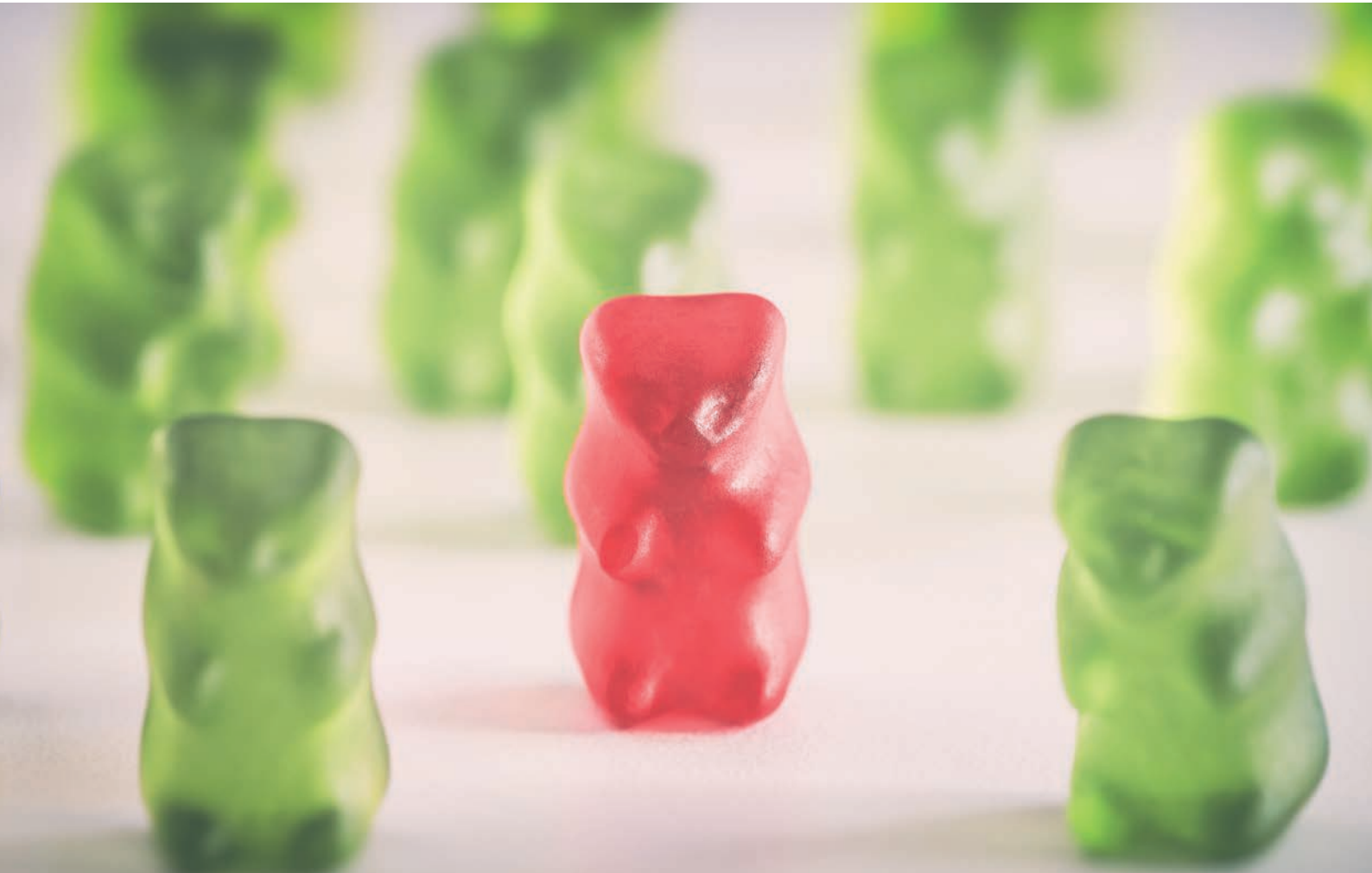


Machine Learning im Compliance-Umfeld

Die False Positives effizient reduzieren

Im Compliance-Bereich kann Machine Learning dazu beitragen, die Anzahl sog. False-Positive-Meldungen deutlich zu verringern und damit für eine Entlastung und höhere Produktivität der Compliance-Mitarbeiter sorgen. Die Autoren beschreiben konkrete ML-Ansätze, die für den Anwendungsfall im Bereich der Namensprüfung relevant sind.



In den vergangenen zwei Jahrzehnten hat die Compliance-Funktion in Banken an Bedeutung deutlich gewonnen. Dies ist nicht zuletzt vor dem Hintergrund der zahlreichen Geldwäscheskandale und Embargo- bzw. Sanktionsregimes zu verstehen. Insbesondere im Hinblick auf die Prävention von Geldwäsche und Terrorismusfinanzierung haben sich inzwischen auch quantitativ orientierte Herangehensweisen etabliert, insbesondere auch der risikobasierte Ansatz, der bereits 2007 von der Financial Action Task Force (FATF) empfohlen und 2014 für den Bankensektor konkretisiert wurde.

Die zunehmend technisch-quantitative Ausrichtung der Prozesse rund um die Geldwäscheprävention erlaubt damit auch den Einsatz fortgeschrittener Methoden und Analysemöglichkeiten wie etwa Machine Learning (ML). Ein einfaches Beispiel dafür ist der Einsatz von ML-Methoden im Kontext der durch das Geldwäschegesetz geforderten Kundenüberprüfung.

Die Identifikation und laufende Überwachung von Neu- und Bestandskunden im Rahmen eines sogenannten KYC (Know Your Customer/Client)-Prozesses ist ein zentraler Bestandteil der Anforderungen aus dem Geldwäschegesetz. Ein Bestandteil des KYC-Prozesses ist die Namensprüfung (Name Matching) gegen verschiedene Listen, etwa Sanktionslisten, Embargolisten, PEP-Listen (PEP bezeichnet politische exponierte Personen) und ggf. institutsspezifische Black Lists.

Beim Name Matching kommt es häufig zu einer großen Anzahl von „falschen Treffern“, sogenannten False Positives. Da durch die entsprechenden Mitarbeiter einer Compliance-Abteilung sämtlich Treffer eines Prüflaufs zu bearbeiten sind, führen diese False Positives zu erhöhten Aufwänden und minimieren die verfügbare Zeit, um die „echten Treffer“ zu analysieren.

Mithilfe von ML ist es zum Beispiel möglich, die Anzahl der False Positives deutlich zu verringern und damit für eine Entlastung und höhere Produktivität der Compliance-Mitarbeiter zu sorgen. Wie dies konkret erreicht werden kann, ist Inhalt der folgenden Abschnitte.

Überblick Machine Learning

Künstliche Intelligenz (KI) ist ein etabliertes Forschungsgebiet mit ersten Arbeiten zu künstlichen neuronalen Netzen in den 1940er-Jahren. Als Geburtsstunde der Künstlichen Intelligenz wird allerdings oft das „Summer Research Project on Artificial Intelligence“ angesehen, das 1956 am Dartmouth College in Hanover stattfand. Das aktuell große Interesse an den Konzepten und Methoden der KI kann durch folgende Entwicklungen der letzten Jahre begründet werden:

- KI-Anwendungen profitieren durch eine inzwischen große Anzahl kostenlos verfügbarer (Open Source-) Toolkits und Bibliotheken,
- Speicherkapazität und Rechenleistung moderner Computer und Cloud-Anbieter ermöglichen die performante Implementierung von KI-Methoden,
- die große Menge und Verfügbarkeit von Daten erlaubt die effiziente Anwendung von KI-Ansätzen, zum Beispiel zum Training von künstlichen neuronalen Netzen.

Der Begriff „Maschinelles Lernen“ (ML) als ein Teilgebiet der KI beschreibt Methoden, die mithilfe von Lernprozessen Zusammenhänge in Datensätzen erkennen, um darauf aufbauend Vorhersagen zu treffen¹. Dabei lassen sich drei verschiedene ML-Ansätze unterscheiden:

1. Unsupervised Learning
2. Supervised Learning
3. Reinforcement Learning

Im Kontext des Unsupervised Learning wird versucht, Muster in bestehenden Datensätzen zu erkennen und daraus Kategorien abzuleiten. Die Mustererkennung wird dabei nicht vorgegeben, sondern der Algorithmus nimmt eigenständig eine Kategorisierung bzw. Clusterung der Datensätze vor. Prominente Algorithmen sind der K-Means-Algorithmus und die Latent-Dirichlet Analyse.

Im Rahmen des Supervised Learnings werden Algorithmen anhand kategorisierter Datensätze trainiert. Der Trainingserfolg wird mithilfe eines Testdatensatzes überprüft, um die Güte des trainierten Modells / Algorithmus beurteilen zu können. Das eigentliche Lernen erfolgt auf dem Trainingsdatensatz, während die Beurteilung des trainierten Modells mit einem Testdatensatz durchgeführt wird.

Reinforcement Learning orientiert sich am menschlichen Lernverhalten. Ein Agent erlernt selbstständig eine Strategie, um eine Belohnung / einen Gewinn zu maximieren. Hierzu werden meistens Temporal-Difference-Learning-Algorithmen eingesetzt, die als Q-Learning Methoden bekannt sind. Q beschreibt in dieser Methode den Nutzen als Funktion eines Zustands und einer Aktion.

Für den später diskutierten Anwendungsfall „Name Matching Customer“ (NMC) ist vor allem der Ansatz des überwachten Lernens von Bedeutung. Konkret werden wir anhand von Beobachtungen einen Random-Forest-Algorithmus trainieren und einsetzen.

Random Forests and Decision Trees

Random Forests können als ein Ensemble von Entscheidungsbäumen verstanden werden. Im Folgenden betrachten wir deshalb zunächst das Konzept der Entscheidungsbäume.

Entscheidungsbäume finden Anwendung in Regressions- und Klassifikationsproblemen. Wir beschränken uns in diesem Unterabschnitt auf die Erklärung von Klassifizierungsbäumen. Das Ziel von Entscheidungsbäumen (Decision Trees) ist es, eine existierende Datenmenge mittels hierarchischer Entscheidungen zu gruppieren bzw. zu unterteilen.

Der einfachste Entscheidungsbaum besteht aus einem Knoten und zwei Blättern. Der Knoten enthält eine logische binäre Regel, die eine Zuordnung der Daten, auf die der Entscheidungsbaum angewendet wird, eindeutig einem der beiden Blätter zuweist. Ein Blatt eines Entscheidungsbaumes ist daher als Antwort auf die vorangegangene Entscheidung zu verstehen. Die Abbildung » 1 zeigt exemplarisch ein Ensemble von Entscheidungsbäumen T_i , die aus Unterteilmengen eines Datensets erzeugt wurden. Farblich hervorgehoben sind die Ergebnisse der binären Entscheidungen.

Im vorliegenden Anwendungsfall liegt die Herausforderung in der Bestimmung von geeigneten Attributen, die eine Klassifizierung durch eine Entscheidungsregel möglich machen. Vielfach ist die explizite Vorgabe einer Entscheidungsregel sehr schwierig, daher verwendet man statistische Algorithmen. Einer der bekanntesten Algorithmen ist ID3 (Iterative Dichotomiser 3) und dessen Weiterentwicklung C4.5²

Die Kernidee des Algorithmus ist die Auswahl eines Attributs α anhand des Informationsgehalts. Der Informationsgehalt (information gain) $IG(\mathbf{M}, \alpha)$ eines Attributs ist die Differenz der Entropie $s(\mathbf{M})$ der zugrunde liegenden Datenmenge \mathbf{M} und der mittleren Entropie $s(\mathbf{M}|\alpha)$ für die fixierte Auswahl des Attributs α . Mit jeder weiteren Auswahl eines Attributs wird der Entscheidungsbaum vergrößert.

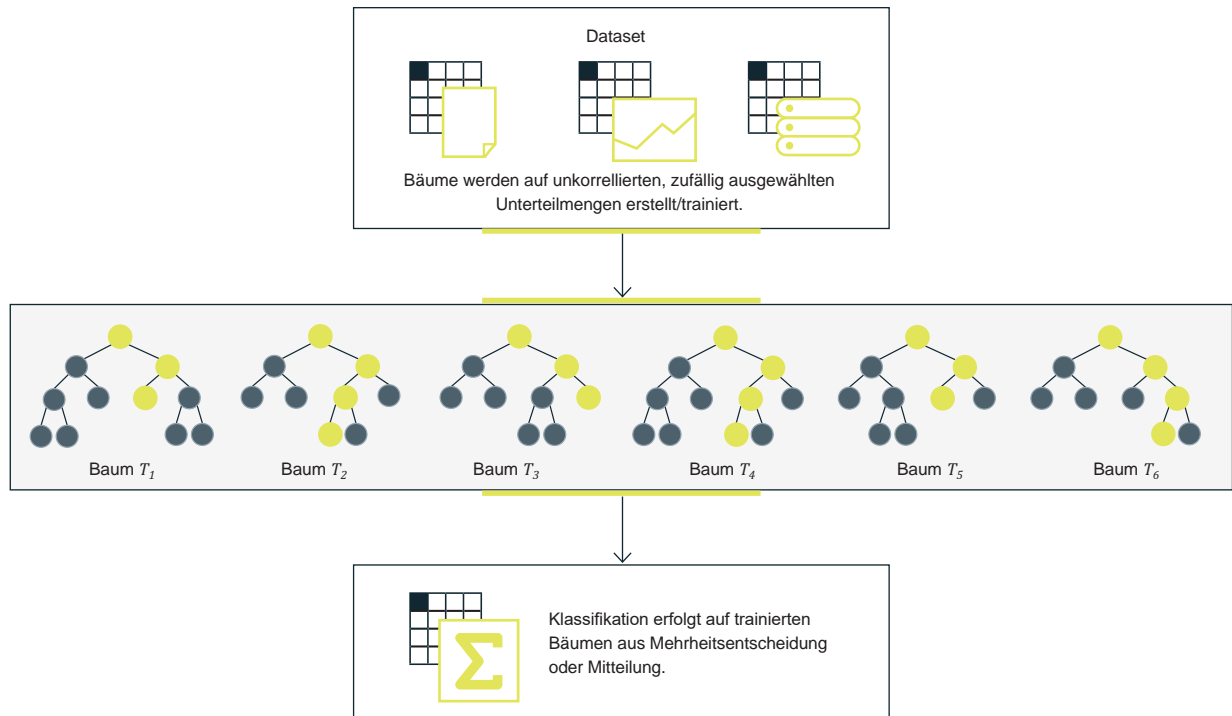
Es gibt auch andere statistische Verfahren, die sich allerdings nur nachrangig für das vorliegende Anwendungsbeispiel NMC eignen. Diese Verfahren basieren auf der Quadratsumme der Residuen (Residual Sum Of Squares, RSS). Eine binäre Entscheidung, den Datensatz \mathbf{M}_i aus der Menge \mathbf{M} aller Daten an der Stelle c in die Blätter $B_1(i, x) = \{\mathbf{M}|\mathbf{M}_i < c\}$ und $B_2(i, x) = \{\mathbf{M}|\mathbf{M}_i > c\}$ zu unterteilen, wird optimiert, indem die Summe der Residuen aus den beiden Blättern für die Anzahl aller Datensätze i und Stelle c minimiert wird. Diese Strategie kann rekursiv auf jede neu entstandene Unterteilmenge angewendet werden, sodass sich eine Baumstruktur ausbildet³.

Die Tiefe von Entscheidungsbäumen und damit einhergehend auch der Detailgrad der Entscheidungen kann limitiert werden, indem eine zulässige Untergrenze für die Zuordnung zu einer finalen Unterkategorie getroffen wird. Das Optimum dieser Untergrenze eines Entscheidungsbaums wird über sog. Pruning-Verfahren⁴ bestimmt. Diese Verfahren wurden entwickelt, um Entscheidungsbäume zu erzeugen, die nicht übermäßig stark auf den genutzten Trainings-Datensatz angepasst (overfitted) sind. Daher steigt bei der Anwendung von Pruning-Verfahren die Treffsicherheit einer richtigen Zuordnung, da die Komplexität reduziert und der Entscheidungsbaum vereinfacht wird.

Die Vorteile von Entscheidungsbäumen: Sie sind einfach zu visualisieren, schnell verständlich und können an menschliche

01 | Exemplarische Darstellung eines Ensembles zufällig erzeugter Klassifizierungsbäume

Farblich hervorgehoben ist die Aggregationslogik, in der per Mehrheitsentscheidung / Mittelung aus dem Ergebnis einzelner Entscheidungsbäume eine Klassifikation getroffen wird.



Quelle: d-fine GmbH.

Entscheidungen geknüpft werden, ohne dass mathematisches Experten-Wissen notwendig ist. Sie stellen eine einfache Möglichkeit dar, hohe, intransparente granulare Datenmengen logisch und nachvollziehbar zu strukturieren, sodass aus einer granularen, quantitativen Ebene eine einfacher zugängliche qualitative Entscheidungsebene erzeugt wird.

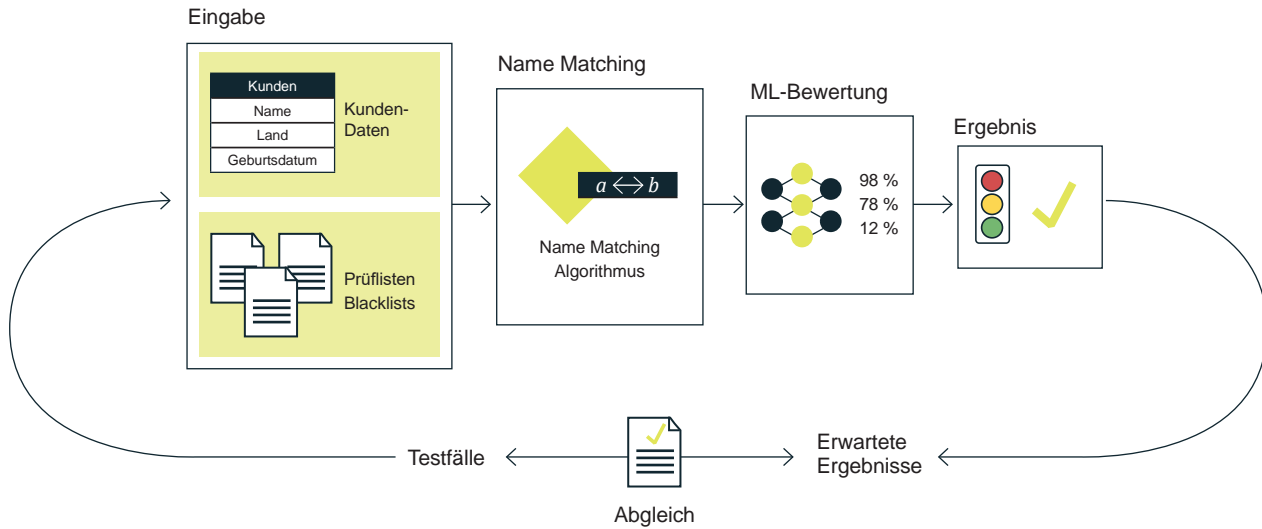
Klassische Entscheidungsbäume stoßen jedoch oft an Grenzen, insbesondere, wenn die für die Knoten erzeugten Regeln sehr sensitiv gegenüber den verwendeten Inputdaten sind, leidet die Stabilität der Vorhersagegenauigkeit. Eine Entscheidung hängt stark von der Verteilung der Inputdaten ab. Ändert sich die Verteilung, so kann der Baum instabil werden, da eine nachträgliche Korrektur der Hierarchie des Baums im Regelfall nicht möglich ist, ohne den ganzen Baum neu zu erzeugen.

Eine Möglichkeit, die Stabilität einer Klassifikation zu erhöhen, ist, verschiedene Modelle, in diesem Fall Entscheidungs-

bäume, zu kombinieren, bzw. zu mitteln (Bootstrap Aggregation bzw. Bagging). Hierbei liefert jedes Modell ein Ergebnis für eine Stichprobe, bzw. Unterteilmenge aus der Gesamtheit aller Daten. Die einzelnen Ergebnisse können zusätzlich, zum Beispiel anhand der Größe einer Stichprobe, gewichtet werden. Dies erzeugt wesentlich stabilere Vorhersagen in der Gesamtklassifikation, da die Mittelung über die einzelnen Ergebnisse die Varianz verringert. Datenanomalien und Verteilungen können jedoch zu einer zu speziellen Kategorisierung durch einen einzelnen Entscheidungsbaum führen. Diese spezielle Kategorisierung funktioniert dann möglicherweise für die gegebene Datenbasis, kann aber eventuell bei neu dazu kommenden Datensätzen versagen.

Diese Einschränkung kann mit der Verwendung von **Random-Forest-Modellen** kontrolliert werden. Random Forests greifen die Idee der Kombination von mehreren Entschei-

02 | Name Matching ergänzt um die Bewertung mit Machine Learning



Quelle: ACTICO GmbH.

Entscheidungsbäumen auf. Wichtig ist jedoch, dass die Entscheidungsbäume untereinander nicht korreliert sein sollen. Einzelne Entscheidungsbäume werden daher auf der Basis zufällig gewählter Untermengen der gesamten Datenmenge erstellt. Eine zufällig ausgewählte Stichprobe der ursprünglichen Daten hat den Vorteil, dass aus den Daten nicht die am stärksten vorhandene Kategorie herangezogen wird, sondern auch kleinere Kategorien in der Stichprobe stärker vertreten sein können und so in die Klassifikation stärker mit einbezogen werden können.³

Die Güte von Random Forests und Entscheidungsbäumen kann mit dem Out-Of-Bag-Fehler (OOB-Error) beschrieben werden. Hierbei wird der Anteil der Daten genutzt, der für die Erstellung des Entscheidungsbaums nicht berücksichtigt ist, um einen Fehler für die Vorhersage einer richtigen Klassifikation zu treffen.

Eine Standard-Methode, um die Qualität eines Entscheidungsbaums zu beschreiben, ist die Verwendung einer Entropiefunktion $s = \sum_{j=1}^M p_j \log(p_j)$.

Hier ist p_j die Wahrscheinlichkeit, mit der ein Datensatz M_j einer Klassifikation j zugeordnet ist. Die Entropie ist minimal, wenn alle Daten in einer Klasse zusammenfallen. Die Entropie eines binären Baums, bestehend aus einer Wurzel und zwei

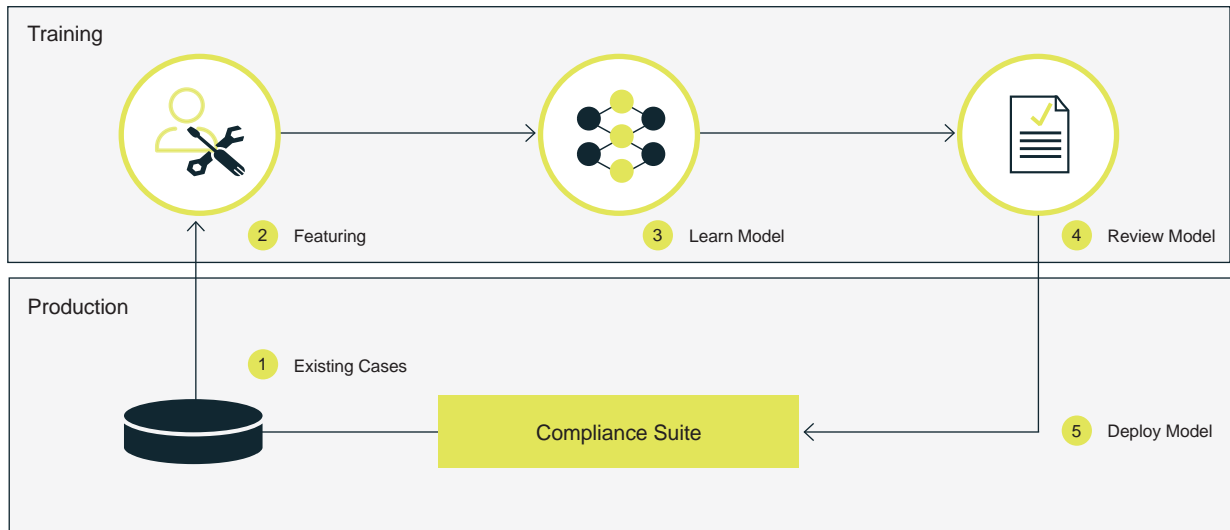
Blättern, ist maximal, wenn die zu klassifizierenden Daten zu gleichen Teilen auf die beiden Blätter entfallen.

Ein alternativer Ansatz zur Bestimmung der Qualität bzw. der Unreinheit (Impurity) eines Entscheidungsbaums ist der Gini-Koeffizient. In der Praxis liefern Gini-Koeffizient und Entropie üblicherweise sehr ähnliche Ergebnisse, sodass es in der Regel ausreichend ist, sich auf ein Impurity-Kriterium zu beschränken.

Im Compliance-Umfeld können sich die zugrunde liegenden Daten schnell ändern. So können Kundendaten und Prüflisten, auf denen beispielsweise kriminelle, prominente oder politisch exponierte Personen geführt sind, in einem bestehenden Modell zu sehr guten Ergebnissen führen. Bei einer Aktualisierung der Datengrundlage besteht allerdings die Gefahr, dass ein ursprünglicher Entscheidungsbaum nicht mehr zum gewünschten Ergebnis einer sinnvollen Klassifikation führt.

Daher eignen sich Random Forests besonders, um im Compliance-Umfeld angewendet zu werden. Sie bilden ein wichtiges Werkzeug, um statistische Korrelationen zwischen Datensätzen zu erzeugen und hierdurch in standardisierten Prozessen in der Überprüfung von Neu- und Bestandskunden zu unterstützen. Zufällig erzeugte Entscheidungs-Hierarchien behalten immer eine Ungenauigkeit und können auch logisch falsche Korrelationen nutzen.

03 | Erstellung eines Modells mithilfe von Machine Learning



Quelle: ACTICO GmbH.

Anwendungsbeispiel

In diesem Kapitel wird gezeigt, wie Machine Learning im Bereich Compliance angewendet werden kann. Eine bekannte Compliance Suite vergleicht Kundendaten gegen Prüflisten. Auf den Listen werden beispielsweise Kriminelle und Terroristen geführt, aber auch Personen mit politischem Einfluss (PEPs). Der Vergleich erfolgt für potenzielle Neukunden vor der Eröffnung einer Geschäftsbeziehung im Rahmen des KYC- bzw. Client Due Diligence (CDD)-Prozesses. Für Bestandskunden findet der Vergleich nach relevanten Änderungen an den Kundenstammdaten oder an den gelisteten Personen regelmäßig statt.

Der Vergleichs-Algorithmus nutzt Namen, Länder (Domicile, Nationalitäten) und Geburtsdaten, um mögliche Übereinstimmungen zu finden. Dabei findet bei Namen auch ein unscharfer Vergleich (ähnlicher Name) statt. Wird eine mögliche Übereinstimmung festgestellt, dann wird diese in der Software abgeklärt, indem ein Bearbeiter dokumentiert, ob diese eine tatsächliche Übereinstimmung ist. » 2

Für diese Software wurde der Vergleichs-Algorithmus so optimiert, dass möglichst alle tatsächlichen Übereinstimmungen (True Positives) gefunden, aber trotzdem möglichst wenig nicht-übereinstimmende Meldungen (False Positives)

erzeugt werden. Aktuell wird diese Optimierung durch Machine Learning weiter verbessert. Nach dem Vergleich mit dem Algorithmus findet eine automatische Bewertung mit einem gelernten Modell statt. Dieses sagt voraus, wie wahrscheinlich eine mögliche Übereinstimmung auch in der Abklärung als tatsächliche Übereinstimmung dokumentiert werden wird. Dies erlaubt es, die möglichen Übereinstimmungen priorisiert abzuklären.

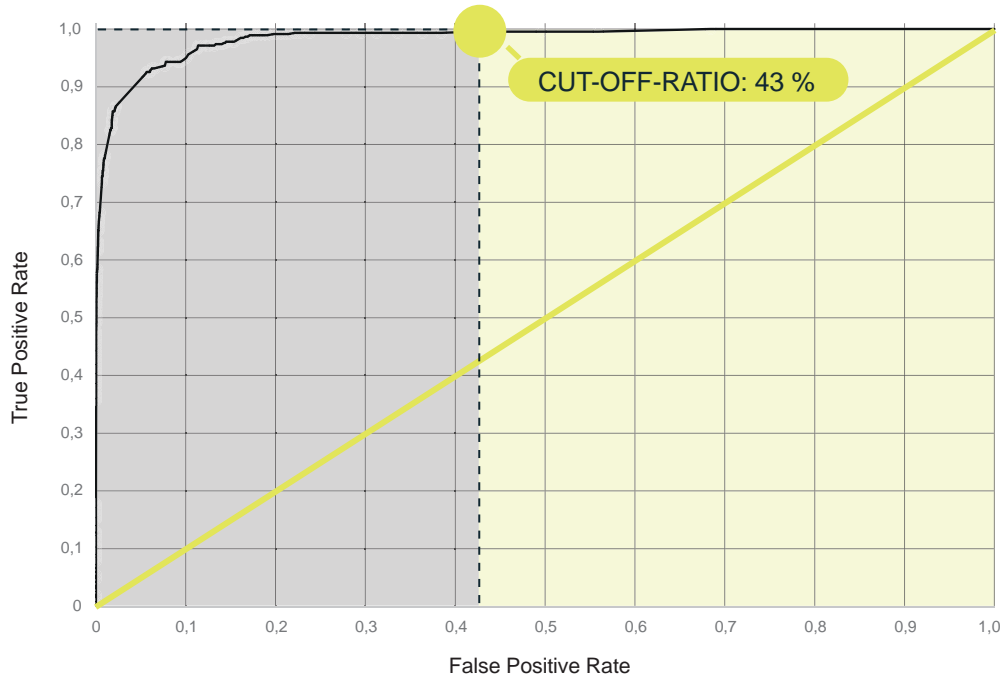
Bei der Abklärung wird der Bearbeiter mit einer Übersicht der gefundenen Übereinstimmungen und Ähnlichkeiten unterstützt. Diese zeigt ihm an, welche Daten wie präzise mit einem Eintrag in der Prüfliste übereinstimmen.

Diese Darstellung benutzt der Bearbeiter dann – auch mithilfe zusätzlicher Recherchen, um zu entscheiden, ob es sich um eine tatsächliche Übereinstimmung handelt.

Lernen eines Modells für NMC

Im Fall von Name Matching haben die Anwender bereits zahlreiche mögliche Übereinstimmungen abgeklärt. Das Ergebnis ist mit den anderen Daten zum Fall in der Datenbank des Systems dokumentiert. Die Abbildung » 3 zeigt, wie aus den Eigenschaften bestehender Fälle ein Modell gelernt und eingeführt werden kann:

04 | Grenzwertoptimierung für ein gelerntes Modell



Quelle: ACTICO GmbH.

Aus den in der Datenbank (1) enthaltenen Fällen mit Ergebnis werden im Feature Engineering (2) die Eigenschaften (Features) und das erwartete Ergebnis (Label) extrahiert. Die extrahierten Daten werden in Trainingsdaten und Testdaten aufgeteilt. Mit Verfahren des überwachten Lernens werden dann Modelle aus den Trainingsdaten erstellt (3). Die Modelle werden mit den Testdaten überprüft (4). Nach dem Review können Modelle in den Betrieb übernommen (5) werden.

Die für das ML-Modell verwendeten Features sind:

- Die Information, welche Vergleiche durch den Algorithmus durchgeführt wurden, z. B. der Vergleich des Nachnamens des Kunden mit den Nachnamen auf der Liste oder der Vergleich des Vor- und Nachnamens des Kunden mit einem Alias auf der Liste.
- Die Information, mit welchem Ergebnis der Vergleich durchgeführt wurde, zum Beispiel Gleichheit, Ähnlichkeit, etc.
- Nicht als Features enthalten sind die tatsächlichen Daten zum Kunden und zum Listeneintrag. Dies geschieht zum

Datenschutz, damit die Features keine Informationen enthalten, mit denen der Kunde identifiziert werden könnte.

Bewertung des gelernten Modells

Die Bewertung eines Modells kann unterschiedlich erfolgen. Wird wie hier eine Unterteilung in zwei Klassen vorgenommen, so kann dies als Grenzwertoptimierungskurve (Receiver Operating Characteristic, ROC) dargestellt werden. » 4

Je stärker die blaue Kurve im Diagramm von der Diagonalen nach oben links abweicht, umso besser konnten die Fälle klassifiziert werden. Werden in diesem Beispiel die Fälle priorisiert nach der von ML ermittelten Wahrscheinlichkeit für die Übereinstimmung bearbeitet, dann wird nach ca. 40 Prozent der Fälle keine weitere tatsächliche Übereinstimmung mehr gefunden.

In der Praxis wurden Modelle zunächst mit den Daten von sechs Kunden erstellt. In allen Fällen standen mehr als 25.000 Datensätze zur Verfügung. Dabei wurden mit Random-Forests-Modelle erstellt, die die Fälle gut klassifizieren. Es konnten in

der Regel 30 bis 40 Prozent der Abklärungen zuverlässig als nicht-übereinstimmend eingestuft werden. Im Einzelfall (wie oben) auch mehr.

Zusammenfassung

Die Anwendung von Machine-Learning-Techniken in der Finanzindustrie, insbesondere auch im Risikocontrolling und in der Compliance-Abteilung von Banken, umfasst inzwischen ein breites Spektrum an Einsatzgebieten. Durch den Einsatz von ML kann hier eine Klassifizierung bzw. Priorisierung von Treffern in der Namensprüfung erreicht werden. Durch die Identifizierung von False Positives und den Ausschluss dieser nicht-relevanten Treffer aus der Folgebearbeitung kann der Aufwand bei der Trefferanalyse deutlich reduziert werden.

Der in diesem Artikel behandelte NMC-Prozess ist nur ein möglicher Anwendungsfall von KI- bzw. ML-Ansätzen im Compliance-Umfeld. So gibt es auch Ansätze, das Reputationsrisiko zu kontrollieren, indem Compliance-Risiken durch Analyse unstrukturierter Kommunikationsdaten gehandhabt werden. Die Analyse von Kommunikationsdaten mit ML-Methoden kann aber auch auf wesentlich komplexere Szenarien, wie der Betrugsprävention und der Vermeidung von Insider-Handel, angewendet werden. Über Compliance-spezifische Themenfelder hinaus gibt es auch Anwendungsfälle, die sich mit der Kreditrisikoüberwachung beschäftigen⁵ [DobrikovGraf2017].

Autoren



Dr. Ulrich Lechner ist Manager bei der d-fine GmbH. Der Diplom-Physiker zeichnet als Co-Head des Competence Center Compliance u. a. auch für Projekte zur Prävention von Geldwäsche und Terrorismusfinanzierung verantwortlich.



Dr. Marcel Langenberg ist theoretischer Physiker im gleichen Unternehmen. Sein Schwerpunkt liegt in der Entwicklung von kundenspezifischen Machine-Learning-Anwendungen.



Thomas Ohlemacher ist Produktmanager bei ACTICO GmbH. Der Diplominformatiker hat seinen Fokus auf Software für Compliance-Themen bei Privat- und Investment-Banken gelegt.

- 1 Vgl. K.P. Murphy: „Machine learning: a probabilistic perspective“ MIT Press, 2012.
- 2 Vgl. J.R. Quinlan: „Induction of Decision Trees. Mach. Learn. 1, 1986 sowie „C4.5: Programs for Machine Learning“, 1993.
- 3 Vgl. G. James, D. Witten, T. Hasti, and R. Tibshirani: „An Introduction to Statistical Learning“, 2017.
- 4 Vgl. L. A. Breslow and D. W. Aha: „Simplifying Decision Trees: A Survey“, The Knowledge Engineering Review, Vol 12, 1997.
- 5 Vgl. T. Dobrikov und F. Graf: „Nachrichten in Frühwarnsystemen und dem Kreditrisikomanagement“, Zeitschrift für das gesamte Kreditwesen, Heft 09/2017, sowie T. Dobrikov, F. Graf, S. Stadelmann, S. Ulsamer: „Kontrolle des Reputationsrisikos: Management von Compliance-Risiken durch Analyse unstrukturierter Kommunikationsdaten“, FIRM Jahrbuch 2019.