

# *A Multi-Task Learning Approach for News Classification and Number of Reader Prediction*

Helna Freecenta<sup>1</sup>, Chastine Fatichah<sup>2</sup>

<sup>1,2</sup>*Informatics Department, Institut Teknologi Sepuluh Nopember, Indonesia*

<sup>1</sup>6025221035@student.its.ac.id (\*)

<sup>2</sup>chastine@its.ac.id

Received: 2025-01-04; Accepted: 2025-01-30; Published: 2025-07-04

*Abstract*— The large volume of online news content presents a challenge in effectively managing and organizing information, especially regarding enhancing literacy rates in Indonesia. As the amount of news articles continues to grow, there is a need for a robust system that can categorize news and predict the number of readers to assess its impact on literacy. This study introduces a Multi-Task Learning (MTL) approach, utilizing data from online websites to simultaneously address news classification and reader prediction tasks. Cross-entropy loss is applied in the model to handle the class imbalance issue. The research compares two performance MTL architectures, the Dense architecture and the CNN architecture. The experiments assess the models' abilities to classify news and predict reader numbers. The results show that the Dense architecture outperforms the CNN architecture, achieving 94% accuracy and a 99% AUC-ROC score, whereas the CNN model achieved 91% accuracy and a 98% AUC-ROC score. This study highlights the effectiveness of the Dense architecture in classifying online news and predicting reader engagement. The findings provide valuable insights for enhancing news sorting systems and could contribute to improving literacy initiatives in Indonesia by offering more accurate predictive models for online news consumption. The results indicate that integrating Multi-Task Learning into news classification systems can enhance content management and offer a deeper understanding of public interaction with news.

*Keywords*— Multi-Task Learning; CNN; Classification News; Prediction News.

## I. INTRODUCTION

The Industrial Revolution, also known as Industry 4.0 [1], marks a significant shift driven by technologies such as the Internet of Things (IoT), cloud computing, big data, robotics, Artificial Intelligence (AI), machine vision, and machine learning. These technologies are crucial to developing modern systems that enhance efficiency, inform decision-making, and facilitate problem-solving. As industries increasingly rely on automation and data-driven solutions, the challenge lies in effectively leveraging these technologies to address complex problems that exceed human capabilities. Traditional approaches to designing security, education, and manufacturing systems are often insufficient, highlighting the need for Industry 4.0 technologies. Among these, AI [2] plays a pivotal role by simulating human cognitive abilities, enabling systems to exercise and make decisions independently. The advent of AI machine learning (ML) and deep learning (DL) subsets has further revolutionized industries by empowering systems to process vast amounts of data, solve complex tasks, and enhance overall performance. As a result, AI-driven innovations have become invaluable, particularly in information technology, where the ability to analyze and interpret large-scale data is crucial for operational success.

The rapid advancement in information technology has transformed the way news is consumed, particularly with the rise of digital platforms. The overwhelming volume of daily online articles makes it crucial to categorize news content effectively and predict the number of readers for each article. News categorization enables readers to quickly locate articles

of interest, while the accurate prediction of reader engagement can help content creators and marketers optimize their strategies.

Classifying news articles into relevant categories allows readers to easily find topics of interest while accurately predicting the number of readers. This can help content creators and marketers optimize their strategies. This process involves assigning articles to predefined categories based on their content, such as politics, sports, entertainment, finance, and other relevant topics. Traditional classification methods often rely on rule-based systems or manual tagging, which can be time-consuming and prone to errors [3]. However, the advantage of using machine learning, deep learning, and automated classification systems has significantly improved accuracy and scalability [4].

Alongside classification, predicting [5][6] the number of readers for a particular news article has gained attention in media analytics. Understanding how various factors, such as the article's content, headline, and external factors like the time of publication, influence reader engagement can provide valuable insights for news agencies. This prediction problem is particularly challenging due to the variability in online content reader behavior.

The evolution of news classification methods parallels advancements in technology. A study [7] aims to address the problem of clickbait detection by fine-tuning the Bidirectional Encoder Representations from Transformers (BERT) model, specifically applied to the Indonesian news headlines dataset called CLICK-ID. The research leverages IndoBERT, an advanced BERT-based language model trained explicitly for

the Indonesian language, as the pre-trained model. The effectiveness of the BERT-based classifiers is evaluated by comparing IndoBERT's performance with other pre-trained models, as well as testing it against two word-vector-based methods (bag-of-words and TF-IDF) and five machine learning classifiers (Naive Bayes, K-Nearest Neighbors, Support Vector Machine, Decision Tree, and Random Forest). The evaluation results demonstrate that all fine-tuned IndoBERT classifiers outperform the word-vector-based machine learning models in distinguishing clickbait from non-clickbait headlines. The best-performing model was IndoBERTBASE, which, after two training phases, achieved an accuracy of 0.8247. This result outperformed the SVM classifier with the bag-of-words model, which achieved an accuracy of 0.7607, representing a 6% improvement over the baseline.

Multi-task learning has also been researched by Pengfei Liu for text classification of product reviews on Amazon.com across various categories, including books, DVDs, and electronics. This method introduces an adversarial training model that combines discriminative networks to differentiate between shared and task-specific features. This forces the model to learn a cleaner separation between the two types of features and can enhance overall performance. The study conducted 16 experiments and showed the effectiveness of the proposed method [8].

[9] studied the research of classification text using multi-task learning; the study discusses the growing issue of cyberbullying on social media platforms. With the increasing usage of social media, detecting cyberbullying has become a significant challenge due to the vast volume of online content and the diverse forms of abusive behavior. The authors propose a multi-task learning (MTL) framework to address this challenge effectively by leveraging the power of MTL. The framework can simultaneously perform various tasks related to cyberbullying detection, such as sentiment analysis, hate speech detection, and offensive language identification, improving the model's overall performance [9], [10].

This research introduces a multi-task learning (MTL) approach that concurrently categorizes news articles and forecasts reader engagement. The proposed model aims to enhance classification accuracy and prediction efficiency by sharing parameters between these interrelated tasks. MTL is known to be highly effective when applied to connected tasks, as the shared representations enable better generalization of new, unseen data.

The motivation for this dual-output approach stems from the importance of classifying news articles into relevant categories and understanding their reach and dissemination among readers. Knowing the reader count is crucial for determining the effectiveness of news communication and its overall spread. The model aims to enhance classification and prediction performance by utilizing shared parameters.

Using Indonesian news headlines instead of English-language issues would provide a more localized, culturally grounded perspective, as noted in prior research [11]. This could offer insights into the unique challenges, priorities, and discussions within the Indonesian media landscape. However,

working with Indonesian headlines would require translation or specialized language skills for non-Indonesian speakers, adding an extra layer of complexity compared to using English-language sources. The model's classification output focuses on assigning news articles to predefined categories based on their content and features, enabling a better understanding of the thematic structure of the news corpus. Conversely, the regression output predicts the number of users or readers likely to engage with each article, providing valuable insights into its potential reach and popularity.

By integrating these two outputs, the model delivers a comprehensive analysis of news articles from both categorical and quantitative standpoints. The classification output facilitates the effective organization and labeling of articles, while the regression output assesses their potential impact on the audience. This dual-output approach empowers news organizations to make informed, data-driven decisions concerning content creation, distribution, and user engagement strategies.

The primary objective of this research is to evaluate the performance of an MTL model with two outputs: news classification and reader prediction. By evaluating various architectures, the aim is to identify the optimal configuration that simultaneously yields the best performance for both tasks. Unlike prior studies that focus on applying MTL to a single task, this research aims to enhance overall work efficiency concurrently by leveraging MTL for both tasks.

## II. RESEARCH METHODOLOGY

The flowchart in Fig.1 illustrates the research steps and methodologies employed.

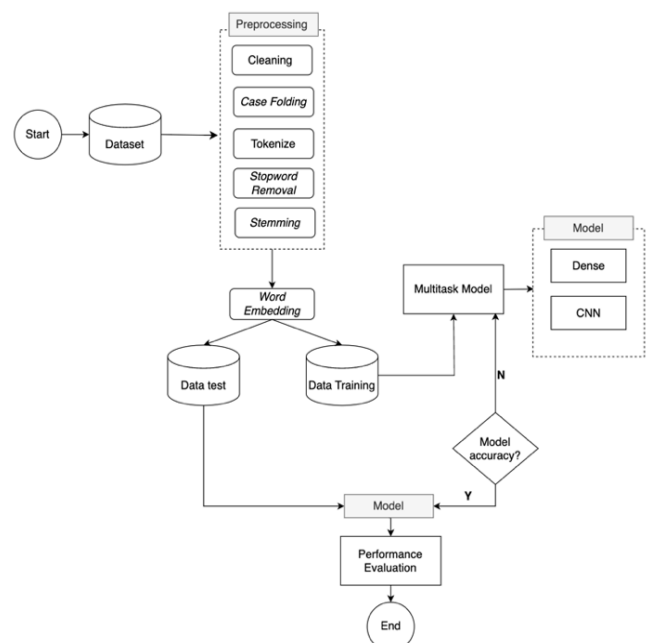


Fig.1 Research flowchart

The process began with the collection of a dataset comprising article titles and user information. The dataset was then passed

through pre-processing, which included cleaning punctuation by removing punctuation marks from the text data to eliminate noise and standardize the text format. Case folding converts all text to lowercase to ensure consistency and avoid treating the same words in different cases as separate entities. Tokenizing breaks down the text into individual words or tokens, the basic unit for further processing and analysis. Remove stopwords, eliminating common words that do not convey significant meaning, such as articles, prepositions, and conjunctions, to focus on the important content words. Stemming reduces words to their base or root form, grouping different variations of the same word and thereby reducing the dimensionality of the text data.

After the pre-processing, the dataset was processed for word embedding using Word2vec. After that, the dataset was divided into training data and testing data with an 80:20 ratio. Following this division, the multi-task architecture was implemented using both Dense and Convolutional Neural Network (CNN) architectures.

#### A. Dataset

The dataset used in this study was derived from internal sources from some online websites company [12], containing 194.635 article titles with classifications shown in Table I. Classification classes of articles in this research were divided into five categories, namely, finance, news, automotive, sports, and celebrity, with each class comprising a different article title to simplify the classification process, the dataset was processed handling imbalance using class weights for the classification task.

TABLE I  
 DISTRIBUTION OF ARTICLES IN EACH CATEGORY

Categories	Count of Articles
Sports	82,664
Celebrity	45,860
News	35,984
Finance	26,587
Automotive	3,540

#### B. Word Embedding

Word2Vec [13] includes two methods: Continuous Bag of Words (CBOW) and Skip-Gram. These approaches utilize neural networks to map the words to target variables, which are also words. Both methods represent words using weighted vectors. Word2Vec can capture semantic and syntactic relationships between words in a vector space.

#### C. Architecture

1) *Dense Multi-task Learning*: Multi-task learning (MTL) methods are usually divided into two main approaches, which depend on how the model shares its representation in hard and soft parameters. Most techniques use a single branching point for task-specific heads [14]. However, this simple branching approach may not always provide optimal results, sparking interest in more intricate architectural designs [15]. In soft parameter sharing, each task is assigned a unique set of parameters, utilizing a feature-sharing mechanism that enables communication between functions, as shown in Fig.2.

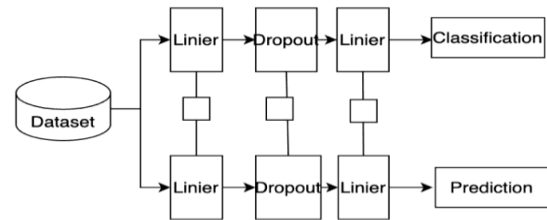


Fig.2 Dense Multi-task Model

2) *Convolutional Neural Network (CNN) Multi-task Learning*: CNN models [16], [17] are inspired by the functioning of the human brain. CNNs are particularly effective at processing data with grid-like structures. The architecture of CNNs operates in three dimensions: one for processing text and signals, two for images or audio, and three for analyzing video content. CNNs are also highly effective in text classification. In this case, the input consists of word vectors created using a word concatenation technique. The architecture used in this study is illustrated in Fig.3.

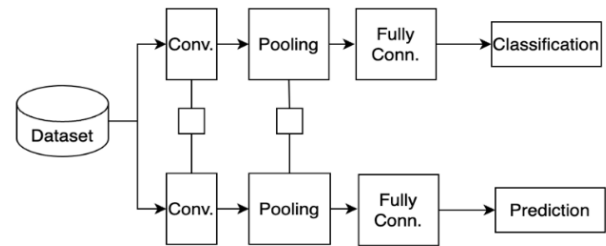


Fig.3 CNN Multi-task Model

#### D. Model Evaluation

The training accuracy from each experiment was examined using the training dataset in the final stage. Each task is evaluated differently in terms of effectiveness. For classification in this research, use accuracy and AUC-ROC [18][19]. The formulation proven was used to determine classification.

AUC-ROC is the true positive rate (TPR), also known as recall or Sensitivity, which indicates the percentage of positive cases that the model correctly identifies, as shown in Equation (1). The false positive rate (FPR) in Equation (2) quantifies the proportion of negative cases incorrectly classified as positive by the model.

$$TPR = \frac{\text{True Positives (TP)}}{\text{False Positive (FP)} + \text{True Negatives (TN)}} \quad (1)$$

$$FPR = \frac{\text{False Positive (FP)}}{\text{False Positives (FP)} + \text{True Negatives (TN)}} \quad (2)$$

$$\text{Accuracy} = \frac{\text{True Positives (TP)} + \text{True Negatives (TN)}}{\text{Total Samples}} \quad (3)$$

Accuracy is shown in Equation (3) and is used to evaluate the performance of the classification model. In other words, accuracy measures how often the model is correct across all classification predictions.

Alongside classification, to evaluate the model for task prediction in this research, using Equation (4) for Root Mean Squared Error (RMSE), Equation (5) to find Mean Absolute Error (MAE), and Equation (6) to find Mean Squared Error (MSE) [20]. The formulation provided in the Equation was used to determine the prediction.

$$MSE = \sqrt{\frac{\sum_{t=1}^n (x_t - f_t)^2}{n}} \quad (4)$$

$$MAE = \frac{\sum_{t=1}^n (y_t - \hat{y}_t)}{n} \quad (5)$$

$$RMSE = \sqrt{\frac{\sum_{t=1}^n (x_t - f_t)^2}{n}} \quad (6)$$

### III. RESULT AND DISCUSSION

This section presents the research results obtained based on the data analysis that has been carried out. The results displayed include the main findings obtained from the experiments, tests, or analyses carried out

#### A. Pre-processing

Before proceeding with further data analysis, a pre-processing stage is necessary to ensure the quality of the data used in the research. Table II shows the distribution of articles across different categories after handling class imbalance in both the training data and the test data. The dataset is split into 80% training data and 20% testing data. The data training includes categories such as sports (66,136 articles) and celebrity (36,665 articles), with the highest number of instances. The data test consists of a smaller portion of the data, with sports and celebrities having the most significant counts. In contrast, the automotive sector has the smallest representation in both sets, with 2,825 articles in the training data and 715 in the testing data.

TABLE II  
DATA TRAIN AND TEST

Categories	Dataset	
	Train	Test
Sports	36,665	9,195
Celebrity	21,269	5,318
News	28,813	7,171
Finance	2,825	715
Automotive	66,136	16,528

TABLE III  
RESULTS PRE-PROCESSING

Article Title	Metrics Evaluation			
	Tokenize	Stopword Removal	Stemming	Result
9 Produk Israel yang Dijual di Indonesia	['produk', 'israel', 'yang', 'dijual', 'di', 'indonesia']	['produk', 'israel', 'dijual', 'indonesia']	['produk', 'israel', 'jual', 'indonesia']	produk israel jual indonesia

Article Title	Metrics Evaluation			
	Tokenize	Stopword Removal	Stemming	Result
Masyarakat at Bisa Ajukan Sendiri sebagai Penerima BLT BBM Rp600.000, Begini Caranya	['masyarakat', 'bisa', 'ajukan', 'sendiri', 'sebagai', 'penerima', 'blt', 'bbm', 'begini', 'caranya']	['masyarakat', 'ajukan', 'penerima', 'blt', 'bbm']	['masyarakat', 'aju', 'terima', 'blt', 'bbm']	masyarakat at aju terima blt bbm
Masyarakat at Bisa Ajukan Sendiri sebagai Penerima BLT BBM Rp600.000, Begini Caranya	['masyarakat', 'bisa', 'ajukan', 'sendiri', 'sebagai', 'penerima', 'blt', 'bbm', 'begini', 'caranya']	['masyarakat', 'ajukan', 'penerima', 'blt', 'bbm']	['masyarakat', 'aju', 'terima', 'blt', 'bbm']	masyarakat at aju terima blt bbm

The research categorizes news into different sections and predicts readership based on data from two years of news titles. Four text pre-processing methods were employed to clean and enhance the data while preserving crucial information. Table III presents a sample of the post-processing steps, illustrating how these techniques were applied to enhance data quality and prepare it for subsequent analysis.

#### B. Word Embedding

Table IV shows the results of using Word2Vec in each article title embedded using Word2Vec in 100 dimensions.

TABLE IV  
RESULTS WORD2VEC

Article Title	Word2vec				
	1	2	...	99	100
produk israel jual indonesia	0.01	1.81	...	0.11	0.18
masyarakat aju terima blt bbm	0.78	0.62	...	-	0.26
masyarakat aju terima blt bbm	0.78	0.62	...	-	0.26
brand kenal keluarga cendana	-1.08	0.68	...	0.03	-
peristiwa bom ledak bandara soetta orang luka	-0.75	0.75	...	-	-

#### C. Model Performance Evaluation

The research evaluates the performance of multi-task models using different architectures and varying epochs. Fig.4 shows the AUC-ROC progression for Dense MTL and CNN MTL models across training epochs. The Dense MTL model consistently outperforms the CNN MTL model, maintaining a higher AUC-ROC score at epochs 15, 20, and 25. At epoch 15, the Dense MTL AUC-ROC is 0.993 compared to 0.99 for CNN

MTL. By epoch 20, the Dense MTL score increases to 0.994, while the CNN MTL score drops to 0.988. At the final epoch 25, the Dense MTL achieves a score of 0.994, while the CNN MTL recovers to 0.99 but remains below the Dense MTL's performance. The Dense MTL model exhibits superior classification performance, as indicated by the AUC-ROC metric.

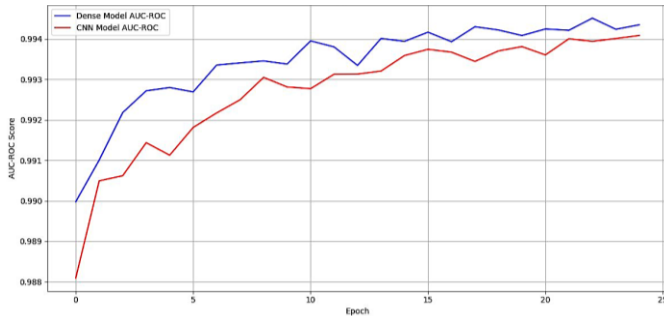


Fig.4 AUC-ROC Graph

The results in Table V indicate that the Dense architecture consistently outperforms the CNN architecture in terms of AUC-ROC and is more accurate for the classification task. Across all epochs, the Dense model achieves higher AUC-ROC values (0.995) and accuracy (94%) compared to CNN, which shows a slightly lower AUC-ROC (0.994) and accuracy (91%).

TABLE V  
 PERFORMANCE EVALUATION OF CLASSIFICATION

Epochs	Multi-task Architecture	Metrics Evaluation	
		AUC-ROC	Accuracy
15	Dense	0.993	93%
	CNN	0.99	91%
20	Dense	0.994	93%
	CNN	0.993	90%
25	Dense	0.995	94%
	CNN	0.994	91%

The results shown in Table VI indicate that the Dense architecture consistently outperforms the CNN model in terms of RMSE, MAE, and MSE across all epochs. While the CNN model shows lower RMSE and MAE in the first two epochs, the Dense model performs better overall in the third epoch despite a slight increase in error metrics. The Dense model achieves an RMSE of 0.040 and an MAE of 0.031 at epoch 25, which remains acceptable considering the increase in complexity. This suggests that the Dense architecture provides more reliable and stable performance for the task, with its slight increase in error not overshadowing its consistent ability to handle classification and prediction tasks.

TABLE VI  
 PERFORMANCE EVALUATION OF PREDICTION

Epochs	Multi-task Architecture	Metrics Evaluation		
		RMSE	MAE	MSE
15	Dense	0.025	0.017	0.0006

Epochs	Multi-task Architecture	Metrics Evaluation		
		RMSE	MAE	MSE
20	CNN	0.018	0.008	0.0003
	Dense	0.019	0.01	0.0004
	CNN	0.017	0.008	0.0003
25	Dense	0.04	0.031	0.0016
	CNN	0.02	0.013	0.0004

For classification, the Dense architecture outperforms the CNN model, achieving higher accuracy (94% vs. 91%) and a higher AUC-ROC (0.9943 vs. 0.9898), indicating better performance in categorizing news articles correctly. The Dense model demonstrates a superior ability to distinguish between different news categories.

When predicting the number of readers, the Dense model also performs better, with a lower RMSE of 0.0254 compared to the CNN's RMSE of 0.0178 at epoch 15, and it continues to maintain this trend through epoch 25. Although the CNN model performs better at earlier epochs in terms of error metrics, the Dense model ultimately provides a more robust and consistent performance across classification and regression tasks, particularly in terms of AUC-ROC and accuracy.

Table VII The model's overall accuracy across all categories demonstrates its effectiveness in classifying content, with all categories achieving a minimum accuracy of 89.92%. The model is particularly adept at categorizing sports and automotive content; however, further refinement is needed for categories such as news and finance, where accuracy is slightly lower.

TABLE VII  
 ACCURACY OF EACH CATEGORY

Categories	Accuracy
Celebrity	91.67
Finance	90.90
News	89.92
Automotive	93.01
Sports	98.04

The highest accuracy was achieved for the sports category, with an accuracy of 98.04%, indicating that the model excels in classifying sports-related content. This can be attributed to the distinct and uniform nature of sports articles, which enables the model to quickly identify key features that differentiate them from other categories. The results of the classification of news categories and prediction of users for each article are shown in Table VIII.

TABLE VIII  
 CLASSIFICATION AND PREDICTION RESULTS

Article_Title	Comparison Results			
	Actual Users	Actual Kanal	Predicted Users	Predicted Kanal
<i>pecat sekdes cantik dugem tenggak miras pamer joget lincah</i>	1,113	celebrity	926	celebrity

Article_Title	Comparison Results			
	Actual Users	Actual Kanal	Predicted Users	Predicted Kanal
<i>mantan main bintang timnas spanyol bawa luis milla persib bandung nomor eks penggawa manchester united</i>	12,240	news	12,051	news
<i>blt minyak goreng bonus cair kategori langsung antar rumah</i>	31,156	sports	31,130	sports
<i>kenal adrian gan desainer indonesia modernisasi baju cheongsam</i>	1,113	celebrity	866	celebrity
<i>hamil cek tekan darah</i>	1,113	news	883	news
<i>megawati tibatiba semprot mendagri tito karnavian</i>	1,113	celebrity	1,055	celebrity

#### IV. CONCLUSION

In conclusion, this research demonstrates the effectiveness of combining Multi-Task Learning (MTL) with CNN and Dense architectures for news classification and reader prediction tasks. The study highlights the advantages of Dense architecture by comparing it with these two architectures. The Dense model, particularly when trained for 25 epochs, outperformed the CNN model, achieving a remarkable AUC-ROC score of 99% and an accuracy of 94%. This superior performance showcases the potential of Dense models in efficiently handling the dual tasks of classification and prediction.

The results underline the value of multi-task learning in improving both the accuracy of news categorization and the predictability of reader engagement. Overall, this research provides valuable insights into the application of MTL in online news management, suggesting that the Dense architecture is a highly effective choice for such tasks. Future work could focus on further optimizing and extending this model to enhance its applicability to other domains, thereby ensuring a broader impact on content curation and resource allocation in digital media.

Although the model performs well across most categories, there is room for improvement, especially in the News and Finance categories. Investigating the misclassified articles in these categories may provide further insights into the model's limitations. Enhancing the model's ability to distinguish between the broader and more diverse content in these categories could improve overall classification performance.

#### REFERENCES

- [1] M. Javaid, A. Haleem, R. P. Singh, S. Rab, and R. Suman, "Exploring impact and features of machine vision for progressive industry 4.0 culture," 2022. doi: 10.1016/j.sintl.2021.100132.
- [2] Y. Li *et al.*, "Artificial intelligence-powered pharmacovigilance: A review of machine and deep learning in clinical text-based adverse drug event detection for benchmark datasets," *J Biomed Inform*, vol. 152, p. 104621, Apr. 2024, doi: 10.1016/J.JBI.2024.104621.
- [3] E. HASANÇEBİ and A. YILMAZ, "Changing Gatekeepers in the New Media Age: An Analysis of Internet Newspapers and Twitter," *Erciyes İletişim Dergisi*, vol. 11, no. 1, pp. 1–14, Jan. 2024, doi: 10.17680/erciyesiletisim.1352735.
- [4] M. I. Syafoah and L. Lestandy, "Emotional Text Classification Using TF-IDF (Term Frequency-Inverse Document Frequency) And LSTM (Long Short-Term Memory)," 2022.
- [5] H. Huang, J. Mao, R. Liu, W. Lu, T. Tang, and L. Liu, "MTLMetro: A Deep Multi-Task Learning Model for Metro Passenger Demands Prediction," *IEEE Transactions on Intelligent Transportation Systems*, vol. 25, no. 9, pp. 11805–11820, 2024, doi: 10.1109/TITS.2024.3373565.
- [6] B. L. Ye, S. Zhu, L. Li, and W. Wu, "Short-term traffic flow prediction at isolated intersections based on parallel multi-task learning," *Systems Science and Control Engineering*, vol. 12, no. 1, 2024, doi: 10.1080/21642583.2024.2316160.
- [7] Diyah Utami Kusumaning Putri and Dinar Nugroho Pratomo, "Clickbait Detection of Indonesian News Headlines using Fine-Tune Bidirectional Encoder Representations from Transformers (BERT)," *Inform : Jurnal Ilmiah Bidang Teknologi Informasi dan Komunikasi*, vol. 7, no. 2, pp. 162–168, Jul. 2022, doi: 10.25139/inform.v7i2.4686.
- [8] P. Liu, X. Qiu, and X. Huang, "Adversarial multi-task learning for text classification," in *ACL 2017 - 55th Annual Meeting of the Association for Computational Linguistics, Proceedings of the Conference (Long Papers)*, Association for Computational Linguistics (ACL), 2017, pp. 1–10. doi: 10.18653/v1/P17-1001.
- [9] G. Xingyi and H. M. Adnan, "Potential cyberbullying detection in social media platforms based on a multi-task learning framework," *International Journal of Data and Network Science*, vol. 8, no. 1, pp. 25–34, Dec. 2024, doi: 10.5267/j.ijdns.2023.10.021.
- [10] M. S. Jahan and M. Oussalah, "A systematic review of hate speech automatic detection using natural language processing," *Neurocomputing*, vol. 546, p. 126232, Aug. 2023, doi: 10.1016/J.NEUCOM.2023.126232.
- [11] K. Denistia and R. H. Baayen, "The morphology of Indonesian: Data and quantitative modeling."
- [12] Internal Company Dataset, "Dataset of Indonesian News Database," 2022.
- [13] E. Yoshua and W. Maharani, "Depression Detection of Users in Social-Media Twitter Using Decision Tree with Word2Vec," *Inform : Jurnal Ilmiah Bidang Teknologi Informasi dan Komunikasi*, vol. 9, no. 1, pp. 95–100, Feb. 2024, doi: 10.25139/inform.v9i1.7617.
- [14] R. Wang and K. Sun, "TIMIT Speaker Profiling: A Comparison of Multi-task learning and Single-task learning Approaches," Apr. 2024.
- [15] S. Vandenhende, S. Georgoulis, B. De Brabandere, and L. Van Gool, "Branched Multi-Task Networks: Deciding What Layers To Share," Apr. 2019.
- [16] B. Qu *et al.*, "Multi-task CNN Behavioral Embedding Model For Transaction Fraud Detection," Nov. 2024.
- [17] D. Yang *et al.*, "An efficient multi-task learning CNN for driver attention monitoring," *Journal of Systems Architecture*, vol. 148, p. 103085, Mar. 2024, doi: 10.1016/J.SYSARC.2024.103085.
- [18] J. S. Aguilar-Ruiz and M. Michalak, "Classification performance assessment for imbalanced multiclass data," *Sci Rep*, vol. 14, no. 1, Dec. 2024, doi: 10.1038/s41598-024-61365-z.
- [19] A. M. Carrington *et al.*, "Deep ROC Analysis and AUC as Balanced Average Accuracy, for Improved Classifier Selection, Audit and Explanation," *IEEE Trans Pattern Anal Mach Intell*, vol. 45, no. 1, pp. 329–341, Jan. 2023, doi: 10.1109/TPAMI.2022.3145392.
- [20] Z. Chen, V. Badrinarayanan, C.-Y. Lee, and A. Rabinovich, "GradNorm: Gradient Normalization for Adaptive Loss Balancing in Deep Multi-task Networks," Nov. 2017.

This is an open-access article under the [CC-BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.

