

Datenanalyse, Hypothesentests und Entwicklung eines prädiktiven Deep Learning Modells für Auftragseingänge durch kontextbewusste Multi-Source-Datenfusion

Jakob Müller

Zusammenfassung—Diese Arbeit behandelt die Identifikation von Mustern und Zusammenhängen, speziell Korrelation, Kovarianz und Granger-Kausalität (Präzedenz), zwischen dem Auftragseingang und unternehmensinternen Vertriebsdaten, wie Kundenaccounts, Verkaufschancen und Aufträgen, sowie externen Daten wie Wirtschaftsindikatoren, Trenddaten (Aufrufe in Suchmaschinen wie Google), Nachrichten und Aktienkursen inklusive der Entwicklung prädiktiver Deep-Learning-Modelle für den Auftragseingang unter Anwendung des Konzepts der Multi-Source-Datenfusion, um die Genauigkeit und Erklärbarkeit der Prognosen für zukünftige Auftragseingänge und damit die Geschäftssituation zu erhöhen. Dafür werden zunächst unternehmensinterne mit externen Daten zu einem multimodalen Datensatz integriert. Anschließend wird eine explorativ-deskriptive und diagnostische Datenanalyse durchgeführt, um Hypothesen und Zusammenhänge im Datensatz mit dem Auftragseingang zu identifizieren und mit statistischen Methoden, darunter Korrelations-, Kovarianz- und Granger-Kausalitätstests, zu prüfen. Darauf aufbauend werden prädiktive Deep Learning Modelle unterschiedlicher Architekturen für die Auftragseingangsvorhersage entwickelt und verglichen und dabei das Konzept der kontextbewussten Multi-Source-Datenfusion angewendet. Aufbauend auf aktuellen Arbeiten zu Transformer Modellen, Deep-Learning-Architekturen und der Fusion mehrerer Datenquellen werden Large Language Models und Text Embeddings genutzt, um Textdaten mit unternehmensinternen und externen Daten zu integrieren, die Änderungen in der wirtschaftlichen Entwicklung und den Marktbedingungen widerspiegeln. Ziel ist es, die Leistungsfähigkeit dieser Technologien auf die Vorhersage von Auftragseingängen zu übertragen und somit die Genauigkeit der Vorhersage und die Entscheidungsfindung in Unternehmen zu verbessern.

Index Terms—Auftragseingangsvorhersage, Explorative Datenanalyse, Prädiktive Modelle, Deep Learning, Transformer Modelle, Multi-Source-Datenfusion, Large Language Models, Text Embeddings, Wirtschaftsindikatoren.

I. EINFÜHRUNG

MIT der schnellen Entwicklung der künstlichen Intelligenz und der Emergenz digitaler Ökonomien erhält die globale Wirtschaft zunehmenden Aufschwung. Wirtschaftsindikatoren wie Aktienkurse reflektieren Änderungen in der wirtschaftlichen Entwicklung und in Marktbedingungen.

Diese Arbeit ist als Studienarbeit am DHBW Center for Advanced Studies in Heilbronn entstanden. Die hier zum Ausdruck gebrachten Meinungen sind ausschließlich die des Autors. Es wird keine Garantie übernommen. Der Benutzer trägt das gesamte Risiko.

Problemstellung

Datengetriebene Prognose- und Entscheidungsmodelle sind in der modernen Wirtschaft wichtig, um Unternehmensentscheidungen in Bezug auf Produktion, Lagerhaltung und Ressourcenplanung zu informieren. Auftragseingangsvorhersagen ermöglichen es Unternehmen, zukünftige Nachfrage präzise zu prognostizieren, Ressourcen effizient zu planen und strategische Entscheidungen zu optimieren, um Wettbewerbsvorteile zu sichern. Viele Unternehmen verlassen sich bezüglich ihrer operativen und strategischen Entscheidungen vor allem auf intuitive Vorhersagen von Markttrends und Nachfrageschwankungen durch Experten. In Zeiten von Volatilität und schnellem Wandel ist die Geschäftssituationsanalyse und die Präzision von Vorhersagen kritisch für den Erfolg, um frühzeitig neue Situationen zu erkennen und effektive Maßnahmen zu treffen.

Ziel der Arbeit

Ziel dieser Arbeit ist die Identifikation von Mustern und Zusammenhängen zwischen dem Auftragseingang und internen Vertriebsdaten, wie Kundenaccounts, Verkaufschancen und Aufträgen, sowie externen Daten wie Wirtschaftsindikatoren, Trenddaten (Aufrufe in Suchmaschinen), Nachrichten und Aktienkursen, um die Genauigkeit und Erklärbarkeit der Prognosen für zukünftige Auftragseingänge und damit der Geschäftssituation zu erhöhen.

Forschungsfrage und Hypothesen

Die Arbeit beschäftigt sich im Kern mit der Frage, welche Muster und Zusammenhänge zwischen dem Auftragseingang und den genannten internen und externen Daten bestehen, wie die Auftragslage vorausschauend sinnvoll eingeschätzt und wie die zukünftige Entwicklung des Auftragseingangs vorhergesagt werden kann. Die Ergebnisse werden auf Basis der folgenden Hypothesen analysiert, die auf den gegebenen Anwendungsfall zur Auftragseingangsvorhersage limitiert sind:

- 1) **Relevanz interner und externer Daten:** Die Integration von externen Trend-, Wirtschafts- und Ereignisdaten neben den unternehmensinternen Daten verbessert signifikant die Vorhersagegenauigkeit von Auftragseingängen gegenüber Modellen, die ausschließlich interne oder externe Daten verwenden.

- 2) **Pragmatismus komplexer Machine Learning Modelle für die Auftragseingangsvorhersage:** Die Modellkomplexität und Genauigkeit stehen in einem optimalen Verhältnis, sodass komplexere Modelle, welche maschinelles Lernen einsetzen, genauer sind als vereinfachte statistische Modelle und die intuitive Vorhersage von Auftragseingängen durch Experten, ohne dabei an praktischer Anwendbarkeit zu verlieren, wenn Erklärbarkeit durch XAI-Techniken gewährleistet wird.
- 3) **Relevanz und Effektivität von Multi-Source-Datenfusionsmechanismen für die Auftragseingangsvorhersage:** Durch die Übertragung dieser Konzepte auf die Vorhersage von Auftragseingängen kann die Genauigkeit und Verlässlichkeit der Prognosen verbessert werden.
- 4) Weitere noch auszuarbeitende Hypothesen zum Zusammenhang von Daten mit dem Auftragseingang, speziell auch zu externen Einflüssen

METHODIK

Nach einer vorläufigen Literaturrecherche zur Thematik wird zunächst der OECD-Datensatz erstellt und beschafft. Dieser Datensatz bildet die Grundlage für die erste Phase der Analysen und Experimente. Im Anschluss daran werden spezifische Funktionen zur Prüfung der Hypothesen entwickelt und unmittelbar anhand des OECD-Datensatzes getestet.

Darauf folgt die Implementierung des Codes für die Modellexperimente. Diese Experimente umfassen alle Schritte der Modellentwicklung, vom Modelltraining bis hin zur Evaluation. Wie die Analysen werden auch die initialen Modellexperimente zuerst ausschließlich mit dem OECD-Datensatz durchgeführt. In der nächsten Phase wird der Datensatz erweitert und die Analysen und Modellexperimente auf den so entstandenen Datensätzen durchgeführt, um alle Experimente anschließend zu vergleichen und dabei die entdeckten Zusammenhänge und die Modellperformance zu erforschen.

Der erweiterte Datensatz entsteht durch die Integration unternehmensinterner und externer Datenquellen, um ein umfassendes multimodales Datenset zu erstellen. Datenquellen umfassen neben der API des OECDs das unternehmensinterne Salesforce-System mit diversen Vertriebsdaten sowie weitere externe APIs wie Google Trends, NewsAPI und Yahoo Finance. Programmiert wird in Python mit Jupyter Notebook unter Verwendung der nachfolgend genannten wesentlichen Bibliotheken.

Insgesamt gliedert sich die systematischen Vorgehensweise zur Erkennung von Mustern und Zusammenhänge mit dem Auftragseingang sowie der Entwicklung von Vorhersagemodellen in drei Phasen: (1) Explorativ-deskriptive Datenanalyse, (2) diagnostische Analyse und Hypothesentests (3) Modellentwicklung, -validierung und -erklärung.

A. Explorativ-deskriptive Datenanalyse (EDA)

Die explorativ-deskriptive Datenanalyse bildet die Grundlage der Hypothesenbildung. Zunächst werden die internen Unternehmensdaten und externen Wirtschaftsdaten aggregiert und visuell sowie statistisch analysiert. Ziel der EDA ist

es, die Bedeutung und Struktur der Daten zu verstehen und Hypothesen zu bilden.

B. Diagnostische Analyse und Hypothesentests

Auf Basis der Ergebnisse der EDA werden spezifische Hypothesen formuliert. Diese Hypothesen zielen darauf ab, die Beziehungen zwischen den verschiedenen Datensätzen und dem Auftragseingang darzustellen und zu testen. Konkret geht es um die Beziehungstypen Korrelation, Kovarianz und Granger-Kausalität (bzw. Präzedenz). Ein Beispiel für eine solche Hypothese könnte sein: „Ein Anstieg des Bruttoinlandsprodukts führt zu einem proportionalen Anstieg der Auftragseingänge.“ Für das Testen der Hypothesen werden Korrelations-, Kovarianz- und Granger-Kausalitätstest angewendet. Dafür soll die Bibliothek statsmodels genutzt werden.

C. Modellentwicklung und -erklärung

Basierend auf den validierten Hypothesen werden prädiktive Modelle entwickelt. Hierfür stehen fortschrittliche Machine Learning-Algorithmen wie Long Short-Term Memory (LSTM)-Netzwerke, Convolutional Neural Networks (CNNs) und Transformer-Modelle zur Auswahl, die zumindest in einem neuronalen Netz mit Fully-Connected-, LSTM-, CNN- und Transformer-Encodern kombiniert und in einem Fully-Connected-Decoder zusammengeführt werden sollen. Die Modelle werden darauf trainiert, auf Basis der Werte der Feature-Zeitreihen der letzten 12 Monate die zukünftigen Auftrags-eingänge der folgenden 12 Monate vorherzusagen, dann mit geeigneten Metriken evaluiert und mit Baseline-Modellen verglichen. Für die Modellentwicklung werden die Bibliotheken Keras und Scikit-Learn verwendet.

D. Modellerklärung

Abschließend wird die Erklärbarkeit der Modelle durch Explainable AI (XAI) Techniken geprüft, um die Entscheidungen der Modelle nachvollziehbarer zu machen. Genutzt wird dafür die Bibliothek Shap.

II. VORGEHENSWEISE

Literaturrecherche

Zunächst erfolgt eine systematische Literaturrecherche zur Untersuchung von bestehenden Ansätzen und Methoden für die Analyse und Prognose von Finanzdaten in ähnlichen Anwendungsbereichen unter Verwendung von Datenbanken (IEEE Explore, Google Scholar, Researchgate, arXiv), Verlagen (Springer, Wiley, De Gruyter) und spezifischen Journals für Data Science, KI und den Finanzbereich (The Economist, New Scientist, The Economic Journal, American Economic Journal), um ein fundiertes Verständnis des Themenbereichs zu erlangen, bestehende Modelle und Methoden zu verstehen und Forschungslücken zu identifizieren.

Datenerhebung und Datenintegration

Hier werden relevante Unternehmensdaten (z. B. Verkaufszahlen, Kundenfeedback) gesammelt und passende externe Datenquellen (z. B. Wirtschaftsindikatoren, Trenddaten, Nachrichtenkanäle) identifiziert.

Datenexploration und -vorverarbeitung

In diesem Schritt werden Verständnis, Bereinigung und Aufbereitung der erhobenen Daten für die nachfolgende Analyse und Modellierung erzielt, um die Datenqualität und damit die Effektivität der Analysen und Modelle zu optimieren.

Explorativ-deskriptive und diagnostische Datenanalyse

Anschließend erfolgt die Analyse der erhobenen Daten auf Muster und Zusammenhänge mit dem Auftragseingang zur Formulierung und zum Testen von Hypothesen, um das Verständnis der Daten zu vertiefen. Methodisch können deskriptive Statistik, Korrelationsanalysen und Hypothesentests eingesetzt werden. Um Hypothesen über Muster und Zusammenhänge im gegebenen Datensatz zu generieren, werden typische statistische Darstellungen und Kennwerte verwendet.

Modellentwicklung

Danach erfolgen Entwicklung, Auswahl und Test verschiedener maschineller Lernalgorithmen und Modelle (auszuwählen und ggf. in Hypothesen integrieren) zur Erstellung von Modellen für die Vorhersage von Auftragseingängen sowie die Auswahl des besten Modells.

Test und Evaluation

In diesem Rahmen werden geeignete Metriken für die Genauigkeit des Modells festgelegt und Vorhersagen anhand von Testdaten der letzten Zeiträume durchgeführt sowie mit bestehenden Prognosen zur Bewertung und Verbesserung der Prognosefähigkeit verglichen.

Modellerklärung und Interpretation

Zur Interpretation der Modellentscheidungen und zur Transparenzsteigerung werden XAI-Techniken (Explainable AI) genutzt, um Vertrauen in das Modell zu schaffen und Einsichten für Entscheidungsträger zu geben.

Erweiterung der Datenbasis

Nachdem diese Schritte zunächst für den Datensatz mit den OECD-Daten durchgeführt wurden, sollen dann weitere Datensätze durch die Ergänzung mit Daten aus den oben genannten Quellen erstellt und die entwickelten Analysen und Modellexperimente darauf angewendet und die Ergebnisse mit jeweils allen Datensätzen verglichen werden. Um die News-Daten in die Vorhersage aufzunehmen, werden diese zunächst nach Suchbegriffen von NewsAPI (oder, falls Zugriff besteht, auch von NewscatcherAPI) abgerufen und nach Monat ihrer Veröffentlichung geordnet. Dann werden für jeden Monat mithilfe eines Large Language Models eine Zusammenfassung der Artikel des jeweiligen Monats mit den für die Auftrags-eingangsvorhersage relevanten Themen und Aspekten erstellt. Für jeden Monat wird dann zusätzlich eine Zusammenfassung der Zusammenfassungen der Vormonate erstellt, um auch die Einflüsse aus der Vergangenheit abzudecken. Die Embeddingd beider Zusammenfassungen, die dann für jeden Monat erstellt sind, werden dann in das Modell integriert.

Zusammenfassung, Diskussion und Ausblick

Abschließend folgt die Zusammenfassung und kritische Reflexion der Ergebnisse und Ableitung erster unternehmerischer Handlungsempfehlungen auf Basis der Daten sowie die Formulierung von Handlungsempfehlungen und Perspektiven für zukünftige Forschung und Entwicklung.

III. ZEITPLAN

Ein erster Stand für die Integration der OECD-Daten und der historischen Daten des Auftragseingangs sowie der Korrelations-, Kovarianz- und Granger-Kausalitätsanalysen und der Modellentwicklung-, -evaluation und -erklärung ist entwickelt. Es folgt die Prüfung und das Refinement dieses Prozesses sowie die Erweiterung des Datensatzes und die systematische Durchführung und Auswertung der Modelle mit allen erweiterten Datensätzen bis Mitte Dezember. Ein erster Stand der Arbeit ist bis Mitte Januar zu erwarten. Der Abgabetermin ist der 22.02.25.

IV. GLIEDERUNGSENTWURF

- 1) Abstract / Zusammenfassung
- 2) Executive Summary
- 3) Einleitung
 - a) Schlüsselbegriffe
 - b) Trends und Entwicklungen in den Bereichen Finanzen, Controlling und Business Data Analytics
 - c) Revolutionierung der Finanzindustrie durch Big Data und fortgeschrittene Analysen und Transformation von Geschäftsmodellen, Entscheidungsprozessen und Strategien (bezogen auf Unternehmen) durch Risikokontrolle, Marktanalyse, Sentimentanalyse und Deep Learning
 - d) Erfolgreiche Datenanalyse für das Controlling
 - e) Forschungsgegenstand
 - f) Motivation und Forschungsinteresse
 - g) Ziel der Arbeit
 - h) Forschungsfrage
 - i) Hypothesen
 - j) Methodik und Vorgehensweise
 - i) Darstellung der Themen für den Grundlagenteil
 - ii) Literaturrecherche
 - iii) Datenerhebung
 - iv) Datenexploration und -vorverarbeitung
 - v) Datenanalyse
 - vi) Modellentwicklung
 - vii) Evaluation
 - viii) Modellerklärung und Interpretation
- 4) Grundlagen
 - a) Kurz: Controlling (Definition, Ziele, Kernaspekte)
 - b) Optional: Wirtschaftspolitik und Wirtschaftsindikatoren, Theorie des Effizienten Marktes, Informationen im Wirtschaftssystem
 - c) Optional: Operations Research, Entscheidungsfindung, Entscheidungsinformation und Strategie. Wie werden Vorhersagen und Analysen zum aktuellen Stand der Geschäftssituation, wie etwa in

Form von Zusammenfassungen von News oder auch anderen Texten aus der Branche, genutzt, um optimale strategische und Operative Entscheidungen in der Unternehmensplanung (z.B. Supply Chain Management und Controlling) zu treffen?

- d) Kurz: Sales (Ziele, Prozess und Daten, Erfolgsmessung, Beitrag von AE-Prognosen)
- e) Data Science (Definition, Ziele, Kernaspekte) und Data Science Prozess (CRISP-DM)
- f) Descriptive, Diagnostic, Predictive und Prescriptive Analytics
- g) Kurz: Statistische Methoden der Explorativen Datenanalyse: Plots, Kennzahlen, Hypothesentests, Korrelationsanalyse
- h) Fokus: Neuronale Netze (Architekturen und Schichten, Trainingskonzepte inklusive Mathematik: Aufbau, Forward Propagation, Backpropagation, Gradient Descent, Optimierer)
- i) Fokus: Explainable AI
- j) Fokus: Evaluationsmetriken

5) Durchführung

- a) Vorgehensweise
 - i) Literaturrecherche
 - ii) Datenerhebung
 - A) Beschreibung der Datenquellen
 - B) Darstellung der Datenintegration
 - iii) Datenexploration und –vorverarbeitung
 - iv) Datenanalyse
 - v) Modellentwicklung
 - vi) Evaluation
 - vii) Modellerklärung und Interpretation
- b) Ergebnisse
 - i) Darstellung und Interpretation der Kausalitätstests.
 - ii) Beantwortung der Frage, welche Indikatoren mit dem Auftragseingang korrelieren bzw. im Sinne der Tests eine zeitliche Ursächlichkeit indizieren.
 - iii) Darstellung und Interpretation der Evaluationsergebnisse für jedes Modell und für jeden Datensatz
 - iv) Beantwortung der Frage, welche Datensätze die Vorhersage des Auftragseingangs im Vergleich zur ausschließlichen Verwendung von Wirtschaftsindikatoren verbessern.
 - v) Beantwortung der Frage, ob Deep Learning-Modelle für die Vorhersage des Modells sinnvoll sind basierend auf ihrer Performance im Vergleich zu einfachen Modellen (Linear Regression und XGBoost)

6) Zusammenfassung

- a) Zusammenfassung der Ergebnisse
- b) Kritische Würdigung und Diskussion
- c) Ausblick und zukünftige Arbeiten

V. FORSCHUNGSSTAND

Die Vorhersage von Aktienkursen ist ein bekannter Anwendungsfall, der künstliche Intelligenz und maschinelles Lernen nutzt, um zukünftige Kursentwicklungen basierend auf historischen Daten und Marktindikatoren zu prognostizieren. Diese Modelle berücksichtigen eine Vielzahl von Variablen, darunter Wechselkurse, Änderungen in steuerpolitischen Regelungen, Marktstimmung und makroökonomische Bedingungen, um präzise Vorhersagen zu treffen [1]. Ähnlich wie bei der Aktienkursvorhersage ist die Vorhersage von Auftragseingängen für Unternehmen von entscheidender Bedeutung, da sie die Grundlage für strategische und operative Entscheidungen bildet.

Datengetriebene Prognose- und Entscheidungsmodelle sind in der modernen Wirtschaft wichtig, um Unternehmensentscheidungen in Bezug auf Produktion, Lagerhaltung und Ressourcenplanung zu informieren. Auftragseingangsvorhersagen ermöglichen es Unternehmen, zukünftige Nachfrage präzise zu prognostizieren, Ressourcen effizient zu planen und strategische Entscheidungen zu optimieren, um Wettbewerbsvorteile zu sichern. Viele Unternehmen verlassen sich bezüglich ihrer operativen und strategischen Entscheidungen vor allem auf intuitive Vorhersagen von Markttrends und Nachfrageschwankungen durch Experten. In Zeiten von Volatilität und schnellem Wandel ist die Präzision dieser Vorhersagen kritisch für den Erfolg. Ziel der Arbeit ist die Identifikation relevanter Muster und Zusammenhänge zwischen Unternehmensdaten und externen Daten und dem Auftragseingang zur Erhöhung der Genauigkeit und Erklärbarkeit der Prognosen für zukünftige Auftragseingänge

Wirtschafts- und Finanzprognosen, wie Aktienkurs- oder Auftragseingangsvorhersagen, werden durch Faktoren wie Wechselkurse, Steuerpolitik, Marktstimmung und makroökonomische Bedingungen beeinflusst [2]. Der Kapitalmarkt ist nach der Theorie des Effizienten Marktes sogar dadurch charakterisiert, dass er alle verfügbaren Informationen reflektiert [3]. Infolgedessen sind die Analyse und die Vorhersage solcher Variablen mit erheblichen Herausforderungen verbunden. Sowohl die quantitative Analyse, die Faktoren wie Aktienkurse, Einkommen und Bilanzen nutzt, als auch die qualitative Analyse, die das Management, Produkte und Strategien beleuchtet, sind wichtig, um erfolgreiche Investmentstrategien am Kapitalmarkt zu entwickeln und daher ist es auch sinnvoll, umfassende Informationen inklusive quantitativer Zahlen und qualitativer Beschreibungen zu sammeln, um Trends am Kapitalmarkt vorherzusagen [2]. Weil Unternehmen im Allgemeinen Teil des Wirtschaftssystems sind, das der Kapitalmarkt abbildet, gelten diese Überlegungen auch für die Vorhersage anderer wesentlicher Finanzdaten wie den Auftragseingang.

Traditionelle Vorhersagemodelle wie ARIMA und Exponential Smoothing zeigen ihre Grenzen insbesondere in hochvolatilen und komplexen Umgebungen [4]. Fortschrittlichere Modelle wie LSTM, CNNs, Transformer und Embedding Transformations bieten jedoch vielversprechende Ansätze zur Verbesserung der Vorhersagegenauigkeit. Diese Modelle

ermöglichen es, komplexe Muster und Beziehungen in großen und heterogenen Datensätzen zu erkennen [1].

Aktuelle Herausforderungen beinhalten die Integration und Verarbeitung vielfältiger Datenquellen, die Handhabung von Informationslücken in Daten, die die tatsächliche Situation nicht vollständig beschreiben, wie es wirtschaftlichen Anwendungen der Fall ist, sowie die Skalierbarkeit und Interpretierbarkeit der Modelle. Ziel dieser Arbeit ist es weiterhin, diese Herausforderungen durch die Entwicklung eines prädiktiven Deep Learning Modells für Auftragseingänge zu adressieren, das kontextbewusste Multi-Source-Datenfusion nutzt und so in der Lage ist, wesentlich mehr Informationen zur allgemeinen wirtschaftlichen Situation über die Fusion mit verschiedenen Textquellen und der Anwendung von Large Language Models (LLMs) und Embedding Transformations abzubilden.

Durch die Übertragung dieser Konzepte auf die Vorhersage von Auftragseingängen kann die Genauigkeit und Verlässlichkeit der Prognosen verbessert werden. Die Integration verschiedener Datenquellen wie Unternehmensdaten und Wirtschaftsindikatoren sowie die Anwendung moderner Lernverfahren hat das Potenzial, die Vorhersagekraft wesentlich zu verbessern und bietet daher ein relevantes und zeitgemäßes Forschungsthema.

VI. VORLÄUFIGE LITERATURRECHERCHE

Zum aktuellen Zeitpunkt wurde eine kompakte Literaturrecherche durchgeführt, um ein erstes Verständnis für das Thema zu erzielen und Forschungslücken zu identifizieren. Nachfolgend ist die Methodik beschrieben, bevor im Abschnitt VIII die Ergebnisse in Form einer Darstellung relevanter Literatur enthalten sind.

A. Schlüsselbegriffe

Zunächst wurden folgende zum Thema passende Schlüsselbegriffe definiert:

- Sales Predictions, Sales forecasting, predictive models for sales
- Machine Learning, predictive analytics, data analytics, pattern recognition, hypothesis testing, data mining, explorative data analysis
- Finance, controlling, business analytics, economic forecasting, business predictions, enterprise data, external sources
- Economic indicators, sales data, stock market, stock movement
- Data preprocessing, data integration, data management, data collection, data analysis
- Performance metrics, evaluation, explainable artificial intelligence, xai, interpretability techniques
- Future trends, developments, market trends, techniques, algorithms
- Systematic review, literature review, empirical study, development, implementation

B. Ein- und Ausschlusskriterien

Anschließend wurden folgende Ein- und Ausschlusskriterien definiert:

- Sprache: Studien, die in Englisch verfasst sind, da dies die vorherrschende Wissenschaftssprache ist und die breiteste Zugänglichkeit gewährleistet.
- Publikationszeitraum: Studien, die seit 2018 veröffentlicht wurden, um sicherzustellen, dass die in der Analyse verwendeten Technologien und Methoden aktuell sind. Ausnahmen sind Artikel, die grundlegende und bedeutende Konzepte beschreiben.
- Studientyp
 - Empirische Studien, die quantitative Analysemethoden verwenden und keine Editorials, Kommentare oder Meinungsartikel.
 - Fallstudien, die spezifische Beispiele für die Anwendung von Data Science und KI in der Vorhersage von Auftragseingängen darstellen.
 - Übersichtsarbeiten und Meta-Analysen, die einen Überblick über den aktuellen Forschungsstand zu Vorhersagen im Finanzbereich geben.
 - Keine Patente und kommerziellen Werbematerialien, die nicht einer wissenschaftlichen Peer-Review unterzogen wurden.
- Thematische Relevanz
 - Studien, die sich explizit mit der Entwicklung von prädiktiven Modellen für Auftragseingänge oder ähnliche wirtschaftliche Prognoseaufgaben oder Aufgaben, die für die Entwicklung eines prädiktiven Modells nützlich sein können, oder der Identifikation von Mustern und Zusammenhängen durch Datenanalyse beschäftigen.
 - Studien, Artikel und Bücher, die sich mit Methoden der Statistik, Datenanalyse, Data Science und KI beschäftigen, die für den Gegenstand der Arbeit relevant sind.
 - Keine Studien, die keine Verbindung zwischen ihren Forschungsergebnissen und Mustern, Zusammenhängen oder der Vorhersage von Auftrags-eingängen auf Basis interner und externer Datenquellen herstellen können oder zu denen sich keine herstellen lässt.
 - Studien, die maschinelles Lernen und KI-Methoden in der Vorhersage von Unternehmenskennzahlen verwenden.
 - Forschungsarbeiten, die die Integration von Unternehmensdaten und externen Daten wie Konjunkturindikatoren untersuchen.
 - Artikel, die sich mit Explainable AI (XAI) im Kontext von Prognosemodellen beschäftigen.

C. Literatursuche

Anschließend wurden die Bibliotheken von ArXiv, Google Scholar, Science Direct und IEEE Explore durchsucht. Dafür wurden die oben genannten Suchbegriffe zu Sucheingaben kombiniert. So wurden entsprechend der Ein- und Ausschlusskriterien 61 relevante Suchergebnisse gefunden, von denen die fünf relevantesten in diesem Exposé aufgenommen wurden und im Abschnitt VIII dargestellt sind.

Die so durchgeführte Literaturrecherche wird im Rahmen der Studienarbeit noch erweitert.

VII. GRUNDLAGENLITERATUR

- Zu Statistik: Bamberg, Baur und Krapp, *Statistik - Eine Einführung für Wirtschafts- und Sozialwissenschaftler*, 2022 [5] und Bourier, *Beschreibende Statistik*, 2022 [6].
- Zu Data Mining: Olson und Lauhoff, *Deskriptives Data-Mining*, 2023 [7].
- Zu Statistik, Data Mining und maschinelles Lernen: Yang, *Introduction to Algorithms for Data Mining and Machine Learning*, 2019 [8] und Liu, McGree, Ge et al. *Computational and Statistical Methods for Analysing Big Data with Applications*, 2016 [9].
- Zu Konzepten der Künstlichen Intelligenz (KI): Görz, Schmid und Braun, *Handbuch der Künstlichen Intelligenz*, 2021 [10].
- Zu mathematischen Grundlagen der Data Science: Knoblauch, *Mathematik für Informatik und Data Science*, 2024 [11].

VIII. VERWANDTE ARBEITEN

Für die Entwicklung eines prädiktiven Modells nutzt diese Arbeit Erkenntnisse und Konzepte der Arbeit von Mo, Yan und Xiao, *Background-aware Multi-source Fusion Financial Trend Forecasting Mechanism*, 2024 [1] und behandelt darauf aufbauend das Thema der Vorhersage von Unternehmensfinanzkennzahlen, speziell den Auftragseingang. Der Artikel beschreibt einen neuen Mechanismus zur Vorhersage von Trends, speziell im Bereich Finanzen und Wirtschaft am Anwendungsfall der Aktienkursvorhersage, der ein LLM, insbesondere das MacBERT-Modell, einsetzt, um Informationen aus Texten wie Aktienkursbewertungen und Berichten aus der Politik zu nutzen und diese mit Aktienpreisdaten zu integrieren. Diese kombinierten Merkmale werden dann für das Training neuronaler Netze mit unterschiedlichen Architekturen eingesetzt, um umfassende Marktanalysen zu ermöglichen. Die Architektur des Konzepts ist in Abbildung 1 dargestellt. Experimente zeigen, dass der Mechanismus die Genauigkeit von Aktienbewegungsvorhersagen verbessert, was auf die Einbeziehung der Berichte über das Sprachmodell zurückzuführen ist. Der Artikel liefert der Arbeit Modellierungs- und Trainingskonzepte sowie Mechanismen für die Integration unterschiedlicher Datenquellen, Anreize für die Behandlung der Erklärbarkeit von Vorhersagen und Implementierungsstrategien.

Der Artikel von Wasserbacher und Spindler, "Machine learning for financial forecasting, planning and analysis: recent developments and pitfalls," 2022 [4] bietet für die Arbeit eine umfassende Einführung in den Einsatz von maschinellem Lernen für Finanzprognosen, Planung und Analyse (FP&A). Dabei wird auch thematisiert, dass die Anwendung von maschinellem Lernen für Planungs- und Ressourcenallokationsaufgaben im Kontrast zu Prognoseaufgaben problematisch sein kann, da diese Aufgaben kausale Inferenz erfordern, was schwierig zu modellieren ist. Im Artikel wird ein Framework vorgestellt, um kausale Zusammenhänge zu behandeln und die Auswirkungen der Anwendung von Prognosen zu schätzen.

Der Artikel von Benidis, Rangapuram, Flunkert et al. "Deep Learning for Time Series Forecasting: Tutorial and Literature Survey," 2022 [12] bietet eine umfassende Einführung und einen Überblick über den Einsatz von Deep-Learning-Methoden zur Zeitreihenvorhersage mit einer Einführung in die Vorhersage von Zeitreihen, einer detaillierten Darstellung der wichtigsten Elemente moderner tiefer Lernmodelle und einer Übersicht über den aktuellen Stand der Forschung und Anwendung im Bereich der Zeitreihenvorhersage. Dabei werden gängige Modelle und Techniken vorgestellt, darunter Recurrent Neural Networks (RNNs), CNNs und Transformers, sowie State Space Models und probabilistische Ansätze.

Der Artikel von Cheriyan, Ibrahim, Mohanan et al. "Intelligent Sales Prediction Using Machine Learning Techniques," 2018 [13] beschäftigt sich ähnlich wie diese Arbeit mit der Entwicklung eines intelligenten Vorhersagesystems für Verkaufszahlen mittels maschineller Lerntechniken. Ziel ist es ebenfalls, zuverlässige Verkaufsprognosen zu erstellen, die Unternehmen bei der Ressourcenplanung und Entscheidungsfindung unterstützen. Der Artikel beschreibt verschiedene Techniken und Modelle, die zur Verbesserung der Effektivität von Verkaufsprognosen eingesetzt werden können. Die Evaluation verschiedener Modelle im Rahmen der Studie ergibt, dass der Gradient Boost Algorithm die höchste Vorhersagegenauigkeit bietet. Im Unterschied zu dieser Arbeit werden ausschließlich interne quantitative Daten für Modellierung und Vorhersage verwendet. Interessant ist sowohl, dass der Gradient Boosted Trees (GBT) Algorithms eine Vorhersagegenauigkeit von 98 % erreicht und somit auch zu prüfen, wie diese Methode für den im Rahmen dieser Arbeit gegebenen Anwendungsfall funktioniert.

LITERATUR

- [1] F. Mo, S. Yan und Y. Xiao, *Background-aware Multi-source Fusion Financial Trend Forecasting Mechanism*, 2024. arXiv: 2407.00904 [cs.LG]. Adresse: <https://arxiv.org/abs/2407.00904>.
- [2] X. Zhang, Y. Zhang, S. Wang, Y. Yao, B. Fang und P. S. Yu, "Improving stock market prediction via heterogeneous information fusion," *Knowledge-Based Systems*, Jg. 143, S. 236–247, 2018, ISSN: 0950-7051. DOI: <https://doi.org/10.1016/j.knsys.2017.12.025>. Adresse: <https://www.sciencedirect.com/science/article/pii/S0950705117306032>.
- [3] E. F. Fama, "Efficient Capital Markets: A Review of Theory and Empirical Work," *The Journal of Finance*, Jg. 25, Nr. 2, S. 383–417, 1970, ISSN: 00221082, 15406261. Adresse: <http://www.jstor.org/stable/2325486> (besucht am 24. 07. 2024).
- [4] H. Wasserbacher und M. Spindler, "Machine learning for financial forecasting, planning and analysis: recent developments and pitfalls," *Digital Finance*, Jg. 4, Nr. 1, S. 63–88, 2022, ISSN: 2524-6186. DOI: 10.1007/s42521-021-00046-2. Adresse: <https://doi.org/10.1007/s42521-021-00046-2>.
- [5] G. Bamberg, F. Baur und M. Krapp, *Statistik - Eine Einführung für Wirtschafts- und Sozialwissenschaftler*. Berlin, Boston: De Gruyter Oldenbourg, 2022, ISBN: 9783110759327. DOI: <https://doi.org/10.1515/9783110759327>.
- [6] G. Bourier, *Beschreibende Statistik, Praxisorientierte Einführung – Mit Aufgaben und Lösungen*, 14. Aufl. Springer Vieweg Berlin, Heidelberg, 2022, S. XI, 279, ISBN: 978-3-658-37020-6. Adresse: <https://doi.org/10.1007/978-3-658-37021-3>.
- [7] D. L. Olson und G. Lauhoff, *Deskriptives Data-Mining*, 1. Aufl. Springer Gabler Cham, 2023, S. XII, 147, ISBN: 978-3-031-21273-4. DOI: 10.1007/978-3-031-21274-1. Adresse: <https://doi.org/10.1007/978-3-031-21274-1>.
- [8] X.-S. Yang, *Introduction to Algorithms for Data Mining and Machine Learning*. Academic Press, 2019, S. 173, ISBN: 978-0-12-817216-2. DOI: <https://doi.org/10.1016/C2018-0-02034-4>. Adresse: <https://www.sciencedirect.com/book/9780128172162/introduction-to>

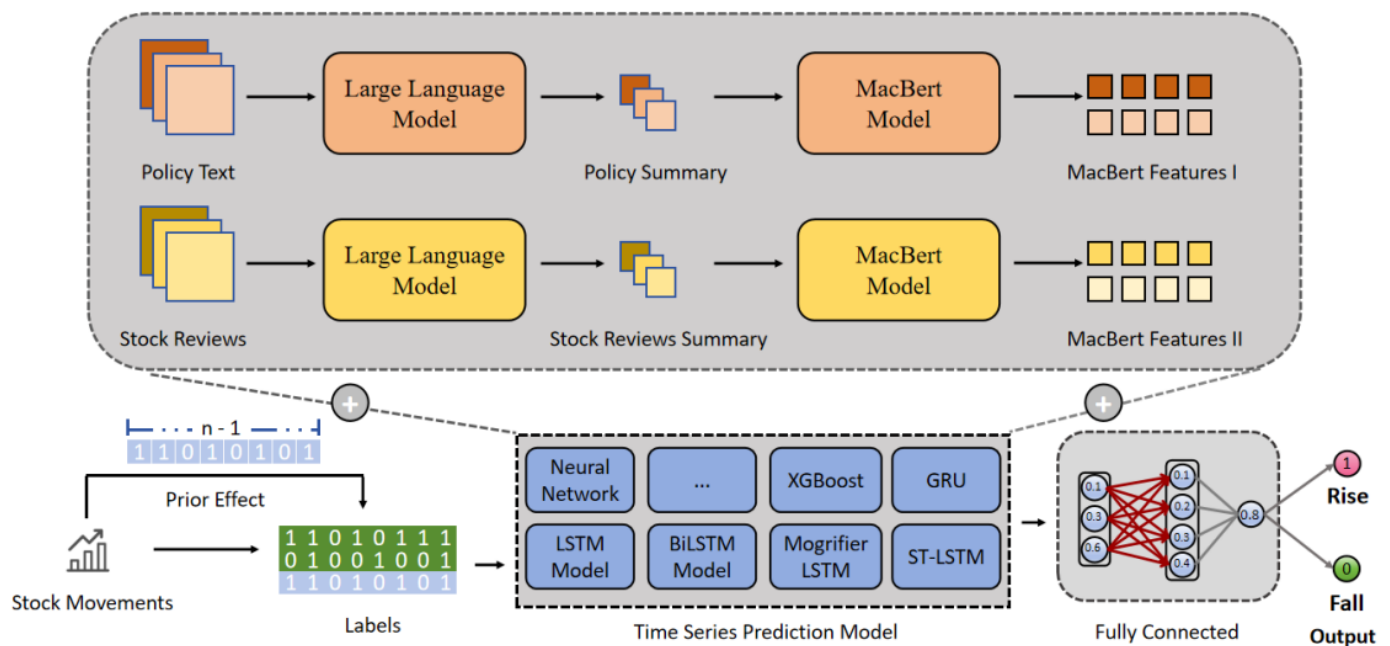


Abbildung 1: Modellarchitektur mit Implementierung kontextbewusster Multi-Source Datenfusion aus [1]

- algorithms-for-data-mining-and-machine-learning?via=ihub=#book-description.
- [9] S. Liu, J. McGree, Z. Ge und Y. Xie, *Computational and Statistical Methods for Analysing Big Data with Applications*. Academic Press, 2016, S. 194, ISBN: 978-0-12-803732-4. DOI: <https://doi.org/10.1016/C2015-0-00198-8>. Adresse: <https://www.sciencedirect.com/book/9780128037324/computational-and-statistical-methods-for-analysing-big-data-with-applications#book-description>.
- [10] G. Görz, U. Schmid und T. Braun, Hrsg., *Handbuch der Künstlichen Intelligenz*. Berlin, Boston: De Gruyter Oldenbourg, 2021, ISBN: 9783110659948. DOI: [doi:10.1515/9783110659948](https://doi.org/10.1515/9783110659948). Adresse: <https://doi.org/10.1515/9783110659948>.
- [11] A. Knoblauch, *Mathematik für Informatik und Data Science, Eine fundierte Einführung in Logik, Analysis, Lineare Algebra und Stochastik für Künstliche Intelligenz und Maschinelles Lernen* (Studienbücher Informatik), 1. Aufl. Springer Vieweg Berlin, Heidelberg, 2024, S. XVIII, 416, Der/die Herausgeber bzw. der/die Autor(en), exklusiv lizenziert an Springer-Verlag GmbH, DE, ein Teil von Springer Nature 2024, ISBN: 978-3-662-69478-7. Adresse: <https://doi.org/10.1007/978-3-662-69479-4>.
- [12] K. Benidis, S. S. Rangapuram, V. Flunkert et al. "Deep Learning for Time Series Forecasting: Tutorial and Literature Survey," *ACM Computing Surveys*, Jg. 55, Nr. 6, S. 1–36, Dez. 2022, ISSN: 1557-7341. DOI: [10.1145/3533382](https://doi.org/10.1145/3533382). Adresse: <http://dx.doi.org/10.1145/3533382>.
- [13] S. Cheriyan, S. Ibrahim, S. Mohanan und S. Treesa, "Intelligent Sales Prediction Using Machine Learning Techniques," in *2018 International Conference on Computing, Electronics & Communications Engineering (iCCECE)*, 2018, S. 53–58. DOI: [10.1109/iCCECOME.2018.8659115](https://doi.org/10.1109/iCCECOME.2018.8659115).