Human Environment and Transport Inspectorate
Ministry of Infrastructure and Water Management

Universiteit Leiden
Leiden Institute of Advanced Computer Science

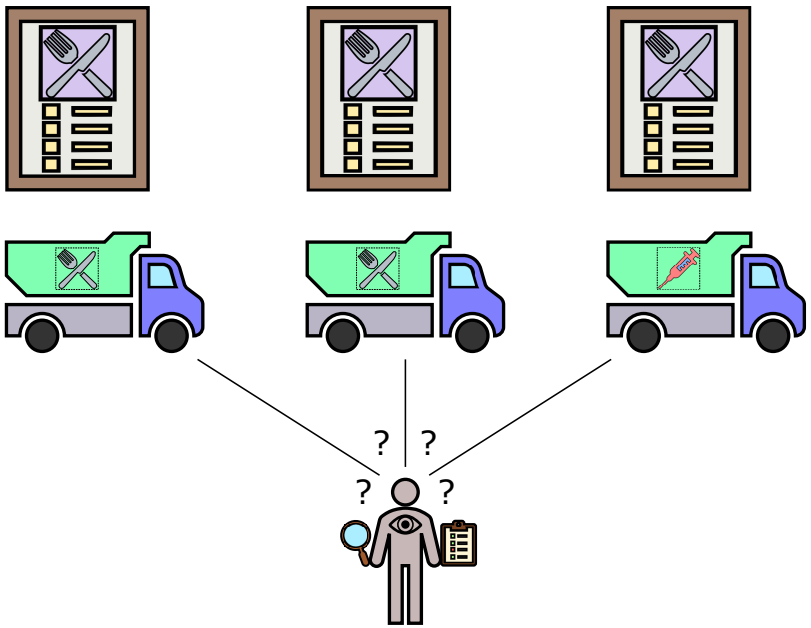
Data-driven Risk Assessment in Infrastructure Networks

António Pereira Barata[1,2], Gerrit-Jan de Bruin[1,2], Frank Takes[1], Cor Veenman[1,3], and Jaap van den Herik[1]
{a.p.pereira.barata, g.j.de.bruin}@liacs.leidenuniv.nl

[1]Leiden Centre of Data Science
Leiden Institute for Advanced Computer Science, Leiden University
[2]Human Environment and Transport Inspectorate,
Ministry of Infrastructure and Water Management
[3]Data Science Department, TNO

ICT.open conference, March 2018

Waste Shipment Regulation



- Ensure proper waste handling
- Enforcement by ILT
  - Notices
  - Deposit
  - Inspections
- Possible company misconduct
  - Cheaper label

Goal

Use outlier detection for finding **mislabeled** notices.

- Strongly deviating notices (within same reported class)

Data

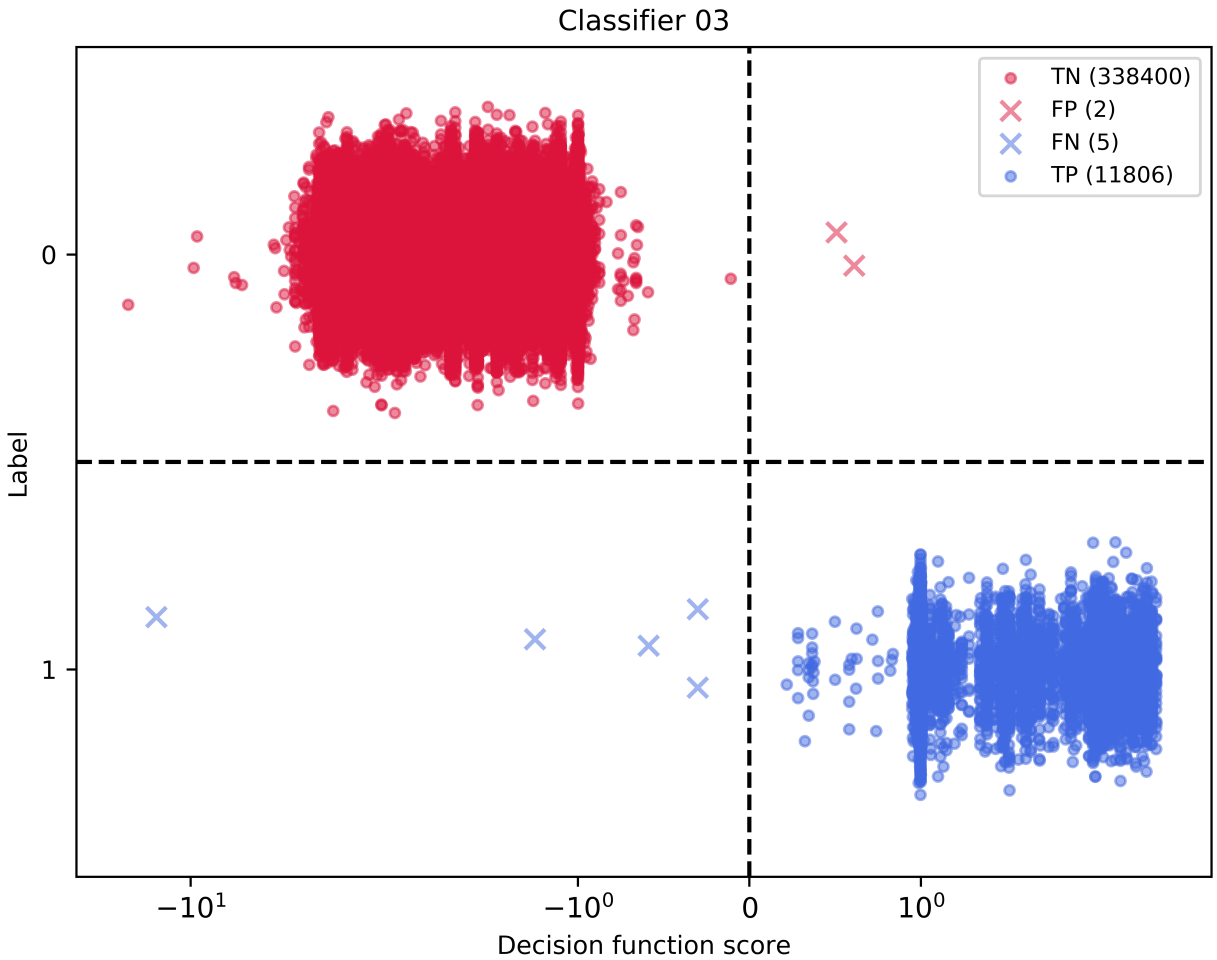
Labeled data

- 20 waste types
- Features
  - Company name
  - Waste tonnage
  - Border crossing
  - Processor location
  - ...

Approach

- Target = waste category
- Linear SVM classifier
- One-vs-rest
- Five-fold train-test split
- Average precision metric
- Lowest scoring false negatives

Results


Classifier 03
TN (338400)
FP (2)
FN (5)
TP (11806)
Label
0
1
-10^1
-10^0
0
10^0
Decision function score

Future work

- Compare results with inspection data.
- Cooperate with domain experts (inspectors).

- Application of data-driven approaches to other data problems.

Human Environment and Transport Inspectorate
Ministry of Infrastructure and Water Management

Universiteit Leiden
Leiden Institute of Advanced Computer Science

Thank you for your attention!
Are there any questions?