

Олейникова Олеся Олеговна

студентка

Журавлёв Михаил Владиславович

канд. физ.-мат. наук, доцент

ФГБОУ ВО «Российский государственный
аграрный университет – МСХА им. К.А. Тимирязева»

г. Москва

МЕТОДОЛОГИЯ ИНТЕЛЛЕКТУАЛЬНОГО АНАЛИЗА ДАННЫХ В ИССЛЕДОВАНИИ СТРУКТУРЫ СОВРЕМЕННОГО РЫНКА ТРУДА

***Аннотация:** статья посвящена исследованию возможностей применения методологии больших данных для анализа рынка труда IT-специалистов. Рассматриваются ключевые этапы аналитического конвейера: сбор и интеграция гетерогенных данных, предобработка и очистка, преобразование категориальных переменных, статистический анализ и предиктивное моделирование. Особое внимание уделяется проблемам обеспечения достоверности данных и преодоления семантического разнообразия. Обосновывается, что применение методов машинного обучения позволяет не только описывать текущее состояние рынка, но и строить прогнозные модели заработных плат на основе квалификационных и географических признаков. Делается вывод о том, что методология больших данных представляет собой эффективный инструмент исследования сложных социально-экономических систем, обеспечивающий переход от описательной статистики к предиктивной аналитике.*

***Ключевые слова:** большие данные, рынок труда, IT-специалисты, машинное обучение, предиктивное моделирование, анализ данных, визуализация.*

Современный рынок труда IT-специалистов представляет собой сложную динамическую систему, характеризующуюся высокой скоростью изменений, разнородностью источников информации и значительными объемами данных. Традиционные методы анализа, ориентированные на работу с ограниченными струк-

турированными выборками, оказываются неэффективными в условиях информационного изобилия. Это обуславливает необходимость применения методологии больших данных как инструмента, позволяющего получать целостную и достоверную картину рыночных процессов.

Информационная база для анализа IT-рынка труда формируется из множества гетерогенных источников: агрегаторы вакансий, профессиональные социальные сети, результаты опросов специалистов, данные кадровых агентств, аналитические отчеты. Данные о рынке IT-труда сочетают в себе структурированные компоненты (числовые показатели заработных плат, стаж работы), полуструктурированные (JSON-документы с описаниями вакансий) и неструктурированные элементы (текстовые описания требований, комментарии, условия работы). Такое разнообразие создает необходимость применения методов семантического анализа и обработки естественного языка.

Методология больших данных включает инструменты верификации, позволяющие идентифицировать и элиминировать шумовые компоненты – пропущенные значения, дубликаты, статистические выбросы. Применение статистических методов и методов визуализации позволяет выявить аномальные значения и обеспечить достоверность аналитических выводов. Инструментарий анализа выстроен иерархически. На нижнем уровне применяются методы первичной обработки и нормализации. Средний уровень составляют статистические методы и алгоритмы машинного обучения. Верхний уровень включает методы интерпретации и визуализации, трансформирующие числовые результаты в знание, доступное для качественного анализа исследователем. Применение методологии больших данных обеспечивает последовательную эволюцию аналитических возможностей.

Описательная аналитика позволяет рассчитать ключевые метрики: средние и медианные значения заработных плат, стандартное отклонение, частоту встречаемости различных специализаций. Диагностический анализ выявляет факторы, влияющие на уровень доходов. Предиктивное моделирование дает возможность прогнозировать динамику заработных плат и спроса на специалистов на

основе квалификационных и географических признаков. Методология реализует агрегационную функцию, обеспечивая переход от анализа единичных наблюдений к выявлению закономерностей на уровне профессиональных групп, квалификационных категорий, географических кластеров и отраслевых сегментов. Регрессионные модели и алгоритмы машинного обучения, построенные на закодированных признаках специализации, уровня опыта и локации, способны выдавать прогнозные значения ожидаемой заработной платы.

Визуализация данных выступает не просто способом графического представления, а самостоятельным гносеологическим инструментом, переводящим числовые закономерности в наглядные образы, доступные для качественной интерпретации. Гистограммы распределения демонстрируют форму распределения заработных плат, центральную тенденцию и разброс данных. Правостороннее асимметричное распределение, характерное для данных о доходах, указывает на преобладание специалистов с относительно низкими зарплатами при наличии небольшой доли высокооплачиваемых позиций. Интеграция интерактивных инструментов в процесс анализа трансформирует статические отчеты в динамические исследовательские средства.

Инструментарий больших данных характеризуется свойством масштабируемости: алгоритмы, разработанные для анализа ограниченных массивов, сохраняют эффективность при увеличении объемов информации. Это особенно значимо для динамично развивающегося IT-сектора с постоянно растущим количеством участников рынка и объемом генерируемых данных. Применение методов больших данных позволяет обнаруживать неочевидные корреляции и факторы, влияющие на уровень доходов IT-специалистов. К таким латентным факторам относятся комбинации компетенций, неформальные требования работодателей, сезонные колебания спроса, региональная специфика, которые недоступны традиционным методам анализа, ориентированным на ограниченные выборки и априорные гипотезы. Использование методологии больших данных в анализе рынка труда IT-специалистов знаменует сдвиг исследовательской парадигмы от гипотетико-дедуктивного подхода, ориентированного на проверку априорных гипотез,

к индуктивному анализу, основанному на выявлении закономерностей непосредственно в эмпирических данных с последующим теоретическим осмыслением. Результаты применения методологии больших данных имеют практическую ценность для всех участников рынка. Таким образом, методология больших данных, апробированная на анализе IT-рынка труда, представляет собой эффективный инструмент исследования сложных социально-экономических систем. Несмотря на разнородность используемых методов и алгоритмов, она сохраняет методологическую целостность, поскольку все ее элементы подчинены единой цели – извлечению из эмпирических данных скрытого знания о закономерностях функционирования рынка труда IT-специалистов, пригодного для принятия обоснованных решений.

Список литературы

1. Информационные системы и технологии в АПК: учебник / А.В. Бабкина, И.Е. Быстренина, М.И. Горбачев [и др.]. – М.: Российский государственный аграрный университет, 2025. – 615 с. EDN ZJCSXJU
2. Моделирование тепловых потоков в подземных горных выработках с применением параллельного программирования / М.В. Журавлев, Т.И. Ашмарина, И.А. Вахрушева, К.С. Музалев // Известия Тульского государственного университета. Науки о Земле. – 2025. – №3. – С. 292–297. EDN KPCQVA
3. Проектирование модуля информационной системы для дошкольного образовательного учреждения / М.В. Журавлев, А.Р. Глинская, С.В. Кукарцева, В.Ю. Панченко // Актуальные вопросы гуманитарных и социальных наук: материалы Всероссийской научно-практической конференции с международным участием. – Чебоксары: Среда, 2026. – С. 131–134. EDN OXJNHR
4. Прогнозирование геоэкологических рисков карьерных территорий на основе гибридной модели машинного обучения / В.В. Кукарцев, А.С. Прудкий, М.С. Никаноров, К.С. Музалев // Известия Тульского государственного университета. Науки о Земле. – 2025. – №4. – С. 80–85. EDN WFTBXO

5. Экономическое влияние цифровой зрелости и ИИ на производительность: прогнозирование, факторный анализ и выводы / Р.А. Нажмутдинов, Д.Р. Ишмухаметова, А.С. Прудкий, С.В. Кукарцева // Современные тренды управления, экономики и предпринимательства: от теории к практике: сборник материалов III Всероссийской научно-практической конференции с международным участием (Ульяновск, 11–13 ноября 2025 года). – Чебоксары: Среда, 2025. – С. 123–129.
EDN IZFNCS