Norwegian Labour Inspection Authority

NTNU
Norwegian University of Science and Technology

A Dataset for Efforts Towards Achieving the Sustainable Development Goal of Safe Working Environments

Eirik Lund Flogard Ole Jakob Mengshoel

Neural Information Processing Systems 2022

Labour inspections

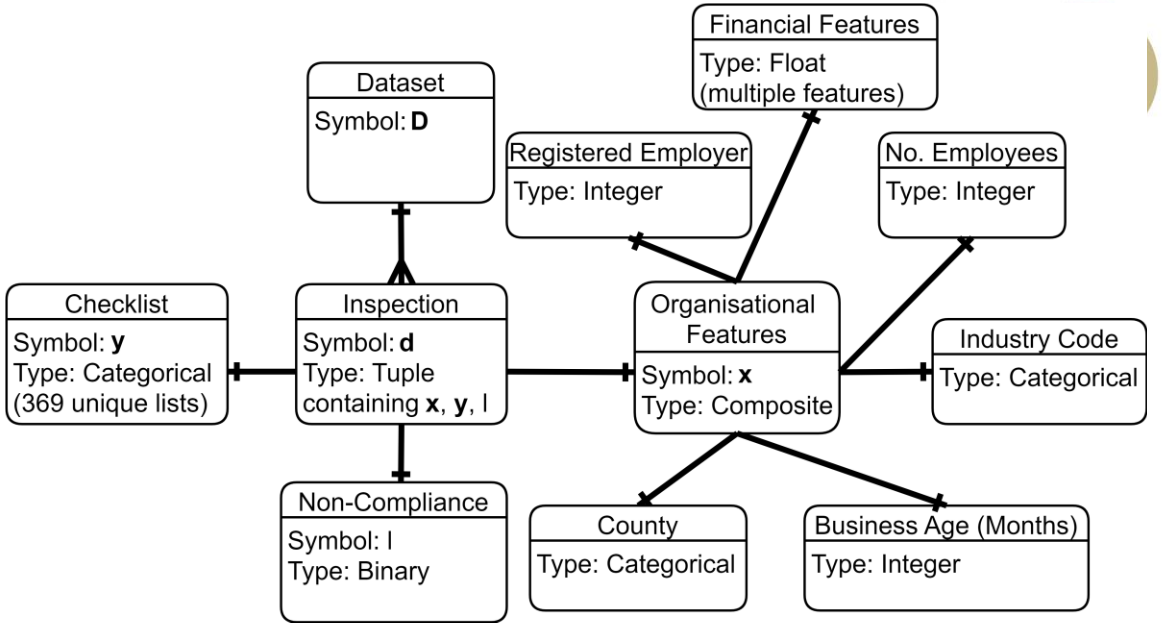
Labour inspections are conducted in order to promote safe working environments(SDG 8)

Research suggests that machine learning can be used to improve labour inspections.

We introduce a new dataset called the Labour Inspection Checklists Dataset (LICD), which could be used to build ML models

Labour Inspection Checklists Dataset

- The dataset contains the results of 63634 inspections conducted by the Norwegian Labour Inspection Authority (NLIA).
- Each instance in the dataset represents a past inspection and contains
  - An organisation, which is the target for the inspection.
  - Target 1: A checklist which is used to survey the inspected organisation for non-compliance.
  - Target 2: A binary indicator of whether or not non-compliance was found in the inspected organisation.

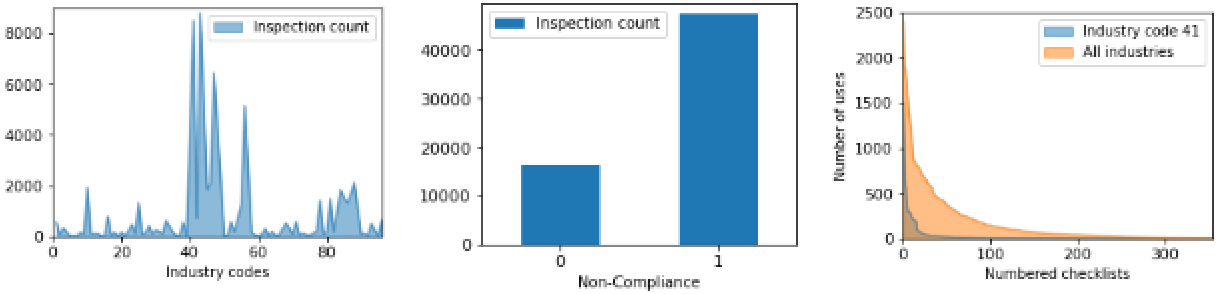


NTNU Open Research Data:
https://doi.org/10.18710/7U6TZP

Arbeidstilsynet

Analysis of the dataset

- Summary of the analysis:
  - Labour inspections are industry-oriented and most inspections are focused on industry codes from 50 to 60.
  - Non-compliance found in 74% of the inspections.
  - The use of checklists follows a long-tailed distribution.



(a) Histogram of inspections over industry codes. (b) Distribution of non-compliance in LICD. (c) Histogram of checklists.

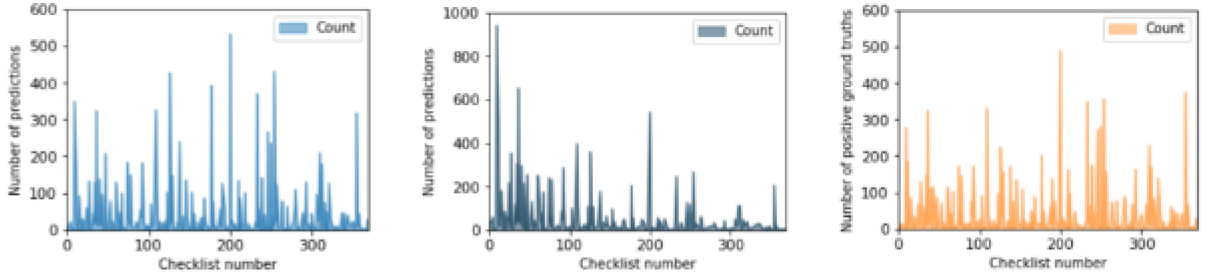
Figure 2: Histograms of inspections, non-compliance and checklists with discrete unit bins on the horizontal axes. The vertical axes on the figures represent the number of occurrences in LICD.

Experiment 1: Selecting Checklists

- Let there be a selection of N checklists and a target organisation $\boldsymbol{x}$. Given the N possible checklists, select the best checklist y to survey the target organisation $\boldsymbol{x}$.
- In this setting, the best checklist is the checklist that its user considers to be most relevant for surveying $\boldsymbol{x}$.

Method	*Mutual Info*				*Anova F*				*Time*
	Bal. Acc	*Acc*	*Prec*	*Rec*	*Bal. Acc*	*Acc*	*Prec*	*Rec*	
LR	.01±0	.02±.01	0±0	0±0	.01±0	.03±.01	0±0	.01±0	381
NBC	.01±0	.01±0	.01±0	.01±0	.02±0	.01±.01	.01±0	.02±0	**13.6**
DT	**.06±.01**	**.09±.01**	**.05±0**	**.04±.01**	**.06±.01**	**.09±.02**	**.05±0**	**.05±.01**	33.2
k-NN	.05±.01	.08±.01	.05±.01	.03±0	.04±0	.07±.01	.04±0	.03±0	40.0
AdaBoost	.01±0	.05±.02	0±0	.01±0	.02±0	.04±.02	0±0	.02±0	619
GradientBoost	.03±.01	.06±.02	.03±.01	.02±.01	.04±.01	.08±.02	.03±0	.03±0	22316
MLP	.01±.01	.04±.02	.01±.01	.01±0	.02±.01	.05±.02	.02±.01	.02±.01	338

Table 1: Prediction performance with average standard deviations and run times for CLSP on LICD. Times are measured in seconds.



(a) Distribution of predictions for DT. **(b)** Distribution of predictions for k-NN. **(c)** Distribution of ground truth labels.

Figure 3: Distributions on the evaluation set of a random paired 80-20 training-evaluation split. The horizontal axes represent the identifiers for 369 possible checklists (classes). The vertical axes on the figures represent the number of observations for each class in the evaluation set.

Experiment 2: Classifying Non-compliance

- Binary classification: The class label $c \in \{0,1\}$ belongs to a Bernoulli distribution.
- Given a checklist y and a target organization $\boldsymbol{x}$:
  - classify into compliant ($c = 0$) versus non-compliant ($c = 1$) to any of the regulations given by the content of y.

Method	χ^2					*Anova F*					*Time*
	Bal. Acc	*Acc*	*Prec*	*Rec*	*Auc*	*Bal. Acc*	*Acc*	*Prec*	*Rec*	*Auc*	
LR	.42±.02	.44±.02	.68±.02	.46±.02	.41±.02	.46±.02	.45±.02	.72±.02	.43±.02	.47±.02	3.56
NBC	.56±.04	.49±.11	.72±.20	.42±.18	.57±.04	.57±.02	.53±.02	.81±.02	.48±.02	.59±.02	**1.06**
DT	.54±.01	.45±.02	.59±.01	.35±.03	.57±.01	.51±0	.30±0	.20±0	.08±.01	.56±.01	14.2
k-NN	.58±.02	**.53±.04**	.81±.01	**.49±.07**	.61±.02	.57±.02	.52±.02	.79±.06	.47±.03	.61±.02	100
AdaBst	**.58±.01**	.51±.03	**.82±.01**	.44±.07	**.63±.02**	**.62±.01**	**.57±.02**	**.84±.01**	**.52±.04**	**.68±.02**	241
GradBst	**.58±.01**	.50±.03	**.82±.01**	.43±.06	**.63±.02**	**.62±.01**	**.57±.02**	**.84±.01**	.51±.04	**.68±.02**	1352
MLP	.53±.01	.40±.03	.78±.02	.27±.05	.53±.01	.54±.01	.39±.02	.82±.02	.23±.05	.56±.02	27.7

Table 2: Results with average standard deviations for NCP on LICD. The average time in seconds per cross-validation is shown on the far right column.

Conclusion

- We introduced a new dataset called LICD.
  - 63634 past inspections carried out by the Norwegian Labour Inspection Authority.
  - Consists of 575 features and 2 target variables.
- We introduced two problems which can be solved via ML.
- Strong prediction performance on the two problems can be difficult to achieve.
- Future work could:
  - Investigate ML or feature selection methods, to improve classification performance.
  - Explore other variants or even combinations of the two problems.

Thank you for listening