actico
Smarter Decisions

d-fine

Whitepaper

Machine Learning in Compliance: So optimieren Banken und Finanzdienstleister das Screening von Kunden und PEPs im KYC-Prozess.

INHALT

_01 Einleitung

In den vergangenen zwei Jahrzehnten hat die Compliance-Funktion in Banken an Bedeutung deutlich hinzugewonnen. Dies ist nicht zuletzt vor dem Hintergrund der zahlreichen Geldwäscheskandale und Embargo- bzw. Sanktionsregimes zu verstehen.

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.

TECHNISCH-QUANTITATIVE AUSRICHTUNG BEIM KUNDENSCREENING

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.

PRÜFUNG GEGEN SANKTIONS- UND PEP-LISTEN IM KYC-PROZESS

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 politisch exponierte Personen) und ggf. institutsspezifische Black Lists.

ZEIT FÜR ECHTE TREFFER GEWINNEN, FALSE POSITIVES REDUZIEREN

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

HÖHERE PRODUKTIVITÄT DER MITARBEITENDEN

Mit Hilfe 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.

DAS ERWARTET SIE IM WHITEPAPER

Abschnitt 2 gibt einen knappen Überblick zum Thema ML, bevor in Abschnitt 3 die konkreten Ansätze, die für den Anwendungsfall im Name Matching relevant sind, in kurzer Form vorgestellt werden. Anschließend werden der konkrete Anwendungsfall und die Erkenntnisse aus dem damit verknüpften Projekt dargestellt. Eine kurze Zusammenfassung bildet den Abschluss.

_02 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, USA, 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.

MACHINE LEARNING ALS TEILGEBIET DER KÜNSTLICHEN INTELLIGENZ

Der Begriff „Maschinelles Lernen“ (ML) als ein Teilgebiet der KI beschreibt Methoden, die mit Hilfe von Lernprozessen Zusammenhänge in Datensätzen erkennen, um darauf aufbauend Vorhersagen zu treffen [Murphy2012]. Dabei lassen sich drei verschiedene ML-Ansätze unterscheiden:

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

UNSUPERVISED 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.

SUPERVISED LEARNING

Im Rahmen des überwachten Lernens werden Algorithmen anhand kategorisierter Datensätze trainiert. Der Trainingserfolg wird mit Hilfe 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

Reinforcement Learning orientiert sich am menschlichen Lernverhalten. Ein Agent erlernt selbstständig eine Strategie, um eine Belohnung/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 in Abschnitt 4 diskutierten Anwendungsfall „Name Matching Customer“ ist vor allem der Ansatz des überwachten Lernens von Bedeutung. Konkret werden wir anhand von Beobachtungen einen Random-Forest-Algorithmus trainieren und einsetzen.

_03 Random Forests und Decision Trees

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

ENTSCHEIDUNGSBÄUME

Entscheidungsbäume finden Anwendung in Regressions- und Klassifikationsproblemen. Da wir uns in diesem Artikel mit den Anwendungsfall „Name Matching Customer“ beschäftigen, beschränken wir uns in diesem Unterabschnitt auf die Erklärung von Klassifizierungsbäumen.

Das Ziel von Entscheidungsbäumen (Decision Trees) ist 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. Abbildung 1 zeigt exemplarisch ein Ensemble von Entscheidungsbäumen T_i die aus Unterteilmengen eines Datasets erzeugt wurden. Farblich hervorgehoben sind die Ergebnisse der binären Entscheidungen.

STATISTISCHE ALGORITHMEN ZUR AUSWAHL VON ATTRIBUTEN ANHAND DES INFORMATIONSGEHALTS

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 C4.5 [Quinlan1986, Quinlan1993].

Die Kernidee des Algorithmus ist die Auswahl eines Attributs a anhand des Informationsgehaltes. Der Informationsgehalt (information gain) $IG(M,a)$ eines Attributes ist die Differenz der Entropie $s(M)$ der zugrundeliegenden Datenmenge M und der mittleren Entropie s(M|a) für die fixierte Auswahl des Attributs a. 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 Name Matching Customer (NMC) eignen. Diese Verfahren basieren auf der Quadratsumme der Residuen (Residual Sum Of Squares, RSS). Eine binäre Entscheidung den Datensatz M_i aus der Menge M aller Daten an der Stelle c in die Blätter $B_1\ (i,x)=\{M|M_i<c\}$ und $B_2\ (i,x)=\{M|M_i>c\}$ 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 ausbildet [JamesWitten2017].

TREFFSICHERHEIT DURCH REDUZIERTE KOMPLEXITÄT DES ENTSCHEIDUNGSBAUMS

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 Entscheidungsbaumes wird über sog. Pruning-Verfahren [BreslowAha97] 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.

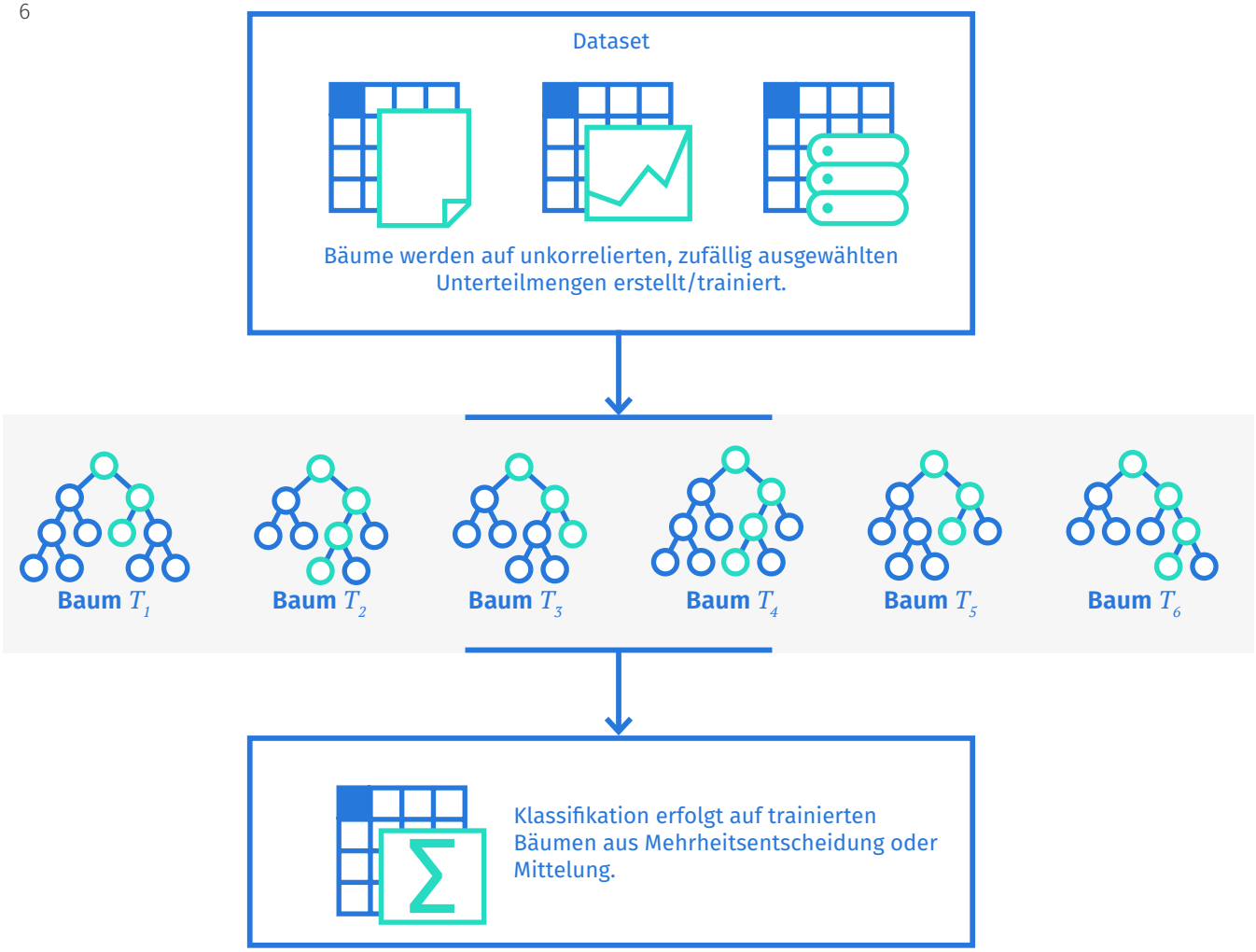


Abb. 1 Exemplarische Darstellung eines Ensembles zufällig erzeugter Klassifizierungsbäume. Farblich hervorgehoben ist die Aggregationslogik in der per Mehrheitsentscheidung/Mittelung aus dem Ergebnis einzelner Entscheidungsbäume eine Klassifikation getroffen wird.

ENTSCHEIDUNGSBÄUME AN MENSCHLICHE ENTSCHEIDUNGEN KNÜPFEN OHNE MATHEMATISCHES EXPERTEN-WISSEN

Die Vorteile von Entscheidungsbäumen sind: Entscheidungsbäume sind einfach zu visualisieren, schnell verständlich und können an menschliche Entscheidungen geknüpft werden, ohne dass mathematisches Experten-Wissen notwendig ist. Entscheidungsbäume sind eine einfache Möglichkeit 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 Baumes 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 Entscheidungsbäume, 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.

RANDOM FORESTS

Im vorhergehenden Abschnitt wurde der Fall eines einfachen Entscheidungsbaums erklärt. 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, versagt aber möglicherweise bei neu dazu kommenden Datensätzen.

RANDOM-FOREST-MODELLE KOMBINIEREN MEHRERE ENTSCHEIDUNGSBÄUME

Diese Einschränkung kann mit der Verwendung von Random-Forest-Modellen kontrolliert werden. Random Forests greifen die Idee der Kombination von mehreren Entscheidungsbäumen auf, vgl. Abbildung 1. 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 [JamesWitten2017]. Die Güte von Random Forests und Entscheidungsbäumen kann mit dem Out-Of-Bag-Fehler (OOB-Error) beschrieben werden. Hierbei wird der Anteil der Daten, der für die Erstellung des Entscheidungsbaums nicht berücksichtigt ist – Out Of Bag – genutzt, um einen Fehler für die Vorhersage einer richtigen Klassifikation zu treffen.

Eine Standard-Methode, um die Qualität eines Entscheidungsbaumes 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_i einer Klassifikation j zugeordnet ist. Die Entropie ist minimal, wenn alle Daten in einer Klasse zusammenfallen. Die Entropie eines binären Baumes, 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, so dass es in der Regel ausreichend ist, sich auf ein Impurity-Kriterium zu beschränken.

RANDOM FORESTS EIGNEN SICH BESONDERS GUT FÜR COMPLIANCE

Im Compliance-Umfeld können sich die zu Grunde 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 Entscheidungshierarchien behalten immer eine Ungenauigkeit und können auch logisch falsche Korrelationen nutzen.

WIE MITARBEITENDE IN COMPLIANCE DIE KLASSIFIZIERUNG VERBESSERN

ACTICO stellt zu diesem Zweck eine in sich geschlossene Compliance Suite zur Verfügung, die sich vollständig in bestehende Systeme integrieren lässt. Sie gibt der bearbeitenden Person stets die Möglichkeit, eine Klassifizierung im Einzelfall zu überprüfen und mit manuellem Feedback die zukünftige Klassifikation iterativ zu verbessern.

Im nächsten Abschnitt werden wir basierend auf dem bereits theoretischen skizzierten Hintergrund im Detail auf Funktionsumfang und Funktionalität der ACTICO Software *Name Matching Customer* eingehen.

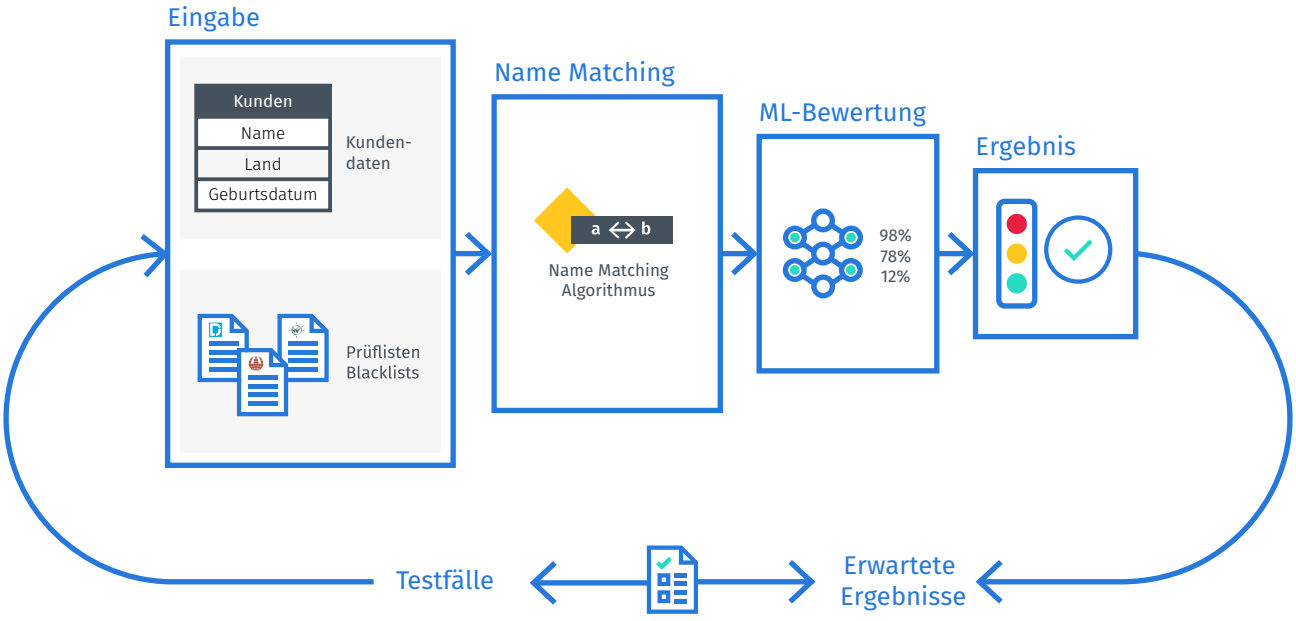
_04 Anwendungsbeispiel für den Abgleich von Kundendaten gegen Prüflisten mit der Unterstützung von Machine Learning

In diesem Kapitel wird dargestellt, wie Machine Learning (ML) im Bereich Compliance angewendet werden kann. Gezeigt wird die Namensprüfung mit dem Softwaremodul *Name Matching Customer (NMC)* der Compliance Suite von ACTICO. Für Machine Learning werden die Werkzeuge des Machine-Learning-Moduls der Compliance Suite eingesetzt. Es unterstützt unter anderem die vorgängig beschriebene Strategie des überwachten Lernens mit Random Forests und auch anderen Algorithmen wie zum Beispiel Deep Learning mit neuronalen Netzen.

SICHERER KYC-PROZESS DURCH SCREENING VON KUNDENDATEN

Das Modul *Name Matching Customer* der Compliance Suite vergleicht Kundendaten gegen Prüflisten. Personen werden aus unterschiedlichen Gründen gelistet. Auf den Listen werden beispielsweise Kriminelle und Terroristen geführt, aber auch Personen mit politischem Einfluss (politisch exponierte Personen, 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 (Domizile, 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 sie eine tatsächliche Übereinstimmung ist.



Die Software überprüft Neu- und Bestandskunden auf Übereinstimmungen mit Einträgen in Sanktions- und PEP-Listen. Machine Learning bewertet die Übereinstimmungen mit Hilfe eines gelernten Modells.

Abb. 2 Name Matching ergänzt um die Bewertung mit Machine Learning.

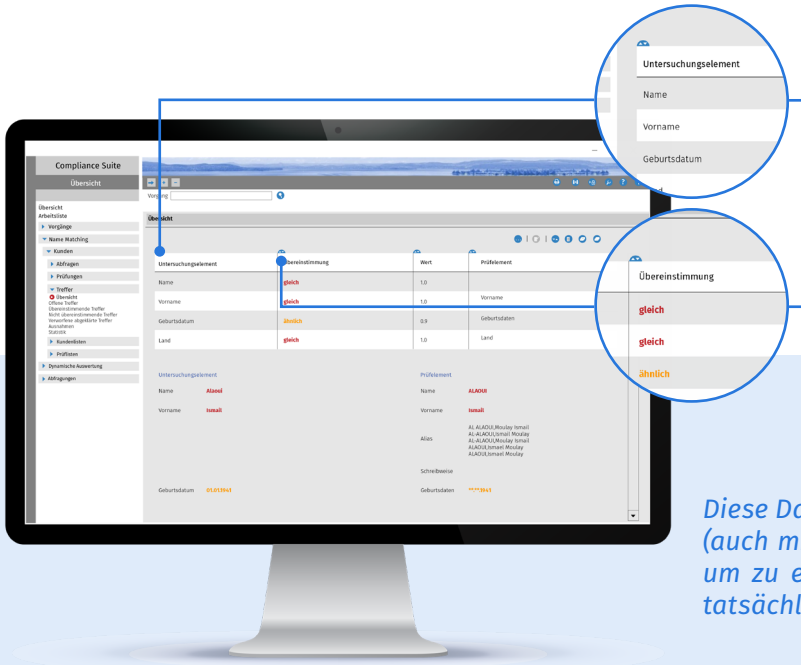
KONZENTRATION AUF TATSÄCHLICHE TREFFER (TRUE POSITIVES)

Die Software ist bei vielen Kunden seit Jahren im Einsatz. Dabei 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.

Trefferwahrscheinlichkeit in %
Qualität der Übereinstimmungsmerkmale zwischen Kundendaten und Sanktionslisteneintrag.

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

Abb. 3 Auflistung der Prüfungsergebnisse für den Bearbeiter, priorisiert nach der Bewertung mit ML.



Hier werden die Untersuchungselemente wie Name, Vorname, Geburtsdatum angezeigt.

Der Bearbeiter sieht, welche Elemente mit Einträgen auf der Sanktionsliste identisch oder ähnlich sind.

Diese Darstellung benutzt der Bearbeiter (auch mit Hilfe zusätzlicher Recherchen), um zu entscheiden, ob es sich um eine tatsächliche Übereinstimmung handelt.

Abb. 4 Darstellung einer möglichen Übereinstimmung zur Abklärung durch den Bearbeiter.

LERNEN EINES MODELLS FÜR DEN NAMENSABGLEICH

Im Fall von Name Matching Customer haben die Anwender der Software bereits zahlreiche mögliche Übereinstimmungen abgeklärt. Das Ergebnis der Abklärung ist mit den anderen Daten zum Fall in der Datenbank des Systems dokumentiert.

Das folgende Schaubild zeigt, wie aus den Eigenschaften bestehender Fälle ein Modell gelernt und eingeführt werden kann:

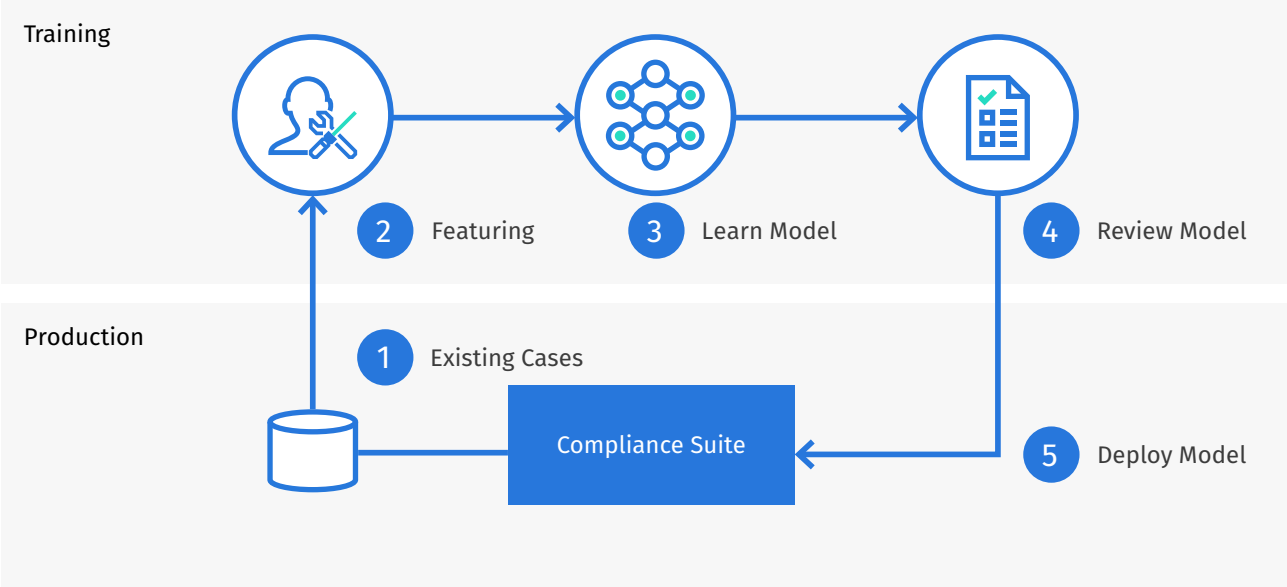


Abb. 5 Erstellung eines Modells mit Hilfe von Machine Learning in der ACTICO Compliance Suite.

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

Die für das ML-Modell verwendeten Features sind:

- Die Information, welche Vergleiche durch den Algorithmus durchgeführt wurden, zum Beispiel 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.

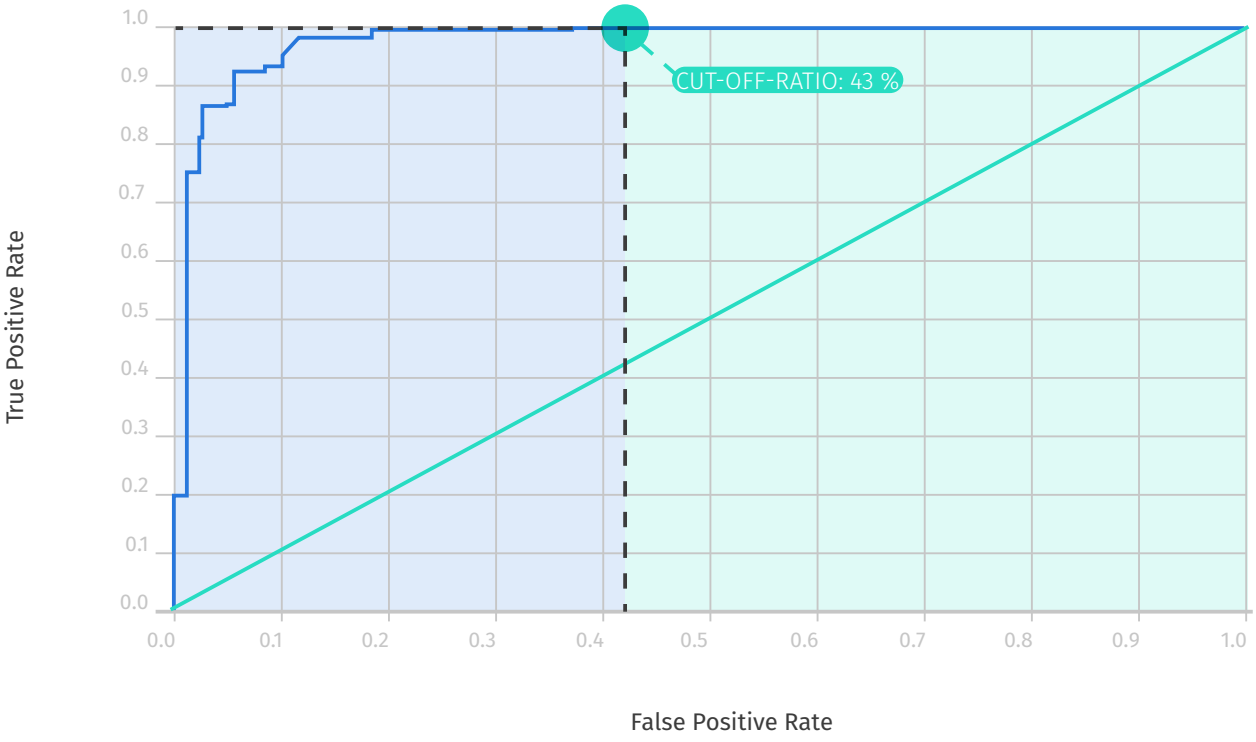


Abb. 6 Compliance-Abteilungen sparen mit Machine Learning rund 57 % ihrer Abklärungen.

Bis zu **57 %** der Abklärungen sparen.

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. 43 % 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 Forest Modelle erstellt, die die Fälle gut klassifizieren. Es konnten in der Regel 30 % bis 40 % der Abklärungen zuverlässig als nicht-übereinstimmend eingestuft werden. Im Einzelfall (wie oben) auch mehr.

_05 Zusammenfassung

Die Anwendung von ML-Techniken in der Finanzindustrie, insbesondere auch im Risikocontrolling und in der Compliance-Abteilung von Banken, umfasst inzwischen ein breites Spektrum an Einsatzgebieten. In diesem Artikel haben wir uns auf das sog. überwachte Lernen mit Random Forests konzentriert und einen darauf basierenden Anwendungsfall aus dem Compliance-Umfeld näher beleuchtet. 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 Prozess beim Kundenscreening 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 die Analyse unstrukturierter Kommunikationsdaten gehandhabt werden [DobrikovGraf2019].

Die Analyse von Kommunikationsdaten mit ML-Methoden kann aber auch auf wesentlich komplexere Szenarien, wie der Betrugsprävention und Vermeidung von Insiderhandel, angewendet werden. Über Compliance-spezifische Themenfelder hinaus gibt es auch Anwendungsfälle, die sich mit der Kreditrisikoüberwachung beschäftigen [DobrikovGraf2017].

_06 Quellen

[Bishop2006] C. Bishop, Pattern Recognition and Machine Learning. Springer, 2006

[BreslowAha97] L. A. Breslow and D. W. Aha, Simplifying Decision Trees: A Survey, The Knowledge Engineering Review, Vol 12 (1), 1997.

[Buxmann2019] P. Buxmann, H.Schmidt (Hrsg.), Künstliche Intelligenz. Springer Gabler, 2019

[Murphy2012] K.P. Murphy, Machine learning: a probabilistic perspective. MIT Press, 2012

[Raschka2018] S. Raschka, V. Mirjalili, Machine Learning mit Python und Scikit-Learn und TensorFlow. mitp, 2018

[DobrikovGraf2017] T. Dobrikov und F. Graf, Nachrichten in Frühwarnsystemen und dem Kreditrisikomanagement, Zeitschrift für das gesamte Kreditwesen, Heft 09/2017

[DobrikovGraf2019] T. Dobrikov, F. Graf, S. Stadelmann, S. Ulsamer, Kontrolle des Reputationsrisikos: Management von Compliance-Risiken durch Analyse unstrukturierter Kommunikationsdaten, FIRM Jahrbuch 2019

[JamesWitten2017] G. James, D. Witten, T. Hasti, and R. Tibshirani, An Introduction to Statistical Learning, Springer New York Heidelberg Dordrecht London, 2017 DOI 10.1007/978-1-4614-7138-7

[Quinlan1986] Quinlan, J. R. 1986. Induction of Decision Trees. Mach. Learn. 1, 1 (Mar. 1986), 81–106

[Quinlan1993] Quinlan, J. R. C4.5: Programs for Machine Learning. Morgan Kaufmann Publishers, 1993.

AUTOREN

d-fine:
Dr. Ulrich Lechner
Dr. Marcel Langenberg

ACTICO:
Thomas Ohlemacher

_07 Abbildungsverzeichnis

d-fine ist ein europäisches Beratungsunternehmen mit Fokus auf analytische und quantitative Herausforderungen und die Entwicklung nachhaltiger technologischer Lösungen. Die Kombination aus hunderten naturwissenschaftlich geprägten Mitarbeitern und langjähriger Praxiserfahrung ermöglicht passgenaue, effiziente und nachhaltige Umsetzungen für unsere mehr als zweihundert Kunden aus allen Wirtschaftsbereichen.

Mehr Information unter: www.d-fine.com

ACTICO ist ein führender internationaler Anbieter von Lösungen für intelligente Automatisierung und digitale Entscheidungsfindung. Seine skalierbare Software kombiniert in einzigartiger Weise Regeltechnologie mit maschinellem Lernen und ist durchgängig revisionssicher. So können Unternehmen aller Größen umfangreiche Datenmengen verarbeiten und KI-gestützte sowie regelbasierte Entscheidungen in Echtzeit treffen und automatisieren. ACTICO steigert durch Intelligente Automatisierung den Business Value seiner Kunden, indem operative Entscheidungen verbessert werden.

ACTICO wurde 1997 gegründet und zählt heute mehr als 100 Unternehmen in 30 Ländern zu seinen Kunden, darunter Volkswagen Financial Services, ING und KfW. Der Hauptsitz befindet sich in Immenstaad am Bodensee (Deutschland), weitere Standorte sind in Chicago (USA) und Singapur.

Mehr Information unter www.actico.de

ACTICO

Europa
ACTICO GmbH
Ziegelei 5
88090 Immenstaad
Germany

Amerika
ACTICO Corp.
141 W. Jackson Blvd.
Ste 300A
Chicago, IL 60604
USA

Asien & Pazifik
ACTICO Pte. Ltd.
#11 - 04, The Arcade
11 Collyer Quay
049317 Singapore

info@actico.com
www.actico.com

actico
Smarter Decisions

02/2020