

ИСПОЛЬЗОВАНИЕ ML-МОДЕЛИРОВАНИЯ В ИССЛЕДОВАНИИ СОЦИАЛЬНЫХ РЕАКЦИЙ НА ИНФОРМАЦИОННОЕ ВОЗДЕЙСТВИЕ

Якимова А.И., студентка группы СГНЗ-41Б

Московский государственный технический университет им. Н.Э. Баумана

*Научный руководитель: Катков О.Н., кандидат исторических наук,
доцент кафедры «Информационная аналитика и политические технологии»*

Аннотация: Данная статья рассматривает применение машинного обучения для исследования реакций, формируемых в результате информационного воздействия на социальные группы. Описываются основные принципы ML-моделирования, его преимущества и ограничения. Приводится пример использования кластеризации для анализа социальных реакций.

Ключевые слова: Машинное обучение, социальные реакции, Data Science, информационное воздействие, кластеризация.

В современном мире социальные реакции на информационное воздействие играют ключевую роль в функционировании общества. Они определяют взаимодействие между людьми, организациями и государствами. Исследование социальных реакций является актуальной задачей, которая требует применения современных методов и технологий. Одним из таких методов является ML-моделирование.

Машинное обучение (англ. Machine learning, ML) – класс методов искусственного интеллекта, характерной чертой которых является обучение на массивах данных, формируемых в ходе решения множества сходных задач [5].

Процесс машинного обучения включает в себя несколько этапов:

1. **Сбор данных.** На этом этапе собираются данные, которые будут использоваться для обучения модели. Чем больше будет информации, тем более точным будет результат.

2. **Предварительная обработка данных.** Данные, собранные на первом этапе, часто нуждаются в предварительной обработке, например, очистке от шума, кодировании переменных, разделении данных для удобного дальнейшего использования.

3. **Выбор модели.** На этом этапе выбирается модель, которая будет использоваться для обучения. Существует множество моделей машинного обучения, таких как линейная регрессия, деревья решений, нейронные сети и т. д. Выбор модели зависит от типа данных, специфики задачи и требований к ней.

4. **Обучение модели.** Происходит обучение на основе собранных данных. Сначала задается процент данных от дата сета, которые

будут использованы для обучения, оставшаяся часть станет основой для оценки качества обученной модели.

5. **Оценка модели.** После обучения модель необходимо оценить, чтобы определить её точность и эффективность. Для оценки модели используются различные метрики, такие как точность, полнота, F-мера и так далее.

6. **Применение модели.** Если модель удовлетворяет требованиям точности и эффективности, она может быть применена для прогнозирования значений целевой переменной на новых данных.

Процедуры выбора, обучения и оценки модели повторяется до тех пор, пока результат ее применения не достигнет необходимой точности и эффективности.

Модели, созданные с использованием метода машинного обучения способны стать мощным инструментом для анализа социальных реакций. Они позволяют исследовать поведение пользователей информационных каналов, в том числе социальных сетей, их эмоции и настроения, а также выявлять закономерности и тенденции реагирования, возникающие в ответ на информационное воздействие.

Социальные действия (реакции) – это определённая система поступков, средств и методов, используя которые, индивид или социальная группа стремятся изменить поведение, взгляды или мнение других индивидов или групп [2].

В рамках настоящей работы предлагается выделять два вида социальных реакций:

1. **Внешние социальные реакции** – поведенческие, то есть проявляющиеся в действиях и поступках человека.

2. **Внутренние социальные реакции** – эмоциональные и когнитивные, то есть связанные с чувствами, эмоциями и мышлением человека.

Примером применения ML-моделирования в исследовании социальных реакций может послужить кластеризация процесса «действие – реакция». Методом кластеризации выбран DBSCAN (Density Based Spatial Clustering of Applications with Noise) – плотностной алгоритм пространственной кластеризации с присутствием шума, популярный алгоритм кластеризации, используемый в анализе данных в качестве одной из замен метода k-средних. Преимущество метода в том, что он не требует предварительных предположений о числе кластеров. Однако для его применения необходимо настроить два дополнительных параметра: `eps` и `min_samples`. Данные параметры – это соответственно максимальное расстояние между соседними точками и минимальное число точек в окрестности (количество соседей), когда можно говорить, что эти экземпляры данных образуют один кластер.

Данный метод возможно интерпретировать под анализ социальных процессов – двумерная выборка данных будет содержать в себе идентификатор субъекта, воспринимающего информационное воздействие (ось X) и его реакцию на действие (ось Y).

Шум будет отображать погрешности системы и человеческий фактор – аномальную реакцию принимающего информацию субъекта, не соответствующую формирующемуся в результате информационного воздействия эмоциональному или ментальному состоянию.

Параметр `eps` отражает допустимый уровень плотности выявленных групп, по достижению которого признается факт формирования кластера.

Параметр `min_samples` – минимальное количество человек в кластере. Кластеры – группы людей, которые среагировали на информационное воздействие сходным образом. На основе этих параметров был написан код на языке Python для построения кластера (рис. 1).

```
import numpy as np
import matplotlib.pyplot as plt
from sklearn.cluster import KMeans, AffinityPropagation, MeanShift, AgglomerativeClustering, DBSCAN
from sklearn.datasets import make_blobs
import os

file_path = "C:\\Users\\Mi\\Downloads\\s1.txt"
ram_capacity = 2**20

if file_path.endswith('.txt') and os.path.isfile(file_path) and os.path.getsize(file_path) <= ram_capacity:
    data = np.loadtxt(file_path)

dbscan = DBSCAN(eps=60000, min_samples=200)
dbscan.fit(data)
dbscan_clusters = dbscan.labels_

plt.figure(figsize = (20, 10))

plt.scatter(data[:, 0], data[:, 1], c=dbscan_clusters, cmap='jet', marker='o')

plt.xlabel("X")
plt.ylabel("Y")
plt.title("Центроиды кластеров на основе модели DBSCAN")
plt.grid(True)

plt.show()
```

Рис. 1. Пример кода кластеризации на языке Python

В результате работы предложенной программы в анализируемом тестовом массиве данных выделено 15 кластеров – социальных групп, распределенных по реакциям. Не входящие в кластеры точки – люди, которые нельзя отнести ни к одной из групп, так как продемонстрированные ими реакции слишком далеки от доминирующих в выделенных кластерах.

В ходе исследования было подтверждено, что машинное обучение можно рассматривать в качестве перспективного инструмента анализа и прогнозирования социальных реакций. Оно позволяет выявлять скрытые закономерности в больших объёмах данных и использовать их для принятия обоснованных решений.

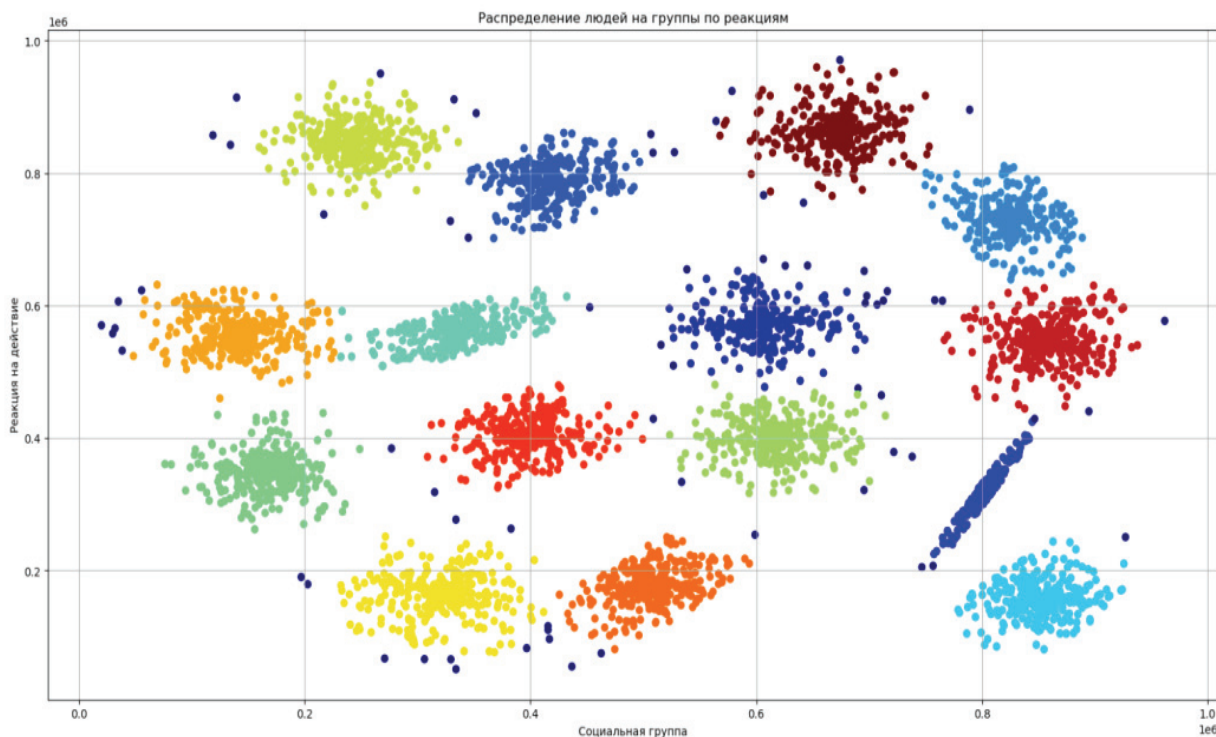


Рис. 2. Разделение людей на группы по реакциям

В статье было рассмотрено, как ML-моделирование может быть применено для изучения социальных реакций, а также приведены примеры использования этого метода. В частности, было показано, как кластеризация может помочь выделить группы людей с похожими реакциями на определённые события или стимулы.

Таким образом, ML-моделирование, представляя собой перспективный инструмент исследования социальных реакций и может стать важным элементом в арсенале социологов, маркетологов и других специалистов, работающих с данными о поведении людей.

Литература и источники

1. Айвазян С.А., Енюков И.С., Мешалкин Л.Д. Прикладная статистика: основы моделирования и первичная обработка данных. М.: Финансы и статистика, 1983.
2. Громов И.А., Мацкевич А.Ю., Семёнов В.А. Западная социология. СПб.: ООО «Издательство ДНК», 2003. С. 532.
3. Информационная аналитика и информационно-аналитические технологии в контексте социального управления. МГТУ имени Н.Э. Баумана. Москва, 2023.
4. Катков О.Н. Специфика организации практической подготовки специалистов в области информационной аналитики в условиях увеличивающегося технологического многообразия. В сборнике: Информационная аналитика и информационно-аналитические технологии в контексте социального управления. Москва, 2024. С. 51–56.

5. Лекун Я. Как учится машина. Революция в области нейронных сетей и глубокого обучения. М.: Альпина нон-фикшн, 2021.
6. Парсонс Т. О структуре социального действия. М.: Академический проект, 2000.
7. Флах П. Машинное обучение. М.: ДМК Пресс, 2015. 400 с.
8. Формы, методы и технологии профилактики и противодействия проникновению идеологии экстремизма и терроризма в образовательную среду / Ремарчук В.Н., Ремарчук В.Н., Бочарников И.В., Семикин Г.И., Белозёров В.К., Овсянникова О.А., Смутьский С.В., Зеленков М.Ю., Гейреханов Г.П., Галаганова С.Г., Пятибратова И.В., Худышева М.К., Пиканина Ю.М., Миронов А.С., Лазарев С.В. Москва, 2019.