BaFin

Bundesanstalt für
Finanzdienstleistungsaufsicht

Big Data und künstliche Intelligenz:

Prinzipien für den Einsatz von Algorithmen in Entscheidungsprozessen

15. Juni 2021

© Pixabay/Gert Altmann

Inhaltsverzeichnis

Big Data und künstliche Intelligenz:
Prinzipien für den Einsatz von Algorithmen in Entscheidungsprozessen

Auf dem Finanzmarkt besteht großes Interesse am Einsatz von Big Data und Artificial Intelligence (BDAI) – und damit allgemein von (komplexen) Algorithmen.[1] Die Unternehmen sehen darin zum Beispiel die Chance, Risiken besser einzuschätzen und Kosten zu senken. Profitieren können auch deren Kundinnen und Kunden. Aber für beide Seiten ist der Einsatz von BDAI auch mit Risiken verbunden, die es zu kontrollieren gilt.

Künstliche Intelligenz als Verbindung von Machine Learning und Big Data | Technisch definiert die BaFin den Begriff der künstlichen Intelligenz (KI) in ihrer Studie „Big Data trifft auf künstliche Intelligenz" als Kombination aus großen Datenmengen (Big Data), Rechenressourcen und maschinellem Lernen (Machine Learning – ML).[2] Beim maschinellen Lernen wird Computern auf Basis spezieller Algorithmen die Fähigkeit verliehen, aus Daten und Erfahrungen zu lernen. Im Vergleich zu regelbasierten Verfahren erfolgt das Lernen ohne dass der Programmierer bzw. die Programmiererin vorgibt, welche Ergebnisse aus bestimmten Datenkonstellationen wie abzuleiten sind. Dieses Verständnis vom Begriff der künstlichen Intelligenz ist auch Teil der Definition des Financial Stability Boards (FSB).[3]

Unter Algorithmen versteht die BaFin Handlungsvorschriften, die in der Regel in ein Computerprogramm integriert sind und ein (Optimierungs-) Problem oder eine Klasse von Problemen lösen. Neben der Unterscheidung nach der Art der Algorithmen (wie wird das Problem technisch gelöst) lassen sich Anwendungen des ML auch nach Ergebnistypen (man unterscheidet grundsätzlich zwischen Klassifikation, Regression und Clustering) und Datentypen (spezielle Ansätze existieren z.B. für Text, Sprache und Bilddaten) differenzieren.

Derzeit keine klare Abgrenzung von klassischen Verfahren möglich | Die obenstehende Definition von künstlicher Intelligenz ermöglicht jedoch keine trennscharfe Abgrenzung von klassischen statistischen Verfahren und dabei verwendeten Algorithmen. Die bestehende Definition entsprechend weiterzuentwickeln, zählt zu den Herausforderungen, vor der Aufsicht, Regulierung und vor allem Standardsetzer stehen.

Hohe Komplexität, kurze Rekalibrierungszyklen und hohe Automatisierung | Dennoch lassen sich drei wesentliche Merkmale nennen, die moderne Methoden und Anwendungen von künstlicher Intelligenz charakterisieren und bei einer Betrachtung der Risiken eine Rolle spielen: eine hohe Komplexität des zugrundeliegenden Algorithmus, kurze Rekalibrierungszyklen und ein hoher Grad an Automatisierung:

[1] BaFin (2018): Big Data trifft auf Künstliche Intelligenz: Herausforderungen und Implikationen für Aufsicht und Regulierung von Finanzdienstleistungen, URL: https://www.bafin.de/dok/10985478
[2] A.a.O., S. 25 f.
[3] FSB (2017): Artificial intelligence and machine learning in financial services: Market developments and financial stability implications, S. 3 f., URL: https://www.fsb.org/2017/11/artificial-intelligence-and-machine-learning-in-financial-service/

- Algorithmen des maschinellen Lernens sind häufig komplexer als die, die bei klassischen statistischen Verfahren verwendet werden. Diese höhere Komplexität, die sich insbesondere bei Verfahren wie künstlichen neuronalen Netzen findet, macht die Nachvollziehbarkeit und Überprüfbarkeit der algorithmischen Ergebnisse schwierig.

- Die Kombination aus selbstlernenden Algorithmen des maschinellen Lernens und täglich neu verfügbaren (Massen-)Daten führt dazu, dass die Rekalibrierungszyklen von Modellen und Algorithmen immer kürzer werden; die Grenzen zwischen Kalibrierung und Validierung verschwimmen.

- Algorithmen werden immer mehr zur Automatisierung auch teilweise nicht standardisierter Prozesse und Entscheidungen eingesetzt, die schnell und in großer Stückzahl (Skalierung) erfolgen.

Minimalanforderungen für den Einsatz von KI als Diskussionsgrundlage | Da es bislang keine trennscharfe Definition der künstlichen Intelligenz gibt, formuliert die BaFin in dieser Veröffentlichung allgemeine Prinzipien für den Einsatz von Algorithmen in Entscheidungsprozessen[4] von Finanzunternehmen, insbesondere von solchen Algorithmen, die die zuvor genannten Merkmale aufweisen.

Die Prinzipien stellen vorläufige Überlegungen zu aufsichtlichen Mindestanforderungen für den Einsatz von künstlicher Intelligenz dar. Sie sind damit eine Diskussionsgrundlage für den Austausch mit diversen Stakeholdern. Zugleich sollen sie aber bereits jetzt den von der BaFin beaufsichtigen Unternehmen als Orientierungshilfe dienen. Dabei wird die oben vorgestellte Arbeitsdefinition zugrunde gelegt. Vor allem gelten die Prinzipien für solche Algorithmen und algorithmischen Entscheidungsprozesse, die sich wie oben beschrieben, durch hohe Komplexität, kurze Rekalibrierungszyklen und einen hohen Automatisierungsgrad auszeichnen. Für diese Veröffentlichung gilt jedoch: Die Prinzipien schließen nicht aus, dass für bestimmte regulierte Tätigkeiten bereits strengere Regulierung bzw. Verwaltungspraxis einschlägig ist. Diese ist vorrangig zu beachten.

I. Konzeptioneller Rahmen

Betrachtung des gesamten Prozesses | Ob Algorithmen brauchbare Ergebnisse erzeugen, hängt davon ab, wie beaufsichtigte Unternehmen sie in Entscheidungsprozesse einbetten: Ein für einen bestimmten Kontext geeigneter Algorithmus kann in einer anderen Situation zu

[4] Dies umfasst auch den Einsatz von Algorithmen in entscheidungsvorbereitenden Prozessen wie zum Beispiel der Quantifizierung und Bewertung von Risiken.

unbrauchbaren Ergebnissen führen. Darüber hinaus sind die Ergebnisse eines Algorithmus abhängig von den verfügbaren Daten und deren Qualität. Deswegen richtet sich der aufsichtliche Fokus auf den gesamten algorithmenbasierten Entscheidungsprozess – von der Datenquelle bis zur Datensenke und der Einbindung in den Geschäftsprozess.

Keine generelle Billigung von Algorithmen | Algorithmenbasierte Entscheidungsprozesse werden – für sich genommen – in der Regel nicht von der BaFin gebilligt.[5] Stattdessen prüft und beanstandet die Aufsicht diese Prozesse risikoorientiert und anlassbezogen zum Beispiel bei Erlaubnisverfahren, in der laufenden Aufsicht und der Missstandsaufsicht.

In begründeten Ausnahmefällen, etwa bei internen Modellen, die Banken und Versicherer verwenden, um ihre regulatorischen Kapitalanforderungen zu ermitteln, überprüft die BaFin, ob die verwendeten algorithmischen Verfahren geeignet sind. Dabei geht es unter anderem um Methodik, Kalibrierung und Validierung.

Risikoorientiert, proportional und technologieneutral | Das aufsichtliche und regulatorische Grundprinzip „gleiches Geschäft, gleiches Risiko, gleiche Regeln" fordert auch bei der Beaufsichtigung von algorithmischen Entscheidungsprozessen einen risikoorientierten, proportionalen und technologieneutralen Ansatz.

Eine intensivere Aufsicht ist dementsprechend dann angebracht, wenn mit dem Einsatz eines Algorithmus in Entscheidungsprozessen (zusätzliche) Risiken verbunden sind.[6] Ein typisches Risiko ist die hohe Skalierbarkeit der Prozesse und damit auch der potenziellen Fehler. Bei vielen algorithmischen Entscheidungsprozessen kommen, wie oben beschrieben, auch eine hohe Komplexität, kurze Rekalibrierungszyklen und ein hoher Automatisierungsgrad hinzu. Dadurch kann es für Unternehmen und Aufsicht erschwert werden, die Entscheidungsprozesse nachzuvollziehen und zu prüfen. Außerdem droht beispielsweise die Gefahr, dass die algorithmischen Ergebnisse verzerrt sind (Bias).[7]

Bestehende Regeln werden ergänzt, präzisiert und weiterentwickelt | Mit den im Folgenden ausgeführten Prinzipien sollen bestehende Regulierung und Verwaltungspraxis ergänzt und konkretisiert werden. Die Prinzipien stellen vorläufige Überlegungen zu aufsichtlichen Mindestanforderungen für den Einsatz von künstlicher Intelligenz dar. Zugleich sollen sie aber bereits jetzt den von der BaFin beaufsichtigen Unternehmen als Orientierungshilfe dienen. Viele dieser Prinzipien sind hierbei nicht gänzlich neu, sondern entwickeln punktuell die prinzipienorientierte und technologieneutrale Regulierung weiter.

[5] BaFin (2020): Generelle Billigung von Algorithmen durch die Aufsicht? Nein, aber es gibt Ausnahmen, BaFinJournal 03/2020, URL: https://www.bafin.de/dok/13783136

[6] Solche Risiken können auch durch Auffälligkeiten im laufenden Aufsichtsbetrieb oder Hinweise sichtbar werden.

[7] Hinzu kommen die mit dem jeweiligen Entscheidungsprozess ohnehin verbundenen Risiken – unabhängig vom Einsatz eines Algorithmus.

Für diese Veröffentlichung gilt jedoch: Die Prinzipien schließen nicht aus, dass für bestimmte regulierte Tätigkeiten bereits strengere Regulierung bzw. Verwaltungspraxis einschlägig ist. Diese sind vorrangig zu beachten. Die Anwendung der Prinzipien führt nicht zu einer Befreiung von geltenden gesetzlichen und aufsichtlichen Vorgaben. Selbstverständlich ist es aber denkbar, dass sich die laufende Verwaltungspraxis weiterentwickelt.

Im Folgenden werden zunächst übergeordnete Prinzipien für die Verwendung von Algorithmen in Entscheidungsprozessen vorgestellt (Kapitel 2). Diese Prinzipien sind für die der Erstellung und die Anwendung des Algorithmus wichtig. Im Anschluss widmet sich Kapitel 3 den spezifischen Prinzipien für die Entwicklungsphase. In Kapitel 4 finden sich die spezifischen Prinzipien für die Anwendungsphase.

II. Übergeordnete Prinzipien

Klare Verantwortung der Geschäftsleitung | Die Geschäftsleitung ist verantwortlich für die unternehmensweiten Strategien und Leitlinien bzw. Richtlinien zum Einsatz von algorithmenbasierten Entscheidungsprozessen. Darin sollten sowohl die Potenziale solcher Prozesse als auch deren Grenzen und Risiken berücksichtigt und klar benannt werden.

Für materielle unternehmerische Entscheidungen ist der Vorstand stets selbst verantwortlich, auch dann, wenn sie auf Algorithmen fußen. Dies setzt zum einen ein adäquates technisches Verständnis bei der Geschäftsleitung voraus. Zum anderen müssen die Berichtslinien und Berichtsformate so gestaltet sein, dass eine risikoadäquate und adressatengerechte Kommunikation gesichert ist – und zwar von der Ebene des Modellierens bis hin zur Geschäftsleitung.

Schon im Jahr 2017 hat die BaFin die Aufnahme von IT-Spezialistinnen und -Spezialisten in die Geschäftsleitung beaufsichtigter Unternehmen vereinfacht. Ziel war es, die IT-Kompetenz in den Geschäftsleitungen der Unternehmen weiter zu fördern.

Die unternehmensweite Strategie zum Einsatz von algorithmenbasierten Entscheidungsprozessen sollte sich auch in der IT-Strategie widerspiegeln. Soweit erforderlich, müssen auch in den unabhängigen Kontrollfunktionen (z.B. in der Compliance und der internen Revision) entsprechendes Fachwissen vorhanden sein.

Adäquates Risiko- und Auslagerungsmanagement | Es ist auch Aufgabe der Geschäftsleitung, ein an den Einsatz von algorithmenbasierten Entscheidungsprozessen adaptiertes Risikomanagement zu etablieren. Werden Anwendungen von einem Dienstleister bezogen, muss die Geschäftsleitung zudem ein effektives Auslagerungs- bzw.

Ausgliederungsmanagement einrichten. Hierbei sind Verantwortungs-, Berichts- und Kontrollstrukturen klar festzulegen.

Bei der Etablierung eines adäquaten Risikomanagements ist es sinnvoll, die Risiken eines algorithmischen Entscheidungsprozesses zu berücksichtigen: Risikomitigierende Maßnahmen und Prozesse sollten nach dem Verursacherprinzip genau da ansetzen, wo ein Risiko seinen Ursprung hat. Ferner sollten Wahrscheinlichkeit und potenzielles Ausmaß von Schäden durch fehlerhafte algorithmenbasierte Entscheidungen klar analysiert und die Ergebnisse dieser Analyse dokumentiert werden. Zudem sollte ein übergreifendes Rahmenwerk im Unternehmen etabliert werden, das eine Aufstellung aller algorithmenbasierten Entscheidungsprozesse und deren wechselseitige Abhängigkeit berücksichtigt (z.B. „Model Risk Management Framework").

Use-case: Anwendung von Telematik-Tarifen in der KFZ-Versicherung

Ein KFZ-Versicherer bietet bestimmten Kundenkreisen eine Telematik-Option an. Zu- oder Abschläge auf den Basistarif leitet er aus versicherungsnehmerspezifischen Risikoprofilen ab. Für die Ermittlung dieser Risikoprofile ermittelt er fortlaufend Telematik-Daten wie Geschwindigkeit und GPS-Ortung und wertet sie mittels künstlicher Intelligenz aus.

Bei der Verwendung von personenbezogenen Daten muss er die Datenschutzregeln (u.a. Artikel 5 Datenschutzgrundverordnung – DSGVO) beachten (Prinzip **„Datenschutzregeln beachten")**. Sowohl bei der Entwicklung als auch bei der Anwendung des Algorithmus muss er gewährleisten, dass diese Daten nur einem sachbezogenen engeren Personenkreis zugänglich sind und nicht an unberechtigte Dritte gelangen.

Im Sinne des Prinzips **„Adäquates Risiko- und Auslagerungsmanagement"** muss das Unternehmen sicherstellen, dass die verwendeten Datenübermittlungsschnittstellen sicher und funktionsfähig sind. Dabei muss es auf einen ausreichenden Schutz vor Datenmanipulation achten. Zu berücksichtigen ist auch, dass an den Schnittstellen häufig Drittanbieter für die Speicherung in einer Cloud eingebunden sind.

Wendet das Unternehmen algorithmenbasierte Telematik an, muss es auch Notmaßnahmen vorsehen, falls technische Probleme auftreten oder das System ausfällt (Prinzip „**Etablierung von Notmaßnahmen"**). Für den Fall, dass bei einem Systemausfall keine konkreten Daten mehr live übermittelt werden können, muss das Unternehmen in der Lage sein, zum Beispiel alternative Aufzeichnungssysteme einzusetzen oder für die Ausfallszeit die Rabatte pauschalieren.

Bias vermeiden | Bei algorithmenbasierten Entscheidungsprozesse muss ein Bias, also die systematische Verzerrung von Ergebnissen, vermieden werden. Hintergrund: Einen Bias gilt es zum einen zu vermeiden, um unternehmerische Entscheidungen treffen zu können, die auf nicht systematisch verzerrten Ergebnissen fußen; zum anderen, um eine systematische, auf einem Bias fußende Benachteiligung einzelner Kundengruppen und daraus resultierende Reputationsrisiken auszuschließen.

Es handelt sich hierbei um ein übergeordnetes Prinzip, da solche Verzerrungen von der Entwicklung des Prozesses bis hin zu seiner Anwendung auftreten können. Beispiele: Es werden für das jeweilige Anwendungsgebiet keine Daten in ausreichender Qualität und Quantität verwendet; bestimmte Merkmale werden unsachgemäß ausgeschlossen oder übergewichtet; eigentlich korrekte algorithmische Ergebnisse werden systematisch falsch interpretiert.

Dem Verursacherprinzip entsprechend gilt es, das Risiko eines Bias dort zu identifizieren, wo er entstehen kann, dieses Risiko zu analysieren und entweder auszuschließen oder zumindest zu mitigieren.

Gesetzlich untersagte Differenzierung ausschließen | Für einige Finanzdienstleistungen ist zudem gesetzlich festgelegt, dass bestimmte Merkmale nicht zur Differenzierung – also zur Risiko- und Preiskalkulation – herangezogen werden dürfen. Werden dennoch systematisch Konditionen auf Basis solcher Merkmale gestaltet, besteht die Gefahr der Diskriminierung. Eine solche Gefahr besteht auch, wenn man diese Merkmale durch eine Approximation ersetzt.

Damit verbunden wären erhöhte Reputationsrisiken und ggf. Rechtsrisiken. Auch die BaFin sähe sich unter Umständen veranlasst, Maßnahmen zu ergreifen, zum Beispiel in der Missstandsaufsicht.

Die Unternehmen sollten (statistische) Überprüfungsprozesse etablieren, die Diskriminierung ausschließen.

III. Spezifische Prinzipien für die Entwicklungsphase

Datenstrategie und Datengovernance | Abhängig vom Anwendungsbereich und von den Merkmalen des Algorithmus müssen Daten in ausreichender Qualität und Quantität verwendet werden. Die Unternehmen müssen über ein überprüfbares Verfahren (Datenstrategie) verfügen, das die fortwährende Datenbereitstellung gewährleistet und definiert, welche Ansprüche jeweils an Qualität und Quantität der Daten erfüllt sein müssen. Die Datenstrategie muss in einer Datengovernance umgesetzt und Zuständigkeiten müssen klar umrissen werden.

Datenschutzregeln beachten | Jede Verwendung von Daten in algorithmenbasierten Entscheidungsprozessen muss konform sein mit den geltenden Datenschutzbestimmungen. Datenschutzrechtliche Vorgaben für die Nutzung von Daten sollten bereits bei der Planung algorithmischer Entscheidungsprozesse berücksichtigt werden. So sind insbesondere die Offenlegungspflichten gegenüber betroffenen Personen zu beachten.

Korrekte, robuste und reproduzierbare Ergebnisse sicherstellen | Das übergeordnete Ziel ist, korrekte und robuste Ergebnisse sicherzustellen. Die Ergebnisse eines Algorithmus sollten zudem reproduzierbar sein. Der Anwender sollte also zum Beispiel bei einer späteren Überprüfung durch einen unabhängigen Dritten in der Lage sein, die Ergebnisse zu reproduzieren.

Diese Reproduzierbarkeit gewährleistet zum einen ein Mindestmaß an unternehmensinterner, aber auch externer Nachvollziehbarkeit und Überprüfbarkeit der Ergebnisse. Zum anderen setzt sie Sorgfalt und Genauigkeit bei der Auswahl des Algorithmus, bei der Kalibrierung und der Dokumentation voraus.

Dokumentation zur internen und externen Nachvollziehbarkeit | Eine hinreichende Dokumentation ist Voraussetzung dafür, dass Algorithmen und die zugrundeliegenden Modelle überprüft werden können – vom Unternehmen selbst und von Abschlussprüfern und Aufsicht. Hierbei sind insbesondere drei Schritte zu berücksichtigen:

Schritt 1: Die Modellauswahl muss dokumentiert werden. Hierbei sind mindestens die folgenden Prüfungen zu berücksichtigen: Erstens muss geprüft werden, ob das Modell grundsätzlich für die konkrete Anwendung geeignet ist. Zweitens sollten statistische Erwägungen zur Modellgüte herangezogen werden. Drittens sollte auch die Modellkomplexität und damit die Interpretierbarkeit und Überprüfbarkeit betrachtet werden. Insbesondere ist eine Verbesserung der Prognosegüte gegen die damit potenziell einhergehende Erhöhung der Modellkomplexität abzuwägen. Die Entscheidung für ein komplexes Modell sollte in jedem Fall erklärt und begründet werden.

Schritt 2: Die Kalibrierung und das Training des Modells sind zu dokumentieren. So sind relevante Kalibrierungsdetails, zum Beispiel die Wahl von sog. Tuning-Parametern, zu dokumentieren und zu begründen. Werden Modelle mit Hilfe von Trainingsdaten kalibriert, muss die Auswahl dieser Daten dokumentiert werden.

Schritt 3: In diesem Schritt muss die Modellvalidierung beschrieben werden. Details zur Modellvalidierung finden sich im folgenden Prinzip „Angemessene Validierungsprozesse".

Die Dokumentation der Modellauswahl, -kalibrierung und -validierung lässt sich unter Umständen nicht klar voneinander trennen. Der Hintergrund ist folgender: Die Modellgüte kann mitunter erst nach Erstkalibrierung und Validierung bestimmt werden. Zudem ist eine Modellauswahl erst über den Vergleich verschiedener Modelle möglich.

Use-Case: Ergänzende Informationsauswertung für das Kreditrating

Ein Kreditinstitut hat ein Verfahren etabliert, mit dem es Kreditratings und Ausfallwahrscheinlichkeiten von Unternehmen ermittelt. Dieses Verfahren ergänzt es nun mit einem Verfahren, das mit Hilfe von Natural Language Processing (NLP) die jährlichen Geschäftsberichte der Unternehmen auswertet und nach Schlüsselbegriffen sucht, die Rückschlüsse auf das Kreditrating erlauben. In diesem neuen Verfahren werden Random-Forest-Ansätze des Machine-Learnings eingesetzt, also zufällig erzeugte Entscheidungsbäume, deren Vorhersagequalität durch ein maschinelles Lernverfahren optimiert wird.

Die auf diese Weise gewonnene Einschätzung ergänzt das etablierte Kreditrating, das auf quantitativen Unternehmensinformationen basiert. Für die endgültige Einstufung des Unternehmens werden beide Ergebnisse kombiniert.

Das Kreditinstitut hat vor der endgültigen Einstufung allerdings eine Kontrolle vorgesehen: Weichen das etablierte und das um den ML-Ansatz erweiterte Kreditrating deutlich voneinander ab, trifft ein Kreditanalyst bzw. eine Kreditanalystin die endgültige Entscheidung. Dabei vergleicht prüft er die Validität deserwirderden die ML-Ergebnisses anhand mit einer unabhängigen Experteneinschätzung verglichen (spezifisches Prinzip **„Putting the human in the loop"**).

Für seine Experteneinschätzung benötigt der Analyst Einblick in die Klassifikation, also die vom ML-Verfahren verwendeten Schlüsselwörter und ihre Auswirkung auf die Einstufung. Das Verfahren muss daher im Sinne des Prinzips **„Dokumentation zur internen und externen Nachvollziehbarkeit"** die Ergebnisse aus dem Trainingsverfahren in einer für den Experten geeigneten Form dokumentieren.

Angemessene Validierungsprozesse | Jeder Algorithmus sollte vor Übernahme in den operativen Betrieb einen angemessenen Validierungsprozess durchlaufen. Diese Initialvalidierung sollte stets eine unabhängige, nicht in die ursprüngliche Modellierung eingebundene Funktion bzw. Person vornehmen oder zumindest begutachten.

Außerdem ist festzulegen und zu dokumentieren, in welchen regelmäßigen Abständen ein Algorithmus erneut einer Validierung unterzogen werden muss (laufende Validierung). Neben einem regelmäßigen Turnus mit angemessenen Abständen sind Faktoren zu benennen, die zu einer Ad-hoc-Validierung des Algorithmus und damit potenziell zu dessen Neu-Kalibrierung oder der Auswahl eines alternativen Algorithmus führen.

Solche Faktoren können zum Beispiel sein: eine systematische Veränderung der Inputdaten, externe (makroökonomischer) Schocks, eine Veränderung der rechtlichen Voraussetzungen, unter denen ein Algorithmus betrieben wird, und ein Feedback aus der Outputphase wie das Brechen eines Limits (zum letztgenannten Punkt siehe Prinzip „Intensive Freigabe- und Feedbackprozesse“).

Verwendung von relevanten Daten zur Kalibrierung und Validierung | Eine Kalibrierung oder Validierung mit unpassenden oder fehlerhaften Daten kann die Performance des Algorithmus im Livebetrieb systematisch beeinträchtigen.[8] Deshalb sollte ein besonderer Fokus auf die Auswahl dieser Daten gerichtet und die Auswahl sorgfältig dokumentiert werden. Die Daten müssen für den jeweiligen Anwendungsbereich relevant und repräsentativ sein. Sie müssen also zum Beispiel Informationen zu allen relevanten Untergruppen enthalten. Andernfalls kann es aufgrund von unausgewogenen Daten bei der Kalibrierung oder Validierung zu einem Bias in der Modellierung kommen. Ein Bias muss bereits bei der Datenaufbereitung vermieden werden, also etwa bei der Datenaggregation (siehe hierzu auch das übergeordnete Prinzip **Bias vermeiden**).

Abhängig von der Reichweite und dem Risikogehalt der Entscheidung, für die ein Algorithmus verwendet wird, sollten außerdem verschiedene Maßnahmen zur späteren Nachvollziehbarkeit und Überprüfbarkeit der Kalibrierung und Validierung getroffen werden: In sehr risikosensitiven Entscheidungsprozessen sollten die hierfür verwendeten Daten gespeichert und vorgehalten werden. In weniger risikosensitiven Entscheidungsprozessen sollten mindestens der Auswahlprozess und die Zusammensetzung dokumentiert werden (z.B. über aussagekräftige zusammenfassende Statistiken insbesondere zur Verteilung der entsprechenden Daten). Es sollte auf jeden Fall sichergestellt werden, dass eine nachträgliche Nachvollziehbarkeit zur internen Kontrolle, Qualitätssicherung und Revision noch möglich sind, solange der Algorithmus im Einsatz ist.

[8] Bei bestimmten Algorithmen, etwa solchen des unüberwachten Lernens, wird klassischerweise kein Datensatz zur Kalibrierung des Algorithmus (Testdatensatz) verwendet. Umso wichtiger ist in solchen Fällen die Validierung und Auswahl der für die Validierung verwendeten Datensätze.

IV. Spezifische Prinzipien für die Anwendung

Interpretation und Verwertung algorithmischer Ergebnisse für die Entscheidungsfindung

Die Ergebnisse des Algorithmus müssen interpretiert und in Entscheidungsprozesse eingebunden werden. Dies kann automatisch geschehen, oder aber, indem Experten einbezogen werden. In jedem Fall muss ein funktionierender Mechanismus etabliert werden, der unter anderem ausreichende Kontrolle, Feedbackloops und Anpassungsregelungen zur Entwicklungsphase umfasst. Wichtig ist auch zu evaluieren, inwieweit die Interaktion mit anderen Algorithmen zu einer Risikoaggregation führt und ob die Verwendung der Algorithmen in Einklang steht mit dem gesamten Risikomanagement des Unternehmens.

„Putting the human in the loop“ | Die Beschäftigten sollten in die Interpretation und Verwertung der algorithmischen Ergebnisse für die Entscheidungsfindung angemessen eingebunden werden. Wie stark sie eingebunden werden, sollte davon abhängen, wie geschäftskritisch der Entscheidungsprozess ist und mit welchen Risiken er behaftet ist (zu möglichen Risiken und zum Aufbau eines adäquaten Risikomanagements siehe auch Kapitel 2 „Übergeordnete Prinzipien“). Die Einbindung sollte darüber echten Nutzen bringen und sich nicht darauf beschränken, dass jede algorithmische Entscheidung lediglich freigegeben werden muss. Ein Beispiel für eine effektive Einbindung ist der unten genannte Limitprozess. Er steuert die Intensität der Einbindung in Abhängigkeit davon, ob und wie sehr algorithmische Ergebnisse von der Norm abweichen. Außerdem bietet es sich in geschäftskritischen Prozessen an, Zeitfenster zu definieren, in denen eine Entscheidung noch revidiert werden kann und menschliche Intervention noch möglich ist.

Use-Case: Sanktionsscreening in der Geldwäscheerkennung

Beim Sanktionsscreening im Rahmen der Geldwäscheerkennung muss unter anderem sichergestellt werden, dass eine Transaktion nicht gegen Embargos verstößt. Hierzu gleichen die beaufsichtigten Unternehmen in der Regel Zahlungsdaten und Kundendaten mit Sanktions- und Embargolisten ab. Vor dem Einsatz von BDAI nahmen in dem vorliegenden Use-Case zwei unabhängig voneinander agierende Compliance-Beschäftigte diesen Abgleich vor. Ein dritter Mitarbeiter wurde in kritische Fälle eingebunden, nämlich wenn die beiden zuerst prüfenden Compliance-Beschäftigten zu unterschiedlichen Ergebnissen kamen.

Die Prüfungstätigkeit eines der beiden Compliance-Beschäftigten wird nun eins-zu-eins durch einen Algorithmus ersetzt, die Prüfungstätigkeit des zweiten und dritten Mitarbeiters verändert sich jedoch nicht. Das heißt, es findet weiterhin auch eine unabhängige menschliche Prüfung statt. Die freigewordenen Personalressourcen lassen sich auf diese Weise gezielt in die Bearbeitung von kritischen Fällen investieren, in denen der Algorithmus und der (menschliche) Zweitprüfer zu unterschiedlichen Ergebnissen kommen. Obwohl der Algorithmus in diesem Fall einen der menschlichen Bearbeiter vollständig ersetzt, findet im Sinne des Prinzips **„Putting the human in the loop“** eine effektive und risikosensitive menschliche Einbindung statt.

Der besondere Fokus auf kritische Fälle (inklusive der Einbindung eines weiteren Bearbeiters), in denen Algorithmus und menschlicher Zweitbearbeiter zu unterschiedlichen Ergebnisse, kommen, kann zudem als ein **„intensiver Freigabe- und Feedbackprozess“** verstanden werden (siehe hierzu das gleichnamige Prinzip).

Intensive Freigabe- und Feedbackprozesse | Bei der Verwendung algorithmisch erzeugter Ergebnisse in Entscheidungsprozessen sollten risikoorientiert vorab die Situationen klar definiert sein, die einen intensiveren Freigabeprozess nach sich ziehen. Dies kann zum Beispiel in Form eines stufenweisen Vorgehens mit Hilfe von Schwellenwerten erfolgen: Ab der ersten Stufe sollte die Freigabe durch einen Menschen auch bei ansonsten automatischen Prozessen erfolgen. Ab der zweiten Stufe sollte eine solche Freigabe nur noch nach zusätzlicher Überprüfung der Inputdaten auf mögliche Besonderheiten (z.B. Ausreißer) möglich sein. In der letzten Stufe sollte eine Freigabe zunächst nicht erfolgen („Stopping-Rule“), sondern es sollte ein Signal zur Überprüfung des Modells (Ad-hoc-Validierung mit der möglichen Folge einer Modellrekalibrierung oder der Auswahl eines gänzlich neuen Modells) und zur Unterbrechung des Prozesses gegeben werden. Eine Freigabe ist dann erst nach dieser intensiven Überprüfung und ggf. Änderung der Ergebnisse möglich. Ein solcher stufenweiser Schwellenwertprozess kann das Risiko von Fehlentscheidungen eines algorithmischen Entscheidungsprozesses verringern und zugleich die Ergebnisqualität aufgrund eines fortwährenden Feedbackmechanismus langfristig verbessern.

Etablierung von Notmaßnahmen | Unternehmen sollten Maßnahmen vorsehen, mit denen sich der Geschäftsbetrieb aufrechterhalten ließe, falls es zu Problemen bei algorithmenbasierten Entscheidungsprozessen kommt. Das gilt zumindest für geschäftskritische Anwendungen. Ein Beispiel: Ein Schwellenwert wird im Rahmen des oben beschriebenen Stufenmodells überschritten, und der Algorithmus muss vor weiteren Einsätzen zunächst erneut intensiv validiert werden.

Laufende Validierung, übergeordnete Evaluation und entsprechende Anpassung | In der praktischen Anwendung müssen Algorithmen laufend validiert werden, um die Funktionalität und Abweichungen anhand von festgelegten Parametern zu überprüfen und ggf. Anpassungen vorzunehmen. Die Validierung ist besonders dann notwendig, wenn neue oder unvorhergesehene interne oder externe Risiken auftreten, die bei der Erstellung der Algorithmen nicht berücksichtigt werden konnten. Wenn neue Algorithmen eingesetzt werden, sollten zudem in einer übergeordneten Evaluation Interaktion und Aggregation der Risiken regelmäßig überprüft werden. Bei der laufenden Validierung, der übergeordneten Evaluation und entsprechenden Anpassungen gelten die in Kapitel 3 aufgestellten Prinzipien entsprechend. Idealerweise sollte eine interne oder externe Revision den regelmäßigen Evaluations- und Anpassungsprozess prüfen. Auf diese Weise werden Funktionalität und Risiken der Algorithmen in der praktischen Anwendung unabhängig beurteilt. Insgesamt lassen sich die Risiken beim Einsatz von Algorithmen reduzieren, indem das Unternehmen eine unabhängige interne oder externe Kontrollfunktion als zusätzliche Instanz hinzuzieht.

Use-Case: Verwaltung von Fonds

Kapitalverwaltungsgesellschaften können Algorithmen bei der Verwaltung von Fonds einsetzen. Investitions- und Desinvestitionsentscheidungen für einen Fonds (kollektive Portfolioverwaltung) trifft dann nicht mehr ausschließlich ein Portfoliomanager. Sie erfolgen vielmehr auf Basis eines für diesen Fonds entwickelten Algorithmus. In dem hier beschriebenen Use-Case erfolgt zum Beispiel die algorithmenbasierte Aktienauswahl nach einem quantitativen Multi-Faktor-Ansatz, der das Risiko-/Ertragsprofil des Fonds optimieren soll.

Dabei wird zunächst mit Hilfe eines quantitativen Modells das Anlageuniversum anhand fundamentaler Parameter, Analysteneinschätzungen und der historischen Wertentwicklung analysiert und festgelegt. Das quantitative Modell wird regelmäßig überprüft und verbessert (spezifisches Prinzip **„Laufende Validierung, übergeordnete Evaluation und entsprechende Anpassung“**). Dies schließt ein, dass neue Kennzahlen hinzugefügt oder bestehende modifiziert werden.

Nachdem das Anlageuniversum festgelegt wurde, wird es einem quantitativen Screening mit einem Multi-Strategie-Modell unterzogen. Dieses Modell analysiert und bewertet Wertpapiere innerhalb des Anlageuniversums auf Basis quantitativer Strategien und berücksichtigt dabei verschiedene Faktoren.

Jede Aktie wird bewertet und innerhalb einer quantitativen Strategie gerankt. Bei der oben genannten **„Interpretation und Verwertung der algorithmischen Ergebnisse zur Entscheidungsfindung“** nimmt der Portfoliomanager die vom Modell am besten bewerteten Einzeltitel aller Strategien in das Fondsportfolio auf und prüft die Zusammensetzung des Portfolios quartalsweise. Zur Abfederung kurzfristiger Markteffekte kann der Portfoliomanager die Faktoren und/oder die Strategiegewichtung anpassen (spezifisches Prinzip **„Intensive Freigabe- und Feedbackprozesse“)**.

V. Einbettung der Prinzipien in internationale Regulierungsvorhaben

Die Prinzipien stellen vorläufige Überlegungen zu aufsichtlichen Mindestanforderungen für den Einsatz von künstlicher Intelligenz dar. Sie sind damit eine Diskussionsgrundlage für den Austausch mit diversen Stakeholdern. Damit sind Marktteilnehmer und die Wissenschaft gemeint, aber vor allem auch andere nationale und internationale Aufsichtsbehörden und Standardsetzer. Die BaFin will in diesem Austausch weiterhin eine aktive Rolle einnehmen und den Diskussionsprozess voranbringen.

Die Europäische Kommission hat in ihrer Digital Finance Strategy[9] erklärt, dass sie bis spätestens 2024 gemeinsam mit den Europäischen Aufsichtsbehörden (European Supervisory Authorities – ESAs) klarstellen möchte, ob und wie bestehende Finanzmarktregulierung bei der Verwendung von BDAI anzuwenden ist. Bereits vor dieser Ankündigung sind bei den ESAs diverse Arbeitsgruppen etabliert worden, die sich mit den Chancen und Risiken von BDAI beschäftigen – beispielsweise die EBA Task Force on IT, die EIOPA InsurTech Task Force und das ESMA Financial Innovation Standing Committee.[10] Das vorliegende Prinzipienpapier soll einen wichtigen Beitrag dazu leisten, die gemeinsamen Arbeiten der ESAs mit der Kommission – auch in den genannten Arbeitsgruppen – voranzubringen. Gleiches gilt für die Arbeiten der globalen Standardsetzer, etwa des Financial Stability Boards (FSB), des Basel Committee on Banking Supervision (BCBS), der International Association of Insurance Supervisors (IAIS) und der International Organization of Securities Commissions (IOSCO).

Darüber hinaus gibt es diverse Arbeiten rund um BDAI bei nationalen und internationalen Standardsetzern wie dem DIN- bzw. dem ISO-Gremium. Da künstliche Intelligenz und maschinelles Lernen noch nicht trennscharf definiert sind, sind von diesen Gremien weitere Arbeiten zu erwarten. Ferner sind weitere (über die Finanzmarktregulierung hinausgehende) technische Standards für die Anwendung von BDAI voraussehbar. Auch diese Entwicklungen wird die BaFin beobachten und, wann immer dies geboten ist, aktiv begleiten.

[9] Europäische Kommission (2020): MITTEILUNG DER KOMMISSION AN DAS EUROPÄISCHE PARLAMENT, DEN RAT, DEN EUROPÄISCHEN WIRTSCHAFTS- UND SOZIALAUSSCHUSS UND DEN AUSSCHUSS DER REGIONEN über eine Strategie für ein digitales Finanzwesen in der EU, URL: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52020DC0591
[10] European Banking Authority, European Insurance and Occupational Pensions Authority, European Securities and Markets Authority.