



Sirius Digital Hack

Подготовили: ПРОНайк



план

01

Актуальность

02

Цель

03

Идея

Модель для

04

кластеризации

05

Модель для определения

тональности
Пользовательский

06

интерфейс



01

Актуальность





Анализ анкет увольняющихся сотрудников является очень важным для любой компании, особенно в условиях высокой текучести кадров. Компании сталкиваются с растущими затратами на найм и обучение новых сотрудников, что делает понимание причин увольнений критически важным.

02

Цель



Цель заключается в создании сервиса для обработки анкет увольняющихся сотрудников, который позволит эффективно оценивать и анализировать данные о причинах увольнений с использованием методов машинного обучения и обработки естественного языка.

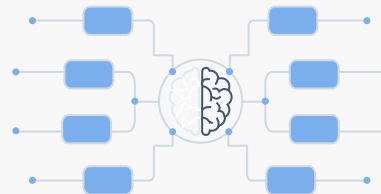


03

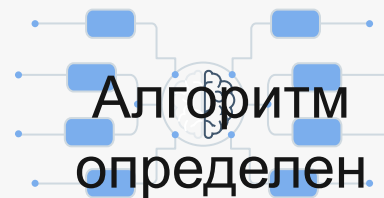
Идея



Алгоритм
кластеризации



диаграммы



Алгоритм
определен.



ия
тональность
и



текст анкет



сервис



04

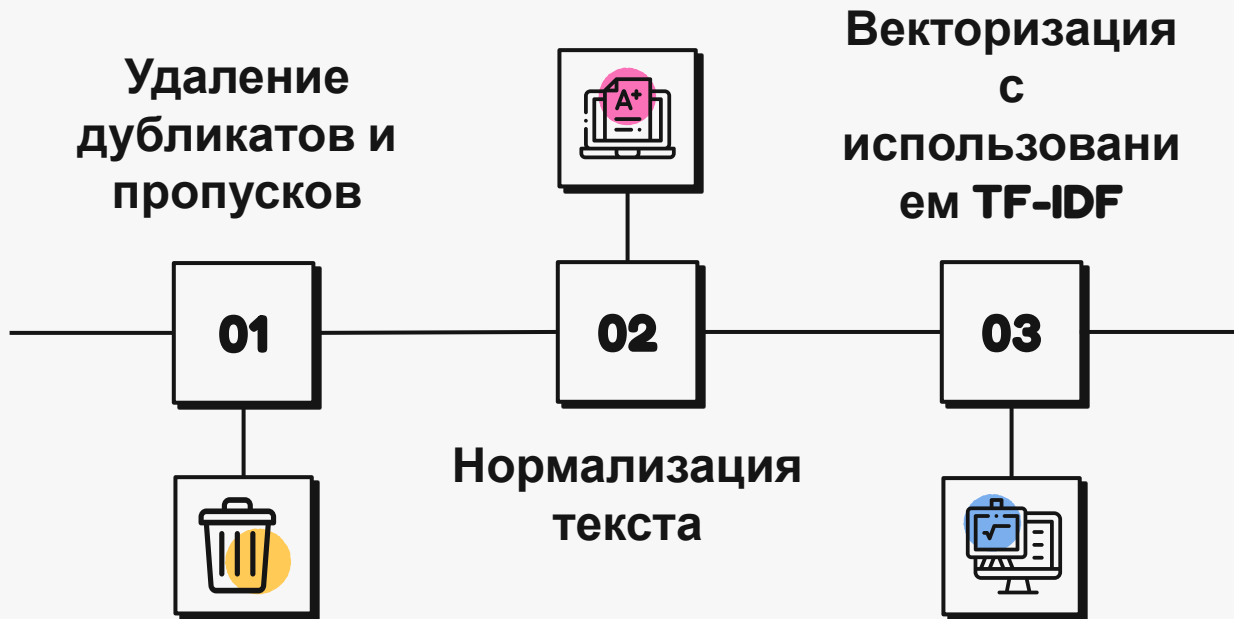
**Модель для
кластеризации**

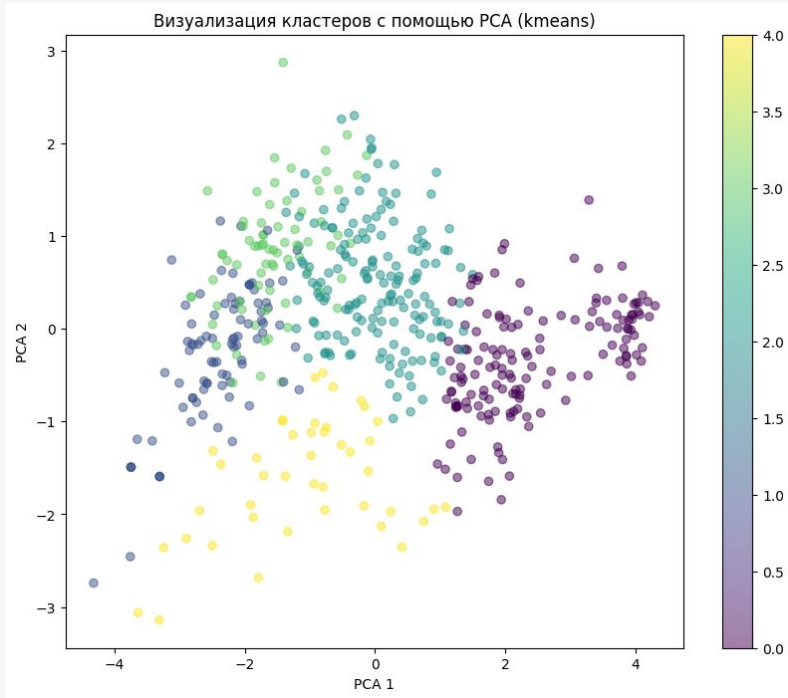


Шаги кластеризации



Предобработка





Кластеризация на 8 классов

Cluster silhouette=0.14

Кластеризация вопросов про увольнение

Кластер 0: целеполагание

Кластер 1: обстоятельство

Кластер 2: месторасположение

Кластер 3: предложение

Кластер 4 : переходит



Определение тональности



Для анализа тональности
использовалась модель
`rubert-base-cased-sentiment` —
предобученная языковая модель на
основе BERT.

Обычно для моделей, основанных на
BERT, точность может достигать 85-90%
на хорошо подготовленных тестовых
наборах



Пользовательский интерфейс

Технологии

- Python
- Flask

Анализ анкет при увольнении

Разработано командой ПРОНайк

Загрузите файл для анализа

Обзор... Файл не выбран.

Анализировать

Анализ тональности

neural -> 322

positive -> 126

negative -> 36



Основные причины ухода

Cluster 0 (обеспечения) -> 23

Cluster 1 (график) -> 755

Cluster 2 (петрозаводск) -> 48

Cluster 3 (профессионального) -> 97

Cluster 4 (процесс) -> 59



Результаты

Был создан сервис для анализа анкетирования увольняющихся работников. Удалось достичь достаточно оптимального решения, которое содержит в себе баланс между скоростью и качеством



Show-case

