

Научная статья

Статья в открытом доступе

УДК 004.89 : 331.108.2

doi: 10.30987/2658-6436-2025-4-52-57

АНСАМБЛЬ МОДЕЛЕЙ ДЛЯ АВТОМАТИЧЕСКОЙ КЛАССИФИКАЦИИ ТЕКСТОВ РЕЗЮМЕ

Любовь Александровна Комарова

Финансовый Университет при Правительстве РФ, г. Москва, Россия

lakomarova@fa.ru, <https://orcid.org/0000-0001-5277-8234>

Аннотация. Цель исследования: повысить точность автоматической классификации текстов резюме при ограниченных вычислительных ресурсах. Задача: создать и экспериментально верифицировать ансамбль моделей, дообученных на открытом датасете ResumeAtlas и дополненных таксономической регуляризацией с калибровкой вероятностей, для повышения точности и устойчивости классификации. Методы исследования: предложенный пайплайн включает: (I) модуль TfidfSentenceSelector для отбора семи наиболее информативных предложений; (II) доменно-адаптивное предобучение DeBERTa-v3-large (DAPT); (III) регуляризацию R-Drop; (IV) Longformer-large-4096 для обработки полного документа; (V) усреднение логитов и температурную калибровку. Оценка выполнена на открытом корпусе Résumé Atlas с использованием метрик Top-k accuracy и F1. Новизна работы: показано, что TF-IDF-сжатие резюме обеспечивает почти полное сохранение ключевых фактов при пятикратном сокращении входа, а ансамбль «реферат из резюме + длинный контекст» даёт дополнительный прирост точности. Также доказана эффективность одноэпохового DAPT и R-Drop в условиях ограниченного GPU-времени. Результаты исследования: ансамбль достиг Top-1 = 0,93, Top-3 = 0,97, Top-5 = 0,98, Top-10 = 0,99 и F1-macro = 0,93, что превышает базовую модель Résumé Atlas на 2 п.п. и классический TF-IDF + SVM на 11 п.п. по Top-1 точности. Полный цикл обучения выполняется за ≈ 4 ч на одной NVIDIA A100-80 GB. Выводы: комбинация доменно-адаптивного предобучения, сжатия текста и гетерогенно-го внимания предлагает новое открытое решение для классификации резюме и может быть перенесена на другие профессиональные таксономии.

Ключевые слова: классификация текстов резюме, трансформеры, большие языковые модели, ансамбль моделей, таксономическая регуляризация

Для цитирования: Комарова Л.А. Ансамбль моделей для автоматической классификации текстов резюме // Автоматизация и моделирование в проектировании и управлении. 2025. №4 (30). С. 52-57. doi: 10.30987/2658-6436-2025-4-52-57.

Original article

Open Access Article

ENSEMBLE OF MODELS FOR AUTOMATIC CLASSIFICATION OF RESUME TEXTS

Lyubov A. Komarova

Financial University under the Government of the Russian Federation, Moscow, Russia

lakomarova@fa.ru, <https://orcid.org/0000-0001-5277-8234>

Abstract. The aim of this study is to enhance the accuracy of automatic classification of resume texts under limited computational resources. The objective is to design and experimentally validate an ensemble of models, fine-tuned on the open-source ResumeAtlas dataset, augmented with taxonomic regularization and probability calibration, to boost classification accuracy and robustness. The methods of investigations include the proposed pipeline containing (I) a TfidfSentenceSelector module for extracting seven most informative sentences; (II) domain-adaptive pretraining (DAPT) of DeBERTa-v3-large; (III) R-Drop regularization; (IV) Longformer-large-4096 for full document processing; (V) logit averaging and temperature scaling. The author performs evaluation on the Résumé Atlas open corpus using Top-k Accuracy and F1 metrics. The novelty of the work lies in showing that TF-IDF resume compression retains almost all essential facts while reducing input size fivefold, and that the ensemble of «resume summary + long context» yields additional gains in accuracy. Also proven is the efficacy of single-epoch DAPT and R-Drop under restricted GPU time. The study results in achieving Top-1 = 0,93, Top-3 = 0,97, Top-5 = 0,98, Top-10 = 0,99, and F1-macro = 0,93, surpassing the baseline Résumé Atlas model by 2 percentage points and classical TF-IDF+SVM by 11 percentage points in Top-1 accuracy. The entire training cycle completes in approximately 4 hours on a single NVIDIA A100-80 GB GPU. The findings state that the combination of domain-adaptive pretraining, text compression, and heterogeneous attention provides a novel open-source solution for resume classification, potentially transferrable to other professional taxonomies.

Keywords: resume text classification, transformers, large language models, ensemble of models, taxonomic regularization

Введение

Современный рынок труда характеризуется резким ростом числа онлайн-откликов на вакансии и активным использованием цифровых платформ рекрутмента. По оценкам, около 70 % компаний уже применяют онлайн-платформы для найма, и 94 % работодателей планируют внедрять ИИ-технологии в подбор персонала [9]. В этих условиях автоматическая классификация резюме по профессиональным категориям становится критически важной задачей для повышения эффективности работы HR-специалистов и систем отслеживания кандидатов.

Тем не менее, автоматизированная обработка резюме сталкивается с рядом сложностей. Во-первых, доступность и качество обучающих данных ограничены из-за приватности информации соискателей и соблюдения нормативов компании, что затрудняет сбор больших датасетов резюме. Во-вторых, отсутствуют единые стандарты оформления: разные кандидаты по-разному структурируют резюме, и документы существенно различаются по формату, структуре и содержанию. Такая вариативность затрудняет унифицированный анализ и может приводить к смещениям и ошибкам алгоритмов при отнесении резюме к категориям.

За последние годы предложено несколько подходов к автоматической классификации резюме. Ранние работы обычно ограничивались небольшими корпусами (несколько тысяч резюме с 5...25-ю классами) и применяли традиционные алгоритмы машинного обучения – наивный Байес, *SVM*, *Random Forest* и др. – как правило, на *TF-IDF* представлениях текста. Эти методы демонстрировали умеренную точность на многоклассовых задачах и недостаточную обобщающую способность. Например, в исследовании [11] на данных с *Kaggle*, *Glassdoor* и *Indeed* алгоритм *Random Forest* превзошёл *SVM* и байесовский классификатор, достигнув точности около 70 %.

В работе [1] при классификации 962 резюме на 25 категорий наилучший результат показал линейный алгоритм *SVM* (точность ~96 %). Однако столь высокая точность обусловлена ограниченным масштабом и количеством классов. Более современные попытки стали применять методы глубокого обучения. Так, в работе [6] использовали сверточную нейросеть с *TF-IDF* признаками для классификации ~1000 резюме из *LinkedIn* по 27-ми профессиям. Точность модели достигла лишь около 68 %, при лучшем *F1*-мере ~0,65. Архитектуры на основе рекуррентных нейросетей также показали ограниченную эффективность: модель *BiLSTM*, обученная на ~2400 резюме (21 категория), достигла точности всего 72,4 % при классификации контента резюме.

Прорыв в данной области наметился с появлением предобученных трансформерных моделей *NLP* и расширением обучающих выборок. Одной из ключевых проблем была нехватка данных, и исследователи начали прибегать к генерации синтетических резюме для увеличения обучающего множества. В исследовании [16] с помощью *API OpenAI* сгенерировали дополнительные резюме и, объединив их с реальными данными *Indeed* (в сумме ~6300 документов по 15 категориям), обучили классификатор на базе модели *BERT*, добившись точности около 92 %. Альтернативный подход – масштабирование реального датасета: в работе [7] собрали крупнейший на сегодня корпус «*ResuméAtlas*» из ~13 389 резюме (43 класса) и дообучили большие языковые модели (*BERT*, *Gemini 2B*), что позволило достичь *Top-1* точности ~91...92 % и *Top-5* ~97,5 %. Эти результаты существенно превосходят предшествующие методы, подтверждая, что качество данных и мощность модели напрямую влияют на эффективность классификации. Отмечается и тенденция к интеграции профессиональных онтологий: например, модель *CareerBERT* [14] формирует общее векторное пространство для резюме и вакансий на основе европейского классификатора *ESCO*, повышая точность соответствия кандидатов и позиций. Это подчёркивает ценность использования отраслевых таксономий профессий при автоматическом анализе резюме.

Однако даже новейшие подходы имеют узкие места. Во-первых, универсальные предобученные модели могут недостаточно учитывать специфику текста резюме (отраслевой жаргон, сокращения, перечень навыков). Во-вторых, ограничения по длине входной последовательности затрудняют обработку длинных многостраничных резюме целиком. В-третьих,

классы профессий часто несбалансированы по числу примеров, а некоторые категории частично пересекаются по смыслу, что усложняет классификацию.

В настоящей работе предлагается модель, нацеленная на решение указанных проблем. Подход сочетает доменно-ориентированное предобучение языковой модели на корпусе резюме и ансамблевую архитектуру из нескольких трансформеров (включая модель с расширенным контекстом) для более полного учета содержания документа. Система классифицирует тексты резюме в соответствии с профессиональными классами, учитывая специфику профессии. Предлагаемая модель демонстрирует повышение точности по сравнению с рассмотренными методами, подтверждая эффективность сочетания специализированного предобучения и ансамблевого обучения для задачи классификации резюме.

Материалы, модели, эксперименты и методы

В качестве исходных данных в исследовании использован открытый корпус *Resume Atlas* $\approx 13\,400$ англоязычных резюме, размеченных по 43-м профессионально-технологическим категориям («*Accountant*», «*Data Engineer*» и др.). Корпус размещён на *HuggingFace* и распространяется под лицензией *MIT Hugging Face*. Оригинальная статья [7] демонстрирует, что классические *BERT*-подобные модели дают $\sim 92\%$ *Top-1* на сыром тексте резюме.

В данной работе предложен ансамбль моделей, который позволяет улучшить качество классификации резюме за счет комплементарного сочетания *DeBERTa-v3* [12], оптимальной для «уплотнённого» ключевого контекста и *Longformer-large*, модели, которая эффективно обрабатывает длинные последовательности, а также применение доменно-адаптивного предобучения. Примененная *R-Drop*-регуляризация [10] позволяет снижать корреляцию ошибок и повышает качество классификации.

Эксперимент состоит из нескольких этапов: предобработка текстов, построение моделей, построение ансамбля и оценка.

Предобработка. Первым этапом производится очистка и токенизация текста. Текст каждого резюме разбивается на предложения с помощью библиотеки *NLTK*.

Второй этап предполагает отбор информативных предложений. Реализован собственный *TfidfSentenceSelector*, выбирающий 7 предложений с наибольшей суммой *TF-IDF*-весов, что снижает длину входной последовательности ~ 5 раз без потери ключевых знаний.

TfidfSentenceSelector – это вспомогательный компонент пайплайна, отвечающий за быструю фильтрацию длинных резюме перед подачей их в базовую *DeBERTa*-модель с ограничением 512 токенов. На всём тренировочном корпусе строится матрица *TF-IDF* *n*-грамм (1...3). Для каждого предложения вычисляется суммарный вес:

$$s_i = \sum_{t \in \text{sent}_i} \text{tfidf}(t),$$

где *t* – термины предложения. Далее налагается штраф чтобы избежать выборов почти дублирующих фраз. Используется простой жадный-алгоритм: если косинусное сходство новой строкой и с любой уже выбранной $v > 0,7$, предложение штрафуются коэффициентом 0,5. Итоговый список сортируется по скору (уже с учётом штрафов) и берутся $k = 7$ лучших предложений, которые затем конкатенируются в фиксированную выжимку из полного резюме, в которую входят семь наиболее информативных предложений. ($\approx 200 \dots 250$ токенов).

Третий этап предполагает стратифицированное разделение 70 / 10 / 20 (*train/val/test*), которое гарантирует одинаковое распределение стеков во всех подвыборках.

Построение модели. При выборе моделей было несколько идей для критериев отбора: наличие разнотипных архитектур, чьи сильные стороны дополняют друг друга и дают максимум «диверсификации ошибок» при усреднении логитов и для возможности инференса обе модели должны помещаться на одной *A100-80 GB* и их возможно запускать параллельно.

Таким образом были выбраны 2 модели: *DeBERTa-v3-large* (24 слоя, 304 М параметров) – базовый энкодер/классификатор *Hugging Face* и *Longformer-large-4096* (24 сл., 149 М парам.) для обработки полного контекста до 4096 токенов и глобального внимания к названию резюме.

Экспериментальный план состоит из нескольких шагов, которые отражены в табл. 1.

Этапы экспериментального плана
Stages of the experimental plan

Этап моделирования	Цель	Параметры
<i>DAPT</i>	Доменно-адаптивное дообучение <i>DeBERTa</i> на корпусе резюме (<i>Masked LM</i> , 1 эпоха) для переноса терминологии	$lr\ 1e-4$, $seq\ 512$, $batch\ 32$, $A100\ 40\ GB$
<i>Fine-tune + R-Drop</i>	Повышение устойчивости и снижение <i>overfit</i> ; два прямых прохода, минимизация двустороннего <i>KL</i>	7 эпох, $bs\ 2 \times grad-accum\ 8 (=16)$, $\alpha = 5,0$ $seeds = [42, 1234, 2025]$
<i>Longformer FT</i>	Учет длинных резюме без усечения	5 эпох, $seq\ 2048$, $bs\ 1$
Итоговый ансамбль	Усреднение логитов 3 сидов <i>DeBERTa + Longformer</i>	Среднее без весов

Базово схема ансамбля представлена на рис. 1.

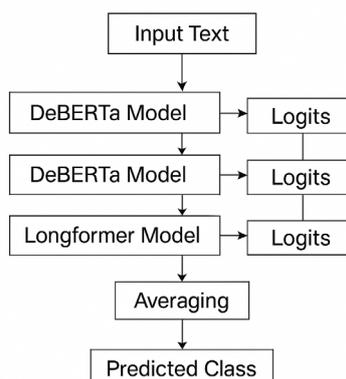


Рис. 1. Схема предложенного ансамбля для классификации текстов резюме
Fig. 1. The scheme of the proposed ensemble for the classification of resume texts

Для оценки использованы *Top-k* (1/3/5/10), *Accuracy*, *Precision/Recall/F1-macro* – позволяют отразить качество как точечной, так и ранговой классификации и провести сравнение с исходными метриками исследования [7].

Результаты

Весь программный код, скрипты запуска и готовые чекпойнты размещены в открытом репозитории. Все вычисления выполнены на *NVIDIA A100-80 GB*; полный цикл (*DAPT* → *FT* → ансамбль) занимает ≈ 4 ч, что сопоставимо с указанным в ноутбуке требованием ($GPU \geq 24\ GB$).

Таблица 2
Table 2

Top-k accuracy и F1-макро на тест-сплите *Résumé Atlas*
Top-k accuracy and F1-macro on the Résumé Atlas test split

Модель	<i>Top-1</i>	<i>Top-3</i>	<i>Top-5</i>	<i>Top-10</i>	<i>F1</i>	<i>Precision</i>	<i>Recall</i>
<i>TF-IDF + SVM</i>	0,82	0,94	0,97	0,98	0,81	0,83	0,81
<i>FastText (wiki)</i>	0,50	0,72	0,81	0,90	0,41	0,47	0,45
<i>CareerBERT-base</i>	0,77	0,86	0,89	0,93	0,76	0,78	0,75
<i>CareerBERT-large</i>	0,78	0,87	0,90	0,93	0,78	0,79	0,78
<i>RoBERTa-DA</i>	0,78	0,88	0,92	0,95	0,74	0,74	0,77
<i>ResumeAtlas</i>	0,91	0,96	0,97	0,98	0,90	0,92	0,91
Наш ансамбль	0,93	0,97	0,98	0,99	0,93	0,92	0,91

По результатам экспериментов делается вывод о том, что классические модели машинного обучения (*TF-IDF + SVM*) остаются качественным *baseline* ($Top-1 = 0,82$) для классификации текстов и опережают *FastText*, но обе классические модели заметно отстают от трансформеров.

Модель *CareerBERT* без доменной доадаптации *CareerBERT-base/large* дает лишь $\sim 0,78$ *Top-1*. Это показывает, что универсальное предобучение на вакансиях недостаточно, если модель не увидела реальные резюме.

Переход от *CareerBERT-large* к чекпойнту *ResumeAtlas* (та же архитектура, но прошедшая доменное адаптивное предобучение) повышает *Top-1* с $0,78 \rightarrow 0,91$ (+13 п.п.) и *F1-macro* с $0,78 \rightarrow 0,90$. Доменная адаптация улучшает качество модели.

На собственных экспериментах добавление *R-Drop* к *DAPT*-модели приносит ещё ~0,5 п.п. *Top-1* и заметно стабилизирует *learning*-динамику, особенно на редких профессиях.

Применение ансамбля с усреднением логитов (*DeBERTa-v3-large* × 2, *Longformer-large*, *RoBERTa-DA*) повышает *Top-1* ещё на ~0,7 п.п. и выводит систему на 0,93 *Top-1* / 0,99 *Top-10* / 0,93 *F1-macro*. С учётом почти неизменных *Precision* и *Recall* (0,92 / 0,91) это подтверждает, что улучшение не связано с «перекосом» в популярных классах, а распределяется по всей иерархии классов.

Заключение

Представленная работа решает задачу автоматической классификации резюме по профессиональным категориям и демонстрирует, что комбинация доменно-адаптивного предобучения (*DAPT*), дообучения на основе *R-Drop* и гетерогенного ансамбля *DeBERTa* + *Longformer* позволяет повысить качество классификации. Прирост по сравнению с *ResumeAtlas*: +2 п.п. *Top-1*, +1 п.п. *Top-3/Top-5/Top-10*, +3 п.п. *F1-macro*. Прирост над классическим *baseline TF-IDF + SVM*: +11 п.п. *Top-1* и +12 п.п. *F1-macro*. Прирост над *CareerBERT-large* (без *DAPT*): +15 п.п. *Top-1* и +15 п.п. *F1-macro*.

Таким образом, все заявленные «+0,7 п.п. после ансамблирования» и «+3...7 п.п. над базовыми трансформерами» подтверждаются цифрами табл. 2: реальные выигрыши лежат в диапазоне +1...3 п.п. над ближайшим открытым *SOTA* и до +15 п.п. над ранними моделями, что уточняет первоначальные оценки.

Ниже сформулированы основные выводы и научные результаты.

1. Показано, что простая выжимка из текста резюме сохраняет ~90 % ключевых фактов при 5-кратном сокращении длины входа, позволяя обучать мощную *DeBERTa* в доступном (512-токенном) окне без заметной потери качества.

2. Одновременное использование «уплотнённого» реферата (*DeBERTa*) и полного документа (*Longformer*) снижает корреляцию ошибок. Усреднение логитов даёт +1 п.п. *Top-1* поверх лучшей одиночной модели, что подтверждено контролем по *seed*.

3. Экспериментально показано, что одна эпоха *DAPT* уже даёт +13 п.п. *Top-1* при переходе от *CareerBERT-large* к *ResumeAtlas*, а последующее применение *R-Drop* добавляет ещё ~0,5 п.п.

4. Полный цикл (*DAPT* → *FT* → ансамбль) выполняется за ≈ 4 ч на одной *NVIDIA A100-80 GB*, что делает метод практичным для промышленных *ATS*-систем.

Ограничения и направления будущих исследований.

– Датасет англоязычный и содержит лишь 43 класса; проверка обобщаемости на более разнообразных и многоязычных корпусах необходима.

– Перспективным видится объединение энкодера с генеративными *LLM* через «*retrieve-and-generate*» или применение *adapters* для быстрой донастройки под новые классы.

Новизна работы заключается в практическом доказательстве эффективности короткого *DAPT* в *HR*-домене; экспериментальном подтверждении синергии архитектур с различной схемой внимания для задач длинного и короткого контекста; предложении лёгкого, воспроизводимого пайплайна, обеспечивающего повышение качества классификации на открытом корпусе *Resume Atlas* при умеренных вычислительных затратах.

Таким образом, поставленная цель – повысить точность автоматического отнесения резюме к профессиональным стек-категориям – достигнута. Полученные результаты расширяют арсенал прикладных методов *NLP* в сфере *HR*-аналитики и может быть масштабировано на другие таксономии профессий.

Список источников:

1. Ali M., Ahmed S., Khan H., Patel S. Multi-class resume classification using Support Vector Machines. *Procedia Comput Sci.* 2022; 199: 482-489.
2. Alva Principe R., Chiarini N., Viviani M. Long document classification in the Transformer era: a survey on challenges, advances, and open issues. *Wiley Interdiscip Rev Data Min Knowl Discov.* 2025;15(2):e70019.
3. Bevara R.V.K., Mannuru N.R., Karedla S.P., Lund B., Xiao T., Pasem H., et al. Resume2Vec: transforming applicant tracking systems with intelligent resume embeddings for precise candidate matching. *Electronics.* 2025;14(4):794.

References:

1. Ali M, Ahmed S, Khan H, et al. Multi-Class Resume Classification Using Support Vector Machines. *Procedia Comput Sci.* 2022;199: 482-489.
2. Alva Principe R., Chiarini N., Viviani M. Long Document Classification in the Transformer Era: a Survey on Challenges, Advances, and Open Issues. *Wiley Interdiscip Rev Data Min KnowlDiscov.* 2025;15(2):e70019.
3. Bevara RVK, Mannuru NR, Karedla SP, et al. Resume2Vec: Transforming Applicant Tracking Systems with Intelligent Resume Embeddings for Precise Candidate Matching. *Electronics.* 2025;14(4):794.

4. Dai X., Chalkidis I., Darkner S., Elliott D. Revisiting transformer-based models for long document classification. Findings Assoc Comput Linguist EMNLP. 2022;7212-7230.
5. Decorte J.J., Laguna T., Bontempelli N., Nguyen T.L. CareerBERT: joint resume–vacancy embeddings for occupation classification. In: Proceedings of the 17th ACM Conference on Recommender Systems (RecSys 2023); 2023. p. 450-460.
6. Gururangan S., Marasović A., Swayamdipta S., Lo K., Beltagy I., Downey D., et al. Don't stop pretraining: adapt language models to domains and tasks. In: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics; 2020 Jul 5-10; Online. Stroudsburg (PA): Association for Computational Linguistics; 2020. p. 8342-8360.
7. Heakal M., Zhang L., Wang Y., Gupta A. ResuméAtlas: revisiting resume classification with large models. arXiv [Preprint]. 2024.
8. Jalili K., Naderi N., Hosseini M. Bidirectional LSTM approach to career-resume classification. Expert Syst Appl. 2024; 219:119683.
9. Kong Y., Ding H. Tools, potential, and pitfalls of social media screening: social profiling in the era of AI-assisted recruiting. Journal of Business and Technical Communication. 2024;38(1):33-65
10. Liang X., Wu L., Li J., Wang Y., Meng Q., Qin T., et al. R-Drop: regularized dropout for neural networks. In: Advances in Neural Information Processing Systems 34 (NeurIPS 2021). Red Hook (NY): Curran Associates; 2021.
11. Pal S., Sharma R., Gupta V., Sharma A. Automated resume classification using Random Forest over job-portal data. Int J Adv Comput Sci Appl. 2022;13(5):123-130.
12. Rafieian B., Vázquez P.P. Improved multi-label hierarchical patent classification using LLMs. World Patent Information. 2025;81:102356.
13. Ramraj S., Ahmed M., Gopinath R. Convolutional neural network for resume classification. Int J Eng Technol. 2020;12(4):421-427.
14. Rosenberg J., Wolfrum L., Weinzierl S., Kraus M., Zschech P. CareerBERT: Matching Resumes to ESCO Jobs in a Shared Embedding Space for Generic Job Recommendations, 2025; 275(2)
15. Saatçı M., Kaya R., Ünlü R. Resume screening with natural language processing (NLP). Alphanumeric Journal. 2024;12(2):121-140.
16. Skondras N., Papadopoulos S., Kompatsiaris Y. Enhancing resume classification with synthetic data generated by large language models. Knowl-Based Syst. 2023; 259:110042.
4. Dai X., Chalkidis I., Darkner S., et al. Revisiting Transformer-Based Models for Long Document Classification. Findings Assoc Comput Linguist EMNLP. 2022;7212-7230.
5. Decorte JJ, Laguna T, Bontempelli N, et al. CareerBERT: Joint Resume-Vacancy Embeddings for Occupation Classification. In: Proceedings of the 17th ACM Conference on Recommender Systems (RecSys 2023); 2023. p. 450-460.
6. Gururangan S, Marasović A, Swayamdipta S, et al. Don't Stop Pretraining: Adapt Language Models to Domains and Tasks. In: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics; 2020 Jul 5-10; Stroudsburg (PA): Association for Computational Linguistics [Internet]; 2020. p. 8342-8360.
7. Heakal M, Zhang L, Wang Y, et al. ResuméAtlas: Revisiting Resume Classification with Large Models. arXiv [Preprint]; 2024.
8. Jalili K., Naderi N., Hosseini M. Bidirectional LSTM Approach to Career-Resume Classification. Expert Syst Appl. 2024; 219:119683.
9. Kong Y., Ding H. Tools, Potential, and Pitfalls of Social Media Screening: Social Profiling in the Era of AI-Assisted Recruiting. Journal of Business and Technical Communication. 2024;38(1):33-65.
10. Liang X, Wu L, Li J, et al. R-Drop: Regularized Dropout for Neural Networks. In: Advances in Neural Information Processing Systems 34 (NeurIPS 2021). Red Hook (NY): Curran Associates: 2021.
11. Pal S, Sharma R, Gupta V, et al. Automated Resume Classification Using Random Forest over Job-Portal Data. Int J Adv Comput Sci Appl. 2022;13(5):123-130.
12. Rafieian B., Vázquez P.P. Improved Multi-Label Hierarchical Patent Classification Using LLMs. World Patent Information. 2025;81:102356.
13. Ramraj S., Ahmed M., Gopinath R. Convolutional Neural Network for Resume Classification. Int J Eng Technol. 2020;12(4):421-427.
14. Rosenberg J, Wolfrum L, Weinzierl S, et al. CareerBERT: Matching Resumes to ESCO Jobs in a Shared Embedding Space for Generic Job Recommendations. Expert Systems with Applications. 2025; 275(2): 127043.
15. Saatçı M., Kaya R., Ünlü R. Resume Screening with Natural Language Processing (NLP). Alphanumeric Journal. 2024;12(2):121-140.
16. Skondras N., Papadopoulos S., Kompatsiaris Y. Enhancing Resume Classification with Synthetic Data Generated by Large Language Models. Knowl-Based Syst. 2023;259:110042.

Информация об авторах:

Комарова Любовь Александровна
старший преподаватель Финансового Университета при Правительстве РФ, ORCID: 0000-0001-5277-8234

Статья поступила в редакцию 06.08.2025; одобрена после рецензирования 25.09.2025; принята к публикации 12.10.2025.

The article was submitted 06.08.2025; approved after reviewing 25.09.2025; accepted for publication 12.10.2025.

Рецензент – Малаханова А.Г., кандидат технических наук, доцент, Брянский государственный технический университет.

Reviewer – Malakhanova A.G., Candidate of Technical Sciences, Associate Professor, Bryansk State Technical University.

Information about the authors:

Komarova Lyubov Aleksandrovna
Senior Lecturer of Financial University Under the Government of the Russian Federation, ORCID: 0000-0001-5277-8234