

ПРОГНОЗИРОВАНИЕ ИНДЕКСОВ СОЦИАЛЬНО-ПОЛИТИЧЕСКОЙ НЕСТАБИЛЬНОСТИ С ИСПОЛЬЗОВАНИЕМ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ*

Сергей Георгиевич Шульгин

Российская академия народного хозяйства и государственной службы при Президенте РФ;
Национальный исследовательский университет
«Высшая школа экономики»

В статье мы анализируем возможность использования методов машинного обучения для краткосрочного предсказания, как агрегированного индекса социально-политической нестабильности, так и отдельных индексов социально-политической дестабилизации. Анализ мы проводили на основе обновленной расширенной базы данных, которую использовали для анализа факторов, в наибольшей степени влияющих на социально-политическую нестабильность. Мы предсказывали различные показатели нестабильности на $(i+1)$ -й год с использованием данных, доступных на i -й год. В отличие от анализа факторов, наиболее влияющих на нестабильность, когда из независимых факторов исключаются все возможные альтернативные оценки нестабильности (альтернативные зависимые переменные), для целей предсказания на будущее мы используем все доступные данные, как по зависимым, так и по оценкам нестабильности на текущий момент (включая значения предсказываемого индекса за текущий и предыдущие периоды). Для анализа факторов, которые влияют на будущую нестабильность, мы также оценили класс моделей, предсказывающих нестабильность без учета информации по текущим уровням нестабильности (как анализируемого индекса, так и его альтернативным оценкам). Предварительные результаты показывают возможность предсказания на следующий период (на один год вперед) для индекса «Политические убийства» с точностью 0,08 и AUC =

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

0,80, индекса «*Политические забастовки*» с точностью 0,08 и $AUC = 0,83$, индекса «*Партизанские действия*» с точностью 0,09 и $AUC = 0,90$, индекса «*Правительственные кризисы*» с точностью 0,12 и $AUC = 0,80$, индекса «*Политические репрессии*» с точностью 0,09 и $AUC = 0,88$, индекса «*Массовые беспорядки*» с точностью 0,16 и $AUC = 0,84$, индекса «*Государственные перевороты и попытки переворотов*» с точностью 0,106 и $AUC = 0,87$, индекса «*Антиправительственные демонстрации*» с точностью 0,16 и $AUC = 0,84$, агрегированного индекса социально-политической дестабилизации на год вперед с точностью 0,21 и $AUC = 0,86$. Однако требуется дополнительный анализ для исключения возможностей «утечки» информации – эффекта, связанного с тем, что в момент предсказания все данные уже известны и оценки предсказанного прошлого были сделаны в момент проведения анализа, что может привести к тому, что наши оценки будущего будут обладать существенно худшими предсказательными возможностями, чем ретроспективные предсказания прошлого.

Модель

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

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

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

В такой формулировке наша задача 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 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:

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

где $q(X)$ – описывает дерево, вершинами которого являются правила относительно значений X . Функция $q(X)$ ставит в соответствие для определенной точки 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 значениями (y_i) 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).

Данные

В качестве исходных данных о нестабильности мы используем данные *Cross National Time Series (CNTS)*, *Global Terrorism Database (GTB)* и базы данных государственных переворотов.

База данных *The Cross National Time Series (CNTS)* – это результат работы по сбору и систематизации данных, начатой Артуром Банксом (Banks, Wilson 2018) в 1968 г. в Университете штата Нью-Йорк в Бингемтоне, обобщения архива данных *The Statesman's Yearbook*, публикуемых с 1864 г. В базе содержатся данные по более чем 200 странам и годовые значения переменных начиная с 1815 г., за исключением периодов двух мировых войн 1914–1918 и 1940–1945 гг.

В данной работе мы используем в качестве зависимых переменных данные, описывающие различные аспекты внутренних конфликтов (domestic). Эти данные получены из анализа страновых событий по 8 различным подкатегориям:

- Политические убийства (*Assassinations, domestic1*).
- Политические забастовки (*General Strikes, domestic2*).
- Партизанские действия (*Guerrilla Warfare, domestic3*).

- Правительственные кризисы (*Government Crises*, domestic4).
- Политические репрессии (*Purges*, domestic5).
- Массовые беспорядки (*Riots*, domestic6).
- перевороты и попытки переворотов (*Revolutions*, domestic7).
- Антиправительственные демонстрации (*Anti-Government Demonstrations*, domestic8).

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

К политическим забастовкам (*General Strikes*, domestic2) относятся забастовки, в которых участвовало 1000 или более работников, более одного работодателя и при этом выдвигались требования, направленные против национальной политики, правительства или органов власти.

К партизанским действиям (*Guerrilla Warfare*, domestic3) относится любая вооруженная деятельность, диверсии или взрывы, совершаемые независимыми группами граждан или нерегулярными вооруженными силами, которые направлены на свержение нынешнего режима.

К правительственным кризисам (*Government Crises*, domestic4) относятся любые ситуации, которые грозят привести к падению текущего режима – за исключением вооруженных переворотов, напрямую направленных на это.

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

К массовым беспорядкам (*Riots*, domestic6) относятся любые демонстрации или столкновения, связанные с использованием насилия, в которых принимали участие более 100 граждан.

К переворотам и попыткам переворотов (*Revolutions*, domestic7) относятся любые незаконные или связанные с принуждением изменения в правящей элите, а также любые попытки таких изменений. Переменная «Перевороты и попытки переворотов» также учитывает все удачные и неудачные вооруженные восстания, целью которых является получение независимости от центрального правительства.

К антиправительственным демонстрациям (*Anti-Government Demonstrations*, domestic8) относятся любые мирные публичные собрания, в которых принимает участие 100 человек и более, а в качестве основной цели проведения является выражение несогласия с политикой правительства или власти за исключением демонстраций с выраженной направленностью против иностранных государств.

Все перечисленные 8 подкатегорий используются при построении общего индекса социально-политической стабилизации (domestic9). Для этого составители базы данных *CNTS* присвоили каждой подкатегории определенный вес (см. Табл. 1).

Табл. 1. Веса подкатегорий, используемых при построении индекса социально-политической стабилизации

Подкатегория	Название переменной	Вес в индексе социально-политической стабилизации (domestic9)
Политические убийства (<i>Assassinations</i>)	cnts_domestic1	25
Политические забастовки (<i>General Strikes</i>)	cnts_domestic2	20
Партизанские действия (<i>Guerrilla Warfare</i>)	cnts_domestic3	100
Правительственные кризисы (<i>Government Crises</i>)	cnts_domestic4	20
Политические репрессии (<i>Purges</i>)	cnts_domestic5	20
Массовые беспорядки (<i>Riots</i>)	cnts_domestic6	25
Перевороты и попытки переворотов (<i>Revolutions</i>)	cnts_domestic7	150
Антиправительственные демонстрации (<i>Anti-Government Demonstrations</i>)	cnts_domestic8	10

Индекс социально-политической стабилизации (*Weighted Conflict Measure*, domestic9) рассчитывается по формуле (4):

$$domestic9 = \frac{\sum_{i=1}^8 w_i cnts_domestic_i}{8} * 100, \quad (4)$$

где w_i – веса, приведенные в последнем столбце Табл. 1.

Кроме показателя domestic9 для анализа мы построили переменную domestic9 с лагом (cnts_domestic9_prev), которая показыва-

ет общее значение страновой нестабильности в предыдущем году. Также мы построили упреждающую переменную (*cnts_domestic9_pext*) для оценки общего уровня нестабильности в будущем году.

Помимо данных *CNTS*, в качестве объясняемой переменной мы используем два индикатора из *Global Terrorism Database* (START 2016). Мы используем переменные:

n_terror_attack – количество террористических атак,

Nkill – количество убитых.

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

Из базы данных государственных переворотов (Marshall M., Marshall D. 2016) для независимых переменных мы взяли для анализа переменную:

coup_detat_failed_coup_detat – государственные перевороты и попытки переворотов (аналог переменной *cnts_domestic8*).

База данных государственных переворотов охватывает временной период с 1960 по 2016 г.

Всего в качестве зависимых (объясняемых, целевых) для данного анализа было отобрано 14 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. Для переменной *n_terror_attack* пороговым значением было выбрано $N=50$.

Оценка моделей

Для анализируемых переменных мы построили переменную, содержащую будущее (на следующий год) значение, и создали модель, в которой попытались предсказать это значение.

Мы построили модель, где в качестве независимых переменных использовали базу данных, из которой исключили всю информацию о состоянии зависимой переменной в текущий момент времени, а также все другие зависимые переменных, то есть переменные, которые мы классифицировали как наблюдаемые характеристики нестабильности. Мы предсказываем будущую нестабильность, как

если бы мы пытались анализировать, какие факторы влияют на текущую нестабильность, и пытались исключить любые проявления нестабильности. Мы пытались оценить значение нестабильности на будущий год по данным независимых переменных текущего года.

Для оценки (тренировки) модели градиентного бустинга нужно определиться с параметрами алгоритма. 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 используется в качестве тестовой, а остальные k-1 в качестве обучающей. Процедура оценки модели с этим разбиением повторяется k раз, так чтобы каждая из k-частей побывала тестовой выборкой. Мы выбрали k равным 5 и для каждой зависимой переменной провели оценку модели с помощью процедуры кросс-валидации 20 раз (каждый раз с новым разбиением на 5 случайных подвыборок). В результате для каждой из 12 зависимых переменных мы получили 100 оценок моделей, в каждой из которых оценивается значимость независимых факторов.

Параметр глубины деревьев (*max.depth*) эмпирическим путем выбран равным 5. Параметр скорости сходимости (*eta*) был выбран 0,15. Параметр *early_stopping_rounds* мы приняли за 20.

Для оценки качества оценивания мы использовали функцию ошибок, в которой оценивается доля полученных оценок, отличающихся от истинных (наблюдаемых) значений, и параметр *AUC* (*Area Under Curve*).

Табл. 2. Оценки качества 100 моделей на обучающих и тестовых выборках с использованием 302 независимых переменных

Переменная	Описание	Число факторов	Обучающая выборка (train)		Тестовая выборка (test)	
			Error	AUC	Error	AUC
cnts_domestic1_nxt	Политические убийства (Assassinations)	302	0,072 (0,006)	0,891 (0,050)	0,081 (0,006)	0,801 (0,026)
cnts_domestic2_nxt	Политические забастовки (<i>General Strikes</i>)	302	0,062 (0,010)	0,918 (0,057)	0,075 (0,006)	0,830 (0,027)
cnts_domestic3_nxt	Партизанские действия (<i>Guerrilla Warfare</i>)	302	0,057 (0,015)	0,975 (0,022)	0,093 (0,006)	0,903 (0,011)

Окончание Табл. 2

Переменная	Описание	Число факторов	Обучающая выборка (train)		Тестовая выборка (test)	
			Error	AUC	Error	AUC
cnts_domestic4_nxt	Правительственные кризисы (<i>Government Crises</i>)	302	0,102 (0,009)	0,915 (0,041)	0,123 (0,007)	0,796 (0,015)
cnts_domestic5_nxt	Политические репрессии (<i>Purges</i>)	302	0,072 (0,008)	0,942 (0,022)	0,087 (0,005)	0,878 (0,018)
cnts_domestic6_nxt	Массовые беспорядки (<i>Riots</i>)	302	0,119 (0,018)	0,922 (0,030)	0,155 (0,007)	0,836 (0,012)
cnts_domestic7_nxt	Перевороты и попытки переворотов (<i>Revolutions</i>)	302	0,074 (0,011)	0,961 (0,016)	0,106 (0,007)	0,866 (0,012)
cnts_domestic8_nxt	Антиправительственные демонстрации (<i>Anti-Government Demonstrations</i>)	302	0,113 (0,017)	0,932 (0,026)	0,155 (0,005)	0,842 (0,008)
cnts_domestic9_nxt	Индекс социально-политической стабилизации	302	0,144 (0,014)	0,934 (0,012)	0,211 (0,007)	0,856 (0,007)
coup_detat_failed_coup_detat_nxt	Государственные перевороты и попытки переворотов	302	0,026 (0,001)	0,932 (0,028)	0,028 (0,002)	0,860 (0,020)
n_terror_attack_nxt	Количество террористических атак	302	0,004 (0,004)	0,999 (0,001)	0,025 (0,004)	0,978 (0,007)
nkill_nxt	Количество убитых	302	0,066 (0,013)	0,980 (0,008)	0,121 (0,007)	0,921 (0,006)
cnts_domestic9_next_nxt	Индекс социально-политической стабилизации (на 2 года вперед)	302	0,154 (0,014)	0,926 (0,013)	0,221 (0,008)	0,844 (0,007)

Для оценки возможностей прогнозирования мы разбили выборку на 2 периода, период с 2015 г. (включительно) использовали в качестве проверочной выборки (validation), то есть не использовали эти данные ни в обучающей выборке, ни в тестовой, а использовали только для оценки качества прогнозирования.

Табл. 3. Оценки качества 100 моделей на обучающих, тестовых выборках с оценкой прогноза на проверочной выборке с использованием 302 независимых переменных

Переменная	Число факторов	Обучающая выборка (train)		Тестовая выборка (test)		Проверочная выборка (validation)	
		Error	AUC	Error	AUC	Error	AUC
cnts_domes tic1_nxt	302	0,071 (0,008)	0,899 (0,052)	0,082 (0,005)	0,801 (0,029)	0,056 (0,005)	0,795 (0,040)
cnts_domes tic2_nxt	302	0,061 (0,005)	0,888 (0,050)	0,070 (0,005)	0,813 (0,031)	0,169 (0,003)	0,727 (0,043)
cnts_domes tic3_nxt	302	0,056 (0,011)	0,977 (0,010)	0,092 (0,005)	0,900 (0,011)	0,181 (0,011)	0,895 (0,010)
cnts_domes tic4_nxt	302	0,100 (0,014)	0,916 (0,051)	0,124 (0,006)	0,798 (0,022)	0,085 (0,003)	0,632 (0,018)
cnts_domes tic5_nxt	302	0,072 (0,007)	0,938 (0,029)	0,087 (0,005)	0,878 (0,017)	0,098 (0,003)	0,759 (0,033)
cnts_domes tic6_nxt	302	0,122 (0,016)	0,901 (0,043)	0,151 (0,007)	0,818 (0,015)	0,326 (0,020)	0,806 (0,014)
cnts_domes tic7_nxt	302	0,079 (0,016)	0,955 (0,028)	0,112 (0,006)	0,857 (0,017)	0,008 (0,003)	0,829 (0,059)
cnts_domes tic8_nxt	302	0,116 (0,013)	0,919 (0,026)	0,154 (0,006)	0,824 (0,011)	0,410 (0,025)	0,808 (0,017)
cnts_domes tic9_nxt	302	0,146 (0,019)	0,931 (0,017)	0,217 (0,009)	0,848 (0,007)	0,233 (0,017)	0,883 (0,006)
coup_detat_ failed_coup_ detat_nxt	302	0,028 (0,001)	0,922 (0,030)	0,029 (0,003)	0,854 (0,019)	0,002 (0,000)	0,577 (0,224)
n_terror_ attack_nxt	302	0,005 (0,004)	0,999 (0,002)	0,023 (0,004)	0,978 (0,007)	0,078 (0,006)	0,923 (0,014)
nkill_nxt	302	0,066 (0,015)	0,980 (0,010)	0,124 (0,008)	0,917 (0,008)	0,197 (0,016)	0,837 (0,020)

Для анализа качества прогноза на проверочной (validation) мы сравниваем результаты полученного прогноза с тремя простейшими прогнозными сценариями: инерционный, нулевой, случайный (с учетом вероятности).

В инерционном сценарии прогноз на следующий интервал делается на основе сегодняшнего значения («завтра будет как вчера»). Существует нулевой сценарий, когда нигде не предсказываются нестабильные ситуации (всегда «0»), так как для многих переменных характерно «доминирование нулевых данных». При случайном прогнозе (с учетом вероятности) предсказываем случайное значение, с вероятностью нестабильности, пропорциональной наблюдаемой выборочной вероятности нестабильности.

Табл. 4. Оценка прогноза на проверочной выборке с использованием простейших прогнозных сценариев

Переменная	Инерционный		Нулевой сценарий		Случайный (с учетом вероятности)	
	Error	AUC	Error	AUC	Error	AUC
cnts_domestic1_nxt	0,076 (0,000)	0,645 (0,000)	0,051 (0,000)	0,500 (0,000)	0,096 (0,009)	0,499 (0,020)
cnts_domestic2_nxt	0,187 (0,000)	0,622 (0,000)	0,173 (0,000)	0,500 (0,000)	0,285 (0,014)	0,500 (0,020)
cnts_domestic3_nxt	0,121 (0,000)	0,828 (0,000)	0,232 (0,000)	0,500 (0,000)	0,358 (0,017)	0,498 (0,022)
cnts_domestic4_nxt	0,145 (0,000)	0,558 (0,000)	0,084 (0,000)	0,500 (0,000)	0,153 (0,011)	0,501 (0,020)
cnts_domestic5_nxt	0,111 (0,000)	0,677 (0,000)	0,098 (0,000)	0,500 (0,000)	0,174 (0,012)	0,501 (0,021)
cnts_domestic6_nxt	0,274 (0,000)	0,723 (0,000)	0,478 (0,000)	0,500 (0,000)	0,496 (0,023)	0,512 (0,019)
cnts_domestic7_nxt	0,017 (0,000)	0,495 (0,000)	0,007 (0,000)	0,500 (0,000)	0,014 (0,004)	0,502 (0,025)
cnts_domestic8_nxt	0,242 (0,000)	0,762 (0,000)	0,623 (0,000)	0,500 (0,000)	0,468 (0,021)	0,502 (0,022)
cnts_domestic9_nxt	0,184 (0,000)	0,792 (0,000)	0,746 (0,000)	0,500 (0,000)	0,378 (0,018)	0,502 (0,021)
coup_detat_failed_ coup_detat_nxt	0,007 (0,000)	0,497 (0,000)	0,002 (0,000)	0,500 (0,000)	0,003 (0,002)	0,499 (0,001)
n_terror_attack_nxt	0,064 (0,000)	0,916 (0,000)	0,092 (0,000)	0,500 (0,000)	0,168 (0,011)	0,499 (0,019)
nkill_nxt	0,214 (0,000)	0,789 (0,000)	0,256 (0,000)	0,500 (0,000)	0,383 (0,017)	0,498 (0,020)

Инерционный сценарий обладает высокой предсказательной силой, это свидетельствует, что для предсказания важны случаи с переключением режима во времени и можно пытаться анализировать

модель в разностях (и пытаться предсказать изменение, а не текущую категорию).

Для прогнозирования мы можем себе позволить использовать данные не только по независимым переменным, но и все данные, доступные на данный момент, включая текущее значение самой переменной. Прогностическая способность инерционного сценария как раз показывает, что текущее значение исключительно важно для предсказания.

Табл. 5. Оценки качества 100 моделей на обучающих, тестовых выборках с оценкой прогноза на проверочной выборке с использованием всех (373) зависимых и независимых переменных

Переменная	Число факторов	Обучающая выборка (train)		Тестовая выборка (test)		Проверочная выборка (validation)	
		Error	AUC	Error	AUC	Error	AUC
cnts_domes tic1_nxt	373	0,064 (0,005)	0,906 (0,049)	0,080 (0,005)	0,815 (0,025)	0,066 (0,009)	0,835 (0,029)
cnts_domes tic2_nxt	373	0,059 (0,004)	0,905 (0,040)	0,071 (0,004)	0,821 (0,023)	0,158 (0,006)	0,787 (0,018)
cnts_domes tic3_nxt	373	0,067 (0,007)	0,946 (0,019)	0,090 (0,006)	0,899 (0,011)	0,112 (0,007)	0,930 (0,009)
cnts_domes tic4_nxt	373	0,101 (0,011)	0,903 (0,041)	0,126 (0,005)	0,798 (0,018)	0,087 (0,003)	0,675 (0,024)
cnts_domes tic5_nxt	373	0,067 (0,006)	0,940 (0,026)	0,087 (0,005)	0,878 (0,015)	0,099 (0,004)	0,779 (0,042)
cnts_domes tic6_nxt	373	0,121 (0,010)	0,896 (0,026)	0,149 (0,006)	0,826 (0,010)	0,282 (0,011)	0,827 (0,007)
cnts_domes tic7_nxt	373	0,085 (0,007)	0,921 (0,022)	0,107 (0,006)	0,866 (0,012)	0,008 (0,002)	0,781 (0,049)
cnts_domes tic8_nxt	373	0,119 (0,011)	0,902 (0,028)	0,155 (0,006)	0,824 (0,011)	0,289 (0,019)	0,850 (0,008)
cnts_domes tic9_nxt	373	0,169 (0,015)	0,913 (0,015)	0,218 (0,008)	0,851 (0,007)	0,172 (0,010)	0,906 (0,004)
coup_detat_ failed_coup_ _detat_nxt	373	0,026 (0,001)	0,926 (0,027)	0,029 (0,003)	0,856 (0,021)	0,002 (0,001)	0,681 (0,228)
n_terror_att ack_nxt	373	0,008 (0,004)	0,992 (0,009)	0,024 (0,003)	0,974 (0,013)	0,066 (0,009)	0,962 (0,009)
nkill_nxt	373	0,084 (0,014)	0,966 (0,014)	0,124 (0,008)	0,918 (0,008)	0,169 (0,009)	0,891 (0,010)

Сопоставим качество прогнозов (на проверочной выборке), полученных с помощью простейшего инерционного сценария, модели, построенной только с использованием независимых переменных и модели, построенной с учетом текущих наблюдаемых переменных, описывающих нестабильность в текущий момент времени.

Табл. 6. Сопоставления качества прогнозов моделей с использованием всех зависимых и независимых переменных

Переменная	Инерционный		Прогноз по 302 независимым переменным		Прогноз по 373 независимым и зависимым переменным	
	Error	AUC	Error	AUC	Error	AUC
cnts_domestic1_nxt	0,076 (0,000)	0,645 (0,000)	0,056 (0,005)	0,795 (0,040)	0,066 (0,009)	0,835 (0,029)
cnts_domestic2_nxt	0,187 (0,000)	0,622 (0,000)	0,169 (0,003)	0,727 (0,043)	0,158 (0,006)	0,787 (0,018)
cnts_domestic3_nxt	0,121 (0,000)	0,828 (0,000)	0,181 (0,011)	0,895 (0,010)	0,112 (0,007)	0,930 (0,009)
cnts_domestic4_nxt	0,145 (0,000)	0,558 (0,000)	0,085 (0,003)	0,632 (0,018)	0,087 (0,003)	0,675 (0,024)
cnts_domestic5_nxt	0,111 (0,000)	0,677 (0,000)	0,098 (0,003)	0,759 (0,033)	0,099 (0,004)	0,779 (0,042)
cnts_domestic6_nxt	0,274 (0,000)	0,723 (0,000)	0,326 (0,020)	0,806 (0,014)	0,282 (0,011)	0,827 (0,007)
cnts_domestic7_nxt	0,017 (0,000)	0,495 (0,000)	0,008 (0,003)	0,829 (0,059)	0,008 (0,002)	0,781 (0,049)
cnts_domestic8_nxt	0,242 (0,000)	0,762 (0,000)	0,410 (0,025)	0,808 (0,017)	0,289 (0,019)	0,850 (0,008)
cnts_domestic9_nxt	0,184 (0,000)	0,792 (0,000)	0,233 (0,017)	0,883 (0,006)	0,172 (0,010)	0,906 (0,004)
coup_detat_failed_coup_detat_nxt	0,007 (0,000)	0,497 (0,000)	0,002 (0,000)	0,577 (0,224)	0,002 (0,001)	0,681 (0,228)
n_terror_attack_nxt	0,064 (0,000)	0,916 (0,000)	0,078 (0,006)	0,923 (0,014)	0,066 (0,009)	0,962 (0,009)
nkill_nxt	0,214 (0,000)	0,789 (0,000)	0,197 (0,016)	0,837 (0,020)	0,169 (0,009)	0,891 (0,010)

Из приведенных в Табл. 6 данных видно, что модели с учетом данных о текущем состоянии выигрывают по показателям качества прогнозирования и оба варианта моделей дают более качественные

предсказания, чем инерционный прогноз (завтра как вчера). Основной проблемой использования данных о зависимых переменных является тот факт, что в них может неявно содержаться «утечка» из будущего – оценки нестабильности для различных периодов могли делаться в один и тот же момент времени и в момент оценки информация о прошлом и будущем была в распоряжении автора этой оценки. При построении прогноза есть смысл исключать все возможные «утечки» из будущего и использовать максимально устойчивую к ним модель с учетом независимых переменных.

* * *

Предварительные результаты показывают возможность предсказания на следующий период (на один год вперед) для индекса «Политические убийства» с точностью 0,08 и AUC = 0,80, индекса «Политические забастовки» с точностью 0,08 и AUC = 0,83, индекса «Партизанские действия» с точностью 0,09 и AUC = 0,90, индекса «Правительственные кризисы» с точностью 0,12 и AUC = 0,80, индекса «Политические репрессии» с точностью 0,09 и AUC = 0,88, индекса «Массовые беспорядки» с точностью 0,16 и AUC = 0,84, индекса «Государственные перевороты и попытки переворотов» с точностью 0,106 и AUC = 0,87, индекса «Антиправительственные демонстрации» с точностью 0,16 и AUC = 0,84, агрегированного индекса социально-политической дестабилизации на год вперед с точностью 0,21 и AUC = 0,86. Однако требуется дополнительный анализ для исключения возможностей «утечки» информации – эффекта, связанного с тем, что в момент предсказания все данные уже известны и оценки предсказанного прошлого были сделаны в момент проведения анализа, что может привести к тому, что наши оценки будущего будут обладать существенно худшими предсказательными возможностями, чем ретроспективные предсказания прошлого.

Библиография

- Banks A. S., Wilson K. A. 2018.** *Cross-National Time-Series Data Archive. Databanks International.* Jerusalem, Israel. URL: <http://www.databanksinternational.com>.

- Chen T., Guestrin C. 2016.** Xgboost: A scalable tree boosting system. *Proceedings of the 22nd acmsigkdd international conference on knowledge discovery and data mining*. ACM. Pp. 785–794.
- Kuhn M. 2008.** Building Predictive Models in R Using the caret Package. *Journal of Statistical Software* 28(5): 1–26.
- Marshall M. G., Marshall D. R. 2016.** Coup D'État Events, 1946–2015. *Codebook*. Center for Systemic Peace.
- START** [National Consortium for the Study of Terrorism and Responses to Terrorism]. **2016.** *Global Terrorism Database*. College Park, MD: National Consortium for the Study of Terrorism and Responses to Terrorism. URL: <https://www.start.umd.edu/gtd/>.