MATHEMATICAL MODELING AND COMPUTING, Vol. 12, No. 3, pp. 767–778 (2025)

Mathematical modeling of multi-label classification of job descriptions using transformer-based neural networks

Podolchak N., Tsygylyk N., Petlovanyi M.

Lviv Polytechnic National University,
12 S. Bandera Str., 79013, Lviv, Ukraine

(Received 12 June 2025; Revised 31 July 2025; Accepted 3 August 2025)

This article presents the mathematical modeling of the multi-label classification task of job description texts aimed at the automatic detection of working conditions and social benefits, which can enhance communication efficiency between employers and job seekers. The proposed approach is based on the use of the transformer-based BERT neural network, pre-trained on a multilingual corpus. The dataset was constructed by collecting job postings from the three largest Ukrainian job search platforms: Work.ua, Robota.ua, and Jooble.org. The collected texts were augmented with artificially generated examples using large language models to ensure class balance. An architecture was implemented for fine-tuning the BERT model in a multi-label classification mode using the Binary Cross-Entropy loss function. To determine the optimal training configuration, a comparative analysis of four popular optimizers (SGD, AdaGrad, RMSprop, AdamW) was conducted under various learning rate values. The model's performance was evaluated using precision, recall, and F1-score metrics. The experimental results demonstrated that the highest classification quality was achieved using the AdamW optimizer with an appropriately selected learning rate. The novelty of the study lies in combining transformer architecture with an applied task in the field of job description text processing, which enables increased informativeness of postings and automation of preliminary analysis of working conditions. The proposed approach can serve as a foundation for developing tools in HR systems and can be integrated into recruitment platforms to improve the relevance of job postings to the needs of target audiences.

Keywords: *mathematical modeling; neural networks; BERT; multi-label classification; optimization algorithms; job description; transformer architecture; natural language processing (NLP).*

2010 MSC: 68T50, 68T05, 90C59 **DOI:** 10.23939/mmc2025.03.767

1. Introduction

In the modern world, employment plays a key role in ensuring social well-being and economic growth. Despite the large number of job openings, the job search process often proves to be inefficient, especially for certain population groups such as students, women, veterans, and individuals over the age of 50. One of the main challenges that hinders decision-making by job seekers is the lack of essential information in job postings, such as working conditions, benefits, and advantages. This creates a rationale for the automated analysis of job descriptions to identify or supplement missing information.

Recent advances in natural language processing (NLP) and transformer-based neural networks - particularly BERT-based models-open new opportunities for addressing such challenges. Models of this class are capable of effectively processing text by considering the context both to the left and right of each word, which enables them to perform complex classification tasks. Unlike large language models, BERT-type transformers are less resource-intensive and better suited for solving specialized applied problems, especially under conditions of limited data availability.

The aim of this study is to develop and analyze a mathematical model for multi-label classification of job descriptions using a transformer-based neural network, taking into account various model training optimization methods.

To achieve this goal, the following tasks must be accomplished:

— collect and preliminarily analyze textual data from open online sources containing job descriptions;
— design and implement a software module for automated extraction of relevant information from web resources;
— develop a text data augmentation algorithm to balance the dataset using synthetic examples;
— adapt the BERT transformer model to the task of multi-label classification using different optimization techniques;
— evaluate the effectiveness of the chosen optimizers using key classification quality metrics — precision, recall, and F1-score;
— summarize the results, considering practical implications for improving job description quality.

The scientific novelty of this study lies in the development of a mathematical model for multi-label classification of textual job descriptions based on a transformer neural network, adapted to the specifics of short, application-oriented texts. The proposed approach not only enhances the informativeness of job postings by automatically identifying key working conditions but also provides a foundation for building analytical modules within modern HR systems.

2. Literature review

Over the past decade, there has been a rapid advancement in natural language processing (NLP) methods, which has significantly contributed to the widespread adoption of machine learning for text analysis tasks. Particular attention is now given to approaches that enable the automatic extraction of relevant features from text, enhancing its informativeness, structure, and appeal to target users. At the same time, traditional statistical or manual feature engineering techniques are becoming increasingly obsolete, especially in applied tasks involving unstructured or semi-structured texts — such as job postings or résumés.

One of the key technological breakthroughs in the field of NLP has been the introduction of transformer architecture, which marked the beginning of a new era of highly accurate language models. The most prominent among these is BERT (Bidirectional Encoder Representations from Transformers), proposed by Devlin et al. [1], which processes context on both the left and right sides of each word in a sentence. This allows for a deeper understanding of text structure and semantics.

Several improved architectures have been developed based on BERT, including RoBERTa, ALBERT, and DistilBERT [2], which demonstrate high accuracy across a wide range of NLP tasks. Leon et al. [3] proposed a sentence-embedding-based approach for hierarchical skill classification, which is widely applied in the context of job descriptions. Al-Smadi [4] implemented a hybrid DeBERTa-BiLSTM model for multi-label classification of Arabic medical queries. Chen et al. [5] applied BERT for classifying COVID-19-related literature, showcasing the architecture's versatility across domains.

In the HR field, the use of BERT has become increasingly prevalent. Tran et al. [6] developed a job title prediction system based on job text using a combination of BERT, Bi-GRU, and CNN. Bhola et al. [7] implemented an extreme multi-label skill classification framework and introduced the Correlation Aware Bootstrapping training strategy. Qin et al. [8] designed an extended BERT-based architecture for résumé classification according to skills, experience, and job type. Cheng et al. [9] explored strategies to improve the performance of pre-trained transformers for multi-label classification involving a large number of classes — an especially relevant task in the HR domain.

Other studies highlight the application of transformer-based models for classifying text within professional taxonomies. In the study by Zhang et al. [10], the Job2Vec method was proposed to benchmark job titles by predicting links within a job title graph using collective representation learning based on multiple aspects, including graph topology, description semantics, career transition balance, and transition duration. While effective, this approach fails to fully address multi-label classification scenarios. Studies [11,12] also examine multi-label classification approaches in the domains of educational testing and legal analysis, further illustrating the broad applicability of transformer technologies.

In parallel with the development of BERT, there has been a rapid expansion of large language models (LLMs) such as GPT. Despite their impressive performance in text generation tasks [13], the use of LLMs for precise classification tasks is often excessively resource-intensive. Tarekegn et al. [14] emphasize that LLMs are more suitable for tasks related to text generation or corpus augmentation, particularly in cases involving label imbalance. Studies [15, 16] explore the use of LLMs as tools for generating synthetic texts in HR contexts, which can also be employed to balance class distributions in training datasets.

Several studies have explored transformer models in academic and scientific classification contexts. Morales-Hernández et al. [17] evaluated transformers for classification based on the Sustainable Development Goals (SDGs), confirming the effectiveness of BERT and its derivatives. Hinojosa Lee et al. [18] highlighted the importance of selecting appropriate evaluation metrics for multi-label classification, demonstrating the sensitivity of the F1-score to class imbalance. This finding is supported by study [19], which provides examples of using micro- and macro-metrics in the classification of climate-related documents.

An essential technical component in model fine-tuning is the choice of optimizer. Loshchilov & Hutter [20] introduced the AdamW algorithm, which differs from the classical Adam by incorporating weight decay directly into the weight update process. Subsequent studies, such as [21], have demonstrated the superior performance of AdamW compared to standard optimizers like SGD, RMSprop, and AdaGrad. Liu et al. [22] proposed the Query2Label method, which improves performance in multi-label learning by enabling more compact query representations. Furthermore, studies [23, 24] examine the impact of various learning rates and optimization strategies on model stability and convergence, underscoring the importance of hyperparameter tuning in training transformer-based models.

It is worth noting that, despite the widespread application of transformer models in classification tasks, the use of multi-label classification for job descriptions-aimed at improving their social relevance and appeal to candidates-remains underexplored. This highlights both the scientific novelty and the practical significance of the present study.

3. Research methodology

The study is based on a step-by-step methodology for developing a multi-label classification model for job descriptions using the BERT (Bidirectional Encoder Representations from Transformers) transformer neural network — one of the most widely used models in natural language processing (NLP). BERT employs the transformer mechanism, which enables the model to consider the context of each word in a sentence from both the left and the right. This feature is crucial for deeper content analysis and the identification of semantic dependencies (Figure 1).

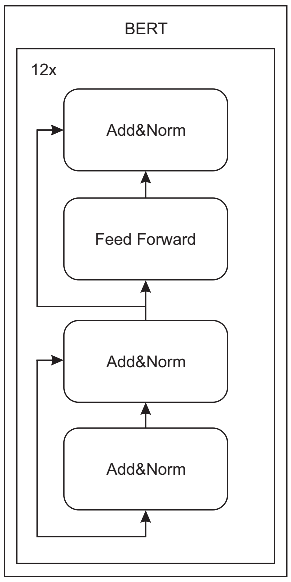

Fig. 1. Architecture of the BERT (Bidirectional Encoder Representations from Transformers) model.

In machine learning classification tasks, three main types are distinguished:

— binary classification — each object (text) can be assigned to one of two classes (e.g., "has a benefit" / "does not have a benefit");
— multi-class classification — each object belongs to only one of several possible classes (e.g., a single job type from a fixed list);
— multi-label classification — each object can simultaneously belong to multiple independent classes, which is relevant when a job posting may contain references to several working conditions (e.g., meals, flexible schedule, health insurance, etc.) (Figure 2).

This study substantiates the use of multi-label classification, as job descriptions often include several relevant characteristics at once, each of which is important for effective analysis and the development of accurate recommendations.

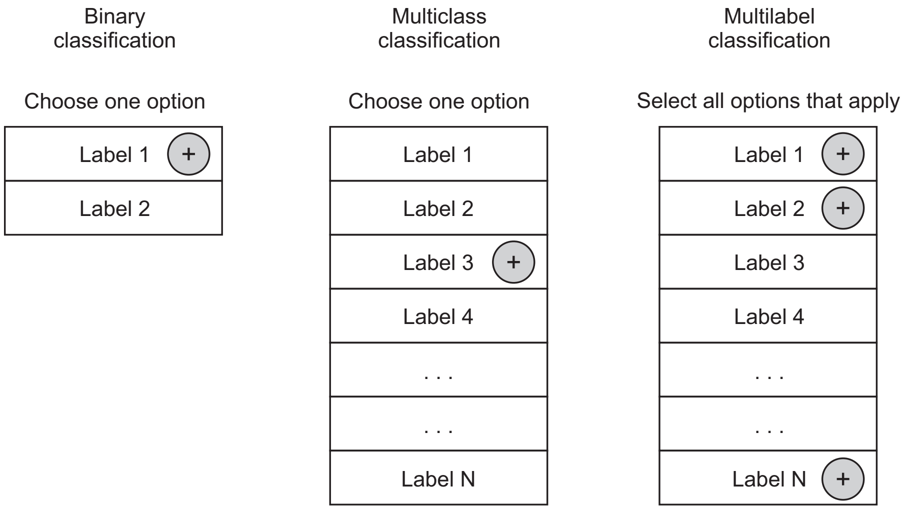

Fig. 2. The difference between binary, multi-class, and multi-label classifications.

The model trained in this research consisted of the following core components:

— a pre-trained BERT model – the fundamental transformer block that employs a multi-layer attention mechanism to process input textual data. This block enables the model to learn contextual word representations;
— a classification head – after the input text passes through BERT, its vector representations are passed to an additional fully connected layer designed to predict the relevant labels.

The methodology consists of three key components:

— analysis of job seekers' expectations in the labor market through a survey;
— creation of a data corpus for modeling;
— training and evaluation of the classification model's performance.

To determine job seekers' expectations regarding desirable working conditions, an online survey was conducted using Google Forms. The survey included 454 respondents. Although data on place of residence was not collected, it is highly likely that most respondents were residents of Lviv or the Lviv region.

The survey focused on identifying the factors that respondents consider important in job postings, including:

— flexible working hours;
— opportunities for career growth;
— official employment;
— health insurance;
— meal compensation;
— corporate culture;
— a motivational atmosphere.

Based on the survey results, a set of target characteristics was identified to serve as labels in the multi-label classification task.

The data source for the training corpus included three of the largest Ukrainian job search platforms:

— Work.ua;
— Robota.ua;
— Jooble.org.

Job postings were selected from the following categories: electrician, driver, construction worker, machine operator, packer, bricklayer, agronomist, mechanic, and logistician. Primary data collection was carried out using the Selenium library for Python, which enables automated access to web pages and the extraction of structured information from HTML.

Particular attention was given to detecting mentions of the following working conditions and benefits in job descriptions:

— maternity leave;
— transportation reimbursement;
— meals;
— access to a gym or compensation for a gym membership;
— flexible schedule;
— inclusiveness.

Given that the distribution of benefits and bonuses was uneven (some labels appeared infrequently), large language models (LLMs) were used to balance the dataset. The generation of additional texts made it possible to increase the number of examples for underrepresented classes.

As a result, after data augmentation, the corpus consisted of more than 800 job descriptions. All texts were in Ukrainian. For further training, the dataset was split into an 80:20 ratio into training and test sets.

To implement the classification task, a pre-trained BERT model (bert-base-multilingual-cased) was used. The model architecture included:

— a base transformer component that generates contextual word representations;
— a fully connected classification layer with a sigmoid activation function for multi-label prediction;
— a Binary Cross-Entropy (BCE) loss function, which is suitable for independently predicting the presence of each label in the text. The loss function is defined as:

$$L_{\mathrm{BCE}} = -\frac{1}{n}\sum_{i=1}^{n}\left(Y_i \cdot \log \widehat{Y}_i + (1-Y_i)\cdot \log(1-\widehat{Y}_i)\right),$$

where L_{BCE} is the Binary Cross-Entropy loss function, Y_i is the true (reference) label for class (Figure 3), and $\widehat{Y}_i$ is the probability predicted by the classifier for the label.

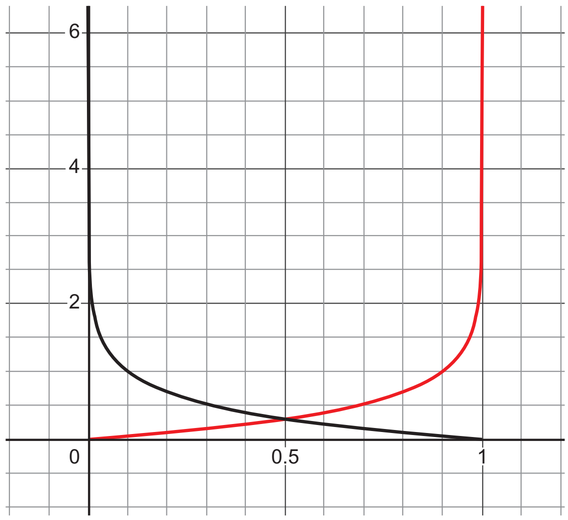

Fig. 3. Dependence of the Binary Cross-Entropy (BCE) loss value on the difference between predicted and true values $\widehat{Y}_i$ and Y_i, where the red curve corresponds to $Y_i = 0$, and black – $Y_i = 1$.

The model was trained using a standard algorithm over a fixed number of epochs, with monitoring of the Binary Cross-Entropy loss value on the training and validation sets (Figure 4).

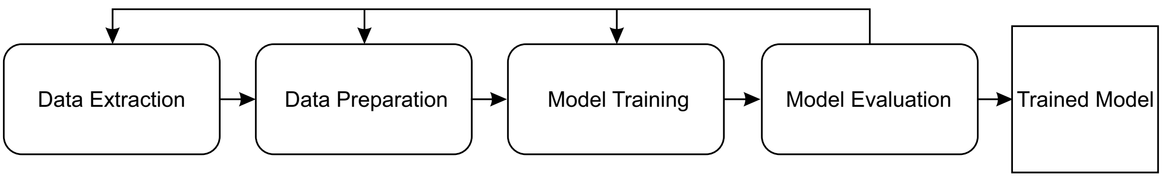

Fig. 4. Standard training algorithm for neural networks.

In addition, the model was trained using various optimizers that perform gradual weight updates to minimize the loss function. The general formula for weight updates is as follows:

$$W_{\mathrm{new}} = W_{\mathrm{old}} - \alpha\frac{dj}{dW},$$

where W is model weights, α is learning rate, $\frac{dj}{dW}$ is the gradient of the loss function.

In this study, four optimizers were tested:

— SGD (Stochastic Gradient Descent) — updates weights by simply shifting in the direction of the gradient;

— AdaGrad — adapts the learning rate for each parameter, decreasing it for frequently changing features;
— RMSprop — takes into account the running average of the squares of previous gradients, reducing oscillations;
— AdamW — combines the benefits of adaptive methods with regularization (weight smoothing through weight decay).

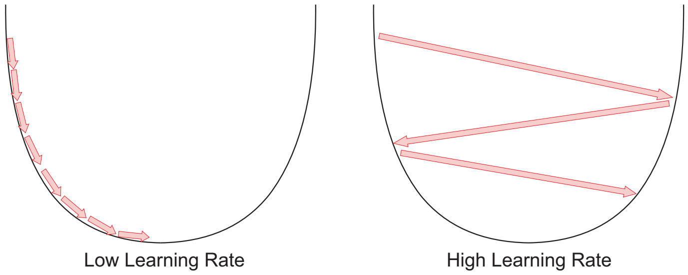

Fig. 5. High and low learning rates.

During the experiment, two learning rate values were tested: $1 \cdot 10^{-4}$ and $2 \cdot 10^{-5}$. The learning rate determines the step size for updating weights after each iteration. Specifically, a high learning rate enables the model to approach the loss function minimum quickly but may cause instability or "overshooting" of the global minimum. A low learning rate ensures stable and precise weight updates but significantly slows down training and may result in convergence to a local minimum (Figure 5).

At the same time, adaptive optimizers such as AdamW or RMSprop automatically adjust the learning rate during training, combining rapid convergence in the early stages with fine-tuning in the later phases.

4. Research results

During an online survey of 454 respondents, data were collected regarding expected working conditions. Based on these results, a set of target labels was created for a multi-label classification task. The most prioritized conditions among the respondents were flexible working hours, opportunities for career advancement, and official employment. The ranking of the 16 most popular conditions is presented in Figure 6.

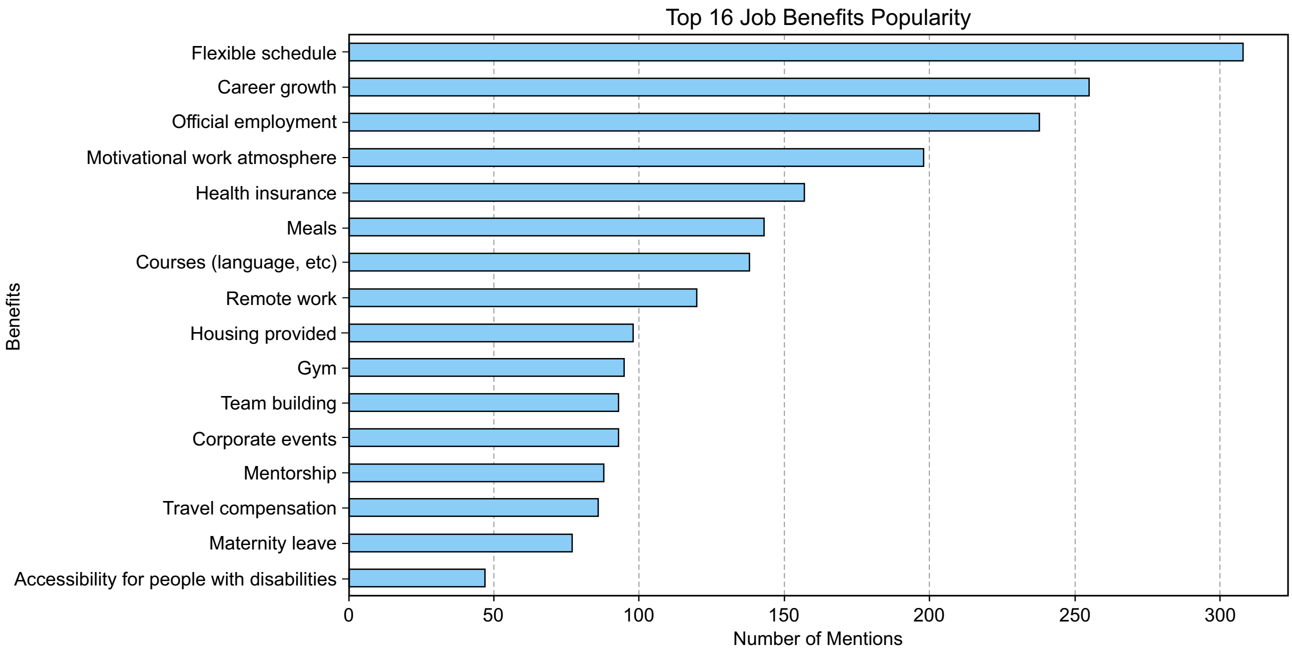

Fig. 6. Popularity of working conditions among respondents (Top 16).

After constructing the training corpus and augmenting the data, the model was trained using the BERT transformer architecture. To evaluate the performance of the developed model, the following metrics were applied:

— Precision — the proportion of correctly predicted positive instances among all instances predicted as positive. A high precision value indicates that the model makes fewer false-positive predictions

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{TN}};$$

— Recall — the proportion of correctly identified positive instances among all actual positive instances. A high recall indicates the model's ability to detect most of the relevant cases

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}};$$

— F1-score — the harmonic mean of precision and recall, which is particularly important in cases of class imbalance. The F1-score is especially useful when dealing with imbalanced classes, as it provides an accurate measure of the model's performance

$$\text{F1} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}.$$

These metrics were calculated on the test set after the completion of training.

To improve classification quality, a series of experiments was conducted using four different optimizers: SGD, AdaGrad, RMSprop, and AdamW. Most of the optimizers in this study did not demonstrate satisfactory training results. The best performance was achieved by AdamW and RMSprop with a learning rate of $lr = 2 \cdot 10^{-5}$, as well as AdaGrad with $lr = 1 \cdot 10^{-4}$. Therefore, these optimizers were selected for further testing. The total number of epochs for each configuration was 20 (Figure 7). The results are summarized in Table 1.

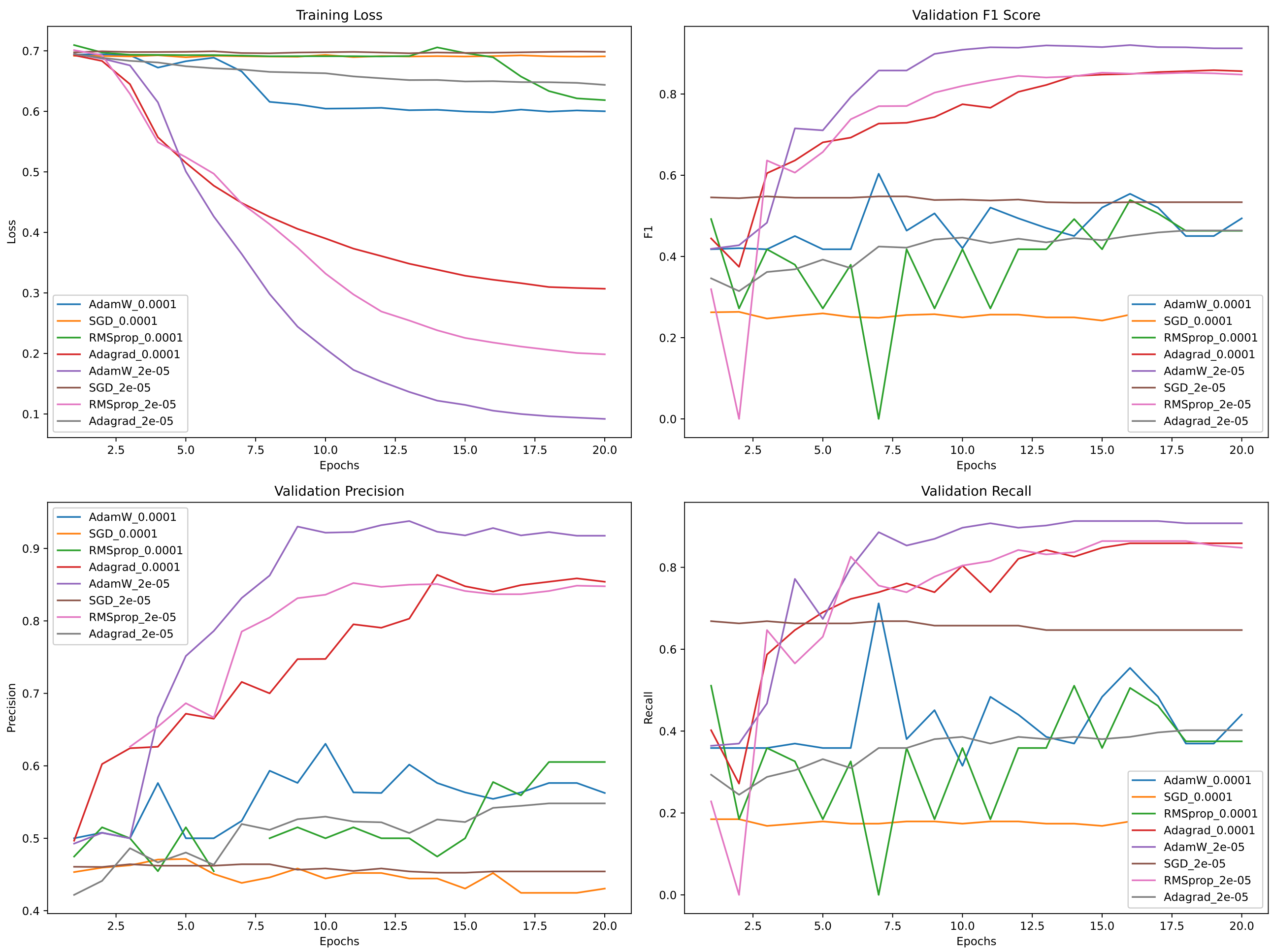

Fig. 7. Training results of the neural network over 20 epochs using different optimizers and learning rates.

Table 1. Comparison of optimizers based on classification performance.

Optimizer	Learning rate	Precision	Recall	F1-score
AdamW	$1 \cdot 10^{-4}$	0.4621	0.3333	0.3873
	$2 \cdot 10^{-5}$	0.9432	0.9071	0.9248
SGD	$1 \cdot 10^{-4}$	0.3224	0.4958	0.3907
	$2 \cdot 10^{-5}$	0.4765	0.4426	0.4589
RMSprop	$1 \cdot 10^{-4}$	0.4621	0.3333	0.3873
	$2 \cdot 10^{-5}$	0.7784	0.7869	0.7826
AdaGrad	$1 \cdot 10^{-4}$	0.8324	0.8415	0.8370
	$2 \cdot 10^{-5}$	0.5278	0.3115	0.3918

To monitor training effectiveness, the Binary Cross-Entropy (BCE) loss function was used. The first graph in Figure 7 shows the loss dynamics during model training with the AdamW optimizer at a learning rate of $2 \cdot 10^{-5}$, which yielded the best classification performance. The loss function gradually decreased over the course of training, indicating stable model convergence.

For a deeper analysis of classification errors, a confusion matrix was constructed for each label. This allows not only for the evaluation of overall classification accuracy but also for an analysis of the distribution of correct and incorrect predictions for each feature.

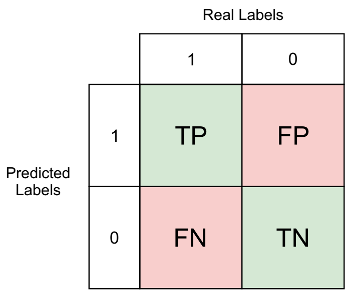

Fig. 8. Confusion matrix.

Each matrix includes the following indicators:

— TP (True Positive) — the number of cases where the model correctly predicted the presence of a label;
— TN (True Negative) — the number of cases where the model correctly predicted the absence of a label;
— FP (False Positive) — cases where the model incorrectly predicted the presence of a label;
— FN (False Negative) — cases where the model failed to detect a label that was actually present in the text.

The confusion matrix is shown in Figure 8. For the best-performing model configuration (AdamW, learning rate $2 \cdot 10^{-5}$) confusion matrices for each label are presented in Figure 9.

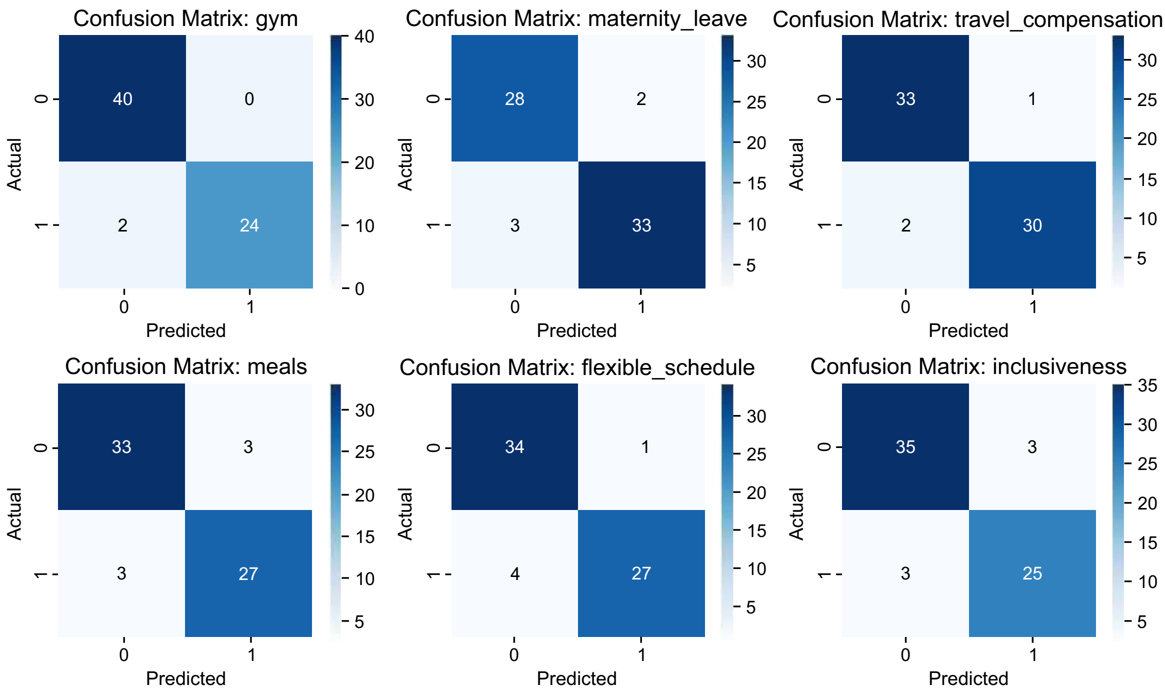

Fig. 9. Confusion matrices for individual labels, AdamW, learning rate $2 \cdot 10^{-5}$.

5. Discussion of results

The modeling results demonstrated that using the BERT transformer model for multi-label classification of Ukrainian-language job vacancy texts is both appropriate and effective. All experiments focused on building a model capable of simultaneously predicting the presence of multiple working conditions within a single vacancy text — a setup that aligns with real market conditions, as each job listing may offer various benefits to potential candidates.

The classification outcomes revealed significant differences in performance across different combinations of optimizers and learning rates. The highest F1-score (0.9248) was achieved by the model trained with the AdamW optimizer and a learning rate of $2 \cdot 10^{-5}$, indicating a very high level of agreement between predicted and actual labels. This is particularly critical for multi-label classification tasks, where even partial errors can significantly impact how a vacancy is perceived by candidates. This is especially relevant for features such as flexible working hours, health insurance, or transportation reimbursement — often decisive factors in a candidate's decision-making process.

Conversely, other combinations — particularly SGD with a high learning rate — proved ineffective. This confirms the widely accepted understanding that non-adaptive optimizers in transformer models require careful tuning of hyperparameters, and even slight deviations from optimal values can lead to loss of accuracy and instability in the loss function.

In real-world deployment, these differences are crucial: misclassifying job content can reduce audience engagement and result in ineffective recruitment campaigns. The loss curve during training showed a stable decline in Binary Cross-Entropy (BCE) over all 20 epochs, with no signs of overfitting. This indicates a well-chosen model configuration, including architecture depth, batch size, learning rate, and number of training epochs.

The stability observed during training is particularly valuable given the limited size of the training set (fewer than 1 000 examples), which is a common feature of real-world applied NLP tasks in the Ukrainian-language context. Analysis of the confusion matrix constructed for the best-performing model configuration revealed varying levels of classification accuracy across labels. Specifically, labels such as "flexible schedule," "meals," and "career advancement" were predicted with high precision, indicating a clear semantic representation of these concepts in job postings. In contrast, features like "inclusivity" or "maternity leave," which appeared infrequently in the corpus, often resulted in false positives or false negatives. This may be attributed not only to frequency issues but also to the contextual variability in how such conditions are phrased across different postings. This observation also highlights the importance of further model refinement through semantic normalization of input text (e.g., unifying synonymous formulations) or expanding the dataset with specifically crafted examples.

The AdamW optimizer, which combines adaptive gradient scaling and weight decay regularization, demonstrated the highest performance among all optimizers tested. This result aligns with current best practices in training transformer-based models, where AdamW has become the de facto standard for fine-tuning large models [20]. AdaGrad also showed relatively high F1 performance, although its performance significantly declined at lower learning rates — consistent with its known characteristic of accumulating gradients and shrinking the learning rate to extremely low values over time.

The practical significance of this study lies in the potential integration of the developed model into recruitment systems, particularly for:

- automatic preliminary evaluation of the quality of job descriptions;
- detection of "weak spots" in job ads that may reduce candidate engagement;
- tailoring job text to specific target audiences — such as young professionals, persons with disabilities, or parents of young children;
- generating analytics by label categories to assess labor market trends.

Key strengths of the study include:

- a focus on Ukrainian-language content, ensuring the model's relevance to the national labor market and addressing the scarcity of Ukrainian-language NLP tools;

— the integration of empirical sociological data (survey results) into the model architecture, enhancing the alignment of outcomes with real candidate expectations;
— the combination of deep language models with practical applied tasks, demonstrating the potential of modern AI solutions in both public and corporate human resource management;
— the flexibility of the proposed approach, which can be adapted to other languages, domains (e.g., education, healthcare), document types (e.g., resumes, cover letters), or even internal HR audit needs.

Moreover, the proposed solution not only enhances the efficiency of recruitment processes but also serves as a foundation for the development of transparent, inclusive, and human-centered digital HR platforms. Despite the high performance achieved, several limitations must be acknowledged when interpreting the findings and planning further research stages.

First, some label categories suffered from limited training data. Although the dataset was augmented using large language models, infrequent features such as "inclusivity" or "accessible workplace" remained underrepresented. This may have negatively affected classification quality for those labels, as evidenced by the confusion matrix analysis. Second, the survey data used to define target labels did not include questions about the respondents' geographical location. The absence of regional segmentation prevents an analysis of how working condition preferences may vary across different areas, even though such variation could significantly affect the relevance of certain labels. Third, the model did not account for the publication time of job postings. Given that job ad styles and content can evolve over time — influenced by labor market dynamics, seasonality, or macroeconomic factors — the structure of job descriptions may shift, and this temporal aspect should be incorporated in future versions of the model.

In light of these limitations, future research directions should include:
— expanding and refining the dataset to ensure balanced representation across all target labels, particularly through the collection of low-frequency examples;
— applying domain-specific fine-tuning of BERT on Ukrainian-language job ads and HR-related documents to better adapt the model to stylistic and semantic nuances in the field;
— utilizing attention visualization to interpret which text fragments influence the assignment of each label, thus improving model transparency and business acceptability;
— assessing correlations between specific labels and actual application response rates, enabling predictive analytics of job post effectiveness in practical use.

Taken together, these directions point not only to the refinement of multi-label classification models in the HR domain, but also to the development of comprehensive intelligent decision-support systems for recruitment based on Ukrainian-language content.

6. Conclusion

This study developed and tested an approach for multi-label classification of Ukrainian-language job vacancy texts based on the BERT transformer model. By incorporating real candidate expectations gathered through an online survey, the research established a representative set of target labels, thereby enhancing the practical value of the model. The system effectively detects the presence or absence of key working conditions in job descriptions, enabling the automation of vacancy quality assessment, identification of weak spots in job postings, and adaptation of content to specific target audiences.

The best classification results were achieved using the AdamW optimizer with a learning rate of $2 \cdot 10^{-5}$, reaching an F1-score of 0.9248 and demonstrating stable loss dynamics without signs of overfitting. The comparison of different optimizers highlighted the advantages of adaptive methods over classical approaches, aligning with current trends in training large language models.

Confusion matrix analysis confirmed the model's differentiated classification performance depending on label representation in the training set, emphasizing the importance of data balance. The proposed model has high practical utility for the HR industry, as it can be integrated into recruitment systems to improve job description quality, analyze labor market trends, and design inclusive recruitment strategies.

A key contribution of this study is its focus on Ukrainian-language content, which remains underrepresented in current NLP solutions. The findings lay the groundwork for further research, including domain-specific fine-tuning of BERT on specialized HR corpora, visualizing attention mechanisms to improve model interpretability, and evaluation of the impact of job features on application conversion rates.

[1] Devlin J., Chang M.-W., Lee K., Toutanova K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. Preprint arXiv:1810.04805 (2018).

[2] Liu Y., Ott M., Goyal N., Du J., Joshi M., Chen D., Levy O., Lewis M., Zettlemoyer L., Stoyanov V. RoBERTa: A Robustly Optimized BERT Pretraining Approach. Preprint arXiv:1907.11692 (2019).

[3] Leon F., Gavrilescu M., Floria S.-A., Minea A. A. Hierarchical Classification of Transversal Skills in Job Advertisements Based on Sentence Embeddings. Information. **15** (3), 151 (2024).

[4] Al-Smadi B. S. DeBERTa-BiLSTM: A Multi-Label Classification Model of Arabic Medical Questions Using Pre-Trained Models and Deep Learning. Computers in Biology and Medicine. **170**, 107921 (2024).

[5] Chen Q., Du J., Allot A., Lu Z. LitMC-BERT: Transformer-Based Multi-Label Classification of Biomedical Literature with an Application on COVID-19 Literature Curation. IEEE/ACM Transactions on Computational Biology and Bioinformatics. **19** (5), 2584–2595 (2022).

[6] Tran H. T., Vo H. H. P., Luu S. T. Predicting Job Titles from Job Descriptions with Multi-label Text Classification. Preprint arXiv:2112.11052 (2021).

[7] Bhola A., Halder K., Prasad A., Kan M.-Y. Retrieving Skills from Job Descriptions: A Language Model Based Extreme Multi-label Classification Framework. Proceedings of the 28th International Conference on Computational Linguistics. 5832–5842 (2020).

[8] Cheng X., Lin H., Wu X., Shen D., Yang F., Liu H., Shi N. MLTR: Multi-Label Classification with Transformer. Proceedings of the 2022 IEEE International Conference on Multimedia and Expo (ICME). 1–6 (2022).

[9] Chang W.-C., Yu H.-F., Zhong K., Yang Y., Dhillon I. Taming Pretrained Transformers for Extreme Multi-label Text Classification. KDD '20: Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. 3163–3171 (2020).

[10] Zhang D., Liu J., Zhu H., Liu Y., Wang L., Wang P., Xiong H. Job2Vec: Job Title Benchmarking with Collective Multi-View Representation Learning. Proceedings of the 28th ACM International Conference on Information and Knowledge Management (CIKM '19). 2763–2771 (2019).

[11] Zhang W., Jiang Y., Fang Y., Pan S. Hierarchical Contrastive Learning for Multi-label Text Classification. Scientific Reports. **15**, 14101 (2025).

[12] Wang G., Du Y., Jiang Y., Liu J., Li X., Chen X., Gao H., Xie C., Lee Y. Label-text Bi-attention Capsule Networks Model for Multi-label Text Classification. Neurocomputing. **588**, 127671 (2024).

[13] Brown T. B., Mann B., Ryder N., Subbiah M., Kaplan J., Dhariwal P., et al. Language Models are Few-Shot Learners. Preprint arXiv:2005.14165 (2020).

[14] Tarekegn A. N., Ullah M., Cheikh F. A. Deep Learning for Multi-Label Learning: A Comprehensive Survey. Preprint arXiv:2401.16549 (2024).

[15] Nissa N. K., Yulianti E. Multi-label text classification of Indonesian customer reviews using bidirectional encoder representations from transformers language model. International Journal of Electrical and Computer Engineering. **13** (5), 5641–5652 (2023).

[16] Li B., Li S., Qiu W., Liu T. Bert-based model for Ceramic Emotional Text Recognition and Multi-label Classification. Computer Life. **12** (1), 12–20 (2024).

[17] Morales-Hernández R. C., Jagüey J. G., Becerra-Alonso D. A Comparison of Multi-Label Text Classification Models in Research Articles Labeled with Sustainable Development Goals. IEEE Access. **10**, 123456–123469 (2022).

[18] Lee M. C. H., Braet J., Springael J. Performance Metrics for Multilabel Emotion Classification: Comparing Micro, Macro, and Weighted F1-Scores. Applied Sciences. **14** (21), 9863 (2024).

[19] Liu J., Chang W.-C., Wu Y., Yang Y. Deep Learning for Extreme Multi-label Text Classification. SIGIR '17: Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval. 115–124 (2017).

[20] Loshchilov I., Hutter F. Decoupled Weight Decay Regularization. Preprint arXiv:1711.05101 (2017).

[21] Touvron H., Cord M., Douze M., Massa F., Sablayrolles A., Hervé J. Training Data-Efficient Image Transformers & Distillation Through Attention. Preprint arXiv:2012.12877 (2021).

[22] Liu S., Zhang L., Yang X., Su H., Zhu J. Query2Label: A Simple Transformer Way to Multi-Label Classification. Preprint arXiv:2107.10834 (2021).

[23] Tian X., Qin Y., Huang R., Chen Y. A Label Information Aware Model for Multi-label Text Classification. Neural Processing Letters. **56**, 242 (2024).

[24] Google Developers. Classification: Accuracy, Recall, Precision, and Related Metrics. Google Machine Learning Crash Course (2024).
https://developers.google.com/machine-learning/crash-course/classification/precision-and-recall.

Математичне моделювання багатоміткової класифікації описів вакансій із використанням нейронних мереж на основі трансформерів

Подольчак Н., Цигилик Н., Петльований М.

Національний університет "Львівська політехніка",
вул. С. Бандери, 12, 79013, м. Львів, Україна

У статті представлено математичне моделювання задачі багатоміткової класифікації текстів описів вакансій з метою автоматичного виявлення умов праці та соціальних пільг, що здатне підвищити ефективність комунікації між роботодавцями та тими, хто шукає роботу. Запропонований підхід ґрунтується на використанні нейронної мережі BERT, навченої на багатомовній основі. Для формування набору даних було зібрано вакансії з трьох найбільших українських платформ пошуку роботи: Work.ua, Robota.ua та Jooble.org. Зібрані тексти були доповнені штучно згенерованими прикладами за допомогою великих мовних моделей для забезпечення збалансованості класів. Реалізовано архітектуру для донавчання моделі BERT у режимі багатоміткової класифікації з використанням функції втрат Binary Cross-Entropy. Для визначення оптимальної конфігурації навчання проведено порівняльний аналіз чотирьох популярних оптимізаторів (SGD, AdaGrad, RMSprop, AdamW) при різних значеннях швидкості навчання. Ефективність моделі оцінювалася за метриками точності (precision), повноти (recall) та F1-міри. Результати експериментів показали, що найвища якість класифікації досягається при використанні оптимізатора AdamW з правильно підібраною швидкістю навчання. Наукова новизна дослідження полягає в поєднанні трансформерної архітектури з прикладною задачею обробки текстів описів вакансій, що дозволяє підвищити інформативність оголошень і автоматизувати попередній аналіз умов праці. Запропонований підхід може стати основою для створення інструментів у HR-системах і бути інтегрованим у платформи з працевлаштування з метою підвищення релевантності вакансій потребам цільової аудиторії.

Ключові слова: *математичне моделювання; нейронні мережі; BERT; багатоміткова класифікація; алгоритми оптимізації; опис вакансій; архітектура трансформерів; обробка природної мови (NLP).*