

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



Työsuojelurahasto
Arbetskyddsfonden
The Finnish Work Environment Fund

KONECRANES[®]

prefere
phenolics

Metso:Outotec

sappi

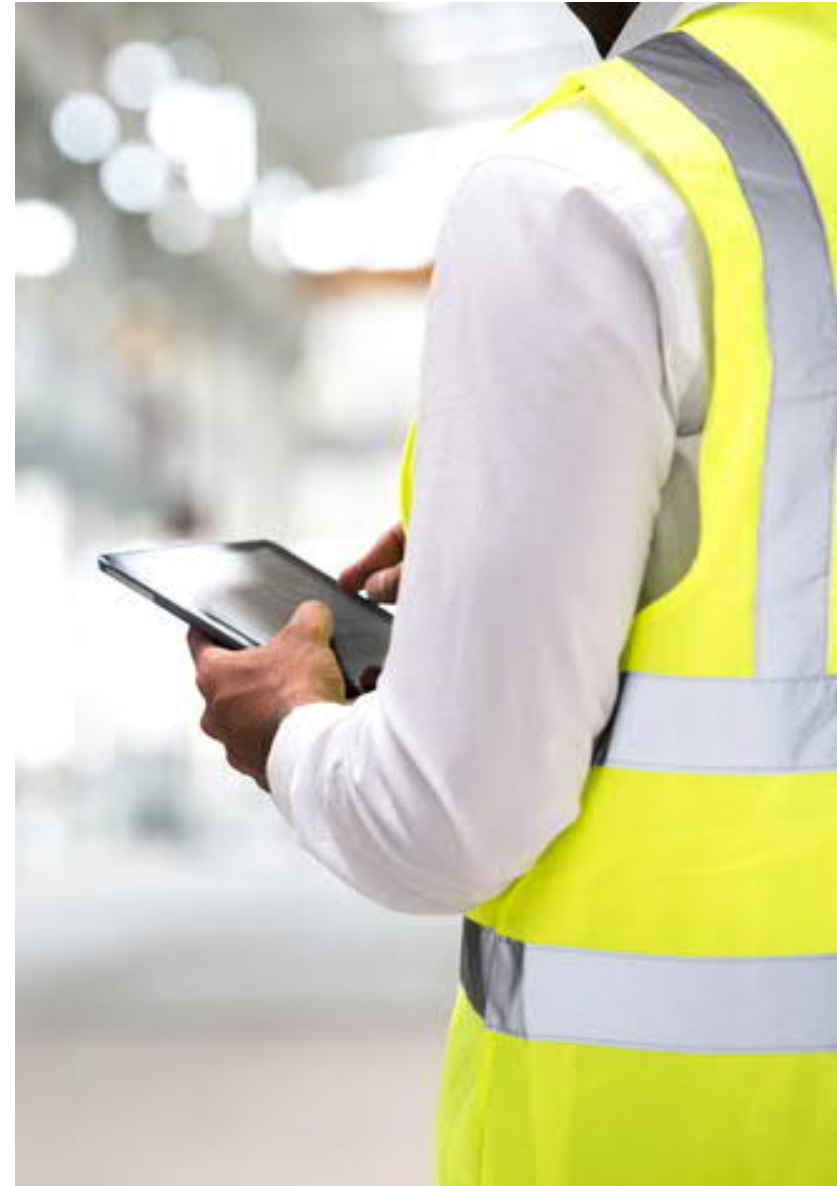
3.10.2022

| © Finnish Institute of Occupational Health |

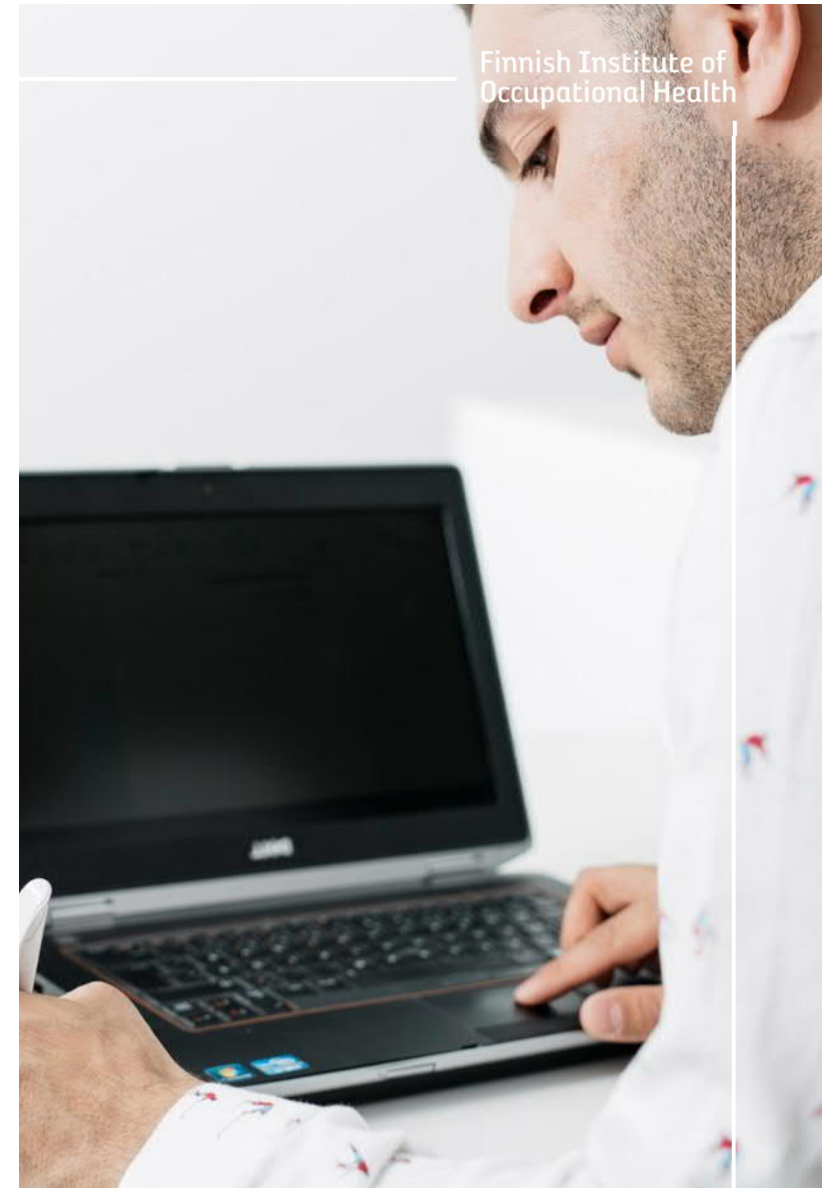
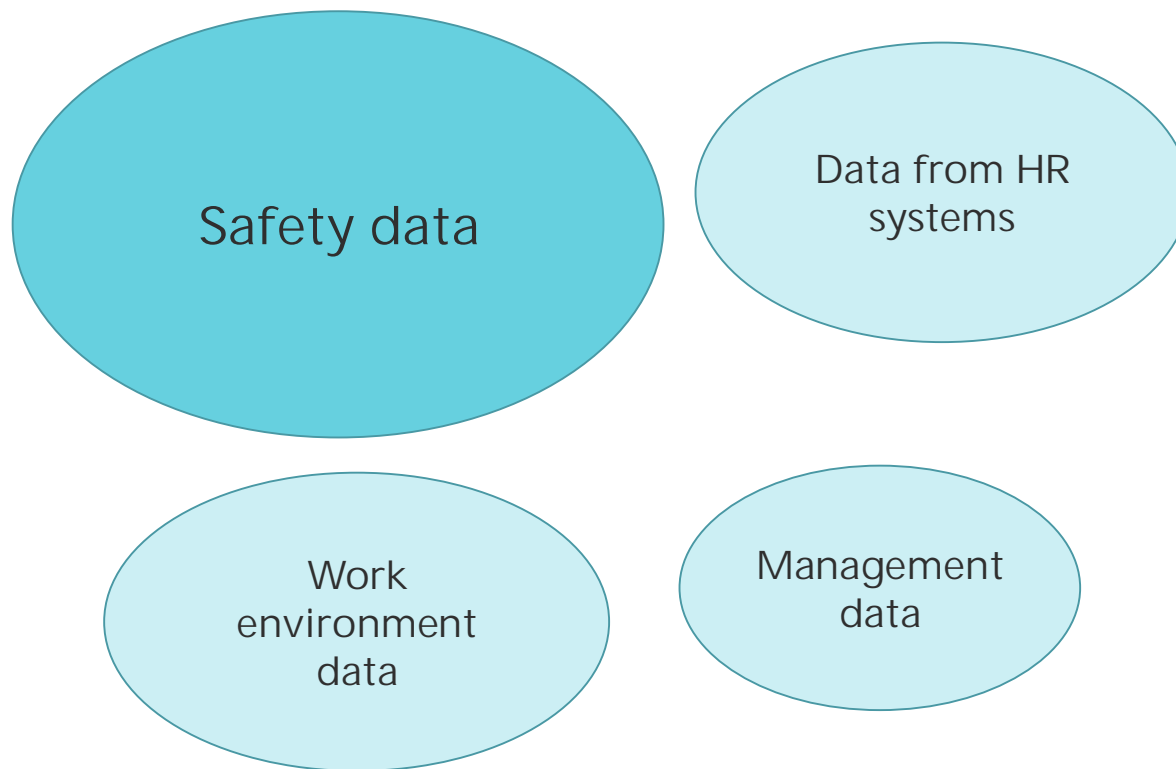


The aims of the research project

- 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



Data comes in various types and formats

Text
- Descriptive
- Finnish

Tabular Data

Dates of Events

Word-Documents

Structured data
(SQL)

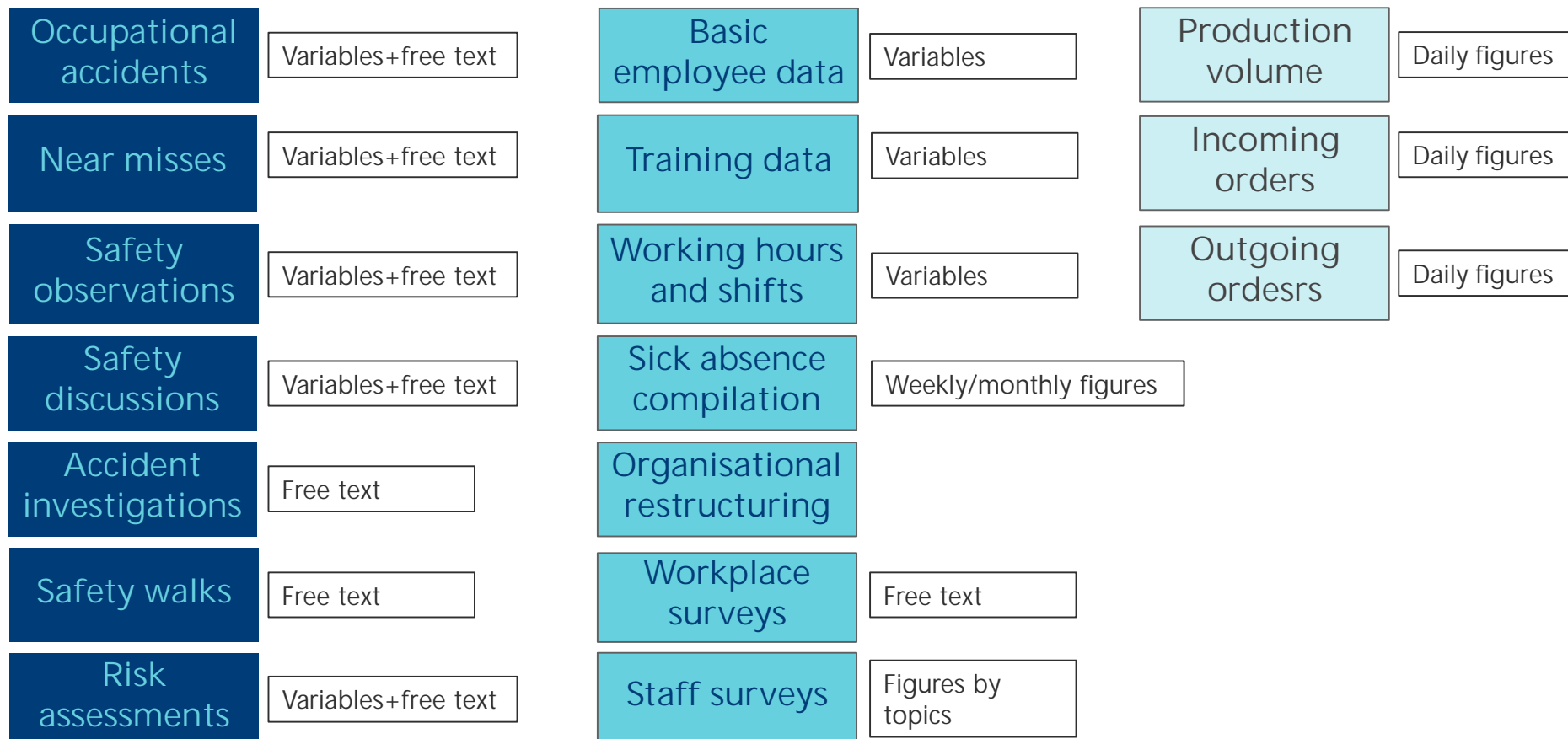
PDFs

Excel-tables

Images



Exemplification of research data



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



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?

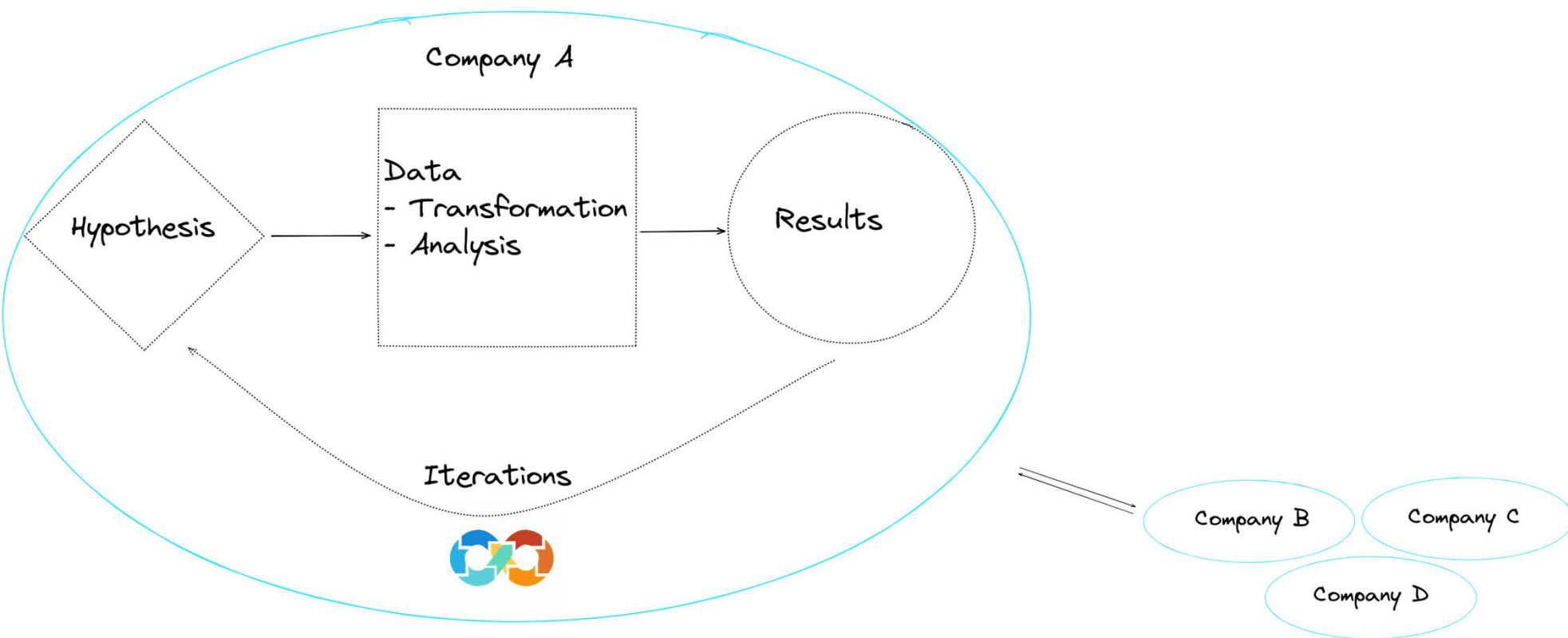
Goal 3

- 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

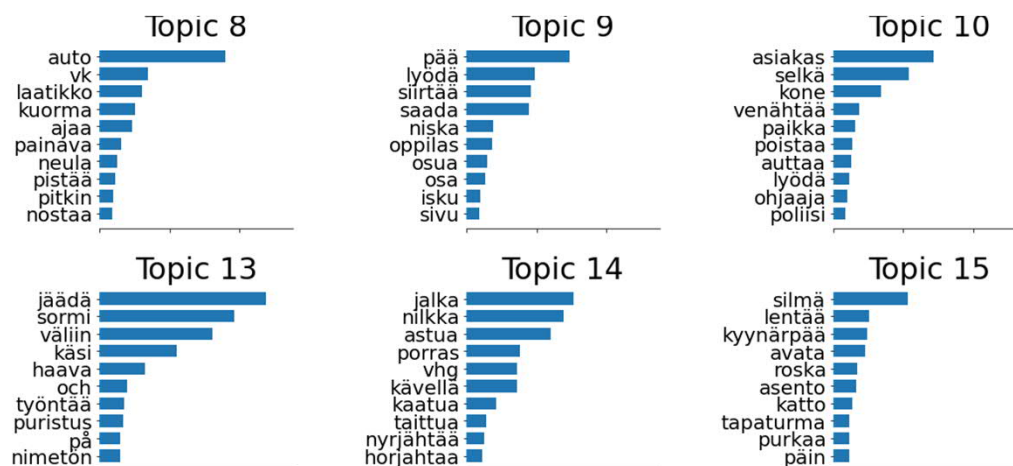


Examples of analyses answering the goal 2

- 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

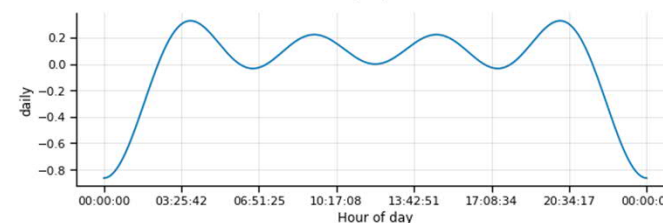
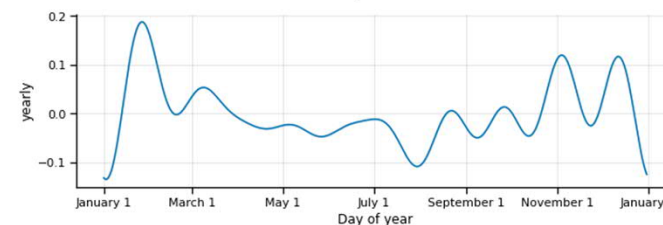
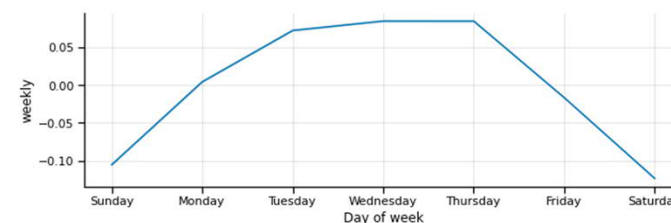
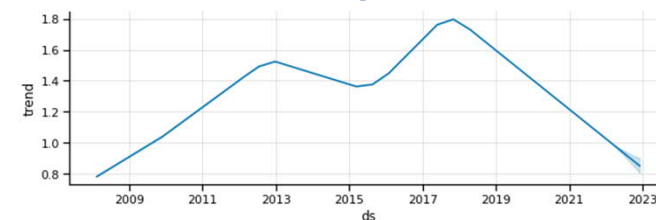
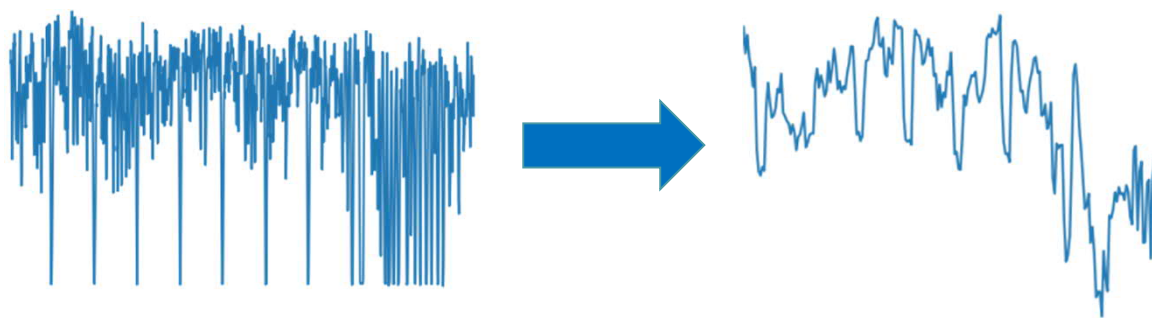
Topic Modeling of Finnish Text using Latent Dirichlet Allocation (LDA)

- 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



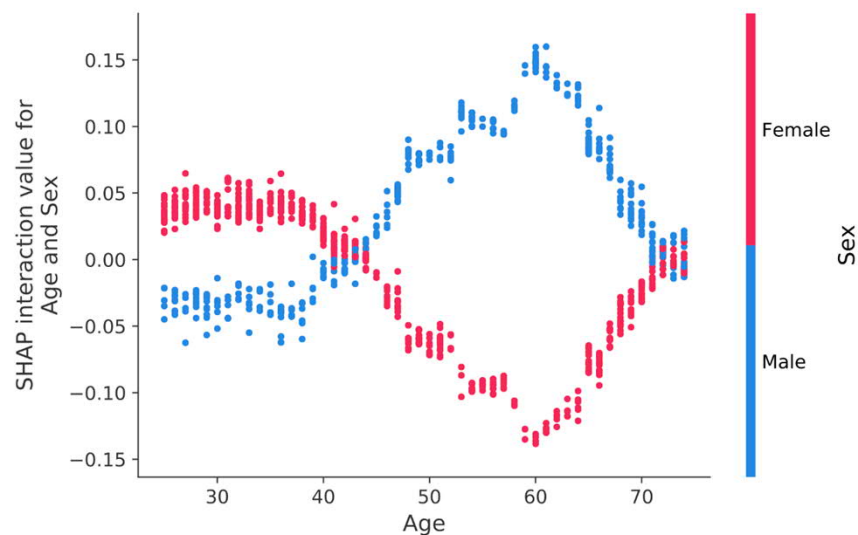
Time Series Forecasting using Facebook's Prophet model

- 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



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?

3.10.2022

APPENDIX 4. HUMAN FACTORS CHART – HF TOOL

Individual's actions and characteristics

1. Competence, mastery of work
2. Situation awareness (perception, memory, decision-making, response/execution)
3. Working along instructions and agreed procedures
4. Understanding the bigger picture/overall situation
5. Proacting, preconceptions and assuring assumptions
6. Workload (overload/unload) and means for managing it
7. Vigilance, alertness, fatigue symptoms
8. Life situation, anxiety, level of (long-term) stress
9. Age; quality and quantity of work experience
10. Health, work ability
11. Motivation, attitudes
12. Emotional state and reactions, mood

Work actions, characteristics of work

13. Quality and contents of work; work demands
14. Quantity of work; time pressure, having to rush
15. Work organization, work distribution, job descriptions; clarity
16. Usability and functionality of devices, software, technology
17. Procedures and instructions; functionality, clarity and being up-to-date
18. Opportunities to influence one's work and working conditions
19. Feedback on work, professional appreciation
20. Opportunity/ability to evaluate and develop one's work processes
21. Assuring competence (training, exercises, other ways of learning)
22. Work hygiene factors, physical work environment, working conditions, occupational hygiene factors (noise, ventilation, lighting, temperature; layout)

Mastering dynamic, complex situations

Group-level factors

23. Shared understanding of the situation among all group members
24. Knowledge of all group members is used
25. Communication within group (e.g. are misunderstandings, misinterpretations and mishearings corrected)
26. Structure and cohesion of group, group dynamics (social relations, atmosphere, mutual support)
27. Communication between different groups (deck, engine room, VTS, pilot, tugboats, icebreakers, harbour, other ships); model maritime glossary; language skills
28. Information flow, communication practices, incl. change of watch, change of shift
29. Decision-making in group (e.g. role of the watch officer)

Organization-level factors

30. Management and leadership; structure, styles
31. Organization/safety culture
32. Co-operation between different organization levels and units (e.g. office, ship, technology, quality and safety, manning)
33. Understanding ship safety as a whole throughout the shipping company's management
34. Decisions made (incl. resources; personnel, equipment)
35. Change management (personnel, systems)
36. Co-operation with partners, e.g. shipping companies, authorities
37. The company's support for ship operations (SMS/DPA)

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

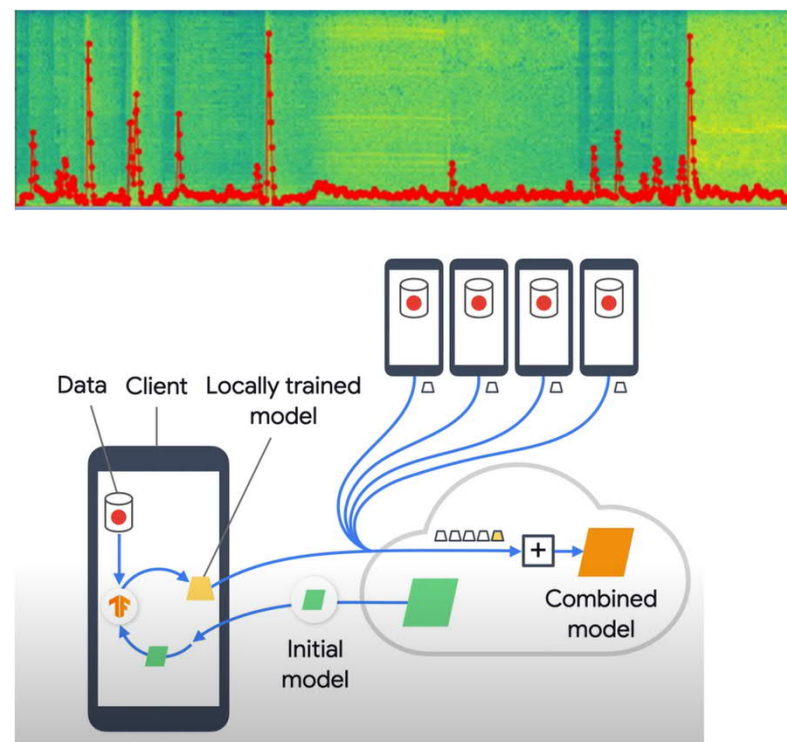
Preliminary findings

- 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

Thank You!

www.ttl.fi/en/aisafety #aisafety

maria.tiikkaja@ttl.fi



ttl.fi



@tyoterveys
@fioh



tyoterveyslaitos



tyoterveys



Tyoterveyslaitos

