

# АВТОМАТИЗИРОВАННЫЙ АНАЛИЗ НОВОСТЕЙ ДЛЯ ВЫЯВЛЕНИЯ ТЕНДЕНЦИЙ РАЗВИТИЯ СОЦИАЛЬНЫХ РИСКОВ

*Смолина Е.Р., студентка группы СГНЗ-11М*

*Московский государственный технический университет им Н.Э. Баумана*

*Научный руководитель: Гришнова Е.Е., доктор политических наук,  
профессор кафедры «Информационная аналитика и политические технологии»  
grishnova@bmstu.ru*

**Аннотация:** В статье раскрываются возможности применения модели BERT для анализа новостных данных и выявления тенденций развития социальных рисков. Разработанный инструмент позволяет эффективно обрабатывать большие объемы информации, анализировать взаимосвязи между событиями и социальными показателями, предоставляя актуальные данные для мониторинга и управления рисками.

**Ключевые слова:** информационная аналитика, анализ, социальные риски, машинное обучение.

Социальные риски тесно связаны с трансформацией системы жизненных ориентиров человека и представляют собой «явления и процессы социальной сферы общества, которые несут опасность социальной дестабилизации и являются весьма сложными по своей структуре, со специфическим функциям и последствиями» [5]. Социальные риски следует рассматривать как значимые угрозы общественного характера, источники которых носят деструктивный характер. Стоит отметить, что самостоятельная защита от таких рисков зачастую оказывается невозможной ввиду их социального масштаба и природы.

Понятие «социальный риск» значительно расширяется за счет анализа его взаимосвязи с процессами социальных изменений. В современных условиях рост социальных рисков обусловлен увеличением уровня неопределенности, который сопровождается социальной нестабильностью, а также сложностью или даже невозможностью точной оценки вероятных последствий различных действий. Ключевым подходом к решению социальных проблем на уровне общества становится диагностика, прогнозирование и предупреждение приоритетных социальных угроз и факторов, влияющих на национальную безопасность и устойчивое развитие. В связи с этим социальный риск необходимо рассматривать как особую категорию риска, формирующуюся в рамках социальной сферы.

В условиях быстрого изменения социальных процессов и растущего объема информации эти риски требуют не только глубокого анализа, но и применения новых подходов к их исследованию [6]. Современные

работы в области социальных наук показывают возрастающую потребность в разработке эффективных методов анализа социальных рисков [9]. В связи с этим становится необходимо внедрение автоматизированных инструментов, способных эффективно обрабатывать и структурировать данные.

В данной статье будут описаны результаты создания инструмента для автоматизированного анализа новостей, что поможет своевременно выявлять тенденции развития социальных рисков. Было решено использовать в разработке методы машинного обучения, так как они предоставляют широкие возможности анализа данных. Машинное обучение (machine learning) фокусируется на разработке и изучении алгоритмов, которые способны анализировать данные и выполнять предсказательный анализ на их основе [14]. Эти алгоритмы создают модели, используя тренировочный набор данных, и на основании этого делают прогнозы или принимают решения, представленные в виде выходных результатов. Преимущество данных методов заключается в способности выявлять скрытые закономерности и тенденции, которые могут быть не замечены при использовании традиционных методов анализа.

Одной из ключевых технологий в этой сфере является обработка естественного языка (NLP). Современные модели NLP способны понимать контекст слов и предложений, классифицировать тексты, извлекать из них ключевую информацию и выявлять скрытые взаимосвязи между различными социальными явлениями [14]. В современных условиях переизбытка информации технологии NLP становятся мощным инструментом и позволяют решать широкий спектр прикладных задач. Интеграция таких моделей в системы анализа и прогнозирования социальных рисков позволит получать более точные и актуальные данные, что улучшит процесс принятия решений и управления социальными рисками.

В нашем инструменте была использована модель BERT (Bidirectional Encoder Representations from Transformers), основанная на архитектуре трансформеров и разработанная компанией Google [15]. BERT обучается на огромном количестве текста и использует механизмы внимания для понимания контекста слов в предложениях, что позволяет модели более точно интерпретировать значение слов в зависимости от текста. Рассмотрим подробнее, почему именно BERT стала ключевым инструментом в нашем проекте.

Во-первых, BERT обладает глубоким пониманием контекста, анализируя слова в предложении с учётом как левого, так и правого контекста. Это позволяет модели более точно интерпретировать значение каждого слова.

Во-вторых, эта модель может быть адаптирована для широкого спектра задач. Для нас ее преимущество заключалось в zero-shot классификации, где модель способна работать с новыми категориями данных без предварительного обучения на них. Это важно для создания универсального инструмента анализа социальных рисков и расширяет сферу её применения в исследованиях.

В-третьих, BERT показала выдающиеся результаты в многочисленных NLP задачах, демонстрируя большую эффективность по сравнению с другими моделями [16].

Перейдем к практической реализации нашего инструмента. На Рис. 1 представлена архитектура программы.

Архитектура включает следующие основные компоненты:

1. Фронтенд – интерфейс взаимодействия с пользователем, позволяющий отправлять запросы и отображать результаты анализа. Он обеспечивает удобный доступ к функционалу инструмента для аналитиков и других пользователей.

2. API – центральный компонент, обеспечивающий связь между всеми модулями системы. Он принимает запросы от фронтенда, передает их в модули обработки данных и возвращает результаты пользователю.

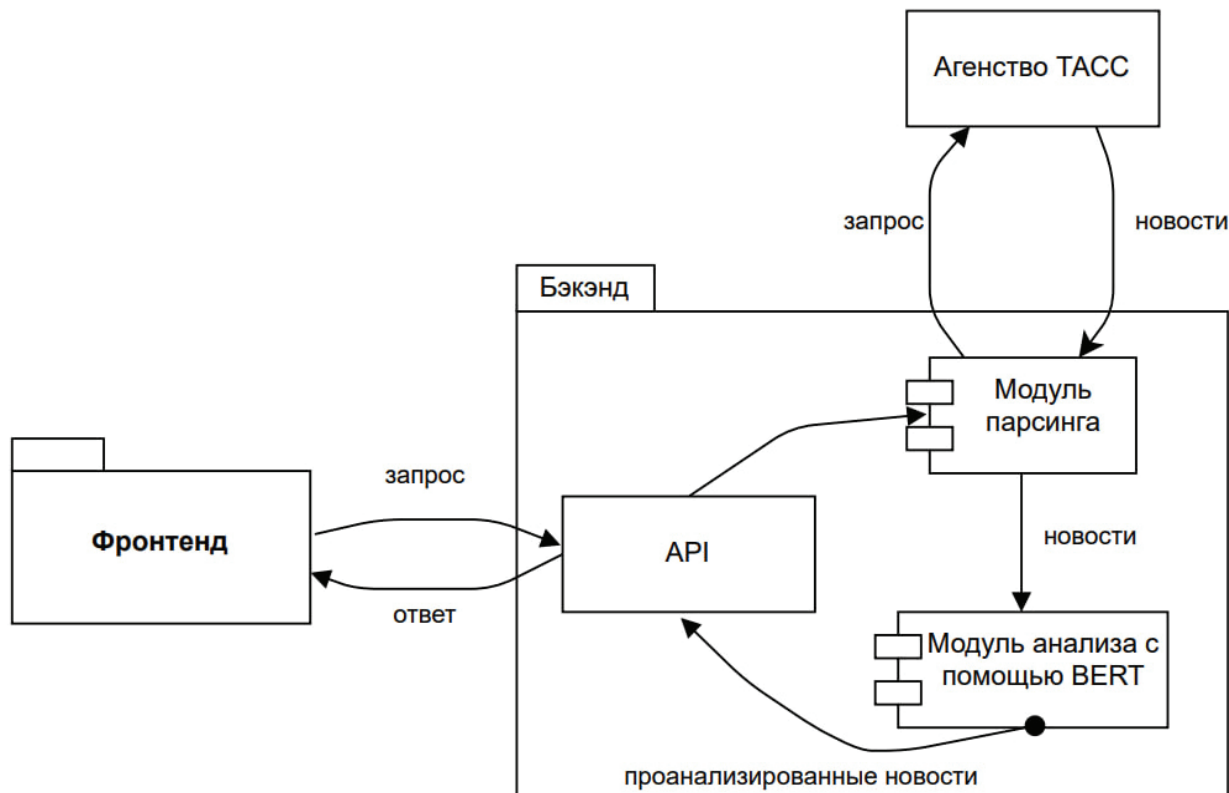


Рис. 1. Архитектура инструмента

3. Модуль парсинга данных – компонент, отвечающий за сбор данных из внешних источников, таких как новостные агентства. В рамках исследования использовалось взаимодействие с новостным агентством ТАСС для получения актуальных данных. Архитектура позволяет добавить взаимодействие и с другими новостными агентствами, что делает разработанный инструмент гибким.

4. Модуль анализа – ключевой компонент, где происходит обработка текстов с использованием модели BERT. Новости классифицируются по заданным категориям, определяемым аналитиком, а также выделяются скрытые взаимосвязи между событиями и их потенциальное влияние на социальные показатели. Данный модуль отвечает за классификацию социальных рисков на основе неявно заданных категорий и отображение вероятности влияния той или иной новости на развитие социального риска.

5. Блок результатов – формирует окончательный отчет и отправляет его обратно через API на фронтенд для отображения пользователю. Отчет содержит структурированную информацию, включая вероятности принадлежности новостей к различным категориям рисков.

Для удобства использования разработанного инструмента, важно обеспечить доступ к его функциям для информационных аналитиков через интегрируемый в их рабочие процессы интерфейс. В этой связи был создан веб-сайт, который позволяет аналитикам взаимодействовать с системой напрямую через браузер. Интерфейс инструмента представлен на Рис. 2.

Добавьте пары классификации (две обязательны):

|         |         |         |
|---------|---------|---------|
| Класс 1 | Класс 2 | Удалить |
| Класс 1 | Класс 2 | Удалить |
| Класс 1 | Класс 2 | Удалить |

Добавить пару

Анализировать

Сводка по всем новостям

Результаты анализа

Текст новости

*Рис. 2. Интерфейс разработанного инструмента*

В результате был разработан сайт, который на основе постоянно обновляемых новостей позволяет выявлять их взаимосвязь с социальными рисками. Инструмент может быть использован для мониторинга социальных рисков в режиме реального времени, предоставляя анали-

тикам и аналитическим центрам актуальную и точную информацию для поддержки принятия решений. Полученная разработка подходит для любых видов социальных рисков, позволяя самостоятельно вводить параметры классификации.

Для демонстрации возможностей разработанного инструмента был проведен анализ одного из демографических социальных рисков – снижения рождаемости, наблюдаемого в последние годы. По мнению экспертов, с 2016 года Россия вошла во вторую волну депопуляции, которая формируется, прежде всего, вследствие устойчивого снижения рождаемости [1].

Стоит уточнить, что для исследования с помощью инструмента аналитикам необходимо самостоятельно определять ключевые параметры, такие как типы социальных рисков и влияющие на них факторы. Так, для анализа снижения рождаемости было выделено 6 ключевых факторов:

- Количество браков и разводов: стабильные супружеские отношения создают благоприятные условия для рождения и воспитания детей.
- Уровень безработицы: экономическая нестабильность и высокий уровень безработицы могут приводить к откладыванию планов по рождению детей.
- Доступность медицинского обслуживания: уверенность в качестве медицинской помощи и своевременности ее получения должна повысить средний коэффициент рождаемости.
- Средняя продолжительность жизни: высокая продолжительность жизни может сопровождаться снижением рождаемости.
- Среднедушевой уровень доходов населения: увеличение доходов семьи способствует улучшению условий жизни и повышению готовности к рождению детей.
- Число семей, улучшивших жилищные условия: для рождения и воспитания детей нужно комфортные жилищные условия.

При этом выбор указанных факторов обоснован их существенным влиянием на демографические процессы, что подтверждается научными исследованиями [1, 2, 7, 9, 12]. Для анализа также необходимо явно указать как положительную динамику развития фактора, так и отрицательную. Например, при добавлении пар классификации следует указать в одном поле «доступность медицины уменьшается», а в другом «доступность медицины увеличивается». Далее инструмент автоматически обрабатывал входные данные и возвращал вероятностные оценки принадлежности новостей к каждой из категорий (Рис. 3).

Добавьте пары классификации (две обязательны):

|  |  |         |
|--|--|---------|
| Число семей получивших жилье уменьшается   | Число семей получивших жилье увеличивается | Удалить |
| Младенческая смертность растет             | Младенческая смертность падает             | Удалить |
| Количество браков уменьшается              | Количество браков растет                   | Удалить |
| Количество разводов уменьшается            | Количество разводов растет                 | Удалить |
| Уровень безработицы уменьшается            | Уровень безработицы растет                 | Удалить |
| Доступность медицины уменьшается           | Доступность медицины увеличивается         | Удалить |
| Средняя продолжительность жизни падает     | Средняя продолжительность жизни растет     | Удалить |
| Уровень дохода уменьшается                 | Уровень дохода растет                      | Удалить |
| Величина прожиточного минимума уменьшается | Величина прожиточного минимума растет      | Удалить |

[Добавить пару](#)

[Анализировать](#)

**Сводка по всем новостям**

| Пара классификации  | Вероятность влияния на первый класс | Вероятность влияния на второй класс |
|---|-------------------------------------|-------------------------------------|
| Число семей получивших жилье уменьшается / Число семей получивших жилье увеличивается | 0.46                                | 0.54                                |
| Младенческая смертность растет / Младенческая смертность падает                       | 0.44                                | 0.56                                |
| Количество браков уменьшается / Количество браков растет                              | 0.52                                | 0.48                                |
| Количество разводов уменьшается / Количество разводов растет                          | 0.55                                | 0.45                                |
| Уровень безработицы уменьшается / Уровень безработицы растет                          | 0.72                                | 0.28                                |
| Доступность медицины уменьшается / Доступность медицины увеличивается                 | 0.32                                | 0.68                                |
| Средняя продолжительность жизни падает / Средняя продолжительность жизни растет       | 0.65                                | 0.35                                |
| Уровень дохода уменьшается / Уровень дохода растет                                    | 0.52                                | 0.48                                |
| Величина прожиточного минимума уменьшается / Величина прожиточного минимума растет    | 0.30                                | 0.70                                |

*Рис. 3. Результат работы инструмента*

Процесс сбора и обработки данных занял не более 3 минут. В результате анализа видно, что новости положительно влияют на «число семей, получивших жилье» с вероятностью 46%, тогда как вероятность их негативного влияния составляет 54%. Аналогично для пары «доступность медицины уменьшается / доступность медицины увеличивается» вероятности составили 32% и 68%, это свидетельствует о том, что доступность медицинского обслуживания растет.

В инструменте можно также посмотреть результаты анализа конкретных новостей. Это делает результаты анализа прозрачными для аналитиков, а также позволяет контролировать точность полученных данных. Ниже представлены примеры новостей, полученных в результате обработки (рис. 4).

|   |   |            |                                 |            |            |   |            |                              |            |
|---|---|------------|---------------------------------|------------|------------|---|------------|------------------------------|------------|
| На юге Приморья запустят видеомониторинг дальневосточных тюленей: Проект будет представлять собой сеть из видеокамер на солнечных батареях, работающих на основе искусственного интеллекта МТС, которые будут реагировать только на детенышей тюленей                       | Нейтрально  | Нейтрально | Нейтрально                      | Нейтрально | Нейтрально | Нейтрально                                | Нейтрально | Нейтрально                   | Нейтрально |
| Более 140 аппаратов ИВЛ закупят для больниц Подмосковья в 2024 году: На эти цели предусмотрено свыше 440 млн рублей   | Нейтрально  | Нейтрально | Нейтрально                      | Нейтрально | Нейтрально | Скорее доступность медицины увеличивается | Нейтрально | Нейтрально                   | Нейтрально |
| Власти Москвы рассказали о потенциальном росте производительности труда в городе: Заммэра столицы, глава департамента экономической политики и развития Мария Багреева отметила, что столица является крупнейшей экономикой России, формирующей одну пятую экономики страны | Нейтрально  | Нейтрально | Нейтрально                      | Нейтрально | Нейтрально | Нейтрально                                | Нейтрально | Скорее Уровень дохода растет | Нейтрально |
| В Крыму число МСП и самозанятых выросло в два раза с 2016 года: Число тех, кто занимается малым и средним предпринимательством, составило 220 тыс. человек  | Скорее Число семей получивших жилье увеличивается | Нейтрально | Скорее Количество браков растет | Нейтрально | Нейтрально | Нейтрально                                | Нейтрально | Скорее Уровень дохода растет | Нейтрально |

**Рис. 4. Пример классификации новостей**

Рассмотрим некоторые из результатов обработки:

1. Новость о запуске видеомониторинга дальневосточных тюленей не относится к новостному фону, связанному с социально-демографическими рисками и факторами, влияющими на них, и разработанный инструмент классифицировал эту новость как нейтральную.

2. Новость о запуске 140 аппаратов ИВЛ для больниц Подмосковья в 2024 году была классифицирована как нейтральная для большинства факторов, однако было отмечено небольшое положительное влияние на доступность медицинской помощи, что логично.

3. Новость о потенциальном росте производительности труда в Москве, по оценке инструмента, оказывает нейтральное влияние на большинство факторов, но есть тенденция к росту уровня дохода.

4. Новость о росте числа МСП и самозанятых в Крыму имеет положительное влияние сразу на несколько факторов: увеличение дохода, увеличение числа семей, получивших жилье, и увеличение количества браков. Действительно рост числа малых и средних предприятий (МСП) и самозанятых приводит к увеличению доходов населения, а

увеличение доходов позволяет улучшать жилищные условия. Это, в свою очередь, способствует большей уверенности пар в будущем и способствует росту числа браков.

Эти примеры подтверждают, что разработанный инструмент способен эффективно выявлять ключевые тенденции и изменения в социально-экономических показателях на основе анализа новостей. Таким образом, использование модели BERT и интеграция её в разработанный инструмент обеспечивают высокую точность анализа и актуальность получаемых данных, что способно упростить и ускорить анализ и управление социальными рисками.

В рамках проведённого исследования был разработан инструмент для анализа новостных данных на основе модели BERT, который продемонстрировал высокую эффективность в выявлении тенденций, связанных с развитием социальных рисков.

Инструмент может быть применён для мониторинга социальных рисков в режиме реального времени, предоставляя аналитикам и государственным органам мощный инструмент для контроля, управления социальными рисками и поддержки принятия решений. Возможность гибкой настройки параметров анализа позволяет адаптировать инструмент под различные задачи, расширять список факторов и категорий. Важной особенностью инструмента является веб-интерфейс и модульная архитектура, которые делают его удобным и доступным для пользователей, а также упрощают его дальнейшее внедрение в системы мониторинга и анализа.

Таким образом, разработанный инструмент улучшает процесс анализа социальных данных и создаёт основу для дальнейшего развития и автоматизации технологий управления социальными рисками.

#### **Литература и источники:**

1. Бочарников И.В., Овсянникова О.А. Риски и вызовы информационной работы при проведении специальных военных и полицейских операций // Вестник Академии военных наук. 20224. № 1. С. 22–29.
2. Ворочаева В.А., Лотокова В.А. Влияние демографической ситуации на экономический рост России // Научные труды Вольного экономического общества России, 2019. – № 3. – С. 227–238.
3. Ивершинь А.В., Кожевина С.Г. Анализ факторов рождаемости в России // Интерэкспо Гео-Сибирь. 2021. № 4.
4. Информационная аналитика в современном социально-политическом процессе: теория и практика / Ремарчук В.Н., Бочарников И.В., Артемьев А.А., Галаганова С.Г., Гришнова Е.Е., Егоров В.Г., Карась Р.А., Катков О.Н., Ламинина О.Г., Смутьский С.В., Шевчун В.Н., Урсул В.И. Москва, 2024.
5. Калининкова М.В. Социальные риски в регионе. Учебное пособие. Саратов: Изд-во ООО Издательский Центр «Наука», 2016. – 96 с.
6. Кобакин М.В. Социальная диагностика рисков в информационно-телекоммуникационной сети интернет: сочетание современных технологий и тра-



диционных методов // *Инноватика и экспертиза: научные труды.* – 2019. – № 1 (26). – С. 130–138.

7. *Международные конфликты XXI века.* Зеленков М.Ю., Бочарников И.В. Учебник для студентов высших учебных заведений, обучающихся по укрупненной группе специальностей и направлений 41.03.00 «Политические науки и регионоведение» (квалификация (степень) «бакалавр») / Сер. Высшее образование: Бакалавриат. Москва, 2019.

8. Невежин В.П., Журавлева К.С., Свитина Е.О. Изучение влияния различных факторов на рождаемость в России с использованием эконометрического моделирования // *Экономика и социум.* 2015. № 6-4 (19).

9. Никулина М.А., Джамалова Б.В., Колодиев М.Ю., Шулова Е.Ю. Социальные риски в современном российском обществе: философская рефлексия // *СИСП.* 2018. – № 2-1. – С. 48–65.

10. Пашихина Е.В. Влияние социально-экономических факторов на рождаемость населения в российской федерации // *Хроноэкономика,* № 7 (28), 2020, С. 104–109.

11. Ремарчук В.Н. Управление смыслами как инструмент современной политики: технологии, вероятные последствия // *Этносоциум и межнациональная культура.* 2019. № 2 (128). С. 9–21.

12. *Украинский кризис в условиях трансформации современного миропорядка: тенденции развития, угрозы и вызовы для России /* Баранов В.П., Бартош А.А., Бочарников И.В., Дульнев П.А., Караваев И.Н., Кардаш И.Л., Карпович О.Г., Корабельников А.А., Кулаков А.А., Манойло А.В., Овсянникова О.А., Петренко А.И., Ремарчук В.Н., Стригунов К.С., Сурма И.В., Суханов П.В. (2-е издание, исправленное) Москва, 2022.

13. Филимонова И.В., Ивершинь А.В., Комарова А.В., Кривошеева О.И. Факторы влияния на решение о рождении ребёнка и количестве детей у женщин в России // *Народонаселение.* 2023. № 1 (26). – С. 55–69.

14. Цитульский А.М., Иванников А.В., Рогов И.Л. NLP– обработка естественных языков // *StudNet.* 2020. № 6. С. 467–475.

15. Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. // *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies,* pp. 4171–4186.

16. Rogers, A., Kovaleva, O., & Rumshisky, A. A primer in BERTology: what we know about how BERT works. // *Transactions of the Association for computational linguistics,* 2021, № 8, pp. 842–866.