

MOTRA 06/23

SPOTLIGHT

FEATURE

Künstliche Intelligenz in der Radikalisierungsforschung: Synopse von Grundlagenwissen, Weiterbildungsangeboten und Limitationen

Prof. Dr. Dennis Klinkhammer

Klinkhammer, D. (2023): Künstliche Intelligenz in der Radikalisierungsforschung: Synopse von Grundlagenwissen, Weiterbildungsangeboten und Limitationen. MOTRA-Spotlight 06/23. Hrsg. von: MOTRA-Verbund, Wiesbaden.

<https://doi.org/10.57671/motra-2023006>

AUTOR

Dennis Klinkhammer ist Fachhochschullehrer für empirische Sozialforschung an der FOM Hochschule und lehrt neben den fächerübergreifenden Grundlagen der Statistik insbesondere die quantitativen Datenanalyseverfahren in den Programmiersprachen R und Python.

✉ dennis.klinkhammer@fom.de

Disclaimer

Die im MOTRA-Spotlight veröffentlichten Beiträge spiegeln die Meinungen und Einschätzungen der Verfasserinnen und Verfasser wider.

Künstliche Intelligenz in der Radikalisierungsforschung: Synopse von Grundlagenwissen, Weiterbildungsangeboten und Limitationen

Einleitung

Im Rahmen des Verbundprojektes MOTRA, auf der einschlägigen Konferenz MOTRA-K sowie in der Publikationsreihe MOTRA-Monitor wurden bereits erste Beiträge aus dem Kontext der Radikalisierungsforschung mit methodischem Bezug zur Künstlichen Intelligenz (KI) vorgestellt und diskutiert. Diese Beiträge beschränken sich dabei nicht nur auf den Versuch eines deskriptiven Monitorings von Social Media Daten, sondern versuchen darüber hinaus die Wechselbeziehungen zwischen den möglichen Faktoren der Online- und Offline-Radikalisierung möglichst kausalanalytisch einzufangen und zu diesem Zweck systematisch abzubilden.

Dabei wird insbesondere im MOTRA-Monitor auf unterschiedliche Klassifikationsalgorithmen sowie Large Language Models zur automatisierten Erkennung und Systematisierung von Sprache verwiesen. Zweitere bilden dabei u.a. eine der technologischen Grundlagen von generativer KI, wie sie bspw. ChatGPT von OpenAI verwendet.

Während die einschlägigen Projekte in diesem formalen Rahmen hinsichtlich des methodischen Vorgehens als äußerst versiert bezeichnet werden können, insbesondere in Bezug auf die wissenschaftlichen Gütekriterien Objektivität, Reliabilität und Validität, zeichnet sich oftmals im informalen Rahmen ab, bspw. im Gespräch unter Kolleginnen und Kollegen, dass eine allgemeine Einführung in diese Thematik unter Ausdifferenzierung der verschiedenen Anwendungsmöglichkeiten wünschenswert wäre. Entsprechend widmet sich dieser Beitrag dem Grundlagenwissen zur KI, einschlägigen Weiterbildungsangeboten und Limitationen.

Grundlagenwissen

In diesem Abschnitt werden zunächst zentrale Begriffe aus dem Kontext der KI in Verbindung mit einschlägigen Anwendungsbeispielen aus der Radikalisierungsforschung – insbesondere unter Rückgriff auf Social Media Daten – vorgestellt. Dazu sind in einem ersten Schritt KI als übergeordneter Begriff und die dazugehörigen Techniken und Ansätze des maschinellen Lernens, der neuronalen Netze und des sogenannten Deep Learnings als Teilbereiche auszudifferenzieren (Abbildung 1).

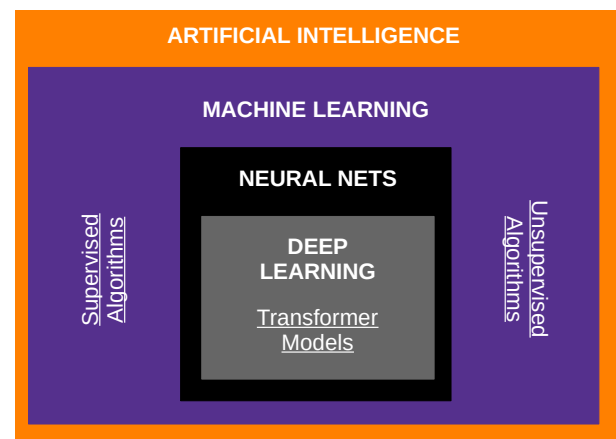


Abb. 1: Teilbereiche der Künstlichen Intelligenz

Künstliche Intelligenz

KI ist ein übergeordneter Begriff, der sich auf die Programmierung von Maschinen / Systemen bezieht, die in der Lage sind, ansatzweise menschenähnliche Intelligenz und Verhaltensweisen zu zeigen. Dabei umfasst KI verschiedene Techniken und Ansätze, von denen maschinelles Lernen, neuronale Netze und Deep Learning wichtige Teilbereiche sind:

Maschinelles Lernen

Das maschinelle Lernen ist ein Teilgebiet der KI, das sich mit der Entwicklung von Algorithmen befasst, die es Computern ermöglichen, aus Daten zu lernen und somit Vorhersagen oder Entscheidungen zu treffen. Es basiert auf statisti-

schen Methoden und Algorithmen, die es einem System ermöglichen, Muster und Strukturen in den Daten zu erkennen und daraus zu lernen. Beim maschinellen Lernen wird zwischen überwachtem und unüberwachtem maschinellen Lernen unterschieden. Dabei handelt es sich um zwei grundlegende Ansätze des maschinellen Lernens, die sich in Bezug auf das Vorhandensein von gelabelten Trainingsdaten unterscheiden und unterschiedliche Ziele verfolgen:

Beim überwachten Lernen werden Algorithmen mit gelabelten Trainingsdaten trainiert. Gelabelt bedeutet, dass jeder Datenpunkt im Trainingsdatensatz mit einer Zielvariablen oder einem „Label“ versehen ist, die bereits bekannt ist. Das Ziel besteht darin, einen Algorithmus zu entwickeln, der auf der Grundlage dieser gelabelten Daten Muster und Beziehungen erkennt und dann in der Lage ist, neue, unbekannte Datenpunkte zu klassifizieren oder vorherzusagen. Ein Beispiel für überwachtes Lernen ist die Klassifikation von Social Media Kommentaren als strafrechtlich relevant oder strafrechtlich nicht relevant. Der Algorithmus wird hierzu mit Social Media Kommentaren trainiert, die bereits als strafrechtlich relevant oder strafrechtlich nicht relevant gekennzeichnet sind. Basierend auf den Merkmalen der Social Media Kommentare lernt der Algorithmus, Muster zu erkennen, und kann dann neue, unbekannte Social Media Kommentare entsprechend klassifizieren.

Im Gegensatz dazu bezieht sich unüberwachtes Lernen auf das Lernen aus ungelabelten Daten. Hier liegen keine vorgegebenen Labels oder Zielvariablen vor. Das Ziel besteht darin, verborgene Strukturen, Muster oder Beziehungen in den Daten zu entdecken, ohne dass der Algorithmus über spezifische Informationen über die Daten verfügt. Ein Beispiel für unüberwachtes Lernen ist die Clusteranalyse. Der Algorithmus analysiert die Daten und sucht nach natürlichen Gruppierungen oder Clustern basierend auf ähnlichen Merkmalen, bspw. dem Verhalten von Nutzerinnen und Nutzern auf Social Media Plattformen. Dabei werden keine vordefinierten Klassen oder Labels ver-

wendet, sondern der Algorithmus erkennt selbstständig die Gruppierungen in den Daten.

Neuronale Netze

Neuronale Netze sind ein spezifischer Ansatz des maschinellen Lernens, der von der Funktionsweise des menschlichen Gehirns inspiriert ist. Sie bestehen aus miteinander verbundenen künstlichen Neuronen oder Knoten, die in Schichten angeordnet sind. Jedes Neuron empfängt Eingaben, führt Berechnungen durch und gibt Ausgaben weiter. Durch das Trainieren eines neuronalen Netzes mit Beispieldaten können neuronale Netze Muster und Zusammenhänge in den Daten erkennen, Vorhersagen treffen oder spezifische Aufgaben ausführen. Während die Algorithmen des maschinellen Lernens unterschiedliche Strukturen aufweisen und verschiedene mathematische Methoden verwenden, um Muster in den Daten zu identifizieren, sind neuronale Netze aufgrund ihrer im Vergleich zu einem Algorithmus flexibleren Architektur und der Verwendung von mehreren Schichten von Neuronen in der Lage, komplexere Probleme zu bewältigen.

Dabei gilt, dass Algorithmen des maschinellen Lernens vielfältige Modelle und Ansätze umfassen, während neuronale Netze eine spezielle Art von Modell sind, deren Flexibilität und Komplexität zu leistungstärkeren Modellen führen können. Dies hängt neben der Architektur eines neuronalen Netzes mit dessen Trainingsprozess zusammen. Ein Trainingsprozess in neuronalen Netzen wird u.a. als Backpropagation bezeichnet. Dabei wird der Fehler zwischen den vorhergesagten Ausgaben des neuronalen Netzes und den tatsächlichen Ausgaben verwendet, um die Gewichte sukzessive anzupassen und die Abweichungen zwischen diesen zu minimieren. Durch wiederholtes Iterieren über die Trainingsdaten passt sich das neuronale Netz an und verbessert somit seine Fähigkeit, genaue Vorhersagen zu treffen. Insbesondere vor dem Kontext oftmals umfangreicher und dynamischer Social Media Daten können neuronale Netze die Komplexität des menschlichen Verhaltens der Nutzerinnen und Nutzer auf Social Media Plattformen zielführender analysieren.

Deep Learning

Deep Learning ist eine fortgeschrittene Methode des maschinellen Lernens, die auf mehrschichtigen neuronalen Netzen basiert. Diese ermöglichen es einem Modell, tiefe Hierarchien von Merkmalen oder Abstraktionen zu lernen, indem mehrere Schichten von Neuronen verwendet werden. Im Vergleich zu herkömmlichen Methoden des maschinellen Lernens kann Deep Learning komplexere Probleme lösen und eine höhere Genauigkeit bei der Mustererkennung und Vorhersage liefern. Deep Learning hat in vielen unterschiedlichen Anwendungsbereichen, wie bspw. der Bild- und Spracherkennung oder dem Natural Language Processing (NLP) in den letzten zehn Jahren bedeutende Fortschritte ermöglicht und stellt über sogenannte Transformer-Modelle die Grundlage für generative KI, wie bspw. ChatGPT von OpenAI, dar.

In einem zweiten Schritt werden insgesamt sechs ausgewählte Bestandteile des maschinellen Lernens, der neuronalen Netze und des Deep Learnings erläutert, um deren zugrundeliegende Funktionsweise anschaulich darstellen zu können (Abbildung 2).

(1) Basics

Zwei elementare Bestandteile für die unterschiedlichen Techniken und Anwendungen im Bereich der KI sind (mindestens) eine Programmier-

sprache und fundierte Statistikenkenntnisse. Nachfolgend werden die Programmiersprache Python und die zentrale Bedeutung von Statistikenkenntnissen hervorgehoben:

Python ist eine der am häufigsten verwendeten Programmiersprachen im Bereich der KI im Allgemeinen und des maschinellen Lernens sowie von neuronalen Netzen im Besonderen. Es bietet eine breite Palette von Bibliotheken und Frameworks, die speziell für KI bezogene Aufgaben entwickelt wurden und den standardmäßigen Funktionsumfang der Programmiersprache Python erweitern. Insbesondere NumPy, Pandas, Scikit-learn, TensorFlow und PyTorch sind grundlegende Bibliotheken und Frameworks für KI bezogene Aufgaben. Dadurch wird u.a. das Laden, Vorverarbeiten und Modellieren von Daten sowie das Trainieren von Modellen des maschinellen Lernens und von neuronalen Netzen vereinfacht. Python ist auch bekannt für seine einfache Syntax und Lesbarkeit, was die Entwicklung und Wartung erleichtert. Es bietet darüber hinaus eine aktive Entwicklergemeinschaft und umfangreiche Ressourcen, zu denen Dokumentationen, Tutorials und Codebeispiele zählen, die bei der Lösung von Problemen und der Erweiterung des notwendigen Wissens hilfreich sein können.

Statistikenkenntnisse sind darüber hinaus unerlässlich, um das Verständnis der zugrunde

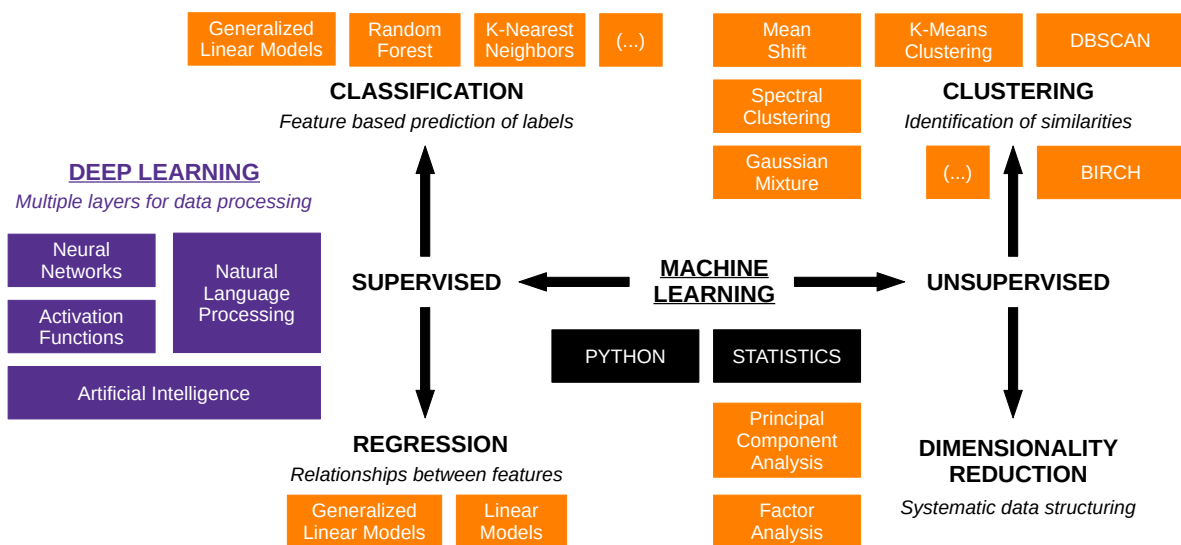


Abb. 2: Bestandteile der Künstlichen Intelligenz

liegenden Prinzipien und Konzepte des maschinellen Lernens und von neuronalen Netzen zu vertiefen. Sie helfen dabei, die verschiedenen Algorithmen des maschinellen Lernens oder die Funktionsweise der verschiedenen Aktivierungsfunktionen innerhalb von neuronalen Netzen zu verstehen und deren Anwendungsgebiete, Stärken und Einschränkungen zu erkennen. Sie sind ebenfalls wichtig, um den Prozess der erforderlichen Vorverarbeitungsschritte wie Feature Engineering, Datentransformation, Normalisierung und Aufteilung von Trainings- und Testdaten besser verstehen und umsetzen zu können. Darüber hinaus sind statistische Konzepte wie Wahrscheinlichkeitstheorie, Hypothesentests, Regressionsanalyse und Modellvalidierung wichtige Grundlagen für die Bewertung und Verbesserung von Modellen des maschinellen Lernens und von neuronalen Netzen.

(2) Regression als Methode des überwachten maschinellen Lernens

Die Regressionsanalyse ist eine statistische Methode im überwachten maschinellen Lernen, die verwendet wird, um eine Beziehung zwischen einer abhängigen Variablen (auch Zielvariable genannt) und einer oder mehreren unabhängigen Variablen herzustellen. Die Regressionsanalyse ermöglicht es, Vorhersagen für die abhängige Variable basierend auf den Werten der unabhängigen Variablen zu treffen (Abbildung 3).

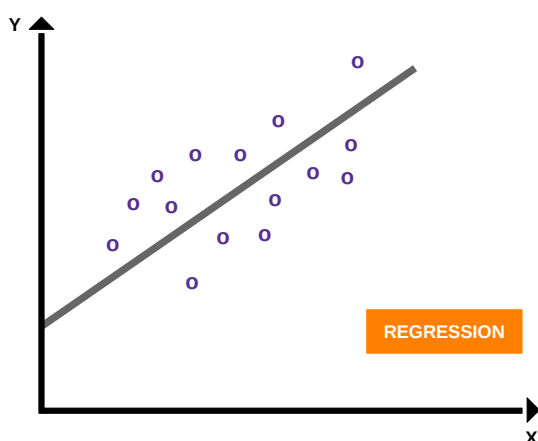


Abb. 3: Funktionsweise eines Regressionsalgorithmus

Es gibt verschiedene Arten von Regressionsmodellen, je nach den Eigenschaften der Daten und der Beziehung zwischen den Variablen. Beispiele für die verschiedenen Arten von

Regressionsmodellen sind neben der linearen Regression auch die multiple lineare Regression, die polynomiale Regression und die logistische Regression. Welches Regressionsmodell am besten zur Modellierung der gegebenen Daten geeignet ist, hängt dabei insbesondere von der Skalierung der Daten und dem Anwendungskontext ab. Das Regressionsmodell sucht dabei nach einer mathematischen Funktion, die die Beziehung zwischen den unabhängigen Variablen und der abhängigen Variable am besten beschreibt. Dies wird erreicht, indem das Modell die besten Koeffizienten schätzt, die die Funktion definieren. Gute Regressionsmodelle können verwendet werden, um Vorhersagen für neue, unbekannte Daten zu treffen und werden u.a. anhand der Metriken Bestimmtheitsmaß (R^2) sowie der mittlere quadratische Fehler (MSE) bewertet. Indem die Werte der unabhängigen Variablen in das Modell eingegeben werden, kann die abhängige Variable vorhergesagt und der Einfluss der unabhängigen Variablen geschätzt werden. Die logistische Regression kann dabei auch für eine Klassifikation herangezogen werden.

(3) Klassifikation als Methode des überwachten maschinellen Lernens

Klassifikationsverfahren im überwachten maschinellen Lernen werden verwendet, um Daten in vordefinierte Klassen oder Kategorien zu klassifizieren. Diese Verfahren ermöglichen es, auf der Grundlage von Eingabevariablen (Features) Vorhersagen über die Klassenzugehörigkeit neuer Datenpunkte zu treffen (Abbildung 4).

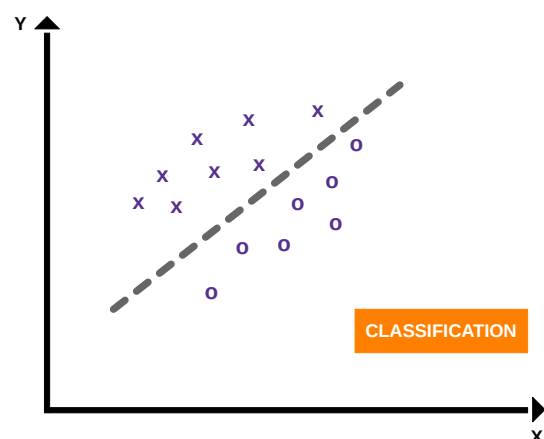


Abb. 4: Funktionsweise eines Klassifikationsalgorithmus

Es gibt verschiedene Klassifikationsalgorithmen, die je nach Art der Daten und der Problemstellung verwendet werden können. Beispiele für Klassifikationsalgorithmen sind die logistische Regression, Random Forest oder K-Nearest Neighbors. Die Auswahl des richtigen Klassifikationsalgorithmus hängt von verschiedenen Faktoren ab, einschließlich der Art der Daten, der Verfügbarkeit von gelabelten Daten, der Komplexität des Problems und der erforderlichen Vorhersagegenauigkeit. Es ist wichtig, verschiedene Algorithmen zu vergleichen und diejenigen auszuwählen, die am besten zu den spezifischen Daten passen. Das Training kann iterativ erfolgen, indem der Algorithmus die Modellparameter anpasst, um die Vorhersagegenauigkeit zu verbessern. Verschiedene Metriken, darunter die Precision, der Recall, der F1-Score und die sogenannte Receiver Operating Characteristic Kurve werden verwendet, um die Qualität der Klassifikation zu bewerten. Sobald das Modell bewertet wurde und eine akzeptable Leistung aufweist, kann es verwendet werden, um Vorhersagen für neue, unbekannte Datenpunkte zu treffen. Indem die Merkmale der neuen Daten in den Klassifikator eingegeben werden, wird die Klassenzugehörigkeit vorhergesagt.

(4) Clustering als Methode des unüberwachten maschinellen Lernens

Die Clusteranalyse ist eine Methode des unüberwachten maschinellen Lernens, die verwendet wird, um natürliche Gruppierungen oder Cluster in einem Datensatz zu identifizieren. Das Ziel besteht darin, ähnliche Datenpunkte in Gruppen zusammenzufassen, während die Datenpunkte in verschiedenen Gruppen möglichst unterschiedlich sein sollten. Die Clusteranalyse basiert nicht auf vorab bekannten Labels oder Klassen, sondern erkennt automatisch die Struktur oder Muster in den Daten (Abbildung 5).

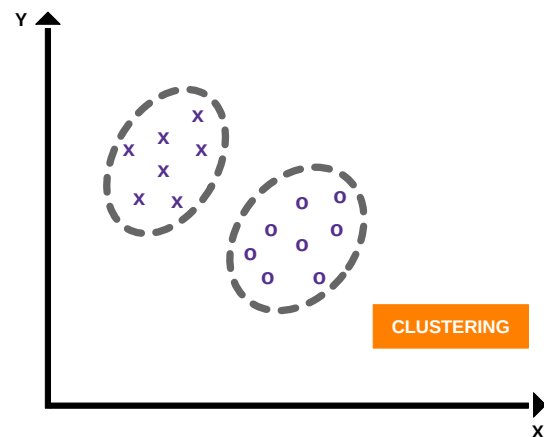


Abb. 5: Funktionsweise eines Clusteringalgorithmus

Es gibt verschiedene Algorithmen für die Clusteranalyse, darunter u.a. Gaussian Mixture, Spectral Clustering, Mean Shift, K-Means Clustering, DBSCAN und Birch. Je nach Art der Daten und der gewünschten Clusterstruktur wird der passende Algorithmus ausgewählt. Nach der Bildung der Cluster können verschiedene Metriken verwendet werden, um die Qualität des Clusterings zu bewerten. Dazu gehören beispielsweise die Silhouettenkoeffizienten, die internen Validitätsindizes oder externe Maße wie der Adjusted Rand Index (ARI) bei vorhandenen Labels.

Nach dem Clustering-Prozess werden die gebildeten Cluster analysiert und interpretiert. Dies kann beinhalten, die Charakteristika und Gemeinsamkeiten der Datenpunkte innerhalb jedes Clusters zu untersuchen und sie zu beschreiben. Visualisierungstechniken wie Scatterplots oder Dendrogramme können verwendet werden, um die Clusterstruktur zu visualisieren.

(5) Dimensionality Reduction als Methode des unüberwachten maschinellen Lernens

Principal Component Analysis (PCA) und Factor Analysis (FA) sind zwei gängige dimensionsreduzierende Verfahren des unüberwachten

maschinellen Lernens, die verwendet werden, um die Dimensionalität von Datensätzen zu reduzieren, indem sie die Anzahl der Merkmale oder Variablen verringern (Abbildung 6).

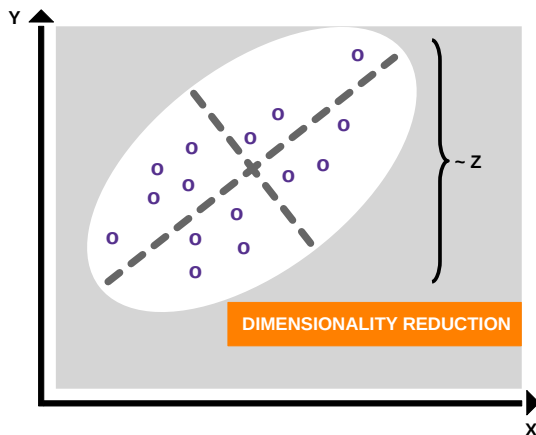


Abb. 6: Funktionsweise eines Dimensionsreduktionsalgorithmus

Obwohl beide Verfahren das Ziel haben, die Daten zu vereinfachen, unterscheiden sie sich in ihrer zugrunde liegenden Annahme und Vorgehensweise:

PCA ist eine lineare Methode, die darauf abzielt, die Varianz in den Daten zu maximieren und die Dimensionalität durch Transformation der ursprünglichen Merkmale in neue, unkorrelierte Variablen, die als Hauptkomponenten bezeichnet werden, zu reduzieren. Die Hauptkomponenten werden in einer absteigenden Reihenfolge ihres Beitrags zur Gesamtvarianz der Daten angeordnet. Durch die Reduzierung der Dimensionalität auf Hauptkomponenten können die Daten komprimiert werden, wobei die größtmögliche Varianz beibehalten wird. Dies ermöglicht eine effizientere Analyse und Visualisierung der Daten.

FA ist ein probabilistisches Modell, das versucht, die zugrunde liegenden latenten Faktoren zu identifizieren, welche die beobachteten Variablen in einem Datensatz erklären. Es geht davon aus, dass die beobachteten Variablen durch eine Kombination von gemeinsamen Faktoren und spezifischen Fehlerkomponenten erklärt werden können. Durch die Reduzierung der Dimensionalität auf die relevanten Faktoren kann eine FA die gemeinsame Variabilität

in den Daten erfassen und gleichzeitig die Rausch- oder Fehlerkomponenten eliminieren. Dies ermöglicht eine einfachere Interpretation der Daten und kann in einigen Fällen dazu beitragen, versteckte Muster oder Strukturen zu enthüllen.

(6) Neuronale Netze und Deep Learning

Neuronale Netze sind von biologischen Gehirnen inspiriert und können im Sinne eines Deep Learnings verwendet werden, um komplexe Aufgaben des maschinellen Lernens zu lösen (Abbildung 7).

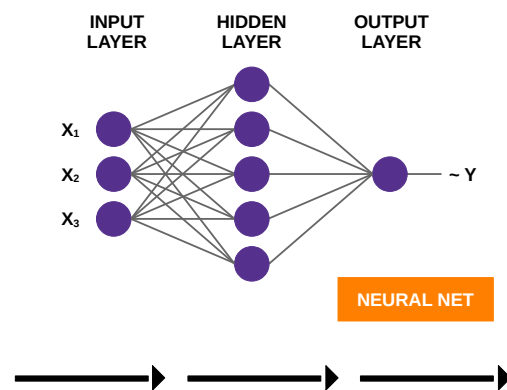


Abb. 7: Funktionsweise eines neuronalen Netzes

Die Funktionsweise eines neuronalen Netzes kann grob in drei Schritte unterteilt werden: Eingabe, Verarbeitung und Ausgabe.

Die Eingabe eines neuronalen Netzes besteht aus einer Sammlung von Datenpunkten oder Merkmalen, die als Eingabevektor repräsentiert werden. Diese Eingabevektoren können vielfältige Informationen enthalten, wie bspw. Bilder, Texte oder numerische Werte. Jeder Eingabevektor wird den Eingangsschichten des neuronalen Netzes zugeführt.

Die Verarbeitung der Eingabe erfolgt durch mehrere Schichten von Neuronen, die als versteckte Schichten oder auch Zwischenschichten bezeichnet werden. Jedes Neuron in einer Schicht ist mit den Neuronen der vorherigen Schicht verbunden. Jede Verbindung zwischen Neuronen hat ein Gewicht, das die Stärke der Verbindung repräsentiert. Die Neuronen in den Zwischenschichten wenden lineare und

nichtlineare Transformationen auf die Eingabevektoren an, indem sie die gewichteten Summen der Eingaben berechnen und sie dann durch eine Aktivierungsfunktion laufen lassen. Die Aktivierungsfunktionen ermöglichen es dem Netzwerk, die Zusammenhänge zu erfassen und komplexe Muster zu lernen.

Das Ergebnis der Verarbeitung wird an die Ausgangsschicht des neuronalen Netzes weitergeleitet. Die Ausgangsschicht enthält Neuronen, die die endgültige Ausgabe des Netzwerks repräsentieren. Je nach Art des Problems kann die Ausgabe des neuronalen Netzes verschiedene Formen annehmen. Zum Beispiel kann es sich um eine Klassifikation (bspw. Erkennung von Objekten in einem Bild), eine Regression (bspw. Vorhersage von Aktivitäten auf Social Media Plattformen) oder eine Sequenzgenerierung (bspw. Erzeugung von Text) handeln.

Die Funktionsweise eines neuronalen Netzes beruht wie beim maschinellen Lernen auf dem Lernen aus Daten. Durch ein Verfahren namens „Training“ werden die Gewichte der Verbindungen zwischen den Neuronen angepasst, um das Netzwerk in die Lage zu versetzen, Muster und Beziehungen in den Trainingsdaten zu erkennen und daraus Vorhersagen oder Entscheidungen zu treffen. Das Training erfolgt normalerweise mit Hilfe von Optimierungsalgorithmen wie dem Backpropagation-Algorithmus, der den Fehler zwischen den vorhergesagten Ausgaben des Netzwerks und den tatsächlichen Ausgaben minimiert.

Einschlägige Weiterbildungsangebote

KI-bezogene Weiterbildungsangebote werden oftmals unter dem Schlagwort Data Science sowohl in institutionalisierter als auch nicht-institutionalisierter, sowie direkter und indirekter Form angeboten. Auf Seite der institutionalisierten und direkten Weiterbildungsangebote kann u. a. auf *Certified Data Scientist Basic Level* und konsekutive Vertiefungen wie *Certified Data Scientist Specialized in Data Analytics* sowie *Certified Data Scientist Specialized in Big Data Analytics* verwiesen werden,

welche regelmäßig von der Fraunhofer Gesellschaft angeboten werden. Die Weiterbildungsangebote dauern dabei zwischen einigen Tagen bis zu mehreren Wochen.

Institutionalisierte und indirekte Formen sind insbesondere dem Bereich der Massive Open Online Courses zuzuordnen. Hier stellen vornehmlich Hochschulen über einschlägige Weiterbildungsplattformen wie bspw. Coursera das erforderliche Fach- und Methodenwissen für ein Selbststudium mit der Möglichkeit angeleiteter und bewerteter Assignments zur Verfügung. In vielen Fällen sind die Inhalte der indirekt über Weiterbildungsplattformen angebotenen Materialien dabei deckungsgleich mit den Materialien bei einer direkten Teilnahme an den Hochschulen selbst. Exemplarisch kann hier auf das Weiterbildungsangebot *Machine Learning* der Stanford University verwiesen werden, welches sowohl auf der Weiterbildungsplattform Coursera, aber auch auf den Internetseiten der Stanford University im Selbststudium absolviert und bei erfolgreicher Teilnahme mit einem entsprechenden Zertifikat abgeschlossen werden kann.

Inhaltlich weisen verschiedene Weiterbildungsangebote dabei teilweise deutliche Überschneidungen auf, insbesondere wenn es um die Vermittlung des zugrundeliegenden mathematischen Fach- und Methodenwissens geht. Exemplarisch kann hier auf die Weiterbildungsangebote *Mathematics for Machine Learning* des Imperial College London und *Data Science Math Skills* der Duke University, beide auf der Weiterbildungsplattform Coursera angeboten, verwiesen werden. Diese vermitteln gleichsam die Grundlagen für die Analyse linearer und nichtlinearer Funktionen und das forschungsmethodische Vorgehen bezüglich Vektoren, bspw. als Grundlage der Algorithmen für Clusteranalysen. Während das Weiterbildungsangebot der Duke University dabei zusätzlich die Wahrscheinlichkeitstheorie thematisiert, vertieft das Weiterbildungsangebot des Imperial College London insbesondere die Besonderheiten multivariater Analyseverfahren, die bspw. im privatwirtschaftlich

organisierten Weiterbildungsangebot *Deep Learning* von DeepLearning.AI eine entsprechende Grundlage darstellen. Darüber hinaus bietet DeepLearning.AI in Kooperation mit OpenAI ein Weiterbildungsangebot zum *ChatGPT Prompt Engineering for Developers* sowie *LangChain for LLM Application Development* an, um generative KI und einschlägige Sprachmodelle versierter nutzen zu können.

Auch im bundesdeutschen Hochschulkontext treten Weiterbildungsangebote aus dem Data Science Kontext zunehmend in Erscheinung. Erste *Empfehlungen für Masterstudiengänge Data Science* an deutschen Hochschulen sind von der Gesellschaft für Informatik e.V. ausgesprochen worden und verweisen darin auf das erforderliche Fach- und Methodenwissen. Beispielsweise stellt der Bachelorstudiengang *Data Science* der Justus-Liebig-Universität Gießen im Curriculum die deutlichen Parallelen zu den Grundlagen der Informatik in den Vordergrund der akademischen Ausbildung.

Ähnliche Parallelen zeigen sich auch hinsichtlich der einschlägigen Data Science Weiterbildungsangebote, die sich hauptsächlich hinsichtlich der verwendeten Programmiersprachen unterscheiden. Hier sind u.a. das mehrere Monate umfassende Weiterbildungsangebot *Data Science Specialization* der Johns Hopkins University sowie das kompaktere Weiterbildungsangebote *Machine Learning* der Stanford University zu nennen. Ersteres setzt auf die Programmiersprache R und Zweiteres auf Python. In den zugrundeliegenden Curricula werden dabei gleichermaßen die Prozesse der Datenaufbereitung, einschlägige Algorithmen und die anschließende Visualisierung von Befunden thematisiert.

Insgesamt sind sowohl die Komplexität des erforderlichen Grundlagenwissens als auch die Mehrstufigkeit der hierfür erforderlichen Weiterbildungsangebote ausschlaggebend für die mit entsprechenden Stellenausschreibungen einhergehenden Qualifikationsprofile. In der Regel umfasst bspw. eine einmalige akademi-

sche Ausbildung nicht das gesamte Grundlagenwissen, so dass eine kontinuierliche und aktive Auseinandersetzung mit den Weiterbildungsangeboten zum sich stets aktualisierenden Thema KI sowie den aus dem Data Science Kontext stammenden Grundlagen geboten ist.

Limitationen

Beim Einsatz von KI, maschinellem Lernen, neuronalen Netzen und Deep Learning in der Radikalisierungsforschung sollten stets die damit in Verbindung stehenden Limitationen berücksichtigt werden:

Datenabhängigkeit

KI-Modelle erfordern große Mengen qualitativ hochwertiger Daten, um effektiv zu funktionieren. Der Mangel an ausreichenden oder repräsentativen Daten kann die Leistung und die Fähigkeit des Modells einschränken.

Begrenzte Generalisierung

KI-Modelle haben oft Schwierigkeiten, über die spezifischen Daten hinaus zu generalisieren, auf denen sie trainiert wurden. Wenn das Modell mit Daten konfrontiert wird, die sich deutlich von den Trainingsdaten unterscheiden, kann es zu ungenauen oder fehlerhaften Vorhersagen kommen.

Erklärbarkeit

KI-Modelle auf Basis neuronaler Netze können aufgrund ihrer komplexen Struktur und der großen Anzahl von Parametern schwer verständlich sein. Dies kann zu einer Herausforderung werden, insbesondere in sensiblen Bereichen wie der Radikalisierungsforschung, in der Entscheidungen nachvollziehbar sein müssen.

Ethik und Bias

KI-Modelle sind anfällig für die Übernahme von Vorurteilen und Diskriminierung, die in den Trainingsdaten vorhanden sind. Wenn die Daten beispielsweise bestimmte Bevölkerungsgruppen unterrepräsentieren oder voreingenommene Muster aufweisen, können die Modelle diese Vorurteile verstärken.

Sicherheit und Datenschutz

KI-Modelle können anfällig für Angriffe sein, bei denen Angreifer absichtlich manipulierte Daten einspeisen, um das Modell zu täuschen oder Fehlverhalten zu verursachen. Darüber hinaus müssen Datenschutzaspekte berücksichtigt werden, insbesondere wenn sensible oder persönliche Informationen in den Daten enthalten sind.

Ressourcenbedarf

KI-Modelle, insbesondere tiefe neuronale Netze, erfordern erhebliche Rechenressourcen, um effizient trainieren zu können. Dies kann finanzielle und infrastrukturelle Herausforderungen mit sich bringen und den Einsatz in bestimmten Umgebungen einschränken.

Rechtliche und regulatorische Aspekte

Der Einsatz von KI-Modellen wirft eine Vielzahl von rechtlichen und regulatorischen Fragen auf, insbesondere im Hinblick auf Haftung, Verantwortlichkeit und Datenschutz. Es ist wichtig, diese Aspekte zu berücksichtigen und angemessene Rahmenbedingungen zu schaffen.

Unzureichendes Grundlagenwissen

Als letzte – ebenso wichtige – Limitation wird auf ein unzureichendes Grundlagenwissen der Anwenderinnen und Anwender von KI-Modellen selbst, aber auch von Entscheidungsträgerinnen und Entscheidungsträgern im Kontext von Behörden und Organisationen mit Sicherheitsaufgaben verwiesen.

Eine nicht zielführende Erwartungshaltung, unzureichendes Feintuning von KI-Modellen oder unpassende – und dem erforderlichen Grundlagenwissen nicht angemessene Stellenausschreibungen einschließlich der vorgesehenen Vergütung zur Gewinnung von qualifiziertem Fachpersonal (bspw. sei auf die oftmals ausgeschriebene Entgeltgruppe 12 beim Bundesamt für Verfassungsschutz für Data Scientists verwiesen) – können die Folge sein.

Diese Limitationen sollten daher bei der Entwicklung, Implementierung und Nutzung von KI-Modellen berücksichtigt werden, um deren Wirksamkeit, Fairness, Sicherheit und ethische Verantwortung sicherzustellen.

Impressum

MOTRA-Verbundpartner



Gefördert vom



Bundesministerium
für Bildung
und Forschung

Bundesministerium
des Innern
und für Heimat

Angaben gemäß § 5 TMG

Dr. Uwe Kemmesies
Äppelallee 45
65203 Wiesbaden

Vertreten durch

Dr. Uwe Kemmesies
E-Mail: motra@bka.bund.de

Verantwortlich für den Inhalt nach § 55 Abs. 2 RStV

Dr. Uwe Kemmesies
Äppelallee 45
65203 Wiesbaden

Redaktion

Isabelle Holz

Layout

studio halvar

Forschungskordinator (MOTRA-Verbund)

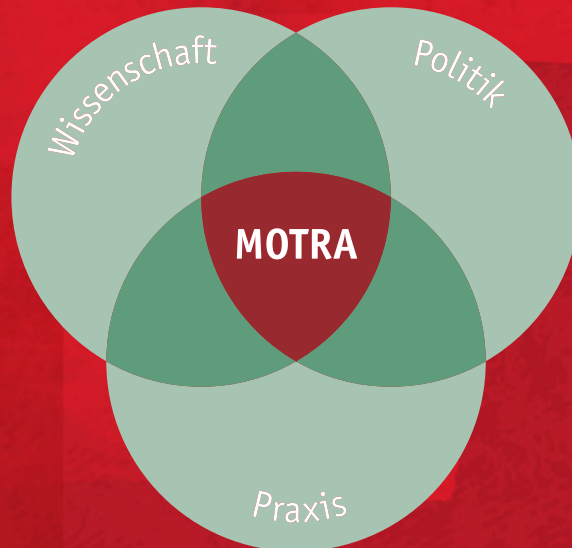
Dr. Uwe Kemmesies

Kontakt

Fragen zu MOTRA allgemein: motra@bka.bund.de
Fragen zum Inhalt dieser Spotlight-Ausgabe an:
Dennis Klinkhammer/dennis.klinkhammer@fom.de

MOTRA

Monitoringsystem und Transferplattform Radikalisierung



Der MOTRA-Forschungsverbund vereint in einem Spitzenforschungscluster neun institutionelle Partner aus renommierten universitären und außeruniversitären Forschungseinrichtungen sowie unterschiedlichen wissenschaftlichen Fachdisziplinen. MOTRA sucht den kooperativen Dialog auf Augenhöhe mit weiteren wissenschaftlichen Forschungsverbänden und Netzwerkpartnern aus Praxis und Politik und verfolgt dabei zwei zentrale Zielstellungen:

1. Monitoring

Es soll durch MOTRA ein Monitoringsystem aufgebaut werden, mittels dessen das politisch und/oder religiös begründete Radikalisierungsgeschehen in Deutschland auf einer breiten Datenbasis fortlaufend beobachtet wird. Ziel ist es die Verbreitung, Entwicklungstrends sowie begünstigende Konstellationen für Extremismus und Radikalisierung besser erfassen zu können.

2. Wissenstransfer

Weiterhin verfolgt MOTRA den Aufbau und die Etablierung einer multifunktionalen Austauschplattform zur Gestaltung eines direkteren Wissenstransfers zwischen den Handlungsfeldern wissenschaftlicher Forschung, sozialer Praxis und Politikgestaltung.

 motra@bka.bund.de

 www.motra.info

 twitter.com/MOTRAVERBUND

