

Одно из решений конкурса Salary prediction

Антон Панкратов

Избавляемся от лишних символов и регистра

```
#!/bin/bash  
sed '1!s/[_()\.\/\*- \&+;\/?]/ /g' $1 | \  
sed "1!s/'/□/g" | \  
sed -e '1!s/\(.*\)/\L\1/' > $2
```

Генерируем признаки для VW

- ▶ Стеммим слова
- ▶ Записываем их в три группы:
 1. текст объявления
 2. бинарные признаки - contract type, contract time
 3. все остальное
- ▶ Будем искать \sqrt{salary}

Итоговый алгоритм

- ▶ `loss_function quantile` - оптимизируем непонятно что
- ▶ `b 24` - битовая точность
- ▶ `hash strings`
- ▶ `passes 20`
- ▶ `l 300`
- ▶ `nn 10` - автокодировщик
- ▶ `l1 1e - 7`
- ▶ `l2 1e - 7`
- ▶ `ngram 2`
- ▶ `skips 1`

The End