Терёшина Влада Валерьевна
канд. экон. наук, доцент
ФГБОУ ВО «МИРЭА – Российский технологический университет»
г. Москва
Савченкова Елена Валерьевна
канд. экон. наук, доцент
ФГОБУ ВО «Финансовый университет при Правительстве Российской Федерации»
г. Москва
Терёшин Кирилл Александрович
студент
ФГБОУ ВО «МИРЭА – Российский технологический университет»
г. Москва

DOI 10.31483/r-171008

ПРОГНОЗИРОВАНИЕ СРЕДНЕСПИСОЧНОЙ ЧИСЛЕННОСТИ РАБОТНИКОВ МЕТОДАМИ МАШИННОГО ОБУЧЕНИЯ

Аннотация: *в статье исследуется прогнозирование среднесписочной численности работников организаций по видам деятельности по российским официальным данным. Построена линейная регрессия с временными признаками и проведено сравнение с прогнозом по последнему наблюдению. Рассчитана средняя абсолютная ошибка на отложенных данных. Обсуждаются применимость результатов и ограничения короткой выборки.*

Ключевые слова: *машинное обучение, прогнозирование, регрессия, временной ряд, ошибка прогноза.*

В работах по сценарному планированию рассматривается применение моделей машинного обучения для работы с неопределённостью [1, с. 361–372]. Опыт исследования российского рынка показывает значимость анализа изменения

Content is licensed under the Creative Commons Attribution 4.0 license (CC-BY 4.0)

условий и сценариев [2, с. 4889–4906]. Цель исследования – оценить применимость простой обучаемой модели к выбранному показателю, сопоставив её с прозрачным базовым прогнозом.

Анализ среднесписочной численности работников организаций по видам деятельности важен для оценки изменений экономической среды и подготовки управленческих решений. Прогноз позволяет заранее рассмотреть диапазон возможных результатов, однако его качество необходимо оценивать на данных, которые не использовались при обучении модели. Исследование выполнено на материалах Республики Карелия. Из статистических данных выбраны двенадцать видов деятельности. Входом модели является численность работников в отрасли за предыдущий год, ответом – численность в следующем году.

Таблица 1

Исходные данные о среднесписочной численности работников организаций Республики Карелия по видам экономической деятельности в 2020–2024 гг., тыс. человек

Вид экономической деятельности	2020	2021	2022	2023	2024
Сельское, лесное хозяйство, охота, рыболовство и рыбоводство	7,4	7,2	7,3	7,1	7,0
Добыча полезных ископаемых	6,5	6,8	7,2	7,2	7,7
Обрабатывающие производства	22,1	22,0	21,0	20,6	19,8
Обеспечение электрической энергией, газом и паром; кондиционирование воздуха	9,0	8,7	8,3	8,3	8,0
Водоснабжение; водоотведение, организация сбора и утилизации отходов, деятельность по ликвидации загрязнений	2,4	2,2	2,0	2,5	2,1
Строительство	8,5	8,3	8,3	8,2	7,9
Торговля оптовая и розничная; ремонт автотранспортных средств и мотоциклов	20,8	20,1	20,2	19,7	19,8
Транспортировка и хранение	15,7	15,3	15,0	14,9	14,8
Деятельность гостиниц и предприятий общественного питания	4,0	3,6	3,6	4,0	4,1
Образование	22,5	22,1	21,7	21,7	21,6
Деятельность в области здравоохранения и социальных услуг	19,5	19,7	19,2	19,0	18,9
Деятельность в области культуры, спорта, организации досуга и развлечений	3,9	3,8	4,0	3,6	3,4

**Источник: статистический сборник Карелиястат, таблица 3.3, с. 27. Доступ к таблице: официальная публикация статистических данных https://10.rosstat.gov.ru/*

Содержимое доступно по лицензии Creative Commons Attribution 4.0 license (CC-BY 4.0)

Данные с ответами 2021–2023 гг. формируют обучение; 2024 г. оставлен для проверки по всем выбранным отраслям.

Линейная регрессия обучается методом наименьших квадратов. Её параметры выбираются по обучающему периоду так, чтобы уменьшить сумму квадратов разностей между наблюдаемыми и рассчитанными значениями. При проверке используются только данные, доступные до прогнозируемого значения. Базовый метод принимает прогноз равным последнему известному значению показателя.

Точность оценивается по средней абсолютной ошибке:

$$MAE = (1/n) \times \Sigma |y_i - \hat{y}_i| \quad (1)$$

где y_i – фактическое значение,

$\hat{y}_i$ – прогноз,

n – число контрольных наблюдений.

Меньшая MAE означает меньшую среднюю абсолютную ошибку именно на рассматриваемой проверочной выборке.

Проверка отделена по времени, чтобы не переносить сведения о будущем в обучение. Такой порядок не устраняет влияние случайного выбора контрольного периода. Исследования прогнозирования экономических показателей методами машинного обучения также указывают на необходимость сопоставления моделей и проверки их ограничений [3, с. 1241–1252].

Таблица 2

Сопоставление фактических и модельных значений

Период / группа	Факт	Модель
Сельское	7,00	7,04
Добыча	7,70	7,14
Обработка	19,80	20,27
Энергетика	8,00	8,22
Вода	2,10	2,53

В обучающей выборке использовано 36 наблюдений, в контрольной – 12. Средняя абсолютная ошибка модели составила 0,27 тыс. чел., базового прогноза – 0,26 тыс. чел. Результат относится только к выбранному контрольному периоду и набору объектов.

Content is licensed under the Creative Commons Attribution 4.0 license (CC-BY 4.0)

На выбранной контрольной выборке ошибка линейной модели больше ошибки метода последнего наблюдения. Это сравнение важно: даже сложный алгоритм должен показывать практическое преимущество перед простым правилом. Результат нельзя автоматически переносить на другие годы или экономические условия.

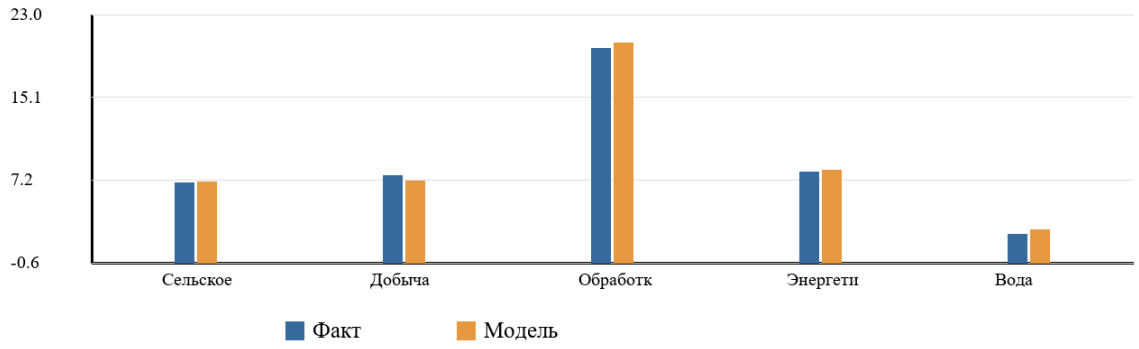


Рис. 1. Фактические и прогнозные значения (тыс. чел.)

Различие между отдельными фактическими и прогнозными значениями в таблице и на рисунке показывает, что средний показатель ошибки не описывает все случаи одинаково. Для решения прикладной задачи необходимо отдельно рассматривать величину неблагоприятных отклонений, единицы измерения и условия, в которых готовится прогноз.

Использование ретроспективных данных требует соблюдения временной последовательности: признаки формируются на основании уже опубликованных наблюдений, а значения проверяемого периода не участвуют в подборе параметров. Дополнительное тестирование на нескольких непересекающихся периодах позволило бы уточнить устойчивость оценки.

Это показатель работников организаций, а не все занятые региона. Состав отраслей неоднороден; общая регрессия не отражает региональные вакансии, демографию и изменения классификации. Одна проверка на одном году не гарантирует устойчивости. Для прикладного применения целесообразно дополнить модель экономически обоснованными признаками и сравнить её с несколькими альтернативами. При разработке сценариев необходимо учитывать возможность

Содержимое доступно по лицензии Creative Commons Attribution 4.0 license (CC-BY 4.0)

структурных изменений, на которую обращается внимание в исследованиях российских финансовых рынков.

В результате можно сформулировать следующие выводы. На российских официальных данных выполнен ретроспективный эксперимент по прогнозированию среднесписочной численности работников организаций по видам деятельности. Средняя абсолютная ошибка модели составила 0,27, а базового правила – 0,26 тыс. чел. Оценка демонстрирует результат конкретного расчёта, но не является гарантией точности будущих прогнозов. Поэтому следующий этап исследования – расширение ряда и повторение проверки на последовательных периодах.

Список литературы

1. Ефремов, В. С. Совершенствование сценарного планирования инструментами машинного обучения / В. С. Ефремов, И. Никитинс // Информатизация в цифровой экономике. – 2024. – Т. 5, № 3. – DOI 10.18334/ide.5.3.121501. – URL: https://elibrary.ru/item.asp?id=74511222 (дата обращения: 15.11.2026). – Текст : электронный.

2. Журавлева, О. В. Сценарии развития российского фондового рынка в условиях усиления санкционного давления / О. В. Журавлева, О. Н. Митрофанова, М. В. Шашкин // Экономика, предпринимательство и право. – 2024. – Т. 14, № 9. – DOI 10.18334/epp.14.9.121678. – URL: https://elibrary.ru/item.asp?id=73162090 (дата обращения: 15.11.2026). – Текст : электронный.

3. Матыцын, А. В. Применение LSTM-модели к прогнозированию CPI и уровня инфляции на примере России / А. В. Матыцын // Вопросы инновационной экономики. – 2022. – Т. 12, № 2. – 1300 с. – DOI 10.18334/vinec.12.2.114845. – URL: https://elibrary.ru/item.asp?id=48939626 (дата обращения: 15.11.2026). – Текст : электронный.

Content is licensed under the Creative Commons Attribution 4.0 license (CC-BY 4.0)