

Анализ демографии, а также демографического кризиса Российской Федерации с помощью методов анализа данных и машинного обучения

А. М. Тен, email: andreiten1@gmail.com
В. В. Мокшин, email: vladimir_kgtu@mail.ru

Казанский национальный исследовательский технический университет
им. А. Н. Туполева - КАИ

***Аннотация.** В данной работе рассматриваются методы анализа данных и машинного обучения в области исследования демографии населения Российской Федерации*

***Ключевые слова:** Машинное обучение, анализ данных, искусственный интеллект, Python, Классификация, демография*

Введение

В настоящее время в России отмечается особая проблема демографии населения. Идет полное вымирание коренного населения Российской Федерации, что приведет к трагедии. В будущем будет наблюдаться острая проблема нехватки населения. Для сохранения численности населения на одном уровне нужен суммарный коэффициент рождаемости около 2,1 рождений на женщину в течение жизни, но рождаемость в стране не достигает этого уровня. Наиболее резкое падение коэффициента рождаемости (до 1,157) произошло после дефолта 1998 года, затем рождаемость стала медленно расти, а с 2007 года в связи с вводом материнского капитала более быстрыми темпами — если в 2006 году на 1000 человек приходилось 10,3 новорождённых, то в 2014 году уже 13,3, а с 2015 года начала снижаться, в 2020 году на 1000 человек приходилось уже 9,8 новорождённых. Число новорождённых уменьшилось с 1 479 637 в 2006 году до 1 436 541 в 2020 году, а это спад на 2,9 %. Суммарный коэффициент рождаемости в 1999—2020 годах вырос с 1,157 до 1,505. Многие аналитики связывают причины демографического кризиса со снижением доходов населения. В остальном, аналогичные демографические тенденции обычны для индустриальных и постиндустриальных обществ, как в развитых, так и в развивающихся странах, где уже завершился демографический переход.

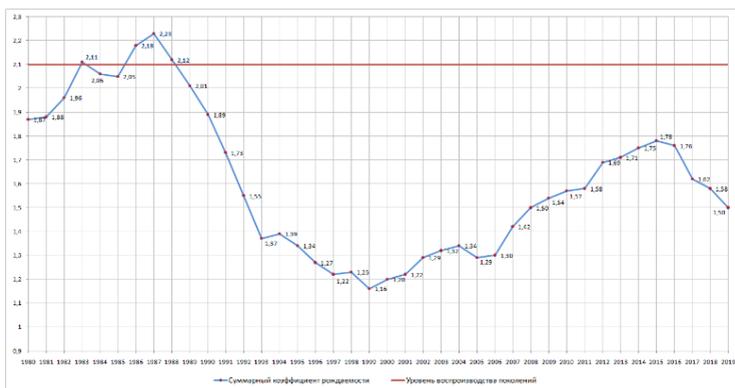


Рис. 1. Суммарный коэффициент рождаемости в России в 1980 – 2020 годах

2. Анализ данных на примере датасета за 1990 – 2020 годы

В качестве примера, возьмем статистику официальных данных за период 1990-2020 года. Набор данных содержит следующие поля: год, регион/область, естественный прирост населения на 1000 человек, кол-во рождения на 1000 человек, кол-во смертей на 1000 человек, прирост мигрирующего населения на 1000 человек.

Таблица

Основные характеристики

На 1000 человек	Естественный прирост (x1)	Кол-во рождения (x2)	Кол-во смертей (x3)	Прирост мигрирующего населения (x4)	Население (x5)
Общее кол-во	2317	2317	2315	2330	2330
Среднее	-2.344022	11.549245	13.901382	68.474013	69.148704
Стандартное отклонение	6.064389	3.233035	3.499475	10.453587	13.349880
Минимальное	-15.9	6.2	3.2	37.5	23.63
Максимальное	24.8	29.9	24.9	97.53	100



Рис. 2. График среднегодовой частоты рождений смертей

Из данных видно что рождаемость впервые превысила смертность только в 2012 году.

В регионах с высоким процентом городского населения общий демографический вес меньше, чем в регионах с небольшим показателем. В больших городах семьи имеют меньше детей, чем в городах и деревнях. Также мы видим, что в первые 10 лет нового века общий демографический вес был значительно ниже, чем в другое время. Это может быть связано с низкой рождаемостью и высокой смертностью в это время.

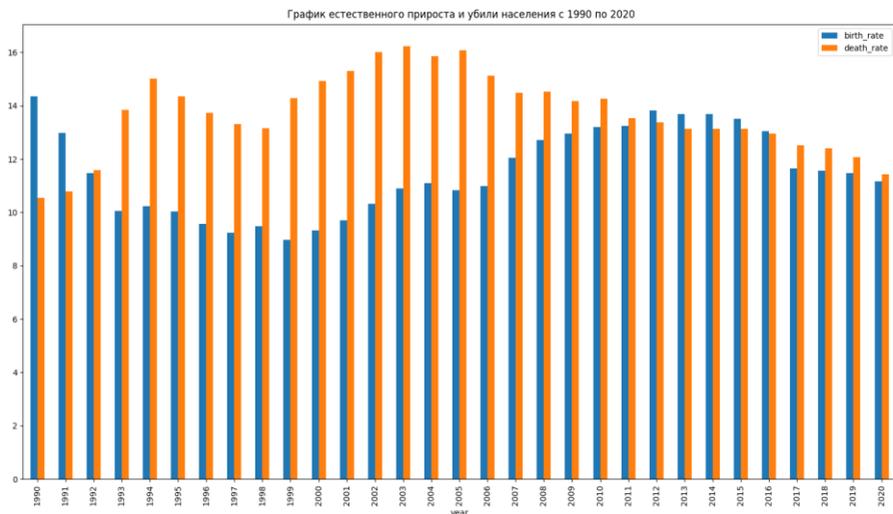


Рис. 3. График естественного прироста и убыли населения

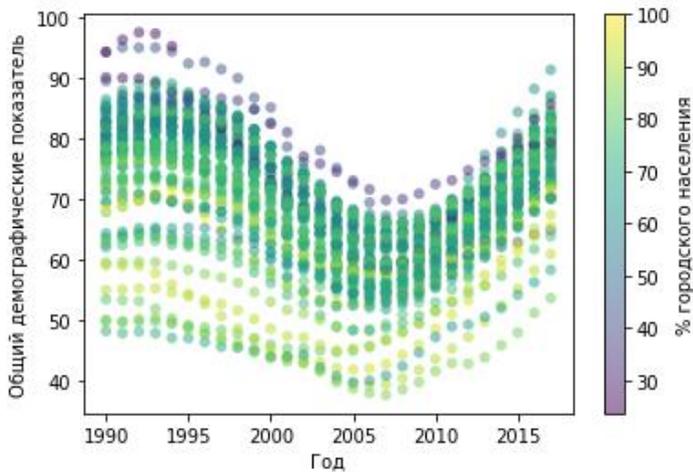


Рис. 4. Демографические показатели среди городского населения

В регионах с высоким процентом городского населения общий демографический вес меньше, чем в регионах с небольшим показателем. В больших городах семьи имеют меньше детей, чем в городах и деревнях. Также мы видим, что в первые 10 лет нового века общий демографический вес был значительно ниже, чем в другое время. Это может быть связано с низкой рождаемостью и высокой смертностью в это время.

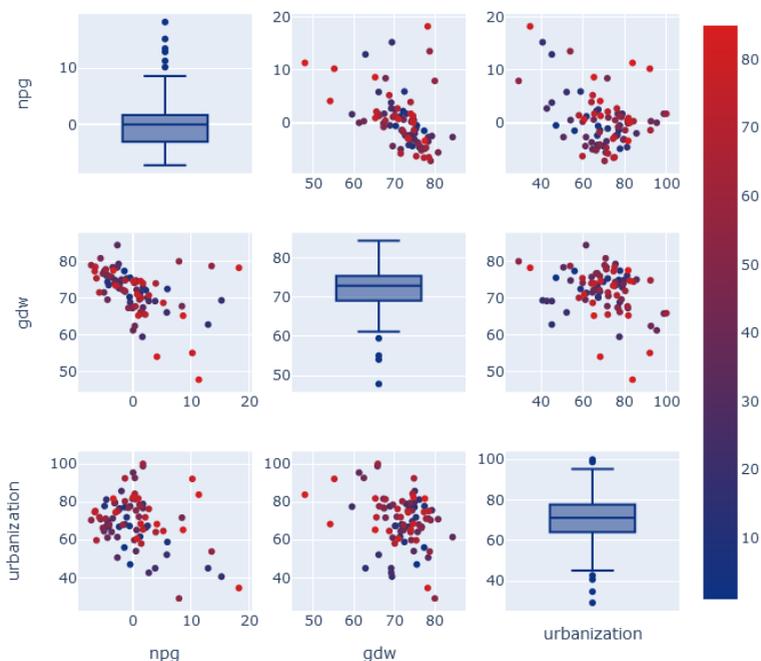


Рис. 5. Диаграмма рассеяния

Можно также получить список республик с самой высокой рождаемостью

	birth_rate	death_rate
region		
Chechen Republic	24.642857	5.480952
Tuva Republic	21.722581	11.677419
Republic of Ingushetia	19.023077	4.407692
Republic of Dagestan	18.219355	6.025806
Altai Republic	17.241290	12.300000
Sakha (Yakutia) Republic	15.606452	9.106452
Nenets Autonomous Okrug	15.158065	10.800000
Republic of Buryatia	14.525806	12.329032
Yamalo-Nenets Autonomous Okrug	14.158065	5.370968
Zabaykalsky Krai	14.032258	13.354839

Рис. 6. Список республик с самой высокой рождаемостью

3. Корреляционный анализ

Корреляционный анализ, совокупность основанных на математической теории корреляции методов обнаружения корреляционной зависимости между двумя случайными признаками или факторами.

Данный метод обработки статистических данных весьма популярен в экономике, астрофизике и социальных науках (в частности в психологии и социологии), хотя сфера применения коэффициентов корреляции обширна: контроль качества промышленной продукции, металловедение, агрохимия, гидробиология, биометрия и прочие. В различных прикладных отраслях приняты разные границы интервалов для оценки тесноты и значимости связи.

Популярность метода обусловлена двумя моментами: коэффициенты корреляции относительно просты в подсчете, их применение не требует специальной математической подготовки. В сочетании с простотой интерпретации, простота применения коэффициента привела к его широкому распространению в сфере анализа статистических данных.

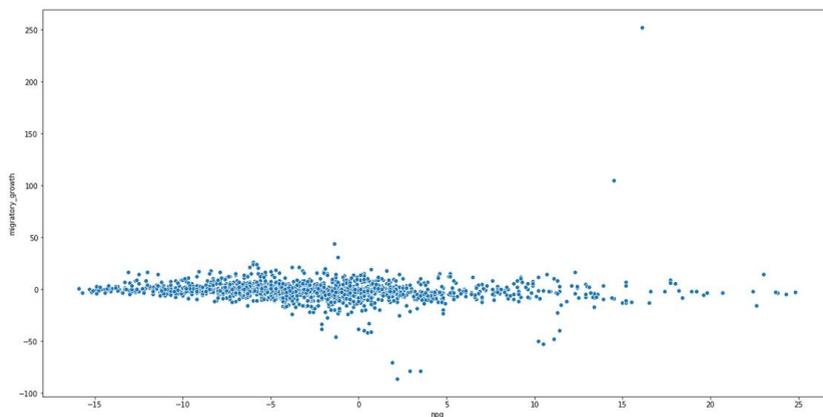


Рис. 7. Корреляция смертности и естественного прироста населения

Заключение

В данной статье была рассмотрена проблема демографии России, а также проведен основной анализ на наборе данных с 1990 по 2020 год населения России. С помощью методов машинного обучения можно построить и спрогнозировать модели, на которые можно будет ориентироваться в будущем, для устранения проблемы.

Список литературы

2. Барсегян, А. А. Анализ данных и процессов : учеб. пособие / А. А. Барсегян, М. С. Куприянов, С. И. Елизаров. – 3-е изд., перераб и доп. – СПб. : БХВ-Петербург, 2009. – 512 с.
3. Data Mining: Информация [Электронный ресурс]. – Режим доступа : <https://www.intuit.ru/studies/courses/6/6/lecture/198?page=3/>
4. Метод формирования модели анализа сложной системы Мокшин В.В., Якимов И.М. Информационные технологии. 2011. № 5. с. 46-51.