31. Januar 2023

Sebastian Knopf

Mat.-Nr.: 68390

Masterarbeit

**Entwicklung und Evaluation eines auf künstliche**

**Intelligenz gestützten Systems zur Betriebslenkung**

**von Linienbussen im Störungsfall**

**Betreuung**

M. Sc. Waldemar Titov,   
Dirk Weißer (VDV)

**Verantwortlicher Hochschullehrer**

Prof. Dr.-Ing. Thomas Schlegel

**Aufgabenstellung**

(DE) Kurzfristige Streckensperrungen oder erhöhtes Verkehrsaufkommen gefährden die Fahrplanstabilität und sorgen dadurch für Frust und Hemmnisse im öffentlichen Personenverkehr. Besonders kleine und mittelständische Verkehrsunternehmen verfügen in den meisten Fällen nicht über eine Betriebsleitstelle, die geeignete Umleitungsfahrwege anweisen und dadurch den Betrieb aufrechterhalten kann. In dieser Masterarbeit soll untersucht werden, inwieweit die Anordnung von Umleitungen durch ein Betriebsleitsystem unterstützt durch künstliche Intelligenz (KI) automatisiert werden kann. Nach einer Literaturrecherche zu den Themen Betriebsleitsystem und KI werden passende Verfahren ausgewählt und anhand von Daten aus der Praxis miteinander verglichen. Eine Evaluation prüft anhand ausgewählter Kennzahlen und im Vergleich mit Expertenentscheidungen prüfen Leistungsfähigkeit und Praxistauglichkeit der gewählten Verfahren. Eine kritische Diskussion, eine Zusammenfassung und ein Ausblick runden die Masterarbeit ab.

(EN) Unplanned service interruptions or traffic jams affect the public transport service stability negatively and thus customers feel frustrated and act self-consciously regarded to public transport services. Most smaller and medium-sized traffic companies do not have a control center which may arrange a re-routing and keep the service running. The aim of this work is to find procedures which enable a vehicle control system to arrange re-routing of public transport vehicles automatically based on artificial intelligence (AI). To achieve this, different approaches of AI are selected based on literature research and compared to each other based to datasets from the operation of several bus agencies. An evaluation confirms performance and productivity of the AI procedures found. A critical discussion and a summary show possible commitments in practice and complement the whole work.

**Betreuender Hochschullehrer**

Prof. Dr.-Ing. Thomas Schlegel

**Erklärung**

Hiermit erkläre ich, die vorliegende Arbeit mit dem Titel

**Entwicklung und Evaluation eines auf künstliche Intelligenz gestützten Systems zur Betriebslenkung von Linienbussen im Störungsfall**

selbstständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt sowie Zitate kenntlich gemacht zu haben.

Karlsruhe, den 31. Januar 2023

Sebastian Knopf

Inhaltsverzeichnis

[Abkürzungsverzeichnis 0](#_Toc113947996)

[Abbildungsverzeichnis 0](#_Toc113947997)

[Tabellenverzeichnis 0](#_Toc113947998)

[1 Einleitung 2](#_Toc113947999)

[1.1 Motivation 3](#_Toc113948000)

[1.2 Zielsetzung 5](#_Toc113948001)

[1.3 Vorgehensweise 5](#_Toc113948002)

[2 Grundlagen 6](#_Toc113948003)

[2.1 Definition künstlicher Intelligenz 6](#_Toc113948004)

[2.2 Künstliche Intelligenz und maschinelles Lernen 8](#_Toc113948005)

[3 Literaturverzeichnis 10](#_Toc113948006)

[Anhang 12](#_Toc113948007)

# Abkürzungsverzeichnis

KI *Künstliche Intelligenz*

KNN *k-Nearest-Neighbours Algorithmus*

ML *Maschinelles Lernen oder Maschinenlernen*

# Abbildungsverzeichnis

[**Abbildung 1:** Darstellung einer klassischen und einer unscharfen Menge 10](https://d.docs.live.net/aeec904eeb5479b8/Dokumente/Studium/11%20WS2223/02%20Masterarbeit/02%20Repository/Ausarbeitung_MA_Knopf.docx#_Toc114070395)

[**Abbildung 2:** Graphische Darstellung der Überdeckung unscharfer Mengen 11](https://d.docs.live.net/aeec904eeb5479b8/Dokumente/Studium/11%20WS2223/02%20Masterarbeit/02%20Repository/Ausarbeitung_MA_Knopf.docx#_Toc114070396)

# Tabellenverzeichnis

# Einleitung

Häufig gilt der öffentliche Personenverkehr noch immer als unzuverlässig, wenig flexibel und zu kompliziert. Unvorhergesehene Ereignisse und Streckensperrungen beeinträchtigen die Fahrplanstabilität und bestätigen damit unnötigerweise das Bild des unzuverlässigen, unflexiblen öffentlichen Personenverkehrs. Auswertungen nach dem Ende des im Sommer 2022 von der Deutschen Bundesregierung initiierten 9-Euro-Tickets zeigen, dass neben dem Fahrpreis besonders auch die Angebotsqualität sowohl bezogen auf die Verfügbarkeit als auch auf die vorhandene Fahrgastinformationen erheblichen Einfluss auf die Verkehrsmittelverlagerung haben (vgl. Krämer 2022).

Durch die Marktöffnung im öffentlichen Personenverkehr sind insbesondere in Busnetzen immer häufiger mittelständische Verkehrsunternehmen mit der Betriebsdurchführung betraut. Bedingt durch den hohe und stetig wachsenden Kostendruck leisten sich kleine und mittelständische Verkehrsunternehmen selten eine Betriebsleitstelle, sodass insbesondere bei kurzfristigen Störungen nicht zeitgerecht reagiert und beispielsweise Umleitungen von zentraler Stelle angeordnet werden können. In den meisten Fällen kommt der Betrieb im Störungsfall kurzzeitig zum Erliegen. Selbst nach Ende der Störung kann es mitunter Stunden dauern, bis alle Fahrzeuge bedingt durch fehlende Wendezeiten in den Wagenumläufen die geplanten Fahrten wieder planmäßig durchführen können. Eine zeitgemäße Information der Fahrgäste unterbleibt in den meisten Fällen komplett, was wiederrum zu Frustration und Abneigung auf Seiten der Fahrgäste führt.

Im Rahmen dieser Masterarbeit soll untersucht werden, inwieweit sich Daten aus der Planung als auch aus dem Betrieb als Entscheidungsgrundlage für dispositive Maßnahmen eignen und wie ein Betriebsleitsystem unterstützt durch künstliche Intelligenz damit selbstständig in die Lage versetzt werden kann, den Betrieb auch im Störungsfall möglichst aufrecht zu erhalten.

## Motivation

Aus technischer Sicht betrachtet sind viele Lösungsansätze für Teilprobleme bereits verfügbar. Besonders in größeren Verkehrsunternehmen, die im Regelfall dann auch über eine über die gesamte Betriebszeit besetzte Betriebsleitstelle verfügen, sind Lösungen zur Unterstützung des Leitstellenpersonals im Störungsfall vielfach auch implementiert.

Bei kleineren und mittelständischen Verkehrsunternehmen ohne Betriebsleitstelle werden wertvolle Daten aus den Bordrechnern der Fahrzeuge hingegen nicht genutzt, obwohl genau diese eine gute Basis für zukünftige Entscheidungen sein könnten. Diese Entscheidungen könnten sowohl im Umfeld der Betriebslenkung als auch bei der Fahrgastinformation eine Basis zur stabilen Aufrechterhaltung des Betriebes sein. Was dies konkret bedeuten könnte, soll an einem selbst erlebten Beispiel aus dem Alltag verdeutlicht werden:

An einem Samstagvormittag möchte ich öffentliche Verkehrsmittel zu einer großen Veranstaltung nutzen. Erfreulicherweise fährt der kombinierte Stadt- und Regionalverkehr auch am Wochenende im 30-Minuten-Takt. Am Busbahnhof angekommen, sehe ich wie sich eine Busfahrerin etwas aufgeregt mit einem Kollegen unterhält. Beim Einstieg in den Bus erzählt mir die motivierte Busfahrerin, dass meine Zielhaltestelle wegen dem hohen Verkehrsaufkommen nicht mehr angefahren wird und sie mich stattdessen an einer alternativen Haltestelle absetzen wird, die zu Fuß aber dieselbe Entfernung bedeutet. Beim Ausstieg zeigt sie mir noch die Haltestelle, an der ich in den Bus zur Rückfahrt einsteigen solle.

Stunden später finde ich mich pünktlich zur Rückfahrt an der besagten Haltestelle ein, doch auch 20 Minuten nach der planmäßigen Abfahrtszeit taucht kein Bus auf. Einschlägige Informationsmedien und die lokale Fahrplanauskunft helfen mir kein Stück weiter, da sie weder Echtzeitdaten noch eine vernünftige Information über eine eventuelle Umleitung enthalten. Erst 30 Minuten später taucht ein Kleinbus auf, der dann noch einen stark von der Fahrplanauskunft abweichenden Weg zurück zum Busbahnhof fährt und mich schließlich mit einer Verspätung von rund 40 Minuten an den Busbahnhof zurückbringt.

Das hier beschriebene Problem gehört oftmals zum Alltag einer jeden Person, die den öffentlichen Personenverkehr täglich nutzt.

Was ist hier passiert? Die folgende stichpunktartige Analyse soll eine wahrscheinliche Antwort auf diese Frage liefern.

* Ausschlaggebend für die Umleitung war das starke Verkehrsaufkommen an einer Stelle im Verlauf der Buslinie.
* Eine Einhaltung des regulären Linienweges hätte dafür gesorgt, dass sich je Folgefahrt die Verspätung mit einem Delta von +10 Minuten erhöht hätte, da Wendezeiten in diesem Ausmaß im geplanten Wagenumlauf nicht enthalten sind.
* Da es keine Betriebsleitstelle, die korrigierend von zentraler Stelle eingreifen kann und auch kein Konzept für einen solchen Störfall gibt, entschied das Fahrpersonal vor Ort eigenständig, dass die Verspätung durch eigenständige Änderung des Linienweges vermieden wird. Zwar erfolgte eine Absprache und der den beiden aktuell diensthabenden Personen, ob diese aber auch an die Ablösung weitergegeben wurde, bleibt fraglich. Auch andere Fahrzeuge, die an dieser Absprache nicht beteiligt waren, wurden nicht informiert.
* Durch das eigenständige Entscheiden des Fahrpersonals kamen unterschiedliche Umleitungswege zur Anwendung, die jedoch weder einheitlich noch für Fahrgäste kommuniziert wurden. Die Folge davon war, dass die Fahrpersonale jeweils ihre eigenen, nicht einheitlichen Umleitungswege bedienten, die jedoch nicht aufeinander abgestimmt waren.
* Das Fehlen einer zentralen Anweisung sorgte letztendlich für den willkürlichen Ausfall eines Linienabschnittes oder gar einer ganzen Fahrt. Die ausbleibende Fahrgastinformation machte das Chaos perfekt.

Das Wissen über technische Möglichkeiten und das eigene Interesse, Lösungen zur Behebung dieser längst bekannten Alltagsprobleme bei der Nutzung öffentlicher Verkehrsmittel zu liefern, bilden die Motivation zu dieser Masterarbeit.

## Zielsetzung

Im Rahmen dieser Masterarbeit sollen Möglichkeiten erörtert werden, die ein Betriebsleitsystem in die Lage versetzen, selbstständig Umleitungen im Störungsfall anzuordnen. Dabei sollen möglichst Eingangsdaten verwendet werden, die ohnehin vorhanden sind oder mit wenig Aufwand erzeugt werden können. Ziel ist ein funktionaler Prototyp, welcher nach Kenntnis über eine Störung selbstständig eine möglichst optimale Umleitung der betroffenen Fahrzeuge anordnet und darüber hinaus auch für die Fahrgastinformation eingesetzt werden kann. Im Kern sollen dabei folgende Forschungsfragen beantwortet werden:

* Welche Daten müssen in welcher Qualität verfügbar sein, um diese zum Training einer künstlichen Intelligenz verwenden zu können?
* Welche Strategien eignen sich für den Einsatz zur automatischen Anordnung einer Umleitung von Linienbussen in einem Betriebsleitsystem?
* Bietet ein solches auf künstliche Intelligenz gestütztes Betriebsleitsystem Potenzial für einen zeitnahen Einsatz in der Praxis?

## Vorgehensweise

Für das grundlegende Verständnis der Arbeit werden in **Kapitel 2** zunächst Grundlagen der künstlichen Intelligenz in einer umfangreichen Literaturrecherche ermittelt. Ein Blick hinter die Kulissen einiger Verkehrsunternehmen soll ergänzend einen Überblick über den aktuellen Stand der Technik liefern und darüber hinaus offenlegen, welche Daten abgreifbar sind.

Die folgenden **Kapitel 3** und **4** beschäftigen sich jeweils mit der Auswahl zweier zum Vergleich geeigneter KI-Verfahren und der Implementierung des Funktionsprototypen.

Eine Evaluation in **Kapitel 5** basierend auf qualitativen Daten stellt die beiden KI-Verfahren auf den Prüfstand und zeigt, ob diese überhaupt geeignet sind und welches Optimierungspotenzial sie bieten. Meinungen von Fachleuten verschiedener Verkehrsunternehmen fließen darüber hinaus als subjektive Einschätzung in die Evaluation ein und schärfen gewonnene Erkenntnisse bezogen auf die Praxistauglichkeit.

In **Kapitel 6** werden die Ergebnisse der Masterarbeit zusammengefasst und zur Diskussion gestellt.

# Grundlagen

In diesem Kapitel sollen zunächst theoretische und technische Grundlagen der künstlichen Intelligenz erörtert und zusammengefasst werden. Die Literaturrecherche bildet anschließend die Grundlage für die Auswahl zweier geeigneter KI-Strategien, die im weiteren Verlauf der Masterarbeit weiter betrachtet werden sollen. Im Zweiten Teil des Kapitels werden bestehende Lösungen bei verschiedenen Verkehrsunternehmen beleuchtet. Der Blick geht hierbei bewusst über den Horizont der Linienbusse hinaus, um eventuell gesammelte Erfahrungen aus verwandten Bereichen des Verkehrs und der Logistik transferieren zu können.

## Definition künstlicher Intelligenz

Der Begriff der künstlichen Intelligenz (KI) spielt in dieser Masterarbeit eine Schlüsselrolle. Sie soll als Entscheidungsträger in einem Betriebsleitsystem eingesetzt werden. Doch wann genau agiert ein System *intelligent*? Nachfolgend werden einige Definitionsansätze beleuchtet, welche diese Frage beantworten und eine Arbeitsdefinition für künstliche Intelligenz im Kontext der Betriebslenkung im öffentlichen Personenverkehr liefern sollen. Grundsätzlich problematisch hierbei ist, dass der Begriff der Intelligenz selbst nicht definiert ist. So existiert keine einheitliche Definition von Intelligenz, welche als Referenz zur Beurteilung eines Systems dienen kann (vgl. Rimscha 2008, S. 105; Mainzer 2019, S. 2).

Die wohl bekannteste und gleichwohl älteste Definition für den Begriff der künstlichen Intelligenz stammt von Turing aus den 1950er Jahren. Dem sogenannten Turing-Test zur Folge ist ein System dann als intelligent anzusehen, wenn eine beobachtende Person nicht mehr in der Lage ist zu unterscheiden, ob sie mit einer Maschine oder einer natürlichen Person interagiert (vgl. Mainzer 2019, S. 10). Diese Aussage wurde entsprechend oft auch hinterfragt. Bekannt wurde die Kritik von Ada Lovelace, die behauptete, eine Maschine könne ausschließlich de Zweck dienen, für den sie letztendlich programmiert wurde (vgl. Liggieri und Müller 2019, S. 304). Noch eindrucksvoller wird die Kritik am Beispiel des „Chinesischen Zimmers“ nach Searle (1980). Hierbei handelt es sich um ein gedankliches Experiment mit Ursprung in der Psychologie, bei dem eine Person in einem Raum steht und Fragen auf Chinesisch gestellt bekommt. Als einziges Hilfsmittel verfügt die Person über ein Regelbuch, mit dessen Hilfe sie die Fragen ebenfalls auf Chinesisch beantworten muss. Eine außenstehende Person könnte zu dem Schluss kommen, dass die Person im Raum die Sprache Chinesisch beherrscht, obwohl sie in Wahrheit nur einen fest definierten Regelsatz abarbeitet und überhaupt nicht über Kenntnisse der chinesischen Sprache verfügt (vgl. Liggieri und Müller 2019, S. 304). Diese beiden Gegenargumente zeigen deutlich, dass die fehlende äußere Unterscheidbarkeit zwischen einem Menschen und einer Maschine noch nicht zwangsläufig eine Interaktion mit einem scheinbar intelligenten Individuum von bedeuten. Im letztgenannten Beispiel wäre dann Intelligenz im Spiel, wenn die Person innerhalb des Raumes sich tatsächlich Kenntnisse der chinesischen Sprache angeeignet hätte.

Einen weiteren, deutlich konkreteren Definitionsansatz liefert Mainzer (2019). Er betrachtet Intelligenz nicht als absolute, binäre Eigenschaft, die durch ein Individuum entweder erfüllt oder nicht erfüllt sein kann, sondern relativiert den Begriff. Statt einem absoluten Vorhandensein oder Nichtvorhandensein der Eigenschaft Intelligenz, wird diese in Graden angegeben und bereits nicht mehr auf beliebige Individuen, sondern schon auf „[Computer]Systeme“ (Mainzer 2019, S. 3) bezogen. Demzufolge hängt der Intelligenzgrad eines Systems „vom Grad der Selbstständigkeit, vom Grad der Komplexität des Problems und dem Grad der Effizienz der Problemlösung“ (Mainzer 2019, S. 3) ab. Diese Definition eröffnet in Bezug auf ein Betriebsleitsystem neue Möglichkeiten. So können Selbstständigkeit, Komplexität des Problems und Effizienz gemessen und in alternativen Systemen miteinander verglichen werden.

Noch kürzer fasst sich Rimscha (2008) bei seiner Definition. Diese besagt sinngemäß, dass ein System dann als intelligent angesehen werden kann, wenn selbstständig Probleme lösen kann, ohne vorher von einem Menschen gesagt zu bekommen, wie genau das Problem gelöst werden soll (vgl. Rimscha 2008, S. 105). Dieser Auffassung ähnelt auch die Definition von Nahrstedt (2012). Ein System im Kontext der künstlichen Intelligenz ist demnach als sogenanntes Expertensystem anzusehen, wenn die gelieferten Lösungen „von der Qualität her von denen eines menschlichen Experten nicht zu unterscheiden sind“ (Nahrstedt 2012, S. 246).

Der Begriff des Expertensystems taucht auch in der historischen Forschung zur künstlichen Intelligenz auf. Während bereits im Barockzeitalter versucht wird, Gedankengut und erlerntes Wissen allein auf Rechnungen zurückzuführen, wächst der Wunsch nach einer Universalsprache. Diese *Lingua Universalis* soll es unabhängig vom Inhalt möglich machen, Wissen in einem mathematischen Kalkül auszudrücken (vgl. Mainzer 2019, S. 8). Die Ergebnisse dieser Forschungen waren jedoch bis Ende der 1960er Jahre kaum praktisch einsatzbar, weshalb sich in den 1970er Jahren der Begriff des Expertensystems etablierte. Hierrunter versteht sich die Abgrenzung vom „allgemeinen Problemlöser“ (Mainzer 2019, S. 11) hin zu Expertenwissen in einem in sich geschlossenen und überschaubaren Bereich. Ein Expertensystem verfügt damit über begrenztes Wissen zu einem spezifischen Sachgebiet und kann mit diesem Wissen entsprechend automatisch Schlussfolgerungen ziehen. Erst diese Eingrenzung ermöglicht erste praxistaugliche Anwendungen basierend auf künstlicher Intelligenz (vgl. Mainzer 2019, S. 12).

Die Kernaussagen dieser Definitionen lassen sich zu einer Arbeitsdefinition für den Begriff des intelligenten Betriebsleitsystems zusammenfassen, die eine Bewertung des Funktionsprototypen aus Kapitel 3 ermöglichen und damit die Überprüfbarkeit sicherstellt.

Ein intelligentes Betriebsleitsystem ist demnach ein System, dass selbstständig in der Lage ist, komplexe Probleme effizient zu lösen, ohne dabei ausschließlich vordefinierte Ergebnisse aus einer endlichen Menge auszuwählen, wobei die gelieferten Ergebnisse hinsichtlich ihrer Qualität nicht von denen eines Menschen unterscheidbar sein dürfen.

Die folgenden Stichpunkte beschreiben beispielhaft, wie ein solches intelligentes Betriebsleitsystem in der Praxis ausgeprägt sein könnte:

* Die Menge der Ergebnisse ist nicht von vorneherein endlich, sondern richtet sich nach messbaren, relevanten Einflussgrößen und bildet dabei auch Schnittmengen von möglichen Kombinationen von Fahrwegen (Eigenständige Lösungsfindung)
* Die Erkennung einer Störungssituation soll vom Betriebsleitsystem automatisch basierend auf bekanntem Wissen erfolgen (Selbstständigkeit)
* Die erwartete Lösung bezieht sich auf die Anordnung von Umleitungen, nicht jedoch auf weitere Dispositive Maßnahmen wie beispielsweise Kurzwenden. Umleitungen sollen hingegen von nahezu beliebigem Umfang sein können (Komplexität)
* Die Lösung soll im Optimalfall schneller als von einem Menschen, zumindest aber in angemessen kurzer Zeit geliefert werden (Effizienz)

## Künstliche Intelligenz und maschinelles Lernen

Die Begriffe künstliche Intelligenz (KI) und maschinelles Lernen oder auch Maschinenlernen genannt (ML) tauchen oft im selben Kontext auf. In den letzten Jahren werden diese Schlüsselwörter dem Anschein nach besonders häufig dann verwendet, wenn ein Projekt nach außen als besonders innovativ oder forschungsintensiv dargestellt werden soll. Stellenweise werden beide Begriffe auch als Synonym verwendet. In diesem Kapitel werden KI, ML und verwandte Begriffe aus diesem Fachgebiet voneinander abgegrenzt.

Zur Verdeutlichung der Abgrenzung zwischen KI und ML sei das folgende Beispiel konstruiert:

In einem Eingabefeld können Nutzende Haltestellennamen eingeben und aus einem Dropdown mit Vorschlägen die gewünschte Haltestelle auswählen. Dabei werden Tippfehler in gewissen Grenzen toleriert und Vorschläge auch dann angezeigt, wenn keine exakte Übereinstimmung gefunden wurde. Werden Abkürzungen verwendet, die zu keiner direkten Übereinstimmung führen, merkt das System sich, welche Haltestelle die Nutzenden nach ihrer Eingabe aus den Vorschlägen auswählen und setzt diesen Vorschlag beim nächsten Suchvorgang im Ranking nach oben.

Die nachfolgend ausgeführten Leitfragen lauten nun:

* Handelt es sich bereits um eine Form der KI?
* Kommen hierbei Ansätze aus dem ML zum Einsatz?

**Handelt es sich bereits um eine Form der KI?**

Das System ist in der Lage, kleinere Schreibfehler zu korrigieren und den Nutzenden trotzdem passende Suchvorschläge anzuzeigen. Wird so beispielsweise „Kallsruh Hbf“ statt „Karlsruhe Hbf“ eingegeben, ist das System dennoch in der Lage, die korrekte Haltestelle „Karlsruhe Hbf“ zu ermitteln. Würde einem Menschen mit entsprechendem Fachwissen dieselbe Aufgabe stellen, wäre auch dieser in der Lage trotz des offensichtlichen Fehlers passende Ergebnisse vorzuschlagen.

Eine Möglichkeit, um solche Ergebnisse zu erhalten, ist die Verwendung sogenannter *unscharfer Mengen*. Im Gegensatz zu einer klassischen Menge, welche bezogen auf das obige Beispiel eine einfache Liste aller verfügbaren Haltestellennamen im Klartext wäre, würde eine unscharfe Menge die Haltestellennamen unterteilt in Buchstabenblöcken zu jeweils drei Buchstaben enthalten. Diese Buchstabenblöcke werden auch als *Trigramme* bezeichnet. Aus dem Wort „Karlsruhe Hbf“ würden sich demnach die Trigramme

KAR ARL RLS LSR SRU RUH UHE HBF

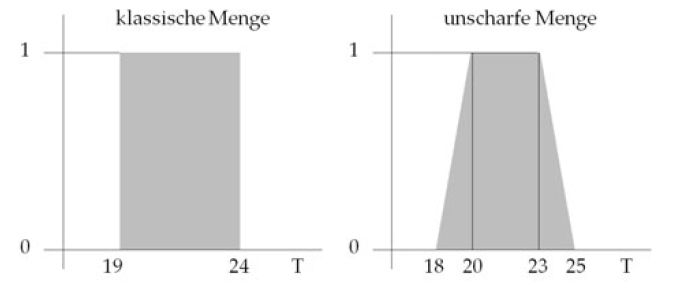
ergeben.

Der durch die Nutzenden eingegebene Text wird ebenfalls in seine Trigramme zerlegt und für jedes Element in der unscharfen Menge ermittelt, wie viele Trigramme aus der Eingabezeichenfolge im jeweiligen Element der unscharfen Menge enthalten sind. Das Ergebnis ist ein Überdeckungsgrad auf dem Intervall [0; 1].

Am Beispiel der Eingabezeichenfolge „Kallsruh Hbf“ ergäben sich die Trigramme

KAL ALL LLS **LSR** **SRU** **RUH** **HBF**

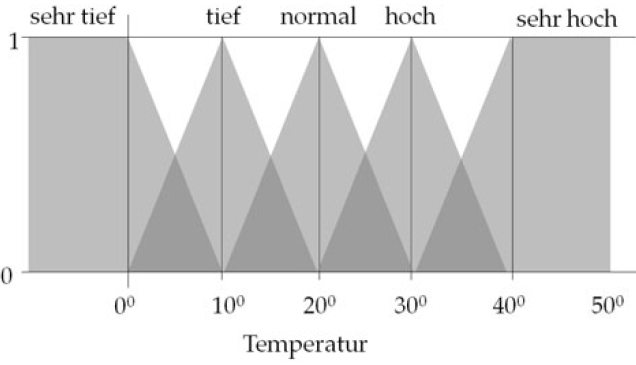
wovon die fett markierten Trigramme auch im entsprechenden Element der unscharfen Menge enthalten sind. Durch den entstehenden Überdeckungsgrad > 0 kann mit einer gewissen Wahrscheinlichkeit davon ausgegangen werden, dass mit der Eingabezeichenfolge eigentlich das korrespondierende Element der unscharfen Menge gesucht ist, wenngleich keine hundertprozentige Übereinstimmung mit der Eingabezeichenfolge vorliegt.

Die folgende Abbildung zeigt die graphische Darstellung einer klassischen und einer unscharfen Menge am Beispiel von Temperaturen im Vergleich:

**Abbildung 1:** Darstellung einer klassischen und einer unscharfen Menge

(Quelle: Nahrstedt 2012, S. 237)

Am Beispiel der Temperaturen werden die Klassifikationen „tief“, „normal“ und „hoch“ durch Verwendung der unscharfen Mengen nicht exakt abgegrenzt, sondern bilden das tatsächliche Empfinden verschiedener Menschen ab. So liegt eine „niedrige“ Temperatur bei einer Person schon bei 15°, bei einer anderen Person aber erst bei 5°.

Diese Überdeckung von verschiedenen Temperaturempfinden ist in der folgenden Abbildung graphisch dargestellt:

**Abbildung 2:** Graphische Darstellung der Überdeckung unscharfer Mengen

(Quelle: Nahrstedt 2012, S. 238)

Dasselbe Prinzip lässt sich auf die Suche nach Haltestellennamen anwenden. Die Zerlegung der Eingabezeichenfolge und der tatsächlichen Haltestellennamen bilden mehrere unscharfe Mengen, die sich zu einem gewissen Grad überdecken können und damit eine teilweise Übereinstimmung feststellbar machen. Wäre die Suche hingegen in einer klassischen Menge durchgeführt worden, hätte dies zu keinem Ergebnis geführt, da die Eingabezeichenfolge mit keinem der Elemente exakt übereinstimmt. Dieses Verfahren wird auch als *unscharfe Suche* oder *Fuzzy-Suche* bezeichnet und ist eine bekannte KI-Anwendung (vgl. Engfer 2002, S. 3).

Die erste Bedingung, um von einer KI gemäß der zuvor festgelegten Arbeitsdefinition ausgehen zu können ist damit erfüllt. Die Qualität der Ergebnisse der KI ist im Nachhinein nicht von der Qualität der Ergebnisse eines Menschen mit entsprechendem Fachwissen zu unterscheiden. Folglich darf beim verwendeten Beispiel von einer Form der KI ausgegangen werden.

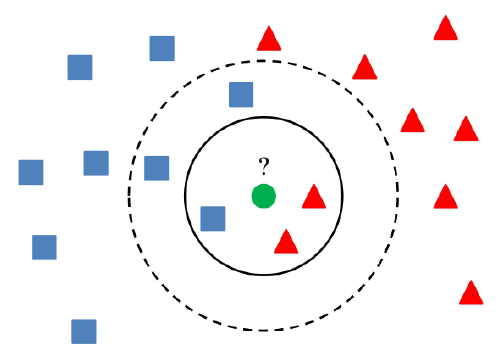
**Kommen hierbei Ansätze aus dem ML zum Einsatz?**

Um diese Frage beantworten zu können, ist es zunächst nötig, sich mit den Grundlagen des ML zu befassen. Grundsätzlich ist ML eine Möglichkeit, die „Programmen die Möglichkeit gibt, mit Hilfe von Daten zu lernen, ohne explizit programmiert zu werden“ (Niebler 2018, S. 10). Insgesamt wird im ML zwischen dem *überwachten Lernen* und dem *unüberwachten Lernen* unterschieden.

Beim überwachten Lernen werden in *Trainingsdaten* enthaltene Eigenschaften, sogenannte *Features,* extrahiert und einem Ergebnis, dem sogenannten *Label,* zugeordnet. Um zu verhindern, dass das Modell ausschließlich mit den Trainingsdaten funktioniert, wird das Modell nach Abschluss des Trainings auf einen *Testdatensatz* angewandt. Auf diesem Weg kann die Performance ermittelt werden, welche das Modell mit unbekannten Daten erreicht. Die Überanspassung des Modells an die Trainingsdaten wird auch als *Overfitting* bezeichnet (vgl. Niebler 2018, S. 11).

Das unüberwachte Lernen verarbeitet hingegen grundsätzlich unbekannte Daten mit dem Ziel, bislang unbekannte Zusammenhänge innerhalb dieser Daten zu finden oder Inhalte nach verschiedenen Merkmalen zu ordnen. Ein nachgelagerter Test mit Testdaten erfolgt nicht. (vgl. Niebler 2018, S. 11).

Das konstruierte Beispiel kann anhand dieser Definitionen nicht exakt eingeordnet werden. Tendenziell trifft die Beschreibung des überwachten Lernens eher zu. Die Nutzung des Systems kann dabei als fortlaufendes Training betrachtet werden. Die Trigramme der Eingabezeichenfolge werden dem jeweils von den Nutzenden tatsächlich gewählten Suchvorschlag zugeordnet. Die Kombination aus Trigrammen entspricht dabei den Features, der tatsächlich gewählte Suchvorschlag dem Feature. Das typischerweise beim überwachten Lernen folgende Testing entfällt an dieser Stelle allerdings. Die verarbeiteten Daten werden direkt gespeichert und beim nächsten Suchvorgang weiterverarbeitet. Eine Überanspassung ist dann möglich, wenn die Trigramme der Eingabezeichenfolge bereits aus anderen Eingaben enthalten und dadurch einem anderen Suchvorschlag zugeordnet sind als dem eigentlich gewünschten. Entsprechend könnte dem Beispiel auch der k-Nearest-Neighbours Algorithmus (KNN) zu Grunde liegen, welcher dem unüberwachten Lernen zuzuordnen ist (vgl. Niebler 2018, S. 11). Der KNN ordnet Daten dabei anhand ihrer Features in Dimensionen ein und sucht entsprechend die k Nachbarn, deren Features am nächsten am Ausgangsdatensatz liegen. Anhand dieser k Nachbarn wird dann versucht, eine Aussage über den Ausgangsdatensatz zu treffen.

Die folgende Abbildung verdeutlicht die Funktionsweise des KNN anhand einem Datensatz über den eine Aussage basierend auf den nächsten k Datensätzen aus zwei verschiedenen Klassen getroffen werden soll:

**Abbildung 3:** Graphische Darstellung des k-Nearest-Neighbours Algorithmus

(Quelle: Alaliyat 2022, S. 38)

In grün dargestellt ist der Datensatz, über den eine Aussage getroffen werden soll. In rot und blau respektive als Dreieck und als Quadrat dargestellt, sind die bereits klassifizierten Datensätze. Der KNN ermittelt nun die k Datensätze, die am nächsten liegen und verwendet diese als Grundlage für eine Aussage über den neuen Datensatz (vgl. Alaliyat 2022, S. 38). Je größer k ist, desto größer ist im Regelfall auch der Suchraum.

Zusammenfassend lässt sich festhalten, dass

Literaturverzeichnis

Alaliyat, Saleh (2022): Video -based Fall Detection in Elderly’s Houses.

Engfer, Andreas (2002): Fuzzy Logik. Ausarbeitung zur Vorlesung Ausarbeitung zur Vorlesung „Methoden der Künstlichen Intelligenz“. Ausarbeitung. Fachhochschule Furtwangen, Furtwangen. Fachbereich Wirtschaftsinformatik.

Krämer, Andreas (2022): exeo OpinionTRAIN. 9-Euro-Ticket: Blick zurück und nach vorne - Nutzerprofil, Nutzung und Bewertungen. exeo Strategic Consulting AG. Bonn, 08.09.2022. Online verfügbar unter https://www.researchgate.net/publication/363417545\_exeo\_OpinionTRAIN\_IV\_9\_EUR\_Ticket\_Post\_I\_220907, zuletzt geprüft am 10.09.2022.

Liggieri, Kevin; Müller, Oliver (2019): Mensch-Maschine-Interaktion. Stuttgart: J.B. Metzler.

Mainzer, Klaus (2019): Künstliche Intelligenz – Wann übernehmen die Maschinen? Berlin, Heidelberg: Springer Berlin Heidelberg.

Nahrstedt, Harald (2012): Algorithmen für Ingenieure. Wiesbaden: Vieweg+Teubner Verlag.

Niebler, Paul (2018): Datenbasiert Entscheiden. Ein Leitfaden Für Unternehmer und Entscheider. Unter Mitarbeit von Dominic Lindner. Wiesbaden: Gabler (Essentials Ser). Online verfügbar unter https://ebookcentral.proquest.com/lib/kxp/detail.action?docID=5566817.

Rimscha, Markus von (2008): Algorithmen kompakt und verständlich. Lösungsstrategien am Computer. 1. Aufl. Wiesbaden: Vieweg + Teubner (Programmiersprachen, Datenbanken und Softwareentwicklung).

Searle, John R. (1980): Minds, brains, and programs. In: *Behav Brain Sci* 3 (3), S. 417–424. DOI: 10.1017/S0140525X00005756.

# Anhang