

## Gendered Pronoun Resolution

К какому существительному относится  
местоимение в тексте?

# Задача

Len Cundell moved from the Bungalow stables at Chilton to Blewburton Hall stables at Aston Tirrold. His best horse was Noble Star who, as a four year old, won the Cesarewitch in 1931. His cousins Ken Cundell and Frank Cundell, and his nephew Jack Waugh, the son of Len's sister Letitia, also became successful racehorse trainers and gained experience in his yard.

$$\text{logloss} = -\frac{1}{N} \sum_{i=1}^N \sum_{j=1}^M y_{ij} \log(p_{ij}),$$

# Примеры сложных случаев

Analyze Predictions dataset tv\_lastName6

validation(tv\_lastName6)

Stage: t-['get\_val': None], ['get\_test'] Visualizer: html\_highlighter Analyzers: certainty\_analyzer

Trusted correct(190)

Untrusted incorrect(24)

76: to ratify the repeal of the eighteenth amendment in 1933. **ritchie** tried, but failed to achieve nomination as president during the election of 1932, and declined an offer by **franklin chris roosevelt** to be **his** running mate.

96: gradually deteriorates and she is hospitalized. **mr. verma** tells **arun** that **sofia** has cancer and is in the hospital. he goes to the hospital, only to find out that **sofia** is none other than **lucy**. he tells **her** wife about **lucy** and

237: lives in corvallis, oregon, with her veterinarian husband, **tim barraud**. she is the mother of singer and musician **courtney love**, and the daughter of novelist **paula fox**. **linda** was born to **paula fox** when **she** was 20, the result

317: position after **larry campbell** left the position. **lopez** then ran in the 1992 election to replace **john turner** in oregon senate district 2, where **lopez** then lived after redistricting. on july 2, 1992, **he** was appointed to the

483: from scotland touching delinquents and sequestrations." a petition having been presented to the council of state by **john, earl of crawford**, order was made september 1, 1653, that **mr. phelps** examine **his** books as to what was

511: that was to transfer." after **cruthers'** decision to leave **usma** he called **derek andy**, the new head coach of **rmu mens hockey team**. **andy**, formerly assistant coach for air force, had recruited **ryan** before **his** decision to attend

512: were parents of the actors **bab christensen** and **pelle christensen**. she had her stage d'but at det nye teater in copenhagen in 1911, in the play kongens hjerte, written by **nancy ring** (**mary's** mother). **she** played at the national

550: **francesco carlo johnson**, an italian nobleman of mixed english and italian parentage, who had been living in the london area at the time of the murder. suggestions about the count's possible involvement seem to have originated in a

1065: **jan** the "pearl in haarlem's crown". **houbraken** saw a painting by **de bray** of **david** and the return of the ark of the covenant in the collection of **arnold van wood** in amsterdam, dated 1697, that **he** admired for its realistic

1243: the journey to the powder river country. he left the camp and returned with 80 lodges to the arkansas river to seek peace with the coloradoans. despite threats against **his** life, captain **silas brown** testified against **sammy**.

1247: the maylie family, as in the novel. it is **nancy's** informing rose, and a highly sceptical **mr john**, of **oliver's** whereabouts that results in her demise at the hands of **sykes**. her ghost continues to haunt **him** when he returns to

1262: **katrina** (**suzan farmer**) and their newborn child, but he is still subject to his mother's evil influence. when **she** is involved in two terrifying deaths, **david's** mind snaps; although he is already mentally twisted by **carrie's** treatment, **david**

1276: poorer background the two became close and were frequently seen collecting fossils together. **miller** encouraged the young **williams** to read about geology and understand the science behind the fossils **she** collected and sold.

1302: property was then sold to **john calcraft**, mp for **rochester**. on **john calcraft's** death in 1772, the estate passed to his son **john**, who sold it in 1788 to **johnjohn disney brown**, on whose death in 1796 it passed to **his** son **henry brown**.

1451: marry an actor. **barnaid kate** tells **charles** why **carole** feels the way she does about actors: **carole's** mother, **lillian grey**, was with a touring show in 1913 when the handsome star, **john beaumont** raised **her** from the chorus to

1497: would have occurred had the other twin survived in his place. during **his** run, there was also a stand-alone issue, the hellblazer annual #1, exploring **constantine's** ancestry, and featuring the video to **john's** punk band, **mucous**

1510: **dave turner** field goal, a safety touch and a **tim johnson** touchdown, to move into a 34-30 lead. with just under two minutes left to go in the fourth quarter the riders went ahead 40-33 after **turner** kicked **his** third field goal of

1532: **her** book, reform, respite, ritual: an archaeology of institutions; the **magdalene society** philadelphia, 1800-1850, that the women in philadelphia's asylum "sought a refuge and a respite from disease, the prison or almshouse, unhelp

1704: man, **raj**, and they fall in love with each other. **raj** then breaks her heart when he tells her that his mother, **malti**, who works as a maidservant in **harry's** household, has always favored **jacob** over **him** and has made him promise

1896: in 1954, her family moved to s"o paulo, brazil. at the age of 17 she married **jos' walker**; she has three children from this marriage: **victoria mickey walker**, **pedro walker** and **nancy walker goodman**. **she** started creating her

2048: his female half around the house, he calls himself **mrs. hyde**, saying she is **jekyll's** widowed sister. **dr. jekyll** soon falls in love with **susan spencer**, the upstairs tenant, but sister **hyde** develops a lust for **her** brother.

2069: arranges to use the friend's place in mexico to hide out. unknown to her, the syndicate is monitoring the hotel room. the next morning, **donna** drives off toward the us-mexico border, unaware **sandra** and **erica** know **her** destination.

2214: was a new zealand singer of maori descent. at one stage she was regarded as new zealand's top entertainer. **clark** was born in taumarunui to famous maori rugby player **ron clark** and **betty**. **her** first husband was professional

2373: (born **ashley ann erickson** on july 24, 1985) is a former american pornographic actress. her stage name in part comes from her parents initially wanting to call **her** "leslie" and in part as an homage to **lisa marie presley**.

Parameters

Level: 0,7

Filters

Trusted incorrect(19)

Untrusted correct(13)

Statistics Data

Apply

# Оригинальная статья

BERT->attention heads

<https://arxiv.org/abs/1810.05201>

# Наше решение: общая архитектура

- Воспроизведение лучшего решения из оригинальной статьи
- Поиск слоев с лучшими результатами
- Сеть, комбинирующая значения снятые с attention heads
- Проблема: bert плохо работает когда имена не входят в словарь

# Наше решение: тюнинг имен

- Замена неизвестных имен на известные
- Максимальные attention score из токенов имен
- Поиск вхождения токенов более близких к местоимению
- Замена имен, зависящая от пола
- Замена фамилий

# Наше решение: улучшение результатов

- Gradient boosting поверх attention heads как альтернатива -> Cudo's to P. Pleskov
- Blending NN+Gradient Boosting
- Test time augmentation (не включено в финальное решение)

# Инструменты и командная работа

Большое количество воспроизводимых экспериментов

Сохранение данных по параметрам и метрикам

Сравнение экспериментов друг с другом

Передачи экспериментов другим участникам команды

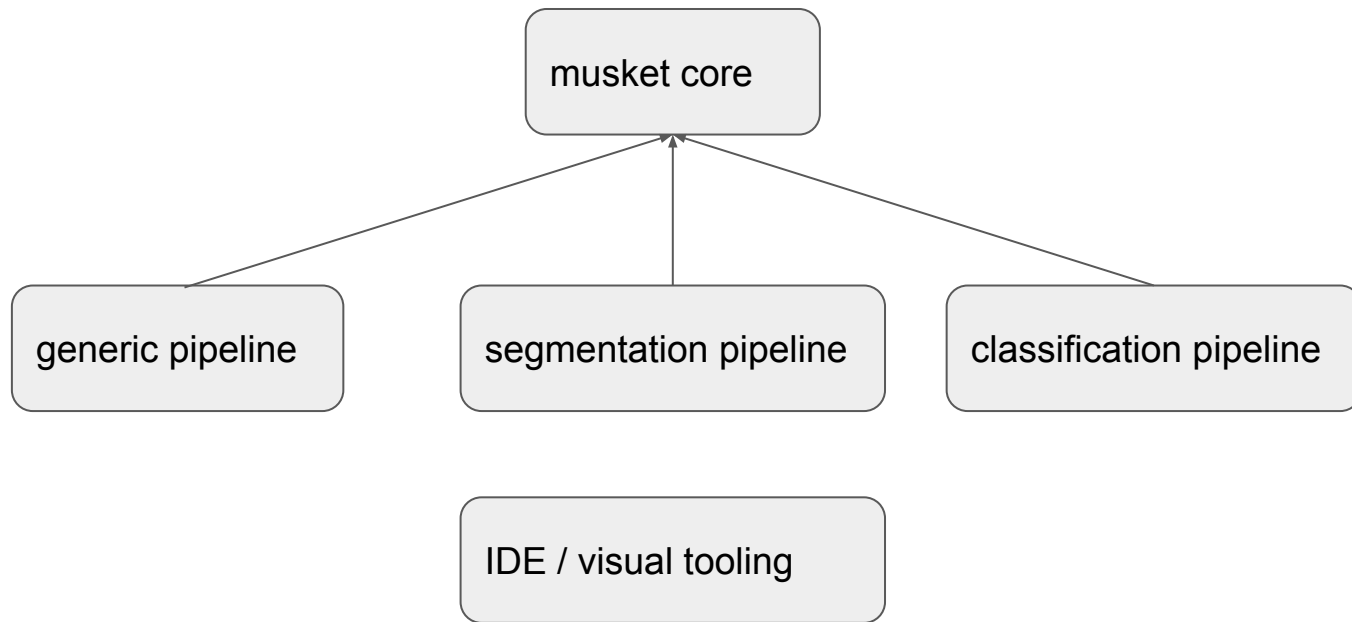


## Декларативное задание

### YAML

```
architecture: Xception
augmentation:
  Fliplr: 0.5
  Flipud: 0.5
classes: 28
activation: sigmoid
weights: imagenet
shape: [256, 256, 4]
optimizer: Adam
batch: 40
lr: 0.001
copyWeights: true
metrics:
  - binary_accuracy
  - macro_f1
primary_metric: val_macro_f1
primary_metric_mode: max
callbacks:
  EarlyStopping:
    patience: 15
    monitor: val_macro_f1
    mode: max
    verbose: 1
  CyclicLR:
    base_lr: 0.00001
    max_lr: 0.0001
    mode: triangular2
    step_size: 10000
loss: binary_crossentropy
stages:
  - epochs: 100
```

# Тулинг



<https://github.com/musket-ml>

# Как выглядели эксперименты

Experiments			
Filter: <input type="text"/>			
Name	Completed	Score	Description
er holdout:0,4	<input type="checkbox"/>		None(primary metric: val_log_loss architecture: net dataset: train=(get_val=null, get_test=null), holdout=(get_dev=null))
test_near_replace_unknowns_1_5_folds_01 holdout:0,4	<input type="checkbox"/>		None(primary metric: val_log_loss architecture: net dataset: train=(get_val=null, get_test=null), holdout=(get_dev=null))
tv_lastName5 holdout:0,4	<input type="checkbox"/>		None(primary metric: val_log_loss architecture: net dataset: train=(get_val=null, get_test=null), holdout=(get_dev=null))
tv_lastName6 holdout:0,4	<input type="checkbox"/>	Error	None(primary metric: val_log_loss architecture: net dataset: train=(get_val=null, get_test=null), holdout=(get_part0_labeled=null))
tv_alot_of_seeds seeds:200	<input type="checkbox"/>	Error	None(primary metric: val_log_loss architecture: net dataset: get_ensemble=[])
tv_lastName4 holdout:0,4	<input type="checkbox"/>		None(primary metric: val_log_loss architecture: net dataset: train=(get_val=null, get_test=null), holdout=(get_dev=null))
tv2 holdout:0,4	<input checked="" type="checkbox"/>	0,2969	None(primary metric: val_log_loss architecture: net dataset: train=(get_val=null, get_test=null), holdout=(get_dev=null))
tv_lastName3 holdout:0,4	<input checked="" type="checkbox"/>	0,2989	None(primary metric: val_log_loss architecture: net dataset: train=(get_val=null, get_test=null), holdout=(get_dev=null))
tv2-11 holdout:0,4	<input checked="" type="checkbox"/>	0,299	None(primary metric: val_log_loss architecture: net dataset: train=(get_val=null, get_test=null), holdout=(get_dev=null))
tv_lastName holdout:0,4	<input checked="" type="checkbox"/>	0,2994	None(primary metric: val_log_loss architecture: net dataset: train=(get_val=null, get_test=null), holdout=(get_dev=null))
tv25 holdout:0,4	<input checked="" type="checkbox"/>	0,2996	None(primary metric: val_log_loss architecture: net dataset: train=(get_val=null, get_test=null), holdout=(get_dev=null))
tv2-14 holdout:0,4	<input checked="" type="checkbox"/>	0,3016	None(primary metric: val_log_loss architecture: net dataset: train=(get_val=null, get_test=null), holdout=(get_dev=null))
tv4 holdout:0,4	<input checked="" type="checkbox"/>	0,3019	None(primary metric: val_log_loss architecture: net dataset: train=(get_val=null, get_test=null), holdout=(get_dev=null))
tv_lastName2 holdout:0,4	<input checked="" type="checkbox"/>	0,3021	None(primary metric: val_log_loss architecture: net dataset: train=(get_val=null, get_test=null), holdout=(get_dev=null))
tv23 holdout:0,4	<input checked="" type="checkbox"/>	0,303	None(primary metric: val_log_loss architecture: net dataset: train=(get_val=null, get_test=null), holdout=(get_dev=null))
tv22 holdout:0,4	<input checked="" type="checkbox"/>	0,303	None(primary metric: val_log_loss architecture: net dataset: train=(get_val=null, get_test=null), holdout=(get_dev=null))
tv3 holdout:0,4	<input checked="" type="checkbox"/>	0,3032	None(primary metric: val_log_loss architecture: net dataset: train=(get_val=null, get_test=null), holdout=(get_dev=null))
test_near_replace_unknown_all_maxToken_01 holdout:0,4	<input checked="" type="checkbox"/>	0,3037	None(primary metric: val_log_loss architecture: net dataset: train=(get_val=null, get_test=null), holdout=(get_dev=null))
tv24 holdout:0,4	<input checked="" type="checkbox"/>	0,3047	None(primary metric: val_log_loss architecture: net dataset: train=(get_val=null, get_test=null), holdout=(get_dev=null))
223 holdout:0,4	<input checked="" type="checkbox"/>	0,3048	None(primary metric: val_log_loss architecture: net dataset: train=(get_val=null, get_test=null), holdout=(get_dev=null))
cased_test holdout:0,4	<input checked="" type="checkbox"/>	0,3054	None(primary metric: val_log_loss architecture: net dataset: train=(get_val=null, get_test=null), holdout=(get_dev=null))
dede33 holdout:0,4	<input checked="" type="checkbox"/>	0,3056	None(primary metric: val_log_loss architecture: net dataset: train=(get_val=null, get_test=null), holdout=(get_dev=null))
dede44 holdout:0,4	<input checked="" type="checkbox"/>	0,307	None(primary metric: val_log_loss architecture: net dataset: train=(get_val=null, get_test=null), holdout=(get_dev=null))
tvLastName444 holdout:0,4	<input checked="" type="checkbox"/>	0,3071	None(primary metric: val_log_loss architecture: net dataset: train=(get_val=null, get_test=null), holdout=(get_dev=null))
tv26 holdout:0,4	<input checked="" type="checkbox"/>	0,3073	None(primary metric: val_log_loss architecture: net dataset: train=(get_val=null, get_test=null), holdout=(get_dev=null))
tv21 holdout:0,4	<input checked="" type="checkbox"/>	0,3077	None(primary metric: val_log_loss architecture: net dataset: train=(get_val=null, get_test=null), holdout=(get_dev=null))
test_near_replace_unknowns_1_01 holdout:0,4	<input checked="" type="checkbox"/>	0,3092	None(primary metric: val_log_loss architecture: net dataset: train=(get_val=null, get_test=null), holdout=(get_dev=null))

# Как выглядел код

tv\_lastName6

[Launch Experiment](#) [Validate](#) [Analyze Data](#) [Analyze Predictions](#) [Create Duplicate](#)

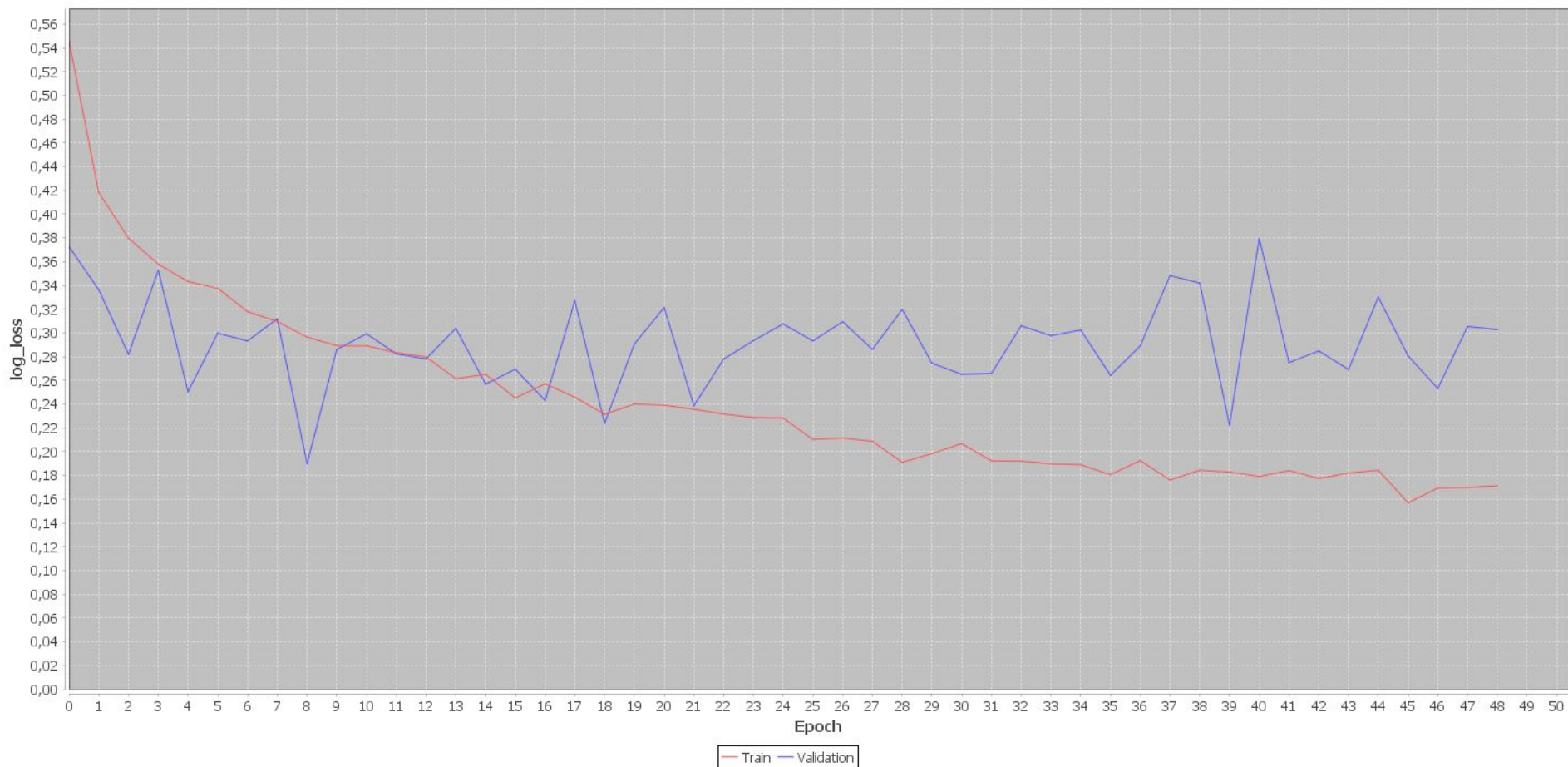
```
declarations:
  net:
    - flatten
    - dropout: 0.5
    - dense: [144, relu]
    - dense: [3, softmax]
  preprocess:
    - try_search_near:
    - replace_unknowns_preprocessor:
        gender: true
        cased: false
        lastNames: true
    - bert_encode:
        maxToken: true
    - disk-cache:
        split: true
    - extract_token_bert_scores_3class:
        startlayer: 9
  inference_batch: 128
  folds_count: 10
  preprocessing: preprocess
  stratified: true
  testSplit: 0.4
  #num_seeds: 10
  architecture: net
  optimizer: Adam #Adam optimizer is a good default choice
  batch: 10 #Our batch size will be 16
```

```
batch: 10 #Our batch size will be 16
metrics: #We would like to track some metrics
  - binary_accuracy
  - binary_crossentropy
  - log_loss
primary_metric: val_log_loss #and the most interesting metric is val_binary_a
primary_metric_mode: min
callbacks: #Let's configure some minimal callbacks
  EarlyStopping:
    patience: 40
    monitor: val_binary_accuracy
    verbose: 1
  ReduceLROnPlateau:
    patience: 16
    factor: 0.5
    monitor: val_binary_accuracy
    mode: auto
    cooldown: 5
    verbose: 1
loss: log_loss #We use simple binary_crossentropy loss
stages:
  - epochs: 50 #Let's go for 100 epochs
dataset:
  train:
    get_val:
    get_test:
  holdout:
    get_dev:
```

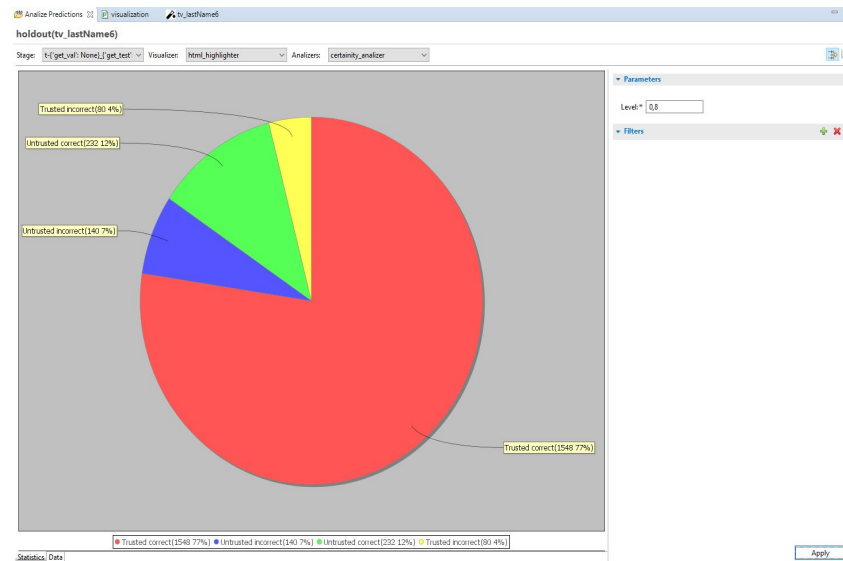
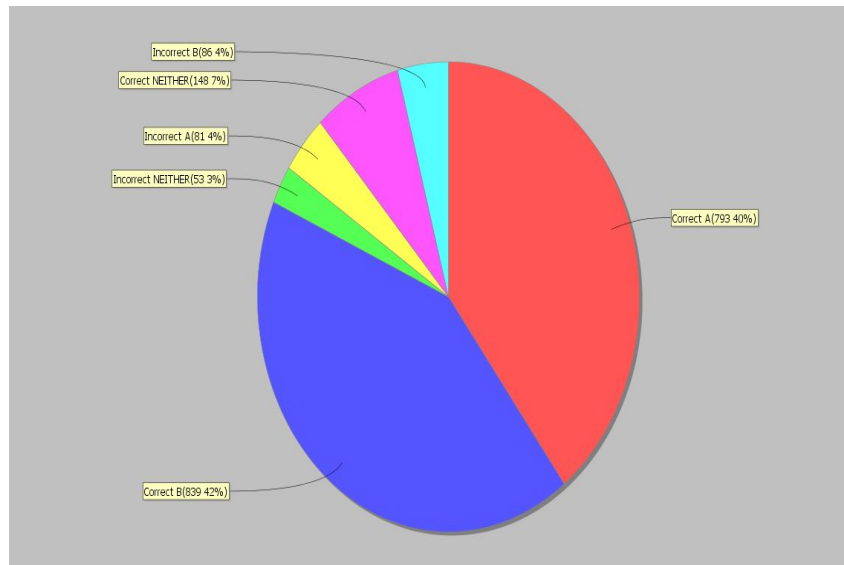
config.yaml Results and Logs Errors

log: fold:8 stage:0

Metric: log\_loss

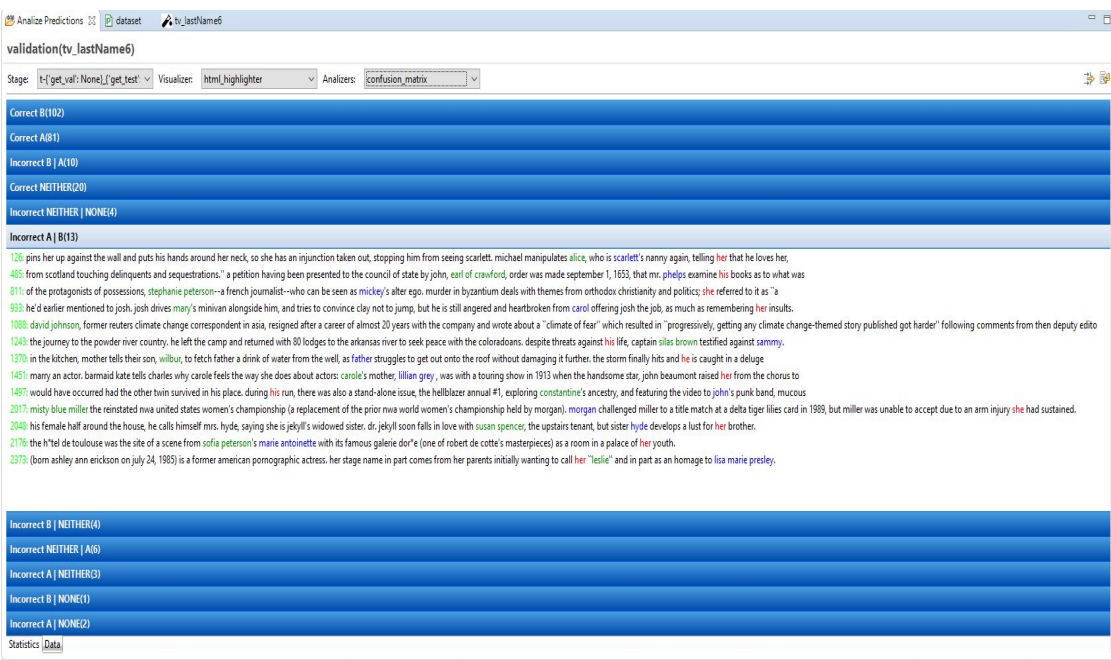
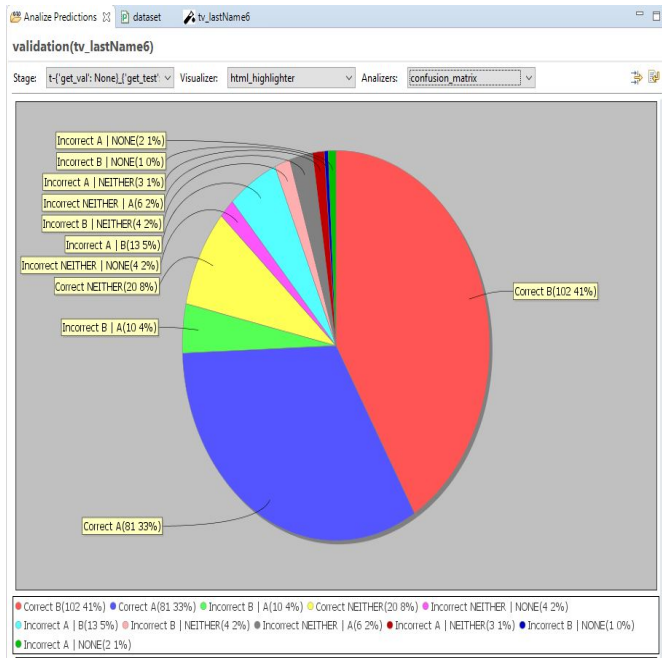


# Как мы смотрели на данные?





# Анализ ошибок



# Примеры препроцессоров

```
@preprocessing.dataset_preprocessor
def stack_flags(input):
    return do_stack_flags(input)

def do_stack_flags(input):
    pFlags = np.expand_dims(input[0], axis=1)
    aFlags = np.expand_dims(input[1], axis=1)
    bFlags = np.expand_dims(input[2], axis=1)
    vTokens = input[3]
    x = np.concatenate((pFlags, aFlags, bFlags, vTokens), axis=1)
    return x

@preprocessing.dataset_preprocessor
def filter_punctuation(input):
    return do_filter_punctuation(input)

def do_filter_punctuation(input):
    (pFlags, aFlags, bFlags, wTokens, text, labels, offsets, wikiRef)
    _wTokens = input[3]
    wordIndices = [i for i in range(len(_wTokens)) if not re.match("[_ -]*$", _wTokens[i]) and re.match("[A-Za-z0-9_-]*$", _wTokens[i])]
    pFlags = input[0][wordIndices]
    aFlags = input[1][wordIndices]
    bFlags = input[2][wordIndices]
    wTokens = input[3][wordIndices]
    text = input[4]
    labels = input[5]
    offsets = input[6]
    wikiRef = input[7]
    result = (pFlags, aFlags, bFlags, wTokens, text, labels, offsets, wikiRef)
    return result

@preprocessing.dataset_transformer
def bert_encode(input, maxToken=False, maxOccurrence=False, norm=None, cased=False):
    if cased:
        return bert_encodeCased(input, maxToken, maxOccurrence, norm)
    namePref = "google-bert-encode" + ("_maxToken" if maxToken else "") + ("_maxOccurrence" if maxOccurrence else "") + ("_norm" if norm is not None else "")
    name = input.name + f"_{namePref}_" if input.name is not None else namePref
    result = BasicDataSet(name, bert_encode_init, input, {"maxToken": maxToken, "norm": norm, "maxOccurrence": maxOccurrence})
    return result

@preprocessing.dataset_transformer
def bert_encodeCased(input, maxToken=False, maxOccurrence=False, norm=None):
    namePref = "google-bert-encode_cased" + ("_maxToken" if maxToken else "") + ("_maxOccurrence" if maxOccurrence else "") + ("_norm" if norm is not None else "")
    name = input.name + f"_{namePref}_" if input.name is not None else namePref
    result = BasicDataSet(name, bert_encode_init_cased, input, {"maxToken": maxToken, "norm": norm, "maxOccurrence": maxOccurrence})
    return result
```



# Чего не хватило до золота

- Времени: большую часть соревнования мы работали не над ним, а над framework, и это сыграло свою роль
- Чуть меньше глупых ошибок приоритизации

# Очевидные улучшения

- Test time aug (простейший вариант улучшает на 0.002)
- Суммировать результаты активации для всех вхождений имен вместо использования активации ближайшего
- Дополнительные фишки, такие как слова из wikipedia ссылки могут использоваться при отсутствии явного победителя
- Использовать несколько pre-trained BERT-ов
- Bert fine-tuning