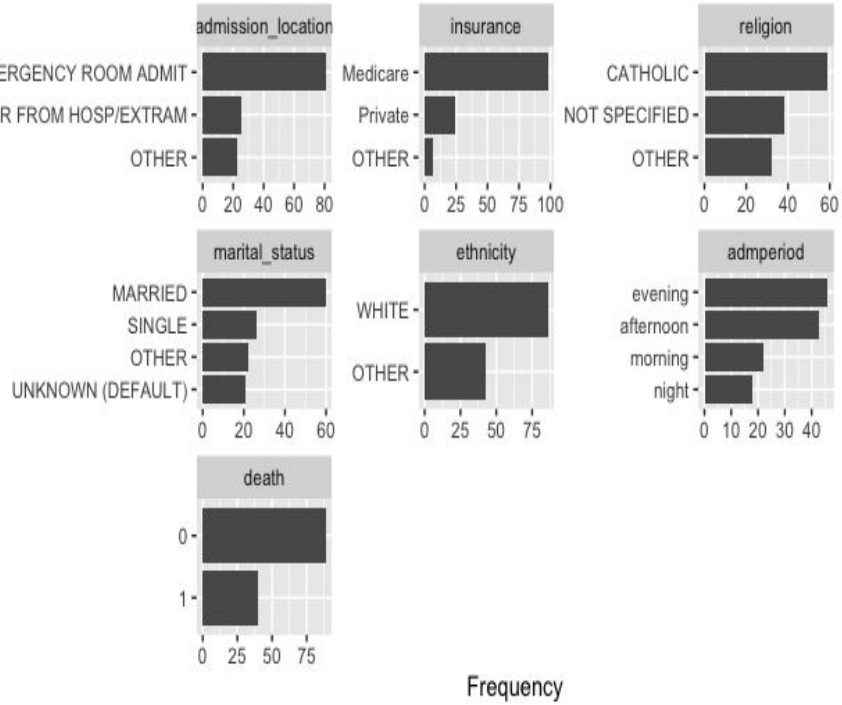
Don't judge patient by the cover

Yevhenii Vinichenko - data extraction and engineering
Mikołaj Malec - data engineering; modeling
Patryk Wrona - sick

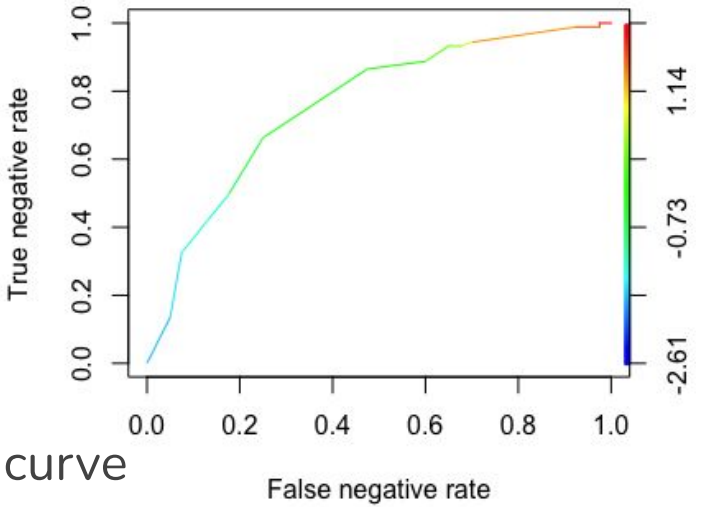
Data description and selection

- ADMISSIONS.csv table from MIMIC Demo was used
- group variables that represent time by day parts
- delete columns that don't meet the assumptions of the task
- group categories that have less than 20 cases
- generalise all null types with one



Methodology

- one hot encoding
- application of logarithmic model
- selecting significant features (AIC)
- select splitting value based on the false-positive rate visualize by ROC curve
- penalty function: weighted accuracy taking into account the proportion of fatalities
- cross-validation of the smaller model on the set
- compare the result with a dummy model



Results and discussion

```
      our_model always.false always.true
Cost: 0.5237375     0.511883    0.488117
> area_under_curve_cv
[1] 0.2935059
```

- Unfortunately model is worse than dummy models, which means that information are not enough (or very small sample or not enough features)
- We modeled problem by classification tree but it had even worse score (decisions tree have tendency to overfitting)
- We could find any evidence suggesting that patients personal feachers are enough to predict patients survivability, but more data is needed