**Описание проблемы**

На маркетплейсе, где взаимодействуют фрилансеры-исполнители и заказчики цифровых услуг, выявлена проблема мошенничества. Мошенники обманывают неопытных заказчиков, предлагая перенести общение за пределы платформы, получать оплату вне сайта и не выполнять работу. Проверка сообщений модераторами вручную слишком трудозатратна, а автоматические алгоритмы требуют постоянных доработок.

Необходимо автоматизировать процесс обнаружения мошеннических сообщений, включая скрытые способы передачи контактов, чтобы защитить пользователей платформы.

**Варианты решений**

1. **Модерация вручную.**
   * **Плюсы:** Высокая точность за счет человеческого анализа.
   * **Минусы:** Низкая масштабируемость, высокая стоимость ресурсов, медленная обработка.
2. **Классические правила и регулярные выражения.**
   * **Плюсы:** Простота реализации, высокая скорость обработки.
   * **Минусы:** Сложно покрыть все скрытые варианты передачи контактов, высокая вероятность ложных срабатываний.
3. **Использование машинного обучения.**
   * **Плюсы:** Автоматическое обучение на реальных данных, способность выявлять скрытые паттерны.
   * **Минусы:** Требуется сбор и разметка данных, сложность настройки, необходимость вычислительных ресурсов.
4. **Гибридный подход: ML + регулярные выражения.**
   * **Плюсы:** Совмещение автоматизации и специфического анализа, высокая гибкость.
   * **Минусы:** Требуется разработка и поддержка обеих частей.

**Выбранное решение**

Для решения задачи выбран гибридный подход, который сочетает машинное обучение и регулярные выражения.

1. **Машинное обучение:**
   * Используется для обучения модели на разметке сообщений (мошеннические и обычные фразы).
   * Модель способна выявлять скрытые паттерны, которые невозможно описать с помощью правил.
2. **Регулярные выражения:**
   * Применяются для обнаружения специфичных форматов передачи контактов (номера телефонов, электронные адреса, кодированные контакты).
3. **Этапы реализации:**
   * Сбор и разметка данных (мошеннические и обычные фразы, триггерные слова).
   * Обучение модели на разметке данных.
   * Интеграция регулярных выражений для предобработки сообщений.
   * Тестирование и настройка системы на данных из маркетплейса.
4. **Ожидаемые результаты:**
   * Повышение точности обнаружения мошеннических сообщений.
   * Минимизация ложных срабатываний.
   * Защита интересов пользователей через автоматизированный контроль.

**Реализация на python**

Файл learn\_model2.py (python3.8) обучает модель на наборах мошеннических и обычных фраз плюс мошеннических триггерах.

spam\_trigger.txt - мошеннические триггеры

spam\_words.txt – мошеннические фразы

good\_words.txt – обычные фразы

custom\_regex\_model.pkl – обученная модель с метриками.

*Качество модели:*

precision recall f1-score support

0 0.60 1.00 0.75 3

1 1.00 0.60 0.75 5

accuracy 0.75 8

macro avg 0.80 0.80 0.75 8

weighted avg 0.85 0.75 0.75 8

Файл use\_model2.py принимает обученную модель и мошеннические триггеры, запрашивает фразу для анализа и выдает результат классификации и метрику точности для этой классификации.

**Листинг работы скриптов**

python.exe e:/Code/!util/for\_job\_kwork/learn\_model2.py

Введите путь к CSV файлу с мошенническими фразами: E:\Code\!util\for\_job\_kwork\spam\_words.txt

Введите путь к CSV файлу с триггерными словами: E:\Code\!util\for\_job\_kwork\spam\_trigger.txt

Качество модели:

precision recall f1-score support

0 0.60 1.00 0.75 3

1 1.00 0.60 0.75 5

accuracy 0.75 8

macro avg 0.80 0.80 0.75 8

weighted avg 0.85 0.75 0.75 8

Модель сохранена по пути: E:\Code\!util\for\_job\_kwork\custom\_regex\_model.pkl

Путь к модели скопирован в буфер обмена.

python.exe e:/Code/!util/for\_job\_kwork/use\_model2.py

Введите путь к модели (custom\_regex\_model.pkl): E:\Code\!util\for\_job\_kwork\custom\_regex\_model.pkl

Введите путь к CSV файлу с триггерами (точка с запятой): E:\Code\!util\for\_job\_kwork\spam\_trigger.txt

Модель загружена из: E:\Code\!util\for\_job\_kwork\custom\_regex\_model.pkl

Введите фразу для анализа (или 'exit' для выхода): Мне нужно больше информации о вашем проекте

Обычная фраза с уверенностью 0.61

Введите фразу для анализа (или 'exit' для выхода): Сколько экранов будет в приложении

Обычная фраза с уверенностью 0.64

Введите фразу для анализа (или 'exit' для выхода): Свяжемся в скайпе: ivan\_petrov\_123

Мошенническая фраза с уверенностью 0.57