DEUTSCHE BUNDESBANK EUROSYSTEM

BaFin
Bundesanstalt für Finanzdienstleistungsaufsicht

Maschinelles Lernen in Risikomodellen – Charakteristika und aufsichtliche Schwerpunkte

Begleitender Fragebogen zur Konsultation des Diskussionspapiers von BaFin und Deutscher Bundesbank

Die BaFin und die Deutsche Bundesbank haben kürzlich das Papier **„ Maschinelles Lernen in Risikomodellen – Charakteristika und aufsichtliche Schwerpunkte“** veröffentlicht. Gerne möchten wir das Papier als Grundlage nutzen, um mit Ihnen in einen Dialog zum Thema Machine learning in Risikomodellen zu treten. Daher laden wir alle Stakeholder ein, unsere Thesen und Erkenntnisse im Rahmen der vorliegenden Konsultation kritisch zu hinterfragen und um das in Ihrem Haus vorhandene Fachwissen anzureichern.

Dazu möchten wir Sie bitten, die untenstehenden Fragen zu beantworten. Die Fragen stehen im Kontext zu den entsprechenden Ausführungen in dem Papier und sollten zum besseren Verständnis nicht isoliert betrachtet werden. Zur Beantwortung dieser Fragen sind die entsprechenden Antwortboxen vorgesehen. Gerne können Sie jeden Themenkomplex auch über die von uns gestellten Leitfragen hinaus um eigene Sichtweisen und Erkenntnisse ergänzen; bitte verwenden Sie dafür jeweils die separate Box. Selbstverständlich ist es auch möglich, sich nur zu ausgewählten Themenkomplexen zu äußern – Sie müssen also nicht jede Box befüllen. Die Fragen dienen der Strukturierung der Konsultation, sollen den angestrebten Dialog aber nicht über Gebühr einengen. Maßgebliche weitere Erkenntnisse und Anregungen mit Kontext zu der Studie, die sich aber Ihrer Meinung nach nicht oder nicht ausreichend in den Fragen widerspiegeln, können Sie daher bei Bedarf unter 4. eingeben.

Es ist beabsichtigt, die eingereichten Stellungnahmen auszuwerten und im Nachgang eine anonymisierte und aggregierte Auswertung im Internet zu veröffentlichen. Auch auf Grundlage dieser Erkenntnisse werden BaFin und Bundesbank entscheiden, ob und welche der Fragestellungen sie weiter verfolgt. Die eingereichten Stellungnahmen werden jedoch nicht einzeln veröffentlicht. Schicken Sie uns bitte das gesamte Word-Dokument bis zum 30.09.2021 als Anlage einer formlosen E-Mail an folgende Adressen: Konsultation-11-21@bafin.de und ai-b3@bundesbank.de.

Die hier aufgeworfenen Fragen sollen eine Diskussion anstoßen. Sie implizieren nicht, dass die übergeordneten Bereiche Finanzstabilität, Markt- und Unternehmensaufsicht oder der kollektive Verbraucherschutz durch den Einsatz von Machine Learning-Technologien in Risikomodellen per se gefährdet wären.

Bitte geben Sie unbedingt zunächst die nachstehenden Informationen an. Stellungnahmen ohne diese Angaben können leider bei der Auswertung nicht berücksichtigt werden.

Institution (Unternehmen, Interessenvertreter/Verband, Aufseher/Regulierer, etc.):

Name: Deutscher Sparkassen- und Giroverband e.V. für Die Deutsche Kreditwirtschaft

Adresse:

Ansprechpartner und Kontaktdetails: Dr. Wiebke Lüke | Wiebke.Lueke@dsgv.de

1. Charakteristika von ML

- Halten Sie den Ansatz für zielführend, auf eine strikte Definition von ML-Methoden zu verzichten, und stattdessen die Aufsichts- und Prüfungspraxis anwendungsorientiert an den einzelnen Charakteristika der eingesetzten Methoden auszurichten?

- Welche weiteren Charakteristika von ML-Methoden können aus Ihrer Sicht für die Aufsichtspraxis oder auch die interne Modelle-Governance von Bedeutung sein?

- Welche Charakteristika gehören aus Ihrer Sicht nicht in diese Übersicht?

- In welchen relevanten Anwendungsbereichen kommen ML-Methoden bei Ihnen zum Einsatz bzw. wo planen Sie deren Umsetzung?

Ihre Antwort auf obenstehende Fragen:

1. Die von der BaFin verwendete Darstellung mit den Schiebereglern erscheint uns zweckdienlich. Diese Darstellung greift auf, dass sich das Umfeld des ML in ständiger Weiterentwicklung befindet. Es ist daher eine breit gefasste Definition wünschenswert, die einer anwendungsorienten Ausrichtung folgt und dabei die Charakterisitika einbezieht. Eine strikte Definition von ML-Methoden halten wir nicht für zielführend.

2. Wir verstehen ML als Methode, die ein mathematisches Modell mit freien Parametern optimal an gegebene Daten anpasst, so dass der zu erwartende Fehler bei der Anwendung auf neue Daten minimiert wird. Dabei sind die zu erklärenden Zusammenhänge nicht kausal, d.h. ein direkter Zusammenhang zwischen Eingangsvariablen und dem Modellergebnis ist nicht explizit darstellbar.
In Bezug nehmend auf „Tabelle 1: Charakteristika von ML-Methoden" ist die Performance des Modells ein weiteres Charakteristikum mit Bedeutung. Eine höhere Performance wird in der Regel durch komplexere Modelle und unter Einbeziehung von zusätzlichen Datenquellen erreicht, was auch für die Aufsichtspraxis von Bedeutung sein kann. Des Weiteren halten wir Plausibilität für ein bedeutendes Charakteristikum.

3. Die im Konsultationspapier genannten Charakteristika sehen wir als wichtige Aspekte im Prozess einer Modellentwicklung/-anwendung an, jedoch nicht als spezifische Charakteristika von ML-Methoden. Die zugrundeliegenden mathematischen Modelle und die Daten sind nicht notwendigerweise komplexer als bei „traditionellen" Verfahren. Auch die Frage der Frequenz ist nicht unbedingt charakteristisch für ML Verfahren, sondern ergibt sich aus dem Anwendungsfall.
Die Charakteristika „Auslagerung" und „IT-Infrastruktur" spielen bei der Auswahl von geeigneten ML-Methoden und deren Aufsicht nur eine untergeordnete Rolle.

4. Bisher sind die tatsächlichen Einsatzgebiete eingeschränkt. Anwendungsfälle lassen sich jedoch im ganzen Bankbetrieb finden Es gibt jedoch Bestrebungen ML-Methoden unterstützend bei der Modellentwicklung einzusetzen (bspw. Variablen Selektion) und Benchmarkingmodelle zu bestehenden Kreditrisikomodellen zu entwickeln. Zur Messung und Vorhersagen können ML-Modelle in nahezu allen klassischen Risikoarten, aber zunehmend auch im Compliance / Surveillance-Bereich (Geldwäsche- / Betrugserkennung) sowie in der Kundenbetreuung (Portfoliomanagement, Vertriebssteuerung) Einsatz finden. Eruiert wird der Einsatz derzeit ebenfalls als unterstützende Methode sowie bei der Bewertung komplexer Derivate (bspw. Kalibrierung hochdimensionaler Volatilitätsoberflächen).

Gibt es zu diesem Themenkomplex etwas über die Fragen hinaus, das Sie uns mitteilen möchten:

2. Aufsichtlicher Ansatz

2.1 Aufsichtspraxis hat Bestand

- Liegen aus Ihrer Sicht bereits aufsichtliche Anforderungen in Regelwerken vor, die für den Einsatz von ML-Methoden hinderlich erscheinen? Ergeben sich aus Ihrer Sicht Widersprüche zwischen prudentiellen Regelungen für Säule 1- und Säule 2-Modelle und dem Entwurf der KI-Verordnung? Bitte geben Sie entsprechende Referenzen auf die entsprechenden Regelwerke an und erläutern Sie die Herausforderungen.

- Inwiefern sind die Anforderungen der EBA/GL/2020/06 mit Bezug auf die automatisierte Kreditwürdigkeitsprüfung und Kreditvergabe auch für andere ML-Methoden in Säule 2 (MaRisk) aus Ihrer Sicht passend und sollten übernommen werden?

- Sehen Sie weitere Punkte, bei denen aus Ihrer Sicht eine Anpassung der bisherigen Aufsichtspraxis erforderlich ist, um ML-Verfahren und die damit verbundenen Risiken angemessen zu würdigen?

- Gehen mit ML-Methoden spezifische Risiken für die IT-Implementierung und das Auslagerungsmanagement einher? Sind z. B. sog. „Adversarial Attacks" im Finanzbereich denkbar und sollten ML-Methoden besonders dagegen geschützt werden?

Ihre Antwort auf obenstehende Fragen:

1.
Die derzeitige Prüfungs- und Anmeldepraxis fordert ein Maß an Nachvollziehbarkeit und Messbarkeit, die nicht geboten scheint.
Ergänzend möchten wir darauf hinweisen, dass wir die in der BaFin Unterlage „Auslegungs- und Anwendungshinweise gemäß §51 Abs. 8 GwG" formulierten Anforderungen bezüglich der Beschreibung der Auswahl und Beschaffenheit von Datenverarbeitungssystemen goutieren (Kapitel 6.2.1). Eine diesbezügliche Verwendung im Kontext der Fragen dieser Konsultation erscheint uns zweckdienlich.

2.
ML-Methoden werden nicht zur Aushebelung der menschlichen Kontrolle eingesetzt. Die regulatorischen Anforderungen werden auch in Zukunft erfüllt. Dadurch ändert der Einsatz von ML nichts. EbA-Konformität ist nach unserer Auffassung nicht betroffen.
Durch alternative ML-Methoden sollten sich die Anforderungen an die Datengrundlage (bspw. Vollständigkeit, Korrektheit, Konsistenz, etc.) nicht ändern. Daher können die Anforderungen übernommen werden.
Die dezidierten Regelungen bzgl. Säule 1-Modelle, wie sie im EGIM niedergelegt sind (sowohl Entwicklung, Monitoring als auch Validierung) könnten die in Frage kommenden Methoden u.U. einschränken, da ggf. nicht in jeder ML-Methode alle geforderten Güteparameter ableitbar sind.
Grundsätzlich sollten ML-Verfahren analog zu etablierten statistischen Modellen behandelt werden, so dass im Allgemeinen keine Anpassung der Aufsichtspraxis notwendig erscheint. Einzelne Aspekte (bspw. Adaptivität) scheinen jedoch die Variabilität von Modellen zu erweitern, so dass sich hier eine Prüfungspraxis noch etablieren muss.

3.
Nein.

4. Für IT-Implemetierungen sehen wir keine ML-spezifischen Risiken. Die bisherigen Anforderungen gelten bereits für die klassisch verwendeten ML-Modelle, es besteht daher kein Anpassungsbedarf. ML ist kein neues Phänomen. Allerdings hat sich die Dynamik der genutzten Datengrundlage stark verändert, daher ist es zwingend nötig, etwa den Prozess zur Abnahme von Modellen zu beschleunigen.

Adversarial Attacks auf alle ML-Systeme sind grundsätzlich möglich. Die Schutzbedürftigkeit ergib sich aus der Sensitivität der im Modell repräsentierten Daten bzw. der Relevanz der vom ML-System unterstützten Prozesses. Zudem können ML-Anwendungen sinvoll in der Cloud betrieben werden. Die Wahrscheinlichkeit, dass Banken Auslagerung nutzen, steigt. Gleichwohl können Banken ihre eigenen Anwendungen auf einer

ausgelagerten Cloud-Infrastruktur betreiben. Entscheiden ist hier aber nicht die Lokation der IT-Implementierung, sondern die Modell- /Methodenhoheit. Diese muss bei den Banken verbleiben. Ein Betrieb der ML kann in einer fremden Infrakstruktur erfolgen, ohne dass daraus ML-Risiken entstehen. Wir möchten an dieser Steller auf das Positionspapier der European Cloud User Coalition verweisen, welches sich ausführlich den Aspekten der Cloud widmet (https://ecuc.group/papers/ecuc_position_paper_may_2021.pdf).

Gibt es zu diesem Themenkomplex etwas über die Fragen hinaus, das Sie uns mitteilen möchten:

2.2 Methoden laden zur Datengläubigkeit ein

- Welche Herausforderungen sehen Sie bei der Auswahl der Daten und bei der Sicherstellung der Datenqualität von ML-Methoden?
- Welche Aspekte der Datenqualität werden durch die Anwendung von ML-Methoden aus Ihrer Sicht erleichtert?

Ihre Antwort auf obenstehende Fragen:
1. Die Anforderungen an die Datenqualität sind nicht spezifisch für bestimmte Methoden und wurden auch bisher schon in den Modellentwicklungsprozessen adressiert. Die Expertise und Erfahrung der Datenmodellierer gerade auch bei der Auswahl der Daten und der Datenqualität insbesondere für komplexe Modelle entscheidend. Eine Vereinfachung der Datenauswahl und -qualität durch ML-Methoden sehen wir nicht. Durch alternative ML-Methoden sollten sich die Anforderungen an die Datengrundlage (bspw. Vollständigkeit, Korrektheit, Konsistenz, etc.) nicht ändern.

2. Wir sehen einen Vorteil, wenn i.R. von KI/ML viele Daten (so auch Metadaten) herangezogen werden, da durch die Robustheit von ML-Modellen gestärkt werden kann.

ML-Methoden können dazu verwendet werden, um fehlende Datensätze intelligent zu ersetzen und damit die Modellqualität zu steigern. Auch kann es dazu verwendet werden, Anomalien (z.B. systematische Verzerrungen in den Daten) zu erkennen, welche durch die standardisierten DQ Checks nicht erkannt werden.

Gibt es zu diesem Themenkomplex etwas über die Fragen hinaus, das Sie uns mitteilen möchten:

2.3 Erklärbarkeit rückt in den Fokus

- Welchen Einfluss hat die „Blackbox"-Eigenschaft Ihres Erachtens auf die Validierung der Verfahren?
- Welche Bedeutung messen Sie dem Trade-off zwischen Performance und Erklärbarkeit bei?
- Bieten XAI-Techniken aus Ihrer Sicht (immer) einen Ausweg aus der „Blackbox"? Welche Verfahren haben sich als vielversprechend herausgestellt und bei welchen ML-Methoden?

- Wie sollte aus Ihrer Sicht eine der Methode nachgelagerte XAI in die Validierung einbezogen werden?

Ihre Antwort auf obenstehende Fragen:

1. Die für die Modellentscheidung verantwortliche Beziehung lässt sich i.d.R. durch einen mathematischen Prozeß beschreiben. Diese kann jedoch – bedingt durch die Komplexität der zugrundliegenden Daten und Zusammenhänge – sehr umfangreich sein. Das „Blackbox-Dogma" ist insofern nicht ganz korrekt. Ein besseres Bild ist unseres Erachtens ein komplexes Uhrwerk, bei dem nur ersichtlich ist, welches Rädchen mit welchem in Berührung ist, aber der ganze Mechanismus in seinem Zusammenspiel nicht ohne weiteres verstanden werden kann. Ähnlich kann ein ML-Modell lokal auch immer durch ein einfaches lineares Modell approximiert werden. Die Validierung von „Blackbox" Eigenschaft kann durch Nutzung von XAI-Ansätzen erleichtert werden.

2. Je nach Anwendungsfall kann eher die Erklärbarkeit oder eher die Performance des Modells als wichtig erachtet werden. Die eigentliche Herausforderung liegt darin, die Komplexität in das ML-Verfahren an die Komplexität des Problems anzupassen. Einen Trade-off zwischen Erklärbarkeit und Performance sehen wir nicht. Entscheidend ist, ob ein Ergebnis unter Verwendung von ML plausibel ist. Erklärbarkeit sollte nicht mit kausal gleichgesetzt werden. Erklärbarkeit wird beispielsweise durch ein detailliertes Verständnis primärer Modelltreiber und deren Sensivitäten erreicht. Erklärbarkeit kann demnach auch dazu dienen, das Modell zu verbessern. Jedoch bedeutet eine leichte Erklärbarkeit nicht zwangsläufig, dass das Model per se besser ist.

3. Eine Blackbox beschreibt ein komplexes Modell, welches nicht mehr vollumgreifend von Menschen nachvollzogen werden kann. Alle XAI Methoden funktionieren nach dem Prinzip, dass ein vereinfachtes Modell auf einem komplexen Modell trainiert wird, um dessen Komplexität zu reduzieren und für den Menschen greifbar zu machen. Sie sind nicht geeignet, hinreichende „Erklärungen" zu liefern, weil die Modelle in der Regel nicht-kausal sind. Wohl aber sind sie gut geeignet, im Sinne von notwendigen Bedingungen und Transparenz, Plausibilisierungsansätze zu liefern und ein Modell somit zu falsifizieren. Diesbezüglich bieten XAI Techniken einen Ausweg aus der Blackbox. Man muss allerdings entscheiden, wie viel verbleibende Komplexität man in dem Modell akzeptiert. Dabei handelt es sich um ein Trade-Off. Feature-Importances oder Shapley Values liefern wertvolle Hinweise und sollten auch in der Anwendung selbst den Nutzern zur Verfügung gestellt werden. In der tiefergehenden Analyse können sie etwa geclustert werden und so zum Auffinden von Erklärungsmustern und approximativen Kausalzusammenhängen genutzt werden.

4. XAI Methoden können zur Plausibilisierung der Validierungsergebnisse herangezogen werden, um folgenden Fragestellungen nachzugehen – sind die in dem Modell enthaltenen Variablen aussagekräftig oder wie ist deren Einfluss auf das Modellergebnis? Grundsätzlich muss das Ziel sein, so weit wie möglich aus dem statistischen Modell und seine XAI-Komponenten zu einem quasi-kausalen Verständnis des zugrundeliegenden Prozesses zu kommen und dieses Wissen dann direkt für die Modellierung zu nutzen. Auch jüngere Verfahren zum Test auf Kausalität (siehe z.B.: https://medium.com/syncedreview/yoshua-bengio-team-proposes-causal-learning-to-solve-the-ml-model-generalization-problem-762c31b51e04) liefern hier wichtige Erkenntnisse und sollten bei der Prüfung von ML-Modellen betrachtet werden.

Gibt es zu diesem Themenkomplex etwas über die Fragen hinaus, das Sie uns mitteilen möchten:

Gemeinsames Verständnis von Aufsicht und Kreditwirtschaft, ab wann anzeigepflichtige Modelländerungen vorliegen, um schneller in den Genehmigungsprozess einsteigen zu können.

Grundsätzlich ist bei Entwicklung, Produktion und Validierung und relativ unabhängig vom Einsatzzweck ganz besonders auf die Erklärbarkeit der Risikokennzahlen anhand des ML-Modells zu achten. Hier sind in den letzten Jahren vermehrt Entwicklungen aufgetreten, die der notwendigen Erklärbarkeit Rechnung tragen, so dass reine „Blackbox"-Verfahren eher die Minderheit bis sehr selten sein sollten.

Wie bereits bei klassischen Methoden sind die Anforderungen des Modellrisikomanagements zu übertragen: So sollten Validierungshandlungen zwar äußerst elaboriert sein, gleichwohl sind die gleichen Ansprüche bereits bei der Modellkonstruktion und –entwicklung anzusetzen.

Hinsichtlich des Trade-Offs zwischen Performanz und Erklärbarkeit kommt der für den jeweiligen Anwendungsfall vernünftigerweise modellierbaren Anzahl an Parametern eine besondere Bedeutung zu. Diese kann einem geeigneten Modellmonitoring unterzogen werden.

2.4 Adaptivität: Modelländerungen sind schwerer zu erkennen

- Welche Fragen ergeben sich aus Ihrer Sicht zur Aufsichtspraxis in Bezug auf Modellanpassungen bei ML-Methoden?

- Sehen Sie für bestimmte ML-Methoden die Notwendigkeit sehr häufiger Retrainings?

- Werden ML-Methoden Ihres Erachtens eine Anpassung der Modell-Governance notwendig machen? Wie arbeiten klassische Modellierungseinheiten, Validierer und neue „Data Science"-Einheiten zusammen?

Ihre Antwort auf obenstehende Fragen:
1. Modellanpassungen sollten auf den zentralen Rechenprozess des Modells beschränkt werden. Änderungen an der IT-Architektur und an den nicht-ML-Funktionen sollten nicht als Modellanpassung eingestuft werden. Wie oft und unter welchen Bedingungen möchte die Aufsicht prüfen? Wer muss eingebunden werden? Wie oft darf das Modell geändert werden? Wann muss ein Modell geändert werden? Muss ein Institut aktiv werden, wenn ein Modell nicht geändert wird?

Durch Methoden wie Cross-Validierung oder Out-of-sample Testing kann die Wahrscheinlichkeit, nicht das globale Optimum zu erreichen, reduziert werden. In den meisten Verfahren kann die statistische Komponente durch das Setzen einer „Seed" fixiert werden, um eine Reproduzierbarkeit der Ergebnisse zu ermöglichen. Es ist daher wichtig, ein On-going Monitoring von (ML-)Modellen sicherzustellen. Es ist hierzu aber ausreichend, sich am bestehenden Rahmen der Regulierung und Modellgenehmigung zu bewegen; einer Erweiterung bedarf es nicht. Spezifische Aspekte werden die Bank über ihre individuelle Governance sicherstellen – die bestehen Anforderungen an die Governance sind ausreichen und stellen dies sicher.

2. Ein Großteil der ML-Methoden überlässt dem Anwender die Entscheidung ob und in welcher Frequenz das Modell an neue Daten angepasst werden soll. Auch heute finden bereits regelmäßige Rekalibrierungen der Modelle statt, sodass die Abgrenzung zwischen Modellpflege und aufsichtlich zu beruteilender Modelländerung bereits heute relevant ist. Die Notwendigkeit häufiger Retrainings ist sicherlich stark abhängig von dem jeweiligen Anwendungsfall. Letztendlich sollte die Notwendigkeit nicht häufiger gegeben sein als bei bestehenden Modellen. Zudem sollten Modelländerungen für ML-Methoden nicht auf einzelne Parameter, Meta-Parameter oder den Output des Modells abstellen, sondern auf das Lernverfahren in seiner Gesamtheit, einschließlich aller Meta-Parameter. D.h. die Optimierung über verschiedene Modellansätze sollte als Teil des Lernverfahrens angesehen werden. Nur wenn dieses Lernverfahren geändert wird, sollte dies als Modelländerung klassifiziert werden. Voraussetzung dafür ist, dass das Lernverfahren inklusive aller Meta-Parameter umfassend validiert wurde und die Leistung des Modells ständig überwacht wird. Bei guten Modellen ist davon auszugehen, dass Parameteranpassungen inhärent durchgeführt werden, indem Daten der jüngsten Vergangenheit direkt als Feature benutzt werden. Ein Modell, das häufig rekalibriert werden muss, ohne dass eine strukturelle Veränderung plausibel wäre, sollte dagegen auf temporäres Overfitting überprüft werden. Sicher ist es hier in der Regel nicht einfach zu erkennen, wo die Grenze zwischen echter Dynamik und temporärem Overfitting verläuft; dies sollte ein zentrales Thema der Modellvalidierung sein.

3. Die Modell-Governance kann unverändert bleiben. An der bewährten Arbeitsteilung zwischen Modellierung und Validierung sollte festgehalten werden. Eine Notwendigkeit für Anpassungen sehen wir nicht. Jedoch müssen klassische Modellierungseinheiten nicht getrennt von neuen „Data Science" Einheiten arbeiten. Allgemein muss ein gemeinsames Verständnis für die neuen Methoden auf Entwicklungs- und Validierungsseite entwickelt werden.

Gibt es zu diesem Themenkomplex etwas über die Fragen hinaus, das Sie uns mitteilen möchten:

3. **Zum Schluss möchten wir Sie einladen, uns Ihre Gedanken, Erfahrungen und Lösungsansätze zu BDAI-Themen mit aufsichtlicher und regulatorischer Relevanz zu nennen, die von den obigen Fragen nicht hinreichend abgedeckt werden.**

Ihre Antwort auf obenstehende Fragen:

Einige Banken haben bereits umfangreiche Ausarbeitungen zur Umsetzung der Ethikrichtlinien zur vertrauenswürdigen KI der Europäischen Kommission formuliert, in denen konkrete Implementierungsvorschläge der dort angeführten Anforderungen ausgearbeitet sind. Diese decken auch die hier angesprochenen Fragen zum großen Teil ab. Eine finale abgestimmte Standardsetzung für die Umsetzung in interne Prozesse ist allerdings per heute noch nicht überall realisiert. Wir sind aber gern bereit und sehr daran interessiert, unsere Ideen, Erkenntnisse und praktischen Erfahrungen in den Konsultationsprozess einzubringen sowie im Rahmen eines direkten Austausches vorzustellen und zu erörtern.