The future role of big data and machine learning for health and safety inspection efficiency

Øyvind Dahl

Senior Researcher, SINTEF Digital – Safety and Reliability, Trondheim, Norway

1. Introduction

Inspections are probably the most important policy instrument that governmental labour inspectorates use to ensure that companies take the necessary steps to comply with occupational health and safety regulations. However, the effect that inspections have depend on several different factors. One fundamental factor is the process of selecting inspection objects, i.e. companies to inspect. In principle, there are at least three different selection approaches available. The first approach is to inspect all companies regardless of potential risk, company size, type of industry or any other criteria. The second approach involves selecting enterprises based on random sampling, where every company, regardless of any characteristic has an equal probability of being selected. As regards preventive and economic conditions both these methods are usually seen as ineffective (Blanc, 2013). Thus, most labour inspectorates select objects on basis of a third approach, namely the risk-based approach. In brief, the risk-based approach implies to select inspection objects on basis of risk level.

Although the risk-based approach is an essential principle for most modern labour inspectorates, there are substantial challenges with applying it in practice. The main reason for this is that sufficiently fine-grained methods for risk-analysis are lacking (Mischke et al., 2013). Without appropriate methods to make risk-based prioritization possible, the risk-based approach runs the risk of becoming a governmental policy statement without tangible practical consequences. Hence, there is a need to develop methods which allow for targeting high-risk companies (Weil, 2008).

Most labour inspectorates collect, and store huge amounts of data related to their inspection objects and their inspection activities. Thus, the inspectorates potentially possess large and rapidly growing volumes of data, today referred to by the term 'big data'. Big data, combined with machine learning technology, is at an increasing rate used for different predictive purposes by learning from hidden trends in the data. For example, the predictive value of big data and machine learning techniques are being tested out in such diverse areas as cancer prognosis and patient outcome, bankruptcy prediction, oil price prediction, tax fraud detection, crime prediction and stock market forecasting. The fundamental question in the current paper, however, is whether big data and machine learning technology also could be a promising avenue for labour inspectorates to solve the challenge of targeting high-risk inspection objects?

2. Risk-based targeting

According to the best practice principles for regulatory policy, outlined by OECD (2014), risk analysis and risk assessment should be the basis for targeting inspection objects for labour inspectorates. This means that companies should be selected for inspection on the basis of assessments of the probability and consequence of risk elements such as accidents, harmful exposure, and illegal working conditions. The fundament of risk-based targeting is the recognition that, due to limited inspection resources, it is not possible to control all risk areas and all risk objects. With regard to the labour inspection authorities' health and safety inspections, this means that some problem areas must be prioritized above others. Furthermore, that some companies must be prioritized for inspection and others not.

The principle of risk-based targeting is not a new one. Nearly 50 years ago, in the Robens committee's evaluation of the UK system for supervision of safety and health at work, the risk-based approach (combined with self-regulation) was introduced as an ideal in the process of modernising regulatory inspection (Robens, 1972). To ensure cost-effective use of inspection resources, the Robens Report recommended the regulatory authority to concentrate its resources selectively on the most serious problem areas and to prioritize companies and problems that had been identified through systematic analysis of all available data related to health and safety, e.g. statistics of accidents, technical information and the inspectorates' local knowledge.

The recommendations of the Robens Report have been extensively adopted by labour inspectorates internationally and in the EU member states. The spread of the risk-based approach means that most modern labour inspectorates have embraced the idea of pulling back resources from low-risk objects and to concentrate more enforcement resources on objects with the highest risks. In order to make this possible, some type of data analysis is necessary. Analytical methods for identifying high-risk industries and risk-exposed clusters of workers are usually well developed. Such risk-based analyses are typically grounded on national statistics related to for example occupational diseases, work-related accidents and occupational exposures. The analyses constitute the fundament for inspection campaigns, strategic plans and national, and even international, priority areas.

Far less common than the broader risk-based analyses are methods that make prioritization across companies *within* an industry possible. Among labour inspectorates, a common approach to target concrete risk-exposed companies is to rely on the local knowledge of the inspectors. Other labour inspectorates, such as those of Denmark and Sweden, have explored the usability of risk-ranking systems based on additive scales. By using additive scales, each company is given a risk-score based on several company characteristics (e.g. size, type of industry, registered accidents etc.) that are added to form a sum score, and those with the highest sum are prioritized for inspection. However, the problem with using such additive scales is that they display relatively low predictive validity, i.e. the score is not particularly appropriate for separating high-risk enterprises from the low-risk ones.

3. Big data and machine learning

The process of making prioritizations across companies resembles finding needles in a haystack. In this case, the haystack potentially consists of hundreds of thousands of possible inspection objects, but only a certain amount of these objects are needles, i.e. above a given level of tolerable risk. Finding needles in a haystack is to a large degree what big data and machine learning is all about.

The main objective of machine learning algorithms is to provide a statistical model which can be utilized to perform predictions, classifications, estimations or similar tasks. Within the field of for example cancer prediction, researchers have for more than three decades utilized machine learning algorithms to predict cancer susceptibility, cancer recurrence and cancer survival. Thematically, cancer prediction is pretty far from risk-based targeting of inspection objects. However, both are examples of predictive challenges, or needles in haystack problems.

The two main common types of machine learning algorithms are *supervised learning* and *unsupervised learning*. In supervised learning, the algorithm consists of a dependent variable (e.g. risk level) which is to be predicted from a set of independent variables. Accurate predictions, of course, requires high correlations between the independent variables and the dependent variable. In unsupervised learning, there is no dependent variable to predict, but the objective of the algorithm is to cluster the data into groups (e.g. different risk groups) by similarity. In contrast to additive scales, such as those explored by the labour inspectorates of Denmark and Sweden, the algorithms used in machine learning progressively improve their predictions primarily by trial and error. This means that the machine learns from past success (correct predictions) and errors (wrong predictions) and attempts to capture this knowledge to make more precise predictions based on the feedback received.

4. Utilizing big data and machine learning in selecting inspection objects

Supervised and unsupervised learning algorithms requires a sufficient volume of data, both with regard to the number of observations and the number of variables, usually referred to as 'features'. As already noted, most labour inspectorates collect, and store huge amounts of data related to their inspection objects and their inspection activities. The available data typically consist of company specific features such as number of employees, company age, industrial grouping, number of previous inspections, results of previous inspections, notifications of accidents etc. Furthermore, the amount of data increases day by day as results from new inspections are added. In principle then, the challenge of targeting high-risk companies by utilizing big data should, at least at first glance, be well suited for machine learning algorithms. Despite this, there have been few such attempts. There are, however, some very few notable exceptions which all illustrate that big data and machine learning could be highly relevant for labour inspectorates to solve the challenge of targeting high-risk inspection objects.

A first example is a research study which explores the suitability of machine learning methodologies for the prediction of workplace accidents, or more specifically; floor-level falls

(Matías et al., 2008). Despite its relatively precise predictions, the drawback of this study is that the features included in the algorithms are not the type of data that labour inspectorates normally possess (e.g. use of personal protective equipment and housekeeping). Furthermore, floor-level falls represent only a tiny piece of workplace risks that labour inspectorates are concerned with.

A second example is also a research study (Hajakbari and Minaei-Bidgoli, 2014). This study developed a scoring system for predicting the risk of occupational accidents. Moreover, the study concluded that it is possible to predict the risk of different types of occupational accidents relatively precisely on basis of some general company characteristics (e.g. a company's main activity, gender distribution, number of employees etc.). Furthermore, the study concluded that the algorithm could be utilized to identify workplaces that need periodic health and safety inspections. The data used in this study were retrieved from the database of a labour inspectorate. The drawback of the study, however, is again that workplace accidents represent only one of many workplace risks that labour inspectorates deal with. Furthermore, a particular problem with relying on injury statistics is that such data are known to be highly vulnerable to underreporting.

A third example is a tool developed by the Norwegian Labour Inspection Authority (NLIA) to assist inspectors in selecting enterprises with regard to risk (Dahl et al., 2018). The tool, named The Risk Group Prediction Tool (RGPT), differentiates between four groups of enterprises based on predicted risk. These are *lowest risk*, *low-risk*, *high-risk* and *highest risk*. The higher the risk group of a given company, the higher is the probability that a future inspection in this company will identify serious deviations from the health and safety regulations. The group assignment is made visible to the inspectors via the NLIA's internal web-based user interface. Hence, when targeting companies for inspection, the inspectors are informed about the companies' risk group and are thus allowed to make risk-informed selections.

The RGPT is built on predictive modelling by means of a machine learning algorithm using so-called binary logistic regression analysis. On basis of the regression model, all companies in Norway (roughly 230,000) are assigned to one of the four risk groups. This is done in two steps. In the first step, the regression model predicts the probability that a future inspection will identify serious deviations from the health and safety regulations. In the second step, the model uses the predicted probability value for risk-group assignment.

Initially, the tool was developed on basis of registrations from roughly 35,000 health and safety inspections carried out by the NLIA. However, the predictions made by the tool gradually and automatically become more precise as the number of inspections increases. This means that the algorithm adjusts itself based on the feedback (correct or faulty predictions) it receives when new inspections are carried out and registered in NLIAs database.

The RGPT falls within the class of supervised learning algorithms, where health and safety inspections resulting in serious deviations (dependent variable) are to be predicted from a set of company characteristics (features). The features that the RGPT uses are general company characteristics such as company size, industrial group, number of previous inspections, results from previous inspections, company age, geographical localization, notifications of accidents etc. The predictive validity of the tool is checked every month, and the experience this far (after roughly 18 months of testing) is that the algorithm manages to target companies with a high risk extremely precisely. This means that there are few false positives and few false negatives. I.e. few inspections within the lowest risk group result in identification of serious deviations, whereas the vast majority of inspections within the highest risk group result in identification of serious deviations. The low- and high-risk groups fall between the two extremes.

The tool developed by the Norwegian Labour Inspection Authority demonstrates that it is possible to target inspection objects by means of utilizing big data and machine learning. Similar machine learning approaches have also been tested out among at least two other European labour inspectorates with promising results; the Swedish Work Environment Authority and the Dutch Inspectorate SZW (Ridemar, 2018; Jacobusse and Veenman, 2016). However, the Norwegian tool is not necessarily transferable to other labour inspectorates. This depends on how data is stored, data quality, data access and database structure. Furthermore, targeting companies on basis of the tool involves acceptance of the way risk is defined and operationalized in the algorithm. As described, the tool is based on a definition of risk which implies that the higher the risk group of a given company, the higher is the probability that an inspection in this company will identify serious deviations from the health and safety regulations. This means that the tool primarily is concerned with so-called *management and control risks*, and not *inherent risks*. Whereas management and control risks arise from a company's ability and willingness to manage risk (e.g. by means of complying to the relevant regulations), inherent risks are those that arise from the nature of a business' activities (e.g. fall from heights, chemical exposure, musculoskeletal strain etc.).

In practice, management and control risks and inherent risks are related. This, however, does not imply that the two types of risks necessarily are highly empirically correlated. Hence, to rely blindly on tools which target companies based on the one type of risks might lead you to miss the other, and vice versa. Within the Norwegian regulatory regime, this challenge is solved by emphasizing inherent risks when identifying priority areas, risk-exposed clusters of workers and high-risk industries, whereas management and control risks are emphasized when targeting concrete companies.

5. Challenges

The fact that management and control risks on the one side, and inherent risks on the other are not necessarily empirically correlated, leads us to another, and probably even more severe challenge with applying big data and machine learning algorithms in risk-based targeting. The three examples of machine learning tools above are all examples of one-dimensional targeting. I.e. targeting based on one particular definition and operationalization of risk. Risks in the world of work, however, are not of one particular type. Hence, enforcing authorities are concerned with multiple types of risks, e.g. accidents, chemical exposure, biological exposure, psychosocial threats, musculoskeletal risk factors, social dumping etc. Within these types of risks, there are even more subtypes. Developing risk models that manage to capture this variety is highly challenging, because the different types of risks do not necessarily correlate. Hence, capturing this variety is quite different from predicting the probability of one particular type of risk (Dahl et al., 2018).

A second, but related challenge, makes the task of risk-based targeting even more complex. This is the so-called political pitfall (Black, 2010). Even though machine learning algorithms are dynamic, in the sense that they can learn from and adapt to successes and errors, they can not take consideration of different political point of views. Firstly, the political context is fickle. Thus, what types of risks that are worthy of prioritization today might not be worthy of prioritization tomorrow. Secondly, the political context is multifaceted. Thus, different stakeholders, e.g. politicians, employers, employees, the media and the public, hold different views on which types of risks are worthy of prioritization. This illustrates that risk in the world of work is not necessarily an objective entity, but a social construction.

A third challenge, worthy of consideration, is related to the fact that even though labour inspectorates possess huge amounts of data related to their inspection objects, these data are usually attached to the company level, but the company level is not necessarily the key unit to consider (see e.g. Gunningham and Sinclair, 2007). In a database, a unique company is typically identified by a unique identifier such as an organisation number. In order for a machine learning algorithm to assign a given predicted risk value to a given company it is dependent on unique identifiers. However, all potential inspection objects are not automatically identifiable by a unique identifier. Within the construction industry for example, it is not necessarily a concrete company that is targeted for inspection, but a temporary construction site. There are at least two challenges related to such temporariness. Firstly, construction sites and other temporary locations of work might not be identifiable by unique identifiers. Secondly, even if they were identifiable, the temporariness implies that a machine learning algorithm might not even be given the chance to learn from its predictive successes and errors before the construction site is history and the companies that made up the site have moved into new constellations at a new site.

6. Concluding remarks

The challenges described above illustrate that there are some significant difficulties related to targeting high-risk inspection objects by utilizing big data and machine learning techniques. However, these challenges do not in any way erase the usefulness of such techniques within a risk-based approach. Rather, the challenges illustrate that risk-based targeting will probably not benefit from relying completely on machine learning algorithms. The Norwegian example above illustrate this. Rather than allowing the algorithm to pick and choose objects directly, the inspectors are allowed to make risk-informed decisions on basis of the predictions that the algorithm make. This implies a combination of artificial and human intelligence, where both complement the strengths of each other. When it comes to predictions of complex social events in general, combining the two types of intelligence is probably a necessity.

7. References

Black, J., 2010. Risk-based Regulation: Choices, Practices and Lessons Being Learnt. In OECD, 2010 (Ed.) Risk and Regulatory Policy. Improving the Governance of Risk. OECD, Paris.

Blanc, F., 2013. Inspection reforms: why, how, and with what results. OECD, Paris.

Dahl, Ø., Kilskar, S.S., Skarholt, K., Rosness, R., 2018. Risikobasert tilsyn i de nordiske arbeidstilsynene. [Risk-based labour inspections in the Nordic countries.] Nordic Council of Ministers, Copenhagen.

Gunningham, N., Sinclair, D., 2007. Factors impinging on the effectiveness of the Mines Inspectorate. The Australian National University, Canberra.

Hajakbari, M.S., Minaei-Bidgoli, B., 2014. A new scoring system for assessing the risk of occupational accidents: A case study using data mining techniques with Iran's Ministry of Labor data. Journal of Loss Prevention in the Process Industries 32, pp. 443-453.

Jacobusse, G., Veenman, C., 2016. On selection bias with imbalanced classes. International Conference on Discovery Science, Bari, pp. 325–340.

Matías, J.M., Rivas, T., Martín, J.E., Taboada, J., 2008. A machine learning methodology for the analysis of workplace accidents. International Journal of Computer Mathematics 85, pp. 559-578.

Mischke, C., Verbeek, J.H., Job, J., Morata, T.C., Alvesalo-Kuusi, A., Neuvonen, K., Clarke, S., Pedlow, R.I., 2013. Occupational safety and health enforcement tools for preventing occupational diseases and injuries. Cochrane Database of Systematic Reviews 2013, Issue 8. Art. No.: CD010183.

OECD, 2014. OECD Best practice principles for regulatory policy. Regulatory enforcement and inspections. The Organisation for Economic Co-operation and Development, Paris.

Ridemar, A., 2018. Decision support for SWEA inspections. KTH Royal Institute of Technology, Stockholm.

Robens, L., 1972. Safety and Health at work: Report of the Committee 1970–1972. Her Majesty's Stationery Office, London.

Weil, D., 2008. A strategic approach to labour inspection. International Labour Review (147), s. 349–375.