



# Ранжирование вакансий

---

Решение задачи от команды Punk Butterfly



01

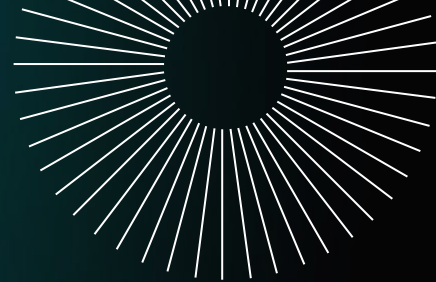
▶▶▶▶

# Анализ данных

---

Как работали с таблицами

# На что обращали внимание



## Релевантность

Некоторые фичи по нашему мнению оказались недостаточно значимыми



## Разметка данных

Большое внимание было уделено разметке обучающих данных

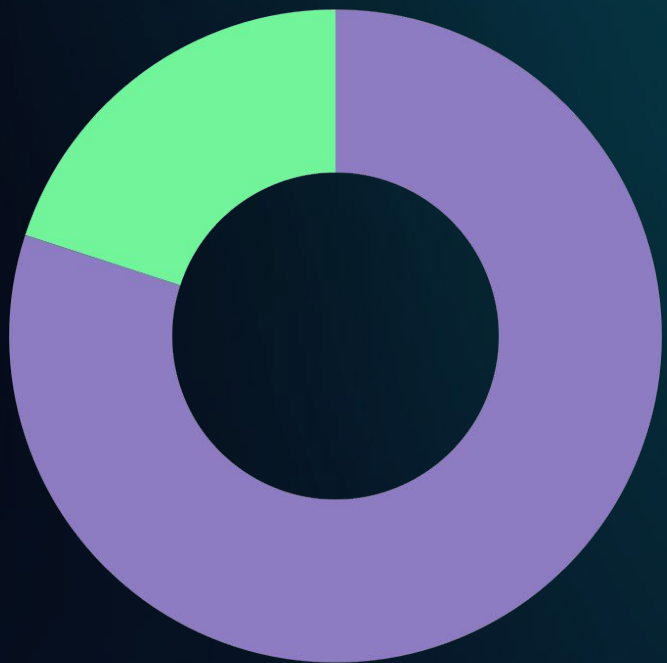


## Однотипность анкет

Удалили одинаково пустые анкеты



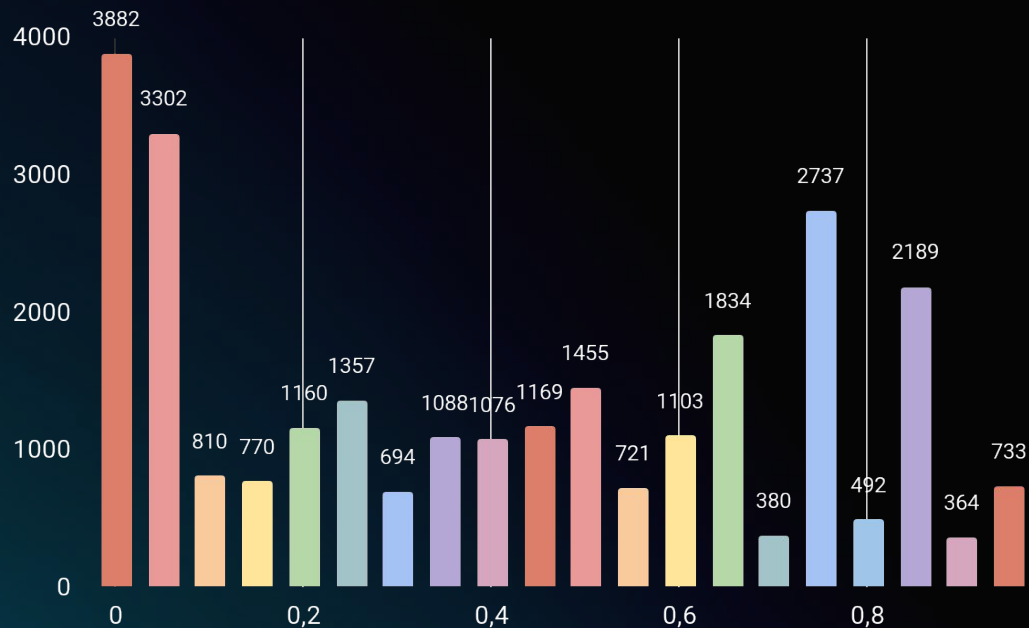
# Соотношение дубликатов



Уникальные значения 20%

Дубликаты 80%

# Распределение размеченных таргетов



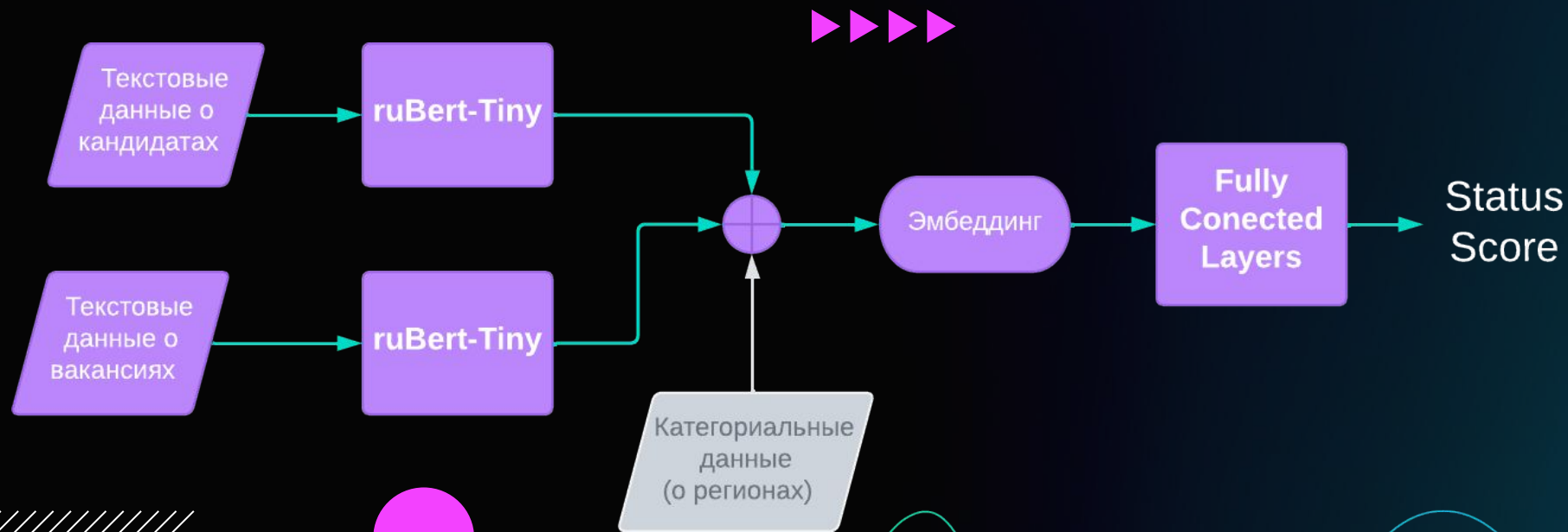


02

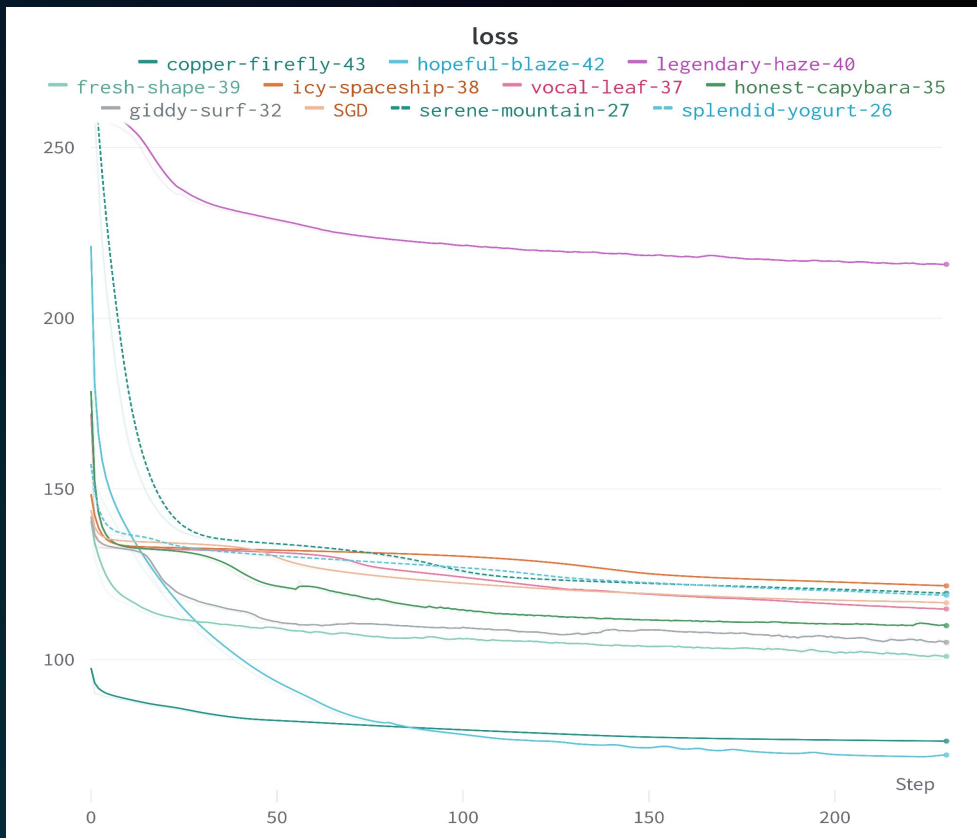
# Построение модели

Какой подход использовали

# Pipeline модели



# Обучение



## Что выбрать...

Мы сделали очень много запусков обучения модели, чтобы подобрать лучшие параметры

## Результаты обучения

Итоговые метрики:  
**MAE ~ 0.22**  
**MSE ~ 0.09**






# 03



## Запуск веб-сервиса

---

Как реализовали



# Streamlit



## Input parameters

### Vacancy

Job ID

 - +

Job status

Title

Region

### Candidate

Position

Sex



## You can upload a csv files

Jobs.csv



Drag and drop file here

Limit 200MB per file

Browse files

Candidates.csv



Drag and drop file here

Limit 200MB per file

Browse files

Candidates\_education.csv

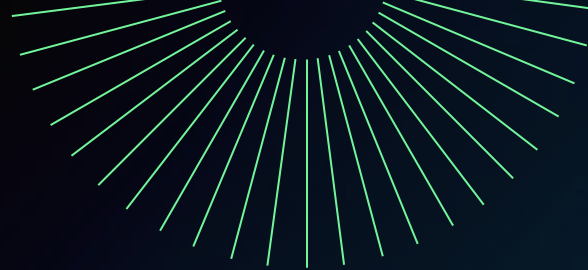
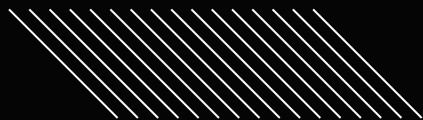


Drag and drop file here

Limit 200MB per file

Browse files





# ИТОГИ

---



# Github команды

