



FOM Hochschule für Oekonomie & Management

Hochschulzentrum Münster

Hausarbeit

im Studiengang Big Data & Business Analytics

zur Erlangung des Grades eines

Master of Science (M. Sc.)

über das Thema

Ansätze zu Optimierung und Automatisierung des OSMI-Index

von

Fiete Ostkamp, Verena Rakers und Artur Gergert

Betreuer : Prof. Dr. Rüdiger Buchkremer

Matrikelnummer : 557851, 536491, 562394

Abgabedatum : 30. August 2022

Inhaltsverzeichnis

Abbildungsverzeichnis	III
Tabellenverzeichnis	IV
Abkürzungsverzeichnis	V
Symbolverzeichnis	VI
1 Einleitung	1
1.1 Problemstellung	1
1.2 Zielsetzung	1
1.3 Vorgehen	2
2 Theoretische Grundlagen	3
2.1 Online Sensory Marketing Index	3
2.2 Deep Learning	4
2.2.1 Image Classification	5
2.2.2 Convolutional Neural Network	5
2.3 Natural Language Processing	6
2.3.1 Named Entity Recognition	6
2.3.2 Vorgehen bei der NER	7
3 Praktische Umsetzung	9
3.1 OSMI-Index	9
3.2 Datenvorbereitung	10
3.3 Modellentwicklung	11
3.3.1 Image Classification Modell	11
3.4 Klassifizierung von Webseiten	11
3.4.1 Datenerhebung	12
3.4.2 Datenverarbeitung	13
3.4.3 Training des Modells	14
3.4.4 NER-Modell	15
3.5 Dashboard zur Darstellung der Ergebnisse	16
4 Fazit	17
Anhang	19

Abbildungsverzeichnis

Abbildung 1: Aufbau des Convolutional Neural Networks	5
Abbildung 2: Beispiel der Named Entity Recognition (NER) Technik	7
Abbildung 3: Zielbild des Consulting-Projekts	9
Abbildung 4: Annotationsstand Textlabeling	11
Abbildung 5: Annotationsstand Textlabeling	11
Abbildung 7: Vergleich der Trainings- und Validierungsmetriken mit und ohne Data Augmentation	14

Tabellenverzeichnis

Tabelle 1: Evaluierung mittels Online Sensory Marketing Index (OSMI) am Beispiel der Webseite <i>www.mymuesli.de</i>	4
---	---

Abkürzungsverzeichnis

CNN	Convolutional Neural Network
FN	False Negative
FP	False Positive
FP	False Positive
NER	Named Entity Recognition
NLP	Natural Language Processing
OSMI	Online Sensory Marketing Index
TP	True Positive
TF-IDF	Term Frequency-Inverse Document Frequency

Symbolverzeichnis

1 Einleitung

Für Unternehmen wird es immer wichtiger die potenziellen Kunden gezielt anzusprechen und zum Kauf anzuregen. Eine Möglichkeit kann hierbei das multisensorische Marketing bieten, welches zur Beeinflussung des Kaufverhaltens mehrere Sinne anspricht. Eine Umfrage der Mood Media Corporation¹, die unter 10.000 Menschen weltweit durchgeführt wurde, hat ergeben, dass 75% der Kunden sich länger in einem Geschäft aufhalten würden, wenn durch Musik oder Düfte eine positive Atmosphäre geschaffen wird.² Multisensorisches Marketing anzuwenden kann demnach einen positiven Effekt auf das Kaufverhalten haben.

1.1 Problemstellung

Die Anwendung von multisensorischem Marketing findet zurzeit mehrheitlich in stationären Geschäften statt. Da der Online-Handel weiterhin an Wachstum gewinnt, sollte auch hier geprüft werden inwiefern multisensorisches Marketing eingesetzt werden kann. Ein von K. Hamacher entwickelter Index, der Online Sensory Marketing Index (OSMI) setzt an dieser Thematik an. Es ist ein Evaluationsmodell, welches Webseiten anhand ihrer multisensorischen Kommunikationsqualitäten bewertet. Allerdings gibt es noch Verbesserungspotentiale in der Nutzung des OSMI als Instrument zur Erfassung der multisensorischen Eigenschaften von Webseiten. So wird der Online Sensory Marketing Index bisher nur manuell und subjektiv ermittelt, wohingegen eine automatische Bewertung die Ergebnisse objektiver machen würde und den Ansatz skalierbarer machen würde. Die Ergebnisse werden außerdem lediglich in einer Tabelle dargestellt und könnten daher für einen Endanwender noch zugänglicher aufbereitet werden. Des Weiteren wird bei der Beurteilung einer Webseite nicht nach Produktwebseiten und nicht-Produktseiten unterschieden und die Textanalyse mittels Term Frequency-Inverse Document Frequency (TF-IDF) erfasst die sensorischen Wörter nicht im Satzkontext. Dadurch können einige Entitäten weniger genau erfasst werden, da grammatikalische Konstellation wie Negationen nicht berücksichtigt werden.

1.2 Zielsetzung

Aus der Problemstellung lässt sich daher nachfolgende Zielsetzung für diese Ausarbeitung ableiten. Es soll eine Webseite erstellt werden, in derer unterschiedliche Produktwebsei-

¹ steiner2022.

² moodmedia2019.

ten automatisiert hinsichtlich ihrer textlichen Ansprache an die Sinne des Konsumenten aufgeführt werden. Durch eine Evaluierung wird dargelegt wie gut oder wie schlecht die Webseite die Sinne Tasten, Riechen, Hören, Schmecken und Sehen anspricht. Darauf aufbauend könnten dann die Webseiten-Betreiber die Webseiten optimieren, um die Kunden besser anzusprechen.

1.3 Vorgehen

Ein erster Schritt in der Umsetzung der Zielsetzung ist es die relevanten produktspezifischen Webseiten herauszufiltern. Mittels einer Bildklassifizierung soll dies umgesetzt werden, um diese für die weitere Bearbeitung verwenden zu können. Andere Unterseiten zum Unternehmen oder zu Karrieremöglichkeiten fließen somit nicht mehr in eine Bewertung ein. Im zweiten Schritt werden durch Natural Language Processing (NLP) die Texte auf den Produktseiten analysiert. Es werden unter Einbezug des Satzkontextes gezielt Wörter markiert, die auf eines der fünf Sinne hindeuten. Eine Beurteilung dieses Ergebnisses wird anhand eines Benchmarkings ermittelt und auf einer Webseite zur Verfügung gestellt.

2 Theoretische Grundlagen

Im folgenden Teil der Ausarbeitung werden die theoretischen Grundlagen zum Verständnis der Ausarbeitung erläutert. Zunächst wird der OSMI als Hauptaugenmerk erklärt. Es folgen der Oberbegriff Deep Learning und die dazugehörige Methode der Bildklassifizierung. Darauf folgt das Natural Language Processing mit der Named Entity Recognition.

2.1 Online Sensory Marketing Index

Der Fokus dieser Ausarbeitung liegt auf dem Online Sensory Marketing Index (OSMI). Es handelt sich dabei um ein Bewertungsinstrument, welches im Jahr 2018 entwickelt wurde. Ziel des OSMI ist es, dass E-Commerce Webseiten im Hinblick auf deren sensorische Gestaltung bewertet werden und dementsprechend Verbesserungspotentiale aufgezeigt werden.³ Unter einer sensorischen Gestaltung werden Parameter gefasst, die sich nach den Sinnen eines Menschen richten. Diese sind die Haptik, die Olfaktorik, die Akustik, die Gustatorik und die Visualität.⁴ Die Bewertung im OSMI wird anhand verschiedener Indikatoren der einzelnen Sinne vorgenommen. Nachfolgend sind beispielhafte Indikatoren der einzelnen Sinne aufgeführt, die im OSMI Berücksichtigung finden:

- Haptik: 3D-Bilder, Video, Endowment-Effekt, ...
- Olfaktorik: Mentale Simulation von Düfte, Schlüsselbegriffe für Düfte
- Auditive: Töne/Musik, Sprecher/Stimme, Schlüsselbegriffe für Akustik, ...
- Gustatorische: Mental Simulation des Schmeckens, Verwendung von Farbschemata, Produktoption & Oberflächen, ...
- Visuelle: Farbgebung, dynamische Bilder, Oberfläche, ...⁵

Die Evaluation der einzelnen Indikatoren wird in Form einer Tabelle dargestellt. Ein Beispiel ist am Ende dieses Kapitels aufgeführt (Abbildung 1). Für eine Webseite wird jeder Indikator der fünf Sinne einzeln bewertet und auf Basis dessen letztendlich für jeden der fünf Parameter ein Wert zwischen 0 und 1 ermittelt. Eine Gewichtung zwischen den einzelnen Indikatoren wird dabei nicht vorgenommen.

Im letzten Schritt zur Ermittlung des OSMI wird über die Parameter der fünf Sinne das arithmetische Mittel berechnet. Dieser Wert liegt ebenfalls zwischen 0 und 1 und bildet

³ hamacher2018.

⁴ zou2018.

⁵ hamacher2018.

den OSMI einer Webseite. Dabei gilt, dass die Webseite stärker die Sensorik anspricht, wenn der OSMI-Index näher an einer 1 liegt. Liegt der Wert näher an einer 0 werden schlussfolgernd wichtige Elemente zur Ansprache der Sensorik nicht genutzt.

⁶ Die nachfolgende Tabelle zeigt ein Beispiel aus der Masterarbeit von K.Hamacher zum OSMI der Webseite *www.mymuesli.de*.

Tabelle 1: Evaluierung mittels OSMI am Beispiel der Webseite *www.mymuesli.de*

	HAPTİK			OLFAKTORİK			AKUSTİK			GUSTATORİK			VISUALİTÄT		
	0-3	H1	0	0-3	O1	1	0-3	A1	0	0-3	G1	3	0-3	V1	2
	0-1	H2	0	0-2	O2	0	0-1	A2	0	0-1	G2	1	0-1	V2	1
	0-2	H3	0				0-1	A3	1	0-1	G3	0	0-1	V3	0
	0-3	H4	3				0-3	A4	0	0-1	G4	0	0-1	V4	0
	0-3	H5	1							0-1	G5	0	0-1	V5	0
	0-1	H6	0										0-1	V6	0
	0-1	H7	0												
	0-2	H8	2												
	0-1	H9	0												
	0-3	H10	1												
	0-1	H11	0												
	0-1	H12	0												
	22			5			8			7			8		
\bar{x}	0,32			0,20			0,13			0,57			0,38		
OSMI = 1/5 (0,32+0,20+0,13+0,57+0,38) = 0,32															

Quelle: hamacher2018

2.2 Deep Learning

Die Image Classification ist ein Anwendungsfall für die Technik des Deep Learning. Als Teilgebiet des Machine Learning bedient sich das Deep Learning mehrschichtiger neuronaler Netze, welche auf Basis großer Datenmengen bereits Erlerntes mit neuen Inhalten verknüpft und somit wiederum erneut lernt.⁷ Deep Learning macht es möglich, dass eine Maschine selbstständig die Gewichtungen seines verwendeten Modells verbessert. Anwendungsbereiche für Deep Learning können beispielsweise im Customer Support, in der medizinischen Versorgung oder selbstfahrende Autos sein.⁸ Ein weiterer Anwendungsbereich kann die Objekterkennung sein, welcher nachfolgend beschrieben wird.

⁶ hamacher2018.

⁷ plaat2022.

⁸ handa2021.

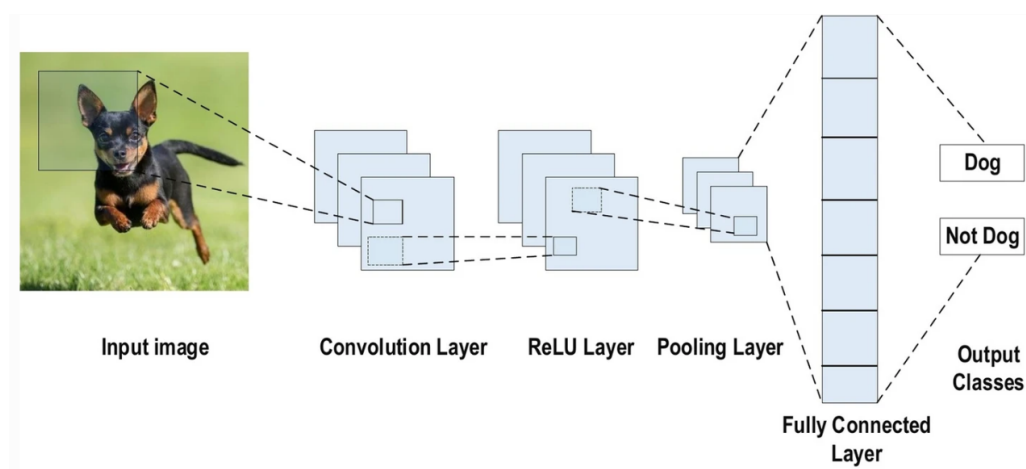
2.2.1 Image Classification

Bei der Image Classification handelt es sich um eine Technik, bei der ein Objekt auf einem Bild einer bestimmten Klasse zugeordnet wird. Als Ziel kann die genaue Identifizierung der Bildmerkmale verstanden werden.⁹ Eine Methode, die zur Image Classification eingesetzt werden kann ist das Convolutional Neural Network (CNN)¹⁰, welches nachfolgend näher erläutert wird.

2.2.2 Convolutional Neural Network

Das CNN ist eine besondere Form eines künstlichen neuronalen Netzes. Es zeichnet sich durch mehrere Schichten wie den Convolutional-Layer, den Pooling-Layer und den Fully-Connected Layer aus.¹¹ Die Architektur ist in der Abbildung 1 dargestellt.

Abbildung 1: Aufbau des Convolutional Neural Networks



Quelle: alzubaidiReviewDeepLearning2021

In den Schichten der Convolution werden die Eingabebilder analysiert und Merkmale (zum Beispiel Linien oder Kanten) der Bilder erkannt und extrahiert. Die Verarbeitung dieser Merkmale erfolgt in Matrizen, die Feature Maps genannt werden.¹²

Die Feature Maps als Ausgabe der Convolutional Layer werden darauf in den Pooling Layer eingegeben. Die Größe der Bilder wird in dieser Schicht reduziert und durch Methoden

⁹ internationalconferenceoncomputercommunicationandinformatics2021.

¹⁰ internationalconferenceonadvancesinelectronics2018.

¹¹ internationalconferenceonadvancesinelectronics2018.

¹² Goodfellow-2016.

wie dem Max-Pooling nur die wesentlichen Merkmale weitergegeben.¹³ Alle anderen Informationen sind für die Verarbeitung überflüssig und werden verworfen. Ergebnis dieser Schicht ist die gleiche Anzahl an Feature Maps, allerdings in komprimierter Form.

Der Fully-Connected Layer bildet den Abschluss der CNN-Architektur. Jeder Knoten in der Ausgabeschicht ist hier direkt mit einem Knoten der vorgelagerten Schicht verbunden.¹⁴ Mithilfe einer Linearkombination und einer Aktivierungsfunktion wird schlussendlich das Bild klassifiziert. Die Zuordnung zu einer Klasse erfolgt dabei auf Basis einer Wahrscheinlichkeit.¹⁵

2.3 Natural Language Processing

Die Named Entity Recognition ist eine Technik, welche dem Oberbegriff des Natural Language Processing zugeordnet wird. Das NLP beschäftigt sich mit der Verarbeitung und dem Verständnis der menschlichen Sprache durch einen Computer. Ziel dieser Techniken ist es eine direkte Kommunikation zwischen dem Menschen und dem Computer mithilfe der Sprache herzustellen. Die Fachgebiete der Computerlinguistik, Informatik, Kognitionswissenschaft und künstlicher Intelligenz werden hier vereint. Anwendungsfälle können beispielsweise die Extraktion der Bedeutung von Sätzen oder Satzteilen oder auch die Erkennung von Satzzusammenhängen sein.¹⁶ Die populärste Funktion des NLP ist die Named Entity Recognition (NER).¹⁷

2.3.1 Named Entity Recognition

Die Named Entity Recognition hat zur Aufgabe automatisiert Entitäten in einem Textdokument zu finden und zu klassifizieren.¹⁸ Letzteres geschieht anhand zuvor definierter Kategorien. Eine Entität muss dabei nicht unbedingt nur aus einem einzelnen Wort bestehen, sondern kann auch eine Reihe von Wörtern umfassen, solange sie sich auf dieselbe Sache beziehen.¹⁹

¹³ Goodfellow-2016.

¹⁴ aggarwal2021.

¹⁵ tang2022.

¹⁶ deng2018.

¹⁷ strobl2022.

¹⁸ kulkarni2022.

¹⁹ strobl2022.

2.3.2 Vorgehen bei der NER

In der NER-Technik werden zwei Schritte durchgeführt, um zu Ergebnissen zu gelangen:

1. Erkennen einer Entität
2. Kategorisierung der Entität²⁰

Im ersten Schritt wird geprüft, ob eine Wortfolge eine Entität bildet. Die Anfangs- und Endgrenzen der Entitäten werden hierbei festgelegt. Der zweite Schritt hat zum Ziel die zuvor definierte Entität in eine der zuvor definierten Klassen einzuordnen.²¹ Häufig verwendete Klassen sind beispielsweise Orte, Namen oder Organisationen, welche unter die generischen Kategorien fallen. Des Weiteren gibt es domänenspezifische Kategorien wie zum Beispiel Proteine, Enzyme oder Gene.²²

Zur Durchführung der beiden Schritte gibt es verschiedene Ansätze, die angewendet werden können. Ausgehend von den annotierten Datensätzen, die beispielsweise aufgrund von manuell erstellten Regeln oder auf Basis von Kontextähnlichkeiten generiert werden, werden Modelle mithilfe von Machine Learning entwickelt. Für zuvor ungesehene Daten ermitteln die Modelle Vorhersagemodelle zur Erkennung und Kategorisierung der Entitäten. Eine weitere Möglichkeit ist der Einsatz von Deep Learning zur NER. In diesem Ansatz können auch nicht-lineare Zusammenhänge erkannt und gelernt werden.²³

Ein Beispiel für ein Textdokument, in dem die NER Technik angewendet wurde, ist in Abbildung 2 dargestellt.

Abbildung 2: Beispiel der NER Technik

performance vehicle that excites with unique power delivery, tantalizing engine sound, excellent handling and a
 •SOUND

direct driving sensation. the proverbial exclusivity also comes as standard, as do outstanding quality and

precise workmanship. in brief: all mercedes-amg models embody the levels of engineering and craftsmanship
 •PRODUCT

Quelle: Eigene Darstellung

Um die Qualität der NER-Technik zu bewerten, werden die Kennzahlen Precision, Recall und der F-Score hinzugezogen.²⁴ Zur Ermittlung dieser muss zunächst die Anzahl der

²⁰ zong2021.

²¹ bast2016.

²² li.2018.

²³ li.2018.

²⁴ li.2018.

Entitäten mit verschiedenen Ausprägungen ermittelt werden (False Positive (FP), False Negative (FN), True Positive (TP)).

- FP: Eine Entität wurde erkannt, obwohl sie keine darstellt
- FN: Eine Entität wurde nicht erkannt, obwohl sie eine darstellt
- TP: Eine Entität wurde richtig erkannt²⁵

Die Precision zeigt das Verhältnis zwischen richtig erkannten Entitäten und der Gesamtheit der identifizierten Entitäten an. Die Formel lautet wie folgt:

$$Precision = \frac{TP}{(TP + FP)} \quad (1)$$

Der Recall stellt den Anteil der richtig erkannten Entitäten an der Gesamtheit aller möglichen Entitäten dar.

$$Recall = \frac{TP}{(TP + FN)} \quad (2)$$

Der F-Score ist die Kennzahl, die die Precision und den Recall zu einem harmonischen Mittel vereint:

$$F - Score = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall} \quad (3)$$

²⁵ **larner2021.**

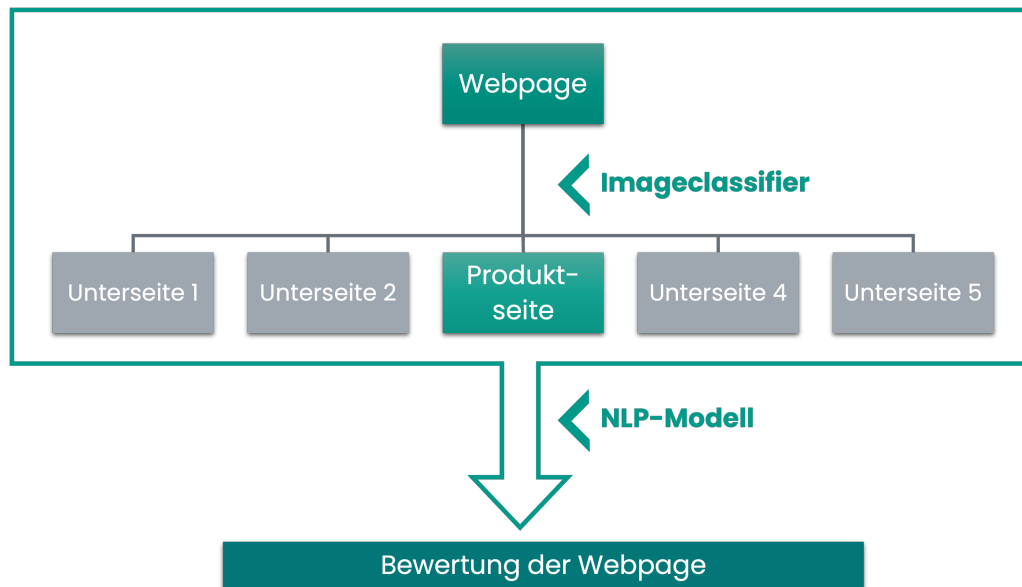
²⁶ **wang2021.**

3 Praktische Umsetzung

In der praktischen Umsetzung werden die Grundlagen auf den Anwendungsfall angewendet. Hierfür wird zunächst der OSMI-Index erläutert. Es folgen die Bildklassifizierung und die Umsetzung der Named Entity Recognition (NER). Zum Abschluss des praktischen Teils folgt ein Kapitel mit Consultinganteil.

3.1 OSMI-Index

Abbildung 3: Zielbild des Consulting-Projekts



Quelle: Eigene Darstellung

Um den OSMI-Index zu optimieren und zu automatisieren, wurde die in Abbildung 3 Pipeline aufgestellt mithilfe dessen eine Website in Bezug auf den OSMI verarbeitet wird. Der erste Schritt stellt dabei eine Vorverarbeitung der zu analysierenden Website dar, indem ein Image Classifier die Haupt- und Unterseiten der Website in Produkt- und Nicht-Produktseiten unterteilt. Dies ist aus Sicht der Autoren dieser Hausarbeit ein essenzieller Schritt, da so sichergestellt wird, dass die Bewertung des OSMI-Index lediglich auf Basis der relevanten URLs vorgenommen wird. Im darauf folgenden Schritt wird auf die textlichen Inhalte der relevanten Produktseiten ein NLP-Modell angewendet, welches die Entitäten des OSMI-Index erkennt und aus zählt. Im abschließenden Schritt werden die Ergebnisse der Website-Analyse im Rahmen eines Dashboards visualisiert. Nachfolgend wird die Umsetzung der soeben beschriebenen Schritte im Detail beschrieben und erläutert.

3.2 Datenvorbereitung

Sowohl in der Literatur als auch in bereits etablierten, gut entwickelten und veröffentlichten Modellen ist das Gebiet des Sensory Marketings sehr rar vertreten. Das bedeutet auch, dass im Hinblick auf den jüngst publizierten OSMI-Index ebenfalls bisher wenig technische Entwicklungen vorgenommen worden sind. Zwar existieren in der Literatur diverse NLP-Modelle wie beispielsweise das Spacy-Modell²⁷ oder der sogenannte BERT²⁸ und von beiden erwähnten Modellen Spezialisierungen wie beispielsweise SciSpacy, BioBERT, ESG-BERT, allerdings ist bisher kein NLP-Modell auf Marketingkontexte trainiert und publiziert worden. Daher besteht die Notwendigkeit zur Entwicklung eines eigenen Modells im Kontext des Consulting-Auftrags. Um im Vorfeld der Website-Bewertung jedoch lediglich nur die relevanten URLs mit Produktinhalt zu selektieren, wurde für den zu entwickelnden Image Classifier ebenfalls ein von Grund auf neu zu entwickelndes Modell benötigt, da es innerhalb der Literatur bislang kein derartig existierendes Modell publiziert wurde auf das zurückgegriffen werden konnte.

Die Trainingsdaten wurden mithilfe von Doccano generiert. Doccano²⁹ ist eine Open Source Software, welche zur Annotation von Text und Bilddaten entwickelt wurde. Mit diesem Instrument ist es also möglich Textpassagen und Bilder hochzuladen und Annotationen vorzunehmen, um Daten zur Entwicklung von Aufträgen wie Named Entity Recognition, Textzusammenfassungen, Sentimentanalysen, Image Classification, etc. zu generieren.

Im Rahmen des Consulting-Projekts wurden zwei Doccano-Projekte gestartet. Ein Projekt wurde für das Annotieren von Textpassagen zur Entwicklung eines Named Entity Recognition Modells erstellt. Im zweiten Projekt wurden Screenshots von Produktseiten und Nicht-Produktseiten annotiert um ein Classification Modell erstellen zu können.

Die Textpassagen, die annotiert wurden, sind im Rahmen des Consulting-Projekts zur Verfügung gestellt worden und entstammen Webscraping-Ergebnissen von Vorgängerprojekten. Hier wurden die fünf Sinne *Sight*, *Smell*, *Sound*, *Taste* und *Touch* als Label festgelegt, da dies die Entitäten sind die später auch für den OSMI nach Hamacher notwendig sind.

Abbildung 4 stellt ein Balkendiagramm über den Annotationsstand dar, das die Häufigkeit der einzelnen Labels aufzeigt. Diese Verteilung der Daten bildet die Basis auf der das Modell zur Erkennung von Entitäten trainiert wurde. Es ist zu erkennen, dass die Entität

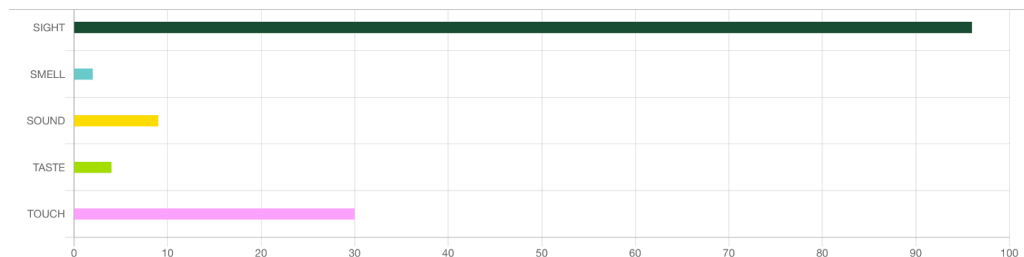
²⁷ spacy2017.

²⁸ devlin2018.

²⁹ doccano-2018.

Sight am häufigsten annotiert worden ist, während die Trainings- und Testdaten für die Entitäten *Smell*, *Sound* und *Taste* kaum vertreten sind.

Abbildung 4: Annotationsstand Textlabeling



Quelle: Eigene Darstellung

Die zu annotierenden Bilddaten von Produkt- und Nicht-Produktseiten wurden im Rahmen des behandelten Projektes erneut gesammelt (siehe auch Abschnitt 3.4.2). Abbildung 5 zeigt, dass zur Modellentwicklung für den Image Classifier eine annähernde Gleichverteilung zwischen Produkt- und Nicht-Produktseiten zugrunde gelegt werden konnte. Der Gesamtdatensatz beläuft sich auf rund 500 annotierten Bildern von Homepages.

Abbildung 5: Annotationsstand Textlabeling



Quelle: Eigene Darstellung

3.3 Modellentwicklung

3.3.1 Image Classification Modell

3.4 Klassifizierung von Webseiten

Es soll ein Modell entwickelt werden, dass in der Lage ist zwischen Produkt und nicht-Produktwebseiten zu unterscheiden. Wie Kapitel 2.2.1 deutlich geworden ist, kann dies von Klassifizierungsmodellen geleistet werden. Im Folgenden wird das Vorgehen mit den Schritten der Datenerhebung (3.4.1), der Datenverarbeitung (3.4.2) und des Modelltrainings (3.4.3) näher beschrieben.

3.4.1 Datenerhebung

Aus vergangenen Arbeiten von Studierenden steht bereits eine Datengrundlage zur Verfügung mit Textauszügen, dem HTML Quelltext und heruntergeladenen Bildern von Unternehmenswebseiten. Grundsätzlich sind dies die Elemente, die die semantischen und inhaltlichen Informationen der Webseite beinhalten, allerdings fehlen zwei für das äußere Erscheinungsbild der Webseite notwendige Elemente: Javascript, das für das dynamische Verhalten der Webseite verantwortlich ist und CSS, das für das Styling der HTML-Elemente verwendet wird. Wären diese Elemente vorhanden, so könnten sie - insofern sie in der richtigen Struktur vorliegen - dafür verwendet werden auch nachträglich, offline das Erscheinungsbild der Webseiten zu reproduzieren.

Vor diesem Hintergrund konnten daher die bereits gesammelten Daten nicht verwendet werden und es war notwendig eigene Screenshots mit einem Webscrawler zu erstellen. Für die Klassifizierung werden jedoch Bilder beziehungsweise Screenshots benötigt, die die gerenderte, visuelle Struktur der Webseite erfassen. Dies ist in den bestehenden Daten nicht gegeben und machte daher eine erneute Erhebung notwendig.

Für diesen Zweck wurde zunächst ein Webcrawler auf Basis von Scrapy³⁰ erstellt,³¹ der alle in der Vergangenheit gescrapten Webseiten aufruft und einen Screenshot der jeweiligen Seite erstellt. Auf diese Weise wurden circa 10.000 Bilder gesammelt, die als Eingangsdaten für das Modell verwendet werden können.

Zwei Probleme sind bei der Datenerhebung zutage getreten. Webseiten sind - gerade im Bereich von Unternehmen und Marketing - stetigen Veränderungen ausgesetzt. Dies zeigte sich auch in der erneuten Erfassung nach über zwei Jahren, da für einige Urls die Webseite nicht mehr verfügbar war (siehe auch Abbildung 6). Ein zweiter Aspekt der dynamischen Natur von Webseiten ist die Veränderung des Inhalts, der sich rückwirkend nicht so einfach feststellen lässt wie eine 404 Status Page, der dennoch aber nicht ausgeschlossen werden kann.

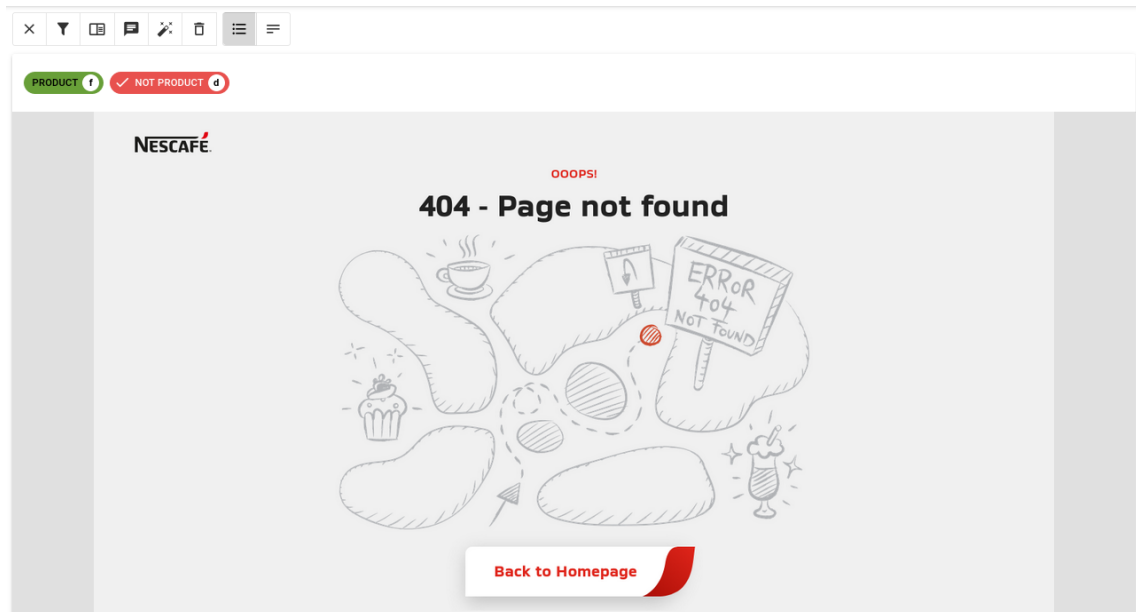
Das zweite wesentliche Problem war technischer Natur und bezog sich auf die Ladezeiten der Webseiten. In einigen Fällen ist es vorgekommen, dass auch nach einer Wartezeit von 3 Sekunden, noch nicht alle Elemente der Webseite - insbesondere Bilder - geladen waren. Die gewählte Wartezeit sollte vor dem Hintergrund der Internetverbindung und der Ladezeit typischer Webseiten (Zitat) eigentlich ausreichend sein, um eine vollständig geladene Seite zu erfassen. Die Autoren haben darauf verzichtet sie weiter zu erhöhen, da

³⁰ zotero-328.

³¹ ostkamp2022a.

auch mit der gewählten Zeit bei 10000 Seiten ein Durchgang bereits 8 Stunden dauert und jede weitere Sekunde eine Verlängerung von 3 Stunden bedeutet.

Abbildung 6: Bildklassifizierung in Doccano



Quelle: Eigene Darstellung

Da es sich bei der Klassifizierung um eine Supervised Learning Methode handelt, müssen die Daten daraufhin den Kategorien *Product* und *Non Product* zugeordnet werden. Dafür werden die Daten in die Annotationssoftware Doccano³² geladen und dann, wie in Abbildung 6 dargestellt, den Kategorien zugeordnet.

3.4.2 Datenverarbeitung

Im Schritt der Datenverarbeitung werden die Eingangsdaten in eine Form gebracht, in der sie für das Training des Modells verwendet werden können. Dazu werden die Bilder anhand des Exports der Annotationen aus Doccano in den Labeln gleichnamige Ordner verschoben.

Da die Screenshots als *.png* Bilder vorliegen, verfügen sie über einen Alpha-Channel der für die Verwendung im Modell hinderlich ist. Aus diesem Grund wurde dieser in einem weiteren Schritt entfernt. Darüber hinaus werden die Bilder im ersten Layer des Modells auf eine Größe von 224x224 Pixeln herunterskaliert, was eine gängige Größe für Convolutional Neural Networks darstellt.³³

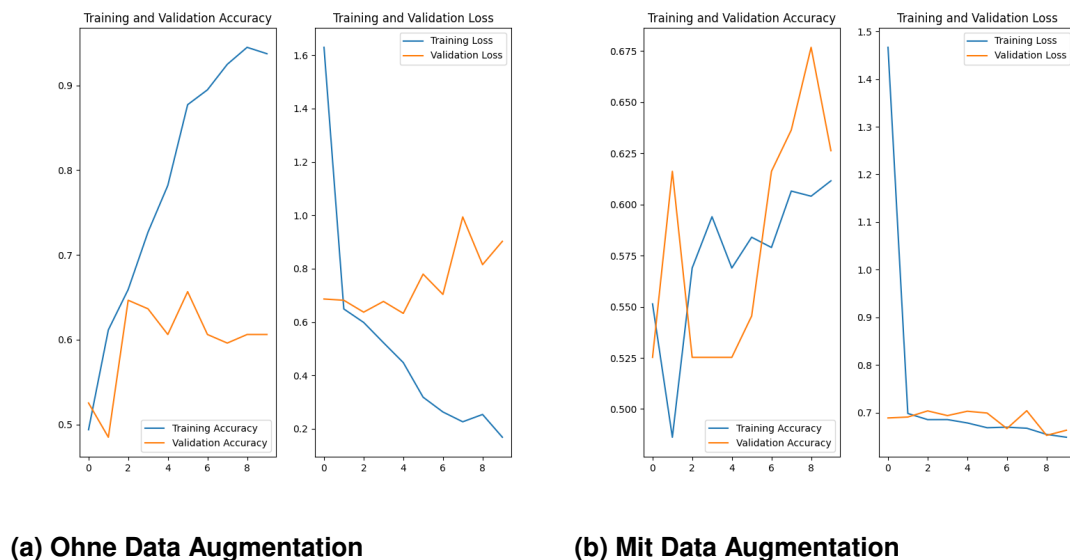
³² doccano-2018.

³³ ghosh2019.

3.4.3 Training des Modells

Nachdem die Daten, wie in Kapitel 3.4.2 beschrieben, verarbeitet wurden, können Sie im weiteren Verlauf für das Training eines Modells verwendet werden. Dazu wurden die Daten wie im Deep Learning üblich, in Train und Validation Sets aufgeteilt um einen späteren Vergleich auf unbekannten Daten zu ermöglichen. Aufgrund der geringen Menge der Eingangsdaten wurde *Data Augmentation* verwendet um durch leichte Mutationen der Bilder wie etwa Zooming oder eine vertikale Spiegelung Overfitting zu reduzieren.

Abbildung 7: Vergleich der Trainings- und Validierungsmetriken mit und ohne Data Augmentation



Quelle: Eigene Darstellung

Abbildung 7 stellt jeweils die Trainings- und Validierungsmetriken für die *Accuracy* und den *Loss* gegenüber für das Training mit und ohne erweiterte Daten. Dabei gilt, dass die *Accuracy* gegen eins gehen sollte und der *Loss* gegen null. Es ist erkennbar, dass die Graphen in beiden Abbildungen relativ früh auseinander gehen, was darauf hindeutet, dass ...

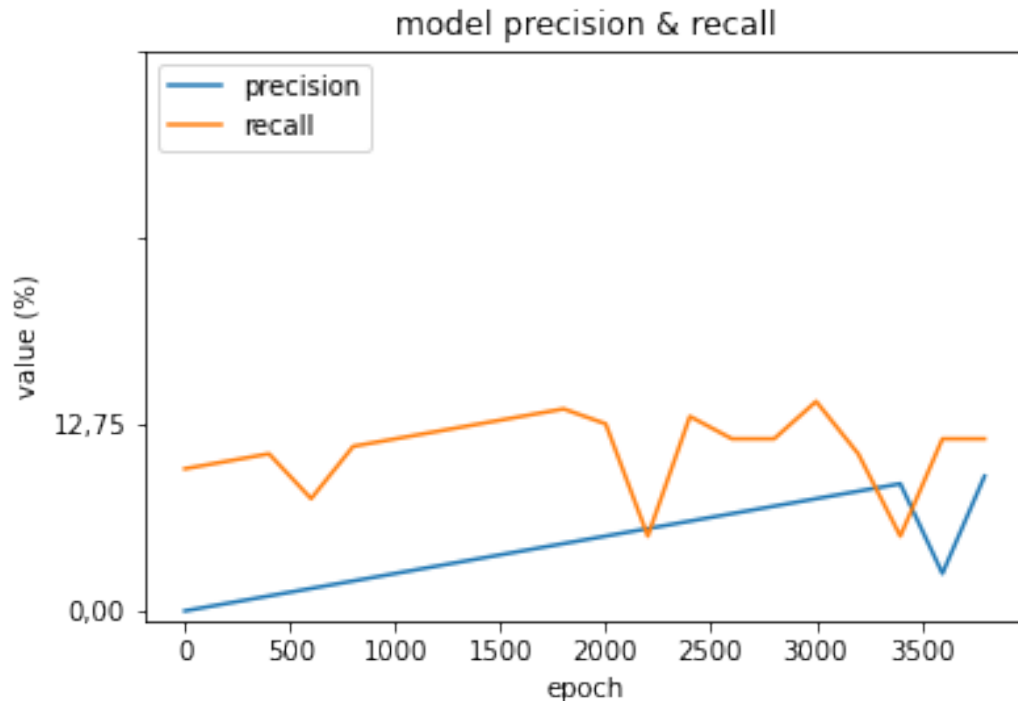
Dennoch ist erkennbar, dass die *Data Augmentation* einen klaren Effekt auf den Trainingsverlauf hat. In Abbildung 7b bleiben die Training und Validation Accuracy über alle Epochen hinweg nah beieinander. In Abbildung 7a gehen beide Kurven nach der zweiten Epoche auseinander. Dies deutet darauf hin, dass das Modell auf den Trainingsdatensatz overfitted und demnach nicht die richtigen Features der Bilder erlernt.

Insgesamt muss gesagt werden, dass auch mit Data Augmentation die erzielten Genauigkeiten und Losses weit von einem idealen Ergebnis entfernt liegen. So ist die beste erzielte Genauigkeit mit 67.5% für ein Klassifizierungsmodell mit lediglich zwei Labels noch sehr gering und würde wahrscheinlich auf Basis der geringen Menge an Trainingsdaten schlecht generalisieren beziehungsweise auf andere anwendbar sein.

3.4.4 NER-Modell

Um den OSMI-Index automatisiert bewerten zu können ist es unter anderem notwendig die textbasierten Inhalte einer Website auf sensorische Inhalte prüfen zu können. Hierfür eignet sich ein NER-Modell, welches die sensorischen Wörter bzw. Begrifflichkeiten als eigene Entität erkennt am meisten. Da das Anlernen eines NER-Modells über ein Supervised Learning Verfahren vorgenommen wird, bedarf es einer Datengrundlage, die vorannotierte Textpassagen enthält auf Basis dessen das Modell trainiert werden kann. Die Datengrundlage wurde bereits aus vergangenen Projekten durch die Extraktion von Rohtexten der jeweiligen Produktwebseiten vorgenommen. Im Rahmen dieses Consulting-Projektes wurden die Rohdaten genutzt, um diese vorzufiltern und anschließend innerhalb von Doccano zu labeln. Die gelabelten Textpassagen stellten anschließend die Datengrundlage für das Training des NER-Modells dar. Um es den Annotierern im Vorfeld einfacher zu machen, wurden die Daten vorgefiltert auf mögliche Sensory Words.

Anschließend wurden die annotierten Daten aus Doccano heraus exportiert und in das Trainingsformat für Spacy transformiert. Das NER-Modell wurde im nächsten Schritt auf der Datenbasis trainiert. Die Ergebnisse des Modelltrainings sind in Abbildung 8 dargestellt.

Abbildung 8: Ergebnisse des NER-Modell-Trainings

Quelle: Eigene Darstellung

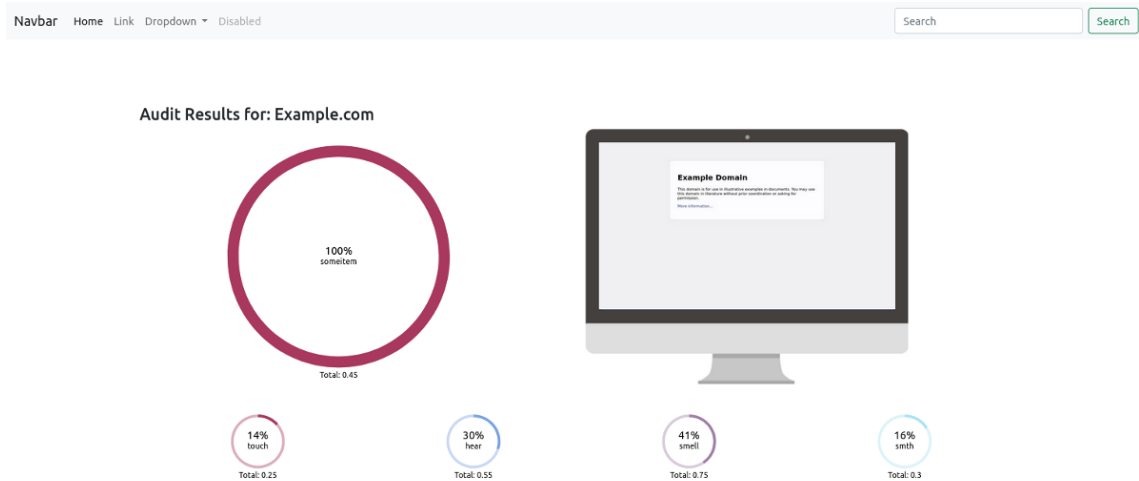
Hier sind die Metriken *Accuracy* und *Recall* je Trainingsepoche abgetragen. Je besser das Modell trainiert ist, desto mehr nähern sich beide Metriken einem Wert von eins bzw. einhundert Prozent an. Die Abbildung zeigt, dass sowohl die eine als auch die andere Metrik weit vom Optimalwert entfernt sind, was für eine geringe Modellgüte spricht.

3.5 Dashboard zur Darstellung der Ergebnisse

Nachdem das Modell zur Image Classification die relevanten Produktseiten einer Webseite identifiziert hat und das NER-Modell alle relevanten Textpassagen auf die fünf Sinnesentitäten geprüft und diese ausgezählt hat, werden die Ergebnisse in einem Dashboard ausgegeben und können entsprechend bewertet werden. Dieses zeigt die Anzahl der erkannten Entitäten aufgeteilt auf die fünf Sinne an und vergleicht diese mit dem Branchendurchschnitt. Um den jeweiligen Branchendurchschnitt ermitteln zu können, muss vorab eine entsprechende Datenbasis aufgebaut werden, indem vorab Webseiten wie zu Anfang des Kapitels aufgezeigt durch den Image Classifier selektiert und durch das NER-Modell bewertet werden. Die Ergebnisse sind jeweils inklusive der Brancheninformationen

zu speichern. Abbildung 9 zeigt einen möglichen Prototypen, der im Rahmen des Consulting Projektes entwickelt wurde.

Abbildung 9: Dashboard Prototyp



Quelle: Eigene Darstellung

4 Fazit

Das Consulting-Projekt hatte zum Ziel einen Beitrag zur automatisierten und objektiven Quantifizierung des OSMI-Index zu leisten. Hierfür wurde das in Kapitel 3.1 vorgestellte Zielbild erarbeitet und entsprechend umgesetzt. Zum Abgabestand der vorliegenden Projektarbeit sind die einzelnen Ansätze und Modelle entwickelt worden. Es wurde sowohl der Image Classifier zur Identifikation der relevanten Produktseiten programmiert als auch das NER-Modell entwickelt, welches die sensorischen Entitäten aus den Texten heraus extrahieren kann. Darüber hinaus wurde der Prototyp einer Webpage entworfen, welcher die Ergebnisse der Analyse für den Endnutzer zugänglich machen soll indem sie die zu bewertende Website im Vergleich zum Branchenschnitt und dem optimalen OSMI-Wert der Branche darstellt.

Die größte Programmierarbeit des Zielbildes wurde im Rahmen des Consulting-Projektes durch die Autoren dieser Projektarbeit geleistet. Die zur Verfügung stehenden Trainings-

und Testdaten sind jedoch nicht ausreichend genug gewesen, um präzise Modelle zu entwickeln. Das heißt in Bezug auf den Image Classifier, dass dieser Produktseiten von Nicht-Produktseiten mit rund 60% Genauigkeit unterscheidet, was nur unwesentlich besser ist als die Chancen eines Münzwurfes für „Kopf“ oder „Zahl“. Das NER-Modell ist vergleichsweise sogar noch weniger genau. Es erreicht eine *Accuracy* von **15%**, wodurch es in diesem Zustand noch nicht für einen produktiven Einsatz verwendbar ist. Zwar standen der Projektgruppe für beide Modelle ausreichend viele Rohdaten zur Verfügung, jedoch wurden sie nicht alle annotiert. Dafür standen im Rahmen der Projektdauer zu zeitlich zu wenige Ressourcen zur Verfügung.

Was ist noch offen? Mehr Trainingsdaten zur Optimierung der Modelle Ende-zu-Ende Integration der Pipeline, um das Zielbild zu komplementieren Funktion zur Berechnung des OSMI-Indexes auf Basis unserer Modellinputs Ausblick / Entwicklungspotential: Speech-to-Text-Implementierung, um Sound und Videodateien verarbeiten zu können

Dem Anspruch der vollständig automatisierten Erfassung des OSMI konnte die Arbeit ebenfalls nicht ganz gerecht werden. So existieren alle dafür notwendigen Teile wie das Scraping der Webseite,³⁴ die Bildklassifizierung und die Named Entity Recognition des Inhalts sowie die Darstellung der Ergebnisse,^{35,36} jedoch wurden sie nicht miteinander integriert.

In dieser Arbeit wurden Text und Bilder zur automatisierten Berechnung des OSMI-Indexes verwendet. Was jedoch nicht berücksichtigt wurde, sind dynamische Medien wie Video und Ton. Diese sind ebenfalls für den OSMI relevant und könnten daher in zukünftigen Arbeiten betrachtet werden um zur Berechnung des OSMI den vollständigen Inhalt einer Webseite zu betrachten.

³⁴ ostkamp2022a.

³⁵ ostkamp2022b.

³⁶ ostkamp2022c.

Anhang

Ehrenwörtliche Erklärung

Hiermit versichere ich, dass die vorliegende Arbeit von mir selbstständig und ohne unerlaubte Hilfe angefertigt worden ist, insbesondere dass ich alle Stellen, die wörtlich oder annähernd wörtlich aus Veröffentlichungen entnommen sind, durch Zitate als solche gekennzeichnet habe. Ich versichere auch, dass die von mir eingereichte schriftliche Version mit der digitalen Version übereinstimmt. Weiterhin erkläre ich, dass die Arbeit in gleicher oder ähnlicher Form noch keiner Prüfungsbehörde/Prüfungsstelle vorgelegen hat. Ich erkläre mich damit **einverstanden/nicht einverstanden**, dass die Arbeit der Öffentlichkeit zugänglich gemacht wird. Ich erkläre mich damit einverstanden, dass die Digitalversion dieser Arbeit zwecks Plagiatsprüfung auf die Server externer Anbieter hochgeladen werden darf. Die Plagiatsprüfung stellt keine Zurverfügungstellung für die Öffentlichkeit dar.

Münster, 30.8.2022

(Ort, Datum)

A handwritten signature in black ink, consisting of a large, stylized 'H' followed by a series of loops and a final flourish.

(Eigenhändige Unterschrift)