

ОПРЕДЕЛЕНИЕ НАБОРА НАИБОЛЕЕ ЗНАЧИМЫХ ФАКТОРОВ, ВЛИЯЮЩИХ НА ЭПИЗОДЫ ТЕРРОРИСТИЧЕСКОЙ АКТИВНОСТИ И ДРУГИЕ АСПЕКТЫ ПОЛИТИЧЕСКОЙ НЕСТАБИЛЬНОСТИ, С ПОМОЩЬЮ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ*

Сергей Георгиевич Шульгин

Российская академия народного хозяйства и государственной службы при Президенте РФ

Максим Слав

Национальный исследовательский университет «Высшая школа экономики»

В работе предлагается метод для отбора факторов, наиболее важных для анализа и прогнозирования эпизодов террористической активности и других аспектов политической нестабильности, связанных с возможными человеческими жертвами. В качестве источников данных по нестабильности используются данные The Cross National Time Series (CNTS), Global Terrorism Database (GTD) и Coups d'Etat, 1946–2018 (Center for Systemic Peace). В качестве источников данных для независимых переменных используются базы данных Всемирного банка, ООН, а также Polity IV, Maddison Database, Worldwide Governance Indicators и др. После тщательного анализа из общей базы данных, включающей более 600 переменных, были отобраны 58 факторов, для которых в литературе была показана теоретическая значимость для анализа террористической активности. Для оценки наиболее важных факторов мы используем метод градиентного бустинга, с помощью которого мы ранжируем переменные по их важности для анализа и предсказания эпизодов террористической активности и других аспектов политической нестабильности, связанных с возможными человеческими жертвами.

* Исследование выполнено при поддержке Российского научного фонда (проект № 18-18-00254).

Системный мониторинг глобальных и региональных рисков 2021 190–218

1. Введение

Среди факторов, которые могут оказывать влияние на социально-политические процессы дестабилизации, можно выделить несколько групп:

- Макроэкономические характеристики страны, среди которых можно выделить такие показатели, как ВВП, ВВП на душу населения и их темпы роста, дефицит бюджета, показатели неравенства, инфляция, уровень сбережений и инвестиционной активности, показатели внешней торговли и множество других.

- Социальные характеристики общества, такие как уровень образования, характеристики национального состава, характеристики, описывающие распространение и роль различных религий, уровень доверия в обществе, приверженность тем или иным ценностям, роль и активность отдельных классов, уровень распространения технологий, характеристики, описывающие распространенность различных языков и многие другие.

- Характеристики, описывающие модель социального и политического устройства, такие как тип политического режима, тип правовой системы, характеристики формальных и неформальных институтов, характеристики системы государственного управления.

- Демографические характеристики – факторы, связанные с численностью населения и ее структурным составом, долей в общей численности отдельных групп населения (например, молодежи), интенсивностью миграционных процессов (как внутренних, так и внешних), показателями рождаемости и смертности.

- Географические и исторические характеристики, описывающие географическое положение, расположение стран относительно друг друга, наличие в прошлом периода, проведенного совместно с другими современными странами в составе некоторого ранее единого государства.

Любая классификация не будет полной или достаточно строгой – существуют факторы, которые будут относиться сразу ко многим типам характеристик и влиять через разные механизмы: например, распространение информационных технологий одновременно влияет на макроэкономические показатели, на социально-экономические характеристики общества и т. п.

В зависимости от того, как авторы подходят к анализу процессов социально-политической нестабильности, они подчеркивают важность тех или иных факторов. Обзор некоторых подходов к анализу нестабильности можно найти, например, в работах А. В. Коротаева с коллегами (Коротаев и др. 2020; Слинько и др. 2018); можно также привести достаточно многочисленные примеры работ, анализирующих отдельные факторы нестабильности (Esty *et al.* 1998; Цирель 2012; 2015; Коротаев, Зинькина 2012; Малков и др. 2012; Przeworski *et al.* 2000 и др.).

Другим подходом к анализу социально-политической нестабильности является применение методов машинного обучения и анализ больших данных. Это подход, при котором для анализа процессов нестабильности используются данные, собранные на микроуровне (то есть на уровне отдельных событий, высокой географической детализации, поведения отдельных людей и т. п.), и для работы с которыми используется класс моделей, условно называемых методами машинного обучения (Donnay 2017; Connelly *et al.* 2016; Donnay *et al.* 2016; Соррочк *et al.* 2016) и ряд других.

В данной работе анализируются данные, собранные на страновом уровне, что позволяет работать с широким набором независимых факторов, собранных из разных источников.

2. Методология

После тщательного анализа из страновой панельной базы данных социально-политической нестабильности, включающей более 600 переменных, были отобраны 58 переменных, для которых в литературе была показана теоретическая значимость для анализа террористической активности. В целом эти факторы можно разделить на следующие группы.

1. Экономические факторы.

К этим факторам относится ВВП на душу населения; инфляция (как индекс потребительских цен); общий уровень безработицы по данным национальных агентств и Международной организации труда; доля безработной молодежи в общей численности молодежи 20–29 лет; доля безработной молодежи во взрослом населении; коэффициент Джини; уровень глобализации, измеряемый через индекс глобализации (Gygli *et al.* 2019), а также через долю импорта или экспорта в ВВП страны; наконец, разнообразные индикаторы

несостоятельности государства (state failure): низкий ИЧР, низкая совокупная протяженность дорожной сети, высокая доля голодающего населения, низкая доля населения, имеющего доступ к канализации, высокая младенческая смертность.

Обычно экономические показатели линейно и интуитивно понятно связаны с числом терактов: чем хуже общий уровень жизни (выше неравенство, инфляция и безработица и т. д.), тем больше число терактов (Mousseau 2011; Caruso, Gavrilova 2012; Wilson, Piazza 2013; Coggins 2015). Однако известно, что ВВП на душу населения криволинейно влияет на число терактов, сначала увеличивая его, а потом уменьшая (Enders, Hoover, Sandler 2016). Влияние уровня глобализации на число терактов неоднозначно, нередко этот показатель демонстрирует отрицательную связь с числом терактов (Caruso, Schneider 2011; Freytag *et al.* 2011; Kis-Katos, Liebert, Schulze 2011; Brockhoff *et al.* 2015; Gaibullov *et al.* 2017), но иногда эта связь оказывается положительной (Bove, Böhmelt 2016).

Для экономических факторов использовались в основном данные, предоставленные Всемирным Банком (World Bank 2021); показатель ВВП на душу населения рассчитывался дважды, на основе данных не только из базы данных Всемирного банка, но и из базы данных проекта Мэддисона (Bolt *et al.* 2018).

2. Социально-демографические факторы.

К социальным факторам относятся этническая, религиозная и языковая фракционализации; этническая дискриминация; гендерное неравенство; миграция, в том числе общая численность мигрантов в стране и их доля в населении страны; уровень образования, измеряемый и как грамотность, и как охват первичным, вторичным и третичным образованием (начальная школа, средняя школа, университетское образование), и как среднее число лет обучения.

К демографическим факторам относятся ожидаемая продолжительность жизни; «молодежный бугор» – доля молодежи в численности взрослого населения; общая численность населения; плотность населения; уровень урбанизации; скорость урбанизации. Молодежь измерялась и как люди в возрастах 15–29 лет, и как люди в возрастах 20–29 лет, и как люди в возрастах 15–24 лет.

Образование, урбанизация, «молодежный бугор» и старение населения отражают уровень модернизации страны. Как правило,

они сопряжены с увеличением нестабильности на первых этапах модернизации и снижением нестабильности на последних этапах (Mousseau 2011; Østby *et al.* 2019; Weber 2019; Slav *et al.* 2021). По мнению некоторых, высокая доля мигрантов, наличие дискриминации и фракционализации повышают вероятность терактов (Freitag *et al.* 2011; Lee 2011; Tausch *et al.* 2011; Findley, Young 2012; Victoroff *et al.* 2012; Lafree, Bersani 2014; Gaibulloev *et al.* 2017; Younas, Sandler 2017). Наконец, такие переменные, как численность населения и плотность населения способствуют терроризму из того простого принципа, что чем больше страна, тем легче в ней затеряться.

Данные по фракционализации были получены из работы А. Алесины и соавторов (Alesina *et al.* 2003), данные по этнической дискриминации – из базы данных *Ethnic Power Relations* (Cederman *et al.* 2005), данные по гендерному неравенству и среднее число лет обучения – из базы данных Программы развития ООН (UNDP 2019), индикаторы миграции – из базы данных Всемирного Банка (World Bank 2021), данные по грамотности – из базы данных CNTS (Banks, Wilson 2020), а охват образованием различных уровней – из базы данных V-Dem (Coppedge *et al.* 2018). Демографические данные в целом брались из баз данных ООН, но данные по общей численности населения брались и из базы данных Отдела народонаселения ООН, и из базы данных ЮНЕСКО, и из базы данных CNTS (United Nations Population Division 2018, 2020; Banks, Wilson 2020; UNESCO Institute for Statistics 2021).

3. Политические факторы.

В эту группу факторов мы включаем внутривнутриполитические факторы, внешнеполитические факторы и факторы, связанные с наличием дестабилизации в прошлом. К первым относятся фракционализация элит (The Fund for Peace 2021), число лет со времени последней смены режима, верховенство права, а также уровень коррупции. Ко вторым относятся общая площадь территории страны, финансирование со стороны США, членство в ГАТТ/ВТО и наличие «холодной войны». К третьим относятся экономические забастовки, массовые беспорядки и антиправительственные демонстрации по базе данных CNTS (Banks, Wilson 2020).

При фракционализации элит, элиты чаще используют теракты как инструмент политической борьбы (Foster *et al.* 2013). Долго-

вечность политического режима связывается с устойчивостью институтов и способностью перевести политическую борьбу в легальное русло (Caruso, Schneider 2011; Findley, Young 2012; Wilson, Piazza 2013; Gaibulloev *et al.* 2017). Уровень коррупции имеет смешанное влияние на число терактов и пока не показывает постоянных результатов; верховенство права же отрицательно влияет на число терактов во всех исследованиях (Coggins 2015; George 2018; Paul, Vagchi 2018).

Членство в международных организациях и отношения с США имеют тот же эффект, что и индикаторы глобализации из группы экономических характеристик; размер страны имеет тот же эффект, что и общая численность населения – в больших странах затруднен контроль над населением (Gassebner, Luechinger, 2011). Холодная война имеет неоднозначный эффект на число терактов: разные авторы предлагают противоположные результаты по этой переменной (Young, Dugan 2011; Aksoy, Carter, Wright 2012; Choi, Luo 2013; Wilson, Piazza 2013; Bove, Böhmelt 2016).

Уровень коррупции получен из базы данных Transparency International (Transparency International 2020), уровень верховенства права – из базы данных Всемирного Банка (World Bank 2021). Территория страны была взята из базы данных CNTS (Banks, Wilson 2020), а финансирование со стороны США – с сайта Государственного Департамента США (U.S. Agency for International Development (USAID) and U.S. Department of State on behalf of United States Government agencies reporting foreign assistance 2020).

Мы выделяем в отдельную группу зависимых переменных набор факторов, которые отражают те или иные формы социально-политической нестабильности. Для поиска наиболее важных факторов мы тренируем (оцениваем) множество моделей, в каждой из которых лишь одна из зависимых переменных используется в качестве целевой объясняемой переменной. При этом ни одна из зависимых переменных никогда не включается в число факторов в оцениваемые модели.

2.1. Модель

Для заданного набора данных D определены n точек данных, при этом каждая точка данных – это набор из объясняемой (зависимой) переменной y_i и множества из m независимых факторов X_i :

$$D = \{(y_i, X_i)\} \quad (|D| = n, X_i \in \mathbb{R}^m, y_i \in \mathbb{R}) \quad (1)$$

где \mathbb{R} – стандартное обозначение для множества действительных чисел.

В такой формулировке наша задача состоит в том, чтобы среди всего множества независимых факторов X выделить такое его подмножество, то есть отдельные его факторы, которые оказываются наиболее важными для объяснения y .

В данной работе мы используем метод, при котором мы пытаемся найти оценку зависимой переменной y_i в форме K аддитивных функций

$$\hat{y}_i = \sum_{k=1}^K f_k(X_i) \quad (2)$$

$f_k(X_i)$ – функция, которая принадлежит к подмножеству классификационных и регрессионных деревьев (*CART – Classification and Regression Tree*).

Класс функций, которые определяются как:

$$CART = \{f(X) = w_{q(X)}\} \quad (q: \mathbb{R}^m \rightarrow T; w \in \mathbb{R}^T) \quad (3)$$

где $q(X)$ описывает дерево, вершинами которого являются правила относительно значений X . Функция $q(X)$ ставит в соответствие определенной точке данных X_i , определенный лист (конечную вершину) (T). Листья в *CART* описывают результат классификации, которым присвоены веса w . Аппроксимирующая функция $f_k(X)$ определяется структурой дерева $q(X)$ и весами листьев w .

Процесс обучения (тренировки) модели сводится к минимизации функционала L , в которой суммируется ошибка между оцененными (\hat{y}_i) и реальными (y_i) значениями зависимой переменной,

а также учитывается сложность (размерность) *CART* функции. Вторая часть функционала L – это элемент так называемой регуляризации, подход, с помощью которого мы контролируем сложность *CART*-функции и пытаемся найти самую простую структуру из возможных *CART*-функций.

Для минимизации функционала L используется последовательный (итеративный) процесс, где на каждой итерации оценивается градиент в направлении минимизации L . (Подробнее описание функционала и алгоритма оптимизации см.: Chen, Guestrin 2016.)

Использование моделей градиентного бустинга (GBM) не требует нормализации данных для корректной работы и хорошо работает без предварительной обработки входных данных. В работе мы использовали GBM для всех оценок, которые производили с помощью библиотеки XGBoost (Chen, Guestrin 2016).

Данный метод успешно применяется для широкого класса задач, связанного с отбором наиболее важных переменных в задачах с высокой размерностью – например, в отборе оптимальных характеристик соискателей для предсказания для них наиболее интересных и релевантных вакансий (Volkovs, Wei Yu, Poutanen 2017), или предсказании того, какие наиболее значимые аффилиации авторов влияют на факт, что их статьи принимаются на основные авторитетные научные конференции в области машинного обучения, больших данных и т. п. (Sandulescu, Chiru 2016), или анализе физических данных CERN, полученных на Большом Адронном Коллайдере, в попытках найти факторы, влияющие на вероятность наблюдения редкого физического явления распада тау-лептона на три мюона ($\tau \rightarrow 3\mu$) (Mironov, Guschin 2015) и во многих других приложениях.

2.2. Данные

В качестве исходных данных по эпизодам **террористической активности и другим аспектам политической нестабильности, связанным с возможными человеческими жертвами**, используются данные Global Terrorism Database (GTD), несколько переменных из базы данных Cross National Time Series (CNTS), а также переменная «индекс кровавой дестабилизации», построенная с помощью методов факторного анализа на переменных CNTS.

База данных The Cross National Time Series (CNTS) – это результат работы по сбору и систематизации данных, начатой Артуром Бэнксом (Banks, Wilson 2018) в 1968 г. в Университете штата Нью-Йорк в Бингемтоне. Он обобщил архив данных ежегодников *The Statesman's Yearbook*, публикуемых с 1864 г. В базе содержатся данные по более чем 200 странам. База данных содержит годовые

значения переменных, начиная с 1815 г. В базе данных исключены периоды двух мировых войн 1914–1918 и 1940–1945 гг.

В данной работе мы используем в качестве зависимых переменных показатели, описывающие различные аспекты внутренних конфликтов (domestic). Эти данные получены из анализа страновых событий по различным подкатегориям:

- Политические убийства (Assassinations, domestic1)
- Партизанские действия (Guerrilla Warfare, domestic3)
- Политические репрессии (Purges, domestic5)

К Политическим убийствам (Assassinations, domestic1) относятся любые политически мотивированные убийства или покушения на убийства высших правительственных чиновников или политиков.

К Партизанским действиям (Guerrilla Warfare, domestic3) относится любая вооруженная деятельность, диверсии или взрывы, совершаемые независимыми группами граждан или нерегулярными вооруженными силами, направленные на свержение действующего режима.

К Политическим репрессиям (Purges, domestic5) относятся любые систематические устранения политической оппозиции (лишения свободы или убийства) среди действующих членов режима или политической оппозиции.

Помимо данных CNTS, в качестве объясняемой переменной мы используем два индикатора из Глобальной базы данных по терроризму (START 2019). Мы используем переменные:

n_terror_attack – количество террористических атак,

Nkill – количество убитых.

База содержит данные с 1970 (в анализируемой версии по 2015 г. включительно).

Всего в качестве зависимых (объясняемых, целевых) переменных для данного анализа было отобрано 6 переменных. Все зависимые переменные мы представили в форме бинарного классификатора, с помощью которого моделировалось наличие или отсутствие в данном году в данной стране нестабильности по анализируемой переменной. Точки данных, в которых значение переменной было больше 0, были классифицированы как факт нестабильности.

В Табл. 1 приведена статистика по 6 зависимым переменным и число случаев наличия и отсутствия нестабильности:

Табл. 1. Статистика по числу случаев нестабильности (ее отсутствия)

Переменная	Общее число N	Число случаев нестабильности	Число случаев отсутствия нестабильности	Пропущенные данные
n_terror_attack	9784	3830	5954	9611
nkill	9784	2650	7134	9611
cnts_domestic1	12712	1088	11624	6683
cnts_domestic3	12712	1729	10983	6683
cnts_domestic5	12712	1217	11495	6683
GK1_1992_world	5100	481	4619	14295

Источник данных: расчеты авторов.

2.3. Выбор параметров модели и тренировка моделей

Для оценки (тренировки) модели градиентного бустинга необходимо выбрать набор параметров, определяющих работу алгоритма. Одной из главных проблем, которые необходимо решить при оценке, является проблема переобучения модели (*over-fitting*). Переобучение выражается в том, что при большом количестве данных и степеней свободы модель может очень точно описать существующие закономерности на обучающей выборке (*training set*), однако полученные закономерности могут оказаться неприменимы за пределами обучающей выборки.

Для решения этой проблемы мы используем подход кросс-валидации (*cross-validation*), когда из имеющихся данных выделяется обучающая (*train*) и тестовая (*test*) выборки. Обучающая выборка используется для тренировки моделей. Тестовая выборка используется только для анализа полученных результатов (и не участвует в процессе обучения). Процесс обучения модели итеративный, и на каждой итерации мы анализируем качество полученной оценки на тестовой выборке и принимаем решение об остановке дальнейшего обучения модели в случае, когда за определенное число последних итераций не произошло улучшение результатов оценки в тестовой выборке.

Тестовую и обучающую выборки мы формируем с помощью процедуры *k-fold* кросс-валидации, когда вся выборка разбивается на *k* случайных частей (Kuhn 2008) и одна из этих частей использу-

ется в качестве тестовой, а остальные $k-1$ частей – в качестве обучающей выборки. Процедура оценки модели с этим разбиением повторяется k раз, так чтобы каждая из k частей побывала тестовой выборкой.

В данной работе мы выбрали k равным 5 и для каждой зависимой переменной провели оценку модели с помощью процедуры кросс-валидации 100 раз (каждый раз с новым разбиением на 5 случайных подвыборок). В результате для каждой из 6 зависимых переменных мы получили 100 оценок моделей, в каждой из которых оценивается значимость независимых факторов.

Складывая для каждого независимого фактора оценки его важности во всех 100 оценках модели, мы получаем результирующую агрегированную оценку значимости каждого независимого фактора для анализируемой зависимой переменной.

Параметр глубины деревьев (*max.depth*) эмпирическим путем выбран равным 5. Уменьшение упрощает структуру (желаемое свойство), однако качество оценки модели падает. Увеличение глубины в целом улучшает качество оценки модели – но только для обучающей выборки, в тестовой выборке улучшений качества оценки с увеличением глубины не наблюдается. Полученные результаты робастны относительно большого диапазона параметров глубины.

Параметр скорости сходимости (*eta*) был выбран 0,15. Анализировался диапазон различных параметров скорости сходимости. Более низкие значения потенциально позволяют достигнуть более высокой точности модели. С учетом контроля процесса обучения на тестовой выборке, этот параметр влияет, в основном, на скорость работы. В результате более низкие значения *eta* не дают выигрыша в точности модели в тестовой выборке и замедляют процесс тренировки.

Использовалась функция ошибок, в которой оценивается доля полученных оценок, отличающихся от истинных (наблюдаемых) значений.

Как уже упоминалось выше, каждый процесс обучения контролировался на тестовой выборке. Для параметра *early_stopping_rounds* мы использовали значение 50.

Тренировка модели – это итеративная процедура построения классификатора, когда на каждом шаге к существующему классу классификаторов добавляется новая CART функция так, чтобы минимизировать ошибку оценки зависимой переменной (стараясь поддерживать максимально простую структуру CART).

На Рис. 1 представлена динамика логарифмической функции ошибок. По горизонтальной оси откладывается номер итерации, по вертикальной оси – значения логарифмической функции классификации на данной итерации. На Рис. 1 приводятся результаты оценок для 100 различных кросс-валидаций, каждой из которых соответствует своя кривая ошибок обучающей (синяя сплошная линия) и тестовой выборки (красная прерывистая линия). На Рис. 1 также приводится среднее значение, полученное по всем 100 оценкам. Так, для обучающей выборки (train Error) это значение равно 0,081, то есть на обучающей выборке модель в среднем дает оценку индекса социально-политической стабилизации, которая ошибается в 8,1 % случаев (относительно реального наличия фактов террористических атак в данной стране в данном году). Можно заметить, что с ростом числа итераций (усложнения модели) значение логарифмической функции ошибок на обучающей выборке (синие сплошные кривые) стабильно снижается и, если продолжать обучение, можно достичь более высокой точности. Однако для нас критерием точности модели является ошибка на тестовой выборке (test-error) – в среднем для переменной `n_terror_attack` она оказывается равной 0,166. С использованием параметра `early_stopping_rounds` мы прерываем дальнейшую тренировку модели, если за последние 50 итераций не было улучшений в оценках на тестовой выборке (test-error).

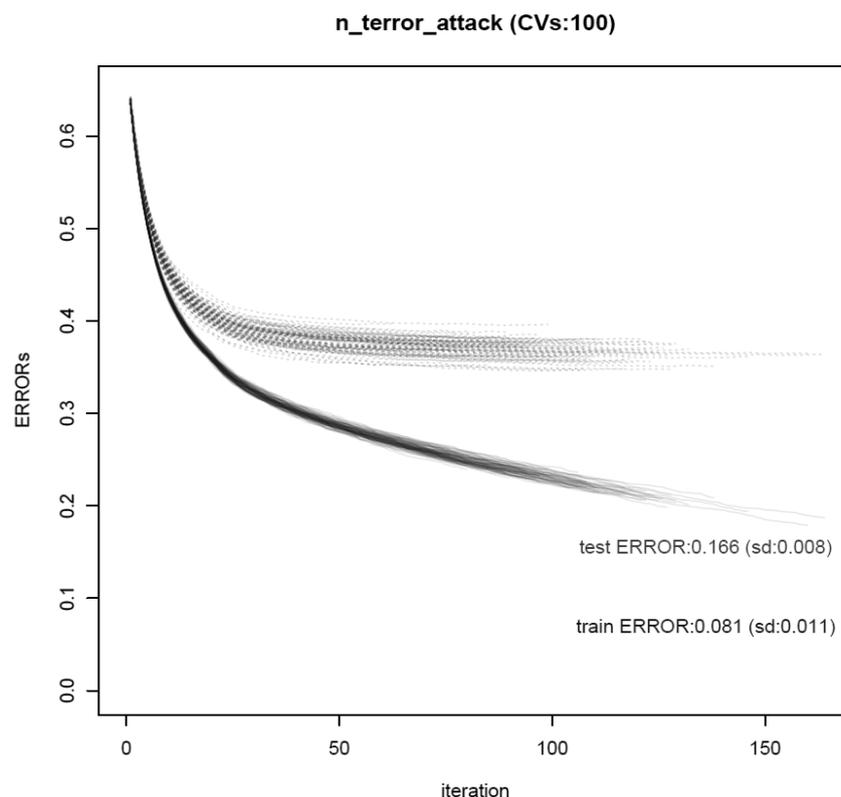


Рис. 1. Динамика логарифмической функции ошибок на обучающей и тестовой выборках для количества террористических атак (`n_terror_attack`)

Примечание: на графике приводится среднее значение и стандартное отклонение (в скобках) для ошибок, рассчитанное для 100 оценок.

Качество полученных оценок мы также можем оценить с помощью ROC-кривых, которые показывают, как соотносятся частота ложноположительных оценок (false-positive rate – FPR) с частотой истинно-положительными оценок (true-positive rate – TPR). На Рис. 2 представлены ROC-кривые для 100 оценок модели на обучающих и тестовых выборках для количества террористических атак (`n_terror_attack`).

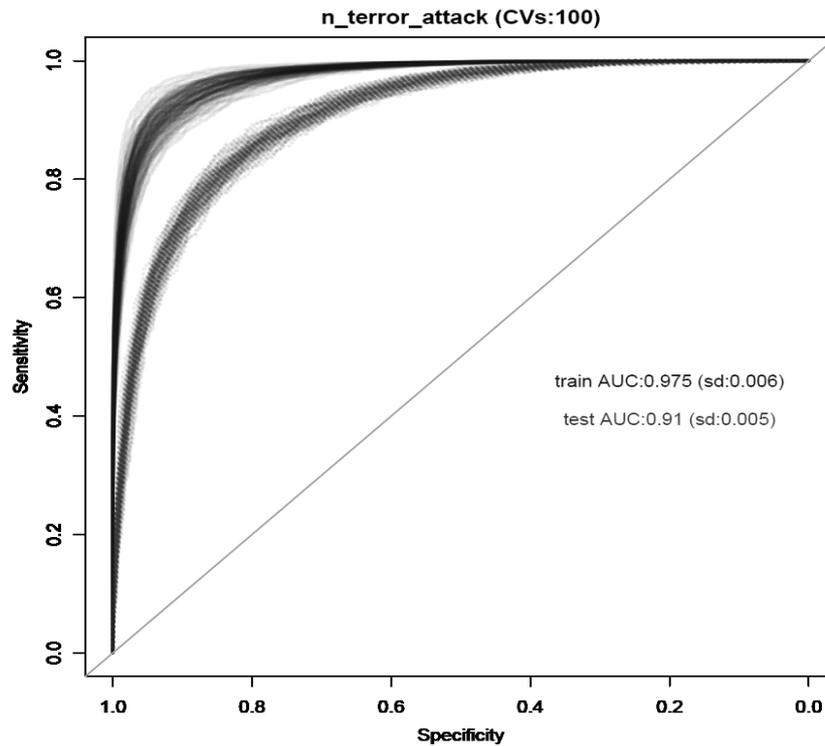


Рис. 2. ROC-кривые для оценок модели на обучающих и тестовых выборках для количества террористических атак (n_terror_attack)

Примечание: на графике приводится среднее значение и стандартное отклонение (в скобках) показателя AUC рассчитанное для 100 прогнозных оценок.

На Рис. 2 по горизонтальной оси отложен коэффициент специфичности (Specificity), который рассчитывается как единица минус частота ложно-положительных оценок ($1 - FPR$), а по вертикальной оси отложен коэффициент чувствительности (Sensitivity) который равен коэффициенту истинно-положительных оценок (TPR). Идеальный классификатор – это классификатор, у которого сочетаются нулевые ложно-положительные ошибки (специфичность = 1) и 100%-ная чувствительность (то есть $TPR = 1$). Модель, которая случайным образом пытается угадать значение, будет соответство-

вать горизонтальной прямой $Sensitivity = 1 - Specificity$ (или $TPR = FPR$). ROC-кривая для модели описывает сочетания TPR и FPR для полученных оценок. Чем ближе ROC-кривая к горизонтальной прямой, тем хуже работает модель, тем меньше объясняющая способность модели и ниже качество полученных оценок. В качестве интегрального критерия качества модели на ROC-кривой используется показатель *площади под ROC-кривой* ($AUC = Area Under Curve$). Для идеального классификатора $AUC = 1$, для абсолютно неинформативной модели (горизонтальной прямой) $AUC = 0,5$. На Рис. 2 видно, что модель на обучающей выборке обладает высокими прогностическими возможностями, но и на тестовой выборке ее прогностические способности оказываются ненулевыми ($AUC = 0.856$).

Полученные результаты оценок ошибок и AUC для выбранных нами 6 измерений нестабильности приведены в Табл. 2, а подробные графики динамики ошибок и ROC-кривые для всех 6 зависимых переменных приводятся в приложениях 1 и 2 соответственно.

Табл. 2. Оценки качества моделей на обучающих и тестовых выборках

Переменная	Обучающая выборка (train)		Тестовая выборка (test)	
	Error	AUC	Error	AUC
GK1_1992_world	0.010 (0.006)	0.999 (0.001)	0.056 (0.006)	0.953 (0.010)
cnts_domestic1	0.060 (0.004)	0.949 (0.011)	0.081 (0.005)	0.827 (0.015)
cnts_domestic3	0.048 (0.008)	0.983 (0.005)	0.098 (0.005)	0.901 (0.008)
cnts_domestic5	0.063 (0.006)	0.963 (0.008)	0.088 (0.006)	0.881 (0.010)
n_terror_attack	0.081 (0.011)	0.975 (0.006)	0.166 (0.008)	0.910 (0.005)
nkill	0.074 (0.009)	0.976 (0.006)	0.143 (0.007)	0.908 (0.006)

Примечание: приводится среднее значение и стандартное отклонение (в скобках) ошибок и показателя AUC рассчитанное для 20 прогнозных оценок.

2.4. Оценка значимости факторов

После тренировки (оценки) моделей градиентного бустинга (по 20 моделей на каждую из 6 зависимых переменных), для каждой полученной оценки мы анализировали значимость независимых факторов. Важность факторов в модели градиентного бустинга оценивается по трем параметрам:

- *gain* – описывает относительный вклад соответствующего фактора в модель, рассчитанный путем оценки вклада фактора для каждого дерева в модели;
- *cover* – оценивает, для какой доли в исходных данных (для скольких точек данных) анализируемая переменная влияет на результат классификации;
- *frequency* – оценивает, сколько раз независимый фактор используется для разделения данных по всем деревьям (количество вершин в дереве принятия решений, в которых используются значения анализируемого фактора).

Каждый из этих факторов может быть использован для оценки относительной важности переменной. Итоговый индекс значимости переменных представлен в формуле (4).

$$importance = (gain + cover + frequency)/3 \quad (4)$$

Размерность коэффициента важности такова, что количественно коэффициент важности (*importance*) на уровне 1, можно интерпретировать как 1 %-ный вклад в объясняющую способность модели (через вклад в выигрыш от использования, покрытие данных, частоту использования).

3. Описание результатов

Результаты анализа важности независимых факторов мы представляем следующим образом. Сначала мы представляем агрегированные результаты важности всех факторов для всех 6 зависимых переменных. Далее кратко перечисляем наиболее важные факторы по каждому из 6 направлений дестабилизации.

Тренировка 100 бустинговых моделей кросс-валидации позволяет выделить следующие основные независимые факторы, влияющие на переменную «Количество террористических атак, в ед.» (*n_terror_attack*):

1. Численность населения, тыс. человек (population_unesco [13.4]),
2. Площадь, км² (cnts_area1 [6.6]),
3. Число лет со времени последней смены режима, в ед. (p_durable [3.9]),
4. Плотность населения (0,1), количество человек на км² (cnts_pop_density [3.2]).

Вторая группа факторов (с коэффициентом importance > 1,5): Финансовая помощь со стороны США, в долл. (us_foreign_aid), Доля городского населения, в % (share_urban_population_UN), Инфляция, измеряемая через индекс потребительских цен, в % (относительно предыдущего года) (consumer_price_index), Доля импорта товаров и услуг в ВВП, в % (import_percent_gdp), Коэффициент Джини (уровень неравенства), в ед. (Gini), Доля ПИИ в ВВП, в % (fdi_inward_annual_growth), Доля валовых внутренних инвестиций в ВВП, в % (gross_fixed_capital_formation_percent_gdp), Темпы роста ВВП на душу населения по ППС, в постоянных ценах 2017 г., в % (gdp_per_capita_annual_growth), Доля населения студенческого возраста, охваченного высшим образованием (vdem_undp_tertiary_school_enrollment), Антиправительственные демонстрации (cnts_domestic8).

В третью по важности группу входят следующие 11 факторов: ВВП на душу населения по ППС, в постоянных ценах 2017 г. (gdp_per_capita_PPP_Mad), Индекс глобализации, в ед. (index_globalization), Уровень безработицы (по данным МОТ), в % (unemployment_rate_ILO), Фракционность элит (factionalized_elites), Количество лет обучения, в годах (mean_year_schooling), Средняя продолжительность жизни, в годах (life_expectancy), Доля экспорта товаров и услуг в ВВП, в % (export_percent_gdp), Грамотность, % (cnts_literacy), Доля населения, которое подвергается активной, преднамеренной и целенаправленной дискриминации со стороны государства с целью исключения их из политической власти (erg_discriminated), ВВП на душу населения по ППС, в постоянных ценах 2017 г. (gdp_per_capita_PPP_WB), Индекс верховенства права (wdi_rule_of_law).

Данные факторы совместно объясняют 71 % общей значимости модели. Второй группой факторов объясняется 24 %, факторами третьей группы 20 % общей значимости.

Для переменной «Количество убитых» (**nkill**) основными факторами в порядке убывания важности являются:

1. Численность населения, тыс. человек (population_unesco [13.9]),
2. Площадь, км² (cnts_area1 [5.9]),
3. Число лет со времени последней смены режима, в ед. (p_durable [4.1]),
4. Финансовая помощь со стороны США, в долл. (us_foreign_aid [3.7]).

Вторая группа факторов: плотность населения (0,1), количество человек на км² (cnts_pop_density), коэффициент Джини (уровень неравенства), в ед. (Gini), инфляция, измеряемая через индекс потребительских цен, в % (относительно предыдущего года) (consumer_price_index), доля импорта товаров и услуг в ВВП, в % (import_percent_gdp), доля городского населения, в % (share_urban_population_UN), индекс глобализации, в ед. (index_globalization), доля ПИИ в ВВП, в % (fdi_inward_annual_growth), доля валовых внутренних инвестиций в ВВП, в % (gross_fixed_capital_formation_percent_gdp), индекс потребительских цен, в % (относительно предыдущего года) (inflation_consumer_prices), фракционность элит (factionalized_elites), доля населения, которое подвергается активной, преднамеренной и целенаправленной дискриминации со стороны государства с целью исключения их из политической власти (epg_discriminated).

Следующие по важности переменные: уровень безработицы (по данным МОТ), в % (unemployment_rate_ILO), средняя продолжительность жизни, в годах (life_expectancy), доля экспорта товаров и услуг в ВВП, в % (export_percent_gdp), темпы роста ВВП на душу населения по ППС, в постоянных ценах 2017 г., в % (gdp_per_capita_annual_growth), ВВП на душу населения по ППС, в постоянных ценах 2017 г. (gdp_per_capita_PPP_WB), уровень безработицы (по данным национальных служб), в % (unemployment_rate_NAT), дельта городского населения, в ед. (delta_urban_population_UN), доля населения в возрасте от 15 до 24 лет в общей численности населения, в % (0,01) (share_15_24), доля населения студенческого возраста, охваченного высшим образованием (vdem_undp_tertiary_school_enrollment), доля голодающих в общей численности населения, в % (share_hungry_population).

Приведенные выше переменные совместно отвечают за 71 % общей значимости модели. Вторая группа факторов отвечает за 26 %, третья – 17 %.

После анализа результатов оценивания 100 моделей кросс-валидации для зависимой **переменной «Политические убийства» (cnts_domestic1)** были выделены следующие факторы:

1. Площадь, км² (cnts_area1 [6.7]),
2. Число лет со времени последней смены режима, в ед. (p_durable [6]),
3. Антиправительственные демонстрации (cnts_domestic8 [4.6]),
4. Плотность населения (0,1), количество человек на км² (cnts_pop_density [4.4]).

Вторая группа переменных: финансовая помощь со стороны США, в долл. (us_foreign_aid), массовые беспорядки (cnts_domestic6), численность населения, тыс. человек (population_unesco), индекс потребительских цен, в % (относительно предыдущего года) (inflation_consumer_prices), коэффициент Джини (уровень неравенства), в ед. (Gini), темпы роста ВВП на душу населения по ППС, в постоянных ценах 2017 г., в % (gdp_per_capita_annual_growth), доля городского населения, в % (share_urban_population_UN), ВВП на душу населения по ППС, в постоянных ценах 2017 г. (gdp_per_capita_PPP_Mad), инфляция, измеряемая через индекс потребительских цен, в % (относительно предыдущего года) (consumer_price_index), индекс верховенства права (wdi_rule_of_law), доля валовых внутренних инвестиций в ВВП, в % (gross_fixed_capital_formation_percent_gdp), уровень младенческой смертности на 1000 живорождений (vdem_infant_mortality_rate), доля населения среднего школьного возраста, охваченного средним образованием (vdem_undp_secondary_school_enrollment), доля населения студенческого возраста, охваченного высшим образованием (vdem_undp_tertiary_school_enrollment).

Следующие по важности факторы (с коэффициентом importance > 1): грамотность, % (cnts_literacy), дельта городского населения, в ед. (delta_urban_population_UN), доля ПИИ в ВВП, в % (fdi_inward_annual_growth), доля экспорта товаров и услуг в ВВП, в % (export_percent_gdp), средняя продолжительность жизни, в годах (life_expectancy), доля импорта товаров и услуг в ВВП, в %

(import_percent_gdp), доля населения в возрасте от 15 до 24 лет, в % (0,01) (share_15__24).

Приведенные факторы совместно объясняют 75 % общей значимости модели. Второй группой факторов объясняется 42 %, факторами третьей группы – 12 %.

Тренировка 100 бустинговых моделей кросс-валидации позволяет выделить следующие основные независимые факторы, влияющие на переменную «**Партизанские действия**» (cnts_domestic3):

1. Площадь, км² (cnts_area1 [9.3]),
2. Численность населения, тыс. человек (population_unesco [4.9]),
3. Финансовая помощь со стороны США, в долл. (us_foreign_aid [4.9]),
4. Плотность населения (0,1), количество человек на км² (cnts_pop_density [4.4]).

Вторая группа факторов (с коэффициентом importance > 1,5): Число лет со времени последней смены режима, в ед. (p_durable), грамотность, % (cnts_literacy), массовые беспорядки (cnts_domestic6), коэффициент Джини (уровень неравенства), в ед. (Gini), доля населения среднего школьного возраста, охваченного средним образованием (vdem_undp_secondary_school_enrollment), ВВП на душу населения по ППС, в постоянных ценах 2017 г. (gdp_per_capita_PPP_Mad), доля населения, которое подвергается активной, преднамеренной и целенаправленной дискриминации со стороны государства с целью исключения их из политической власти (erg_discriminated), доля населения студенческого возраста, охваченного высшим образованием (vdem_undp_tertiary_school_enrollment), дельта городского населения, в ед. (delta_urban_population_UN), инфляция, измеряемая через индекс потребительских цен, в % (относительно предыдущего года) (consumer_price_index), доля валовых внутренних инвестиций в ВВП, в % (gross_fixed_capital_formation_percent_gdp), доля городского населения, в % (share_urban_population_UN), уровень младенческой смертности на 1000 живорождений (vdem_infant_mortality_rate), доля населения начального школьного возраста, охваченного начальным образованием (vdem_undp_primary_school_enrollment), доля экспорта товаров и услуг в ВВП, в % (export_percent_gdp), доля населения в возрасте от 15 до 24 лет в общей численности населения, в % (0,01) (share_15__24).

В третью по важности группу входят следующие 5 факторов: средняя продолжительность жизни, в годах (*life_expectancy*), темпы роста ВВП на душу населения по ППС, в постоянных ценах 2017 г., в % (*gdp_per_capita_annual_growth*), доля импорта товаров и услуг в ВВП, в % (*import_percent_gdp*), количество лет обучения, в годах (*mean_year_schooling*), индекс глобализации, в ед. (*index_globalization*).

Данные факторы совместно объясняют 76 % общей значимости модели. Второй группой факторов объясняется 43 %, факторами третьей группы 9 % общей значимости.

Тренировка 100 бустинговых моделей позволяет выделить следующие основные факторы, влияющие на зависимую переменную **«Политические репрессии» (*cnts_domestic5*)**:

1. Площадь, км² (*cnts_areal* [8.1]),
2. Число лет со времени последней смены режима, в ед. (*p_durable* [6.5]),
3. Индекс глобализации, в ед. (*index_globalization* [6.1]),
4. Инфляция, измеряемая через индекс потребительских цен, в % (относительно предыдущего года) (*consumer_price_index* [5.3]).

Во вторую группу входят следующие 14 переменных: плотность населения (0,1), количество человек на км² (*cnts_pop_density*), массовые беспорядки (*cnts_domestic6*), доля населения среднего школьного возраста, охваченного средним образованием (*vdem_undp_secondary_school_enrollment*), антиправительственные демонстрации (*cnts_domestic8*), грамотность, % (*cnts_literacy*), ВВП на душу населения по ППС, в постоянных ценах 2017 г. (*gdp_per_capita_PPP_Mad*), доля ПИИ в ВВП, в % (*fdi_inward_annual_growth*), численность населения, тыс. человек (*population_unesco*), финансовая помощь со стороны США, в долл. (*us_foreign_aid*), доля населения студенческого возраста, охваченного высшим образованием (*vdem_undp_tertiary_school_enrollment*), доля импорта товаров и услуг в ВВП, в % (*import_percent_gdp*), темпы роста ВВП на душу населения по ППС, в постоянных ценах 2017 г., в % (*gdp_per_capita_annual_growth*), темпы роста городского населения, в ед. (*delta_urban_population_UN*), фракционность элит (*factionalized_elites*).

Следующие по важности переменные: доля населения в возрасте от 20 до 29 лет в общей численности населения, в % (0,01) (*share_20_29*), доля населения начального школьного возраста,

охваченного начальным образованием (`vdem_undp_primary_school_enrollment`), коэффициент Джини (уровень неравенства), в ед. (`Gini`), доля населения, которое подвергается активной, преднамеренной и целенаправленной дискриминации со стороны государства с целью исключения их из политической власти (`epg_discriminated`), доля городского населения, в % (`share_urban_population_UN`), уровень младенческой смертности на 1000 живорождений (`vdem_infant_mortality_rate`), число лет со времени последней смены режима (`pol_durable`).

Данные факторы совместно объясняют 80 % общей значимости модели. Второй группой факторов объясняется 42 %, факторами третьей группы 12 % общей значимости.

Тренировка 100 бустинговых моделей кросс-валидации позволяет выделить следующие основные независимые факторы влияющих на переменную «Индекс кровавой дестабилизации» (**GK1_1992_world**):

1. Численность населения, тыс. человек (`population_unesco` [17]),
2. Финансовая помощь со стороны США, в долл. (`us_foreign_aid` [6.5]),
3. Фракционность элит (`factionalized_elites` [5.4]),
4. Доля городского населения, в % (`share_urban_population_UN` [3.2]).

Вторая группа факторов (с коэффициентом `importance > 1,5`): плотность населения (0,1), количество человек на км² (`cnts_pop_density`), чистая миграция, в ед. (`net_migration`), число лет со времени последней смены режима, в ед. (`p_durable`), уровень безработицы (по данным МОТ), в % (`unemployment_rate_ILO`), индекс верховенства права (`wdi_rule_of_law`), площадь, км² (`cnts_area1`), доля ПИИ в ВВП, в % (`fdi_inward_annual_growth`), доля экспорта товаров и услуг в ВВП, в % (`export_percent_gdp`), инфляция, измеряемая через индекс потребительских цен, в % (относительно предыдущего года) (`consumer_price_index`), ВВП на душу населения по ППС, в постоянных ценах 2017 г. (`gdp_per_capita_PPP_WB`), количество лет обучения, в годах (`mean_year_schooling`), дельта городского населения, в ед. (`delta_urban_population_UN`), темпы роста ВВП на душу населения по ППС, в постоянных ценах 2017 г., в % (`gdp_per_capita_annual_growth`).

В третью по важности группу входят следующие 8 факторов: доля валовых внутренних инвестиций в ВВП, в % (`gross_fixed_capital_formation_percent_gdp`), доля населения, которое подвергается активной, преднамеренной и целенаправленной дискриминации со стороны государства с целью исключения их из политической власти (`egr_discriminated`), коэффициент Джини (уровень неравенства), в ед. (`Gini`), доля импорта товаров и услуг в ВВП, в % (`import_percent_gdp`), индекс потребительских цен, в % (относительно предыдущего года) (`inflation_consumer_prices`), индекс глобализации, в ед. (`index_globalization`), индекс борьбы с коррупцией (`wdi_control_of_corruption`), грамотность, % (`cnts_literacy`).

Данные факторы совместно объясняют 78 % общей значимости модели. Второй группой факторов объясняется 31 %, факторами третьей группы 15 % общей значимости.

По результатам оценки по 100 моделей для **всех 6 зависимых переменных** мы выделили следующие наиболее важные независимые факторы:

1. Численность населения, тыс. человек (`population_unesco` [9.2]),
2. Площадь, км² (`cnts_area1` [6.5]),
3. Число лет со времени последней смены режима, в ед. (`p_durable` [4.6]),
4. Финансовая помощь со стороны США, в долл. (`us_foreign_aid` [4.1]),
5. Плотность населения (0,1), количество человек на км² (`cnts_pop_density` [3.8]),
6. Инфляция, измеряемая через индекс потребительских цен, в % (относительно предыдущего года) (`consumer_price_index` [3]),
7. Коэффициент Джини (уровень неравенства), в ед. (`Gini` [2.7]),
8. Массовые беспорядки (`cnts_domestic6` [2.5]),
9. Индекс глобализации, в ед. (`index_globalization` [2.5]),
10. Доля городского населения, в % (`share_urban_population_UN` [2.5]).

Во вторую группу входят следующие 8 переменных: грамотность, % (`cnts_literacy`), фракционность элит (`factionalized_elites`), доля ПИИ в ВВП, в % (`fdi_inward_annual_growth`), ВВП на душу населения по ППС, в постоянных ценах 2017 г. (`gdp_per_capita_PPP_Mad`), антиправительственные демонстрации (`cnts_domestic8`), темпы роста ВВП на душу населения по ППС, в постоянных ценах 2017 г., в % (`gdp_per_capita_annual_growth`), Доля валовых внут-

ренных инвестиций в ВВП, в % (`gross_fixed_capital_formation_percent_gdp`), доля импорта товаров и услуг в ВВП, в % (`import_percent_gdp`).

Следующие по важности переменные: доля населения среднего школьного возраста, охваченного средним образованием (`vdem_undp_secondary_school_enrollment`), дельта городского населения, в ед. (`delta_urban_population_UN`), индекс потребительских цен, в % (относительно предыдущего года) (`inflation_consumer_prices`), доля населения, которое подвергается активной, преднамеренной и целенаправленной дискриминации со стороны государства с целью исключения их из политической власти (`emp_discriminated`), доля населения студенческого возраста, охваченного высшим образованием (`vdem_undp_tertiary_school_enrollment`), доля экспорта товаров и услуг в ВВП, в % (`export_percent_gdp`), уровень безработицы (по данным МОТ), в % (`unemployment_rate_ILO`).

Данные факторы совместно объясняют 72 % общей значимости модели. Второй группой факторов объясняется 17 %, факторами третьей группы 13 % общей значимости.

Обсуждение результатов и заключение

Используя методы градиентного бустинга, мы отобрали набор переменных, которые оказываются наиболее важными для оценки индикаторов политической нестабильности.

Среди наиболее важных факторов выделяются переменные, которые описывают историю существования и устойчивости режима (число лет со времени последней смены режима, возраст государственности и независимость); переменные, характеризующие тип режима (индекс регулирования участия в политике, комбинированная оценка Polity IV); переменные, характеризующие структуру населения и занятости (население в возрасте от 0 до 4 лет (за 5 лет), доля занятых в промышленности), а также переменные, отражающие состояние мировой конъюнктуры (цена на золото, индекс потребительских цен).

В текущей версии мы использовали модели, в которых анализируется ограниченное число аспектов нестабильности (измеряемых с помощью индексов нестабильности CNTS и еще трех дополнительных переменных). Дальнейшее уточнение, формализация понятия нестабильности поможет уточнить набор переменных для других измерений нестабильности.

Для отбора наиболее важных факторов, влияющих на нестабильность, мы использовали модель «common pool», в которой единицей наблюдения является точка страна-год. Построение моделей, ориентированных на предсказание нестабильности, позволит еще в большей степени уточнить набор наиболее важных факторов.

Библиография

- Коротаев А. В., Зинькина Ю. В. 2012.** Структурно-демографические факторы «арабской весны». *Протестные движения в арабских странах. Предпосылки, особенности, перспективы* / Ред. И. В. Следзевский, А. Д. Саватеев. М.: ЛИБРОКОМ/URSS. С. 28–40.
- Коротаев А. В., Медведев И. А., Слинько Е. В., Шульгин С. Г. 2020.** Эффективность систем глобального мониторинга рисков социально-политической дестабилизации: опыт систематического анализа. *Социологическое обозрение* 19(2): 143–197.
- Малков С. Ю., Коротаев А. В., Исаев Л. М., Кузьминова Е. В. 2013.** О методике оценки текущего состояния и прогноза социальной нестабильности: опыт количественного анализа событий Арабской весны. *Полис. Политические исследования* 4: 137–162.
- Слинько Е. В., Мещерина К. В., Шульгин С. Г., Зинькина Ю. В., Блюга С. Э., Коротаев А. В. 2018.** Измерение внутривнутриполитических дестабилизационных процессов: типы нестабильности и их связь с социально-политическими и экономическими факторами. *Системный мониторинг глобальных и региональных рисков. Социально-политическая и экономическая дестабилизация: анализ страновых и региональных ситуаций в мир-системном аспекте: ежегодник* / Отв. ред. Л. Е. Гринин, А. В. Коротаев, К. В. Мещерина. Волгоград: Учитель. С. 95–114.
- Цирель С. В. 2012.** Условия возникновения революционных ситуаций в арабских странах. *Системный мониторинг глобальных и региональных рисков. Арабская весна 2011 года* / Отв. ред. А. В. Коротаев, Ю. В. Зинькина, А. С. Ходунов. М.: ЛКИ/URSS. С. 162–173.
- Цирель С. В. 2015.** К истокам украинских революционных событий 2013–2014 гг. *Системный мониторинг глобальных и региональных рисков. Украинский разлом* / Отв. ред. Л. Е. Гринин, А. В. Коротаев, Л. М. Исаев, А. Р. Шишкина. Волгоград: Учитель. С. 254–279.
- Aksoy D., Carter D. B., Wright J. 2012.** Terrorism in dictatorships. *Journal of Politics* 74(3): 810–826.
- Alesina A., Devleescauwer A., Easterly W., Kurlat S. 2003.** Fractionalization. *Journal of Economic Growth* 8(2): 155–194.

- Astashkin A., Chuvilin K. 2017.** Syntax description synthesis using gradient boosted trees. In: *2017 20th Conference of Open Innovations Association (FRUCT)*. IEEE: 32–39.
- Banks A. S., Wilson K. A. 2018.** *Cross-National Time-Series Data Archive*. Israel: Databanks International. URL: <https://www.cntsdata.com/>.
- Banks A. S., Wilson K. A. 2020.** *Cross-National Time-Series Data Archive*. Jerusalem: Databanks International.
- Bove V., Böhmelt T. 2016.** Does immigration induce terrorism?. *Journal of Politics* 78(2): 572–588.
- Breiman L. 2001.** Random forests. *Machine learning* 45(1): 5–32.
- Brockhoff S., Krieger T., Meierrieks D. 2015.** Great Expectations and Hard Times: The (Nontrivial) Impact of Education on Domestic Terrorism. *The Journal of Conflict Resolution* 59(7): 1186–1215.
- Caruso R., Gavrilova E. 2012.** Youth unemployment, terrorism and political violence, evidence from the Israeli/Palestinian conflict. *Peace Economics, Peace Science and Public Policy* 18(2): 1–37.
- Caruso R., Schneider F. 2011.** The socio-economic determinants of terrorism and political violence in Western Europe (1994–2007). *European Journal of Political Economy* 27: 37–49.
- Cederman L., Min B., Wimmer A. 2005.** *The Ethnic Power Relations (EPR) dataset*. Zürich: ETH.
- Center for Systemic Peace 2019.** *Coups d'Etat, 1946–2018*. Vienna: Center for Systemic Peace. URL: <http://www.systemicpeace.org/inscrdata.html>.
- Chen T., Guestrin C. 2016.** *Xgboost: A scalable tree boosting system*. Proceedings of the 22nd acm SIGKDD international conference on knowledge discovery and data mining. August 13–17, 2016, San Francisco: 785–794.
- Choi S. W., Luo S. 2013.** Economic Sanctions, Poverty, and International Terrorism: An Empirical Analysis. *International Interactions* 39(2): 217–245.
- Coggins B. L. 2015.** Does State Failure Cause Terrorism? An Empirical Analysis (1999–2008). *Journal of Conflict Resolution* 59(3): 455–483.
- Connelly R., Playford C. J., Gayle V., Dibben C. 2016.** The Role of Administrative Data in the Big Data Revolution. *Social Science Research* 59: 1–12. DOI: 10.1016/j.ssresearch.2016.04.015.
- Coppedge M., Gerring J., Knutsen C. H., Lindberg S. I., Skaaning S. E., Teorell J., Ziblatt D. 2018.** *V-Dem [Country-Year/Country-Date] Dataset v8*. Gothenburg: V-Dem Institute.
- Coppock A., Guess A., Ternovski J. 2016.** When Treatments are Tweets: A Network Mobilization Experiment over Twitter. *Political Behavior* 38(1): 105–128. DOI: 10.1007/s11109-015-9308-6.

- Donnay K. 2017.** Big Data for Monitoring Political Instability. *International Development Policy / Revue internationale de politique de développement*. 8(8.1). URL: <http://journals.openedition.org/poldev/2468> (accessed: 01.08.2021).
- Donnay K., Dunford E., McGrath E. C., Backer D., Cunningham D. E. 2016.** *MELTT: Matching Event Data by Location, Time and Type*. Chicago: Annual Conference of the Midwest Political Science Association.
- Enders W., Hoover G. A., Sandler T. 2016.** The Changing Nonlinear Relationship between Income and Terrorism. *Journal of Conflict Resolution* 60(2): 195–225.
- Esty D., Goldstone J. A., Gurr T. R., Harff B., Levy M., Dabelko G. D., Surko P., Unger A. N. 1999.** State Failure Task Force Report: Phase II Findings. *Environmental Change & Security Project Report 5*: 49–75.
- Findley M. G., Young J. K. 2012.** More Combatant Groups, More Terror?: Empirical Tests of an Outbidding Logic. *Terrorism and Political Violence* 24(5): 706–721.
- Foster D. M., Braithwaite A., Sobek D. 2013.** There can be no compromise: Institutional inclusiveness, fractionalization and domestic terrorism. *British Journal of Political Science* 43(3): 541–557.
- Freedom House 2017.** *Freedom in the World*. Washington, DC: Freedom House. URL: <https://freedomhouse.org/report/methodology-freedom-world-2017>.
- Freytag A., Krüger J. J., Meierrieks D., Schneider F. 2011.** The origins of terrorism: Cross-country estimates of socio-economic determinants of terrorism. *European Journal of Political Economy* 27: 5–16.
- Friedman J. H. 2001.** Greedy function approximation: a gradient boosting machine. *Annals of statistics* 29(5): 1189–1232.
- Gaibulloev K., Piazza J. A., Sandler T. 2017.** Regime types and terrorism. *International Organization* 71(3): 491–522.
- Gassebner M., Luechinger S. 2011.** Lock, stock, and barrel: A comprehensive assessment of the determinants of terror. *Public Choice* 149(3): 235–261.
- George J. 2018.** State failure and transnational terrorism: An empirical analysis. *Journal of Conflict Resolution* 42(3): 471–495.
- Gygli S., Haelg F., Potrafke N., Sturm J. E. 2019.** The KOF Globalisation Index – Revisited. *Review of International Organizations* 14(3): 543–574.
- Inklaar R., Jong de H., Bolt J., Van Zanden J. 2018.** *Rebasing Maddison: new income comparisons and the shape of long-run economic development*. Groningen: Groningen Growth and Development Centre.
- Kis-Katos K., Liebert H., Schulze G. G. 2011.** On the origin of domestic and international terrorism. *European Journal of Political Economy* 27: 17–36.

- Kuhn M. 2008.** Building Predictive Models in R Using the caret Package. *Journal of Statistical Software* 28(5): 1–26.
- Lafree G., Bersani B. E. 2014.** County-level correlates of terrorist attacks in the United States. *Criminology and Public Policy* 13(3): 455–481.
- Lee A. 2011.** Who becomes a terrorist? Poverty, education, and the origins of political violence. *World Politics* 63(2): 203–245.
- Marshall M. G., Marshall D. R. 2016.** *Coup D'État Events, 1946-2015 Codebook*. Vienna: Center for Systemic Peace.
- Mironov V., Guschin A. 2015.** *1st place of the CERN LHCb experiment Flavour of Physics competition*. URL: <http://blog.kaggle.com/2015/11/30/flavour-of-physics-technical-write-up-1st-place-go-polar-bears/>.
- Mousseau M. 2011.** Urban poverty and support for Islamist terror: Survey results of Muslims in fourteen countries. *Journal of Peace Research* 48(1): 35–47.
- Østby G., Urdal H., Dupuy K. 2019.** Does Education Lead to Pacification? A Systematic Review of Statistical Studies on Education and Political Violence. *Review of Educational Research* 89(1): 46–92.
- Paul J. A., Bagchi A. 2018.** Does terrorism increase after a natural disaster? An analysis based upon property damage. *Defence and peace economics* 29(4): 407–439.
- Polity IV Annual Time-Series, 1800–2018.** URL: <http://www.systemicpeace.org/polity/polity4.htm>.
- Przeworski A., Alvarez M. E., Cheibub J. A., Limongi F. 2000.** *Democracy and Development; Political Institutions and Well-Being in the World, 1950–1990*. New York: Cambridge University Press.
- Sandulescu V., Chiru M. 2016.** Predicting the future relevance of research institutions-The winning solution of the KDD Cup 2016. *Cornell University: arXiv preprint arXiv:1609.02728*.
- Slav M., Smyslovskikh E., Novikov V., Kolesnikov I., Korotayev A. 2021.** Deprivation, instability, and propensity to attack: how urbanization influences terrorism. *International Interactions*. DOI: 10.1080/03050629.2021.1924703.
- START [National Consortium for the Study of Terrorism and Responses to Terrorism]. 2019.** *Global Terrorism Database*. College Park, MD: National Consortium for the Study of Terrorism and Responses to Terrorism. URL: <https://www.start.umd.edu/gtd/>.
- Tausch N., Becker J. C., Spears R., Christ O., Saab R., Singh P., Siddiqui R. N. 2011.** Explaining radical group behavior: Developing emotion and efficacy routes to normative and nonnormative collective action. *Journal of Personality and Social Psychology* 101(1): 129–148.

- The Fund for Peace 2021.** *Fragile States Index*. Washington, D.C: The Fund for Peace.
- Tianqi C., He T., Benesty M., Khotilovich V., Tang Y., Cho H., Chen K., Mitchell R., Cano I., Zhou T., Li M., Xie J., Lin M., Geng Y., Li Y. 2018.** *Xgboost: Extreme Gradient Boosting*. R package version 0.71.2. URL: <https://CRAN.R-project.org/package=xgboost>.
- Transparency International 2020.** *Corruption Perception Index*. Berlin: Transparency International. URL: <https://www.transparency.org/en/cpi/2020/index/nzl>.
- U.S. Agency for International Development (USAID), U.S. Department of State on behalf of United States Government agencies reporting foreign assistance 2020.** *ForeignAssistance.gov*.
- UNDP [United Nations Population Division] 2018.** *World Urbanization Prospects Database*. New York: United Nations.
- UNDP [United Nations Population Division] 2019.** *Human Development Report 2019: Beyond income, beyond averages, beyond today: Inequalities in human development in the 21st century*. New York: United Nations.
- UNDP [United Nations Population Division] 2020.** *World Urbanization Prospects Database*. New York: United Nations.
- UNESCO Institute for Statistics 2021.** *UIS Statistics*. Quebec: UNESCO.
- Victoroff J., Adelman J. R., Matthews M. 2012.** Psychological Factors Associated with Support for Suicide Bombing in the Muslim Diaspora. *International Society of Political Psychology* 33(6): 791–809.
- Volkovs M., Yu G. W., Poutanen T. 2017.** Content-based Neighbor Models for Cold Start Recommender Systems. *Proceedings of RecSys Challenge*: 1–6.
- Weber H. 2019.** Age structure and political violence: a re-assessment of the “youth bulge” hypothesis. *International Interactions* 45(1): 80–112.
- Wilson M. C., Piazza J. A. 2013.** Autocracies and Terrorism: Conditioning Effects of Authoritarian Regime Type on Terrorist Attacks. *American Journal of Political Science* 57(4): 941–955.
- World Bank 2020.** *World Development Indicators Online*. Washington DC: World Bank. URL: <http://data.worldbank.org/indicator/>.
- Younas J., Sandler T. 2017.** Gender Imbalance and Terrorism in Developing Countries. *Journal of Conflict Resolution* 61(3): 483–510.
- Young J. K., Dugan L. 2011.** Veto players and terror. *Journal of Peace Research* 48(1): 19–33.