

Источники идей проектов:

1. [сайт](#) с темами проектов
2. [SemEval2021](#)
3. [соревнование](#) с конференции Диалог
4. [всякие соревнования](#)
5. ещё соревнования - <https://www.drivendata.org/competitions/>

?Enron Email Dataset - подумать над проектом на его основе

Обработка естественного языка

1. чат-боты;
2. генерация текстов в стиле определённого автора;
3. квест-игры;
4. вопросно-ответные системы;
5. спам-детекторы, распознавание фейковых новостей;
6. sentiment-анализ, распознавание эмоций;

Компьютерное зрение

1. определение гриба/животного/пола человека по фото;
2. актуальное: mask detector;
3. разметка лиц;

Медицинская диагностика

Рекомендательные системы

1. Рекомендательная система для статей - [соревнование](#) на Kaggle

2. предсказание лучшего маршрута на основе датасета Uber
3. сегментация покупателей и подбор оптимального предложения/товара
4. Рекомендательная система для фильмов
5. <https://zindi.africa/competitions/akeed-restaurant-recommendation-challenge/leaderboard> - рекомендация ресторанов

Разработка датасетов

1. Расширение датасета NER для новых сущностей, пока есть Pers, Org, Media, Loc, Geopolit (https://www.labinform.ru/pub/named_entities/descr_ne.htm)
2. Разработка датасетов под конкретную задачу CV

Предсказания, основанные на статистических данных

1. предсказание погоды
2. предсказание курса доллара
3. предсказание цен на товары/недвижимость
4. covid :)