

Gottfried Wilhelm
Leibniz Universität Hannover
Fakultät für Elektrotechnik und Informatik
Institut für Praktische Informatik
Fachgebiet Software Engineering

Untersuchung des Einflusses von
Domänenwissen auf den
Erklärungsbedarf der Nutzenden von
Software

Bachelorarbeit

im Studiengang Informatik

von

Jonas Bohnstedt

Prüfer: Prof. Kurt Schneider

Zweitprüfer: Dr. Jil Klünder

Betreuer: Martin Obaidi

Hannover, 27.03.2024

Erklärung der Selbstständigkeit

Hiermit versichere ich, dass ich die vorliegende Bachelorarbeit selbständig und ohne fremde Hilfe verfasst und keine anderen als die in der Arbeit angegebenen Quellen und Hilfsmittel verwendet habe. Die Arbeit hat in gleicher oder ähnlicher Form noch keinem anderen Prüfungsamt vorgelegen.

Hannover, den 27.03.2024

Jonas Bohnstedt

Zusammenfassung

Die zunehmende Digitalisierung beeinflusst unseren Alltag immer mehr. Bei vielen ist sie seit Jahren am Arbeitsplatz, in der Freizeit oder in der Kommunikation tief verankert. Dabei gewinnt effektive Kommunikation zwischen der Informationsflut zunehmend an Bedeutung. Dies gilt vor allem für Software, die bei Personen für wichtige Entscheidungen sorgen und wo das Verständnis des Erklärungsbedarfs entscheidend ist, um ein möglichst gelungenes Nutzererlebnis zu gewährleisten.

In dieser Bachelorarbeit wurde eine Onlinestudie im Bereich der Erklärbarkeit von Software durchgeführt. Dabei wurde abgefragt, welche Software die Teilnehmer häufig nutzen, wie sie ihr Domainwissen einschätzen und wie häufig und wie tief sie Erklärungen für ihre genannte Software wünschen. Zur Erstellung der Onlinestudie, wurde ein Konzept erarbeitet, welches den systematischen Erhalt von Erklärungsdetailliertheit und dem Domänenwissen der Nutzergruppen ermöglicht. Die Onlinestudie fokussiert sich primär auf die Abfrage von subjektivem und objektivem Domänenwissen, die gewünschte Erklärungshäufigkeit sowie die Detailliertheit von Erklärungstexten in Softwaresystemen. Um Zusammenhänge zwischen den Umfrageergebnissen und dem Erklärungsbedarf zu identifizieren, wurde eine Korrelations- und eine Clusteranalyse durchgeführt. Beide Analysen ergaben keinen starken Zusammenhang zwischen Domänenwissen und Erklärungsbedarf. Der Erklärungsbedarf wurde dabei in fünf verschiedene Typen eingeteilt. Diese Typen zeigen, dass der Erklärungsbedarf entweder von Domäne zu Domäne unterschiedlich entsteht, oder Personen einen grundlegend gleichen Erklärungsbedarf haben. Durch die Korrelations- und Clusteranalysen ist aufgefallen, dass das Domänenwissen sich von den demografischen Daten beeinflussen lässt. Und dass mit dem Domänenwissen die Sicherheit in der Anwendung steigt. Eine zusätzliche Erkenntnis ist, dass die Teilnehmer Erklärungstexte mit wenig detaillierten Erklärungen als Erklärungsform priorisieren.

Abstract

The increasing digitization is exerting a growing influence on our daily lives. For many, it has been deeply embedded in the workplace, leisure activities, or communication for years. Effective communication amidst the deluge of information is becoming increasingly crucial. This is particularly true for software that aids individuals in making important decisions, where understanding the need for explanations is crucial to ensuring the best possible user experience.

This bachelor thesis conducted an online study in the realm of explainability in software. It queried participants on the software they frequently use, their assessment of their domain knowledge, and how often and how deeply they desire explanations for their mentioned software. To create the online study, a concept was devised to systematically capture the level of explanation detail and the domain knowledge of user groups. The study primarily focuses on eliciting subjective and objective domain knowledge, desired frequency and detail of explanations in software systems. To identify correlations between survey results and the need for explanations, correlation analysis and cluster analysis were performed. Neither analysis revealed a connection between domain knowledge and the need for explanations. The need for explanations was categorized into five different types. These types indicate that the need for explanations either varies across domains or that individuals share a fundamentally similar need for explanations. Through correlation and cluster analysis, it was noticed that domain knowledge can be influenced by demographic data and that application security increases with domain knowledge. An additional insight is that participants prioritize explanation texts with less detailed explanations as a form of explanation.

Inhaltsverzeichnis

1	Einleitung	1
1.1	Problemstellung	1
1.2	Lösungsansatz	2
1.3	Struktur der Arbeit	2
2	Grundlagen	3
2.1	Erklärbarkeit	3
2.2	Detailliertheit und Erklärungstexte	5
2.3	Domänenwissen	6
2.4	Korrelationsanalyse	7
2.5	Clusteranalyse	9
2.6	Python-Bibliotheken	11
3	Verwandte Arbeiten - Abgrenzung	13
3.1	Verwandte Arbeiten	13
3.2	Abgrenzung zu verwandten Arbeiten	14
4	Studiendesign	15
4.1	Konzept	16
4.1.1	Herangehensweise	16
4.1.2	Online-Studie	16
4.2	Analyseaufbau	17
4.2.1	Datenvorbereitung	17
4.2.2	Forschungsziele und Fragen	18
4.2.3	Hypothesenübersicht	19
4.2.4	Variablenübersicht	22
4.2.5	Cluster	22
4.3	Abfrage allgemeine Meinung der Teilnehmer	23
5	Ergebnisse	25
5.1	Randdaten	25
5.2	Ergebnisse der Korrelationsanalyse	27
5.3	Ergebnisse der Clusteranalyse	29

6	Diskussion	35
6.1	Beantworten der Forschungsfragen	35
6.2	Interpretation	37
7	Validität	39
8	Zusammenfassung und Ausblick	41
8.1	Zusammenfassung	41
8.2	Ausblick	42
A	Ein Anhang	43
A.1	Clusteranalyse	43
A.1.1	Silhouettenkoeffizienten	43
A.1.2	Dendrogramme	45
A.1.3	Streudiagramme	46

Kapitel 1

Einleitung

Die zunehmende Digitalisierung hat dazu geführt, dass hochwertige Software ein zentraler Baustein des digitalen Wandels geworden ist [11]. Das Digitale durchdringt alle Lebensbereiche und verursacht dadurch neue ingenieurtechnische Herausforderungen [11]. Dabei entsteht die Frage, wie die Qualität einer Software zu definieren ist [11]. Das Forschungsfeld der Erklärbarkeit ist eines von vielen Bereichen des Software-Engineerings und beschäftigt sich unter anderem mit dieser Frage in Bezug auf Qualität [11]. Dazu gehören die Selbsterklärbarkeit, Effektivität sowie viele weitere Qualitätsaspekte [8]. Die Identifizierung von Erklärungsbedarfen der Nutzer hinsichtlich der entsprechenden Erklärbarkeitsanforderungen ist als Grundlage dessen zu sehen [3]. Hauptbestandteil dieser Arbeit ist, beeinflussende Variablen und Faktoren auf die gewünschte Erklärungstiefe von Nutzern von Software zu erkennen. Eine eigens durchgeführte Online-Studie dient als Grundlage zur Analyse, wie häufig und in welcher Tiefe Erklärungen zu Software gewünscht werden. Die Studie zielt darauf ab, Einblicke in die Bedürfnisse der Nutzenden zu gewinnen und insbesondere besser abzuschätzen, wie detailliert Erklärungen für Softwarefunktionalitäten mit und ohne Domänenwissen bereitzustellen sind. Dieses Forschungsvorhaben positioniert sich im Kontext von Explainability [9] [8] und Requirements Engineering [3] im Software-Engineering.

1.1 Problemstellung

Erklärungsbedarf entsteht für jeden Nutzer individuell an anderen Stellen im System und kann sich stärker oder schwächer ausprägen [22]. Die Identifizierung des Ursprungs ist somit für die Verbesserung der Benutzerfreundlichkeit relevant [10]. Nach einer Analyse soll klarer werden, ob die Erklärungen in Softwaresystemen ausreichen, oder ob sie zu undetailliert sind. Ebenso wird festgestellt, ob Domänenwissen bei Nutzern förderlich ist und wie dies die Erklärbarkeit in verschiedenen Detaillierungsgraden beeinflusst.

1.2 Lösungsansatz

Für die Untersuchung wurde eine Onlineumfrage erstellt, in der die Teilnehmer einschätzen sollten, ob sie viel oder wenig Domänenwissen zu zwei Domänen haben. Da die Selbsteinschätzung als subjektive Herangehensweise zu uneindeutig wäre, wird genauso das möglichst objektive Domänenwissen der Teilnehmer ermittelt. Zusätzlich wurden weitere Einschätzungen abgefragt, um mehrere Faktoren und Variablen für die Auswertung zu sammeln. Wie subjektive, objektive, demografische und Erfahrungswert-Daten der Teilnehmer aus den jeweiligen Domänen zusammenhängen, wird anschließend mit einer Korrelationsanalyse untersucht. Im Fokus steht die Überprüfung des Zusammenhangs von subjektivem und objektivem Domänenwissen in Bezug auf den Erklärungsbedarf. Zudem wird eine Clusteranalyse durchgeführt, um Datengruppierungen in den Umfragedaten zu identifizieren. Die Clusteranalyse stellt somit den dynamischen Teil dar. Es sollen Cluster in jedweder Kombination aus den Variablen hinsichtlich der gewünschten Erklärungstiefe gefunden werden.

1.3 Struktur der Arbeit

Im Folgenden wird die Struktur der Arbeit kurz zusammengefasst:

In Kapitel 2 werden alle benötigten Konzepte und genutzten Methoden erklärt. Das Kapitel 3 beschäftigt sich mit den verwandten Arbeiten, welche ähnliche Literatur, Arbeiten und Paper zu den Themen Erklärbarkeit, Korrelation und Clusteranalyse behandeln. Ebenfalls erfolgt eine Abgrenzung zu verwandten Arbeiten, die sich mit der Erklärbarkeit im KI Bereich beschäftigen. In Kapitel 4 wird die Durchführung der Online-Studie und die Clusteranalyse behandelt. Zudem werden die Forschungsfragen aufgestellt, welche in der Korrelationsanalyse benötigt werden. Das folgende Kapitel 5 beschäftigt sich mit den Ergebnissen aus beiden Analysen und behandelt Randdaten aus der Umfrage. Das Kapitel 6 beantwortet die Forschungsfragen mit den gewonnenen Ergebnissen aus Kapitel 5 und interpretiert diese. Im 7 Kapitel werden die möglichen *Threats to Validity* [37] betrachtet. Im letzten Kapitel 8 kommen alle Ergebnisse nochmals zusammen. Des Weiteren wird ein Ausblick auf die Zukunft im Kontext dieser Arbeit gegeben.

Kapitel 2

Grundlagen

In den Grundlagen werden Begriffe wie Erklärbarkeit, Erklärungsbedarf, Erklärungstexte, Detailliertheit und Domänenwissen erläutert. Zudem werden die genutzten Methoden zur Auswertung des Umfrage-Datensatzes erklärt. Abschließend werden alle genutzten Python-Bibliotheken für die Clusteranalyse aufgelistet.

2.1 Erklärbarkeit

Die Erklärbarkeit ist sowohl Teilgebiet des Requirement Engineerings [3]. Dabei bezieht man sich auf Erklärungstexte als auch das Verhalten eines Softwaresystems. Der Erklärungsbedarf, der in dieser Arbeit Hauptbestandteil ist, leitet sich aus diesem Forschungsfeld ab und beschreibt die Ab- und Anwesenheit des Verständnisses von Software(-Funktionen). Dabei bezieht man sich nicht nur auf den Nutzer. Ihm soll das Verhalten einer Software erklärt werden. Den Entwicklern muss die Resultatermittlung klar werden. In Bezug auf die Gesellschaft wird Erklärbarkeit für die ethische Vertretbarkeit benötigt. Generell betrachtet man den Input, das Training und Gesetze, die Rahmenbedingungen bilden. Überall benötigt es klar definierte Erklärungen über die Durchführung, das Verhalten und die Ein- und Ausgabe.

Erklärungsbedarf kann entstehen, wenn dem Nutzer etwas unklar ist oder die Software (technische) Probleme aufweist. Dies lässt sich in verschiedene Abstufungen unterteilen, welche sich dem jeweiligen Betrachtungspunkt annähern. Gründe dafür können folgende sein:

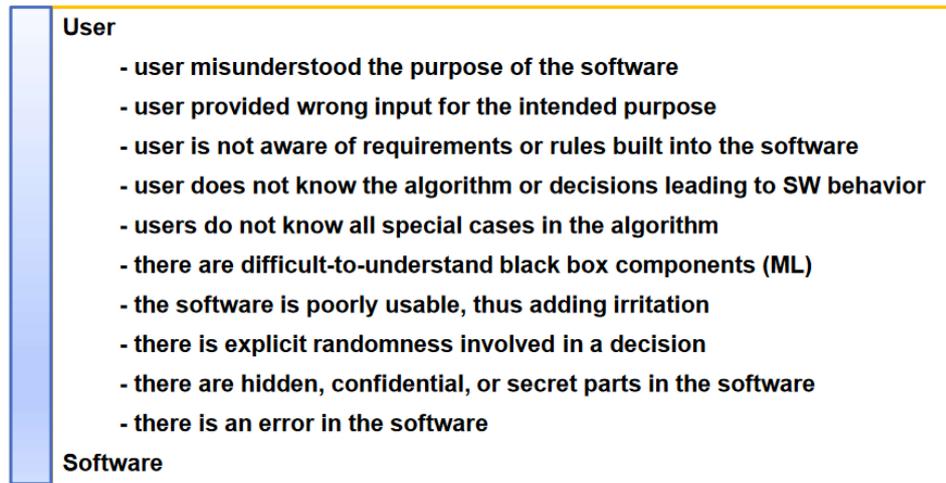


Abbildung 2.1: Gründe von Erklärungsbedarf aus dem Antrag zu softXplain²

Die Erklärbarkeit wird auch oft im KI-Kontext als Interpretierbarkeit gesehen. Diese Arbeit grenzt sich dagegen ab, 3.1 da hier die SW-Explainability in Bezug auf den Nutzer im Fokus steht.

Aus dem Requirements Engineering sind funktionale [25] und nicht-funktionale Anforderungen (non functional requirements: NFRs) bekannt [5]. Die Erklärbarkeit wird dabei den nicht-funktionalen Anforderungen zugeordnet. Dies bedeutet, es muss zuerst ein System existieren, das man auf Erklärbarkeit testen kann. Erklärbarkeit kann sich zudem negativ, neutral oder positiv auf NFRs auswirken. Im Folgenden wird nach Einschätzung von gefundener Literatur und Experten ein Spannungsfeld zwischen Erklärbarkeit und NFRs aufgelistet.

	Literatur	Experte
Komplexität		-
Privatsphäre	+	-
Sicherheit	+/-	-
Testbarkeit	+	

Tabelle 2.1: Spannungsfeld mit anderen NFRs [4]

Für die Messbarkeit von Erklärbarkeit werden Qualitätsaspekte und Metriken benötigt [8]. Damit ein zufriedenstellendes Ergebnis erfolgt, ist es wichtig, die eigentlich gesetzten Ziele zu berücksichtigen. Dies kann durch die GQM (Goal Question Metric) [34] Herangehensweise geschehen. Die Ziele

²softXplain: Anforderungen für selbst-erklärende Software
<https://gepris.dfg.de/gepris/projekt/470146331>

sind dabei die priorisierten Q-Aspekte, wovon die Sicherheit und Nutzbarkeit nicht unter der Erklärbarkeit leiden sollten.

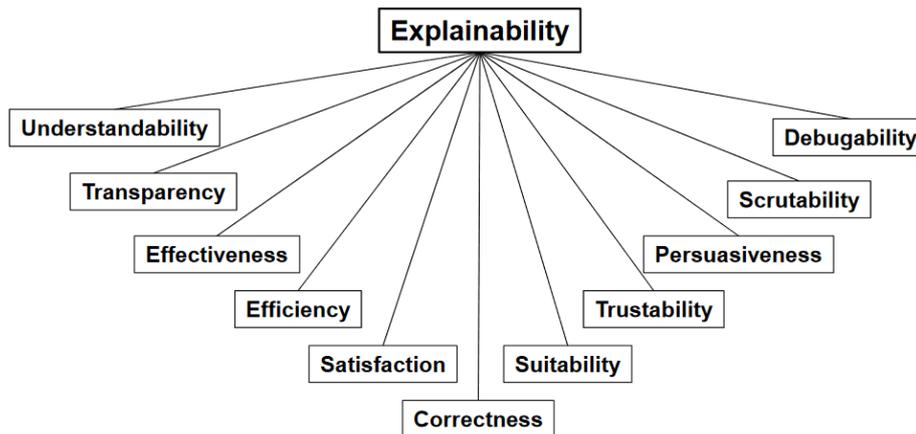


Abbildung 2.2: Q-Aspekte für Erklärbarkeit [7]

2.2 Detailliertheit und Erklärungstexte

Im Zentrum dieser Arbeit stehen der Begriff Erklärungsbedarf und die Forschung auf Zusammenhänge mit Domänenwissen, oder anderen Variablen. Dazu gehören viele Begrifflichkeiten, die im Verlauf immer wieder erwähnt werden. Dieses Kapitel dient der Klärung und Definition dieser Begriffe. Vor der Definition wird klargestellt, dass sich diese auf Erklärungstexte beziehen. Jene Texte kommen überall in Software vor und sind die grundlegende Verständnisübermittlung einzelner Funktion. Sie dienen ebenfalls in Form von Schrift zur Definition eines Verhaltens von Software. Dies kann die Beschreibung eines Buttons sein oder die Übersicht über den Programmablauf.

Erklärungstexte unterscheiden sich dabei in ihrer Detailliertheit. Nicht nur die Länge des Erklärungstextes beschreibt die Detailliertheit, sondern auch die Art und die Fülle der Informationen. Informationen über die Funktion oder das Verhalten können dabei sehr oberflächlich, oder tiefgehend vorhanden sein. Dafür habe ich fünf Stufen definiert. Sehr wenig detaillierte Erklärungstexte beinhalten nur grundlegende Informationen, die möglicherweise nicht ausreichend, um eine Funktion vollständig zu verstehen. Dafür sind sie kompakt und sollen nur das Nötigste beinhalten. Bei wenig detaillierten Erklärungen werden einige zusätzliche Informationen angeboten, welche bereits erste Tipps enthalten können. Jedoch können noch Verständnislücken bestehen. Detaillierte Erklärungstexte bieten dabei eine

weitaus umfassendere Beschreibung und können Beispiele mit einbinden. Viel detaillierte Texte begeben sich erstmals in beträchtlicher Tiefe der Software. So werden technische Details über Funktionen und Verhalten mitgeteilt. Und sehr detaillierte Erklärungstexte vertiefen dieses technische Wissen. Sie geben eine ausführliche Beschreibung über die gesamte Funktion, erklären Schnittstellen, weisen auf Erweiterungen und Anpassungen hin oder beinhalten viele Fachbegriffe. So kann aus einem simplen Erklärungstext „Googles Software Gemini nutzt KI für Formulierungen von Texten“ eine weitaus höhere und tiefgehende, definierte Erklärung werden „Goggles KI-Software Gemini nutzt LLMs (Large Language Models) und ist das erste Modell, was mit einem Benchmark Score von 90,0 % menschliche Experten in MMLU (Massive Multitask Language Understanding) schlägt“ [28]. Zusammenfassend lässt sich sagen, dass Erklärungsbedarf bzw. dessen Bildung aus der gewünschten Erklärungstiefe resultiert. Das Projekt softXplain³, welches den Hintergrund dieser Arbeit beschreibt, thematisiert dabei folgendes. Es ist wichtig, eine Software entsprechend selbsterklärend zu gestalten. Zudem müssen nach der Identifizierung von Erklärungsbedarf dafür entsprechende Erklärbarkeitsanforderungen sowie Erklärungen passend formuliert werden.

2.3 Domänenwissen

Domänenwissen bezieht sich auf spezifisches Fachwissen oder Expertise in einem bestimmten Bereich oder einer Disziplin, das es einer Person ermöglicht, fundierte Entscheidungen zu treffen, Probleme zu lösen oder in diesem speziellen Bereich effektiv zu agieren [29]. Im Softwarekontext ist Domänenwissen das Verständnis und die Kenntnis über spezifische Branchen, Problembereiche oder Fachgebiete, die von der Software adressiert werden soll. Im Rahmen dieser Arbeit wird Domänenwissen in fünf Abstufungen skaliert. Sehr geringes Domänenwissen beschreibt eine Person mit einer minimalen Verbundenheit und Erfahrung mit der Software. Geringes Domänenwissen ist bei Anwendern mit wenig Erfahrung, aber bereits ersten wenigen Erfolgen vorhanden. Durchschnittliches Domänenwissen betrifft die meisten Personen in Bezug auf einen Bereich oder eine Software. Dabei ist es ihnen möglich, die Software für ihre Zwecke gut zu nutzen. Wenn es allerdings um tiefgehendere Funktionen geht, dann müssen Personen dieser Einstufung häufig auf externe Hilfen zurückgreifen. Ein hohes Domänenwissen besteht bei Personen mit viel Erfahrung und Nutzungszeit. Dabei ist es ihnen möglich, komplexere Aufgaben mit der Software erfolgreich alleine zu erledigen. Sehr hohes Domänenwissen ist definiert durch einen Nutzer oder Entwickler, der die Funktionen der Software nahe zu in- und auswendig

³softXplain Nutzerstudie:

<https://www.pi.uni-hannover.de/de/se/forschung/projekte/forschungsprojekte-detailansicht/projects/softxplain>

kennt und diese mit Leichtigkeit bedienen kann. Er trifft so gut wie nie auf Probleme und ist in der Lage, dem durchschnittlichen Nutzer zu helfen.

2.4 Korrelationsanalyse

Mit der Korrelationsanalyse sollen Korrelationen zwischen den exportierten Umfragedaten gefunden werden. Dabei ist die Korrelation eine statistische Maßeinheit, welche aussagt, wie zwei unterschiedliche Variablen zueinander in Verbindung stehen [12]. Unter den Korrelationsanalysen gibt es verschiedene Arten. In dieser Arbeit werden primär der Pearson Chi²-Test[39] (Kontingenzanalyse) und die Spearman Korrelation genutzt [31]. Der Chi²-Test bildet eine Kreuztabelle aus zwei zu vergleichenden Variablen. Das Ergebnis der Spearman Korrelation liegt in einem Wertebereich von -1 bis +1 [20]. Wenn der Wert -1 ist, steht dies für eine perfekt negative Korrelation. Dies bedeutet, dass, wenn der Wert der einen Variable steigt, dass der Wert der anderen Variable sinkt[20]. Falls der Wert wiederum +1 ist, steht dies für eine perfekt positive Korrelation, was bedeutet, dass wenn der Wert der einen Variable steigt, sich auch der Wert der anderen Variable erhöht[20]. Bei einem Korrelationskoeffizienten von 0 existiert keine Korrelation [20]. Für beide Korrelationen sind einzelne Vorbedingungen zu beachten. Der Datensatz eines Pearson Chi²-Test muss aus kontinuierlichen kategorialen (nominal- oder ordinalskaliert) Daten bestehen[39]. Der Datensatz der Spearman-Korrelationsanalyse besteht aus ordinalen oder metrisch skalierten Daten, zwischen welchen ein linearer Zusammenhang vorliegt [20]. Hier ist ein monotonen Wachstum ausreichend [20]. Bei dem Pearson Chi²-Test (X^2) wird die Teststatistik durch folgende Formel berechnet[39]:

$$X^2 = \sum_{i=1}^I \sum_{j=1}^J \frac{(f_{ij} - f'_{ij})^2}{f_{ij}}$$

mit

f_{ij} = beobachte Häufigkeit in einer Zelle

f'_{ij} = erwartete Häufigkeit in einer Zelle

i = Laufindex über Spalten (I = Anzahl Spalten)

j = Laufindex über Zeilen (J = Anzahl Zeilen)

Die Prüfung, ob die Abweichung zwischen beobachteten und erwarteten Häufigkeiten signifikant ist, lässt sich aus der Chi²-Verteilung deuten. Anhand einer Tabelle mit kritischen Werten⁴ wird durch die Freiheitsgrade (df) abgelesen, welchen Schwellenwert X^2 übersteigen muss, um statistisch signifikant zu sein. Die Freiheitsgrade berechnen sich wie folgt[39]:

⁴Chi-Quadrat-Verteilung: https://www.uibk.ac.at/econometrics/einf/tab_chisq.pdf

$$df = (I - 1) \cdot (J - 1)$$

Um die Stärke des Zusammenhangs zu deuten, wird Cramers V und/oder der Kontingenzkoeffizient CC verwendet. Der Wert von beiden liegt zwischen 0 und 1 [39]. Cramers V kann für jegliche Tabellengrößen verwendet werden, was für den Datensatz der Umfrageergebnisse in dieser Arbeit benötigt wird [39]. Cramers V berechnet sich wie folgt [39]:

$$V = \sqrt{\frac{X^2}{n \cdot \min(I-1, J-1)}} \quad \text{mit } n = \text{Stichprobengre}$$

Um die Stärke des Zusammenhangs mittels des V-Werts zu deuten, wird folgende Tabelle verwendet:

Betrag von V	Stärke des Zusammenhangs
$0,00 \leq V < 0,10$	kein Zusammenhang
$0,10 \leq V < 0,30$	geringer Zusammenhang
$0,30 \leq V < 0,50$	mittlerer Zusammenhang
$0,50 \leq V < 0,70$	hoher Zusammenhang
$0,70 \leq V < 1,00$	sehr hoher Zusammenhang

Tabelle 2.2: Cramers V Zusammenhangstärke [20]

Der Kontingenzkoeffizient berechnet sich wie folgt [19]:

$$CC = \sqrt{\frac{X^2}{X^2 + n}}$$

$$CC_{\max} = \sqrt{\frac{M-1}{M}} \quad \text{mit } M = \min(I, J)$$

$$CC_{\text{norm}} = CC_{\max} \cdot CC$$

Die Spearman (Rang-)Korrelation ist eine Pearson Korrelation, bei der der Variablenwert x_i, y_i nicht direkt aus dem Datensatz mit in die Rechnung einfließt, sondern zunächst der jeweilige Rang r_{x_i}, r_{y_i} bestimmt wird [31]. Diese Ränge können auf- oder absteigend und mit oder ohne doppelt vorkommenden Werten gelistet werden [31]. Die Formel zur Berechnung der Spearman Korrelation sieht wie folgt aus [31]:

$$r_s = \frac{\sum_{i=1}^{n_r} (r_{x_i} - \bar{r}_x)(r_{y_i} - \bar{r}_y)}{\sqrt{\sum_{i=1}^{n_r} (r_{x_i} - \bar{r}_x)^2} \sqrt{\sum_{i=1}^{n_r} (r_{y_i} - \bar{r}_y)^2}}$$

Ob eine gefundene Spearman Korrelation wirklich signifikant ist, hängt neben dem eigentlichen Korrelationskoeffizienten r_s ebenfalls von der verwendeten Stichprobengröße n_{r_s} ab [31]. Dafür wird der t-Test mit den

Freiheitsgraden (degree of freedom) $df = n_{r_s} - 2$ durchgeführt [31].

$$t = r \sqrt{\frac{df}{1-r^2}}$$

Der Signifikanzwert p kann auf Basis der zweiseitigen t-Verteilung berechnet werden [31] [32]. Falls der p-Wert darunter liegt, ist die Korrelation auf dem Niveau $\alpha = 0,05$ (2-seitig) signifikant oder $\alpha = 0,01$ hochsignifikant [21]. Für den Korrelationskoeffizienten r_s stellten Kendall and Smith [18] eine Methode auf, aus der sich die statistische Signifikanz ermitteln lässt. Mittels der Kritische-Werte-Tabelle $r_s(\alpha, n)$ kann zusätzlich für die Signifikanz geprüft werden, ob $|r_s| \geq r_s(\alpha, n)$ ist ⁵. Wenn $|r_s|$ über $\alpha = 0,05$ und/oder $\alpha = 0,01$ liegt, ist eine Korrelation nach dem Niveau von α statistisch signifikant ⁵. Um eine Aussage über den Zusammenhang zweier Variablen bei der Spearman Korrelation treffen zu können, wird in dieser Arbeit die Effektstärke von Cohen verwendet [6][27]. Dabei ist r_s wie folgt unterteilt:

Betrag von r_s	Stärke des Effekts
$r_s < 0,10$	kein Effekt
$r_s \geq 0,10$	schwacher Effekt
$r_s \geq 0,30$	mittlerer Effekt
$r_s \geq 0,50$	starker Effekt

Tabelle 2.3: Cohen(1992) Effektstärke [6][27]

Für beide Korrelations-Methoden lassen sich die Nullhypothesen aufstellen, welche aussagen, ob zwischen zwei Variablen keine Korrelation vorliegt [21]. Ist $r_s = 0$ bzw. $X^2 = 0$ oder besteht kein Signifikanz zwischen den Variablen, wird die Nullhypothese angenommen [20]. Im Fall $r_s \neq 0$ bzw. $X^2 \neq 0$ und einer Signifikanz wird die Nullhypothese verworfen [20].

2.5 Clusteranalyse

Es gibt viele Clusteranalysemethoden, aus denen je nach Anwendungszweck gewählt werden kann[24]. In folgerndem Abschnitt wird die hierarchische Clusteranalyse behandelt. Sie ist ein unüberwachtes maschinelles Lernverfahren [26]. Dabei ist es das Ziel, zusammenhängende Strukturen und Gruppierungen von sogenannten *Clustern* in den Daten zu finden. Der Hauptgrund, warum hier die hierarchische Clusteranalyse ausgewählt wurde, ist, dass durch das unüberwachte Arbeiten die Cluster-Anzahl zu Beginn nicht bekannt sein muss [14]. Am Ende entstehen immer zwei Cluster [14]. Dabei wird zwischen dem agglomerativem und dem diversen Clustering unterschieden [14]. Agglomeratives Clustering, auch hierarchisches Clustering

⁵Statistical Table H: Critical Values for the Spearman's correlation:
<https://www.zoology.ubc.ca/bio300/StatTables.pdf>

genannt, arbeitet mit dem Bottom-Up Prinzip [26]. Jeder Punkt bildet ein eigenes Cluster. Diese werden sukzessiv von dem Fusionierungsalgorithmus bestimmenden Cluster zusammengeführt [14]. Das Verfahren nutzt dabei ein Distanzmaß zur Distanzbestimmung zwischen den Clustern [14]. In dieser Arbeit wird die euklidische Distanz dafür genutzt [14]. Der Euklidisch-Abstand ergibt sich aus der Wurzel der Summe der quadrierten Differenzen zwischen zwei Punkten [13]. Weitere übliche Distanzmaße sind die Manhattan Distanz oder die Maximumdistanz. Bei der Manhattan Distanz wird die Summe der absoluten Differenzen verwendet [13]. Die Maximumdistanz wird mit dem Maximalwert der absoluten Differenzen berechnet [13]. Anschließend wird der Fusionierungsalgorithmus, auch Linkage genannt, mit den errechneten Distanzen auf die Cluster angewendet. Dadurch wird bestimmt, welche Cluster zusammengeführt werden [14]. Es gibt drei Linkage-Methoden [14]:

- Single linkage: Misst den Abstand zwischen den nächstgelegenen Mitgliedern der Cluster.
- Complete linkage: Misst den Abstand zwischen den am weitesten entfernten Mitgliedern.
- Comparison of centroids: Misst den Abstand zwischen den Mittelpunkten der Cluster.

In der letzten Iteration bleiben die zwei angesprochenen Cluster übrig [14]. Dieser hierarchische Ablauf lässt sich in einem Dendrogramm anzeigen und ablesen [14]. Der Silhouettenkoeffizient bestimmt dabei die Stärke der gefundenen Struktur und beschreibt, welche weiteren End-Cluster-Anzahlen ebenfalls eine starke Aussagekraft hätten.

SK	Stärke
$SK \leq 0,25$	keine substanzielle Struktur gefunden
$0,25 < SK \leq 0,50$	schwache, möglicherweise künstliche Struktur gefunden
$0,50 < SK \leq 0,70$	sinnvolle Struktur gefunden
$0,70 < SK \leq 1$	starke Struktur gefunden

Tabelle 2.4: Aussagekraft des Silhouettenkoeffizienten nach Rousseeuw [30]

2.6 Python-Bibliotheken

Für die Clusteranalyse wurden folgende Bibliotheken verwendet:

pandas ⁶ wurde zum csv-Daten einlesen genutzt, sodass Python den Datensatz interpretieren kann.

sklearn ⁷ beinhaltet eine Bibliothek für maschinelles Lernen und stellt prädiktive Datenanalysen zur Verfügung. Im Kontext dieser Arbeit wurde sie für die Durchführung der Clusteranalyse und somit auch zur Berechnung des Silhouettenkoeffizient und der PCA genutzt.

scipy ⁸ stellt Optimierungs- und mathematische Algorithmen zur Verfügung. Die Bibliothek wurde in dieser Arbeit für das Erstellen der Dendrogramme genutzt.

matplotlib.pyplot ⁹ dient als Teil der matplotlib zur Erstellung von Diagrammen.

numpy ¹⁰ erweitert Python um multidimensionale Arrays.

collections ¹¹ ist eine Unterklasse von Wörterbüchern zum Zählen hashbarer Elemente. Elemente werden als Wörterbuchschlüssel und ihre Anzahl als Wörterbuchwerte gespeichert.

plotly.express ¹² wird zum Darstellen der gefundenen Cluster nach der Analyse genutzt. Da eine html-Datei erzeugt wird, können interaktiv mit dem Hovern über die einzelnen Punkte zusätzliche Informationen abgelesen werden.

⁶<https://pandas.pydata.org/>

⁷<https://scikit-learn.org/stable/index.html>

⁸<https://scipy.org/>

⁹https://matplotlib.org/stable/api/pyplot_summary.html

¹⁰<https://numpy.org/>

¹¹<https://docs.python.org/3/library/collections.html>

¹²<https://plotly.com/python/plotly-express/>

Kapitel 3

Verwandte Arbeiten - Abgrenzung

In diesem Kapitel werden verwandte Arbeiten behandelt, die Themeninhalte beinhalten, welche auch in dieser Arbeit Anwendung gefunden haben. Darunter zählt der Erklärungsbedarf, die Korrelationsanalyse und die Clusteranalyse.

3.1 Verwandte Arbeiten

Chazette et al.[4] haben eine Literaturstudie durchgeführt, um festzustellen, wie sich Erklärungsbedarf definiert. Es werden ein Katalog, eine neue Definition und ein Modell für Erklärungsbedarf geboten.

Deters et al.[8] behandeln in ihrem Paper die Evaluation von Explainability in Softwaresystemen. Da die Evaluierung von Explainability von denjenigen Qualitätszielen abhängt, die die Erklärungen erreichen sollen, ist sie nicht trivial. Um dieses Problem anzugehen, werden die bereits etablierte Methode der Expertenbewertung mit zielorientierten Heuristiken kombiniert.

Das Paper von Drosten et al. [9] betont die wachsende Bedeutung der Einbindung von Erklärungen in moderne Softwaresysteme, um deren Komplexität zu verringern und sie für Endbenutzer verständlicher zu machen. Es präsentiert eine Taxonomie für Erklärbarkeitsbedürfnisse in verschiedenen Arten von Software und untersucht, wie sich der Bedarf an Erklärungen zwischen diesen Systemen unterscheidet.

Kupczyk[23] hat in seiner Masterarbeit eine Studie zur Untersuchung von der „Automatisierte Detektion von Erklärungsbedarf in Nutzerfeedback zu Software“ durchgeführt. Dabei wurden verschiedene Machine Learning und Deep Learning Modelle sowie eine regelbasierte Filtermethode getestet. Als teilweise präziser, aber mit weniger *Recall* als beim Deep Learning Modell hat sich die Filtermethode herausgestellt. Beide Modelle wurden in eine GUI integriert, um App- und Playstore-Reviews zu analysieren.

Kuckartz et al.[20] haben in ihrem Buch „Statistik: Eine verständliche Einführung“ verschiedenste statistische Tests und Verfahren grundlegend erläutert. Es kommen unter anderem der t-Test und die Identifikation von Zusammenhängen mittels Korrelationsanalysen vor. Dabei werden intervall, ordinal und nominalskalierte Variablen behandelt. Ebenfalls für diese Arbeit relevant ist die Likert-Skala, welche für multiple ordinale Variablen verwendet wird.

Siebertz et al.[31] gehen in ihrem Buch „Statistische Versuchsplanung“ die Pearson und Spearman-Rangkorrelationsanalyse durch. Es wird erklärt, wie Zusammenhänge anhand von verschiedenen Algorithmen zur linearen und nicht-linearen Korrelation und unterschiedlichen Auswirkungen von Fehlmessungen in Analysedaten gefunden werden können.

Janssen und Laatz[16] haben das Buch „Statistische Datenanalyse mit SPSS“ veröffentlicht, in dem unter anderem Clusteringverfahren erläutert werden. Das Kapitel Clusteranalyse beschäftigt sich mit den theoretischen Grundlagen und der praktischen Anwendung. Eine hierarchische Clusteranalyse wird in der Software SPSS durchgeführt.

3.2 Abgrenzung zu verwandten Arbeiten

Da die Erklärbarkeit als Grundlage dieser Arbeit dient, ist hierbei wichtig zu definieren, an welchem Bereich sich die Arbeit orientiert. Eine Erklärung wird auch als Interpretation oder Deutung von Handlungen, Ergebnissen, Verhalten usw. angesehen. Dieser Bereich fällt in die Explainability von KI. Die hier vorliegende Arbeit möchte sich nur im Rahmen der *SW-Explainability* und nicht der *Interpretability* bewegen.

Eine Arbeit aus dem KI-Bereich ist die „Diskriminierung im Machine Learning und Erklärbarkeit von Algorithmen“ von Jackszis[15]. Die Arbeit beschäftigt sich mit der Problematik, dass Algorithmen oft nicht erklärbar sind. Dies erschwert die Erkennung von Diskriminierung und verringert die Vertrauenswürdigkeit eines Systems. Es wurden die Wechselwirkungen zwischen Bias und Transparenz analysiert, und wie diese miteinander interagieren und zur Erhöhung der Softwarequalität beitragen können.

Kapitel 4

Studiendesign

In diesem Kapitel wird das geplante Konzept zur Durchführung und Auswertung der eigens erstellten Onlineumfrage im Detail erläutert. Dies dient zur Vorbereitung der Ergebnisse. Dabei werden die Forschungsfragen und Hypothesen aufgestellt. Für die Umfrage wurden zwei Domänen im Softwarebereich, welche möglichst vielen Nutzer aus ihrem Alltag bekannt sind, ausgewählt. Dies wurde entschieden, damit die Nutzer mehr Antwortmöglichkeiten haben und eine größere Anzahl verwendbarer Daten entstehen. Dabei ist es irrelevant, ob ein Teilnehmer kaum oder sehr viel Domänenwissen vorzuweisen hat. Um die Software-Favoriten der Teilnehmer herauszufinden, wurden als Orientierungshilfen bekannte SW-Produkte aus den beiden Domäne ausgesucht. Den Teilnehmern steht es offen, weitere oder keine anzugeben.

Office-Domäne

Textbearbeitungs-, Tabellenkalkulations- und Kommunikationsprogramme sind häufig Teil des Arbeitsplatzes [2]. Software in diesem Bereich ist also stets im allgemeinen Bewusstsein der Teilnehmer vorhanden. Spätestens während der Corona-Pandemie 2020-2022 hat sich durch das häufige Arbeiten im Home-Office sowie das Zurückziehen aus dem öffentlichen Leben der Arbeitsplatz zunehmend digitalisiert [2]. Da diese Domäne weit verbreitet ist, lässt sich gut prüfen, ob auch viele Personen mit ihr vertraut sind.

Browser-Domäne

Browser sind genauso wie Office-Software weitverbreitet [35]. Sie dienen nicht nur zur Visualisierung von Webseiten [35]. In vielen Browsern sind Zusatzfunktionen wie Adblocker oder Passwortmanager bereits integriert [35]. Durch ihre Vielseitigkeit und die vielen, sich ständig anpassenden Funktionen [35] bietet diese Domäne viel Spielraum für Wissensabfragen.

4.1 Konzept

Im Rahmen dieser Arbeit wurde ein Konzept zum Erhalt von Teilnehmerdaten einer Online-Studie in Bezug auf ihr Domänenwissen und zur Analyse dieser Daten erarbeitet. Dieses Konzept dient als Orientierungshilfe und Übersicht zur Durchführung des Vorhabens. Es unterteilt sich in zwei Bereiche, die **Herangehensweise** für die Durchführung dieser Arbeit und die Durchführung einer **Online-Studie** in Bezug auf Erklärungsbedarf in Abhängigkeit zu Domänenwissen.

4.1.1 Herangehensweise

Dieses Unterkapitel behandelt die Methode zur Durchführung der Arbeit. Die Herangehensweise ist in drei große Felder eingeteilt: Einarbeitungsphase, Umfrage, Analyse und Evaluation. Während der Einarbeitungsphase wurden sich mit den Grundlagen des Software-Engineerings im Bereich der Erklärbarkeit (und teilweise des Requirements Engineerings) vertraut gemacht 2.1. Diese sind entscheidend, um die Ermittlung des Erklärungsbedarfs präziser zu gestalten und passenden Fragen zu formulieren. Als Umfrageplattform wurde Lime Survey¹ verwendet. Aus dieser Vorarbeit ist die Umfrage entstanden. Mit den Umfrageergebnissen werden durch eine Korrelationsanalyse Verbindungen und Beziehungen zwischen den einzelnen 4.2.3 Hypothesen evaluiert. Die Clusteranalyse dient zur Gruppierung von den einzelnen Variablen in Bezug auf die Erklärungstiefe.

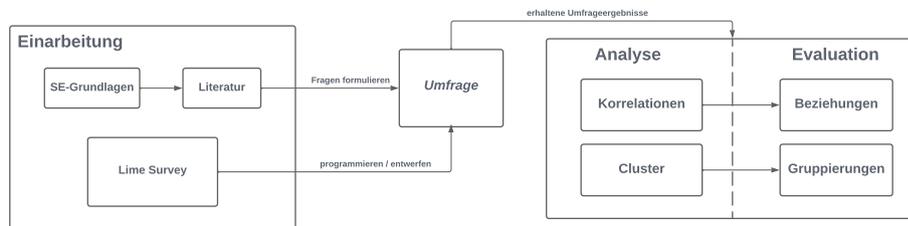


Abbildung 4.1: Herangehensweise zur Durchführung der Arbeit

4.1.2 Online-Studie

In diesem Abschnitt geht es um die Durchführung einer Online-Studie. Dabei liegt der Bezug auf dem Erklärungsbedarf in Abhängigkeit vom Domänenwissen. Dieser Prozess soll reproduzierbar gemacht werden. Die zu untersuchenden Domänen werden von Beginn an bestimmt. Diese Domänen sollten genügend Spielraum für gut zu formulierende Frage bieten. Im

¹Lime Survey: <https://www.limesurvey.org/de>

Optimalfall setzen sich Personen, die Experten in diesen Domänen sind, zusammen, um Fragen zu formulieren. Diese sollen das subjektiv empfundene und wirklich vorhandene (objektive) Domänenwissen abfragen. Aus Domänen und Fragen entsteht die Online-Studie. Im nächsten Schritt werden die Umfrageergebnisse exportiert und ausgewertet. Dafür bietet sich eine Korrelationsanalyse an, da diese Verbindungen aus den Forschungsfragen überprüft. Anschließend wird eine Clusteranalyse durchgeführt und ebenfalls ausgewertet. Bei Bedarf können die Ergebnisse der Clusteranalyse mit einer Regressionsanalyse weiter vertieft werden [38]. Mit den Ergebnissen sollte sich der Einfluss des Domänenwissens oder anderer getesteter Variablen auf den Erklärungsbedarf der Teilnehmer einordnen lassen. Die folgende Grafik wurde mittels der Flow-Methode erstellt [33].

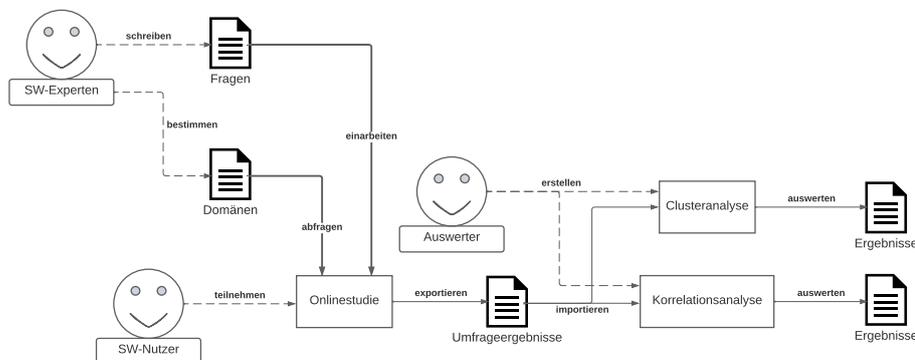


Abbildung 4.2: Durchführung und Auswertung der Online-Studie

4.2 Analyseaufbau

Im Folgenden wird die Durchführung der Analyse mit dem zugrunde liegenden Wissen zum verwendeten Datensatz und den Analysemethoden vorgestellt.

4.2.1 Datenvorbereitung

Damit die erhobenen Umfragedaten in der Korrelations- und Clusteranalyse vernünftig verwendet werden können, mussten einzelne Ausschnitte dessen extrahiert werden. Der Umfragedatensatz wurde als csv-Datei exportiert und als Datenbank in SQLite² eingearbeitet. Für einzelne Variablenspalten wurden SQL-Anfragen geschrieben.

²SQLite Datenbank <https://www.sqlite.org/>

4.2.2 Forschungsziele und Fragen

Das allgemeine Ziel dieser Arbeit ist es zu untersuchen, inwiefern das Domänenwissen der Nutzer von Software einen Einfluss auf die Erklärungen hat. Hierbei sollen die Abhängigkeiten von den Umfrageergebnissen zueinander geprüft werden. Falls keine Zusammenhänge identifiziert werden, gilt die Nullhypothese. Dazu stellen sich folgende Forschungsfragen:

RQ1: *Wie hängen das von den Teilnehmern eigenständig eingeschätzte Domänenwissen und das objektiv ermittelte Domänenwissen miteinander zusammen?*

Diese Untersuchung zielt darauf ab, das Ausmaß der Übereinstimmung der subjektiven Selbsteinschätzung mit dem objektiv ermittelten Domänenwissen zu erforschen. Bei einem hohen Maß an Übereinstimmung würde dies auf ein ebenso passendes Verständnis der betreffenden Domäne hinweisen.

RQ2: *Wie hängt das Domänenwissen mit dem Erklärungsbedarf zusammen?*

Diese Forschungsfrage soll klären, ob Domänenwissen einen Einfluss auf den Erklärungsbedarf hat. Der Erklärungsbedarf definiert sich durch die präferierte Erklärungstiefe, welche aus einer gewissen Erklärungsform stammt. Im Kontext dieser Arbeit werden nur Erklärungstexte, also Textformen, erläutert 2.2. Darüber hinaus existieren weitere Formen, die der Vollständigkeit halber auch betrachtet werden. Die Erklärungsform steht somit für die Repräsentation der Erklärung.

RQ3.1: *Wie hängt die Sicherheit im Umgang mit der bevorzugten Software der Teilnehmer mit dem Erklärungsbedarf zusammen?*

Als zusätzliche Teilfrage dient folgende:

RQ3.2: *Wie hängt die Sicherheit im Umgang mit der bevorzugten Software der Teilnehmer mit dem Domänenwissen zusammen?*

Falls ein geschickter Umgang der Teilnehmer mit der von ihnen bevorzugten Software einen Einfluss auf das Domänenwissen haben sollte, ist anzunehmen, dass das Domänenwissen dieser Teilnehmer höher ausgeprägt ist.

RQ4: *Wie hängen die demografischen Daten mit dem Erklärungsbedarf zusammen?*

Ziel ist es, herauszufinden, inwieweit sich demografische Faktoren auf die gewünschte Erklärungstiefe und Erklärungsform der beiden Domänen auswirken, und ob sich Zusammenhänge zeigen lassen.

RQ5: *Wie ist die Zufriedenheit im Umgang mit der bevorzugten Software der Teilnehmer mit dem Erklärungsbedarf verbunden?*

Dies ist eine zusätzliche Abfrage zum Erklärungsbedarf, um weitere mögliche Parallelen zwischen der Zufriedenheit im Umgang mit der bevorzugten Software und der gewünschten Erklärungstiefe und Erklärungsform zu ermitteln.

4.2.3 Hypothesenübersicht

Um die Forschungsfragen besser zu beantworten, wird jede Frage in mehrere Nullhypothesen aufgeteilt. Die Signifikanz der Hypothesen wird auf dem für statistische Zwecke üblichen Signifikanzlevel $\alpha = 0,05$ mit der Bonferroni Korrektur geprüft [36]. Wenn vier untergeordnete Hypothesen zusammenkommen, beträgt das korrigierte Niveau $\alpha_{\text{corr}} = \frac{0,05}{4} = 0,0125$ [36]. Jede Variable, die die Forschungsfragen beinhalten, wird auf ihre Verbindung mit anderen Variablen geprüft, sodass alle relevanten Kombinationen abgedeckt werden.

Hypothesen		RQ
H1 ₀	Das subjektive Domänenwissen entspricht nicht dem objektiven Domänenwissen.	RQ1
H1.1 ₀	Das subjektive Domänenwissen entspricht nicht dem objektiven Domänenwissen im Kontext der Office-Domäne.	RQ1
H1.2 ₀	Das subjektive Domänenwissen entspricht nicht dem objektiven Domänenwissen im Kontext der Browser-Domäne.	RQ1
H2 ₀	Das objektive Domänenwissen beeinflusst nicht den Erklärungsbedarf.	RQ2
H2.1 ₀	Das objektive Domänenwissen beeinflusst nicht die präferierte Erklärungsform im Kontext der Office-Domäne.	RQ2
H2.2 ₀	Das objektive Domänenwissen beeinflusst nicht die präferierte Erklärungsform im Kontext der Browser-Domäne.	RQ2
H2.3 ₀	Das objektive Domänenwissen beeinflusst nicht die gewünschte Erklärungstiefe im Kontext der Office-Domäne.	RQ2
H2.4 ₀	Das objektive Domänenwissen beeinflusst nicht die gewünschte Erklärungstiefe im Kontext der Browser-Domäne.	RQ2
H3 ₀	Die Anwendungssicherheit beeinflusst nicht den Erklärungsbedarf.	RQ3.1
H3.1 ₀	Die Anwendungssicherheit beeinflusst nicht die präferierte Erklärungsform im Kontext der Office-Domäne.	RQ3.1
H3.2 ₀	Die Anwendungssicherheit beeinflusst nicht die präferierte Erklärungsform im Kontext der Browser-Domäne.	RQ3.1
H3.3 ₀	Die Anwendungssicherheit beeinflusst nicht die gewünschte Erklärungstiefe im Kontext der Office-Domäne.	RQ3.1
H3.4 ₀	Die Anwendungssicherheit beeinflusst nicht die gewünschte Erklärungstiefe im Kontext der Browser-Domäne.	RQ3.1
H4 ₀	Die Anwendungssicherheit beeinflusst nicht das objektive Domänenwissen.	RQ3.2
H4.1 ₀	Die Anwendungssicherheit beeinflusst nicht das objektive Domänenwissen im Kontext der Office-Domäne.	RQ3.2
H4.2 ₀	Die Anwendungssicherheit beeinflusst nicht das objektive Domänenwissen im Kontext der Browser-Domäne.	RQ3.2

Tabelle 4.1: 1. Hypothesenübersicht

H50	Die demografischen Daten haben keinen Einfluss auf den Erklärungsbedarf.	RQ4
H5.1 ₀	Das Teilnehmerge schlecht hat keinen Einfluss auf die präferierte Erklärungsform der Office-Domäne.	RQ4
H5.2 ₀	Das Teilnehmerge schlecht hat keinen Einfluss auf die präferierte Erklärungsform der Browser-Domäne.	RQ4
H5.3 ₀	Das Teilnehmerge schlecht hat keinen Einfluss auf die gewünschte Erklärungstiefe der Office-Domäne.	RQ4
H5.4 ₀	Das Teilnehmerge schlecht hat keinen Einfluss auf die gewünschte Erklärungstiefe der Browser-Domäne.	RQ4
H5.5 ₀	Das Teilnehmeralter hat keinen Einfluss auf die präferierte Erklärungsform der Office-Domäne.	RQ4
H5.6 ₀	Das Teilnehmeralter hat keinen Einfluss auf die präferierte Erklärungsform der Browser-Domäne.	RQ4
H5.7 ₀	Das Teilnehmeralter hat keinen Einfluss auf die gewünschte Erklärungstiefe der Office-Domäne.	RQ4
H5.8 ₀	Das Teilnehmeralter hat keinen Einfluss auf die gewünschte Erklärungstiefe der Browser-Domäne.	RQ4
H5.9 ₀	Die Teilnehmerberufserfahrung hat keinen Einfluss auf die präferierte Erklärungsform der Office-Domäne.	RQ4
H5.10 ₀	Die Teilnehmerberufserfahrung hat keinen Einfluss auf die präferierte Erklärungsform der Browser-Domäne.	RQ4
H5.11 ₀	Die Teilnehmerberufserfahrung hat keinen Einfluss auf die gewünschte Erklärungstiefe der Office-Domäne.	RQ4
H5.12 ₀	Die Teilnehmerberufserfahrung hat keinen Einfluss auf die gewünschte Erklärungstiefe der Browser-Domäne.	RQ4
H60	Die Nutzerzufriedenheit beeinflusst den Erklärungsbedarf nicht.	RQ5
H6.1 ₀	Die Nutzerzufriedenheit beeinflusst nicht die gewünschte Erklärungsform im Kontext der Office-Domäne.	RQ5
H6.2 ₀	Die Nutzerzufriedenheit beeinflusst nicht die gewünschte Erklärungsform im Kontext der Browser-Domäne.	RQ5
H6.3 ₀	Die Nutzerzufriedenheit beeinflusst nicht die gewünschte Erklärungstiefe im Kontext der Office-Domäne.	RQ5
H6.4 ₀	Die Nutzerzufriedenheit beeinflusst nicht die gewünschte Erklärungstiefe im Kontext der Browser-Domäne.	RQ5
Beziehungen zwischen den Hypothesen und Variablen		
H1	Domänenwissen ($D(f, d)$)	
H2	Domänenwissen ($D(f^2, d)$), Erklärungsform (E_{form}), Erklärungstiefe (E_{tiefe}), Domänen (d)	
H3	Anwendungssicherheit (S_{sw}), Erklärungsform (E_{form}), Erklärungstiefe (E_{tiefe}), Domänen (d)	
H4	Anwendungssicherheit (S_{sw}), Domänenwissen ($D(f^2, d)$), Domänen (d)	
H5	Teilnehmerge schlecht (d_{gesch}), Teilnehmeralter (d_{alter}), Teilnehmerberufserfahrung (d_{job}), Erklärungsform (E_{form}), Erklärungstiefe (E_{tiefe}), Domänen (d)	
H6	Nutzerzufriedenheit ($N_{zufrieden}$), Erklärungsform (E_{form}), Erklärungstiefe (E_{tiefe}), Domänen (d)	

Tabelle 4.2: 2. Hypothesenübersicht

4.2.4 Variablenübersicht

Variable	Name	Skalierung	Wertebereich	Beschreibung
f	Domänenwissen-Formen	nominal	$\{f_1; f_2\}$	subjektiv, objektiv
d	Domänen	nominal	$\{d_1; d_2\}$	Office, Browser
$D(f, d)$	Domänenwissen	ordinal	1..5	in Form von f und für Domäne d
S_{sw}	Anwendungssicherheit	ordinal	1..5	
E_{form}	Erklärungsform	ordinal	1..5	Repräsentation der Erklärung
E_{tiefe}	Erklärungstiefe	ordinal	1..5	Detailliertheit des Erklärungstextes
d_{geschl}	Teilnehmergeschlecht	nominal	$\{1; 2; 3\}$	weiblich, männlich, divers
d_{alter}	Teilnehmeralter	metrisch	$n \in \mathbb{N}_0$	
d_{job}	Teilnehmerberufserfahrung	nominal	$\{1; \dots; 8\}$	repräsentiert Berufsfelder
$N_{zufrieden}$	Nutzerzufriedenheit	ordinal	1..5	

Tabelle 4.3: Variablenübersicht

Die Erklärungsform ist wie die Erklärungstiefe (2.2 Detailliertheit und Erklärungstexte) in fünf Stufen unterteilt: 1. keine spezielle Form oder andere, 2. simple Form, z. B. als kurze Textanleitungen, 3. komplexere Form, z. B. Video-Tutorials (GIFs), 4. aufwendigere Form, z. B. interaktive Schulungen und 5. sehr zeitintensive Form, z. B. Live-Chat-Support. Die Erklärungsform steigt in ihrer Detailliertheit und Komplexität wie die Erklärungstiefe an. Die Berufsfelder sind wie folgt definiert: {Bildung; Finanzen; Gesundheit; IT/Technologie; Rechtswesen; Student; in keinem; andere}. Im Verlaufe der Arbeit werden D_{sub} und D_{obj} genutzt. Sie stehen für subjektives Domänenwissen $D(f1, d)$ und objektives Domänenwissen $D(f2, d)$ und sind lediglich Abkürzungen.

4.2.5 Cluster

In dieser Arbeit wurde die agglomerative Clusteranalyse verwendet, da vor der Clusteranalyse nicht die genaue Anzahl der möglichen Cluster bekannt ist. Wie in 2.5 Clusteranalyse beschrieben, wird durch mehrere Iterationen die Bildung von zwei Clustern garantiert. Die Clusteranalyse wurde zur Identifizierung von Gruppierungen unter allen Variablen und in Bezug auf die gewünschte Erklärungstiefe durchgeführt. Als Input dient dabei eine csv-Datei, die als Spaltennamen die jeweilige Variable und in ihren Zeilen die Fälle beinhaltet. Ein sogenannter Fall [14] entspricht im Kontext der Umfragedaten einem Teilnehmer und seinen Angaben, welcher durch seine *ID* einzigartig ist. Insgesamt wurden drei Clusteranalysen durchgeführt. Die erste Clusteranalyse behandelt alle Variablen der Office-Domäne und wird mit *OfficeMain* abgekürzt. Die zweite behandelt alle Variablen der Browser-Domäne und wird mit *BrowserMain* abgekürzt. Damit sollen grobe Tendenzen und Zusammenhänge festgestellt werden. Für die jeweilige Domäne sieht die Eingabe wie folgt aus:

id	d _{geschl}	d _{alter}	d _{job}	D _{sub}	D _{obj}	S _{sw}	E _{form}	E _{tiefe}	N _{zufrieden}
1	1	23	6	3	5	4	1	3	4
2	1	26	6	5	5	4	1	1	4
...

Tabelle 4.4: Variablen für Clusteranalyse OfficeMain und BrowserMain

Die dritte Analyse wird mit *ETiefe* abgekürzt und versucht beeinflussende Variablen auf die gewünschte Erklärungstiefe zu finden. Da die Korrelationsanalyse nicht jede Variable abdeckt, wurde dies dynamisch mit der Clusteranalyse behandelt. Damit kommt nicht nur das Domänenwissen als beeinflussender Faktor auf den Erklärungsbedarf infrage, sondern alle Variablen, einschließlich der demografischen Daten. Die Eingabe sieht dabei wie folgt aus:

id	E _{tiefe_{office}}	E _{tiefe_{browser}}
1	3	1
2	1	1
...

Tabelle 4.5: Variablen für Clusteranalyse ETiefe

Die PCA wurde, mit den bereitgestellten Methoden aus 2.6 Python-Bibliotheken durchgeführt. Der Sourcecode, welcher im Nachhinein leicht abgewandelt wurde, stammt aus der Bachelorarbeit „Clusteranalyse zwischen Stimmung und Erklärbarkeitsanforderungen von Nutzern“ von Niklas Jösten [17].

4.3 Abfrage allgemeine Meinung der Teilnehmer

Als zusätzliche Einordnung des Domänenwissens in Bezug auf Erklärungsbedarf wurden die Teilnehmer abschließend über ihre Zustimmung zu dementsprechenden allgemeinen Aussagen gefragt. In diesem Kontext geht es um Anwender von Software. Die erste Aussage A: „*Eine Person mit mehr Domänenwissen hat weniger Erklärungsbedarf.*“ konfrontiert den Teilnehmer am Ende der Umfrage mit dem Hauptthema der Arbeit. Damit sollen ein allgemeines Stimmungsbild eingefangen und weitere Meinungsdaten gesammelt werden. Die zweite Aussage B: „*Erklärungen sollten für Personen mit weniger Domänenwissen detaillierter sein.*“ versucht die generelle Einschätzung über die gewünschte Erklärungstiefe einzufangen. In der dritten Aussage C: „*Mehr Domänenwissen hilft Personen dabei, ihre Ziele schneller zu erreichen.*“ liegt der Schwerpunkt auf der Bedeutung von Domänenwissen für die persönliche und berufliche Entwicklung. Die letzte Aussage D: „*Software aus den letzten Jahren ist verständlicher geworden.*“ befasst sich mit der

allgemeinen Frage, ob Software für den Anwender immer komplexer wird, oder ob sich der Trend eher in eine nutzerfreundlichere Richtung entwickelt. Die Usability hat im Software-Engineering eine relevante Qualität³. Mit den subjektiven Einschätzungen der Teilnehmer ist es schwierig, diese Aussagen zu bestätigen. Dennoch dient sie hier auch wieder für ein allgemeines Stimmungsbild.

³Usability aus der ISO 25010 <https://iso25000.com/>

Kapitel 5

Ergebnisse

In diesem Kapitel werden hauptsächlich die Ergebnisse aus der Korrelations- und Clusteranalyse behandelt, anschließend aufgelistet und zusammengefasst. Am Anfang beschäftigt sich das Kapitel mit solchen Umfragedaten, die nicht in den Analysen vorkamen.

5.1 Randdaten

An der Umfrage haben 58 Personen teilgenommen, davon sind 11 Frauen, 46 Männer und eine Person divers. Das Durchschnittsalter beträgt 33,5 Jahre, wovon 18 Jahre das jüngste und 84 Jahre das älteste ist. Jeder Teilnehmer sollte zusätzlich seine Berufserfahrung angeben. Dabei sind folgende Berufsfelder vorgegeben worden: Bildung, Finanzen, Gesundheitswesen, IT/Technologie, Rechtswesen. Zusätzlich konnte angegeben werden, ob ein Teilnehmer Student ist. Falls kein Berufsfeld zutrifft oder ein Teilnehmer kein Student ist, ist die Option „keinem“ anzugeben. Ebenfalls konnten die Teilnehmer sonstige Berufsfelder angeben. Dabei sind folgende Angaben gemacht worden: Mediengestalter Bild und Ton, Elektrotechnik, Architektur, Ingenieurwissenschaften, Industrie, Literatur-Sprach- und Kulturwissenschaft, Wirtschaft, Forstwissenschaft. In allen Berufsfeldern, die als „sonstige“ angegeben wurden, arbeitet oder hat jeweils nur ein Teilnehmer Berufserfahrung. Das Gesamtbild der Berufsfelder sieht wie folgt aus:

Berufsfeld	Verteilung
Bildung	9
Finanzen	3
Gesundheitswesen	1
IT/Technologie	13
Rechtswesen	3
Student	28
keinem	1
andere	8

Tabelle 5.1: Berufsfelderverteilung

Im Gesamtbild lässt sich klar erkennen, dass der Großteil der Teilnehmer aus Studenten und Personen, die im IT/Technologie-Bereich tätig sind, besteht. Zusätzlich wurde ermittelt, dass fünf Studenten bereits ein Berufsfeld erlernt haben oder in einem arbeiten. In beiden Domänen wurden Softwareprodukte als Orientierungshilfe vorgegeben. Die Teilnehmer sollten ankreuzen, welche Produkte sie im Alltag oder auf der Arbeit verwenden. Wenn *keine* Software angekreuzt wurde, ist für die Teilnehmer die Umfrage beendet und es entstanden keine weiteren Daten. Die Verteilung sieht wie folgt aus:

Office	Verteilung
branchenspezifische Software	27
Google Workspace (Docs, Sheets, Slides)	25
Microsoft Office Suite (Word, Excel, Powerpoint usw.)	40
Projektmanagement-Tools (z. B. Jira, Trello)	15
keine	5
sonstige	6

Browser	Verteilung
Chrome	27
Edge	12
Firefox	36
Internet Explorer	4
Opera	10
Safari	23
keine	1
sonstige	6

Tabelle 5.2: Softwareprodukte Verteilung

Am Ende der Umfrage sollten die Teilnehmer ihre Zustimmung zu Aussage A-D angeben. Folgende Aussagen wurde gestellt:

- A:** *Eine Person mit mehr Domänenwissen hat weniger Erklärungsbedarf.*
B: *Erklärungen sollten für Personen mit weniger Domänenwissen detaillierter sein.*
C: *Mehr Domänenwissen hilft Personen dabei, ihre Ziele schneller zu erreichen.*
D: *Software aus den letzten Jahren ist verständlicher geworden.*

Die meisten stimmen den Aussagen zu. In wenigen Fällen gibt es gar keine Zustimmung. Aussage B hat die meisten „stimme überhaupt nicht zu“ und „stimme eher nicht zu“. Aussage C hat die höchste Zustimmung. Die Meinungsverteilung bildet sich wie folgt zusammen:

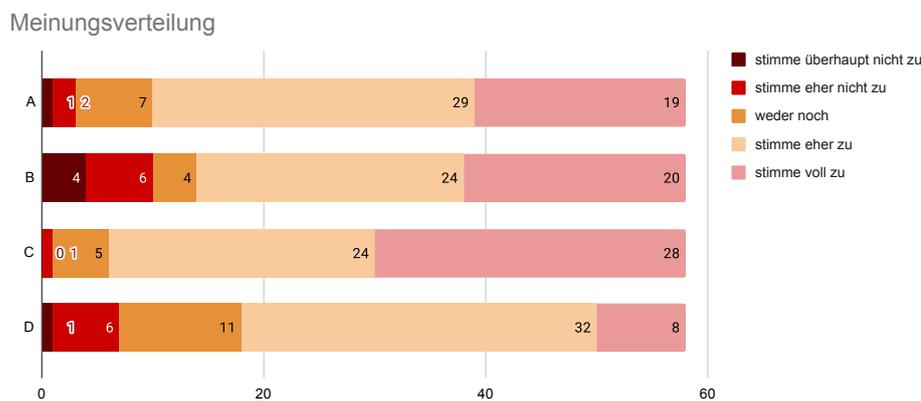


Abbildung 5.1: Meinungsverteilung der Aussagenabfrage

5.2 Ergebnisse der Korrelationsanalyse

Die statistischen Ergebnisse der Korrelationsanalyse ergeben sich aus den getesteten 4.2.3 Nullhypothesen. Dabei werden stets zwei Variablen Var_1 und Var_2 auf einen signifikanten Zusammenhang überprüft. Der Korrelationskoeffizient r_s und das Ergebnis des Chi-Quadrates X^2 sind dabei die Grundlage für die Einstufung auf die Effektstärke oder den jeweiligen Zusammenhang. Für die Spearman Korrelation wird die Effektstärke nach Cohen genutzt [6]. Das Chi-Quadrat verwendet den errechneten Cramers V-Wert, um den Zusammenhang einzustufen [20]. Wenn der p-Wert unter dem angepasste Signifikanzlevel

$$\alpha_{\text{corr}} = \frac{0,05}{ANZ(RQ_i)}$$

nach der Bonferroni Korrektur liegt, wird die Nullhypothese verworfen [36]. Hierbei ist $ANZ(RQ_i)$ die Anzahl der aufgestellten Hypothesen pro Forschungsfrage ist. Wenn mehr als die Hälfte der untergeordneten Nullhypothesen verworfen werden, dann wird auch die übergeordnete Nullhypothese verworfen. Die Hypothesen werden jeweils für die Office- wie auch Browser-Domäne getestet. Die Variablen sind der Variablenübersicht 4.2.4 zu entnehmen.

H ₀	Domäne	Var ₁	Var ₂	$ r_s $ oder X^2	p-Wert	Bonferroni	Zusammenhang
H1.1 ₀	Office	D _{sub}	D _{obj}	0,6201	0,0000002	verwerfen	stark
H1.2 ₀	Browser	D _{sub}	D _{obj}	0,5654	0,0000037	verwerfen	stark
H2.1 ₀	Office	D _{obj}	E _{form}	-0,4891	0,0000977	verwerfen	mittel
H2.2 ₀	Browser	D _{obj}	E _{form}	-0,1113	0,4053816	nicht verwerfen	schwach
H2.3 ₀	Office	D _{obj}	E _{tiefe}	-0,0683	0,6100771	nicht verwerfen	kein Effekt
H2.4 ₀	Browser	D _{obj}	E _{tiefe}	-0,0260	0,8459299	nicht verwerfen	kein Effekt
H3.1 ₀	Office	S _{sw}	E _{form}	-0,3433	0,0083325	verwerfen	mittel
H3.2 ₀	Browser	S _{sw}	E _{form}	-0,0970	0,4686234	nicht verwerfen	kein Effekt
H3.3 ₀	Office	S _{sw}	E _{tiefe}	-0,0317	0,8128447	nicht verwerfen	kein Effekt
H3.4 ₀	Browser	S _{sw}	E _{tiefe}	0,1186	0,3750410	nicht verwerfen	schwach
H4.1 ₀	Office	S _{sw}	D _{obj}	0,5923	0,0000009	verwerfen	stark
H4.2 ₀	Browser	S _{sw}	D _{obj}	0,5694	0,0000030	verwerfen	stark
H5.1 ₀	Office	d _{gesch}	E _{form}	7,9633	0,4370547	nicht verwerfen	gering
H5.2 ₀	Browser	d _{gesch}	E _{form}	12,3144	0,1377139	nicht verwerfen	mittel
H5.3 ₀	Office	d _{gesch}	E _{tiefe}	8,9156	0,3494653	nicht verwerfen	gering
H5.4 ₀	Browser	d _{gesch}	E _{tiefe}	3,2845	0,9152552	nicht verwerfen	gering
H5.5 ₀	Office	d _{alter}	E _{form}	-0,0581	0,6646178	nicht verwerfen	kein Effekt
H5.6 ₀	Browser	d _{alter}	E _{form}	-0,0821	0,5400545	nicht verwerfen	kein Effekt
H5.7 ₀	Office	d _{alter}	E _{tiefe}	0,1375	0,3031455	nicht verwerfen	schwach
H5.8 ₀	Browser	d _{alter}	E _{tiefe}	0,0369	0,7832352	nicht verwerfen	kein Effekt
H5.9 ₀	Office	d _{job}	E _{form}	16,7422	0,9535075	nicht verwerfen	gering
H5.10 ₀	Browser	d _{job}	E _{form}	14,6708	0,9817029	nicht verwerfen	gering
H5.11 ₀	Office	d _{job}	E _{tiefe}	11,7823	0,9969024	nicht verwerfen	gering
H5.12 ₀	Browser	d _{job}	E _{tiefe}	14,6708	0,9817029	nicht verwerfen	gering
H6.1 ₀	Office	N _{zufrieden}	E _{form}	-0,3299	0,0114312	verwerfen	mittel
H6.2 ₀	Browser	N _{zufrieden}	E _{form}	-0,1855	0,1631928	nicht verwerfen	schwach
H6.3 ₀	Office	N _{zufrieden}	E _{tiefe}	0,2123	0,1094930	nicht verwerfen	schwach
H6.4 ₀	Browser	N _{zufrieden}	E _{tiefe}	0,1002	0,4539858	nicht verwerfen	schwach

Tabelle 5.3: Ergebnisse der Korrelationsanalyse

H1₀: Die Selbsteinschätzung der Teilnehmer trifft das objektiv ermittelte Domänenwissen ziemlich genau. Auch nach der Bonferroni Korrektur liegt der p-Wert deutlich unter dem α_{corr} -Wert. Um zu testen, ob diese Korrelation hochsignifikant ist, wird $\alpha = 0,01$ als neues Signifikanzlevel genommen. Somit ist $\alpha_{\text{corr}} = \frac{0,01}{2} = 0,005$. Die p-Werte von H1.1₀ und H1.2₀ liegen weiterhin unter α_{corr} und zeigen, dass H1₀ insgesamt hochsignifikant ist. Die Effektstärke nach Cohen zeigt ebenfalls einen starken Zusammenhang auf. Somit werden H1₀ und alle untergeordneten Nullhypothesen verworfen.

H2₀: Das objektive Domänenwissen korreliert nur in der Office-Domäne mit der gewünschten Erklärungsform. Der p-Wert ist mit 0,00009 so gering, dass die Korrelation ebenfalls für $\alpha = 0,01$ und $\alpha_{\text{corr}} = \frac{0,01}{4} = 0,0025$ hochsignifikant ist. Somit wird H2.1₀ verworfen. Die anderen Nullhypothesen dieser Forschungsfrage korrelieren nicht. Daher lässt sich H2₀ nicht verwerfen.

H3₀: Die Anwendungssicherheit korreliert in der Office-Domäne mit der gewünschten Erklärungsform nur für $\alpha = 0,05$ und mit $\alpha_{\text{corr}} = \frac{0,05}{4} = 0,0125$ signifikant. H3.1₀ wird verworfen. Alle anderen untergeordneten Nullhypothesen aus H3₀ werden nicht verworfen. H3₀ wird ebenfalls nicht verworfen.

H4₀: Die Nullhypothese H4₀ wird verworfen, da H4.1₀ und H4.2₀ mit $p < \alpha_{\text{corr}}$ für $\alpha_{\text{corr}} = \frac{0,01}{2} = 0,005$ hochsignifikant sind und ebenfalls verworfen werden. Beide zeigen einen starken Effekt zwischen der Sicherheit in der Anwendung und dem objektiven Domänenwissen auf.

H5₀: Die demografischen Daten weisen in keiner Hypothese statistische Signifikanz auf. In der Browser-Domäne gibt es unter dem Geschlecht und der gewünschter Erklärungsform keine Signifikanz, aber durch Cramers V einen mittleren Zusammenhang. Das Alter hat aus der Office-Domäne einen schwachen Zusammenhang mit der gewünschten Erklärungstiefe. Die Signifikanz ist für die demografischen Daten auch ohne Bonferri Korrektur nicht vorhanden. Die Nullhypothese H5₀ wird also nicht verworfen.

H6₀: Die Zufriedenheit in der Nutzung der gewählten Software korreliert in der Office-Domäne signifikant mit der gewünschten Erklärungsform. Hierbei wird $\alpha_{\text{corr}} = \frac{0,05}{4} = 0,0125$ mit $p = 0,0114$ unterschritten. H6.1₀ wird verworfen. Die übergeordnete Nullhypothese H6₀ wird durch H6.2₀, H6.3₀ und H6.4₀ nicht verworfen.

5.3 Ergebnisse der Clusteranalyse

Die agglomerative Clusteranalyse unterteilt sich in zwei Clusteranalysen über alle Variablen, jeweils für die Domänen und eine Clusteranalyse in Bezug auf die Erklärungstiefe. Insgesamt wurden also drei Clusteranalysen durchgeführt.

Für *OfficeMain* und *BrowserMain* bilden sich immer die Durchschnittsalter von 24,1 in Cluster 1 und 62,5 in Cluster 2. Das Domänenwissen ist in Cluster 1 mit den jüngeren Teilnehmern im Schnitt höher als das der älteren Teilnehmer aus Cluster 2. Die Anwendungssicherheit ist ebenfalls in beiden Domänen im ersten Cluster höher als in Cluster 2. Dies geht auch mit den Daten der Korrelationsanalyse in einher. Dort gibt es einen hochsignifikanten Zusammenhang zwischen dem Domänenwissen und der Anwendungssicherheit. Die gewünschte Erklärungsform ist in der Office-Domäne im ersten Cluster nur leicht niedriger als im zweiten Cluster. In der Browser-Domäne ist dies genau andersherum. Zwischen den ersten Clustern aus beiden Domänen existiert kein nennenswerter Unterschied für die gewünschten Erklärungsform. Die gewünschte Erklärungstiefe ist im ersten Cluster immer etwas niedriger als im zweiten Cluster. Die Geschlechter und die Berufsfelder sind zwischen beiden Domänen gleich verteilt. Im ersten Cluster sind immer alle ITler und Studenten enthalten. Die folgende Tabelle

fasst die Mittelwerte und Verteilung vom Teilnehmergeschlecht wie auch der Berufserfahrung zusammen:

	Office		Browser	
	Cluster 1	Cluster 2	Cluster 1	Cluster 2
d_{alter}	24,1	62,5	24,1	62,5
D_{sub}	3,4	3	3,9	3
D_{obj}	3,9	2,8	4,6	3,3
S_{sw}	3,6	3	4,5	3,6
E_{form}	1,7	2	1,6	1,3
E_{tiefe}	1,9	2,6	2,2	2,5
$N_{\text{zufrieden}}$	3,5	3,3	4,1	3,8
d_{geschl}	Cluster 1	Cluster 2	Cluster 1	Cluster 2
Frauen	6	5	6	5
Männer	36	10	36	10
Divers	1	0	1	0
d_{job}	Cluster 1	Cluster 2	Cluster 1	Cluster 2
Bildung	3	6	3	6
Finanzen	1	2	1	2
Gesundheit	0	1	0	1
IT/Technologie	12	0	12	0
Rechtswesen	1	2	1	2
Student	23	0	23	0
keinem	1	0	1	0
andere	2	4	2	4

Tabelle 5.4: Mittelwerte und Verteilungen der Domänen-Clusteranalysen

Für alle drei Clusteranalysen wurde ein Silhouettenkoeffizient gefunden, der auf eine starke Struktur hinweist. Nach der Interpretation der Tabelle 2.4 liegt ab $SK > 0,70$ eine starke Struktur vor. Für die ersten zwei Clusteranalysen wurden jeweils zwei Cluster gefunden. Die dritte Clusteranalyse $ETiefe$ behandelt die Erklärungstiefe aus beiden Domänen und versucht häufige Einordnungen von Erklärungstiefe-Kombinationen zu finden. Da hierbei zwei Cluster wenig sinnvoll erscheinen, wurde die Clusteranalyse $ETiefe$ ein erstes Mal durchgeführt, um den Silhouettenkoeffizient für die ersten fünf aussagekräftigsten Clusteranzahlen zu finden. Die niedrigste Clusteranzahl, welche größer als zwei Cluster ist und den stärksten Silhouettenkoeffizient hat, ist fünf Cluster mit $SK \approx 0,8914$.

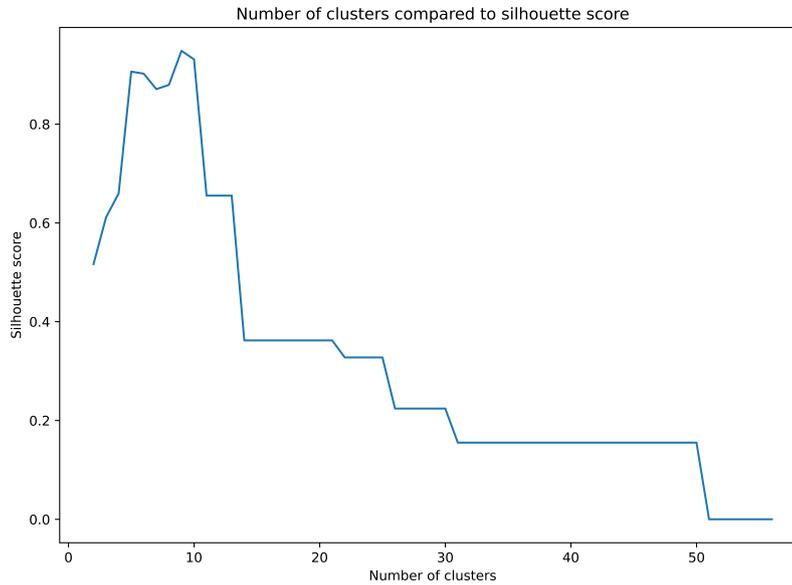


Abbildung 5.2: Silhouettenkoeffizient der ETiefe Clusteranalyse

Analyse	Cluster	Silhouettenkoeffizient
OfficeMain	2	0.7951
BrowserMain	2	0.7986
ETiefe	5	0.8914

Tabelle 5.5: Clusteranzahl und Silhouettenkoeffizient

Das erste Cluster der *ETiefe* Clusteranalyse hat hauptsächlich *detaillierte* (3) Einträge. Das zweite und dritte Cluster wechseln sich zwischen *sehr wenig detailliert* (1) und *detailliert* (3) jeweils in der Domäne ab. Personen aus diesen Clustern wünschen sich im Schnitt immer eine sehr wenig *detaillierte* (1) und die andere eine *detaillierte* (3) Erklärung. Das vierte Cluster besteht nur aus zwei Personen, welche sich im Office-Bereich *sehr detaillierte* (5) und im Browser-Bereich *viel detaillierte* (4) Erklärungen wünschen. Im letzten Cluster sollen Erklärungstexte möglichst simple sein. Es werden in beiden Domänen nur *sehr wenig detaillierte* (1) Erklärungen bevorzugt. Das Durchschnittsalter hebt sich im vierten Cluster von allen anderen mit 60 Jahren ab. Die Werte der anderen Variablen sind in allen Clustern ähnlich verteilt.

	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5
Mittelwert von E_{tiefe}	3,07	2	3,27	4,5	1
Einträge	19	10	11	2	16

Tabelle 5.6: Clusterverteilung für ETiefe

Für die Untersuchung der gefundenen Cluster wurden diese als Dendrogramm und Streudiagramm visualisiert. Im Dendrogramm lässt sich der Verlauf der Clusterbildung besser erkennen. Im Streudiagramm wird die Verteilung der einzelnen Cluster dargestellt. Auf den Achsen des Streudiagramms sind die Komponenten 1 (pc1) und 2 (pc2) angegeben. Diese sind Ergebnisse aus der PCA und geben den Prozentsatz der abgedeckten Daten an.

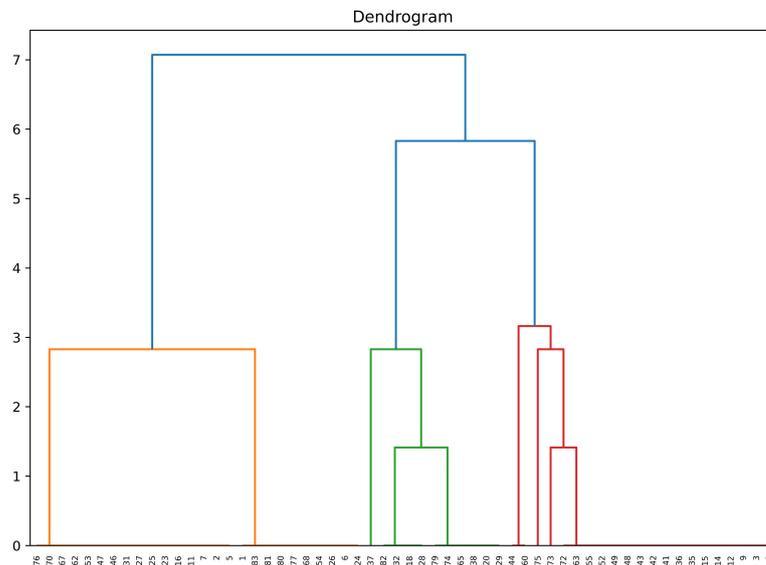


Abbildung 5.3: Dendrogramm der ETiefe Clusteranalyse

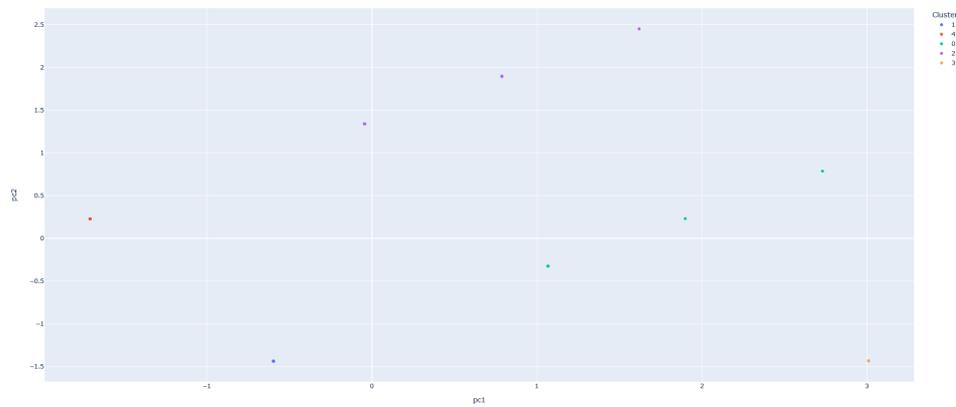


Abbildung 5.4: Streudiagramm der ETiefe Clusteranalyse

Kapitel 6

Diskussion

In diesem Kapitel werden die Forschungsfragen aus Kapitel 4 mit den Ergebnissen aus dem vorherigen Kapitel beantwortet und anschließend interpretiert.

6.1 Beantworten der Forschungsfragen

RQ1: *Wie hängen das von den Teilnehmern eigenständig eingeschätzte Domänenwissen und das objektiv ermittelte Domänenwissen miteinander zusammen?*

Die hochsignifikante Korrelation zwischen subjektivem und objektivem Domänenwissen mit einem starken Effekt, zeigt deutlich, dass sich die Teilnehmer sehr gut in ihrem Domänenwissen eingeschätzt haben. Für die Clusteranalyse gilt, dass die Mittelwerte des subjektiven und objektiven Domänenwissens keine großen Differenzen in den Clustern aufweisen. Es lässt sich also beobachten, dass das eigenständig eingeschätzte Domänenwissen und das objektive ermittelte Domänenwissen durch einen starken Zusammenhang miteinander verbunden sind.

RQ2: *Wie hängt das Domänenwissen mit dem Erklärungsbedarf zusammen?*

Aus der Korrelationsanalyse geht hervor, dass keine signifikante Korrelation zwischen Erklärungstiefe und Domänenwissen besteht. Zwischen der gewünschten Erklärungsform und dem Domänenwissen aus der Office-Domäne besteht ein mittelstarker Zusammenhang. Dieser hat für die Forschungsfrage zu wenig Aussagekraft, um eine generelle Verbindung zwischen Domänenwissen und Erklärungsbedarf herzustellen. Aus der Clusteranalyse geht hervor, dass das Domänenwissen sich vielmehr nach den demografischen Daten (außer dem Geschlecht) orientiert und nicht nach der gewünschten Erklärungstiefe. Somit ist das Domänenwissen, wenn auch nur sehr schwach, mit dem Erklärungsbedarf verbunden.

RQ3.1: *Wie hängt die Sicherheit im Umgang mit der bevorzugten Software der Teilnehmer mit dem Erklärungsbedarf zusammen?*

Die Anwendungssicherheit hat einen sehr ähnlichen Bezug zum Erklärungsbedarf. Wie beim Domänenwissen fällt nur der Fall der gewünschten Erklärungsform aus der Office-Domäne mit einem mittelstarken Zusammenhang auf. Aus der Clusteranalyse geht hervor, dass die gewünschte Erklärungstiefe sich nicht nach der Anwendungssicherheit orientiert. Die Sicherheit im Umgang mit der bevorzugten Software der Teilnehmer ist, wenn auch nur sehr schwach, mit dem Erklärungsbedarf verbunden.

RQ3.2: *Wie hängt die Sicherheit im Umgang mit der bevorzugten Software der Teilnehmer mit dem Domänenwissen zusammen?*

Zwischen Anwendungssicherheit und Domänenwissen besteht ein starker Zusammenhang. Die hochsignifikante Korrelation mit einem starken Effekt weist auf eine tiefe Verbindung hin. Auch in den Clustern ist deutlich zu erkennen, dass sich die Anwendungssicherheit am Domänenwissen orientiert und somit auch an den demografischen Daten (außer dem Geschlecht). Je höher das Domänenwissen ausgeprägt ist, desto sicherer ist der Umgang mit Software aus dieser Domäne. Es lässt sich also sagen, dass die Sicherheit im Umgang mit der bevorzugten Software der Teilnehmer sehr stark mit dem Domänenwissen verbunden ist.

RQ4: *Wie hängen die demografischen Daten mit dem Erklärungsbedarf zusammen?*

Zwischen demografischen Daten und dem Erklärungsbedarf gibt es keine signifikante Korrelation. Zudem liegt überwiegend kein Effekt oder nur ein geringerer Zusammenhang vor. Auch die Clusteranalyse kann keinen aussagekräftigen Zusammenhang zwischen demografischen Daten und Erklärungsbedarf herstellen. Dort, wo geringe bis mittlere Zusammenhänge existieren, steht der Erklärungsbedarf mit dem Alter oder der beruflichen Erfahrung in einer schwachen Verbindung. Zusammenfassend lässt sich sagen, dass die demografischen Daten keinen starken Zusammenhang mit dem Erklärungsbedarf haben.

RQ5: *Wie ist die Zufriedenheit im Umgang mit der bevorzugten Software der Teilnehmer mit dem Erklärungsbedarf verbunden?*

Aus der Korrelationsanalyse geht hervor, dass nur die gewünschte Erklärungsform aus der Office-Domäne einen signifikanten Zusammenhang mit der Nutzerzufriedenheit hat. Dabei ist der Effekt nur mittelstark ausgeprägt. In der *OfficeMain* und *BrowserMain* Clusteranalyse fällt eine leichte Verbindung zwischen höherem Domänenwissen und höherer Nutzerzufriedenheit auf, zwischen der Zufriedenheit und der Erklärungstiefe allerdings nicht. Somit ist die Zufriedenheit im Umgang mit der bevorzugten Software der Teilnehmer nur sehr schwach mit dem Erklärungsbedarf

verbunden.

6.2 Interpretation

Aus der Korrelationsanalyse und Clusteranalyse geht hervor, dass der Erklärungsbedarf auf keine der getesteten Variablen einen starken Zusammenhang aufweist. Daraus lässt sich schließen, dass externe Faktoren den Erklärungsbedarf beeinflussen. Die Umfragedaten haben sich hauptsächlich auf das Domänenwissen, die Häufigkeit des Erklärungsbedarfs und die gewünschte Erklärungsform als auch die Erklärungstiefe bezogen. Dabei steht der geringe Zusammenhang nicht zwangsläufig für einen immer gleichbleibenden Erklärungsbedarf, welcher sich nicht durch andere Faktoren beeinflussen lässt. Das ist ein Anreiz, um an diesem Punkt in der Forschung anzusetzen und ein breites Feld an Faktoren zu untersuchen. Eine unklare Benutzeroberfläche, die Nutzer verwirrt oder Schwierigkeiten beim Verständnis der Funktionalitäten verursacht, könnte im Erklärungsbedarf eine Rolle spielen. Komplexere Algorithmen oder Geschäftslogiken, die schwer zu durchschauen sind, könnten ebenfalls einen Einfluss auf den Erklärungsbedarf haben. Oder technische Fehler und unerwartetes Verhalten führen dazu, dass Nutzer zusätzliche Informationen benötigen, um ein Problem zu verstehen und zu beheben.

Die Durchführung der *ETiefe* Clusteranalyse konnte fünf verschiedene aussagekräftige *Typen* von Personen für die gewünschte Erklärungstiefe in beiden Domänen finden. Zwei der fünf Cluster beinhalten Personen, die in einer Domäne keine tiefgehenden Erklärungen wünschen, dafür aber in der anderen wiederum detaillierte. Das zeigt, dass die Domänen untereinander auch nochmal ganz unterschiedlich zu betrachten sind. Wenn es Faktoren in der einen Domäne gibt, heißt das nicht zwangsläufig, dass diese auch in anderen Domänen den Erklärungsbedarf beeinflussen. Die übrigen drei Cluster sind deutlich einseitiger verteilt. Dort kommen größtenteils nur Personen, die unabhängig von der Domäne wenig oder viel Erklärungsbedarf haben, vor. Die Entwickler von Softwareprodukten sollten ihre Kunden kennen, damit sie die Erklärungstiefe dem Kunden anpassen können.

Das Domänenwissen beeinflusst den Erklärungsbedarf nur sehr schwach, dafür hängen Domänenwissen und die Sicherheit in der Anwendung fest zusammen. Je mehr Domänenwissen zu einer Software vorhanden ist, desto sicherer und vertrauter können Nutzer mit ihr umgehen. Es zeigt sich also, dass Wissen über die Funktionen einen leichteren Umgang ermöglichen kann. Vor allem jüngere Menschen haben im Durchschnitt ein höheres Domänenwissen als ältere Personen. Dies wird an dem Generationenunterschied liegen. Viele junge Menschen sind mit dem stets voranschreitenden digitalen Wandel aufgewachsen. Personen im Alter von 50 Jahre haben diesen erst viel später im Leben erlebt. Dazu muss gesagt werden, dass es hierbei nur um

Mittelwerte handelt und durchaus mehrere Fälle mit hohem Domänenwissen im höheren Alter vorkommen. Ähnlich sieht es für die Berufsfelder aus. Auch hier sticht die Gruppe von Informatikern und Studenten besonders mit hohem Domänenwissen heraus. Personen aus dem technologischen Berufsfeld können selbst Entwickler von Software sein, welche wiederum gut dazu in der Lage sind, sich in fremde Softwaresysteme einzuarbeiten. Auch Studenten nutzen mehr oder weniger digitale Medien für Studienzwecke [1] und haben dadurch ein hohes Domänenwissen für Office- oder Browser-Software. Womöglich ist aber auch hier das Alter entscheidend.

Die letzte Erkenntnis ist, dass die textuelle Form für Erklärungen in beiden Domänen am besten abschneidet. Die meisten Teilnehmer bevorzugen es, kurze Texte mit Erklärungen zu lesen und verzichten somit auf interaktive Varianten der Erklärung. Auch wenn sich durch eine Schulung oder ein Video Anwendungsbeispiele als Förderung des Verständnisses durchführen lassen, könnte dies in den meisten Fällen im Konflikt mit der vorhandenen Zeit der Nutzer stehen. Kurze und wenig detaillierte Erklärungstexte sind am beliebtesten.

Kapitel 7

Validität

Im Folgenden werden die *Threats to Validity* nach Wohlin [37] auf die Inhalte der Arbeit angewandt. Dabei werden die Threats in *Construct*, *Conclusion* und *External* kategorisiert. Construct beschreibt die Fehler, welche schon beim Planen des Experiments passiert sind oder sich nicht vermeiden ließen. Ein Threat to Conclusion Validity lässt sich der Robustheit der Ergebnisse zuordnen. Für External zählt, ob die Ergebnisse auch im anderen Kontext gelte.

Die eigenhändig aufgestellte Bewertungsmetrik des objektiven Domänenwissens und alle ordinalen Skalierungen wurden nicht von einer zweiten Instanz überprüft. Trotz mehrfacher Prüfung der daraus entstanden Daten ist es möglich, dass Fehler übersehen und dadurch das Ergebnis beeinflusst wurde. Die Ergebnisse sind auch von keiner zweiten Instanz geprüft worden, was die Wahrscheinlichkeit von Fehler hätte minimieren können. Es besteht in der Bewertungsmetrik des objektiven Domänenwissens und den Umfragedaten ein *Threat of Construct Validity*.

Die Einschränkung der Variablen aus der Korrelations- und Clusteranalyse kann das Ergebnis verfälschen. Auch wenn hier die Variablen mit dem Fokus auf Erklärbarkeit gewählt wurden, zeigen die Analysen, dass es immer noch unklar ist, welche Faktoren den Erklärungsbedarf wirklich beeinflussen. Hier liegt ein *Threat of Construct Validity* vor.

Durch die wenigen abgefragten Domänen lassen sich nur eingeschränkte Aussagen über das Verhalten des Erklärungsbedarfs zwischen den Domänen treffen. Die Auswahl dieser Domänen könnte die Endergebnisse ebenfalls stark beeinflussen. Hier liegt erneut ein *Threat of Construct Validity* vor.

Die Bonferroni-Korrektur senkt das Risiko eine Typ-I-Fehler (falsch, positiv) zu erzeugen, was einen Typ-II-Fehler (falsch, negativ) hervorrufen kann. Hier liegt also ein *Threat to Conclusion Validity* vor.

Die demografischen Daten aus der Umfrage bilden nicht den Schnitt der Gesellschaft ab. Das Durchschnittsalter der Umfrageteilnehmer liegt bei 33,5 Jahren. Ebenfalls ist die Verteilung der Geschlechter mit 11 Frauen,

46 Männern und einer diversen Person global gesehen nicht repräsentativ. In der beruflichen Verteilung sind 48 % Studenten, was die gesellschaftliche Verteilung ebenfalls nicht repräsentiert. Somit besteht ein *Threat to external Validatio*, da diese Ergebnisse nicht auf die Allgemeinheit übertragbar sind.

Die Teilnehmeranzahl mit 58 Personen ist zu klein, um generellere Aussagen über die Ergebnisse zu treffen. Auch hier besteht ein *Threat to external Validatio*.

Kapitel 8

Zusammenfassung und Ausblick

In diesem Kapitel wird ein abschließender Überblick über die gesamte Studie mit den wichtigsten Erkenntnissen aus den Analysen wiedergeben.

8.1 Zusammenfassung

Das Ziel dieser Arbeit war es, den Zusammenhang zwischen Domänenwissen und Erklärungsbedarf zu untersuchen. Zum Erreichen dieses Ziels wurde eine Online-Umfrage mit dem Fokus auf zwei Domänen durchgeführt. Dabei wurde abgefragt, welche Software die Teilnehmer häufig nutzen, wie sie ihr Domänenwissen einschätzen und wie häufig und wie tief sie für ihre genannte Software Erklärungen wünschen. Um das subjektive Domänenwissen der Teilnehmer effektiv zu prüfen, wurden Kurzfragen zu Software aus den Domänen Office und Browser gestellt, um ihr tatsächliches Verständnis in diesen Domänen zu erfassen. Zur Überprüfung von Zusammenhängen und Verbindungen in den Umfrageergebnissen wurden sowohl eine Korrelationsanalyse als auch eine Clusteranalyse durchgeführt. Aus beiden Analysen ist kein starker Zusammenhang zwischen Domänenwissen und dem Erklärungsbedarf festgestellt worden. Der Erklärungsbedarf wurde in fünf verschiedene Typen eingeteilt. Diese Typen zeigen, dass der Erklärungsbedarf entweder von Domäne zu Domäne unterschiedlich entsteht oder Personen einen grundlegend gleichen Erklärungsbedarf haben. Bei der Korrelations- und Clusteranalyse ist aufgefallen, dass sich das Domänenwissen von den demografischen Daten beeinflussen lässt, und dass mit dem Domänenwissen die Sicherheit in der Anwendung steigt. Eine zusätzliche Erkenntnis ist, dass die Teilnehmer Erklärungstexte mit wenig detaillierten Erklärungen als Erklärungsform priorisieren.

8.2 Ausblick

Neben einem größeren Datensatz wäre es in Zukunft sinnvoller, mehrere Domänen in der Online-Studie abzufragen. Im übrigen sollte ein größerer Teil der Gesellschaft durch die Umfragedaten abgedeckt werden, um generellere Aussagen treffen zu können und eine vielseitigere Clusterbildung zu ermöglichen. Die Auswahl der Domänen sowie die Überprüfung des objektiven Domänenwissens sollten unter der Leitung von Experten geschehen. Mit dem Einschreiten einer zweiten oder dritten Instanz verliert die Durchführung an *Threats to Validity*. Der Umfang an Variablen sollte in beiden Analysen steigen, um ein feineres Bild zu erzeugen und möglichst viele Faktoren, die sich auf den Erklärungsbedarf auswirken könnten, einzubinden.

Beim Chi-Quadrat-Test sind viele Einträge der nominal oder ordinal Skalar selten vorgekommen. Für derartig kleine Stichproben empfiehlt es sich, den Kolmogorov-Smirnov-Test auf Normalverteilung zu nutzen. Auch das Clusteringverfahren könnte durch eine logistische Regressionsanalyse auf Basis der Clusteranalyse erweitert werden, um mehr Einsichten zu bekommen.

Anhang A

Ein Anhang

A.1 Clusteranalyse

A.1.1 Silhouettenkoeffizienten

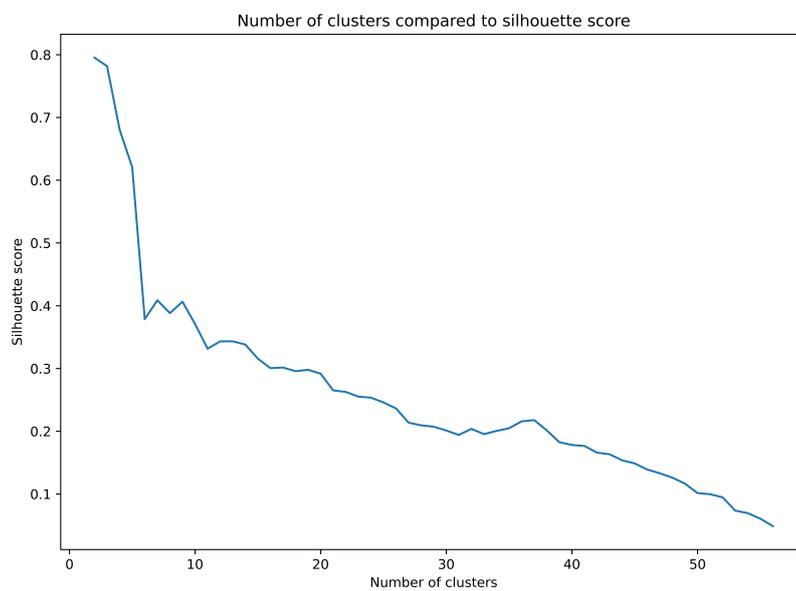


Abbildung A.1: Silhouettenkoeffizient der OfficeMain Clusteranalyse

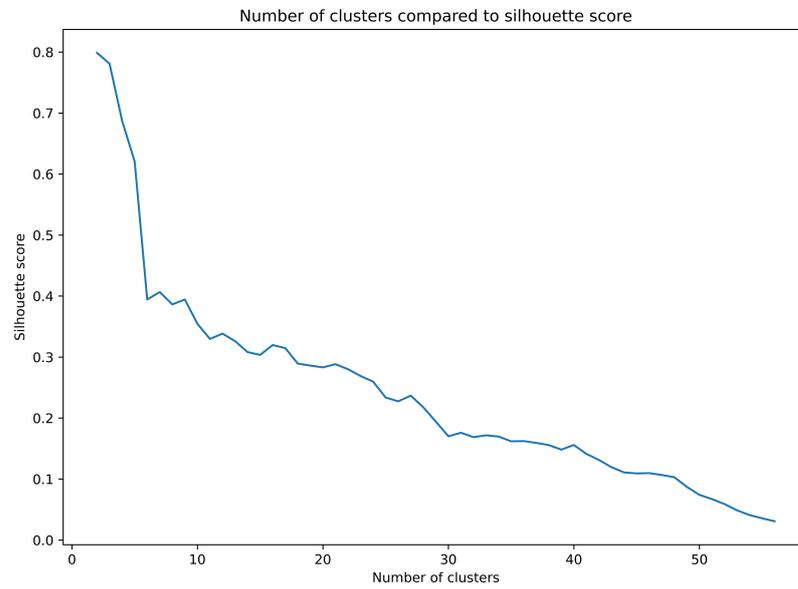


Abbildung A.2: Silhouettenkoeffizient der BrowserMain Clusteranalyse

A.1.2 Dendrogramme

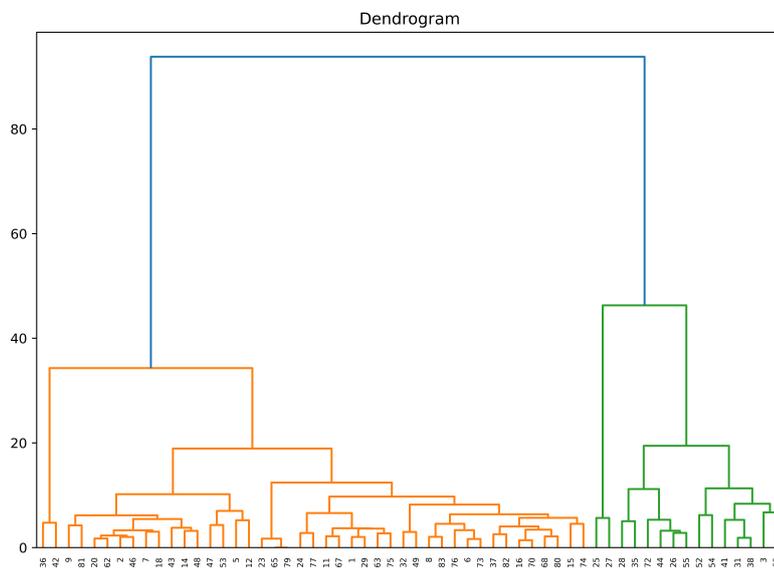


Abbildung A.3: Dendrogramm der OfficeMain Clusteranalyse

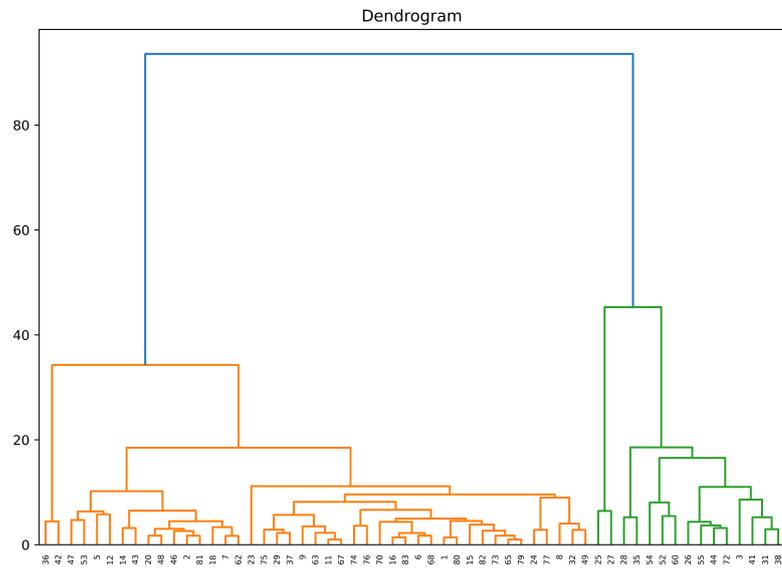


Abbildung A.4: Dendrogramm der BrowserMain Clusteranalyse

A.1.3 Streudiagramme

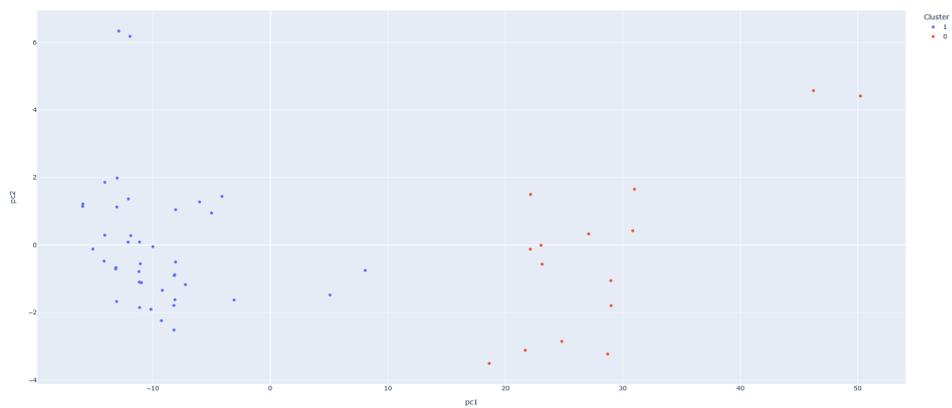


Abbildung A.5: Streudiagramm der OfficeMain Clusteranalyse

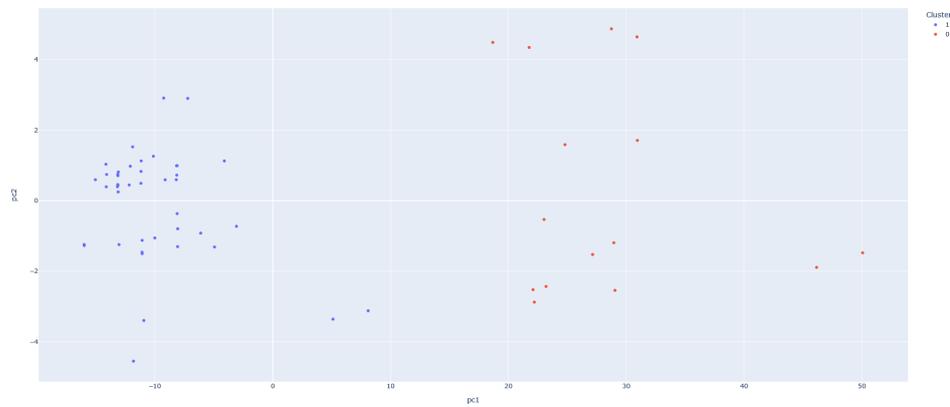


Abbildung A.6: Streudiagramm der BrowserMain Clusteranalyse

Literaturverzeichnis

- [1] M. Bond, V. I. Marín, C. Dolch, S. Bedenlier, and O. Zawacki-Richter. Digital transformation in german higher education: student and teacher perceptions and usage of digital media. *International journal of educational technology in higher education*, 15(1):1–20, 2018.
- [2] T. Carstensen. Zwischen homeoffice, neuer präsenz und care. die räumliche und digitale neuordnung von arbeit. *WSI-Mitteilungen*, 76(1):3–9, 2023.
- [3] L. Chazette. Requirements engineering für erklärbare systeme. In *Ausgezeichnete Informatikdissertationen 2022 (Band D23)*, pages 31–40. Gesellschaft für Informatik e.V., Bonn, 2023.
- [4] L. Chazette, W. Brunotte, and T. Speith. Exploring explainability: a definition, a model, and a knowledge catalogue. In *2021 IEEE 29th international requirements engineering conference (RE)*, pages 197–208. IEEE, 2021.
- [5] L. Chung and J. C. S. do Prado Leite. *On Non-Functional Requirements in Software Engineering*, pages 363–379. Springer Berlin Heidelberg, Berlin, Heidelberg, 2009.
- [6] J. Cohen. A power primer. *American Psychological Association*, 112(1):155–159, 1992.
- [7] H. Deters. Criteria and metrics for the explainability of software. *Masterarbeit, Gottfried Wilhelm Leibniz Universität Hannover, FG Software Engineering*, 2022.
- [8] H. Deters, J. Droste, and K. Schneider. A means to what end? evaluating the explainability of software systems using goal oriented heuristics. In *Proceedings of the 27th International Conference on Evaluation and Assessment in Software Engineering*, page 329–338, New York, NY, USA, 2023. Association for Computing Machinery.
- [9] J. Droste, H. Deters, M. Obaidi, and K. Schneider. Explanations in everyday software systems: Towards a taxonomy for explainability

- needs. In *2024 IEEE 32nd international requirements engineering conference (RE)*. IEEE, 2024.
- [10] M. Fechner. Konzept und implementierung einer komponente zur untersuchung des erklärungsbedarfs von software. *Bachelorarbeit, Leibniz Universität Hannover*, 2023.
- [11] M. Felderer, R. Reussner, and B. Rumpe. Software engineering und software-engineering-forschung im zeitalter der digitalisierung: Ein beitrag über das aktuelle und zukünftige selbstverständnis des software engineering. *Informatik Spektrum*, 44:82–94, 2021.
- [12] N. J. Gogtay and U. M. Thatte. Principles of correlation analysis. *Journal of the Association of Physicians of India*, 65(3):78–81, 2017.
- [13] U. Grimmer and H.-J. Mucha. Datensegmentierung mittels clusteranalyse. *Data Mining: Theoretische Aspekte und Anwendungen*, pages 109–141, 1998.
- [14] M. Halkidi, Y. Batistakis, and M. Vazirgiannis. On clustering validation techniques. *Journal of intelligent information systems*, 17:107–145, 2001.
- [15] A. Jackszis. Diskriminierung im machine learning und erklärbarkeit von algorithmen. *Arbeitspapier des Lehrgebiets Datenbanken und E-Business, Hochschule Düsseldorf Fachbereich Medien*, 2021.
- [16] J. Janssen, W. Laatz, J. Janssen, and W. Laatz. Clusteranalyse. *Statistische Datenanalyse mit SPSS: Eine anwendungsorientierte Einführung in das Basissystem und das Modul Exakte Tests*, pages 489–519, 2013.
- [17] N. Jösten. Clusteranalyse zwischen stimmung und erklärbarkeitsanforderungen von nutzern. *Bachelorarbeit, Gottfried Wilhelm Leibniz Universität Hannover, FG Software Engineering*, 2024.
- [18] M. G. Kendall, S. F. Kendall, and B. B. Smith. The distribution of spearman’s coefficient of rank correlation in a universe in which all rankings occur an equal number of times. *Biometrika*, pages 251–273, 1939.
- [19] M. Kohl and F. Münch. Statistik teil 5: Kontingenzkoeffizienten. *Kardiotechnik*, page 12, 2023.
- [20] U. Kuckartz, S. Rädiker, T. Ebert, and J. Schehl. *Korrelation: Zusammenhänge identifizieren*, pages 207–237. VS Verlag für Sozialwissenschaften, Wiesbaden, 2013.

- [21] U. Kuckartz, S. Rädiker, T. Ebert, J. Schehl, U. Kuckartz, S. Rädiker, T. Ebert, and J. Schehl. Korrelation: Zusammenhänge identifizieren. *Statistik: Eine verständliche Einführung*, pages 207–237, 2013.
- [22] M. Kuhnke. Identifizierung von erklärungsbedarf via user-feedback-analyse. *Masterarbeit, Leibniz Universität Hannover, Fachgebiet Software Engineering*, 2020.
- [23] D. Kupczyk. Automatisierte detektion von erklärungsbedarf in nutzerfeedback zu software. *Masterarbeit, Gottfried Wilhelm Leibniz Universität Hannover, FG Software Engineering*, 2023.
- [24] T. S. Madhulatha. An overview on clustering methods, 2012. 1205.1117, archivePrefix: arXiv, cs.DS.
- [25] R. Malan, D. Bredemeyer, et al. Functional requirements and use cases. *Bredemeyer Consulting*, pages 335–1653, 2001.
- [26] F. Nielsen. *Introduction to HPC with MPI for Data Science*. Springer, 2016.
- [27] N. Peter, T. Kleinjung, R. Lippuner, B. Boecking, P. Brueggemann, and B. Mazurek. Deutschsprachige versionen des tinnitus functional index. *Hno*, 70(3):187–192, 2022.
- [28] S. Pichai and D. Hassabis. Introducing gemini: our largest and most capable ai model. <https://blog.google/technology/ai/google-gemini-ai/#performance>, 2023. Abruf am 20.03.2024, 16:40 Uhr.
- [29] J. Rapp. Einbindung von domänenexperten in die interaktive verfeinerung von clustering-resultaten. *Masterarbeit, Universität Stuttgart, Institut für Parallele und Verteilte Systeme*, 2022.
- [30] P. J. Rousseeuw. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of computational and applied mathematics*, 20:53–65, 1987.
- [31] K. Siebertz, D. van Bebber, T. Hochkirchen, K. Siebertz, D. v. Bebber, and T. Hochkirchen. Korrelationsanalyse. *Statistische Versuchsplanung: Design of Experiments (DoE)*, pages 381–394, 2017.
- [32] K. Stange. Ueber einen zweiseitigen test für die korrelationszahl einer zweidimensionalen normalverteilung. *Statistische Hefte*, 14(3):206–236, 1973.
- [33] K. Stapel and K. Schneider. Flow-methode-methodenbeschreibung zur anwendung von flow. *arXiv preprint arXiv:1202.5919*, 2012.

- [34] R. van Solingen (Revision), V. Basili (Original article, 1994 ed.), G. Caldiera (Original article, 1994 ed.), and H. D. Rombach (Original article, 1994 ed.). *Goal Question Metric (GQM) Approach*. John Wiley & Sons, Ltd, 2002.
- [35] T. Walter and T. Walter. Der webclient—browser. *Kompodium der Web-Programmierung: Dynamische Web-Sites*, pages 63–70, 2008.
- [36] E. W. Weisstein. Bonferroni correction. *Herausgeber [https://mathworld.wolfram.com/-A Wolfram Web Resource](https://mathworld.wolfram.com/-A%20Wolfram%20Web%20Resource). <https://mathworld.wolfram.com/BonferroniCorrection.html>*, 2004. Abruf am 27.03.2024, 03:50 Uhr.
- [37] C. Wohlin, P. Runeson, M. Höst, M. C. Ohlsson, B. Regnell, and A. Wesslén. *Experimentation in software engineering*. Springer Science & Business Media, 2012.
- [38] B. Zhang. Regression clustering. In *Third IEEE International Conference on Data Mining*, pages 451–458. IEEE, 2003.
- [39] U. Zürich. Pearson chi-quadrat-test (kontingenzanalyse). https://www.methodenberatung.uzh.ch/de/datenanalyse_spss/zusammenhaenge/pearsonzush.html, 2023. Abruf am 19.03.2024, 19:00 Uhr.