

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

# Stimmungsanalyse in Meetings von studentischen Softwareprojekten aus verschiedenen Blickwinkeln

Sentiment Analysis in Meetings of Student Software  
Projects from Different Viewpoints

## Masterarbeit

im Studiengang Informatik

von

Jacqueline Christina Klose

Prüfer: Prof. Dr. rer. nat. Kurt Schneider  
Zweitprüferin: Dr. rer. nat. Jil Ann-Christin Klünder  
Betreuerin: Dr. rer. nat. Jil Ann-Christin Klünder

Hannover, 21. Oktober 2021



# Erklärung der Selbstständigkeit

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

Hannover, den 21. Oktober 2021

---

Jacqueline Christina Klose



# Zusammenfassung

## **Stimmungsanalyse in Meetings von studentischen Softwareprojekten aus verschiedenen Blickwinkeln**

Die Kommunikation im Team spielt eine große Rolle, wenn es um die Motivation und Produktivität von Softwareentwicklern geht. Diese werden durch die Stimmung der einzelnen Teammitglieder beeinflusst. Aus diesem Grund werden zunehmend Methoden entwickelt, um die Stimmung zu analysieren. Für die Abfrage des emotionalen Empfindens wurde ein Instrument namens PANAS entwickelt. Dieses ist eine Sammlung von Begriffen, die in positive und negative Emotionen eingeteilt sind. Zu diesem Instrument gibt es Erweiterungen mit zusätzlichen psychologischen Abfragen, um die Emotionen noch genauer festzustellen. Für eine andere Feststellung der Emotion, wurde die Stimmungsanalyse entwickelt. Mithilfe dieser werden Aussagen in positiv, negativ oder neutral klassifiziert. Diese Methode wird für die Analyse im Team in den Meetings eingesetzt um somit die Polarität der Aussagen feststellen zu können. Die Annahme, dass mithilfe dieser Methoden die Stimmung beziehungsweise die Polarität der Aussagen im Team festgestellt werden können, wird in dieser Masterarbeit als Voraussetzung angenommen. Speziell geht es hierbei um die Untersuchung auf einen Zusammenhang zwischen der Stimmung vor und nach dem Meeting und den im Meeting getroffenen Aussagen und deren Polarität. Dafür wurden Meetings aus dem Modul Softwareprojekt von der Leibniz Universität Hannover untersucht und analysiert. Anschließend wurden Tests angewendet, um den Zusammenhang zwischen den Aussagen im Meeting und den Stimmungen vor und nach dem Meeting bestimmen zu können. Das Ergebnis der Analyse ist, dass mit dem vorhandenen Datensatz Aussagen zu dem positiven Anteil der Aussagen im Meeting und der Stimmung vor dem Meeting, sowie zu dem positiven Anteil der Aussagen im Meeting und den aufgabenbezogenen Konflikten und

der Mitarbeiterzufriedenheit getroffen werden konnten. Bei diesen drei Kombinationen besteht ein mittelstarker Zusammenhang.

# Abstract

## **Sentiment Analysis in Meetings of Student Software Projects from Different Viewpoints**

Team communication plays a major role when it comes to the motivation and productivity of software developers. These are influenced by the mood of the individual team members. For this reason, methods are increasingly being developed to analyze mood. An instrument called PANAS has been developed for querying emotional feelings. This is a collection of terms divided into positive and negative emotions. There are extensions to this instrument with additional psychological queries to determine emotion even more accurately. For another determination of the emotion, the mood analysis was developed. With the help of this, statements are classified as positive, negative or neutral. This method is used for the analysis in the team in the meetings in order to be able to determine the polarity of the statements. The assumption that the mood or the polarity of the statements in the team can be determined with the help of these methods is assumed as a prerequisite in this master thesis. Specifically, this involves an investigation of a connection between the mood before and after the meeting and the statements made in the meeting and their polarity. For this purpose, meetings from the module software project of the Leibniz University Hannover were examined and analyzed. Tests were then applied to be able to determine the correlation between the statements made in the meeting and the moods before and after the meeting. The result of the analysis is that with the available data set, statements could be made about the positive proportion of statements in the meeting and the mood before the meeting, as well as the positive proportion of statements in the meeting and the task-related conflicts and employee satisfaction. There is a moderately strong correlation for these three combinations.



# Inhaltsverzeichnis

<b>1</b>	<b>Einleitung</b>	<b>1</b>
1.1	Motivation . . . . .	1
1.2	Zielsetzung . . . . .	2
1.3	Struktur der Arbeit . . . . .	5
<b>2</b>	<b>Grundlagen</b>	<b>7</b>
2.1	Stimmungsanalyse . . . . .	7
2.1.1	SEnti-Analyzer . . . . .	9
2.1.2	GerVader . . . . .	10
2.2	PANAS . . . . .	11
2.3	SVM . . . . .	12
2.4	GLM . . . . .	15
<b>3</b>	<b>Datensätze</b>	<b>17</b>
3.1	Studentisches Softwareprojekt an der LUH . . . . .	17
3.2	Datenerhebung . . . . .	18
3.3	Resultierende Datensätze . . . . .	19
3.3.1	Meetingtranskripte . . . . .	19
3.3.2	Stimmungsbefragung . . . . .	19
3.3.3	Verarbeitung der Stimmungsbefragung . . . . .	20
<b>4</b>	<b>Qualitative Untersuchung</b>	<b>23</b>
4.1	Verarbeitung der Transkript-Datensätze . . . . .	23
4.1.1	Erweiterung des SEnti-Analyzer's . . . . .	23
4.1.2	Anwendung der Analysetools . . . . .	25
4.1.3	Herausfiltern der Übereinstimmungen . . . . .	27
4.2	Auswertung der Klassifizierung des SEnti-Analyzer . . . . .	28
4.3	Auswertung der Klassifizierung von GerVader . . . . .	29
4.4	Gemeinsame Auswertung . . . . .	30
4.4.1	Statistische Metrik . . . . .	30
4.4.2	Auswertung der übereinstimmenden Aussagen . . . . .	31

4.5	Erkenntnisse der Auswertungen . . . . .	33
<b>5</b>	<b>Quantitative Untersuchung</b>	<b>35</b>
5.1	Untersuchung mit einer SVM und einer GLM . . . . .	35
5.1.1	Vorbereitung . . . . .	36
5.1.2	Untersuchung Forschungsfrage 1 . . . . .	36
5.2	Statistische Untersuchung der Forschungsfragen . . . . .	40
5.2.1	Auswahl der Statistischen Tests . . . . .	41
5.2.2	Untersuchung mit den Statistischen Tests . . . . .	43
5.3	Ergebnisse . . . . .	47
<b>6</b>	<b>Diskussion</b>	<b>51</b>
6.1	Beantwortung der Forschungsfragen . . . . .	51
6.2	Validität . . . . .	55
6.3	Grenzen der Arbeit . . . . .	56
6.4	Verwandte Arbeiten . . . . .	56
<b>7</b>	<b>Zusammenfassung und Ausblick</b>	<b>61</b>
7.1	Zusammenfassung . . . . .	61
7.2	Ausblick . . . . .	62
<b>A</b>	<b>Anhang</b>	<b>65</b>

# Abbildungsverzeichnis

2.1	Decision Boundary . . . . .	14
2.2	Lineare Regression . . . . .	15
3.1	Positive and Negative Affect Schedule (PANAS)-Affekte und psychologischen Abfragen im Durchschnitt von Team 1 . . . . .	21
3.2	PANAS-Affekte und psychologischen Abfragen im Durchschnitt von Team 20 . . . . .	21
4.1	Polaritäre Anteile der Analyse des SEnti-Analyzer zum Team 1 . . . . .	25
4.2	Polaritäre Anteile der Analyse des SEnti-Analyzer zum Team 20 . . . . .	25
4.3	Polaritäre Anteile der Analyse von GerVader zum Team 1	26
4.4	Polaritäre Anteile der Analyse von GerVader zum Team 20 . . . . .	26
4.5	Ergebnisse der Meetinganalyse mit dem SEnti-Analyzer	29
4.6	Ergebnisse der Meetinganalyse mit GerVader . . . . .	30
4.7	Polaritäre Anteile der Gemeinsamkeiten der Analyse- tools zum Team 1 . . . . .	32
4.8	Polaritäre Anteile der Gemeinsamkeiten der Analyse- tools zum Team 20 . . . . .	32
4.9	Übereinstimmende Klassifikationen . . . . .	33
5.1	Confusion Matrix . . . . .	40
A.1	PANAS-Affekte und psychologischen Abfragen im Durchschnitt von Team 5 . . . . .	67
A.2	PANAS-Affekte und psychologischen Abfragen im Durchschnitt von Team 16 . . . . .	67
A.3	PANAS-Affekte und psychologischen Abfragen im Durchschnitt von Team 21 . . . . .	67

A.4	PANAS-Affekte und psychologischen Abfragen im Durchschnitt von Team 34 . . . . .	67
A.5	PANAS-Affekte und psychologischen Abfragen im Durchschnitt von Team 37 . . . . .	68
A.6	Polaritäre Anteile der Analyse des SEnti-Analyzer zum Team 5 . . . . .	68
A.7	Polaritäre Anteile der Analyse von GerVader zum Team 5	68
A.8	Polaritäre Anteile der Analyse des SEnti-Analyzer zum Team 16 . . . . .	69
A.9	Polaritäre Anteile der Analyse von GerVader zum Team 16 . . . . .	69
A.10	Polaritäre Anteile der Analyse des SEnti-Analyzer zum Team 21 . . . . .	69
A.11	Polaritäre Anteile der Analyse von GerVader zum Team 21 . . . . .	69
A.12	Polaritäre Anteile der Analyse des SEnti-Analyzer zum Team 34 . . . . .	70
A.13	Polaritäre Anteile der Analyse von GerVader zum Team 34 . . . . .	70
A.14	Polaritäre Anteile der Analyse des SEnti-Analyzer zum Team 37 . . . . .	70
A.15	Polaritäre Anteile der Analyse von GerVader zum Team 37 . . . . .	70
A.16	Polaritäre Anteile der Gemeinsamkeiten der Analyse-tools zum Team 5 . . . . .	71
A.17	Polaritäre Anteile der Gemeinsamkeiten der Analyse-tools zum Team 16 . . . . .	71
A.18	Polaritäre Anteile der Gemeinsamkeiten der Analyse-tools zum Team 21 . . . . .	71
A.19	Polaritäre Anteile der Gemeinsamkeiten der Analyse-tools zum Team 34 . . . . .	71
A.20	Polaritäre Anteile der Gemeinsamkeiten der Analyse-tools zum Team 37 . . . . .	72

# Tabellenverzeichnis

2.1	PANAS-Items . . . . .	12
4.1	Anzahl der polaritären Zuordnung der Analysetools im Vergleich von Team 1 . . . . .	27
4.2	Anzahl der polaritären Zuordnung der Analysetools im Vergleich von Team 20 . . . . .	27
4.3	Übereinstimmungsquote ausgewählter Teams mit Gesamtsumme . . . . .	31
5.1	Konfigurationen der Maschinenmodelle für Ansatz 1 . .	37
5.2	Konfigurationen der Maschinenmodelle für Ansatz 2 . .	38
5.3	Konfigurationen der Maschinenmodelle für Ansatz 3 . .	38
5.4	Konfigurationen der Maschinenmodelle für Ansatz 4 . .	39
5.5	Skala zur Analyse eines Zusammenhangs . . . . .	42
5.6	Zusammenfassung der statistischen Untersuchung der Forschungsfrage 1 . . . . .	48
5.7	Zusammenfassung der statistischen Untersuchung der Forschungsfrage 2 . . . . .	48
5.8	Zusammenfassung der statistischen Untersuchung der Forschungsfrage 3 . . . . .	49
A.1	Abfragen der psychologischen Befragung im Softwareprojekt mit den jeweiligen Kategorien . . . . .	65
A.2	Anzahl der polaritären Zuordnung der Analysetools im Vergleich von Team 5 . . . . .	66
A.3	Anzahl der polaritären Zuordnung der Analysetools im Vergleich von Team 16 . . . . .	66
A.4	Anzahl der polaritären Zuordnung der Analysetools im Vergleich von Team 21 . . . . .	66
A.5	Anzahl der polaritären Zuordnung der Analysetools im Vergleich von Team 34 . . . . .	66

A.6	Anzahl der polaritären Zuordnung der Analysetools im Vergleich von Team 37 . . . . .	67
-----	--	----

# Abkürzungsverzeichnis

**LUH** Leibniz Universität Hannover

**HCI** Institut für Mensch-Maschine-Kommunikation

**PSY** Institutes für Psychologie

**PANAS** Positive and Negative Affect Schedule

**SVM** Support Vector Machine

**SVC** Support Vector Classifier

**SVR** Support Vector Regressor

**GLM** Generalized Linear Model

**SK** soziale Konflikte

**AK** aufgabenbezogene Konflikte

**MZ** Mitarbeiterzufriedenheit

**NA** Negative Affects

**PA** Positive Affects

**PositiveT1** positive PANAS-Stimmung zum Zeitpunkt t1

**NegativeT1** negative PANAS-Stimmung zum Zeitpunkt t1

**PositiveT2** positive PANAS-Stimmung zum Zeitpunkt t2

**NegativeT2** negative PANAS-Stimmung zum Zeitpunkt t2



# Kapitel 1

## Einleitung

Die Entwicklung von Software hat heutzutage viele Aspekte, "weil die Entwicklung von großen Softwaresystemen viel mehr ist als das reine Programmieren. Es muss nämlich der ganze Prozess von der ersten Idee bis zum fertigen System bewältigt werden." [5]

### 1.1 Motivation

Ein Aspekt der Softwareentwicklung ist die Arbeitsweise. Wird diese betrachtet, gibt es zunehmend Entwicklungsteams statt Einzelpersonen [29]. In diesen Teams gibt es wiederum viele Aspekte zu betrachten, wenn ein Team so erfolgreich wie möglich sein soll [19]. Probleme hierbei liegen am häufigsten an mangelnder oder fehlerhafter Kommunikation [27]. Diese Probleme müssen gelöst werden, um Schwierigkeiten zu beheben. Die Frage, die sich dabei stellt, ist, wie der Ursprung des Problems gefunden werden kann.

Für eine gute Zusammenarbeit untereinander und eine ständige Verbesserung der Teamarbeit spielen soziale Faktoren eine wichtige Rolle [3]. Dabei ist es wichtig alle Dinge auszuwerten, die auf die Motivation und die Produktivität der Teammitglieder schließen lassen und somit Punkte zu finden, die verbessert werden können. Ist die Stimmung im Team beispielsweise gut, wirkt sich das auch auf die Motivation und die allgemeine Zufriedenheit aus [21]. Auch negative Stimmungen haben dabei einen Einfluss gezeigt, wie 2016 in einem Paper von Bakshi et al. [3] veröffentlicht wurde.

Um diese Stimmung zu analysieren, werden zunehmend Tools und Befragungen eingesetzt. In einem Paper von D'Andrea et al. [11] wurden zahlreiche dieser Tools mit kleinen Beschreibungen aufgelistet. Es gibt beispielsweise SentiStrenght, ein Tool, welches Aussagen anhand

eines Lexikons klassifiziert oder Sentiment140, mit welchem Tweets klassifiziert werden können oder PANAS-t mit welchem die Stimmung in Kommentaren auf Twitter analysiert wird. Diese Tools analysieren auf unterschiedliche Art und Weise, aber sie haben das gemeinsame Ziel die Stimmung zu analysieren.

Nun gibt es zum einen die Feststellung der Stimmung im Team und zum anderen die Analyse, wie die getätigten Aussagen zu klassifizieren sind. Daraus ergibt sich die Frage eines Zusammenhangs zwischen diesen unterschiedlichen Analysen. Dieser Zusammenhang soll in dieser Arbeit analysiert werden.

## 1.2 Zielsetzung

Ziel dieser Masterarbeit ist es die Stimmung von Teammitgliedern festzustellen, indem Auswertungen über die Stimmung nach bestimmten Kriterien und Auswertungen über die polaritären Aussagen in den Meetings mithilfe von Tools gemacht werden und einen Zusammenhang dieser Analysen zu untersuchen. Die Polarität von Aussagen steht in diesem Kontext für die Klassifikation von positiven, negativen oder neutralen Aussagen. Dies wird durch folgende Aussage aus einem Paper von Wolfgrube bestätigt: "Es kann dabei in zwei Polaritäten unterschieden werden: positiv und negativ. Ist ein Wort oder eine Aussage keiner der beiden Polaritäten zuzuordnen gilt sie als neutral." [35] Der zu untersuchende Aspekt ist, ob Zusammenhänge zwischen den einzelnen Datensätzen bestehen. Mithilfe dieser Untersuchung kann festgestellt werden, ob die Stimmung der Teammitglieder von Aussagen im Meeting beeinflusst werden oder ob die Aussagen im Meeting von der Stimmung beeinflusst werden. In der vorliegenden Arbeit soll dies analysiert und ausgearbeitet werden.

Für eine detaillierte Beantwortung wurden die folgenden Forschungsfragen erstellt, die in dieser Arbeit qualitativ und quantitativ untersucht werden sollen. Die zentrale Forschungsfrage wird durch die anderen drei Forschungsfragen in der Diskussion der Ergebnisse beantwortet. Diese ist wie folgt aufgestellt:

### Zentrale Forschungsfrage

Welcher Zusammenhang besteht zwischen der Stimmung im Team und den polaritären Aussagen in Meetings?

Die zentrale Forschungsfrage behandelt dem gesamten Zusammenhang der Stimmung im Team und den Polaritäten der Aussagen im Meeting. Mit diesem Zusammenhang können Probleme im Team aufgezeigt werden. Hierbei ist es wissenswert, ob die Stimmung vor dem Meeting die Aussagen im Meeting beeinflusst, oder ob die Aussagen im Meeting die Stimmung nach dem Meeting beeinflusst. Diese Fragen wurden als Forschungsfragen 1 bis 3 aufgestellt, um einzeln beantwortet werden zu können. Mit der einzelnen Beantwortung werden die unterschiedlichen Bereiche der zentralen Forschungsfrage abgedeckt. Diese sind die Veränderung der Stimmung über das Meeting hinweg, die Polarität der Aussagen durch vorherige Stimmung im Team und die Polarität der Stimmung durch Aussagen im Meeting.

#### Forschungsfrage 1

Wie verändert sich die Stimmung im Team nach dem Meeting, wenn dieses viele negative oder positive Aussagen hatte?

Mit der Forschungsfrage 1 wird der Teil der zentralen Forschungsfrage abgedeckt, der sich auf die Veränderung der Stimmung von vor dem Meeting auf nach dem Meeting im Bezug auf die polaritären Aussagen im Meeting bezieht. Dadurch wird der Verlauf der Stimmung über das Meeting hinweg abgedeckt. Besteht ein Zusammenhang der verbesserten Stimmung, weil im Meeting viele positive oder negative Aussagen aufgetreten sind, oder verschlechtert sich die Stimmung dadurch. Diese Fragen auf einen Zusammenhang sollen mit der Forschungsfrage 1 beantwortet werden.

#### Forschungsfrage 2

Besteht ein Zusammenhang zwischen der Stimmung vor dem Meeting und dem Anteil positiver oder negativer Aussagen im Meeting?

Die Forschungsfrage 2 bezieht sich auf die Polaritäten der Aussagen im Bezug auf die Stimmung vor dem Meeting. Wenn die Stimmung vorher negativ war, gibt es einen Zusammenhang auf die Aussagen im Meeting? In einem Paper von Kauffeld et al. wird von einer Verbreitung negativer Stimmung durch verbale Äußerungen gesprochen: " Verbale Äußerungen in Teammeetings wie z. B. Jammern sind ansteckend, d. h.

auf ein Jammern folgt überzufällig häufig ein weiteres Jammern." [22] Gerade solch eine Aussage soll mit dieser Forschungsfrage untersucht werden, und zwar ob ein Zusammenhang zwischen der Stimmung vor dem Meeting und der polaritären Ausrichtung des Meetings besteht.

### Forschungsfrage 3

Welche anderen psychologische Abfragen stehen im Zusammenhang mit der Polarität im Meeting?

Die Forschungsfrage 3 bezieht sich auf andere psychologische Abfragen. Diese sind die sozialen und aufgabenebezogene Konflikte und die Mitarbeiterzufriedenheit. Mit der Forschungsfrage soll beispielsweise beantwortet werden, ob es einen Zusammenhang gibt, wenn das Meeting gut oder schlecht verlief, dass dann auch beispielsweise abgefragte Konflikte verstärkt vertreten sind. Ebenfalls beantworten soll diese Frage, ob ein Zusammenhang besteht, wenn das Meeting positiv oder negativ verlief, dass die abgefragte Zufriedenheit im Team stark vorhanden ist. Diese psychologischen Abfragen gehen auf die Beeinflussung des Wohlbefindens der einzelnen Mitarbeiter ein. Geringe Konflikte und eine hohe Zufriedenheit tragen zur Motivation der Mitarbeiter bei. [12]

Diese Forschungsfragen werden mithilfe von unterschiedlichen Tools, wie dem Stimmungsanalysetool GerVader und der PANAS-Befragung, die die Stimmung vor und nach einem Meeting und die Polarität von Aussagen im Meeting analysieren, und anderen Auswertungsmethoden, wie beispielsweise statistischer Analyse, beantwortet.

Die Polarität des Datensatzes der Aussagen im Meeting (Abschnitt 3.3.1) wird mithilfe eines Tools namens SEnti-Analyzer bestimmt, welches im Fachgebiet Software Engineering an der Leibniz Universität entwickelt wurde beziehungsweise für Erweiterungen noch in der Entwicklung steht. Anschließend werden diese zugeordneten Polaritäten mit einem zweiten Tool namens GerVader analysiert, um bei Übereinstimmung den Zuordnungen der Tools die unabhängig voneinander gleich klassifizierten Aussagen nutzen zu können. Dieses Tool ist ein öffentlich zugängliches Tool, wessen Herkunft in Abschnitt 2.1.2 genauer beschrieben wird.

Die Analyse der Stimmung der Teams wird mit einer Auswertung

der Befragung des Teams durchgeführt. Die Befragung wurde durch eine Sammlung von Begriffen namens PANAS (Abschnitt 2.2) durchgeführt. Anschließend werden die Ergebnisse quantitativ und qualitativ in einen Zusammenhang gebracht, sodass Aussagen im Bezug auf die Forschungsfragen getroffen werden können.

## 1.3 Struktur der Arbeit

In Kapitel 2 werden die Grundlagen zu dieser Arbeit erklärt. Hierzu zählen die verwendeten Tools und Praktiken.

Kapitel 3 enthält die Erläuterung der Datensätze, die in dieser Masterarbeit verwendet werden. Zudem wird in diesem Kapitel die Verarbeitung des einen Datensatzes beschrieben. In Kapitel 4 wird dann die qualitative Untersuchung der Datensätze durchgeführt. Hierzu zählen die Erweiterung des Analysetools von Marc Herrmann [14], die Verarbeitung der Transkripte mit den zwei Analysetools und die Bewertung der genutzten Datensätze.

Kapitel 5 enthält die Untersuchung der Forschungsfragen durch verschiedene Verfahren. Hierbei werden unter anderem eine Support Vector Machine (SVM), sowie statistische Verfahren genutzt. In Kapitel 6 werden die Ergebnisse zusammengefasst und diskutiert und es werden verwandte Arbeiten aufgeführt. Kapitel 7 enthält die Zusammenfassung der Arbeit und mögliche Ausblicke für weitere Arbeiten.



# Kapitel 2

## Grundlagen

In den folgenden Abschnitten werden die Grundlagen für diese Masterarbeit erläutert. Hierbei wird insbesondere auf die Stimmungsanalyse eingegangen, da diese eines der Hauptthemen in dieser Arbeit ist. Anschließend wird ein Analysetool namens SEnti-Analyzer vorgestellt, welches in einer vorangegangenen Bachelorarbeit [15] entwickelt wurde. Hierfür werden die genutzten Arbeiten und Paper bezüglich des Analysetools vorausgesetzt. Für die weitere Analyse in dieser Masterarbeit wird PANAS als Verfahren für die Stimmung und GerVader als Tool für die Polaritäten von Aussagen genauer erklärt. Anschließend werden die Support Vector Machine und das Generalized Linear Model erklärt, da diese für die quantitative Analyse genutzt werden.

### 2.1 Stimmungsanalyse

Die Stimmungsanalyse, oder auch Sentiment Analysis, wird heutzutage in vielen Themengebieten eingesetzt. Diese sind beispielsweise die Analyse von Beiträgen auf Plattformen im Internet wie Twitter, von allgemeinen Beiträgen zu den zu analysierenden Themengebieten oder im Bereich der Softwareentwicklung Beiträge zu Problemen mit einer Programmiersprache. Mit der Stimmungsanalyse können Stimmungen von Aussagen analysiert werden. Grundsätzlich geht es um die Analyse der Polarität im Kommunikationsverlauf verschiedenen Kontaktpersonen. In dieser Masterarbeit ist besonders die Stimmungsanalyse im Bezug auf die Softwareentwicklung wichtig, da der zu analysierende Datensatz aus diesem Bereich stammt (vergl. Abschnitt 3.3.1). Hierbei sollen die Teams analysiert werden, um diese auf einen Zusammenhang mit der Stimmung vor uns nach dem Meeting zu testen.

## Funktion

Bei der Stimmungsanalyse werden Aussagen von Personen in drei verschiedene Polaritäten eingeteilt. Eine Aussage kann *Positiv*, *Neutral* oder *Negativ* sein. Die Zuordnung entscheidet sich anhand eines Lexikons bei der lexikonbasierten Methode oder anhand einer Klassifikation bei der Methode des maschinellen Lernens.

Das Lexikon, welches bei der lexikonbasierten Methode genutzt wird, enthält ein System von Begriffzuweisungen. Dabei sind die Begriffe mit ihren Polaritäten und der jeweiligen Gewichtung aufgelistet. Mithilfe des Lexikons wird jedem Wort einer Aussage eine Polarität zugeordnet. Die Werte liegen hierbei beispielsweise bei dem Tool SentiStrenght<sup>1</sup> im negativen zwischen -1 und -5(stark negativ) und im positiven zwischen +1 und +5(stark positiv). Diese Skala ist jedoch variabel, da anderen Tools auf eine andere Zahlenspanne verwenden wie VADER<sup>2</sup>, welches eine Skala von -1 bis +1 hat. Im Anschluss an die Bewertung werden die Polaritäten der einzelnen Teile der Aussage zusammengefügt. Dadurch entsteht eine gänzliche Polaritätszuordnung zu der entsprechenden Aussage.

## Sprache

Das verwendete Lexikon enthält jeweils nur eine Sprache, was bedeutet, dass beispielsweise ein Lexikon mit deutscher Sprache nicht für Aussagen auf englisch genutzt werden kann. Dies liegt daran, dass in dem Lexikon nur englischsprachige Wörter enthalten sind, wie dies bei SentiStrenght der Fall ist [28].

Die Tools könnten dennoch verwendet werden, wenn das Lexikon ausgetauscht wird [32] [31] oder die einzelnen Aussagen übersetzt werden würden. Bei wörtlicher Übersetzung lässt dies allerdings weitere Probleme auftreten, wie beispielsweise den Verlust der Umgangssprache und eventuelle Ironie in einigen Sätzen [10].

## Stimmungsanalyse im SE-Bereich

Viele bekannte Stimmungsanalysetools wurden nicht für den Gebrauch im Bereich des Software-Engineerings entwickelt. Hierzu zählen beispielsweise Rapidminer<sup>3</sup> oder HubSpot<sup>4</sup>. Ein nicht für den Bereich

---

<sup>1</sup><http://sentistrength.wlv.ac.uk/>

<sup>2</sup><http://www.nltk.org/api/nltk.sentiment.vader.html>

<sup>3</sup><https://rapidminer.com/>

<sup>4</sup><https://www.hubspot.com/>

der SE entwickeltes Tool stellt in dem Sinne ein Problem dar, dass in diesem Bereich beispielsweise das Wort "Problem" nicht negativ gemeint ist [8].

Die Stimmungsanalyse im Bereich SE ist wichtig, da hierbei die Stimmung im Team analysiert werden kann und bei Problemen eine schnelle Reaktion möglich ist. Dies ist bei Projekten im Bereich der Softwareentwicklung notwendig [27].

Das Fachgebiet Software Engineering an der Leibniz Universität Hannover hat diesbezüglich ein Analysetool für Audioaufzeichnungen in Meetings entwickelt. Dieses wird in Kapitel 2.1.1 genauer erläutert.

### 2.1.1 SEnti-Analyzer

Der SEnti-Analyzer ist ein Tool, welches im Rahmen der Bachelorarbeit von Marc Herrmann [14] an der Leibniz Universität Hannover im Fachgebiet Software Engineering entwickelt wurde. Dieses Tool wurde gebaut, um Audiodateien, die in Meetings aufgenommen wurden, zu transkribieren und im Anschluss auf Polaritäten auszuwerten. Vorgänger zu diesem Tool war die Masterarbeit von Julian Horstmann [16], welche im Anschluss in ein Paper [18] zusammengefasst wurde.

Die Namensgebung des SEnti-Analyzer's fand im Rahmen einer Publikation von Jil Klünder und Marc Herrmann statt [15].

Bei der Anwendung des Tools gibt es drei Möglichkeiten.

1. Eine Audiodatei zur Auswertung besprechen.
2. Eine besprochene Audiodatei speichern.
3. Eine vorhandene Audiodatei zur Auswertung auswählen.

#### Ablauf des Algorithmus

Beim Ausführen des Tools ab dem Zeitpunkt, zu dem eine Audiodatei ausgewählt oder besprochen wurde, wird diese zuerst eingelesen und in einer Text-Datei abgespeichert. Anschließend wird diese in eine CSV-Datei umgewandelt und mithilfe dieser Datei analysiert. Die Analyse geschieht über eine lexikonbasierte Stimmungsanalyse. Das bedeutet, dass die Aussagen nacheinander den Algorithmus durchlaufen und den einzelnen Abschnitten jeder Aussage eine Polarität zugewiesen wird. Anschließend werden die Aussagen mit ihren jeweiligen Polaritäten abgespeichert.

Nach dem kompletten Ablauf entstehen folgende fünf Dateien:

- transkript.txt
- ergebnis.txt
- ergebnis.csv
- statements.csv
- anzahl.csv

"transkript.txt" enthält die unausgewerteten Aussagen, die aus der Audiodatei gelesen wurden. "ergebnis.txt" und "ergebnis.csv" enthalten jeweils die Aussagen mit ihrer Klassifizierung nach positiv, negativ und neutral. Hierbei ist die CSV-Datei die konvertierte Textdatei in CSV-Format. "statements.csv" beinhaltet die einzelnen Aussagen des Meetings mit zugehörigen Polaritäten und anderen Parametern im Detail. "anzahl.csv" beinhaltet die Anzahl der insgesamt ausgewerteten Aussagen und die Anzahl der Aussagen mit der Zuordnungen wie viele positiv, negativ und neutral sind mit den jeweilig passenden Prozentsätzen dazu.

### 2.1.2 GerVader

GerVader ist ein Stimmungsanalysetool, welches von der ursprünglichen englischen Version "VADER" zu einem deutschen Tool umgewandelt wurde [30]. VADER und somit auch GerVader ist ein lexikonbasiertes Stimmungsanalysetool. Für die Änderung von VADER zu GerVader wurde lediglich das eingesetzte englische Lexikon gegen ein deutsches Lexikon ersetzt.

#### **Funktion**

Die Eingabe von GerVader erfolgt durch eine ".tsv"-Dateien. Hierbei ist die Anordnung besonders wichtig. Es gibt drei Spalten, wobei die erste die Zeilennummern, die zweite die Vorklassifizierungen der Aussagen und die dritte Spalte die Aussagen enthält. Mithilfe dieser Datei analysiert GerVader die Aussagen. Hierbei werden die Aussagen mit dem Lexikon bewertet und demnach mit den Polaritäten klassifiziert. Als Ergebnis wird eine genaue Auflistung der übereinstimmenden und nicht übereinstimmenden Aussagen in unterschiedlichen Dateien ausgegeben.

## 2.2 PANAS

PANAS ist eine weitere Methode Stimmungen auszuwerten. Hierbei handelt es sich um die Stimmung, die eine Person selbst empfindet. Ausgeschrieben bedeutet PANAS "Positive and Negative Affect Schedule". Es ist eine Sammlung von Begriffen, um die derzeitige Stimmung einer Person identifizieren zu können. Die Analyse wird mithilfe einer Befragung durchgeführt, wobei es zwei unterschiedlichen Ausprägungen in der Tiefe der Befragung gibt. Die kurze Version enthält weniger Begriffe in den positiven sowie negativen Bereichen, als die lange Version. Diese beinhaltet zehn positive und zehn negative Begriffe. Beide Versionen werden anhand einer 5-Punkte-Zahlenskala bewertet. Eine 1 (schwach zustimmend) ist die niedrigste Zustimmung und eine 5 (stark zustimmend) die höchste Zustimmung, dabei können nur ganze Zahlen angegeben werden [34]. Für diese Masterarbeit wurde die lange Version mit den deutschen Begriffen verwendet. Eine Auflistung ist in Tabelle 2.1 zu finden.

Die Befragung wird bei Einzelpersonen durchgeführt, in Kombination mit den Befragungen der anderen Teammitglieder gibt die Auswertung die Gesamtstimmung im Team an. Für die Auswertung werden zuerst von jeder Person im Team die positiven und die negativen Begriffe zu je einer Zahl zusammengefasst. Hierbei wird der Median der einzelnen Bewertungen gebildet.

Im Anschluss können diese zusammengefassten Werte aus dem gesamten Team für eine Berechnung des Mittelwerts genutzt werden. So entsteht für das befragte Team ein durchschnittlicher Wert für die positive, sowie ein Wert für die negative Stimmung.

### Auswertung

Für die Auswertung der PANAS-Daten in einem Team wird ein eindeutiges Schema verwendet. Zuerst wird für jeden der 20 Begriffe der Median aus dem Team gebildet. Dies verhindert, dass Ausreißer die Werte stark beeinflussen. Anschließend wird aus allen Positive Affects (PA) und allen Negative Affects (NA) je ein Mittelwert gebildet, sodass ein Mittelwert für die positiven Stimmungen und ein Mittelwert für die negativen Stimmungen entsteht.

Soll nun herausgefunden werden, ob die Stimmung im Team positiv oder negativ ist, so wird der negative Wert vom positiven abgezogen. Anschaulich kann dies wie folgt dargestellt werden:

$$s = \emptyset PA_{Team} - \emptyset NA_{Team} \quad \begin{cases} \text{positiv} & \text{falls } s > 0 \\ \text{neutral} & \text{falls } s = 0 \\ \text{negativ} & \text{falls } s < 0 \end{cases}$$

$s$  ist hierbei die Stimmung, *positiv*, *negativ* und *neutral* stehen für die Polarität und die Namen PA und NA sind der Tabelle 2.1 zu entnehmen.

Nr.	Deutsch	Englisch	Dimension
1	aktiv	active	PA
2	bekümmert	distressed	NA
3	interessiert	interested	PA
4	freudig erregt	excited	PA
5	verärgert	upset	NA
6	stark	strong	PA
7	schuldig	guilty	NA
8	erschrocken	scared	NA
9	feindselig	hostile	NA
10	angeregt	inspired	PA
11	stolz	proud	PA
12	gereizt	irritable	NA
13	begeistert	enthusiastic	PA
14	beschämt	ashamed	NA
15	wach	alert	PA
16	nervös	nervous	NA
17	entschlossen	determined	PA
18	aufmerksam	attentive	PA
19	durcheinander	jittery	NA
20	ängstlich	afraid	NA

Anmerkung. PA = Positiver Affekt, NA = Negativer Affekt.

Tabelle 2.1: Items der Positive and Negative Affect Schedule (PANAS) [6]

## 2.3 SVM

Dieser Abschnitt erklärt die SVM genauer, dies ist wichtig für das Verständnis der quantitativen Untersuchung in Kapitel 5. In einem Buch mit den zehn besten Algorithmen fürs Data Mining wird die SVM

als eines der robustesten und akkuratsten Methoden für diesen Zweck beschrieben: "Support vector machines (SVMs), including support vector classifier (SVC) and support vector regressor (SVR), are among the most robust and accurate methods in all well-known data mining algorithms." [36]

### Funktionsweise

Für die Teilung der Klassen werden Support-Vektoren erstellt. Diese sollen möglichst weit auseinander liegen, um die beste Teilung der Daten zu finden. Die Decision Boundary (Trennlinie zwischen den Klassen) liegt genau an der Position, wo der Support Vektor des dichtesten Punktes beider Klasse die größte Distanz enthält. Um diese Support-Vektoren zu berechnen wird die Formel eines linearen Klassifizierers genutzt. Dieser beinhaltet einen Gewichtsvektor  $w$ , die eingegebenen Daten  $x_i$  und ein festes Bias  $b$ . Die Formel des Klassifizierers sieht wie folgt aus:

$$w * x_i + b = 0$$

Mit dieser Formel erhalten alle Elemente, die zu der einen Klasse gehören, einen Wert mit negativem Vorzeichen und alle Elemente der anderen Klasse einen Wert mit positivem Vorzeichen. Wurde die Klassifikation für jedes Element durchgeführt, können zu jedem Element Geraden gebildet werden. Anschließend wird für jede Gerade die Distanz zur anderen Klasse berechnet, wobei die Decision Boundary die größte Distanz zu beiden Klassen hat. Verdeutlicht wird dies in Abbildung 2.1. Hier ist zu erkennen, dass die roten und die blauen Punkte zwei unterschiedliche Klassen darstellen. Die Support Vektoren liegen auf den am dichtesten zueinander liegenden Punkten der beiden Klassen. Der größte Abstand dieser zwei Vektoren ist die Decision Boundary, die im Punkt -5 auf der Y-Achse beginnt.

Support Vector Classifier (SVC) und Support Vector Regressor (SVR) sind hierbei unterschiedliche Methoden, die angewendet werden können, um Daten unterschiedliche zu klassifizieren. Die aufgeführte Abbildung 2.1 ist hierbei eine Darstellung des Classifiers.

### SVC

Ein SVC ist der standardmäßige Klassifizierer einer SVM. Die Klassifizierung gibt hierbei Aufschluss über die Zugehörigkeit zu einer der

---

<sup>5</sup><https://scikit-learn.org/stable/modules/svm.html>

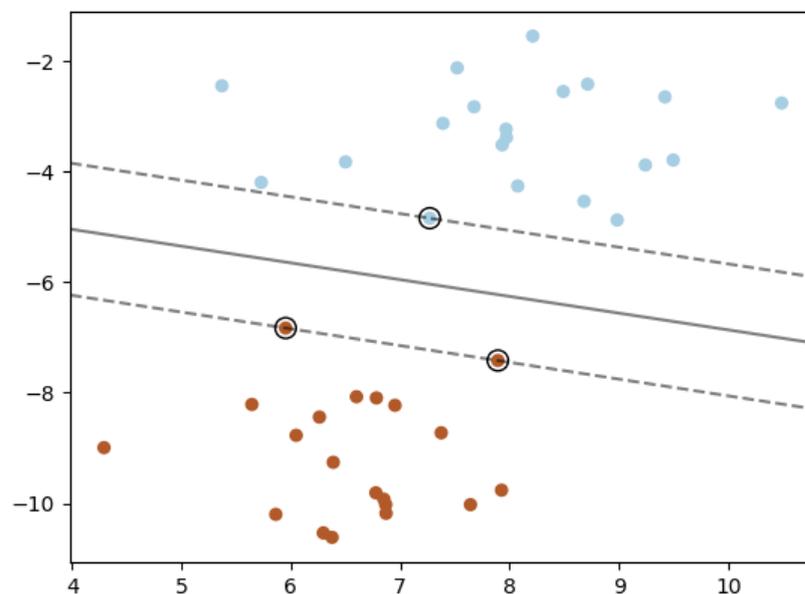


Abbildung 2.1: Beispielhafte Darstellung einer Decision Boundary mit Support Vektoren<sup>5</sup>

beiden Klassen, die der Classifier auswertet. Nach der Klassifizierung wird ein trainiertes Modell ausgegeben. Daraus kann beispielsweise ein Wert wie die Accuracy abgeleitet werden. Bei einer sehr geringen Accuracy hat die SVM viele Elemente falsch klassifiziert. Bei der Bewertung, wann eine Accuracy hoch ist, kommt es darauf an, wie viele mögliche Ergebnisse hätten geliefert werden können. Liegt die Accuracy bei 50% und kann zwei unterschiedlichen Ergebnisse liefern, so kann die SVM jedes mal eins der Ergebnisse errechnen und liegt mit einer Wahrscheinlichkeit von 50% richtig. Bei einer Accuracy von 100% klassifiziert die SVM jedes Element richtig.

## SVR

Der SVR ist eine Methode, die ähnlich wie die SVC funktioniert. Hierbei wird nicht der größte Abstand zwischen den Support Vektoren und der Decision Boundary gesucht, sondern der kleinste Abstand. Dementsprechend verläuft die Gerade durch die Klassen durch und teilt diese nicht auf. Je enger die Punkte an der Geraden liegen, umso stärker ist der Zusammenhang. In der folgenden Abbildung( 2.2) ist

dies genauer zu erkennen.

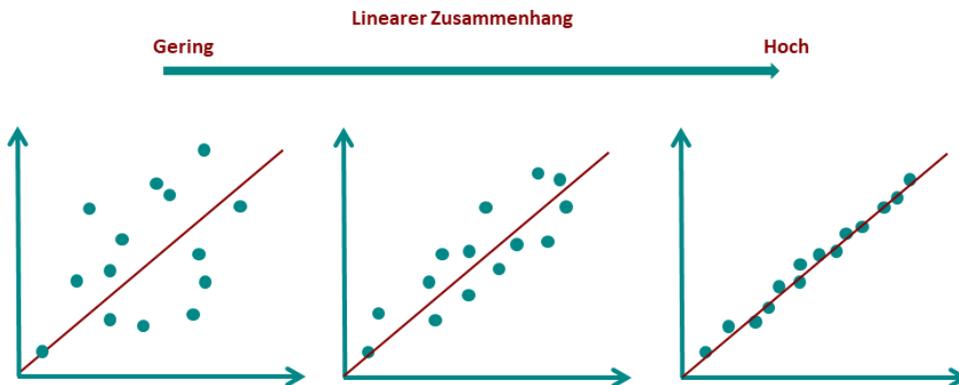


Abbildung 2.2: Beispielhafte Darstellung einer Lineare Regression von schwachem bis zum starken Zusammenhang<sup>6</sup>

Die linke Abbildung zeigt einen geringen Zusammenhang, da dort die Streuung an der Geraden sehr groß ist. Die rechte Abbildung hingegen hat einen hohen Zusammenhang, da die Elemente sehr nah an der Geraden liegen. Mithilfe dieser Darstellung kann demnach die Stärke des Zusammenhangs abgeschätzt werden.

## 2.4 GLM

Generalized Linear Model (GLM) ist eine Regressionsanalyse, die beispielsweise auf unterschiedliche Verteilungen eingestellt werden kann. Grundsätzlich geht beispielsweise die SVR von einer Normalverteilung aus. Bei der GLM lassen sich hierfür Verteilungen einstellen. Beispiele hierfür wäre die Binominalverteilung oder die Poissonverteilung, die als Verteilung genutzt werden können.

Ein wichtiger Kennwert des GLM ist die Log-Likelihood. Dieser Wert ist ein Indikator, um anzugeben, ob ein Modell trainiert werden konnte oder nicht. Für ein sehr gut trainiertes Modell, sollte dieser Wert etwa bei "0" liegen. Je höher oder niedriger die Log-Likelihood ist, umso schlechter ist die Verschiebung der Eingaben. Bei einer Log-Likelihood von "Minus Unendlich" oder "Plus Unendlich" konnte kein Modell trainiert werden.

Ein weiterer wichtiger Kennwert ist die Signifikanz der unabhängigen Variablen. Bei dem GLM wird diese für die unabhängigen Variablen

<sup>6</sup><https://datatab.de/tutorial/lineare-regression>

berechnet. Grundsätzlich sind alle Variablen, die ein Signifikanzniveau größer als 5% haben, nicht für das Modell signifikant. Als Signifikanztest wird in der Mathematik beispielsweise der P-Wert verwendet. Die Signifikanz ist dafür da, um einen Fehler von maximal 5% herauszufinden. Aus diesem Grund wird die Signifikanz auch als Irrtumswahrscheinlichkeit bezeichnet.

Mit dem GLM soll also vor allem getestet werden, ob ein Modell existiert und wie gut der Zusammenhang der Variablen und deren Abhängigkeiten ist.

# Kapitel 3

## Datensätze

In diesem Kapitel wird auf die in dieser Arbeit wichtigen Datensätze eingegangen. Hierfür wird das Softwareprojekt der Leibniz Universität Hannover (LUH) erläutert und es wird auf die Datenerhebung eingegangen. Anschließend werden die resultierenden Datensätze präsentiert, mit denen in den folgenden Kapiteln weiter gearbeitet wird.

### 3.1 Studentisches Softwareprojekt an der LUH

Das Softwareprojekt<sup>1</sup> an der LUH ist ein Modul im Bachelorstudium im Studiengang Informatik. In der Regelstudienzeit findet es im 5.Semester. Es hat eine Dauer von drei Monaten und es gibt 8 Leistungspunkte für die Studierenden. Das zuständige Fachgebiet der LUH ist das Fachgebiet Software Engineering.

Die "Kunden" der Projekte kommen jedes Jahr aus unterschiedlichen Bereichen. In den vergangenen Jahren kamen die Projekte aus externen sowie internen Unternehmen und Instituten. Beispiele hierfür sind die Polizei Hannover oder das Institut für Mensch-Maschine-Kommunikation (HCI) von der LUH. Bei nicht ausreichender Anzahl an Projekten, werden vom Fachgebiet interne sowie virtuelle Projekte hinzugefügt. Jedes Projekt wird in der Regel zweimal vergeben, dies deckt die hohe Anzahl an Teilnehmenden ab. Der Kunde betreut hierbei zwei unterschiedliche Gruppen und hat am Ende des Projekts die Wahl, ob er eines dieser Projekte einsetzen und eventuell weiterentwickeln möchte.

Die Auswahl der Teammitglieder wird zugelost. Jeder Studierende

---

<sup>1</sup><https://www.pi.uni-hannover.de/de/se/lehre/swp/>

kann drei Wünsche äußern. Zusätzlich werden Stärken, wie gute Programmierkenntnisse einer Sprache oder gute Führungsqualität in Projekten, sowie Schwächen aufgeschrieben, um ein passendes Projekt und Team zu finden.

Die Durchführung des Projekts ist auf drei Monate beschränkt. Die ersten zwei Wochen sind für die Ausarbeitung der Spezifikation vorgesehen. Hierbei erarbeitet das Team gemeinsam mit dem Kunden die gewünschten Anforderungen und Features für das Projekt aus. Anschließend folgen zwei Iterationen. Die Beschreibung welche Funktionen in welcher Iteration erledigt werden sollte, wurde ebenfalls in der Spezifikation festgehalten. Die erste Iterationsphase ist dafür da, um grundlegende Features des Projekts zu erstellen. Hierbei wird nach der Projektmanagement Methode Scrum vorgegangen, in der wöchentlich ein Meeting mit dem Kunden stattfindet. Dieses Meeting ist dafür da, um dem Kunden den bisherigen Fortschritt zu zeigen und auf eventuelle Anpassungen zu reagieren.

In der zweite Iteration findet größtenteils die Implementierung des Projekts statt. Alle Erweiterungen und Funktionen, die laut der Spezifikation bis zum Ende des Projekts fertig sein müssen, mussten in dieser Phase umgesetzt werden. Die wöchentlichen Meetings mit dem Kunden finden weiterhin statt, sodass dieser weitere Änderungswünsche äußern konnte. Am Ende der beiden Iterationen ist das eigentliche Softwareprojekt vorbei. Die Phase danach ist lediglich für die Fehlerbehebung gedacht.

Im Anschluss an das Projekt wurde dieses in einer Abschlusspräsentation im Lichthof der LUH vorgestellt. Hierbei konnten die Projekte von allen Personen, die an dem Tag anwesend waren ausprobiert werden.

## 3.2 Datenerhebung

Die Datenerhebung fand im Rahmen des Softwareprojekts statt. Diese wurde von dem Fachgebiet Software Engineering in Zusammenarbeit mit der Abteilung Arbeits-, Organisations- und Sozialpsychologie des Institutes für Psychologie (PSY) von der Technischen Universität Braunschweig durchgeführt [24]. Die Erhebung fand in mehreren Zyklen statt. Der in dieser Arbeit verwendete Erhebungszeitraum beginnt mit der ersten Aufzeichnung im Jahr 2012/13 und endet mit den Aufzeichnungen im Jahr 2015/16.

Während des ersten Meetings im Softwareprojekt der einzelnen Teams wurde die Aufzeichnung für den ersten Datensatz gemacht. Die Teams

bestanden aus drei bis fünf Personen. Das Meeting wurde als Video mit Audio aufgezeichnet, wobei in einigen Fällen das Audio separat aufgenommen wurde, weil dadurch eine bessere Qualität entstand. Diese Audiodatei wurde anschließend manuell von der PSY verschriftlicht. Vor und nach dem Meeting wurde jeweils die Befragung für den zweiten erhobenen Datensatz durchgeführt.

### 3.3 Resultierende Datensätze

In dieser Masterarbeit werden zwei unterschiedliche Datensätze verwendet. Beide Datensätze bestehen aus 42 Teams, wobei diese Teams je nach Teamnummer übereinstimmen. Das bedeutet, dass Team 1 im 1. Datensatz (Meetingtranskripte) auch Team 1 aus dem 2. Datensatz (Stimmungsbefragung) ist.

In den folgenden Unterabschnitten werden die Datensätze genauer beschrieben. In den weiterführenden Kapiteln werden diese Datensätze jedoch für die anschließende Analyse weiter aufbereitet und verarbeitet.

#### 3.3.1 Meetingtranskripte

Der erste Datensatz besteht aus Meetingtranskripten. Diese entstanden während der ersten Besprechung in Softwareprojekten von 42 Teams an der Leibniz Universität Hannover. Die Aufnahmen wurde im den Jahren 2012/13 bis 2015/16 gemacht. Vier Audiodateien konnten nicht transkribiert werden, da diese eine schlechte Qualität hatten und deshalb keine Aussagen verstanden werden konnten. Zu diesen Teams gibt es deshalb keine Transkripte. Von den 42 Audiodateien blieben demnach 38 Transkripte, die weiterverarbeitet werden konnten.

Im Rahmen der Bachelorarbeit von David Schiller [23] wurden die Transkripte analysiert und überarbeitet. Dadurch sind vereinzelt weitere Transkripte weggefallen, da diese durch entfernen von unbrauchbaren Aussagen keinen weiteren Inhalt enthielten. Insgesamt sind zur Auswertung für diese Masterarbeit 35 Transkripte geblieben.

#### 3.3.2 Stimmungsbefragung

Der zweite Datensatz besteht aus Daten, die bei den selben Teams gesammelt wurden, wie bei dem erste Datensatz. Hierbei handelt es sich um die Dokumentation zur PANAS-Befragung (vergleiche Abschnitt 2.2) der jeweiligen Teammitglieder mit Teamzuordnung.

Diese Befragung wurden jeweils vor und nach dem ersten Meeting des Softwareprojekts, in dem die Transkripte des ersten Datensatzes aufgenommen wurden, durchgeführt. Die entstandenen Daten sind die *positive PANAS-Stimmung zum Zeitpunkt t1 (PositiveT1)*, die *negative PANAS-Stimmung zum Zeitpunkt t1 (NegativeT1)*, die *positive PANAS-Stimmung zum Zeitpunkt t2 (PositiveT2)* und die *negative PANAS-Stimmung zum Zeitpunkt t2 (NegativeT2)*.

Zusätzlich zur PANAS-Befragung gab es nach dem Meeting noch erweiterte Fragen bezüglich einiger psychologischen Merkmale (vergleiche Tabelle A.1). Dies war beispielsweise "Wie offensichtlich sind persönliche Konflikte in der Gruppe". Einzuordnen sind diese in drei Kategorien. *soziale Konflikte (SK)*, *aufgabenbezogene Konflikte (AK)* und *Mitarbeiterzufriedenheit (MZ)*. Die psychologischen Fragen, sowie die PANAS-Bewertungen, fanden anhand von Punkten statt, wobei 1 die niedrigste Übereinstimmung und 5 die höchste Übereinstimmung ist [24]. Lediglich die Mitarbeiterzufriedenheit wurde mit einer anderen Skala abgefragt. Hierbei ist die maximale Übereinstimmung mit der Zahl 7 begrenzt. Die Werte wurden zusammen je Teammitglied in einer Excel-Datei gespeichert und konnten daraus verwendet werden.

### 3.3.3 Verarbeitung der Stimmungsbefragung

Der Datensatz aus der Befragung mit den zehn positiven und zehn negativen PANAS-Begriffen vor und nach dem Meeting, sowie den zusätzlichen psychologischen Befragungen, mussten für die Analyse zusammengefasst werden, um eine Auswertung vornehmen zu können. Für die Auswertung wurde ein Python-Skript geschrieben, welches den Datensatz einliest und anschließend Mediane der einzelnen Abfragen pro Team bildet. Dies ist das typische Vorgehen, um PANAS-Angaben auszuwerten (siehe Abschnitt 2.2). So hat jedes Team für jeden PANAS-Begriff und für jede andere psychologische Angabe einen gemeinsamen Wert. Aus den zehn positiven und zehn negativen PANAS-Begriffen musste dann pro Team und Zeitpunkt der Mittelwert gebildet werden, sodass pro Team folgende sieben Werte entstanden:

- $\emptyset$  soziale Konflikte (SK)
- $\emptyset$  aufgabenbezogene Konflikte (AK)
- $\emptyset$  Mitarbeiterzufriedenheit (MZ)
- $\emptyset$  positive PANAS-Stimmung zum Zeitpunkt t1 (PositiveT1)

- $\emptyset$  negative PANAS-Stimmung zum Zeitpunkt t1 (NegativeT1)
- $\emptyset$  positive PANAS-Stimmung zum Zeitpunkt t2 (PositiveT2)
- $\emptyset$  negative PANAS-Stimmung zum Zeitpunkt t2 (NegativeT2)

Aus diesen gebildeten Mittelwerten aus den jeweiligen Medianen wurde im Anschluss pro Team ein Säulendiagramm gebildet. In Abbildung 3.1 ist das Säulendiagramm für Team 1 zu finden. Hier sind die sieben unterschiedlichen Säulen zu erkennen. Im Vergleich zur Abbildung 3.2 variieren die Ergebnisse von *NegativeT1* zu *NegativeT2* und von *PositiveT1* zu *PositiveT2*.

Diese Variation ist das Auswahlkriterium für die Team-Repräsentanten. Da die Forschungsfragen auf die verschiedenen Zeitpunkte vor und nach dem Meeting eingehen, sind hierbei die unterschiedlichen Ausprägungen der Veränderungen interessant. Dies ergibt neun unterschiedliche Kombinationen, wovon sieben in dem Datensatz auftreten. Bei Team 1 (Abbildung 3.1) beispielsweise sinkt die negative Stimmung vom Zeitpunkt t1 (vor dem Meeting) zum Zeitpunkt t2 (nach dem Meeting), die positive Stimmung jedoch steigt. Beim Team 20 (Abbildung 3.2) bleiben beide Stimmungen gleich. Die jeweiligen Repräsentanten der einzelnen Kombinationen sind beispielhaft in Abbildung 3.1 und 3.2 zu finden, die weiteren Repräsentanten befinden sich im Anhang Abbildung A.1 bis A.5.

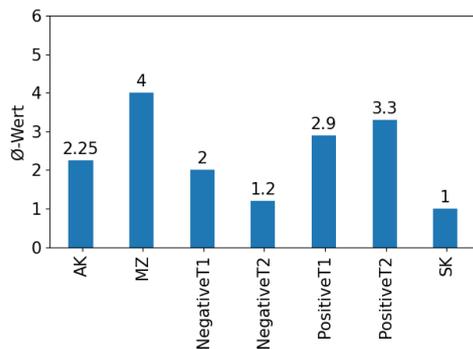


Abbildung 3.1: PANAS-Affekte und psychologischen Abfragen im Durchschnitt von Team 1

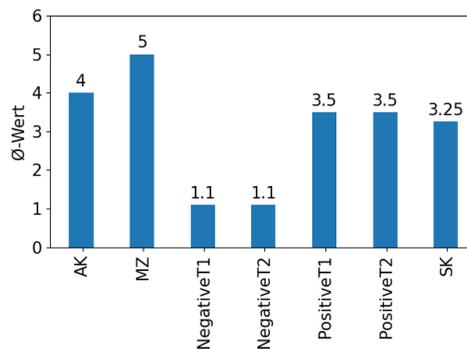


Abbildung 3.2: PANAS-Affekte und psychologischen Abfragen im Durchschnitt von Team 20



# Kapitel 4

## Qualitative Untersuchung

Die Analyse der Transkripte wurde mithilfe von zwei unterschiedlichen Tools durchgeführt. In der Bachelorarbeit von David Schiller [23] wurden diese auf ihre Genauigkeit und Richtigkeit bewertet. Mithilfe der Tools entstanden im Rahmen dieser Masterarbeit die im folgenden Abschnitt dargestellten Diagramme zur Auswertung der Transkripte mit ihren klassifizierten Aussagen. Diese Analyse ist dafür da, um den Datensatz qualitativ zu untersuchen. Anschließend werden die Übereinstimmungen der von den zwei Tools klassifizierten Aussagen dargestellt und ebenfalls auf ihre Qualität untersucht. Im letzten Abschnitt werden diese Erkenntnisse genauer analysiert und bewertet.

### 4.1 Verarbeitung der Transkript-Datensätze

In diesen Abschnitten wird die Verarbeitung der Transkripte mit den zwei Analysetools erklärt und anschaulich dargestellt. Weiter wird hier die Übereinstimmung der Aussagen von beiden Tools analysiert, um die unabhängig voneinander gleich klassifizierten Aussagen hervorzuheben.

#### 4.1.1 Erweiterung des Senti-Analyzer's

Das entwickelte Tool aus der Bachelorarbeit von Marc Herrmann [14] musste zur Verarbeitung, der für diese Masterarbeit zu analysierende Meetingtranskripte (Abschnitt 3.3), erweitert werden. Die bestehende Analyse ist dafür da, um neue Audiodateien aufzunehmen oder bestehende Audiodateien zu analysieren. Der zu analysierende Datensatz für diese Masterarbeit besteht aus CSV-Dateien, welche eingelesen und

ausgewertet werden.

Für diese Form der Auswertung musste ein Teil des Codes erweitert werden. Dies musste gemacht werden, damit nicht nur Audiodateien zur Analyse eingelesen, sondern bereits vorhandene Transkripte ebenfalls analysiert werden können. In der Datei "Sprache.py" befinden sich die Main-Methode. Mithilfe dieser Methode wird das Tool ausgeführt. Für die ausschließliche Textanalyse kann der erste Teil des Tools weggelassen werden, in welchem die Audiodateien eingelesen und als Text in eine CSV-Datei gespeichert werden.

Dem Tool wurde zur Vereinfachung dieser Umsetzung eine neue Python-Datei hinzugefügt, welche "Text.py" benannt wurde. In dieser wurde der gegebene Datensatz eingelesen, die benötigten Informationen heraus gelesen und im vorher festgelegtem Format abgespeichert. Nach diesem Schritt konnte die Verarbeitung durch den bestehenden Quellcode aufgerufen werden.

Zur Erleichterung der Auswertung von 35 Meetingtranskripten aus dem Datensatz, wurde dem Code eine Schleife hinzugefügt, sodass automatisch alle Transkriptdateien aus einem angegebenen Ordner analysiert werden können. Voraussetzung ist, dass alle Dateien in dem Ordner CSV-Dateien sind, die die Spalten-Überschrift "Transcript" enthalten, welche den zu analysierenden Transkriptinhalt haben.

Der Aufruf zum Ausführen des Algorithmus auf dem Testordner "Testtexte" lautet wie folgt:

```
python Text.py -d TestTexte
```

Des Weiteren ist es möglich nur einzelne CSV-Dateien einzulesen. Dies funktioniert mit dem folgenden Aufruf, wobei "Text.csv" durch jeden beliebigen Namen der CSV-Datei ersetzt werden kann:

```
python Text.py -f Test.csv
```

Mithilfe dieser Veränderungen wurden die Meetingtranskripte analysiert und ausgewertet. Die Ergebnisse wurden in einem Ordner "AusgabeText" unter Unterordnern mit dem jeweiligen Namen der CSV-Datei gespeichert. Aus der eingelesenen "Test.csv" entstand somit der Ordner "Test" mit den Inhalten "anzahl.csv", "ergebnis.csv", "ergebnis.txt" und "statements.csv". Die Erweiterung wurde in der "README.txt"-Datei erweitert. Diese einzelnen Ausgabedateien des Senti-Analyzers wurden zuvor in Kapitel 2.1.1 genauer erläutert.

### 4.1.2 Anwendung der Analysetools

Der Datensatz mit den Transkripten aus Abschnitt 3.3 ergab, nach der Verwendung des SEnti-Analyzer's aus Abschnitt 4.1.1, pro Team drei Dateien. Davon wird eine Datei, namens "anzahl.csv", für die Darstellung der Ergebnisse genutzt. In dieser Datei stehen die Anzahl positiver, neutraler und negativer Aussagen, die der SEnti-Analyzer klassifiziert hat. Aus diesen Ergebnissen wurden Kreisdiagramme erstellt, da dadurch die Ergebnisse veranschaulicht werden können. Die Diagramme, die zu den Repräsentanten passen, die in Abschnitt 3.3.3 ausgewählt wurden, sind in Abbildung 4.1 und 4.2, sowie im Anhang ab Abbildung A.6 zu finden. Es werden nur diese Diagramme aufgeführt, da in dem passenden Datensatz der Befragung bei diesen Diagrammen die unterschiedlichen Änderungen auftreten und somit die Analysen beispielhaft belegt werden können.

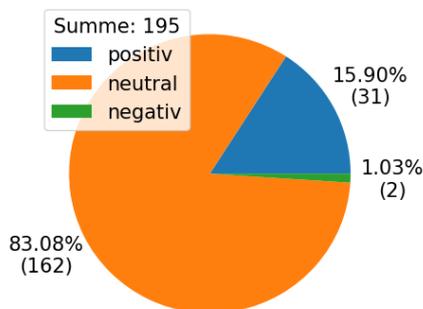


Abbildung 4.1: Polaritäre Anteile der Analyse des SEnti-Analyzer zum Team 1

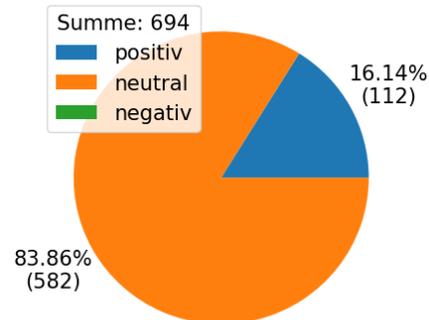


Abbildung 4.2: Polaritäre Anteile der Analyse des SEnti-Analyzer zum Team 20

Für die zweite Klassifikation der Daten wurde GerVader (Abschnitt 2.1.2) genutzt. Hierfür wurde die Datei "ergebnisse.csv" aus dem SEnti-Analyzer weiterverwendet.

In diesem Tool konnten nicht die Transkripte verwendet werden, da GerVader eine Vorklassifikation als Eingabe benötigt. Das bedeutet, dass jeder Aussage vorab eine Polarität zugewiesen sein muss. Dies ist von Vorteil, da dadurch nicht nur die Klassifikationsergebnisse erzeugt wurden, sondern gleichzeitig ein Vergleich von unterschiedlichen Klassifikationen zwischen dem SEnti-Analyzer und GerVader aufgezeigt werden können.

Aus der Klassifikation von GerVader aus positiven, negativen und

neutralen Aussagen wurden ebenfalls Kreisdiagramme erstellt. Diese sind in Abbildungen 4.3 bis 4.4 aufgeführt. Weitere Abbildungen sind im Anhang ab Abbildung A.7 zu finden. Die prozentualen Angaben wurden auf zwei Stellen hinter dem Komma gerundet.

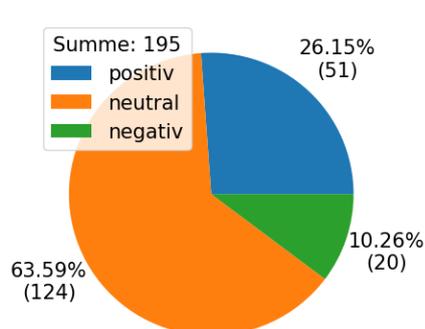


Abbildung 4.3: Polaritäre Anteile der Analyse von GerVader zum Team 1

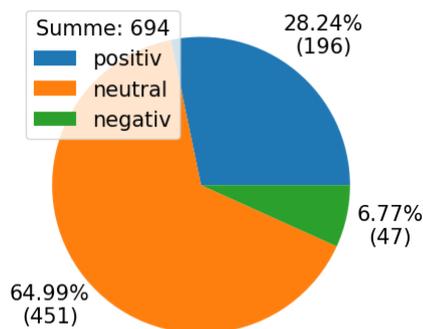


Abbildung 4.4: Polaritäre Anteile der Analyse von GerVader zum Team 20

In den Abbildungen 4.1 und 4.3 zum Team 1 sind einige Unterschiede zu erkennen. Der Anteil negativer Aussagen liegt bei GerVader bei 20 Aussagen. Der SEnti-Analyzer hat dagegen nur zwei Aussagen als negativ klassifiziert. Diese Aussagen müssen in ihrer Polarität nicht übereinstimmen. Das bedeutet, dass im Bezug auf das vorherige Beispiel von Team 1 auch 22 Aussagen von den zwei Klassifizierern als negativ bewertet worden sein können, dies aber durch die Anzahl der negativen Aussagen nicht festzustellen ist. Auf diese Auswertung wird in den folgenden Abschnitten noch genauer eingegangen. Der Anteil positiver Aussagen ist bei GerVader, wie auch schon der Anteil der negativen Aussagen, im Vergleich zum SEnti-Analyzer größer. Beim Team 20 mit den Abbildungen 4.2 und 4.4 verhält sich die Verteilung ähnlich.

Insgesamt wurde die gleiche Anzahl an Aussagen bewertet, sodass keine Aussagen von GerVader als ungültig erkannt worden war. Dies kann an der Aufbereitung des Datensatzes liegen, die in der Bachelorarbeit von David Schiller [23] gemacht wurde.

### 4.1.3 Herausfiltern der Übereinstimmungen

Nach den Analysen der Transkripte mit zwei unterschiedlichen Tools ist eine Analyse der Übereinstimmung sinnvoll. Der Grund hierfür ist, dass die übereinstimmenden Abschnitte der Klassifikation in den Kreisdiagrammen des SEnti-Analyzers nicht zwangsläufig die äquivalenten Aussagen in der Klassifikation vom GerVader sein müssen. Selbst bei gleichen prozentualen Anteilen muss nicht jede Aussage auch identisch klassifiziert sein. Durch das Herausfiltern der Übereinstimmung wird die Klassifikation der Aussagen, die gleich klassifiziert wurden, bestätigt.

Für diesen Teil der Analyse zeigen die Tabellen 4.1 und 4.2 sowie Tabellen A.2 bis A.6 im Anhang, die Anzahl der übereinstimmenden, sowie unterschiedlichen Aussagen der Repräsentanten. Diese Auswertung kam, wie zuvor erwähnt, mithilfe von GerVader während der Analyse als Nebenergebnis heraus.

		GerVader		
		negativ	positiv	neutral
SEnti-Analyzer	negativ	0	0	2
	positiv	1	7	23
	neutral	19	44	99

Tabelle 4.1: Anzahl der polaritären Zuordnung der Analysetools im Vergleich von Team 1

		GerVader		
		negativ	positiv	neutral
SEnti-Analyzer	negativ	0	0	0
	positiv	1	50	61
	neutral	46	146	390

Tabelle 4.2: Anzahl der polaritären Zuordnung der Analysetools im Vergleich von Team 20

In Tabelle 4.1 und 4.2 lässt sich deutlich erkennen, dass es einige Aussagen gab, die unterschiedlich von den zwei Tools erkannt wurden. Bei Team 1 gab es beispielsweise insgesamt 195 Aussagen, von denen laut dem SEnti-Analyzer 162 und laut GerVader 124 Aussagen neutral klassifiziert wurden. Übereinstimmend wurden jedoch nur 99 Aussagen neutral klassifiziert. Diese 99 Aussagen werden in ihrer Klassifikation

durch zwei unterschiedliche Analysetools bestätigt und somit wird in dieser Arbeit davon ausgegangen, dass diese Klassifikationen übernommen werden können. Die unterschiedliche Klassifikation der anderen Aussagen hingegen widerlegt die Richtigkeit der Klassifikationen, da somit nicht klar ist, welche Klassifikation die richtige ist.

## 4.2 Auswertung der Klassifizierung des SEnti-Analyzer

Mithilfe der Ergebnisse des SEnti-Analyzer wurden, wie in Abschnitt 4.1 beschrieben, Kreisdiagramme erstellt. An diesen lässt sich erkennen, dass der Großteil der Diagramme mit neutral klassifiziert wurde. Dies lässt auf eine sachliche Umgebung in dem Meeting schließen. Der geringste Anteil besteht aus negativen Klassifikationen. In einigen Fällen wurde sogar keine Aussage als negativ klassifiziert. Als Beispiel einer Klassifizierung lässt sich folgender Satz aus dem Datensatz nehmen:

"genau, die doofe suche"

Dieser wurde mit dem SEnti-Analyzer als positiv klassifiziert. Zusammengefasst aus allen Teams ergibt sich das folgende Boxplot-Diagramm. Dieses beinhaltet alle stattgefundenen Meetings aus den Transkripten mit der Klassifikation mit dem SEnti-Analyzer. An diesem Diagramm ist die Polaritätsverteilung der Aussagen im Durchschnitt zu erkennen und bildet somit einen Überblick über die gesamten Meetings ab.

Die Anzeige mit den negativen Aussagen lässt sich nicht gut darstellen, da nur sehr wenige negative Aussagen klassifiziert wurden und somit die Darstellung eines Boxplots, der gegen null konvergieren, sehr schwierig ist.

Wie in Abbildung 4.5 zu sehen, ist der Großteil der gesamten Meetings neutral, hierbei gibt es nur wenige Ausreißer. Dies spiegelt sich auf den einzelnen Kreisdiagrammen aus den vorherigen Abschnitten wider. Durchschnittlich sind etwa 79% der Aussagen im den Meetings neutral, wobei sich der Großteil zwischen 72% und 82% befindet.

### 4.3. AUSWERTUNG DER KLASSIFIZIERUNG VON GERVADER29

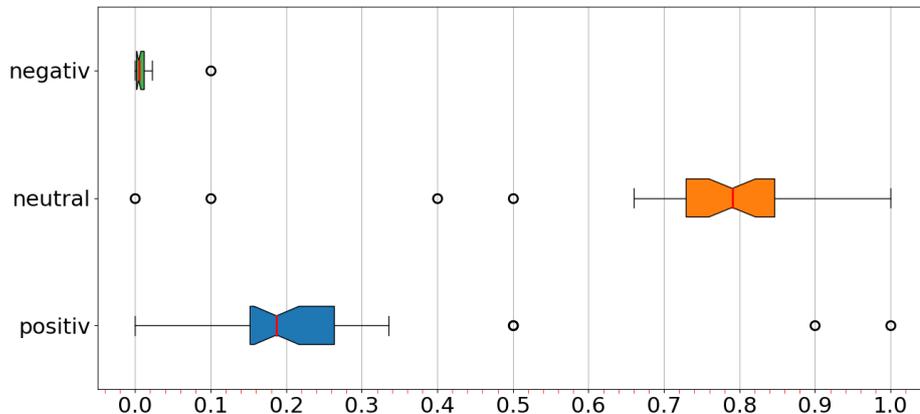


Abbildung 4.5: Ergebnisse der Meetinganalyse mit dem SEnti-Analyzer über alle 35 Meetings

## 4.3 Auswertung der Klassifizierung von GerVader

Die Ergebnisse bei GerVader unterscheiden sich nur geringfügig von denen des SEnti-Analyzer's. Wie schon in den Kreisdiagrammen der Repräsentanten aus Abschnitt 4.1 zu erkennen ist, gibt es bei der Analyse mit GerVader mehr negativ klassifizierte Aussagen, als beim SEnti-Analyzer. Ein Beispiel hierfür wäre der folgende Satz aus dem Datensatz, welcher bereits beim SEnti-Analyzer genannt wurde:

"genau, die doofe suche"

GerVader hat diese Aussage als negativ klassifiziert, der SEnti-Analyzer hat diese Aussage als positiv klassifiziert. Die negative Klassifizierung liegt hierbei an den einzelnen Begriffen, denn das Wort "doof" hat bei GerVader eine stärkere negative Polarität in dem Lexikon, als das Wort "genau".

Aus diesem Grund ist in Abbildung 4.6 ein Boxplot von den negativen Anteilen an Aussagen im Meeting zu erkennen. Der Median der Aussagen liegt bei etwa 8%. Auch die positiven und neutralen Anteile sind leicht verschoben. Der Anteil neutraler Aussagen liegt bei etwa 71% im Median und der Anteil positiver Aussagen liegt bei etwa 22%.

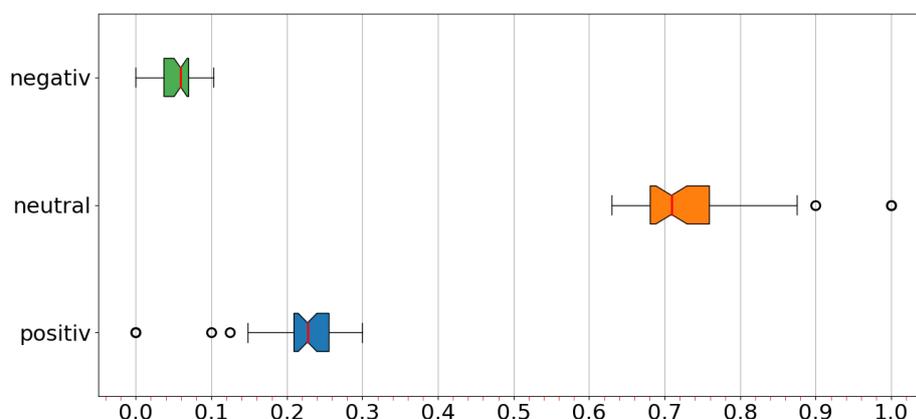


Abbildung 4.6: Ergebnisse der Meetinganalyse mit GerVader über alle 35 Meetings

Die Gesamtauswertung ergab im Vergleich zum SEnti-Analyzer wenige Veränderungen. Der hohe Anteil neutraler Aussagen lässt dennoch auf einen sachlichen Umgang im Meeting schließen. Der Anteil negativer Aussagen ist auch bei der Auswertung mit GerVader so gering, dass dies sich auf einzelne Aussagen im Meeting beschränken lässt. Zudem gibt es bei dem negativen Boxplot keine Ausreißer.

## 4.4 Gemeinsame Auswertung

In Abschnitt 4.1 wurden Tabellen mit den Unterschieden der Klassifikationen vom SEnti-Analyzer im Bezug auf GerVader aufgeführt. Mit diesen wird die Bestimmung der einzelnen klassifizierten Aussagen in ihrer Aussagekraft bestärkt. Aus diesem Grund werden in folgendem Abschnitt nur die Aussagen bewertet, die in beiden Analyseverfahren gleich klassifiziert wurden. Die anderen werden ausgelassen.

### 4.4.1 Statistische Metrik

In der Bachelorarbeit von David Schiller [23] wurde die Accuracy der einzelnen Stimmungsanalysetools ausgewertet. Beim SEnti-Analyzer lag diese bei 50,67%, GerVader hatte eine Accuracy von 57,33% bei dem überarbeiteten Datensatz.

Diese Werte liegen etwas unter der Übereinstimmung der gemeinsam klassifizierten Aussagen der beiden in dieser Arbeit verwendeten

Analysertools von etwa 60%. Dieser Wert kann aus der folgenden Tabelle entnommen werden. Diese zeigt die Werte der übereinstimmenden Klassifikationen im Bezug auf die gesamten Aussagen, anhand der ausgewählten Repräsentanten.

	Aussagen insgesamt	Aussagen gemeinsam	Übereinstimmung in %
Team 1	195	106	54
Team 5	942	570	61
Team 16	301	179	60
Team 20	694	440	63
Team 21	630	434	69
Team 34	284	184	65
Team 37	582	341	59
Gesamt	3628	2254	62

Tabelle 4.3: Übereinstimmungsquote ausgewählter Teams mit Gesamtsumme

Die Auswertung mit dem Stimmungsanalysetool GerVader zeigt, durch das in dieser Masterarbeit angewendet Vorgehen, dass die Klassifizierung des Senti-Analyzer's in die Analyse von GerVader eingegeben wurden, Gemeinsamkeiten und Unterschiede der Klassifizierung zum Senti-Analyzer auf. Eine Aussagen, die unterschiedlich klassifiziert wurde, ist das Wort "hm" oder "mh". Dieses wurde beim Senti-Analyzer als positiv klassifiziert, von GerVader jedoch als neutral. Ebenfalls die Aussage "ja" wurde beim Senti-Analyzer als positiv und bei GerVader als neutral klassifiziert. Das Wort "achso" wurde bei GerVader als neutral klassifiziert, beim Senti-Analyzer als negativ. Diese Begriffe sind häufig in den Gesprächen aufgetreten und dementsprechend in jedem auftretenden Fall unterschiedlich klassifiziert worden. Durch diese Begriffe und ähnliche Interjektionen verringert sich die Übereinstimmungsrate der beiden Analysetools.

#### 4.4.2 Auswertung der übereinstimmenden Aussagen

Durch die Tabellen in Abschnitt 4.1 und den dort enthaltenen übereinstimmenden Klassifikationen, aus den beiden Analysetools, ergeben sich die Kreisdiagramme in Abbildung 4.7 und 4.8, sowie im Anhang

ab Abbildung A.7. Die prozentualen Angaben wurden auf zwei Stellen hinter dem Komma gerundet.

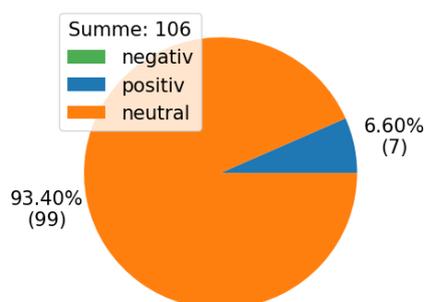


Abbildung 4.7: Polaritäre Anteile der Gemeinsamkeiten der Analysetools zum Team 1

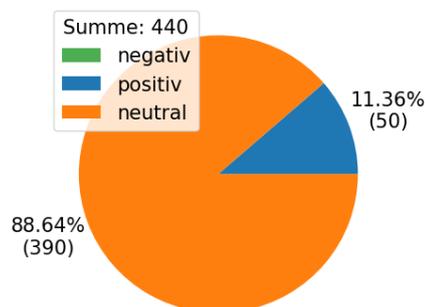


Abbildung 4.8: Polaritäre Anteile der Gemeinsamkeiten der Analysetools zum Team 20

In diesen zwei Diagrammen fällt auf, dass keine Aussage der übereinstimmenden Aussagen negativ klassifiziert wurde. Auf dieses Ergebnis deutet die Auswertung der beiden Stimmungsanalysetools, in Abschnitt 4.2 und 4.3, ebenfalls hin. Dort wird eine sehr geringe, gegen null konvergierende, negative Klassifikation angezeigt. Dies kann vor allem an der geringen negativen Klassifikation beim SEnti-Analyzer liegen, da alle Aussagen, die vom SEnti-Analyzer negativ klassifiziert wurde, vom GerVader ebenfalls negativ klassifiziert werden müssten, um den Anteil der negativen Klassifikation des SEnti-Analyzers nicht zu verringern. Den größten negativen Anteil hat Team 37 mit 2,23% (Abbildung A.14). Alle anderen Anteile an negativen Aussagen liegen unter diesem Wert, wobei vier der Teams aus den Repräsentanten keine negative Klassifikation aufweisen.

Insgesamt haben drei der repräsentativen Teams einen negativen Anteil nach der Auswertung der übereinstimmenden Aussagen. Hierbei sind die Anteile so gering, dass maximal 1,5% der Aussagen negativ klassifiziert wurden.

Das entstandene Boxplot-Diagramm zu den übereinstimmenden Aussagen aller Teams der beiden Analysetools ist in Abbildung 4.9 angezeigt.

Hier ist zu erkennen, dass der negative Anteil sehr gering ist und somit kaum darstellbar ist. Der positive Anteil liegt im Schnitt bei etwa 10% (vergleiche Abbildung 4.9).

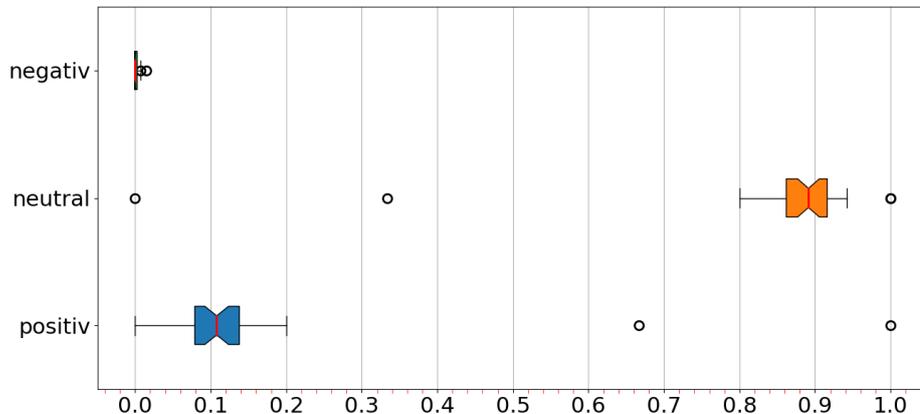


Abbildung 4.9: Übereinstimmende Klassifikationen von GerVader und dem Senti-Analyzer über alle Meetings

## 4.5 Erkenntnisse der Auswertungen

In diesem Kapitel wurden die Meetingtranskripte aus dem Softwareprojekt verarbeitet und mit den zwei Stimmungsanalysetools Senti-Analyzer und GerVader analysiert. Anschließend wurden die von beiden Tools gleich klassifizierten Aussagen ausgewertet. Der Großteil der mit dem Senti-Analyzer und GerVader klassifizierten Aussagen wurde neutral klassifiziert. Bei dem Senti-Analyzer liegt der Durchschnitt bei etwa 79%, bei GerVader bei etwa 71%. Der Anteil positiver Aussagen liegt bei beiden Tools bei etwa 20%, wobei GerVader einen etwas größeren Anteil positiv klassifiziert hat. Auch bei dem negativen Anteil hat GerVader einen größeren Anteil an Aussagen negativ klassifiziert. Beim Senti-Analyzer ist dieser Teil kaum darstellbar als Boxplot, da die Werte gegen null konvergieren. Bei GerVader liegt der negative Anteil bei etwa 8%.

Für eine Bestätigung der Klassifizierung wurden anschließend die einzelnen Aussagen geprüft, ob sie von beiden Analysetools gleich klassifiziert worden sind. Hierbei ergab sich, dass etwa 60% der Klassifizierungen der Aussagen übereinstimmen. Durch diesen Vergleich haben sich die Werte der Anteile an positiv, negativ und neutral klassifizierten Aussagen verschoben. Der Anteil neutraler Aussagen ist auf etwa 89% angestiegen. Der Anteil negativer Aussagen ist fast komplett verschwunden. Dies kann anhand der Repräsentanten verglichen werden. Hier haben nur drei der sieben Repräsentanten einen

negativen Anteil, wobei zuvor bei den einzelnen Klassifikationen nur ein Team bei der Analyse mit dem SEnti-Analyzer keinen negativen Anteil hatte. Der positive Anteil ist ebenfalls gesunken, wobei im Durchschnitt ein Anteil von etwa 10% bestehen bleibt.

# Kapitel 5

## Quantitative Untersuchung

In diesem Kapitel geht es um die Untersuchung der Zusammenhänge der zwei Datensätze aus den vorangegangenen Kapiteln. Diese stehen im Bezug auf die in der Einleitung genannten Forschungsfragen. Die in Abschnitt 4.4 erarbeitete Auswertung wird hierbei als ausschlaggebend für die Stimmungsanalyse der Meetings verwendet, da diese Anteile an Aussagen durch die gleiche Klassifikation bestätigt wurden. Für die Stimmung vor und nach dem Meeting werden die Daten aus Abschnitt 3.3.3 verwendet.

Für die quantitative Untersuchung gibt es unterschiedliche Möglichkeiten. In den folgenden Abschnitten wird zum einen eine SVM und eine GLM zur Untersuchung verwendet, zum anderen wird die Untersuchung mithilfe von Hypothesen durchgeführt.

### 5.1 Untersuchung mit einer SVM und einer GLM

Zur Untersuchung der Zusammenhänge wurden die Daten mit einer SVM verarbeitet und überprüft. Anschließend wurden die unabhängigen Variablen mithilfe der GLM evaluiert. Im folgenden wird die Vorgehensweise der Schritte und die Ergebnisse genauer beschrieben. Die Modelle wurde konzipiert, um die Forschungsfrage 1 (Abschnitt 1.2) auswerten zu können. Die Veränderung sind lediglich die Eingaben und die erwartete Ausgabe. Die wichtigen Daten sind je Team die positiven und negativen Werte vor und nach dem Meeting aus der PANAS-Befragung, sowie die Polaritäten der Stimmungsanalyse zu den Aussagen im Meeting.

Die verwendeten Daten sind die Daten aus allen Teams und nicht nur

den Repräsentanten.

### 5.1.1 Vorbereitung

Bei der Vorbereitung der Daten für die Modelle gab es verschiedene Vorgehensweisen. Zuerst mussten die Daten in Form von Arrays gebracht werden, welche in die Modelle als Trainings- und Testdaten, sowie als erwartetes Ergebnis, eingebunden werden mussten. Als Eingabe könnte hierbei beispielsweise die positive und negative Stimmung vor dem Meeting und die drei Anteile an polaritären Aussagen im Meeting eingegeben werden. Das erwartete Ergebnis wäre demnach für die Feststellung einer Veränderung die positiven oder negative Stimmung nach dem Meeting. Bei dem erwarteten Ergebnis muss es sich um Integer-Zahlen handeln, da nur solche von den beiden Modellen ausgewertet werden können. Um dies gewährleisten zu können, musste bei den PANAS-Daten eine Anpassung vorgenommen werden. Wie in der Verarbeitung der Daten angegeben (Abschnitt 3.3.3), liegen die Werte zwischen 1 und 5, wobei durch die Mittelwertberechnung des Teams Zahlen mit vielen Nachkommastellen berechnet wurden. Um diese Zahlen sinnvoll auf eine ganze Zahl zu bringen wurden die Zahlen auf die vierte Nachkommastelle gerundet und mit 10.000 multipliziert. Dadurch entstanden ganze Zahlen, die trotzdem in gleichem Verhältnis zueinander stehen, da dies mit allen vier Werten pro Team gemacht wurde.

Eine weitere Aufbereitung musste bei den Datensätzen vorgenommen werden. Bei der PANAS-Befragung wurde jedes Team befragt, weshalb es Daten zu jedem Teilnehmenden der Teams gibt. Bei der Stimmungsanalyse in den Meetings wurde allerdings nicht jedes Team ausgewertet, was darauf zurückführt, dass die Qualität der Audioaufnahme, die im Anschluss transkribiert wurde, sehr schlecht war und deshalb kein Transkript entstanden ist. Aus diesem Grund sind, wie in Abschnitt 3.3 erwähnt, nur 35 auszuwertende Transkripte übrig.

Diese Feststellung führte dazu, dass die Teams, von denen es keine Transkripte gibt aus den PANAS-Daten für die Auswertung entfernt wurden. Dadurch entstand ein Datensatz mit 35 Teams, der ausgewertet werden konnte.

### 5.1.2 Untersuchung Forschungsfrage 1

Bei der Forschungsfrage 1 in Abschnitt 1.2 geht es um die Veränderung der Stimmung im Team nach dem Meeting. Hierbei soll geprüft werden,

ob ein Zusammenhang der Stimmungsveränderung besteht, wenn die Polaritäten der Aussagen besonders in eine Richtung tendieren.

Mithilfe der SVM und der GLM soll dies beurteilt werden. Hierfür wurden als erwartetes Ergebnis verschiedene Daten eingegeben, wie die erwartete Stimmung zu Zeitpunkt nach dem Meeting (t2) und die Eigenklassifikation aus Ansatz 3. Zusätzlich wurden unterschiedliche Konfigurationen vorgenommen, wie die GLM-family als Binominal oder Poisson und das Modell als Classifier (SVC) und als Regressor (SVR). Die Modelle erhielten die Daten der positiven und negativen Stimmung, sowie die Daten der Polaritäten aus dem Meeting als Eingabe. Da eine SVM nur ein erwartetes Ergebnis verarbeiten kann, wurde mit dem positiven Teil der Stimmungen angefangen, um zu analysieren, ob ein Zusammenhang der Daten vorhanden ist.

Bei jedem Ansatz wurden unterschiedliche Konfigurationen vorgenommen, welche in den Tabellen zu den Ansätzen aufgelistet sind.

### Ansatz 1

Teams Modell	35 SVC
GLM-family erwartetes Ergebnis	Binominal Stimmung t2

Tabelle 5.1: Konfigurationen der Maschinenmodelle für Ansatz 1

Dieser Ansatz war der erste Ansatz, der feststellen sollte, ob eine SVM und eine GLM sinnvoll mit dem vorhandenen Datensatz trainiert werden können. Die nötigen Konfigurationen sind in Tabelle 5.1 aufgeführt. Bei der Ausführung des Codes wurden verschiedene Werte ausgegeben, welche in den Grundlagen Abschnitt 2 genauer erläutert wurden.

Bei der SVM wurde als Modell der SVC genutzt. Die SVM hat nach dem Training einen Wert von 0,14 für die Accuracy ausgegeben. Dies lässt darauf schließen, dass die SVM mit den Daten kein Modell trainieren konnte.

Eine weitere Bestätigung hierfür war die Log-Likelihood der GLM. Dieser Wert lag bei "Minus Unendlich", sollte aber bei einem richtig trainierten Modell null sein.

Zu diesen Werten kommt hinzu, dass alle unabhängigen Variablen vom Modell mit dem Signifikanztest mit "0%" gewichtet wurden. Das bedeutet, dass das Modell alle als gleich wichtig für die Auswertung

betrachtet. Es lässt sich dadurch jedoch keine Auswertung über die Zusammenhänge schließen, da durch die Aussage der Log-Likelihood kein richtiges Modell erkannt wurde.

### Ansatz 2

Teams Model	35 SVR
GLM-family erwartetes Ergebnis	Binominal Stimmung t2

Tabelle 5.2: Konfigurationen der Maschinenmodelle für Ansatz 2

In diesem Ansatz wurde im Vergleich zum Ersten das Modell geändert. Hier würde statt des SVC eine SVR verwendet. Die Aussage der SVM verschlechterte sich mit der Accuracy auf -0.02. Das heißt, dass hier ebenfalls kein Modell gefunden werden konnte.

Mit der GLM wurde auch in diesem Modell eine Log-Likelihood von "Minus Unendlich" berechnet und auch hier waren alle Variablen gleich gewichtet. Mit dieser Kombination konnte also auch kein Erkenntnis gezogen werden.

### Ansatz 3

Teams Model	35 SVC und SVR
GLM-family erwartetes Ergebnis	Poisson Stimmung t2

Tabelle 5.3: Konfigurationen der Maschinenmodelle für Ansatz 3

Bei Ansatz 3 wurde ausschließlich die GLM-family geändert. Aus diesem Grund blieben die SVM-Ergebnisse identisch zu denen aus dem Ansatz 1 und 2. Die Log-Likelihood hat sich ein wenig verbessert, wobei die Zahl dennoch außerhalb des gewünschten Bereichs blieb. Nämlich bei -3089,1. Die Wichtigkeit der unabhängigen Variablen blieb ebenfalls gleich zu den vorherigen Ansätzen. Es konnte also auch bei diesem Ansatz kein richtiges Modell trainiert werden und somit keine Kenntnis über gültige Zusammenhänge geschlossen werden.

**Ansatz 4**

Teams	35
Model	SVC
GLM-family	Binominal
erwartetes Ergebnis	Eigenklassifikation

Tabelle 5.4: Konfigurationen der Maschinenmodelle für Ansatz 4

Bei Ansatz 4 wurde das erwartete Ergebnis im Vergleich zu den vorherigen Ansätzen verändert. Für das Ergebnis wurde eine Skala von -1, 0 und 1 zur Klassifikation erstellt. Hierfür wurden die PANAS-Werte manuell auf diese Klassifikation ausgewertet. "-1" erhielt hierbei jedes Team, welches eine negative Veränderung von mehr als -0,35 hat. Diese Zahl setzt sich wie folgt zusammen:

$$(\text{NegativeT1} - \text{NegativeT2}) + (\text{PositiveT2} - \text{PositiveT1})$$

Die Klassifikation "1" erhielten die Teams, die einen Wert von mehr als 0,35 hatten. Alle anderen erhielten die Klassifikation "0". 0,35 wurde als Grenzwert im positiven, sowie negativen Bereich festgelegt, weil diese Veränderung die Zahlenspanne in etwa drei gleich große Teile aufspaltet.

Das entstandene Array wurde als erwartetes Ergebnis der SVM gesetzt. Diese gab eine Accuracy von etwa 57% aus. Dieser Wert ist besser als das Ergebnis in den meisten anderen Ansätzen zuvor und lässt darauf schließen, dass ein Modell erkannt werden konnte.

Das Ergebnis der GLM sagt hingegen das Gegenteil aus. Die Log-Likelihood liegt bei "Minus Unendlich", also konnte kein Modell trainiert werden. Die Signifikanz liegt bei allen unabhängigen Variablen bei mehr als 5%, das bedeutet, dass laut dem Modell keine der Variablen für die Auswertung wichtig ist. Auch hier konnte demnach kein Zusammenhang bzw. eine Auswertung gemacht werden.

Ein zusätzlicher Beweis hierfür ist die entstandene Confusion-Matrix. In Abbildung 5.1 erkennt man deutlich, dass das Modell immer neutral als Ergebnis setzt, was die Accuracy von 57% erklärt, da 57,14% als richtig neutral klassifiziert worden sind.

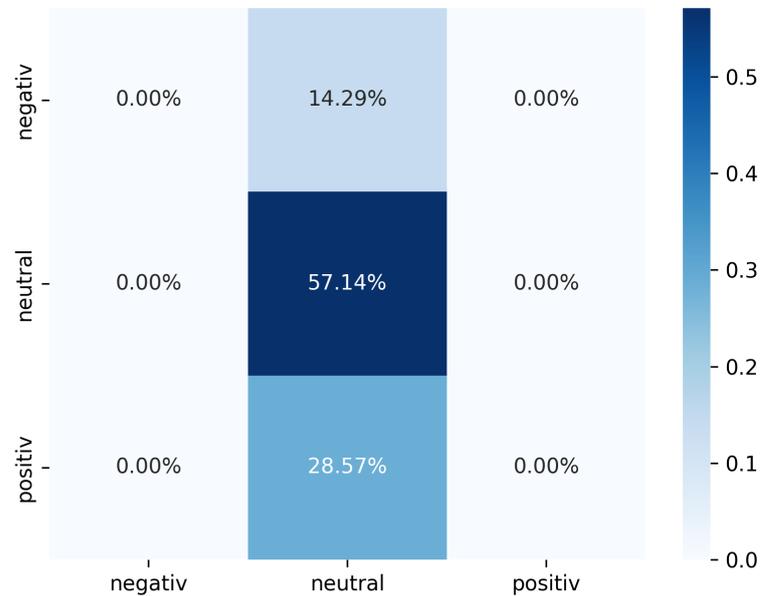


Abbildung 5.1: Confusion Matrix zum Ansatz 4

### Zusammenfassung der Ansätze

Zusammenfassend ist zu sagen, dass mithilfe einer SVM kein Modell für die erste Forschungsfrage trainiert werden kann. Somit kann keine Auswertung auf einen Zusammenhang der unterschiedlichen Datensätze gegeben werden.

## 5.2 Statistische Untersuchung der Forschungsfragen

In diesem Kapitel werden die Forschungsfragen manuell untersucht und ausgewertet. Hierfür wird auf jede Forschungsfrage im einzelnen eingegangen. Die Forschungsfragen wurden mithilfe von statistischen Tests untersucht, um den Zusammenhang der einzelnen Komponenten herauszufiltern.

### 5.2.1 Auswahl der Statistischen Tests

Statistische Tests haben viele verschiedene Anwendungsgebiete und Ausprägungen in der Untersuchung. Für die Untersuchung dieser Masterarbeit sind vor allem Tests auf Korrelation wichtig, um einen Zusammenhang zwischen den Daten zu beweisen. Mithilfe dieser Tests soll die Korrelation zwischen den Stimmungen im Team und den Polaritäten der Aussagen bestimmt werden. Bei der Wahl der unterschiedlichen Tests sind die Datensätze, die eingegeben werden sollen, von Bedeutung. Je nach Skalentyp der Daten können unterschiedliche statistische Tests verwendet werden. Bei dem Test auf Korrelation gibt es beispielsweise für die unterschiedlichen Skalentypen unterschiedliche Tests. Bei einer Nominalskala kann der Phi-Coefficient verwendet werden. Bei metrischen Skalen der Pearson Correlation Coefficient. [33] Die Daten aus dieser Masterarbeit sind Daten einer Ordinalskala, da sie Häufigkeiten darstellen und dabei die Rangordnung bzw. Reihenfolge beibehalten werden muss. Dementsprechend müssen Tests für diesen Skalentyp ausgewählt werden. Diese sind beispielsweise der Pearson Correlation Coefficient Test oder der Spearman's Rho Test. Welche Tests angewendet werden sollte, hängt zusätzlich davon ab, ob die Werte normalverteilt sind. Hierfür gibt es Tests, welche die Verteilung angeben.

#### Test auf Normalverteilung

Es gibt unterschiedliche Tests, die weitere unterschiedliche Bedingungen an die einzugebenden Daten stellen. Hierbei kann eine Bedingung die Normalverteilung sein. Um festzustellen, ob die Daten Normalverteilt sind, gibt es den sogenannten Test nach Kolmogorov-Smirnov<sup>1</sup>. Mithilfe dieses Tests wurde herausgefunden, dass in den Datensätzen dieser Masterarbeit nur die positiven Stimmungen zum Zeitpunkt vor dem Meeting normalverteilt sind. Alle anderen Datensatzteile sind nicht normalverteilt. Dementsprechend konnten nur Tests genutzt werden, die ohne eine Normalverteilung auskommen.

#### Test auf Korrelation

Tests auf Korrelation werden genutzt, um den Zusammenhang einer Veränderung von einem Zeitpunkt zu einem anderen, in Abhängigkeit von eventuellen Einflüssen, zu bestimmen. Beim Aussuchen des richtigen Tests ist es wichtig, ob die vorhandenen Werte normalverteilt

---

<sup>1</sup><https://www.socscistatistics.com/tests/kolmogorov/default.aspx>

sind. Bei vorhandener Normalverteilung kann bei ordinalskalierten Daten der Pearson Correlation Coefficient berechnet werden. Wenn es keine Normalverteilung der Daten gibt, kann der Spearman's Rho Test angewendet werden. Da die Daten nicht normalverteilt sind, wurde in dieser Masterarbeit der Spearman's Rho Test<sup>2</sup> verwendet. Die Formel für diesen Test sieht wie folgt aus:

$$r_s = 1 - \frac{6 \sum D^2}{N^3 - N} \quad (5.1)$$

$r_s$  = Korrelationskoeffizient

D = Differenz zwischen den Rängen der eingegebenen Werte

N = Anzahl an Werten

$r_s$  kann einen Wert zwischen +1 und -1 annehmen, wobei bei 0 kein Zusammenhang, bei +1 einen starken positiven Zusammenhang und bei -1 einen starken negativen Zusammenhang besteht. Die genaue Skala, welche in dieser Arbeit verwendet wird, sieht wie folgt aus:

r =1	perfekter Zusammenhang
0,7 < r < 0,99	sehr starker Zusammenhang
0,5 < r < 0,69	starker Zusammenhang
0,3 < r < 0,59	mittelstarker Zusammenhang
0,2 < r < 0,29	schwacher Zusammenhang
r <=0,19	kein Zusammenhang

Tabelle 5.5: Skala zur Analyse eines Zusammenhangs<sup>3</sup>

### P-Wert Berechnung

Bei der Berechnung des Korrelationskoeffizienten wird zusätzlich ein sogenannter p-Wert berechnet, welcher die Signifikanz angibt. Mithilfe der Signifikanz können aufgestellte Hypothesen überprüft werden. "Vor der Datenerhebung wird eine maximale Irrtumswahrscheinlichkeit festgelegt (Signifikanzniveau  $\alpha$ ), die den Fehler 1. Art, nämlich die Nullhypothese abzulehnen, obwohl sie richtig ist, begrenzt. Häufig gewählte Niveaus sind  $\alpha = 0,05$  und  $\alpha = 0,01$ . Ist der p-Wert kleiner als das festgelegte Signifikanzniveau, so liegt statistische Signifikanz zum

<sup>2</sup><https://www.socscistatistics.com/tests/spearman/default.aspx>

<sup>3</sup><https://www.2ask.de/Interpretation-des-Korrelationskoeffizienten-720d.html>

Niveau  $\alpha$  vor." [4]

Liegt der  $p$ -Wert also unter dem Signifikanzniveau, kann die Nullhypothese abgelehnt werden. Liegt er über dem Signifikanzniveau kann der Alternativhypothese nicht zugestimmt werden bzw. es kann keine Aussage auf einen Zusammenhang getroffen werden.

## 5.2.2 Untersuchung mit den Statistischen Tests

In diesem Abschnitt werden die einzelnen Forschungsfragen mit den statistischen Tests untersucht. Die Durchführung der Tests wurde auf der Website Social Science Statistic<sup>4</sup> durchgeführt. Dies ist eine Oberfläche, um unterschiedliche statistische Tests durchführen zu können, ohne andere Frameworks einbinden zu müssen..

### Untersuchung der Forschungsfrage 1

Wie im vorherige Abschnitt beschrieben, wird bei der Forschungsfrage 1 der Zusammenhang zwischen der Stimmung vor dem Meeting in Kombination mit den Polaritäten im Meeting und der Stimmung nach dem Meeting untersucht. In diesem Abschnitt wird statistisch geprüft, ob ein Zusammenhang besteht.

Für diese Forschungsfrage wurden eine Hypothese aufgestellt, auf die der Test auf Korrelation namens Spearman's Rho 5.1 angewendet wurde.

H1 Es besteht ein Zusammenhang zwischen den positiven Aussagen im Meeting und einer veränderten Stimmung im Team.

H0 Es besteht kein Zusammenhang zwischen den positiven Aussagen im Meeting und einer veränderten Stimmung im Team.

Für die Eingabe des Tests wurden als X-Wert die Anteile an positiven Aussagen im Meeting in % angegeben und als Y-Wert der  $\delta$ -Wert der PANAS-Stimmungen. Dieser lautet: *(PositiveT2-NegativeT2)-(PositiveT1-NegativeT1)*. Mit diesem Wert kann die Veränderung vom ersten zum zweiten Zeitpunkt dargestellt werden. Der  $r_s$ -Wert liegt bei 0,16557, der  $p$ -Wert liegt bei 0.34185. Der  $P$ -Wert sagt aus, dass der Hypothese H1 nicht zugestimmt werden kann, weil der Test nicht signifikant ist. Demnach kann H0 nicht widerlegt werden. Der  $r_s$ -Wert ist demnach nicht relevant.

Dieselbe Hypothese wurde für den Anteil an negativen Aussagen im

---

<sup>4</sup><https://www.socscistatistics.com/tests/>

Meeting aufgestellt. Hierbei wurde ebenfalls der Spearman's Rho Test angewendet.

H1 Es besteht ein Zusammenhang zwischen den negativen Aussagen im Meeting und einer veränderten Stimmung im Team.

H0 Es besteht kein Zusammenhang zwischen den negativen Aussagen im Meeting und einer veränderten Stimmung im Team.

Als Eingabe wurde für den X-Wert der Anteil negativer Aussagen im Meeting in % genutzt und für Y ebenfalls das  $\delta$ , wie in der vorherigen Hypothese. Bei der Durchführung ergab  $p = 0.14081$ . Das heißt, dass die Aussage nicht signifikant ist und demnach H1 nicht zugestimmt werden kann. Der  $r_s$ -Wert von  $-0.25407$  ist nicht relevant.

### Untersuchung der Forschungsfrage 2

Die Forschungsfrage 2 handelt von der Polarität der Aussagen im Meeting, abhängig von der Stimmung, die zuvor im Team ausgewertet wurde. Der Zusammenhang der Stimmung vor dem Meeting und den Aussagen im Meeting wird mit den folgenden Hypothesen überprüft. Für die Untersuchung wird der Spearman's Rho Test angewendet.

H1 Es gibt einen Zusammenhang zwischen der Stimmung im Team vor dem Meeting und dem Anteil positiver Aussagen im Meeting.

H0 Es gibt keinen Zusammenhang zwischen der Stimmung im Team vor dem Meeting und dem Anteil positiver Aussagen im Meeting.

Die Eingabe für diese Überprüfung der ersten Hypothese waren zum einen die negative Stimmung subtrahiert von der positiven Stimmungen zum Zeitpunkt vor dem Meeting und zum anderen der Anteil positiver Aussagen in %. Die Formel für den ersten Wert sieht wie folgt aus: *PositiveT1* - *NegativeT1*. Der P-Wert, der sich bei der Untersuchung mit dem Spearman's Rho Test ergeben hat liegt bei  $0.00958$ . Das bedeutet, dass die Hypothese H1 signifikant ist und somit die Hypothese H0 widerlegt werden konnte. Der  $r_s$ -Wert liegt bei  $0.43187$ . Dies sagt aus, dass der Zusammenhang zwischen den beiden eingesetzten Variablen mittelstark ist, sie aber dennoch einen positiven Zusammenhang haben.

## 5.2. STATISTISCHE UNTERSUCHUNG DER FORSCHUNGSFRAGEN<sup>45</sup>

Die zweite Hypothese lautet wie folgt:

H1 Es gibt einen Zusammenhang zwischen der Stimmung im Team vor dem Meeting und dem Anteil negativer Aussagen im Meeting.

H0 Es gibt einen Zusammenhang zwischen der Stimmung im Team vor dem Meeting und dem Anteil negativer Aussagen im Meeting.

Für die Untersuchung wurde für diese Hypothese der Gegensatz in der Polarität der Aussagen zur vorherigen genutzt. Hierbei wurde für den X-Wert ebenfalls *PositiveT1* - *NegativeT1* genutzt. Für den Y-Wert wurde der negative Anteil an Aussagen im Meeting in % genutzt. Bei dieser Analyse ergab sich ein p-Wert von 0.94181, was bedeutet, dass die Aussage von H1 nicht signifikant ist und demnach H0 nicht widerlegt werden kann. Der  $r_s$ -Wert von -0.0128 ist nicht relevant.

### Untersuchung der Forschungsfrage 3

Die Forschungsfrage 3 befasst sich mit den psychologischen Abfragen, die zusätzlich zur PANAS-Befragung nach dem Meeting durchgeführt wurde. Hierbei stellt sich die Frage, ob die Stimmung im Meeting und die Stimmung in den psychologischen Abfragen einen Zusammenhang haben. War die Stimmung im Meeting beispielsweise gut, hat dies einen Zusammenhang mit der Mitarbeiterzufriedenheit. Diese Fragen sollen mithilfe dem Spearman's Rho Test beantwortet werden.

H1 Es gibt einen Zusammenhang zwischen den positiven Aussagen im Meeting und den aufgabenbezogenen Konflikten.

H0 Es gibt keinen Zusammenhang zwischen den positiven Aussagen im Meeting und den aufgabenbezogenen Konflikten.

Für die Beantwortung der ersten Hypothese, die einen Teil der Forschungsfrage 3 beantworten soll, wurden als Variablen in den Test der Anteil positiver Aussagen im Meeting und die Stimmung der aufgabenbezogenen Konflikte nach dem Meeting verwendet. Hierbei ergab sich ein p-Wert von 0.02752. Demnach kann der Aussage H1 zugestimmt werden. Der Zusammenhang liegt hierbei bei -0.37255 und liegt somit im unteren Bereich des mittelstarken Zusammenhangs. Die folgende zweite aufgestellte Hypothese bezieht sich ebenfalls auf die aufgabenbezogenen Konflikte. Hierbei wurde im Vergleich zur ersten Hypothese bei der Eingabe nur die Werte für X geändert, diese wurden durch die negativen Anteile der Aussagen im Meeting ausgetauscht.

H1 Es gibt einen Zusammenhang zwischen den negativen Aussagen im Meeting und den aufgabenbezogenen Konflikten.

H0 Es gibt keinen Zusammenhang zwischen den negativen Aussagen im Meeting und den aufgabenbezogenen Konflikten.

Für diese Hypothese ergab sich ein p-Wert von 0.60015. Die Hypothese H1 ist nicht signifikant und ihr kann nicht zugestimmt werden. Der  $r_s$ -Wert von 0.09175 ist demnach nicht relevant. Die nächsten zwei aufgestellten Hypothesen beziehen sich auf die sozialen Konflikte. Demnach ist bei beiden Berechnungen mit dem Spearman's Rho Test der Y-Wert die Stimmung über die sozialen Konflikte nach dem Meeting. Die folgende Hypothese enthält als X-Wert den Anteil an positiver Aussagen im Meeting.

H1 Es gibt einen Zusammenhang zwischen den positiven Aussagen im Meeting und den sozialen Konflikten.

H0 Es gibt keinen Zusammenhang zwischen den positiven Aussagen im Meeting und den sozialen Konflikten.

Das Ergebnis des p-Wertes liegt bei 0.46118, dementsprechend ist die Aussage nicht signifikant. Der  $r_s$ -Wert von -0.12872 ist nicht relevant.

Für die zweite Hypothese im Bezug auf die sozialen Konflikte, wurde als X-Wert der Anteil negativer Aussagen genutzt.

H1 Es gibt einen Zusammenhang zwischen den negativen Aussagen im Meeting und den sozialen Konflikten.

H0 Es gibt keinen Zusammenhang zwischen den negativen Aussagen im Meeting und den sozialen Konflikten.

Diese Hypothese ergab nach der Anwendung dem Spearman's Rho Test ein p-Wert von 0.49156. Diese Aussage ist demnach ebenfalls nicht signifikant und auch hier ist der  $r_s$ -Wert von 0.12021 nicht relevant.

Die letzten zwei Hypothesen beziehen sich auf die Mitarbeiterzufriedenheit als Y-Eingabe. In der ersten der zwei Hypothesen wird wieder der Anteil positiver Aussagen als X-Wert für der Spearman's Rho Test eingesetzt.

H1 Es gibt einen Zusammenhang zwischen den positiven Aussagen im Meeting und der Mitarbeiterzufriedenheit.

H0 Es gibt keinen Zusammenhang zwischen den positiven Aussagen im Meeting und der Mitarbeiterzufriedenheit.

Als Ergebnis ergab sich für den p-Wert 0.00733. Die Aussage ist also statistisch signifikant. Der Zusammenhang ist mittelstark, da der  $r_s$ -Wert bei 0.44543 liegt.

Die zweite Hypothese der Mitarbeiterzufriedenheit hat als X-Wert den Anteil negativer Aussagen im Meeting.

H1 Es gibt einen Zusammenhang zwischen den negativen Aussagen im Meeting und der Mitarbeiterzufriedenheit.

H0 Es gibt keinen Zusammenhang zwischen den negativen Aussagen im Meeting und der Mitarbeiterzufriedenheit.

Hier ergab sich für den p-Wert ein Wert von 0.71768, was bedeutet, dass die Hypothese H1 nicht signifikant ist. Auch der  $r_s$ -Wert ist sehr gering, wobei er durch die fehlende Signifikanz keine Relevanz hat.

## 5.3 Ergebnisse

Nachdem im vorherigen Kapitel die Untersuchung einzelner Hypothesen zu den Forschungsfragen durchgeführt wurde, wird in diesem Kapitel die zuvor entstandenen Ergebnisse erfasst und ausgewertet. Hierfür wird auf die einzelnen Forschungsfragen eingegangen. Zum Schluss wird die zentrale Forschungsfrage beantwortet.

### Auswertung der Forschungsfrage 1

Bei Forschungsfrage 1 wurden zwei unterschiedliche Methoden angewendet, um diese zu beantworten. Die erste Methode war die Untersuchung mit einer SVM und einer GLM. Bei dieser Methode konnte kein Zusammenhang festgestellt werden. Alle vier durchgeführten Ansätze konnten kein Modell bilden.

Die zweite Methode war die Untersuchung mithilfe von Hypothesen. Diese wurde ebenfalls für alle anderen Forschungsfragen genutzt. Zusammenfassend sind folgende Ergebnisse herausgekommen:

X-Wert	Y-Wert	Signifikant	Zusammenhang
Anteile positiver Aussagen im Meeting in %	(PositiveT2 - NegativeT2) (PositiveT1 - NegativeT1)	nein	nicht relevant
Anteile negativer Aussagen im Meeting in %	(PositiveT2 - NegativeT2) (PositiveT1 - NegativeT1)	nein	nicht relevant

Tabelle 5.6: Zusammenfassung der statistischen Untersuchung der Forschungsfrage 1

Wie der Tabelle 5.6 zu entnehmen ist, konnte bei beiden Hypothesen der jeweiligen Alternativhypothese H1 nicht zugestimmt werden. Demnach konnte H0 nicht widerlegt werden und es muss weiterhin angenommen werden, dass es keinen Zusammenhang zwischen den positiven und negativen Aussagen aus dem Meeting und der Veränderung der Stimmungen von vor dem Meeting zu nach dem Meeting gibt.

### Auswertung der Forschungsfrage 2

Die Forschungsfrage 2 wurde mithilfe von zwei Hypothesen beantwortet. Die Ergebnisse wurden in folgender Tabelle zusammengefasst:

X-Wert	Y-Wert	Signifikant	Zusammenhang
(PositiveT1 - NegativeT1)	Anteile positiver Aussagen im Meeting in %	ja	0,43187
(PositiveT1 - NegativeT1)	Anteile negativer Aussagen im Meeting in %	nein	nicht relevant

Tabelle 5.7: Zusammenfassung der statistischen Untersuchung der Forschungsfrage 2

Den Hypothesen zufolge besteht zwischen den positiven Aussagen im Meeting und der Stimmung vor dem Meeting ein Zusammenhang. Dieser liegt bei 0,43187. Demnach existiert ein mittelstarker Zusammenhang. Über einen Zusammenhang mit den negativen Aussagen im

Meeting kann keine Aussage getroffen werden. Die Forschungsfrage 2 konnte mit den vorhandenen Daten nur teilweise beantwortet werden.

### Auswertung der Forschungsfrage 3

Bei der Forschungsfrage 3 wurden sechs Hypothesen zur Überprüfung aufgestellt. Jeweils zwei Hypothesen testen dabei die drei unterschiedlichen psychologischen Abfragen. Diese zwei Hypothesen werden jeweils in die positiven und negativen Anteile der Aussagen im Meeting aufgeteilt. Die Ergebnisse sehen wie folgt aus:

X-Wert	Y-Wert	Signifikant	Zusammenhang
Anteile positiver Aussagen im Meeting in %	AK	ja	-0,37255
Anteile negativer Aussagen im Meeting in %	AK	nein	nicht relevant
Anteile positiver Aussagen im Meeting in %	SK	nein	nicht relevant
Anteile negativer Aussagen im Meeting in %	SK	nein	nicht relevant
Anteile positiver Aussagen im Meeting in %	MZ	ja	0,44543
Anteile negativer Aussagen im Meeting in %	MZ	nein	nicht relevant

Tabelle 5.8: Zusammenfassung der statistischen Untersuchung der Forschungsfrage 3

Der Tabelle 5.8 ist zu entnehmen, dass nur bei zwei der aufgestellten Hypothesen eine Zustimmung erfolgen konnte und die Nullhypothese widerlegt werden konnte. Dabei handelt es sich um den Zusammenhang zwischen dem Anteil positiver Aussagen im Meeting und den aufgabenbezogenen Konflikten beziehungsweise der Mitarbeiterzufriedenheit.

Der Zusammenhang mit den aufgabenbezogenen Konflikten liegt bei  $-0,37255$  und ist somit am unteren Rand der mittelstarken Zusammenhangs einzuordnen. Der Zusammenhang mit der Mitarbeiterzufriedenheit liegt bei  $0,44543$ , dementsprechend ist dieser mittelstark positiv, aber etwas stärker, als bei den aufgabenbezogenen Konflikten. Die Forschungsfrage kann demnach nur in zwei von sechs Fällen sicher beantwortet werden, für die anderen vier Fälle kann keine Aussage getroffen werden.

# Kapitel 6

## Diskussion

In dieser Arbeit geht es um die Untersuchung auf einen Zusammenhang zwischen der Stimmung vor und nach dem Meeting, sowie den Polaritäten der Aussagen im Meeting. In diesem Kapitel werden die Forschungsfragen, die in der Kapitel 4 und 5 untersucht wurden, beantwortet und diskutiert. Hierfür wird unter anderem auf die Einflussfaktoren der Analyse eingegangen. Anschließend werden Grenzen der Arbeit, sowie verwandte Arbeiten aufgeführt.

### 6.1 Beantwortung der Forschungsfragen

In den vorangehenden Kapiteln wurden unterschiedliche Untersuchungen gemacht, um die aufgestellten Forschungsfragen zu beantworten. Hierbei wurde eine qualitative Untersuchung der Daten, sowie eine quantitative Untersuchung mit statistischen Tests durchgeführt. Für die Untersuchung wurden Teams als Repräsentanten ausgewählt. Diese wurden ausgewählt, da sie jeweils die Kombination an absteigend, aufsteigend oder gleichbleibender Stimmung von den Zeitpunkten vor und nach dem Meeting bei den positiven und negativen PANAS-Abfragen hatten. Insgesamt müssten es neun Repräsentanten sein, allerdings gibt es zu zwei Kombinationen keinen Repräsentanten, weshalb sieben Repräsentanten aufgeführt wurden. In diesem Abschnitt sollen die zwei Untersuchungen nun zusammengeführt und auf ein Ergebnis gebracht werden, damit die Forschungsfragen beantwortet werden können.

Beginnend mit Forschungsfrage 1. Diese zielt auf die Veränderung der Stimmung bei positiven oder negativen Aussagen im Meeting ab. Werden hierzu die Ergebnisse der qualitativen Untersuchung

betrachtet, ist zu erkennen, dass die größte Veränderung in der Stimmung bei Team 1 und 37 auftritt. (siehe Abbildung 3.1 und A.5). Die Veränderung liegt bei 0,9. Team 1 hat dabei keine negativen Aussagen im Meeting und der positive Anteil liegt bei 6,6%, also deutlich unter dem Median der gesamten positiven Aussagen von etwa 10% (vergleiche Abbildung 4.9 und 4.7). Team 37 hat stattdessen den höchsten Anteil an negativen Aussagen von den Repräsentanten mit 1,47% und einen etwa im Schnitt liegenden positiven Anteil von 10,85% (vergleiche Abbildung A.20). Team 20 hingegen hat keine Veränderung in der Stimmung, der Anteil positiver Aussagen liegt mit 11,36% über dem Schnitt und es gibt keinen Anteil negativer Aussagen (vergleiche Abbildung 4.8 und 3.2). Diese Ergebnisse haben von dieser Betrachtungsweise her keinen Zusammenhang zwischen der Veränderung der Stimmung und den positiven oder negativen Aussagen im Meeting.

Diese Auffälligkeiten werden durch die quantitative Untersuchung bestätigt. Diese ergab, dass die Hypothesen über den Zusammenhang der positiven oder negativen Aussagen und der Veränderung der Stimmung nicht bestätigt werden konnten und somit die Forschungsfrage wie folgt beantwortet werden kann: Es gibt keine Signifikanz für eine auftretende Veränderung der Stimmung im Team bei positiven oder negativen Aussagen.

Bei Forschungsfrage 2 geht es um die Stimmung im Team und die dadurch ausgelösten Anteile positiver oder negativer Aussagen. Werden hierzu erneut die Repräsentanten betrachtet, so ist zu erkennen, dass Team 20 und 37 die beste Stimmung im Team vor dem Meeting haben, Wird die Diferrenz der negativen und positiven Stimmungen berechnet, so liegt die Stimmung bei Team 20 bei +2,4 und die Stimmung bei Team 37 bei +2,35. Demnach ist die Stimmung bei Team 20 minimal höher als bei Team 37. Team 1 hat die schlechteste Stimmung mit einer Differenz von +0,9 (vergleiche Abbildung 3.1, 3.2 und A.5). Wird hierzu die Auswertung der Aussagen im Meeting verglichen, fällt auf, dass Team 1 und 20 keinen negativen Anteil an Aussagen haben, Team 37 hat einen Anteil von 1,47%. Die positiven Anteile von Team 20 und 37 liegen über dem Median der positiven Anteile aller Teams, wobei der Anteil bei Team 20 mit 11,36% etwas größer ist als bei Team 37 mit 10,85%. Hieran ist zu erkennen, dass sich am positiven Anteil ein Zusammenhang erkennen lässt. Je positiver die Stimmung vor dem Meeting, umso höher ist der Anteil positiver Aussagen.

Diese Entwicklung wird ebenfalls durch die quantitative Untersuchung bestätigt. Die Hypothese über den Zusammenhang zwischen der Stimmung vor dem Meeting und dem Anteil positiver Aussagen ließ sich bestätigen. Das bedeutet, dass dieser Zusammenhang besteht. Die Hypothese über den Zusammenhang zwischen der Stimmung vor dem Meeting und dem Anteil negativer Aussagen hingegen ließ sich nicht bestätigen. Daraus lässt sich für die Beantwortung der Forschungsfrage schließen, dass die Stimmung vor dem Meeting und der Anteil positiver Aussagen einen Zusammenhang haben, der Zusammenhang zwischen der Stimmung vor dem Meeting und dem Anteil negativer Aussagen sich jedoch anhand der Daten nicht bestätigen ließ.

Die Forschungsfrage 3 handelt von dem Zusammenhang der positiven oder negativen Aussagen im Meeting und den psychologischen Abfragen nach dem Meeting. Zur Beantwortung dieser Forschungsfrage müssen die drei psychologischen Abfragen getrennt voneinander betrachtet werden. Beginnend mit den aufgabenbezogenen Konflikten ist zu erkennen, dass diese bei Team 37 und bei Team 20 mit einem Wert von 4 am höchsten sind. In diesen Teams gibt es also am meisten aufgabenbezogene Konflikte. In Team 5 gibt es hingegen am wenigsten Konflikte mit einem Wert von 2 (vergleiche Abbildung 3.2, A.1 und A.5). Wird hierbei auf den Zusammenhang mit den Anteilen positiver oder negativer Aussagen eingegangen, ist zu erkennen, dass Team 37 und 20 einen geringeren Anteil an positiven Aussagen haben als Team 5 (vergleiche Abbildung A.16, 4.8 und A.20). Weniger positive Anteile an Aussagen bedeutet also höhere aufgabenbezogene Konflikte. Dennoch liegen alle drei über dem Median der positiven Anteile aller Teams von etwa 10%. Der negative Anteil der Aussagen weist bei diesen Repräsentanten nicht auf einen Zusammenhang, da Team 20 keinen Anteil negativer Aussagen besitzt, Team 5 und 37 hingegen schon. Dies widerspricht den aufgabenbezogenen Konflikten, da für einen Zusammenhang Team 20 und 37 einen ähnlichen negativen Anteil haben müssten und Team 5 einen abweichenden.

Ähnlich sieht dies bei der statistischen Untersuchung aus. Hier ließ sich die Aussage, dass ein Zusammenhang zwischen den aufgabenbezogenen Konflikten und dem Anteil positiver Aussagen besteht, bestätigen. Über den Zusammenhang mit dem negativen Anteil konnte keine Aussage getroffen werden.

Wird die Mitarbeiterzufriedenheit betrachtet, hat Team 37 den niedrigsten Wert mit 3,38. Team 21 und 34 haben hingegen den höchsten Wert mit 5,25 (vergleiche Abbildung A.3, A.4 und A.5). Den

negativen Anteilen der Aussagen zur Folge, lässt sich bei diesen drei Repräsentanten ein Muster erkennen, Team 21 und 34 haben keine negativen Aussagen in ihrer Analyse, Team 37 hingegen schon. Bei den positiven Anteilen liegen die Anteile von Team 34 und 37 im selben Bereich bei etwa 11%. Team 21 hat hierbei den geringsten positiven Anteil an Aussagen mit 5,76%. Dies lässt auf keinen Zusammenhang schließen.

Diese Vermutung widerlegt allerdings die statistische Untersuchung. Hierbei wird zwischen dem Anteil positiver Aussagen und der Mitarbeiterzufriedenheit ein Zusammenhang festgestellt. Zwischen dem Anteil negativer Aussagen, konnte jedoch kein Zusammenhang festgestellt werden.

Zuletzt wird auf die sozialen Konflikte eingegangen. Hierbei sind die Werte bei den Repräsentanten Team 1 und 16 am niedrigsten mit einem Wert von 1. Bei Team 37 und 20 sind sie am höchsten mit einem Wert von 3,25 (vergleiche Abbildung 3.1, A.2, 3.2 und A.5). Wird dies wieder in einen Zusammenhang mit dem positiven und negativen Anteilen der Aussagen gebracht, so lässt sich zeigen, dass Team 16 und 1 einen kleineren Anteil an positiven Aussagen haben, als der Median der gesamten Team von etwa 10%. Team 20 und 37 hingegen haben einen größeren Anteil, sodass hierbei ein Zusammenhang erkennbar ist. Dieser ist jedoch nicht der erwartete, denn je höher der Anteil positiver Aussagen, desto mehr soziale Konflikte sind aufgetreten. Beim negativen Anteil lässt sich kein Zusammenhang erkennen. Team 1 und 20 haben keinen negativen Anteil, Team 16 und 37 haben einen Anteil von etwa 1%.

Das Ergebnis der statistischen Untersuchung widerspricht in der Untersuchung auf einen Zusammenhang mit dem positiven Anteil den aufgeführten Beobachtungen. Hier ließ sich keine Signifikanz feststellen, weshalb die Aussage nicht bestätigt werden konnte. Beim negativen Anteil entspricht es der zuvor genannten Beobachtung, hier konnte ebenfalls kein Zusammenhang festgestellt werden.

Bei der Beantwortung von Forschungsfrage 3 mit den statistischen Tests wurden mehr Hypothesen benötigt, als sie in den vorangegangenen Forschungsfragen aufgestellt wurden. Dies liegt daran, dass alle drei psychologischen Abfragen einzeln überprüft werden müssen, um Aufschluss über den Zusammenhang geben zu können. Zwei der insgesamt sechs Hypothesen konnten wie bereits erwähnt mit der statistischen Untersuchung bestätigt werden. Die anderen vier konnten mit den vorhandenen Daten nicht bewiesen werden.

Die Forschungsfrage 3 kann demnach wie folgt beantwortet werden:

Es besteht ein Zusammenhang zwischen den positiven Aussagen im Meeting und den aufgabenbezogenen Konflikten sowie der Mitarbeiterzufriedenheit. Über die restlichen psychologischen Abfragen auch in Kombination mit den negativen Aussagen kann keine Aussage getroffen werden.

Die zentrale Forschungsfrage soll nun mithilfe der beantworteten Forschungsfragen 1 - 3 beantwortet werden. Hierbei fällt auf, dass keine Aussage zu den negativen Anteilen des Meeting im Zusammenhang mit jeglicher Stimmung vor oder nach dem Meeting getroffen werden kann. Es kann zudem nicht zu jedem Anteil positiver Aussagen im Meeting im Zusammenhang mit den Stimmungen davor oder danach getroffen werden. Die Stimmung vor dem Meeting hat einen Zusammenhang mit den positiven Aussagen im Meeting und auch die aufgabenbezogenen Konflikte, sowie die Mitarbeiterzufriedenheit nach dem Meeting, haben einen Zusammenhang mit den positiven Aussagen im Meeting. Demnach besteht nur in diesen drei Fällen ein Zusammenhang zwischen der Stimmung im Team und den polaritären Aussagen im Meeting. Zu den anderen Fällen kann mit den vorhandenen Daten keine Aussage getroffen werden.

## 6.2 Validität

Liest man die Diskussion der Ergebnisse, stellt sich die Frage, ob die Aussagen in dieser Form getroffen werden können. Zuerst ist wichtig zu wissen, dass die Hypothesen, die nicht angenommen werden konnten, nicht gleichzeitig widerlegt wurden. Die aufgestellten Hypothesen konnten nicht bewiesen werden, sodass die Nullhypothese ( $H_0$ ) nicht widerlegt werden konnte und somit weiterhin gültig ist. Es kann aber trotzdem sein, dass die Alternativhypothesen bewiesen werden könnten. Dies könnte beispielsweise an fehlerhaften oder zu wenigen Daten liegen, um eindeutige Schlüsse ziehen zu können. Diesen Eindruck vermittelt beispielsweise die SVM, die zur Untersuchung der Forschungsfrage 1 genutzt werden sollte. Hierbei konnte kein ausreichend gut trainiertes Modell gefunden werden, was vermutlich an fehlenden Daten liegt. Die Analyse der Daten enthalten nur wenige Fehler. Um dies behaupten zu können, wurden bei der Analyse der Aussagen zwei unterschiedliche Analysetools verwendet und anschließend nur die Daten genutzt, die eine Übereinstimmung haben. Durch dieses Vorgehen wurde die Richtigkeit der Klassifikation der Polaritäten der

Aussagen im Meeting unterstützt.

### 6.3 Grenzen der Arbeit

In dieser Arbeit konnten einige Untersuchungen nicht durchgeführt werden. Zum einen wurde die Bachelorarbeit von David Schiller [23] genutzt, um die zu verwendenden Analysetools auszuwählen, wobei eines der Tools vorgegeben war. Hierzu könnte für eine Verbesserung der Accuracy eine Analyse durchgeführt werden, um herauszufinden, welche Kombination aus zwei Analysetools die beste Accuracy ergibt. Mit einem erweiterten Datensatz könnten die in dieser Arbeit durchgeführten Untersuchungen erneut durchgeführt und Veränderungen festgestellt werden. Zudem könnte dadurch die SVM ein Modell trainieren, wenn das Problem an einem zu kleinen Datensatz lag.

Mit den vorhandenen Daten kann nicht sichergestellt werden, dass diese auch in anderen Projekten so zu analysieren ist. Die Teams, die an dem Softwareprojekt an der LUH teilgenommen haben, sind Studierende, welche mit großer Wahrscheinlichkeit noch nie in anderen Projekten gearbeitet haben. Der Erfahrungsgrad solcher Projekte war demnach sehr klein. Zudem kannten sich die Teammitglieder teilweise noch nicht und es war das erste Meeting, welches mit dem Kunden stattfand. Im Laufe des Projekts können sich die Daten verändern, weil der Umgang mit bekannten Kollegen anders sein kann, als mit unbekanntem.

### 6.4 Verwandte Arbeiten

Innerhalb einer Masterarbeit ist es oftmals interessant zu hören, wie das Thema entstanden ist. Dabei spielen sowohl die Grundlagen dieser Arbeit, als auch andere Ansätze und Ideen eine wichtige Rolle.

An der LUH im Fachgebiet Software Engineering werden seit einiger Zeit unterschiedliche Themen bearbeitet. Aus diesem Themen ist das Thema dieser Masterarbeit entstanden. Zu den wichtigsten Themen zählt die Masterarbeit von Julian Horstmann [16]. Diese setzte den Grundstein für den entstandenen SEnti-Analyzer. In der Bachelorarbeit von Marc Herrmann [14] wurde dieser überarbeitet und weiterentwickelt, sodass eine Analyse von Audiodateien möglich ist. In einer anderen Bachelorarbeit am Fachgebiet Software Engineering von Christian Meyer [20] wurde eine Verbesserung des Tools von Julian Horstmann mithilfe eines anderen Stimmungsanalysetools namens Senti4SD durchgeführt. Dieses Tool wurde in einem Paper von Calefato

et al [8] vorgestellt. Es soll eine Verbesserung der Stimmungsanalyse im Bezug auf den Bereich Software Engineering sein.

Die Idee für die Untersuchung der Softwareprojekte mithilfe von PANAS-Befragungen, anderen psychologischen Befragungen und der Meetinganalyse entstand am Fachgebiet Software Engineering an der LUH in Zusammenarbeit mit dem Institutes für Psychologie von der technischen Universität Braunschweig. Über diese Datenerhebung wurde 2019 ein Paper veröffentlicht [5], welches die Studie aufführt.

Die Daten wurden mitsamt einiger Analysetools in der Bachelorarbeit von David Schiller [23] aufgegriffen. Hierfür wurde der Datensatz aufbereitet und die unterschiedlichen Tools auf Richtigkeit analysiert. In eine ähnliche Richtung geht die Masterarbeit von Alexander Specht. In dieser wurden Analysetools untersucht, die nicht speziell für den Bereich Software Engineering entwickelt wurden. Mithilfe von Tests sollten diese verbessert werden, damit sie auch im Bereich der Software Entwicklung genutzt werden konnten.

Ideen zu ähnlichen Ansätzen wurden in den vergangenen Jahren an unterschiedlichen Stellen analysiert. Eine Initiative für die Untersuchung von Audioaufzeichnungen wurde als Open-Source-Projekt von Mozilla gestartet namens Common Voice<sup>1</sup>. Hierbei kann jede Person Texte aufnehmen oder anhören. Beim Anhören soll festgestellt werden, ob diese verständlich sind. Erst bei mehrfacher Bestätigung wird der jeweilige gesprochene Satz in einen Datenpool aufgenommen, mit dem weitere Sprachanalysen durchgeführt werden können.

Eine weitere Idee zur Analyse von Kommunikation entstand 2018 von Schuh et al. [26] außerhalb der Softwareentwicklung. Bei dieser Analyse sollte aggressive Kommunikation in Internetbeiträgen erkannt werden. Dies zeigt einen anderen Lösungsweg für eine Analyse von negativer Kommunikation auf, als die Methode, die in dieser Masterarbeit verwendet wurde.

Alghalibi et al. [2] hat sich wiederum mit der Analyse von Emotionen in Internetbeiträgen, besonders bei Twitter, beschäftigt. Hierbei sollte eine Echtzeit-Bestimmung von Stimmungen entstehen, welche laut Alghalibi et al. [2] beispielsweise für das Ausbauen des Marketing genutzt werden kann.

Diese Ansätze wurden alle außerhalb des Bereichs der Softwareentwicklung analysiert. In einem weiteren Paper von Calefato

---

<sup>1</sup><https://commonvoice.mozilla.org/de>

et al. [9] wird explizit auf die Verwendung von domänenspezifischer Verwendung der Stimmungsanalyse eingegangen. Bei der Untersuchung kam heraus, dass durch angepasste Lexikas die Analyse im Bereich der Softwareentwicklung verbessert werden konnte.

Zu diesem Schluss kamen ebenfalls Jongeling et al. [17] in ihrem Paper von 2017. Sie untersuchten einen Datensatz aus dem Bereich der Softwareentwicklung mit herkömmlichen Tools. Dabei ergab sich, dass es keine Übereinstimmung der Klassifizierung mit domänenspezifischer Klassifizierung als auch unter den herkömmlichen Tools gab und Jonheling et al. [17] in ihrem Paper zu dem Schluss kamen, dass eine Spezifizierung der Domäne wichtig für eine korrekte Klassifikation ist.

Eine andere Sichtweise für die Lösung von Problemen im Entwicklungsteam wurde 2018 von Shakeri et al. [1] veröffentlicht. Hierbei handelt es sich um ein Tool namens ELICA, welches Kommunikationsprobleme zwischen Analysten und Stakeholdern analysiert und verbessern soll.

Die Problemanalyse für Verzögerung von globalen Softwareprojekten wurde in einem Paper von Herbsleb et al. [13] durchgeführt. Dabei wurde untersucht, wie lange die einzelnen Abschnitte der Software in ihrer Produktion bei verteilten Teams benötigten. Als Ergebnis und Grund gaben Herbsleb et al. [13] an, dass durch die Verteilung mehr Personen an den Projekten beteiligt sind und somit die Zeit für die Produktion verdoppelt wurde.

Ein weiteres Problem im Entwicklungsteam ist der Verlust der Produktivität. Zu diesem Thema wurde 2015 von Ortu et al. [21] ein Paper veröffentlicht. Dies handelt von der Untersuchung von Emotionen und Stimmung des Teams und ihren Einfluss auf die Produktivität. Diesbezüglich wurde ein Zusammenhang zwischen den Stimmungen und Emotionen und der Zeit der Produktion anhand von aktiven Jira-Projekten untersucht. Dies kam zu dem Schluss, dass positive Emotionen die Zeit der Bearbeitung verringern und somit die Produktivität steigern.

Auch im Bereich der Softwareprojekte von Studierenden wurden zuvor schon einige Untersuchungen durchgeführt. 2015 untersuchten Bruegge et al. [7] Softwareprojekte an der Technischen Universität München. In ihrem Paper stellten sie eine Vorgehensweise vor, die es ermöglichen sollte die Fähigkeiten den Studierenden zu verbessern, damit solche Softwareprojekte ebenfalls kommerziell verwendbar sind. Die Verbesserung sollte mithilfe von Tools unterstützt werden, die

sowohl Teamarbeit als auch die Kommunikation im Team verbessern sollte. In ihrem Paper beschrieben Bruegge et al. [7] den Erfolg der Studie.

Hinsichtlich der in dieser Masterarbeit analysierten Stimmung in studentischen Softwareprojekten, wurde von Schneider et al. [25] bereits eine Analyse über den Verlauf der Konflikte während studentischer Softwareprojekte gemacht. Dabei kam heraus, dass soziale Konflikte im Verlauf des Projekts linear anstiegen. aufgabenbezogene Konflikte hingegen stiegen zum Anfang des Projekts an, wurden zum Ende hin jedoch deutlich weniger, bildeten also einen Bogen.

Diese genannten Arbeiten sind ein wesentlicher Punkt, wieso das Thema dieser Masterarbeit entstanden ist und sie zeigen auf, wie viele Erweiterungsmöglichkeiten neuer Themen in den Bereichen entstehen können.



# Kapitel 7

## Zusammenfassung und Ausblick

### 7.1 Zusammenfassung

In dieser Masterarbeit wurde der Zusammenhang zwischen polaritären Aussagen im Meeting und der Stimmung im Team vor und nach dem Meeting untersucht. Hierfür wurde eine zentrale Forschungsfrage aufgestellt, welche sich in drei weitere vertiefende Forschungsfragen aufgliedert. Die einzelnen Forschungsfragen behandeln die unterschiedlichen Betrachtungswinkel der möglichen Zusammenhänge.

Im ersten Schritt wurden die Daten, die im Rahmen des Softwareprojekts an der LUH gesammelt wurden, aufbereitet und analysiert. Hierfür wurden unterschiedliche Verfahren genutzt, für die Analyse der Stimmung vor und nach dem Meeting wurden die zuvor gesammelten Daten zusammengefasst. Dies geschah mithilfe eines Verfahrens, welches typischerweise für PANAS-Befragungen genutzt wird. Zuerst wird pro Teammitglied ein Wert für die positiven und ein Wert für die negativen Stimmungen mithilfe des Mittelwerts gebildet. Anschließend wurden alle positiven und alle negativen Werte eines Teams mithilfe der Median-Berechnung zusammengefasst. Dasselbe Vorgehen wurde für die psychologischen Abfragen genutzt.

Für die Analyse des Meetings wurden zuerst die Daten verwendet, welche im Rahmen der Bachelorarbeit von David Schiller [23] aufbereitet wurden. Durch diese Aufbereitung konnten 35 von den ursprünglich 42 Teams für die Auswertung genutzt werden. Von den anderen Teams waren keine zu analysierenden Aussagen mehr

vorhanden, da diese durch mangelnde Qualität weggefallen sind. Anschließend wurde für die Analyse ein zweites Tool ausgewählt. Die Nutzung des Analysetools SEnti-Analyzer vom Fachgebiet Software Engineering stand zuvor fest, da diese Masterarbeit eine Folgearbeit der Arbeit ist, in der das Analysetool entwickelt wurde. Um eine zusätzliche Validität des Tools zu erhalten, wurde aus der Analyse der Bachelorarbeit von David Schiller [23] das Analysetool mit der höchsten Accuracy ausgewählt, welches deutsche Aussagen einlesen konnte: GerVader. Mithilfe dieser beiden Tools wurde der Datensatz dann analysiert.

Im nächsten Schritt wurde der Datensatz von den Meetingtranskripten qualitativ analysiert. Hierfür wurden Vergleiche aufgezeigt, wie die Verarbeitung mit den beiden Analysetools gemacht wurde und welche Ergebnisse dabei entstanden sind. Dabei wurden diese so zusammengefasst, dass nur die Aussagen, die in beiden Tools mit der gleichen Polarität bestimmt wurden, für die quantitative Analyse weiterverwendet wurden. Dies wurde gemacht, damit die Aussagekraft der quantitativen Untersuchung aussagekräftiger wird.

Im Anschluss an diesen Schritt wurde die quantitative Analyse durchgeführt. Hierbei wurden zuerst eine SVM und ein GLM mit den Daten trainiert, um die Forschungsfrage 1 zu beantworten. Mit dem Datensatz konnte jedoch kein Modell trainiert werden, sodass dann eine statistische Untersuchung durchgeführt wurde. Für diese Untersuchung wurden zu jeder der drei Forschungsfragen Hypothesen aufgestellt, die beantwortet werden sollten. Diese Hypothesen deckten die drei Forschungsfragen dahingehend ab, dass am Ende die zentrale Forschungsfrage aus den Ergebnissen beantwortet werden konnte.

## 7.2 Ausblick

In dieser Arbeit wurden einige Untersuchungen durchgeführt, welche auf die Stimmung vor und nach dem Meeting und den Zusammenhang mit den Polaritäten der Aussagen im Meeting eingehen. Bei den Ergebnissen stellt sich die Frage, ob die Ergebnisse erweiterbar sind. Für weiterführende Arbeiten könnten mehr Datensätze verwendet werden, um die Untersuchungsergebnisse zu bestärken. Dies könnte andere Projekte beinhalten, innerhalb und außerhalb der LUH. Dies wären innerhalb der LUH beispielsweise andere Labore wie das

XP-Lab, in welchen ebenfalls Softwareprojekte durchgeführt werden. Außerhalb würde dieser Ansatz die Möglichkeit bieten, den Vergleich von der Untersuchung von Softwareprojekten aus dem Studium zu kommerziellen Softwareprojekten durchzuführen.

Des weiteren könnte die Untersuchung auch für weiterführende Meetings durchgeführt werden. So kann eine Veränderung im Laufe der Durchführung des Projekts untersucht werden. Hierfür wäre eine Erweiterung des Datensatzes im Bezug auf die Daten mehrerer Jahre interessant, um zum einen den Vergleich der unterschiedlichen Studienjahre zu sehen und zum anderen einen großen Datensatz nutzen zu können, um die Probleme dieser Arbeit, dass der Datensatz sehr klein war, zu umgehen.

Die Erweiterung der Daten könnte zusätzlich durch besseres Equipment unterstützt werden. In dieser Arbeit wurde erwähnt, dass einige Audioaufzeichnungen durch schlechte Qualität nicht transkribiert werden konnten. Mit verbessertem Equipment könnte dies umgangen werden.

Durch die Erweiterung des Datensatzes könnte das Problem der genutzten SVM umgangen werden. Zu vermuten ist, dass das nicht mögliche trainieren eines Modells an einem zu geringen Datensatz lag. Mit einem erweiterten Datensatz könnte also beispielsweise eine SVM oder ein neuronales Netz nutzbar werden.



# Anhang A

## Anhang

Abfrage	Kategorie
Wie viele Reibereien gibt es zwischen den Gruppenmitgliedern?	SK
Wie offensichtlich sind persönliche Konflikte in der Gruppe?	SK
Wie viele Spannungen gibt es zwischen den Gruppenmitgliedern?	SK
Wie viele emotionale Konflikte gibt es zwischen den Gruppenmitgliedern?	SK
Wie oft sind sich die Gruppenmitglieder uneinig?	AK
Wie häufig gibt es Ideenkonflikte in der Gruppe?	AK
Wie viele die Arbeit betreffende Konflikte gibt es in der Gruppe?	AK
In welchem Ausmaß gibt es Meinungsverschiedenheiten in der Gruppe?	AK
Mit dem Diskussionsverlauf bin ich insg. zufrieden.	MZ
Die Gruppendiskussion war für mich sinnvoll verwendete Zeit.	MZ
Mit dem Diskussionsergebnis bin ich insg. zufrieden.	MZ
Das Diskussionsergebnis ist klar und eindeutig.	MZ

Anmerkung: SK = soziale Konflikte, AK = aufgabenbezogene Konflikte, MZ = Mitarbeiterzufriedenheit

Tabelle A.1: Abfragen der psychologischen Befragung im Softwareprojekt mit den jeweiligen Kategorien

		GerVader		
SEnti-Analyzer		negativ	positiv	neutral
negativ		1	1	11
positiv		3	70	164
neutral		52	141	499

Tabelle A.2: Anzahl der polaritären Zuordnung der Analysetools im Vergleich von Team 5

		GerVader		
SEnti-Analyzer		negativ	positiv	neutral
negativ		1	1	4
positiv		4	14	38
neutral		16	59	164

Tabelle A.3: Anzahl der polaritären Zuordnung der Analysetools im Vergleich von Team 16

		GerVader		
SEnti-Analyzer		negativ	positiv	neutral
negativ		0	1	2
positiv		0	25	43
neutral		44	106	409

Tabelle A.4: Anzahl der polaritären Zuordnung der Analysetools im Vergleich von Team 21

		GerVader		
SEnti-Analyzer		negativ	positiv	neutral
negativ		0	0	1
positiv		0	20	14
neutral		20	65	164

Tabelle A.5: Anzahl der polaritären Zuordnung der Analysetools im Vergleich von Team 34

		GerVader		
		negativ	positiv	neutral
SEnti-Analyzer	negativ	5	0	8
	positiv	1	37	71
	neutral	33	128	299

Tabelle A.6: Anzahl der polaritären Zuordnung der Analysetools im Vergleich von Team 37

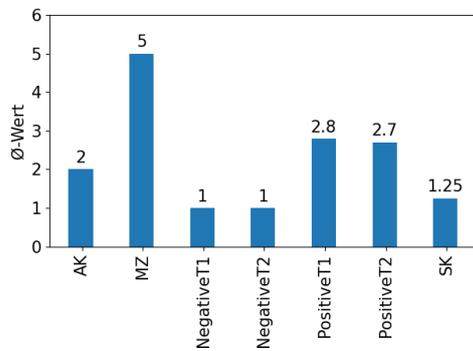


Abbildung A.1: PANAS-Affekte und psychologischen Abfragen im Durchschnitt von Team 5

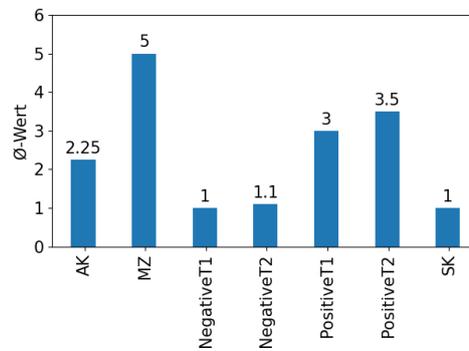


Abbildung A.2: PANAS-Affekte und psychologischen Abfragen im Durchschnitt von Team 16

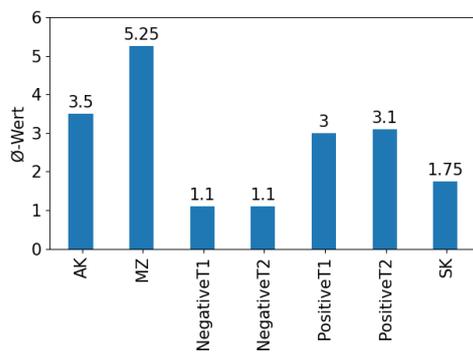


Abbildung A.3: PANAS-Affekte und psychologischen Abfragen im Durchschnitt von Team 21

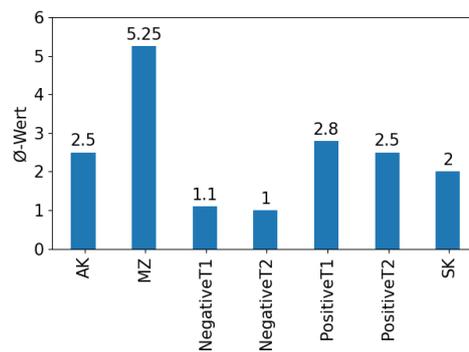


Abbildung A.4: PANAS-Affekte und psychologischen Abfragen im Durchschnitt von Team 34

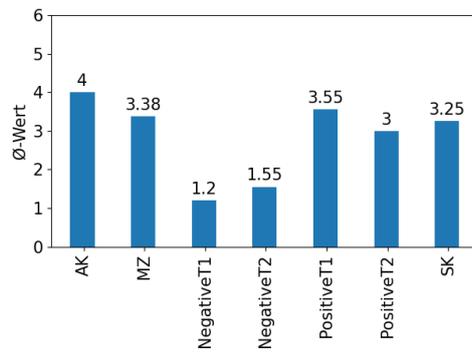


Abbildung A.5: PANAS-Affekte und psychologischen Abfragen im Durchschnitt von Team 37

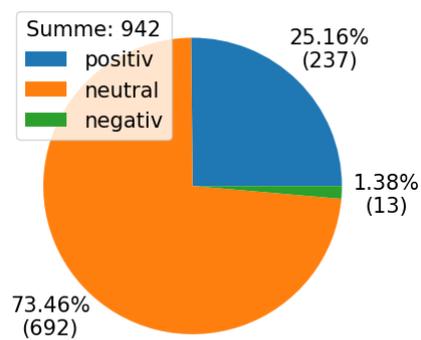


Abbildung A.6: Polaritäre Anteile der Analyse des Senti-Analyser zum Team 5

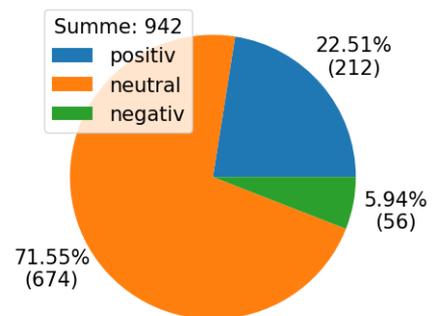


Abbildung A.7: Polaritäre Anteile der Analyse von GerVader zum Team 5

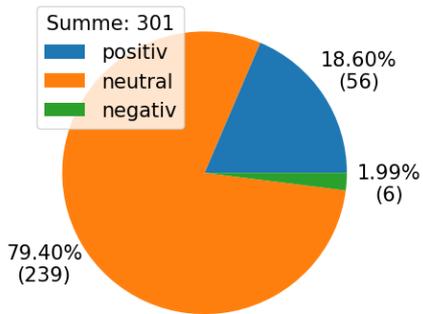


Abbildung A.8: Polaritäre Anteile der Analyse des SEnti-Analyser zum Team 16

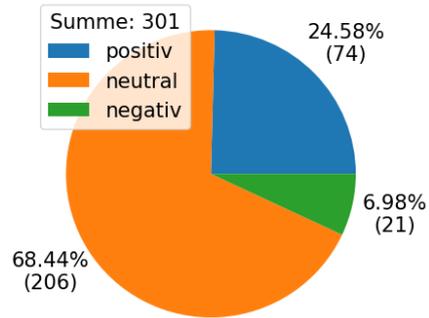


Abbildung A.9: Polaritäre Anteile der Analyse von GerVader zum Team 16

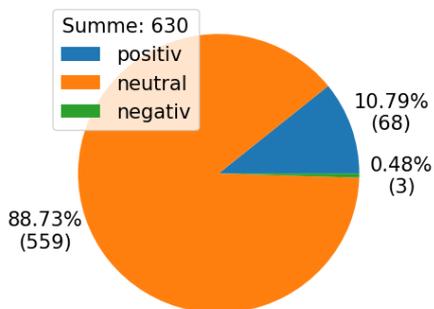


Abbildung A.10: Polaritäre Anteile der Analyse des SEnti-Analyser zum Team 21

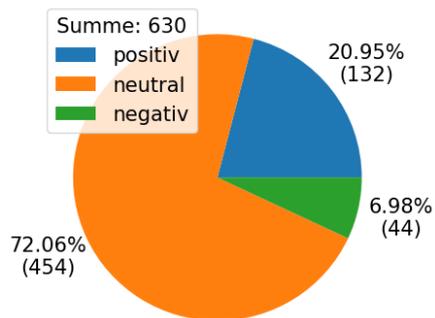


Abbildung A.11: Polaritäre Anteile der Analyse von GerVader zum Team 21

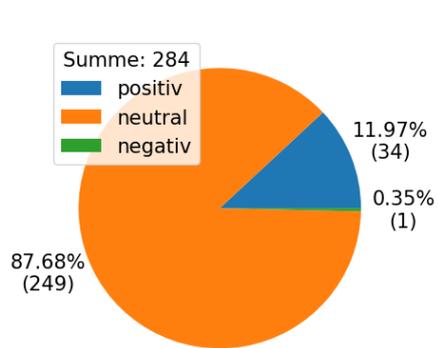


Abbildung A.12: Polaritäre Anteile der Analyse des Senti-Analyser zum Team 34

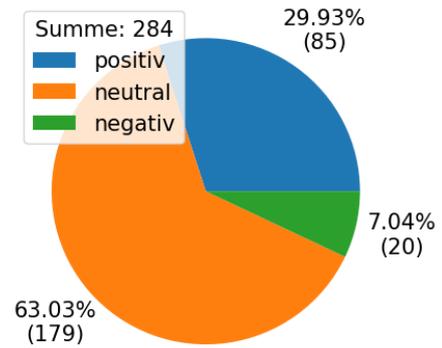


Abbildung A.13: Polaritäre Anteile der Analyse von GerVader zum Team 34

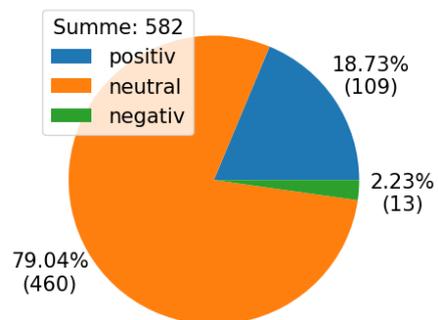


Abbildung A.14: Polaritäre Anteile der Analyse des Senti-Analyser zum Team 37

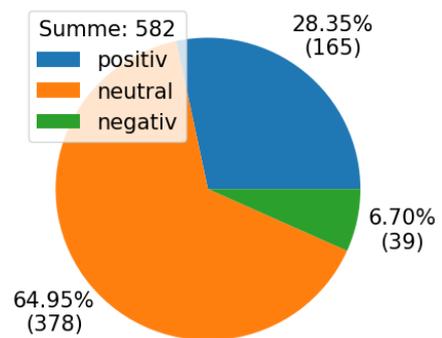


Abbildung A.15: Polaritäre Anteile der Analyse von GerVader zum Team 37

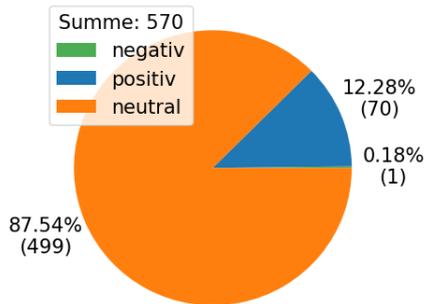


Abbildung A.16: Polaritäre Anteile der Gemeinsamkeiten der Analysetools zum Team 5

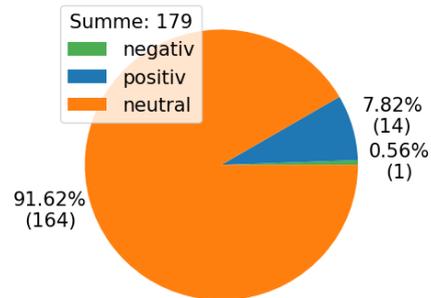


Abbildung A.17: Polaritäre Anteile der Gemeinsamkeiten der Analysetools zum Team 16

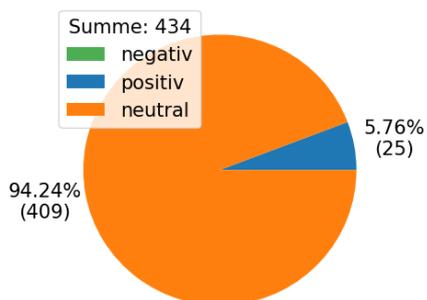


Abbildung A.18: Polaritäre Anteile der Gemeinsamkeiten der Analysetools zum Team 21

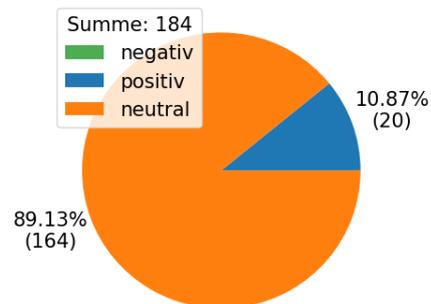


Abbildung A.19: Polaritäre Anteile der Gemeinsamkeiten der Analysetools zum Team 34

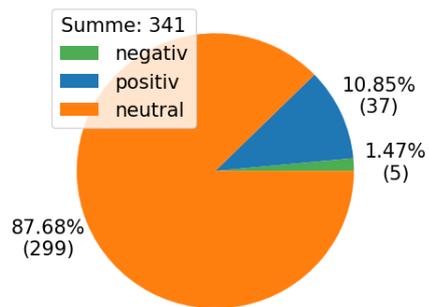


Abbildung A.20: Polaritäre Anteile der Gemeinsamkeiten der Analysetools zum Team 37

# Literaturverzeichnis

- [1] Z. S. H. Abad, V. Gervasi, D. Zowghi, and K. Barker. Elica: an automated tool for dynamic extraction of requirements relevant information. In *2018 5th International Workshop on Artificial Intelligence for Requirements Engineering (AIRE)*, pages 8–14. IEEE, 2018.
- [2] M. Alghalibi, A. Al-Azzawi, and K. Lawonn. Deep tweets analyzer model for twitter mood visualization and prediction based deep learning approach. *International Journal of Computer and Communication Engineering*, 8:1–17, 2019.
- [3] R. K. Bakshi, N. Kaur, R. Kaur, and G. Kaur. Opinion mining and sentiment analysis. In *2016 3rd international conference on computing for sustainable global development (INDIACom)*, pages 452–455. IEEE, 2016.
- [4] R. Bender and S. Lange. Was ist der p-wert? *DMW-Deutsche Medizinische Wochenschrift*, 132(S 01):e15–e16, 2007.
- [5] H. Brandt-Pook and R. Kollmeier. *Softwareentwicklung kompakt und verständlich*. Springer, 2015.
- [6] B. Breyer and M. Bluemke. Deutsche version der positive and negative affect schedule panas (gesis panel). Technical report, Zusammenstellung sozialwissenschaftlicher Items und Skalen (ZIS), 2016.
- [7] B. Bruegge, S. Krusche, and L. Alperowitz. Software engineering project courses with industrial clients. *ACM Transactions on Computing Education (TOCE)*, 15(4):1–31, 2015.
- [8] F. Calefato, F. Lanubile, F. Maiorano, and N. Novielli. Sentiment polarity detection for software development. *Empirical Software Engineering*, 2018.

- [9] F. Calefato, N. Novielli, and F. Lanubile. The challenges of sentiment detection in the social programmer ecosystem. In *Proceedings of the 7th International Workshop on Social Software Engineering*, pages 33–40, 2015.
- [10] R. Cohn-Gordon and N. Goodman. Lost in machine translation: A method to reduce meaning loss. *arXiv preprint arXiv:1902.09514*, 2019.
- [11] A. D’Andrea, F. Ferri, P. Grifoni, and T. Guzzo. Approaches, tools and applications for sentiment analysis implementation. *International Journal of Computer Applications*, 125(3), 2015.
- [12] D. Graziotin, X. Wang, and P. Abrahamsson. Happy software developers solve problems better: psychological measurements in empirical software engineering. *PeerJ*, 2:e289, 2014.
- [13] J. D. Herbsleb and A. Mockus. An empirical study of speed and communication in globally distributed software development. *IEEE Transactions on software engineering*, 29(6):481–494, 2003.
- [14] M. Herrmann. Automatische klassifikation von aussagen in meetings von entwicklungsteams. Bachelorthesis, Leibniz Universität Hannover, March 2021.
- [15] M. Herrmann and J. Klünder. From textual to verbal communication: Towards applying sentiment analysis to a software project meeting. *arXiv preprint arXiv:2108.01985*, 2021.
- [16] J. Horstmann. Computer-gestützte analyse des kommunikationsverhaltens in entwicklerteams unter berücksichtigung digitaler medien. Masterthesis, Leibniz Universität Hannover, October 2019.
- [17] R. Jongeling, P. Sarkar, S. Datta, and A. Serebrenik. On negative results when using sentiment analysis tools for software engineering research. *Empirical Software Engineering*, 22(5):2543–2584, 2017.
- [18] J. Klünder, J. Horstmann, and O. Karras. Identifying the mood of a software development team by analyzing text-based communication in chats with machine learning. In *International Conference on Human-Centred Software Engineering*, pages 133–151. Springer, 2020.

- [19] R. E. Kraut and L. A. Streeter. Coordination in software development. *Commun. ACM*, 38(3):69–81, Mar. 1995.
- [20] C. Meyer. Verbesserung von evolutionären algorithmen zur klassifikation von schriftlicher kommunikation in entwicklungsteams. Bachelorthesis, Leibniz Universität Hannover, November 2020.
- [21] M. Ortu, B. Adams, G. Destefanis, P. Tourani, M. Marchesi, and R. Tonelli. Are bullies more productive? empirical study of affectiveness vs. issue fixing time. In *2015 IEEE/ACM 12th Working Conference on Mining Software Repositories*, pages 303–313. IEEE, 2015.
- [22] H. F. K. Paulsen and S. Kauffeld. Ansteckungsprozesse in gruppen: Die rolle von geteilten gefühlen für gruppenprozesse und-ergebnisse. *Gruppe. Interaktion. Organisation. Zeitschrift für Angewandte Organisationspsychologie (GIO)*, 47(4):357–364, 2016.
- [23] D. Schiller. Untersuchung von zusammenhängen zwischen stimmungen und interaktionen von meetings in softwareprojekten. Bachelorthesis, Leibniz Universität Hannover, October 2019.
- [24] K. Schneider, J. Klünder, F. Kortum, L. Handke, J. Straube, and S. Kauffeld. Positive affect through interactions in meetings: The role of proactive and supportive statements. *Journal of Systems and Software*, 143:59–70, 2018.
- [25] K. Schneider, O. Liskin, H. Paulsen, and S. Kauffeld. Media, mood, and meetings: Related to project success? *ACM Transactions on Computing Education (TOCE)*, 15(4):1–33, 2015.
- [26] T. Schuh and S. Dreiseitl. Evaluating novel features for aggressive language detection. In *International Conference on Speech and Computer*, pages 585–595. Springer, 2018.
- [27] G. Stepanek. *Software Projects Secrets: Why Projects Fail*. APress, 2012.
- [28] M. Thelwall, K. Buckley, G. Paltoglou, D. Cai, and A. Kappas. Sentiment strength detection in short informal text. *Journal of the American society for information science and technology*, 61(12):2544–2558, 2010.

- [29] D. A. Tietze and T. Schümmer. *Kooperative Softwareentwicklung*, pages 264–275. Springer Berlin Heidelberg, Berlin, Heidelberg, 2001.
- [30] K. Tymann, M. Lutz, P. Palsbröcker, and C. Gips. Gervader-a german adaptation of the vader sentiment analysis tool for social media texts. In *LWDA*, pages 178–189, 2019.
- [31] D. Vilares, M. Thelwall, and M. A. Alonso. The megaphone of the people? spanish sentistrength for real-time analysis of political tweets. *Journal of Information Science*, 41(6):799–813, 2015.
- [32] A. G. Vural, B. B. Cambazoglu, P. Senkul, and Z. O. Tokgoz. A framework for sentiment analysis in turkish: Application to polarity detection of movie reviews in turkish. In *Computer and Information Sciences III*, pages 437–445. Springer, 2013.
- [33] K. Völkl and C. Korb. *Deskriptive Statistik*, chapter Variablen und Skalenniveaus, pages 15–19. Springer VS, Wiesbaden, 2018.
- [34] D. Watson and H. Levin-Aspenson. *Positive and Negative Affect Schedule (PANAS)*, pages 1–3. Springer International Publishing, Cham, 2017.
- [35] M. Wolfgruber. *Sentiment Analyse mit lokalen Grammatiken*. PhD thesis, Ludwig-Maximilians-Universität München, 2015.
- [36] X. Wu and V. Kumar, editors. *The Top Ten Algorithms in Data Mining (1st ed.)*, chapter SVM: Support Vector Machines, page 208. Chapman and Hall/CRC, 2009.