В компании реализуется множество проектов в области обработки естественного языка (Natural Language Processing) - от классификации обращений клиентов в контакт-центр и разработки диалоговой системы в чатах до распознавания речи в телефонных разговорах.

В качестве тестового задания предлагается решить классическую задачу NLP - классификацию обращений. За основу взять открытый корпус обращений в систему народного контроля республики Татарстан.

**Задание.** Реализовать модель для определения класса сообщения. Примеры классов: благоустройство територии, свалки, сельское хозяйство и т.д.

Для обучения и валидации можно использовать следующий открытый датасет: [ссылка](https://git.asi.ru/tasks/world-ai-and-data-challenge/intelligent-algorithm-for-automated-processing-of-citizenss-requests). Датасет региона Татарстан.

При выполнении задания обратить внимание на:

1. Предобработку данных
2. Исследование различных векторных моделей для работы с текстом: простых статистических и нейросетевых
3. Выбор метрики качества
4. Выбор схемы валидации, оценка устойчивости модели
5. Сравнение результатов моделей, анализ ошибок моделей
6. Использование дополнительной информации для улучшения качества классификации. (внешние словари, гео данные, предобученные модели)

По итогу задания: оформить наглядный отчет о результатах.

Результаты реализации (код) и отчет загрузить в свой репозиторий на github.

Прислать ссылку на github на e-mail: dmitrinyuv1@intersvyaz.net