

IfM-Materialien

Bedrohungen und Chancen frühzeitig erkennen –
Entwicklung eines Früherkennungskonzepts

Rodi Akalan, Siegrun Brink, Annette Icks und Hans-Jürgen Wolter

Impressum

Herausgeber

Institut für Mittelstandsforschung Bonn
Maximilianstr. 20, 53111 Bonn

Telefon +49/(0)228 / 72997 - 0
Telefax +49/(0)228 / 72997 - 34

www.ifm-bonn.org

Ansprechpartner

Rodi Akalan
Siegrun Brink
Annette Icks
Hans-Jürgen Wolter

IfM-Materialien Nr. 303

ISSN 2193-1852 (Internet)
ISSN 2193-1844 (Print)

Bonn, 08.12.2023

Das IfM Bonn ist eine Stiftung des privaten Rechts.

Gefördert durch:



Bundesministerium
für Wirtschaft
und Klimaschutz

aufgrund eines Beschlusses
des Deutschen Bundestages

Ministerium für Wirtschaft,
Industrie, Klimaschutz und Energie
des Landes Nordrhein-Westfalen



Bedrohungen und Chancen frühzeitig erkennen – Entwicklung eines Früherkennungskonzepts

Early detection of threats and opportunities – Developing an early detection concept

Rodi Akalan, Siegrun Brink, Annette Icks und Hans-Jürgen Wolter

IfM-Materialien Nr. 303

Zusammenfassung

Der Mittelstand ist derzeit mit vielfältigen Krisen konfrontiert. Eine frühe Erkennung relevanter Herausforderungen und Chancen ermöglicht es den mittelständischen Unternehmen und der Wirtschaftspolitik, sich darauf vorzubereiten und die geeigneten Rahmenbedingungen zu setzen. Gegenwärtig erfolgt die Früherkennung zumeist anhand von Konjunkturindikatoren, die i.d.R. anhand konkreter Zahlenwerte Rückschlüsse auf die zukünftige Wirtschaftsentwicklung ziehen. Eine systematische Auswertung wirtschaftsrelevanter Textdaten erfolgt nicht. Hier setzt das in der vorliegenden Studie entwickelte innovative Früherkennungskonzept an, das KI-gestützt Textdaten aus Medien und Wirtschaft effizient analysiert und Themen extrahiert. Mithilfe von Praxistests zeigen wir, dass das Konzept zuverlässig funktioniert und relevante Themen frühzeitig erkennen kann.

Schlagwörter: *Früherkennung, Themen, Topic Modeling, Maschinelles Lernen*

Abstract

Currently, SMEs are confronted with several crises. The early detection of relevant threats and opportunities enables SMEs and economic policymakers to be prepared and set the appropriate framework conditions. Early detection is often based on economic indicators that develop forecasts with the help of structured data (usually numbers). Moreover, information from text data is not analysed systematically in current economic indicators. Hence, the present study develops an innovative early detection concept that efficiently analyses text data from the media and economy. The AI-based concept can analyse large amounts of text data and extract relevant topics. With the help of multiple tests with real data, we show that the concept delivers reliable results and can recognise relevant topics at an early stage.

JEL: M20, O10

Keywords: *early detection, topics, topic modeling, machine learning*

Inhalt

Verzeichnis der Abbildungen	II
Verzeichnis der Übersicht	II
1 Einleitung	1
2 Definition und Analyse existierender Früherkennungsindikatoren	3
3 Methodisches Vorgehen bei der Entwicklung eines Früherkennungskonzepts	7
4 Praxistest des Früherkennungskonzepts	11
5 Fazit	17
Literatur	18
Anhang	21

Verzeichnis der Abbildungen

Abbildung 1:	Merkmale zur Klassifikation von Wirtschaftsindikatoren	3
Abbildung 2:	Verteilung der Wirtschaftsindikatoren auf Themenbereiche	4
Abbildung 3:	Entwicklungsschritte für das Früherkennungskonzept	8
Abbildung 4:	Topic Modeling-Ergebnisse der Unternehmensdaten	12
Abbildung 5:	Topic Modeling-Ergebnisse der Verbandspressemitteilungen	13
Abbildung 6:	Topic Modeling-Ergebnisse der Nachrichtentitel	14
Abbildung 7:	Intensität der Berichterstattung zum Thema "Energie"	15
Abbildung 8:	Monatliche Veränderung der Intensität der Berichterstattung zum Thema "Energie"	16

Verzeichnis der Übersicht

Übersicht 1:	Auswahl an Konjunkturindikatoren	21
--------------	----------------------------------	----

1 Einleitung

Mittelständische Unternehmen finden sich zunehmend im komplexen Spannungsfeld sich überlagernder Entwicklungen wieder. Krisen wie die Corona-Pandemie, die Energiekrise oder geopolitische Konflikte bedrohen die wirtschaftliche Entwicklung von Unternehmen (vgl. Brink/Icks 2022). Auf der anderen Seite können z.B. infolge stetiger Transformationsprozesse im Falle einer erfolgreichen Umgestaltung der Wirtschaftsstruktur durchaus neue Chancen für Unternehmen erwachsen. Beispielhaft genannt sei hier die Entwicklung des Ruhrgebietes nach Ende der Montanindustrie (vgl. Goch 1999). Veränderungen des Umfelds – seien sie positiver oder negativer Natur – können durch externe Schocks auftreten, die das gesellschaftliche, politische oder wirtschaftliche System erheblich und langfristig stören (vgl. Schubert/ Klein 2020). Beispiele hierfür sind die globale Finanzkrise ab 2007 oder die Corona-Pandemie ab 2020 (vgl. Busch-Heizmann et al. 2022). Alternativ können anstehende Veränderungen zunächst ein Randthema sein, um dann über einen längeren Zeitraum immer mehr an Bedeutung zu gewinnen (z.B. Strukturwandel, Fachkräftemangel). Die frühzeitige Identifikation derartiger Entwicklungen ist einerseits für die mittelständischen Unternehmen wichtig, damit sie sich rechtzeitig auf anstehende Veränderungen vorbereiten können. Sie ist aber andererseits auch Voraussetzung dafür, dass die Politik die Ausgestaltung der Mittelstandspolitik antizipativ anpassen kann.

Auf einzelwirtschaftlicher Ebene versuchen deshalb immer mehr Unternehmen, ihre Resilienz gegenüber Umfeldveränderungen zu stärken (vgl. Schwartz/ Gerstenberger 2020). Neben aktivem Risikomanagement bedeutet dies, Krisensituationen frühzeitig zu erkennen, um darauf zu reagieren. Dabei nutzen Unternehmen bereits verschiedene Methoden zur Krisenfrüherkennung, wie beispielsweise die Kennzahlenanalyse (vgl. Giesen 2022). Auch auf gesamtwirtschaftlicher Ebene werden häufig makroökonomische Kennzahlen verwendet, um Entwicklungen frühzeitig zu identifizieren (vgl. Hinze 2003). So wird beispielsweise die Zinsstrukturkurve als ein Indikator für die konjunkturelle Entwicklung der Wirtschaft genutzt (vgl. Klimonczyk 2016). Bestehende einzelwirtschaftliche und gesamtwirtschaftliche Methoden zur Krisenfrüherkennung stellen zumeist auf eine Kennzahlenanalyse auf aggregierter Ebene ab, ohne Entwicklungen konkret zu benennen. Hier setzt die vorliegende Studie an und stellt einen innovativen Ansatz zur frühzeitigen Erkennung von mittelstands- und mittelstandspolitisch relevanten Themen vor. Dabei greifen wir auf das sog. Text

Mining, eine Vorgehensweise zur effizienten Analyse großer Textmengen, zurück (vgl. Tan 1999).

2 Definition und Analyse existierender Früherkennungsindikatoren

Angesichts der gesamtwirtschaftlichen Konsequenzen wirtschaftlicher Krisen wie der Finanzkrise 2008/2009 oder auch der COVID 19-Pandemie 2020/2021 stellt sich die Frage, wie derartige Entwicklungen frühzeitig erkannt werden können. Ein wichtiges Instrument sind dabei Wirtschaftsindikatoren, mit deren Hilfe ökonomische Entwicklungen eingeschätzt und mögliche Trendveränderungen erkannt werden sollen. Solche Indikatoren lassen sich anhand verschiedener Kriterien klassifizieren (vgl. Abbildung 1).

Abbildung 1: Merkmale zur Klassifikation von Wirtschaftsindikatoren



© IfM Bonn 23 2301 001

Quelle: IfM Bonn 2023.

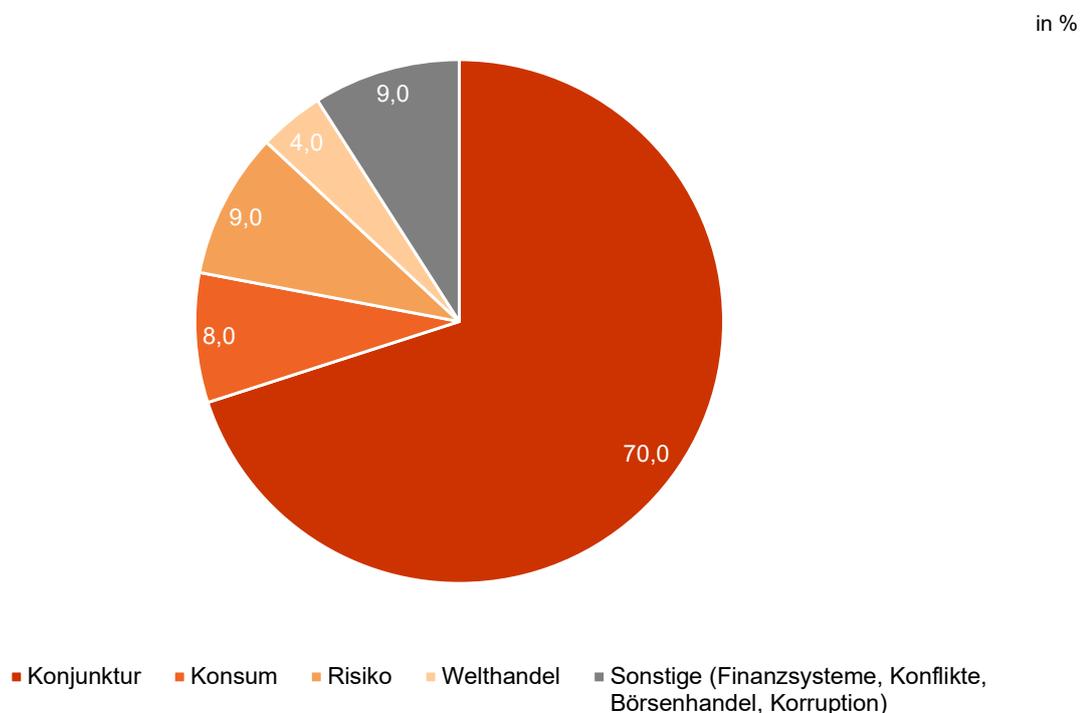
Grundsätzlich kann zwischen Früh-, Präsenz- und Spätindikatoren unterschieden werden (vgl. Giesen 2002). Während die beiden Letzteren gegenwärtige bzw. zurückliegende Entwicklungen erfassen, versuchen Frühindikatoren zukünftige Tendenzen zu antizipieren (vgl. Krystek 2007), weshalb sie für uns von besonderem Interesse sind.

Früherkennung bezeichnet dabei die Möglichkeit, "latente Bedrohungen und Chancen bereits in einem sehr frühen Stadium ihrer Entwicklung zu erkennen" (vgl. Krystek 2007, S. 50). Dabei sind zwei Merkmale für eine Früherkennung charakteristisch: Zum einen wird auf "latente" Entwicklungen abgestellt, die

meist verdeckt entstehen und nicht unmittelbar wahrnehmbar sind. Zum anderen ist die Frühzeitigkeit ein wichtiges Merkmal. Die Früherkennung soll folglich eine gewisse Vorhersagekraft für zukünftige Entwicklungen besitzen.

Bei der Früherkennung kommen verschiedene Informationssysteme (sog. Früherkennungssysteme) zum Einsatz, die sich grundsätzlich in zwei Arten unterteilen lassen (vgl. Krystek 2007): Zum einen sind hier kennzahlenorientierte Früherkennungssysteme zu nennen, die sich auf wirtschaftliche Kennzahlen und Hochrechnungen fokussieren (vgl. Giesen 2002). Wirtschaftliche Kennzahlen zeigen jedoch häufig vergangene, bereits eingetretene Entwicklungen an, so dass kaum eine Früherkennung latenter Entwicklungen möglich ist (vgl. Krystek/Müller-Stewens 1993). Neben dem kennzahlenorientierten Früherkennungssystemen wird zur Erfassung latenter Entwicklungen auch auf indikatorbasierte Früherkennungssysteme zurückgegriffen. Dabei wird versucht, mit Hilfe von Indikatoren nicht direkt wahrnehmbare Entwicklungen abzubilden (vgl. Hauff 2009, S.13). Das können konjunkturelle Entwicklungen, aber auch die Entwicklung von Risiken oder des privaten Konsums sowie des Welthandels sein (vgl. Abbildung 2).

Abbildung 2: Verteilung der Wirtschaftsindikatoren auf Themenbereiche



© IfM Bonn 23 2301 002

Quelle: Eigene Berechnungen basierend auf einer Literaturrecherche.

Als Datenquellen zur Berechnung der Indikatoren werden sowohl realwirtschaftliche (z.B. Arbeitslosenzahlen, BIP) als auch umfragebasierte Daten herangezogen (z.B. Einkaufsmanagerindex) (vgl. Benner/ Meier 2003). Zu den bekanntesten umfragebasierten Frühindikatoren für die konjunkturelle Entwicklung in Deutschland zählt der ifo-Geschäftsklimaindex des ifo-Instituts, der auf einer Befragung von über 9.000 Industrie-, Bau- und Einzelunternehmen beruht (vgl. ifo-Institut 2023). Dieser Indikator setzt sich als geometrisches Mittel aus der Beurteilung der aktuellen Lage und den Geschäftserwartungen der nächsten sechs Monate zusammen (vgl. ifo-Institut 2023). Neben dem ifo-Geschäftsklimaindex zählt auch der ZEW-Index – auch als ZEW-Konjunkturerwartungen bekannt – des Leibniz-Zentrums für Europäische Wirtschaftsforschung zu den wichtigsten Konjunkturindikatoren für Deutschland. Dieser Index basiert auf einer Expertenbefragung von 350 Finanzanalysten aus Banken, Versicherungen und großen Industrieunternehmen, die die aktuelle konjunkturelle Lage und die zukünftige Entwicklung aus gesamtwirtschaftlicher Sicht einschätzen (vgl. ZEW 2023).

Neben den umfragebasierten Frühindikatoren wie dem ifo-Geschäftsklimaindex und dem ZEW-Index kombinieren andere Indikatoren die Ergebnisse von Umfragen mit statistischen Daten. Die Kombination unterschiedlichster Informationen hat den Vorteil, dass diese Indikatoren weniger volatil sind (vgl. Hinze 2003). Zu diesen Indikatoren zählt unter anderem der von der Commerzbank herausgegebene Early-Bird-Index. Dieser setzt sich aus dem kurzfristigen Realzins, einem Indikator für die preisliche Wettbewerbsfähigkeit, und dem Einkaufsmanagerindex für das verarbeitende Gewerbe in den USA zusammen und berücksichtigt damit die wirtschaftliche Entwicklung Deutschlands und der US-Konjunktur (vgl. Rossen 2012; Fischer 2014). Neben Forschungsinstituten und Finanzdienstleistern geben auch internationale Organisationen oder staatliche Institutionen wie das Statistische Bundesamt eigene Wirtschaftsindikatoren heraus. Eine Übersicht über verschiedene Wirtschaftsindikatoren mit dem Themenfokus Konjunkturentwicklung, ihre Herausgeber und Datenquellen bietet Übersicht 1 im Anhang.

Es existiert eine Vielzahl von Wirtschaftsindikatoren zur frühzeitigen Identifikation wirtschaftlicher Entwicklungen. Ein wesentlicher Nachteil dieser Art der Früherkennung ist die Beschränkung der Analyse auf strukturierte Daten. Vor dem Hintergrund, dass 80 bis 95 % aller Informationen als unstrukturierte Daten (d.h. Daten in nicht maschinell lesbarem Format wie z.B. Textdaten) vorliegen (vgl. Berger et al. 2019; Gandomi/ Haider 2015), bleiben folglich eine Vielzahl

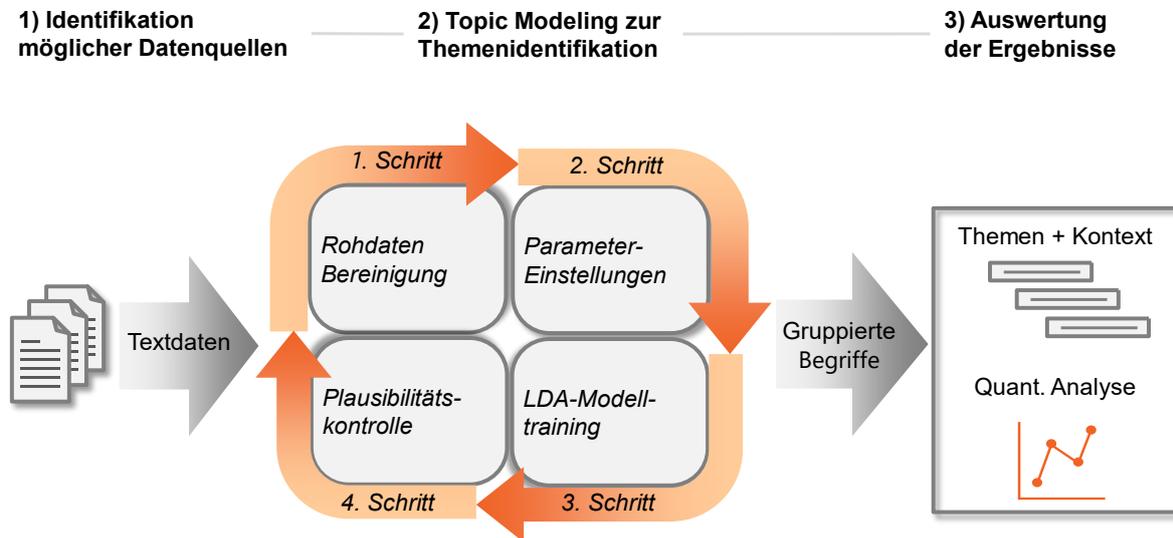
von Informationen ungenutzt. Des Weiteren können diese Indikatoren konstruktionsbedingt nur zur Prognose von wirtschaftlichen Entwicklungen genutzt werden. Zur Identifizierung konkreter Themen eignen sie sich, eben wegen der Nichtberücksichtigung von Textdaten, nicht. Das im Folgenden von uns entwickelte Früherkennungskonzept basiert daher auf Textdaten. Dadurch können konkrete Themen als Chancen und Bedrohungen für den Mittelstand identifiziert werden, die die zukünftige wirtschaftliche Entwicklung beeinflussen könnten und damit für Unternehmen und die Wirtschaftspolitik besonders wichtig sein werden.

3 Methodisches Vorgehen bei der Entwicklung eines Früherkennungskonzepts

Die Analyse existierender Wirtschaftsindikatoren hat aufgezeigt, dass unstrukturierte Daten in Textform bisher keine besondere Rolle bei der Früherkennung spielen. Dabei deuten erste Studien das große Informationspotential von Textdaten an. So hat ein Forschungsprojekt der Europäischen Zentralbank (EZB) gezeigt, dass Wirtschaftsnachrichten sowohl die aktuelle als auch teilweise die zukünftige Wirtschaftsentwicklung eines Landes abbilden können (vgl. Ferrari/Le Mezo 2020). Die EZB-Studie weist nach, dass US-Wirtschaftsnachrichten die, durch die Finanzkrise ausgelöste, Rezession im Jahr 2008 frühzeitig angezeigt haben. Auch weitere innovative Frühindikatoren nutzen Textdaten als Grundlage für ihre Vorhersagen. Eine Studie des Internationalen Währungsfonds (IWF) entwickelte basierend auf Nachrichtendaten in 20 Ländern einen Frühindikator für mögliche Finanzkrisen (vgl. Huang et al. 2019). Die OECD nutzte Textdaten von Unternehmenswebseiten, um die grüne Transformation von kleinen und mittleren Unternehmen zu messen (vgl. Ahrend/Menon 2023). Diese Studienbeispiele unterstreichen die Rolle von Textdaten als wertvolle Informationsquelle. Im Kontext der Mittelstandspolitik wurden diese Quellen jedoch bisher nicht systematisch genutzt. Im weiteren Verlauf dieser Studie wird überprüft, inwiefern die Analyse von Textdaten speziell bei der Früherkennung von mittelstands- und mittelstandspolitisch relevanten Themen behilflich sein kann.

Die computergestützte Analyse von Textdaten wird als "Text Mining" bezeichnet und hat zum Ziel, Informationen aus textbasierten Quellen zu extrahieren (vgl. Berger et al. 2019). Das Text Mining ist die Grundlage unseres Früherkennungskonzepts zur Themenidentifikation, welches im Rahmen von drei wesentlichen Schritten entwickelt wird (vgl. Abbildung 3).

Abbildung 3: Entwicklungsschritte für das Früherkennungskonzept



Quelle: IfM Bonn 2023.

(1) Identifikation möglicher Datenquellen

Eine wertvolle Informationsquelle sind die Medien, die in Studien bereits eine hohe Vorhersagekraft für Entwicklungen bewiesen haben (vgl. z.B. Ferrari/Le Mezo 2020). In Bezug auf den Mittelstand sind dabei insbesondere Wirtschaftsnachrichten und politische Nachrichten von Bedeutung. Die Medienlandschaft hat jedoch eine Vielzahl unterschiedlicher Akteure. So gibt es neben (unabhängigen) Journalisten, auch Lobbyverbände, Wirtschaftsinstitute, staatliche und nicht staatliche Organisationen, die Nachrichten (z.B. in Form von Pressemitteilungen) herausgeben. Jede Quelle kann zur Analyse genutzt werden und bietet das Potential, unterschiedliche Perspektiven in den Medien zu erfassen. Eine zweite wichtige Datenquelle stellen Unternehmensdaten in Form von Geschäftsberichten großer Unternehmen dar. Diese beinhalten einen Chancen- und Risikobericht (Teil des Lageberichts) in welchem die Unternehmen auf zukünftige Entwicklungen in ihren Geschäftsbereichen eingehen (vgl. Köhlbrandt et al. 2020). Aufgrund ihrer Größe müssen die meisten mittelständischen Unternehmen zwar keinen Chancen- und Risikobericht veröffentlichen. Dennoch können die angesprochenen Themen aus den Berichten der veröffentlichungspflichtigen Unternehmen auch für den Mittelstand relevant sein, da der Mittelstand oft in die Wertschöpfungs-

ketten größerer Unternehmen eingebunden ist. Schließlich ist als dritte Datenquelle die Wissenschaft zu nennen. Forschungsthemen können ein Frühindikator für neue Technologien und Innovationen in der Wirtschaft sein. Insgesamt bietet sich eine breite Datenbasis aus Medien-, Unternehmens-, und Wissenschaftsinformationen für das Text Mining an. Dies bietet zum einen unterschiedliche Perspektiven und ermöglicht auch einen Vergleich der Themenschwerpunkte. Die unterschiedlichen Daten können dabei entweder über Datenbanken oder durch Web Scraping (Extrahieren von Text aus Internetseiten) abgerufen werden.

(2) Topic Modeling zur Themenidentifikation

Im Rahmen des Text Minings bietet sich für unser Forschungsziel der Themenidentifikation die Analysemethode des Topic Modelings an, welche es ermöglicht, aus großen Textmengen frühzeitig Themenfelder zu extrahieren (vgl. Berger et al. 2019). Der dabei zugrunde liegende Algorithmus "Latent Dirichlet Allocation (LDA)" ist ein Wahrscheinlichkeitsmodell, der auf Worthäufigkeiten und auf dem gemeinsamen Auftreten von Wörtern in einem Dokument basiert (vgl. Blei et al 2003). So kann der LDA-Algorithmus bspw. in großen Textmengen das häufige gemeinsame Auftreten der Begriffe "Energie", "Strom", "Gas", "Energiepreise", "Stromkosten" basierend auf Wahrscheinlichkeitsberechnungen erkennen und diese entsprechend als ein zusammengehöriges Themenfeld identifizieren. Der LDA-Algorithmus gehört zu den Verfahren des unüberwachten Maschinellen Lernens – ein Teilbereich der künstlichen Intelligenz (vgl. Gump/Schneider 2021). Das LDA-basierte Topic Modeling ist ein iterativer Prozess, der aus vier Phasen besteht: In der ersten Phase müssen die Rohdaten der Textquelle für die Analyse vorbereitet werden, indem irrelevante Textelemente, wie Stoppwörter, Satzzeichen, Zahlen und HTML-Tags entfernt werden. Außerdem werden in diesem Schritt die Begriffe in Nomen, Verben und Adjektive kategorisiert (vgl. Berger et al 2019). In der zweiten Phase werden die LDA-Algorithmus Parameter eingestellt, die auf statistischen Berechnungen beruhen. Da der LDA-Algorithmus die Themenanzahl nicht selbstständig bestimmen kann, muss in dieser Phase zudem die vermutete Anzahl der Themenfelder in einer Datenquelle vorab festgelegt werden. Um die Anzahl der Themen näherungsweise bestimmen zu können, werden verschiedene Gütetests (wie z.B. der Coherence Score) genutzt (vgl. Stevens et al. 2012). In der dritten Phase findet das Modelltraining statt. Anhand von Worthäufigkeiten und Wahrscheinlichkeitsberechnungen werden Wörter innerhalb der Textdaten gruppiert (vgl. Blei et al. 2003). Das Ergebnis sind Wortgruppen, die wahrscheinlich zu

einem Themenbereich gehören. Diese werden in der vierten Phase einem Plausibilitätscheck unterzogen, der prüft, ob die Worte innerhalb einer Gruppe zusammen ein sinnvolles Thema ergeben. Je nachdem wie die Beurteilung ausfällt, muss eine weitere Iteration des Prozesses durchlaufen werden, um die Ergebnisse bspw. durch eine bessere Datenvorbereitung (Phase 1) oder modifizierte LDA-Parameter (Phase 2) zu verbessern. Im Endergebnis sollten die generierten Wortgruppen sinnvolle Themen ergeben, sich nicht zu stark überschneiden und inhaltlich die wesentlichen Informationen aus der Datenquelle abbilden.

(3) Auswertung der Ergebnisse

Im dritten Entwicklungsschritt folgt die Auswertung der Topic Modeling Ergebnisse. Hier werden zunächst die Wortgruppen in Themen eingeteilt. Anschließend kommen je nach Datenquelle eine qualitative oder eine quantitative Auswertung in Betracht. Im Rahmen der qualitativen Auswertung kann z.B. der Kontext der Themen näher untersucht werden. Dazu durchlaufen die Passagen mit Themenbezug erneut die Topic Modeling-Analyse und es wird untersucht mit welchen Verben und Adjektiven die Themen beschrieben werden, sodass ein positives, neutrales, oder negatives Stimmungsbild des Themenfeldes (Sentiment) erfasst wird. Als weitere qualitative Auswertungsmöglichkeit bieten sich Expertengespräche an, in welchen die identifizierten Themen diskutiert werden. Demgegenüber fokussiert sich die quantitative Auswertung auf die große Datenmenge beim Topic Modeling. Haben die Daten einen Panelcharakter, sind also über mehrere Jahre erhebbar, können Analysen über die zeitliche Entwicklung von Themenfeldern durchgeführt werden. Bei Nachrichtendaten bietet es sich an, die Intensität der Berichterstattung zu einem Thema im Zeitablauf zu beobachten (z.B. die Anzahl der Berichte zu einem Thema relativ zur Gesamtberichterstattung). Bei Geschäftsberichten kann die Anzahl der Nennungen eines Themas über die Zeit Aufschluss über dessen Entwicklung geben. Insgesamt kann die zeitliche Entwicklung als Indikator für die Wichtigkeit von Themen genutzt werden. Erhöht sich z.B. die Intensität der Berichterstattung schlagartig, kann dies ein Indiz dafür sein, dass das Thema stark an Bedeutung gewinnt. Sinken die Nennungen eines Themas über die Zeit, kann dies ein Hinweis darauf sein, dass das Thema an Relevanz verliert.

4 Praxistest des Früherkennungskonzepts

In diesem Kapitel wird das entwickelte Früherkennungskonzept einem ersten Praxistest unterzogen, um zu überprüfen, ob es mit realen Daten valide Ergebnisse liefert. Als Datenquellen dienen die Chancen- und Risikoberichte von 57 zufällig gezogenen deutschen Unternehmen für das Geschäftsjahr 2022. Diese sind überwiegend dem Verarbeitenden Gewerbe zuzuordnen. Daneben wurden 328 Pressemitteilungen vom Januar 2023 bis Oktober 2023 herangezogen, die von drei großen Verbänden in Deutschland, dem Bundesverband der Deutschen Industrie (BDI), Verband der Chemischen Industrie (VCI) und dem Zentralverband des Deutschen Handwerks (ZDH) publiziert wurden. Schließlich greifen wir noch auf Artikelüberschriften der Wirtschaftsnachrichten für den Zeitraum vom Oktober 2021 bis zum März 2022 zurück. Dieser Zeitraum wurde bewusst gewählt, um die Vorhersagekraft des Früherkennungskonzepts zu testen. Insgesamt gehen 4.823 Titel in die Analyse ein. Alle Daten wurden dabei von den Websites der Unternehmen, Verbänden und Medienanbietern extrahiert. Im Folgenden werden die Analyseergebnisse je Datenquelle im Detail präsentiert.

In den Chancen- und Risikoberichten der 57 Unternehmen werden die zukünftigen Entwicklungen und Herausforderungen der individuellen Geschäftsbereiche dargelegt (vgl. Köhlbrandt et al. 2020). Im ersten Durchlauf der Topic Modeling-Analyse werden zunächst nur die Nomen in den Textdaten analysiert, um die diskutierten Themenfelder zu identifizieren. Aus den Daten konnten die Themen "Rohstoffe und Energie", "Klimaschutz", und "Informationssicherheit" als drei wesentliche Themenfelder extrahiert werden (vgl. Abb. 4)¹.

¹ Die Benennung der Themenfelder für alle Datenquellen erfolgte durch die Autoren.

Abbildung 4: Topic Modeling-Ergebnisse der Unternehmensdaten

Chancen- und Risikoberichte: Identifizierte Themen (Auswahl)



Quelle: IfM Bonn 2023.

Um den Kontext der Themen besser einzuordnen, wurde anschließend eine zweite Topic Modeling-Analyse je Themenfeld durchgeführt, welches neben Nomen auch Verben und Adjektive berücksichtigte. Im Themenfeld "Rohstoffe und Energie" wurde festgestellt, dass Unternehmen Geschäftsrisiken aufgrund von hohen Energiepreisen sowie von Ressourcenengpässen (beides verursacht durch den Ukraine-Krieg) sehen. So schreibt die BASF beispielsweise: "Weiter steigende Energie-, Rohstoffpreise sowie Rohstoffknappheiten bei einigen Produkten und Wertschöpfungsketten, vor allem bedingt durch den Krieg in der Ukraine, können den Margendruck weiter erhöhen." (vgl. BASF 2022, S. 161). Zum Thema "Klimaschutz" zeigt die Kontextanalyse zum einen, dass Risiken durch den Klimawandel diskutiert werden und zum anderen welche aktiven Maßnahmen dagegen unternommen werden. So schreibt das Unternehmen Jost, ein Zulieferer für die Nutzfahrzeugindustrie, beispielsweise "Wir wollen den Energieeinsatz und die CO₂-Emissionen bei der Herstellung unserer Produkte erheblich reduzieren." (vgl. Jost 2022, S. 61). Beim Thema "Informationssicherheit" kann festgestellt werden, dass Unternehmen eine zunehmende Bedrohung durch Cyberkriminalität feststellen und entsprechend in ihre IT-Sicherheit investieren. So benennt das Unternehmen Schweizer Electronic, ein Hersteller von Leiterplatten, "[...] die allgemein beobachtbare Dynamik im Bereich

Cyberangriffe und Cyberkriminalität als wesentliche Risiken zu nennen. [...] Schweizer unternimmt regelmäßig technische und organisatorische Maßnahmen, um das Risiko von Beeinträchtigungen, Ausfälle von IT-Systemen, Cyberangriffe [...] zu mittigeren." (vgl. Schweizer Electronic 2022, S. 79). Insgesamt zeigt der Test des Früherkennungskonzepts mit den 57 Unternehmensdaten plausible Ergebnisse, die sich durch die Berichtsinhalte validieren lassen.

Analog zu den Unternehmensberichten werden auch die Verbandspressemitteilungen untersucht. Als Topthemen werden hier "Energiekosten", "Fachkräfteausbildung", und "Klimafreundliche Technologien" identifiziert (siehe Abbildung 5).

Abbildung 5: Topic Modeling-Ergebnisse der Verbandspressemitteilungen

Verbandspressemitteilungen: Identifizierte Themen (Auswahl)



Quelle: IfM Bonn 2023.

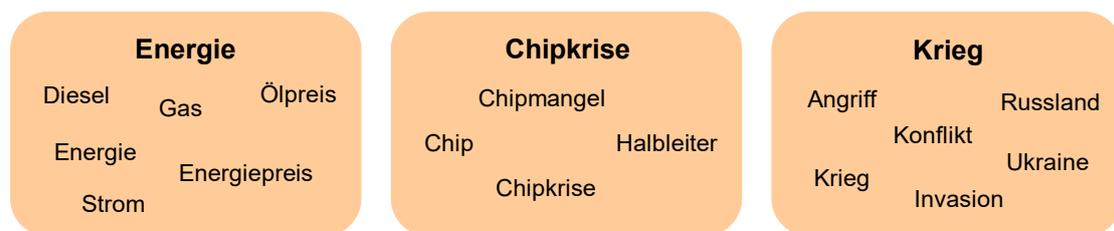
Die Kontextanalyse für das Thema "Energiekosten" unterstreicht die großen Belastungen für Unternehmen aufgrund stark gestiegener Energiepreise. Das Thema „Fachkräfteausbildung“ ist ein besonders stark diskutiertes Thema beim Verband der Handwerksbetriebe. Insbesondere die Ausbildung neuer Fachkräfte wird als große Aufgabe in der Branche unterstrichen. Beim Themenblock "Klimafreundliche Technologien" werden verschiedene technologische Ansätze zur Erreichung der Klimaneutralität diskutiert, wie alternative Antriebe oder

Wasserstofftechnologien. Insgesamt liefert auch dieser Konzepttest nachvollziehbare Ergebnisse.

Im Rahmen der Analyse der Titel der Wirtschaftsnachrichten konnten die Themen "Energie", "Chipkrise", und "Krieg" als relevante Themenblöcke extrahiert werden (siehe Abbildung 6).

Abbildung 6: Topic Modeling-Ergebnisse der Nachrichtentitel

Wirtschaftsnachrichten-Titel: Identifizierte Themen (Auswahl)



© IfM Bonn 23 2301 006

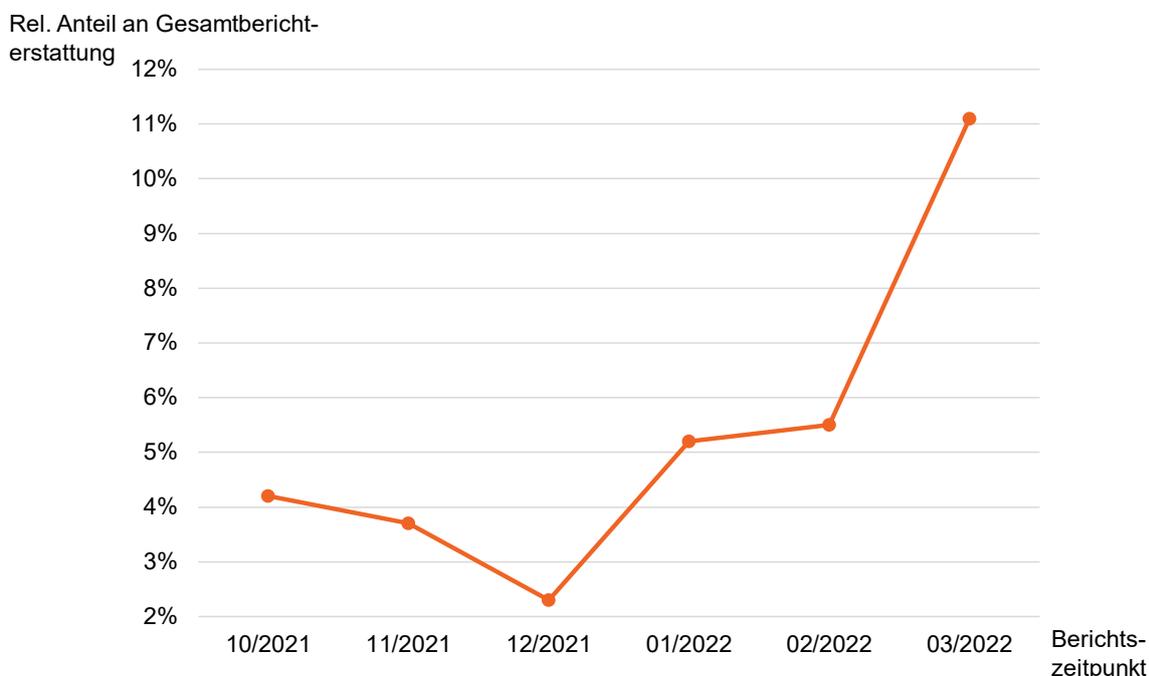
Quelle: IfM Bonn 2023.

Insgesamt liefern auch die Titel von Wirtschaftsnachrichten plausible Ergebnisse im Rahmen des Konzepttests. Als Auswertungsmethode wird ein quantitativer Ansatz gewählt, da wir zum einen eine große Datenmenge haben und zum anderen die inhaltliche Kürze von Titeln eine qualitative Analyse ineffektiv macht. Dabei soll die zeitliche Entwicklung eines Themas analysiert werden, um die Vorhersagekraft von Nachrichtentitel zu analysieren. Konkret wird der Fokus auf das Thema "Energie" gesetzt. Zunächst untersuchen wir, wie sich die Intensität der Berichterstattung² (d.h. der relative Anteil des Themas Energie an der

² Intensität Berichterstattung Thema X im Monat j = $\frac{\text{Summe Titel mit Thema X im Monat j}}{\text{Gesamttitelanzahl im Monat j}}$

Gesamtberichterstattung) über den betrachteten Zeitraum entwickelt hat (siehe Abbildung 7).

Abbildung 7: Intensität der Berichterstattung zum Thema "Energie"



© IfM Bonn 23 2301 007

Quelle: IfM Bonn 2023.

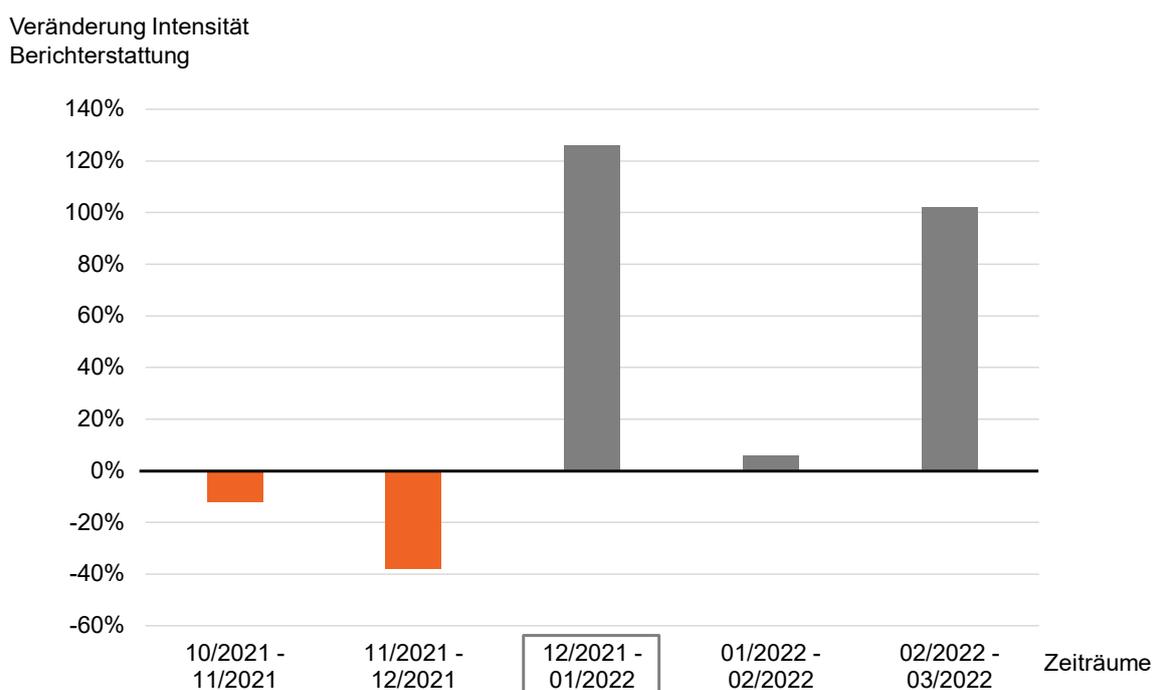
Auf den ersten Blick ist zu beobachten, dass sich die Intensität der Berichterstattung zum Thema "Energie" zwischen Oktober 2021 und Februar 2022 zunächst halbiert und dann wieder verdoppelt. Im Endergebnis sich also nur minimal verändert. Diese Veränderungen können nicht per se einem offensichtlichen Ereignis zugeordnet werden und könnten einem zu erwartenden Grundrauschen innerhalb der Daten zugeordnet werden. Lediglich im März sieht man den erwarteten schlagartigen Anstieg der Berichterstattung aufgrund des Ukraine-Krieges.

Eine absolute Betrachtung der Berichterstattungsintensität ist jedoch nur eine Seite der Medaille. Auf der anderen Seite können wir die monatliche Veränderung der Berichterstattungsintensität zum Thema "Energie" analysieren³. Hier bietet sich folgendes Bild (siehe Abbildung 8): Interessanterweise beobachten

³
$$\text{Veränderung Intensität der Berichterstattung zum Thema X} = \frac{[\text{Intensität Berichterstattung Thema X im Monat } j] - [\text{Intensität Berichterstattung Thema X im Monat } j-1]}{[\text{Intensität Berichterstattung Thema X im Monat } j-1]}$$

wir den höchsten Anstieg der Berichterstattung zum Thema Energie zum Jahreswechsel 2021/2022. Dieser starke Anstieg ist ein Indiz dafür, dass das Energiethema zum Jahreswechsel immens an Bedeutung gewinnt.

Abbildung 8: Monatliche Veränderung der Intensität der Berichterstattung zum Thema "Energie"



© IfM Bonn 23 2301 008

Quelle: IfM Bonn 2023.

Um mögliche Verzerrungen auszuschließen, haben wir untersucht, ob besondere Faktoren wie z.B. ein sehr kalter Winter oder eine wichtige OPEC-Entscheidung, diesen sprunghaften Anstieg verursacht haben. Dies war nicht der Fall. Besonders hervorzuheben ist, dass der sprunghafte Anstieg vor dem Ukraine-Krieg und der resultierenden Energiekrise stattfindet. Das unterstreicht die gute Qualität von Nachrichtendaten als Frühindikator.

Insgesamt zeigt unser Früherkennungskonzept im Praxistest mit allen drei Datenquellen plausible Ergebnisse. Im nächsten Schritt muss das Konzept mit größeren Datenmengen getestet werden.

5 Fazit

Die bisher im Rahmen der Früherkennung genutzten Verfahren basieren vor allem auf der Analyse strukturierter Daten. In Textform gebundene Informationen bleiben zumeist unberücksichtigt. Zur Schließung dieser Lücke haben wir im Rahmen der vorliegenden Studie ein Früherkennungskonzept entwickelt und erprobt, das in der Lage ist, aus Textquellen effizient Themenfelder zu extrahieren. Der experimentelle Charakter der Studie ermöglichte es uns, das State of the Art-Verfahren des maschinellen Lernens anzuwenden und für unser Forschungsziel zu optimieren. Ein ausgiebiger Praxistest zeigt, dass unser Früherkennungskonzept valide Ergebnisse zur frühzeitigen Erkennung von Themen und Entwicklungen liefert. Das ermöglicht Mittelstand und Wirtschaftspolitik, proaktiv auf Veränderungen zu reagieren.

Allerdings konnte sich der Praxistest zwangsläufig nur auf vergangene Ereignisse beziehen, wie z.B. die Sicherheit und Bezahlbarkeit der Energieversorgung im Gefolge des Ukraine-Krieges. In einem nächsten Schritt ist zu prüfen, inwieweit sich das Früherkennungskonzept, dann gespeist aus zusätzlichen, größeren Datenquellen, tatsächlich auch für eine echte Vorhersage eignet.

Literatur

- Ahrend, R.; Menon, C. (2023): Which SMEs are greening? – Cross-country evidence from one million websites, 5th session of the Committee for SMEs and Entrepreneurship, 9. – 10. November 2023.
- BASF (2022): Geschäftsbericht 2022. URL: <https://archiv.geschaeftsberichte-download.de/2022/BASF11.pdf>, S. 161 (Abgerufen: 23.10.2023).
- Benner, J.; Meier, C. (2003): Prognosegüte alternativer Frühindikatoren für die Konjunktur in Deutschland, Institut für Weltwirtschaft, Kiel.
- Berger, J.; Humphreys, A.; Ludwig, S.; Moe, W.; Netzer, O.; Schweidel, D. (2020): Uniting the Tribes: Using Text for Marketing Insight, *Journal of Marketing* 84(1), S. 1-25.
- Blei, D.; Ng, A.; Jordan, M. (2003): Latent Dirichlet Allocation, *Journal of Machine Learning Research* 3, S. 993-1022.
- Brink, S.; Icks, A. (2022): Zukunftspanel 2022: Klima und Energie gewinnen als Herausforderungen an Bedeutung, IfM Bonn.
- Busch-Heizmann, A.; Krabel, S.; Wangler, L. (2022): Die Resilienz von Unternehmen und Wirtschaft in Zeiten externer Schocks, Berlin.
- Ferrari, M.; Le Mezo, H. (2020): Using information in newspaper articles as an indicator of real economic activity, *Economic Bulletin Issue* 2.
- Fischer, M. (2014): Serie Frühindikatoren (VII), Earlybird-Indikator: Früher Vogel, URL: <https://www.wiwo.de/politik/konjunktur/serie-fruehindikatoren-vii-earlybird-indikator-frueher-vogel/10306798.html> (Abgerufen: 16.10.2023).
- Gandomi, A.; Haider, M. (2015): Beyond the hype: Big data concepts, methods, and analytics, *International Journal of Information Management*, S. 137-144.
- Giesen, K. (2022): Quick Guide Krisenfrüherkennung im Unternehmen: Umsetzungsansätze für § 1 StaRUG.
- Goch, S. (1999): Strukturwandel im Ruhrgebiet-Eine Erfolgsgeschichte?, *Forschungen und Forschungsberichte* 22, S. 159-190.
- Gumpp, T.; Schneider, M. (2021): Methoden der Künstlichen Intelligenz in der Rechtswissenschaft.

Hauff, S. (2009): Konzeptionen der Früherkennung, Diskussionspapiere des Schwerpunktes Unternehmensführung am Fachbereich BWL der Universität Hamburg.

Hinze, J. (2003): Prognoseleistung von Frühindikatoren: Die Bedeutung von Frühindikatoren für Konjunkturprognosen - Eine Analyse für Deutschland, HWWA Discussion Paper, No. 236, Hamburg Institute of International Economics (HWWA), Hamburg.

Huang, C.; Simpson, S.; Ulybina, D.; Roitman, A. (2019): News-based Sentiment Indicators, IMF Working Papers 273.

Ifo-Institut (2023): ifo Geschäftsklima Deutschland, URL: <https://www.ifo.de/umfrage/ifo-geschaeftsklima-deutschland> (Abgerufen: 12.10.2023).

Jost (2022): Geschäftsbericht, URL: <https://archiv.geschaeftsberichte-download.de/2022/JST400.pdf>, S.61 (Abgerufen am 23.10.2023).

Klimonczyk, S. (2016): Die Vorhersagekraft von Zinsstrukturkurven für das Wirtschaftswachstum. Ein Ländervergleich anhand zweier Modelle

Köhlbrandt, J.; Gleißner, W.; Günther, T. (2020): Corporate Finance (07/08), Umsetzung gesetzlicher Anforderungen an das Risikomanagement in DAX- und MDAX-Unternehmen: Eine empirische Studie zur Erfüllung der gesetzlichen Anforderungen nach den §§ 91 und 93 AktG., S. 248-258.

Krystek, U.; Müller-Stewens, G. (1993): Frühaufklärung für Unternehmen: Identifikation und Handhabung zukünftiger Chancen und Bedrohungen.

Krystek, U. (2007): ZfCM (Sonderheft 2/2007), Strategische Früherkennung, 50–59.

Rossen, A. (2012): Frühindikatoren: gute Vorlaufeigenschaften, Wirtschaftsdienst 92 (8), S. 575–576.

Schubert, K.; Klein, M. (2020): Das Politiklexikon, 7. Auflage, Bonn.

Schwartz, M.; Gerstenberger, J. (2020): KfW Research Fokus Volkswirtschaft (Nr.286), Corona-Krise hat den Mittelstand fest im Griff, aber Widerstandsfähigkeit (noch) hoch.

Schweizer Electronic (2022): Geschäftsbericht, URL: <https://archiv.geschaeftsberichte-download.de/2022/515623.pdf>, S.79 (Abgerufen am 23.10.2023).

Seiler, C.; Wohlrabe, K. (2013): Das ifo Geschäftsklima und die deutsche Konjunktur, ifo Schnelldienst, ifo Institut - Leibniz-Institut für Wirtschaftsforschung an der Universität München, Vol. 66, Iss. 18, pp. 17-21.

Stevens, K.; Kegelmeyer, P.; Andrzejewski, D.; Buttler, D. (2012): Exploring Topic Conherence over many models and many topics.

Tan, A. (1999): Text Mining: The state of the art and the challenges.

ZEW (2023): ZEW-Finanzmarkttest, URL: <https://www.zew.de/forschung/projekte/zew-finanzmarkttest> (Abgerufen:12.10.2023).

Anhang

Übersicht 1: Auswahl an Konjunkturindikatoren

Indikator	Herausgeber	Themenfokus	Datenquelle	Art
<i>Bruttoinlandsprodukt (BIP)-Prognose</i>	Bundesregierung	Konjunktur insgesamt	Modellrechnungen	Früh
<i>Early-Bird-Index</i>	Commerzbank	Konjunktur insgesamt	Modellrechnungen	Früh
<i>Konsum (Privat)</i>	GfK-Konsumklima-Index	Einzelne Konjunktur-Elemente	Umfragebasiert	Früh
<i>Stromverbrauchsdaten-Index</i>	ifo Institut	Einzelne Konjunktur-Elemente	Realwirtschaftlich	Früh/ Präsenz
<i>ifo-Geschäftsklimaindex</i>	ifo-Institut	Konjunktur insgesamt	Umfragebasiert	Früh
<i>Gemeinschaftsprognose</i>	ifo-Institut; IfW Kiel; RWI	Konjunktur insgesamt	Modellrechnungen; Expertenbeurteilungen	Früh
<i>Kiel Trade Indicator</i>	IfW Kiel	Einzelne Konjunktur-Elemente	Realwirtschaftlich	Früh/ Präsenz
<i>Passanzahlen Konsumaktivität</i>	IfW Kiel	Einzelne Konjunktur-Elemente	Realwirtschaftlich	Präsenz
<i>Einkaufsmanagerindex (Manufacturing PMI)</i>	IHS Markit (S&P Global); Institute for Supply Management (ISM)	Einzelne Konjunktur-Elemente	Umfragebasiert	Früh
<i>Konjunkturampel</i>	Institut für Makroökonomie und Konjunkturforschung (IMK)	Konjunktur insgesamt	Realwirtschaftlich; Umfragebasiert	Früh
<i>KfW-ifo-Mittstandsbarometer</i>	KfW-Bank; ifo-Institut	Konjunktur insgesamt	Umfragebasiert	Früh
<i>Auftragseingänge</i>	Statistisches Bundesamt	Einzelne Konjunktur-Elemente	Realwirtschaftlich	Früh/ Präsenz

Fortsetzung Übersicht 1

Indikator	Herausgeber	Themenfokus	Datenquelle	Art
<i>Bruttoanlageinvestitionen</i>	Statistisches Bundesamt	Einzelne Konjunktur-Elemente	Realwirtschaftlich	Früh/ Präsenz
<i>Preisindizes</i>	Statistisches Bundesamt	Einzelne Konjunktur-Elemente	Realwirtschaftlich	Früh/ Präsenz
<i>Außenhandelsbilanz</i>	Statistisches Bundesamt	Einzelne Konjunktur-Elemente	Realwirtschaftlich	Präsenz
<i>Gemeldete Stellen</i>	Statistisches Bundesamt	Einzelne Konjunktur-Elemente	Realwirtschaftlich	Präsenz
<i>LKW-Maut-Fahrleistungsindex</i>	Statistisches Bundesamt	Einzelne Konjunktur-Elemente	Realwirtschaftlich	Früh/ Präsenz
<i>Arbeitslosenquote</i>	Statistisches Bundesamt; weitere Institute	Einzelne Konjunktur-Elemente	Modellrechnung; Realwirtschaftlich	Früh/ Präsenz
<i>ZEW-Index</i>	ZEW Mannheim	Konjunktur insgesamt	Umfragebasiert	Früh