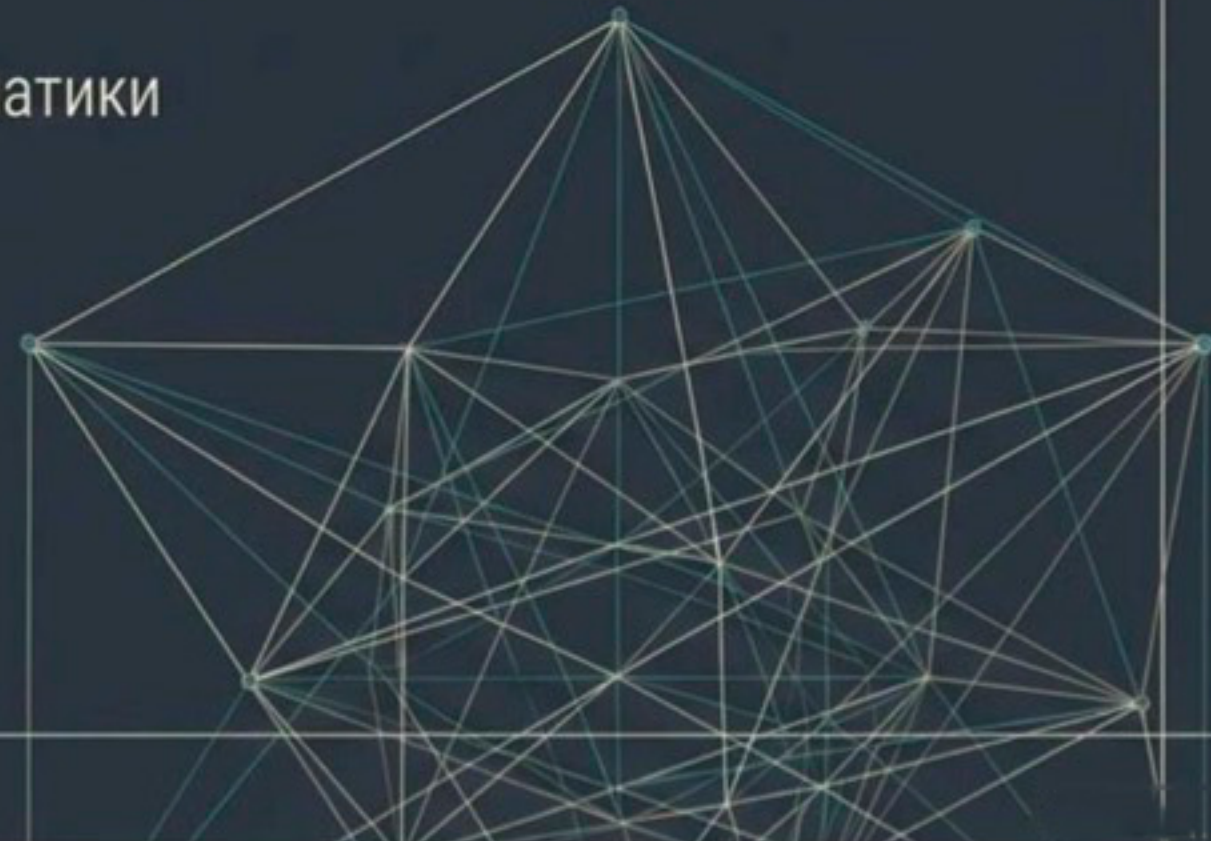
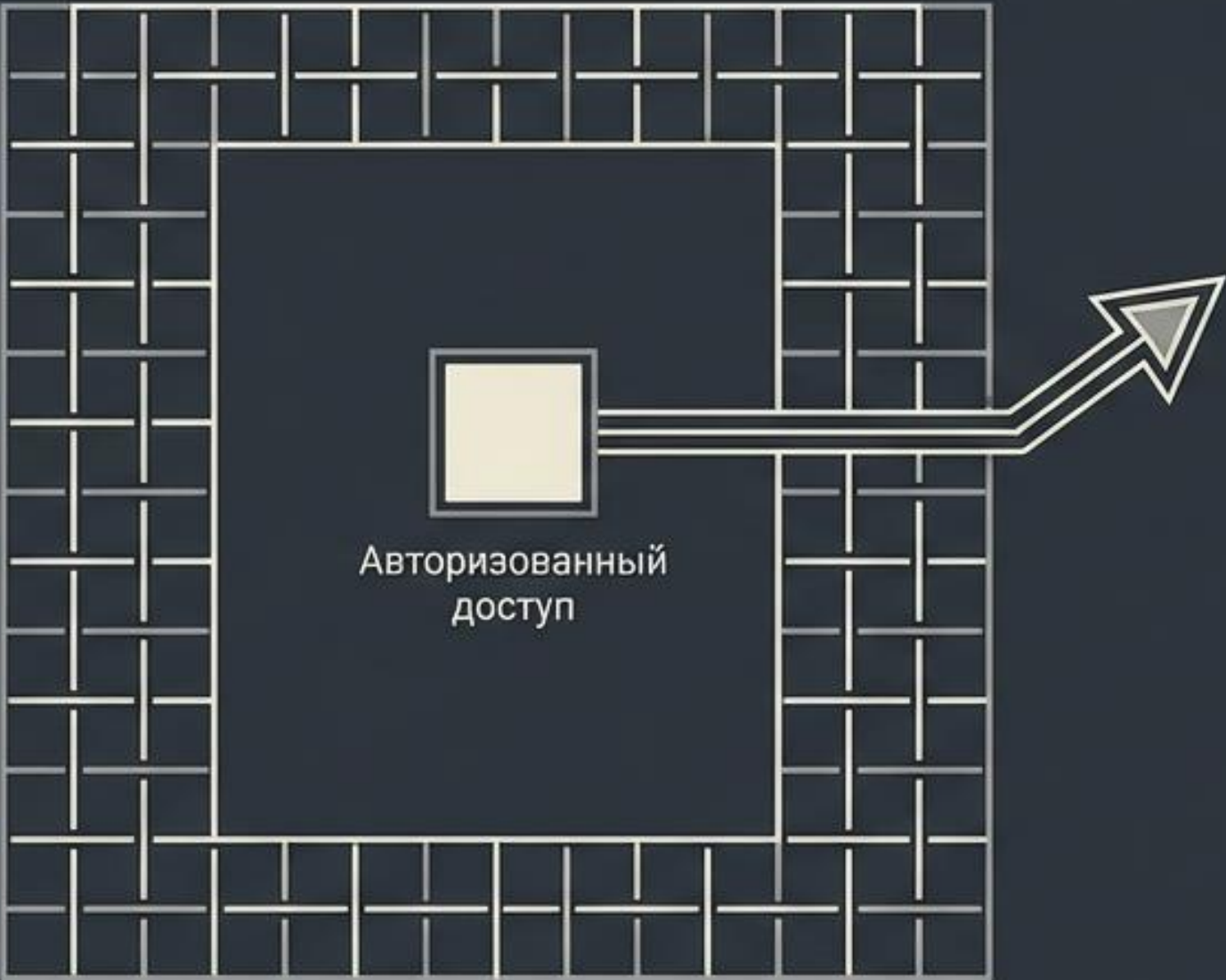


Обнаружение редкой аномальной активности пользователей информационной системы методами машинного обучения

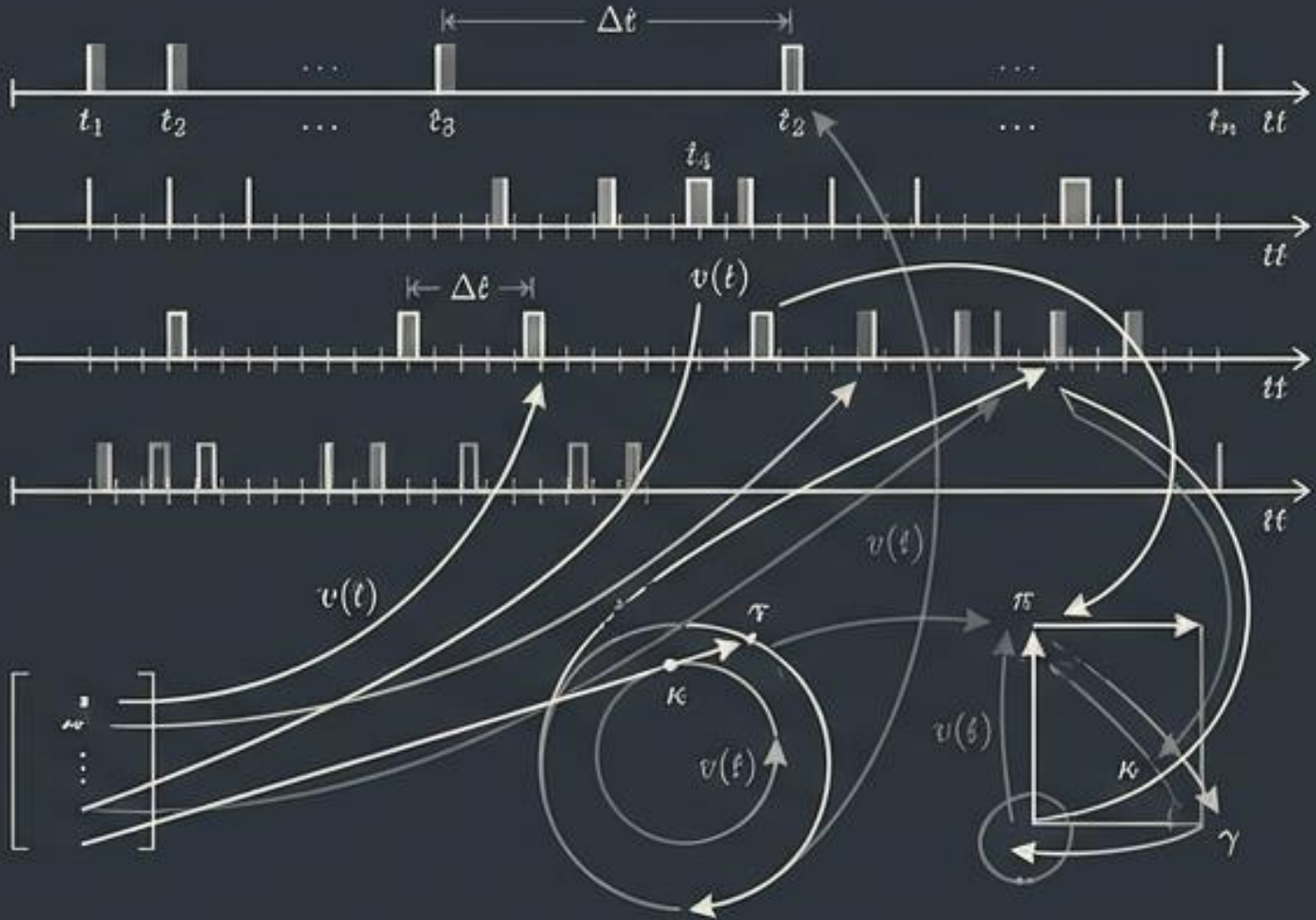
О.И. Шелухин, А.В. Осин
Московский технический университет связи и информатики



Скрытая угроза: Ограничения традиционной защиты

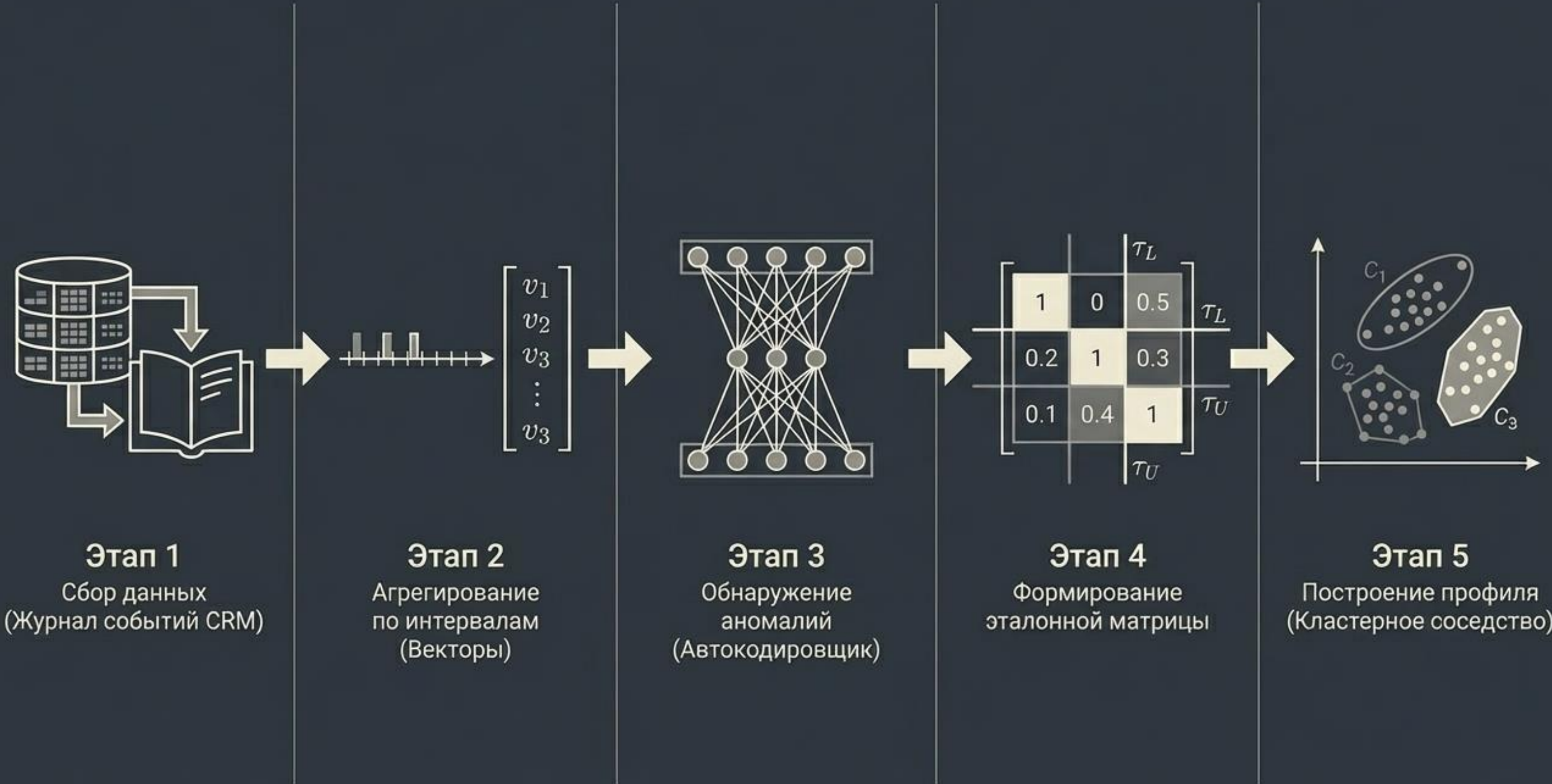


ПРОБЛЕМА Инсайдеры обладают легитимным доступом к CRM-системам. Традиционная периметральная защита неэффективна против авторизованных пользователей, знающих внутренние процессы.



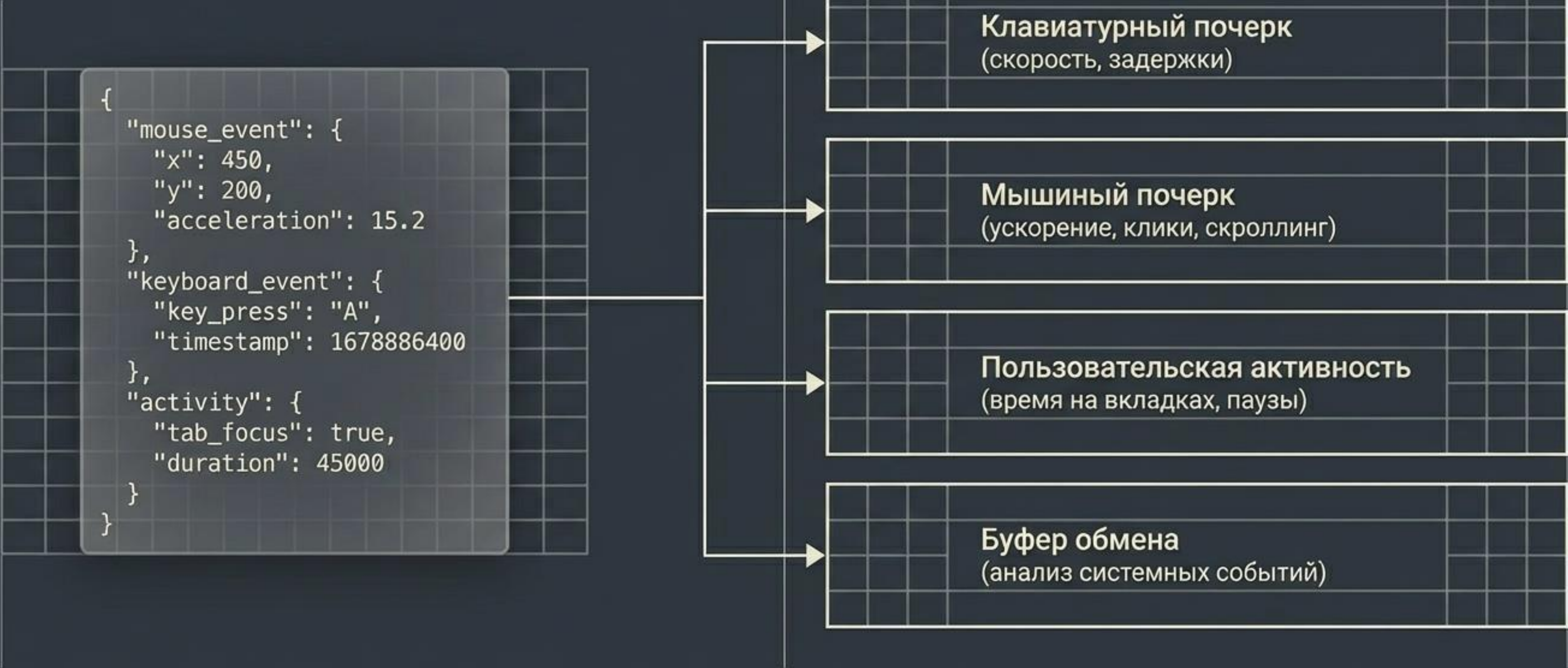
| | |
|----------------|---|
| ЗАДАЧА | Выявление скрытых отклонений от нормального поведенческого профиля в условиях, когда действия формально допустимы, но структурно аномальны. |
| РЕШЕНИЕ | Непрерывный анализ цифрового следа с использованием архитектуры глубокого обучения (Автокодировщики) и топологического кластерного анализа. |

Обобщенный алгоритм обнаружения аномального поведения



Этап 1: Сбор сырых данных и парсинг журналов

Механика: Фиксация действий пользователя в CRM-системе в реальном времени через сниффер в формате JSON.



Математическая формализация векторов признаков

Мышь: $\{a_i^{\text{ДМ}}\}, i=1..I, I=11$

(координаты, среднее ускорение, клики)

Активность: $\{a_k^{\text{Да}}\}, k=1..K, K=7$

(доли активного/неактивного времени)

$$A = \left\{ \underbrace{a_i^{\text{ДМ}}, a_j^{\text{ДК}}}_{\text{Клавиатура}}, \underbrace{a_k^{\text{Да}}, a_l^{\text{ДБ}}}_{\text{Буфер и события}}, a_m^{\text{ДВ}} \right\}$$

Клавиатура: $\{a_j^{\text{ДК}}\}, j=1..J, J=7$

(скорость нажатия, длительность)

Буфер и события: $\{a_l^{\text{ДБ}}\}, \{a_m^{\text{ДВ}}\}$

(содержимое, системные вызовы)

Этап 2: Агрегирование и стабилизация шкал



Агрегирование: Преобразование асинхронного потока событий в унифицированные 5-минутные интервалы. Формируется многомерный вектор.

Нормализация признаков:

Roboto Light, vs light ash gray.

Необходима для стабилизации диапазонов (пиксели/сек vs миллисекунды) перед подачей в нейронную сеть.

$$\tilde{x}_{ij} = \frac{x_{ij} - \min(X_{\cdot j})}{\max(X_{\cdot j}) - \min(X_{\cdot j})}$$

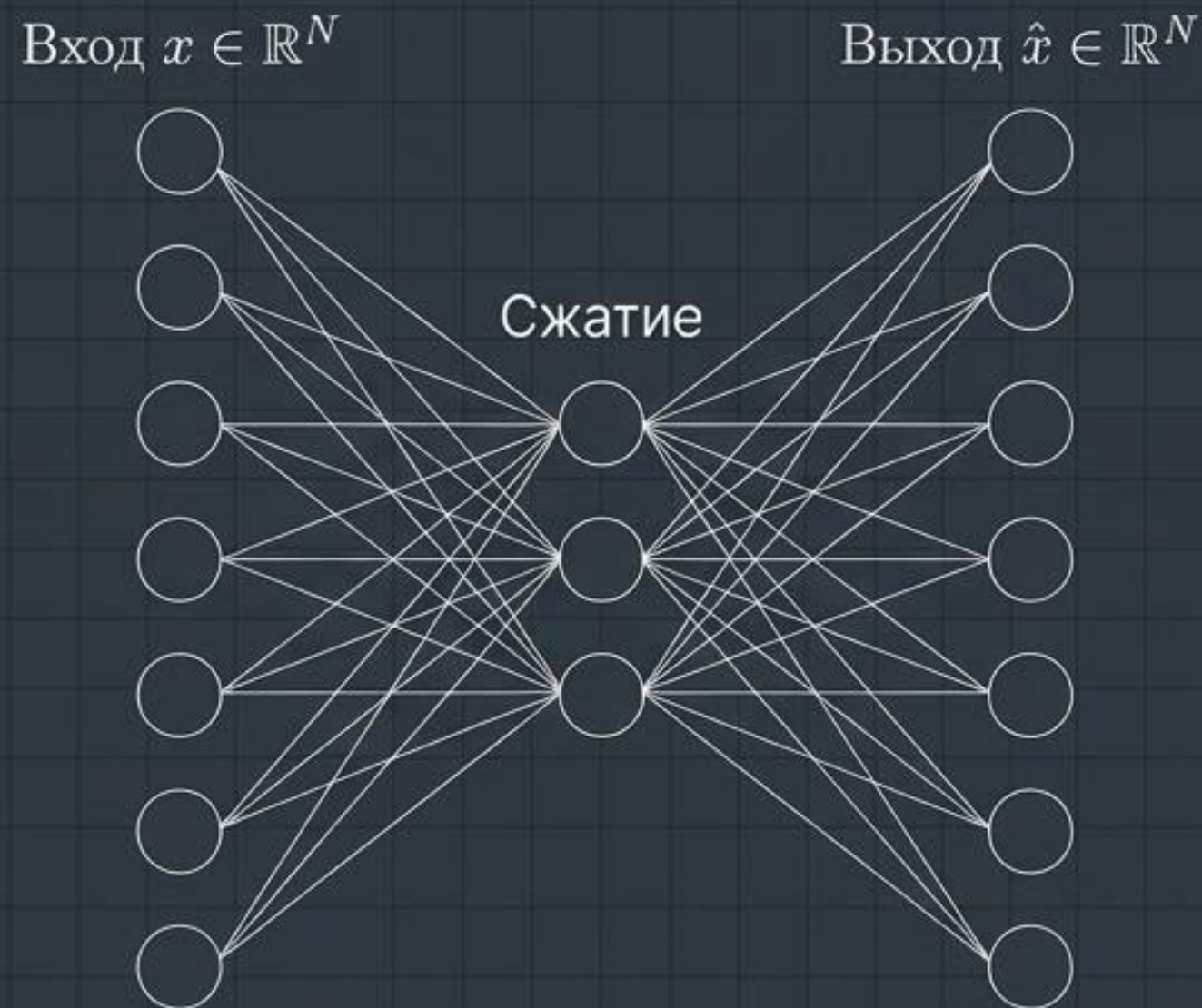
Этап 3: Архитектура обнаружения аномалий (ИНС)

Концепция:

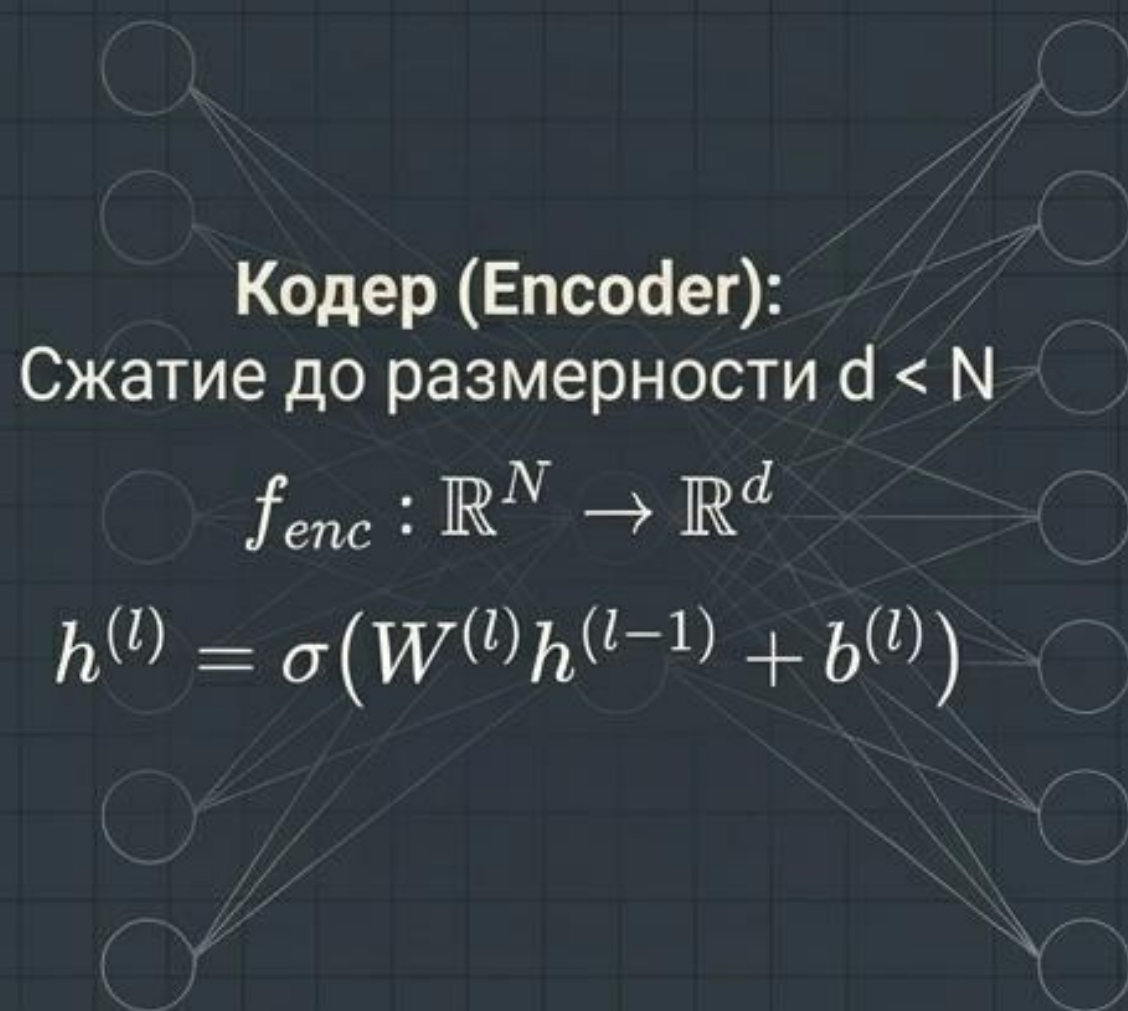
Применение искусственных нейронных сетей (Автокодировщиков) для обработки агрегированных данных каждой вкладки CRM.

Логика выявления:

1. Автокодировщик обучается сжимать рутинные (нормальные) данные в скрытое пространство низкой размерности и восстанавливать их.
2. При подаче аномального вектора сеть не может корректно его реконструировать.
3. Неудачная реконструкция генерирует высокую математическую ошибку на выходе.

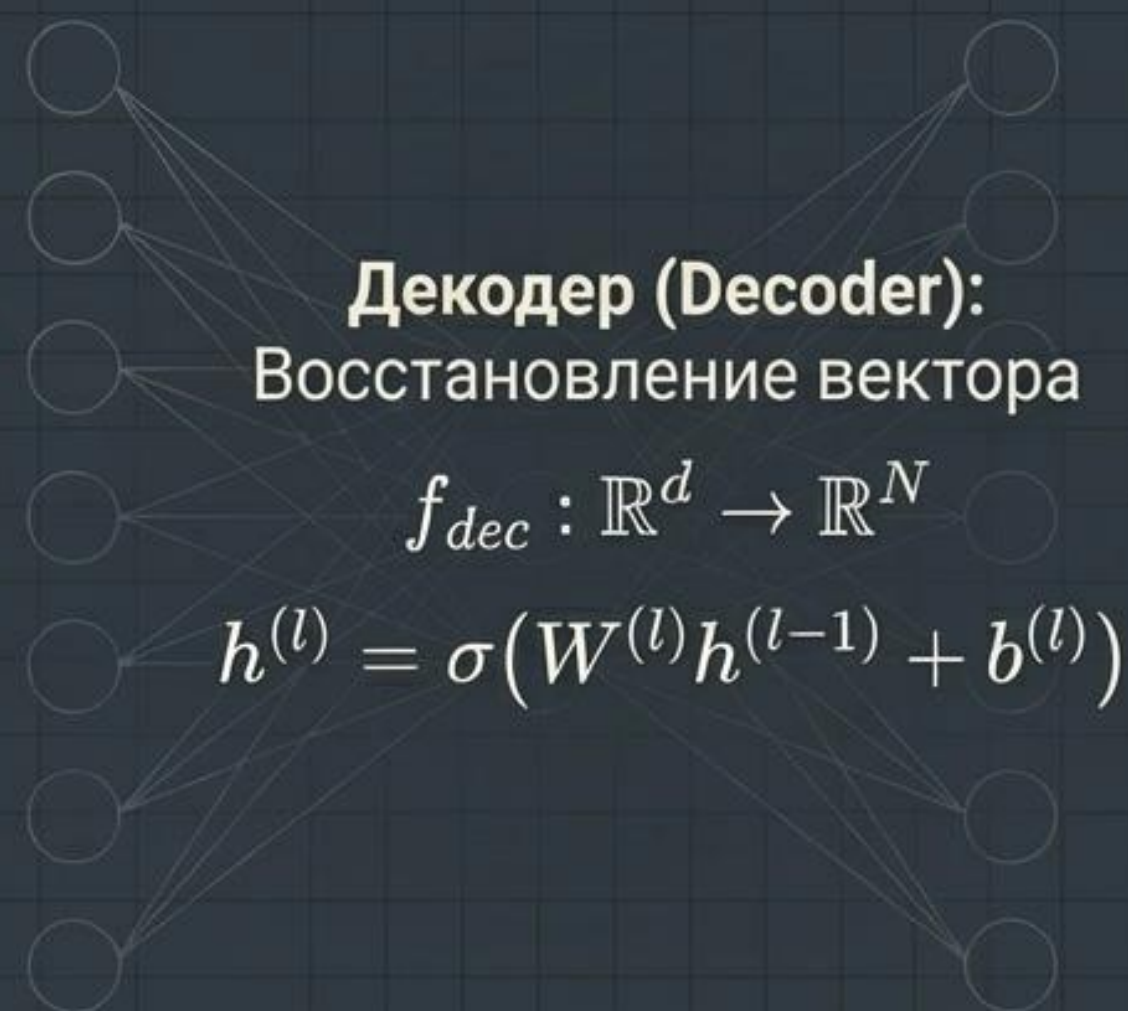


Математический аппарат Автокодировщика



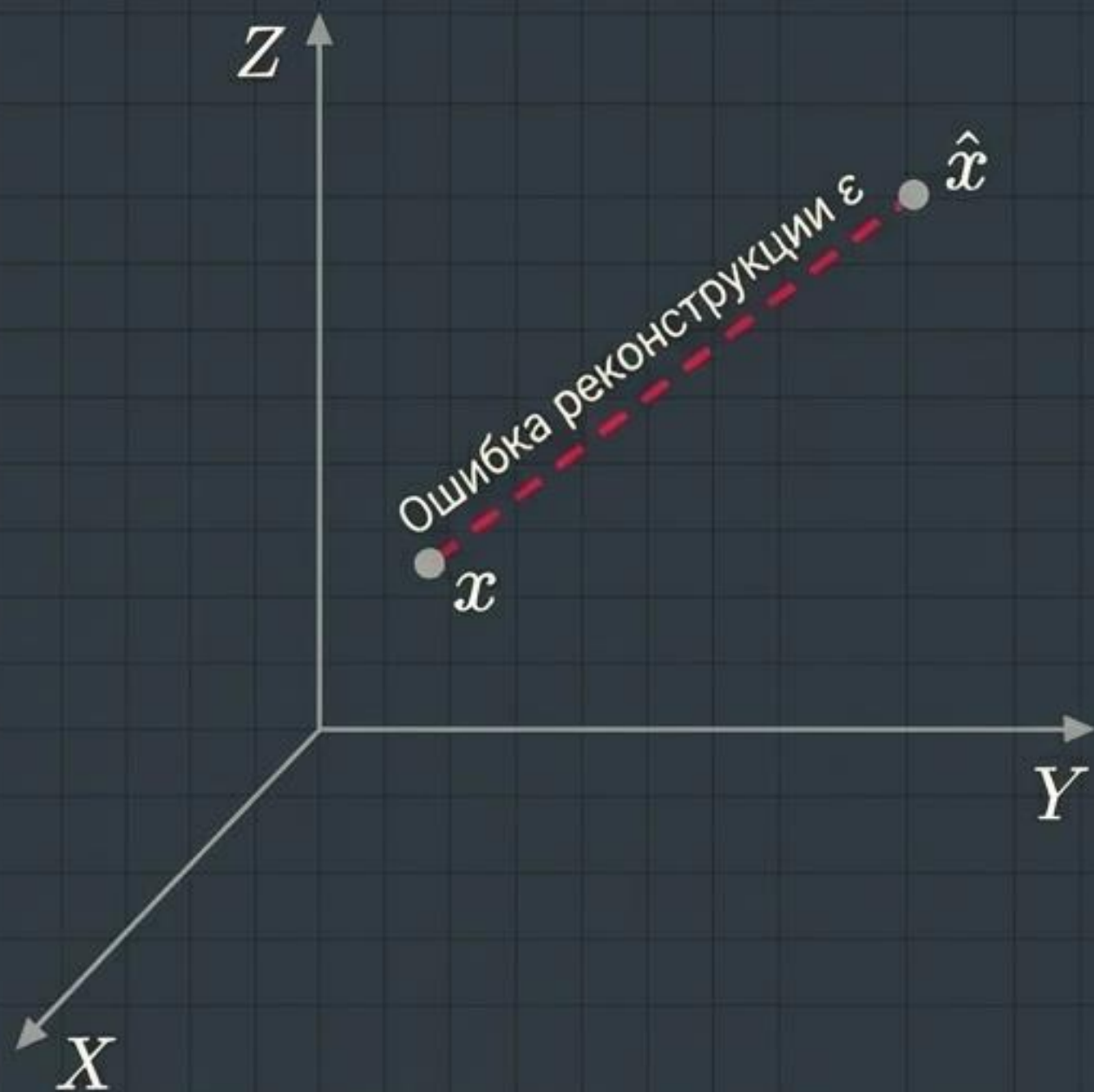
**Скрытое
представление
(Bottleneck)**

Вектор $z \in \mathbb{R}^d$
(Латентное
пространство
признаков)



Функции активации σ : Используются нелинейные функции (ReLU, tanh, sigmoid), симметричные для декодера.

Оценка аномальности: Функция потерь



Среднеквадратичная ошибка (MSE):
Функция потерь выступает не только метрикой обучения, но и фактическим маркером отклонения.

$$L(x, \hat{x}) = \frac{1}{N} \sum_{i=1}^N (x_i - \hat{x}_i)^2$$

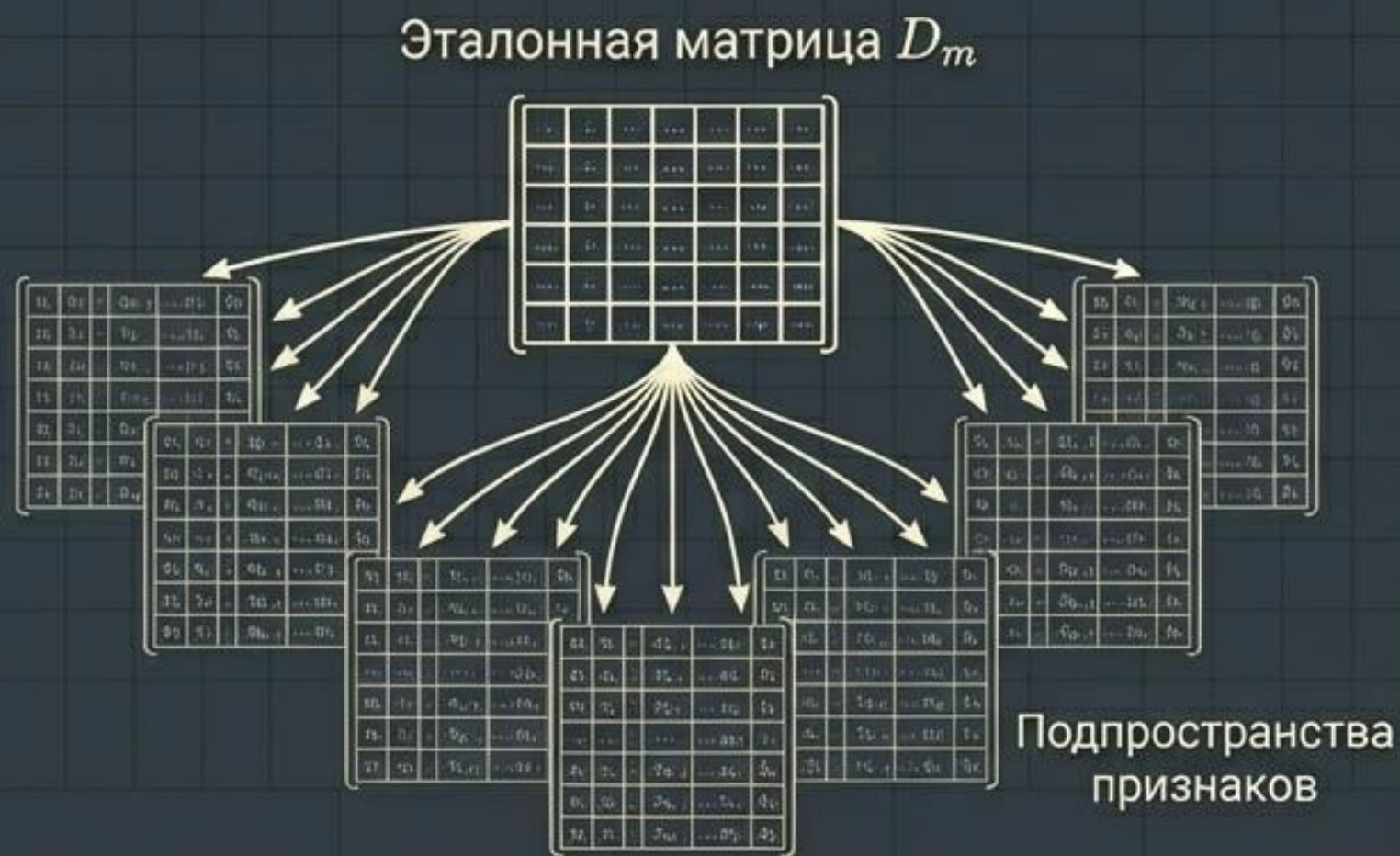
Интерпретация:
Ошибка реконструкции $\varepsilon_i = L(x_i, \hat{x}_i)$. Чем выше значение ε_i , тем сильнее исследуемый 5-минутный интервал отклоняется от усвоенной "нормы" поведения пользователя.

Этап 4: Адаптивная классификация состояний



Этап 5: Бутстрэп и генерация подпространств

Сборка базы: Формирование сбалансированной эталонной матрицы $D_m = \{(x_i, y_i) \mid y_i = m\}$, где $m \in \{0, 1, 2\}$.



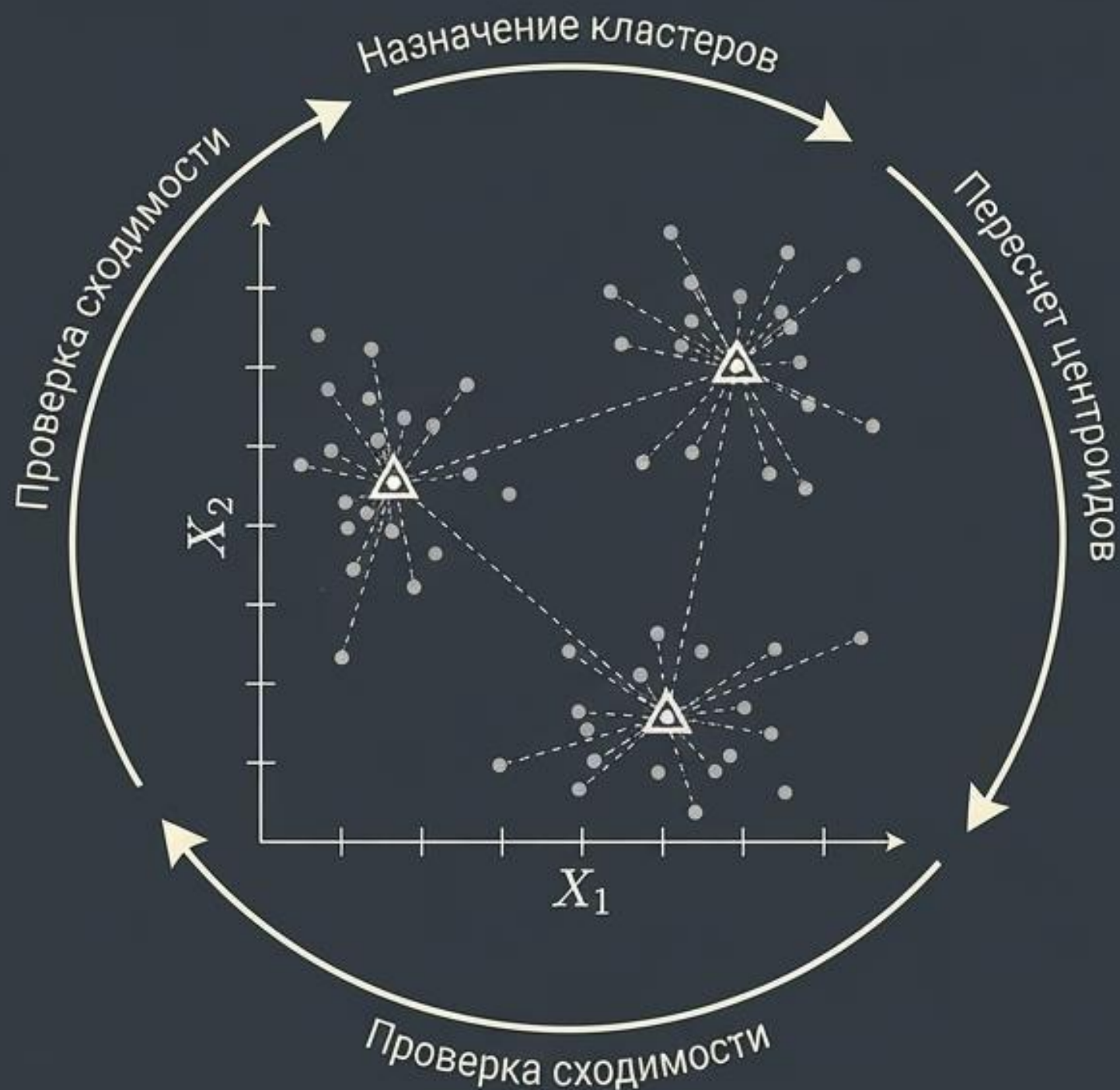
Метод Бутстрэпа (Bootstrap): Для выявления скрытых корреляций анализируются различные комбинации признаков.

Математика комбинаторики: Множество комбинаций из N атрибутов длины от K_{\min} до K_{\max} :

$$\Omega = \bigcup_{k=K_{\min}}^{K_{\max}} \mathbb{C}_N^k$$

Каждая комбинация формирует уникальное подпространство для поиска аномального 'соседства'.

Кластерное соседство: Итеративный алгоритм K-means



Кластеризация подматриц D_C : Поиск структуры внутри каждой комбинации признаков.

Целевая функция минимизации квадратов расстояний:

$$J(M, U, C) = \sum_{i=1}^c \sum_{j=1}^d u_{ij} d^2(m_j, c_i) \rightarrow \min$$

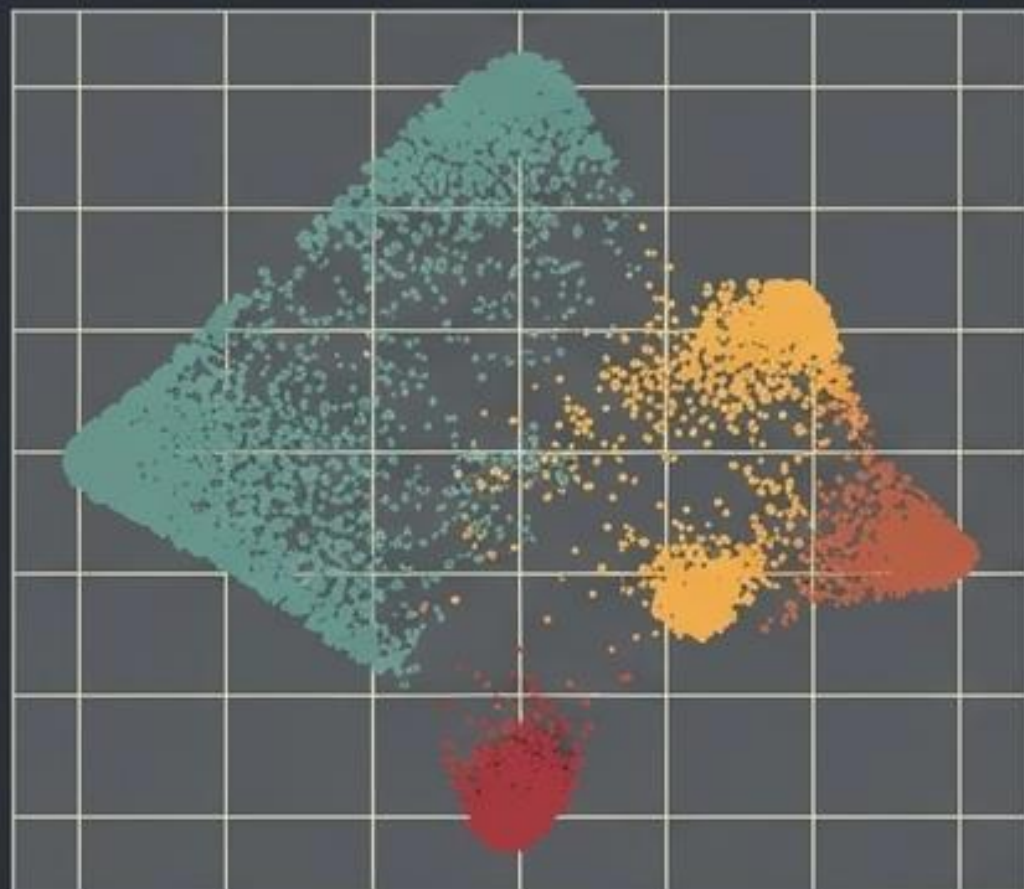
Матрица принадлежности U :

$$u_{ij} = 1, \text{ если } d(m_j, c_i) = \min_k d(m_j, c_k), \text{ иначе } 0.$$

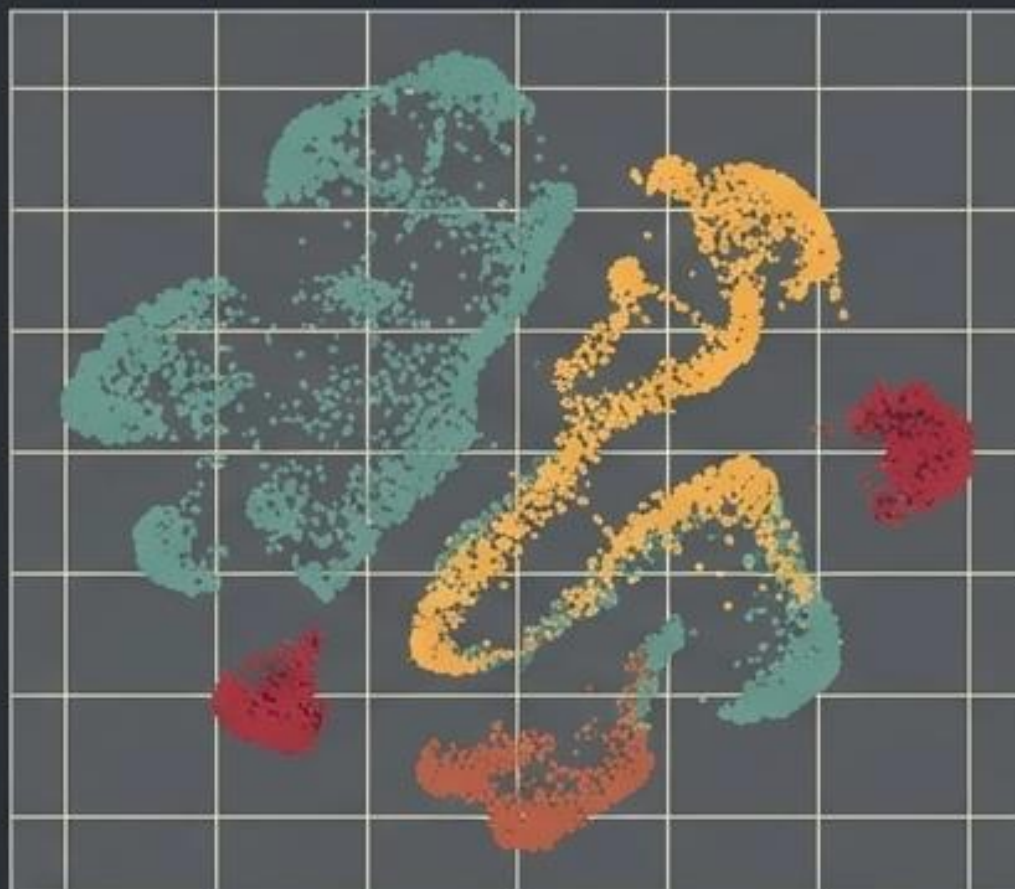
Оценка профиля: Определение доли нормальных, переходных и аномальных объектов в кластере, куда попало новое наблюдение.

Топология поведения: Визуализация мета-кластеров

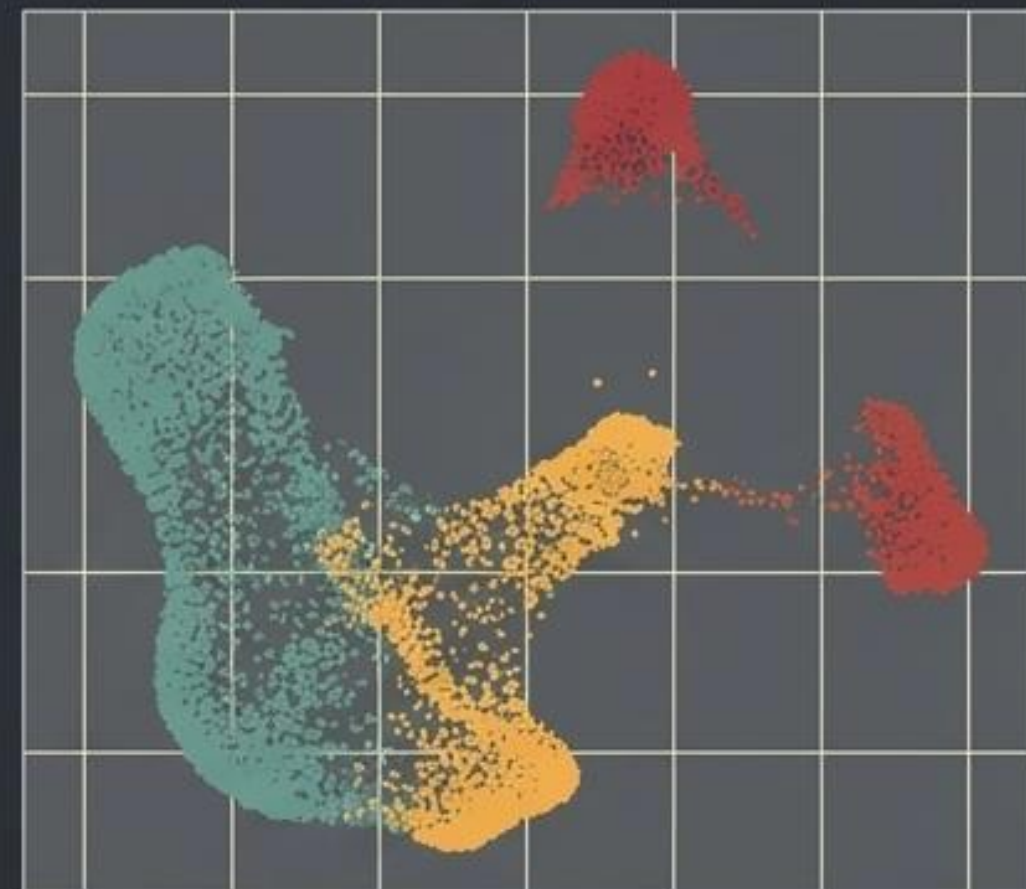
PCA



t-SNE



UMAP



Проекция пространств: Трансформация подпространств $D_c \in R^{T \times r}$ на 2D-плоскость для анализа плотности: $z_i = Proj_{2D}(x_i^C)$.

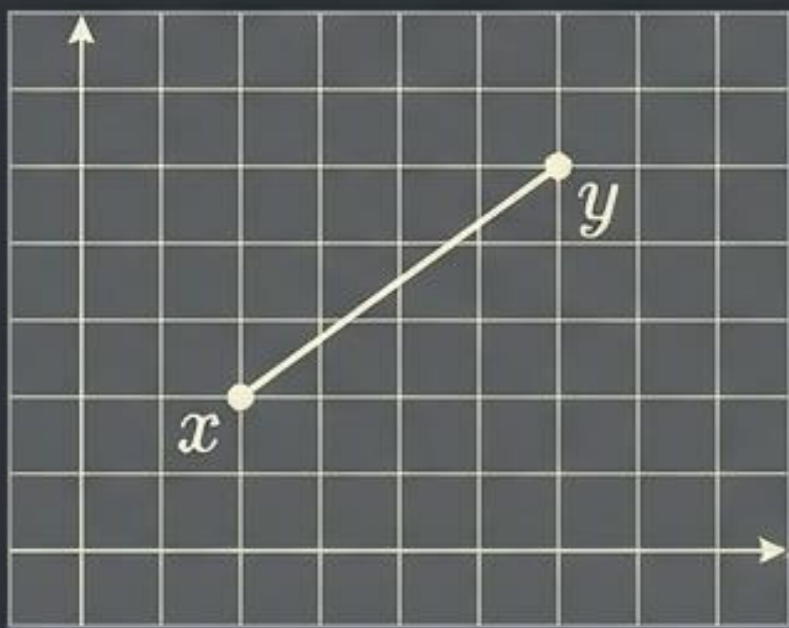
Алгоритмы снижения размерности (PCA, t-SNE, UMAP) позволяют оценить визуальную геометрию мета-кластеров.

Интерпретация профиля: Поведенческий профиль инсайдера представляет собой процент соседства его текущих действий с ранее размеченными аномалиями по десяткам тысяч подпространств.

Идентификация угрозы: Метрики сходства

Принятие решения: Сравнение полученного "отпечатка" (signature) объекта $S(x_{new})$ с эталонной матрицей Θ .

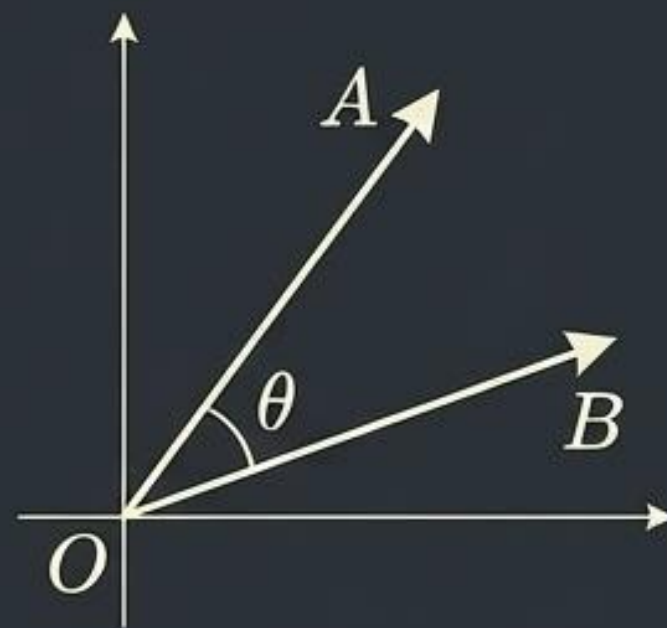
Евклидово расстояние



(Оценка абсолютного масштаба отклонений)

$$d_{euclidean}(x, y) = \sqrt{\sum_i (x_i - y_i)^2}$$

Косинусное сходство



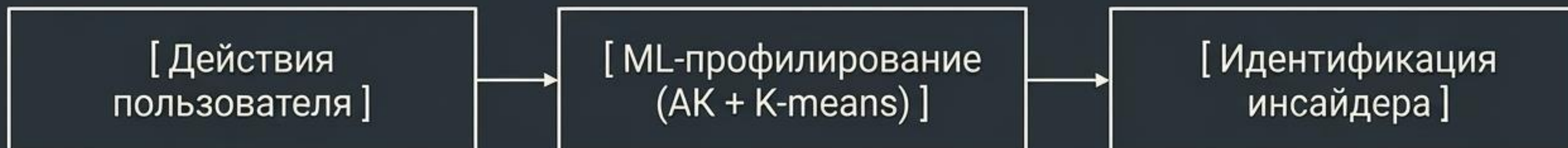
(Оценка структурного паттерна поведения)

$$similarity = \cos(\theta) = \frac{A \cdot B}{\|A\| \|B\|}$$

Вывод: Высокое структурное сходство с аномальными кластерами маркирует инсайдера даже при отсутствии явных нарушений правил доступа.

Заключение и выводы

- **Строгая методология:** Разработан математически обоснованный 5-этапный алгоритм выявления инсайдеров в CRM-системах на основе анализа микроповедения (клавиатура, мышь, тайминги).
- **Вероятностное профилирование:** Успешный переход от жестких статических правил безопасности к адаптивной топологии (Автокодировщик + Бутстрэп + K-means).
- **Раннее выявление:** Предложенный метод метрик сходства позволяет фиксировать формально легитимные, но структурно аномальные действия, предотвращая утечки данных на ранних стадиях.



Контакты

Московский технический университет связи и информатики,
Кафедра "Информационная безопасность"

Адрес: Москва, ул. Авиамоторная, д.8А, 111024
Телефон: +7(495)957-77-99 (доб.136)
e-mail: ib@mtuci.ru

д.т.н., проф. Шелухин Олег Иванович
o.i.shelukhin@mtuci.ru

к.т.ц, доц. Осин Андрей Владимирович
a.v.osin@mtuci.ru

