

# Sentimentanalyse der Medienberichterstattung und Online-Kommentare im Kontext der Schweizer Parlamentswahlen 2023

Im Rahmen des Moduls “Applied Data Science I”  
an der Digital Business University of Applied Sciences DBU Berlin

|                        |                                     |
|------------------------|-------------------------------------|
| Autor:                 | Simon Sahli                         |
| Studiengang:           | Data Science und Management (M.Sc.) |
| Immatrikulation:       | 190252                              |
| Adresse:               | CH-3065 Bern                        |
| E-Mail:                | simon.sahli@student.dbuas.de        |
| Datum der Einreichung: | 31.10.2023                          |

## **Abstract**

Diese Studienarbeit analysierte die Tonalität und Meinungsvielfalt in der Berichterstattung von zwei führenden Schweizer Online-Medien – srf.ch und 20min.ch – im Kontext der Schweizer Parlamentswahlen 2023. Durch die Verwendung moderner Sentimentanalyse-Techniken wurden 303 Artikel und insgesamt 16'305 dazugehörige Nutzerkommentare untersucht. Die Ergebnisse zeigten, dass beide Medienplattformen eine tendenziell neutrale Berichterstattung an den Tag legten. Im Gegensatz dazu wiesen die Online-Kommentare eine breitere Meinungsvielfalt und stärkere Emotionalität auf. Zusätzlich lieferte die Untersuchung Einblicke in das Verhalten der Nutzer, insbesondere in Bezug auf bevorzugte Zeiten für Kommentare. Während der Analyse wurden Webscraping-Techniken eingesetzt, um relevante Daten von den beiden Plattformen zu gewinnen. Die Arbeit dient als Einblick in die Dynamik zwischen Medienberichterstattung und Leserreaktionen in einem politischen Kontext und unterstreicht die Bedeutung von Online-Kommentaren als Spiegelbild der öffentlichen Meinung.

# Inhaltsverzeichnis

|          |  |           |
|----------|--|-----------|
| <b>1</b> | <b>Einleitung</b>                      | <b>4</b>  |
| <b>2</b> | <b>Daten und Methoden</b>              | <b>5</b>  |
| 2.1      | Sentimentanalyse . . . . .             | 5         |
| 2.2      | Webscraping . . . . .                  | 6         |
| 2.3      | Explorative Datenanalyse EDA . . . . . | 8         |
| <b>3</b> | <b>Ergebnisse</b>                      | <b>13</b> |
| <b>4</b> | <b>Diskussion</b>                      | <b>19</b> |
| <b>5</b> | <b>Fazit</b>                           | <b>21</b> |
| <b>6</b> | <b>Quellenverzeichnis</b>              | <b>22</b> |
| <b>A</b> | <b>Appendix</b>                        | <b>23</b> |
| A.1      | Code . . . . .                         | 23        |

# 1 Einleitung

Am 22. Oktober 2023 fanden in der gesamten Schweiz die Parlamentswahlen statt (Bundeskanzlei 2023)[2]. Nach der Wahl wird das neu formierte Parlament im Dezember die Mitglieder des Bundesrats sowie den Bundeskanzler oder die Bundeskanzlerin bestimmen. Seit den letzten nationalen Wahlen sind bereits vier Jahre vergangen. Während dieser Zeit hat sich sowohl in der Schweiz als auch weltweit viel verändert. Entsprechend ausgeprägt findet die Berichterstattung in Online-Newsportalen statt.

Im Kontext bevorstehender Wahlen und Abstimmungen initiieren nicht nur Bundesrat und Behörden, sondern auch Parteien, Verbände und NGOs zielgerichtete Kampagnen, um die Stimmberechtigten mit ihren Argumenten zu gewinnen. Doch wie ausgewogen gestaltet sich die Berichterstattung der Schweizer Medien? Diese Frage war bereits Gegenstand einer Studie, die vom Forschungszentrum Öffentlichkeit und Gesellschaft (fög) der Universität Zürich[7] im September 2023 veröffentlicht wurde. Die Studie analysierte die Berichterstattung zu 44 Volksabstimmungen im Zeitraum zwischen 2018 und 2023. Nach Angaben des Autors der Studie positionieren sich die meisten Medien, insbesondere die grossen öffentlichen Medien SRF und RTS sowie die Pendlermedien von 20 Minuten, nicht an den extremen Polen des politischen Spektrums, sondern tendieren eher zur Mitte.

Die vorliegende Studienarbeit zielt darauf ab, die Online-Berichterstattung im Umfeld der Wahlen am 22. Oktober zu untersuchen und die Erkenntnisse der Studie der Universität Zürich zu überprüfen. Dabei kommen Methoden des Webscraping und Text Mining zum Einsatz. Diese Techniken helfen, Artikel im Kontext der Wahlen automatisiert zu erfassen und auszuwerten. Der Schwerpunkt der Arbeit liegt auf einer Sentimentanalyse, die sowohl die Artikel als auch Nutzerkommentare untersucht, um positive oder negative Haltungen in den Inhalten und Diskussionen zu identifizieren. Die Analyse fokussiert sich auf zwei führende Medien, die ihre Beiträge ohne Kostenzugang veröffentlichen: [www.srf.ch](http://www.srf.ch) und [www.20min.ch](http://www.20min.ch).

Entsprechend werden folgende Hypothesen aufgestellt, welche im Rahmen der Studienarbeit untersucht werden:

- H1: Die Berichterstattung der untersuchten Online-Medien (SRF.ch und 20min.ch) im Kontext der Parlamentswahlen 2023 zeigt eine neutrale Ausrichtung in Bezug auf Tonalität.
- H2: Kommentare von Leser und Leserinnen zu Beiträgen über die Wahlen 2023 zeichnen sich durch eine grössere Meinungsvielfalt und emotionalere Tonalität aus als die zugrunde liegenden Artikel. Abhängig vom jeweiligen Nachrichtenmedium und aktuellen Geschehnissen können die Intensität und Richtung der geäusserten Meinungen variieren.

Dieser erste Teil der Studienarbeit stellt die Motivation und Problemstellung kurz dar, um daraus sieben vorgestellte Hypothesen abzuleiten, die mit Hilfe einer Datenbasis analysiert werden sollen. Im zweiten Teil wird diese Datenbasis im Rahmen einer explorativen Datenanalyse genauer untersucht. Zudem werden die Analysemethoden und Techniken, welche für die Auswertung der Hypothesen verwendet werden, vorgestellt. Im dritten Kapitel der Arbeit werden die Ergebnisse präsentiert. Die Arbeit schliesst mit einer Einordnung der Ergebnisse in eine mögliche Diskussion mit Handlungsempfehlungen.

Eine Übersicht über das Projekt, der Report selbst als auch die Datengrundlage ist unter folgendem Link einsehbar:

[https://www.simonsahli.ch/project\\_wahlen2023.html](https://www.simonsahli.ch/project_wahlen2023.html)

## 2 Daten und Methoden

### 2.1 Sentimentanalyse

In der Studie von Haselmayer und Jenny (2017)[4] wird die Sentimentanalyse detailliert erläutert. Laut den Autoren erfasst die Sentimentanalyse die Polarität oder Tonalität von Texten, indem sie Ausdrücke identifiziert und bewertet.

Im Englischen existieren bereits zahlreiche bewertete Wörterbücher und Modelle, wie die in der Arbeit von DiBattista (2021)[3] vorgestellten NLTK/VADER (regelbasiert) oder Flair (Einbettung in ein vorab trainiertes Datenmodell).

Haselmayer und Jenny erläutern weiterhin, dass im nicht-englischsprachigen Raum eine computer-gestützte Sentimentanalyse teilweise nicht verfügbar ist oder die Ergebnisse nicht zufriedenstellend sind. In solchen Fällen kann auf traditionelles Encoding von Texten zurückgegriffen werden, bei dem Menschen Texte oder eine Auswahl von Sätzen lesen und bewerten. Es existieren Crowdcoding-Services, die eine ressourcenschonende Bewertung von Big-Data-Projekten ermöglichen.

Die vorliegende Arbeit verfolgt einen alternativen Ansatz. Dank stetig verbesserter Übersetzungsdienste wird der Datensatz zunächst ins Englische übersetzt. Dies lässt sich mit einer Python-Bibliothek wie “deep-translator” von Baccouri[1] automatisieren, die auf APIs verschiedener Übersetzungsdienste zugreift. Obwohl dieser Übersetzungsschritt sehr rechenintensiv ist und Nuancen der deutschen Sprache möglicherweise nicht vollständig übertragen werden können, bietet der übersetzte Datensatz dennoch einen geeigneten Kompromiss im Vergleich zu einer eigens erstellten Bibliothek oder einem traditionellen manuellen Encoding. So kann auf bereits vorhandene Datenmodelle und Bibliotheken in englischer Sprache zurückgegriffen werden. Dies sollte im Hinblick auf die Genauigkeit für diese Arbeit ausreichend sein.

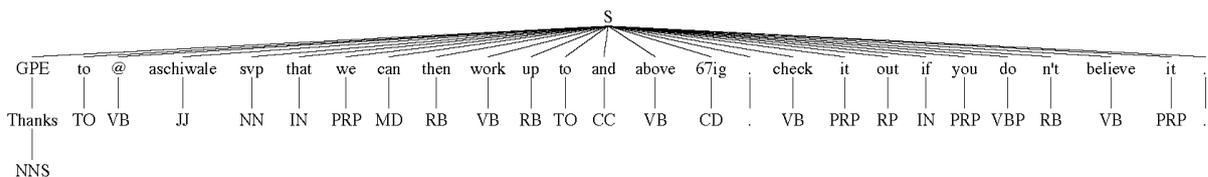
Für die eigentliche Sentiment-Analyse werden zwei Modelle gewählt und verglichen:

- **VADER** (regelbasiert) nach Hutto und Gilbert (2014)[5] – Valence Aware Dictionary and sEntiment Reasoner. Hier werden alle Texte in ihre einzelnen Bausteine aufgebrochen, neu gruppiert und nach einem festgelegten Regelsystem bewertet. Stoppwörter werden entfernt, sodass je Text ein Gesamtscore errechnet wird, der die Gefühlslage des Textes widerspiegeln sollte. Der Score bewegt sich in einem Wertebereich von -1 (negative Tonalität) bis zu 1 (positive Tonalität). Es ist zu beachten, dass dieses Modell nicht auf die Beziehungen zwischen den Wörtern eingehen und daher Konzepte wie Ironie nicht erfassen kann.

Beispiel eines Kommentars:

*“@aschiwale svp sei dank, dass wir dann bis und über 67ig arbeiten können. schaut es euch an wenn ihr es nicht glaubt.”*

Aufteilung in einzeln übersetzte Tokens:



- **RoBERTa** (NLA-Transformationsmodell) nach Liu et al. (2019)[6]. Mit Hilfe dieses Modells können der Kontext einzelner Textbausteine ermittelt und in die Berechnung der Tonalität einbezogen werden. Das RoBERTa-Modell ist in der Lage, Wortverbindungen zu interpretieren, und stützt sich dabei auf ein vortrainiertes Datenmodell, auf das im Programmcode zugegriffen wird. Eine Herausforderung bei der Verwendung von RoBERTa besteht in der Begrenzung der Textlänge. Jedes Textelement (Token), das in das Modell eingegeben wird, durchläuft eine Reihe von Transformationen. Aufgrund der Architektur des Modells sowie technischer Einschränkungen hinsichtlich Speicher- und Rechenkapazität wurde daher ein Token-Limit festgelegt. Entsprechend mussten die Artikeltexte dafür in einzelne Abschnitte unterteilt werden, die zunächst separat in das RoBERTa-Modell integriert wurden. Nach einer aufwendigen Berechnung wurden die Abschnitte mit ihren RoBERTa-Scores wieder zusammengeführt.

## 2.2 Webscraping

Um eine Sentimentanalyse durchführen zu können, bedarf es Daten. Diese wurden in dieser Arbeit mit Hilfe von Webscraping selbst erhoben. Gemäss Munzert (2014)[8] ist Webscraping eine datenwissenschaftliche Technik, bei der Skripte zur Extraktion strukturierter Daten aus Webseiten eingesetzt werden. Für diese Arbeit wurden die Webseiten 20min.ch und SRF.ch ausgewählt. Beide

Plattformen verfügen über eine äusserst hohe Reichweite und stellen ihre Online-Artikel kostenlos zur Verfügung. Alternativen mit hoher Reichweite aus der privaten Mediengruppe “Tamedia” wären zwar interessant gewesen, relevante Online-Artikel sind jedoch zum grössten Teil hinter einer Paywall versteckt und somit nicht direkt zugänglich.

Mithilfe der Python-Bibliothek “BeautifulSoup” von Richardson (n.d.)[9] lassen sich die Artikelseiten in ihrer HTML- bzw. XML-Struktur analysieren und automatisiert nach Informationen innerhalb der Architektur durchsuchen.

So wurde beispielsweise bei 20min.ch festgestellt, dass Artikel zu den Wahlen im Oktober 2023 unter folgender URL dynamisch zusammengetragen werden: <https://www.20min.ch/schweiz/wahlen-2023>. Mit Hilfe von “BeautifulSoup” können innerhalb dieses dynamischen Zusammenzugs alle <a>-Elemente – also Verlinkungen – auf andere Webseiten gesammelt werden. Diese werden in einem zweiten Schritt so gefiltert, dass ausschliesslich Links zu Artikelseiten verbleiben. Eine selbst erstellte Funktion durchläuft die Sammlung an Links von Online-Artikeln und speichert die entsprechende HTML-Seite des jeweiligen Artikels lokal ab. Jede Seite wird dabei mit einer eindeutigen ID gekennzeichnet.

In einem zweiten Schritt wird mithilfe eines Skripts untersucht, ob der jeweilige Online-Artikel einen Kommentarbereich aktiviert hat, da dies nicht bei jedem Artikel der Fall ist. Hier zeigt sich die erste Schwierigkeit beim Webscraping: der Unterschied zwischen dynamischem und statischem Inhalt von Webseiten. Die lokal abgespeicherten HTML-Seiten sind eine Momentaufnahme des Zustandes der Webseite zum Zeitpunkt des Scrapings und ermöglichen die Analyse der Architektur und das Herausfiltern von Informationen. Dies funktioniert gut für den eigentlichen Artikeltext. Leser-Kommentare hingegen sind – bis auf die dargestellten Top-Kommentare – dynamische JavaScript-Bereiche innerhalb der Artikel-Seiten, die erst durch manuelles Scrollen oder Klicken auf “weitere Kommentare anzeigen” Inhalte vom Server des Anbieters laden und im Quellcode der Seite ersichtlich machen. Bei 20min.ch sind die vollständigen User-Kommentare zudem auf einer separaten URL abgespeichert, sodass ein Zwischenschritt und die Anwendung einer weiteren Python-Bibliothek erforderlich sind. Mithilfe des “Selenium”-Pakets (n.d.)[10] kann die Interaktion mit dem Webbrowser automatisiert und ein Browser aufgerufen werden, der die Artikelseiten mit Kommentarbereich selbstständig aufruft. Der simulierte Browser klickt selbstständig auf “alle Kommentare anzeigen” oder scrollt ans Ende des Bildschirmfensters, bis möglichst alle Kommentare bzw. Webseiten-Inhalte geladen sind. Anschliessend wird die Seite bzw. deren Inhalt (die User-Kommentare) lokal abgespeichert. Dies funktioniert nicht immer vollständig, da bei bestimmten Seiten beispielsweise das automatisierte Laden von weiteren User-Kommentaren aufgrund von Pop-Up-Fenstern (zum Akzeptieren von Cookies) regelmässig zusammenbrach. Zudem konnten bei SRF.ch im Gegensatz zu 20min.ch die Antworten auf User-Kommentare nicht vollständig geladen werden. Um diese auswerten zu können, hätte der

Selenium-Browser bei jedem einzelnen Kommentar mit Antwort auf “Antworten anzeigen” klicken müssen, was jedoch nicht umgesetzt wurde.

Das technische Verfahren für SRF.ch gleicht im Grundsatz dem für 20min.ch, jedoch gibt es aufgrund der individuellen Architektur der Webseiten-Inhalte einige Abweichungen. Diese Unterschiede stellen auch die grösste technische Herausforderung im Rahmen der Studienarbeit dar. Für eine Analyse von mehreren Newsanbietern ist es notwendig, die jeweilige Architektur der Webseiten-Elemente im Quellcode genau zu analysieren. Damit das automatisierte Scraping der Daten auch funktioniert, sind regelmässige Tests erforderlich.

Eine weitere Herausforderung bestand in der langen Rechenzeit. Durch die Interaktion mit dem Webbrowser via Selenium-Paket und das manuelle Abrollen von Kommentarseiten zum Laden der dynamischen Webseiten-Inhalte liefen die einzelnen Skripte teilweise mehrere Stunden. Dies ist in Bezug auf Effizienz und Ressourcenschonung definitiv nicht die ideale Lösung.

Eine letzte Schwierigkeit ergab sich aus ethischen Überlegungen und dem Datenschutz. In der Regel sind Autorennamen und andere Metadaten, die öffentlich in einem Artikel erscheinen, nicht Gegenstand von Datenschutzanforderungen. Dennoch ist es wichtig, bei der Verwendung solcher Daten ethische Überlegungen anzustellen und die Privatsphäre bzw. Würde der Autoren zu respektieren – insbesondere wenn aus den gesammelten Daten und Visualisierungen Schlüsse gezogen werden können. Deshalb wurden die Namen – obwohl sie durch das Rekonstruieren der Studienarbeit erneut von den Webseiten der Anbieter heruntergeladen werden könnten – in den Illustrationen der Arbeit anonymisiert.

### **2.3 Explorative Datenanalyse EDA**

Im Rahmen der Wahlen vom 22. Oktober 2023 wurden insgesamt 61 Artikel von 14 Autorinnen und Autoren von 20min.ch sowie 242 Artikel von 69 Autorinnen und Autoren von SRF.ch gespeichert und aufbereitet. Zudem umfassten die Artikel insgesamt 16'305 User-Kommentare, wobei 15'429 davon von 20min.ch stammten. Die durchschnittliche Länge eines Artikels von 20min.ch zum Thema der Wahlen lag bei 4'214 Zeichen, bei SRF.ch betrug die durchschnittliche Länge 3'301 Zeichen. Die systematische Überprüfung der Newsportale auf neue Artikel begann am 29. September. Um die Studienarbeit fristgerecht bis Ende Oktober einreichen zu können, wurden Artikel, die bis zum 23. Oktober veröffentlicht wurden, berücksichtigt.

Abbildung 1 illustriert nicht nur die Verteilung der Artikel, sondern auch einen grundlegenden Unterschied zwischen den beiden Newsportalen: Während bei 20min.ch bei der Mehrheit (80.3 %) der archivierten Artikel die Kommentarfunktion aktiv war, konnte bei SRF.ch nur bei 22.3 % der Artikel die eigene Meinung geteilt werden. Abbildung 2 visualisiert die Anzahl der Kommentare in Abhängigkeit von Wochentag und Uhrzeit. Auffällig ist das gesteigerte Kommentaraufkommen

an Wochenenden. Während der Arbeitswoche sind vor allem der Freitag, Montag und Donnerstag hervorzuheben, wobei in diesen Fällen Kommentare vermehrt zu Pausenzeiten abgegeben wurden.

Abbildung 1: Online-Artikel mit und ohne Kommentarbereich

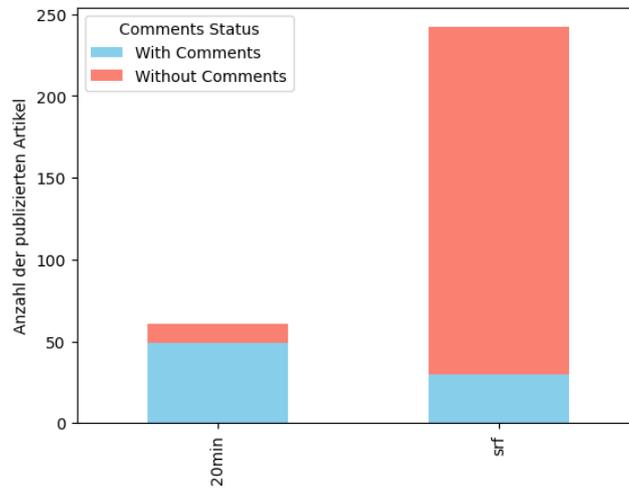
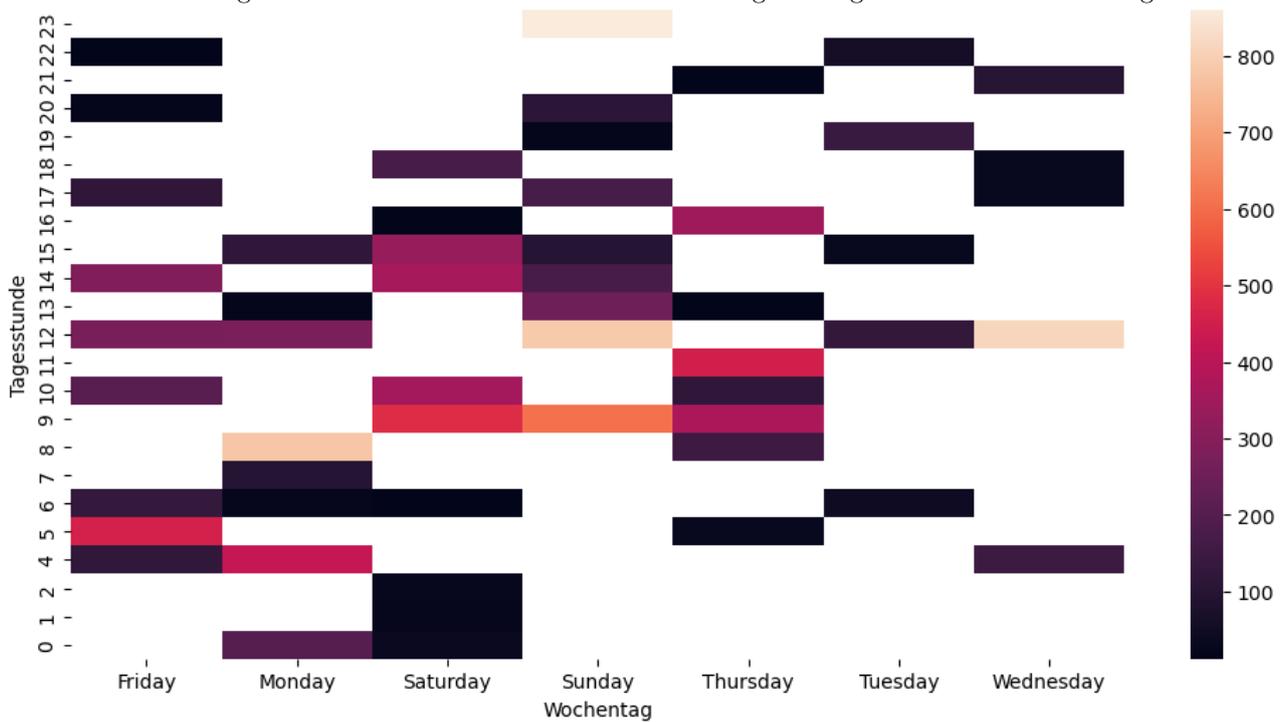


Abbildung 2: Anzahl an Nutzerkommentaren in Bezug auf Tagesstunde und Wochentag



Die Newsdienste veröffentlichen ihre Artikel im Verlauf der Woche unterschiedlich. Wie Abbildung 3 zeigt, wird während der Arbeitswoche besonders am Freitag viel publiziert. Das Wochenende ist ebenfalls ein beliebter Zeitpunkt für Veröffentlichungen, insbesondere weil die Wahlen am Wochenende stattfanden und zahlreiche Artikel am Wahlwochenende vorbereitet und veröffentlicht wurden. Das geringe Kommentaraufkommen am Dienstag und Mittwoch spiegelt sich auch in den Daten zur Artikelveröffentlichung wider: Am Dienstag wurde von 20min und SRF am wenigsten publiziert.

20min.ch veröffentlicht bereits um 4 Uhr morgens einen Grossteil seiner Artikel, wahrscheinlich um Lese-

rinnen und Leser vor Arbeitsbeginn zu erreichen. Bei SRF.ch liegt der Hauptveröffentlichungszeitpunkt später am Vormittag. Gegen Ende des Arbeitstages steigt, ähnlich wie bei 20min.ch, auch bei SRF.ch die Anzahl der Veröffentlichungen im Vergleich zu anderen Tageszeiten an.

Abbildung 3: Publizierter Artikel zu den Wahlen vom 22. Oktober nach Uhrzeit und Wochentag

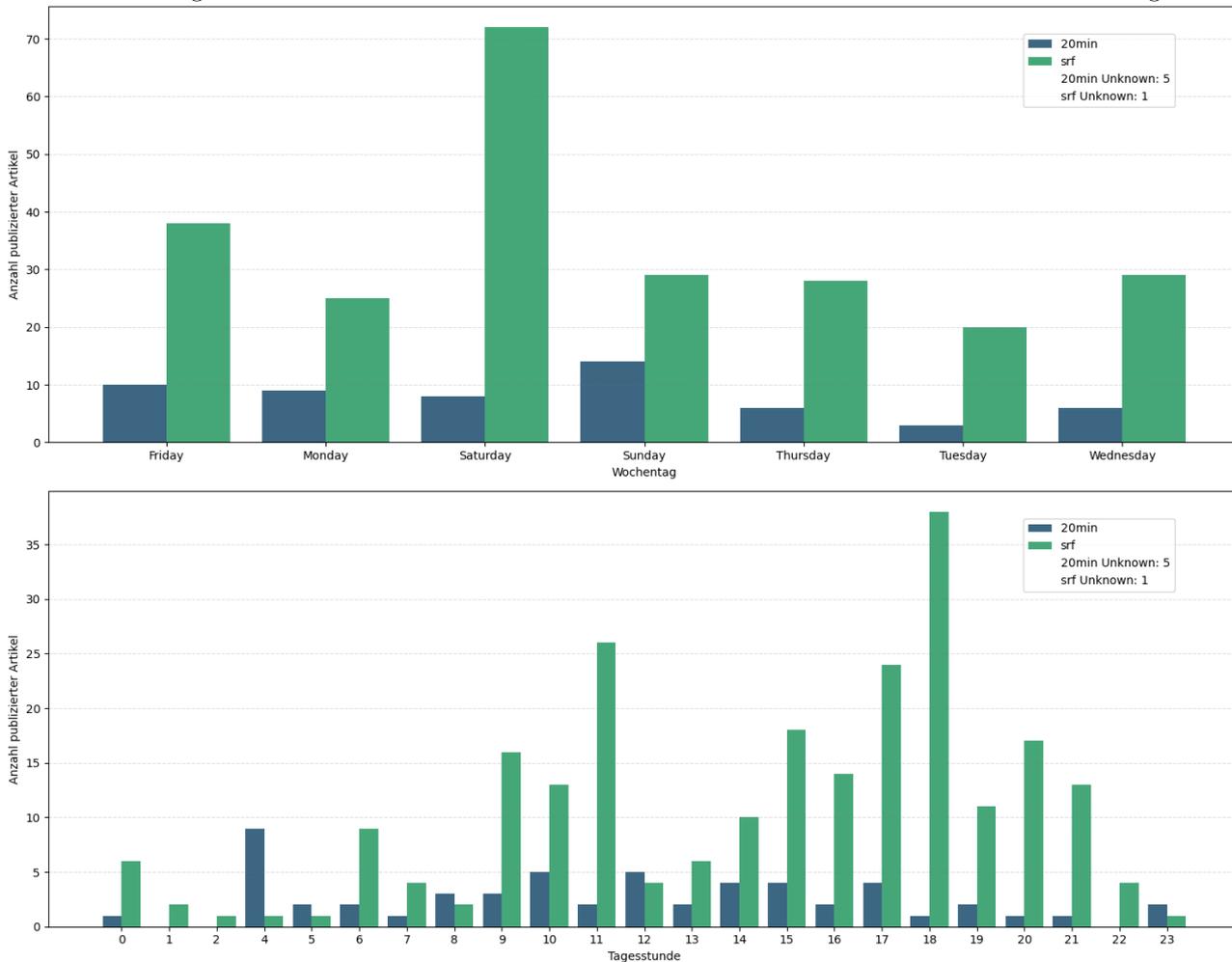


Abbildung 4: Publizierter Artikel zu den Wahlen vom 22. Oktober 2023

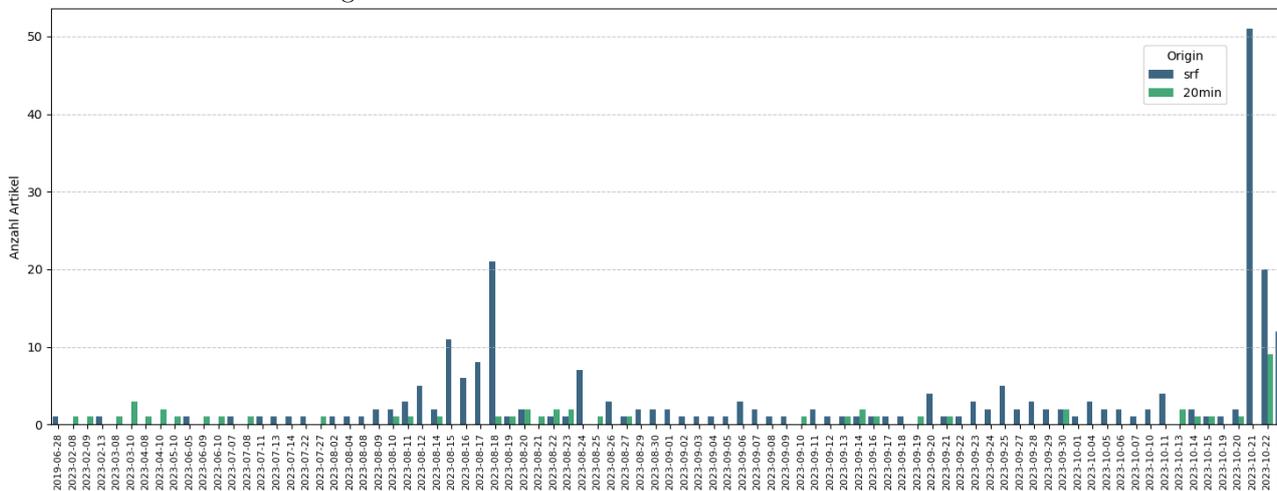


Abbildung 4 illustriert die Verteilung der veröffentlichten Artikel zu den Wahlen am 22. Oktober. Auffallend ist, dass SRF.ch bereits früh mit der Berichterstattung begonnen hat. So wurde schon ein Artikel im Anschluss an die Wahlen 2019 für das Jahr 2023 vorbereitet. Über den Sommer hinweg folgten aus der Redaktion von SRF.ch weitere Artikel – insbesondere nach den Sommerferien und vor dem Versand der Wahlunterlagen im August wurden sehr viele Artikel veröffentlicht: Allein am 18. August waren es 20 Online-Artikel zu den Wahlen. Eine genauere Untersuchung zeigte, dass es sich hier um einzelne Artikel-Seiten für entsprechende Kantone handelte, die im Verlauf des Oktobers bzw. am Wahltag aktualisiert wurden.

Es ist interessant festzustellen, dass für die Berichterstattung im Hinblick auf die Wahlen 2023 über beide Newsportale hinweg eine Vielzahl an Autoren und Autorinnen verantwortlich waren, von denen die meisten lediglich einen Artikel veröffentlicht haben. Nur eine Minderheit der Autoren und Autorinnen hat, wie in Abbildung 5 dargestellt, regelmässig Artikel geschrieben und veröffentlicht. Des Weiteren wird ersichtlich, dass bei SRF.ch sowie 20min.ch einige Artikel ohne Autorennennung existieren. Das bedeutet, dass bei der automatisierten Suche nach Autorennamen diese bei diesen Artikelseiten nicht extrahiert werden konnten. Eine genauere Untersuchung hat ergeben, dass es sich hier bei SRF.ch häufig um Beitragsseiten handelt, die sich auf Sendungen der Tagesschau beziehen und die vermittelten Inhalte der Tagesschau verschriftlichen – jedoch ohne Angabe von Autoren und Autorinnen.

Ein letzter erwähnenswerter Aspekt im Rahmen der ersten explorativen Datenanalyse in Bezug auf Autoren und Autorinnen wird in Abbildung 6 dargestellt. Diese zeigt die durchschnittliche Anzahl an Kommentaren je Autor bzw. Autorin und verdeutlicht, dass nur eine Minderheit der Autoren und Autorinnen für den Grossteil der Leserkommentare verantwortlich ist.

Im nachfolgenden Kapitel werden die hier dargelegten ersten grafischen Eindrücke in Verbindung mit den Texten der Artikel und den Kommentaren der User gesetzt. Ziel ist es, zu analysieren, welche Schlussfolgerungen im Hinblick auf die eingangs formulierten Hypothesen gezogen werden können.

Abbildung 5: Veröffentlichte Artikel der einzelnen AutorInnen im Zeitverlauf (ab Juli 2023)

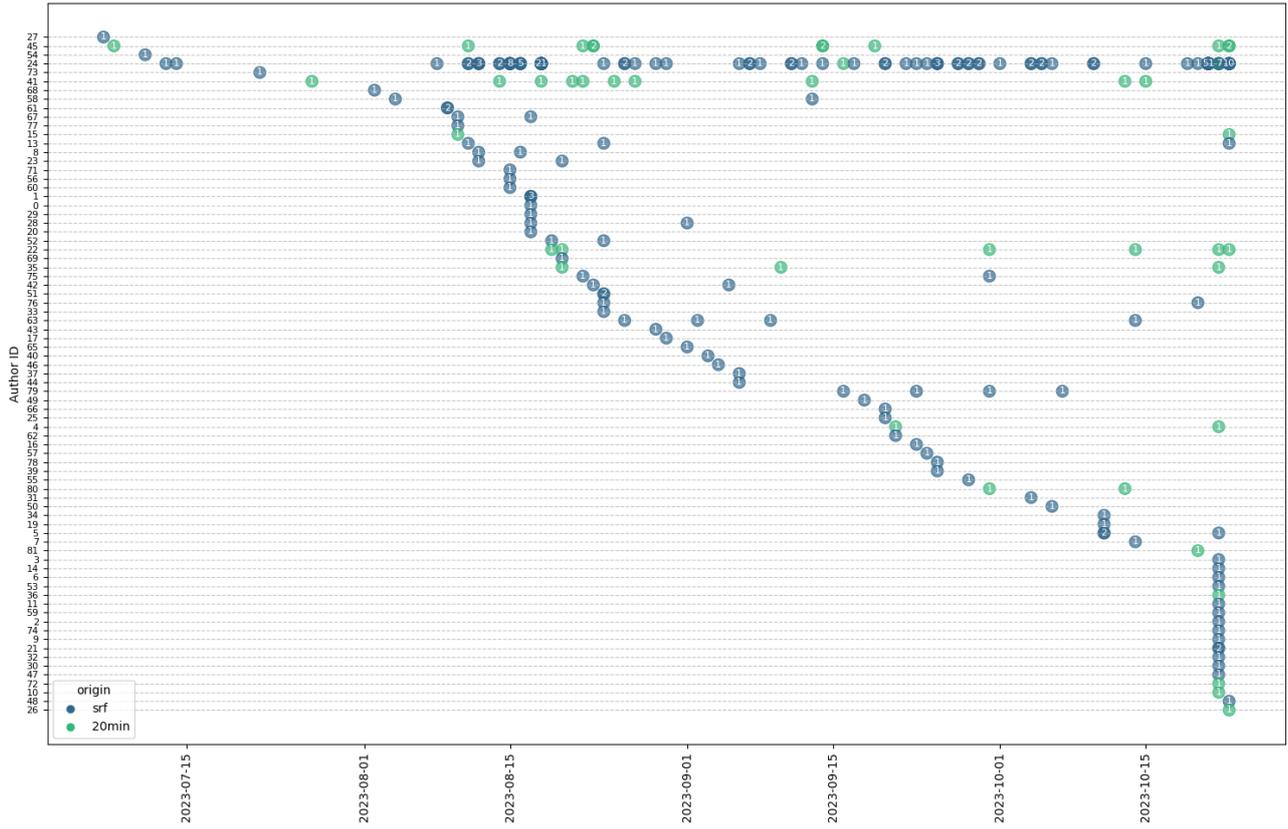
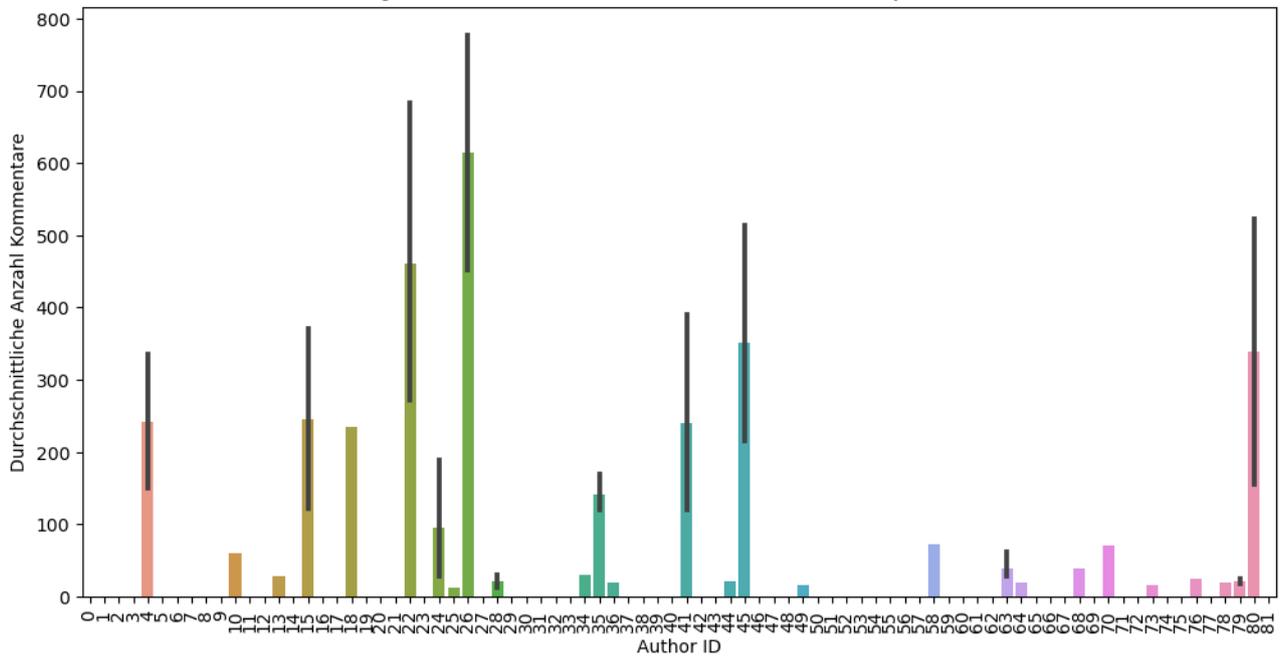


Abbildung 6: Durchschnittliche Anzahl Kommentare je AutorIn

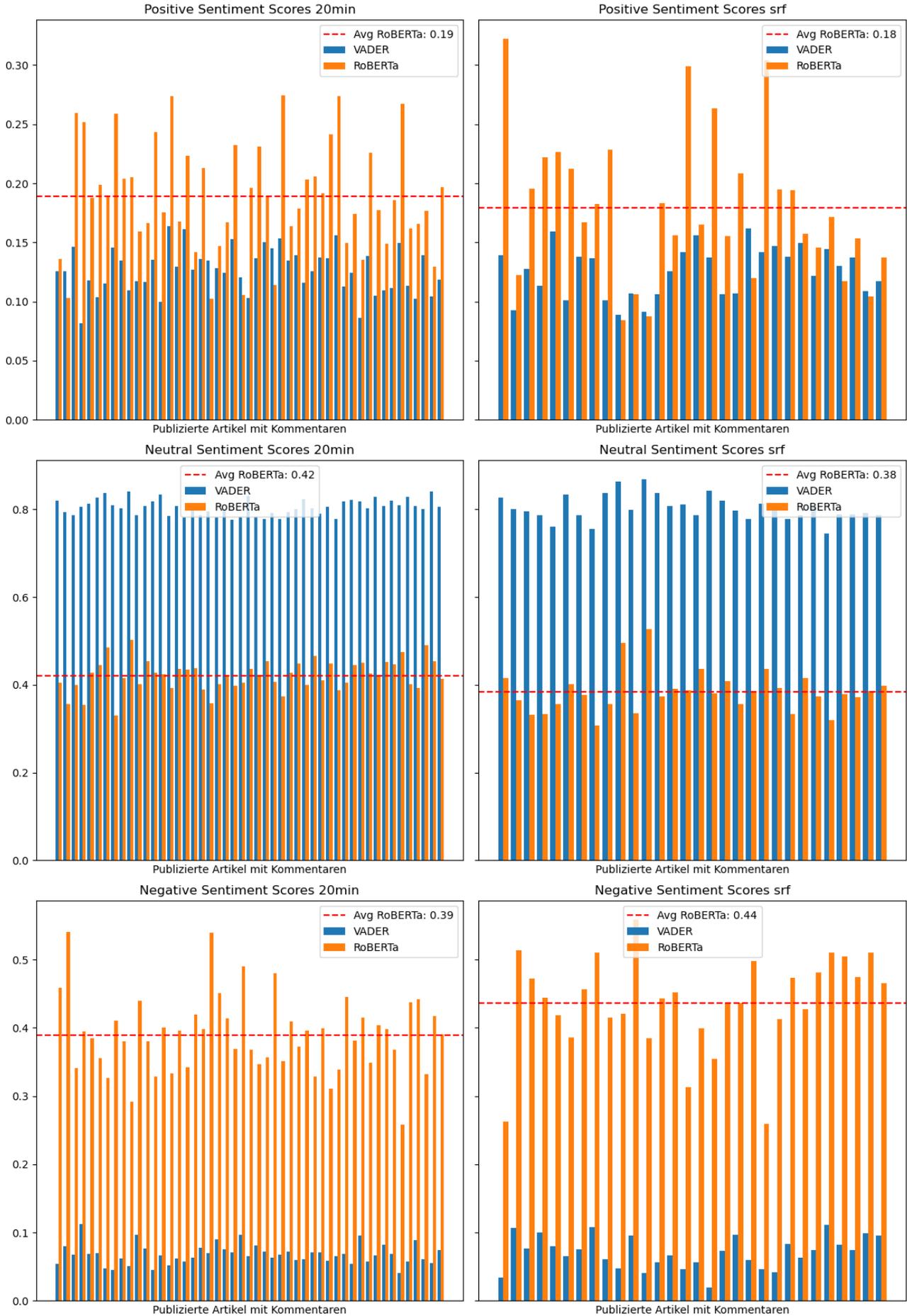


### 3 Ergebnisse

Um die Hypothesen H1 und H2 überprüfen zu können, wurde die Tonalität der veröffentlichten Texte detailliert analysiert. Für diesen Zweck wurden die automatisiert extrahierten Artikeltexte sowie die Kommentare von Leserinnen und Lesern ins Englische übersetzt. Anschliessend wurden sie mithilfe von bewerteten Wörterbüchern und Transformationsmodellen aus Python-Bibliotheken wie NLTK/VADER und RoBERTa analysiert (siehe Kapitel 2.1). Mit dieser Sentimentanalyse lässt sich ableiten, welche Haltung – positiv, negativ oder neutral – in den Artikeln und den Leserkommentaren zu den Wahlen vorherrscht.

Abbildung 7 zeigt die durchschnittlichen Sentiment-Scores (positiv, neutral, negativ) der Leserkommentare je Artikel von 20min.ch (links) und SRF.ch (rechts). Die blau dargestellten Werte repräsentieren die durchschnittlichen VADER-Scores, während die RoBERTa-Scores in Orange dargestellt sind. Die rot gestrichelte Linie markiert den durchschnittlichen RoBERTa Score über alle Artikel hinweg. Auffällig ist, dass VADER und RoBERTa nicht immer übereinstimmen. Bei positiven oder neutralen Sentiments sind die VADER-Scores oft höher, während sie bei negativen Sentiments niedriger sind als die RoBERTa-Werte. Dies könnte darauf hindeuten, dass VADER positive oder neutrale Ausdrücke stärker betont, während RoBERTa dank eines tieferen Kontextverständnisses negative Aspekte intensiver wahrnimmt. Es sollte jedoch betont werden, dass weder VADER noch RoBERTa als absolut korrekt angesehen werden sollten. Die Sentiment-Scores der Kommentare unterscheiden sich erheblich je nach Artikel, was darauf hindeutet, dass unterschiedliche Artikel verschiedene Emotionen in den Kommentaren hervorrufen. Faktoren wie Artikelinhalt, Zielgruppe oder Moderationsrichtlinien könnten diese Bewertungen beeinflussen. Interessant ist, dass die durchschnittlichen negativen Sentiment-Scores bei 20min.ch (0.39 basierend auf RoBERTa) etwas niedriger sind als bei SRF.ch (0.44). Das könnte bedeuten, dass die Kommentare von 20min.ch bezogen auf die Wahlen allgemein weniger negativ waren als jene von SRF.ch.

Abbildung 7: Gegenüberstellung der Stimmung in Kommentaren bei 20min.ch und SRF.ch



In der weiteren Analyse liegt der Fokus ausschliesslich auf den RoBERTa Scores. Dies zeigt auch Abbildung 8, die das Sentiment (negativ, positiv, neutral) der Leserkommentare beider Nachrichtenplattformen zeitlich gegenüberstellt. Die Grafik verdeutlicht, dass es sowohl bei 20min.ch (durchgestrichen) als auch bei SRF.ch (gestrichelt) im dargestellten Zeitraum deutliche Schwankungen in den Sentiment-Scores gibt. Dies ist nachvollziehbar, da sich die Stimmung in den Kommentarsektionen wohl je nach Nachrichtenlage änderte.

Weiterhin zeigen die negativen Kommentare sowohl für 20min.ch als auch für SRF.ch im Durchschnitt einen höheren Score als die positiven. Dies könnte mehrere Gründe haben. Ein Grund könnte sein, dass negative Meinungen oft mit stärkeren Worten ausgedrückt werden, was zu einem erhöhten Sentiment-Score führt. Direkt nach dem Wahlsonntag verzeichnet die grüne Linie (positive Kommentare von 20min) einen Anstieg, während die rote Linie (negative Kommentare von 20min) einen Rückgang zeigt. Im Gegensatz dazu zeigt die rote gestrichelte Linie (negative Kommentare von SRF) direkt nach dem 22. Oktober einen markanten Anstieg. Dies deutet darauf hin, dass die Negativität in den Kommentaren auf SRF.ch nach den Wahlen gestiegen ist. Möglicherweise spiegelt dies die allgemeine Stimmung oder die Wahlergebnisse wider. So ging die SVP als grosse Gewinnerin aus den Wahlen hervor. Eine darauf aufbauende Hypothese könnte sein, dass die kommentierenden Leserinnen und Leser bei 20min generell mit den Wahlergebnissen zufriedener waren – im Gegensatz zu jenen, die auf SRF.ch kommentierten.

Abbildung 8: Vergleich der durchschnittlichen RoBERTa Sentiment-Scores der veröffentlichten Kommentare im

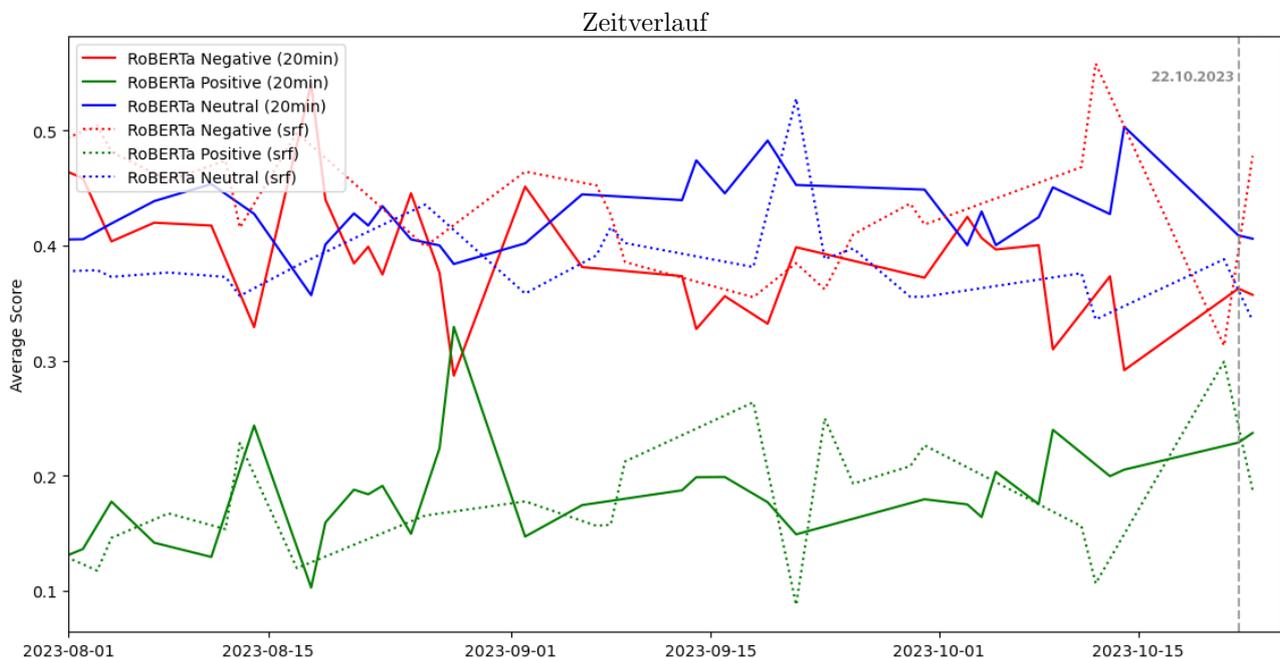
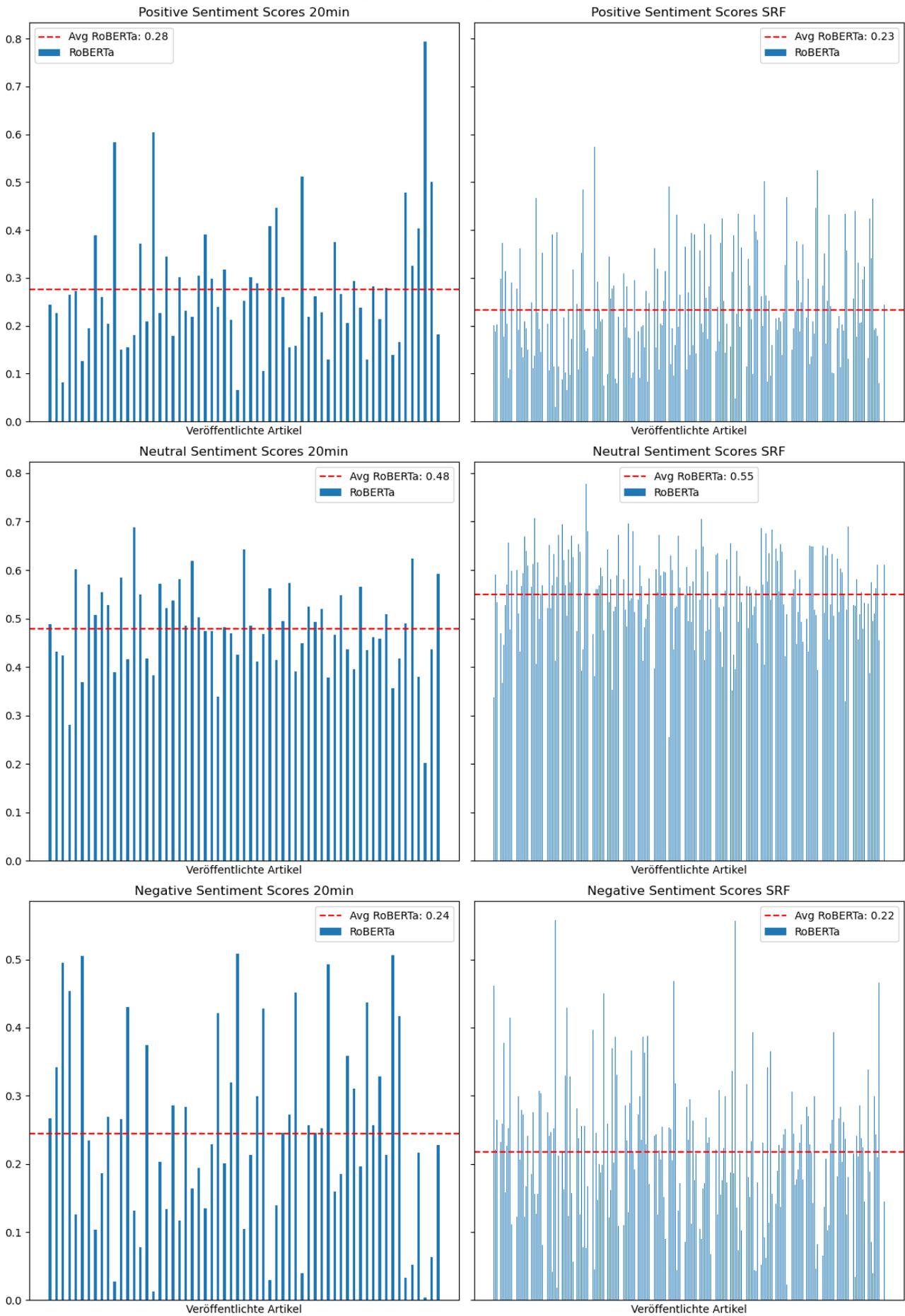


Abbildung 9: Gegenüberstellung der Stimmung in Artikeln bei 20min.ch und SRF.ch



Die Tonalität der Artikeltexte selbst wurde ebenfalls analysiert. Falls sich die beiden Medien in ihrer Berichterstattung zu den Wahlen im Oktober 2023 neutral verhalten haben, sollte sich dies in den Sentiment-Scores der Artikeltexte widerspiegeln. Ein neutraler Sentiment-Score deutet darauf hin, dass der Artikel insgesamt weder eine positive noch negative Stimmung vermittelt. Doch “neutral” in der Tonalität bedeutet nicht zwangsläufig “objektiv”. Eine detaillierte qualitative Analyse der Artikeltexte bezüglich ihrer Objektivität wäre daher empfehlenswert. Diese Arbeit konzentriert sich jedoch ausschliesslich auf die Tonalität.

Abbildung 9 zeigt die durchschnittlichen Sentiment-Scores von RoBERTa für die Artikeltexte – für 20min.ch (links in Abbildung) und SRF.ch (rechts in Abbildung). Die durchschnittlichen positiven Sentiment-Scores liegen leicht über 0.2. Die neutralen Scores sind deutlich höher und liegen bei RoBERTa für 20min.ch im Durchschnitt bei 0.48 und für srf.ch bei 0.55. Bei beiden Medienquellen liegen die negativen Scores für die gesamte Berichterstattung über die Wahlen unter 0.3.

Es fällt auf, dass die positiven Scores der Artikeltexte im Vergleich zu den positiven Scores der Leserkommentare höher ausfallen. Das könnte bedeuten, dass die Artikel tendenziell positiver gestimmt sind als die Feedbacks der Leserschaft. Die neutralen Scores der Artikeltexte sind signifikant höher. Dies könnte darauf hinweisen, dass die Kommentare der Leserschaft eine stärkere Meinungsfärbung aufweisen, während die Artikel im Vergleich dazu objektiver verfasst sind. Die negativen Scores sind in den Kommentaren höher als in den Artikeln. Dies könnte darauf hindeuten, dass die negativen Meinungen oder Reaktionen in den Kommentaren nicht direkt in den Artikeln wiedergegeben werden, was wiederum ein Indikator für eine gewisse Objektivität der Artikel sein könnte.

Zusammengefasst zeigen die Ergebnisse der Sentiment-Analyse, ohne den konkreten Inhalt der Texte und Meinungen zu berücksichtigen, dass Artikel tendenziell neutraler oder positiver wirken, während Leserkommentare oft stärker gefärbte und häufig negative Meinungen reflektieren. Das Ziel dieser Sentiment-Analyse war die Überprüfung folgender Hypothesen:

H1: Die Berichterstattung der untersuchten Online-Medien (SRF.ch und 20min.ch) im Kontext der Parlamentswahlen 2023 zeigt eine neutrale Ausrichtung in Bezug auf Tonalität.

H2: Kommentare von Leser und Leserinnen zu Beiträgen über die Wahlen 2023 zeichnen sich durch eine grössere Meinungsvielfalt und emotionalere Tonalität aus als die zugrunde liegenden Artikel. Abhängig vom jeweiligen Nachrichtenmedium und aktuellen Geschehnissen können die Intensität und Richtung der geäusserten Meinungen variieren.

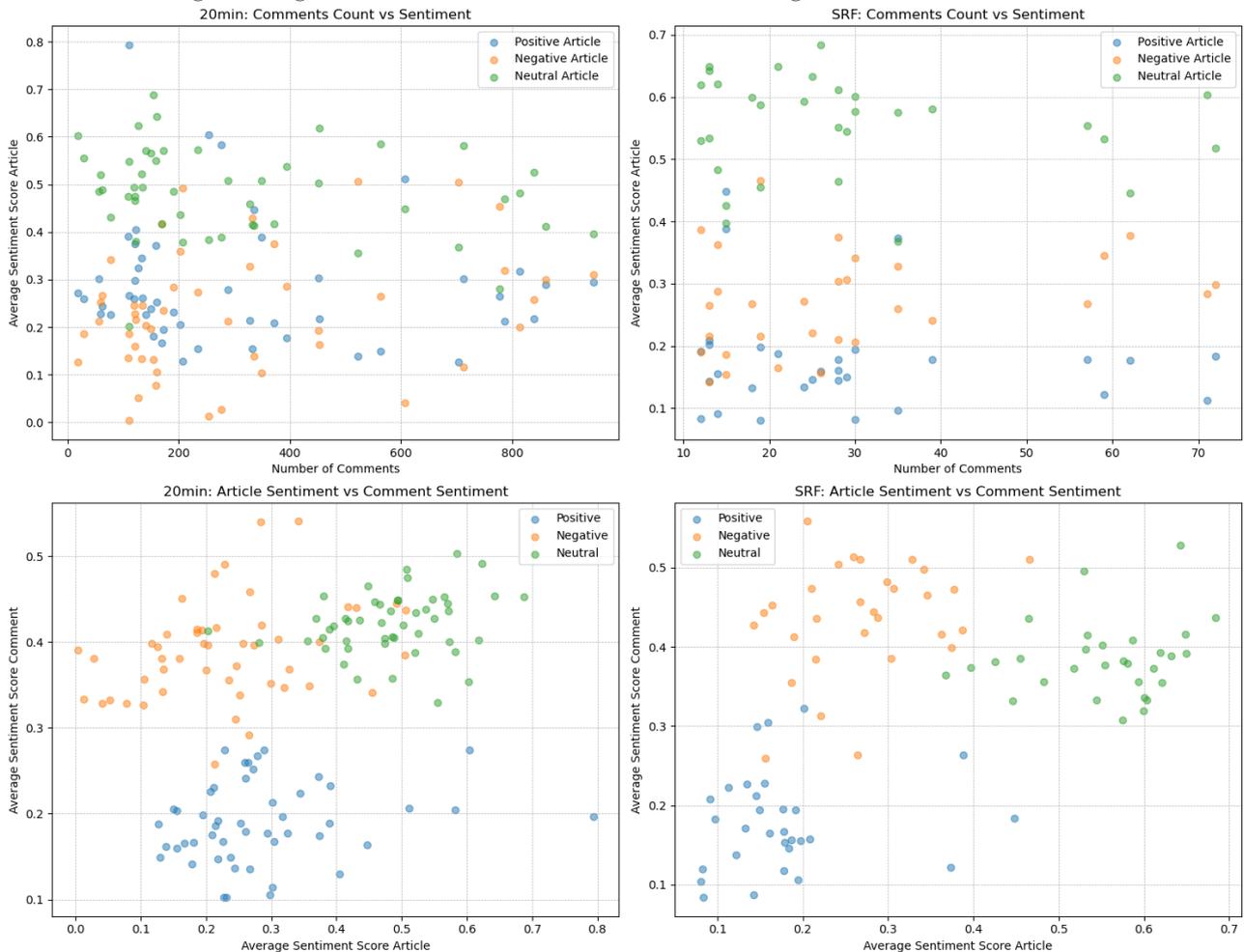
Basierend auf den Analysen lassen sich H1 und H2 teilweise bestätigen. Beide Medien weisen in vielen ihrer Artikel deutlich neutrale Sentiment-Scores auf, was auf eine tendenzielle Neutralität schliessen lässt. Dennoch gibt es auch Artikel mit ausgeprägt positiven oder negativen Scores, sodass man nicht von einer konstanten Neutralität in der Tonalität sprechen kann. Eine nachfolgende Überprüfung der Artikel auf ihre Objektivität wäre sinnvoll. Die Daten unterstützen Hypothese

H2. Die Sentiment-Scores der Kommentare zeigen eine grössere Varianz und stärkere Ausschläge im Vergleich zu den Scores der Artikel. Dies deutet darauf hin, dass Leserinnen und Leser in ihren Kommentaren eine breite Palette an Meinungen und Emotionen zum Ausdruck bringen. Zudem unterscheiden sich die beiden Medien hinsichtlich des durchschnittlichen Sentiment-Scores, was nahelegt, dass Nutzermeinungen je nach Plattform variieren können. Eine detailliertere Analyse wäre nötig, um festzustellen, inwiefern aktuelle Nachrichten diese Meinungen beeinflussen.

Im zugehörigen Programmcode finden sich zusätzliche Analysen und Abbildungen. Unter anderem wurde versucht, die Anzahl der Kommentare in Relation zum durchschnittlichen Sentiment-Score der Artikeltexte zu setzen. Dabei ging es um die Frage, ob besonders positive oder negative Artikeltexte zu einer erhöhten Anzahl an Leserkomentaren führen. Eine erste grafische Interpretation deutete jedoch nicht auf einen linearen Zusammenhang hin (siehe Abbildung 10). Ebenso zeigte sich keine klare Korrelation zwischen dem durchschnittlichen Sentiment-Score der Kommentare und dem der Artikel.

Des Weiteren wurde auf eine tiefgehende Untersuchung der Sentiment-Scores nach einzelnen Autorinnen und Autoren verzichtet. Ursprünglich stand die Frage im Raum, ob bestimmte Autorinnen oder Autoren hinsichtlich ihrer Tonalität einen stärkeren Einfluss aufweisen als ihre Kolleginnen und Kollegen. Allerdings ergab die explorative Datenanalyse, dass die meisten von ihnen lediglich einen Artikel im Kontext der Wahlen verfasst hatten. Eine sinnvolle Analyse, ob spezifische Autorinnen oder Autoren massgeblich die Richtung der Sentiment-Scores beeinflussen, wäre erst bei mehreren Publikationen pro Person möglich. Bei nur einem Artikel pro Autorin oder Autor könnten festgestellte Unterschiede in der Tonalität eher auf den Kontext des jeweiligen Artikels als auf den individuellen Schreibstil der schreibenden Person zurückgeführt werden.

Abbildung 10: Vergleich der Artikel- und Kommentarstimmung zwischen 20min.ch und SRF.ch



## 4 Diskussion

Die durchgeführte Sentiment-Analyse hat einen ersten Einblick in die Art und Weise der Berichterstattung über die Wahlen vom Oktober 2023 zweier Online-Medien ermöglicht. 20min.ch sowie SRF.ch weisen eine sehr hohe Reichweite auf und tragen in der Berichterstattung von wichtigen Ergebnissen eine entsprechende Verantwortung. Die beobachteten, überwiegend neutralen Sentiment-Scores der Artikel weisen darauf hin, dass SRF.ch sowie 20min.ch bemüht sind/waren, möglichst ausgewogen über die Wahlen zu berichten. Dies ist von grösserer Bedeutung, da unparteiische Informationen die Meinungsbildung der Leserschaft verzerren kann. Die Leserkommentare zeigen, wie erwartet, eine grössere emotionale Bandbreite und Meinungsvielfalt. Dies spiegelt die persönlichen Überzeugungen, Gefühle und Reaktionen der Leserschaft wider und ist ein Indikator für ein engagiertes und reaktionäres Publikum.

Abbildungen 11, 12 sind sogenannte “Word-Clouds”, welche exemplarisch die meist genannten Wörter der Kommentare mit sehr hohen positiven (11) wie auch negativen (12) RoBERTa Scores aufzeigen.



## 5 Fazit

Die Schweizer Parlamentswahlen von 2023 und die darauffolgende Medienberichterstattung zogen die Aufmerksamkeit zahlreicher Bürgerinnen und Bürger auf sich. Im Rahmen dieser Studienarbeit basieren die Analysen auf Webscraping und Text Mining von Online-Beiträgen der Medienportale [www.srf.ch](http://www.srf.ch) und [www.20min.ch](http://www.20min.ch), mit dem Ziel, das Medienecho rund um die Wahlen in Bezug auf die Tonalität zu bewerten.

Die Ergebnisse dieser Arbeit stützen teilweise die Befunde des Forschungszentrums Öffentlichkeit und Gesellschaft der Universität Zürich. Es wird angenommen, dass viele Medien im politischen Zentrum angesiedelt sind. Dies konnte in meinen Analysen bis zu einem gewissen Grad bestätigt werden, denn zwei der grossen Plattformen strebten eine neutrale Berichterstattung an. Ein Vorbehalt ist hier jedoch angebracht, da der eigentliche Inhalt der Texte nicht auf Objektivität geprüft wurde. Zukünftige Untersuchungen könnten die Texte indexieren, um die Ausgewogenheit in der Erwähnung einzelner Kandidaten oder Parteien zu überprüfen. Der Fokus dieser Arbeit lag einzig auf der Tonalität, die anhand von Sentiment-Scores zweier Modelle ermittelt wurde.

Die Analyseergebnisse dieser Studienarbeit bestätigen Hypothese H1 teilweise, indem sie zeigen, dass die Berichterstattung der untersuchten Online-Medien (SRF.ch und 20min.ch) im Allgemeinen eine neutrale Tonalität im Kontext der Parlamentswahlen 2023 aufweist. Weiter zeigt eine Untersuchung von Nutzerkommentaren, dass Hypothese H2 bestätigt werden kann. Die Kommentare von Lesern und Leserinnen zu den Beiträgen über die Wahlen 2023 zeichnen sich deutlich durch eine grössere Meinungsvielfalt und emotionalere Tonalität aus als die eigentlichen Artikel. Es unterstreicht die signifikante Rolle und den Einfluss von sozialen Medien und Online-Kommentaren in der aktuellen politischen Diskussionskultur. Während die Artikel versuchen, Neutralität und Ausgewogenheit zu wahren, geben die Kommentare der Leserschaft oft ein unverfälschteres Bild der öffentlichen Meinung wider, mit all ihrer Vielfalt und Emotionalität.

In der aktuellen Ära, in der das digitale Zeitalter Informationsverbreitung und -konsum revolutioniert, ist die Bedeutung von Online-Medien in der politischen Landschaft der Schweiz nicht zu unterschätzen. Politikerinnen, Politiker, Journalistinnen und Journalisten müssen die Notwendigkeit von Transparenz und Ausgewogenheit in der Berichterstattung anerkennen.

Abschliessend sei betont, dass eine fortlaufende Überwachung und Analyse der Medienberichterstattung entscheidend ist, um sicherzustellen, dass die Öffentlichkeit gut informiert ist und fundierte Entscheidungen treffen kann. Mit dieser Studienarbeit soll vor allem das Verständnis für die Rolle der Medien in der Schweizer Demokratie vertieft und Inspiration für zukünftige Forschungen in diesem Bereich gegeben werden.

## 6 Quellenverzeichnis

- [1] Baccouri, Nidhal. n.d. “Deep-Translator: A Flexible Free and Unlimited Python Tool to Translate between Different Languages in a Simple Way Using Multiple Translators.” OS Independent. Python. Accessed October 24, 2023. [https://github.com/nidhaloff/deep\\_translator](https://github.com/nidhaloff/deep_translator).
- [2] Bundeskanzlei. 2023. “Das Informationsangebot der Behörden.” ch.ch — Wahlen 2023. Accessed October 29, 2023. <https://www.ch.ch/de/wahlen2023/organisation-der-wahlen/das-informationsangebot-der-behorden/#die-wahlanleitung>.
- [3] DiBattista, Jeremy. 2021. “The Best Python Sentiment Analysis Package (+1 Huge Common Mistake).” Medium. August 27, 2021. <https://towardsdatascience.com/the-best-python-sentiment-analysis-package-1-huge-common-mistake-d6da9ad6cdeb>.
- [4] Haselmayer, Martin, and Marcelo Jenny. 2017. “Sentiment Analysis of Political Communication: Combining a Dictionary Approach with Crowdcoding.” *Quality & Quantity* 51 (6): 2623–46. <https://doi.org/10.1007/s11135-016-0412-4>.
- [5] Hutto, C., and Eric Gilbert. 2014. “VADER: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text.” *Proceedings of the International AAAI Conference on Web and Social Media* 8 (1): 216–25. <https://doi.org/10.1609/icwsm.v8i1.14550>.
- [6] Liu, Yinhan, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. “RoBERTa: A Robustly Optimized BERT Pretraining Approach.” arXiv. <https://doi.org/10.48550/arXiv.1907.11692>.
- [7] Media Relations - fög. 2023. “Die grossen Medien berichten politisch ausgewogen.” September 25, 2023. <http://www.news.uzh.ch/de/articles/media/2023/Medienstudie-foeg.html>.
- [8] Munzert, Simon. 2014. *Automated Data Collection with R: A Practical Guide to Web Scraping and Text Mining*. Chichester, West Sussex, United Kingdom: Wiley.
- [9] Richardson, Leonard. n.d. “Beautifulsoup4: Screen-Scraping Library.” Python. Accessed October 24, 2023. <https://www.crummy.com/software/BeautifulSoup/bs4/>.
- [10] “Selenium.” n.d. POSIX. Python. Accessed October 24, 2023. <https://www.selenium.dev>.

# A Appendix

## A.1 Code

Der für diese Arbeit verwendete Programmcode ist in vier Python Jupyter Notebooks dokumentiert und online verfügbar:

1 – Webscraping 20min.ch:

[https://www.simonsahli.ch/code/Wahlen23\\_1-Webscraping\\_20min.ipynb](https://www.simonsahli.ch/code/Wahlen23_1-Webscraping_20min.ipynb)

2 – Webscraping SRF.ch:

[https://www.simonsahli.ch/code/Wahlen23\\_2-Webscraping\\_srf.ipynb](https://www.simonsahli.ch/code/Wahlen23_2-Webscraping_srf.ipynb)

3 – Sentimentanalyse:

[https://www.simonsahli.ch/code/Wahlen23\\_3-Sentiment-analysis.ipynb](https://www.simonsahli.ch/code/Wahlen23_3-Sentiment-analysis.ipynb)

4 – EDA:

[https://www.simonsahli.ch/code/Wahlen23\\_4-EDA.ipynb](https://www.simonsahli.ch/code/Wahlen23_4-EDA.ipynb)

Um die Analysen dieser Arbeit vollständig nachzuvollziehen, sollten die Notebooks in der angegebenen Reihenfolge betrachtet werden.