ЦЕНТР АРХИВНЫХ ДОКУМЕНТОВ,
МЕТОДИЧЕСКИХ МАТЕРИАЛОВ И СТАТИСТИЧЕСКОЙ
ОТЧЕТНОСТИ В СФЕРЕ ОБРАЗОВАНИЯ

Анализ факторов трудоустройства выпускников ВУЗов в сфере информационных технологий

2026

Санкт–Петербург

УДК 331.5:378:004

Скрипкина О. С.[1], Караев Ф. Ш.[1], Терещенков Д. И.[2]
Научный руководитель: Моторина И. Ю., к.э.н. [1]
[1] СПб ГКУ «Центр архивных документов, методических материалов и статистической отчетности в сфере образования», Санкт-Петербург, Россия
[2] Северо-Западный институт управления — филиал РАНХиГС, Санкт-Петербург, Россия

АНАЛИЗ ФАКТОРОВ ТРУДОУСТРОЙСТВА ВЫПУСКНИКОВ ВУЗОВ В СФЕРЕ ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ

Аннотация. Целью исследования является выявление и количественная оценка факторов, влияющих на долю трудоустроенных выпускников ИТ-направлений. Исследование опирается на объединенный массив данных Роструда, Росстата и ЕМИСС за 2019–2024 гг. В качестве базовой модели использована линейная регрессия, в качестве основной - алгоритм градиентного бустинга CatBoost. Для интерпретации результатов применен метод SHAP. Результаты исследования показали, что наиболее оптимальной для описания данных является модель CatBoost. SHAP-анализ выявил, что ведущими факторами трудоустройства являются масштаб выпуска (численность когорты), доля студентов, совмещающих учебу с работой (прокси практического опыта), год выпуска (макроэкономический шок) и институциональный фактор (университет). Региональные факторы оказались менее значимы по сравнению с организационными параметрами программ. В соответствии с полученными результатами в работе сформулированы рекомендации для образовательных организаций, студентов и работодателей.

Ключевые слова: рынок труда, трудоустройство выпускников, ИТ-сектор, машинное обучение, линейное моделирование.

Трудоустройство выпускников образовательных организаций является ключевым индикатором качества подготовки кадров и согласованности образовательной системы с потребностями рынка труда. В сфере информационных технологий (ИТ) этот процесс осложняется быстрым устареванием навыков и высокой нелинейностью взаимосвязей между характеристиками программ, региональным контекстом и макроэкономической конъюнктурой. В связи с этим простое сопоставление числа выпускников и числа вакансий уже не дает достаточного объяснения. Необходимо понимать, какие характеристики образовательной программы, выпускника и региона действительно связаны с более высокой долей занятых после окончания обучения.

Таким образом, цель исследования заключается в выявлении и количественной оценке факторов, влияющих на уровень трудоустройства выпускников образовательных организаций по ИТ-направлениям подготовки.

Проблема исследования состоит в том, что факторы трудоустройства выпускников не действуют изолированно. Уровень образования связан с возрастом и типом программы, региональные условия связаны с уровнем

зарплат и безработицы, а показатели самой образовательной программы могут одновременно отражать как качество подготовки, так и структуру спроса на рынке труда. Традиционные эконометрические подходы, опирающиеся на линейную регрессию, зачастую оказываются несостоятельными при анализе социально-экономических данных подобного рода. Социально-экономические показатели (доходы, безработица, возраст) часто обладают выраженной нелинейностью и взаимодействием (эффектами модификации). Кроме того, наличие категориальных признаков с высоким кардиналом (названия вузов, коды специальностей) требует специфических подходов к кодированию.

Эмпирическую базу составили открытые наборы данных о трудоустройстве и зарплатах выпускников по направлениям подготовки и образовательным организациям, данные мониторинга качества подготовки кадров в образовательных учреждениях, показатели эффективности российских вузов, показатели ЕМИСС и Росстата по доходам населения и безработице, а также классификационные сведения о регионах Российской Федерации [1-4, 6].

Итоговый датасет охватывает период 2019-2024 гг. и содержит 119 080 наблюдений по 20 признакам. Признаки сгруппированы в четыре блока: образовательные (вуз, специальность, уровень образования, численность выпуска), демографические (пол, средний возраст, доля совмещающих работу с учебой), региональные (статус региона, средняя зарплата, безработица) и временные (год выпуска).

В рамках исследования были проведены корреляционный анализ и оценка мультиколлинеарности данных. Проверка на мультиколлинеарность (VIF) выявила сильную корреляцию между различными зарплатными показателями (VIF > 100), а также средним возрастом и годом выпуска. Сравнение коэффициентов корреляции Пирсона и Спирмена показало, что связь между численностью выпуска и долей трудоустройства не является линейной (коэффициент Пирсона близок к 0, тогда как Спирмена составляет -0,224). Это свидетельствует о том, что в работе предпочтительнее использовать нелинейные алгоритмы, при этом, по мнению авторов, для сравнительного анализа целесообразно провести базовое моделирование на основе множественной линейной регрессии. Для нелинейного моделирования авторами применен алгоритм CatBoostRegressor, эффективно работающий с категориальными переменными и нелинейным взаимодействием [7-9]. Интерпретация результатов проведена с применением метода SHAP [10].

Для линейной регрессии категориальные признаки с небольшим числом уникальных значений были закодированы методом one-hot encoding, а признаки с большим числом категорий – через target encoding. Числовые признаки были масштабированы. Данные были разделены на обучающую и тестовую выборки в пропорции 80/20. Такая схема позволила проверить,

насколько простая линейная зависимость способна объяснить долю трудоустроенных выпускников.

Результаты линейной регрессии оказались неудовлетворительными для практической задачи. Средняя абсолютная ошибка (MAE) составила 8,05 процентного пункта, среднеквадратичная ошибка (RMSE) – 10,47 процентного пункта (таблица 1). Для показателя доли трудоустроенных такая ошибка существенна: она может менять интерпретацию эффективности образовательной программы, особенно если сравниваются программы со схожими значениями. Кроме того, линейная модель чувствительна к мультиколлинеарности.

Модель CatBoostRegressor обучалась на тех же содержательных группах признаков. Ключевые параметры включали функцию потерь RMSE, глубину дерева, скорость обучения, регуляризацию L2, количество итераций и случайность при выборе разбиений. Обучение проводилось с разделением на обучающую, валидационную и тестовую части, а также с использованием кросс-валидации. Такой подход снижает риск случайного завышения качества модели на одном разбиении.

Таблица 1. Метрики качества моделей

Модель	R^2	MAE	RMSE	Интерпретация
Линейная регрессия	0,322	8,05	10,47	Ошибка слишком высока для практической интерпретации
CatBoostRegressor	0,6942	5,4574	7,0142	Качество признано удовлетворительным для агрегированных социально-экономических данных

Итоговые метрики CatBoost оказались заметно лучше линейной регрессии (таблица 1). На тестовой выборке R^2 составил 0,6942, MAE – 5,4574 процентного пункта, RMSE – 7,0142 процентного пункта. Среднее значение R^2 на кросс-валидации составило 0,6789. Это означает, что модель объясняет почти 70 % вариации доли трудоустроенных выпускников на тестовой выборке. Для социально-экономических данных с агрегированными наблюдениями такой результат можно считать удовлетворительным.

Сравнение моделей показывает, что переход от линейной регрессии к CatBoost уменьшил среднюю абсолютную ошибку примерно с 8,05 до 5,46 процентного пункта. Улучшение связано с тем, что CatBoost учитывает взаимодействия между признаками и нелинейные зависимости. Например, влияние специальности может зависеть от уровня образования, влияние региона – от статуса региона и зарплатной среды, а влияние численности выпуска – от типа образовательной программы.

При этом результат CatBoost не является идеальным. Ошибка около 5,46 процентного пункта показывает, что часть вариации остается необъясненной. Это ожидаемо, поскольку в данных отсутствуют индивидуальные факторы:

качество портфолио, опыт стажировок, уровень владения конкретными технологиями, активность поиска работы, soft skills, качество карьерного центра и прямые контакты с работодателями. Модель работает с доступными агрегированными признаками, а не с полной биографией выпускника.

По встроенной важности признаков CatBoost на первом месте оказалось количество выпускников с importance 11,50. Далее следуют университет (11,32), доля совмещающих работу с учебой (9,34), регион (8,58), средний возраст (7,65), год выпуска (7,58), код специальности (7,40), средняя зарплата трудоустроенных (7,14), специальность (6,47), группа специальностей (4,94), статус региона (4,87), уровень образования (3,62), пол (3,26), средняя зарплата в регионе (3,21) и средняя безработица в регионе (3,10).

Такая структура важности показывает, что модель использует не один доминирующий фактор, а комбинацию образовательных, демографических, региональных и временных признаков. Однако наиболее значимыми оказались признаки, связанные с траекторией выпускников: численность выпуска, университет, доля совмещающих работу с учебой, код специальности и специальность. Это согласуется с гипотезой о том, что для ИТ-направлений практический опыт часто важнее общей региональной конъюнктуры.

После обучения CatBoost была проведена интерпретация результатов с помощью SHAP-анализа (рисунок 1). Для расчета SHAP-значений использовалась выборка 2000 строк из тестовой части. Такой размер выборки позволил получить устойчивую картину влияния признаков без чрезмерной вычислительной нагрузки. Основной показатель для глобальной интерпретации – среднее абсолютное SHAP-значение, показывающее среднюю силу влияния признака на прогноз модели.

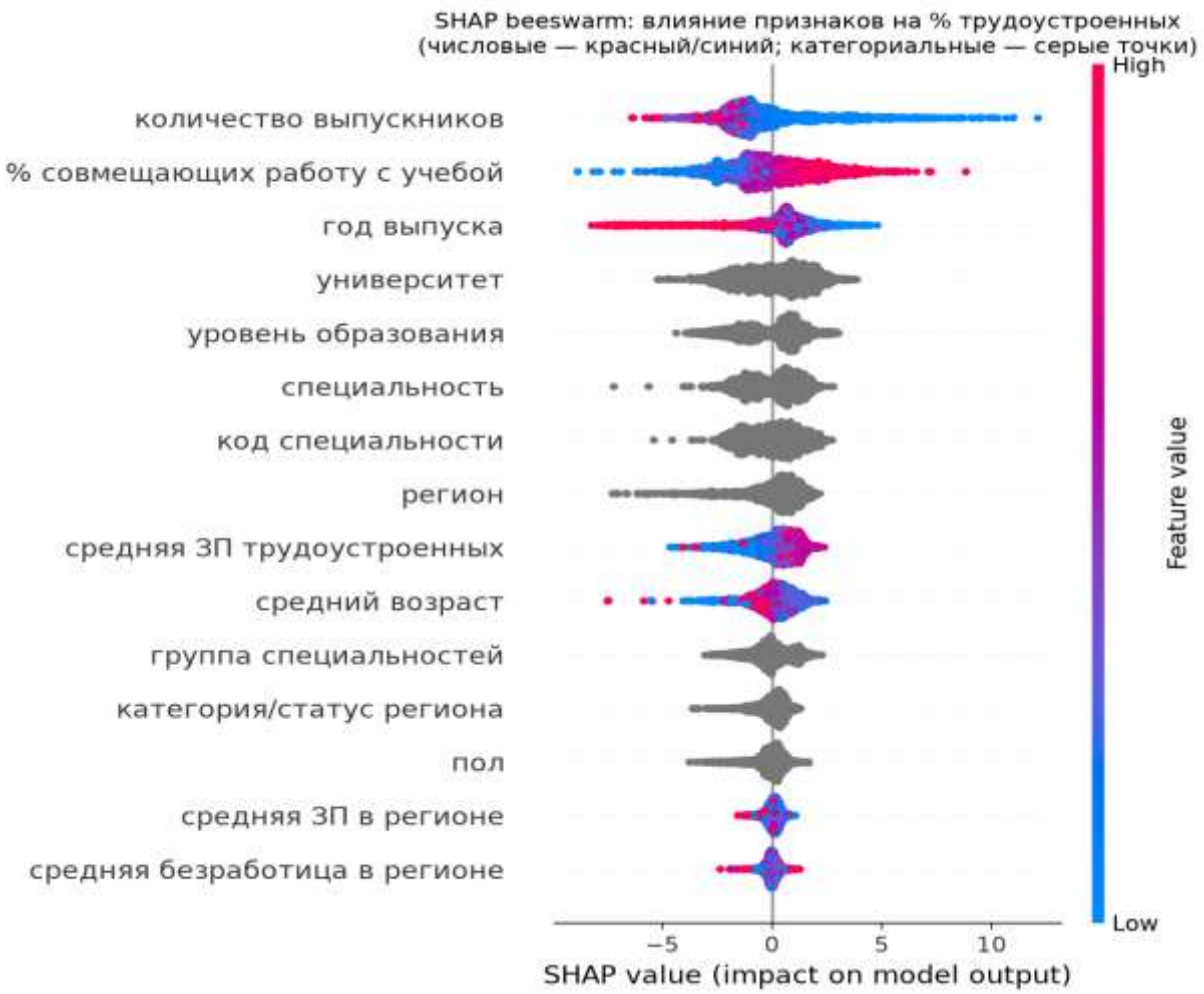

Рисунок 1. SHAP beeswarm: влияние признаков на прогноз модели

Результаты SHAP свидетельствуют, что ключевое влияние на трудоустройство выпускников оказывают не только региональные характеристики рынка труда, но и особенности образовательной программы. Наиболее важными оказываются масштаб выпуска – лидирующая роль фактора «количество выпускников» (mean |SHAP| = 1,873) получает дополнительное объяснение при анализе динамики приема на ИТ-направления. Как показано на рисунке 2, численность зачисленных выросла с 45 000 чел. в 2016 г. до 90 000 чел. в 2024 г., то есть вдвое за восемь лет [5]. Кроме того, наиболее важными факторами являются практическая занятость во время учебы, год выпуска, университет и уровень образования. Такой вывод имеет практическое значение, поскольку образовательные организации могут воздействовать именно на организацию программ, карьерное сопровождение и связь обучения с практикой.

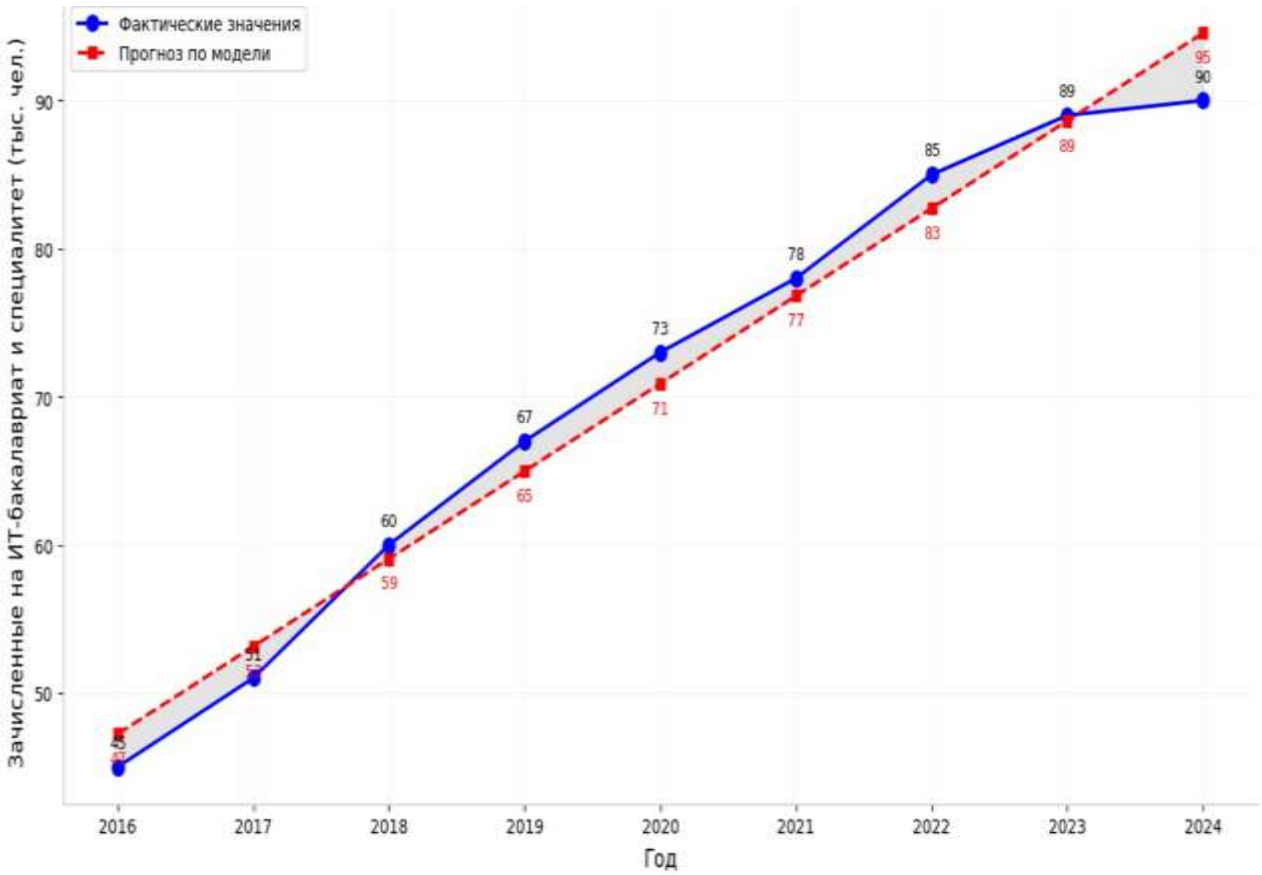

Рисунок 2. Фактические и прогнозные значения численности зачисленных на ИТ-направления в вузы России (на основе данных Tadvisor, 2016-2024 гг.)

Анализ SHAP-значений позволил ранжировать факторы по степени их влияния на долю трудоустроенных выпускников (таблица 2).

Таблица 2. Топ факторов трудоустройства по результатам SHAP-анализа

Ранг	**Признак (фактор)**	**mean |SHAP| (сила влияния)**	**mean SHAP (направление)**	**Содержательная интерпретация**
1	Количество выпускников (масштаб программы)	1,873	−0,121	Массовые программы демонстрируют более низкий процент трудоустройства. Эффект связан с размыванием адресного карьерного сопровождения и высокой внутренней конкуренцией среди выпускников одной когорты.
2	Доля совмещающих работу с учебой (прокси	1,642	−0,061	Раннее погружение в индустрию — ключевой драйвер занятости. Влияние нелинейно и зависит от

Ранг	Признак (фактор)	mean |SHAP| (сила влияния)	mean SHAP (направление)	Содержательная интерпретация
	практического опыта)			контекста (специальности, вуза). Для ИТ-рынка практический опыт часто перевешивает академические показатели.
3	Год выпуска (макроэкономическая конъюнктура)	1,487	0,043	Модель чувствительна к рыночным циклам. Временной фактор агрегирует структурные шоки (пандемия 2020–2021 гг., перестройка сектора 2022–2024 гг.) и циклические колебания спроса на Junior-кадры.
4	Университет (институциональный фактор)	1,292	0,054	Название вуза аккумулирует скрытые эффекты: качество карьерных сервисов, партнерские связи с бизнесом, репутационный капитал и неформальные сети. Выступает классическим «сигналом» для работодателя (теория М. Спенса).
5	Уровень образования	1,167	−0,082	Агрегированный отрицательный знак отражает структурный эффект: программы бакалавриата/специалитета (49,7% выборки) имеют иную динамику трудоустройства по сравнению с магистратурой и требуют отдельного анализа.
6	Специальность	1,01	0,005	Подтверждает гетерогенность внутри ИТ-сферы: программная инженерия, анализ данных и информационная безопасность формируют разные рынки труда и требования нанимателей.
7	Код специальности	0,996	−0,034	Технический детерминант, уточняющий профиль подготовки внутри укрупненных групп направлений.

Ранг	Признак (фактор)	mean |SHAP| (сила влияния)	mean SHAP (направление)	Содержательная интерпретация
8	Регион	0,984	0,022	Географический фактор значим, но уступает образовательным переменным. В условиях распространенности удаленной занятости территориальная привязка выпускников ослабевает.
9	Средняя ЗП трудоустроенных	0,882	0,11	Выступает индикатором качества рабочих мест. Высокие стартовые зарплаты коррелируют с более полным и профильным трудоустройством когорты.
10	Средний возраст	0,735	0,079	Прокси-переменная для траектории обучения: более старшие когорты (магистры, второе высшее) чаще имеют подтвержденный опыт и устойчивые карьерные связи.

Полученные результаты позволяют сформулировать рекомендации для ключевых участников рынка труда, основанные на принципах синергии образования, бизнеса и государства.

1. Для образовательных организаций:

- интеграция образовательных программ с рынком труда. В силу того, что одним из ведущих факторов является доля совмещающих работу с учебой, образовательным учреждениям необходимо встроить практическую занятость в образовательную траекторию. Это может включать стажировки, проектные семестры, партнерские программы с ИТ-компаниями, реальные кейсы в учебных дисциплинах;
- масштабирование карьерной инфраструктуры. При увеличении набора на ИТ-направления, что само по себе не гарантирует высокого процента трудоустройства, учебное заведение обязано пропорционально развивать карьерные центры, оказывающие индивидуальные консультации, помощь в подготовке резюме, к собеседованию, базы стажировок. Иначе увеличение выпуска может приводить к снижению относительного результата трудоустройства. Кроме того, следует отметить, что карьерное сопровождение должно начинаться не на выпускном курсе, а значительно раньше. Для ИТ-направлений уже на младших курсах можно формировать проектную культуру, знакомить студентов с требованиями работодателей, развивать

портфолио и объяснять различия между учебным заданием и промышленной разработкой. Если студент впервые сталкивается с рынком труда только после выпуска, он теряет время и уступает тем, кто начал профессиональную траекторию раньше;

- обновление и корректировка образовательных программ в соответствии с требованиями на рынке труда. При этом фундаментальные дисциплины должны сохраняться, но связываться с практическими задачами. Тогда студент понимает, зачем ему теория, а работодатель видит, что выпускник способен применять ее в работе.

Образовательным организациям следует использовать данные о трудоустройстве не как отчетный показатель, а как инструмент обратной связи. Если по отдельной программе доля трудоустроенных ниже ожидаемой, необходимо анализировать не только учебный план, но и практику, партнерства, работу карьерного центра, структуру выпускников и региональный контекст. Такой анализ должен проводиться регулярно, поскольку рынок ИТ меняется быстрее, чем цикл обновления многих образовательных программ.

2. Для обучающихся:

- накопление практического опыта. Студентам ИТ-направлений целесообразно уже во время обучения формировать портфолио, участвовать в проектах, стажировках, хакатонах, командной разработке и открытых инициативах;
- умение представить результаты обучения на языке работодателя. Для выпускников начального уровня работодатель часто оценивает не только диплом, но и наличие репозиториев, описания проектов, сертификатов, учебных достижений и примеров кода. Это не заменяет реальных компетенций, но помогает сделать их видимыми.

Студентам также важно учитывать региональную структуру рынка труда. Если регион имеет ограниченное число ИТ-вакансий начального уровня, необходимо заранее рассматривать дистанционные форматы, стажировки в других городах, релокацию или работу с федеральными компаниями. При этом дистанционная занятость не отменяет конкуренции: она расширяет рынок, но одновременно увеличивает число претендентов.

3. Для работодателей и HR-специалистов:

- развитие социального партнерства с образовательными организациями. Доля совмещения работы и учебы как значимый фактор указывает на важность раннего профессионального опыта. Работодателям целесообразно развивать стажировочные программы, junior-позиции, проектные школы и партнерство с образовательными организациями, чтобы снижать разрыв между обучением и реальной работой;
- снижение входных барьеров. Если кандидат не имеет большого коммерческого стажа, но демонстрирует завершенные проекты, участие в

командной разработке и способность объяснить принятые решения, он может быть перспективным кандидатом. Слишком жесткие фильтры по стажу могут отсекать сильных молодых специалистов и усиливать проблему входного барьера на рынок труда. Для начальных позиций важнее оценивать обучаемость, базовую техническую подготовку, способность работать в команде и наличие практических проб. Если компания требует от junior-кандидата опыт, характерный для middle-специалиста, она сужает воронку найма и усиливает разрыв между образованием и рынком труда;

- пересмотр формата тестовых заданий. Для выпускников начального уровня слишком сложные или не связанные с будущей работой задания могут плохо измерять реальную пригодность кандидата. Более эффективны задания, приближенные к рабочим ситуациям: анализ небольшого набора данных, исправление ошибки, разработка простого API, оформление SQL-запроса, написание документации или обсуждение архитектурного решения на базовом уровне.

Исследование доказало, что успешность трудоустройства выпускников не является автоматическим следствием востребованности ИТ-отрасли в целом. Она жестко детерминирована организационным дизайном образовательной программы, масштабом выпуска и, что наиболее важно, степенью ранней практической интеграции студента в индустрию. Применение связки CatBoost и SHAP позволило перейти от констатации фактов к выявлению скрытых нелинейных паттернов на рынке труда ИТ-специалистов.

Дальнейшее развитие исследования может быть направлено на обогащение предикторного пространства данными о стажировках, хакатонах и интеграцию балансовых моделей сопоставления выпуска с реальными требованиями работодателей.

Список литературы:

1. Единая межведомственная информационно-статистическая система. Показатель 43062 // ЕМИСС. — URL: https://www.fedstat.ru/indicator/43062
2. Единая межведомственная информационно-статистическая система. Показатель 57039 // ЕМИСС. — URL: https://fedstat.ru/indicator/57039
3. Трудоустройство и зарплаты выпускников по направлениям подготовки // Роструд; обработка «Если быть точным». — URL: https://tochno.st/datasets/
4. Алтунина, В. В. Классификация регионов Российской Федерации в контексте пространственной поляризации / В. В. Алтунина, Д. А. Анучина // Экономика, предпринимательство и право. — 2022. — Т. 12, № 5. — С. 1453–1474.
5. ВУЗы по информационным технологиям // Tadviser. — URL: https://www.tadviser.ru/index.php/
6. Основные тенденции и прогнозы трудоустройства выпускников профессиональных образовательных организаций, подведомственных Комитету по образованию Санкт-Петербурга (2019-2026 гг.) // СПб ГКУ «Центр архивных документов, методических материалов и статистической отчетности в сфере образования». — URL: https://portal.gkuoa.ru/main-page/a-materials/

7. CatBoost: обзор алгоритма и примеры применения // Хабр. — URL: https://habr.com/ru/companies/otus/articles/778714/
8. CatBoost documentation. — URL: https://catboost.ai/docs/en/
9. Prokhorenkova, L. CatBoost: unbiased boosting with categorical features / L. Prokhorenkova, G. Gusev, A. Vorobev, A. V. Dorogush, A. Gulin. — URL: https://arxiv.org/abs/1706.09516
10. SHAP documentation. — URL: https://shap.readthedocs.io/en/latest/