

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

Clusteranalyse zwischen Stimmung und Erklärbarkeitsanforderungen von Nutzern

**Clusteranalysis of User Sentiment and Explainability
Requirements**

Bachelorarbeit

im Studiengang Informatik

von

Niklas Jösten

Prüfer: Prof. Kurt Schneider

Zweitprüfer: Dr. Jil Klünder

Betreuer: Martin Obaidi

Hannover, 26.02.2024

Erklärung der Selbstständigkeit

Hiermit versichere ich, dass ich die vorliegende Bachelorarbeit selbstständig 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 26.02.2024

Niklas Jösten

Zusammenfassung

In der sich ständig verändernden digitalen Welt, geprägt von einer zunehmenden Informationsflut, gewinnt effektive Kommunikation zunehmend an Bedeutung. Dies gilt insbesondere für Software, wo das Verständnis des Erklärungsbedarfs von Personen entscheidend ist, um ein optimales Nutzungserlebnis zu gewährleisten.

Das Hauptziel dieser Bachelorarbeit bestand darin, einen Zusammenhang zwischen dem Erklärungsbedarf, der Stimmung und den demografischen Daten zu identifizieren. Zur Erreichung dieses Ziels wurden eine Korrelationsanalyse unter Verwendung von Cramer's V sowie eine agglomerative Clusteranalyse durchgeführt.

Die Ergebnisse der Korrelationsanalyse zeigten höchstens einen geringen Zusammenhang zwischen dem Erklärungsbedarf und dem Stimmungsbild einer Person. Die Untersuchung des Zusammenhangs zwischen dem Erklärungsbedarf und den demografischen Daten ergab maximal einen schwachen Zusammenhang. Durch die Clusteranalyse konnten Cluster unter Berücksichtigung der demografischen Daten gebildet werden, aus denen unterschiedliche Merkmale identifiziert wurden. Es wurde festgestellt, dass zwischen den Clustern ein deutlicher Altersunterschied besteht. Jüngere Befragte zeigten einen höheren Erklärungsbedarf bei der Interaktion mit der Software und dem Systemverhalten, während ältere Befragte mehr Interesse an Metainformationen und dem Betrieb der Software zeigten.

Abstract

Clusteranalysis of User Sentiment and Explainability Requirements

In the constantly evolving digital world, characterized by an increasing flood of information, effective communication is gaining importance. This is particularly true for software, where understanding the explanatory needs of individuals is crucial to ensure an optimal user experience.

The main goal of this bachelor thesis was to identify a correlation between the explanatory needs, mood, and demographic data. To achieve this objective, a correlation analysis using Cramer's V and an agglomerative cluster analysis were conducted.

The results of the correlation analysis indicated, at most, a weak correlation between the explanatory needs and the mood of an individual. The examination of the relationship between explanatory needs and demographic data revealed, at most, a weak connection. Through the cluster analysis, clusters were formed considering demographic data, from which different characteristics were identified. It was observed that there is a significant age difference between the clusters. Younger respondents exhibited a higher need for explanations, particularly in interacting with the software and system behavior, while older respondents showed more interest in meta-information and the operation of the software.

Inhaltsverzeichnis

1	Einleitung	1
1.1	Problemstellung	1
1.2	Lösungsansatz	1
1.3	Struktur der Arbeit	2
2	Grundlagen	3
2.1	Erklärbarkeit	3
2.2	Stimmung	4
2.3	Korrelationsanalyse	7
2.4	Explorative Faktorenanalyse	10
2.5	Dimensionsreduktion	11
2.6	Hierachische Clusteranalyse	11
2.7	Datensatz	12
2.8	Python Bibliotheken	12
3	Verwandte Arbeiten	15
3.1	Abgrenzung zu verwandten Arbeiten	16
4	Studiendesign	19
4.1	Formulierung der Forschungsfragen	19
4.2	Datensatz	22
4.3	Korrelationsanalyse	23
4.4	EFA	23
4.5	Clusteranalyse	23
5	Ergebnis	27
5.1	Ergebnisse der Korrelationsanalyse	27
5.2	Ergebnisse der EFA	29
5.3	Ergebnisse der Clusteranalyse	30
6	Diskussion	37
6.1	Beantwortung der Forschungsfragen	37
6.2	Interpretation	39

7	Validität	41
8	Zusammenfassung und Ausblick	43
8.1	Zusammenfassung	43
8.2	Ausblick	43
A	Ein Anhang	45
A.1	Clusteranalyse	45
A.1.1	Silhouettenkoeffizient	45
A.1.2	Dendrogramme	49
A.1.3	Streudiagramme mit Boxplots der Cluster	53
A.1.4	Kruskal-Wallis Test	71

Kapitel 1

Einleitung

In der digitalen Welt des ständigen Wandels und einer ansteigenden Informationsflut wird eine effektive Kommunikation immer relevanter[6]. Insbesondere bei Software ist es entscheidend den Erklärungsbedarf von Personen zu verstehen um ein optimales Nutzungserlebnis zu gewährleisten[9]. Vor diesem Hintergrund rückt die Beziehung zwischen Erklärungsbedarf und Stimmung der Nutzenden in den Fokus dieser Arbeit. Diese zielt darauf ab den Zusammenhang zwischen Erklärungsbedarf und Stimmung zu untersuchen. Zudem wird in diesem Kontext der Zusammenhang von Erklärungsbedarf und demografischen Daten untersucht.

1.1 Problemstellung

Erklärungsbedarf und das automatisierte erkennen dessen wird im Bereich des Software Engineerings, insbesondere im Bereich des Requirements Engineerings, viel beforscht[6, 7, 21, 37]. Diesen in Zusammenhang zu Stimmung und demografischen Daten zu setzen ist dennoch weitgehend unerforscht. Diese Arbeit soll beginnen diese Lücke zu schließen.

Diese Forschung ist unter Anderem nützlich für Softwareentwickler, Unternehmen, Bildungseinrichtungen und soziale Netzwerke. Diese sind in der Lage, Kundenfeedback auf die untersuchten Eigenschaften zu analysieren. Diese können, entsprechend der gefundenen Eigenschaften, in Nutzergruppen eingeteilt werden. Mit diesen ist es möglich, das Nutzungserlebnis der Kunden zu personalisieren und an deren Anforderungen anzupassen, wodurch das Nutzungserlebnis gesteigert werden könnte.

1.2 Lösungsansatz

Für die Untersuchung wird eine Korrelationsanalyse mit Cramer's V durchgeführt. Durch diese wird untersucht, inwieweit Erklärungsbedarf und das Stimmungsbild von Personen zusammenhängen. Außerdem wird der

Zusammenhang zwischen Erklärungsbedarf und den demografischen Daten untersucht, um weitere mögliche Einflüsse zu erkennen. Zudem wird eine agglomerative Clusteranalyse durchgeführt, um Gruppen in den Befragten zu identifizieren. Diese Gruppen könnten als Grundlage für die Nutzergruppen genutzt werden.

1.3 Struktur der Arbeit

Im Folgenden wird die Struktur der Arbeit kurz zusammengefasst:

Diese Arbeit ist wie folgt strukturiert. In Kapitel 2 werden alle benötigten Konzepte und genutzten Methoden erklärt. Das Kapitel 3 beschäftigt sich mit den verwandten Arbeiten. Diese umfassen andere Arbeiten und Paper die das Thema Erklärbarkeit, Stimmung, Korrelation und Clusteranalyse behandeln. Kapitel 4 werden die Forschungsfragen aufgestellt und es wird die Durchführung der Korrelationsanalyse, der explorativen Faktor Analyse und der Clusteranalyse erklärt. In Kapitel 5 werden die Ergebnisse der Analysen dargestellt. Das Kapitel 6 beantwortet die Forschungsfragen mit den Ergebnissen aus Kapitel 5 und interpretiert diese. In Kapitel 7 werden die möglichen *Threats to Validity* nach Wohlin[40] betrachtet. Im letzten Kapitel 8 werden die Ergebnisse nochmals zusammengefasst. Zusätzlich wird ein Ausblick auf zukünftige Arbeiten gegeben.

Kapitel 2

Grundlagen

In diesem Kapitel werden alle nötigen Grundlagen erklärt, welche für diese Arbeit benötigt werden. Es wird Erklärungsbedarf und Stimmung definiert, die genutzten Methoden erklärt. Zum Ende wird der Datensatz erklärt, auf dem diese Arbeit aufbaut ist.

2.1 Erklärbarkeit

Erklärbarkeit gehört zum Teilgebiet des Requirement Engineerings[7]. Dabei wird Erklärbarkeit den nicht-funktionalen Anforderungen zugeordnet. Das bedeutet, es muss erst ein System existieren, das man auf Erklärbarkeit testen kann. Dabei kann Erklärbarkeit wie folgt definiert werden. Nach Köhl et al.[22] macht der Zugang zu Erklärungen das System erklärbar. Diese Definition lässt nach Chazette et al.[7] außer acht was erklärt werden soll. Außerdem haben sie herausgearbeitet, dass es keine eindeutige Definition gibt[7]. Dazu fanden Köhl et al. heraus, dass es nicht nur ein Arte des Erklärungsbedarfes gibt[22]. Basierend darauf haben Chazette et al.[7] in ihrer Arbeit mithilfe den Ergebnissen von Köhl et al.[22], Definitionen aus anderer Literatur, und Ergebnissen aus Workshops mit Philosophen und Psychologen eine abstrakte Definition erarbeitet, die wie folgt aussieht:

A system S is explainable with respect to an aspect X of S relative to an addressee A in context C if and only if there is an entity E (the explainer) who, by giving a corpus of information I (the explanation of X), enables A to understand X of S in C .

Die Erklärungen könnten an verschiedenen Stellen im System notwendig sein[7]:

Interne Logik: Wenn der Nutzer nicht versteht, wie das System Informationen verarbeitet oder Muster erkennt.

Interne Modelle: Falls der Benutzer nachvollziehen möchte, wie bestimmte Modelle im System funktionieren, um Vorhersagen oder Antworten zu generieren.

Intentionen: Wenn Unklarheit darüber besteht, warum das System eine bestimmte Aktion durchführt oder eine spezifische Antwort gibt.

Verhalten: Bei Fragen zur Reaktion des Systems auf bestimmte Eingaben oder Anfragen des Benutzers.

Entscheidungen: Wenn der Nutzer verstehen möchte, warum das System eine bestimmte Entscheidung getroffen hat, insbesondere wenn es um komplexe Probleme geht.

Die Form und Ausführlichkeit der Erklärung variiert dabei abhängig vom Kontext, in dem die Frage aufkommt, sowie von den spezifischen Anforderungen, die eine umfassende Antwort erfordern[7]. Mittels Fragebögen oder Nutzerstudien kann überprüft werden, ob der Adressat die vermittelten Informationen verstanden hat[7]. In dem noch nicht veröffentlichtem Projekt softXplain wird Erklärbarkeit in expliziten und impliziten Erklärungsbedarf eingeteilt¹. Expliziter Erklärungsbedarf wird in diesem Projekt durch die W-Fragen oder Satzbausteine wie „muss/will/möchte“ in Verbindung mit „herausfinden/verstehen“ identifiziert². Die befragte Person bringt also explizit zum Ausdruck, dass sie etwas verstehen möchte³. Impliziter Erklärungsbedarf wird durch Stichwörter wie „nicht ersichtlich“, „komisch“, „plötzlich“, „fraglich“ oder Ähnliches identifiziert⁴. Die befragte Person muss Unverständnis oder Überraschung zum Ausdruck bringen, bzw. muss deutlich werden, dass etwas nicht nachvollziehbar war⁵.

2.2 Stimmung

Stimmung setzt sich aus Emotionen zusammen. Shaver et al.[33] sortiert in seinem Klassifikationsmodell die Emotionen entsprechend einer positiven oder negativen Stimmung. Man sieht in der übergeordneten Kategorie die Aufteilung in positive und negative Emotionen. Diese teilen sich weiter

¹softXplain: Anforderungen für selbst-erklärende Software <https://gepris.dfg.de/gepris/projekt/470146331>

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

³softXplain: Anforderungen für selbst-erklärende Software <https://gepris.dfg.de/gepris/projekt/470146331>

⁴softXplain: Anforderungen für selbst-erklärende Software <https://gepris.dfg.de/gepris/projekt/470146331>

⁵softXplain: Anforderungen für selbst-erklärende Software <https://gepris.dfg.de/gepris/projekt/470146331>

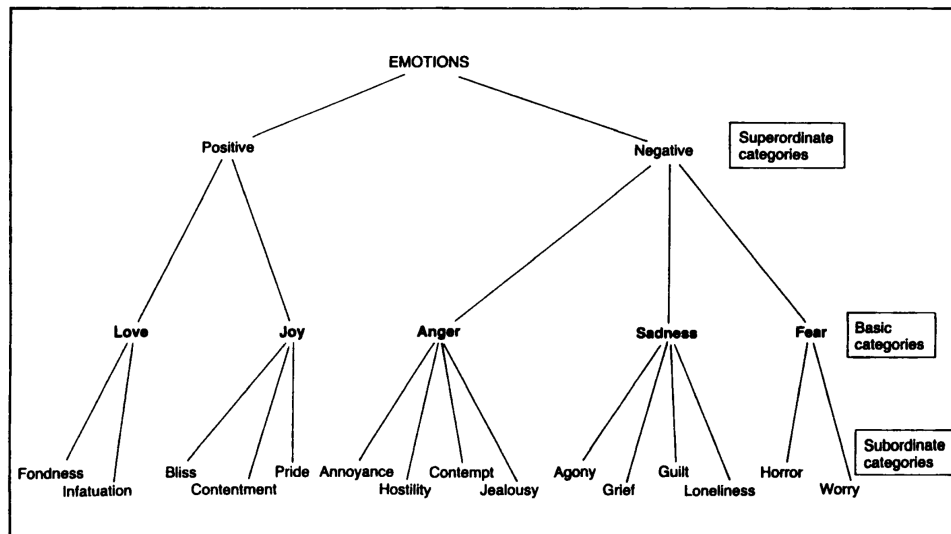


Abbildung 2.1: Vereinfachtes Modell nach Shaver et al.[11]

auf in die Basiskategorien. Die positive Kategorie setzt sich aus Liebe(Love) und Freude(Joy) zusammen. Die negative Kategorie teilt sich in Wut(Anger), Trauer(Sadness) und Angst(Fear). Im Bereich des Software Engineering wird dieses Modell zur automatischen Erkennung von Stimmung in Datensätzen[28, 29]. Alternativ kann für die Bestimmung der Stimmung eine Stimmungsskala wie in 2.1 von Underwood und Froming[36] verwendet werden. Diese ist von Bohner et al.[5] ins deutsche übersetzt worden.

Die befragten Personen müssen diese 15 Aussagen anhand einer hier sieben stufigen Skala bewerten⁶. Die Ergebnisse werden für die beiden Subskalen aufaddiert. SS steht dabei für die Stimmung und R steht für Reaktivität. Reaktivität beschreibt wie schnell sich die Stimmung einer Person verändert. Die Items mit einer negativen Polung müssen entsprechend folgender Formel vor der Aufsummierung umgepolt werden.

$$\text{umgepolter Wert} = \text{Länge der Skala} + 1 - \text{Wert} \quad (2.1)$$

Der Wertebereich für Stimmung liegt zwischen 9-63. Ein höherer Wert steht entsprechend für eine positivere Stimmung und ein niedrigerer Wert für eine negativere Stimmung. Der Wertebereich der Reaktivität liegt zwischen 6-42. Ein höherer Wert steht dabei für eine schnellere Stimmungsänderung und ein niedrigerer Wert steht für eine konstantere Stimmung.

⁶softXplain: Anforderungen für selbst-erklärende Software <https://gepris.dfg.de/gepris/projekt/470146331>

Nr.	Item	Polung	Subskala
1	Manchmal pendelt meine Stimmung mehrmals zwischen glücklich und traurig in einer einzigen Woche.	+	R
2	Ich fühle mich meist ziemlich fröhlich.	+	S
3	Meine Stimmung ist oft bedrückt.	-	S
4	Ich sehe im Allgemeinen mehr die Sonnenseiten des Lebens.	+	S
5	Verglichen mit meinen Freunden gehen meine Stimmungen weniger rauf und runter.	-	R
6	Ich bin selten in wirklicher Hochstimmung.	-	S
7	Manchmal schwankt meine Stimmung sehr schnell hin und her.	+	R
8	Ich fühle mich meist so, als ob ich vor Freude übersprudeln würde.	+	S
9	Meine Stimmungen sind sehr konsistent; sie ändern sich fast nie.	-	R
10	Ich halte mich für eine glückliche Person.	+	S
11	Verglichen mit meinen Freunden denke ich weniger optimistisch über das Leben.	-	S
12	Ich bin eine Person, die sich oft ändert.	+	R
13	Ich bin nicht so fröhlich wie die meisten Leute.	-	S
14	Ich bin weniger von meinen Stimmungen abhängig als die meisten Leute, die ich kenne.	+	R
15	Meine Freunde scheinen oft zu glauben, dass ich unglücklich bin.	-	S

Tabelle 2.1: Übernommen von Bohner et al. [5]

2.3 Korrelationsanalyse

Korrelation ist eine statistische Maßeinheit und beschreibt wie sich zwei verschiedene Variablen zueinander verhalten[13]. Korrelationsanalyse ist folglich eine angewandte statistische Methode. Dabei gibt es verschiedenen Arten. Der Pearson Korrelationskoeffizient[10] oder der Spearman Korrelationskoeffizient[35] unterscheiden zum Beispiel zwischen positiver und negativer Korrelation. Der Wertebereich begrenzt sich bei beiden auf -1 bis +1[20]. Dabei steht -1 für eine perfekte negative Korrelation und bedeutet, dass wenn der Wert der einen Variable steigt der Wert der anderen fällt[20]. Dabei steht +1 für eine perfekte positive Korrelation und bedeutet, dass wenn der Wert der einen Variable steigt, der Wert der anderen auch steigt[20]. Wenn der Korrelationskoeffizient bei 0 ist, liegt keine Korrelation vor[20]. Für Pearson muss der Datensatz aus kontinuierlichen, nominalen Daten bestehen, zwischen welchen ein linearer Zusammenhang vorhanden sein muss, damit dieser erkannt wird[20]. Bei Spearman muss der Datensatz aus ordinal oder metrischen skalierten Daten bestehen und es ist ausreichend wenn ein monotonen Wachstum vorliegt[20]. Andere Ansätze können mit nominal skalierten Daten arbeiten. Üblicherweise wird dafür Cramer's V benutzt. Der Nachteil bei Cramer's V ist, dass nur eine Aussage, ob eine Korrelation vorliegt, getroffen werden kann[20]. Dabei ist nicht zuzuordnen, ob es sich um eine positive oder negative Korrelation handelt[20]. Cramer's V kann dabei Werte von 0 für keinen Zusammenhang bis 1 für perfekter Zusammenhang annehmen[20]. Um Cramer's V zu berechnen muss zuerst eine Matrix aufgestellt werden[20]. Die Werte in der Matrix sind die Anzahl der Einträge im Datensatz, welche sowohl auf die linke, als auch die obere Kategorie zutreffen. Hier einmal an dem Beispiel Erklärungsbedarf(e/i/y/n) und gesprochener Sprache veranschaulicht:

Indifferenztabelle				
	klein	mittel	groß	
vormittags	7,368421053	16,57894737	11,05263158	35
nachmittags	6,315789474	14,21052632	9,473684211	30
abends	6,315789474	14,21052632	9,473684211	30
	20	45	30	95

Tabelle 2.2: Anzahl der je auf beide Kategorien zutreffenden Einträge

Mit dieser Matrix stellt man zuerst die Indifferenzmatrix auf. Die Werte in dieser Matrix sind so angepasst als wären die Werte normalverteilt. Dafür wird folgende Formel für alle Werte die nicht Summen der Spalten oder Zeilen sind benutzt:

$$\text{Wert} = \text{Summe der Spalte} * \text{Summe der Zeile} / \text{Summe aller Werte} \quad (2.2)$$

Indifferenztabelle				
	klein	mittel	groß	
vormittags	7,368421053	16,57894737	11,05263158	35
nachmittags	6,315789474	14,21052632	9,473684211	30
abends	6,315789474	14,21052632	9,473684211	30
	20	45	30	95

Tabelle 2.3: Indifferenztabelle für das Beispiel

Mit diesen Matrizen kann Chi-Quadrat[4], für alle Werte die nicht Summen der Spalten oder Zeilen sind, berechnet werden. Dafür verwendet man die Formel:

$$X^2 = \sum_{i=1}^m \sum_{j=1}^k \frac{(n_{ij} - \tilde{n}_{ij})^2}{\tilde{n}_{ij}}$$

X^2	Chi-Quadrat
m	Gesamtanzahl der Zeilen
k	Gesamtanzahl der Spalten
n_{ij}	absolute Häufigkeit der Merkmalskombination in i-Zeile und j-Spalte (beobachteter Wert)
\tilde{n}_{ij}	erwarteter Wert der absoluten Häufigkeit der Merkmalskombination in i-Zeile und j-Spalte

Tabelle 2.4: Bedeutung der Variablen aus der Formel für Chi-Quadrat

Vereinfacht in natürlicher Sprache bedeutet das:

$$X^2 = \sum \frac{(\text{beobachteter} - \text{erwarteter Wert})^2}{\text{erwarteter Wert}}$$

Mit Chi-Quadrat berechnet, kann zusätzlich der Chi-Quadrat-Test[30] durchgeführt werden. Dieser gibt an, ob die gewählten Kategorien unabhängig voneinander sind[30]. Dafür wird Nullhypothese aufgestellt, welche besagt, dass es keinen Zusammenhang zwischen den Variablen gibt[30]. Um diese zu widerlegen kann der p-Wert berechnet werden[30]. Sollte der p-Wert gleich oder unterhalb des gewählten Signifikanzniveaus α liegen, wird die Nullhypothese verworfen[30]. Zusätzlich kann bei einer Vielzahl von Korrelationsanalysen eine Bonferroni-Korrektur durchgeführt werden[39]. Diese ist eine Methode um das Signifikanzniveau an die Anzahl der durchgeführten Tests anzupassen[39]. Diese Methode wird angewandt um der alpha-Fehler-Inflation, dem Typ-I-Fehler, entgegenzuwirken[39]. Das Ziel ist es dadurch die Zahl an Ablehnung tatsächlich korrekter Nullhypothesen (falsch-positiv Befunde) zu verringern[39]. Das adjustierte Signifikanzniveau berechnet sich mit der Formel:

$$\alpha_{adj} = \frac{\alpha}{\text{Anzahl durgeführter Tests}} \quad (2.3)$$

Um Cramer's V zu berechnen, werden die berechneten Werte für Chi-Quadrat, die Anzahl an Freiheitsgraden (Degrees of Freedom) und die

Gesamtzahl der Stichprobe in folgende Formel eingesetzt:

$$V = \sqrt{\frac{x^2}{n \times (df(c, r))}} \quad (2.4)$$

V	Cramer's V
X^2	Chi-Quadrat
n	Gesamtanzahl (der Stichprobe)
df	Degrees of Freedom, berechnet durch: $df = \min(c-1, r-1)$
c	Anzahl an Zeilen der Matrix
r	Anzahl an Spalten der Matrix

Tabelle 2.5: Bedeutung der Variablen aus der Formel für Cramer's V

In dem Beispiel würde das wie folgt aussehen:

Chi Quadrat	klein	mittel	groß	
vormittags	0,939849624	0,705931495	3,314536341	4,96031746
nachmittags	0,274122807	0,043859649	0,029239766	0,347222222
abends	0,274122807	1,247563353	3,223684211	4,74537037
	1,488095238	1,997354497	6,567460317	10,05291005
df	2		Cramers V	0,230021853

Tabelle 2.6: Beispielergebnis für Cramer's V

Das Ergebnis aus den Berechnungen wird nach folgender Tabelle, genau wie Pearsons Korrelationskoeffizient, interpretiert:

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 2.7: Übernommen aus „Statistik: eine verständliche Einführung“[20]

In diesem Fall besteht kein Zusammenhang zwischen Erklärungsbedarf(e/i/y/n) und der gesprochenen Sprache.

2.4 Explorative Faktorenanalyse

Explorative Faktorenanalyse (EFA) ist eine statistische Methode, mit der man den zugrundeliegenden Strukturen zwischen Variablen identifizieren kann[8, 24]. EFA basiert dabei auf der Konfirmatorische Faktorenanalyse. Das Ziel von EFA ist es, Faktoren zu finden, die den beobachteten Zusammenhang zwischen Variablen aufzeigen. Zuerst muss eine Korrelationsmatrix aufgestellt werden um die Beziehung zwischen den Variablen zu bekommen. Bevor man eine EFA durchführt, sollte evaluiert werden, ob die Daten für ein Faktorenanalyse geeignet sind[34]. Dafür kann Bartlett's Test und den Kaiser-Meyer-Olkin (KMO) Test verwendet werden[34]. Bei Bartlett's Test handelt es sich um einen statistischen Test. Es wird die Nullhypothese aufgestellt, dass die Korrelationsmatrix eine Identitätsmatrix ist[34]. Es wird dabei die Korrelationsmatrix mit der Identitätsmatrix verglichen und das Ergebnis ist der p-Wert[2, 26]. Bei einem p-Wert unter 0.05 ist eine statistisch signifikante Korrelation in den Daten und man verwirft die Nullhypothese[34]. Die Daten sind somit für eine EFA geeignet[34]. Wenn der p-Wert über 0.05 liegt kann die Nullhypothese nicht verworfen. Die Daten

demnach nicht für eine EFA geeignet. Der KMO Test berechnet die Eignung des Datensatzes für die EFA. Das Ergebnis wird in einem Wertebereich von 0 bis 1 angegeben[19]. Interpretiert wird dieser wie folgt:

KMO	Interpretation
$0.8 \leq KMO \leq 1$	Die Daten sind gut geeignet.
$0.7 \leq KMO < 0.8$	Die Daten sind mittelmäßig geeignet.
$0.6 \leq KMO < 0.7$	Die Daten sind mäßig geeignet.
$0 \leq KMO < 0.6$	Die Daten sind nicht gut geeignet.

Tabelle 2.8: Interpretation des Ergebnisses aus dem KMO Test[34]

2.5 Dimensionsreduktion

In der Wissenschaft werden immer mehr Daten aus der realen Welt gesammelt und abgespeichert[12]. Diese Mengen an Daten auszuwerten kann einen sehr großen Aufwand darstellen[12]. Um die Auswertung zu vereinfachen, forscht der Bereich Data Science unter anderem an der Dimensionsreduktion von Datensätzen[12]. Das Ziel ist es, die Dimensionalität (die Anzahl Variablen pro Datenpunkt) zu reduzieren, sodass die wichtigsten Informationen erhalten bleiben[12]. Zum Beispiel kann dafür die Hauptkomponentenanalyse verwendet werden. Hauptkomponentenanalyse oder im englischen Principal Component Analysis (PCA) beschreibt eine statistische Methode mit der die Dimension eines hochdimensionalen Datensatzes reduziert werden kann. Dabei werden aus den Daten durch lineare Kombination neue Variablen gebildet, welche Trends und Muster aus den originalen Daten beibehalten. Diese Variablen sind die Hauptkomponenten[14]. Hauptkomponenten bestehen aus linearen Kombinationen der ursprünglichen Variablen. Um diese lineare Kombination zu erhalten wird zuerst eine Kovarianzmatrix aufgestellt. Aus dieser werden die Eigenvektoren berechnet und nach der maximalen Varianz angeordnet[31]. Somit erhält man eine annähernde Abbildung der originalen Daten durch die Hauptkomponenten[14]. Diese können zur weiteren Analyse verwendet werden oder zur Visualisierung[12].

2.6 Hierarchische Clusteranalyse

Hierarchische Clusteranalyse ist ein unüberwachtes maschinelles Lernverfahren[27]. Ziel ist es zusammenhängende Strukturen in den Daten zu finden[27]. Da die Hierarchische Clusteranalyse unüberwacht arbeitet, muss die Anzahl der resultierenden Cluster nicht bei Beginn feststehen[16]. Hierbei unterscheidet man zwischen agglomerativem und divisivem Clustering[16]. Das agglomerative Verfahren arbeitet nach dem Bottom-Up Prinzip. Dabei bildet jeder Punkt ein eigenes Cluster. Diese

werden schrittweise abhängig vom Fusionierungsalgorithmus bestimmten Cluster zusammengeführt[16]. Das divisive Verfahren funktioniert nach dem Top-Down Prinzip[16]. Alle Datenpunkte bilden am Anfang eine großes Cluster[16]. Dieses wird schrittweise in neue Cluster aufgeteilt[16].

Beide Verfahren benötigen ein Distanzmaß um die Distanz zwischen den Clustern zu errechnen. Üblicherweise wird zum Beispiel die euklidische Distanz oder die Manhattan-Distanz[3] verwendet. Zusätzlich muss ein Fusionierungsalgorithmus (engl. Linkage) genutzt werden[27]. Dieser bestimmt anhand der errechneten Distanz, welche Cluster beim agglomerativen Verfahren zusammengeführt oder beim divisivem Verfahren aufgeteilt werden. Linkage-Verfahren mit einem einfachen Kriterium sind zum Beispiel: Single-Linkage(minimaler Abstand), Complete-Linkage(maximale Distanz) oder Average-Linkage(durchschnittliche Distanz)[27]. Ein komplexeres Linkage Verfahren ist Ward's minimales Varianzkriterium (engl. Ward-Linkage)[38]. Bei diesem werden die zwei Cluster, die zusammen die niedrigste Varianz aufweisen, zusammengeschlossen, bzw. geteilt[27].

2.7 Datensatz

Als Grundlage für alle durchgeführten Analysen dienten die Ergebnisse aus dem Projekt softXplain aus dem Jahr 2022 geleitet von Kurt Schneider⁽⁷⁾. Bei dem Projekt wurden die 83 befragten Personen nach drei Software Beispielen gefragt bei denen die Befragten Erklärungsbedarf hatten. Diese Antworten wurden untersucht und nach dem Erklärungsbedarf gelabelt. Es wurde unterschieden zwischen explizitem (e) und implizitem (i) Erklärungsbedarf und keinem Erklärungsbedarf (n), sowie für den Fall „es besteht Erklärungsbedarf, aber dieser ist nicht eindeutig festzustellen“ (y). Bei der Identifikation von explizitem oder implizitem Erklärungsbedarf wurde zusätzlich die Art des Erklärungsbedarfs gelabelt.

Zusätzlich wurden 15 Fragen(siehe Tabelle 2.12.1) zu Stimmung und Reaktivität nach Bohner et al.[5] gestellt, welche diese aus der Arbeit von Underwood et al. [36] ins deutsche Übertragen haben.

2.8 Python Bibliotheken

Die Python Skripte, welche für die Durchführung der Tests, Analysen und zur Virtualisierung verwendet wurden, benutzen folgende Bibliotheken:

pandas⁸ erlaubt Tabellen aus Excel einzulesen und als sogenannten

⁷softXplain: Anforderungen für selbst-erklärende Software <https://gepris.dfg.de/gepris/projekt/470146331>

⁸<https://pandas.pydata.org/>

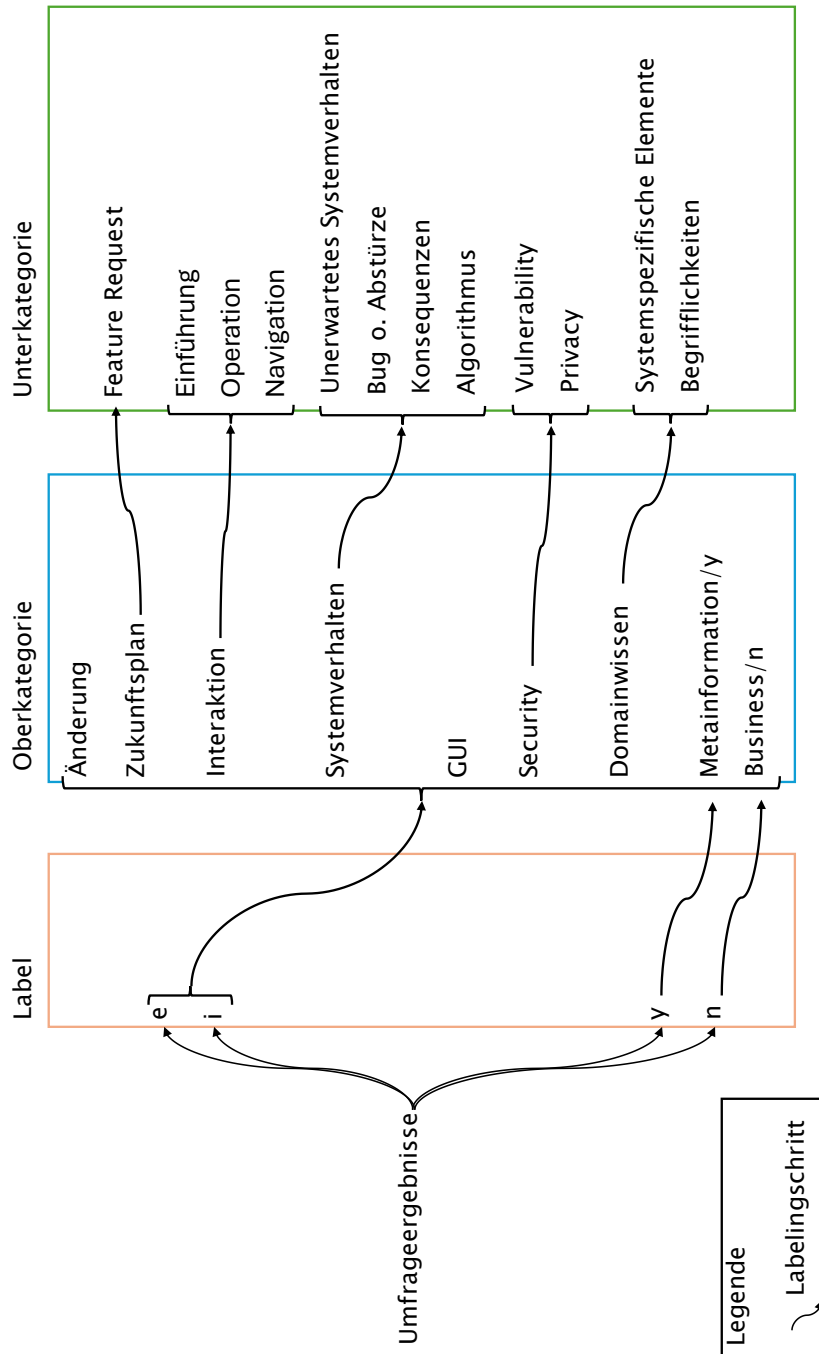


Abbildung 2.2: Veranschaulichung zur Aufteilung des Erklärungsbedarfes

DataFrame zu speichern. Dieser kann von Python bearbeitet werden. Hier wurde diese Bibliothek hauptsächlich genutzt um den Datensatz zusammenzuführen.

numpy ⁹ fügt Python multidimensionale Arrays hinzu.

matplotlib.pyplot ¹⁰ ist Teil der matplotlib. Diese stellt Funktionen zur Verfügung um Diagramme zu erstellen.

sklearn ¹¹ ist eine Bibliothek für maschinelles Lernen. Diese stellt Methoden für prädiktive Datenanalyse zur Verfügung. Hier wird diese Bibliothek für die Durchführung der Clusteranalyse, sowie zur Durchführung einer PCA genutzt.

scipy ¹² stellt die Algorithmen für Optimierung, Integration, Interpolation, Eigenwertprobleme, algebraische Gleichungen, Differentialgleichungen, Statistik und weiter bereit. Genutzt wird diese Bibliothek hier für die Erstellung des Dendrogramms und für die Durchführung des Kruskal-Wallis h-Tests.

FactorAnalyzer ¹³ wird für die EFA verwendet. Diese bietet Funktionen um eine EFA mit verschiedenen Rotationen durchzuführen. Außerdem enthält FactorAnalyzer Funktionen um Bartletts Test und den KMO Test durchzuführen.

plotly.express ¹⁴ wird genutzt um die gefundenen Cluster nach der Clusteranalyse darzustellen. Diese Diagramme bieten den Vorteil, dass Informationen beim Hovering über einen Punkt in Diagramm einfacher abgelesen werden können.

⁹<https://numpy.org/>

¹⁰https://matplotlib.org/stable/api/pyplot_summary.html

¹¹<https://scikit-learn.org/stable/index.html>

¹²<https://scipy.org/>

¹³<https://pypi.org/project/factor-analyzer/>

¹⁴<https://plotly.com/python/plotly-express/>

Kapitel 3

Verwandte Arbeiten

Dieses Kapitel behandelt verwandte Arbeiten, welche sich mit den Themen Erklärungsbedarf, Stimmung, Korrelationsanalyse oder Clusteranalyse beschäftigen, die auch in dieser Arbeit Anwendung gefunden haben.

Chazette et al.[7] haben in dem Paper „Exploring Explainability: A Definition, a Model, and a Knowledge Catalogue“ zusammengefasst, wie Erklärungsbedarf definiert ist. Darin bieten sie eine neue Definition für Erklärungsbedarf, ein Modell sowie einen Katalog für Erklärungsbedarf.

Unterbusch et al.[37] haben in ihrem Paper untersucht, welchen Erklärungsbedarf Nutzende von verschiedenen Apps haben und wie dieser automatisiert erkannt werden kann. Die Autoren haben eine Taxonomie für die Erklärungsbedürfnisse entwickelt, für die ein Datensatz von 1730 App Reviews aus acht verschiedenen Apps manuell codiert wurde. Außerdem haben sie Ansätze zur automatisierten Erkennung von Erklärungsbedarf untersucht.

Kupczyk[21] hat sich in seiner Masterarbeit mit der automatisierten Erkennung von Erklärungsbedarf beschäftigt. Dafür hat er verschiedene Machine Learning und Deep Learning Modelle trainiert und verwendet. Zusätzlich hat er eine regelbasierte Methode entwickelt, welche stellenweise in einem unausgeglichenen Datensatz eine höhere Präzision erreichen konnte als die Deep Learning Modelle. Allerdings war der Recall bei dem regelbasierten Modell schlechter. Die Deep Learning Modelle lieferten insgesamt stabilere Ergebnisse, aber die Performance bezüglich der Präzision auf dem unausgewogenem Datensatz viel im Vergleich zu der regelbasierten Methode schlechter aus.

Lin et al.[23] haben versucht ein automatisches Empfehlungsprogramm zu entwickeln, welches Softwarebibliotheken empfiehlt. Dafür haben sie ein eigenes Stimmungsanalysetool entwickelt. Dieser Ansatz ist allerdings gescheitert. Stattdessen haben sie mehrere Stimmungsanalysetools auf die Genauigkeit in verschiedenen Datensätzen getestet. Sie sind zu dem Ergebnis gekommen, dass keines der getesteten Stimmungsanalysetools wegen starken

Limitierungen für eine Nutzung geeignet waren.

Obaidi et al.[29] haben eine Studie zu Stimmungsanalysetools in Bereich des Software Engineerings durchgeführt. Ziel dieses Papers ist es Entwicklern und Interessengruppen bei der Auswahl von Stimmungsanalysetools für einen spezifischen Zweck zu helfen. Sie fassen dafür 106 wissenschaftliche Arbeiten zusammen, welche sich mit Anwendungsbereich, Zweck, verwendeten Datensätzen, Ansätzen zur Entwicklung von Stimmungsanalysetools, Verwendung bereits vorhandener Tools und Herausforderungen, mit denen Forscher konfrontiert sind, befassen.

Martensen[25] hat in seiner Bachelorarbeit untersucht, durch welche Faktoren die Stimmung in Entwicklerteams beeinflusst wird. Dabei ist Martensen zu dem Ergebniss gekommen, dass zwischen den Faktoren und der abgegebenen Stimmung kein Zusammenhang besteht. Außerdem verschiebt eine mehrfache Teilnahme den Fokus und es kommt zu einer stärkeren Abweichung bei der Stimmungsabgabe. Zudem führen Konflikte innerhalb des Teams zu einer negativeren Wahrnehmung und Stimmungsabgabe. Genutzt hat Martensen dabei unter anderem die Spearman Rang Korrelation.

Bachmann und Bernstein[1] haben in ihrem Paper untersucht, ob eine Korrelation zwischen Qualität und Eigenschaften der Prozessdaten und dem Fehlerbehebungsprozess besteht und ob eine Korrelation zwischen der an Prozessdaten gemessene Prozessqualität und der Produktqualität besteht. Sie konnten zeigen, dass es eine Korrelation gibt zwischen Qualität und Eigenschaften der Prozessdaten und dem Fehlerbehebungsprozess. Zusätzlich konnten sie auch zeigen dass die Prozessdatenqualitätsmessungen einen Einfluss auf die Produktqualität hat.

Herrmann[17] hat in seiner Masterarbeit die Wahrnehmung von 94 Informatikern mithilfe explorativer Datenanalyse untersucht. Das Ziel war es einen Anhaltspunkt zu schaffen, wie eine Kalibrierung von Stimmungsanalysetools für die Nutzung in industriellen Anwendungen aussehen könnte. Dafür wurde die Stimmung der Befragten untersucht. Für die Analysen hat er die Daten auf eine Korrelation untersucht. Die Ergebnisse hat er für eine Clusteranalyse verwendet.

3.1 Abgrenzung zu verwandten Arbeiten

Erklärungsbedarf und Stimmung sind Bereiche des Software Engineerings in denen viel geforscht wird. In den vorgestellten Papern werden diese Bereiche nicht zusammen untersucht. Chazette et al.[7], Unterbusch et al.[37] und Kupczyk[21] haben in deren Papern nur den Erklärungsbedarf und die automatisierte Erkennung dessen untersucht. Obaidi et al.[29], Lin et al.[23], Martensen[25] und Herrmann[17] haben Stimmungsanalysetools untersucht und entwickelt. Das Ziel dieser Arbeit ist es diese beiden Bereichen zu verbinden und auf eine Korrelation zu untersuchen. Als

Grundlage dient ein aus dem Projekt softXplain¹ erstellter Datensatz. Um dieses Ziel zu erreichen werden verschiedene Methoden angewandt. Es wird eine Korrelationsanalyse mit Cramer's V durchgeführt um eine Korrelation zwischen Erklärungsbedarf und Stimmung, Reaktivität und anderen demografischen Daten zu untersuchen. Diese Daten stammen aus der Umfrage, die während dem Projekt softXplain² durchgeführt wurde. Außerdem wird eine EFA und eine Clusteranalyse auf dem Datensatz durchgeführt um darin zusammenhängende Faktoren und Gruppen zu finden.

¹softXplain: Anforderungen für selbst-erklärende Software <https://gepris.dfg.de/gepris/projekt/470146331>

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

Kapitel 4

Studiendesign

In diesem Kapitel werden die Forschungsfragen aufgestellt und die genutzten Methoden um diese zu beantworten erklärt. Das erste Unterkapitel beschreibt woher die Daten für den verwendeten Datensatz stammen und wie dieser für diese Arbeit angepasst wurde. Das zweite Unterkapitel beschreibt das Vorgehen für die Korrelationsanalyse mit Cramer's V und erklärt warum diese Methode gewählt wurde. Das dritte Unterkapitel befasst sich mit dem Vorgehen bei der explorativen Faktor Analyse. Das letzte Unterkapitel erklärt warum sich für die agglomerative Clusteranalyse entschieden wurde und wie diese durchgeführt wurde.

4.1 Formulierung der Forschungsfragen

In dieser Arbeit wird der Erklärungsbedarf und das Stimmungsbild der befragten Personen aus dem Projekt softXplain¹ untersucht. Ziel ist es einen Zusammenhang zwischen dem Erklärungsbedarf und dem Stimmungsbild einer Person zu identifizieren. Zusätzlich wird der Zusammenhang von Erklärungsbedarf und demografischen Daten einer Person untersucht. Außerdem wird der Datensatz auf mit einer Clusteranalyse untersucht. Die Eigenschaften der gefundenen Cluster werden anschließend ausgewertet und verglichen.

RQ1: *Wie stark ist der Zusammenhang zwischen dem Erklärungsbedarf und dem Stimmungsbild der befragten Personen?*

RQ1.1: *Wie stark ist der Zusammenhang zwischen dem Erklärungsbedarf und der Stimmung der befragten Personen?*

RQ1.2: *Wie stark ist der Zusammenhang zwischen dem Erklärungsbedarf und der Reaktivität der befragten Personen?*

¹softXplain: Anforderungen für selbst-erklärende Software <https://gepris.dfg.de/gepris/projekt/470146331>

- RQ2:** *Wie stark ist der Zusammenhang zwischen dem Erklärungsbedarf und den demografischen Daten der befragten Personen?*
- RQ2.1:** *Wie stark ist der Zusammenhang zwischen dem Erklärungsbedarf und dem Alter der befragten Personen?*
- RQ2.2:** *Wie stark ist der Zusammenhang zwischen dem Erklärungsbedarf und der Generation der befragten Personen?*
- RQ2.3:** *Wie stark ist der Zusammenhang zwischen dem Erklärungsbedarf und dem Geschlecht der befragten Personen?*
- RQ2.4:** *Wie stark ist der Zusammenhang zwischen dem Erklärungsbedarf und dem Beruf der befragten Personen?*
- RQ2.5:** *Wie stark ist der Zusammenhang zwischen dem Erklärungsbedarf und der gesprochenen Sprache der befragten Personen?*
- RQ3.1:** *Wie viele Cluster ergeben sich aus den untersuchten Variablen?*
- RQ3.2:** *Welche Eigenschaften unterscheiden die identifizierten Cluster voneinander?*

Für die Beantwortung der **RQ1** und **RQ2** wurden mehrere Hypothesen aufgestellt 4.1.

Diese Hypothesen sollen durch die Korrelationsanalyse mit Cramer's V beantwortet werden. Es werden Korrelationsanalysen für die in Tabelle 4.2 aufgeführten Paarungen durchgeführt:

Für die Beantwortung von **RQ3** wurde eine agglomerative Clusteranalyse durchgeführt.

4.2 Datensatz

Ich habe zuerst die Daten aus dem Projekt softXplain² geordnet und in eine Datei zusammengeführt. Dafür habe ich alle gelabelten Kategorien des Erklärungsbedarfes aus MaxQDA in mehrere Excel-Dateien exportiert. Außerdem wurden die IDs, Antworten der befragten Personen, sowie Apps bei denen der Erklärungsbedarf bestand als je eine weitere Excel-Datei exportiert. Diese wurden in mehreren Schritten über IDs, Antworten und Software mit dem gelabelten Erklärungsbedarf zusammengeführt. Zusätzlich zu der MaxQDA Datei enthält softXplain zwei weitere Excel-Dateien. Die Datei mit den vollständigen Umfrageergebnissen enthält alle Antworten, die während der Umfrage zu Erklärungsbedarf, Stimmung und demografischen Daten beantwortet wurden. In der anderen Datei wurden die Antworten, mit der entsprechenden ID und Software, nach der explizitem (e), implizitem (i), „es besteht Erklärungsbedarf, aber dieser ist nicht eindeutig festzustellen“ (y) und keinem (n) Erklärungsbedarf gelabelt. Diese Label konnten entsprechend der ID, Software und Antwort hinzugefügt werden. Die Stimmungsdaten und die demografischen Daten aus der Datei mit den vollständigen Umfrageergebnissen konnten anhand der ID ergänzt werden. Zusätzlich musste bei Einträgen mit "Business/n" und "Metainformationen/y" das Label entsprechend angepasst werden, da diese im ersten Schritt des Labelings noch mit explizitem oder implizitem Erklärungsbedarf gelabelt waren. Anschließend konnte ich die Stimmung und Reaktivität für jede Person berechnen. Die Stimmung und Reaktivität wurden, wie in Grundlagen 2.2, umgepolt. An der Befragung haben insgesamt 83 Personen teilgenommen. Von diesen haben nicht alle Angaben zu Stimmung und Reaktivität ausgefüllt und wurden entsprechend aussortiert. Es blieben insgesamt 66 von den 83 Personen mit vollständigen Daten. Für EFA, PCA und die Clusteranalyse habe ich den Datensatz nochmals erweitert, da diese Verfahren nicht mit nominal skalierten Werten arbeiten können. Die nominalen Werte aus den Umfrageergebnissen des Projekts softXplain³, wie beispielsweise Geschlecht, Sprache oder Beruf, werden danach als „0“ und „1“ repräsentiert. Die „0“ steht dabei für „trifft auf diese Person nicht zu“ und die „1“ steht entsprechend für „trifft auf diese Person zu“. Für das Beispiel „Sprache“ wird dieser in die vorkommenden Ausprägungen „deutsch“ und „englisch“ aufgeteilt. Der Eintrag „deutsch“ enthält somit die Werte „0“ für spricht kein deutsch und „1“ entsprechend für spricht deutsch. Dieses Vorgehen wurde für die Einträge Geschlecht und Beruf durchgeführt.

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

³softXplain: Anforderungen für selbst-erklärende Software <https://gepris.dfg.de/gepris/projekt/470146331>

4.3 Korrelationsanalyse

In dieser Arbeit hab ich mich für die Cramer's V entschieden, da in dem Datensatz die Korrelation von Kategorien mit nominalen Daten untersucht wird, die keine natürliche Rangfolge haben. Cramer's V kann dabei Werte von 0 für keinen Zusammenhang bis 1 für perfekter Zusammenhang annehmen. Es kann daher nicht erkannt werden, ob es sich um eine negative oder positive Korrelation handelt. Die Berechnungen für Cramer's V wurden in Excel, wie in dem Beispiel aus Grundlagen, durchgeführt. Anschließend wird ein Bonferroni-Korrektur durchgeführt.

4.4 EFA

Um EFA durchzuführen habe ich ein Python Skript geschrieben. Um zu testen, ob der Datensatz für eine EFA verwendet werden kann, habe ich Bartlett's Test und den KMO Test durchgeführt. Bartlett's Test gibt dabei einen p-Wert zwischen 0 und 1 aus, der besagt ob die Variablen in den Daten miteinander korrelieren. Dabei wird die Identitätsmatrix mit der Korrelationsmatrix verglichen. Die Nullhypothese besagt dabei, dass die Korrelationsmatrix die selbe wie die Identitätsmatrix ist. Dies bedeutet, dass keine Korrelation vorliegt. Wenn der p-Wert unter 0.05 liegt kann diese Nullhypothese verworfen werden. Folglich können die Daten für eine EFA verwendet werden.

Der KMO Test liefert eine Aussage darüber, wie geeignet die Daten für eine EFA sind. KMO betrachtet dabei die Varianz innerhalb der Variablen. KMO gibt dabei einen Wert zwischen 0 und 1 an. Wenn der Wert größer als 0.6 ist, sind die Daten gut für die Faktoranalyse geeignet und es kann zu der EFA übergegangen werden. Wenn der Wert kleiner als 0.6 ist sollte keine EFA mit den Daten durchführen.

Sofern beide Tests zu dem Ergebnis kommen, dass die Daten für eine EFA geeignet sind, kann die EFA durchgeführt werden. Dafür wird das Modell auf die Daten angewandt und aus den Ergebnissen kann geschlussfolgert werden, wie viele Faktoren relevant sind. Diese wurden nach dem Kaiser-Guttman-Kriterium[15, 18] bestimmt. Dafür ist es nötig die Eigenwerte für jeden Faktor berechnen. Sofern der Eigenwert größer als „1“ ist kann der Faktor behalten werden. Die Anzahl der Faktoren kann in der Clusteranalyse als Startpunkt für die Anzahl an Clustern verwendet werden.

4.5 Clusteranalyse

Ich habe mich hier für die Agglomerative Clusteranalyse entschieden, da vor der Clusteranalyse noch nicht bekannt ist, wie viele Cluster gefunden werden können. Dieses Verfahren nimmt zum Start an, dass jeder Punkt ein

Cluster ist und führt dann die Cluster, die sich am nächsten sind, zusammen, bis alle in einem Cluster sind. Es werden insgesamt sechs Clusteranalysen durchgeführt. Um die Clusteranalysen durchzuführen, habe ich je ein Skript geschrieben. Das Skript führt eine vollständige Clusteranalyse durch, um ein Dendrogramm zu erstellen. Der Silhouettenkoeffizient[32] wird verwendet um festzustellen wie viele Cluster sinnvoll sind. Dieser gibt auf einer Skala von „0“ bis „1“ an wie gut die Zuordnung zu den Clustern ist. Die Anzahl, für die der Silhouettenkoeffizient am höchsten war, wird in der Clusteranalyse verwendet. Nach der Durchführung der Clusteranalyse wird eine PCA durchgeführt um die Dimensionalität auf zwei Dimensionen zu reduzieren. Dadurch lassen sich die Ergebnisse in einem Streudiagramm darstellen. Nach der Clusteranalyse wird ein Signifikanztest durchgeführt, um die Daten der Cluster auf ihre Signifikanz zu testen. Dafür wird der Kruskal-Wallis Test verwendet, da die Daten nicht normalverteilt sind. Außerdem kann der Kruskal-Wallis Test auf mehr als zwei Gruppen angewendet werden. Der Kruskal-Wallis Test stellt die Nullhypothese auf, dass es keinen Unterschied zwischen den Medianen der untersuchten Gruppen gibt. Dieses Skript gibt am Ende aus, ob die Nullhypothese verworfen werden kann. Der Ergebnis des Tests muss dabei kleiner als 0.05 sein. Um die Cluster zu visualisieren habe ich eine PCA durchgeführt um die Daten auf zwei Komponenten zu reduzieren. Dadurch können die gefundenen Cluster in einem Streudiagramm dargestellt werden. Die PCA wurde mit den bereitgestellten Methoden aus Python Bibliothek sklearn durchgeführt.

Nr.	Hypothese	RQ
$H1_1$	Es besteht eine Korrelation zwischen dem Erklärungsbedarf (ja/nein) und der Stimmung der befragten Personen.	RQ1
$H1_2$	Es besteht eine Korrelation zwischen dem Erklärungsbedarf (e,i,y,n) und der Stimmung der befragten Personen.	RQ1
$H1_3$	Es besteht eine Korrelation zwischen dem Erklärungsbedarf (Oberkategorien) und der Stimmung der befragten Personen.	RQ1
$H1_4$	Es besteht eine Korrelation zwischen dem Erklärungsbedarf (Unterkategorien) und der Stimmung der befragten Personen.	RQ1
$H2_1$	Es besteht eine Korrelation zwischen dem Erklärungsbedarf (ja/nein) und der Reaktivität der befragten Personen.	RQ1
$H2_2$	Es besteht eine Korrelation zwischen dem Erklärungsbedarf (e,i,y,n) und der Reaktivität der befragten Personen.	RQ1
$H2_3$	Es besteht eine Korrelation zwischen dem Erklärungsbedarf (Oberkategorien) und der Reaktivität der befragten Personen.	RQ1
$H2_4$	Es besteht eine Korrelation zwischen dem Erklärungsbedarf (Unterkategorien) und der Reaktivität der befragten Personen.	RQ1
$H3_1$	Es besteht eine Korrelation zwischen dem Erklärungsbedarf (ja/nein) und der Generation der befragten Personen.	RQ2
$H3_2$	Es besteht eine Korrelation zwischen dem Erklärungsbedarf (e,i,y,n) und der Generation der befragten Personen.	RQ2
$H3_3$	Es besteht eine Korrelation zwischen dem Erklärungsbedarf (Oberkategorien) und der Generation der befragten Personen.	RQ2
$H3_4$	Es besteht eine Korrelation zwischen dem Erklärungsbedarf (Unterkategorien) und der Generation der befragten Personen.	RQ2
$H4_1$	Es besteht eine Korrelation zwischen dem Erklärungsbedarf (ja/nein) und dem Geschlecht der befragten Personen.	RQ2
$H4_2$	Es besteht eine Korrelation zwischen dem Erklärungsbedarf (e,i,y,n) und dem Geschlecht der befragten Personen.	RQ2
$H4_3$	Es besteht eine Korrelation zwischen dem Erklärungsbedarf (Oberkategorien) und dem Geschlecht der befragten Personen.	RQ2
$H4_4$	Es besteht eine Korrelation zwischen dem Erklärungsbedarf (Unterkategorien) und dem Geschlecht der befragten Personen.	RQ2
$H5_1$	Es besteht eine Korrelation zwischen dem Erklärungsbedarf (ja/nein) und dem Beruf der befragten Personen.	RQ2
$H5_2$	Es besteht eine Korrelation zwischen dem Erklärungsbedarf (e,i,y,n) und dem Beruf der befragten Personen.	RQ2
$H5_3$	Es besteht eine Korrelation zwischen dem Erklärungsbedarf (Oberkategorien) und dem Beruf der befragten Personen.	RQ2
$H5_4$	Es besteht eine Korrelation zwischen dem Erklärungsbedarf (Unterkategorien) und dem Beruf der befragten Personen.	RQ2
$H6$	Es besteht eine Korrelation zwischen dem Erklärungsbedarf (e,i,y,n) und der Sprache der befragten Personen.	RQ2
$H7$	Es besteht eine Korrelation zwischen dem Erklärungsbedarf (e,i,y,n) und dem Alter der befragten Personen.	RQ2
$H8$	Es besteht eine Korrelation zwischen dem Erklärungsbedarf (ja/nein) und der Stimmungsabgabe der befragten Personen.	RQ1

Tabelle 4.1: Aufgestellte Hypothesen für die Korrelationsanalyse

Kategorie 1	Kategorie 2
Erklärungsbedarf (ja/nein)	Stimmung
Erklärungsbedarf (e,i,y,n)	Stimmung
Erklärungsbedarf (Oberkategorien)	Stimmung
Erklärungsbedarf (Unterkategorien)	Stimmung
Erklärungsbedarf (ja/nein)	Reaktivität
Erklärungsbedarf (e,i,y,n)	Reaktivität
Erklärungsbedarf (Oberkategorien)	Reaktivität
Erklärungsbedarf (Unterkategorien)	Reaktivität
Erklärungsbedarf (ja/nein)	Generation
Erklärungsbedarf (e,i,y,n)	Generation
Erklärungsbedarf (Oberkategorien)	Generation
Erklärungsbedarf (Unterkategorien)	Generation
Erklärungsbedarf (ja/nein)	Geschlecht
Erklärungsbedarf (e,i,y,n)	Geschlecht
Erklärungsbedarf (Oberkategorien)	Geschlecht
Erklärungsbedarf (Unterkategorien)	Geschlecht
Erklärungsbedarf (ja/nein)	Beruf
Erklärungsbedarf (e,i,y,n)	Beruf
Erklärungsbedarf (Oberkategorien)	Beruf
Erklärungsbedarf (Unterkategorien)	Beruf
Erklärungsbedarf (ja/nein)	Sprache
Erklärungsbedarf (e,i,y,n)	Sprache
Erklärungsbedarf (Oberkategorien)	Sprache
Erklärungsbedarf (Unterkategorien)	Sprache
Erklärungsbedarf (ja/nein)	Alter
Erklärungsbedarf (e,i,y,n)	Alter
Erklärungsbedarf (Oberkategorien)	Alter
Erklärungsbedarf (Unterkategorien)	Alter

Tabelle 4.2: Paarungen für die Korrelationsanalyse, Erklärungsbedarf (e, i, y, n) setzt sich aus explizitem (e), implizitem (i), „es besteht Erklärungsbedarf, aber dieser ist nicht eindeutig festzustellen“ (y) und keinem (n) Erklärungsbedarf zusammen

Kapitel 5

Ergebnis

In diesem Kapitel werden die Ergebnisse der Korrelationsanalyse beschrieben. Die Ergebnisse von Bartlett's Test und des KMO Tests werden dargestellt und zeigen, dass keine EFA durchgeführt werden konnte. Am Ende werden die Ergebnisse der Clusteranalyse dargestellt und die Eigenschaften der gefundenen Cluster analysiert.

5.1 Ergebnisse der Korrelationsanalyse

Die Ergebnisse der Korrelationsanalyse sind in der Tabelle 5.1 aufgeführt. In dieser sind die verglichenen Kategorien mit der dazugehörigen Stichprobengröße aufgelistet. Außerdem werden die Werte für Chi-Quadrat und der dazugehörige p-Wert. Der p-Wert wird für den Chi-Quadrat-Unabhängigkeitstest verwendet. Für diesen wurde die Nullhypothese „Die Merkmale von Kategorie 1 und Kategorie 2 sind stochastisch unabhängig“ aufgestellt. Diese wird verworfen wenn das Signifikanzniveau α unter 0.05 liegt. Das Signifikanzlevel wird durch die Bonferroni-Korrektur für die **RQ1** auf 0,00625 und **RQ2** auf 0,0025 angepasst um den Typ-I-Fehler zu verringern. Diese Werte ergeben sich durch eine Division des Signifikanzlevel von 0,05 mit der Anzahl aufgestellter Hypothesen pro Forschungsfrage. Die Nullhypothese für Erklärungsbedarf (ja/nein) und Reaktivität, Erklärungsbedarf (ja/nein) und Geschlecht, Erklärungsbedarf (Oberkategorien) und Sprache, sowie Erklärungsbedarf (ja/nein) und Alter werden nach der Bonferroni-Korrektur nicht mehr verworfen. Zusätzlich werden die Freiheitsgrade (df), die Werte für Cramer's V und die zugehörige Interpretation aufgelistet.

Es konnte insgesamt in 18 von 28 ein geringer Zusammenhang festgestellt werden. Der stärkste Zusammenhang mit einem Wert von 0,2960 besteht hier zwischen Erklärungsbedarf (Unterkategorien) und dem Geschlecht. Der geringste Zusammenhang besteht hier zwischen dem Erklärungsbedarf (ja/nein) und der Sprache mit einem Wert von 0,0240. Auffällig ist außer-

Kategorie 1	Kategorie 2	Größe	Chi Quadrat	p-Wert	H0	Bonferroni	df	Cramer's V	Interpretation	RQ
Erklärungsbedarf (ja/ nein)	Stimmung	845	0,4973	0,4807	nicht verwerfen	nicht verwerfen	1	0,0243	kein Zusammenhang	RQ1
Erklärungsbedarf (e.i.y.n)	Stimmung	845	11,984	0,0025	verwerfen	verwerfen	2	0,0842	kein Zusammenhang	RQ1
Erklärungsbedarf (Oberkategorien)	Stimmung	591	10,7648	0,0046	verwerfen	verwerfen	2	0,0954	kein Zusammenhang	RQ1
Erklärungsbedarf (Unterkategorien)	Stimmung	254	28,4269	0	verwerfen	verwerfen	2	0,2366	geringer Zusammenhang	RQ1
Erklärungsbedarf (ja/ nein)	Reaktivität	848	4,7903	0,0286	verwerfen	nicht verwerfen	1	0,0752	kein Zusammenhang	RQ1
Erklärungsbedarf (e.i.y.n)	Reaktivität	541	49,4675	0	verwerfen	verwerfen	3	0,1746	geringer Zusammenhang	RQ1
Erklärungsbedarf (Oberkategorien)	Reaktivität	591	67,9421	0	verwerfen	verwerfen	5	0,1516	geringer Zusammenhang	RQ1
Erklärungsbedarf (Unterkategorien)	Reaktivität	257	88,0612	0	verwerfen	verwerfen	5	0,2618	geringer Zusammenhang	RQ1
Erklärungsbedarf (ja/ nein)	Generation	1001	2,3595	0,1245	nicht verwerfen	nicht verwerfen	1	0,0486	kein Zusammenhang	RQ2
Erklärungsbedarf (e.i.y.n)	Generation	1001	65,0324	0	verwerfen	verwerfen	3	0,1472	geringer Zusammenhang	RQ2
Erklärungsbedarf (Oberkategorien)	Generation	710	47,6751	0	verwerfen	verwerfen	4	0,1296	geringer Zusammenhang	RQ2
Erklärungsbedarf (Unterkategorien)	Generation	291	43,7202	0	verwerfen	verwerfen	4	0,1938	geringer Zusammenhang	RQ2
Erklärungsbedarf (ja/ nein)	Geschlecht	1001	7,7654	0,0053	verwerfen	nicht verwerfen	1	0,0881	kein Zusammenhang	RQ2
Erklärungsbedarf (e.i.y.n)	Geschlecht	1001	36,5815	0	verwerfen	verwerfen	2	0,1352	geringer Zusammenhang	RQ2
Erklärungsbedarf (Oberkategorien)	Geschlecht	710	55,3042	0	verwerfen	verwerfen	2	0,1973	geringer Zusammenhang	RQ2
Erklärungsbedarf (Unterkategorien)	Geschlecht	291	50,9886	0	verwerfen	verwerfen	2	0,2960	geringer Zusammenhang	RQ2
Erklärungsbedarf (ja/ nein)	Beruf	704	33,2531	0	verwerfen	verwerfen	1	0,2173	geringer Zusammenhang	RQ2
Erklärungsbedarf (e.i.y.n)	Beruf	835	41,2115	0	verwerfen	verwerfen	3	0,1283	geringer Zusammenhang	RQ2
Erklärungsbedarf (Oberkategorien)	Beruf	710	134,32	0	verwerfen	verwerfen	3	0,1539	geringer Zusammenhang	RQ2
Erklärungsbedarf (Unterkategorien)	Beruf	291	102,5821	0	verwerfen	verwerfen	3	0,2320	geringer Zusammenhang	RQ2
Erklärungsbedarf (ja/ nein)	Sprache	1001	0,5747	0,4484	nicht verwerfen	nicht verwerfen	1	0,0240	kein Zusammenhang	RQ2
Erklärungsbedarf (e.i.y.n)	Sprache	1001	1,2927	0,2556	nicht verwerfen	nicht verwerfen	1	0,0359	kein Zusammenhang	RQ2
Erklärungsbedarf (Oberkategorien)	Sprache	710	5,4446	0,0196	verwerfen	nicht verwerfen	1	0,0876	kein Zusammenhang	RQ2
Erklärungsbedarf (Unterkategorien)	Sprache	291	14,9412	0,0001	verwerfen	verwerfen	1	0,2266	geringer Zusammenhang	RQ2
Erklärungsbedarf (ja/ nein)	Alter	978	7,7848	0,0053	verwerfen	nicht verwerfen	1	0,0892	kein Zusammenhang	RQ2
Erklärungsbedarf (e.i.y.n)	Alter	978	62,9083	0	verwerfen	verwerfen	3	0,1464	geringer Zusammenhang	RQ2
Erklärungsbedarf (Oberkategorien)	Alter	710	44,9723	0	verwerfen	verwerfen	6	0,1027	geringer Zusammenhang	RQ2
Erklärungsbedarf (Unterkategorien)	Alter	291	42,5877	0	verwerfen	verwerfen	6	0,1562	geringer Zusammenhang	RQ2

Tabelle 5.1: Ergebnisse der Korrelationsanalyse

dem, dass die Stichprobengröße von Erklärungsbedarf (Oberkategorien) zu Erklärungsbedarf (Unterkategorien) immer um mehr als die Hälfte sinkt. Dies lässt sich durch das Wegfallen der Werte für Metainformationen/*y* und *Business/n* erklären, da für diese keine Unterkategorie existiert. Zudem sieht man, dass statistisch unabhängig Kategorien nicht miteinander korrelieren. In den Fällen die stochastisch signifikant konnte in bis auf zwei Fällen ein geringer Zusammenhang festgestellt. Bei den zwei Fällen handelt es sich um die Korrelation zwischen Stimmung und Erklärungsbedarf (e,i,y,n), sowie Stimmung und Erklärungsbedarf (Oberkategorien).

5.2 Ergebnisse der EFA

Zuerst wurden die Daten in sechs Gruppen eingeteilt. Diese bestehen aus den in der Korrelationsanalyse betrachteten Daten, welche in Simple, Main und Sub aufgeteilt wurden und jeweils nochmal durch CS („Complete Set“) erweitert werden. Simple beschreibt die Gruppe bestehend aus dem explizitem (e), implizitem (i), „es besteht Erklärungsbedarf, aber dieser ist nicht eindeutig festzustellen“ (y) und keinem (n) Erklärungsbedarf und der Stimmung. Main beschreibt die Gruppe mit den gelabelten Oberkategorien des Erklärungsbedarfes und der Stimmung. Sub beschreibt die Gruppe bestehend aus den gelabelten Unterkategorien des Erklärungsbedarfes und der Stimmung. Der Zusatz CS steht für die Erweiterung durch die demografischen Daten.

Main	Stimmung, Reaktivität, <i>Änderung, Business/n, Domainwissen, GUI, Interaktion, Metainformation/y, Security, Systemverhalten, Zukunftsplan</i>
MainCS	Alter, de, en, männlich, weiblich, divers, Studium, Arbeit, Sonstiges, Stimmung, Reaktivität, <i>Änderung, Business/n, Domainwissen, GUI, Interaktion, Metainformation/y, Security, Systemverhalten, Zukunftsplan</i>
Simple	Stimmung, Reaktivität, e, i, y, n
SimpleCS	Alter, de, en, männlich, weiblich, divers, Studium, Arbeit, Sonstiges, Stimmung, Reaktivität, e, i, y, n
Sub	Stimmung, Reaktivität, <i>Algorithmus, Begrifflichkeiten, Bug o. Abstürze, Einführung, Konsequenzen, Navigation, Operation, Privacy, Systemspezifische Elemente, Unerwartetes Systemverhalten, Vulnerability</i>
SubCS	Alter, de, en, männlich, weiblich, divers, Studium, Arbeit, Sonstiges, Stimmung, Reaktivität, <i>Algorithmus, Begrifflichkeiten, Bug o. Abstürze, Einführung, Konsequenzen, Navigation, Operation, Privacy, Systemspezifische Elemente, Unerwartetes Systemverhalten, Vulnerability</i>

Tabelle 5.2: Begriffserklärung für durchgeführte Clusteranalysen

In der folgenden Tabelle sind die Testergebnisse für die sechs Gruppen aufgelistet. Bartlett’s Test4.4 kam dabei nur für die Gruppe Simple zu dem Ergebnis, dass die Daten nicht signifikant für eine EFA sind. Der KMO Test4.4 kam für alle Gruppen zu dem Ergebnis, dass die Daten nicht geeignet

sind für eine EFA.

EFA	Bartlett's Test	KMO Test
Main	signifikant	nicht adäquat
MainCS	signifikant	nicht adäquat
Simple	nicht signifikant	nicht adäquat
SimpleCS	signifikant	nicht adäquat
Sub	signifikant	nicht adäquat
SubCS	signifikant	nicht adäquat

Tabelle 5.3: Ergebnisse der Tests für eine EFA

Aufgrund dieser Ergebnisse wurde die EFA nicht durchgeführt.

5.3 Ergebnisse der Clusteranalyse

Es wurden insgesamt sechs agglomerative Clusteranalysen durchgeführt. Diese werden mit den bereits in der Korrelationsanalyse untersuchten Variablen durchgeführt. Für die Clusteranalyse habe ich die Daten wie für die EFA in sechs Gruppen aus Tabelle 5.2 aufgeteilt. Für jede dieser Clusteranalysen wurden die in die vier in den Grundlagen 2.6 beschriebenen Linkage Methoden getestet. Um die am besten geeignete Methode auszuwählen wurde der Silhouettenkoeffizient abhängig von den Clustern bestimmt. Dieser gibt die Güte der gefundenen Cluster an. Die Ergebnisse sind in folgender Tabelle zusammengefasst. Zusätzlich enthält die Tabelle die Ergebnisse der PCA. Diese geben den Prozentsatz der durch die Komponenten 1 (pc1) und 2 (pc2) abgedeckten Daten an.

	Linkage	Cluster	pc1	pc2	Silhouettenkoeffizient
Simple	average	2	52,57	27,96	0,464
SimpleCS	average	2	87,23	6,33	0,6085
Main	Complete	2	29,15	20,28	0,6134
MainCS	Complete	2	94,93	1,41	0,6615
Sub	Complete	2	25,75	20,28	0,5271
SubCS	single	2	97,05	0,63	0,7533

Tabelle 5.4: Übersicht über Linkage Methode und zugehörige Ergebnisse

Der Silhouettenkoeffizient wird nach der Tabelle 5.5 interpretiert. Demnach konnte eine starke Struktur in der SubCS Gruppe gefunden werden. Eine sinnvolle Struktur wurde für die Gruppen SimpleCS, Main, MainCS und Sub gefunden. Für die Gruppe Simple konnte nur eine schwache Struktur gefunden werden.

Für jede der sechs Gruppen wurde zusätzlich der Verlauf des Silhouettenkoeffizienten visualisiert. Aus diesen lässt sich ablesen, dass eine

SK	Interpretation
$0.70 < SK \leq 1.00$	Eine starke Struktur wurde gefunden
$0.50 < SK \leq 0.70$	Eine sinnvolle Struktur wurde gefunden
$0.25 < SK \leq 0.50$	Eine schwache, möglicherweise künstliche Struktur wurde gefunden
$< SK \leq 0.25$	Keine substanziale Struktur wurde gefunden

Tabelle 5.5: Wertebereiche des Silhouettenkoeffizient und dessen Interpretation nach Rousseeuw[32]

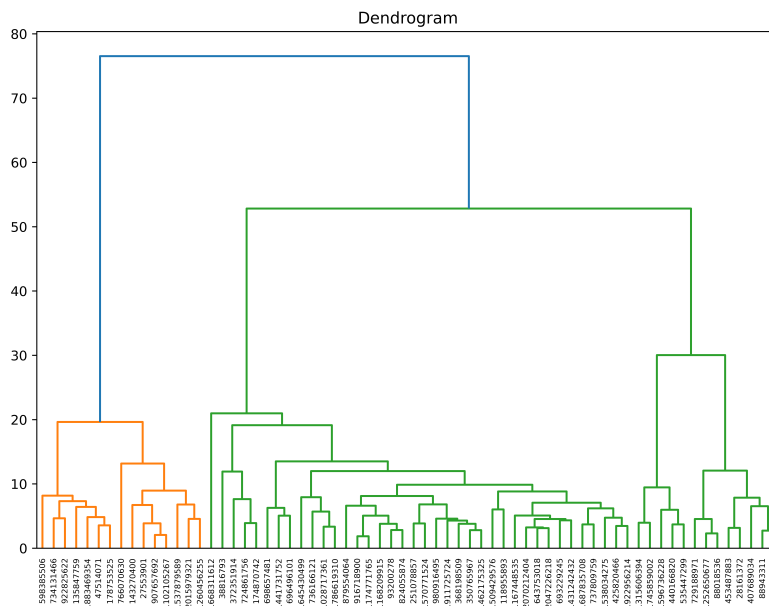


Abbildung 5.1: Dendrogramm für MainCS

Anzahl von zwei Clustern für alle Gruppen ein globales Maximum in den Diagrammen darstellen. Zur Untersuchung der gefundenen Cluster wurden diese als Dendrogramm und als Streudiagramm visualisiert. Das Dendrogramm für die Gruppe MainCS sieht wie folgt aus: Man sieht in diesem Fall eine klare Trennung zwischen einem kleineren Cluster bestehend aus 15 Personen auf der linken Seite und einem größeren Cluster bestehend aus den restlichen 51 Personen. Das Streudiagramm wird durch die Komponenten PC1 und PC2 aus 5.4 aufgespannt und sieht für diese Gruppe wie folgt aus:

Die beiden Cluster sind hier farblich gekennzeichnet. Hier sieht man, dass die Trennung der Cluster ab dem pc1 Wert von 20 stattgefunden hat. Auch anhand dieser Darstellung könnten die Cluster anders aufgeteilt werden. Der Bereich von einem pc1-Wert von 10 bis 20 könnte dem Cluster 1 zugeordnet werden, wodurch ein größerer Abstand zwischen den Clustern erreicht werden

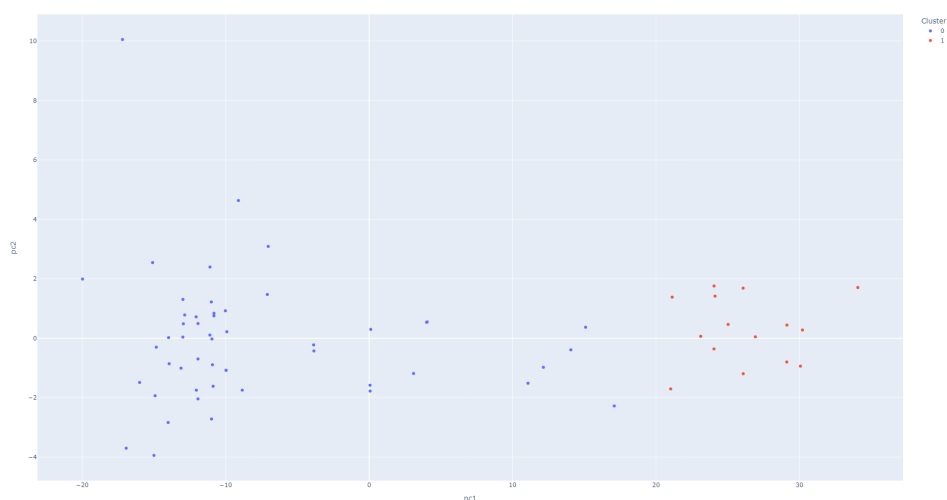


Abbildung 5.2: Streudiagramm der MainCS

würde. Außerdem ist der Punkt oben links sehr weit von der Mitte des Cluster 0 entfernt. Um die Eigenschaften der gefundenen Cluster zu vergleichen, habe ich diese als Boxplot dargestellt.

Hier sieht man einen deutlichen Altersunterschied zwischen den beiden Clustern. Der Median für Cluster 0 liegt bei 28, während der Median für Cluster 1 bei 64 liegt. Außerdem ist der Anteil an Studierenden in Cluster 0 höher als bei Cluster 1 bei 0 liegt. Die Stimmungswert liegt für Cluster 0 bei 4,777 und liegt damit im neutralen Bereich, während der Wert für Cluster 1 bei 5,555 und liegt im positiven Bereich. Die Reaktivität ist bei beiden Clustern ungefähr auf dem selben Niveau mit 3.833 bei Cluster 0 und 3.5 bei Cluster 1. Die Werte für den Erklärungsbedarf unterscheiden sich hauptsächlich an den Stellen *Interaktion*, *Metainformation/y* und *Systemverhalten*. *Interaktion* und *Systemverhalten* sind in Cluster 0 höher und *Metainformation/y* ist in Cluster 1 höher.

Die gefundenen Cluster werden mit dem Kruskal-Wallis Test^{4.5} überprüft. Für diesen wird die Nullhypothese „Alle Gruppen haben die selbe Verteilung“ aufgestellt. Um den Test durchzuführen und die Nullhypothese zu bestätigen oder zu verwerfen, wurde ein weiteres Python Skript verwendet. Allerdings konnten die Werte von der Variable „divers“ nicht untersucht werden, da zu diesem Zeitpunkt keine Person mit dieser Eigenschaft mehr in den Daten vorhanden war. Die Ergebnisse aus dem Kruskal-Wallis Test werden in den folgender Tabellen aufgeführt.

Die Nullhypothese wird verworfen für die Variablen Alter, weiblich, Arbeit, *Interaktion* und *Zukunftsplan*. Das bedeutet, die Variablen der Cluster haben nicht die selbe Verteilung. Die Variablen weiblich und

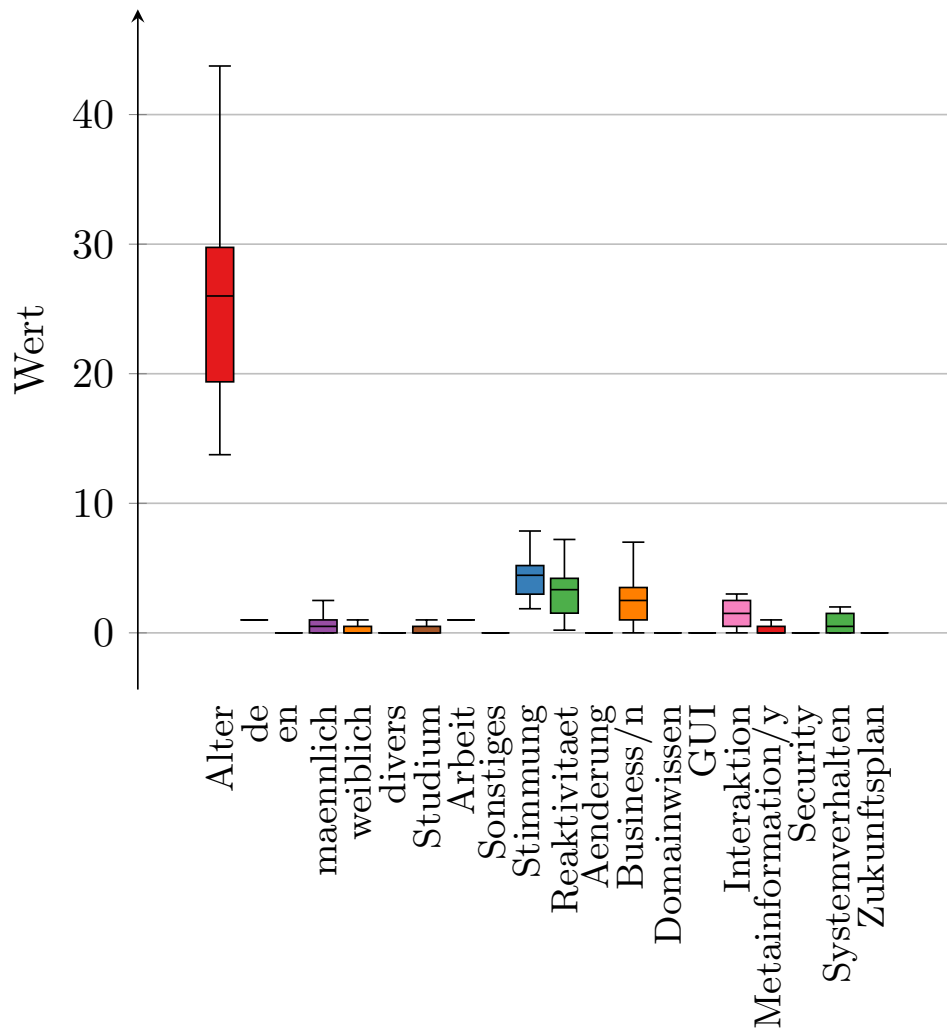


Abbildung 5.3: Boxplot des ersten Clusters aus MainCS

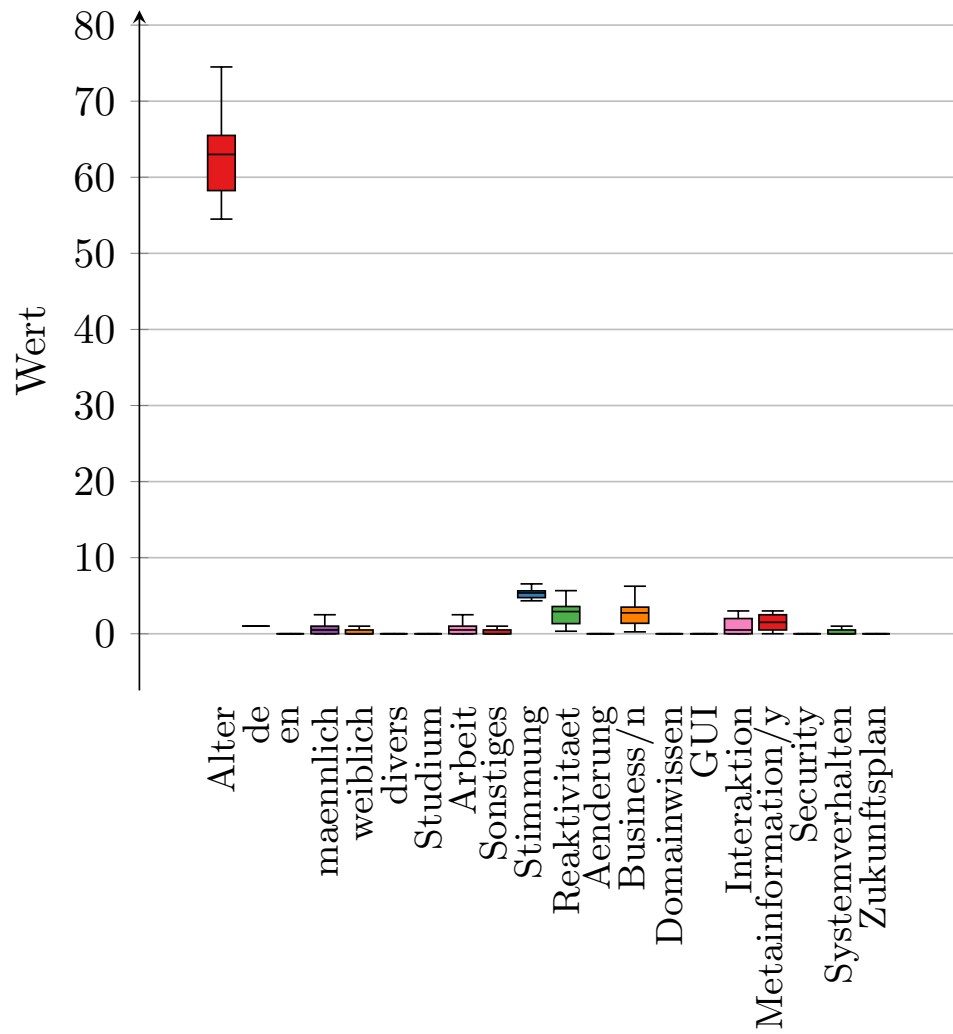


Abbildung 5.4: Boxplot des zweiten Clusters aus MainCS

MainCS	p-Wert	H0
Alter	0	verwerfen
deutsch	0,1044	nicht verwerfen
englisch	0,4283	nicht verwerfen
männlich	0,4283	nicht verwerfen
weiblich	0,0028	verwerfen
Studium	0,0843	nicht verwerfen
Arbeit	0	verwerfen
Sonstiges	0,1012	nicht verwerfen
Stimmung	0,2232	nicht verwerfen
Reaktivität	0,8289	nicht verwerfen
<i>Änderung</i>	0,7184	nicht verwerfen
<i>Business/n</i>	0,5289	nicht verwerfen
<i>Domainwissen</i>	0,1047	nicht verwerfen
<i>GUI</i>	0,3317	nicht verwerfen
<i>Interaktion</i>	0,0002	verwerfen
<i>Metainformation/y</i>	0,7588	nicht verwerfen
<i>Security</i>	0,1715	nicht verwerfen
<i>Systemverhalten</i>	0,4583	nicht verwerfen
<i>Zukunftsplan</i>	0	verwerfen

Tabelle 5.6: Ergebnisse des Kruskal-Wallis Tests für MainCS

Zukunftsplan sind zwar nicht gleichverteilt, unterscheiden sich aber zwischen den beiden Clustern nicht.

Im folgenden fasse ich die Ergebnisse der anderen Clusteranalysen kurz zusammen. Die Dendrogramme A.1.2 und Streudiagramme sowie die dazugehörigen Boxplots A.1.3 und Ergebnisse aus dem Kruskal-Wallis Test A.1.4 sind im Anhang zu finden. Die Clusteranalyse für die Gruppe Main ergab keine aussagekräftigen Cluster A.7. Das eine Cluster bestand aus einer einzelnen Person und das andere Cluster aus den restlichen 65 Personen. Dies zeigt sich auch in dem dazugehörigen Streudiagrammen A.13. Man sieht eine große weit gestreute Gruppe auf der linken Seite und den einzelnen Punkt auf der rechten Seite des Diagramms. Der Kruskal-Wallis Test ergab für alle Variablen, dass diese gleich verteilt sind. Folglich ist es nicht sinnvoll diese Cluster auf verschiedene Eigenschaften zu untersuchen. Die Clusteranalyse für die Gruppe Simple kam zu einem ähnlichen Ergebnis A.19. Auch hier wurde ein Cluster bestehend aus einer Person identifiziert und das andere Cluster enthält die restlichen 65 Personen A.19. Der Kruskal-Wallis Test kam wieder zu dem Ergebnis, dass alle Variablen gleich verteilt sind. Das bedeutet auch hier ist es nicht sinnvoll die Eigenschaften der Cluster weiter zu untersuchen. Die Clusteranalyse für die Gruppe SimpleCS konnte sinnvollere Cluster bestimmen A.9. Diese setzen sich aus 51 Personen in Cluster 0 und 15

Personen in Cluster 1 zusammen. Für diese konnten die Eigenschaften für die Cluster betrachtet werden. Cluster 0 besteht aus einer eher jüngeren Gruppe mit einem Median von 27 im Vergleich zu Cluster 1 mit einem Median von 64. Entsprechend enthält Cluster 1 auch die Studierenden, während Cluster 0 keine Studierenden enthält. Stimmung und Reaktivität unterscheiden sich in dieser Gruppe kaum voneinander. Es fällt aber auf, dass der implizite Erklärungsbedarf (i) bei Cluster 0 stärker ausgeprägt als bei Cluster 1. Die Clusteranalyse für Sub fällt wieder ähnlich wie bei Main und Simple aus. Es wird ein Cluster aus einer Person identifiziert und die restlichen 65 Personen bilden das andere Cluster. Es ist hier auch wieder nicht sinnvoll diese auf Eigenschaften zu untersuchen. Die Clusteranalyse für SubCS ist dafür eindeutiger. Es können zwei deutlich getrennte Cluster identifiziert werden. Auch bei dem dazugehörigen Streudiagramm ist eine klare Trennung ersichtlich. Auch hier ist wie in MainCS und SimpleCS, das Alter eine entscheidende Eigenschaft. Für Cluster 0 liegt der Median bei 27 und bei Cluster 1 bei 62. Dieser Unterschied erklärt auch wieder, dass Cluster 1 keine Studierenden beinhaltet im Gegensatz zu Cluster 0. Bei Cluster 0 besteht auch ein breiter aufgestellter Erklärungsbedarf bestehend aus den Variablen *Bug o. Abstürze*, *Navigation*, *Operation* und *Unerwartetes Systemverhalten*. Bei Cluster 1 besteht der Erklärungsbedarf hauptsächlich aus *Operation*. Auch der Kruskal-Wallis Test bestätigt nochmal diese Unterschiede bei den Variablen Alter, Studium, Sonstiges, und *Unerwartetes Systemverhalten*.

Kapitel 6

Diskussion

In diesem Kapitel werden die Forschungsfragen mit den Ergebnissen aus Kapitel 5 beantwortet und anschließend interpretiert.

6.1 Beantwortung der Forschungsfragen

RQ1: *Wie stark ist der Zusammenhang zwischen dem Erklärungsbedarf und dem Stimmungsbild der befragten Personen?*

Diese Frage wurde in zwei untergeordnete Forschungsfragen aufgeteilt.

RQ1.2: *Wie stark ist der Zusammenhang zwischen dem Erklärungsbedarf und der Stimmung der befragten Personen?*

Zwischen dem Erklärungsbedarf und der Stimmung der befragten Personen wurde größtenteils kein Zusammenhang festgestellt. Nur zwischen den Unterkategorien des Erklärungsbedarf und der Stimmung konnte ein geringer Zusammenhang festgestellt werden.

RQ1.2: *Wie stark ist der Zusammenhang zwischen dem Erklärungsbedarf und der Reaktivität der befragten Personen?*

Zwischen Erklärungsbedarf und der Reaktivität konnte insgesamt ein geringer Zusammenhang festgestellt werden.

Für den Erklärungsbedarf und das Stimmungsbild ergibt sich somit kein bis ein geringer Zusammenhang.

RQ2: *Wie stark ist der Zusammenhang zwischen dem Erklärungsbedarf und den demografischen Daten der befragten Personen?*

Diese Frage wurde in die Forschungsfragen RQ2.1 bis RQ2.5 aufgeteilt.

RQ2.1: *Wie stark ist der Zusammenhang zwischen dem Erklärungsbedarf und dem Alter der befragten Personen?*

Zwischen dem Erklärungsbedarf und dem Alter wurde insgesamt ein geringer Zusammenhang gefunden.

RQ2.2: *Wie stark ist der Zusammenhang zwischen dem Erklärungsbedarf und der Generation der befragten Personen?*

Zwischen dem Erklärungsbedarf und der Generation konnte insgesamt ein geringer Zusammenhang festgestellt werden.

RQ2.3: *Wie stark ist der Zusammenhang zwischen dem Erklärungsbedarf und dem Geschlecht der befragten Personen?*

Zwischen dem Erklärungsbedarf und dem Geschlecht wurde insgesamt ein geringer Zusammenhang festgestellt.

RQ2.4: *Wie stark ist der Zusammenhang zwischen dem Erklärungsbedarf und dem Beruf der befragten Personen?*

Zwischen dem Erklärungsbedarf und dem Beruf besteht ein geringer Zusammenhang.

RQ2.5: *Wie stark ist der Zusammenhang zwischen dem Erklärungsbedarf und der gesprochenen Sprache der befragten Personen?*

Zwischen dem Erklärungsbedarf und der gesprochenen Sprache besteht insgesamt eher kein Zusammenhang.

Insgesamt ergibt sich daher ein eher geringer Zusammenhang zwischen dem Erklärungsbedarf und den demografischen Daten.

RQ3.1: *Wie viele Cluster ergeben sich aus den untersuchten Variablen?*

Es konnten bei den sechs durchgeführten Clusteranalysen je zwei Cluster identifiziert werden. Die Clusteranalysen unter Einbeziehung der demografischen Daten konnten nach Interpretation des Silhouettenkoeffizienten eine stärkere Struktur in den Clustern finden. Die gefundenen Cluster dieser Clusteranalysen, welche zusätzlich demografische Daten betrachteten, waren am aussagekräftigsten.

RQ3.2: *Welche Eigenschaften unterscheiden die identifizierten Cluster voneinander?*

Die Clusteranalyse ohne demografische Daten konnte keine aussagekräftige Cluster bilden, da je eines Cluster aus einer einzelnen Person bestand. Die Eigenschaften aus den Clustern ohne demografische Daten wurden daher nicht weiter betrachtet. Die Eigenschaften der Cluster, die demografische Daten betrachteten, unterschieden sich hauptsächlich in Alter und Beruf. Das erste Cluster bestand dabei aus einer Altersgruppe mit Medianen von 27 bis 28 Jahren und beinhaltete die Studierenden. Das zweite Cluster enthielt die Altersgruppe mit Medianen von 62 bis 64 Jahren und dem Beruf Sonstiges. Der implizite Erklärungsbedarf des ersten Clusters war bei einer der Analysen höher. Das erste Cluster zeigte zudem einen höheren Bedarf an Erklärung zu *Interaktion* und dessen Unterkategorien *Navigation* und *Operation*, sowie zu *Systemverhalten* und dessen Unterkategorien *Bug o. Abstürze* und *Unerwartetes Systemverhalten*. Das zweite Cluster zeigte einen höheren Erklärungsbedarf an *Metainformation/y* und der zusätzlichen Unterkategorie *Operation*.

6.2 Interpretation

Die Ergebnisse der Korrelationsanalyse haben gezeigt, dass maximal ein geringer Zusammenhang zwischen Erklärungsbedarf und Stimmung, sowie ein geringer Zusammenhang zwischen Erklärungsbedarf und demografischen Daten besteht. Dieser geringe Zusammenhang weist nicht notwendigerweise auf eine mangelnde Bedeutung hin, sondern sollte als Anreiz zu weiterer qualitativer Forschung oder vertiefenden Analysen verstanden werden. Das Ziel sollte es sein, die Gründe für diese Beziehungen besser zu verstehen.

Zudem konnten durch die Clusteranalyse Cluster identifiziert werden, die Unterschiede beim Erklärungsbedarf und dem Alter der befragten Personen gezeigt hat. Diese Ergebnisse können genutzt werden, sodass bei der Softwareentwicklung die gegebenen Erklärungen an die Nutzerzielgruppe angepasst werden. Für die Zielgruppe um 28-Jährige sollte der Fokus auf Erklärungen zu Interaktionen und Systemverhalten liegen. Bei den Interaktionen sollte dabei der Fokus auf Erklärungen zu der Navigation und Operation des Systems liegen. Die Erklärungen zu dem Systemverhalten sollten so angepasst werden, dass der Erklärungsbedarf zu Bugs oder Abstürzen und unerwartetem Systemverhalten verringert wird. Ich denke, dass mehrere Faktoren für dieses Ergebnis verantwortlich sind. Die jüngere Generation wird meistens mit der digitalen Welt aufgewachsen sein und verbringen einen erheblichen Teil ihrer Zeit in der digitalen Welt. Das kann ein Grund sein, die Komplexität des Systems und das Verhalten dessen besser verstehen zu wollen. Ich könnte mir zusätzlich vorstellen, dass die Lernbereitschaft und Neugier bei der jüngeren Zielgruppe ausgeprägter ist und daher die Interaktionsmöglichkeiten hinterfragt werden. Außerdem denke ich, dass die jüngere Generation die Konsequenz einer Aktion in der digitalen Welt erfahren möchte, welches den Erklärungsbedarf zu unerwartetem Systemverhalten erklären könnte.

Die ältere Gruppe um die 64-Jährigen benötigt eher Erklärungen zu Metainformationen oder unidentifizierbaren Faktoren, sowie zur Operation. Der Erklärungsbedarf an Operation hängt wahrscheinlich mit der fehlenden Erfahrung im Umgang mit digitalen Anwendungen zusammen. Dazu kommt wahrscheinlich die Angst, die Software falsch zu bedienen. Ein weiterer Grund könnte die fehlende Intuition sein, welche die jüngere Gruppe besitzt, die in einer digitalisierten Welt aufgewachsen ist. Der Punkt der Metainformationen lässt vermuten, dass diese Personen ihren Erklärungsbedarf nicht ausreichend beschreiben konnten oder nicht auf die Software bezogen haben.

Allerdings sollte beachtet werden, dass die gefundenen Ergebnisse nur für die befragten Personen des Projekts softXplain¹ gelten und daher nicht auf die Allgemeinheit übertragbar sind.

¹softXplain: Anforderungen für selbst-erklärende Software <https://gepris.dfg.de/gepris/projekt/470146331>

Kapitel 7

Validität

Im Folgenden werden die *Threats to Validity* nach Wohlin[40] betrachtet.

Der genutzte Datensatz aus dem Projekt softXplain¹ bildet nicht den Schnitt der Gesellschaft hinsichtlich der demografischen Daten ab. Der Median der Personen, die an der Umfrage teilgenommen haben, lag bei Alter 28 Jahren und bilden daher nicht die gesamte Gesellschaft ab. Es besteht hier ein *Threat to external Validity*, da diese Ergebnisse nicht auf die Allgemeinheit übertragbar sind.

Die aus dem Projekt softXplain² übernommene Taxonomie für die Einordnung des Erklärungsbedarfes ist bisher nicht in der Praxis validiert worden. Es werden daher die *Threats* der Taxonomieerstellung übernommen. Hierbei handelt es sich um einen *Threat of Construct Validity*.

Es besteht, besonders bei der manuell durchgeführten Korrelationsanalyse, der *Threat to internal Validity*. Obwohl mehrmals die Ergebnisse überprüft wurden, ist es möglich, dass Fehler übersehen wurden und somit das Ergebnis beeinflusst wurde. Außerdem hat keine zweite Instanz diese Ergebnisse überprüft. Dies würde die Wahrscheinlichkeit von Fehlern minimieren. Da dies nicht der Fall ist, besteht hier der *Threat of Construct Validity*

Die Skala, mit der die Stimmung von Zahlenwerten zu positiv, neutral und negativ konvertiert wurde, orientiert sich an dem Paper von Lin et al.[23]. Da diese Umwandlung ad hoc entwickelt wurde, besteht der *Threat to Construct Validity*.

Die aus dem Datensatz übernommenen Kategorien für den Erklärungsbedarf *Metainformationen/y* und *Business/n* repräsentieren je zwei möglicherweise unterschiedliche Konzepte und können daher zu einer Verzerrung der Ergebnisse, sowie der Interpretation führen. Es liegt der *Threat to Construct Validity* vor.

¹softXplain: Anforderungen für selbst-erklärende Software <https://gepris.dfg.de/gepris/projekt/470146331>

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

Mit der Durchführung der Bonferroni-Korrektur senkt man das Risiko einer Typ-I-Fehlers (falsch positiv), wodurch es zu einer Erhöhung des Typ-II-Fehlers (falsch negativ) kommen kann. Daher liegt hier eine *Threat to Conclusion Validity* vor.

Kapitel 8

Zusammenfassung und Ausblick

Dieses Kapitel dient dazu, einen abschließenden Überblick über die Inhalte und die wichtigsten Ergebnisse der Arbeit zu geben.

8.1 Zusammenfassung

Das Ziel dieser Arbeit war es einen Zusammenhang zwischen Erklärungsbedarf, Stimmung und demografischen Daten zu finden. Um dieses Ziel zu erreichen wurde eine Korrelationsanalyse mit Cramer's V und eine agglomerative Clusteranalyse durchgeführt.

Aus den Ergebnissen der Korrelationsanalyse konnte maximal ein geringer Zusammenhang zwischen dem Erklärungsbedarf und dem Stimmungsbild einer Person abgeleitet werden. Die Untersuchung des Zusammenhangs zwischen dem Erklärungsbedarf und den demografischen Daten ergab maximal einen geringen Zusammenhang. Durch die Clusteranalyse ließen sich, unter der Berücksichtigung der demografischen Daten, Cluster bilden. Von diesen konnten unterschiedliche Eigenschaften identifiziert werden. Es konnte gezeigt werden, dass es zwischen den Clustern einen deutlichen Altersunterschied gibt. Die jüngeren Befragten hatten einen stärkeren Erklärungsbedarf, wenn es um die Interaktion mit der Software und dem Verhalten des Systems ging. Die älteren Befragten waren eher an Metainformationen und der Operation der Software interessiert.

8.2 Ausblick

In zukünftigen Arbeiten wäre es sinnvoll ein Datensatz zu erstellen, welcher einen größeren Teil der Gesellschaft abbilden kann. Dadurch sollten die Korrelationen präziser werden und es könnten die Ergebnisse der Clusteranalyse nochmals bestätigt oder verfeinert werden.

Anstatt mit Cramer's V die Korrelation zu untersuchen, könnten die nominalen Werte wie für die Clusteranalyse angepasst werden. Dies ermöglicht die Verwendung anderer Korrelationsanalysemethoden, wie der Person Korrelationskoeffizient oder Rangkorrelationskoeffizient von Spearman, welche beide in Kapitel 2.3 beschrieben wurden. Dadurch ist es möglich zu bestimmen, wie die betrachteten Variablen zusammenhängen.

Aus den gefundenen Clustern können in zukünftigen Arbeiten zudem Personas erstellt werden. Dies hat den Vorteil, dass bei der Softwareentwicklung die Eigenschaften und Bedürfnisse der Nutzenden verdeutlicht werden. Die Entwicklung der Software kann zielgerichteter stattfinden und Annahmen zu Personen können vermieden werden, da es eine belegte Datengrundlage gibt.

Anhang A

Ein Anhang

A.1 Clusteranalyse

A.1.1 Silhouettenkoeffizient

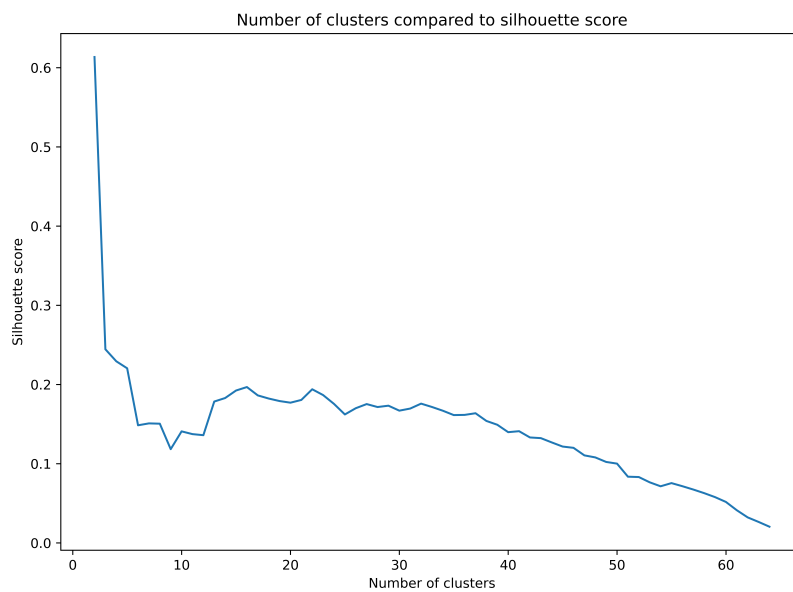


Abbildung A.1: Silhouettenkoeffizient für Main

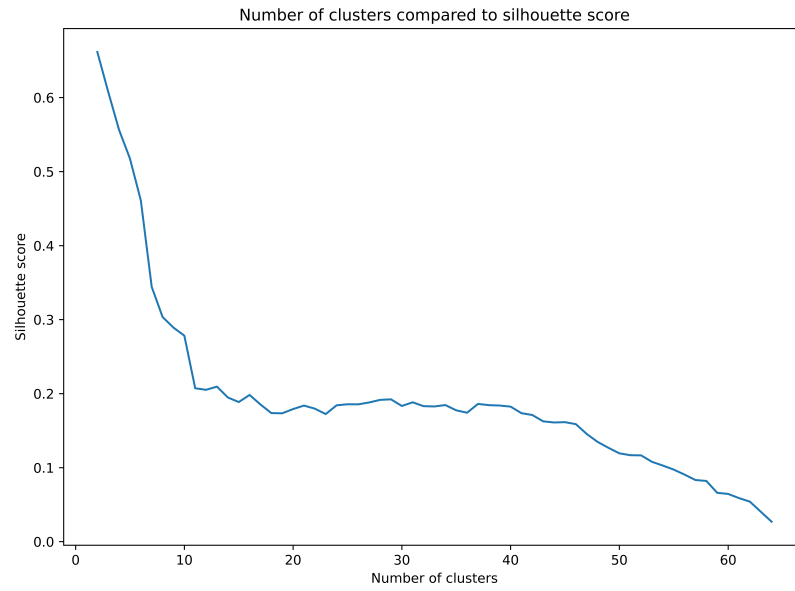


Abbildung A.2: Silhouettenkoeffizient für MainCS

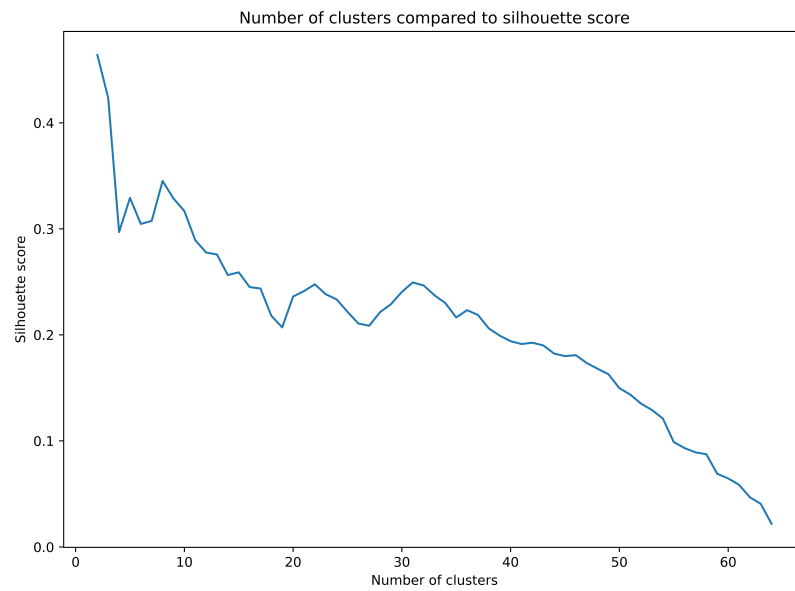


Abbildung A.3: Silhouettenkoeffizient für Simple

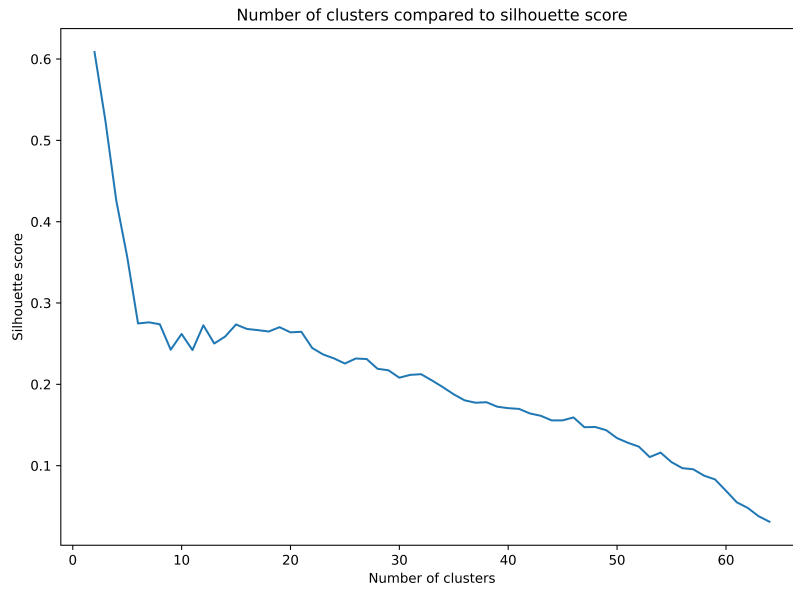


Abbildung A.4: Silhouettenkoeffizient für SimpleCS

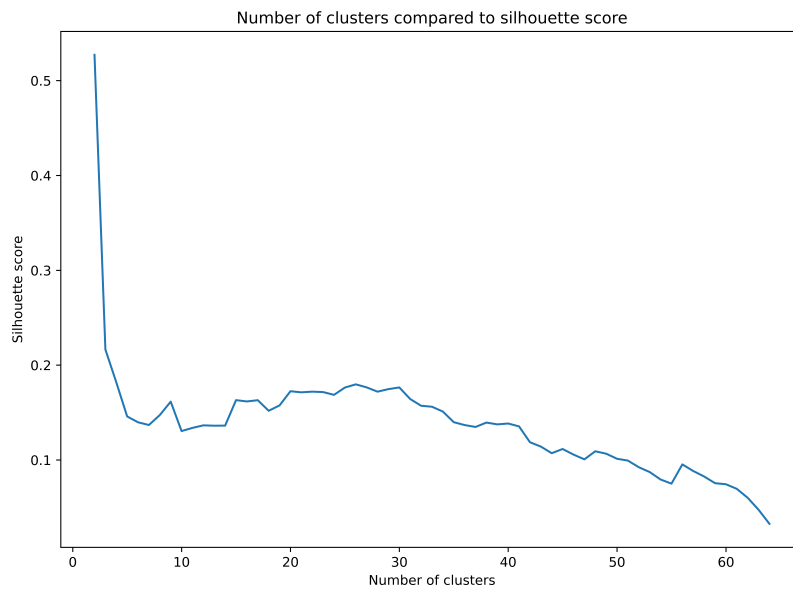


Abbildung A.5: Silhouettenkoeffizient für Sub

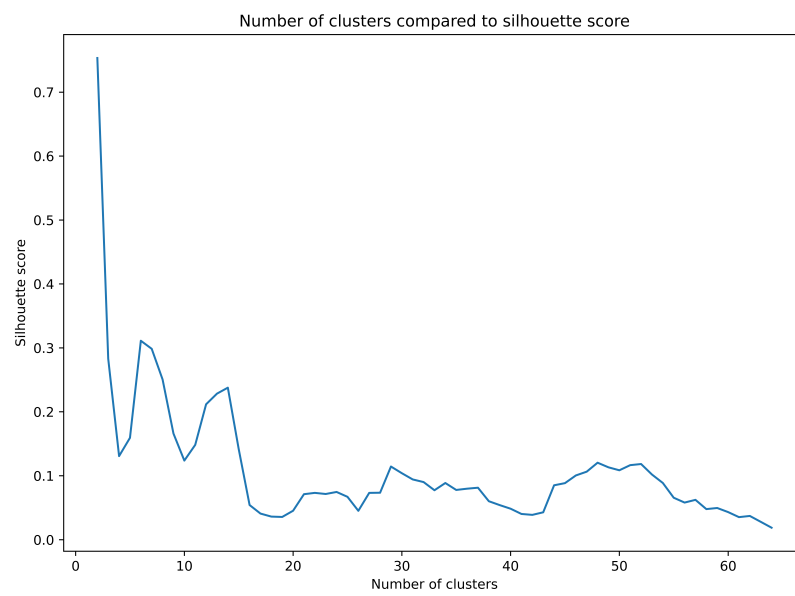


Abbildung A.6: Silhouettenkoeffizient für SubCS

A.1.2 Dendrogramme

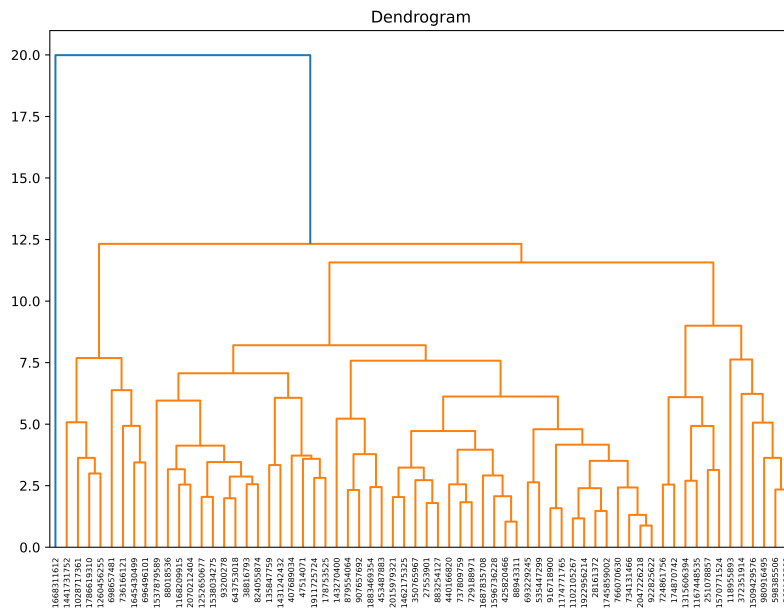


Abbildung A.7: Dendrogramm für Main

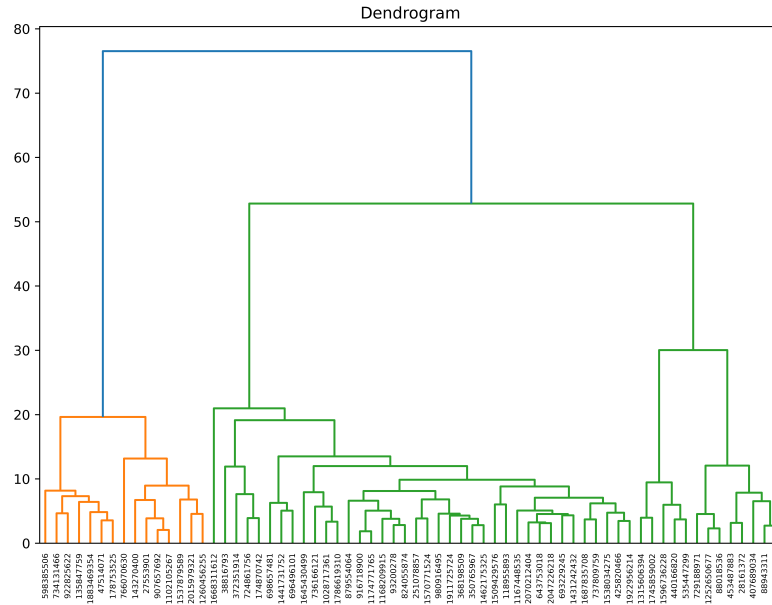


Abbildung A.8: Dendrogramm für MainCS

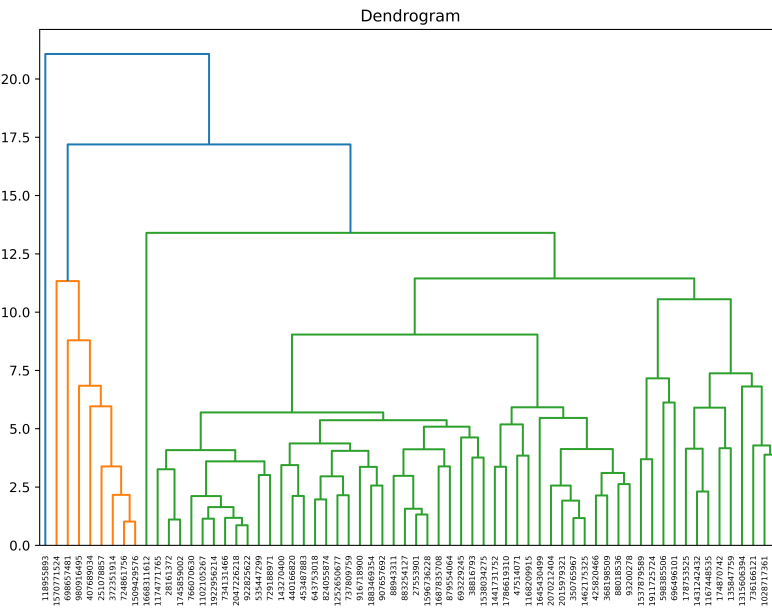


Abbildung A.9: Dendrogramm für Simple

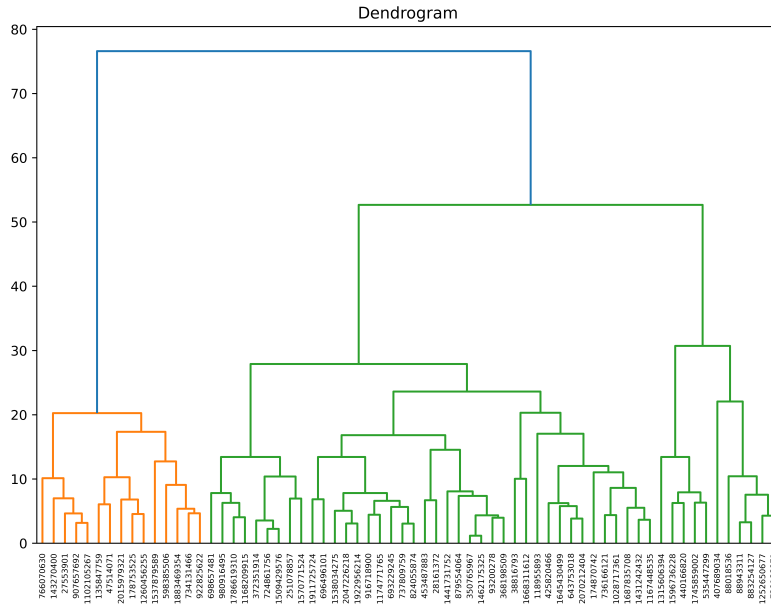


Abbildung A.10: Dendrogramm für SimpleCS

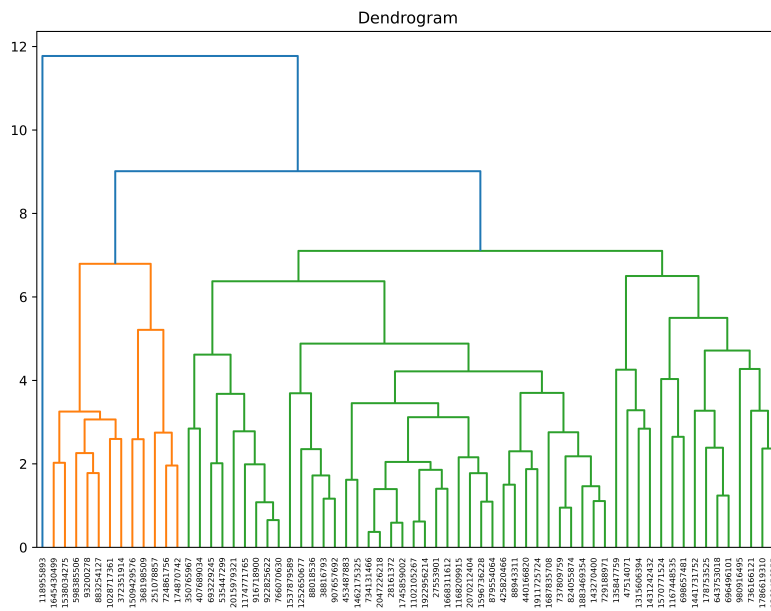


Abbildung A.11: Dendrogramm für Sub

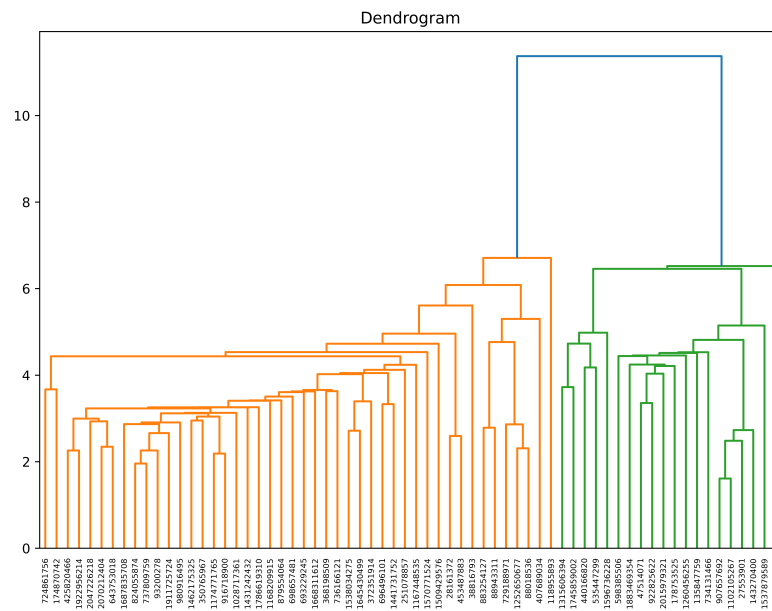


Abbildung A.12: Dendrogramm für SubCS

A.1.3 Streudiagramme mit Boxplots der Cluster

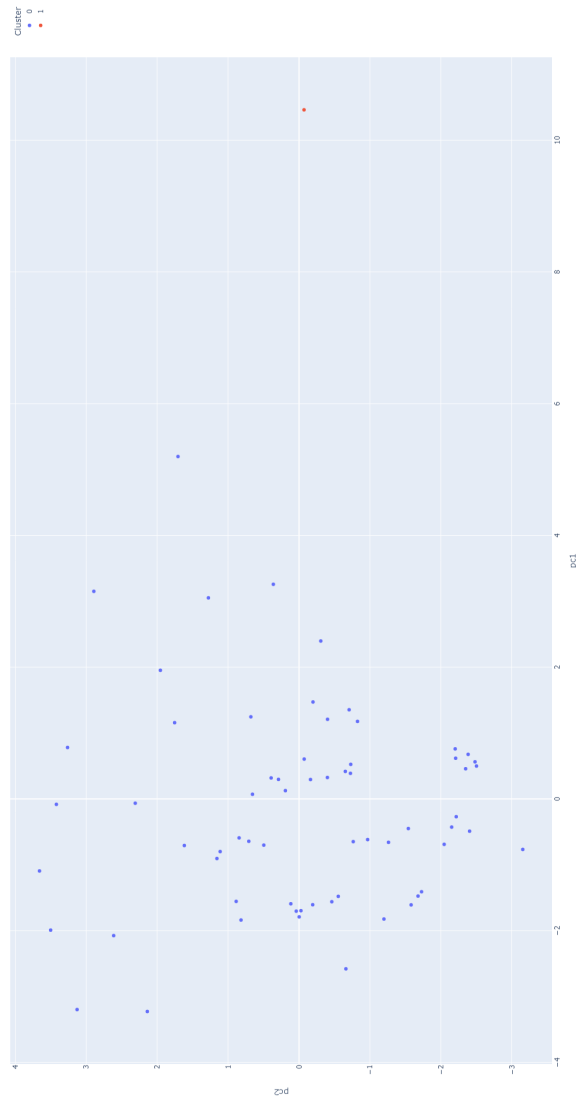


Abbildung A.13: Streudiagramm der Main

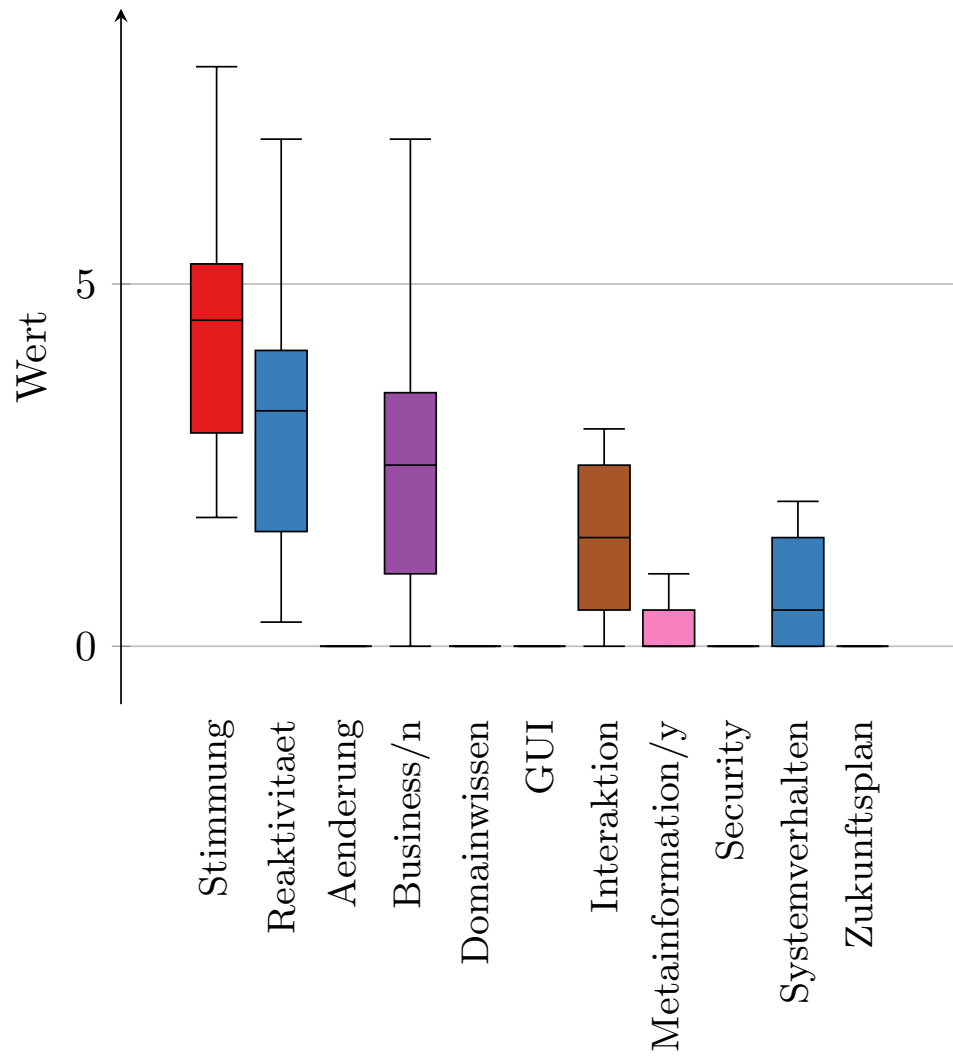


Abbildung A.14: Boxplot des ersten Clusters aus Main

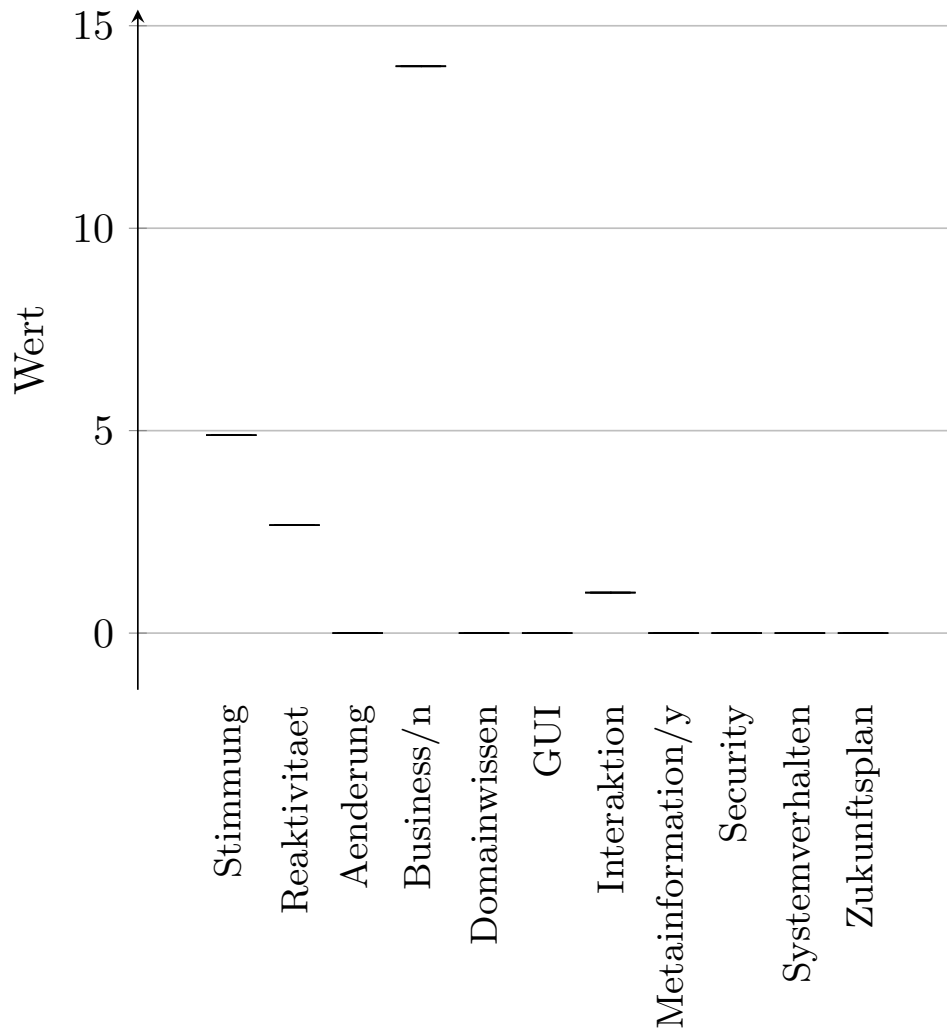


Abbildung A.15: Boxplot des zweiten Clusters aus Main

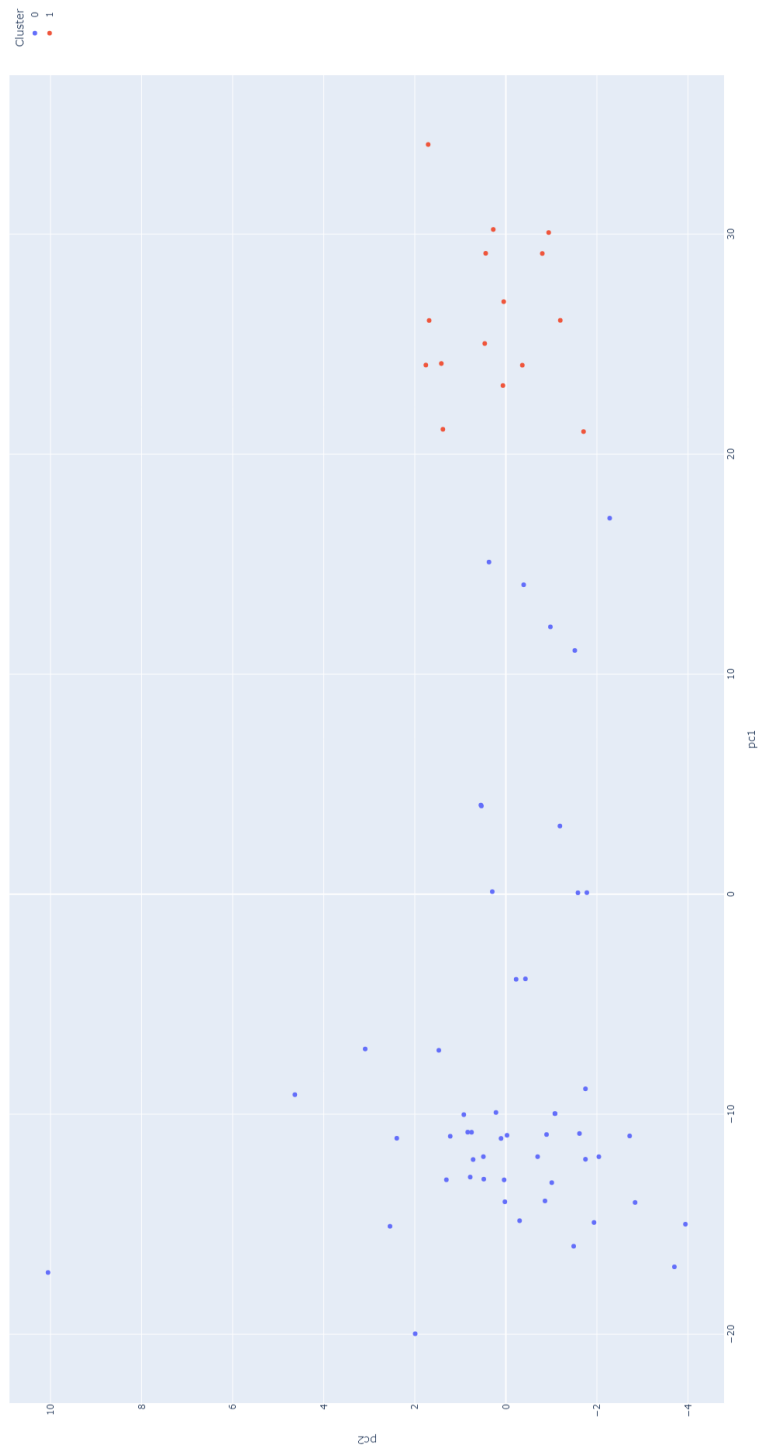


Abbildung A.16: Streudiagramm der MainCS

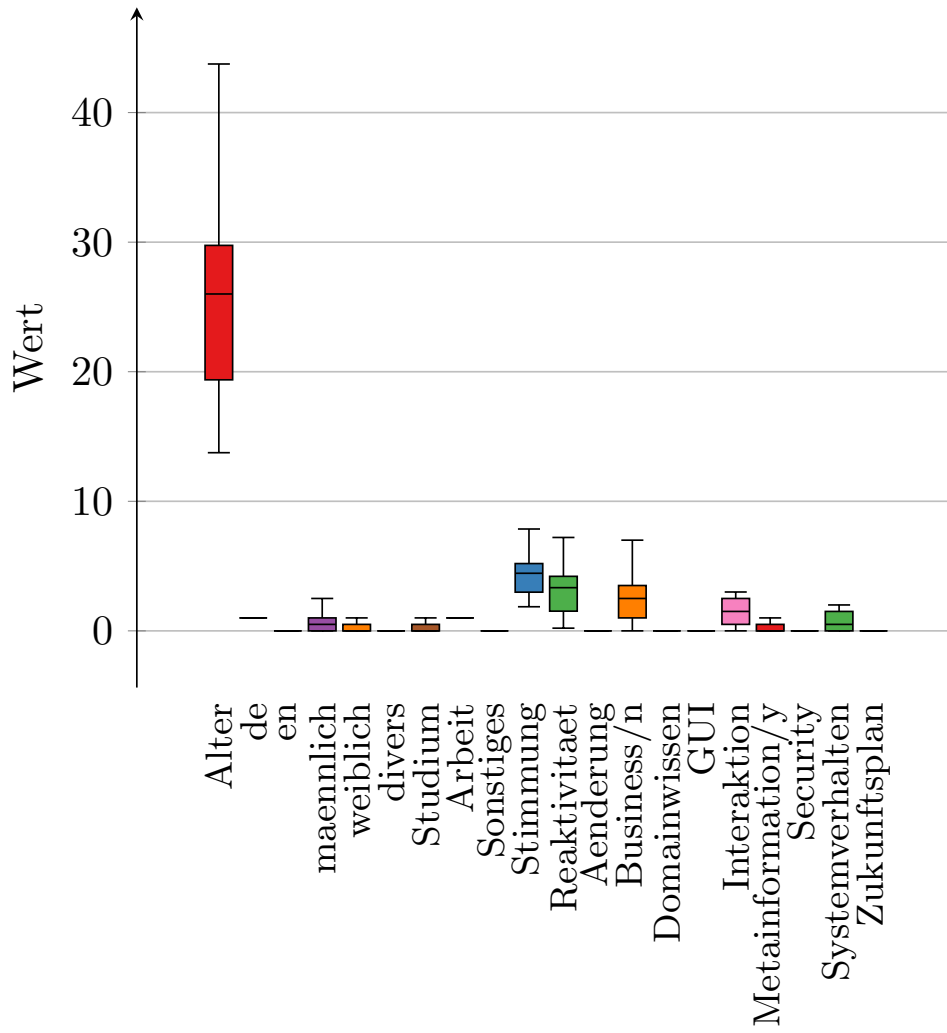


Abbildung A.17: Boxplot des ersten Clusters aus MainCS

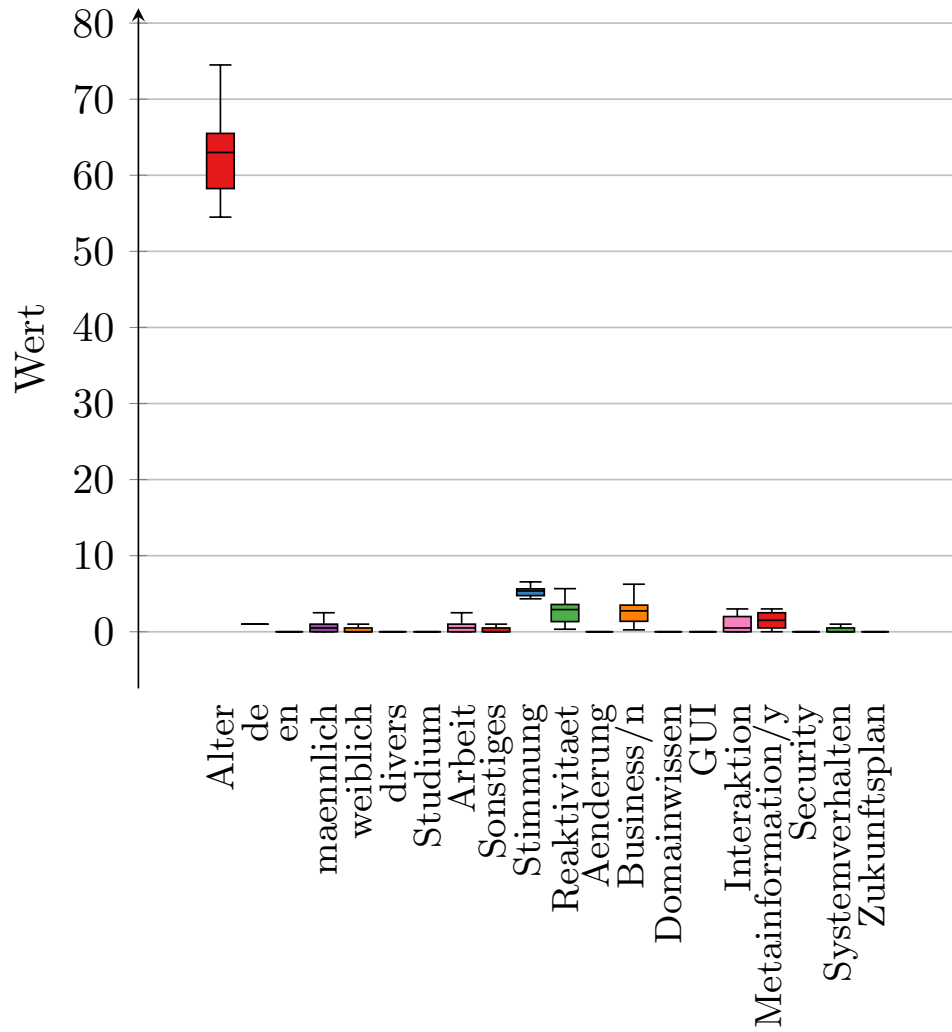


Abbildung A.18: Boxplot des zweiten Clusters aus MainCS

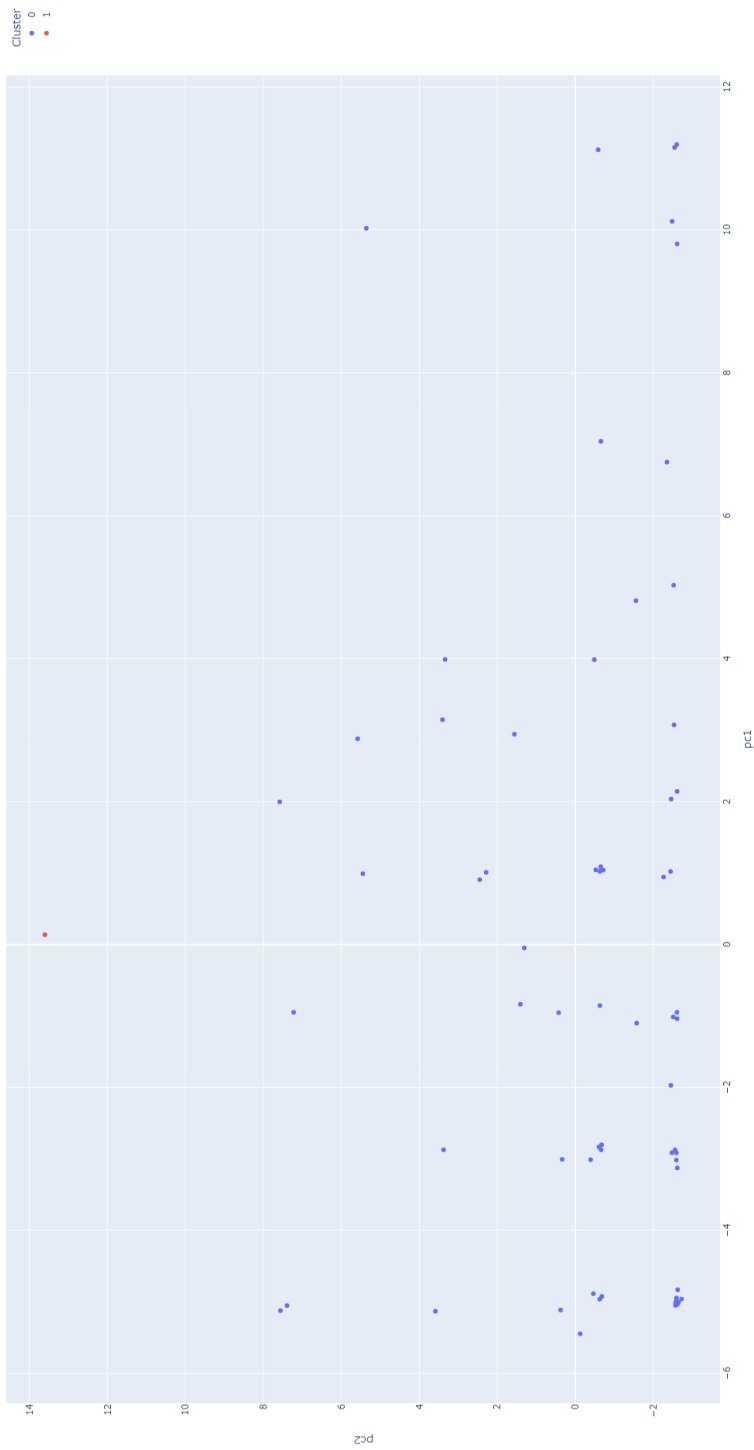


Abbildung A.19: Streudiagramm der Simple

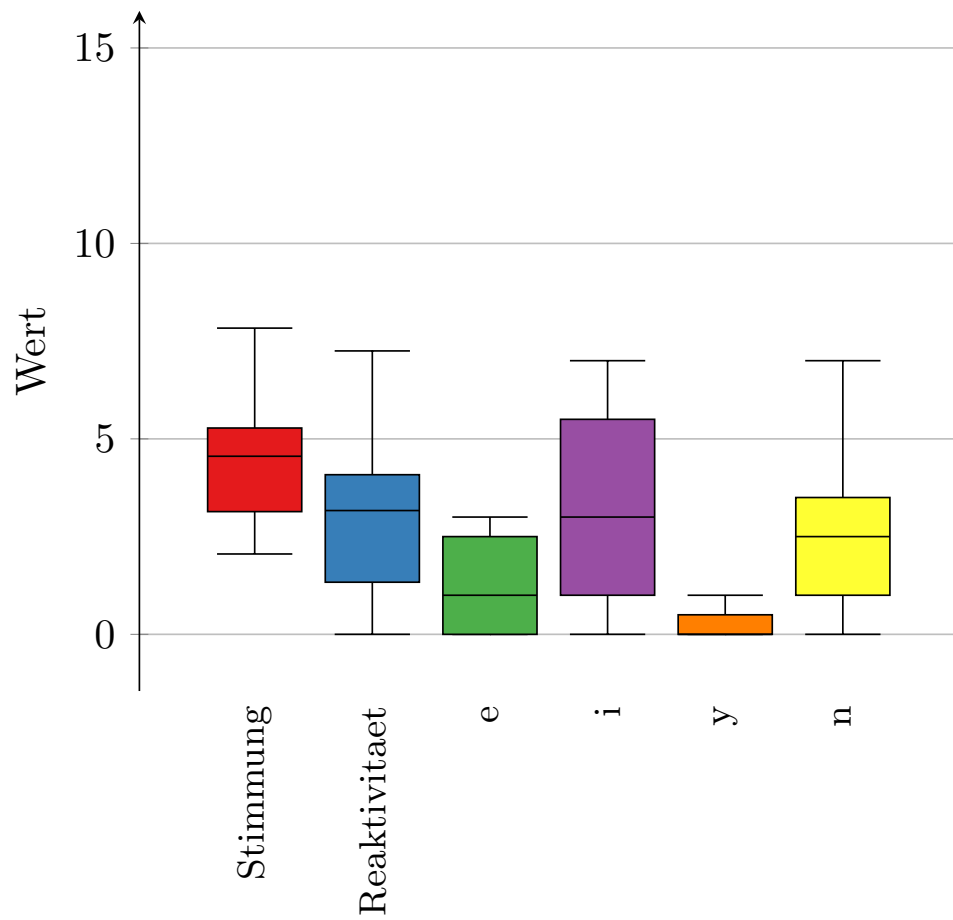


Abbildung A.20: Boxplot des ersten Clusters aus Simple

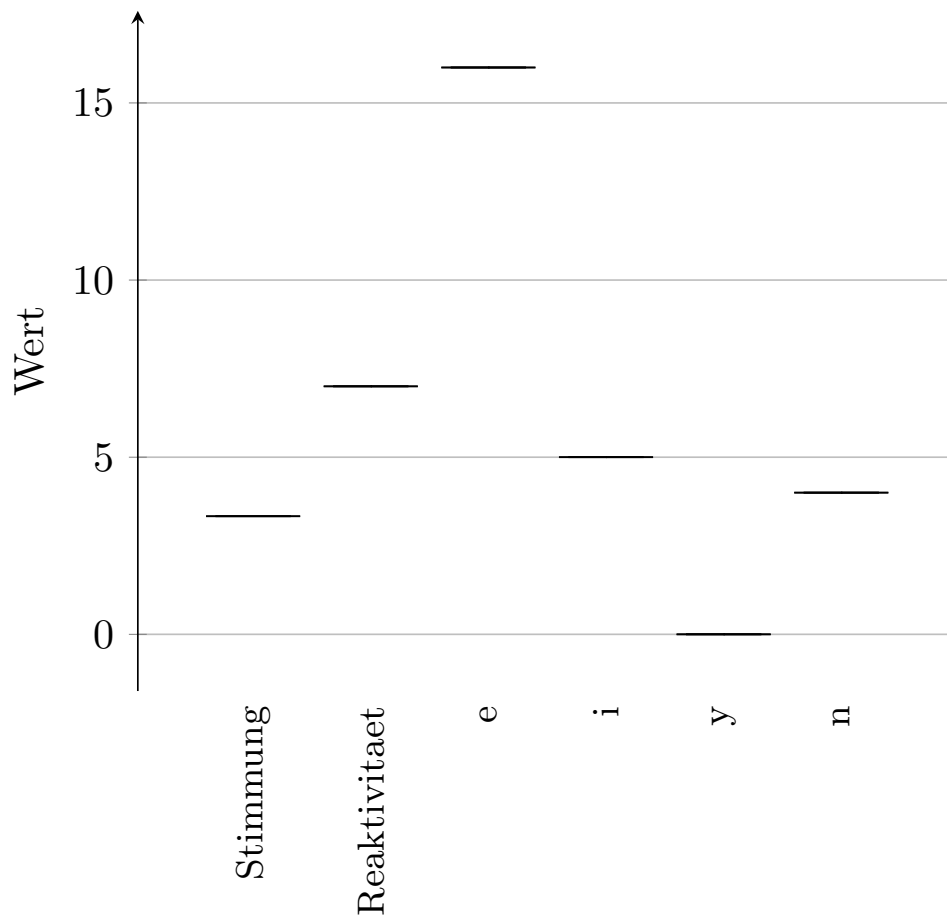


Abbildung A.21: Boxplot des zweiten Clusters aus Simple

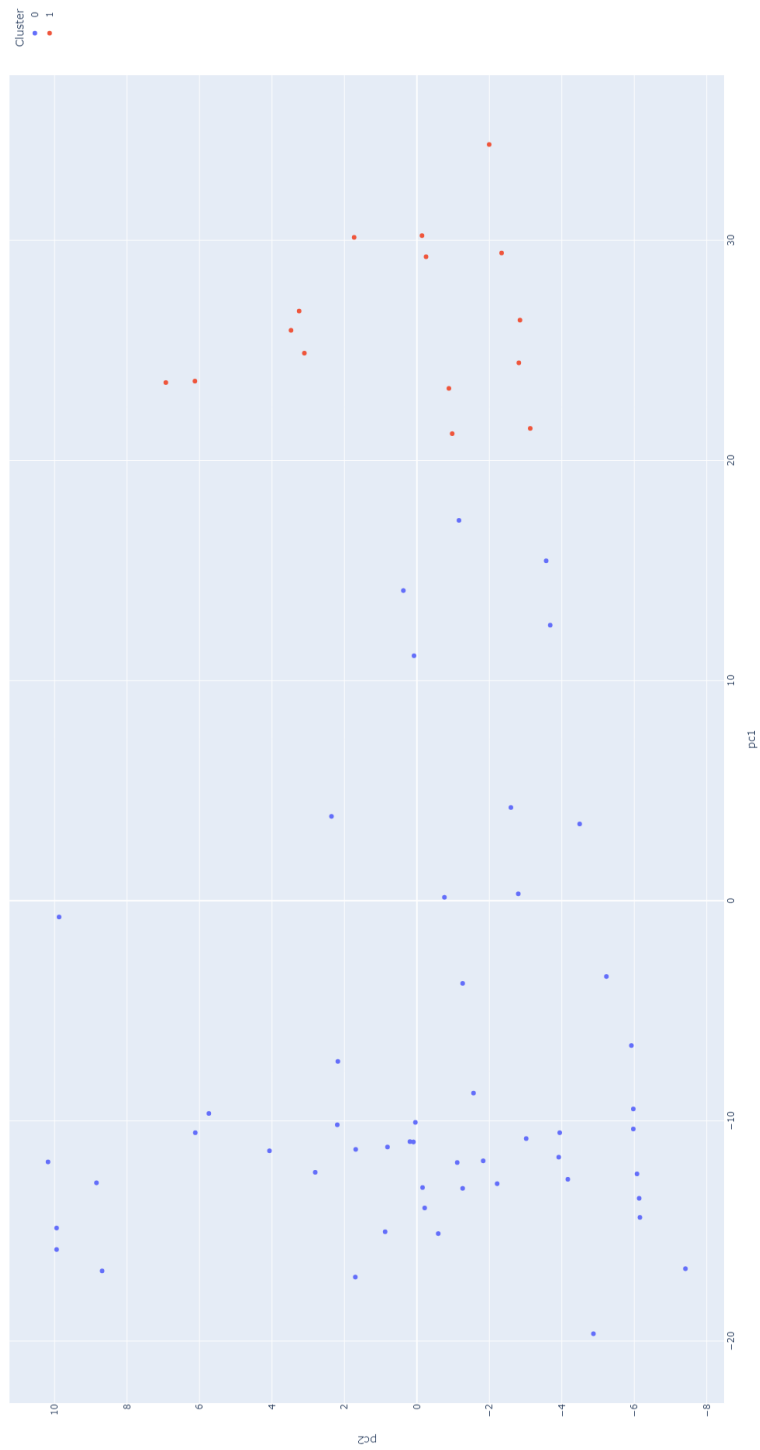


Abbildung A.22: Streudiagramm der SimpleCS

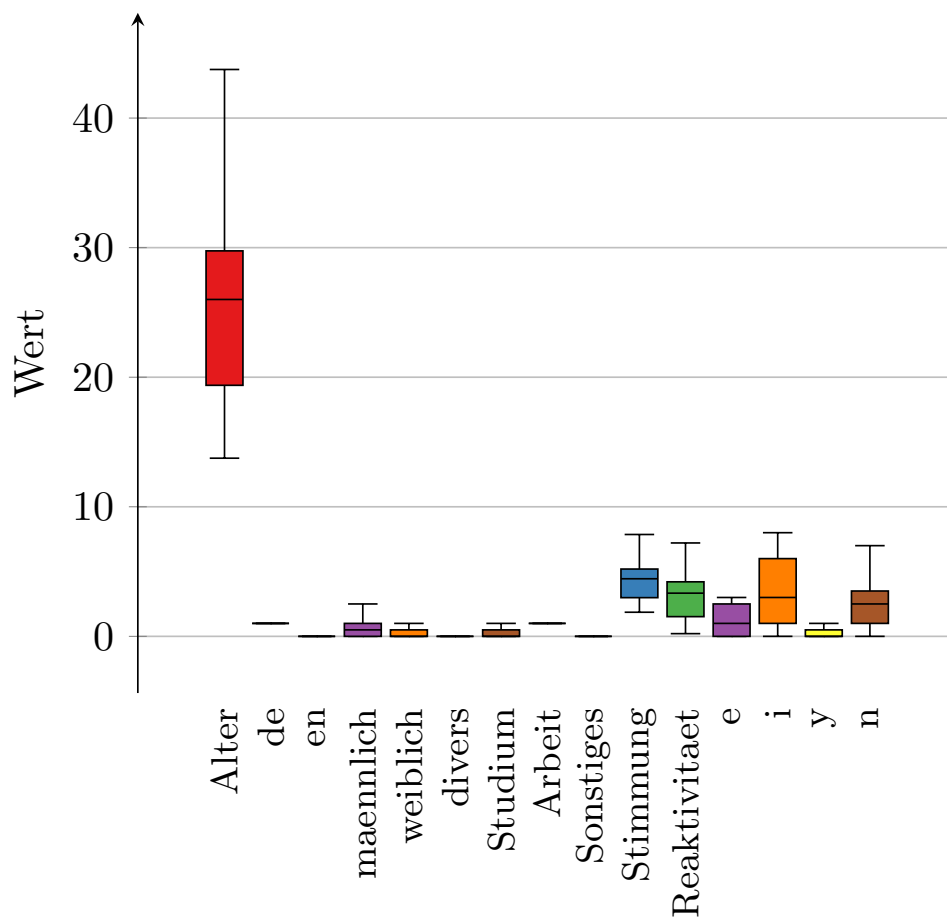


Abbildung A.23: Boxplot des ersten Clusters aus SimpleCS

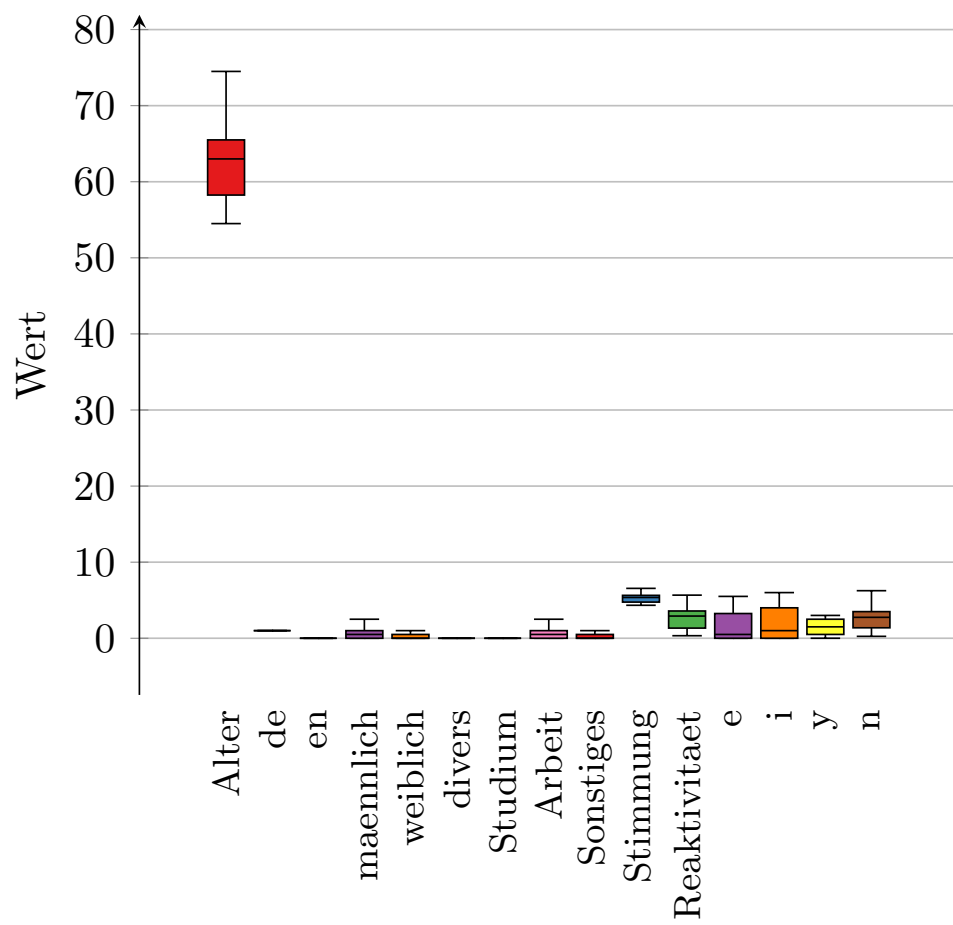


Abbildung A.24: Boxplot des zweiten Clusters aus SimpleCS

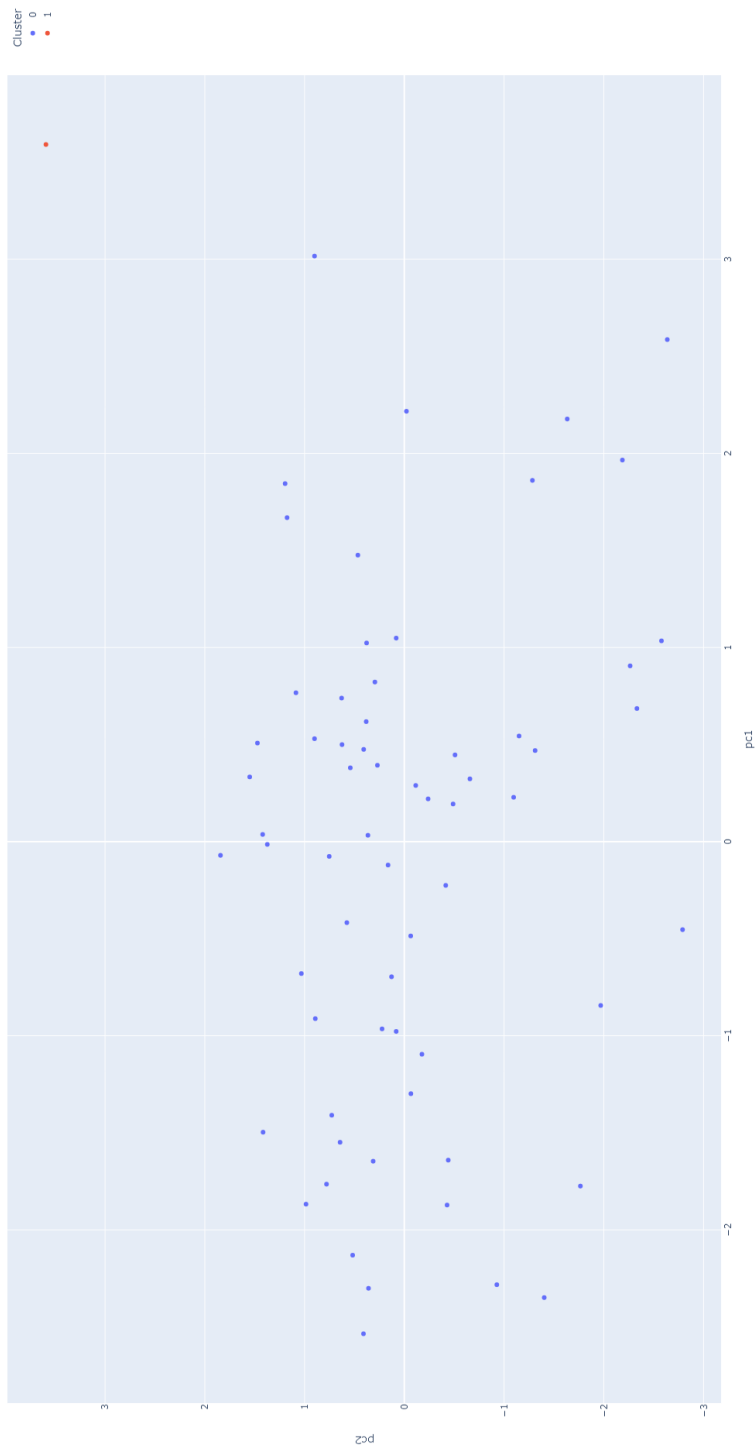


Abbildung A.25: Streudiagramm der Sub

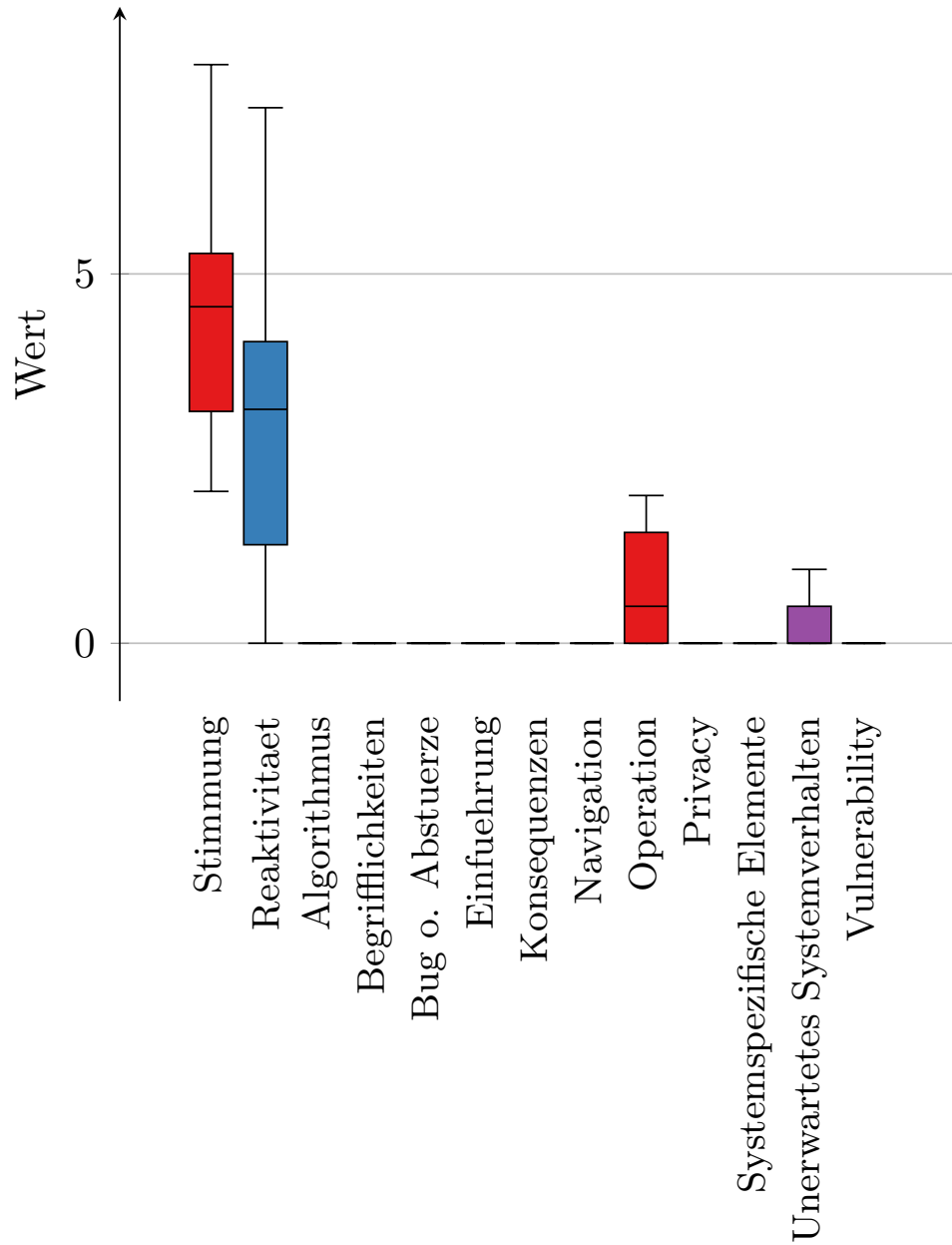


Abbildung A.26: Boxplot des ersten Clusters aus Sub

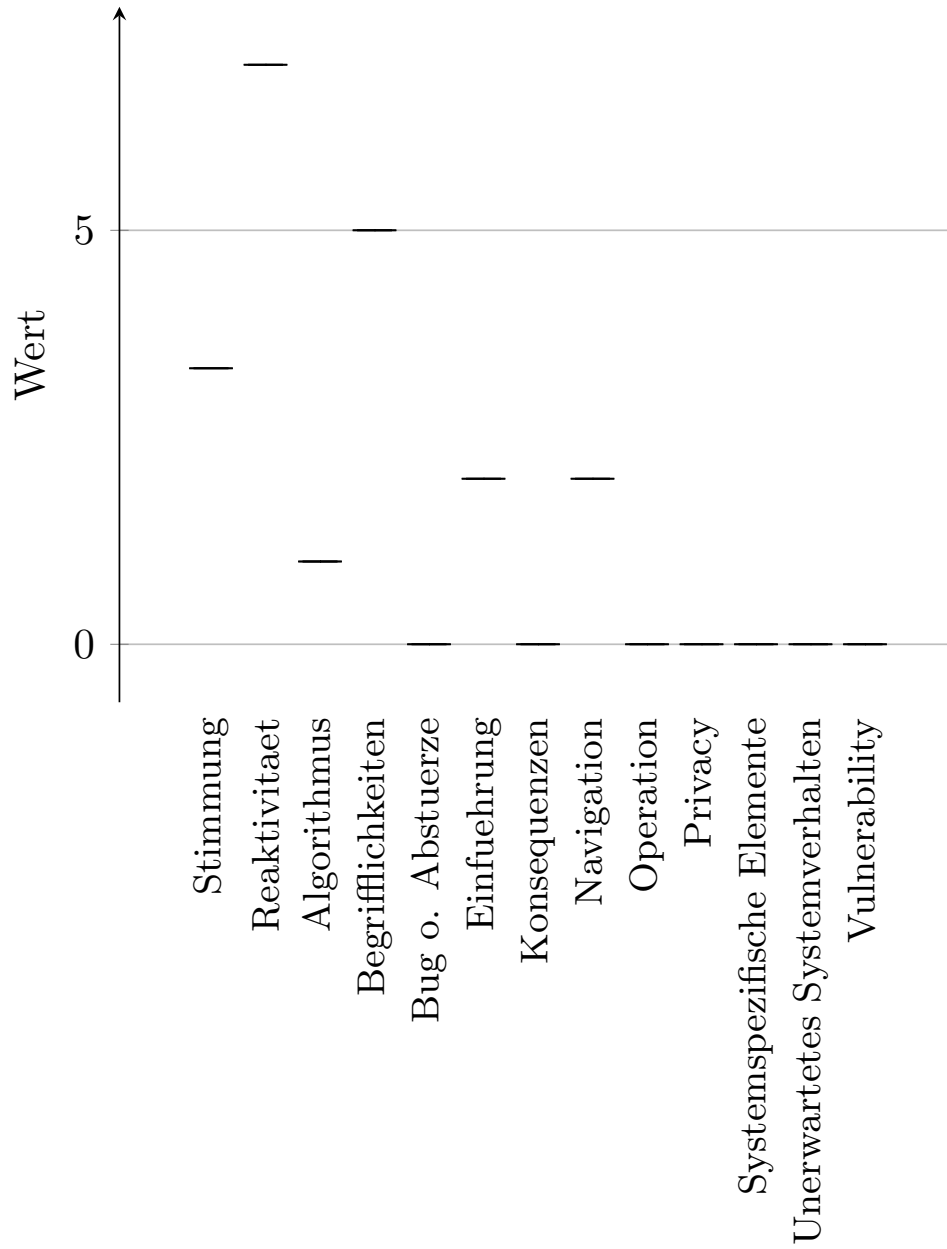


Abbildung A.27: Boxplot des zweiten Clusters aus Sub

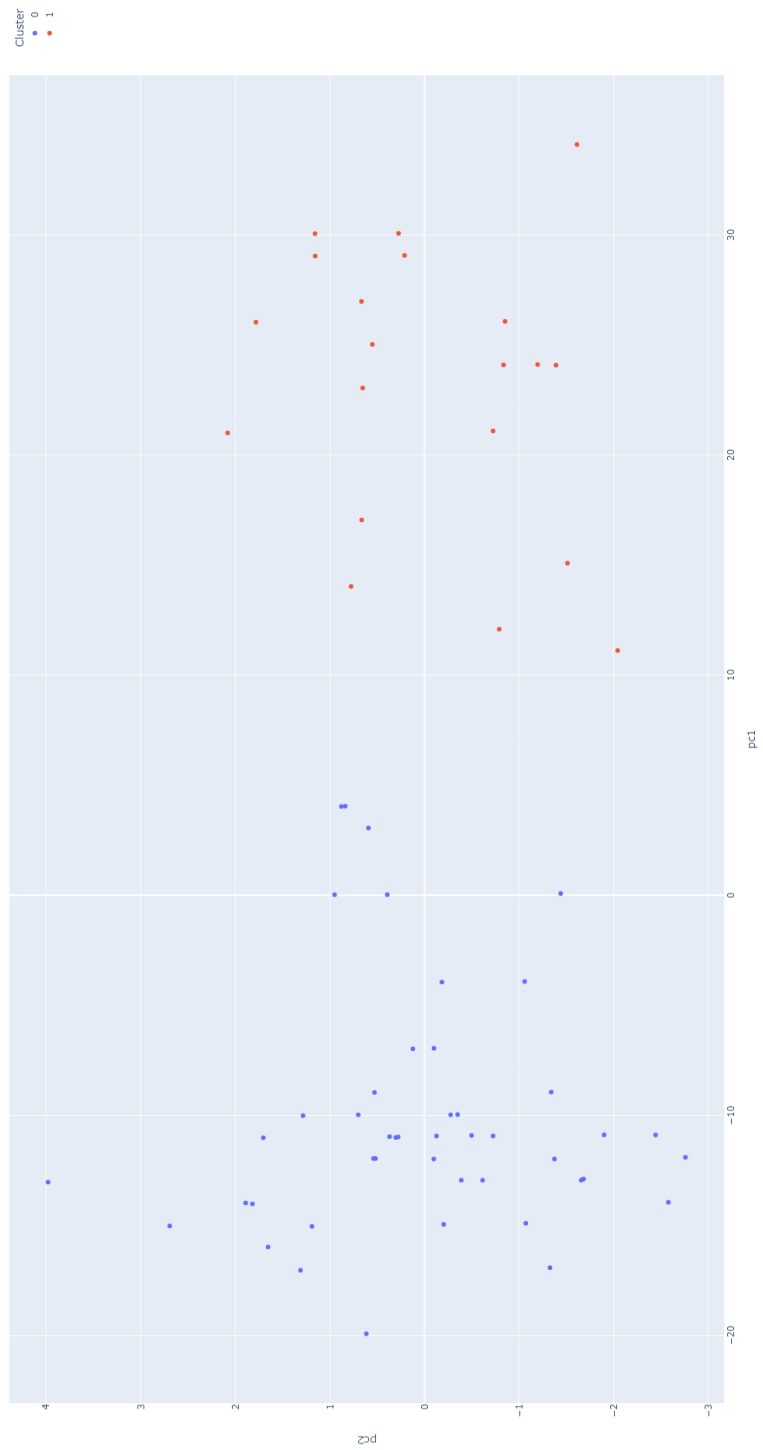


Abbildung A.28: Streudiagramm der SubCS

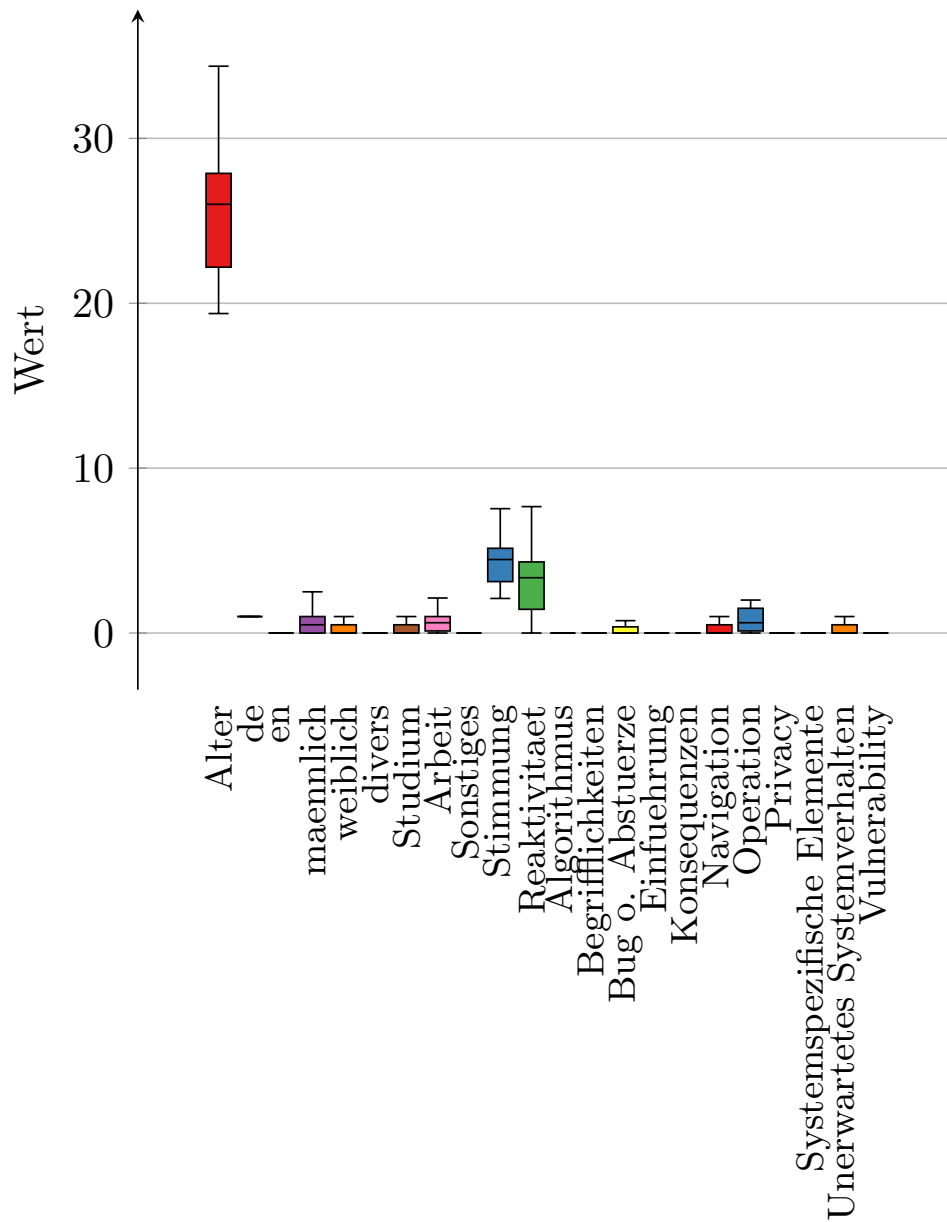


Abbildung A.29: Boxplot des ersten Clusters aus SubCS

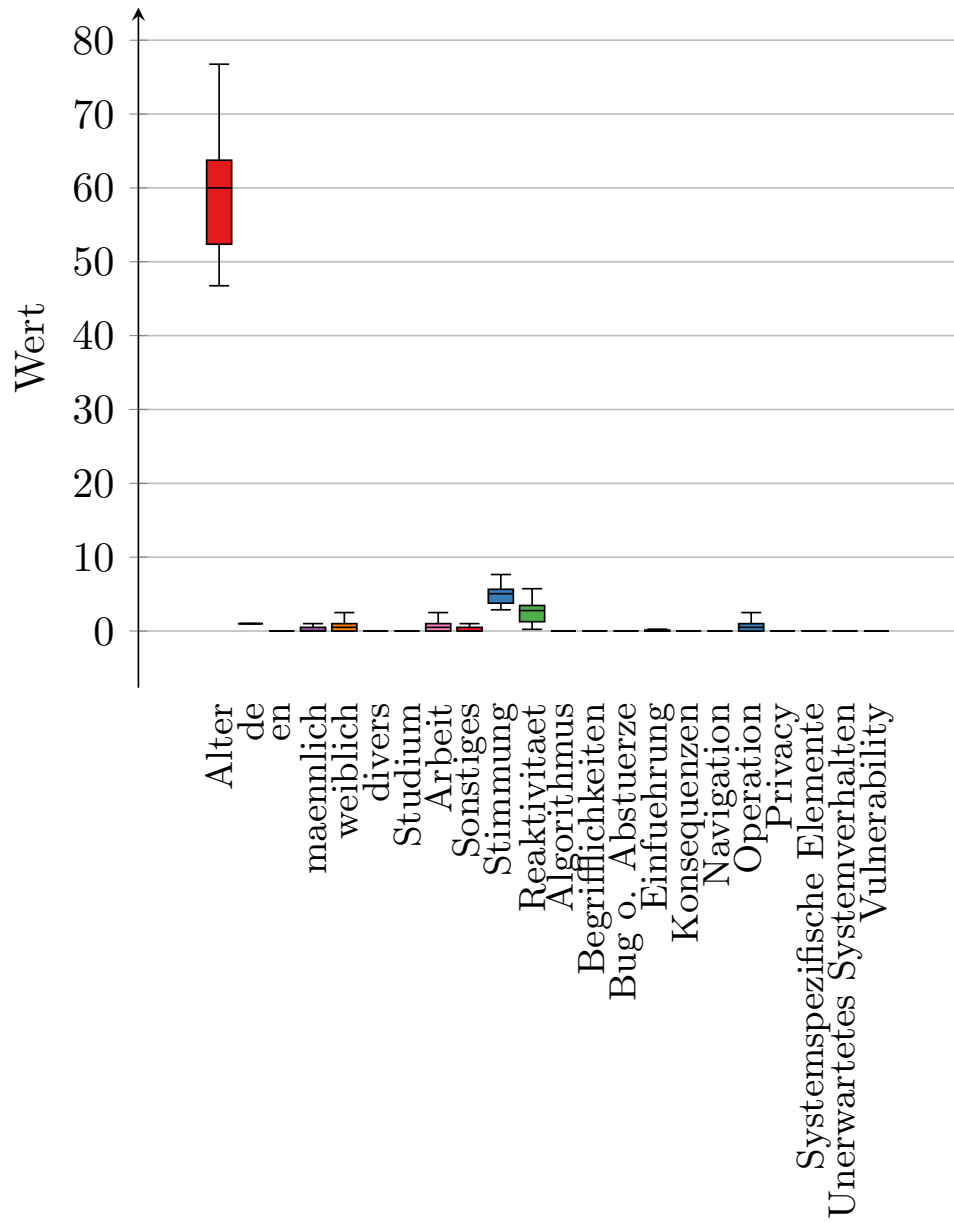


Abbildung A.30: Boxplot des zweiten Clusters aus SubCS

A.1.4 Kruskal-Wallis Test

Main	p-Wert	H0
Stimmung	0,9581	nicht verwerfen
Reaktivität	0,3716	nicht verwerfen
Änderung	0,7108	nicht verwerfen
Business/n	0,0805	nicht verwerfen
Domainwissen	0,6223	nicht verwerfen
GUI	0,7106	nicht verwerfen
Interaktion	0,6099	nicht verwerfen
Metainformation/y	0,3608	nicht verwerfen
Security	0,6558	nicht verwerfen
Systemverhalten	0,2916	nicht verwerfen
Zukunftsplan	0,6215	nicht verwerfen

Tabelle A.1: Ergebnisse des Kruskal-Wallis Tests für Main

MainCS	p-Wert	H0
Alter	0	verwerfen
deutsch	0,1044	nicht verwerfen
englisch	0,4283	nicht verwerfen
männlich	0,4283	nicht verwerfen
weiblich	0,0028	verwerfen
Studium	0,0843	nicht verwerfen
Arbeit	0	verwerfen
Sonstiges	0,1012	nicht verwerfen
Stimmung	0,2232	nicht verwerfen
Reaktivität	0,8289	nicht verwerfen
Änderung	0,7184	nicht verwerfen
Business/n	0,5289	nicht verwerfen
Domainwissen	0,1047	nicht verwerfen
GUI	0,3317	nicht verwerfen
Interaktion	0,0002	verwerfen
Metainformation/y	0,7588	nicht verwerfen
Security	0,1715	nicht verwerfen
Systemverhalten	0,4583	nicht verwerfen
Zukunftsplan	0	verwerfen

Tabelle A.2: Ergebnisse des Kruskal-Wallis Tests für MainCS

Simple	p-Wert	H0
Stimmung	0,1638	nicht verwerfen
Reaktivität	0,0876	nicht verwerfen
e	0,0735	nicht verwerfen
i	0,8317	nicht verwerfen
y	0,3608	nicht verwerfen
n	0,5907	nicht verwerfen

Tabelle A.3: Ergebnisse des Kruskal-Wallis Tests für Simple

SimpleCS	p-Wert	H0
Alter	0	verwerfen
deutsch	0,1044	nicht verwerfen
englisch	0,1044	nicht verwerfen
männlich	0,4283	nicht verwerfen
weiblich	0,4283	nicht verwerfen
Studium	0,0028	verwerfen
Arbeit	0,0843	nicht verwerfen
Sonstiges	0	verwerfen
Stimmung	0,1012	nicht verwerfen
Reaktivität	0,2232	nicht verwerfen
e	0,8662	nicht verwerfen
i	0,1311	nicht verwerfen
y	0,0002	verwerfen
n	0,7184	nicht verwerfen

Tabelle A.4: Ergebnisse des Kruskal-Wallis Tests für SimpleCS

Sub	p-Wert	H0
Stimmung	0,1638	nicht verwerfen
Reaktivität	0,0876	nicht verwerfen
Algorithmus	0,0255	verwerfen
Begrifflichkeiten	0,0002	verwerfen
Bug o. Abstürze	0,5744	nicht verwerfen
Einführung	0,0258	verwerfen
Konsequenzen	0,7518	nicht verwerfen
Navigation	0,0314	verwerfen
Operation	0,2052	nicht verwerfen
Privacy	0,7108	nicht verwerfen
Systemspezifische Elemente	0,7106	nicht verwerfen
Unerwartetes Systemverhalten	0,4888	nicht verwerfen
Vulnerability	0,7995	nicht verwerfen

Tabelle A.5: Ergebnisse des Kruskal-Wallis Tests für Sub

SubCS	p-Wert	H0
Alter	0	verwerfen
deutsch	0,2461	nicht verwerfen
englisch	0,2461	nicht verwerfen
männlich	0,0606	nicht verwerfen
weiblich	0,0606	nicht verwerfen
Studium	0,0003	verwerfen
Arbeit	0,4658	nicht verwerfen
Sonstiges	0,0002	verwerfen
Stimmung	0,1067	nicht verwerfen
Reaktivität	0,0309	verwerfen
Algorithmus	0,42	nicht verwerfen
Begrifflichkeiten	0,1284	nicht verwerfen
Bug o. Abstürze	0,709	nicht verwerfen
Einführung	0,7597	nicht verwerfen
Konsequenzen	0,0927	nicht verwerfen
Navigation	0,1157	nicht verwerfen
Operation	0,1688	nicht verwerfen
Privacy	0,7769	nicht verwerfen
Systemspezifische Elemente	0,596	nicht verwerfen
Unerwartetes Systemverhalten	0,0351	verwerfen
Vulnerability	0,8132	nicht verwerfen

Tabelle A.6: Ergebnisse des Kruskal-Wallis Tests für SubCS

Literaturverzeichnis

- [1] A. Bachmann and A. Bernstein. When process data quality affects the number of bugs: Correlations in software engineering datasets. In *2010 7th IEEE Working Conference on Mining Software Repositories (MSR 2010)*, pages 62–71. IEEE, 2010.
- [2] M. S. Bartlett. Properties of sufficiency and statistical tests. *Proceedings of the Royal Society of London. Series A-Mathematical and Physical Sciences*, 160(901):268–282, 1937.
- [3] B. D. Basic. *Distance Measures*, pages 397–398. Springer Berlin Heidelberg, Berlin, Heidelberg, 2011.
- [4] V. Benning. Chi-quadrat verstehen und berechnen, Apr. 2023. zuletzt besucht am: 23.02.2024, <https://www.scribbr.de/statistik/chi-quadrat/>.
- [5] G. Bohner, S. E. Hormuth, and N. Schwarz. Die stimmungsskala: Vorstellung und validierung einer deutschen version des „mood survey“. *Diagnostica*, 37(2):135–148, 1991.
- [6] W. Brunotte, L. Chazette, V. Klös, and T. Speith. Quo vadis, explainability?—a research roadmap for explainability engineering. In *International Working Conference on Requirements Engineering: Foundation for Software Quality*, pages 26–32. Springer, 2022.
- [7] 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.
- [8] R. Cudeck. Exploratory factor analysis. In *Handbook of applied multivariate statistics and mathematical modeling*, pages 265–296. Elsevier, 2000.
- [9] 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*, pages 329–338, 2023.
- [10] E. C. Fieller and E. S. Pearson. Tests for rank correlation coefficients: Ii. *Biometrika*, pages 29–40, 1961.
- [11] K. W. Fischer, P. R. Shaver, and P. Carnochan. How emotions develop and how they organise development. *Cognition and emotion*, 4(2):81–127, 1990.
- [12] M. Garzon, C.-C. Yang, D. Venugopal, N. Kumar, K. Jana, and L.-Y. Deng. *Dimensionality Reduction in Data Science*. Springer, 2022.
- [13] N. J. Gogtay and U. M. Thatte. Principles of correlation analysis. *Journal of the Association of Physicians of India*, 65(3):78–81, 2017.
- [14] M. Greenacre, P. J. Groenen, T. Hastie, A. I. d’Enza, A. Markos, and E. Tuzhilina. Principal component analysis. *Nature Reviews Methods Primers*, 2(1):100, 2022.
- [15] L. Guttman. Some necessary conditions for common-factor analysis. *Psychometrika*, 19(2):149–161, 1954.
- [16] M. Halkidi. *Hierarchical Clustering*, pages 1–5. Springer New York, New York, NY, 2016.
- [17] M. Herrmann. Analyse der wahrnehmung von stimmung in softwareprojekten durch explorative datenanalyse. *Leibniz Universität Hannover*, 2023.
- [18] H. F. Kaiser. Analytic determination of common factors. *Am Psychol*, 14(7):425, 1959.
- [19] H. F. Kaiser. An index of factorial simplicity. *psychometrika*, 39(1):31–36, 1974.
- [20] U. Kuckartz, S. Rädiker, T. Ebert, and J. Schehl. *Statistik: eine verständliche Einführung*. Springer-Verlag, 2013.
- [21] D. Kupczyk. Automatisierte detektion von erklärungsbedarf in nutzerfeedback zu software. *Leibniz Universität Hannover*, 2023.
- [22] M. A. Köhl, K. Baum, M. Langer, D. Oster, T. Speith, and D. Bohlender. Explainability as a non-functional requirement. In *2019 IEEE 27th International Requirements Engineering Conference (RE)*, pages 363–368, 2019.

- [23] B. Lin, F. Zampetti, G. Bavota, M. Di Penta, M. Lanza, and R. Oliveto. Sentiment analysis for software engineering: How far can we go? In *Proceedings of the 40th international conference on software engineering*, pages 94–104, 2018.
- [24] L. Luo, C. Arizmendi, and K. M. Gates. Exploratory factor analysis (efa) programs in r. *Structural Equation Modeling: A Multidisciplinary Journal*, 26(5):819–826, 2019.
- [25] J. Martensen. Analyse von einflüssen in der sentimenterkennung von entwicklern. *Leibniz Universität Hannover*, 2022.
- [26] A. Navlani, 2019. zuletzt besucht am: 23.02.2024 ,<https://www.datacamp.com/tutorial/introduction-factor-analysis>.
- [27] F. Nielsen. *Introduction to HPC with MPI for Data Science*. Springer, 2016.
- [28] N. Novielli, F. Calefato, D. Dongiovanni, D. Girardi, and F. Lanubile. Can we use se-specific sentiment analysis tools in a cross-platform setting? In *Proceedings of the 17th International Conference on Mining Software Repositories*, pages 158–168, 2020.
- [29] M. Obaidi, L. Nagel, A. Specht, and J. Klünder. Sentiment analysis tools in software engineering: A systematic mapping study. *Information and Software Technology*, page 107018, 2022.
- [30] K. Pearson. X. on the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 50(302):157–175, 1900.
- [31] M. Ringnér. What is principal component analysis? *Nature biotechnology*, 26(3):303–304, 2008.
- [32] 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.
- [33] P. Shaver, J. Schwartz, D. Kirson, and C. O’connor. Emotion knowledge: further exploration of a prototype approach. *Journal of personality and social psychology*, 52(6):1061, 1987.
- [34] N. Shrestha. Factor analysis as a tool for survey analysis. *American Journal of Applied Mathematics and Statistics*, 9(1):4–11, 2021.
- [35] C. Spearman. The proof and measurement of association between two things. *American Journal of Psychology*, 1961.

- [36] B. Underwood and W. J. Froming. The mood survey: A personality measure of happy and sad moods. *Journal of Personality Assessment*, 44(4):404–414, 1980.
- [37] M. Unterbusch, M. Sadeghi, J. Fischbach, M. Obaidi, and A. Vogelsang. Explanation needs in app reviews: Taxonomy and automated detection. In *2023 IEEE 31st International Requirements Engineering Conference Workshops (REW)*, pages 102–111. IEEE, 2023.
- [38] J. H. Ward Jr. Hierarchical grouping to optimize an objective function. *Journal of the American statistical association*, 58(301):236–244, 1963.
- [39] E. W. Weisstein. Bonferroni correction. <https://mathworld.wolfram.com/>, 2004.
- [40] C. Wohlin, P. Runeson, M. Höst, M. C. Ohlsson, B. Regnell, and A. Wesslén. *Experimentation in software engineering*. Springer Science & Business Media, 2012.