

Применение технологий больших данных для отслеживания социально-политических и экономических ситуаций в регионах России

Галин Л.Р., магистрант кафедры «Прикладная информатика», Московский политехнический университет, Москва, Россия.

Рабинович А.Е., кандидат экономических наук, доцент кафедры «Прикладная информатика», Московский политехнический университет, Москва, Россия.

Аннотация. В статье рассматривается проблема исследования и отслеживания настроения в обществе, а также социально-политических и экономических ситуаций в регионах Российской Федерации. Описываются типичные проблемы регионов и предлагаются методики их решения. По этим методикам предлагается модель анализа данных и отслеживания социально-политических и экономических ситуаций в регионах России. На основании результатов анализа предлагаются рекомендации по принятию решений в проблемных регионах РФ.

Ключевые слова: социально-политическая, экономическая ситуация, большие данные, анализ данных, анализ тональности, корреляционный анализ, дерево решений.

Application of big data technologies for tracking socio-political and economic situations in Russian regions

Galin L.R., magistrate of the «Applied Informatics» Department at the Moscow polytechnic University, Moscow, Russia

Rabinovich AE, candidate of economic sciences, Associate Professor of the «Applied Informatics» Department at the Moscow polytechnic University, Moscow, Russia

Annotation. The article deals with the problem of studying and tracking the mood in society, as well as socio-political and economic situations in the regions of the Russian Federation. Typical problems of regions are described and methods of their solution are offered. According to these methods, a model of data analysis and tracking of socio-political and economic situations in the regions of Russia is proposed. Based on the results of the analysis, recommendations for decision-making in problem regions of the Russian Federation are proposed.

Keywords: socio-political, economic situation, big data, data analysis, sentiment analysis, correlation analysis, decision tree.

Введение

Российская Федерация – большая и сложная в управлении страна, не только из-за обширной территории, но и вследствие больших объемов информации, которую необходимо постоянно собирать, хранить, обрабатывать и принимать решения на основе ее анализа. Так, по данным Федеральной службы государственной статистики в современной России в 85 регионах проживают около 146 миллионов людей [2] и в каждом регионе есть свой спектр проблем, которые могут быть не видны из федерального центра. В связи с этим возникает большой интерес к исследованиям региональных проблем, не только со стороны научного сообщества, но и со стороны органов государственной власти. Всё это обуславливает необходимость своевременного мониторинга и прогнозирования региональных проблем. Кроме того, решение проблем в регионах в значительной степени определяет развитие страны. Также своевременный анализ данных поможет прогнозировать и предотвращать нежелательные ситуации в регионах. От того, как федеральные и региональные власти реагируют на изменяющуюся внешнюю и внутреннюю среду, зависит уровень развития региона и Российской Федерации в целом.

Для предотвращения большинства проблем в регионах, для планирования и реализации стратегии развития России, целесообразно использовать анализ той информации, которую оставляют о себе граждане, пользующиеся

разнообразными информационными ресурсами (цифровой след). А применение методов анализа больших данных позволит прогнозировать и предотвращать большинство угроз в регионах.

В настоящее время проводится множество исследований, посвящённых новым способам отслеживания социально-политических и экономических ситуаций в регионах России [10], но, несмотря на это, проблема некоторой напряженности в регионах до сих пор остается нерешенной. Многие исследователи считают, что без взаимодействия с губернаторами, главами администрации субъектов Российской Федерации невозможно получить объективную картину и отслеживать социально-политические и экономические ситуации регионов РФ [7]. Тем не менее, во многих регионах проблемы возникают как раз из-за вышестоящих лиц. Для снижения доли субъективности, в данной работе не будут привлекаться мнения отдельных лиц, а будет использоваться анализ настроения граждан по их же запросам в сети Интернет.

Российская Федерация является многонациональной страной, и она интегрирована в мировую экономику. Поэтому в эпоху глобального кризиса многие мировые проблемы отражаются и на России. Например, проблема экологии, миграции, безработицы, проблема межнациональных отношений также имеют место и в Российской Федерации [1]. Все эти проблемы усложняют управление регионами и страной в целом. Из-за этого может расти недовольство граждан в отдельных регионах, которое потом может трансформироваться в широкомасштабные акции (митинги) и протестное голосование. Чаще всего в России выездные проверки осуществляются в связи с уже произошедшими событиями или при возникновении сложных ситуаций в регионах, и все начинают работать только тогда, когда уже что-то произошло. Это явно свидетельствует о том, что анализ и прогнозирование социально-политических ситуаций в регионах России работает плохо, либо не работает вообще. Поэтому актуальность исследования, направленного на анализ и отслеживание социально-политических и экономических ситуаций в регионах РФ трудно переоценить.

По данным Росстата 95% жителей России пользуются мобильными телефонами или смартфонами, и 81% пользователей выходят в Сеть ежедневно, почти половина населения (49%) зарегистрированы в социальных сетях и активно ими пользуются [5]. Благодаря всему этому накапливаются большие данные, которые требуют специальных методик извлечения, хранения, обработки и анализа и интерпретации. Источниками этих данных являются в том числе социальные сети и интернет-платформы, где пользователи добровольно оставляют цифровые отпечатки и подробную информацию не только о себе, но и об окружающей их среде. Применяя технологии Big Data всех к этим данным, которые создаются можно проводить их анализ и отслеживать неблагоприятные ситуации в регионах РФ.

Постановка задачи исследования

На основании вышеизложенного ставится задача: выявить зависимость между настроениями жителей регионов (через содержание запросов в справочные правовые системы и информационно-поисковые системы) и социально-политической и экономической обстановкой в соответствующих регионах РФ.

Решение поставленной задачи сводится к выполнению ряда действий:

1. Определение множеств данных (Data Set) для анализа.
2. Сбор данных (запросов жителей). *Производится сбор исходной информации (данных).*
3. Исследование данных. *Определение исходной выборки для анализа.*
4. Предобработка исходных данных. *Производится подготовка данных к их последующему использованию. А именно, удаляются не нужные (не используемые) поля в исходной выборке, данные преобразуются в более подходящий для обработки формат, а также производится повышение качества данных.*
5. Анализ тональности (эмоциональной окраски) исходной выборки. *Используя технологию машинного обучения, производится обучение сети на основе данных, полученных в предыдущих этапах.*

6. Оценка настроения граждан по результатам анализа тональности.
7. Определение проблемных регионов по настроению граждан.
8. Сбор информации о проблемных регионах Российской Федерации.
9. Оценка социально-политических и экономических ситуаций в проблемных регионах России.
10. Выработка рекомендаций по мероприятиям в проблемных регионах.

Методы решения поставленной задачи

Описание структуры данных. Данные для исследования были взяты из открытых источников.

Data Set 1. Запросы жителей в справочно-правовые системы (СПС). Данные о содержаниях запросов жителей (то, что они искали, чем они интересовались) можно извлекать с сайтов справочно-правовых систем. В этих файлах хранится большое количество информации, но в анализе будут использоваться только наиболее информативные поля, а именно:

1. Текст обращения.
2. Регион.

Data Set 2. Данные о проблемных регионах, доступные в открытых источниках сети Интернет. Данное множество будет определять следующий набор полей: Регион; Численность населения; Средний возраст (средний возраст населения по регионам РФ); Медианная зарплата; Средняя зарплата; Прожиточный минимум (в регионах на душу населения); Бедные (Численность населения с денежными доходами ниже величины регионального прожиточного минимума, в процентах от общей численности населения субъекта РФ); Преступлений (Общее количество преступлений, зарегистрированных за год, на 10 000 жителей региона); Экономические преступления (Количество преступлений экономической направленности на 10 000 жителей региона); Численность врачей (Численность врачей, оказывающих медицинские услуги населению, на 10 000 человек региона); Смертность (Число умерших, на 1000 человек в год); Общий прирост населения (Общий прирост на 1000 человек за год); Прирост в городах (Общий прирост

населения в города на 1000 человек за год); Прирост в сёлах (Общий прирост населения на 1000 человек за год в сельской местности); Общая демографическая нагрузка (Сколько граждан нетрудоспособного возраста (детей и пожилых) приходится на 100 граждан трудоспособного возраста (мужчины 16–65 лет и женщины 16–60 лет)); Детские садики (Количество мест в детских садах на 1000 детей дошкольного возраста); Школы (Количество школ из топ-300 школ России на миллион детей); Плохие дороги (Доля протяженности автомобильных дорог общего пользования регионального или межмуниципального значения, не отвечающих нормативным требованиям, в общей протяженности автомобильных дорог общего пользования регионального значения в %); Жильё (Сколько квадратных метров приходится на душу населения), Рождаемость (Число родившихся на 1000 человек в год).

Также целесообразно использовать показатели Раскрываемость преступлений, Рождаемость и Естественный прирост населения.

Раскрываемость преступлений вычисляется по следующей формуле (1)

$$R_p = \frac{P_n}{P} * 100\% , \quad (1)$$

где P_n – количество нераскрытых преступлений, P – общее количество преступлений, зарегистрированных правоохранительными органами за год.

Естественный прирост населения в регионах РФ вычисляется по формуле (2).

$$E_p = R - S , \quad (2)$$

где R – коэффициент рождаемости, S – коэффициент смертности;

Безработица вычисляется по формуле (3)

$$B = \frac{B_k}{F} * 100\% , \quad (3)$$

где B_k – количество безработных, F – численность рабочей силы (занятых и безработных);

По данным из Data Set 1 оцениваются ситуации в регионах Российской Федерации. Исходные данные хранятся в виде *.csv файла (рис. 1).

1	А	В	С	Д	Е	Г	Н	И	К	Л	М	Н		
2	Номер за	Номер ди	IP-адрес	ID сессии	Дата начала	Время на	Имя поль	Регион	Поиск	Время по	Размер п	Время фо	Размер а	Врем
2	1				01.08.2019	00:10:35		Курганская область	почему не повышают зарплату?					
3	2				01.08.2019	00:05:30		Курганская область	Кому обращаться если не вывозят мусор					
4	3				01.08.2019	00:03:16		Курганская область	в больнице нас не приняли					
5	4				01.08.2019	00:00:38		Курганская область	gixtve e yfc ybxtuj yt vtyzntncz					
6	5				01.08.2019	00:00:18		Курганская область	арендатор накопил долги и исчез. Что делать?					
7	6				01.08.2019	00:00:23		Курганская область	Новые законы России					
8	7				01.08.2019	00:30:08		Курганская область	куда идти, если уволили?					
9	8				31.07.2019	23:03:14		Курганская область	как провест газ в Кургане					
10	9				31.07.2019	23:58:20		Курганская область	Кто должен заниматься стройкой и ограждением детских площа					
11	10				01.08.2019	00:00:50		Курганская область	оформление пособия					
12	11				31.07.2019	23:01:12		Курганская область	жилищный кодекс					
13	12				01.08.2019	01:01:11		Курганская область	из-за чего могут отказать в выдаче кредита					
14	13				01.08.2019	00:31:26		Курганская область	детские пособия					
15	14				01.08.2019	01:00:42		Курганская область	почему					
16	15				01.08.2019	01:00:37		Курганская область	мусор не вывозят неделями что делать?					
17	16				31.07.2019	23:00:12		Курганская область	трудовой кодекс РФ					
18	17				01.08.2019	01:10:19		Курганская область	пособие, как и рассчитать?					
19	18				01.08.2019	01:01:12		Курганская область	как регистрируются юридические лица					
20	19				01.08.2019	01:00:28		Курганская область	что делать если задерживают зарплату?					
21	20				01.08.2019	01:00:25		Курганская область	В больницах полный бардак! Что делать?					
22	21				01.08.2019	01:00:25		Курганская область	Когда повысят зарплату?					
23	22				01.08.2019	01:00:39		Курганская область	авыивыивавав					

Рис. 1 – Запросы жителей в СПС в формате csv

Как видно из рис. 1, в данной выборке присутствует много неинформативных полей и информации, а также некорректные запросы, которые только внесут путаницу при анализе данных. Поэтому необходима предобработка и последующая оценка качества этих данных. На рынке программных средств представлено множество решений для анализа больших данных, в том числе и свободно распространяемых, что является существенным фактором во многих ситуациях.

Одна их таких программ – Deductor Studio Academic, бесплатная версия аналитической платформы Deductor от российской компании BaseGroup.

Программа Deductor Academic предоставляет множество инструментов интеллектуального анализа данных. Так, помощью инструмента Качество данных [3], можно оценить пригодность исходных данных для анализа (рис. 2).

Качество данных	Резюме
1,0000	Пригоден
0,6239	Пригоден
0,7975	Предобраб...
0,9996	Пригоден
0,9926	Пригоден
1,0000	Пригоден
0,5220	Предобраб...
0,0725	Предобраб...
0,0168	Предобраб...
0,2374	Предобраб...
0,7246	Пригоден
0,1096	Предобраб...
0,1685	Предобраб...
0,2531	Предобраб...
0,2882	Предобраб...
0,0071	Предобраб...
0,1711	Предобраб...
0,0044	Предобраб...
0,8105	Пригоден
0,9937	Пригоден
0,9999	Пригоден
0,5547	Пригоден
0,9898	Пригоден
0,9889	Пригоден
0,0000	Непригоден
0,7473	Предобраб...

Рис. 2 – Качество данных

Далее необходимо провести исследование исходных данных. На этом этапе меняется формат и кодировка данных, удаляются ненужные поля, например, имена пользователей, IP-адрес, время пользования и т.д. Все они не понадобятся при анализе данных, т.к. будут оцениваться социально-политические и экономические ситуации в регионах в целом, а не отдельное состояние граждан этого региона. Подготовленные исходные данные приведены на рисунке 3.

Анализ тональности запросов целесообразно проводить с помощью анализа исходных данных и модели «мешка слов». В данной модели представления текстовых документов каждое встречающееся слово рассматривается как независимая случайная величина без привязки к содержанию текста и связи с другими словами

Регион	Поиск
Курганская область	Трудовой кодекс. Оплата труда
Курганская область	Кому обращаться если не вывозят мусор
Курганская область	в больнице нас не приняли
Курганская область	Новые законы России
Курганская область	куда идти, если уволили?
Курганская область	как провест газ в Кургане
Курганская область	Кто должен заниматься стройкой и ограждением
Курганская область	оформление пособия
Курганская область	жилищный кодекс
Курганская область	из-за чего могут отказать в выдаче кредита
Курганская область	арендатор накопил долги и исчез. Что делать?
Курганская область	почему не повышают зарплату?
Курганская область	детские пособия
Курганская область	мусор не вывозят неделями что делать?
Курганская область	трудовой кодекс РФ
Курганская область	пособие, как и рассчитать?
Курганская область	как регистрируются юридические лица
Курганская область	что делать если задерживают зарплату?
Курганская область	В больницах полный бардак! Что делать?
Курганская область	Когда повысят зарплату?

Рис. 3 – Исходная выборка для анализа данных

«Вес» слова пропорционален частоте его встречаемости в тексте. В результате анализа по различным показателям регионы были разделены на два класса (позитивные и негативные запросы).

По анализу тональности к неблагоприятным были отнесены следующие регионы: Республика Саха (Якутия), Алтайский край, Республика Карелия, Республика Мордовия, Республика Марий Эл, Архангельская область, Кабардино-Балкарская Республика, Республика Бурятия, Еврейская автономная область, Саратовская область, Курганская область, Пензенская область, Республика Алтай, Республика Калмыкия, Республика Ингушетия, Забайкальский край, Карачаево-Черкесская Республика, Республика Тыва.

Полученные результаты коррелируют с докладами Правительства [4] и данными социологических опросов [9]. Это позволяет сделать вывод, что данный инструментарий можно использовать в дальнейшем для анализа текста и классификации обращений граждан.

При анализе проблемных регионов Российской Федерации использовались официальные источники информации, публикуемые Федеральной службой государственной статистики [8], ЕМИСС – Единой межведомственной информационно-статистической системой [6], рейтинговым агентством «Эксперт», BASETOP – Топ Рейтинги Мира, порталом криминальной статистики, а также размещаемые в открытом доступе на официальных сайтах регионов и Правительства регионов.

После оценки качества исходных данных и предварительной обработки, необходимо выполнить их импорт в аналитическую программу Deductor (рис. 4).

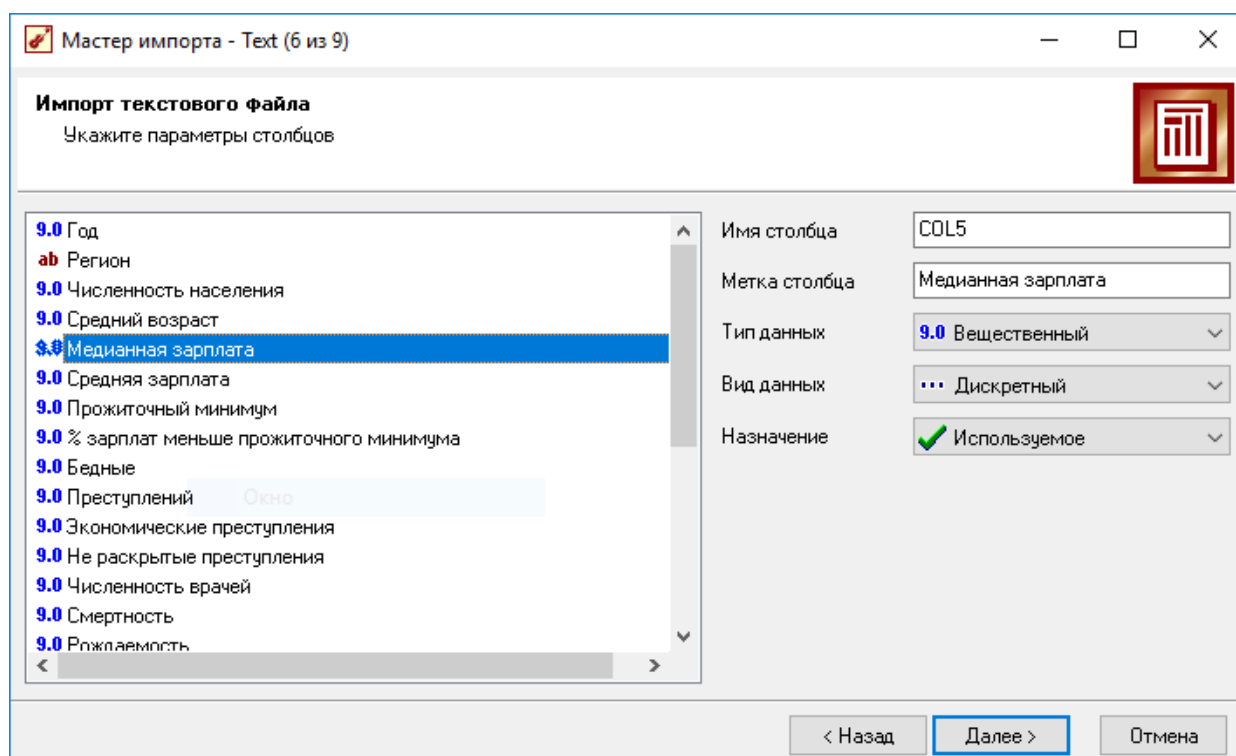


Рис. 4. – Импорт исходных данных

Обработанные исходные данные по проблемным регионам РФ приведены на рис. 5.

Регион	Численность населения	Средний возраст	Медианная зарплата	Средняя зарплата	Прожиточный минимум	% зарплата меньше прожиточно	Бедные	Преступлений	Экономические преступления	Не раскрытые преступления	Численность врачей	Смертность	Рождаемость	Естественный прирост
▶ Алтайский край	2332813	40,08	23480	29162	9908	29,3	17,5	169,32	4,11	34,88	43	14	10,8	10,8
Архангельская область	1100290	40,15	41030	50724	13110	11,4	14,1	171,09	8,84	39,33	54,86	13,2	10,5	10,5
Бурятия	983273	35,5	31260	39139	11549	15,1	18,5	239,91	8,52	41,86	46,38	10,7	14,5	14,5
Еврейская автономная область	159913	37,63	34538	40989	14342	23,1	24,9	219,95	4,02	41,03	36,54	13,3	11,7	11,7
Забайкальский край	1065785	35,9	33456	42443	12283	21,9	21,4	250,16	8,69	42,81	51,59	11,7	13,4	13,4
Ингушетия	497393	30,65	19954	24769	10461	25	31,6	33,43	4,25	43,28	46,47	3,2	16,3	16,3
Кабардино-Балкария	866219	36,46	19796	26110	11801	35,9	24,1	77,84	7,62	33,45	45,3	8,5	12,8	12,8
Калмыкия	272647	37,17	21984	26342	10119	34,5	30,6	109	8,89	26,88	53,2	9,9	11	11
Карачаево-Черкесия	465563	37,64	20788	26782	9775	38,4	24,6	86,89	11,23	23,07	41,01	9,2	11	11
Карелия	618056	40,48	34922	42996	14019	16,6	17,3	180,09	7,24	42,96	51,19	14,6	10,3	10,3
Курганская область	834701	40,9	24717	30114	9895	28,4	19,3	209,27	8,01	36,76	29,53	15,2	11,1	11,1
Марий Эл	680380	39,19	23739	29145	9830	24,8	22,8	107,38	6	38,85	36,26	12,4	11,9	11,9
Мордовия	795504	41,61	24052	28245	9181	19,5	18,9	98,47	8,66	26,61	53,06	13,5	8,5	8,5
Пензенская область	1318103	42,14	25563	29892	9544	22,1	14,6	88,1	4,54	26,52	41,66	14,1	8,9	8,9
Республика Алтай	218866	33,99	24872	30335	10162	35,2	25,1	213,4	7,79	24,79	47,74	9,7	15,8	15,8
Республика Тыва	324423	29,49	27822	35488	10778	23,5	41,5	366,28	6,75	57,25	53,4	8,7	21,9	21,9
Саратовская область	2440815	41,11	24155	30115	9521	22,8	17,4	116,3	6,45	42,56	50,02	13,6	9,5	9,5

Рис. 5 – Исходные данные для оценки ситуаций в регионах РФ

С помощью программы Deductor произведен корреляционный анализ полученных данных, результат показан на рис. 6.

Входные поля		Корреляция с выходными полями	Входные поля		Корреляция с выходными полями	Входные поля		Корреляция с выходными полями
№	Поле		№	Поле		№	Поле	Преступлений
1	Численность населения	0,425	1	Численность населения	-0,592	1	Численность населения	-0,152
2	Средний возраст	0,844	2	Средний возраст	-0,843	2	Средний возраст	-0,351
3	Медианная зарплата	0,430	3	Медианная зарплата	-0,329	3	Медианная зарплата	0,540
4	Средняя зарплата	0,375	4	Средняя зарплата	-0,294	4	Средняя зарплата	0,568
5	Прожиточный минимум	0,111	5	Прожиточный минимум	-0,080	5	Прожиточный минимум	0,356
6	% зарплат меньше прожиточного ...	-0,370	6	% зарплат меньше прожиточного ...	0,396	6	% зарплат меньше прожиточного ...	-0,272
7	Бедные	-0,708	7	Преступлений	0,272	7	Бедные	0,272
8	Преступлений	0,190	8	Экономические преступления	-0,032	8	Экономические преступления	0,064
9	Экономические преступления	-0,035	9	Не раскрытые преступления	0,332	9	Не раскрытые преступления	0,578
10	Не раскрытые преступления	-0,096	10	Численность врачей	0,147	10	Численность врачей	0,122
11	Численность врачей	-0,219	11	Смертность	-0,708	11	Смертность	0,190
12	Рождаемость	-0,653	12	Рождаемость	0,799	12	Рождаемость	0,569
13	Естественный прирост	-0,901	13	Естественный прирост	0,831	13	Естественный прирост	0,225
14	Безработица	-0,878	14	Безработица	0,742	14	Безработица	0,046
15	Общий прирост населения	-0,827	15	Общий прирост населения	0,587	15	Общий прирост населения	-0,052
16	Прирост в городах	-0,740	16	Прирост в городах	0,433	16	Прирост в городах	-0,303
17	Прирост в сёлах	-0,747	17	Прирост в сёлах	0,619	17	Прирост в сёлах	0,201
18	Общая демографическая нагрузка	0,566	18	Общая демографическая нагрузка	-0,197	18	Общая демографическая нагрузка	0,517
19	Плохие дороги	0,396	19	Жильё	-0,813	19	Детские сашки	0,006

Рис. 6 – Результаты корреляционного анализа

Результаты анализа показывают, что в регионах РФ имеет место высокая зависимость смертности от среднего возраста населения, бедности населения, безработицы, общего прироста населения, и средняя зависимость от зарплаты, плохих дорог, уровня преступности и численности врачей. Выявлена высокая зависимость бедности в регионах от безработицы, прироста в сёлах, жилищных условий, естественного и общего прироста населения и возраста населения, и средняя зависимость от зарплаты, преступлений, не раскрытых преступлений, прироста в городах и общей демографической нагрузки. Установлена высокая зависимость уровня преступности в регионах РФ от зарплаты, раскрываемости преступлений, рождаемости, общей демографической нагрузки и средняя зависимость от прожиточного минимума в регионах, бедности населения и прироста в городах.

Анализируя эти данные средствами Data Mining с помощью инструмента — [Дерево решений], были оценены социально-политические и экономические ситуации в проблемных регионах Российской Федерации.

Вектор переменных, подаваемый на вход дерева может содержать как числовые, так и порядковые, а также категориальные переменные. В любом случае в каждом узле разбиение идет только по *одной переменной*. Если переменная числового типа, то в узле формируется правило вида:

$$x_i \geq c, \quad (4)$$

где c – некоторый порог обучающей выборки, x_i – значения переменной.

Пороговое значение обучающей выборки для смертности, преступности, экономической преступности и т.п. в регионах Российской Федерации будет вычисляться по следующей формуле:

$$c = \frac{\sum_{i=1}^N x_i}{N}, \quad (5)$$

где x_i – значения переменной, i – коэффициенты переменной, N – количество данных. В рассматриваемом случае $N = 85$, т.к. мы оцениваем ситуацию внутри страны, и для общего анализа используются данные по всем регионам РФ.

С помощью дерева решений получены следующие результаты. Наиболее неблагоприятные регионы по естественному приросту населения: Алтайский край, Архангельская область, Еврейская автономная область, Карелия, Курганская область, Марий эл, Мордовия, Пензенская область, Саратовская область. В этих регионах Смертность больше Рождаемости. В Алтайском крае, Еврейской автономной, Курганской и Пензенской областях остро стоит вопрос здравоохранения. Судя по полученным данным, смертность в Архангельской, Курганская, Саратовской областях, Карелии, Марий Эл, Мордовии связана с большим количеством плохих дорог.

Архангельская область, Карелия, Бурятия, Забайкальский Край, Республика Алтай, Республика Алтай являются лидерами по преступности. Судя по полученным данным, высокая преступность возникает из-за бедности, безработицы и плохой работы органов внутренних дел (низкая раскрываемость преступлений) в регионах.

В Алтайском Крае, Пензенской области, Бурятии, Забайкальском крае, Ингушетии, Кабардино-Балкарии, Калмыкии, Карачаево-Черкесии, Республике Алтай и Республике Тыва остро стоит вопрос нехватки школ с рейтингом выше среднего и детских садиков.

Жилищные проблемы наблюдаются в следующих регионах: Алтайском Крае, Еврейской автономной области. Курганской области, Бурятии, Забайкальском Крае, Ингушетии, Кабардино-Балкарии, Калмыкии, Карачаево-

Черкесии, Республике Алтай и Республике Тыва. Судя по полученным данным, эти проблемы связаны высоким уровнем экономической преступностью в данных регионах.

По полученным результатам можно сделать выводы и предложить варианты решений. Например, в Алтайском крае, Еврейской автономной, Курганской и Пензенской областях больше выделить средств на здравоохранение (увеличить количество врачей, построить больницы), также дополнительно выделить средства для борьбы с экономической преступностью, чаще проводить проверки в этих регионах. Также в Алтайском Крае, Еврейской автономной области, Курганской области, Бурятии, Забайкальском Крае, Ингушетии, Кабардино-Балкарии, Калмыкии, Карачаево-Черкесии, Республике Алтай и Республике Тыва улучшить жилищные условия, принять больше мер для борьбы с экономической преступностью.

В Марий Эл, Мордовии, Бурятии, Забайкальском крае, Ингушетии, Кабардино-Балкарии, Республике Алтай, Республике Тыва больше выделить средства для развития сельской местности, т.к. в этих регионах наблюдается вымирание сельского населения – это связано с безработицей и миграцией сельской местности в городскую. Например, в Бурятии, Ингушетии, Кабардино-Балкарии, Республике Алтай, Республике Тыва в отличие от других регионов развивается сельская местность. Нужно поддерживать развитие сельской местности, но также следует обратить внимание на то, что в этих регионах (кроме Кабардино-Балкарии) отрицательный прирост населения в городах – это связано с безработицей в городах и бедностью в регионе. Для исправления всех этих ситуаций также нужно принять меры. Например, выделить больше средств из федерального бюджета и т.д.

В Архангельской области, Карелии, Бурятии, Забайкальском Крае, Республике Алтай, Республика Тыва принять больше мер и привлечь больше лиц для борьбы с преступностью, выделить больше средств из федерального бюджета и открыть больше рабочих мест в этих регионах, а также принять меры для исправления социального неравенства в этих субъектах.

В Алтайском Крае, Пензенской области, Бурятии, Забайкальском крае, Ингушетии, Кабардино-Балкарии, Калмыкии, Карачаево-Черкесии, Республике Алтай и Республике Тыва выделять средства на образование (строительство хороших школ, увеличение зарплаты учителям/преподавателям), на строительство детских садилов.

Как уже было сказано ранее, полученные данные коррелируются с докладами Правительства [4], с социологическими опросами [9]. Это позволяет сделать вывод, что данный инструментарий можно использовать в дальнейшем для анализа больших данных и отслеживания социально-политических и экономических ситуаций в регионах Российской Федерации.

Заключение

В рамках данной статьи была рассмотрена возможность применения технологий больших данных для отслеживания социально-политических и экономических ситуаций в регионах Российской Федерации. Была обоснована актуальность данной темы и была предложена методика решения поставленной задачи на основе данных из открытых источников и с использованием аналитической программы Deductor. Применение технологий Big Data для отслеживания социально-политических и экономических ситуаций в регионах является важным инструментом в принятии решений по развитию регионов и страны в целом.

Библиографический список

1. Актуальные проблемы России [Электронный ресурс]// Свобода слова. — URL: <https://svobos.ru/> (дата обращения: 20.09.2019).
2. Демография [Электронный ресурс]// Официальный сайт «Федеральной службы государственной статистики». — URL: <https://www.gks.ru/> (дата обращения: 14.09.2019).
3. Качество данных [Электронный ресурс] // BaseGroup Labs Технологии анализа данных. — URL: <https://basegroup.ru/deductor/function/algorithm/data-quality> (дата обращения: 06.10.2019).

4. Медведев назвал самые проблемные регионы России [Электронный ресурс]// Московский комсомолец. — URL: <https://www.mk.ru/> (дата обращения: 21.09.2019).»

5. Официальные статистические материалы [Электронный ресурс]// Официальный сайт «Федеральной службы государственной статистики». — URL: <https://www.gks.ru/> (дата обращения: 21.09.2019).

6. Официальные статические показатели [Электронный ресурс]// Государственная статистика. — URL: <https://www.fedstat.ru/> (дата обращения: 13.10.2019).

7. Полосин А.В., Барклянский Ю.А., Гречишников С.Е. Проект создания системы анализа и прогнозирования социально-политической ситуации в регионах Российской Федерации: методологическое обоснование // Территория роста. — URL: <http://territoria-rosta.ru/> (дата обращения: 15.09.2019).»

8. Регионы России [Электронный ресурс]// Официальный сайт «Федеральной службы государственной статистики». — URL: <https://www.gks.ru/> (дата обращения: 12.10.2019).

9. Рейтинг регионов России по уровню жизни 2019 [Электронный ресурс] // Топ Рейтинги Мира. — URL: <https://basetop.ru/> (дата обращения: 10.10.2019).

10. «Твердохлебова Е. А. Социально-экономическое прогнозирование как основа управления развитием территории» // Молодой ученый. — 2013. — №7. — С. 249-255. — URL: <https://moluch.ru/> (дата обращения: 14.09.2019).

References

1. Actual problems of Russia [Electronic resource] // Freedom of speech. — URL: <https://svobos.ru/> (accessed date: 09/20/2019).

2. Demographics [Electronic resource] // Official site of the «Federal State Statistics Service». — URL: <https://www.gks.ru/> (accessed date: 09/14/2019).

3. Data quality [Electronic resource] // BaseGroup Labs Data analysis technologies. — URL: <https://basegroup.ru/deductor/function/algorithm/data-quality>

(accessed: 06.10.2019).

4. Medvedev named the most problematic regions of Russia [Electronic resource] // Moskovsky Komsomolets. – URL: <https://www.mk.ru/> (accessed: 09/21/2019).

5. Official statistical materials [Electronic resource] // Official site of the Federal State Statistics Service. – URL: <https://www.gks.ru/> (accessed: 09/21/2019).

6. Official static indicators [Electronic resource] // State statistics. – URL: <https://www.fedstat.ru/> (accessed: October 13, 2019).

7. Polosin A.V., Barklyansky Yu.A., Grechishnikov S.E. The project of creating a system of analysis and forecasting of the socio-political situation in the regions of the Russian Federation: methodological justification // Territory of growth. – URL: <http://territoria-rosta.ru/> (accessed: 09/15/2019).

8. Regions of Russia [Electronic resource] // Official site of the «Federal State Statistics Service». – URL: <https://www.gks.ru/> (accessed: October 12, 2019).

9. Rating of Russian regions by living standards 2019 [Electronic resource] // Top World Ratings. – URL: <https://basetop.ru/> (accessed date: 10/10/2019).

10. Tverdokhlebova E.A. Socio-economic forecasting as the basis for managing the development of the territory // Young scientist. – 2013. – № 7. – S. 249-255. – URL: <https://moluch.ru/> (accessed: 09/14/2019).