***Метод оценки риска социального конфликта на основе критерия подобия***

***Актуальность.*** В настоящий момент оценка похожести городов (муниципальных образований) является актуальной проблемой и может применяться в различных задачах, затрагивающих инфраструктурные, архитектурные и другие вопросы [1, 2, 3, 4]. Интерес к данному исследовательскому направлению в первую очередь вызвано возможностью использования оценки подобия для решения задач планирования в градостроительной сфере [4]. В свою очередь в градостроительной сфере большое значение уделяется социальным конфликтам, возникающим в рамках реализации тех или иных проектов [5, 6]. Таким образом, прогноз социальных рисков является актуальной задачей как в градостроительной сфере, так и в области социально-экономического планирования.

Мотивы и факторы современных социальных конфликтов являются предметом серьезного изучения в связи с их большой значимостью на общественно-политическое состояние, которое в свою очередь неразрывно взаимосвязано с социально-экономическими процессами [7-11]. Поскольку множество механизмов данной сферы остаются неизвестными, дискуссионными либо же плохо изученными, актуальность исследований не вызывает сомнений, что также отражается в увеличении количества современных работе по рассматриваемой тематике [7, 8].

Целью исследования является разработка метода оценки и/или прогноза социального риска для произвольного муниципального образования. Суть предлагаемого подхода заключается в оценке похожести не только социально-экономического состояния, но и структуры населения (человеческого капитала) произвольного муниципального образования (МО) к некоторым показателям, характерным для возникновения различного типа социальных конфликтов. Проще говоря, основа подхода кроется в потенциале использования похожести между двумя МО, где один из них испытал социальный конфликт, как один из маркеров вероятной похожей социальной реакции как минимум на тот же раздражитель.

***Обзор проблематики.*** Существуют различные подходы изучения социальных конфликтов. До сих пор наиболее популярным является рассмотрение данной проблематики в контексте социологических и психологических исследований [7, 8]. Однако, необходимо отметить, что точность прогнозов, которые основаны только лишь на социологических данных, не демонстрируют высокую точность [7].

Также актуальным является подход, основанный на математическом моделировании коллективного поведения [8]. Разрабатываются различные агентные системы, позволяющие анализировать динамику протеста на основе процесса имитации. Например, популярен вариант представления протестных движений в контексте эпидемиологической модели [8].

Естественно, в последние годы набирает популярность использование машинного обучения в вопросе исследования и прогноза социальных конфликтов [9, 10, 11]. Эти исследования, как правило, базируются на анализе общественных настроений, которое осуществляется путём мониторинга социальных сетей и СМИ [9, 10, 11].

В работе [9] протесты исследуются на основе анализа стилистики статей СМИ, освещавших рабочие забастовки, которые имели место в Румынии (2018-2025). В контексте текущего исследования особый интерес вызывает данные в пользу гипотезы о «пространственной-временной» связанности протестов, т.е. протест в одном месте повышает вероятность протестов в соседних территориях [9]. Этот вывод перекликается с результативностью эпидемиологического подхода анализа динамики протестов. Однако, в работе предлагается подход, рассматривающий как минимум пространственную часть с точки зрения похожести экономического и демографического пространства территорий, которая зачастую хоть и вызвана географической близостью, но при этом не определяется только локационной составляющей.

В работе [10] для прогноза социального конфликта использовались данные из социальных сетей. В рамках исследования была предложена модернизированная нейронная сеть, которая обучалась на данных из соц. сети twitter (ныне X.com) на примере протестов в Гонконге 2019 года. Показана результативность данного подхода на примере исследуемых данных относительно других алгоритмов машинного обучения [10].

В исследование [11] рассматривались протесты в Латинской Америке (2012 – 2014) и в качестве данных использовались не только данные социальности сети twitter (ныне X.com), но и из других источников (новостные, базы данных политических акций, данные сети TOR, а также экономическая составляющая). Важно отметить, что в качестве экономического параметра использовалось только курс валюты, что хоть и является интересным параметром, тем не менее не является достаточным для отражения социально-экономического состояния некоторой территории внутри страны.

Несмотря на эволюцию протестов с точки зрения их мотивов (уход от чисто материалистических побуждений), значимость социально-экономических факторов до сих пор остаётся под весомым вниманием - “misery matters” [12, 13]. Так, в исследовании [12] рассматривается корреляция между ВВП и протестной активностью. Работа [13] исследует корреляции между социально-экономическим состоянием общим и индивидуальным (респондентов) с готовностью их участия в протестах. В работе [13] находятся подтверждения гипотезы о большем влиянии общей депрессивности, нежели индивидуальной. Такой вывод поддерживает текущее исследование, поскольку оценка похожести МО - это не столько про “misery matters”, сколько про “ environment matters”. В рамках исследования предлагается подход оторванный от субъективных данных, оценивающий похожесть не только по социально-экономическим индикаторам, но и демографической структуры (человеческий капитал).

***Общий набор данных.*** Исследуемый датасет включает в себя социально-экономические и демографические данные более 1800 муниципальных образований за период времени с 2014 по 2022 год. При этом данные МО за отдельный год являются независимыми примерами в датасете. Таким образом, итоговый набор включает в себя более 9000 примеров.

Все социально-экономические индикаторы, кроме показателей средних значений, нормированы на душу населения для лучшей оценки с точки зрения обеспеченности населения. Демографические данные также были нормированы как долевые значения конкретной когорты (возрастной группы) от общего числа населения определенного МО, что позволяет оценивать человеческий капитал с позиции структуры, а не количества.

Отобранные социально-экономические индикаторы представлены таблице 1. Этот набор факторов был выбран исходя из опыта их использования в задаче прогнозирования миграции, где была продемонстрирована их результативность. Таким образом, эти факторы позволяют адекватно оценивать социально-экономическое состояние муниципального образования.

1. Average number of employers in organizations - avgemployers (ppl.)

2. Average salary - avgsalary (rub.)

3. Shopping area - shoparea (sq.m.)

4. Number of seats in café, bars and restaurants - foodseats (num.)

5. Retail turnover - retailturnover (thnd. rub.)

6. Live area per capita - livarea (sq.m.)

7. Number of sporting venues - sportsvenue (num.)

8. Number of services (barbershops, repairs) - servicesnum (num.)

9. Length of roads - roadslen (km.)

10. Number of livestock - livestock (num.)

11. Productivity of land (vegetables) - harvest (centners)

12. Agricultural production - agrprod (thnd. rub.)

13. Number of healthcare organizations - hospitals (num.)

14. Number of places in preschool organizations - preschool (num.)

15. Volume of self-produced goods - factoriescap (thnd. rub.)

***Набор социальных конфликтов.*** Непосредственным источником показателей, характерных для социального конфликта, является состояние определенного МО в период возникновения конфликтной ситуации. Таким образом, фундаментальной составляющей рассматриваемого подхода является набор МО (конфликтов), относительно которых и будет оцениваться похожесть.

В настоящий момент сформирован набор наиболее резонансных социальных конфликтов в РФ в период с 2013 по 2023 год. Набор включает в себя 21 конфликт в различных муниципальных образованиях, которые сопровождались массовым собранием людей и широкой оглаской в СМИ.

Отбор именно острых конфликтов связан с тем, что их анализ имеет потенциал ярко продемонстрировать тенденции социально-экономических и демографических индикаторов, предрасполагающих к возникновению социальной напряжённости. То есть эти примеры можно рассматривать в качестве локальных «экстремумов» абстрактной функции, отражающей социальный конфликт.

На рисунке ниже проиллюстрированы демографические данные этих МО на момент социального конфликта (см. рис. 1). Отобранные конфликты были классифицированы по трём группам: «анти-мусорные», «этнические» и «анти-промышленные». Так, на рисунке 1 наблюдаются некоторые отличительные особенности конфликтов. Например, мун. образования, в которых имели место этнические конфликты, отличаются наибольшей долей молодежи среди населения (возрастные группы 20-24, 25-29 и 30-34, см. рис. 1).



Рис. 1. Сравнительный анализ демографической структуры на момент возникновения социального конфликта определенного типа

В муниципальных образованиях с «анти-мусорными» конфликтами наблюдается иная картина, а именно наименьшая доля молодежи в населении (см. рис. 1). При этом эти МО также отличаются высокой долей возрастного населения (группы 60-64, 65-69).

При рассмотрение социально-экономического состояния также можно наблюдать некоторые тенденции. Так, на рисунке 2 видно, что в развлекательной и сельскохозяйственной сфере МО, отличившееся этническим конфликтом, имеют наихудшие показатели (множество красных зон, низкое значение индикатора), тогда как в МО с другими конфликтами по этим показателям более благоприятная ситуация. В особенности это касается МО с «анти-мусорными» конфликтами, которые, судя по данным, наиболее сбалансированы (см. рис. 2).



Рис. 2. Сравнительный анализ социально-экономических факторов в год возникновения социального конфликта определенного типа

Таким образом, отличительные особенности между определенными конфликтами наблюдаются как с точки зрения социально-экономических показателей, так и демографической структуры. Несомненно, это является весомым доводом в пользу потенциала использования похожести этих компонент как меры риска возникновения социального конфликта.

***Критерий похожести.*** Каждое муниципальное образование в наборе данных определяется социально-экономической и демографической компонентой. Так, социально-экономическую составляющую можно представить в виде кортежа:

,

где это значение определенного социально-экономического индикатора (см. табл. 1).

Демографическая структура в свою очередь может быть представлена как два отдельных кортежа, отражающих долю людей определенного пола, следующего вида:

.

Здесь число женщин в определенной когорте, общее количество женщин в населении и *k* количество когорт. Представленное деление позволяет получить значение процентной доли женщин относительно всей популяции. Идентичное выражение используется и для представления мужчин:

,

где – число мужчин в определенной когорте, а общее количество мужчин в населении.

Для оценки подобия между двумя муниципальными образованиями предлагается использовать совокупную сумму среднеквадратического отклонения социально-экономических и демографических показателей:

где – значение социально-экономического фактора мун. образования *x, n* – количество факторов, – доля человек в когорте мун. образования *x, k –* количество когорт и – коэффициенты значимости для каждой компоненты.

В качестве примера рассмотрим использование этого критерия для поиска наиболее похожих МО на Красноармейский муниципальный район, где в станице Полтавской в 2022 году произошли широкие протесты против мусорного полигона **[ссылка на данные из СМИ].** Для этого необходимо провести оценку отклонения по критерию (1) всех МО из исследуемого датасета относительно Красноармейского муниципального района (МР) образца 2022 года.

Результаты оценки критерия представлены на рисунке 3 (слева), где все МО отсортированы по возрастанию их отклонения (непохожести) от Красноармейского МР. При этом значение критерия (1) было нормализовано в диапазоне от 0 до 1. Из таблицы видно, что второе место по похожести занимает Красноармейский МР, но образца 2017 года. Это подтверждает адекватность критерия, поскольку очевидно, что среди наиболее похожих ожидаемо наблюдать исследуемое МО, но за разные периоды времени. На рисунке 3 (справа) также представлен график, демонстрирующий значение критерия (1) для всех МО. Из графика видно, что диапазон отклонения критерия от 0 до 0,1 достигается к ~300 примеру, тогда как диапазон от 0,1 до 0,2 пересекается уже приблизительно к ~5000 примеру. Это говорит о том, что с точки зрения душевых показателей в РФ наблюдается однородность, т.е. существует большое количество МО не сильно отличных друг от друга с точки зрения экономического и демографического состояния, что в целом является нормальным в рамках одной страны. Данный графический анализ может также являться показателем уникальности исследуемого МО, поскольку чем более выпуклым будет начало кривой, тем меньше существует похожи МО.



Рис. 3. Оценка похожести на Красноармейский МР за 2022 год по критерию (1)

В качестве следующего шага необходимо проанализировать полученные результаты похожести социально-экономической и демографический среды в контексте идентификации потенциальных социальных конфликтов, что и является главной целью текущего исследования. К сожалению, в Курганинском МР (1-е место по похожести, см. рис. 3) не удалось найти значимых конфликтов в рамках анализа СМИ. Здесь важно отметить, что с точки зрения идеи метода важна не только похожесть, но и наличие схожего раздражителя.

При анализе Тихорецкого МР образца 2022 года (3-е место по похожести) удалось найти довольно масштабный социальный протест (см. рис. 4, слева), который произошел станице Алексеевской на фоне отключения водопроводной воды. На митинге приняли участие более 200 человек. Это существенное число участников как для протестных акций в России, так и для самой станицы, которая насчитывает чуть более 3000 человек.

Однако, в данном случае можно лишь говорить о совпадение социального напряжения в контексте похожести соц-эко. и демо. состояния, тогда как в первую очередь представляет интерес возможность похожей реакции на одинаковый раздражитель. И подобный случай произошел в пос. Каменный (население ~1000 человек, Тихорецком МР, см. рис. 4), где местные жители записали видеообращение с требованием разобраться с работой мусорного полигона. Тем не менее это произошло в Тихорецком МР образца 2024 года, которого нет в датасете.



Рис. 4. Митинги в Тихорецком муниципальном районе в 2022 году (слева) и 2024 (справа)

При анализе Белореченского МР за 2019 год (26-е место, см. рис. 3) был обнаружен социальный протест вокруг мусорного полигона, но в 2021 году (см. рис. 5). В митинге учувствовали около 500 человек. Белореченский МР образца 2021 также присутствует в датасете и занимает 50 место по похожести. Таким образом, от похожих МО в действительности можно ожидать схожей реакции на определенные социальные раздражители.



Рис. 5. Митинг в Белореченске за закрытие мусорного полигона в 2021 году

Как видно из примера, оценка похожести по даже по одному конфликту представляет интерес, поскольку даёт возможность оценить близость МО к какому-либо специфическому социальному конфликту. Однако, такой подход не позволяет проанализировать МО с точки зрения их потенциальной похожести на индикаторы, характерные большинству социальных конфликтов, что в свою очередь можно рассматривать как оценку социальной напряженности.

Таким образом, используя оценку похожести в качестве базиса, существует возможность реализовать подход, позволяющий оценить наличие муниципальных образований похожих сразу на несколько социальных конфликтов как одного типа, так и различного.

***Метод оценки относительной вероятности риска социального конфликта.*** В качестве основы для реализации совокупной оценки потенциала к социальному конфликту предлагается использовать систему ранжирования. Подобная система позволит осуществить некоторую рейтинговую оценку похожести всех муниципальных образований датасета относительно каждого социального конфликта по-отдельности. Сумму этих оценок в контексте критерия похожести будет являться мерой так называемого социального риска.

Упрощенная схема предлагаемого метода представлена на рисунке 6. Очевидно, что теперь наряду с полнотой и точностью критерия похожести, используемых факторов, а также качества самих данных, фундаментальное значение на результативность будет оказывать и принцип ранжирования. Существует множество способов выполнения ранжирования согласно уровню подобия.

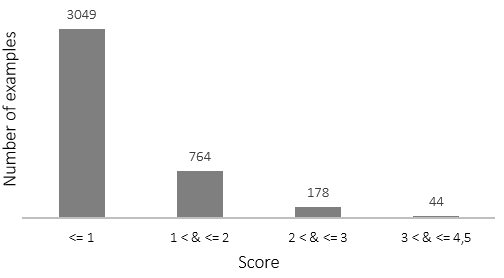
Так, наиболее простой вариант – выполнять прогрессивное ранжирование исходя из занятого места по похожести на основе шкалы от 1 (наиболее похожее, 1-е место) до 0 (наименее похожее, последнее место). Однако, из за некоторой однородности, которая наблюдается у мун. образ. в РФ, такой вариант будет давать слишком большое преимущество средним мун. образ., которые более или менее похожи на немалую группу других образований.

В логике текущего исследования, наиболее похожие мун. образования могут иметь схожую предрасположенность к конфликту. Анализ графиков с критерием похожести демонстрирует, что к более или менее близким можно отнести первые 300 похожих примеров (см. рис. 3). Таким образом, предлагается использовать ранжирование на основе топ-300, который заключается в следующей системе оценивания: 1 балл – оценка для МО с 1 по 100 место по похожести; 0.5 балл – за 100 по 200 место; 0.2 – балла за близость на уровне с 200 по 300 место.



Рис. 6. Упрощенная схема подхода по оценке социального риска

Полное распределение оценок системы ранжирования представлено на диаграмме. Почти половина примеров из датасета (4035 мун. образований) получила какую-либо оценку, т.е. оказалась в некоторой близости (4) к конфликтному образованию. Тем не менее подавляющая большинство примеров получила оценку равную 1 или меньше (3049 примеров, см. рис.), тогда как наиболее высокую совокупную оценку более 3 баллов получили 44 примера.



Итак, наивысшую суммарную оценку 4.5 получил Туапсинский муниципальный район за 2022 и 2021 год. В целом, высший балл 1 (вхождение в топ-100 по похожести) относительно конфликтного мун. образования 18 (Анапа) и 19 (Геленджик) не является удивтельным, поскольку все они входят в Краснодарский Край и являются близкими географически. Однако, как видно из таблицы, Туапсинский район получил 0 относительно Красноармейского МО (кейс 2), т.е. не вошел даже в топ-300 наиболее похожих, несмотря на то, что Красноармейский тоже находится в Краснодарском Крае и является локационно близким образованием для Туапсинского района.

В этой связи как раз больший интерес представляет Новоалтайск за 2021 год (3е место, см. табл.), который получил 0.5 балла относительно мун. образ. 18 (Анапа) и 0.2 балла относительно мун. образ. 19 (Геленджик), несмотря на удаленность от них более чем в 4000 километров (длина дороги). Эти результаты в очередной раз подтверждают один из тезисов исследования (см. секцию 2) согласно которой социально-экономическое и демографическое пространство не определяется локационной близостью.

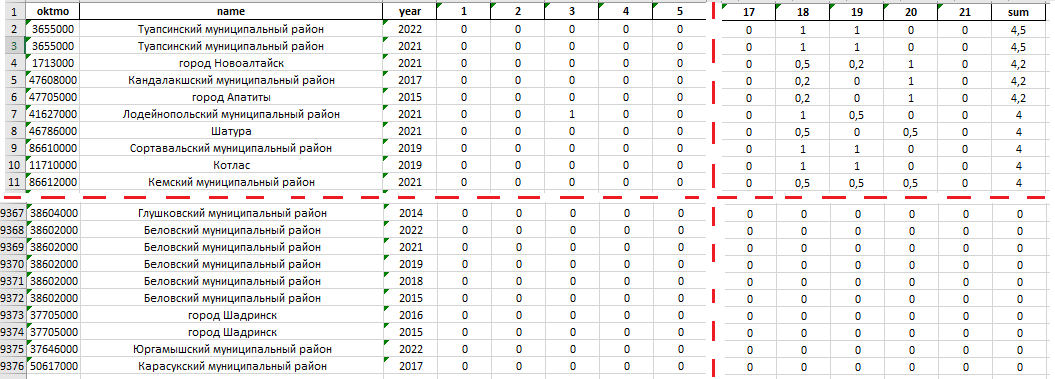


Рис. 7. Оценка относительного потенциала к социальному конфликту

Если в условиях оценки похожести на одно МО (см. раздел 4.1), в котором был определенный конфликт, исследовалась предрасположенность к конкретному социальному конфликту (похожий отклик на схожий триггер), то, как уже было отмечено, совокупная похожесть уже рассматривается как некоторый показатель потенциала конфликта как такового. Таким образом, в ходе анализа будут рассматриваться серьезные конфликты (сопровождающиеся массовым собранием людей) по любым темам.

В Туапсинском районе в 2021 году (второе место) в результате поиска действительно удалось найти один значимый политический протест в городе Туапсе (население ~60000 человек) в поддержку Алексея Навального (см. рис.). В 2022 году каких-либо протестов в данном районе не было зафиксировано. Следует отметить, что оценка демонстрирует вероятный потенциал, то есть для социального конфликта требует еще и мотив. И вот как раз серьезный социальный конфликт в Тупасинском районе произошел в 2024 году (к сожалению, в датасете отсутствуют данные за этот год), где несколько сотен человек вышли на протест против генплана в городе Джубга (население ~7000). Как раз против генплана протестовали в Анапе и Геленджике, на которых очень похож Туапсинский МО.

В городе Новоалтайск (население ~70000 человек) удалось найти два заметных социальный конфликта, которые произошли в 2018 и 2019 годах. Согласно оценке метода, Новоалтайск образца 2018 года также получил довольно высокую оценку потенциала социального протеста (2.9 балла, 75-е место). В 2018 году там случился социальный протест против повышения пенсионного возраста, а в 2019 году в Новоалтайске был протест против закрытия роддома. К сожалению, данных Новоалтайска за 2019 нет в датасете, поэтому отсутствует оценка уровня потенциала социального протеста.

В Канадалакшском муниципальном районе образца 2017 год (4е-место, 4.2 балла) в городе Кандалакша (~30000 человек) действительно был обнаружен серьезный социальный протест против двойной тарификации как раз в 2017 году. В протесте приняло участие более 100 человек.

Также был проанализирован и город Апатиты (население ~50000 человек). В рамках поиска удалось найти социальный конфликт в 2017 году, где более 100 учёных вышли на протест с требованием увеличить финансирование науки. По оценке потенциала социального протеста, город Апатиты в 2017 году занимает 13 место с оценкой 3,9.

Анализ демонстрирует, что метод оценки потенциала социального протеста способен обеспечивать выделение муниципальных образований, которые действительно могут отличаться социальной напряженностью и готовностью активного выражения недовольства. Рассмотренные протесты происходили в малых по меркам РФ городах (население меньше 100 тысяч) и при этом не имели широкого освещения в СМИ, поэтому обнаружение протестов на этих территориях может считаться показательным.

***Значимость факторов и прогноз динамики.*** Разработанный метод осуществляет детерминированное вычисление оценки социального риска по социально-экономическим и демографическим показателям за конкретный момент времени. Соответственно, нельзя вычислить оценку, например, за будущий год (или за тот по которому нет данных).

В условиях распространения использования машинного обучения как в целом, так и в прогнозе социальных конфликтов, следует рассмотреть возможность применения данного подхода в контексте реализованного метода оценки потенциала социального конфликта. Использование машинного обучения позволит, с одной стороны, оценить значимость набора факторов, а с другой, сделать прогноз динамики изменения оценки социального риска.

Поскольку датасет содержат данные по муниципальным образованиям за определенные годы, то простым вариантом имплементации машинного обучения является прогноз социального риска на следующий год, используя при этом социально-экономическим и/или демографическим данные за предыдущий год. Здесь важно отметить, что приятая система оценок ранжирования является простой и понятной для человека, но при этом будет малоэффективной для обучения прогностической модели. В связи с этим вместо принятых оценок, топ-300 следует представить шкалой от 1 до 0, где 1 место получает 1 балл, а 301 – 0, что сделает оценки более гладкими. Данный решение внесет несущественные изменения с точки зрения оценки муниципальных образований, но при этом повысит качество обучения моделей.

Предлагается разработать две независимые модели, предполагающие прогноз оценки потенциала отдельно на основе социально-экономических факторов и отдельно на основе демографических данных. Это сделано с целью сфокусировать внимание на значимости факторов, тогда как их совместное использование может перекрывать уровень значимости друг друга.

Итоговый датасет для обучения модели содержит приблизительно 7000 примеров. Сокращение количество примеров вызвано тем, что для обучения модели необходимо знать оценку потенциала социального конфликта за следующий год. Таким образом, все примеры 2022 года исключатся из датасета, поскольку нельзя рассчитать оценку за 2023 год. Это касается и определенных примеров, не имеющих данных за следующий год, даже за принятый период времени.

Для начальных исследований было принято решения использовать метод случайного леса. Это простой и популярный подход, который еще и позволяет определить значимость факторов на основе критерия Джини. С целью усилить надежность полученных оценок значимости факторов, а также точности модели, было реализовано 50 циклов обучения с вычислением средних оценок.

**Источники**

[1] A Similarity Approach to Cities and Features. 2022. URL: <https://www.researchgate.net/publication/358646626_A_Similarity_Approach_to_Cities_and_Features>

[2] Analysis of the Uniqueness and Similarity of City Landscapes Based on Deep Style Learning. 2021. URL: <https://doi.org/10.3390/ijgi10110734>

[3] Exploring venue-based city-to-city similarity measures. 2013. URL: <https://www.researchgate.net/publication/257212810_Exploring_venue-based_city-to-city_similarity_measures>

[4] The scenario method in urban planning. 2022. URL: <https://www.researchgate.net/publication/276321581_The_scenario_method_in_urban_planning>

[5] Are land use conflicts a barrier to sustainable city development? Evidence from Chattogram District of Bangladesh. 2025. URL: <https://www.researchgate.net/publication/388926768_Are_land_use_conflicts_a_barrier_to_sustainable_city_development_Evidence_from_Chattogram_District_of_Bangladesh>

[6] Conflicts in urban peripheries in Europe. 2023. URL: <https://www.researchgate.net/publication/373236871_Conflicts_in_urban_peripheries_in_Europe>

[7] Davydov D. A. (2022) Dynamics of Mass Protest Actions in Modern Russia: An Event Study. Monitoring of Public Opinion: Economic and Social Changes. No. 5. P. 72–93. https:// doi.org/10.14515/monitoring.2022.5.2199 (In Russ.)

[8] Petrovskii S, Shishlenin M, GlukhovA (2025) Understanding street protests: from amathematical model to protest management. PLoS ONE 20(4): e0319837. <https://doi.org/10.1371/journal.pone.0319837>

[9] Nicula, Alexandru-Sabin and Cretan, Remus and Simionescu, Mihaela and Oancea, Bogdan and Dragan, Alexandru, The City as Stage: Labour Protest, Sentiment, and Machine Learning. DOI: <http://dx.doi.org/10.2139/ssrn.5357390>

[10] Iyda, J.J., Geetha, P. An improved deep belief neural network based civil unrest event forecasting in twitter. Appl Intell 53, 5714–5731 (2023). https://doi.org/10.1007/s10489-022-03746-3

[11] Korkmaz, G., Cadena, J., Kuhlman, C.J. et al. Multi-source models for civil unrest forecasting. Soc. Netw. Anal. Min. 6, 50 (2016). <https://doi.org/10.1007/s13278-016-0355-8>

[12] Korotayev, A. V., Sawyer, P. S., & Romanov, D. M. (2021). Socio-Economic Development and Protests: A Quantitative Reanalysis. Comparative Sociology, 20(2), 195 222. <https://doi.org/10.1163/15691330-bja10030>