

Machine Learning im Retourenmanagement von Hamide Kahraman

Arbeitspapier des Lehrgebiets
Datenbanken und E-Business
No. 2/2023

herausgegeben von
Thomas C. Rakow



Bild: Radschläger auf Campus HSD.
© Thomas Rakow.

© Die Autorin 2023

Open Access: Dieser Beitrag wird unter der Creative Commons Namensnennung - Weitergabe unter gleichen Bedingungen 4.0 International Lizenz (<https://creativecommons.org/licenses/by-sa/4.0/deed.de>) veröffentlicht.

Zum Geleit

Das Retourenaufkommen im B2C-Handel hat sich zu einem ökonomischen sowie ökologischen Problem entwickelt. 2021 wurden in Deutschland 1,3 Milliarden Artikel zurückgeschickt.¹ Die Kosten allein für Transport und Bearbeitung betragen etwa € 3.700.000.000. Zwar können im Mittel 93% der retournierten Artikel als neuwertig verkauft werden, wobei kaum die ursprünglichen Preise realisiert werden können. Geschätzt 795.000 Tonnen CO₂-Ausstoß wurden durch die Retouren verursacht, was etwa dem CO₂-Ausstoß eines Zementwerks entspricht.²

Anreiz genug, um neuere Methoden wie das Machine Learning zur Verminderung von Retouren einzusetzen. In dieser Arbeit werden wissenschaftliche Artikel zum Retourenmanagement recherchiert, ob und welche Techniken eingesetzt werden. Frau Kahraman kategorisiert die Ansätze passend und beschreibt die Ergebnisse der jeweiligen Methode. Die angegebene Literatur liefert einen Überblick zum derzeitigen Stand der Forschung im Bereich Retourenmanagement mit Machine Learning.

Frau Kahraman hat diese Studienarbeit als Prüfungsleistung im B.Sc. Medieninformatik der Hochschule Düsseldorf erstellt.

Düsseldorf, 22.06.2023

Professor Dr.-Ing. Thomas C. Rakow

¹ Retourenforschung.de, Otto-Friedrich-Universität Bamberg, Lehrstuhl für Allgemeine Betriebswirtschaftslehre, insb. Produktion und Logistik: Ergebnisse des Europäischen Retourentachos veröffentlicht. Pressemitteilung 20.09.2022. <http://www.retourenforschung.de/info-ergebnisse-des-europaeischen-retourentachos-veroeffentlicht.html>.

² mdr.de: Top 30 CO₂-Verursacher in der Industrie: Fünf davon in Mitteldeutschland. Stand: 07. Juni 2023, 10:21 Uhr. <https://www.mdr.de/wissen/treibhausgas-emissionen-in-deutschland-industrie-eisen-und-stahlerzeugung100.html>.

Zusammenfassung

Machine Learning – eine Disziplin der IT, die in letzter Zeit immens an Traktion gewonnen hat. Durch die verzeichneten Entwicklungen, wie z.B. autonomes Fahren, ChatGPT und noch vielen mehr, haben sich in zahlreichen Sparten neue Perspektiven eröffnet. Die vorliegende Arbeit untersucht die Einsatzmöglichkeiten von Machine Learning in der Retourenmanagement Domäne, welche bislang noch wenig erforscht wurden. Gerade der boomende Online-Handel und die daraus resultierende Masse an Retouren verdeutlichen die Notwendigkeit einer Intervention mit dem Ziel, Retouren, die sowohl den Unternehmen als auch der Umwelt schaden, zu minimieren. Ob klassische Ansätze wie logistische Regression und lineare Diskriminanzanalyse oder state-of-the-art Methoden wie neuronale Netzwerke, Entscheidungsbäume, Ensemble Machine Learning, Graphen, etc. – Machine Learning bietet der Wirtschaft viele Werkzeuge, die für die Lösung des Retourenproblems eingesetzt werden können. Die Auswahl der Methoden hängt von dem Ziel und den verfügbaren Daten ab, aber die schiere Anzahl der Möglichkeiten verdeutlicht, dass mehrere Wege für die Reduzierung von Retouren eingeschlagen werden können. Abhängig von Know-How und technischen Voraussetzungen können auch einfachere Methoden, die keine allzu hohe Rechenleistung bzw. wenig Expertise benötigen, den Unternehmen äußerst nützliche Resultate liefern. Die vorliegende Arbeit versucht den aktuellen wissenschaftlichen Status-Quo darzustellen, um dadurch die den Online-Händlern verfügbaren Möglichkeiten aufzuzeigen.

Schlagwörter: *Machine Learning, Retourenmanagement, Retourenminimierung*

Abstract

Machine Learning, as an IT discipline, has gained immense traction in recent years. Developments such as autonomous driving, ChatGPT and many others, introduce fresh perspectives to numerous sectors. This research paper explores potential machine learning applications in the field of returns management, which have been examined only little thus far. The tremendous growth of e-commerce and the ensuing volume of returns highlight the necessity of a seller-side intervention to reduce returns, that harm both companies and the environment. Ranging from classical approaches such as logistic regression and linear discriminant analysis to state-of-the-art methods including neural networks, decision trees, ensemble machine learning, graphs, etc. – machine learning offers the e-commerce industry a myriad of tools that can be utilised to address the returns issue. The selection of the method depends highly on the objective and the available data, but the sheer number of options illustrates that multiple paths can be pursued to mitigate returns. Depending on know-how and technical prerequisites, even simpler methods that do not require excessive computing power or expertise in the subject matter, can provide businesses with valuable insights. This paper attempts to depict the current scientific status quo to demonstrate available options to online retailers.

Keywords: *machine learning, returns management, returns mitigation*

Inhaltsverzeichnis

1	Einleitung	1
2	Methodik der Recherche	3
3	Ergebnisse	7
3.1	Klassifikation.....	7
3.2	Feature Extraction	13
3.3	Graphen	15
3.4	Causal Machine Learning	18
3.5	Rule-Based Machine Learning	20
4	Fazit	21
4.1	Lineare Methoden.....	23
4.2	Neuronale Netzwerke	23
4.3	Deep Learning	24
4.4	Entscheidungsbäume	25
4.4.1	Ensemble Machine Learning	26
4.4.2	Gradient Boosting.....	26
4.4.3	Random Forest / Tree	27
4.4.4	Regressionsbäume	28
4.5	Support Vector Machines	28
4.6	Rule-Based Machine Learning	29
4.7	Causal Machine Learning	30
4.8	Graphen	31

5	Ausblick	32
6	Quellenverzeichnis	35
7	Abbildungsverzeichnis	40
8	Glossar	41

1 Einleitung

Der weltweite Online-Handel erfreut sich seit Jahren einem starken Zuwachs, so stieg der Umsatz von 2,382 Milliarden US-Dollar im Jahr 2017 auf 5,717 Milliarden US-Dollar in 2022 (Cramer-Flood, 2022, zit. nach Statista, 2022a). Die Corona-Pandemie verlieh dem E-Commerce einen zusätzlichen Schub (OECD, 2020). Durch die Einschränkungen im Alltag haben sich auch am Einkaufsverhalten der Menschen gezwungenermaßen Änderungen ergeben. Laut einer Statista „Umfrage zur Änderung des Einkaufsverhaltens während der Corona-Krise 2020 und 2021“ (Statista, 2021) haben 36 % der an der Umfrage teilgenommenen Menschen angegeben, „dass sie Einkäufe, die sie normalerweise im Geschäft erledigt hätten, online getätigt haben“ (Statista, 2021). Während die Online-Händler mehr Verkäufe verzeichnen konnten, ist die Kehrseite der Medaille jedoch nicht allzu rosig. Denn ein Umsatzwachstum bringt auch eine Steigerung der Retouren mit sich. Allein im Jahre 2020 wurden in Deutschland schätzungsweise 315 Millionen Pakete retourniert (Asdecker, Felch, Karl, 2022, zit. nach Statista, 2022b). Insbesondere durch kompetitive Strategien der Online-Händler, die mit kostenlosem Versand und in einigen Fällen sogar mit kostenlosen Retouren werben, werden die Kunden dazu verleitet, eher riskantere bzw. unbedachte Käufe zu tätigen, die oftmals auch retourniert werden. Eine Vielzahl an Studien haben bewiesen, dass der kostenlose Versand bzw. die kostenlose Retoure einen großen Einfluss auf die Retourenquote haben (Shehu, et al., 2020, S. 3) (Frei, et al., 2022, S. 1). Hinzu kommen sehr großzügige Rückgabebedingungen, die den Kunden u.a. ein 30-tägiges Rückgaberecht einräumen. Diese Faktoren führen zu mehr Retouren, die ein Dorn im Auge des E-Commerce sind.

Ziel der vorliegenden Arbeit ist es nicht, Retouren im Allgemeinen zu betrachten oder den Einfluss von verschiedenen Unternehmensstrategien auf die Retourenrate zu analysieren. Vielmehr liegt das Augenmerk auf einer der spannendsten Entwicklungen in der IT – dem Machine Learning und den möglichen Einsatzgebieten im Bereich des Retourenmanagements. Der Begriff des Retourenmanagements umfasst alle Aktivitäten beginnend von der Initiierung der Retoure bis hin zur Bearbeitung der Retoure im Warenhaus des betroffenen Händlers. Es wird eine bewusste Abgrenzung zu Reverse Logistics vorgenommen, welches auch die Rückführung von Materialien und Ressourcen analysiert (Ahsan, et al., 2021, S.

12). Da die Menge und Qualität der Retouren nur sehr schwierig vorausgesagt werden können, ist der Einsatz von Algorithmen unabdingbar, um aus den vorhandenen Daten Erkenntnisse zu ziehen, die bei der Prognose von zukünftigen Fällen eingesetzt werden können.

Künstliche Intelligenz (KI) und Machine Learning haben in letzter Zeit immens an Traktion gewonnen. Neben Anwendungen wie z.B. autonomes Fahren, Gesichtserkennung, virtuelle Assistenten zeigt insbesondere der Durchbruch von Systemen wie z.B. ChatGPT und OpenAI die Wichtigkeit von künstlicher Intelligenz und Machine Learning. Während die beiden Begriffe oftmals unberechtigterweise synonym eingesetzt werden, unterscheidet die vorliegende Arbeit die Begriffe. Machine Learning bildet einen Teilbereich von künstlicher Intelligenz. Den wichtigsten Grundbaustein für Machine Learning stellen dabei die Daten dar, denn anhand dieser lernt das System und optimiert sein Verhalten (Frochte, 2019, S. 13). Die Algorithmen ermitteln Muster in den Daten, die u.a. für das Aufstellen von Prognosen eingesetzt werden (SAP, o. J.). Künstliche Intelligenz hingegen weist ähnliche intellektuelle Fähigkeiten wie ein Mensch auf (Frochte, 2019, S. 13) und ist in der Lage Aufgaben eigenständig zu erledigen und dynamisch auf neue Gegebenheiten zu reagieren (Stadler, 2022). Die künstliche Intelligenz umfasst also mehr als das Machine Learning.

In einem ersten Schritt werden die eingesetzten Recherchemethoden kurz umrissen. Neben den zur Rate gezogenen Datenbanken und Quellen werden ebenfalls die für die Ermittlung der Artikel verwendeten Suchbegriffe aufgezeigt. Dieser Abschnitt enthält eine Zusammenstellung der ermittelten Artikel in tabellarischer Form, der bereits eine erste Einteilung in die Themenbereiche zu entnehmen ist. Der nächste Abschnitt geht im Detail auf die im Vorfeld ausgewählten elf Artikel ein. Diese werden in einem Fazit evaluiert. Zum Abschluss soll ein Ausblick einen Blick in die Zukunft werfen und mögliche Entwicklungen aufgreifen. Dies erfolgt u.a. auch unter Zuhilfenahme von Limitationen der erläuterten Artikel. Basierend auf diesen soll analysiert werden, welche Entwicklungsmöglichkeiten in Zukunft denkbar wären.

Abschließend soll an dieser Stelle vermerkt werden, dass es nicht möglich ist alle in dieser Arbeit aufgeführten Termini und Konzepte zu erklären, da es sonst den Rahmen sprengen würde. Begriffe, die im Text nicht erläutert werden, werden im Glossar grob definiert.

2 Methodik der Recherche

Für die systematische Literaturrecherche wurden unterschiedliche Literaturdatenbanken hinzugezogen. In einem ersten Schritt wurden die im weiteren Verlauf aufgeführten Suchbegriffe in Google Scholar eingegeben. Ausgehend von den Artikeltiteln sowie dem Publikationsjahr wurden die potenziell relevanten Publikationen aufgerufen und überflogen. Dabei wurden nur diejenigen Artikel in die engere Auswahl genommen, die nicht vor mehr als acht Jahren publiziert wurden. Somit soll gewährleistet werden, dass der letzte wissenschaftliche Stand des Untersuchungsgegenstandes akkurat abgebildet wird. Die Online-Kataloge der Hochschule Düsseldorf sowie der Heinrich-Heine-Universität wurden ebenfalls konsultiert. Außerdem wurde die Plattform Research Gate nach relevanten Artikeln durchsucht.

Die Erarbeitung der Schlagwörter erfolgte in mehreren Phasen. Erst wurden grobe Suchbegriffe, wie Retourenmanagement und Machine Learning genutzt. Unter Verwendung von Synonymen und spezifischeren Begriffen wurde anschließend ein Netz aus Schlagwörtern gewoben. Die folgenden Suchbegriffe wurden im Rahmen der Recherche eingesetzt.

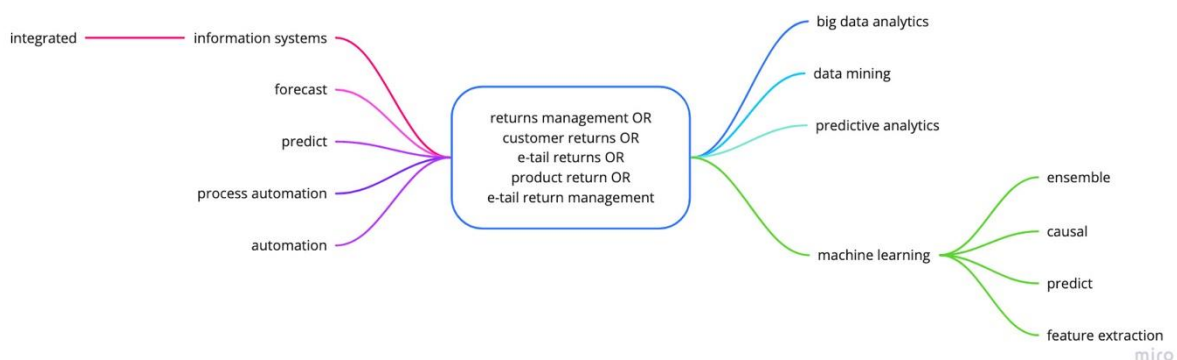


Abbildung 1 Englische Suchbegriffe

Die im mittleren Bereich der Abbildungen 1 und 2 zu sehenden Begriffe wurden mit einer oder-Verknüpfung versehen und anschließend nacheinander mit den unmittelbar folgenden Knoten mit einer und-Verknüpfung verbunden. Eine beispielhafte Suchanfrage ist returns management OR customer returns OR e-tail returns OR product return OR e-tail return management AND data mining.

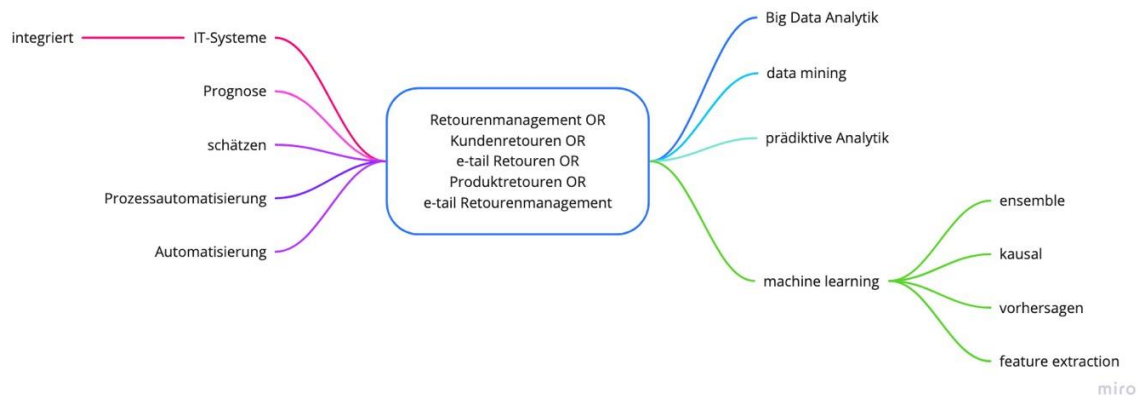


Abbildung 2 Deutsche Suchbegriffe

Die in die engere Auswahl genommenen Artikel wurden auf ihre Bibliografie hin untersucht und das so genannte Snowballing Verfahren wurde angewandt. Da die ausgewählten Artikel bereits eine hohe Relevanz für die Thematik der vorliegenden Arbeit aufwiesen, lag es nahe, dass die für die Verfassung der Artikel genutzten Publikationen auch einen Mehrwert für diese Arbeit bilden könnten. Anhand der Titel wurden die Veröffentlichungen in den jeweiligen Bibliografien eingegrenzt.

Die im Rahmen der Recherche ermittelten Artikel werden in nachfolgender Tabelle nach Erscheinungsjahr sortiert aufgeführt. Neben den Gründen, weshalb gerade diese Beiträge in die engere Auswahl genommen wurden, werden auch die in den Artikeln behandelten Themen aufgeführt. Eine große Rolle spielt die Spalte Kategorie. Die hier aufgeführten Begriffe sind im weiteren Verlauf der Arbeit wiederzufinden. Anhand dieser Kategorien werden die Publikationen geclustert und erläutert. Die in der Wissenschaft zentralen Texte werden als solche in der Spalte Leitartikel kenntlich gemacht.

Die Artikel wurden aufgrund ihrer Relevanz im Hinblick auf die Thematik der vorliegenden Arbeit ausgesucht. Alle aufgeführten Veröffentlichungen und Ausarbeitungen behandeln Machine Learning Methoden und Algorithmen im Retourenmanagement-Kontext. Trotz der immensen Traktion der Themen Machine Learning und künstliche Intelligenz sind diese im Zusammenhang mit Retourenmanagement weniger erforscht worden. Dies spiegelt sich auch in der nachfolgenden Literaturtabelle wider. Der Vollständigkeit halber werden neben peer-reviewed Artikeln auch Masterarbeiten aufgelistet, auf die im weiteren Verlauf nicht näher eingegangen wird.

Leitartikel	Auswahlgründe	Themen	Kategorie	Autoren	Titel	Veröffentlichung	Jahr	Anzahl Zitierungen	Link
	Minimierung von Retouren durch Causal Machine Learning und Green Nudge.	Green Nudge, Causal Machine Learning, Causal Forest, Retourenminimierung, Rückgeberichtlinien	Causal Machine Learning	von Zahn, M., Bauer, K., Mihale-Wilson, C., Jagow, J., Speicher, M., Hinz, O.	The smart green nudge: Reducing product returns through enriched digital footprints & causal machine learning	SAFE Working Paper, No. 2022 363 SSRN Electronic Journal		-	SSRN
	Einsatz von Ensemble Machine Learning für die Prognose von Retouren.	Ensemble Machine Learning, Prediction Modelling, Retourenquote	Ensemble Machine Learning	Tüylü, A. N. A., Eroglu, E.	The prediction of product return rates with ensemble machine learning algorithms	Journal of Engineering Research	2022	1	Kuwait Journals
	Einsatz von Machine Learning für die Erarbeitung eines Entscheidungssystems zur Verbesserung der Kundenzufriedenheit.	Ensemble Machine Learning, Boosting, LightGBM, Random Forest, Entscheidungsbäume, Retouren, Kundenzufriedenheit	Klassifikation	Cosar Sogukkuyu, D. Y., Senvar, O., Aysoysal, B., Varol, M. A., Polat, M. F., Sertbas, S., Caglar, G., et al.	A Supervised Learning Algorithms for Consumer Product Returns Case Study for FLO Offline Stores	Intelligent and Fuzzy Systems: Digital Acceleration and The New Normal-Proceedings of the INFUS 2022 Conference, Volume 2 (Vol. 505, 190)	2022	-	Springer Link
	Einsatz von unsupervised Machine Learning für die Prognose von Retouren.	Unsupervised Machine Learning, Logistic Regression, Retouren	Logistische Regression	Granov, A.	Customer loyalty, return and churn prediction through machine learning methods: for a Swedish fashion and e-commerce company	Master's thesis, Umeå University	2021	-	Diva Portal
	Prognose von Retouren (ML) basierend auf dem Warenkorb eines Kunden.	Entscheidungssystem, binäre Klassifikation, LightGBM, Retouren	Klassifikation	Hofmann, A., Gwinner, F., Fuchs, K., Winkelmann, A.	An Industry-Agnostic Approach for the Prediction of Return Shipments	AMCIS	2020	1	ResearchGate
	Einsatz von Machine Learning für die Prognose von Retouren.	Data Mining, XGBoost, LightGBM, CatBoost, TabNet, Decision Tree Algorithm, Retouren	Klassifikation	Al Imran, A., Amin, M. N.	Predicting the return of orders in the e-tail industry accompanying with model interpretation	Procedia Computer Science, 176, 1170-1179	2020	4	Science Direct
X	Prognose von Retourenmissbrauch mittels ML und prädiktiver Analytik.	Klassifikation, Missbrauch von Retouren	Klassifikation	Ketzenberg, M. E., Abbey, J. D., Heim, G. R., Kumar, S.	Assessing customer return behaviors through data analytics	Journal of Operations Management, 66(6), 622-645	2020	23	Wiley
	Einsatz von ML für die Prognose von Retouren.	Machine Learning, Retouren	Logistische Regression	Maljaars, M.	An advice to improve the short-term return forecast	Master's thesis, University of Twente	2020	-	University of Twente
	Einsatz von Machine Learning für die Prognose von Retouren.	Machine Learning, Regression Analysis, Data Mining	Klassifikation	TÜYLÜ, A. N. A., Eroglu, E.	Using Machine Learning Algorithms For Forecasting Rate of Return Product In Reverse Logistics Process	Alphanumeric Journal, 7(1), 143-156	2019	4	Dergipark

Leitartikel	Auswahlgründe	Themen	Kategorie	Autoren	Titel	Veröffentlichung	Jahr	Anzahl Zitiertungen	Link
	Einsatz von Machine Learning für die Klassifikation von Retouren.	Data Mining, Rule-Based Machine Learning, JRip, Retouren	Rule-Based Machine Learning	Seewald, A. K., Wernbacher, T., Pfeiffer, A., Denk, N., Platzer, M., Berger, M., Winter, T.	Towards Minimizing e-Commerce Returns for Clothing	ICAART (2), 801-808	2019	7	Semantic Scholar
	Ermittlung von Retouren mittels Kaufhistorie des Kunden.	Machine Learning, Hybrid Graph, Retouren	Graph	Zhu, Y., Li, J., He, J., Quanz, B. L., Deshpande, A. A.	A Local Algorithm for Product Return Prediction in E-Commerce	IJCAI, 3718-3724	2018	12	IJCAI
X	Einsatz von Machine Learning für die Prognose von Retouren basierend auf dem Warenkorb eines Kunden.	Machine Learning, Data Mining, Hyper Graph, K-NN, JacWght, JacNorm, Retouren	Graph	Li, J., He, J., Zhu, Y.	E-tail product return prediction via hypergraph-based local graph cut	Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, 519-527	2018	24	ACM
X	Prognose von Retouren durch ML und Produktfotos.	Machine Learning, Feature Extraction, Histogramm, Gabor Filter, CNN	Feature Extraction	Dzyabura, D., El Kihal, S., Ibragimov, M.	Leveraging the power of images in predicting product return rates	SSRN Electronic Journal, 1-33	2018	14	NYU
	Einsatz von Data Analysis Methoden für die Prognose von Retouren.	Ensemble Machine Learning, Decision Tree, Neuronal Networks, Logistische Regression, Retourenmanagement	Klassifikation	Asdecker, B., Karl, D.	Big data analytics in returns management-Are complex techniques necessary to forecast consumer returns properly?	2nd International Conference on Advanced Research Methods and Analytics Proceedings	2018	4	INIS
	Einsatz von ML für die Prognose von Retouren basierend auf der Kauf- bzw. Retourenhistorie des Kunden.	Network Science, Machine Learning, Retouren	Bipartite Heterogeneous Information Network	Joshi, T., Mukherjee, A., Ippadi, G.	One size does not fit all: Predicting product returns in e-commerce platforms	IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, 926-927	2018	8	IEEE
	Möglicher Einsatz von ML für die Verbesserung des Retourenprozesses.	Retourenmanagement, Machine Learning, Retouren	(Un)Supervised Machine Learning	Westmaas, D.	Product returns in the e-commerce market, explorative study into the applicability of Machine Learning	Research paper Master thesis	2018	-	TU Delft
X	Prognose von Retouren.	Machine Learning, Prädiktive Analytik, Retouren	Predictive Model Selection	Ma, J., Kim, H. M.	Predictive model selection for forecasting product returns	Journal of Mechanical Design, 138(5)	2016	15	University of Illinois
X	Einsatz von Machine Learning für real-time Prognosen von Retouren.	Machine Learning, Feature Extraction, Logistic Regression, Support Vector Machine, Retourenbedingungen	Feature Extraction	Urbanke, P., Kranz, J., Kolbe, L.	Predicting product returns in e-commerce: the contribution of mahalanobis feature extraction	Thirty Sixth International Conference on Information Systems, Forth Worth	2015	17	ResearchGate

Tabelle 1 Rechercheergebnisse

3 Ergebnisse

Elf der in der Literaturtabelle aufgeführten Artikel werden nachfolgend aufgegriffen und erläutert. In einer ersten Analyse konnten die thematischen Schwerpunkte Klassifikation, Feature Extraction, Graphen, Causal Machine Learning und Rule-Based Machine Learning ermittelt werden. In den thematischen Unterkapiteln werden aufgrund ihrer Bedeutsamkeit zunächst die Leitartikel aufgegriffen. Alle weiteren Artikel werden nach ihrem Erscheinungsjahr sortiert aufgeführt und näher betrachtet. Für die Themenbereiche Klassifikation und Graphen gibt es jeweils einen Leitartikel. Das Thema Feature Extraction weist zwei Leitartikel auf, während die verbleibenden zwei Themenbereiche, Causal Machine Learning und Rule-Based Machine Learning keine Leitartikel haben.

3.1 Klassifikation

Die nachfolgend erläuterten fünf Artikel befassen sich mit der Klassifikation. Die Klassifikation ist ein fundamentaler Bereich des supervised Machine Learning. Vereinfacht ausgedrückt ordnet die Klassifikation den Untersuchungsgegenstand in bestimmte bereits definierte Kategorien. Die hier erläuterten Artikel betrachten Retouren und ihre Prognose bzw. Handhabung als ein binäres Klassifikationsproblem. Die eingesetzten Algorithmen ordnen die untersuchte Variable, wie z.B. bestimmte Produkte oder Warenkörbe in die Klassen „Retoure“ bzw. „keine Retoure“ ein.

Ketzenberg et al. (2020) erarbeiten in ihrem Artikel ein System für die Prognose und Verhinderung von Retourenmissbrauch. Sie untersuchen insbesondere Eigenschaften und Verhaltensstrukturen von Kunden, die Rückgabebedingungen missbrauchen. Die Prognosen erfolgen bei Initiierung der Retoure und werden aus Sicht der Wirtschaftlichkeit vorgenommen. Dabei werden gezielt Unterschiede zwischen Kunden, die Rückgabebedingungen missbrauchen und denen, die berechtigterweise Retouren initiieren, identifiziert, denn nicht alle Retouren erweisen sich für die Unternehmen als unwirtschaftlich. Die Wirtschaftlichkeit

berechnet sich aus der Gewinnspanne, den Erstattungen, dem Restwert der retournierten Artikel sowie den Bearbeitungskosten der Retouren.

Für ihre Untersuchungen setzen die Autoren sieben Jahre (1999-2006) umfassende Daten eines amerikanischen Händlers ein. Die Retourenrate in dieser Zeitspanne betrug 15,7 %. Trotz der Unausgewogenheit der Daten (weniger Retouren als keine Retouren), einer so genannten Klassen-Imbalance, werden keine korrigierenden Aktionen, wie z.B. Over- bzw. Undersampling vorgenommen. Die Autoren berufen sich dabei auf wissenschaftliche Studien, die durch den Einsatz von Over- bzw. Undersampling nur minimale Performanzunterschiede feststellen konnten.

Zur Lösung des binären Klassifikationsproblems werden logistische Regression, Support Vector Machine, Random Forest und neuronales Netzwerk in Kombination mit den Regularisierungsmethoden Lasso, Ridge Regression und Elastic Net eingesetzt und observiert. Die für die logistische Regression nicht relevanten Variablen werden mittels Rückwärtselimination schrittweise entfernt. Der Random Forest, die Support Vector Machine und das neuronale Netzwerk werden einer Justierung unterzogen, um die bestmöglichen Hyperparameter zu ermitteln. Der Random Forest mit fünf Features und 1000 Bäumen, die Support Vector Machine mit einem linearen Kernel, und das neuronale Netzwerk mit acht Neuronen in einer versteckten Schicht sind, gemessen anhand der Metrik Accuracy, am performantesten.

Für den Vergleich der Modelle wird die Metrik AUC (Area under the Curve) hinzugezogen. Das neuronale Netzwerk und der Random Forest erzielen mit einer AUC von 0,993 bzw. 0,992 die besten Resultate. Zur weiteren Analyse der Effizienz der Algorithmen wird Backtesting durchgeführt, in dem alle historischen Retourendaten zum Einsatz kommen. Auch hier übertrifft das neuronale Netzwerk performanztechnisch die anderen betrachteten Algorithmen und eignet sich daher für die Ermittlung von Retourenmissbrauch.

Die Autoren stellen fest, dass Kunden, die nachsichtige Rückgabebedingungen ausnutzen, mehr Bestellungen mit teureren Produkten über eine große Bandbreite an Kategorien aufgeben und eine Vielzahl an Retouren einleiten. Dies schlägt sich in den Erstattungsbeträgen wieder, die 12,6 % höher ausfallen als die der legitimen Retouren. Retouren sind aber grund-

sätzlich nicht problematisch, denn Kunden, die keine Retouren veranlassen, sind aus Wirtschaftlichkeitssicht weniger profitabel als Kunden mit legitimen Retouren, da diese zu weniger Käufen tendieren.

Cosar Sogukkuyu et al. (2022) erarbeiten in ihrer Studie ein Entscheidungssystem für die effizientere Handhabung des Retourenprozesses. Das System entscheidet über die Annahme bzw. Verweigerung von Retouren basierend auf Faktoren wie Kundensegment, verstrichene Zeit zwischen Kauf und Retoure, Hersteller- und Produktinformationen. Die Daten für die Studie werden den Autoren von einem Schuhhersteller zur Verfügung gestellt. Diese umfassen neben CRM-Daten, auch Informationen zum Umsatz und den Produkten.

Für die Aufstellung des Entscheidungssystems kommen Ensemble Machine Learning Methoden, insbesondere Boosting Methoden zum Einsatz. Ensemble Machine Learning kombiniert verschiedene Modelle, auch Lerner genannt, um die Performanz des Gesamtsystems zu verbessern. Das Boosting trainiert dabei sequenziell schwache Modelle und eliminiert schrittweise Fehler, um sie so zu einem starken Modell zu wandeln. Die auf Entscheidungsbäumen basierenden Algorithmen LightGBM, Random Forest und Decision Trees werden für die Klassifikation eingesetzt. Entscheidungsbäume sind baumartig aufgebaut und bestehen aus Knoten, Ästen und Blättern. Ausgehend vom Wurzelknoten führen Äste, die verschiedene Entscheidungen darstellen, zu weiteren Ebenen und enden in den Blättern, den Ergebnissen. Das Zusammenfügen mehrerer Entscheidungsbäume bildet einen Random Forest. Dieser betrachtet die Ergebnisse der einzelnen Entscheidungsbäume und kombiniert sie zu einem Gesamtergebnis. LightGBM ist ein Gradient Boosting Framework und unterscheidet sich von den anderen Entscheidungsbäumen in der Richtung, in die die Bäume wachsen. Hier entfaltet sich der Baum in vertikaler Richtung.

Die Performanz der Modelle wird mittels Precision, Recall und F1-Score gemessen. Die Precision betrachtet lediglich die richtig und falsch positiven Prognosen, während der Recall die richtig Positiven und falsch Negativen berücksichtigt. Der F1-Score errechnet sich aus Precision und Recall. Mit einer Precision von 74 % und einem F1-Score von 72 % erzielt LightGBM das beste Ergebnis. In anderen Worten, das LightGBM prognostiziert 74 % der Fälle korrekt und wird daher von den Autoren für den Einsatz im Entscheidungssystem empfohlen.

Hofmann et al. (2020) erarbeiten ein Entscheidungssystem, welches, direkt in den Web-Shop eingebunden, Prognosen zu potenziellen Retouren in Echtzeit erstellen kann. Ausschlaggebend für die Prognose ist dabei der Warenkorb eines Kunden.

Die Bestell- und Retourendaten eines deutschen Großhandels werden für die Erarbeitung des Systems genutzt. Diese Daten werden einer Aufbereitung und Bereinigung unterzogen, bevor sie für die entsprechenden Algorithmen vorbereitet werden können. Aufgrund der geringen durchschnittlichen Retourenrate von 5,25 % wird bei einer Kopie der Originaldaten ein Oversampling eingesetzt, um der Unausgewogenheit der Daten (weniger Retouren als keine Retouren) entgegenzuwirken. Hierbei wird die Anzahl der Retouren künstlich erhöht.

Ein statistisches Modell wird zu Vergleichszwecken aufgesetzt anhand dessen die Performanz der Machine Learning Methoden gemessen werden kann. Dieses berechnet für jeden Artikel die Retourenwahrscheinlichkeit, indem es die Anzahl der Retouren durch die Anzahl der Bestellungen dividiert. Für das Machine Learning Modell wird auf jedem Artikel im Warenkorb eine One-Hot-Codierung vorgenommen, welche zu einer vektoriellen Darstellung führt. Zur Lösung des binären Klassifikationsproblems kommen die Algorithmen in dem LightGBM Framework, insbesondere Gradient Boosting, Random Forest und DART zum Einsatz. Für die Konfiguration dieser auf Entscheidungsbäumen basierenden Algorithmen werden 60 Blätter, eine Binary Log Loss Funktion und eine Lernrate von 0,1 verwendet.

Die drei Machine Learning Algorithmen werden jeweils mit ausgeglichenen (Oversampling) und unausgeglichenen Daten trainiert. Die Performanz der Algorithmen wird durch Receiver Operator Characteristics (ROC) Kurven, sowie die Metriken Precision und Recall ermittelt. Dabei erzielen die Modelle, die mit ausgeglichenen Daten trainiert wurden, bessere Resultate. Die Algorithmen Gradient Boosting und DART schneiden im Vergleich am besten ab.

In dem erarbeiteten Entscheidungssystem kommen je nach Confidence Score zwei Modelle zum Einsatz. Dieser Klassifizierungsschwellenwert gibt eine Selbsteinschätzung des Modells hinsichtlich der Genauigkeit der Prognose wieder. Ein mit ausgeglichenen Daten trainiertes Modell wird im Zusammenspiel mit einem niedrigeren Klassifizierungsschwellenwert eingesetzt, um Warnungen für mögliche Retouren auszugeben, die anschließend manuell überprüft und verarbeitet werden müssen. Ein Modell mit einer hohen

Precision soll in Prognosen mit einem hohen Confidence Score zum Einsatz kommen und ohne menschliche Intervention Demarketing Strategien einleiten.

Al Imran et al. (2020) setzen in ihrer Studie vier unterschiedliche Data Mining Techniken ein, um diejenigen zu ermitteln, die sich besonders gut für die Prognose von Retouren eignen. Die Autoren nutzen für ihre Analysen Daten eines bekannten Online-Händlers. Da die erhaltenen Daten weniger Retouren enthalten, liegt eine ungleichmäßige Verteilung vor. Für die Eliminierung der Unausgewogenheit setzen die Autoren auf einer Kopie der Originaldaten Undersampling ein, wobei Daten zur Erreichung einer Balance entfernt werden.

Zur Steigerung der Performanz wird auf den Daten Feature Engineering angewandt. Hierbei werden aus den bereits vorhandenen Attributen, den so genannten Features weitere Features mit hohem Informationsgehalt hergeleitet. Für die weitere Analyse werden insgesamt fünfzehn Features, davon acht kategorische und sieben numerische Features erarbeitet. Die kategorischen Features werden mittels eines James-Stein Encoders in numerische Werte umgewandelt. Die Aufbereitung der numerischen Features erfolgt durch eine Powertransformation.

Die Autoren untersuchen die Algorithmen XGBoost, LightGBM, CatBoost, TabNet und Decision Tree, wobei letzteres als Bezugsgröße dient. Bei den ersten drei Algorithmen handelt es sich um Ensemble Boosting Algorithmen, die auf Entscheidungsbäumen basieren. TabNet ist ein Deep Learning Netzwerk, welches neuronale Netze auf tabellarischen Daten anwendet. Die für die Algorithmen wichtigen Hyperparameter werden mittels einer Bayes'schen Optimierung ermittelt.

Die Evaluation der Modelle erfolgt über die Metriken Recall, ROC-AUC Score und F2-Score. Die Algorithmen werden sowohl mit den Originaldaten als auch mit den ausgeglichenen Daten trainiert. In beiden Szenarien übertrifft TabNet alle anderen Algorithmen. Zur Ermittlung der aussagekräftigsten Features wird eine Analyse des TabNet Modells sowie eine Morris Sensitivity Analyse vorgenommen. Beide Analysen ermitteln die Features Ort der Bestellung und Zahlungskondition als am wirkungsvollsten. So werden die meisten Retouren (56 %) aus großen Städten initiiert. Auch die gewählte Zahlungskondition spielt eine besondere Bedeutung, denn 98 % der Retouren weisen die Zahlungskondition „Nachnahme“ auf.

Asdecker et al. (2018) vergleichen in ihrem Artikel die Performanz unterschiedlicher Modelle zur Prognose von Retouren, um auch kleinen Unternehmen Möglichkeiten für den Einsatz von Machine Learning aufzuzeigen.

Der Prognose von Retouren liegt ein binäres Klassifikationsproblem zugrunde. Die Autoren untersuchen die Methoden binäre logistische Regression, lineare Diskriminanzanalyse, Multilayer Perceptron, C5.0 Algorithmus und Ensemble Machine Learning. Die binäre logistische Regression ermittelt die Wahrscheinlichkeit, mit der der Untersuchungsgegenstand in eine der beiden evaluierten Kategorien fällt. Die lineare Diskriminanzanalyse erfordert die Festlegung mindestens zweier Gruppen und berechnet für jede Beobachtung einen Wert, anhand dessen die Beobachtung einer der festgelegten Gruppen zugeordnet werden kann. Der Multilayer Perceptron, ein vorwärtsgerichtetes neuronales Netzwerk, zeichnet sich dadurch aus, dass die Informationen nur in eine Richtung fließen und kein Feedback erfolgt. Dieser wird durch Backpropagation trainiert, d.h. die Ausgabe des Netzwerks wird mit den bekannten Werten verglichen. Der C5.0 Algorithmus erstellt einen Entscheidungsbaum, der die Beobachtungen am besten klassifiziert. Ensemble Machine Learning kombiniert mehrere Algorithmen, um die Prognosefähigkeit des Systems zu stärken. In diesem Zusammenhang setzen die Autoren künstliche neuronale Netzwerke, C5.0, QUEST, CART, CHAID, binäre logistische Regression, lineare Diskriminanzanalyse, Bayes'sches Netzwerk und Nearest Neighbor ein.

Die ein Jahr umfassenden Daten (2012 / 2013) werden von einem deutschen Online-Händler zur Verfügung gestellt. Diese beinhalten neben Informationen zu den Bestellungen auch Informationen zu den Retouren. Die Retourenrate betrug in diesem Zeitraum 59,8 %.

Die Performanz der Algorithmen wird mittels der Total Absolute Error (TAE) Metrik gemessen, welches die Anzahl der falschen Prognosen wiedergibt. Die Ensemble Machine Learning Methode in der Kombination der Decision Tree Algorithmen C5.0, CHAID und QUEST erzielen die besten Resultate und sagen 68,45 % der Fälle korrekt vorher. Die weniger komplexen Methoden binäre logistische Regression (66,79 %) und lineare Diskriminanzanalyse (66,6 %) erzielen ebenfalls gute Ergebnisse. Die Autoren argumentieren, dass nicht immer die komplexeste Methode genutzt werden muss, um ein System für die Prog-

nose von Retouren aufzusetzen. Dies öffnet insbesondere kleineren Unternehmen mit eingeschränkten Ressourcen und minimalem Data Mining Wissen Türen für den Aufbau eines Systems für die Vorhersage von Retouren.

3.2 Feature Extraction

Die nachfolgenden zwei Artikel befassen sich jeweils mit Feature Extraction und stellen Leitartikel für den Themenbereich dar. Feature Extraction Methoden gruppieren Variablen zu Features und verringern durch diesen Prozess die eingesetzte Datenmenge. Ein Verlust von relevanten Informationen wird dabei vermieden (Usability, o. J.).

Dzyabura et al. (2018) setzen in ihrem Artikel Machine Learning ein, um die Retourenrate eines neuen Produktes anhand der Produktfotos zu ermitteln. Dies ermöglicht den Unternehmen frühzeitig geeignete Schritte für die Verbesserung des Produktes bzw. die erneute Evaluierung des angesetzten Preises einzuleiten. Die benötigten Daten werden ihnen von einem europäischen Omnichannel Händler zur Verfügung gestellt. Diese umfassen online und offline Daten aus den Jahren 2013 bis 2016.

Für die Prognose werden die Features Preis, Kategorie, Farbe und Produktbilder hinzugezogen. Um die in den Produktfotos enthaltenen Informationen in den Machine Learning Algorithmus einfließen lassen zu können, wenden die Autoren zunächst Methoden der Feature Extraction an. Hier kommen Farb-Histogramme, Gabor Filter und Convolutional Neural Networks (CNN) zum Einsatz. Ersteres ermöglicht die Reduzierung der Dimensionalität und die Erfassung des Verhältnisses zwischen Retourenrate und Produktfarbe. Gabor Filter isolieren die Periodizitäten innerhalb eines Bildes und eignen sich, um strukturelle Besonderheiten z.B. eines Kleidungsstücks (gestreift bzw. kariert) zu ermitteln. Das CNN erstellt Prognosen basierend auf den Features. Ausgehend von dem Originalbild in der ersten Schicht, wendet das System in den weiteren Schichten nichtlineare Transformationen an, um jeweils neue Features zu erhalten. Aufgrund der unzureichenden Größe der vorhandenen Daten werden Schichten bereits trainierter CNNs eingesetzt. Für die Prognose von Retourenquoten wird XGBoost, ein Gradient Boosted Regression Tree Modell verwendet.

Insgesamt testen die Autoren neun Modelle, die sie mit verschiedenen Eigenschaften trainieren. Die Performanz der Modelle wird mit der Accuracy Metrik festgestellt. Alle Modelle nutzen die Features Preis und Produktkategorie. Durch das Hinzunehmen von der Feature Farbe und den Produktfotos, verbessert sich die Prognose um ca. 30 %. Die beste Vorhersagegenauigkeit erzielt das Modell, welches neben den Features Preis und Kategorie, Farb-Histogramme und CNN einsetzt. Durch den Einsatz von ausschließlich dem Kunden während des Bestellvorgangs zugänglichen Daten stellen die Autoren fest, dass Retouren nicht nur wegen der fehlenden physischen Betrachtung des erworbenen Artikels generiert werden. Produkte, die offline weniger gut verkauft werden, resultieren bei mehr Verkäufen online in mehr Retouren.

Urbanke et al. (2015) erarbeiten in ihrer Studie ein Entscheidungssystem für die Prognose von Retouren durch den Einsatz von Feature Extraction. Dieses wird eingesetzt, um Muster im Konsumverhalten festzustellen, die Anzeichen von hohen Retourenraten sind. Die Prognose findet noch vor Abschluss der Bestellung statt und ermöglicht dem Händler einzugreifen und gezielt Demarketing Strategien einzusetzen. Hierunter fallen Strategien wie z.B. die Eingrenzung der verfügbaren Zahlungskonditionen (nur Vorkasse), die bewusste Erhöhung der Lieferzeit sowie der Einsatz von Green Nudges, technischen Interventionen, wie Pop-Up Fenster, die den Kunden vor Abschluss der Bestellung über die ökologischen Auswirkungen von Retouren informieren.

Die fünf Monate umfassenden Daten (Juli – November 2014) wurden von einem deutschen Online-Händler zur Verfügung gestellt. Die Retourenrate betrug in dieser Zeitspanne 57,3 %. Für die Berechnung der historischen Retourenrate eines Kunden werden Verkaufs- und Retourendaten der vorangegangenen zwölf Monate mit einbezogen. Die Daten werden in einer sparsen Matrix, in der nur von Null abweichende Werte gespeichert werden, dargestellt, um bei der Verarbeitung die Belastung von Speicher und CPU zu mindern.

Für die Minimierung der Dimensionalität von schwachbesetzten Matrizen wird ein Algorithmus erarbeitet. Ziel des Mahalanobis Feature Extraction Algorithmus ist es, den Mahalanobis Abstand zwischen dem observierten Wert und dem aus der Nullhypothese erwarteten Wert zu maximieren, so dass die Wahrscheinlichkeit, dass die Nullhypothese zutrifft, minimiert wird. Die Nullhypothese besagt, dass die Korrelationskoeffizienten ein Produkt aus der zufälligen Umstrukturierung der Daten sind. Hierfür wird ein stochastisches

Gradientenverfahren mit Minibatch Aktualisierungen verwendet. Der Algorithmus wird zusammen mit einem Adaptive Boosting Klassifikationsalgorithmus eingesetzt.

Der Mahalanobis Feature Extraction Algorithmus und der Adaptive Boosting Klassifikationsalgorithmus werden mit unterschiedlichen Methoden und Algorithmen verglichen. Die Feature Extraction bzw. Feature Selection Methoden Hauptkomponentenanalyse, lineare Diskriminanzanalyse, Singulärwertzerlegung, Merkmalsselektion, Random Projection und nicht-negative Matrix Faktorisierung werden in die Analysen mit einbezogen. Aus dem Bereich der Classifier kommen die Klassifikationsalgorithmen CART, Extremely Randomized Trees, Gradient Boosting, lineare Diskriminanzanalyse, logistische Regression, Random Forest und Linear Kernel Support Vector Machine zum Einsatz. Die Feature Extraction Techniken werden jeweils mit allen Classifiern kombiniert und gegenübergestellt.

Die Performanz wird mit den Metriken Korrelationskoeffizient, Precision, Recall und ROC-Kurve ermittelt. Die Kombination aus Adaptive Boosting und Mahalanobis Feature Extraction erzielt die besten Resultate. Bei einer Interventionsquote von 10 % wären 84,8 % der Interventionen des Systems berechtigt. Abhängig von der eingesetzten Demarketing Strategie würde dies zu einer Reduzierung der Retouren von 14,1 % führen.

3.3 Graphen

Die zwei nächsten Artikel beschäftigen sich mit den zwei Graphen Arten Hypergraphen bzw. hybride Graphen. Hypergraphen sind spezielle Graphen, die mit Hyperkanten ausgestattet sind. Diese sind in der Lage nicht nur zwei Knoten, sondern mehrere Knoten miteinander zu verbinden (TH Köln, o. J.). Der Artikel von Li et al. (2018) bildet für den Themenbereich einen Leitartikel und wird daher als Erster aufgeführt.

Li et al. (2018) erarbeiten in ihrer Studie HyperGo, ein System, welches die Retourenintention des Kunden nach Zusammenstellen des Warenkorb ermittelt. Diese wird auf Warenkorb- und Produktebene erfasst, d.h. die Dual-Level Retouren Prognose bestimmt in einem ersten Schritt den Warenkorb, der eine potenzielle Rücksendung enthält und ermittelt in einem zweiten Schritt, welches der Produkte zurückgesendet werden könnte. Die Feststellung der Retourenintention noch vor Abschluss der Bestellung ermöglicht es den Händler

unterschiedliche Demarketing Schritte einzuleiten. Die Daten für die Studie werden den Autoren von zwei Omnichannel Händlern zur Verfügung gestellt.

HyperGo basiert auf einem Hypergraphen, der die Kauf- und Retourenhistorie repräsentiert. Die Knoten stellen die Warenkörbe und die Kanten ein Produkt dar. Die Kanten verbinden alle Warenkörbe, die das entsprechende Produkt enthalten und clustern sie somit.

Ein Local Graph Cut Algorithmus wird auf dem Hypergraphen angewandt, um ähnliche Warenkörbe zu dem zu untersuchenden Warenkorb zu ermitteln. Er basiert auf dem Random-Walk und untersucht, beginnend von einem bestimmten Warenkorb, die unmittelbare Umgebung dieses Warenkorbs, um ähnliche Warenkorbkompositionen zu ermitteln. Die sich bildenden Cluster bestehen aus mehreren verbundenen Knoten, den Warenkörben und sind nur mit wenigen Kanten mit dem Rest des Hypergraphen verbunden. Als Parameter erwartet der Algorithmus einen Hypergraphen, den zu betrachtenden Warenkorb, die maximale Anzahl an Kanten mit denen der lokale Cluster mit dem Rest des Hypergraphen verbunden ist, und einen positiven Integerwert. Der Algorithmus gibt eine Reihe von Warenkorbknoten innerhalb des identifizierten lokalen Clusters zurück.

Zum Vergleich werden die auf Ähnlichkeit basierenden Modelle K-NN, JacWght und JacNorm hinzugezogen. K-NN findet die nächsten Nachbarwarenkörbe und schätzt die Retourenwahrscheinlichkeit unter Hinzunahme der gesamten Retouren ein. JacWght errechnet die Ähnlichkeit mit anderen Warenkörben, indem es einen Jaccard Index einsetzt. JacNorm stellt die Prognose von Retouren ähnlich wie JacWght auf, berücksichtigt aber die unterschiedlichen Retourenraten, die den verschiedenen Warenkorbkompositionen (nur diverse Produkte bzw. ähnliche Produkte) zugrunde liegen. Die durchschnittliche Retourenrate von Warenkörben mit mehreren ähnlichen Produkten ist höher als die von Warenkörben mit nur verschiedenen Produkten.

Die Bewertung der Effizienz von HyperGo wird anhand der Laufzeit vorgenommen. Für die Wirksamkeit des Systems werden die Metriken Precision, Recall, ROC und AUC hinzugezogen. HyperGo erzielt in den Experimenten die besten Resultate mit einem AUC-Wert von 0,7 und einer Precision von 81 %. Es modelliert die den historischen Daten zugrunde liegenden Informationen effizienter als die anderen untersuchten Methoden. Um den Einfluss der Eingabeparameter auf die Performanz von HyperGo zu analysieren, wird eine Parameter Sensitivity Analyse durchgeführt. Die Parameter Größe des Output Clusters sowie die

maximale Anzahl an Kanten mit denen der lokale Cluster mit dem Rest des Hypergraphen verbunden ist, werden mit unterschiedlichen Werten belegt und getestet. Der die Größe des Output Clusters kontrollierende Parameter erzielt die besten Resultate bei sechs Warenkörben. Bei einer geringeren Größe ist die für die Dual-Level Prognose genutzte Anzahl an Warenkörben zu gering, um präzise Prognosen aufzustellen. Je größer die Cluster Größe wird, desto mehr verrauschte Daten werden in die Prognose mit einbezogen. Die maximale Anzahl der Kanten mit denen der lokale Cluster mit dem Rest des Hypergraphen verbunden ist, hat einen Einfluss auf die Qualität des Output Clusters. Ein zu hoher Wert bringt keine Änderung mit sich, ein geringerer Wert erhöht die Qualität des Output Clusters. Dieser Wert sollte jedoch einen bestimmten Schwellenwert nicht unterstreiten, da dies zu größeren Output Clustern führt. Die Laufzeit wird durch den Vergleich von CPU-Zeit in Sekunden und der Anzahl an Knoten im Output Cluster bewertet. Die CPU-Zeit skaliert linear in Bezug auf die Anzahl der Knoten in dem zurückgegebenen Cluster, so dass sich das Modell besonders für den Einsatz mit großen Datenmengen eignet.

Zhu et al. (2018) nutzen in ihrer Studie HyGraph, einen gewichteten hybriden Graphen, und den LoGraph Algorithmus, um Prognosen zu Retouren aufzustellen. Dabei sollen Kunden identifiziert werden, die ein bestimmtes Produkt mit hoher Wahrscheinlichkeit retournieren werden. Der Graph modelliert das Kauf- und Retourenverhalten von Kunden und setzt sich aus den zwei Knotenarten – Kunden und Produkte zusammen. Zwischen den Kunden- und Produktknoten verlaufen unterschiedliche Kanten, die verschiedene Sachverhalte darstellen sollen. So stellen gerichtete Kanten, historische Käufe ohne Retouren, und ungerichtete Kanten, historische Käufe mit Retouren dar. Innerhalb der Produkt- bzw. Kundenknoten symbolisieren ungerichtete Kanten Ähnlichkeiten zwischen Kunden bzw. Produkten. Das System wird unter der Annahme aufgesetzt, dass Kunden, die bereits Retouren eingeleitet haben, viel eher ein ähnliches Produkt retournieren werden als Kunden, die das Produkt nicht zurückgeschickt haben.

Für die Identifizierung der entsprechenden Kundengruppe wird LoGraph, ein auf einen Random-Walk basierender Algorithmus, eingesetzt. Dieser ermittelt die Neigung eines Kunden ein bestimmtes Produkt zu retournieren. Als Eingabeparameter benötigt der Algorithmus einen hybriden Graphen, den zu untersuchenden Produktknoten, die maximale Anzahl der Kunden in dem lokalen Cluster, die maximale Anzahl an Kanten mit denen der lokale Cluster mit dem Rest des Graphen verbunden ist, die Größe des zurückgegebenen

Clusters, die Laufzeit und die stationäre Verteilung. Der Random-Walk beginnt bei dem Produktknoten, dessen Retouren ermittelt werden soll und generiert eine Sequenz von Vektoren. Zu jedem neuen Zeitpunkt geht der Algorithmus einen Schritt weiter. Dabei werden die Sets mit den größten Übereinstimmungen ermittelt und innerhalb dieses Sets wird geprüft, ob es die gesuchten Kunden enthält. Der Algorithmus findet Cluster in der Nähe des Eingabeknotens, in dem es sich nur einen kleinen Bereich in unmittelbarer Umgebung des Ausgangsknotens ansieht. Die Laufzeit des Algorithmus hängt von der Größe des Output Clusters ab und nicht von der Größe des gesamten Graphen, so dass es für die Verarbeitung großer Daten einsetzbar ist.

Daten von drei europäischen Omnichannel Händlern werden genutzt, um die Performanz des Systems zu validieren. Die erhaltenen Daten werden in Trainings-, Validierungs- und Testdaten unterteilt. Die maximale Anzahl an Kundenknoten wird auf 1000 gesetzt und nur die Produkte werden berücksichtigt, die auch in den Trainingsdaten enthalten sind, da sie sonst nicht getestet werden können.

Die Performanz des Algorithmus wird mit Baseline und ADNI verglichen. Ersteres stellt Prognosen ausschließlich unter Berücksichtigung der Ähnlichkeit von Produkten auf. ADNI inkludiert die Neigung eines Kunden für Retouren mittels bipartiter Graphen. Für den Vergleich werden die Metriken Precision und Recall hinzugezogen, wobei Precision höher gewichtet wird als Recall. LoGraph erzielt mit einer Precision von 82 %, 86 % bzw. 75 % die besten Resultate, da sich der Algorithmus neben den Produktähnlichkeiten auch die Informationen aus den historischen Bestell- und Retourendaten zu Nutze macht. Der LoGraph Algorithmus eignet sich in Kombination mit einem hybriden Graphen für die Prognose von Retouren.

3.4 Causal Machine Learning

Der nachfolgende Artikel befasst sich mit Causal Machine Learning, ein Themengebiet, in dem die Modelle kausale Zusammenhänge betrachten. Kausalität beschreibt, wie ein Prozess oder ein Ereignis zu dem Entstehen eines anderen Prozesses bzw. Ereignisses führt (Gonfalonieri, 2020).

Von Zahn et al. (2022) erarbeiten in ihrem Artikel ein System zur Reduzierung von Retouren. Für die Aufstellung des Systems kombinieren die Autoren einen so genannten Green Nudge mit einer Data Enrichment Strategie und Causal Machine Learning. Ziel ist es, ausgewählte Kunden u.a. noch vor Abschluss der Bestellung technischen Interventionen in Form von beispielsweise Pop-Up Fenstern auszusetzen, mittels derer sie über die ökologischen Auswirkungen von Retouren aufmerksam gemacht werden. Diese Green Nudges sollen den Kunden animieren seine bzw. ihre Bestellung so aufzugeben, dass im Nachhinein keine Artikel retourniert werden müssen.

Die generelle Wirksamkeit von Green Nudges wird zunächst anhand von A/B Tests in einem sieben-wöchigen Feldexperiment überprüft und validiert. Während grundsätzlich eine Reduzierung der Retouren von 3,8 % und eine Gewinnmaximierung von 8,7 % beobachtet werden konnte, konnte das gewünschte Kundenverhalten nicht über alle Kundengruppen hinweg observiert werden. Zur Vermeidung von negativen Auswirkungen des Green Nudges aufgrund der Vielschichtigkeit der Kunden wird für die Ermittlung von zielführenden Kundengruppen (umweltbewusst, viele Retouren) ein Causal Forest eingesetzt.

Der Causal Forest wird mit einem so genannten angereicherten digitalen Fußabdruck trainiert. Hierzu werden die aus dem Experiment gewonnenen Daten, wie z.B. technische Informationen (Browser, IP-Standort), Warenkorbzusammensetzung, Einkaufsverhalten und Retourenhistorie mit öffentlich zugänglichen Daten fusioniert. Die öffentlich zugänglichen Daten beinhalten Informationen über den sozio-ökonomischen Status (z.B. durchschnittliches Haushaltseinkommen), Einstellungen zur Umwelt (u.a. Anteil an Umweltaktivisten) und ökologisches Verhalten in bestimmten Regionen. Demographische Informationen wurden aus Datenschutzgründen nicht berücksichtigt.

Der trainierte Causal Forest stellt Prognosen hinsichtlich des konditionalen durchschnittlichen Behandlungseffekts (CATE) auf. In anderen Worten es quantifiziert, wie sich der Gewinn, der durch einen Kunden erwirtschaftet wurde, ändert, wenn dieser einem Green Nudge ausgesetzt wird. So soll der Green Nudge nur den Kunden angezeigt werden, die einen positiven CATE-Wert aufweisen. Dieses System führt zu einer Reduzierung der Retouren von 6,7 % sowie einer Gewinnsteigerung von 12,5 %, da nur umweltbewusste Kunden angesprochen werden, die durch die technischen Interventionen ihre Bestellungen so aufgeben, dass keine Ware retourniert werden muss. Durch die bewusste Selektierung der

anzusprechenden Kunden, kommen die Kunden, bei denen der Green Nudge zu einem Abbruch der Bestellung führen könnte, gar nicht erst in Berührung mit dem Green Nudge. So werden dem Händler Gewinneinbußen durch den Green Nudge erspart.

3.5 Rule-Based Machine Learning

Der letzte aufgeführte Artikel beschäftigt sich mit Rule-Based Machine Learning. Regeln helfen dem Modell entsprechende Entscheidungen zu fällen. Das Rule-Based Machine Learning unterscheidet bei der Aufstellung der Regeln zwischen zwei möglichen Vorgehensweisen, „divide and conquer“ bzw. „separate and conquer“. Erstgenannter Ansatz erzeugt Regeln in Form eines Entscheidungsbaumes, in dem es die Trainingsdaten in immer kleinere Subsets aufteilt (Lantz, 2019, S. 127). „Separate and conquer“ hingegen generiert wenn-dann Regeln. Hierbei stellt das Modell zunächst eine Regel auf, die einen Teil der Trainingsdaten abdeckt und lernt rekursiv weitere Regeln bis alle Trainingsdaten durch die aufgestellten Regeln abgedeckt sind (Fürnkranz, 1999, S. 3).

Seewald et al. (2019) untersuchen in ihrem Artikel den Einsatz von Rule-Based Machine Learning für die Klassifikation von Retouren. Dabei fundieren die Autoren ihre Analysen auf Daten eines österreichischen Unternehmens.

Für die Klassifikation von Retouren setzen die Autoren JRip, einen Rule Learning Algorithmus ein, der präzise und einfach auswertbare Regel Sets generiert. Es handelt sich hierbei um eine auf Java basierte Implementierung von RIPPER. Für das Training des Modells werden 20 % der verfügbaren Daten verwendet, die mit den Labels Retoure und keine Retoure versehen werden. Die restlichen 80 % sowie die Daten, die in den Trainingsdaten als keine Retoure eingestuft wurden, werden für die Evaluation des Modells eingesetzt. Um die Regeln so akkurat wie möglich zu halten, werden nur Daten aus den Jahren 2017 und 2018 verwendet. Im Zuge der Datenvorbereitung werden sehr prädiktive Eigenschaften aus den Daten entfernt. Eine zweifache Kreuzvalidierung wird für die erste Evaluierung der Trainingsdaten vorgenommen.

JRip generiert elf Regeln, die Retouren mit einer Precision von 49,5 %, einem Recall von 78,4 % und einem F-Score von 0,607 klassifizieren. Die von den Regeln erkannten Muster

beziehen meist den Zeitpunkt, ab dem das Produkt zum Erwerb angeboten wird, einen Warennummernbereich, um bestimmte Warengruppen auszuschließen sowie einen Startpreis ein. Bei den Auftragsdaten ist die Kombination aus Zustelltag und Versandoption ein Indiz für mögliche Retouren. So besagt eine der Regeln, dass wenn eine an einem Freitag mit Express Versand bestellte Ware erst an einem Dienstag zugestellt wird, diese mit hoher Wahrscheinlichkeit (85,3 %) retourniert wird.

Vier der elf Regeln verwenden nur Produktdaten, fünf verwenden neben den Produktdaten auch Auftragsdaten, eine lässt auch anonymisierte Kundendaten mit einfließen. Die Produktdaten bilden dabei einen wichtigen Bereich für die Regeln. Zur Erzielung von akkuraten Resultaten müssen die Regeln in einer bestimmten Reihenfolge angewandt werden.

4 Fazit

Bekanntermaßen führen viele Wege nach Rom – dies gilt auch für den Untersuchungsgegenstand der vorliegenden Arbeit. Die hier aufgeführten Studien setzen jeweils unterschiedliche Methoden und Algorithmen ein, um Prognosen im Bereich von Retouren aufzustellen. Auch wenn sich die Arbeiten minimal in der Zielsetzung unterscheiden, wie z.B. die Prognose von Retouren basierend auf Warenkorbkomposition oder die Prognose von Retourenmissbrauch, ist das übergeordnete Ziel doch das Gleiche, nämlich die Minimierung von Retouren.

Nachfolgend werden die untersuchten Methoden verglichen. Hierzu werden die eingesetzten Methoden zusammen mit den im Machine Learning zentralen Metriken tabellarisch zusammengefasst. Die Empfehlungen der Autoren sind hervorgehoben. Es sei an dieser Stelle vermerkt, dass nicht alle Studien Werte zu allen Metriken liefern, denn nicht alle Metriken eignen sich für die Interpretation der Ergebnisse. In solchen Fällen werden die entsprechenden Stellen in der Tabelle mit blanken Stellen versehen. Werte, die mit einem Sternchen gekennzeichnet sind, wurden anhand der vorhandenen Werte errechnet. Die nicht aufgeführten Studien liefern zu den Metriken keinerlei Werte und werden somit nicht aufgelistet. Kurze Erläuterungen zu den genutzten Metriken finden sich im Glossar wieder.

Autor	Algorithmus	Precision	Recall	F1-Score	TAE	AUC	Accuracy
Ketzenberg, et al., 2020	Neuronales Netzwerk	-	-	-	-	0,993	-
	Random Forest	-	-	-	-	0,992	-
	Logistische Regression	-	-	-	-	0,979	-
	Support Vector Machines	-	-	-	-	0,972	-
Cosar Soguk-kuyu, et al., 2022	LightGBM	0,74	0,7	0,72	-	-	-
	Decision Tree	0,72	0,69	0,7	-	-	-
	Random Tree	0,71	0,67	0,69	-	-	-
Al Imran, et al., 2020	TabNet	0,619	0,831*	0,71*	-	0,67	0,66
	XGBoost	0,637	0,736*	0,683*	-	0,658	0,658
	LightGBM	0,655	0,723*	0,687*	-	0,671	0,671
	CatBoost	0,604	0,72*	0,657*	-	0,624	0,624
	Decision Tree	0,626	0,72*	0,67*	-	0,645	0,645
Asdecker, et al., 2018	Ensemble Machine Learning	-	-	-	0,315	-	-
	C5.0	-	-	-	0,316	-	-
	Künstliches Neuronales Netzwerk	-	-	-	0,324	-	-
	Binäre logistische Regression	-	-	-	0,332	-	-
	Lineare Diskriminanzanalyse	-	-	-	0,334	-	-
Dzyabura, et al., 2018	RGB + Deep-learned + Gabor	-	-	-	-	-	0,4638
	RGB + Deep-learned	-	-	-	-	-	0,4626
	Gabor + Deep-learned	-	-	-	-	-	0,4532
	RGB + Gabor	-	-	-	-	-	0,4474
	Deep-learned	-	-	-	-	-	0,4464
	RGB	-	-	-	-	-	0,4412
	Gabor	-	-	-	-	-	0,4216
Li, et al., 2018	HyperGo	0,69	0,55	0,612*	-	0,67	-
	JacNorm	0,64	0,53	0,58*	-	0,64	-
	JacWght	0,56	0,77	0,65*	-	0,56	-
	k-NN	0,55	0,78	0,65*	-	0,55	-
Zhu, et al., 2018	LoGraph	0,868	0,285	0,429*	-	-	-
	ADNI	0,733	0,162	0,265*	-	-	-
	Baseline	0,7	0,212	0,325*	-	-	-
Seewald, et al., 2019	JRip	0,495	0,784	0,607	-	-	-
Urbanke, et al., 2015	Mahalanobis + AdaBoost	0,848	0,141	0,243*	-	-	-
	Mahalanobis + Random Forest	0,844	0,144	0,247*	-	-	-
	Mahalanobis + Extremely Randomized Trees	0,843	0,144	0,245*	-	-	-
	Mahalanobis + CART	0,823	0,142	0,242*	-	-	-
	Mahalanobis + logistische Regression	0,820	0,142	0,241*	-	-	-
	Mahalanobis + lineare Diskriminanzanalyse	0,820	0,142	0,241*	-	-	-
	Mahalanobis + Gradient Boosting	0,818	0,143	0,243*	-	-	-

Tabelle 2 Vergleich der eingesetzten Algorithmen und Methoden

Ein eins-zu-eins Vergleich lässt sich aufgrund der vielen abweichenden Faktoren nicht vornehmen. Denn Faktoren wie Branche, Unternehmen, Größe und Alter der Daten, Kundenverhalten, Produkte, etc. haben einen großen Einfluss auf die Performanz der eingesetzten Algorithmen. Vielmehr kann die vorliegende Arbeit versuchen einen groben

Vergleich vorzunehmen. Hierzu werden die meisten der in den Studien eingesetzten Algorithmen basierend auf ihrer Natur gebündelt und nicht nach Studie. Diese werden untereinander gegenübergestellt. Eine Empfehlung kann allerdings nicht ausgesprochen werden, da die Gegebenheiten der jeweiligen Studie die Auswahl der Algorithmen stark beeinflussen. Die aufgeführten Konzepte werden nur sehr grob angeschnitten, da eine tiefgehende Analyse und Beschreibung den Rahmen dieser Arbeit sprengen würden.

4.1 Lineare Methoden

Die logistische Regression zählt zu den linearen Methoden. Aufgrund ihrer Einsatzgebiete in Regressions- und Klassifikationsproblemen wird sie am meisten eingesetzt (Joshi, 2023, S. 51).

Drei der untersuchten Studien setzen die logistische Regression ein. Um die statistisch nicht benötigten Variablen schrittweise zu entfernen, kann, wie bei Ketzenberg et al. (2020) eine Rückwärtselimination eingesetzt werden (S. 642). Dies ist insbesondere in Situationen mit Multikollinearität nötig, d.h. wenn mehrere Prädiktoren untereinander stark korrelieren (Statistik Guru, o. J.). In allen drei Studien erreicht die logistische Regression sehenswerte Ergebnisse, in Urbanke et al. (2015, S. 12) und in Asdecker et al. (2018, S. 44) liegt sie etwas weniger als zwei Prozentpunkte unter der performantesten Methode. Die Autoren zeigen somit auf, dass nicht immer die komplexesten und rechenintensivsten Methoden zum Einsatz kommen müssen und dass auch weniger komplexe Methoden den Unternehmen einen großen Mehrwert bringen können.

4.2 Neuronale Netzwerke

Künstliche neuronale Netzwerke entstanden in Anlehnung an den Aufbau und die Funktion des menschlichen Gehirns. Den wichtigsten Bestandteil bilden dabei die Neuronen, die Signale von anderen Neuronen erhalten, diese bearbeiten und anschließend im Netzwerk weiterleiten (Joshi, 2023, S. 57), siehe Abbildung 3.

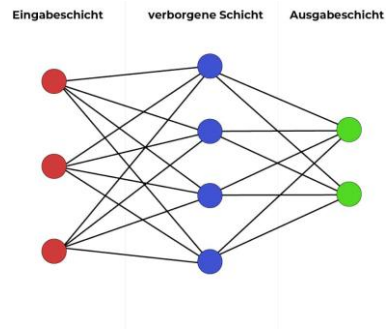


Abbildung 3 Aufbau eines neuronalen Netzwerks (Wuttke, o. J.)

Während in Ketzenberg et al. (2020) künstliche neuronale Netzwerke in ihrer Basisform vorkommen, setzen Asdecker et al. (2018) ein vorwärtsgerichtetes neuronales Netzwerk, den so genannten Multilayer Perceptron ein (S. 42). Die Justierung neuronaler Netzwerke erfolgt über die Anzahl der versteckten Schichten und die Anzahl der Knoten in jeder Schicht. Da diese einen großen Einfluss auf die Performanz des Systems haben, müssen sie sorgfältig ermittelt werden. In Ketzenberg et al. (2020) erzielt das künstliche neuronale Netzwerk mit acht Neuronen und einer versteckten Schicht eine AUC von 0,993 und wird daher für den Einsatz zur Prognose von Retourenmissbrauch empfohlen (S. 642). Ein ähnliches Setup, mit sechs Neuronen und einer versteckten Schicht erzielt in Asdecker et al. (2018) einen TAE von 33,37 % (S. 44). Aufgrund der nur gemäßigten Resultate nehmen Asdecker et al. (2018) den Multilayer Perceptron nicht in die engere Auswahl.

4.3 Deep Learning

Im Themenbereich des Deep Learning kommen Convolutional Neural Networks und TabNet auf, die im Folgenden kurz aufgegriffen werden.

Ein Convolutional Neural Network (CNN) ist ein Deep Learning Algorithmus, welches besonders in der Bildverarbeitung eingesetzt wird (Joshi, 2023, S. 157). Eine etwas detaillierte Erläuterung kann dem Glossar entnommen werden. Die Abläufe innerhalb eines CNNs werden in Abbildung 4 grob dargestellt.

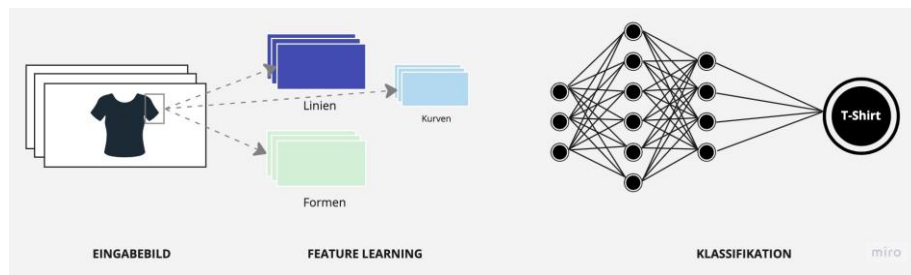


Abbildung 4 Convolutional Neural Network (CNN)

Dzyabura et al. (2018) setzen CNNs in einen etwas anderen Teilbereich des Retourenmanagements ein. Sie erarbeiten ein System, welches die Retourenrate eines bestimmten Produktes noch vor dessen Launch ermittelt. Hierzu setzen die Autoren Produktfotos ein. Für das Extrahieren von Attributen werden bereits vortrainierte CNNs eingesetzt, da die zur Verfügung stehende Datenmenge nicht ausreichte, um das CNN selbst zu trainieren (Dzyabura, et al., 2018, S. 16). Durch die Hinzunahme von Produktfotos verbessert sich die Accuracy ihres Modells um ca. 30 % (Dzyabura, et al., 2018, S. 19).

TabNet ist ein Deep Learning Netzwerk, welches neuronale Netze auf tabellarischen Daten anwendet (Al Imran, et al., 2020, S. 1175) und sich dabei auf die wichtigsten Features fokussiert (Llango, 2021). In Al Imran et al. (2020) ermittelt das Modell die beiden Features „order_location“ und „payment_method“ (S. 1178) als am aussagekräftigsten, da diese bereits ein Indiz für die Retourenwahrscheinlichkeit sind. So weisen 98 % der Retouren eine Zahlungskondition auf, bei der die Kunden nicht in Vorkasse treten müssen. In Al Imran et al. (2020) übertrifft das TabNet alle anderen untersuchten Algorithmen (S. 1176) und zeigt, dass Unternehmen durch den Einsatz von TabNet die für ihren Einsatzgebiet aussagekräftigsten Attribute erarbeiten können.

4.4 Entscheidungsbäume

Entscheidungsbäume sind baumartig aufgebaut und bestehen aus Knoten, Ästen und Blättern. Ausgehend vom Wurzelknoten führen Äste, die verschiedene Entscheidungen darstellen, zu weiteren Ebenen und enden in den Blättern, den Ergebnissen (Cosar Sogukkuyu, et al., 2022, S. 193). Anhand dem in Abbildung 5 dargestellten Entscheidungsbaum können beispielsweise Retourenanfragen angenommen bzw. abgelehnt werden.

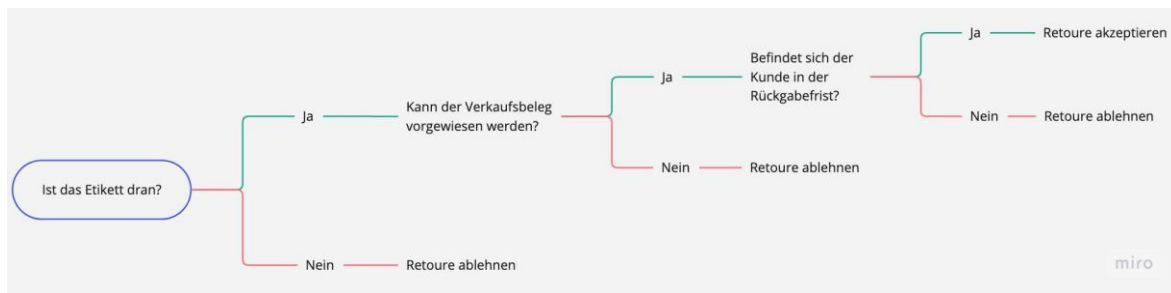


Abbildung 5 Entscheidungsbaum

Die Entscheidungsbäume in Cosar Sogukkuyu et al. (2022) erzielen nur moderate Ergebnisse (S. 195) und auch in Al Imran et al. (2020, S. 1177) überzeugt die Performanz der Entscheidungsbäume nicht. Vielmehr scheinen die auf Entscheidungsbäumen basierenden Algorithmen, die nachfolgend kurz aufgegriffen werden, besser abzuschneiden als die Entscheidungsbäume in ihrer Basisform.

4.4.1 Ensemble Machine Learning

Ensemble Machine Learning kombiniert verschiedene Lerner (Modelle), um die Performanz des Gesamtsystems zu verbessern (Cosar Sogukkuyu, et al., 2022, S. 193). Asdecker et al. (2018) kombinieren den C5.0 Algorithmus, CHAID und QUEST und erreichen in ihrer Studie korrekte Prognosen von 68,45 % (S. 44). Die aufgeführten Algorithmen werden für die Generierung von Entscheidungsbäumen eingesetzt. Kurze Definitionen zu den Algorithmen werden im Glossar aufgeführt. Ensemble Machine Learning ist aufgrund seiner Arbeitsweise in der Lage, Prognosen mit einer hohen Precision aufzustellen, da es die Stärken der einzelnen Modelle gezielt kombiniert, um ein performantes Gesamtsystem aufzustellen.

4.4.2 Gradient Boosting

Das Gradient Boosting fügt sequenziell schwache Lerner hinzu, um den Fehler in der Prognose immer weiter zu minimieren (Joshi, 2023, S. 86). In Hofmann et al. (2020) erzielt AdaBoost, welches zu den Gradient Boosting Verfahren gehört das beste Resultat (S. 7). Gradient Boosting im Zusammenspiel mit Mahalanobis Feature Extraction schneidet in Urbanke et al. (2015) allerdings am schlechtesten ab.

LightGBM, XGBoost und CatBoost sind weitere Gradient Boosting Modelle, die in den untersuchten Studien zum Einsatz kommen. Da XGBoost zusammen mit Regressionsbäumen behandelt wird, soll es an dieser Stelle nicht aufgegriffen werden.

LightGBM Bäume wachsen vertikal, was zu einem größeren Abbau von Verlusten und somit zu einer besseren Accuracy führt (Saha, 2023). Im Vergleich mit normalen Entscheidungsbäumen und einem Random Forest, erzielt LightGBM in der Studie von Cosar Sogukkuyu et al. (2022) das beste Resultat. 74 % der Fälle werden korrekt prognostiziert (Cosar Sogukkuyu, et al., 2022, S. 195). In Al Imran et al. (2020) erzielt LightGBM eine Precision von nur 65,5 % (S. 1176).

4.4.3 Random Forest / Tree

Ein Random Forest besteht aus einer Reihe von Entscheidungsbäumen. Dieser betrachtet die Ergebnisse der einzelnen Entscheidungsbäume und kombiniert sie zu einem Gesamtergebnis (Cosar Sogukkuyu, et al., 2022, S. 193).

In vier der betrachteten Studien kommt der Random Forest zum Einsatz. Die Ergebnisse unterscheiden sich sehr stark. Während der Random Tree in Cosar Sogukkuyu et al. (2022, S. 195) und der Random Forest in Hofmann et al. (2020, S. 8) im Vergleich mit den anderen untersuchten Methoden am schlechtesten abschneiden, erreicht es in Urbanke et al. (2015, S. 12) und in Ketzenberg et al. (2020, S. 642) die zweitbesten Resultate.

Ketzenberg et al. (2020) sind die einzigen Autoren, die auf die Justierung des Random Forests eingehen. Hierzu observieren sie verschiedene Werte für die Anzahl der zufällig ausgewählten Features, sowie die Anzahl der Bäume, aus denen der Random Forest zusammengestellt wird. In diesem Fall wird die Performanz der Justierung anhand der Accuracy gemessen und ergibt, dass fünf Features und 1000 Bäume die beste Performanz aufweisen (Ketzenberg, et al., 2020, S. 642). Dies sollte jedoch mit Vorsicht genossen werden, da die Autoren in ihrer Studie eine Klassen-Imbalance haben und diese nicht korrigieren. Die Metrik Accuracy sollte erst in den Fällen betrachtet werden, in denen die untersuchten Klassen ausgewogen sind (Oppermann, 2021). Im Vergleich mit den anderen eingesetzten Methoden wird AUC hinzugezogen und der Random Forest erzielt mit einer AUC von 0,992 die zweitbesten Resultate (Ketzenberg, et al., 2020, S. 642).

4.4.4 Regressionsbäume

Ein Regressionsbaum teilt die verfügbaren Daten in mehrere Bereiche, die anhand von bestimmten Variablen gebündelt werden. Beim Boosting werden schrittweise mehrere einfache Modelle zusammengetragen, die mit dem Residuum des aktuellen Modells trainiert werden (Dzyabura, et al., 2018, S. 16 f.).

Dzyabura et al. (2018) setzen in ihrer Studie XGBoost, ein Gradient Boosted Regressionsbaum Modell ein, um Prognosen zur Retourenrate aufzustellen. Das zusammen mit einem CNN eingesetzte Modell erreicht eine Accuracy von 46,38 % (Dzyabura, et al., 2018, S. 19). Mit einer im untersuchten Zeitraum vorliegenden Retourenrate von 53 % (Dzyabura, et al., 2018, S. 8) sind die Daten größtenteils ausgewogen, so dass die Accuracy eingesetzt werden kann, um die Performanz des Modells akkurat zu deuten. 46,38 % der Prognosen des Systems sind korrekt, mehr als die Hälfte der Prognosen stimmen jedoch nicht. Hier kommt die Frage auf, ob Produktfotos zur Genüge kenntlich machen können, ob ein Produkt retourniert werden wird oder nicht. Oftmals tragen neben der Meinungsänderung des Kunden aufgrund einer impulsiven Bestellung (Pei, et al., 2018, S. 301), einer nachträglichen kognitiven Dissonanz (post-purchase dissonance) (Asdecker, et al., 2017, S. 4193) auch die Diskrepanz zwischen den Erwartungen des Kunden und dem erhaltenen Produkt (Aussehen, Qualität, etc.) (Sun et al., 2021) zu einer Retoure bei.

4.5 Support Vector Machines

Support Vector Machine Algorithmen werden für die Lösung binärer Klassifikationsprobleme eingesetzt (Joshi, 2023, S. 89). Der Algorithmus ermittelt eine Trennebene in einem n-dimensionalen Raum mit dem maximal möglichen Abstand, welcher die Datenpunkte eindeutig klassifiziert (Gandhi, 2018).

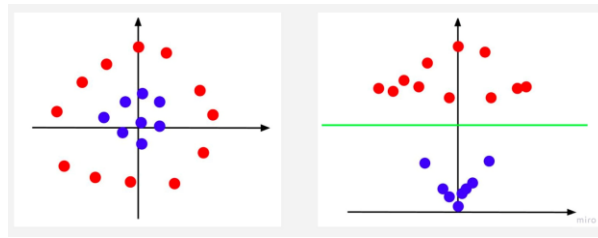


Abbildung 6 Support Vector Machine (Kafesu, 2023)

Abbildung 6 (links) zeigt eine mögliche Darstellung nicht-linearer Datenpunkte in einem n -dimensionalen Raum, wobei die blauen Datenpunkte Retouren und die roten Datenpunkte keine Retouren darstellen. In diesem Zustand ist es für die Support Vector Machine schwierig, einen neuen Datenpunkt eindeutig zu klassifizieren, da es nicht ohne Weiteres die Trennebene mit dem maximal möglichen Abstand zwischen den Klassen ermitteln kann. Hier kommt die Kernel-Funktion zum Tragen, die den Zustand in der linken Abbildung in einen höherdimensionalen Raum überführt, in der sich, wie in Abbildung 6 (rechts) zu sehen, eine Trennebene zwischen den betrachteten Klassen eindeutiger platzieren lässt (Kafesu, 2023).

Während Urbanke et al. (2015) die Support Vector Machines für die Prognose von Retouren einsetzen, nutzen sie Ketzenberg et al. (2020) für die Prognose und Verhinderung von Retourenmissbrauch. Urbanke et al. (2015) verwenden den Korrelationskoeffizienten, auch „Person’s r “ genannt, für die Evaluation der Performanz der Algorithmen. In Verbindung mit der Mahalanobis Feature Extraction erreicht der Algorithmus einen Korrelationskoeffizienten von 0,3 (Urbanke, et al., 2015, S. 10) also einen mittleren Zusammenhang zwischen den Prognosen und den eigentlichen Labels. Die Support Vector Machines mit einem linearen Kernel erzielen in der Studie von Ketzenberg et al. (2020, S. 642) einen AUC-Wert von 0,972, was die präzise Genauigkeit des Modells bestätigt. Sowohl Urbanke et al. (2015) als auch Ketzenberg et al. (2020, S. 642) stellen fest, dass der mit einem linearen Kernel justierte Algorithmus am performantesten ist und auch für große Datensätze gut einsetzbar ist.

4.6 Rule-Based Machine Learning

Rule-Based Machine Learning Algorithmen setzen Regeln auf, anhand derer u.a. die Klassifikation erfolgen kann. Das von Seewald et al. (2019) eingesetzte Rule-Based Machine Learning generiert für die Klassifikation von Retouren aus den Trainingsdaten wenn-dann

Regeln. Anhand dieser Regeln kann das System weitere Datensätze entsprechend den Labels „Retoure“ bzw. „keine Retoure“ zuordnen. 78,4 % der Instanzen konnten von JRip korrekt identifiziert werden (Seewald, et al., 2019, S. 806), wobei der etwas niedrige Precision Wert auf eine hohe Anzahl an falsch Positiven Zuordnungen hindeutet. In diesem Fall lag JRip bei einer Einstufung als „Retoure“ in nur 49,5 % der Fälle korrekt (Seewald, et al., 2019, S. 806). Gerade im Bereich von Retourenprognosen sollte ein hoher Precision Wert angestrebt werden. Die Autoren nutzen für ihre Analysen lediglich Daten von 24 Monaten (Seewald, et al., 2019, S. 806), wobei eine ungewöhnliche Aufteilung der Daten vorgenommen wird. Anstatt der üblichen 80-20 Aufteilung in Trainings- und Test- bzw. Validierungsdaten, wird eine 20-80 Aufteilung vorgenommen (Seewald, et al., 2019, S. 805). Da die von JRip aufgestellten Regeln alle Trainingsdaten abdecken, kann ein Grund für die niedrige Precision diese ungewöhnliche Aufteilung sein. Dem Algorithmus stehen viel weniger Daten für das Training zur Verfügung, so dass die aufgestellten Regeln für den Rest der Daten nicht sehr repräsentativ sind. Aufgrund der gemäßigten Performanz des Modells sollten weitere Experimente durchgeführt werden, um festzustellen, ob sich durch andere Daten und eine abweichende Vorbereitung die Resultate ändern. Somit kann ermittelt werden, inwieweit Rule-Based Machine Learning für die Prognose von Retouren eingesetzt werden kann.

4.7 Causal Machine Learning

Einen sehr spannenden Bereich untersuchen von Zahn et al. (2022) in ihrer Studie. Die Autoren setzen in ihren Analysen Causal Machine Learning gekoppelt mit einem Smart Green Nudge ein. Causal Machine Learning eignet sich, um die Auswirkungen des Green Nudges zu erfassen. Der verwendete Causal Forest erkennt Heterogenitäten im Behandlungseffekt und ermittelt so Kundengruppen, bei denen die eingesetzten minimal-invasiven Interventionen (Green Nudge) zum gewünschten Ergebnis führen und bei welchen Kunden es einen gegensätzlichen Effekt hat (von Zahn, et al., 2022, S. 17). So können diejenigen Kunden erfasst werden, die aufgrund ihrer Bestellhistorie als Retoureninitiiierer auffallen, die aber zudem eine gewisse umweltbewusste Disposition haben. Dieses System führt zu einer Reduzierung der Retourenrate um 6,7 % (von Zahn, et al., 2022, S. 21), welches laut den Autoren gleichzusetzen ist mit einer jährlichen Reduzierung des CO₂

Ausstoßes um 315 Tausend Tonnen. Gleichzeitig führt es zu einer Gewinnsteigerung von 12,5 % (von Zahn, et al., 2022, S. 22). Die Studie zeigt, dass nicht alle Kundengruppen gleich behandelt werden dürfen und gewisse Abgrenzungen nötig sind. Andernfalls kann es dem Unternehmen sogar schaden.

4.8 Graphen

Für die Prognose von Retouren setzen zwei der Studien Graphen ein. Ein hybrider Graph kommt in Zhu et al. (2018, S. 3718) zum Einsatz. Diese können als eine Art Erweiterung konventioneller Graphen aufgefasst werden, da Knoten nicht nur Entitäten, sondern auch Beziehungen modellieren können (siehe Abbildung 7). Hypergraphen hingegen, die von Li et al. (2018) verwendet werden, weisen Kanten auf, die beliebig viele Knoten miteinander verbinden können. Abbildung 8 stellt einen Hypergraphen exemplarisch dar.

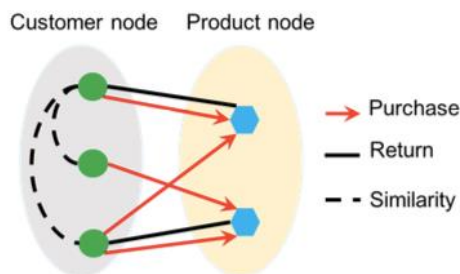


Abbildung 7 Hybrider Graph (Zhu et al., 2018)

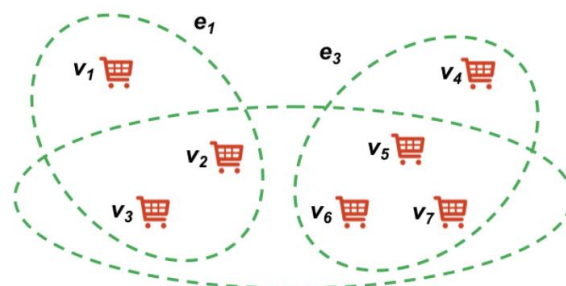


Abbildung 8 Hypergraph (Li et al., 2018)

So sind beide Graphen Arten in der Lage komplexe Beziehungen zwischen mehreren Entitäten zu modellieren (Li, et al., 2018, S. 520). Diese Besonderheit erschwert aber die Verarbeitung im Vergleich mit konventionellen Graphen, so dass die Autoren die Effizienz ihrer Modelle über die Laufzeit evaluieren. Die aus der Laufzeit gewonnenen Erkenntnisse verdeutlichen, dass die erarbeiteten Modelle für die Verarbeitung großer Datenmengen einsetzbar sind.

Für die Ermittlung der Warenkörbe bzw. Produkte werden in beiden Studien auf Random-Walk basierende Algorithmen eingesetzt, die jeweils die unmittelbare Umgebung des

Ausgangsknoten betrachten (Zhu, et al., 2018, S. 3720) (Li, et al., 2018, S. 521). In Li et al. (2018) erreicht das Modell eine Precision von 81 % (S. 525) während in Zhu et al. (2018) der Algorithmus einen Precision Wert von 82 % (S. 3723) erzielt. Durch die spezielle Anordnung der Daten im Graphen und die Untersuchung der unmittelbaren Umgebung kann das Modell schnell eine Prognose mit einer hohen Precision aufstellen. Dies ist insbesondere in real-time Situationen ein großer Vorteil, da die Prognose während des Bestellvorgangs innerhalb kürzester Zeit erfolgen muss, um z.B. das Käuferlebnis des Kunden nicht negativ zu beeinflussen.

Die in diesem Kapitel aufgeführten Erläuterungen machen deutlich, dass zur Aufstellung von Prognosen zu Kundenretouren unterschiedliche Methoden eingesetzt werden können. Hierbei müssen nicht immer komplexe Methoden zum Einsatz kommen, denn die logistische Regression eignet sich besonders gut für binäre Klassifikationsprobleme. Weitere Methoden sind Ensemble Machine Learning Methoden (z.B. Random Forest, Gradient Boosting), die sequenziell weitere Entscheidungsbäume hinzufügen, um ein performantes Modell zu erstellen. Neuronale Netzwerke und Support Vector Machines sind ebenfalls mächtige Werkzeuge, wenn es um Klassifikationsprobleme geht.

5 Ausblick

Retouren sind ein oftmals unerwünschtes Nebenprodukt des boomenden Online-Handels. Sie haben nicht nur Auswirkungen auf die Wirtschaftlichkeit eines Unternehmens, denn gerade die Bearbeitung von Retouren treibt die Kosten für die Unternehmen in die Höhe, sondern auch auf die Umwelt. Strategien für die Minimierung von Retouren sind daher unerlässlich. Der sich stetig weiterentwickelnde Bereich des Machine Learning bietet den Unternehmen viele Möglichkeiten, die vollends ausgeschöpft werden sollten.

Spannende neue Entwicklungen bzw. Abzweigung des Machine Learning, wie zum Beispiel Causal Machine Learning oder Automated Machine Learning können, wenn gezielt eingesetzt, den Unternehmen einen großen Mehrwert bringen. Insbesondere das Causal Machine Learning eignet sich besonders gut, um in der Wirtschaft die richtigen Unternehmensentscheidungen zu treffen. Denn oftmals kommt es auf die kausalen Zusammenhänge an,

die einen Sachverhalt besser erklären (Eichenberger, 2020). Im Rahmen dieser wissenschaftlichen Vertiefung konnte nur eine Studie ermittelt werden, die sich mit Causal Machine Learning befasst. Die Möglichkeiten können mit Automated Machine Learning noch ein Stück vergrößert werden, denn durch AutoML lässt sich der Machine Learning Prozess automatisieren. Die hier zum Einsatz kommenden Algorithmen können u.a. die besten Modelle und Hyperparameter für ein bestimmtes Problem erarbeiten und einsetzen. Aufgrund der Vielschichtigkeit der zur Verfügung stehenden Daten könnte AutoML einen besonders großen Einfluss auf die Entwicklung zukünftiger Modelle haben.

Die in dieser Arbeit untersuchten Studien haben größtenteils einen theoretischen Charakter und die Implementierung in Unternehmen sowie die daraus resultierenden Ergebnisse kommen dabei zu kurz. Hier sollten in Zukunft mehr A/B Experimente durchgeführt werden, um die Prognosen in real-life Situationen zusammen mit den möglichen Demarketing Strategien zu testen. Die Resultate der Studien sind vielversprechend, diese müssten in einem nächsten Schritt in einzelnen Unternehmen implementiert und analysiert werden. Insbesondere die zum Einsatz kommenden Demarketing Strategien müssen durch A/B Tests evaluiert werden. Die Notwendigkeit einer Evaluation zeigt die Studie von Zahn et al. (2022). Urbanke et al. (2015) unterscheiden in ihrer Studie zwischen milden Interventionen (Überzeugung auf moralischer Ebene), harten Interventionen (Abbruch der Transaktion) und einer Kombination aus beiden (Eingrenzung der Zahlungsoptionen) (S. 15). Basierend auf dem Produkt sowie der Branche greifen die vorgeschlagenen Strategien jedoch nicht bei allen Kundengruppen. So stellen von Zahn et al. (2022) in ihrem Feldexperiment mit Green Nudges fest, dass nicht alle Kundengruppen das gewünschte Verhalten (keine Retouren) an den Tag legen (S. 20). Ganz im Gegenteil, einige Kundengruppen scheinen die Intervention als zu invasiv empfunden zu haben, denn sie gaben ihre Bestellung gar nicht erst auf. Gerade aus diesem Grund ist es unabdingbar, Experimente durchzuführen, um die möglichen Interventionen und Kundengruppen zu ermitteln, die dem Unternehmen keinen wirtschaftlichen Schaden entstehen lassen.

Eine weitere Limitierung der betrachteten Studien basiert auf dem so genannten „cold start problem“ (Hofmann, et al., 2020, S. 9). In vielen Studien werden historische Daten bezogen auf ein bestimmtes Produkt verwendet. Probleme ergeben sich mit neu ins Sortiment aufgenommenen Produkten, da diese keine historischen Daten haben, auf die sich der Algorithmus stützen kann. In solchen Fällen lassen sich keine Prognosen aufstellen, welche speziell für

die Unternehmen ein großer Nachteil sein kann. Recherchen und Analysen zu dieser Limitierung müssen durchgeführt werden.

Die analysierten Studien befassen sich beispielsweise nur mit dem Produkt, welches natürlich eine große Rolle für die Entscheidung zur Retoure spielt. Es ist allerdings nicht der einzige Faktor, der betrachtet werden sollte. Rückgabebedingungen und der Retourenprozess beeinflussen neben dem Käuferverhalten ebenfalls die Retoureninitiierung. Zahlreiche Studien haben bewiesen, dass die Nachsichtigkeit der Rückgabebedingungen die Entscheidung eines Kunden zum Kauf (von Zahn, et al., 2022, S. 4) sowie zur Initiierung einer Retoure leiten. Je nachsichtiger die Rückgabebedingungen sind, desto eher und schneller lässt sich der Kunde zur Retoure hinreißen. Ebenso verhält es sich mit dem Retourenprozess als solches. Wenn dem Kunden mehr Verpflichtungen, wie z.B. das Ausdrucken des Retourenlabels, die Abgabe des Pakets bei der Post, aufgebürdet werden, desto mehr sinkt die Bereitschaft des Kunden all dies auf sich zu nehmen und das Produkt zu retournieren. Selbstverständlich ist die Sachlage nicht ganz so schwarz und weiß, wie es gerade dargestellt wurde, denn diese beiden Punkte können auch einen Einfluss auf den Absatz im Allgemeinen haben (von Zahn, et al., 2022, S. 1). Die Studien sollten jedoch Aspekte der Rückgabebedingungen sowie des Retourenprozesses mitberücksichtigen.

Wie bereits zu Beginn aufgeführt, stellen Retouren eine große Belastung für die Umwelt dar. Gerade unter diesem Gesichtspunkt sollten mehr Studien durchgeführt werden, damit die Kunden mittels Machine Learning dazu animiert werden ihr Einkaufsverhalten bewusster zu gestalten. Die ökologischen Auswirkungen von Retouren müssen den Kunden viel öfter vergegenwärtigt werden, um eine Reduzierung der Retouren herbeizuführen. Von Zahn et al. (2022) zeigen, dass dies nicht nur möglich ist, sondern auch zu einer Reduzierung von Retouren und einer Maximierung des Gewinns führt (S. 21 f.). Für diese Studien ist es unabdingbar soziologische und psychologische Aspekte zu berücksichtigen, die natürlich nicht zu Lasten des Datenschutzes kommen dürfen.

6 Quellenverzeichnis

- Ahsan, et al., 2021 Ahsan, K., Rahman, S.: A systematic review of e-tail product returns and an agenda for future research. *Industrial Management & Data Systems*, 1-31, 2021.
- Al Imran, et al., 2020 Al Imran, A., Amin, M. N.: Predicting the return of orders in the e-tail industry accompanying with model interpretation. *Procedia Computer Science*, 176, 1170-1179, 2020.
- Asdecker, et al., 2017 Asdecker, B., Karl, D., Sucky, E.: Examining Drivers of Consumer Returns in E-Tailing with Real Shop Data. *Proceedings of the 50th Annual Hawaii International Conference on System Sciences (HICSS)*, 4192-4201, 2017.
- Asdecker, et al., 2018 Asdecker, B., Karl, D.: Big data analytics in returns management - Are complex techniques necessary to forecast consumer returns properly? *2nd International Conference on Advanced Research Methods and Analytics (CARMA 2018) Proceedings*, 39-46, 2018.
- Cosar Sogukkuyu, et al., 2022 Cosar Sogukkuyu, D. Y., Senvar, O., Aysoysal, B., Yigit, E., Derelioglu, V., Varol, M. A., Polat, M. F., Sertbas, S., Caglar, G., Kocas, B., Tasoglu, K., Demirkale, H.: A Supervised Learning Algorithms for Consumer Product Returns Case Study for FLO Offline Stores. *Intelligent and Fuzzy Systems: Digital Acceleration and The New Normal-Proceedings of the INFUS 2022 Conference, Volume 2 (Vol. 505)*, 190-196, 2022.
- Data Scientist, 2023 Data Scientist: Cross-Validation: Bedeutung für Machine Learning. 2023, letzter Zugriff 10.04.2023. <https://datascientest.com/de/kreuzvalidierungsverfahren-definition-und-bedeutung-fur-machine-learning>
- Dzyabura, et al., 2018 Dzyabura, D., El Kihal, S., Ibragimov, M.: Leveraging the power of images in predicting product return rates. *SSRN Electronic Journal*, 1-33, 2018.
- Eichenberger, 2020 Eichenberger, L.: Machine Learning Goes Causal I: Why Causality Matters. 2020, letzter Zugriff 19.02.2023. <https://www.statworx.com/en/content-hub/blog/machine-learning-goes-causal-i-why-causality-matters/>
- Frei, et al., 2022 Frei, R., Jack, L., Krzyzaniak, S.-A.: Mapping product returns processes in multichannel retailing: challenges and opportunities. *Sustainability*, 14(3), 1382, 1-17, 2022.

- Frochte, 2019 Frochte, J.: Maschinelles Lernen: Grundlagen und Algorithmen in Python. München: Carl Hanser Verlag GmbH Co KG. 2019.
- Fürnkranz, 1999 Fürnkranz, J.: Separate-and-conquer rule learning. Artificial Intelligence Review, 13, 3-54, 1999.
- Gandhi, 2018 Gandhi, R.: Support Vector Machine — Introduction to Machine Learning Algorithms. 2018, letzter Zugriff 15.02.2023. <https://towardsdatascience.com/support-vector-machine-introduction-to-machine-learning-algorithms-934a444fca47>
- Godoy, 2018 Godoy, D.: Understanding binary cross-entropy / log loss: a visual explanation. 2018, letzter Zugriff 09.04.2023. <https://towardsdatascience.com/understanding-binary-cross-entropy-log-loss-a-visual-explanation-a3ac6025181a>
- Gonfalonieri, 2020 Gonfalonieri, A.: Introduction to Causality in Machine Learning. 2020, letzter Zugriff 10.02.2023. <https://towardsdatascience.com/introduction-to-causality-in-machine-learning-4cee9467f06f>
- Hofmann, et al., 2020 Hofmann, A., Gwinner, F., Fuchs, K., Winkelmann, A.: An Industry-Agnostic Approach for the Prediction of Return Shipments. Americas Conference on Information Systems, 1-10, 2020.
- Joshi, 2023 Joshi, A. V.: Machine Learning and Artificial Intelligence. Springer. 2023.
- Kafesu, 2023 Kafesu, A.: Unterstützung von Vector Machine (SVM) beim maschinellen Lernen. 2023, letzter Zugriff 18.06.2023. <https://geekflare.com/de/support-vector-machine/>
- Ketzenberg, et al., 2020 Ketzenberg, M. E., Abbey, J. D., Heim, G. R., Kumar, S.: Assessing customer return behaviors through data analytics. Journal of Operations Management, 66(6), 622-645, 2020.
- Lantz, 2019 Lantz, B.: Machine learning with R: expert techniques for predictive modeling. Birmingham: Packt Publishing. 2019.
- Li, et al., 2018 Li, J., He, J., Zhu, Y.: E-tail product return prediction via hypergraph-based local graph cut. Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, 519-527, 2018.
- Llango, 2021 Llango, V.: Tabnet — Deep Learning for Tabular data: Architecture Overview. 2021, letzter Zugriff 10.04.2023.

- <https://vigneshwarilango.medium.com/tabnet-deep-learning-for-tabular-data-architecture-overview-448ced8f8cfc>
- Luber, et al., 2016 Luber, S., Litzel, N.: Was ist Data Mining? 2016, letzter Zugriff 04.02.2023. <https://www.bigdata-insider.de/was-ist-data-mining-a-593421/>
- Microsoft, 2023 Microsoft: Hyperparameter tuning a model (v2). 2023, letzter Zugriff 26.01.2023. <https://learn.microsoft.com/en-us/azure/machine-learning/how-to-tune-hyperparameters?view=azure-ml-api-2>
- OECD, 2020 OECD: E-commerce in the times of COVID-19. 2020.
- Oppermann, 2021 Oppermann, A.: Accuracy, Precision, Recall, F1-Score und Specificity. 2021, letzter Zugriff 11.03.2023. <https://artemoppermann.com/de/accuracy-precision-recall-f1-score-und-specificity/>
- Pei, et al., 2018 Pei, Z., Paswan, A.: Consumers' Legitimate And Opportunistic Product Return Behaviors in Online Shopping. Journal of Electronic Commerce Research, Vol. 19, No. 4, 301-319, 2018.
- Saha, 2023 Saha, S.: XGBoost vs LightGBM: How Are They Different. 2023, letzter Zugriff 05.03.2023. <https://neptune.ai/blog/xgboost-vs-lightgbm>
- SAP, o. J. SAP: Was ist maschinelles Lernen? Ohne Jahresangabe, letzter Zugriff 17.02.2023. <https://www.sap.com/germany/insights/what-is-machine-learning.html>
- Seewald, et al., 2019 Seewald, A. K., Wernbacher, T., Pfeiffer, A., Denk, N., Platzer, M., Berger, M., Winter, T.: Towards Minimizing e-Commerce Returns for Clothing. ICAART (2), 801-808, 2019.
- Shehu, et al., 2020 Shehu, E., Papias, D., Neslin, S. A.: Free Shipping Promotions and Product Returns. Journal of Marketing Research. 57(4), 640-658, 2020.
- Spektrum, o. J. Spektrum: bipartiter Graph. Ohne Jahresangabe, letzter Zugriff 05.04.2023. <https://www.spektrum.de/lexikon/mathe/matik/bipartiter-graph/2289>
- Stadler, 2022 Stadler, M.-L.: Künstliche Intelligenz. 2022, letzter Zugriff 03.04.2023. <https://mindsquare.de/knowhow/kuenstliche-intelligenz/>

- Statista, 2018 Statista: Retouren im Online-Handel: Statista-Dossier zum Thema Retouren im Online-Handel. Statista, 2018.
- Statista, 2021 Statista: Anteil der Befragten, die ihre Einkäufe online getätigt haben statt im Geschäft, in Deutschland in den Jahren 2020 und 2021. 2021, letzter Zugriff 08.03.2023. <https://de.statista.com/statistik/daten/studie/1113346/umfrage/aenderung-des-einkaufsverhaltens-waehrend-der-corona-krise/>
- Statista, 2022a Statista: E-Commerce-Umsatz im Einzelhandel weltweit in den Jahren 2017 bis 2021 sowie eine Prognose bis 2026. 2022, letzter Zugriff 18.02.2023. <https://de.statista.com/statistik/daten/studie/244110/umfrage/globaler-umsatz-von-e-commerce/>
- Statista, 2022b Statista: Geschätzte Anzahl der Retouren von Paketen und Artikeln in Deutschland in den Jahr 2018 bis 2020. 2022, letzter Zugriff 08.03.2023. <https://de.statista.com/statistik/daten/studie/1082408/umfrage/retouren-von-paketen-und-artikeln-in-deutschland/>
- Statistik Guru, o. J. Statistik Guru: Multiple Lineare Regression. Ohne Jahresangabe, letzter Zugriff 03.04.2023. <https://statistikguru.de/spss/multiple-lineare-regression/voraussetzung-multikollinearitaet.html>
- Sun, et al., 2021 Sun, M., Chen, J., Tian, Y., Yan, Y.: The impact of online reviews in the presence of customer returns. International journal of production economics, 232, 107929, 1-13, 2021.
- TH Köln, o. J. TH Köln: Hypergraph. Ohne Jahresangabe, letzter Zugriff 17.01.2023. <https://wikis.gm.fh-koeln.de/Datenbanken/Hypergraph>
- Urbanke, et al., 2015 Urbanke, P., Kranz, J., Kolbe, L.: Predicting product returns in e-commerce: the contribution of mahalanobis feature extraction. Thirty Sixth International Conference on Information Systems, Forth Worth 2015, 1-19, 2015.
- Usability, o. J. Usability in Germany e.V.: Feature Extraction (KI). Ohne Jahresangabe, letzter Zugriff 28.01.2023. <https://www.kompetenzzentrum-usability.digital/definition/feature-extraction-ki>
- van Heeswijk, 2022 van Heeswijk, W.: Regulate Your Regression Model With Ridge, LASSO and ElasticNet. 2022, letzter Zugriff 07.03.2023. <https://towardsdatascience.com/regulate-your-regression-model-with-ridge-lasso-and-elasticnet-92735e192e34>

- Vinayak, et al., 2015 Vinayak, R. K., Gilad-Bachrach, R.: Dart: Dropouts meet multiple additive regression trees. *Artificial Intelligence and Statistics*, 489-497, 2015.
- von Zahn, et al., 2022 von Zahn, M., Bauer, K., Mihale-Wilson, C., Jagow, J., Speicher, M., Hinz, O.: The smart green nudge: Reducing product returns through enriched digital footprints & causal machine learning. *SAFE Working Paper*, No. 363, 2022.
- Wuttke, o. J. Wuttke, L.: Künstliche Neuronale Netzwerke: Definition, Einführung, Arten und Funktion. Ohne Jahresangabe, letzter Zugriff 18.06.2023. <https://datasolut.com/neuronalenetzwerke-einfuehrung/>
- Zhu, et al., 2018 Zhu, Y., Li, J., He, J., Quanz, B. L., Deshpande, A. A.: A Local Algorithm for Product Return Prediction in E-Commerce. *IJCAI*, 3718-3724, 2018.

7 **Abbildungsverzeichnis**

Abbildung 1 Englische Suchbegriffe.....	3
Abbildung 2 Deutsche Suchbegriffe.....	4
Abbildung 3 Aufbau eines neuronalen Netzwerks.....	24
Abbildung 4 Convolutional Neural Network.....	25
Abbildung 5 Entscheidungsbaum.....	26
Abbildung 6 Support Vector Machine.....	29
Abbildung 7 Hybrider Graph.....	31
Abbildung 8 Hypergraph.....	31

8 Glossar

Accuracy:
$$\frac{\text{richtig positiv} + \text{richtig negativ}}{\text{richtig positiv} + \text{richtig negativ} + \text{falsch positiv} + \text{falsch negativ}}$$

Die Accuracy wird bei Klassifikationsproblemen eingesetzt und zeigt die Fähigkeit des Modells auf zwischen den betrachteten Klassen zu unterscheiden. Je höher der Wert, desto zutreffender die Klassifikation des Modells (Oppermann, 2021).

AUC: Area under the curve ist der Bereich unterhalb der ROC-Kurve und zeigt die Fähigkeit des Modells auf zwischen den betrachteten Klassen, z.B. Retoure und keine Retoure, zu unterscheiden. Je höher der Wert, desto zutreffender die Klassifikation des Modells (Joshi, 2023, S. 219).

Binary Log Loss Funktion: wird in logistischen Regressionsproblemen eingesetzt und dient dazu, die Leistung des Modells zu beurteilen (Godoy, 2018).

Bipartiter Graph: die Knoten eines bipartiten Graphen können in zwei disjunkte Mengen aufgeteilt werden, die innerhalb einer Menge nicht miteinander verbunden sind (Spektrum, o. J.).

C5.0 Algorithmus: Industriestandard für die Generierung von Entscheidungsbäumen (Lantz, 2019, S. 124).

CHIAD: Chi-Squared Automatic Interaction Detector ist ein Algorithmus für die Erstellung von Entscheidungsbäumen (Joshi, 2023, S. 75). Wie der Name suggeriert, nutzt der Algorithmus die Chi-Quadrat-Verteilung, um einen Datensatz mit der Menge an Daten zu vergleichen und eine Prognose hinsichtlich der Zugehörigkeit mit der gewünschten statistischen Signifikanz aufzustellen (Joshi, 2023, S. 81).

CNN: Convolutional Neural Network besteht aus einem Convolutional Layer, welches aus einer Reihe von Kernels besteht, einer Rectified Linear Unit, die alle negativen Werte im Output des Convolutional Layers in 0 umwandelt, und einem Pooling Layer (Joshi, 2023, S. 154 f.), in dem die Informationen verallgemeinert und verdichtet werden (Frochte, 2019, S. 235).

DART: Dropouts meet multiple additive regression trees, ein Algorithmus, der in jeder Boosting Runde einen nicht relevanten Bereich des Regressionsbaums eliminiert (Vinayak, et al., 2015).

Data Mining: durch den Einsatz von Algorithmen werden Muster und Zusammenhänge in einem Datenpool ermittelt (Luber, et al., 2016).

Elastic Net: eine Regularisierungsmethode, welche in der Regression eingesetzt wird. Es reduziert die Anzahl der Features und vergibt den verbleibenden Features eine aussagekräftigere Gewichtung (van Heeswijk, 2022).

$$\mathbf{F1-Score:} \frac{2 * \text{richtig positiv}}{2 * \text{richtig positiv} + \text{falsch positiv} + \text{falsch negativ}}$$

Der F1-Score errechnet sich aus Precision und Recall. Je höher der Wert ausfällt, desto besser schneidet das Modell ab (Oppermann, 2021).

Hyperparameter: Parameter, die den Lernprozess des Modells steuern. Ein Hyperparameter ist beispielsweise die Anzahl der versteckten Schichten in neuronalen Netzwerken (Microsoft, 2023).

Korrelationskoeffizient: auch Person's r genannt. Die Werte liegen zwischen -1 und 1 und geben an, inwieweit eine Korrelation zwischen dem untersuchten Feature und dem Ergebnis vorliegt (Joshi, 2023, S. 186).

Lasso Regularisierung: Lasso steht für Least absolute shrinkage and selection operator und ist eine Regularisierungsmethode, die bei Regressionsproblemen eingesetzt wird, um die aussagekräftigsten Features zu selektieren. Hierzu wird ein sogenannter Strafterm eingesetzt, der die nicht relevanten Features auf 0 setzt und somit die Komplexität des Modells reduziert (van Heeswijk, 2022).

$$\mathbf{Precision:} \frac{\text{richtig positiv}}{\text{richtig positiv} + \text{falsch positiv}}$$

Verdeutlicht die vom Modell richtig prognostizierten Resultate und betrachtet lediglich die positiven Prognosen (richtig Positive, falsch Positive). Precision wird im Zusammenhang mit der binären Klassifikation eingesetzt (Oppermann, 2021).

$$\mathbf{Recall:} \frac{\text{richtig positiv}}{\text{richtig positiv} + \text{falsch negativ}}$$

Beschreibt die Fähigkeit des Modells, positive Resultate zu lokalisieren und berücksichtigt neben den richtig Positiven auch die falsch Negativen. Diese Metrik kommt in der Klassifikation zum Einsatz (Oppermann, 2021).

Ridge Regression: ist eine Regularisierungsmethode, die bei Regressionsproblemen eingesetzt wird. Ziel ist es, die Varianz des Modells zu verringern (van Heeswijk, 2022).

ROC: Receiver Operating Characteristics. Bei binären Klassifikationsproblemen trennt ein Threshold die beiden involvierten Klassen. Dieser Threshold hat einen unmittelbaren Einfluss auf die Metriken Precision und Recall, die sich gegensätzlich verhalten. Je besser die Precision wird, desto schlechter wird der Recall und umgekehrt. Um einen optimalen Threshold zu ermitteln, werden die Precision und Recall Werte bei einem bestimmten Threshold ermittelt und in Form einer Kurve dargestellt, der ROC (Joshi, 2023, S. 219).

Supervised Machine Learning: der Algorithmus wird mittels bereits gelabelter Daten trainiert und wird bei Regressions- und Klassifikationsproblemen eingesetzt (Frochte, 2019, S. 20 f.).

TAE: Total Absolute Error ist eine Metrik, die auf numerische Fehlerwerte zielt und für Regressionsanalysen eingesetzt wird. Je kleiner der TAE-Wert ist, desto besser ist die Performanz des Modells (Joshi, 2023, S. 215).

Zweifache Kreuzvalidierung: hierbei wird der verfügbare Datenpool halbiert, wobei beide Teile in unterschiedlichen Schritten jeweils zu Trainings- und Testzwecken eingesetzt werden. Die Ergebnisse werden anschließend kombiniert, um die Leistung des Modells zu bewerten (Data Scientist, 2023).