

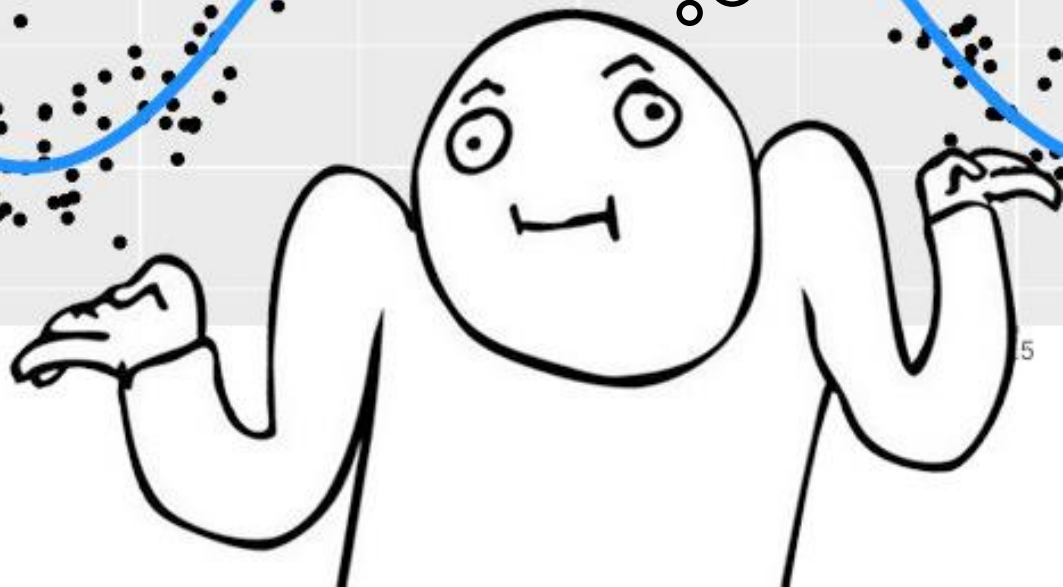
Предсказание компенсации
навыков на основе резюме

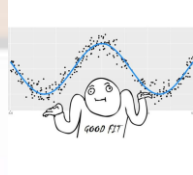
GoodFit

-5.0

5

5.0





Задача:

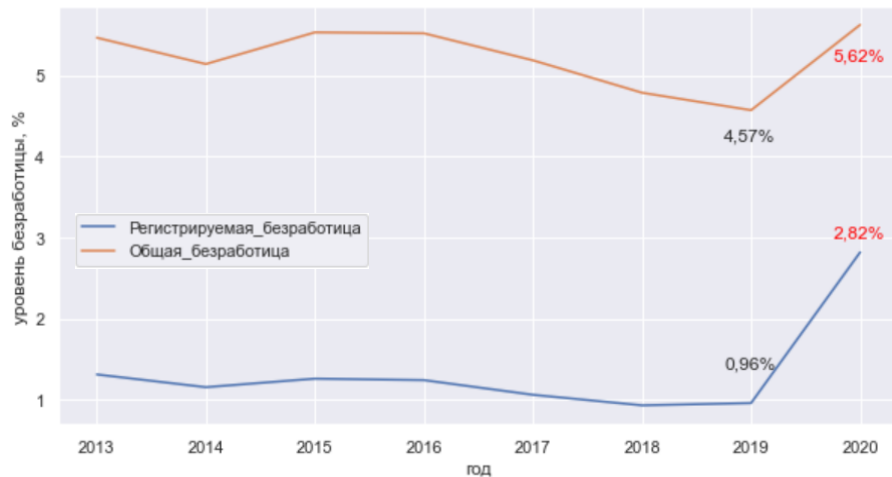
Разработать модель, позволяющую давать соискателям адекватную их резюме оценку предполагаемой заработной платы, чтобы ускорить для них процесс поиска работы

Медианное время поиска работы = **8,6** мес*

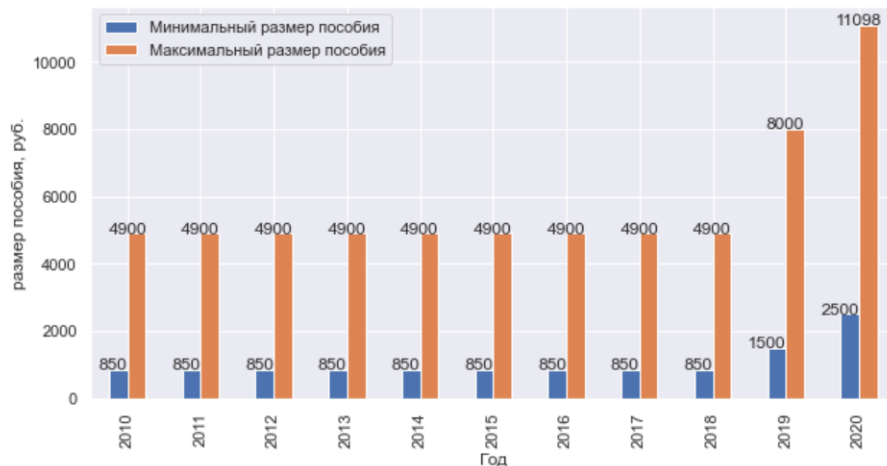
*на основе анализа данных портала trudvsem

GET A JOB

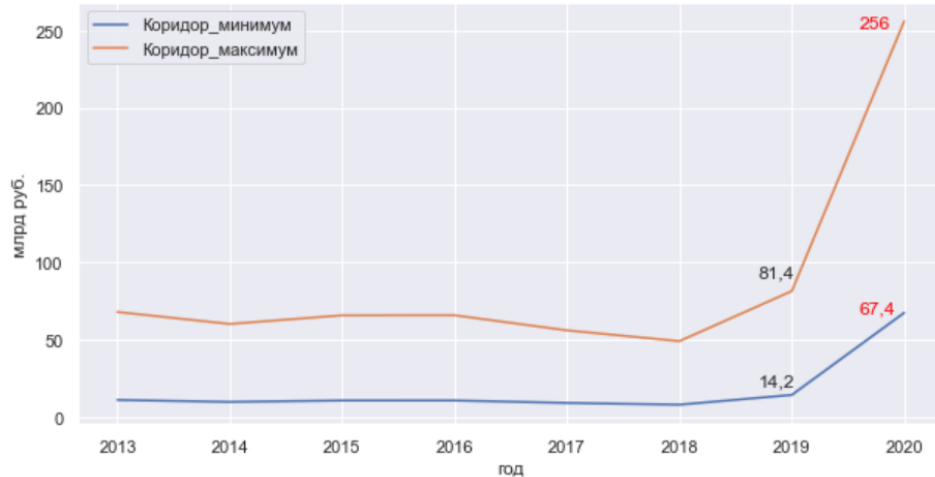
Динамика безработицы в РФ за 2013-2020 гг.



Статистика изменений размера пособия по безработице в РФ за 2010-2020 гг.



Предполагаемый коридор выплат пособий по безработице в РФ за 2013-2020 гг.



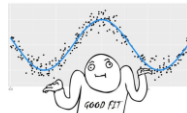
Рост безработицы + увеличение выплат =>
взрывной рост затрат на выплату пособий

Сокращение времени поиска работы на 1 месяц
позволит снизить затраты на :

32,236 – 284,626 млн руб. в год

и получить дополнительно:

~96,89 млн руб. ндфл



Коэффициенты различия в группах по категориям в регионах

Определяют минимальный
уровень оплаты

МРОТ

уровень цен

районные коэффициенты

готовность к релокации

готовность к командировкам

готовность к переобучению

пол

ИЗВЛЕЧЕННЫЕ ДАННЫЕ

Образование

число учебных заведений

число уникальных учебных
заведений

мин/макс дата конца обучения

Социально-экономический портрет региона

рабочая сила

безработица

доля городского населения

численность населения по городам

количество юр. лиц

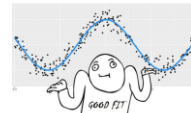
Трудовой опыт

число записей о работе

число уникальных должностей

время простоя без работы

число уникальных мест работы



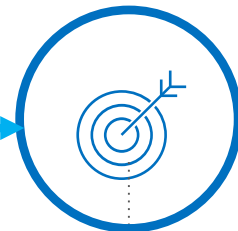
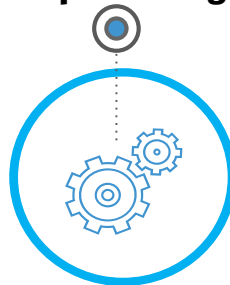
Обработка
текстовых данных

Text Data



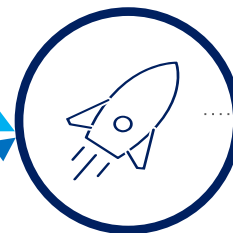
Извлечение
признаков, работа
с аномальными
значениями

Preprocessing



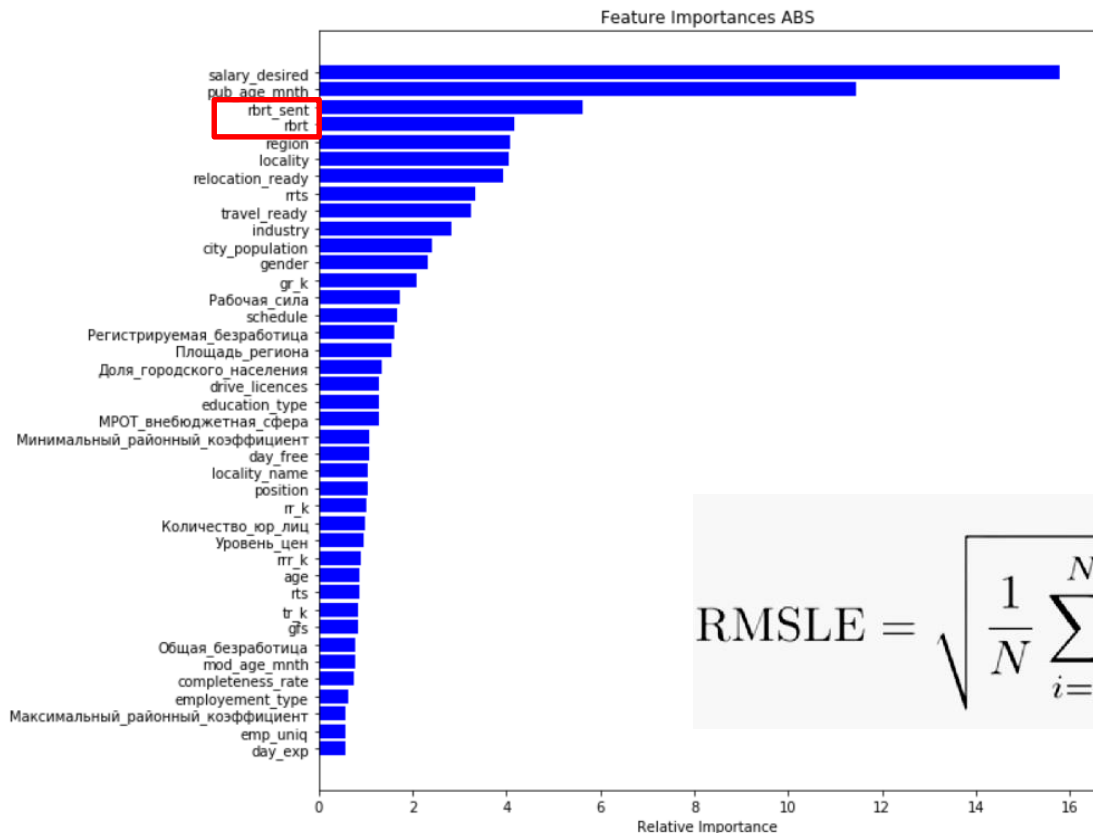
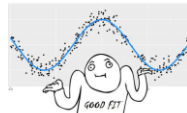
NLP

Обучение
архитектуры Bert
от DeepPavlov

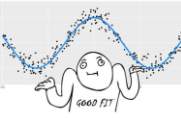


Catboost

Кодирование
категориальных
признаков,
построение
прогноза



$$\text{RMSLE} = \sqrt{\frac{1}{N} \sum_{i=1}^N (\log(y_i + 1) - \log(\hat{y}_i + 1))^2}$$



CatBoost



NumPy



SciPy

PYCARRET

IP[y]: IPython
Interactive Computing

pandas

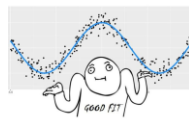
$$y_{it} = \beta x_{it} + \mu_i + \epsilon_{it}$$



HYPEROPT



GoodFit team



Глеб
Ерофеев



Иван
Стразов



Артемий
Саранцев



Екатерина
Гладкова