# Task formulation

The large amount of accumulated production data forms an expert layer valuable for analysis, which can serve as a basis both for the creation of automated tools that use this data and for the creation of a wide range of support and decision-making systems for human experts in various fields.

We need to develop a semantic model that allows us to efficiently define generalized classes for tasks that do not have this information.

# Dataset description

The dataset contains information about approximately 716 thousand tasks. For each of the tasks, information about its name in the construction plan is available, as well as partially specified information about the task hierarchy and generalized name classes to which these tasks belong.

Tasks from the dataset are presented with following information:

- *work_name* - Text name of the task in the construction plan (without preprocessing).

- *generalized_work_class* - Information about the generalized class of work names to which the task belongs.

- *global_work_class* - Information about the highest level of task name generalization.

- *upper_works* - Information about the hierarchy of names of objects and work blocks, within which this task was performed.

3

# Data exploration (1)

Some info about dataset after cleaning of empty values and rows:

- *generalized_work_class* :
    - number of unique generalized classes is 207
    - most frequent class in generalized classes is "Монтаж мк": 26081
- *global_work_class* :
    - number of unique global classes is 56
    - most frequent class in global classes is "Монтаж": 38286
- *work_name* :
    - number of unique work names is 8292
    - most frequent work name in work_name is "пнр": 7498

4

Generalized work classes distrib

Work names distrib over gen classes
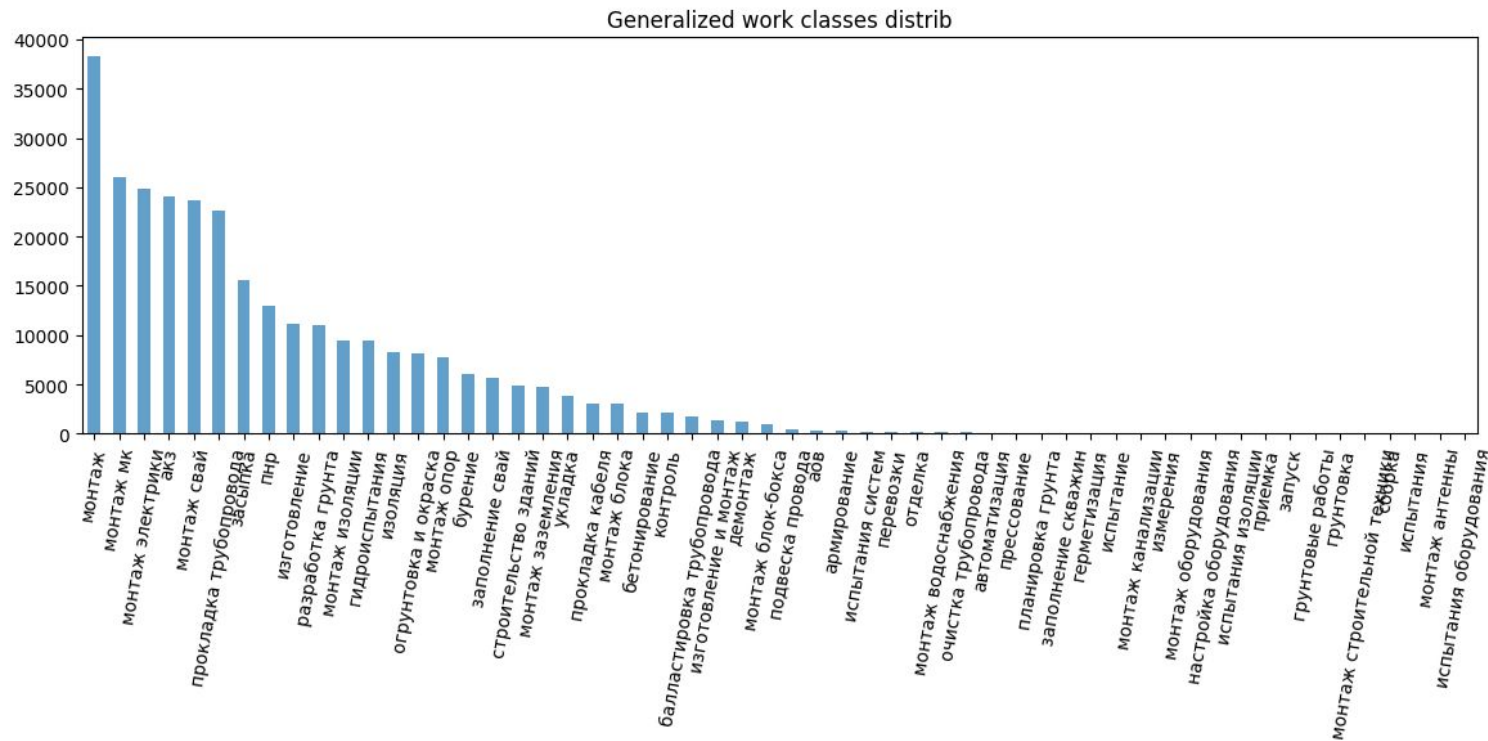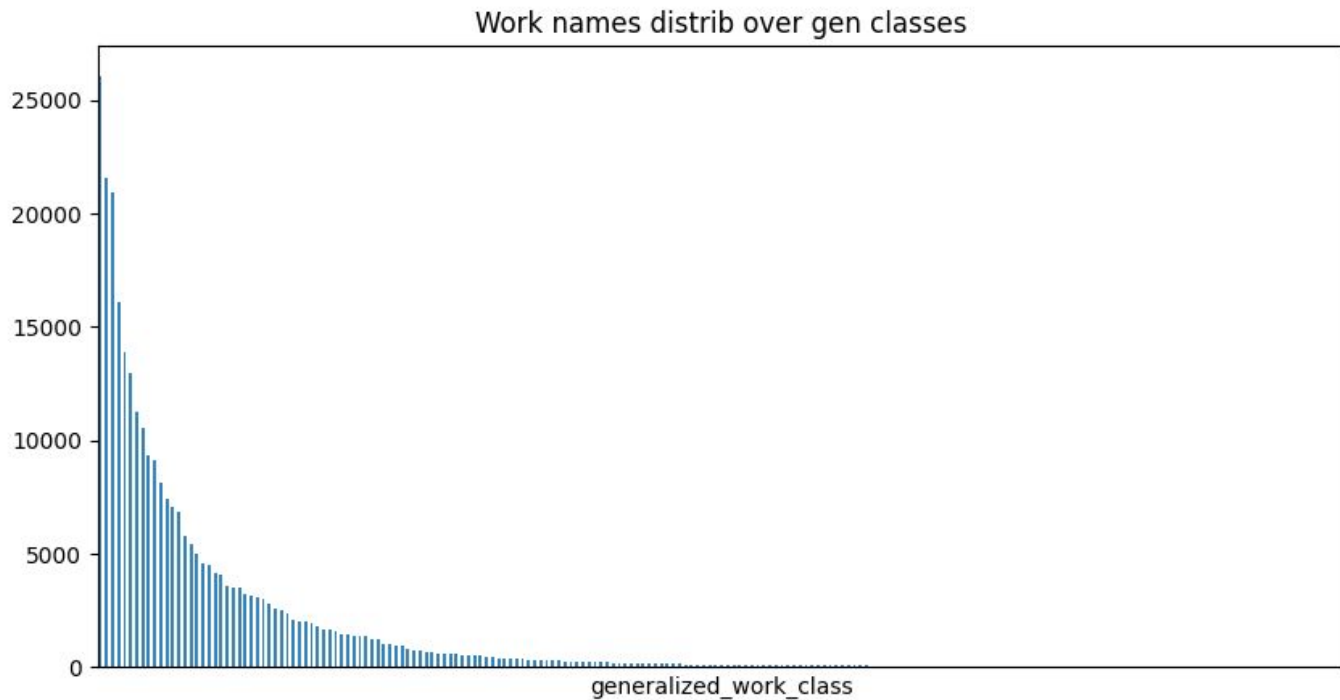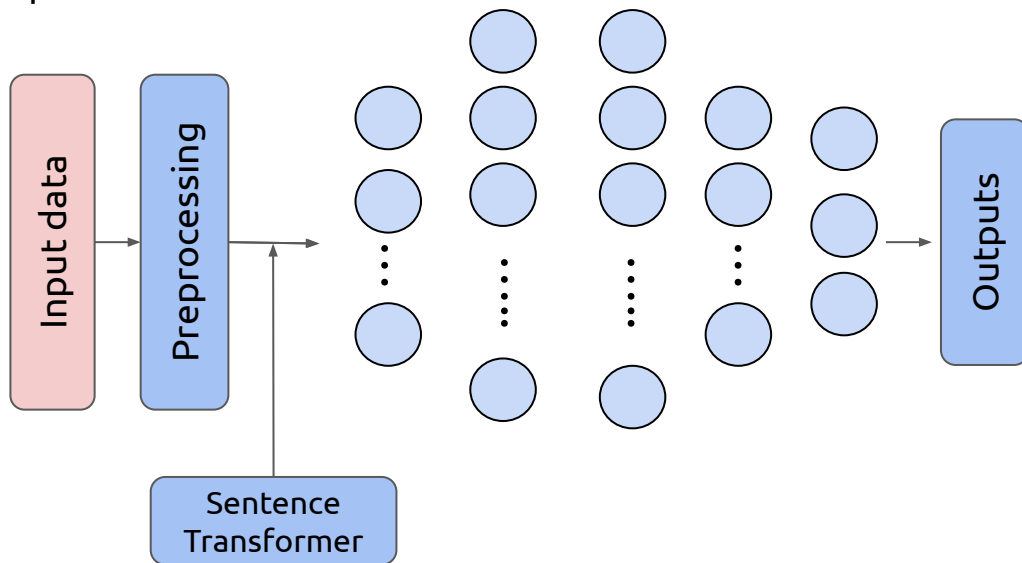
# Model description

For current task we have decided to regular MLP model for our multiclass classification problem



```
Model: "sequential"

Layer (type)                    Output Shape
=================================================
dense (Dense)                   (None, 512)

dense_1 (Dense)                 (None, 1024)

dropout (Dropout)               (None, 1024)

dense_2 (Dense)                 (None, 512)

dense_3 (Dense)                 (None, 207)

=================================================
```

7

# Prediction and Metrics

## Prediction examples:

```
Predicted generalized_work_class --> обратная засыпка --> Actual generalized_work_class --> обратная засыпка
Associated global_work_class --> засыпка --> Actual global_work_class --> засыпка


Predicted generalized_work_class --> обратная засыпка --> Actual generalized_work_class --> обратная засыпка
Associated global_work_class --> засыпка --> Actual global_work_class --> засыпка


Predicted generalized_work_class --> изоляция резервуара --> Actual generalized_work_class --> изоляция резервуара
Associated global_work_class --> изоляция --> Actual global_work_class --> изоляция


Predicted generalized_work_class --> монтаж оборудования --> Actual generalized_work_class --> монтаж оборудования
Associated global_work_class --> монтаж --> Actual global_work_class --> монтаж


Predicted generalized_work_class --> огрунтовка мк --> Actual generalized_work_class --> огрунтовка мк
Associated global_work_class --> огрунтовка и окраска --> Actual global_work_class --> огрунтовка и окраска


Predicted generalized_work_class --> монтаж мк --> Actual generalized_work_class --> монтаж мк
Associated global_work_class --> монтаж мк --> Actual global_work_class --> монтаж мк
```

## Metrics:

*F1 score:* 0.9949
*Accuracy:* 0.9948
*Balanced Accuracy:* 0.9595

# Thank you for your attention!

iT'sMOre than a
UNIVERSITY