

## **Применение методов машинного обучения для автоматизированного анализа форм взаимодействия горожан с городской средой**

**Холмогоров И.М.** (Федеральное государственное автономное образовательное учреждение высшего образования “Национальный исследовательский университет ИТМО”, Санкт-Петербург)

**Научный руководитель – к.т.н., Митягин С.А.** (Федеральное государственное автономное образовательное учреждение высшего образования “Национальный исследовательский университет ИТМО”, Санкт-Петербург)

В рамках данной работы предлагается перспективный метод автоматизированного анализа форм взаимодействия горожан с городской средой. Входными данными для него являются комментарии из социальных сетей, оставленные жителями определенной локации. Сущность метода состоит в применении машинного обучения для декомпозиции текстового массива на смысловые элементы.

### **Введение.**

В настоящее время основным недостатком применяемой методики полевых исследований городского пространства является необходимость задействовать значительные кадровые ресурсы на протяжении долгого времени, причем их объем прямо зависит от площади исследуемого участка. Между тем, предлагаемый в рамках данной работы метод, основанный на применении машинного обучения для анализа активности населения в социальных сетях, позволяет решать аналогичные задачи удаленно и с минимальными необходимыми трудозатратами. Данная технология имеет потенциал дополнить ныне доступный массив информации о городской среде новыми данными.

### **Основная часть.**

Метод состоит из следующих этапов: сбор данных, обработка данных, семантический анализ текста, привязка к городским сущностям. На этапе сбора данных выбирается сообщество в одной из социальных сетей, посвященное определенной локации, например, отдельному микрорайону. Из этого сообщества с помощью официальных API выгружаются комментарии пользователей за определенный временной промежуток. Полученный массив текстовых данных проходит предварительную обработку, в ходе которой дробится на отдельные массивы, сообщения разбиваются на предложения, проводится очистка от шума (рекламные сообщения, пустые сообщения, сообщения, состоящие из “смайликов” и “эмодзи”). Строки обработанного массива подвергаются декомпозиции на смысловые элементы с помощью семантического анализатора текста на базе SyntaxNet; происходит выделение слов, соответствующих формам взаимодействия людей с городской средой. Методом поиска по ключевым словам часть слов-действий привязывается к определенным городским сущностям согласно контексту их исходных предложений.

### **Выводы.**

Таким образом, метод позволяет в короткие сроки получить релевантную информацию о жителях и территории для последующего ее использования на начальных этапах планирования преобразований или в социальных исследованиях.