 5, no. 1, pp. 1–22, 2016.

[6] Y. Meng, Y. Zhang, J. Huang, Y. Zhang, C. Zhang, and J. Han, “Hierarchical topic mining via joint spherical tree and text embedding,” in *Proceedings of the 26th ACM SIGKDD international conference on knowledge discovery & data mining*, 2020, pp. 1908–1917.

[7] X. Li *et al.*, “Weighted multi-label classification model for sentiment analysis of online news,” in *2016 International conference on big data and smart computing (bigcomp)*, IEEE, 2016, pp. 215–222.

[8] S. Feng, H. Li, and J. Qiao, “Hierarchical multi-label classification based on LSTM network and Bayesian decision theory for LncRNA function prediction,” *Sci Rep*, vol. 12, no. 1, p. 5819, 2022.

[9] P. Kherwa and P. Bansal, “Topic modeling: a comprehensive review,” *EAI Endorsed transactions on scalable information systems*, vol. 7, no. 24, 2019.

[10] A. N. Tarekegn, M. Giacobini, and K. Michalak, “A review of methods for imbalanced multi-label classification,” *Pattern Recognit*, vol. 118, p. 107965, 2021.

[11] H. Wang, W. Liu, A. Bocchieri, and Y. Li, “Can multi-label classification networks know what they don’t know?,” *Adv Neural Inf Process Syst*, vol. 34, pp. 29074–29087, 2021.

[12] R. Wang, R. Ridley, W. Qu, and X. Dai, “A novel reasoning mechanism for multi-label text classification,” *Inf Process Manag*, vol. 58, no. 2, p. 102441, 2021.

[13] Y. Liu, K. Wen, Q. Gao, X. Gao, and F. Nie, “SVM based multi-label learning with missing labels for image annotation,” *Pattern Recognit*, vol. 78, pp. 307–317, 2018.

[14] D. Urgun and C. Singh, “A hybrid Monte Carlo simulation and multi label classification method for composite system reliability evaluation,” *IEEE Transactions on Power Systems*, vol. 34, no. 2, pp. 908–917, 2018.

[15] Y. Hua, L. Mou, and X. X. Zhu, “Recurrently exploring class-wise attention in a hybrid convolutional and bidirectional LSTM network for multi-label aerial image classification,” *ISPRS journal of photogrammetry and remote sensing*, vol. 149, pp. 188–199, 2019.

[16] M. Röder, A. Both, and A. Hinneburg, “Exploring the space of topic coherence measures,” in *Proceedings of the eighth ACM international conference on Web search and data mining*, 2015, pp. 399–408.

[17] G. Wu and J. Zhu, “Multi-label classification: do Hamming loss and subset accuracy really conflict with each other?,” *Adv Neural Inf Process Syst*, vol. 33, pp. 3130–3140, 2020.

[18] R. Yacouby and D. Axman, “Probabilistic extension of precision, recall, and f1 score for more thorough evaluation of classification models,” in *Proceedings of the first workshop on evaluation and comparison of NLP systems*, 2020, pp. 79–91.