



ПРОБЛЕМЫ УПРАВЛЕНИЯ РАЗВИТИЕМ ЧЕЛОВЕЧЕСКОГО ПОТЕНЦИАЛА

УДК [314.3:004](470.53)

DOI: 10.17072/2218-9173-2021-2-204-221

ПРОГНОЗИРОВАНИЕ ДЕМОГРАФИЧЕСКИХ ПОКАЗАТЕЛЕЙ В СФЕРЕ РОЖДАЕМОСТИ НАСЕЛЕНИЯ: ИНЕРЦИОННЫЙ ПРОГНОЗ VERSUS ПРОГНОЗ НА ОСНОВЕ МАШИННОГО ОБУЧЕНИЯ

Н. Ю. ЗУБАРЕВ

Пермский государственный национальный исследовательский университет,
г. Пермь, Россия

Д. Д. ФЕДУЛОВА

Пермский государственный национальный исследовательский университет,
г. Пермь, Россия

Для цитирования:

Зубарев Н. Ю., Федулова Д. Д. Прогнозирование демографических показателей в сфере рождаемости населения: инерционный прогноз versus прогноз на основе машинного обучения // *Ars Administrandi* (Искусство управления). 2021. Т. 13, № 2. С. 204–221. DOI: 10.17072/2218-9173-2021-2-204-221.

АННОТАЦИЯ

Введение: статья посвящена сравнению методов прогнозирования демографических показателей в сфере рождаемости населения. **Цель:** определить точность методов прогнозирования демографических показателей в сфере рождаемости населения на основании сравнения прогнозных значений, полученных с помощью инерционного метода и метода машинного обучения (на данных Пермского края). **Методы:** статистический анализ, графический анализ, инерционный метод прогнозирования, метод машинного обучения на основе градиентного бустинга (Catboost) с использованием программной среды *Google Collab* и языка программирования *Python* версии 3.7. **Результаты:** получены прогнозные значения показателя «абсолютная численность родившихся». Средняя ошибка отклонения прогнозных значений от фактических для метода инерционного прогнозирования составила 11,9 %, а для метода машинного обучения – 19,85 %. Выявлены особенности формирования прогнозных значений для каждого метода и обоснованы высокие значения отклонений. **Выводы:** метод инерционного прогнозирования оказался более точным, чем метод

машинного обучения, при этом оба метода показали низкую точность. Это связано с ограничениями датасета, доступного и используемого для расчетов, а также с ухудшением демографической ситуации в сфере рождаемости в Пермском крае и резким изменением динамики прогнозируемого показателя.

Ключевые слова: рождаемость населения; машинное обучение; инерционный прогноз; прогнозирование

ВВЕДЕНИЕ

Воспроизводство населения представляет собой комплексный процесс, объединяющий две составляющие: процесс рождаемости и смертности населения. Проблема низкой рождаемости характерна для Российской Федерации на протяжении всей новейшей истории. Вкупе с высокой смертностью низкая рождаемость привела к интенсивной депопуляции и отрицательному естественному приросту (Курносова, 2019). Поэтому государственная демографическая политика в сфере рождаемости направлена на увеличение численности населения, прежде всего за счет увеличения количества родившихся.

Для изучения процесса рождаемости населения в демографии используют систему определенных показателей, базовым в которой является показатель «абсолютная численность родившихся». На его основании рассчитываются относительные и специализированные показатели. Они дают представление о росте численности населения и служат для оценки медицинского, социального и демографического благосостояния населения. Показатели рождаемости необходимы как для общего анализа процесса рождаемости, так и для различных демографических исследований, а также для работы органов власти в сфере здравоохранения и социальной политики.

Управление процессом рождаемости и на федеральном, и на региональном уровнях тесно сопряжено с необходимостью разработки прогнозов абсолютной численности родившихся как основного результирующего показателя проводимой демографической политики и эффективности мер воздействия.

ТЕОРЕТИЧЕСКИЕ ОСНОВЫ ИССЛЕДОВАНИЯ

Демографический прогноз в сфере рождаемости – это обоснованное научными методами предвидение будущей демографической ситуации, прежде всего абсолютной численности родившихся и ее структуры (Архангельский и Елизаров, 2016, с. 524; Андреев, 1999, с. 176). В основе прогноза, как правило, лежат перспективные расчеты, построенные на некоторых общих гипотезах относительно будущей динамики показателей рождаемости (Тарасова, 2015, с. 186).

В данной статье будут рассмотрены два метода прогнозирования показателей: метод инерционного прогнозирования и метод машинного обучения на основе градиентного бустинга. Методам факторного анализа и корреляционно-регрессионного моделирования будет посвящена отдельная статья.

Инерционному прогнозированию свойственно выполнение расчетов в нескольких вариантах. Иногда прогнозом называют только общую оценку будущей численности населения (Архангельский и Елизаров, 2016, с. 525). Основным методом расчета является метод передвижки возрастов, в западной литературе его чаще называют методом компонент (Cohort Component Method) (Сафиуллин и др., 2014; Юзаева, 2014; Smith et al., 2013; Riiman et al., 2019; Thatkar et al., 2018; Sitkovskiy, 2019). Часто исследователи, определив основную тенденцию на снижение или увеличение численности населения, экстраполируют ее в форме среднего темпа роста / снижения на будущие показатели (Шубат и Киселева, 2016, с. 330). Несмотря на кажущуюся простоту данного метода, иногда он дает достаточно точные прогнозы.

Росстат, как правило, рассчитывает три варианта прогноза на основе гипотез относительно будущих тенденций рождаемости (низкий, средний и высокий). Низкий вариант прогноза основан на экстраполяции существующих демографических тенденций, высокий вариант является нормативным и ориентирован на достижение целей, определенных в Концепции демографической политики Российской Федерации на период до 2025 года¹. Средний вариант прогноза считается наиболее реалистичным, в нем учтены сложившиеся демографические тенденции и принимаемые меры демографической политики².

Точность прогноза зависит от точности исходной информации. К основным источникам помимо статистических данных относят данные переписей населения, результаты выборочных исследований. При этом важную роль играют текущие демографические тенденции, определяющие тренд и динамику изменения показателей рождаемости. Точность и достоверность прогноза чрезвычайно чувствительна к продолжительности периода, на который делается прогноз (краткосрочный – до 5 лет, среднесрочный – от 5 до 30 лет, долгосрочный – свыше 30 лет). Чем меньше период прогноза, тем точнее результат инерционного прогнозирования.

Ошибки в определении демографических тенденций могут дать существенное отклонение прогноза, снизив его ценность, и привести к значительным ошибкам в реализации социально-экономической политики государства и регионов. Одним из способов определения гипотез возможной динамики рождаемости, используемой Росстатом, является экстраполяция тренда, имевшего место в предшествующий период. Данный прием особенно чувствителен к длине динамического ряда, который должен быть не короче прогнозируемого периода. Использование этого способа определения гипотез возможной динамики показателей рождаемости целесообразно применять при незначительном отклонении прогнозируемых показателей от тренда. Прогноз перспективной динамики показателей рождаемости, основанный на экстраполя-

¹ Об утверждении Концепции демографической политики Российской Федерации на период до 2025 года [Электронный ресурс]: Указ Президента Рос. Федерации от 09.10.2007 № 1351. URL: <https://base.garant.ru/191961/> (дата обращения: 10.04.2021).

² Демографический ежегодник России. 2019 [Электронный ресурс]: стат. сб. М.: Росстат, 2019. 252 с. URL: https://rosstat.gov.ru/storage/mediabank/Dem_ejegod-2019.pdf (дата обращения: 10.04.2021).

ции их трендов, часто может подвергаться замедлению изменения величины показателя по мере приближения к пороговым значениям.

Второй метод, рассмотренный в статье – это метод машинного обучения на основе градиентного бустинга (Catboost), который позволяет создавать множество алгоритмов (деревьев решения), которые, в свою очередь, способны научиться принимать решения и строить прогнозы на основе данных (Brink et al., 2016; Ibrahim et al., 2020; Hancock and Khoshgoftaar, 2020). Машинное обучение – сравнительно новый метод, при этом для демографической науки он редко применялся ранее (Соловьев и Соловьев, 2018, с. 948). Между тем данный метод постепенно завоевывает популярность в других разделах экономической науки (Kauffman et al., 2017; Köhn and Lessmann, 2020; Zhang et al., 2020).

Идея бустинга состоит в комбинации функций с невысокой обобщающей способностью, которые строятся в ходе итеративного процесса, на каждом шаге которого новая модель обучается с использованием данных об ошибках предыдущих итераций. Итоговая результирующая функция представляет собой линейную комбинацию базовых, слабых моделей. В итоге данный метод создает решающую модель прогнозирования в виде ансамбля слабых моделей, обычно деревьев решений. Он строит модель последовательно, оптимизируя произвольную дифференцируемую функцию потерь.

В рамках данной статьи мы определим точность методов прогнозирования демографических показателей в сфере рождаемости населения на основании сравнения прогнозных значений, полученных с помощью описанных выше методов.

При этом важным остается вопрос выбора факторов, под влиянием которых формируются показатели рождаемости. Разнообразное сочетание факторов обуславливает уровень рождаемости населения (Малева и Синявская, 2006).

Н. Н. Подольная и коллеги разделяют факторы рождаемости на следующие группы (Подольная и др., 2015):

- природно-биологические факторы – экологическая обстановка, детородный возраст, наследственность и т.д.;
- социально-экономические факторы – уровень жизни населения, система здравоохранения, занятость населения, войны и т.д.;
- демографические факторы – половой, брачный, территориальный, возрастной и т.д. состав населения.

И. С. Лунева и соавторы рассматривают основные факторы, которые влияют на рождаемость в современном обществе (Лунева и др., 2019):

- изменение социальной роли женщины;
- тенденция к малодетным семьям;
- время выбора даты рождения;
- сложное экономическое положение страны;
- рост уровня безработицы.

Иные факторы выделяет С. В. Щепотьева (Щепотьева, 2010):

- материальное положение семьи;
- возраст;

- семейное положение;
- число детей в семье;
- уровень образованности;
- тип населенного пункта;
- религия;
- занятость женщин;
- отношение к семейному образу жизни.

В. А. Карпова разделяет факторы, влияющие на рождаемость, на медико-биологические и социально-психологические. Первые определяют физическое состояние организма человека и влияют на выполнение человеком репродуктивной функции. К медико-биологическим относятся такие факторы, как состояние репродуктивного здоровья населения и возраст матери. Социально-психологические факторы определяют моральную и психологическую готовность человека к реализации репродуктивной функции. Среди них выделяются следующие: внеэкономическая потребность в детях и родительский инстинкт. Медико-биологическая группа факторов является наиболее важной, поскольку от нее зависит репродуктивное здоровье, напрямую влияющее на возможность зачатия и рождения ребенка, безопасность течения беременности, а также здоровье ребенка и матери.

Факторы, которые позитивно влияют на уровень рождаемости, рассматривает В. И. Шарин (Шарин, 2019). Это:

- пропаганда семей и семейных ценностей;
- налоговые послабления и небольшие дотации для семей с детьми;
- снижение пенсионного возраста;
- материнский социальный пакет;
- усложнение процедуры разводов;
- скидки на железнодорожные перевозки;
- получение новой профессии в период ухода за ребенком;
- запрет абортот;
- снижение минимально допустимого возраста вступления в брак;
- коммерческое суррогатное материнство;
- вспомогательные репродуктивные технологии.

Согласно указу Президента Российской Федерации 2007 года «Об утверждении Концепции демографической политики Российской Федерации на период до 2025 года»³, на уровень рождаемости влияют следующие факторы:

- низкий уровень доходов семей;
- отсутствие нормальных жилищных условий;
- рост числа неполных семей;
- ориентация на малодетность;
- тяжелый физический труд части работающих женщин;
- условия труда, которые не отвечают санитарно-гигиеническим нормам;
- низкий уровень репродуктивного здоровья;
- высокое число абортов.

³ Об утверждении Концепции демографической политики...

Подводя итог обзору различных перечней факторов, влияющих на рождаемость населения, обобщим в табл. 1 показатели, рассчитываемые органами статистики и органами власти на регулярной основе, выражающие отдельные факторы, для их дальнейшего использования в разработке прогноза.

Таблица 1 / Table 1

Статистические показатели, выражающие факторы влияния на рождаемость населения в субъекте Российской Федерации / Statistic indicators reflecting the factors influencing fertility in a Russian region

№	Показатель
1	Численность городского населения, чел.
2	Численность сельского населения, чел.
3	Число зарегистрированных браков, шт.
4	Число зарегистрированных разводов, шт.
5	Число прерываний беременности, шт.
6	Число мужчин в трудоспособном возрасте, чел.
7	Численность женщин в возрасте 15–19 лет, чел.
8	Численность женщин в возрасте 20–24 лет, чел.
9	Численность женщин в возрасте 25–29 лет, чел.
10	Численность женщин в возрасте 30–34 лет, чел.
11	Численность женщин в возрасте 35–39 лет, чел.
12	Численность женщин в возрасте 40–44 лет, чел.
13	Численность женщин в возрасте 45–49 лет, чел.
14	Численность женщин в возрасте 50–54 лет, чел.
15	Уровень безработицы в возрасте 15–72 лет, %.
16	Уровень занятости в возрасте 15–72 лет, %.
17	Среднемесячная номинальная начисленная заработная плата работников организаций, руб.
18	Среднедушевые денежные доходы населения в месяц, руб.
19	Численность населения с денежными доходами ниже величины прожиточного минимума, в % от общей численности населения субъекта Российской Федерации.

Источник: составлено авторами.

Таким образом, для построения прогноза абсолютной численности родившихся, в случае с методом инерционного прогнозирования, мы будем использовать простейшую экстраполяцию предыдущей тенденции и динамики показателя, а в случае с методом машинного обучения используем показатели из табл. 1 для построения результирующей прогнозной модели. Построим прогнозы на данных Пермского края за период с 2009 по 2019 годы. Для сопоставления точности прогнозов используем период с 2018 по 2019 годы в качестве прогнозного.

РЕЗУЛЬТАТЫ ИССЛЕДОВАНИЯ

Прогноз с использованием метода инерционного прогнозирования

Для получения прогнозных значений были рассчитаны темпы роста/снижения целевого показателя – «абсолютная численность родившихся» и поделены на количество лет, за которые проводились расчеты. Как говорилось ранее, метод инерционного прогнозирования чувствителен к резкому преломлению существующих тенденции и сдвигам в динамике изменения показателей. С 2017 по 2019 годы произошло стремительное ухудшение ситуации с рождаемостью населения в Пермском крае и темпы снижения увеличились до двузначных значений. Средняя ошибка отклонения прогнозных значений от фактических – 11,9 %.

Таблица 2 / Table 2

Фактические и прогнозные значения, полученные методом инерционного прогнозирования / Actual and predicted values obtained through inertial method

Год	Абсолютная численность родившихся, чел		Темп роста/(снижения)
	(факт)	(прогноз)	
2009	36 043	-	Базовый год
2010	37 304	-	3,49 %
2011	37 015	-	(0,77 %)
2012	38 931	-	5,17 %
2013	38 637	-	(0,75 %)
2014	38 837	-	0,51 %
2015	38 761	-	(0,19 %)
2016	37 258	-	(3,87 %)
2017	31 915	-	(14,34 %)
2018	29 526	31535	Среднеарифметический темп роста / (снижения) (1,19 %)
2019	26 617	31159	

Источник: составлено авторами.

Прогноз с использованием метода машинного обучения

В отличие от инерционного метода прогнозирования, для формирования прогнозных значений целевого показателя модели машинного обучения необходимы данные показателей, влияющих на процесс рождаемости населения в Пермском крае. Поэтому метод является более трудозатратным и более требовательным к данным.

Таблица 3 / Table 3

Статистические данные Пермского края для построения модели машинного обучения /
Statistic data for the Perm Region used to build a machine learning model

Годы/показатели	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019
Родившихся всего	36 043	37 304	37 015	38 931	38 637	38 837	38 761	37 258	31 915	29 526	26 617
Городское население	1 999 733	1 980 762	1 975 474	1 969 985	1 978 502	1 985 784	1 991 808	1 991 998	1 993 520	1 988 243	1 980 789
Сельское население	660 626	667 892	658 076	661 088	655 959	650 370	645 224	642 411	638 577	634 879	630 011
Число зарегистрированных браков	1 131	1 180	1 098	995	1 075	1 179	1 073	833	968	970	797
Число зарегистрированных разводов	878	922	775	824	966	996	753	801	870	941	861
Число прерываний беременности	32 341	28 197	26 282	23 818	22 602	20 501	17 889	16 338	15424	15286	14356
Число мужчин в трудоспособном возрасте	844 592	831 260	817 521	806 782	799 573	790 644	780 374	766 994	756 495	745 779	735 502
Численность женщин в возрасте 15–19 лет	88 043	80 253	73 548	69 629	66 115	64 864	63 527	63 357	63 513	65 074	66 026
Численность женщин в возрасте 20–24 лет	116 934	113 812	109 057	102 021	94 290	86 018	78 992	72 909	68 968	65 214	62314
Численность женщин в возрасте 25–29 лет	105 774	107 909	109 630	112 301	114 084	112 812	110 368	106 491	99 631	92 003	86241
Численность женщин в возрасте 30–34 лет	106 303	105 014	103 031	101 702	101 471	102 846	104 884	106 765	109 396	110 907	111852
Численность женщин в возрасте 35–39 лет	95 384	98 372	100 680	102 263	102 973	103 838	102 591	100 863	99 515	98 935	97258
Численность женщин в возрасте 40–44 лет	88 235	86 310	86 480	87 991	90 914	93 174	96 213	98 447	99 884	100 501	101300
Численность женщин в возрасте 45–49 лет	111 306	106 601	99 788	94 024	89 246	85 991	84 221	84 522	86 067	88 717	90 776

Годы/показатели	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019
Численность женщин в возрасте 50–54 лет	116 293	114 563	114 445	113 911	111 510	107 916	103 515	97 186	91 557	86 874	83 498
Уровень безработицы в возрасте 15–72 лет (%)	9,9	8,3	7,5	6,3	6,5	5,8	6,3	5,8	6,1	5,4	5,1
Уровень занятости в возрасте 15–72 лет (%)	61,5	62,7	63,6	63,5	62,3	60,8	61,8	63	61,5	61,2	59,5
Среднемесячная номинальная начисленная заработная плата работников организаций (руб.)	15 228	17 438	18 773	21 821	24 716	27 102	28 528	30 651	32 952	35 802	39 210
Среднедушевые денежные доходы населения (в месяц; руб.)	17 975	19 834	21 307	23 329	26 054	28 315	32 043	28 400	28 655	28 708	30 588
Численность населения с денежными доходами ниже величины прожиточного минимума (в процентах от общей численности населения субъекта Российской Федерации)	14,6	13,8	14,4	12,2	11,4	12	12,6	15,3	15,1	14,9	13,9

Источник: составлено авторами на основании данных Росстата и Пермьстата⁴.

⁴ *Офис*, сайт Территор. органа Федер. службы гос. статистики по Перм. краю (Пермьстат) [Электронный ресурс]. URL: <http://www.permstat.gks.ru> (дата обращения: 18.04.2021); *офис*, сайт Федер. службы гос. статистики (Росстат) [Электронный ресурс]. URL: http://www.gks.ru/free_doc/new_site/vvp/met-37.htm (дата обращения: 18.04.2021).

Прогнозные значения за 2018 и 2019 годы можно увидеть в табл. 4. Модель машинного обучения была настроена по умолчанию и дополнительно не модифицировалась. Лучшая итерация модели – 999, средняя ошибка отклонения прогнозных значений от фактических – 19,85 %.

Таблица 4 / Table 4

Фактические и прогнозные значения, полученные методом машинного обучения / Actual and predicted values obtained through machine learning method

Год	Абсолютная численность родившихся, чел.	
	(факт)	(прогноз)
2009	36 043	–
2010	37 304	–
2011	37 015	–
2012	38 931	–
2013	38 637	–
2014	38 837	–
2015	38 761	–
2016	37 258	–
2017	31 915	–
2018	29 526	33 188
2019	26 617	33 885

Источник: составлено авторами.

В процессе построения модели были рассчитаны значения важности показателей (feature importance), которые представлены в табл. 5.

Таблица 5 / Table 5

**Значения важности показателей модели машинного обучения /
Relevance value indicators for machine learning model**

№	Показатель	Важность показателя для модели
1	Численность городского населения, чел.	5,25
2	Численность сельского населения, чел.	6,50
3	Число зарегистрированных браков, шт.	3,79
4	Число зарегистрированных разводов, шт.	2,19
5	Число прерываний беременности, шт.	7,08
6	Число мужчин в трудоспособном возрасте, чел.	5,49
7	Численность женщин в возрасте 15–19 лет, чел.	2,08
8	Численность женщин в возрасте 20–24 лет, чел.	5,66
9	Численность женщин в возрасте 25–29 лет, чел.	6,87
10	Численность женщин в возрасте 30–34 лет, чел.	6,46
11	Численность женщин в возрасте 35–39 лет, чел.	8,77
12	Численность женщин в возрасте 40–44 лет, чел.	6,99

№	Показатель	Важность показателя для модели
13	Численность женщин в возрасте 45–49 лет, чел.	3,64
14	Численность женщин в возрасте 50–54 лет, чел.	6,97
15	Уровень безработицы в возрасте 15–72 лет, %.	2,23
16	Уровень занятости в возрасте 15–72 лет, %.	5,06
17	Среднемесячная номинальная начисленная заработная плата работников организаций, руб.	5,38
18	Среднедушевые денежные доходы населения в месяц, руб.	3,23
19	Численность населения с денежными доходами ниже величины прожиточного минимума, в % от общей численности населения субъекта.	6,16

Источник: рассчитано авторами.

Как видно из табл. 5, наиболее важными показателями для формирования прогнозного значения являются численность женщин в возрасте 35–39 лет, а также число прерываний беременности. В целом, все 19 показателей демонстрируют невысокие значения важности для модели.

Сравнение полученных прогнозных значений

Оба метода показали плохие результаты, о чем свидетельствуют высокие значения ошибки отклонения прогнозных значений от фактических (табл. 6). При этом более точные прогнозные значения были получены с помощью метода инерционного прогнозирования, даже несмотря на описанные выше сложности, связанные с резким ухудшением тенденции в сфере рождаемости и динамики изменения показателя абсолютной численности родившихся. Метод машинного обучения продемонстрировал крайне высокие значения отклонения прогнозных значений от фактических, что объясняется требовательностью данного метода к наличию большого объема данных за продолжительный промежуток времени. При этом с ростом периодов качество генерации прогнозных значений моделью будет возрастать.

Таблица 6 / Table 6

Сравнительная таблица полученных прогнозных значений /
Comparison of the obtained predicted values

Год	Абсолютная численность родившихся, чел.			Ошибка отклонения прогнозных значений от фактических, %	
	Факт	Инерционный прогноз	Прогноз на основе модели машинного обучения	Инерционный прогноз	Прогноз на основе модели машинного обучения
2018	29 526	31 535	33 188	6,80	12,40
2019	26 617	31 159	33 885	17,06	27,30

Источник: рассчитано авторами.

Наглядная разница между точностью прогнозных значений представлена на рисунке.

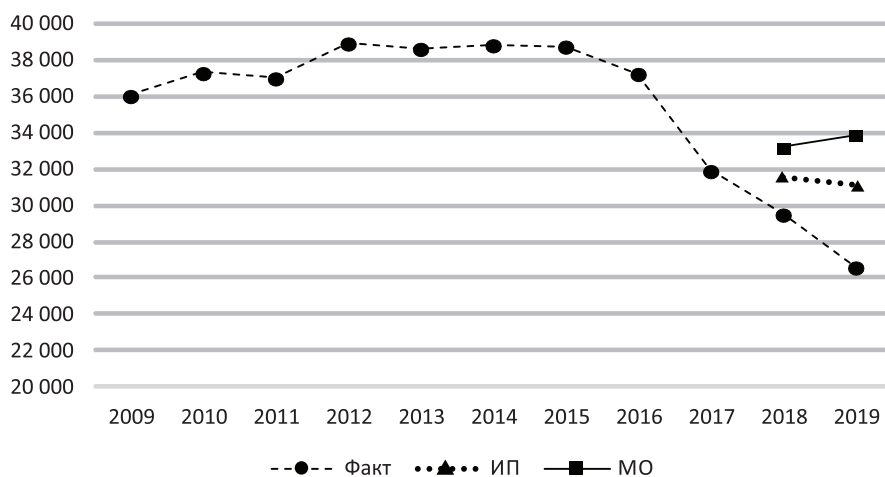


Рис. Фактические и прогнозные значения показателя «абсолютная численность родившихся», полученных с помощью двух методов на данных Пермского края за период с 2009 по 2019 годы / Fig. Actual and predicted values for “absolute number of births” obtained through two methods based on the data for 2009–2019

Источник: составлено авторами на основе собственных расчетов и данных Пермьстата⁵.

ЗАКЛЮЧЕНИЕ

Таким образом, по итогам сравнения прогнозных значений, полученных с помощью инерционного метода и метода машинного обучения (на данных Пермского края), можно сделать вывод, что оба метода показали достаточно низкую точность, поэтому нельзя однозначно рекомендовать ни один из них в качестве основного для целей прогнозирования демографических показателей в сфере рождаемости населения.

Метод инерционного прогнозирования является простым и не требовательным к объему данных, при этом он весьма чувствителен к резкому изменению прогнозируемых показателей. Поэтому его можно использовать для прогнозирования демографических показателей в сфере рождаемости на данных регионов, имеющих ярко выраженную демографическую тенденцию и демонстрирующих относительно стабильную динамику ее изменения.

Метод машинного обучения требователен к длине временных рядов и наличию статистических показателей, выражающих факторы влияния на процесс рождаемости. Поэтому его применение в ближайшие годы для указанных целей будет осложнено нехваткой данных и их незначительным временным периодом. Но, учитывая специфику метода, он позволяет потенциально полу-

⁵ Официальный сайт Территориального органа Федеральной службы государственной статистики по Пермскому краю (Пермьстат) [Электронный ресурс]. URL: <http://www.permstat.gks.ru> (дата обращения: 18.04.2021).

чать более точные данные по мере накопления статистики за продолжительный период времени.

В качестве альтернативных методов, позволяющих получать прогнозные значения демографических показателей в сфере рождаемости населения, в будущих публикациях рассмотрим факторный анализ и корреляционно-регрессионное моделирование.

Библиографический список

Андреев Е. М. Современный демографический кризис и прогнозы населения России // Мир России. 1999. Т. 8, № 4. С. 175–186.

Архангельский В. Н., Елизаров В. В. Демографические прогнозы в современной России: анализ результатов и выбор гипотез // Научные труды: Институт народнохозяйственного прогнозирования РАН. 2016. № 1. С. 524–544.

Карпова В. А. Демографические факторы, влияющие на рождаемость // Сб. материалов IX Всерос. науч.-практ. конф. молодых ученых с междунар. участием «Россия молодая» / Под ред. С. В. Костюк. Кемерово: Кузбас. гос. техн. ун-т им. Т. Ф. Горбачева, 2017. 4 с.

Курносова Е. А. Прогнозирование демографических показателей в РФ // Цифровая экономика: проблемы и перспективы развития: сб. науч. статей Межрегион. науч.-практ. конференции: в 2-х т. Т. 2 / Отв. ред. С. В. Мамонтова. Курск: Юго-Запад. гос. ун-т, 2019. С. 237–242.

Лунева И. С., Иванова О. Ю., Хардилов А. В. и др. Факторы, влияющие на рождаемость в современной России // Российский вестник акушера-гинеколога. 2019. Т. 19, № 2. С. 14–20. DOI: 10.17116/rosakush20191902114.

Малева Т. М., Синявская О. В. Социально-экономические факторы рождаемости в России: эмпирические измерения и вызовы социальной политике [Электронный ресурс] // Spero. Социальная политика: экспертиза, рекомендации, обзоры. Электрон. журн. 2006. №. 5. С. 70–98. URL: <http://www.demoscope.ru/weekly/2007/0309/analit02.php> (дата обращения: 20.04.2021).

Подольная Н. Н., Бикеева М. В., Катынь А. В. Источники данных статистики рождаемости // Подольная Н. Н., Бикеева М. В., Катынь А. В. Демография и статистика населения. Саранск: Морд. гос. ун-т, 2015. С. 21–23.

Сафиуллин Л. Р., Шакирова Л. Р., Габдрахманов Н. К. и др. Метод передвижки возрастов в прогнозировании численности населения (на примере Республики Татарстан) // Вестник Самарского государственного экономического университета. 2014. № 11. С. 135–140.

Соловьев А. И., Соловьев С. А. О подходах к анализу демографических данных с использованием машинного обучения // Современные информационные технологии и ИТ-образование. 2018. Т. 14, № 4. С. 947–959. DOI: 10.25559/SITITO.14.201804.947-959.

Тарасова О. В. К вопросу о применении математических методов изучения демографических процессов // Теория и практика общественного развития. 2015. № 24. С. 186–188.

Шарин В. И. Социально-экономические факторы роста рождаемости // Достойный труд – основа стабильного общества: материалы XI Междунар.

науч.-практ. конф. / Под ред. Р. А. Долженко, М. Н. Шавровской. Екатеринбург: Урал. гос. экон. ун-т, 2019. С. 68–71.

Шубат О. М., Киселева О. И. Методы прогнозирования численности населения: опыт критического анализа // Проблемы моделирования социальных процессов: Россия и страны АТР: материалы II Всерос. науч.-практ. конф. с междунар. участием / Отв. ред. И. Г. Кузина. Владивосток: Дальневосточ. федер. ун-т, 2016. С. 328–331.

Щепотьева С. В. Факторы дифференциации репродуктивных установок и рождаемости населения России // Научное обозрение. Серия 1: экономика и право. 2010. № 6. С. 56–60.

Юзаева Ю. Р. Прогноз численности населения Оренбургской области в разрезе городской и сельской местности методом «передвижки возрастов» // Экономика, статистика и информатика. Вестник УМО. 2014. № 5. С. 155–160.

Brink H., Richards J., Fetherolf M. Real-world machine learning. Shelter Island: Manning Publications, 2016. 264 p.

Hancock J., Khoshgoftaar T. CatBoost for Big Data: An interdisciplinary review // Journal of Big Data. 2020. Vol. 7. № 94. DOI: 10.21203/rs.3.rs-54646/v1.

Ibrahim A. A., Ridwan R. L., Muhamme M. M. et al. Comparison of the CatBoost classifier with other machine learning methods // International Journal of Advanced Computer Science and Applications. 2020. Vol. 11, № 11. P. 738–748.

Kauffman R. J., Kim K., Lee S-Y. T. et al. Combining machine-based and econometrics methods for policy analytics insights // Electronic Commerce Research and Applications. 2017. Vol. 25. P. 115–140. DOI: 10.1016/j.elerap.2017.04.004.

Köhn D., Lessmann S. Predicting online shopping behaviour from clickstream data using deep learning // Expert Systems with Applications. 2020. Vol. 150. № 113342. DOI: 10.1016/j.eswa.2020.113342.

Riiman V., Wilson A., Milewicz R. et al. Comparing artificial neural network and cohort-component models for population forecasts // Population Review. 2019. Vol. 58, № 2. P. 100–116. DOI: 10.1353/prv.2019.0008.

Sitkovskiy A. M. Changing the structure of the working-age population projected by the cohort component method (on the example of a megacity) // Актуальные вопросы гуманитарных и общественных наук: сб. науч. тр. всерос. науч.-практ. конф. с междунар. участием. Ч. 1 / Под ред. О. В. Архиповой, А. И. Климина. СПб.: ООО «Фора-принт», 2019. С. 260–269.

Smith S. K., Tayman J., Swanson D. A. Overview of the cohort-component method // Smith S. K., Tayman J., Swanson D. A. A practitioner's guide to state and local population projections. New York, London: Springer Dordrecht Heidelberg, 2013. P. 45–50.

Thatkar P., Pawar D. D., Ingale K. Y. Cohort-component method for projection of population of India // International Journal of Scientific Research. 2018. Vol. 7, № 6. P. 387–389.

Zhang H., Zeng R., Chen L. et al. Research on personal credit scoring model based on multi-source data // Journal of Physics: Conference Series. 2020. № 1437. DOI: 10.1088/1742-6596/1437/1/012053.

Информация об авторах

Зубарев Николай Юрьевич – канд. экон. наук, ст. преподаватель кафедры государственного и муниципального управления, ст. преподаватель кафедры мировой и региональной экономики, экономической теории ФГАОУ ВО «Пермский государственный национальный исследовательский университет», 614990, Россия, г. Пермь, ул. Букирева, 15

ORCID: 0000-0002-9021-4058

ResearcherID: D-6310-2017

Электронный адрес: nu_zubarev@mail.ru

Федулова Дарья Дмитриевна – бакалавр 4 курса кафедры государственного и муниципального управления ФГАОУ ВО «Пермский государственный национальный исследовательский университет», 614990, Россия, г. Пермь, ул. Букирева, 15

ORCID: 0000-0003-0509-9241

ResearcherID: –

Электронный адрес: fedylovadaria@icloud.com

Статья получена редакцией 3 мая 2021 года

Принята к печати 31 мая 2021 года

UDC [314.3:004](470.53)

DOI: 10.17072/2218-9173-2021-2-204-221

FORECASTING FERTILITY DEMOGRAPHIC INDICATORS: INERTIAL METHOD VERSUS MACHINE LEARNING METHOD

Nikolai Yu. Zubarev

Perm State University, 15 Bukirev str., Perm, 614990, Russia

ORCID: 0000-0002-9021-4058

ResearcherID: D-6310-2017

Электронный адрес: nu_zubarev@mail.ru

Daria D. Fedulova

Perm State University, 15 Bukirev str., Perm, 614990, Russia

ORCID: 0000-0003-0509-9241

ResearcherID: –

Электронный адрес: fedylovadaria@icloud.com

For citation:

Zubarev, N. Yu. and Fedulova, D. D. (2021), "Forecasting fertility demographic indicators: Inertial method versus machine learning method", *Ars Administrandi*, vol. 13, no. 2, pp. 204–221, doi: 10.17072/2218-9173-2021-2-204-221.

Abstract

Introduction: the article is devoted to the comparison of methods for predicting demographic indicators of fertility rate. **Objectives:** to determine the accuracy of methods for forecasting demographic indicators of fertility rate through comparison of the predicted values obtained using the inertial method and the machine learning method (based on data for the Perm Region). **Methods:** statistical analysis, graphical analysis, inertial forecasting method, machine learning method based on gradient boosting (CatBoost) using the *Google Collab* software environment and the *Python* programming language version 3.7. **Results:** forecast values of the “absolute number of births” indicator were obtained, with the average deviation error for the inertial forecasting method of 11.9 %, and for the machine learning method – 19.85 %. The specificity of predicted values formation for each method has been revealed and high values of deviations substantiated. **Conclusions:** the inertial forecasting method turned out to be more accurate than the machine learning method, while both methods showed low accuracy. This is due to the limitations of the dataset available and used for calculations, as well as the deterioration of the demographic situation in terms of fertility in the Perm Region and a sharp change in dynamics of the predicted indicator.

Keywords: fertility; machine learning; inertial forecast; forecasting

References

- Andreev, E. M. (1999), “Modern demographic crisis and forecasts of the population of Russia”, *Universe of Russia*, vol. 8, no. 4, pp. 175–186.
- Arkhangel'ski, V. N. and Elizarov, V. V. (2016), “Demographic forecasts in modern Russia: Analysis of results and choice of hypotheses”, *Nauchnye trudy: The Institute of Economic Forecasting of the Russian Academy of Sciences*, no. 1, pp. 524–544.
- Karpova, V. A. (2017), “Demographic factors affecting the birth rate”, in Kostyuk, S. V. (ed.), *Sbornik materialov IX vserossiiskoi nauchno-prakticheskoi konferentsii molodykh uchenykh s mezhdunarodnym uchastiem “Rossiya molodaya”* [Problems of modeling social processes: Russia and the countries of the Asia-Pacific region: Materials of the II All-Russian scientific and practical conference with international participation], T.F. Gorbachev Kuzbass State Technical University, Kemerovo, Russia, 4 p.
- Kurnosova, E. A. (2019), “Forecasting demographic indicators in the Russian Federation”, in Mamontova, S. V. (ed.), *Tsifrovaya ekonomika: problemy i perspektivy razvitiya: sbornik nauchnykh statei Mezhhregional'noi nauchno-prakticheskoi konferentsii: v 2-kh tomakh* [Digital economy: Problems and development prospects: Proceedings of the Interregional scientific and practical conference: in 2 vol. Vol. 2], South-West State University, Kursk, Russia, pp. 237–242.
- Luneva, I. S., Ivanova, O. Yu., Khardikov, A. V. et al. (2019), “Factors influencing the birth rates in modern Russia”, *Russian bulletin of obstetrician-gynecologist*, vol. 19, no. 2, pp. 14–20, doi: 10.17116/rosakush20191902114.
- Maleva, T. M. and Sinyavskaya, O. V. (2006), “Social and economic factors of fertility in Russia: Empirical dimensions and challenges to social policy”, *Spero. Social*

Policy: Expertise, Recommendations, Overviews, no. 5, pp. 70–98 [Online], available at: <http://www.demoscope.ru/weekly/2007/0309/analit02.php> (Accessed April 20, 2021).

Podolnaya, N. N., Bikeeva, M. V. and Katyn, A. V. (2015), “Sources of fertility statistics”, in Podolnaya, N. N., Bikeeva, M. V. and Katyn, A. V., *Demografiya i statistika naseleniya* [Demography and population statistics], Mordovian State University, Saransk, Russia, pp. 21–23.

Safiullin, L. R., Shakirova, L. R., Gabdrakhmanov, N. K. et al. (2014), “The method of moving ages in predicting the population size (on the example of the Republic of Tatarstan)”, *Vestnik of Samara State University of Economics*, no. 11, pp. 135–140.

Solovev, A. I. and Solovev, S. A. (2018), “On approaches to the analysis of demographic data using machine learning”, *Modern Information Technologies and IT Education*, vol. 14, no. 4, pp. 947–959, doi: 10.25559/sitito.14.201804.947-959.

Tarasova, O. V. (2015), “On the question of the use of mathematical methods for studying demographic processes”, *Theory and practice of social development*, no. 24, pp. 186–188.

Sharin, V. I. (2019), “Social and economic factors of fertility growth”, in Dolzhenko, R. A. and Shavrovskaya, M. N. (eds.), *Dostoinyi trud – osnova stabil'nogo obshchestva: materialy XI Mezhdunarodnoi nauchno-prakticheskoi konferentsii* [Decent work is the basis of a stable society: Materials of the XI International scientific and practical conference], Ural State University of Economics, Yekaterinburg, Russia, pp. 68–71.

Shubat, O. M. and Kiseleva, O. I. (2016), “Methods of forecasting the population size: The experience of critical analysis”, in Kuzina, S. G. (ed.), *Problemy modelirovaniya sotsial'nykh protsessov: Rossiya i strany ATR: Materialy II Vserossiiskoi nauchno-prakticheskoi konferentsii s mezhdunarodnym uchastiem* [Problems of modeling social processes: Russia and the countries of the Asia-Pacific region: Materials of the II All-Russian scientific and practical conference with international participation], Far Eastern Federal University, Vladivostok, Russia, pp. 328–331.

Shchepot'eva, S. V. (2010), “Factors of differentiation of reproductive attitudes and fertility of the population of Russia”, *Scientific review, Series 1: Economics and Law*, no. 6, pp. 56–60.

Yuzaeva, J. R. (2014), “The forecast of the population of the Orenburg region in the context of urban and rural areas by the method of “advancing ages”, *Economy, Statistics and Informatics. Bulletin of UMO*, no. 5, pp. 155–160.

Brink, H., Richards, J. and Fetherolf, M. (2017), *Real-world machine learning*, Manning Publications, Shelter Island, NY, US.

Hancock, J. and Khoshgoftaar, T. (2020), “CatBoost for Big Data: An interdisciplinary review”, *Journal of Big Data*, vol. 7, no. 94, doi: 10.21203/rs.3.rs-54646/v1.

Ibrahim, A. A., Ridwan, R. L., Muhamme, M. M. et al. (2020), “Comparison of the CatBoost classifier with other machine learning methods”, *International Journal of Advanced Computer Science and Applications*, vol. 11, no. 11, pp. 738–748.

Kauffman, R. J., Kim, K., Lee, S-Y. T. et al. (2017), “Combining machine-based and econometrics methods for policy analytics insights”, *Electronic Commerce Research and Applications*, vol. 25, pp. 115–140, doi: 10.1016/j.elerap.2017.04.004.

Köhn, D. and Lessmann, S. (2020), "Predicting online shopping behaviour from clickstream data using deep learning", *Expert Systems with Applications*, vol. 150, no. 113342, doi: 10.1016/j.eswa.2020.113342.

Riiman, V., Wilson, A., Milewicz, R. et al. (2019), "Comparing artificial neural network and cohort-component models for population forecasts", *Population Review*, vol. 58, no. 2, pp. 100–116, doi: 10.1353/prv.2019.0008.

Sitkovskiy, A. M. (2019), "Changing the structure of the working-age population projected by the cohort component method (on the example of a megacity)", in Arkhipova, O. V. and Klimin, A. I. (eds.), *Aktual'nye voprosy gumanitarnykh i obshchestvennykh nauk: sbornik nauchnykh trudov vserossiiskoi nauchno-prakticheskoi konferentsii s mezhdunarodnym uchastiem. Chast' 1* [Topical issues of the humanities and social Sciences: Proceedings of the All-Russian scientific and practical conference with international participation. Part 1], Ltd "Fora-print", St. Petersburg, Russia, pp. 260–269.

Smith, S. K., Tayman, J. and Swanson, D. A. (2013), "Overview of the cohort-component method", in Smith, S. K., Tayman, J. and Swanson, D. A., *A practitioner's guide to state and local population projections*, Springer Dordrecht Heidelberg, NY, NYC, US, London, UK, pp. 45–50.

Thatkar, P., Pawar, D. D. and Ingale, K. Y. (2018), "Cohort-component method for projection of population of India", *International Journal of Scientific Research*, vol. 7, no. 6, pp. 387–389.

Zhang, H., Zeng, R., Chen, L. et al. (2020), "Research on personal credit scoring model based on multi-source data", *Journal of Physics: Conference Series*, no. 1437, doi: 10.1088/1742-6596/1437/1/012053.

Received May 3, 2021

Accepted May 31, 2021