



NLP НЕЙРОСЕТИ
В ЗАЩИТЕ ДАННЫХ:

ОПЫТ MAKVES DCAP

ОБО МНЕ



Tg: @LaHundo

Старший специалист по
машинному обучению

Makves (входит в группу
компаний «Гарда»)

CV, NLP, ASR etc

Open source contributor

Амбассадор Яндекс
Практикума

Выпускник DLS МФТИ

СОКРАЩЕНИЯ

DCAP



Data-Centric Audit and Protection

NLP



Natural Language Processing

NER



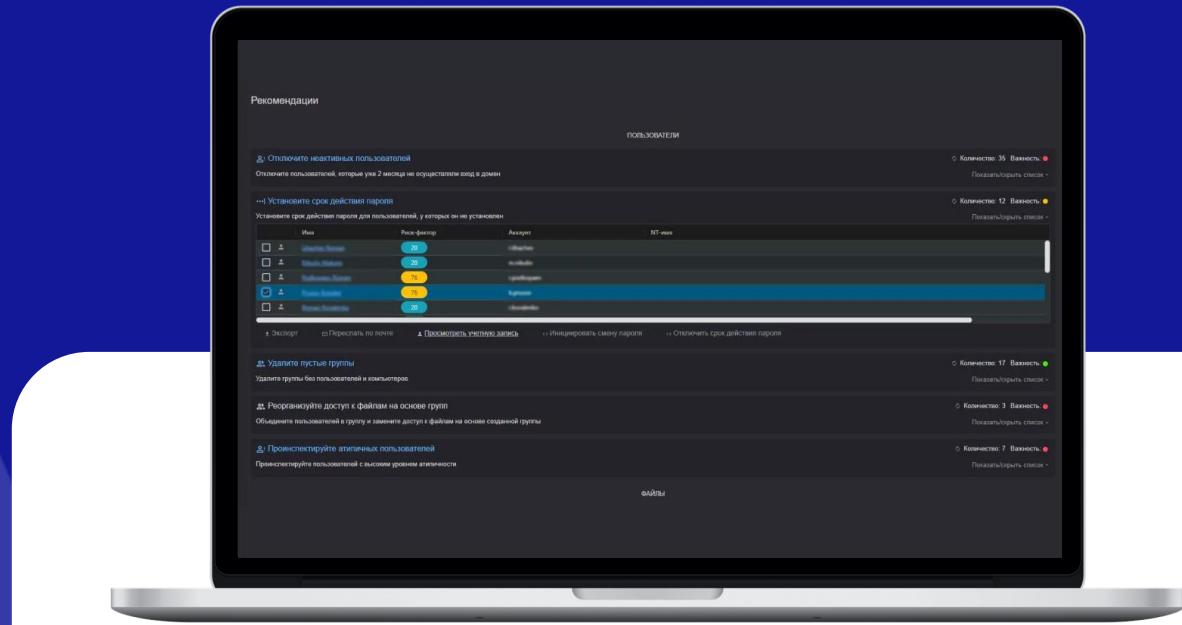
Named Entity Recognition

SOTA

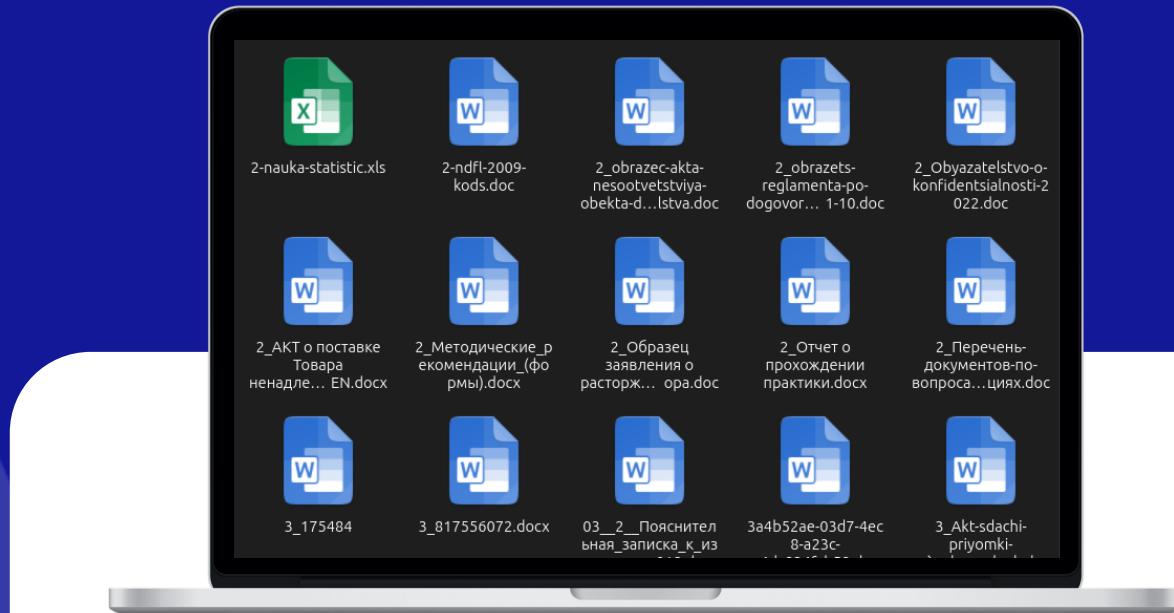


State of the Art

ЗАДАЧА



НЕСТРУКТУРИРОВАННЫЕ ДАННЫЕ



| Название файла | Категория | Информационные объекты |
|-------------------------|---------------------|-------------------------------|
| ДОГОВОР_ФИНАЛ.DOC | Договор | Иванов И. И. |
| НОВЫЙ ФАЙЛ (3).RTF | Техническое задание | ООО Ромашка |
| РОМАШКА_ПЛАТЁЖКА.XLSX | Квитанция | ООО Ромашка, Иванов И. И. |
| ПЕРЕДЕЛКА_НОВАЯ.DOCX | Другое | Петров П. П. |
| ПРАВКИ ЗАКАЗЧИК.DOCX | Другое | Иванов И. И. |
| ДОГОВОР_ФИНАЛ_ФИНАЛ.DOC | Договор | ООО Иванов и партнёры |

КЛАССИФИКАЦИЯ

КЛАССИФИКАЦИЯ

Количество файлов по выборкам



КЛАССИФИКАЦИЯ

Количество файлов по выборкам



| | rubert-tiny2 | CatBoost |
|--------------------------|---------------------|-----------------|
| Accuracy val | 98,3 % | 97,9 % |
| Accuracy test | 92,7 % | 91,8 % |

КЛАССИФИКАЦИЯ

Количество файлов по выборкам



| | rubert-tiny2 | CatBoost |
|---------------|--------------|----------|
| Accuracy val | 98,3 % | 97,9 % |
| Accuracy test | 92,7 % | 91,8 % |
| Время | 347 сек | 2 сек |

NAMED ENTITY RECOGNITION

NER

| | Время поиска |
|------------|---------------------|
| Spacy | 1319 сек |
| WiKiNEural | 1090 сек |
| SlovNet | 344 сек |

NER

| | Время поиска |
|------------|----------------|
| Spacy | 1319 сек |
| WikiNEural | 1090 сек |
| SlovNet | 344 сек |

| | factru | gareev | slip |
|-------------------|--------|--------|-------|
| SlovNet F1 ORG | 0,825 | 0,899 | 0,723 |

NER

| | Время поиска |
|------------------------------|---------------|
| Spacy | 1319 сек |
| WikiNEural | 1090 сек |
| SlovNet | 344 сек |
| SlovNet (быстрее) | 46 сек |

| | factru | gareev | slip |
|---------------------------|--------|--------|-------|
| SlovNet F1 ORG | 0,825 | 0,899 | 0,723 |

НОРМАЛИЗАЦИЯ

Иванов Иван Иванович является одним из ведущих специалистов ООО Ромашка в области информационных технологий. Он обладает обширным опытом работы и глубокими знаниями в своей сфере деятельности. Иванов И.И. успешно руководит проектами, связанными с разработкой программного обеспечения и управлением ИТ-инфраструктурой. Благодаря своему профессионализму и умению находить инновационные решения, он внес значительный вклад в развитие ООО Ромашка.

ВЫВОДЫ



SOTA не всегда
подходит



Сбор данных =
инсайты



Метрики могут
просесть

MAKVES

Российский разработчик
программного обеспечения

