

ИНТЕГРАЦИЯ МАШИННОГО ОБУЧЕНИЯ В ТРАДИЦИОННУЮ ЭКОНОМИЧЕСКУЮ СТАТИСТИКУ

М.С. Барышников

Территориальный орган Федеральной службы государственной статистики по Саратовской области, специалист 1 разряда, г. Саратов, Россия

64.baryshnikovms@rosstat.gov.ru

Экономика является одной из наиболее активно использующих машинное обучение среди всех сфер деятельности. Экономические циклы, с их периодическими колебаниями экономической активности, представляют собой идеальную область применения методов машинного обучения для анализа и прогнозирования экономических тенденций.

Экономические циклы являются фундаментальной характеристикой развития национальных экономик современных стран. Эти периодические колебания экономической активности играют ключевую роль в формировании экономической стабильности и роста [1]. Для обеспечения устойчивого экономического развития критически важно не только понять причины возникновения экономических циклов, но и развить способность точно предсказывать начало новых циклов.

Понимание перспектив экономического развития имеет важное значение для бизнес-планов и государственной экономической политики. Методы машинного обучения становятся все более востребованными в этом контексте благодаря своей способности анализировать сложные и крупномасштабные данные.

Традиционные эконометрические модели, основанные на простых линейных связях, все чаще дополняются или заменяются более гибкими алгоритмами машинного обучения. Эти алгоритмы могут выявлять скрытые закономерности в данных и учитывать множество факторов, влияющих на экономическую динамику.

Машинное обучение становится все более востребованным инструментом в экономике благодаря своей способности автоматизировать анализ и обработку больших объемов данных. Оно позволяет сократить трудозатраты на рутинные процессы и создавать адаптивные решения без необходимости написания комплексного алгоритма для каждой задачи [2]. Основное отличие машинного обучения заключается в его способности самостоятельно определять зависимости между входными и выходными данными.

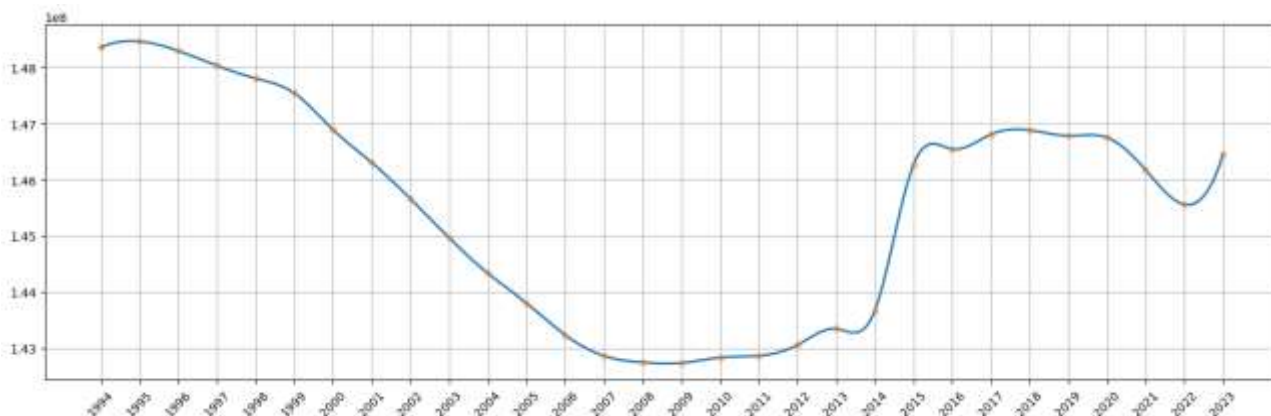


Рисунок 1. График численности постоянного населения России

Цель данного исследования состоит в том, чтобы предсказать численность населения России на 2024 год, а затем сравнить эти прогнозы с фактическими данными, предоставленными Росстатом [3]. Для этого будет проведен анализ показателей численности населения с 1994 по 2023 год, показанных на рис. 1.

Алгоритмы машинного обучения используют выборку данных для построения функции, которая затем может быстро классифицировать новые входные данные или прогнозировать вероятные значения. Для реализации этих методов в данном исследовании использовался язык программирования Python с библиотекой sklearn, предоставляющей широкий спектр алгоритмов машинного обучения, включая линейную и полиномиальную регрессию.

При работе с прогнозированием экономических циклов можно выделить два основных типа методов машинного обучения: классификацию и регрессию. Методы классификации используются для определения функции, которая оптимально разделяет входные данные на группы, позволяя предсказывать вероятность принадлежности к определенному экономическому циклу. Методы регрессии направлены на точное предсказание следующего элемента последовательности экономических показателей, что позволяет прогнозировать конкретные значения в будущем. Оба типа методов работают по схожему принципу: сначала обучается функция на выборке данных, а затем эта функция применяется к новым данным для получения прогнозов [4].

Линейная регрессия представляет собой базовый метод машинного обучения, необходимый для понимания сути регрессионного подхода. Цель линейной регрессии – обучить линейную функцию, которая будет иметь наименьшую ошибку сходства с действительной численностью населения России, то есть найти линейную функцию, наиболее приближенную к реальному значению численности населения.

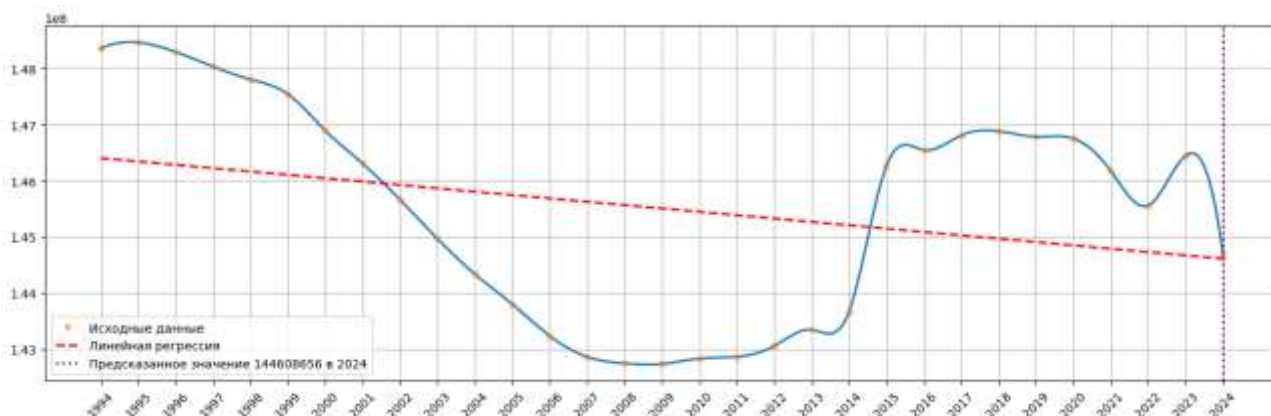


Рисунок 2. Результат прогнозирования с помощью метода линейной регрессии

На рис. 2 представлен прогноз численности населения Российской Федерации до 2024 года, основанный на методе линейной регрессии. Коэффициент детерминации в этом случае равен 0,88. Предсказанное значение 144,6 млн отличается от фактического.

Хотя этот метод позволяет делать достаточно точные прогнозы на короткий срок, он имеет значительные ограничения при прогнозировании долгосрочных демографических тенденций. Основные проблемы линейной регрессии связаны с линейной зависимостью, которая не всегда точно отражает реальные демографические процессы. Кроме того, сложно учесть непредсказуемые факторы, влияющие на численность населения, такие как изменения в рождаемости, смертности и миграции. Также существует вероятность изменения тренда в

долгосрочной перспективе, и исторические данные могут не отражать будущие события и изменения в стране.

Учитывая эти ограничения, для более точных прогнозов на длительные периоды необходимо использовать более сложные методы машинного обучения.

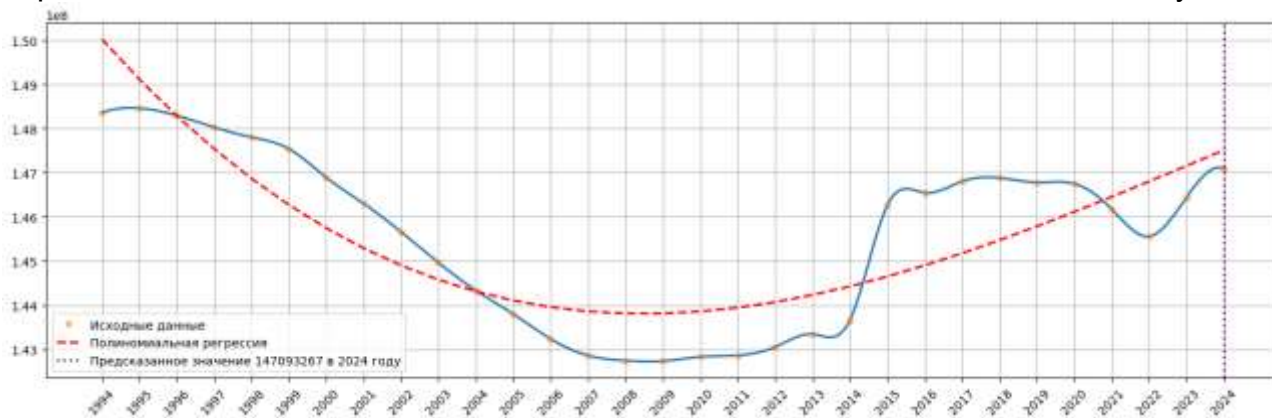


Рисунок 3. Результат вычисления полинома с помощью метода полиномиальной регрессии

Полиномиальная регрессия, показанная на рис. 3 значительно повышает точность прогнозирования по сравнению с линейной. Однако предсказанное значение 147,1 млн всё равно не совпадает с фактическим.

Процесс работы алгоритма полиномиальной регрессии похож на линейную, но с одним важным отличием: он включает трансформацию исходной линейной функции в более сложную полиномиальную. Это преобразование позволяет модели лучше адаптироваться к особенностям демографического развития страны, что существенно улучшает качество предсказаний [5].

Коэффициент детерминации для полиномиальной регрессии достигает высокого значения 0,95, что свидетельствует о высокой степени соответствия между прогнозируемыми и фактическими данными о численности населения.

Хотя модель полиномиальной регрессии хорошо описывает общую структуру графика численности населения, ее предсказательная сила оказывается ограниченной. Эту ограниченность можно проиллюстрировать путем эксперимента с сокращением исходных данных и попыткой предсказания отсутствующих элементов

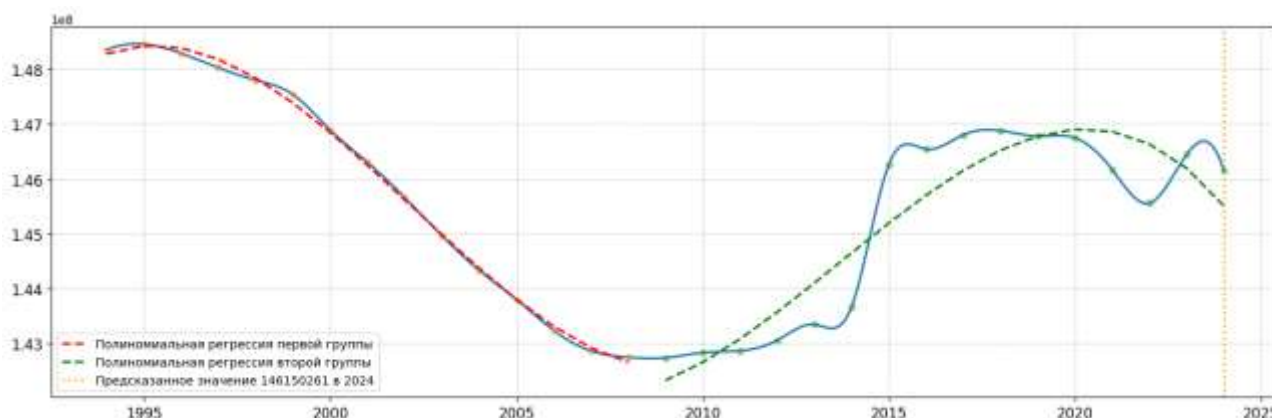


Рисунок 4. Результат прогнозирования с помощью метода полиномиальной регрессии

(рис. 4).

В данном случае коэффициент детерминации для предсказанного участка составляет 0,82, что считается удовлетворительным показателем. Однако стоит отметить, что предсказанное значение составило 146,15 млн человек. Согласно данным Росстата, численность населения на 1 января 2024 года составляет 146,2 млн человек. Это почти полное совпадение подтверждает высокую точность и эффективность методов машинного обучения в прогнозировании демографических трендов.

Таким образом, результаты исследования демонстрируют высокую эффективность и точность методов машинного обучения в прогнозировании демографических трендов. Сравнительный анализ предсказанных и фактических данных подтверждает, что использование машинного обучения позволяет значительно улучшить качество прогнозов и повысить их достоверность. Это открывает новые перспективы для применения данных методов в различных областях, связанных с анализом и прогнозированием социальных и экономических процессов.

Список использованной литературы

1. Веселов Д.И. Основные проблемы промышленных предприятий в условиях цифровизации // Прогрессивная экономика. 2024. № 3. С. 5–13.

2. Поленок М.В., Бондаренко С.В., Козлова И.Р., Юркова О.Н. О методах машинного обучения при принятии управленческих решений в области здравоохранения / Вызовы цифровой экономики: тренды развития в условиях последствий пандемии COVID-19: Сборник статей IV Всероссийской научно-практической конференции, приуроченной к Году науки и технологий в России. Брянск: Брянский государственный инженерно-технологический университет. 2021. С. 225–229.

3. Федеральная служба государственной статистики. URL: <https://rosstat.gov.ru>.

4. Маслакова П.И., Осенний В.В., Бурда А.Г. Направления использования машинного обучения и искусственного интеллекта в экономики и перспективы их развития // Цифровизация экономики: направления, методы, инструменты : Сборник материалов II всероссийской научно-практической конференции. Краснодар: Кубанский государственный аграрный университет имени И.Т. Трубилина, 2020. С. 51–54.

5. Давыдов, А. Н. использование машинного обучения в экономике // Современные проблемы лингвистики и методики преподавания русского языка в ВУЗе и школе. 2022. № 38. С. 845–848.