

Применение технологий больших данных для исследования влияния социально-экономических факторов на состояние психического здоровья населения

Рабинович А.Е., к.э.н., доцент кафедры «Прикладная информатика»

Московский Политехнический университет, Россия, г. Москва

Седова Е.А., магистр направления «Системная аналитика больших данных»

Московский Политехнический университет, Россия, г. Москва

Аннотация. В статье обоснована актуальность и важность проведения анализа влияния социально-экономических факторов на состояние здоровья населения стран с использованием технологии Big Data, а также определены основные методы, которые позволят решить описанную проблему. На основании анализа корреляции факторов может быть предложена прогнозная модель, которая позволит решить описанную проблему и представить рекомендации о том, какие профилактические меры позволят минимизировать негативные последствия изменения факторов.

Ключевые слова: технологии Big Data, психическое здоровье, статистика самоубийств, методы обработки больших данных, прогнозная модель, корреляция.

The use of big data technologies to study the influence of socio-economic factors on the state of mental health of the population

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

Sedova E.A., Master of Science in Big Data Systems Analytics
Moscow Polytechnic University, Russia, Moscow

Annotation. The article substantiates the relevance and importance of analyzing the impact of socio-economic factors on the health of the population of countries using Big Data technology, as well as identifies the main methods that will solve the described problem. Based on the analysis of the correlation of factors, a predictive model can be proposed that will solve the described problem and provide recommendations on what preventive measures will minimize the negative consequences of changing factors.

Keywords: Big Data technologies, mental health, suicide statistics, big data processing methods, predictive model, correlation.

Одним из общепризнанных фактов является то, что существует прямая зависимость между экономическим состоянием страны и здоровьем ее населения. Население страны является производительным ресурсом, который формирует эффективную экономику, соответственно, истощение трудового потенциала экономики страны является одной из важнейших проблем, которую необходимо решать. Только психически и физически здоровый человек способен в полной мере выполнять свои социально-экономические функции, соответственно повышение уровня здоровья населения может оказать положительное влияние на все сферы деятельности государства [1].

Одной из важных составляющих здоровья является здоровье психическое, которое характеризует состояние благополучия, при котором человек может эффективно реализовывать свой потенциал, справляться со стрессовыми ситуациями, продуктивно работать и вносить свой вклад в жизнь общества и в целом страны, в которой он живет. И, несмотря на то, что часто под понятием «здоровый человек» подразумевается физически здоровый человек, нельзя забывать о том, что в современных реалиях возрастает количество случаев нарушения психического и психосоциального здоровья. По статистике Всемирной Организации Здравоохранения депрессивные расстройства занимают четвертое место среди причин потери

трудоспособности, после инфекционных, сердечно-сосудистых и респираторных заболеваний [2].

Но на состояние здоровья оказывает влияние не только экономическое состояние страны. Существует ряд богатых экономически развитых стран с высоким уровнем жизни, где достаточно большой процент самоубийств, например, Бельгия, Франция и Финляндия. Исходя из этого можно сделать вывод о том, что существует ряд факторов, влияющих на состояние психического здоровья человека, которые невозможно явно определить с первого взгляда. Другими словами, существуют скрытые зависимости, определение которых является важным шагом к пониманию причин, по которым происходит ухудшение состояния психического здоровья населения тех или иных стран.

В настоящее время в рамках социально-гигиенического мониторинга проводится большое количество исследований, выявляющих влияние экологических факторов на здоровье населения стран, а также на заболеваемость и смертность. Однако, в свете проводимых реформ и возникающих кризисных явлений, большой интерес вызывает изучение корреляции социально-экономических факторов и здоровья населения, на основании которых может производиться разработка проектов и решений в управленческой среде для последующего улучшения качества жизни населения в различных странах.

Проблема исследования влияния различных факторов на психическое здоровье населения была описана еще в конце прошлого столетия в программе Всемирной организации здравоохранения «Здоровье – 21: политика достижения здоровья для всех в Европейском регионе ВОЗ», в которой улучшение состояния психического здоровья отмечено как задача первостепенной важности. В данной программе указывается, что «улучшение психического здоровья – особенно снижение числа самоубийств – требует самого пристального внимания к вопросам укрепления и охраны психического здоровья на протяжении всей жизни, в частности, в находящихся в

неблагоприятном социально-экономическом положении группах населения» [3].

Существует большое количество факторов, которые могут влиять на состояние психического здоровья населения. Наибольшая их часть относится к уровню социально-экономического развития страны. Недостаточное внимание государства к анализу факторов, влияющих на психическое здоровье населения, может привести к ухудшению состояния всех государственных сфер. Задачами первостепенной важности является анализ и оценка динамики изменения факторов, для разработки профилактических мер по улучшению данных показателей.

Одним из показателей, по которому можно судить о состоянии психического здоровья населения страны, является статистика самоубийств. По статистике, ежедневно в мире добровольно уходят из жизни около 3-х тыс. человек, ежегодно – около 1 млн. человек. Данная проблема приобретает более масштабный характер за счет того, что число попыток самоубийств превышает число самоубийств более чем в 15 раз. Тем более угрозы и попытки самоубийства обычно бывают повторными. По различным данным, от 20 до 60% людей, покончивших с собой, ранее уже предпринимали попытки самоубийства. За последние 50 лет количество самоубийств выросло примерно на 50%, особенно среди подростков и молодых людей из промышленно-развитых стран. На несформировавшуюся личность подростков оказывают большое влияние средства массовой информации. Вследствие этого часто происходит сведение счетов с жизнью в качестве подражания любимым героям из фильмов, книг компьютерных игр и реалити-шоу. Это является серьезной причиной для изучения влияния различных факторов, в том числе и социально-экономических, на состояние психического здоровья различных групп людей в разные промежутки времени. Есть большое количество скрытых факторов, которые могут оказывать влияние, которые невозможно определить, не осуществляя глубинный анализ данной темы.

Лучше всего динамика изменения факторов прослеживается на ретроспективных данных за продолжительные промежутки времени. Для того, чтобы сделать корректные выводы, необходимо анализировать различные источники данных, которые чаще всего разнородны по структуре. Сфера здравоохранения накапливает большие объемы данных, чаще всего неструктурированных, на основании анализа которых можно составить статистику влияния различных факторов на состояние здоровья населения различных стран. Извлечение информации из неструктурированных данных позволит выявить закономерности, на основании которых можно будет делать выводы о том, изменение каких факторов в наибольшей степени влияет на состояние психического здоровья населения. На основе данных выводов можно будет разработать модель, которая сможет прогнозировать изменение факторов. На основе полученных прогнозов может быть разработан комплекс управленческих решений со стороны государства по улучшению социально-экономических показателей и уменьшению негативного влияния на состояние здоровья граждан.

Одной из основных проблем, препятствующих глубокому и детальному изучению данной темы, является ограниченный доступ к источникам данных, в том числе медицинских. Несмотря на это, в открытых источниках можно найти наборы данных, достаточных для проведения ретроспективного анализа, позволяющего проследить тенденции изменения состояния психического здоровья населения в корреляции с социально-экономическими факторами, характеризующими состояние страны в определенный период времени.

Для решения поставленной задачи было проанализировано несколько источников больших данных:

1. Статистика самоубийств в разных странах с 1960 по 2017 годы [4];
2. Статистические данные о здоровье населения: содержит около 300 различных показателей по каждой стране (например, динамика населения, репродуктивное здоровье, иммунизация, инфекционные заболевания, коэффициент смертности при родах и т. д.) [5];

3. Статистика индекса человеческой свободы: содержит около 79 индикаторов личной и экономической свободы в различных областях в разных странах [6].

Существует несколько проблем, которые необходимо будет решить:

1. Определить какие данные из найденных источников следует включить в анализ;
2. Разрешить проблему с неоднородностью данных, для того чтобы привести их к виду, пригодному для анализа;
3. Определить методы, по которым будет проводиться анализ данных.

Для того, чтобы определить какие данные из источников необходимо включить в анализ, необходимо подробно рассмотреть каждый из них.

Основным источником данных, на основе анализа которого будут строиться выводы, является статистика самоубийств в различных странах. Данный источник содержит в себе порядка 28 тысяч записей, распределенных на 12 колонок, в которых содержится информация о стране (количество населения, показатели ВВП в год и на душу населения, индекс человеческого развития), а также информация о возрастных группах и половой принадлежности людей, покончивших жизнь самоубийством.

Данный источник содержит в себе неструктурированную информацию, на основании которой можно вычислить основной показатель, по которому будет производиться анализ – количество самоубийств на 100 тысяч населения. После определения данного показателя появится возможность построить график его распределения по годам для каждой страны, а также сравнить его с показателем ВВП на душу населения, что позволит выдвинуть гипотезу о том, каким образом влияет динамика изменения данного показателя на уровень самоубийств для каждой страны.

Для проведения более детального анализа был выбран источник данных, содержащий в себе статистические данные о здоровье населения различных стран, который содержит в себе около 90 тысяч записей и 300 различных показателей по каждой стране. После выявления экстремальных значений из

первого источника можно отследить динамику изменения показателей здоровья населения для данной страны в благоприятные и неблагоприятные периоды. Анализ ситуации в стране может быть произведен только в совокупности факторов, влияющих на здоровье населения, что включает в себя как физическую, так и психологическую составляющую.

Наконец, для того, чтобы детальнее раскрыть тему влияния на психическое здоровье именно социально-экономических факторов, был выбран источник данных, содержащий в себе статистику индекса человеческой свободы в различных странах. Индекс человеческой свободы является всеобъемлющим индексом, который отражает степень свобод в отдельных сферах человеческой жизни и деятельности. Данный индекс включает в себя 79 показателей личных, гражданских и экономических свобод в 12 областях человеческой деятельности: верховенство права, свобода передвижения, безопасность, свобода выбора идентичности и личных отношений, свобода религиозной практики и др. Размеры и динамические изменения данных показателей может оказывать большое влияние на психическое здоровье населения.

К настоящему моменту разработано большое количество методов обработки больших данных в различных областях медицины: исследование эффективности лечения в медицинских учреждениях, нейронные сети и системы поддержки принятия решений для постановки диагноза на основании симптомов, разработка алгоритмов и предиктивного моделирования лекарственных средств, массовый скрининг и предупреждение выявления эпидемий и др. [7]. Однако для того, чтобы сформировать понимание о причинах заболеваний, в том числе и психического характера, очень важно обратить внимание на анализ корреляции факторов, влияющих на состояние здоровья населения. Знание о том, какие социально-экономические факторы в наибольшей степени оказывают влияние на здоровье населения, может использоваться для прогнозирования динамики.

Построение прогнозной модели позволит более детально изучить проблему самоубийств, и позволит сделать выводы о том, какие профилактические меры и управленческие решения позволят минимизировать негативные последствия изменения факторов. Для того, чтобы проанализировать влияние всех вышеперечисленных факторов, необходимо построить модель, которая сможет просчитать зависимость изменения данных показателей и динамики самоубийств в различных странах. После расчета корреляции можно будет сделать вывод о том, какие факторы в наибольшей степени влияют на изменение результирующего показателя – количества самоубийств, следовательно, сделать выводы о том, какие данные следует использовать для построения прогнозной модели.

Для решения поставленной проблемы, и проблем похожего характера, помимо построения модели анализа необходима разработка программной реализации данной модели, которая позволит в автоматическом режиме просчитывать корреляцию и по определенному алгоритму отбирать наиболее значимые факторы для построения модели. На основании отобранных факторов будет производиться автоматическое построение прогнозной модели. Разработанное программное средство может стать инструментом для построения прогнозных моделей.

В рамках выполнения данной задачи имеет смысл разделить процесс на несколько составляющих:

1. Разработка алгоритма отбора наиболее значимых факторов из загруженных источников данных на основании методов статистического анализа данных;
2. Разработка алгоритма обучения системы для построения прогнозной модели;
3. Определение функциональной структуры разрабатываемой системы;
4. Разработка функциональных модулей системы на языках программирования высокого уровня, таких как Python;
5. Отладка разработанного функционала, тестирование системы.

Технологии прогнозных моделей опираются на методы обработки больших массивов данных (Big Data). Прогнозная аналитика задействует множество статистических методов, методов интеллектуального анализа данных, позволяет анализировать как текущие данные, так и данные за прошлые периоды, на основе анализа которых составляются прогнозы о будущих событиях [8]. Разработанная прогнозная модель позволит наметить сценарии будущего изменения наиболее важных показателей, которые в наибольшей степени оказывают влияние на изменение состояния психического здоровья населения. На основании полученных выводов можно проводить формирование и разработку мер для минимизации ущерба от изменения наиболее важных факторов, что будет способствовать решению поставленной проблемы.

Библиографический список

1. Самутин К.А. Здоровье населения как составной элемент экономической политики государства // Российское предпринимательство. – 2012. – Том 13. – № 11. – С. 131-136.

2. Всемирная организация здравоохранения и Колумбийский университет. Групповая интерперсональная терапия (ИПТ) при депрессии [Group Interpersonal Therapy (IPT) for Depression]. Женева: Всемирная организация здравоохранения; 2018.

3. Здоровье-21: Основы политики достижения здоровья для всех в Европейском регионе ВОЗ: введение / Европейское региональное бюро ВОЗ, - 5 изд. 1999.

4. World Health Organization URL: https://www.who.int/mental_health/prevention/suicide/suicideprevent/en/ (дата обращения: 21.10.2019).

5. Health Nutrition And Population Statistics // The World Bank URL: <https://datacatalog.worldbank.org/dataset/health-nutrition-and-population-statistics> (дата обращения: 30.10.2019).

6. Human Freedom Index // CATO Institute URL: <https://www.cato.org/human-freedom-index-new> (дата обращения: 29.10.2019).

7. Антонова Н.Е., Большаков А.А., Киселев В.В., Лобанов В.В. Применение гибридной интеллектуальной системы поддержки принятия решений в стабилотрии // Вестник ТГТУ. – 2014. – №2. URL: <https://cyberleninka.ru/article/n/primenenie-gibridnoy-intellektualnoy-sistemy-podderzhki-prinyatiya-resheniy-v-stabilometrii> (дата обращения: 04.11.2019).

8. Прогнозная аналитика (Big Data) // Marketing Logic URL: <https://marketing-logic.ru/bigdata> (дата обращения: 06.11.2019).

References

1. Kirill Samutin (2012). Public Health as an Integral Element of the National Economic Policy [Zdorovye naseleniya kak sostavnoy element ekonomicheskoy politiki gosudarstva]. Russian Journal of Entrepreneurship, 13(11). (in Russian).

2. World Health Organization and Columbia University. Group Interpersonal Therapy (IPT) for Depression [Group Interpersonal Therapy (IPT) for Depression]. Geneva: World Health Organization; 2018.

3. Health 21: A health for all policy framework for the WHO European Region: introduction / WHO Regional Office for Europe, 5th ed. 1999.

4. World Health Organization URL: https://www.who.int/mental_health/prevention/suicide/suicideprevent/en/ (accessed date: 21.10.2019).

5. Health Nutrition And Population Statistics // The World Bank URL: <https://datacatalog.worldbank.org/dataset/health-nutrition-and-population-statistics> (accessed date: 30.10.2019).

6. Human Freedom Index // CATO Institute URL: <https://www.cato.org/human-freedom-index-new> (дата обращения: 29.10.2019).

7. Antonova N.E., Bolshakov A.A., Kiselev V.V., Lobanov V.V. Application of a hybrid intelligent decision support system in stabilometry // Vestnik TSTU. – 2014. – №2. URL: <https://cyberleninka.ru/article/n/primenenie-gibridnoy->

intellektualnoy-sistemy-podderzhki-prinyatiya-resheniy-v-stabilometrii (accessed 04.11.2019).

8. Predictive analytics (Big Data) // Marketing Logic URL: <https://marketing-logic.ru/bigdata> (accessed: 06.11.2019).