

# 1 Einleitung

## 1.1 Ausgangssituation und Problemstellung

Text-Mining umfasst die systematische Untersuchung von unstrukturierten textbasierten Daten (s. BIEMANN ET AL. 2022a, S. 3). Im Zentrum des Text-Minings steht die Verarbeitung und Analyse von Textdaten, um implizites Wissen (z. B. Muster, Trends und relevante Informationen) herauszufiltern, die in den rohen Textdaten verborgen sind (s. JO 2019, S. 3).

Die Technik wird zunehmend in komplexen Untersuchungen wie der Recherche zu den Panama Papers, einer Enthüllung von Offshore-Finanzaktivitäten, eingesetzt. (vgl. SRIVASTAVA U. SINGH 2018). Nach MUNZINGER spielte Text-Mining eine entscheidende Rolle, indem es durch die systematische Analyse und Auswertung großer Mengen von Textdaten die Identifikation beteiligter Personen und Unternehmen ermöglichte. Zusätzlich erlaubte es das Clustering von Dokumenten nach thematischen Schwerpunkten, was effektive Ausgangspunkte für vertiefende Recherchen schuf. (s. MUNZINGER 2018, S. 1ff.) In Anwendungsbereichen wie dem investigativen Journalismus ist Text-Mining bereits weit verbreitet.

In produzierenden Unternehmen ist das Potenzial von Textdaten, die einen Großteil der unstrukturierten Daten dieser ausmachen, bisher weitestgehend ungenutzt geblieben (s. ANSARI ET AL. 2021, S. 373). Ein Grund liegt darin, dass Unternehmen davon profitieren, sich durch das Outsourcing auf ihr Kerngeschäft zu fokussieren (s. ADKISON 2000, S. 69; PAUDEL U. SHARMA 2020, S. 37). Unterstützende Funktionen, wie zum Beispiel die IT-Abteilung wurden deshalb in der Vergangenheit in großem Umfang ausgelagert (s. LACITY ET AL. 2009, S. 130). Diese verfügen über Grundlagen in den Bereichen Mathematik, Statistik und IT, die heute für die Umsetzung von datenbasierten Anwendungsfällen, inklusive Text-Mining-Anwendungsfällen, fehlen (vgl. WICHT ET AL. 2021, S. 270ff.). Dadurch besteht ein Mangel an Kenntnissen zur Identifizierung eigener Anwendungsfälle. Fachbereiche konzentrieren sich auf ihr Kerngeschäft. Ihnen fällt es schwer darüber hinaus technisch unterstützte Lösungen für bestehende Herausforderungen zu identifizieren. Darüber hinaus müssen bei der Lösungsentwicklung spezifische Anforderungen für ein Lastenheft abgeleitet werden (s. HICKING U. VÖLKEL 2022, S. 250). In Bezug auf unbekannte Technologien wie das Text-Mining birgt dies für produzierende Unternehmen die Gefahr von Projektmisserfolgen durch unzureichende Anforderungsdefinition (s. FUCHS U. SAUER 2016, S. 177).

Angesichts dieser Situation haben produzierende Unternehmen verschiedene Optionen, um die Nutzung von Textdaten in ihren Betriebsabläufen oder zur Verbesserung ihrer Produkte zu realisieren. Die Weiterentwicklung der Fähigkeiten der bestehenden Mitarbeiter durch Schulungen und Weiterbildungen sind aufgrund des fehlenden Vorwissens der Mitarbeiter in Bereichen wie IT, Mathematik und Statistik nicht reali-

sierbar (s. PFEIFFER 2020, S. 476ff.). Produzierende Unternehmen haben darüber hinaus die Möglichkeit, entweder Mitarbeiter zu Rekrutieren oder individuelle Text-Mining-Anwendungen von externen Dienstleistern entwickeln zu lassen (vgl. KÖNINGER 2021). Sowohl bei der Auswahl externer Dienstleister als auch neuem Personal müssen produzierende Unternehmen deren Fähigkeiten bewerten und herausfinden, inwieweit der Dienstleister oder Bewerber die Fähigkeiten besitzen, ihren Anwendungsfall umsetzen zu können (vgl. HEß ET AL. 2020, S. 181). Diese Bewertung funktioniert nach STOYAN, wenn ein Mindestmaß an Fähigkeiten im eigenen Unternehmen vorhanden sind. Fehlen die Fähigkeiten in Gänze, ist die Beurteilung herausfordernd. (s. STOYAN 2004, S. 125) Folglich müssen produzierende Unternehmen die eingesetzten Text-Mining-Methoden für ihre Anwendungsfälle kennen, um die Fähigkeitslücken zufriedenstellend zu schließen.

## 1.2 Zielsetzung

In dieser Dissertation wird die in Kapitel 1.1 beschriebene Problemstellung untersucht, die sich mit der Umsetzung von Text-Mining-Anwendungsfällen in produzierenden Unternehmen befasst. Ziel des Vorhabens ist es daher, die Schließung der bestehenden Fähigkeitslücke in Bezug auf die Entwicklung von Text-Mining-Anwendungen methodisch zu unterstützen. Adressaten der vorliegenden Dissertation sind demnach Geschäftsführer, Fachbereichsleiter und IT-Leiter in Großunternehmen sowie Geschäftsführer in kleinen und mittelständischen Unternehmen. Diese werden in der Konzeptions- und Planungsphase für die Entwicklung von Text-Mining-Anwendungen unterstützt. Die Ergebnisse der Dissertation ermöglichen es, aus generischen Anwendungsfällen die relevanten Anforderungen für die Entwicklung individueller Text-Mining-Anwendungen abzuleiten. Darüber hinaus unterstützen die Ergebnisse, die dafür relevanten Text-Mining-Methoden zu bestimmen. Mit deren Hilfe werden anschließend Fähigkeiten sowie Fähigkeitslevel abgeleitet. Ausgangspunkt sind Prozesse oder Produkte eines Unternehmens, für die eine Text-Mining-Anwendung entwickelt werden soll. Aufbauend auf der identifizierten Problemstellung und der Zielsetzung der Dissertation ergibt sich folgende Forschungsfrage:

Wie können Text-Mining-Anwendungsfälle in produzierenden Unternehmen systematisch umgesetzt werden?

Zur Beantwortung der Forschungsfrage werden in dieser Dissertation die nachfolgenden Unterrforschungsfragen beantwortet:

- Wie können Text-Mining-Anwendungsfälle in produzierenden Unternehmen beschrieben werden?
- Wie können Methoden zur Umsetzung von Text-Mining-Anwendungsfällen beschrieben werden?
- Wie können aus den Text-Mining-Anwendungsfällen Wirkbeziehungen zu den Text-Mining-Methoden hergestellt werden?

### **1.3 Wissenschaftstheoretische Einordnung und Struktur der Untersuchung**

Das gegenwärtige Dissertationsthema befasst sich mit ingenieurwissenschaftlichen, betriebswirtschaftlichen und wirtschaftsinformatischen Themenstellungen, die eine wissenschaftstheoretische Verortung erfordern. Die wissenschaftstheoretische Kategorisierung basiert auf der von ULRICH U. HILL dargelegten Unterscheidung zwischen Formal- und Realwissenschaften. Die Klassifizierung ermöglicht erst die detaillierte Analyse eines spezifischen, einer wissenschaftlichen Kategorie zugehörigen Problems (vgl. ULRICH U. HILL 1976, S. 305). Formalwissenschaften fokussieren sich auf die Konstruktion von Sprachen und befassen sich mit Zeichensystemen sowie den dahinterliegenden Regeln für den Gebrauch von Symbolen. Zu den Formalwissenschaften gehören Philosophie, Mathematik und Logik. Realwissenschaften beschäftigen sich mit der Beschreibung, Erklärung und Gestaltung von in der Realität existierenden Objekten, Ereignissen und Systemen (vgl. ULRICH U. HILL 1976, S. 305). Abhängig vom verfolgten Ziel innerhalb der Realwissenschaften unterscheidet man zwischen Grundlagen- und Handlungswissenschaften. Die Grundlagenwissenschaft, als reine Wissenschaft, strebt nach theoretischer Erkenntnis und zielt darauf ab, Zusammenhänge zwischen realen Objekten zu erklären (vgl. ULRICH U. HILL 1976, S. 305). Im Gegensatz dazu stehen praktische Ziele im Fokus der Handlungswissenschaften. Angewandte Wissenschaften, insbesondere in den Ingenieurwissenschaften, streben nach der Gestaltung technischer Systeme. Der Anspruch an die Realitätsnähe der angewandten Wissenschaften manifestiert sich in ihrer praktischen Anwendbarkeit. Handlungsempfehlungen, Entscheidungsunterstützung und Anleitungen sind Ergebnisse, die zur Gestaltung der Realität beitragen (vgl. DEINDL 2013, S. 5; LEHNER ET AL. 1995, S. 20). Während die Formalwissenschaften sich mit der logischen Wahrheit eines Konstrukts befassen und keinen Bezug zur Realität haben, müssen die Ergebnisse der Realwissenschaften faktisch begründet sein (vgl. ULRICH U. HILL 1976, S. 305f.). Die Disziplinen der Ingenieur- und Betriebswissenschaften, sowie die Wirtschaftsinformatik, sind der Realwissenschaft, insbesondere der angewandten Wissenschaft, zuzuordnen (vgl. LEHNER ET AL. 2008; ABTS U. MÜLDER 2017, S. 5ff.). Basierend auf der in Kapitel 1.2 dargelegten Zielsetzung wird ein Praxisproblem im Kontext der Umsetzung von Text-Mining-Anwendungsfällen identifiziert. Somit ist die vorliegende Dissertation dem Bereich der angewandten Handlungswissenschaften zuzuordnen.

Diese Kategorisierung führt zur Notwendigkeit, die Herausforderungen der Subjektivität und Kommunikation zu meistern. Erstere bezieht sich darauf, wie individuelle Erfahrungen und Kenntnisse einen Wahrnehmungsfilter bilden, der die selektive Wahrnehmung von Realitätsausschnitten ermöglicht (vgl. ULRICH U. HILL 1976, S. 306). Letztere bezieht sich darauf, dass Erkenntnisse, die nicht mit präzisen Begriffen artikuliert werden können, als gegenstandslos und daher unberücksichtigt gelassen werden müssen. Der Abstraktionsgrad ist so zu wählen, dass Erkenntnisse an Klarheit gewinnen und gleichzeitig in präziser Sprache formuliert werden. Diese beiden

Herausforderungen sind im Kontext des Forschungsaspekts des Entdeckungszusammenhangs zu bewältigen. Dieser Zusammenhang repräsentiert den konzeptionellen Rahmen eines wissenschaftlichen Forschungsprozesses und stellt als heuristisches Problem nicht den Anspruch, durch analytische Verfahren die ideale Lösung zu finden, sondern vielmehr durch die Reduktion der Untersuchungsdauer eine praktikable Lösung zu präsentieren, basierend auf dem Kriterium der Zweckmäßigkeit (vgl. ULRICH u. HILL 1976, S. 305f.).

Um heuristische Probleme zu adressieren, wird der systemtheoretische Ansatz nach ULRICH u. HILL herangezogen. Dieser interdisziplinäre Ansatz berücksichtigt nicht die reine Theorie, sondern zusätzlich die praktische Anwendung. Er ist relevant für ingenieurwissenschaftliche Fragestellungen, da er auf Geschlossenheit und Mathematisierung verzichtet. Durch seine terminologische Funktion basierend auf einem interdisziplinären Begriffssystem, kann das Kommunikationsproblem gelöst werden. Die heuristische Funktion bietet notwendige Strukturmodelle zur Beantwortung der Forschungsfragen, während die Integrationsfunktion das Einbeziehen verschiedener Einflussfaktoren und Variablen im Kontext einer interdisziplinären Betrachtung ermöglicht (vgl. ULRICH u. HILL 1976, S. 308). Aufgrund seiner Interdisziplinarität, Offenheit und Praxisnähe wird dieser Ansatz in diesem Dissertationsvorhaben angewendet.

Da das skizzierte Problem dem Bereich der angewandten Wissenschaften zugeordnet ist, folgt es dem von ULRICH u. HILL beschriebenen Forschungsprozess. Dieser umfasst terminologisch-deskriptive (Festlegung und Anwendung eines einheitlichen Begriffssystems), empirisch-induktive (Erkenntnisgewinn durch Schlussfolgerungen aufgrund des Aufdeckens von Korrelationen) und analytisch-deduktive (logische Konstruktion von Modellen) Forschungsaktivitäten (vgl. ULRICH u. HILL 1976, S. 347f.). Um die Herausforderung zu bewältigen, abstrahierte Aussagen aus Einzelbeobachtungen induktiv abzuleiten, folgt das Dissertationsvorhaben dem Ansatz des kritischen Rationalismus nach POPPER. Der kritische Rationalismus zielt darauf ab, durch Induktion gewonnene Erkenntnisse zu begründen (vgl. POPPER 2010, S. 17). Dazu wird nicht die logische, sondern die faktische bzw. empirische Wahrheit überprüft. Eine Hypothese gilt als wahr, bis sie durch Einzelbeobachtungen falsifiziert wird (vgl. ULRICH u. HILL 1976, S. 345f.).

Der Forschungsprozess wird als iterativer Lernprozess durchgeführt, bei dem kontinuierliche Fragen an die Realität gestellt und zu beantworten versucht werden. Gewonnene Erkenntnisse werden kritisch hinterfragt, was zu angepassten Fragestellungen führt und den Prozess erneut in Gang setzt (vgl. KUBICEK 1977, S. 13ff.). Dieser Lernprozess basiert auf Vorwissen und Erfahrung (vgl. KUBICEK 1977, S. 19ff.). Das initierende Praxisproblem wird unter Einbeziehung eines gemischten Kreises aus Wissenschaft und Praxis im Sinne der partizipativen Aktionsforschung gelöst (vgl. KUBICEK 1977, S. 282). Erstens werden die Ansprechpartner in Beratungsprojekten in den Forschungsprozess einbezogen. Zweitens wird die Expertise von Partnern der Forschungsprojekte „VoBAKI“ (s. A.2.2) und „Techrad“ (s. A.7) für die Erarbeitung der Ergebnisse dieser Dissertation genutzt.

Gemäß dem Aufbau von ULRICH u. HILL orientiert sich die Kapitelstruktur an den Forschungsaktivitäten (Abbildung 1-1). Kapitel 1 beschreibt die Ausgangssituation, Problemstellung und Zielsetzung. Kapitel 2 umfasst die Beschreibung der verwendeten Terminologie, Begriffsdefinitionen und die Abgrenzung des Untersuchungsbereichs. Kapitel 3 legt den aktuellen Stand der Erkenntnisse dar. Die Forschungslücke wird anhand der bewerteten Erkenntnisse aufgezeigt. In Kapitel 4 werden die detaillierte Vorgehensweise und deren methodische Bestandteile beschrieben. Kapitel 5 führt zur Identifikation von Text-Mining-Anwendungsfällen in produzierenden Unternehmen und deren Beschreibung. Kapitel 6 beschreibt aktuell verfügbare Text-Mining-Methoden in strukturierter Form. Die Ergebnisse der Beschreibungsmodelle werden in Kapitel 7 zusammengeführt, um Wirkungszusammenhänge zwischen den einzelnen Bestandteilen zu identifizieren. Kapitel 8 beinhaltet die eingeschränkt empirisch-induktive Evaluation der Ergebnisse durch Anwendung in zwei Fallstudien. Die Untersuchung wird in Kapitel 9 mit einer Zusammenfassung und einem Ausblick abgeschlossen.

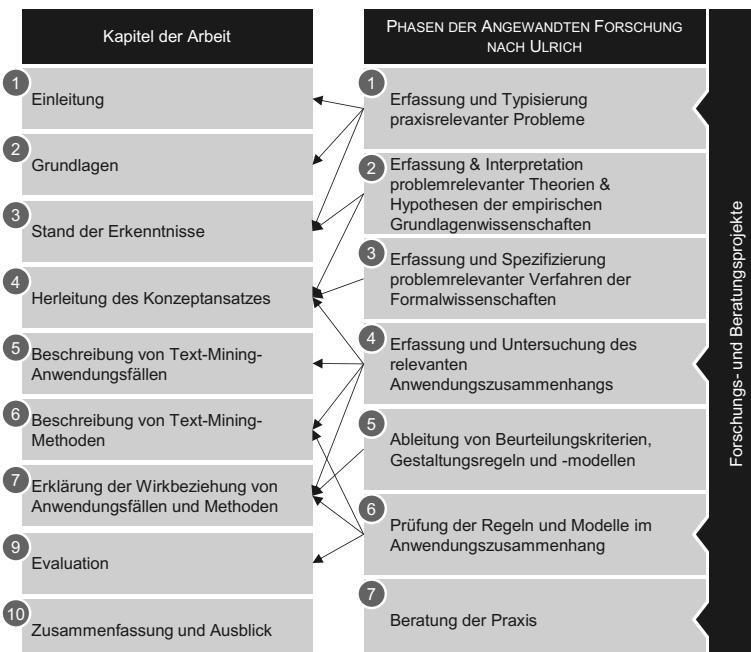


Abbildung 1-1: Gliederung der Arbeit und angewendete Forschungsaktivitäten (eigene Darstellung)