DEUTSCHE BUNDESBANK
EUROSYSTEM

BaFin
Bundesanstalt für Finanzdienstleistungsaufsicht

Maschinelles Lernen in Risikomodellen – Charakteristika und aufsichtliche Schwerpunkte

Antworten auf das Konsultationspapier

Inhalt

I. Rückmeldungen aus Versicherungs- und Bankenbereich

BaFin und Bundesbank haben im Juli 2021 das Konsultationspapier „Maschinelles Lernen in Risikomodellen – Charakteristika und aufsichtliche Schwerpunkte" veröffentlicht und Rückmeldungen aus dem Versicherungs- und Bankenbereich eingesammelt.

Der Fokus der Konsultation lag auf der Solvenzaufsicht, und hier insbesondere auf der Anwendung von ML-Methoden an Stellen, die aufsichtlich besonders relevant sind. Dazu zählen einerseits – als Ausnahme von dem Prinzip, dass Algorithmen keiner Genehmigungspflicht unterliegen – ML-Methoden, die Gegenstand aufsichtlicher Prüfungen und Erlaubnisverfahren sind, also in internen Modellen zur Berechnung der regulatorischen Eigenmittelanforderungen (Säule 1). Darüber hinaus sind ML-Methoden relevant, wenn sie im Risikomanagement in Säule 2 zum Einsatz kommen.

ML bisher kaum bei Modellen in Säule 1 im Einsatz

Auf die Konsultation sind Rückmeldungen von Verbänden, u. a. der Spitzenverbände der Banken- und Versicherungswirtschaft, sowie Rückmeldungen einzelner Versicherungen und Banken sowie Unternehmensberatungen und Technologieunternehmen eingegangen.

Es hat sich gezeigt, dass sich auf Techniken des Machine Learning beruhende Verfahren aktuell schon in vielen Anwendungsbereichen antreffen lassen. Im Fokus der aktuellen Anwendungen stehen laut den Rückmeldungen das Risikomanagement (bspw. Geldwäsche- und Betrugserkennung, Analysen im Kreditprozess), Vertriebs- und Anlageprodukte, die Bepreisung von Produkten und einzelne Anwendungen in anderen Risikobereichen. ML-Methoden für Risikomodelle der Säule 1 finden sich bisher nur vereinzelt; sie werden in manchen Rückmeldungen jedoch als vielversprechend eingeschätzt. ML-Methoden werden auch als Unterstützungs- oder Challengertool herangezogen.

Die folgenden Kapitel fassen die Antworten auf die Konsultation zusammen.

II. Charakteristika von ML

Überwiegende Zustimmung zum Charakteristikaansatz

Das Konsultationspapier hat keine allgemeingültige Abgrenzung von ML-Methoden entwickelt, sondern Aufsichtspraxis, Prüfungstechnik und -intensität darauf ausgerichtet, ob und welche ML-Charakteristika bei einer konkret zu prüfenden Methodik vorliegen und wie stark diese ausgeprägt sind. Für die Beurteilung von Modellen unter Säule 1 und 2 stellt dies aus Sicht von BaFin und Bundesbank eine für aufsichtliche Zwecke hinreichende Differenzierung dar.

Die Charakteristika wurden anhand der drei Dimensionen des **AI/ML-Szenarios** gruppiert: Zusammen beschreiben die **Methodik und Datengrundlage** die Komplexität und damit das mit dem ML-Verfahren verbundene Modellrisiko. Die **Nutzung des Outputs** beinhaltet, welchen Stellenwert das Verfahren innerhalb des Risikomanagements einnimmt. Die Prüfungsintensität richtet sich dabei nicht nach der Unterscheidung zwischen Eigenentwicklung und **Auslagerung** oder der zugrundeliegenden **IT-Infrastruktur.**

Die Antworten auf die Konsultation bekräftigen den vorgestellten Ansatz. Es herrscht breite Übereinstimmung darin, keine explizite Definition vorzugeben: Zum einen werde eine Definition der Vielfalt der Verfahren und der stetigen Weiterentwicklung nicht gerecht. Zum anderen ließen sich ML-Verfahren und klassische Verfahren nicht eindeutig abgrenzen. Ebenfalls wird befürwortet, dass sich das Aufsichtshandeln technologieneutral nach den eingesetzten Verfahren richtet. Im Mittelpunkt sollten nicht pauschale Anforderungen stehen, sondern konkrete Anwendungsfälle. Das AI/ML-Szenario entspricht diesem technologieneutralen und einzelfallbezogenen Blickwinkel.

Vereinzelt werden weitere mögliche Charakteristika genannt. Dazu zählen die Auswirkungen von Modellschwächen und die Performance als Charakteristikum von ML. Einzelne Rückmeldungen führen Charakteristika an, die für sich genommen nicht als ML-spezifisch eingeschätzt werden. Dazu zählt die Datengrundlage, die eher durch die Verfügbarkeit der Daten getrieben werde als von der Methode. Genannt wurden auch die Automatisierung, Adaptivität und Komplexität, die mitunter auch bei klassischen Verfahren höher sein können als bei ML.

Auslagerung und IT-Infrastruktur sollten nach Ansicht der Konsultationsteilnehmer nicht als ML-spezifische Charakteristika gesehen werden, weil sie für die Auswahl und Bewertung des ML-Verfahrens grundsätzlich nicht herangezogen würden und einem eigenen, von ML unabhängigen, Aufsichtsrahmen unterlägen. Die Rückmeldungen haben gezeigt, dass keine spezifischen Risiken von ML für die IT-Implementierung und das Auslagerungsmanagement identifiziert werden können, auch wenn die grundsätzliche Bedeutung von Cloud und Auslagerung mit ML steige.

III. Aufsichtlicher Ansatz

1. Aufsichtspraxis hat Bestand

Das Konsultationspapier führte aus, dass in Säule 1 umfangreiche Regeln zur Überprüfung und Genehmigung interner Modelle vorliegen, die technologieneutral formuliert sind und somit auch die Risiken von ML-Methoden adressieren. In Säule 2 liefern prinzipienorientierte Anforderungen an das Risikomanagement und die IT eine solide Grundlage.

Die Rückmeldungen auf die Konsultation stimmen weitgehend darin überein, dass die bisherige technologieneutrale und risikoorientierte Regulierung auch für ML einen ausreichenden Rahmen bildet. Gerade bei der Säule 1 bestehe bereits ein umfangreiches Regelwerk, das sich ebenfalls für ML heranziehen lasse. Teilweise werden die Anforderungen in Bezug auf die Nachvollziehbarkeit und Messbarkeit von ML-Verfahren kritisch gesehen.

Das aufsichtliche Regelwerk wird als ausreichend eingeschätzt

Rückmeldungen aus dem Bankenbereich bewerten die Anforderungen der EBA/GL/2020/06 in Bezug auf ML als angemessen, technologieneutral und risikoorientiert. Dementsprechend werden auch keine Bedenken gegen eine Umsetzung und Anwendung der Regelungen für alle ML-Methoden in der Säule 2 vorgebracht.

2. Methoden laden zur Datengläubigkeit ein

Die Datenqualität ist gemäß dem Konsultationspapier bereits ein Schwerpunkt aufsichtlichen Handelns. Die Charakteristika von ML-Methoden machen allerdings deutlich, dass die Datengrundlage hier insbesondere als Ausgangspunkt und Erfolgsfaktor anzusehen ist. Unstrukturierte Daten können mittlerweile durch bzw. für ML-Methoden erschlossen werden. Ferner erlauben ML-Methoden Berechnungen unter Einbeziehung zahlreicher Einflussgrößen, was wiederum zum Problem des „Overfittings" führen kann. Es wird dem Modellierer einfach gemacht, ML zügig auf eine große Datengrundlage anzuwenden. Bei der Verwendung großer Datenmengen muss deren Qualität fortlaufend sichergestellt werden. Dies betrifft neben der Modellentwicklung und -validierung auch die Modellanwendung.

Die Antworten auf die Konsultation beschreiben einhellig, dass alle verwendeten Daten so repräsentativ wie möglich sein müssen und der Einsatz von ML-Methoden daher keine speziellen und neuartigen Anforderungen aufwerfe. Die Rückmeldungen weisen auf den hohen manuellen Aufwand für die Auswahl geeigneter Daten und die erforderliche Expertise hin: Ein tiefes Verständnis für die verwendeten Daten nehme an Bedeutung zu. Vor allem Versicherungsunternehmen weisen darauf hin, dass regulatorische Vorgaben die nutzbare Datenmenge verringerten z. B. durch die geforderte Konzentration auf nur wenige singuläre Ereignisse, oder dass Trends die Vergangenheitsdaten ungeeignet für Prognosen machten. Datengläubigkeit sei schon jetzt ein Thema, und eine hohe Erklärbarkeit der Modelle bzw. eine Vielfalt an möglichen Modellen könne dem entgegenwirken.

Die Rückmeldungen besagen, dass Datenqualität aktuell schon von entscheidender Bedeutung und vielschichtig zu betrachten sei. Es wird angeführt, dass Softwaretools (u.a. ML-basiert) helfen könnten, nicht nur die Qualität bereits genutzter Datengrundlagen zu sichern, sondern auch neuartige Datenquellen wie etwa unstrukturierte Daten zu erschließen.

3. Erklärbarkeit rückt in den Fokus

Eine der Thesen des Konsultationspapiers betrifft die Erklärbarkeit von Modellen. Je komplexer und höherdimensional der vom Modell abbildbare Hypothesenraum ist, desto schwieriger wird es, den funktionalen Zusammenhang zwischen Input und Output verbal oder durch mathematische Formeln zu beschreiben, und desto weniger sind die Berechnungen durch Modellierer, Anwender, Validierer und Aufseher im Detail nachvollziehbar. Dies führt ggf. auch zu einer erschwerten Überprüfung der Validität der Modellergebnisse. Zusätzlich kann die Akzeptanz bei den Anwendern leiden. Während diese „Blackbox"-Eigenschaft zum Beispiel durch eine höhere Vorhersagegüte gerechtfertigt werden kann, entsteht ein potenziell höheres Modellrisiko. Um diesem angemessen Rechnung zu tragen, wurden Techniken der Explainable AI (XAI) entwickelt. Auch wenn XAI-Techniken aus Sicht der Aufsicht vielversprechend erscheinen, um die „Blackbox"-Eigenschaft abzumildern, stellen diese Ansätze selbst Modelle mit Annahmen und Schwächen dar und befinden sich vielfach noch in der Erprobungsphase.

Aus Sicht der Konsultationsrückmeldungen muss die Reproduzierbarkeit von Berechnungsergebnissen gewährleistet sein, wobei sich die Interpretierbarkeit von Modellergebnissen mit technischen Mitteln verbessern ließe. Die Erklärbarkeit sollte die Auswahl des Modellierungsansatzes treiben. Eine wichtige Aufgabe der Modellvalidierung ist gemäß den Konsultationsantworten die Plausibilisierung der Modelle, wobei die Komplexität der Modelle auch die Komplexität der Validierung bestimme. Als unkritisch werden Modelle gesehen, deren Ergebnisse zeitlich sehr stabil sind und bei denen Sprünge in den Ergebnissen mit eindeutig identifizierbaren Sondereffekten erklärt werden können. Auch bei komplexen Modellen wird die Validierung mit geeigneten Datensätzen und Methoden als möglich betrachtet. Überwiegend wird die Meinung vertreten, dass nicht jeder Zwischenschritt erklärbar sein müsse, sondern nur das Gesamtergebnis.

Erklärbarkeit ist ein zentrales Kriterium in der Anwendung von AI/ML

Die meisten Antworten fordern, dass der Trade-off zwischen Performance und Erklärbarkeit je nach Anwendungsfall zu untersuchen sei. XAI verlange einen so signifikanten Aufwand in der Umsetzung, dass die Performance des Modells außerordentlich hoch sein müsse, um diese zu rechtfertigen. Die Antworten schwanken zwischen „XAI bietet grundsätzlich keinen Ausweg aus der Blackbox-Problematik" und „XAI behebt die Blackbox". XAI wird nicht als Allheilmittel gesehen, da es lediglich ausgewählte Modelleigenschaften untersuche, der Nutzen als Basis für eine Expertenbewertung wird aber anerkannt. Aufsichtliche Standards für XAI können nach Meinung der Teilnehmer noch nicht festgelegt werden, da sich die Methode in rasanter Entwicklung befinde.

Nutzen und Aufwand von XAI müssen abgeglichen werden

Der Einsatz einer nachgelagerten XAI könne nur stichprobenartig erfolgen. Daher sollte XAI ein Teil der Validierung vor Inbetriebnahme des Modells und im laufenden Betrieb sein. XAI könne auch dazu dienen, die wesentlichen Variablen zu erkennen und damit ein nahezu kausales Modell zu konstruieren.

4. Adaptivität: Modelländerungen sind schwerer zu erkennen

Institute und Unternehmen sind dazu verpflichtet, der Aufsicht Änderungen von Säule 1 Modellen zu melden und diese ggf. erst nach Genehmigung in Betrieb zu nehmen. Die Grenze zwischen regelmäßiger Modellpflege und Modelländerung ist fließend und führt immer wieder zu Diskussionen mit der Aufsicht, zumal der Begriff der „Modelländerung" auch vom jeweiligen aufsichtlichen Kontext abhängig ist. Das Konsultationspapier hat hier mehrere Beispiele gegeben. Die Flexibilität und die teils hochfrequente Adaptivität von ML-Verfahren erschweren jedoch eine aufsichtlich unverzichtbare klare Unterscheidung zwischen Anpassungen und Änderungen. Ebenfalls sollte die Notwendigkeit einer hochfrequenten Adaptivität grundsätzlich gut begründet werden. Auch wenn viele ML-Methoden in den nicht genehmigungspflichtigen Aufsichtsbereich der Säule 2 fallen und sich somit eine größere Flexibilität für Retrainings und Änderungen des Modells ergibt, bleiben bereits bestehende Anforderungen zum Beispiel aus der MaRisk gültig. Aus Sicht der Aufsicht ist es dennoch entscheidend, trotz dieser Flexibilität den Trainingszyklus an den Anwendungsfall anzupassen und entsprechend zu begründen, um ein Gleichgewicht zwischen Aktualität hinsichtlich der Daten und Erklär- sowie Validierbarkeit zu schaffen.

In den Konsultationsantworten wird das Spannungsfeld zwischen Modellpflege/Kalibrierung und Modelländerung beleuchtet und auf das Dauerthema des Modelldrift hingewiesen. Die Aufsicht wird gebeten, den Begriff der Modelländerung weiter zu konkretisieren und Genehmigungsverfahren für Modelle generell zu beschleunigen, um Wettbewerbsnachteile zu verhindern. Mögliche Definitionen und Beispiele für Modellpflege und -änderung werden gegeben: So sollte das Training eines Verfahrens zum Beispiel keine Modelländerung sein (auch nicht die dabei vorgenommenen Anpassungen von Hyperparametern), jedoch könne der iterative Data-Science-Prozess Triggerpunkte für eine Modelländerung beinhalten.

Grundsätzliche Besonderheiten beim Retraining werden nicht gesehen; diese seien je nach Datenverfügbarkeit aber im Einzelfall zu klären. Hierbei gibt es laut den Rückmeldungen einen fließenden Übergang zwischen Modellpflege und -änderung, und wesentliche Ergebnisänderungen aus diesen Gründen müssten erklärbar sein. Speziell die Versicherungswirtschaft erwartet für Risikomodelle kein häufigeres Retraining als für „klassische Ansätze", wogegen für andere Anwendungsgebiete wie etwa Betrugserkennung ein häufiges Retraining erforderlich sein könne. Als Anforderung an das Retraining wird eine Qualitätsverbesserung des Modells formuliert.[1]

Modell-Governance wird wichtiger. Der Begriff der Modelländerung ist zu schärfen

Bezüglich organisatorischer Anpassungen zur Nutzung von ML-Methoden verdeutlichen die Antworten sehr unterschiedliche Vorstellungen bis hin zum Verschmelzen von Data-Science- und Modellierungseinheiten. Die Trennung von Modellierung und Validierung solle beibehalten, deren Zusammenarbeit jedoch intensiviert werden. Vielen Antworten gemeinsam ist der Wunsch nach einer klaren Modell-Governance.

1 Siehe dazu auch die EBA-Konsultation https://www.eba.europa.eu/eba-consults-machine-learning-internal-ratings-based-models

IV. Ausblick

Die Ergebnisse der Konsultation sind der Grundstein, um in den Dialog mit Unternehmen einzutreten. Damit soll, ergänzend zu den allgemeinen BDAI Prinzipien, Klarheit bei der Entwicklung und Anwendung von ML-Methoden im Kontext aufsichtlich relevanter Modelle in Säule 1 und Säule 2 geschaffen werden.

Gerade dort, wo die Konsultation ein breiteres Meinungsbild aufgezeigt hat, beispielsweise beim Thema Erklärbarkeit und der Adaptivität der Modelle, sind weitere Diskussionen notwendig. Gewisse Kombinationen der Charakteristika können eine intensivere aufsichtliche Auseinandersetzung erfordern als andere. Der Fokus der Aufsicht liegt darauf, die Überprüfbarkeit der Risikomodelle zu erhalten.

Mit dem weiteren Austausch soll sichergestellt werden, dass Klarheit bezüglich der aufsichtlichen Erwartungen besteht und sich die Erwartungen technologieneutral in das bestehende Regelwerk zu Säule 1 und 2 einfinden. Unternehmen erhalten damit ein regulatorisches Umfeld, das Investitionen in ML-Methoden ermöglicht und sie in die Lage versetzt, die Risiken der Methoden so früh wie möglich zu adressieren.

Um internationale Ansätze möglichst zu harmonisieren und sektorübergreifend gleichlautende Anforderungen an den Einsatz von ML-Methoden zu stellen, werden die Ergebnisse der Konsultation auch bei der Umsetzung der Digital Finance Strategy der EU Kommission[2] eingebracht und mit anderen europäischen Aufsichtsbehörden diskutiert.

2 EU Kommission, 2020, „Strategie für ein digitales Finanzwesen in der EU“,
online abrufbar: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52020DC0591

Anhang – Ausgewählte Rückmeldungen

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?

  „Eine strikte Definition von ML-Methoden halten wir nicht für zielführend."

  „Wir begrüßen den Vorschlag, die Aufsichts- und Prüfungspraxis anwendungsorientiert an den einzelnen Charakteristika der eingesetzten Methoden auszurichten. Mit diesem Ansatz ist es möglich, die Aufsichts- und Prüfungspraxis zwischen etablierten statistischen Methoden und neuartigen ML-Methoden abzustufen."

  „Aufgrund der Vielzahl an Ansätzen und Forschungsgruppen, die sich im Umfeld ML ausgebildet haben und derzeit entwickeln, erscheint eine einheitliche Definition zum einen unrealistisch und zum anderen wenig hilfreich."

  „Die strikte Definition von Modellcharakteristika macht keinen Sinn, denn die ML Methoden stellen nur einen Algorithmus zum Berechnen von Modellfunktionen dar."

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

  „Der Umgang mit Open Source Software, die verstärkt für ML-Methoden zur Verfügung steht, sollte geklärt werden"

  „Die Performance des Modells ist ein weiteres Charakteristikum mit Bedeutung. […] Des Weiteren halten wir Plausibilität für ein bedeutendes Charakteristikum."

  „In Verbindung mit der Erklärbarkeit und Anpassungsfähigkeit stellt aus unserer Sicht auch die Reproduzierbarkeit einen wichtigen Aspekt bzw. Charakteristik dar."

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

  „Datenquellen und Datenmengen spielen unseres Erachtens für die Unterscheidung klassischer und moderner ML-Methoden nur eine untergeordnete Rolle."

  „Die Interpretierbarkeit der Ergebnisse ist ein wichtiger Punkt, welcher aus unserer Sicht in Ihren Ausführungen berücksichtigt wird. Dies als gesondertes Charakteristikum aufzunehmen ist daher nicht notwendig."

  „Die Auslagerung von Prozessen ist ein generell mögliches Verfahren, welches schon heute stark durch regulatorische Anforderungen beeinflusst ist. Hierbei spielen Transparenz und Governance eine relevante Rolle, die unabhängig vom Einsatz von ML-Methodik bewertet werden muss."

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

  „Im Rahmen der Validierung zum internen Modell haben einige Unternehmen bereits erste ML-Methoden integriert. Neuronale Netze oder etwaige andere ML-Methoden werden unseres Wissens nach noch nicht für die offizielle Risikokapitalberechnung verwendet. […] Bereits zum Einsatz kommen solche Modelle bereits im Vertrieb und bei allgemeinen Betriebsabläufen. Geplant ist die Weiterentwicklung von Elementen des internen Risikomodells."

  „Bisher sind die tatsächlichen Einsatzgebiete eingeschränkt. Anwendungsfälle lassen sich jedoch im gesamten Bankbetrieb finden."

  „Wir sehen in vielen Bereichen derzeit keinen Anwendungsbereich von maschinellem Lernen, sondern eher „klassische" Verfahren oder (naturwissenschaftliche) Simulationsansätze."

  „Aktuell finden ML-Methoden nur begleitend zu klassischen Risikomodellentwicklungen als Challenger-Modelle Anwendung."

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.

  „Die derzeitige Prüfungs- und Anmeldepraxis fordert ein Maß an Nachvollziehbarkeit und Messbarkeit, die nicht geboten scheint."

  „Wir teilen die Einschätzung von BaFin und Deutscher Bundesbank, dass die bestehende Regulierung und Aufsichtspraxis aufgrund ihrer Technologieneutralität bereits einen sehr guten Rahmen für den Einsatz von ML in Risikomodellen bereitstellen."

  „Aus unserer Sicht bestehen keine Bedenken bei der Anwendung von ML-Methoden im Hinblick auf die bestehenden aufsichtlichen Anforderungen. Auch steht der Entwurf der KI-Verordnung im Einklang mit den Regelungen für Säule 1- und Säule 2-Modelle und es sind keine direkten Widersprüche zu erkennen."

  „Wir halten es für sehr wichtig, die Wettbewerbsfähigkeit der Versicherungswirtschaft nicht durch ein Übermaß an Regulierung zu beeinträchtigen."

  „Aktuell sehen wir keine Regelungen, welche den Einsatz von ML-Methoden grundsätzlich verhindern würden."

- 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?

  „Die entsprechenden Anforderungen in Kapitel 4.3.4 der EBA/GL/2020/06 an die automatisierte Kreditwürdigkeitsprüfung und Kreditvergabe werden in ihrem Umfang als angemessen beurteilt."

  „Die Anforderungen der EBA/GL/2020/06 mit Bezug auf die automatisierte Kreditwürdigkeitsprüfung und Kreditvergabe sollen einen hinreichenden Schutzmechanismus bieten. Die geforderten Maßnahmen bieten grundsätzlich eine gute Ausgangslage für die Anwendung von ML-Methoden in der Praxis."

- 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?

  „Nein."

  „Bei ML-Verfahren kommt dem Schutz der Trainings-, Validierungs- und Testdaten sowie des Quellcodes der KI-Anwendung eine wichtige Rolle zu. […] Weitere Punkte, die für die Aufsicht relevant sein könnten, sind: ein sich ggf. stärker verändernder Datenumfang, die Volatilität der Ergebnisse, sowie mangelnde Dokumentation zu Algorithmen."

  „Wir sehen keine weiteren Punkte, die eine Anpassung der bisherigen Aufsichtspraxis erfordern."

  „Wir sehen das eher durch Ausbildung/Fachkenntnisse/Standards und Richtlinien abgebildet."

- 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?

  „Aus numerischer Sicht hat die Implementierung der ML-Algorithmen eine starke Analogie zu bisher eingesetzten Methoden, so dass neben obiger Datenschutz-Thematik keine weiteren spezifischen IT-Risiken neu zu klassifizieren sind (ggf. sind bereits bestehende Risiken anzupassen)."

  „Für IT-Implementierungen sehen wir keine ML-spezifischen Risiken. Die bisherigen Anforderungen gelten bereits für die klassisch verwendeten ML-Modelle. […] 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 Prozesse."

  „Da der Kreis an Personen mit Zugriff auf die für die Risikorechnung relevanten Datenströme in der Regel begrenzt und klar umrissen ist und eine Vielzahl an Kontrollen und Prüfinstanzen etabliert sind, sehen wir keine größere Gefährdung durch „Adversarial Attacks". Darüber hinaus und durchaus unabhängig von den verwendeten Methoden nimmt die Bedeutung von IT-Risiken durch den gestiegenen Datenaustausch und -verarbeitung zu."

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?

  „Datengläubigkeit ist nicht erst ein Phänomen bei ML/KI, sondern existiert auch in der klassischen Umgebung."

  „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 ist, 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."

  „Wichtig ist, dass die verwendeten Datensätze so weit wie möglich den Echtdaten entsprechen, damit sie repräsentativ zum „use case" bleiben. Je mehr (hauptsächlich aus Datenschutzgründen) anonymisiert wird, desto mehr entfernen sich die Datensätze und darauf trainierte Modelle von den Echtdaten, was zu veränderten ML-Modellen und zusätzlichen Risiken führen kann."

  „Die Möglichkeit der Verwendung von unbearbeiteten Rohdaten in vielen ML-Verfahren und das Bestreben vieler Institute zur Prozessautomatisierung in allen Bereichen, kann dazu verleiten, Variablen ohne jegliche manuelle Interaktion in Modelle aufzunehmen ungeachtet der zugrunde liegenden Datenqualität oder – noch bedenklicher – ohne die Sinnhaftigkeit der Variablen für das zugrunde liegende Problem zu plausibilisieren."

  „Es ist bei der Anwendung von stark datengestützten Verfahren wie Maschinellem Lernen kritisch zu hinterfragen, ob die Daten tatsächlich geeignet sind um eine Extrapolation auf die für Kapitalanforderungen relevanten Szenarien durchzuführen."

- Welche Aspekte der Datenqualität werden durch die Anwendung von ML-Methoden aus Ihrer Sicht erleichtert?

  „Die Qualitätssicherung und der Schutz der Daten sind bereits jetzt im Fokus."

  „Wir sehen einen Vorteil, wenn i.R. von KI / ML viele Daten (auch Metadaten) herangezogen werden, da durch diese 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."

  „Je breiter die verwendete Datenbasis ist, desto geringer ist der erwartete Performanceverlust des Modells bei „Ausfall" eines Merkmals / einer Datenquelle."

  „ML-Methoden sind in der Lage, im gewissen Maße auch Störungen bzw. Verzerrungen in den Daten zu erkennen und auch bei unvollständigen/gestörten Daten sinnvolle Ergebnisse zu berechnen. Nichtsdestotrotz ist der Datenaufbereitungsprozess vor dem eigentlichen Modellentwicklungsprozess von großer Bedeutung, so dass hier stets eine hohe Qualität gegeben sein muss."

2.3 Erklärbarkeit rückt in den Fokus

- Welchen Einfluss hat die „Blackbox"-Eigenschaft Ihres Erachtens auf die Validierung der Verfahren?

  „Der „Blackbox"-Charakter rührt von der hohen Komplexität des Modells her, so dass Kausalzusammenhänge nicht transparent dargestellten werden können. Dennoch gehen wir davon aus, dass auch zukünftig beim Einsatz von ML-Methoden ein nicht vom Zufall getriebenes sondern reproduzierbares Ergebnis gewährleistet sein muss. Die Interpretierbarkeit der Ergebnisse lässt sich mit geeigneten Methoden (z.B. Sensitivitätsanalyse) verbessern."

  „Auf reine Blackbox-Modelle, ohne für diese mithilfe von XAI-Techniken zumindest eine prinzipielle Erklärbarkeit zu erreichen, sollte vernünftigerweise stets verzichtet werden, insbesondere aber im Rahmen von Risikomodellen."

  „Solange die Modellergebnisse zeitlich auch stabil verifizierbar sind und Sondereffekte (temporal & fachlich) in ausreichender Form mit verprobt worden sind, wird die Blackbox-Eigenschaft als unkritisch gesehen."

  „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 Mechanismus in seiner Gesamtheit und 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 die Nutzung von XAI-Ansätzen erleichtert werden."

  „Der Einfluss der „Blackbox"-Eigenschaft auf die Validierung sollte in jedem Einzelfall betrachtet werden."

  „ML/KI-basierte Ergebnisse oder die Geeignetheit von Methoden müssen für Spezialisten erklärbar bzw. plausibel sein, nicht aber jeder einzelne Arbeitsschritt oder einzelne Zwischenergebnisse. Auch klassisch erstellte Statistiken sind nicht per se aussagekräftig, sondern deren Ergebnisse und Methoden müssen erst interpretiert werden."

  „Der Prozess des Erklärbar-machens der Blackbox (Open-the-Blackbox) sollte bei der Verwendung moderne ML-Verfahren Teil sowohl der Initialvalidierung direkt im Anschluss an die Modellentwicklung als auch der regelmäßigen Validierung während der Modelllaufzeit sein."

- Welche Bedeutung messen Sie dem Trade-off zwischen Performance und Erklärbarkeit bei?

  „Je nach Anwendungsfall kann entweder die Erklärbarkeit oder die Performance des Modells als wichtiger 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."

  „Die Abwägung zwischen Performance und Erklärbarkeit ist sinnvoller Bestandteil der Modellauswahl unabhängig davon, ob es sich um ein ML-Modell oder ein klassisches Modell handelt."

  „Der „Trade-Off" zwischen Performance und Erklärbarkeit ist von großer Bedeutung für die Praxis und je nach Anwendung muss dies abgewogen werden."

- Bieten XAI-Techniken aus Ihrer Sicht (immer) einen Ausweg aus der „Blackbox"? Welche Verfahren haben sich als vielversprechend herausgestellt und bei welchen ML-Methoden?

  „Da XAI-Techniken in der Umsetzung durchaus einen erheblichen Mehraufwand zur Validierung klassischer Modelle bedeuten, sollte diesem Mehraufwand ein signifikanter Performancegewinn gegenüberstehen."

  „Aus unserer Sicht sind XAI-Techniken kein Ausweg aus der Blackbox. Verschiedene XAI-Techniken können lediglich Features hervorheben, die für eine Vorhersage primär entscheidend waren."

  „XAI-Techniken sind angemessene Tools, um die Interpretation von Modellen zu fördern."

  „Die Erfahrung wird erst noch zeigen müssen, welche Methoden für die Einschätzung des Verhaltens von Algorithmen am besten geeignet sind. Die Entwicklung weiterer Methoden zur Analyse des Verhaltens von Algorithmen ist derzeit ein dynamisches Feld - insbesondere für Algorithmen aus dem BDAI-Bereich. Dementsprechend ist es aus unserer Sicht zu früh, dafür Standards festlegen zu wollen."

  „Alle Ansätze bergen das Risiko, den Ergebnissen zu leicht zu vertrauen. In jedem Fall wird es wichtig sein, sich intensiv mit den Modellen und Inputdaten zu beschäftigen und Ergebnisse der XAI-Methoden intensiv zu verifizieren."

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

  „Eine nachgelagerte XAI kann genutzt werden, um zu überprüfen, ob eine Entscheidung eines Modells korrekt getroffen wurde. […] Dies dient lediglich für eine manuelle Überprüfung, die stichprobenhaft angewendet werden könnte, um auch dokumentieren zu können, dass eine gewisse Qualität vorliegt."

  „Eine regelmäßige Anwendung von XAI-Techniken ist empfehlenswert, um die ML-Modelle besser zu verstehen. XAI-Techniken können sowohl bei der Entwicklung als auch als Ex-Post-Tool einbezogen 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?

  „Eine Aktualisierung von Daten sowie Veränderungen der (Hyper-) Parameter, die sich bei ML-Modellen allein aus dem Training der Modelle ergeben, stellen dagegen keine Modellanpassung dar, da diese eher Teile der Anwendung des Modells im Rahmen der aktuellen Markt-/Geschäftslage darstellen."

  „Auch bisher stellt eine neue Kalibrierung, sofern sie den festgelegten Qualitätsstandards genügt, u.U. keine Modelländerung dar. Insofern wäre es naheliegend, dass auch bei ML-Methoden, sofern diese gewissen Qualitätsstandards genügen, keine Modelländerung vorliegt. Auch dies ist natürlich anwendungsabhängig."

  „Der bestehende Regulierungsrahmen und die Modellgenehmigung genügen hierfür; einer Erweiterung bedarf es nicht. Spezifische Aspekte stellen die Banken über ihre individuelle Governance sich."

  „Seitens der Aufsicht sollten klare Kriterien definiert werden, nach denen die Modellpflege von genehmigungspflichtigen Modelländerungen abgegrenzt werden kann."

„Das Training (inkl. dem Hyperparameter-Tuning) ist ein regelmäßiger Prozess und darf nicht als Modelländerung im aufsichtlichen Sinne eingestuft werden. Dies gehört zur (jährlichen) Re-Kalibrierung."

„Sollten nicht statt Veränderungen in den verwendeten Risikofaktoren vielmehr Veränderungen in den Gewichten der verwendeten Informationsarten für die Abgrenzung von Modellpflege oder Modelländerung betrachtet werden? Der Austausch eines Merkmals sollte nicht zwingend als Modelländerung aufgefasst werden, wenn das ersetzende Merkmal eine sehr ähnliche Information beauskunftet."

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

  „Unabhängig von der gewählten Methodik, d.h. auch für klassische Regressionsmodelle, werden regelmäßige Retrainings der Modelle als sinnvoll erachtet."

  „Außerdem ist zu beachten, dass „data science" aus iterativen Zyklen besteht. Es stellt sich die Frage, ab wann ein Zyklus als abgeschlossen gilt, und ab wann ein neuer beginnt. Dies könnte dann ein Triggerpunkt für eine Modellanpassung der Aufsichtsbehörde sein. Sehr häufige „retrainings" sind aus unserer Sicht nicht notwendig. Der treibende Faktor der Veränderung von Modellergebnissen sind eher die Datengrundlagen und deren Volatilitäten / Erweiterungen."

  „Es finden regelmäßige Rekalibrierungen der Modelle statt, sodass die Abgrenzung zwischen Modellpflege und aufsichtlich zu beurteilender Modelländerung schon heute relevant ist. Die Notwendigkeit häufiger Retrainings ist stark abhängig von dem jeweiligen Anwendungsfall. Sie sollte 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 ein Lernverfahren geändert wird, sollte dies als Modelländerung klassifiziert werden."

  „Per se wird ein Retraining nur durchgeführt, wenn dadurch eine höhere Qualität erwartet wird. Das Retrainings wird aber nicht persistiert, wenn das neu trainierte Modell keine höhere Qualität aufweist."

  „Bei Säule 1 und 2 Anwendungen wäre unsere Einschätzung, dass dort ein Re-Training im Normalfall nicht sehr oft vonnöten sein würde (im Gegensatz zu Anwendungen, die z.B. Kundenbetreuung / -service beinhalten)."

  „Insbesondere bei Methoden, die ein häufiges Re-Training der Modelle erfordern, ist es unserer Ansicht nach wichtig, sicherzustellen, dass hierdurch keine großen Sprünge in den Ergebnisvariablen auftreten. Treten dennoch Sprünge auf, sollte das im Modell-Monitoring erkennbar sein und begründet werden können."

- Werden ML-Methoden Ihres Erachtens eine Anpassung der Modell-Governance notwendig machen?

  „Aufgrund der Technologieneutralität der Regulierung und Aufsichtspraxis sind generell bei dem Einsatz von ML-Methoden in internen Modellen keine Änderungen am Aufsichtsvorgehen nötig."

  „Bei einer solchen Modell - Governance werden alle Parteien involviert, denn das gesamte Vorgehen von Data Load bis zum Model deployment muss erklärt und dokumentiert werden, sodass jede Partei nachvollziehen kann, was Input für ein Modell ist und welchen Output es liefert."

„Die Modell-Governance kann unverändert bleiben."

„Aus unserer Sicht werden durch ML-Methoden die bestehenden Aspekte der Modell-Governance noch deutlicher hervortreten als bisher."

- Wie arbeiten klassische Modellierungseinheiten, Validierer und neue „Data Science"-Einheiten zusammen?

  „Die Validierungsfunktion hat die Aufgabe, eine unabhängige Prüfung der Modelle, Prozesse und Berechnungen durchzuführen. Hierunter sollte auch die aktuelle Methodik geprüft werden. Der Einsatz von ML-Methoden kann sowohl von Modellierungseinheiten als auch von Validierungseinheiten erfolgen. Im Optimalfall findet ein reger Austausch zwischen diesen Einheiten statt, ggf. auch unter Einbezug weiterer „Data Science"-Einheiten (falls diese bereits implementiert wurden)."

  „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."

  „Data Science- und Modellierungseinheiten werden bei produktiver Nutzung eines ML-Modells miteinander verschmelzen, während die Validierer wie bisher unabhängig davon agieren. Vorstellbar ist, dass der Know-How Transfer für eine erfolgreiche Validierung verstärkt durch die Data Science-Einheit erfolgt."

  „Unabhängig von ML-Methoden ist die enge Zusammenarbeit von Modellierungseinheiten und Validierungseinheit zu empfehlen. Dabei sollte die Validierungseinheit auf konstruktive Art und Weise die entwickelten Modelle, Prozesse und Ergebnisse auf Angemessenheit prüfen. […] „Data Science" stellt für die Versicherungsunternehmen teilweise ein neues Feld dar, so dass noch nicht zwingend eigene „Data Science" Teams fest verankert sind. Ein regelmäßiger Austausch bzgl. „Best Practice" und Anwendungen (z.B. im Rahmen von „Use Cases") zwischen allen Abteilungen sollte in jedem Fall stattfinden."

Impressum

Herausgeber
Deutsche Bundesbank
Postfach 10 06 02, 60006 Frankfurt am Main
www.bundesbank.de

Bundesanstalt für Finanzdienstleistungsaufsicht (BaFin)
Graurheindorfer Straße 108, 53117 Bonn
Marie-Curie-Straße 24 – 28, 60439 Frankfurt am Main
www.bafin.de