

## РАЗРАБОТКА ПОДХОДА К ОЦЕНКЕ СОЦИАЛЬНОГО САМОЧУВСТВИЯ НА ОСНОВЕ РЕЗУЛЬТАТОВ СЕНТИМЕНТ-АНАЛИЗА КОММЕНТАРИЕВ ПОЛЬЗОВАТЕЛЕЙ В СОЦИАЛЬНЫХ СЕТЯХ

Огурцова Е.В. (Национальный исследовательский университет ИТМО)

Научный руководитель – к.соц.н., начальник ОМИ ЦТЭП ИДУ Видясова Л.А.

(Национальный исследовательский университет ИТМО)

В данной работе приводятся возможные этапы оценки социального самочувствия на основе результатов сентимент-анализа комментариев пользователей в социальных сетях, раскрываются такие понятия как социальное самочувствие, сентимент-анализ. Автором приводятся субъекты оценки социального самочувствия в России и за рубежом, а также раскрываются недостатки применяемых методов сбора данных.

### **Введение.**

Социальное самочувствие один из важных показателей текущего эмоционального состояния общества. В средствах массовой информации можно встретить определение социального самочувствия как эмоциональной оценки людьми своего общественного положения, уровня удовлетворения социально-экономических и духовных потребностей. Также, некоторые исследователи, как, например, Бочканова Е.Н., определяют социальное самочувствие как целостную, относительно устойчивую эмоциональную реакцию субъекта на воздействие социальной среды и условия его жизнедеятельности. В данной работе под социальным самочувствием понимается субъективная оценка человеком окружающей его реальности, своего социального положения, динамики экономической и политической ситуации, а также ее влияния на настроение человека и принимаемые им решения.

На данный момент оценкой социального самочувствия в России занимаются такие исследовательские организации как Всероссийский центр изучения общественного мнения, Фонд развития гражданского общества, а также Левада-Центр. За рубежом более распространено понятие «благополучия», которое сходно по своему содержанию с социальным самочувствием, и его оценкой активно занимается Институт Гэллага (American Institute of Public Opinion). Упомянутые оценки проводятся с разной периодичностью, имеют разные наборы вопросов к респондентам, и в целом подход к оценке социального самочувствия у исследователей различается, но объединяют их применяющиеся при сборе данных методы — так называемые «конвенциональные» методы, включающие в себя телефонный опрос, личное интервью и прочее, что ещё можно условно отнести к стандартным социологическим методам. Одним из недостатков таких методов является их ресурсозатратность: требуются большие трудо- и времязатраты со стороны исследователя, а, кроме того, высок шанс ошибок со стороны интервьюеров и неискренности респондентов на полевом этапе. Таким образом на данный момент не существует устоявшегося подхода к оценке социального самочувствия посредством использования современных технологических решений, которые в свою очередь ускорили бы процесс сбора и обработки данных, позволили бы исследовательским организациям и органам власти получать актуальную информацию о текущем состоянии общества.

### **Основная часть.**

Перспективой в вопросе эмоционального анализа больших данных, по мнению Ненько А. и Петровой М., обладает анализ семантических данных — анализ текстов комментариев и хэштегов. Поскольку социальное самочувствие это всегда субъективная и эмоциональная оценка, то, безусловно, стоит рассмотреть анализ текстов, а именно анализ тональности комментариев пользователей, применительно к оценке социального самочувствия.

Анализ тональности текстов — это класс методов контент-анализа в компьютерной лингвистике, основная задача которого заключается в классификации текста по его

настроению. Наиболее распространены двухчастная классификация, где тексты классифицируются как негативный или позитивный, и трехчастная, где к уже упомянутым вариантам добавляется вариант классификации текста как нейтрального.

В общем виде алгоритм оценки социального самочувствия на основе результатов сентимент-анализа комментариев пользователей в социальных сетях можно описать следующим образом:

**1-й этап:** выбор источников данных о социальном самочувствии населения.

Поскольку социальное самочувствие присуще каждому отдельному индивиду, на первом этапе важно определить масштаб исследования — будет ли оцениваться социальное самочувствие на уровне страны, региона, города или, может быть, района. Исходя из масштаба исследования стоит выбирать источники данных. Так, например, для исследования регионального масштаба источниками могут быть группы ВКонтакте региональных органов исполнительной и законодательной власти, группы Центров управления регионом и прочие источники соответствующие уровню исследования. Кроме того, в выборку можно включать группы городов-столиц регионов.

**2-й этап:** сбор данных о социальном самочувствии населения.

В данной работе под данными о социальном самочувствии населения понимаются комментарии, которые оставляют люди в социальных сетях. На этапе сбора данных рекомендуется применять различные программы-парсеры, которые обращаются к веб-серверам для запроса данных и последующей их обработке — Scrapper API, Octoparse, DataOx, ScrapingBot, Wintr и др.

**3-й этап:** предварительная обработка данных.

Этап включает в себя удаление стоп-слов, токенизацию, стематизацию, частеречную разметку. Этот этап предполагает чистку и подготовку данных к анализу.

**4-й этап:** выбор подхода к определению тональности комментариев.

Принято выделять четыре основных подхода к определению тональности высказывания: подход, основанный на правилах; подход, основанный на словарях; машинное обучение с учителем; машинное обучение без учителя. Подходы на основе правил базируются на заданных вручную правилах классификации и эмоционально размеченных словарях. Правила обычно строятся на основе эмоциональных ключевых слов. Подходы на основе машинного обучения автоматически извлекают признаки из текста.

**5-й этап:** оценка социального самочувствия на основе результатов сентимент-анализа комментариев пользователей в социальных сетях.

В зависимости от имеющихся данных и выбранного подхода к определению тональности строится дизайн исследования. В целом можно говорить о вынужденном data-driven подходе в такого рода исследованиях, как существенном минусе и сложности для исследователей: исследователь должен отталкиваться не от теории и операционализации понятий, а от имеющихся данных.

## **Выводы.**

Таким образом описанный подход может быть применен в исследованиях социального самочувствия населения, а также учтен при разработке сервиса оценки социального самочувствия населения на основе анализа комментариев пользователей из социальных сетей.