FINANSTILSYNET

N O T A T

The Danish Financial Supervisory Authority (the Danish FSA)

13 September 2019

Recommendations when using supervised machine learning

1. Introduction

The financial sector is increasingly using new technologies such as supervised machine learning. There is great potential in this, and the Danish FSA expects that the technology will find wide application in the sector. The Danish FSA wants to contribute to positive developments in this field. However, companies should not be uncritical in their use of "new" technologies.

Between August 2018 to March 2019, the Danish FSA and e-nettet specifically discussed how companies can use supervised machine learning to estimate the sales price of residential real estate. The discussions took place in the Danish FSA's regulatory sandbox, FT Lab.

What is machine learning?
Machine learning is a subset of artificial intelligence. Machine learning can be briefly described as algorithms that process and learn from data, then use them to make informed decisions.

Based on experiences the Danish FSA has gained, including from FT Lab, the purpose of this paper is to draw attention to some of the risks brought about by supervised machine learning. In addition, the purpose of the paper is to guide companies and inspire them to make the right decisions in relation to dealing with those risks.

The paper lists a number of issues that companies covered by financial legislation should consider when using supervised machine learning. The paper is a preliminary and non-exhaustive catalogue of what companies in the financial sector should consider on an ongoing basis if they use supervised machine learning.

Thus, the paper must be seen as a supplement to the requirements that companies must otherwise observe if they are covered by financial

regulations, such as specific requirements for corporate use of models. This applies, for example, to the requirements for the use of models by institutions when calculating their own funds.

What is supervised machine learning?
Supervised machine learning is a subset of machine learning, where variables for input and output are known. Based on known variables, the optimal coupling and weighting is derived between all input variables and the output variable. This context can then be used to describe new input examples.

In a number of cases, machine learning can be included in the modelling work of financial institutions. The application of other legislation will depend on the specific activities that the individual company performs using machine learning. Companies must therefore keep in mind that rules for general risk management when using models, such as those laid down by CRR[1], continue to apply. This paper will not go into more detail on these aspects.

The Danish FSA will not take a stand on which tools or models a company should use. The paper does not prescribe specific standards, but lays down recommendations when using supervised machine learning. The requirements will be higher in the case of activities that are significant, either for the business model of a company, for risk management or for consumers, than if the activities are less significant.

Companies should not involve the Danish FSA in all use of supervised machine learning. However, when a company wants to use machine learning to perform a regulated activity, it may require permission or waiver from the Danish FSA in specific cases. In such cases, companies should at least have examined the considerations described in this paper before requesting permission or waiver.

Basically, the Danish FSA's considerations indicate that the use of supervised machine learning by financial institutions does not differ significantly from other statistical analysis or activity in general. In any event, companies must ensure that internal processes support reassuring operations with adequate risk management, reporting, financial reporting and customer service tailored to the activities performed or offered.

2. Background

The Danish FSA and e-nettet have investigated how companies can use supervised machine learning by means of a neural network to estimate the sales price of residential real estate within the framework of FT Lab, the regulatory sandbox. The Danish FSA and e-nettet examined how processes,

[1] Regulation (EU) No 575/2013 of the European Parliament and of the Council of 26 June 2013

development and results can be described, documented and explained when it comes to complex models based on supervised machine learning.

Valuation of residential real estate with supervised machine learning implies that the model is trained on historical data, where both input (such as floor area, distance to schools, location on noise maps) and output, i.e. the actual selling price (the price a property is actually sold at), are known. The model looks for links and weighting between the input variables to find an optimum correlation to best explain the output variable, the actual sales price. When the model has become sufficiently good at matching historical data (i.e. when the predictions are close enough to the sales price), the model can be used to estimate a possible sales price on any residential real estate. In other words, the model can give an idea of what a specific property with specific characteristics (input) can be sold for (output).

Use of other variants of machine learning should likely take into account other specific conditions that are not addressed here. However, the Danish FSA expects that the general principles in the paper can be applied broadly to all use of machine learning.

What are neural networks?
Neural networks denote a number of algorithms which attempt to mimic the processes performed by biological neural networks, such as the human brain. Neural networks "learn" to perform tasks by considering specific examples without being pre-programmed with the criteria for solving the task. In this way, they can be used for such purposes as image recognition, as they can "learn" to identify images. The classic example is a picture of a cat. The network learns to identify these images by analysing training data that are manually labelled as either "cat" or "not cat". Using these results, the network can identify cats on new images. Neural networks do so without prior knowledge of cats, e.g. being aware that they have fur, whiskers and tails. Instead, the network automatically generates identification properties from the learning material it processes.

The paper not only reflects the actual discussions that the Danish FSA has had with e-nettet within the framework of FT Lab, but is a summary of considerations it has so far made, inter alia, in various international contexts.

3. Purpose for using supervised machine learning and description of the model

Machine learning models can be used to solve a wide range of tasks in the financial sector. For example: monitoring suspicious transactions, asset valuation, and automating customer advice. Depending on the intended specific use of the model, certain considerations are necessary. It is therefore crucial that a financial institution that wants to use a machine learning model first defines the purpose of the model.

The purpose assessment is key because all subsequent model choices should be decided in the light of the purpose. Companies will make many choices during a model's lifecycle, and all such choices should be made on the basis of the purpose. The purpose of a given model should therefore be clearly described. The description must, at minimum, state which task a model performs, why the solution is best achieved with machine learning, and which internal or external stakeholders the model should benefit.

Training, use, and updating of a model must support the model's purpose. It is therefore important that the purpose is formulated so that all stakeholders understand it. This requires a specific and clear wording that is largely devoid of technical terms. Such a wording will also ensure that everyone involved, including decision-makers, understands why the specific purpose is best achieved through the use of a specific model.

At the same time, and before the model is put into production, it should be clear to the organisation why a specific task can best be carried out by means of machine learning. This is particularly important if the activity performed is specifically regulated, for example by financial regulation. This can stipulate specific requirements that the company must take into account.

A company can develop machine learning models for a specific purpose or as part of an explorative process, where the company searches available internal data to find undetected links. While there may be value in a purely explorative approach to a company's internal data to find such links, financial institutions should carefully consider and describe the purpose of a model before putting it into operation. This will ensure that a model is not used until it is clear what it can contribute and on what basis this is decided.

Recommendations for using supervised machine learning implies that:

- companies have clearly described the purpose of their model, so that it is able to make decisions on an adequate basis at a later date
- description of the model supports sufficient understanding at all relevant levels of the organisation.

4. Governance (model development, application and updating), policies and business procedures

Financial institutions must already comply with governance requirements. They must, for example develop policies and procedures in a number of areas, in accordance with, inter alia, the Executive Orders on Management and Control. These requirements must ensure that companies are operated properly and handle all relevant risks. Machine learning has the potential to

create new types of risks, which the corporate governance structure should be able to accommodate.

Development and use of machine learning models should basically follow the same procedures and methods as a company's other IT development. Although there may be a need to adapt company procedures and methods to account for the distinctive characteristics of machine learning models, companies should not develop such models without including the general knowledge of IT development that they already possess.

There are several model types and ways to optimise them. The development of models should therefore be documented and validated on an ongoing basis. Logging changes to models should also be done systematically. Companies should thus be able to document choices and options left out during the development, operation, and updating of a given model.

In choosing between different model types and documentation hereof, companies should define criteria for selection. The criteria should include more than just performance. They should also relate to robustness, cf. section 7, and explainability, cf. section 9. The model's limitations, including in the quality of input data, should also be described to ensure that it is clear to the model's stakeholders under what circumstances it should be used.

Such conditions depend on the model's purpose and the relationship that the model describes or explains. Companies should consider which factors can affect the model and how these are best handled. For example, there will be a difference between whether a model must explain a market where the same (relatively few) factors are significant over a long period, such as the housing market, or a market where the importance of factors are fluctuating more, such as the financial markets.

Companies should also consider use of machine learning in ongoing risk management and implement clear guidelines for managerial follow-up. This means that they should ensure an appropriate governance structure to handle the use of the technology. They should have (or implement) adequate measures for IT security and for handling e.g. cyber attacks that attempt to influence the model or input data.

In other words, machine learning should be part of the usual governance structures that already exist in the areas where technology can be used. If new processes arise when a company uses machine learning that it has not considered in its usual governance structure, it will be relevant to revisit and update the existing set-up. This must ensure that the company can also handle the new processes. Companies should therefore consider to what extent machine learning gives rise to adjustments in policies, instructions and business procedures, as well as reporting and control procedures.

Recommendations for using supervised machine learning implies that:

- its use entails the necessary adjustments to the company's usual governance
- companies document and log the choices that were made and the options that were not chosen during the model's life cycle, so that the model's history is assured
- the model's limitations are thoroughly described, including in terms of data quality.

5. Data processing

Machine learning is not a new phenomenon. The technology has been known for many decades. However, in recent years, it has become considerably more widespread as companies have increasingly been given access to the necessary computing power and data volume. In particular, the availability of larger and more complex data sets imposes a number of requirements on financial institutions that use machine learning.

Companies should actively consider the data the model uses in connection with the purpose of using machine learning. They should describe how it ensures data quality and stability in obtaining data. Particularly, if data is obtained from third parties, it is important that they consider how the company can ensure the quality of data and that the company has stable access. The approach to this work should be risk-based, so that the most critical data sources are best protected. These considerations require companies to identify data needs in relation to the specific model.

It may also be beneficial to have a list of alternative data sources in the event of an external supplier being unable to provide the required data at all times. The more important the data is for the model, the more necessary alternative data sources become.

Companies should document the process of handling all data included in the model. This includes how data is processed prior to use. This applies, for example, to whether outliers are removed, variables normalised, or unstructured data are divided into intervals. As part of this work, companies should also describe why data is processed and how data may be annotated.

Recommendations for using supervised machine learning implies that:

- companies define their data needs

- they ensure that data quality and stability in delivery can be maintained at a satisfactory level
- they document their data handling process.

6. Training the model

When designing a model based on machine learning, companies should choose how to optimise the model. Optimisation should directly support the intended purpose. For example, if the purpose of the model is to estimate output on individual observations, training should minimise deviations at individual level. If, on the other hand, the purpose is to estimate outputs grouped in a larger portfolio, training should minimise deviations at group or portfolio level.

Companies should therefore be able to explain the choices they make when optimising the model. Using machine learning is contingent on companies being able to describe the connection between the model's purpose and the approach used for training and optimisation.

Companies should describe any effects from data management in connection to model training. Are there, for example, discarded data that may lead the model to perform better or worse in certain data segments, or does the chosen optimisation mean that specific data points are attributed greater weight than others? In any event, companies should always be able to describe how the model's training best supports the purpose of using machine learning.

A supervised machine learning model often requires very large amounts of data to achieve results that are significantly better than those achieved with more classic statistical models.

The large amount of data should be divided into several subsets to ensure that a model is consistent across data. A model should initially be trained on one data set, *the training set*. Every time the model has been optimised, it should be validated on a data set not included in the training, *the validation set* in this case. When the model achieves sufficient precision on the validation set, it should be tested on a further data set that has not previously been included in the training, *the test set*. This approach minimises the risk of overfitting[2] the model, which may mean that the model is not accurate. This type of risk already exists in classic statistical models, but in machine learning the risk is increased.

Companies should therefore consider how the data used for training, validation, and testing are divided and used in order to best support the

[2] Overfitting means that a model describes a data set so accurately that the model is no longer suitable for explaining further or future data. A model thereby loses its functionality.

model's training and thus its purpose. The division of data should allow the model to describe the reality the company wishes to describe and not just the training set on which the model was developed.

Recommendations for using supervised machine learning implies that:

- companies organise training of their model to ensure optimum support of its purpose, including the choice of optimisation method
- they can describe whether data processing could have influenced the training and optimisation of the model
- data are divided into training, validation, and testing
- the division of data is generally based on the model's purpose.

7. Performance and robustness

In some cases, machine learning models can improve results and increase accuracy compared to more traditional statistical models. Such improvement may involve more accurate results, faster results, or identification of relationships that were not previously known. Using a common term, these improvements of existing models can be referred to as *performance*. 'Performance' covers the model's ability to accurately estimate a desired parameter. It is key to the use of supervised machine learning that a model is performing well, but it is equally important that the model is robust.

In addition to estimating accurately, a model should be robust to changes in data and other external influences. The model should be able to resist malicious attempts to influence its outcome. For example, it could be a user trying to cheat a model – directly or by changing details in input data. Financial institutions should consider the risk of abuse when using machine learning.

Robustness also includes the model's ability to handle changes in the reality the model is trying to describe, i.e. its ability to process updated data. This ability should ensure that the model is sufficiently consistent over time. Output from a model that is sufficiently robust should therefore generally not change significantly from one model version to the next, or if testing with data from different time periods. Companies should therefore ensure that versioning of specific models is documented and archived. This will make it possible to restore previous results in order to investigate why any significant change in the model has occurred.

There may be situations where it is necessary to balance the relationship between performance and robustness. In such a situation, companies should particularly consider how and on what basis the trade-off between performance and robustness occurs. They should also be able to document why they made their decision.

In relation to such balancing, it is important that companies use their domain knowledge, i.e. in-depth knowledge of their products, services, and target groups available in the company. If a model is accurate for a period (high performance) without any apparent explanation, it may be a sign that the model has poor robustness.

Recommendations for using supervised machine learning implies that:

- companies have actively addressed the model's robustness
- they have a proven approach to versioning their models
- they consider the risk of abuse
- they can document any trade-offs between performance and robustness.

8. Accountability

Companies should make decisions based on results from a machine learning model in the same context as other decisions are made. Such a decision will thus generally have to meet the same requirements as a decision made in the traditional way. For a financial institution, this means making decisions at the same management level, whether or not a machine learning model was involved in the process.

Accountability therefore implies that managerial approval of a model and its use is located at the level in the company already responsible for the specific activity to which the model relates. The formal responsibility for machine learning models in a company should always be rooted at a suitably high level of management.

Recommendations for using supervised machine learning implies that:

- companies ensure that the models are approved at a sufficiently high level of management
- responsibility for using machine learning is well defined.

9. Explainability

Machine learning currently involves processes and results that may be more difficult to explain than classic statistical methods. This means that companies should consider which methods can be used to ensure proper use of the technology.

Explainability implies that the company is able to explain and understand why a machine learning model has produced a given result. This can be done, for example, by means of various statistical or mathematical tools. Explainability

is an essential prerequisite for control with the use of a machine learning model and for using it reassuringly.

Companies should continuously assess the risk of inappropriate outcomes and effects in connection with maintenance of the model. For example, the company should include sensitivity analyses of the model's components and assess whether weighting of the main elements makes sense in its procedures for the model's development, maintenance, and use. A model should generally produce results that are intuitive and consistent with economic theory.

Companies should consider how they continuously address their model's results to ensure the results contribute to decision-making in an appropriate manner. Models can have many different designs and applications. Therefore, it is not currently possible to provide an exhaustive list of tools, methods, or the like that will always improve explainability. However, as a minimum, industry-standard tools for explainability of machine learning models prevalent at the time should be used. In addition, companies should use well-known statistical methods, e.g. back-tests of results and sensitivity analyses, to assess the model weights of parameters.

In connection with this, it is also important to consider the model's purpose. The tools a company can use to explain a model's results can vary, depending on the purpose of the model. For example, if it is not possible to assess whether a model emphasises the most logical input, it should be possible to explain why the model can still be used to fulfil the purpose. A high degree of explainability will be especially relevant if machine learning is used for activities that are subject to financial regulation or for activities that have a direct impact on consumers.

A model that is used for making decisions that affect individuals must allow the persons concerned to receive an explanation of the basis of that decision.

Recommendations for using supervised machine learning implies that:

- companies can explain how a model works and what, in particular, is the basis for its results
- explanations should be available to relevant stakeholders, including the company's own management
- tools for explainability are used to the appropriate extent and that companies can, for example, show which components are weighted highest for specific outcomes of the model
- individuals can receive an explanation of the basis of a decision that affects them.

10. Data ethics, bias and fairness

Companies must necessarily consider data ethics as a supplement to their technological design and approach to machine learning[3].

Data ethics is responsible use of data. It is based, inter alia, on the principles of the data protection legislation, which includes the General Data Protection Regulation and the Data Protection Act. Financial legislation does not yet contain explicit requirements for data ethics that go beyond the good business practice rules. Companies should nevertheless make a number of ethical considerations both before and during the process. Incorrect ethical decisions can incur costs for the company and involve significant operational risks, especially to reputation. In addition, there is increasing focus on data ethics nationally and internationally.

When a company processes personal data, the controller must comply with the rules in the General Data Protection Regulation. This also includes the principles for processing personal data listed in Article 5. The controller is responsible for and must be able to document that the principles have been complied with.

An important area in data ethics is problems with bias in data. Bias can have many sources and can give rise to inappropriate outcomes of the model. It is therefore important that companies using machine learning actively consider how they can reduce the effect of bias in the model's design and use.

For example, bias may arise from data containing variables that are considered discriminatory, such as gender or ethnicity. Bias can also occur indirectly through interactions between several variables, which are not in themselves discriminatory.

The latter situation can be difficult to test statistically, and development of the model and evaluation of results should therefore substantially involve experts with domain knowledge of the given topic. Bias should be identified and eliminated to the extent possible.

The development process should involve experts from the technology and business sides of the company to ensure anchoring in the company in general and to ensure that several different types of experts have reviewed the model to identify any inappropriate outcomes.

---

[3] The VLAK government's fact sheet on data ethics for business, 29 January 2019: https://em.dk/media/12932/faktaark_dataetiske-initiativer.pdf, and recommendations from the Expert Group on Data Ethics, November 2018: https://em.dk/media/12191/ekspertgruppens-afrapportering-inkl-anbefalinger.pdf

Companies must also consider fairness in addition to bias. Fairness in this context covers the current understanding in society of what is right and wrong. Fairness is thus a cultural and fluid concept. A model can thus be free of bias from data sources and model development, but nevertheless lead to outcomes which are not considered to be fair to certain customer segments for example. Companies should be able to document how they have addressed fairness.

Recommendations for using supervised machine learning implies that:

- companies have actively considered bias, looked into how the risk of inappropriate outcomes can be minimised and can document this
- they have actively considered fairness and can document this.

11. Transparency

Using machine learning may involve a risk that companies make decisions on a basis that is difficult to verify. This can cause insecurity among anyone subject to such decisions. In addition, ambiguity can hide errors in the model that could have a negative effect on a larger group of recipients than individual human errors may have.

Companies working with machine learning should therefore consider how they inform their stakeholders of this issue. This is becoming increasingly necessary when a model has an impact on decisions that directly or indirectly affect individuals, especially consumers. Companies must note the rules in Article 22 of the General Data Protection Regulation on automatic individual decisions, including profiling[4].

It must be possible for the individual consumer to find out which processes can be expected to form the basis for processing his/her dealings with the company in question.

Recommendations for using supervised machine learning implies that:

- companies make available sufficient information on the use of machine learning to the affected parties
- individuals can gain insight into which processes form the basis for processing their dealings.

[4] Article 22 of Regulation (EU) 2016/679 (General Data Protection Regulation) gives the data subject the right not to be subject to a decision based solely on automated processing, including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her. This does not apply if, for example, it is established in a contract or the data subject consents.

Annex 1

Overall recommendations

Purpose

Recommendations for using supervised machine learning implies that:

- companies have clearly described the purpose of their model, so that it is able to make decisions on an adequate basis at a later date
- description of the model supports sufficient understanding at all relevant levels of the organisation.

Governance

Recommendations for using supervised machine learning implies that:

- its use entails necessary adjustments to the company's usual governance
- companies document and log the choices that were made and the options that were not chosen during the model's life cycle, so that the model's history is assured
- the model's limitations are thoroughly described, including in terms of data quality.

Data processing

Recommendations for using supervised machine learning implies that:

- companies define their data needs
- they ensure that data quality and stability in delivery can be maintained at a satisfactory level
- they document their data handling process.

Training the model

Recommendations for using supervised machine learning implies that:

- companies organise training of their model to ensure optimum support of its purpose, including the choice of optimisation
- they can describe whether data processing could have influenced the training and optimisation of the model
- data are divided into training, validation, and testing
- the division of data is generally based on the model's purpose.

Performance and robustness

Recommendations for using supervised machine learning implies that:

- companies have actively addressed the model's robustness
- they have a proven approach to versioning their models
- they consider the risk of abuse
- they can document any trade-offs between performance and robustness.

Accountability

Recommendations for using supervised machine learning implies that:

- companies ensure that the models are approved at a sufficiently high level of management
- responsibility for using machine learning is well defined.

Explainability

Recommendations for using supervised machine learning implies that:

- companies can explain how a model works and what, in particular, is the basis for its results
- explanations should be available to relevant stakeholders, including the company's own management
- tools for explainability are used to the appropriate extent and that companies can, for example, show which components are weighted highest for specific outcomes of the model
- individuals can receive an explanation of the basis of a decision that affects them.

Data ethics

Recommendations for using supervised machine learning implies that:

- companies have actively considered bias, looked into how the risk of inappropriate outcomes can be minimised and can document this
- they have actively considered fairness and can document this.

Transparency

Recommendations for using supervised machine learning implies that:

- companies make available sufficient information on the use of machine learning to the affected parties
- individuals can gain insight into which processes form the basis for processing their dealings.