

ПОСТРОЕНИЕ ПРЕДСКАЗАТЕЛЬНОГО ИНДЕКСА РЕВОЛЮЦИЙ: ОПЫТ ИСПОЛЬЗОВАНИЯ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ*

Илья Александрович Медведев

Национальный исследовательский университет «Высшая школа экономики»

Андрей Витальевич Кортаев

Национальный исследовательский университет «Высшая школа экономики»;
Московский государственный университет имени М. В. Ломоносова

В статье представлена попытка прогноза революционных событий на основе анализа целого ряда факторов, разделенных по нескольким блокам: социальные, экономические, политические, демографические и так далее. В качестве основного метода анализа был выбран подход с использованием машинного обучения. Новый подход позволил авторам обойти множество проблем, с которыми сталкивались предыдущие исследователи в этой области, и использовать большое количество переменных для анализа. В статье революционные события разделяются на два типа – вооруженные и невооруженные, и для каждого из типов революционных событий строится отдельная предсказательная модель и, соответственно, свой индекс революционных событий.

Введение

За последние несколько лет задача прогнозирования в социальных науках значительно трансформировалась за счет стремительного открытия новых методов анализа. Область применения новых методов достаточно обширна, но основные направления новых исследований можно разделить на два основных вида: развитие в сфере анализа и обработки текстов, и анализ большого объема статистической информации (Bohr, Dunlap 2018; Molina, Garip 2019; Mutzel

* Работа выполнена в рамках Программы развития Междисциплинарной научно-образовательной школы Московского университета «Математические методы анализа сложных систем» при поддержке Российского научного фонда (проект № 18-18-00254).

2015; Keuschnigg *et al.* 2018). Поскольку нас интересуют именно методы, связанные с прогнозированием, то фокус нашей работы будет направлен преимущественно на новые методы, связанные с работой со статистической информацией. Некоторые авторы выделяют два этапа прорыва в работе со статистическими данными. В первую очередь это прорыв 80-х гг. XX в., который связан с активным внедрением методов регрессионного, кластерного и факторного анализа в социальные науки, вторым же прорывом называют то, что происходит с 2015 г. и продолжается до сих пор (McFarland *et al.* 2016; Benbouzid 2018; Williamson 2021).

Данный прорыв связан преимущественно с работой в области больших объемов информации, их последовательной и структурной обработкой, получением из них важной новой информации. В нашем исследовании мы воспользуемся некоторыми новыми методами в области машинного обучения, которые позволили бы нам работать с большим количеством показателей и нестандартными типами распределения зависимых переменных. Такие преимущества новых методов являются для нас принципиальными, поскольку задача построения индексов нестабильности обычно связана с оценкой большого количества факторов. Большие организации, оценивающие риски дестабилизации в разных странах, часто прибегают к оценкам экспертов и прочим методам снижения размерности данных (Коротаев, Медведев и др. 2018, 2020; Коротаев, Айсин и др. 2020). Наш же подход отличается тем, что в своих моделях мы не будем использовать методы по снижению размерности данных, что позволит учесть все показатели стран, без потери информации при агрегировании.

Начиная с середины прошлого века, происходит сильный всплеск интереса к изучению нестабильности в различных регионах мира. Активно изучаются причины, виды, последствия нестабильности (см., например: Goldstone *et al.* 2003, 2010; Lawson 2019; Коротаев и др. 2017). Было отмечено, что проявления нестабильности, особенно в больших странах, можно наблюдать практически каждый год (Ortiz *et al.* 2013); но в то время как, например, наличие одного протеста не говорит о наличии реальной нестабильности в стране, наличие террористических атак может уже быть достаточно значимым для любой страны. Таким образом, возникла необходимость в квантификации различных типов нестабильности, чтобы

иметь возможность отделять одни факты политического насилия от других, разделяя их по степени опасности и последствиям. Одной из самых серьезных попыток в квантификации актуальной нестабильности является работа создателей базы CNTS (*Cross-National Time Series Database*), разделивших нестабильность на 8 различных категорий (Banks, Wilson 2020). Используемая авторами CNTS классификация основана на актуальной литературе по изучению нестабильности. В то же время, авторы использовали лишь эмпирические критерии по квантификации нестабильности.

В нашем же исследовании перед нами стоит задача по составлению индекса нестабильности, отражающего не просто наличие и количество тех или иных событий, но и интенсивность нестабильности, которая имеет одинаковую природу и последствия. Поскольку составление индекса не может происходить в отрыве от реально наблюдаемых событий, недостаточно просто составить индекс, который отражал бы нестабильность постфактум, но необходимо иметь возможность предсказывать наличие нестабильности в будущем, основываясь на актуальных данных. Таким образом, перед нами стоит сразу две задачи. В первую очередь, это составление индекса актуальной нестабильности, который мог бы отражать нестабильность, имеющую одинаковые основания и последствия. Во-вторых, это составление индекса, учитывающего актуальные характеристики страны, для прогнозирования будущей нестабильности.

Методология построения индекса

Задача построения индексов нестабильности не нова и уже реализуется многими исследователями и исследовательскими организациями. Большинство из них используют методы математического анализа для задачи предсказания нестабильности. Их модели преимущественно построены на небольшом числе очень емких переменных, которые отражают целый ряд характеристик развития страны (см., например: Акаева и др. 2013; Гринин и др. 2014; Гринин, Коротаев 2009; Коротаев, Божевольнов и др. 2011; Коротаев, Малков и др. 2012; Коротаев, Ходунов и др. 2012; Esty *et al.* 1998; Alesina *et al.* 2003; Goldstone 2001; 2002; 2011a; 2011b; 2014a; 2014b; Goldstone *et al.* 1991, 2003, 2010; Gurr 1968; Korotayev, Khaltourina 2006; Zinkina, Korotayev 2014a; 2014b; Turchin,

Korotayev 2006; Mesquida, Weiner 1999; Moller 1968; Chambers 2012; Owen 2017; Slinko *et al.* 2017; Pickard 2018). Такой подход плох преимущественно тем, что хотя мы и используем ограниченный набор очень сильных переменных, мы в то же время теряем значительную часть информации о стране и процессе ее развития.

Существует и иные подходы к построению индекса нестабильности. Помимо чисто математического подхода, о котором мы говорили, существует и подход с использованием группировок и экспертных оценок. Здесь стоит вспомнить наши предыдущие работы по анализу различных индексов социально-политической нестабильности (Коротаев, Медведев и др. 2018; 2020). Основным отличием таких подходов является составление индекса на основе создания некоторого количества субиндексов, которые впоследствии взвешиваются для построения финального индекса. Например, такой подход использует *State Fragility Index (SFI)*, созданный Университетом Джорджа Мейсона и Центром системного мира (*Center for Systemic Peace*) (Marshall, Elzinga-Marshall 2017; 2018). Так, они выделяют 4 субиндекса: Индекс безопасности, Политический индекс, Экономический индекс и Социальный индекс. При построении этих индексов они используют набор показателей, которые оцениваются исходя из распределения основных переменных. Таким образом, они получают набор оценок, которые потом обсуждаются и корректируются экспертами. Данный индекс является показательным примером построения индекса нестабильности на основе категориальных и экспертных оценок. Таким же подходом пользуются, например, создатели *Country Indicators for Foreign Policy: Failed and Fragile States* (Country Indicators for Foreign Policy 2006).

Существует также и третий подход к составлению индекса нестабильности – метод комбинированных оценок. Данный подход характеризуется тем, что в нем используется как блок математического моделирования, так и блок экспертных оценок. В качестве результатов математического моделирования, как правило, используют набор предсказаний ближайшей нестабильности на основе анализа нескольких блоков переменных. Далее полученные результаты используются в финальной математической модели, куда в качестве дополнительных переменных добавляются и экспертные оценки по каждому кейсу и проводится результирующий анализ,

который включает в себя как математический анализ различного набора переменных, так и экспертную оценку. Ярким представителем такого подхода является *Fragile States Index* (Fund for Peace 2017; 2019a; 2019b).

Мы же в своей работе не будем использовать ни один из существующих подходов по построению индексов нестабильности, а предложим собственный способ построения такового. Во многом, мы все же будем придерживаться генеральной линии, которую используют индексы, построенные на математическом анализе, однако привнесем несколько существенных отличий. В первую очередь, мы откажемся от фактического разделения переменных на те или иные субиндексы. Все переменные, которые мы будем использовать в нашем анализе, будут использованы в рамках математической модели одновременно. Это позволит нам отразить в нашей модели кросс-взаимодействия переменных из разных категорий, чего не позволяет сделать ни один из существующих подходов. В то же время, мы не будем использовать экспертные оценки. Использование экспертных оценок обусловлено в первую очередь тем, что в рамках предыдущих подходов, мы не имели возможности оценить значимое количество переменных, а экспертная оценка позволяла соединить в себе множество различных не учитываемых показателей. Однако экспертные оценки нестабильны, а сделать их универсальными для каждой страны не представляется возможным, особенно на большом временном отрезке. Таким образом, мы приняли решение не использовать уже существующие подходы, а использовать современные методы математического анализа, которые позволяют если не избавиться, то значительно нивелировать многие ограничения, связанные с предыдущими подходами.

Для задачи построения индекса мы воспользуемся методами машинного обучения, которые уже на протяжении большого периода времени используются для предсказания различных показателей, в том числе покупательского спроса, покупательских предпочтений, логистики закупок и так далее (Gammerman, Vovk 2007; Bohanec *et al.* 2017). Данные модели хорошо себя показали в качестве средств прогноза, в первую очередь, из-за ряда существенных математических улучшений по сравнению, например, с регрессионным моделированием. При построении индекса мы воспользовались моделями «машинного обучения с учителем» (*supervised*

learning methods), предполагающими, что мы уже имеем предсказываемый результат (в нашем случае, наличие революционных событий), и задача модели состоит в том, чтобы подобрать оптимальные параметры и коэффициенты для всех переменных, чтобы наиболее точно предсказывать результирующую переменную.

Здесь стоит отдельно остановиться на принципе работы нашей модели. Мы будем использовать модели машинного обучения, построенные на основе градиентного бустинга. Градиентный бустинг, в свою очередь, построен на основе моделей случайного леса (*random forest*), которые, в свою очередь, построены на основе модели дерева решений (*decision tree*). (Grönlund *et al.* 2020; Lundberg *et al.* 2020). Модель дерева решений имеет в своем основании принцип разбиения принятия решений на последовательность принятых решений на основании набора переменных (см. Рис. 1).

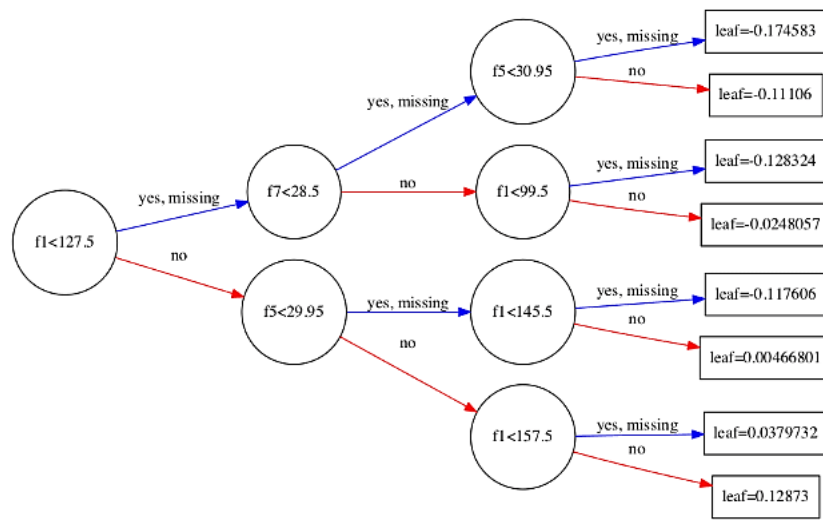


Рис. 1. Визуализация дерева принятия решений

Как мы можем видеть, внутри круга находится условие, а стрелками разных цветов указывается принцип перехода от одного условия к другому. В конечном итоге, мы получим большое дерево, концом которого является так называемый лист (*leaf*), указывающий нам на результат, к которому будет стремиться предсказанное значение переменной. Важно отметить, что на Рис. 1 представлено

ограниченное дерево, построенное на основе 7 переменных, из которых в данном дереве участвуют только 3 переменные (Brownlee 2016). В то же время, для наших задач такая блок-схема была бы значительно обширнее и включала в себя множество переменных. Однако такой подход очень универсализирован и не дает нам должного уровня специфичности. В первую очередь, это связано с тем, что в рамках одного дерева не всегда имеется возможность использовать все переменные, так как модель подбирает такое дерево, чтобы полученный результат имел наименьшее значение функции ошибок. В то же время, некоторые специфические кейсы плохо обрабатывались такой моделью и плохо предсказывались (Mentch, Zhou 2019). Исследователи начали использовать ансамбли решающих деревьев. Данная идея заключалась в построении сразу нескольких решающих деревьев, с разным набором предсказывающих условий и используемых переменных (Kashyap 2019) – см. Рис. 2.

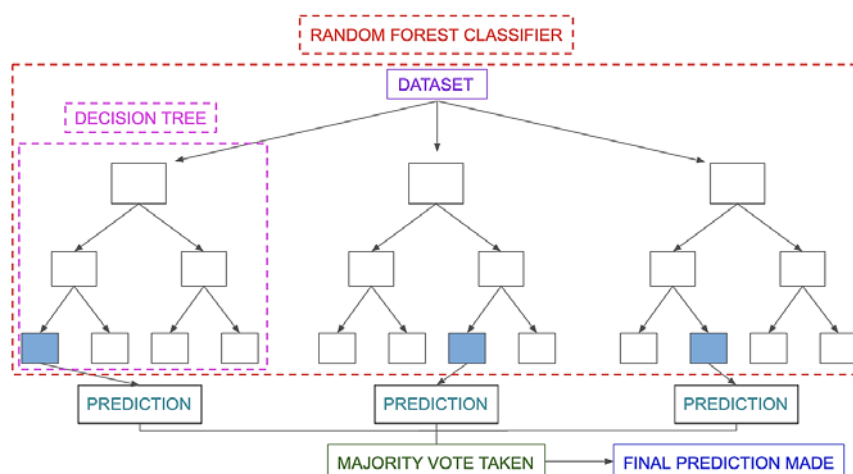


Рис. 2. Визуализация модели случайного леса

Далее, каждое из построенных деревьев дает свое предсказание относительно результирующей переменной, а финальный результат выбирается на основе выбора наиболее часто предсказанной категории. Такой подход дает возможность учитывать различные условия и сочетания переменных при построении предсказывающей модели.

Градиентный бустинг же берет за основу идею модели случайного леса, но значимо модернизирует его. Само название «градиентный бустинг» такие модели получили за счет применения методов по анализу и построению различных решающих деревьев. Если в модели случайного леса каждое новое дерево строилось случайным образом, то для моделей градиентного бустинга исследователи применили метод поиска параметров по градиентному спуску (Bakouregui *et al.* 2021) – см. Рис. 3.

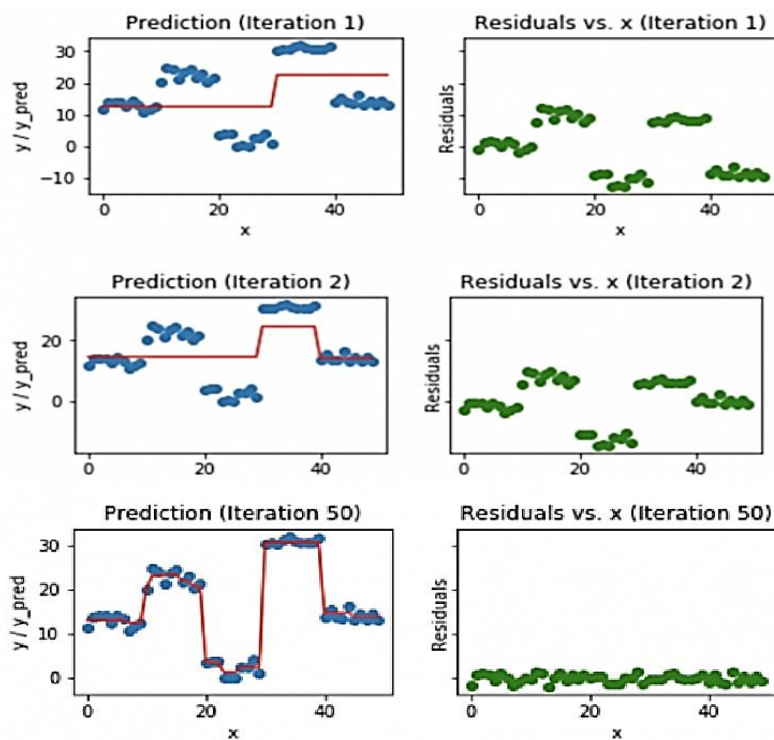


Рис. 3. Визуализации работы градиентного бустинга на первой, второй и пятидесятой итерации (слева изначальные данные и функция модели, справа график ошибок)

Этот подход предполагает, что каждое новое дерево оценивается по отношению к предыдущему на основе сетки параметров. И каждое новое дерево дает нам все более и более точное предсказание.

Такой подход позволяет значительно улучшить качество предсказания результирующей переменной, устраняя многие ошибки, которые делали предыдущие модели. Однако существует несколько видов моделей машинного обучения, которые представлены различными исследователями. Далее мы определимся с тем, какую модель будем использовать для решения наших задач на основе анализа их обучения и функции ошибок.

Данные и выбор модели

Перед тем, как заняться выбором модели машинного обучения и описанием набора данных для обучения, стоит описать нашу основную зависимую переменную. Отдельно стоит отметить работу Э. Ченовет и Дж. Ульфельдера, где анализировались различные причины появления невооруженных дестабилизационных событий (Chenoweth, Ulfelder 2017). Для нас эта работа важна преимущественно своим подходом к определению невооруженных революционных событий. Авторы выделяют набор характеристик, которые отражают принадлежность события к ненасильственной революции:

- Не менее 1000 видимых участников, активно участвующих в конфликте.
- Участники совершают несколько действий, наблюдаемых в согласованной последовательности (то есть не разовое событие, а серия событий, наблюдаемых в течение одной или более недель).
- Общая кампания длится более одной недели.
- Действующие лица заявляют о максималистской цели. К максималистским целям Э. Ченовет относит свержение режима, изгнание оккупационной армии, отделение или самоопределение.
- Основным методом достижения цели является невооруженный протест.
- Протестующие участвуют в неинституциональных, как правило, незаконных актах гражданского неповиновения, например, сидячие забастовки, бойкоты и акты отказа от сотрудничества.

В свою очередь, для наших задач мы используем в качестве основы подход Ченовет и Ульфельдера, однако модернизируем его. Наибольшее изменение заключается в том, что мы будем рассматривать не только невооруженные протесты, но также и вооруженные, но будем рассматривать их отдельно. Мы также добавили

данные о революциях в XXI в. в изначальную базу данных. Тут стоит отметить, что количество таких случаев достаточно мало. Посмотрим, для примера, на количество невооруженных революций в нашем датасете (см. Рис. 4).



Рис. 4. Количество невооруженных революций в общей базе данных

Хотя наши модели и методы их оценки нечувствительны к пропорциональности датасета, все же такая диспропорция может значительно сказаться на нашем результате. Также отметим, что задача предсказания революции в конкретный год очень сложна, так как триггером для подобного рода дестабилизации могут выступить самые разные события, если оно произойдет на фоне глубоких структурных проблем в обществе (см., например: Голдстоун 2015). Многие исследователи отмечают, что задача предсказать революцию в конкретный год очень сложна, однако предсказать революцию на определенном временном отрезке – задача намного более реализуемая (Asongu, Nwachukwu 2016). В своей работе мы воспользуемся подходом Э. Эмануэла и З. Обермайера, которые предсказывали вероятность болезней с помощью методов градиентного бустинга (Obermeier, Emanuel 2016). Для обучения модели авторы расширили временной этап проявления болезни с определенного

года на период в 3 года, механически добавляя случаи на год до и год после нахождения болезни. Таким образом авторам удалось значительно улучшить предсказательную силу своей модели. Мы воспользуемся аналогичным подходом, однако расширим временной период на 5 лет: 2 года до революции и 2 года после революции. Таким образом, мы сможем предсказывать не конкретный год революции, а вероятность возникновения революции в течение 5 лет. Однако, чтобы в нашу прогностическую модель не проникли косвенные данные о революции из других переменных, то возьмем изначально нашу зависимую переменную с лагом в 1 год. Так мы будем указывать наличие революции за год и в последующие 3 года после революции. Посмотрим на характер нашего датасета после внесения таких изменений (см. Рис. 5).



Рис. 5. Количество невооруженных революций в общей базе данных с интерполяцией данных на 5-летний период

Так, мы преодолели пятипроцентный порог зависимой переменной для нашего датасета. Это позволит модели обучаться более правильно без значимых искажений и случайных выбросов данных.

Для нашего исследования мы воспользуемся несколькими базами данных, агрегирующими различные социально-политические

показатели. В качестве основной базы данных мы использовали базу *Cross National Time Series Database* (Banks, Wilson 2021). Эта база данных включает в себя более чем 150 показателей по 200 странам за период с 1815 по 2020 год. Дополнительно мы использовали данные МВФ (IMF 2019), Всемирного банка (World Bank 2020) и др. В конечном итоге мы будем использовать более 240 показателей. Анализировать такое количество переменных крайне трудно и не входит в задачи этой статьи, однако мы воспользуемся подходом составителей предыдущих индексов при интерпретации полученных результатов, а именно, мы будем анализировать не отдельные переменные, а блоки наших зависимых переменных, которые мы разобьем по нескольким категориям:

- экономические переменные;
- демографические переменные;
- политические переменные;
- социальные переменные.

Далее для каждого из предсказаний мы будем вычислять вклад каждой из используемых переменных в модель и суммировать этот вклад в рамках блоков и далее будем интерпретировать их. Перейдем далее к выбору нашей модели машинного обучения.

Воспользуемся нашим подходом к выбору моделей, который мы описывали в работе по анализу факторов революционных событий (Медведев, Коротаев 2021). Мы будем тестировать несколько подходов, чтобы определить наиболее релевантную модель. Для этого мы будем оценивать модели по ряду параметров. В первую очередь, мы будем оценивать модели, изолируя 25 % нашего датасета в качестве тестовой выборки, и будем проверять данные, основываясь на трех метриках (ROC-AUC, Logloss, Accuracy). Обучение модели приостанавливалось на основе оценки AUC-score. Выбор этих метрик обусловлен характером нашей зависимой переменной, а именно бинарной переменной. Во-вторых, мы будем проводить кросс-валидацию наших моделей. Кросс-валидация необходима, чтобы исключить эффект случайности разбиения исходной выборки на тренировочную и тестовую. Принцип кросс-валидации заключается в последовательном разделении моделей на N блоков, где каждый из блоков поочередно будет становиться тестовой выборкой, а остальные будут использоваться в качестве обучающей. Данные по метрике AUC будут суммироваться и де-

литься на количество прогнанных моделей. Таким образом будет получен средний показатель AUC по всему изначальному датасету. В-третьих, мы будем производить так называемый поиск по сетке или *GridSearch*. Этот метод предполагает, что мы будем инициализировать модели не только с изначальными параметрами обучения, но будем использовать несколько вариантов обучения модели для поиска наиболее оптимального.

Посмотрим теперь более детально на результаты наших моделей (см. Табл. 1).

Табл. 2. Сравнительная таблица AUC-score различных моделей

<i>Название модели</i>	<i>Модель с валидацией</i>	<i>Модель с кросс-валидацией</i>	<i>Модель с кросс-валидацией и GridSearch</i>
Decision Tree	0,56	0,57	0,56
Random Forest	0,6	0,7	0,73
XGBoost	0,58	0,63	0,81
Catboost	0,77	0,83	0,86

Как мы можем видеть, модель градиентного бустинга от компании Яндекс подходит нам наилучшим образом на основе данных тестового использования моделей. Также отметим, что модель *Catboost* обладает значимыми достоинствами, в частности, значимым качеством модели даже без использования поиска по сетке параметров. Более того, данная модель хорошо работает с небольшим размером датасета, а наш датасет с более чем 19 тысячами наблюдений является достаточно малым для задачи машинного обучения (Dorogush, Ershov, Gulin 2018; Prokhorenkova *et al.* 2018). Посмотрим теперь на процесс обучения модели с валидацией (см. Рис. 6).

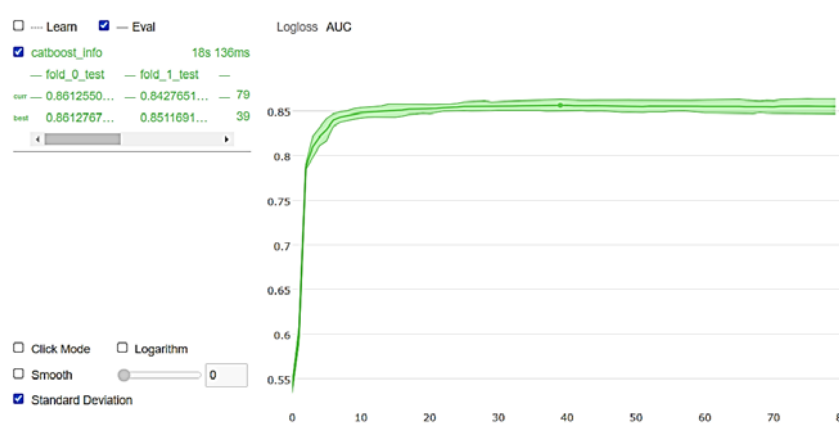


Рис. 6. Процесс обучения модели Catboost с кросс-валидацией

Как мы можем видеть, модель обучалась равномерно, без значимых искажений. Мы также можем видеть, что модель не обучилась слишком рано или слишком поздно, что говорит о хорошем подборе параметров при обучении. После того, как мы определились с нашей зависимой переменной, определили наш датасет, способ его анализа и используемую модель машинного обучения, перейдем непосредственно к анализу полученных результатов.

В наших предыдущих работах мы уже обращали внимание, что различные макрзоны нестабильности могут обладать собственными уникальными предикторами, вызывающими нестабильность (Коротаев и др. 2015; Коротаев, Гринин и др. 2021а: 266–289; 2021б: 199–247; Медведев, Коротаев 2019а; 2019б; 2020). Особенно эти различия значимы для афразийской макрзоны. Стоит отметить, что данный регион является одним из наиболее нестабильных. В связи с этим, в нашей работе мы будем также анализировать отдельно предсказания для данного региона. Это также поможет нам проверить адекватность работы нашей предсказательной модели.

Анализ вероятности начала невооруженных революционных выступлений

Для начала обратимся к результатам наших предсказаний для невооруженных революционных выступлений. Посмотрим для нача-

ла результаты по миру и проанализируем адекватность предсказанных нами значений. Ниже приведены страны, имеющие наибольший риск невооруженных революционных выступлений в ближайшие 5 лет (см. Табл. 2):

Табл. 2. Топ стран мира, имеющих наибольший риск начала невооруженных революционных выступлений в ближайшие пять лет (от наибольшей предсказанной нестабильности к меньшей)

<i>Название страны</i>	<i>Предсказанный индекс риска начала невооруженных революционных выступлений</i>
Ливан	24,36
Гондурас	18,64
Алжир	9,55
Боливия	5,35
Черногория	3,64
Албания	2,09
Ирак	1,9
Чили	1,34
Израиль	1,31
Гаити	1,22
Непал	0,78
Египет	0,61
Судан	0,6
Франция	0,56
Мали	0,52
Гватемала	0,5
Марокко	0,5
Иран	0,46
Индия	0,46

Окончание Табл. 2

<i>Название страны</i>	<i>Предсказанный индекс риска начала невооруженных революционных выступлений</i>
Мьянма	0,43
Колумбия	0,43
Афганистан	0,43
Доминиканская Республика	0,35
Индонезия	0,35
Узбекистан	0,34

Примечания: отметим, что применительно к странам мир-системного центра речь идет о риске начала аналогов революции или революционных движений без революции (таких, как движение «Желтых жилетов» во Франции). Отметим также, что в процессе обучения нашего датасета мы отрезали все данные начиная с 2019 г., а предсказания строились на основе данных за 2020 г., которые в обучении модели не участвовали вовсе.

Обратим внимание, что на первую строчку в нашем рейтинге попал Ливан, где уже действительно много лет наблюдается нестабильность. В том числе, говорят и о вероятности гражданской войны (Lenta 2020). Также в топ стран с высокой предсказанной нестабильностью попал, например, Афганистан, где революция (правда, вооруженная) уже фактически совершилась в 2021 г.; проведенный анализ позволяет предположить, что за ней может последовать попытка еще одной революции – на этот раз невооруженной (анти-талибской?). В целом же не вызывает удивления (ср.: Grinin, Korotaev 2022), что в списке стран с наибольшими рисками начала невооруженных революционных выступлений абсолютно доминируют страны мир-системной полупериферии, а представителей как мир-системного ядра, так и мир-системной периферии там довольно мало.

Обратимся теперь к основным показателям, которые влияли на такой результат нашей модели (см. Табл. 3).

Табл. 3. Оценка групп переменных по уровню кумулятивной значимости в модели

<i>Группа переменных</i>	<i>Кумулятивная значимость</i>
Экономические переменные	27,3 %
Демографические переменные	13,1 %
Политические переменные	25,6 %
Социальные переменные	34,0 %

Заметим, что наиболее сильной группой факторов возникновения невооруженных революционных выступлений оказались социальные и экономические переменные. В то же время, демографический фактор при возникновении невооруженных революционных выступлений оказался не таким значимым. Это также соотносится с теми результатами, что мы получали ранее (Коротаев, Гринин и др. 2021а: 266–289; 2021б: 199–247; Медведев, Коротаев 2019а; 2019б; 2020). Отметим, что в группу социальных факторов попадали такие показатели, как уровень угнетения отдельных слоев населения, индекс свободы вероисповедания и так далее. Однако стоит также проверить результаты нашей модели и на отдельном макро-регионе. Посмотрим результаты предсказания модели для афразийской макрзоны нестабильности (см. Табл. 4).

Табл. 4. Топ стран, имеющих наибольший риск начала невооруженных революционных выступлений в ближайшие 5 лет (от наибольшей предсказанной нестабильности к меньшей)

<i>Название страны</i>	<i>Предсказанный индекс риска начала невооруженных революционных выступлений</i>
Ливан	24,36
Алжир	9,56
Ирак	1,91
Израиль	1,32
Египет	0,62

Продолжение Табл. 4

<i>Название страны</i>	<i>Предсказанный индекс риска начала вооруженных революционных выступлений</i>
Судан	0,60
Мали	0,53
Марокко	0,50
Иран	0,46
Афганистан	0,43
Узбекистан	0,34
Киргизия	0,33
Пакистан	0,29
Кот-д'Ивуар	0,25
Сенегал	0,25
Центрально-Африканская Республика	0,22
Того	0,20
Буркина-Фасо	0,20
Грузия	0,19
Сирия	0,15
Либерия	0,14
Казахстан	0,13
Тунис	0,13
Палестина	0,12
Гамбия	0,11
Турция	0,11
Гана	0,10
Нигерия	0,10
Армения	0,09

Окончание Табл. 4

<i>Название страны</i>	<i>Предсказанный индекс риска начала вооруженных революционных выступлений</i>
Иордания	0,08
Гвинея-Бисау	0,07
Бенин	0,05
Гвинея	0,05
Мавритания	0,05
Нигер	0,04
Кувейт	0,03
Азербайджан	0,03
Чад	0,03
Бахрейн	0,02
Южный Судан	0,02
Оман	0,02
Джибути	0,02
Эритрея	0,02
Йемен	0,02
Сомали	0,01
Таджикистан	0,01
Туркменистан	0,01
Сьерра-Леоне	0,01
Саудовская Аравия	0,01
Катар	0,01
Объединенные Арабские Эмираты	0,01
Ливия	0,01

Заметим, что в топ стран действительно попали те страны, где в последние несколько лет наблюдается крайне высокий уровень нестабильности. Отметим, что самые первые строки здесь в основном занимают страны со средними (для макрозоны) значениями подушевого ВВП и частично демократическими режимами. В то же время, отдельно стоит обратить внимание на самые нижние строчки нашего рейтинга. Там преобладают очень богатые страны с монархическими полностью авторитарными режимами (а в разоренной гражданской войной Ливии сейчас, похоже, тоже не до «ненасильственных» революций). Кроме того, в число наиболее стабильных стран, имеющих наименьший риск вооруженных революционных выступлений, попали достаточно маленькие страны, что также говорит об адекватности использованных нами моделей.

Анализ результатов по предсказанию вооруженных революционных выступлений

Обратимся теперь к результатам наших предсказаний для вооруженных революционных выступлений. Посмотрим для начала результаты по миру и проанализируем адекватность предсказанных нами значений. Ниже приведен перечень стран с наибольшей вероятностью начала новых вооруженных революционных выступлений в ближайшие 5 лет (см. Табл. 5).

Табл. 5. Топ стран мира с наибольшей вероятностью начала вооруженных революционных выступлений в ближайшие 5 лет (от наибольшей предсказанной нестабильности к наименьшей)

<i>Название страны</i>	<i>Предсказанный индекс нестабильности</i>
Индия	3,22
Ангола	2,31
Чад	2,31
Нигер	2,15
Бурунди	2,07
Южный Судан	1,96

Окончание Табл. 5

Название страны	Предсказанный индекс неустойчивости
Мозамбик	1,95
Афганистан	1,92
Сирия	1,89
Пакистан	1,83
Йемен	1,82
Центральноафриканская Республика	1,79
Сомали	1,78
Ирак	1,73
Египет	1,68
Мьянма	1,64
Мали	1,58
Бангладеш	1,56
Ливия	1,50

Заметим, что в топ стран, наиболее подверженных вооруженным революционным выступлениям, попали другие страны, отличающиеся от наших предыдущих результатов. Интересно отметить, что в топ стран с наивысшими рисками опять попала Индия, которая также оказывалась на вершинах рейтинга в предыдущем году. Это также неудивительно, поскольку Индия уже длительное время демонстрирует заметные проявления неустойчивости (Gettleman and Abi-Nabib 2019; BBC 2019). Нигер, Чад и Мали также не удивляют своим попаданием в топ-20 наиболее неустойчивых стран. В этих регионах уже длительное время состояние стран можно охарактеризовать как *failed state* – недееспособное государство (Bassou, Ouassif 2019). В целом же, в списке стран с наибольшими рисками начала вооруженных революционных выступлений абсолютно доминируют не страны мир-системной полупериферии со средними подушевыми доходами (как это было применительно к «ненасильственным» революциям), а именно периферийные страны с низки-

ми подушевыми доходами, высокой рождаемостью, очень молодым крайне малообразованным слабо урбанизированным населением (ср. Коротаев, Устюжанин 2021).

Обратимся теперь и к анализу групп переменных, которые внесли наибольший вклад в предсказания (см. Табл. 6).

Табл. 6. Оценка групп переменных по уровню кумулятивной значимости в модели

<i>Группа переменных</i>	<i>Кумулятивная значимость</i>
Экономические переменные	26,1 %
Демографические переменные	19,9 %
Политические переменные	36,1 %
Социальные переменные	17,9 %

Отметим, что и относительный вклад групп переменных сильно изменился. В первую очередь, отметим, что на первое место вышли переменные, которые отражают характер актуального типа режима в стране. Очень значимыми оказались и экономические показатели. Социальные переменные же, напротив, потеряли значимую долю важности в модели. Демографические переменные незначительно увеличили значимость, но не так сильно, как другие переменные.

Обратимся, наконец, к списку стран из афразийской макрзоны нестабильности (см. Табл. 7).

Табл. 7. Страны афразийской макрзоны с наибольшей вероятностью начала вооруженных революционных выступлений в ближайшие 5 лет (от наибольшей предсказанной нестабильности к наименьшей)

<i>Название страны</i>	<i>Предсказанный индекс нестабильности</i>
Чад	2,31
Нигер	2,15
Южный Судан	1,96
Афганистан	1,92
Сирия	1,89
Пакистан	1,83

Продолжение Табл. 7

<i>Название страны</i>	<i>Предсказанный индекс нестабильности</i>
Йемен	1,82
Центрально-Африканская Республика	1,79
Сомали	1,78
Ирак	1,73
Египет	1,68
Мали	1,58
Ливия	1,50
Иран	1,45
Ливан	1,35
Сенегал	1,28
Буркина-Фасо	1,24
Нигерия	1,23
Турция	1,12
Саудовская Аравия	1,06
Тунис	1,05
Гвинея-Бисау	1,02
Алжир	1,00
Мавритания	0,97
Таджикистан	0,96
Того	0,94
Судан	0,90
Гвинея	0,90
Гана	0,89
Туркменистан	0,87
Азербайджан	0,87
Узбекистан	0,84

Окончание Табл. 7

<i>Название страны</i>	<i>Предсказанный индекс неустойчивости</i>
Эритрея	0,83
Кот-д'Ивуар	0,83
Палестина	0,81
Кувейт	0,80
Иордания	0,80
Казахстан	0,77
Либерия	0,73
Сьерра-Леоне	0,72
Бахрейн	0,69
Гамбия	0,68
Марокко	0,68
Кыргызстан	0,67
Бенин	0,67
Катар	0,60
Джибути	0,55
Оман	0,54
Израиль	0,54
Армения	0,52
Объединенные Арабские Эмираты	0,43
Грузия	0,28

Говоря об отличиях от предыдущей переменной, заметим, что значительно ниже в списке находится Грузия. Кыргызстан и Израиль также значительно опустились в списке. В то же время Афганистан оказался значительно выше чем в предыдущий раз. Результаты также соотносятся со группировкой переменных по кумулятивной значимости – действительно, страны с очень нестабильной политической и экономической обстановкой вышли на первые строчки нашего индекса, в то время как такие страны, как Грузия, где про-

блемы имеют скорее социальный характер, опустились вниз в нашем рейтинге. Отметим также, что для очень богатых стран макрозоны с монархическими режимами характерны низкие риски как вооруженной, так и невооруженной революционной дестабилизации.

В целом же, среди «африканских стран» с наибольшими рисками начала вооруженных революционных выступлений доминируют не страны со средними (для макрозоны) значениями подушевого ВВП и частично демократическими режимами (как это было применительно к «ненасильственным» революциям), а (так же, как и для мира в целом) страны с крайне низкими подушевыми доходами, высокой рождаемостью, очень молодым крайне малообразованным, слабо урбанизированным населением и гибридными или авторитарными режимами (ср.: Коротаев, Устюжанин 2021).

Заключение

Подводя итог исследования, мы можем говорить о том, что наша попытка использовать новый подход к построению индексов революционных событий оказалась сравнительно успешной. В первую очередь, стоит сказать, что использование современных методов анализа данных позволяет устранить множество проблем, с которыми сталкивались предыдущие исследователи. В частности, нам удалось использовать множество переменных в рамках одной модели, не опасаясь проблемы мультиколлинеарности. Мы также удачно протестировали подход по предсказанию нестабильности не на конкретный год, а на временной период, что дало вполне осмысленные результаты. Анализ топа стран показывает, что в списки наиболее нестабильных стран действительно попадают те случаи, где многие исследователи также ожидают появления революций. Также модель хорошо показала себя, определяя страны с недееспособными государствами.

В целом, удалось показать, что наибольшая вероятность начала *невооруженных* революционных выступлений в странах африканской макрозоны нестабильности характерна для стран со средними (для макрозоны) значениями подушевого ВВП и частично демократическими режимами. В то же самое время, наибольшая вероятность начала *невооруженных* революционных выступлений в странах данной макрозоны характерна для стран с крайне низкими по-

душевыми доходами, высокой рождаемостью, очень молодым крайне малообразованным, слабо урбанизированным населением и гибридными или авторитарными режимами.

Библиография

- Акаева Б. А., Коротаев А. В., Исаев Л. М., Шишкина А. Р. (ред.). 2013.** *Системный мониторинг глобальных и региональных рисков. Центральная Азия: новые вызовы.* М.: Ленанд/URSS.
- Голдстоун Дж. А. 2015.** *Революции. Очень краткое введение.* М.: Изд-во Ин-та Гайдара.
- Гринин Л. Е., Коротаев А. В. 2009.** Урбанизация и политическая нестабильность: к разработке математических моделей политических процессов. *Полис. Политические исследования* 4: 34–52.
- Гринин Л. Е., Коротаев А. В., Исаев Л. М., Шишкина А. Р. 2014.** Риски дестабилизации в контексте нарастающей неопределенности в «афразийской» зоне. *Системный мониторинг глобальных и региональных рисков. Центральная Азия: новые вызовы* / Отв. ред. Б. А. Акаев, А. В. Коротаев, Л. М. Исаев, А. Р. Шишкина. М.: Ленанд/URSS. С. 4–10.
- Коротаев А. В., Айсин М. Б., Зинькина Ю. В., Романов Д. М. 2020.** К прогнозированию структурно-демографических рисков социально-политической дестабилизации в странах Афразийской макрзоны нестабильности и Африки южнее Сахеля. *Системный мониторинг глобальных и региональных рисков* / Отв. ред. Л. Е. Гринин, А. В. Коротаев, Д. А. Быканова. М.: Моск. ред. изд-ва «Учитель». С. 200–238.
- Коротаев А. В., Божевольнов Ю. В., Гринин Л. Е., Зинькина Ю. В., Малков С. Ю. 2011.** Ловушка на выходе из ловушки. Логические и математические модели. *Проекты и риски будущего. Концепции, модели, инструменты, прогнозы* / Ред. А. А. Акаев, А. В. Коротаев, Г. Г. Малинецкий, С. Ю. Малков. М.: Красанд/URSS. С. 138–164.
- Коротаев А. В., Гринин Л. Е., Исаев Л. М., Билюга С. Э., Васькин И. А., Слинько Е. В., Шишкина А. Р., Мещерина К. В. 2017.** *Дестабилизация: глобальные, национальные, природные факторы и механизмы.* М.: Моск. ред. изд-ва «Учитель».
- Коротаев А. В., Гринин Л. Е., Малков С. Ю., Исаев Л. М., Билюга С. Э., Шишкина А. Р., Иванов Е. А., Мещерина К. В., Васькин И. А., Медведев И. А., Романов Д. М. 2021а.** *Катализаторы политических переворотов: от акций протеста к смене власти. Анализ и моделирование процессов дестабилизации на примере модернизирующихся социально-политических систем.* М.: Ленанд/URSS.

- Коротаев А. В., Гринин Л. Е., Малков С. Ю., Исаев Л. М., Филин Н. А., Билюга С. Э., Зинькина Ю. В., Слинько Е. В., Шишкина А. Р., Шульгин С. Г., Мещерина К. В., Айсин М. Б., Иванов Е. А., Кокликов В. О., Медведев И. А., Романов Д. М., Слав М., Сойер П. С. 2021б.** *Социально-политическая дестабилизация в странах афразийской макрзоны нестабильности: количественный анализ и прогнозирование рисков.* М.: Ленанд/URSS.
- Коротаев А. В., Исаев Л. М., Руденко М. А. 2015.** Формирование афразийской зоны нестабильности. *Восток. Афро-Азиатские общества: история и современность* 2: 88–99.
- Коротаев А. В., Малков С. Ю., Бурова А. Н., Зинькина Ю. В., Ходунов А. С. 2012.** Ловушка на выходе из ловушки. Математическое моделирование социально-политической дестабилизации в странах мир-системной периферии и события Арабской весны 2011 г. *Моделирование и прогнозирование глобального, регионального и национального развития* / Ред. А. А. Акаев, А. В. Коротаев, Г. Г. Малинецкий, С. Ю. Малков. М.: ЛИБРОКОМ/URSS. С. 210–276.
- 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.
- Коротаев А. В., Медведев И. А., Шульгин С. Г., Слинько Е. В., Билюга С. Э., Малков С. Ю., Халтурина Д. А., Дербилова Е. В., Зинькина Ю. В., Романов Д. М. 2018.** Системы глобального мониторинга рисков социально-политической дестабилизации: опыт систематического обзора. *Системный мониторинг глобальных и региональных рисков. Социально-политическая и экономическая дестабилизация: анализ страновых и региональных ситуаций в мир-системном аспекте* / Отв. ред. Л. Е. Гринин, А. В. Коротаев, К. В. Мещерина. Волгоград: Учитель. С. 5–94.
- Коротаев А. В., Устюжанин В. В. 2021.** О структурно-демографических факторах вооруженных исламистских революционных выступлений в странах группы G5 Сахель. *Системный мониторинг глобальных и региональных рисков: ежегодник. Т. 12: Революционные процессы в афразийской макрзоне нестабильности и их глобальный контекст* / Отв. ред. Л. Е. Гринин, А. В. Коротаев, Д. А. Быканова. Волгоград: Учитель.
- Коротаев А. В., Ходунов А. С., Бурова А. Н., Малков С. Ю., Халтурина Д. А., Зинькина Ю. В. 2012.** Социально-демографический анализ Арабской весны. *Системный мониторинг глобальных и регио-*

- нальных рисков: ежегодник / Отв. ред. А. В. Коротаев, Ю. В. Зинькина, А. С. Ходунов. М.: Ленанд/URSS С. 28–76.
- Медведев И. А., Коротаев А. В. 2019а.** К использованию методов машинного обучения для ранжирования факторов дестабилизации в странах мир-системной периферии. *Системный мониторинг глобальных и региональных рисков: ежегодник* / Отв. ред: Л. Е. Гринин, А. В. Коротаев, К. В. Мещерина. Волгоград: Учитель. С. 227–242.
- Медведев И. А., Коротаев А. В. 2019б.** К построению индекса социально-политической дестабилизации в различных мир-системных зонах. *Системный мониторинг глобальных и региональных рисков: ежегодник* / Отв. ред. Л. Е. Гринин, А. В. Коротаев, К. В. Мещерина. Волгоград: Учитель. С. 120–136.
- Медведев И. А., Коротаев А. В. 2020.** К построению индекса социально-политической дестабилизации в различных мир-системных зонах (Ч. 2). *Системный мониторинг глобальных и региональных рисков: ежегодник* / Отв. ред. Л. Е. Гринин, А. В. Коротаев, Д. А. Быканова. Волгоград: Учитель. С. 436–457.
- Медведев И. А., Коротаев А. В. 2021.** Структурные факторы мирной и вооруженной революционной смены власти: опыт анализа методами машинного обучения. *Системный мониторинг глобальных и региональных рисков. Революционные процессы в африканской макрорегионе нестабильности и их глобальный контекст* / Отв. ред. Л. Е. Гринин, А. В. Коротаев, Д. А. Быканова. Волгоград: Учитель.
- Alesina A., Devleeschauwer A., Easterly W., Kurlat S., Wacziarg R. 2003.** Fractionalization. *Journal of Economic Growth* 8(2): 155–194.
- Asongu, S. A., Nwachukwu, J. C. 2016.** Revolution Empirics: Predicting the Arab Spring. *Empirical Economics* 51(2): 439–482.
- Bakouregui A. S., Mohamed H. M., Yahia A., Benmokrane B. 2021.** Explainable Extreme Gradient Boosting Tree-based Prediction of Load-carrying Capacity of FRP-RC Columns. *Engineering Structures* 245: <https://doi.org/10.1016/j.engstruct.2021.112836>.
- Banks A. S., Wilson K. A. 2021.** *Cross-National Time-Series Data Archive*. Jerusalem: Databanks International. URL: <https://www.cntsdata.com/>.
- Bassou A., el Ouassif A. 2019.** Understanding Terrorism and Organized Crime in Light of Fragile States: Case Study on Niger, Mali and Chad. *Evolving Human Security Challenges in the Atlantic Space* / Ed. by N. Severiano Teixeira, D. Marcos. Washington, DC: Brookings Institution Press. Pp. 23–40.

- BBC News. 2019.** Six Killed on Deadly Day of India Protests. *BBC News* 11.01. URL: <https://www.bbc.com/news/world-asia-india-50872900>.
- Benbouzid B. 2018.** Values and Consequences in Predictive Machine Evaluation. A Sociology of Predictive Policing. *Science & Technology Studies* 31.
- Bohanec M., Borštnar M. K., Robnik-Šikonja M. 2017.** Explaining machine learning models in sales predictions. *Expert Systems with Applications* 71: 416–428.
- Bohr J., Dunlap R. E. 2018.** Key Topics in environmental sociology, 1990–2014: results from a computational text analysis. *Environmental Sociology* 4(2): 181–195.
- Brownlee J. 2016.** How to Visualize Gradient Boosting Decision Trees With XGBoost in Python. *Machine Learning Mastery*. URL: <https://machinelearningmastery.com/visualize-gradient-boosting-decision-trees-xgboost-python/>.
- Chambers R. 2012.** *Revolutions in Development Inquiry*. London: Routledge.
- Chenoweth E., Ulfelder J. 2017.** Can structural conditions explain the onset of nonviolent uprisings? *Journal of Conflict Resolution* 61(2): 298–324.
- Country Indicators for Foreign Policy. 2006.** *Failed and Fragile States Project*. Ottawa: Carleton University. URL: <https://carleton.ca/cifp/wp-content/uploads/1145.pdf>.
- Dorogush A. V., Ershov V., Gulin A. 2018.** *CatBoost: Gradient Boosting with Categorical Features Support*. URL: <http://arxiv.org/abs/1810.11363>.
- Esty D., Goldstone J. A., Gurr T. R., Harff B., Levy M., Dabelko G. D., Surko P., Unger A. N. 1998.** State Failure Task Force Report: Phase II Findings. *Environmental Change & Security Project Report* 5: 49–72.
- Fund for Peace. 2017.** *Fragile States Index methodology and cast framework*. Washington, DC: The Fund for Peace. URL: <https://fragilestatesindex.org/wp-content/uploads/2017/05/FSI-Methodology.pdf>.
- Fund for Peace. 2019a.** *Fragile States Index Annual Report 2019*. Washington, DC: The Fund for Peace. URL: <https://fundforpeace.org/wp-content/uploads/2019/04/9511904/fragilestatesindex.pdf>.
- Fund for Peace. 2019b.** *Fragile States Index Database*. Washington, DC: The Fund for Peace. URL: <https://fragilestatesindex.org/excel/>.
- Gamerman A., Vovk V. 2007.** Hedging predictions in machine learning. *The Computer Journal* 50(2): 151–163.
- Gettleman J., Abi-Habib M. 2019.** As Protests Rage on Citizenship Bill, is India Becoming a Hindu Nation? *The New York Times* 16.12. URL:

- <https://www.nytimes.com/2019/12/16/world/asia/india-citizenship-protests.html>.
- Goldstone J. 2001.** Toward a Fourth Generation of Revolutionary Theory. *Annual Review of Political Science* 4: 139–187.
- Goldstone J. 2002.** Population and Security: How Demographic Change Can Lead to Violent Conflict. *Journal of International Affairs* 56(1): 3–21.
- Goldstone J. 2011a.** Cross-class Coalitions and the Making of the Arab Revolts of 2011. *Swiss Political Science Review* 17(4): 457–462.
- Goldstone J. 2011b.** Understanding the Revolutions of 2011: Weakness and Resilience in Middle Eastern Autocracies. *Foreign Affairs* 90(3): 8–16.
- Goldstone J. 2014a.** Protests in Ukraine, Thailand and Venezuela: What Unites Them? *Russia Direct* 21: 46–53.
- Goldstone J. 2014b.** *Revolutions: A Very Short Introduction*. Oxford: Oxford University Press.
- Goldstone J., Bates R., Epstein D., Gurr T., Lustik M., Marshall M., Ulfelder J., Woodward M. 2010.** A Global Model for Forecasting Political Instability. *American Journal of Political Science* 54(1): 190–208.
- Goldstone J., Gurr T. R., Moshiri F. (eds.). 1991.** *Revolutions of the Late Twentieth Century*. Boulder, CO: Westview Press.
- Goldstone J., Gurr T., Harff B., Levy M., Marshall M., Bates R., Epstein D., Kahl C., Surko P., Ulfelder J., Unger Jr. A. 2003.** *State Failure Task Force Report: Phase III Findings*. McLean, VA: Science Applications International Corporation (SAIC).
- Grinin L., Korotayev A. 2022.** Revolutions, Counterrevolutions, and Democracy. *Handbook of Revolutions in the 21st Century: The New Waves of Revolutions, and the Causes and Effects of Disruptive Political Change* / Ed. by J. Goldstone, L. Grinin, A. Korotayev. Cham: Springer. URL: https://doi.org/10.1007/978-3-030-86468-2_4.
- Grønlund A., Kamra L., Larsen K. G. 2020.** Margins are insufficient for explaining gradient boosting. *arXiv preprint arXiv:2011.04998*.
- Gurr T. R. 1968.** A Causal Model of Civil Strife: A Comparative Analysis Using New Indices. *American Political Science Review* 62: 1104–1124.
- International Monetary Fund = IMF. 2019.** *World Economic Outlook: Global Manufacturing Downturn, Rising Trade Barriers*. Washington, D.C.: IMF.
- Kashyap K. 2019.** Machine Learning- Decision Trees and Random Forest Classifiers. *Medium*. URL: <https://medium.com/analytics-vidhya/machine-learning-decision-trees-and-random-forest-classifiers-81422887a544>.

- Keuschnigg M., Lovsjö N., Hedström P. 2018.** Analytical Sociology and Computational Social Science. *Journal of Computational Social Science* 1(1): 3–14.
- Korotayev A., Khaltourina D. 2006.** *Introduction to Social Macrodynamics: Secular Cycles and Millennial Trends in Africa*. Moscow: KomKniga, URSS.
- Lawson G. 2019.** *Anatomies of revolution*. Cambridge: Cambridge University Press.
- Lenta.ru 2020.** Взрывная волна. Ливан на грани гражданской войны: как взрыв в порту стал детонатором народного протеста. *Lenta.ru* 24.08. URL: <https://lenta.ru/articles/2020/08/24/lebanon/>.
- Lundberg S. M., Erion G., Chen H., DeGrave A., Prutkin J. M., Nair B., Lee S. I. 2020.** From local explanations to global understanding with explainable AI for trees. *Nature machine intelligence* 2(1): 56–67.
- Marshall M. G., Elzinga-Marshall G. 2017.** *Global Report 2017: Conflict, Governance, and State Fragility*. Vienna: Center for Systemic Peace. URL: <http://www.systemic-peace.org/vlibrary/GlobalReport2017.pdf>.
- Marshall M. G., Elzinga-Marshall G. 2018.** *State Fragility Index and Matrix 2018*. Vienna: Center for Systemic Peace. URL: <http://www.systemicpeace.org/inscr/SFMatrix2018c.pdf>.
- McFarland D. A., Lewis K., Goldberg A. 2016.** Sociology in the Era of Big Data: The Ascent of Forensic Social Science. *The American Sociologist* 47(1): 12–35.
- Mentch L., Zhou S. 2019.** Randomization as Regularization: a Degrees of Freedom Explanation for Random Forest Success. arXiv preprint. *arXiv*: 1911.00190.
- Mesquida C. G., Weiner N. I. 1999.** Male Age Composition and Severity of Conflicts. *Politics and the Life Sciences* 18: 113–117.
- Molina M., Garip F. 2019.** Machine learning for sociology. *Annual Review of Sociology* 45: 27–45.
- Moller H. 1968.** Youth as a Force in the Modern World. *Comparative Studies in Society and History* 10: 238–260.
- Mützel S. 2015.** Facing big data: Making Sociology Relevant. *Big Data & Society* 2(2): <https://doi.org/10.1177/2053951715599179>.
- Obermeyer Z., Emanuel E. J. 2016.** Predicting the Future – Big Data, Machine Learning, and Clinical Medicine. *The New England Journal of Medicine* 375(13): 1216.

- Ortiz I., Burke S., Berrada M, Cortes H. 2013.** *World Protests 2006–2013*. Rochester, NY: Social Science Research Network.
- Owen S. 2017.** Monitoring Social Media and Protest Movements: Ensuring Political Order through Surveillance and Surveillance Discourse. *Social Identities* 23(6): 688–700.
- Pickard S. 2018.** Governing, Monitoring and Regulating Youth Protest in Contemporary Britain. *Governing Youth Politics in the Age of Surveillance* / Ed. by M. T. Grasso, J. L. Bessant. London: Routledge. Pp. 93–106.
- Prokhorenkova L., Gusev G., Vorobev A., Dorogush A. V., Gulin A. 2018.** CatBoost: Unbiased Boosting with Categorical Features. *Advances in neural information processing systems* 31. URL: <https://papers.nips.cc/paper/2018/hash/14491b756b3a51daac41c24863285549-Abstract.html>.
- Slinko E., Bilyuga S., Zinkina J., Korotayev A. 2017.** Regime Type and Political Destabilization in Cross-National Perspective: A Re-Analysis. *Cross-Cultural Research* 51(1): 26–50.
- Turchin P., Korotayev A. 2006.** Population Density and Warfare: A Reconsideration. *Social Evolution & History* 5(2): 121–158.
- Williamson B. 2021.** Digital Policy Sociology: Software and Science in Data-intensive Precision Education. *Critical Studies in Education* 62(3): 354–370.
- World Bank. 2020.** *World Development Indicators Database*. URL: <https://data.worldbank.org/indicator>.
- Zinkina J., Korotayev A. 2014a.** Explosive Population Growth in Tropical Africa: Crucial Omission in Development Forecasts (Emerging Risks and Way Out). *World Futures* 70(4): 271–305.
- Zinkina J., Korotayev A. 2014b.** Projecting Mozambique’s Demographic Futures. *Journal of Futures Studies* 19(2): 21–40.