Finnish Institute of Occupational Health

Developing safety management with data mining – new understanding on factors enhancing and weakening safety

Maria Tiikkaja et al.

AI Safety research project

- Developing safety management with data mining
- 9/2020 – 3/2023

Työsuojelurahasto
Arbetarskyddsfonden
The Finnish Work Environment Fund

- Funded by the Finnish work environment fund, Finnish Institute of Occupational Health (FIOH) and participating companies
- More info: www.ttl.fi/en/aisafety

KONECRANES
Metso:Outotec
prefere phenolics
sappi

The aims of the research project

- To use machine learning to detect silent signals behind occupational accidents and safety deviations
- To improve the focus of the collection of safety related data in companies
- To develop novel data-driven ways to improve occupational safety management in the framework of human factors (HF)
- To give knowledge to data-driven safety research by determining what format and types of data sets are useful for integration

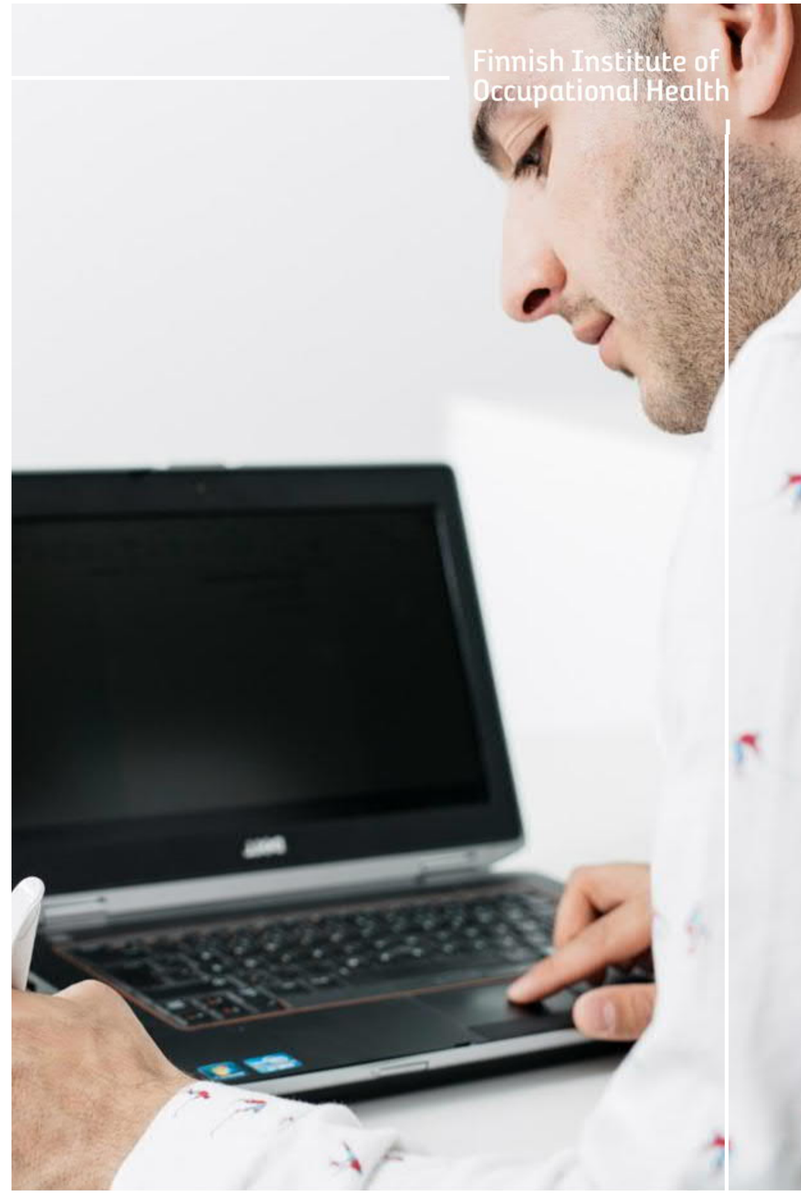
Diverse datasets are analyzed and integrated

Safety data

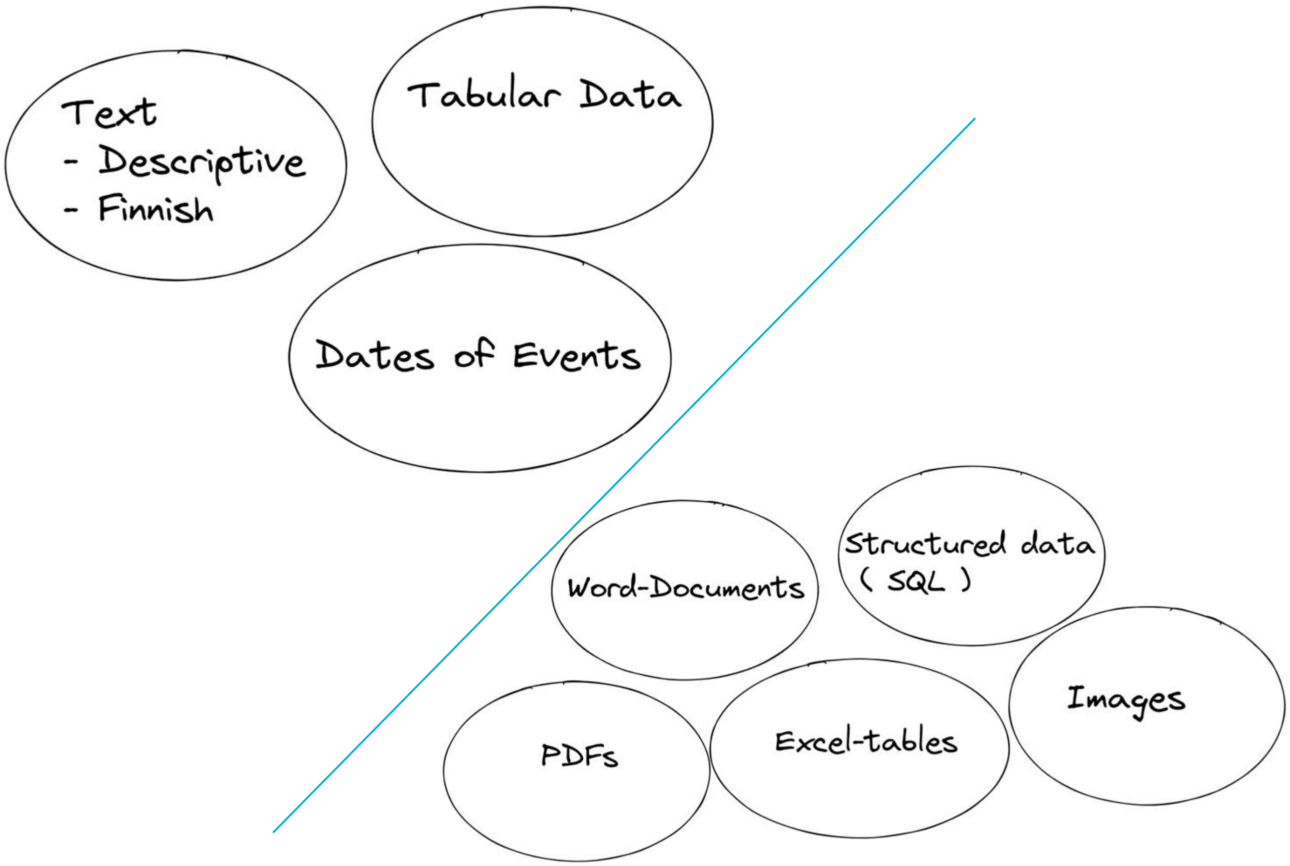
Data from HR systems

Work environment data

Management data



Data comes in various types and formats



Exemplification of research data

Occupational accidents	Variables+free text	Basic employee data	Variables	Production volume	Daily figures
Near misses	Variables+free text	Training data	Variables	Incoming orders	Daily figures
Safety observations	Variables+free text	Working hours and shifts	Variables	Outgoing ordesrs	Daily figures
Safety discussions	Variables+free text	Sick absence compilation	Weekly/monthly figures		
Accident investigations	Free text	Organisational restructuring			
Safety walks	Free text	Workplace surveys	Free text		
Risk assessments	Variables+free text	Staff surveys	Figures by topics		

Research goals

1. To produce information on data delivery and use in research
2. To study if better analytics on occupational accidents and near misses can be enabled by integrating heterogenous data
3. To identify information gaps in the informal text descriptions of accident reports and in other text data
4. To develop data-based monitoring and prediction of occupational accidents and to produce information to support decision making

Finnish Institute of Occupational Health

Research questions

Goal 1

- How can organisations/workplaces securely and ethically give their employees' personal data to research?
- How does ethical and data protection questions affect on the integration of datasets?

Goal 2

- What other available datasets than safety data could explain safety situations (accidents and near misses)?
- What kind of groups (clusters) can be made for different safety situations and what and why stays outside of them?
- What silent signals of occupational safety can be identified data-driven?

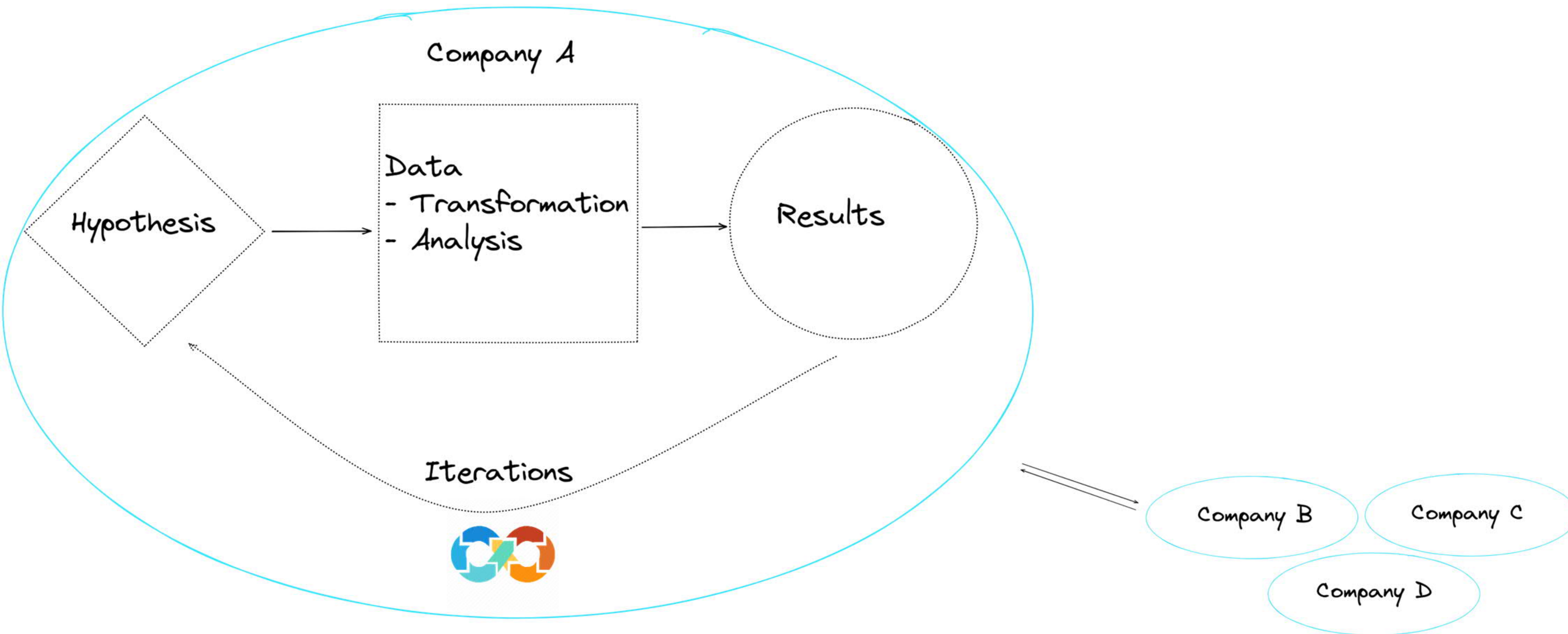
Goal 3

- Are similar accidents recurring in safety reports?
- Are the workplaces getting enough information on the causes of the accidents to prevent them happening again?

Goal 4

- How can machine learning methods be used to predict the development of safety?
- Can the safety development reliably be predicted by integrating safety data with other data sets?
- What kind of knowledge management processes and systems are worthwhile to build when trying to predict safety development?

Flowchart of Analysis



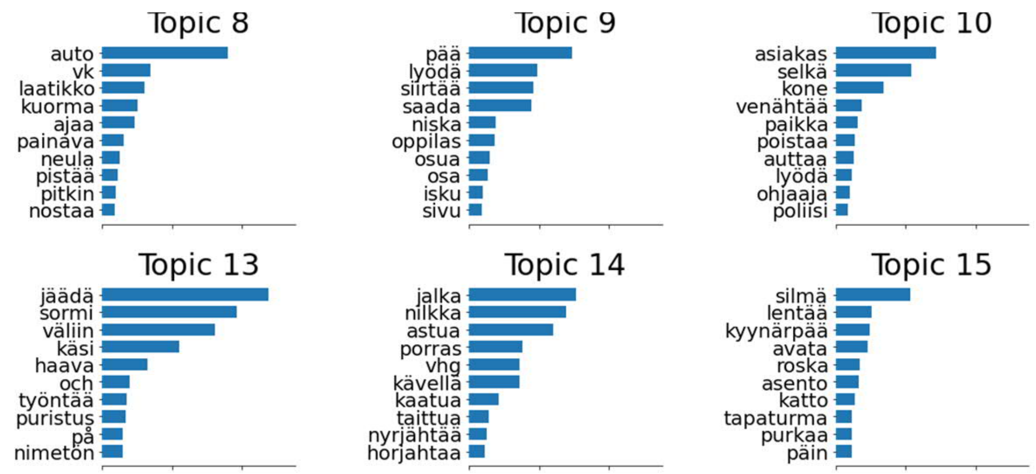
Finnish Institute of Occupational Health

Examples of analyses answering the goal 2

- Correlation of variables x, y (all the viable variables in our data) with occupational accidents and near misses (= unwanted occurrences)
- The number of participants in safety briefings predict the occurrence of unwanted occurrences in a workplace
- The number of safety observations correlates with the number of unwanted occurrences
- Correlation between safety observations and safety walks
- Correlation between production volume and absences due to illness
- Correlation between production volume and occupational accidents
- Correlation between the number of staff at work with unwanted occurrences
- Correlation between the amount of safety training and unwanted occurrences

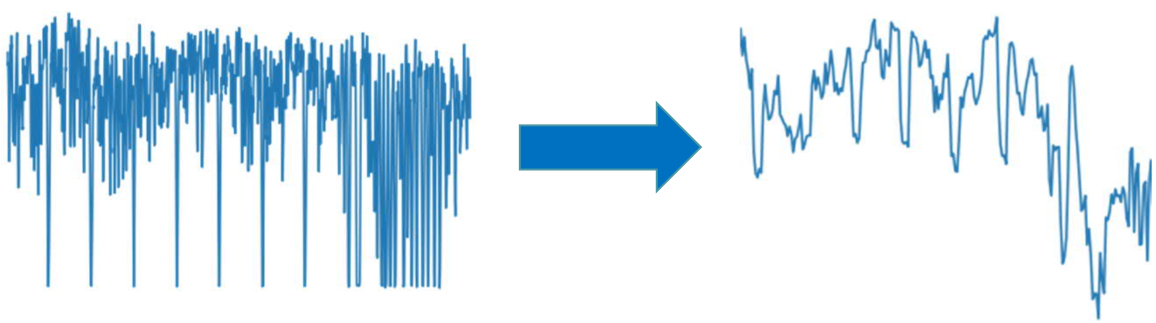
Topic Modeling of Finnish Text using Latent Dirichlet Allocation (LDA)

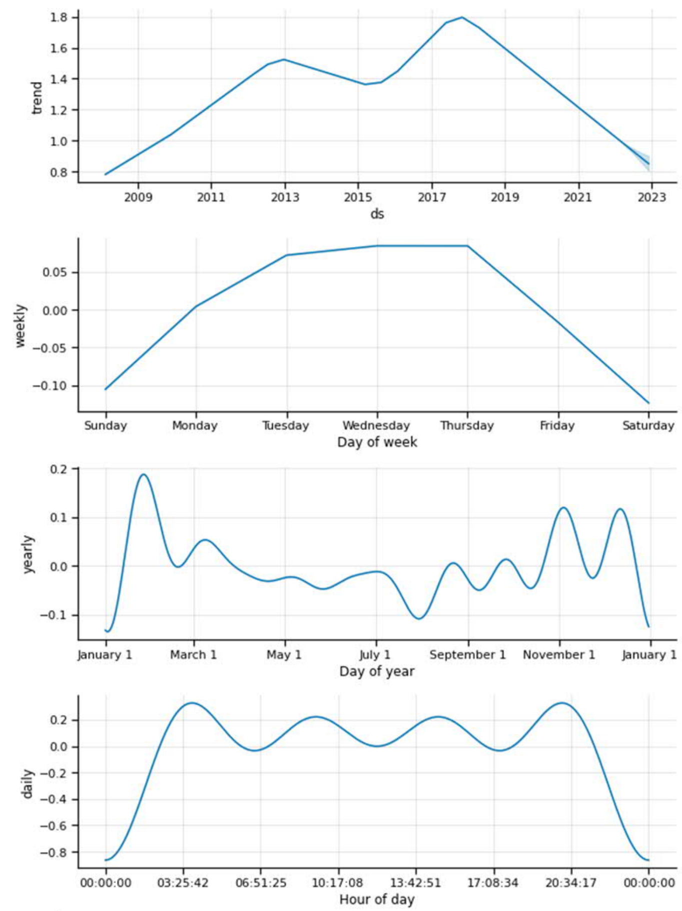
- Extract 1,000 different accident classes from training data
  - Hand accidents, slip accidents etc
  - Training data from the Finnish Workers’ Compensation Center
  - 468,609 accident descriptions between 2003 and 2019
- Apply the model to company-specific data and to timeline



Time Series Forecasting using Facebook's Prophet model

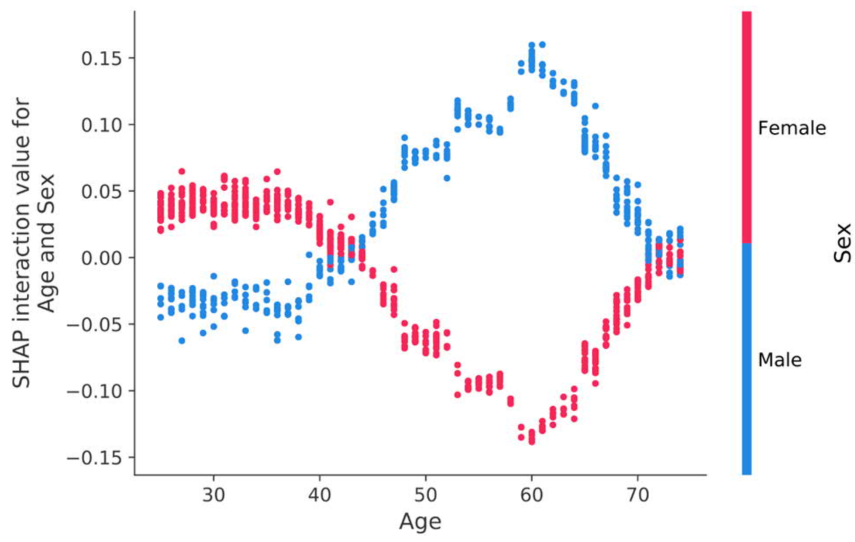
- Build various Safety Status indices of accidents and near-miss observations
  - Indices computed from accident seriousness / days of sick-leave
- Extract trends and seasonalities





Explaining the XGBoost model using Shapley Values to extract most important variables as ”silent signals”

- *A prediction can be explained by assuming that each feature value of the instance is a “player” in a game where the prediction is the payout. Shapley values – a method from coalitional game theory – tells us how to fairly distribute the “payout” among the features.*
  quote from https://christophm.github.io/interpretable-ml-book/shapley.html
- Each player is a unique feature that goes into the model (eg. day of week, production amount, number of workers working at that date, days without accidents).
- Now we can find which features play important role and focus on those better and study the interactions of the features



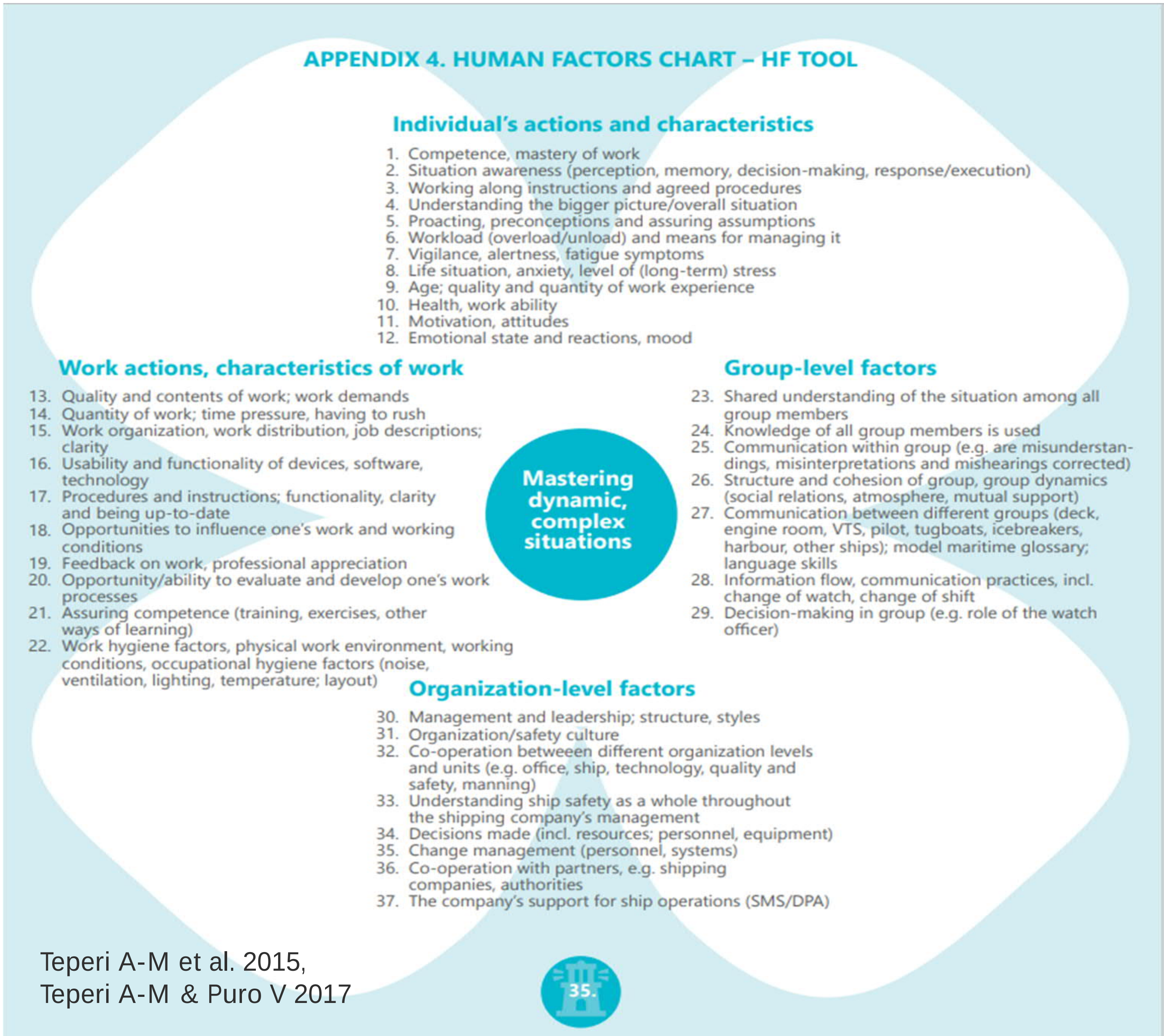
Finnish Institute of Occupational Health

Challenges

- Finnish Language
  - Comes from different language family than English, French, German, Swedish
  - Less models available for deep learning
- Unstructured way of preserving data
  - Data is situated in hand-curated excels that rely on the expertise of the persons working there
  - Data preserved in structured databases but no way to extract the data in the raw format
- The people in charge of safety rarely data experts
- Different companies give the same data as different levels
  - Company A: Person XXX sick 2020/12/12
  - Company B: Relative number of sick-leaves in 2020/12

Qualitative analysis of safety data

- Using human factors (HF) framework and HF tool to analyse occupational accidents, near misses and safety observations
- Answering the research question: Are the workplaces getting enough information on the causes of the accidents to prevent them happening again?



Teperi A-M et al. 2015,
Teperi A-M & Puro V 2017

HF analysis of occupational accident investigations

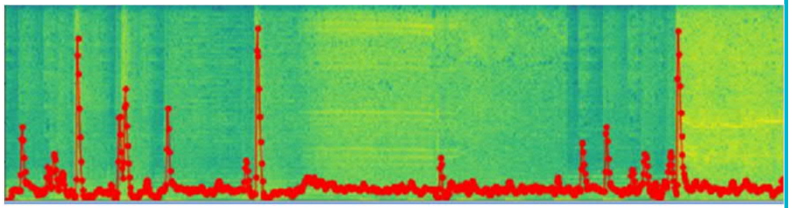
- Text data on accident descriptions and reports (80 investigations in total)
- Theory-oriented classification of the causes and corrective measures identified on the reports -> which HF tool items can be seen (both supporting and detracting factors on safety)
- Overall assessment of the investigation
    - how system versus individual-oriented is the analysis
    - are enough information obtained on the causes to prevent similar incidents to reoccur
    - how just and fair the description is

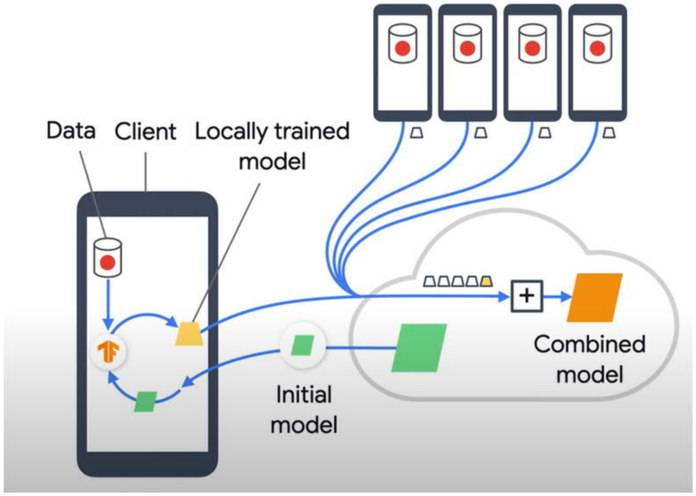
Preliminary findings

- Causes and corrective measures were centered mainly on individual and work level factors while group and organisation level factors were not systematically raised
- ”Human behavior”, ”human error”, and ”carelessness” named as root causes
- Corrective measures often direct
- Positive factors not systematically registered
- In single investigations deviations have offered wider learning windows to organisations’ operation and fluency and safety of work

Using AI to answer our goals

- Silent Signals
  - We want to identify the critical features that play a role in weakening and strengthening safety status
  - By extracting these silent signals computationally allows us to focus the use of the domain expertise to study these phenomena
- Way to monitor changes in safety status
  - Provide the companies a way how to monitor the changes in their safety status
  - Using our findings from their own data





More on the progress of the research and discussion on the topic on workshop:

New Possibilities for Knowledge Management of Occupational Safety in the Framework of Vision Zero

- Organised by NIVA (the Nordic Institute for Advanced Training in Occupational Health)
- 8-11th of November 2022 in Helsinki, Finland
- https://niva.org/course/new-possibilities-for-knowledge-management-of-occupational-safety-in-the-framework-of-vision-zero/

Research team at FIOH

Maria Tiikkaja

Eero Lantto

Tuula Räsänen

Kimmo Sirén

Jukka Kärkimaa

Vuokko Puro

Tarja Heikkilä

Simo Virtanen

Henriikka Kannisto

Finnish Institute of Occupational Health

Thank You!

www.ttl.fi/en/aisafety #aisafety

maria.tiikkaja@ttl.fi

ttl.fi

@tyoterveys
@fioh

tyoterveyslaitos

tyoterveys

Tyoterveyslaitos