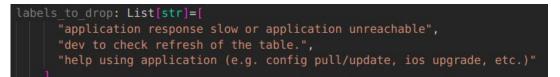# Automated Ticketing System II

Dominik Macko

Tasks Reminder

- Text data of tickets
- Supervised tasks:
    - Team classification
    - Time to resolve prediction
- Unsupervised task:
    - Ticket similarity

Completed Steps

- Data Analysis and Cleaning
- Simple models

Data Analysis and Cleaning

1. Clean labels
   a. Remove sample when row is missing
   b. Calculate minutes to resolve ticket
   c. Drop some noisy teams and unify others

```python
labels_to_drop: List[str]=[
    "application response slow or application unreachable",
    "dev to check refresh of the table.",
    "help using application (e.g. config pull/update, ios upgrade, etc.)"
],
```

```python
token_unifiers: List[str]=[ # if such token appears then unify it around it
    "acs", "action", "addtac", "cisco", "advocacy", "global",
    "gtac", "incm", "netbrain", "igems",
    "security", "unified", "service", "gess"
],
```

Data Analysis and Cleaning

1. Clean labels
   a. Remove sample when row is missing
   b. Calculate minutes to resolve ticket
   c. Drop some noisy teams and unify others
2. Merge data and remove duplicates

Data Analysis and Cleaning

1. Clean labels
   a. Removesample when row is missing
   b. Calculate minutes to resolve ticket
   c. Drop some noisy teams and unify others
2. Merge data and remove duplicates
3. Split labeled data (6:2:2)

Data Analysis and Cleaning

1. Clean labels
   a. Remove sample when row is missing
   b. Calculate minutes to resolve ticket
   c. Drop some noisy teams and unify others
2. Merge data and remove duplicates
3. Split labeled data (6:2:2)
4. Analyzed text data for most frequent tokens, ..

Supervised Models

- tf-idf vectorization followed by:
    - regularized regression or logistic regression
    - naive bayes classifier
    - linear support vector
    - random forest

Unsupervised Models

- tf-idf followed by cosine similarity
- bm25 search function ("tf-idf on steroids" used in text retrieval)

Next Steps

1. Use pretrained word embedding methods
2. Use Transformers and Sentence Transformers
3. Evaluate all the models
   a. For ticket similarity label some test data
4. Deploy a simple app

Thank you for your attention!